US 20250263449A1

(19) United States
(12) Patent Application Publication
SCHEU et al.
(10) Pub. No.: US 2025/0263449 A1
(43) Pub. Date: Aug. 21, 2025

(54) **MULTIVALENT PROTEINS AND SCREENING METHODS**

(71) Applicant: **c/o VALINK THERAPEUTICS LTD**, London (GB)

(72) Inventors: **Arne Hagen August SCHEU**, Greater London (GB); **Irsyad Noor Abadi Bin KHAIRIL ANUAR**, Greater London (GB); **Ying Ting Sheryl LIM**, Greater London (GB)

(73) Assignee: **c/o VALINK THERAPEUTICS LTD**, London (GB)

(21) Appl. No.: **18/283,353**

(22) PCT Filed: **Mar. 24, 2022**

(86) PCT No.: **PCT/GB2022/050750**

§ 371 (c)(1),
(2) Date: **Sep. 21, 2023**

(30) **Foreign Application Priority Data**

Mar. 24, 2021 (GB) 2104104.1

Publication Classification

(51) **Int. Cl.**

C07K 14/315	(2006.01)
C07K 14/47	(2006.01)
C07K 14/525	(2006.01)
C07K 14/78	(2006.01)
C07K 16/46	(2006.01)
G01N 33/50	(2006.01)
G01N 33/68	(2006.01)

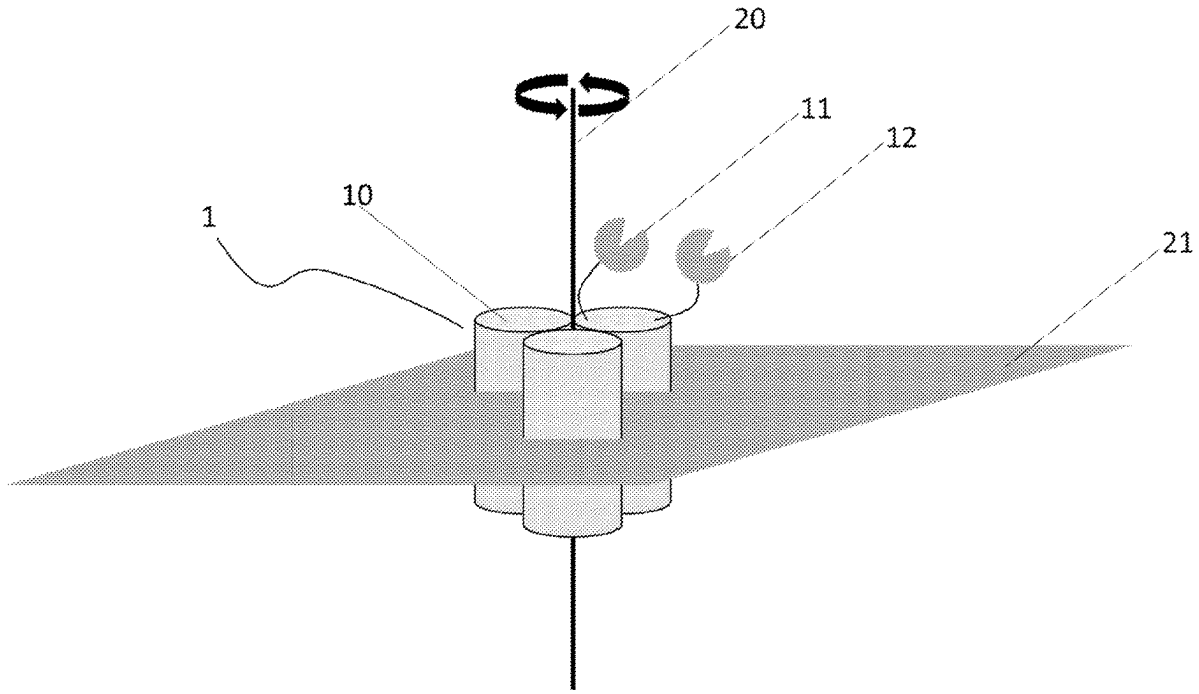
(52) **U.S. Cl.**

CPC ***C07K 14/315*** (2013.01); ***C07K 14/4703*** (2013.01); ***C07K 14/472*** (2013.01); ***C07K 14/525*** (2013.01); ***C07K 14/78*** (2013.01); ***C07K 16/46*** (2013.01); ***G01N 33/5008*** (2013.01); ***G01N 33/6845*** (2013.01); *C07K 2319/70* (2013.01); *G01N 2500/10* (2013.01)

(57) **ABSTRACT**

Provided herein are multivalent protein scaffolds useful as therapeutics, and useful in identifying new therapeutic compounds. The invention also relates to multi-domain polypeptide constructs having multiple binding domains and a structural domain. Also provided herein are methods of using the provided multivalent protein scaffolds to identify new candidate therapeutics, and new therapeutics thereby identified.

Specification includes a Sequence Listing.



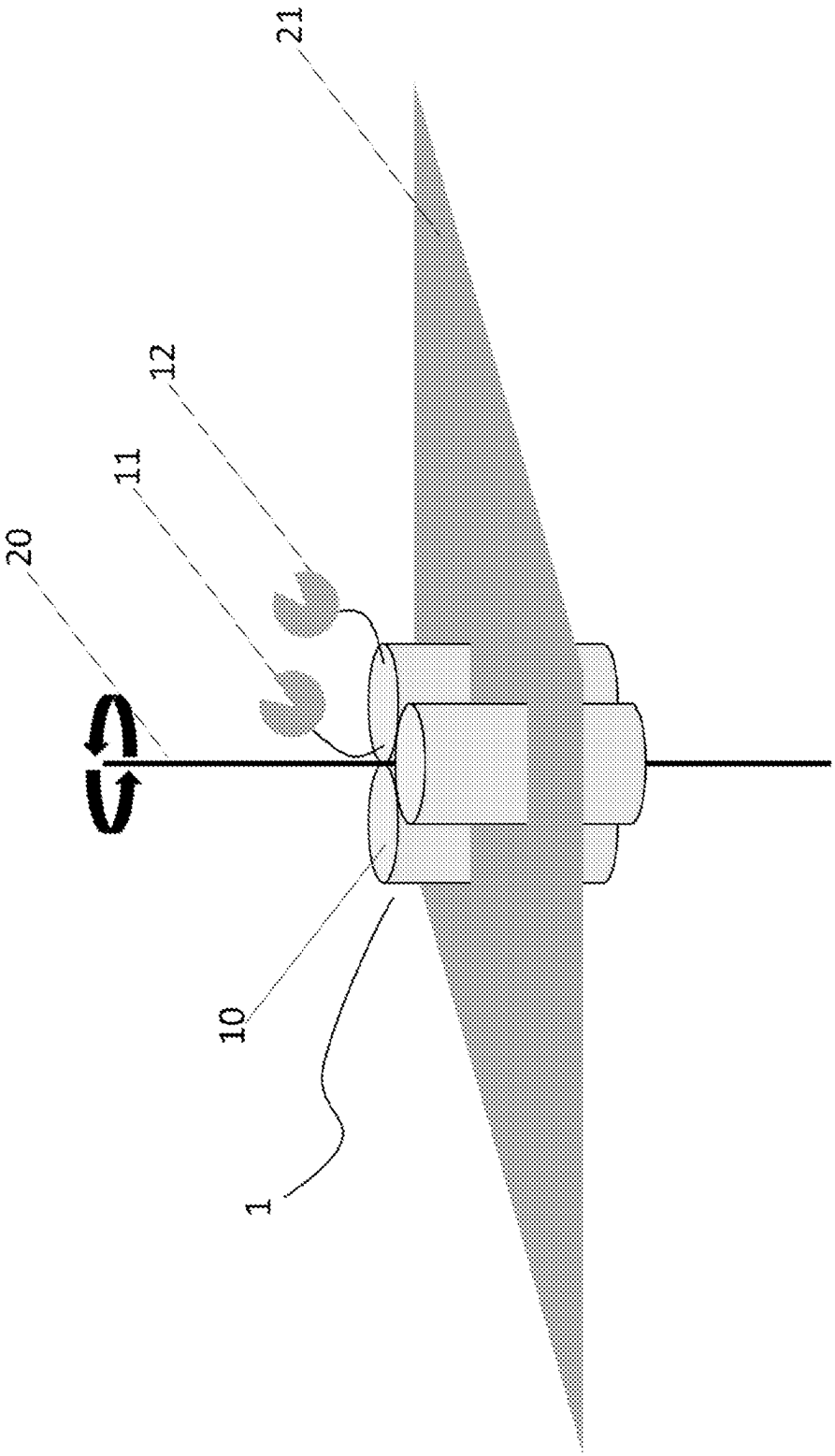


Figure 1

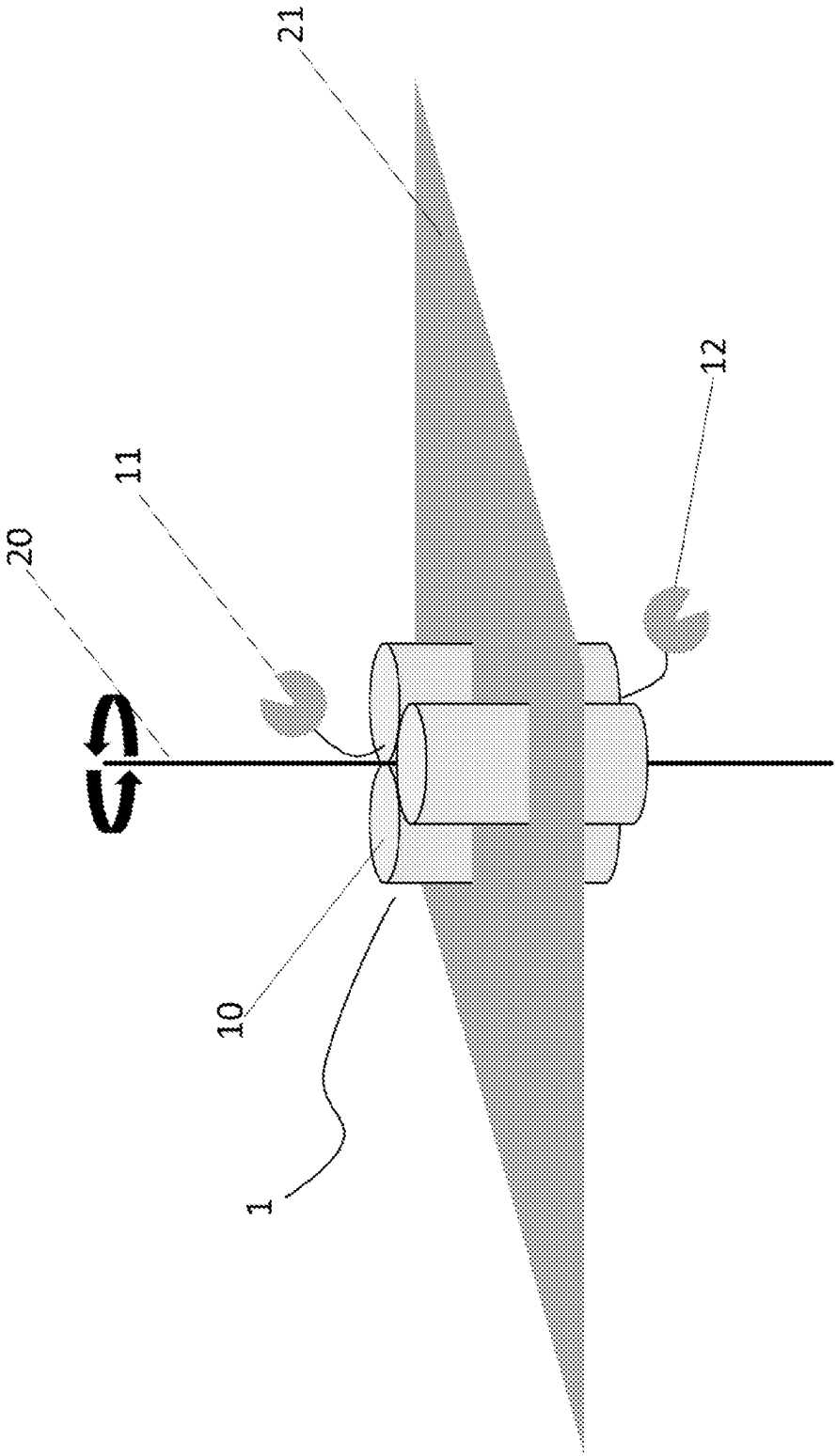


Figure 2

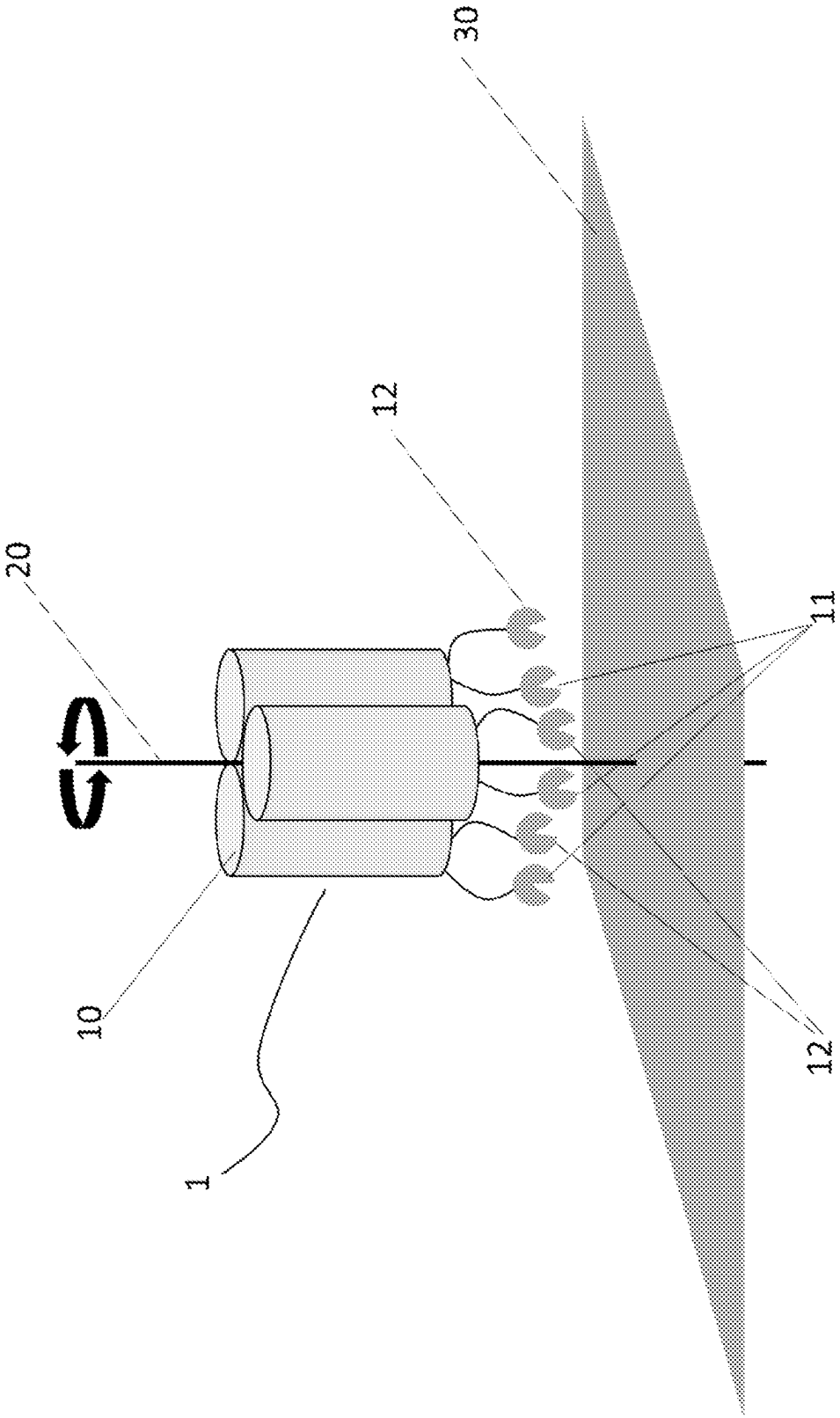


Figure 3

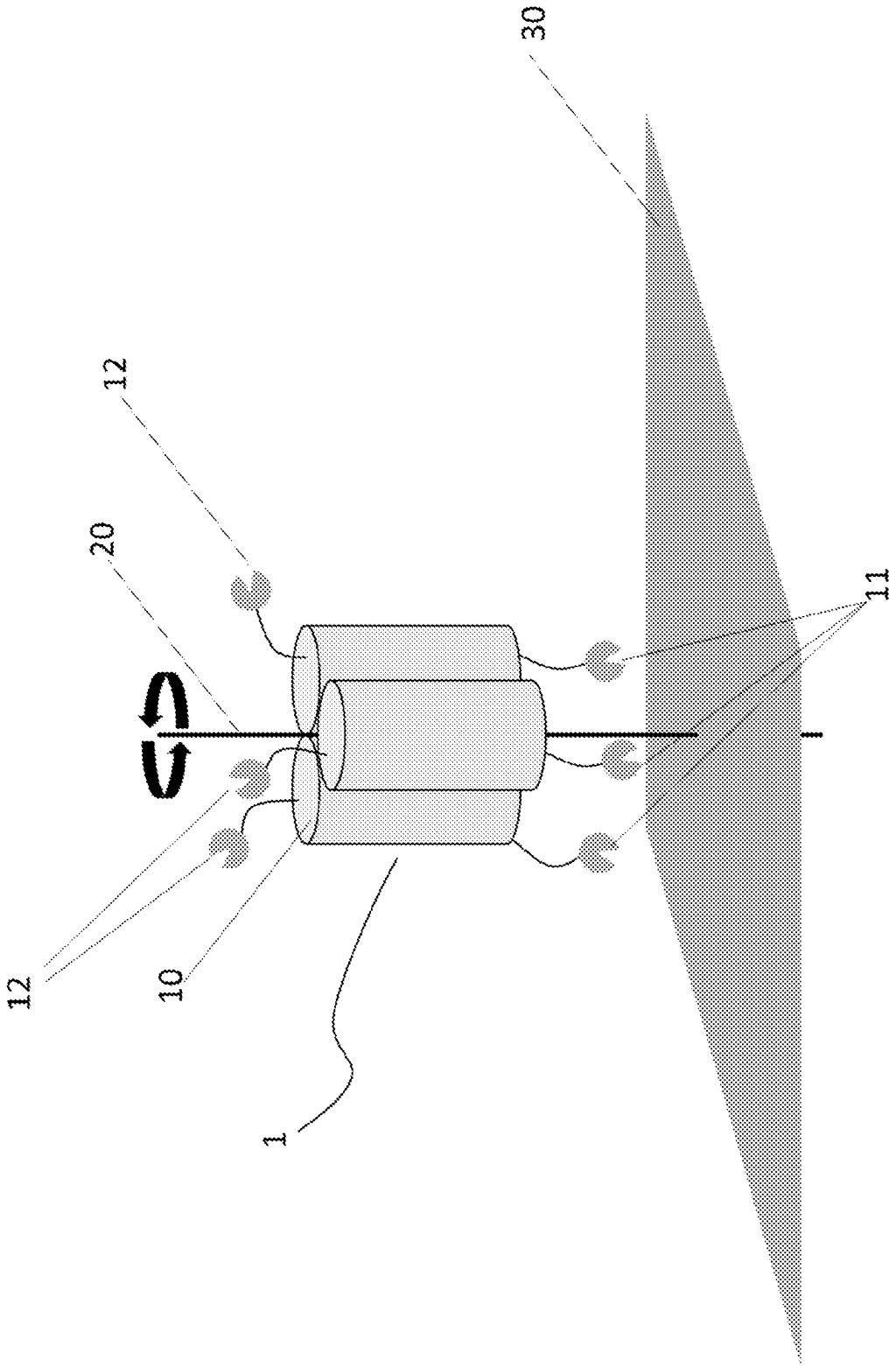


Figure 4

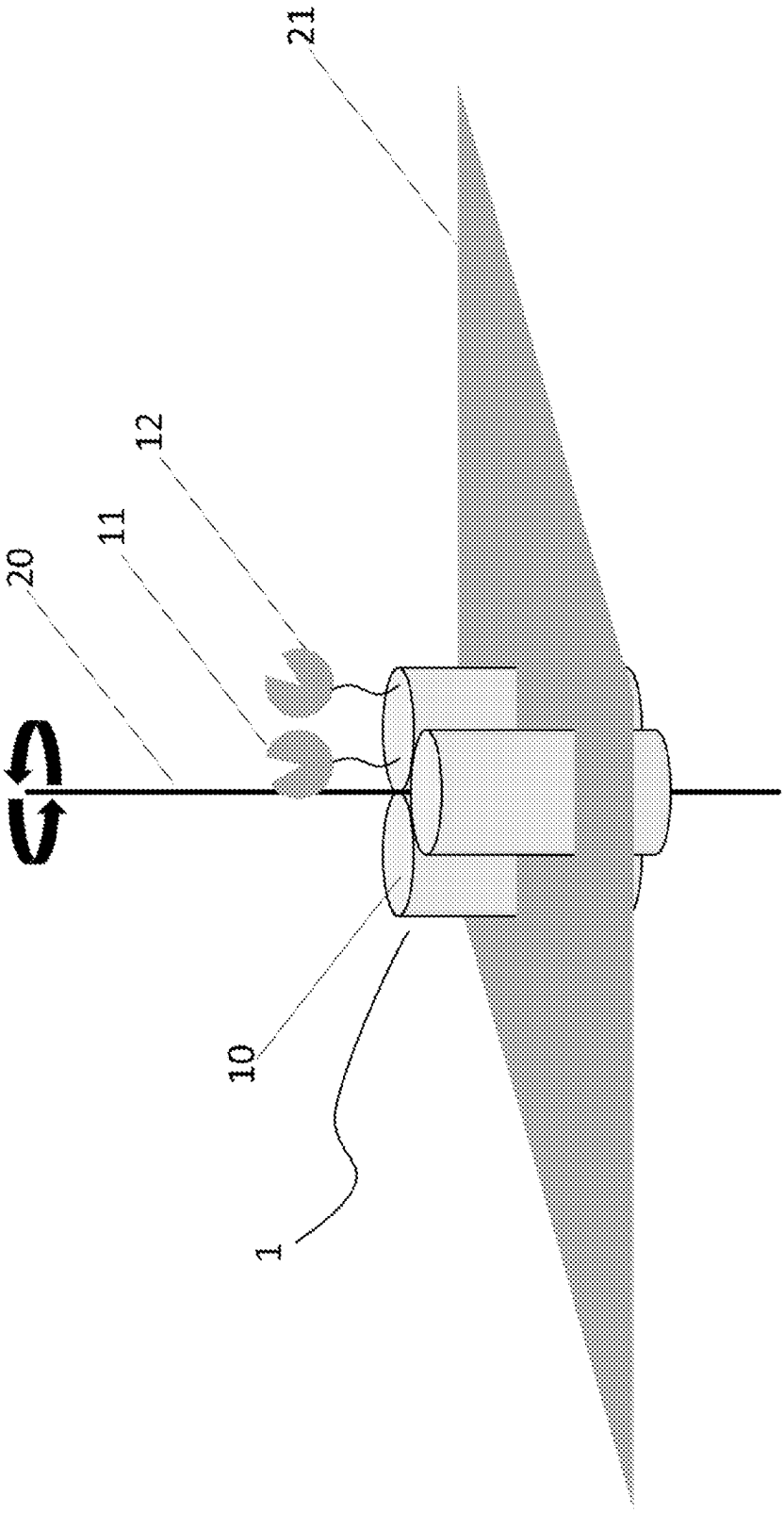


Figure 5

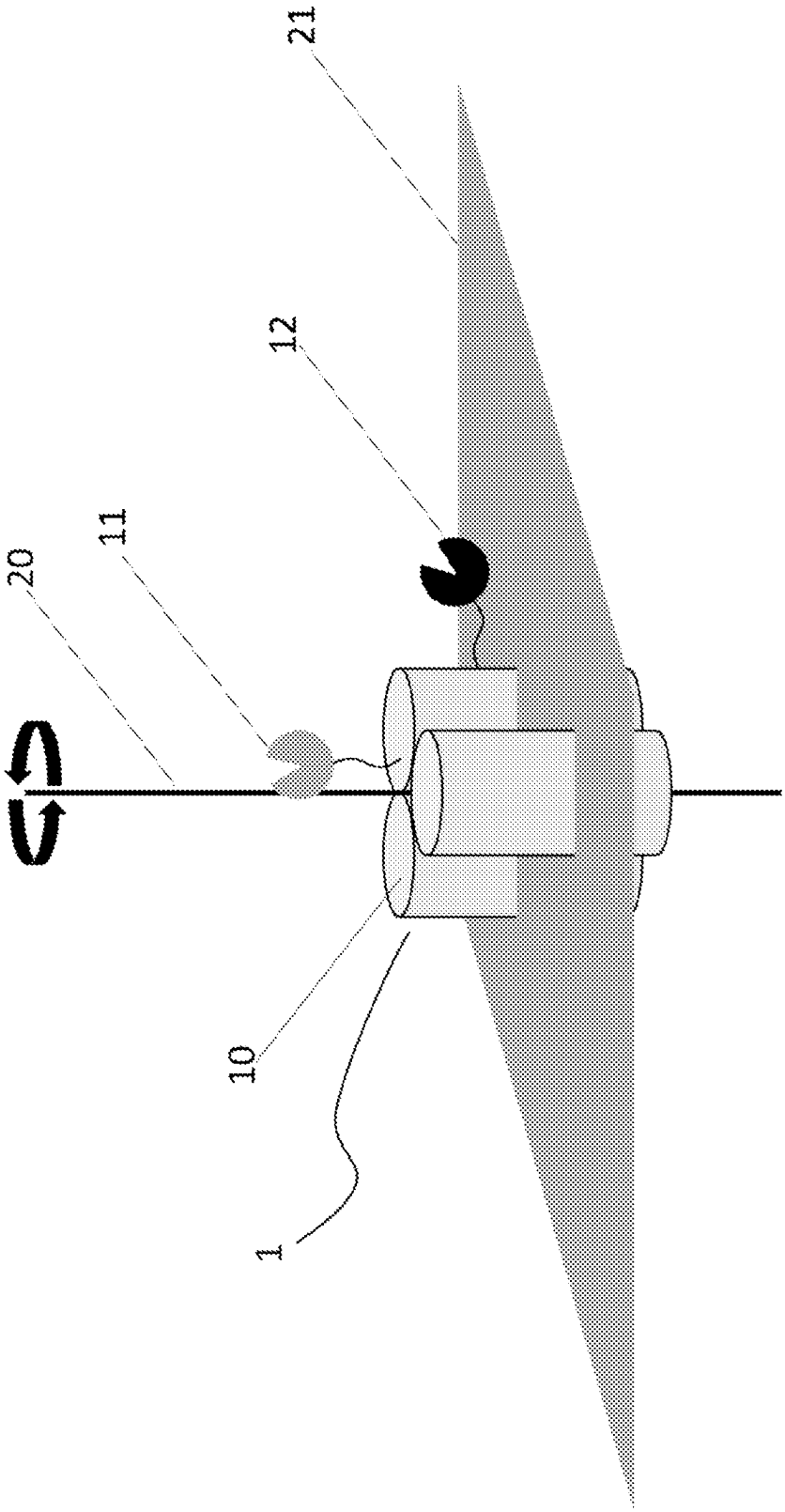


Figure 6

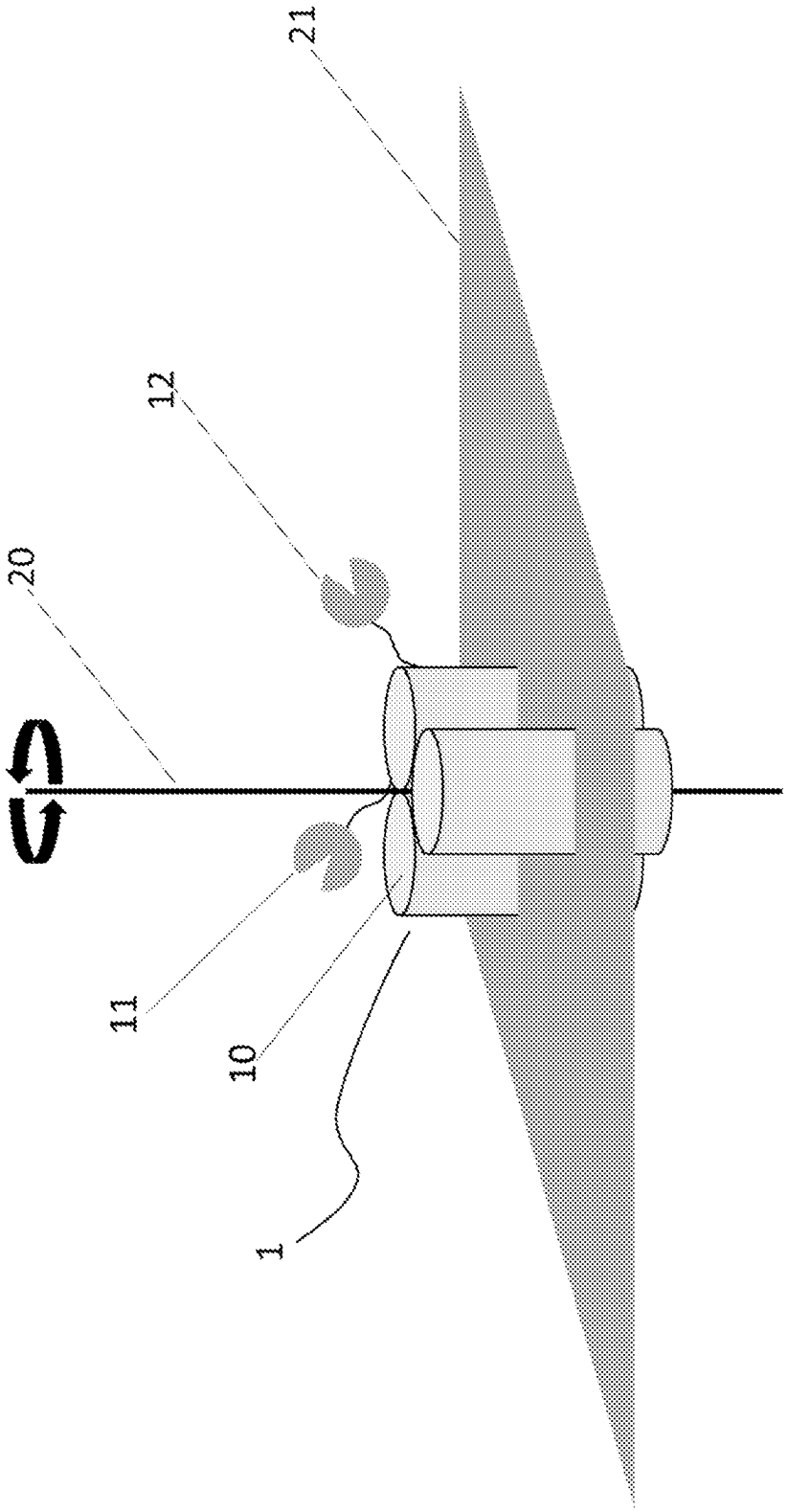


Figure 7

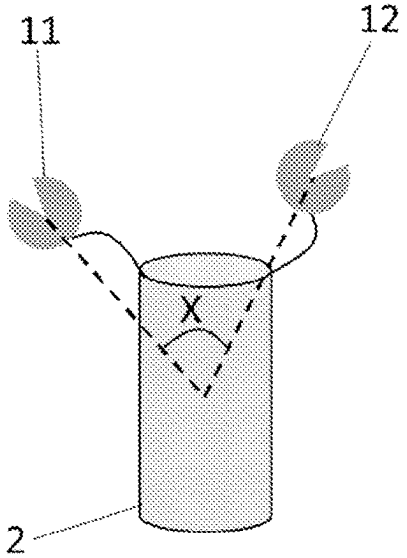


Figure 8

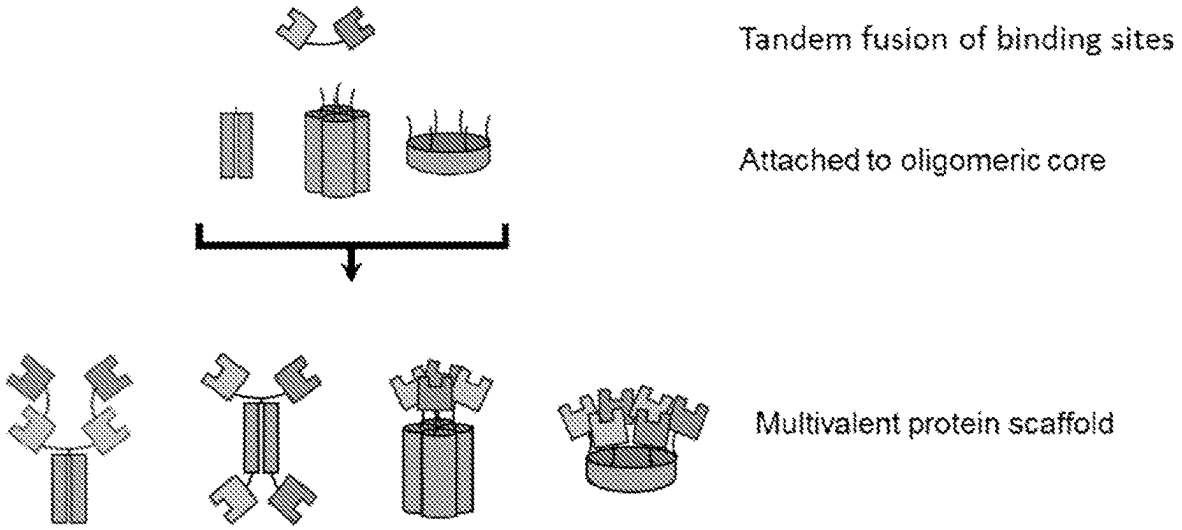


Figure 9

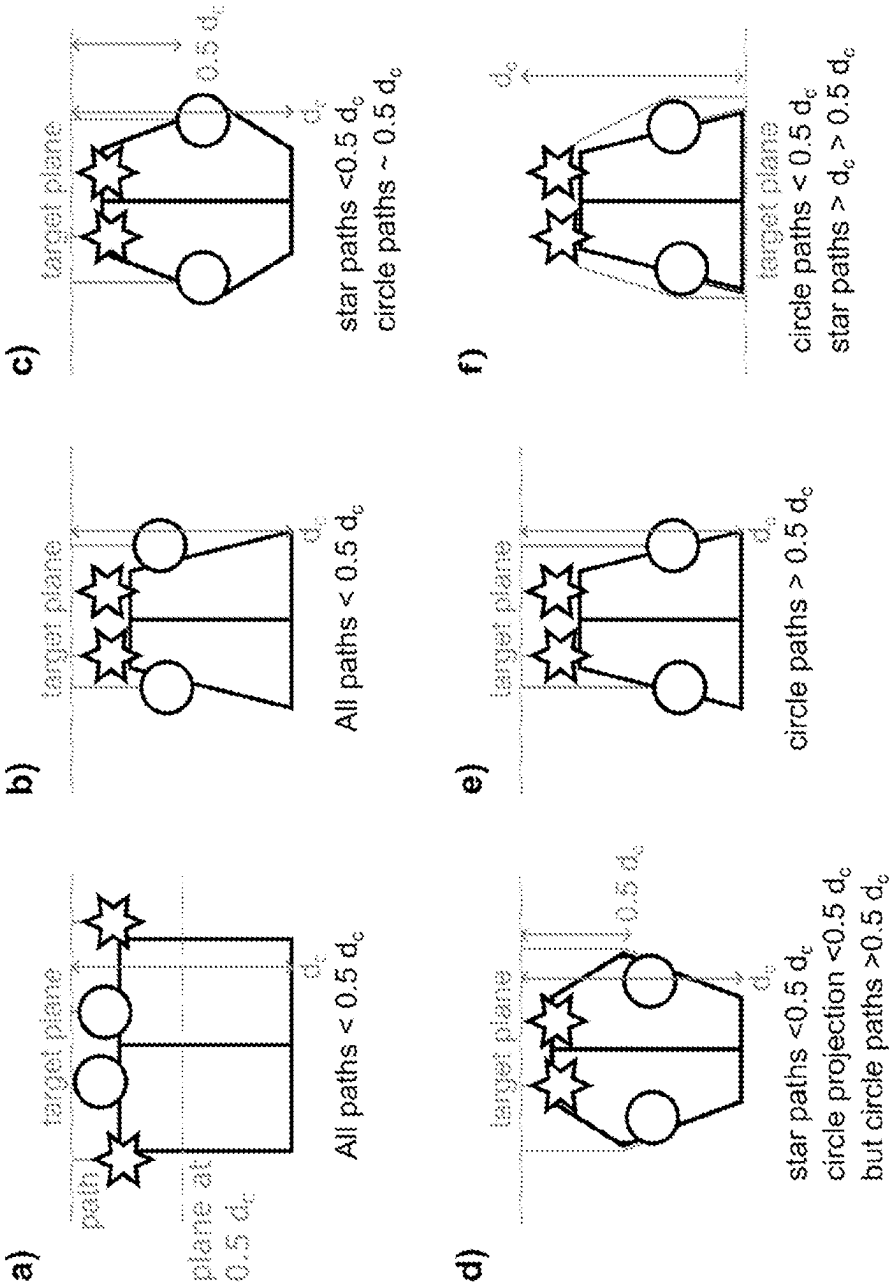

a)
target plane
path
plane at
0.5 d_c
d_c
All paths < 0.5 d_c
b)
target plane
d_c
All paths < 0.5 d_c
c)
target plane
0.5 d_c
d_c
star paths <0.5 d_c
circle paths ~ 0.5 d_c
d)
target plane
0.5 d_c
d_c
star paths <0.5 d_c
circle projection <0.5 d_c
but circle paths >0.5 d_c
e)
target plane
d_c
circle paths > 0.5 d_c
f)
d_c
target plane
circle paths < 0.5 d_c
star paths > d_c > 0.5 d_c

Figure 10

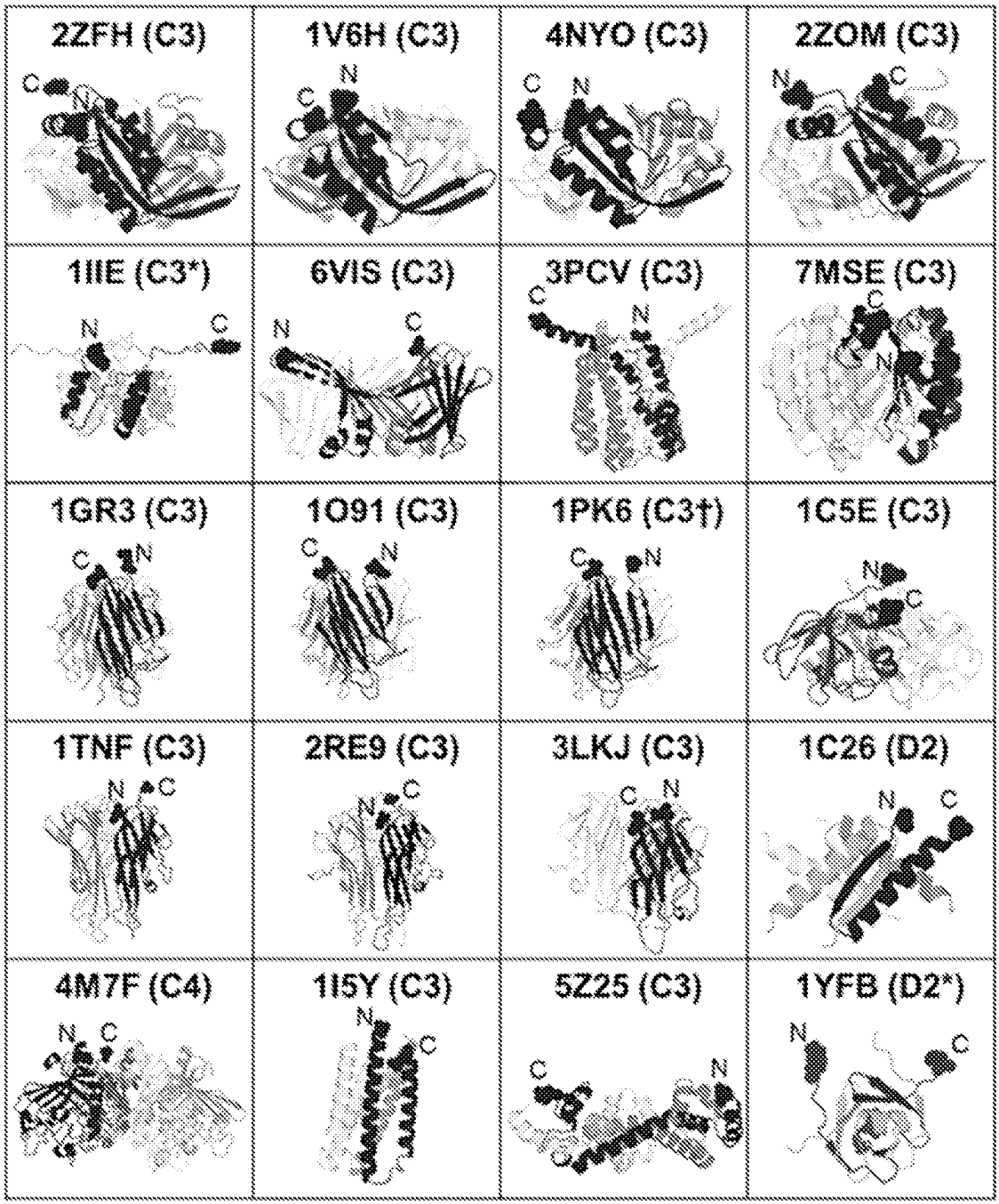

2ZFH (C3)
1V6H (C3)
4NYO (C3)
2ZOM (C3)
1IIE (C3*)
6VIS (C3)
3PCV (C3)
7MSE (C3)
1GR3 (C3)
1O91 (C3)
1PK6 (C3†)
1C5E (C3)
1TNF (C3)
2RE9 (C3)
3LKJ (C3)
1C26 (D2)
4M7F (C4)
1I5Y (C3)
5Z25 (C3)
1YFB (D2*)

Figure 11

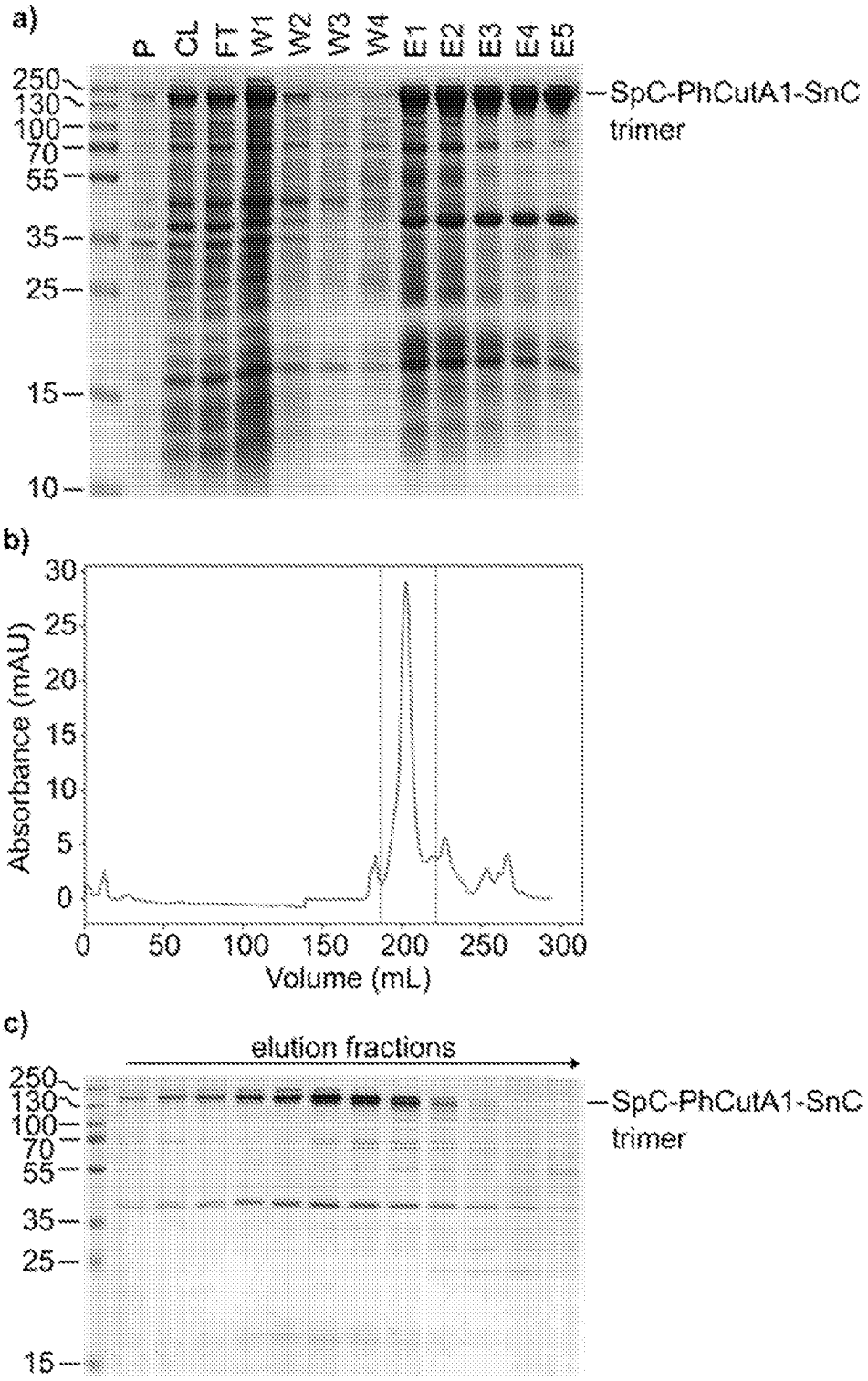

a)
P
CL
FT
W1
W2
W3
W4
E1
E2
E3
E4
E5
250
130
100
70
55
35
25
15
10
SpC-PhCutA1-SnC trimer
b)
Absorbance (mAU)
30
25
20
15
10
5
0
0
50
100
150
200
250
300
Volume (mL)
c)
elution fractions
250
130
100
70
55
35
25
15
SpC-PhCutA1-SnC trimer

Figure 12

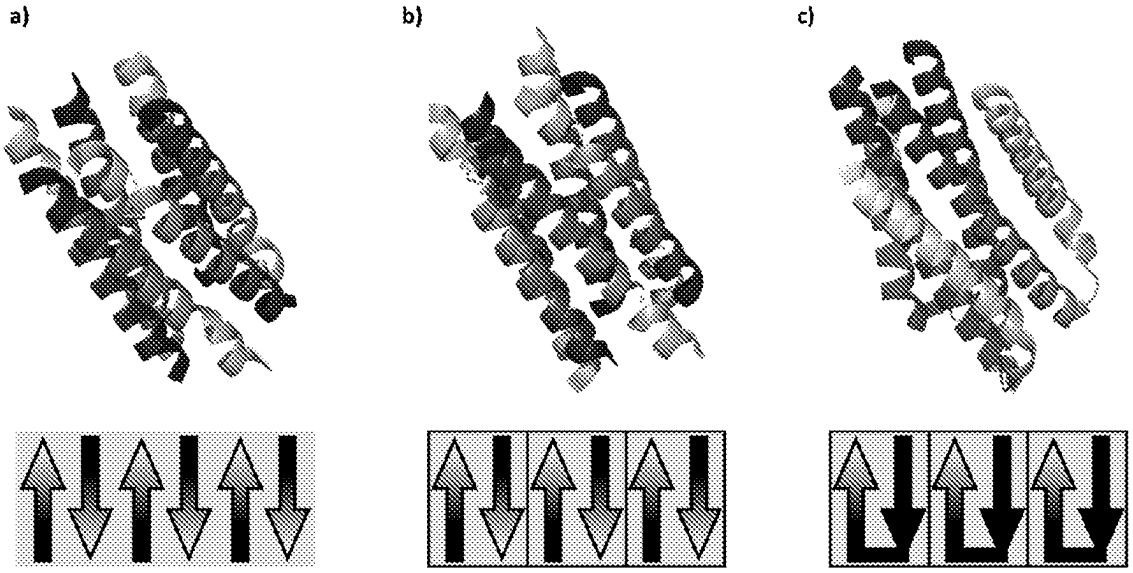


Figure 13

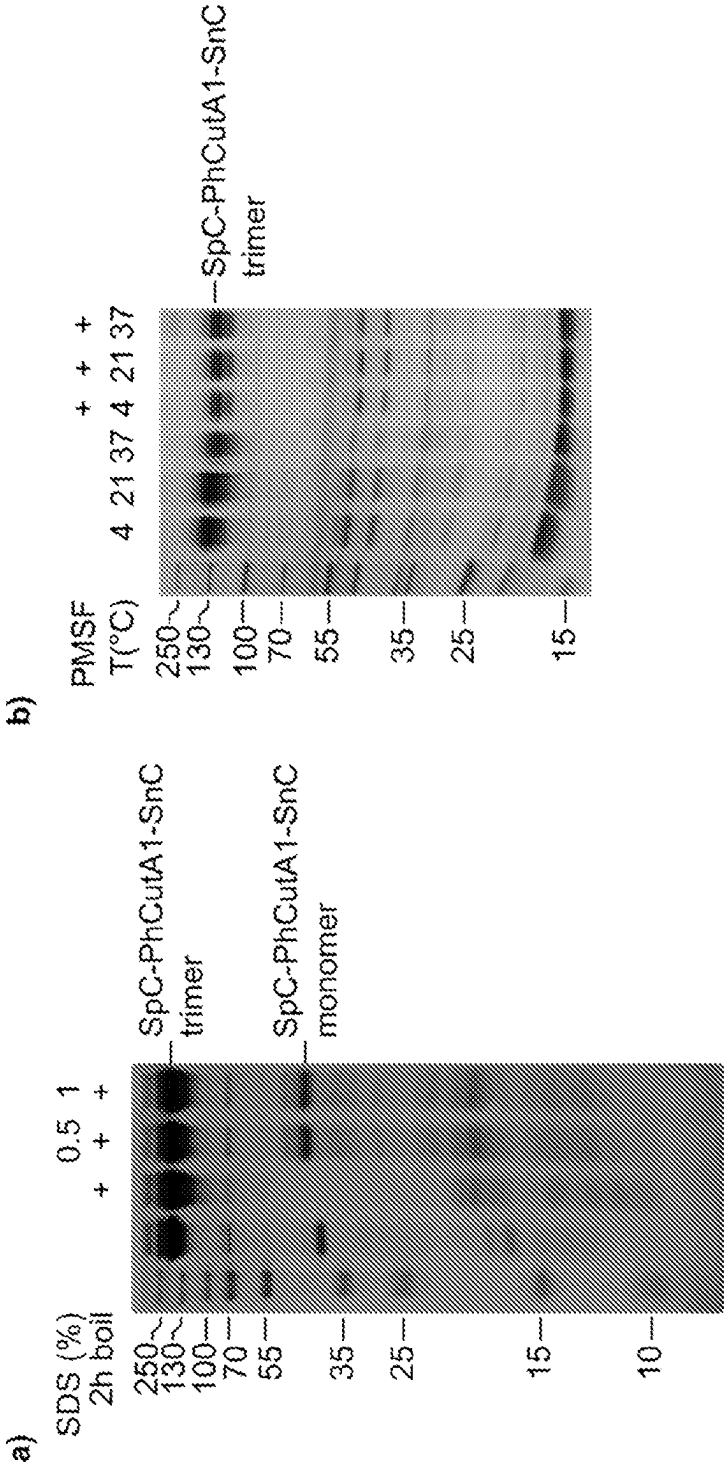

a)
SDS (%)
2h boil
0.5 1
+ + +
250
130
100
70
55
35
25
15
10
SpC-PhCutA1-SnC
trimer
SpC-PhCutA1-SnC
monomer
b)
PMSF
T(°C)
4 21 37 4 21 37
+ + +
250
130
100
70
55
35
25
15
SpC-PhCutA1-SnC
trimer

Figure 14

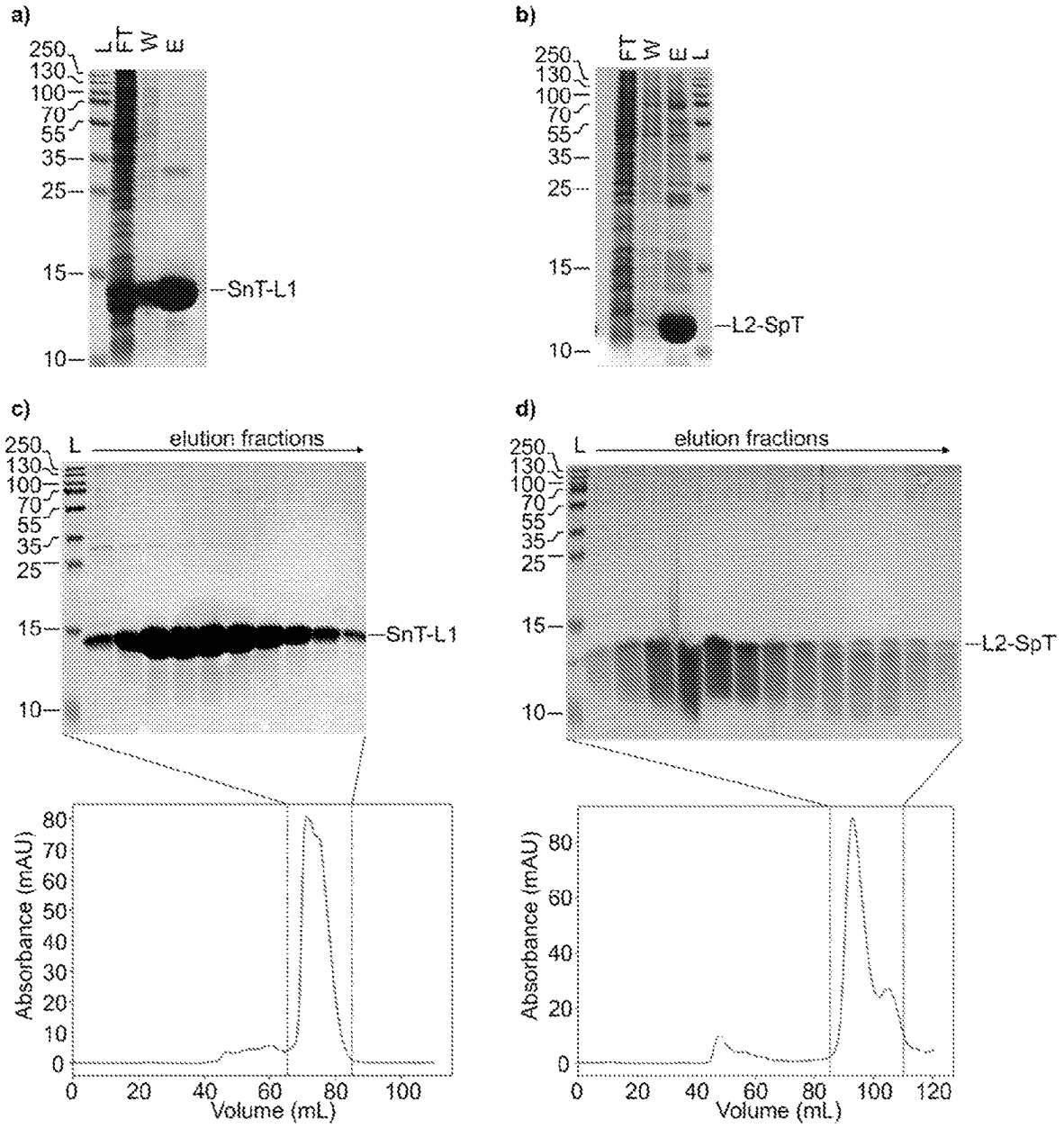

a)
L
FT
W
E
250
130
100
70
55
35
25
15
10
SnT-L1
b)
FT
W
E
L
250
130
100
70
55
35
25
15
10
L2-SpT
c)
L
elution fractions
250
130
100
70
55
35
25
15
10
SnT-L1
Absorbance (mAU)
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
Volume (mL)
d)
L
elution fractions
250
130
100
70
55
35
25
15
10
L2-SpT
Absorbance (mAU)
80
60
40
20
0
0
20
40
60
80
100
120
Volume (mL)

Figure 15

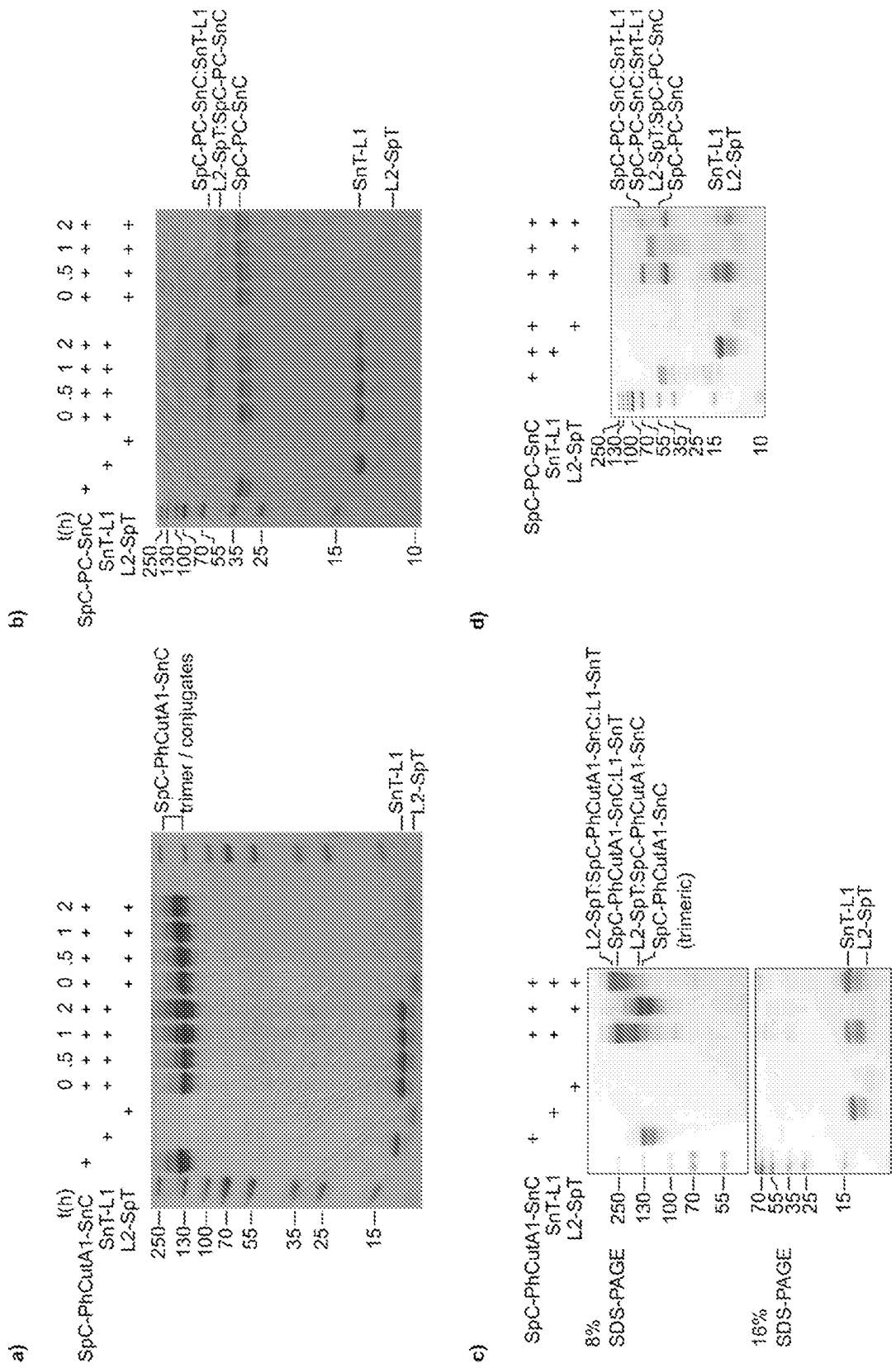

a)
t(h)
SpC-PhCutA1-SnC
SnT-L1
L2-SpT
0 .5 1 2 0 .5 1 2
250
130
100
70
55
35
25
15
SpC-PhCutA1-SnC
trimer / conjugates
SnT-L1
L2-SpT
b)
t(h)
SpC-PC-SnC
SnT-L1
L2-SpT
0 .5 1 2 0 .5 1 2
250
130
100
70
55
35
25
15
10
SpC-PC-SnC:SnT-L1
L2-SpT:SpC-PC-SnC
SpC-PC-SnC
SnT-L1
L2-SpT
c)
SpC-PhCutA1-SnC
SnT-L1
L2-SpT
8%
SDS-PAGE
250
130
100
70
55
L2-SpT:SpC-PhCutA1-SnC:L1-SnT
SpC-PhCutA1-SnC:L1-SnT
L2-SpT:SpC-PhCutA1-SnC
SpC-PhCutA1-SnC
(trimeric)
16%
SDS-PAGE
70
55
35
25
15
SnT-L1
L2-SpT
d)
SpC-PC-SnC
SnT-L1
L2-SpT
250
130
100
70
55
35
25
15
10
SpC-PC-SnC:SnT-L1
SpC-PC-SnC:SnT-L1
L2-SpT:SpC-PC-SnC
SpC-PC-SnC
SnT-L1
L2-SpT

Figure 16

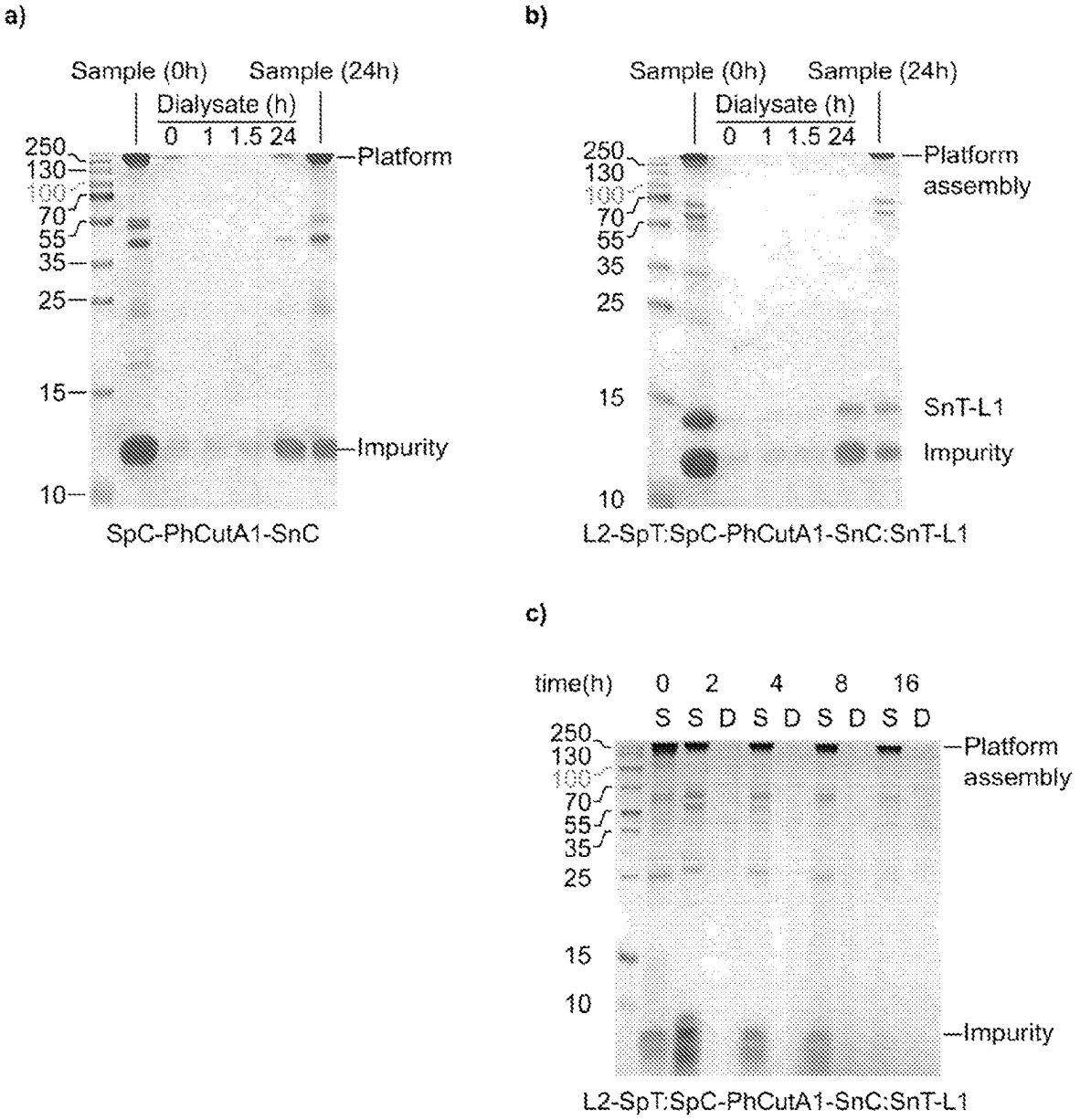

a)
Sample (0h)
Sample (24h)
Dialysate (h)
0 1 1.5 24
250
130
100
70
55
35
25
15
10
Platform
Impurity
SpC-PhCutA1-SnC
b)
Sample (0h)
Sample (24h)
Dialysate (h)
0 1 1.5 24
250
130
100
70
55
35
25
15
10
Platform assembly
SnT-L1
Impurity
L2-SpT:SpC-PhCutA1-SnC:SnT-L1
c)
time(h) 0 2 4 8 16
S S D S D S D S D
250
130
100
70
55
35
25
15
10
Platform assembly
Impurity
L2-SpT:SpC-PhCutA1-SnC:SnT-L1

Figure 17

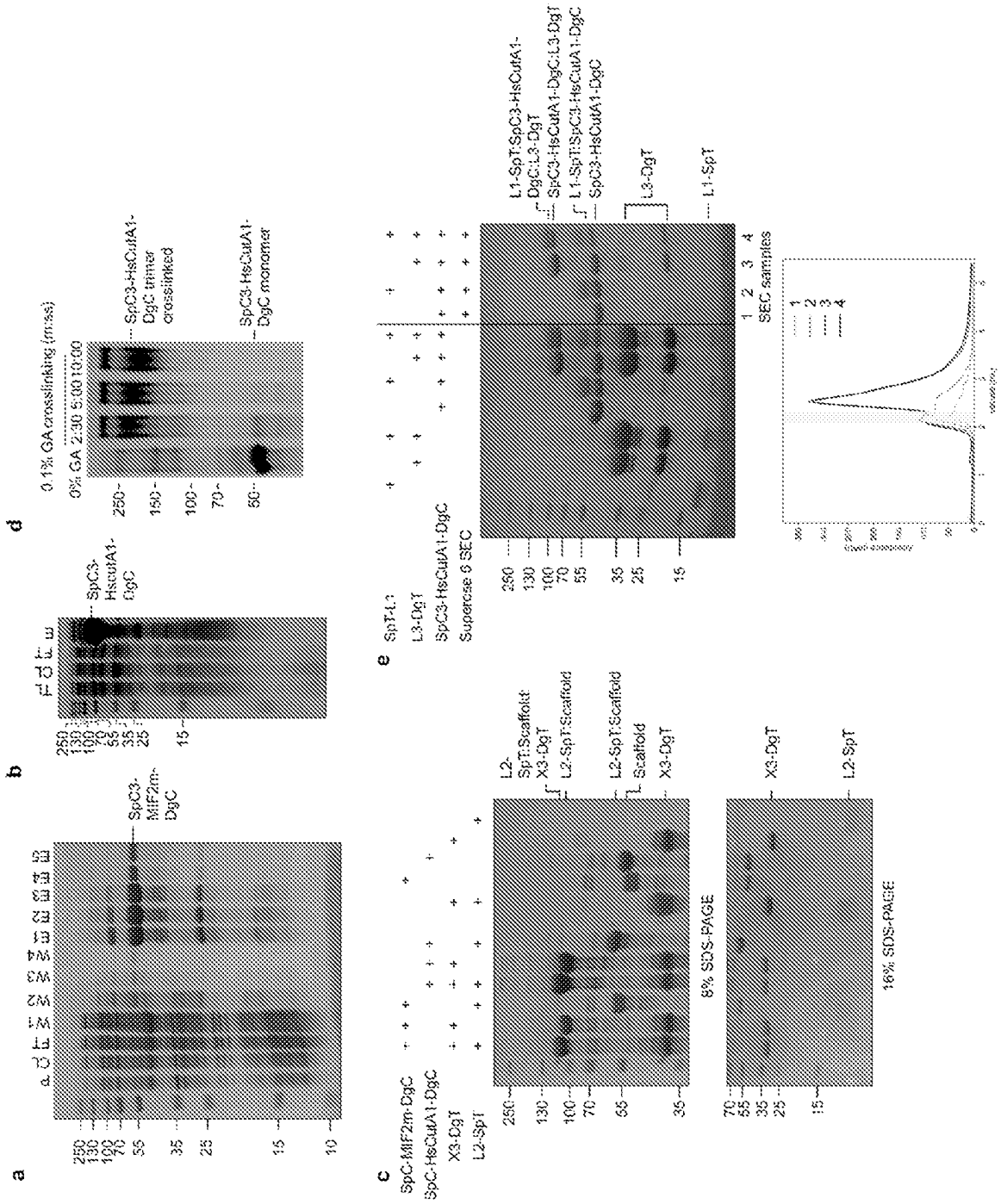

a
SpC3-MIF2m-DgC
b
SpC3-HsCutA1-DgC
d
0.1% GA crosslinking (min:ss)
0% GA 2:30 5:00 10:00
SpC3-HsCutA1-DgC trimer crosslinked
SpC3-HsCutA1-DgC monomer
c
SpC-MIF2m-DgC
SpC-HsCutA1-DgC
X3-DgT
L2-SpT
L2-SpT:Scaffold
L2-SpT:Scaffold:X3-DgT
L2-SpT:Scaffold
Scaffold
X3-DgT
8% SDS-PAGE
16% SDS-PAGE
e
SpT-L3
L3-DgT
SpC3-HsCutA1-DgC
Superose 6 SEC
L1-SpT:SpC3-HsCutA1-DgC:L3-DgT
SpC3-HsCutA1-DgC:L3-DgT
L1-SpT:SpC3-HsCutA1-DgC
SpC3-HsCutA1-DgC
L3-DgT
L1-SpT
SEC samples

Figure 18

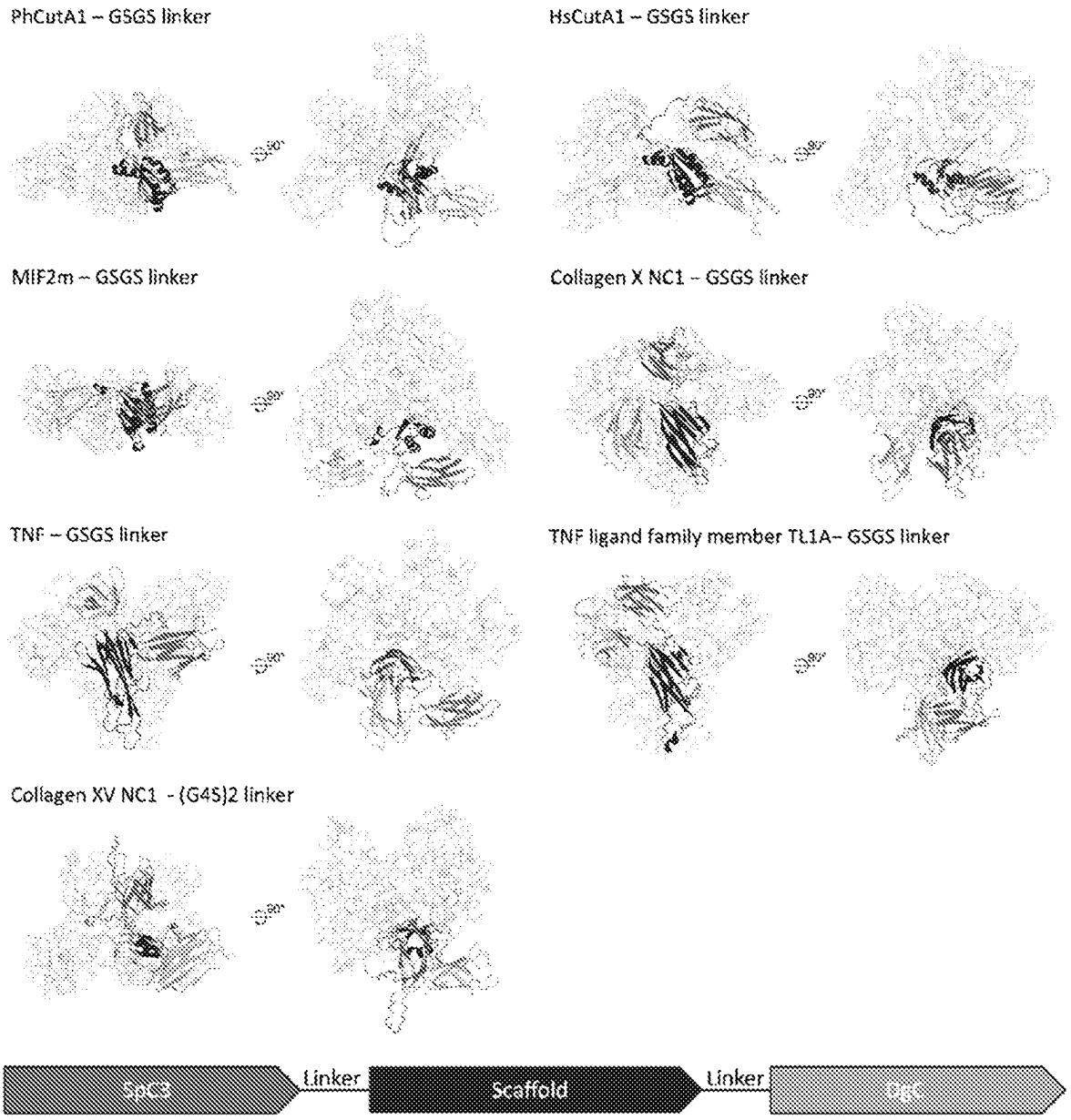

PhCutA1 – GSGS linker
HsCutA1 – GSGS linker
MIF2m – GSGS linker
Collagen X NC1 – GSGS linker
TNF – GSGS linker
TNF ligand family member TL1A– GSGS linker
Collagen XV NC1 - (G4S)2 linker
90°
SpC3
Linker
Scaffold
Linker
SpC

Figure 19

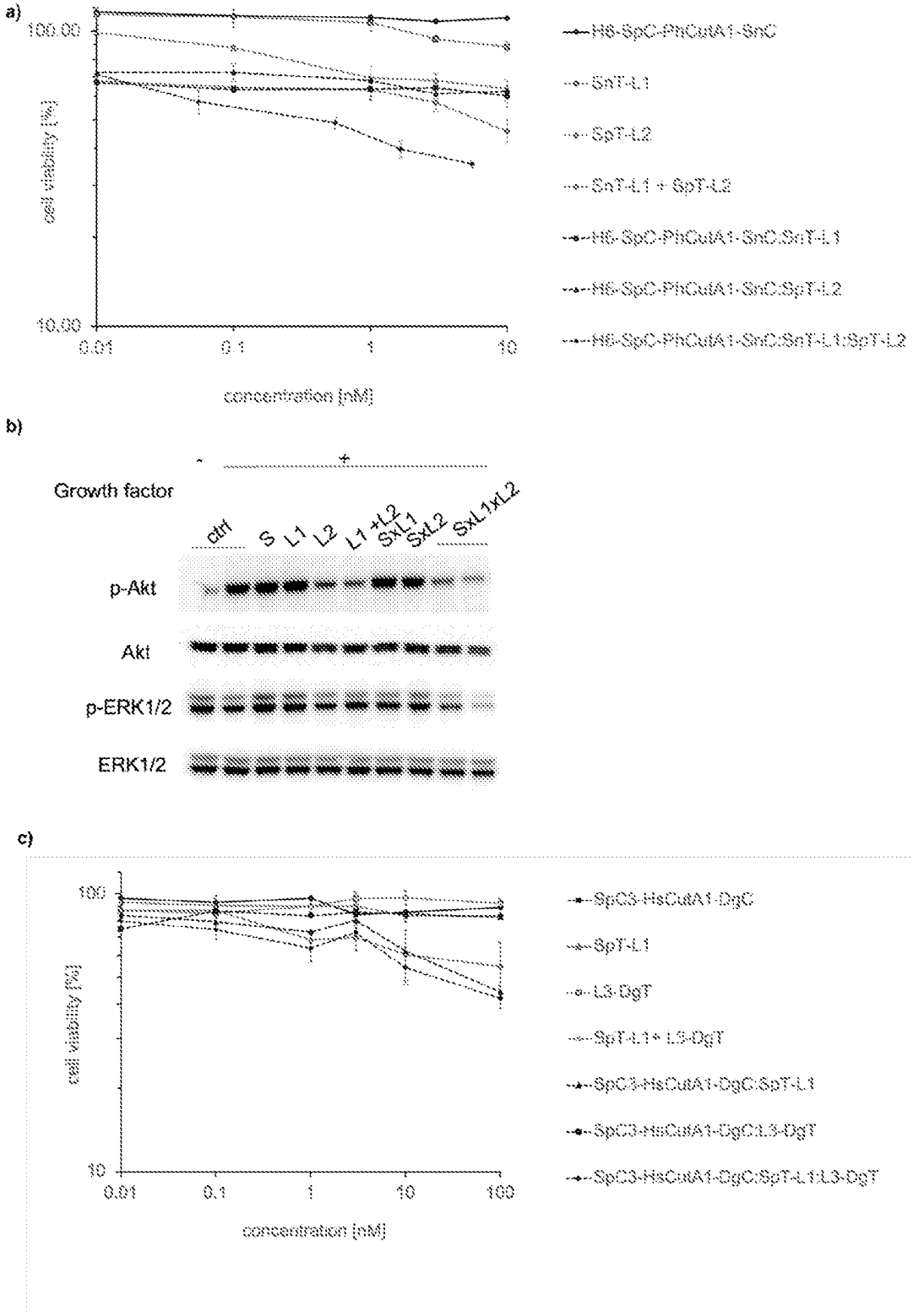

a)
100.00
10.00
cell viability [%]
0.01
0.1
1
10
concentration [nM]
H6-SpC-PhCutA1-SnC
SnT-L1
SpT-L2
SnT-L1 + SpT-L2
H6-SpC-PhCutA1-SnC:SnT-L1
H6-SpC-PhCutA1-SnC:SpT-L2
H6-SpC-PhCutA1-SnC:SnT-L1:SpT-L2
b)
-
+
Growth factor
ctrl
S
L1
L2
L1 +L2
SxL1
SxL2
SxL1xL2
p-Akt
Akt
p-ERK1/2
ERK1/2
c)
100
10
cell viability [%]
0.01
0.1
1
10
100
concentration [nM]
SpC3-HsCutA1-DgC
SpT-L1
L3-DgT
SpT-L1+ L3-DgT
SpC3-HsCutA1-DgC:SpT-L1
SpC3-HsCutA1-DgC:L3-DgT
SpC3-HsCutA1-DgC:SpT-L1:L3-DgT

Figure 20

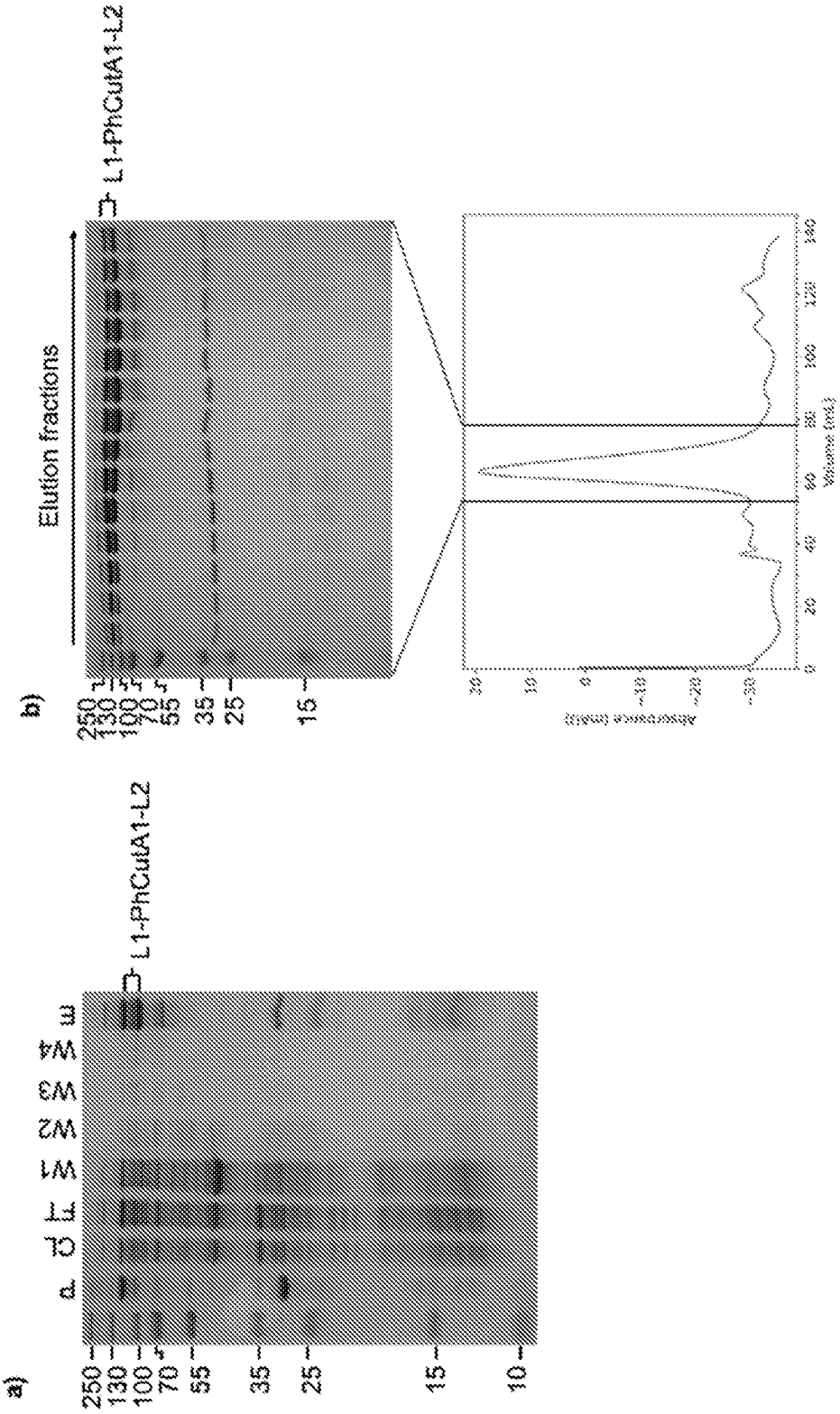

a)
250
130
100
70
55
35
25
15
10
P
CL
FT
W1
W2
W3
W4
E
L1-PhCutA1-L2
b)
Elution fractions
250
130
100
70
55
35
25
15
L1-PhCutA1-L2
Absorbance (mAU)
30
10
0
-10
-20
-30
0
20
40
60
80
100
120
140
Volume (mL)

Figure 21

MULTIVALENT PROTEINS AND SCREENING METHODS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a U.S. National Phase application under 35 U.S.C. § 371 of International Application No. PCT/GB2022/050750, filed Mar. 24, 2022, which claims benefit to GB 2104104.1, filed Mar. 24, 2021, which are hereby incorporated by reference in their entirety, and to which applications we claim priority.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII text file format and is hereby incorporated by reference in its entirety. Said ASCII text file copy, created on Aug. 16, 2024, is named 265479_000002_SL.txt and is 90,420 bytes in size.

FIELD

[0003] The present invention relates to multivalent protein scaffolds and their use as a modular system for phenotypic screening of combinations of target molecules and as therapeutics. The invention also relates to multi-domain polypeptide constructs comprising multiple binding domains and a structural domain. The invention also relates to methods of identifying new therapeutics using the described protein scaffolds and to therapeutics that can be identified in this manner.

BACKGROUND

[0004] There is an ongoing need to identify new therapeutics for many pathological conditions.

[0005] Protein-based therapies have offered an attractive approach to address many common diseases. Such therapeutics have proven to have high clinical success and many protein therapeutics have been approved by regulators for clinical use around the world. Protein-based therapeutics can operate in various ways: for example, by replacing a protein that is deficient or abnormal; by augmenting existing pathways; by providing novel functions or activities with therapeutic utility; by interfering with a molecule or organism; and by delivering other compounds or proteins, such as a radionuclide, cytotoxic drug, or effector proteins. Therapeutic proteins can be grouped based on their physical and structural properties, and can for example be divided into antibody-based drugs, Fc fusion proteins, anticoagulants, blood factors, bone morphogenetic proteins, engineered protein scaffolds, enzymes, growth factors, hormones, interferons, interleukins, thrombolytics, and the like. Therapeutic proteins can also be classified based on their molecular mechanism of activity. For example, monoclonal antibodies typically operate by binding non-covalently to targets. Enzymes may affect covalent bonds in targets. Other proteins may exert activity without any specific interaction, e.g. serum albumin.

[0006] One class of protein-based therapeutics that has enjoyed clinical success are therapeutic antibodies. Antibodies, also known as immunoglobulins (Ig), have been assessed as potential treatments for many disease conditions. For example, monoclonal antibody therapy has been employed to treat diseases including rheumatoid arthritis, multiple sclerosis, psoriasis, and various forms of cancer. Marketed antibody therapeutics include muromomab, abciximab, rituximab, daclizumab, basiliximab, palivizumab, infliximab, trastuzumab, etanercept, gemtuzumab, alemtuzumab, ibritomomab, adalimumab, alefacept, omalizumab, tositumomab, efalizumab, cetuximab, bevacizumab, natalizumab, ranibizumab, panitumumab, eculizumab, and certolizumab.

[0007] Antibodies typically comprise four polypeptide chains, forming an Fc region and two antigen-binding (Fab) regions. Each Fab region contains variable regions (Fv) that form the paratope and contact the antigen. Naturally occurring antibodies typically display symmetrical binding at the variable regions. However, this limitation means that typically only a single receptor type (or other target) can be targeted by a given antibody at the variable regions.

[0008] To seek to address this, much interest recently has turned to bispecific antibodies. Bispecific antibodies differ from conventional monospecific antibodies in that each of the two Fab sites binds to two different antigens. Bispecific antibodies are often classed as being either Ig-like or non-Ig-like, the latter of which may consist of chemically linked Fab regions.

[0009] Bispecific antibodies are being actively researched for clinical use. Two examples of marketed bispecific antibodies include Blinatumomab, sold under the brand name Blincyto and which comprises both a CD3 site for targeting T cells and a CD19 site for targeting B cells, with utility in treating Philadelphia chromosome-negative relapsed or refractory acute lymphoblastic leukemia; and Emicizumab (sold under brand name Hemlibra), which targets both clotting factors IXa and X, and which is used in the treatment of hemophilia A. Bispecific antibodies are commonly used to bind to multiple cell types at the same time, for example, by simultaneously binding tumor cell receptors and recruiting cytotoxic immune cells.

[0010] Despite the promise offered by some bispecific antibodies, problems remain. Antibody therapeutics are associated with high production costs, not least due to their size and complex post-translational modification chemistry, including complex glycosylation patterns. Antibody production necessitates the use of very large cultures of mammalian cells followed by extensive purification steps, leading to extremely high production costs and limiting the wide use of these drugs. Antibodies have also been associated with poor tumor targeting, limiting their use in treating cancer (for example, studies have shown that in murine xenograft models less than 20% of administered antibody typically interacts with the tumour). The Fc portion of antibodies, for example IgG antibodies, can interact with various receptors expressed at the surface of several cell types, which increase their retention in the circulation. Their large size can also lead to slow diffusion in vivo. IgG-like antibodies can be immunogenic, leading to detrimental downstream immune reactions via Fc-receptor activation. Regarding bispecific antibodies specifically, the Ig-like approach of "knobs into holes" that has been described previously is not readily adaptable for screening large numbers of antigen-binding domains, due to a lack of modularity. Furthermore, the bispecific antibody approach is practically limited to screening of Fv/Fab regions, and thus is not used to investigate the therapeutic potential of other, non-immunoglobulin protein domains. The non-Ig-like approach of tandem fusions is more adaptable, but not readily scalable.

[0011] Blanco-Toribio et al (MAbs. 2013 Jan. 1; 5 (1): 70-79) describes the generation and characterization of monospecific and bispecific hexavalent trimerbodies. The molecules, termed "trimerbodies," use a modified version of the N-terminal trimerization region of human collagen XVIII noncollagenous 1 (NC1) domain flanked by two flexible linkers as trimerizing scaffold. By fusing single-chain variable fragments (scFv) with the same or different specificity to both N- and C-terminus of the trimerizing scaffold domain, the authors produced monospecific or bispecific hexavalent molecules that were efficiently secreted as soluble proteins by transfected mammalian cells. A bispecific anti-laminin x anti-CD3 N-/C-trimerbody was found to be trimeric in solution. One drawback of this method is that the required use of transfection is not always a convenient production method.

[0012] WO-A-2020/0188346 describes bispecific antigen-binding proteins wherein two antigen-binding domains ("ABD") are covalently bound to a fusion protein formed of two or more domains that form an isopeptide linkage with the antigen-binding proteins. The isopeptide-linkage forming domains are typically catcher domains such as Spy-catcher ("SC"), and the resulting bispecific proteins are in the format ABD-SC-SC-ABD.

[0013] Brune et al 2017 (Bioconjugate Chem. 2017, 28, 5, 1544-1551) describes a plug-and-display synthetic assembly using orthogonal reactive proteins for twin antigen immunization. The authors prepared a dually addressable synthetic nanoparticle by engineering the multimerizing coiled-coil IMX313 and two orthogonally reactive split proteins. The construct is in the format SpyCatcher-IMX-SnoopCatcher and provides a modular platform, whereby SpyTag-antigen and SnoopTag-antigen can be multimerized on opposite faces of the particle simply upon mixing.

[0014] Accordingly, there is a need for a new paradigm to allow rapid, scalable and adaptable identification and design of new protein therapeutics which overcome some or all of these problems. In particular, there is a need for new platform technologies to rival traditional antibody platforms.

SUMMARY OF THE INVENTION

[0015] The present inventors have recognized the issues described above. It has now been recognized that non-antibody assembly platforms can be used to provide a customizable, reproducible, scalable and adaptable scaffold for screening, identifying and developing novel therapeutics. The approach described herein allows for multiple different protein geometries, valencies and/or functionalities to be assessed for potential therapeutic benefit.

[0016] Part of the inventors' approach was to develop protein constructs with favourable properties. These constructs can be prepared recombinantly, by expression as a fusion protein, or the component domains can be joined by other means known in the art such as chemical conjugation. In particular, the inventors identified that a polypeptide can advantageously be modified at both N and C termini to provide a polypeptide with two modified termini. The modifications are typically the addition of polypeptide domains that are each able to bind to a target molecule, for example an antigen-binding region or an isopeptide bond-forming region. A different target molecule may be bound by the N and C terminus, to provide a so-called bispecific binding construct. The resulting protein construct is able to bind to the target molecule at the modified N terminus and at the modified C terminus. The inventors have, in particular, engineered protein constructs wherein the N and C termini have the same general orientation, and the resulting construct is able to bind to the binding partner of each terminus when the binding partners are in the same general space and orientation, for example when bound to a solid surface such as a plate or bead, or on the surface of a cell. This may be referred to as providing modified N and C termini of a single polypeptide chain in a "cis" orientation. Typically, a cis-oriented bispecific construct is provided. Two of more of these protein constructs can combine to form an oligomeric protein.

[0017] These protein constructs allow for the creation of a combinatorial system that can be used to screen for useful combinations of effector moieties such as binding regions (for example antigen binding regions). Furthermore, once a useful combination has been identified, the construct can be modified to remove (or replace e.g. with a linker) the features necessary for the combinatorial screen, thereby providing a simpler protein construct with the identified favourable combination of binding regions. These constructs may find particular utility as therapeutic, diagnostic or analytical agents, as monomers or as oligomers of more than one construct.

[0018] Accordingly, provided herein is a multivalent protein scaffold comprising:

- **[0019]** an oligomeric core comprising a plurality of subunit monomers; and
- **[0020]** at least two first binding sites orthogonal to at least two second binding sites;
- **[0021]** wherein said first binding sites and said second binding sites are positioned on the same face of the scaffold.

[0022] Also provided is a multivalent protein scaffold comprising:

- **[0023]** an oligomeric core comprising a plurality of subunit monomers;
- **[0024]** at least one first binding site orthogonal to at least one second binding site;
- **[0025]** wherein said first binding site(s) and said second binding site(s) are positioned on the same face of the scaffold; and
- **[0026]** wherein said first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target; and said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target.

[0027] Further provided is a multivalent protein scaffold comprising:

- **[0028]** an oligomeric core comprising a plurality of subunit monomers;
- **[0029]** at least one first binding site orthogonal to at least one second binding site;

wherein said first binding site(s) and said second binding site(s) are positioned on the same face of the scaffold; and

- **[0030]** wherein said oligomeric core does not comprise an Fc region of an antibody.

[0031] Preferably, the oligomeric core comprises at least three subunit monomers. More preferably, the oligomeric core comprises from 3 to 6 subunit monomers.

[0032] Preferably, in one embodiment the subunit monomers are non-covalently attached together. Preferably, in another embodiment the subunit monomers are covalently attached together. Preferably, when the subunit monomers

are covalently attached together, the subunit monomers are genetically fused together. In some embodiments, the subunit monomers are expressed as a single polypeptide chain from a recombinant nucleic acid.

[0033] In one embodiment the oligomeric core is preferably a homooligomeric core. In such embodiments, preferably each monomer in the oligomeric core comprises at least one first binding site and at least one second binding site, and wherein the at least one first binding site is orthogonal to the at least one second binding site. Preferably, in one aspect each monomer comprises a first binding site attached at a first terminus of said monomer and a second binding site at a second terminus of said monomer. Preferably the first terminus and the second terminus of each monomer are positioned on the same face of said monomer. Preferably, in another aspect, each monomer comprises a first binding site attached at a first terminus of said monomer and a second binding site attached to said first binding site.

[0034] In another embodiment the oligomeric core is preferably a hetero-oligomeric core. Preferably, in such embodiments said core comprises at least one first subunit monomer comprising a first binding site, and at least one second subunit monomer comprising a second binding site, and wherein the first binding site is orthogonal to the second binding site.

[0035] Preferably, in the invention, the protein scaffold provided herein typically is such that each subunit monomer comprises less than 300 amino acids; preferably less than 200 amino acids; more preferably less than 150 amino acids. Preferably, the oligomeric core has a molecular weight of less than about 150 kDa, preferably less than about 100 kDa; more preferably less than about 70 kDa.

[0036] In some embodiments, the oligomeric core does not comprise an Fc region of an antibody. In some embodiments, the oligomeric core does not comprise a CH2 domain. In some embodiments, the oligomeric core does not comprise a CH3 domain. In some embodiments, the oligomeric core does not comprise a CH2 domain and does not comprise a CH3 domain.

[0037] Preferably, the oligomeric core and/or the scaffold does not generate an immune response when administered to a human subject. This is described in more detail herein.

[0038] In some embodiments, the oligomer core and/or scaffold, or the structural domain, does not generate a deleterious immune response when administered to a human subject. For example, an active B cell or T cell response is not raised against the structural domain, and/or the structural domain does not specifically bind to immunoglobulin receptors or activate antibody-dependent cell medicated toxicity (ADCC).

[0039] Preferably, the oligomeric core comprises a soluble multimerising structural element of a multimeric protein. Preferably, the multimeric protein comprises a collagen NC (noncollagenous) domain (e.g. an NC1 domain), a CutAl, a Clq head domain, a TNF, a p53, a fibrinogen, a C4, *Bacillus subtillus* AbrB or a homolog or paralog thereof.

[0040] Preferably, the multimeric protein comprises a Collagen VIII NC1 (noncollagenous) domain, a Collagen X NC1 (noncollagenous) domain, a Clq head domain, a CutAl protein, a Macrophage Migration Inhibitory Factor (MIF) or Macrophage Migration Inhibitory Factor 2 (MIF-2), a Tumor Necrosis Factor (TNF), a TNF family member including TLIA or CD40L, or a homolog or paralog thereof.

[0041] Preferably, the multimerising structural element comprises a polypeptide have at least 30% or at least 50% amino acid identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 29, SEQ ID NO: 60, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 42, SEQ ID NO: 31, SEQ ID NO: 58 or SEQ ID NO: 19.

[0042] Preferably, the first binding site and/or said second binding site comprises a protein domain. Typically, said first binding site comprises a first protein domain and said second binding site comprises a second protein domain. Preferably, the first binding site and/or second binding site is genetically fused to the subunit monomer(s) to which they are attached to form a single polypeptide chain.

[0043] Preferably, the first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target. Preferably, the said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target. More preferably, the first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target and the said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target. Preferably said first protein domain is capable of forming an isopeptide bond with said first polypeptide target and said second protein domain is capable of forming an isopeptide bond with said second binding target.

[0044] Preferably, said first binding site and said second binding site each comprise a different split ligand-binding protein domain. More preferably, one of said first binding site and said second binding site comprises a split *Streptococcus pyogenes* fibronectin-binding protein domain and the other of said first binding site and said second binding site comprises a split *Streptococcus pneumoniae* adhesin domain.

[0045] Preferably, said first and said second binding site each independently have at least 50% amino acid identity to any one of SEQ ID NOs: 4-9, 11-13, 23 or 15-18. In some embodiments, said first and said second binding site each independently have at least 60%, at least 70%, at least 80% or at least 90% amino acid identity to any one of SEQ ID NOs: 4-9, 11-13, 23 or 15-18.

[0046] Also provided herein is a protein complex comprising a protein scaffold as described herein, wherein the first binding site is bound to a first polypeptide target attached to a first effector moiety, and the second binding site is bound to a second polypeptide target attached to a second effector moiety.

[0047] Preferably, in said complex the first binding site/polypeptide target pair and the second binding site/polypeptide target pair are each independently selected from the following combinations: (i) any one of SEQ ID NO: 4, 6 or 8 with any one of SEQ ID NOs: 5, 7 or 9; (ii) SEQ ID NO: 12 with SEQ ID NO: 13 or 15; (iii) SEQ ID NO: 5 with SEQ ID NO: 11; (iv) SEQ ID NO: 15 with SEQ ID NO: 16); (v) SEQ ID NO: 17 with SEQ ID NO: 18; or (vi) SEQ ID NO: 23 with SEQ ID NO: 16).

[0048] Also provided is a screening platform comprising a library, wherein said library comprises a plurality of populations of protein complexes as described herein, wherein the populations of protein complexes each comprise a different combination of first effector moieties, second effector moieties and/or oligomeric core.

[0049] Also provided is a method for identifying a therapeutic drug or drug analog, the method comprising:

[0050] providing a protein complex as described herein;

[0051] contacting the protein complex with a biological system; and

[0052] measuring whether the protein complex induces a desired change in a property of the biological system;

[0053] and optionally further comprising selecting a protein complex that induces a desired change in a property of the biological system.

[0054] The method may further comprise synthesizing a therapeutic drug or drug candidate comprising the oligomeric core of the scaffold of the protein complex of the identified therapeutic drug analog attached to the first and second effector moieties of said protein complex.

[0055] Also provided is a therapeutic drug candidate obtainable according to the methods described herein.

[0056] Also provided is a therapeutic drug obtainable according to the methods described herein.

[0057] Also provided is a therapeutic drug or drug candidate comprising or consisting of one or more constructs or polypeptides as described herein.

[0058] Also provided herein is a therapeutic drug or drug candidate, comprising an oligomeric core comprising a plurality of subunit monomers attached to one or more first effector moieties and one or more second effector moieties; wherein said one or more first effector moieties and said one or more second effector moieties are positioned on the same face of the oligomeric core; and wherein (i) said one or more first effector moieties comprise two or more first effector moieties and said one or more second effector moieties comprise two or more second effector moieties; and/or (ii) said oligomeric core does not comprise an antibody or antibody fragment. Preferably, the oligomeric core of the therapeutic drug counterpart is as described in more detail herein.

[0059] Preferably, in the provided therapeutic drug candidate, the oligomeric core comprises a plurality of subunit monomers and: (i) each subunit monomer comprises a collagen NC1 domain, a CutAl, a Clq domain, a TNF, a p53, a fibrinogen, a C4, *Bacillus subtillus* AbrB or a homolog or paralog thereof; and/or (ii) each subunit monomer comprises a multimerising structural element comprising a polypeptide having at least 50% amino acid identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 19.

[0060] Preferably, in the provided therapeutic drug or drug candidate, the oligomeric core comprises a plurality of subunit monomers and: (i) each subunit monomer comprises a Collagen VIII NC1 (noncollagenous) domain, a Collagen X NC1 (noncollagenous) domain, a Clq head domain, a CutAl protein, a Macrophage Migration Inhibitory Factor (MIF) or Macrophage Migration Inhibitory Factor 2 (MIF-2), a Tumor Necrosis Factor (TNF), a TNF family member including TL1A or CD40L or a homolog or paralog thereof; and/or (ii) each subunit monomer comprises a multimerising structural element comprising a polypeptide having at least 50% amino acid identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 29, SEQ ID NO: 60, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 42, SEQ ID NO: 31, SEQ ID NO: 58 or SEQ ID NO: 19.

[0061] In certain aspects, the invention provides a polypeptide comprising a first binding domain at the N terminus and a second binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain. The polypeptide is typically a single engineered polypeptide chain, expressed as a fusion protein from a recombinant nucleic acid. Typically, the first binding domain and second binding domain are able to bind to their targets when the target molecules are expressed on a single cell or immobilised onto a plate or single bead. This is sometimes described herein as providing the first and second binding domains in "cis" orientation.

[0062] In some embodiments, the first binding domain and second binding domain are different antigen-binding domains. The construct is then a bispecific construct.

[0063] In some embodiments, the first binding domain and/or the second binding domain is a protein or peptide capable of specific binding with a biological molecule. This may be a signalling molecule capable of specific interaction with a binding partner, such as a p protein or peptide ligand or a receptor, for example a cytokine or a cell surface receptor.

[0064] In other embodiments, the first binding domain and the second binding domain are catcher domains (i.e. split ligand-binding protein domains) that are each able to form an isopeptide linkage with a cognate peptide. These cognate peptides are often referred to as tag peptides, for example a SpyTag forms an isopeptide bond with a SpyCatcher domain as is known in the art and as discussed below. Typically, the cognate peptide for the first binding domain is different from the cognate peptide for the second binding domain. However, in some embodiments the cognate peptide may be the same for both first and second binding domains. The polypeptide with the catcher at each terminus may be provided separately from the molecule (e.g protein) comprising the tag, for example as part of a kit. In some embodiments, each tag peptide is covalently attached to its cognate catcher domain, optionally wherein one or both cognate peptides are linked to the first and/or second catcher domain by an isopeptide bond. In some embodiments, one or both cognate peptide tags are present as a fusion polypeptide with an effector moiety, typically an antigen binding domain. The linkage of the catcher to its cognate peptide tag therefore links the effector moiety (e.g. antigen binding domain) to its binding domain.

[0065] In some embodiments, the polypeptide comprises a first binding domain at the N terminus and a second binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain, the first binding domain and the second binding domain are catcher domains that are each able to form an isopeptide linkage with a cognate peptide, wherein the first catcher domain is linked to its cognate peptide tag by an isopeptide bond and wherein the second catcher domain is linked to its cognate peptide tag by an isopeptide bond. In some embodiments, each peptide tag is attached to an antigen binding domain.

[0066] The polypeptide of the previous four paragraphs is useful as a monomer, or combined to form an oligomer.

[0067] In some aspects, an oligomer is provided comprising two or more polypeptides as defined above and elsewhere herein.

[0068] In some aspects, a polypeptide or oligomer as defined in the preceding paragraphs comprises the features described elsewhere herein. For example, the structural domain of the polypeptide construct is typically the "subunit monomer" as described extensively herein, so definitions and description of the subunit monomer apply to the struc-

tural domain. Similarly, the first binding domain and second binding domain are typically the first binding site and second binding site as described elsewhere herein, so definitions and description of the first and second binding sites apply to the first binding domain and second binding domain of the polypeptide construct.

[0069] The present invention allows for the highly adaptable screening of large numbers of effector moieties. The effector moieties may be any protein domain and are not limited to Fab/Fv regions or other antigen-binding domains. The present invention also may also be used to investigate effects of molecules achieved by higher valency interactions that would not be observed using conventional bispecific antibody or other approaches. Accordingly, the invention provides a system for high throughput screening of bispecific combinations of molecules, that can pick up effects that are only observed through higher-valency interactions. The invention also provides new therapeutic candidates which may be identified according to the methods provided herein. The therapeutic candidates provide the benefits of multiple functionalities and increased valency.

[0070] Limited use of multivalent protein scaffolds has been described in the art. One attempt described by Brune et al., Bioconjugate Chemistry, 28 (5), pp. 1544-1551, relates to the use of a heptameric assembly in which antigens were assembled on opposing faces of an IMX313 heptameric core as a potential malaria vaccine. However, such constructs are not suitable for same-cell multi-receptor engagement due to opposing display of the attached antigens, limiting their applicability in same-cell binding for treating disorders such as cancer and autoimmune disorders.

[0071] Accordingly, there is a need for new and/or improved methods for phenotypic screening of combinations of effector moieties (also referred to herein as ligands) and for methods of developing new therapeutics. Prior methods do not allow for increased valency of the combinations of functionalities investigated, and/or do not recognize the synergistic benefits of presenting antigens on the same face of a protein scaffold. There is also a need for new and/or improved therapeutics, including the therapeutics provided herein which can be designed and identified according to the present disclosure.

BRIEF DESCRIPTION OF THE FIGURES

[0072] FIG. **1** is a schematic showing a multivalent protein scaffold as described herein in which a first binding site and a second binding site are positioned on the same face of the oligomeric core and thus on the same face of the multivalent protein scaffold.

[0073] FIG. **2** is a schematic showing a multivalent protein scaffold as described herein in which a first binding site and a second binding site are positioned on opposite faces of the oligomeric core and thus on opposite faces of the multivalent protein scaffold.

[0074] FIG. **3** is a schematic showing a multivalent protein scaffold as described herein in which a plurality of first binding sites and a plurality of second binding sites are positioned on the same face of the oligomeric core and thus on the same face of the multivalent protein scaffold for engagement with a surface.

[0075] FIG. **4** is a schematic showing a multivalent protein scaffold as described herein in which a plurality of first binding sites and a plurality of second binding sites are positioned on opposite faces of the oligomeric core and thus on opposite faces of the multivalent protein scaffold and therefore cannot simultaneously engage with a surface.

[0076] FIG. **5** is a schematic showing a multivalent protein scaffold as described herein in which a first binding site and a second binding site are positioned on the same face of the oligomeric core and thus on the same face of the multivalent protein scaffold in a front-front orientation.

[0077] FIG. **6** is a schematic showing a multivalent protein scaffold as described herein in which a first binding site and a second binding site are positioned on the same face of the oligomeric core and thus on the same face of the multivalent protein scaffold in a front-side orientation.

[0078] FIG. **7** is a schematic showing a multivalent protein scaffold as described herein in which a first binding site and a second binding site are positioned on the same face of the oligomeric core and thus on the same face of the multivalent protein scaffold in an orientation intermediate between a front-front orientation and a front-side orientation.

[0079] FIG. **8** is a schematic showing the angle (X) that is formed between a first binding site and a second binding site attached to a subunit monomer of the oligomeric core of a multivalent protein scaffold as described herein.

[0080] FIG. **9** is a schematic showing an embodiment of the invention in which a tandem fusion of first and second binding sites is attached to an oligomeric core as described herein allowing production of a multivalent protein scaffold. Various different geometries and stoichiometries can be produced using the methods disclosed herein.

[0081] FIG. **10** illustrates a 2D exposition of fusion sites in "cis". a) For a given target plane, the longest cross-section of the core protein via an orthogonal line drawn from that target plane is determined, featuring a distance d_c. A parallel plane crossing the midpoint of this cross-section is shown. Here, sites for protein conjugation (stars, circles) are considered to preferentially be in "cis" if, for all conjugation sites, the distance of the shortest path from a conjugation site to the target plane that does not intersect with the protein surface is less than a certain % of the distance d_c, such as less than 50% of d_c. b) An example protein in which all binding sites are in cis. c) An example protein in which all binding sites are in cis to each other according to a), as all second binding sites (circles) feature minimal path lengths that are just below the threshold of 50% d_c. d) Even though the projected distances (crossing the protein) from all binding sites to the target plane are the same as in c), this protein features a geometry that obstructs the shortest paths from all second binding sites (circles), therefore these binding sites are not accessible to and/or do not fall onto the same side of the scaffold. e,f) The binding sites are too far apart from each other to be considered cis towards any target plane.

[0082] FIG. **11** provides an overview of protein structures referred to herein. Protein structures for given PDB IDs are visualized as cartoon, with chains in differing hues. N-terminus and C-terminus are each annotated for a single monomer, and approximately oriented towards a binding surface. The symmetry of the protein structure is shown in parenthesis, e.g. cyclic C3 symmetry, cyclic C4 symmetry, and dihedral D2 symmetry. *: protein symmetry is estimated from the NMR structure. +1PK6 is a heteromer featuring C1 symmetry, however the domains are homomeric and their arrangement resembles C3 symmetry. By selection of a multimeric protein core with suitable monomer configuration, multiple binding sites can be projected per monomer to form a single binding surface. In addition to recombinant

fusion, such a protein could then be utilised for modular assembly, for instance by recombinant fusion at N-terminus and C-terminus with SpyCatcher and SnoopCatcher or DogCatcher, to quickly confer multivalency or other properties onto suitably modified peptides or proteins, for instance, via a recombinant fusion of SpyCatcher at the N-terminus and SnoopCatcher at the C-terminus of the monomer of an oligomeric core protein. Examples include: (1) HsCutAl , a human copper-binding protein with high thermostability, featuring C3 geometry with both N- and C-termini of each monomer in close proximity to each other, projecting onto a single plane in the assembled trimer (PDB ID 2ZFH). (2) The hyper-thermostable homologue PhCutAl from *Pyrococcus horikoshii* (PDB ID 4NYO) which has high structural similarity to HsCutAl. (3) NC1-domain of Collagen X (PDB ID: 1GR3), (4) NC1-domain of Collagen VIII (PDB ID: 1091), (5) Macrophage Migration Inhibitory Factor 2 (PDB ID: 7MSE), (6) Tumor Necrosis Factor (PDB ID: 1TNF), (7) TNF-like protein TLIA (PDB ID: 2RE9).

[0083] FIG. **12** shows that cis-oriented multimeric protein complexes can be readily expressed and prepared by standard protein purification methods. a) Ni-NTA purification of H6-SpC-PhCutAl-SnC [SEQ ID NO: 21]. H6-SpC-PhCutAl-SnC is readily expressed in *E. coli* BL21 (DE3) and retains an intact trimeric structure characteristic to hyper-stable PhCutAl, even after boiling with SDS-PAGE loading buffer. Washes 1 and 2 were performed with 10 column volumes equilibration buffer (50 mM Tris, pH 7.8; 300 mM NaCl; 10 mM imidazole). Washes 3 and 4 were performed with 10 column volumes of wash buffer (50 mM Tris, pH 7.8; 300 mM NaCl; 30 mM imidazole). All elution steps were performed with 2 column volumes of elution buffer (50 mM Tris, pH 7.8; 300 mM NaCl; 200 mM imidazole). Samples were analysed using 12% SDS-PAGE gels with Coomassie staining. P-Lysate pellet; CL-Clarified lysate; FT-Flow-through; W-Wash; E-Elution. b,c) Result of size exclusion chromatography using HiLoad 16/600 Superdex 200 pg for H6-SpC-PhCutAl-SnC. b) UV A280 absorbance chromatogram with the H6-SpC-PhCutAl-SnC peak highlighted. c) SDS-PAGE of the H6-SpC-PhCutAl-SnC 2 mL fractions from the highlighted region in the above chromatogram.

[0084] FIG. **13** shows transition from a dihedral hexamer to a circularly symmetric trimeric core protein for cis-oriented display. a) A homo-hexameric antiparallel coiled-coil features N-termini and C-termini on two opposing sides of the protein assembly (PDB ID: 5W0J). b) A heteromeric assembly can be derived from a homomeric assembly by point mutagenesis, e.g. by introduction or modification of salt-bridge formation, "locking" the assembly in one orientation (PDB ID: 5VTE). c) By linking the termini of a heteromeric assembly, a homomeric assembly suitable for cis-oriented display would be derived (compare to HIV GP41, PDB ID: 115Y). Black to white (structure, schematic) and arrow direction (schematic): N-terminus to C-terminus. Structures visualised in PyMOL.

[0085] FIG. **14** shows that SpC-PhCutAl-SnC is a highly stable trimeric protein. a) Samples of SpC-PhCutAl-SnC were heated at 97° C. for 2 h in either 0%, 0.5% or 1% SDS in addition to PBS, before adding SDS-loading dye. Samples were resolved on 12% SDS-PAGE and stained with Coomassie. A control sample that was not heated and had no SDS present is shown in the first lane. At increased SDS concentrations, partial monomerisation of the trimer was observed, confirming that the trimer is not covalently cross-linked. b) SpC-PhCutAl-SnC was retained after prolonged storage. Protein aliquots were stored at 4° C., ambient temperature (21° C.), or 37° C. for 7 days before preparing samples with SDS-loading buffer and resolving on SDS-PAGE. Compared to storage at 4° C., the protein showed little sign of degradation at 21° C. to 37° C. Optionally, the protease inhibitor PMSF was added, with similar effect.

[0086] FIG. **15** shows the preparation of Spy Tagged and SnoopTagged ligand components for modular assembly to platform proteins. a) Purification of SnT-LI from 200 mL BL21 (DE3) culture using Ni-NTA chromatography. Each washing step was performed with 5 mL Ni-NTA wash buffer. Each elution step was performed with 2 mL of Ni-NTA elution buffer. Samples were analyzed using 12% SDS-PAGE gels with Coomassie staining. Pel.-Lysate pellet; FT-Flow-through; W-Wash; E-Elution; L-Ladder. b) Purification of L2-SpT as in a). c) Size exclusion chromatography of SnT-L1 after Ni-NTA purification. UV A280 absorbance chromatogram for SnT-LI from AKTA Pure 25 with a HiLoad Superdex 16/600 75 pg column. Inset: SDS-PAGE of the SnT-L1 2 mL fractions from the highlighted region in the chromatogram. d) Size exclusion chromatography of L2-SpT as in c).

[0087] FIG. **16** shows that SpC-PhCutAl-SnC facilitates stable trimerization of SpyTagged and SnoopTagged proteins. a) H6-SpC-PhCutAl-SnC was conjugated with a 1:2:2 molar excess of SnT-L1 or L2-SpT for the indicated time before samples were supplemented with SDS-loading buffer and all samples were denatured by boiling at 95° C. for 5 min. Samples were resolved on 8% and 16% SDS-PAGE gels followed by Coomassie staining. We observed time-dependent conjugation of SpC-PhCutAl-SnC to SnT-LI or L2-SpT, with consumption of ligand components. b) Conjugation of SpC—PC-SnC with SnT-L1 and L2-SpT as in a). c) Conjugation of SpC-PhCutAl-SnC with SnT-LI and L2-SpT. SpC-PhCutAl-SnC was incubated with SnT-L1 and/or L2-SpT at 1:2:2 molar excess and samples were incubated at 25° C. for 64 h. Samples were analysed using 16% SDS-PAGE gels with Coomassie staining. Notably, SpC-PhCutAl-SnC and SnT-L1/L2-SpT conjugate to completion while retaining characteristic hyper-thermostability of PhCutAl. d) Conjugation of SpC—PC-SnC with SnT-LI and L2-SpT as in c). SpC—PC—SnC and SnT-L1/L2-SpT conjugation was performed to completion.

[0088] FIG. **17** shows that a high-molecular weight scaffold enables post-assembly purification via dialysis. a) Ni-NTA purified SpC-PhCutAl-SnC and b) SpC-PhCutAl-SnC: SnT-L1: L2-SpT assembly were dialysed utilizing a 96-well format. a) SpC-P2-SnC was purified using Ni-NTA chromatography and the elution fractions were combined and concentrated. b) Protein conjugation was performed for SpC-PhCutAl-SnC, SnT-L1 and L2-SpT, at 25° C. for 2 h. Dialysis was performed using a 100 kDa MWCO membrane with a 1:1 ratio of sample to dialysis buffer in a HTDialysis 96-well block. The dialysis was performed with orbital shaking at ambient temperature. PBS dialysate was changed every 30 min up to 90 min. 24 h of dialysis show removal of protein impurities (a-b) and unconjugated ligands (b) to equilibrium (at 1:1 ratio of sample and dialysis buffer). Samples were boiled with reducing SDS-loading dye, resolved on 12% SDS-PAGE and visualized by Coomassie staining. c) Alternatively, SpC-PhCutAl-SnC: SnT-L1: L2-SpT conjugate was purified via dialysis in 12-well plate

format, featuring a 1:30 ratio of sample to dialysis buffer. A 1:1 conjugation was set up for SpC-PhCutAl-SnC, SnT-L1 and L2-SpT at 25° C. for 24 h. Dialysis was performed using a HTDialysis 12-well block and a 100 kDa MWCO cellulose membrane over 16 h at ambient temperature with no agitation. A sample volume of 100 μL and a PBS dialysate volume of 3 mL was used, both of which contained 1×PMSF. Sample and dialysate were taken at the following timepoints: 2 h, 4 h, 8 h and 16 h. Samples were analysed using a 14% SDS-PAGE gel and Coomassie staining. S=dialysed sample, D=dialysate (PBS).

[0089] FIG. **18** shows changes in species core components (PhCutAl to MIF2m or HsCutAl), protein components for conjugation (SpC/SnC to SpC3/DgC), and variable linker lengths (GGGGSGGGGSGGGGS for MIF2m and GGGGS for HsCutA), highlighting the potential for rapid prototyping a,b) Samples from Ni-NTA purification of H6-SpC3-HsCutAl-DgC or H6-SpC3-MIF2m-DgC. TL-Total lysate; P-lysate pellet; CL—cleared lysate; FT-flow-through; W-Wash; E-Elution. Samples were analyzed using SDS-PAGE gels with Coomassie staining. c) Both SpC3-MIF2m-DgC and SpC3-HsCutAl-DgC are capable of rapid conjugation to DogTagged or SpyTagged proteins. Platform proteins were incubated at 25° C. for 16 h in 1:1.5:1.5 molar ratio. d) Incubation of H6-SpC3-HsCutAl-DgC with 0.1% glutaraldehyde demonstrates crosslinking of trimeric protein in solution. 10 μM H6-SpC3-HsCutAl-DgC was crosslinked for 0-20 min using 0.1% glutaraldehyde. The reaction was incubated at 37° C. and stopped by the addition of 100 mM Tris, pH 8.8. Trimeric crosslinked species are rapidly formed; upon prolonged incubation, monomeric H6-SpC3-HsCutAl-DgC is consumed, with some formation of crosslinked species at the molecular weight expected for a crosslink between two trimers. e) SpC3-HsCutAl-DgC was reacted with L1-SpT and L3-DgT at a 1:2:2 molar ratio for 16 h at 25° C. Each sample of SpC3-HsCutAl-DgC only, L1-SpT: SpC3-HsCutAl-DgC, SpC3-HsCutAl-DgC: L3-DgT, and L1-SpT: SpC3-HsCutAl-DgC: L3-DgT were then subjected to size exclusion chromatography with Superose 6 Increase 5/150 GL column, showing an increase in hydrodynamic radius of each protein scaffold or protein assembly. Peak fractions from each chromatogram peak from the shaded area were loaded onto the SDS page gel to show only the scaffold and assembly proteins, with the excess ligands removed. These samples were used as the input for FIG. **20***c*.

[0090] FIG. **19** shows how Alphafold v2.0 was utilized to predict cis-orientation of fusion proteins. The highest ranked model of each SpC3-Scaffold-DgC assembly as visualized in PyMOL. A single chain is highlighted showing SpyCatcher3 (medium grey), Scaffold (dark grey) and DogCatcher (light grey). Additional chains are presented as a transparent surface with a cartoon in white. A GSGS linker was used between the catchers and scaffolds for all simulations. In the case of Collagen XV NC1 the structural prediction collapsed using the GSGS linker. Therefore, the prediction was repeated using a (GGGGS) 2 linker and this structure is presented here. Monomer sequences used for Alphafold predictions: TLIA-SEQ ID NO: 32; Col XV NC1-SEQ ID NO: 33; MIF2m-SEQ ID NO: 34; HsCutAl-SEQ ID NO: 54; PhCutAl-SEQ ID NO: 55; Col X NC1-SEQ ID NO: 56; TNF-SEQ ID NO: 57.

[0091] FIG. **20** shows suitability of the scaffold for in vitro testing. a) H6-SpC-PhCutAl-SnC was conjugated with two different ligands, SnT-LI and L2-SpT. Serum-starved NCI-N87 cells were grown for 7 days in the presence of a relevant growth factor and dual-conjugated assembly (H6-SpC-PhCutAl-SnC: SnT-L1: L2-SpT), single-conjugated assemblies (H6-SpC-PhCutAl-SnC: SnT-L1, H6-SpC-PhCutAl-SnC: L2-SpT) as well as ligand only controls (SnT-L1, L2-SpT, SnT-L1+L2-SpT), followed by MTT cell viability measurements. Control antibodies against L1, L2 and L1+L2 at 10 nM were used as controls (data not shown) and resulted in similar results compared to single- and dual-conjugated assembly samples; error bars show technical replicates of n=3. b) Fully assembled scaffold H6-SpC-PhCutAl-SnC: SnT-L1: L2-SpT represses activation of Akt and Erk1/2. NCI-N87 cells were treated with scaffold only (S; H6-SpC-PhCutAl-SnC), ligands only (L1; SnT-L1 and L2; L2-SpT) and single-(SxL1, SxL2) and dual-conjugated (SxL1xL2) assemblies for 1 h and showed repression of downstream activation of Akt/ERK signalling with the full assembly H6-SpC-PhCutAl-SnC: SnT-L1: L2-SpT. c) H6-SpC-HsCutAl-SnC was conjugated with two different ligands, SnT-L1 and L3-SpT. Serum-starved NCI-N87 cells were treated for 2 days with dual-conjugated assembly (H6-SpC-HsCutA1-SnC: SnT-L1: L3-SpT), single-conjugated assemblies (H6-SpC-HsCutA1-SnC: SnT-L1, H6-SpC-HsCutAl-SnC: L3-SpT) as well as ligand only controls (SnT-L1, L3-SpT, SnT-L1+L3-SpT), followed by MTT cell viability measurements; error bars show technical replicates of n=3.

[0092] FIG. **21** shows the purification of L1-PhCutAl-L2 as a direct fusion multi-domain polypeptide both by Ni-NTA and size exclusion chromatography. a) L1-PhCutAl-L2 is readily expression in *E. coli* BL21 (DE3) and is purified by Ni-NTA chromatography using HisPur resin (ThermoFisher). Wash 1 and 2 were performed with 10 column volumes equilibration buffer (50 mM Tris, pH 7.8, 300 mM NaCl, 10 mM imidazole). Wash 3 and 4 were performed with 2 column volumes of wash buffer (50 mM Tris, pH 7.8, 300 mM NaCl, 30 mM imidazole). All elution steps were performed with 2 column volumes of elution buffer (50 mM Tris, pH 7.8, 300 mM NaCl, 200 mM imidazole). Samples were analysed on a 12% SDS-PAGE gel and stained with Coomassie. P-Lysate Pellet, CL-Cleared Lysate, FT-Flow-through, W-Wash, E-Elution. b) Result of size exclusion chromatography using HiLoad 16/600 Superdex 200 pg for LI-PhCutAl-L2 b) UV A280 absorbance chromatogram with the L1-PhCutAl-L2 peak highlighted. Inset: SDS-PAGE of the L1-PhCutAl-L2 2 mL fractions from the highlighted region in the above chromatogram.

DETAILED DESCRIPTION

[0093] The present invention will be described with respect to particular embodiments and with reference to certain drawings but the invention is not limited thereto but only by the claims. Any reference signs in the claims shall not be construed as limiting the scope. Of course, it is to be understood that not necessarily all aspects or advantages may be achieved in accordance with any particular embodiment of the invention. Thus, for example those skilled in the art will recognize that the invention may be embodied or carried out in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other aspects or advantages as may be taught or suggested herein.

[0094] In addition as used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a scaffold" includes two or more scaffolds, reference to "an oligomer" includes two or more such oligomers and the like.

[0095] All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

Definitions

[0096] The following terms or definitions are provided solely to aid in the understanding of the invention. Unless specifically defined herein, all terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook et al., Molecular Cloning: A Laboratory Manual, 4th ed., Cold Spring Harbor Press, Plainsview, New York (2012); and Ausubel et al., Current Protocols in Molecular Biology (Supplement 114), John Wiley & Sons, New York (2016), for definitions and terms of the art. The definitions provided herein should not be construed to have a scope less than understood by a person of ordinary skill in the art.

[0097] "About" as used herein when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of +20% or =10%, more preferably +5%, even more preferably +1%, and still more preferably +0.1% from the specified value, as such variations are appropriate to perform the disclosed methods. The term "amino acid" in the context of the present disclosure is used in its broadest sense and is meant to include organic compounds containing amine (NH2) and carboxyl (COOH) functional groups, along with a side chain (e.g., a R group) specific to each amino acid. In some embodiments, the amino acids refer to naturally occurring L a-amino acids or residues. The commonly used one and three letter abbreviations for naturally occurring amino acids are used herein: A=Ala; C-Cys; D-Asp; E-Glu; F-Phe; G-Gly; H=His; I=Ile; K=Lys; L-Leu; M-Met; N=Asn; P=Pro; Q-Gln; R-Arg; S-Ser; T=Thr; V=Val; W=Trp; and Y=Tyr (Lehninger, A. L., (1975) Biochemistry, 2d ed., pp. 71-92, Worth Publishers, New York). The general term "amino acid" further includes D-amino acids, retro-inverso amino acids as well as chemically modified amino acids such as amino acid analogues, naturally occurring amino acids that are not usually incorporated into proteins such as norleucine, and chemically synthesised compounds having properties known in the art to be characteristic of an amino acid, such as B-amino acids. For example, analogues or mimetics of phenylalanine or proline, which allow the same conformational restriction of the peptide compounds as do natural Phe or Pro, are included within the definition of amino acid. Such analogues and mimetics are referred to herein as "functional equivalents" of the respective amino acid. Other examples of amino acids are listed by Roberts and Vellaccio, The Peptides: Analysis, Synthesis, Biology, Gross and Meiehofer, eds., Vol. 5 p. 341, Academic Press, Inc., N.Y. 1983, which is incorporated herein by reference.

[0098] The terms "polypeptide", and "peptide" are interchangeably used herein to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues is a synthetic non-naturally occurring amino acid, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers. Polypeptides can also undergo maturation or post-translational modification processes that may include, but are not limited to: glycosylation, proteolytic cleavage, lipidization, signal peptide cleavage, propeptide cleavage, phosphorylation, and such like. A peptide can be made using recombinant techniques, e.g., through the expression of a recombinant or synthetic polynucleotide. A recombinantly produced peptide it typically substantially free of culture medium, e.g., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

[0099] The term "protein" is used to describe a folded polypeptide having a secondary or tertiary structure. The protein may be composed of a single polypeptide, or may comprise multiple polypepties that are assembled to form a multimer. The multimer may be a homooligomer, or a heterooligmer. The protein may be a naturally occurring, or wild type protein, or a modified, or non-naturally, occurring protein. The protein may, for example, differ from a wild type protein by the addition, substitution or deletion of one or more amino acids.

[0100] A "variant" of a protein encompass peptides, oligopeptides, polypeptides, proteins and enzymes having amino acid substitutions, deletions and/or insertions relative to the unmodified or wild-type protein in question and having similar biological and functional activity as the unmodified protein from which they are derived. The term "amino acid identity" as used herein refers to the extent that uences are identical on an amino acid-by-amino acid basis over a window of comparison. Thus, a "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical amino acid residue (e.g., Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

[0101] For all aspects and embodiments of the present invention, a "variant" typically has at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% complete sequence identity to the amino acid sequence of the corresponding wild-type protein. Sequence identity can also be to a fragment or portion of the full length polynucleotide or polypeptide. Hence, a sequence may have only 50% overall sequence identity with a full length reference sequence, but a sequence of a particular region, domain or subunit could share 80%, 90%, or as much as 99% sequence identity with the reference sequence.

[0102] The term "wild-type" refers to a gene or gene product isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designed the "normal" or "wild-type" form of the gene. In contrast, the term "modified", "mutant" or "variant" refers to a gene or gene product that displays modifications in sequence (e.g., substitutions, truncations, or insertions), post-translational modifications and/or functional properties (e.g., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally occurring mutants can be isolated; these

are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product. Methods for introducing or substituting naturally-occurring amino acids are well known in the art. For instance, methionine (M) may be substituted with arginine (R) by replacing the codon for methionine (ATG) with a codon for arginine (CGT) at the relevant position in a polynucleotide encoding the mutant monomer. Methods for introducing or substituting non-naturally-occurring amino acids are also well known in the art. For instance, non-naturally-occurring amino acids may be introduced by including synthetic aminoacyl-tRNAs in the IVTT system used to express the mutant monomer. Alternatively, they may be introduced by expressing the mutant monomer in *E. coli* that are auxotrophic for specific amino acids in the presence of synthetic (i.e. non-naturally-occurring) analogues of those specific amino acids. They may also be produced by naked ligation if the mutant monomer is produced using partial peptide synthesis. Conservative substitutions replace amino acids with other amino acids of similar chemical structure, similar chemical properties or similar side-chain volume. The amino acids introduced may have similar polarity, hydrophilicity, hydrophobicity, basicity, acidity, neutrality or charge to the amino acids they replace. Alternatively, the conservative substitution may introduce another amino acid that is aromatic or aliphatic in the place of a pre-existing aromatic or aliphatic amino acid. Conservative amino acid changes are well-known in the art and may be selected in accordance with the properties of the 20 main amino acids as defined in Table 1 below. Where amino acids have similar polarity, this can also be determined by reference to the hydropathy scale for amino acid side chains in Table 2.

TABLE 1

Chemical properties of amino acids

Ala	aliphatic, hydrophobic, neutral	Met	hydrophobic, neutral
Cys	polar, hydrophobic, neutral	Asn	polar, hydrophilic, neutral
Asp	polar, hydrophilic, charged (−)	Pro	hydrophobic, neutral
Glu	polar, hydrophilic, charged (−)	Gln	polar, hydrophilic, neutral
Phe	aromatic, hydrophobic, neutral	Arg	polar, hydrophilic, charged (+)
Gly	aliphatic, neutral	Ser	polar, hydrophilic, neutral
His	aromatic, polar, hydrophilic, charged (+)	Thr	polar, hydrophilic, neutral
Ile	aliphatic, hydrophobic, neutral	Val	aliphatic, hydrophobic, neutral
Lys	polar, hydrophilic, charged(+)	Trp	aromatic, hydrophobic, neutral
Leu	aliphatic, hydrophobic, neutral	Tyr	aromatic, polar, hydrophobic

TABLE 2

Hydropathy Scale

Side Chain	Hydropathy
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8

TABLE 2-continued

Hydropathy Scale

Side Chain	Hydropathy
Cys	2.5
Met	1.9
Ala	1.8
Gly	−0.4
Thr	−0.7
Ser	−0.8
Trp	−0.9
Tyr	−1.3
Pro	−1.6
His	−3.2
Glu	−3.5
Gln	−3.5
Asp	−3.5
Asn	−3.5
Lys	−3.9
Arg	−4.5

[0103] A mutant or modified protein, monomer or peptide can also be chemically modified in any way and at any site. A mutant or modified monomer or peptide may be chemically modified by attachment of a molecule to one or more cysteines (cysteine linkage), attachment of a molecule to one or more lysines, attachment of a molecule to one or more non-natural amino acids, enzyme modification of an epitope or modification of a terminus. Suitable methods for carrying out such modifications are well-known in the art. The mutant of modified protein, monomer or peptide may be chemically modified by the attachment of any molecule. For instance, the mutant of modified protein, monomer or peptide may be chemically modified by attachment of a dye or a fluorophore.

Polypeptide Constructs

[0104] The invention relates in part to multi-domain polypeptide constructs that are useful when two or more are combined to form an oligomeric protein and may, in some embodiments, also be useful as monomers. The multi-domain polypeptide is typically engineered to combine domains that do not exist together in nature. In some embodiments, 3, 4, 5 or 6 polypeptide constructs are combined to form an oligomer. In some embodiments, 3 constructs are combined to form a trimer, for example a homotrimer.

[0105] The description provided herein describes an oligomeric core of subunit monomers. The subunit monomer is typically the structural domain of the multi-domain polypeptide construct. The oligomerisation of these structural domains in turn may form the core of a multivalent protein scaffold.

[0106] Accordingly, discussion of the features of the subunit monomer also applies to the disclosure and definition of the structural domain of the individual polypeptide constructs.

[0107] Similarly, the first binding domain and second binding domain of the polypeptide construct may form the first binding site and second binding site as described elsewhere herein, or may form the first effector moiety and the second effector moiety as described elsewhere herein, depending on the context. For example, when the binding domain is an isopeptide bond-forming "catcher" domain (or other binding site as described herein) then it is a binding

site as described elsewhere, below. When the binding domain is e.g. an antibody, an antigen-binding fragment, an antibody mimic, a protein or peptide ligand, a protein or peptide signalling molecule (e.g. cytokine) a biological receptor or other molecule described as an effector moiety herein, then the binding domain is an effector moiety as described elsewhere herein.

[0108] Accordingly, definitions and description of the first and second binding sites, or of the first and second effector moieties, apply as appropriate to the first binding domain and second binding domain of the polypeptide construct.

[0109] In some embodiments, a polypeptide construct comprises a first binding domain at the N terminus and a second binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain. The N terminus refers to the terminal amino acid residue at the amino terminus of a polypeptide. The C terminus refers to the terminal amino acid residue at the carboxy terminus of a polypeptide. Typically, the first binding domain and second binding domain are able to bind to their targets when the target molecules are expressed on a single cell or immobilised onto a plate or single bead. This is sometimes described herein as providing the first and second binding domains in "cis" orientation. Typically, the first binding domain and the second binding domain are able to bind to their cell targets on the surface of a single cell, and cluster both targets in the cell membrane. A cis orientation can therefore be preferential for some cis acting agents (i.e. that can act on a single cell). Cis orientation of bispecific antibodies is discussed in Dickopf et al (Computational and Structural Biotechnology Journal, Volume 18, 2020, Pages 1221-1227), along with the converse "trans" orientation.

[0110] In the context of geometry, "cis-orientation" is used herein to refer to a spatial arrangement in which two components are (from Latin) "on the same side" of a plane, as opposed to "trans" or "trans-orientation" in the context of geometry in which two components are "across" (from Latin), i.e. on different sides of a plane, similar to cis-trans-isomerism and previously used to describe bispecific antibody architecture. This geometric definition is distinct from "cis-acting" and "trans-acting" in the context of biological effect, in which a single bispecific molecule acts on a single or adjacent cell (cis) or on distinct cell populations (trans), e.g. recruiting an effector cell to a target cell. There is active interest and effort towards the exploration of different format geometries (e.g. see Dengl et al Nat Commun. 2020; 11:4974). "Cis-orientation" can also be beneficial in certain "trans-acting" bispecifics, e.g. by reducing intermolecular distance (Dickopf et al, Computational and Structural Biotechnology Journal, Volume 18, 2020, pp. 1221-1227). Cis-orientation is of particular interest in multivalent cis-orientation towards cis-acting bispecifics via multivalent binding or clustering of targets on a single cell (compare for example to bispecific tandem-fusion described by Veggiani et al Biochemistry Jan. 19, 2016, 113 (5) 1202-1207 or higher-order monospecific clustering Khairil Anuar et al, Nature Communications volume 10, Article number: 1734 (2019)).

[0111] The function of the structural domain is to provide a defined structural support for the binding domains. Advantageously, the structural domain can ensure that the binding domains have the desired orientation so that they can bind their targets, typically with both binding domains in the cis orientation. The constructs can therefore present a single binding surface.

[0112] In certain embodiments, the attachment site for the binding domains on the structural domain (oligomeric core) allows binding even for short linkers.

[0113] The structural domain may be any polypeptide domain comprising a defined secondary structure, typically an alpha helix or a beta sheet. In some particularly advantageous embodiments, the structural domain has its N and C termini in the same spatial region, for example substantially adjacent or adjacent to each other. Attaching the binding domains to the termini of the structural domain then provides the two binding domains substantially adjacent in the three-dimensional conformation. In some embodiments, the N and C termini are oriented to face in substantially the same direction. As described elsewhere herein, the provision of N and C termini that are adjacent in space can result in the binding domains being situated on the same face of the construct, or on the same face of the oligomer comprising multiple constructs. Accordingly, the constructs typically present a single binding surface. The constructs typically present the binding regions in cis oreientation.

[0114] The structural domains may comprise a single polypeptide chain, or may be formed of two or more separate polypeptide chains that associate to form a single structural domain, for example two anti-parallel (N—C C—N) alpha helices or two or more beta strands that associate to form a beta sheet. In some embodiments, two or more polypeptide chains with appropriate characteristics are identified and then fused, typically by recombinant means to form a single polypeptide chain (i.e. a fusion protein), but also by chemical conjugation or bonding to form a single covalent molecule.

[0115] The structural domain is different from the two binding domains. Therefore, when the binding domains are catcher polypeptides such as SpyCatcher, DogCatcher or SnoopCatcher, the structural domain is not a catcher polypeptide.

[0116] Typically, the structural domain does not comprise a CH2 domain. Typically, the structural domain does not comprise a CH3 domain. In some embodiments, the structural domain does not comprise a CH2 domain and does not comprise a CH3 domain.

[0117] A number of exemplary structural domains have been identified by the inventors, as discussed below and in the Examples. In some embodiments, the structural domain comprises or consists of the Collagen X NC1 domain (SEQ ID NO:2), or a polypeptide with at least 50%, at least 60%, at least 70% or at least 80%, for example at least 90% or at least 95% identity thereto. In some embodiments, the structural domain comprises or consists of the Collagen VIII NCI domain (SEQ ID NO:3), or a polypeptide with at least 50%, at least 60%, at least 70% or at least 80%, for example at least 90% or at least 95% identity thereto.

[0118] In some embodiments, the structural domain comprises or consists of a CutAl polypeptide (e.g. SEQ ID NO: 1 or SEQ ID NO:19), or a polypeptide with at least 50%, at least 60%, at least 70% or at least 80%, for example at least 90% or at least 95% identity thereto.

[0119] In certain embodiments, the structural element comprises or consists of a polypeptide with at least 50% amino acid identity, for example at least 90% identity, to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 19, SEQ ID NO: 29, SEQ ID NO: 60, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 42, SEQ ID NO: 31, or SEQ ID NO: 58.

[0120] The Collagen IV C4 domain (also known as the collagen IV NC1 domain) (PDB ID: 1m3d, SEQ ID NO: 49) is a suitable structural domain, or a polypeptide with at least 50%, at least 60%, at least 70% or at least 80%, for example at least 90% or at least 95% identity thereto.

[0121] Collagen NC1 domains can generally be used as structural domains according to the invention, including the NC1 domains from Collagen IV, Collagen VIII and Collagen X.

[0122] However, not all collagen NC1 domains are appropriate as structural domains. In particular NC1 domains from Collagen XV and from Collagen XVIII do not have the required orientation.

[0123] In certain embodiments, the structural domain comprises human macrophage migration inhibitory factor (MIF) (PDB ID: 1CA7, or SEQ ID NO: 25, or with Y99G mutation PDB ID: 60Y8) or human macrophage migration inhibitory factor 2 (MIF2) (PDB ID: 7MSE, or SEQ ID NO: 26, or with S62A and F99A mutation SED ID NO: 27) or a homolog or paralog thereof.

[0124] In certain embodiments, the structural domain comprises TNF family proteins including TNF (PDB ID: 1TNF, SEQ ID NO: 42), TLIA (PDB ID: 2re9, SEQ ID NO: 31) or CD40L (PDB ID: 31kj, SEQ ID NO: 58).

[0125] Further examples of structural domains include suitably modified antiparallel coiled-coil hexamer (PDB ID: 5WOJ, see example 4, SEQ ID NO: 43), HIV-1 GP41 core (PDB ID: 115Y or SEQ ID NO: 44), cytochrome c555 (PDB ID: 5Z25 or SEQ ID NO: 45), MHC Class II associated chaperonin and targeting protein invariant chain (Ii) (PDB ID: liie or SEQ ID NO: 46), p53 (PDB ID: 1C26 or SEQ ID NO: 47); a fibrinogen-like domain (PDB ID: 4M7F or SEQ ID NO: 48), a *Bacillus subtilis* AbrB (PDB ID: 1YFB or SEQ ID NO: 50), bacteriophage lambda head protein D (e.g. PDB ID: 1C5E or PDB ID: 1C5E or SEQ ID NO: 51); the domain-swapped trimer variant of HCRBPII (PDB ID: 6VIS or SEQ ID NO: 52); the TIL reovirus attachment protein sigmal (chain A,B,C of PDB ID: 40 DB or SEQ ID NO: 53).

[0126] A polypeptide construct according to the invention comprises a first binding domain and a second binding domain, in addition to the structural domain.

[0127] In certain embodiments, the binding domains are able to form an isopeptide linkage with a cognate peptide, for example the various catcher domains that are well-known in the art. Constructs comprising these isopeptide bond-forming domains are particularly well-suited to screening of different pairs of effector molecules such as antigen binding proteins. As discussed elsewhere herein, many combinations of effector molecules can be linked to the constructs comprising isopeptide bond forming binding domains, via isopeptide-forming peptide tags. Accordingly, the constructs comprising binding domains able to form an isopeptide linkage with a cognate peptide are particularly useful as a drug discovery platform.

[0128] The aspects of the invention relating to isopeptide bond-formation are generally exemplified with reference to a larger molecule (domain), typically referred to as a catcher, attached to the structural domain and a smaller polypeptide or peptide, typically referred to as the tag, forming part of the binding region (e.g. antigen-biding domain) of interest. However, all aspects and embodiments can be performed in the reverse orientation wherein the larger (e.g. catcher) molecule forms part of the binding region (e.g. antigen-binding domain) of interest and the smaller tag peptide forms the binding domain attached to the structural domain.

[0129] In some embodiments, the first binding domain and the second binding domain in the polypeptide construct are effector molecules such as antigen-binding domains. In these embodiments, the constructs are particularly suited for use as diagnostic, analytical or therapeutic agents.

[0130] In some embodiments, an interesting or effective pair of antigen-binding regions is identified using the drug discovery platform of the invention (e.g, wherein the construct comprises isopeptide bond-forming binding domains) and the construct is then expressed without the isopeptide bond-forming binding regions, and with the identified combination of antigen-binding domains (or other effector moiety) connected directly to a structural domain without the intermediary catcher domains on the structural domain and without the peptide tags on the antigen-binding domains. For the avoidance of doubt, these direct fusion constructs may still comprise a linker region between the terminal residue of the structural domain and the terminal residue of the or each effector moiety (e.g. antigen-binding region), as described in detail elsewhere herein.

[0131] Accordingly, one aspect of the invention provides a system for large-scale high-throughput screening of many possible combinations of effector molecules using combinatorial pairs of tagged effector proteins, and the combinations identified as useful can be used in the form that they were provided in the screening construct or converted into a simpler format (e.g. for therapeutic candidates) by creating a direct-fusion of the effector molecules (e.g. antigen-binding regions) onto the same structural domain as was used in the drug discovery platform. This provides a simple, fast and reliable technology to identify and develop bispecific and multispecific agents.

[0132] Antigen-binding domains are a typical domain that can be used and applied according to the invention. In certain aspects of the invention, antigen-binding domains comprise a peptide tag that can form an isopeptide bond, such as a Spy Tag or a SnoopTag, and can be bound by an isopeptide bond to a construct comprising a cognate catcher domain, for example to create a platform for a combinatorial or modular screen. In other aspects of the invention, the construct of the invention comprises a first antigen-binding domain at the N terminus and a second antigen-binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain. There may optionally be a linker sequence between one or each of the antigen-binding domains and the structural domain. As discussed below, suitable peptide linkers for use in connecting a binding domain (binding site) to a structural domain (monomer subunit) are typically between 1 and 100, 1 and 50, 1 and 25, 1 and 20, 1 and 15, or 1 to 10 amino acids in length. An example of a linker sequence is GSGS, GGGGS, GGGGSGGGGS, or GGGGSGGGGGGGGS.

[0133] The antigen-binding domain is typically an antigen-binding fragment of an antibody. An antigen-binding antibody fragment is not a full-length intact antibody, and typically lacks at least the CH2 and/or CH3 domains. Such antigen-binding fragments are well-known, and include a Fab, F(ab') 2, Fv, or a single chain Fv fragment (scFv). Antigen binding fragments typically comprise the CDRs (typically six CDRs) required for antigen binding, and the framework residues necessary for correct CDR structure. In some embodiments, the antigen-binding domain comprises

a heavy (H) chain variable domain sequence (VH), and a light (L) chain variable domain sequence (VL).

[0134] In some embodiments, the antigen-binding region can be a single domain antibody. Single domain antibodies (sdAb) can include antibodies whose complementary determining regions are part of a single domain polypeptide. Examples include, but are not limited to, heavy chain antibodies, antibodies naturally devoid of light chains, single domain antibodies derived from conventional 4-chain antibodies, engineered antibodies and single domain scaffolds other than those derived from antibodies. Single domain antibodies may be any of the art, or any future single domain antibodies. Single domain antibodies may be derived from species including, but not limited to mouse, human, camel, llama, fish, shark, goat, rabbit, and bovine. A single domain antibody may be a naturally occurring single domain antibody known as heavy chain antibody devoid of light chains. Such single domain antibodies are disclosed in WO-A-94/04678, for example. For clarity reasons, this variable domain derived from a heavy chain antibody naturally devoid of light chain is sometimes referred to as a VHH or nanobody to distinguish it from the conventional VH of four chain immunoglobulins. Such a VHH molecule can be derived from antibodies raised in Camelidae species, for example in camel, llama, dromedary, alpaca and guanaco. Other species besides Camelidae may produce heavy chain antibodies naturally devoid of light chain; such VHHs are within the scope of the invention.

[0135] The antigen-binidng domain may also comprise or consist of an antibody mimetic, such as an affibody or a DARPin. An affibody is, as known in the art, is a small polypeptide containing three alpha helices typically with around 58 amino acids and having a molecular mass of about 6 kDa. As is known in the art, a DARPIN (designed ankyrin repeat protein) is a genetically engineered antibody mimetic protein typically exhibiting highly specific and high-affinity target protein binding.

[0136] The binding domain may comprise naturally occurring ligands, such as cytokines, as an alternative to an antigen-binding domain.

[0137] When two different antigen binding domains are present on a bispecific molecule, they typically bind to different epitopes. This may be different epitopes on the same target molecule, or may be different epitopes on different target molecules. In some embodiments, the epitopes are both on therapeutic targets, wherein binding of an antigen-binding domain to the therapeutic target modifies a biological mechanism, typically a pathological mechanism, for therapeutic benefit.

[0138] In some embodiments, the antigen binding domains may each agonise a biological target. In some embodiments, the antigen binding domains may each antagonise a biological target. In some embodiments, one antigen-binding domain may agonise a first target and the other antigen-binding domain may antagonise a second target.

[0139] In some embodiments, a construct may comprise two binding regions that bind to the same epitope but have different affinities for that epitope. In some embodiments, a construct may comprise two binding regions that bind to the same epitope, optionally with different affinities, and the two binding regions have a different format, for example one antigen-binding region is an scFv and the second is a Fab.

[0140] Multi-domain polypeptide constructs of the invention typically have the following format, depicted in the N to C orientation according to the usual convention:

[0141] (Binding domain 1)-Linker1-Structural Domain-Linker2-(Binding domain 2) wherein Linker1 and Linker2 are optional linker sequences, optionally between 1 and 20 amino acids, for example GSGS. An optional purification tag, for example a His tag, for example 6xHis, can be incorporated at either end of the construct.

[0142] Binding domains may be the same or typically are different. Binding domains may typically be catcher polypeptides, or antigen-binding domains. A number of illustrative constructs are provided below.

[0143] SpC-Linker1-CutAl-Linker2-SpC
[0144] SnC-Linker1-CutAl-Linker2-SnC
[0145] SnC-Linker1-CutAl-Linker2-SpC
[0146] SpC-Linker1-CutAl-Linker2-SnC
[0147] SpC3-Linker1-CutAl-Linker2-DgC
[0148] scFv-Linker1-CutAl-Linker2-ScFv
[0149] Fab-Linker1-CutAl-Linker2-Fab
[0150] ScFv-Linker1-CutAl-Linker2-Fab
[0151] Fab-Linker1-CutAl-Linker2-ScFv
[0152] nanobody 1-Linker1-CutAl-Linker2-nanobody2
[0153] nanobody-Linker1-CutAl-Linker2-DgC
[0154] SpC3-Linker1-CutAl-Linker2-nanobody

wherein SpC is SpyCatcher, SpC3 is SpyCatcher003, SnC is SnoopCatcher, Fab is an antibody "fragment antigen binding" and scFv is a single chain Fv. In each case, the Linker1 and Linker2 linkers are optional. The CutAl sequences may be human or from *Pyrococcus horikoshii*, or a homologue from another species, or have at least 30%, at least 50%, at least 70% or at least 90% identity to the human or *Pyrococcus horikoshii* sequene.

[0155] Further illustrative constructs are:

[0156] SpC-Linker1-NC1-Linker2-SpC
[0157] SnC-Linker1-NC1-Linker2-SnC
[0158] SnC-Linker1-NC1-Linker2-SpC
[0159] SpC-Linker1-NC1-Linker2-SnC
[0160] scFv-Linker1-NC1-Linker2-ScFv
[0161] Fab-Linker1-NC1-Linker2-Fab
[0162] ScFv-Linker1-NC1-Linker2-Fab
[0163] Fab-Linker1-NC1-Linker2-ScFv
[0164] nanobody 1-Linker1-NC1-Linker2-nanobody2
[0165] nanobody-Linker1-NC1-Linker2-DgC
[0166] SpC3-Linker1-NC1-Linker2-nanobody

wherein NC1 is a collagen NC1 domain from Collagen VIII or Collagen X. In each case, the Linker1 and Linker2 linkers are optional.

[0167] Further illustrative constructs comprise the macrophage migration inhibitory factor (MIF) (SEQ ID NO: 25) or macrophage migration inhibitory factor 2 (MIF2) (SEQ ID NO: 26) or S62A F99A mutant of MIF2 (MIF2m, SEQ ID NO: 27) as the structural domain, including:

[0168] SpC-Linker1-MIF2-Linker2-SpC
[0169] SnC-Linker1-MIF2-Linker2-SnC
[0170] SnC-Linker1-MIF2-Linker2-SpC
[0171] SpC-Linker1-MIF2-Linker2-SnC
[0172] scFv-Linker1-MIF2-Linker2-ScFv
[0173] Fab-Linker1-MIF2-Linker2-Fab
[0174] ScFv-Linker1-MIF2-Linker2-Fab
[0175] Fab-Linker1-MIF2-Linker2-ScFv
[0176] nanobody 1-Linker1-MIF2-Linker2-nanobody2
[0177] nanobody-Linker1-MIF2-Linker2-DgC
[0178] SpC3-Linker1-MIF2-Linker2-nanobody

[0179] Any suitable structural domain, in particular any of the structural domains described herein, may be used in these constructs instead of the exemplary structural domains provided above. Accordingly, these exemplary formats are described for use with structural domains in general and with the structural domains described herein.

[0180] The orientation of the binding domains in the multi-domain construct, in monomeric or oligmeric form, can be assessed functionally using a variety of assays. A selection of exemplary assays are described below.

[0181] FRET: To demonstrate cis-orientation of the selected scaffolds compared to non-cis-oriented proteins, a FRET assay can be performed. The scaffold polypeptide with Catcher components, for example SpC3-HsCutAl-DgC, can be conjugated to fluorescent protein FRET pairs fused to the respective Tag pairs, for example mCherry (6+)-SpT3-H6 and H6-DgT-mCitrine (4-). Upon conjugation of the tagged FRET pairs to the scaffold protein, the emission of the acceptor FRET protein can be measured via standard fluorescence reading and compared to the sensitised emission of the donor FRET protein. Protein scaffolds that show preferential cis-orientation will show higher acceptor emission, whereas protein scaffolds that show preferential trans-orientation will show higher donor sensitised emission.

[0182] SPR: To demonstrate that cis-oriented scaffolds conjugated to suitable ligands may preferentially bind to targets on the same plane compared to non-cis-oriented proteins, an SPR experiment can be performed. Target proteins against the scaffold-conjugated ligands, for example the targets against L1 and L2 in SpC-PhCutAl-SnC: SnT-LI: L2-SpT, can be immobilized on the surface of the SPR sensor chip, either together, or separately as L1-target only or L2-target only for controls. Cis-oriented or non-cis-oriented scaffolds conjugated to L1 and L2 are then loaded to the SPR sensor chip with the L1- and/or L2-targets and the conjugated assemblies' binding to the immobilised targets on the chip is determined. Assemblies with cis-orientation is expected to show highly measurable binding to both L1- and L2-targets when both targets are immobilised on the same chip, whereas assemblies with non-cis-orientation is expected to not have highly measurable binding to such chip. Both assembly types are expected to have highly measurable binding to immobilised L1- or L2-targets only.

[0183] SEC-MALS: To demonstrate the native oligomeric state of the scaffold and assembly proteins in solution, a SEC-MALS experiment can be performed. Scaffold and assembly proteins can be prepared as described in the methods section. The samples are then injected into an FPLC machine coupled to a MALS machine and detector to separate the samples by size and the native protein mass is approximated by calculations of light scattering. The oligomeric state of the proteins can then be derived by dividing the native protein mass by predicted monomeric mass calculated through softwares such as ProtParam. Both scaffold and assembly proteins, for example SpC-PhCutAl-SnC and SpC-PhCutAl-SnC: SnT-L1: L2-SpT, are expected to show oligomeric state of 3.

[0184] Crosslinking and LC-MS/MS: To demonstrate the simultaneous binding of two targets by a cis-oriented conjugated assembly, for example the targets against L1 and L2, targets-expressing cells can be incubated with a biotinylated version of the conjugated assembly, biotin-SpC-PhCutAl-SnC: SnT-L1: L2-SpT, and subsequently the binding between targets and biotinylated-assembly can be crosslinked with BS3 (bis (sulfosuccinimidyl) suberate), followed by cell lysis and the extraction of the crosslinked target-assembly complex via streptavidin. This complex is then tryptic-digested to be feed into LC-MS/MS to confirm the binding of L1 and L2 to their respective targets. The same methodology can be run for non-cis-oriented assemblies, whereby the output of the LC-MS/MS data can be compared between the two protein assemblies. It is expected that cis-oriented assemblies may preferentially bind to both L1- and L2-targets, whereas non-cis-oriented assemblies may preferentially bind to either L1- or L2-targets only.

Multivalent Protein Scaffold

[0185] One aspect of the invention relates to a modular system for screening target molecules. The system allows for the multivalent presentation of the target molecules. In one aspect, a multivalent protein scaffold is provided. The multivalent protein scaffold comprises an oligomeric core comprising a plurality of subunit monomers. The multivalent protein scaffold also comprises at least one first binding site orthogonal to at least one second binding site. Suitable binding sites are described in more detail herein.

[0186] The scaffold acts as a platform to which other molecules may be bound. Different combinations of molecules may be bound to the scaffold in a modular fashion. The scaffold allows multivalent binding of the molecules. Generally, the molecules bound to the scaffold have a potential therapeutic benefit, and the scaffold bound to the molecules can be used to investigate whether multivalent assemblies of different molecules may have a desired effect. For example, different combinations of potential anti-cancer polypeptides may be attached to the scaffold and the resulting assembly can then be used in a screening assay to see if the combination has an effect on, such as binding to and causing the death of, a cancer cell. Once a combination of molecules has been identified, a therapeutic drug candidate may be produced by modifying the multivalent protein scaffold so that it is directly attached to the identified molecules, rather than by using a modular system. The multivalent protein scaffold 'presents' the molecules on the same face of the scaffold, thereby allowing all of the molecules to potentially interact with a target cell.

[0187] The provided scaffold typically comprises at least two first binding sites and at least two second binding sites. The presence of at least two first and at least two second binding sites allows the scaffold to be used to screen for multivalent interactions, that may not necessarily be seen when screening with, for example, a bispecific antibody format. It also allows to test different scaffolds as a "toolbox" with regards to the same combination. Accordingly, provided herein is a multivalent protein scaffold comprising:

- **[0188]** an oligomeric core comprising a plurality of subunit monomers; and
- **[0189]** at least two first binding sites orthogonal to at least two second binding sites;
- **[0190]** wherein said first binding sites and said second binding sites are positioned on the same face of the scaffold.

[0191] The provided scaffold typically comprises binding sites capable of forming covalent bonds to their respective targets. Covalent bonds lead to strong irreversible association. Complexes generated by covalently attaching the scaffold of the invention to targets for the binding sites are

physically robust and can be readily produced. Such complexes can be produced in high yields and with high homogeneity. Accordingly the biological response produced when such complexes are administered to a biological system such as a subject as described herein are reproducible and controllable.

[0192] Accordingly, also provided herein is a multivalent protein scaffold comprising:

[0193] an oligomeric core comprising a plurality of subunit monomers;

[0194] at least one first binding site orthogonal to at least one second binding site;

[0195] wherein said first binding site(s) and said second binding site(s) are positioned on the same face of the scaffold; and wherein said first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target; and said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target.

[0196] As explained herein, the scaffold provided herein has significant advantages over conventional antibodies including bispecific antibodies. The oligomeric core of the provided scaffold typically does not comprise an Fc region of an antibody. In some embodiments, the oligomeric core does not comprise a CH2 region. In some embodiments, the oligomeric core does not comprise CH3 region. In some embodiments, the oligomeric core does not comprise a CH2 region and does not comprise a CH3 region. The use of immunoglobulin domains, typically constant domains such as in an Fc region, of an antibody, typically may not display the advantages of the provided scaffold described herein. For example, a bispecific antibody lacks the modularity of the present invention and may not be useful for investigating multivalent interactions.

[0197] Accordingly, also provided herein is a multivalent protein scaffold comprising:

[0198] an oligomeric core comprising a plurality of subunit monomers;

[0199] at least one first binding site orthogonal to at least one second binding site;

[0200] wherein said first binding site(s) and said second binding site(s) are positioned on the same face of the scaffold; and wherein said oligomeric core does not comprise an Fc region of an antibody.

[0201] Also provided herein is a multivalent protein scaffold comprising:

[0202] an oligomeric core comprising a plurality of subunit monomers;

[0203] at least one first binding site orthogonal to at least one second binding site;

[0204] wherein said first binding site(s) and said second binding site(s) are positioned on the same face of the scaffold; and wherein said oligomeric core does not comprise a CH2 domain of an antibody, or does not comprise a CH3 domain of an antibody, or does not comprise a CH2 domain and does not comprise a CH3 domain.

[0205] The multivalent protein scaffold comprises an oligomeric core and at least one first binding site and at least one second binding site. The multivalent protein scaffold may also comprise other features, such as linkers, domain insertions and/or functional groups as described in more detail herein.

[0206] Preferably, the diameter of the multivalent protein scaffold is less than less than about about 100 nm, e.g. less than about 50 nm, e.g. less than about 25 nm, e.g. less than about 10 nm. Preferably, the height of the multivalent protein scaffold is less than less than about 100 nm, e.g. less than about 50 nm, such as less than about 30 nm, e.g. less than about 20 nm for example less than about 10 nm. The multivalent protein scaffold is preferably from 1 to 500 Å in size, such as from 2 to 250 Å, e.g. from about 10 to about 100 Å, such as from about 20 to about 80 Å.

[0207] Preferably, the multivalent protein scaffold itself preferably essentially does not induce an immune response in a biological system, cell culture or subject, such as in a human subject. In other words, typically in the absence of binding sites on the oligomeric core and/or effector moieties attached to the binding sites of the protein scaffold, no immune response (or essentially no immune response; e.g an immune response no greater than when a non-immunogenic protein is administered) is induced when the protein scaffold is administered to a biological system such as a human subject. For example, administration of the protein scaffold (in the absence of binding sites on the multivalent protein scaffold and/or effector moieties attached to the binding sites of the multivalent protein scaffold) to a biological system (e.g. a subject as defined herein) typically does not induce a biological system's innate or adaptive immunity. For example, the protein scaffold typically does not induce activation of the complement system, B cells, T cells, natural killer cells, mast cells, basophils, eosinophils, neutrophils, dendritic cells or macrophages.

[0208] The multivalent protein scaffold preferably does not comprise an antibody or antibody fragment; although as described further herein antibodies and/or antibody fragments may be attached as effector moieties to the scaffold. The multivalent protein scaffold (e.g. absent any effector moieties) more preferably does not comprise an Fc region of an antibody. An Fc region is the tail region of an antibody that interacts with cell surface receptors called Fc receptors and some proteins of the complement system. In some cases, the multivalent protein scaffold does not comprise an immunoglobulin constant region. In some embodiments, the multivalent protein scaffold does not comprises a CH2 domain. In some embodiments, the multivalent protein scaffold does not comprises a CH3 domain. In some embodiments, the multivalent protein scaffold does not comprises a CH2 domain and does not comprise a CH3 domain.

[0209] Preferably the multivalent protein scaffold is thermodynamically stable. Preferably the multivalent protein scaffold is stable at a temperature of from about 0 to about 100° C., e.g. from about 4° C. to about 90° C., such as from about 10° C. to about 50° C. e.g. from about 20 to about 38° C. e.g., from about 25 to about 37° C. In other words, preferably the multivalent protein scaffold does not dissociate into its substituent subunit monomers, and/or the binding sites do not disassociate from the oligomeric core, when in aqueous solution at a temperature of from about 0 to about 100° C. (e.g. from about 4° C. to about 90° C., such as from about 10° C. to about 50° C. e.g. from about 20 to about 38° C. e.g., from about 25 to about 37° C.); for example, at least 90%, such as at least 95%, e.g. at least 99%, such as at least 99.9%, e.g. at least 99.99% or 99.999% of the multivalent protein scaffold does not dissociate into its substituent subunit monomers, and/or the binding sites do not disassociate from the oligomeric core, when in aqueous

solution at such a temperature. More preferably the multivalent protein scaffold is stable at a temperature of from about 0 to about 100° C., e.g. from about 4° C. to about 90° C., such as from about 10° C. to about 50° C., e.g. from about 20 to about 38° C. e.g., from about 25 to about 37° C. The multivalent protein scaffold preferably has a lifetime of at least 10 minutes, more preferably at least one hour, e.g. at least one day, such as at least one week, e.g. at least one month or at least one year, when determined at a temperature of from about 0° C. to about 100° C., e.g. from about 4° C. to about 90° C., such as from about 10° C. to about 50° C. e.g. from about 20 to about 38° C. e.g., from about 25 to about 37° C.

[0210] Preferably, the interactions between the constituents of the multivalent protein scaffold is not a weak transient interaction. Weak transient complexes show a dynamic mixture of different oligomeric states in vivo, whereas strong transient complexes change their quaternary state only when triggered by, for example, ligand binding. Weak transient interactions are characterized by a dissociation constant (K_D) in the micromolar range and lifetimes of seconds. Strong transient interactions, stabilized by binding of an effector molecule, may have a longer lifetime and have a lower K_D in the nanomolar range. The constituents of the multivalent protein scaffold preferably interact with at least strong transient reactions, more preferably the constituents of the multivalent protein scaffold form a permanent interaction. A permanent interaction means that the multivalent protein scaffold does not disassociate into its constituents under normal conditions, for example between 20° C. and 40° C. and between pH 6 and pH 8. A multivalent protein scaffold in which the constituent parts form a permanent interaction typically disassociates only under denaturing conditions that denature the tertiary structure of the subunit monomers themselves.

[0211] Preferably, therefore, the constituents of the multivalent protein scaffold interact with a K_D of less than 1 μM, e.g. less than 100 nM, more preferably less than 10 nM, at a temperature of from about 0° C. to about 100° C., e.g. from about 4° C. to about 90° C., such as from about 10° C. to about 50° C. e.g. from about 20 to about 38° C. e.g., from about 25 to about 37° C.

[0212] Preferably, the multivalent protein scaffold and the constituent parts thereof is stable to proteases. For example, the multivalent protein scaffold and the constituents of the multivalent protein scaffold may be exposed to proteases, such as trypsin, without loss of the tertiary or quaternary structure of the scaffold. The multivalent protein scaffold and the constituent parts of the multivalent protein scaffold are stable to proteases for at least 1 hour, e.g. at least 2 hours, e.g. at least 4 hours, e.g. at least 8 hours, such as at least 24 hours, or more, at a temperature of from about 10 to about 40° C., e.g. from about 20 to about 38° C. e.g., from about 25 to about 37° C.

Oligomeric Core

[0213] As explained above, the multivalent protein scaffold provided herein comprises an oligomeric core comprising a plurality of subunit monomers. A subunit monomer of the oligomeric core is typically the structural domain of the polypeptide construct as described elsewhere herein.

[0214] Any suitable number of subunit monomers may be used. For example, the oligomeric core may comprise from about 2 to about 20 subunit monomers, e.g. from about 2 to about 10 subunit monomers, more preferably 3 to 7 subunit monomers, more preferably 3 to 6 subunit monomers. For example, the oligomeric core may comprise two, three, four, five, six, seven, eight, nine or 10 subunit monomers. Preferably, the oligomeric core comprises at least 3 subunit monomers. Most preferably the oligomeric core comprises three subunit monomers. Preferably, the oligomeric core does not comprise or consist of 7 subunit monomers.

[0215] Preferably, the subunit monomers have rotational symmetry when multimerised, such as three-fold rotational symmetry, four-fold rotational symmetry, five-fold rotational symmetry, six-fold rotational symmetry, or seven-fold rotational symmetry. The oligomeric core may have C2, C3, C4, D2, C5, C6, D3, C7, C8, D4, C9, C10, D5, C11, C12, D6 or T symmetry.

[0216] Some or all of the subunit monomers in the oligomeric core may be non-covalently attached together. Some or all of the subunit monomers in the oligomeric core may be covalently attached together. The oligomeric core may comprise a mix of subunit monomers attached covalently and non-covalently. For example, the oligomeric core may comprise a first monomer covalently bound to a second monomer to form a heterodimer. The oligomeric core may comprise at least two such heterodimers non-covalently bound together. For example, the oligomeric core may comprise three non-heterodimers non-covalently attached together, wherein each heterodimer comprises two monomers covalently bound together.

[0217] Some or all of the subunit monomers in the oligomeric core may be attached by non-covalent interactions. Suitable non-covalent interactions include, but are not limited to, electrostatic interactions, such as ionic bonds, hydrogen bonds and halogen bonds, Van der Waals forces, such as dipole-dipole interactions, x-x stacking, cation-x interactions, anion-π interactions or polar-x interactions.

[0218] Some or all of the subunit monomers in the oligomeric core may be covalently attached together. When subunit monomers are covalently attached together, the monomer is typically the amino acid sequence corresponding to the original or naturally occurring monomeric domain.

[0219] Two or more subunit monomers may be covalently linked by disulphide bonds. Disulphide bonds typically form between cysteine residues in polypeptides. Artificial amino acids having free thiol groups may also participate in disulphide bond formation.

[0220] Two or more subunit monomers may be covalently attached via chemical cross linking. Cross linking reagents include homobifunctional crosslinking reagents, heterobifunctional crosslinking reagents, and photoreactive crosslinking reagents. Homobifunctional crosslinking reagents have identical reactive groups at either end. Examples of homobifunctional crosslinking reagents include disuccinimidyl suberate (DSS), disuccinimidyl tartrate (DST) and dithiobis succinimidyl propionate (DSP). Some common examples of sulfhydryl-to-sulfhydryl crosslinkers include BMOE and DTME. Heterobifunctional crosslinking reagents possess two different reactive groups and can be used to link dissimilar functional groups. Examples of heterobifunctional crosslinking reagents include MDS (m-Maleimidobenzoyl-N-hydroxysuccinimide ester), GMBS (N-y-Maleimidobutyryloxysuccinimide ester), EMCS (N-(ε-Maleimidocaproyloxy) succinimide ester) and sulfo-EMCS (N-(ε-Maleimidocaproyloxy) sulfo succinim-

ide ester). Photoreactive crosslinking reagents are heterobifunctional crosslinkers that become reactive only upon exposure to ultraviolet or visible light. Two classes of common photoreactive chemical groups are aryl-azides and diazirines. Aryl azides (N-((2-pyridyldithio)ethyl)-4-azidosalicylamide) are widely used. Upon exposure to 250-350 nm UV light, these reagents can facilitate the formation of a nitrene group that may set off an addition reaction with the double bonds. Additionally, these crosslinkers may initiate the production of C—H insertion products or react with a nucleophile. Some common crosslinking reagents that belong to this group include ANB-NOS (N-5-Azido-2-nitrobenzyloxysuccinimide) and Sulfo-SANPAH. NHS-ester diazirines or azipentanoates contain a photoactivatable diazirine ring and an N-hydroxysuccinimide (NHS) ester which efficiently reacts with primary amino groups in neutral to basic buffers to form stable amide bonds. They exhibit better photostability compared to the phenyl azide group and can be easily activated with long-wave ultraviolet light (330 to 370 nm) to produce carbene intermediates that form covalent bonds with any peptide backbones or amino acid side chains within the spacer arm distance.

[0221] More preferably, two or more subunit monomers in the oligomeric core may be genetically fused together. Subunit monomers are genetically fused together if they are expressed are encoded in a single polynucleotide sequence such that they are expressed in a single polypeptide chain. Accordingly, when the subunit monomers are genetically fused together, the oligomeric core may comprise a single polypeptide chain.

[0222] Genetically fused subunit monomers may be genetically fused together via peptide linkers. Suitable peptide linkers for use in linking subunit monomers are amino acid sequences include those that can act as a hinge region between subunit monomers, thus allowing them to fold independently from one another and providing sufficient flexibility to allow the subunit monomers to retain their ability to multimerise. The length, flexibility and hydrophilicity of the peptide linker are typically designed such that subunit monomers can readily assemble to form the oligomeric core. Preferably, subunit monomers linked by peptide linkers can assemble to form an oligomeric core wherein the interaction between adjacent subunit monomers is substantially identical to the interaction between the same subunit monomers when not.

[0223] Suitable peptide linkers for use in connecting monomer subunits of the oligomeric core are typically between 1 and 100, 1 and 50, 1 and 25, 1 and 20, 1 and 15, or 1 to 10 amino acids in length. The linkers may, for example, be composed of one or more of the following amino acids: lysine, serine, arginine, proline, glycine and alanine. Examples of suitable flexible peptide linkers are stretches of 2 to 20, such as 4, 6, 8, 10 or 16, serine and/or glycine amino acids. Examples of rigid linkers are stretches of 2 to 30, such as 4, 6, 8, 16 or 24, proline amino acids. Examples of suitable linkers include, but are not limited to, the following: GGGS, PGGS, PGGG, RPPPPP, RPPPP, VGG, RPPG, PPPP, RPPG, PPPPPPPPP, PPPPPPPPPPPPPPPPG, GGG, SG, SGSG, SGSGSG, SGSGSGSG, SGSGSGSGSG and SGSGSGSGSGSGSGSG wherein G is glycine, P is proline, R is arginine, S is serine and Vis valine. Additional exemplary linkers include GSGS, GGGGS, GGGGSGGGGS and GGGGSGGGGSGGGGS. Appropriate linking groups may be designed using conventional modelling techniques. The linker is typically sufficiently flexible to allow the monomers, or subunits thereof, to assemble into their respective protein oligomers.

[0224] Preferably, the total molecular weight of the oligomeric core is less than about 1000 kDa, such as less than about 500 kDa e.g. less than about about 250 kDa. The total molecular weight of the oligomeric core is preferably from about 10 kDa to about 1000 kDa, such as from about 10 kDa to about 500 kDa, e.g from about 10 kDa to about 250 kDa, such as from about 10 kDa to about 150 kDa. The total molecular weight of the oligomeric core is more preferably from about 20 kDa to about 150 kDa.

[0225] Preferably, the diameter of the oligomeric core is less than less than about 100 nm, e.g. less than about 50 nm, such as less than about 30 nm, e.g. less than about 20 nm for example less than about 10 nm. Preferably, the height of the oligomeric core is less than less than about 100 nm, e.g. less than about 50 nm, such as less than about 30 nm, e.g. less than about 20 nm for example less than about 10 nm. The oligomeric core is preferably from 1 to 50 nm in size, such as from 2 to 40 nm e.g. from about 2 to about 20 nm, such as from about 5 to about 10 nm.

[0226] Preferably the oligomeric core is thermodynamically stable. Preferably the oligomeric core is stable at a temperature of from about 0 to about 50° C. In other words, preferably the oligomeric core does not spontaneously dissociate into its substituent monomers when in solution at a temperature of from about 0 to about 50° C. More preferably the oligomeric core is stable at a temperature of from about 10 to about 40° C., e.g. from about 20 to about 38° C. e.g., from about 25 to about 37° C.

[0227] Preferably, the subunit monomers stably interact to form the oligomeric core. The interactions between subunit monomers is preferably not a weak transient interaction. Weak transient complexes typically show a dynamic mixture of different oligomeric states in vivo, whereas strong transient complexes change their quaternary state only when triggered by, for example, ligand binding and exist in a single predominant oligomeric state (e.g. at least 90%, such as at least 95%, e.g. at least 99%, such as at least 99.9%, e.g. at least 99.99% or 99.999% of the complex may exist in a stable oligomeric state under standard conditions). Weak transient interactions are characterized by a dissociation constant (K_D) in the micromolar range and lifetimes of seconds. Strong transient interactions, stabilized by binding of an effector molecule, typically have a longer lifetime and have a lower K_D in the nanomolar range. The subunit monomers more preferably interact with at least strong transient reactions, more preferably form a permanent interaction. A permanent interaction means that the oligomer does not or substantially does not disassociate into its constituent subunit monomers under normal conditions (for example when in aqueous solution at a temperature of from about 0 to about 100° C.; under these conditions typically at least 90%, such as at least 95%, e.g. at least 99%, such as at least 99.9%, e.g. at least 99.99% or 99.999% of the oligomeric core does not dissociate), usually the oligomeric core only disassociates when denaturing conditions are used that denature the tertiary structure of the subunit monomers themselves. Accordingly, preferably, the subunit monomers multimerise with a K_D of less than 1 μM, e.g. less than 100 nM, more preferably less than 10 nM. The oligomeric core typically has a lifetime of at least 10 minutes, more prefer-

ably at least one hour, e.g. at least one day, such as at least one week, e.g. at least one month or at least one year. Lifetime may be determined at any suitable temperature, such as from about 0° C. to about 100° C., e.g. from about 4° C. to about 90° C., such as from about 10° C. to about 50° C. e.g. from about 20 to about 38° C. e.g., from about 25 to about 37° C.

[0228] Preferably, the oligomeric core is stable to proteases. For example, the oligomeric core may be exposed to dilute concentrations of a protease, such as trypsin, for a limited period of time, such as 4 hours, without loss of the tertiary of quaternary structure of the oligomeric core.

[0229] Preferably the oligomeric core is human, or humanised. A human oligomeric core is a multimeric region of a human protein. A humanised oligomeric core is an oligomeric core which is a multimeric region of a non-human protein, which has been modified to more closely resemble the corresponding multimeric region of a human protein. A humanised oligomeric core may comprise at least 50% amino acid identity to the amino acid sequence of the multimeric region of the corresponding human protein, such as at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98% or at least 99% amino acid identity. The corresponding multimeric region of a human protein is the multimeric region of a human protein with the greatest amino acid sequence identity to the humanised oligomeric core. This information can be identified using a Blast search (blast.ncbi.nlm.nih.gov) and limiting the search results to *Homosapiens*.

[0230] Preferably, the oligomeric core of the multivalent protein scaffold does not itself induce an immune response in a biological system, cell culture or subject (such as a non-human or human subject). In other words, typically in the absence of binding sites on the oligomeric core and/or effector moieties attached to the binding sites of the oligomeric core, no immune response is induced when the oligomeric core is administered to a biological system. For example, administration of the oligomeric core (in the absence of binding sites on the oligomeric core and/or effector moieties attached to the binding sites of the oligomeric core) to a biological system (e.g. a subject as defined herein) typically does not induce a biological system's innate or adaptive immunity. For example, the oligomeric core typically does not induce activation of the complement system, B cells, T cells, natural killer cells, mast cells, basophils, eosinophils, neutrophils, dendritic cells or macrophages. However, molecules attached to the oligomeric core may be designed specifically to generate an immune response in a human subject.

[0231] Preferably, the oligomeric core of the protein scaffold does not comprise an antibody or antibody fragment; although as described further herein antibodies and/or antibody fragments may be attached as effector moieties to the oligomeric core. The oligomeric core (e.g. absent any effector moieties) more preferably does not comprise an Fc region of an antibody. In some cases, the oligomeric core does not comprise an immunoglobulin constant region.

[0232] In some embodiments the oligomeric core of the multivalent protein scaffold may be a homooligomeric core, i.e. the oligomeric core may comprise only one type of monomer. The homooligomeric core may comprise two or more, such as three or more, four or more, five or more, six or more, or seven or more, identical subunit monomers.

[0233] In some other embodiments the oligomeric core of the multivalent protein scaffold may be a heterooligomeric core, i.e. the oligomeric core may comprise more than one type of monomer. The different types of subunit monomer are capable of forming an oligomeric core. In other words, the subunit monomers are capable of attaching together. The heterooligomeric core may comprise two or more, such as three or more, four or more, five or more, six or more, or seven or more, different subunit monomers. For example, if a heterooligomeric core comprises three monomers, it may comprise two first monomers and one second monomer; or it may comprise one first monomer, one second monomer, and one third monomer. If a heterooligomeric core comprises four monomers it may comprise two first monomers and two second monomers; two first monomer, one second monomer and one third monomer; or one first monomer, one second monomer, one third monomer and one fourth monomer. Preferably, when the oligomeric core is a heterooligomeric core, the oligomeric core comprises two types of subunit monomer; i.e. for a heterooligomeric core comprising monomers of type A and type B, and having n subunit monomers in total, said heterooligomer core may have a stoichiometry of (AaBb), wherein a +b=n. When the oligomeric core is a heterooligomeric core, the subunit monomers comprised therein may be modified such that a first type of subunit monomer binds preferentially to a second type of subunit monomer rather than another monomer of the first type (in other words, for a heterooligomeric core comprising monomers of type A and monomers of type B, the heterooligomeric core is of the form ABABAB . . . rather than AAABBB . . .).

[0234] In still further embodiments the oligomeric core may comprise a plurality of multimeric subunits. For example, two monomers may be fused (e.g. as a tandem fusion) and the monomer fusions may assemble to form the oligomeric core. In this regard, the fused monomers may be the same or different.

[0235] For example, two or more identical monomers may be fused, with the fusion product assembling with further identical fusion products to form a homooligomeric core. The oligomeric core may comprise a plurality of homodimers, for example wherein the homodimer is 'AA', the oligomeric core may comprise 'AA', 'AAAA', or 'AAAAAA' etc.

[0236] Alternatively, two or more different monomers may be fused together, with the fusion product assembling with further identical fusion products to form an oligomeric core. As used herein, such a core is typically considered as a homooligomeric core wherein each fusion product is considered as the monomer subunit. The oligomeric core may comprise a plurality of heterodimers, for example wherein the heterodimer is 'AB', the oligomeric core may comprise 'AB', 'ABAB', or 'ABABAB' etc.

[0237] Alternatively, two or more identical monomers may be fused, with the fusion product assembling with further non-identical fusion products to form a heterooligomeric core. The oligomeric core may comprise a plurality of homodimers, for example wherein a first homodimer is 'AA' and a second homodimer is 'BB', the oligomeric core may comprise 'AABB', etc.

[0238] Alternatively, two or more different monomers may be fused together, with the fusion product assembling with further non-identical fusion products to form a heterooligomeric core. The oligomeric core may comprise a plu-

rality of heterodimers, for example wherein a first heterodimer is 'AB' and a second heterodimer is 'CD', the oligomeric core may comprise 'ABCD', etc.

[0239] A homooligomeric core comprises a plurality of subunit monomers, wherein each monomer comprises at least one first binding site and at least one second binding site. For example, if a homooligomeric core comprises three subunit monomers, the oligomeric core will comprise at least three first binding sites and at least three second binding sites. If the homooligomeric core comprises four, five, six or seven subunit monomers, the oligomeric core will comprise at least four first binding sites and at least four second binding sites; at least five first binding sites and at least five second binding sites; at least six first binding sites and at least six second binding sites; or at least seven first binding sites and at least seven second binding sites; respectively. Thus, the multivalent protein scaffold preferably comprises at least two first binding sites and at least two second binding sites, i.e, wherein all of first binding sites are the same as the other first binding sites, and all of the second binding sites are the same as the other second binding sites. The multivalent protein scaffold more preferably comprises at least three first binding sites and at least three second binding sites. In some embodiments the multivalent protein scaffold comprises at least four, at least five, at least six, at least seven, or at least eight of each first and second binding sites.

[0240] A heterooligomeric core comprises a plurality of subunit monomers comprising at least two types of subunit monomer, wherein one type of subunit monomer comprises at least one first binding site and a second type of subunit monomer comprises at least one second binding site. For example, if a heterooligomeric core comprises three subunit monomers, it may comprise three different binding sites, or it may comprise two first binding sites and one second binding site. If a heterooligomeric core comprises four subunit monomers, it may comprise four different binding sites, or it may comprise two first binding sites and one second binding site and one third binding site, or two first binding sites and two second binding site.

[0241] Binding sites are described in more detail herein.

Monomers

[0242] As explained herein, the multivalent protein scaffold provided herein comprises an oligomeric core comprising a plurality of subunit monomers. A subunit monomer is typically the structural domain of the multi-domain polypeptide construct as described elsewhere herein.

[0243] Each subunit monomer (excluding any binding site(s) attached thereto, as described in more detail herein) preferably comprises less than 300 amino acids, preferably less than 200 amino acids, more preferably less than 150 amino acids. For example, each subunit monomer (excluding any binding site(s) attached thereto) preferably has a molecular weight of less than 40 kDa, such as less than 30 kDa, such as less than 20 kDa. Protein scaffolds as described herein which comprise such monomers may be of relatively low mass allowing efficient diffusion in vivo. They are typically capable of being expressed and correctly folded in bacterial cell expression or yeast cell expression systems. Such expression systems can often yield far higher yields than mammalian cell cultures typically required to produce antibodies.

[0244] The subunit monomers preferably do not comprise or consist of an antibody or antibody fragment. The oligomeric core or subunit monomer preferably does not comprise or consist of an Fc region of an antibody. In some embodiments, the subunit monomer does not comprise or consist of a CH2 domain. In some embodiments, the subunit monomer does not comprise or consist of a CH3 domain. In some embodiments, the subunit monomer does not comprise a CH2 domain and does not comprise a CH3 domain.

[0245] Preferably each monomer subunit of the oligomeric core is human, or humanised. A human monomer is a monomer of a human oligomeric protein. A humanised monomer is a monomer of a non-human oligomeric protein, which has been modified to more closely resemble a monomer of the corresponding human protein. A humanised monomer may thus comprise at least 50% amino acid identity to the amino acid sequence of the corresponding human protein, such as at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98% or at least 99% amino acid identity. Typically, a human or humanised protein will not cause a deleterious immune response in a patient to which it is administered.

[0246] The subunit monomers comprised in the oligomeric core preferably each comprise a multimerising structural element, which is the structural and/or functional features of the subunit monomer that allow the subunit monomers to multimerise.

[0247] The multimerising structural element may be a protein domain. The multimerising structural element of the oligomeric core may be a multimerisation domain of a naturally-occurring multimeric protein, or a de novo multimeric domain. A protein domain is an autonomously folding unit of a protein. A multimerisation domain is typically a protein domain that is involved in protein-protein interactions with another protein domain. A multimerising structural element is preferably soluble, such that the monomer and oligomeric core is soluble.

[0248] In view of the disclosure herein, those skilled in the art will be able to identify suitable multimerising structural elements for use in the present invention. For example, the skilled person can identify multimeric proteins. For example, numerous multimeric proteins are listed in databases such as the NCBI databases (www.ncbi.nlm.nih.gov) and the Protein Data Bank (PDB; www.rscb.org) which can be searched for multimeric proteins having rotational symmetry axes.

[0249] Preferably, the multimeric proteins identified are homooligomers, such as homodimers, homotrimers, homotetramers, homopentamers, homohexamers, homoheptamers and so on. The multimeric proteins may be heterooligomers, such as heterodimers, heterotrimers, heterotetramers, heteropentamers, heterohexamers, heteroheptamers and so on. Functional and or structural information can be used to identify which domains of the multimeric protein are responsible for multimerisation, i.e. the multimerisation domains

[0250] The multimerising structural element preferably comprises the multimerisation interface of the multimerisation domains (i.e. the structural or functional element of the multimerisation domain that allows the domains to multimerise). Other aspects of the multimerisation domains may be modified without affecting the multimerisation of the subunit monomers. Thus, the subunit monomers of the oligomeric core preferably comprise a multimerising struc-

tural element. The subunit monomers preferably comprise at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% amino acid identity to the multimerisation domain, from which they are derived. The subunit monomers retain the ability to form a multimer (i.e. an oligomeric core).

[0251] Preferably, the subunit monomers of the oligomeric core comprise a soluble multimerising structural element of a multimeric protein. Soluble domains are preferred, over, for example, multimerisation domains found within membranes. Preferably, each of the subunit monomers of the oligomeric core comprise a soluble multimerising structural element of a soluble multimeric protein.

[0252] A multimerising structural element may be derived from any multimeric protein of suitable symmetry (e.g. rotational or dihedral symmetry, as described in more detail herein), such as a collagen (e.g. a collagen NC1 domain), a CutA, a TNF, a p53, a fibrinogen, a C4, *Bacillus subtillus* AbrB or a homolog or paralog thereof.

[0253] Preferably, a subunit monomer may comprise a monomer of, or the multimerisation domain of, a protein selected from: collagen X (PDB ID: 1GR3) (e.g. the NC1 domain thereof), collagen VIII (PDB ID: 1091) (e.g. the NC1 domain thereof), a Clq head domain (for example, PDB ID: 1PK6 is the globular head of human Clq), or a CutA (copper tolerance A) protein such as the CutAl proteins from *Pyrococcus horikoshii, Homo sapiens* (PDB ID: 2ZFH), *Thermus* thermophiles (PDB ID: 1V6H); *Oryza sativa* (PDB ID: 2ZOM); or *Shewanella* sp. SIB1 (PDB ID: 3AHP); or a polypeptide having at least 30% or at least 50% amino acid sequence identity to any one of the preceding polypeptides; more preferably at least 60% amino acid sequence identity, at least 70%, at least 80%, at least 90%, at least 95% or at least 97%, at least 98% or at least 99% amino acid sequence identity to any one of the preceding polypeptides.

[0254] In some embodiments, a subunit monomer may comprise or consist of collagen X (PDB ID: 1GR3, or SEQ ID NO: 2) (e.g. the NC1 domain thereof), collagen VIII (PDB ID: 1091, or SEQ ID NO: 3) (e.g. the NCI domain thereof), a heteromeric Clq head domain (for example, PDB ID: 1PK6 is the globular head of human Clq, see SEQ ID NO: 36-38), or a CutA (copper tolerance A) protein such as the CutAl proteins from *Pyrococcus horikoshii* (PDB ID: 4YNO, or SEQ ID NO: 1), *Homo sapiens* (PDB ID: 2ZFH, or SEQ ID NO: 19), *Thermus thermophilus* (PDB ID: 1V6H, or SEQ ID NO: 39), *Oryza sativa* (PDB ID: 2ZOM, or SEQ ID NO: 40); or *Shewanella* sp. SIB1 (PDB ID: 3AHP, or SEQ ID NO: 41); TNF-like protein TL1A (PDB ID: 2RE9, or SEQ ID NO: 31); TNF (PDB ID: 1TNF, or SEQ ID NO: 42); TNF family protein CD40L (PDB ID: 3LKJ, or SEQ ID NO: 58); human macrophage migration inhibitory factor (MIF) (PDB ID: 1CA7, or SEQ ID NO: 25, or with Y99G mutation PDB ID: 60Y8); human macrophage migration inhibitory factor 2 (MIF2) (PDB ID: 7MSE, or SEQ ID NO: 26, or with S62A and F99A mutation SED ID NO: 27) or a homolog or paralog thereof.

[0255] Other multimerising domains include the multimerising domains of: antiparallel coiled-coil hexamer (PDB ID: 5WOJ, see example 4, SEQ ID NO: 43), HIV-1 GP41 core (PDB ID: 115Y or SEQ ID NO: 44), cytochrome c555 (PDB ID: 5Z25 or SEQ ID NO: 45), MHC Class II associated chaperonin and targeting protein invariant chain (Ii) (PDB ID: liie or SEQ ID NO: 46); p53 (PDB ID: 1C26 or SEQ ID NO: 47); a fibrinogen-like domain (PDB ID: 4M7F or SEQ ID NO: 48); a Collagen IV C4 (PDB ID: 1LI1 or SEQ ID NO: 49); a *Bacillus subtilis* AbrB (PDB ID: 1YFB or SEQ ID NO: 50); or a polypeptide having at least 50% amino acid sequence identity to any one of the preceding polypeptides; more preferably at least 60% amino acid sequence identity, such as at least 70%, at least 80%, at least 90%, at least 95% or at least 97%, at least 98% or at least 99% amino acid sequence identity to any one of the preceding polypeptides.

[0256] Other multimerising domains include the multimerising domains of: bacteriophage lambda head protein D (e.g. PDB ID: 1C5E or PDB ID: 1C5E or SEQ ID NO: 51); the domain-swapped trimer variant of HCRBPII (PDB ID: 6VIS or SEQ ID NO: 52); the TIL reovirus attachment protein sigmal (chain A,B,C of PDB ID: 40 DB or SEQ ID NO: 53); or a polypeptide having at least 50% amino acid sequence identity to any one of the preceding proteins; more preferably at least 60% amino acid sequence identity, such as at least 70%, at least 80%, at least 90%, at least 95% or at least 97%, at least 98% or at least 99% amino acid sequence identity to any one of the preceding proteins.

[0257] Preferably, in one embodiment, the oligomeric core may comprise monomers derived from the multimerising structural element of *Pyrococcus horikoshii* CutAl. Accordingly, the or each subunit monomer may comprise or consist of a polypeptide having at least 30%, such as at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99% or 100%, amino acid identity to the amino acid sequence of SEQ ID NO: 1.

[0258] As discussed above and elsewhere herein, CutAl (e.g. *Pyrococcus horikoshii*) is a typical structural domain of the multi-domain polypeptide construct.

[0259] Preferably, in another embodiment, the oligomeric core may comprise monomers derived from the multimerising structural element of collagen X NC1. Accordingly, the or each subunit monomer may comprise or consist of a polypeptide having at least 30%, such as at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99% or 100%, amino acid identity to the amino acid sequence of SEQ ID NO: 2.

[0260] As discussed above and elsewhere herein, Collagen X NC1 is a typical structural domain of the multi-domain polypeptide construct.

[0261] Preferably, in another embodiment, the oligomeric core may comprise monomers derived from the multimerising structural element of collagen VIII NC1. Accordingly, the or each subunit monomer may comprise or consist of a polypeptide having at least 30%, such as at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99% or 100%, amino acid identity to the amino acid sequence of SEQ ID NO: 3.

[0262] As discussed above and elsewhere herein, Collagen VIII NC1 is a typical structural domain of the multi-domain polypeptide construct.

[0263] Preferably, in another embodiment, the oligomeric core may comprise monomers derived from the multimerising structural element of CutAl from *Homo sapiens*. Accordingly, the or each subunit monomer may comprise or consist of a polypeptide having at least 30%, such as at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least

95%, at least 97%, at least 98%, at least 99% or 100%, amino acid identity to the amino acid sequence of SEQ ID NO: 19.

[0264] As discussed above and elsewhere herein, CutAl (e.g human) is a typical structural domain of the multi-domain polypeptide construct.

[0265] Preferably, in another embodiment, the oligomeric core may comprise monomers derived from the multimerising structural element of MIF or MIF-2. Accordingly, the or each subunit monomer may comprise or consist of a polypeptide having at least 30%, such as at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99% or 100%, amino acid identity to the amino acid sequence of or SEQ ID NO: 25 or SEQ ID NO: 26 or SEQ ID NO: 27.

[0266] As discussed above and elsewhere herein, MIF is a typical structural domain of the multi-domain polypeptide construct. As discussed above and elsewhere herein, MIF-2 is a typical structural domain of the multi-domain polypeptide construct.

[0267] Preferably, in another embodiment, the oligomeric core may comprise monomers derived from the multimerising structural element TNF family proteins including TNF and

[0268] TNF-like TLIA or CD40L. Accordingly, the or each subunit monomer may comprise or consist of a polypeptide having at least 30%, such as at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99% or 100%, amino acid identity to the amino acid sequence of SEQ ID NO: 42 or SEQ ID NO: 31 or SEQ ID NO: 58

[0269] As discussed above and elsewhere herein, TNF is a typical structural domain of the multi-domain polypeptide construct. As discussed above and elsewhere herein, TNF-like protein are typical structural domains of the multi-domain polypeptide construct.

[0270] When the oligomeric core is a heterooligomeric core, each monomer in the oligomeric core may be derived from the same protein. For example, each monomer in the oligomeric core may be derived from one of the proteins described above. The oligomeric core may be heteromeric due to differences in the binding sites attached to the monomer subunits, even if the multimerising domains of each monomer subunit are identical. The oligomeric core may be heteromeric due to differences in the multimerising domains of the monomer subunits, even if all monomer subunits are derived from the same protein.

[0271] Those skilled in the art will appreciate that the monomer subunits of the oligomeric core of the multivalent protein scaffold provided herein may be further altered to provide additional functionality or beneficial properties.

[0272] For example, a monomer may be a fragment, derivative or variant of a monomer or multimerising structural element described herein. As those skilled in the art will appreciate, fragments of amino acid sequences include deletion variants of such sequences wherein one or more, such as at least 1, 2, 5, 10, 20, 50 or 100 amino acids are deleted. Deletion may occur at the C-terminus or N-terminus of the native sequence or within the native sequence. Typically, deletion of one or more amino acids does not influence the residues immediately surrounding the multimerising structural element of a subunit monomer.

[0273] Derivatives of amino acid sequences include post-translationally modified sequences including sequences which are modified in vivo or ex vivo. Many different protein modifications are known to those skilled in the art and include modifications to introduce new functionalities to amino acid residues, modifications to protect reactive amino acid residues or modifications to couple amino acid residues to chemical moieties such as reactive functional groups on linkers or substrates (surfaces) for attachment to such amino acid residues.

[0274] Derivatives of amino acid sequences also include addition variants of such sequences wherein one or more, such as at least 1, 2, 5, 10, 20, 50 or 100 amino acids are added or introduced into the native sequence. Addition may occur at the C-terminus or N-terminus of the native sequence or within the native sequence. Typically, addition of one or more amino acids does not influence the residues immediately surrounding the multimerising structural element of a subunit monomer.

[0275] Variants of amino acid sequences include sequences wherein one or more amino acid such as at least 1, 2, 5, 10, 20, 50 or 100 amino acid residues in the native sequence are exchanged for one or more non-native residues. Such variants can thus comprise point mutations or can be more profound e.g. native chemical ligation can be used to splice non-native amino acid sequences into partial native sequences to produce variants of native enzymes. Variants of amino acid sequences include sequences carrying naturally occurring amino acids and/or unnatural amino acids.

[0276] Variants, derivatives and functional fragments of the aforementioned amino acid sequences typically retain the ability of the wild-type sequence to oligomerise. Preferably, variants, derivatives and functional fragments of the aforementioned sequences have improved properties, such as increased stability, reduced toxicity, additional functionalities including binding sites, etc, compared to the wild-type or native sequence.

Binding Sites

[0277] The multivalent protein scaffold comprises at least one first binding site and at least one second binding site. In some embodiments, typically in the modular system useful to identify useful combinations of effector molecules in drug discovery. the at least one first binding site is orthogonal to the at least one second binding site. In other words, the chemistry by which the first binding site binds to its target (the first target) it orthogonal to the chemistry by which the second binding site binds to its target (the second target). Thus, the first target will bind to the first binding site but will not bind to the second binding site; and the second target will bind to the second binding site but will not bind to the first binding site. Accordingly, the term orthogonal is given its usual meaning in the field of protein-protein interactions, and the first binding interaction (i.e. the first binding site and the first ligand) is independent of the second binding interaction (i.e. the second binding site and the second ligand).

[0278] The binding sites of the multivalent protein scaffold allows the scaffold to be used as a modular system to bind effector moieties. The first and second binding sites bind to their cognate target on the effector moieties.

[0279] The first and second binding sites may be incorporated into the multivalent protein scaffold provided herein in any suitable manner. In one embodiment the first and second binding sites are provided as a tandem fusion which is attached as described herein to a or each monomer of the oligomeric core to form the multivalent protein scaffold. For

comparison, SEQ ID NO: 22 is an example of two binding sites (described herein) provided as a fusion linked by an &H linker.

[0280] Binding sites are described in more detail below.

[0281] The interaction between a binding site and its target may be a non-covalent interaction. Preferably, the or each binding site can form a covalent bond to its respective target. A reactive functional group may be present naturally in the subunit monomer or effector moiety, or may be introduced, e.g. by genetic manipulation or by chemical modification of the monomer. The reactive group may originate from a non-natural amino acid incorporated into the monomer during its synthesis or expression, e.g. during cell-free expression, e.g. via in vitro transcription/translation.

[0282] A binding site on the multivalent protein scaffold may bind to its target via a reactive group. Any suitable reactive group can be used. For example, a reactive group may be an amine-reactive group; a carboxyl-reactive group; a sulfhydryl-reactive group or a carbonyl-reactive group. A reactive group may comprise a cysteine-reactive group. A reactive group may comprise a maleimide, an azide, a thiol, an alkyne, an NHS ester or a haloacetamide.

[0283] A reactive group may be a group capable of reacting with a non-natural amino acid such as 4-azido-L-phenylalanine (Faz) and any one of the amino acids numbered 1-71 in FIG. **1** of Liu C. C. and Schultz P. G., Annu. Rev. Biochem., 2010, 79, 413-444. Such groups are particularly useful when corresponding non-natural amino acids are comprised in the binding site and the cognate target.

[0284] A reactive group may be a click chemistry group. Click chemistry is a term first introduced by Kolb et al. in 2001 to describe an expanding set of powerful, selective, and modular building blocks that work reliably in both small- and large-scale applications (Kolb HC, Finn, MG, Sharpless KB, Click chemistry: diverse chemical function from a few good reactions, Angew. Chem. Int. Ed. 40 (2001) 2004-2021). They have defined the set of stringent criteria for click chemistry as follows: "The reaction must be modular, wide in scope, give very high yields, generate only inoffensive by-products that can be removed by non-chromatographic methods, and be stereospecific (but not necessarily enantioselective). The required process characteristics include simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation. Purification if required must be by non-chromatographic methods, such as crystallization or distillation, and the product must be stable under physiological conditions".

[0285] For example, the first and second binding sites may comprise orthogonal click chemistry reagents. Suitable examples of click chemistry include, but are not limited to, the following:

[0286] (a) copper-free variant of the 1,3 dipolar cycloaddition reaction, where an azide reacts with an alkyne under strain, for example in a cyclooctane ring;

[0287] (b) the reaction of an oxygen nucleophile on one linker with an epoxide or aziridine reactive moiety on the other;

[0288] (c) the Staudinger ligation, where the alkyne moiety can be replaced by an aryl phosphine, resulting in a specific reaction with the azide to give an amide bond;

[0289] (d) nitrone dipole cycloaddition;

[0290] (e) norbornene cycloaddition;

[0291] (f) oxanorbornadiene cycloaddition;

[0292] (g) tetrazine ligation;

[0293] (h) [4+1] Cycloaddition;

[0294] (i) tetrazole photoclick chemistry; and

[0295] (j) quadricyclane ligation.

[0296] A reactive group may be a haloacetamide, for example, iodoacetamide, bromoacetemide or chloroacetamide.

[0297] A reactive group may be selected from a vinyl group, TCO, tetrazine and a strained alkyne; DBCO; an activated acid e.g. an acid chloride; and piperazine and reactive amines. Host-guest chemistry can also be used to provide the reaction between a binding site and its target. For example, a binding site may comprise a ligand for binding to a metal complex, and the target comprises a metal complex, or vice-versa. Thus, a binding site may comprise a metal complex which can interact non-covalently via chelation or supramolecular association with its target containing a site that can act as a ligand to complex with the modifier molecule by forming a stable association; or vice-versa.

[0298] A reactive group may be any of those disclosed in Sakamoto and Hamachi, "*Recent progress in chemical modification of proteins*", Anal. Sci 2019 (35) 5-27; or Mckay and Finn, "*Click chemistry in complex mixtures: bioorthogonal bioconjugation*", Chem. Biol. 2014, 21 (9) 1075-1101, both of which are hereby incorporated by reference in their entirety.

[0299] A binding site of the multivalent protein scaffold preferably comprises a polypeptide, such as a protein domain. More preferably, the first binding site comprises a first protein domain and said second binding site comprises a second protein domain.

[0300] When the first binding site comprises a first protein domain and the second binding site comprises a second protein domain, the first binding site and/or the second binding site is preferably genetically fused to the subunit monomer(s) to which they are attached to form a single polypeptide chain. Typically, the first binding site and/or the second binding site is expressed as a single polypeptide chain with the subunit monomer(s) to which they are attached, for example as a fusion protein from a recombinant nucleic acid molecule. This can be beneficial in that the multivalent protein scaffold can be expressed ready for binding effector moieties without further chemical modification needed to, for example, attach a click chemistry reagent. The attachment between a protein binding site and the protein to which it is attached, e.g. the monomer subunit of the oligomeric core of the multivalent protein scaffold, is described below.

[0301] The first binding site may comprise a first protein domain capable of forming a non-covalent bond to a first polypeptide target; and said second binding site may comprise a second protein domain capable of forming a non-covalent bond to a second polypeptide target. More preferably, the first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target; and said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target. Any suitable covalent bond can be formed, with examples above. Preferably, the first protein domain is capable of forming an isopeptide bond with the

first polypeptide target and the second protein domain is capable of forming an isopeptide bond with the second binding target. An isopeptide bond is an amide bond that can form for example between the carboxyl group of one amino acid and the amino group of another. At least one of these joining groups is typically part of the side chain of one of these amino acids.

[0302] Preferably, the first binding site and the second binding site each comprise a different split protein domain, such as a split ligand-binding protein domain. As used herein a ligand-binding protein domain is a domain of a protein-binding ligand. Any suitable protein can be used, however proteins which natively are stabilised by an intra-strand covalent bond such as an isopeptide bond are particularly beneficial. In such cases, a portion of the protein containing the isopeptide bond donor residue is split from the portion of the peptide containing the isopeptide bond receiver residue. The two protein fragments can be attached, e.g. by genetic fusion, to further polypeptides such as a monomer of an oligomeric core and/or a polypeptide target as described herein. Contacting the two separate fragments leads to the creation of the isopeptide bond which attaches, typically irreversibly, the two fragments together. Accordingly, the split protein approach for producing binding sites and complementary tags is particularly useful. Pairs of such binding site/tags are typically orthogonal as the fragment of one protein will bind preferentially or solely to its native partner (i.e. the complementary portion of the protein from which it was derived) over any other potential partner. These principles are discussed in e.g. Reddington & Howarth, Curr. Op. Chem. Biol. 29, 94-99 (2015) and Keeble et al, PNAS 2019 116 (52) 26523.

[0303] Preferably, one of said first binding site and said second binding site comprises a split *Streptococcus pyogenes* fibronectin-binding protein domain and the other of said first binding site and said second binding site comprises a split *Streptococcus pneumoniae* adhesin domain.

[0304] Preferably, the first protein domain and the first polypeptide target, and the second protein domain and the second polypeptide target, may each comprise a peptide linker pair, such as those disclosed in WO 2016/193746 A1, WO 2018/197854 A1, WO 2018/189517 A1, Keeble et al. (PNAS 116 (52), 2019:26523-26533), Fierer et al. (PNAS 111 (13), 2014: E1176-E1181).

[0305] Preferably, said first and said second binding site each independently have at least 50% amino acid identity to any one of SEQ ID NOs: 4-9, 11-13, 23 or 15-18.

[0306] Preferably, said first binding site/polypeptide target pair and the second binding site/polypeptide target pair are each independently selected from the following combinations: (i) any one of SEQ ID NO: 4, 6 or 8 with any one of SEQ ID NOs: 5, 7 or 9; (ii) SEQ ID NO: 12 with SEQ ID NO: 13 or 15; (iii) SEQ ID NO: 5 with SEQ ID NO: 11; (iv) SEQ ID NO: 15 with SEQ ID NO: 16); (v) SEQ ID NO: 17 with SEQ ID NO: 18; or (vi) SEQ ID NO: 23 with SEQ ID NO: 16).

[0307] More preferably, the first protein domain and the first polypeptide target, and the second protein domain and the second polypeptide target are each independently selected from the following pairs:

Binding group	Target	
SpyCatcher (SEQ ID NO: 4)	SpyTag (SEQ ID NO: 5)	
SpyCatcher (SEQ ID NO: 4)	SpyTag002 (SEQ ID NO: 7)	
SpyCatcher (SEQ ID NO: 4)	SpyTag003 (SEQ ID NO: 9)	
SpyTag (SEQ ID NO: 5)	SpyCatcher (SEQ ID NO: 4)	
SpyTag002 (SEQ ID NO: 7)	SpyCatcher (SEQ ID NO: 4)	
SpyTag003 (SEQ ID NO: 9)	SpyCatcher (SEQ ID NO: 4)	
SpyCatcher002 (SEQ ID NO: 6)	SpyTag002 (SEQ ID NO: 7)	
SpyCatcher002 (SEQ ID NO: 6)	SpyTag (SEQ ID NO: 5)	
SpyCatcher002 (SEQ ID NO: 6)	SpyTag003 (SEQ ID NO: 9)	
SpyTag002 (SEQ ID NO: 7)	SpyCatcher002 (SEQ ID NO: 6)	
SpyTag (SEQ ID NO: 5)	SpyCatcher002 (SEQ ID NO: 6)	
SpyTag003 (SEQ ID NO: 9)	SpyCatcher002 (SEQ ID NO: 6)	
SpyCatcher003 (SEQ ID NO: 8)	SpyTag003 (SEQ ID NO: 9)	
SpyCatcher003 (SEQ ID NO: 8)	SpyTag (SEQ ID NO: 5)	
SpyCatcher003 (SEQ ID NO: 8)	SpyTag002 (SEQ ID NO: 7)	
SpyTag003 (SEQ ID NO: 9)	SpyCatcher003 (SEQ ID NO: 8)	
SpyTag (SEQ ID NO: 5)	SpyCatcher003 (SEQ ID NO: 8)	
SpyTag002 (SEQ ID NO: 7)	SpyCatcher003 (SEQ ID NO: 8)	
SpyTag (SEQ ID NO: 5)	K-tag (SEQ ID NO: 11)	Binding mediated by
K-tag (SEQ ID NO: 11)	SpyTag (SEQ ID NO: 5)	SpyLigase: (SEQ ID NO: 10)
SnoopCatcher (SEQ ID NO: 12)	SnoopTag (SEQ ID NO: 13)	
SnoopTag (SEQ ID NO: 13)	SnoopCatcher (SEQ ID NO: 12)	
SnoopCatcher (SEQ ID NO: 12)	SnoopTagJr (SEQ ID NO: 15)	
SnoopTagJr (SEQ ID NO: 15)	SnoopCatcher (SEQ ID NO: 12)	
SnoopTagJr (SEQ ID NO: 15)	DogTag (SEQ ID NO: 16)	Binding mediated by
DogTag (SEQ ID NO: 16)	SnoopTagJr (SEQ ID NO: 15)	SnoopLigase (SEQ ID NO: 14)
Pilin-C (SEQ ID NO: 17)	IsopepTag (SEQ ID NO: 18)	
IsopepTag (SEQ ID NO: 18)	Pilin-C (SEQ ID NO: 17)	
DogCatcher (SEQ ID NO: 23)	DogTag (SEQ ID NO: 16)	

The protein domain and the targeting domain may have at least 50% amino acid identity, such as at least 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or 100% amino acid identity with the sequences set forth above, whilst retaining the ability of the protein domain to specifically bind to the targeting domain.

[0308] In some embodiments, the first binding site is a protein domain and the first target is a tag which binds to the first protein domain; and the second binding site is a protein domain and the second target is a tag which binds to the second protein domain. In some embodiments, the first binding site is a tag and the first target is a protein domain which binds to the first tag; and the second binding site is a tag and the second target is a protein domain which binds to the second tag. In some embodiments, the first binding site is a protein domain and the first target is a tag which binds to the first protein domain; and the second binding site is a tag and the second target is a protein domain which binds to the second tag. Preferably both the first and second binding sites are protein domains and the first and second targets are tags which specifically bind to the first and second protein domains, respectively.

[0309] More preferably the first protein domain and the first polypeptide target, and the second protein domain and the second polypeptide target are each independently selected from the following pairs:

Binding group	Target	
SpyCatcher (SEQ ID NO: 4)	SpyTag (SEQ ID NO: 5)	
SpyCatcher (SEQ ID NO: 4)	SpyTag002 (SEQ ID NO: 7)	
SpyCatcher (SEQ ID NO: 4)	SpyTag003 (SEQ ID NO: 9)	
SpyCatcher002 (SEQ ID NO: 6)	SpyTag002 (SEQ ID NO: 7)	
SpyCatcher002 (SEQ ID NO: 6)	SpyTag (SEQ ID NO: 5)	
SpyCatcher002 (SEQ ID NO: 6)	SpyTag003 (SEQ ID NO: 9)	
SpyCatcher003 (SEQ ID NO: 8)	SpyTag003 (SEQ ID NO: 9)	
SpyCatcher003 (SEQ ID NO: 8)	SpyTag (SEQ ID NO: 5)	
SpyCatcher003 (SEQ ID NO: 8)	SpyTag002 (SEQ ID NO: 7)	
SpyTag (SEQ ID NO: 5) K-tag (SEQ ID NO: 11)	K-tag (SEQ ID NO: 11) SpyTag (SEQ ID NO: 5)	Binding mediated by SpyLigase: (SEQ ID NO: 10)
SnoopCatcher (SEQ ID NO: 12)	SnoopTag (SEQ ID NO: 13)	
SnoopCatcher (SEQ ID NO: 12)	SnoopTagJr (SEQ ID NO: 15)	
SnoopTagJr (SEQ ID NO: 15) DogTag (SEQ ID NO: 16)	DogTag (SEQ ID NO: 16) SnoopTagJr (SEQ ID NO: 15)	Binding mediated by SnoopLigase (SEQ ID NO: 14)
Pilin-C (SEQ ID NO: 17)	IsopepTag (SEQ ID NO: 18)	

The protein domain and the targeting domain may have at least 50% amino acid identity, such as at least 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or 100% amino acid identity with the sequences set forth above, whilst retaining the ability of the protein domain to specifically bind to the targeting domain.

[0310] The binding groups and targets above may be divided into the following subgroups: Subgroup A:

[0311] SpyCatcher (SEQ ID NO: 4)/SpyTag (SEQ ID NO:5);

[0312] Spy Catcher (SEQ ID NO: 4)/SpyTag002 (SEQ ID NO:7);

[0313] SpyCatcher (SEQ ID NO: 4)/SpyTag003 (SEQ ID NO:9);

[0314] SpyCatcher002 (SEQ ID NO: 6)/SpyTag (SEQ ID NO:5);

[0315] SpyCatcher002 (SEQ ID NO: 6)/SpyTag002 (SEQ ID NO:7);

[0316] SpyCatcher002 (SEQ ID NO: 6)/SpyTag003 (SEQ ID NO:9);

[0317] SpyCatcher003 (SEQ ID NO: 6)/SpyTag003 (SEQ ID NO:5);

[0318] SpyCatcher003 (SEQ ID NO: 6)/SpyTag003 (SEQ ID NO:7);

[0319] SpyCatcher003 (SEQ ID NO: 8)/SpyTag003 (SEQ ID NO:9);

[0320] SpyTag (SEQ ID NO:5)/K-tag (SEQ ID NO: 11) (mediated by SpyLigase: (SEQ ID NO: 10))

Subgroup B:

[0321] SnoopCatcher (SEQ ID NO: 12)/SnoopTag (SEQ ID NO: 13);

[0322] SnoopCatcher (SEQ ID NO: 12)/SnoopTagJr (SEQ ID NO: 15);

[0323] SnoopTagJr (SEQ ID NO: 15)/DogTag (SEQ ID NO: 16) (mediated by SnoopLigase (SEQ ID NO: 14)/-

[0324] DogCatcher (SEQ ID NO: 23)/DogTag (SEQ ID NO: 16)

Subgroup C:

[0325] Pilin-C(SEQ ID NO: 17)/IsopepTag (SEQ ID NO: 18)

[0326] Preferably, the first binding site/target pair is selected from subgroup A and the second binding site/target pair is selected from subgroup B and subgroup C; or the first binding site/target pair is selected from subgroup B and the second binding site/target pair is selected from subgroup A and subgroup C; or the first binding site/target pair is selected from subgroup C and the second binding site/target pair is selected from subgroup A and subgroup B.

[0327] Still more preferably, the first protein domain-polypeptide target pair, and the second protein domain-polypeptide target may be selected from the group consisting of (i) SpyCatcher/SpyTag and SnoopCatcher/SnoopTag; (ii) SpyCatcher002/SpyTag002 and SnoopCatcher/SnoopTag; (iii) SpyCatcher003/SpyTag003 and SnoopCatcher/SnoopTag; (iv) SpyCatcher/SpyTag and Pilin-C/Isopeptag; (v) SpyCatcher002/SpyTag002 and Pilin-C/Isopeptag; (vi) SpyCatcher003/SpyTag003 and Pilin-C/Isopeptag; (vii) Pilin-C/Isopeptag and SnoopCatcher/

SnoopTag; (viii) Spy Catcher/Spy Tag and SnoopTagJr/DogTag; (ix) SpyCatcher002/SpyTag002 and SnoopTagJr/DogTag; (x) SpyTag/K-tag and SnoopCatcher/SnoopTag; (xi) SpyTag/K-tag and SnoopTagJr/DogTag; (xii) SpyTag/K-tag and Pilin-C/Isopeptag; and (xiii) SnoopTagJr/DogTag and Pilin-C/Isopeptag; and (xiv) SpyCatcher003/SpyTag003 and DogCatcher/DogTag; and (xv) SpyCatcher003/SpyTag002 and DogCatcher/DogTag; and (xvi) SpyCatcher003/SpyTag and DogCatcher/DogTag; (xvii) SpyCatcher003/SpyTag003 and SnoopCatcher/SnoopTagJr; and (xviii) SpyCatcher003/SpyTag002 and SnoopCatcher/SnoopTagJr; and (xix) SpyCatcher003/SpyTag and SnoopCatcher/SnoopTagJr.

[0328] Those skilled in the art will appreciate that when SpyLigase/SpyTag/K-tag, or SnoopLigase/SnoopTagJr/DogTag are used the 'ligase' catalyses the attachment of the two tags. Accordingly, the first binding site and the first polypeptide target, and the second binding site and the second polypeptide target are selected from the two 'tags'. The ligase may be added exogenously to catalyse the attachment of the two 'tags', or may be associated with the multivalent protein scaffold, non-covalently or covalently, such as genetically fused with the multivalent protein scaffold. It is interchangeable which of the tags is/are comprised within the multivalent protein scaffold.

[0329] Other binding site/tag pairs include SdyTag/SdyCatcher (Tan et al, PLOS One 11 (1) e0165074) and the Cpe0147439-563/Cpe0147565-587 pair derived from *Clostridium perfringens* cell-surface adhesin protein Cpe0147 (Young et al, Chem Comm. 53 (9) 1502).

[0330] "Specifically binds" as used herein in the context of binding between a binding site and its target, refers to the ability of a binding site to bind to its complementary binding site with greater affinity than it binds to an unrelated control. The unrelated control may be an unrelated control protein. For example, SnoopCatcher specifically binds to SnoopTag with greater affinity than it binds to an unrelated control protein. The binding is preferably covalent, such as the formation of an isopeptide bond. Preferably, the control protein is bovine serum albumin, and the binding site binds to the complementary binding site with an affinity that is at least 10, at least 50, at least 100, at least 500, or at least 1000 times greater than the control protein. Affinity may be determined by methods known in the art. For example, affinity may be determined by ELISA assay, biolayer interferometry, surface plasmon resonance, kinetic methods or equilibrium/solution methods. The skilled person will recognize which pairs of binding sites specifically bind to produce a protein complex that can be used in the methods of the invention.

[0331] The at least one first binding site(s) and the at least one second binding site(s) preferably do not comprise an antibody or antibody fragment. The at least one first binding site(s) and the at least one second binding site(s) more preferably do not comprise an antigen binding fragment of an antibody, such as a Fab, or a Fc region.

[0332] In all discussions concerning a 'binding site' which binds to a 'target' herein, the skilled person will readily understand that the chemical binding groups are reversible, In other words, whereas binding above has been described in terms of reaction of a reactive group A at the binding site with a corresponding reactive group B of the target for that binding site, equivalent chemistry is possible wherein reactive group B at the binding site reacts with reactive group A of the target.

[0333] When the multivalent protein scaffold comprises one or more binding sites which are protein domains e.g. protein domains attached to monomer subunits of the oligomeric core of the multivalent protein scaffold, the protein domains may be attached to the multivalent protein scaffold (e.g. attached to monomer subunits of the oligomeric core of the multivalent protein scaffold) by any suitable means.

[0334] A binding site may be attached to the multivalent protein scaffold (e.g. attached to monomer subunits of the oligomeric core of the multivalent protein scaffold) by a linker. In one embodiment the same linker may be used at each terminus of a subunit monomer of the oligomeric core. In another embodiments a different linker may be used at each terminus of a subunit monomer of the oligomeric core.

[0335] The binding site is preferably covalently attached to the oligomeric core (or subunit monomer). The covalent linkage may for example be a peptide bond, a disulphide bond or a click chemistry linkage. More preferably, the covalent linkage comprises at least one amino acid, i.e. a peptide linker, and forms part of the same polypeptide chain as the subunit monomer at the binding site.

[0336] A peptide linker used to attach a binding site to the monomer subunit of the oligomeric core of the multivalent protein scaffold may be genetically fused to the subunit monomer and/or the binding site. A linker is genetically fused if the linker is expressed as a single construct with the subunit monomer and/or the binding site from a single polynucleotide coding sequence. The length, flexibility and hydrophilicity of the peptide linker are typically designed such that the binding sites may be positioned on the same face of the oligomeric core or multivalent protein scaffold. The peptide linker typically allows for directional tethering of the binding sites.

[0337] Suitable peptide linkers for use in connecting a binding site to a monomer subunit are typically between 1 and 100, 1 and 50, 1 and 25, 1 and 20, 1 and 15, or 1 to 10 amino acids in length. The linkers may, for example, be composed of one or more of the following amino acids: lysine, serine, arginine, proline, glycine and alanine. Examples of suitable flexible peptide linkers are stretches of 2 to 20, such as 4, 6, 8, 10 or 16, serine and/or glycine amino acids. Examples of rigid linkers are stretches of 2 to 30, such as 4, 6, 8, 16 or 24, proline amino acids. Examples of suitable linkers include, but are not limited to, the following: GGGS, PGGS, PGGG, RPPPPP, RPPPP, VGG, RPPG, PPPP, RPPG, PPPPPPPPP, PPPPPPPPPPPPPP SGSGSGSGSG and SGSGSGSGSGSGSGSG wherein G is glycine, P is proline, R is arginine, S is serine and V is valine. Other exemplary linkers include GSGS, GGGGS, GGGGSGGGGS and GGGGSGGGGSGGGGS. Appropriate linking groups may be designed using conventional modelling techniques. The linker is typically sufficiently flexible to allow the binding site and the monomer subunit to assemble into their respective protein oligomers.

[0338] The oligomeric core of the multivalent protein scaffold preferably comprises at least one first binding site at a terminus of a subunit monomer, and at least one second binding site at a terminus of a subunit monomer. For example, a first binding site may be positioned at the first

terminus of a subunit monomer, and a second binding site may be positioned at the second terminus of a subunit monomer.

[0339] The termini are preferably determined by reference to termini of the subunit monomers of the oligomeric core, not including any linker or binding site. More preferably, the terminus of a subunit monomer is selected from the N-terminus and/or the C-terminus of the subunit monomer. If a binding site (e.g. a protein domain) forms part of the same polypeptide as a subunit monomer, the N-terminus and C-terminus preferably relate to the amino acids corresponding to the respective termini of the monomer in the absence of the binding site. Similarly, if a linker forms part of the same polypeptide as a subunit monomer, the N-terminus and C-terminus preferably relate to the amino acids corresponding to the respective termini of the monomer in the absence of the linker.

[0340] Preferably, the termini of the subunit monomers to which the binding sites are attached are on the same face of the oligomeric core or the multimeric protein scaffold, as defined in more detail above.

[0341] In some cases, the oligomeric core only comprises a single terminus of each subunit monomer on a given face. In this case, the at least one first binding site and the at least one second binding site are typically both attached to the same terminus thereby being on the same face of the multivalent protein scaffold. For example, the oligomeric core may comprise a plurality of subunit monomers, wherein each subunit monomer comprises a first binding site attached at a first terminus of said monomer and a second binding site attached to said first binding site. The oligomeric core may comprise a plurality of subunit monomers, wherein the at least one first binding site(s) are attached at a first terminus of a first subunit monomer and the at least one second binding sites(s) are attached at a second terminus of a second subunit monomer (e.g. a heterooligomeric core). The oligomeric core may comprise a combination of the methods of attachment described above.

[0342] The subunit monomers in the oligomeric core of the multivalent protein scaffold preferably each comprise two termini on a single face of the monomer (and thereby on a single face of the oligomeric core and multivalent protein scaffold). The two termini are preferably the N-terminus and C-terminus of the monomer polypeptide. Each monomer more preferably comprises a first binding site attached at a first terminus of said monomer and a second binding site at a second terminus of said monomer. The first and second binding sites may be the N-terminus and the C-terminus, respectively, or the C-terminus and N-terminus respectively.

[0343] Sometimes, a monomer may comprise more than one binding site at each terminus. For example, a subunit monomer may comprise at each terminus: (i) a first binding site attached at a terminus of said monomer and a second binding site attached to said first binding site, or vice-versa, (ii) a first binding site attached at a terminus of said monomer and at least one further first binding site attached to said first binding site, (iii) a second binding site attached at a terminus of said monomer and at least one further second binding site attached to said second binding site, (iv) a single first or second binding site attached at a terminus of said monomer.

Positioning of Binding Sites on the Multivalent Protein Scaffold

[0344] As described above, in the multivalent protein scaffold provided herein the at least one first binding site(s) and the at least one second binding site(s) are positioned on the same face of the scaffold. Likewise, in typical embodiments of the multi-domain polypeptide construct, the first binding domain and the second binding domain are positioned on the same face of the polypeptide construct.

[0345] By being positioned on the same face of the multivalent protein scaffold (or multi-domain polypeptide construct), the at least one first binding site(s) and the at least one second binding site(s) are arranged such that the effector moieties ultimately bound to the multivalent protein scaffold via the binding sites can interact with their respective biological targets (e.g. receptors on cell surfaces) on a single surface or plane.

[0346] The at least one first binding site(s) and the at least one second binding site(s) are positioned on the same face of the multivalent protein scaffold (or multi-domain polypeptide construct). Preferably, the at least one first binding site(s) and the at least one second binding site(s) are positioned on the same face of the oligomeric core. Preferably, the at least one first binding site(s) and the at least one second binding site(s) are positioned on the same face of the subunit monomer(s) to which they are attached. As described herein, a subunit monomer is typically the structural domain of the multi-domain polypeptide construct.

[0347] The term "the at least one first binding site(s) and the at least one second binding site(s) are positioned on the same face of the scaffold" can be understood as follows, with reference to FIGS. **1** and **2**.

[0348] The multivalent protein scaffold (1) or oligomeric core (10) comprises a conceptual rotational symmetry axis (20) corresponding to the number of monomers in the core. For example, a homotrimeric core comprises a C3 symmetry axis. For example, a homooligomeric pentameric core comprises a C5 symmetry axis. Heterooligomeric cores similarly comprise a conceptual rotational axis that runs through the centre of the oligomeric core and parallel to the interfaces between each subunit, for example, the rotational axis for a heterodimer runs through the oligomeric core and parallel to the length of the interface between the monomers; the rotational axis for a heterotrimer runs through the oligomeric core and as parallel as possible to the length of the at least two interfaces between the monomers. A plane (21) can be defined as being perpendicular or approximately perpendicular (e.g. between about 80° and about 100°, such as between about 85° and about 95° e.g. between about 88° and about 92°, e.g. about) 90° to the rotational axis and running through the centre of the oligomeric core. The at least one first binding site(s) (11) and the at least one second binding site(s) (12) are positioned on the same side of that plane and thus on the same face of the multivalent protein scaffold (1). This is illustrated schematically in FIG. **1**, which depicts a trimeric oligomeric core and in which only one first binding site and one second binding site is depicted for clarity. By contrast, FIG. **2** depicts the contrasting situation where the at least one first binding site(s) (11) and the at least one second binding site(s) (12) are positioned on opposite side of the plane (21) and thus on opposite faces of the multivalent protein scaffold (1).

[0349] Those skilled in the art will appreciate that the conceptual symmetry axis remains even when monomers of the oligomeric core are linked e.g. by being covalently fused as described herein.

[0350] Thus, the "same face of the protein scaffold" may be the solvent accessible surface of the multivalent protein scaffold on one side of the plane perpendicular to the highest-order rotational symmetry axis of the oligomeric core of the multivalent protein scaffold and running though the centre of the multivalent protein scaffold. Similarly, a face of the oligomeric core may be the solvent accessible surface of the oligomeric core (preferably defined in the absence of the binding sites attached thereto) on one side of the plane perpendicular to the highest-order rotational symmetry axis of the oligomeric core and running though the centre of the oligomeric core.

[0351] Preferably, the face of the multivalent protein scaffold is the solvent-accessible portion of the multivalent protein scaffold which makes contact with a single surface, e.g. the surface of a cell such as a cell wall, cell membrane or protein complex. Thus, with reference to the schematic depiction in FIG. **3** (which for clarity depicts a plurality of first and second binding sites (11, 12) attached to the oligomeric core (10) of the multivalent protein scaffold (1)), the at least one first binding site(s) (11) and the at least one second binding site(s) (12) are preferably positioned on the multivalent protein scaffold (1) so that they can both contact the said surface (30). This is not possible if the at least one first binding site(s) (11) and the at least one second binding site(s) (12) are positioned on opposite faces of the multivalent protein scaffold (1), as illustrated schematically in FIG. **4**. For example, the first binding site(s) may be capable of contacting the surface (30) but the second binding sites may not be capable of contacting the surface (30).

[0352] The at least one first binding site(s) and the at least one second binding site(s) are preferably arranged in a front-front orientation or a side-front orientation, or any position in between. As used herein, a front-front orientation refers to both the first binding site(s) and the second binding site(s) both being positioned on the same face of the multivalent protein scaffold with attachment substantially parallel to the rotational symmetry axis through the multivalent protein scaffold. This is illustrated schematically in FIG. **5**. A side-front orientation refers to one of the first binding site(s) and the second binding site(s) being positioned on a face of the multivalent protein scaffold with attachment substantially parallel to the rotational symmetry axis through the multivalent protein scaffold; and the other of the first binding site(s) and the second binding site(s) both being positioned on the same face of the multivalent protein scaffold with attachment substantially perpendicular to the rotational symmetry axis through the multivalent protein scaffold (i.e. substantially parallel to plane (21)). This is illustrated schematically in FIG. **6**. Of course, any position between these extremes can be used, for example where the first binding site(s) and/or the second binding site(s) are positioned on the same face of the multivalent protein scaffold with attachment at approximately 45° to the rotational symmetry axis through the multivalent protein scaffold. This is depicted schematically in FIG. **7**.

[0353] Those skilled in the art will appreciate that the angles between the one or more first binding site(s) (11) and axis (20) and the angles between the one or more second binding site(s) (12) and axis (20) need not be the same. For example, the one or more first binding site(s) (11) may be on the "front" of the multivalent protein scaffold and the one or more second binding site(s) (12) may be on the "side" of the multivalent protein scaffold; i.e. in front-side orientation as described above. Alternatively, the one or more first binding site(s) (11) may be on the "side" of the multivalent protein scaffold and the one or more second binding site(s) (12) may be on the "front" of the multivalent protein scaffold; i.e. in side-front orientation as described above. Both the one or more first binding site(s) (11) and the one or more second binding site(s) (12) may be on the "front" of the multivalent protein scaffold; i.e. in front-front orientation as described above.

[0354] Subject to both the first and second binding site(s) being on the same face of the multivalent protein scaffold, when both first and second binding sites are attached to a given subunit monomer (2), typically the angle formed between the first and second binding site(s) and the centre of that monomer (depicted X in FIG. **8**) is at most at an 160° angle, e.g. at most a 140° angle, e.g. at most a 120° angle, such as at most a 100° angle or at most a 90° angle. Typically, the angle formed between the first and second binding site(s) and the centre of that monomer is at least a 10° angle, e.g. at least a 20° angle, e.g. at least a 30° angle, such as at least a 45° angle or at least a 60° angle.

[0355] Structures may also be visualized by placing a flat target plane in a 3D coordinate system in an arbitrary position such that it does not intersect with the surface of an oligomeric core as determined from protein structural data (NMR, X-Ray) or structure prediction. For each fusion site, the distance of the shortest path to the target plane that does not intersect with the surface of an oligomeric core (other than at the originating fusion site) may be determined. Preferably, for a given structure a position of a target plane can be found such that all such shortest paths are less than 50%, 45%, or 40% of the longest protein's cross-section orthogonal to the target plane.

[0356] In some embodiments, the maximum shortest path length to contact the same plane is less than 100 nm, e.g. less than 50 nm, e.g. less than 20 nm, e.g. less than 10 nm, e.g. less than 5 nm, e.g. less than 2 nm. In a preferred situation, all shortest path lengths to the target plane end within a circular area on the target plane with a radius less than 50 nm, such as less than 25 nm, such as less than 10 nm, such as less than 5 nm.

[0357] In some embodiments, cis-orientation of proteins fused to a scaffold core can be determined by way of structural prediction. In addition to distance alone, Alphafold can take linker geometry and interactions between fused binding domains and scaffold core protein into account. Preferably, the scaffold core is predicted to retain its oligomerization properties even after fusion to preferred binding sites via linkers, preferably via short linkers, such as via GSGS, such as via GGGGS, such as via GGGGSGGGGS, such as via GGGGSGGGGSGGGGS, while the binding sites are predicted to be presented approximately in a cis geometry.q

Domain Insertions

[0358] The multivalent protein scaffold comprises an oligomeric core comprising a plurality of subunit monomers, and at least one first binding site orthogonal to at least one second binding site, preferably wherein the at least one first binding site and the at least one second binding site are

positioned on the same face of the multivalent protein scaffold. The multivalent protein scaffold may further comprise a domain insertion. A domain insertion is a protein domain. A domain insertion may be located on the same face of the multivalent protein scaffold as the binding sites, or on a different face.

[0359] In some cases, wherein the oligomeric core and/or the subunit monomers comprise at least one free terminus that is not attached a binding site, the multivalent protein scaffold comprises a domain insertion at the free terminus. A domain insertion may also be located within loop regions of the oligomeric core, thus many be located within a loop region of a subunit monomer. Preferably, the multimeric protein scaffold comprises at least one domain insertion at an opposite face of the oligomeric core and/or multimeric protein scaffold to the face at which the binding sites are positioned, for example, within 90° of the opposite end of the axis.

[0360] As used herein, a domain insertion is a polypeptide sequence encoding a protein domain, i.e. an autonomously folding functional unit of a protein. The domain insertion does not interfere with the structure and folding of the oligomeric core or the binding sites.

[0361] A domain insertion preferably has an effector function. The domain insertion may comprise an antibody, an antibody fragment or an antigen-binding fragment, such as an antigen-binding fragment capable of binding to CD3 or CD16. For example, a domain insertion may bind be an immune modulatory protein, such as a cytokine, or a chemotherapeutic agent, or a cancer immunotherapy agent (i.e. a treatment that makes use of a subject's immune system to treat cancer). The domain insertion may form a protein that induces cell death when contacted with a biological system. The domain insertion may induce apoptosis, increase an anti-tumor response, or have other beneficial activity. The domain insertion may have complement-inhibiting or complement-stimulating activity.

[0362] For the avoidance of doubt, the domain insertion is typically made into the structural domain of the multi-domain polypeptide construct as described herein.

Protein Complexes

[0363] Also provided herein is a protein complex comprising a multivalent protein scaffold as described in more detail herein, attached to at least one first effector moiety and at least one second effector moiety. Each first effector moiety is attached to a first target bound to a first binding site on the multivalent protein scaffold. Each second effector moiety is attached to a second target bound to a second binding site on the multivalent protein scaffold.

[0364] The targets are preferably polypeptide targets, more preferably a partner of the peptide linker pairs described above.

[0365] Each effector moiety is attached to the target by which it is bound to the multivalent protein scaffold. The first and second effector moieties may be the same or different, preferably different. Any of the attachment routes described above in the context of binding sites may be used. Conventional organic chemistry routes accessible to those skilled in the art may be used to directly attach each effector moiety to the target by which it is bound to the multivalent protein scaffold. Suitable chemical approaches are described in textbooks such as March's Advanced Organic Chemistry (Wiley 2020).

[0366] Preferably, each effector moiety is covalently linked to the target. More preferably, the target is a polypeptide target and may be genetically fused to the effector moiety. In other words, preferably an or each effector moiety is genetically fused to a polypeptide target by being encoded in the same polynucleotide such that they are expressed as a single polypeptide chain. The effector may be genetically fused to a first polypeptide target, a cleavage site, and a second polypeptide target, wherein the first polypeptide target is orthogonal to the second polypeptide target. The cleavage site may be a TEV cleavage site. When both the first and second polypeptide targets are present on the effector moiety, only the terminal polypeptide target is functional (i.e. able to bind its cognate binding site on the multivalent protein scaffold). The cleavage site may be employed to separate the terminal polypeptide target, such that only a single target is present. After complete conjugation of the effector moiety, the terminal polypeptide target can then be deployed specifically.

[0367] An effector moiety is preferably a protein domain. The protein domain is preferably a soluble protein domain. The protein domain preferably comprises a domain of a secreted protein or an extracellular domain of a transmembrane protein. The protein domain more preferably comprises an extracellular domain of a cell-surface receptor or a ligand of a cell-surface receptor, such as a human cell-surface receptor.

[0368] An effector moiety is preferably a moiety which exert a therapeutic effect when contacted with a biological system. For example, an effector moiety may be an immune modulatory protein, such as a cytokine, or a chemotherapeutic agent, or a cancer immunotherapy agent (i.e. a treatment that makes use of a subject's immune system to treat cancer). The effector moiety may induce cell death when contacted with a biological system. The effector moiety may induce apoptosis, increase an anti-tumor response, or have other beneficial activity. The effector moiety may have complement-inhibiting or complement-stimulating activity. The effector moiety may result in altered gene expression, receptor internalization, cytokine release, cell death, or susceptibility to therapeutic molecules.

[0369] In one embodiment an effector moiety may be a synthetic organic or inorganic molecule. A suitable molecule may be a chemotherapeutic agent. A suitable molecule may be a toxic agent, e.g. an agent having an EC50 of less than about 100 μM, e.g. less than about 10 μM, such as less than about 1 μM or less than about 100 nM; wherein EC50 is the concentration of the agent required to result in 50% cell toxicity when assessed in a suitable cell assay. A suitable cell assay may be for example a sulforhodamine B (SRB) assay.

[0370] A suitable synthetic molecule may be an enzyme activator or inhibitor. A suitable molecule may be an inhibitor of one or more of serine/threonine/tyrosine kinases, matrix metalloproteinases (MMPs), heat shock proteins (HSPs), and proteasomes. A suitable molecule may act as an alkylating agent (e.g. nitrogen mustards, nitrosoureas, tetrazines, aziridines, cisplatins and derivatives thereof); an antimetabolite (e.g. anti-folates, fluoropyrimidines, deoxynucleoside analogues and thiopurines); an anti-microtubule agent (e.g. a *vinca* alkyloid or taxanes); a topoisomerase inhibitor (e.g an inhibitor of topoisomerase I, e.g. irinotecan and topotecan; a topoisomerase II poison such as etoposide, doxorubicin, mitoxantrone and teniposidel or a topoisomerase II inhibitor such as novobiocin, merbarone, and

aclarubicin) or a cytotoxic antibiotic (e.g. anthracyclines and bleomycins). A suitable molecule may have a molecular mass of from about 50 to about 5000 g/mol, e.g. from about 100 to about 1000 g/mol such as from about 250 to about 500 g/mol.

[0371] In another embodiment the effector moiety preferably comprises an antibody or an antigen-binding fragment thereof. The term 'antibody or an antigen-binding fragment thereof' as used herein in relation to effector moieties may relate to whole antibodies (i.e. comprising the elements of two heavy chains and two light chains inter-connected by disulphide bonds) as well as antigen-binding fragments thereof. Antibodies typically comprise immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. As used herein, the term "specifically binds" or "immunoreacts with" in the context of the interaction of an antibody or fragment thereof with an antigen, means that the antibody reacts with one or more antigenic determinants of the desired antigen preferentially compared with other polypeptides. Each heavy chain is comprised of a heavy chain variable region (abbreviated herein as HCVR or VH) and at least one heavy chain constant region. Each light chain is comprised of a light chain variable region (abbreviated herein as LCVR or VL) and a light chain constant region. The variable regions of the heavy and light chains contain a binding domain that interacts with an antigen. The VH and VL regions can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDR), interspersed with regions that are more conserved, termed framework regions (FR). Antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, dAb (domain antibody), single chain, Fab, Fab' and F(ab') 2 fragments, scFvs, and Fab expression libraries. An antibody may for example be selected from the group consisting of single chain antibodies, single chain variable fragments (scFvs), variable fragments (Fvs), fragment antigen-binding regions (Fabs), recombinant antibodies, monoclonal antibodies, fusion proteins comprising the antigen-binding domain of a native antibody or an aptamer, single-domain antibodies (sdAbs), also known as VHH antibodies, nanobodies (Camelid-derived single-domain antibodies), shark IgNAR-derived single-domain antibody fragments called VNAR, diabodies, triabodies, Anticalins, aptamers (DNA or RNA) and active components or fragments thereof.

[0372] A "Fab fragment" (also referred to as fragment antigen-binding, or Fab region) contains the constant domain (CL) of the light chain and the first constant domain (CH1) of the heavy chain along with the variable domains VL and VH on the light and heavy chains respectively. The variable domains comprise the complementarity determining loops (CDR, also referred to as hypervariable region) that are involved in antigen-binding. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxy terminus of the heavy chain CHI domain including one or more cysteines from the antibody hinge region.

[0373] A "Single-chain Fv" or "scFv" includes the VH and VL domains of an antibody, wherein these domains are present in a single polypeptide chain. In one embodiment, the Fv polypeptide further comprises a polypeptide linker between the VH and VL domains which enables the scFv to form the desired structure for antigen-binding. For a review of scFv see Pluckthun in The Pharmacology of Monoclonal Antibodies, vol. 113, Rosenburg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994). As an example of scFv fragments, antibody scFv fragments are described in WO93/16185; U.S. Pat. Nos. 5,571,894; and 5,587,458.

[0374] The effector moiety may be a Fab region of a therapeutic antibody. For example, the effector moiety may be the Fab region of a monoclonal antibody such as muromomab, abciximab, rituximab, daclizumab, basiliximab, palivizumab, infliximab, trastuzumab, etanercept, gemtuzumab, alemtuzumab, ibritomomab, adalimumab, alefacept, omalizumab, tositumomab, efalizumab, cetuximab, bevacizumab, natalizumab, ranibizumab, panitumumab, eculizumab, or certolizumab.

[0375] The effector moiety may target any receptor associated with a pathological condition, e.g. a pathological condition described herein. For example, the effector moiety may target any receptor the binding of which is associated with a clinical benefit. For example, hormone receptors.

[0376] In some embodiments the effector moiety may have a target (e.g. a receptor) not previously known to be associated with a pathological condition. For instance, targeting of such receptors has been found to have therapeutic benefit in some cases.

[0377] Those skilled in the art will appreciate that the protein complexes provided herein can therefore be used to simultaneously engage two targets within the biological system thereby contacted simultaneously. The targets may for example be from the same cell. For example, the protein complexes provided herein can be used to bind to two different types of receptor on the same cell surface.

[0378] The protein complexes provided herein typically comprise a plurality of first binding sites and a plurality of second binding sites on the multivalent protein scaffold; and thus can bind a plurality of first effector moieties and a plurality of second effector moieties. This is particularly beneficial as such "high valency" compounds may allow for improved or previously unseen effector functions. It has previously been shown that multiple copies of a single effector moiety may lead to an improved therapeutic response when contacted with a biological system (e.g. Brune et al. (above); and Khairil Anuar et al., Nature communications 10.1 (2019): 1-13). However the contacting of multiple copies of a plurality of different effector moieties is a complex technical challenge that is solved by the protein complexes provided herein.

[0379] In some embodiments an effector function may arise only as a result of the interaction of a combination of effector moieties with the biological system contacted therewith. For example, a first effector moiety (e.g. an effector moiety attached to the first binding site) may be shown to have therapeutic effects only in combination with a second effector moiety (e.g. an effector moiety attached to the second binding site) in circumstances where neither the first nor second effector moiety have therapeutic efficacy alone.

[0380] Those skilled in the art will appreciate that the platforms and methods identified herein allow new combinations of therapeutically useful effector moieties to be screened and useful candidates identified.

Screening Platform

[0381] Also provided herein is a screening platform. The screening platform comprises a library, wherein said library comprises a plurality of populations of protein complexes of the invention, wherein the populations of protein complexes

each comprise a different combination of first effector moieties, second effector moieties and/or oligomeric core. Such a library is also provided herein.

[0382] For example, the library may be used to screen new combinations of effector moieties. Thus, the library may comprise a plurality of samples of different protein complexes. Each sample may be homogeneous, i.e. each sample may contain just one type of protein complex. Each sample may be different to each other sample. Thus, each sample may contain protein complexes comprising a different combination of first and second effector moieties compared to the combination of first and second effector moieties in the protein complexes of each other sample. For example, the library may comprise from about 1 or about 2 to about 1,000,000 samples, e.g. from about 10 to about 100,000 samples, e.g. from about 50 to about 50,000 samples, such as from about 100 to about 10,000 samples, e.g. from about 500 to about 1,000 samples. Each sample may comprise a different type of protein complex, wherein the protein complex in each sample has a different combination of first effector moieties, second effector moieties and oligomeric cores compared to the protein complexes in all other samples.

[0383] In some embodiments the library may be a "ID" library. Thus, in some embodiments all samples in the library may have the same or substantially the same oligomeric core and first effector moiety and may differ in terms of the second effector moiety. In other embodiments all samples in the library may have the same or substantially the same oligomeric core and second effector moiety and may differ in terms of the first effector moiety. In some other embodiments all samples in the library may have the same or substantially the same first and second effector moieties and may differ in terms of the oligomeric core. A polypeptide (e.g. the oligomeric core, or the first or second polypeptide binding site) which is substantially the same as a given polypeptide may for example have at least 90% sequence identity to the given polypeptide, e.g. at least 95% sequence identity such as at least 97%, 98%, 99%, 99.9% or 99.99% sequence identity to the given polypeptide. A polypeptide (e.g. the oligomeric core, or the first or second polypeptide binding site) which is substantially the same as a given polypeptide may for example differ from the given polypeptide by comprising one or more sequence additions, deletions or insertions or variations as described herein. A polypeptide (e.g. the oligomeric core, or the first or second polypeptide binding site) which is substantially the same as a given polypeptide may for example differ from the given polypeptide in terms of post-translational modifications made to the polypeptide, e.g. its glycosylation or phosphorylation pattern.

[0384] In some embodiments the library may be a "2D" library. Thus, in some embodiments all samples in the library may have the same or substantially the same oligomeric core and differ in terms of the combination of the first effector moiety and the second effector moiety. In other embodiments all samples in the library may have the same or substantially the same first effector moiety and differ in terms of the combination of the oligomeric core and the second effector moiety. In some other embodiments all samples in the library may have the same or substantially the same second effector moiety and may differ in terms of the combination of the oligomeric core and the first effector moiety.

[0385] In some embodiments the library may be a "3D" library. Thus in some embodiments all samples in the library may differ in terms of the combination of oligomeric core, the first effector moiety and the second effector moiety.

[0386] The screening platform may also comprise other constituent parts in addition to the library. For example, the screening platform may comprise any or all of:

[0387] a biological system for contacting with the samples in the library;

[0388] a detector system for detecting changes in the biological system resulting from contacting the biological system with samples in the library;

[0389] reagents and/or buffer solutions; and

[0390] optical, electrical or spectroscopic means for detecting changes reported by the detector system

[0391] The biological system may be a cell culture, such as a mammalian cell culture, preferably a human cell culture, more preferably an immune cell culture and/or a cancer-cell line culture. The biological system may be a biological sample, such as blood sample, serum sample, plasma sample, or a sample of tissue or organ. The biological sample include tumor samples, cells, cell lysates, urine, amniotic fluid and other biological fluids. The biological sample is preferably mammalian. The sample may be human or non-human.

[0392] The detector system may be any suitable detector system. The detector system may be a dye or stain, e.g. a cell viability stain. Suitable stains may include for example trypan blue, (fluorescein diacetate)-green, propidium iodide, Hoechst 33258, and the like.

[0393] Reagents include components required for cell viability, including cell growth media components; and may include therapeutic molecules.

[0394] Buffers include aqueous compositions which may comprise e.g. buffer salts. Preferred buffer salts which can be used include Tris; phosphate; citric acid/Na_2HPO_4; citric acid/sodium citrate; sodium acetate/acetic acid; Na_2HPO_4/NaH_2PO_4; imidazole (glyoxaline)/HCl; sodium carbonate/sodium bicarbonate; ammonium carbonate/ammonium bicarbonate; MES; Bis-Tris; ADA; aces; PIPES; MOPSO; Bis-Tris Propane; BES; MOPS; TES; HEPES; DIPSO; MOBS; TAPSO; Trizma; HEPPSO; POPSO; TEA; EPPS; Tricine; Gly-Gly; Bicine; HEPBS; TAPS; AMPD; TABS; AMPSO; CHES; CAPSO; AMP; CAPS and CABS. Selection of appropriate buffers for a desired pH is routine to those skilled in the art, and guidance is available at e.g. http://www.sigmaaldrich.com/life-science/core-bioreagents/biological-buffers/learning-center/buffer-reference-center.html. Buffer salts are preferably used at concentrations of from 1 mM to 1 M, preferably from 10 mM to 100 mM such as about 50 mM in solution.

[0395] Means for detecting changes reported by the detector system include microscopes (optical or electronic), electrical means such as electrophysiology (e.g. patch clamp) apparatus; and spectroscopic means such as equipment for UV/VIS spectroscopy, NMR spectroscopy, mass spectroscopy, IR spectroscopy, Raman spectroscopy, circular dichroism spectroscopy, etc.

Methods

[0396] Also provided is a method for identifying a therapeutic drug analog, the method comprising:

[0397] providing a protein complex as described herein;

[0398] contacting the protein complex with a biological system; and

[0399] measuring whether the protein complex induces a desired change in a property of the biological system.

[0400] The method may optionally further comprise selecting a protein complex that induces a desired change in a property of the biological system.

[0401] Also provided is a method for identifying a therapeutic combination of effector molecules (e.g. antigen-binding domains), the method comprising:

[0402] providing a protein complex as described herein;

[0403] contacting the protein complex with a biological system; and

[0404] measuring whether the protein complex induces a desired change in a property of the biological system.

[0405] The biological system may be a cell culture, such as a mammalian cell culture, preferably a human cell culture, more preferably an immune cell culture and/or a cancer-cell line culture.

[0406] The biological system may be a biological sample, such as blood sample, serum sample, plasma sample, or a sample of tissue or organ. The biological sample include tumor samples, cells, cell lysates, urine, amniotic fluid and other biological fluids. The biological sample is preferably mammalian. The sample may be human or non-human.

[0407] The change in the property of the biological system can be any change associated with the desired activity of the intended therapeutic. In some embodiments the desired change is cell death. This can be particularly used when developing cancer therapeutics.

[0408] Other changes include changes in effector functions. Thus, the method may comprise a step of measuring whether the protein complex induces an effector function in the biological system.

[0409] Changes in effector functions may include altered gene expression, altered protein modifications for example phosphorylation, receptor internalization, cytokine release, cell death, susceptibility to therapeutic molecules, etc. The effector function may be high affinity binding to the biological sample, which can be measured by a range of techniques, such as ELISA. High affinity binding to a target biological system may allow effector domains, discussed above, to specifically effect the targeted biological system, which may be a particular cell type, such as a cancer cell.

[0410] The effector function can be assessed with reference to a control. The control may be the protein complex without effector moieties.

[0411] The control may be a protein complex with only a single type of effector moiety attached (i.e. only one type of effector moiety is attached to the multivalent protein scaffold). In this case, the method can be used to identify effector moieties with a "synergistic function" or "synergistic biological function", which refers to an effector function or level of effector function that: is not observed with individual fusion protein components until a bispecific multivalent protein complex is used; or higher or lower activity in comparison to the activity observed when the first and second effector moieties of the protein complex are employed individually, i.e. activity which is only observed when both effector moieties are used together in the complex.

[0412] The method may further comprise a step of identifying the molecules of the biological system that are bound by the effector moieties of the protein complex. The method may preferably comprise selecting a combination of effector moieties that specifically bind to the same molecules of the biological system as the selected protein complex, such as the effector moieties themselves.

[0413] The method may further comprise synthesising a therapeutic drug candidate or drug comprising the selected combination of effector moieties, or analogs thereof. The therapeutic drug candidate or drug may comprise the oligomeric core and effector moieties of the therapeutic drug analog, but wherein the binding site and target functionalities are replaced with a covalent linkage such as a genetic fusion as described in more detail herein. The therapeutic drug or drug candidate may comprise the same oligomeric core as the therapeutic drug analog identified in the disclosed methods. Alternatively the therapeutic drug candidate may comprise a different oligomeric core as the therapeutic drug analog identified in the disclosed methods. The therapeutic drug candidate may have an oligomeric core chosen or designed in order to impart an additional therapeutic benefit, for example a further effector function.

[0414] Also provided herein is a therapeutic drug candidate obtainable according to the disclosed methods.

Therapeutic Drug Candidate, Therapeutic Drug

[0415] Further provided is a therapeutic drug candidate, comprising an oligomeric core comprising a plurality of subunit monomers attached to one or more first effector moieties and one or more second effector moieties; wherein said one or more first effector moieties and said one or more second effector moieties are positioned on the same face of the oligomeric core; and wherein (i) said one or more first effector moieties comprise two or more first effector moieties and said one or more second effector moieties comprise two or more second effector moieties; and/or (ii) said oligomeric core does not comprise an antibody or antibody fragment.

[0416] A therapeutic drug is also provided with the same features.

[0417] Typically, the oligomeric core is an oligomeric core as described in more detail herein. Typically, the subunit monomers are as described in more detail herein. Typically, the first effector moieties and second effector moieties are as described in more detail here. The first and second effector moieties may be attached to the subunit monomers of the oligomeric core in any appropriate manner, including by any of the attachment means described herein. In some embodiments the attachment of the first and second effector moieties comprises first and second binding sites and first and second polypeptide targets as described herein. However in other embodiments the attachment of the first and second effector moieties does not comprise first and second binding sites and first and second polypeptide targets as described herein and may instead comprise a simple covalent attachment, such as a genetic fusion and/or a click chemistry linkage as described herein.

Specific Embodiments

[0418] In a first preferred aspect, provided herein is:

[0419] A multivalent protein scaffold comprising an oligomeric core comprising a plurality (e.g. from 3 to 6, preferably 3) of monomers each comprising an amino acid sequence having at least 30% or at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to the amino acid sequence of SEQ ID NO: 1; wherein each monomer comprises a first binding site and a second binding site, wherein the first binding site is orthogonal to the second binding site, and wherein the first binding site and the second binding site each independently have at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to any one of SEQ ID NOs: 4-9, 11-13, 23 or 15-18; and wherein each first binding site and each second binding site are independently genetically fused to the monomer to which they are attached. Preferably, one of the first and second binding sites has at least 50% amino acid identity to SEQ ID NO: 4, 6 or 8 and the other has at least 50% amino acid identity to SEQ ID N: 12. A preferred multivalent protein scaffold of this aspect comprises monomers of SEQ ID NO: 21, or a fragment thereof (e.g. comprising residues 14-348 thereof).

[0420] A protein complex comprising the multivalent protein scaffold of the first aspect, wherein the first binding site is bound to a first polypeptide target attached to a first effector moiety; and the second binding site is bound to a first polypeptide target attached to a second effector moiety; wherein the first and second effector moieties may be the same or different, preferably different; and wherein the first binding site/polypeptide target pair and the second binding site/polypeptide target pair are each independently selected from the following combinations: (i) any one of SEQ ID NO: 4, 6 or 8 with any one of SEQ ID NOs: 5, 7 or 9; (ii) SEQ ID NO: 12 with SEQ ID NO: 13 or 15; (iii) SEQ ID NO: 5 with SEQ ID NO: 11; (iv) SEQ ID NO: 15 with SEQ ID NO: 16); (v) SEQ ID NO: 17 with SEQ ID NO: 18; or (vi) SEQ ID NO: 23 with SEQ ID NO: 16).

[0421] A screening platform comprising a library comprising a plurality of populations of protein complexes of the first aspect, wherein each population comprises a different combination of first and second effector moieties.

[0422] A therapeutic drug candidate comprising an oligomeric core comprising a plurality (e.g. from 3 to 6, preferably 3) of monomers each comprising an amino acid sequence having at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to the amino acid sequence of SEQ ID NO: 1; wherein each monomer is directly attached (e.g. as a genetic fusion) to a first effector moiety and a second effector moiety; wherein the first and second effector moieties may be the same or different, preferably different; and preferably wherein each monomer is directly attached via a polypeptide linker to the first effector moiety and the second effector moiety attached thereto.

[0423] In a second preferred aspect, specifically provided herein is:

[0424] A multivalent protein scaffold comprising an oligomeric core comprising a plurality (e.g. from 3 to 6, preferably 3) of monomers each comprising an amino acid sequence having at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to the amino acid sequence of SEQ ID NO: 2; wherein each monomer comprises a first binding site and a second binding site, wherein the first binding site is orthogonal to the second binding site, and wherein the first binding site and the second binding site each independently have at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to any one of SEQ ID NOs: 4-9, 11-13, 23 or 15-18; and wherein each first binding site and each second binding site are independently genetically fused to the monomer to which they are attached. Preferably, one of the first and second binding sites has at least 50% amino acid identity to SEQ ID NO: 4, 6 or 8 and the other has at least 50% amino acid identity to SEQ ID N: 12. A preferred multivalent protein scaffold of this aspect comprises monomers of SEQ ID NO: 20, or a fragment thereof (e.g. comprising residues 14-380 thereof).

[0425] A protein complex comprising the multivalent protein scaffold of the second aspect, wherein the first binding site is bound to a first polypeptide target attached to a first effector moiety; and the second binding site is bound to a first polypeptide target attached to a second effector moiety; wherein the first and second effector moieties may be the same or different, preferably different; and wherein the first binding site/polypeptide target pair and the second binding site/polypeptide target pair are each independently selected from the following combinations: (i) any one of SEQ ID NO: 4, 6 or 8 with any one of SEQ ID Nos: 5, 7 or 9; (ii) SEQ ID NO: 12 with SEQ ID NO: 13 or 15; (iii) SEQ ID NO: 5 with SEQ ID NO: 11; (iv) SEQ ID NO: 15 with SEQ ID NO: 16); (v) SEQ ID NO: 17 with SEQ ID NO: 18; or (vi) SEQ ID NO: 23 with SEQ ID NO: 16).

[0426] A screening platform comprising a library comprising a plurality of populations of protein complexes of the second aspect, wherein each population comprises a different combination of first and second effector moieties.

[0427] A therapeutic drug candidate comprising an oligomeric core comprising a plurality (e.g. from 3 to 6, preferably 3) of monomers each comprising an amino acid sequence having at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to the amino acid sequence of SEQ ID NO: 2; wherein each monomer is directly attached (e.g. as a genetic fusion) to a first effector moiety and a second effector moiety; wherein the first and second effector moieties may be the same or different, preferably different; and preferably wherein each monomer is directly attached via a polypeptide linker to the first effector moiety and the second effector moiety attached thereto.

[0428] In a third preferred aspect, specifically provided herein is:

[0429] A multivalent protein scaffold comprising an oligomeric core comprising a plurality (e.g. from 3 to 6, preferably 3) of monomers each comprising an amino acid sequence having at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to the amino acid sequence of SEQ ID NO: 3; wherein each monomer comprises a first binding site and a second binding site, wherein the first binding site is orthogo-

nal to the second binding site, and wherein the first binding site and the second binding site each independently have at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to any one of SEQ ID Nos: 4-9, 11-13, 23 or 15-18; and wherein each first binding site and each second binding site are independently genetically fused to the monomer to which they are attached. Preferably, one of the first and second binding sites has at least 50% amino acid identity to SEQ ID NO: 4, 6 or 8 and the other has at least 50% amino acid identity to SEQ ID N: 12.

[0430] A protein complex comprising the multivalent protein scaffold of the third aspect, wherein the first binding site is bound to a first polypeptide target attached to a first effector moiety; and the second binding site is bound to a first polypeptide target attached to a second effector moiety; wherein the first and second effector moieties may be the same or different, preferably different; and wherein the first binding site/polypeptide target pair and the second binding site/polypeptide target pair are each independently selected from the following combinations: (i) any one of SEQ ID NO: 4, 6 or 8 with any one of SEQ ID NOs: 5, 7 or 9; (ii) SEQ ID NO: 12 with SEQ ID NO: 13 or 15; (iii) SEQ ID NO: 5 with SEQ ID NO: 11; (iv) SEQ ID NO: 15 with SEQ ID NO: 16); (v) SEQ ID NO: 17 with SEQ ID NO: 18; or (vi) SEQ ID NO: 23 with SEQ ID NO: 16).

[0431] A screening platform comprising a library comprising a plurality of populations of protein complexes of the third aspect, wherein each population comprises a different combination of first and second effector moieties.

[0432] A therapeutic drug candidate comprising an oligomeric core comprising a plurality (e.g. from 3 to 6, preferably 3) of monomers each comprising an amino acid sequence having at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to the amino acid sequence of SEQ ID NO: 3; wherein each monomer is directly attached (e.g. as a genetic fusion) to a first effector moiety and a second effector moiety; wherein the first and second effector moieties may be the same or different, preferably different; and preferably wherein each monomer is directly attached via a polypeptide linker to the first effector moiety and the second effector moiety attached thereto.

[0433] In a fourth preferred aspect, specifically provided herein is:

[0434] A multivalent protein scaffold comprising an oligomeric core comprising a plurality (e.g. from 3 to 6, preferably 3) of monomers each comprising an amino acid sequence having at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to the amino acid sequence of SEQ ID NO: 19; wherein each monomer comprises a first binding site and a second binding site, wherein the first binding site is orthogonal to the second binding site, and wherein the first binding site and the second binding site each independently have at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to any one of SEQ ID NOs: 4-9, 11-13, 23 or 15-18; and wherein each first binding site and each second binding site are independently genetically fused to the monomer to which they are attached.

[0435] Preferably, one of the first and second binding sites has at least 50% amino acid identity to SEQ ID NO: 4, 6 or 8 and the other has at least 50% amino acid identity to SEQ ID N: 12.

[0436] A protein complex comprising the multivalent protein scaffold of the fourth aspect, wherein the first binding site is bound to a first polypeptide target attached to a first effector moiety; and the second binding site is bound to a first polypeptide target attached to a second effector moiety; wherein the first and second effector moieties may be the same or different, preferably different; and wherein the first binding site/polypeptide target pair and the second binding site/polypeptide target pair are each independently selected from the following combinations: (i) any one of SEQ ID NO: 4, 6 or 8 with any one of SEQ ID Nos: 5, 7 or 9; (ii) SEQ ID NO: 12 with SEQ ID NO: 13 or 15; (iii) SEQ ID NO: 5 with SEQ ID NO: 11; (iv) SEQ ID NO: 15 with SEQ ID NO: 16); (v) SEQ ID NO: 17 with SEQ ID NO: 18; or (vi) SEQ ID NO: 23 with SEQ ID NO: 16).

[0437] A screening platform comprising a library comprising a plurality of populations of protein complexes of the fourth aspect, wherein each population comprises a different combination of first and second effector moieties.

[0438] A therapeutic drug candidate comprising an oligomeric core comprising a plurality (e.g. from 3 to 6, preferably 3) of monomers each comprising an amino acid sequence having at least 50% amino acid identity (e.g. at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99% or 100% % amino acid identity) to the amino acid sequence of SEQ ID NO: 4; wherein each monomer is directly attached (e.g. as a genetic fusion) to a first effector moiety and a second effector moiety; wherein the first and second effector moieties may be the same or different, preferably different; and preferably wherein each monomer is directly attached via a polypeptide linker to the first effector moiety and the second effector moiety attached thereto.

[0439] In a fifth preferred aspect, specifically provided herein is:

[0440] A polypeptide comprising a first binding domain at the N terminus and a second binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain and wherein the first antigen binding domain and a second antigen binding domain are able to bind to their targets when the target molecules are expressed on a single cell or immobilised onto a plate or single bead.

[0441] An oligomer of polypeptides, wherein each polypeptide in the oligomer comprises or consists of a polypeptide comprising a first binding domain at the N terminus and a second binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain and wherein the first antigen binding domain and a second antigen binding domain are able to bind to their targets when the target molecules are expressed on a single cell or immobilised onto a plate or single bead.

[0442] A polypeptide comprising a first binding domain at the N terminus and a second binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain, wherein the first binding domain and second binding domain are able to bind to their targets when the targets molecules are expressed on a single cell or immobilised onto a plate or single bead. The first binding domain and the second binding domain are catcher

domains that are each able to form an isopeptide linkage with a cognate peptide. These cognate peptides are often referred to as tag peptides, for example a SpyTag forms an isopeptide bond with a SpyCatcher domain as is known in the art and as discussed above. The cognate peptide for the first binding domain is different from the cognate peptide for the second binding domain.

[0443] An oligomer of polypeptides, wherein each polypeptide in the oligomer comprises or consists of a polypeptide comprising a first binding domain at the N terminus and a second binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain, wherein the first binding domain and second binding domain are able to bind to their targets when the targets molecules are expressed on a single cell or immobilised onto a plate or single bead. The first binding domain and the second binding domain are catcher domains that are each able to form an isopeptide linkage with a cognate peptide. These cognate peptides are often referred to as tag peptides, for example a Spy Tag forms an isopeptide bond with a SpyCatcher domain as is known in the art and as discussed above. The cognate peptide for the first binding domain is different from the cognate peptide for the second binding domain.

Additional Aspects of the Disclosure

[0444] Also provided herein is a polynucleotide encoding at least one monomer of the oligomeric core of the multivalent protein scaffold as described in more detail herein. Also provided herein is a polynucleotide encoding the multi-domain polypeptide construct as described in more detail herein, comprising a first binding domain, a second binding domain and a structural domain

[0445] Further provided is a vector comprising the polynucleotide; a cell comprising the vector; and a method of producing the monomer, oligomeric core and/or multivalent protein scaffold, comprising culturing the cell in a medium to produce the protein scaffold.

[0446] Further provided is a vector comprising the polynucleotide; a cell comprising the vector; and a method of producing the multi-domain polypeptide construct, comprising culturing the cell in a medium to produce the multi-domain polypeptide.

[0447] Selection of appropriate polynucleotide sequences to encode the at least one monomer; of appropriate expression vectors; and of appropriate cells in which to express the monomer, oligomeric core and/or multivalent protein scaffold is routine to those of skill in the art.

Therapeutic Efficacy

[0448] The protein complexes, therapeutic drug analogs and therapeutic drug candidates provided herein are therapeutically useful. The multi-domain polypeptide construct is typically therapeutically useful. These provided substances are also referred to herein as "therapeutic protein complexes".

[0449] The present invention therefore provides therapeutic protein complexes and constructs as described herein, for use in medicine. The present invention provides therapeutic protein complexes as described herein, for use in treating the human or animal body. The present invention provides therapeutic protein constructs as described herein, for use in treating the human or animal body.

[0450] The present invention provides a method of treating a human or animal in need of such treatment, comprising administering to the human or animal in need of treatment a protein complex, multi-domain polypeptide construct (in monomeric or oligomeric form), therapeutic drug analog, therapeutic drug candidate, or drug as described herein.

[0451] Also provided is a pharmaceutical composition comprising one or more therapeutic protein complexes as described herein together with a pharmaceutically acceptable carrier or diluent. Typically, the composition contains up to 85 wt % of a therapeutic protein complexes of the invention. More typically, it contains up to 50 wt % of a therapeutic protein complex of the invention. Preferred pharmaceutical compositions are sterile and pyrogen free.

[0452] Also provided is a pharmaceutical composition comprising one or more multi-domain polypeptide constructs as described herein together with a pharmaceutically acceptable carrier or diluent. Typically, the composition contains up to 85 wt % of a therapeutic multi-domain polypeptide construct of the invention. More typically, it contains up to 50 wt % of a therapeutic multi-domain polypeptide construct of the invention. Preferred pharmaceutical compositions are sterile and pyrogen free.

[0453] The composition of the invention may be provided as a kit comprising instructions to enable the kit to be used in the methods described herein or details regarding which subjects the method may be used for.

[0454] As explained above, the therapeutic protein complexes and constructs provided herein are useful in treating or preventing various disorders. Disorders for treatment using the provided therapeutic protein complexes may include cancer, autoimmune disorders (e.g. ankolysing spondylitis), psoriasis, eye disorders such as age-related macular degeneration, multiple sclerosis, cardiovascular disorders, infections including viral and bacterial infections, Crohn's disease, Rheumatoid arthritis, osteoarthritis, Alzheimer's disease, transplant and allograft rejection, etc. hematopoietic stem cell disorders, and the like. More broadly, therapeutic protein complexes as provided herein find utility in treating any and all conditions also treated using antibodies, particularly bispecific antibodies.

[0455] Cancer, e.g. acute lymphoblastic leukemia, acute myeloid leukemia, adrenocortical carcinoma, aids-related lymphoma, primary CNS lymphoma, anal cancer, astrocytomas, brain cancer, basal cell carcinoma, bile duct cancer, bladder cancer, bone cancer (e.g. ewing sarcoma, osteosarcoma and malignant fibrous histiocytoma), breast cancer, bronchial tumors, medulloblastoma and other CNS embryonal tumors, cervical cancer, chronic lymphocytic leukemia, chronic myelogenous leukemia, chronic myeloproliferative neoplasms, colorectal cancer, craniopharyngioma, endometrial cancer, ependymoma, esophageal cancer, esthesioneuroblastoma, ewing sarcoma, extragonadal germ cell tumor, intraocular melanoma, retinoblastoma, fallopian tube cancer, gallbladder cancer, gastric cancer, gastrointestinal carcinoid tumor, gastrointestinal stromal tumors (gist), germ cell tumors, extragonadal germ cell tumors, ovarian germ cell tumors, testicular cancer, gestational trophoblastic disease, hairy cell leukemia, hepatocellular cancer, histiocytosis, langerhans cell, hodgkin lymphoma, hypopharyngeal cancer, intraocular melanoma, islet cell tumors, pancreatic neuroendocrine tumors, kaposi sarcoma, kidney (renal cell) cancer, langerhans cell histiocytosis, laryngeal cancer, leukemia, liver cancer, lung cancer (non-small cell, small cell,

pleuropulmonary blastoma, and tracheobronchial tumor), lymphoma, malignant fibrous histiocytoma of bone and osteosarcoma, merkel cell carcinoma, mesothelioma, mouth cancer, multiple endocrine neoplasia syndromes, multiple myeloma/plasma cell neoplasms, mycosis fungoides, myelodysplastic syndromes, myelodysplastic/myeloproliferative neoplasms, myelogenous leukemia, neuroblastoma, non-hodgkin lymphoma, oropharyngeal cancer, osteosarcoma and undifferentiated pleomorphic sarcoma, pancreatic cancer, pancreatic neuroendocrine tumors (islet cell tumors), papillomatosis, paraganglioma, parathyroid cancer, penile cancer, pharyngeal cancer, pheochromocytoma, pituitary tumor, prostate cancer, rectal cancer, retinoblastoma, rhabdomyosarcoma, t-cell lymphoma, testicular cancer, throat cancer, thymoma and thymic carcinoma, thyroid cancer, tracheobronchial tumors and the like is particularly suitable to being treated with the therapeutic protein complexes provided herein. Autoimmune disorders amenable to being treated with the therapeutic protein complexes provided herein include rheumatoid arthritis, systemic lupus erythematosus (lupus), inflammatory bowel disease (IBD), multiple sclerosis (MS), Type 1 diabetes mellitus, Guillain-Barre syndrome, chronic inflammatory demyelinating polyneuropathy, psoriasis, Graves' disease, Hashimoto's thyroiditis, Myasthenia gravis and vasculitis.

[0456] The therapeutic protein complexes provided herein may be used as standalone therapeutic agents. Alternatively, they may be used in combination with other active agents such as chemotherapeutic agents. For example, they may be used in combination with an EGFR inhibitor (for instance erlotinib, gefitinib, lapatinib or cetuximab), an immunotherapy (for instance pembrolizumab or nivolumab), a tumour-agnostic therapy (for instance larotrectinib) or a chemotherapy (for instance 5-fluorouracil, cisplatin or docetaxel).

[0457] When used to treat a cancer, the therapeutic protein complexes provided herein may be used in alleviating, ameliorating or preventing aggravation of the symptoms of the cancer. Typically, treating a cancer may comprise reducing progression of the cancer, e.g. increasing progression free survival. Treating a cancer may comprise preventing or inhibiting growth of a tumour associated with the cancer. Treating a cancer may comprise preventing metastasis of the cancer. Preferably, treating a cancer may comprise reducing the size of a tumour associated with the cancer. As such, the treatment may cause tumour regression in the cancer. Treating a cancer may comprise reducing the number of tumours or lesions present in the patient. When the treatment reduces the size of a tumour associated with the cancer, the size of the tumour is typically reduced from base line by at least 10%. Base line is the size of the tumour at the date treatment with the compound is first started. The size of the tumour is typically as measured in accordance with version 1.1 of the RECIST criteria (for instance as described in Eisenhauer et al, European Journal of Cancer 45 (2009) 228-247).

[0458] The response to the treatment with the compound may be complete response, partial Response or stable disease, in accordance with version 1.1 of the RECIST criteria. Preferably, the response is partial response or complete response. The treatment may achieve progression free survival for at least 60 days, at least 120 days or at least 180 days.

[0459] The reduction in tumour size may be greater 20%, greater than 30% or greater than 50% reduction relative to base line. The reduction in tumour size may be observed after 30 days of treatment or after 60 days of treatment.

[0460] The therapeutic protein complexes provided herein may also be useful in treating infection, such as infection caused by Gram-positive and/or Gram-negative bacteria; and viral infections. The therapeutic protein complexes provided herein may be designed to interact with pathogens such as bacteria, fungi and viruses.

[0461] As explained here, the therapeutic protein complexes provided herein are useful in treating or preventing various disorders. The present invention therefore provides a therapeutic protein complex as provided herein for use in medicine. The invention also provides the use of a therapeutic protein complex as provided herein in the manufacture of a medicament. The invention also provides compositions and products comprising the therapeutic protein complexes provided herein. Such compositions and products are also useful in treating or preventing disorders. The present invention therefore provides a composition or product as defined herein for use in medicine. The invention also provides the use of a composition or product of the invention in the manufacture of a medicament.

[0462] Also provided is a method of treating a subject in need of such treatment, said method comprising administering to the subject a therapeutic protein complex provided herein. In some embodiments the subject suffers from or is at risk of suffering from one of the disorders disclosed herein.

[0463] In one aspect, the subject is a mammal, in particular a human. However, it may be non-human. Preferred non-human animals include, but are not limited to, primates, such as marmosets or monkeys, commercially farmed animals, such as horses, cows, sheep or pigs, and pets, such as dogs, cats, mice, rats, guinea pigs, ferrets, gerbils or hamsters. The subject can be any animal that is capable of being infected by a bacterium.

[0464] A subject is typically a human patient. The patient may be male or female. The age of the patient is typically at least 18 years, for instance from 30 to 70 years or from 40 to 60 years. The subject may also be paediatric or adolescent, for example between 6 months and 11 years or between 12 years and 17 years.

[0465] A therapeutic protein complex, polypeptide construct or composition of the invention can be administered to the subject in order to prevent the onset or reoccurrence of one or more symptoms of the disorder. This is prophylaxis. In this embodiment, the subject can be asymptomatic. A prophylactically effective amount of the agent or formulation is administered to such a subject. A prophylactically effective amount is an amount which prevents the onset of one or more symptoms of the disorder.

[0466] A therapeutic protein complex, polypeptide construct or composition of the invention can be administered to the subject in order to treat one or more symptoms of the disorder. In this embodiment, the subject is typically symptomatic. A therapeutically effective amount of the agent or formulation is administered to such a subject. A therapeutically effective amount is an amount effective to ameliorate one or more symptoms of the disorder.

[0467] The therapeutic protein complex, polypeptide construct or composition of the invention may be administered in a variety of dosage forms. Thus, it can be administered orally, for example as tablets, troches, lozenges, aqueous or oily suspensions, dispersible powders or granules. The

therapeutic protein complex or composition of the invention may also be administered parenterally, whether subcutaneously, intravenously, intramuscularly, intrasternally, transdermally or by infusion techniques. The therapeutic protein complex, polypeptide construct or composition may also be administered as a suppository. Preferably, the compound, composition or combination may be administered via inhaled (aerosolised) or intravenous administration, most preferably by inhaled (aerosolised) administration.

[0468] The therapeutic protein complex or composition of the invention is typically formulated for administration with a pharmaceutically acceptable carrier or diluent. For example, solid oral forms may contain, together with the active compound, diluents, e.g. lactose, dextrose, saccharose, cellulose, corn starch or potato starch; lubricants, e.g. silica, talc, stearic acid, magnesium or calcium stearate, and/or polyethylene glycols; binding agents; e.g. starches, arabic gums, gelatin, methylcellulose, carboxymethylcellulose or polyvinyl pyrrolidone; disaggregating agents, e.g. starch, alginic acid, alginates or sodium starch glycolate; effervescing mixtures; dyestuffs; sweeteners; wetting agents, such as lecithin, polysorbates, laurylsulphates; and, in general, non toxic and pharmacologically inactive substances used in pharmaceutical formulations. Such pharmaceutical preparations may be manufactured in known manner, for example, by means of mixing, granulating, tableting, sugar coating, or film coating processes.

[0469] The therapeutic protein complex, polypeptide construct or composition of the invention may be formulated for inhaled (aerosolised) administration as a solution or suspension. The therapeutic protein complex or composition of the invention may be administered by a metered dose inhaler (MDI) or a nebulizer such as an electronic or jet nebulizer. Alternatively, the therapeutic protein complex or composition of the invention may be formulated for inhaled administration as a powdered drug, such formulations may be administered from a dry powder inhaler (DPI). When formulated for inhaled administration, the therapeutic protein complex or composition of the invention may be delivered in the form of particles which have a mass median aerodynamic diameter (MMAD) of from 1 to 100 μm, preferably from 1 to 50 μm, more preferably from 1 to 20 μm such as from 3 to 10 μm, e.g. from 4 to 6 μm. When the therapeutic protein complex or composition of the invention is delivered as a nebulized aerosol, the reference to particle diameters defines the MMAD of the droplets of the aerosol. The MMAD can be measured by any suitable technique such as laser diffraction.

[0470] Liquid dispersions for oral administration may be syrups, emulsions and suspensions. The syrups may contain as carriers, for example, saccharose or saccharose with glycerine and/or mannitol and/or sorbitol.

[0471] Suspensions and emulsions may contain as carrier, for example a natural gum, agar, sodium alginate, pectin, methylcellulose, carboxymethylcellulose, or polyvinyl alcohol. The suspension or solutions for intramuscular injections or inhalation may contain, together with the active compound, a pharmaceutically acceptable carrier, e.g. sterile water, olive oil, ethyl oleate, glycols, e.g. propylene glycol, and if desired, a suitable amount of lidocaine hydrochloride.

[0472] Solutions for inhalation, injection or infusion may contain as carrier, for example, sterile water or preferably they may be in the form of sterile, aqueous, isotonic saline solutions. Pharmaceutical compositions suitable for delivery by needleless injection, for example, transdermally, may also be used.

[0473] A therapeutically or prophylactically effective amount of the therapeutic protein complex or composition of the invention is administered to a subject. The dose may be determined according to various parameters, especially according to the compound used; the age, weight and condition of the subject to be treated; the route of administration; and the required regimen. Again, a physician will be able to determine the required route of administration and dosage for any particular subject. A typical daily dose is from about 0.01 to 100 mg per kg, preferably from about 0.1 mg/kg to 50 mg/kg, e.g. from about 1 to 10 mg/kg of body weight, according to the activity of the specific inhibitor, the age, weight and conditions of the subject to be treated, the type and severity of the disease and the frequency and route of administration. Preferably, daily dosage levels are from 5 mg to 2 g.

[0474] When the therapeutic protein complex or composition of the invention is administered to a subject in combination with another active agent, the dose of the other active agent can be determined as described above. The dose may be determined according to various parameters, especially according to the agent used; the age, weight and condition of the subject to be treated; the route of administration; and the required regimen. Again, a physician will be able to determine the required route of administration and dosage for any particular subject. A typical daily dose is from about 0.01 to 100 mg per kg, preferably from about 0.1 mg/kg to 50 mg/kg, e.g. from about 1 to 10 mg/kg of body weight, according to the activity of the specific agent, the age, weight and conditions of the subject to be treated, the type and severity of the disease and the frequency and route of administration. Preferably, daily dosage levels are from 5 mg to 2 g.

[0475] The protein complexes, therapeutic drug analogs, and therapeutic drug candidates provided herein are also useful in diagnostic methods. The polypeptide constructs, and drugs provided herein are also useful in diagnostic methods. Accordingly, provided herein are protein complexes, therapeutic drug analogs or therapeutic drug candidates, or polypeptide constructs or drugs, as described herein for use in a method of diagnosing a pathology in a subject. The subject may be a subject as described in more detail herein. The pathology may be a pathology as described herein. The method may comprise contacting the protein complex, therapeutic drug analog or therapeutic drug candidate with a sample obtained from the subject (e.g. a biological fluid such as blood, serum, urine, or cerebrospinal fluid; and virology swab samples, biopsy and necropsy tissues) and detecting a change characteristic of the pathology occurs in the sample in the presence of the protein complex, therapeutic drug analog or therapeutic drug candidate compared to in the absence thereof.

[0476] The invention includes at least the following numbered embodiments:

[0477] 1. A multivalent protein scaffold comprising:

- **[0478]** an oligomeric core comprising a plurality of subunit monomers; and
- **[0479]** at least two first binding sites orthogonal to at least two second binding sites;

wherein said first binding sites and said second binding sites are positioned on the same face of the scaffold.

[0480] 2. A multivalent protein scaffold comprising:

[0481] an oligomeric core comprising a plurality of subunit monomers;

[0482] at least one first binding site orthogonal to at least one second binding site;

wherein said first binding site(s) and said second binding site(s) are positioned on the same face of the scaffold; and

[0483] wherein said first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target; and said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target

[0484] 3. A multivalent protein scaffold comprising:

[0485] an oligomeric core comprising a plurality of subunit monomers;

[0486] at least one first binding site orthogonal to at least one second binding site; wherein said first binding site(s) and said second binding site(s) are positioned on the same face of the scaffold; and

[0487] wherein said oligomeric core does not comprise an Fc region of an antibody.

[0488] 4. A protein scaffold according to any one of the preceding embodiments, wherein the oligomeric core comprises at least three subunit monomers,

[0489] wherein preferably the oligomeric core comprises from 3 to 6 subunit monomers.

[0490] 5. A protein scaffold according to any one of the preceding embodiments, wherein said subunit monomers are non-covalently attached together.

[0491] 6. A protein scaffold according to embodiments 1 to 4, wherein said subunit monomers are covalently attached together;

[0492] wherein preferably said subunit monomers are genetically fused together.

[0493] 7. A protein scaffold according to any one of the preceding embodiments, wherein said oligomeric core is a homooligomeric core.

[0494] 8. A protein scaffold according to any one of the preceding embodiments, wherein each monomer in the oligomeric core comprises at least one first binding site and at least one second binding site, and wherein the at least one first binding site is orthogonal to the at least one second binding site.

[0495] 9. A protein scaffold according to any one of the preceding embodiments, wherein each monomer comprises a first binding site attached at a first terminus of said monomer and a second binding site at a second terminus of said monomer.

[0496] 10. A protein scaffold according to any one of the preceding embodiments, wherein the first terminus and the second terminus of each monomer are positioned on the same face of said monomer.

[0497] 11. A protein scaffold according to any one of embodiments 1 to 8, wherein each monomer comprises a first binding site attached at a first terminus of said monomer and a second binding site attached to said first binding site.

[0498] 12. A protein scaffold according to any one of embodiments 1 to 6, wherein said oligomeric core is a hetero-oligomeric core.

[0499] 13. A protein scaffold according to embodiment 12, wherein said core comprises at least one first subunit monomer comprising a first binding site, and at least one second subunit monomer comprising a second binding site, and wherein the first binding site is orthogonal to the second binding site.

[0500] 14. A protein scaffold according to any one of the preceding embodiments, wherein:

[0501] i) each subunit monomer comprises less than 300 amino acids;

[0502] preferably wherein each subunit monomer comprises less than 200 amino acids;

[0503] more preferably wherein each subunit monomer comprises less than 150 amino acids; and/or

[0504] ii) the oligomeric core has a molecular weight of less than about 150 kDa, preferably less than about 100 kDa; more preferably less than about 70 kDa.

[0505] 15. A protein scaffold according to any one of the preceding embodiments, wherein the oligomeric core does not comprise an Fc region of an antibody.

[0506] 16. A protein scaffold according to any one of the preceding embodiments, wherein the oligomeric core comprises a soluble multimerising structural element of a multimeric protein.

[0507] 17. A protein scaffold according to embodiment 16, wherein the multimeric protein comprises a collagen NC1 domain, a CutAl, a Clq domain, a TNF, a p53, a fibrinogen, a C4, *Bacillus subtillus* AbrB or a homolog or paralog thereof.

[0508] 18. A protein scaffold according to embodiment 16 or embodiment 17, wherein the multimerising structural element comprises a polypeptide have at least 50% amino acid identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 19.

[0509] 19. A protein scaffold according to any one of the preceding embodiments, wherein said first binding site and/or said second binding site comprises a protein domain;

[0510] wherein preferably said first binding site comprises a first protein domain and said second binding site comprises a second protein domain; and the first binding site and/or second binding site is genetically fused to the subunit monomer(s) to which they are attached to form a single polypeptide chain.

[0511] 20. A protein scaffold according to any one of the preceding embodiments, wherein said first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target; and said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target;

[0512] wherein preferably said first protein domain is capable of forming an isopeptide bond with said first polypeptide target and said second protein domain is capable of forming an isopeptide bond with said second binding target.

[0513] 21. A protein scaffold according to embodiment 20, wherein said first binding site and said second binding site each comprise a different split ligand-binding protein domain;

[0514] wherein preferably one of said first binding site and said second binding site comprises a split *Streptococcus pyogenes* fibronectin-binding protein domain and the other of said first binding site and said second binding site comprises a split *Streptococcus pneumoniae* adhesin domain.

[0515] 22. A protein scaffold according to embodiment 21, wherein said first and said second binding site each inde-

pendently have at least 50% amino acid identity to any one of SEQ ID Nos: 4-9, 11-13 or 15-18.

[0516] 23. A protein complex comprising a protein scaffold according to any one the preceding embodiments, wherein the first binding site is bound to a first polypeptide target attached to a first effector moiety, and the second binding site is bound to a second polypeptide target attached to a second effector moiety.

[0517] 24. The protein complex according to embodiment 23, wherein the first binding site/polypeptide target pair and the second binding site/polypeptide target pair are each independently selected from the following combinations: (i) any one of SEQ ID NO: 4, 6 or 8 with any one of SEQ ID NOs: 5, 7 or 9; (ii) SEQ ID NO: 12 with SEQ ID NO: 13 or 15; (iii) SEQ ID NO: 5 with SEQ ID NO: 11; (iv) SEQ ID NO: 15 with SEQ ID NO: 16); (v) SEQ ID NO: 17 with SEQ ID NO: 18.

[0518] 25. A screening platform comprising a library, wherein said library comprises a plurality of populations of protein complexes according to embodiment 23 or embodiment 24, wherein the populations of protein complexes each comprise a different combination of first effector moieties, second effector moieties and/or oligomeric core.

[0519] 26 A method for identifying a therapeutic drug analog, the method comprising:

- **[0520]** providing a protein complex according to embodiment 23 or embodiment 24;
- **[0521]** contacting the protein complex with a biological system; and
- **[0522]** measuring whether the protein complex induces a desired change in a property function of the biological system;
- **[0523]** and optionally further comprising selecting a protein complex that induces a desired change in a property of the biological system.

[0524] 27. A method according to embodiment 26, further comprising:

- **[0525]** synthesizing a therapeutic drug candidate comprising the oligomeric core of the scaffold of the protein complex of the identified therapeutic drug analog attached to the first and second effector moieties of said protein complex.

[0526] 28. A therapeutic drug candidate obtainable according to the method of embodiment 26 or embodiment 27.

[0527] 29 A therapeutic drug candidate, comprising an oligomeric core comprising a plurality of subunit monomers attached to one or more first effector moieties and one or more second effector moieties; wherein said one or more first effector moieties and said one or more second effector moieties are positioned on the same face of the oligomeric core; and

- **[0528]** wherein (i) said one or more first effector moieties comprise two or more first effector moieties and said one or more second effector moieties comprise two or more second effector moieties; and/or (ii) said oligomeric core does not comprise an antibody or antibody fragment.

[0529] 30. The therapeutic drug candidate of embodiment 29, wherein the oligomeric core is as defined in any one of embodiments 1 to 22.

[0530] 31. The therapeutic drug candidate of embodiment 29 or embodiment 30, wherein the oligomeric core comprises a plurality of subunit monomers and wherein: (i) each subunit monomer comprises a collagen NC1 domain, a CutAl, a Clq domain, a TNF, a p53, a fibrinogen, a C4, *Bacillus subtillus* AbrB or a homolog or paralog thereof; and/or (ii) each subunit monomer comprises a multimerising structural element comprising a polypeptide having at least 50% amino acid identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or SEQ ID NO: 19.

[0531] The following examples illustrate the invention. They do not however limit the invention in any way. In particular, there are many assays for protein binding and so a negative result in any specific assay is not determinative. The invention is defined according to the claims.

[0532] In summary, Example 1 details the production of constructs comprising a Collagen X NC1 structural domain and SnoopCatcher and SpyCatcher domains at the N and C termini, which are then covalently linked via isopeptide bonds to SpyTagged and SnoopTagged therapeutic polypeptides, and the isopeptide-bound constructs are then oligomerised to form a homotrimer. Example 2 provides the materials and methods used in the subsequent examples. Example 3 overviews the design of constructs according to the invention, identifies multiple components in suitable geometry, including C3 geometry, and demonstrates purification of SpC-PhCutAl-SnC (SEQ ID NO: 22). Example 4 illustrates how other assembly geometries can be modified to meet design criteria for preferred constructs of the invention. Example 5 highlights the exceptional stability of PhCutAl-derived components, as well as firm evidence for multimerization as predicted by its structure. Example 6 purifies SpyTagged/SnoopTagged components and decorates the resulting platform with tagged proteins. Example 7 demonstrates that after modular assembly, proteins can be cleaned up in a scalable fashion, herein utilizing the large size in solution via Dialysis against a 100-kDa membrane. Example 8 demonstrates that modular assembly enables rapid prototyping, including development of a HsCutAl-derived platform as a transition from PhCutAl-derived platform. Example 9 uses Alphafold to predict cis-orientation and contrast to non-cis IMX and Collagen XVIII NCI assemblies. Example 10 provides cell data showing that the assembled platform can be used for in vitro screening to elucidate the efficacy and downstream effects of ligands. Example 10 also shows that multi-domain polypeptide of PhCutAl incorporating effector moieties can be readily produced.

EXAMPLES

Example 1

[0533] A polynucleotide sequence encoding a monomer of collagen NC1 attached by a linker at the N-terminus to one of SpyCatcher and SnoopCatcher and at the C-terminus to the other of SpyCatcher and SnoopCatcher is synthesized according to methods described in Sambrook et al., Molecular Cloning: A Laboratory Manual, 4th ed., Cold Spring Harbor Press, Plainsview, New York (2012). The polynucleotide sequence is expressed in a cellular expression system to produce the polypeptide fusion. The polypeptide fusion is allowed to oligomerise such that the SpyCatcher and SnoopCatcher binding sites are on the same face of the oligomerised trimeric construct.

[0534] Samples of the construct are contacted with a panel of different first and second therapeutic polypeptides each bound to SpyTag and SnoopTag, respectively, causing the SpyTagged polypeptide to form a covalent isopeptide bond

with each SpyCatcher moiety and the SnoopTagged polypeptide to form a covalent isopeptide bond with each SnoopCatcher moiety, thereby producing a library of collagen X NCI constructs, wherein each monomer in the construct is bound to each of two different therapeutic polypeptides, and the trimeric construct thus comprises three copies of each polypeptide. Each sample comprises a different combination of first and second therapeutic polypeptides.

[0535] Each sample in the library is assessed for its ability to trigger a biological reaction in a biological system, such as to cause cell death in a sample of cancer cells.

[0536] The samples from the library which are most effective in causing cancer cell death are noted. The combination of therapeutic polypeptides in such samples is noted.

[0537] A polynucleotide encoding a monomer of an oligomeric protein such as a monomer of collagen NC1, linked at the N-terminus to one of the therapeutic polypeptides and linked at the C-terminus to the other of the therapeutic peptides is synthesized as described in Sambrook et al., Molecular Cloning: A Laboratory Manual, 4th ed., Cold Spring Harbor Press, Plainsview, New York (2012). The polynucleotide is expressed to produce a polypeptide monomer consisting of a fusion of the oligomeric protein monomer and the first and second therapeutic polypeptides. The monomer is allowed to oligomerise. The monomer is tested against the biological system. The biological reaction observed for the initial construct, e.g. the causation of cell death in a sample of cancer cells, is observed.

[0538] The Example can be performed with a polynucleotide sequence encoding a monomer of CutAl attached by a linker at the N-terminus to one of SpyCatcher and SnoopCatcher and at the C-terminus to the other of SpyCatcher and SnoopCatcher.

Example 2—Methods

[0539] Selection of scaffold protein components: Protein structures that meet the design criteria were selected from the Protein Data Bank (PDB). The use of adequate search filters (such as provided via http://www.rcsb.org/pdb/) enabled geometry-based prescreening, after which candidate structures were further inspected via protein structure visualization and by reference to biochemical properties described in relevant literature.

[0540] Prediction of assembled protein structures: The multimer 3D-structure of protein sequences were predicted via an AlphaFold v2.0 implementation in Colab notebooks provided by Mirdita et al., 2021, bioRxiv, DOI: 10.1101/2021.08.15.456425v3 using the AlphaFold2-multimer-v2 model_type parameter, with default parameters for all proteins aside from IMX-containing SpC3-IMX-DgC SEQ ID NO: 35. For IMX-containing SpC3-IMX-DgC SEQ ID NO: 35, template_mode was set as pdb70 to conserve computational resources.All proteins other than SpC3-IMX-DgC were submitted as trimers, with SpC3-IMX-DgC submitted as a heptamer. Terminal linker and tag sequences were removed prior to prediction. The highest-ranked model was visualized.

[0541] Molecular cloning: Plasmids encoding recombinant proteins were provided by Twist Biosciences or ProteoGenix. DNA fragments and oligonucleotides were synthesised by Integrated DNA Technologies (IDT). Constructs for L1-PhCutAl-L2, DgT-X3 and SpC-HsCutAl-DgC were assembled through standard cloning procedures. To introduce synthesised DNA fragments into plasmid backbones, to introduce point mutations, or to make other adjustments to recombinant sequences, DNA was amplified using standard polymerase chain reaction (PCR) followed by standard cloning methods, including restriction cloning. Assembled constructs were transformed into *E. coli* NEB 5-alpha cells. Putative positive clones were grown overnight, and DNA was isolated from bacterial pellets via miniprep. Samples were sent for Sanger sequencing to Source Bioscience for sequence validation and alignment was performed using Benchling's molecular biology suite (www.benchling.com).

[0542] Protein purification: To obtain protein of SnT-L1 (16.1 kDa), L2-SpT (9.7 kDa), DgT-X3 (26.9 kDa), SpC-PhCutAl-SnC (39.0 kDa), SpC3-MIF2m-DgC (39.6 kDa) and SpC3-HsCutAl-DgC (40.5 kDa) DNA (synthesised by ProteoGenix, Twist Bioscience, or through standard cloning procedures) encoding for both proteins was transformed into BL21 (DE3) cells. Colonies were used to inoculate LB cultures with 50 g/mL Kanamycin at 37° C. with 160-220 rpm shaking. Overnight cultures were diluted 1:100 into LB or 2×YT media supplemented with 50 μg/mL Kanamycin. Cultures were grown at 37° C. with 160-220 rpm shaking before induction of protein expression with 0.2-0.4 μM IPTG at OD 0.6-0.8 (LB) or 1.6-2.0 (2×YT). For platform proteins (SpC-PhCutAl-SnC, SpC3-MIF2m-DgC, SpC3-HsCutAl-DgC), samples were incubated for further 4 h at 37° C. with 160-200 rpm shaking.

[0543] For tagged proteins (SnT-L1, L2-SpT, DgT-X3) samples were incubated for 16 h at 18° C. with 160-200 rpm shaking. Cells were collected by centrifugation at 5000×g for 15 min at 4° C. and pellets stored at −20° C. For protein purification, cell pellets were resuspended in Ni-NTA equilibration buffer (50 mM Tris, pH 7.8; 300 mM NaCl, 10 mM imidazole) supplemented with 1 mM PMSF, complete EDTA-free protease inhibitor cocktail and benzonase (5 U/mL). Samples were sonicated with an Ultrasonic Processor using a 20 mm probe and an amplitude of 20%, for 9-12 minutes pulsing on 2 seconds/off 4 seconds and spun down at 16,000x g for 30 min. Supernatants were retained for Ni-NTA chromatography.

[0544] Proteins were purified from cell lysates using pre-equilibrated HisPur Ni-NTA gravity-flow columns. Protein lysates were loaded to the resin. The resin was washed with 50 mM Tris, pH 7.8; 300 mM NaCl, 10 mM imidazole and subsequently with 50 mM Tris, pH 7.8; 300 mM NaCl, 30 mM imidazole until absorbance of the flow-through factions at 280 nm approached baseline. His-tagged proteins were eluted from the resin with two resin-bed volumes of Elution Buffer (50 mM Tris, pH 7.8; 300 mM NaCl, 200 mM imidazole) until the absorbance of the elution fractions at 280 nm approached baseline. Eluates were analysed by SDS-PAGE followed by Coomassie staining. Following Ni-NTA purification, samples in a high concentration of imidazole were dialysed into PBS using SnakeSkin™ Dialysis Tubing at 3K MWCO. An appropriate length of tubing was determined based on the total elution volume and was hydrated with Milli-Q water. Sample was transferred to the SnakeSkin™ Dialysis Tubing and placed in 5 L PBS overnight at 4° C. on a magnetic stirrer. After 16 h the buffer was replaced a further two times with 2 h incubations. Estimated protein concentrations were calculated from measurement of absorbance at 280 nm using an Implen NanoPhotometer N60 with reduced extinction coefficients predicted by ProtParam. L3-DgT was produced by Absolute Antibody.

[0545] Size exclusion chromatography: Following Ni-NTA purification, proteins were optionally purified further using size exclusion chromatography, as shown in the Figures. This was performed using an ÄKTA Pure 25 (Cytiva) with a HiLoad Superdex 75 pg (SnT-L1 and L2-SpT) or a HiLoad Superdex 200 pg (SpC-PhCutAl-SnC and L1-PhCutAl-L2) column. The column was first equilibrated with one column volume of PBS. Prior to loading, the protein sample was concentrated to ~1 mL and injected onto the column using a 2 mL injection loop. The sample was then separated using a flow rate of 1 mL min-1 with 2 mL fraction size collection. 20 μl samples were taken from each fraction, corresponding to major elution peaks for SDS-PAGE analysis. Fractions corresponding to the peak for the protein of interest were pooled and used in downstream applications or stored at −20° C.

[0546] HsCutAl assemblies were purified using a Superose 6 Increase 5/150 column. The column was first equilibrated with one column volume of PBS. Prior to loading, the protein sample was prepared to ~100 μL and injected onto the column via a Hamilton 700 Microliter Syringe. The sample was then separated using a flow rate of 0.3 mL min-1 with 0.1 mL fraction size collection. 10 μL samples were taken from each fraction, corresponding to major elution peaks for SDS-PAGE analysis. Fractions corresponding to the peak for the protein of interest were pooled and used in downstream applications or stored at −20° C.

[0547] Protein quantification by BCA assay: Prior to conjugation, samples were quantified using a BCA Protein Assay kit (ThermoFisher) according to the manufacturer's instructions. BSA standards were diluted in 1×PBS. Purified protein was diluted 1/5, 1/10 and 1/20 in 1×PBS to ensure concentration was within the linear range of the assay. Following incubation with the BCA reagent, absorbance at 562 nm was measured on a BMG FluoSTAR Omega plate reader. Concentration in mg/mL was interpolated based on the standard curve.

[0548] Catcher-based protein conjugation: Platform conjugation with relevant ligands were performed between 1:1:1-1:2:2 of platform: ligand: ligand ratios, according to the specific assays as detailed in the relevant figures. The conjugation reactions proceeded at 25° C. for 16 h, 24 h or 64 h, according to relevant assays, and samples were analysed by SDS-PAGE (8%, 16%) followed by Coomassie staining.

[0549] Post-assembly dialysis: Confirmatory dialysis of conjugated assembly H6-SpC-PhCutAl-SnC: SnT-L1: L2-SpT to remove excess substrates was performed using an HTDialysis 12-well block with a 100 kDa MWCO cellulose membrane at room temperature. Prior to dialysis, the membrane was hydrated for 60 min in sterile Milli-Q, replaced by 20% ethanol for 20 min and washed in sterile Milli-Q water twice before use. Both sample and dialysate contained 1×PMSF. Samples of the assembled platform and dialysate were taken at 2 h, 4 h, 8 h and 16 h during dialysis and analysed by SDS-PAGE (14%) followed by Coomassie staining.

[0550] Preparatory dialysis of conjugated assemblies as input for cell assays were done as above, but with the following modifications: no PMSF was added, and the dialysis was performed at room temperature overnight.

[0551] HsCutA1 crosslinking: To crosslink H6-SpC3-HsCutAl-DgC, 10 μM of the protein was incubated with 0.1% glutaraldehyde in PBS for 20 min at 37° C. Samples were taken at the indicated intervals during the reaction and 100 mM Tris pH 8.8 was added to stop the reaction. Samples were analysed by SDS-PAGE (12%) followed by Coomassie staining.

[0552] Cell culture: NCI-N87 (CRL-5822) and A-431 (CRL-1555) cells were obtained from ATCC and routinely cultured in RPMI and DMEM, respectively, supplemented with 10% FCS and 5% Penicillin/Streptomycin.

[0553] Cell viability assay: 2000 NCI-N87 cells/well were seeded into 96-well plates and grown in DMEM supplemented with 10% FCS for 24 h before starvation in DMEM medium containing 0.2% FCS for 24 h. Cells were then treated or mock-treated with various concentrations (0.01-100 nM) of protein assemblies with two ligands (H6-SpC-PhCutAl-SnC: SnT-L1: L2-SpT) or one ligand only (H6-SpC-PhCutAl-SnC: SnT-L1, H6-SpC-PhCutAl-SnC: L2-SpT). Scaffold only (H6-SpC-PhCutAl-SnC), ligands only (SnT-L1, L2-SpT, SnT-L1+L2-SpT) and monoclonal control antibodies against both ligands were used as controls. All antibodies were diluted in starvation medium. 1 h after the start of the treatment, assay-relevant growth factors were added to a final concentration of 30 ng/ml with a final FCS concentration of 2%. Cells were then grown for 7 days, and surviving fractions were measured using the MTT assay. 20 μL of 5 mg/mL 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) in PBS were added to each well containing 200 μL medium. Incubation for 3 h was followed by media aspiration, formazan dissolution in 100% DMSO and absorbance reading at 570 nm. Surviving fractions were normalised to mock-treated controls.

[0554] Western blotting: Stability assay: H6-SpC-PhCutAl-SnC and H6-SpC-PhCutAl-SnC: SnT-L1: L2-SpT (purified using size exclusion chromatography) were incubated in PBS and complete media containing 10% FCS for 4 days at 4° C. or 37° C. Samples were run on 4-20% Tris-Glycine SDS-Page gels and transferred to nitrocellulose membranes using the iBlot 2 system (Thermo Fisher). Membranes were probed with an antibody against anti-polyHistidine (Sigma-Aldrich, cat no. A7058, 1:2000) and signals were detected using ECL. Akt/ERK signalling: 1.5× 10^{-6} NCI-N87 cells were seeded into T25 flasks and grown in DMEM supplemented with 10% FCS and 5% Penicillin/Streptomycin medium for 24 h before starvation in medium containing 0.2% FCS for 24 h. Cells were then treated or mock-treated with 25 nM of protein assemblies containing both (H6-SpC-PhCutAl-SnC: SnT-L1: L2-SpT) or one ligand (H6-SpC-PhCutAl-SnC: SnT-L1, H6-SpC-PhCutAl-SnC: L2-SpT). Scaffold only (H6-SpC-PhCutAl-SnC), ligands only (SnT-L1, L2-SpT, SnT-L1+L2-SpT) and commercially available monoclonal antibodies against the two target proteins were used as controls. 1 h after the start of the treatment, cells were stimulated by adding assay-relevant growth factors to a final concentration of 50 ng/ml and incubated for 1 h with FCS at 0.2% throughout the treatment. Whole cells extracts were prepared using RIPA buffer supplemented with phosphatase and protease inhibitors. 25 μg of protein per lane were separated on 4-12% Bis-Tris gels and transferred to nitrocellulose membranes using the iBlot2 system according to manufacturer's instructions. Membranes were probed with antibodies against p-Akt (Cell Signalling, cat no. 4060, 1:2000), Akt (Cell Signalling, cat no. 2920, 1:2000), p-ERK 1/2 (Cell Signalling, cat no. 9101, 1:1000), ERK1/2 (Cell Signalling, cat no. 4695S, 1:1000) . . . . Signals were detected using ECL.

[0555] Cell killing assay: To investigate protein assemblies targeting L1 and L3, NCI-N87 cells (30,000 cells/well) were seeded into 96-well plates and grown in complete media (DMEM, supplemented with 10% FCS and 5% Penicillin/Streptomycin) for 24 h, followed by starvation in medium containing 0.2% FCS and 5% Penicillin/Streptomycin for 24 h. Cells were then treated with various concentrations (0.01-100 nM) of assembled proteins H6-SpC3-HsCutAl-DgC: L1-SpT: L3-DgT, H6-SpC3-HsCutAl-DgC: L1-SpT and H6-SpC3-HsCutAl-DgC: L3-DgT scaffold only (H6-SpC3-HsCutAl-DgC) and ligands only (L1-SpT, L3-DgT) and incubated for 40 h. Surviving fractions were measured by MTT assay and normalised to mock-treated control cells.

Example 3—Suitable Selection of Recombinant Protein Components Enables Simple Preparation of Protein Complexes Featuring Binding in Cis-Oriented Geometry

[0556] The inventors recognised that a multimeric protein component in which each monomer features a C-terminus and N-terminus that are in proximity to each other or to the termini of other monomers in the same complex can be utilised to project binding sites towards a single binding surface in "cis-orientation". Publicly available protein structures were filtered for keywords and/or geometric parameters to identify protein structures with suitable geometry to arrive at multimeric protein complexes in "cis-orientation" via recombinant fusion of binding sites at monomer termini, or ligands to such a protein complex (FIG. **11**).

[0557] The inventors identified a number of suitable domains, including: Collagen X NC1 domain (PDB ID: 1GR3), Collagen VIII NC1 domain (PDB ID: 1091), CutAl (copper tolerance A) proteins from various species (such as the CutAl proteins from *Pyrococcus horikoshii* (PDB ID: 4YNO), *Homo sapiens* (PDB ID: 2ZFH), *Thermus* thermophiles (PDB ID: 1V6H); *Oryza sativa* (PDB ID: 2ZOM); or *Shewanella* sp. SIB1

[0558] (PDB ID: 3AHP), Clq head domain (PDB ID: 1PK6), TNF-like protein TLIA (PDB ID: 2RE9), TNF (PDB ID: 1TNF), MIF (PDB ID: 1CA7), MIF2 (PDB ID: 7MSE), and other protein structures described herein or depicted in FIG. **11**.

[0559] The inventors were able to readily express and prepare PhCutAl from *Pyrococcus horikoshii* (SEQ ID NO: 1) fused to SpyCatcher (SEQ ID NO: 4)N-terminally (via a GSGS linker) and SnoopCatcher (SEQ ID NO: 12)C-terminally (via a GSGS linker) recombinantly in *E. coli*, further featuring a His-tag N-terminal of SpyCatcher (SEQ ID NO: 21, FIG. **12**). Notably, this construct "H6-SpyCatcher-PhCutAl-SnoopCatcher" or "SpC-PhCutAl-SnC" featured hyper-thermostable trimerisation as apparent in resistance to boiling in denaturing SDS-loading buffer, a property characteristic to intact PhCutAl (Tanaka et al., 2006, FEBS Letters, 580 (17), pp. 4224-4230). Therefore, multimerisation properties of the core protein PhCutAl were successfully imparted onto the SpC-PhCutAl-SnC scaffold.

Example 4—A Protein Complex Suitable for "Cis-Oriented" Display Via Recombinant Fusion N-Terminal and C-Terminal of Monomer Proteins can be Derived from Heteromeric Protein Complexes or Dihedral Protein Assemblies

[0560] In addition to protein structures or domains already featuring geometry suitable for "cis-oriented" display by recombinant fusion at N-terminal and C-terminal sites, the inventors also identified proteins from which such components could be derived.

[0561] With a suitable linker between C-terminus and N-terminus of antiparallel coiled-coils in PDB ID 5w0j (Spencer & Hochbaum, 2017, Biochemistry, 56 (40), pp. 5300-5308), the structure mimics the structure of circularly symmetric HIV GP41 (PDB ID 1i5y) (FIG. **13** a,c). Herein, a heteromeric antiparallel coiled-coil assembly (PDB ID 5vte) derived in the same publication as PDB ID 5w0j (Spencer & Hochbaum, 2017) provides an example of how simple (rational) mutagenesis could mitigate inversion of the termini by encouraging uniform assembly. Coiled-coil proteins are easy to design and can benefit from designed properties such as pH-sensitivity (Nagarkar et al., 2020, Peptide Science, 112 (5), e24180) or as bioactive protein switches (Langan et al., 2019, Nature, 572 (7768), pp. 205-210).

Example 5—CutAl Proteins Retain Trimeric Structure after Recombinant Fusion to Catcher Proteins

[0562] PhCutAl is a highly thermostable protein that retains trimeric structure even after boiling in denaturing SDS-loading buffer. As shown in FIG. **14**, a protein band was observed between 130-250 kDa indicating stable trimerisation of PhCutAl even after recombinant fusion to SpC and SnC, confirming that the overall complex assembles according to our structure-based design. To further validate that this observed band was dependent on correct PhCutAl assembly, we used harsh denaturing conditions, upon which we observed partial monomerisation of SpC-PhCutAl-SnC.

Example 6: The Use of Modular Components Enables Rapid Assembly of Ligand Proteins and Other Effectors onto Scaffold Proteins of Choice to Impart Complex Geometry

[0563] To demonstrate the feasibility of modular component assembly, Spy Tagged, SnoopTagged and DogTagged ligands were designed and expressed. Spy/Snoop-tagged protein components SnT-L1 and L2-SpT are ligands against common cellular antigens. Ni-NTA purification resulted in clean-up of both ligand proteins (FIG. **15** a-b), which was optionally followed by size exclusion chromatography (FIG. **15** c-d). These components were used to confirm that conjugation of tagged ligands to modular platforms containing Catcher proteins yields fully assembled platforms with ligands attached.

[0564] After incubation of SnT-L1 and/or L2-SpT with SpC-PhCutAl-SnC or control protein SpC—PC-SnC, we observed that conjugation was able to go to completion for all samples (FIG. **16***a-c*). Notably, hyper-thermostable trimerisation of SpC-PhCutAl-SnC was retained even during decoration with both SnT-LI and L2-SpT. Conjugation of H6-SpC-PhCutAl-SnC (117 kDa as trimer) and H6-SpC—PC-SnC (31 kDa) with SnT-LI and with L2-SpT resulted in a band shift corresponding to the added molecular weight (16.1 kDa per monomer SnT-L1, 9.7 kDa per monomer L2-SpT). Simultaneous conjugation of both ligands to either platform resulted in a significant band shift (expected 25.8 kDa per monomer) (FIG. **15** c-d). Furthermore, the consumption of ligands was observed in a reduction of ligand intensity (added in excess of platform).

Example 7: Integration of Modular Assembly with Simple Post-Assembly Clean-Up Enables the Manufacture of Uniform Drug Candidates for Downstream Analysis

[0565] To validate a simple and effective method for the purification of assembled drug candidates in an automation-compatible manner, the inventors investigated the suitability of a reusable, 96- and 12-well high-throughput dialysis device for drug candidate purification. The inventors have demonstrated that the modular assembly of SpC-PhCutAl-SnC with SnT-L1 and L2-SpT can be purified via high-throughput dialysis for 16 h with regular buffer changes. For this, assembly was performed with a protein component ratio of 1:1:1 and samples were incubated at 25° C. for 16 h. Dialysis was performed using a 12-well high-throughput dialysis block with a 100 kDa MWCO cellulose membrane at room temperature. Both sample and dialysate contained 1×PMSF to avoid protein degradation during dialysis. Samples of the assembled platform and dialysate were taken at 2 h, 4 h, 8 h and 16 h during dialysis and analysed by SDS-PAGE to demonstrate the removal of excess ligands and low molecular impurities over time (FIG. **17**).

[0566] This demonstrates that a large molecular weight assembly is present and stable in solution. Its properties enable a simple purification protocol. Furthermore, this methodology is scalable and automation-compatible for final protein assembly workflow, which can then be used for downstream in vitro analysis.

Example 8: Modular Component Design Enables Rapid Prototyping and a Gradual Transition of Assembly Platforms Optimised for Early Screening to Assembly Platforms Optimised for More Downstream Therapeutic Validation

[0567] With PhCutAl, the inventors have demonstrated the design, production, assembly and purification of a bacterial scaffold protein for use in in vitro screening (FIG. **11**, FIG. **12**). To further demonstrate the ability to rapidly change platform components, including for potential drug utility, we have also identified, cloned, expressed and purified scaffolds based on the human homologue of the CutAl protein, HsCutAl (FIG. **11**) and a mutant of the human Macrophage Migration Inhibitory Factor 2 (MIF2), MIF2m (FIG. **11**). The SpC3-HsCutAl-DgC (SEQ ID NO: 24) platform features SpyCatcher003 (SEQ ID NO: 8) and DogCatcher (SEQ ID NO: 23) for seamless modular conjugation to tagged-ligands fused to HsCutAl (SEQ ID NO: 29, truncated to retain some natural amino acids beyond the oligomeric core as linkers), representing a human variant for in vitro validation and downstream therapeutic validation. SpC3-MIF2m-DgC (SEQ ID NO: 28) features a different scaffold with similar C3 geometry and with longer linkers (GGGGSGGGGSGGGGS) compared to SpC-PhCutAl-SnC (GSGS) and Sp3-HsCutAl-DgC (GGGGS). To prepare SpC3-HsCutAl-DgC and SpC3-MIF2-DgC, BL21 (DE3) cells were used for protein expression, followed by Ni-NTA gravity flow column purification (FIG. **18***a*). As with PhCutAl , platforms derived from HsCutAl and MIF2m were readily prepared to be available for ligand assembly.

[0568] HsCutAl is a stable protein with a near identical fold as PhCutAl (FIG. **11**), however it is readily denatured to a monomer during boiling in SDS-loading buffer (FIG. **18**). Upon incubation with the crosslinking agent glutaraldehyde, we observed covalent crosslinking of monomeric subunits of SpC3-HsCutAl-DgC with an approximately threefold increase in apparent molecular weight, confirming that the protein is a trimer in solution (FIG. **18***c*) and correctly assembles as predicted from the protein structure. Other ligands were also conjugated to SpC3-HsCutAl-DgC, and the different samples of unconjugated scaffold only, conjugated scaffold+one ligand, and conjugated scaffold+two ligands were then subjected to size exclusion chromatography to show the increase in hydrodynamic radii as the assembly complexes increase in size and removal of excess ligands.

Example 9: The Capacity of Platform Designs to Produce "Cis-Orientation" can be Explored Computationally

[0569] Despite preselection of core components with suitable arrangement of N-terminal and C-terminal fusion sites, the introduction of effector or ligand proteins as well as the connection of such domains via linkers of varying length and flexibility can affect fusion protein geometry. The inventors used an implementation of Alphafold v2.0 optimized for multivalent protein assembly to predict the orientation of varying Catcher components conjugated onto different core proteins. Herein, PhCutAl , HsCutAl, Col X NC1, TNF and TL1A were predicted to assume cis-oriented display of Catcher components as a stable trimer (FIG. **19**). For PhCutAl, this confirms the crystal structure of PhCutAl alone (FIG. **11**) and various experiments (FIG. **12**, FIG. **14**). For comparison, we also tested core proteins which were previously utilized in "trans-oriented" designs, namely IMX313 (Brune et al., 2017, Bioconjugate Chemistry, 28 (5), pp. 1544-1551) and the NC1 domain of Collagen XV (Lobo et al., 2006, International Journal of Cancer, 119 (2), pp. 455-462). Herein, SpC3-IMX-DgC featuring GSGS linkers (SEQ ID 35) was not able to produce an intact core structure. Such a display may result in steric clashes upon conjugation with SpyTagged/DogTagged proteins; notably, Brune et al introduced a prolonged, rigid linker between IMX and SnoopCatcher, separating the orthogonal SpyCatcher proteins. Like IMX, SpC3-Collagen XV NC1-DgC featuring GSGS linkers (SEQ ID 33) was not able to produce an intact core structure. Upon replacing GSGS linkers (as in SEQ ID 33) with longer (G4S) 2 linkers, we observed a staggered confirmation of catchers in SpC3-Collagen XV NC1-DgC.

Example 10: The Assembled Platform can be Used for In Vitro Screening to Elucidate on the Efficacy of Ligands

[0570] We tested if PhCutAl fully conjugated with ligands against two different targets involved in cell proliferation is able to inhibit growth factor-induced cell growth (FIG. **20***a*). NCI-N87 cells were treated or mock-treated with indicated concentrations of protein assemblies containing two ligands (H6-SpC-PhCutAl-SnC: SnT-L1: L2-SpT) or one ligand only (H6-SpC-PhCutAl-SnC: SnT-L1, H6-SpC-PhCutAl-SnC: L2-SpT). Scaffold only (H6-SpC-PhCutAl-SnC), ligands only (SnT-L1, L2-SpT, SnT-L1+L2-SpT) and monoclonal control antibodies against receptor targets of L1 or L2 were used as controls. 1 h after the start of the treatment, cells were stimulated by addition of a relevant growth factor. Cell viability was measured using the MTT assay 7 days

later. Surviving fractions were normalised to mock-treated control cells. Treatment with only scaffold or ligands has no significant impact on cell viability while treatment with the conjugated assembly with only one ligand component results in a modest reduction of cell survival (around 60% cell viability at 10 nM of either protein conjugate). Treatment with the fully assembled assembly resulted in significantly decreased cell numbers, with only 35% of viable cells present after growth in the presence of 10 nM of the full assembly. This was very similar to the survival rate we observed after treatment with a combination of the monoclonal control antibodies at the same concentrations (10 nM). This suggests that the constructs are blocking both targets on the same cell. To further confirm this, a control consisting of H6-SpC-PhCutAl-SnC: SnT-L1, H6-SpC-PhCutAl-SnC: L2-SpT in the same treatment sample could be added for future assays.

[0571] We further investigated if the fully-conjugated assembly represses downstream activation of Akt and Erk1/2 (FIG. **20***b*). After starving NCI-N87 cells for 24 h in medium containing 0.2% FCS, they were treated with scaffold only or conjugated assembly of proteins SnT-L1 and L2-SpT for 1 h and then stimulated with growth factors for 1 h before preparation of whole cell extracts. Binding of the two used ligands to their receptors results in activation of two downstream signalling pathways and in the phosphorylation of their main effectors, Akt and Erk1/2. Phosphorylation levels of these two proteins were analyzed by Western Blotting. Erkl is constitutively phosphorylated in NCI-N87 cells even in the absence of growth factor stimulation. However, fully-conjugated assembly (H6-SpC-PhCutAl-SnC:SnT-L1: L2-SpT) significantly represses Erk phosphorylation. Growth factor-induced activation of Akt is prevented by unbound L2 at this concentration. Importantly, phosphorylation levels decrease further in the presence of the full assembly.

[0572] Once targets for drug development are confirmed, direct fusion of the ligands to the core protein as a multidomain polypeptide can also be achieved without the usage of Tag and Catcher components for drugs that are validated for clinical testing (FIG. **21**).

[0573] We also investigated whether a different scaffold, SpC-HsCutAl-SnC, conjugated to another pair of ligands (SpT-L1, L3-DgT) causes apoptosis within 2 days. (FIG. **20***c*). After starving NCI-N87 cells for 24 h in medium containing 0.2% FCS, they were treated with scaffold only, ligand only or conjugated assembly of proteins SpT-LI and L3-DgT for 48 h in starving medium. Cell viability was measured by MTT assay and surviving fractions were normalised to mock-treated control cells. Treatment with scaffold or either ligand only did not result in reduced cell viability. The highest decrease in cell survival was observed after treatment with the fully assembled scaffold SpC-HsCutAl-DgC: SpT-L1: L3-DgT.

Brief Description of the Informal Sequence Listing

[0574] SEQ ID NO: 1 shows the amino acid sequence of a monomer of the CutAl protein from *Pyrococcus horikoshii* (PhCutAl).

[0575] SEQ ID NO: 2 shows the amino acid sequence of a monomer of the Collagen X NC1 protein domain.

[0576] SEQ ID NO: 3 shows the amino acid sequence of a monomer of the Collagen VIII protein.

[0577] SEQ ID NO: 4 shows the amino acid sequence of "SpyCatcher". This is also referred to herein as "SpyCatcher 001".

[0578] SEQ ID NO: 5 shows the amino acid sequence of "SpyTag". This is also referred to herein as "SpyTag 001".

[0579] SEQ ID NO: 6 shows the amino acid sequence of "SpyCatcher 002".

[0580] SEQ ID NO: 7 shows the amino acid sequence of "SpyTag 002".

[0581] SEQ ID NO: 8 shows the amino acid sequence of "SpyCatcher 003".

[0582] SEQ ID NO: 9 shows the amino acid sequence of "SpyTag 003".

[0583] SEQ ID NO: 10 shows the amino acid sequence of "SpyLigase".

[0584] SEQ ID NO: 11 shows the amino acid sequence of "K-Tag".

[0585] SEQ ID NO: 12 shows the amino acid sequence of "SnoopCatcher".

[0586] SEQ ID NO: 13 shows the amino acid sequence of "SnoopTag".

[0587] SEQ ID NO: 14 shows the amino acid sequence of "SnoopLigase".

[0588] SEQ ID NO: 15 shows the amino acid sequence of "SnoopTagJr".

[0589] SEQ ID NO: 16 shows the amino acid sequence of "DogTag".

[0590] SEQ ID NO: 17 shows the amino acid sequenc of Pilin-C.

[0591] SEQ ID NO: 18 shows the amino acid sequence of "Isopeptag".

[0592] SEQ ID NO: 19 shows the amino acid sequence of a monomer of the human CutAl protein.

[0593] SEQ ID NO: 20 shows the amino acid sequence of a monomer of the his-tagged construct H6-SpyCatcher-NC1-SnoopCatcher.

[0594] SEQ ID NO: 21 shows the amino acid sequence of a monomer of the his-tagged construct H6-SpyCatcher-PhCutAl-SnoopCatcher.

[0595] SEQ ID NO: 22 shows the amino acid sequence of a H6-SpyCatcher-aH_Linker-SnoopCatcher construct.

[0596] SEQ ID NO: 23 shows the amino acid sequence of "DogCatcher".

[0597] SEQ ID NO: 24 shows the amino acid sequence of a monomer of the His-tagged construct H6-SpyCatcher003-HsCutAl-DogCatcher, with HsCutAl truncated as in SEQ ID NO: 29.

[0598] SEQ ID NO: 25 shows the amino acid sequence of a monomer of macrophage migration inhibitory factor (MIF).

[0599] SEQ ID NO: 26 shows the amino acid sequence of a monomer of human macrophage migration inhibitory factor 2 (MIF2).

[0600] SEQ ID NO: 27 shows the amino acid sequence of a monomer of human macrophage migration inhibitory factor 2 (MIF2) with mutations S62A and F99A of MIF (MIF2m).

[0601] SEQ ID NO: 28 shows the amino acid sequence of a monomer of the His-tagged construct H6-SpyCatcher003-MIF2m-DogCatcher with mutations S62A and F99A of MIF2 (MIF2m).

[0602] SEQ ID NO: 29 shows the amino acid sequence of a monomer of the human CutAl truncated based on the

resolved structure in PDB ID: 2zfh, representing an intermediate truncation between SEQ ID NO: 19 and SEQ ID NO: 60.

[0603] SEQ ID NO: 30 shows the amino acid sequence of the His-tagged fusion of DogTag to a variant (first described in Grünberg et al, 2013) of mCitrine fluorescent protein DgT-X3.

[0604] SEQ ID NO: 31 shows the amino acid sequence of a monomer of the TNF-like protein TLIA.

[0605] SEQ ID NO: 32 shows the amino acid sequence of a monomer of the construct SpyCatcher003-TL1A-DogCatcher as used in structural prediction.

[0606] SEQ ID NO: 33 shows the amino acid sequence of a monomer of the construct SpyCatcher003-Col XV NC1-DogCatcher as used in structural prediction.

[0607] SEQ ID NO: 34 shows the amino acid sequence of a monomer of the construct SpyCatcher003-MIF2m-DogCatcher as used in structural prediction with mutations S62A and F99A of MIF2.

[0608] SEQ ID NO: 35 shows the amino acid sequence of a monomer of the construct SpyCatcher003-IMX-DogCatcher as used in structural prediction.

[0609] SEQ ID NO: 36 shows the amino acid sequence of Chain A of heterotrimeric Clq head domain.

[0610] SEQ ID NO: 37 shows the amino acid sequence of Chain B of heterotrimeric Clq head domain.

[0611] SEQ ID NO: 38 shows the amino acid sequence of Chain C of heterotrimeric Clq head domain.

[0612] SEQ ID NO: 39 shows the amino acid sequence of a monomer of the CutAl from *Thermus Thermophilus* HB8.

[0613] SEQ ID NO: 40 shows the amino acid sequence of a monomer of the CutAl from *Oryza sativa.*

[0614] SEQ ID NO: 41 shows the amino acid sequence of a monomer of the CutAl from *Shewanella* sp. SIB1.

[0615] SEQ ID NO: 42 shows the amino acid sequence of a monomer of the tumor necrosis factor (TNF).

[0616] SEQ ID NO: 43 shows the amino acid sequence of a monomer of the antiparallel coiled coil hexamer.

[0617] SEQ ID NO: 44 shows the amino acid sequence of a monomer of the HIV-1 GP41 core.

[0618] SEQ ID NO: 45 shows the amino acid sequence of a monomer of a circular permutant of cytochrome c555.

[0619] SEQ ID NO: 46 shows the amino acid sequence of a monomer of the MHC class II-associated invariant chain.

[0620] SEQ ID NO: 47 shows the amino acid sequence of a monomer of the p53.

[0621] SEQ ID NO: 48 shows the amino acid sequence of a monomer of a fibrinogen-like domain.

[0622] SEQ ID NO: 49 shows the amino acid sequence of a monomer of the Collagen IV NC1 domain.

[0623] SEQ ID NO: 50 shows the amino acid sequence of a monomer of the *Bacillus subtilis* ArbB.

[0624] SEQ ID NO: 51 shows the amino acid sequence of a monomer of the phage lambda head protein D.

[0625] SEQ ID NO: 52 shows the amino acid sequence of a monomer of a domain-swapped trimer variant of HCRB-PII.

[0626] SEQ ID NO: 53 shows the amino acid sequence of a monomer of the TIL reovirus attachment protein signal.

[0627] SEQ ID NO: 54 shows the amino acid sequence of a monomer of the construct SpyCatcher003-HsCutAl-DogCatcher as used in structural prediction.

[0628] SEQ ID NO: 55 shows the amino acid sequence of a monomer of the construct SpyCatcher003-PhCutAl-DogCatcher as used in structural prediction.

[0629] SEQ ID NO: 56 shows the amino acid sequence of a monomer of the construct SpyCatcher003-Col X NC1-DogCatcher as used in structural prediction.

[0630] SEQ ID NO: 57 shows the amino acid sequence of a monomer of the construct SpyCatcher003-TNF-DogCatcher as used in structural prediction.

[0631] SEQ ID NO: 58 shows the amino acid sequence of a monomer of the TNF family protein CD40 ligand (CD40L).

[0632] SEQ ID NO: 59 shows the amino acid sequence of a monomer of human leukotriene C4 synthase.

[0633] SEQ ID NO: 60 shows the amino acid sequence of a monomer of the human CutAl as resolved in PDB ID: 2zfh, representing a truncation of SEQ ID NO: 19.

SEQ ID NO: 1
MIIVYTTFPDWESAEKVVKTLLKERLIACANLREHRAFYWWEGKIEEDKEVGAILKTREDLWEELKERIKELH

PYDVPAIIRIDVDDVNEDYLKWLIEETKK

SEQ ID NO: 2
TGMPVSAFTVILSKAYPAIGTPIPFDKILYNRQQHYDPRTGIFTCQIPGIYYFSYHVHVKGTHVWVGLYKNGT

PVMYTYDEYTKGYLDQASGSAIIDLTENDQVWLQLPNAESNGLYSSEYVHSSFSGELVAPM

SEQ ID NO: 3
EMPAFTAELTVPFPPVGAPVKFDKLLYNGRQNYNPQTGIFTCEVPGVYYFAYHVHCKGGNVWVALFKNNEPMM

YTYDEYKKGELDQASGSAVLLLRPGDQVFLQMPSEQAAGLYAGQYVHSSFSGYLLYPM

SEQ ID NO: 4
VDTLSGLSSEQGQSGDMTIEEDSATHIKESKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKY

TFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHI

SEQ ID NO: 5
AHIVMVDAYKPTK

SEQ ID NO: 6
VTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATAITFTVNEQGQVTVNGEATKGDAHT

-continued

SEQ ID NO: 7
VPTIVMVDAYKRYK

SEQ ID NO: 8
VTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHT

SEQ ID NO: 9
RGVPHIVMVDAYKRYK

SEQ ID NO: 10
GQSGDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVN

GKATKGGSGGSGGSGEDSATHI

SEQ ID NO: 11
ATHIKFSKRD

SEQ ID NO: 12
KPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAF

QIVNGEVRDVTSIVPQDIPATYEFTNGKHYITNEPIPPK

SEQ ID NO: 13
KLGDIEFIKVNK

SEQ ID NO: 14
VNKNDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPPGYKPVQN

KPIVAFQIVNGEVRDVTSIVPPGVPATYEFT

SEQ ID NO: 15
KLGSIEFIKVNK

SEQ ID NO: 16
DIPATYEFTDGKHYITNEPIPPK

SEQ ID NO: 17
ATTVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTTVNEDGNKFKGVALNTPMTKVTYTNSDKGG

SNTKTAEFDESEVTFEKPGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVATYIVGYKEGSKVP

IQFKNSLDSTTLTVKKKVSGTGGDRSKDENFGLTLKANQYYKASEKVMIEKTTKGGQAPVQTEASIDQLYHFT

LKDGESIKVTNLPVGVDYVVTEDDYKSEKYTTNVEVSPQDGAVKNIAGNSTEQETSTDKDMTI

SEQ ID NO: 18
TDKDMTITFTNKKDAE

SEQ ID NO: 19
MSGGRAPAVL LGGVASLLLS FVWMPALLPV ASRLLLLPRV LLTMASGSPP

TQPSPASDSG SGYVPGSVSA AFVTCPNEKV AKEIARAVVE KRLAACVNLI

PQITSIYEWK GKIEEDSEVL MMIKTQSSLV PALTDFVRSV HPYEVAEVIA

LPVEQGNFPY LQWVRQVTES VSDSITVLP

SEQ ID NO: 20
MGSSHHHHHHSSGVDTLSGLSSEQGQSGDMTIEEDSATHIKESKRDEDGKELAGATMELRDSSGKTISTWISD

GQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIGSGSTGMPVSAFTVILSKAY

PAIGTPIPFDKILYNRQQHYDPRTGIFTCQIPGIYYFSYHVHVKGTHVWVGLYKNGTPVMYTYDEYTKGYLDQ

ASGSAIIDLTENDQVWLQLPNAESNGLYSSEYVHSSFSGFLVAPMGSGSKPLRGAVESLQKQHPDYPDIYGAI

DQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIVPQDIPATYEF

TNGKHYITNEPIPPK

SEQ ID NO: 21
MGSSHHHHHHSSGVDTLSGLSSEQGQSGDMTIEEDSATHIKESKRDEDGKELAGATMELRDSSGKTISTWISD

GQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIGSGSMIIVYTTFPDWESAEK

-continued

VVKTLLKERLIACANLREHRAFYWWEGKIEEDKEVGAILKTREDLWEELKERIKELHPYDVPAIIRIDVDDVN

EDYLKWLIEETKKGSGSKPLRGAVESLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFEN

SEPAGYKPVQNKPIVAFQIVNGEVRDVTSIVPQDIPATYEFTNGKHYITNEPIPPK

SEQ ID NO: 22

MGSSHHHHHHSSGVDTLSGLSSEQGQSGDMTIEEDSATHIKESKRDEDGKELAGATMELRDSSGKTISTWISD

GQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIGSGSPANLKALEAQKQKEQR

QAAEELANAKKLKEQLEKGSGSKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKY

RLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIVPQDIPATYEFTNGKHYITNEPIPPK

SEQ ID NO 23

KLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLIENSE

PPGYKPVQNKPIVSFRIVDGEVRDVTSIVPQ

SEQ ID NO 24

MGSSHHHHHHSSGVTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISD

GHVKDFYLYPGKYTFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGGGGSMASGSPPTQPSPASD

SGSGYVPGSVSAAFVTCPNEKVAKEIARAVVEKRLAACVNLIPQITSIYEWKGKIEEDSEVLMMIKTQSSLVP

ALTDFVRSVHPYEVAEVIALPVEQGNFPYLQWVRQVTESVSDSITVLPGGGGSKLGEIEFIKVDKTDKKPLRG

AVFSLQKQHPDYPDIYGAIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVDG

EVRDVTSIVPQ

SEQ ID NO: 25

PMFIVNTNVPRASVPDGFLSELTQQLAQATGKPPQYIAVHVVPDQLMAFGGSSEPCALCSLHSIGKIGGAQNR

SYSKLLCGLLAERLRISPDRVYINYYDMNAANVGWNNSTFA

SEQ ID NO: 26

PFLELDTNLPANRVPAGLEKRLCAAAASILGKPADRVNVTVRPGLAMALSGSTEPCAQLSISSIGVVGTAEDN

RSHSAHFFEFLTKELALGQDRILIRFFPLESWQIGKIGTVMTEL

SEQ ID NO: 27

PFLELDTNLPANRVPAGLEKRLCAAAASILGKPADRVNVTVRPGLAMALSGSTEPCAQLSIASIGVVGTAEDN

RSHSAHFFEFLTKELALGQDRILIRFAPLESWQIGKIGTVMTEL

SEQ ID NO: 28

MGSSHHHHHHSSGVTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATMELRDSSGKTISTWISD

GHVKDFYLYPGKYTFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGGGGGGGGSGGGGSPFLEL

DTNLPANRVPAGLEKRLCAAAASILGKPADRVNVTVRPGLAMALSGSTEPCAQLSIASIGVVGTAEDNRSHSA

HFFEFLTKELALGQDRILIRFAPLESWQIGKIGTVMTFLGGGGSGGGGSGGGGSKLGEIEFIKVDKTDKKPLR

GAVFSLQKQHPDYPDIYGAIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVD

GEVRDVTSIVPQ

SEQ ID NO: 29

MASGSPPTQPSPASDSGSGYVPGSVSAAFVTCPNEKVAKEIARAVVEKRLAACVNLIPQITSIYEWKGKIEED

SEVLMMIKTQSSLVPALTDFVRSVHPYEVAEVIALPVEQGNFPYLQWVRQVTESVSDSITVLP

SEQ ID NO: 30

MSHHHHHHGSTGLEVLFQGPTGSSDIPATYEFTDGKHYITNEPIPPKGGGGSGGGGSVSKGEELFTGVVPILV

ELDGDVNGHKFSVSGEGEGDATYGKLELKFICTTGKLPVPWPTLVTTFGYGLMCFARYPDHMKQHDFFKSAMP

EGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYDYNSHNVYIMADKQKNGIK

VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLEYQSALSKDPNEKRDHMVLAEFVTAEGITLGMDE

LYK

-continued

SEQ ID NO: 31
LRADGDKPRAHLTVVRQTPTQHFKNQFPALHWEHELGLAFTKNRMNYTNKFLLIPESGDYFIYSQVTERGMTS

ECSEIRQAGRPNKPDSITVVITKVTDSYPEPTQLLMGTKSVCEVGSNWFQPIYLGAMESLQEGDKLMVNVSDI

SLVDYTKEDKTFFGAFLL

SEQ ID NO: 32
VTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGSGSLRADGDKPRAHLTVVRQTPTQHFKNQFPA

LHWEHELGLAFTKNRMNYTNKFLLIPESGDYFIYSQVTFRGMTSECSEIRQAGRPNKPDSITVVITKVTDSYP

EPTQLLMGTKSVCEVGSNWFQPIYLGAMESLQEGDKLMVNVSDISLVDYTKEDKTFFGAFLLGSGSKLGEIEF

IKVDKTDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPV

QNKPIVSFRIVDGEVRDVTSIVPQ

SEQ ID NO: 33
VTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAGSGSVTAFSNMDDMLQKAHLVIEGTFIYLRDSTEF

FIRVRDGWKKLQLGELIPIPADSPPPPALSSNPGSGSKLGEIEFIKVDKTDKKPLRGAVESLQKQHPDYPDIY

GAIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVDGEVRDVTSIVPQ

SEQ ID NO: 34
VTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGSGSPFLELDTNLPANRVPAGLEKRLCAAAASI

LGKPADRVNVTVRPGLAMALSGSTEPCAQLSIASIGVVGTAEDNRSHSAHFFEFLTKELALGQDRILIRFAPL

ESWQIGKIGTVMTFLGSGSKLGEIEFIKVDKTDKKPLRGAVESLQKQHPDYPDIYGAIDQNGTYQDVRTGEDG

KLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVDGEVRDVTSIVPQ

SEQ ID NO: 35
VTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGSGSKKQGDADVCGEVAYIQSVVSDCHVPTAEL

RTLLEIRKLFLEIQKLKVELQGLSKEGSGSKLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDIYGAIDQNG

TYQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVDGEVRDVTSIVPQ

SEQ ID NO: 36
QPRPAFSAIRRNPPMGGNVVIFDTVITNQEEPYQNHSGRFVCTVPGYYYFTFQVLSQWEICLSIVSSSRGQVR

RSLGFCDTTNKGLFQVVSGGMVLQLQQGDQVWVEKDPKKGHIYQGSEADSVESGELIFPS

SEQ ID NO: 37
TQKIAFSATRTINVPLRRDQTIREDHVITNMNNNYEPRSGKFTCKVPGLYYFTYHASSRGNLCVNLMRGRERA

QKVVTFCDYAYNTFQVTTGGMVLKLEQGENVFLQATDKNSLLGMEGANSIFSGELLEPD

SEQ ID NO: 38
KFQSVFTVTRQTHQPPAPNSLIRFNAVLTNPQGDYDTSTGKFTCKVPGLYYFVYHASHTANLCVLLYRSGVKV

VTFCGHTSKTNQVNSGGVLLRLQVGEEVWLAVNDYYDMVGIQGSDSVFSGELLFPD

SEQ ID NO: 39
MEEVVLITVPSEEVARTIAKALVEERLAACVNIVPGLTSIYRWQGEVVEDQELLLLVKTTTHAFPKLKERVKA

LHPYTVPEIVALPIAEGNREYLDWLRENTG

SEQ ID NO: 40
STTVPSIVVYVTVPNKEAGKRLAGSIISEKLAACVNIVPGIESVYWWEGKVQTDAEELLIIKTRESLLDALTE

HVKANHEYDVPEVIALPIKGGNLKYLEWLKNSTR

SEQ ID NO: 41
KPEQLLIFTTCPDADIACRIATALVEAKLAACVQIGQAVESIYQWDNNICQSHEVPMQIKCMTTDYPAIEQLV

ITMHPYEVPEFIATPIIGGFGPYLQWIKDNSPS

-continued

SEQ ID NO: 42
RTPSDKPVAHVVANPQAEGQLQWLNRRANALLANGVELRDNQLVVPSEGLYLIYSQVLFKGQGCPSTHVLLTH

TISRIAVSYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPIYLGGVFQLEKGDRLSAEINRPDYLLFAESGQVY

FGIIAL

SEQ ID NO: 43
ELAQAFKEIAKAFKEIAKAFEFIAQAIE

SEQ ID NO: 44
IVQQQNNLLRAIEAQQHLLQLTVWAIKQLQARSGGRGGWMEWDREINNYTSLIHSLIEESQ

SEQ ID NO: 45
VDPAKEAIMKPQLTMLKGLSDAELKALADEILRIAKQAQEKQQQDVAKAIFQQKGCGSCHQANVDTVGPSLAK

IAQAYAGKEDQLIKFLKGEAPAI

SEQ ID NO: 46
YGNMTEDHVMHLLQNADPLKVYPPLKGSFPENLRHLKNTMETIDWKVFESWMHHWLLFEMSRHSLEQKPTDAP

PK

SEQ ID NO: 48
SRPRDCLDVLLSGQQDDGVYSVFPTHYPAGFQVYCDMRTDGGGWTVFQRREDGSVNFFRGWDAYRDGFGRLTG

EHWLGLKRIHALTTQAAYELHVDLEDFENGTAYARYGSFGVGLFSVDPEEDGYPLTVADYSGTAGDSLLKHSG

MRFTTKDRDSDHSENNCAAFYRGAWWYRNCHTSNLNGQYLRGAHASYADGVEWSSWTGWQYSLKFSEMKIRPV

R

SEQ ID NO: 49
VDHGFLVTRHSQTIDDPQCPSGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKESTMPFLFCNINNVCNFA

SRNDYSYWLSTPEPMPMSMAPITGENIRPFISRCAVCEAPAMVMAVHSQTIQIPPCPSGWSSLWIGYSFVMHT

SAGAEGSGQALASPGSCLEEFRSAPFIECHGRGTCNYYANAYSFWLATIERSEMFKKPTPSTLKAGELRTHVS

RCQVCMRRT

SEQ ID NO: 50
FMKSTGIVRKVDELGRVVIPIELRRTLGIAEKDALEIYVDDEKIILKKYKPN

SEQ ID NO: 51
SDPAHTATAPGGLSAKAPAMTPLMLDTSSRKLVAWDGTTDGAAVGILAVAADQTSTTLTFYKSGTFRYEDVLW

PEAASDETKKRTAFAGTAISIV

SEQ ID NO: 52
TRDFNGTWEMESNENFEGYMKALDIDFATRKIAVRLTFTDVIDQDGDNFKTKATSTELNYDEDETVGVEFDEY

TKSLDNRHVKALVTWEGDVLVCVQKGEKENRGWKKWIEGDKLYLELTCGDQVCRQVFKKK

SEQ ID NO: 53
LPTYRYPLELDTANNRVQVADRFGMRTGTWTGQLQYQHPQLSWRANVTLNLMKVDDWLVLSFSQMTTNSIMAD

GKFVINFVSGLSSGWQTGDTEPSSTIDPLSTTFAAVQFLNNGQRIDAFRIMGVSEWTDGELEIKNYGGTYTGH

TQVYWAPWTIMYPCNV

SEQ ID NO: 54
VTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGSGSMASGSPPTQPSPASDSGSGYVPGSVSAAF

VTCPNEKVAKEIARAVVEKRLAACVNLIPQITSIYEWKGKIEEDSEVLMMIKTQSSLVPALTDFVRSVHPYEV

AEVIALPVEQGNFPYLQWVRQVTESVSDSITVLPGSGSKLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDI

YGAIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVDGEVRDVTSIVPQ

SEQ ID NO: 55
VTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGSGSMIIVYTTFPDWESAEKVVKTLLKERLIAC

-continued

ANLREHRAFYWWEGKIEEDKEVGAILKTREDLWEELKERIKELHPYDVPAIIRIDVDDVNEDYLKWLIEETKK

GSGSKLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLI

ENSEPPGYKPVQNKPIVSFRIVDGEVRDVTSIVPQ

SEQ ID NO: 56

VTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGSGSTGMPVSAFTVILSKAYPAIGTPIPEDKIL

YNRQQHYDPRTGIFTCQIPGIYYFSYHVHVKGTHVWVGLYKNGTPVMYTYDEYTKGYLDQASGSAIIDLTEND

QVWLQLPNAESNGLYSSEYVHSSFSGELVAPMGSGSKLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDIYG

AIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVDGEVRDVTSIVPQ

SEQ ID NO: 57

VTTLSGLSGEQGPSGDMTTEEDSATHIKESKRDEDGRELAGATMELRDSSGKTISTWISDGHVKDFYLYPGKY

TFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHTGSGSRTPSDKPVAHVVANPQAEGQLQWLNRRAN

ALLANGVELRDNQLVVPSEGLYLIYSQVLFKGQGCPSTHVLLTHTISRIAVSYQTKVNLLSAIKSPCQRETPE

GAEAKPWYEPIYLGGVFQLEKGDRLSAEINRPDYLLFAESGQVYFGIIALGSGSKLGEIEFIKVDKTDKKPLR

GAVFSLQKQHPDYPDIYGAIDQNGTYQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVD

GEVRDVTSIVPQ

SEQ ID NO: 58

QIAAHVISEASSKTTSVLQWAEKGYYTMSNNLVTLENGKQLTVKRQGLYYIYAQVTFCSNREASSQAPFIASL

CLKSPGRFERILLRAANTHSSAKPCGQQSIHLGGVFELQPGASVFVNVTDPSQVSHGTGFTSFGLLKL

SEQ ID NO: 59

MKDEVALLAAVTLLGVLLQAYFSLQVISARRAFRVSPPLTTGPPEFERVYRAQVNCSEYFPLFLATLWVAGIF

FHEGAAALCGLVYLFARLRYFQGYARSAQLRLAPLYASARALWLLVALAALGLLAHFLPAALRAALLGRLRTL

SEQ ID NO: 60

SGYVPGSVSAAFVTCPNEKVAKEIARAVVEKRLAACVNLIPQITSIYEWKGKIEEDSEVLMMIKTQSSLVPAL

TDFVRSVHPYEVAEVIALPVEQGNFPYLQWVRQVT

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 78

<210> SEQ ID NO 1
<211> LENGTH: 102
<212> TYPE: PRT
<213> ORGANISM: Pyrococcus horikoshii

<400> SEQUENCE: 1

Met Ile Ile Val Tyr Thr Thr Phe Pro Asp Trp Glu Ser Ala Glu Lys
1 5 10 15

Val Val Lys Thr Leu Leu Lys Glu Arg Leu Ile Ala Cys Ala Asn Leu
20 25 30

Arg Glu His Arg Ala Phe Tyr Trp Trp Glu Gly Lys Ile Glu Glu Asp
35 40 45

Lys Glu Val Gly Ala Ile Leu Lys Thr Arg Glu Asp Leu Trp Glu Glu
50 55 60

Leu Lys Glu Arg Ile Lys Glu Leu His Pro Tyr Asp Val Pro Ala Ile
65 70 75 80

Ile Arg Ile Asp Val Asp Asp Val Asn Glu Asp Tyr Leu Lys Trp Leu
85 90 95

-continued

```
Ile Glu Glu Thr Lys Lys
            100


<210> SEQ ID NO 2
<211> LENGTH: 134
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

Thr Gly Met Pro Val Ser Ala Phe Thr Val Ile Leu Ser Lys Ala Tyr
1               5                   10                  15

Pro Ala Ile Gly Thr Pro Ile Pro Phe Asp Lys Ile Leu Tyr Asn Arg
            20                  25                  30

Gln Gln His Tyr Asp Pro Arg Thr Gly Ile Phe Thr Cys Gln Ile Pro
        35                  40                  45

Gly Ile Tyr Tyr Phe Ser Tyr His Val His Val Lys Gly Thr His Val
    50                  55                  60

Trp Val Gly Leu Tyr Lys Asn Gly Thr Pro Val Met Tyr Thr Tyr Asp
65                  70                  75                  80

Glu Tyr Thr Lys Gly Tyr Leu Asp Gln Ala Ser Gly Ser Ala Ile Ile
                85                  90                  95

Asp Leu Thr Glu Asn Asp Gln Val Trp Leu Gln Leu Pro Asn Ala Glu
            100                 105                 110

Ser Asn Gly Leu Tyr Ser Ser Glu Tyr Val His Ser Ser Phe Ser Gly
        115                 120                 125

Phe Leu Val Ala Pro Met
    130


<210> SEQ ID NO 3
<211> LENGTH: 131
<212> TYPE: PRT
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 3

Glu Met Pro Ala Phe Thr Ala Glu Leu Thr Val Pro Phe Pro Pro Val
1               5                   10                  15

Gly Ala Pro Val Lys Phe Asp Lys Leu Leu Tyr Asn Gly Arg Gln Asn
            20                  25                  30

Tyr Asn Pro Gln Thr Gly Ile Phe Thr Cys Glu Val Pro Gly Val Tyr
        35                  40                  45

Tyr Phe Ala Tyr His Val His Cys Lys Gly Gly Asn Val Trp Val Ala
    50                  55                  60

Leu Phe Lys Asn Asn Glu Pro Met Met Tyr Thr Tyr Asp Glu Tyr Lys
65                  70                  75                  80

Lys Gly Phe Leu Asp Gln Ala Ser Gly Ser Ala Val Leu Leu Leu Arg
                85                  90                  95

Pro Gly Asp Gln Val Phe Leu Gln Met Pro Ser Glu Gln Ala Ala Gly
            100                 105                 110

Leu Tyr Ala Gly Gln Tyr Val His Ser Ser Phe Ser Gly Tyr Leu Leu
        115                 120                 125

Tyr Pro Met
    130


<210> SEQ ID NO 4
<211> LENGTH: 113
```

-continued

<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 4

Val Asp Thr Leu Ser Gly Leu Ser Ser Glu Gln Gly Gln Ser Gly Asp
1 5 10 15

Met Thr Ile Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
20 25 30

Asp Glu Asp Gly Lys Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
35 40 45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly Gln Val Lys
50 55 60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65 70 75 80

Pro Asp Gly Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu
85 90 95

Gln Gly Gln Val Thr Val Asn Gly Lys Ala Thr Lys Gly Asp Ala His
100 105 110

Ile

<210> SEQ ID NO 5
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 5

Ala His Ile Val Met Val Asp Ala Tyr Lys Pro Thr Lys
1 5 10

<210> SEQ ID NO 6
<211> LENGTH: 113
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 6

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1 5 10 15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
20 25 30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
35 40 45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
50 55 60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65 70 75 80

Pro Asp Gly Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu
85 90 95

Gln Gly Gln Val Thr Val Asn Gly Glu Ala Thr Lys Gly Asp Ala His
100 105 110

Thr

<210> SEQ ID NO 7
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

-continued

<400> SEQUENCE: 7

Val Pro Thr Ile Val Met Val Asp Ala Tyr Lys Arg Tyr Lys
1 5 10

<210> SEQ ID NO 8
<211> LENGTH: 113
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 8

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1 5 10 15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
20 25 30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
35 40 45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
50 55 60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65 70 75 80

Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
85 90 95

Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His
100 105 110

Thr

<210> SEQ ID NO 9
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 9

Arg Gly Val Pro His Ile Val Met Val Asp Ala Tyr Lys Arg Tyr Lys
1 5 10 15

<210> SEQ ID NO 10
<211> LENGTH: 95
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 10

Gly Gln Ser Gly Asp Gly Lys Glu Leu Ala Gly Ala Thr Met Glu Leu
1 5 10 15

Arg Asp Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly Gln
20 25 30

Val Lys Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr
35 40 45

Ala Ala Pro Asp Gly Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val
50 55 60

Asn Glu Gln Gly Gln Val Thr Val Asn Gly Lys Ala Thr Lys Gly Gly
65 70 75 80

Ser Gly Gly Ser Gly Gly Ser Gly Glu Asp Ser Ala Thr His Ile
85 90 95

<210> SEQ ID NO 11
<211> LENGTH: 10

-continued

<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 11

Ala Thr His Ile Lys Phe Ser Lys Arg Asp
1 5 10

<210> SEQ ID NO 12
<211> LENGTH: 112
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pneumoniae

<400> SEQUENCE: 12

Lys Pro Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp
1 5 10 15

Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asn
20 25 30

Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Lys Asn Leu Ser Asp
35 40 45

Gly Lys Tyr Arg Leu Phe Glu Asn Ser Glu Pro Ala Gly Tyr Lys Pro
50 55 60

Val Gln Asn Lys Pro Ile Val Ala Phe Gln Ile Val Asn Gly Glu Val
65 70 75 80

Arg Asp Val Thr Ser Ile Val Pro Gln Asp Ile Pro Ala Thr Tyr Glu
85 90 95

Phe Thr Asn Gly Lys His Tyr Ile Thr Asn Glu Pro Ile Pro Pro Lys
100 105 110

<210> SEQ ID NO 13
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pneumoniae

<400> SEQUENCE: 13

Lys Leu Gly Asp Ile Glu Phe Ile Lys Val Asn Lys
1 5 10

<210> SEQ ID NO 14
<211> LENGTH: 104
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pneumoniae

<400> SEQUENCE: 14

Val Asn Lys Asn Asp Lys Lys Pro Leu Arg Gly Ala Val Phe Ser Leu
1 5 10 15

Gln Lys Gln His Pro Asp Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln
20 25 30

Asn Gly Thr Tyr Gln Asn Val Arg Thr Gly Glu Asp Gly Lys Leu Thr
35 40 45

Phe Lys Asn Leu Ser Asp Gly Lys Tyr Arg Leu Phe Glu Asn Ser Glu
50 55 60

Pro Pro Gly Tyr Lys Pro Val Gln Asn Lys Pro Ile Val Ala Phe Gln
65 70 75 80

Ile Val Asn Gly Glu Val Arg Asp Val Thr Ser Ile Val Pro Pro Gly
85 90 95

Val Pro Ala Thr Tyr Glu Phe Thr
100

-continued

<210> SEQ ID NO 15
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pneumoniae

<400> SEQUENCE: 15

Lys Leu Gly Ser Ile Glu Phe Ile Lys Val Asn Lys
1 5 10

<210> SEQ ID NO 16
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pneumoniae

<400> SEQUENCE: 16

Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asp Gly Lys His Tyr Ile Thr
1 5 10 15

Asn Glu Pro Ile Pro Pro Lys
20

<210> SEQ ID NO 17
<211> LENGTH: 282
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 17

Ala Thr Thr Val His Gly Glu Thr Val Val Asn Gly Ala Lys Leu Thr
1 5 10 15

Val Thr Lys Asn Leu Asp Leu Val Asn Ser Asn Ala Leu Ile Pro Asn
20 25 30

Thr Asp Phe Thr Phe Lys Ile Glu Pro Asp Thr Thr Val Asn Glu Asp
35 40 45

Gly Asn Lys Phe Lys Gly Val Ala Leu Asn Thr Pro Met Thr Lys Val
50 55 60

Thr Tyr Thr Asn Ser Asp Lys Gly Gly Ser Asn Thr Lys Thr Ala Glu
65 70 75 80

Phe Asp Phe Ser Glu Val Thr Phe Glu Lys Pro Gly Val Tyr Tyr Tyr
85 90 95

Lys Val Thr Glu Glu Lys Ile Asp Lys Val Pro Gly Val Ser Tyr Asp
100 105 110

Thr Thr Ser Tyr Thr Val Gln Val His Val Leu Trp Asn Glu Glu Gln
115 120 125

Gln Lys Pro Val Ala Thr Tyr Ile Val Gly Tyr Lys Glu Gly Ser Lys
130 135 140

Val Pro Ile Gln Phe Lys Asn Ser Leu Asp Ser Thr Thr Leu Thr Val
145 150 155 160

Lys Lys Lys Val Ser Gly Thr Gly Gly Asp Arg Ser Lys Asp Phe Asn
165 170 175

Phe Gly Leu Thr Leu Lys Ala Asn Gln Tyr Tyr Lys Ala Ser Glu Lys
180 185 190

Val Met Ile Glu Lys Thr Thr Lys Gly Gly Gln Ala Pro Val Gln Thr
195 200 205

Glu Ala Ser Ile Asp Gln Leu Tyr His Phe Thr Leu Lys Asp Gly Glu
210 215 220

-continued

```
Ser Ile Lys Val Thr Asn Leu Pro Val Gly Val Asp Tyr Val Val Thr
225                 230                 235                 240

Glu Asp Asp Tyr Lys Ser Glu Lys Tyr Thr Thr Asn Val Glu Val Ser
                245                 250                 255

Pro Gln Asp Gly Ala Val Lys Asn Ile Ala Gly Asn Ser Thr Glu Gln
            260                 265                 270

Glu Thr Ser Thr Asp Lys Asp Met Thr Ile
        275                 280


<210> SEQ ID NO 18
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 18

Thr Asp Lys Asp Met Thr Ile Thr Phe Thr Asn Lys Lys Asp Ala Glu
1               5                   10                  15


<210> SEQ ID NO 19
<211> LENGTH: 179
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

Met Ser Gly Gly Arg Ala Pro Ala Val Leu Leu Gly Gly Val Ala Ser
1               5                   10                  15

Leu Leu Leu Ser Phe Val Trp Met Pro Ala Leu Leu Pro Val Ala Ser
            20                  25                  30

Arg Leu Leu Leu Leu Pro Arg Val Leu Leu Thr Met Ala Ser Gly Ser
        35                  40                  45

Pro Pro Thr Gln Pro Ser Pro Ala Ser Asp Ser Gly Ser Gly Tyr Val
    50                  55                  60

Pro Gly Ser Val Ser Ala Ala Phe Val Thr Cys Pro Asn Glu Lys Val
65                  70                  75                  80

Ala Lys Glu Ile Ala Arg Ala Val Val Glu Lys Arg Leu Ala Ala Cys
                85                  90                  95

Val Asn Leu Ile Pro Gln Ile Thr Ser Ile Tyr Glu Trp Lys Gly Lys
            100                 105                 110

Ile Glu Glu Asp Ser Glu Val Leu Met Met Ile Lys Thr Gln Ser Ser
        115                 120                 125

Leu Val Pro Ala Leu Thr Asp Phe Val Arg Ser Val His Pro Tyr Glu
    130                 135                 140

Val Ala Glu Val Ile Ala Leu Pro Val Glu Gln Gly Asn Phe Pro Tyr
145                 150                 155                 160

Leu Gln Trp Val Arg Gln Val Thr Glu Ser Val Ser Asp Ser Ile Thr
                165                 170                 175

Val Leu Pro


<210> SEQ ID NO 20
<211> LENGTH: 380
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the his-tagged construct
      H6-SpyCatcher-NC1-SnoopCatcher

<400> SEQUENCE: 20
```

-continued

```
Met Gly Ser Ser His His His His His His Ser Ser Gly Val Asp Thr
1               5                   10                  15

Leu Ser Gly Leu Ser Ser Glu Gln Gly Gln Ser Gly Asp Met Thr Ile
            20                  25                  30

Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg Asp Glu Asp
        35                  40                  45

Gly Lys Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp Ser Ser Gly
    50                  55                  60

Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly Gln Val Lys Asp Phe Tyr
65                  70                  75                  80

Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala Pro Asp Gly
                85                  90                  95

Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu Gln Gly Gln
            100                 105                 110

Val Thr Val Asn Gly Lys Ala Thr Lys Gly Asp Ala His Ile Gly Ser
        115                 120                 125

Gly Ser Thr Gly Met Pro Val Ser Ala Phe Thr Val Ile Leu Ser Lys
    130                 135                 140

Ala Tyr Pro Ala Ile Gly Thr Pro Ile Pro Phe Asp Lys Ile Leu Tyr
145                 150                 155                 160

Asn Arg Gln Gln His Tyr Asp Pro Arg Thr Gly Ile Phe Thr Cys Gln
                165                 170                 175

Ile Pro Gly Ile Tyr Tyr Phe Ser Tyr His Val His Val Lys Gly Thr
            180                 185                 190

His Val Trp Val Gly Leu Tyr Lys Asn Gly Thr Pro Val Met Tyr Thr
        195                 200                 205

Tyr Asp Glu Tyr Thr Lys Gly Tyr Leu Asp Gln Ala Ser Gly Ser Ala
    210                 215                 220

Ile Ile Asp Leu Thr Glu Asn Asp Gln Val Trp Leu Gln Leu Pro Asn
225                 230                 235                 240

Ala Glu Ser Asn Gly Leu Tyr Ser Ser Glu Tyr Val His Ser Ser Phe
                245                 250                 255

Ser Gly Phe Leu Val Ala Pro Met Gly Ser Gly Ser Lys Pro Leu Arg
            260                 265                 270

Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr Pro Asp Ile
        275                 280                 285

Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asn Val Arg Thr Gly
    290                 295                 300

Glu Asp Gly Lys Leu Thr Phe Lys Asn Leu Ser Asp Gly Lys Tyr Arg
305                 310                 315                 320

Leu Phe Glu Asn Ser Glu Pro Ala Gly Tyr Lys Pro Val Gln Asn Lys
                325                 330                 335

Pro Ile Val Ala Phe Gln Ile Val Asn Gly Glu Val Arg Asp Val Thr
            340                 345                 350

Ser Ile Val Pro Gln Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly
        355                 360                 365

Lys His Tyr Ile Thr Asn Glu Pro Ile Pro Pro Lys
    370                 375                 380
```

<210> SEQ ID NO 21
<211> LENGTH: 348

-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the his-tagged construct
H6-SpyCatcher-PhCutA1-SnoopCatcher

<400> SEQUENCE: 21

Met Gly Ser Ser His His His His His His Ser Ser Gly Val Asp Thr
1 5 10 15

Leu Ser Gly Leu Ser Ser Glu Gln Gly Gln Ser Gly Asp Met Thr Ile
20 25 30

Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg Asp Glu Asp
35 40 45

Gly Lys Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp Ser Ser Gly
50 55 60

Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly Gln Val Lys Asp Phe Tyr
65 70 75 80

Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala Pro Asp Gly
85 90 95

Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu Gln Gly Gln
100 105 110

Val Thr Val Asn Gly Lys Ala Thr Lys Gly Asp Ala His Ile Gly Ser
115 120 125

Gly Ser Met Ile Ile Val Tyr Thr Thr Phe Pro Asp Trp Glu Ser Ala
130 135 140

Glu Lys Val Val Lys Thr Leu Leu Lys Glu Arg Leu Ile Ala Cys Ala
145 150 155 160

Asn Leu Arg Glu His Arg Ala Phe Tyr Trp Trp Glu Gly Lys Ile Glu
165 170 175

Glu Asp Lys Glu Val Gly Ala Ile Leu Lys Thr Arg Glu Asp Leu Trp
180 185 190

Glu Glu Leu Lys Glu Arg Ile Lys Glu Leu His Pro Tyr Asp Val Pro
195 200 205

Ala Ile Ile Arg Ile Asp Val Asp Asp Val Asn Glu Asp Tyr Leu Lys
210 215 220

Trp Leu Ile Glu Glu Thr Lys Lys Gly Ser Gly Ser Lys Pro Leu Arg
225 230 235 240

Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr Pro Asp Ile
245 250 255

Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asn Val Arg Thr Gly
260 265 270

Glu Asp Gly Lys Leu Thr Phe Lys Asn Leu Ser Asp Gly Lys Tyr Arg
275 280 285

Leu Phe Glu Asn Ser Glu Pro Ala Gly Tyr Lys Pro Val Gln Asn Lys
290 295 300

Pro Ile Val Ala Phe Gln Ile Val Asn Gly Glu Val Arg Asp Val Thr
305 310 315 320

Ser Ile Val Pro Gln Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly
325 330 335

Lys His Tyr Ile Thr Asn Glu Pro Ile Pro Pro Lys
340 345

<210> SEQ ID NO 22

-continued

<211> LENGTH: 280
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: H6-SpyCatcher-aH_Linker-SnoopCatcher construct

<400> SEQUENCE: 22

Met Gly Ser Ser His His His His His His Ser Ser Gly Val Asp Thr
1 5 10 15

Leu Ser Gly Leu Ser Ser Glu Gln Gly Gln Ser Gly Asp Met Thr Ile
20 25 30

Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg Asp Glu Asp
35 40 45

Gly Lys Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp Ser Ser Gly
50 55 60

Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly Gln Val Lys Asp Phe Tyr
65 70 75 80

Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala Pro Asp Gly
85 90 95

Tyr Glu Val Ala Thr Ala Ile Thr Phe Thr Val Asn Glu Gln Gly Gln
100 105 110

Val Thr Val Asn Gly Lys Ala Thr Lys Gly Asp Ala His Ile Gly Ser
115 120 125

Gly Ser Pro Ala Asn Leu Lys Ala Leu Glu Ala Gln Lys Gln Lys Glu
130 135 140

Gln Arg Gln Ala Ala Glu Glu Leu Ala Asn Ala Lys Lys Leu Lys Glu
145 150 155 160

Gln Leu Glu Lys Gly Ser Gly Ser Lys Pro Leu Arg Gly Ala Val Phe
165 170 175

Ser Leu Gln Lys Gln His Pro Asp Tyr Pro Asp Ile Tyr Gly Ala Ile
180 185 190

Asp Gln Asn Gly Thr Tyr Gln Asn Val Arg Thr Gly Glu Asp Gly Lys
195 200 205

Leu Thr Phe Lys Asn Leu Ser Asp Gly Lys Tyr Arg Leu Phe Glu Asn
210 215 220

Ser Glu Pro Ala Gly Tyr Lys Pro Val Gln Asn Lys Pro Ile Val Ala
225 230 235 240

Phe Gln Ile Val Asn Gly Glu Val Arg Asp Val Thr Ser Ile Val Pro
245 250 255

Gln Asp Ile Pro Ala Thr Tyr Glu Phe Thr Asn Gly Lys His Tyr Ile
260 265 270

Thr Asn Glu Pro Ile Pro Pro Lys
275 280

<210> SEQ ID NO 23
<211> LENGTH: 104
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pneumoniae

<400> SEQUENCE: 23

Lys Leu Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys Lys
1 5 10 15

Pro Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr
20 25 30

-continued

```
Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp Val
        35                  40                  45

Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp Gly
    50                  55                  60

Lys Tyr Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro Val
65                  70                  75                  80

Gln Asn Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val Arg
                85                  90                  95

Asp Val Thr Ser Ile Val Pro Gln
            100


<210> SEQ ID NO 24
<211> LENGTH: 376
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the His-tagged construct
      H6-SpyCatcher003-HsCutA1-DogCatcher, with HsCutA1 truncated as in
      SEQ ID NO: 29

<400> SEQUENCE: 24

Met Gly Ser Ser His His His His His His Ser Ser Gly Val Thr Thr
1               5                   10                  15

Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp Met Thr Thr
            20                  25                  30

Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg Asp Glu Asp
        35                  40                  45

Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp Ser Ser Gly
    50                  55                  60

Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys Asp Phe Tyr
65                  70                  75                  80

Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala Pro Asp Gly
                85                  90                  95

Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu Asp Gly Gln
            100                 105                 110

Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His Thr Gly Gly
        115                 120                 125

Gly Gly Ser Met Ala Ser Gly Ser Pro Pro Thr Gln Pro Ser Pro Ala
    130                 135                 140

Ser Asp Ser Gly Ser Gly Tyr Val Pro Gly Ser Val Ser Ala Ala Phe
145                 150                 155                 160

Val Thr Cys Pro Asn Glu Lys Val Ala Lys Glu Ile Ala Arg Ala Val
                165                 170                 175

Val Glu Lys Arg Leu Ala Ala Cys Val Asn Leu Ile Pro Gln Ile Thr
            180                 185                 190

Ser Ile Tyr Glu Trp Lys Gly Lys Ile Glu Glu Asp Ser Glu Val Leu
        195                 200                 205

Met Met Ile Lys Thr Gln Ser Ser Leu Val Pro Ala Leu Thr Asp Phe
    210                 215                 220

Val Arg Ser Val His Pro Tyr Glu Val Ala Glu Val Ile Ala Leu Pro
225                 230                 235                 240

Val Glu Gln Gly Asn Phe Pro Tyr Leu Gln Trp Val Arg Gln Val Thr
                245                 250                 255

Glu Ser Val Ser Asp Ser Ile Thr Val Leu Pro Gly Gly Gly Gly Ser
```

-continued

```
            260                 265                 270

Lys Leu Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys Lys
        275                 280                 285

Pro Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr
    290                 295                 300

Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp Val
305                 310                 315                 320

Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp Gly
                325                 330                 335

Lys Tyr Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro Val
            340                 345                 350

Gln Asn Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val Arg
        355                 360                 365

Asp Val Thr Ser Ile Val Pro Gln
    370                 375


<210> SEQ ID NO 25
<211> LENGTH: 114
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

Pro Met Phe Ile Val Asn Thr Asn Val Pro Arg Ala Ser Val Pro Asp
1               5                   10                  15

Gly Phe Leu Ser Glu Leu Thr Gln Gln Leu Ala Gln Ala Thr Gly Lys
            20                  25                  30

Pro Pro Gln Tyr Ile Ala Val His Val Val Pro Asp Gln Leu Met Ala
        35                  40                  45

Phe Gly Gly Ser Ser Glu Pro Cys Ala Leu Cys Ser Leu His Ser Ile
    50                  55                  60

Gly Lys Ile Gly Gly Ala Gln Asn Arg Ser Tyr Ser Lys Leu Leu Cys
65                  70                  75                  80

Gly Leu Leu Ala Glu Arg Leu Arg Ile Ser Pro Asp Arg Val Tyr Ile
                85                  90                  95

Asn Tyr Tyr Asp Met Asn Ala Ala Asn Val Gly Trp Asn Asn Ser Thr
            100                 105                 110

Phe Ala


<210> SEQ ID NO 26
<211> LENGTH: 117
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

Pro Phe Leu Glu Leu Asp Thr Asn Leu Pro Ala Asn Arg Val Pro Ala
1               5                   10                  15

Gly Leu Glu Lys Arg Leu Cys Ala Ala Ala Ala Ser Ile Leu Gly Lys
            20                  25                  30

Pro Ala Asp Arg Val Asn Val Thr Val Arg Pro Gly Leu Ala Met Ala
        35                  40                  45

Leu Ser Gly Ser Thr Glu Pro Cys Ala Gln Leu Ser Ile Ser Ser Ile
    50                  55                  60

Gly Val Val Gly Thr Ala Glu Asp Asn Arg Ser His Ser Ala His Phe
65                  70                  75                  80
```

-continued

```
Phe Glu Phe Leu Thr Lys Glu Leu Ala Leu Gly Gln Asp Arg Ile Leu
                85                  90                  95

Ile Arg Phe Phe Pro Leu Glu Ser Trp Gln Ile Gly Lys Ile Gly Thr
            100                 105                 110

Val Met Thr Phe Leu
        115


<210> SEQ ID NO 27
<211> LENGTH: 117
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 27

Pro Phe Leu Glu Leu Asp Thr Asn Leu Pro Ala Asn Arg Val Pro Ala
1               5                   10                  15

Gly Leu Glu Lys Arg Leu Cys Ala Ala Ala Ala Ser Ile Leu Gly Lys
            20                  25                  30

Pro Ala Asp Arg Val Asn Val Thr Val Arg Pro Gly Leu Ala Met Ala
        35                  40                  45

Leu Ser Gly Ser Thr Glu Pro Cys Ala Gln Leu Ser Ile Ala Ser Ile
    50                  55                  60

Gly Val Val Gly Thr Ala Glu Asp Asn Arg Ser His Ser Ala His Phe
65                  70                  75                  80

Phe Glu Phe Leu Thr Lys Glu Leu Ala Leu Gly Gln Asp Arg Ile Leu
                85                  90                  95

Ile Arg Phe Ala Pro Leu Glu Ser Trp Gln Ile Gly Lys Ile Gly Thr
            100                 105                 110

Val Met Thr Phe Leu
        115


<210> SEQ ID NO 28
<211> LENGTH: 377
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the His-tagged construct
      H6-SpyCatcher003-MIF2m-DogCatcher with mutations S62A and F99A of
      MIF2 (MIF2m)

<400> SEQUENCE: 28

Met Gly Ser Ser His His His His His His Ser Ser Gly Val Thr Thr
1               5                   10                  15

Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp Met Thr Thr
            20                  25                  30

Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg Asp Glu Asp
        35                  40                  45

Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp Ser Ser Gly
    50                  55                  60

Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys Asp Phe Tyr
65                  70                  75                  80

Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala Pro Asp Gly
                85                  90                  95

Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu Asp Gly Gln
            100                 105                 110

Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His Thr Gly Gly
```

-continued

```
        115                 120                 125

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Pro Phe Leu
    130                 135                 140

Glu Leu Asp Thr Asn Leu Pro Ala Asn Arg Val Pro Ala Gly Leu Glu
145                 150                 155                 160

Lys Arg Leu Cys Ala Ala Ala Ala Ser Ile Leu Gly Lys Pro Ala Asp
                165                 170                 175

Arg Val Asn Val Thr Val Arg Pro Gly Leu Ala Met Ala Leu Ser Gly
            180                 185                 190

Ser Thr Glu Pro Cys Ala Gln Leu Ser Ile Ala Ser Ile Gly Val Val
        195                 200                 205

Gly Thr Ala Glu Asp Asn Arg Ser His Ser Ala His Phe Phe Glu Phe
    210                 215                 220

Leu Thr Lys Glu Leu Ala Leu Gly Gln Asp Arg Ile Leu Ile Arg Phe
225                 230                 235                 240

Ala Pro Leu Glu Ser Trp Gln Ile Gly Lys Ile Gly Thr Val Met Thr
                245                 250                 255

Phe Leu Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
            260                 265                 270

Ser Lys Leu Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys
        275                 280                 285

Lys Pro Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp
    290                 295                 300

Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp
305                 310                 315                 320

Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp
                325                 330                 335

Gly Lys Tyr Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro
            340                 345                 350

Val Gln Asn Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val
        355                 360                 365

Arg Asp Val Thr Ser Ile Val Pro Gln
    370                 375


<210> SEQ ID NO 29
<211> LENGTH: 136
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

Met Ala Ser Gly Ser Pro Pro Thr Gln Pro Ser Pro Ala Ser Asp Ser
1               5                   10                  15

Gly Ser Gly Tyr Val Pro Gly Ser Val Ser Ala Ala Phe Val Thr Cys
            20                  25                  30

Pro Asn Glu Lys Val Ala Lys Glu Ile Ala Arg Ala Val Val Glu Lys
        35                  40                  45

Arg Leu Ala Ala Cys Val Asn Leu Ile Pro Gln Ile Thr Ser Ile Tyr
    50                  55                  60

Glu Trp Lys Gly Lys Ile Glu Glu Asp Ser Glu Val Leu Met Met Ile
65                  70                  75                  80

Lys Thr Gln Ser Ser Leu Val Pro Ala Leu Thr Asp Phe Val Arg Ser
                85                  90                  95
```

-continued

```
Val His Pro Tyr Glu Val Ala Glu Val Ile Ala Leu Pro Val Glu Gln
            100                 105                 110

Gly Asn Phe Pro Tyr Leu Gln Trp Val Arg Gln Val Thr Glu Ser Val
        115                 120                 125

Ser Asp Ser Ile Thr Val Leu Pro
    130                 135


<210> SEQ ID NO 30
<211> LENGTH: 295
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: His-tagged fusion of DogTag to a variant
      (first described in Grunberg et al, 2013) of mCitrine
      fluorescent protein DgT-X3

<400> SEQUENCE: 30

Met Ser His His His His His His Gly Ser Thr Gly Leu Glu Val Leu
1               5                   10                  15

Phe Gln Gly Pro Thr Gly Ser Ser Asp Ile Pro Ala Thr Tyr Glu Phe
            20                  25                  30

Thr Asp Gly Lys His Tyr Ile Thr Asn Glu Pro Ile Pro Pro Lys Gly
        35                  40                  45

Gly Gly Gly Ser Gly Gly Gly Gly Ser Val Ser Lys Gly Glu Glu Leu
    50                  55                  60

Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn
65                  70                  75                  80

Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr
                85                  90                  95

Gly Lys Leu Glu Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val
            100                 105                 110

Pro Trp Pro Thr Leu Val Thr Thr Phe Gly Tyr Gly Leu Met Cys Phe
        115                 120                 125

Ala Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala
    130                 135                 140

Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp
145                 150                 155                 160

Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu
                165                 170                 175

Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn
            180                 185                 190

Ile Leu Gly His Lys Leu Glu Tyr Asp Tyr Asn Ser His Asn Val Tyr
        195                 200                 205

Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile
    210                 215                 220

Arg His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln
225                 230                 235                 240

Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His
                245                 250                 255

Tyr Leu Glu Tyr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg
            260                 265                 270

Asp His Met Val Leu Ala Glu Phe Val Thr Ala Glu Gly Ile Thr Leu
        275                 280                 285

Gly Met Asp Glu Leu Tyr Lys
```

-continued

```
    290                 295


<210> SEQ ID NO 31
<211> LENGTH: 164
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 31

Leu Arg Ala Asp Gly Asp Lys Pro Arg Ala His Leu Thr Val Val Arg
1               5                   10                  15

Gln Thr Pro Thr Gln His Phe Lys Asn Gln Phe Pro Ala Leu His Trp
            20                  25                  30

Glu His Glu Leu Gly Leu Ala Phe Thr Lys Asn Arg Met Asn Tyr Thr
        35                  40                  45

Asn Lys Phe Leu Leu Ile Pro Glu Ser Gly Asp Tyr Phe Ile Tyr Ser
    50                  55                  60

Gln Val Thr Phe Arg Gly Met Thr Ser Glu Cys Ser Glu Ile Arg Gln
65                  70                  75                  80

Ala Gly Arg Pro Asn Lys Pro Asp Ser Ile Thr Val Val Ile Thr Lys
                85                  90                  95

Val Thr Asp Ser Tyr Pro Glu Pro Thr Gln Leu Leu Met Gly Thr Lys
            100                 105                 110

Ser Val Cys Glu Val Gly Ser Asn Trp Phe Gln Pro Ile Tyr Leu Gly
        115                 120                 125

Ala Met Phe Ser Leu Gln Glu Gly Asp Lys Leu Met Val Asn Val Ser
    130                 135                 140

Asp Ile Ser Leu Val Asp Tyr Thr Lys Glu Asp Lys Thr Phe Phe Gly
145                 150                 155                 160

Ala Phe Leu Leu


<210> SEQ ID NO 32
<211> LENGTH: 389
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the construct
      SpyCatcher003-TL1A-DogCatcher

<400> SEQUENCE: 32

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1               5                   10                  15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
            20                  25                  30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
        35                  40                  45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
    50                  55                  60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65                  70                  75                  80

Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
                85                  90                  95

Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His
            100                 105                 110

Thr Gly Ser Gly Ser Leu Arg Ala Asp Gly Asp Lys Pro Arg Ala His
        115                 120                 125
```

-continued

```
Leu Thr Val Val Arg Gln Thr Pro Thr Gln His Phe Lys Asn Gln Phe
    130                 135                 140

Pro Ala Leu His Trp Glu His Glu Leu Gly Leu Ala Phe Thr Lys Asn
145                 150                 155                 160

Arg Met Asn Tyr Thr Asn Lys Phe Leu Leu Ile Pro Glu Ser Gly Asp
                165                 170                 175

Tyr Phe Ile Tyr Ser Gln Val Thr Phe Arg Gly Met Thr Ser Glu Cys
            180                 185                 190

Ser Glu Ile Arg Gln Ala Gly Arg Pro Asn Lys Pro Asp Ser Ile Thr
        195                 200                 205

Val Val Ile Thr Lys Val Thr Asp Ser Tyr Pro Glu Pro Thr Gln Leu
    210                 215                 220

Leu Met Gly Thr Lys Ser Val Cys Glu Val Gly Ser Asn Trp Phe Gln
225                 230                 235                 240

Pro Ile Tyr Leu Gly Ala Met Phe Ser Leu Gln Glu Gly Asp Lys Leu
                245                 250                 255

Met Val Asn Val Ser Asp Ile Ser Leu Val Asp Tyr Thr Lys Glu Asp
            260                 265                 270

Lys Thr Phe Phe Gly Ala Phe Leu Leu Gly Ser Gly Ser Lys Leu Gly
        275                 280                 285

Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys Lys Pro Leu Arg
    290                 295                 300

Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr Pro Asp Ile
305                 310                 315                 320

Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp Val Arg Thr Gly
                325                 330                 335

Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp Gly Lys Tyr Arg
            340                 345                 350

Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro Val Gln Asn Lys
        355                 360                 365

Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val Arg Asp Val Thr
    370                 375                 380

Ser Ile Val Pro Gln
385


<210> SEQ ID NO 33
<211> LENGTH: 287
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the construct SpyCatcher003-Col XV
      NC1-DogCatcher

<400> SEQUENCE: 33

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1               5                   10                  15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
            20                  25                  30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
        35                  40                  45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
    50                  55                  60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
```

-continued

```
65                  70                  75                  80

Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
                85                  90                  95

Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala Gly
            100                 105                 110

Ser Gly Ser Val Thr Ala Phe Ser Asn Met Asp Asp Met Leu Gln Lys
        115                 120                 125

Ala His Leu Val Ile Glu Gly Thr Phe Ile Tyr Leu Arg Asp Ser Thr
    130                 135                 140

Glu Phe Phe Ile Arg Val Arg Asp Gly Trp Lys Lys Leu Gln Leu Gly
145                 150                 155                 160

Glu Leu Ile Pro Ile Pro Ala Asp Ser Pro Pro Pro Pro Ala Leu Ser
                165                 170                 175

Ser Asn Pro Gly Ser Gly Ser Lys Leu Gly Glu Ile Glu Phe Ile Lys
            180                 185                 190

Val Asp Lys Thr Asp Lys Lys Pro Leu Arg Gly Ala Val Phe Ser Leu
        195                 200                 205

Gln Lys Gln His Pro Asp Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln
    210                 215                 220

Asn Gly Thr Tyr Gln Asp Val Arg Thr Gly Glu Asp Gly Lys Leu Thr
225                 230                 235                 240

Phe Thr Asn Leu Ser Asp Gly Lys Tyr Arg Leu Ile Glu Asn Ser Glu
                245                 250                 255

Pro Pro Gly Tyr Lys Pro Val Gln Asn Lys Pro Ile Val Ser Phe Arg
            260                 265                 270

Ile Val Asp Gly Glu Val Arg Asp Val Thr Ser Ile Val Pro Gln
        275                 280                 285


<210> SEQ ID NO 34
<211> LENGTH: 342
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the construct
      SpyCatcher003-MIF2m-DogCatcher

<400> SEQUENCE: 34

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1               5                   10                  15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
            20                  25                  30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
        35                  40                  45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
    50                  55                  60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65                  70                  75                  80

Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
                85                  90                  95

Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His
            100                 105                 110

Thr Gly Ser Gly Ser Pro Phe Leu Glu Leu Asp Thr Asn Leu Pro Ala
        115                 120                 125
```

-continued

```
Asn Arg Val Pro Ala Gly Leu Glu Lys Arg Leu Cys Ala Ala Ala Ala
    130                 135                 140

Ser Ile Leu Gly Lys Pro Ala Asp Arg Val Asn Val Thr Val Arg Pro
145                 150                 155                 160

Gly Leu Ala Met Ala Leu Ser Gly Ser Thr Glu Pro Cys Ala Gln Leu
                165                 170                 175

Ser Ile Ala Ser Ile Gly Val Val Gly Thr Ala Glu Asp Asn Arg Ser
            180                 185                 190

His Ser Ala His Phe Phe Glu Phe Leu Thr Lys Glu Leu Ala Leu Gly
        195                 200                 205

Gln Asp Arg Ile Leu Ile Arg Phe Ala Pro Leu Glu Ser Trp Gln Ile
    210                 215                 220

Gly Lys Ile Gly Thr Val Met Thr Phe Leu Gly Ser Gly Ser Lys Leu
225                 230                 235                 240

Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys Lys Pro Leu
                245                 250                 255

Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr Pro Asp
            260                 265                 270

Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp Val Arg Thr
        275                 280                 285

Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp Gly Lys Tyr
    290                 295                 300

Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro Val Gln Asn
305                 310                 315                 320

Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val Arg Asp Val
                325                 330                 335

Thr Ser Ile Val Pro Gln
            340


<210> SEQ ID NO 35
<211> LENGTH: 280
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the construct
      SpyCatcher003-IMX-DogCatcher

<400> SEQUENCE: 35

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1               5                   10                  15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
            20                  25                  30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
        35                  40                  45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
    50                  55                  60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65                  70                  75                  80

Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
                85                  90                  95

Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His
            100                 105                 110

Thr Gly Ser Gly Ser Lys Lys Gln Gly Asp Ala Asp Val Cys Gly Glu
        115                 120                 125
```

-continued

```
Val Ala Tyr Ile Gln Ser Val Val Ser Asp Cys His Val Pro Thr Ala
    130                 135                 140

Glu Leu Arg Thr Leu Leu Glu Ile Arg Lys Leu Phe Leu Glu Ile Gln
145                 150                 155                 160

Lys Leu Lys Val Glu Leu Gln Gly Leu Ser Lys Glu Gly Ser Gly Ser
                165                 170                 175

Lys Leu Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys Lys
            180                 185                 190

Pro Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr
        195                 200                 205

Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp Val
    210                 215                 220

Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp Gly
225                 230                 235                 240

Lys Tyr Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro Val
                245                 250                 255

Gln Asn Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val Arg
            260                 265                 270

Asp Val Thr Ser Ile Val Pro Gln
        275                 280


<210> SEQ ID NO 36
<211> LENGTH: 133
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 36

Gln Pro Arg Pro Ala Phe Ser Ala Ile Arg Arg Asn Pro Pro Met Gly
1               5                   10                  15

Gly Asn Val Val Ile Phe Asp Thr Val Ile Thr Asn Gln Glu Glu Pro
            20                  25                  30

Tyr Gln Asn His Ser Gly Arg Phe Val Cys Thr Val Pro Gly Tyr Tyr
        35                  40                  45

Tyr Phe Thr Phe Gln Val Leu Ser Gln Trp Glu Ile Cys Leu Ser Ile
    50                  55                  60

Val Ser Ser Ser Arg Gly Gln Val Arg Arg Ser Leu Gly Phe Cys Asp
65                  70                  75                  80

Thr Thr Asn Lys Gly Leu Phe Gln Val Val Ser Gly Gly Met Val Leu
                85                  90                  95

Gln Leu Gln Gln Gly Asp Gln Val Trp Val Glu Lys Asp Pro Lys Lys
            100                 105                 110

Gly His Ile Tyr Gln Gly Ser Glu Ala Asp Ser Val Phe Ser Gly Phe
        115                 120                 125

Leu Ile Phe Pro Ser
    130


<210> SEQ ID NO 37
<211> LENGTH: 132
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 37

Thr Gln Lys Ile Ala Phe Ser Ala Thr Arg Thr Ile Asn Val Pro Leu
1               5                   10                  15
```

-continued

```
Arg Arg Asp Gln Thr Ile Arg Phe Asp His Val Ile Thr Asn Met Asn
            20                  25                  30

Asn Asn Tyr Glu Pro Arg Ser Gly Lys Phe Thr Cys Lys Val Pro Gly
        35                  40                  45

Leu Tyr Tyr Phe Thr Tyr His Ala Ser Ser Arg Gly Asn Leu Cys Val
    50                  55                  60

Asn Leu Met Arg Gly Arg Glu Arg Ala Gln Lys Val Val Thr Phe Cys
65                  70                  75                  80

Asp Tyr Ala Tyr Asn Thr Phe Gln Val Thr Thr Gly Gly Met Val Leu
                85                  90                  95

Lys Leu Glu Gln Gly Glu Asn Val Phe Leu Gln Ala Thr Asp Lys Asn
            100                 105                 110

Ser Leu Leu Gly Met Glu Gly Ala Asn Ser Ile Phe Ser Gly Phe Leu
        115                 120                 125

Leu Phe Pro Asp
    130


<210> SEQ ID NO 38
<211> LENGTH: 129
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 38

Lys Phe Gln Ser Val Phe Thr Val Thr Arg Gln Thr His Gln Pro Pro
1               5                   10                  15

Ala Pro Asn Ser Leu Ile Arg Phe Asn Ala Val Leu Thr Asn Pro Gln
            20                  25                  30

Gly Asp Tyr Asp Thr Ser Thr Gly Lys Phe Thr Cys Lys Val Pro Gly
        35                  40                  45

Leu Tyr Tyr Phe Val Tyr His Ala Ser His Thr Ala Asn Leu Cys Val
    50                  55                  60

Leu Leu Tyr Arg Ser Gly Val Lys Val Val Thr Phe Cys Gly His Thr
65                  70                  75                  80

Ser Lys Thr Asn Gln Val Asn Ser Gly Gly Val Leu Leu Arg Leu Gln
                85                  90                  95

Val Gly Glu Glu Val Trp Leu Ala Val Asn Asp Tyr Tyr Asp Met Val
            100                 105                 110

Gly Ile Gln Gly Ser Asp Ser Val Phe Ser Gly Phe Leu Leu Phe Pro
        115                 120                 125

Asp


<210> SEQ ID NO 39
<211> LENGTH: 103
<212> TYPE: PRT
<213> ORGANISM: Thermus thermophilus

<400> SEQUENCE: 39

Met Glu Glu Val Val Leu Ile Thr Val Pro Ser Glu Glu Val Ala Arg
1               5                   10                  15

Thr Ile Ala Lys Ala Leu Val Glu Glu Arg Leu Ala Ala Cys Val Asn
            20                  25                  30

Ile Val Pro Gly Leu Thr Ser Ile Tyr Arg Trp Gln Gly Glu Val Val
        35                  40                  45
```

-continued

```
Glu Asp Gln Glu Leu Leu Leu Leu Val Lys Thr Thr Thr His Ala Phe
    50                  55                  60

Pro Lys Leu Lys Glu Arg Val Lys Ala Leu His Pro Tyr Thr Val Pro
65                  70                  75                  80

Glu Ile Val Ala Leu Pro Ile Ala Glu Gly Asn Arg Glu Tyr Leu Asp
                85                  90                  95

Trp Leu Arg Glu Asn Thr Gly
            100


<210> SEQ ID NO 40
<211> LENGTH: 107
<212> TYPE: PRT
<213> ORGANISM: Oryza sativa

<400> SEQUENCE: 40

Ser Thr Thr Val Pro Ser Ile Val Val Tyr Val Thr Val Pro Asn Lys
1               5                   10                  15

Glu Ala Gly Lys Arg Leu Ala Gly Ser Ile Ile Ser Glu Lys Leu Ala
            20                  25                  30

Ala Cys Val Asn Ile Val Pro Gly Ile Glu Ser Val Tyr Trp Trp Glu
        35                  40                  45

Gly Lys Val Gln Thr Asp Ala Glu Glu Leu Leu Ile Ile Lys Thr Arg
    50                  55                  60

Glu Ser Leu Leu Asp Ala Leu Thr Glu His Val Lys Ala Asn His Glu
65                  70                  75                  80

Tyr Asp Val Pro Glu Val Ile Ala Leu Pro Ile Lys Gly Gly Asn Leu
                85                  90                  95

Lys Tyr Leu Glu Trp Leu Lys Asn Ser Thr Arg
            100                 105


<210> SEQ ID NO 41
<211> LENGTH: 106
<212> TYPE: PRT
<213> ORGANISM: Shewanella sp. SIB1

<400> SEQUENCE: 41

Lys Pro Glu Gln Leu Leu Ile Phe Thr Thr Cys Pro Asp Ala Asp Ile
1               5                   10                  15

Ala Cys Arg Ile Ala Thr Ala Leu Val Glu Ala Lys Leu Ala Ala Cys
            20                  25                  30

Val Gln Ile Gly Gln Ala Val Glu Ser Ile Tyr Gln Trp Asp Asn Asn
        35                  40                  45

Ile Cys Gln Ser His Glu Val Pro Met Gln Ile Lys Cys Met Thr Thr
    50                  55                  60

Asp Tyr Pro Ala Ile Glu Gln Leu Val Ile Thr Met His Pro Tyr Glu
65                  70                  75                  80

Val Pro Glu Phe Ile Ala Thr Pro Ile Ile Gly Gly Phe Gly Pro Tyr
                85                  90                  95

Leu Gln Trp Ile Lys Asp Asn Ser Pro Ser
            100                 105


<210> SEQ ID NO 42
<211> LENGTH: 152
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
```

-continued

<400> SEQUENCE: 42

Arg Thr Pro Ser Asp Lys Pro Val Ala His Val Val Ala Asn Pro Gln
1 5 10 15

Ala Glu Gly Gln Leu Gln Trp Leu Asn Arg Arg Ala Asn Ala Leu Leu
20 25 30

Ala Asn Gly Val Glu Leu Arg Asp Asn Gln Leu Val Val Pro Ser Glu
35 40 45

Gly Leu Tyr Leu Ile Tyr Ser Gln Val Leu Phe Lys Gly Gln Gly Cys
50 55 60

Pro Ser Thr His Val Leu Leu Thr His Thr Ile Ser Arg Ile Ala Val
65 70 75 80

Ser Tyr Gln Thr Lys Val Asn Leu Leu Ser Ala Ile Lys Ser Pro Cys
85 90 95

Gln Arg Glu Thr Pro Glu Gly Ala Glu Ala Lys Pro Trp Tyr Glu Pro
100 105 110

Ile Tyr Leu Gly Gly Val Phe Gln Leu Glu Lys Gly Asp Arg Leu Ser
115 120 125

Ala Glu Ile Asn Arg Pro Asp Tyr Leu Leu Phe Ala Glu Ser Gly Gln
130 135 140

Val Tyr Phe Gly Ile Ile Ala Leu
145 150

<210> SEQ ID NO 43
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the antiparallel coiled coil hexamer

<400> SEQUENCE: 43

Glu Leu Ala Gln Ala Phe Lys Glu Ile Ala Lys Ala Phe Lys Glu Ile
1 5 10 15

Ala Lys Ala Phe Glu Phe Ile Ala Gln Ala Ile Glu
20 25

<210> SEQ ID NO 44
<211> LENGTH: 61
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 44

Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln
1 5 10 15

His Leu Leu Gln Leu Thr Val Trp Ala Ile Lys Gln Leu Gln Ala Arg
20 25 30

Ser Gly Gly Arg Gly Gly Trp Met Glu Trp Asp Arg Glu Ile Asn Asn
35 40 45

Tyr Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln
50 55 60

<210> SEQ ID NO 45
<211> LENGTH: 96
<212> TYPE: PRT
<213> ORGANISM: Aquifex aeolicus VF5

<400> SEQUENCE: 45

-continued

```
Val Asp Pro Ala Lys Glu Ala Ile Met Lys Pro Gln Leu Thr Met Leu
1               5                   10                  15

Lys Gly Leu Ser Asp Ala Glu Leu Lys Ala Leu Ala Asp Phe Ile Leu
            20                  25                  30

Arg Ile Ala Lys Gln Ala Gln Glu Lys Gln Gln Gln Asp Val Ala Lys
        35                  40                  45

Ala Ile Phe Gln Gln Lys Gly Cys Gly Ser Cys His Gln Ala Asn Val
    50                  55                  60

Asp Thr Val Gly Pro Ser Leu Ala Lys Ile Ala Gln Ala Tyr Ala Gly
65                  70                  75                  80

Lys Glu Asp Gln Leu Ile Lys Phe Leu Lys Gly Glu Ala Pro Ala Ile
                85                  90                  95


<210> SEQ ID NO 46
<211> LENGTH: 75
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 46

Tyr Gly Asn Met Thr Glu Asp His Val Met His Leu Leu Gln Asn Ala
1               5                   10                  15

Asp Pro Leu Lys Val Tyr Pro Pro Leu Lys Gly Ser Phe Pro Glu Asn
            20                  25                  30

Leu Arg His Leu Lys Asn Thr Met Glu Thr Ile Asp Trp Lys Val Phe
        35                  40                  45

Glu Ser Trp Met His His Trp Leu Leu Phe Glu Met Ser Arg His Ser
    50                  55                  60

Leu Glu Gln Lys Pro Thr Asp Ala Pro Pro Lys
65                  70                  75


<210> SEQ ID NO 47

<400> SEQUENCE: 47

000


<210> SEQ ID NO 48
<211> LENGTH: 220
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 48

Ser Arg Pro Arg Asp Cys Leu Asp Val Leu Leu Ser Gly Gln Gln Asp
1               5                   10                  15

Asp Gly Val Tyr Ser Val Phe Pro Thr His Tyr Pro Ala Gly Phe Gln
            20                  25                  30

Val Tyr Cys Asp Met Arg Thr Asp Gly Gly Gly Trp Thr Val Phe Gln
        35                  40                  45

Arg Arg Glu Asp Gly Ser Val Asn Phe Phe Arg Gly Trp Asp Ala Tyr
    50                  55                  60

Arg Asp Gly Phe Gly Arg Leu Thr Gly Glu His Trp Leu Gly Leu Lys
65                  70                  75                  80

Arg Ile His Ala Leu Thr Thr Gln Ala Ala Tyr Glu Leu His Val Asp
                85                  90                  95

Leu Glu Asp Phe Glu Asn Gly Thr Ala Tyr Ala Arg Tyr Gly Ser Phe
            100                 105                 110
```

-continued

```
Gly Val Gly Leu Phe Ser Val Asp Pro Glu Glu Asp Gly Tyr Pro Leu
        115                 120                 125

Thr Val Ala Asp Tyr Ser Gly Thr Ala Gly Asp Ser Leu Leu Lys His
    130                 135                 140

Ser Gly Met Arg Phe Thr Thr Lys Asp Arg Asp Ser Asp His Ser Glu
145                 150                 155                 160

Asn Asn Cys Ala Ala Phe Tyr Arg Gly Ala Trp Trp Tyr Arg Asn Cys
                165                 170                 175

His Thr Ser Asn Leu Asn Gly Gln Tyr Leu Arg Gly Ala His Ala Ser
            180                 185                 190

Tyr Ala Asp Gly Val Glu Trp Ser Ser Trp Thr Gly Trp Gln Tyr Ser
        195                 200                 205

Leu Lys Phe Ser Glu Met Lys Ile Arg Pro Val Arg
    210                 215                 220


<210> SEQ ID NO 49
<211> LENGTH: 228
<212> TYPE: PRT
<213> ORGANISM: Bos taurus

<400> SEQUENCE: 49

Val Asp His Gly Phe Leu Val Thr Arg His Ser Gln Thr Ile Asp Asp
1               5                   10                  15

Pro Gln Cys Pro Ser Gly Thr Lys Ile Leu Tyr His Gly Tyr Ser Leu
            20                  25                  30

Leu Tyr Val Gln Gly Asn Glu Arg Ala His Gly Gln Asp Leu Gly Thr
        35                  40                  45

Ala Gly Ser Cys Leu Arg Lys Phe Ser Thr Met Pro Phe Leu Phe Cys
    50                  55                  60

Asn Ile Asn Asn Val Cys Asn Phe Ala Ser Arg Asn Asp Tyr Ser Tyr
65                  70                  75                  80

Trp Leu Ser Thr Pro Glu Pro Met Pro Met Ser Met Ala Pro Ile Thr
                85                  90                  95

Gly Glu Asn Ile Arg Pro Phe Ile Ser Arg Cys Ala Val Cys Glu Ala
            100                 105                 110

Pro Ala Met Val Met Ala Val His Ser Gln Thr Ile Gln Ile Pro Pro
        115                 120                 125

Cys Pro Ser Gly Trp Ser Ser Leu Trp Ile Gly Tyr Ser Phe Val Met
    130                 135                 140

His Thr Ser Ala Gly Ala Glu Gly Ser Gly Gln Ala Leu Ala Ser Pro
145                 150                 155                 160

Gly Ser Cys Leu Glu Glu Phe Arg Ser Ala Pro Phe Ile Glu Cys His
                165                 170                 175

Gly Arg Gly Thr Cys Asn Tyr Tyr Ala Asn Ala Tyr Ser Phe Trp Leu
            180                 185                 190

Ala Thr Ile Glu Arg Ser Glu Met Phe Lys Lys Pro Thr Pro Ser Thr
        195                 200                 205

Leu Lys Ala Gly Glu Leu Arg Thr His Val Ser Arg Cys Gln Val Cys
    210                 215                 220

Met Arg Arg Thr
225
```

-continued

<210> SEQ ID NO 50
<211> LENGTH: 52
<212> TYPE: PRT
<213> ORGANISM: Bacillus subtilis

<400> SEQUENCE: 50

Phe Met Lys Ser Thr Gly Ile Val Arg Lys Val Asp Glu Leu Gly Arg
1 5 10 15

Val Val Ile Pro Ile Glu Leu Arg Arg Thr Leu Gly Ile Ala Glu Lys
20 25 30

Asp Ala Leu Glu Ile Tyr Val Asp Asp Glu Lys Ile Ile Leu Lys Lys
35 40 45

Tyr Lys Pro Asn
50

<210> SEQ ID NO 51
<211> LENGTH: 95
<212> TYPE: PRT
<213> ORGANISM: Escherichia virus Lambda

<400> SEQUENCE: 51

Ser Asp Pro Ala His Thr Ala Thr Ala Pro Gly Gly Leu Ser Ala Lys
1 5 10 15

Ala Pro Ala Met Thr Pro Leu Met Leu Asp Thr Ser Ser Arg Lys Leu
20 25 30

Val Ala Trp Asp Gly Thr Thr Asp Gly Ala Ala Val Gly Ile Leu Ala
35 40 45

Val Ala Ala Asp Gln Thr Ser Thr Thr Leu Thr Phe Tyr Lys Ser Gly
50 55 60

Thr Phe Arg Tyr Glu Asp Val Leu Trp Pro Glu Ala Ala Ser Asp Glu
65 70 75 80

Thr Lys Lys Arg Thr Ala Phe Ala Gly Thr Ala Ile Ser Ile Val
85 90 95

<210> SEQ ID NO 52
<211> LENGTH: 133
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 52

Thr Arg Asp Phe Asn Gly Thr Trp Glu Met Glu Ser Asn Glu Asn Phe
1 5 10 15

Glu Gly Tyr Met Lys Ala Leu Asp Ile Asp Phe Ala Thr Arg Lys Ile
20 25 30

Ala Val Arg Leu Thr Phe Thr Asp Val Ile Asp Gln Asp Gly Asp Asn
35 40 45

Phe Lys Thr Lys Ala Thr Ser Thr Phe Leu Asn Tyr Asp Glu Asp Phe
50 55 60

Thr Val Gly Val Glu Phe Asp Glu Tyr Thr Lys Ser Leu Asp Asn Arg
65 70 75 80

His Val Lys Ala Leu Val Thr Trp Glu Gly Asp Val Leu Val Cys Val
85 90 95

Gln Lys Gly Glu Lys Glu Asn Arg Gly Trp Lys Lys Trp Ile Glu Gly
100 105 110

Asp Lys Leu Tyr Leu Glu Leu Thr Cys Gly Asp Gln Val Cys Arg Gln
115 120 125

-continued

```
Val Phe Lys Lys Lys
    130


<210> SEQ ID NO 53
<211> LENGTH: 162
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 53

Leu Pro Thr Tyr Arg Tyr Pro Leu Glu Leu Asp Thr Ala Asn Asn Arg
1               5                   10                  15

Val Gln Val Ala Asp Arg Phe Gly Met Arg Thr Gly Thr Trp Thr Gly
            20                  25                  30

Gln Leu Gln Tyr Gln His Pro Gln Leu Ser Trp Arg Ala Asn Val Thr
        35                  40                  45

Leu Asn Leu Met Lys Val Asp Asp Trp Leu Val Leu Ser Phe Ser Gln
    50                  55                  60

Met Thr Thr Asn Ser Ile Met Ala Asp Gly Lys Phe Val Ile Asn Phe
65                  70                  75                  80

Val Ser Gly Leu Ser Ser Gly Trp Gln Thr Gly Asp Thr Glu Pro Ser
                85                  90                  95

Ser Thr Ile Asp Pro Leu Ser Thr Thr Phe Ala Ala Val Gln Phe Leu
            100                 105                 110

Asn Asn Gly Gln Arg Ile Asp Ala Phe Arg Ile Met Gly Val Ser Glu
        115                 120                 125

Trp Thr Asp Gly Glu Leu Glu Ile Lys Asn Tyr Gly Gly Thr Tyr Thr
    130                 135                 140

Gly His Thr Gln Val Tyr Trp Ala Pro Trp Thr Ile Met Tyr Pro Cys
145                 150                 155                 160

Asn Val


<210> SEQ ID NO 54
<211> LENGTH: 361
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the construct
      SpyCatcher003-HsCutA1-DogCatcher

<400> SEQUENCE: 54

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1               5                   10                  15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
            20                  25                  30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
        35                  40                  45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
    50                  55                  60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65                  70                  75                  80

Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
                85                  90                  95

Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His
            100                 105                 110
```

-continued

```
Thr Gly Ser Gly Ser Met Ala Ser Gly Ser Pro Pro Thr Gln Pro Ser
        115                 120                 125

Pro Ala Ser Asp Ser Gly Ser Gly Tyr Val Pro Gly Ser Val Ser Ala
    130                 135                 140

Ala Phe Val Thr Cys Pro Asn Glu Lys Val Ala Lys Glu Ile Ala Arg
145                 150                 155                 160

Ala Val Val Glu Lys Arg Leu Ala Ala Cys Val Asn Leu Ile Pro Gln
                165                 170                 175

Ile Thr Ser Ile Tyr Glu Trp Lys Gly Lys Ile Glu Glu Asp Ser Glu
            180                 185                 190

Val Leu Met Met Ile Lys Thr Gln Ser Ser Leu Val Pro Ala Leu Thr
        195                 200                 205

Asp Phe Val Arg Ser Val His Pro Tyr Glu Val Ala Glu Val Ile Ala
    210                 215                 220

Leu Pro Val Glu Gln Gly Asn Phe Pro Tyr Leu Gln Trp Val Arg Gln
225                 230                 235                 240

Val Thr Glu Ser Val Ser Asp Ser Ile Thr Val Leu Pro Gly Ser Gly
                245                 250                 255

Ser Lys Leu Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys
            260                 265                 270

Lys Pro Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp
        275                 280                 285

Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp
    290                 295                 300

Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp
305                 310                 315                 320

Gly Lys Tyr Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro
                325                 330                 335

Val Gln Asn Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val
            340                 345                 350

Arg Asp Val Thr Ser Ile Val Pro Gln
        355                 360


<210> SEQ ID NO 55
<211> LENGTH: 327
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the construct
      SpyCatcher003-PhCutA1-DogCatcher

<400> SEQUENCE: 55

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1               5                   10                  15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
            20                  25                  30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
        35                  40                  45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
    50                  55                  60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65                  70                  75                  80

Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
                85                  90                  95
```

-continued

```
Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His
            100                 105                 110

Thr Gly Ser Gly Ser Met Ile Ile Val Tyr Thr Thr Phe Pro Asp Trp
        115                 120                 125

Glu Ser Ala Glu Lys Val Val Lys Thr Leu Leu Lys Glu Arg Leu Ile
    130                 135                 140

Ala Cys Ala Asn Leu Arg Glu His Arg Ala Phe Tyr Trp Trp Glu Gly
145                 150                 155                 160

Lys Ile Glu Glu Asp Lys Glu Val Gly Ala Ile Leu Lys Thr Arg Glu
                165                 170                 175

Asp Leu Trp Glu Glu Leu Lys Glu Arg Ile Lys Glu Leu His Pro Tyr
            180                 185                 190

Asp Val Pro Ala Ile Ile Arg Ile Asp Val Asp Asp Val Asn Glu Asp
        195                 200                 205

Tyr Leu Lys Trp Leu Ile Glu Glu Thr Lys Lys Gly Ser Gly Ser Lys
    210                 215                 220

Leu Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys Lys Pro
225                 230                 235                 240

Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr Pro
                245                 250                 255

Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp Val Arg
            260                 265                 270

Thr Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp Gly Lys
        275                 280                 285

Tyr Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro Val Gln
    290                 295                 300

Asn Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val Arg Asp
305                 310                 315                 320

Val Thr Ser Ile Val Pro Gln
                325


<210> SEQ ID NO 56
<211> LENGTH: 359
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the construct SpyCatcher003-Col X
      NC1-DogCatcher

<400> SEQUENCE: 56

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1               5                   10                  15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
            20                  25                  30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
        35                  40                  45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
    50                  55                  60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65                  70                  75                  80

Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
                85                  90                  95

Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His
```

-continued

```
            100                 105                 110

Thr Gly Ser Gly Ser Thr Gly Met Pro Val Ser Ala Phe Thr Val Ile
        115                 120                 125

Leu Ser Lys Ala Tyr Pro Ala Ile Gly Thr Pro Ile Pro Phe Asp Lys
    130                 135                 140

Ile Leu Tyr Asn Arg Gln Gln His Tyr Asp Pro Arg Thr Gly Ile Phe
145                 150                 155                 160

Thr Cys Gln Ile Pro Gly Ile Tyr Tyr Phe Ser Tyr His Val His Val
                165                 170                 175

Lys Gly Thr His Val Trp Val Gly Leu Tyr Lys Asn Gly Thr Pro Val
            180                 185                 190

Met Tyr Thr Tyr Asp Glu Tyr Thr Lys Gly Tyr Leu Asp Gln Ala Ser
        195                 200                 205

Gly Ser Ala Ile Ile Asp Leu Thr Glu Asn Asp Gln Val Trp Leu Gln
    210                 215                 220

Leu Pro Asn Ala Glu Ser Asn Gly Leu Tyr Ser Ser Glu Tyr Val His
225                 230                 235                 240

Ser Ser Phe Ser Gly Phe Leu Val Ala Pro Met Gly Ser Gly Ser Lys
                245                 250                 255

Leu Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys Lys Pro
            260                 265                 270

Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp Tyr Pro
        275                 280                 285

Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp Val Arg
    290                 295                 300

Thr Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp Gly Lys
305                 310                 315                 320

Tyr Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro Val Gln
                325                 330                 335

Asn Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val Arg Asp
            340                 345                 350

Val Thr Ser Ile Val Pro Gln
        355


<210> SEQ ID NO 57
<211> LENGTH: 377
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: monomer of the construct
      SpyCatcher003-TNF-DogCatcher

<400> SEQUENCE: 57

Val Thr Thr Leu Ser Gly Leu Ser Gly Glu Gln Gly Pro Ser Gly Asp
1               5                   10                  15

Met Thr Thr Glu Glu Asp Ser Ala Thr His Ile Lys Phe Ser Lys Arg
            20                  25                  30

Asp Glu Asp Gly Arg Glu Leu Ala Gly Ala Thr Met Glu Leu Arg Asp
        35                  40                  45

Ser Ser Gly Lys Thr Ile Ser Thr Trp Ile Ser Asp Gly His Val Lys
    50                  55                  60

Asp Phe Tyr Leu Tyr Pro Gly Lys Tyr Thr Phe Val Glu Thr Ala Ala
65                  70                  75                  80
```

-continued

```
Pro Asp Gly Tyr Glu Val Ala Thr Pro Ile Glu Phe Thr Val Asn Glu
                85                  90                  95

Asp Gly Gln Val Thr Val Asp Gly Glu Ala Thr Glu Gly Asp Ala His
            100                 105                 110

Thr Gly Ser Gly Ser Arg Thr Pro Ser Asp Lys Pro Val Ala His Val
        115                 120                 125

Val Ala Asn Pro Gln Ala Glu Gly Gln Leu Gln Trp Leu Asn Arg Arg
    130                 135                 140

Ala Asn Ala Leu Leu Ala Asn Gly Val Glu Leu Arg Asp Asn Gln Leu
145                 150                 155                 160

Val Val Pro Ser Glu Gly Leu Tyr Leu Ile Tyr Ser Gln Val Leu Phe
                165                 170                 175

Lys Gly Gln Gly Cys Pro Ser Thr His Val Leu Leu Thr His Thr Ile
            180                 185                 190

Ser Arg Ile Ala Val Ser Tyr Gln Thr Lys Val Asn Leu Leu Ser Ala
        195                 200                 205

Ile Lys Ser Pro Cys Gln Arg Glu Thr Pro Glu Gly Ala Glu Ala Lys
    210                 215                 220

Pro Trp Tyr Glu Pro Ile Tyr Leu Gly Gly Val Phe Gln Leu Glu Lys
225                 230                 235                 240

Gly Asp Arg Leu Ser Ala Glu Ile Asn Arg Pro Asp Tyr Leu Leu Phe
                245                 250                 255

Ala Glu Ser Gly Gln Val Tyr Phe Gly Ile Ile Ala Leu Gly Ser Gly
            260                 265                 270

Ser Lys Leu Gly Glu Ile Glu Phe Ile Lys Val Asp Lys Thr Asp Lys
        275                 280                 285

Lys Pro Leu Arg Gly Ala Val Phe Ser Leu Gln Lys Gln His Pro Asp
    290                 295                 300

Tyr Pro Asp Ile Tyr Gly Ala Ile Asp Gln Asn Gly Thr Tyr Gln Asp
305                 310                 315                 320

Val Arg Thr Gly Glu Asp Gly Lys Leu Thr Phe Thr Asn Leu Ser Asp
                325                 330                 335

Gly Lys Tyr Arg Leu Ile Glu Asn Ser Glu Pro Pro Gly Tyr Lys Pro
            340                 345                 350

Val Gln Asn Lys Pro Ile Val Ser Phe Arg Ile Val Asp Gly Glu Val
        355                 360                 365

Arg Asp Val Thr Ser Ile Val Pro Gln
    370                 375


<210> SEQ ID NO 58
<211> LENGTH: 141
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 58

Gln Ile Ala Ala His Val Ile Ser Glu Ala Ser Ser Lys Thr Thr Ser
1               5                   10                  15

Val Leu Gln Trp Ala Glu Lys Gly Tyr Tyr Thr Met Ser Asn Asn Leu
            20                  25                  30

Val Thr Leu Glu Asn Gly Lys Gln Leu Thr Val Lys Arg Gln Gly Leu
        35                  40                  45

Tyr Tyr Ile Tyr Ala Gln Val Thr Phe Cys Ser Asn Arg Glu Ala Ser
    50                  55                  60
```

-continued

```
Ser Gln Ala Pro Phe Ile Ala Ser Leu Cys Leu Lys Ser Pro Gly Arg
65                  70                  75                  80

Phe Glu Arg Ile Leu Leu Arg Ala Ala Asn Thr His Ser Ser Ala Lys
                85                  90                  95

Pro Cys Gly Gln Gln Ser Ile His Leu Gly Gly Val Phe Glu Leu Gln
            100                 105                 110

Pro Gly Ala Ser Val Phe Val Asn Val Thr Asp Pro Ser Gln Val Ser
        115                 120                 125

His Gly Thr Gly Phe Thr Ser Phe Gly Leu Leu Lys Leu
    130                 135                 140


<210> SEQ ID NO 59
<211> LENGTH: 146
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 59

Met Lys Asp Glu Val Ala Leu Leu Ala Ala Val Thr Leu Leu Gly Val
1               5                   10                  15

Leu Leu Gln Ala Tyr Phe Ser Leu Gln Val Ile Ser Ala Arg Arg Ala
            20                  25                  30

Phe Arg Val Ser Pro Pro Leu Thr Thr Gly Pro Pro Glu Phe Glu Arg
        35                  40                  45

Val Tyr Arg Ala Gln Val Asn Cys Ser Glu Tyr Phe Pro Leu Phe Leu
    50                  55                  60

Ala Thr Leu Trp Val Ala Gly Ile Phe Phe His Glu Gly Ala Ala Ala
65                  70                  75                  80

Leu Cys Gly Leu Val Tyr Leu Phe Ala Arg Leu Arg Tyr Phe Gln Gly
                85                  90                  95

Tyr Ala Arg Ser Ala Gln Leu Arg Leu Ala Pro Leu Tyr Ala Ser Ala
            100                 105                 110

Arg Ala Leu Trp Leu Leu Val Ala Leu Ala Ala Leu Gly Leu Leu Ala
        115                 120                 125

His Phe Leu Pro Ala Ala Leu Arg Ala Ala Leu Leu Gly Arg Leu Arg
    130                 135                 140

Thr Leu
145


<210> SEQ ID NO 60
<211> LENGTH: 108
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 60

Ser Gly Tyr Val Pro Gly Ser Val Ser Ala Ala Phe Val Thr Cys Pro
1               5                   10                  15

Asn Glu Lys Val Ala Lys Glu Ile Ala Arg Ala Val Val Glu Lys Arg
            20                  25                  30

Leu Ala Ala Cys Val Asn Leu Ile Pro Gln Ile Thr Ser Ile Tyr Glu
        35                  40                  45

Trp Lys Gly Lys Ile Glu Glu Asp Ser Glu Val Leu Met Met Ile Lys
    50                  55                  60

Thr Gln Ser Ser Leu Val Pro Ala Leu Thr Asp Phe Val Arg Ser Val
65                  70                  75                  80
```

-continued

His Pro Tyr Glu Val Ala Glu Val Ile Ala Leu Pro Val Glu Gln Gly
85 90 95

Asn Phe Pro Tyr Leu Gln Trp Val Arg Gln Val Thr
100 105

<210> SEQ ID NO 61
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 61

Gly Ser Gly Ser
1

<210> SEQ ID NO 62
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 62

Gly Gly Gly Gly Ser
1 5

<210> SEQ ID NO 63
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 63

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10

<210> SEQ ID NO 64
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 64

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10 15

<210> SEQ ID NO 65
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 65

Gly Gly Gly Ser
1

<210> SEQ ID NO 66
<211> LENGTH: 4
<212> TYPE: PRT

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 66

Pro Gly Gly Ser
1

<210> SEQ ID NO 67
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 67

Pro Gly Gly Gly
1

<210> SEQ ID NO 68
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 68

Arg Pro Pro Pro Pro Pro
1 5

<210> SEQ ID NO 69
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 69

Arg Pro Pro Pro Pro
1 5

<210> SEQ ID NO 70
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 70

Arg Pro Pro Gly
1

<210> SEQ ID NO 71
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 71

Pro Pro Pro Pro
1

<210> SEQ ID NO 72

-continued

<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 72

Pro Pro Pro Pro Pro Pro Pro Pro Pro
1 5

<210> SEQ ID NO 73
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 73

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
1 5 10

<210> SEQ ID NO 74
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 74

Ser Gly Ser Gly
1

<210> SEQ ID NO 75
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 75

Ser Gly Ser Gly Ser Gly
1 5

<210> SEQ ID NO 76
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 76

Ser Gly Ser Gly Ser Gly Ser Gly
1 5

<210> SEQ ID NO 77
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 77

Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly
1 5 10

-continued

```
<210> SEQ ID NO 78
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Linker

<400> SEQUENCE: 78

Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly
1               5                   10                  15
```

1. A multivalent protein scaffold comprising:

an oligomeric core comprising a plurality of subunit monomers;

at least one first binding site orthogonal to at least one second binding site;

wherein said first binding site(s) and said second binding site(s) are positioned on the same face of the scaffold; and

wherein said first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target; and said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target.

2. A protein scaffold according to any one of the preceding claims, wherein each monomer in the oligomeric core comprises at least one first binding site and at least one second binding site, and wherein the at least one first binding site is orthogonal to the at least one second binding site.

3. A protein scaffold according to claim **1** or claim **2**, wherein each monomer comprises a first binding site attached at a first terminus of said monomer and a second binding site at a second terminus of said monomer.

4. A protein scaffold according to any of claims **1-3**, wherein the first terminus and the second terminus of each monomer are positioned on the same face of said monomer and/or on the same face of an oligomer of said monomer.

5. A protein scaffold according to any one of the preceding claims, wherein said first binding site comprises a first protein domain capable of forming a covalent bond to a first polypeptide target; and said second binding site comprises a second protein domain capable of forming a covalent bond to a second polypeptide target;

wherein said first protein domain is capable of forming an isopeptide bond with said first polypeptide target and said second protein domain is capable of forming an isopeptide bond with said second binding target.

6. A protein scaffold according to claim **5**, wherein said first binding site and said second binding site each comprise a different split ligand-binding protein domain;

wherein preferably one of said first binding site and said second binding site comprises a split *Streptococcus pyogenes* fibronectin-binding protein domain and the other of said first binding site and said second binding site comprises a split *Streptococcus pneumoniae* adhesin domain.

7. A protein scaffold according to claim **6**, wherein said first and said second binding site each independently have at least 50% amino acid identity to any one of SEQ ID NOs: 4-9, 11-13, 23 or 15-18.

8. A protein scaffold according to any preceding claim, wherein the oligomeric core comprises at least three subunit monomers,

wherein preferably the oligomeric core comprises from 3 to 6 subunit monomers.

9. A protein scaffold according to any preceding claim, wherein said subunit monomers are non-covalently attached together.

10. A protein scaffold according to claims **1** to **8**, wherein said subunit monomers are covalently attached together;

wherein preferably said subunit monomers form a recombinant fusion protein.

11. A protein scaffold according to any one of the preceding claims, wherein said oligomeric core is a homooligomeric core.

12. A protein scaffold according to any one of claims **1** to **6**, wherein said oligomeric core is a hetero-oligomeric core.

13. A protein scaffold according to any one of the preceding claims, wherein:

i) each subunit monomer comprises less than 300 amino acids;

preferably wherein each subunit monomer comprises less than 200 amino acids;

more preferably wherein each subunit monomer comprises less than 150 amino acids; and/or

ii) the oligomeric core has a molecular weight of less than about 150 kDa, preferably less than about 100 kDa; more preferably less than about 70 kDa.

14. A protein scaffold according to any one of the preceding claims, wherein the oligomeric core (i) does not comprise an Fc region of an antibody, or (ii) does not comprise a CH2 domain, or (iii) does not comprise a CH3 domain, or (iv) does not comprise a CH2 domain or a CH3 domain.

15. A protein scaffold according to any one of the preceding claims, wherein the oligomeric core comprises a soluble multimerising structural element of a multimeric protein.

16. A protein scaffold according to claim **15**, wherein the multimeric protein comprises a Collagen VIII NC1 (noncollagenous) domain, a Collagen X NCI (noncollagenous) domain, a Clq head domain, a CutAl protein, a Macrophage Migration Inhibitory Factor (MIF or MIF-2), a Tumor Necrosis Factor (TNF) or a homolog or paralog thereof.

17. A protein scaffold according to claim **15** or claim **16**, wherein the multimerising structural element comprises a polypeptide have at least 30% or at least 50% amino acid identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 19, SEQ ID NO: 29, SEQ ID NO: 60, SEQ ID

NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 42, SEQ ID NO: 31, or SEQ ID NO: 58.

18. A protein complex comprising a protein scaffold according to any one the preceding claims, wherein the first binding site is bound to a first polypeptide target attached to a first effector moiety, and the second binding site is bound to a second polypeptide target attached to a second effector moiety.

19. The protein complex according to claim **18**, wherein the first binding site/polypeptide target pair and the second binding site/polypeptide target pair are each independently selected from the following combinations: (i) any one of SEQ ID NO: 4, 6 or 8 with any one of SEQ ID NOs: 5, 7 or 9; (ii) SEQ ID NO: 12 with SEQ ID NO: 13 or 15; (iii) SEQ ID NO: 5 with SEQ ID NO: 11; (iv) SEQ ID NO: 15 with SEQ ID NO: 16); (v) SEQ ID NO: 17 with SEQ ID NO: 18; or (vi) SEQ ID NO: 23 with SEQ ID NO: 16).

20. A screening platform comprising a library, wherein said library comprises:

a plurality of populations of protein complexes according to claim **18** or claim **19**, wherein the populations of protein complexes each comprise a different combination of first effector moieties, second effector moieties and/or oligomeric core; or

a plurality of protein scaffolds according to any of claims **1** to **17**, a plurality of first effector moieties that are able to bind specifically to the first binding site, and a plurality of second effector moieties that are able to bind specifically to the second binding site.

21. A method for identifying a therapeutic drug or drug analog, the method comprising:

providing a protein complex according to claim **18** or claim **19**;

contacting the protein complex with a biological system; and

measuring whether the protein complex induces a desired change in a property function of the biological system;

and optionally further comprising selecting a protein complex that induces a desired change in a property of the biological system.

22. A method according to claim **21**, further comprising:

synthesizing a therapeutic drug or drug candidate comprising the oligomeric core of the scaffold of the protein complex of the identified therapeutic drug or drug analog attached to the first and second effector moieties of said protein complex.

23. A therapeutic drug or drug candidate obtainable or obtained according to the method of claim **21** or claim **22**.

24. A therapeutic drug or drug candidate, comprising:

(a) an oligomeric core comprising a plurality of subunit monomers attached to one or more first effector moieties and one or more second effector moieties; wherein said one or more first effector moieties and said one or more second effector moieties are positioned on the same face of the oligomeric core; and

wherein (i) said one or more first effector moieties comprise two or more first effector moieties and said one or more second effector moieties comprise two or more second effector moieties; and/or (ii) said oligomeric core does not comprise an antibody or antibody fragment; or

(b) a monomeric polypeptide attached to one or more first effector moieties and one or more second effector moieties; wherein said one or more first effector moieties and said one or more second effector moieties are positioned on the same face of the monomeric polypeptide; and

wherein (i) said one or more first effector moieties comprise two or more first effector moieties and said one or more second effector moieties comprise two or more second effector moieties; and/or (ii) said oligomeric core does not comprise an antibody or antibody fragment.

25. The therapeutic drug or drug candidate of claim **24**(a), wherein the oligomeric core is as defined in any one of claims **1** to **17**.

26. The therapeutic drug or drug candidate of claim **24** or claim **25**, wherein:

(a) each subunit monomer comprises a Collagen VIII NC1 (noncollagenous) domain, a Collagen X NC1 (noncollagenous) domain, a Clq head domain, a CutAl protein, a Macrophage Migration Inhibitory Factor (MIF or MIF-2), a Tumor Necrosis Factor (TNF) or a homolog or paralog thereof; and/or (ii) each subunit monomer comprises a multimerising structural element comprising a polypeptide having at least 50% amino acid identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 19 or SEQ ID NO: 29, SEQ ID NO: 60, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 42, SEQ ID NO: 31, or SEQ ID NO: 58; or

(b) the monomeric polypeptide comprises Collagen VIII NC1 (noncollagenous) domain, a Collagen X NC1 (noncollagenous) domain, a Clq head domain, a CutAl protein, a Macrophage Migration Inhibitory Factor (MIF or MIF-2), a Tumor Necrosis Factor (TNF) or a homolog or paralog thereof.

27. A polypeptide comprising a first binding domain at the N terminus and a second binding domain at the C terminus, wherein the first and second binding domains are separated by a structural domain, and wherein the first binding domain and second binding domain are able to bind to their targets when expressed on a single cell or immobilised onto a plate or single bead.

28. A polypeptide according to claim **27**, wherein the structural domain is CutAl, an MIF or MIF-2, a TNF or TNF-like protein TLIA or CD40L, or an NC1 from Collagen VIII or Collagen X.

29. A polypeptide according to claim **27** or claim **28**, wherein the first binding domain and second binding domain are different antigen-binding domains, optionally wherein one or both antigen-binding domain is an antigen-binding fragment of an antibody, optionally an scFv or a Fab, or is a single domain antibody (sdAb), or other antibody mimetics or scaffolds selected to bind specific targets, or other proteins or peptides capable of specific binding with a biological molecule.

30. A polypeptide according to claim **27** or claim **28**, wherein the first binding domain and second binding domain are catcher domains each able to form an isopeptide linkage with a cognate peptide, optionally wherein the cognate peptide for the first binding domain is different from the cognate peptide for the second binding domain.

31. A polypeptide according to claim **30** wherein each cognate peptide is attached to an antigen-binding domain, optionally wherein one or both cognate peptides are linked to the first and/or second catcher domain by an isopeptide bond.

32. An oligomer comprising two or more polypeptides according to any of claims **27** to **31**.

33. A polypeptide according to any of claims **27** to **31** or an oligomer according to claim **32**, wherein the polypeptide or oligomer comprises the features of any of claims **1** to **26**.

* * * * *