Polypeptides comprising monomer domains that bind to c-Met, or portions thereof, are provided.
Figure 1

LDL-receptor family

β-Propeller domain

EGF-like domain

LDL receptor class A domain

LDLR

ApoER2

VLDLR

LRP2

LRP1
**Figure 4**

Ligands recognized by LDL-receptor family

<table>
<thead>
<tr>
<th>Proteases</th>
<th>Proteins involved in lipoprotein metabolism</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor IXa</td>
<td>apoB100</td>
<td>albumin</td>
</tr>
<tr>
<td>pro-uPA</td>
<td>apoE</td>
<td>transthyretin</td>
</tr>
<tr>
<td>t-PA</td>
<td>apoJ (clusterin)</td>
<td>β-Amyloid precursor protein</td>
</tr>
<tr>
<td>plasminogen</td>
<td>apoH (β₂-glycoprotein I)</td>
<td>RAP</td>
</tr>
<tr>
<td>MMP-9</td>
<td>Lp(a)</td>
<td>complement C3</td>
</tr>
<tr>
<td></td>
<td>hepatic lipase</td>
<td>lactoferrin</td>
</tr>
<tr>
<td></td>
<td>lipoprotein lipase</td>
<td>thyroglobulin</td>
</tr>
<tr>
<td></td>
<td>IDL</td>
<td>thrombospondin</td>
</tr>
<tr>
<td></td>
<td>VLDL</td>
<td>saposin precursor</td>
</tr>
<tr>
<td></td>
<td>β-VLDL</td>
<td>reelin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>insulin</td>
</tr>
<tr>
<td><strong>Inhibitors</strong></td>
<td></td>
<td>parathyroid hormone (PTH)</td>
</tr>
<tr>
<td>α₂-macroglobulin</td>
<td></td>
<td>aprotinin</td>
</tr>
<tr>
<td>PAI-1</td>
<td></td>
<td>α-amylose</td>
</tr>
<tr>
<td>TFPI</td>
<td></td>
<td>C1q</td>
</tr>
<tr>
<td>pancreatic trypsin inhibitor</td>
<td></td>
<td>α₁-microglobulin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β₂-microglobulin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>odorant-binding protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td>epidermal growth factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prolactin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lysozyme</td>
</tr>
<tr>
<td></td>
<td></td>
<td>connective tissue growth factor (CTGF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cytochrome c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>seminal vesicle secretory protein II</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clara cell secretory protein (CCSP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cubulin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>factor VIII</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Complexes</th>
<th>Non-human</th>
<th>Antibiotics</th>
</tr>
</thead>
<tbody>
<tr>
<td>protease/</td>
<td>pseudomonas exotoxin A</td>
<td>gentamicin</td>
</tr>
<tr>
<td>α₁-antitrypsin</td>
<td>circumsporozoite protein</td>
<td>polymyxin B</td>
</tr>
<tr>
<td>protein C inhibitor</td>
<td>trichosanthin</td>
<td></td>
</tr>
<tr>
<td>protease nexin-1</td>
<td>ricin A</td>
<td></td>
</tr>
<tr>
<td>antithrombin</td>
<td>saporin</td>
<td></td>
</tr>
<tr>
<td>C1-inhibitor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>thrombin/heparin cofactor II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cathepsin G/α₁-antichymotrypsin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vitamin-carrier complexes</th>
<th>Viruses</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>vitamin D-bp, vitamin D</td>
<td>HRV2 (Rhino)</td>
<td></td>
</tr>
<tr>
<td>retinol-bp, vitamin A</td>
<td>HCV (Flavi)</td>
<td></td>
</tr>
<tr>
<td>transcobalamin, vitamin B12</td>
<td>BVDV (Flavi)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5

Select monomer domain binders

Select polypeptide Multimers binders

Combinatorial Multimers
Figure 8

- 78% V A D N Q R P S G S A E Q R
- 96% G N G S D P K K K
- 96% I P Y A L L V G M
- 68% T H S G E R P A H E Q R
- 66% 1 2 3 4 5 6 7 8 C 1 2 3 4 5 6
- 9% C S H P G F E Y R M W W M
### FIG. 9A

|    | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A  | 0.0 | 7.1 | 12.3| 3.2 | 1.9 | 1.9 | 7.1 | 0.0 | 15.8| 1.5 | 0.0 | 1.5 | 0.0 | 1.0 | 3.7 | 7.3 | 9.4 |
| C  | 100.0| 0.0 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 100.0| 0.0 | 0.0 | 0.0 | 1.5 | 99.5| 0.0 | 0.0 | 0.0 |
| D  | 0.0 | 5.2 | 4.5 | 19.5| 2.6 | 0.0 | 0.0 | 0.0 | 0.0 | 8.3 | 10.5| 2.3 | 0.8 | 0.0 | 0.0 | 3.7 | 1.6 | 4.7 |
| E  | 0.0 | 4.5 | 9.7 | 2.6 | 37.0| 0.0 | 3.9 | 0.0 | 3.8 | 0.0 | 0.8 | 9.8 | 0.0 | 0.0 | 1.6 | 7.3 | 7.3 |
| F  | 0.0 | 0.6 | 0.0 | 3.9 | 1.9 | 76.0| 0.6 | 0.0 | 0.8 | 0.0 | 0.0 | 0.8 | 0.0 | 0.0 | 0.5 | 1.6 | 2.1 | 2.6 |
| G  | 0.0 | 13.0| 3.2 | 16.2| 1.9 | 0.0 | 0.0 | 0.0 | 12.8| 0.8 | 72.2| 2.3 | 0.0 | 0.0 | 3.1 | 5.2 | 4.2 |
| H  | 0.0 | 1.9 | 1.3 | 5.2 | 3.9 | 1.9 | 3.9 | 0.0 | 3.0 | 0.8 | 4.5 | 9.0 | 0.0 | 0.0 | 4.7 | 3.1 | 5.8 |
| I  | 0.0 | 0.0 | 1.3 | 0.0 | 0.0 | 0.0 | 3.9 | 0.0 | 0.0 | 0.0 | 0.0 | 2.3 | 0.0 | 0.0 | 63.9| 0.0 | 5.2 | 3.7 |
| J  | 0.0 | 3.9 | 3.9 | 1.9 | 1.9 | 0.6 | 7.0 | 0.0 | 11.3| 0.0 | 3.0 | 9.0 | 0.0 | 2.1 | 2.1 | 9.9 | 3.7 |
| K  | 0.0 | 8.4 | 4.5 | 0.0 | 1.3 | 3.9 | 3.9 | 0.0 | 1.5 | 0.0 | 1.5 | 4.5 | 0.0 | 11.0| 5.2 | 12.0| 3.1 |
| L  | 0.0 | 0.6 | 0.6 | 0.0 | 1.3 | 0.6 | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.8 | 0.0 | 0.0 | 0.0 | 0.5 | 3.1 |
| M  | 0.0 | 1.9 | 0.6 | 13.6| 1.3 | 0.0 | 1.9 | 0.0 | 5.3 | 0.0 | 51.9| 5.3 | 3.8 | 0.0 | 0.0 | 5.2 | 4.2 |
| N  | 0.0 | 10.4| 33.8| 1.3 | 0.0 | 0.0 | 9.7 | 0.0 | 4.5 | 0.0 | 0.0 | 0.0 | 0.0 | 45.5| 7.3 | 2.1 |
| P  | 0.0 | 10.4| 2.6 | 1.9 | 30.5| 0.6 | 21.4| 0.0 | 5.3 | 0.0 | 2.3 | 9.8 | 0.0 | 0.0 | 2.6 | 5.8 | 4.7 |
| Q  | 0.0 | 7.1 | 3.2 | 3.2 | 1.3 | 0.6 | 18.2| 0.0 | 7.5 | 0.0 | 4.5 | 32.3| 0.0 | 0.0 | 1.0 | 9.4 | 11.5|
| R  | 0.0 | 18.8| 11.7| 16.2| 3.9 | 0.6 | 9.7 | 0.0 | 15.0| 25.6| 3.0 | 3.0 | 0.0 | 0.0 | 14.7| 10.5| 20.4|
| S  | 0.0 | 5.2 | 2.6 | 5.8 | 3.2 | 1.3 | 7.8 | 0.0 | 2.3 | 6.0 | 0.8 | 6.0 | 0.0 | 1.6 | 2.6 | 0.5 | 5.8 |
| T  | 0.0 | 0.6 | 1.9 | 0.6 | 0.0 | 0.6 | 1.3 | 0.0 | 2.3 | 0.0 | 0.0 | 3.0 | 0.0 | 0.0 | 17.3| 0.5 | 3.1 |
| U  | 0.0 | 0.0 | 0.0 | 0.6 | 1.9 | 0.0 | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 | 0.0 |
| V  | 0.0 | 0.0 | 0.0 | 0.6 | 2.6 | 5.8 | 5.2 | 0.0 | 0.0 | 0.8 | 0.0 | 0.0 | 0.0 | 0.5 | 2.6 | 1.0 | 2.1 |

100.0 85.7 67.5 76.6 73.4 81.2 81.8 100.0 81.2 94.0 77.4 75.9 99.5 92.1 70.7 80.1 60.2
1 9 4 6 3 2 7 1 8 4 2 6 1 3 4 10 6
100.0 73.4 67.5 65.6 67.5 76.0 66.2 100.0 68.4 77.4 72.2 60.9 99.5 63.9 60.2 60.2 59.2
1 7 4 4 2 1 5 1 6 2 1 4 1 1 2 7 6
CAD61944  C  PPCHFPCGAAGTSGATA  CYLPADR  CNYQTFCADG  ADERR  C
CAD61944  C  QPGNFRCRDBK  CVYETMV  CDGQPDADG  SDEMD  C
ENSG00000181006  C  PEITDLCRDKK  CIASHLL  CDYKPDSCDR  SDEAH  C
ENSP00000320248  C  NNRTFKCQGDI  CFRKQNAKCDGTVDQPDQ  SDEBG  C
ENSP00000277547  C  PPCHHHCQNKV  CVEPQQL  CDGSQDIQDL  SDENPLT  C
ENSP00000320022  C  KQCHLACGDL  CVPQEQL  CFPSQCCAGG  EDRAA  C
ENSP00000313222  C  PONSFSGNSQ  CVTVKVPEDQDQEDGSDG  SDEAH  C
Figure 13

Pan library against cells, tissue, immobilized cell-extract or serum

Pan

Repeat panning 2-3 times

Amplify phage

Wash & elute

Final eluate

Monomer or multimer library
Figure 15

Stain membrane using cDNA phage specific MAb (HRP labeled)
Add substrate
Cut positive spots out of membrane and add to PCR tube
PCR amplify cDNA and MB sequences
Sequence PCR fragments
Format Variations

Figure 16

Effector Sequence
Targeting Domains

Half-life + Avidity
Avidity + Agonism
Recruit Function

Avidity

T-cell

Cell
Figure 19

Best Monomer

3 $10^8$ Libraries

3 Pools of Improveds

1 Library of Combineds

Few Improved Monomers

Few Improved Trimers
Figure 20

Library of Monomers → Best Monomers → Best Multimers → Optimized Monomer → Linker-Optimized Dimer → Optimized Trimer
C-MET KINASE BINDING PROTEINS

BACKGROUND OF THE INVENTION

[0001] Hepatocyte Growth Factor/Scatter Factor (HGF/SF) is a mesenchyme-derived pleiotropic factor, which regulates cell growth, cell motility, and morphogenesis of various types of cells and mediates epithelial-mesenchymal interactions responsible for morphogenic tissue interactions during embryonic development and organogenesis. Although HGF was originally identified as a potent mitogen for hepatocytes, it has also been identified as an angiogenic growth factor.

[0002] Met was first identified in the 1980s as an oncogene and is the receptor for HGF. The proto-oncogene c-Met, was found to encode a receptor tyrosine kinase. In response to HGF treatment a range of activities are observed: phosphorylation of receptor, docking of signaling intermediates Gab-1/Grb2, culminating in activation of kinases such as P13K, ERK1 and 2, and AKT. These activities aid in cell growth, survival, migration, and neovascularisation.

[0003] Inappropriate expression or signaling of the receptor tyrosine kinase Met and its ligand Hepatocyte Growth Factor/Scatter Factor (HGF/SF) is associated with an aggressive phenotype and poor clinical prognosis for a wide variety of solid human tumors.

[0004] Four lines of evidence cement the case for a role of c-Met in cancer:

[0005] First, mouse and human cell lines that ectopically overexpress HGF and/or Met become tumorigenic and metastatic in athymic nude mice. Secondly, downregulation of Met or HGF expression in human tumour cells decreases their tumorigenic potential. Mouse models that express the receptor or ligand as a transgene develop various types of tumour and metastatic tumors. Third, a large number of studies show that HGF and/or Met are frequently expressed in carcinomas, in other types of human solid tumours and in their metastases, and that HGF and/or Met over- or misexpression often correlates with poor prognosis. Fourth, unequivocal evidence that implicates Met in human cancer is provided by the activating mutations that have been discovered in both sporadic and inherited forms of human renal papillary carcinomas.

BRIEF SUMMARY OF THE INVENTION

[0006] The present invention provides a polypeptide comprising a monomer domain that binds to c-Met. In some embodiments, the monomer domain:

[0007] is a non-naturally-occurring monomer domain consisting of 30 to 50 amino acids;

[0008] comprises at least one disulfide bond; and

[0009] binds to an ion.

[0010] In some embodiments, binding of at least one monomer domain to c-Met inhibits dimerization of Met. In some embodiments, at least one monomer domain binds to the Sema domain of c-Met, thereby preventing binding of Met ligands to c-Met.

[0011] In some embodiments, the polypeptide comprises at least one and no more than six monomer domains. In some embodiments, the polypeptide comprises at least two monomer domains and the monomer domains are linked by a linker. In some embodiments, the linker is a peptide linker.

[0012] In some embodiments, the monomer domains are each between 35 to 45 amino acids.

[0013] In some embodiments, the polypeptide comprises at least one monomer domain with binding specificity for a blood factor, thereby increasing the serum half-life of the polypeptide compared to a polypeptide lacking the blood factor monomer domain. In some embodiments, the blood factor is serum albumin, an immunoglobulin or an erythrocyte.

[0014] In some embodiments, each monomer domain comprises two disulfide bonds. In some embodiments, each monomer domain comprises three disulfide bonds.

[0015] In some embodiments, the ion is a metal ion. In some embodiments, the ion is a calcium ion.

[0016] In some embodiments, at least one of the monomer domains is derived from a LDL-receptor class A domain. In some embodiments, at least one of the monomer domains is derived from an EGF-like domain.

[0017] In some embodiments, the monomer comprises an amino acid sequence in which at least 10% of the amino acids in the sequence are cysteine; and/or at least 25% of the amino acids are non-naturally-occurring amino acids.

[0018] The present invention also provides methods for identifying a polypeptide that binds to c-Met. In some embodiments, the method comprises,

[0019] screening a library of polypeptides for affinity to c-Met; and

[0020] selecting a polypeptide comprising at least one monomer domain that binds to c-Met, wherein the monomer domain:

[0021] is a non-naturally-occurring monomer domain;

[0022] comprises at least one disulfide bond; and

[0023] binds to an ion.

[0024] In some embodiments, the selecting step comprises selecting a polypeptide that reduces HGF-mediated cell proliferation and/or migration. In some embodiments, the method further comprises selecting a polypeptide that inhibits tumor growth in an animal.

[0025] In some embodiments, the monomer domain comprises an amino acid sequence in which at least 10% of the amino acids in the sequence are cysteine; and/or at least 25% of the amino acids are non-naturally-occurring amino acids.

[0026] In some embodiments, the method further comprises

[0027] linking the monomer domain in the selected polypeptide to a second monomer domain to form a library of multimers, each multimer comprising at least two monomer domains;

[0028] screening the library of multimers for the ability to bind to c-Met; and

[0029] selecting a multimer that binds c-Met.
In some embodiments, the method further comprises linking the monomer domain in the selected polypeptide to a second monomer domain to form a library of multimers, each multimer comprising at least two monomer domains;

screening the library of multimers for the ability to bind to a target molecule other than the c-Met; and

selecting a multimer that binds to the target molecule.

In some embodiments, the method further comprises a step of mutating at least one monomer domain, thereby providing a library comprising mutated monomer domains.

In some embodiments, the library of monomer domains is expressed as a phage display, ribosome display or cell surface display.

In some embodiments, the polypeptide comprises at least two monomer domains and the monomer domains are linked by a linker. In some embodiments, the linker is a peptide linker. In some embodiments, the linker is between 4 to 12 amino acids long.

In some embodiments, the monomer domains are each between 35 to 45 amino acids.

In some embodiments, each monomer domain comprises two disulfide bonds. In some embodiments, each monomer domain comprises three disulfide bonds.

In some embodiments, the ion is a metal ion. In some embodiments, the ion is a calcium ion.

In some embodiments, at least one of the monomer domains is derived from a LDL-receptor class A domain. In some embodiments, at least one of the monomer domains is derived from an EGF-like domain.

In some embodiments, the monomer domain comprises an amino acid sequence in which at least 10% of the amino acids in the sequence are cysteine; and/or at least 25% of the amino acids are non-naturally-occurring amino acids.

The present invention also provides polynucleotides encoding a polypeptide comprising a monomer domain that binds to c-Met, wherein the monomer domain:

is a non-naturally-occurring monomer domain consisting of 30 to 50 amino acids;

comprises at least one disulfide bond; and

binds to an ion.

DEFINITIONS

"Met" also referred to as "c-Met," refers to the Hepatocyte Growth Factor/Scatter Factor (HGF/SF)-binding receptor tyrosine kinase. In response to HGF treatment a range of activities are observed: phosphorylation of receptor, docking of signaling intermediates Gab-1/Grb2, culminating in activation of kinases such as P13K, ERK1 and 2, and AKT. These activities aid in cell growth, survival, migration, and neovascularisation. See, e.g., Birchmeier et al., Mol. Cell Biol. 4:915-925 (2003). The amino acid sequence of Met is known and is displayed in SEQ ID NO:1. See, e.g., Park et al., Proc. Natl. Acad. Sci. USA 84(18):6379 (1987).

The term “monomer domain” or “monomer” is used interchangeably herein refer to a discrete region found in a protein or polypeptide. A monomer domain forms a native three-dimensional structure in solution in the absence of flanking native amino acid sequences. Monomer domains of the invention will often bind to a target molecule. For example, a polypeptide that forms a three-dimensional structure that binds to a target molecule is a monomer domain. As used herein, the term “monomer domain” does not encompass the complementarity determining region (CDR) of an antibody.

The term “loop” refers to that portion of a monomer domain that is typically exposed to the environment by the assembly of the scaffold structure of the monomer domain protein, and which is involved in target binding. The present invention provides three types of loops that are identified by specific features, such as, potential for disulfide bonding, bridging between secondary protein structures, and molecular dynamics (i.e., flexibility). The three types of loop sequences are a cysteine-defined loop sequence, a structure-defined loop sequence, and a B-factor-defined loop sequence.

As used herein, the term “cysteine-defined loop sequence” refers to a subsequence of a naturally occurring monomer domain-encoding sequence that is bound at each end by a cysteine residue that is conserved with respect to at least one other naturally occurring monomer domain of the same family. Cysteine-defined loop sequences are identified by multiple sequence alignment of the naturally occurring monomer domains, followed by sequence analysis to identify conserved cysteine residues. The sequence between each consecutive pair of conserved cysteine residues is a cysteine-defined loop sequence. The cysteine-defined loop sequence does not include the cysteine residues adjacent to each terminus. Monomer domains having cysteine-defined loop sequences include the LDL receptor A-domains, EGF-like domains, sushi domains, Fibronectin type 1 domains, and the like. Thus, for example, in the case of LDL receptor A-domains represented by the consensus sequence, CX_{X_1}CX_{X_2}CX_{X_3}CX_{X_4}C (see also FIG. 9), wherein X_0, X_0, X_3, and X_6 each represent a cysteine-defined loop sequence comprising the designated number of amino acids.

As used herein, the term “structure-defined loop sequence” refers to a subsequence of a monomer-domain encoding sequence that is bound at each end by subsequences that each form a secondary structure. Secondary structures for proteins with known three dimensional structures are identified in accordance with the algorithm STRIDE for assigning protein secondary structure as described in Frishman, D. and Argos, P. (1995) “Knowledge-based secondary structure assignment,” Proteins, 23(4):566-79 (see also http://mr.pmc.ac.uk/Registered/Option/stridle.html at the World Wide Web). Secondary structures for proteins with unknown or uncharacterized three dimensional structures are identified in accordance with the algorithm described in Jones, D. T. (1999), “Protein secondary structure prediction based on position-specific scoring matrices,” J. Mol. Biol., 292:195-202 (see also McGuilly, L.

[0052] The term “B-factor-defined loop sequence” refers to a subsequence of at least three amino acid residues of a monomer-domain encoding sequence in which the B-factors for the alpha carbons in the B-factor-defined loop are among the 25% highest alpha carbon B factors in the entire monomer domain. Typically the average alpha-carbon B-factor for the subsequence is at least about 65. As used herein, the term “B-factor” (or “temperature factor” or “Debye-Waller factor”) is derived from X-ray scattering data. The B-factor is a factor that can be applied to the X-ray scattering term for each atom, or for groups of atoms, that describes the degree to which electron density is spread out B-factors employed in the practice of the present invention may be either isotropic or anisotropic. The term “average alpha-carbon B-factor” refers to:

\[
\frac{\sum_{i=1}^{n} B\text{-factor}_{\alpha C}}{n}
\]

where \(n\) corresponds to the number of residues in the loop, and is at least 3, and \(B\text{-factor}_{\alpha C}\) is the B-factor for the alpha carbon of amino acid residue \(i\) of the loop.

[0053] The term “multimer” is used herein to indicate a polypeptide comprising at least two monomer domains. The separate monomer domains in a multimer can be joined together by a linker. A multimer is also known as a combinatorial mosaic protein or a recombinant mosaic protein.

[0054] The term “family” and “family class” are used interchangeably to indicate proteins that are grouped together based on similarities in their amino acid sequences. These similar sequences are generally conserved because they are important for the function of the protein and/or the maintenance of the three dimensional structure of the protein. Examples of such families include the LDL Receptor A-domain family, the EGF-like family, and the like.

[0055] The term “ligand,” also referred to herein as a “target molecule,” encompasses a wide variety of substances and molecules, which range from simple molecules to complex targets. Target molecules can be proteins, nucleic acids, lipids, carbohydrates or any other molecule capable of recognition by a polypeptide domain. For example, a target molecule can include a chemical compound (i.e., non-biological compound such as, e.g., an organic molecule, an inorganic molecule, or a molecule having both organic and inorganic atoms, but excluding nucleotides and proteins), a mixture of chemical compounds, an array of spatially localized compounds, a biological macromolecule, a bacteriophage peptide display library, a polypeptide display library, an extract made from a biological material such as bacteria, plants, fungi, or animal (e.g., mammalian) cells or tissue, a protein, a toxin, a peptide hormone, a cell, a virus, or the like. Other target molecules include, e.g., a whole cell, a whole tissue, a mixture of related or unrelated proteins, a mixture of viruses or bacterial strains or the like. Target molecules can also be defined by inclusion in screening assays described herein or by enhancing or inhibiting a specific protein interaction (i.e., an agent that selectively inhibits a binding interaction between two predetermined polypeptides).

[0056] The term “linker” is used herein to indicate a moiety or group of moieties that joins or connects two or more discrete separate monomer domains. The linker allows the discrete separate monomer domains to remain separate when joined together in a multimer. The linker moiety is typically a substantially linear moiety. Suitable linkers include polypeptides, polynucleic acids, peptide nucleic acids and the like. Suitable linkers also include optionally substituted alkylene moieties that have one or more oxygen atoms incorporated in the carbon backbone. Typically, the molecular weight of the linker is less than about 2000 daltons. More typically, the molecular weight of the linker is less than about 1500 daltons and usually is less than about 1000 daltons. The linker can be small enough to allow the discrete separate monomer domains to cooperate, e.g., where each of the discrete separate monomer domains in a multimer binds to the same target molecule via separate binding sites. Exemplary linkers include a polynucleotide encoding a polypeptide, or a polypeptide of amino acids or other non-naturally occurring moieties. The linker can be a portion of a native sequence, a variant thereof, or a synthetic sequence. Linkers can comprise, e.g., naturally occurring, non-naturally occurring amino acids, or a combination of both.

[0057] The term “separate” is used herein to indicate a property of a moiety that is independent and remains independent even when complexed with other moieties, including for example, other monomer domains. A monomer domain is a separate domain in a protein because it has an independent property that can be recognized and separated from the protein. For instance, the ligand binding ability of the A-domain in the LDLR is an independent property. Other examples of separate include the separate monomer domains in a multimer that remain separate independent domains even when complexed or joined together in the multimer by a linker. Another example of a separate property is the separate binding sites in a multimer for a ligand.

[0058] As used herein, “directed evolution” refers to a process by which polynucleotide variants are generated, expressed, and screened for an activity (e.g., a polypeptide with binding activity) in a recursive process. One or more candidates in the screen are selected and the process is then repeated using polynucleotides that encode the selected candidates to generate new variants. Directed evolution involves at least two rounds of variation generation and can include 3, 4, 5, 10, 20 or more rounds of variation generation and selection. Variation can be generated by any method known to those of skill in the art, including, e.g., by error-prone PCR, gene recombination, chemical mutagenesis, and the like.

[0059] The term “shuffling” is used herein to indicate recombination between non-identical sequences. In some embodiments, shuffling can include crossover via homologous recombination or via non-homologous recombination, such as via cre/lox and/or flp/frt systems. Shuffling can be
carried out by employing a variety of different formats, including for example, in vitro and in vivo shuffling formats, in silico shuffling formats, shuffling formats that utilize either double-stranded or single-stranded templates, primer based shuffling formats, nucleic acid fragmentation-based shuffling formats, and oligonucleotide-mediated shuffling formats, all of which are based on recombination events between non-identical sequences and are described in more detail or referenced herein below, as well as other similar recombination-based formats. The term “random” as used herein refers to a polynucleotide sequence or an amino acid sequence composed of two or more amino acids and constructed by a stochastic or random process. The random polynucleotide sequence or amino acid sequence can include framework or scaffolding motifs, which can comprise invariant sequences.

The term “pseudo-random” as used herein refers to a set of sequences, polynucleotide or polypeptide, that have limited variability, so that the degree of residue variability at some positions is limited, but any pseudorandom position is allowed at least some degree of residue variation.

The terms “polypeptide,” “peptide,” and “protein” are used herein interchangeably to refer to an amino acid sequence of two or more amino acids.

The term “amino acid” refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxypyridone, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfoxide. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. “Amino acid mimetics” refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

“Conservative amino acid substitution” refers to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

The phrase “nucleic acid sequence” refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes chromosomal DNA, self-replicating plasmids and DNA or RNA that performs a primarily structural role.
to a multimer, indicates that the multimer comprises a linker and a monomer that are not found in the same relationship to each other in nature (e.g., they form a non-naturally occurring fusion protein).

[0072] A “non-naturally-occurring amino acid” in a protein sequence refers to any amino acid other than the amino acid that occurs in the corresponding position in an alignment with a naturally-occurring polypeptide with the lowest smallest sum probability where the comparison window is the length of the monomer domain queried and when compared to the non-redundant (“nr”) database of Genbank using BLAST 2.0 as described herein.

[0073] “Percentage of sequence identity” is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue 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 and multiplying the result by 100 to yield the percentage of sequence identity.

[0074] The terms “identical” or percent “identity,” in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be “substantially identical.” This definition also refers to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 amino acids or nucleotides in length, or more preferably over a region that is 75-100 amino acids or nucleotides in length.

[0075] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0076] A “comparison window”, as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1970) Adv. Appl. Math. 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman (1988) Proc. Nat’l. Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (see, e.g., Ausabel et al., Current Protocols in Molecular Biology (1995 supplement)).

[0077] One example of a useful algorithm is the BLAST 2.0 program, which is described in Altschul et al. (1990) J. Mol. Biol. 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0).

For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=−4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) Proc. Natl. Acad. Sci. USA 86:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=−4, and a comparison of both strands.

[0078] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5877). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

BRIEF DESCRIPTION OF THE DRAWINGS

[0079] FIG. 1 schematically illustrates the type, number and order of monomer domains found in members of the
LDL-receptor family. These monomer domains include β-Propeller domains, EGF-like domains and LDL receptor class A-domains. The members shown include low-density lipoprotein receptor (LDLR), ApoE Receptor 2 (ApoER2), very-low-density lipoprotein receptor (VLDLR), LDLR-related protein 2 (LRP2) and LDLR-related protein 1 (LRP1).

[0080] FIG. 2 schematically illustrates the alignment of partial amino acid sequence from a variety of the LDL-receptor class A-domains (SEQ ID NOS: 103, 100, 65, 117, 128, 21, 29, 39, 30, 77, 58, 50, and 14, respectively in order of appearance) to demonstrate the conserved cysteines.

[0081] FIG. 3, panel A schematically illustrates an example of an A-domain. Panel A schematically illustrates conserved amino acids in an A-domain of about 40 amino acids long. The conserved cysteine residues are indicated by C, and the negatively charged amino acids are indicated by a circle with a minus (−) sign. Circles with an “H” indicate hydrophobic residues. Panel B schematically illustrates two folded A-domains connected via a linker. Panel B also indicates two calcium binding sites, dark circles with Ca²⁺, and three disulfide bonds within each folded A-domain for a total of 6 disulfide bonds.

[0082] FIG. 4 indicates some of the ligands recognized by the LDL-receptor family, which include inhibitors, proteases, protease complexes, vitamin-carrier complexes, proteins involved in lipoprotein metabolism, non-human ligands, antibodies, viruses, and others.

[0083] FIG. 5 schematically illustrates a general scheme for identifying monomer domains that bind to a ligand, isolating the selected monomer domains, creating multimers of the selected monomer domains by joining the selected monomer domains in various combinations and screening the multimers to identify multimers comprising more than one monomer that binds to a ligand.

[0084] FIG. 6 is a schematic representation of another selection strategy (guided selection). A monomer domain with appropriate binding properties is identified from a library of monomer domains. The identified monomer domain is then linked to monomer domains from another library of monomer domains to form a library of multimers. The multimer library is screened to identify a pair of monomer domains that bind simultaneously to the target. This process can then be repeated until the optimal binding properties are obtained in the multimer.

[0085] FIG. 7 shows the multimerization process of monomer domains. The target-binding monomer hits are amplified from a vector. This mixture of target-binding monomer domains is then cleaved and mixed with an optimal combination of linker and stopper oligonucleotides. The multimers that are generated are then cloned into a suitable vector for the second selection step for identification of target-binding multimers.

[0086] FIG. 8 depicts common amino acids in each position of the A domain. The percentages above the amino acid positions refer to the percentage of naturally-occurring A domains with the inter-cysteine spacing displayed. Potential amino acid residues in bold depicted under each amino acid position represent common residues at that position. The final six amino acids, depicted as lighter-colored circles, represent linker sequences. The two columns of italicized amino acid residues at positions 2 and 3 of the linker represent amino acid residues that do not occur at that position. Any other amino acid (e.g., A, D, E, G, H, I, K, L, N, P, Q, R, S, T, and V) may be included at these positions.

[0087] FIG. 9 displays the frequency of occurrence of amino acid residues in naturally-occurring A domains with the following spacing between cysteines: Cx,Cx,Cx,Cx,Cx,Cx,Cx,Cx, C (SEQ ID NO: 199).

[0088] FIG. 10 depicts an alignment of A domains (SEQ ID NO: 1-197). At the top and the bottom of the figure, small letters (a-q) indicate conserved residues.

[0089] FIG. 11 depicts linkage of domains via partial linkers.

[0090] FIG. 12 is a graphical representation of the regions of sequence identity between the sequences of two different selected clones and known human sequences from a database. The horizontal bars indicate areas of sequence identity between the selected clone and the human sequence and the numbers indicate the exact amino acid numbers that define the region of identity. The vertical arrow depicts an acceptable crossover sequence.

[0091] FIG. 13 illustrates screening a library of monomer domains against multiple ligands displayed on a cell.

[0092] FIG. 14 illustrates identification of monomers that were selected to bind to one of a plurality of ligands.

[0093] FIG. 15 illustrates an embodiment for identifying polynucleotides encoding ligands and monomer domains.

[0094] FIG. 16 illustrates monomer domain and multimer embodiments for increased avidity. While the figure illustrates specific gene products and binding affinities, it is appreciated that these are merely examples and that other binding targets can be used with the same or similar conformations.

[0095] FIG. 17 illustrates monomer domain and multimer embodiments for increased avidity. While the figure illustrates specific gene products and binding affinities, it is appreciated that these are merely examples and that other binding targets can be used with the same or similar conformations.

[0096] FIG. 18 illustrates various possible antibody-monomer or multimer conformations. In some embodiments, the monomer or multimer replaces the Fab fragment of the antibody.

[0097] FIG. 19 illustrates a method for intradomain optimization of monomers.

[0098] FIG. 20 illustrates a possible sequence of multimer optimization steps in which optimal monomers and then multimers are selected followed by optimization of monomers, optimization of linkers and then optimization of multimers.

[0099] FIG. 21 illustrates four possible ways to recombine monomer and/or multimer libraries to introduce new variation.

[0100] FIG. 22 depicts a possible conformation of a multimer of the invention comprising at least one monomer domain that binds to a half-life extending molecule and other monomer domains binding to two other different
molecules. In the Figure, two monomer domains bind to a first target molecule and a separate monomer domain binds to a second target molecule.

**DETAILED DESCRIPTION OF THE INVENTION**

I. Introduction

[0101] The present invention provides for non-naturally-occurring proteins that bind to c-Met. Generally, the proteins of the present invention comprise a domain that binds to c-Met. These domains may be readily identified using a variety of polypeptide scaffolds to generate a plurality of polypeptide variants and then selecting a variant that binds to c-Met. The present invention therefore also provides for selecting a protein that binds to c-Met. Proteins that bind c-Met are useful, e.g., for treating individuals with solid tumors that express c-Met. The polypeptides of the invention are also useful to detect tissues in which Met is expressed and can be used to target molecules to those tissues.

[0102] c-Met is inactive in its resting monomer state and dimer formation results in receptor activation (often even in absence of ligand binding). The mature form of the receptor consists of a solely extracellular α chain and a longer β chain encompassing the remainder of the extracellular domain, a transmembrane domain and a cytoplasmic tail. The cytoplasmic tail contains the juxtamembrane domain, a kinase domain and docking sites for signaling intermediates. The α chain and the first 212 amino acids of the β also known as the Sema domain (Kong-Beltran, et al., Cancer Cell 6:75-84 (2004), are sufficient for binding to HGF. The rest of the extracellular portion of the β chain consists of a cysteine-rich C domain and four repeats of an unusual immunoglobulin domain. Accordingly, in some embodiments, the polypeptides of the invention comprise at least one monomer domain that inhibits dimerization of c-Met α and β chains and/or functions as an antagonist to prevent ligands of c-Met from binding and/or activating c-Met.

[0103] While the present invention provides for polypeptides comprising single domains, multimers of the domains may also be synthesized and used. In some embodiments, all of the domains of the multimer bind c-Met. In some of these embodiments, each of the domains are identical and bind to the same portion (i.e., “epitope”) of c-Met. For example, in some embodiments, the monomer domains bind to the Sema domain of c-Met. In other embodiments, at least some of the domains in the multimer bind to different portions of c-Met. In yet other embodiments, at least some of the domains of the polypeptide bind to a molecule or molecules other than c-Met.

II. Monomers

[0104] Monomer domains can be polypeptide chains of any size. In some embodiments, monomer domains have about 25 to about 500, about 30 to about 200, about 30 to about 100, about 35 to about 50, about 35 to about 100, about 90 to about 200, about 30 to about 250, about 30 to about 60, about 9 to about 150, about 100 to about 150, about 25 to about 50, or about 30 to about 150 amino acids. Similarly, a monomer domain of the present invention can comprise, e.g., from about 30 to about 200 amino acids; from about 25 to about 180 amino acids; from about 40 to about 150 amino acids; from about 50 to about 130 amino acids; or from about 75 to about 125 amino acids. Monomer domains can typically maintain a stable conformation in solution, and are often heat stable, e.g., stable at 95° C. for at least 10 minutes without losing binding affinity. Sometimes, monomer domains can fold independently into a stable conformation. In one embodiment, the stable conformation is stabilized by ions (e.g., such as metal or calcium ions). The stable conformation can optionally contain disulfide bonds (e.g., at least one, two, or three or more disulfide bonds). The disulfide bonds can optionally be formed between two cysteine residues. In some embodiments, monomer domains, or monomer domain variants, are substantially identical to the sequences exemplified (e.g., LDL A domains, EGF domains, etc., or otherwise referenced herein.

[0105] The invention provides monomer domains that bind to c-Met or a portion thereof. A portion of a polypeptide can be, e.g., at least 5, 10, 15, 20, 30, 50, 100, or more contiguous amino acids of the polypeptide.


[0107] Monomer domains that are particularly suitable for use in the practice of the present invention are (1) β sandwich domains; (2) β-barrel domains; or (3) cysteine-rich domains comprising disulfide bonds. Cysteine-rich domains employed in the practice of the present invention typically do not form an α helix, a β sheet, or a β-barrel structure. Typically, the disulfide bonds promote folding of the domain into a three-dimensional structure. Usually, cysteine-rich domains have at least two disulfide bonds, more typically at least three disulfide bonds. In some embodiments, at least 5, 10, 15 or 20% of the amino acids in a monomer domain are cysteines.

[0108] Domains can have any number of characteristics. For example, in some embodiments, the domains have low or no immunogenicity in an animal (e.g., a human). Domains can have a small size. In some embodiments, the domains are small enough to penetrate skin or other tissues. Domains can have a range of in vivo half-lives or stabilities.

[0109] Illustrative monomer domains suitable for use in the practice of the present invention include, e.g., an EGF-like domain, a Kringle-domain, a fibronectin type I domain, a fibronectin type II domain, a fibronectin type III domain, a PAN domain, a Gla domain, a SRCR domain, a Kunitz/Bovine pancreatic trypsin Inhibitor domain, a Kazal-type serine protease inhibitor domain, a Trefoil (P-type) domain, a von Willebrand factor type C domain, an Anaphylatoxin-like domain, a CUB domain, a thyroglobulin type I repeat, LDL-receptor class A domain, a Sushi domain, a Link domain, a Thrombospondin type I domain, an Immunoglobulin-like domain, a C-type lectin domain, a MAM domain, a von Willebrand factor type A domain, a Somatomedin B domain, a WAP-type four disulfide core domain, a f5/8 type C domain, a Hemopexin domain, an SH2 domain, an SH3 domain, a Laminin-type EGF-like domain, a C2 domain, and other such domains known to those of ordinary skill in the art, as well as derivatives and/or variants thereof. In some embodiments, the monomer domain is not the C2 domain. FIG. 1 schematically diagrams various kinds of monomer domains found in molecules in the LDL-receptor family.
In some embodiments, suitable monomer domains (e.g., domains with the ability to fold independently or with some limited assistance) can be selected from the families of protein domains that contain β-sandwich or β-barrel three dimensional structures as defined by such computational sequence analysis tools as Simple Modular Architecture Research Tool (SMART), see Shultz et al., SMART: a web-based tool for the study of genetically mobile domains, (2000) Nucleic Acids Research 28(1):231-234 or CATH (see Pearl et al., Assigning genomic sequences to CATH, (2000) Nucleic Acids Research 28(1):277-282).

In another embodiment, monomer domains of the present invention include domains other than a fibronectin type III domain, an anticalin domain and a Ig-like domain from CTLA-4. Some aspects of these domains are described in WO 01/64942 entitled “Protein scaffolds for antibody mimics and other binding proteins” by Lipovsek et al., published on Sep. 7, 2001, WO99/16873 entitled “Anticalins” by Beste et al., published Apr. 8, 1999 and WO 00/60070 entitled “A polypeptide structure for use as a scaffold” by Desmet et al., published on Oct. 12, 2000.

As described supra, monomer domains are optionally cysteine rich. Suitable cysteine rich monomer domains include, e.g., the LDL receptor class A domain (“A-domain”) or the EGF-like domain. The monomer domains can also have a cluster of negatively charged residues. Optionally, the monomer domains contain a repeated sequence, such as YWTD (SEQ ID NO: 198) as found in the β-Propeller domain.

Other features of monomer domains can include the ability to bind ligands (e.g., as in the LDL receptor class A domain, or the CUB domain (complement C1r/C1s, Uegf, and bone morphogenic protein-1 domain)), the ability to participate in endocytosis or internalization (e.g., as in the cytoplasmic tail of the LDL receptor or the cytoplasmic tail of Megalin), the ability to bind an ion (e.g., Ca^{2+} binding by the LDL receptor A-domain), and/or the ability to be involved in cell adhesion (e.g., as in the EGF-like domain). Monomer domains that bind ions to maintain their secondary structure include, e.g., A-domain, EGF domain, EF Hand (e.g., as those found in present in calmodulin and troponin C), Cadherin domain, C-type lectin, C2 domain, Annexin, GlA-domain, Trombospondin type 3 domain, all of which bind calcium, and zinc fingers, which bind ions to maintain their secondary structure. For example, C11HC4 type (RING finger), Integrase Zinc binding domain, PHD finger, GATA zinc finger, FYVE zinc finger, B-box zinc finger), which bind zinc. Without intending to limit the invention, it is believed that ion-binding provides stability of secondary structure while providing sufficient flexibility to allow for numerous binding conformations depending on primary sequence.

Characteristics of a monomer domain can include the ability to fold independently and the ability to form a stable structure. Thus, the structure of the monomer domain is often conserved, although the polynucleotide sequence encoding the monomer need not be conserved. For example, the A-domain structure is conserved among the members of the A-domain family, while the A-domain nucleic acid sequence is not. Thus, for example, a monomer domain is classified as an A-domain by its cysteine residues and is affinity for calcium, not necessarily by its nucleic acid sequence. See, FIG. 2.

As described herein, monomer domains may be selected for the ability to bind to targets other than the target that a homologous naturally occurring domain may bind. Thus, in some embodiments, the invention provides monomer domains (and multimers comprising such monomers) that do not bind to the target or the class or family of target proteins that a homologous naturally occurring domain may bind.

Specifically, the A-domains (sometimes called “complement-type repeats”) contain about 30-50 or 30-65 amino acids. In some embodiments, the domains comprise about 35-45 amino acids and in some cases about 40 amino acids. Within the 30-50 amino acids, there are about 6 cysteine residues. Of the six cysteines, disulfide bonds typically are found between the following cysteines: C1 and C3, C2 and C5, C4 and C6. The cysteine residues of the domain are disulfide linked to form a compact, stable, functionally independent moiety. See, FIG. 3. Clusters of these repeats make up a ligand binding domain, and differential clustering can impart specificity with respect to the ligand binding.

Exemplary A-domain sequences and consensus sequences are depicted in FIGS. 2, 3 and 8. FIG. 9 displays location and occurrence of residues in A domains with the following spacing between cysteines. In addition, FIG. 10 depicts a number of A domains and provides a listing of conserved amino acids. One typical consensus sequence useful to identify A domains is the following: C-V[VLMA]-X(m)-C-[DNH]-X(n)-[DENQHT]-C-X(m)-X(n)-[STADE]-[DEH]-[DE]-X(n)-C (SEQ ID NO: 200), where the residues in brackets indicate possible residues at one position. “X(n)” indicates number of residues. These residues can be any amino acid residue. Parentheticals containing two numbers refers to the range of amino acids that can occupy that position (e.g., “[DE]-X(n)-C” means that the amino acids DE are followed by 1, 2, 3, 4, or 5 residues, followed by C). This consensus sequence only represents the portion of the A domain beginning at the third cysteine. A second consensus is as follows: C-X(n)-X(n)-C-X(n)-X(n)-X(n)-[ND]-X(n)-[DE]-[DE]-X(n)-[ND]-X(n)-C (SEQ ID NO: 201). The second consensus predicts amino acid residues spanning all six cysteine residues. In some embodiments, A domain variants comprise sequences substantially identical to any of the above-described sequences.

Additional exemplary A domains include the following sequence:

```
C_x3.13C_yX2.15C_x0.7C_x0.7C_yD(D,N)X_cX_yC_yDDEX_2
```

wherein C is cysteine, X represents between n and m number of independently selected amino acids, and (D,N) indicates that the position can be either D or N; and wherein C_yC_yC_yC_yC_yC_y form disulfide bonds.

In some embodiments, the monomer domain is an LDL receptor class A domain monomer comprising the following sequence:

```
CaX_n0.7C_yN4.5C_yX_5C_yD_DC_yC_yC_yC_yC_y
```

wherein X is defined as follows:

The table above indicates alternative amino acid residues at each position of the LDL receptor class A monomer
domain. For example, there can be either 6 or 7 amino acids between cystein C1 and cystein C2. The upper left box of the table indicates alternative amino acid residues at each position if there are 6 amino acids between C1 and C2. The bottom left box in the table indicates alternative amino acid residues of there are seven amino acids between C1 and C2. In all cases, the amino acid for one position (e.g., X1) is selected independently of the amino acids selected for remaining positions (e.g., X2, X3, etc.)

Exemplary proteins containing A-domains include, e.g., complement components (e.g., C6, C7, C8, C9, and Factor I), serine proteases (e.g., enteropeptidase, matriptase, and corin), transmembrane proteins (e.g., ST7, LRPS, LRPS5, and LRPS6) and endocytic receptors (e.g., Sortilin-related receptor, LDL-receptor, VLDLR, LRPI, LRPS2, and ApoER2). A domains and A domain variants can be readily employed in the practice of the present invention as monomer domains and variants thereof. Further description of A domains can be found in the following publications and references cited therein: Howell and Hertz, The LDL receptor gene family: signaling functions during development, (2001) Current Opinion in Neurobiology 11:74-81; Herz (2001), supra; Krieger, The “best” of cholesterol, the “worst” of cholesterol: A tale of two receptors, (1998) PNAS 95: 4077-4080; Goldstein and Brown, The Cholesterol Quatet, (2001) Science, 292: 1310-1312; and, Moestrup and Verrout, Megalin-and Cubilin-Mediated Endocytosis of Protein-Bound Vitamins, Lipids, and Hormones in Polarized Epithelia, (2001) Ann. Rev. Nutr. 21:407-28.

Other examples of monomer domains can be found in the protein Cubilin, which contains EGF-type repeats and CUB domains. The CUB domains are involved in ligand binding, e.g., some ligands include intrinsic factor (IF)-vitamin B12, receptor associated protein (RAP), Apo A-I, Transferrin, Albumin, Ig light chains and calcium. See, Moestrup and Verrout, supra.


Exemplary EGF monomer domains include the sequence:

C_nX_m-C_yC_w-C_zC_w,C_{n+1}C_{w+t}C_{w+t+1}C_{w+t+2}C_{t}

wherein C is cysteine, X_m represents between n and m number of independently selected amino acids; and

wherein C_nC_{w+t}C_{w+t+1}C_{w+t+2}C_{t} form disulfide bonds.

In some embodiments, the monomer domain is an EGF domain monomer comprising the following sequence:

C_nX_m-C_yC_w-C_zC_w,C_{n+1}C_{w+t}C_{w+t+1}C_{w+t+2}C_{t}

wherein X is defined as follows:

EGF monomer domains are sometimes between 25-45 amino acids and typically 30-39 amino acids.

Polynucleotides (also referred to as nucleic acids) encoding the monomer domains are typically employed to make monomer domains via expression. Nucleic acids that encode monomer domains can be derived from a variety of different sources. Libraries of monomer domains can be prepared by expressing a plurality of different nucleic acids encoding naturally occurring monomer domains, altered monomer domains (i.e., monomer domain variants), or a combinations thereof. For example, libraries may be designed in which a scaffold of amino acids remain constant (e.g., an LDL A receptor domain, EGF domain) while the intervening amino acids in the scaffold comprise randomly generated amino acids.

The invention provides methods of identifying monomer domains that bind to a selected or desired ligand or mixture of ligands. In some embodiments, monomer domains are identified or selected for a desired property (e.g., binding affinity) and then the monomer domains are formed into multimers. See, e.g., FIG. 5. For those embodiments, any method resulting in selection of domains with a desired property (e.g., a specific binding property) can be used. For example, the methods can comprise providing a plurality of different nucleic acids, each nucleic acid encoding a monomer domain; translating the plurality of different nucleic acids, thereby providing a plurality of different monomer domains; screening the plurality of different monomer domains for binding of the desired ligand or a mixture of ligands; and, identifying members of the plurality of different monomer domains that bind the desired ligand or mixture of ligands.

As mentioned above, monomer domains can be naturally-occurring or altered (non-natural variants). The term “naturally occurring” is used herein to indicate that an object can be found in nature. For example, natural monomer domains can include human monomer domains or optionally, domains derived from different species or sources, e.g., mammals, primates, rodents, fish, birds, reptiles, plants, etc. The natural occurring monomer domains can be obtained by a number of methods, e.g., by PCR amplification of genomic DNA or cDNA.

Monomer domains of the present invention can be naturally-occurring domains or non-naturally occurring variants. Libraries of monomer domains employed in the practice of the present invention may contain naturally-occurring monomer domain, non-naturally occurring monomer domain variants, or a combination thereof.

Monomer domain variants can include ancestral domains, chimeric domains, randomized domains, mutated domains, and the like. For example, ancestral domains can be based on phylogenetic analysis. Chimeric domains are domains in which one or more regions are replaced by corresponding regions from other domains of the same family. For example, chimeric domains can be constructed by combining loop sequences from multiple related domains.
of the same family to form novel domains with potentially lowered immunogenicity. Those of skill in the art will recognized the immunologic benefit of constructing modified binding domain monomers by combining loop regions from various related domains of the same family rather than creating random amino acid sequences. For example, by constructing variant domains by combining loop sequences or even multiple loop sequences that occur naturally in human LDL-receptor class A-domains, the resulting domains may contain novel binding properties but may not contain any immunogenic protein sequences because all of the exposed loops are of human origin. The combining of loop amino acid sequences in endogenous context can be applied to all of the monomer constructs of the invention. Thus the present invention provides a method for generating a library of chimeric monomer domains derived from human proteins, the method comprising: providing loop sequences corresponding to at least one loop from each of at least two different naturally occurring variants of a human protein, wherein the loop sequences are polynucleotide or polypeptide sequences; and covalently combining loop sequences to generate a library of at least two different chimeric sequences, wherein each chimeric sequence encodes a chimeric monomer domain having at least two loops. Typically, the chimeric domain has at least four loops, and usually at least six loops. As described above, the present invention provides three types of loops that are identified by specific features, such as, potential for disulfide bonding, bridging between secondary protein structures, and molecular dynamics (i.e., flexibility). The three types of loop sequences are a cysteine-defined loop sequence, a structure-defined loop sequence, and a B-factor-defined loop sequence.

[0139] Randomized domains are domains in which one or more regions are randomized. The randomization can be based on full randomization, or optionally, partial randomization based on natural distribution of sequence diversity.

[0140] Compositions of nucleic acids and polypeptides are included in the present invention. For example, the present invention provides a plurality of different nucleic acids wherein each nucleic acid encodes at least one monomer domain or immuno-domain.

[0141] The present invention also provides recombinant nucleic acids encoding one or more polypeptides comprising a plurality of monomer domains and/or immuno-domains, which monomer domains are altered in order or sequence as compared to a naturally occurring polypeptide. For example, the naturally occurring polypeptide can be selected from the group consisting of: an EGF-like domain, a Kringle-domain, a fibronectin type I domain, a fibronectin type II domain, a fibronectin type III domain, a PAN domain, a Gla domain, a SRCR domain, a Kunitz/Bovine pancreatic trypsin Inhibitor domain, a Kazal-type serine protease inhibitor domain, a Trefoil (P-type) domain, a von Willebrand factor type C domain, an Anaphylatoxin-like domain, a CUB domain, a thyroglobulin type I repeat, LDL-receptor class A domain, a Sushi domain, a Link domain, a Thrombospondin type I domain, an Immunoglobulin-like domain, a C-type lectin domain, a MAM domain, a von Willebrand factor type A domain, a Somatomedin B domain, a WAP-type four disulfide core domain, a F5/8 type C domain, a Hemopexin domain, an SH2 domain, an SH3 domain, a Laminin-type EGF-like domain, a C2 domain and variants of one or more thereof. In another embodiment, the naturally occurring polypeptide encodes a monomer domain found in the Pfam database and/or the SMART database.

[0142] All the compositions of the present invention, including the compositions produced by the methods of the present invention, e.g., monomer domains and/or immuno-domains, as well as multimers and libraries thereof can be optionally bound to a matrix of an affinity material. Examples of affinity material include beads, a column, a solid support, a microarray, other pools of reagent-supports, and the like.

III. Multimers

[0143] Methods for generating multimers are a feature of the present invention. Multimers comprise at least two monomer domains. For example, multimers of the invention can comprise from 2 to about 10 monomer domains, from 2 and about 8 monomer domains, from about 3 and about 10 monomer domains, about 7 monomer domains, about 6 monomer domains, about 5 monomer domains, or about 4 monomer domains. In some embodiments, the multimer comprises 3 or at least 3 monomer domains. In some embodiments, the multimers have no more than 2, 3, 4, 5, 6, 7, or 8 monomer domains. In view of the possible range of monomer domain sizes, the multimers of the invention may be, e.g., less than 100 kD, less than 90 kD, less than 80 kD, less than 70 kD, less than 60 kD, less than 50 kD, less than 40 kD, less than 30 kD, less than 25 kD, less than 20 kD, less than 15 kD, less than 10 kD or may be smaller or larger. In some cases, the monomer domains have been pre-selected for binding to the target molecule of interest (e.g., Met).

[0144] In some embodiments, each monomer domain specifically binds to one target molecule (e.g., Met). In some of these embodiments, each monomer binds to a different position (analogous to an epitope) on a target molecule. Multiple monomer domains that bind to the same target molecule results in an avidity effect resulting in improved avidity of the multimer for the target molecule compared to each individual monomer. In some embodiments, the multimer has an avidity of at least about 1.5, 2, 3, 4, 5, 10, 20, 50 or 100 times the avidity of a monomer domain alone. In some embodiments, at least one, two, three, four or more (e.g., all) monomers of a multimer bind an ion such as calcium or another ion. Multimers can comprise a variety of combinations of monomer domains. For example, in a single multimer, the selected monomer domains can be identical or different. In addition, the selected monomer domains can comprise various different monomer domains from the same monomer domain family, or various monomer domains from different domain families, or optionally, a combination of both.

[0145] Multimers that are generated in the practice of the present invention may be any of the following:

[0146] (1) A homo-multimer (a multimer of the same domain, i.e., A1-A1-A1-A1);

[0147] (2) A hetero-multimer of different domains of the same domain class, e.g., A1-A2-A3-A4. For example, hetero-multimer include multimers where A1, A2, A3 and A4 are different non-naturally occurring variants of a particular LDL-receptor class A domains, or where some of A1, A2, A3, and A4 are naturally-occurring variants of a LDL-receptor class A domain (see, e.g., FIG. 10).
(0148) A hetero-multimer of domains from different monomer domain classes, e.g., A1-B2-A2-B1. For example, where A1 and A2 are two different monomer domains (either naturally occurring or non-naturally occurring) from LDL-receptor class A, and B1 and B2 are two different monomer domains (either naturally occurring or non-naturally occurring) from class EGF-like domain.

(0149) In another embodiment, the multimer comprises monomer domains with specificities for different target molecules. For example, in some embodiments, the multimers of the invention comprises 1, 2, 3, or more monomer domains that bind to Met and at least one monomer domain that binds to a second target molecule. Exemplary target molecules include, e.g., a serum molecule that extends the serum half-life of the multimer, EGFR gene family members, VEGF receptors, PDGF receptor, other receptor tyrosine kinases, integrins, other molecules implicated in tumorigenesis, or markers of tumor tissue. Exemplary molecules that extend the serum half-life of a multimer include, e.g., red blood cells, IgG, and serum albumin such as HSA.

(0150) Multimer libraries employed in the practice of the present invention may contain homo-multimers, hetero-multimers of different monomer domains (natural or non-natural) of the same monomer class, or hetero-multimers of monomer domains (natural or non-natural) from different monomer classes, or combinations thereof.

(0151) Monomer domains, as described herein, are also readily employed in a immuno-domain-containing heteromultimer (i.e., a multimer that has at least one immuno-domain variant and one monomer domain variant). Thus, multimers of the present invention may have at least one immuno-domain such as a minobody, a single domain antibody, a single chain variable fragment (ScFv), or a Fab fragment; and at least one monomer domain, such as, for example, an EGF-like domain, a Kringle-domain, a fibronectin type I domain, a fibronectin type II domain, a fibronectin type III domain, a PAN domain, a Gla domain, a SRCR domain, a Kunitz/Bovine pancreatic trypsin inhibitor domain, a Kazal-type serine protease inhibitor domain, a Trefoil (P-type) domain, a von Willebrand factor type C domain, an Anaphylatoxin-like domain, a CUB domain, a thyroglobulin type I repeat, LDL-receptor class A domain, a Sushi domain, a Link domain, a Thrombospondin type I domain, an Immunoglobulin-like domain, a C-type lectin domain, a MAM domain, a von Willebrand factor type A domain, a Somatomedin B domain, a WAP-type four disulfide core domain, a F5/F8 type C domain, a Hemopexin domain, an SH2 domain, an SH3 domain, a Laminin-type EGF-like domain, a C2 domain, or variants thereof.

(0152) Domains need not be selected before the domains are linked to form multimers. On the other hand, the domains can be selected for the ability to bind to a target molecule before being linked into multimers. Thus, for example, a multimer can comprise two domains that bind to one target molecule and a third domain that binds to a second target molecule.

(0153) The multimers of the present invention may have the following qualities: multivalent, multispecific, single chain, heat stable, extended serum and/or shelf half-life. Moreover, at least one, more than one or all of the monomer domains may bind an ion (e.g., a metal ion or a calcium ion), at least one, more than one or all monomer domains may be derived from LDL receptor A domains and/or EGF-like domains, at least one, more than one or all of the monomer domains may be non-naturally occurring, and/or at least one, more than one or all of the monomer domains may comprise 1, 2, 3, or 4 disulfide bonds per monomer domain. In some embodiments, the multimers comprise at least two (or at least three) monomer domains, wherein at least one monomer domain is a non-naturally occurring monomer domain and the monomer domains bind calcium. In some embodiments, the multimers comprise at least 4 monomer domains, wherein at least one monomer domain is non-naturally occurring, and wherein:

(0154) a. each monomer domain is between 30-100 amino acids and each of the monomer domains comprise at least one disulfide linkage; or

(0155) b. each monomer domain is between 30-100 amino acids and is derived from an extracellular protein; or

(0156) c. each monomer domain is between 30-100 amino acids and binds to a protein target.

(0157) In some embodiments, the multimers comprise at least 4 monomer domains, wherein at least one monomer domain is non-naturally occurring, and wherein:

(0158) a. each monomer domain is between 35-100 amino acids; or

(0159) b. each domain comprises at least one disulfide bond and is derived from a human protein and/or an extracellular protein.

(0160) In some embodiments, the multimers comprise at least two monomer domains, wherein at least one monomer domain is non-naturally occurring, and wherein each domain is:

(0161) a. 25-50 amino acids long and comprises at least one disulfide bond; or

(0162) b. 25-50 amino acids long and is derived from an extracellular protein; or

(0163) c. 25-50 amino acids and binds to a protein target; or

(0164) d. 35-50 amino acids long.

(0165) In some embodiments, the multimers comprise at least two monomer domains, wherein at least one monomer domain is non-naturally-occurring and:

(0166) a. each monomer domain comprises at least one disulfide bond; or

(0167) b. at least one monomer domain is derived from an extracellular protein; or

(0168) c. at least one monomer domain binds to a target protein.

(0169) A significant advantage of the present invention is that known ligands, or unknown ligands can be used to select the monomer domains and/or multimers. No prior information regarding ligand structure is required to isolate the monomer domains of interest or the multimers of interest. The monomer domains and/or multimers identified can have biological activity, which is meant to include at least specific binding affinity for a selected or desired ligand, and,
in some instances, will further include the ability to block the binding of other compounds, to stimulate or inhibit metabolic pathways, to act as a signal or messenger, to stimulate or inhibit cellular activity, and the like. Monomer domains can be generated to function as ligands for receptors where the natural ligand for the receptor has not yet been identified (orphan receptors). These orphan ligands can be created to either block or activate the receptor top which they bind.

A single ligand can be used, or optionally a variety of ligands can be used to select the monomer domains and/or multimers. A monomer domain of the present invention can bind a single ligand or a variety of ligands. A multimer of the present invention can have multiple discrete binding sites for a single ligand, or optionally, can have multiple binding sites for a variety of ligands.

In some embodiments, the multimer comprises monomer domains with specificities for different proteins. The different proteins can be related or unrelated. Examples of related proteins including members of a protein family or different serotypes of a virus. Alternatively, the monomer domains of a multimer can target different molecules in a physiological pathway (e.g., different blood coagulation proteins). In yet other embodiments, monomer domains bind to proteins in unrelated pathways (e.g., two domains bind to blood factors, two other domains bind to inflammation-related proteins and a fifth binds to serum albumin). In another embodiment, a multimer is comprised of monomer domains that bind to different pathogens or contaminants of interest. Such multimers are useful as a single detection agent capable of detecting for the possibility of any of a number of pathogens or contaminants.

In some embodiments, the multimers of the invention bind to the same or other multimers to form aggregates. Aggregation can be mediated, for example, by the presence of hydrophobic domains on two monomer domains, resulting in the formation of non-covalent interactions between two monomer domains. Alternatively, aggregation may be facilitated by one or more monomer domains in a multimer having binding specificity for a monomer domain in another multimer. Aggregates can also form due to the presence of affinity peptides on the monomer domains or multimers. Aggregates can contain more target molecule binding domains than a single multimer.

Multimers with affinity for both a cell surface target and a second target may provide for increased avidity effects. In some cases, membrane fluidity can be more flexible than protein linkers in optimizing (by self-assembly) the spacing and valency of the interactions. In some cases, multimers will bind to two different targets, each on a different cell or one on a cell and another on a molecule with multiple binding sites. See, e.g., FIGS. 16 and 17.

In some embodiments, the monomers or multimers of the present invention are linked to another polypeptide to form a fusion protein. Any polypeptide in the art may be used as a fusion partner, though it can be useful if the fusion partner forms multimers. For example, monomers or multimers of the invention may, for example, be fused to the following locations or combinations of locations of an antibody:

1. At the N-terminus of the VH1 and/or VL1 domains, optionally just after the leader peptide and before the domain starts (framework region 1);
2. At the N-terminus of the CH1 or CL1 domain, replacing the VH1 or VL1 domain;
3. At the N-terminus of the heavy chain, optionally after the CH1 domain and before the cysteine residues in the hinge (Fc-fusion);
4. At the N-terminus of the CH3 domain;
5. At the C-terminus of the CH3 domain, optionally attached to the last amino acid residue via a short linker;
6. At the C-terminus of the CH2 domain, replacing the CH3 domain; and/or
7. At the C-terminus of the CL1 or CH1 domain, optionally after the cysteine that forms the interchain disulfide; or
8. At the C-terminus of the VH1 or VL1 domain. See, e.g., FIG. 18.

In some embodiments, the monomer or multimer domain is linked to a molecule (e.g., a protein, nucleic acid, organic small molecule, etc.) useful as a pharmaceutical. Exemplary pharmaceutical proteins include, e.g., cytokines, antibodies, chemokines, growth factors, interleukins, cell-surface proteins, extracellular domains, cell surface receptors, cytokotoxins, etc. Exemplary small molecule pharmaceuticals include small molecule toxins or therapeutic agents.

In some embodiments, the monomer or multimers are selected to bind to a tissue- or disease-specific target protein. Tissue-specific proteins are proteins that are expressed exclusively, or at a significantly higher level, in one or several particular tissue(s) compared to other tissues in an animal. As c-Met is expressed at significant levels in the liver, monomer domains that bind to Met may be used to target other molecules, including other monomer domains, to the liver. This may be used to target liver-specific diseases, for example, by targeting therapeutic or toxic molecules to the liver. An example of a liver disease that can be treated is hepatocellular carcinoma. Similarly, disease-specific proteins are proteins that are expressed exclusively, or at a significantly higher level, in one or several diseased cells or tissues compared to other non-diseased cells or tissues in an animal.

In some embodiments, the monomers or multimers that bind to the target protein are linked to the pharmaceutical protein or small molecule such that the resulting complex or fusion is targeted to the specific tissue or disease-related cell(s) where the target protein (e.g., Met) is expressed. Monomers or multimers for use in such complexes or fusions can be initially selected for binding to the target protein and may be subsequently selected by negative selection against other cells or tissue (e.g., to avoid targeting bone marrow or other tissues that set the lower limit of drug toxicity) where it is desired that binding be reduced or eliminated in other non-target cells or tissues. By keeping the pharmaceutical away from sensitive tissues, the therapeutic window is increased so that a higher dose may be administered safely. In another alternative, in vivo panning can be performed in animals by injecting a library of monomers or multimers into an animal and then isolating the monomers or multimers that bind to a particular tissue or cell of interest.
The fusion proteins described above may also include a linker peptide between the pharmaceutical protein and the monomer or multimers. A peptide linker sequence may be employed to separate, for example, the polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Fusion proteins can also be expressed as recombinant proteins in an expression system by standard techniques.

Multimers or monomer domains of the invention can be produced according to any methods known in the art. In some embodiments, E. coli comprising a pET-derived plasmid encoding the polypeptides are induced to express the protein. After harvesting the bacteria, they may be lysed and clarified by centrifugation. The polypeptides may be purified using Ni-NTA agarose elution and refolded by dialysis. Misfolded proteins may be neutralized by capping free sulfhydryls with iodoacetic acid. Q sepharose elution, butyl sepharose FF, SP sepharose elution, Q sepharose elution, or SP sepharose elution may be used to purify the polypeptides.

IV. Linkers

Monomer domains can be joined by a linker to form a multimer. For example, a linker may be positioned between each separate discrete monomer domain in a multimer.

Joining the selected monomer domains via a linker can be accomplished using a variety of techniques known in the art. For example, combinatorial assembly of polynucleotides encoding selected monomer domains can be achieved by restriction digestion and re-ligation, by PCR-based, self-priming overlap reactions, or other recombinant methods. The linker can be attached to a monomer before the monomer is identified for its ability to bind to a target multimer or after the monomer has been selected for the ability to bind to a target multimer.

The linker can be naturally-occurring, synthetic or a combination of both. For example, the synthetic linker can be a randomized linker, e.g., both in sequence and size. In one aspect, the randomized linker can comprise a fully randomized sequence, or optionally, the randomized linker can be based on natural linker sequences. The linker can comprise, e.g., a non-polypeptide moiety, a polynucleotide, a polypeptide or the like.

A linker can be rigid, or flexible, or a combination of both. Linker flexibility can be a function of the composition of both the linker and the monomer domains that the linker interacts with. The linker joins two selected monomer domains, and maintains the monomer domains as separate discrete monomer domains. The linker can allow the separate discrete monomer domains to cooperate yet maintain separate properties such as multiple separate binding sites for the same ligand in a multimer, or e.g., multiple separate binding sites for different ligands in a multimer.

Choosing a suitable linker for a specific case where two or more monomer domains (i.e., polypeptide chains) are to be connected may depend on a variety of parameters including, e.g. the nature of the monomer domains, the structure and nature of the target to which the polypeptide multimer should bind and/or the stability of the peptide linker towards proteolysis and oxidation.

The present invention provides methods for optimizing the choice of linker once the desired monomer domains/variants have been identified. Generally, libraries of multimers having a composition that is fixed with regard to monomer domain composition, but variable in linker composition and length, can be readily prepared and screened as described above.

Typically, the linker polypeptide may predominantly include amino acid residues selected from the group consisting of Gly, Ser, Ala and Thr. For example, the peptide linker may contain at least 75% (calculated on the basis of the total number of residues present in the peptide linker), such as at least 80%, e.g. at least 85% or at least 90% of amino acid residues selected from the group consisting of Gly, Ser, Ala and Thr. The peptide linker may also consist of Gly, Ser, Ala and/or Thr residues only. The linker polypeptide should have a length, which is adequate to link two monomer domains in such a way that they assume the correct conformation relative to one another so that they retain the desired activity, for example as antagonists of a given receptor.

A suitable length for this purpose is a length of at least one and typically fewer than about 50 amino acid residues, such as 2-25 amino acid residues, 5-20 amino acid residues, 5-15 amino acid residues, 8-12 amino acid residues or 11 residues. Similarly, the polypeptide encoding a linker can range in size, e.g., from about 2 to about 15 amino acids, from about 3 to about 15, from about 4 to about 12, about 10, about 8, or about 6 amino acids. In methods and compositions involving nucleic acids, such as DNA, RNA, or combinations of both, the polynucleotide containing the linker sequence can be, e.g., between about 6 nucleotides and about 45 nucleotides, between about 9 nucleotides and about 45 nucleotides, between about 12 nucleotides and about 36 nucleotides, about 30 nucleotides, about 24 nucleotides, or about 18 nucleotides. Likewise, the amino acid residues selected for inclusion in the linker polypeptide should exhibit properties that do not interfere significantly with the activity or function of the polypeptide multimer. Thus, the polypeptide linker should on the whole not exhibit a charge which would be inconsistent with the activity or function of the polypeptide multimer, or interfere with internal folding, or form bonds or other interactions with amino acid residues in one or more of the monomer domains which would seriously impede the binding of the polypeptide multimer to the target in question.

In another embodiment of the invention, the peptide linker is selected from a library where the amino acid residues in the peptide linker are randomized for a specific set of monomer domains in a particular polypeptide multimer. A flexible linker could be used to find suitable combinations of monomer domains, which is then optimized using this random library of variable linkers to obtain linkers with optimal length and geometry. The optimal linkers may contain the minimal number of amino acid residues of the right type that participate in the binding to the target and restrict the movement of the monomer domains relative to each other in the polypeptide multimer when not bound to the target.

The use of naturally occurring as well as artificial peptide linkers to connect polypeptides into novel linked

[0198] One example where the use of peptide linkers is widespread is for production of single-chain antibodies where the variable regions of a light chain (V_L) and a heavy chain (V_H) are joined through an artificial linker, and a large number of publications exist within this particular field. A widely used peptide linker is a 15mer consisting of three repeats of a Gly-Gly-Gly-Ser (SEQ ID NO: 240) amino acid sequence ((Gly,Ser)). Other linkers have been used, and phage display technology, as well as, selective infective phage technology has been used to diversify and select appropriate linker sequences (Tang et al. (1996), J. Biol. Chem. 271, 15682-15686; Hennecke et al. (1998), Protein Eng. 11, 405-410). Peptide linkers have been used to connect individual chains in hetero- and homo-dimeric proteins such as the T-cell receptor, the lambda Cr repressor, the P22 phage Arc repressor, IL-12, TSH, FSH, IL-5, and interferon-γ. Peptide linkers have also been used to create fusion polypeptides. Various linkers have been used and in the case of the Arc repressor phage display has been used to optimize the linker length and composition for increased stability of the single-chain protein (Robinson and Sauer (1998), Proc. Natl. Acad. Sci. USA 95, 5929-5934).

[0199] Another type of linker is an intein, i.e. a peptide stretch which is expressed with the single polypeptide, but removed post-translationally by protein splicing. The use of inteins is reviewed by F. S. Gimble in Chemistry and Biology, 1998, Vol 5, No. 10 pp 251-256.

[0200] Still another way of obtaining a suitable linker is by optimizing a simple linker, e.g. (Gly,Ser)_n (SEQ ID NO: 240), through random mutagenesis.

[0201] As mentioned above, it is generally preferred that the peptide linker possess at least some flexibility. Accordingly, in some embodiments, the peptide linker contains 1-25 glycine residues, 5-20 glycine residues, 5-15 glycine residues or 8-12 glycine residues. The peptide linker will typically contain at least 50% glycine residues, such as at least 75% glycine residues. In some embodiments of the invention, the peptide linker comprises glycine residues only.

[0202] The peptide linker may, in addition to the glycine residues, comprise other residues, in particular residues selected from the group consisting of Ser, Ala and Thr, in particular Ser. Thus, one example of a specific peptide linker includes a peptide linker having the amino acid sequence Gly-Xaa-Gly-Yaa-Xaa-Gly (SEQ ID NO: 203), wherein each Xaa is independently selected from the group consisting Ala, Val, Leu, Ile, Met, Phe, Trp, Gly, Ser, Thr, Cys, Tyr, Asn, Gin, Lys, Arg, His, Asp and Glu, and wherein x and y are each integers in the range from 1-5. In some embodiments, each Xaa is independently selected from the group consisting of Ser, Ala, and Thr, in particular Ser. More particularly, the peptide linker has the amino acid sequence Gly-Gly-Gly-Xaa-Gly-Gly-Gly-Gly-Xaa-Gly-Gly-Gly (SEQ ID NO: 204), wherein each Xaa is independently selected from the group consisting Ala, Val, Leu, Ile, Met, Phe, Trp, Pro, Gly, Ser, Thr, Cys, Tyr, Asn, Gin, Lys, Arg, His, Asp and Glu. In some embodiments, each Xaa is independently selected from the group consisting of Ser, Ala and Thr, in particular Ser.

[0203] In some cases it may be desirable or necessary to provide some rigidity into the peptide linker. This may be accomplished by including proline residues in the amino acid sequence of the peptide linker. Thus, in another embodiment of the invention, the peptide linker comprises at least one proline residue in the amino acid sequence of the peptide linker. For example, the peptide linker has an amino acid sequence, wherein at least 25%, such as at least 50%, e.g. at least 75%, of the amino acid residues are proline residues. In one particular embodiment of the invention, the peptide linker comprises proline residues only.

[0204] In some embodiments of the invention, the peptide linker is modified in such a way that an amino acid residue comprising an attachment group for a non-polypeptide moiety is introduced. Examples of such amino acid residues may be a cysteine residue (to which the non-polypeptide moiety is then subsequently attached) or the amino acid sequence may include an in vivo N-glycosylation site (thereby attaching a sugar moiety (in vivo) to the peptide linker).

[0205] In some embodiments of the invention, the peptide linker comprises at least one cysteine residue, such as one cysteine residue. Thus, in some embodiments of the invention the peptide linker comprises amino acid residues selected from the group consisting of Gly, Ser, Ala, Thr and Cys. In some embodiments, such a peptide linker comprises one cysteine residue only.

[0206] In a further embodiment, the peptide linker comprises glycine residues and cysteine residue, such as glycine residues and cysteine residues only. Typically, only one cysteine residue will be included per peptide linker. Thus, one example of a specific peptide linker comprising a cysteine residue, includes a peptide linker having the amino acid sequence Gly-Glu-Cys-Gly (SEQ ID NO: 205), wherein n and m are each integers from 1-12, e.g., from 3-9, from 4-8, or from 4-7. More particularly, the peptide linker may have the amino acid sequence GGGGG-C-GGGGG (SEQ ID NO: 206).

[0207] This approach (i.e. introduction of an amino acid residue comprising an attachment group for a non-polypeptide moiety) may also be used for the more rigid proline-containing linkers. Accordingly, the peptide linker may comprise proline and cysteine residues, such as proline and cysteine residues only. An example of a specific proline-containing peptide linker comprising a cysteine residue, includes a peptide linker having the amino acid sequence Pro-Cys-Pro (SEQ ID NO: 207), wherein n and m are each integers from 1-12, preferably from 3-9, such as from 4-8 or from 4-7. More particularly, the peptide linker may have the amino acid sequence PPPPP—C—PPP (SEQ ID NO: 208).

[0208] In some embodiments, the purpose of introducing an amino acid residue, such as a cysteine residue, comprising an attachment group for a non-polypeptide moiety is to subsequently attach a non-polypeptide moiety to said residue. For example, non-polypeptide moieties can improve the serum half-life of the polypeptide multimer. Thus, the cysteine residue can be covalently attached to a non-polypep-
tide moiety. Preferred examples of non-polypeptide moieties include polymer molecules, such as PEG or mPEG, in particular mPEG as well as non-polypeptide therapeutic agents.

[0209] The skilled person will acknowledge that amino acid residues other than cysteine may be used for attaching a non-polypeptide to the peptide linker. One particular example of such other residue includes coupling the non-polypeptide moiety to a lysine residue.

[0210] Another possibility of introducing a site-specific attachment group for a non-polypeptide moiety in the peptide linker is to introduce an in vivo N-glycosylation site, such as one in vivo N-glycosylation site, in the peptide linker. For example, an in vivo N-glycosylation site may be introduced in a peptide linker comprising amino acid residues selected from the group consisting of Gly, Ser, Ala and Thr. It will be understood that in order to ensure that a sugar moiety is in fact attached to said in vivo N-glycosylation site, the nucleotide sequence encoding the polypeptide multimer must be inserted in a glycosylating, eukaryotic expression host.

[0211] A specific example of a peptide linker comprising an in vivo N-glycosylation site is a peptide linker having the amino acid sequence Gly-Gys-Asn-Xaa-Ser/Thr-Gly-m (SEQ ID NO: 209), preferably Gly-Gys-Asn-Xaa-Thr-Gly-m (SEQ ID NO: 210), wherein Xaa is any amino acid residue except proline, and wherein n and m are each integers in the range from 1-8, preferably in the range from 2-5.

[0212] Often, the amino acid sequences of all peptide linkers present in the polypeptide multimer will be identical. Nevertheless, in certain embodiments the amino acid sequences of all peptide linkers present in the polypeptide multimer may be different. The latter is believed to be particular relevant in case the polypeptide multimer is a polypeptide tri-mer or tetra-mer and particularly in such cases where an amino acid residue comprising an attachment group for a non-polypeptide moiety is included in the peptide linker.

[0213] Quite often, it will be desirable or necessary to attach only a few, typically only one, non-polypeptide moieties/moiety (such as mPEG, a sugar moiety or a non-polypeptide therapeutic agent) to the polypeptide multimer in order to achieve the desired effect, such as prolonged serum-half life. Evidently, in case of a polypeptide tri-mer, which will contain two peptide linkers, only one peptide linker is typically required to be modified, e.g. by introduction of a cysteine residue, whereas modification of the other peptide linker will typically not be necessary at. In this case all (both) peptide linkers of the polypeptide multimer (tri-mer) are different.

[0214] Accordingly, in a further embodiment of the invention, the amino acid sequences of all peptide linkers present in the polypeptide multimer are identical except for one, two or three peptide linkers, such as except for one or two peptide linkers, in particular except for one peptide linker, which has/hae an amino acid sequence comprising an amino acid residue comprising an attachment group for a non-polypeptide moiety. Preferred examples of such amino acid residues include cysteine residues of in vivo N-glycosylation sites.

[0215] A linker can be a native or synthetic linker sequence. An exemplary native linker includes, e.g., the sequence between the last cysteine of a first LDL receptor A domain and the first cysteine of a second LDL receptor A domain can be used as a linker sequence. Analysis of various A domain linkages reveals that native linkers range from at least 3 amino acids to fewer than 20 amino acids, e.g., 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18 amino acids long. However, those of skill in the art will recognize that longer or shorter linker sequences can be used. An exemplary A domain linker sequence is depicted in FIG. 8. In some embodiments, the linker is a 6-mer of the following sequence A1A2A3A4A5A6 (SEQ ID NO: 244), wherein A1 is selected from the amino acids A, P, T, Q, E and K; A2 and A3 are any amino acid except C, F, Y, W, or M; A4 is selected from the amino acids S, G and R; A5 is selected from the amino acids H, P, and R; and A6 is the amino acid, T.

[0216] Methods for generating multimers from monomer domains can include joining the selected domains with at least one linker to generate at least one multimer, e.g., the multimer can comprise at least two of the monomer domains and the linker. The multimer(s) is then screened for an improved avidity or affinity or altered specificity for the desired ligand or mixture of ligands as compared to the selected monomer domains. A composition of the multimer produced by the method is included in the present invention.

[0217] In other methods, the selected multimer domains are joined with at least one linker to generate at least two multimers, wherein the two multimers comprise two or more of the selected monomer domains and the linker. The two or more multimers are screened for an improved avidity or affinity or altered specificity for the desired ligand or mixture of ligands as compared to the selected monomer domains. Compositions of two or more multimers produced by the above method are also features of the invention.

[0218] Typically, multimers of the present invention are a single discrete polypeptide. Multimers of partial linker-domain-partial linker moieties are an association of multiple polypeptides, each corresponding to a partial linker-domain-partial linker moiety.

[0219] Suitable linkers employed in the practice of the present invention include an obligate heterodimer of partial linker moieties. The term “obligate heterodimer” (also referred to as “affinity peptides”) refers herein to a dimer of two partial linker moieties that differ from each other in composition, and which associate with each other in a non-covalent, specific manner to join two domains together. The specific association is such that the two partial linkers associate substantially with each other as compared to associating with other partial linkers. Thus, in contrast to multimers of the present invention that are expressed as a single polypeptide, multimers of domains that are linked together via heterodimers are assembled from discrete partial linker-monomer-partial linker units. Assembly of the heterodimers can be achieved by, for example, mixing. Thus, if the partial linkers are polypeptide segments, each partial linker-monomer-partial linker unit may be expressed as a discrete peptide prior to multimer assembly. A disulfide bond can be added to covalently lock the peptides together following the correct non-covalent pairing. A multimer containing such obligate heterodimers is depicted in FIG. 11. Partial linker moieties that are appropriate for forming obligate heterodimers include, for example, polynucleotides, polypeptides, and the like. For example, when the
When the partial linker comprises a DNA binding motif, each monomer domain has an upstream and a downstream partial linker (i.e., Lp-domain-Lp, where “Lp” is a representation of a partial linker) that contains a DNA binding protein with exclusively unique DNA binding specificity. These domains can be produced individually and then assembled into a specific multimer by the mixing of the domains with DNA fragments containing the proper nucleotide sequences (i.e., the specific recognition sites for the DNA binding proteins of the partial linkers of the two desired domains) so as to join the domains in the desired order. Additionally, the same domains may be assembled into many different multimers by the addition of DNA sequences containing various combinations of DNA binding protein recognition sites. Further randomization of the combinations of DNA binding protein recognition sites in the DNA fragments can allow the assembly of libraries of multimers. The DNA can be synthesized with backbone analogs to prevent degradation in vivo.

Suitable linkers employed in the practice of the present invention include an obligate heterodimer of partial linker moieties. The term “obligate heterodimer” (also referred to as “affinity peptides”) refers herein to a dimer of two partial linker moieties that differ from each other in composition, and which associate with each other in a non-covalent, specific manner to join two domains together. The specific association is such that the two partial linkers associate substantially with each other as compared to associating with other partial linkers. Thus, in contrast to multimers of the present invention that are expressed as a single polypeptide, multimers of domains that are linked together via heterodimers are assembled from discrete partial linker-monomer-partial linker units. Assembly of the heterodimers can be achieved by, for example, mixing. Thus, if the partial linkers are polypeptide segments, each partial linker-monomer-partial linker unit may be expressed as a discrete peptide prior to multimer assembly. A disulfide bond can be added to covalently lock the peptides together following the correct non-covalent pairing. A multimer containing such obligate heterodimers is depicted in FIG. 11. Partial linker moieties that are appropriate for forming obligate heterodimers include, for example, polynucleotides, polypeptides, and the like. For example, when the partial linker is a polypeptide, binding domains are produced individually along with their unique linking peptide (i.e., a partial linker) and later combined to form multimers. See, e.g., Madden, M., Aldwin, L., Gallop, M. A., and Stemmer, W. P. C. (1993) Peptide linkers: Unique self-associative high-affinity peptide linkers. Thirteenth American Peptide Symposium, Edmonton, Canada (abstract). The spatial order of the binding domains in the multimer is thus mandated by the heterodimeric binding specificity of each partial linker. Partial linkers can contain terminal amino acid sequences that specifically bind to a defined heterologous amino acid sequence. An example of such an amino acid sequence is the Hydra neuropetide head activator as described in Bodemuller et al., *The neuropetide head activator loses its biological activity by dimerization*, (1986) EMBO J. 5(8):1825-1829. See, e.g., U.S. Pat. No. 5,491,074 and WO 94/28173. These partial linkers allow the multimer to be produced first as monomer-partial linker units or partial linker-monomer-partial linker units that are then mixed together and allowed to assemble into the ideal order based on the binding specificities of each partial linker. Alternatively, monomers linked to partial linkers can be contacted to a surface, such as a cell, in which multiple monomers can associate to form higher avidity complexes via partial linkers. In some cases, the association will form via random Brownian motion.

When the partial linker comprises a DNA binding motif, each monomer domain has an upstream and a downstream partial linker (i.e., Lp-domain-Lp, where “Lp” is a representation of a partial linker) that contains a DNA binding protein with exclusively unique DNA binding specificity. These domains can be produced individually and then assembled into a specific multimer by the mixing of the domains with DNA fragments containing the proper nucleotide sequences (i.e., the specific recognition sites for the DNA binding proteins of the partial linkers of the two desired domains) so as to join the domains in the desired order. Additionally, the same domains may be assembled into many different multimers by the addition of DNA sequences containing various combinations of DNA binding protein recognition sites. Further randomization of the combinations of DNA binding protein recognition sites in the DNA fragments can allow the assembly of libraries of multimers. The DNA can be synthesized with backbone analogs to prevent degradation in vivo.

Those of skill in the art can readily identify monomer domains with a desired property (e.g., binding affinity). For those embodiments, any method resulting in selection of domains with a desired property (e.g., a specific binding property) can be used. For example, the methods can comprise providing a plurality of different nucleic acids, each nucleic acid encoding a monomer domain, translating the plurality of different nucleic acids, thereby providing a plurality of different monomer domains, screening the plurality of different monomer domains for binding of the desired ligand or a mixture of ligands; and, identifying members of the plurality of different monomer domains that bind the desired ligand or mixture of ligands.

In addition, any method of mutagenesis, such as site-directed mutagenesis and random mutagenesis (e.g.,
chemical mutagenesis) can be used to produce monomer domains, e.g., for a monomer domain library. In some embodiments, error-prone PCR is employed to create variants. Additional methods include aligning a plurality of naturally occurring monomer domains by aligning conserved amino acids in the plurality of naturally occurring monomer domains; and, designing the non-naturally occurring monomer domain by maintaining the conserved amino acids and inserting, deleting or altering amino acids around the conserved amino acids to generate the non-naturally occurring monomer domain. In one embodiment, the conserved amino acids comprise cysteines. In another embodiment, the inserting step uses random amino acids, or optionally, the inserting step uses portions of the naturally occurring monomer domains. The portions could ideally encode loops from domains from the same family. Amino acids are inserted or exchanged using synthetic oligonucleotides, or by shuffling, or by restriction enzyme-based recombination. Human chimeric domains of the present invention are useful for therapeutic applications where minimal immunogenicity is desired. The present invention provides methods for generating libraries of human chimeric domains. Human chimeric monomer domain libraries can be constructed by combining loop sequences from different variants of a human monomer domain, as described above. The loop sequences that are combined may be sequence-defined loops, structure-defined loops, B-factor-defined loops, or a combination of any two or more thereof.

[0225] Alternatively, a human chimeric domain library can be generated by modifying naturally-occurring human monomer domains at the amino acid level, as compared to the loop level. In some embodiments, to minimize the potential for immunogenicity, only those residues that naturally occur in protein sequences from the same family of human monomer domains are utilized to create the chimeric sequences. This can be achieved by providing a sequence alignment of at least two human monomer domains from the same family of monomer domains, identifying amino acid residues in corresponding positions in the human monomer domain sequences that differ between the human monomer domains, generating two or more human chimeric monomer domains, wherein each human chimeric monomer domain sequence consists of amino acid residues that correspond in type and position to residues from two or more human monomer domains from the same family of monomer domains. Libraries of human chimeric monomer domains can be employed to identify human chimeric monomer domains that bind to a target of interest by: screening the library of human chimeric monomer domains for binding to a target molecule, and identifying a human chimeric monomer domain that binds to the target molecule. Suitable naturally-occurring human monomer domain sequences employed in the initial sequence alignment step include those corresponding to any of the naturally-occurring monomer domains described herein.

[0226] Domains of human monomer variant libraries of the present invention (whether generated by varying loops or single amino acid residues) can be prepared by methods known to those having ordinary skill in the art. Methods particularly suitable for generating these libraries are split-pool format and trinucleotide synthesis format as described in WO01/23401.

[0227] In some embodiments, monomer domains of the invention are screened for potential immunogenicity by:

[0228] providing a candidate protein sequence;

[0229] comparing the candidate protein sequence to a database of human protein sequences;

[0230] identifying portions of the candidate protein sequence that correspond to portions of human protein sequences from the database; and

[0231] determining the extent of correspondence between the candidate protein sequence and the human protein sequences from the database.

[0232] In general, the greater the extent of correspondence between the candidate protein sequence and one or more of the human protein sequences from the database, the lower the potential for immunogenicity is predicted as compared to a candidate protein having little correspondence with any of the human protein sequences from the database. A database of human protein sequences that is suitable for use in the practice of the invention method for screening candidate proteins can be found at ncbi.nlm.nih.gov/BLAST/BLAST.cgi at the World Wide Web (in addition, the following web site can be used to search short, nearly exact matches: cbi.nlm.nih.gov/blast/BLAST.cgi?CMD=Web&LAYOUT=

TwoWindows&AUTO_FORMAT=

Semiauto&ALIGNMENTS=50&ALIGNMENT_VIEW=

Pairwise&CLIENT=web&DATABASE=

nr&DESCRIPTIONS=100&ENTREZ_QUERY=

(none)&EXPECT=1000&FORMAT_OBJECT=

Alignment&FORMAT_TYPE=HTML&NCBI_GI=

on&PAGE=Nucleotides&PROGRAM=blastn&SERVICE=

plain&SET_DEFAULTS.x=29&SET_DEFAULTS.y=

6&SHOW_OVERVIEW=on&WORD_SIZE=

7&END_OF_HTTPGET=YES&SHOW_LINKOUT=Yes at the World Wide Web). The method is particularly useful in determining whether a crossover sequence in a chimeric protein, such as, for example, a chimeric monomer domain, is likely to cause an immunogenic event. If the crossover sequence corresponds to a portion of a sequence found in the database of human protein sequences, it is believed that the crossover sequence is less likely to cause an immunogenic event. An example of the comparison step is depicted in FIG. 12, which shows a comparison of two candidate protein sequences to human sequences from a database. The horizontal lines indicate where the human protein sequences from the database are identical to the candidate protein sequence.

[0233] Information pertaining to portions of human protein sequences from the database can be used to design a protein library of human-like chimeric proteins. Such library can be generated by using information pertaining to "crossover sequences" that exist in naturally occurring human proteins. The term "crossover sequence" refers herein to a sequence that is found in its entirety in at least one naturally occurring human protein, in which portions of the sequence are found in two or more naturally occurring proteins. Thus, recombination of the latter two or more naturally occurring proteins would generate a chimeric protein in which the chimeric portion of the sequence actually corresponds to a sequence found in another naturally occurring protein. The crossover sequence contains a chimeric junction of two consecutive amino acid residue positions in which the first
Amino acid position is occupied by an amino acid residue identical in type and position found in a first and second naturally occurring human protein sequence, but not a third naturally occurring human protein sequence. The second amino acid position is occupied by an amino acid residue identical in type and position found in a second and third naturally occurring human protein sequence, but not the first naturally occurring human protein sequence. In other words, the “second” naturally occurring human protein sequence corresponds to the naturally occurring human protein in which the crossover sequence appears in its entirety, as described above.

In some embodiments, a library of human-like chimeric proteins is generated by: identifying human protein sequences from a database that correspond to proteins from the same family of proteins; aligning the human protein sequences from the same family of proteins to a reference protein sequence; identifying a set of subsequences derived from different human protein sequences of the same family, wherein each subsequence shares a region of identity with at least one other subsequence derived from a different naturally occurring human protein sequence; identifying a chimeric junction from a first, a second, and a third subsequence, wherein each subsequence is derived from a different naturally occurring human protein sequence, and wherein the chimeric junction comprises two consecutive amino acid residue positions in which the first amino acid position is occupied by an amino acid residue common to the first and second naturally occurring human protein sequence, but not the third naturally occurring human protein sequence, and the second amino acid position is occupied by an amino acid residue common to the second and third naturally occurring human protein sequence, and generating human-like chimeric protein molecules each corresponding in sequence to two or more subsequences from the set of subsequences, and each comprising one of more of the identified chimeric junctions.

Thus, for example, if the first naturally-occurring human protein sequence is A-B-C, and the second is B-C-D-E, and the third is D-E-F, then the chimeric junction is C-D. Alternatively, if the first naturally-occurring human protein sequence is D-E-F-G, and the second is B-C-D-E-F, and the third is A-B-C-D, then the chimeric junction is D-E. Human-like chimeric protein molecules can be generated in a variety of ways. For example, oligonucleotides comprising sequences encoding the chimeric junctions can be recombined with oligonucleotides corresponding in sequence to two or more subsequences from the above-described set of subsequences to generate a human-like chimeric protein, and libraries thereof. The reference sequence used to align the naturally occurring human proteins is a sequence from the same family of naturally occurring human proteins, or a chimera or other variant of proteins in the family.

Nucleic acids encoding fragments of naturally-occurring monomer domains can also be mixed and/or recombined (e.g., by using chemically or enzymatically-produced fragments) to generate full-length, modified monomer domains. The fragments and the monomer domain can also be recombined by manipulating nucleic acids encoding domains or fragments thereof. For example, ligation of a nucleic acid construct encoding fragments of the monomer domain can be used to generate an altered monomer domain.

Altered monomer domains can also be generated by providing a collection of synthetic oligonucleotides (e.g., overlapping oligonucleotides) encoding conserved, random, pseudorandom, or a defined sequence of peptide sequences that are then inserted by ligation into a predetermined site in a polynucleotide encoding a monomer domain. Similarly, the sequence diversity of one or more monomer domains can be expanded by mutating the monomer domain(s) with site-directed mutagenesis, random mutation, pseudorandom mutation, defined kernal mutation, codon-based mutation, and the like. The resultant nucleic acid molecules can be propagated in a host for cloning and amplification. In some embodiments, the nucleic acids are shuffled.

The present invention also provides a method for recombining a plurality of nucleic acids encoding monomer domains and screening the resulting library for monomer domains that bind to the desired ligand or mixture of ligands or the like. Selected monomer domain nucleic acids can also be back-crossed by shuffling with polynucleotide sequences encoding neutral sequences (i.e., having insubstantial functional effect on binding), such as for example, by back-crossing with a wild-type or naturally-occurring sequence substantially identical to a selected sequence to produce native-like functional monomer domains. Generally, during back-crossing, subsequent selection is applied to retain the property, e.g., binding to the ligand.

In some embodiments, the monomer library is prepared by shuffling. In such a case, monomer domains are isolated and shuffled to combinatorially recombine the nucleic acid sequences that encode the monomer domains (recombination can occur between or within monomer domains, or both). The first step involves identifying a monomer domain having the desired property, e.g., affinity for a certain ligand. While maintaining the conserved amino acids during the recombination, the nucleic acid sequences encoding the monomer domains can be recombined, or recombined and joined into multimers.

A significant advantage of the present invention is that known ligands, or unknown ligands can be used to select the monomer domains and/or multimers. No prior information regarding ligand structure is required to isolate the monomer domains of interest or the multimers of interest. The monomer domains and/or multimers identified can have biological activity, which is meant to include at least specific binding affinity for a selected or desired ligand, and, in some instances, will further include the ability to block the binding of other compounds, to stimulate or inhibit metabolic pathways, to act as a signal or messenger, to stimulate or inhibit cellular activity, and the like. Monomer domains can be generated to function as ligands for receptors where the natural ligand for the receptor has not yet been identified (orphan receptors). These orphan ligands can be created to either block or activate the receptor to which they bind.

A single ligand can be used, or optionally a variety of ligands can be used to select the monomer domains and/or multimers. A monomer domain of the present invention can bind a single ligand or a variety of ligands. A multimer of the present invention can have multiple discrete binding sites for a single ligand, or optionally, can have multiple binding sites for a variety of ligands.

The invention also includes compositions that are produced by methods of the present invention. For example,
the present invention includes monomer domains selected or identified from a library and/or libraries comprising monomer domains produced by the methods of the present invention.

[0243] The present invention also provides libraries of monomer domains and libraries of nucleic acids that encode monomer domains. The libraries can include, e.g., about 100, 250, 500 or more nucleic acids encoding monomer domains, or the library can include, e.g., about 100, 250, 500 or more polypeptides that encode monomer domains. Libraries can include monomer domains containing the same cysteine frame, e.g., A-domains or EGF-like domains.

[0244] In some embodiments, variants are generated by recombining two or more different sequences from the same family of monomer domains (e.g., the LDL receptor class A domain). Alternatively, two or more different monomer domains from different families can be combined to form a multimer. In some embodiments, the multimers are formed from monomers or monomer variants of at least one of the following family classes: an EGF-like domain, a Kringle domain, a fibronectin type I domain, a fibronectin type II domain, a fibronectin type III domain, a PAN domain, a Gla domain, a SRCR domain, a Kunitz/Bovine pancreatic trypsin Inhibitor domain, a Kazal-type serine protease inhibitor domain, a Trefoll (P-type) domain, a von Willebrand factor C domain, an Anaphylatoxin-like domain, a CUB domain, a thyroglobulin type I repeat, LDL-receptor class A domain, a Sushi domain, a Link domain, a Thrombospondin type I domain, an Immunoglobulin-like domain, a C-type lectin domain, a MAM domain, a von Willebrand factor type A domain, a Somatomedin B domain, a WAP-type four disulfide core domain, a F5/8 type C domain, a Hemopexin domain, an SH2 domain, an SH3 domain, a Laminin-type EGF-like domain, a C2 domain and derivatives thereof. In another embodiment, the monomer domain and the different monomer domain can be included in one or more domains found in the Pfam database and/or the SMART database. Libraries produced by the methods above, one or more of cell(s) comprising one or more members of the library, and one or more displays comprising one or more members of the library are also included in the present invention.

[0245] Optionally, a data set of nucleic acid character strings encoding monomer domains can be generated e.g., by mixing a first character string encoding a monomer domain, with one or more character string encoding a different monomer domain, thereby producing a data set of nucleic acids character strings encoding monomer domains, including those described herein. In another embodiment, the monomer domain and the different monomer domain can include one or more domains found in the Pfam database and/or the SMART database. The methods can further comprise inserting the first character string encoding the monomer domain and the one or more second character string encoding the different monomer domain in a computer and generating a multimer character string(s) or library(s), thereof in the computer.

[0246] The libraries can be screened for a desired property such as binding of a desired ligand or mixture of ligands. For example, members of the library of monomer domains can be displayed and prescreened for binding to a known or unknown ligand or a mixture of ligands. The monomer domain sequences can then be mutagenized (e.g., recombined, chemically altered, etc.) or otherwise altered and the new monomer domains can be screened again for binding to the ligand or the mixture of ligands with an improved affinity. The selected monomer domains can be combined or joined to form multimers, which can then be screened for an improved affinity or avidity or altered specificity for the ligand or the mixture of ligands. Altered specificity can mean that the specificity is broadened, e.g., binding of multiple related viruses, or optionally, altered specificity can mean that the specificity is narrowed, e.g., binding within a specific region of a ligand. Those of skill in the art will recognize that there are a number of methods available to calculate avidity. See, e.g., Mammen et al., *Angew Chem Int Ed*: 37:2754-2794 (1998); Muller et al., *Anal Biochem*: 261:149-158 (1998).

VI. Selection of Monomer Domains that Bind c-Met

[0247] Preliminary screens can be conducted by screening for agents capable of binding to c-Met, as at least some of the agents so identified are likely c-Met modulators (e.g., antagonists or agonists). The binding assays usually involve contacting a c-Met protein (or a fragment thereof such as a fragment comprising the SEMA domain or the α chain) with one or more test agents (i.e., monomers or multimers of the invention) and allowing sufficient time for the protein and test agents to form a binding complex. Any binding complexes formed can be detected using any of a number of established analytical techniques. Protein binding assays include, but are not limited to, immunohistochemical binding assays, flow cytometry or other assays. The c-Met protein utilized in such assays can be naturally expressed, cloned or synthesized.

[0248] The screening methods of the invention can be performed as in vitro or cell-based assays. Cell based assays can be performed in any cells in which c-Met is expressed. Cell-based assays may involve whole cells or cell fractions containing a c-Met receptor to screen for agent binding or modulation of activity of c-Met by the agent. Exemplary cell types that can be used according to the methods of the invention include, e.g., any mammalian cells, as well as fungal cells, including yeast, and bacterial cells. Cells can be primary cells or tumor cells or other types of immortal cell lines. Of course, c-Met can be expressed in cells that do not endogenously contain c-Met.

[0249] c-Met activity assays may also be used to identify a modulator (antagonist or agonist) of c-Met. In these embodiments, one or more test agents are contacted to a cell expressing c-Met and then tested for an activity of c-Met. Exemplary c-Met activities include HGF-dependent or constitutive kinase activity. See, e.g., Christensen et al., *Cancer Res*: 63:7345-7355 (2003). In other embodiments, downstream molecular events can also be monitored to determine signaling activity. For example, c-Met induces cell growth (proliferation and survival), cell motility, invasion and morphology changes. In addition, c-Met indirectly mediates phosphorylation of Gab-1, Akt, signal transducer and activator of transcription 3, phospholipase Cγ, and focal adhesions kinase, among others. See, e.g., Christensen et al., *Cancer Res*: 63:7345-7355 (2003).

[0250] In some embodiments, activity assays are also used to confirm that identified antagonist monomers or multimers (i.e., that compete with HGF) lack agonist activity (i.e., that they do not activate c-Met in the absence of HGF or another agonist).
Agents that are initially identified by any of the foregoing screening methods can be further tested to validate the apparent activity. Such studies may be conducted with suitable animal models. The basic format of such methods involves administering a lead compound identified during an initial screen to an animal that serves as a model for humans and then determining if c-Met is in fact modulated and/or the disease or condition is ameliorated. The animal models utilized in validation studies generally are mammalian of any kind. Specific examples of suitable animals include, but are not limited to, primates, mice and rats.

Selection of monomer domains that bind c-Met from a library of domains can be accomplished by a variety of procedures. For example, one method of identifying monomer domains which have a desired property (e.g., binding c-Met) involves translating a plurality of nucleic acids, where each nucleic acid encodes a monomer domain, screening the polypeptides encoded by the plurality of nucleic acids, and identifying those monomer domains that, e.g., bind to a desired ligand or mixture of ligands, thereby producing a selected monomer domain. The monomer domains expressed by each of the nucleic acids can be tested for their ability to bind to the ligand by methods known in the art (i.e., panning, affinity chromatography, FACS analysis).

As mentioned above, selection of monomer domains can be based on binding to a ligand such as c-Met, or a fragment thereof or other target molecule (e.g., lipid, carbohydrate, nucleic acid and the like). Other molecules can optionally be included in the methods along with the target, e.g., ions such as Ca\(^{2+}\).

When a monomer domain of the invention is selected based on its ability to bind to a ligand, the selection basis can include selection based on a slow dissociation rate, which is usually predictive of high affinity. The valency of the ligand can also be varied to control the average binding affinity of selected monomer domains. The ligand can be bound to a surface or substrate at varying densities, such as by including a competitor compound, by dilution, or by other method known to those in the art. High density (valency) of predetermined ligand can be used to enrich for monomer domains that have relatively low affinity, whereas a low density (valency) can preferentially enrich for higher affinity monomer domains.

A variety of reporting display vectors or systems can be used to express nucleic acids encoding the monomer domains and/or multimers of the present invention and to test for a desired activity. For example, a phage display system is a system in which monomer domains are expressed as fusion proteins on the phage surface (Pharmacia, Milwaukee Wis.). Phage display can involve the presentation of a polypeptide sequence encoding monomer domains on the surface of a filamentous bacteriophage, typically as a fusion with a bacteriophage coat protein.

Generally in these methods, each phage particle or cell serves as an individual library member displaying a single species of displayed polypeptide in addition to the natural phage or cell protein sequences. The nucleic acids are cloned into the phage DNA at a site which results in the transcription of a fusion protein, a portion of which is encoded by the plurality of the nucleic acids. The phage containing a nucleic acid molecule undergoes replication and transcription in the cell. The leader sequence of the fusion protein directs the transport of the fusion protein to the tip of the phage particle. Thus, the fusion protein that is partially encoded by the nucleic acid is displayed on the phage particle for detection and selection by the methods described above and below. For example, the phage library can be incubated with a predetermined ligand such as c-Met or a fragment thereof, so that phage particles which present a fusion protein sequence that binds to the ligand can be differentially partitioned from those that do not present polypeptide sequences that bind to the predetermined ligand. For example, the separation can be provided by immobilizing the predetermined ligand. The phage particles (i.e., library members) which are bound to the immobilized ligand are then recovered and replicated to amplify the selected phage subpopulation for a subsequent round of affinity enrichment and phage replication. After several rounds of affinity enrichment and phage replication, the phage library members that are thus selected are isolated and the nucleotide sequence encoding the displayed polypeptide sequence is determined, thereby identifying the sequence(s) of polypeptides that bind to the predetermined ligand. Such methods are further described in PCT patent publication Nos. 91/17271, 91/18980, and 91/19818 and 93/08278.

Examples of other display systems include ribosome displays, a nucleotide-linked display (see, e.g., U.S. Pat. Nos. 6,281,344; 6,194,550, 6,207,446, 6,214,553, and 6,258,558), polysome display, cell surface displays and the like. The cell surface displays include a variety of cells, e.g., E. coli, yeast and/or mammalian cells. When a cell is used as a display, the nucleic acids, e.g., obtained by PCR amplification followed by digestion, are introduced into the cell and translated. Optionally, polypeptides encoding the monomer domains or the multimers of the present invention can be introduced, e.g., by injection, into the cell.

The monomer and multimer libraries of the invention can be screened for a desired property such as binding of a desired ligand (e.g., c-Met) or mixture of ligands. For example, members of the library of monomer domains can be displayed and prescreened for binding to a known or unknown ligand or a mixture of ligands. The monomer domain sequences can then be mutagenized (e.g., recombined, chemically altered, etc.) or otherwise altered and the new monomer domains can be screened again for binding to the ligand or the mixture of ligands with an improved affinity. The selected monomer domains can be combined or joined to form multimers, which can then be screened for an improved affinity or avidity or altered specificity for the ligand or the mixture of ligands. Altered specificity can mean that the specificity is broadened, e.g., binding of multiple related ligands, or optionally, altered specificity can mean that the specificity is narrowed, e.g., binding within a specific region of a ligand. Those of skill in the art will recognize that there are a number of methods available to calculate avidity. See, e.g., Mammen et al., Angew. Chem. Int. Ed. 37:2754-2794 (1998); Muller et al., Anal. Biochem. 261:149-158 (1998).

Those of skill in the art will recognize that the steps of generating variation and screening for a desired property can be repeated (i.e., performed recursively) to optimize results. For example, in a phage display library or other like format, a first screening of a library can be performed at a relatively lower stringency, thereby selected as many par-
articles associated with a target molecule as possible. The selected particles can then be isolated and the polynucleotides encoding the monomer or multimer can be isolated from the particles. Additional variations can then be generated from these sequences and subsequently screened at higher affinity. FIG. 7 illustrates a generic cycle of selection and generation of variation.

[0260] All the compositions of the present invention, e.g., monomer domains as well as multimers and libraries thereof can be optionally bound to a matrix of an affinity material. Examples of affinity material include beads, a column, a solid support, a microarray, other pools of reagent-supports, and the like.

[0261] When multimers capable of binding relatively large targets are desired, they can be generated by a “walking” selection method. This method is carried out by providing a library of monomer domains and screening the library of monomer domains for affinity to a first target molecule. Once at least one monomer that binds to the target is identified, that monomer is covalently linked to a new library or each remaining member of the original library of monomer domains. This new library of multimers (dimers) is then screened for multimers that bind to the target with an increased affinity, and a multimer that binds to the target with an increased affinity can be identified. The “walking” monomer selection method provides a way to assemble a multimer that is composed of monomers that can act additively or even synergistically with each other given the restraints of linker length. This walking technique is very useful when selecting for and assembling multimers that are able to bind large target proteins with high affinity. The walking methods can be repeated to add more monomers thereby resulting in a multimer comprising 2, 3, 4, 5, 6, 7, 8 or more monomers linked together.

[0262] In some embodiments, the selected multimer comprises more than two domains. Such multimers can be generated in a step fashion, e.g., where the addition of each new domain is tested individually and the effect of the domains is tested in a sequential fashion. See, e.g., FIG. 6. In an alternate embodiment, domains are linked to form multimers comprising more than two domains and selected for binding without prior knowledge of how smaller multimers, or alternatively, how each domain, bind.

[0263] The methods of the present invention also include methods of evolving monomers or multimers. As illustrated in FIG. 21, intra-domain recombination can be introduced into monomers across the entire monomer or by taking portions of different monomers to form new recombined units. Interdomain recombination (e.g., recombining different monomers into or between multimers) or recombination of modules (e.g., multiple monomers within a multimer) may be achieved. Inter-library recombination is also contemplated.

[0264] Methods for evolving monomers or multimers can comprise, e.g., any or all of the following steps: providing a plurality of different nucleic acids, where each nucleic acid encoding a monomer domain; translating the plurality of different nucleic acids, which provides a plurality of different monomer domains; screening the plurality of different monomer domains for binding of the desired ligand (e.g., c-Met) or mixture of ligands; identifying members of the plurality of different monomer domains that bind the desired ligand or mixture of ligands, which provides selected monomer domains; joining the selected monomer domains with at least one linker to generate at least one multimer, wherein the at least one multimer comprises at least two of the selected monomer domains and the at least one linker; and, screening the at least one multimer for an improved affinity or avidity or altered specificity for the desired ligand or mixture of ligands as compared to the selected monomer domains.

[0265] Variation can be introduced into either monomers or multimers. An example of improving monomers includes intra-domain recombination in which two or more (e.g., three, four, five, or more) portions of the monomer are amplified separately under conditions to introduce variation (for example by shuffling or other recombination methods) in the resulting amplification products, thereby synthesizing a library of variants for different portions of the monomer. By locating the 5′ ends of the middle primers in a “middle” or ‘overlap’ sequence that both of the PCR fragments have in common, the resulting “left” side and “right” side libraries may be combined by overlap PCR to generate novel variants of the original pool of monomers. These new variants may then be screened for desired properties, e.g., panned against a target or screened for a functional effect. The “middle” primer(s) may be selected to correspond to any segment of the monomer, and will typically be based on the scaffold or one or more consensus amino acids within the monomer (e.g., cysteines such as those found in A domains).

[0266] Similarly, multimers may be created by introducing variation at the monomer level and then recombining monomer variant libraries. On a larger scale, multimers (single or pools) with desired properties may be recombined to form longer multimers. In some cases variation is introduced (typically synthetically) into the monomers or into the linkers to form libraries. This may be achieved, e.g., with two different multimers that bind to two different targets, thereby eventually selecting a multimer with a portion that binds to one target and a portion that binds a second target. See, e.g., FIG. 21.

[0267] Additional variation can be introduced by inserting linkers of different length and composition between domains. This allows for the selection of optimal linkers between domains. In some embodiments, optimal length and composition of linkers will allow for optimal binding of domains. In some embodiments, the domains with a particular binding affinity(s) are linked via different linkers and optimal linkers are selected in a binding assay. For example, domains are selected for desired binding properties and then formed into a library comprising a variety of linkers. The library can then be screened to identify optimal linkers. Alternatively, multimer libraries can be formed where the effect of domain or linker on target molecule binding is not known.

[0268] Methods of the present invention also include generating one or more selected multimers by providing a plurality of monomer domains. The plurality of monomer domains is screened for binding of a desired ligand or mixture of ligands. Members of the plurality of domains that bind the desired ligand or mixture of ligands are identified, thereby providing domains with a desired affinity. The identified domains are joined with at least one linker to generate the multimers, wherein each multimer comprises at
least two of the selected domains and the at least one linker; and, the multimers are screened for an improved affinity or avidity or altered specificity for the desired ligand or mixture of ligands as compared to the selected domains, thereby identifying the one or more selected multimers.

**[0269]** Multimer libraries may be generated, in some embodiments, by combining two or more libraries or monomers or multimers in a recombinase-based approach, where each library member comprises as recombination site (e.g., a lox site). A larger pool of molecularly diverse library members in principle harbor more variants with desired properties, such as higher target-binding affinities and functional activities. When libraries are constructed in phage vectors, which may be transformed into *E. coli*, library size (10^2–10^5) is limited by the transformation efficiency of *E. coli*. A recombinase/recombination site system (e.g., the Cre-loxP system) and in vivo recombination can be exploited to generate libraries that are not limited in size by the transformation efficiency of *E. coli*.

**[0270]** For example, the Cre-loxP system may be used to generate dimer libraries with 10^6, 10^7, 10^8, 10^9, or greater diversity. In some embodiments, *E. coli* as a host for one naïve monomer library and a filamentous phage that carries a second naïve monomer library are used. The library size in this case is limited only by the number of infective phage (carrying one library) and the number of infectible *E. coli* cells (carrying the other library). For example, infecting 10^12 *E. coli* cells (1 L of O(600w=1) with >10^17 phage could produce as many as 10^12 dimer combinations.

**[0271]** Selection of multimers can be accomplished using a variety of techniques including those mentioned above for identifying monomer domains. Other selection methods include, e.g., a selection based on an improved affinity or avidity or altered specificity for the ligand compared to selected monomer domains. For example, a selection can be based on selective binding to specific cell types, or to a set of related cells or protein types (e.g., different virus serotypes). Optimization of the property selected for, e.g., avidity of a ligand, can then be achieved by recombinating the domains, as well as manipulating amino acid sequence of the individual monomer domains or the linker domain or the nucleotide sequence encoding such domains, as mentioned in the present invention.

**[0272]** One method for identifying multimers can be accomplished by displaying the multimers. As with the monomer domains, the multimers are optionally expressed or displayed on a variety of display systems, e.g., phage display, ribosome display, polyethylene display, nucleotide-linked display (see, e.g., U.S. Pat. Nos. 6,281,344, 6,194, 550, 6,207,446, 6,214,553, and 6,258,558) and/or cell surface display, as described above. Cell surface displays can include but are not limited to *E. coli*, yeast or mammalian cells. In addition, display libraries of multimers with multiple binding sites can be panned for avidity or affinity or altered specificity for a ligand or for multiple ligands.


**[0274]** Other variations include the use of multiple binding compounds, such that monomer domains, multimers or libraries of these molecules can be simultaneously screened for a multiplicity of ligands or compounds that have different binding specificity. Multiple predetermined ligands or compounds can be concomitantly screened in a single library, or sequential screening against a number of monomer domains or multimers. In one variation, multiple ligands or compounds, each encoded on a separate bead (or subset of beads), can be mixed and incubated with monomer domains, multimers or libraries of these molecules under suitable binding conditions. The collection of beads, comprising multiple ligands or compounds, can then be used to isolate, by affinity selection, selected monomer domains, selected multimers or library members. Generally, subsequent affinity screening rounds can include the same mixture of beads, subsets thereof, or beads containing only one or two individual ligands or compounds. This approach affords efficient screening, and is compatible with laboratory automation, batch processing, and high throughput screening methods.

**[0275]** In another embodiment, multimers can be simultaneously screened for the ability to bind multiple ligands, wherein each ligand comprises a different label. For example, each ligand can be labeled with a different fluorescent label, contacted simultaneously with a multimer or multimer library. Multimers with the desired affinity are then identified (e.g., by FACS sorting) based on the presence of the labels linked to the desired label.

**[0276]** Libraries of either monomer domains or multimers (referred to the following discussion for convenience as “affinity agents”) can be screened (i.e., panned) simultaneously against multiple ligands in a number of different formats. For example, multiple ligands can be screened in a simple mixture, in an array, displayed on a cell or tissue (e.g., a cell or tissue provides numerous molecules that can be bound by the monomer domains or multimers of the invention), and/or immobilized. The libraries of affinity agents can optionally be displayed on yeast or phage display
systems. Similarly, if desired, the ligands (e.g., encoded in a cDNA library) can be displayed in a yeast or phage display system.

[0277] Initially, the affinity agent library is panned against the multiple ligands. Optionally, the resulting “hits” are panned against the ligands one or more times to enrich the resulting population of affinity agents. See, e.g., FIG. 13.

[0278] If desired, the identity of the individual affinity agents and/or ligands can be determined. In some embodiments, affinity agents are displayed on phage. Affinity agents identified as binding in the initial screen are divided into a first and second portion. See, e.g., FIG. 14. The first portion is infected into bacteria, resulting in either plaques or bacterial colonies, depending on the type of phage used. The expressed phage are immobilized and then probed with ligands displayed in phage selected as described below.

[0279] The second portion is coupled to beads or otherwise immobilized and a phage display library containing at least some of the ligands in the original mixture is contacted to the immobilized second portion. Those phage that bind to the second portion are subsequently eluted and contacted to the immobilized phage described in the paragraph above. Phage-phage interactions are detected (e.g., using a monoclonal antibody specific for the ligand-expressing phage) and the resulting phage polynucleotides can be isolated.

[0280] In some embodiments, the identity of an affinity agent-ligand pair is determined. For example, when both the affinity agent and the ligand are displayed on a phage or yeast, the DNA from the pair can be isolated and sequenced. In some embodiments, polynucleotides specific for the ligand and affinity agent are amplified. Amplification primers for each reaction can include 5’ sequences that are complementary such that the resulting amplification products are fused, thereby forming a hybrid polynucleotide comprising a polynucleotide encoding at least a portion of the affinity agent and at least a portion of the ligand. The resulting hybrid can be used to probe affinity agent or ligand (e.g., cDNA-encoded) polynucleotide libraries to identify both affinity agent and ligand. See, e.g., FIG. 15.

[0281] The above-described methods can be readily combined with “walking” to simultaneous generate and identify multiple multimers, each of which bind to a ligand in a mixture of ligands. In these embodiments, a first library of affinity agents (monomer domains or multimers) are panned against multiple ligands and the eluted affinity agents are linked to the first or a second library of affinity agents to form a library of multimeric affinity agents (e.g., comprising 2, 3, 4, 5, 6, 7, 8, 9, or more monomer), which are subsequently panned against the multiple ligands. This method can be repeated to continue to generate larger multimeric affinity agents. Increasing the number of monomer domains may result in increased affinity and avidity for a particular target. For example, the inventors have found that trimers of monomer domains that bind CD28 have a higher affinity than dimmers, which in turn have a higher affinity than single CD28-binding monomer domains alone. Of course, at each stage, the panning is optionally repeated to enrich for significant binders. In some cases, walking will be facilitated by inserting recombination sites (e.g., lox sites) at the ends of monomers and recombinating monomer libraries by a recombinase-mediated event.

[0282] The selected multimers of the above methods can be further manipulated, e.g., by recombining or shuffling the selected multimers (recombination can occur between or within multimers or both), mutating the selected multimers, and the like. This results in altered multimers which then can be screened and selected for members that have an enhanced property compared to the selected multimer, thereby producing selected altered multimers.

[0283] In view of the description herein, it is clear that the following process may be followed. Naturally or non-naturally occurring monomer domains may be recombined or variants may be formed. Optionally the domains initially or later are selected for those sequences that are less likely to be immunogenic in the host for which they are intended. Optionally, a phage library comprising the recombined domains is panned for a desired affinity. Monomer domains or multimers expressed by the phage may be screened for IC₅₀ for a target. Hetero- or homo-multimeric multimers may be selected. The selected polypeptides may be selected for their affinity to any target, including, e.g., hetero- or homomultimeric targets.

[0284] Linkers, multimers or selected multimers produced by the methods indicated above and below are features of the present invention. Libraries comprising multimers, e.g., a library comprising about 100, 250, 500 or more members produced by the methods of the present invention or selected by the methods of the present invention are also included. Libraries of the recombinant polypeptides are also a feature of the present invention, e.g., a library comprising about 100, 250, 500 or more different recombinant polypeptides.

[0285] Compositions of the present invention can be bound to a matrix of an affinity material, e.g., the recombinant polypeptides. Examples of affinity material include, e.g., beads, a column, a solid support, and/or the like.

VII. Therapeutic and Prophylactic Treatment Methods

[0286] The present invention also includes methods of therapeutically or prophylactically treating a disease or disorder by administering in vivo or ex vivo one or more nucleic acids or polypeptides of the invention described above (or compositions comprising a pharmaceutically acceptable excipient and one or more such nucleic acids or polypeptides) to a subject, including, e.g., a mammal, including a human, primate, mouse, pig, cow, goat, rabbit, rat, guinea pig, hamster, horse, sheep; or a non-mammalian vertebrate such as a bird (e.g., a chicken or duck), fish, or invertebrate.

[0287] Antagonists of c-Met are useful treating cancers in which Met is expressed. Exemplary cancers include bladder, breast, cervical, colorectal, esophageal, gastric, head and neck, kidney, liver, lung, nasopharyngeal, ovarian, pancreatic, gall bladder, prostate or thyroid cancer, osteosarcoma, synovial sarcoma, rhabdomyosarcoma, MFI/fibrosarcoma, Kaposi’s sarcoma, multiple myeloma, lymphomas, adult T-cell leukemia, glioblastomas, astrocytomas, melanoma, mesothelioma, and Wilm’s tumor.

[0288] In one aspect of the invention, in ex vivo methods, one or more cells or a population of cells of interest of the subject (e.g., tumor cells, tumor tissue sample, organ cells, blood cells, cells of the skin, lung, heart, muscle, brain, mucosae, liver, intestine, spleen, stomach, lymphatic system, cervix, vagina, prostate, mouth, tongue, etc.) are
obtained or removed from the subject and contacted with an amount of a selected monomer domain and/or multimer of the invention that is effective in prophylactically or therapeutically treating the disease, disorder, or other condition. The contacted cells are then returned or delivered to the subject to the site from which they were obtained or to another site (e.g., including those described above) of interest in the subject to be treated. If desired, the contacted cells can be grafted onto a tissue, organ, or system site (including all described above) of interest in the subject or contacting said one or more cells or population of cells with a polynucleotide construct comprising a nucleic acid sequence of the invention that encodes a biologically active polypeptide of interest (e.g., a selected monomer domain and/or multimer) that is effective in prophylactically or therapeutically treating the disease, disorder, or other condition. The one or more cells or population of cells is contacted with a sufficient amount of the polynucleotide construct and a promoter controlling expression of said nucleic acid sequence such that uptake of the polynucleotide construct (and promoter) into the cell(s) occurs and sufficient expression of the target nucleic acid sequence of the invention results to produce an amount of the biologically active polypeptide, encoding a selected monomer domain and/or multimer, effective to prophylactically or therapeutically treat the disease, disorder, or condition. The polynucleotide construct can include a promoter sequence (e.g., CMV promoter sequence) that controls expression of the nucleic acid sequence of the invention and/or, if desired, one or more additional nucleotide sequences encoding at least one or more of another polypeptide of the invention, a cytokine, adjuvant, or co-stimulatory molecule, or other polypeptide of interest.

[0289] The invention also provides in vivo methods in which one or more cells or a population of cells of interest of the subject are contacted directly or indirectly with an amount of a selected monomer domain and/or multimer of the invention effective in prophylactically or therapeutically treating the disease, disorder, or other condition. In direct contact/administration formats, the selected monomer domain and/or multimer is typically administered or transferred directly to the cells to be treated or to the tissue site of interest (e.g., tumor cells, tumor tissue sample, organ cells, blood cells, cells of the skin, lung, heart, muscle, brain, mucosa, liver, intestine, spleen, stomach, lymphatic system, cervix, vagina, prostate, mouth, tongue, etc.) by any of a variety of formats, including topical administration, injection (e.g., by using a needle or syringe), or vaccine or gene gun delivery, pushing into a tissue, organ, or skin site. The selected monomer domain and/or multimer can be delivered, for example, intramuscularly, intradermally, subcutaneously, orally, intraperitoneally, intravenously, or placed within a cavity of the body (including, e.g., during surgery), or by inhalation or vaginal or rectal administration.

[0290] In vivo indirect contact/administration formats, the selected monomer domain and/or multimer is typically administered or transferred indirectly to the cells to be treated or to the tissue site of interest, including those described above (such as, e.g., skin cells, organ systems, lymphatic system, or blood cell system, etc.), by contacting or administering the polypeptide of the invention directly to one or more cells or population of cells from which treatment can be facilitated. For example, tumor cells within the body of the subject can be treated by contacting cells of the blood or lymphatic system, skin, or an organ with a sufficient amount of the selected monomer domain and/or multimer such that delivery of the selected monomer domain and/or multimer to the site of interest (e.g., tissue, organ, or cells of interest or blood or lymphatic system within the body) occurs and effective prophylactic or therapeutic treatment results. Such contact, administration, or transfer is typically made by using one or more of the routes or modes of administration described above.

[0291] In another aspect, the invention provides ex vivo methods in which one or more cells of interest or a population of cells of interest of the subject (e.g., tumor cells, tumor tissue sample, organ cells, blood cells, cells of the skin, lung, heart, muscle, brain, mucosa, liver, intestine, spleen, stomach, lymphatic system, cervix, vagina, prostate, mouth, tongue, etc.) are obtained or removed from the subject and transformed by contacting said one or more cells or population of cells with a polynucleotide construct comprising a nucleic acid sequence of the invention that encodes a biologically active polypeptide of interest (e.g., a selected monomer domain and/or multimer) that is effective in prophylactically or therapeutically treating the disease, disorder, or other condition. The polynucleotide construct can be directly administered or transferred to cell(s) suffering from the disease or disorder (e.g., by direct contact using one or more of the routes or modes of administration described above). Alternatively, the polynucleotide construct can be indirectly administered or transferred to cell(s) suffering from the disease or disorder by first directly contacting non-diseased cell(s) or other diseased cells using one or more of the routes or modes of administration described above with a sufficient amount of the polynucleotide construct comprising the
nucleic acid sequence encoding the biologically active polypeptide, and a promoter controlling expression of the nucleic acid sequence, such that uptake of the polynucleotide construct (and promoter) into the cell(s) occurs and sufficient expression of the nucleic acid sequence of the invention results to produce an amount of the biologically active polypeptide effective to prophylactically or therapeutically treat the disease or disorder, and whereby the polynucleotide construct or the resulting expressed polypeptide is transferred naturally or automatically from the initial delivery site, system, tissue or organ of the subject's body to the diseased site, tissue, organ or system of the subject's body (e.g., via the blood or lymphatic system). Expression of the target nucleic acid occurs naturally or can be induced (as described in greater detail below) such that an amount of expressed polypeptide is sufficient and effective to treat the disease or condition at the site or tissue system. The polynucleotide construct can include a promoter sequence (e.g., CMV promoter sequence) that controls expression of the nucleic acid sequence and/or, if desired, one or more additional nucleotide sequences encoding at least one or more of another polypeptide of the invention, a cytokine, adjuvant, or co-stimulatory molecule, or other polypeptide of interest.

In each of the in vivo and ex vivo treatment methods as described above, a composition comprising an excipient and the polypeptide or nucleic acid of the invention can be administered or delivered. In one aspect, a composition comprising a pharmaceutically acceptable excipient and a polypeptide or nucleic acid of the invention is administered or delivered to the subject as described above in an amount effective to treat the disease or disorder.

In another aspect, in each in vivo and ex vivo treatment method described above, the amount of polynucleotide administered to the cell(s) or subject can be an amount such that uptake of said polynucleotide into one or more cells of the subject occurs and sufficient expression of said nucleic acid sequence results to produce an amount of a biologically active polypeptide effective to enhance an immune response in the subject, including an immune response induced by an immunogen (e.g., antigen). In another aspect, for each such method, the amount of polypeptide administered to cell(s) or subject can be an amount sufficient to enhance an immune response in the subject, including that induced by an immunogen (e.g., antigen).

In yet another aspect, in an in vivo or in vivo treatment method in which a polynucleotide construct (or composition comprising a polynucleotide construct) is used to deliver a physiologically active polypeptide to a subject, the expression of the polynucleotide construct can be induced by using an inducible on- and off-gene expression system. Examples of such on- and off-gene expression systems include the Tet-On™ Gene Expression System and Tet-Off™ Gene Expression System (see, e.g., Clontech Catalog 2000, pg. 110-111 for a detailed description of each such system), respectively. Other controllable or inducible on- and off-gene expression systems are known to those of ordinary skill in the art. With such system, expression of the target nucleic acid of the polynucleotide construct can be regulated in a precise, reversible, and quantitative manner. Gene expression of the target nucleic acid can be induced, for example, after the stable transfected cells containing the polynucleotide construct comprising the target nucleic acid are delivered or transferred to or made to contact the tissue site, organ or system of interest. Such systems are of particular benefit in treatment methods and formats in which it is advantageous to delay or precisely control expression of the target nucleic acid (e.g., to allow time for completion of surgery and/or healing following surgery; to allow time for the polynucleotide construct comprising the target nucleic acid to reach the site, cells, system, or tissue to be treated; to allow time for the graft containing cells transformed with the construct to become incorporated into the tissue or organ onto or into which it has been spliced or attached, etc.).

VIII. Additional Multimer Uses

[0298] The potential applications of multimers of the present invention are diverse and include any use where an affinity agent is desired.

[0299] The present invention provides a method for extending the serum half-life of a protein, including, e.g., a multimer of the invention or a protein of interest in an animal. The protein of interest can be any protein with therapeutic, prophylactic, or otherwise desirable functionality. This method comprises first providing a monomer domain that has been identified as a binding protein that specifically binds to a half-life extender such as a blood-carried molecule or cell, such as serum albumin (e.g., human serum albumin), IgG, red blood cells, etc. The half-life extender-binding monomer is then covalently linked to another monomer domain that has a binding affinity for the protein of interest (e.g., Met). This complex formation results in the half-life extension protecting the multimer and/or bound protein(s) from proteolytic degradation and/or other removal of the multimer and/or protein(s) and thereby extending the half-life of the protein and/or multimer. One variation of this use of the invention includes the half-life extender-binding monomer covalently linked to the protein of interest. The protein of interest may include a monomer domain, a multimer of monomer domains, or a synthetic drug. Alternatively, monomers that bind to either immunoglobulins or erythrocytes could be generated using the above method and could be used for half-life extension.

[0300] The half-life extender-binding multimers are typically multimers of at least two domains, chimeric domains, or mutagenized domains (i.e., one that binds to Met and one that binds to the blood-carried molecule or cell). Suitable domains include all of those described herein, that are further screened and selected for binding to a half-life extender. The half-life extender-binding multimers are generated in accordance with the methods for making multimers described herein, using, for example, monomer domains pre-screened for half-life extender-binding activity. For example, some half-life extender-binding LDL receptor class A-domain monomers are described in Example 2 below. The serum half-life of a molecule can be extended to be, e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 400, 500 or more hours.

[0301] In some cases, a pair of monomers or multimers are selected to bind to the same target (i.e., for use in sandwich-based assays). To select a matched monomer or multimer pair, two different monomers or multimers typically are able to bind the target protein simultaneously. One approach to identify such pairs involves the following:

[0302] (1) immobilizing the phage or protein mixture that was previously selected to bind the target protein
(2) contacting the target protein to the immobilized phage or protein and washing;

(3) contacting the phage or protein mixture to the bound target and washing; and

(4) eluting the bound phage or protein without eluting the immobilized phage or protein.

One use of the multimers or monomer domains of the invention is to replace antibodies or other affinity agents in detection or other affinity-based assays. Thus, in some embodiments, monomer domains or multimers are selected against the ability to bind components other than a target in a mixture. The general approach can include performing the affinity selection under conditions that closely resemble the conditions of the assay, including mimicking the composition of a sample during the assay. Thus, a step of selection could include contacting a monomer domain or multimer to a mixture not including the target ligand and selecting against any monomer domains or multimers that bind to the mixture. Thus, the mixtures (absent the target ligand, which could be depleted using an antibody, monomer domain or multimer) representing the sample in an assay (serum, blood, tissue, cells, urine, semen, etc.) can be used as a blocking agent. Such subtraction is useful, e.g., to create pharmaceutical proteins that bind to their target but not to other serum proteins or non-target tissues.

For example, the invention can be used in the application for creating antagonists, where the selected monomer domains or multimers block the interaction between two proteins, e.g., the α and β chains of Met and/or between Met and HGF. Optionally, the invention can generate agonists. For example, multimers binding two different proteins, e.g., enzyme and substrate, can enhance protein function, including, for example, enzymatic activity and or substrate conversion.

In some embodiments, the monomer domains are used for ligand inhibition, ligand clearance or ligand stimulation. Possible ligands in these methods, include, e.g., HGF.

If inhibition of ligand binding to a receptor is desired, a monomer domain is selected that binds to the ligand (e.g., HGF) at a portion of the ligand that contacts the ligand’s receptor, or that binds to the receptor at a portion of the receptor that binds contacts the ligand, thereby preventing the ligand-receptor interaction. The monomer domains can optionally be linked to a half-life extender, if desired.

Ligand clearance refers to modulating the half-life of a soluble ligand in, bodily fluid. For example, most monomer domains, absent a half-life extender, have a short half-life. Thus, binding of a monomer domain to the ligand will reduce the half-life of the ligand, thereby reducing ligand concentration by clearing the ligand through the kidney so long as the complex is no larger than the maximum size able to pass through the kidney (less than about 50 or 40 kDa). The portion of the ligand (e.g., HGF) bound by the monomer domain will generally not matter, though it may be beneficial to bind the ligand at the portion of the ligand that binds to its receptor (e.g., Met), thereby further inhibiting the ligand’s effect. This method is useful for reducing the concentration of any molecule in the bloodstream.

Alternatively, a multimer comprising a first monomer domain that binds to a half-life extender and a second monomer domain that binds to a portion of the ligand that does not bind to the ligand’s receptor can be used to increase the half-life of the ligand.

In another embodiment, a multimer comprising a first monomer domain that binds to the ligand and a second monomer domain that binds to the receptor can be used to increase the effective affinity of the ligand for the receptor.

In another embodiment, multimers comprising at least two monomers that bind to receptors are used to bring two receptors into proximity by both binding the multimer, thereby activating the receptors.

Further examples of potential uses of the invention include monomer domains, and multimers thereof, that are capable of drug binding (e.g., binding radiolabeled tracers for targeting, pharmaceutical binding for half-life extension of drugs, controlled substance binding for overdose treatment and addiction therapy), immune function modulating (e.g., immunogenicity blocking by binding such receptors as CTLA-4, immunogenicity enhancing by binding such receptors as CD80, or complement activation by Fc type binding), and specialized delivery (e.g., slow release by linker cleavage, electrotransport domains, dimerization domains, or specific binding to cell entry domains, clearance receptors such as FcR, oral delivery receptors such as plgR for trans-mucosal transport, and blood-brain transfer receptors such as transferring).

Further examples of potential uses of the invention include monomer domains, and multimers thereof, that are capable of drug binding (e.g., binding radiolabeled tracers for targeting, pharmaceutical binding for half-life extension of drugs, controlled substance binding for overdose treatment and addiction therapy), immune function modulating (e.g., immunogenicity blocking by binding such receptors as CTLA-4, immunogenicity enhancing by binding such receptors as CD80, or complement activation by Fc type binding), and specialized delivery (e.g., slow release by linker cleavage, electrotransport domains, dimerization domains, or specific binding to cell entry domains, clearance receptors such as FcR, oral delivery receptors such as plgR for trans-mucosal transport, and blood-brain transfer receptors such as transferring).

In further embodiments, monomers or multimers can be linked to a detectable label (e.g., Cy3, Cy5, etc.) or linked to a reporter gene product (e.g., CAT, luciferase, horseradish peroxidase, alkaline phosphatase, GFP, etc.).

Monomers or multimers of the invention that bind to Met may also be used in diagnostic and predictive applications in which is is useful to detect Met. For example, detection of Met can be used to predict prognosis of breast cancer, wherein higher abundance of Met than in a normal tissue indicates a poor prognosis. See, e.g., U.S. Pat. No. 6,673,559.

IX. Further Manipulating Monomer Domains and/or Multimer Nucleic Acids And Polypeptides

As mentioned above, the polypeptide of the present invention can be altered. Descriptions of a variety of diversity generating procedures for generating modified or altered nucleic acid sequences encoding these polypeptides are described herein and the references cited therein.

Another aspect of the present invention includes the cloning and expression of monomer domains, selected monomer domains, multimers and/or selected multimers coding nucleic acids. Thus, multimer domains can be synthesized as a single protein using expression systems well known in the art. General texts which describe molecular biological techniques useful herein, including the use of vectors, promoters and many other topics relevant to expressing nucleic acids such as monomer domains, selected monomer domains, multimers and/or selected multimers, include Berg et al., Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, Calif. (Berger); Sambrook et al., Molecular Cloning—A Laboratory Manual (2nd Ed.), Vol.

[0319] The present invention also relates to the introduction of vectors of the invention into host cells, and the production of monomer domains, selected monomer domains, multimers and/or selected multimers of the invention by recombinant techniques. Host cells are genetically engineered (i.e., transduced, transformed or transfected) with the vectors of this invention, which can be, for example, a cloning vector or an expression vector. The vector can be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the monomer domain, selected monomer domain, multimer and/or selected mutimer gene(s) of interest. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, e.g., Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley-Liss, New York and the references cited therein.

[0320] As mentioned above, the polypeptides of the invention can also be produced in non-animal cells such as plants, yeast, fungi, bacteria and the like. Indeed, as noted throughout, phage display is an especially relevant technique for producing such polypeptides. In addition to Sambrook, Berger and Ausubel, details regarding cell culture can be found in Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, N.Y.; Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, Fla.

[0321] The present invention also includes alterations of monomer domains, immuno-domains and/or multimers to improve pharmacological properties, to reduce immunogenicity, or to facilitate the transport of the multimer and/or monomer domain into a cell or tissue (e.g., through the blood-brain barrier, or through the skin). These types of alterations include a variety of modifications (e.g., the addition of sugar-groups or glycosylation), the addition of PEG, the addition of protein domains that bind a certain protein (e.g., HSA or other serum protein), the addition of proteins fragments or sequences that signal movement or transport into, out of and through a cell. Additional components can also be added to a multimer and/or monomer domain to manipulate the properties of the multimer and/or monomer domain. A variety of components can also be added including, e.g., a domain that binds a known receptor (e.g., a Fe-region protein domain that binds a Fe receptor), a toxin(s) or part of a toxin, a prodomain that can be optionally cleaved off to activate the multimer or monomer domain, a reporter molecule (e.g., green fluorescent protein), a component that bind a reporter molecule (such as a radionucide for radiotherapy, biotin or avidin) or a combination of modifications.

X. Animal Models

[0322] Another aspect of the invention is the development of specific non-human animal models in which to test the immunogenicity of the monomer or multimer domains. The method of producing such non-human animal model comprises: introducing into at least some cells of a recipient non-human animal, vectors comprising genes encoding a plurality of human proteins from the same family of proteins, wherein the genes are each operably linked to a promoter that is functional in at least some of the cells into which the vectors are introduced such that a genetically modified non-human animal is obtained that can express the plurality of human proteins from the same family of proteins.

[0323] Suitable non-human animals employed in the practice of the present invention include all vertebrate animals, except humans (e.g., mouse, rat, rabbit, sheep, and the like). Typically, the plurality of members of a family of proteins includes at least two members of that family, and usually at least ten family members. In some embodiments, the plurality includes all known members of the family of proteins. Exemplary genes that can be used include those encoding monomer domains, such as, for example, members of the LDL receptor class A-domain family, the EGF-like domain family, as well as the other domain families described herein.

[0324] The non-human animal models of the present invention can be used to screen for immunogenicity of a monomer or multimer domain that is derived from the same family of proteins expressed by the non-human animal model. The present invention includes the non-human animal model made in accordance with the method described above, as well as transgenic non-human animals whose somatic and germ cells contain and express DNA molecules encoding a plurality of human proteins from the same family
of proteins (such as the monomer domains described herein), wherein the DNA molecules have been introduced into the transgenic non-human animal at an embryonic stage, and wherein the DNA molecules are each operably linked to a promoter in at least some of the cells in which the DNA molecules have been introduced.

0325 An example of a mouse model useful for screening LDL receptor class A-domain derived binding proteins is described as follows. Gene clusters encoding the wild type human LDL receptor class A-domain monomers are amplified from human cells using PCR. Almost all of the 200 different A-domains can be amplified with only three separate PCR amplification reactions of about 7 kb each. These fragments are then used to generate transgenic mice according to the method described above. The transgenic mice will recognize the human A-domains as "self," thus mimicking the "selfness" of a human with regard to A-domains. Individual A-domain-derived monomers or multimers are tested in these mice by injecting the A-domain-derived monomers or multimers into the mice, then analyzing the immune response (or lack of response) generated. The mice are tested to determine if they have developed a mouse anti-human response (MAHR). Monomers and multimers that do not result in the generation of a MAHR are likely to be non-immunogenic when administered to humans.

0326 Historically, MAHR test in transgenic mice is used to test individual proteins in mice that are transgenic for that single protein. In contrast, the above described method provides a non-human animal model that recognizes an entire family of human proteins as "self," and that can be used to evaluate a huge number of variant proteins that each are capable of vastly varied binding activities and uses.

XI. Kits

0327 Kits comprising the components needed in the methods (typically in an unmixed form) and kit components (packaging materials, instructions for using the components and/or the methods, one or more containers (reaction tubes, columns, etc.)) for holding the components are a feature of the present invention. Kits of the present invention may contain a multimer library, or a single type of monomer or multimer. Kits can also include reagents suitable for promoting target molecule binding, such as buffers or reagents that facilitate detection, including detectably-labeled molecules. Standards for calibrating a ligand binding to a monomer domain or the like, can also be included in the kits of the invention.

0328 The present invention also provides commercially valuable binding assays and kits to practice the assays. In some of the assays of the invention, one or more ligand is employed to detect binding of a monomer domain, immunodomains and/or multimer. Such assays are based on any known method in the art, e.g., flow cytometry, fluorescent microscopy, plasmon resonance, and the like, to detect binding of a ligand(s) to the monomer domain and/or multimer.

0329 Kits based on the assay are also provided. The kits typically include a container, and one or more ligand. The kits optionally comprise directions for performing the assays, additional detection reagents, buffers, or instructions for the use of any of these components, or the like. Alternatively, kits can include cells, vectors, (e.g., expression vectors, secretion vectors comprising a polypeptide of the invention), for the expression of a monomer domain and/or a multimer of the invention.

0330 In a further aspect, the present invention provides for the use of any composition, monomer domain, immunodomain, multimer, cell, cell culture, apparatus, apparatus component or kit herein, for the practice of any method or assay herein, and/or for the use of any apparatus or kit to practice any assay or method herein and/or for the use of cells, cell cultures, compositions or other features herein as a therapeutic formulation. The manufacture of all components herein as therapeutic formulations for the treatments described herein is also provided.

XII. Integrated Systems

0331 The present invention provides computers, computer readable media and integrated systems comprising character strings corresponding to monomer domains, selected monomer domains, multimers and/or selected multimers and nucleic acids encoding such polypeptides. These sequences can be manipulated by in silico recombination methods, or by standard sequence alignment or word processing software.

0332 For example, different types of similarity and considerations of various stringency and character string length can be detected and recognized in the integrated systems herein. For example, many homology determination methods have been designed for comparative analysis of sequences of biopolymers, for spell checking in word processing, and for data retrieval from various databases. With an understanding of double-helix pair-wise complement interactions among 4 principal nucleobases in natural polynucleotides, models that simulate annealing of complementary homologous polynucleotide strings can also be used as a foundation of sequence alignment or other operations typically performed on the character strings corresponding to the sequences herein (e.g., word-processing manipulations, construction of figures comprising sequence or subsequence character strings, output tables, etc.). An example of a software package with DBs for calculating sequence similarity is BLAST, which can be adapted to the present invention by inputting character strings corresponding to the sequences herein.

0333 BLAST is described in Altschul et al., (1990) J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (available on the World Wide Web at ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the
word hits in each direction are halted when: the cumulative alignment score falls off by the quantity \( X \) from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters \( W, T \), and \( X \) determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength \( W \) of 11, an expectation \( E \) of 10, a cutoff of 100, \( M=5 \), \( N=-4 \), and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength \( W \) of 3, an expectation \( E \) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

[0334] An additional example of a useful sequence alignment algorithm is PILJEF. PILJEF creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILJEF uses a simplification of the progressive alignment method of Feng & Doolittle, (1987) J. Mol. Evol. 35:351-360. The method used is similar to the method described by Higgins & Sharp, (1989) CABIOS 5:151-153. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster can then be aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences can be aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program can also be used to plot a dendrogram or tree representation of clustering relationships. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison. For example, in order to determine conserved amino acids in a monomer domain family or to compare the sequences of monomer domains in a family, the sequence of the invention, or coding nucleic acids, are aligned to provide structure-function information.

[0335] In one aspect, the computer system is used to perform “in silico” sequence recombination or shuffling of character strings corresponding to the monomer domains. A variety of such methods are set forth in “Methods For Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics” by Selifonov and Stemmer, filed Feb. 5, 1999 (U.S. Ser. No. 60/118,854) and “Methods For Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics” by Selifonov and Stemmer, filed Oct. 12, 1999 (U.S. Ser. No. 09/416,375). In brief, genetic operators are used in genetic algorithms to change given sequences, e.g., by mimicking genetic events such as mutation, recombination, death and the like. Multi-dimensional analysis to optimize sequences can be also be performed in the computer system, e.g., as described in the ‘375 application.

[0336] A digital system can also instruct an oligonucleotide synthesizer to synthesize oligonucleotides, e.g., used for gene reconstruction or recombination, or to order oligonucleotides from commercial sources (e.g., by printing appropriate order forms or by linking to an order form on the Internet).

[0337] The digital system can also include output elements for controlling gene synthesis (e.g., based upon a sequence or an alignment of a recombinant, e.g., recombinant, monomer domain as herein), i.e., an integrated system of the invention optionally includes an oligonucleotide synthesizer or an oligonucleotide synthesis controller. The system can include other operations that occur downstream from an alignment or other operation performed using a character string corresponding to a sequence herein, e.g., as noted above with reference to assays.

EXAMPLES

[0338] The following example is offered to illustrate, but not to limit the claimed invention.

Example 1

[0339] This example describes selection of monomer domains and the creation of multimers.

[0340] Starting materials for identifying monomer domains and creating multimers from the selected monomer domains and procedures can be derived from any of a variety of human and/or non-human sequences. For example, to produce a selected monomer domain with specific binding for a desired ligand or mixture of ligands, one or more monomer domain gene(s) are selected from a family of monomer domains that bind to a certain ligand. The nucleic acid sequences encoding the one or more monomer domain gene can be obtained by PCR amplification of genomic DNA or cDNA, or optionally, can be produced synthetically using overlapping oligonucleotides.

[0341] Most commonly, these sequences are then cloned into a cell surface display format (i.e., bacterial, yeast, or mammalian (COS) cell surface display; phage display) for expression and screening. The recombinant sequences are transfected (transduced or transformed) into the appropriate host cell where they are expressed and displayed on the cell surface. For example, the cells can be stained with a labeled (e.g., fluorescently labeled), desired ligand. The stained cells are sorted by flow cytometry, and the selected monomer domains encoding genes are recovered (e.g., by plasmid isolation, PCR or expansion and cloning) from the positive cells. The process of staining and sorting can be repeated multiple times (e.g., using progressively decreasing concentrations of the desired ligand until a desired level of enrichment is obtained). Alternatively, any screening or detection method known in the art that can be used to identify cells that bind the desired ligand or mixture of ligands can be employed.

[0342] The selected monomer domain encoding genes recovered from the desired ligand or mixture of ligands binding cells can be optionally recombined according to any of the methods described herein or in the cited references. The recombinant sequences produced in this round of diversification are then screened by the same or a different method to identify recombinant genes with improved affinity for the desired or target ligand. The diversification and selection process is optionally repeated until a desired affinity is obtained.

[0343] The selected monomer domain nucleic acids selected by the methods can be joined together via a linker sequence to create multimers, e.g., by the combinatorial
assembly of nucleic acid sequences encoding selected monomer domains by DNA ligation, or optionally, PCR-based, self-priming overlap reactions. The nucleic acid sequences encoding the multimers are then cloned into a cell surface display format (i.e., bacterial, yeast, or mammalian (COS) cell surface display; phage display) for expression and screening. The recombinant sequences are transduced (transduced or transformed) into the appropriate host cell where they are expressed and displayed on the cell surface. For example, the cells can be stained with a labeled, e.g., fluorescently labeled, desired ligand or mixture of ligands. The stained cells are sorted by flow cytometry, and the selected multimers encoding genes are recovered (e.g., by PCR or expansion and cloning) from the positive cells. Positive cells include multimers with an improved avidity or affinity or altered specificity to the desired ligand or mixture of ligands compared to the selected monomer domain(s). The process of staining and sorting can be repeated multiple times (e.g., using progressively decreasing concentrations of the desired ligand or mixture of ligands until a desired level of enrichment is obtained). Alternatively, any screening or detection method known in the art that can be used to identify cells that bind the desired ligand or mixture of ligands can be employed.

[0344] The selected multimer encoding genes recovered from the desired ligand or mixture of ligands binding cells can be optionally recombined according to any of the methods described herein or in the cited references. The recombinant sequences produced in this round of diversification are then screened by the same or a different method to identify recombinant genes with improved avidity or affinity or altered specificity for the desired or target ligand. The diversification and selection process is optionally repeated until a desired avidity or affinity or altered specificity is obtained.

Example 2

[0345] This example describes the selection of monomer domains that are capable of binding to Human Serum Albumin (HSA). The phage particles were mixed with 1/5 volumes 20% w/v PEG 8000, 15% w/v NaCl, and was incubated for several hours at 4°C. After centrifugation (20 minutes, 10000 g, 4°C) the precipitated phage particles were dissolved in 2 ml of cold TBS (50 mM Tris, 150 mM NaCl, pH 8.0) containing 2 mM CaCl₂. The solution was incubated on ice for 30 minutes and was distributed into two 1.5 ml reaction vessels. After centrifugation to remove undissolved components (5 minutes, 15000 g, 4°C) the supernatants were transferred to a new reaction vessel. Phage particles were reprecipitated by adding 1/5 volumes 20% w/v PEG 8000, 15% w/v NaCl and incubation for 60 minutes on ice. After centrifugation (30 minutes, 15000 g, 4°C) and removal of the supernatants, the precipitated phage particles were dissolved in a total of 1 ml TBS containing 2 mM CaCl₂. After incubation for 30 minutes on ice the solution was centrifuged as described above. The supernatant containing the phage particles was used directly for the affinity enrichment.

[0348] Affinity enrichment of phage was performed using 96 well plates (Maxisorp, NUNC, Denmark). Single wells were coated for 12 h at RT by incubation with 150 µl of a solution of 100 µg/ml human serum albumin (HSA, Sigma) in TBS. Binding sites remaining after HSA incubation were saturated by incubation with 250 µl 2% w/v bovine serum albumin (BSA) in TBS (TBS with 0.1% w/v Tween 20) for 2 hours at RT. Afterwards, 40 µl of the phage solution, containing approximately 5×10¹⁴ phage particles, were mixed with 80 µl TBS containing 3% BSA and 2 mM CaCl₂ for 1 hour at RT. In order to remove non binding phage particles, the wells were washed 5 times with 1 ml per wash using 130 µl TBS containing 2 mM CaCl₂.

[0349] Phage bound to the well surface were eluted either by incubation for 15 minutes with 130 µl of 0.1 M glycine/HCl pH 2.2 or in a competitive manner by adding 130 µl of 500 µg/ml HSA in TBS. In the first case, the pH of the elution fraction was immediately neutralized after removal from the well by mixing the eluate with 30 µl 1 M Tris/HCl pH 8.0.

[0350] For the amplification of phage, the eluate was used to infect E. coli K91BluKan cells (F⁺). 50 µl of the eluted phage solution were mixed with 50 µl of a preparation of cells and incubated for 10 minutes at RT. Afterwards, 20 ml LB medium containing 20 µg/ml tetracycline were added and the infected cells were grown for 36 h at 22°C, 250 rpm. Afterwards, the cells were sedimented (10 minutes, 5000 g, 4°C). Phage were recovered from the supernatant by precipitation as described above. For the repeated affinity enrichment of phage particles the same procedure as described in this example was used. After two subsequent rounds of panning against HSA, random colonies were picked and tested for their binding properties against the used target protein.

Example 3

[0351] This example describes the determination of biological activity of monomer domains that are capable of binding to HSA.

[0352] In order to show the ability of an HSA binding domain to extend the serum half life of a protein in vivo, the following experimental setup was performed. A multimeric A-domain, consisting of an A-domain which was evolved for binding HSA (see Example 2) and a streptavidin binding A-domain was compared to the streptavidin binding A-do-
main itself. The proteins were injected into mice, which were either loaded or not loaded (as control) with human serum albumin (HSA). Serum levels of α-domain proteins were monitored.

Therefore, an A-domain, which was evolved for binding HSA (see Example 1) was fused on the genetic level with a streptavidin binding A-domain multimer using standard molecular biology methods (see Maniatis et al.). The resulting genetic construct, coding for an A-domain multimer as well as a hexahistidine tag and a HA tag, were used to produce protein in E. coli. After refolding and affinity tag mediated purification the proteins were dialyzed several times against 150 mM NaCl, 5 mM Tris pH 8.0, 100 μM CaCl₂, and sterile filtered (0.45 μM).

Two sets of animal experiments were performed. In a first set, 1 ml of each prepared protein solution with a concentration of 2.5 μM were injected into the tail vein of separate mice and serum samples were taken 2, 5 and 10 minutes after injection. In a second set, the protein solution described before was supplemented with 50 mg/ml human serum albumin. As described above, 1 ml of each solution was injected per animal. In case of the injected streptavidin binding A-domain dimer, serum samples were taken 2, 5 and 10 minutes after injection, while in case of the trimer, serum samples were taken after 10, 30 and 120 minutes. All experiments were performed as duplicates and individual animals were assayed per time point.

In order to detect serum levels of A-domains in the serum samples, an enzyme linked immunosorbent assay (ELISA) was performed. Therefore, wells of a maxisorp 96 well microtiter plate (NUNC, Denmark) were coated with each 1 μg anti-His₆-antibody in TBS containing 2 mM CaCl₂ for 1 h at 4 °C. After blocking remaining binding sites with casein (Sigma) solution for 1 h, wells were washed three times with TBS containing 0.1% Tween and 2 mM CaCl₂. Serial concentration dilutions of the serum samples were prepared and incubated in the wells for 2 h in order to capture the α-domain proteins. After washing as before, anti-HA-tag antibody coupled to horse radish peroxidase (HRP) (Roche Diagnostics, 25 μg/ml) was added and incubated for 2 h. After washing as described above, HRP substrate (Pierce) was added and the detection reaction developed according to the instructions of the manufacturer. Light absorption, reflecting the amount of α-domain protein present in the serum samples, was measured at a wavelength of 450 nm. Obtained values were normalized and plotted against a time scale.

Evaluation of the obtained values showed a serum half life for the streptavidin binding A-domain of about 4 minutes without presence of HSA respectively 5.2 minutes when the animal was loaded with HSA. The trimer of A-domains, which contained the HSA binding A-domain, exhibited a serum half life of 6.3 minutes without the presence of HSA but a significantly increased half life of 38 minutes when HSA was present in the animal. This clearly indicates that the HSA binding A-domain can be used as a fusion partner to increase the serum half life of any protein, including protein therapeutics.

Example 4

This example describes experiments demonstrating extension of half-life of proteins in blood.

To further demonstrate that blood half-life of proteins can be extended using monomer domains of the invention, individual monomer domain proteins selected against monkey serum albumin, human serum albumin, human IgG, and human red blood cells were added to aliquots of whole, heparinized human or monkey blood.

The following list provides sequences of monomer domains analyzed in this example.
IG156  CLSSEFQCCSSGRCIPLAWVDGDND<RDDSDEKSCPRT
RBCA  CRSSQFQCNDSRICIPGRWRCDGDNDCQDGSDETGCGDSHILPFSTPGPST
RBCB  CPGAEPPCKNGGQLPVWLCDGVNCDLDGSDEKGCGRPGPGATSAAPAA
RBC11  CPPDEFPCKNGQCPPQDCLCDGVNCDLDGSDEKDCGRPGPGATSAAPAA
CSA-A8  CGAGQFPCKNHCLPLNLLCDGVNDEDNSDEPSLCKALT
Blood aliquots containing monomer protein were then added to individual dialysis bags (25,000 MWCO), sealed, and stirred in 4 L of Tris-buffered saline at room temperature overnight.

Anti-foxAHis antibody was immobilized by hydrophobic interaction to a 96-well plate (Nunc). Serial dilutions of serum from each blood sample were incubated with the immobilized antibody for 3 hours. Plates were washed to remove unbound protein and probed with α-HA-HRP to detect monomer.

Monomers identified as having long half-lives in dialysis experiments were constructed to contain either an HA, FLAG, E-Tag, or myc epitope tag. Four monomers were pooled, containing one protein for each tag, to make two pools.

One monkey was injected subcutaneously per pool, at a dose of 0.25 mg/kg/monomer in 2.5 mL total volume in saline. Blood samples were drawn at 24, 48, 96, and 120 hours. Anti-foxAHis antibody was immobilized by hydrophobic interaction to a 96-well plate (Nunc). Serial dilutions of serum from each blood sample were incubated with the immobilized antibody for 3 hours. Plates were washed to remove unbound protein and separately probed with α-HA-HRP, α-FLAG-HRP, α-E-Tag-HRP, and α-myc-HRP to detect the monomer.

The following illustrates a comparison between commercial antibodies and an anti-IgG multimer:

<table>
<thead>
<tr>
<th>Drug</th>
<th>Mol. Wt.</th>
<th>Human T1/2 Dosing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rebif IFN-β</td>
<td>23 kD</td>
<td>69 hrs Weekly 3x</td>
</tr>
<tr>
<td>Pegasys IFN-α-PEG</td>
<td>40 kD</td>
<td>78 hrs Weekly</td>
</tr>
<tr>
<td>Rituxan CD20 Antibody</td>
<td>150 kD</td>
<td>78 hrs Weekly</td>
</tr>
<tr>
<td>Enbrel TNF-α-Fe</td>
<td>150 kD</td>
<td>103 hrs Weekly 2x</td>
</tr>
<tr>
<td>Multimer Anti-IgG</td>
<td>5 kD</td>
<td>120 hrs Weekly 1–2x</td>
</tr>
<tr>
<td>Herceptin Her2 Antibody</td>
<td>150 kD</td>
<td>144 hrs Weekly</td>
</tr>
<tr>
<td>Remicade TNF-α Antibody</td>
<td>150 kD</td>
<td>216 hrs Monthly 5x</td>
</tr>
<tr>
<td>Humira TNF-α Antibody</td>
<td>150 kD</td>
<td>236 hrs Monthly 2x</td>
</tr>
</tbody>
</table>

Example 5

This example describes in vivo intra-protein recombination to generate libraries of greater diversity.

A monomer-encoding plasmid vector (pCK-derived vector; see below), flanked by orthologous loxP sites, was recombined in a Cre-dependent manner with a phage vector via its compatible loxP sites. The recombined phage vectors were detected by PCR using primers specific for the recombinate construct. DNA sequencing indicated that the correct recombinate product was generated.

Reagents and Experimental Procedures

pCK-cre-lox-Mb-loxP. This vector has two particularly relevant features. First, it carries the cre gene, encoding the site-specific DNA recombinase Cre, under the control of Pcre. Cre was PCR-amplified from p705-cre (from Gene-Bridges) with cre-specific primers that incorporated XbaI (5) and SfiI (3) at the ends of the PCR product. This product was digested with XbaI and SfiI and cloned into the identical sites of pCK, a bla*, Cm derivative of pCK110919-HC-Bla (pACYC ori), yielding pCK-cre.

The second feature is the naïve A domain library flanked by two orthologous loxP sites, loxP(wild-type) and loxP(FAS), which are required for the site-specific DNA recombination catalyzed by Cre. See, e.g., Siegel, R. W., et al. FEBS Letters 505:467-473 (2001). These sites rarely recombine with another. loxP sites were built into pCK-cre sequentially. 5'-phosphorylated oligonucleotides loxP(K) and loxP(Krc), carrying loxP(WT) and EcoRI and HinDIII-compatible overhangs to allow ligation to digested EcoRI and HinDIII-digested pCK, were hybridized together and ligated to pCK-cre in a standard ligation reaction (T4 ligase; overnight at 16 C).

The resulting plasmid was digested with EcoRI and Sphi and ligated to the hybridized, 5'-phosphorylated oligos loxP(L) and loxP(Lrc), which carry loxP(FAS) and EcoRI and Sphi-compatible overhangs. To prepare for library construction, a large-scale purification (Qiagen MAXI prep) of pCK-cre-lox-Pl(wo)-loxP(FAS) was performed according to Qiagen's protocol. The Qiagen-purified plasmid was subjected to CsCl gradient centrifugation for further purification. This construct was then digested with Sphi and BglIII and ligated to digested naïve A domain library insert, which was obtained via a PCR-amplification of a preexisting A domain library pool. By design, the loxP sites and Mb are in-frame, which generates Mbs with loxP-encoded linkers. This library was utilized in the in vivo recombination procedure as detailed below.

fUSESHA-Mb-lox-lox vector. The vector is a derivative of fUSE5 from George Smith's laboratory (University of Missouri). It was subsequently modified to carry an HA tag for immunodetection assays. loxP sites were built into fUSESHA sequentially. 5'-phosphorylated oligonucleotides loxP(I) and loxP(Irc), carrying loxP(WT), a string of stop codons and Xmal and SfiI-compatible overhangs, were hybridized together and ligated to Xmal- and SfiI-digested fUSESHA in a standard ligation reaction (New England Biolabs T4 ligase; overnight at 16 C).

The resulting phage vector was next digested with Xmal and Sphi and ligated to the hybridized oligos loxP(I) and loxP(Irc), which carry loxP(FAS) and overhangs compatible with Xmal and Sphi. This construct was digested with Xmal/SfiI and then ligated to pre-cut (Xmal/SfiI) naïve A domain library insert (PCR product). The stop codons are located between the loxP sites, preventing expression of gIII and consequently, the production of infectious phase.

The ligated vector/library was subsequently transformed into an E. coli host bearing a gIII-expressing plasmid that allows the rescue of the fUSESHA-Mb-lox-lox phage, as detailed below.

pCK-gIII. This plasmid carries gIII under the control of its native promoter. It was constructed by PCR-amplifying gIII and its promoter from VCSM13 helper phage (Stratagene) with primers gIII promoter EcoRI and gIII promoter HinDIII. This product was digested with EcoRI and HinDIII and cloned into the same sites of pCK110919-HC-Bla. As gIII is under the control of its own promoter, gIII expression is presumably constitutive. pCK-gIII was transformed into E. coli EC100 (Epicentre).
In vivo recombination procedure. In summary, the procedure involves the following key steps: a) Production of infective (i.e. rescue) of fUSE5HA-Mb-lox-fox library with an E. coli host expressing gIII from a plasmid; b) Cloning of 2nd library (pCK) and transformation into Fc TG1 E. coli; c) Infection of the culture carrying the 2nd library with the rescued fUSE5HA-Mb-fox phage library.

a. Rescue of phage vector. Electrocompetent cells carrying pCK-gIII were prepared by a standard protocol. These cells had a transformation frequency of 4×10^9/µg DNA and were electroporated with large-scale ligations (~5 µg vector DNA) of fUSE5HA-fox vector and the naive A domain library insert. After individual electroporations library size of approximately 10^8. After an hour-long recovery period at 37 C with shaking, the electroporated cells are diluted to OD_595~0.05 in 2xYT (plus 20 µg/mL chloramphenicol) and grown to mid-log phase at 37 C before infection by fUSE5HA-Mb-fox-fox.

b. Cloning of the 2nd library and transformation into an E. coli host. The ligated pCK/naive A domain library is electroporated into a bacterial Fc host, with an expected

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>loxP(K)</td>
</tr>
<tr>
<td>[P-5' agcttataactttcgatagaaaggtatatacctctgtgctgcaggtcag]</td>
</tr>
<tr>
<td>loxP(K.rc)</td>
</tr>
<tr>
<td>[P-5' atacctgtatagactcagtcagagctatatacctctgtgctgcaggtcag]</td>
</tr>
<tr>
<td>loxP(L)</td>
</tr>
<tr>
<td>[P-5' atacctgtatagactcagtcagagctatatacctctgtgctgcaggtcag]</td>
</tr>
<tr>
<td>loxP(L.rc)</td>
</tr>
<tr>
<td>[P-5' atacctgtatagactcagtcagagctatatacctctgtgctgcaggtcag]</td>
</tr>
<tr>
<td>loxP(I)</td>
</tr>
<tr>
<td>[P-5' ccgggcagtcaggtgtatatacctctgtgctgcaggtcag]</td>
</tr>
<tr>
<td>loxP(I.rc)</td>
</tr>
<tr>
<td>[P-5' ccgggcagtcaggtgtatatacctctgtgctgcaggtcag]</td>
</tr>
<tr>
<td>loxP(J)</td>
</tr>
<tr>
<td>[5' ccgggcagtcaggtgtatatacctctgtgctgcaggtcag]</td>
</tr>
<tr>
<td>loxP(J.rc)</td>
</tr>
<tr>
<td>[5' ccgggcagtcaggtgtatatacctctgtgctgcaggtcag]</td>
</tr>
<tr>
<td>gIIIPromoter_EcoRI</td>
</tr>
<tr>
<td>[5' atggcacctaattgtctggtcagac]</td>
</tr>
<tr>
<td>gIIIPromoter_HindIII</td>
</tr>
<tr>
<td>[5' gattccagtttaaactttctggcag]</td>
</tr>
</tbody>
</table>

(100 ng DNA/electroporation) with ~70 µL cells/cuvette, 930 µL warm SOC media were added, and the cells were allowed to recover with shaking at 37 C for 1 hour. Next, tetracycline was added to a final concentration of 0.2 µg/mL, and the cells were shaken for ~45 minutes at 37 C. An aliquot of this culture was removed, 10-fold serially diluted and plated to determine the resulting library size (1.8×10^8). The remaining culture was diluted into 2x500 mL 2xYT (with 20 µg/mL chloramphenicol and 20 µg/mL tetracycline to select for pCK-gIII and the fUSE5HA-based vector, respectively) and grown overnight at 30 C.

Rescued phage were harvested using a standard PEG/NaCl precipitation protocol. The titer was approximately 1×10^12 transducing units/mL.

b. Cloning of the 2nd library and transformation into an E. coli host. The ligated pCK/naive A domain library is electroporated into a bacterial Fc host, with an expected

Example 6

This example describes construction of an EGF-based monomer library.

The CaEGF domain library, E3, encodes a protein domain of 36-43 amino acids having the following pattern:

X(5)C1-X(4)-C2-X(4,5)-C3-X(8)-C4-X(1)-C5-X(8,12)-C6

The table below describes for each position which amino acids are encoded in the library based upon the natural diversity of human calcium binding EGF domains:

The library of DNA sequences, E3, encoding monomeric calcium binding EGF domains, was created by assembly PCR as described in Stemmer et al., Gene 164:49-53 (1995). The oligonucleotides used in this PCR reaction are in two groups, 1 and 2. They are:
Group 1:
1. 5'-AAAAGCCCTCGAGGGCCTGGGATATG-3'
2. 5'-'-CCCTGACCCACCAACAHCCCGGAGGCGAAGYCGRCRMACATT
   ATYAYATCTDYTACATTCGACCACC-3'
3. 5'-CCCTGACCCACCAACAHCCCGGAGGCGAAGYCGRCRMACATT
   ATYAYATCTDYTACATTCGACCACC-3'
4. 5'-CCCTGACCCACCAACAHCCCGGAGGCGAAGYCGRCRMACATT
   ATYAYATCTDYTACATTCGACCACC-3'
5. 5'-CCCTGACCCACCAACAHCCCGGAGGCGAAGYCGRCRMACATT
   ATYAYATCTDYTACATTCGACCACC-3'
6. 5'-TGAAATTTTGCTATGATGGTTGGCTAACCACTTCTCAAGTCTTCGCCCC
   CAGAACCCCTGGCACCTGACAACACCACA-3'

Group 2:
1. 5'-'-ACGTYTGTACCCGATGATGGTTGGCACCACCC-3'
2. 5'-CCCTGACCCACCAACAHCCCGGAGGCGAAGYCGRCRMACATT
   ATYAYATCTDYTACATTCGACCACC-3'
3. 5'-CCCTGACCCACCAACAHCCCGGAGGCGAAGYCGRCRMACATT
   ATYAYATCTDYTACATTCGACCACC-3'
4. 5'-AAAAGGCCCTCGAGGGCCTGGGATATG-3'


[0386] Following the separate PCRs of the Group 1 and 2 oligonucleotides, the Group 1 PCR fragments were digested with BpmI and group 2 PCR fragments were digested with BsrDI. Digestion products were purified using Qiagen Qiaquick columns and then ligated together. The ligated DNA was then amplified in a PCR using two primers. These are:

5'-AAAAGCCCTCGAGGGCCTGGGATATG-3'
5'-AAAAGCCCTCGAGGGCCTGGGATATG-3'

[0387] The PCR products were purified with Qiagen Qiaquick columns and digested with Sfil. The digested product was purified with Qiagen Qiaquick columns. The DNA fragments were ligated into the Sfil restriction sites of phage display vector fus5-HA(G4S)4, a derivative of fus5 carrying an in-frame HA-epitope and a glycine, sorine flexible linker. The amino acids which comprise the flexible linker are: S-G-G-G-S-G-G-G-S-G-G-G-S-G-G-G-G-G. The ligation mixture was electroporated into Transformation Max™ EC100™ electrocompetent E. coli cells. Transformed E. coli cells were grown overnight at 37°C in 2xYT medium containing 20 μg/ml tetracycline. The resulting library contained 2x10⁹ independent clones. Phage particles were purified from the culture medium by PEG-precipitation. The titer of the phage was 1.3x10¹²/ml. The sequences of 24 individual clones were determined and these were consistent with the library design.

Example 7

[0388] This example describes construction of an EGF-based monomer library.

[0389] FIG. 19 illustrates the process of intradomain optimization by recombination. Shown is a three-fragment PCR overlap reaction, which recombines three segments of a single domain relative to each other. One can use two, three, four, or more fragment overlap reactions in the same way as illustrated. This recombination process has many applications. One application is to recombine a large pool of hundreds of previously selected clones without sequence information. All that is needed for each overlap to work is one known region of (relatively) constant sequence that exists in the same location in each of the clones (fixed site approach). For A domains, typically these clones would have been derived from a library in which 20-25 amino acids distributed over all five inter-cysteine segments were randomized. The intra-domain recombination method can also be performed on a pool of sequence-related monomer domains by standard DNA recombination (e.g., Stemmer, Nature 370:389-391(1994)) based on random fragmentation and reassembly based on DNA sequence homology, which does not require a fixed overlap site in all of the clones that are to be recombined.

[0390] Another application of this process is to create multiple separate, naïve (meaning unpanned) libraries in each of which only one of the intercysteine loops is randomized, to randomize a different loop in each library. After panning of these libraries separately against the target, the selected clones are then recombined. From each panned library only the randomized segment is amplified by PCR and multiple randomized segments are then combined into a single domain, creating a shuffled library which is panned and/or screened for increased potency. This process can also be used to shuffle a small number of clones of known sequence.
[0391] Any common sequence may be used as crossover points. For A domains or other cysteine-containing monomers, the cysteine residues are logical places for the crossover. However, there are other ways to determine optimal crossover sites, such as computer modeling. Alternatively, residues with highest entropy, or the least number of intramolecular contacts, may also be good sites for crossovers.

[0392] An exemplary method of generating libraries comprised of proteins with randomized inter-cysteine loops is presented below. In this example, in contrast to the separate loop, separate library approach described above, multiple intercysteine loops are randomized simultaneously in the same library.

[0393] An A domain NNK library encoding a protein domain of 39-45 amino acids having the following pattern was constructed:

[0394] C1—X(n,6)-E1-F-R1-C2-A-X(2,4)-G1-R2-C3-P—S1-S2-W-V-C4-D1-G2-E2-D2-D3-C5-G3-D4-G4-S3-D5-E3-X(4,6)—C6;

where,

[0395] C1-C6: cysteines;

[0396] X(n): sequence of n amino acids with any residue at each position;

[0397] E1-E3: glutamine;

[0398] F: phenylalanine;

[0399] R1-R2: arginine;

[0400] A: alanine;

[0401] G1-G4: glycine;

[0402] I: isoleucine;

[0403] P: proline;

[0404] S1-S3: serine;

[0405] W: tryptophan;

[0406] V: valine;

[0407] D1-D5: aspartic acid; and

[0408] C1-C3, C2-C5 & C4-C6 form disulfides.

[0409] The library was constructed by creating a library of DNA sequences, containing tyrosine codons (TAT) or variable non-conserved codons (NNK), by assembly PCR as described in Stemmer et al., Gene 164:49-53 (1995). Compared to the native A-domain scaffold and the design that was used to construct library A1 (described previously) this approach: 1) keeps more of the existing residues in place instead of randomizing these potentially critical residues, and 2) inserts a string of amino acids of variable length of all 20 amino acids (NNK codon), such that the average number of inter-cysteine residues is extended beyond that of the natural A domain or the A1 library. The rate of tyrosine residues was increased by including tyrosine codons in the oligonucleotides, because tyrosines were found to be over-represented in antibody binding sites, presumably because of the large number of different contacts that tyrosine can make. The oligonucleotides used in this PCR reaction are:

1. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
2. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
3. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
4. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
5. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
6. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
7. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
8. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
9. 5'-ATATCCCCCGCGGCTCGAGGGCTCGTGGTATGTGGTGCCGATGGCGAATTCGGA-3'
10. 5'-ATACCCAAGAAGAGCGTAGTTATCATGTGCCHNNNTGCGACTCGGAATTC-3'
11. 5'-ATACCCAAGAAGAGCGTAGTTATCATGTGCCHNNNTGCGACTCGGAATTC-3'
12. 5'-ATACCCAAGAAGAGCGTAGTTATCATGTGCCHNNNTGCGACTCGGAATTC-3'
13. 5'-ATACCCAAGAAGAGCGTAGTTATCATGTGCCHNNNTGCGACTCGGAATTC-3'
14. 5'-ATACCCAAGAAGAGCGTAGTTATCATGTGCCHNNNTGCGACTCGGAATTC-3'
15. 5'-ATACCCAAGAAGAGCGTAGTTATCATGTGCCHNNNTGCGACTCGGAATTC-3'
16. 5'-ATACCCAAGAAGAGCGTAGTTATCATGTGCCHNNNTGCGACTCGGAATTC-3'
17. 5'-ATACCCAAGAAGAGCGTAGTTATCATGTGCCHNNNTGCGACTCGGAATTC-3'
18. 5'-ACCGCTCTTCTTTGATGTGACCGGGAGCAGATTGTGACCGGACCTGACGAG-3'
The library was constructed though an initial round of 10 cycles of PCR amplification using a mixture of 4 pools of oligonucleotides, each pool containing 400 pmols of DNA. Pool 1 contained oligonucleotides 1-9, pool 2 contained 10-17, pool 3 contained only 18 and pool 4 contained 19-27. The fully assembled library was obtained through an additional 8 cycles of PCR using pool 1 and 4. The library fragments were digested with Xmal and Sfil. The DNA fragments were ligated into the corresponding restriction sites of phage display vector fuse5-HA, a derivative of fuse5 carrying an in-frame HA-epitope. The ligation mixture was electroported into TransforMax™ EC100™ electrocompeptent E. coli cells resulting in a library of 2×10⁶ individual clones. Transformed E. coli cells were grown overnight at 37°C in 2xYT medium containing 20 µg/ml tetracycline. Phage particles were purified from the culture medium by PEG-precipitation and a titer of 1.1×10¹⁵/ml was determined. Sequences of 24 clones were determined and were consistent with the expectations of the library design.

Example 8

The assays may include binding assays to a target or agonist or antagonist potency determination of the multimer in functional protein- or cell-based assays.

[0413] The monomeric domain(s) of the single best trimer are then optimized as a second step. Homomultimers are easiest to optimize, since only one domain sequence exists, though heteromultimers may also be synthesized. For homomultimers, an increase in binding by the multimer compared to the monomer is an avidity effect.

[0414] After optimization of the domain sequence itself (e.g., by recombining or NNK randomization) and phage panning, the improved monomers are used to construct a dimer with a linker library. Linker libraries may be formed, e.g., from linkers with an NNK composition and/or variable sequence length.

[0415] After panning of this linker library, the best clones (e.g., determined by potency in the inhibition or other functional assay) are converted into multimers composed of multiple (e.g., two, three, four, five, six, seven, eight, etc.) sequence-optimized domains and length- and sequence-optimized linkers.
Theoretical Space by the physical library. Further, in the case of non-overlapping categories such as the Top and Bottom categories, half-domain sequences selected against different targets can be combined into a single sequence which would be able to bind simultaneously or alternatively to the selected targets. In either case, creating binding sites that occupy only half a domain allows for the creation of molecules that are half as large and would have half the number of immunogenic epitopes, reducing the risk of immunogenicity.

Structural Classification of A-domain Positions

A canonical A-domain sequence is shown below with high-diversity positions represented as an X. Positions that belong to either the Top, Bottom, Loop 1, or Loop 2 categories are designated with a star.

Example 10

This example describes screening for monomers or multimers that bind c-Met.

Phage libraries are panned through several rounds either on solid support (i.e., Maxisorb plates) or in solution (i.e., Dynal Streptavidin beads). The panning is repeated on each enriched library several times in order to select for a pool with the enrichment and maximal diversity.

1. Round 1 (Maxisorb Plates or Dynal Beads)

1. Coating Target

A. Coating plates: Six wells/library or coated with c-Met (approximately 0.5 μg/well) using 100 μL/well of 5 μg/mL pure c-Met diluted in TBS [pH 7.5]/2 mM CaCl₂. In addition, one negative control well/library is coated with TBS [pH 7.5]/2 mM CaCl₂ only. The plates are then incubated for 1.5 hrs. at room temperature with shaking.

B. Coating beads: 20 μL Dynal streptavidin beads (M~280) plus 5 μg biotinylated target are incubated in 500 μL TBS [pH 7.5]/2 mM CaCl₂ and rotated at room temperature for 1 hr in Eppendorf tubes. As a negative control, 20 μL Dynal streptavidin beads without target are incubated in 500 μL TBS [pH 7.5]/2 mM CaCl₂, and rotated at room temperature for 1 hr. Dynal beads are typically washed at least twice with TBS [pH 7.5]/2 mM CaCl₂ before adding target and beads are coated in bulk.

2.Blocking

A. Blocking Plates: Coating solution is removed and wells are washed one time with 200 μL/well of TBS [pH 7.5]/2 mM CaCl₂ and 250 μL/well of 1% BSA (protease free) in TBS [pH 7.5]/2 mM CaCl₂ and incubated for 1 hr. at room temperature with shaking. Alternative reagents can be used to block if BSA contains proteins closely related to the target of interest (for example reagents such as casein or milk).

B. Blocking beads: Coating solution is removed and beads are washed twice with TBS [pH 7.5]/2 mM CaCl₂; add 500 μL 1% BSA (protease free) in TBS [pH 7.5]/2 mM CaCl₂ and rotated for 1 hr. at room temperature.

3. Washes

A. Wash Plates: Wells are washed three times with 200 μL/well of TBS [pH 7.5]/2 mM CaCl₂ to remove excess target.

B. Wash Beads: Beads are washed three times with 1000 μL TBS [pH 7.5]/2 mM CaCl₂ to remove excess target. Beads are allowed to collect on a magnet for a few minutes after each wash to avoid bead loss.

4. Phage Addition

A. Phage addition to plates: About 1000 library equivalents are added in 1% nonfat dry milk/0.2% BSA (protease free) or (appropriate blocking agent) in TBS [pH 7.5]/2 mM CaCl₂ and incubated at room temperature for 2 hours with shaking. In rounds 2-3, 100 μL total of harvested phage is added to 7 wells (6 target+1 negative control) diluted in phage addition buffer.

B. Phage addition to beads: About 1000 library equivalents are added in 500 μL 1% non-fat dry milk+100 μL 1% BSA (protease free) in TBS [pH 7.5]/2 mM CaCl₂ and incubated with rotation at room temperature for 2 hr. In rounds 2-3, 100 μL total of harvested phage are added to beads.

5. Washes

A. Washing Plates: The plates are washed eight times with 200 μL/well of TBS [pH 7.5]/2 mM CaCl₂/0.1% Tween-20 over a period of 10 minutes.
B. Washing Beads: The beads are washed with 800 µl of TBS [pH 7.5]/2 mM CaCl₂/0.1% Tween-20 over a period of 30-45 mins. Bead resuspension is facilitated by dispensing wash buffer directly onto collected beads or by pipetting up and down.

Conditions for Stringent Washes (Options)

- a. 800 µl of TBS [pH 7.5]/2 mM CaCl₂/0.1% Tween-20 at 37°C;
- b. 800 µl of TBS [450 mM NaCl, pH 7.5]/2 mM CaCl₂/0.1% Tween-20 at room temperature;
- c. Wash beads normally 6-8x, then add unlabeled 1 µg of e-Met for 1 hour at room temperature or 37°C. Phage that survive this wash are retained for elution/infection;
- d. 1% milk/0.2% BSA with or without 1 M urea/37°C (high stringency).

6. Competition (Optional):

A. Competition on Plates: Phage are incubated with 100 µl/well of 50 µg/ml or 5 µg/ml (10x over target) of c-Met competitor in TBS [pH 7.5]/2 mM CaCl₂ for 1 hr. at room temperature, with shaking.

B. Competition on Beads: Phage are incubated with 10 µg e-Met in 500 µl TBS [pH 7.5]/2 mM CaCl₂ for 1 hr. at room temperature with shaking.

7. Phage Elution

A. Elution off of Plates: 100 µl/well of 10 mg/mL trypsin in TBS [pH 7.5]/2 mM CaCl₂ is added and incubated at 37°C for 30 min; with shaking.

B. Elution off of Beads: 100 µl 10 mg/ml trypsin TBS [pH 7.5]/2 mM CaCl₂ is added to beads and incubated at 37°C for 30 min; shaking in Eppendorf rack at 37°C.

C. Alternative elution/infection: 200 µL of log-phase BlueKan K91 cells at OD₆₀₀=0.5 are added to each well (for plates) or to aspirated beads. The infection is allowed to proceed for 30 minutes at 37°C without shaking. Next, the 200 µL volumes are pooled and added to ~3 mL of 2xYT/0.2 µg/mL tetracycline and shaken for 15 minutes.

8. Infection: (Same for Plate and Bead Protocol)

An appropriate volume of BlueKan K91 (in 2xYT/40 µg/mL kanamycin) are grown to OD₆₀₀ ~0.5-0.6. When the culture reaches this OD₆₀₀, culture may be placed on ice prior to use, although the time on ice is generally minimized.

A. In 50 mL sterile conical tube, the eluted phage is mixed with 5 mL of BlueKan K91 culture and incubated at 37°C for 25 minutes without shaking, covering sterile conical tubes with AirPore tape in order to allow for aeration.

B. Tetracycline is added to a final concentration of 0.2 µg/mL and shaken for 15 minutes at 37°C.

C. A 30 µL aliquot is sampled for titering and diluted 10-fold (10⁻⁵ to 10⁻⁵) in 2xYT, plated in 8 µL dilution in spots on 2xYT/20 µg/mL tetracycline plates and incubated plates overnight at 30°C or 37°C. The remaining volume of 10⁻⁵-10⁻⁴ dilutions is plated to obtain single colonies for phage ELISAs.

D. Infected cultures are diluted ~10-fold into 50 mL 2xYT/20 µg/mL tetracycline and incubated with shaking at 30°C overnight to saturation.

9. Titer sample of input phage used in the current round of panning (same for Plate and Bead Protocol)

A. 100-fold serial dilutions of harvested phage (10⁻⁴ to 10⁻¹) are made in 2xYT.

B. 100 µl/well of OD₆₀₀ 0.5-0.6 BlueKan K91 is added to 6 wells of 96-well polypropylene plate.

C. 10 µl of diluted phage is added to wells containing 100 µl of BlueKan K91 cells.

D. Phage are incubated at 37°C for 25 minutes without shaking, with the plates covered with AirPore tape to allow for aeration.

E. Tetracycline is added to a final concentration of 0.2 µg/mL and the plate is shaken for 15 minutes at 37°C.

F. 8 µL of each dilution (10⁻⁴ to 10⁻¹⁰) is plated onto a dry 2xYT agar/20 µg/mL tetracycline plate.

G. Plates are incubated at 30°C or 37°C overnight.

10. Harvesting Phage (Same for Plate and Bead Protocols)

A. Cultures are centrifuged overnight at 7000 rpm in disposable 50 mL tubes for 25 min to pellet cells.

B. A standard PEG/NaCl precipitation protocol is performed by adding 1/5 volume of a 20% PEG/15% NaCl stock to culture supernatant; mixing and incubating on ice for 45 minutes -1 hour.

C. The culture is then centrifuged at 7000 rpm for 40 minutes to pellet phage and the supernatant is removed.

D. The phage pellet is resuspended in 1 mL TBS [pH 7.5]/2 mM CaCl₂ and transferred to an Eppendorf tube and centrifuged at 13 K rpm for at least 2 minutes to pellet insoluble material.

E. Supernatant is transferred to fresh tube and 1/5 volume of PEG/NaCl is added, vortexed and incubated on ice for 5 mins.

F. The supernatant is then centrifuged at 13 K rpm for at least 2 minutes, removed and purified phage is resuspended in up to 1 mL TBS [pH 7.5]/2 mM CaCl₂ and stored at 4°C.

II. Round 2 and Round 3 Panning

The 2nd and 3rd round panning conditions are generally the same as round 1 described above, except the coated target (i.e., c-Met) amount is decreased to 2 to 4 fold for each next round, and the plates (and beads) are washed more times for each next round of panning.

III. Optional Intra-Domain Recombination
and then the entire monomer is fused by overlapping PCR and amplified. Once the fragments are fused together and amplified the DNAs are then cut by a restriction enzyme to allow for cloning and ligated into cut fuse5 phage vector. Recombined monomer libraries are then panned two more rounds.

IV. Analysis of Panning Output (Same for Plate and Bead Protocols)

Phage ELISAs: For each output to be analyzed (typically Rounds 2, 3 and 4, if applicable), independent clones are inoculated in 1 mL (2×YT/20 µg/mL tetracycline) cultures grown in costar 96-well polypropylene deep-well plates, leaving inoculating tips in and shaken overnight at 37°C. Cells are pelleted by centrifugation at 3600 rpm for 15 minutes. Culture supernatants are retained and ELISAs are performed as discussed below.

96-well Maxisorp plate are coated with 50 µL/well of 50 µg/mL (2.5 µg/well) Streptavidin is diluted in TBS [pH 7.5]/2 mM CaCl₂ and the plate is incubated at 37ºC for 1 hr with shaking. Plates are washed three times with 200 µL/well of TBS [pH 7.5]/2 mM CaCl₂. Wells are blocked with 200 µL/well of 1% BSA (fraction V) and the covered plate is incubated at RT for 1 hr with shaking. The plate is washed three times with TBS [pH 7.5]/2 mM CaCl₂. The 96-well Maxisorp plate is coated with 100 µL/well of 1 µg/mL (0.1 µg/well) c-Met diluted in TBS [pH 7.5]/2 mM CaCl₂ or 100 µL/well buffer only (negative control). The plate is incubated at RT for 1 hr with shaking. Plates are washed three times with TBS [pH 7.5]/2 mM CaCl₂. Phage is added to wells. First, 70 µL of 1% Milk/0.2% BSA [pH 7.5]/2 mM CaCl₂/0.02% Tween-20 is added to the wells, then 30 µL of each phage supernatant is added from spun down overnight 1 mL culture to 1 positive (+SA-bn-CD40 L) well and to 1 negative (+SA only) well. Covered plates are incubated at RT for 1.5 hrs with shaking.

Plates are washed four times with TBS [pH 7.5]/2 mM CaCl₂/0.02% Tween-20. Note: 100 µL/well of α-M13-HRP detection antibody diluted 1:5000 in TBS [pH 7.5]/2 mM CaCl₂/0.02% Tween-20 is added. Plates are washed three times with cold TBS [pH 7.5]/2 mM CaCl₂/0.02% Tween-20. 100 µL/well of TMB/H₂O₂ mixture diluted 1:1 is added.

Color is allowed to turn blue until OD₄₂₀ reaches ~1.0. The reaction is stopped with 100 µL/well 2N H₂SO₄. Positive wells should change in color from blue to yellow. Once the reaction is stopped it can be read on ELISA plate reader OD₄₂₀ using SoftMaxPro software.

Selected phage are amplified overnight by infection of K91 cells. The phage are then purified by PEG/NaCl precipitation of culture supernatants. To clone the monomer or multimer sequences into the expression vector, pEve, approximately 10⁷ phage are amplified by 20 cycles of PCR as follows:

5 µL 10× Buffer
8 µL 2.5 mM dNTPs
5 µL 10 nM VS—For primer (5'-AACATCTGGCCGGTCCGCTACCTACGATATGTCCCGGA-3')
5 µL 10 nM EveNut primer (5'-AAAAAGGCCCCCA-GAGCGGCTTCTCGAATGAC-3')
1 µL Phage
26 µL H₂O
0.5 µL LA Taq (1 unit) TAKARA
Cycle 20×[94°C/10 sec.-50°C/30 sec.-72°C/30 sec.]

The PCR is then run on a 3% Agarose gel, the approximately 200 bp product is purified with a QIAquick column (Qiagen), digested with Sfi I (NEB), purified again with a QIAquick column and then ligated with T4 DNA Ligase (NEB) to the Sfi I digested vector, pEve. The ligation is used to transform electro-competent BL21 (DE3) E. coli and plated onto 2×YT plates containing Kanamycin at 40 µg/mL. Following growth, approximately 6000 individual clones are picked into 1×YT/Kan and grown overnight— included on the plate are wells that will serve as positive and negative controls.

Monomers may be purified in parallel in E. coli BL21DE3 (transformed with monomer-variants cloned in pETHC) in a parallel manner and subsequently purified by boiling.

Optionally, the selected pEve/monomer clone plasmid DNAs are digested in separate reactions with BsrDI and Bpml (NEB). The 10.1 kb BsrDI and 2.9 kb Bpml bands are isolated from 1% Agarose gels and purified with QIAquick columns. The DNAs are ligated together with T4 DNA Ligase (NEB). The ligation is purified with a QIAquick column and amplified as described for Phage Sub-cloning. The 400 bp product is then purified, ligated and transformed as described for Phage Sub-cloning. Protein can be expressed and purified from these and screened using the same methods described above.

Additional libraries referred to as “walking libraries” are generated by ligating the selected monomers with the full representation of random domains (generated using the same methods used to generate the monomer libraries). These “walking libraries” are then panned and screened as described above.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques, methods, compositions, apparatus and systems described above can be used in various combinations. All publications, patents, patent applications, or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, or other document were individually indicated to be incorporated by reference for all purposes.
1. A polypeptide comprising a monomer domain that binds to c-Met, wherein the monomer domain:
   - is a non-naturally-occurring monomer domain consisting of 30 to 50 amino acids;
   - comprises at least one disulfide bond; and
   - binds to an ion.
2. The polypeptide of claim 1, wherein binding of at least one monomer domain to c-Met inhibits dimerization of Met.
3. The polypeptide of claim 1, wherein at least one monomer domain binds to the Sena domain of c-Met, thereby preventing binding of Met ligands to c-Met.
4. The polypeptide of claim 1, wherein the polypeptide comprises at least one and no more than six monomer domains.
5. The polypeptide of claim 1, wherein the polypeptide comprises at least two monomer domains and the monomer domains are linked by a linker.
6. The polypeptide of claim 5, wherein the linker is a peptide linker.
7. The polypeptide of claim 6, wherein the linker is between 4 to 12 amino acids long.
8. The polypeptide of claim 1, wherein the monomer domains are each between 35 to 45 amino acids.
9. The polypeptide of claim 1, wherein the polypeptide comprises at least one monomer domain with binding specificity for a blood factor, thereby increasing the serum half-life of the polypeptide compared to a polypeptide lacking the blood factor monomer domain.
10. The polypeptide of claim 9, wherein the blood factor is serum albumin, an immunoglobulin or an erythrocyte.
11. The polypeptide of claim 1, wherein each monomer domain comprises two disulfide bonds.
12. The polypeptide of claim 11, wherein each monomer domain comprises three disulfide bonds.
13. The polypeptide of claim 1, wherein the ion is a metal ion.
14. The polypeptide of claim 1, wherein the ion is a calcium ion.
15. The polypeptide of claim 1, wherein at least one of the monomer domains is derived from a LDL-receptor class A domain.
16. The polypeptide of claim 1, wherein at least one of the monomer domains is derived from an EGF-like domain.
17. The polypeptide of claim 1, wherein the monomer comprises an amino acid sequence in which:
   - at least 10% of the amino acids in the sequence are cysteine; and/or
   - at least 25% of the amino acids are non-naturally-occurring amino acids.
18. A method for identifying a polypeptide that binds to c-Met, the method comprising,
   - screening a library of polypeptides for affinity to c-Met; and
   - selecting a polypeptide comprising at least one monomer domain that binds to c-Met, wherein the monomer domain:
     - is a non-naturally-occurring monomer domain;
     - comprises at least one disulfide bond; and
     - binds to an ion.
19. The method of claim 18, wherein the selecting step comprises selecting a polypeptide that reduces HGF-mediated cell proliferation and/or migration.
20. The method of claim 18, further comprising selecting a polypeptide that inhibits tumor growth in an animal.
21. The method of claim 18, wherein the monomer comprises an amino acid sequence in which:
   - at least 10% of the amino acids in the sequence are cysteine; and/or
   - at least 25% of the amino acids are non-naturally-occurring amino acids.
22. The method of claim 18, further comprising linking the monomer domain in the selected polypeptide to a second monomer domain to form a library of multimers, each multimer comprising at least two monomer domains;
   - screening the library of multimers for the ability to bind to c-Met; and
   - selecting a multimer that binds c-Met.
23. The method of claim 18, further comprising linking the monomer domain in the selected polypeptide to a second monomer domain to form a library of multimers, each multimer comprising at least two monomer domains;
   - screening the library of multimers for the ability to bind to a target molecule other than the c-Met; and
   - selecting a multimer that binds to the target molecule.
24. The method of claim 18, further comprising a step of mutating at least one monomer domain, thereby providing a library comprising mutated monomer domains.
25. The method of claim 18, wherein the library of monomer domains is expressed as a phage display, ribosome display or cell surface display.
26. The method of claim 18, wherein the polypeptide comprises at least two monomer domains and the monomer domains are linked by a linker.

27. The method of claim 26, wherein the linker is a peptide linker.

28. The method of claim 27, wherein the linker is between 4 to 12 amino acids long.

29. The method of claim 18, wherein the monomer domains are each between 35 to 45 amino acids.

30. The method of claim 18, wherein each monomer domain comprises two disulfide bonds.

31. The method of claim 18, wherein each monomer domain comprises three disulfide bonds.

32. The method of claim 18, wherein the ion is a metal ion.

33. The method of claim 18, wherein the ion is a calcium ion.

34. The method of claim 18, wherein at least one of the monomer domains is derived from a LDL-receptor class A domain.

35. The method of claim 18, wherein at least one of the monomer domains is derived from an EGF-like domain.

36. The method of claim 18, wherein the monomer domain comprises an amino acid sequence in which:

   at least 10% of the amino acids in the sequence are cysteine; and/or

   at least 25% of the amino acids are non-naturally-occurring amino acids.

37. A polynucleotide encoding the polypeptide of claim 1.

* * * * *