

(51) International Patent Classification:
C12N 15/10 (2006.01)(21) International Application Number:
PCT/EP2020/054697(22) International Filing Date:
21 February 2020 (21.02.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1902392.8 21 February 2019 (21.02.2019) GB
16/282,155 21 February 2019 (21.02.2019) US(71) Applicant: CAMBRIDGE ENTERPRISE LIMITED
[GB/GB]; The Old Schools, Trinity Lane, Cambridge Cam-
bridgeshire CB2 1TN (GB).(72) Inventors: ITZHAKI, Laura; Department of Pharmacol-
ogy, University of Cambridge, Tennis Court Road, Cam-
bridge Cambridgeshire CB2 1PD (GB). ROWLING, Pam;
Department of Pharmacology, University of Cambridge,
Tennis Court Road, Cambridge Cambridgeshire CB2 1PD
(GB). LADDS, Graham; Department of Pharmacology,
University of Cambridge, Tennis Court Road, Cambridge
Cambridgeshire CB2 1PD (GB). PEREZ RIBA, Alber-
to; Department of Pharmacology, University of Cambridge,
Tennis Court Road, Cambridge Cambridgeshire CB2 1PD
(GB). GOYENECHEA CORZO, Beatriz; PolyProx Ther-
apeutics Ltd, Building 910 (Jonas Webb), Babraham Re-
search Campus, Cambridge Cambridgeshire CB22 3AT
(GB). MABBITT, Joseph; PolyProx Therapeutics Ltd,
Building 910 (Jonas Webb), Babraham Research Campus,
Cambridge Cambridgeshire CB22 3AT (GB). BARDELLI,
Marco; PolyProx Therapeutics Ltd, Building 910 (Jonas
Webb), Babraham Research Campus, Cambridge Cam-
bridgeshire CB22 3AT (GB). GILBERT, Simon; PolyProx
Therapeutics Ltd, Building 910 (Jonas Webb), Babraham
Research Campus, Cambridge Cambridgeshire CB22 3AT
(GB).(74) Agent: MEWBURN ELLIS LLP; City Tower, 40 Basing-
hall Street, London Greater London EC2V 5DE (GB).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).**Published:**

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))

(54) Title: MODULAR BINDING PROTEINS

(57) Abstract: This invention relates to modular proteins that are capable of binding to one or more target molecules. The modular binding proteins comprise two or more repeat domains, such as tetratricopeptide repeat domains; inter-repeat loops linking the repeat domains; and one or more heterologous peptide ligands that bind to a target molecule. Each peptide ligand is located in an inter-repeat loop or at the N or C terminus of the modular binding protein and one or more of the heterologous peptide ligands comprises at least one amino acid sequence set out in Table 8 or Table 9 or a variant thereof. Modular binding proteins with various configurations and methods for their production and use are provided.



Modular Binding ProteinsField

This invention relates to modular binding proteins and their production and uses.

5 Background

A priority area in medicine, particularly cancer research, is the expansion of the 'druggable' proteome, which is currently limited to narrow classes of molecular targets. For example, protein-protein interactions (PPIs) are fundamental to all biological processes and represent a large proportion of potential drug targets, but they are not readily amenable to conventional small molecule inhibition. The architecture of tandem repeat proteins has tremendous scope for rational design (Kobe & Kajava 2000, Longo & Blaber, 2014, Rowling et al., 2015). The key features of tandem repeat proteins are relatively small size, modularity and extremely high stability (and therefore recombinant production) without the need of disulphide bonds. Individual consensus-designed repeats are self-compatible and can be put together in any order; function is therefore also modular, which means that multiple functions can be independently designed and incorporated in a combinatorial fashion within a single molecule (WO2017106728).

Novel repeat protein functions, e.g. DARPinS (Tamaskovic et al., 2012), have been developed based on the natural type of PPI interface of these proteins i.e. spanning many repeat units to create an extended, high-affinity binding interface for the target. Mutations have been introduced into the surface residues in the tetratricopeptide (TPR) repeats of the cytosolic receptor peroxin 5 (Sampathkumar et al. (2008) J. Mol. Biol., 381, 867-880). Binding of peptide ligands to peroxin 5 is shown to be mediated by residues located in several different TPR repeats. The interactions of TPR containing protein kinesin-1 with different cargo proteins has also been reported (Zhu et al PLoS One 2012 7 3 e33943). The specificity and stability of ankyrin repeat proteins has been modified through the introduction of mutations into ankyrin repeat sequences (Li et al (2006) Biochemistry 45 15168-15178).

Summary

The present inventors have found that modular proteins capable of binding to one or more target molecules can be generated by displaying peptidyl binding motifs, such as short linear motifs (SLiMs), on modular scaffolds. These modular binding proteins may be useful, for example as single- or multi-function protein therapeutics.

An aspect of the invention provides a modular binding protein comprising;

- (i) two or more repeat domains,
- (ii) inter-repeat loops linking said repeat domains; and
- (iii) one or more peptide ligands, each said peptide ligand being located in an inter-repeat loop or at the N or C terminus of the modular binding protein.

Preferably, the one or more peptide ligands comprise an amino acid sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617).

In other embodiments, a modular binding protein may comprise an amino acid sequence set out in Table 13 (SEQ ID NOs: 1230-1304), or a variant thereof.

5 In some preferred embodiments, the modular binding protein may comprise a first peptide ligand that binds a first target molecule and a second peptide ligand that binds a second target molecule. One of the first or second target molecules may be an E3 ubiquitin ligase. One of the first and second peptide ligands may comprise an amino acid sequence set out in Table 8 (SEQ ID NOs: 277-388) and the other of the first and second peptide ligands may comprise an amino acid sequence set out in Table 9 (SEQ ID NOs: 470-617).

Another aspect of the invention provides a method of producing a modular binding protein comprising; inserting a first nucleic acid encoding a peptide ligand into a second nucleic acid encoding two or more repeat domains linked by inter-repeat loops, to produce a chimeric nucleic acid encoding
15 a modular binding protein as described herein; and
expressing said chimeric nucleic acid to produce the modular binding protein.

Preferably the peptide ligand comprises an amino acid sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617).

20 Another aspect of the invention provides a method of producing a modular binding protein that binds to a first target molecule and a second target molecule comprising;
a nucleic acid encoding two or more repeat domains linked by inter-repeat loops, and
incorporating into said nucleic acid a first nucleotide sequence encoding a first peptide ligand
25 that binds to a first target molecule and a second nucleotide sequence encoding a second peptide ligand that binds to a second target molecule to generate a nucleic acid encoding a modular binding protein comprising said first and second peptide ligands, wherein said peptide ligands are located in an inter-repeat loop or at the N or C terminus of the modular binding protein; and
expressing the nucleic acid to produce said protein.

30 Preferably the first and second peptide ligands comprise amino acid sequences set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617).

In some preferred embodiments, one of the first or second target molecules is an E3 ubiquitin ligase.

35 Another aspect of the invention provides a library comprising modular binding proteins, each modular binding protein in the library comprising;

- (i) two or more repeat domains,
- (ii) inter-repeat loops linking said repeat domains; and
- 40 (iii) one or more peptide ligands, each said peptide ligand being located in an inter-repeat

loop or at the N or C terminus of the modular binding protein,
wherein at least one amino acid residue in the peptide ligands in said library is
diverse.

- 5 Preferably the one or more peptide ligands comprise an amino acid sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617).

Another aspect of the invention provides a library comprising a first and a second sub-library of modular binding proteins, each modular binding protein in the first and second sub-libraries

10 comprising;

- (i) two or more repeat domains,
- (ii) inter-repeat loops linking said repeat domains;
- (iii) a first peptide ligand comprising an amino acid sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and

- 15 (iv) a second peptide ligand comprising at least one diverse amino acid residue,
wherein the peptide ligand in the modular binding proteins in the first sub-library binds to a first target molecule and is located in one of (i) an inter-repeat loop; (ii) the N terminus or (iii) the C terminus of the modular binding protein, and

- 20 the peptide ligand in the modular binding proteins in the second sub-library binds to a second target molecule and is located in another of (i) an inter-repeat loop; (ii) the N terminus or (iii) the C terminus of the modular binding protein.

Another aspect of the invention provides a method of producing a library of modular binding proteins comprising;

- 25 (a) providing a population of nucleic acids encoding a diverse population of modular binding proteins comprising

- (i) two or more repeat domains,
- (ii) inter-repeat loops linking said repeat domains; and
- (iii) a first peptide ligand comprising an amino acid sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and

- 30 (iv) a second peptide ligand comprising at least one diverse amino acid residue, each said peptide ligand being located in an inter-repeat loop or at the N or C terminus of the modular binding protein, and

- 35 (b) expressing said population of nucleic acids to produce the diverse population, thereby producing a library of modular binding proteins.

Another aspect of the invention provides a method of screening a library comprising;

- (a) providing a library of modular binding proteins, each modular binding protein in the library comprising;

- 40 (i) two or more repeat domains,

- (ii) inter-repeat loops linking said repeat domains;
 - (iii) a first peptide ligand comprising an amino acid sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and
 - (iv) a second peptide ligand comprising at least one diverse amino acid residue,
- wherein the first and second peptide ligands are located in the inter-repeat loop, at the N terminus or at the C terminus of the protein,
- (b) screening the library for modular binding proteins which display a binding activity, and
 - (c) identifying one or more modular binding proteins in the library which display the binding activity.

Other aspects and embodiments of the invention are described in more detail below.

Brief Description of Figures

Figure 1 shows the thermostability of consensus-designed tetratricopeptide (CTPR) proteins containing loop- or helix-grafted binding motifs: Thermal denaturation, monitored by circular dichroism, of 2-repeat RTPR (a CTPR in which lysine residues have been replaced with arginine residues) proteins: RTPR2 (in diamonds), RTPR2 containing a loop binding-module (circles) and RTPR2 containing a helix binding-module (squares). All samples are at 20 μ M in 10 mM sodium phosphate buffer pH 7.4, 150 mM NaCl.

Figure 2 shows the thermostability of CTPR proteins of increasing length containing an increasing number of binding modules (alternating with blank modules): Thermal denaturation curves, monitored by circular dichroism, of TPR proteins containing 1, 2, 3 and 4 loops comprising a tankyrase-binding sequence: 1TBP-CTPR2, 2TBP-CTPR4, 3TBP-CTPR6, 4TBP-CTPR8. All samples are at 20 μ M in 10 mM sodium phosphate buffer pH 7.4, 150 mM NaCl.

Figure 3 shows an example of helix grafting. Figure 3A (i) shows the crystal structures of SOS1 (son-of-sevenless homologue 1) bound to KRAS (Kirsten rat sarcoma) (PDB 1NVU, Margarit et al. Cell (2003) 112(5):685-95), and (ii) shows the SOS1 helix grafted onto a helix at the N-terminus of a CTPR2 protein. The modelled structure of SOS-RTPR2 is shown, and the sequence of the helix is given with the key KRAS-binding residues in grey and the residues that form the interface with the CTPR helices in black. (iii) shows the modelled structure of SOS-TPR2 in complex with KRAS. Figure 3B shows binding of SOS-TPR2 to KRAS measured by competitive fluorescence polarization (FP). The complex between mant-GTP and KRAS was pre-formed, and 0.1-300 μ M SOS-RTPR2 was then titrated in to the complex, displacing the mant-GTP from KRAS resulting in a decrease in FP. EC₅₀ is 3 μ M.

Figure 4 shows another example of helix grafting. Figure 4A shows the modelled structure of the Mdm2 (Mouse double minute 2 homolog) N-terminal domain in complex with the p53-TPR2 comprising the Mdm2-binding helix of p53 grafted onto a helix at the C-terminus of a CTPR2 protein.

Figure 4B shows an ITC analysis of the interaction between p53-TPR2 and Mdm2 N-terminal domain. The N-terminal domain of Mdm2 was titrated into the cell containing 10µM p53-TPR2.

Figure 5 shows an example of single and multivalent loop-grafted CTPRs. Figure 5A shows an ITC analysis of the interaction between a series of tankyrase-binding loop-grafted CTPR2 proteins (TBP-CTPR2) and the substrate-binding ARC4 (ankyrin-repeat cluster) domain of tankyrase. There is an enhancement of both binding affinity and dissociation constant with increasing number of binding modules. Figure 5B shows native gel analysis (using a native gel in Tris-Glycine buffer pH 8.0, 40 µM protein concentration) of multivalent TBP-CTPR proteins expressed as fusion constructs with the foldon trimerisation domain (Boudko et al 2002; Meier et al. 2004). 1TBP-CTPR2, 2TBP-CTPR4 and 4TBP-CTPR8 (all lacking the foldon domain) were purified and run as monomeric controls. Constructs having the foldon domain run at much higher molecular weights than their monomeric counterparts.

Figure 6 shows an example of loop-grafted CTPRs comprising the 10-residue Skp2-binding sequence derived from p27 grafted into a loop of a CTPR protein (CTPR-p27).

Figure 6A shows that HA-CTPR2-p27 is able to co-IP FLAG-Skp2 from HEK293T cells.

Figure 6B shows *E. coli*-expressed and purified TPR5-p27 inhibits p27 ubiquitination in vitro.

Figure 7 shows another example of loop-grafted CTPRs. Figure 7A shows (left) ITC analysis of the interaction between the Keap1 (Kelch-like ECH-associated protein 1) KELCH domain and a CTPR2 protein containing a loop-grafted Keap1-binding sequence derived from the protein Nrf2 (Nuclear factor (erythroid-derived 2)-like 2) (Nrf-CTPR2). No binding is observed for the blank CTPR2 protein (right). Figure 7B shows that three variants of Nrf-CTPR2 (Nrf-CTPR2 (i), Nrf-CTPR2 (ii), Nrf-CTPR2 (iii)) can co-IP Keap1 from HEK293T cells.

Figure 8 shows live-cell imaging of intracellular delivery of an RTPR achieved by resurfacing (by introducing Arginine residues at surface sites). PC3 (left) and U2OS (right) cells incubated with 10 µM FITC-labelled resurfaced TBP-RTPR2 for 3 hours at 37°C, 5% CO₂. Overlay of DIC (differential interference contrast) and confocal image. Intracellular fluorescence was also observed at lower concentrations of protein.

Figure 9 shows the induced degradation of the target protein beta-catenin by designed hetero-bifunctional RTPRs. Figure 9A shows the beta-catenin levels in cells transfected with either HA-tagged beta-catenin plasmid alone or HA-tagged beta-catenin plasmid together with one of two different hetero-bifunctional RTPR plasmids (LRH1-TPR-p27 and axin-TPR-p27, designed to bind simultaneously to beta-catenin and to E3 ligase SCF^{Skp2}). Figure 9B shows a quantitative analysis of the beta-catenin levels in the presence of different hetero-bifunctional RTPRs designed to bind simultaneously to beta-catenin and to either E3 ligase SCF^{Skp2} or E3 ligase Mdm2. The analysis was performed using densitometry of the bands detected by Western blots corresponding to HA-tagged beta-catenin normalised to actin bands using ImageJ. Negative controls used were single-function

TPRs or blank (non-functional) TPRs.

Figure 10 shows examples of different modular binding protein formats. A modular binding protein may comprise: two repeat domains with a helical target-binding peptide and a helical E3-binding peptide at the N and C termini (Figure 10A); three repeat domains with a helical E3-binding peptide at the C terminus and a target peptide ligand in the first inter-repeat loop from the N terminus (Figure 10B); three repeat domains with a helical target-binding peptide at the N terminus and an E3 peptide ligand in the second inter-repeat loop from the N terminus (Figure 10C), four repeat domains with a target-peptide ligand and an E3 peptide ligand in the first and third inter-repeat loop from the N terminus (Figure 10D).

Figure 11 shows a schematic of a modular binding protein with four peptide ligands located in alternate inter-repeat loops. The binding sites are arrayed at 90° to each other.

Figure 12 shows a schematic of a modular binding protein engineered so that peptide ligands in alternate inter-repeat loops bind adjacent epitopes on the target.

Figure 13 shows the modelled structure of a hetero-bifunctional modular binding protein comprising TPR repeat domains, an LRH1-derived peptide ligand designed to bind target beta-catenin, and a p53-derived N-terminal peptide ligand designed to bind to the E3 ubiquitin ligase mdm2.

Figure 14 shows a schematic of the combinatorial assembly of a module comprising a repeat domain and a terminal helical peptide ligand and a module comprising repeat domains and an inter-repeat loop peptide ligand to generate a modular binding protein.

Figure 15 shows examples of different modular binding protein formats. (i) shows the blank proteins; (ii) shows binding peptides inserted into one or more inter-repeat loops. (iii) shows helical binding peptides at one or both of the termini; (iv) is a combination of loop and helical binding peptides; (v) and (vi) show examples of how multivalency can be achieved.

Figure 16 shows a schematic of the assembly of a modular binding protein by the progressive screening of modular binding proteins comprising modules with a diverse peptide ligand in addition to modules already identified in previous rounds of screening.

Figure 17 shows the effect of designed multi-valent tankyrase-binding TPR proteins on Wnt signalling. HEK293T cells were transfected with TPR-encoding plasmids using Lipofectamine2000. The TPR proteins contained 1-4 copies of a tankyrase-binding peptide (TBP) grafted onto the inter-repeat loop(s). For example, 2TBP-CTPR4 is a protein comprising 4 TPR modules with one TBP grafted onto the loop between the first and second TPR and one between the third and fourth TPR. 'Foldon' indicates a trimeric TPR-foldon fusion protein.

Figure 18 shows characterisation of the size and charge of liposome-encapsulated TPR proteins.

Figure 19 shows the delivery of TPR proteins into cells by liposome encapsulation. FITC dye-labelled liposomes stain the cell membrane upon membrane fusion (red panel), and RITC-labelled TPR protein cargo is then delivered into the cytoplasm. The green panel and red-green merge show that the proteins have entered the cells and are spread diffusely in the cytoplasm.

Figure 20 shows that liposome-encapsulated TPR proteins are not toxic to HEK293T cells at the concentrations used.

Figure 21 shows the effect of designed TPR proteins delivered by liposome encapsulation. The TPR proteins contained a tankyrase-binding peptide. Cells were treated with liposomes for 2 hr.

Figure 22 shows the effect of designed TPR proteins delivered by liposome encapsulation. The TPR proteins contained a tankyrase-binding peptide. Cells were treated with liposomes encapsulating 32 µg protein for variable times (2-8 h) indicated in the figure.

Figure 23 shows the effect of designed hetero-bifunctional TPR proteins on KRAS levels in HEK 293T cells. The TPR proteins contained a binding sequence for KRAS (a non-helical peptide sequence, referred to as KBL, grafted onto an inter-repeat loop of the RTPR) and a degron derived from p27 grafted onto another inter-repeat loop. Cells were transiently transfected with 50 ng or 500ng of TPR encoding plasmids, as indicated, and with KRAS plasmid or empty vector as control. 24 hours post transfection the cells were lysed, and KRAS levels were evaluated by western blot. In dark grey are cells treated transfected with single-function TPR plasmid (containing degron only).

Figure 24 shows the effect of designed hetero-bifunctional TPR proteins (delivered by liposome encapsulation) on KRAS levels. The TPR proteins contained a KRAS-binding peptide and a SCF^{Skp2}-binding peptide to direct KRAS for ubiquitination and subsequent degradation. Cells were treated with liposomes for 2 hr.

Figure 25 shows the effect of designed hetero-bifunctional TPR proteins (delivered by liposome encapsulation) on KRAS levels. The TPR proteins contained a KRAS-binding peptide and a SCF^{Skp2}-binding peptide to direct KRAS for ubiquitination and subsequent degradation. Cells were treated with TPR protein for variable times (2-8 h) indicated in the figure.

Figure 26 shows the effect of hetero-bifunctional TPR proteins targeting endogenous KRAS to the CMA (chaperone-mediated autophagy) pathway. The TPR proteins contained a binding sequence for KRAS (either a grafted helix derived from son-of-sevenless-homolog 1 (SOS) or a non-helical peptide sequence (referred to as 'KBL') displayed in a loop of the RTPR) and targeted for degradation using

two different chaperone-mediated autophagy peptides (referred to as 'CMA_Q' or 'CMA_K') at the N- or C-terminus of the construct. Constructs or empty vector (light grey) were transiently transfected into either HEK293T or DLD1 (colorectal cancer cell line). 24 hours post transfection the cells were lysed, and KRAS levels were evaluated by western blot. Those constructs that resulted in significant reduction in KRAS compared to the empty vector control are shown in white.

Figure 27 shows examples of variations in the linker sequence connecting a peptide ligand to an inter-repeat loop in order to optimise the binding affinity for the target. The example shown is Nrf-TPR, a TPR protein designed to bind to Keap1 (see Fig. 7 of the original patent application). Glycine residues were introduced into the linker to provide flexibility and increased spatial sampling. The introduction of this more flexible linker sequence was found to increase the binding affinity of the Nrf-TPR protein (labelled 'Flexible') when compared with the consensus-like linker sequence. Altering the charge content of the linker sequence ('labelled 'Charged') and altering the conformational properties (based on the predictions of the program CIDER (Holehouse *et al.* Biophys. J. 112, 16-21 (2017)) of the loop by changing the amino acid composition of the linker sequence (labelled 'CIDER-optimised') also affected the Keap1-binding affinity.

Figure 28 shows the results of a screen for Beta Catenin degradation by bispecific CPTR constructs using the "Phospho" peptide ligand that binds to Beta Catenin. The structures and sequences the CPTR constructs are shown in Tables 12 and 13.

Figure 29 shows the results of a screen for Beta Catenin degradation by bispecific CPTR constructs using the "AXIN" peptide ligand that binds to Beta Catenin. The structures and sequences the CPTR constructs are shown in Tables 12 and 13.

Figure 30 shows the results of a screen for Beta Catenin degradation by bispecific CPTR constructs using the "BCL9" peptide ligand that binds to Beta Catenin. The structures and sequences the CPTR constructs are shown in Tables 12 and 13.

Figure 31 shows the results of a screen for Beta Catenin degradation by bispecific CPTR constructs using the "LRH1" peptide ligand that binds to Beta Catenin. The structures and sequences the CPTR constructs are shown in Tables 12 and 13.

Detailed Description

This invention relates to the modular binding proteins that comprise multiple repeat domains. These repeat domains are linked to each other in the polypeptide chain by inter-repeat loops. One or more peptide ligands or binding domains, are located in one or more of the inter-repeat loops and/or in N or C terminal helices of the modular binding protein. The peptide ligands may be to the same or different target molecules and the modular binding protein may be multi-functional and/or multi-valent. The geometrical display of the grafted binding sites may be precisely and predictably tuned by adjusting

the positions of the binding sites and the number and shape of the repeat domains. Modular binding proteins as described herein may be useful in a range of therapeutic and diagnostic applications.

A repeat domain is a repetitive structural element of 30 to 100 amino acids that forms a defined secondary structure. Multiple repeat domains stack sequentially in a modular fashion to form a stable protein, which may for example have a solenoid or toroid structure. Repeat domains may be synthetic or may be naturally-occurring repeats from tandem repeat proteins, or variants thereof.

Due to the identical form of their building blocks, solenoid domains can only assume a limited number of shapes. Two main topologies are possible: linear (or open, generally with some degree of helical curvature) and circular (or closed). Patthy, László (2007). Protein Evolution. Wiley-Blackwell. ISBN 978-1-4051-5166-5.

If the two terminal repeats in a solenoid do not physically interact, it leads to an open or linear structure. Members of this group are frequently rod- or crescent-shaped. The number of individual repeats can range from 2 to over 50. A clear advantage of this topology is that both the N- and C-terminal ends are free to add new repeats and folds, or even remove existing ones during evolution without any gross impact on the structural stability of the entire domain. Kinch LN, Grishin NV (June 2002). Curr. Opin. Struct. Biol. 12 (3): 400–8. doi:10.1016/s0959-440x(02)00338-x. PMID 12127461.

This type of domain is extremely common among extracellular segments of receptors or cell adhesion molecules. A non-exhaustive list of examples include: EGF repeats, cadherin repeats, leucine-rich repeats, HEAT repeats, ankyrin repeats, armadillo repeats, tetratricopeptide repeats, etc. Whenever a linear solenoid domain structure participates in protein-protein interactions, frequently at least 3 or more repetitive subunits form the ligand-binding sites. Thus - while individual repeats might have a (limited) ability to fold on their own – they usually cannot perform the functions of the entire domain alone.

In the case when the N- and C-terminal repeats lie in close physical contact in a solenoid domain, the result is a topologically compact, closed structure. Such domains typically display a high rotational symmetry (unlike open solenoids that only have translational symmetries), and assume a wheel-like shape. Because of the limitations of this structure, the number of individual repeats is not arbitrary. In the case of WD40 repeats (perhaps the largest family of closed solenoids) the number of repeats can range from 4 to 10 (more usually between 5 and 7). (Vogel C, Berzuini C, Bashton M, Gough J, Teichmann SA (February 2004). J. Mol. Biol. 336 (3): 809–23). Kelch repeats, beta-barrels and beta-trefoil repeats are further examples for this architecture.

A repeat domain may have the structure of a solenoid repeat. The structures of solenoid repeats are well known in the art (see for example Kobe & Kajava Trends in Biochemical Sciences 2000; 25(10):509-15). For example, a repeat domain may have an α/α or $\alpha/3_{10}$ (helix-turn-helix or hth) structure, for example a tetratricopeptide repeat structure; $\alpha/\alpha/\alpha$ (helix-turn-helix-turn-helix or hthh)

structure, for example an armadillo repeat structure; a $\beta/\beta/\alpha/\alpha$ structure; a α/β or $3_{10}/\beta$ structure, for example a leucine rich repeat (LRR) structure; a $\beta/\beta/\beta$ structure, for example, an IGF1RL, HPR or PeIC repeat structure; or a β/β structure, for example a serralysin or EGF repeat structure.

- 5 A “scaffold” refers to two or more repeat domains, and a “grafted scaffold” refers to a chimeric protein, which is a continuous polypeptide comprising a scaffold and a heterologous binding site (e.g., a peptide ligand).

10 Ankyrin repeat, one of the most widely existing protein motifs in nature, consists of 30-34 amino acid residues and exclusively functions to mediate protein-protein interactions, some of which are directly involved in the development of human cancer and other diseases. Each ankyrin repeat exhibits a helix-turn-helix conformation, and strings of such tandem repeats are packed in a nearly linear array to form helix-turn-helix bundles with relatively flexible loops. The global structure of an ankyrin repeat protein is mainly stabilized by intra- and inter-repeat hydrophobic and hydrogen bonding interactions.

15 The repetitive and elongated nature of ankyrin repeat proteins provides the molecular bases of the unique characteristics. Examples of ANK repeats suitable for use as described herein are shown in Table 14.

Consensus Sequences for ANK repeats (SMART database, see Table 14) include the following:

20 O04242/1-30 NGHTALHIAASK-----GDEQCVKLLLEHGA-----DPNA (SEQ ID NO: 1)

CONSENSUS/80% .t.sslhhsh.t.....tp.phhphllp.t.....pht (SEQ ID NO: 2)

CONSENSUS/65% pstosLphAstp.....sphphlphLlptss.....shsh (SEQ ID NO: 3)

CONSENSUS/50% sGpTsLHhAsps.....sshdllchLlspus.....slst (SEQ ID NO: 4)

25

The armadillo (Arm) repeat is an approximately 40 amino acid long tandemly repeated sequence motif first identified in the *Drosophila melanogaster* segment polarity gene armadillo involved in signal transduction through wingless. Animal Arm-repeat proteins function in various processes, including intracellular signaling and cytoskeletal regulation, and include such proteins as beta-catenin, the junctional plaque protein plakoglobin, the adenomatous polyposis coli (APC) tumour suppressor protein, and the nuclear transport factor importin-alpha, amongst others [(PUBMED:9770300)].

30 Examples of Arm repeats suitable for use as described herein are shown in Table 15.

Consensus sequences for ARM repeats (SMART database, see Table 15) include the following:

35 IMO2HUMANb PND-KIQAVIDAG-VCRRLLVELLM-----HNDYKVVSPALRA (SEQ ID NO: 5)

CONSENSUS/80% pt...h...hhp.t...hl...lhphlt.....p.pl.t.shhs (SEQ ID NO: 6)

40 CONSENSUS/65% ssp.ptphlhpts...slshLlpLLp.....pts.plhptsshs

(SEQ ID NO: 7)

CONSENSUS/50% `ssc.sppsllcsG.slstLlpLLs.....sscsclppsAstA`

(SEQ ID NO: 8)

5 IMO2HUMANb VGNIVT (SEQ ID NO: 9)

CONSENSUS/80% `ltpls.` (SEQ ID NO: 10)CONSENSUS/65% `LpNlst` (SEQ ID NO: 11)CONSENSUS/50% `LsNlus` (SEQ ID NO: 12)

- 10 Suitable repeat domains may include domains of the Ankyrin clan (Pfam: CL0465), such as ankyrin (PF00023), which may comprise a 30-34 amino-acid repeat composed of two beta strands and two alpha helices; domains of the leucine-rich repeat (LRR) clan (Pfam; CL0022), such as LRR1 (PF00560), which may comprise a 20-30 amino acid repeat composed of an α/β horseshoe fold; domains of the Pec Lyase-like (CL0268) clan, such as pec lyase C (PF00544), which may comprise a
- 15 right handed beta helix; domains of the beta-Roll (CL0592) clan such as Haemolysin-type calcium-binding repeat (PF000353), which may comprise short repeat units (e.g. 9-mers) that form a beta-roll made up of a super-helix of beta-strand-turns of two short strands each, stabilised by Ca^{2+} ions; domains of the PSI clan (CL0630), such as trefoil (PF00088); and domains of the tetratricopeptide clan (CL0020), such as TPR-1 (PR00515), which may comprise a 24 to 90 amino acid repeat
- 20 composed of a helix-turn-helix.

Suitable repeat domains may be identified using the PFAM database (see for example Finn et al Nucleic Acids Research (2016) Database Issue 44:D279-D285).

- 25 In some preferred embodiments, the repeat domain may have the structure of an α/α -solenoid repeat domain, such as a helix-turn-helix. A helix-turn-helix domain comprises two antiparallel α -helices of 12-45 amino acids.

- Suitable helix-turn-helix domains include tetratricopeptide-like repeat domains. Tetratricopeptide-like
- 30 repeats may include domains of the TPR clan (CL0020), for example and Arm domains (see for example Armadillo; PF00514; Huber et al Cell 1997;90: 871-882), HEAT domains (Huntingtin, EF3, PP2A, TOR1; PF02985; see for example Groves et al. Cell. 96 (1): 99-110), PPR domains (pentatricopeptide repeat PF01535; see for example Small (2000) Trends Biochem. Sci. 25 (2): 46-7), TALE domains (TAL (transcription activator-like) effector; PF03377; see for example Zhang et al
- 35 Nature Biotechnology. 29 (2): 149-53) and TPR1 domains (tetratricopeptide repeat-1; PF00515; see for example Blatch et al BioEssays. 21 (11): 932-9).

Other suitable helix-turn-helix domain may be synthetic, for example DHR1 to DHR83 as disclosed in Brunette et al., Nature 2015 528 580-584.

In some preferred embodiments, the helix-turn-helix scaffold may be a tetratricopeptide repeat domain (TPR) (D'Andrea & Regan, 2003) or a variant thereof. TPR repeat domains may include naturally occurring or synthetic TPR domains. Suitable TPR repeat domains are well known in the art (see for example Parmeggiani et al., *J. Mol. Biol.* 427 563-575) and may have the amino acid sequence:

AEAWYNLGNAYYKQGDYQKAIEYYQKALEL-X₁X₂X₃X₄ (SEQ ID NO: 13)

wherein X₁₋₄ are independently any amino acid, preferably X₁ and X₂ being D and P respectively, or may be a variant of this sequence. Other TPR repeat domain sequences are shown in Tables 4-6 and 11 below.

Additional TPR repeat consensus sequences (SMART database, see Table 11 (SEQ ID NOs: 816-1229) include the following:

S75991 ALTLNNIGTI YYAREDYDQA LNYEQALSL SRAV (SEQ ID NO: 14)

CONSENSUS/80% XXhhXthuXh hXXXtphppA htXhpps1tht XpX (SEQ ID NO: 15)

CONSENSUS/65% spshhphGth hhphsphppA lphappAlpl pspX (SEQ ID NO: 16)

CONSENSUS/50% spsatslGps atptucaccA lcsap+ALcl sPss (SEQ ID NO: 17)

Other TPR repeat domain sequences are shown in Tables 4 (SEQ ID NOs 168-181), 6 (SEQ ID NOs 221-243) and 11 (SEQ ID NOs 811-1233) below.

The grouping of amino acids to classes and class abbreviation (the key) used within consensus sequences are shown below

Class	Key	Residues
alcohol	o	S,T
aliphatic	l	I,L,V
any	.	A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
aromatic	a	F,H,W,Y
charged	c	D,E,H,K,R
hydrophobic	h	A,C,F,G,H,I,K,L,M,R,T,V,W,Y
negative	-	D,E
polar	p	C,D,E,H,K,N,Q,R,S,T
positive	+	H,K,R
small	s	A,C,D,G,N,P,S,T,V
tiny	u	A,G,S
turnlike	t	A,C,D,E,G,H,K,N,Q,R,S,T

Preferred TPR domains may include CTPR, RTPRa, RTPRb and KTPRb domains, for example a domain having a sequence shown in Table 4 (SEQ ID NOs 170-183) or Table 6 (SEQ ID NOs: 223-

246) or Table 11 (SEQ ID NOs: 816-1229) or a variant of a sequence shown in Table 4 (SEQ ID NOs 170-183) or Table 6 (SEQ ID NOs: 223-246) or Table 11 (SEQ ID NOs: 816-1229).

In some embodiments, a TPR repeat domain may be a human TPR repeat domain, preferably a TPR repeat domain from a human protein in blood. TPR repeat domains from human blood may have reduced immunogenicity *in vivo*. Suitable human blood TPR repeat domains may include repeat domains from IFIT1, IFIT2 or IFIT3. Other examples of human blood repeat domains identified in the plasma proteome database are shown in Table 5 (SEQ ID NOs; 184-222).

Suitable human blood repeat domains may be identified from the plasma proteome database (Nanjappa et al Nucl Acids Res 2014 Jan;42(Database issue):D959-65) for example by searching for sequences with high sequence identity to the TPR repeat domain using standard sequence analysis tools (e.g. Altschul et al Nucleic Acids Res. 25:3389-3402; Altschul et al FEBS J. 272:5101-5109).

In some embodiments, the two or more repeat domains of a modular binding protein described herein may comprise a small, glutamine-rich, tetratricopeptide repeat protein alpha (SGTA) domain. The STGA domain is involved in protein quality control pathways and comprises three TPR repeat domains and a C-terminal capping helix. The STGA domain binds to the Rpn13 substrate receptor of the proteasome. A modular binding protein comprising the STGA domain may bind to proteasome receptors through the STGA domain. One or more peptide ligands for a target molecule grafted into the STGA domain of a modular binding protein as described herein may allow the target molecule to be delivered directly to the proteasome for degradation by the modular binding protein.

A suitable STGA domain may have the sequence:

EDSAEAERLKTEGNEQMKVENFEAAVHFYGKAIELNPANAVYFCNRAAAYSKLGN YAGAVQDCERAICI
DPAYSKAYGRMGLALSSLNKHVEAVAYYKKALELDPDNETYKSNLKIAELKLEAP

(SEQ ID NO: 18) (domains TPR1-A, TPR1-B, TPR2-A, TPR1-B, TPR3-A, TPR3-B
 (sequentially) solid underlined, C terminal helix dotted underlined)

or may be a variant of this sequence.

A suitable STGA domain may be encoded by the nucleotide sequence:

gaagatagcgcggaagcgggaacgcctgaaaaccgaaggcaacgaacagatgaaagtggaaaactttgaa
 gcggcggtgcatttttatggcaaagcgattgaactgaacccggcgaacgcggtgtatttttgcaaccgc
 gcggcggtatagcaaactgggcaactatgcggcgcggtgcaggattgcgaacgcgcgatttgcatt
 gatccggcggtatagcaaagcgatggccgcctggcgctgagcagcctgaacaaacatgtggaa
 gcgggtggcgattataaaaaagcgctggaactggatccggataacgaaacctataaaagcaacctgaaa
 attgcgggaactgaaactgcgcgaagcgccg (SEQ ID NO: 19)

A variant of a reference repeat domain or peptide ligand sequence set out herein may comprise an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% sequence identity to the reference sequence. Particular amino acid sequence variants may differ from a repeat domain shown above by insertion, addition, substitution or

deletion of 1 amino acid, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more than 10 amino acids. Preferred variants of a TPR repeat domain may comprise one or more conserved residues, for example, 1, 2, 3, 4, 5, 6 or more preferably all of Leu at position 7, Gly or Ala at position 8, Tyr at position 11, Ala at position 20, Ala at position 27, Leu or Ile at positions 28 and 30 and Pro at position 32.

5

Sequence similarity and identity are commonly defined with reference to the algorithm GAP (Wisconsin Package, Accelrys, San Diego USA). GAP uses the Needleman and Wunsch algorithm to align two complete sequences that maximizes the number of matches and minimizes the number of gaps. Generally, default parameters are used, with a gap creation penalty = 12 and gap extension
 10 penalty = 4. Use of GAP may be preferred but other algorithms may be used, e.g. BLAST (which uses the method of Altschul *et al.* (1990) *J. Mol. Biol.* 215: 405-410), FASTA (which uses the method of Pearson and Lipman (1988) *PNAS USA* 85: 2444-2448), or the Smith-Waterman algorithm (Smith and Waterman (1981) *J. Mol Biol.* 147: 195-197), or the TBLASTN program, of Altschul *et al.* (1990) *supra*, generally employing default parameters. In particular, the psi-Blast algorithm (Nucl. Acids Res.
 15 (1997) 25 3389-3402) may be used. Computerized implementations of these algorithms (GAP, BESTFIT, PASTA, and FASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI) are available and publicly available computer software may be used such as ClustalOmega (Söding, J. 2005, *Bioinformatics* 21, 951-960), T-coffee (Notredame *et al.* 2000, *J. Mol. Biol.* (2000) 302, 205-217), Kalign (Lassmann and Sonnhammer 2005, *BMC*
 20 *Bioinformatics*, 6(298)), Genomequest™ software (Gene-IT, Worcester MA USA) and MAFFT (Katoh and Standley 2013, *Molecular Biology and Evolution*, 30(4) 772-780 software. When using such software, the default parameters, e.g. for gap penalty and extension penalty, are preferably used. A preferred example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*,
 25 *Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990), respectively.

Sequence comparison may be made over the full-length of the relevant sequence described herein.

30

For example, a repeat domain may comprise one or more point mutations to facilitate grafting of hydrophobic peptide ligands. For example, aromatic residues in the repeat domain may be substituted for polar or charged residues. Suitable substitutions may be identified in a rational manner, for example using Hidden Markov plots of repeat domain sequences to identify non-aromatic residues that are found in nature in consensus aromatic positions. A suitable TPR repeat domain for grafting
 35 hydrophobic peptide ligand may have the amino acid sequence:

AEAWYNLGNAYYRQGDYQRAIEYYQRALEL-X₁X₂ X₃X₄ (SEQ ID NO: 20)

wherein X₁₋₄ are independently any amino acid, preferably X₁ and X₂ being D and P respectively.

40

In some embodiments, lysine residues in the repeat domain may be replaced by arginine residues to

prevent ubiquitination and subsequent degradation. This may be particularly useful when the modular binding protein comprises an E3 ubiquitin ligase-peptide ligand, for example in a proteolysis targeting chimera (PROTAC). For example, a suitable TPR repeat domain may have the amino acid sequence:

AEALNNLGNVYREQGDYQRAIEYYQRALEL-X₁X₂ X₃X₄ (SEQ ID NO: 21)

- 5 wherein X₁₋₄ are independently any amino acid, preferably X₁ and X₂ being D and P respectively.

In some embodiments, a modular binding protein may comprise a scaffold with the amino acid sequence of residues 1 to 171 of SEQ ID NO: 1230 (PPX172 of Table 13 without the HA Tags). A target peptide ligand and the E3 ligase peptide ligand may be grafted into any two of the N terminal helical grafting site 1 (corresponding to residues 1 to 2 of SEQ ID No 1230; before first CTPR repeat); loop grafting site 2 (corresponding to residues 35 to 36 of SEQ ID No 1230; between first and second CTPR repeat); loop grafting site 3 (corresponding to residues 69 to 70 of SEQ ID NO: 1230m between second and third CTPR repeat); loop grafting site 4 (corresponding to residues 103 to 104 of SEQ ID No 1230; between third and fourth CTPR repeat), loop grafting site 5 (corresponding to residues 137 to 138 SEQ ID No 1230; between fourth and fifth CTPR modules) and loop grafting site 6 (corresponding to residues 167 to 168 SEQ ID No 1230; after last CTPR repeat) DPNN peptides may be placed before and after a peptide ligand to force a turn in the structure.

20 In preferred embodiments, the modular binding protein may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10 or more than 10 repeat domains. Preferably, the modular binding protein comprises 2 to 5 repeat domains. Modular binding proteins with fewer repeat domains may display increased cell penetration. For example, a modular binding protein with 2-3 repeat domains may be useful in binding intracellular target molecule. Modular binding proteins with more repeat domains may display increased stability and functionality. For example, a modular binding protein with 4 or more repeat domains may be useful in binding extracellular target molecules. A modular binding protein with 6 or more repeat domains may be useful in producing long linear molecules for targeting or assembling extracellular complexes in bi- or multivalent formats.

30 In other embodiments, sufficient stability and functionality may be conferred by a single repeat domain with N and C terminal peptide ligands. For example, a modular binding protein may comprise:

- (i) a repeat domain, and
- (ii) peptide ligands at the N and C terminal of the repeat domain.

35 The repeat domains of a modular binding protein may lack binding activity i.e. the binding activity of the modular binding protein is mediated by the peptide ligands and not by residues within the repeat domains.

A peptide ligand is a contiguous amino acid sequence that specifically binds to a target molecule.

40 Suitable peptide ligands that are capable of grafting onto a terminal helix or inter-repeat loop are well-

known in the art and include peptide sequences selected from a library, antigen epitopes, natural protein-protein interactions (helical, extended or turn-like) and short linear motifs (SLiMs). Viral SLiMs (that hijack the host machinery) may be particularly useful because they may display high binding affinities (Davey et al (2011) Trends Biochem. Sci. 36,159-169).

5

A suitable peptide ligand for a target molecule may be selected from a library, for example using phage or ribosome display, or identified or designed using rational approaches or computational design, for example using the crystal structure of a complex or an interaction. In some embodiments, peptide ligands may be identified in an amino acid sequence using standard sequence analysis tools (e.g. Davey et al Nucleic Acids Res. 2011 Jul 1; 39(Web Server issue): W56–W60).

10

Peptide ligands may be 5 to 25 amino acids in length, preferably 8 to 15 amino acids, although in some embodiments, longer peptide ligands may be employed.

15

The peptide ligands and the repeat domains of the modular binding protein are heterologous i.e. the peptide ligand is not associated with the repeat domain in naturally occurring proteins and the binding and repeat domains are artificially associated in the modular binding protein by recombinant means.

20

A modular binding protein described herein may comprise 1 to $n+1$ peptide ligands, where n is the number of repeat domains in the modular binding protein. The number of peptide ligands is determined by the required functionality and valency of the modular binding protein. For example, one peptide ligand may be suitable for a mono-functional modular binding protein and two or more peptide ligands may be suitable for a bi-functional or multi-functional modular binding protein.

25

Modular binding proteins may be monovalent. A target molecule may be bound by a single peptide ligand in a monovalent modular binding protein. Modular binding proteins may be multivalent. A target molecule may be bound by two or more of the same or different peptide ligands in a multivalent modular binding protein.

30

Modular binding proteins may be monospecific. The peptide ligands in a monospecific modular binding protein may all bind to the same target molecule, more preferably the same site or epitope of the target molecule.

35

Modular binding proteins may be multi-specific. The peptide ligands in a multi-specific modular binding protein may bind to different target molecules. For example, a bi-specific modular binding protein may comprise one or more peptide ligands that bind to a first target molecule and one or more peptide ligands that bind to a second target molecule and a tri-specific modular binding protein may comprise one or more peptide ligands that bind to a first target molecule, one or more peptide ligands that bind to a second target molecule and one or more peptide ligands that bind to third target molecule.

40

A bi-specific modular binding protein may bind to the two different target molecules concurrently. This may be useful in bringing the first and second target molecules into close proximity. When the target molecules are located on different cells, concurrent binding of the target molecules to the modular binding protein may bring the cells into close proximity, for example to promote or enhance the interaction of the cells. For example, a modular binding protein which binds to a tumour specific antigen and a T cell antigen, such as CD3, may be useful in bringing T cells into proximity to tumour cells. When the target molecules are from different biological pathways, this may be useful in achieving synergistic effects and also for minimising resistance.

A tri-specific modular binding protein may bind to three different target molecules concurrently. In some embodiments, one of the target molecules may be an E3 ubiquitin ligase. For example, tri-specific modular binding protein may bind to a first target molecule from a first biological pathway and a second target molecule from a second biological pathway as well as an E3 ubiquitin ligase. This may be useful in achieving synergistic effects and also for minimising resistance.

A peptide ligand may be located in an inter-repeat loop of the modular binding protein.

An inter-repeat peptide ligand may comprise 5 to 25 amino acid residues, preferably 8 to 15 amino acids. However, since there is no intrinsic restriction on the size of the inter-loop peptide ligand, longer sequences of more than 25 amino acid residues may be used in some embodiments.

In some embodiments, an unstructured peptide ligand may be inserted into an inter-repeat loop.

One or more, two or more, three or more, four or more or five or more of the inter-repeat loops in the modular binding protein may comprise peptide ligands. The peptide ligands may be located on consecutive inter-repeat loops or may have a different distribution in the inter-repeat loops of the modular binding protein. For example, inter-repeat loops comprising a peptide ligand may be separated in the modular protein by one or more, two or more, three or more or four or more inter-repeat loops which lack a peptide ligand.

A peptide ligand may be connected to an inter-repeat loop directly or via one or more additional residues or linkers. Additional residues or linkers may be useful for example when a peptide ligand requires conformational flexibility in order to bind to a target molecule, or when the amino acid residues that are adjacent to the minimal peptide ligand favourably influence the micro-environment of the binding interface.

Additional residues or linkers may be positioned at the N terminus of the peptide ligand, the C terminus of the peptide ligand, or both. For example, the sequence of an inter-repeat loop containing a peptide ligand may be $[X_{1-1}]-[X_{1-n}]-[X_{1-2}]$ (SEQ ID NO: 22), where each residue denoted by X is

independently any amino acid and may be the same amino acid or a different amino acid to any other residue that is also denoted by X, $[X_{1-n}]$ is the peptide ligand, n is 1 to 100, $[X_{1-i}]$ is a linker and *i* is independently any number between 1 to 10. In some embodiments, D may be preferred at the first position of the linker $[X_{1-i}]$, P may be preferred at the second position of linker $[X_{1-i}]$, D may be preferred at the last position of the linker $[X_{1-z}]$ and/or P may be preferred at the penultimate position of linker $[X_{1-z}]$. Examples of preferred inter-repeat loop sequences may include DP- $[X_{1-n}]$ -PX; DPXX- $[X_{1-n}]$ -XXPX (SEQ ID NO: 23); DPXX- $[X_{1-n}]$ -XPXX (SEQ ID NO: 24); DPXX- $[X_{1-n}]$ -PXXX (SEQ ID NO: 25); PXX- $[X_{1-i}]$ - $[X_{1-n}]$ - $[X_{1-i}]$ -XXPX (SEQ ID NO: 26), DPXX- $[X_{1-i}]$ - $[X_{1-n}]$ - $[X_{1-i}]$ -XPXX (SEQ ID NO: 27), DPXX- $[X_{1-i}]$ - $[X_{1-n}]$ - $[X_{1-i}]$ -PXXX (SEQ ID NO: 28), DPXX- $[X_{1-i}]$ - $[X_{1-n}]$ -XPXX (SEQ ID NO: 29), DPXX- $[X_{1-i}]$ - $[X_{1-n}]$ -XPXX (SEQ ID NO: 30), DPXX- $[X_{1-i}]$ - $[X_{1-n}]$ -XPXX (SEQ ID NO: 31), DPXX- $[X_{1-n}]$ - $[X_{1-i}]$ -XXPX (SEQ ID NO: 32), DPXX- $[X_{1-n}]$ - $[X_{1-i}]$ -XPXX (SEQ ID NO: 33) and DPXX- $[X_{1-n}]$ - $[X_{1-i}]$ -PXXX (SEQ ID NO: 34).

The precise sequence of the residues or linkers used to connect a peptide ligand to an inter-repeat loop depends on the peptide ligand and may be readily determined for any peptide ligand of interest using standard techniques. For example, small, non-hydrophobic amino acids, such as glycine, may be used to provide flexibility and increased spatial sampling, for example when a peptide ligand needs to adopt a specific conformation, or proline residues may be used to increase rigidity, for example, when the peptide ligands are short.

In some preferred embodiments, an inter-repeat peptide ligand may be non-hydrophobic. For example, at least 40% of the amino acids in the peptide ligand may be charged (e.g. D, E, R or K) or polar (e.g. Q, N, H, T, Y, C or W). Alternatively, the repeat domains may be modified to accommodate a hydrophobic peptide ligand, for example by replacing aromatic residues with charged or polar residues.

A peptide ligand may be located at one or both termini of the modular binding protein.

In some embodiments, a peptide ligand located at the N or C terminus may comprise an α -helical structure and may comprise all or part of a half-repeat (i.e. all or part of a single α -helix) that stacks against the adjacent repeat domain. The α -helix of the terminal peptide ligand makes stabilising interactions with the adjacent repeat domain and is stable and folded. Only a few of the positions that structurally define an α -helix are required for the correct interfacial interaction with the adjacent repeat domain. The residues in some of these positions are defined (Tyr (*i*) - Ile (*i*+4) - Tyr (*i*+7) - Leu (*i*+11) for the N-terminal α -helix and Ala (*i*) - Leu (*i*+4) - Ala/Val (*i*+7) for the C-terminal helix), but the remaining positions of the α -helix may be modified to form a helical peptide ligand.

A helical peptide ligand may be located at the N terminus of the protein. The N terminal peptide ligand may be helical and may comprise all or part of the sequence $X_n(X)_{15}-X_1X_2XX$ (SEQ ID NO: 35), preferably all or part of the sequence $X_nXYXXXIXXYXXLXX-X_1X_2XX$ (SEQ ID NO: 36), where each

residue denoted by X is independently any amino acid and may be the same amino acid or a different amino acid to any other residue in the sequence that is also denoted by X, X_1 is independently any amino acid, preferably D, and X_2 is independently any amino acid, preferably P, and n is 0 or any number. In some embodiments, the Y, I, and/or L residues in the N terminal peptide ligand may be substituted for an amino acid residue with similar properties (i.e. a conservative substitution).

A helical peptide ligand may be located at the C terminus of the protein. The C terminal peptide ligand may be helical and may comprise all or part of the sequence $X_n(X)_{15}X_1X_2XX$ (SEQ ID NO: 35), preferably all or part of the sequence $X_1X_2XX-XXAXXXLXX[A$ (SEQ ID NO: 37) or $V]XXXXX-X_n$ (SEQ ID NO: 38), where X is independently any amino acid and may be the same amino acid or a different amino acid to any other residue in the sequence that is also denoted by X, X_1 is independently any amino acid, preferably D, and X_2 is independently any amino acid, preferably P, and n is 0 or any number. In some embodiments, the A, L and/or V residues in the C terminal peptide ligand may be substituted for an amino acid residue with similar properties (i.e. a conservative substitution).

The minimum length of the terminal peptide ligand is determined by the number of residues required to form a helix that binds to the target molecule. There is no intrinsic maximum length of the terminal peptide ligand and n may be any number.

In other embodiments, a peptide ligand located at the N or C terminus may comprise a non-helical structure. For example, a peptide ligand that is an obligate N- or C- terminal domain (for example because the terminal amino or carboxylate group mediates the binding interaction) may be located at the beginning or end of the one or more repeat domains.

In some embodiments, one or more positions in a peptide ligand may be diverse or randomised. A modular binding protein comprising one or more diverse or randomised residues may form a library as described below.

In some embodiments, the N and C terminal peptide ligands may be non-hydrophobic. For example, at least 20% of the amino acids in the peptide ligand may be charged (e.g. D, E, R or K) or polar (e.g. Q, N, H, T, Y, C or W). Alternatively, the helix turn helix scaffold of the repeat domains may be modified, for example by replacing aromatic residues with charged or polar residues in order to accommodate a hydrophobic peptide ligand.

A modular binding protein as described herein may comprise peptide ligands in any arrangement or combination. For example, peptide ligands may be located at both the N and C terminus and optionally one or more inter-repeat loops of a modular binding protein; at the N terminus and optionally one or more loops of a modular binding protein; at the C terminus and optionally one or more loops of a modular binding protein; or in one or more inter-repeat loops of a modular binding protein.

The location of the peptide ligands within a modular binding protein may be determined by rational design, for example using modelling to identify the optimal arrangement for the presentation of two target molecules to each other (e.g. for substrate presentation to an E3 ubiquitin ligase); and/or by
5 screening for example using populations of modular binding proteins with different arrangements of peptide ligands to identify the arrangement which confers the optimal interaction of target molecules.

Suitable target molecules for modular binding proteins described herein include biological macromolecules, such as proteins. The target molecule may be a receptor, enzyme, antigen,
10 oligosaccharide, oligonucleotide, integral membrane protein, transcription factor, transcriptional regulator, G protein coupled receptor (GPCR) or any other target of interest. Proteins that are difficult to target with small molecules, such as PPIs, proteins that accumulate in neurodegenerative diseases and proteins overexpressed in disease conditions, such as cancer, may be particularly suitable target molecules. Target molecules may include α -synuclein; β -amyloid; tau; superoxide dismutase;
15 huntingtin; β -catenin; KRAS; components of superenhancers and other types of transcriptional regulators, such as N-Myc, C-Myc, Notch, aurora A, EWS-FLI1 (Ewing's sarcoma-friend leukemia integration 1), TEL-AML1, TAL1 (T-cell acute lymphocytic leukemia protein 1) and Sox2 ((sex determining region Y)-box 2); tankyrases; phosphatases such as PP2A; epigenetic writers, readers and erasers, such as histone deacetylases and histone methyltransferases; BRD4 and other
20 bromodomain proteins; and kinases, such as PLK1 (polo-like kinase 1), c-ABL (Abelson murine leukemia viral oncogene homolog 1) and BCR (breakpoint cluster region)-ABL.

In some embodiments, a modular binding protein may neutralise a biological activity of the target molecule, for example by inhibiting or antagonising its activity or binding to another molecule or by
25 tagging it for ubiquitination and proteasomal degradation or for degradation via autophagy. In other embodiments, a modular binding protein may activate a biological activity of the target molecule.

In some embodiments, the target molecule may be β -catenin. Suitable peptide ligands that specifically bind to β -catenin are well-known in the art and include β -catenin-peptide ligands derived from axin
30 (e.g. GAYPEYILDIHVYRVQLEL (SEQ ID NO: 39) and variants thereof), Bcl-9 (e.g. SQQLEHRYRSLITLYDIQLML (SEQ ID NO: 40) and variants thereof), TCF7L2 (e.g. QELGDNDLMHFSYESTQD (SEQ ID NO: 41) and variants thereof), ICAT (e.g. YAYQRAIVEYMLRLMS (SEQ ID NO: 42) and variants thereof), LRH-1 (e.g. YEQAIAAYLDALMC (SEQ ID NO: 43) and variants thereof) or APC (e.g. SCSELEALEALELDE (SEQ ID NO: 44) and
35 variants thereof).

In some embodiments, the target molecule may be KRAS. Suitable peptide ligands that specifically bind to KRAS are well-known in the art and include a KRAS-peptide ligand from SOS-1 (e.g. FEGIALTNYLKALEG (SEQ ID NO: 45) and variants thereof) and KRAS-peptide ligands identified by
40 phage display (see for example Sakamoto et al. Biochem. Biophys. Res. Comm. (2017) 484 605-

611).

In some embodiments, the target molecule may be tankyrase. Suitable peptide ligands that specifically bind to tankyrase are well-known in the art and include tankyrase peptide ligands from Axin (e.g. REAGDGEE (SEQ ID NO: 46) and HLQREAGDGEEFRS (SEQ ID NO: 47) or variants thereof).

In some embodiments, the target molecule may be EWS-FLI1. Suitable peptide ligands that specifically bind to EWS-FLI1 are well-known in the art and include the ESAP1 peptide TMRGKKKRTRAN (SEQ ID NO: 48) and variants thereof. Other suitable sequences may be identified by phage display (see for example Erkizan et al. Cell Cycle (2011) 10, 3397-408).

In some embodiments, the target molecule may be Aurora-A. Suitable peptide ligands that specifically bind to Aurora-A are well-known in the art and include Aurora-A binding sequences from TPX2, such as SYSYDAPSDFINFSS (SEQ ID NO: 49) (Bayliss et al. Mol. Cell (2003) 12, 851-62) and Aurora-A binding sequences from N-myc, such as N-myc residues 19-47 or 61-89 (see for example Richards et al. PNAS (2016) 113, 13726-31).

In some embodiments, the target molecule may be N-Myc or C-Myc. Suitable peptide ligands that specifically bind to N-myc or C-myc are well-known in the art and include helical binding sequences from Aurora-A (see for example Richards et al. PNAS (2016) 113, 13726-31).

In some embodiments, the target molecule may be WDR5 (WD repeat-containing protein 5). Suitable peptide ligands that specifically bind to WDR5 are well-known in the art and include the WDR5-interacting motif (WIN) of MLL1 (mixed lineage leukemia protein 1) (see for example Song & Kingston J. Biol. Chem. (2008) 283, 35258-64; Patel et al. J. Biol. Chem. (2008) 283, 32158-61), e.g. EPPLNPHGSARAEVHLRKS (SEQ ID NO: 50) and variants thereof.

In some embodiments, the target molecule may be BRD4 or a Bromodomain protein. Suitable peptide ligands that specifically bind to BRD4 are well-known in the art and include sequences derived from histone protein ligands.

In some embodiments, the target molecule may be a HDAC (histone deacetylase). Suitable peptide ligands that specifically bind to HDAC are well-known in the art and include binding sequences derived from SMRT and other proteins that recruit HDACs to specific transcriptional regulatory complexes or binding sequences derived from histone proteins (see for example Watson et al. Nat. Comm. (2016) 7, 11262; Dowling et al. Biochem. (2008) 47, 13554-63).

In some embodiments, the target molecule may be Notch. Suitable peptide ligands that specifically bind to Notch are well-known in the art and include binding sequences from the N-terminus of MAML1 (mastermind like protein 1), e.g. SAVMERLRRRIELCRRHHST (SEQ ID NO: 51) and variants thereof

(see for example Moellering et al. Nature (2009) 462, 182-8).

In some embodiments, the target molecule may be a Cdk (cyclin-dependent kinase). Suitable peptide ligands that specifically bind to Cdks are well-known in the art and include substrate-based peptides, for example, Cdk2 sequences derived from cyclin A, such as TYTKKQVLRMEHLVLKVLTFDL (SEQ ID NO: 52) and variants thereof (see for example Gondeau et al. J. Biol. Chem. (2005) 280, 13793-800; Mendoza et al. Cancer Res. (2003) 63, 1020-4).

In some embodiments, the target molecule may be PLK1 (polo-like kinase 1). Suitable peptide ligands that specifically bind to PLK1 are well-known in the art and include optimised substrate-derived sequences that bind to the substrate-binding PBD (polo-box domain), such as MAGPMQSEPLMGAKK (SEQ ID NO: 53) and variants thereof.

In some embodiments, the target molecule may be Tau. Suitable peptide ligands that specifically bind to Tau are well-known in the art and include tau-binding sequences derived from alpha- and beta-tubulin, such as KDYEEVGVDSE (SEQ ID NO: 54) and YQQYQDATADEQG (SEQ ID NO: 55) and variants thereof (see for example Maccioni et al. EMBO J. (1988) 7, 1957-63; Rivas et al. PNAS (1988) 85, 6092-6).

In some embodiments, the target molecule may be BCR-ABL. Suitable peptide ligands that specifically bind to BCR-ABL are well-known in the art and include optimized substrate-derived sequences, such as EAIYAAPFAKKK (SEQ ID NO: 56) and variants thereof.

In some embodiments, the target molecule may be PP2A (protein phosphatase 2A). Suitable peptide ligands that specifically bind to PP2A are well-known in the art and include sequences that bind the B56 regulatory subunit, such as LQTIQEEE (SEQ ID NO: 57) and variants thereof (see for example Hetz et al. Mol. Cell (2016), 63 686-95).

In some embodiments, the target molecule may be EED (Embryonic ectoderm development). Suitable peptide ligands that specifically bind to EED are well-known in the art and include helical binding sequences from co-factor EZH2 (enhancer of zeste homolog 2), such as FSSNRQKILERTEILNQEWKQRRIQPV (SEQ ID NO: 58) and variants thereof (see for example Kim et al. Nat. Chem. Biol. (2013) 9, 643-50.)

In some embodiments, the target molecule may be MCL-1 (induced myeloid leukemia cell differentiation protein). Suitable peptide ligands that specifically bind to MCL-1 are well-known in the art and include sequences from BCL2, e.g. KALETLRVGDGVQRNHETAF (SEQ ID NO: 59) and variants thereof (see for example Stewart et al. Nat. Chem. Biol. (2010) 6, 595-601).

In some embodiments, the target molecule may be RAS. Suitable RAS peptide ligands are well-known in the art and include RAS-binding peptides identified by phage display, such as

RRRRCPLYISYDPVCRRRR (SEQ ID NO: 60) and variants thereof (see for example Sakamoto et al. BBRC (2017) 484, 605-11).

In some embodiments, the target molecule may be GSK3 (glycogen synthase kinase 3). Suitable GSK3 peptide ligands are well-known in the art and include substrate-competitive binding sequences such as KEAPPAPPQDP (SEQ ID NO: 61), LSRRPDYR (SEQ ID NO: 62), RREGGMSRPADVGD (SEQ ID NO: 63), and YRRAAVPPSPSLSRHSSPSQDEDEEE (SEQ ID NO: 64) and variants thereof (see for example Ilouz et al. J. Biol. Chem. 281 (2006), 30621–30630. Plotkin et al. J. Pharmacol. Exp. Ther. (2003) 305, 974–980).

In some embodiments, the target molecule may be CtBP (C-terminal binding protein). Suitable CtBP peptide ligands are well-known in the art and include sequences identified from a cyclic peptide library screen, such as SGWTVVRMY (SEQ ID NO: 65) and variants thereof (see for example Birts et al. Chem. Sci. (2013) 4, 3046-57).

Examples of suitable peptide ligands for target molecules that may be used in a modular binding protein as described herein are shown in Tables 2 and 7. Preferred peptide ligands are shown in Table 8 (SEQ ID NOs: 277-388).

In some preferred embodiments, a modular binding protein as described herein may comprise a peptide ligand for an E3 ubiquitin ligase. Examples of suitable E3 ubiquitin ligases include MDM2, SCF^{Skp2}, BTB-CUL3-RBX1, APC/C, SIAH, CHIP, Cul4-DDB1, SCF-family, β -TrCP, Fbw7 and Fbx4.

Suitable peptide ligands for E3 ubiquitin ligases (degrons) are well known in the art and may be 5 to 20 amino acids. For example, a suitable peptide ligand for MDM2 may include a peptide ligand from p53 (e.g. FAAYWNLLSAYG (SEQ ID NO: 66)) and or a variant thereof. A suitable peptide ligand for SCF^{Skp2} may include a peptide ligand from p27 (e.g. AGSNEQEPKKRS (SEQ ID NO: 67)) and variants thereof. A suitable peptide ligand for Keap1-Cul3 may include a peptide ligand from Nrf2 (e.g. DPETGEL (SEQ ID NO: 68)) or a variant thereof. A suitable peptide ligand for SPOP-Cul3 may include a peptide ligand from Puc (e.g. LACDEVTSTTSSSTA (SEQ ID NO: 69)) or a variant thereof. A suitable peptide ligand for APC/C may include the degrons termed ABBA (e.g. SLSSAFHVFEDGNKEN (SEQ ID NO: 70)), KEN (e.g. SEDKENVPP (SEQ ID NO: 71)), or DBOX (e.g. PRLPLGDVSNN (SEQ ID NO: 72)) or a variant thereof. In some instances, a combination of these degrons for may be used (mimicking the bipartite or tripartite degrons found in some natural substrates). A suitable peptide ligand for SIAH may include a peptide ligand from PHYL (e.g. LRPVAMVRPTV (SEQ ID NO: 73)) or a variant thereof. A suitable peptide ligand for CHIP (carboxyl terminus of Hsc70-interacting protein) may include peptide sequences such as ASRMEEVD (SEQ ID NO: 74) (from Hsp90 C-terminus) and GPTIEEVD (SEQ ID NO: 75) (from Hsp70 C-terminus) or a variant thereof. A suitable peptide ligand for β -TrCP may include a degron sequence motif (including phosphomimetic amino acids), such as DDGYFD (SEQ ID NO: 76) or a variant thereof. A

suitable peptide ligand for Fbx4 may include sequences derived from TRF1, such as MPIFWKAHRMSKMG TG (SEQ ID NO: 77) or a variant thereof (see for example Lee et al. *ChemBiochem* (2013) 14, 445-451). A suitable peptide ligand for FBw7 may include degron sequence motifs (including phosphomimetic amino acids), such as LPSGLLEPPQD (SEQ ID NO: 78). A suitable peptide ligand for DDB1-Cul4 may include sequences derived from HBx (hepatitis B virus X protein) and similar proteins from other viruses and from DCAFs (DDB1-CUL4-associated factors) including helical motifs such as ILPKVLHKRTLGL (SEQ ID NO: 79), NFVSWHANRQLGM (SEQ ID NO: 80), NTVEYFTSQQVTG (SEQ ID NO: 81), and NITRDLIRRQIKE (SEQ ID NO: 82) (see for example Li et al. *Nat. Struct. Mol. Biol.* (2010) 17, 105-111).

Examples of suitable peptide ligands for E3 ubiquitin ligases that may be used in a modular binding protein as described herein are shown in Table 3. Preferred peptide ligands for E3 ubiquitin ligases are shown in Table 9 (SEQ ID NOs: 470-617).

A modular binding protein comprising a peptide ligand for an E3 ubiquitin ligase may also comprise a peptide ligand for a target molecule. Binding of the modular binding protein to both the target molecule and the E3 ubiquitin ligase may cause the target molecule to be ubiquitinated by the E3 ubiquitin ligase. Ubiquitinated target molecules are then degraded by the proteasome. This allows the specific targeting of molecules for proteolysis by the modular binding protein. The ubiquitination and subsequent degradation of a target protein has been shown for hetero-bifunctional small molecules (PROTACs; proteolysis targeting chimeras) that bind the target protein and a ubiquitin ligase simultaneously (see for example Bondeson et al. *Nat. Chem. Biol.* 2015; Deshaies 2015; Lu et al. 2015).

In some embodiments, the modular binding protein may lack lysine residues, so that it avoids ubiquitination by the E3 ubiquitin ligase. Examples of modular binding proteins that bind E3 ubiquitin ligase and a target molecule are shown in Tables 1 and 8.

A suitable modular binding protein may comprise an N terminal peptide ligand that binds a target protein, such as β catenin, and a C terminal peptide ligand that binds an E3 ubiquitin ligase. For example, the N terminal peptide ligand may be a β catenin-binding sequence derived from Bcl9 and the C terminal peptide ligand may be an Mdm2-binding sequence derived from p53. Alternatively, a modular binding protein may comprise a C terminal peptide ligand that binds a target protein, such as β catenin, and an N terminal peptide ligand that binds an E3 ubiquitin ligase (see figure 10A).

Another suitable modular binding protein may comprise three repeat domains, a peptide ligand located in an inter-repeat loop that binds a target protein, such as β catenin, and a C terminal peptide ligand that binds an E3 ubiquitin ligase. For example, the inter-repeat loop peptide ligand may be derived from the phosphorylated region of APC (adenomatous polyposis coli) and the C terminal peptide ligand may be an Mdm2-binding sequence derived from p53. Alternatively, the modular

binding protein may comprise a peptide ligand located in an inter-repeat loop that binds an E3 ubiquitin ligase, and a C terminal peptide ligand that binds a target protein, such as β catenin (See figure 10B).

- 5 Another suitable modular binding protein may comprise three repeat domains, an N terminal peptide ligand that binds a target protein, such as β catenin, and a peptide ligand located in an inter-module loop that binds an E3 ubiquitin ligase. For example, the N terminal peptide ligand may be a β catenin-binding sequence derived from LRH1 (liver receptor homolog 1) and the inter-module loop peptide ligand may be a sequence derived from the Skp2-targeting region of p27. Alternatively, the modular binding protein may comprise an N terminal peptide ligand that binds an E3 ubiquitin ligase and a peptide ligand located in an inter-module loop that binds a target protein, such as β catenin (see figure 10C).

- 15 Another suitable modular binding protein may comprise four repeat domains, a first peptide ligand located in an inter-repeat loop that binds an E3 ubiquitin ligase and a second peptide ligand located in an inter-repeat loop that binds a target molecule. The first and second inter-repeat loops may be separate by an inter-repeat loop lacking a peptide ligand. For example, the first peptide ligand may be located in the first inter-repeat loop from the N terminus and the second peptide ligand may be located in the third inter-repeat loop from the N terminus or vice versa.

- 20 In some preferred embodiments, a modular binding protein as described herein may comprise an amino acid shown in Table 10 (SEQ ID NOs: 766-815) or a variant thereof.

- 25 In other preferred embodiments, a modular binding protein as described herein may comprise a peptide ligand that binds to a component of a target-selective autophagy pathway, such as chaperone-mediated autophagy (CMA). The modular binding protein and target molecules bound thereto are thus recognised by the autophagy pathway and the target molecules are subsequently degraded. Suitable components of the CMA pathway include heat shock cognate protein of 70 kDa (hsc70, HSPA8, Gene ID: 3312). Suitable peptide ligands are well known in the art (Dice J.F. (1990). Trends Biochem. Sci. 15, 305-309) and include Lys-Phe-Glu-Arg-Gln (KFERQ) and variants thereof, such as CMA_Q and CMA_K, as described herein. These domains have been demonstrated to be capable of targeting heterologous proteins to the autophagy pathway (Fan, X. et al; (2014) Nature Neuroscience 17, 471-480).

- 35 In some embodiments, a modular binding protein may comprise a scaffold with the amino acid sequence of residues 1 to 171 of SEQ ID NO: 1230 (PPX172 of Table 13 without the HA Tag). A target peptide ligand and the E3 ligase peptide ligand may be grafted into any two of the N terminal helical grafting site 1 and loop grafting sites 2 to 6. Preferably, the target peptide ligand is located in helical grafting site 1, loop grafting site 2, loop grafting site 4 or loop grafting site 5 of the scaffold.
- 40 Preferably, the E3 ligase peptide ligand or other degron may be located in any of loop grafting sites 2

to 6.

For example, the target peptide ligand may be in site 1 and the E3 ligase peptide ligand may be in site 4; the target peptide ligand may be in site 2 and the E3 ligase peptide ligand may be in site 5, the target peptide ligand may be in site 5 and the E3 ligase peptide ligand may be in site 2; the target peptide ligand may be in site 1 and the E3 ligase peptide ligand may be in site 6, the target peptide ligand may be in site 2 and the E3 ligase peptide ligand may be in site 6, the target peptide ligand may be in site 3 and the E3 ligase peptide ligand may be in site 6, the target peptide ligand may be in site 4 and the E3 ligase peptide ligand may be in site 6; the target peptide ligand may be in the site 5 and the E3 ligase peptide ligand may be in site 6; the target peptide ligand may be in site 1 and the E3 ligase peptide ligand may be in site 3; and the target peptide ligand may be in site 2 and the E3 ligase peptide ligand may be in site 4.

Suitable peptide ligands include beta catenin binding ligands, for example helical beta-catenin binding sequence from the protein AXIN, such as ILXXHV and variants thereof; peptides from APC (Adenomatous polyposis coli; "phospho"), such as SEELEALELELDE or GSEELEALELELDEA and variants thereof; peptides from LRH1, such as YEQAIAAYLDALMC and variants thereof, peptides from BCL9, such as TLXXIQXXL, LXTLXXIQ, and SLXXIXXML and variants thereof, and KRAS binding ligands, for example alpha-helical sequences from the protein SOS1 (Son of sevenless homolog 1), such as TNXXKXXE and variants thereof.

Suitable E3 ligase peptide ligands include SCFSkp2 binding sequences from p27, such as AGSNEQEPNR and variants thereof, Cul3-KEAP1 binding sequences from NRF2, such as LDPETGEL and variants thereof, E3 Cul3-SPOP binding sequences from Puc, such as DEVTSTTSS, and variants thereof, SIAH binding sequences from PHYL, such as LRPVAMVRPWVR, and variants thereof, COP1 binding sequences from Trib, such as SDQIVPEYQE, and variants thereof, UBR5 binding sequences from PAM2, such as LSVNAPEFYF, and variants thereof, beta-TRCP binding sequences from CDC25B, such as TEEDDGFVDI, and variants thereof, and MDM2 binding sequences from p53, such as FSXXWXXL and variants thereof.

Examples of modular binding proteins as described herein are shown in Tables 12 and 13. For example, a modular binding protein as described herein may comprise an amino acid sequence shown in Table 13 (SEQ ID NOs: 1230-1304) or a variant of an amino acid sequence shown in Table 13. Variants of a reference sequence are described above and may include variants in which the one or both of the target peptide ligand and the E3 ligase peptide ligand in a reference amino acid sequence of Table 13 are replaced by a different peptide ligand. Suitable peptide ligands are described below. Variants may also include variants in which the scaffold sequence in a reference amino acid sequence of Table 13 is replaced by a different scaffold sequence. Suitable scaffold sequences are described above.

In addition to repeat domains and peptide ligands, a modular binding protein may further comprise one or more additional domains which confer additional functionality, such as targeting domains, intracellular transport domains, stabilising domains or oligomerisation domains. Additional domains may for example be located at the N or C terminus of the modular binding protein or in a loop between repeats.

A targeting domain may be useful in targeting the modular binding protein to a particular destination in vivo, such as a target tissue, cell, membrane or intracellular organelle. Suitable targeting domains include chimeric antigen receptors (CARs).

An intracellular transport domain may facilitate the passage of the modular binding protein through the cell membrane into cells, for example to bind intracellular target molecules. Suitable intracellular transfer domains are well known in the art (see for example Bechara et al FEBS Letters 587 1 (2013) 1693-1702) and include cell-penetrating peptides (CPPs), such as Antennapedia (43–58), Tat (48–60), Cadherin (615–632) and poly-Arg.

A stabilising domain may increase the half-life of the modular binding protein in vivo. Suitable stabilising domains are well known in the art and include Fc domains, serum albumin, unstructured peptides such as XTEN⁹⁸ or PAS⁹⁹ and polyethylene glycol (PEG).

An oligomerisation domain may facilitate the formation of multi-protein complexes, for example to increase avidity against multi-valent targets. Suitable oligomerisation domains include the 'foldon' domain, the natural trimerisation domain of T4 fibrin (Meier et al., J. Mol. Biol. (2004) 344(4):1051-69).

In addition to repeat domains, peptide ligands and optionally one or more additional domains, a modular binding protein may further comprise a cytotoxic or therapeutic agent and/or or detectable label.

Suitable cytotoxic agents include, for example, chemotherapeutic agents, such as methotrexate, auristatin adriamycin, doxorubicin, melphalan, mitomycin C, ozogamicin, chlorambucil, maytansine, emtansine, daunorubicin or other intercalating agents, enzymatically active toxins of bacterial, fungal, plant, or animal origin, such as diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain, ricin A chain, abrin A chain, modeccin A chain, α -amanitin, alpha-sarcin, Aleurites fordii proteins, tubulysins, dianthin proteins, Phytolacca americana proteins (PAPI, PAPII, and PAP-S), Momordica charantia inhibitor, curcumin, croton, Saponaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, pyrrolbenzodiazepines, and the tricothecenes and fragments of any of these. Suitable cytotoxic agents may also include radioisotopes. A variety of radionuclides are available for the production of radioconjugated modular binding proteins including, but not limited to, ⁹⁰Y, ¹²⁵I, ¹³¹I, ¹²³I, ¹¹¹In, ¹³¹In, ¹⁰⁵Rh, ¹⁵³Sm, ⁶⁷Cu, ⁶⁷Ga, ¹⁶⁶Ho, ¹⁷⁷Lu, ¹⁸⁶Re, ¹⁸⁸Re and ²¹²Bi. Conjugates

of a modular binding protein and one or more small anti-cancer molecules, for example toxins, such as a calicheamicin, maytansinoids, a trichothene, and CC1065, and the derivatives of these toxins that have toxin activity, may also be used.

- 5 Suitable therapeutic agents may include cytokines (e.g. IL2, IL12 and TNF), chemokines, pro-coagulant factors (e.g. tissue factor), enzymes, liposomes, and immune response factors.

A detectable label may be any molecule that produces or can be induced to produce a signal, including but not limited to fluorescers, radiolabels, enzymes, chemiluminescers or photosensitizers.

- 10 Thus, binding may be detected and/or measured by detecting fluorescence or luminescence, radioactivity, enzyme activity or light absorbance. Detectable labels may be attached to modular binding proteins using conventional chemistry known in the art.

- 15 There are numerous methods by which the label can produce a signal detectable by external means, for example, by visual examination, electromagnetic radiation, heat, and chemical reagents. The label can also be bound to another specific binding member that binds the modular binding protein, or to a support.

- 20 In some embodiments, a modular binding protein may be configured for display on a particle or molecular complex, such as a cell, ribosome or phage, for example for screening and selection. A suitable modular binding protein may further comprise a display moiety, such as phage coat protein, to facilitate display on a particle or molecular complex. The phage coat protein may be fused or covalently linked to the modular binding protein.

- 25 Modular binding proteins as described herein may be produced by recombinant means. For example, a method of producing a modular binding protein as described herein may comprise expressing a nucleic acid encoding the modular binding protein. A nucleic acid may be expressed in a host cell and the expressed modular binding protein may then be isolated and/or purified from the cell culture.

- 30 In some embodiments, a method may comprise;
 inserting a first nucleic acid encoding a peptide ligand into a second nucleic acid encoding a two or more repeat domains to produce a chimeric nucleic acid encoding a modular binding protein comprising a peptide ligand located in an inter-repeat loop or at the N or C terminus of the modular binding protein; and,

- 35 expressing said chimeric nucleic acid to produce the modular binding protein.

- Preferably, a third nucleic acid encoding a second peptide ligand is inserted into the second nucleic acid to produce a chimeric nucleic acid encoding a chimeric protein comprising a first peptide ligand located at one of the first, second, third and fourth loops of the fibronectin scaffold and a second
40 peptide ligand located at another of the first, second, third and fourth loops of the fibronectin scaffold.

One of the first and second peptide ligands may comprise a sequence set out in Table 8 (SEQ ID NOs: 277-388) and the other of the first and second peptide ligands may comprise a sequence set out in Table 9 (SEQ ID NOs: 470-617).

5

Methods described herein may be useful in producing a modular binding protein that binds to a first target molecule and a second target molecule. For example, a method may comprise;

providing a nucleic acid encoding two or more repeat domains linked by inter-repeat loops, each repeat domain; and

10

incorporating into said nucleic acid a first nucleotide sequence encoding a first peptide ligand that binds to a first target molecule and a second nucleotide sequence encoding a second peptide ligand that binds to a second target molecule to generate a nucleic acid encoding a modular binding protein comprising said first and second peptide ligands, wherein said peptide ligands are located in an inter-repeat loop or at the N or C terminus of the modular binding protein; and

15

expressing the nucleic acid to produce said protein

One of the first and second target molecules may be an E3 ubiquitin ligase. For example, a method may comprise;

20

providing a nucleic acid encoding two or more repeat domains linked by inter-repeat loops, each repeat domain; and

25

incorporating into said nucleic acid a first nucleotide sequence encoding a first peptide ligand that binds to a target molecule and a second nucleotide sequence encoding a second peptide ligand that binds to an E3 ubiquitin ligase to generate a nucleic acid encoding a modular binding protein comprising said first and second peptide ligands, wherein said peptide ligands are located in an inter-repeat loop or at the N or C terminus of the modular binding protein; and

30

expressing the nucleic acid to produce said protein.

The first peptide ligand may comprise a sequence set out in Table 8 (SEQ ID NOs: 277-388) and second peptide ligand may comprise a sequence set out in Table 9 (SEQ ID NOs: 470-617).

35

An isolated nucleic acid encoding a modular binding protein as described herein is provided as an aspect of the invention. The nucleic acid may be comprised within an expression vector. Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. Preferably, the vector contains appropriate regulatory sequences to drive the expression of the nucleic acid in a host cell. Suitable regulatory sequences to drive the expression of heterologous nucleic acid coding sequences in expression systems are well-known in the art and include constitutive promoters, for example viral promoters such as CMV or SV40, and inducible promoters, such as Tet-on controlled promoters. A vector may also comprise sequences, such as origins of replication and selectable markers, which allow for its selection and

40

replication and expression in bacterial hosts such as *E. coli* and/or in eukaryotic cells.

Many techniques and protocols that are suitable for the expression of recombinant modular binding proteins in cell culture and their subsequent isolation and purification are known in the art (see for example *Protocols in Molecular Biology*, Second Edition, Ausubel et al. eds. John Wiley & Sons, 1992; *Recombinant Gene Expression Protocols* Ed RS Tuan (Mar 1997) Humana Press Inc).

A host cell comprising a nucleic acid encoding a modular binding protein as described herein or vector containing such a nucleic acid is also provided as an aspect of the invention. Suitable host cells include bacteria, mammalian cells, plant cells, filamentous fungi, yeast and baculovirus systems and transgenic plants and animals. The expression of proteins in prokaryotic cells is well established in the art. A common bacterial host is *E. coli*. A modular binding protein may also be produced by expression in eukaryotic cells in culture. Mammalian cell lines available in the art for expression of a modular binding protein include Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney cells, NS0 mouse melanoma cells, YB2/0 rat myeloma cells, human embryonic kidney cells (e.g. HEK293 cells), human embryonic retina cells (e.g. PerC6 cells) and many others.

Modular binding proteins as described herein may be used to produce libraries. A suitable library may be screened in order to identify and isolate modular binding proteins with specific binding activity. A library may comprise modular binding proteins, each modular binding protein in the library comprising:

- (i) two or more repeat domains,
- (ii) inter-repeat loops linking said repeat domains; and
- (iii) a first and a second peptide ligand, each said peptide ligand being located in an inter-repeat loop or at the N or C terminus of the modular binding protein,

wherein at least one amino acid residue in the second peptide ligand in said library is diverse.

The first peptide ligand may comprise a sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) or a variant thereof.

The residues at one or more positions in the peptide ligand of the modular binding proteins in the library may be diverse or randomised i.e. the residue located at the one or more positions may be different in different molecules in a population.

For example, 1 to 12 positions within a helical peptide ligand at the N or C terminus of the modular binding proteins in the library may be diverse or randomised. In addition, the non-constrained X_n sequence of the peptide ligand may contain additional diversity. Alternatively or additionally, 1 to n positions within an inter-repeat peptide ligand of the modular binding proteins in the library may be diverse or randomised, where n is the number of amino acids in the peptide ligand.

In some embodiments, peptide ligands may be screened individually and a modular binding protein

progressively assembled from repeat domains comprising peptide ligands identified in different rounds of screening. For example, a library may comprise modular binding proteins, each modular binding protein in the library comprising:

(i) two or more repeat domains,

(ii) inter-repeat loops linking said repeat domains; and

(iii) one or more constant peptide ligands having the same amino acid sequence in each modular binding protein in the library and one or more diverse peptide ligands, preferably one diverse peptide ligand, having a different amino acid sequence in each modular binding protein in the library, said peptide ligands being located in an inter-repeat loop or at the N or C terminus of the modular binding protein.

The constant peptide ligand may comprise a sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) or a variant thereof.

At least one amino acid residue in the diverse peptide ligands in said library may be diverse.

A library may be produced by a method comprising:

(a) providing a population of nucleic acids encoding a diverse population of modular binding proteins comprising

(i) two or more repeat domains,

(ii) inter-repeat loops linking the two or more repeat domains; and

(iii) one or more peptide ligands, each said peptide ligand being located in an inter-repeat loop or at the N or C terminus of the modular binding protein,

wherein one or more residues of a peptide ligand in each modular binding protein is diverse in said library, and

(b) expressing said population of nucleic acids to produce the diverse population, thereby producing a library of modular binding proteins.

The population of nucleic acids may be provided by a method comprising inserting a first population of nucleic acids encoding a diverse peptide ligand into a second population of nucleic acids encoding the two or more repeat domains linked by inter-repeat loops, optionally wherein the first and second nucleic acids are linked with a third population of nucleic acids encoding linkers of up to 10 amino acids.

The nucleic acids may be contained in vectors, for example expression vectors. Suitable vectors include phage-based or phagemid-based phage display vectors.

The nucleic acids may be recombinantly expressed in a cell or in solution using a cell-free *in vitro* translation system such as a ribosome, to generate the library. In some preferred embodiments, the library is expressed in a system in which the function of the modular binding protein enables isolation

of its encoding nucleic acid. For example, the modular binding protein may be displayed on a particle or molecular complex to enable selection and/or screening. In some embodiments, the library of modular binding proteins may be displayed on beads, cell-free ribosomes, bacteriophage, prokaryotic cells or eukaryotic cells. Alternatively, the encoded modular binding protein may be presented within an emulsion where activity of the modular binding protein causes an identifiable change. Alternatively, the encoded modular binding protein may be expressed within or in proximity of a cell where activity of the modular binding protein causes a phenotypic change or changes in the expression of a reporter gene.

Preferably, the nucleic acids are expressed in a prokaryotic cell, such as *E coli*. For example, the nucleic acids may be expressed in a prokaryotic cell to generate a library of recombinant binding proteins that is displayed on the surface of bacteriophage. Suitable prokaryotic phage display systems are well known in the art, and are described for example in Kontermann, R & Dubel, S, *Antibody Engineering*, Springer-Verlag New York, LLC; 2001, ISBN: 3540413545, WO92/01047, US5969108, US5565332, US5733743, US5858657, US5871907, US5872215, US5885793, US5962255, US6140471, US6172197, US6225447, US6291650, US6492160 and US6521404. Phage display systems allow the production of large libraries, for example libraries with 10^8 or more, 10^9 or more, or 10^{10} or more members.

In other embodiments, the cell may be a eukaryotic cell, such as a yeast, insect, plant or mammalian cell.

A diverse sequence as described herein is a sequence which varies between the members of a population i.e. the sequence is different in different members of the population. A diverse sequence may be random i.e. the identity of the amino acid or nucleotide at each position in the diverse sequence may be randomly selected from the complete set of naturally occurring amino acids or nucleotides or a sub-set thereof. Diversity may be introduced into the peptide ligand using approaches known to those skilled in the art, such as oligonucleotide-directed mutagenesis²², *Molecular Cloning: a Laboratory Manual*: 3rd edition, Russell et al., 2001, Cold Spring Harbor Laboratory Press, and references therein).

Diverse sequences may be contiguous or may be distributed within the peptide ligand. Suitable methods for introducing diverse sequences into peptide ligand are well-described in the art and include oligonucleotide-directed mutagenesis (see *Molecular Cloning: a Laboratory Manual*: 3rd edition, Russell et al., 2001, Cold Spring Harbor Laboratory Press, and references therein). For example, diversification may be generated using oligonucleotide mixes created using partial or complete randomisation of nucleotides or created using codons mixtures, for example using trinucleotides. Alternatively, a population of diverse oligonucleotides may be synthesised using high throughput gene synthesis methods and combined to create a precisely defined and controlled population of peptide ligands. Alternatively, "doping" techniques in which the original nucleotide

predominates with alternative nucleotide(s) present at lower frequency may be used.

Preferably, the library is a display library. The modular binding proteins in the library may be displayed on the surface of particles, or molecular complexes such as beads, for example, plastic or resin beads, ribosomes, cells or viruses, including replicable genetic packages, such as yeast, bacteria or bacteriophage (e.g. Fd, M13 or T7) particles, viruses, cells, including mammalian cells, or covalent, ribosomal or other *in vitro* display systems. Techniques for the production of display libraries, such as phage display libraries are well known in the art. Each particle or molecular complex may comprise nucleic acid that encodes the modular binding protein that is displayed by the particle.

In some preferred embodiments, the modular binding proteins in the library are displayed on the surface of a viral particle such as a bacteriophage. Each modular binding protein in the library may further comprise a phage coat protein to facilitate display. Each viral particle may comprise nucleic acid encoding the modular binding protein displayed on the particle. Suitable viral particles include bacteriophage, for example filamentous bacteriophage such as M13 and Fd.

Suitable methods for the generation and screening of phage display libraries are well known in the art. Phage display is described for example in WO92/01047 and US patents US5969108, US5565332, US5733743, US5858657, US5871907, US5872215, US5885793, US5962255, US6140471, US6172197, US6225447, US6291650, US6492160 and US6521404.

Libraries as described herein may be screened for modular binding proteins which display binding activity, for example binding to a target molecule. Binding may be measured directly or may be measured indirectly through agonistic or antagonistic effects resulting from binding. A method of screening may comprise;

(a) providing a library of modular binding proteins, each modular binding protein in the library comprising;

- (i) two or more repeat domains,
- (ii) inter-repeat loops linking said repeat domains; and
- (iii) a first and a second peptide ligand, each said peptide ligand being located in an inter-repeat loop or at the N or C terminus of the modular binding protein, wherein the first peptide ligand comprises a sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and one or more residues of the second peptide ligand are diverse in said library,

(b) screening the library for modular binding proteins which display a binding activity, and (c) identifying one or more modular binding proteins in the library which display the binding activity.

In some embodiments, the modular binding proteins in the library may comprise one peptide ligand with at least one diverse amino acid residue. Conveniently the modular binding proteins in the library comprise two repeat domains. The library may be screened for peptide ligands that bind to a target

molecule. Peptide ligands identified in this fashion can be assembled in a modular fashion to generate a modular binding protein as described herein that is multi-specific.

For example, a first library may be screened for a first peptide ligand that binds to a first target molecule and a second library may be screened for a second peptide ligand that binds to a second target molecule. The first and second peptide ligands are in different locations in the modular binding protein i.e. they are not both N terminal peptide ligands, C terminal peptide ligands or inter-repeat peptide ligands. First and second peptide ligands that bind to the first and second target molecules, respectively, are identified from the first and second libraries. The identified first and second peptide ligands may then be incorporated into a modular binding protein that binds to the first and second target molecules.

A first library may comprise modular binding proteins in the library with a first diverse peptide ligand having at least one diverse amino acid residue. A first peptide ligand that binds to a target molecule may be identified from the first library. Modular binding proteins comprising the first peptide ligand may be used to generate a second library comprising a second diverse peptide ligand having at least one diverse amino acid residue. For example, the modular binding protein from the first library may be modified by addition of a second diverse peptide ligand at the N or C terminal or by the addition of additional repeat domains comprising the second diverse peptide ligand in an inter-repeat loop. A second peptide ligand that binds to the same or a different target molecule may be identified from the second library. Modular binding proteins comprising the first and second peptide ligands may be used to generate a third library comprising a third diverse peptide ligand having at least one diverse amino acid residue. For example, the modular binding protein from the second library may be modified by addition of a third diverse peptide ligand at the N or C terminal or by the addition of additional repeat domains comprising the third diverse peptide ligand in an inter-repeat loop. A third peptide ligand that binds to the same target molecule as the first and/or second peptide ligands or a different target molecule may be identified from the third library. In this way, a modular binding protein containing multiple peptide ligands may be sequentially assembled (see Figure 16).

The use of separate libraries for each peptide ligand allows large numbers of different variants of each peptide ligand to be screened independently and then combined. For example, a phage library of 10^8 - 10^{12} first peptide ligand variants may be combined with a phage library of 10^8 - 10^{12} second peptide ligand variants and a phage library of 10^8 - 10^{12} third peptide ligand variants. In some embodiments, a phage library of 10^8 - 10^{12} N terminal peptide ligand variants may be combined with a phage library of 10^8 - 10^{12} C terminal peptide ligand variants to generate a modular binding protein with N and C terminal peptide ligands.

Screening a library for binding activity may comprise providing a target molecule and identifying or selecting members of the library that bind to the target, or expressing the library in a population of cells and identifying or selecting members of the library that elicit a cell phenotype. The one or more

identified or selected modular binding proteins may be recovered and subjected to further selection and/or screening.

Binding may be determined by any suitable technique. For example, the library may be contacted with the target molecule under binding conditions for a time period sufficient for the target molecule to interact with the library and form a binding reaction complex with a least one member thereof. Binding conditions are those conditions compatible with the known natural binding function of the target molecule. Those compatible conditions are buffer, pH and temperature conditions that maintain the biological activity of the target molecule, thereby maintaining the ability of the molecule to participate in its preselected binding interaction. Typically, those conditions include an aqueous, physiologic solution of pH and ionic strength normally associated with the target molecule of interest.

The library may be contacted with the target molecule in the form of a heterogeneous or homogeneous admixture. Thus, the members of the library can be in the solid phase with the target molecule present in the liquid phase. Alternatively, the target molecule can be in the solid phase with the members of the library present in the liquid phase. Still further, both the library members and the target molecule can be in the liquid phase.

Suitable methods for determining binding of a modular binding protein to a target molecule are well known in the art and include ELISA, bead-based binding assays (e.g. using streptavidin-coated beads in conjunction with biotinylated target molecules, surface plasmon resonance, flow cytometry, Western blotting, immunocytochemistry, immunoprecipitation, and affinity chromatography. Alternatively, biochemical or cell-based assays, such as fluorescence-based or luminescence-based reporter assays may be employed.

Multiple rounds of panning may be performed in order to identify modular binding proteins which display the binding activity. For example, a population of modular binding proteins enriched for the binding activity may be recovered or isolated from the library and subjected to one or more further rounds of screening for the binding activity to produce one or further enriched populations. Modular binding proteins which display binding activity may be identified from the one or more further enriched populations and recovered, isolated and/or further investigated.

In some embodiments, binding may be determined by detecting agonism or antagonism resulting from the binding of a modular binding protein to a target molecule, such as a ligand, receptor or enzyme. For example, the library may be screened by expressing the library in reporter cells and identifying one or more reporter cells with altered gene expression or phenotype. Suitable functional screening techniques for screening recombinant populations of modular binding proteins are well-known in the art

Modular binding proteins which display the binding activity may be further engineered to improve an

activity or property or introduce a new activity or property, for example a binding property such as affinity and/or specificity, an in vivo property such as solubility, plasma stability, or cell penetration, or an activity such as increased neutralization of the target molecule and/or modulation of a specific activity of the target molecule or an analytical property. Modular binding proteins may also be engineered to improve stability, solubility or expression level.

Further rounds of screening may be employed to identify modular binding proteins which display the improved property or activity. For example, a population of modular binding proteins enriched for binding to the target molecule may be recovered or isolated from the library and subjected to one or more further rounds of screening for the improved or new property or activity to produce one or further enriched populations. Optionally, this may be repeated one or more times. Modular binding proteins which display the improved property or activity may be identified from the one or more further enriched populations and recovered, isolated and/or further investigated.

A modular binding protein as described herein may be encapsulated in a liposome, for example for delivery into a cell. Preferred liposomes include fusogenic liposomes. Suitable fusogenic liposomes may comprise a cationic lipid, such as 1, 2-dioleoyl-3-trimethylammoniumpropane (DOTAP), and a neutral lipid, such as dioleoylphosphatidylethanolamine (DOPE) for example in a 1:1 (w/w) ratio. Optionally, a liposome may further comprise an aromatic lipid, such as DiO (3, 3'-dioctadecyloxacarbocyanine perchlorate), DiR (1, 1'-dioctadecyl-3, 3', 3'-tetramethylindotricarbocyanine iodide), N-(4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-sindacene-3-propionyl)-1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine (triethylammonium salt) (BODIPY FL-DHPE), and 2-(4,4-difluoro-5-methyl-4-bora-3a,4a-diazas-indacene-3-dodecanoyl)-1-hexadecanoyl-sn-glycero-3-phosphocholine (BODIPY-C12HPC) for example in a 0.1:1:1 (w/w) ratio relative to the neutral and cationic lipid. Suitable techniques for the encapsulation of proteins in liposomes and their delivery into cells are established in the art (see for example, Kube et al Langmuir (2017) 33 1051-1059; Kolašinac et al (2018) Int. J. Mol. Sci. 19 346).

A method described herein may comprise admixing a modular binding protein or encoding nucleic acid as described herein with a solution of lipids, for example in an organic solvent, such as chloroform, and evaporating the solvent to produce liposomes encapsulating the modular binding protein. Liposome encapsulations comprising a modular binding protein as described herein are provided as an aspect of the invention.

A modular binding protein or encoding nucleic acid as described herein may be admixed with a pharmaceutically acceptable excipient. A pharmaceutical composition comprising a modular binding protein or nucleic acid as described herein and a pharmaceutically acceptable excipient is provided as an aspect of the invention.

The term "pharmaceutically acceptable" as used herein pertains to compounds, materials,

compositions, and/or dosage forms which are, within the scope of sound medical judgement, suitable for use in contact with the tissues of a subject (e.g., human) without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio. Each carrier, excipient, etc. must also be "acceptable" in the sense of being compatible with the other ingredients of the formulation. Suitable carriers, excipients, etc. can be found in standard pharmaceutical texts, for example, Remington's Pharmaceutical Sciences, 18th edition, Mack Publishing Company, Easton, Pa., 1990.

The pharmaceutical composition may conveniently be presented in unit dosage form and may be prepared by any methods well-known in the art of pharmacy. Such methods include the step of bringing the modular binding protein into association with a carrier which may constitute one or more accessory ingredients. In general, pharmaceutical compositions are prepared by uniformly and intimately bringing into association the active compound with liquid carriers or finely divided solid carriers or both, and then if necessary shaping the product.

Pharmaceutical compositions may be in the form of liquids, solutions, suspensions, emulsions, elixirs, syrups, tablets, lozenges, granules, powders, capsules, cachets, pills, ampoules, suppositories, pessaries, ointments, gels, pastes, creams, sprays, mists, foams, lotions, oils, boluses, electuaries, or aerosols.

A modular binding protein, encoding nucleic acid or pharmaceutical composition comprising the modular binding protein or encoding nucleic acid may be administered to a subject by any convenient route of administration, whether systemically/ peripherally or at the site of desired action, including but not limited to, oral (e.g. by ingestion); topical (including e.g. transdermal, intranasal, ocular, buccal, and sublingual); pulmonary (e.g. by inhalation or insufflation therapy using, e.g. an aerosol, e.g. through mouth or nose); rectal; vaginal; parenteral, for example, by injection, including subcutaneous, intradermal, intramuscular, intravenous, intraarterial, intracardiac, intrathecal, intraspinal, intracapsular, subcapsular, intraorbital, intraperitoneal, intratracheal, subcuticular, intraarticular, subarachnoid, and intrasternal; by implant of a depot, for example, subcutaneously or intramuscularly.

Pharmaceutical compositions suitable for oral administration (e.g., by ingestion) may be presented as discrete units such as capsules, cachets or tablets, each containing a predetermined amount of the active compound; as a powder or granules; as a solution or suspension in an aqueous or non-aqueous liquid; or as an oil-in-water liquid emulsion or a water-in-oil liquid emulsion; as a bolus; as an electuary; or as a paste.

Pharmaceutical compositions suitable for parenteral administration (e.g. by injection, including cutaneous, subcutaneous, intramuscular, intravenous and intradermal), include aqueous and non-aqueous isotonic, pyrogen-free, sterile injection solutions which may contain anti-oxidants, buffers, preservatives, stabilisers, bacteriostats, and solutes which render the formulation isotonic with the

blood of the intended recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening agents, and liposomes or other microparticulate systems which are designed to target the compound to cells, tissue or organs. Examples of suitable isotonic vehicles for use in such formulations include Sodium Chloride Injection, Ringer's Solution, or Lactated Ringer's Injection. Typically, the concentration of the active compound in the solution is from about 1 ng/ml to about 10 µg/ml, for example, from about 10 ng/ml to about 1 µg/ml. The formulations may be presented in unit-dose or multi-dose sealed containers, for example, ampoules and vials, and may be stored in a freeze-dried (lyophilised) condition requiring only the addition of the sterile liquid carrier, for example water for injections, immediately prior to use.

It will be appreciated that appropriate dosages of the modular binding protein, can vary from patient to patient. Determining the optimal dosage will generally involve the balancing of the level of diagnostic benefit against any risk or deleterious side effects of the administration. The selected dosage level will depend on a variety of factors including, but not limited to, the route of administration, the time of administration, the rate of excretion of the imaging agent, the amount of contrast required, other drugs, compounds, and/or materials used in combination, and the age, sex, weight, condition, general health, and prior medical history of the patient. The amount of imaging agent and route of administration will ultimately be at the discretion of the physician, although generally the dosage will be to achieve concentrations of the imaging agent at a site, such as a tumour, a tissue of interest or the whole body, which allow for imaging without causing substantial harmful or deleterious side-effects.

Administration in vivo can be effected in one dose, continuously or intermittently (e.g., in divided doses at appropriate intervals). Methods of determining the most effective means and dosage of administration are well known to those of skill in the art and will vary with the formulation used for therapy, the purpose of the therapy, the target cell being treated, and the subject being treated. Single or multiple administrations can be carried out with the dose level and pattern being selected by the physician.

Modular binding proteins described herein may be used in methods of diagnosis or treatment in human or animal subjects, e.g. human. Modular binding proteins for a target molecule may be used to treat disorders associated with the target molecule.

Other aspects and embodiments of the invention provide the aspects and embodiments described above with the term "comprising" replaced by the term "consisting of" and the aspects and embodiments described above with the term "comprising" replaced by the term "consisting essentially of".

It is to be understood that the application discloses all combinations of any of the above aspects and embodiments described above with each other, unless the context demands otherwise. Similarly, the

application discloses all combinations of the preferred and/or optional features either singly or together with any of the other aspects, unless the context demands otherwise.

Modifications of the above embodiments, further embodiments and modifications thereof will be apparent to the skilled person on reading this disclosure, and as such, these are within the scope of the present invention.

All documents and sequence database entries mentioned in this specification are incorporated herein by reference in their entirety for all purposes.

“and/or” where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. For example “A and/or B” is to be taken as specific disclosure of each of (i) A, (ii) B and (iii) A and B, just as if each is set out individually herein.

Certain aspects and embodiments of the invention will now be illustrated by way of example and with reference to the figures described above.

Experiments

1. Methods

1.1 Large-scale protein purification (His-tagged) from *E. coli*

The pRSET B (His-tag) constructs were transformed into chemically competent *E. coli* C41 cells by heat shock and plated on LB-Amp plates. Colonies were grown in 2TY media containing ampicillin (50 micrograms/mL) at 37 °C, 220 rpm until the optical density (O.D.) at 600 nm reached 0.6. Cultures were then induced with IPTG (0.5mM) for 16-20 h at 20 °C or 4 h at 37 °C. Cells were pelleted by centrifugation at 3000 g (4 °C, 10 min) and resuspended in lysis buffer (10 mM sodium phosphate pH 7.4, 150 mM NaCl, 1 tablet of SIGMAFAST protease inhibitor cocktail (EDTA-free per 100 mL of solution), then lysed on a Emulsiflex C5 homogenizer at 15000 psi. Cell debris was pelleted by centrifugation at 15,000 g at 4 °C for 45 min. Ni-NTA beads 50% bed volume (GE Healthcare) (5 mL) were washed once with phosphate buffer (10 mM sodium phosphate pH 7.4, 150 mM NaCl) before the supernatant of the cell lysate was bound to them for 1 hr at 4 °C in batch. The loaded beads were washed three times with phosphate buffer (40 mL) containing 30 mM of imidazole to prevent non-specific interaction of lysate proteins with the beads. Samples were eluted using phosphate buffer with 300 mM imidazole, and purified by size-exclusion chromatography using a HiLoad 16/60 SuperdexG75 column (GE Life-Science) pre-equilibrated in phosphate buffer (10 mM sodium phosphate, pH 7.4, 150 mM NaCl) and proteins separated in isocratic conditions. Purity was checked on NuPage protein gel (Invitrogen), and fractions found to be over 95% pure were pooled. Purified protein was flash-frozen and stored at -80 °C until further use. Concentrations were determined by measuring absorbance at 280 nm and using a calculated extinction coefficient from ExPASy ProtParam (Gasteiger et al. 2005) for each variant. Molecular weight and purity was confirmed using mass spectrometry (MALDI).

1.2 Large-scale protein purification (heat treatment) from *E. coli*

All modular binding proteins described herein are thermally very stable, with melting temperatures above 80°C. This means that the modular binding proteins could be separated from *E. coli* proteins by incubating the cell lysates at 65 °C for 20 min. Very few of the *E. coli* proteins survive such temperatures, and therefore, they will unfold and aggregate. Aggregated proteins were removed by centrifugation, leaving 80-90% pure sample of the desired protein. All our constructs folded reversibly, and therefore could be further purified by methods such as acetone or salt precipitation to remove DNA and other contaminants.

This approach allowed the production of large amounts of functional proteins without expensive affinity purification methods such as antibodies or His tags and is scalable to industrial production and bioreactors.

1.3 Small-scale purification of His-tagged proteins for higher-throughput testing

Plasmids were transformed into *E. coli* C41 cells and plated overnight. 15 mls of 2TY medium (Roche) containing 50 micrograms/ml ampicillin was placed in multiple 50 ml tubes. Several colonies were picked and resuspended in each 15 ml culture. For sufficient aeration it is important to only loosely tighten the lids of the 50 ml tubes. Cells were grown at 37 °C until OD600 of 0.6 and then induced with 0.5 mM IPTG overnight. Cells were pelleted at 3000 g (Eppendorf Centrifuge 5804) and then resuspended in 1ml of BugBuster® cell lysis reagent. Alternatively, sonication in combination with lysozyme and DNase I treatment was used. The lysate was spun at 12000 g for 1 minute to pellet any insoluble protein and cell debris.

The supernatant was added to 100 µl bed volume of pre-washed Ni-NTA agarose beads. The subsequent affinity purification was performed in batch, by washing the beads 4 times with 1 ml of buffer each time (alternatively, Qiagen Ni-NTA Spin Columns can be used). The first wash contained 10% BugBuster® solution and 30 mM imidazole in the chosen buffer. Here we used 50 mM sodium phosphate buffer pH 6.8, 150 mM NaCl. The three successive washes had 30 mM of imidazole in the chosen buffer. Beads were washed thoroughly to remove the detergent present in the BugBuster® solution. Protein was eluted from the beads in a single step using 1 ml of chosen buffer containing 300 mM imidazole. The combination of Bugbuster® and imidazole and the repeat washes in small bead volumes yielded >95% pure protein. Imidazole was removed using a NAP-5 disposable gel-filtration column (GE Healthcare).

1.4 Competition Fluorescence Polarization (FP)

To assay the binding of the designed SOS-TPR protein to KRAS, Competition FP was performed using purified KRAS Q61H mutant and (2'-(or-3')-O-(N-Methylanthraniloyl) Guanosine 5'-Triphosphate, a fluorescent version of GTP, also known as mant-GTP. SOS-TPR was titrated using a 2-fold serial dilution against a 1:1 complex of KRAS Q61H and mant-GTP (1 µM) in a black 96-well

plate (CLS3993 SIGMA). Plates were prepared under reduced light conditions and incubated at room temperature. Readings were taken on the CLARIOstar microplate reader, using an excitation filter at 360 nm and emission filter at 440 nm.

5 1.5 Isothermal Titration Calorimetry (ITC)

ITC was performed at 25°C using a VP-ITC (Microcal). 1TBP-CTPR2, 2TBP-CTPR4, 3TBP-CTPR6 and TNKS2 ARC4 were dialysed into 10 mM sodium phosphate buffer pH 7.4, 150 mM NaCl, 0.5 mM TCEP. Dialysed TNKS2 ARC4 (200 µM) was titrated into the sample cell containing 1TBP-CTPR2 at 20 µM. Similar experiments were performed for 2TBP-CTPR4 and 3TBP-CTPR6. Injections of TNKS2
10 ARC4 into the cell were initiated with a 5 µL injection, followed by 29 injections of 10 µL. The reference power was set at 15 µCal/s with an initial delay of 1000 s and a stirring speed of 485 rpm. Data were fitted using the instrument software a one-site binding model.

15 1.6 Cell culture

HEK293T cells were cultured in Dulbecco's Modified Eagle's Medium (Sigma Aldrich) supplemented with 10% fetal bovine serum and penicillin/streptomycin (LifeTech) at 37°C with 5% CO₂ air supply.

20 1.7 Cell transfection

HEK293T were seeded in 6-well tissue culture plates (500,000 cells per well) and transfected the next
20 day using the Lipofectamine2000 transfection reagent (Invitrogen) according to the manufacturer's protocol.

25 1.8 β-catenin levels western blot assay

HA-β-catenin (1 µg) alone and with various PROTACs (1 µg) was transfected in HEK293T cells in 6-
25 well plates using Lipofectamine2000. After 48 hours of transfection, the cells were lysed in 200 µL of Laemmli buffer. After sample was boiled at 95°C for 20 min proteins were resolved by SDS-PAGE and transferred to a PVDF membrane, and immunoblotting was performed using anti-HA (C29F4, Cell Signaling Technologies) and anti-actin (A2066, Sigma-Aldrich) antibodies. Changes in β-catenin levels were evaluated by the densitometry of the bands corresponding to HA-β-catenin normalised to
30 actin levels using ImageJ.

35 1.9 Liposomal formulation and cytotoxicity assay

To make liposomal formulations of proteins (LFP), lipids (DOTAP (cationic): DOPE (neutral): DiR (aromatic) = 1:1:0.1 w/w) were dissolved in chloroform, and solvent was evaporated under vacuum
35 overnight. Resulting mixed lipid cake was hydrated with 10 mM HEPES pH 7.4, containing 27 µM protein, so that the total lipid concentration is 4 mg/ml. This mixture was vortexed for 2 minutes and then sonicated for 20 minutes at room temperature. Liposomes encapsulating proteins were stored at 4°C until further use. To make empty liposomes (EL, empty liposomes without proteins), lipid cake was hydrated with 10 mM HEPES pH 7.4 without proteins.

An ATP assay was used to investigate whether there is any cytotoxicity associated with EL and LFP. In a typical procedure, 2×10^5 HEK 293T cells/well in 500 μ L of Dulbecco's Modified Eagles Medium (DMEM) supplemented with 10% fetal bovine serum were grown for 24 hours in a 24-well cell culture plate. Cells were incubated with liposome (EL/LFP)-media (DMEM without FBS) mix, having different
5 volumes (0-60 μ L) of EL and LFP, for 15 minutes at 37°C. After washing twice with 1x PBS, 500 μ L of CellTiter-Glo® Reagent (Promega) was added and luminescence was measured using a microplate reader as per the manufacture's protocol. Untreated cells were used as control. Data were obtained from triplicate samples, and the standard deviations were calculated from two independent
10 experiments.

1.10 TOPFLASH assay

The Wnt pathway was activated by treating HEK293T cells with Wnt-conditioned media obtained from L-cells expressing Wnt3A for 8 days. To perform the assay, 10^5 HEK293T cells/well were seeded on a 24-well plate Nunclon Delta Surface plate (NUNC) and incubated overnight at 37°C, 5% CO₂. The
15 following day, cells were transfected with 100 ng of TOPflash TCF7L2-firefly luciferase plasmid, 10 ng of CMV-Renilla plasmid (as internal control) and 100 ng of the corresponding TPR construct. Plasmids were mixed with 0.5 μ L of Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (invitrogen). Transfected cells were allowed to recover for 8 h, then they were treated with Wnt-conditioned media (1:2 final concentration) for a further 16 h. The TOPflash assay
20 was performed using the Dual-Luciferase Reporter Assay System (Promega) (Korinek et al., 1997 Science 275(5307):1784-7) following the manufacturer's instructions. The activities of firefly and Renilla luciferases were measured sequentially from a single sample, using the CLARIOstar plate reader. Relative luciferase values were obtained from triplicate samples dividing the firefly luminescence activity by the CMV-induced Renilla activity, and standard deviation was calculated.

1.11 TOPFLASH assay using liposome encapsulation to deliver designed TPR proteins into the cell

10^5 HEK 293T cells in 500 μ L of Dulbecco's Modified Eagles Medium (DMEM) supplemented with 10% fetal bovine serum were grown overnight in each well of a 24-well cell culture plate. For
30 TOPFLASH reporter assays, 100 ng/well of TOPFLASH plasmid and 10 ng/well of CMV-Renilla plasmid (as internal control) were used to transfect cells in 24-well plates. Cells were transfected with the Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Invitrogen). Transfected cells were allowed to recover for 8 hours, and Wnt signalling was activated by addition of Wnt3A-conditioned media obtained from L-cells. 16 hours post Wnt pathway activation, proteins were
35 delivered into the cells by liposomal treatment. Cells were incubated with liposome (LFP)-media (DMEM without FBS) mix for 15 minutes at 37°C followed by one PBS wash. Wnt3A conditioned media was replaced and cells were incubated for variable time durations (2-8 hours). Following incubation, TOPFLASH assays were performed using the Dual-Luciferase Reporter Assay System (Promega) (Korinek et al., 1997) following the manufacturer's instructions. Relative luciferase values
40 were obtained from triplicate samples (from two independent experiments) by dividing the firefly

luciferase values (from TOPFLASH) by the Renilla luciferase values (from CMV renilla), and standard deviations were calculated.

1.12. Competition fluorescence polarisation (FP) assay to measure the binding of designed Nrf-TPR proteins to Keap1

To measure the binding of the designed Nrf-TPR proteins to Keap1, Competition FP was performed using 384-well black opaque optiplate microplates and a CLARIOstar microplate reader. Nrf-TPR proteins were titrated into a solution containing a mixture of FITC-labelled Nrf2 peptide and Keap1 protein. The prepared plates were incubated for 30 minutes at room temperature before readings were taken.

1.13 Generation of HiBiT tagged Beta Catenin MIA PaCa-2 cell line

Beta Catenin in MIA PaCa-2 cells was tagged with the HiBiT small peptide tag by CRISPR editing and homology directed repair (HDR). Ribonuclearprotein complex of Cas9 enzyme and gRNA

(mU*mG*mA* rCrCrU rGrUrA rArArU rCrArU rCrCrU rUrUrG rUrUrU rUrArG rArGrC rUrArG rArArA rUrArG rCrArA rGrUrU rArArA rArUrA rArGrG rCrUrA rGrUrC rCrGrU rUrArU rCrArA rCrUrU rGrArA rArArA rGrUrG rGrCrA rCrCrG rArGrU rCrGrG rUrGrC mU*mU*mU* rU (SEQ ID NO: 83) was prepared and electroporated into MIA PaCa-2 cells with a single stranded DNA HDR template (CTTTCTGAAGTTCTGTAGGCAGAGTAAAAGTATTTTACCCAACTGGCTTTTTAAACTTCTTACC TAAAGGATGATTTAGCTAATCTTCTTGAACAGCCGCCAGCCGCTCACCAGGTCAGTATCAAACCA GGCCAGCTGATTGCTGTACCTGGAGGCAGCCCATCCATGAGGTCCTGGGCATGCCCCAGATC TGGCA (SEQ ID NO: 84)) using the Lonza 4D Nucleofector system.

Stable pools were aliquoted and frozen after 3 weeks. Aliquots were used on assays until passage 10.

1.14 HiBiT Assay for Endogenous Beta Catenin Degradation

MIA PaCa-2 cells in which endogenous Beta Catenin has been tagged with the small peptide HiBiT (MIA PaCa-2 Beta Catenin HiBiT) were cultured in T175 flasks in DMEM 10% FCS at 37°C, 5% CO₂ and split 1 in 10 on reaching 90% confluency. Flasks were grown to approx. 80% confluency and cells were then split by gentle washing with 20 ml of PBS and incubation with 4 ml of cell dissociation buffer for 5 minutes. Cells were counted and seeded at 5000 cells per well into replicate white solid bottom and black clear bottom plates in 100 µl per well of DMEM 10%FCS. Cells were incubated overnight at 37°C, 5% CO₂.

Cells were transfected with 100 ng per well of pcDNA 3.1 vector containing CPTR constructs using Lipofectamine 3000 using the manufacturer's recommended protocol (Thermofisher Scientific) or with 25nM siRNA (Beta Catenin targeted or Scramble control) using TransIT X2 using the manufacturers recommended protocol (Mirus Bio Inc). 24 hours after transfection the NanoGlo lytic detection assay was used to determine Beta Catenin-HiBiT levels. Plates were equilibrated to room temperature and 80 µl of Nanoglo Lytic reagent was added to each well. After 10 minutes of shaking luminescence was

measured on a GloMax Discover plate reader with an integration time of 2 seconds.

1.14 Immunofluorescent Assay for Expression

MIA PaCa-2 cells in which endogenous Beta Catenin has been tagged with the small peptide HiBiT (MIA PaCa-2 Beta Catenin HiBiT) were cultured in T175 flasks in DMEM 10% FCS at 37°C, 5% CO₂ and split 1 in 10 on reaching 90% confluency. Flasks were grown to approx. 80% confluency and cells were then split by gentle washing with 20 ml of PBS and incubation with 4 ml of cell dissociation buffer for 5 minutes. Cells were counted and seeded at 5000 cells per well into replicate white solid bottom and black clear bottom plates in 100 µl per well of DMEM 10% FCS. Cells were incubated overnight at 37°C, 5% CO₂.

Cells were transfected with 100 ng per well of pcDNA 3.1 vector containing CPTR constructs using Lipofectamine 3000 using the manufacturers recommended protocol (ThermoFisher Scientific). 24 hours after transfection, media was aspirated from the black clear bottom plates and cells were fixed with 4% PFA, quenched with 0.1M glycine in PBS, and permeabilized with IF blocking and permeabilization buffer (1* PBS, 5% FCS, 0.2% Saponin, 0.2 µM filtered). Cells were stained with anti HA primary antibody (HA-Tag (C29F4) Rabbit mAb), followed Goat anti-Rabbit, Alexa Fluor® 555 secondary and Hoescht 33342 Nuclear counterstain. Wells were imaged immediately on an EVOS M5000 Fluorescence Microscope.

2. Results

Tetratricopeptide repeat (TPR) is a 34-residue motif that can be repeated in tandem to generate modular proteins. TPRs are used here as an example of helix-turn-helix tandem-repeats arrays, but any tandem repeat array may be used.

RTPR proteins comprising TPRs were derived from the consensus TPR sequence (CTPR). Two repeats were found to be sufficient to generate a highly stable mini-protein of 68 amino acids (RTPR2). The biophysical properties of two types of engineering strategy; loop insertions and terminal helix grafting, were assessed. The molar ellipticity at 222 nm (a measure of helical secondary structure content) of three different RTPR modules was monitored as a function of increasing temperature. A decrease in the absolute molar ellipticity with increasing temperature indicates a loss of structure and the unfolding of the protein. Even at the highest temperature recorded (85 °C), the RTPR2 protein without insertion was not fully denatured (Figure 1). RTPR2 with a 20-residue unstructured loop between the two repeats showed a small shift to a lower melting temperature (Figure 1), but the protein remains fully folded up to 55 °C. This is well above physiologically relevant temperatures. RTPR2 with an additional N-terminal helix showed an increase in absolute molar ellipticity, indicating that the additional helical domain is folded. Moreover, unlike the loop insertion, the helix domain was capable of stabilising the RTPR2 module, shifting the transition midpoint to above 90 °C (Figure 1). These results showed that the two engineering strategies generated folded and stable modular mini-proteins capable of withstanding high temperatures.

A key feature of the TPR scaffold was its modular nature. This modularity allowed us to display any number of binding modules in tandem to obtain bi- and multi-valent and multi-functional molecules against one, two or more targets. The stability of these proteins was shown to be modular. The stabilities of proteins comprising TBP-CTPR2 (a two-repeat CTPR with a loop insertion that binds to the protein tankyrase (Guettler et al. 2011) repeated in tandem were measured. The TBP-CTPR2-containing proteins had two, four, six, and eight repeats, and they displayed one, two, three and four binding loops, respectively. The helical content of the proteins, monitored by molar ellipticity at 222 nm, was found to increase in proportion to the number of repeats, as did the stability, indicating that they were behaving like classic helical repeat proteins (Figure 2). These results demonstrate that bi- or multi-functional modular binding proteins have a high thermostability.

2.1. Demonstration of proteins with a single binding function grafted onto an alpha-helix

2.1.1 SOS1-TPR, a helix-grafted binding module designed to bind to oncoprotein KRAS

First, we mapped the helix of SOS1 that interacts with KRAS (Margarit et al. 2003 Cell 112 5 685-695) onto the heptad distribution. We matched the heptad positions with the stapled SOS1 helical peptide produced by Leshchiner *et al.* (PNAS 2015 112 (6) 1761-1766) and set the stapled side of the peptide to form the hydrophobic interface with the rest of the TPR protein (Fig. 3A). The length of the helix is important. An N-terminal solvating CTPR helix ends in the sequence DPNN, which forms a short loop that leads into the next repeat. CTPR-mediated "stapling" (constraining) of binding helices therefore occurred through residues Tyr (*i*) - Ile (*i*+4) - Tyr (*i*+7) - Leu (*i*+11), fully stapling a 15-residue helix.

We created a hydrophobic interface between the grafted helix and the adjacent repeat and allowed the formation of the DPNN loop at the C-terminal end of the grafted helix. We then grafted the final sequence onto the crystal structure of a CTPR B helix for further validation of the interaction. Our designed KRAS-binding protein, SOS1-TPR, was docked against KRAS using the Haddock software (de Vries & Bonvin 2011; de Vries et al. 2010). Haddock is a data-driven docking algorithm that uses known information about the interaction for its calculations. The crystal structure of SOS1-KRAS (PDB: 1NVU) (Margarit et al. 2003) was originally used to design the stapled peptide. The active (primary interaction residues) and the passive (5 Å proximity to active) residues were extracted and inputted into the calculations.

Since the initial work, we have found that docking is not necessary to validate new helical modules. The design strategy has a solid theory based on the geometry of α -helices, and a design will be successful as long as the key binding residues have been grafted. These TPR repeats were thus found to be exceptional scaffolds to display binding helices, as they grow linearly in the opposite direction of the helix, thereby avoiding any steric clashes with the target protein.

Next, we monitored KRAS binding using the change in fluorescence polarisation of mant-GTP (2'-/3'-O-(N'-Methylantraniloyl) guanosine-5'-O-triphosphate), a fluorescent analog of GTP (Fig. 3B). The fluorescence of mant-GTP is dependent on the hydrophobicity of its environment (excitation at 360 nm, emission at 440 nm). An increase in fluorescence intensity and fluorescence polarization was observed

previously upon binding to KRAS (Leshchiner et al. 2015). SOS-TPR2 was then titrated into the preformed mant-GTP-KRAS complex. There was a clear decrease in polarisation with increasing concentrations of SOS-TPR2, indicating displacement of mant-GTP upon binding of SOS-TPR2 to KRAS (Figure 3B). Fitting the data gave an EC50 of 3.4 μ M. In contrast, a blank protein, CTPR3, had no effect on the fluorescence polarisation.

2.1.2 p53-TPR, a helix-grafted binding module designed to bind to Mdm2

Many degrons (region within the substrate that is recognized by the E3 ubiquitin ligase) are unstructured. However, p53 binds to the Mdm2 E3 through an alpha helix (Figure 4A). Stapled versions of the p53 helix, as well as circular peptides and grafted coiled coils, have been developed by many groups, and the sequences have been optimised to give nanomolar affinities in some cases (see for example, Ji et al. 2013; Lee et al. 2014; Kritzer et al. 2006). The p53 helix has a favourable geometry to be grafted onto the C-terminal solvating helix of the CTPR scaffold, and moreover the two helices have 30% sequence identity.

Proof of binding of p53-CTPR2 to Mdm2 (N-terminal domain) was obtained using isothermal titration calorimetry (ITC). Mdm2 was titrated into a solution containing 10 μ M of p53-TPR2. ITC measures the heat released upon binding. A high-affinity interaction was observed with a dissociation constant of approximately 50 nM (Figure 4B).

2.2. Demonstration of proteins with a single binding function grafted onto an inter-repeat loop

2.2.1 TPB2-TPR, a loop module designed to bind to oncoprotein tankyrase

First, we introduced the SLiM "3BP2", a sequence that binds to the substrate-binding ankyrin-repeat clusters (ARC) of the protein tankyrase, a multi-domain poly ADP-ribose polymerase that is upregulated in many cancers (Guettler et al. 2011) onto the CTPR scaffold. Grafting SLiMs in folded domains led to an increase of proteolysis resistance; showing the potential to expand the interaction surface through further rational engineering, *in silico* methods and/or directed evolution; controlled geometric arrangement; and bi- or multivalency of interactions.

We tested the binding of 1TBP-CTPR2, 2TBP-CTPR4 and 3TBP-CTPR6 to the ARC4 domain of tankyrase using ITC (Figure. 5A). This technique is particularly useful for these interactions, as it can measure the stoichiometry (n) of the interaction. We showed that n increased with the number of binding loops, meaning that there were as many tankyrase molecules bound to one TBP-CTPR as loops in the protein. Thus, all loops are accessible to the binding partner. Moreover, the binding affinity increases and the off rate decreases with the number of repeats indicative of an avidity effect. This type of multivalent molecule would be particularly useful for full-length tankyrase, as it has four ARC domains capable of binding the 3BP2 peptide.

Multivalency in this system was increased further via oligomerisation of the binding modules by fusing them to the foldon domain of T4 fibrin (Fig. 5B). This trimerisation domain comprises of a C-terminal helix, such as that of p53-CTPR, ending with the foldon domain, a short β -sheet peptide capable of homo-

trimerising. The foldon domain has been shown to be highly stable and independently folded (Boudko et al 2002; Meier et al. 2004). In this way, multiple binding modules can be arranged with specified geometries to inhibit complex multivalent molecules that cannot be targeted with monovalent interactions due to their natural tendency to interact with other multivalent networks with high avidity.

5

2.2.2 Effect of introducing multivalency into a single binding function TPR

We tested the function of multi-valent CTPR proteins containing variable numbers of the “3BP2” motif that binds to the protein tankyrase. (1TBP-CTPR2, 2TBP-CTPR4 and 3TBP-CTPR6 etc.). Multivalency was increased further via oligomerisation of the TPRs by fusing them to the foldon domain of T4 fibrin (1TBP-CTPR2-Foldon, 2TBP-CTPR4-Foldon etc.). Tankyrase is upregulated in many cancers and exerts its effect by upregulating beta-catenin. Therefore, the inhibitory effect of the TBP-grafted TPRs was assayed using a beta-catenin reporter gene assay (TOPFLASH assay). Increasing the number of functional units increased the inhibitory effect of the proteins, as mentioned using a Wnt signalling assay (Figure 17).

15

2.2.3 Skp2-RTPR, a loop module designed to bind to E3 ubiquitin ligase SCF^{Skp2}

Skp2 is the substrate recognition subunit of the SCF^{Skp2} ubiquitin ligase. The Skp2-binding sequence that we inserted into the RTPR loop was based on the previously published degron peptide sequence derived from the substrate p27 that binds to Skp2 in complex with Cks1 (an accessory protein) (Hao et al. 2005). We used only 10 residues of this peptide. Although ideally the Skp2-binding sequence would include a phospho-threonine (as this residue makes some key contacts with Skp2 and Cks1), we instead explored whether we could replace the phospho-threonine with a phosphomimetic (glutamate) without affecting binding affinity. We found using co-immunoprecipitation that the resulting p27-TPR protein was able to bind to Skp2 (Fig. 6A) and that it was able to inhibit the ubiquitination of p27 *in vitro* with a high efficiency indicating a dissociation constant of the order of 30 nM (Fig. 6B). As the peptide adopts a turn-like conformation in its Skp2/Cks1-bound state, constraining it within the RTPR scaffold leads to a large enhancement in binding affinity that outweighs any loss in affinity arising from replacing the phosphothreonine with a phosphomimetic.

20

25

2.2.4 Nrf-TPR, a loop module designed to bind to E3 ubiquitin ligase Keap1-Cul3

Keap1 is the substrate recognition subunit of the Keap1-Cul3 ubiquitin ligase. A Keap1-binding sequence that we inserted into the CTPR loop was based on the previously published degron peptide sequence derived from the Keap1 substrate Nrf2. We found using co-immunoprecipitation that the resulting Nrf-TPR protein was able to bind to Keap1 (Fig. 7A) and that the interaction had a high affinity in the low nanomolar range as measured by ITC analysis (Fig. 7B).

30

35

2.3. Engineering the RTPR scaffold for delivery into the cell

Combining our RTPR sequences with an alternative consensus TPR sequence (Parmeggiani et al. 2015) we included additional solvent-exposed Arginine residues, as such ‘resurfacing’ or ‘supercharging’ has been shown previously to facilitate the entry of proteins into cells (Chapman & McNaughton 2016;

40

Thompson et al. 2012). Figure 8 shows that this approach was successful in delivering a fluorescent-labelled resurfaced TBP-RTPR2 protein into two different cell lines.

2.4. Design of hetero-bifunctional TPRs to direct proteins for ubiquitination and subsequent degradation

The Wnt/ β -catenin signalling pathway is deregulated in many cancers and in neurodegenerative diseases, and therefore β -catenin is an important drug target. There are a large number of known binding sequences (both helical and non-helical) for β -catenin that appear suitable for grafting onto the TPR scaffold, and therefore we chose it as the first target for our design of hetero-bifunctional TPRs to induce protein degradation. We selected Mdm2 and SCF^{Skp2} to test as E3 ubiquitin ligases, as we had successfully generated single-function TPRs to bind to them (Figs. 4 and 6). We generated structural models of some of the hetero-bifunctional molecules and used these as a crude assessment of whether the resulting presentation of β -catenin to the E3 looked appropriate. We then generated a small library of plasmids encoding proteins comprising three or four TPRs functionalized with different combinations of the β -catenin-binding module and the two E3 ligase-binding modules.

We transfected HA-tagged β -catenin plasmid alone or HA-tagged β -catenin plasmid together with one of the various hetero-bifunctional TPR plasmids in HEK293T cells using Lipofectamine2000. After 48 hours of transfection, the cells were lysed, the sample was boiled and proteins were resolved by SDS-PAGE and immunoblotting was performed using anti-HA and anti-actin antibodies. Changes in β -catenin levels were evaluated by the densitometry of the bands corresponding to HA- β -catenin normalised to actin levels (Fig. 9). The results show that a number of the hetero-bifunctional molecules are capable of reducing β -catenin levels by up to 70%. In contrast, neither a blank TPR nor single-function TPRs have any effect on β -catenin levels.

A range of different factors contribute to efficient ubiquitination and target degradation by these hetero-bifunctional molecules, hence the power of screening different combinations of single-function modules and potentially also different lengths of intervening blank modules.

2.5. Using a delivery vehicle to introduce the modular TPR proteins into cells

We encapsulated the designed TPR proteins within fusogenic liposomes made from cationic, neutral, and aromatic lipids, and we showed that they were thereby delivered into cells (Figures 18 and 19). Empty liposomes and liposomes encapsulating TPR proteins are not toxic to the cell (Figure 20).

2.6. Further examples of designed single-function and hetero-bifunctional TPRs

TPR proteins targeting tankyrase (Figures 21 and 22) were delivered into cells using liposome encapsulation, and the effect on Wnt signalling was assayed using a TOPFLASH assay. The results show that the designed TPR proteins are able to inhibit Wnt signalling.

For KRAS, we transfected KRAS plasmid alone or KRAS plasmid together with one of the TPR plasmids in HEK293T cells using Lipofectamine2000. 24 hours post transfection the cells were lysed, and KRAS levels were evaluated by western blot. The results show that the designed hetero-bifunctional TPR is capable of reducing KRAS levels (Figure 23).

In addition, TPR proteins targeting KRAS were delivered into cells using liposome encapsulation, and the effect on HiBiT-tagged KRAS levels was evaluated the HiBiT luminescence assay. The results show that the designed hetero-bifunctional TPR proteins are able to reduce KRAS levels (Figures 24 and 25).

2.7 Hetero-bifunctional TPRs to direct KRAS for degradation via chaperone-mediated autophagy (CMA)

Hetero-bifunctional TPR proteins were designed to target endogenous KRAS for degradation via CMA (Figure 26). TPR constructs or empty vector (light grey) were transiently transfected into either HEK293T or DLD1 (colorectal cancer cell line) using Lipofectamine2000. 24 hours post transfection the cells were lysed, and KRAS levels were evaluated by western blot. The designed hetero-bifunctional TPRs that resulted in reduction of KRAS levels compared to the empty vector control are shown in white.

2.8 Variations in the linker sequence connecting a peptide ligand to an inter-repeat loop

The linker sequence connecting a peptide ligand to an inter-repeat loop was varied in order to optimise the binding affinity for the target for Nrf-TPR, a TPR protein designed to bind to the protein Keap1 (see Fig. 7). Glycine residues were introduced into the linker to provide flexibility and increased spatial sampling. The introduction of this more flexible linker sequence was found to increase the binding affinity of the Nrf-TPR protein (labelled 'Flexible') when compared with the consensus-like linker sequence. Altering the charge content of the linker sequence ('labelled 'Charged') and altering the conformational properties (based on the predictions of the program CIDER (Holehouse et al. Biophys. J. 112, 16-21 (2017)) of the loop by changing the amino acid composition of the linker sequence (labelled 'CIDER-optimised') also affected the Keap1-binding affinity (Figure 27).

2.9 Screen for Beta Catenin degradation by bispecific CPTR construct using "Phospho" targeting sequence

MIA PaCa-2 cells in which endogenous Beta Catenin has been tagged with the small peptide HiBiT were transfected with Scrambled or Beta Catenin targeted siRNA, or with CPTR scaffold based constructs utilising the Phospho targeting together with single function controls. After 24 hours Beta Catenin abundance was quantified using the lytic HiBiT assay. Mean and SEM of at least 4 replicates are shown in Figure 28 (Untreated: Untreated cells, Scr: Scrambled siRNA, Lipofectamine: Lipofectamine only treated cells, siRNA: Beta catenin targeted siRNA). The modular design of the constructs is shown in Table 12 and the sequences are shown in Table 13 (SEQ ID NOs: 1230-1304). The bispecific CPTR-based PPX229 caused a 36% reduction in signal relative to lipofectamine only demonstrating its ability to degrade Beta Catenin more effectively than both its single function controls (PPX239 3%, PPX244 14%). Positive control Beta Catenin targeted siRNA caused a 52% reduction in

signal relative to scrambled RNA.

2.10 Screen for Beta Catenin degradation by bispecific CPTR construct using "AXIN" targeting sequence

5 MIA PaCa-2 cells in which endogenous Beta Catenin has been tagged with the small peptide HiBiT were transfected with Scrambled or Beta Catenin targeted siRNA, or with CPTR scaffold based constructs utilising the AXIN targeting sequence together with single function controls. After 24 hours Beta Catenin abundance was quantified using the lytic HiBiT assay. Mean and SEM of at least 4 replicates are shown in Figure 29 (Untreated: Untreated cells, Src: Scrambled siRNA, Lipofectamine: Lipofectamine only treated cells, siRNA: Beta catenin targeted siRNA). The modular design of the
10 constructs is shown in Table 12 and the sequences are shown in Table 13 (SEQ ID NOs: 1230-1304). The bispecific CPTR-based PPX197 caused a 37% reduction in signal relative to lipofectamine only demonstrating its ability to degrade Beta Catenin more effectively than both its single function controls (PPX235 11%, PPX240 6%). The bispecific CPTR-based PPX202 caused a 36% reduction in signal
15 relative to lipofectamine only demonstrating its ability to degrade Beta Catenin more effectively than both its single function controls (PPX235 11%, PPX241 8%). The bispecific CPTR-based PPX225 caused a 33% reduction in signal relative to lipofectamine only demonstrating its ability to degrade Beta Catenin more effectively than both its single function controls (PPX235 11%, PPX244 14%).

2.11 Screen for Beta Catenin degradation by bispecific CPTR constructs using BCL9 targeting sequence

MIA PaCa-2 cells in which endogenous Beta Catenin has been tagged with the small peptide HiBiT were transfected with Scrambled or Beta Catenin targeted siRNA, or with CPTR scaffold based constructs utilising the BCL9 targeting together with single function controls. After 24 hours Beta
25 Catenin abundance was quantified using the lytic HiBiT assay. Mean and SEM of at least 4 replicates are shown in Figure 30 (Untreated: Untreated cells, Scr: Scrambled siRNA, Lipofectamine: Lipofectamine only treated cells, siRNA: Beta catenin targeted siRNA). The modular design of the constructs is shown in Table 12 and the sequences are shown in Table 13 (SEQ ID NOs: 1230-1304). The bispecific CPTR-based PPX226 caused a 36% reduction in signal relative to lipofectamine only
30 demonstrating its ability to degrade Beta Catenin more effectively than both its single function controls (PPX237 9%, PPX244 14%). Positive control Beta Catenin targeted siRNA caused a 52% reduction in signal relative to scrambled RNA.

2.12 Screen for Beta Catenin degradation by bispecific CPTR constructs using "LRH1" targeting sequence

35 MIA PaCa-2 cells in which endogenous Beta Catenin has been tagged with the small peptide HiBiT were transfected with Scrambled or Beta Catenin targeted siRNA, or with CPTR scaffold based constructs utilising the LRH1 targeting together with single function controls. After 24 hours Beta Catenin abundance was quantified using the lytic HiBiT assay. Mean and SEM of at least 4 replicates
40 shown in Figure 31 (Untreated: Untreated cells, Scr: Scrambled siRNA, Lipofectamine: Lipofectamine only treated cells, siRNA: Beta catenin targeted siRNA). The bispecific CPTR-based PPX203 caused

a 36% reduction in signal relative to lipofectamine only (LIPO) demonstrating its ability to degrade Beta Catenin more effectively than both its single function controls (PPX241 8%, PPX236 12%). Positive control Beta Catenin targeted siRNA caused a 52% reduction in signal relative to scrambled RNA.

5

The bispecific CPTR-based PPX227 caused a 30% reduction in signal relative to lipofectamine only (LIPO) demonstrating its ability to degrade Beta Catenin more effectively than both its single function controls (PPX241 3%, PPX244 14%). Positive control Beta Catenin targeted siRNA caused a 52% reduction in signal relative to scrambled RNA.

10

2.13 Expression of CPTR constructs in MIA PaCa-2 24 hours after transfection.

MIA PaCa-2 cells were transfected with CPTR scaffold based constructs PPX197, PPX202, PPX203, PPX225, PPX226, PPX227 and PPX229 which contain HA tags. After 24 hours cells were fixed, permeabilised and stained for HA tag. Hoescht was used as a nuclear counterstain to obtain images of HA and Hoescht staining for transfected cells (not shown). HA was detected in cells transfected with PPX197, PPX202, PPX203, PPX225, PPX226, PPX227 and PPX229 indicating that CPTR-based molecules are expressed in cells.

15

20

Targeted Ubiquitin Ligase	Degron sequence derived from	Targeted protein for Degradation	β -catenin-binding sequence derived from:	Scaffold
Mdm2	p53	β -catenin	axin	RTPR
Mdm2	p53	β -catenin	Bcl-9	RTPR
Mdm2	p53	β -catenin	TCF-4	RTPR
Mdm2	p53	β -catenin	ICAT	RTPR
Mdm2	p53	β -catenin	LRH-1	RTPR
Mdm2	p53	β -catenin	APC	RTPR
SCF ^{skp2}	p27	β -catenin	axin	RTPR
SCF ^{skp2}	p27	β -catenin	Bcl-9	RTPR
SCF ^{skp2}	p27	β -catenin	TCF-4	RTPR
SCF ^{skp2}	p27	β -catenin	ICAT	RTPR
SCF ^{skp2}	p27	β -catenin	LRH-1	RTPR
SCF ^{skp2}	p27	β -catenin	APC	RTPR
BTB-CUL3-RBX1	Nrf2	β -catenin	Bcl-9	RTPR
BTB-CUL3-RBX1	SPOP	β -catenin	Bcl-9	RTPR
APC/C	ABBA	β -catenin	Bcl-9	RTPR
APC/C	KEN	β -catenin	Bcl-9	RTPR
APC/C	DBOX	β -catenin	Bcl-9	RTPR
SIAH	PHYL	β -catenin	Bcl-9	RTPR
BTB-CUL3-RBX1	Nrf2	β -catenin	axin	RTPR
BTB-CUL3-RBX1	SPOP	β -catenin	axin	RTPR
APC/C	ABBA	β -catenin	axin	RTPR
APC/C	KEN	β -catenin	axin	RTPR
APC/C	DBOX	β -catenin	axin	RTPR
SIAH	PHYL	β -catenin	axin	RTPR
BTB-CUL3-RBX1	Nrf2	β -catenin	TCF-4	RTPR
BTB-CUL3-RBX1	Nrf2	β -catenin	APC	RTPR

Table 1

Targeted Molecule	Binding sequence derived from	Introduction to scaffold	Amino acid sequence	DNA sequence, <i>E. coli</i> codon optimised, 5' to 3'	DNA sequence, human codon optimised, 5' - 3'
β -catenin	axin	grafted helix	GAYPEYILDHIVRVQLEL (SEQ ID NO: 39)	GGTGCAATCCGGGAATACATCCTGGATATTCATG TTTATCGTGTTCAGCTGGAACCTG (SEQ ID NO: 85)	GGCGCCTACCCCGAGTACATCCTGGACATCCACGTGTACCCGG GTGACGCTGGAACCTG (SEQ ID NO: 86)
β -catenin	Bcl-9	grafted helix	SQEQLEHRYRSLITLYDIQL ML (SEQ ID NO: 40)	AGCCAAGAACAGCTGGAAACATCGTTATCGTAGCC TGATTACCTGTATGATATTCAGCTGATGCTG (SEQ ID NO: 87)	AGCCAGGAACAGCTGGAAACACACCCGGTACAGAAAGCCTGTATCACC CTGTACGACATCCAGCTGATGCTG (SEQ ID NO: 88)
β -catenin	TCF-4	loop insertion	QELGNDNDELMHFSYESTQD (SEQ ID NO: 41)	CAAGAAGCTGGCGGATAATGATGAACCTGATGCAC TTAGCTATGAAAGCACCCAGGAT (SEQ ID NO: 89)	CAGGAAGCTGGCGGATAACGACGAGCTGATGCACCTTCAGCTAC GAGAGCACCAGGAC (SEQ ID NO: 90)
β -catenin	ICAT	grafted helix	YAYQRAIVEYMLRLMS (SEQ ID NO: 42)	TATGCATATCAGCGTGCCATCGTTGAATATATG TGCGTCTGATGAGC (SEQ ID NO: 91)	TACGCCCTACAGCGGGGCCATCGTGGAGTACATGCTGCGGCTG ATGAGC (SEQ ID NO: 92)
β -catenin	LRH-1	grafted helix	YEQAIAAYLDALMC (SEQ ID NO: 43)	TATGAACAGGCAATTGCGACATATCTGGATGCAC TGATGTGT (SEQ ID NO: 93)	TACGAGCAGGCGCAATTGCCGCTACCTTGGACGCCCTGTATGTGC (SEQ ID NO: 94)
β -catenin	APC	loop insertion	SCSELEALEALELDE (SEQ ID NO: 44)	AGCTGTAGCGAAGAACTGGAAGCCCTGGAAGCAT TAGAACTGGATGAA (SEQ ID NO: 95)	AGCTGTAGCGAGGAACCTGGAAGCTCTGGAAGCACTGGAACCTG GATGAG (SEQ ID NO: 96)
KRAS	SOS-1	grafted helix	FEGLATNVLKALEG (SEQ ID NO: 45)	TTTGAAGGTATTGCACCTGACCAATTAATCTGAAAG CACTGGAAGGT (SEQ ID NO: 97)	TTTGAGGGAATCGCCCTGACCAACTACCTGAAGGCCCTGGAA GGA (SEQ ID NO: 98)
Tankyrase	Axin (8 residue)	loop insertion	REAGDGEE (SEQ ID NO: 46)	CGTGAAGCCGGTGATGGTGAAGAA (SEQ ID NO: 99)	AGAGAAGCTGGCGACGGCGAAGAA (SEQ ID NO: 100)
Tankyrase	Axin (14 residue)	loop insertion	HLQREAGDGEEFRS (SEQ ID NO: 47)	CATCTGCAGCGTGAAGCCGGTGATGGTGAAGAAT TTCGTAGC (SEQ ID NO: 101)	CATCTGCAGCGTGAAGCCGGTGATGGTGAAGAATTTTCGTAGC (SEQ ID NO: 102)

Table 2

Table 3

Targeted Ubiquitin Ligase Complex	Degron sequence	Introduction to scaffold	Amino acid sequence	DNA sequence E. Coli codon optimised 5' to 3'	DNA sequence human codon optimised 5' - 3'
Mdm2	p53	grafted helix	FAAYWILL ^{SEQ ID NO: 120} SAIG	TTTGCAGCCTATTGGAAATCTGCTCAGCGCATATGTT	TTTCGGCCCTACTGGAACCTCTCAGCGCCCTAGCGGT
Mdm2	p53	grafted helix	RFMDY ^{SEQ ID NO: 121} EWGL	CGCTTCATGGATTATTGGGAAGGTCG	CGGTTCATGATTAATCTGGAGAGGCGCTG
Mdm2	p53	grafted helix	TSFAEY ^{SEQ ID NO: 122} WALLAENL	ACCAGCTTTGCGGAGTATGTTGGCCCTGCTGGCGGAGAAATCG	ACCAGCTTTGCGGAGTATGTTGGCCCTGCTGGCGGAGAAATCG
SCFskp2	p27	loop insertion	AGNNEQ ^{SEQ ID NO: 123} EPKKR ^{SEQ ID NO: 67}	CGAGGTAGCAATGAAACAAAGAACCGAAAAACGTAT	GGCGGCTCTAATGAGCAAGAGGCCCAAGAGAGAAAGC
Keap1	Nrf2	loop insertion	DPET ^{SEQ ID NO: 68} GEL	GATCGGAAACCGGTGAACG	GATCTGAAACAGGGGGAACG
SPOP	SBC	loop insertion	LAD ^{SEQ ID NO: 69} EVTSISSIA	CTGGCATGTGATGAAGTTACCGSCACGACGACGACACCCGCA	CTGGCATGTGATGAAGTTACCGSCACGACGACGACCTCCAGGCG
APC/C	ABBA	loop insertion	SLISSA ^{SEQ ID NO: 70} FHVEDGKEN	AGGCTGASCAGCGCATTTTCATGTTTGTGAGAGTGCACAAAGAAAAT	TCCTCTGTCTAGCGCCCTCCACGTCCTTCGAGGAGCGGCAACAAAGAGAG
APC/C	ABOX	loop insertion	QRVL ^{SEQ ID NO: 105}	CACGCGTTCG	CAGAGAGTTCG
APC/C	KEN	loop insertion	SEKEN ^{SEQ ID NO: 71} VPP	AGCAGAGATAAAGAAAATATGTTCCGCCCT	AGCAGAGCAAAAGAAAACGTCGCCACCT
APC/C	D8OX	loop insertion	FRPL ^{SEQ ID NO: 72} GDVSN ^{SEQ ID NO: 72}	CCGGCTCTCCCGCTGGGTGAGTTAGCAATAAT	CCTAGACTGCCTCTCGGAGATGTGTCCAAACAAC
SIAH	PHYL	loop insertion	LRPVM ^{SEQ ID NO: 73} RVET ^{SEQ ID NO: 73}	CTGCGTCGCGGTGCAATAGTTCGTCGACCGGTT	CTAGGCGCTGTGGCTATATGTCGCGCCCTACAGTG
SCF ^{Rbw7}	cyclin-D3	loop insertion	PEQTS ^{SEQ ID NO: 106} EPDVAL ^{SEQ ID NO: 106}	CCGAGACAGACAGCAGCAACGACCGAGATGTTGCAATT	CCTGAGCAGACAGGAGCGCCTACGAGATGTGGCCATC
SCF ^{Rbw8}	myc	loop insertion	EPELEP ^{SEQ ID NO: 107}	GAACCGCCCTCTGGAACCG	GAGCTCCCTCTTTGAACCT
SCF ^{Rbw7}	cyclin E1	loop insertion	SLIPE ^{SEQ ID NO: 108} DR	AGGCTGATTCGGAACCGGAGCGT	TCCTGATCCCCGAGCCCTGACAGA
PIP		loop insertion	QRMT ^{SEQ ID NO: 109} DFIARR ^{SEQ ID NO: 109}	CGCGCTGATGACCGGATTTTATGCAACCTGCTCGT	CGAGAGAGGATGACCGGATTTCTAGCCGACAGAGGCGGG
paramyxovirus SV5-V protein		grafted helix	TVAYET ^{SEQ ID NO: 110} LQQVIG	ACGGTTGCATATTTTAACTCCGACAGGTTTATGTT	ACGTTGCCCTACTTCACACTGACAGAGGTTTACGGC
Hepatitis B virus X protein		grafted helix	ILFV ^{SEQ ID NO: 111} LHLRTVIG		ACGTTGCCCTACTTCACACTGACAGAGGTTTACGGC
Woodchuck Hepatitis virus X protein		grafted helix	ILFV ^{SEQ ID NO: 112} LHLRTVIG	ATTCTGCCTGCGAGTCTCGATGTCGTAACGTTTATGTT	ATTCTGCCTGCGCTGCTGCACCTCGAGACAGTGTATGGA
DDb1-CUL4			NFVW ^{SEQ ID NO: 113} HALQVIG	RAATTTGTTGCATGCGCATGCACTGCGTCAGGTTTATGTT	RAATTTGTTGCATGCGCATGCACTGCGTCAGGTTTATGTT
DDb1-CUL5		grafted helix	NITAD ^{SEQ ID NO: 114} LILRQVIG	AAATTAACCGCAGATCTGATTCCTGCTCAGGTTTATGTT	AAATCACCGCCGACCTGATCCTCGACACAGGTGTACGGA
Unknown	Bongor	loop insertion	RRRG ^{SEQ ID NO: 115}	CGTGGTCTGGT	AGAGAAGAGGCG
Unknown	iNOS	loop insertion	DINN ^{SEQ ID NO: 116}	GACATCAACAC	GACATCAACAAAC
β-TRCP		loop insertion	DEGNYE ^{SEQ ID NO: 117}	GATGAAGGCACTATGAA	GACGAGGGCAATTACGAG
5CF ^{Rbw4}		loop insertion	EEVSV ^{SEQ ID NO: 118} LASEPTV ^{SEQ ID NO: 118}	GAAGAAGAAGTTAGCCTGCGCAAGCAACCGACGATGTTCTGTAT	GAGGAAGAAGTTCTCTCTGGCCAGCGAGCCCTACCGATGTTCAATAT
CHIP	Hsp90	loop insertion	ASRME ^{SEQ ID NO: 119} VD	GCAAGCGGTATGGAAGAAGTTNGAT	GCCAGCAGATGGAAGAAGTTGGAC
CHIP	Hsp70	loop insertion	GPTIE ^{SEQ ID NO: 120} VD	GGTCCGACCATTTGAAGAAGTTNGAT	GGCCTACATATGGAAGAAGTTGGAC

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

5	RTPRc	GCAGAAGCACTGCGTAATCTGGGTCGTGTTTATCGTCGTCAGGGTCGTTATCAGCGTGCA	
	RTPRa-ii-H	GCCGAAGCTTGGTATAATCTGGGGAATGCCTATTACAGACAGGGGGATTATCAGCGCGCC	
	RTPRa-i-E		
10	GCAGAAGCATGGTATAATCTGGGTAATGCATATTATCGCCAGGGTGATTATCAGCGTGCC		
	RTPRa-iii-E	GCAGAAGCATGGTATAATCTGGGCAATGCATATTATCGTCAGGGTGATTATCAGCGTGCC	
	CTPRa-E	GCAGAAGCATGGTATAATCTGGGTAATGCATATTACAAACAGGGCGATTATCAGAAAGCC	
15	CTPRb-E	GCAGAAGCACTGAATAATCTGGGTAATGTTTATCGTGAACAGGGCGATTATCAGAAAGCC	
	RTPRb-E	GCAGAAGCACTGAATAATCTGGGTAATGTTTATCGTGAACAGGGCGATTATCAGCGTGCC	
	RTPRa-ii-E		
20	GCCGAGGCCTGGTATAACCTTGGCAACGCCTATTATCGTCAAGGCGACTACCAGAGAGCA		
	RTPRc-H	GCCGAGGCTCTGAGAAATCTGGGCAGAGTGACAGACGGCAGGGCAGATACCAGCGGGCC	
	CTPRb-H	GCCGAGGCTCTGAACAACCTGGGCAACGTGTACAGAGAGCAGGGCGACTACCAGAAGGCC	
25	RTPRb-H	GCCGAGGCTCTGAACAACCTGGGCAACGTGTACAGAGAGCAGGGCGACTACCAGCGGGCC	
	RTPRa-iv-E	GCCGAGGCCTGGTACAACCTGGGTAACGCCTATTATCGCCAAGGCGACTACCAGCGTGCA	
	CTPRa-H	GCCGAGGCCTGGTACAATCTGGGCAACGCCTACTACAAGCAGGGCGACTACCAGAAGGCC	
30	RTPRa-i-H	GCCGAGGCCTGGTACAACCTGGGCAACGCCTACTACCGGCAGGGCGACTACCAGCGGGCC	
		** ** *	** ** *
35	RTPRc	ATTGAATATTATCGTCGCGCACTGGAATTAGATCCGNNNNNN (SEQ ID NO: 170)	
	RTPRa-ii-H	ATTGAATATTATCAGCGGGCTCTGGAACCTGGATCCTNNNNNN (SEQ ID NO: 171)	
	RTPRa-i-E	ATTGAATATTATCAACGTGCACTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 172)	
	RTPRa-iii-E	ATCGAATATTATCAACGTGCACTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 173)	
	CTPRa-E	ATCGAGTATTATCAAAAAAGCACTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 174)	
40	CTPRb-E	ATCGAATATTATCAAAAAAGCGCTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 175)	
	RTPRb-E	ATTGAATATTATCAACGTGCGCTGGAATTAGATCCGNNNNNN (SEQ ID NO: 176)	
	RTPRa-ii-E	ATCGAATATTACCAGCGTGCCTTAGAATTAGATCCGNNNNNN (SEQ ID NO: 177)	
	RTPRc-H	ATCGAGTATTACCGCAGAGCCCTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 178)	
	CTPRb-H	ATCGAGTATTATCAGAAGGCCCTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 179)	
45	RTPRb-H	ATCGAGTATTATCAGAGAGCCCTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 180)	
	RTPRa-iv-E	ATTGAGTACTACCAACGTGCCCTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 181)	
	CTPRa-H	ATCGAGTATTATCAGAAGGCCCTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 182)	
	RTPRa-i-H	ATCGAGTACTACCAGAGAGCCCTGGAACCTGGACCCGNNNNNN (SEQ ID NO: 183)	
		** ** *	** * ** *
50			

Table 4

Multiple Alignment of DNA sequences of all CTPR and RTPR used in hetero-bifunctional CTPRs and RTPRs to date

5	089	AEAYSNLGNVYKERGQLQEAIIEHYRHALRL	118	NP_858058.1 (SEQ ID NO: 184)
	191	AVAWSNLGCVFNAQGEIWLAIHHFEKAVTL	220	
	327	ADSLNNLANIKREQGNIEEAVRLYRKALEV	356	
	264	NLACVYYEQGLIDLAIIDTYRRRAIEL	288	
10	079	AEAYSNLGNVYKERGQLQEAIIEHYRHALRL	108	NP_858059.1 (SEQ ID NO: 185)
	181	AVAWSNLGCVFNAQGEIWLAIHHFEKAVTL	210	
	317	ADSLNNLANIKREQGNIEEAVRLYRKALEV	346	
	254	NLACVYYEQGLIDLAIIDTYRRRAIEL	278	
15	079	AEAYSNLGNVYKERGQLQEAIIEHYRHALRL	108	NP_003596.2 (SEQ ID NO: 186)
	812	ESFYNLGRGLHQLGLIHLAIHYYQKALEL	840	
	637	AWYGLGMIYYKQEKFSLAEMHFQKALDI	664	
	568	EAWCAAGNCFSLQREHDIAIKFFQRAIQV	596	
20	275	AQSCYSLGNTYTLLQDYEKADYHLKHLAI	304	NP_001247.2 (SEQ ID NO: 188)
	058	YSQLGNAYFYLDHYAKALEYHHHDLTL	084	
	315	GRACWSLGNAYTALGNHDQAMHFAEKHLEI	344	
	247	NMGNIYLKQRNYSKAIKFYRMALD	270	
25	495	ALTNKGNTVFANGDYEKAAEFYKEAL	520	NP_783195.2 (SEQ ID NO: 189)
	238	NMGNIYLKQRNYSKAIKFYRMALD	261	
	486	ALTNKGNTVFANGDYEKAAEFYKEAL	511	
	715	AQAWMNMGGIQHIKGYVSARAYYERLQQL	744	
30	575	AEILSPLGALYYNTGRYEEALQIYQEAAL	604	NP_787057.2 (SEQ ID NO: 191)
	114	AQAANKNGNKYFKAGKYEQAIQCYTEAISL	143	
	610	YNLGKLYHEQGHYEEALSVMYKEAIQ	634	
	586	AQAWMNMGGIQHIKGYVSARAYYERLQQL	615	
35	446	AEILSPLGALYYNTGRYEEALQIYQEA	472	NP_114126.2 (SEQ ID NO: 194)
	018	AETFKEQGNAYYAKKDYNEAYNYYTKAIDM	47	
	300	AKAYARIGNSYFKEEKYKDAIHFNKSL	327	
	231	LGNDAYKKKDFDTALKHYDKAKEL	254	
40	365	NKGNECFQKGDYPQAMKHYTEAI	387	NP_006810.1 (SEQ ID NO: 196)
	028	AETFKEQGNAYYAKKDYNEAYNYYTKAIDM	057	
	228	AYSNLGNAHVFLGRFDVAAEYKKTLLQL	255	
	266	AQACYSLGNTYTLLQDYERAAEYHLRHL	293	
45	28	AEELKTQANDYFKAKDYENAIKFYSQAIEL	57	NP_056412.2 (SEQ ID NO: 198)
	318	DAYKSLGQAYRELGNFEAATESFQKALLL	346	
	1262	ETLKNLAVLSYEGGDFEKAEEYKRAMEI	1290	
	140	GNKYFKQGGYDEAIDCYTKGMD	161	
50	289	GNGFFKEGKYERAIECYTRGI	309	NP_078880.1 (SEQ ID NO: 202)
	600	CWESLGEAYLSRGGYTTALKSFTKASEL	627	
	172	KATYRAGIAFYHLGDYARALRYLQEA	197	
55				NP_055454.1 (SEQ ID NO: 203)
60				NP_689692.2 (SEQ ID NO: 204)

5	174 LGKIHLLLEGDLDKAIEVYKKAVE	196	NP_149017.2 (SEQ ID NO: 205)
	158 LGDLFSKAGDFPRAAEAYQKQLRF	181	NP_038460.3 (SEQ ID NO: 206)
	384 AYFNAGNIYFHHRQFSQASDYFSKALKF 207)	411	NP_001007796.1 (SEQ ID NO:
10	104 EAWNQLGEVYWKKGDVAAAHTCFSGAL	130	NP_612385.1 (SEQ ID NO: 208)
	814 EAWQGLGEVLQAQGQNEAAVDCFLTALRL	842	NP_065191.2 (SEQ ID NO: 209)
15	446 AKLWNNVGHALENEKNFERALKYFLQA	472	NP_861448.1 (SEQ ID NO: 210)
	597 ADLWYNLAIVHIELKEPNEALKNFNRALEL	626	
	251 YRRKGDLDKAIELFQRVLE	269	NP_001540.2 (SEQ ID NO: 211)
20	251 YRRKGDLDKAIELFQRVLE 212)	269	NP_001026853.1 (SEQ ID NO:
	079 AKTYKDEGNDYFKEKDYKKAVISYTEGL	106	NP_004614.2 (SEQ ID NO: 213)
	501 AKVHYNIGKNLADKGNQTAAIRYYREAVRL	530	NP_116202.2 (SEQ ID NO: 214)
25	482 AKVHYNIGKNLADKGNQTAAIRYYREAVRL	511	NP_001073137.1 (SEQ ID NO: 215)
	200 GNELVKKGNHKKAEIKYSESL	220	NP_006800.2 (SEQ ID NO: 216)
30	123 GNEQFKKGDYIEAESSYSRALEM	145	NP_003305.1 (SEQ ID NO: 217)
	438 ESLSLLGFVYKLEGNMNEALEYYERALRL	466	NP_001001887.1 (SEQ ID NO: 218)
	438 ESLSLLGFVYKLEGNMNEALEYYERALRL	466	NP_001539.3 (SEQ ID NO: 219)
35	375 AKTKNNLASAYLKQNKYQQAEEELYKEIL	402	NP_803136.2 (SEQ ID NO: 220)
	564 WFSLGCAYLALDYQGSAKAFQRCVTL	590	NP_060205.3 (SEQ ID NO: 221)
40	306 AESCYQLARSFHVQEDYDQAFQYYYQATQF	335	NP_055448.1 (SEQ ID NO: 222)

Table 5

CTPRa *E. coli* expression codon optimised

GCAGAAGCATGGTATAATCTGGGTAATGCATATTACAAACAGGGCGATTATCAGAAAGCCATCGAGTATTATCAA
AAAGCACTGGAAGTGGACCCGNNNNNN (SEQ ID NO: 223)

5

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPXX (SEQ ID NO: 224)

CTPRa *H. Sapiens* expression codon optimised

GCCGAGGCCTGGTACAATCTGGGCAACGCCTACTACAAGCAGGGCGACTACCAGAAGGCCATCGAGTATTATCAG
AAGGCCCTGGAAGTGGACCCNNNNNN (SEQ ID NO: 225)

10

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPXX (SEQ ID NO: 224)

RTPRa-i *H.sapiens* expression codon optimised

GCCGAGGCCTGGTACAACCTGGGCAACGCCTACTACCGGCAGGGCGACTACCAGCGGGCCATCGAGTACTACCAG
AGAGCCCTGGAAGTGGACCCNNNNNN (SEQ ID NO: 226)

15

AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPXX (SEQ ID NO: 227)

RTPRa-ii *H.sapiens* expression codon optimised

GCCGAAGCTTGGTATAATCTGGGGAATGCCTATTACAGACAGGGGGATTATCAGCGGCCATTGAATATTATCAG
CGGGCTCTGGAAGTGGATCCTNNNNNN (SEQ ID NO: 228)

20

AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPXX (SEQ ID NO: 227)

25

RTPRa-i *E.coli* expression codon optimised

GCAGAAGCATGGTATAATCTGGGTAATGCATATTATCGCCAGGGTGATTATCAGCGTGCCATTGAATATTATCAA
CGTGCACTGGAAGTGGACCCGNNNNNN (SEQ ID NO: 229)

30

AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPXX (SEQ ID NO: 227)

RTPRa-ii *E.coli* expression codon optimised

GCCGAGGCCTGGTATAACCTTGGCAACGCCTATTATCGTCAAGGCGACTACCAGAGAGCAATCGAATATTACCAG
CGTGCGTTAGAATTAGATCCGNNNNNN (SEQ ID NO: 230)

35

AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPXX (SEQ ID NO: 227)

RTPRa-iii *E.coli* expression codon optimised

GCAGAAGCATGGTATAATCTGGGCAATGCATATTATCGTCAGGGTGATTATCAGCGTGCCATCGAATATTATCAA
CGTGCACTGGAAGTGGACCCGNNNNNN (SEQ ID NO: 231)

40

AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPXX (SEQ ID NO: 227)

RTPRa-iv *E. coli* expression codon optimised

- 5 GCCGAGGCCTGGTACAACCTGGGTAACGCCTATTATCGCCAAGGCGACTACCAGCGTGCAATTGAGTACTACCAA
CGTGCCCTGGAAGTGGACCCNNNNNN (SEQ ID NO: 232)

AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPXX (SEQ ID NO: 227)

- 10 CTPRb *E. coli* expression codon optimised

GCAGAAGCACTGAATAATCTGGGTAATGTTTATCGTGAACAGGGCGATTATCAGAAAGCCATCGAATATTATCAA
AAAGCGCTGGAAGTGGACCCNNNNNN (SEQ ID NO: 233)

AEALNNLGNVYREQGDYQKAIEYYQKALEL-DPXX (SEQ ID NO: 234)

15

CTPRb *H. sapiens* expression codon optimised

GCCGAGGCTCTGAACAACCTGGGCAACGTGTACAGAGAGCAGGGCGACTACCAGAAGGCCATCGAGTATTATCAG
AAGGCCCTGGAAGTGGACCCNNNNNN (SEQ ID NO: 235)

- 20 AEALNNLGNVYREQGDYQKAIEYYQKALEL-DPXX (SEQ ID NO: 234)

RTPRb *E. coli* expression codon optimised

GCAGAAGCACTGAATAATCTGGGTAATGTTTATCGTGAACAGGGCGATTATCAGCGTGCCATTGAATATTATCAA
CGTGCGCTGGAATTAGATCCGNNNNNN (SEQ ID NO: 236)

25

AEALNNLGNVYREQGDYQRAIEYYQRALEL-DPXX (SEQ ID NO: 237)

RTPRb *H. sapiens* expression codon optimised

GCCGAGGCTCTGAACAACCTGGGCAACGTGTACAGAGAGCAGGGCGACTACCAGCGGGCCATCGAGTATTATCAG
30 AGAGCCCTGGAAGTGGACCCNNNNNN (SEQ ID NO: 238)

AEALNNLGNVYREQGDYQRAIEYYQRALELDPXX (SEQ ID NO: 239)

RTPRc *E. Coli* expression codon optimised

- 35 GCAGAAGCACTGCGTAATCTGGGTCGTGTTTATCGTCGTCAGGGTCGTTATCAGCGTGCAATTGAATATTATCGT
CGCGCACTGGAATTAGATCCGNNNNNN (SEQ ID NO: 240)

AEALRNLRVYRRQGRYQRAIEYYRRALELDPXX (SEQ ID NO: 241)

- 40 RTPRc *H. Sapiens* expression codon optimised

GCCGAGGCTCTGAGAAATCTGGGCAGAGTGTACAGACGGCAGGGCAGATACCAGCGGGCCATCGAGTATTACCGC

60

AGAGCCCTGGAAGTGGACCCNNNNNN(SAQ ID NO: 242)

AEALRNLRVYRRQGRYQRAIEYYRRALELDPXX(SAQ ID NO: 241)

5 Additional TPR repeat consensus sequences:

S75991 ALTINNIGTI YYAREDYDQA LNYEQALSL SRAV

(SAQ ID NO: 243)

CONSENSUS/80% XXhhXthuXh hXXXtphppA htXhppshtht XpX

(SAQ ID NO: 244)

CONSENSUS/65% spshhphGth hhphsphppA lphappAlpl pspX

(SAQ ID NO: 245)

CONSENSUS/50% spsatslGps atptucacca lcsap+ALcl sPss

(SAQ ID NO: 246)

Class	Key	Residues
alcohol	O	S,T
aliphatic	L	I,L,V
any	.	A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
aromatic	A	F,H,W,Y
charged	C	D,E,H,K,R
hydrophobic	H	A,C,F,G,H,I,K,L,M,R,T,V,W,Y
negative	-	D,E
polar	P	C,D,E,H,K,N,Q,R,S,T
positive	+	H,K,R
small	S	A,C,D,G,N,P,S,T,V
tiny	U	A,G,S
turnlike	T	A,C,D,E,G,H,K,N,Q,R,S,T

Table 6 RTPR and CTPR sequences

Protein Target	Paratope	struct	RSCB no	therapeutic area
Nrp1/2	RASQYFSSYLA	loop	2qqn	anti-angiogenic
Nrp1/2	AREDFRNRRLWYVMDY	helix	2qql	anti-angiogenic
IL18	KASGYSTDYFIY	helix	2yxt	anti-inflammatory
IL15	YRDRRRPS	loop	2xqb	anti-inflammatory
Thyroid stimulating hormone receptor	SGSSSDIGSNYVS	loop	2xwt	
EGF receptor	QQWSSHIFT	loop	3C09	cancer
EGF receptor	ASRDYDYAGRYFDY	helix	3C09	cancer
IL23	QNGHSFPFT	loop	3d85	anti-inflammatory
IL23	YINPYNDGTK	helix	3d85	anti-inflammatory
IL23	ARNWDVAY	helix	3d85	anti-inflammatory
Lymphocyte function-associated antigen 1 (LFA-1)	SGYSFTGHWMN	helix	3eoa	auto-immune
Lymphocyte function-associated antigen 1 (LFA-1)	MIHPSDSETR	helix	3eoa	auto-immune
Lymphocyte function-associated antigen 1 (LFA-1)	ARGIYFYGTTFDY	helix	3eoa	auto-immune
C3b	SGFSFTSSVS	helix	3g6j	anti-inflammatory
C3b	LIYPYNGFN	helix	3g6j	anti-inflammatory
FGF receptor 3	AASGFTFTSTGIS	helix	3grw	multiple myeloma
FGF receptor 3	ARTYGIYDLYVDYTE	helix	3grw	multiple myeloma
IL2	SRDYGYFD	helix	3iu3	anti-inflammatory
IL2	GYSFTRYWMH	helix	3iu3	anti-inflammatory
HER2	QWWWWPST	loop	3n85	breast cancer
HER2	ASGFSIWWSWIH	helix	3n85	breast cancer
membrane-type serine protease 1	YDNNQRPS	loop	3nps	metastasis of carcinomas
membrane-type serine protease 1	TFHIRRYRSGYYDKMDH	helix	3nps	metastasis of carcinomas
beta-secretase	ARGPFSPWVMDY	helix	3r1g	Alzheimer's disease
VEGF-R	TRHDGTNFD	helix	3s35	anti-angiogenic
VEGF-R	QQAKAFPPT	loop	3s37	anti-angiogenic
Irp5/6 receptor	SGHVNAVKNYGY	helix	3sob	bone-loss
hepsin protease	WINTETGS	Helix	3t2n	prostrate cancer
Factor D	WINTYTGE	helix	4d9r	anti-inflammatory
Factor D	GYTFTNYGMN	helix	4d9r	anti-inflammatory

SEQ ID NOs 247 to 276

Table 7

Target protein and binding partner	Grafting site in scaffold	Amino acid sequence	DNA sequence optimised for E. coli expression
β -catenin			
α -catenin	helix	RSKAHVLAASVEQATQNFLKEGEIAKESQ	CGCAGCAAAAAGCGCATGTGCTGGCGGCGAGCGTGGAACAGCGGCCAGAACCTTCTGGAAAAAGCGAACAAGATTGG
α -catenin	helix	RTLTVRLLEPLVTQVTTLV	AAAGAAAAGCCAG CGCACCTGACCGTGGAAACGCTGCTGGAAACGCTGGTGACCCAGGTGACCACCCCTGGTG
APC Membrane recruitment protein	loop	RREQLEAQEARARAEAHAREAHAREAYTREAYG REAYAREAHWTWEAHGREARTREAQA	CGCCGCGAAACAGCTGGAAGCGCAGGAAGCGCGCGCGGCGGAAGCGCATGGCGCGAAGCGCATCTGGGAAGCGCATGGCGCGAAGCGCGCACCCGCGAA GCGCAGGCG
SOX	loop	D..EFDQYL	GATNNNNNGAAITTTGATCAGTATCTG
kindlin 2	loop	QALLDKAKINQ GWLDSSRSLMEODKENELLRF	CAGGCGCTGCTGGATAAAGCGAAAAITTAACCAAGGCTGGCTGATAGCAGCGCAGCCCTGATGSAAACAGGATAAAAGAAAAAC GAAGCGCTGCTGCCTTT
KRAS			CCCCGTGACATCAGCTAC
phage-display library peptide KR-pep1	loop	PLYISY	
Synthetic peptide 225-1	helix	SIEDLHEYWARLNWNLVYA	AGCATTGAAGATCTGCATGAATATTGGCGCGCCTGTGGAACTATCTGTATGTGGCG
Synthetic peptide 225-15a	helix	QASLEELHEYWARLNWNRVA	CAGGCGAGCCTGGAAGAAGTCATGAATATTGGCGCGCCTGTGGAACTATCGCGTGGCG
Synthetic peptide 225-15b	helix	NASIKQLHAYWQRLYAYLAAVA	AACGGGAGCATTAACAGACTGCATGCGTATTGGCAGCGCCTGTATCGGTATCTGGCGGCGGTGGCG
phage-display library peptide KR-pep3	loop	CMWWREICPVWW	TGCATGTGGTGGCGCGAATAATTGCCGGTGTGGTGG
Raf-S	loop	FARKTFKLAF	TTTTGCGCGCAAAACCTTTCTGAAACTGGCGTTT
NF1	loop	ARRFFLDIAD	GCGCGCGCTTCTTTCTGGATATTGCGGAT
Rasin peptide 2	loop	FRWP...RL...	TTTCGCTGGCGNNNNNNNNCCTGNNNNNN
Rasin peptide 1	loop	t.VFXh.p	AGCATTGTGTTTGGCGCGCATGAT
NF1 monobody peptide (74-84)	loop	VGHGOVVVY	TATGGCCATGGCCAGGTGTATTATTAT

farnesyl transferase 1	loop	ENPKQN	GAAACCCGAAACAGAAC
farnesyl transferase 2	loop	DAYECLDASRPW	GATGCGTATGAATGCCTGGATCGAGCGGCCCGTGG
farnesyl transferase 3	loop	K3RDFYH	AAATCCGCGATTCTATCAT
c-Myc			
Aurora A	helix	AGVEHQLRREVEIQSH	GCGGGCTGGAAACATCAGCTGCGCGCGGAAGTGGAAATTCAGAGCCAT
Aurora A	loop	WSVHAPSSRRITpLAGTLDYLPPEMI	TGGAGCTGTCATGCGCGAGCAGCCGCCGCACTGGCGGCACCTGGATTATCTGCCGCGGAAATGATT
Aurora A	helix	TYQETY	ACCTATCAGGAAACCTAT
Omomyc	helix	QAEQKLSEEDLR KRREQLKHLEQLRNSCA	CAGCGGGAAGAAACAGAAACTGAGCGAAGAAGATCTGCTGCGCAACGCGCGAACAGCTGAAACATAAACTGGAAACAGCTGCGCAACAGCTGCGCG
Myc H1 F8A		NELKRSFAALRDQI	AACGAACAGCAACGCGCTTTCGCGCGCTGCGCGATCAGATT
Myc H1 F8A S6A		NELKRAFAALRDQI	AACGAACAGCAACGCGCTTTCGCGCGCTGCGCGATCAGATT
MIP	helix	IREKNHYHRQEVDDLRRQNALLEQVRL	ATTGCGGAAAAAAACCATTTATCATCGCCAGGAAGTGGATGATCTGCGCGCAGAACGCGCTGCTGGAAACAGCAGGTGCGCG
PIN1	loop	FNHITNASQWE	TTTAACCATATTACCAACGCGAGCCAGTGGAA
PIN2	loop	GDIGAFSRQM	GGCGATCTGGCGCGCTTAGCCGCGGCCAGATG
9E10 paratope	loop	RSEFYVYGYNTYYVSAMID	CGCAGCGAATTTTATTATTATGGCAACACCTATTATTATAGCGCGATGGAT
BIN1	loop	QHDYTATDE	CAGCATGATTATACCGCGACCGATGAA
BIN1	loop	QNPEEQDEGW	CAGAACCGGGAAGAACAGGATGAAGGCTGG
BIN1	loop	EKCRGVFENF	GAAAAGTCCGCGCGGTGTTTCGGGAAACTTT
BRD4			
JMJD6	loop	KWTLERLKRKYRN	AAATGGACCCCTGGAACGCTGAAACGTAATAACCGTAAC
murine leukemia virus integrase	loop	TWRVQRSNPLKIRLTR	ACCTGGCGTGTTCAGCGTTCTCAGAACCCGCTGAAAAATCCGCTGACCCGT
Aurora A			
TPX2	loop	SYSYDAPSDFINFSSLLDEGDTQNIDSWFEEKAN	AGCTATAGCTATGATGCGCCGAGCGGATTTTAACTTTAGCAGCCTGGATGATGAAGCGGATACCCAGAACATTGATAGCTG
TPX3	loop	LEN	GTTTGAAGAAAAAGCGAACCTGGAAAAAC
	loop	MSQVKSSYSYDAPSDFINFSSLLD	ATGAGCCAGGTGAAGTCATCTTATTCCTATGATGCCCCCAGCGATTTCATCAATTTTCATCCTTGGATGATGAA
CK2alpha			

CK2beta	loop	RLYGKIHPMAYQLQ	CGCCTGATGGCTTTAAATTCATCCGATGGCGTATCAGCTGCAG
Notch			
MAML1	helix	SAVMERLRRIELCRRHHST	AGCGGGTGATGGAAAGCGCTGGCGCGCGCATTTGAACTGTCCGCGCGCATCAGCAACC
Cdk2			
aptamer library		LVCKSYRLDWEAGALFRSLF	CTGGTGTGCAAAAGCTATGCGCTGGATTGGGAAGCGGGCGGCTGTTTCGAGCCGTGTTT
aptamer library		YSFVHHGFFNFRVSWREMLA	TATAGCTTTGTGCATCATGGCTTTTTTAACTTTCGCGTGAGCTGGCGCGAAATGCTGGCG
peptide	loop	TAALS	ACCGCGGCGCTGAGC
peptide	loop	TALLS	ACCGCGCTGCTGAGC
peptide	loop	LAALS	CTGGCGGCGCTGAGC
peptide	loop	DAALT	GATGGCGCTGACC
peptide	loop	YAALQ	TATCGGGCGCTGCAG
peptide	loop	SKL-RFTGCSC	AGCAAACTGNNNGCTTTACCGGCTGCAGCTGC
RXL peptide	loop	PVKRRLLF	CCGGTGAACGCGCCTGTTTCTG
p21	loop	GRKRQTSMTDFVHSKRRLIFSKRKP	GGCGGCAAAAGCGCGCCAGACCAGCATGACCGATTTTATCATAGCAAAAGCGCGCTGATTTTAGCAACGCAAAACCG
PLK1			
peptide	loop	MAGPMQTSpTPKNAGKK	ATGGCGGGCGCGATGCAGACGACGACCCCGGAAAAACGCGGGGCAAAAAA
PBP1	loop	FDPPLHSpTA	TTTGATCGCGCGTGTCATAGCACCGCG
designed			
peptide	loop	PLHSpTAI	CCGCTGCATAGCACCGCGATT
designed			
peptide	loop	MDSpTPL	ATGGATAGCACCGCGCTG
Emi2	loop	FSQHKpTI	TTTAGCCAGCATAAACCCAGCATT
HEF1	loop	LHYPSpTTALQE	CTGCATTATCCGAGCACCCGCGGCTGCAGGAA
cdc-25	loop	LLCSpTPNGL	CTGCTGTGCAGCACCCCGGAACGGCCTG
BCR-ABL			
proline-rich			
peptide	helix	APSPppppp	GGCGGAGCTATCCGCGCGCGCGCGCG
pp1c			
consensus			
sequence	loop	RV.F	CGCGTGNNTTT
consensus			
sequence	loop	SILK	AGCATTTCTGAAA

KNL1	loop	SRVSVFADTIKVFQT	AGCGCGCGTGAGCTTTGGGATACCAITTAAGTGTTTCAGACC
MCL-1			
NOXA BH3	helix	AELVESATQLRRFGDKLNFROKLL	GCGGAAGTGAAGTGAAGCGGACCCAGCTGCGCGCTTTGGGATAAACTGAACCTTTGCGCCAGAAACTGCTG
MCL-1 BH3	helix	KALETLR.VGD.VQRNHETAF	AAAGCGTGAAACCCCTGCGCNGNNGTGGCGATNNNGTGCAGCGCAACCATGAAACCGCGCTTT
Bcl-2 and BCL-XL			
Bim BH3	helix	IWIQAQLRRIGDEFNAYYARR	ATTGGATTGCGCAGGAAGTGGCGCATTTGGCGATGAATTTAACGCGTATTATGCGGCGCGC
Bak BH3	helix	GQVGRQLAIGDDINR	GGCCAGGTGGCGCGCAGCTGGGATTATTGGCGATGATATAACCGC
Bad BH3	helix	NLWAAQRYGRELRRMSDEFVDSFKK	AACGTGTGGGCGCGCAGCGCTATGGCGCGAACTGGCGCGCATGAGCGATGAATTTGTGGATAGCTTTAAAAAA
Jun			
library-selected peptide	helix	SIAATLEKEANLEKMNKLLAAEIESLLKEKDKLES VLNYHE	AGCATCGCCGCCACCCCTGGGAAGGAGGAGGCCAACCTGGAGAAGATGAACAAGAAGCTGGCGCGGAGATCGAGAGCCCTG CTGAAGGAGAAGGACAAGCTGGAGAGCGTGTGTAACCTACCCAGAG
library-selected peptide	helix	VQEIEQEIELEKRIKQIQEFQIEQQIALL	GTTCAGGAAATCGAACAGGAAATCCAGGAACTGGAAAAACGTATCAACAGATCCAGGAAATTCAGGAAATCGAAACAGC AGATGCGCTGCTG
BFL1			
NOXA BH3	helix	ATQLRRFGDKLNFRO	GCGACCCAGCTGCGCGCTTTGGCGATAAACTGAACCTTCGCCAG
BAX			
Bcl-2 BH3	helix and loop	EIVAKYIHYKLSQRYEWDA	GAAATTGTGCGGAAATATATTTCATTATAAACTGAGCCAGCGCGCTATGAATGGGATGCG
elF4E			
elF4G	helix	KKRYDREFLLGFQF	AAAAACGCTATGATCGCGAAATTCGCTGGGCTTCAGTTT
elF4G	helix and loop	GKKRYDREFLLGFQFIFASMKPEGLPHISDVVL	GGCAAAAAACGCTATGATCGGAAATTTCTGCTGGGCTTCAGTTTATTTTGGAGCATGCAGAAACCGGAAGCCTGCCGCAT ATTAGCGATGTGGTCTG
optimised peptide	helix and loop	TKLIYERAFMKNLRGSPSLQTPPSNVPSCLLRT	ACCAAACTGATTATGAACGCGCTTTATGAAAACTGGCGGCGAGCCCGCTGAGCCAGACCCCGCGGAGCAACGTGCCGAG CTGCTGCTGCGGCGCACCC
Fos			
library-selected peptide	helix	AJARLEERVKTLEIYELRSKANMLREQIAQLGA P	GCGATTGCGCGCTGGAAGAACCGTGAACCCCTGAAAGCGGAAATTTATGAACCTGCGAGCAAGCGAACAATGCTGCGCG AACAGATTGCGCAGCTGGGCGCGCGC

library-selected peptide	helix	<u>A</u> IARLEERV <u>K</u> TLKAEI <u>Y</u> ELQSEAN <u>M</u> LREQIA <u>L</u> GAP	GCGATTGCGCGCTGGAAGAACGGTGAAACCCTGAAAGCGGAAATTTATGAACCTGGCGAGCAAGCGAACATGCTGCGCG AACAGATTGCGCAGCTGGGCGCGCGC
HDAC4			
SMRT corepressor	loop	HIRGSITQGIPRSV	CACATCCGTGGTTCTATCACCCAGGGTATCCCGCGTTCTTACGTT
BCL6			
SMRT and N-CoR corepressors	loop	GRSIEIPR	GGCCGAGCATTGAGAAATTCGCGCG
SMRT and N-CoR corepressors	loop	GLVATVKEAGRSIEIPREL	GGCCTGGTGGACCGTGAAAGAGCGGGCGGCGAGCATTGAAATTCGCGCGAAGAACTG
Tau			
alpha-tubulin	loop	KDYEEVGVDSE	AAAGATTATGAAGAAAGTGGCGTGATAGCGTGGA
beta-tubulin	loop	YQQYQDATADEQG	TATCAGCAGTATCAGATGCGACCGCGGATGAACAGGGC
PD-L1			
HIP1R	loop	DAVRRIEDMMNQARHASSGV	GATGCGGTGCGCGCATTGAAGATATGATGAACCGGCGGCCATGCGAGCAGCGCGTG
KDM4A			
library-selected peptide	loop	YVYNTSRGWRWYT	TACGTTTACAACACCGTTCTGGTTGGCGTTGGTACACC
EGFR			
EGFR			
(juxtamembrane coiled-coil domain)	helix	VRKRTLRLQLQERLVE	GTGCGCAACGCAACCTGCGCGCCTGCTGCAGGAACGCGAACTGGTGGA
RAB25			
RFIP1	helix	ROVRELENYIDRLVRVMEETPNILRIPR	CGCCAGGTGCGGAACTGGAAAACATATATTGATGCGCCTGGTGCGGTGATGGAAGAAACCCCGAACAATTCGCGCATTCGCGCG C

GPCRs and other transmembrane proteins			
PAR1 pepducin	N-term	pal-KKSRALF-NH2 (pal- = palmitoyl; -NH2 = amino group)	synthetic peptide
PAR1 pepducin	N-term	pal-RCLSSSAVANRS-NH2	synthetic peptide
PAR1 pepducin	N-term	pal-RSLSSSAVANRS-NH2	synthetic peptide
PAR1 pepducin	N-term	pal-AVAINRSKKSRALF-NH2	synthetic peptide
PAR1 pepducin	N-term	pal-RCESSAEANRSKKRELf-NH2	synthetic peptide
PAR1 pepducin	N-term	pal-ASSESQRVVSIL-NH2	synthetic peptide
PAR1 pepducin	N-term	pal-ASSASQEVVYSIL-NH2	synthetic peptide
PAR2 pepducin	N-term	pal-RSSAMDENSEKKRKAIAK-NH2	synthetic peptide
PAR2 pepducin	N-term	pal-GDENSEKKRKAIAK-NH2	synthetic peptide
PAR4 pepducin	N-term	pal-SGRRYGHALR-NH2	synthetic peptide
PAR4 pepducin	N-term	pal-ATGAPRLPST-NH2	synthetic peptide
PAR4 pepducin	N-term	pal-RLAHGYRRGS-NH2	synthetic peptide
CXCR1/2 pepducin	N-term	pal-RTLFLKAHMGQKHR-NH2	synthetic peptide
CXCR1/2 pepducin	N-term	pal-LCA- YSRVGRSVTD-NH2 (LCA- = lithocholic acid)	synthetic peptide
CXCR4 pepducin	N-term	pal-HSKGHQKRKALK-NH2	synthetic peptide
CXCR4 pepducin	N-term	pal-MGYQKKLRSMTD-NH2	synthetic peptide
CXCR4 pepducin	N-term	pal-MGYQKKLRSMTDKYRL-NH2	synthetic peptide
SLP3 pepducin	N-term	myristoyl-GMRPYDANKR-NH2	synthetic peptide
SLP3 pepducin	N-term	myristoyl-GRPYDAN-NH2	synthetic peptide
FRP2 pepducin	N-term	pal-KIHKKGMKSSRPLRV-NH2	synthetic peptide
FRP2 pepducin	N-term	pal-KIHKKGMKIS-NH2	synthetic peptide
FRP2 pepducin	N-term	pal-KIHKKGMKSSR-NH2	synthetic peptide

LGR7 pepducin	N-term	pal-KRKALKALIINEKKVQ-H (-H = hydrogen)	synthetic peptide
SMO pepducin	N-term	pal-TFVADWRNSNRY-H	synthetic peptide
SMO pepducin	N-term	pal-TWAWHTSFKALGTTTYPQLSGKTS-H	synthetic peptide
SMO pepducin	N-term	pal-RGVMTLFSIKSNHPGLLSEKAASKINETMLR-H	synthetic peptide
IGF1R pepducin	N-term	pal-RNNSRLNGVLY-NH2	synthetic peptide
CD226 pepducin	N-term	pal-RRERRDLFTE-NH2	synthetic peptide
TRPV1 TRPducin	N-term	pal-MGETVKNKIAQES-NH2	synthetic peptide

SEQ ID NOs 277 to 388 SEQ ID NOs 389 to 469

Table 8 (SEQ ID NOs: 277 to 469)

Ubiquitin Ligase	Degron sequence derived from	Grafting site in scaffold	Amino acid sequence	DNA sequence E. Coli codon optimised 5' to 3'
Mdm2	Consensus	helix	F[V]P[3]W[V]P[2,3][VIL]	TTT[NNNNNNN]EXCEPT CCN[TTGG]NNN[2,3]{CCN}[GTG/ATT/CTG]
Mdm2	p53	helix	EAQWAAL	GAAGCGCAGTGGGGGGGGCTG
Mdm2	p53	helix	FEAQWAAL	TTTGAAGCGCAGTGGGGGGGGCTG
Mdm2	p63	helix	FQHIWDFL	TTTCAGCATATTGGGATTCTG
Mdm2	p73	helix	FEHLWSSL	TTTGAACATCTGTGGAGAGGCTG
SCF[skp2]	Consensus	loop	[DE]PTP.K	NNN[GAT/GAA]NNNACCCGNNNAAA
Cul3-Keap1	Consensus	loop	[DMS].[DES][TNS] GE	[GAC/AAC/AGC]NNN[GAC/GAA/AGC][ACC/AAC/AGC]GGCGAA
Cul3-Keap1	Sequestosome-1	loop	DPSTGEL	GATCGAGCACCGGGGAACTG
Cul3-Keap1	IKK β	loop	NQETGE	AACCAGGAAACCGCGGAA
Cul3-Keap1	APC membrane recruitment protein 1	loop	SPETGE	AGCCCGGAAACCGGGGAA
Cul3-Keap1	Prothymosin alpha	loop	NEENGE	AACGAAAGAAACCGGGGAA
Cul3-Keap1	Nucleosome-remodeling factor subunit BPTF	loop	DPENGE	GATCGGAAACCGGGGAA
Cul3-Keap1	Serine/threonine-protein phosphatase PGAM5, mitochondrial	loop	INVEGE	AACGTGGAAAGCGGCGAA
Cul3-Keap1	Nuclear factor erythroid 2-related factor 2	loop	DEET GE	GATGAAGAAACCGGGGAA
Cul3-Keap1	Partner and localizer of BRCA2	loop	DEETGE	GATGAAGAAACCGGGGAA
Cul3-Keap1_2	Consensus	loop	QD.DLGV	CAGGATNNNGATCTGGGTGTG
Cul3-SPOP	Consensus	loop	[AVP].[ST][ST][ST]	[GGG/GTG/CCG]NNN[AGC/ACC][AGC/ACC][AGC/ACC]
Cul3-SPOP	Map kinase phosphatase	loop	ELDSPSTSSSS	GAACTGGATAGCCCGAGCAGCACCAGCAGCAGCAGC
Cul3-SPOP	Androgen receptor	loop	ASSTT	GGGAGCAGCACCACC
Cul3-SPOP	Map kinase phosphatase	loop	DEVTSSTSSST	GATGAAGTGACCGAGCAGCACCAGCAGCAGCAGC
Cul3-KELCH	Consensus	loop	E.E.E.[AV]DQH	GAANNNGAAGAANNNGAA[CCG/GTG]GATAACCAT
Cul3-KELCH	Serine/threonine-protein kinase WNK1	helix/loop	EPEEPEADQH	GAAACCGGAAGAACCGGAAGCGGATCAGCAT
Cul3-KELCH	Serine/threonine-protein kinase WNK3	helix/loop	ECEEYVDQH	GAATCGGAAGAAACCGGAAGTGTGATCAGCAT

CUI3-KELCH	Serine/threonine-protein kinase WNK4	helix/loop	EPPEPDAQH	GAAACGGAGAACCGGAAGCGGATCAGCAT
CUI3-KELCH	Nuclear factor erythroid 2-related factor 2	helix/loop	ILWRQDIDLGV	ATTCTGTGGCGCCAGBATAATTGATCTGGGGTGS
KELCH actinfilin	Consensus	loop	[AP]P[IMV][IM]V	[GGG/CCG]CCG[ATG/ATG][ATT/ATG]GTG
APC/C	ABBA	loop	[FIV].[L][IMVP][FHY].[DE].[Q.3][DEST]	[TTT/ATT/ATG/CTG]NNN[ATT/CTG/ATG/ATG/CTG/CCG][TTT/CAT/TAT][NNN[Q.3]][GAT/GAA/AGC/ACC]
APC/C	Cyclin-A2 :ABBA	loop	FTIHVD	TTTACCAATTCATGTGGAT
APC/C	Consensus	loop	.KEN.	NNNAAGAAACNNN
APC/C	DBOX consensus	loop	R.L.[L][IMV].	NNNCGTNNNNNCTGNNNNN[CTG/ATT/CTG/ATG]NNN
APC/C	Shugoshin 1 :DBOX	loop	RLSLPKKN	CGCTGAGCCTGAGCCGCGGAAAAAAC
APC/C	Shugoshin 1 :DBOX	loop	RSLKKHON	CGCAGCAGCCTGAAAAAAGATTGCAAC
APC/C	Shugoshin 1 :DBOX	loop	HLSLKDTN	CATCTGAGCCTGAAAGATATTACCAAC
APC/C	Bcl-2-like protein 11 :DBOX	loop	RSPLFIF	CGCAGCCGCTGTTTATTTT
APC/C	Bcl-2-like protein 11 :DBOX	loop	RSLLSR	CGCAGCAGCCTGCTGAGCCGC
APC/C	Securin :DBOX	loop	RKALGTV	CGCAAGCGCTGGGCACCGTG
APC/C	Securin-2 :DBOX	loop	RKALGTV	CGCAAGCGCTGGGCACCGTG
APC/C	Ski-like protein :DBOX	loop	RCLPQV	CGCTGTGCTGCCGCGAGGTG
APC/C	Aurora kinase B :DBOX	loop	RLPLAQV	CGCTGCCGCTGGCGCAGGTG
APC/C	Serine/threonine-protein kinase PLK1 :DBOX	loop	NRKPLTVLN	AACAGGAAGCCCTGACCGTGTGAAC
APC/C	Cyclin-A2 :DBOX	loop	RAALAYL	CGCGCGCGCTGGGGTGCTG
APC/C	G2/mitotic-specific cyclin-B1 :DBOX	loop	PRTALGDIG	CGCGCACCGCGCTGGGCGATATTGGC
APC/C	G2/mitotic-specific cyclin-B3 :DBOX	loop	RSAFEDLTN	CGCAGCGCTTTGAAGATCTGACCAAC
APC/C	S-phase kinase-associated protein 2 :DBOX	loop	HRKHLEIP	CATGCAAAACATCTGCAGGAAATTCCG
APC/C	Nuclear autoantigen Sp-100 :DBOX	loop	RSLGLLS	CGCAGCGCCTGCAGCTGAGC
APC/C	Nucleolar and spindle-associated protein 1 :DBOX	loop	RRLULA	CGCCGCGCCTGATTCTGGCG
APC/C	BRCA1-A complex subunit RAP80 :DBOX	loop	RHCLPTL	CGCCATTGCTGCCGACCCCTG

APC/C	BARD1: DBOX	loop	RNLLHDN	CGCAACCTGCTGCATGATAAC
APC/C	BARD1: DBOX	loop	PAALDR	CGCGCGGCGCTGGATGCGCTG
APC/C	E3 ubiquitin ligase RNF157: DBOX	loop	RKKL	CGCAAAAAAATG
APC/C	E3 ubiquitin ligase RNF157: DBOX	loop	RRRL	CGCGCGCGCTG
APC/C	Nuclear-interacting partner of ALK: DBOX	loop	RARLSS	CGCGCGCGCTGTGACAGCAGC
APC/C	Nuclear-interacting partner of ALK: DBOX	loop	RLLPVE	CGCTGCGCTGTGTGCGGAA
APC/C	Tribbles homolog 3: DBOX	loop	RKKLVLE	CGCAAAAAAATGCTGTGTGAA
APC/C	Anillin: DBOX	loop	RENLRQK	CGCGAAAAAATGACAGCGCAAA
APC/C	Anillin: DBOX	loop	RQPLSEA	CGCGAGCGCTGTGAGCGAAGCG
APC/C	Ninein-like protein: DBOX	loop	RTOLETK	CGCACCCAGCTGGAACCAAA
APC/C	Dual specificity protein kinase TTK: DBOX	loop	RNSLRQT	CGCAACAGCCTGGGCGAGACC
APC/C	Inactive serine/threonine-protein kinase TEX14: DBOX	loop	RYGLHPD	CGCTATGCGCTGCATCCGGAT
APC/C	Bub1b	loop	AKENE	GCAGAAAGAAACGAA
APC/C	Bub1b	loop	SKENW	AGCAAGAGAAACGCTG
APC/C TPR1	Consensus	loop	[ILM]RS	NNN[ATT/CTG/ATG]CTG
SCF ^{hsw/1}	Consensus	loop	[LVNMP]{0,2}pTP..[pspt]	[CTG/ATT/GTT/ATG/CCG][NNN{0,2}]ACCCGNNNNN[AGC/ACC]
SCF ^{hsw/1}	Neurogenic locus notch homolog protein 1	loop	PFLpTPSPPE	CGTTTCTGACCCCGAGCCCGGAA
SCF ^{hsw/1}	Uracil-DNA glycosylase	loop	PGTPPSP5	CGGGCACCCCGCGAGCAGC
SCF ^{hsw/1}	G1/S-specific cyclin-E1	loop	LLpTPPQp5	CTGCTGACCCCGCGCAGAGC
SCF ^{hsw/2}	Consensus	loop	[LVNMP]{0,2}pTP..E	[CTG/ATT/GTT/ATG/CCG][NNN{0,2}]ACCCGNNNNNGAA
SCF ^{hsw/2}	Neurogenic locus notch homolog protein 1	loop	PFLpTPSPPE	CGTTTCTGACCCCGAGCCCGGAA
SCF ^{hsw/2}	G1/S-specific cyclin-E1	loop	SLPpTPDK	AGCTGATTGCGACCCCGGATAAA
SCF ^{hsw/2}	Sterol regulatory element-binding protein 1	loop	SDSEPD	AGCGATAGCGAACCGGAT
SCF ^{hsw/2}	SV40	loop	TPxxE	ACCCGNNNNNGAA

SCF ^{flaw7}	Nuclear factor NF-kappa-B p105 subunit	loop	psGVETpsF	AGCGGGCTGGAACACAGCTTT
SCF ^{flaw7}	E3 ubiquitin ligase RNF157	loop	LKLKSL	CTGAAGAACTGAAAAAAGCCTG
SCF ^{flaw7}	NF-kappa-B inhibitor alpha	loop	psGLDps	AGCGGCTGATAGC
SCF ^{flaw7}	NF-kappa-B inhibitor epsilon	loop	DpsGLEps	GATAGCGCAATTGAAAGC
SCF ^{flaw7}	Programmed cell death protein 4	loop	psSRDGRGDS	AGCAGCGGATAGCGCGCGCGGATAGC
SCF ^{flaw7}	NF-kappa-B inhibitor beta	loop	DpsGLGps	GATAGCGGCTGGGCAAC
SCF_TIR1	Consensus	loop	[VLA][VU][GWPP][VU]...R.	NNN[GTG/CTG/ATT/GCG][GTG/CTG/ATT][GGTTATCCGCG[GTG/CTG/ATT]NNNNNCGTNNN
Cul4-DOB1-	Consensus	loop	[NQ][0.1].[ILMV][ST][DEN][FY][FY].{2.3}[KR][2.3][^DE]	[AAC/CAG[0.1]]NNNNNN[ATT/CTG/ATG/GTG][AGC/ACC][GAC/GAA/AAC][TTT/TAT][TTT/TAT][NNN[0.3]]AAA[CGT[2.3]]NNN [^GAA/GAT]]
Cul4-DOB1-	Consensus	loop	[NQ][0.1].[ILMV][DEN][HMF][F	[AAC/CAG[0.1]]NNNNNN[ATT/CTG/ATG/GTG][ACC][GAC/GAA/AAC][CAT/ATG/TTT/TAT][TTT/TAT/ATG/TTT/TAT][^GAA/GAT]]
Cdt2_2	Consensus	loop	MY].[2.3][KR][2.3][^DE]	
SCF_TRCP1_1	Consensus	loop	DpsG.[2.3][pspt]	GATAGCGGC[NNN[2.3]]AGC/ACC]
SCF_TRCP1_1	SETBP1	loop	DSGIGT	GATAGCGGCATTGGCACC
B-TRCP	Vpu	loop	DSGxDS	GATAGCGGCNNNNNAGC
B-TRCP	RE1-silencing transcription factor	loop	SEGSDDSGL	AGCGAAGGCAGCGATGATAGCGGCTG
B-TRCP	Prolactin receptor	loop	TDGGRGS	ACCGATAGCGGCGCGGCGCAGC
B-TRCP	Protein aurora borealis	loop	DSGYNT	GATAGCGGCTATAACACC
B-TRCP	Vaccinia virus	loop	YSGHLEps	TATAGCGGCAACCTGGAAAGC
SCF ^{flaw2}	G1/S-specific cyclin-D3	loop	psQTSPTDVTAIHL	AGCCAGACCAACACCCGACCGGATGTGACCGGATTGATCTG
SCF ^{flaw3}	G1/S-specific cyclin-D3	loop	PTDVTAL	CCGACCGATGTGACCGCGATT
OPDH VHL 1	consensus	loop	[UL]apt.[6.8][FLVM].[FLVM]	[ATT/CTG][GCGACC][NNN[6.8]]TTT/CTG/ATT/CTG/ATG/ATG/NNN[TTT/CTG/ATT/CTG/ATG]
SCF coil	consensus	loop	[RK][RK].[SL.FLUM].[RK][R][HRK].[RK]	[CGT/AAA][CGT/AAA][NNNAGCCTGNNNTTT][TTT/CTA/ATG/NNN[CGT/AAA][CAT/NNN[CGT/AAA]
SOCS box-Cul5-SP584	VASA	loop	DINNINNIVEDVERKREFYI	GACATCAACACACACACACATCGTTGAAGACGTTGAACGTTAAACGTGAAATCTACATC
UBR5	PAM2	loop	SKLSVNAPEFYP5G	TCTAAACTGTCTGTTAAACGCGCGGAAATCTACCGGTCTGCT
CRL2(KLHDC2)	Usp1	C-terminus	IGLLGG	ATCGGCTGCTGGGTGGT
GID4	Pro/N-degron	N-terminus	PGLW	CCGGGTCTGTGG
TRIM21	Fc fragment	loop	WXW	TGGNNNTGG
TRIM21	Fc fragment	loop	HNH	CATAACCAT
Nedd4	PPxY motif	loop	TAPPVATLG	ACCAGCGCGCGCGCGGCTATGGACCCCTGGGC

Elongin C	Vif	loop	SLxxLxxd	AGCTGNNNNNNNNCTGNNNNNNNNNNATT
Unknown	ID2	loop	SRTPLTLN	AGCGCACCCGCGCTGACCAACCTGAAC
Unknown	ZAP70	loop	DGYTPEP	GATGGCTATACCCGGAACCG
Unknown	SH3R1	loop	RPTAAVTPH	CGCCGACCGCGCGCGTGAACCCGATT
Unknown	ETV1	loop	DEQVFD	GATGAACAGTTTGTGCCGGAAT
Unknown	EPA51	loop	LAPYPMGDGFOL	CTGGCGCGGTATATTCGATGGATGGGAAGATTTTCAGCTG
Unknown	hantavirus	loop	YVGLVW/GVLLTTELIVWAASA	TATGTGGCTGTGTGTGGCGCTGCTGTGACCAACGAACTGATTTGTGTGGCGCGGAGCGCG
CRL4_CDT2_1	SETD8	loop	PKTPSSCDSTN	CCGAAAACCCCGCGAGCAGCTCGATAGCACCAAC
CBL (PTK)	Consensus	loop	[DN].pV[ST].P	[GAT/AAC]NNNTAT[AGC/ACC]NNNCCG
CBL (met)	Consensus	loop	DpYR	GATTATCGT
CBL	SH2B adapter protein 3	loop	RAIDNQVTL	CGCGGATTGATAACCAAGTATACCCCGCTG
CBL	Protein sprouty homolog 1	loop	IRGSNEYTEGPS	ATTGCGCGCAGCAACGAATATACGAAGGCCCGAGC
CBL	Protein sprouty homolog 2	loop	IRNTNEYTEGPT	ATTGCGCAACCAACGAATATACGAAGGCCCGAGC
CBL	Protein sprouty homolog 4	loop	HVENVDHNP	CATGTGAAAACGATTATTGATAACCCGAGC
CBL	Tyrosine-protein kinase SYK	loop	SNPVEPELA	AGCTTTAACCCGTATGAACGGGAACCTGCGC
CBL	Tyrosine-protein kinase ZAP-70	loop	TUNSDGPVPEPA	ACCGTGAACAGCGATGGCTATACCCCGGAACCGCG
CBL	Plexin-A3	loop	IPFLDVRTYAV	ATTCCGTTTCTGGATTATGCAACCTATGCGGTG
CBL	Plexin-A1	loop	IPFLDVRTYAM	ATTCCGTTTCTGGATTATGCAACCTATGCGGTG
CBL	Platelet-derived growth factor receptor alpha	loop	SIFDNLVTLSD	AGCATTTTAAACAGCCCTGTATACCAACCTGAGCGAT
CBL	Platelet-derived growth factor receptor beta	loop	SIFNSLYTILSD	AGCATTTTAAACAGCCCTGTATACCAACCTGAGCGAT
CBL	Tumor necrosis factor receptor superfamily member 15	loop	KGDGGVYSLPP	AAAGCGATGGCGGCTGTATAGCAACCTGCGCGCG
CRL4(COP1/OET)	Trib1	loop	SDQIVPEY	TCTGACCAAGATCGTTCCGGAATAC
SH3RF1	E3 ubiquitin-protein ligase SH3RF1	loop	RPTAAVTPH	CGCCGACCGCGCGGTGACCCGATT
COP-1	Consensus	loop	[D,E][D,E][2,3]VP[DE]	[GAA/GAC][GAA/GAC][NNNNNN/NNNNNNNN][GTGCG][GAA/GAC]
COP-1	Tribbles homolog 1	loop	SDQIVPEY	AGCGATCAGATTGTGCCGGAATAT
SHAH	Consensus	loop	P.A.V.P[VP]	NNNCCGNNNGGNNNGTNNNCCG[NNN EXCEPT CCN]

SIAH	AF4/FMR2 family member 4	loop (beta strand)	KPTAYVRPM	AAACCGACCGCGTATGTGCGCCGCGATG
SIAH	calyculin-binding protein	loop	KPAAVVAPI	AAACCGCGCGCGGTGGTGGCGCGGATT
SIAH	POU domain class 2-associating factor 1	loop	APTAVVLPH	GCGCCGACCGCGGTGTGTGCTGCCGCAT
SIAH	Retrotransposon-derived protein PEG10	loop	PPRALVLPH	CCGCCGCGCGCGCTGTGTGCTGCCGCAT
ERAD-C	CL1	amphipathic helix extension	ACKNWFSLSHFVHL	GCCTGCAAAAACCTGGTTTAGCAGCGCTGAGCCATTTTGTGATTCATCTG
UBR	Nend_Nbox 2	N-terminal extension	⁴ M[0.1][FLYIW][⁴ V]	[ATG[0.1]][TTT/CTG/TAT/TGG/ATT][NNN ⁴ CCG]
UBR	Nend_UBRbox 1	N-terminal extension	⁴ M[0.1][RK][⁴ V]	[ATG[0.1]][AAA/CGT][NNN ⁴ CCG]
UBR	Nend_UBRbox 2	N-terminal extension	⁴ M[0.1][ED]	[ATG[0.1]][GAT/GAA]
UBR	Nend_UBRbox 3	N-terminal extension	⁴ M[0.1][NQ]	[ATG[0.1]][CAG/AAC]
UBR	Nend_UBRbox 4	N-terminal extension	⁴ M[0.1][C]	[ATG[0.1]][TGC]

Other degradation pathways: pathway; protein to which the degron binds	Degrone sequence derived from	Grafting site in scaffold	Amino acid sequence	DNA sequence <i>E. coli</i> codon optimised 5' to 3'
ESRCT; ALIX	Consensus, e.g. HIV Gag	loop	LVP...L, e.g. ELYPLTSLRS	GAACTGTACCGCGCTGACCTCTCTGCGTTCT
ESRCT; ALIX	SIV(mac239) Gag	loop insertion	SREKPYKEVTEDLLHLSLF	AGCGCGAAAAAACCGTATAAAGAGTGAOCGGAAGATCTGCTGCATCTGAACAGCCTGTTT
ESRCT; ALIX	SIV(jagmTan-1) Gag	loop insertion	AAGAYDPARKLEQY	GCGCCGCGCGGTATGATCCGCGCGCAAACTGCTGGAACAGTAT
ESRCT; AP-1	Nef	loop insertion	EKKLL	GAANNNNNNNNNCTGCTG
ESRCT; AP-2	Env	loop insertion	YKKL	TATNNNNNNCTG
ESRCT; AP-1	HIPR1	loop insertion	MDFGSLSLKKQ	ATGGATTTTAGCGGCTGAGCCTGATTAAACTGAAAAAAGAG

ESCRT ₁ , AP-2	consensus	loop insertion	D(E)xxL(L)(I)	GAT(GAC)NNNNNNNNCTGCTG(ATT)
ESCRT ₁ , AP	viral adaptor	loop	SREKPYKEVTEDILHLNSLF	AGCCGCGAAAAACCGTATATAAAGAAGTGAACGGAAGATCTGCTGCATCTGAACAGCCTGTTT
ESCRT ₁ , AP	viral adaptor	loop	AAGAYDPARKLLEQYAKK	GCGGCGGGCGGTATGATCCGGCGGCAAACTGCTGGAACAGTATGCGAAAAAAA
CMA; Hsc70	repeat consensus	loop	KFERQKILDRFFE	AAATTTGAACGCCACAGAGAAAATTCTGGATCAGCGCTTTTITGAA
Autophagy; LC3 / Atg8 family	Consensus LIR (LC3- interacting) / AIM (Atg8 family-interacting) motif			
Autophagy; FIP200	CCPG1	loop	[W/F/V]...[L/V]	(TGG/TTC/TAT)NNNNNN(CTG/ATC/GTG)
Autophagy; LC3	DVL	loop	TASDDSDVTLPPPK	ADCGCTCTGACGACTCTGACATCGTTACCCCTGGAACCGCGAAA
GABARAP	Ankyrin G	loop insertion	EVRDRMILKITI	GAAAGTTGCTGACCGTATGTGGCTGAAAAATCACCATC
		loop insertion	PEDDWIEFSSEIREARQOAAASQSPS	CCGGAAGATGATTGGATTGAAATTTAGCAGCGAAGAAATTCGGAAGCGGCCAGCGCGGCGAGCCAGAGCCCGGAGC

SEQ ID NOS 470 to 617 SEQ ID NOS 618 to 765

- 5 Key
- .

[X]

p

\$

^x

X{x, y}

[vx]

NNN

[NNN/NNN/NNN]

[VNNN]

[NNN{x,y}]
- Any amino acid

Allowed amino acid at the position

Phosphorylated amino acid

C terminal of chain

N-terminal of chain

where x & y are the minimum, maximum of X amino acids at the position

Amino acid not allowed at the position

Any codon

Any one of these codons at the position

Any codon except this

codon, where x & y are the maximum & minimum of codons
- 10
- 15

Table 9

1. Axin-RTPR-ABBA

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPGGSLSSAFHVFEDGNKENGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN
(SEQ ID NO: 766)

2. Axin-RTPR-DBOX

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPGGPRLPLGDVSNNGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID
NO: 767)

3. Axin-RTPR-KEN

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPGGPRLPLGDVSNNGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID
NO: 768)

4. Axin-RTPR-Nrf2

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPGGPRLPLGDVSNNGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID
NO: 769)

5. Axin-RTPR-SIAH

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPGGLRPVAMVRPTVGGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID
NO: 770)

6. Axin-RTPR-SPOP

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGGLACDEVSTTSSSTA
GGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO: 771)

7. Axin-RTPR-p27

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPNNAGSNEQEPKKRSPDAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID
NO: 772)

8. Axin-RTPR-p53

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPNNFAAYWNLLSAYG (SEQ ID NO: 773)

10. Bcl9-RTPR-ABBA

MGSSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYY
RQGDYQRAIEYYQRALELDPGGDPETGELGGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID
NO: 774)

11. Bcl9-RTPR-DBOX-v1

MGSSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYY
RQGDYQRAIEYYQRALELDPGGPRLPLGDVSNNGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ
ID NO: 775)

12. Bcl9-RTPR-DBOX-v2

MGSSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGPRLPLGDVSNNG
GPNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO: 776)

13. Bcl9-RTPR-KEN

MSGAYPEYILD IHVYRVQLELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPGGSLSSAFHVFEDGNKENGPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (S
EQ ID NO: 777)

14. Bcl9-RTPR-Nrf2

MGSSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYY
RQGDYQRAIEYYQRALELDPGGDPETGELGGPNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID
NO: 778)

5 **15. Bcl19-RTPR-p27**

MGSSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYY
RQGDYQRAIEYYQRALELDPNNAAGSNEQEPKKRSPDAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ
ID NO: 779)

10 **16. Bcl19-RTPR-p53**

MGSSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYY
RQGDYQRAIEYYQRALELDPNNFAAYWNLLSAYG (SEQ ID NO: 780)

17. Bcl19-RTPR-SIAH

15 MGSSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYY
RQGDYQRAIEYYQRALELDPGGLRPVAMVRPTVGGPNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ
ID NO: 781)

18. Bcl19-RTPR-SPOP

20 MGSSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGLACDEVSTSTSS
STAGGPNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO: 782)

19. TCF7L2-RTPR-Nrf2

25 MGSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGQELGDNDLMHFSYESTQDGGPNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGDPETGELGGPNAEAWYNLGNAY
YRQGDYQRAIEYYQRALELDPNN (SEQ ID NO: 783)

20. TCF7L2-RTPR-p27

30 MGSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGQELGDNDLMHFSYESTQDGGPNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAAGSNEQEPKKRSPDAEAWYNLGN
NAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO: 784)

21. p27-RTPR-TCF7L2

35 MRGSHHHHHHGLVPRGSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAAGSNEQEPKKRSPDAEAWYNLGN
YYRQGDYQRAIEYYQRALELDPNSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGQELGDNDLMHFSYEST
QDGGPNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNS (SEQ ID NO: 785)

22. TCF7L2-RTPR-p53

40 MGSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGQELGDNDLMHFSYESTQDGGPNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNFAAYWNLLSAYG (SEQ ID
NO: 786)

23. ICAT-RTPR-p27

45 MGSYAYQRAIVEYMLRLMSDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQ
RAIEYYQRALELDPNNAAGSNEQEPKKRSPDAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO:
787)

24. ICAT-RTPR-p53

50 MGSYAYQRAIVEYMLRLMSDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQ
RAIEYYQRALELDPNNFAAYWNLLSAYG (SEQ ID NO: 788)

25. LRH1-RTPR-ABBA

55 MGSYEQAIAYLDALMCDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPGGSEDKENVPPGGPNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO:
789)

26. LRH1-RTPR-DBOX

60 MGSYEQAIAYLDALMCDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPGGPRLPLGDIVSNNGGPNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO:
790)

27. **LRH1-RTPR-KEN**
MGSYEQAIAAYLDALMCDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPGGSEDKENVPPGGPNAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO:
791)
- 5
28. **LRH1-RTPR-Nrf2**
MGSYEQAIAAYLDALMCDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPGGDPETGELGGPNAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO: 792)
- 10
29. **LRH1-RTPR-p27**
MGSYEQAIAAYLDALMCDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPNNAAGSNEQEPKKRSPDAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO:
793)
- 15
30. **LRH1-RTPR-p53**
MGSYEQAIAAYLDALMCDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPNNFAAYWNLLSAYG (SEQ ID NO: 794)
- 20
31. **LRH1-RTPR-SIAH**
MGSYEQAIAAYLDALMCDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPGGLRPVAMVRPTVGGPNAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO:
795)
- 25
32. **LRH1-RTPR-SPOP**
MGSYEQAIAAYLDALMCDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPGGLACDEVSTTTSSSTAGGPNAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID
NO: 796)
- 30
33. **APC-RTPR-Nrf2**
MGSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGSCSEEELEALEALELDEGGPNAAEAWYNLGNAYYRQGDYQ
RAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGDPETGELGGPNAAEAWYNLGNAYYRQ
GDYQRAIEYYQRALELDPNN (SEQ ID NO: 797)
- 35
34. **APC-RTPR-p27**
MGSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGSCSEEELEALEALELDEGGPNAAEAWYNLGNAYYRQGDYQ
RAIEYYQRALELDPNN (SEQ ID NO: 798)
- 40
35. **p27-RTPR-APC**
MRGSHHHHHHGLVPRGSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAAGSNEQEPKKRSPDAEAWYNLGN
AYYRQGDYQRAIEYYQRALELDPRAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGSCSEEELEALEALELDEG
GPNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPRA (SEQ ID NO: 799)
- 45
36. **APC-RTPR-p53**
MGSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGQELGDNDLMHFSYESTQDGGPNAAEAWYNLGNAYYRQ
DYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNFAAYWNLLSAYG (SEQ ID
NO: 800)
- 50
37. **1TBP-CTPR2**
MGSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYYQK
ALELDPRA (SEQ ID NO: 801)
- 55
38. **2TBP-CTPR4**
MGSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYYQK
ALELDPRAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKA
IEYYQKALELDPRA (SEQ ID NO: 802)
- 60
39. **3TBP-CTPR6**
MGSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYYQK
ALELDPRAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKA
IEYYQKALELDPRAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGD

YQKAIEYYQKALELDPRS (SEQ ID NO: 803)

40. 4TBP-CTPR8

MGSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYYQK
ALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAI
EYYQKALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGD
YQKAIEYYQKALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAY
YKQGDYQKAIEYYQKALELDPRS (SEQ ID NO: 804)

41. 1TBP-CTPR2-Foldon (Foldon sequence in bold)

MGSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYYQK
ALELDPRS**AKASLNLANADIKTIQEAGYIPEAPRDGQAYVRKDGEWVLLSTFLRS** (SEQ ID NO: 805)

42. 2TBP-CTPR4-Foldon (Foldon sequence in bold)

MGSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYYQK
ALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAI
EYYQKALELDPRS**AKASLNLANADIKTIQEAGYIPEAPRDGQAYVRKDGEWVLLSTFLRS** (SEQ ID NO:
806)

43. 3TBP-CTPR6-Foldon (Foldon sequence in bold)

MGSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYYQK
ALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAI
EYYQKALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGD
YQKAIEYYQKALELDPRS**AKASLNLANADIKTIQEAGYIPEAPRDGQAYVRKDGEWVLLSTFLRS** (SEQ ID
NO: 807)

44. 4TBP-CTPR8-Foldon (Foldon sequence in bold)

MGSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYYQK
ALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAI
EYYQKALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGD
YQKAIEYYQKALELDPRSAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAY
YKQGDYQKAIEYYQKALELDPRS**AKASLNLANADIKTIQEAGYIPEAPRDGQAYVRKDGEWVLLSTFLRS** (SEQ
ID NO: 808)

45. KBL -RTPR-CMA_Q

MGSIPNPLLGLDGSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNPLYISYDPAEAWYNLGNAYYRQGDYQR
AIEYYQRALELDPRSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNQRFFE (SEQ ID NO: 809)

46. CMA_Q-KBL-RTPR

MGSQRFFEGSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNPLYISYDPAEAWYNLGNAYYRQGDYQRAIEY
YQRALELDPRSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO: 810)

47. CMA_K-KBL-RTPR

MGSKFERQGSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNPLYISYDPAEAWYNLGNAYYRQGDYQRAIEY
YQRALELDPRSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNN (SEQ ID NO: 811)

48. SOS-RTPR-CMA_K

MGSFEGIALTNYLKALEGDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQR
AIEYYQRALELDPRSKFERQ (SEQ ID NO: 812)

49. SOS-RTPR-CMA_Q

MGSFEGIALTNYLKALEGDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPRSIPNPLLGLDKFERQ (SEQ
ID NO: 813)

50. SOS-RTPR-p27

MGSFEGIALTNYLKALEGDPNNAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYRQGDYQR
AIEYYQRALELDPRSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAAGSNEQEPKKRSPDAEAWYNLGNAY
YRQGDYQRAIEYYQRALELDPRS (SEQ ID NO: 814)

51. KBL-RTPR-p27

MGSAAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNPLYISYDPAEAWYNLGNAYYRQGDYQRAIEYYQRALEL

DPRSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNAGSNEQEPKKRSAEAWYNLGNAYYRQGDYQRAIEYYQ
RALELDPNN (SEQ ID NO: 815)

Table 10

	S75991	ALTNNIGTI	YYAREDDYDQA	LNYYEQALSL	SRAV (SEQ ID NO: 816)
5	2407639-54	AEELKEQANE	YFRVKDYDHA	VQYYTQAI DL	SPDT (SEQ ID NO: 817)
	YDAB_MY	WRAWFRLGLA	YDGAGDRRRA	REAIRRAITL	EKNP (SEQ ID NO: 818)
	GSIA_BA	SSALYNLGNC	YEKMGELQKA	AEYFGKSVSI	CKSE (SEQ ID NO: 819)
	E64417	VPMWVKKAEI	LRKLGRYEDA	LLCLNRALEL	KPHD (SEQ ID NO: 820)
	A56519	EPLNNLGHV	CRKLKKYAEA	LDYHRQALVL	IPQN (SEQ ID NO: 821)
10	DMUNKNOWN_1-96	EKVYFNLGML	AMDESSFDEA	EQFFKRAIHL	KADF (SEQ ID NO: 822)
	SPAC6B12_12-516	SEVYNYFGEI	LLDQQKFDDA	VKNFDHAI EL	EKRE (SEQ ID NO: 823)
	NUC2_SC	PESWCILANC	FSLQREHSQA	LKCINRAIQL	DPTF (SEQ ID NO: 824)
	CELZK320	ARICNLQAEI	LYRRREFSQA	VDICKQALAY	HETD (SEQ ID NO: 825)
	NFRA_EC	ANAYVARATI	YRQRHNPAA	VSDLRAALEL	EPNN (SEQ ID NO: 826)
15	CCU2886	GLAWHILAIA	REKTGDFASS	LRAYEAALAL	LPDH (SEQ ID NO: 827)
	S766850	FELHYHLANA	YRROQKWELA	RKHYQKALDE	EILL (SEQ ID NO: 828)
	CELT19A	VAIHFTLGNI	YASIGDYQRA	LNFYYSTLSL	QANF (SEQ ID NO: 829)
	S766851	AEAYQNLAVT	SFKAGLIQES	VDAFQQAIAL	YEQR (SEQ ID NO: 830)
	S75601	YKGWYQQGNT	LWQQGKYNDA	LLSYERALEY	YPGD (SEQ ID NO: 831)
20	YHBM_EC	PEVFNLYGIY	LTQAGNFDA	YEA FDSVLEL	DPTY (SEQ ID NO: 832)
	HSU4203	EKGLYRRGEA	QLLMNEFESA	KGDFEKVLEV	NPQN (SEQ ID NO: 833)
	YKD1_CA	ADLRVGIGHC	FAKMGMMDKA	KTA FERAMEI	EPYN (SEQ ID NO: 834)
	CELZK32	PGSYSLLGDA	FMKVQEPEDA	INFYEQALKM	QSKD (SEQ ID NO: 835)
	S56658	SKGYTRKGAV	QFSMKEYDKA	LETYREGLKH	DPNN (SEQ ID NO: 836)
25	G02540	AQSCYSLGNT	YTLLQDYEKA	IDYHLKHLAI	AQEL (SEQ ID NO: 837)
	CELC55B	ADILCERAE	HILDEDYDSA	IEDYQKATEV	NPDH (SEQ ID NO: 838)
	CELF30H	PVLYRNRAMA	RLKRDDFEGA	QSDCTKALEF	DGAD (SEQ ID NO: 839)
	B55508	VESTSLGLV	YKLKGQEKNA	LFYYEKALRL	TGEM (SEQ ID NO: 840)
	SMU5458	TDCLYHLAVG	FLKLNDYHNA	TIYCQCLLTI	EPDN (SEQ ID NO: 841)
30	BSTHRZ1	GRTLYNIGLC	KNSQSQYEDA	IPYFKRAIAV	FEES (SEQ ID NO: 842)
	PSEPILF	SRVFENLGLV	SLQMKKPAQA	KEYFEKSLRL	NRNQ (SEQ ID NO: 843)
	CELF38B0	AVAWMNLGIS	QMNLLKKYYEA	EKSLKNLLI	RPNS (SEQ ID NO: 844)
	JT06030	PETHYNRGLA	WERLGNVDQA	IADYGRSIAL	DRYY (SEQ ID NO: 845)
	JT06031	YKAYYNRANS	YFQLGQYAQA	IADYNRVLVL	RPDY (SEQ ID NO: 846)
35	S76422	AEAYYQRLS	HYRLEDWAEA	VRDCTEAI RL	RGDL (SEQ ID NO: 847)
	CC23_YE0	RRIWQVLGEC	YSKTGNKVEA	IKCYKRSIKA	SQTV (SEQ ID NO: 848)
	CELF38B1	AHCLFNLGV	L YQRTNRDEMA	MSAWKNATRI	DPSH (SEQ ID NO: 849)
	CC23_YE1	PETCCIIANY	YSARQEHEKS	IMYFRRALTL	DKKT (SEQ ID NO: 850)
	CELF38B2	EQALNNLGNL	LEKSGDSKTA	ESLLARAVTL	RPSF (SEQ ID NO: 851)
40	MEU8731	FEAWAALGKL	YLAQDDKGRA	LDAFRRAEAL	YPQW (SEQ ID NO: 852)
	CELF38B3	IQVLLGVAKC	WMIQKELDKA	LESIEKAQEL	AEGL (SEQ ID NO: 853)
	S75615	AEAYKARGKI	YWQLGEQAQA	IADLEAALAL	FSDQ (SEQ ID NO: 854)
	CELF38B4	VLFHANLGIL	YQRM SRHKEA	ESQYRIVLAL	DSKN (SEQ ID NO: 855)
	CC27_YE0	YNAYYGLGTS	AMKLGQYEEA	LLYFEKARSI	NPVN (SEQ ID NO: 856)
45	S74806	AEVWLSRGLL	LEAMERKEEA	IPSYEKALTL	EPTL (SEQ ID NO: 857)

	CELF38B5	SRVHMQIGSC	HAKHSNFTAA	ENHIKSAIDL	NPTS (SEQ ID NO: 858)
	CC27_YE1	AYAYTLQGHE	HSSNDSSDSA	KTCYRKALAC	DPQH (SEQ ID NO: 859)
	CELC56C	VEALRQKGNE	LFVQKDYKEA	IDAYRDALTR	LDTL (SEQ ID NO: 860)
	F643990	IDLILKVAFT	YFKLKKYKHA	LKYFEKALKL	NPNV (SEQ ID NO: 861)
5	F643991	WKLWKNLGDK	AYLWKAYYEA	LFCYNKALEL	NQNT (SEQ ID NO: 862)
	MTCL_CO	AEFYFRLGES	YFAHHQYTFA	VSYLEQAIDL	FENN (SEQ ID NO: 863)
	MM284231_1-255	TEAFYKISTL	YYQLGDHEL	LSEVRECLKL	DQDH (SEQ ID NO: 864)
	F643992	TELLCKKGYA	LLKLYKRDLA	IKYFEKASEK	DRNN (SEQ ID NO: 865)
	KNLC_CA	AKQLNNLALL	CQNQKYEYEV	EKYKRALEI	YESK (SEQ ID NO: 866)
10	ATAF0	AQCFMHRASA	YRSAGRIAES	IADCNKTLAL	DPSC (SEQ ID NO: 867)
	S585440	VKAFYRRALA	HKGLKNYQKS	LIDLNKVILL	DPSI (SEQ ID NO: 868)
	S585441	LSIIFNRAAC	YLKEGNCSCG	IQDCNRALEL	HPFS (SEQ ID NO: 869)
	A645120	AVILSNIACS	LYSRGLTNKV	IEVYNKAIKI	AEES (SEQ ID NO: 870)
	A645121	CEAFLNLAKA	YERLADFTKA	LQYGKASLEH	PSMD (SEQ ID NO: 871)
15	AB001	AMSHMNIGIC	YDELKEYKKA	SQHLILALEI	FEKS (SEQ ID NO: 872)
	I49564	PSALTNGKNT	VFANGDYEKA	AEFYKEALRN	DSSC (SEQ ID NO: 873)
	OM70_NE	PDIYYHRAQL	HFIKGEFAEA	AKDYQKSIDL	DSDF (SEQ ID NO: 874)
	ATU6213	AEAYCNMGVI	YKNRGDLEMA	ITCYERCLAV	SPNF (SEQ ID NO: 875)
	S75709	PAVWSNRGNS	RVSQNKLEDA	IADFNQAIEL	APEQ (SEQ ID NO: 876)
20	CELF32A	GRCYGS LGNT	YYCLGDYDQS	IHFHKLRLLEL	SQQY (SEQ ID NO: 877)
	CELT09B0	AEQHNTNGKK	CYMNKRYDDA	VDHYSKAIKV	NPLP (SEQ ID NO: 878)
	CELT09B1	VKPLYFLGNV	FLQSKKYSEA	ISCLSKALYH	NAVI (SEQ ID NO: 879)
	YCF3_OD	SYTLNIGLI	YGNNGNYSQA	LEYYHQALEL	NSNL (SEQ ID NO: 880)
	YHR7_YE	PPTYHHRGQM	YFILQDYKNA	KEDFQKAQSL	NPEN (SEQ ID NO: 881)
25	SSN6_YE	PIFWCSIGVL	YYQISQYRDA	LDAYTRAIRL	NPYI (SEQ ID NO: 882)
	FLBA_CAUCR-226	AVLWNTLGTV	LCNIGDAAGS	IVFFDESLRL	APDF (SEQ ID NO: 883)
	BIMA_EM	YNAWYGLGTV	YDKMGKL DFA	EQHFRNAAKI	NPSN (SEQ ID NO: 884)
	CUT9_SC0	ASMCYLRGQV	YTNLSNF DRA	KECYKEALMV	DAKC (SEQ ID NO: 885)
	CC27_HU	YNAWYGLGMI	YYKQEKFS LA	EMHFQKALDI	NPQS (SEQ ID NO: 886)
30	MMUTY0	AAFLYGLGLV	YFYNAFQWA	IRAFQEVLYV	DPNF (SEQ ID NO: 887)
	KNLC_ST	AKTKNNLAAA	YLKQKGYKAA	ETLYKQVLTR	AHER (SEQ ID NO: 888)
	S757090	AFAYNNRGNA	EGGLGNW TSA	LED FQQATAI	APNF (SEQ ID NO: 889)
	C48583	VKAYLRRGIA	YEGMEKWKLA	LEDYTKAQSI	SPGV (SEQ ID NO: 890)
	S757091	TDPYLNRGTA	LEAKGEFKAA	IADYNRVLAV	NPED (SEQ ID NO: 891)
35	S759910	GTSLSNIGLA	YNNLGEREKA	LEYYQEALTM	SRSV (SEQ ID NO: 892)
	S566581	ISYLTNRAAV	YLEMGKFEDC	IKDCEKAVER	GKEL (SEQ ID NO: 893)
	S759911	GAILSNIQYV	YDGLGELNQA	MDYYQQALVL	RREI (SEQ ID NO: 894)
	S566582	ADEAREKGNE	LFKQQKYPEA	TKHYTEAIKR	NPKD (SEQ ID NO: 895)
	LMU73840	VKGYARKGHA	YFWTKQYNRA	LQAYDEGLKV	DPSN (SEQ ID NO: 896)
40	S759912	VFALNRLGTI	YSDFGEKSQA	LAYYEQAIP L	AQQL (SEQ ID NO: 897)
	S566583	AQKEKEAGNA	AYKKKDFETA	IGHYSKALEL	DDED (SEQ ID NO: 898)
	LMU73841	HTSYSNRAAA	YIKLGAFNDA	LKDAEKCIEL	KPDF (SEQ ID NO: 899)
	S759913	GIRLNNIALL	YDSVGEKDEA	LTTYKEALKI	SQET (SEQ ID NO: 900)
	LMU73842	ATELKNKGNE	EFSAAGRYVEA	VNYFSKAIQL	DEQN (SEQ ID NO: 901)
45	S759914	MTTINN LGVA	YDNLGESEKA	LQYYAEALAL	GKSL (SEQ ID NO: 902)

	TPRD_HU	GELMKMKGNE	EFSKERFDIA	IIYYTRAIEY	RPEN (SEQ ID NO: 903)
	LMU73843	ALALKEEGNK	LYLSKKFEEA	LTKYQEAQVK	DPNN (SEQ ID NO: 904)
	S759915	GSVISNIALV	LDGLGQKAQA	LEYYQQAEL	ARLV (SEQ ID NO: 905)
	S60905	IEYLYEKAQL	YRRLKDKENE	LKYYNLALSI	DENK (SEQ ID NO: 906)
5	YHBM_HA	ASVYNYLGLY	LLLEEDYDGA	LDAFNTVFEL	DSGY (SEQ ID NO: 907)
	S759916	GATLNNIGSI	YNALADRKKA	IDFYQQAALV	IRQA (SEQ ID NO: 908)
	LMU73844	AKQKKDEGNQ	YFKEDKFPEA	VAAYTEAIKR	NPAA (SEQ ID NO: 909)
	S75633	ASSLNNLGYL	AFLQGDLPTA	LDYYQRALIL	RQTT (SEQ ID NO: 910)
	CET12D8	AMLHAKRANV	LLKLKRPVAA	IADCDKAISI	NPDS (SEQ ID NO: 911)
10	S58544	CAIYTNRALC	YKLKCFEEA	KQDCDQALQL	ADGN (SEQ ID NO: 912)
	C485830	AVYYTNRAAC	HQQTHMYSLM	VDDCNAAIEI	DPAN (SEQ ID NO: 913)
	BSZ9404	WSAYNNLALA	YFYSGNVVKA	KQTAYEVLSH	NEGN (SEQ ID NO: 914)
	Y366_HA	AKARVELALS	YLQQNNPQLA	KINLDKALQH	DKNY (SEQ ID NO: 915)
	C485831	PEEAKQLGNS	FFKDGKYDQA	AEFYTRAIEL	QTEP (SEQ ID NO: 916)
15	NPRA_BAO	AECHILLGIC	YRRYGEVDQA	IECYSLAHKI	AQII (SEQ ID NO: 917)
	YC37_PO	AKLYNMLGFI	YYKADQKKLA	KNFYERAIEV	DGNY (SEQ ID NO: 918)
	N358_HU	PKAHRFLGLL	YELEENTDKA	VECYRRSVEL	NPTQ (SEQ ID NO: 919)
	S42210	TDVLRSAARF	YYKTHDKDRA	IQLLSQALEL	LPNN (SEQ ID NO: 920)
	CC27_YE	PETWCCIGNL	LSLQKDHDA	IKAFEKATQL	DPNF (SEQ ID NO: 921)
20	HIBN_XEO	AQAHQKLGEV	CIESENYSQA	VEDFLACLNI	QKEH (SEQ ID NO: 922)
	STI1_YEO	ARGYSNRAAA	LAKLMSFPEA	IADCNKAIEK	DPNF (SEQ ID NO: 923)
	STI1_YE1	ADEYKQQGNA	AFTAKDYDKA	IELFTKAIEV	SETP (SEQ ID NO: 924)
	CELF32A0	AQMAYSLANA	LYIGKEVQKA	ITYFQRHLKI	ARSL (SEQ ID NO: 925)
	STI1_YE2	SKSFARIGNA	YHKLGLDKKT	IEYYQKSLTE	HRTA (SEQ ID NO: 926)
25	LACALS	ISIIYQIRIGDS	YAQLGNFENA	ISFLEKSLEF	DEKP (SEQ ID NO: 927)
	CELR05F	SKAWGRMGLA	YSCQNRYEHA	AEAYKKALEL	EPNQ (SEQ ID NO: 928)
	G025400	AKASGNLGNT	LKVLGNFDEA	IVCCQRHLDI	SREL (SEQ ID NO: 929)
	G025401	RRAYSNLGNA	YIFLGEFETA	SEYYKKTLLL	ARQL (SEQ ID NO: 930)
	H64467	PLLYLYKGII	LNKLKGKYNEA	IKYFDKVLEI	NPNI (SEQ ID NO: 931)
30	G025402	SAIYSQLGNA	YFYLHDYAKA	LEYHHHDLTL	ARTI (SEQ ID NO: 932)
	G025403	GRACWSLGNA	YTALGNHDQA	MHFAEKHLEI	SREV (SEQ ID NO: 933)
	PTSR_HU0	YLLWNKLGAT	LANGNQSEEA	VAAYRRALEL	QPGY (SEQ ID NO: 934)
	G025404	GRAFGNLGNT	HYLLGNFRDA	VIAHEQRLLI	AKEF (SEQ ID NO: 935)
	PTSR_HU1	IRSRYNLGIS	CINLGAHREA	VEHFLEALNM	QRKS (SEQ ID NO: 936)
35	G025405	TKQQIEKGLQ	LYQANETGKA	LEIWQQVVER	STEL (SEQ ID NO: 937)
	YCA1_PL0	PKGYYIRKGA	EHGLRQLSNA	EKTYLEGLKI	DPNN (SEQ ID NO: 938)
	PQ01800	PETWLNRLGLA	YYRQGSQAA	IADFDQLLQQ	SPTD (SEQ ID NO: 939)
	CEC34C61	VRARYNLGIS	CMQLSSYDEA	LKHFLSALEL	QKGG (SEQ ID NO: 940)
	CEC34C62	PDLQNALGV	YNLNRNFARA	VDSLKLAISK	NPTD (SEQ ID NO: 941)
40	SSD900_122-109	ADAYYNRGYA	KHVLGQYQAA	ITDYNQAISL	NPEF (SEQ ID NO: 942)
	STI1_YE	SKGYNRLGAA	HLGLGDLDEA	ESNYKKALEL	DASN (SEQ ID NO: 943)
	YCV0_YE	PVGYSNKAMA	LIKLGLEYTQA	IQMCQQGLRY	TSTA (SEQ ID NO: 944)
	HSAB2370_1-631	FNCWESLGEA	YLSRGGYTTA	LKSFTKASEL	NPES (SEQ ID NO: 945)
	S75648	ADAYNLRGVA	YMVIEQYTD	LADFDQAIAL	NPKD (SEQ ID NO: 946)
45	CELT19A0	PLFHYVQGRM	KLLLDHVDKA	IQHLKDAMDK	DPNN (SEQ ID NO: 947)

	CELT25F0	FEGNYNLGLV	SFTQGKYHEC	RELIEKALAA	FPEH (SEQ ID NO: 948)
	CEF10B5	HPMLNNIGHI	ARRQGRLENEA	IMFYQKAIRM	EPKF (SEQ ID NO: 949)
	CELT25F1	VTMLTGMAV	QEALGEYDES	VKLYKRVLDA	ESNN (SEQ ID NO: 950)
	BSAB617_17-292	VQAVFSLTHI	YCKEGKYDKA	VEAYDRGIKS	AAEW (SEQ ID NO: 951)
5	CEM720	ATMLNVLAIV	YRNQENFKDA	AIYLEKALSI	RVQC (SEQ ID NO: 952)
	YCV0_YE0	FEKQKEQONS	LFKQGLYREA	VHCYDQLITA	QPQN (SEQ ID NO: 953)
	YCV0_YE1	AMSQYHMAVA	YRLGLRLGSA	MECCEESMKI	ALQH (SEQ ID NO: 954)
	CELC55B0	YQAIYRRATT	YLAMGRGKAA	IVDLERVLEL	KPDF (SEQ ID NO: 955)
	CELC55B1	YGARIQRGNI	LLKQGELEAA	EADFNIVLNH	DSSN (SEQ ID NO: 956)
10	HSU46570	AKLYCNRGT	NSKLRKLDDA	IEDCTNAVKL	DDTY (SEQ ID NO: 957)
	HSU46571	SILFSNRAAA	RMKQDKKEMA	INDCSKAIQL	NPSY (SEQ ID NO: 958)
	BSU5504	TKGLLNLGNC	FNKQDDVAKA	TVYYQRALEL	GEKI (SEQ ID NO: 959)
	HSU46572	ADALYVRGLC	LYYEDCIEKA	VQFFVQALRM	APDH (SEQ ID NO: 960)
	HSU46573	IRAILRRAEL	YEKTDKLDEA	LEDYKSILEK	DPSI (SEQ ID NO: 961)
15	HSU46574	IKAYLRRAQC	YMDTEQYEEA	VRDYEKVYQT	EKTK (SEQ ID NO: 962)
	LJGLN12	ASACNNLAEF	YRIRKEFDKA	EPLYLEAINI	LEES (SEQ ID NO: 963)
	HSU46575	LKAKKEDGNK	AFKEGNYKLA	YELYTEALGI	DPNN (SEQ ID NO: 964)
	D504537	GRAYYNLGLC	YYNQDLLDPA	IDYFEKAVST	FESS (SEQ ID NO: 965)
	MMUTY	ADTWCSIGVL	YQQQNQPMDA	LQAYICAVQL	DHGH (SEQ ID NO: 966)
20	CELD2020	SAAWTDLGEL	YEKNAQYQDA	LECCKNAMLN	NPVA (SEQ ID NO: 967)
	HSU46576	STRLKEEGNE	QFKKGDYIEA	ESSYSRALEM	CPSC (SEQ ID NO: 968)
	YB05_YE	ESLYANRAAC	ELELKNYRRC	IEDCSKALTI	NPKN (SEQ ID NO: 969)
	HUMFKBP0	VKCLNNLAAS	QLKLDHYRAA	LRSCSLVLEH	QPDN (SEQ ID NO: 970)
	HSU46577	ASYYGNAAT	LMMLGRFREA	LGDAQQSVRL	DDSF (SEQ ID NO: 971)
25	CEM72	AKQLTNLGIV	TQOLEKYEET	ENYFKQALSI	YNRA (SEQ ID NO: 972)
	S75578	SYILYNMALI	HASNGDHEKA	LGLYQEAIEL	NPKM (SEQ ID NO: 973)
	PSEPILF0	RDAYIQLGLG	YLQRGNTQEA	KVPLRKALEI	DPSS (SEQ ID NO: 974)
	S450640	AKSCANLGNI	FKMKGAYNDA	LTFTFKQLDF	AEEL (SEQ ID NO: 975)
	BBCDG5	ARFFNLIGLE	FFKLQGYGPA	IEYFAKNLEI	NPNN (SEQ ID NO: 976)
30	I40554	AKALLNLGDS	YKNMGDQDRA	LYFLIKAVDQ	AKES (SEQ ID NO: 977)
	ACU89981	SRAFHRKGNA	YMKMEKYAEA	IDSYNRALTE	HRNP (SEQ ID NO: 978)
	HSU20361	IKIMQNIQVT	FIQAGQYSDA	INSYEHIMSM	APNL (SEQ ID NO: 979)
	S668420	EDVQLNLAHC	YLEMREYGKA	IENYELVLKK	FDNE (SEQ ID NO: 980)
	HSU20362	GQSWYFLGRC	YSCIGKVQDA	FVSYRQSIDK	SEAS (SEQ ID NO: 981)
35	ACU89983	ALEEKNKGNA	AMSAGDFKAA	VEHYTNAIQH	DPQN (SEQ ID NO: 982)
	S748530	WPALNNIGLI	EYEQGKIDTA	LKRWQEVIKI	DSEQ (SEQ ID NO: 983)
	HSU20363	CRVCCSLGSF	YAQVKDYEKA	LEFPCKAAEL	VNNY (SEQ ID NO: 984)
	ACU89984	SKGYSRKGAA	LCYLGRYADA	KAAYAAGLEV	EPTN (SEQ ID NO: 985)
	S748531	AGIKFTLGNA	YFQKGQYDQA	VEVLLAGLAQ	R PDT (SEQ ID NO: 986)
40	ACU89985	SQQEKEKGND	CFRNAQYPDA	IKHYTEAIRR	NPTD (SEQ ID NO: 987)
	D63875	AEVRLGMGHC	FVKLNKLEKA	RLAFSRALEL	NSKC (SEQ ID NO: 988)
	ACU89986	MTYLTNLAHV	YMEQKNYEEC	VNTCTEAIEV	GRRV (SEQ ID NO: 989)
	YCOA_SY	TVALTALGLA	ARAINGYPEA	IAAYQQALQL	DPND (SEQ ID NO: 990)
	D644170	PVAYALLGQL	YELLGNFNDNA	LECYEKSLGI	EEKF (SEQ ID NO: 991)
45	YCIM_HA	VRASLLLANL	AMLDGQYQQA	VKILENVLEQ	NPDY (SEQ ID NO: 992)

85

	D644171	IPAYIIKANM LRKLGRYEEA LACVNKVLEL KEND (SEQ ID NO: 993)
	YLU28151	AEAWGRLGAC QAQNEKEDPA IRALERCIKL EPGN (SEQ ID NO: 994)
	S74853	PAALFDLGNA YLKLVKYPDA VTAYQKALKA EEQF (SEQ ID NO: 995)
	HGV2_HA	AETYYNLGLA YSFEKRYDNA LEHYQSALDV LEAR (SEQ ID NO: 996)
5	CELRO5F0	PVYFCNRAAA YCRLEQYDLA IQDCRTALAL DPSY (SEQ ID NO: 997)
	YC37_CY	ALIYNSLGYI CSAQEYELA IEYYKKALFY IPDF (SEQ ID NO: 998)
	NCF2_HU	AVAYFQRGML YYQTEKYDLA IKDLKEALIQ LRGN (SEQ ID NO: 999)
	HUMFKBP	IKALFRKGKV LAQQGEYSEA IPILRAALKL EPSN (SEQ ID NO: 1000)
	JC47510	VAVRLNLGLT YLKLCKPDKC IEFCKRVLDV FGNN (SEQ ID NO: 1001)
10	I40567	AKALLNMGNS YNQMSLFS A VPYYHKA IKA AKIS (SEQ ID NO: 1002)
	CEZK856	AIAYQQIATV YEQLGDHQKA LQFGLLASHL DPKT (SEQ ID NO: 1003)
	S756480	PAIYFNRAV HGVNLNNYQGA IDDCSQGILL DPQD (SEQ ID NO: 1004)
	S756481	EEAHYFRGLA YAMVNNYERA LADLNRTIRL NPYN (SEQ ID NO: 1005)
	LEPLIPL	WGAKANLATY YFSAGDFEKS IKLYEEAMKL KDAD (SEQ ID NO: 1006)
15	S756482	VDLLICRGQA QLGLEQPRQA IPDFDRAIEL DPRS (SEQ ID NO: 1007)
	CUT9_SC	AATWANLGH A YRKLKMYDAA IDALNQGLLL STND (SEQ ID NO: 1008)
	PS43_TO0	AHVLLNIAKC WMTEKKLDKT LGVVQKAEEL ADAV (SEQ ID NO: 1009)
	RNU76551	ADSLNNLANI KREQGNIEEA VRLYRKALEV FPEF (SEQ ID NO: 1010)
	S66842	PDPRIGIGLC FWQLKDSKMA IKSWQRALQL NPKN (SEQ ID NO: 1011)
20	HSU58970	SVLYSNRAAC HWKNGNCRDC IKDCTSALAL VPFS (SEQ ID NO: 1012)
	RNU76552	PDAYCNLANA LKEKGSVAEA EDCYNTALRL CPTH (SEQ ID NO: 1013)
	JC4751	EKALFRMQQA HLLRNDHDEA LVYFKKIVAK NPNN (SEQ ID NO: 1014)
	HSU58971	IKPLLRASA YEALEKYPMA YVDYKTVLQI DDNV (SEQ ID NO: 1015)
	RNU76553	YCVSRDLGNL LKALGRLEEA KACYLKAIET QPNF (SEQ ID NO: 1016)
25	HSU58972	VKAFYRRAQA HKALKDYKSS FADISNLLQI EPRN (SEQ ID NO: 1017)
	RNU76554	LDAYINLGNV LKEARIFDRA VAAYLRALSL SPNH (SEQ ID NO: 1018)
	YQGP_BA	HASYYNLALL YAEKNELAQA EKAIQTAVKL KPKE (SEQ ID NO: 1019)
	D907664	ARVSIMMGRV FMAKGEYAKA VESLQRVISQ DREL (SEQ ID NO: 1020)
	AB0010	GKYYKEIAEL YELEQNFEQA IIYFEKAADI YQSE (SEQ ID NO: 1021)
30	KNLC_HU	AATLNNLAVL YGKRGKYKEA EPLCKRALEI REKV (SEQ ID NO: 1022)
	S756010	FWSWYRQADC FRRWGHYAEA LQCYKKSLEI RPND (SEQ ID NO: 1023)
	G02058	VQSLSALGFV YKLEGEKRQA AEYYEKAQKI DPEN (SEQ ID NO: 1024)
	YQCH_BA0	PQAYHDLALI YFKQGGKQQA MDCFRKGIRS AVDF (SEQ ID NO: 1025)
	S756011	YAWYKRGIT LEHLGRYRDA AESYDNACQI QPEN (SEQ ID NO: 1026)
35	S756012	YWATYRLGES YRLSGQYEA A IASFRQTLTI RQDD (SEQ ID NO: 1027)
	S760710	VIALNNLANV YEKKQMVNKA LETYQETLAI EPNN (SEQ ID NO: 1028)
	CC23_YE	TNAWTLMGHE FVELSNSHAA IECYRRAVDI CPRD (SEQ ID NO: 1029)
	S756013	YAWFRQGE L WQQIYRYGRA IACYHRALDV EVDD (SEQ ID NO: 1030)
	S760711	AQEYYQLGSI YLDKKLYSQS INLFQKALKM AEQV (SEQ ID NO: 1031)
40	CEC34C6	ARAWCKLGLA HAENEKDQLA MQAFQKCLQI DAGN (SEQ ID NO: 1032)
	PTSR_PI0	VDWLKLGEV QTQNEKESDG IAALEKCLEL DPTN (SEQ ID NO: 1033)
	PTSR_PI1	VRARYNLGVS FINMGYRKEA VEHLTGISL HEVE (SEQ ID NO: 1034)
	INI6_HU	LESLSLLGFV YKLEGNMNEA LEYYERALRL AADF (SEQ ID NO: 1035)
	S75202	AELFGSMGYL YARQGQFAEA SRSFQQALRV NPNN (SEQ ID NO: 1036)
45	YO91_CA	ARAHRS LGHL LLMDKKFE EA YKHLRRSLEL QPIQ (SEQ ID NO: 1037)

	CC16_YE0	ISIQNLNGHT	YRKLNENEIA	IKCFRCVLEK	NDKN (SEQ ID NO: 1038)
	CC16_YE1	PLVLNEMGVM	YFKKNEFVKA	KKYLKKALEV	VKDL (SEQ ID NO: 1039)
	CC16_YE2	SSLCFLRGKI	YFAQNNFNKA	RDAFREAILV	DIKN (SEQ ID NO: 1040)
	CC16_YE3	AITWFSVATY	YMSLDRISEA	QKYYSKSSIL	DPSF (SEQ ID NO: 1041)
5	ACSC_AC	ADSLGGMGLV	SMRQGDTEAE	RRYFEEAMAA	DPKT (SEQ ID NO: 1042)
	HSU5251	TDVLRSAAKF	YRRKGDLDKA	IELFQRVLES	TPNN (SEQ ID NO: 1043)
	CC16_YE4	TEGYLNLARS	NEKLCEFQKT	ISYCKTCLNM	QGTT (SEQ ID NO: 1044)
	NCU89985_1-41	AIAFKNEGNK	AFAAHDWPKA	IEFYDKAIEL	NDKE (SEQ ID NO: 1045)
	CC16_YE5	AAAWLGFAHT	YALEGEQDQA	LTAYSTASRF	FPGM (SEQ ID NO: 1046)
10	CYP6_YE	AKALYRRGLA	YYHVNDTDMA	LNDLEMATTF	QPND (SEQ ID NO: 1047)
	D638750	VDCYLRLGAM	ARDKGNFYEA	SDWFKEALQI	NQDH (SEQ ID NO: 1048)
	S75685	AQAHCLLGMA	YLKNNMMTMA	KVHIGSAVKL	DPND (SEQ ID NO: 1049)
	S76576	PQLFYGWGLL	ETSKGNYQSA	LNHYDQALAL	DPEF (SEQ ID NO: 1050)
	D638751	VLFFFGLGQM	YIYRGDKENA	SQCFEKVLKA	YPNN (SEQ ID NO: 1051)
15	D638752	IPALLGKACI	SFNKKDYRGA	LAYYKKALRT	NPGC (SEQ ID NO: 1052)
	H644670	AIAWAEKGEI	LYREGKLKKS	LECFDNALKI	NPKD (SEQ ID NO: 1053)
	RLU3940	GRAYRELGFV	ALYRRRFDES	LEYFQQAQDL	NPND (SEQ ID NO: 1054)
	H644671	PDVYVRKARI	LRTLGENDKA	LEYFDKALKL	KPKY (SEQ ID NO: 1055)
	MMU1695	LAAFLNLAMC	YLKLREYNKA	VECCDKALGL	DSAN (SEQ ID NO: 1056)
20	D638753	AESCYQLARS	FHVQEDYDQA	FQYYYQATQF	ASSS (SEQ ID NO: 1057)
	I495640	GRLKVNMGNI	YLKQRNYSKA	IKFYRMALDQ	IPSV (SEQ ID NO: 1058)
	OM70_NE0	ALAYNLRGTF	HCLMGKHEEA	LADLSKSIEL	DPAM (SEQ ID NO: 1059)
	H644672	CQSLLYKGEI	LFKLGRYGEA	LKCLKKVFER	NNKD (SEQ ID NO: 1060)
	D638754	SDVWLNLAHI	YVEQKQYISA	VQMYENCLRK	FYKH (SEQ ID NO: 1061)
25	I495641	AQVLCQIANI	YELMEDPNQA	IEWLMQLISV	VPTD (SEQ ID NO: 1062)
	H644673	IRALMYIIQI	LIYLGRLNQA	LEYTKKALKL	NPDD (SEQ ID NO: 1063)
	OM70_NE1	TQSYIKRASM	NLELGHDPKA	EEDFNKAIEQ	NAED (SEQ ID NO: 1064)
	D638755	VTTSYNLARL	YEAMCEFHEA	EKLYKNILRE	HPNY (SEQ ID NO: 1065)
	CELF38B	AKIHYNLGKV	LGDNGLTKDA	EKNYWNAIKL	DPSY (SEQ ID NO: 1066)
30	JT0603	IPPYINRGNL	YSQQQDHHTA	IQDFTQAITY	DPNR (SEQ ID NO: 1067)
	YHBM_EC0	AQLLYERGV	YDSLGLRALA	RNDFSQALAI	RPDM (SEQ ID NO: 1068)
	OM70_NE2	AAKLKELGNK	AYGSKDFNKA	IDLYSKAIIC	KPDP (SEQ ID NO: 1069)
	I495642	SQALSKLGEL	YDSEGDKSQA	FQYYYESYRY	FPSN (SEQ ID NO: 1070)
	OM70_NE3	PVYYSNRAAC	HNALAQWEQV	VADTTAALKL	DPHY (SEQ ID NO: 1071)
35	PQ0180	YRAYYNRGLA	YDLAQPEQA	IADFQQALER	LPAT (SEQ ID NO: 1072)
	2408032-184	ARAFGRGLGRA	KLSLGDAAAA	ADAYKKGLDF	DPNN (SEQ ID NO: 1073)
	HSU5897	SATYSNRALC	YLVLKQYTEA	VKDCTEALKL	DGKN (SEQ ID NO: 1074)
	TAFKBP7	ISIKLNNAAC	KLKLKDYKEA	EKICSKVLEL	ESTN (SEQ ID NO: 1075)
	YC37_CY0	ILARINLARI	FELKNLKTEA	INMYQEVLI	DPKN (SEQ ID NO: 1076)
40	YC37_PO0	IVALNNLAKI	YEDTKNILKA	EALYDKVLNI	AKSN (SEQ ID NO: 1077)
	D643710	RDWAFNKALA	LRILGRYEEA	RECFFRGLAV	EKHL (SEQ ID NO: 1078)
	ATAC01	SSVHASLGKI	YNQLKQYDKA	VLHFGIALDL	SPSP (SEQ ID NO: 1079)
	HSU2036	TEALYNIGLT	YEKLNRLDEA	LDCFLKLHAI	LRNS (SEQ ID NO: 1080)
	CELK04G1	AVAWSNLGCV	FNSQGEIWL	IHHFEKAVTL	DPNF (SEQ ID NO: 1081)
45	ACU8998	HVLYSNRAAC	YMKLGRVPM	VKDCDKAIEL	SPTF (SEQ ID NO: 1082)

87

	ATAC02	TYAHTLCGHE	FAALEEFEDA	ERCYRKALGI	DTRH (SEQ ID NO: 1083)
	YHE3_PS	PAILDMSGWI	NYRQGLADA	ERYLRQALQR	YPDH (SEQ ID NO: 1084)
	CELK04G2	AVVHGNLACV	YYEQGLIDLA	IDTYKKAIDL	QPHF (SEQ ID NO: 1085)
	SC72_YE	PDVFVRKADC	LLKLRQWEEA	RATCERGLAL	APED (SEQ ID NO: 1086)
5	CELK04G3	AAAHSNLASI	LQQQGLNDA	ILHYKEAIRI	APTF (SEQ ID NO: 1087)
	CELK04G4	ADAHSNLASI	HKDAGNMAEA	IQSYSTALKL	KPDF (SEQ ID NO: 1088)
	CELK04G6	IDAYINLAAA	LVSGGDLEQA	VTAYFNALQI	NPDL (SEQ ID NO: 1089)
	JC4775	VNALKDRAEA	YLIEEMYDEA	IQDYETAQEH	NEND (SEQ ID NO: 1090)
	ATU62130	AEACNNLGVL	YKDRDNLDA	VECYQMALSI	KPNF (SEQ ID NO: 1091)
10	ATU62131	AEAFNNLGVL	YRDAGNITMA	IDAYEECLKI	DPDS (SEQ ID NO: 1092)
	ATU62132	APAYYNLGVV	YSEMMQYDNA	LSCYEKAAL	RPMY (SEQ ID NO: 1093)
	ATU62133	VEAHIGKGIC	LQTQNGKGLA	FDCFSEAIRL	DPHN (SEQ ID NO: 1094)
	YCA1_PL	AQRLKELGNK	CFQEGKYEEA	VKYFSDAITN	DPLD (SEQ ID NO: 1095)
	ATU62134	AQSLNNLGVL	YTVQGMKMDAA	ASMIEKAILA	NPTY (SEQ ID NO: 1096)
15	ATU62135	ADAMYNLGVA	YGEMLKFDMA	IVFYELAFHF	NPHC (SEQ ID NO: 1097)
	RNU7655	AEAYSNLGNV	YKERGQLQEA	IEHYRHALRL	KPDF (SEQ ID NO: 1098)
	ATU62136	AIVLTDLGTS	LKLAGNTQEG	IQKYYEALKI	DPHY (SEQ ID NO: 1099)
	PS43_TO	TEAYLNLARG	HEKLCEFSEA	VAYCRTCLGA	EGGP (SEQ ID NO: 1100)
	D82942	FNVRFRGLVA	LDNLGRFDEA	IDSFKIALAC	VPMK (SEQ ID NO: 1101)
20	YPU2283	SRGLASAARA	YRNEKRWDQA	LALWQSSLKK	DPTN (SEQ ID NO: 1102)
	CEUC55B6_3-59	VAKHLELGSQ	FLARAQFADA	LTQYHAAIEL	DPKS (SEQ ID NO: 1103)
	S755780	AFVYYRDGMS	AQADGEYAEA	LDNYEEALRL	EENP (SEQ ID NO: 1104)
	D86980	FKALYRSGVA	FYHLGDYDKA	LYYLKEARTQ	QPTD (SEQ ID NO: 1105)
	D86981	AKHYGNLGR	YQSMRKFEKA	EEMHIKAIQI	KEQL (SEQ ID NO: 1106)
25	H643320	TSILRQKASI	LEILGKLDEA	LDCVNKILSI	KKDD (SEQ ID NO: 1107)
	S55383	VKALYRRAQA	YTQLADLELA	EVDIKKALEI	DPEN (SEQ ID NO: 1108)
	YAD5_CL	GISYENLAWF	YYLTGKYDKA	IENFEKAISM	GSTN (SEQ ID NO: 1109)
	PPT1_YE	SIYFSNRAFA	HFKVDNFQSA	LNDCEAIKL	DPKN (SEQ ID NO: 1110)
	CELC33H	AQRLKEQNE	AFKKKKYHKA	MTIYSKSLEH	WPDP (SEQ ID NO: 1111)
30	CELC17G	AVLYFNRAAA	QKHLGNLRS	IKDCSMGRKF	DPHT (SEQ ID NO: 1112)
	TPRD_HU0	CLAYCGIGKV	YLKKNRFLEA	LNHFEEKARTL	IYRL (SEQ ID NO: 1113)
	D64417	ASLWYFKGKL	YEKQNKFEAA	LKYNNKAIQL	MPHH (SEQ ID NO: 1114)
	DMU18291	VYSYTLGHE	LVLTEEFDKA	MDYFRAAVVR	DPRH (SEQ ID NO: 1115)
	MXU7705	FIAQGNLQWA	YYKKGEPDRA	VESIKAAVTT	NPNF (SEQ ID NO: 1116)
35	RSMGPGN0	GYAHLTIALA	HLGMSQFQQC	LESFESAMNV	ANET (SEQ ID NO: 1117)
	CEF52H3	PSAYNNRAQA	YRLQNKPEKA	LDDLNEALSL	AGPK (SEQ ID NO: 1118)
	CYP4_BO	TKALYRRAQG	WQGLKEYDQA	LADLKKAQEI	APED (SEQ ID NO: 1119)
	IEFS_HU0	AKAYARIGNS	YFKEEKYKDA	IHFYNKSLAE	HRTF (SEQ ID NO: 1120)
	S76202	TSLQLLLGEI	YLSQNRPDQA	IAIYEAASKV	NGND (SEQ ID NO: 1121)
40	IEFS_HU2	IKGYTRKAAA	LEAMKDYTKA	MDVYQKALDL	DSSC (SEQ ID NO: 1122)
	F64399	YKALFGLGKS	YYLMSDNKNS	IKYFEKVLEL	NPND (SEQ ID NO: 1123)
	IEFS_HU3	MTYITNQAAV	YFEKGDNKNC	RELCEKAIEV	GREN (SEQ ID NO: 1124)
	SSN6_YE0	AKALTSLAHL	YRSRDMFQRA	AELYERALLV	NPEL (SEQ ID NO: 1125)
	YHR7_YE0	AVQLKNRGNH	FFTAKNFNEA	IKYYQYAIEL	DPNE (SEQ ID NO: 1126)
45	IEFS_HU4	GKGYSRKA	LEFLNRFEAA	KRTYEEGLKH	EANN (SEQ ID NO: 1127)

	SSN6_YE1	SDVWATLGHC	YLMLDDLQRA	YNAYQQALYH	LSNP (SEQ ID NO: 1128)
	SSN6_YE2	WDIWFQLGSV	LESMGEWQGA	KEAYEHVLAQ	NQHH (SEQ ID NO: 1129)
	SSN6_YE3	PKLWHGIGIL	YDRYGS LDYA	EEAFKVLLEL	DPHF (SEQ ID NO: 1130)
	HEMY_EC	PLLWSTLGQS	LMKHGEWQEA	SLAFRAALKQ	RPDA (SEQ ID NO: 1131)
5	SSN6_YE4	ATTWYHLGRV	HMIRTDYTAA	YDAFQQAVNR	DSRN (SEQ ID NO: 1132)
	SSN6_YE5	NEIYFRLGII	YKHQGWKSQA	LECFRYILPQ	PPAP (SEQ ID NO: 1133)
	RSMGPGN	FGALAGLGTM	LEQMDRPKPA	LEAYRAALAI	HPhL (SEQ ID NO: 1134)
	CYP7_YE	AKAYYRRGNS	YLKKKRLDEA	LQDYIFCKEK	NPDD (SEQ ID NO: 1135)
	S76685	AIAFFNLAMI	YKAQGNLLEA	IKGYQEAITL	QPDY (SEQ ID NO: 1136)
10	A55346	AEELKTQAND	YFKAKDYENA	IKFYSKAIEL	NPSN (SEQ ID NO: 1137)
	CELC18H	FRALGCLITA	HSEMGKYEDM	LRFVAQSEA	ARQM (SEQ ID NO: 1138)
	S761560	VWAYINRGVA	YYDLGCHQKS	LDDYNQALVI	DSKC (SEQ ID NO: 1139)
	S761561	VSSYILRAEI	HNLLGDFDAA	IEDCFTSLRL	NPMH (SEQ ID NO: 1140)
	NASP_HU	AETHYQLGLA	YGYNSQYDEA	VAQFSKSIEV	IENR (SEQ ID NO: 1141)
15	S761562	HEHLFKLGVE	QYILGNISDA	IDSFTKALEF	NDRF (SEQ ID NO: 1142)
	CELF10C0	AALWVLIGHE	FMEMKNNAAA	CVSYRRAIEI	DPAD (SEQ ID NO: 1143)
	S761563	DNTLYILGSS	FLSLEKYQSA	IDYLDQSLAL	NGNL (SEQ ID NO: 1144)
	C64478	PIDWFNLAYA	LYHLEKYDSA	LEAINEALKI	SPSN (SEQ ID NO: 1145)
	CEF10B50	VDAIASTALC	YAVLGNIDKA	TEFFNKALAI	DPFN (SEQ ID NO: 1146)
20	ATAC98_24-445	VDALYNLGGI	YMDLGRFQRA	SEMYTRVLTV	WPNH (SEQ ID NO: 1147)
	A565190	ASTYSAIGYI	HSLMGNFENA	VDYFHTALGL	RRDD (SEQ ID NO: 1148)
	E644170	VALWYFKXEL	YERLGKLDEA	LKCYEKVIEL	QPHY (SEQ ID NO: 1149)
	E644171	PITWVFGQL	YGMMSGNCDEA	LKCYNKALGI	ENRF (SEQ ID NO: 1150)
	E644172	IKALLSKARI	YERXGNIEAA	IEYYNKAVEN	IHKD (SEQ ID NO: 1151)
25	CELF10C	HRGWYGLGQM	YDIMKMPAYA	LFYYQEAQKC	KPHD (SEQ ID NO: 1152)
	E644173	YLALFLKGLA	LSAKGEIKEA	ITTFEELLSY	ESKN (SEQ ID NO: 1153)
	CYP4_BO0	LSCVLNIGAC	KLKMSDWQGA	VDSCLEALEI	DPSN (SEQ ID NO: 1154)
	CYP4_BO1	SEDLKNIGNT	FFKSQNWEMA	IKKYTKVLRV	VEGS (SEQ ID NO: 1155)
	MMU2783	VNELKEKGNK	ALSAGNIDDA	LQCYSEAIKL	DPQN (SEQ ID NO: 1156)
30	S752020	FNELLRQGKA	YVDNGNFPQA	IAIYQQAAML	DGEN (SEQ ID NO: 1157)
	ENAC0	LETMNNMALA	YEGLGKYKEA	EEIWAQLLEL	ANEC (SEQ ID NO: 1158)
	PFCYT12	AEGLYFLGRA	YMAQDRSADA	AKVFERAVAL	AGRQ (SEQ ID NO: 1159)
	CC16_YE	SEIHCSLGYL	YLKTKKLQKA	IDHLHKSLLY	KPNN (SEQ ID NO: 1160)
	BSATPC1	GSALYNIGNC	YDDKGELDQA	AEYFEKALPV	FEDY (SEQ ID NO: 1161)
35	ECAE00	ARGYAAVAVA	YRNLOQWQNS	LTLWQKALSL	EPQN (SEQ ID NO: 1162)
	ECAE01	AGGYYGLARL	AARNGNFDAG	LDQCQQLRA	CPTN (SEQ ID NO: 1163)
	ECAE02	WFARHLLACF	YYNKRSYNKA	IAFWQRCVEM	SPEF (SEQ ID NO: 1164)
	YAD5_CL0	NSVYRSLGIT	YAKIGDYKKS	EEYLKKALDA	EPEK (SEQ ID NO: 1165)
	PR06_YEAST-701	HKFFFLQLGQI	YHSMGNIEMS	RETYLSGTRL	VPNC (SEQ ID NO: 1166)
40	YAD5_CL1	INDVYSFALS	YHILGEPERA	LKYFLRAVEL	QPNV (SEQ ID NO: 1167)
	YPIA_BA	VNVHQRLAES	LSASGEFEDA	IPWYEKAVDE	NPDP (SEQ ID NO: 1168)
	NCF2_HU0	SRICFNIGCM	YTILKNMTEA	EKAFTRSINR	DKHL (SEQ ID NO: 1169)
	TPU93844_2-211	SVAYFYLGEL	FRLTKRFRKA	DIAYSTAVNF	EPSN (SEQ ID NO: 1170)
	SKI3_YE	VESWVGLGQA	YHACGRIEAS	IKVFDKAIQL	RPSH (SEQ ID NO: 1171)
45	YHJL_EC	SEALGALGQA	YSQKGDRAA	VANLEKALAL	DPHS (SEQ ID NO: 1172)

	ECAE0	ASLYYYRALV	HFHNEKIEEA	RICIDKSLQL	EPRR (SEQ ID NO: 1173)
	C644780	SRWWYVKGYI	YKLGNYKDA	YESFMNALRV	NPKD (SEQ ID NO: 1174)
	OM70_YE	ALALKDKGNQ	FFRNKKYDDA	IKYYNWALEL	KEDP (SEQ ID NO: 1175)
	C644781	SLYYAKKGDI	LYKLGDEEGA	IEAYNKAIKL	NSQN (SEQ ID NO: 1176)
5	C644782	YEDWVTEANY	YLDEGIYDKA	VECYLKALEK	KNTN (SEQ ID NO: 1177)
	C644783	LNALFKAGKI	YLLFGDIDKA	YDAFNEILQQ	NPSH (SEQ ID NO: 1178)
	A64512	AIALSEIAKA	QYNIGMHDEA	LKNYDKAIFI	TEGV (SEQ ID NO: 1179)
	LMU7384	AKGYVRRGAA	LHGMRRYDDA	IAAYEKGLKV	DPSN (SEQ ID NO: 1180)
	C644784	ISTLKSLAIV	LEKSGKIDEA	ITTYTKILKI	VNSL (SEQ ID NO: 1181)
10	PTSR_HU	PDVQCGLGVL	FNLSGEYDKA	VDCFTAALSV	RPND (SEQ ID NO: 1182)
	HSU4657	AETFKEQGNA	YYAKKDYNEA	YNYTKAIDM	CPKN (SEQ ID NO: 1183)
	CELK04G	ADAYSNMGNT	LKEMGDSSAA	IACYNRAIQI	NPAF (SEQ ID NO: 1184)
	YKD1_CA0	PEMLNNVGAL	YMSMKQYEKA	EHFHKRAKER	LEEQ (SEQ ID NO: 1185)
	YKD1_CA1	TLAHYGLGQM	YIHRNEIEEA	IKCFDTVHKR	LPNN (SEQ ID NO: 1186)
15	S619910	FRGYSRLGFA	KYAQGKPEEA	LEAYKKVLDI	EGDN (SEQ ID NO: 1187)
	S619911	AEDLKMQGNK	AMANKDYELA	INKYTEAIKV	LPTN (SEQ ID NO: 1188)
	YKD1_CA2	AEAFYQMGRG	RHAQGQFDGA	YKYYYQARQA	NNGE (SEQ ID NO: 1189)
	SR72_CA	AI IHGQMAYI	LQLQGRTEEA	LQLYNQIIKL	KPTD (SEQ ID NO: 1190)
	CELT09B	PKYYQNRAMC	YFQLNNLKMT	EEDCKRALEL	SPNE (SEQ ID NO: 1191)
20	MXU77058_2-100	PDANNAMGIL	LHLAFRRPDE	AVKHYTEKALE	VRPD (SEQ ID NO: 1192)
	NRFG_EC	SEQWALLGEY	YLWQNDYSNS	LLAYRQALQL	RGEN (SEQ ID NO: 1193)
	NRFF_HA	AETWLQLGEA	YVQNNFEFSA	LVAYSNAEKL	SGSK (SEQ ID NO: 1194)
	CELT25F	ADVWYNIGQI	LVDIGDLVSA	ARSFRIALSH	DPDH (SEQ ID NO: 1195)
	S45064	GQVSLSMGNA	FLGLSVFQKA	LESFEKALRY	AHNN (SEQ ID NO: 1196)
25	S76315	YLRYRALATT	YRDLEKYDLA	MANIDRALAI	TPDN (SEQ ID NO: 1197)
	CEAF16427_7-686	VTVLQNIALA	EFHMQNYNRS	LLFYRKALHL	DPTH (SEQ ID NO: 1198)
	BIMA_EM0	ANVHYLLGKL	YKMLRDKGNA	IKHFTTALNL	DPKA (SEQ ID NO: 1199)
	BIMA_EM1	AAVLCLQGKL	WQAHKEHNKA	VECYAAALKL	NPFM (SEQ ID NO: 1200)
	S76156	FNGFYNRGLA	YYKMGQIKKA	IKDFSDALIV	KPTF (SEQ ID NO: 1201)
30	BIMA_EM2	AVLICIGLIV	LEKMNNPKSA	LIQYNRACTL	APHS (SEQ ID NO: 1202)
	BIMA_EM3	AYGFTLQGHE	YVANEEDYKA	LDAYRSGINA	DSRH (SEQ ID NO: 1203)
	NUC2_SC1	SVLITCIGMI	YERCKDYKKA	LDFYDRACKL	DEKS (SEQ ID NO: 1204)
	JC4348	GRVYNLLGNI	SFDQTNLKKA	LKYYQIAYQL	FDGN (SEQ ID NO: 1205)
	LBAPREL	AMTLNNMGEL	YHYKNEHDKS	KSYFLQSFQL	QQTV (SEQ ID NO: 1206)
35	YO91_CA0	LGTWFNAGYC	AWKLENFKES	TQCYHRCVSL	QPDH (SEQ ID NO: 1207)
	PTSR_YE	PEIQLCLGLL	FYTKDDFDKT	IDCFESALRV	NPND (SEQ ID NO: 1208)
	A60088	ALCLLCFADI	HRHRSDIGKA	LPRYESSLNI	MTEI (SEQ ID NO: 1209)
	BSU55040	VQACYDLAII	FFKQGRKEEA	LDYCVKGKSY	AEKY (SEQ ID NO: 1210)
	ATAC0	PESWCAVGNC	YSLRKDHDTA	LKMFQRAIQI	NERF (SEQ ID NO: 1211)
40	D64371	AEYYYKKGVE	VGNGKDVEKA	LEYFNKAIEL	NPFY (SEQ ID NO: 1212)
	H64332	AHAWYLGRI	LKKLGNIEKA	LDALKMAINL	NENL (SEQ ID NO: 1213)
	S748060	TSAWQNRGSA	LGVMGKLEEA	LANFDEALAQ	NPDD (SEQ ID NO: 1214)
	S748061	AEDWLNLIQ	QAQRGELETA	IASWGAEISL	NPQM (SEQ ID NO: 1215)
	YREC_SY0	LNALLEQGNE	QLTNRNFAQA	VQHYRQALTL	EANN (SEQ ID NO: 1216)
45	YREC_SY1	LAYSLGLATV	QFRAGDYDQA	LVAYRKVLAK	DSNN (SEQ ID NO: 1217)

90

	YREC_SY2	AEFFNALGFN	LAQSGDNRSA	INAYQ ^Q RAT ^Q L	Q ^Q PNN (SEQ ID NO: 1218)
	NPRA_BA	YYYYKFLGLL	YYCKEKYEDA	LEY ^Y KKAEQR	FRSQ (SEQ ID NO: 1219)
	JC47750	TAARL ^Q R ^Q GH ^L	LLK ^Q GKLDEA	EDDFKKVLKS	NPSE (SEQ ID NO: 1220)
	G020580	AFAYTDLANM	YAEGGQYSNA	EDIFRKALRL	ENIT (SEQ ID NO: 1221)
5	PPP5_RA	IKGY ^Y RR ^A AS	NMALGKFRAA	LRDYETVVKV	KPND (SEQ ID NO: 1222)
	YQCH_BA	ATAFFNLGNC	YHKMDNLNKA	ARYIEQ ^Q ALV ^Q	YRKI (SEQ ID NO: 1223)
	CC27_HU2	SVLLCHIGVV	QH ^A LKKSEKA	LDTLNKAIVI	DPKN (SEQ ID NO: 1224)
	S61991	AIYYANRAAA	HSSLKEYDQA	VKDAESAISI	DPSY (SEQ ID NO: 1225)
	YREC_SY	ARIHGALGYA	LSQLGNYSEA	VTAYRRATEL	EDDN (SEQ ID NO: 1226)
10	CC27_HU3	CFTLSLLGHV	YCKTDR ^L LAKG	SECYQ ^Q KSLSL	NPFL (SEQ ID NO: 1227)
	CC27_HU4	SLVYFLIGKV	YKKLGQTHLA	LMNFSWAMD ^L	DPKG (SEQ ID NO: 1228)
	Consensus/60%	spsbhp ^b G.h	abp ^b bscbppA	lphapcAlpl	sspp (SEQ ID NO: 1229)

Table 11 Alignment of TPR Repeat Sequences

15

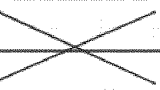
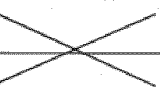
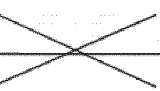
	Target Binder	Scaffold	Degron	Tag	
Structure	1	2	3	4	4
PPX172		CTPR		HA	HA
PPX173		CTPR	P27v1	HA	HA
PPX174		CTPR	P27v2	HA	HA
PPX175		CTPR	NRF2v1	HA	HA
PPX176		CTPR	NRF2v2	HA	HA
PPX177	AXIN	CTPR	P27v1	HA	HA
PPX178	AXIN	CTPR	P27v2	HA	HA
PPX179	BCL9	CTPR	P27v1	HA	HA
PPX181	LRH1	CTPR	P27v1	HA	HA
PPX182	LRH1	CTPR	P27v2	HA	HA
PPX183	Phospho	CTPR	P27v1	HA	HA
PPX184	Phospho	CTPR	P27v2	HA	HA
PPX185	Phospho	CTPR	P27v1	HA	HA
PPX186	Phospho	CTPR	P27v2	HA	HA
PPX187	AXIN	CTPR	NRF2v1	HA	HA
PPX188	AXIN	CTPR	NRF2v2	HA	HA
PPX189	LRH1	CTPR	NRF2v1	HA	HA

Table 12

	Target Binder	Scaffold	Degron	Tag	
Structure	1	2	3	4	4
PPX190	LRH1	CTPR	NRF2v2	HA	HA
PPX191	BCL9	CTPR	NRF2v1	HA	HA
PPX192	BCL9	CTPR	NRF2v2	HA	HA
PPX193	Phospho	CTPR	NRF2v1	HA	HA
PPX194	Phospho	CTPR	NRF2v2	HA	HA
PPX195	Phospho	CTPR	NRF2v1	HA	HA
PPX196	Phospho	CTPR	NRF2v2	HA	HA
PPX197	AXIN	CTPR	SPOP	HA	HA
PPX198	LRH1	CTPR	SPOP	HA	HA
PPX199	BCL9	CTPR	SPOP	HA	HA
PPX200	Phospho	CTPR	SPOP	HA	HA
PPX201	Phospho	CTPR	SPOP	HA	HA
PPX202	AXIN	CTPR	PHYL	HA	HA
PPX203	LRH1	CTPR	PHYL	HA	HA
PPX204	BCL9	CTPR	PHYL	HA	HA
PPX205	Phospho	CTPR	PHYL	HA	HA
PPX206	Phospho	CTPR	PHYL	HA	HA

Table 12 (CONT)

	Target Binder	Scaffold	Degron	Tag	
Structure	1	2	3	4	4
PPX208	LRH1	CTPR	Trib	HA	HA
PPX210	Phospho	CTPR	Trib	HA	HA
PPX211	Phospho	CTPR	Trib	HA	HA
PPX212	AXIN	CTPR	P53	HA	HA
PPX213	LRH1	CTPR	P53	HA	HA
PPX214	BCL9	CTPR	P53	HA	HA
PPX215	Phospho	CTPR	P53	HA	HA
PPX216	Phospho	CTPR	P53	HA	HA
PPX217	Phospho	CTPR	P53	HA	HA
PPX218	Phospho	CTPR	P53	HA	HA
PPX219	LRH1	CTPR	P53	HA	HA
PPX220	Phospho	CTPR	P53	HA	HA
PPX221	BCL9	CTPR	P53	HA	HA
PPX222	AXIN	CTPR	P27	HA	HA
PPX223	BCL9	CTPR	P27v1	HA	HA
PPX224	LRH1	CTPR	P27v1	HA	HA
PPX225	AXIN	CTPR	PAM2	HA	HA

Table 12 (CONT)

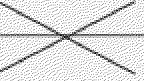
	Target Binder	Scaffold	Degron	Tag	
Structure	1	2	3	4	4
PPX226	BCL9	CTPR	PAM2	HA	HA
PPX227	LRH1	CTPR	PAM2	HA	HA
PPX228	Phospho	CTPR	PAM2	HA	HA
PPX229	Phospho	CTPR	PAM2	HA	HA
PPX230	AXIN	CTPR	CDC25B	HA	HA
PPX231	LRH1	CTPR	CDC25B	HA	HA
PPX232	BCL9	CTPR	CDC25B	HA	HA
PPX233	Phospho	CTPR	CDC25B	HA	HA
PPX234	Phospho	CTPR	CDC25B	HA	HA
PPX235	AXIN	CTPR		HA	HA
PPX236	LRH1	CTPR		HA	HA
PPX237	BCL9	CTPR		HA	HA
PPX238	Phospho	CTPR		HA	HA
PPX239	Phospho	CTPR		HA	HA
PPX240		CTPR	SPOF	HA	HA
PPX241		CTPR	PHYL	HA	HA
PPX242		CTPR	Trib	HA	HA

Table 12 (CONT)

	Target Binder	Scaffold	Degron	Tag	
Structure	1	2	3	4	4
PPX243		CTPR	P53	HA	HA
PPX244		CTPR	PAM2	HA	HA
PPX245		CTPR	CDC25B	HA	HA
PPX246	SOS	RTPR	P27v1	HA	HA
PPX247		RTPR		HA	HA
PPX248	SOS	RTPR		HA	HA
PPX249		RTPR	P27v1	HA	HA

Table 12 (CONT)

Schematic representation of components used to build the CTPR/RTPR constructs.

CTPR: Consensus-designed tetratricopeptide sequence.

RTPR: As above but with the lysine residues replaced with arginine residues.

LRH1: a beta-catenin binding sequence from the protein LRH1

Phospho: a beta-catenin binding sequence from the protein APC (Adenomatous polyposis coli).

AXIN: an alpha-helical beta-catenin binding sequence from the protein AXIN.

BCL9: an alpha-helical beta-catenin binding sequence from the protein BCL9.

SOS: an alpha-helical KRAS-binding sequence from the protein SOS1 (Son of sevenless homolog 1).

P27v1: a degron sequence from the protein p27 that binds the E3 SCFSkp2.

P27v2: a degron sequence from the protein p27 that binds the E3 SCFSkp2.

Puc: a degron sequence from the protein Puc that binds the E3 Cul3-SPOP.

NRF2v1: a degron sequence from the protein NRF2 that binds the E3 Cul3-KEAP1.

NRF2v2: a degron sequence from the protein NRF2 that binds the E3 Cul3-KEAP1.

PHYL: a degron sequence from the protein PHYL that binds the E3 SIAH.

Trib: a degron sequence from the protein Trib that binds the E3 COP1.

PAM2: a degron sequence from the protein PAM2 that binds the E3 UBR5.

CDC25B: a degron sequence from the protein CDC25B that binds the E3 beta-TRCP.

P53: an alpha-helical degron sequence from the protein p53 that binds the E3 MDM2.

Constructs	Amino Acid Sequence	Grafting site for target binding ligand	Degron grafting site
PPX172	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1230)	N/A	N/A
PPX173	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPKDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1231)	N/A	Site 4
PPX174	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPNRPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1232)	N/A	Site 4
PPX175	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1233)	N/A	Site 4
PPX176	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPLLDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1234)	N/A	Site 4
PPX177	MGAYPEYILDIHVYRVQLELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPKDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1235)	Site 1	Site 4
PPX178	MGAYPEYILDIHVYRVQLELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPNRDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1236)	Site 1	Site 4
PPX179	MSQEQLEHRYRSLITLTDIQLMLDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPKDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1237)	Site 1	Site 4
PPX181	MYEQATAAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPKDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1238)	Site 1	Site 4
PPX182	MYEQATAAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPNRDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1239)	Site 1	Site 4
PPX183	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNSEELEALEALELDEDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAGSNEOEPPKDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA (SEQ ID NO: 1240)	Site 2	Site 5

PPX184	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN AGSNEOEPNR DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1241)	Site 2	Site 5
PPX185	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN AGSNEOEPK DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1242)	Site 2	Site 5
PPX186	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN AGSNEOEPNR DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1243)	Site 2	Site 5
PPX187	MGAYPEY ILD IHVYRVQLELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1244)	Site 1	Site 4
PPX188	MGAYPEY ILD IHVYRVQLELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1245)	Site 1	Site 4
PPX189	MYEQAI AAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1246)	Site 1	Site 4
PPX190	MYEQAI AAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1247)	Site 1	Site 4
PPX191	MSQEQLHRYRS LIT LYD IQ LM L DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1248)	Site 1	Site 4
PPX192	MSQEQLHRYRS LIT LYD IQ LM L DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1249)	Site 1	Site 4
PPX193	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1250)	Site 2	Site 5
PPX194	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1251)	Site 2	Site 5
PPX195	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DPETGEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1252)	Site 5	Site 2

	NAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1252)		
PPX196	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN LDPELCELL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1253)	Site 5	Site 2
PPX197	MGAYPEY ILDIHVYRVQ LELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DEVTSTTSSS DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1254)	Site 1	Site 4
PPX198	MYEQAI AAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DEVTSTTSSS DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1255)	Site 1	Site 4
PPX199	MSQEQLHRYRS LITLYDIQL M LDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DEVTSTTSSS DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1256)	Site 1	Site 4
PPX200	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DEVTSTTSSS DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1257)	Site 5	Site 2
PPX201	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DEVTSTTSSS DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1258)	Site 2	Site 5
PPX202	MGAYPEY ILDIHVYRVQ LELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN LRPVAMVRPWVR DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1259)	Site 1	Site 4
PPX203	MYEQAI AAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN LRPVAMVRPWVR DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1260)	Site 1	Site 4
PPX204	MSQEQLHRYRS LITLYDIQL M LDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN LRPVAMVRPWVR DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1261)	Site 1	Site 4
PPX205	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN LRPVAMVRPWVR DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1262)	Site 5	Site 2
PPX206	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN LRPVAMVRPWVR DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1263)	Site 2	Site 5

PPX208	MYEQAI AA YLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNN SDQIVPEYQE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDY A (SEQ ID NO: 1264)	Site 1	Site 4
PPX210	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SDQIVPEYQE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDY A (SEQ ID NO: 1265)	Site 2	Site 5
PPX211	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SDQIVPEYQE DPNNAEAWYN LGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKA LELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDY A (SEQ ID NO: 1266)	Site 5	Site 2
PPX212	MGAYPEY ILD IHVYRVQLELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELD PNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQ KAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWY NLGNAYYKQGDYQKAIEYYQKALELDPNN FAAYWNLLSAY GGSSGSSGYPYDVPD YAYPYDVPDYA (SEQ ID NO: 1267)	Site 1	Site 6
PPX213	MYEQAI AA YLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGN AYYKQGDYQKAIEYYQKALELDPNN FAAYWNLLSAY GGSSGSSGYPYDVPDYAYPY DVPDYA (SEQ ID NO: 1268)	Site 1	Site 6
PPX214	MSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAL ELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQG DYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNN FAAYWNLLSAY GGSSGSSGYPYD VPDYAYPYDVPDYA (SEQ ID NO: 1269)	Site 1	Site 6
PPX215	MYPYDVPDYAYPYDVPDYAGSSGSSGAEAWYNLGNAYYKQGDYQKAIEYYQKALE LDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELD PNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQK AIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN FAAYWN LLSAY G (SEQ ID NO: 1270)	Site 2	Site 6
PPX216	MYPYDVPDYAYPYDVPDYAGSSGSSGAEAWYNLGNAYYKQGDYQKAIEYYQKALE LDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALEALELDE DP NNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQK AIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN FAAYWN LLSAY G (SEQ ID NO: 1271)	Site 3	Site 6
PPX217	MYPYDVPDYAYPYDVPDYAGSSGSSGAEAWYNLGNAYYKQGDYQKAIEYYQKALE LDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGD YQKAIEYYQKALELDPNN SEELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQK AIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN FAAYWN LLSAY G (SEQ ID NO: 1272)	Site 4	Site 6
PPX218	MYPYDVPDYAYPYDVPDYAGSSGSSGAEAWYNLGNAYYKQGDYQKAIEYYQKALE LDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGD YQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEE LEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN FAAYWN LLSAY G (SEQ ID NO: 1273)	Site 5	Site 6
PPX219	MYPYDVPDYAYPYDVPDYASQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGN AYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELD PNN FAAYWNLLSAY G (SEQ ID NO: 1274)	Site 1	Site 3
PPX220	MYPYDVPDYAYPYDVPDYAAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SE ELEALEALELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWY NLGNAYYKQGDYQKAIEYYQKALELDPNN FAAYWNLLSAY G (SEQ ID NO: 1275)	Site 2	Site 4

PPX221	MYPYDVPDYAYPYDVPDYASQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNA YYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELD PNNFAAYWNLLSAYG (SEQ ID NO: 1276)	Site 1	Site 3
PPX222	MGAYPEYILDHIVYRVQLELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELD PNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAAGSNEOEPKKDPNNAEAW YNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1277)	Site 1	Site 3
PPX223	MSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAL ELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAAGSNEOEPKKDPNNA EAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1278)	Site 1	Site 3
PPX224	MYEQATAAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAAGSNEOEPKKDPNNAEAWYNLGNA YYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA* (SEQ ID NO: 1279)	Site 1	Site 3
PPX225	MGAYPEYILDHIVYRVQLELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELD PNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQ KAIEYYQKALELDPNNLSVNAPEFFYP,SDPNNAEAWYNLGNAYYKQGDYQKAIEYY QKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYP YDVPDYA (SEQ ID NO: 1280)	Site 1	Site 4
PPX226	MSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAL ELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGD DYQKAIEYYQKALELDPNNLSVNAPEFFYP,SDPNNAEAWYNLGNAYYKQGDYQKA IEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDY AYPYDVPDYA (SEQ ID NO: 1281)	Site 1	Site 4
PPX227	MYEQATAAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNLSVNAPEFFYP,SDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALE LDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPD YA (SEQ ID NO: 1282)	Site 1	Site 4
PPX228	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNSEELEALEALELDEAEAWYN LGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKA LELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNLSVNAPEFFYP,SDPN NAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (S EQ ID NO: 1283)	Site 2	Site 5
5PPX229	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNLSVNAPEFFYP,SDPNNAEAWY NLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQK ALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNSEELEALEALELD EDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPD YA (SEQ ID NO: 1284)	Site 5	Site 2
PPX230	MGAYPEYILDHIVYRVQLELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELD PNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQ KAIEYYQKALELDPNNTEEDDGFFVD,SDPNNAEAWYNLGNAYYKQGDYQKAIEYYQ KALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPY DVPDYA (SEQ ID NO: 1285)	Site 1	Site 4
PPX231	MYEQATAAYLDALMCDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNTEEDDGFFVD,SDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALEL DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDY A (SEQ ID NO: 1286)	Site 1	Site 4
PPX232	MSQEQLEHRYRSLITLYDIQLMLDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAL ELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGD DYQKAIEYYQKALELDPNNTEEDDGFFVD,SDPNNAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYA YPYDVPDYA (SEQ ID NO: 1287)	Site 1	Site 4
PPX233	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNSEELEALEALELDEDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNTEEDDGFFVD,I	Site 2	Site 5

	DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDY A (SEQ ID NO: 1288)		
PPX234	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN TEEDDGFVD IDPNNAEAWYN LGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKA LELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALELELDE DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDY A (SEQ ID NO: 1289)	Site 5	Site 2
PPX235	MGAYPEY ILDIHVYRVQ LELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELD PNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQ KAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWY NLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1290)	Site 1	N/A
PPX236	MYEQAIAYLDALMC DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGN AYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1291)	Site 1	N/A
PPX237	MSQEQLHRYR SLITLYDIQLML DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKAL ELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQ DYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1292)	Site 1	N/A
PPX238	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SEELEALELELDE DPNNAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGN AYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1293)	Site 2	N/A
PPX239	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKA IEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNL GNAYYKQGDYQKAIEYYQKALELDPNN SEELEALELELDE DPNNAEAWYNLGN AYYKQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1294)	Site 5	N/A
PPX240	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKA IEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN DEVTSTT SSSD DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYK GDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA** (SEQ ID NO: 1295)	N/A	Site 4
PPX241	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKA IEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN LRPVAMV RPWVR DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1296)	N/A	Site 4
PPX242	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKA IEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN SDQIVPE YQED DPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYK GDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1297)	N/A	Site 4
PPX243	MYPYDVPDYAYPYDVPDYAGSSSGAEAWYNLGNAYYKQGDYQKAIEYYQKALE LDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGD YQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAE WYNLGNAYYKQGDYQKAIEYYQKALELDPNN FAAYWNLLSAY G (SEQ ID NO: 1298)	N/A	Site 6
PPX244	MAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKA IEYYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN LSVNAPE FYP SDPNNAEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNAEAWYNLGNAYYK QGDYQKAIEYYQKALELDPNNYPYDVPDYAYPYDVPDYA (SEQ ID NO: 1299)	N/A	Site 4

All constructs were cloned into pCDNA3.1

Bold and Underlined dashed (heavy): Degron (E3 ligase engager)
 Underlined solid: HA Tag

Loop Grafting site 2 – res 35 to 36 SEQ ID No 1230 (Between first and second CTPR modules - DPNN sequence is placed before and after the targeting/degron peptides to facilitate the loop structure)

Loop Grafting site 4 – res 103 to 104 SEQ ID No 1230(Between third and fourth CTPR modules - DPNN sequence is placed before and after the targeting/degron peptides to facilitate the loop structure)

Loop Grafting site 6 – res 177 to 178 SEQ ID No 1230 (after the last CTPR module)

	O04242/1-30	NGHTALHIAA	SK-----GDE	QCVKLLLEHG	A-----DPNA	(SEQ ID NO: 1305)
	O93130/1-30	DGYTPLHYAI	KY-----NTI	RIAKYLLDNG	A-----DMTL	(SEQ ID NO: 1306)
5	VB18_VARV/1-32	TGYTALHCYL	YNN---YFTN	DVLKVLLNHG	V-----DVTI	(SEQ ID NO: 1307)
	Q25328/1-30	DKDTALHLAV	YY-----KNL	QMIKLLIKYG	I-----DVTI	(SEQ ID NO: 1308)
	Q14349_1/1-30	GGWTPIIWAA	EH-----KHI	EVIRMLLTRG	A-----DVTI	(SEQ ID NO: 1309)
	BCL3_HUMAN/1-30	SGRSPLIHAV	EN-----NSL	SMVQLLLQHG	A-----NVNA	(SEQ ID NO: 1310)
	Q24241_4/1-31	DGLTPLHCAS	RS-----GHV	EVIKHLLQQN	AP---ILTK	(SEQ ID NO: 1311)
10	O43150/1-30	KGSTALHYCC	LT-----DNA	ECLKLLLRGK	A-----SIEI	(SEQ ID NO: 1312)
	G3790744/1-30	NKQTSLHYAC	SK-----NHV	EIVKLLIEAD	P----NIIN	(SEQ ID NO: 1313)
	Q25328_3/1-30	EKYTPLHLAA	MS-----KYP	ELIQILLDQG	S-----NFEA	(SEQ ID NO: 1314)
	GLP1_CAEEL/1-30	KGRTALHYAA	MH-----DNE	EMVIMLVRRS	S-----NKDK	(SEQ ID NO: 1315)
	YA2A_SCHPO/1-31	DKMTPLHWSI	VG-----GNL	KCMKLILKEG	GI---PCTA	(SEQ ID NO: 1316)
15	VB04_VACCC/1-30	NRYTPLHYVS	CR-----NKY	DFVKLLISKG	A-----NVNA	(SEQ ID NO: 1317)
	HT16_HYDAT/1-33	GLDTRLHLAC	EE-----KNP	NTVKELLQDS	VIK-ENVNA	(SEQ ID NO: 1318)
	O04703_1/1-30	HRSTAIHVAI	EE-----NHL	EMVKLLLLNG	D-----EIDD	(SEQ ID NO: 1319)
	G3790744_1/1-30	EGNTALHLAC	DE-----NRG	DVAILLVNRG	A-----DMKM	(SEQ ID NO: 1320)
	Q25338_1/1-30	GNLTVLHLAV	ST-----GQI	NIKELLKRG	S-----NIEE	(SEQ ID NO: 1321)
20	Q21920/1-30	TLETPLTIAC	AN-----GHK	DIVELLKKEG	A-----NIEH	(SEQ ID NO: 1322)
	O16568/1-30	DGKTCLHVAI	EN-----LHK	ETIELMIQSG	A-----DKYA	(SEQ ID NO: 1323)
	Q21920_1/1-30	TGDTPLSLAA	RN-----GYI	AIMKMLIEKG	G----DLTA	(SEQ ID NO: 1324)
	O82630/1-30	DGRSPLHLAA	SR-----GYE	DITLYLIQES	V-----DVNI	(SEQ ID NO: 1325)
	O00542/1-31	QSRTLLHHAV	ST-----GSK	DVVRVLLDHA	PP---EILD	(SEQ ID NO: 1326)
25	Q25338_2/1-30	DGSTALHLAV	SG-----RKM	KTVETLLNKG	A-----NLKE	(SEQ ID NO: 1327)
	Q86916/1-32	NGKNLLHMYM	CNF---NVRI	NVIKLLIDSG	V-----NFLQ	(SEQ ID NO: 1328)
	Q23595/1-30	IGQNALHMAA	RV-----GNL	NILKYLLDRL	P-----ELRD	(SEQ ID NO: 1329)
	Q21587/1-30	IGCSALHLCA	EH-----GHY	RMKLLLLQYM	K-----VVEQ	(SEQ ID NO: 1330)
	P87621/1-34	DGNTPLHIVC	SKT---VKNV	DIIDLPLPST	DV--NKQNK	(SEQ ID NO: 1331)
30	YD57_SCHPO/1-33	AGWTPLMISI	NNR---SVPD	NVIEELINRS	DV---DPTI	(SEQ ID NO: 1332)
	O04097_1/1-30	HGRTPLHHCI	SS-----GNH	KFAKILLRRG	A-----RPSI	(SEQ ID NO: 1333)
	P87621_2/1-30	YDSTDFKMAV	EV-----GSI	RCVKYLLDND	I-----ICED	(SEQ ID NO: 1334)
	Q84644_1/1-30	YGNTPLAIAI	ST-----GQC	DMVKLLMNYV	T----NIFD	(SEQ ID NO: 1335)
	O17055_1/1-31	YGD TALHLSC	YS-----GRL	DIVKSILDSS	PT---NIVN	(SEQ ID NO: 1336)
35	VB18_VARV_1/1-3	CGNTPFHLYL	SIE--MCNNI	HMTKMLLTFN	P----NFEI	(SEQ ID NO: 1337)
	HRG_COWPX/1-32	NYCTALQYYI	KSS---HIDI	DIVKLLMKGI	D-----NTAY	(SEQ ID NO: 1338)
	O55222_1/1-30	HGFSPLHWAC	RE-----GRS	AVVEMLIMRG	A-----RINV	(SEQ ID NO: 1339)
	Q28282/1-30	EGRSAFHVVA	SK-----GNL	ECLNAILIHG	V-----DITT	(SEQ ID NO: 1340)
	G3929219_3/1-30	DFMTPLHVAA	ER-----AHN	DVMEVLHKKH	A-----KMNA	(SEQ ID NO: 1341)
40	BCL3_HUMAN_1/1-	DGDTPLHIAV	VQGN-LPAVH	RLVNLFQQGG	R-----ELDI	(SEQ ID NO: 1342)
	BCL3_HUMAN_2/1-	HGQTAHLAC	EH-----RSP	TCLRALLDSA	AP--GTLDL	(SEQ ID NO: 1343)
	O73630/1-30	HGLSPVHWSV	KM-----KNE	KCLVLLVKAG	A-----NVNS	(SEQ ID NO: 1344)
	Q18297_2/1-30	DQNTPMHIVA	SN-----GYL	EMMQLLQKHG	A-----SITQ	(SEQ ID NO: 1345)
	Q94527/1-31	DGLTPLHMAI	RQ-----NKY	DVAKKLISYD	RT---SISV	(SEQ ID NO: 1346)
45	O90757/1-30	KGYTALYYTI	CN-----NNY	DMVCFLLEKN	A-----DISI	(SEQ ID NO: 1347)
	HT16_HYDAT_2/1-	NGRTPVQVVC	FY-----NHA	STLHLLISKG	SA---DFLK	(SEQ ID NO: 1348)

	P87609_1/1-31	NGYTICIAIAI	NE----	SRNI	ELLKMLLCHK	P----	TLDC	(SEQ ID NO: 1349)
	O90757_1/1-30	GGRTSLHLAI	KE-----	RNY	EAAFVLINNG	A----	NVDS	(SEQ ID NO: 1350)
	Q20109_1/1-31	MGWSPLMWAV	YK-----	NHL	DVVDLLVNAK	VN---	ACDK	(SEQ ID NO: 1351)
	SWI6_YEAST_1/1-	HGNTPLHWLT	SI-----	ANL	ELVKHLVKHG	S----	NRLY	(SEQ ID NO: 1352)
5	P93755/1-31	DYRTPLMVAA	TY-----	GS	DVIKLIVSLT	DA---	DVNR	(SEQ ID NO: 1353)
	Q19995/1-31	AGRTPLHYAA	AD-----	PNGE	HMIKVLQKSG	G----	DAFI	(SEQ ID NO: 1354)
	O48738/1-30	DGFTPIHMAA	KE-----	GHV	RIIKEFLKHC	P----	DSRE	(SEQ ID NO: 1355)
	Q89202/1-30	ILLSCIHITI	KN-----	GHV	DMMILLLDYM	T----	STNT	(SEQ ID NO: 1356)
	G3927831_1/1-30	HKQTPLMLAA	MY-----	GRI	SCVKKLAEVG	A----	NILM	(SEQ ID NO: 1357)
10	FEM1_CAEEL_3/1-	HGHTCLMIAS	YR-----	NKV	GIVEELLKTG	I----	DVNC	(SEQ ID NO: 1358)
	Q18297_3/1-30	EGKTAFDIAC	EN-----	DHK	DVARAFLETG	Q----	WKNL	(SEQ ID NO: 1359)
	Q38898_1/1-30	HGVTALQVAM	AE-----	DQM	DMVNLLATNG	A----	DVVC	(SEQ ID NO: 1360)
	HRG_COWPX_2/1-3	NDFNIFTYMK	SK----	NVDI	DLIKVLVEHG	F----	DFSV	(SEQ ID NO: 1361)
	O43988/1-30	EGYTPIHYAV	RE-----	SRI	ETVKFLIKFN	S----	KLNI	(SEQ ID NO: 1362)
15	O43988_1/1-30	QGYTPLYLAA	KA-----	GKT	NFVKYLLSKG	R----	SKKI	(SEQ ID NO: 1363)
	AKR1_YEAST/1-30	EGFTPLHWGT	VK-----	GQP	HVLKYLIQDG	A----	DFFQ	(SEQ ID NO: 1364)
	Q21920_5/1-30	TKDTALTISA	EK-----	GHE	KFVRMLLNGD	A----	AVDV	(SEQ ID NO: 1365)
	P87621_3/1-32	QHKTPLYYLS	GTD---	DEVI	ERINLLVRYG	A----	KINN	(SEQ ID NO: 1366)
	O18270_1/1-30	EGRSCLMWAV	CA-----	GNI	EVINYLIQRE	D----	APKR	(SEQ ID NO: 1367)
20	VC17_VACCC/1-33	RGNNALHCYV	SNK--	CDTDI	KIVRLLLSRG	V----	ERLC	(SEQ ID NO: 1368)
	VC09_VACCC/1-30	HGCSILYHCI	KS-----	HSV	SLVEWLIDNG	A----	DINI	(SEQ ID NO: 1369)
	YIA1_YEAST/1-30	FDNSPLFLAS	LC-----	GHE	AVVKLLLQRG	A----	VCDR	(SEQ ID NO: 1370)
	Q89540_1/1-32	CGRTPLHAYV	QYD---	GVRP	EVVALMLEAG	A----	DVVC	(SEQ ID NO: 1371)
	Q25338_5/1-31	NNWTPLHFAL	YF-----	KKE	DAAKELLKQD	DI---	NLTI	(SEQ ID NO: 1372)
25	O17055_3/1-30	HLLPALHLAA	MI-----	GDS	EMLTILLNSG	A----	NIHV	(SEQ ID NO: 1373)
	O44872/1-30	LGRNALLIAI	EN-----	ENI	EMIELLLDHN	I----	ETGD	(SEQ ID NO: 1374)
	1790447/1-30	ARDTGLFMAM	QR-----	GHM	NVINTIFNAL	P----	TLFN	(SEQ ID NO: 1375)
	HRG_COWPX_3/1-3	YIISHLYSES	DSR--	SCVNP	EVVKCLINHG	I----	NPSS	(SEQ ID NO: 1376)
	Q21587_2/1-30	EKPSPIDIAV	LK-----	DDP	HLLKIVLDAG	A----	NPNA	(SEQ ID NO: 1377)
30	VB18_VARV_2/1-3	EGKTLLHIAC	EY-----	NNT	HVIDYLIRIN	G----	DINA	(SEQ ID NO: 1378)
	VB04_VACCC_1/1-	YGCTLLHRCI	YH(5)	SESYN	ELIKILLNNG	S----	DVDK	(SEQ ID NO: 1379)
	O75407/1-30	LNSTPLHWAT	RQ-----	GHL	SMVVQLMKYG	A----	DPSL	(SEQ ID NO: 1380)
	P87621_4/1-31	FGDSPLTLLI	KT----	LSPA	HLINKLLSTS	N----	VITD	(SEQ ID NO: 1381)
	O54807_1/1-31	EGRTPVHVAI	SN-----	QHS	VIIQLLISHP	NI---	ELSV	(SEQ ID NO: 1382)
35	FEM1_CAEEL_4/1-	QGVDPMLGAA	LS-----	GFL	DVLNVLADQM	PS---	GIHK	(SEQ ID NO: 1383)
	Q84566/1-33	GYRSLQYAL	SWG--	RHGNS	KVIELLLKHG	A----	NVDY	(SEQ ID NO: 1384)
	Q27105_1/1-28	HGRSPLYWAV	YY-----	KNK	EIVELLQSHG	A----	TI--	(SEQ ID NO: 1385)
	O13075/1-30	NGDNVLHLSI	IH-----	LHR	ELVKNNLEVM	P----	DMNY	(SEQ ID NO: 1386)
	G3930527_4/1-30	QRSTPLIIAA	RN-----	GHA	KVVRLLLEHY	R----	VQTQ	(SEQ ID NO: 1387)
40	G3925387_5/1-30	DGHSLLHWAA	LG-----	GNA	DVCQILINENK	I----	NPNV	(SEQ ID NO: 1388)
	O88849_6/1-30	EGRTSFMWAA	GK-----	GND	DVLRTMLSLK	S----	DIDI	(SEQ ID NO: 1389)
	O88849_7/1-30	QGATPLHYAA	QS-----	NFA	ETVKVFLQHP	S----	VKDD	(SEQ ID NO: 1390)
	O17055_4/1-30	EGNQALHYAA	KS-----	GSL	VILNMLIKQV	R----	GTND	(SEQ ID NO: 1391)
	Q18297_5/1-30	NKMTPLLLAC	VH-----	GSQ	EIIQELIKAN	S----	NVTK	(SEQ ID NO: 1392)
45	FEM1_CAEEL_5/1-	NNDTPMHILL	RAR---	EFRK	SLVRALLVRG	T----	WLFA	(SEQ ID NO: 1393)

105

	075762_4/1-30	DGCTPLHYAC	RQ-----GGP	GSVNNLLGFN	V-----SIHS	(SEQ ID NO: 1394)
	014586_1/1-30	SWATGLHLSV	LF-----GHV	ECLLVLLDHN	A-----TINC	(SEQ ID NO: 1395)
	P90784/1-30	NGNTALHLCV	IH-----DKM	DMLDAVLEAG	G-----NIRL	(SEQ ID NO: 1396)
	Q94527_1/1-30	DGDSALHVAC	QQ-----DRA	HYIRPLLGMG	C-----NPNL	(SEQ ID NO: 1397)
5	054807_2/1-31	YLQTPHMAI	AY-----NHP	DVVSIVLEQK	AN---ALHA	(SEQ ID NO: 1398)
	075762_5/1-30	NGWTALHHAS	MG-----GYT	QTMKVILDTN	L-----KCTD	(SEQ ID NO: 1399)
	157038_1/1-31	DGDTPLHLAC	IS-----GSV	DVVAALIRMA	PH---PCLL	(SEQ ID NO: 1400)
	Q02989_4/1-30	NGNTLLHLFS	ST-----GEL	EVVQFLMQNG	A-----NFRL	(SEQ ID NO: 1401)
	060736_1/1-32	QGWTVLHEAV	ST-----GDP	EMVYTVLQHR	DY--HNTSM	(SEQ ID NO: 1402)
10	ANK1_MOUSE_18/1	DGFTPLAVAL	QQ-----GHE	NVVAHLINYG	-----TKGK	(SEQ ID NO: 1403)
	061222_1/1-33	YGRTPLMYAI	MT-----NNR	SVVDAIVGDG	KLA-VVLHK	(SEQ ID NO: 1404)
	Q62422/1-31	AGSTALYWAC	HG-----GHK	DIVEVLFTQP	NV---ELNQ	(SEQ ID NO: 1405)
	Q24241_15/1-30	NGLSALHMAA	QG-----EHD	EAAHLLLDNK	A-----PVDE	(SEQ ID NO: 1406)
	YD57_SCHPO_2/1-	DKRTPLHWAC	SV-----GKV	NTIYFLLKQP	NI---KPDE	(SEQ ID NO: 1407)
15	G3927831_2/1-32	NRRTCLHYAA	YY-----GHA	NCVQAILSAA	QS--SPVAV	(SEQ ID NO: 1408)
	P70770_1/1-30	DNNNALWFAC	FG-----NHY	DLIHLLLAAK	I-----NIDN	(SEQ ID NO: 1409)
	Q94527_2/1-30	AGNTPLHVAV	KE-----EHL	SCVESFLNGV	P-----TVQL	(SEQ ID NO: 1410)
	P72763_4/1-30	QGTTALEWAI	KQ-----QNL	SMVRKLIEHH	A-----PLDL	(SEQ ID NO: 1411)
	060733_2/1-30	EGCTPLHLAC	RK-----GDG	EILVELVQYC	H-----TQMD	(SEQ ID NO: 1412)
20	KBF1_MOUSE/1-30	KSYPOVKICN	YV-----GPA	KVIVQLVTNG	K----NIHL	(SEQ ID NO: 1413)
	088849_9/1-31	ERYTPLDYAL	LG-----ERH	EVIQFMLEHG	AL---SIAA	(SEQ ID NO: 1414)
	082630_1/1-30	LGSTPLLEAI	KN-----GND	RVAALLVKEG	A-----TLNI	(SEQ ID NO: 1415)
	YAHB_ECOLI/1-30	QGKTAITLAS	LY-----QQY	ACVQALIDAG	A-----DINK	(SEQ ID NO: 1416)
	Q84644_3/1-30	REEPILHVIT	SI-----GNV	NMMKTMLDEG	A-----DINV	(SEQ ID NO: 1417)
25	157038_2/1-30	SGRTPLHIAI	EG-----CNE	DLANFLLDEC	E-----KLNL	(SEQ ID NO: 1418)
	1790447_1/1-34	YGCPGLYLAM	QN-----GHS	DIVKVILEAL	PSLAQEINI	(SEQ ID NO: 1419)
	075762_6/1-31	NMMAPLHIAV	QG-----MNN	EVMKVLEHR	TI---DVNL	(SEQ ID NO: 1420)
	LI12_CAEEL_2/1-	DENTPLMLAV	LA-----RRR	RLVAYLMKAG	A-----DPTI	(SEQ ID NO: 1421)
	G3970962_1/1-30	HGYTALMFAA	LS-----GNK	DITWVMLEAG	A-----ETDV	(SEQ ID NO: 1422)
30	VC09_VACCC_2/1-	QDLLLEYVSY	HT-----VYI	NVIKCMIDEG	A-----TLYR	(SEQ ID NO: 1423)
	083807_2/1-31	SGKTPWHWAV	RA-----LDR	DLIKHLVTLG	PP---TQER	(SEQ ID NO: 1424)
	023296/1-30	DGNTALYYAI	EG-----RYL	EMATCLVNAD	K-----DAPF	(SEQ ID NO: 1425)
	G3927831_3/1-30	DRHSVLHVAA	AN-----GQI	EILSLLLERF	T-----NPD	(SEQ ID NO: 1426)
	YB07_FOWPM_3/1-	NGYSPIKMAV	RL-----RDV	EMIKLLMSYN	TY--PDYNY	(SEQ ID NO: 1427)
35	RN5A_MOUSE_2/1-	NGFTAFMEAA	ER-----GNA	EALRFLFAKG	A-----NVNL	(SEQ ID NO: 1428)
	054910_1/1-30	HGDTALHVAC	RR-----QNL	ACACCLLEEQ	P-----EPGR	(SEQ ID NO: 1429)
	044997_4/1-32	GGYEPMHTCY	DHF---VGNA	DCIHLILYRT	S-----DPTE	(SEQ ID NO: 1430)
	000306/1-30	REQTPLFLAA	RE-----GAV	EVAQLLLGLG	A-----AREL	(SEQ ID NO: 1431)
	061222_2/1-30	SGNTAAHYAA	AY-----GFL	DCLKLLASID	D-----NILS	(SEQ ID NO: 1432)
40	074205/1-30	DQFSPLMTAI	SL-----GRL	DIARILLQSG	A-----PLDI	(SEQ ID NO: 1433)
	VB18_VARV_3/1-3	EYIKSRYMLL	KEE---DIDE	NIVSTLLDKG	I-----DPNF	(SEQ ID NO: 1434)
	P87603_1/1-33	HLITPLHSYL	RRD--ESISA	SVLKKVIELG	A-----DRNL	(SEQ ID NO: 1435)
	082630_2/1-30	DHRTPLHVAA	SE-----GFY	VLAIQLVEAS	A-----NVLA	(SEQ ID NO: 1436)
	Q25328_4/1-30	KTDTPLCYAS	EN-----GHF	TVVQYLVSNQ	A-----KVNH	(SEQ ID NO: 1437)
45	JC4356/1-30	AGQTAAMYAA	LF-----KRT	EVLKELTDKG	A-----DLSI	(SEQ ID NO: 1438)

	050999/1-30	NGNPIFTYAI	NV-----KAK	SIINYELITKE	F----NINL	(SEQ ID NO: 1439)
	Q20109_3/1-30	YSSTALMLAT	RG-----NFI	QVVELLLTRE	P----NVNV	(SEQ ID NO: 1440)
	GLP1_CAEEI_2/1-	SERSALHEAV	VN-----KDL	RILRHLLTDK	RLL-KEIDE	(SEQ ID NO: 1441)
	Q86916_1/1-33	NNLSALAHYL	SFN--KNVEP	EIVKILIDSG	S----SITE	(SEQ ID NO: 1442)
5	Q92527_1/1-30	RYNTVLHYAV	CG-----QSL	SLVEKLLEYE	A----DLEA	(SEQ ID NO: 1443)
	O90757_4/1-30	NDYRLLHSAI	TH-----ENK	KMIELLCLHG	I----NINV	(SEQ ID NO: 1444)
	I50404_2/1-32	YGNSLLHLAL	QA-----ADE	EMLRMLLAHL	GS--ATPYL	(SEQ ID NO: 1445)
	Q94400_1/1-31	GGATALQYAV	IN-----GNE	YLVEILTSHA	SI---DVNI	(SEQ ID NO: 1446)
	Q83730/1-32	DGYGPVQTYI	HSK---NVVL	DTLRELVVAG	A----TVHD	(SEQ ID NO: 1447)
10	O44997_5/1-30	NGDTPLHVAC	RF-----AQH	TVAGYVANEK	I----DVDS	(SEQ ID NO: 1448)
	Q25338_6/1-30	NKYLPIHKAI	IN-----DDL	DMVRLFLEKD	P----SLKD	(SEQ ID NO: 1449)
	O90760_3/1-30	TGLTPLDIAM	SC-----NNY	EIISLLVSHV	I----RLDY	(SEQ ID NO: 1450)
	391941_1/1-30	NGHSPLFYAL	WE-----SHV	DVLNALLQRP	L----NLPS	(SEQ ID NO: 1451)
	1790447_2/1-34	TSSHVLYHVM	AN-----GDA	DMLKIVLNAL	PLLIRTCHL	(SEQ ID NO: 1452)
15	Q02979_2/1-30	YKRTPLHYSC	QY-----GLS	EVTKLIKLM	K----EWNI	(SEQ ID NO: 1453)
	O74205_1/1-30	EGNSLLHIAV	KT-----NSL	SITRLLERY	K----SCRE	(SEQ ID NO: 1454)
	Q86916_2/1-32	LYETPLFSCV	EKD---KVVV	EILNFLIENG	S----DINF	(SEQ ID NO: 1455)
	Q94527_3/1-31	DGNALHMAV	LE-----QSV	ELLVLILDAQ	NE---NLTD	(SEQ ID NO: 1456)
	O15084_16/1-30	NWQTPHIAA	AN-----KAV	KCAEALVPLL	S----NVNV	(SEQ ID NO: 1457)
20	YG4X_YEAST_1/1-	DGRIPLHWSV	SF-----QAH	EITSFLLSKM	E----NVNL	(SEQ ID NO: 1458)
	O18270_5/1-30	NGYTALHLAA	MV-----GHE	KVCKILTNPQ	T----SIFP	(SEQ ID NO: 1459)
	Q14349_4/1-30	QQRTPLMEAV	VN-----NHL	EVARYMVQRG	G----CVYS	(SEQ ID NO: 1460)
	JQ1744_2/1-34	KFHTPLHTYV	DKNS-KGVAP	SIIQYMIDRG	V----STAR	(SEQ ID NO: 1461)
	PLU_DROME/1-32	FGQNALTAT	YA-----GHL	TLVKELLRRR	SY--KDFNL	(SEQ ID NO: 1462)
25	O54807_4/1-31	NGNNALHLAV	MH-----GRL	NNIRALLTEC	TV---DAEA	(SEQ ID NO: 1463)
	O15084_17/1-30	ATISPLHLAA	YH-----GHH	QALEVLVQSL	L----DLDV	(SEQ ID NO: 1464)
	Q93203_1/1-30	PAIYPFHLLA	KY-----NHV	ECMQALREHG	F----SINY	(SEQ ID NO: 1465)
	O88849_11/1-31	EGRTPLHFAV	AD-----GNL	TVVDVLTSYE	SC---NITS	(SEQ ID NO: 1466)
	Q17643/1-30	AMMNTLHMAV	AH-----KQR	DIVELLLKNG	Y----DPNT	(SEQ ID NO: 1467)
30	O48738_2/1-30	DEASPLYMAV	EA-----GYH	ELVLKMLESS	S----SPSI	(SEQ ID NO: 1468)
	1790447_3/1-30	QKYSAFELAF	EF-----GHR	VIAELILNTL	N----KMAE	(SEQ ID NO: 1469)
	Q02989_9/1-31	SECNPLHEAA	AY-----AHL	DLVKYFVQER	GI---NPAE	(SEQ ID NO: 1470)
	Q19995_1/1-31	FGCTPLHSAV	VH-----EHT	EIVRYIAGHY	NS---VLNA	(SEQ ID NO: 1471)
	O90760_4/1-30	DRNSLLLAV	KR-----NYI	DVVRYLVKKG	V----DINF	(SEQ ID NO: 1472)
35	Q89340_5/1-31	SGCTPLHRAV	FN-----GHD	ACASMLVNKI	VS---ERPL	(SEQ ID NO: 1473)
	O90757_5/1-31	NKYSMHLFLS	TS----NKYH	NVMAVLLDKG	I----DVNI	(SEQ ID NO: 1474)
	Q20313_1/1-30	QDSDIFLSAC	MS-----GDE	EEVEELLNKG	A----NINT	(SEQ ID NO: 1475)
	Q89202_2/1-29	KGNTALYYAV	DS-----GNM	QTVKLFVKKN	-----WRLM	(SEQ ID NO: 1476)
	YMV8_YEAST_2/1-	KGNTCVHLAL	MK-----GHE	QTLHLLQOF	P----RFIN	(SEQ ID NO: 1477)
40	157038_3/1-30	VAQTPLHLAA	LT-----AQP	NIMRILLLAG	A----EPTV	(SEQ ID NO: 1478)
	AKR1_YEAST_2/1-	QGFNLLHLSV	NSSN-IMLVL	YVLFNVVSKG	LL---DIDC	(SEQ ID NO: 1479)
	Q18297_6/1-30	REKELIHFAA	EK-----GFL	EVLKALVEAG	G----NKNE	(SEQ ID NO: 1480)
	O68219/1-30	DGRTIIHYAA	KD-----GNL	EILQALGRK	S----SYSK	(SEQ ID NO: 1481)
	Q20109_5/1-30	FGNWILTSV	RS-----GNA	AIVRMILDKF	A----DINC	(SEQ ID NO: 1482)
45	Q25328_7/1-30	DGMNFFYYAV	QN-----GHL	NIVKYAMSEK	D----KFEW	(SEQ ID NO: 1483)

	061222_4/1-30	EDNTPLHYAL	TN-----GNL	MLFNMLLDKV	A----NKRN	(SEQ ID NO: 1484)
	Q21587_5/1-30	LLKSALHFAL	LS-----GNH	DLVKFMIANG	S----NVNM	(SEQ ID NO: 1485)
	061222_5/1-30	KKRTPLIHAM	LN-----GQI	HTAAFLLAGK	A----SLTL	(SEQ ID NO: 1486)
	Q21920_13/1-30	RKISPMMAAF	RK-----GHV	EIVKYMVNSA	K----QFPN	(SEQ ID NO: 1487)
5	Q83730_1/1-30	LGNSCLDLAV	LN-----GNK	YMVHRLLRKT	I----TPDA	(SEQ ID NO: 1488)
	Q18297_7/1-30	DTKTVLHTAA	FY-----GNE	SIVRYFIAEG	V----TIDR	(SEQ ID NO: 1489)
	Q18663_1/1-31	EEFTPLHNAV	KM-----GNT	VMAKNFIDSK	SV---WIDE	(SEQ ID NO: 1490)
	Q25328_8/1-31	GLLSALHYAI	LY-----KHD	DVASFLMRSS	NV---NVNL	(SEQ ID NO: 1491)
	018152_4/1-30	YGQHPMMIAA	AN-----GNL	KVLEYLMGIS	D----IINQ	(SEQ ID NO: 1492)
10	054807_6/1-30	HRQTALHLAA	QQ-----DLP	TICSVLLENG	V----DFAA	(SEQ ID NO: 1493)
	073630_3/1-30	NGDTPLHLAV	IH-----GQS	SVIEQLVQII	L----SIPN	(SEQ ID NO: 1494)
	35040_3/1-30	RGHTPLDLTC	ST-----LVK	TLLLNAAQNT	M----EPPL	(SEQ ID NO: 1495)
	Q19995_2/1-33	DGRTPLHHAY	MK-----RRN	ELIEYLLYIC	PDSA-NIKD	(SEQ ID NO: 1496)
	DAPK_HUMAN_3/1-	DADIRLWVNG	CKL--ANRGA	ELLVLLVNHG	Q----GIEV	(SEQ ID NO: 1497)
15	Q25328_9/1-30	KGLAPLLAFS	KK-----GNL	DMVKYLFDDN	A----NVYI	(SEQ ID NO: 1498)
	043150_1/1-33	PDETALHLAV	RSV--DRTSL	HIVDFLVQNS	G----NLDK	(SEQ ID NO: 1499)
	YAR1_YEAST_1/1-	SDSTALHMAA	AN-----GHI	ETVRYILETV	S----RANS	(SEQ ID NO: 1500)
	Q84566_2/1-31	NGHTLLLWSA	EK----SKNM	ELVKYLIHAG	V----EISQ	(SEQ ID NO: 1501)
	S57237/1-30	YGKSPINDAA	EN-----QQV	ECLNVLVQHG	T----SVDY	(SEQ ID NO: 1502)
20	35040_4/1-33	NGDTPLHLAI	IHG--QTSVI	EQIVYVIHHA	Q----DLGV	(SEQ ID NO: 1503)
	Q18587_1/1-30	NGRTAIQIAA	DY-----GQT	SIIAYLISIG	A----NIQD	(SEQ ID NO: 1504)
	Q02989_10/1-30	QENTPITVAI	FA-----NKV	SILNYLVGIG	A----DPNQ	(SEQ ID NO: 1505)
	Q09493_1/1-32	IGLTPLYNYM	LTA---DSND	QVAEILLREA	A----DIGV	(SEQ ID NO: 1506)
	015084_18/1-30	RGRTPIHLSA	AC-----GHI	GVLGALLQSA	A----SMDA	(SEQ ID NO: 1507)
25	Q21920_15/1-33	KKNTPLMEAC	AGD--QGQQA	GVVKLLLSKH	A----EVDV	(SEQ ID NO: 1508)
	083807_6/1-30	DGSTPLHVMA	KR-----GYT	GFVQFLVDRK	V----NLNA	(SEQ ID NO: 1509)
	D1037943_4/1-30	YDATAMHRAA	AK-----GNL	KMVHILLFYK	A----STNI	(SEQ ID NO: 1510)
	YB07_FOWPM_4/1-	DIESELHEAV	EE-----GDV	VKVEELLDG	K----FIND	(SEQ ID NO: 1511)
	045398_1/1-30	DDRTALDIAV	EK-----RDL	KSARICIKGG	A----DVNA	(SEQ ID NO: 1512)
30	049409_1/1-30	DGGVRLMYLA	NE-----GDI	EGIKELIDSG	I----DANY	(SEQ ID NO: 1513)
	083807_8/1-30	KGETPLVLT	DR-----DHR	DLTAYFVSLG	A----DIHA	(SEQ ID NO: 1514)
	S57237_1/1-31	NDVTPVYLAA	QE-----GHL	EVLKFLVLEA	GG---SLYV	(SEQ ID NO: 1515)
	JQ1744_3/1-30	YGFTPLLSAV	YA-----VNL	NFFNHVFKLG	A----DINV	(SEQ ID NO: 1516)
	E1350345_1/1-30	NGRSCLFYAR	SN-----GFR	EVFDMLVTAG	L----SPDY	(SEQ ID NO: 1517)
35	P87611_1/1-35	IGKTALYYYY	ITRSRDKLSL	DVINCLISYE	K----EILY	(SEQ ID NO: 1518)
	Q23595_2/1-30	RGQSLHIIAC	LC-----GHE	HIVRWILNRS	G----SDAI	(SEQ ID NO: 1519)
	P90902_2/1-30	DFWQPVHAAA	CW-----AQP	DLIELLCQYG	G----DIHA	(SEQ ID NO: 1520)
	VC17_VACCC_1/1-	MCKNSLHYII	SS(4)QSLSK	DVIKCLINNN	V----SIHG	(SEQ ID NO: 1521)
	JQ1744_4/1-32	HGNSSLHYIM	RRY---MLCK	SVLDILLSTG	V----DING	(SEQ ID NO: 1522)
40	Q93203_2/1-30	KNILPIHLAA	WN-----GNL	EIVKMLLLQT	M----EKPA	(SEQ ID NO: 1523)
	045398_2/1-30	EEFTPLICAV	MA-----NNH	IMCETLIRSG	V----NCDV	(SEQ ID NO: 1524)
	O00522_1/1-31	DHWAPIHYAC	WY-----GKV	EATRILLEKG	KC---NPNL	(SEQ ID NO: 1525)
	Q02989_11/1-30	SGFTPLHLIS	SS-----TSE	TAAILIRNTN	A----VINI	(SEQ ID NO: 1526)
	YMV8_YEAST_3/1-	NGWSSLHYAS	YH-----GRY	LICVYLIQLG	H----DKHE	(SEQ ID NO: 1527)
45	Q07045_4/1-35	NSNILQHILI	EYMTFDDIDI	HLVECMLAYG	A----VVNK	(SEQ ID NO: 1528)

108

	015084_20/1-30	DGKTPLHMTA	LH-----GRF	SRSQTIIQSG	A----VIDC	(SEQ ID NO: 1529)
	Q18297_8/1-30	YQLTPLHYAA	MK-----SNF	SALHALIKLK	A----DVDA	(SEQ ID NO: 1530)
	024538_2/1-32	NGNTALWYAI	AS-----KHY	SIFRILYQLS	AL--SDPYT	(SEQ ID NO: 1531)
	055014_3/1-30	LLSTALHVAV	RT-----GHY	ECAEHLIACE	A----DLNA	(SEQ ID NO: 1532)
5	045398_3/1-33	QGASALEIAL	CSE--HEKSR	QIAAQLVEKG	A----DVNV	(SEQ ID NO: 1533)
	017055_5/1-30	VHFTALHCAT	YF-----GQE	NAVRTLISAS	A----NLNL	(SEQ ID NO: 1534)
	015084_21/1-30	LKRTPIHAAA	TN-----GHS	ECLRLLLIGNA	E----PQNA	(SEQ ID NO: 1535)
	088849_14/1-33	EGKIPLHWAA	NHK--DPSAV	HTVRCILDAA	P----TESL	(SEQ ID NO: 1536)
	Q02989_13/1-30	RNEYPFYLA	EK-----RYK	DIFDYFVSKD	A----NVNE	(SEQ ID NO: 1537)
10	023295_1/1-30	YGVSSLFVAI	NT-----GDV	SLVKAILKII	G----NKDL	(SEQ ID NO: 1538)
	Q14349_5/1-30	EENICLHWAS	FT-----GSA	AIAEVLLNAR	C----DLHA	(SEQ ID NO: 1539)
	Q12013_1/1-30	RGFNALHCAL	VG-----GDQ	RVICDLILSG	A----NFYE	(SEQ ID NO: 1540)
	Q40785_2/1-30	DDESIVHHTA	SV-----GDA	EGLKKALDGG	A----DKDE	(SEQ ID NO: 1541)
	DAPK_HUMAN_5/1-	GGSSNAVYWAA	RH-----GHV	DTLKFLSENK	C----PLDV	(SEQ ID NO: 1542)
15	018270_6/1-34	RDRSPLHYAA	CK-----VNL	EALRILLFVD	PNGGPDPFGF	(SEQ ID NO: 1543)
	Q86916_3/1-30	RYCNPLFYV	EQ-----DDI	NGVKKWIKFV	N----DCND	(SEQ ID NO: 1544)
	AKR_ARATH_2/1-3	GGLTALHRAI	IG-----KKQ	AITNYLLRES	A----NPFV	(SEQ ID NO: 1545)
	016004_2/1-31	MGETPLHAAV	RA-----DAI	EVFRLLIQNR	ST---QIDA	(SEQ ID NO: 1546)
	016004_3/1-41	DGLTPLMLAV	MR(10)ETHA	HFIEDLITRG	G----DINH	(SEQ ID NO: 1547)
20	YA2A_SCHPO_4/1-	QGFNCLHLAV	HA-----ASP	LLVVYLLHLD	I----SVDL	(SEQ ID NO: 1548)
	Q18297_9/1-30	EKKTPLRMAV	EG-----NHP	ETLKKILQME	K----KNSC	(SEQ ID NO: 1549)
	035433_1/1-31	KGLTPLALAA	SS-----GKI	GVLAYILQRE	IH---EPEC	(SEQ ID NO: 1550)
	054807_8/1-30	KGNPPLWLAL	AS-----NLE	DIASLTVRHG	C----DATC	(SEQ ID NO: 1551)
	P87603_3/1-40	TYLSRLCIVN	KN(8)NNINI	EVVKYLIRSG	S----DISL	(SEQ ID NO: 1552)
25	073579_3/1-30	SLPIQYYWSC	ST-----IDI	EIVKLLIKDV	D----TCRV	(SEQ ID NO: 1553)
	Q21587_6/1-33	RLKSPFVEYF	RSN--DSIDA	KVVKLFITHG	A----KVVM	(SEQ ID NO: 1554)
	A53950_3/1-30	FDKSAFDIAM	EK-----NNT	EILVMLQEAM	Q----NQVN	(SEQ ID NO: 1555)
	Q18663_3/1-31	SGRTPFDLAC	EY-----GQE	KMLERLFSCG	LR---KINF	(SEQ ID NO: 1556)
	Q25328_11/1-31	NNITALHYAA	IL-----GYL	ETTKQLINLK	EI---NANV	(SEQ ID NO: 1557)
30	072760_2/1-36	EYKNILIHVI	YS(4)KPINY	YVLYKLLRKG	A----DPNY	(SEQ ID NO: 1558)
	Q89342_2/1-30	DGSTPLHFIA	RW-----GRK	ICARELITAG	V----EINT	(SEQ ID NO: 1559)
	083807_9/1-30	MGSSALVLAI	KK-----DRS	DLCHELLALG	A----DLFI	(SEQ ID NO: 1560)
	Q83730_2/1-33	DGRYPLLCLL	END--RINTA	RFVRYMIDRG	T----SVYV	(SEQ ID NO: 1561)
	004242_3/1-32	RLDLPVTLCF	AVN---KGDD	FMLHQLLKRG	L----DPNE	(SEQ ID NO: 1562)
35	082490_1/1-30	YGLSPLHLAI	EE-----GQT	RLVLSLLKVD	S----DLVR	(SEQ ID NO: 1563)
	045398_4/1-30	DQFTPAHMAV	SW-----AQN	DVLRALRDHS	A----NLCD	(SEQ ID NO: 1564)
	P87600_4/1-35	NGVDPLMLTM	ENNMLSGHQW	YLVKNILDKR	P----NVNM	(SEQ ID NO: 1565)
	073560_2/1-48	DGLTSLHYVC	KH(17)RAEK	RFIYTIIDHG	A----NINA	(SEQ ID NO: 1566)
	P87603_5/1-33	NIPEILYLYI	QMA--QYVDL	DVIKHFLNNS	V----ILDY	(SEQ ID NO: 1567)
40	Q14678_3/1-29	DGSTALSIAL	EA-----GHK	DIAVLLYAHV	-----NFAK	(SEQ ID NO: 1568)
	Q93318_1/1-38	HGDTAVDLLC	SS(6)NNVRE	NLYIRLLTSG	A----VPNK	(SEQ ID NO: 1569)
	G3930525_6/1-30	YGMTPLLAAS	VT-----GHT	NIVEYLIQEQ	P----GHEQ	(SEQ ID NO: 1570)
	YB07_FOWPM_5/1-	SKRPCVTAMC	YAI--QNNKI	DMVSMFLKRG	A----DSNI	(SEQ ID NO: 1571)
	JQ1744_5/1-31	FGDTCVTLAV	AT-----NHV	YFVTAVLQHK	PS---EQTI	(SEQ ID NO: 1572)
45	VB04_VACCC_2/1-	YGENILHIYS	MD----DANT	NIIIFFLDRV	L----NINK	(SEQ ID NO: 1573)

	O41154_3/1-35	YGKTCIEMAA YY-----GHQ	DILKMMNGKL GI(4) IIND	(SEQ ID NO: 1574)
	P87600_5/1-36	GDMDILYTYF NSPRTRCIKL	DLIKYMVDVG IV---NLNY	(SEQ ID NO: 1575)
	Q21920_17/1-31	DGVNCFMEVA RH-----GSI	DLMSLLVEFT KG---NMPM	(SEQ ID NO: 1576)
	O44997_6/1-30	NGATAMHCAA KY-----GHA	EVFNYFHMKG G----NICA	(SEQ ID NO: 1577)
5	P72763_6/1-30	EGKTPLMLAA MA-----NLA	AVITTIANYS V----QVNR	(SEQ ID NO: 1578)
	Q09493_2/1-31	QGETPLTLAA GI-----PNNR	AVIVSLIGGG A----HVDF	(SEQ ID NO: 1579)
	Q24145_1/1-30	FGCQPLHYAA RS-----KTA	SFIRTLISAQ A----NVQG	(SEQ ID NO: 1580)
	Q17583_2/1-31	NGVSPLMYAT TV-----RSI	NSAKHLVLHR DS---DVNM	(SEQ ID NO: 1581)
	O90757_6/1-30	HVKAPIHVAV ER-----NNI	YGTMLLINRN A----DVNI	(SEQ ID NO: 1582)
10	MBP1_YEAST_1/1-	ELHTAFHWAC SM-----GNL	PIAEALYEAG T----SIRS	(SEQ ID NO: 1583)
	O83807_10/1-30	GGNTPLHLAC EW-----KLT	QAINGILRKG A----EIEA	(SEQ ID NO: 1584)
	G4103857_2/1-32	GLLTPLHLAA GN-----RDSR	DTLELLLMNR YI---KPEL	(SEQ ID NO: 1585)
	O18270_7/1-30	ESVTPLHIAA TK-----QDT	KILKIFVEIL K----TPKV	(SEQ ID NO: 1586)
	O68219_2/1-31	GKKTTLTAEA LT-----SGKY	GVVKALIKNS A----DVNA	(SEQ ID NO: 1587)
15	Q19995_4/1-30	TGSVALILAV LN-----GHK	SIVEAILGYS L----NTLN	(SEQ ID NO: 1588)
	O45398_5/1-29	DGRTPAHIGV RE-----QNV	DGVKILLDAE -----NVEF	(SEQ ID NO: 1589)
	Q12013_3/1-30	NNRTPLLWAA YQ-----GDF	LTVELLLKFG S----TVAW	(SEQ ID NO: 1590)
	O18152_5/1-31	MGQTPLMLAA SQ-----GQL	EAVQFLTCKA KA---DVDA	(SEQ ID NO: 1591)
	Q89540_4/1-32	YGRITLHHLA RTA---KISE	GMVRTLTGLG V----DPAA	(SEQ ID NO: 1592)
20	Q17643_1/1-30	VDQTPLGVAV RS-----QSS	ELIALLIAYG G----DVNL	(SEQ ID NO: 1593)
	Q63618_3/1-31	DGMTPLHAAA QM-----GHN	PVLVWLVSFA DV---SFEQ	(SEQ ID NO: 1594)
	O14586_2/1-30	DFKSPLHCAA WN-----CDH	VLHMMLEAG A----EANL	(SEQ ID NO: 1595)
	Q17583_4/1-31	GNLTALSMAI VC----TPGV	YMVRTLIDLG A----SVNK	(SEQ ID NO: 1596)
	O68219_3/1-32	TGTPALHLAT AA-----GNH	RTAMLLLDKG AP--ATQRD	(SEQ ID NO: 1597)
25	Q93318_2/1-31	DGDTPLHIVA AH-----NDL	GKIYALCETL RK---TMNE	(SEQ ID NO: 1598)
	O54807_9/1-30	DGETALQLAI KH-----QLP	LVVDAICTRG A----DMSV	(SEQ ID NO: 1599)
	Q01317_4/1-30	DNFTPLVHSI VR-----NHL	ECVGRLLERS A----RIDP	(SEQ ID NO: 1600)
	O82490_2/1-30	FINTPLHIAS AS-----GNL	SFAMELMNLK P----SFAR	(SEQ ID NO: 1601)
	YAHB_ECOLI_2/1-	TCLNPFLISC LN-----DDL	TLLRIILPAK P----DLNC	(SEQ ID NO: 1602)
30	Q63618_4/1-30	LDALPVHAAA RS-----GKL	HCLRYLVEEV A----LPAV	(SEQ ID NO: 1603)
	Q89440_2/1-44	MEMYPRHRYK KH(13)DLDM	NVVKELLSNG A----SLTI	(SEQ ID NO: 1604)
	P87601_1/1-38	DGYTCLDVAV DR(6)REAH	L KILEILLREP L----SIDC	(SEQ ID NO: 1605)
	O18152_8/1-31	HGRTCVHVVA ST-----NDV	AAAKMLKRIG GV---NFEA	(SEQ ID NO: 1606)
	1AP7_1/1-30	TGSLPIHLAI RE-----GHS	SVVSFLAPES D----LHHR	(SEQ ID NO: 1607)
35	O17055_6/1-32	FSETPLHAAC TG----GKSI	ELVSFLMKYP GV---DPNY	(SEQ ID NO: 1608)
	Q18297_10/1-32	RLNTVFHIVA LR-----GEP	EYLEMMMDHD PV--EAIKA	(SEQ ID NO: 1609)
	157038_4/1-32	HGNTALHLSC IA-----GEK	QCVRALTEKF GA--TEIHE	(SEQ ID NO: 1610)
	O75762_9/1-30	DKKSPLHFAA SY-----GRI	NTCQRLQDI S----DTRL	(SEQ ID NO: 1611)
	O23296_4/1-31	TGDSILHIAA KW-----GHL	ELVKEIIFEC PC---LLFE	(SEQ ID NO: 1612)
40	TRI9_HUMAN_3/1-	DGDTALHLAV IH-----QHE	PFLDFLLGFS A----GTEY	(SEQ ID NO: 1613)
	Q23595_3/1-32	NGETAAHISA AH-----GDM	RALEILLGGS LK--STMNC	(SEQ ID NO: 1614)
	VC17_VACCC_2/1-	GGSLPIQYYW SFS---TIDI	EIVKLLLIKD V----DTCR	(SEQ ID NO: 1615)
	Q18297_11/1-30	YQKTPLQVAV DS-----GKL	ETCQRLVAKG A----QIES	(SEQ ID NO: 1616)
	NTC4_MOUSE_4/1-	AGRTPLHTAV AA----DARE	VCQLLLASRQ T----SVDA	(SEQ ID NO: 1617)
45	O54807_10/1-30	KGRNFLHVAV QN-----SDI	ESVLFLLISVQ A----NVNS	(SEQ ID NO: 1618)

	O68219_4/1-31	MGDTALHEAL	YSD---NVTE	KCFLKMLKES	R-----KHL	(SEQ ID NO: 1619)
	O18270_8/1-32	VGR TAVFWAC	MG-----GQA	HTLHCMIKEL	GF--EWRTS	(SEQ ID NO: 1620)
	O18152_9/1-31	DGAVALHYAV	TH-----DHV	ELVKHLTNQH	TV---EAKD	(SEQ ID NO: 1621)
	Q18297_12/1-30	DERTPVFIGA	KF-----NAL	SSVEYILDHL	R----KKNK	(SEQ ID NO: 1622)
5	VC09_VACCC_4/1-	DGETPLKAYV	TKKN-NNIKN	DVVILLSSV	D----YKNI	(SEQ ID NO: 1623)
	1AP7_2/1-29	FGKTALQVMM	F-----GSP	AVALELLKQG	A----SPNV	(SEQ ID NO: 1624)
	P87600_6/1-35	LGYTPLTSYI	CTAQ-NYMY	DIIDCLISNK	VL---NMVK	(SEQ ID NO: 1625)
	O75762_10/1-30	CHETMLHRAS	LF-----DHH	ELADYLISVG	A----DINK	(SEQ ID NO: 1626)
	P87603_6/1-35	YSSNIFSUYF	KAHNTIGIDI	KLVKWMITKG	V----DINY	(SEQ ID NO: 1627)
10	O48738_5/1-30	MGETTLHVAA	RA-----GSL	NIVEILVRFI	T----ESSS	(SEQ ID NO: 1628)
	P87611_2/1-30	DFDIFEYIES	DN-----IDV	ELLRLLIAGK	L----EINS	(SEQ ID NO: 1629)
	Q25328_13/1-31	RGTTPLTAV	AE-----NAL	HIAEYLIREK	RQ---DINI	(SEQ ID NO: 1630)
	Q02979_3/1-31	KSSSLHLAT	EW-----NNY	PLHLVLLSSK	RF---DINY	(SEQ ID NO: 1631)
	Q01317_5/1-30	KDLPAMYAA	WE-----GHL	ECMKLLTPAK	K----EKAA	(SEQ ID NO: 1632)
15	Q17343_21/1-30	EGDEPEHHYT	ET-----TTT	TVTKEVIDDS	Q----EMGD	(SEQ ID NO: 1633)
	Q19995_6/1-31	DKIDGIIKSI	KE-----GNL	DKVKELMKTK	KL---AIAR	(SEQ ID NO: 1634)
	P90902_3/1-30	DGSTYLHVAS	AN-----GY	DVAAFLLTCS	V----SPLI	(SEQ ID NO: 1635)
	O68219_5/1-34	YEPHPQHVA	EAV--RTGAV	GVLEHLITTE	VI---SVNE	(SEQ ID NO: 1636)
	Q83730_3/1-30	LDFTPINYCV	IH-----NDR	RTFDYLLERG	A----DPNV	(SEQ ID NO: 1637)
20	E1344043_2/1-30	FGYNALVCAV	KA-----KAM	DCIIAIRDAG	G----FIDA	(SEQ ID NO: 1638)
	O45398_8/1-31	YGENAMLCV	RS-----GSL	DCIRAVADSS	RT---NRYA	(SEQ ID NO: 1639)
	Q12013_4/1-31	YVVDIFIEAA	KD-----GDL	KVVKDVVESG	AV---DINN	(SEQ ID NO: 1640)
	O24382_4/1-31	GDVGHFACVA	VE----QNNL	SLLKEIVRYG	G----DVTL	(SEQ ID NO: 1641)
	4151809_3/1-30	DGMTALHKAA	CA-----RHC	LALTALLDLG	G----SPNY	(SEQ ID NO: 1642)
25	O73579_4/1-36	DESIVQAMLI	NY(4)DMVSI	PIVRCMLDNG	A----TMDK	(SEQ ID NO: 1643)
	O15084_23/1-32	NAFSPLHCAV	IN-----DNE	GAAEMLIDTL	GA--SIVNA	(SEQ ID NO: 1644)
	Q20109_8/1-31	NGESLLTVAV	RS-----GNT	AVAKQLAQLD	PD---AIDE	(SEQ ID NO: 1645)
	O04703_2/1-33	IRTSSLIEAM	KSR--QEDNV	TMMKNFLQHH	K----KLKD	(SEQ ID NO: 1646)
	Q18970_1/1-31	SGMTPAMMCA	KR----SFTM	FPLRLIVRAG	A----DLSL	(SEQ ID NO: 1647)
30	Q18663_4/1-30	DGISACHIA	KD-----GMI	DHFNLIIYH	A----DICL	(SEQ ID NO: 1648)
	O72755_3/1-40	DGMIPVVYCI	HS(7)NITNI	KIIRKLLNLS	RH---APHN	(SEQ ID NO: 1649)
	Q25338_10/1-30	NDRSAMHAVA	YR-----GNN	KIALRFLLN	Q----SIDI	(SEQ ID NO: 1650)
	O54807_12/1-30	EGNTVLLLAY	MK-----GNA	NLCRAIVRSG	V----RLGV	(SEQ ID NO: 1651)
	O45398_10/1-30	RGQTACHVAA	EA-----GAP	EILSKIVEAR	R----GLKW	(SEQ ID NO: 1652)
35	Q21920_18/1-31	NCNTALIIYAA	ST----DGRD	VVREILMTEG	P----KKPD	(SEQ ID NO: 1653)
	Q21587_7/1-29	QLRNVKLAA	LQ-----NQP	EVLELLLGLG	-----EQYD	(SEQ ID NO: 1654)
	VC09_VACCC_5/1-	NGINIVEKYA	TTS---NPNV	DVFKLLLDKG	IP---TCSN	(SEQ ID NO: 1655)
	VC17_VACCC_3/1-	SNINDFDLSS	DN-----IDL	RLKYLIVDK	RI(4)NTNY	(SEQ ID NO: 1656)
	VB04_VACCC_3/1-	SGCTCISEAV	AN-----NNK	IIMEVLLSKR	P----SLKI	(SEQ ID NO: 1657)
40	Q18970_2/1-33	QGYTPIHLAI	QG-----NHV	PLVAYFLLKF	EYA-KDISD	(SEQ ID NO: 1658)
	Q23859_3/1-30	NGFSALHQAV	SS-----DDR	ILMRGLQYEN	I----NVDV	(SEQ ID NO: 1659)
	Q25328_14/1-30	YDKTALDIAI	DA-----KFS	NIVEYLKTKS	G----KFRR	(SEQ ID NO: 1660)
	Q19995_7/1-30	YGMTPIHKAL	LH-----GQT	NTVRFLGRF	P----SCVN	(SEQ ID NO: 1661)
	VB04_VACCC_4/1-	YNETSIYDAV	SY-----NAY	NTLVYLLNRN	G----DFET	(SEQ ID NO: 1662)
45	O88202_1/1-31	VLLPGLALAA	AH----AGDL	DTLQAFVELG	R----DLNL	(SEQ ID NO: 1663)

	E1350208_1/1-30	ELPNPLHEAA	RR-----GNM	DMLAECLRER	V----SVNS	(SEQ ID NO: 1664)
	O54807_13/1-32	ALYSPKKYSA	DVM---SEMA	QIAEALLQAG	A----NPNM	(SEQ ID NO: 1665)
	Q23859_4/1-34	QQQHQLFSAV	EQ-----NDV	TKVKKLSSKK	KISKSNLTS	(SEQ ID NO: 1666)
	O04242_4/1-31	-GDVALYSCV	AVE---ENDP	ELLENIIRYG	G----NVNS	(SEQ ID NO: 1667)
5	Q93203_3/1-30	SDFTLLHHAV	LE-----NQN	QISKLLMDYE	P----LLLL	(SEQ ID NO: 1668)
	Q17583_5/1-30	LHITSLQIAS	LL-----RVD	DIIPLLQRR	A----DVTV	(SEQ ID NO: 1669)
	YIA1_YEAST_1/1-	KNFEELCYSC	RT-----GDM	DNVDRLISTG	V----NVNS	(SEQ ID NO: 1670)
	O35433_2/1-30	FGELPLSLAA	CT-----NQL	AIVKFLQNS	W----QPAD	(SEQ ID NO: 1671)
	O83515_2/1-30	CFEENFIATV	MD-----GNI	DIVNLFIDAG	F----SAAL	(SEQ ID NO: 1672)
10	O23295_4/1-30	NGWTCLSLAA	HI-----GYG	EGVCNLLERS	T----KAAE	(SEQ ID NO: 1673)
	AKR_ARATH_3/1-3	KKWLPLHTLA	AC-----GEF	YLVDSSLKHN	L----DINA	(SEQ ID NO: 1674)
	Q83730_4/1-35	SRTSPLCTVL	SNKDLGNEAE	ALAKQLIDAG	A----DVNA	(SEQ ID NO: 1675)
	O72760_5/1-43	RGNTFLHYFC	IY(12)HREK	KFIKELVKYG	A----DINK	(SEQ ID NO: 1676)
	Q63618_7/1-32	NGATPAHDAA	AT-----GYL	SCLQWLLTQG	GC--RVQEK	(SEQ ID NO: 1677)
15	YG4X_YEAST_4/1-	FNQIPLHRAA	SV-----GSL	KLIELLCGLG	KS---AVNW	(SEQ ID NO: 1678)
	G3786431_4/1-34	EGSTCLHWAA	RC-----GSS	ECVSTILNFP	FPSEFIIEI	(SEQ ID NO: 1679)
	O73579_5/1-32	EGLTPLGAYS	KHR---HVKY	QIVHLLISSY	S----NSSN	(SEQ ID NO: 1680)
	Q01317_6/1-32	ERVTRTFLAS	IN---EGSL	EALKVLLDTG	LV---DIQS	(SEQ ID NO: 1681)
	O73579_6/1-33	DNNYPLHDYF	VNN--NLVDV	NVVRFIVENN	G----HMAV	(SEQ ID NO: 1682)
20	O83807_13/1-31	NQETPLFSAV	KS-----DAA	EVISILLHPQ	AG---NPAL	(SEQ ID NO: 1683)
	LI12_CAEEL_4/1-	SERSALHQAA	AN-----RDF	GMMVYMLNST	KLK-GDIEE	(SEQ ID NO: 1684)
	Q89202_3/1-30	LFIPDNKLAI	DN-----KDI	EMLQALFKYD	I----NIYS	(SEQ ID NO: 1685)
	Q24241_22/1-34	DGNTALHIAS	NLG--YVTVM	ESLKIVTSTS	VI---NSNI	(SEQ ID NO: 1686)
	Q93318_4/1-31	VGDSPLHFAT	AR-----GMN	NMVEALLSKR	EI---RVNE	(SEQ ID NO: 1687)
25	Q25328_17/1-32	HGRTVFHAAA	KS-----GND	KIMFGLTFLA	KST--ELNQ	(SEQ ID NO: 1688)
	P70770_2/1-30	MAIVALMKAT	RE-----GVD	AVVKELIDLG	V----DQAR	(SEQ ID NO: 1689)
	VB04_VACCC_5/1-	HGISLIKLYL	ES(4)EIDNE	HIVRHLIIFD	AV--ESLDY	(SEQ ID NO: 1690)
	VB04_VACCC_6/1-	YGNTPFILLC	KHD---INNV	ELFEICLENA	NI--DSVDF	(SEQ ID NO: 1691)
	O04704_2/1-34	YGDPMYVNL	LTVA-STGNA	TFLEELLKAK	L----DPDI	(SEQ ID NO: 1692)
30	O45398_12/1-31	DSKTPFFNEI	VS----RKDV	ASVEALLAAN	V----DCHV	(SEQ ID NO: 1693)
	TRPL_DROME_2/1-	KSYTRFWGLL	MF(4)VINVI	VLLNLLIAMM	S----NSYA	(SEQ ID NO: 1694)
	JQ1744_6/1-32	YKRQVVEDYL	RKR---YVKA	DVLKTILDSG	I----RLSK	(SEQ ID NO: 1695)
	Q84566_5/1-31	IRHNILFYHV	IR-----DNY	DVVKFIIDNK	LV---DINE	(SEQ ID NO: 1696)
	O49409_2/1-30	WGSTPFADAI	FY-----KNI	DVIKILEIHG	A----KHPM	(SEQ ID NO: 1697)
35	JQ1744_7/1-35	SGHSAMMSYL	INNPDATVHP	EYVSVMVDCG	A----NINQ	(SEQ ID NO: 1698)
	Q02979_4/1-30	GGWTPMEHAV	LR-----GHL	HIADMVQIRD	E----LVTH	(SEQ ID NO: 1699)
	O24538_3/1-27	---NLLCLAA	KR-----NDL	TVMNELLKQG	L----NIDS	(SEQ ID NO: 1700)
	Q89202_4/1-30	GWKTSFYHAV	ML-----NDV	SIVSYFLSEI	P----STFD	(SEQ ID NO: 1701)
	O83807_14/1-30	SGSTPLMVAV	TE-----NCF	ESVRMLIARD	A----SLFS	(SEQ ID NO: 1702)
40	G3786431_5/1-31	ECRTAMYLA	AE-----GHL	EVVKAMTDFK	CT---SIDG	(SEQ ID NO: 1703)
	O62398_2/1-27	TGNTVIHCAI	-----NK	KCLILLMEKF	R----DQTD	(SEQ ID NO: 1704)
	Q25338_11/1-30	NHMAPIHFAA	SM-----GSI	KMLRYLISIK	D----KVSI	(SEQ ID NO: 1705)
	Q23595_4/1-31	CKMLPVHVAA	AQ-----GNI	EFLRAAIKFD	NQ---MVNA	(SEQ ID NO: 1706)
	Q25328_18/1-30	NGQMPIHGAA	MT-----GLL	DVAQAIISID	A----TVVD	(SEQ ID NO: 1707)
45	Q28282_5/1-31	KYDDRLMKAA	ER----GDVE	KVSSILAKKG	I----NPGK	(SEQ ID NO: 1708)

	Q94527_4/1-29	NRDTLLHEVI	SH-----KK	DKLKLAIQTI	Q----VMNY	(SEQ ID NO: 1709)
	AKR1_YEAST_5/1-	PLLTRYHTAC	QR-----GDL	ATVKEMIHGK	LL---EVNN	(SEQ ID NO: 1710)
	391941_4/1-31	SSDLDLHLTL	DD-----IDE	KIIHQRLQEN	KA---AFFF	(SEQ ID NO: 1711)
	Q23595_5/1-30	DRANAIHCAA	YS-----GSV	PVLSHLLNAF	S-----KKKR	(SEQ ID NO: 1712)
5	P87603_8/1-29	SGMTAFHVAV	C-----LRK	DAMKYLLSIG	Y----SIDK	(SEQ ID NO: 1713)
	O72755_4/1-42	QMRTPLHKYL	CK(11)YYNE	KIIDAFIELG	A----DLTI	(SEQ ID NO: 1714)
	O18270_9/1-31	IGATPAHYAA	Q-----FSV	ECLKILFAES	KI--TEVND	(SEQ ID NO: 1715)
	Q90623_5/1-31	EGDTPLDIAE	EE-----AME	ELLQNEVNRQ	GV---DIEA	(SEQ ID NO: 1716)
	O61222_7/1-30	QTETPLHVAA	RA----GRAV	NCTFLMKEML	-----DLEK	(SEQ ID NO: 1717)
10	JQ1744_9/1-33	RGHTPLLVIY	TMG--LFIES	DAITCMIDHG	A----DPLV	(SEQ ID NO: 1718)
	P87621_5/1-32	NNYHILHAYC	GIK---GLDE	RFVEELLHIG	Y----SPNE	(SEQ ID NO: 1719)
	O16229_1/1-32	DGKNAILDAV	RE----RNKP	NLLKALQNGA	FV---DVYN	(SEQ ID NO: 1720)
	VB18_VARV_5/1-3	NNKHAIQLII	DNKENSQYTI	DCLLYILRYI	V----DKNV	(SEQ ID NO: 1721)
	O90757_7/1-30	KHFNMLRKAV	LN-----HDH	NLVNIFIDKN	F----NINI	(SEQ ID NO: 1722)
15	Q93318_5/1-32	TGKTIVHHAV	DKM--DVELL	DFLKTVVNED	-----TFTE	(SEQ ID NO: 1723)
	Q21920_19/1-31	ERDSALTLSA	QK-----GHI	KIVTAIMDYY	EK---NPPQ	(SEQ ID NO: 1724)
	Q02989_18/1-31	ERKTFFDLAI	EN-----GRL	NIVAFAVEKN	KV---NLQA	(SEQ ID NO: 1725)
	A55839_4/1-31	GGRTPLGSAL	LR-----PNP	ILARLLRAHG	AP---EPED	(SEQ ID NO: 1726)
	O60733_5/1-31	KGETVFHYAV	QG-----DNS	QVLQLLGRNA	VA---GLNQ	(SEQ ID NO: 1727)
20	O83807_16/1-31	AGESILHYAA	KV-----ADE	KTLQGLLAMN	RF---GKFL	(SEQ ID NO: 1728)
	O75762_11/1-30	QQASFLHLAL	HN-----KRK	EVVLTIIIRSK	R----WDEC	(SEQ ID NO: 1729)
	391941_5/1-30	VQFDPLNVAC	KF-----NNH	DAAKLLEIR	S----KQNA	(SEQ ID NO: 1730)
	Q94447_2/1-30	KSFTTRFWALL	MF-----GSY	SVINIIVLLN	M----LIAM	(SEQ ID NO: 1731)
	Q18104_2/1-33	TIFEVLAWCN	ETV--KQNDF	DVMETALDSL	N----NLNL	(SEQ ID NO: 1732)
25	O48738_6/1-31	SGKSVIHAAM	KA-----NRR	DILGIVLRQD	PG---LIEL	(SEQ ID NO: 1733)
	O61222_8/1-46	FNRTALHYAF	GN(13)SDPI	AVVSLSSLI	RP--EQIEI	(SEQ ID NO: 1734)
	O73579_8/1-30	RFNNGGYHCY	ET-----ILI	DVFDILSKYM	D----DIDM	(SEQ ID NO: 1735)
	O23296_6/1-31	QGNKHLAHVA	LK----AKSI	GVLDVILDEY	P----SLMD	(SEQ ID NO: 1736)
	Q09493_4/1-30	HGNHEIHQAC	KN-----GLT	KHVEHLLYFG	G----QIDA	(SEQ ID NO: 1737)
30	Q83730_5/1-34	YGFNVLQCYM	IAHV-RSSNV	QILRFLLRHG	V----DSSR	(SEQ ID NO: 1738)
	Q91974_4/1-34	DGDTFLHLAI	IHEE-KALSL	EVIRQAAGDA	A----FLNF	(SEQ ID NO: 1739)
	O61240_6/1-32	MDRLPRDIAN	ER-----LHT	DIVQLLDEYN	LV--RSPNL	(SEQ ID NO: 1740)
	Consensus/60%	psp*sLabAs	pp.....spb	chlcbLlpps	s.....shsh	(SEQ ID NO: 1741)

	IMO2HUMANb	PND-KIQAVI	DAG--VCRRL	VELLM-----	-HNDYKVVSP	ALRAVGNIVT	(SEQ ID NO: 1742)
	MMU34228c	PND-KIQAVI	DAG--VCRRL	VELLM-----	-HNDYKVVSP	ALRAVGNIVT	(SEQ ID NO: 1742)
	cATAF00130	SND-KIQAVI	EAG--VVPRL	IQLLG-----	-HSSPSVLIP	ALRTIGNIVT	(SEQ ID NO: 1743)
5	cATKAPAPRO	SND-KIQAVI	EAG--VVPRL	IQLLG-----	-HSSPSVLIP	ALRTIGNIVT	(SEQ ID NO: 1743)
	ATU69533d	TND-KIQTVI	QAG--VVPKL	VELLL-----	-HHSPSVLIP	ALRTVGNIVT	(SEQ ID NO: 1744)
	AB002533c	GNE-QIQMVI	DSG--IVPHL	VPLLS-----	-HQEVKVQTA	ALRAVGNIVT	(SEQ ID NO: 1745)
	HSSRP11a	GNE-QIQMVI	DSG--IVPHL	VPLLS-----	-HQEVKVQTA	ALRAVGNIVT	(SEQ ID NO: 1745)
	HSSRP1Bc	GNE-QIQMVI	DSG--VVPFL	VPLLS-----	-HQEVKVQTA	ALRAVGNIVT	(SEQ ID NO: 1746)
10	CELF32E10c	GNE-HIQMVI	EAQ--VVTHL	VPLLG-----	-HVDVKVQTA	ALRAVGNIVT	(SEQ ID NO: 1747)
	SRP1YEASTd	PQE-AIQAVI	DVR--IPKRL	VELLS-----	-HESTLVQTP	ALRAVGNIVT	(SEQ ID NO: 1748)
	IMO1HUMANc	PNE-RIGMVV	KTG--VVPQL	VKLLG-----	-ASELPIVTP	ALRAIGNIVT	(SEQ ID NO: 1749)
	bATAF00130	TSE-NTNVII	ESG--AVPIF	IQLLS-----	-SASEDVREQ	AVWALGNVAG	(SEQ ID NO: 1750)
	bATKAPAPRO	TSE-NTNVII	ESG--AVPIF	IQLLS-----	-SASEDVREQ	AVWALGNVAG	(SEQ ID NO: 1750)
15	SLU96718b	TSD-HTAVVI	EQG--AVPIF	VQLLS-----	-SPSDDVREQ	AVWALGNVAG	(SEQ ID NO: 1751)
	ATU69533b	TSD-HTKVVI	DHN--AVPIF	VQLLA-----	-SPSDDVREQ	AVWALGNVAG	(SEQ ID NO: 1752)
	IMO2HUMANc	NSL-QTRIVI	QAR--AVPIF	IELLS-----	-SEFEDVQEQ	AVWALGNIAG	(SEQ ID NO: 1753)
	MMU34228a	NSL-QTRNVI	QAG--AVPIF	IELLS-----	-SEFEDVQEQ	AVWALGNIAG	(SEQ ID NO: 1754)
	AB002533b	TSE-QTQAVV	QSN--AVPLF	LRLHH-----	-SPHQNVCEQ	AVWALGNIIG	(SEQ ID NO: 1755)
20	HSSRP1Bb	TSA-QTQAVV	QSN--AVPLF	LRLLR-----	-SPHQNVCEQ	AVWALGNIIG	(SEQ ID NO: 1756)
	CELF32E10b	TSE-QTQAVV	NAG--AVPLF	LQLLS-----	-CGNLNVCEQ	SVWALGNIIG	(SEQ ID NO: 1757)
	IMO1HUMANa	TSE-QTKAVV	DGG--AIPAF	ISLLA-----	-SPHAHISEQ	AVWALGNIAG	(SEQ ID NO: 1758)
	SRP1YEASTb	TSA-QTKVVV	DAD--AVPLF	IQLLY-----	-TGSVEVKEQ	AIWALGNVAG	(SEQ ID NO: 1759)
	CEF53B24	SQE-VTHLFV	NSN--CLDIL	IKLVR-----	SQ-NARLSSQ	TVWAIANIAA	(SEQ ID NO: 1760)
25	CELF26B13b	STE-QTIAAV	EAG--VTIPL	IHLVS-----	HQ-SAQISEQ	ALWAVANIAG	(SEQ ID NO: 1761)
	AB002533a	RNP-PIDDLI	KSG--ILPIL	VHCLE-----	RDDNPSLQFE	AAWALTNIAS	(SEQ ID NO: 1762)
	HSSRP1Ba	RNP-PIDDLI	KSG--ILPIL	VHCLE-----	RDDNPSLQFE	AAWALTNIAS	(SEQ ID NO: 1763)
	CELF32E10a	RNP-PIDDLI	GSG--ILPVL	VQCLS-----	-STDPNLQFE	AAWALTNIAS	(SEQ ID NO: 1764)
	CELF26B13a	KNP-PIDEVI	HCG--LLQAL	VQALS-----	-VENERVQYE	AAWALTNIVS	(SEQ ID NO: 1765)
30	aATAF00130	QNP-PINEVV	QSG--VVPRV	VKFLS-----	RDDFPKLQFE	AAWALTNIAS	(SEQ ID NO: 1766)
	aATKAPAPRO	QNP-PINEVV	QSG--VVPRV	VKFLS-----	RDDFPKLQFE	AAWALTNIAS	(SEQ ID NO: 1766)
	ATU69533a	RSP-PIEEVI	SAG--VVPRF	VEFLK-----	KEDYPAIQFE	AAWALTNIAS	(SEQ ID NO: 1767)
	SLU96718a	RSP-PIEEVI	AAG--VVPRL	VQFLQ-----	RTDYPQLQFE	AAWALTNIAS	(SEQ ID NO: 1768)
	IMO2HUMANa	PNP-PIDEVI	STPG-VVARF	VEFLK-----	RKENC SLQFE	SAWVL TNIAS	(SEQ ID NO: 1769)
35	SRP1YEASTa	HRP-PIDVVI	QAG--VVPRL	VEFMR-----	ENQPEMLQLE	AAWALTNIAS	(SEQ ID NO: 1770)
	IMO1HUMAN1	KQP-PIDNII	RAG--LIPKF	VSFLGRT---	--DCSPIQFE	SAWALTNIAS	(SEQ ID NO: 1771)
	AB002533e	NQQ-QVQAVI	DAN--LVPMI	IHLDD-----	-KGDFGTQKE	AAWALSNLTI	(SEQ ID NO: 1772)
	HSSRP11c	NQQ-QVQAVI	DAN--LVPMI	IHLDD-----	-KGDFGTQKE	AAWALSNLTI	(SEQ ID NO: 1772)
	CELF32E10e	NQQ-QVQDVF	DAG--IMPVI	IHLDD-----	-RGDFPTQKE	AAWALSNVTI	(SEQ ID NO: 1773)
40	IMO2HUMANf	NRA-QIQTVI	DAN--IFPAL	ISILQ-----	-TAEFRTRKE	AAWALTNATS	(SEQ ID NO: 1774)
	MMU34228e	NRA-QIQTVI	DAN--MFPAL	ISILQ-----	-TAEFRTRKE	AAWALTNATS	(SEQ ID NO: 1775)
	eATAF00130	NAD-QIQAVI	DAG--IIQSL	VWVLQ-----	-SAEFEVKKE	AAWALSNATS	(SEQ ID NO: 1776)
	ATU69533f	NKD-QIQTVV	EAN--LISPL	VSLLO-----	-NAEFDIKKE	AAWALSNATS	(SEQ ID NO: 1777)
	SRP1YEASTf	NTE-QIQAVI	DAN--LIPPL	VKLLS-----	-VAEYKTKKE	ACWALSNASS	(SEQ ID NO: 1778)
45	IMO1HUMANe	RQD-QIQQVV	NHG--LVVPL	VSVLS-----	-KADFKTQKE	AVWAVTNYS	(SEQ ID NO: 1779)
	CELF26B13e	TQK-QIQAVL	DAN--LLPVL	INVLL-----	-SGDHKCQFE	ASWALSNLAQ	(SEQ ID NO: 1780)
	AB002533d	TDE-QTQVVL	NCD--ALSHF	PALLT-----	-HPKEKINKE	AVWFLSNITA	(SEQ ID NO: 1781)
	HSSRP11b	TDE-QTQVVL	NCD--ALSHF	PALLT-----	-HPKEKINKE	AVWFLSNITA	(SEQ ID NO: 1781)
	CELF32E10d	TDE-QTQLVL	DSG--VLRFM	PGLLAH----	--YKEKINKE	AVWFLSNITA	(SEQ ID NO: 1782)
50	IMO1HUMANd	TDE-QTQVVI	DAG--ALAVF	PSLLT-----	-NPKTNIQKE	ATWTMSNITA	(SEQ ID NO: 1783)
	IMO2HUMANe	DDI-QTQVIL	NCS--ALQSL	LHLLS-----	-SPKESIKKE	ACWTISNITA	(SEQ ID NO: 1784)

	MMU34228d	DDI-QTQVIL	NCS--ALQSL	LHLLS-----	-SPKESIKKE	ACWTISNITA	(SEQ ID NO: 1784)
	dATAF00130	DDL-QTQMV	DQQ--ALPCL	LNLLKNN---	--YKKSIIKE	ACWTISNITA	(SEQ ID NO: 1785)
	ATU69533e	DDI-QTQCVI	NSG--ALPCL	ANLLTQN---	--HKKSIIKE	ACWTISNITA	(SEQ ID NO: 1786)
	SRPLYEASTe	NDL-QTQVVI	NAG--VLPAL	RLLLS-----	-SPKENIKKE	ACWTISNITA	(SEQ ID NO: 1787)
5	CELF26B13d	NDS-LTQAVI	DLG--SLDEI	LPLME-----	KTRSSSIVKE	CCWLVSNIIA	(SEQ ID NO: 1788)
	CRU40057b	SPE-LAQSVI	DSG--ALDSL	VTCL-----	EEFDPGVKEA	SAWTLGYIAG	(SEQ ID NO: 1789)
	CRU40057e	SET-LALSVI	AEK--ALPPL	VSALN-----	EEPEDHLKSA	TAWTLGQIGR	(SEQ ID NO: 1790)
	CRU40057d	TPE-LAQAVV	DAG--AVAYL	APLVI-----	-NQDAKLKRQ	VCCALSQIAK	(SEQ ID NO: 1791)
	CRU40057c	NAD-VAQQVV	DAG--AVPLL	VLCVQ-----	-EPESLKRRI	AASALSDISK	(SEQ ID NO: 1792)
10	CTNBMOUSEk	DVH-NRIVIR	GLN--TIPLF	VQLLY-----	-SPIENIQRV	AAGVLCCELAQ	(SEQ ID NO: 1793)
	JC4835h	EAL-NRSIIR	DLN--CIPTF	VQLLY-----	-SEVENIVRV	AAGVLCCELAQ	(SEQ ID NO: 1794)
	CTNBMOUSEj	CPA-NHAPLR	EQG--AIPRL	VQLLVR(20)	GVRMEEIVEG	CTGALHILAR	(SEQ ID NO: 1795)
	HSPGLGNj	CPA-NHAPLQ	EAA--VIPRL	VQLLVK(19)	GVRMEEIVEG	CTGALHILAR	(SEQ ID NO: 1796)
	JC4835g	CPS-NHTPIR	DQG--GLPKL	VQLLMK(17)	GVRMEEIVEG	TVGALHILAR	(SEQ ID NO: 1797)
15	HSP0071a	DNK-VKMEVC	RLG--GIKHL	VDLLD-----	-HRVLEVQKN	ACGALRNLFV	(SEQ ID NO: 1798)
	HSU96136a	DNK-IKAEIR	RQG--GIQLL	VDLLD-----	-HRMTEVHRS	ACGALRNLFV	(SEQ ID NO: 1799)
	HSU51269a	NEG-VKRRVR	QLR--GLPLL	VALLD-----	-HPRAEVRRR	ACGALRNLSY	(SEQ ID NO: 1800)
	P120MOUSEa	NDK-VKTDVA	KLK--GIPIL	VGLLD-----	-HPKKEVHLG	ACGALKNISF	(SEQ ID NO: 1801)
	S60712a	DES-AKQQVY	QLG--GICKL	VDLLR-----	-SPNQNVQQA	AAGALRNLFV	(SEQ ID NO: 1802)
20	HSIRNAUa	KSE-ARKRVN	QLR--GILKL	LQLLK-----	-VQNEDEVQRA	VCGALRNLFV	(SEQ ID NO: 1803)
	CTNBMOUSEh	NYK-NKMMVC	QVG--GIEAL	VRTVLRA---	-GDREDITEP	AICALRHLS	(SEQ ID NO: 1804)
	JC4835e	NPR-NKQVVF	QVG--GIEAL	VRTIINA---	-GDREEITEP	AVCALRHLS	(SEQ ID NO: 1805)
	YEB3YEASTb	NNE-NKLLIV	EMG--GLEPL	INQMM-----	-GDNVEVQCN	AVGCITNLAT	(SEQ ID NO: 1806)
	YEB3YEASTf	HPL-NEGLIV	DAG--FLKPL	VRLLD-----	-YKDSEETQCH	AVSTLRNLAA	(SEQ ID NO: 1807)
25	P120MOUSEb	DQD-NKIAIK	NCD--GVPAL	VRLLRK----	-ARDMDLTEV	ITGTLWNLS	(SEQ ID NO: 1808)
	HSU51269b	DTD-NKAAIR	DCG--GVPAL	VRLLRK----	-ARDNEVREL	ITGTLWNLS	(SEQ ID NO: 1809)
	HSP0071b	TDE-NKIAMK	NVG--GIPAL	LRLLRK----	-SIDAEVREL	VTGVLWNLS	(SEQ ID NO: 1810)
	HSU96136b	NDD-NKIALK	NCG--GIPAL	VRLLRK----	-TTDLEIREL	VTGVLWNLS	(SEQ ID NO: 1811)
	S60712b	STT-NKLETR	RQN--GIREA	VSLLR-----	-RTGNAETQKQ	LTGLLWNLS	(SEQ ID NO: 1812)
30	HSIRNAUb	DND-NKLEVA	ELN--GVPR	LQVLK-----	-QTRDLETKKQ	ITGLLWNLS	(SEQ ID NO: 1813)
	HSU96136c	SVY-IRAAVR	KEK--GRPIL	VELLR-----	-IDNDRVACA	VATALRNMAL	(SEQ ID NO: 1814)
	HSP0071c	AAY-IRGGRP	KRK--GLPIL	VELLR-----	-MDNDRVVSS	GATALRNMAL	(SEQ ID NO: 1815)
	HSU51269c	ATY-IRATVR	KER--GLPVL	VELLQ-----	-SETDKVVRA	VAIALRNLSL	(SEQ ID NO: 1816)
	P120MOUSEc	GRY-IRSALR	QEK--ALSAR	AELLT-----	-SEHERVVKA	ASGALRNLA	(SEQ ID NO: 1817)
35	CTNBMOUSEi	AEM-AQNAVR	LHY--GLPVV	VKLLH-----	-PPSHWPLIKA	TVGLIRNLAL	(SEQ ID NO: 1818)
	JC4835f	AEH-AENGVR	LHY--GIPIL	VKLLN-----	-PPSRWPLIKA	VVGLIRNLGL	(SEQ ID NO: 1819)
	CTNBMOUSEc	HRE-GLLAIF	KSG--GIPAL	VKMLG-----	-SPVDSVLFY	AITTLHNLLL	(SEQ ID NO: 1820)
	JC4835a	HRQ-GLMAIF	KCS--GIPAL	VKLLG-----	-HRIEAVVFY	AITTLHNLLL	(SEQ ID NO: 1821)
	JC6161b	DDS-CAALLA	KSG--IIPAL	IELLN-----	-AQEDDEFVCQ	IIYVFYQMV	(SEQ ID NO: 1822)
40	HSU59919b	DDS-CAALLA	KSG--IIPAL	IELLN-----	-AQEDDEFVCQ	IIYVFYQMV	(SEQ ID NO: 1822)
	SPU38655b	DDS-CAALLA	KSG--IIQSL	IELLN-----	-AQEDDEFVCQ	IVYVFYQMV	(SEQ ID NO: 1823)
	CTNBMOUSEf	CSS-NKPAIV	EAG--GMQAL	GLHLT-----	-DPSQRLVQN	CLWTLRNLSD	(SEQ ID NO: 1824)
	JC4835d	CSS-NKPAIV	EAG--GMQAL	AHYLS-----	-HQSTRLVQN	CLWTLRNLSD	(SEQ ID NO: 1825)
	CELC54D15	SNF-DAPNLV	AFG--GRQIL	ANLLS-----	-HGSPRLVQS	TLETLRNLSD	(SEQ ID NO: 1826)
45	APCHUMAc	DVA-NKATLC	SMKG-CMRAL	VAQL-----	-KSESEDQOV	IASVLRNLSS	(SEQ ID NO: 1827)
	XLU64442c	DVA-NKATLC	SMKS-CMRAL	VAQL-----	-KSESEDQOV	IASVLRNLSS	(SEQ ID NO: 1828)
	DMU77947c	DEN-NKALLC	GQKQ-FMEAL	VAQL-----	-DSAPDDLQV	TASVLRNLSS	(SEQ ID NO: 1829)
	JC6161a	FME-NKNDMV	EMD--IVEKL	VKMIP-----	-CEHEDLLNI	TLRLLNLSS	(SEQ ID NO: 1830)
	HSU59919a	FME-NKNDMV	EMD--IVEKL	VKMIP-----	-CEHEDLLNI	TLRLLNLSS	(SEQ ID NO: 1830)
50	SPU38655a	YVE-NKNEMA	EQQ--IIERL	AKLVP-----	-CDHEDLLNI	TLRLLNLSS	(SEQ ID NO: 1831)
	CTNBMOUSEd	QEG-AKMAVR	LAG--GLQKM	VALLN-----	-KTNVKFLAI	TTDCLQILAY	(SEQ ID NO: 1832)

	JC4835b	QEG-AKMAVR LAL--GLQKM VSLQ-----	-RPKVKFLAI VTDCLQILAY (SEQ ID NO: 1833)
	CTNBMOUSEe	NQE-SKLIIL ASG--GPQAL VNIMR-----	TYTYEKLLWT TSRVLKVLVS (SEQ ID NO: 1834)
	JC4835c	NQE-SKLIIL SSG--GPAEL VRIMR-----	SYTYEKLLYT TCRVLKVLVS (SEQ ID NO: 1835)
	HSP0071d	NME-NAKALA DSG--GIEKL VNITKGRG--	DRSSLKVKA AAQVLNLTWQ (SEQ ID NO: 1836)
5	HSU96136d	NME-NAKALR DAG--GIEKL VGISKSKGD-	-KHSPKVKA ASQVLNSMWQ (SEQ ID NO: 1837)
	IMO1HUMANb	NPA-PPIDAV EQ---ILPTL VRLH-----	-HDDPEVLAD TCWALSYLTD (SEQ ID NO: 1838)
	ATU69533c	KPQ-PHFDQV KP---ALPAL ERLIH-----	-SDAEEALTD ACWALSYSLS (SEQ ID NO: 1839)
	CELF32E10-	DPA-PSPAVV RT---ILPAL SLLIHHQ---	---DTNILID TVWALSYLTD (SEQ ID NO: 1840)
	AB002533z	DPP-PPMETI QE---ILPAL CVLIHHT---	---DVNILVD TVWALSYLTD (SEQ ID NO: 1841)
10	IMO2HUMAN-	DST-MCRDYV LDCN-ILPPL LQLFSK----	-QNRLTMTRN AVWALSNLCS (SEQ ID NO: 1842)
	MMU34228b	DST-MCRDYV LNCN-ILPPL LQLFSK----	-QNRLTMTRN AVWALSNLCS (SEQ ID NO: 1843)
	SRP1YEASTc	DST-DYRDYV LQCN-AMEPI LGLFN-----	-SNKPSLIRT ATWTLSNLCS (SEQ ID NO: 1844)
	IMOB_RAT	EPN-QLKPLV IQ---AMPTL IELMKDP---	---SVVVRDT TAWTVGRICE (SEQ ID NO: 1845)
	HUMNTF9	EPS-QLKPLV IQ---AMPTL IELMKDP---	---SVVVRDT AAWTVGRICE (SEQ ID NO: 1846)
15	APCHUMANe	NED-HRQILR ENN--CLQTL LQHLK-----	-SHSLTIVSN ACGTLWNLSA (SEQ ID NO: 1847)
	XLU64442d	NED-HRQILR ENN--CLQTL LQHLK-----	-SHSLTIVSN ACGTLWNLSA (SEQ ID NO: 1847)
	DMU77947d	CEP-YRQILR QHN--CLAIL LQQLK-----	-SESLTVVSN SCGTLWNLSA (SEQ ID NO: 1848)
	APCHUMANb	DEE-HRHAMN ELG--GLQAI AELLQV(8)N	DHYSITLRRY AGMALTNLTF (SEQ ID NO: 1849)
	XLU64442b	DEE-HRHAMN ELG--GLQAI AELLQV(8)N	DHYSITLRRY AGMALTNLTF (SEQ ID NO: 1850)
20	DMU77947b	DEE-HRHAMC ELG--ALHAI PNLVHL(10)	DQCCNSLRRY ALMALTNLTF (SEQ ID NO: 1851)
	fATAF00130	THD-QIKFMV SQG--CIKPL CDLLT-----	-CPDLKVTV CLEALENILV (SEQ ID NO: 1852)
	dATKAPAPRO	THD-QIKFMV SQG--CIKPL CDLLT-----	-CPDLKVTV CLEALENILV (SEQ ID NO: 1852)
	ATU69533g	SHD-QIKYLV EQG--CIKPL CDLLV-----	-CPDPRIITV CLEALENILK (SEQ ID NO: 1853)
	IMO2HUMANg	SAE-QIKYLV ELG--CIKPL CDLLT-----	-VMDSKIVQV ALNGLENILR (SEQ ID NO: 1854)
25	MMU34228f	SAE-QIKYLV ELG--CIKPL CDLLT-----	-VMDAKIVQV ALNGLENILR (SEQ ID NO: 1855)
	SRP1YEASTg	RPD-IIRYLV SQG--CIKPL CDLLEIA---	---DNRIIEV TLDALENILK (SEQ ID NO: 1856)
	IMO1HUMANf	TVE-QIVYLV HCG--IIEPL MNLLT-----	-AKDTKIILV ILDAISNIFQ (SEQ ID NO: 1857)
	AB002533f	RKD-QVAYLI QQN--VIPPF CNLLT-----	-VKDAQVVQV VLDGLSNILK (SEQ ID NO: 1858)
	HSSRP11d	RKD-QVAYLI QQN--VIPPF CNLLT-----	-VKDAQVVQV VLDGLSNILK (SEQ ID NO: 1858)
30	CELF32E10g	RPN-QVEQMV KLG--VLRPF CAMLSCT---	---DSQIIQV VLDGINNILK (SEQ ID NO: 1859)
	APCHUMANa	SQD-SCISMR QSG--CLPLL IQLLHG(10)	SRGSKEARAR ASAALHNIIH (SEQ ID NO: 1860)
	XLU64442a	SQD-SCIAMR QSG--CLPLL IQLLHG(10)	SRGSKEARAS GSAALDNIIH (SEQ ID NO: 1861)
	DMU77947a	NAQ-SCATLR RSG--CMPLL VQMMH----	A PDNDQEVKRC AEQALHNVVH (SEQ ID NO: 1862)
	CTNBMOUSEl	DKE-AAEAIE AEG--ATAPL TELLH-----	-SRNEGVATY AAALVFRMSE (SEQ ID NO: 1863)
35	JC4835i	DKE-GADAIE REG--ATTIL TELLH-----	-SRNDGIAAY ARAVLFMSE (SEQ ID NO: 1864)
	CELF08F8	FDE-NKIVME QNG--TIEKL LKLFP-----	-IQDPELRKA VIMLLFNFSF (SEQ ID NO: 1865)
	GDS1HUMANc	RNDANCIHNV DNG--IVEKL MDLLDRHVE-	-DGNVTQHA ALSALRNLA (SEQ ID NO: 1866)
	AT81KBGEN4	NDR-NKKIIV DEG--GVSP LRLK-----	ESSSAEGQIA AATALLGAC (SEQ ID NO: 1867)
	S51350	PDKVQRTYYV HQ---ALPSI LNLMDQ---	---SLQVKET TAWCIGRIAD (SEQ ID NO: 1868)
40	CELF26B13c	PDE-QIELAR ESG--VLPHV VAFF-----	-KEAENLVAP ALRTLGNVAT (SEQ ID NO: 1869)
	HSU51269d	DSLNDARSL QAR--GVPAL VALVA-----	SSQSVREKA ASHVLQTVWS (SEQ ID NO: 1870)
	CELF26B13-	SSQLRDYVIK CHG---VEAL MHLMEKV---	DQLGDSHVRT IAWAFSNMCR (SEQ ID NO: 1871)
	YEB3YEASTd	SEE-NRKELV NAG--AVPVL VSLLS-----	-STDPDVQYY CTTALSNIIV (SEQ ID NO: 1872)
	GDS1HUMANa	KNEFMRI PCV DAG--LISPL VQLLN-----	-SKDQEVLLQ TGRALGNICY (SEQ ID NO: 1873)
45	YEB3YEASTe	SDTSYQLEIV RAG--GLPHL VKLIQ-----	-SDSIPLVLA SVACIRNISI (SEQ ID NO: 1874)
	CTNBMOUSEa	NYQDDAELAT R----AIPEL TKLLN-----	-DEDQVVVNK AAVMVHQLSK (SEQ ID NO: 1875)
	ADBHUMAN	LHDINAQMV DQG--FLDSL RDLIA-----	-DSNPMVVAN AVAALSEISE (SEQ ID NO: 1876)
	S60712c	SGMSQLIGLK EK---GLPQI ARLLQ-----	-SGNSDVVRS GASLLSNMSR (SEQ ID NO: 1877)
	HSIRNAUd	PTSVAQTVVQ KES--GLQHT RKMLH-----	-VGDPSPVKKT AISLLRNLSR (SEQ ID NO: 1878)
50	YEB3YEASTa	FAEITEKYVR QVSREVL EPI LILLQ-----	-SQDPQIQVA ACAALGNLAV (SEQ ID NO: 1879)
	HSIRNAUc	GADGRKAMRR CDG--LIDSL VHYVR-(4)D	YQPD DKATEN CVCILHNLSY (SEQ ID NO: 1880)

116

	CTNBMUSEb	KEASRHAIMR	SPQ--MVSAI	VRTMQ-----	NTNDVETARC	TAGTLHNLSH	(SEQ ID NO: 1881)	
	CET19B106	ANQKHLEYVV	ELG--MLSAF	TDLLTCM---	---DVSLVSY	ILDALYLLQ	(SEQ ID NO: 1882)	
	YSPPAA1B	AAG-GDSTVY	EQQ--IIPLL	EQLTK-----	-DNDPDVQYF	ATQALEQTND	(SEQ ID NO: 1883)	
	CRU40057a	NPQ-NIEALQ	QAG--AMALL	RPLLL-----	-DNVPSIQQS	AALALGRLAN	(SEQ ID NO: 1884)	
	D87671	YSDSLTAAMI	DA--VLDEL	PPLISES---	---DMHVSQM	AISFLTTLAK	(SEQ ID NO: 1885)	
5	CELM01E114	SQPDLYGSFV	EVN--GVEIL	LQLLG-----	-HENTDIVCA	TLSSLRELTD	(SEQ ID NO: 1886)	
	HSZYGHOMO	ETPDNCEMFL	NFN--GMKLF	LDCLK-----	E	FPEKQELIRN	MLGLLGNVAE	(SEQ ID NO: 1887)
	YLK3CAEELa	CEEQAREQIR	IYD--GVPTL	LGLLS-----	-IKNSRLQWH	VAWTLAQLAE	(SEQ ID NO: 1888)	
	GDS1HUMANb	DSHSLQAQLI	NMG--VIPTL	VKLLG---I	HCQNAALTEM	CLVAFGNLAE	(SEQ ID NO: 1889)	
	COPBYEASTa	RQDANRTPAL	KAQ--YIELL	MELLS-----	TTTSDEVIFE	TALALTVLSA	(SEQ ID NO: 1890)	
10	CEC48D1	KSSKACTNMI	QAG--IISIS	LSCMEK----	-VESNDITTK	CLELLTNLSS	(SEQ ID NO: 1891)	
	YEB3YEASTc	RDD-NKHKIA	TSG--ALIPL	TKLAKSK---	---HIRVQRN	ATGALLNMTH	(SEQ ID NO: 1892)	
	GDS1BOVINc	KSKDVIKTIV	QSG--GIKHL	VTMATSE---	---HVIMQNE	ALVALALIAA	(SEQ ID NO: 1893)	
	APCHUMAND	ADVNSKKTLR	EVG--SVKAL	MECALEVKK-	-ESTLKSVL	ALWNLSARNP	(SEQ ID NO: 1894)	
	P115_BOVIN	SKK-YRLEV	G	IQ---AMEHL	IHVLQTD---	-RSDSEIIGY	ALDTLYNIIS	(SEQ ID NO: 1895)
15	YD71_SCHPO	PLE-MIQIVT	KMN--IPDHL	TATLPRS---	---SDDIQND	CLFALRQVSK	(SEQ ID NO: 1896)	
	CELB033611	SDL-PAIQEQ	DMKE-SIHCI	VQLIG-----	-CSDVTIVEL	ATGTLRNIGL	(SEQ ID NO: 1897)	
	Consensus/60%	ssp.pbphlb	pss..slshL	lpLLp.....	.p.s.plbp.	tshsLpNls.	(SEQ ID NO: 1898)	

20

Table 15

References

- 5 Bondeson, D.P., Mares, A., Smith, I.E.D., Ko, E., Campos, S., Miah, A.H., Mulholland, K.E., Routly, N., Buckley, D.L., Gustafson, J.L., et al. (2015). Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat. Chem. Biol.* 11, 611–617.
- 10 Boudko, S.P., Londer, Y.Y., Letarov, A. V, Sernova, N. V, Engel, J., and Mesyanzhinov, V. V (2002). Domain organization, folding and stability of bacteriophage T4 fibrin, a segmented coiled-coil protein. *Eur. J. Biochem.* 269, 833–841.
- 15 Brunette, T.J., Parmeggiani, F., Huang, P.-S., Bhabha, G., Ekiert, D.C., Tsutakawa, S.E., Hura, G.L., Tainer, J.A., Baker, D. (2015) Exploring the repeat protein universe through computational protein design. *Nature* 528, 580–584.
- 20 Chapman & McNaughton, B.R. (2016). Scratching the surface: Resurfacing proteins to endow new properties and function. *Cell Chem. Biol.* 23, 543–553.
- D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem. Sci.* 28, 655–662.
- 25 Deshaies, R.J. (2015). Protein degradation: Prime time for PROTACs. *Nat. Chem. Biol.* 11, 634–635.
- de Vries, S.J., and Bonvin, A.M.J.J. (2011). CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS One* 6, e17695.
- 30 de Vries, S.J., van Dijk, M., and Bonvin, A.M.J.J. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* 5, 883–897.
- Guettler, S., LaRose, J., Petsalaki, E., Gish, G., Scotter, A., Pawson, T., Rottapel, R., and Sicheri, F. (2011). Structural basis and sequence rules for substrate recognition by Tankyrase explain the basis for cherubism disease. *Cell* 147, 1340–1354.
- 35 Güthe, S., Kapinos, L., Möglich, A., Meier, S., Grzesiek, S., and Kiefhaber, T. (2004). Very Fast Folding and Association of a Trimerization Domain from Bacteriophage T4 Fibrin. *J. Mol. Biol.* 337, 905–915.
- Hao, B., Zheng, N., Schulman, B.A., Wu, G., Miller, J.J., Pagano, M., Pavletich, N.P. (2005). Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. *Mol. Cell* 20, 9–19.
- 40 Kobe, B. & Kajava, A.V. (2000). When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends in Biochem. Sci.* 25, 509–515.
- 45 Lee, J.-H., Kang, E., Lee, J., Kim, J., Lee, K.H., Han, J., Kang, H.Y., Ahn, S., Oh, Y., Shin, D., et al. (2014). Protein grafting of p53TAD onto a leucine zipper scaffold generates a potent HDM dual inhibitor. *Nat. Commun.* 5, 3814.
- 50 Leshchiner, E.S., Parkhitko, A., Bird, G.H., Luccarelli, J., Bellairs, J.A., Escudero, S., Opoku-Nsiah, K., Godes, M., Perrimon, N., and Walensky, L.D. (2015). Direct inhibition of oncogenic KRAS by hydrocarbon-stapled SOS1 helices. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1761–1766.
- Longo, L.M. & Blaber, M. (2014). Symmetric protein architecture in protein design: to-down symmetric deconstruction. *Methods Mol. Biol.* 1216, 161–82.
- 55 Lu, J., Qian, Y., Altieri, M., Dong, H., Wang, J., Raina, K., Hines, J., Winkler, J.D., Crew, A.P., Coleman, K., et al. (2015). Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem. Biol.* 22, 755–763.
- 60 Margarit, S.M., Sondermann, H., Hall, B.E., Nagar, B., Hoelz, A., Pirruccello, M., Bar-Sagi, D., and Kuriyan, J. (2003). Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* 112, 685–695.

Meier, S., Guthe, S., Kiefhaber, T. and Grzesiek, S. (2004). Foldon, the natural trimerization domain of T4 fibrin, dissociates into a monomeric A-state form containing a stable beta-hairpin: atomic details of trimer dissociation and local beta-hairpin stability from residual dipolar couplings. *J. Mol. Biol.* 344, 1051-1069.

5

Parmeggiani, F., Huang, P.-S., Vorobiev, S., Xiao, R., Park, K., Caprari, S., Su, M., Seetharaman, J., Mao, L., Janjua, H., Montelione, G.T., Hunt, J., Baker, D. (2015) A general computational approach for repeat protein design. *J. Mol. Biol.* 427, 563-575.

10

Rowling, P.J., Sivertsson, E.M., Perez-Riba, A., Main, E. R., Itzhaki, L.S. (2015) *Biochem. Soc. Trans.* 43 881-888.

Tamaskovic, R., Simon, Stefan, N., Scwhill, Pluckthun, A. (2012). Designed ankyrin repeat proteins (DARPs): From research to therapy. *Methods in Enzym.* 503, 101-134.

15

Thompson, D.B., Cronican, J.J., Liu, D.R. (2012). Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells. *Methods Enzymol.* 503, 293-319.

20

Claims

1. A modular binding protein comprising;
 (i) two or more repeat domains,
 5 (ii) inter-repeat loops linking said repeat domains; and
 (iii) one or more heterologous peptide ligands that bind to a target molecule, each said peptide ligand being located in an inter-repeat loop or at the N or C terminus of the modular binding protein,

wherein said one or more heterologous peptide ligands comprise at least one amino acid
 10 sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) or a variant thereof.

2. A modular binding protein according to claim 1 comprising a first peptide ligand that binds a first target molecule and a second peptide ligand that binds a second target molecule.

3. A modular binding protein according to claim 2 wherein the first target molecule is β -catenin, KRAS, c-Myc, BRD4, Aurora-A, CK2alpha, Notch, Cdk2, PLK1, BCR-ABL, PP1c, MCL-1, Bcl-2, BCL-XL, Jun, BFL1, BAX, e1F4E, Fos, HDAC4, BCL6, Tau, PD-L1, KDM4A, EGFR, RAB25, or a transmembrane protein or GPCR.

4. A modular binding protein according to claim 2 or claim 3 comprising a first peptide ligand having an amino acid sequence set out in Table 8 (SEQ ID NOs: 277-388) or a variant thereof.

5. A modular binding protein according to claim 4 wherein the first target molecule is β -catenin and the peptide ligand has the amino acid sequence;

RSKKAHVLAASVEQATQNFLEKGEQIAKESQ (SEQ ID NO: 277),

RTLTVERLLEPLVTQVTTLV (SEQ ID NO: 278),

REQLEAQEARAREAHAREAHAREAYTREAYGREAYAREAHWTWEAHGREARTREAQA (SEQ ID NO: 279),

DX1X2EFDQYL (SEQ ID NO: 280), where X1 and X2 are independently any amino acid,

QALLDKAKINQ (SEQ ID NO: 281), or

GWLDSSRSLMEQDKENEALLRF (SEQ ID NO: 282).

6. A modular binding protein according to claim 4 wherein the first target molecule is KRAS and the peptide ligand has the amino acid sequence;

PLYISY (SEQ ID NO: 283)

SIEDLHEYWARLWNYLYVA (SEQ ID NO: 284)

QASLEELHEYWARLWNYRVA (SEQ ID NO: 285)

NASIKQLHAYWQRLYAYLAAVA (SEQ ID NO: 286)

CMWWREICPVWW (SEQ ID NO: 287)

FARKTFLKLAF (SEQ ID NO: 288)

ARRFFLDIAD (SEQ ID NO: 289)

FRWPX1X2RLX3X4 (SEQ ID NO: 290), where X1 to X4 are independently any amino acid

tX2VFXhX1p (SEQ ID NO: 291), where X1 to X2 are independently any amino acid

5 YGHGQVYYY (SEQ ID NO: 292)

ENPKQN (SEQ ID NO: 293)

DAYECLDASRPW (SEQ ID NO: 294) or

KSRDFYH (SEQ ID NO: 295).

10 7. A modular binding protein according to claim 4 wherein the first target molecule is myc and the peptide ligand has the amino acid sequence;

AGVEHQLRREVEIQSH (SEQ ID NO: 296)

WSVHAPSSRRTTLAGTLDYLPPEMI (SEQ ID NO: 297)

TYQETY (SEQ ID NO: 298)

15 QAEEQKLSEEDLLR KRREQLKHKLEQLRNSCA (SEQ ID NO: 299)

NELKRSFAALRDQI (SEQ ID NO: 300)

NELKRAFAALRDQI (SEQ ID NO: 301)

IREKNHYHRQEVDLRRQNALLEQQVRAL (SEQ ID NO: 302)

FNHITNASQWE (SEQ ID NO: 303)

20 GDLGAFSRGQM (SEQ ID NO: 304)

RSEFYYYGNTYYYSAMD (SEQ ID NO: 305)

QHDYTATDE (SEQ ID NO: 306)

QNPEEQDEGW (SEQ ID NO: 307) or

EKCRGVFPENF (SEQ ID NO: 308).

25

8. A modular binding protein according to any one of claims 2 to 7 wherein the second target molecule is an E3 ubiquitin ligase selected from MDM2, SCF(Skp2), Cul3-Keap1, Cul3-SPOP, Cul3-KELCH, KELCH actinfilin, APC/C, TRP1, SCF^{Fbw7}, SCF_TIR1, Cul4-DDB1-Cdt2_1, SCF_TRCP1_1, β-TRCP, SCF^{Fbw2}, SCF^{Fbw3}, OPDH VHL 1, SCF coil, SOCS box-Cul5-SPSB4, UBR5, CRL2(KLHDC2), GID4, TRIM21, Nedd4, ElonginC, CBL, CRL4_CDT2_1, CBL(PTK), CBL(Met), CRL4(COP1/DET), SH3RF1, COP-1, SIAH, ERAD-C, and UBR.

30

9. A modular binding protein according to any one of claims 2 to 7 wherein the second target molecule is selected from ALIX, AP-1, AP-2, AP, Hsc70, LC3, FIP200, LC3 and GABARAP.

35

10. A modular binding protein according to claim 8 or claim 9 comprising a second peptide ligand having an amino acid sequence set out in Table 9 (SEQ ID NOS: 470-617) or a variant thereof.

11. A modular binding protein according to any one of the preceding claims wherein one or more of the peptide ligands are located in inter-repeat loops.

40

12. A modular binding protein according to claim 11 wherein the one or more peptide ligands are connected to the inter-repeat loops by one or more additional residues.

5 13. A modular binding protein according to any one of the preceding claims wherein a peptide ligand is located at the N terminus.

14. A modular binding protein according to claim 13 wherein the N terminal peptide ligand comprises an α helix.

10

15. A modular binding protein according to claim 13 or 14 wherein the N terminal peptide ligand comprises the sequence $X_n\text{-}Y\text{XXXIXYXXLXX-}X_1X_2\text{XX}$, where residues denoted by X are independently any amino acid, X_1 , and X_2 are independently any amino acid and n is 0 or any number.

15 16. A modular binding protein according to claim 15 wherein X_1 is D.

17. A modular binding protein according to claim 15 or claim 16 wherein X_2 is P.

18. A modular binding protein according to any one of the preceding claims wherein peptide
20 ligand is located at the C terminus.

19. A modular binding protein according to claim 18 wherein the C terminal peptide ligand comprises an α helix.

25 20. A modular binding protein according to claim 19 wherein the C terminal peptide ligand comprises the sequence $X_1X_2\text{XX-XXAXXXLXX[AV]XXXXX-X}_n$, where residues denoted by X are independently any amino acid, X_1 , and X_2 are independently any amino acid and n is 0 or any number.

21. A modular binding protein according to claim 20 wherein X_1 is D.

30

22. A modular binding protein according to claim 20 or claim 21 wherein X_2 is P.

23. A modular binding protein according to any of claims 1 to 10 comprising an N terminal peptide ligand of Table 8 (SEQ ID NOs: 277-388) or a variant thereof and a C terminal peptide ligand of Table
35 9 (SEQ ID NOs: 470-617) or a variant thereof.

24. A modular binding protein according to any of claims 1 to 10 comprising an inter-repeat peptide ligand of Table 8 (SEQ ID NOs: 277-388) or a variant thereof and a C terminal peptide ligand of Table 9 (SEQ ID NOs: 470-617) or a variant thereof.

40

25. A modular binding protein according to any of claims 1 to 10 comprising an inter-repeat peptide ligand of Table 8 (SEQ ID NOs: 277-388) or a variant thereof and an N terminal peptide ligand of Table 9 (SEQ ID NOs: 470-617) or a variant thereof
- 5 26. A modular binding protein according to any of claims 1 to 10 comprising a C terminal peptide ligand of Table 8 (SEQ ID NOs: 277-388) or a variant thereof and an N terminal peptide ligand Table 9 (SEQ ID NOs: 470-617) or a variant thereof
- 10 27. A modular binding protein according to any of claims 1 to 10 comprising an inter-repeat peptide ligand of Table 9 (SEQ ID NOs: 470-617) or a variant thereof and an N terminal peptide ligand of Table 8 (SEQ ID NOs: 277-388) or a variant thereof.
- 15 29. A modular binding protein according to any of claims 1 to 10 comprising an inter-repeat peptide ligand of Table 9 (SEQ ID NOs: 470-617) or a variant thereof and a C terminal peptide ligand of Table 8 (SEQ ID NOs: 277-388) or a variant thereof.
30. A modular binding protein according to any of the preceding claims wherein the repeat domains are helix-turn-helix repeat domains.
- 20 31. A modular binding protein according to claim 30 wherein the repeat domains are tetratricopeptide (TPR) repeat domains.
32. A modular binding protein according to claim 31 wherein the repeat domains have the amino acid sequence Y-X₁X₂X₃X₄;
- 25 wherein Y is an amino acid sequence shown in any of Tables 4 to 6 or a variant thereof and X₁, X₂, X₃, X₄ are independently any amino acid.
33. A modular binding protein according to claim 31 or claim 32 wherein the repeat domains have the amino acid sequence;
- 30 AEAWYNLGNAYYKQGDYQKAIEYYQKALEL-X₁X₂X₃X₄ (SEQ ID NO: 13); or
 AEALNNLGNVYREQGDYQKAIEYYQKALEL-X₁X₂X₃X₄ (SEQ ID NO: 1899); or
 AEAWYNLGNAYYRQGDYQRAIEYYQRALEL-X₁X₂X₃X₄ (SEQ ID NO: 20); or
 AEALNNLGNVYREQGDYQRAIEYYQRALEL-X₁X₂X₃X₄ (SEQ ID NO: 21); or
 AEALRNLGRVYRRQGRYQRAIEYYRRALEL-X₁X₂X₃X₄ (SEQ ID NO: 1900)
- 35 wherein X₁, X₂, X₃, X₄ are independently any amino acid, or a variant thereof.
34. A modular binding protein according to claim 32 or claim 33 wherein X₁ is D.
35. A modular binding protein according to any one of claims 32 to 34 wherein X₂ is P.

36. A modular binding protein according to claim 31 wherein the repeat domains have the amino acid sequence:

EDSAEAERLKTEGNEQMKVENFEAAVHFYGKAIELNPANAVYFCNRAAAYS KLGN YAGAVQ
DCERAICIDPAYSKAYGRMGLALSSLNKHVEAVAYYKKALELDPDNETYKSNL KIAELK LREA

P (SEQ ID NO: 18)

or a variant thereof

37. A modular binding protein according to any one of the preceding claims comprising 2-5 repeat domains.

38. A modular binding protein according to any one of the preceding claims further comprising a targeting domain, intracellular transfer domain, stabilising domain, oligomerisation domain, cytotoxic agent, therapeutic agent and/or a detectable label.

39. A modular binding protein comprising an amino sequence set out in Table 13 (SEQ ID NOs: 1230 to 1304)

40. A nucleic acid encoding a modular binding protein according to any one of claims 1 to 39.

41. An expression vector comprising a nucleic acid according to claim 40.

42. A host cell comprising an expression vector according to claim 41.

43. A method of producing a modular binding protein comprising expressing a nucleic acid according to claim 40 to produce the modular binding protein.

44. A method of producing a modular binding protein comprising;
inserting a first nucleic acid encoding a peptide ligand of Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) into a second nucleic acid encoding two or more repeat domains linked by inter-repeat loops to produce a chimeric nucleic acid encoding a modular binding protein according to any one of claims 1 to 39; and
expressing said chimeric nucleic acid to produce the modular binding protein.

45. A method of producing a modular binding protein that binds to a first target molecule and a second target molecule comprising;
providing a nucleic acid encoding two or more repeat domains linked by inter-repeat loops;
and
incorporating into said nucleic acid a first nucleotide sequence encoding a first peptide ligand that binds to a first target molecule and a second nucleotide sequence encoding a second peptide ligand that binds to a second target molecule to generate a nucleic acid encoding a modular binding

protein comprising said first and second peptide ligands, wherein said peptide ligands are located in an inter-repeat loop or at the N or C terminus of the modular binding protein; wherein one or both of the first and second peptide ligands are set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and

5 expressing the nucleic acid to produce said protein.

46. A method according to claim 45 wherein one of the first or second target molecules is an E3 ubiquitin ligase.

10 47. A library comprising modular binding proteins, each modular binding protein in the library comprising;

 (i) two or more repeat domains,

 (ii) inter-repeat loops linking said repeat domains; and

15 (iii) a first peptide ligand of Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and a second peptide ligand, each said peptide ligand being located in an inter-repeat loop or at the N or C terminus of the modular binding protein, wherein at least one amino acid residue in the second peptide ligand in said library is diverse.

20 48. A library according to claim 47 wherein each modular binding protein in the library comprises one peptide ligand.

49. A library comprising a first and a second sub-library of modular binding proteins, each modular binding protein in the first and second sub-libraries comprising;

 (i) two or more repeat domains,

25 (ii) inter-repeat loops linking said repeat domains; and

 (iii) a peptide ligand of Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and a peptide ligand comprising at least one diverse amino acid residue,

30 wherein the peptide ligand in the modular binding proteins in the first sub-library binds to a first target molecule and is located in one of (i) an inter-repeat loop; (ii) the N terminus or (iii) the C terminus of the modular binding protein, and

 the peptide ligand in the modular binding proteins in the second sub-library binds to a second target molecule and is located in another of (i) an inter-repeat loop; (ii) the N terminus or (iii) the C terminus of the modular binding protein.

35 50. A library according to any one of claims 47-49 wherein the modular binding proteins are according to any one of claims 1 to 39.

51. A library according to any one of claims 47-50 wherein the library is displayed on the surface of particles

52. A method of screening a library comprising;
(a) providing a library of modular binding proteins, each modular binding protein in the library comprising;

- (i) two or more repeat domains
- (ii) inter-repeat loops linking said repeat domains; and
- (iii) a first peptide ligand located in an inter-repeat loop, at the N terminus or at the C terminus of the protein, and
- (iv) a second peptide ligand located in an inter-repeat loop, at the N terminus or at the C terminus of the protein

wherein the first peptide ligand has a sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and the second peptide ligand has at least one diverse amino acid residue,

- (b) screening the library for modular binding proteins which display a binding activity, and
- (c) identifying one or more modular binding proteins in the library which display the binding

activity.

53. A population of nucleic acids encoding a library according to any one of claims 47-51.

54. A method of producing a library comprising expressing a population of nucleic acids according to claim 53.

55. A method of producing a library of modular binding proteins comprising;
(a) providing a population of nucleic acids encoding a diverse population of modular binding proteins comprising

- (i) two or more repeat domains
- (ii) inter-repeat loops linking said repeat domains;
- (iii) a first peptide ligand located in an inter-repeat loop, at the N terminus or at the C terminus of the protein, and
- (iv) a second peptide ligand located in an inter-repeat loop, at the N terminus or at the C terminus of the protein,

wherein the first peptide ligand has a sequence set out in Table 8 (SEQ ID NOs: 277-388) or Table 9 (SEQ ID NOs: 470-617) and the second peptide ligand has at least one diverse amino acid residue, and

(b) expressing said population of nucleic acids to produce the diverse population, thereby producing a library of modular binding proteins.

56. A pharmaceutical composition comprising a modular binding protein according to any one of claims 1 to 39, a nucleic acid according to claim 40 or a vector according to claim 41 and a pharmaceutically acceptable excipient.

57. A method of producing a pharmaceutical composition comprising formulating a modular binding protein according to any one of claims 1 to 39, a nucleic acid according to claim 40 or a vector according to claim 41 with a pharmaceutically acceptable excipient.

58. A population of liposomes comprising a modular binding protein according to any one of claims 1 to 39, a nucleic acid according to claim 40 or a vector according to claim 41 a .

59. A method of producing a population of liposomes comprising admixing a modular binding protein according to any one of claims 1 to 39, a nucleic acid according to claim 40 or a vector according to claim 41 a with a lipid solution and evaporating said solution to produce liposomes encapsulating said modular binding protein, nucleic acid or vector.

60. A modular binding protein according to any one of claims 1 to 39, a nucleic acid according to claim 40 or a vector according to claim 41 for use in a method of diagnosis or treatment in human or animal subject.

61. A modular binding protein according to any one of claims 1 to 39 that binds to a target molecule, a nucleic acid according to claim 40 that encodes a modular binding protein that binds to a target molecule or a vector according to claim 41 for use in the treatment of a disorder associated with the target molecule.

62. Use of a modular binding protein according to any one of claims 1 to 39 that binds to a target molecule, a nucleic acid according to claim 40 that encodes a modular binding protein that binds to a target molecule or a vector according to claim 41 in the manufacture of a medicament for use in the treatment of a disorder associated with the target molecule.

63. A method of treatment of a disorder associated with the target molecule comprising; administering a modular binding protein according to any one of claims 1 to 39 that binds to a target molecule, a nucleic acid according to claim 40 that encodes a modular binding protein that binds to a target molecule or a vector according to claim 41 comprising said nucleic acid to an individual in need thereof.

64. A modular binding protein, nucleic acid or vector for use according to claim 61, use according to claim 62 or method according to claim 63 wherein the disorder is an inflammatory disorder, a neurodegenerative disease, an angiogenic disorder, a bone loss disorder or cancer.

65. A modular binding protein, nucleic acid or vector for use, use or method according to claim 64 wherein the cancer is breast, ovarian, colorectal, gastrointestinal, pancreatic, prostate, thyroid, lung, hepatocellular carcinoma, oesophageal, multiple myeloma, leukemia, T-cell lymphoma, neuroblastoma, glioblastoma multiforme, pleural mesothelioma, myelogenous leukemia (CML), acute

lymphoblastic leukemia (ALL) or acute myelogenous leukemia (AML).

66. A modular binding protein, nucleic acid or vector for use, use or method according to claim 64 wherein the neurodegenerative disease is Alzheimer's disease, Huntington's disease, Parkinson's disease, multiple sclerosis or amyotrophic lateral sclerosis.
67. A modular binding protein, nucleic acid or vector for use, use or method according to claim 64 wherein the inflammatory disorder is an autoimmune disease.
68. A modular binding protein, nucleic acid or vector for use, use or method according to any one of claims 60 to 67 wherein the modular binding protein, nucleic acid or vector is encapsulated in a liposome.

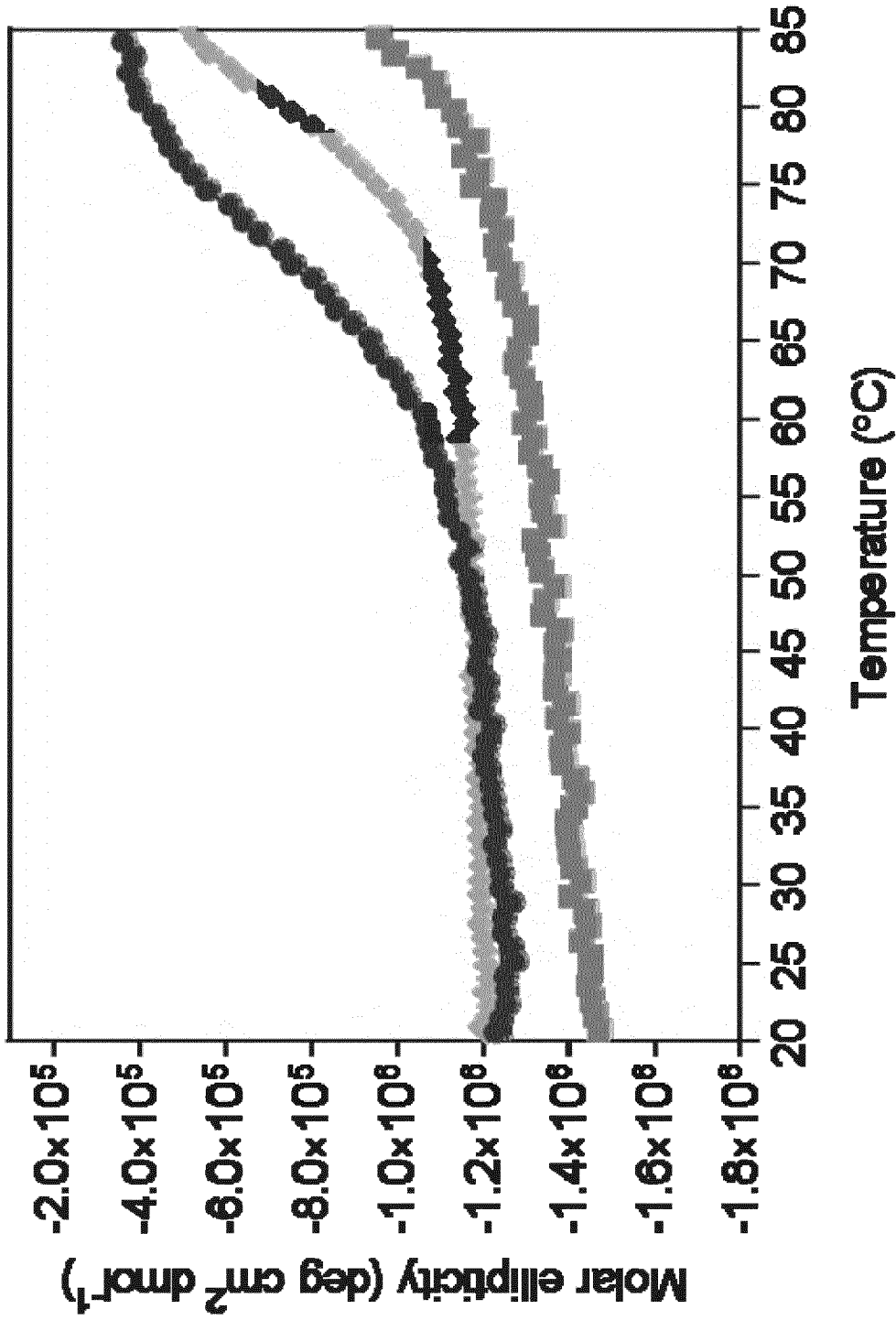


Figure 1

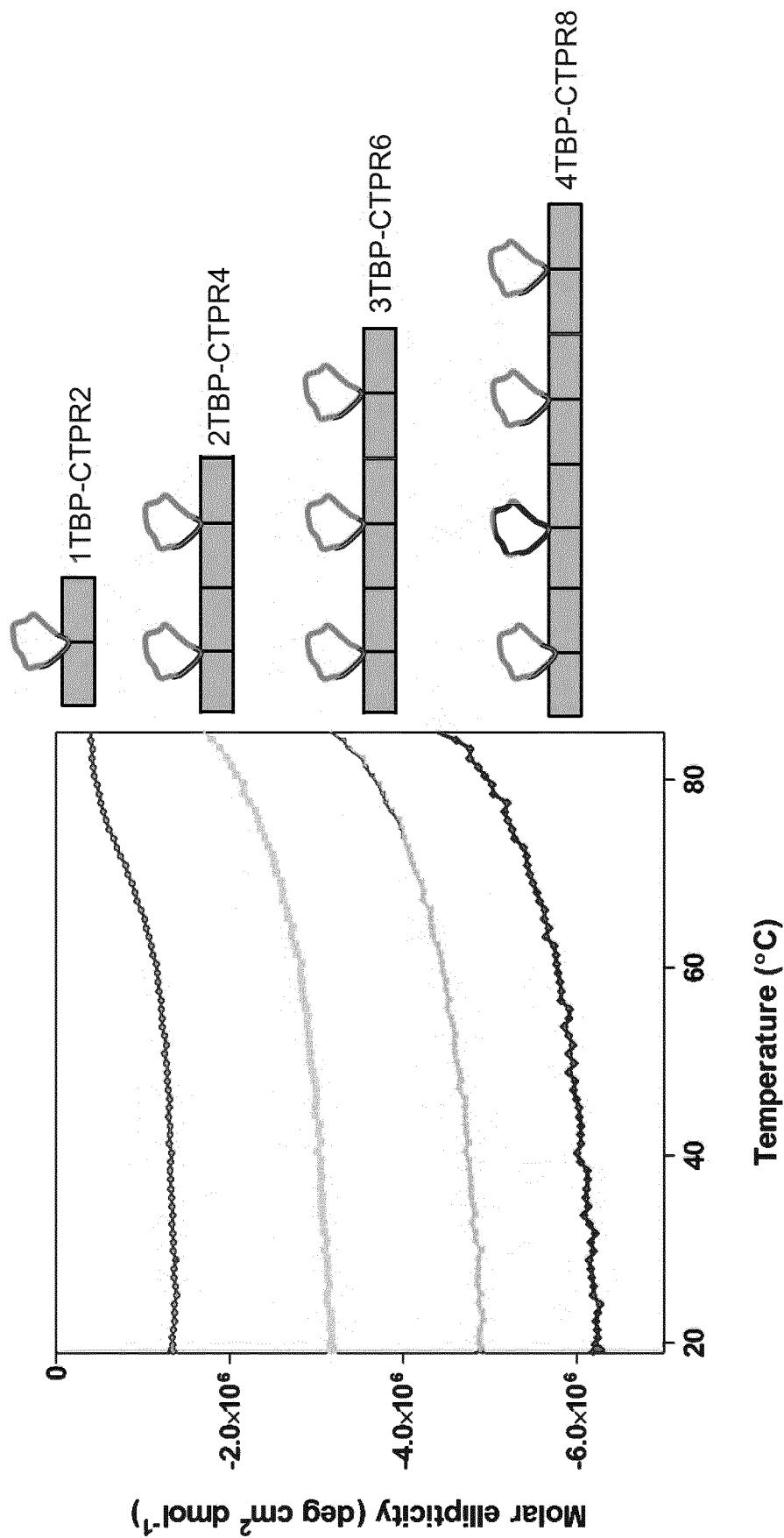


Figure 2

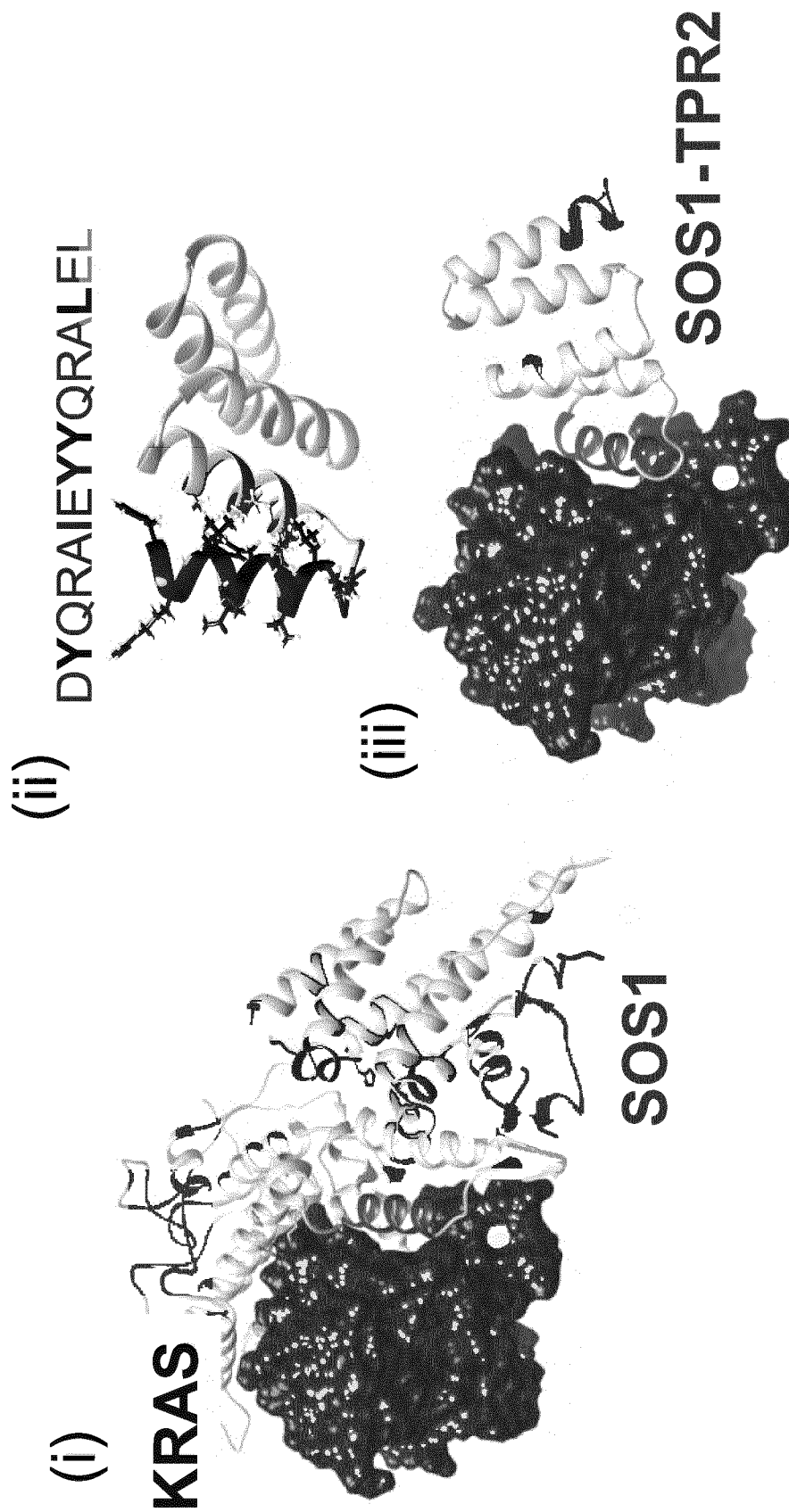


Figure 3A

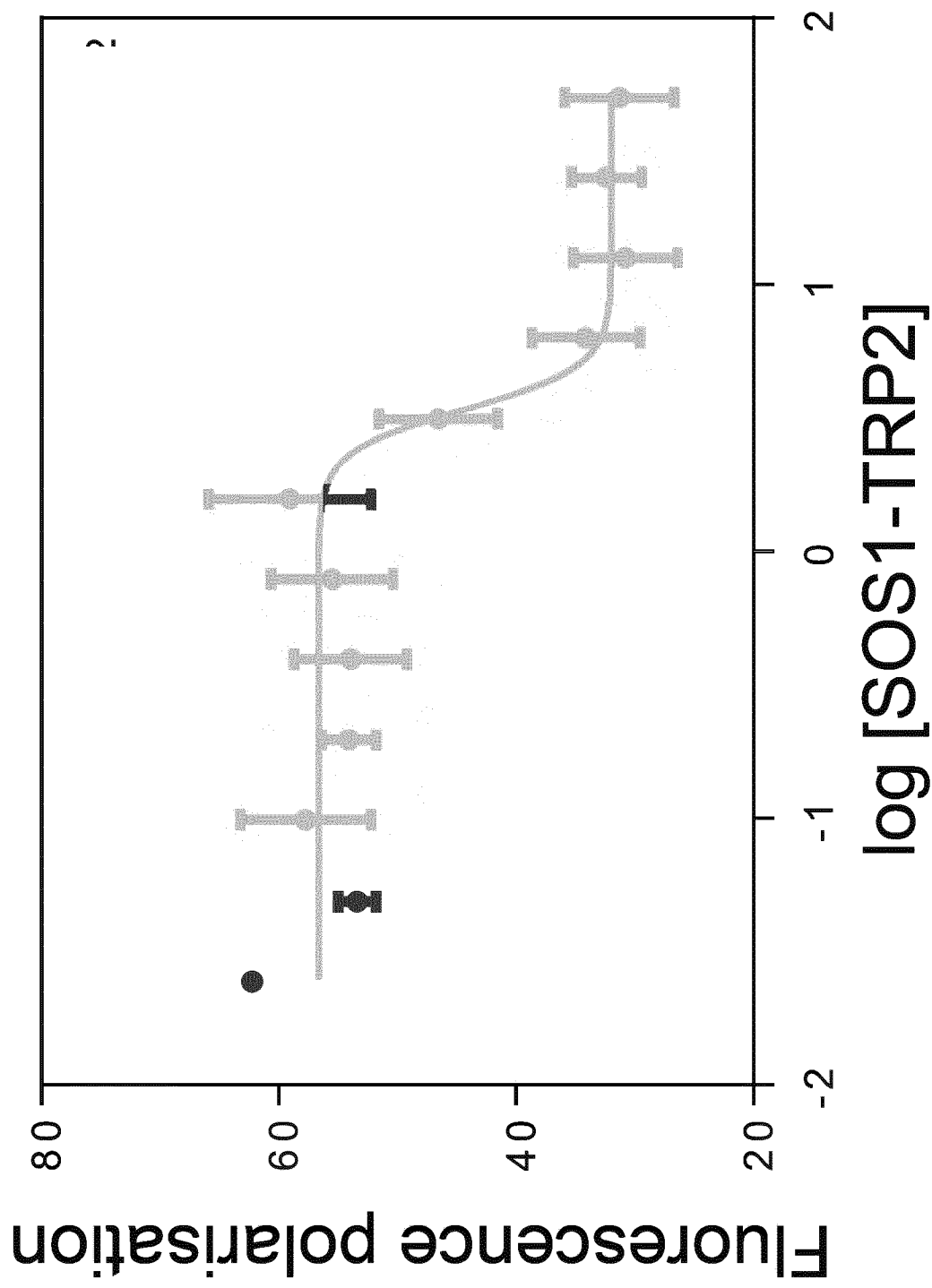


Figure 3B

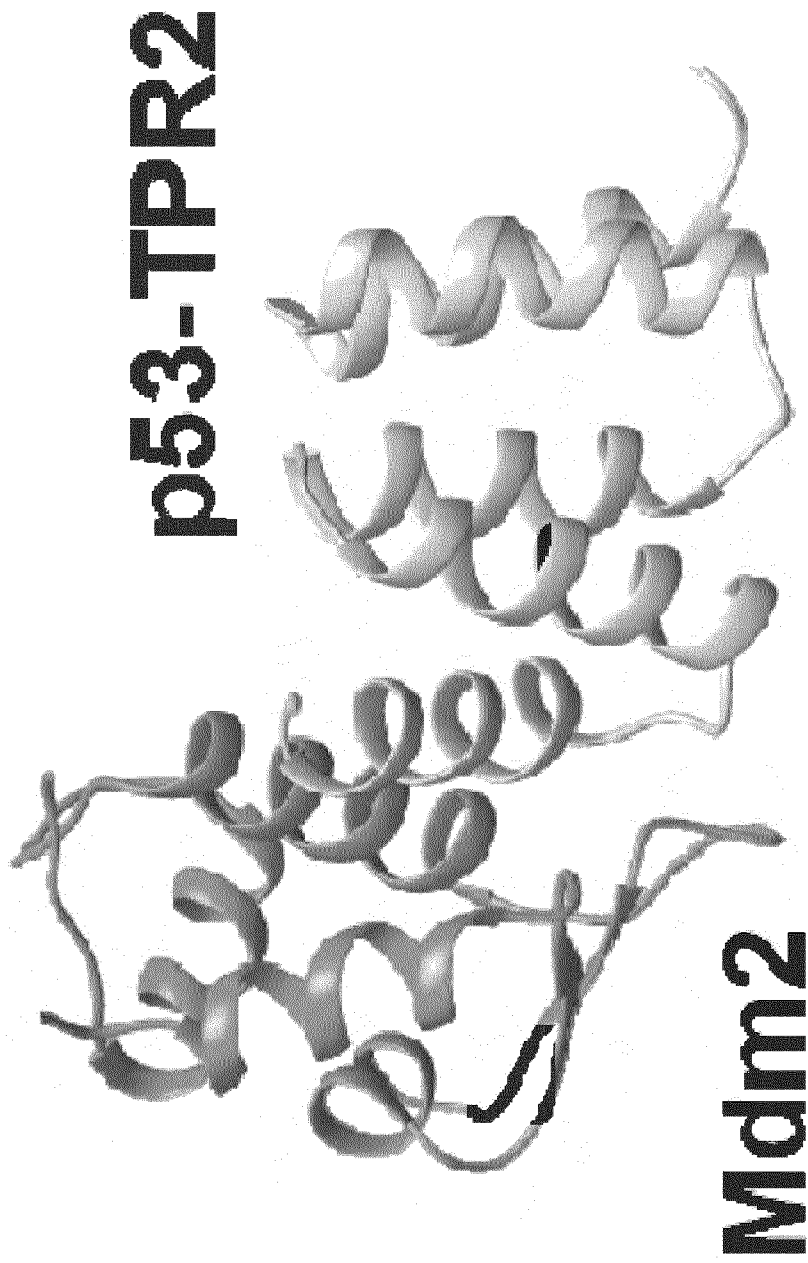


Figure 4A

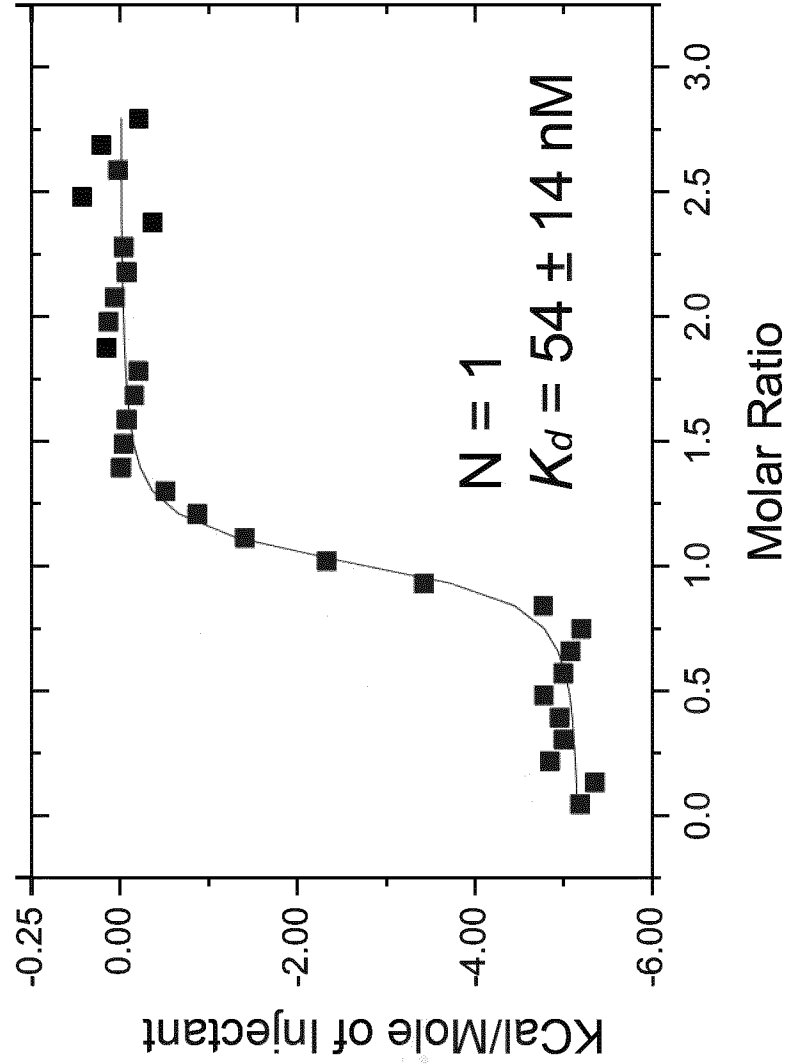


Figure 4B

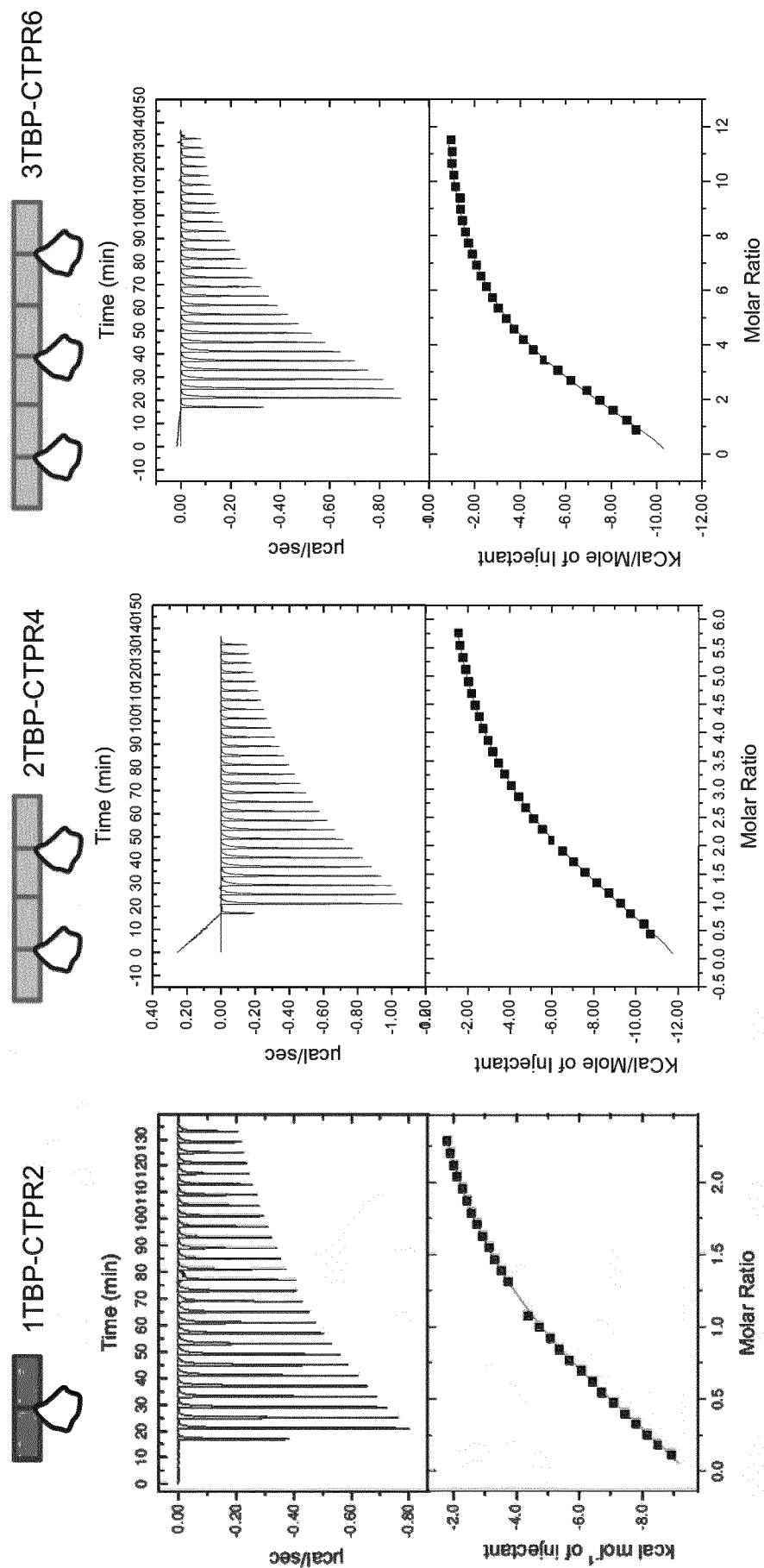


Figure 5A

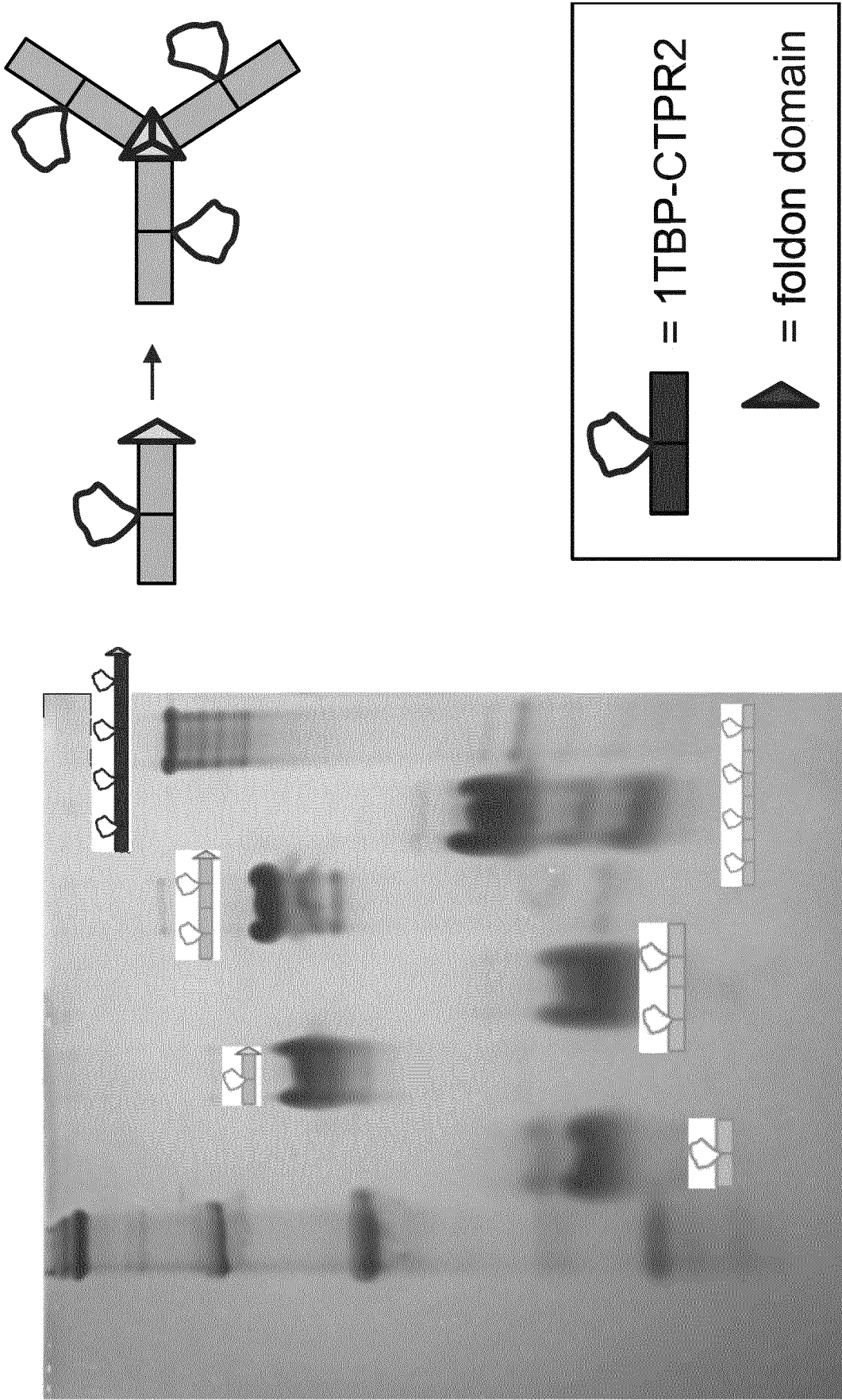


Figure 5B

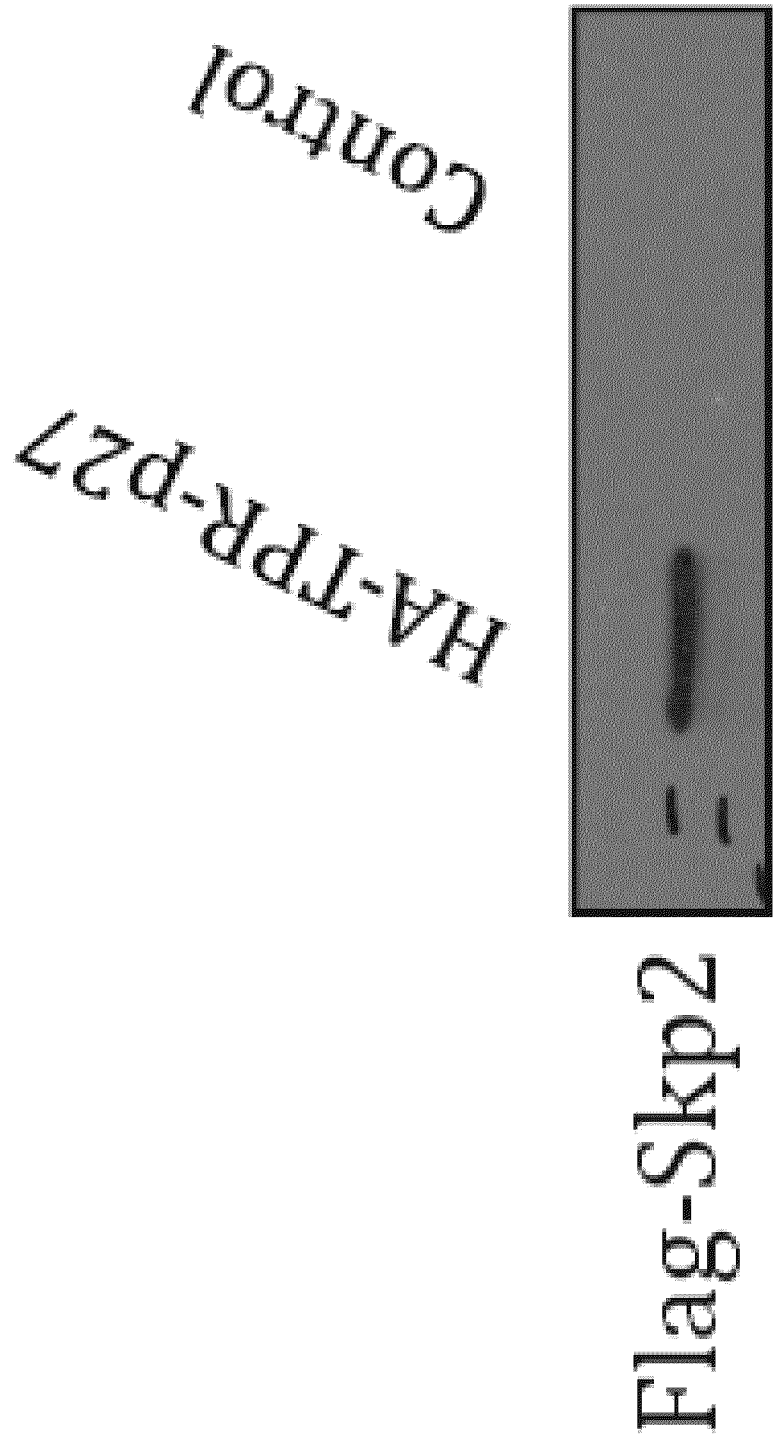


Figure 6A

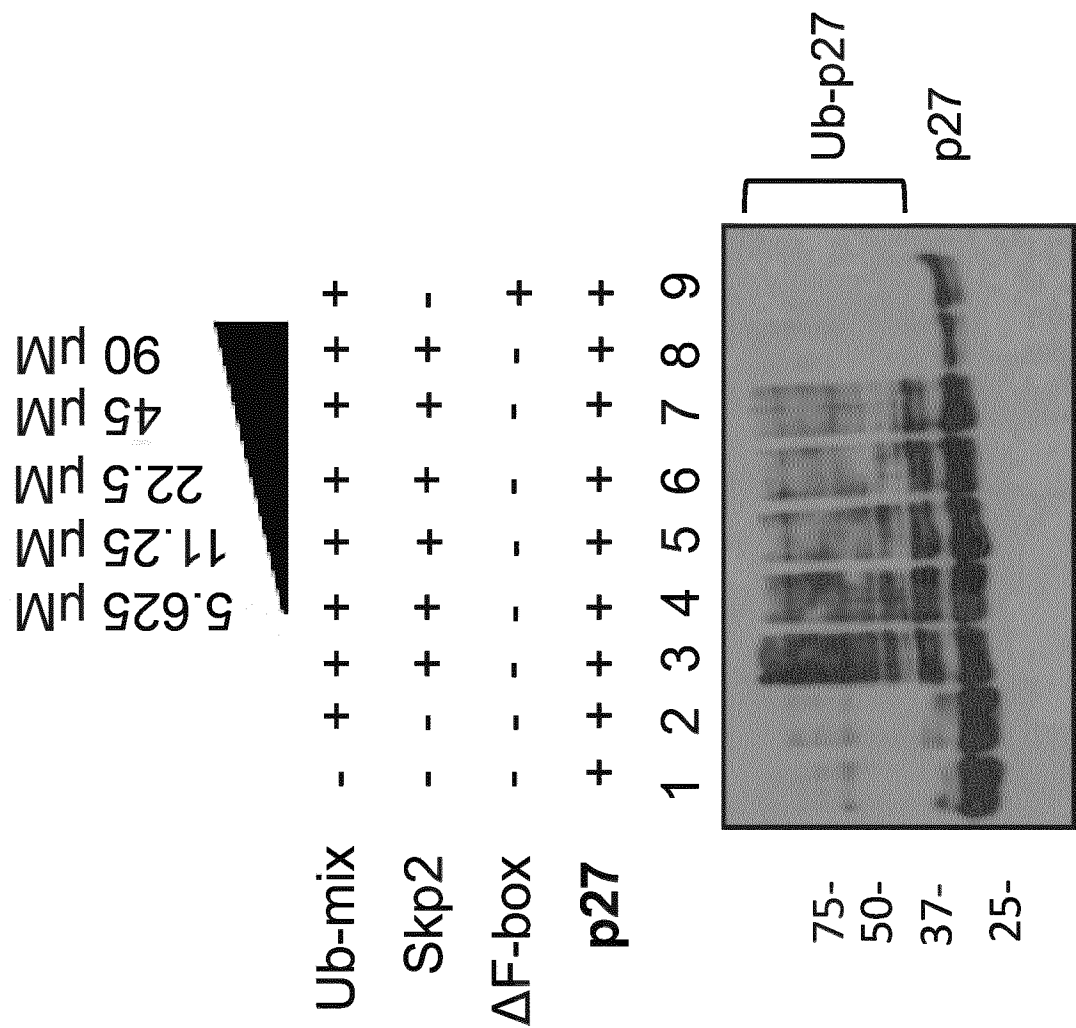


Figure 6B

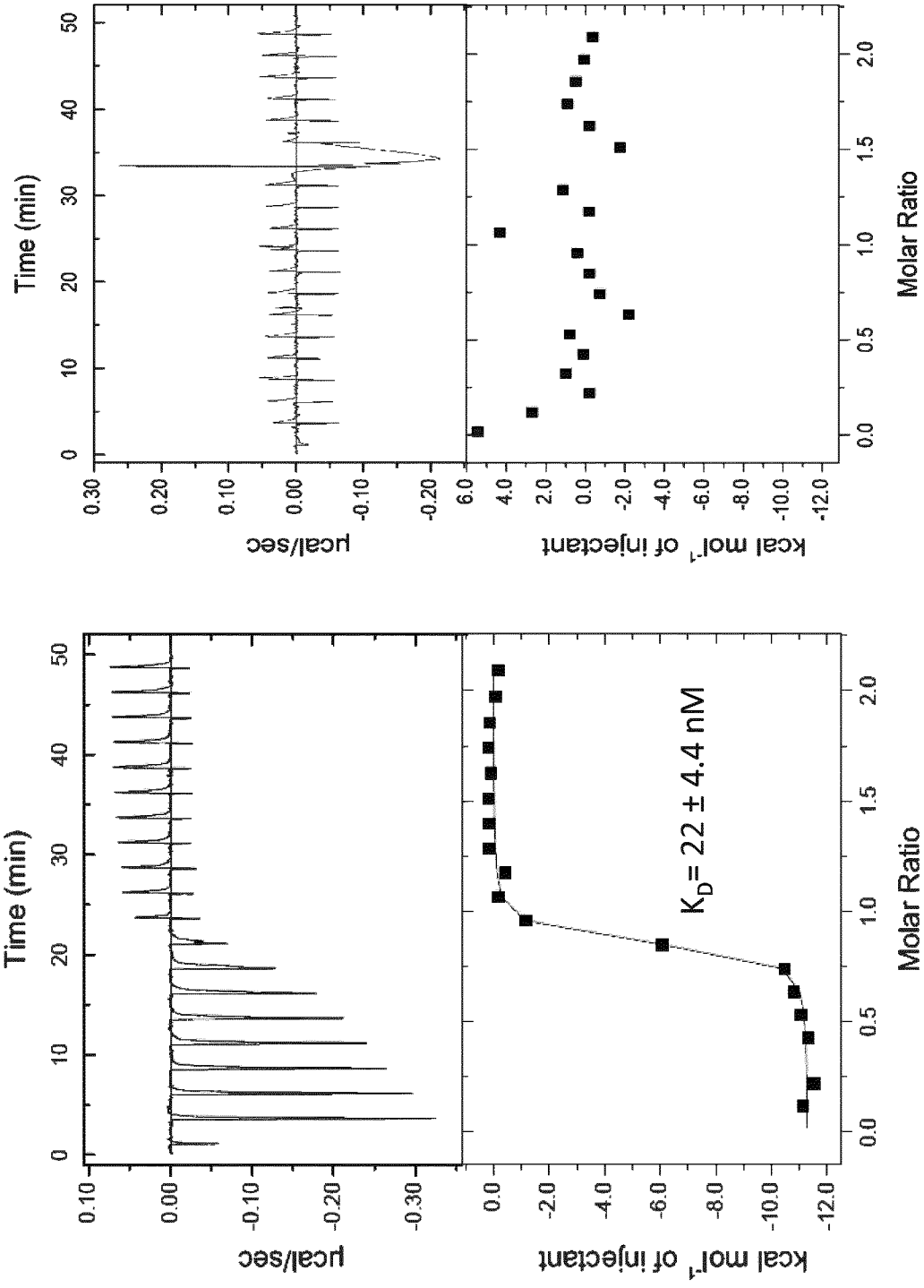


Figure 7A

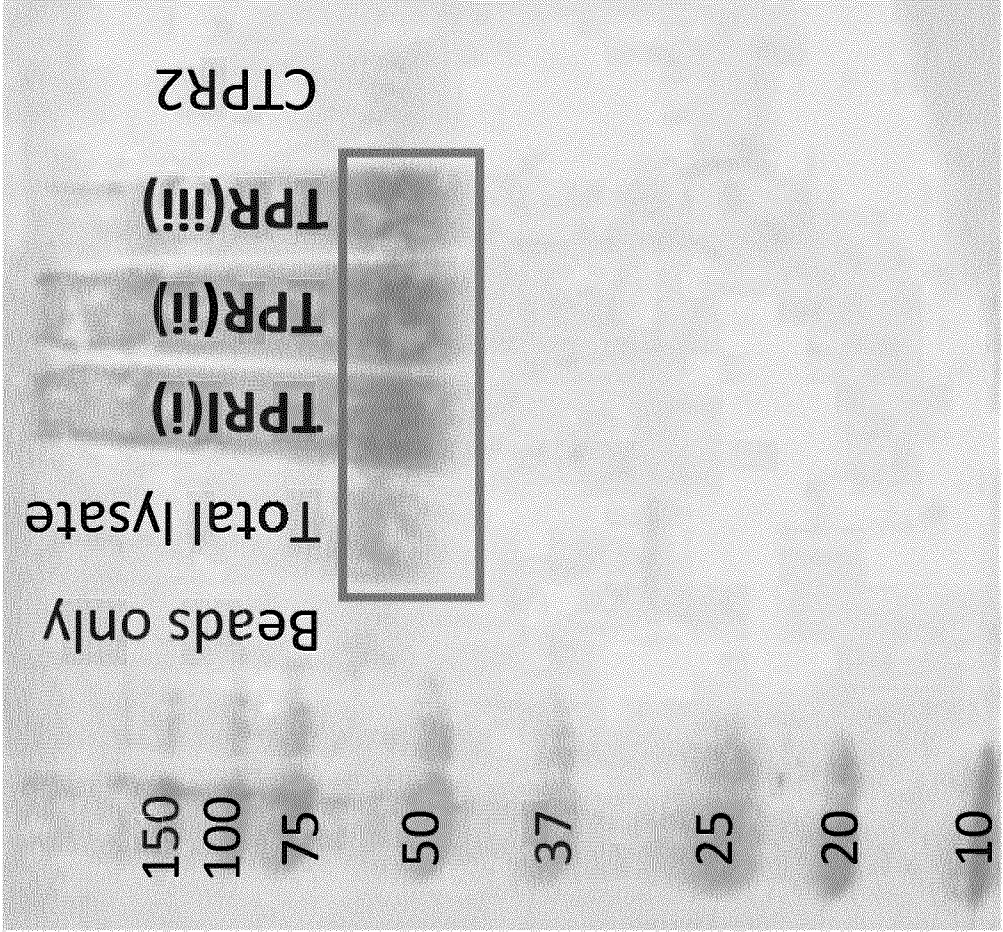


Figure 7B

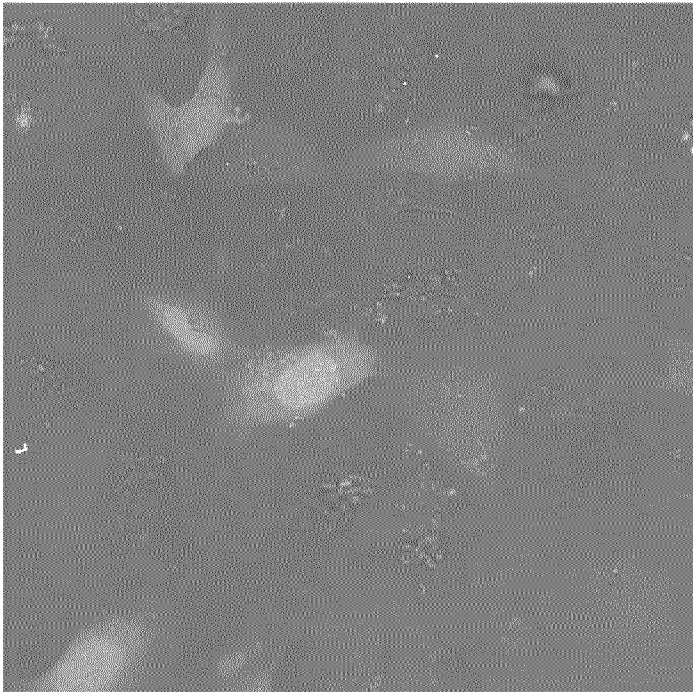
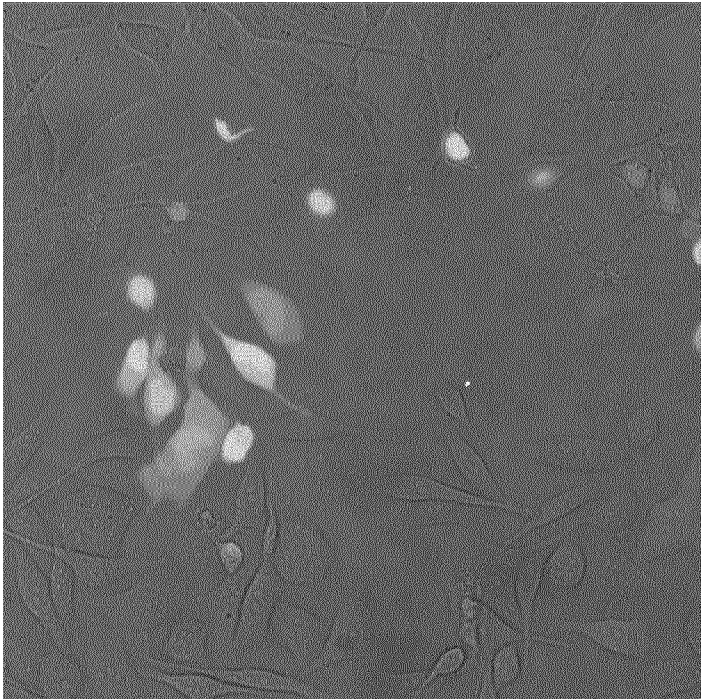


Figure 8



LRH1-TPR-p27	-	-	-	-	+	+
axin-TPR-p27	-	-	+	+	-	-
Skp2	-	+	-	+	-	+
HA-β-catenin	+	+	+	+	+	+

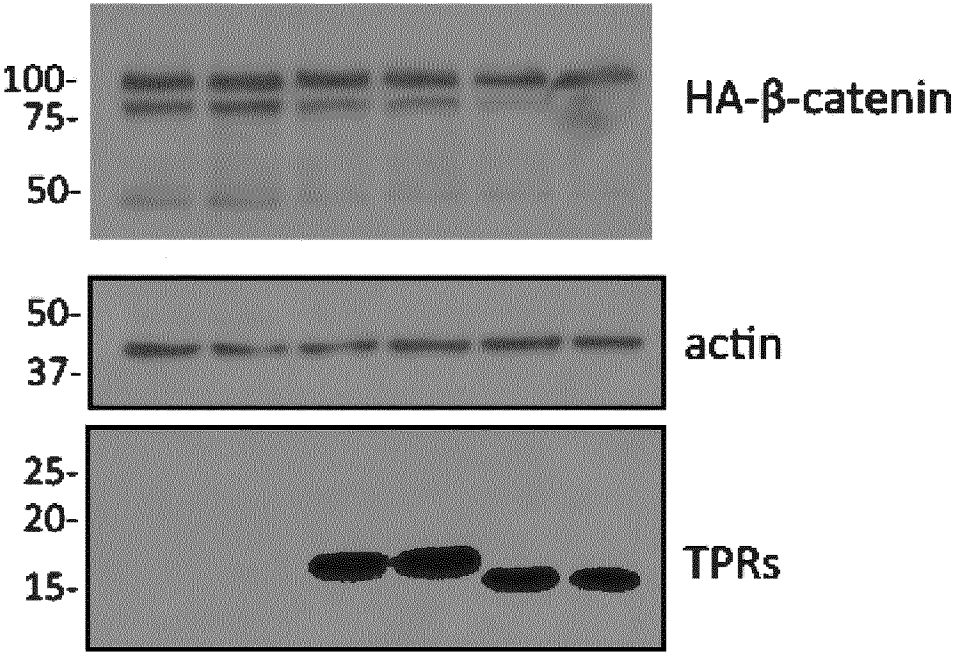


Figure 9A

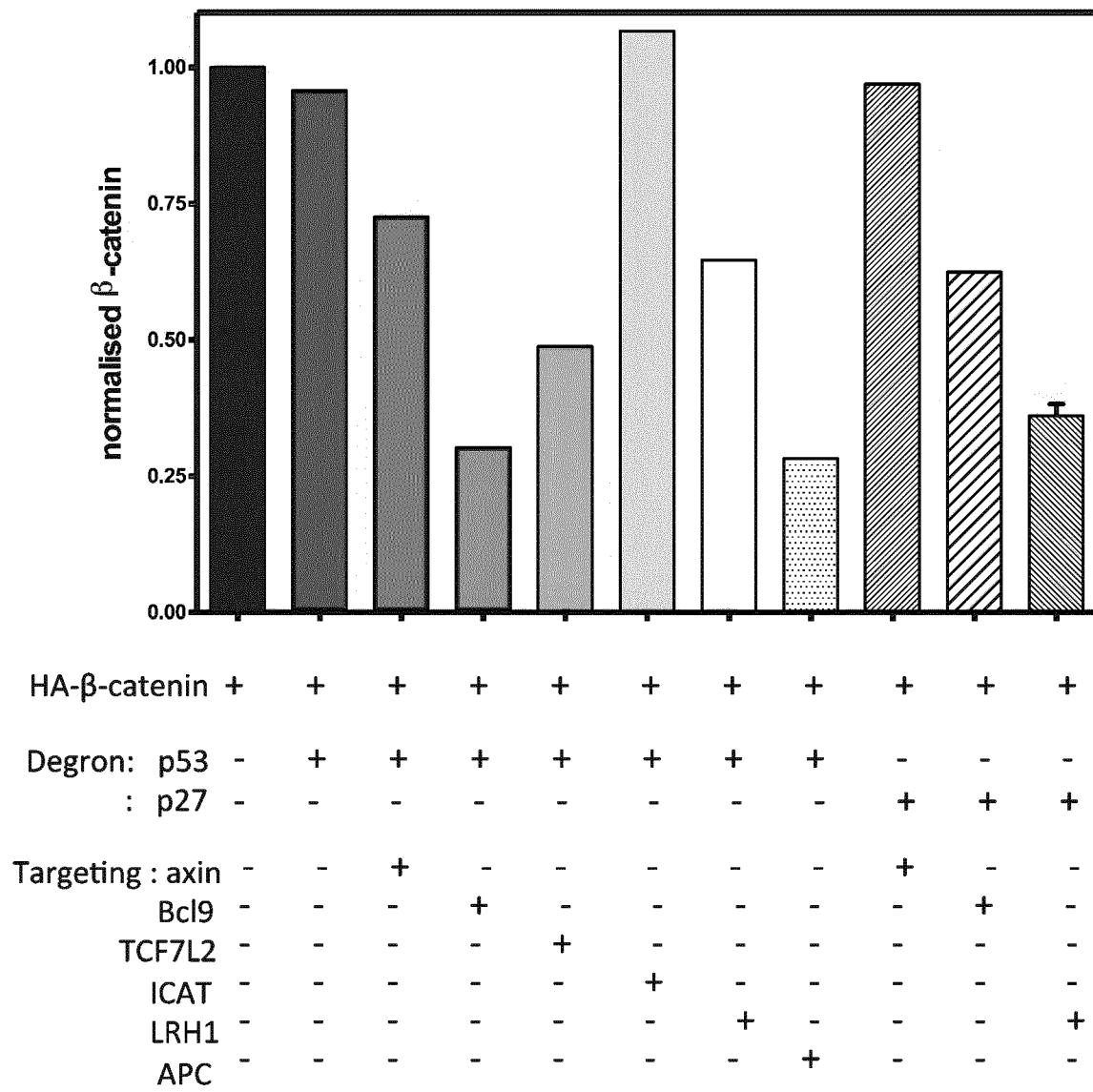


Figure 9B

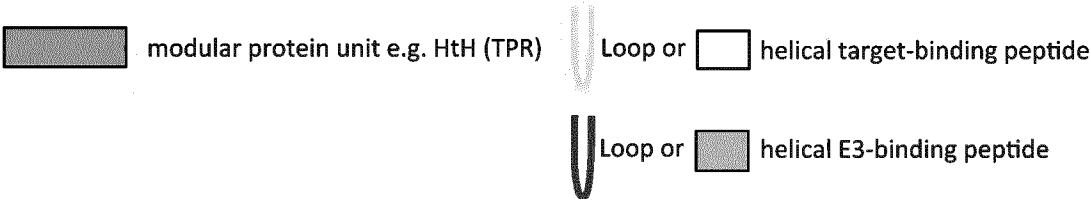


Figure 10A



Figure 10B

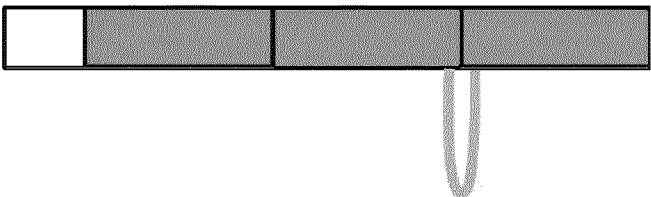


Figure 10C

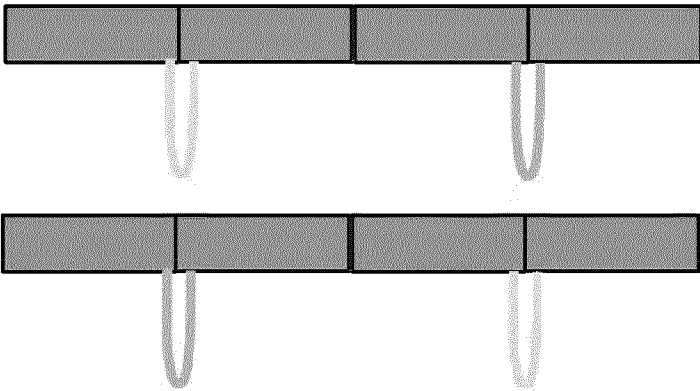


Figure 10D

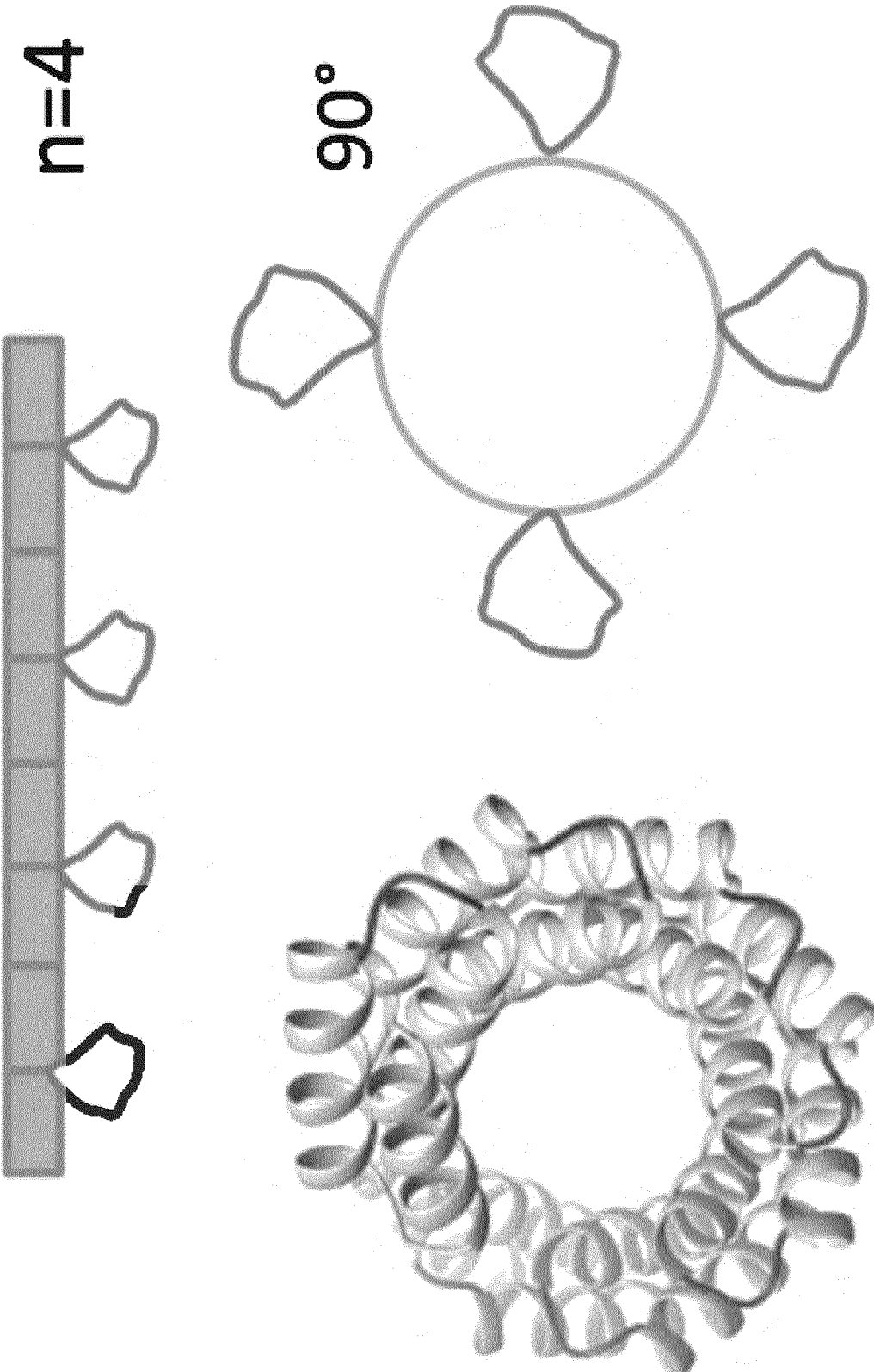


Figure 11

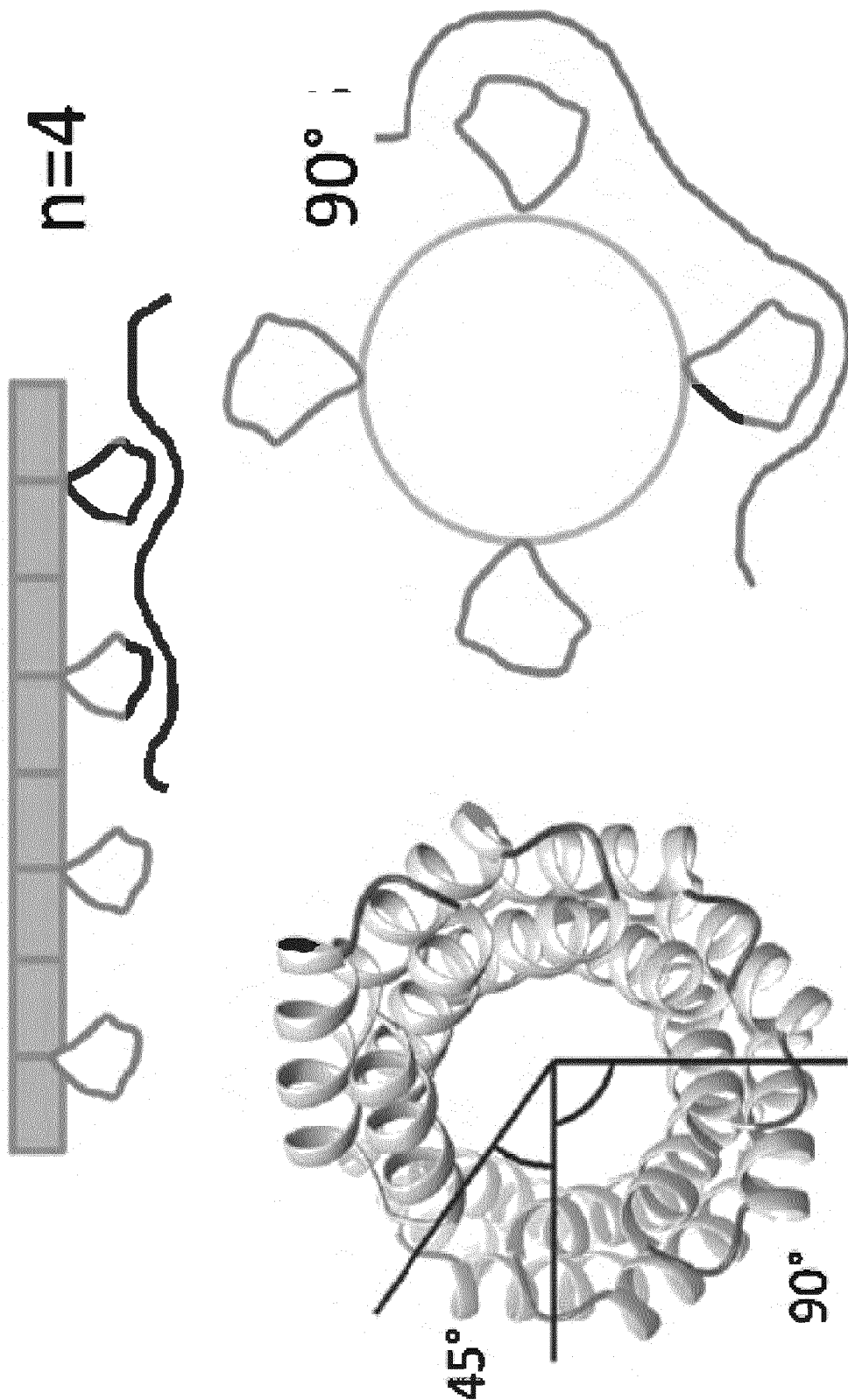


Figure 12



Figure 13

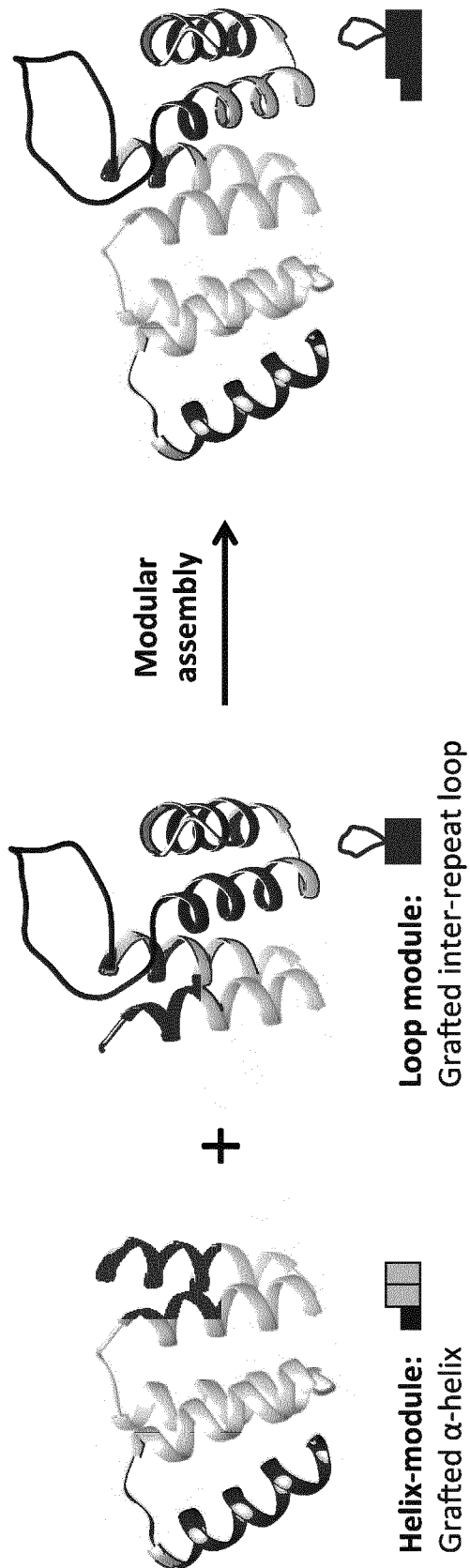


Figure 14

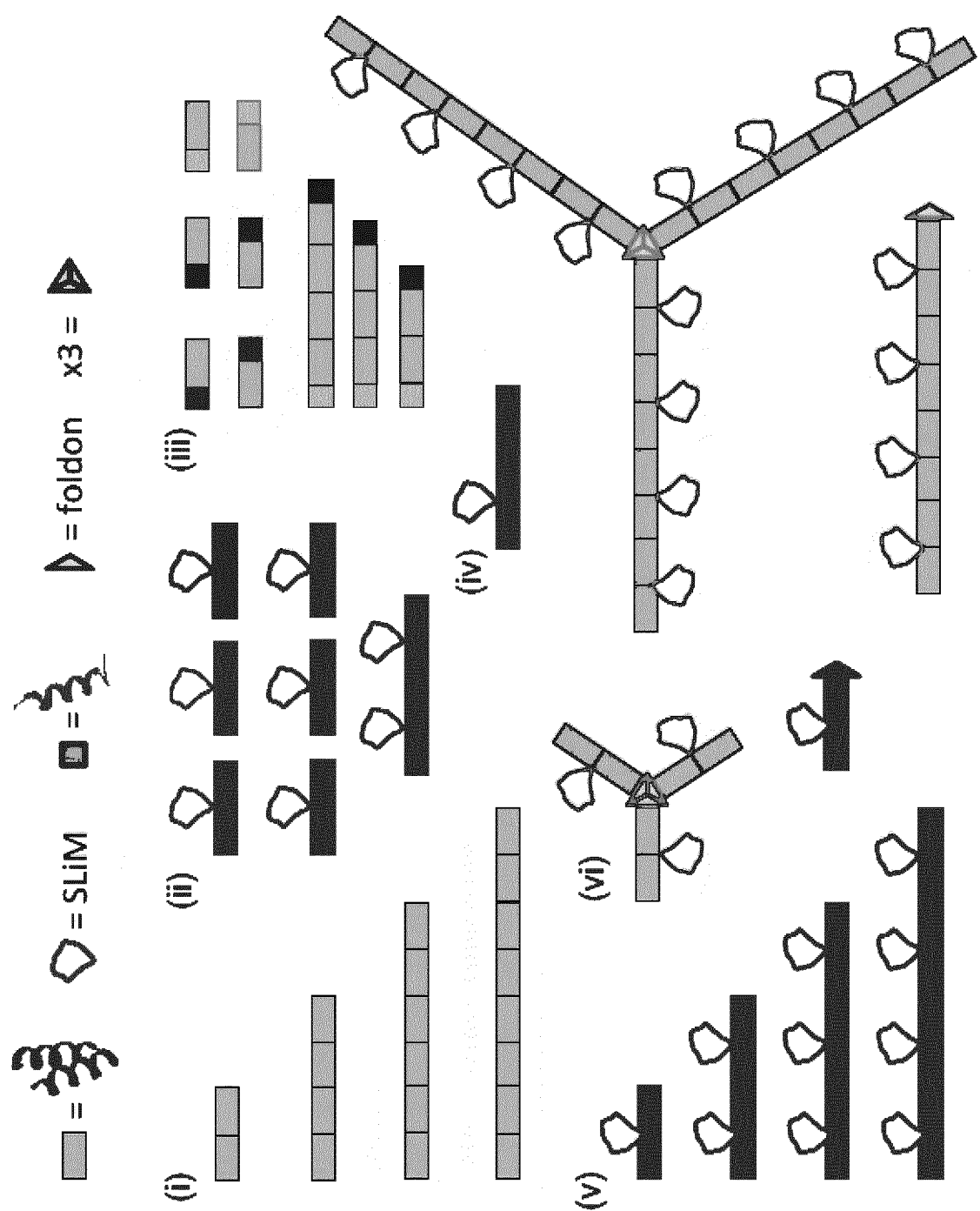


Figure 15

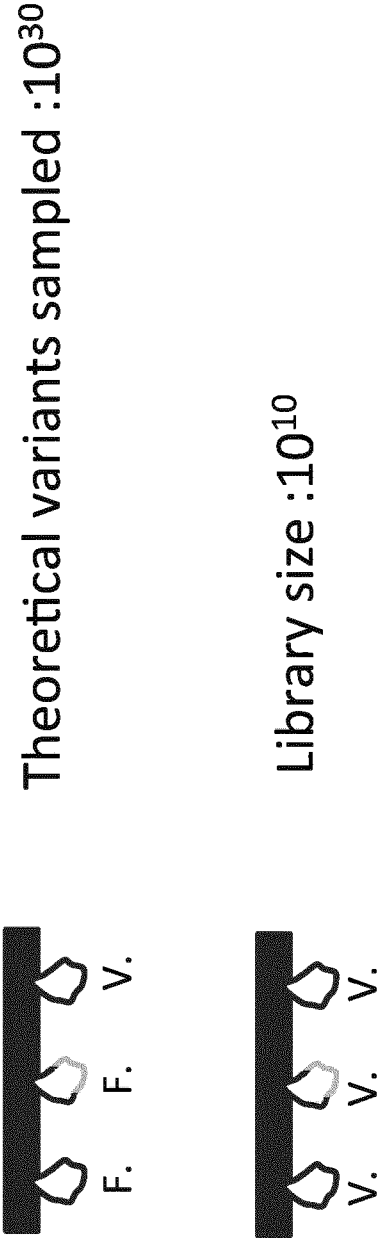
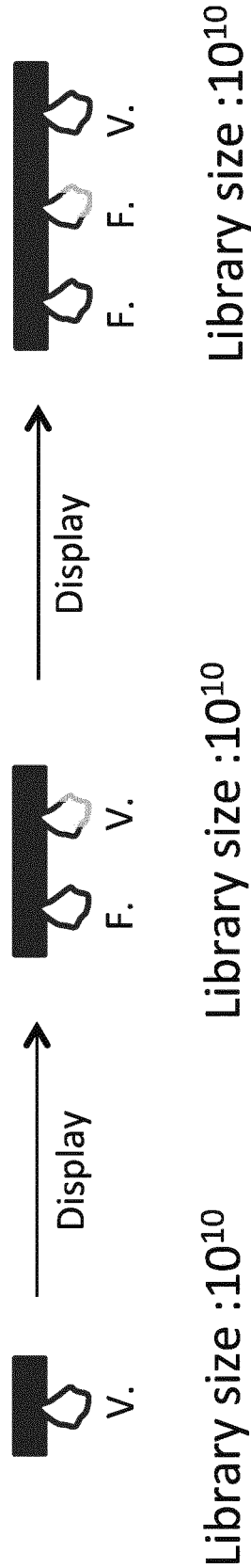


Figure 16

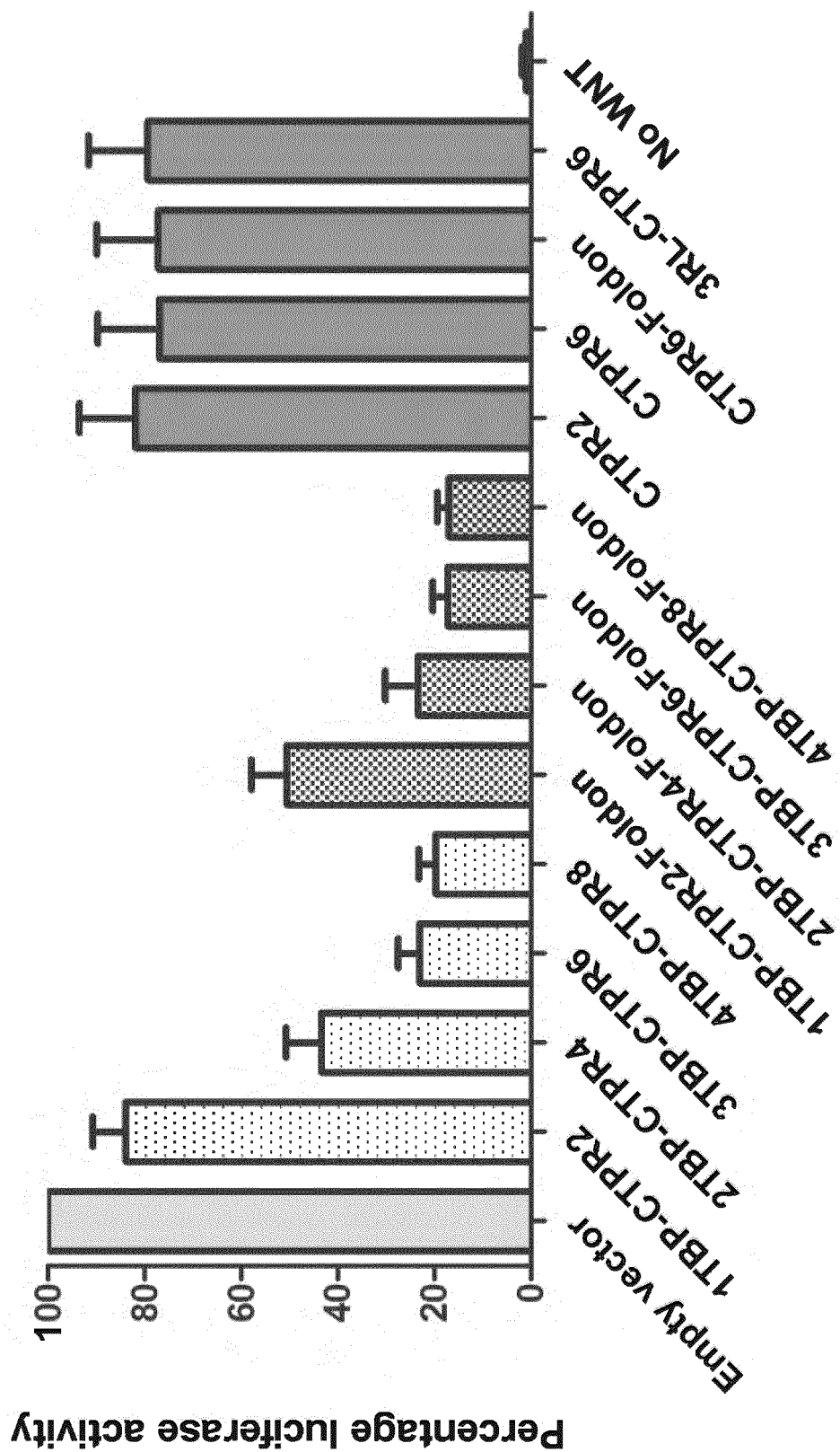
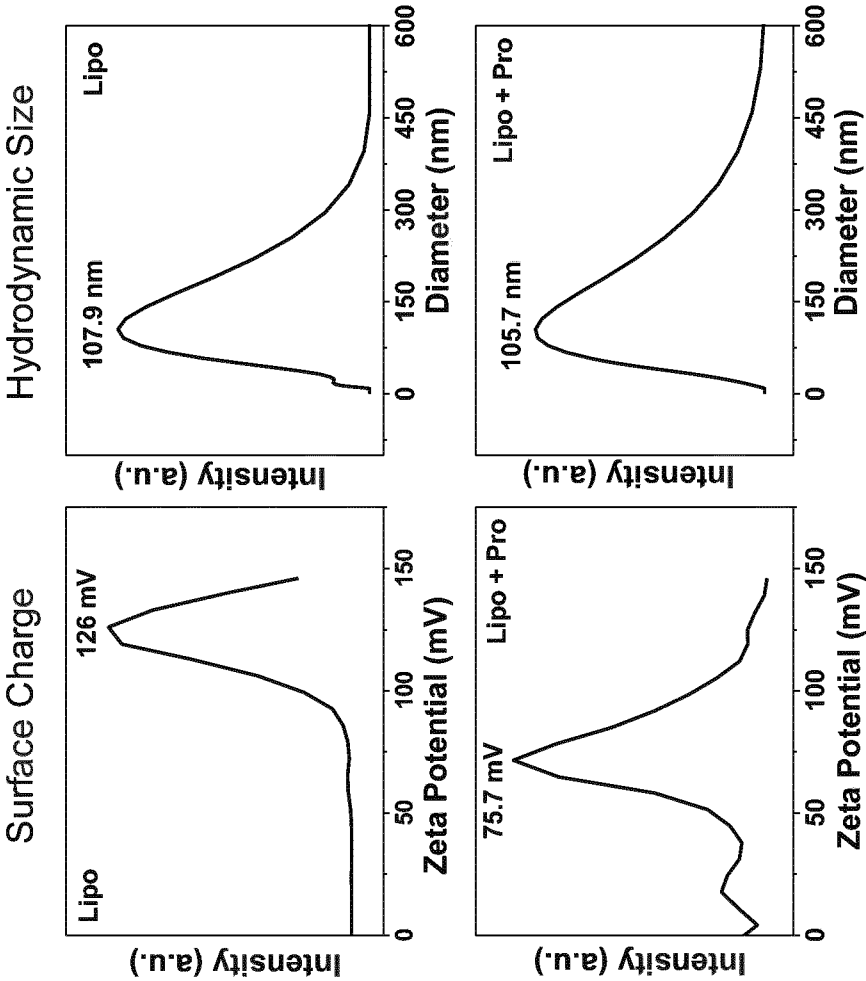


Figure 17



Lipo – Liposomes
Pro – Protein

Figure 18

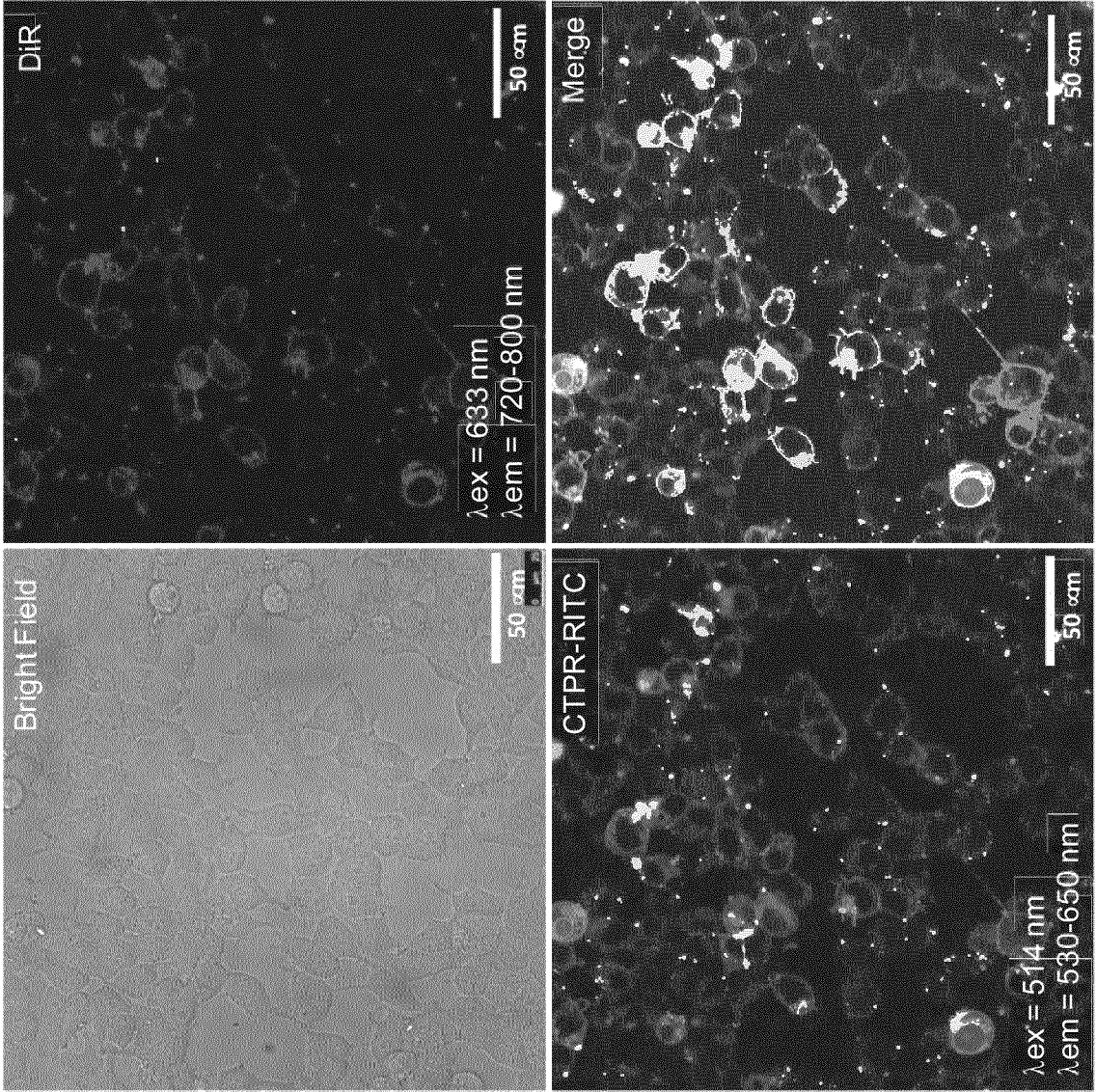


Figure 19

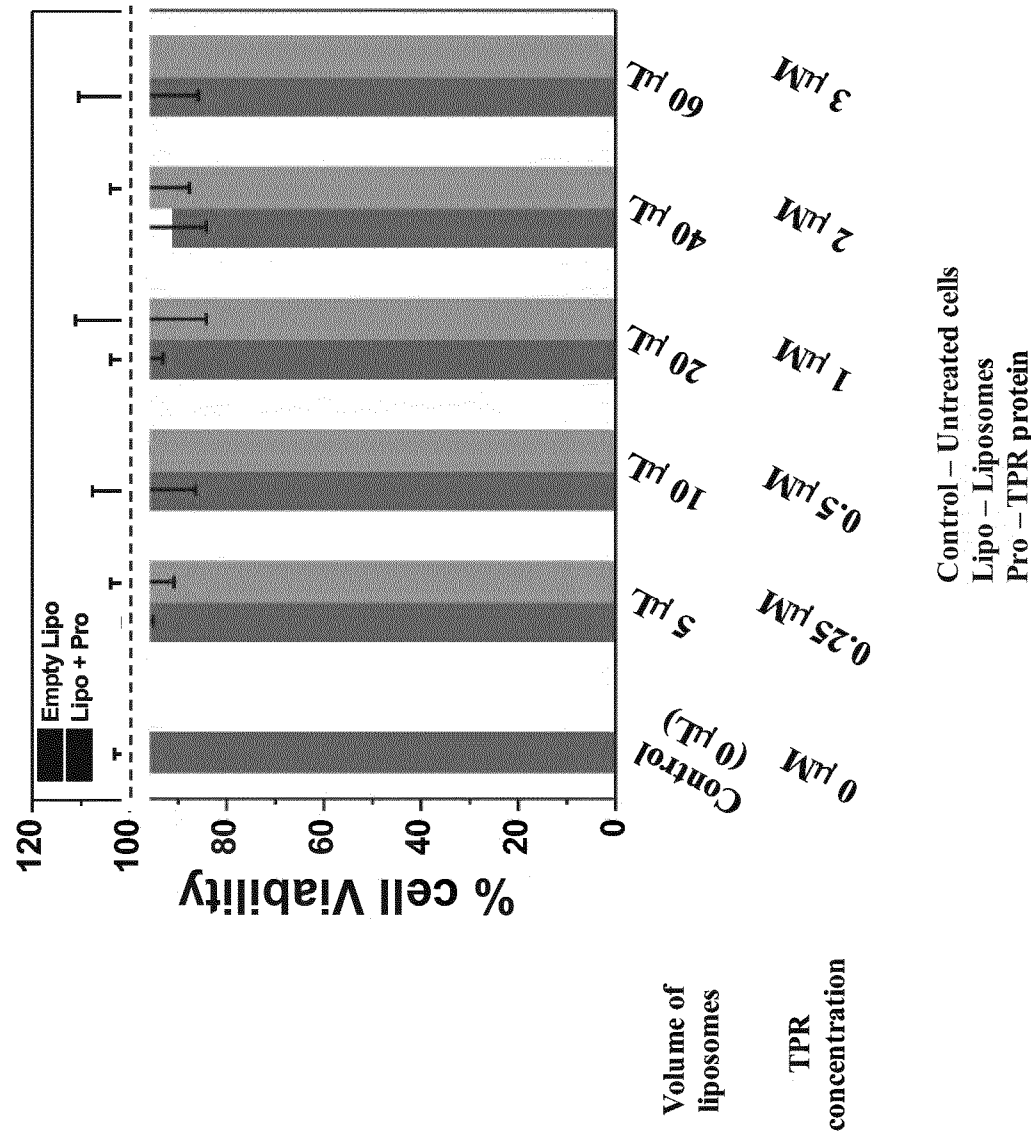


Figure 20

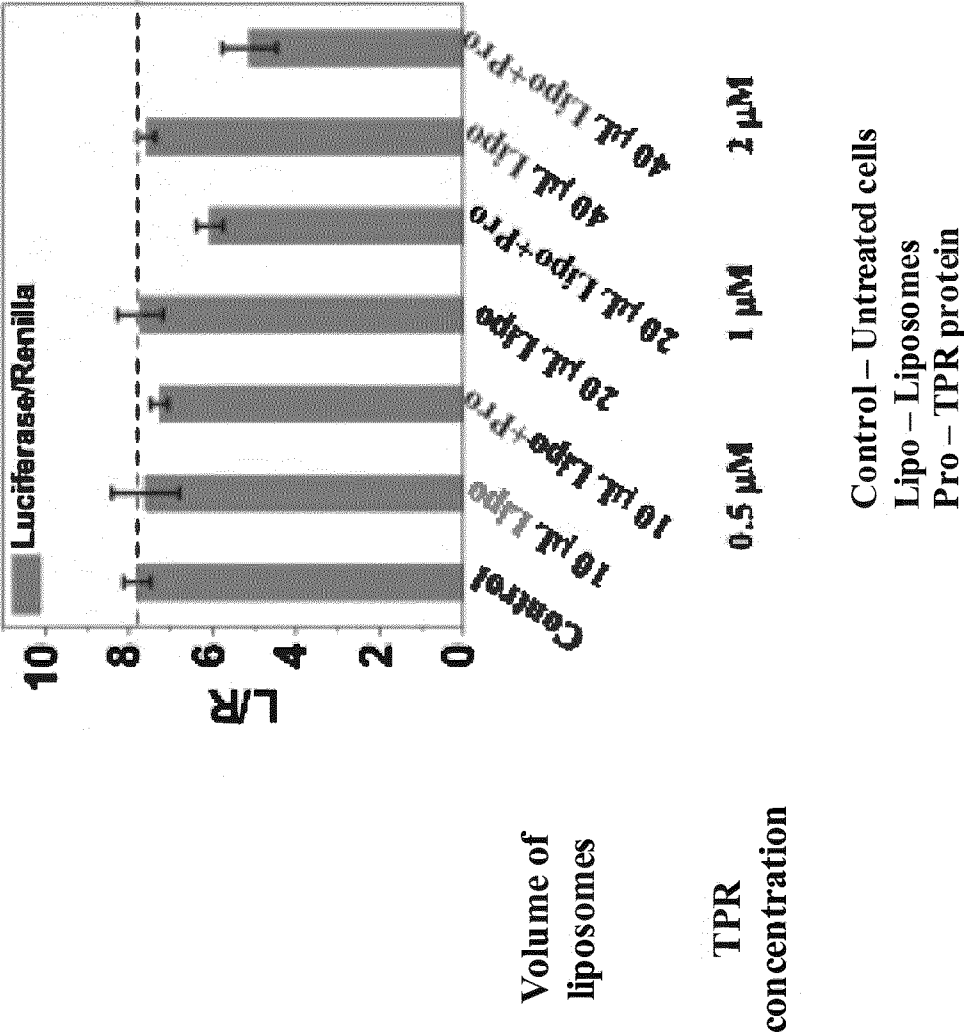


Figure 21

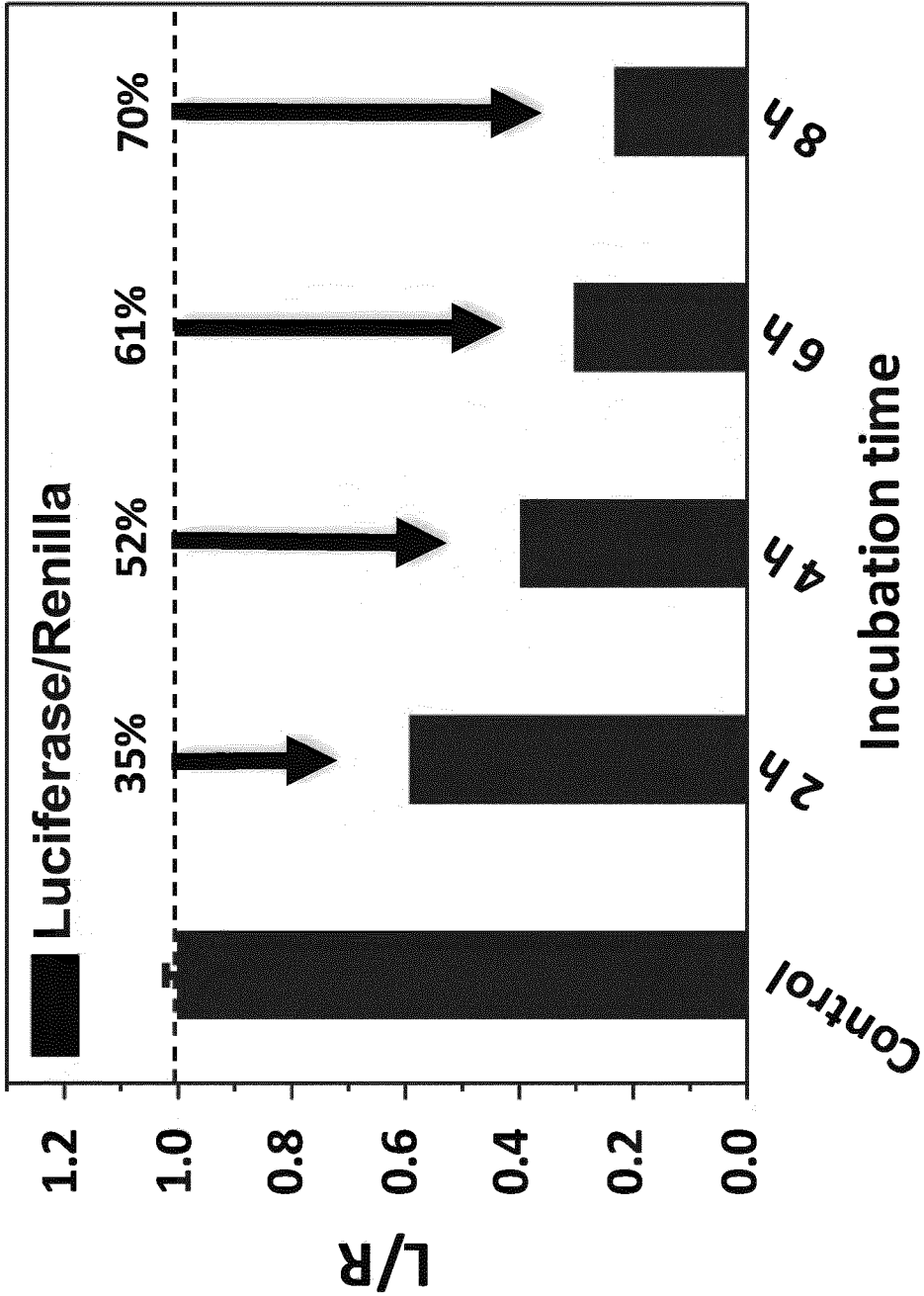


Figure 22

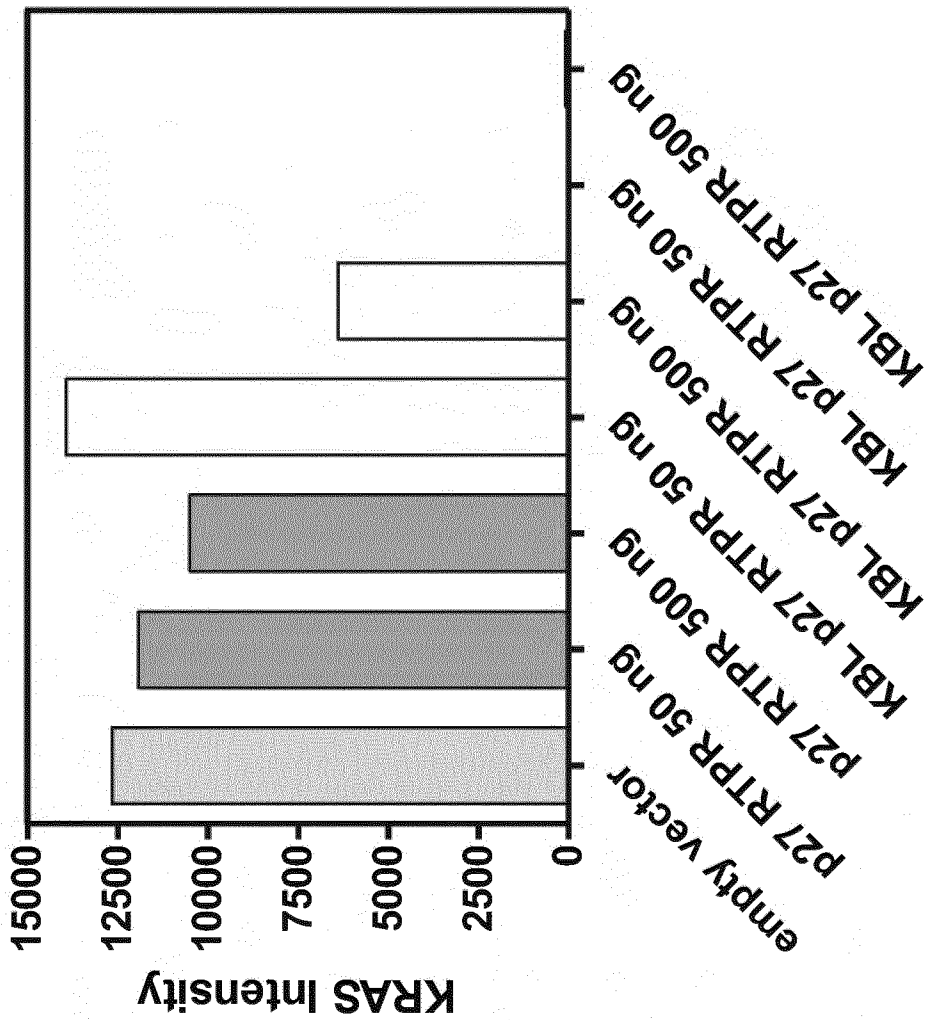


Figure 23

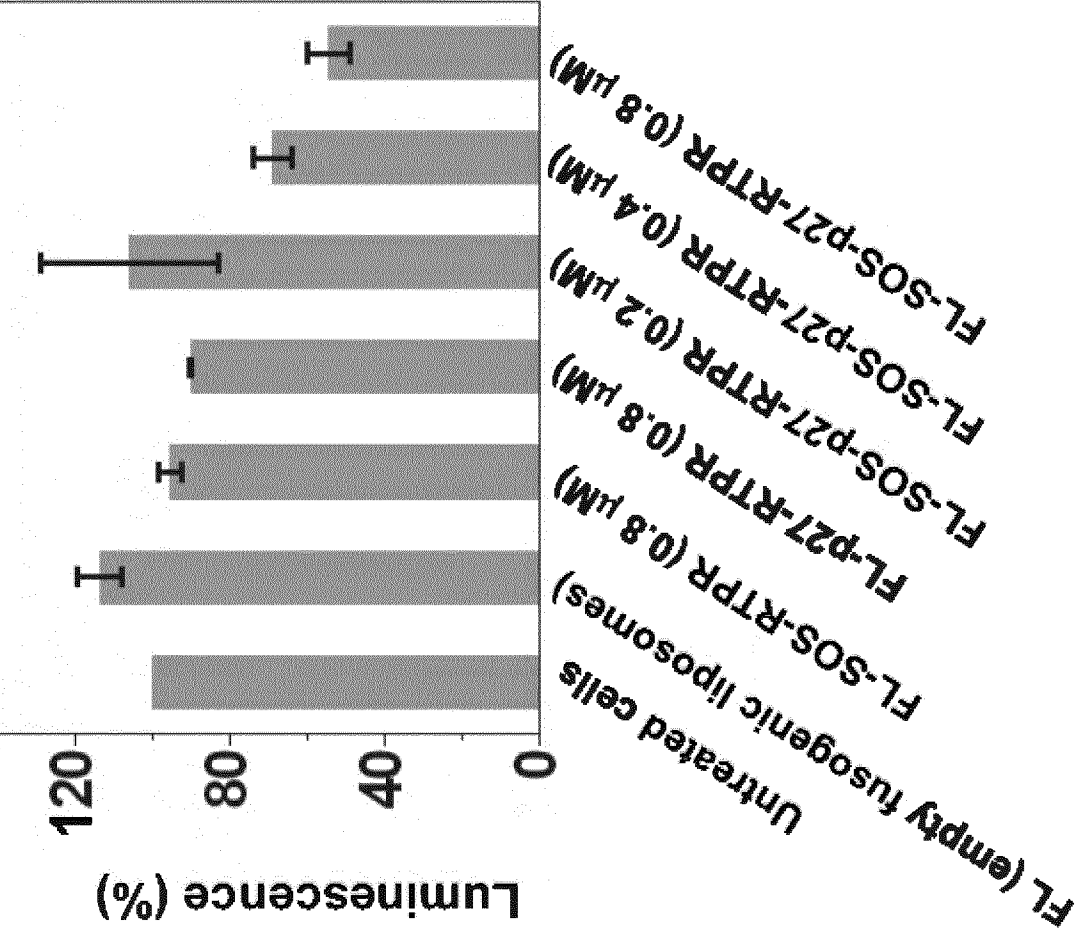


Figure 24

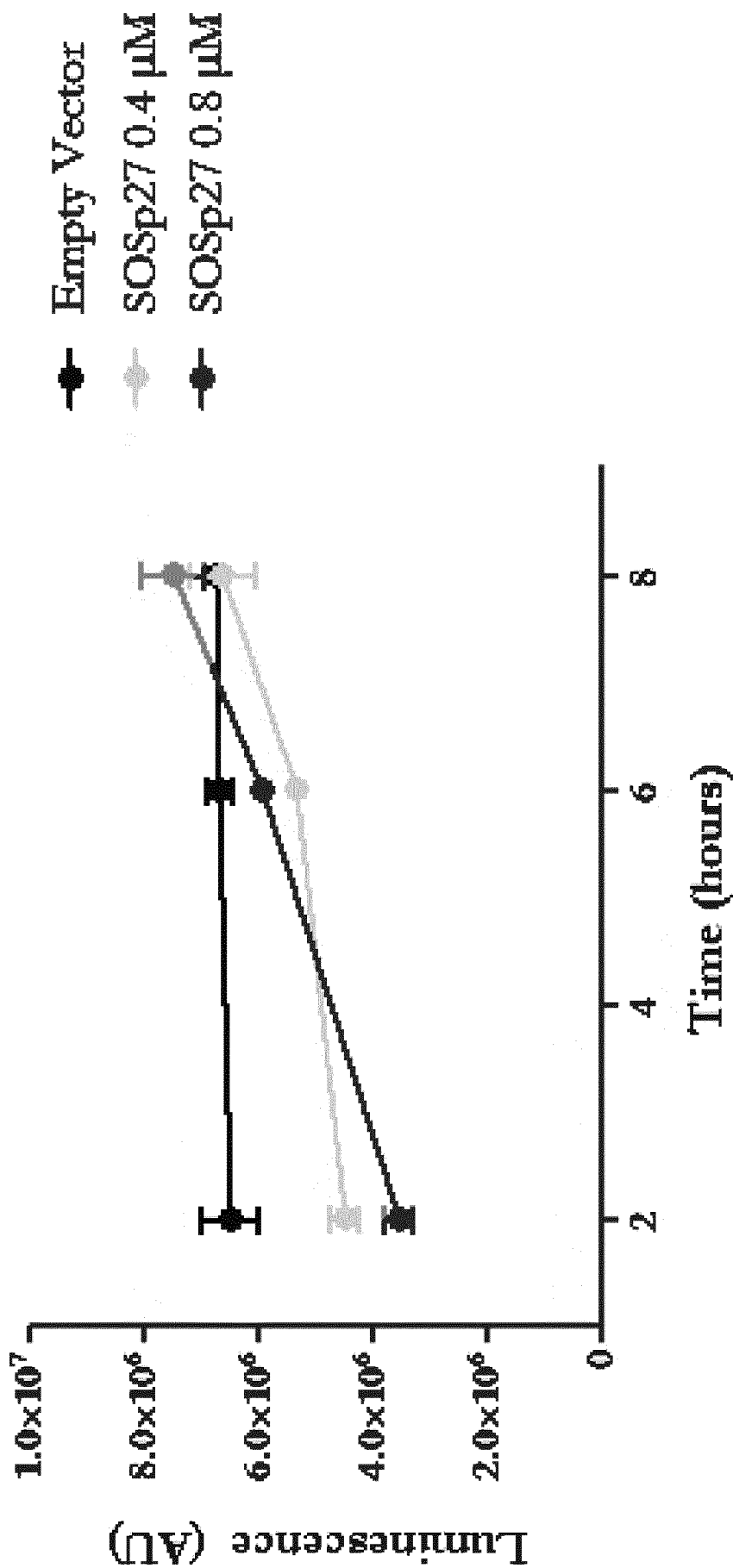


Figure 25

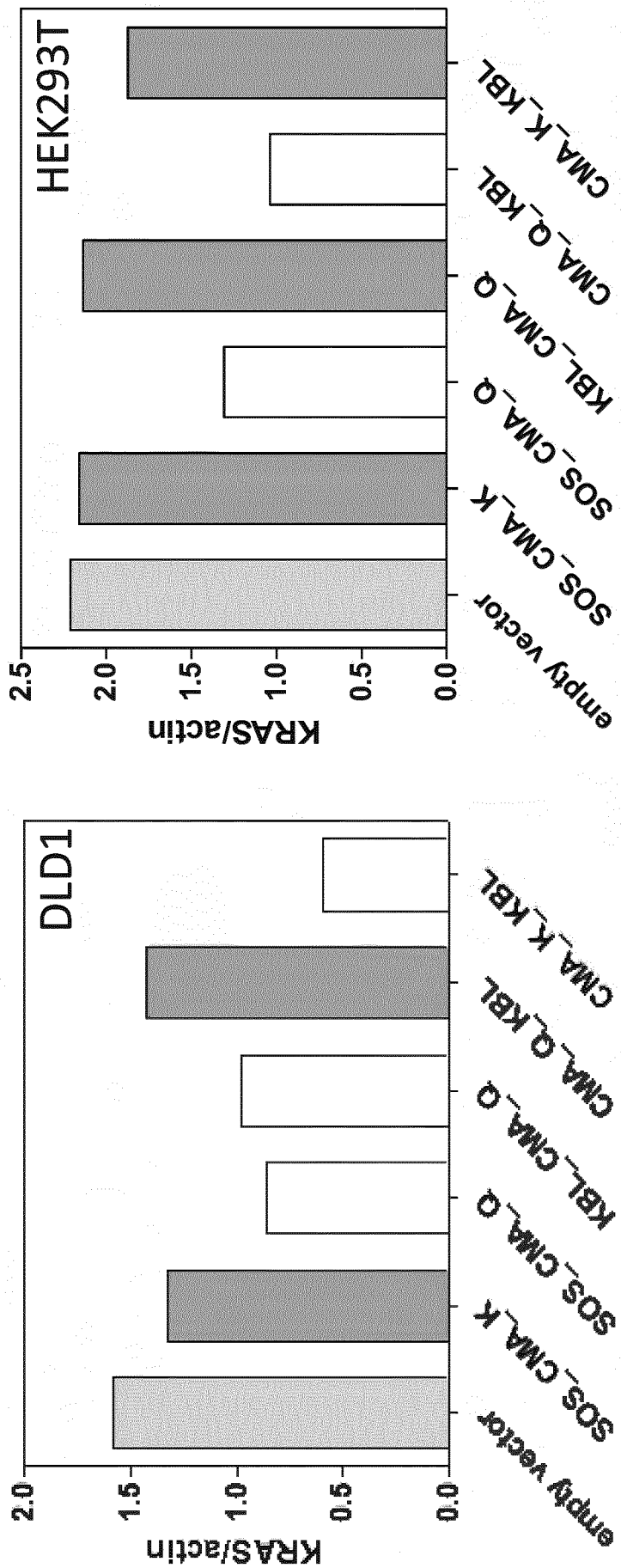
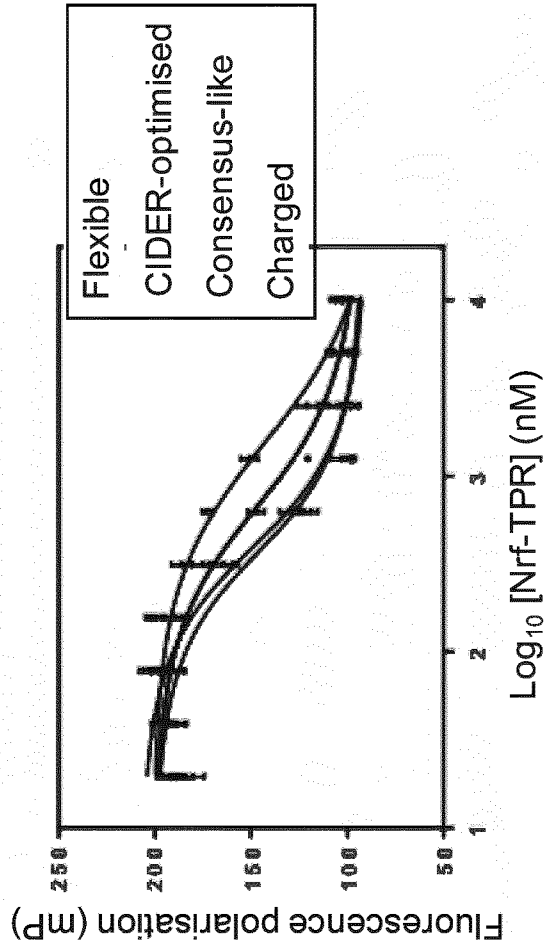
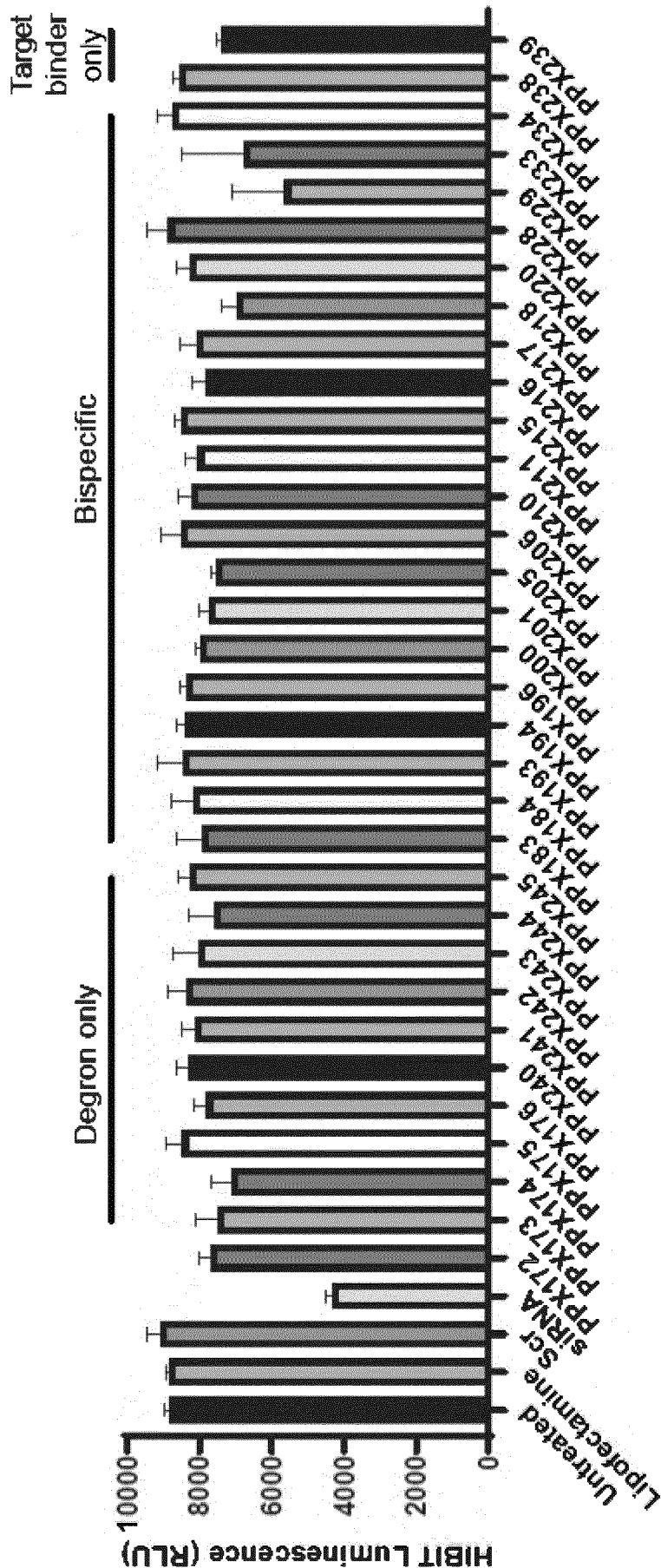


Figure 26



Composition of linker sequence in Nrf-TPR	$K_i \pm SE$ (nM)
Consensus-like	282 ± 64
Flexible	109 ± 18
Charged	817 ± 75
CIDER-optimised	103 ± 15

Figure 27



Sample

Figure 28

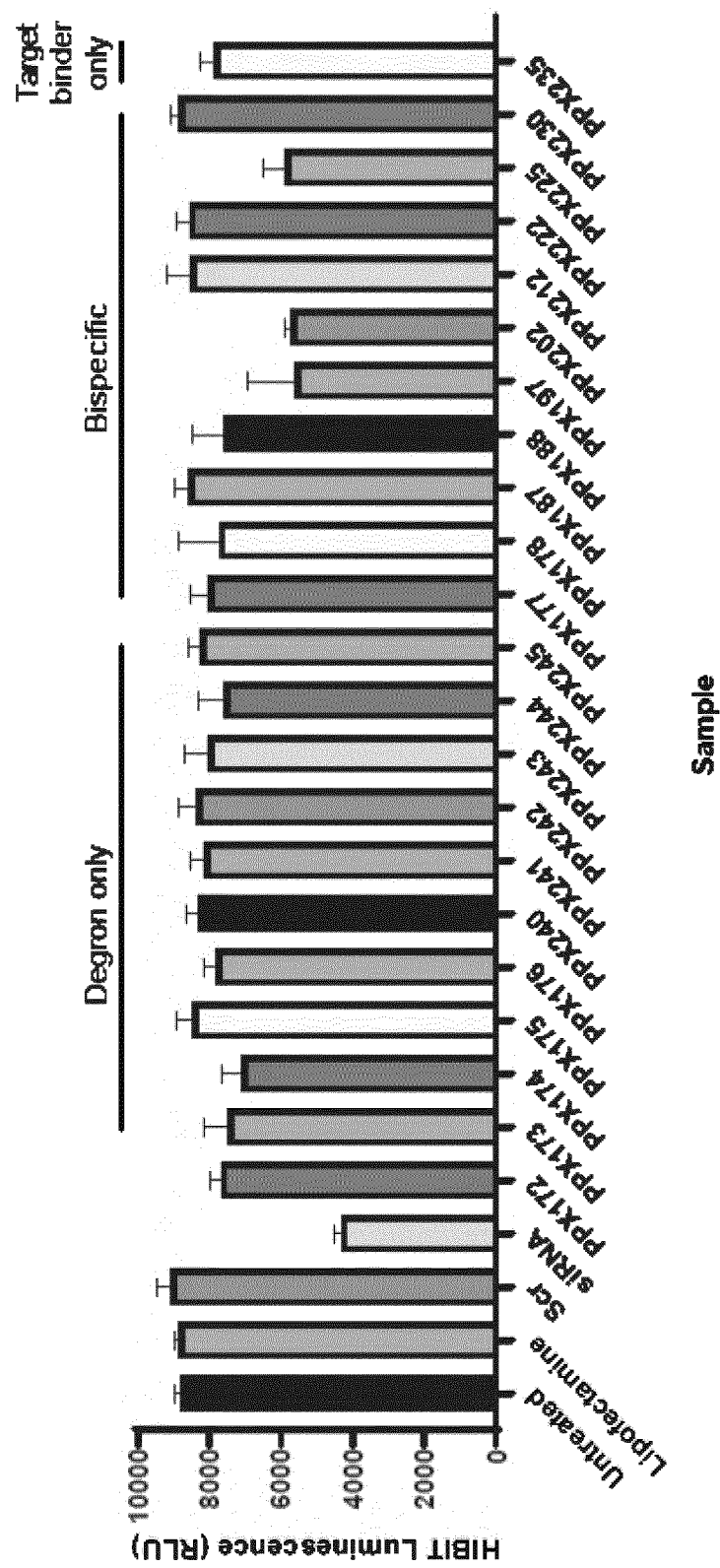


Figure 29

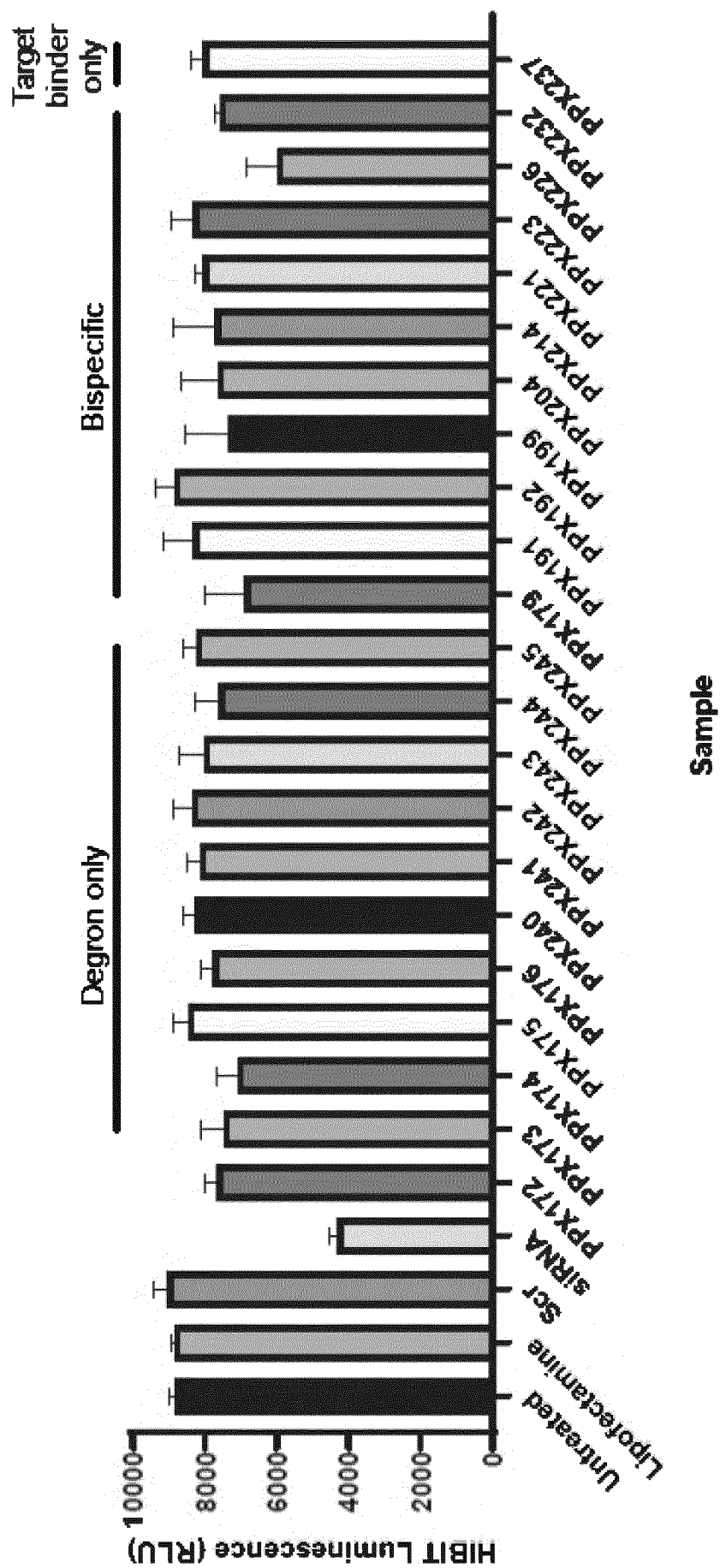


Figure 30

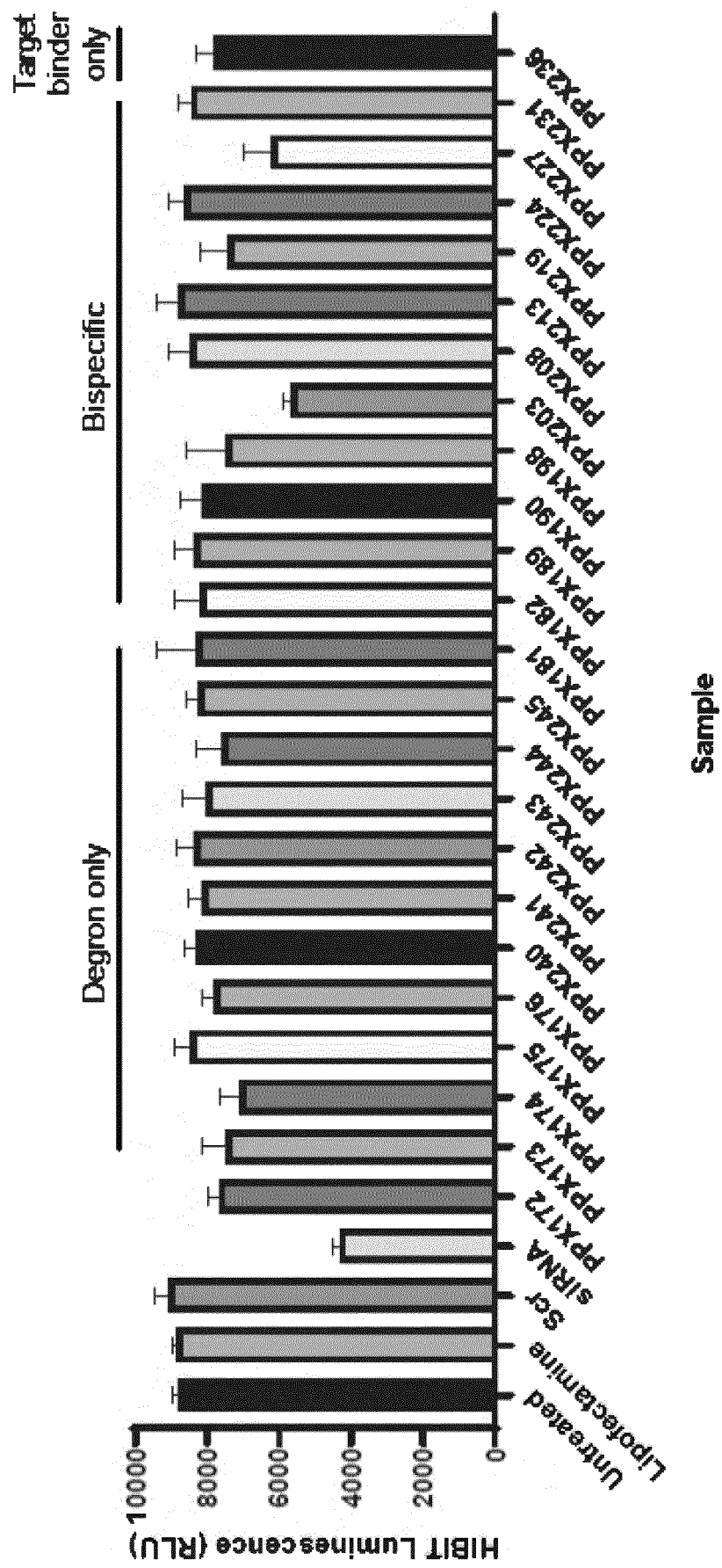


Figure 31

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2020/054697

A. CLASSIFICATION OF SUBJECT MATTER
 INV. C12N15/10
 ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 C12N C40B

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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ANDREAS PL?CKTHUN: "Designed Ankyrin Repeat Proteins (DARPs): Binding Proteins for Research, Diagnostics, and Therapy", ANNUAL REVIEW OF PHARMACOLOGY AND TOXICOLOGY, vol. 55, no. 1, 6 January 2015 (2015-01-06), pages 489-511, XP055217451, ISSN: 0362-1642, DOI: 10.1146/annurev-pharmtox-010611-134654 the whole document ----- -/--	1-68



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

28 April 2020

Date of mailing of the international search report

01/07/2020

Name and mailing address of the ISA/

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

Authorized officer

Keller, Yves

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2020/054697

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	STUMPP M T ET AL: "DARPin: A new generation of protein therapeutics", DRUG DISCOVERY TODAY, ELSEVIER, AMSTERDAM, NL, vol. 13, no. 15-16, 1 August 2008 (2008-08-01), pages 695-701, XP023440383, ISSN: 1359-6446, DOI: 10.1016/J.DRUDIS.2008.04.013 [retrieved on 2008-07-11] the whole document	1-68
X,P	WO 2019/034332 A1 (CAMBRIDGE ENTPR LTD [GB]) 21 February 2019 (2019-02-21) the whole document	1-68
X	BINZ H ET AL: "Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins", JOURNAL OF MOLECULAR BIOLOGY, ACADEMIC PRESS, UNITED KINGDOM, vol. 332, no. 2, 12 September 2003 (2003-09-12), pages 489-503, XP027101280, ISSN: 0022-2836, DOI: 10.1016/S0022-2836(03)00896-9 [retrieved on 2003-09-12] the whole document	1-68
X	JOHANNES SCHILLING ET AL: "From DARPin to LoopDARPin: Novel LoopDARPin Design Allows the Selection of Low Picomolar Binders in a Single Round of Ribosome Display", JOURNAL OF MOLECULAR BIOLOGY, vol. 426, no. 3, 1 February 2014 (2014-02-01), pages 691-721, XP055689492, United Kingdom ISSN: 0022-2836, DOI: 10.1016/j.jmb.2013.10.026 the whole document	1-68
X	BINZ H KASPAR ET AL: "High-affinity binders selected from designed ankyrin repeat protein libraries", NATURE BIOTECHNOLOGY, GALE GROUP INC., NEW YORK, US, vol. 22, no. 5, 1 May 2004 (2004-05-01), pages 575-582, XP002343919, ISSN: 1087-0156, DOI: 10.1038/NBT962 the whole document	1-68
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/054697

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YKELIEN L BOERSMA ET AL: "DARPin and other repeat protein scaffolds: advances in engineering and applications", CURRENT OPINION IN BIOTECHNOLOGY, vol. 22, no. 6, 1 January 2011 (2011-01-01), pages 849-857, XP028397473, ISSN: 0958-1669, DOI: 10.1016/J.COPBIO.2011.06.004 [retrieved on 2011-06-08] the whole document -----	1-68
X	EP 3 004 152 A1 (MOLECULAR PARTNERS AG [CH]) 13 April 2016 (2016-04-13) the whole document abstract; claims 1-15 -----	1-68

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2020/054697

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-68(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-68(partially)

A modular binding protein comprising;two or more repeat domains,
inter-repeat loops linking said repeat domains; and one or more heterologous peptide ligands that bind to a target molecule wherein said one or more heterologous peptide ligands comprise at least the amino acid sequence listed in the first line of Table 8 as well as directly derived subject matter such as constructs, medical uses, methods for obtaining said modular binding protein

2-112. claims: 1-68(partially)

A modular binding protein comprising;two or more repeat domains,
inter-repeat loops linking said repeat domains; and one or more heterologous peptide ligands that bind to a target molecule wherein said one or more heterologous peptide ligands comprise at least the amino acid sequence listed in the Table 8 as well as directly derived subject matter such as constructs, medical uses, methods for obtaining said modular binding protein
(the amino acid of each line starting from the second line of table 8 representing a single invention)

113-260. claims: 1-68(partially)

A modular binding protein comprising;two or more repeat domains,
inter-repeat loops linking said repeat domains; and one or more heterologous peptide ligands that bind to a target molecule wherein said one or more heterologous peptide ligands comprise at least the amino acid sequence listed in the Table 8 as well as directly derived subject matter such as constructs, medical uses, methods for obtaining said modular binding protein
(the amino acid of each line starting from the first line of table 9 representing a single invention)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2020/054697

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2019034332	A1	21-02-2019	
		AU 2018319060 A1	02-04-2020
		CA 3073125 A1	21-02-2019
		CN 111225976 A	02-06-2020
		EP 3668976 A1	24-06-2020
		KR 20200038303 A	10-04-2020
		SG 11202001420V A	30-03-2020
		US 2019309286 A1	10-10-2019
		US 2019352637 A1	21-11-2019
		WO 2019034332 A1	21-02-2019

EP 3004152	A1	13-04-2016	
		CN 105209483 A	30-12-2015
		EP 3004152 A1	13-04-2016
		HK 1213001 A1	24-06-2016
		JP 6486908 B2	20-03-2019
		JP 2016521688 A	25-07-2016
		US 2016251404 A1	01-09-2016
		WO 2014191574 A1	04-12-2014
