Engineering mono- and multi-valent inhibitors on a modular repeat-protein scaffold to target oncogenic Tankyrase

Aurora Diamante

Jesus College

University of Cambridge

December 2020

This dissertation is submitted for the degree of Doctor of Philosophy

To Giorgia and my whole family

DECLARATION

This dissertation is submitted for the degree of Doctor of Philosophy. It is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the text.

It does not exceed the prescribed word limit of 60,000 words.

Aurora Diamante

December 2020

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Professor Laura Itzhaki for allowing me to work in her group. Throughout my time working in her lab, she not only gave me freedom to design and perform different type of experiments and learn about multiple techniques, but she also allowed me to explore other interests that not always fit within a scientific career. Among these, she cultivated my passion for entrepreneurship by allowing me to follow through on her success in setting up the spin-out company PolyProx Therapeutics. At the beginning of this process, she took me along to one of the meetings with Cambridge Enterprise and we attended together the business training programme Impulse 2018. Moreover, she supported my attendance to several conferences, including one in Boston (USA), and she helped with all my funding applications. For all this, and much more, I am extremely grateful!

I would also like to thank each member of the Itzhaki group that made my PhD an enjoyable experience from beginning to end. A special mention goes to Dr Albert Perez-Riba, who helped and supervised me throughout the course of my first BBSRC DTP rotation and during the first year of PhD. He has been key to establish the work explained in this dissertation. I am also grateful to all the other current and previous lab members who have contributed to this project: Dr Piyush K. Chaturbedy, Dr Pamela J. E. Rowling, Dr Janet R. Kumita and Dr Rohan S. Eapen. I also thank our collaborator Dr Marc de la Roche for the useful discussions, protocols and materials he provided.

Alongside the PhD, Cambridge has allowed me to establish new friendships in College, in the Department and within the Rhinos Volleyball Club, mostly. Lots of friends, in particular Matthew W., Sara, Annarita and Francesca, other friends in Italy (Alberto, Valeria, Letizia and Giulia), as well as my housemates Matthew R. and Szymon have supported me during this journey and my deepest gratitude goes to them too.

I am extremely grateful to the BBSRC DTP PhD programme for the financial support throughout the PhD, and Jesus College for additional funding received to attend conferences.

Last but not least, I want to thank every member of my family, including more distant relatives and those that have left us too early. My whole family has constantly supported me throughout my academic studies, with a particular emphasis in the last few months of my PhD. Moreover, during the course of my PhD, I celebrated the extreme joy of becoming the aunt of my niece Giorgia, to whom I dedicate this dissertation.

Abstract

Tandem-repeat proteins are a class of proteins ubiquitous in nature and exploited in recent years in biotechnological and pharmaceutical applications due to their favourable biophysical properties. One such repeat motif, the tetratricopeptide repeat (TPR), has already been exploited for biotechnological applications and here the consensus-designed TPR (CTPR) sequence was used as a scaffold to create novel arrays of binding molecules. The platform allows us to display single and multiple functions with diverse geometrical arrangements by grafting short binding sequences onto the loops between adjacent repeats or at the terminal alpha-helices. As proof of concept, proteins were designed to bind to and inhibit the human tankyrase (TNKS), a key regulatory protein involved in Wnt signalling and overexpressed in cancer and other disorders. For this purpose, a tankyrase-binding peptide (TBP) was grafted between two adjacent repeats to create a solvent-exposed loop. A series of mono- and multi-valent TNKS binders, named TBP-CTPR, was assembled by repeating the TNKS-binding unit in tandem in both a monomeric format and also in a trimeric arrangement. The folding and thermodynamic stability of these TBP-CTPR proteins were characterised and the interaction with TNKS was measured using a range of biophysical approaches. Both the engineered TBP-CTPR proteins and TNKS are multivalent, and the effects of multivalency were explored both in the test tube and in the cell. The results show that the proteins interact to form large assemblies. Moreover, the TBP-CTPR proteins were found to have exceptional activity in inhibiting the Wnt signaling pathway upon delivery by encapsulation in fusogenic liposomes. Lastly, hetero-bifunctional constructs were generated by grafting two different binding sequences onto the CTPR scaffold, and a preliminary analysis of their activities was performed. In conclusion, these results point to the tremendous potential of the CTPR scaffold as a platform to build synthetic protein binders, with a particular focus on multivalent interactions.

Table of Contents

Copyright

Reuse licenses or permissions have been obtained for the following figures:

Figure	Publisher	License number
1.3	Elsevier	4864710379075
1.10	Elsevier	4861340698225

List of Abbreviations

2xYT	2x Yeast Tryptone
Amp	Ampicillin
ANK	Ankyrin (repeat)
ARC1-5	Ankyrin Repeat Clusters from 1 to 5 of Tankyrase
ARC4	Ankyrin Repeat Cluster 4 of Tankyrase
ATP	Adenosine Triphosphate
bp	base pairs
CD	Circular Dichroism
CDR	complementary-determining region (antibody)
CMV	Cytomegalovirus (promoter)
C-terminus	Carboxyl terminus
CTPR	Consensus Tetratricopeptide Repeat Protein
CV	Column volume
Da	Dalton (unit)
DARPin	Designed ankyrin repeat protein
DNA	Deoxyribonucleic acid
dNTPs	deoxyribonucleotides
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FL	full-length
GST	Glutathione S-transferase
HA	Hemagglutinin (tag)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
h	hours
HPS	Histidine, Proline, Serine-rich domain of Tankyrase
IDP	Intrinsically Disordered Proteins
IDR	Intrinsically Disordered Regions
IPTG	Isopropyl ß-D-1-thiogalactoside
LB	Luria-Bertani broth
MALDI	Matrix Assisted Laser Desorption/Ionization
MCS	Multiple Cloning Site

MOPS	3-(N-morpholino)propanesulfonic acid
NAD^+	Nicotinamide adenine dinucleotide
Ni-NTA	Nickel Nitrilotriacetic
N-terminus	Amino terminus
OD_{600}	Optical Density at 600 nm
PAGE	Polyacrylamide Gel Electrophoresis
PARP	poly(adenosine diphosphate-ribose) polymerase
PARylation	poly(ADP-ribosyl)ation
PCR	Polymerase Chain Reaction
PROTAC	Proteolysis Targeting Chimera
PTM	Post-translational modification
SAM	Sterile Alpha Motif of Tankyrase
SDS	Sodium Dodecyl Sulphate
SEC	Size Exclusion Chromatography
SLiMs	Short Linear Motifs
TAE	Tris Acetate EDTA
TCEP	tris(2-carboxyethyl)phosphine
TEM	Transmission Electron Microscopy
TNKS	Tankyrase 1 and Tankyrase 2 (human)
TNKS1	Tankyrase 1 (human)
TNKS2	Tankyrase 2 (human)
TPR	Tetratricopeptide Repeat
Tris	tris(hydroxymethyl)aminomethane

The majority of the results shown in this dissertation
have been published in the following article:

Engineering mono- and multi-valent inhibitors on a modular scaffold

Aurora Diamante, Piyush K. Chaturbedy, Pamela J. E. Rowling, Janet R. Kumita,

Rohan S. Eapen, Stephen H. McLaughlin, Marc de la Roche,

Albert Perez-Riba and Laura S. Itzhaki

Available in Appendix D

Chapter 1

General introduction

1.1 Protein engineering in the last decade

In the last few years, protein biotherapeutics have gained increased attention. Biologic drugs are characterised by much larger binding interfaces than small molecules, thereby increasing target specificity and reducing off-target effects. Although a few decades ago antibodies would have been regarded as the only option available, nowadays a great variety of biotherapeutics based on alternative protein scaffolds have been developed. One such alternative is provided by the class of proteins known as tandem-repeat proteins. The emergence of protein engineering techniques has allowed the diversification of antibody technologies and the conversion of repeat proteins into stable, scaffolding units that can be functionalised to bind to a specific target of interest. This Chapter highlights the properties that have contributed to the success of alternative scaffolds in the past few years, with a particular focus on repeat proteins.

1.1.1 Antibody engineering

Antibodies are protein complexes produced by the immune system to defend the body from the threats of external pathogens and microorganisms. Among the five isotypes of natural, human antibodies, IgG is the most abundant class and consists of two heavy chains (50 kDa each) and two light chains (23.5 kDa each) that assemble to form a Y-shaped complex with a high molecular weight of about 150 kDa (Figure 1.1). Each chain consists of a variable region and a constant region. Selectivity towards a specific antigen is provided by the variable region, which adapts and evolves to maximise its binding affinity against a determined pathogen.

Three, hypermutable, flexible loops in each variable domain give rise to the so-called complementary-determining region (CDR) of an antibody, responsible for its specificity.[1]

In the past few years, antibodies have been extensively modified and engineered for drug development purposes. Starting from humanised antibodies, progressively smaller antibody fragments have been developed to obtain recombinant, monomeric, easier to produce antibody derivatives, with identical binding efficiency but improved tissue penetration (Figure 1.1). Apart from mono-specific antibodies, bi- and tri-specific antibody variants have also been explored to recruit multiple targets or extend their half-life in the blood stream.[2]
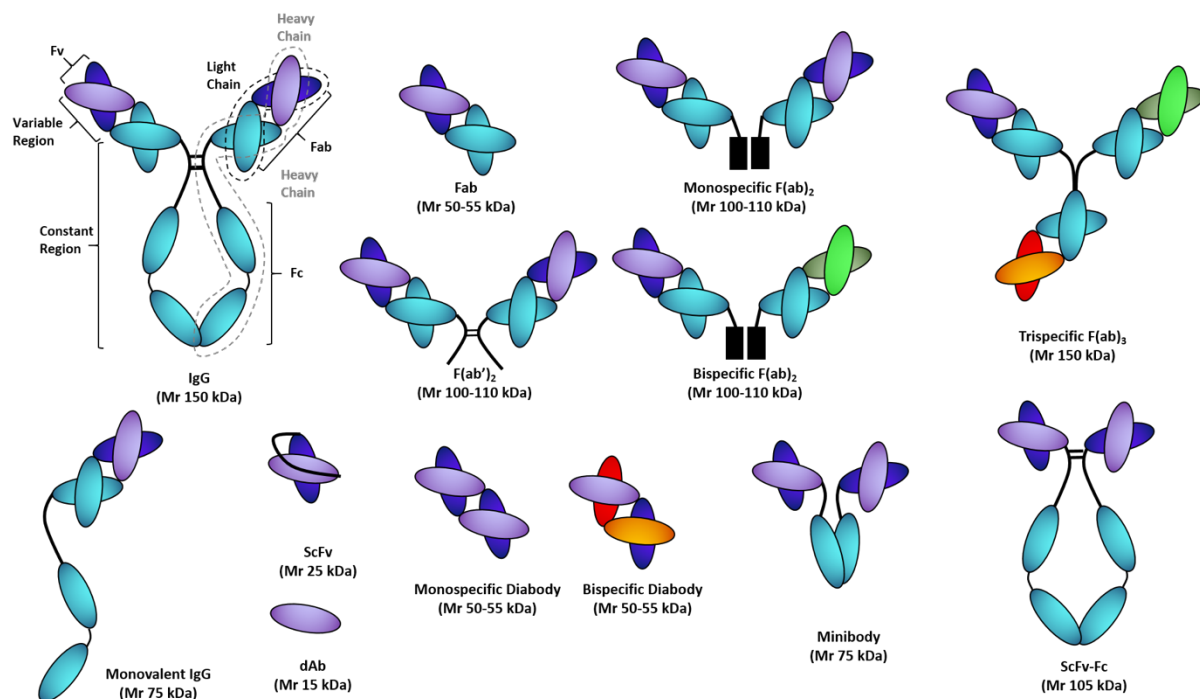
Figure 1.1: Schematic representation of an IgG antibody and its recombinant antibody derivatives of progressively smaller molecular weight size. Figure obtained from the tebu-bio blog.[3]

1.1.2   The concept of multivalency

Protein-protein and antibody-antigen complexes are formed when one or more binding events occur between two or more interacting partners. When describing the binding interaction for complex formation, both affinity and valency should be taken into account. Affinity corresponds to the strength of the interaction at a single site, and valency indicates the number of identical or similar binding sites within a protein, engaging with the same target. Although the majority of natural proteins interact through only one binding site, some proteins have

multiple binding sites for their ligand, and multivalency often counterbalances a weak binding affinity.[4]

Antibodies provide an obvious example of multivalency. Each antibody has between two and ten identical binding sites, depending on the isotype: antibodies with fewer binding sites tend to have high affinities, whereas antibodies with a greater number of binding sites (such as IgM) tend to have low affinities. For intracellular proteins, multivalency has been frequently observed for intrinsically-disordered proteins (IDPs), as their intrinsic structural flexibility allows the epitopes to adapt more easily to the complementary binding surface of the interacting partner. Multivalency has therefore been more extensively studied in this particular class of proteins, and it can occur through different mechanisms:[5]

- Avidity occurs when two or more binding sites located on the ligand bind and complement two or more binding sites on the binding partner. Avidity is therefore only possible in the presence of two multivalent binders. Avidity is described as the accumulated strength generated by multiple affinities: once the ligand has bound one site, the probability of establishing a second binding event is much higher than for the first one, introducing cooperativity. As the number of binding events increases, the affinity between the two interacting proteins also accumulates.

- Allovalency occurs when multiple and identical receptor-binding sites are positioned in tandem on a ligand and they compete for a single binding site on the receptor. Despite only one binding event can occur, the presence and competition by multiple tandem sites creates cooperativity and increases the overall affinity.

- Fuzzy complexes: this term, introduced by Tompa and Fuxreiter in 2007,[6] refers to the ensemble of various conformational states that occur when two or more binding sites on the ligand bind two or more binding sites on the receptor in different combinations. As various binding conformational states are allowed, the fuzzy complex is dynamic in the bound state and individual binding events continuously form and break, while the two interacting partners are still in complex.[7]

Multivalent interactions are a recurrent feature across all biological systems, in the macro as well as in the nano scale. Multivalency is exploited in nature to provide increased molecular affinity and specificity to binding interactions, compared to monovalent counterparts.[4] The many advantages induced by multivalent interactions have therefore led to the development of artificial, multivalent systems. Many multivalent antibody technologies leverage and expand the natural modularity (i.e. multi-domain nature) of immunoglobulins, to achieve higher

valency and dual-targeting within the same construct.[8,9] Likewise, other proteins used as alternative scaffolds, have been designed and used for multivalent recognition, as mentioned below. In alternative, functional peptide motifs have been assembled on synthetic chemical scaffolds,[10,11] DNA[12,13] and protein scaffolds.[14–20]

1.1.3 Antibody versus alternative scaffolds

In the past few years, a wide variety of biologic drugs have been developed as alternatives to antibody therapeutics. These therapeutics are usually based on small protein scaffolds. Alternative scaffolds have similar properties to antibodies in terms of affinity and functionality, but they are differentiated by their greater solubility and stability, smaller size and simplified manufacturing. Other liabilities of antibodies, such as redox sensitivity and the need of targets with poor host homology for effective immunisation, can also be avoided. For these reasons, alternative scaffolds are becoming a new powerful modality in drug development, in particular for targets considered "undruggable" by conventional approaches.[21]

Alternative scaffolds are generally characterised by a relatively small size (less than ~150 amino acids) and an independent folded unit with rigid tertiary structure, providing stability and folding efficiency.[22] Generally, target binding of the scaffold molecule comes from randomised sequences within one or multiple loops (in a similar way to the CRD of antibodies), helices, flat surfaces, cavities or an inserted random peptide.[22] In practical terms, target binding can be achieved in several ways. Binding can be introduced by random mutagenesis within non-conserved residues of the scaffold followed by library generation and screening, most commonly through phage, bacterial, ribosome or direct nucleic acid display. As an alternative, new binding functions can be introduced by rational engineering approaches, where known binding sequences (i.e. loops or α-helices) involved in naturally occurring protein-protein interactions are transferred (grafted) onto a novel scaffold. This method requires the epitope to be a contiguous sequence and is facilitated by having a crystal structure of the native protein-protein interaction, allowing identification of the critical binding residues and their conformation required for binding. In this way, the binding epitope is "cut-and-pasted" onto an appropriate site within the scaffold.[20,22,23] The scaffold thus constrains the grafted peptide in its native, bioactive (binding-competent) conformation. Similar to chemical macrocyclisation or "stapling", the scaffold also provides the grafted peptide with enhanced proteolytic stability relative to the isolated peptide.[22] Another alternative method to introduce target specificity onto a scaffold is by computer-aided design. Protein design algorithms such

as the Rosetta software (from David Baker's lab, Seattle, USA) allow one to screen *in silico* for target binding using a three-step process: grafting, docking, and design. The user-supplied binding fragment is grafted onto a proposed scaffold, and the grafted protein is then tested in silico by docking onto the target. If the fit is deemed to be favourable, the binding interface can be further optimised by introducing additional contact interactions. Nowadays, small proteins can also be designed *de novo* with customised shape, stability and binding specificity towards a target of interest.[24–27] Recently, algorithms such as Rosetta and DeepMind's AlphaFold have advanced to such an extent that it is now possible to predict with atomic-level accuracy the 3D shape of proteins just from their amino-acid sequences.

Finally, as for antibody engineering, multivalency has also been attempted in alternative scaffolds.[28] It is generally achieved by connecting monovalent domains in "beads-on-a-string" manner,[29] by fusing them onto an antibody scaffold[30] or a smaller immunoglobulin fragments or an albumin-binding domain (ABD, which also provides the multivalent construct with extended half-life),[31,32] or by fusing other oligomerisation domains to the protein.[33]

The list of the most relevant examples of engineered alternative protein scaffolds used as novel biotherapeutics is provided in Table 1.1, with an indication of their origin, fold, number of amino acid and stage of development.

Name	Scaffold origin	Fold (AA)	Stage of development
Affibodies	Z domain of protein A	α_3 (58)	Clinical trials (ABY-025)
Affilins	gamma-B crystallin	α/β (183)	Preclinical studies
Affimers (adhiron)	Cystatin	α/β (79)	Preclinical studies
Anticalins	Lipocalins	β-barrel (160-180)	Clinical trials (PRS-050)
DARPINs	Ankyrin repeat	$\alpha_2\beta_2$ repeated (33)	Clinical trials (MP0112)
Fynomers	SH3 domain of Fyn	Mainly β (~60)	Ab fusions (FynomAb)
Monobodies	10th type III domain of fibronectin	β-sandwich (94)	Clinical trials (Pegdinetanib)

Table 1.1: List of the most relevant alternative scaffold proteins used for therapeutic purposes. AA indicates the number of amino acid residues; Ab is an abbreviation for antibody.

1.2   Tandem-repeat proteins

Tandem-repeat proteins (also referred to as repeat proteins) are a broad group of ubiquitous proteins, characterised by the repetition of a peptide unit in tandem. Tandem-repeat proteins

occur in 14% of all known proteins (from both eukaryotes and prokaryotes)[34] and in about one-third of the human proteome.[35] Within this group of proteins, the many variations are determined by the length of the repeating unit: with units ranging from 1 to over 100 amino acid residues, the tandem-repeat proteins represent an extremely diversified group. A more recent analysis of the 3D structures obtained from repeat proteins has allowed the classification of these into five classes depending on the length of their repeats.[36] Proteins with 1-2 amino acid repeats (class I) form crystalline aggregates of unlimited size, which are harmful to living organisms. When the repetitive unit consists of 30-130 amino acids (class V), the repeat folds as an independent stable domain.[36] Here we are interested in tandem-repeat proteins that belong to class III, comprising repeats of 5-40 amino acids. Each repeat folds into a structural motif formed from one to four segments of secondary structure (α-helices, β-strands or a combination of the two) connected by turns or loops. Unlike globular proteins, they pack in a linear fashion to produce regular, elongated, quasi-one-dimensional architectures (Figure 1.2). Moreover, tandem repeats are stabilised by short-range interactions between residues close in sequence either within a repeat or between neighbouring repeats. Stabilising contacts are located at the interface between consecutive repeats, whereas residues involved in target binding are surface-exposed. Tandem-repeat proteins belonging to this subgroup, such as ankyrin repeats (ANK), tetratricopeptide repeats (TPR), armadillo repeats (ARM) and leucine-rich repeats (LRR), generally function as binding partners to other proteins, small molecules or nucleic acids. These protein arrays require at least three repeats to be folded, stable and allow the formation of an extended binding surface.[36] From an evolutionary perspective, repeats are thought to arise from intragenic duplications and recombination events, with the resulting advantage of enlarging the available binding surface area.[37]

Increasing interest in class III of repeat proteins has arisen recently due to their simple, modular, stable and regular architectures. They are also chemically homogeneous due to their repeating nature, and there is no limit to their size or the number of repeats. These characteristics make repeat proteins straightforward both to dissect and to rationally redesign in a LEGO[TM]-like manner.[38] Several types of repeat proteins have recently been functionalised and used as building blocks to create artificial biologics and biomaterial for clinical and biotechnological purposes, respectively.[39–44] Interestingly, new repeat-protein folds, consisting of repeats of a helix-loop-helix-loop motif and not found in nature, have also been *de novo* designed by the Baker lab. These proteins allow one to expand beyond natural repeat protein shapes and to generate completely novel repeat proteins with customised geometries.[45]
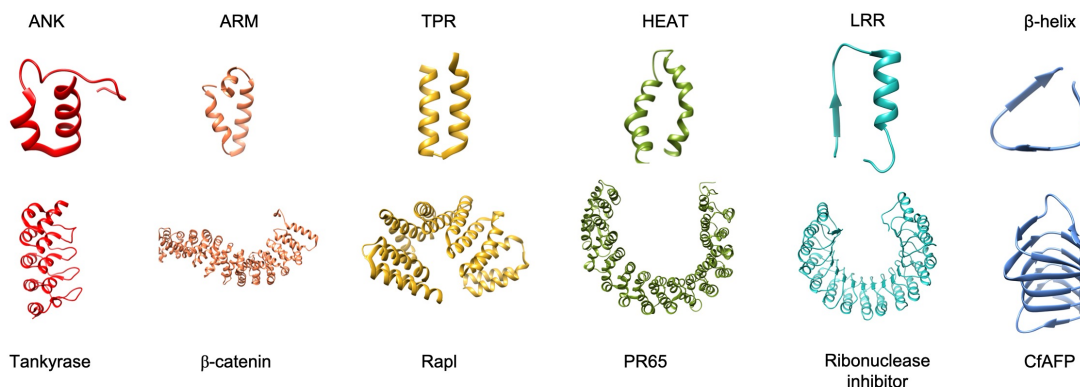
Figure 1.2: Examples of tandem repeat proteins and their repeated motif: ANK – ankyrin (PDB 3TWR)[46], ARM – armadillo (PDB 2Z6H)[47], TPR – tetratricopeptide repeat (PDB 4I1A)[48], HEAT – Huntington, elongation factor 3, protein phosphatase 2A, and the yeast kinase TOR1 (PDB 2IAE)[49], LRR – leucine rich repeat (PDB 2BNH)[50], β-helix (PDB 1L0S)[51]. Images were generated using the software UCSF Chimera.[52]

Among the different types of naturally occurring repeat proteins, ANK repeats have been particularly thoroughly exploited by the Plückthun lab and others for the development of biotherapeutics, giving rise to a new class of antibody-like proteins named DARPins ((consensus-)Designed Ankyrin Repeat Proteins), some of which are already in clinical trials for a variety of diseases.[44] Multivalency in DARPins has been achieved by fusing up to six DARPin modules mostly through flexible peptide linkers, or through more rigid linkers such as leucine zipper and α-helical connectors.[29,53] However, no modular repeat-protein platform has been developed to date that is capable of combinatorial incorporation of multiple binding motifs in a single, stable and robust scaffold as well as presenting them with varied, precise and programmable geometries. The work presented here shows that the repeat protein scaffold adopted during my PhD possesses all the necessary features with which to build such a platform.

1.2.1 The tetratricopeptide repeat (TPR)

One of the simplest structures in class III of tandem-repeat proteins is the tetratricopeptide repeat (TPR), a 34-residue motif comprising two antiparallel α-helices, named helix A and B, connected by a short turn.[54] It was first discovered in 1990 within the *CDC23* gene of *Saccharomyces cerevisiae*[55] and the *nuc2*⁺ gene of *Schizosaccharomyces pombe*.[56] To date, over 5000 TPR-containing proteins have been identified using bioinformatics tools.[57] These proteins are involved in a diverse spectrum of cellular functions (protein transcription, folding,

transport and cell-cycle progression) and in multiprotein complex assemblies in humans and in other organisms.[58] The TPR motif occurs in tandem arrays of three to sixteen or more units,[59] and in nature it mediates protein-protein interactions in a variety of molecular recognition modes.[57,60] The most commonly observed binding mode involves around three repeats forming a concave groove and accommodating a short extended peptide that sits roughly perpendicular to the repeat array.[60] The specificity of each TPR-containing proteins relies on the underlying sequence of each repeat, with a particular attention to the solvent-exposed residues. Looking closer at the TPR sequence within TPR-containing proteins, Sikorski *et al.* identified eight residues that are particularly conserved, at positions 4 (W/L/F), 7 (L/I/M), 8 (G/A/S), 11 (Y/L/F), 20 (A/S/E), 24 (F/Y/L), 27 (A/S/L) and 32 (P/K/E).[55] Amino acids at positions 4, 7, 8 and 11 are present within the helix A; amino acids at positions 20, 24 and 27 are within helix B; and they are all hydrophobic (with few exceptions in some natural protein sequences). The side chains of these residues are clustered along the same face of the corresponding helix, allowing hydrophobic interfacial interactions between the two amphipathic helices. The proline at position 32 is found at the loop connecting consecutive repeats. This residue is key in forcing a turn in direction to occur, leading to a termination of helix B. The backbone of Pro32 also interacts with Trp4 of the following repeat, therefore stabilising the turn.[54]

In 2003, an artificial consensus-designed TPR sequence, named CTPR, was developed by Main *et al.* for biotechnological and clinical purposes.[61] Designing a consensus sequence allows one to explore the overall structure and stability of the TPR motif, without the additional complications of protein-specific differences between repeats. Moreover, residues required for binding are removed, providing an excellent starting-point sequence for structural studies and an ideal scaffold on which to introduce novel functionalities. Following sequence alignment of 1837 TPR motifs from a non-redundant protein database of 107 proteins, the authors selected the most preferred residue for each of the 34 amino acids positions within the TPR.[61] The eight residues previously identified by Sikorski *et al.* to be highly conserved were among those with the highest value of propensity in the Main *et al.* study.[61]

Several changes to this CTPR sequence were introduced to enhance its favourable properties, making it more amenable for protein engineering:

1. At position 10, cysteine was replaced with alanine to avoid unwanted disulphide bond formation;

2. An N-terminal stabilising capping sequence (Gly-Asn-Ser) was introduced;

3. A solvating helix was grafted to the C-terminal end of the TRP, based on capping repeats observed in some natural and previously studied TPR proteins.[54] To enhance

the protein's solubility, four large hydrophobic residues within the solvating helix, one

Trp and three Tyr, which would be packed against the adjacent repeat if they were in

an internal repeat, were mutated to hydrophilic polar residues Lys and Gln.

The final CTPR sequence, together with the capping sequence and the solvating helix is shown in Figure 1.3. This sequence, with few additional mutations detailed in Section 3.2.1, was adopted as the scaffold for the studies presented in this thesis.
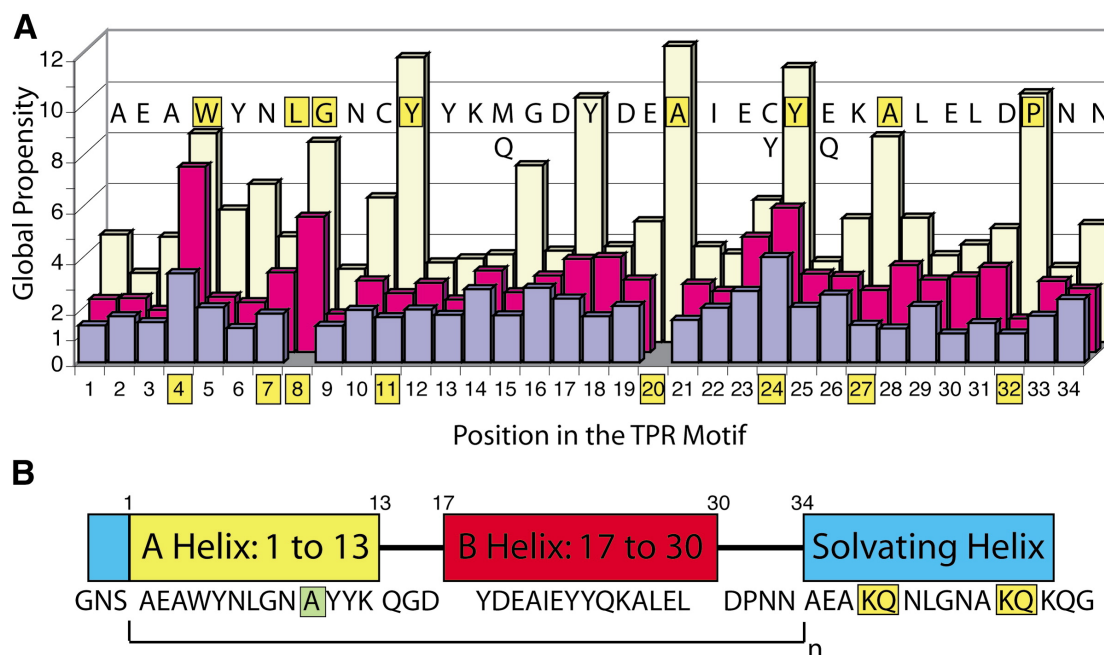
Figure 1.3: The CTPR motif developed by Main *et al.* (A) Histogram showing the global propensity of the three most frequent amino acids (if greater than 1) at each of the 34 amino acid positions of the TPR motif. Residues within boxes and their position correspond to the highly conserved TPR residues. (B) Schematic representation of the CTPR sequence, including an N-terminal capping sequence and a C-terminal solvating helix. Boxes denote residues that have been mutated by the authors. Figure obtained from Main *et al.*[61]

The same authors determined the structures of the CTPR2 and CTPR3 constructs, both by NMR and X-ray crystallography.[61] As for the previous TPR solved structures, the CTPR motif also folds into the expected helix-turn-helix conformation with two antiparallel and amphipathic α-helices per repeat (Figure 1.4A). Most α-helices in nature are amphipathic, meaning that one face of the helix displays hydrophobic residues, while the other is hydrophilic, and this characteristic promotes the correct folding of helical domains. Indeed, the hydrophobic residues remain buried within the structure and interact with other structural elements to avoid contact with an aqueous solvent, whereas the hydrophilic residues are

solvent-exposed. This pattern is observed also in the CTPR helices. Within the helix-turn-helix CTPR motif, the side chains of the eight highly conserved, hydrophobic residues are responsible for the interfacial contacts between helices. In particular, the large, hydrophobic side chains along helix A (Trp4, Leu7 and Tyr11) are placed between the preceding and the following B helices, forming a network of packing interactions with residues from both helices. These inter-repeat interactions lead to an extended hydrophobic core, that stabilises consecutive repeats at their interface and the overall CTPR structure.[61] It is also interesting to note how conserved residues with small side chains (Ala and Gly) accommodate and interact with the larger side chain of hydrophobic residues from the same repeat (Figure 1.4B). This characteristic pattern, known as "knobs-into-holes", is crucial to the formation of stabilising intra-repeat interactions within the CTPR scaffold.[61]
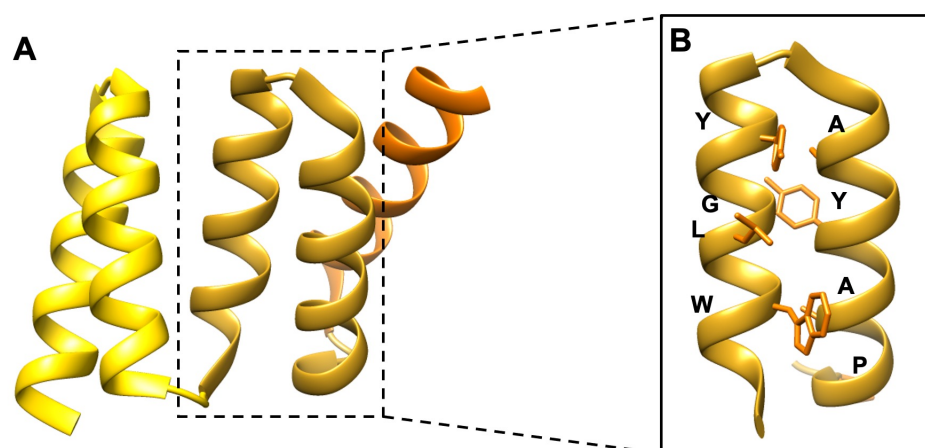
Figure 1.4: CTPR helix-turn-helix motif. (A) Ribbon structure of the CTPR2-Solvating Helix construct obtained by Main *et al.*[61] The CTPR repeats are coloured with different shades of yellow, the solvating helix is in orange. (PDB: 1NA3) (B) CTPR repeat with the side chains of the highly conserved residues visible and labelled. (PDB: 1NA3) Images were generated using the software UCSF Chimera.

Looking at the helical wheel representation, that illustrates the orientation of each residue along the helical axis, it is particularly easy to visualise the amphipathic nature of the helices A and B, with the seven highly conserved residues promoting hydrophobic intra-repeat interactions being indicated (Figure 1.5). In the solvating helix, the four hydrophobic residues on the outward-facing surface are mutated to polar amino acids to enhance the solubility of the protein construct.
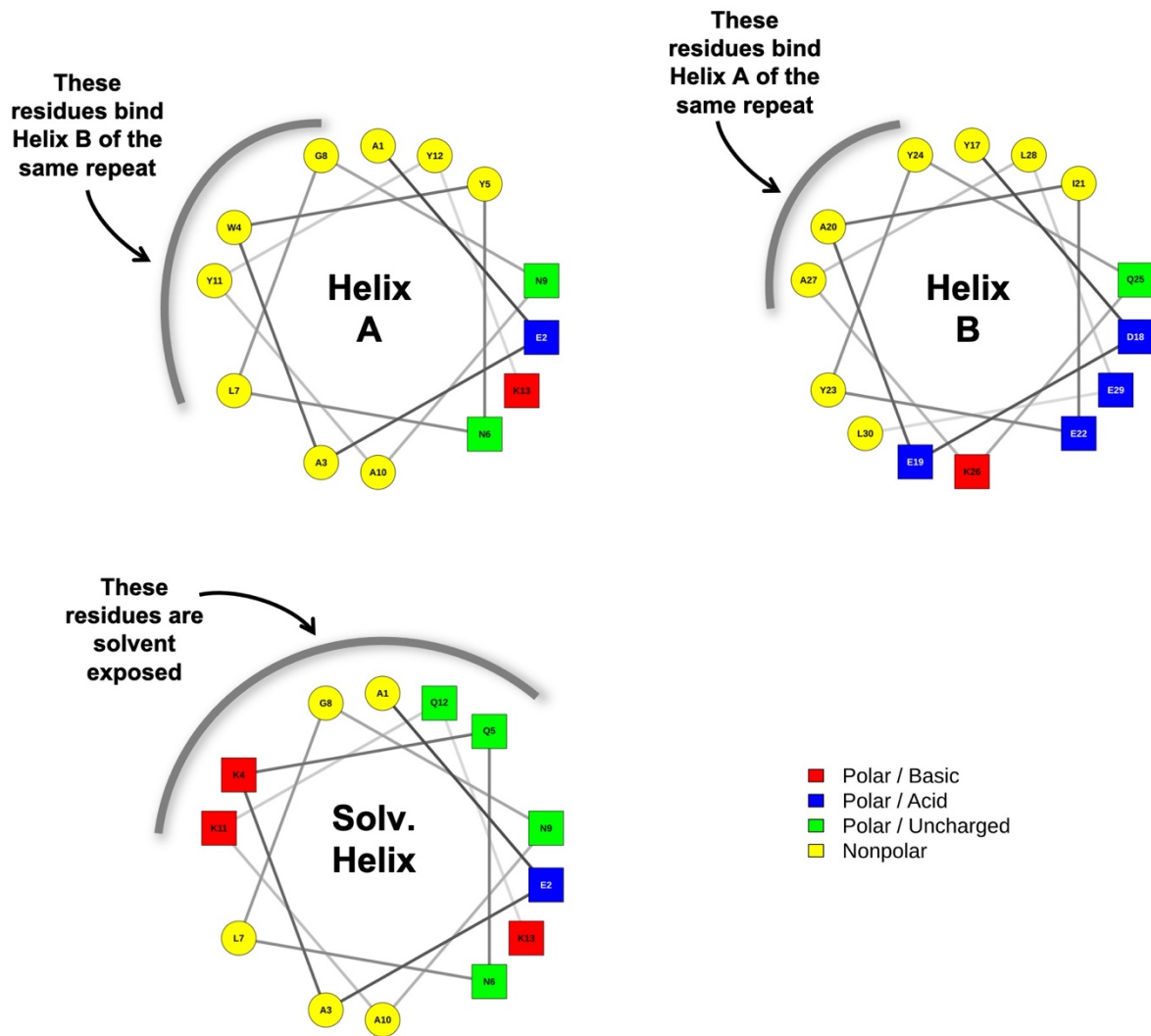
Figure 1.5: Helical wheel representation of Helix A, Helix B (Top) and the Solvating Helix (Bottom). For each helix, the signature residues involved in intra-repeat interactions are indicated with a grey arch. Polar residues are shown as boxes, nonpolar residues as circles. Colour legend is provided in the bottom right corner. Figures were obtained from a the web-based application NetWheels.[62]

As for other repeat proteins in class III, long arrays of TPRs fold into an extended conformation with an overall superhelical architecture, as demonstrated by the crystal structure of TPR-repeat domain of O-linked GlcNAc transferase (OGT).[63] The individual helices within the same repeat are arranged in a slightly twisted manner, tilted from each other by ~24º, resulting in a right-handed superhelical shape. The same has been observed for the consensus TPR proteins, where eight CTPR repeats are required to complete a turn of the superhelix (Figure 1.6).[64] The superhelix presents a regular geometry and consists of two protein surfaces: helices A mainly contribute to the inner concave surface, whereas the outer surface includes residues

from both helices A and B. The concave face is the one that forms the binding interface for each specific target.[64]
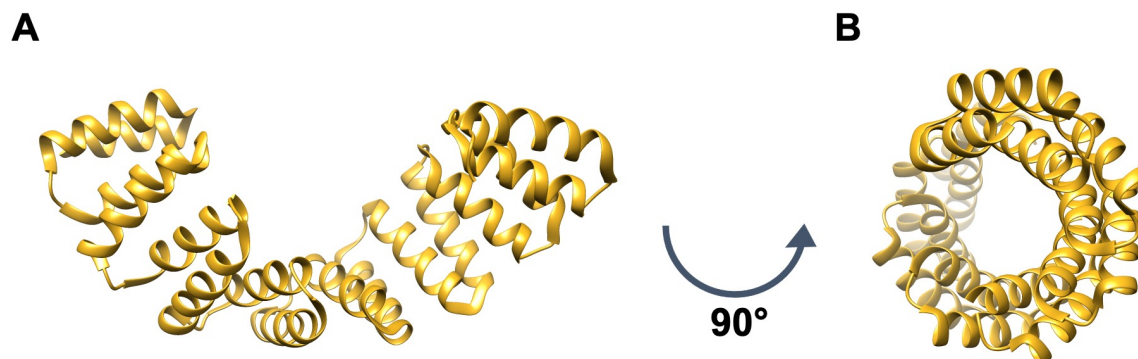
Figure 1.6: Superhelical fold of a CTPR8 protein. (A) Crystal structure of CTPR8 construct obtained by Kajander *et al.*[64] (PDB: 2FO7) (B) Protein in A observed along its axis, after a 90º rotation. Images were generated using the software UCSF Chimera.

The CTPR proteins have been extensively studied by the Regan, Cortajarena, Grove and Main groups not only to understand their biophysical properties, but also to exploit them in biotechnological applications. Artificial CTPR proteins have been shown to have properties favourable for protein engineering, with applications in biomaterial design and biotherapeutics development, among others:

- As a result of the consensus residues, they are extraordinarily stable (much more stable than natural TPR proteins) and their stability increases with increasing numbers of repeats.[61,65] A minimum of two repeats is required for a stable and independently folded construct.

- Due to their high stability, CTPR proteins of different lengths express with very high yields in *Escherichia coli* (typically tens of milligrams per litre of culture) and are monomeric, soluble and well folded proteins.

- The CTPR sequence does not contain cysteine residues, and therefore the protein structure is not affected by the reducing and oxidizing environments within and outside the cell, respectively.

- The modular nature of the architecture means that consensus repeats are self-compatible and can be individually designed and put together in any order.

1.2.2 Functionalising CTPRs

Initial attempts to functionalise the CTPR scaffold have been performed by mimicking how natural TPR proteins bind to their target. The binding mode most represented in the PDB structures and the most well studied corresponds to a short target peptide binding the concave groove formed by 2-3 repeats.[60] Binding affinities are generally modest (low micromolar or weaker). In the first design study, the Hsp90-binding residues of the TPR protein Hop (Hsp Organizing protein) were grafted onto a CTPR scaffold comprising three repeating units to generate a functional Hsp90 inhibitor for therapeutic applications.[39,66,67] The Cortajarena and Regan groups have also produced novel, functional CTPR proteins using rational design approaches and library screening.[40,68,69]

The CTPR scaffold has also been exploited by the Cortajarena lab for the fabrication of bioinspired materials to be used in several biotechnological applications. Firstly, as observed in the crystal lattice of crystallised CTPR proteins,[64,66] CTPR proteins have intrinsic self-assembly properties, which allow them to polymerise in a controlled fashion and form nanostructured biomaterials through head-to-tail and side-to-side interactions, while maintaining their structural and functional properties. Further studies engineered a controlled head-to-tail assembly by introducing reactivities (cysteine residues or thioester groups) at the end of the CTPR module and produce longer polymers through the formation of more stable interactions.[41,70] Secondly, this bottom-up approach led to the assembly of stimuli-responsive gels, ordered protein films and nanometer-scale fibers with specific ligand recognition function and tunability, given the binding specificity of the underlying TPR modules and the possibility of mixing and matching different repeats.[70–72] Importantly, CTPR modules retained their helical structure, and therefore their binding capabilities, within all the different polymeric assemblies.[70–72] These unique features allowed to further expand the applications of CTPR-based systems and fabricate well-ordered bio-organic nanostructured materials with varying properties depending on the functional component conjugated to the scaffold. More recently, CTPR modules have therefore been conjugated to organic functional components or used as templates for stabilising metal nanoclusters.[73–77] For example, photo- and electroactive films were generated by forming CTPR-porphyrin hybrid modules, where the required distance and orientation of porphyrin molecules for efficient π-π stacking interactions between the rings is guaranteed by the structural integrity of the CTPR scaffold.[73] As an alternative, CTPR films have been engineered to immobilise the catalase enzyme, expanding the CTPR applications towards the production of biocatalytic materials.[78] The ability to control the functionalisation of the protein, while maintaining its structure and binding properties, represents, therefore, a

key advantage for the production of CTPR-based hybrid materials that can find applications in drug delivery (for example through their stimuli-dependent encapsulation and release of molecules),[71] synthetic biology and nanoelectronics, among others.[79]

The Itzhaki group has recently shown that the CTPR scaffold can be functionalised in other ways. Rather than mimicking the natural binding mode of a binding interface comprising residues spanning the surface of two-three repeats, the Itzhaki lab has investigated the possibility of extending the 4-residue inter-repeat loop, which corresponds to the DPNN sequence in the CTPR sequence. The results showed that the CTPR scaffold can accommodate extensions in the inter-repeat loop of 25 amino acids or more without compromising the native structure.[80,81] The Itzhaki group then demonstrated that it is possible to functionalise the scaffold by extending the loop with a linear, target-binding peptide.[82] The grafted sequence can maintain its binding properties, and this conformational constraint within a structural loop represents an alternative to peptide stapling. The same approach has previously been used with other protein scaffolds.[15,20,23] In this study, the grafted peptide sequence was derived from the protein Nrf2 (Nuclear factor erythroid-2-related factor 2) and is known to bind to the oncogenic protein Keap1 (Kelch-like ECH-associated protein 1).[82] Following loop grafting, the Keap1-binding affinity was further optimised by varying the loop length and sequence without the need of computational modelling. Therefore, the CTPR scaffold could theoretically accommodate unstructured binding sequences of various lengths that bind their respective targets in an extended conformation. As an alternative, CTPR modules can also be functionalised by helix grafting (an approach that has been used to functionalise scaffolds such as helix bundles – see Chin *et al.* and Sia *et al.* for the earliest examples of this approach[83,84]), as demonstrated in my PhD thesis (Chapter 3). Thus, by combining the possibility to graft binding sequence together with the modularity of the scaffold we obtain the capacity for possibly limitless functionalisation.[82]

1.2.3   Functionalising CTPRs to bind Tankyrase (TNKS)

As proof of concept, in my PhD thesis I explored the possibility of grafting a Tankyrase-binding peptide (TBP) onto the inter-repeat loop to produce a set of monovalent and multivalent CTPR constructs, named TBP-CTPR. The TBP is known to bind and inhibit the human poly(ADP) ribose polymerase (PARP) proteins Tankyrase 1 (TNKS1) and Tankyrase 2 (TNKS2; TNKS when referring to both) according to previous studies,[46] including work performed in the Itzhaki lab,[85] as described in the following sections.

1.3 The PARP family

The human PARP proteins are a large family of 17 enzymes catalysing the poly(ADP) ribosylation (PARylation) of the protein substrate.[86] For the PARylation reaction, a molecule of ß-nicotinamide adenine dinucleotide (NAD$^+$) is used by the enzyme to add a single or multiple ADP-ribose molecules onto the protein substrate via an ester bond with the carboxyl group of acidic residues (aspartic acid and glutamic acid) or, more rarely, cysteine and lysine residues on the target proteins. The reaction results in ADP-ribosylated protein and the release of a molecule of nicotinamide for each NAD$^+$ added (Figure 1.7).[86]
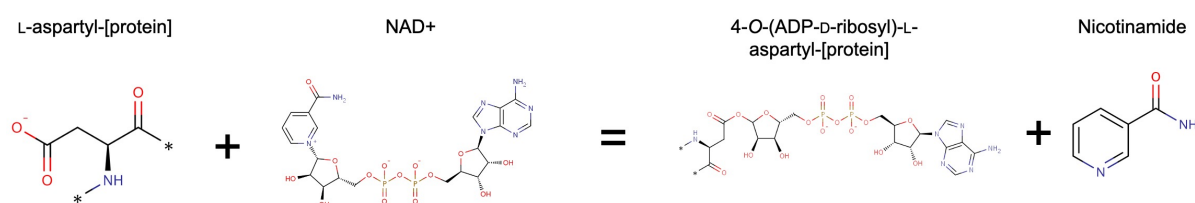
Figure 1.7: PARylation reaction mediated by the PARP enzymes. A molecule of NAD$^+$ is added to the L-aspartyl group of a substrate protein, resulting in a 4-O-ADP-D-ribosylated protein and the release of a molecule of nicotinamide.

Similar to protein ubiquitination, multiple NAD$^+$ molecules can be added to form a linear or a branched chain of poly-ADP-ribosyl units onto the same residue. This post-translational modification determines the fate of the modified substrate protein, in most instances causing a change in its subcellular localisation or its degradation via the proteasome.[86]

1.3.1 Structure of Tankyrase (TNKS)

The human genome encodes two Tankyrase isoforms: Tankyrase 1 (TNKS1, also named ARTD5 or PARP-5a) and Tankyrase 2 (TNKS2, also named ARTD6 or PARP-5b). Tankyrase 1 (<u>T</u>RF1-interacting, <u>anky</u>rin-<u>r</u>elated <u>ADP</u>- ribose polymer<u>ase</u>) was identified in 1998 and named after its interaction with telomeric repeat binding factor-1 (TRF1).[87] Tankyrase 2 was isolated in 2001 as a binding partner of the adaptor protein Grb14 (Growth factor Receptor Bound protein 14).[88] TNKS share 83% overall sequence similarity[88] and high structural homology. TNKS are unique in the PARP family in having a N-terminal ankyrin domain (ANK), segmented into five consecutive ankyrin-repeat cluster (ARC) subdomains.[87] These are followed by a sterile alpha motif (SAM) and a catalytic PARP domain. TNKS1 has an

additional histidine, proline, serine-rich (HPS) domain at its N-terminal end. Each of the five ARC subdomains contains itself five ankyrin repeats. Four of the five ARCs (ARC I, ARC II, ARC IV and ARC V) mediate protein-protein interactions, allowing the formation of large macromolecular complexes between TNKS and substrates.[46,89–91] Substrates recognise TNKS ARCs through a short (8 residue) binding peptide motif having the form **R**XXΦ[D/E]**G**X[D/E], where X is any amino acid, Φ is a small hydrophobic residue, and arginine and glycine at position 1 and 6 respectively are essential for the binding, as demonstrated by alanine scanning mutagenesis.[46] Despite the substantial degree of degeneracy tolerated in some positions of the TNKS binding motifs, all TNKS binding peptides engage the ARCs in an extended conformation, as shown in the crystal structure of TNKS2 ARC4 in complex with the 3BP2 peptide (Figure 1.8).[46] Often, this motif is present in multiple copies within a TNKS substrate to allow simultaneous binding to multiple ARC subdomains in the same TNKS molecule and enhancing affinity by avidity. This effect has been observed in the TNKS substrates Axin1/2 and RNF146.[89,92]
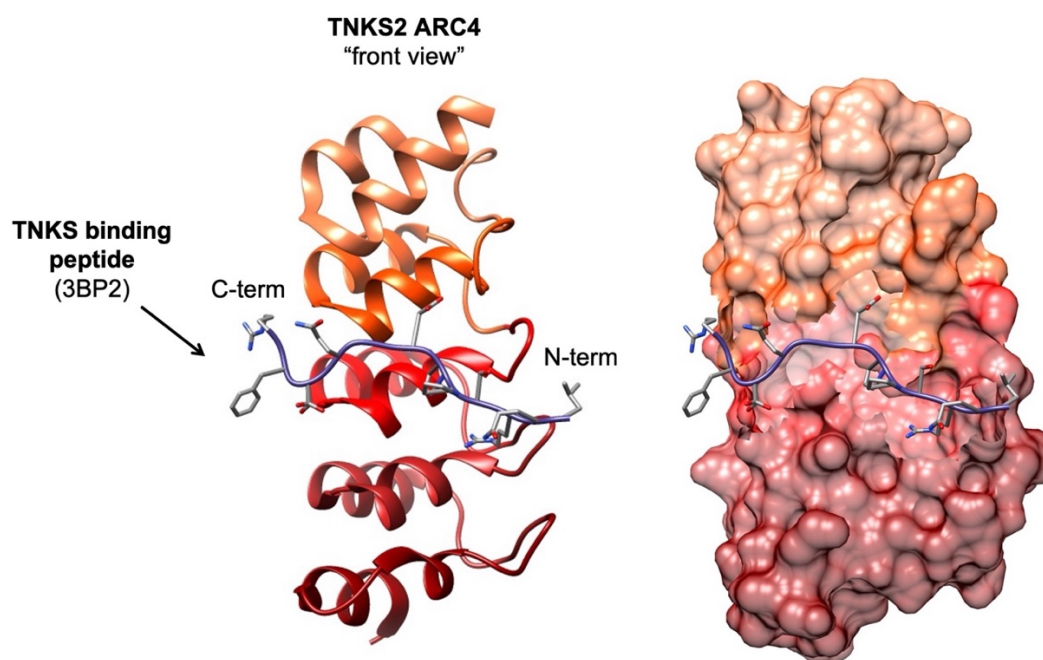
Figure 1.8: Structural representation of the TNKS2 ARC4 domain in complex with the TNKS-binding peptide of the 3BP2 substrate. (Left) Ribbon representation of TNKS2 ARC4:3BP2 complex, with the peptide's side chains shown in stick representation. The five ANK repeats of TNKS2 ARC4 are coloured in shades of red, while the 3BP2 peptide is in purple with the side chains coloured by heteroatoms (oxygen in red, nitrogen in blue) (Right) TNKS2 ARC4:3BP2 complex, with TNKS2 ARC4 in surface representation. (PDB 3TWR) Images were generated using the software UCSF Chimera.

The SAM domain is a five-helix bundle with acidic and basic residues on opposite surfaces, mediating SAM, and therefore TNKS, polymerisation by head-to-tail electrostatic interactions.[93,94] Purified, wild-type SAM domains can, indeed, form filaments *in-vitro*.[93] The PARP domain mediates substrate PARylation once the complex with the substrate has been formed. TNKS self-oligomerisation also results in self-PARylation.[95] The function of the HPS domain of TNKS1 remains unknown. The structure of several domains of TNKS1 and TNKS2 has been determined (Figure 1.9), and the high sequence homology between the two proteins results in almost identical tertiary structural elements. Each ARC subdomain comprises five of the 33-residue ANK tandem repeats, that fold into the characteristic two α-helices connected by a loop and pack against each other to form a superhelical structure. The five ARC subdomains are separated from each other by a semi-conserved, α-helical peptide motif (sequence LLEAAR/K), so that there is not one single, continuous superhelical ANK array.[89] All ARC subdomains of TNKS1, with the exception of ARC4, have been crystallised.[89,96] However, the high flexibility of the overall ARC1-5 domain, in particular in the linker region between ARC3 and ARC4, has prevented crystallisation of it in its entirety. The α-helical LLEAAR/K motifs between ARC1-ARC2 and ARC2-ARC3 are instead more rigid than that between ARC3-ARC4, and they induce the formation of the asymmetric "U" shape observed in the crystal structure of the ARC1-3 domains (Figure 1.9, TNKS1 ARC1-3).[89] For TNKS2, only the ARC4 subdomain has been crystallised.[46] All ARC subdomains, except ARC3, have been crystallised in the presence of binding peptides, which have an extended conformation and make contacts predominantly with the middle ANK repeat (Figure 1.8 and Figure 1.9., with bound peptides indicated). In both proteins, the SAM domain consists of five α-helices arranged in a globular conformation,[93] and the PARP domain folds into α-helices and β-strands.[97,98]
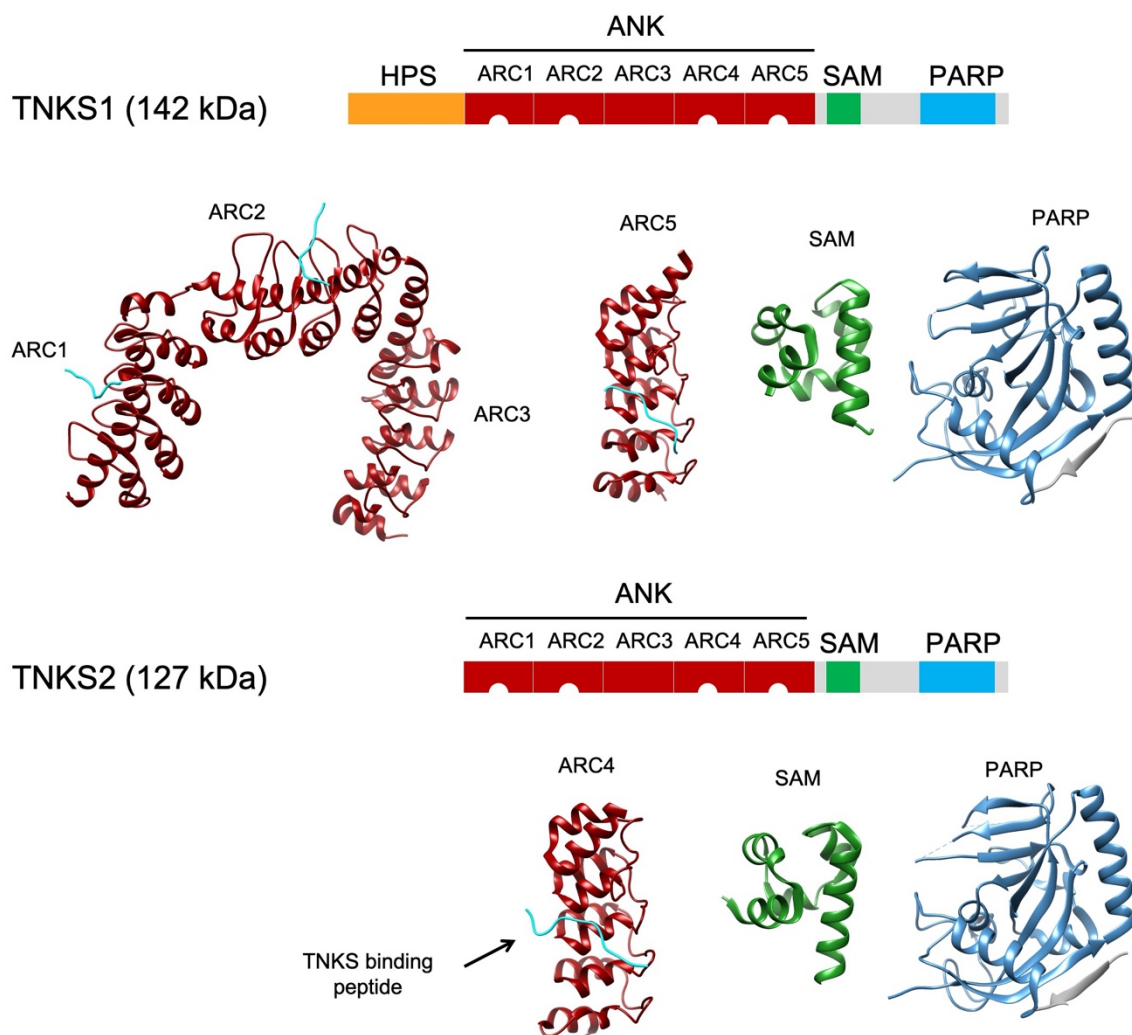
Figure 1.9: TNKS structures. Schematic representation of the structural domains of TNKS1 (top) and TNKS2 (bottom). The TBP binding sites within ARC1, ARC2, ARC4 and ARC5 are represented as semi-circles. The available crystallographic structures of the individual domains are provided in the corresponding colours. For TNKS1, the structures of the ARC1-3 (PDB: 5JHQ), ARC5 (PDB: 5GP7), SAM (PDB: 5JU5) and PARP (PDB: 2RF5) domains are provided. For TNKS2, the structures of the ARC4 (PDS: 3TWR), SAM (PDB: 5JRT) and PARP (PDB: 3KR7) domains are available. The peptides binding to the ARC subdomains are coloured in cyan and the one binding TNKS2 ARC4 is indicated with an arrow. Images were generated using the software UCSF Chimera.

Work from Eisemann *et al.* also revealed that the ARC4-5 subdomains of TNKS1 are highly dynamic and can sample different orientations.[89] This finding suggests that the ANK domain overall acts as a flexible binding platform that can dynamically assume different conformations depending on the relative positioning of the binding peptides in the substrate. This is particularly relevant for multivalent TNKS substrates that have more than one TNKS binding motifs.[89] The same authors, therefore, proceeded to study the multivalent interaction between

TNKS1 ARC1-5 and the two binding motifs of Axin1 and assessed how different pairs of ARC subdomains can cooperatively bind bivalent Axin1 in a 1:1 complex. The results show that an increased binding affinity is achieved when both binding motifs engage with TNKS1 ARC1-5, with a preference for interaction with the ARC2 and ARC5 subdomains. Upon binding, TNKS undergoes a compaction of its ANK domain, which brings different ARC units into close proximity (Figure 1.10). This compact conformation might also place the PARP domain of TNKS closer to Axin1, thereby facilitating Axin1 PARylation. Multivalency, together with sequence specificity and conformational docking mediated by different ARC combinations, might therefore dictate the substrate specificity of TNKS.[89]
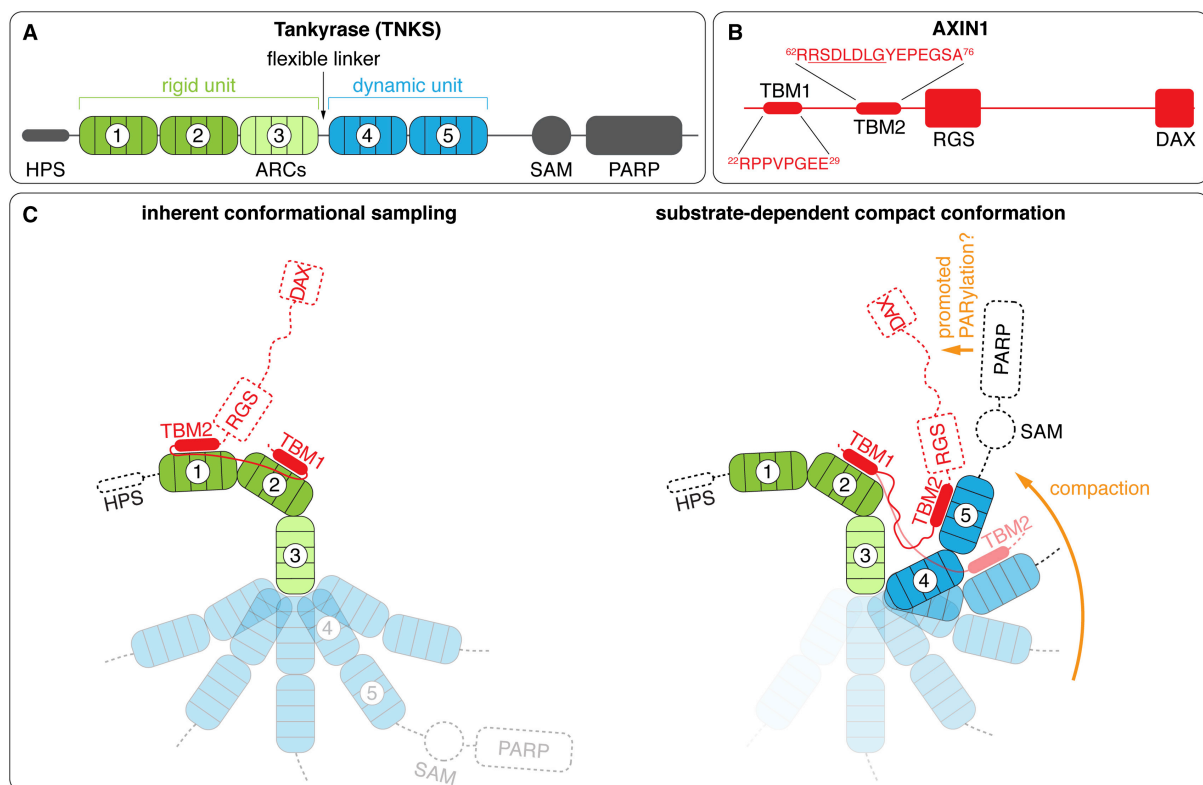
Figure 1.10: TNKS conformational changes upon binding to Axin1. (A) Schematic representation of TNKS1, with the flexible linker region highlighted between ARC3 and ARC4. (B) Schematic representation of Axin1, with the sequence of the two TNKS binding motif (TBM) provided. (C) Schematic representation of the multivalent binding between TNKS1 and Axin1. When Axin1 binds ARC2 and ARC5, a conformational compaction is induced in TNKS. Figure obtained from Guettler.[99]

1.3.2 Localisation and functions of Tankyrase (TNKS)

TNKS proteins have been implicated in the regulation of a large number of cellular processes and have been identified in various subcellular localisations due to the broad spectrum of

proteins with which they interact. Overexpression of TNKS has also been reported in several types of tumours, further validating TNKS involvement in cancer development through different pathways.[100–103] A proteomic analysis of the TNKS interaction network has, indeed, identified over 100 high-confidence interacting proteins for TNKS (Figure 1.11).[104] The same study also highlighted that 85% of the hits overlap between TNKS1 and TNKS2, suggesting that TNKS share redundant or related functions.[104] TNKS1 and TNKS2 ARC subdomains bind to **R**XXΦ[D/E]**G**X[D/E] motifs in their substrates, leading to substrate PARylation.[46,90,91] For most TNKS substrates, including TNKS itself, PARylation causes the protein to be recognised by the PAR-dependent E3 ubiquitin ligase RNF146 (RING finger 146), resulting in protein ubiquitination and degradation via the proteasome.[105] RNF146, however, can also bind TNKS1 directly.[92,106] TNKS1 was initially identified as being involved in telomere maintenance by inhibiting the DNA-binding telomere repeat factor 1 (TRF1) in the nucleus. PARylation of TRF1 inhibits the ability of TRF1 to bind the telomeric repeats, resulting in telomere elongation.[87,91,107,108] More recently, the role of TNKS in the regulation of Wnt pathway activity has received particular attention (further described in Section 1.3.3). TNKS play an essential role in controlling ß-catenin levels (discussed further in the next section), thereby having an effect on many other cellular processes including cell proliferation, embryonic and cancer development.[109] Both TNKS proteins are also involved in insulin-stimulated glucose uptake and homeostasis at the Golgi, through their interaction and co-localisation with IRAP (insulin-responsive aminopeptidase)-containing storage vesicles.[110] They also regulate the tumor-suppressive Hippo signalling pathway through their binding to the angiomotin (AMOT) family of proteins,[111,112] and regulate the LKB1/AMPK (AMP-activated protein kinase) signalling pathway through their interaction with Liver kinase B1 (LKB1).[113] TNKS1 but not TNKS2 appears to be essential for mitotic spindle formation through the binding with NuMA (Nuclear Mitotic Apparatus protein 1)[114,115] and for sister telomere resolution during mitosis.[116] Additionally, TNKS1 regulates DNA repair by PARylating the DNA-PK (DNA-dependent protein kinase), an important effector of the non-homologous end-joining mechanism of DNA repair.[117] TNKS2, instead, regulates the Src signalling via binding to the protein 3BP2. Mutations in the TNKS-binding motif of 3BP2 abolish TNKS2-mediated degradation and underlie a disease called cherubism, characterised by the inflammation of the facial bone.[118] Other TNKS substrates include BLZF1[105], CASC3[105], CPAP at centrosomes[119] and PTEN.[120] However, for many substrates, the details of the TNKS interaction and the biological effects remain unclear.
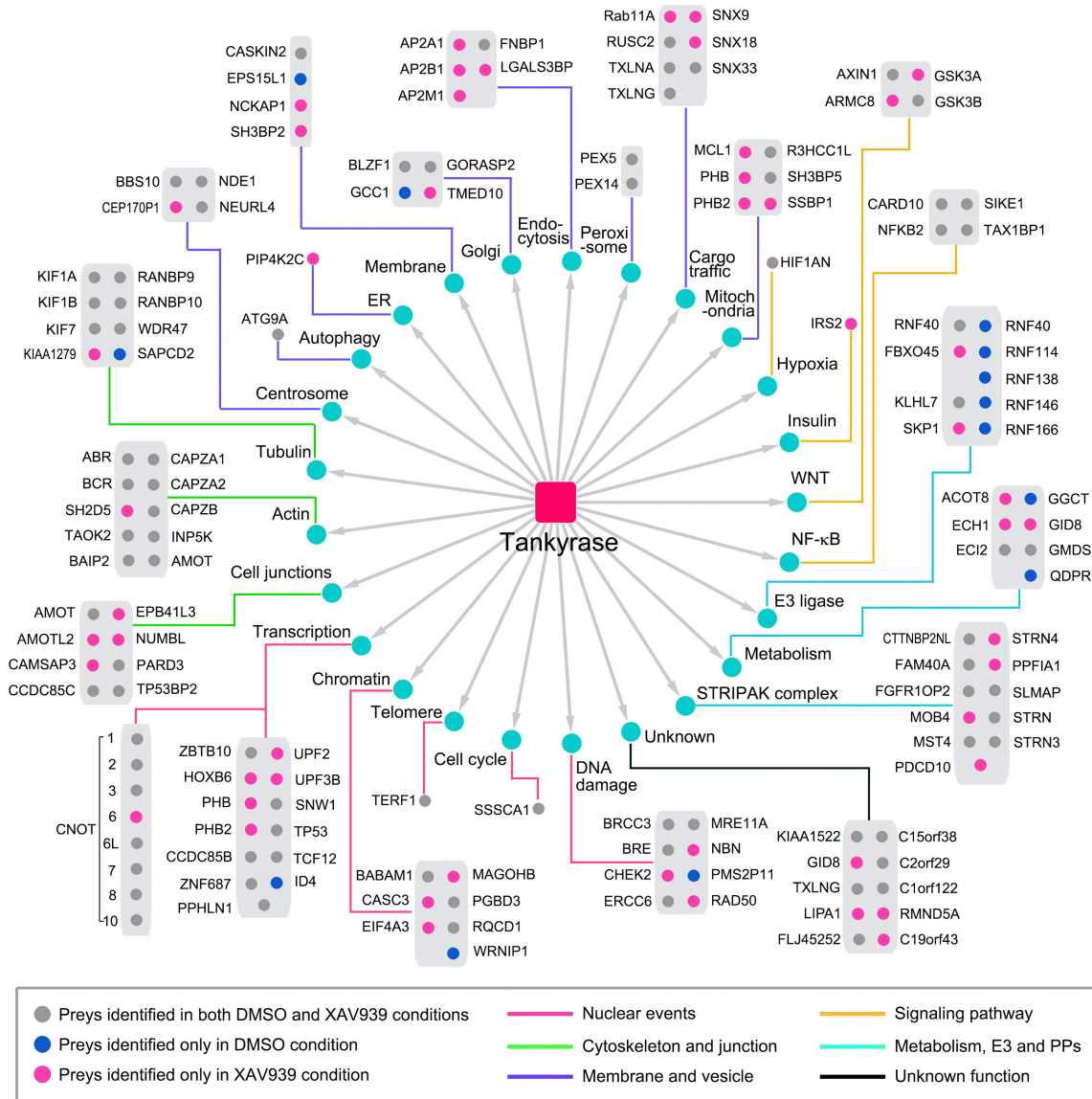
Figure 1.11: TNKS protein interaction network. High-confidence interacting proteins are grouped according to their cellular function and localisation. Figure obtained from Li *et al*.[104]

1.3.3 TNKS in the Wnt signalling pathway

The canonical Wnt signaling pathway is an evolutionarily conserved pathway that plays a major role in tissue homeostasis and regeneration, carcinogenesis and embryogenesis.[121] TNKS controls the Wnt pathway by coordinating two of its major players: Axin1/2 (axis inhibition protein 1/2) and, indirectly, ß-catenin. TNKS binds and PARylates Axin1/2, targeting it for proteasome-mediated destruction and maintaining it at very low levels.[105,109] Axin1/2 is the concentration-limiting component of the ß-catenin destruction complex (ßDC) and directly binds the other core components: APC (adenomatous polyposis coli), GSK3

(kinases glycogen synthase kinase 3) and CK1 (casein kinase 1), and ß-catenin. Under basal signalling conditions (Figure 1.12, left), the ßDC tightly regulates the intracellular levels of ß-catenin by its phosphorylation, which leads to ß-catenin ubiquitination by the E3 ubiquitin ligase SCF$^{β-TrCP}$ and to its proteasome-mediated destruction.[122] Upon binding of a Wnt protein to the extracellular domains of the trans-membrane Frizzled receptor and the co-receptor LRP5/6, the Wnt signaling pathway is activated (Figure 1.12, right). Upon Wnt stimulation, the ßDC is sequestered to the plasma membrane, forming the so-called Wnt signalosome. Although the mechanism is still not fully clear, PARylation of Axin enhances its interaction with the Wnt co-receptor LDL receptor-related protein 6 (LRP6), causing the accumulation of Axin, and thereby the ßDC, to the plasma membrane.[123,124] The ßDC inactivation at the plasma membrane results in cytoplasmic β-catenin accumulation, leading to its translocation into the nucleus and hence the transcription of Wnt target genes through the T-cell factor/lymphoid enhancer factor (TCF/LEF) family of transcription factors.[125–127] The Wnt pathway is the principle driver of colon cancer, where it is dysregulated in around 90% of cases (most often by mutations in APC and less frequently in ß-catenin itself).[128,129]
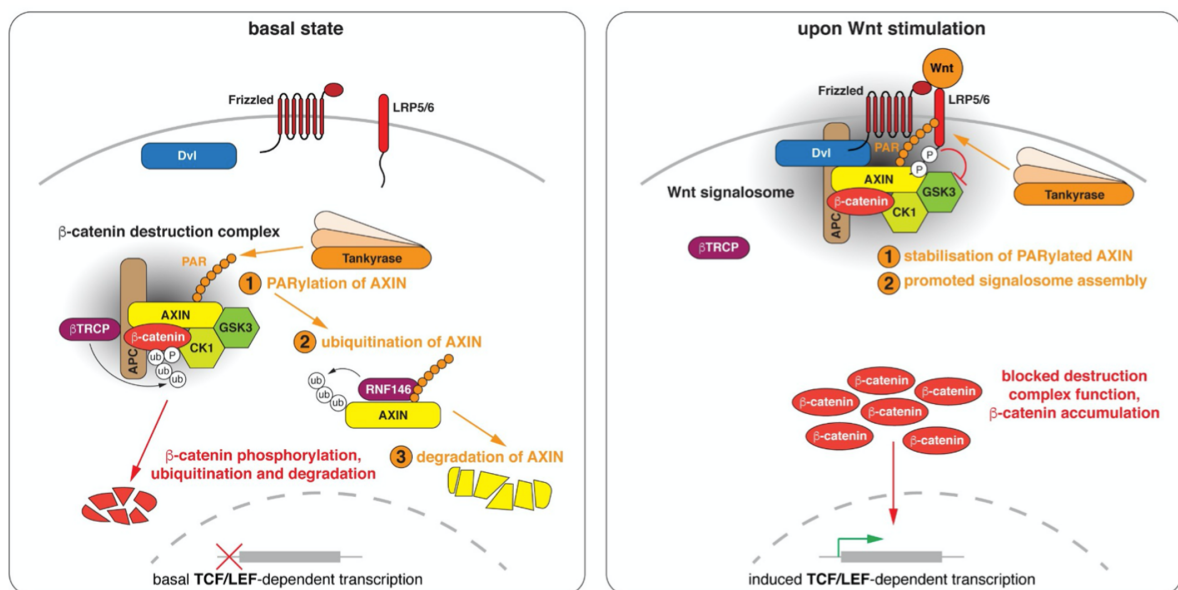
Figure 1.12: Function of TNKS in the Wnt signalling pathway. Schematic representation of the Wnt signalling pathway under basal state (left) and upon Wnt stimulation (right). Under basal state, TNKS PARylates Axin, leading to its proteasomal degradation and the regulation of the ß-catenin destruction complex (ßDC) levels. Upon activation of the Wnt signalling pathway, the ßDC is stabilised at the plasma membrane, resulting in ß-catenin accumulation and its translocation into the nucleus, where can induce the transcription of TCF/LEF-dependent genes. Figure obtained from Mariotti *et al*.[130]

1.3.4 Non-catalytic functions of TNKS

More recently, non-catalytic functions of TNKS have been identified.[93,94] The scaffolding functions of TNKS are mediated by the SAM and ARC domains. Studies have demonstrated that TNKS can promote Wnt signalling independently of its catalytic activity, as PARP-inactive TNKS mutants are still able to activate the Wnt-dependent transcription of the reporter gene. On the other hand, deletions of the SAM or the ARC domains completely abrogated TNKS-dependent reporter activation.[93,94] The same studies also showed that SAM polymerisation induces TNKS self-oligomerisation, and higher-order TNKS structures correlate with enhanced PARP activity. Moreover, SAM-mediated-oligomerisation alters TNKS intracellular localisation, with wild-type TNKS displaying a punctate, cytoplasmic distribution and SAM-mutant TNKS having a more diffused localisation.[93,94] The loss of activity of the SAM-mutant TNKS in the Wnt signalling pathway also led to the discovery that SAM-mediated TNKS polymerisation promotes its interaction with Axin1 in the formation of ß-catenin degradasomes.[93] TNKS scaffolding functions have also been observed in other processes such as pexophagy[104] and apoptosis.[131] Overall, the scaffolding functions of TNKS remain incompletely understood but are expected to extend beyond Wnt signalling, given the varied functions of TNKS-binding proteins.

1.4 Inhibiting TNKS

Due to their important biological roles, the TNKS proteins have become attractive therapeutic targets, and TNKS inhibitors could potentially have a broad clinical utility.[132] They have been mostly developed and tested for diseases associated with Wnt-dependent dysfunctions, in particular colon cancer. Within the Wnt signaling pathway, TNKS inhibition prevents Axin PARylation, leading to Axin accumulation and the formation of the ßDC. Consequently, β-catenin phosphorylation and degradation cause Wnt response gene transcription to be turned off. TNKS are potential anticancer targets additionally because of their involvement in telomere elongation.[133] Within the LKB1-AMPK signalling pathway, instead, TNKS inhibition was shown to activate AMPK and induce anti-tumor as well as anti-diabetic effects in mice.[113] Several TNKS inhibitors have been developed to date, and all drug discovery efforts have been focused on inhibiting the catalytic PARylation activity.[130,132] All small molecule TNKSi developed to date, indeed, are mimetics of NAD^+ and they engage PARP's adenosine-binding subsite, the nicotinamide subsite or both sites. The most well-characterised small molecule inhibitors are XAV939[109], IWR-1 and IWR-2[134], JW55[135], JW74[136], and WIKI4[137], which were

all discovered in phenotypic screens designed to identify Wnt signalling antagonists. A second generation of TNKSi has led to the development of G007-LK.[138]

1.4.1 Effects of TNKS inhibitors (TNKSi)

Almost all colorectal tumours carry mutations in ß-catenin or the ßDC component APC, causing an activation of the Wnt signalling pathway.[128,129] Therefore, several small molecule TNKSi have been tested as a means of selectively modulating dysregulated Wnt signalling.[109,130,138,139] G007-LK and G244-LM were shown to attenuate Wnt signalling in colorectal cancer cell lines and inhibit tumor growth in APC-mutant xenograft models.[138] Similarly, treatment of colorectal cancer cell lines with XAV939 correlated with increased ß-catenin degradation, resulting in reduced proliferation and colony formation rate.[109] XAV939 also seemed to attenuate lung fibrosis in mice.[140] Effective inhibition of the Wnt signalling was also observed when treating colorectal cancer cells with IWR-1 and IWR-2[134], JW55[135], JW74[136] and WIKI4.[137]

However, as the PARP domain is shared with the other PARP family members, the specificity of these small molecule TNKSi for TNKS over other PARP proteins has been found to be quite variable, with none being completely selective for TNKS.[130] Likewise, many other enzymes also use NAD^+ as a co-substrate, and therefore targeting the NAD^+ donor site of TNKS might cause additional off-target effects. This promiscuity, together with the inhibition of Wnt signalling, has been shown to cause cellular toxicity in mice, mostly in the intestine.[138,141] This intestinal toxicity in particular, together with other adverse effects, has prevented most TNKSi from entering clinical trials. Additionally, these TNKSi prevent substrate PARylation as well as TNKS self-PARylation, causing intracellular accumulation of substrates and TNKS itself.[109,138] Lastly, current TNKSi only inhibit the catalytic functions of TNKS but not the well-documented non-catalytic functions.[93,94]

Currently, there is only one ongoing clinical trial testing a new TNKSi in metastatic breast cancer patients. 2X-121 is a small molecule inhibitor of PARP1/2 and TNKS1/2, administered orally on a daily basis, and developed by Oncology Venture A/S. This PARP inhibitor demonstrated clinical activity in a number of solid tumors tested during the prior Phase 1 study.[142] Other ongoing clinical trials are testing the efficacy of olaparib (AZD2281, AstraZeneca), a PARP inhibitor with higher specificity for PARP1 and PARP2 than TNKS.[142,143] Therefore, the development of a highly specific TNKS inhibitor remains to be achieved. As in nature, a high-affinity and high-specificity inhibitor might be obtained through

multivalency and avidity, neither of which are straightforward to realise with conventional small molecules or peptide technologies. Consequently, new approaches are needed, and biologics might represent the answer to this problem.

1.5   Overall thesis motivation and aims

Cancer remains a major cause of death worldwide, and novel more efficient therapies are therefore needed. Driven by this motivation and the absence of effective TNKS inhibitors, the aim of this project was to generate highly specific TNKS inhibitors by targeting the substrate-binding ARC subdomains instead of the PARP domain. Given the fact that the ARC subdomains are only found in the TNKS proteins and that the two TNKS share a high degree of homology, this strategy would allow selective inhibition of both TNKS1 and TNKS2. Guettler *et al*. described a consensus TNKS-binding peptide (TBP) recognised by ARC subdomains 1, 2, 4 and 5.[46] The same consensus TBP sequence was further optimised by previous members of the Itzhaki group to generate a proteolytically stable constrained (using chemical cross-linking) peptide that was shown to have inhibitory functions in the cell.[85] In this project, the TBP sequence was constrained by grafting it onto the loop between consecutive repeats of a CTPR protein to generate a set of potential biotherapeutics as inhibitors of TNKS. As proof of concept, a series of single- and multivalent TNKS-binding proteins were generated by exploiting the modular architecture of the repeat-protein scaffold. These designed proteins were first characterised for their structures, stability and binding and then assayed for their TNKS inhibitory activity, and the applicability of this scaffold as a biotherapeutics platform was explored. Taking our cue from nature, multivalent TNKS-binding proteins were produced to increase binding affinity when they bind to TNKS, which itself is multivalent due to its multiple substrate-binding ARC subdomains. A prime example of TNKS multivalency is provided by its substrate Axin1, which binds TNKS through two separate TNKS-binding peptides.[89,144] Another example is RNF146, which contains five TNKS-binding motifs and can bind multiple ARCs simultaneously.[92] By binding to the ARC subdomains, these designed TBP-CTPR proteins might have the ability to block both catalytic and non-catalytic functions of TNKS, and they may therefore display enhanced activity compared to small molecule inhibitors targeting the catalytic PARP domain. The TBP-CTPR protein series was then expanded to generate bi-specific TNKS degraders, being potentially more effective and acting catalytically. Overall, the modular repetitive architecture of the CTPR proteins can be exploited by grafting onto them one or multiple binding functions and could be used as a platform for

intracellular target inhibition and also target degradation. Thus, this new approach paves the way for the development of repeat-protein arrays as a new class of biologic drugs.

Chapter 2

Materials and Methods

2.1   Buffers and Reagents

All chemicals, enzymes and buffers for DNA purification, gel extraction, clean up were purchased from Sigma-Aldrich, Qiagen or Thermo Scientific unless otherwise specified. All unmodified DNA oligo nucleotides were purchased from Integrated DNA Technologies (IDT). Listed below are recipes for general buffers and solutions. Specific buffers are described in the text were used. Milli-Q standard water was used to make all buffers and all molecular biology work.

- Ampicillin (Amp) stock: 50 mg/mL in H_2O, sterile filtered. Used at 1:1000 dilution
- Dithiothreitol (DTT): 1 M in H_2O, sterile filtered.
- Isopropylthiogalactoside (IPTG): 1 M in H_2O, sterile filtered.
- 50X TAE: 242 g Trizma base, 57.1 mL glacial acetic acid, 18.6 g EDTA to a final volume of 1 L with H_2O. Stored at room temperature and used at 1X for DNA agarose gel electrophoresis and running buffer.
- Tris(2-carboxyethyl)phosphine (TCEP): 0.5 M in H_2O, sterile filtered.
- 10X TGS: 30 g Trizma base, 144 g glycine, 10 g SDS to a final volume of 1 L with H_2O. Stored at room temperature and used at 1X for protein gel electrophoresis and running buffer.

All DNA gels were cast at a percentage of agarose (Appleton Woods) ranging between 0.8% - 1% and stained with SYBR® Safe DNA Gel Stain used at a 1:10000 dilution. All protein gels

were casted at a varying percentage of Acrylamide/Bis-acrylamide, 30% solution between 8% - 15% and stained with Coomassie Brilliant Blue G250 (Sigma-Aldrich).

2.2   *Escherichia coli* strains

Two *Escherichia coli* (*E. coli)* strains were used in this work:

DH5α: K-12 strain derivative optimised for molecular cloning applications. Chemically competent cells were propagated in-house for routine cloning needs; cells for high efficiency transformations were purchased from New England Biolabs (NEB).

C41 (DE3) pLysS: derivative of the BL21 strain used for high-level expression of non-toxic heterologous genes under the T7 promoter. The strain indeed contains the λDE3 lysogen that carries the gene for T7 RNA Polymerase under the control of the lacUV5 promoter. The strain was purchased from Lucigen and propagated in-house following the protocol below. pLysS is a chloramphenicol-resistant plasmid that expresses T7 lysozyme in small amount.

Chemically competent *E. coli* cells were prepared from a glycerol stock of the above-mentioned strains. They were inoculated into 5 mL of LB medium, containing 34 µg/mL of chloramphenicol when the C41(DE3) pLysS was propagated, and incubated at 37ºC in a shaking incubator (200 rpm) overnight. The overnight culture was diluted 1:100 in fresh LB medium (containing antibiotic if appropriate) and grown while shaking at 37ºC until OD_{600} reached 0.25-0.30. The cell suspension was transferred into 50 mL tubes and cooled on ice for 10 minutes to stop growth before centrifugation at 4000 xg for 5 minutes at 4ºC. The pellet in each tube was resuspended thoroughly in 10 mL cold Transformation Buffer I (30 mM KOAc pH 5.8, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 3 mM Hexamine Cobalt Cl, 15% (v/v) glycerol), combined together and incubated on ice for 5 minutes before renewed centrifugation. The cell pellet was resuspended in 5 mL Transformation Buffer II (10 mM MOPS pH 6.5, 10 mM RbCl, 75 mM $CaCl_2$, 15% (v/v) glycerol) and incubated on ice for 15 minutes. The bacterial suspension was dispensed in 50 µL aliquots into pre-chilled microfuge tubes and stored at -80ºC.

All chemically competent *E. coli* cells were transformed with a few ng of DNA (up to 100 ng) by standard heat shock method. The transformed cells were plated onto LB agar plates containing the appropriate antibiotic and incubated at 37ºC overnight.

2.3 Molecular Biology

2.3.1 Plasmids

Two types of *E. coli* expression vectors were used depending on the desired purification method:

pRSet B vector: a commercially available vector (Thermo Fisher) allowing N-terminal Poly-(6)-His tagging of recombinant proteins. It is used for high protein expression under the T7 promoter. It has been altered in-house to replace the Xpress™ Epitope with a Thrombin-cleavage site between the His-tag and the multiple cloning site (MCS).

pGST vector: an in-house variant of the pRSet B, in which the His-tag has been exchanged to a glutathione S-transferase (GST) tag.

For mammalian transfection and expression, two vectors were used:

pcDNA3.1(-) vector: an in-house variant of the commercially available pcDNA3.1(-) vector (Thermo Fisher), which enables C-terminal HA tagging of the gene of interest. It induces high protein expression under the CMV promoter.

HiBiT-pcDNA3.1(-) vector: an in-house variant of the above pcDNA3.1(-) vector, which allows to add a HiBiT tag (developed by Promega) at the N-terminal end of the gene of interest, as well as a HA tag at the C-terminus. The HiBiT tag is followed by a 6-amino acid linker (GSSGGS) to ensure accessibility of the tag, as suggested on the Promega Nano-Glo HiBiT Lytic Detection System Technical Manual.[145]

Inserts encoding for the genes of interest were amplified by polymerase chain reaction (PCR) or purchased as double-stranded gBlocks® Gene Fragments (Integrated DNA Technologies) were codon optimised for enhanced expression in *E. coli*. They were cloned into bacterial expression pRSet B or pGST vectors between BamHI and HindIII restriction sites, following procedures detailed in Section 2.3.2 and Section 2.3.3. Most constructs were then subcloned into the mammalian pcDNA3.1(-) and/or HiBiT-pcDNA3.1(-) vectors between BamHI and EcoRI restriction sites, as explained in Section 2.3.5.

The gene encoding the full-length (FL) human TNKS2 was kindly provided by our collaborator Dr. Marc de la Roche in a pLP dMyc SD plasmid. The DNA sequence encoding the FL TNKS2 protein or the ARC domains was amplified and subcloned into several expression vectors, depending on the need.

DNA plasmids for bacterial expression generated in this work are listed in Table 2.1. The vector maps are provided in Appendix A. The protein sequence of the generated CTPR constructs is provided in Appendix B.

Name	Vector	Promoter	Protein tag	Notes
1TBP-CTPR2	pRSet B	T7	N-terminal His$_6$	Linear
2TBP-CTPR4	pRSet B	T7	N-terminal His$_6$	Linear
3TBP-CTPR6	pRSet B	T7	N-terminal His$_6$	Linear
4TBP-CTPR8	pRSet B	T7	N-terminal His$_6$	Linear
1TBP-CTPR2-foldon	pRSet B	T7	N-terminal His$_6$	Trimeric
2TBP-CTPR4-foldon	pRSet B	T7	N-terminal His$_6$	Trimeric
3TBP-CTPR6-foldon	pRSet B	T7	N-terminal His$_6$	Trimeric
4TBP-CTPR8-foldon	pRSet B	T7	N-terminal His$_6$	Trimeric
CTPR6 [*]	pRSet B	T7	N-terminal His$_6$	Linear, Control
CTPR6-foldon	pRSet B	T7	N-terminal His$_6$	Trimeric, Control
TNKS2 ARC1-3	pRSet B	T7	N-terminal His$_6$	Residues 2-485
TNKS2 ARC1-5	pRSet B	T7	N-terminal His$_6$	Residues 2-797
1TBP-CTPR4-Solv H	pGST	T7	N-terminal GST	Linear, solvating helix
TNKS2 ARC4 [*]	pGST	T7	N-terminal GST	Residues 488-649
TNKS2 ARC1-5	pGST	T7	N-terminal GST	Residues 2-797

Table 2.1: List of DNA plasmids for bacterial expression.

[*] This plasmid was already available in the lab.[85,146]

DNA plasmids for mammalian expression generated in this work are listed in Table 2.2 (single-function constructs) and Table 2.3 (hetero-bifunctional constructs). The vector maps are provided in Appendix A. The protein sequence of the generated CTPR constructs is provided in Appendix B.

Name	Vector	Protein tag	Notes
1TBP-CTPR2	pcDNA3.1(-)	C-terminal HA	Linear
2TBP-CTPR4	pcDNA3.1(-)	C-terminal HA	Linear
3TBP-CTPR6	pcDNA3.1(-)	C-terminal HA	Linear
4TBP-CTPR8	pcDNA3.1(-)	C-terminal HA	Linear
1TBP-CTPR2-foldon	pcDNA3.1(-)	C-terminal HA	Trimeric
2TBP-CTPR4-foldon	pcDNA3.1(-)	C-terminal HA	Trimeric
3TBP-CTPR6-foldon	pcDNA3.1(-)	C-terminal HA	Trimeric
4TBP-CTPR8-foldon	pcDNA3.1(-)	C-terminal HA	Trimeric
CTPR2	pcDNA3.1(-)	C-terminal HA	Linear, Control
CTPR6	pcDNA3.1(-)	C-terminal HA	Linear, Control
3RL-CTPR6	pcDNA3.1(-)	C-terminal HA	Linear, Control
CTPR6-foldon	pcDNA3.1(-)	C-terminal HA	Trimeric, Control
1TBP-CTPR2-mCherry	pcDNA3.1(-)	C-terminal mCherry	Linear
3TBP-CTPR6-mCherry	pcDNA3.1(-)	C-terminal mCherry	Linear
3TBP-CTPR6-foldon-mCherry	pcDNA3.1(-)	C-terminal mCherry	Trimeric
CTPR6-mCherry	pcDNA3.1(-)	C-terminal mCherry	Linear, Control
1TBP-CTPR2	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Linear
2TBP-CTPR4	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Linear
3TBP-CTPR6	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Linear
4TBP-CTPR8	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Linear
1TBP-CTPR2-foldon	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Trimeric
2TBP-CTPR4-foldon	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Trimeric
3TBP-CTPR6-foldon	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Trimeric
4TBP-CTPR8-foldon	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Trimeric
CTPR2	HiBiT-pcDNA3.1(-)	N-terminal HiBiT	Linear, Control

		C-terminal HA	
CTPR6	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Linear, Control
3RL-CTPR6	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Linear, Control
CTPR6-foldon	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	Trimeric, Control
TNKS2 ARC1-eGFP	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal eGFP, HA	
TNKS2 ARC1-5-eGFP	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal eGFP, HA	
TNKS2 FL	HiBiT-pcDNA3.1(-)	N-terminal HiBiT C-terminal HA	
TNKS2 FL (no HA)	HiBiT-pcDNA3.1(-)	N-terminal HiBiT	No HA-tag

Table 2.2: List of DNA plasmids for mammalian expression, encoding single-function and control CTPR constructs, or TNKS2 constructs of various length.

Name	Vector	Protein tag	Notes
2TBP-CTPR4-ABBA	pcDNA3.1(-)	C-terminal HA	Hetero-bifunctional
2TBP-CTPR4-DBOX	pcDNA3.1(-)	C-terminal HA	Hetero-bifunctional
2TBP-CTPR4-KEN	pcDNA3.1(-)	C-terminal HA	Hetero-bifunctional
2TBP-CTPR4-Nrf2	pcDNA3.1(-)	C-terminal HA	Hetero-bifunctional
2TBP-CTPR4-p27	pcDNA3.1(-)	C-terminal HA	Hetero-bifunctional
2TBP-CTPR4-p53	pcDNA3.1(-)	C-terminal HA	Hetero-bifunctional
2TBP-CTPR4-PHYL	pcDNA3.1(-)	C-terminal HA	Hetero-bifunctional
2TBP-CTPR4-SPOP	pcDNA3.1(-)	C-terminal HA	Hetero-bifunctional
2TBP-CTPR4-RTPR2	pcDNA3.1(-)	C-terminal HA	Control
CTPR2-ABBA	pcDNA3.1(-)	C-terminal HA	Control
CTPR2-DBOX	pcDNA3.1(-)	C-terminal HA	Control
CTPR2-KEN	pcDNA3.1(-)	C-terminal HA	Control
CTPR2-Nrf2	pcDNA3.1(-)	C-terminal HA	Control
CTPR2-p27	pcDNA3.1(-)	C-terminal HA	Control
CTPR2-p53	pcDNA3.1(-)	C-terminal HA	Control

| CTPR2-PHYL | pcDNA3.1(-) | C-terminal HA | Control |
| CTPR2-SPOP | pcDNA3.1(-) | C-terminal HA | Control |

Table 2.3: List of DNA plasmids for mammalian expression, encoding hetero-bifunctional and control CTPR constructs.

2.3.2   Molecular cloning into pRSet B and pGST vectors

The standard molecular cloning approach relying on conventional restriction enzymes digestion and ligation was adopted for construct insertion into the pRSet B or pGST vectors, as illustrated in Figure 2.1. Both vectors present a unique BamHI restriction site immediately after the Thrombin site, and a unique HindIII site upstream the stop codon. The insert was generated by amplifying the region of interest (10 ng) by mixing it with 5X Q5 Reaction Buffer (10 µL), forward and reverse primers introducing the restriction sites (2.5 µL, 10 µM), dNTPs (1 µL, 10 mM), Q5 High-Fidelity DNA Polymerase (0.5 µL, NEB) and Milli-Q water to reach a final volume of 50 µL. When TNKS2 was amplified, the 5X Q5 High GC Enhancer buffer was also included in the reaction mixture, due to the high GC content in the TNKS2 gene. PCR was performed on an Eppendorf Mastercycler Nexus PCR Cycler using protocols having annealing temperature and extension duration optimised for each reaction. Melting temperature (T_m) for each set of primers was calculated using the NEB T_m calculator tool).[147] When the gene of interest was not available, a gBlock® Gene Fragment was designed with the appropriate restriction sites at each terminus, preceded and followed by at least 5 bp to enhance digestion activity of restriction enzymes.

Cloning was performed by digesting the insert (150 ng of gBlock® or 1-10 µg of purified PCR product) with FastDigest BamHI and HindIII (1 µL each, Thermo Scientific) according to the manufacturer's protocol. When a PCR product was used, DpnI was also added to the mixture to digest the methylated template DNA. Reaction was stopped by heat inactivation at 85ºC for 20 minutes or by agarose gel electrophoresis followed by gel extraction and purification according to the QIAquick gel extraction protocol. Similarly, the vector (~1 µg) was digested with the same set of restriction enzymes, with the addition of FastAP Thermosensitive Alkaline Phosphatase (1 µL, Thermo Scientific) and purified by gel extraction. All DNA was eluted with Milli-Q water and stored at -20ºC unless used immediately.

Ligation was performed by mixing the digested vector and digested insert in a 1:3 vector:insert molar ratio (total 3 µL) and 1 µL Anza^{TM} T4 DNA Ligase Master Mix (Thermo Scientific) and incubating the reaction for 15 minutes at room temperature. The whole mixture was transformed into NEB® 5-alpha Competent *E. coli* (High Efficiency) (NEB) as specified in Section 2.2. Usually, a couple of individual colonies were propagated in suspension culture for plasmid standard extraction. Correct incorporation of the insert was verified by renewed digestion using the same enzymes and subsequent agarose gel electrophoresis, before sequence confirmation by Sanger sequencing (Mix2Seq Kit, Eurofins Genomics) using the T7 forward and T7 reverse primers separately. Plasmid concentration was measured using the NanoDrop 2000 (Thermo Scientific).

2.3.3   Construction of tandem-repeat gene arrays from single repeat unit

Tandem repeat arrays of higher order, named $_n$TBP-CTPR$_{2n}$ with n between one and four, were generated by concatemerization of single units using a previously described method,[148] already adopted for building long CTPR tandem repeat constructs.[149] This cloning strategy is necessary as a gBlock® with highly repetitive DNA sequence cannot be purchased. The gBlock® encoding the 1TBP-CTPR2 sequence was designed having a BamHI site and a HindIII site at either end to allow directional cloning into pRSet B. It also included a BglII site at the 3' end preceding the two stop codons and the HindIII site. Following digestion and ligation into the pRSet B vector using the BamHI and HindIII restriction enzymes, the 1TBP-CTPR2 gene cassette was PCR amplified using the T7 forward and T7 reverse primers (Figure 2.1A). To generate the larger $_n$TBP-CTPR$_{2n}$ gene constructs, the 1TBP-CTPR2 PCR product (insert) was digested with BamHI and HindIII and ligated into the $_{n-1}$TBP-CTPR$_{2n-2}$ expression vector previously digested with BglII and HindIII. This is possible because, although BamHI and BglII recognise different 6 bp sequences, they produce complementary sticky ends (Figure 2.1B). The ligation product was transformed into NEB® 5-alpha Competent *E. coli,* DNA extracted and sequenced as detailed in Section 2.3.2. Subsequent plasmid digestion and ligation of 1TBP-CTPR2 units yielded progressively larger concatemers until the gene encoding for 4TBP-CTPR8 was generated. With this method, $_n$TBP-CTPR$_{2n}$ constructs of identical repeats at both the DNA and protein level were generated.

Using the same strategy, the trimeric repeat protein array was generated: the insert encoding the foldon domain was digested with BamHI and HindIII and ligated into the $_n$TBP-CTPR$_{2n}$ expression vectors previously digested with BglII and HindIII.
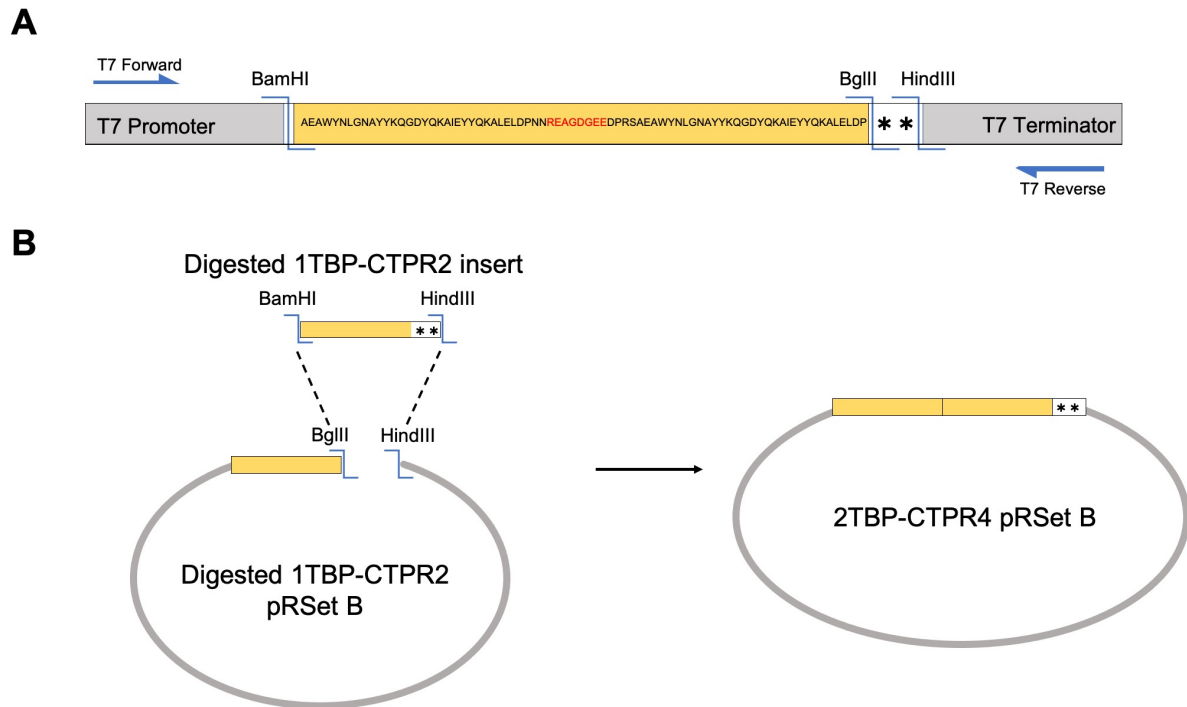
Figure 2.1: Construct design for conventional molecular biology cloning strategies. (A) Schematic representation of the gBlock® construct encoding for 1TBP-CTPR2 (single repeat unit). The construct contains three restriction sites and two stop codons (*). Once inserted into the pRSet B vector, the T7 forward and reverse primers (shown as arrows) can be used for construct PCR amplification and sequencing. (B) The cloning strategy used to generate higher order $_n$TBP-CTPR$_{2n}$ constructs by concatemerisation is illustrated. The insert digested with the BamHI and HindIII restriction enzymes is introduced into the digested vector between the BglII and HindIII restriction sites.

2.3.4 Round-the-Horn site-directed mutagenesis and primer design

An alternative method for directing specific changes at the DNA level such as point mutations, small insertions or deletions is by Round-the-Horn site-directed mutagenesis PCR.[150] The whole plasmid is used as template and is amplified using "back-to-back" non-overlapping primers carrying the mutation to be introduced. Depending on the mutation to be inserted, the primers are designed to contain an internal mismatch, or nucleotides in excess or in defect, therefore generating a point mutation, a small insertion or a deletion after the first amplification cycle, respectively (Figure 2.2). This method has the advantage of introducing mutations at any desired site, without relying on restriction sites present nearby. For Round-the-Horn PCR, 100 µM primers were phosphorylated using T4 Polynucleotide Kinase (Thermo Scientific) in the presence of 1 mM ATP according to the manufacturer's protocol. The enzyme was heat inactivated at 75°C for 10 minutes. PCR was performed as mentioned in Section 2.3.2, allowing a long extension time in each cycle to amplify the entire plasmid (~5 minutes). PCR products

were gel purified, ligated using 1 µL Anza$^{\text{TM}}$ T4 DNA Ligase Master Mix (Thermo Scientific) and transformed into NEB® 5-alpha Competent *E. coli* (High Efficiency) (NEB). As before, successfully mutated plasmids were selected, propagated for plasmid purification and sequenced.
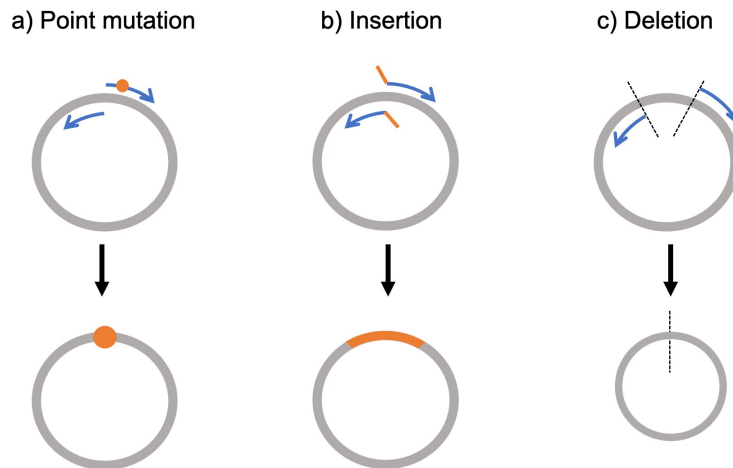
Figure 2.2: Round-the-Horn site-directed mutagenesis. Primer design to introduce a point mutation (a), an insertion (b) or a deletion (c) is illustrated.

2.3.5 Molecular cloning into pcDNA3.1(-) and HiBiT-pcDNA3.1(-)

The backbone of the pcDNA3.1(-) vector (Thermo Scientific) contains a BglII restriction site. Before subcloning each CTPR construct into this vector, the BglII restriction site was mutated by Round-the-Horn site-directed mutagenesis as described in Section 2.3.4 in order to avoid having more than one BglII site in the final plasmid.

The mutated vector was then used as template to introduce the DNA sequence encoding the N-terminal HiBiT tag (licensed from Promega) in frame with the start codon by Round-the-horn PCR, therefore generating the HiBiT-pcDNA3.1(-) vector.

For cloning into the pcDNA3.1(-) and HiBiT-pcDNA3.1(-) vectors, the unique restriction sites BamHI and EcoRI were used. The inserts of interest were amplified by PCR, using a forward primer carrying the BamHI site and a reverse primer introducing the EcoRI site. Both the insert and the vector were digested with BamHI and EcoRI and the same protocol detailed in Section 2.3.2 was followed. Due to the absence of the appropriate restriction sites, eGFP-tagged TNKS2 constructs were instead generated using the NEBuilder HiFi DNA

Assembly Cloning Kit (New England Biolabs), according to the manufacturer's instructions. Correct clones were confirmed by DNA sequencing (Eurofins Genomics).

2.4   Protein expression and purification

Recombinant proteins were expressed with a His$_6$- or a GST-tag to allow purification by affinity chromatography. Few microliters of the pRSet B or pGST expression vector were transformed into chemically competent *E. coli* C41 (DE3) pLysS cells by heat shock and plated on LB-Amp plates. Several colonies were grown at 37ºC in 2xYT medium (Formedium) containing ampicillin (50 µg/mL) and shaking at 200 rpm until the OD$_{600}$ was ~0.6. For all protein constructs, expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.5 mM, for 16-20 h at 20ºC. Cells were harvested by centrifugation at 8000 xg (4ºC, 10 minutes). Purification protocols of His$_6$- or GST-tagged proteins are described below. The buffers used at each purification step are listed in Table 2.4.

2.4.1   Protein solubilisation test

A small aliquot of the above-mentioned culture (1 mL) was pelleted to perform a protein solubilisation test before proceeding with the purification of the entire culture. The bacterial pellet was resuspended in 300 µL of 1X BugBuster® Master Mix (Millipore) to induce cell wall perforation without denaturing soluble proteins. The resuspension was incubated with rotation at room temperature for 30 minutes. The lysate was spun at 12000 xg for 10 minutes to pellet the insoluble fraction and cell debris. The supernatant corresponds to soluble fraction. The insoluble pellet was washed twice in 0.1X BugBuster® Master Mix and centrifuged again. The total lysate, soluble and insoluble fractions were analysed by SDS-PAGE to evaluate the solubility of the protein.

2.4.2   Purification of His$_6$-tagged constructs (pRSet B plasmid)

The bacterial pellet was resuspended in cold Wash Buffer including EDTA-free SIGMAFAST protease inhibitors (Sigma-Aldrich), Lysozyme (Sigma-Aldrich) and DNase I (Sigma-Aldrich). Cells were lysed by high-pressure homogenization using an Emulsiflex C5 homogenizer (Avestin) at pressures between 10000 and 15000 psi. Cell debris and other insoluble fractions were removed by centrifugation at 40000 xg for 40 minutes at 4ºC. All His$_6$-tagged constructs were purified by immobilised metal ion affinity chromatography (IMAC) on a 5 mL HisTrap Excel column (Cytiva) connected to an ÄKTA Pure protein purification system

(Cytiva). The column was washed with 20 CV of Wash Buffer mixed with 4% Elution Buffer (20 mM imidazole final concentration) to prevent nonspecific interaction of lysate proteins to the beads. Proteins were eluted with 5 CV Elution Buffer and the eluted fractions were analysed by SDS-PAGE and pooled accordingly. All His$_6$-tagged $_n$TBP-CTPR$_{2n}$ and CTPR$_n$ proteins were subsequently separated by size-exclusion chromatography (SEC) in isocratic conditions using a HiLoad 16/600 Superdex 75 pg column (Cytiva) pre-equilibrated in Elution Buffer. Proteins to be analysed by size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) were instead purified by SEC using a Superdex 200 10/300 GL column (Cytiva) pre-equilibrated in Elution Buffer.

Occasionally, protein purification of small-scale expression cultures (10-50 mL) was performed in batch using Amintra Ni-NTA resin (Expedeon) using the same wash and elution buffers. The same method applies, with the only difference that the cell pellet is solubilised and lysed in few mL of 1X BugBuster® Master Mix (Millipore) and the supernatant is applied on 1 mL of resin slurry. Beads are pelleted by centrifuging at 1000 xg for 1 minute, to allow washing and elution steps.

His$_6$-tagged TNKS2 proteins were also purified by IMAC on the ÄKTA Pure protein purification system (Cytiva) using the same method. Following affinity chromatography, His$_6$-tagged TNKS2 ARC1-3 and His$_6$-tagged TNKS2 ARC1-5 were further purified by SEC using a HiLoad 16/600 Superdex 75 pg column and HiLoad 16/600 Superdex 200 pg column (Cytiva), respectively, pre-equilibrated in Elution Buffer. The composition of the Wash Buffer and Elution Buffer used for each purification step is listed in Table 2.4.

2.4.3 Expression and purification of GST-tagged variants (pGST plasmid)

E. coli C41 (DE3) pLysS cells expressing GST-tagged recombinant proteins were harvested and lysed as previously described in Section 2.4 and 2.4.2. The supernatant obtained from 1 L culture was incubated for 2 h at 4°C on a rotating mixer with 5 mL slurry of Glutathione Sepharose 4B resin (Cytiva), pre-equilibrated in Wash Buffer. The resin was then washed three times with 50 mL of Wash Buffer, followed by overnight on-matrix cleavage of the fusion protein using 400 units of bovine thrombin (Sigma Aldrich) per litre of culture, at room temperature on a rotating mixer. Cleaved protein was collected by washing the resin with 5 mL Elution Buffer five times. The fractions were combined and concentrated to a final volume of 5 mL. All variants were subsequently separated by SEC in isocratic conditions using a HiLoad 16/600 Superdex 75 pg column (Cytiva) pre-equilibrated in the corresponding Elution Buffer.

The composition of the Wash Buffer and Elution Buffer used for each purification step is listed in Table 2.4.

Protein	Method	Wash Buffer	Elution Buffer
His_6-$nTBP$-$CTPR_{2n}$	HisTrap Excel/ Ni-NTA resin	50 mM Tris-HCl pH 8.0, 150 mM NaCl	50 mM Tris-HCl pH 8.0, 150 mM NaCl, 500mM imidazole
	Size-exclusion		50 mM Tris-HCl pH 8.0, 150 mM NaCl
His_6-TNKS2 ARC1-3	HisTrap Excel	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 2 mM DTT	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 500 mM imidazole, 2 mM DTT
	Size-exclusion		50 mM Tris-HCl, 150 mM NaCl, pH 8.0, 2 mM DTT
His_6-TNKS2 ARC1-5	HisTrap Excel	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 2 mM DTT	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 500 mM imidazole, 2 mM DTT
	Size-exclusion		50 mM Tris-HCl, 150 mM NaCl, pH 8.0, 2 mM DTT
GST-1TBP-CTPR4-SolvH	Glutathione	50 mM Tris·HCl pH 8.0, 300 mM NaCl	50 mM Tris·HCl pH 8.0, 300 mM NaCl
	Size-exclusion		50 mM Tris·HCl pH 8.0, 150 mM NaCl
GST-TNKS2 ARC4	Glutathione	50 mM Tris·HCl pH 8.0, 300 mM NaCl, 2 mM DTT	50 mM Tris·HCl pH 8.0, 300 mM NaCl, 2 mM DTT
	Size-exclusion		50 mM Tris·HCl pH 8.0, 150 mM NaCl, 2 mM DTT

Table 2.4: Buffer system used for the purification of the listed proteins.

2.4.4 Protein concentration, purity and molecular weight determination

Following protein purification, proteins were concentrated using a Vivaspin® centrifugal concentrator of the appropriate size and membrane molecular weight cut-off value. Proteins were flash-frozen and stored at -80ºC until further use. The final purity and identity of all the proteins were determined by SDS-PAGE and mass spectrometry using matrix-assisted laser desorption/ionisation (MALDI, Department of Biochemistry, Cambridge, UK) or liquid

chromatography-mass spectrometry electrospray ionisation (LCMS ESI, Department of Chemistry, Cambridge, UK). MS spectra are provided in Appendix C. Protein absorbance at 280 nm was measured using the NanoDrop 2000 (Thermo Scientific). Protein concentration was calculated according to the Beer-Lambert law, using the theoretical extinction coefficient obtained from the primary amino acid sequence (ExPaSy ProtParam tool)[151] of each protein (Table 2.5).

Protein	Molecular weight (Da)	ε (M^{-1} cm^{-1})
His$_6$-1TBP-CTPR2	11308.23	28880
His$_6$-2TBP-CTPR4	20677.32	57760
His$_6$-3TBP-CTPR6	30046,40	86640
His$_6$-4TBP-CTPR8	39415,48	115520
His$_6$-1TBP-CTPR2-foldon$^{(*)}$	49368.17	112080
His$_6$-2TBP-CTPR4-foldon$^{(*)}$	77475.42	198720
His$_6$-3TBP-CTPR6-foldon$^{(*)}$	105582.67	285360
His$_6$-4TBP-CTPR8-foldon$^{(*)}$	133689.92	372000
His$_6$-CTPR6	26193.75	86640
His$_6$-CTPR6-foldon$^{(*)}$	94024.71	285360
His$_6$-TNKS2 ARC1-3	54853.46	27390
His$_6$-TNKS2 ARC1-5	88086.00	51800
GST-1TBP-CTPR4-Solv H	45402.22 / 19122.60	100620 / 57760
GST-TNKS2 ARC4	43988.59 / 17708.97	55810 / 12950

Table 2.5: Molecular weights and theoretical extinction coefficient of the listed proteins. Extinction coefficient calculated assuming all Cysteine residues being reduced. For GST-tagged constructs, values of molecular weight and ε for each construct before and after Thrombin cleavage are provided. $^{(*)}$ Values of molecular weight and ε for the trimeric construct are indicated.

2.5 SEC-MALS

Purified protein samples (100 µL, protein concentration >1.7 mg/mL) were subjected to SEC-MALS, performed on a SEC Superdex 200 Increase 10/300 GL column (Cytiva) preequilibrated in the corresponding Elution Buffer, in line with a multiangle light scattering

module (DAWN-8+; Wyatt Technologies) and a differential refractometer (Optilab T-rEX; Wyatt Technologies). The light scattering and protein concentration at each point across the peaks in the chromatograph were used to determine the absolute molecular mass from the intercept of the Debye plot using Zimm's model as implemented in the ASTRA v7.3.0.11 software (Wyatt Technologies). To determine interdetector delay volumes, band-broadening constants and detector intensity normalisation constants for the instrument, Bovine Serum Albumin (BSA, Thermo Fisher Scientific) was used as a standard prior-to sample measurement. Data were plotted with the program PRISM (GraphPad Software Inc.). SEC-MALS was performed at the MRC-Laboratory of Molecular Biology (Cambridge, UK) by Dr. Stephen McLaughlin.

2.6 Far-UV Circular Dichroism (CD) spectroscopy and thermal denaturation

All $_n$TBP-CTPR$_{2n}$ protein samples to be analysed by CD were buffer exchanged into 10 mM sodium phosphate buffer pH 7.4, 150 mM NaCl, using a PD-10 Desalting column (Cytiva) and diluted to a final concentration of 20 µM. CD measurements were performed on a Chirascan CD spectrophotometer (Applied Photophysics) at 20°C in a 1 mm pathlength Precision Cell (110-QS, Hellma Analytics). Measurements were acquired at wavelengths between 200 nm and 280 nm using a 1 nm bandwidth at a scan speed of 120 nm per minute. Readings were repeated in triplicate and the data averaged. The molar ellipticity, defined as [θ], was calculated using

$$[\theta] = \frac{\theta}{C \cdot l}$$

Eq. 2.1

where θ is the ellipticity in millidegrees (mdeg), C the sample concentration in molarity (M = mol/L = mol/1000 cm^3), and l the pathlength in centimeters. The value of [θ] has to be divided by 10 to be reported in the conventional units of deg•cm^2•dmol^{-1}, according to the conversion shown below[152]

$$\frac{mdeg}{cm \cdot M} = \frac{mdeg \cdot 1000 \cdot cm^3}{cm \cdot mol} = \frac{deg \cdot cm^2}{mol} = \frac{deg \cdot cm^2}{10\ dmol} = 10^{-1} \cdot deg \cdot cm^2 \cdot dmol^{-1}$$

2.7 Thermal denaturation monitored by CD

The same protein samples used for far-UV CD (Section 2.6) were also analysed by thermal denaturation monitored by CD, using the same instrument and cuvette as above. Temperature

ramp was induced by increasing the temperature of the protein samples from 20ºC to 92ºC in 1ºC steps, and the ellipticity at 222 nm was monitored as a function of temperature. The α-helical content of the sample was followed by plotting the molar ellipticity against temperature. Readings were repeated five times and averaged. Since the heat capacity of the CTPR proteins under measurement is unknown, the molar ellipticity with respect to temperature was fitted to a sigmoidal Boltzmann equation including a sloping baseline term to extract the apparent melting temperature, T_m:

$$Y = \theta_N + \frac{\theta_U - \theta_N}{1 + \exp(\frac{T_m - T}{n})} \qquad \text{Eq. 2.2}$$

where θ_N and θ_U are the ellipticity of the native and denatured states, respectively, and n is the slope of the transition. In a sloping sigmoidal transition, θ_N and θ_U are corrected as

$$\theta_N = \alpha_N + \beta_N T \qquad \text{Eq. 2.3}$$

$$\theta_U = \alpha_U + \beta_U T \qquad \text{Eq. 2.4}$$

where α_N and α_U are the state signal at the lowest and highest temperature, respectively, β_N and β_U are the slopes of the native and unfolded state baseline, respectively.

Following thermal denaturation, the samples were allowed to cool back to 20ºC and the CD spectrum was re-acquired to evaluate the refolding capability of each CTPR protein.

2.8   Equilibrium denaturation monitored by fluorescence spectroscopy

Equilibrium denaturation experiments were performed according to a high-throughput method established in the Itzhaki laboratory.[153] All protein samples were buffer exchanged into 50 mM sodium phosphate buffer, 150 mM NaCl, pH 6.8. A fixed volume (15 μL) of protein sample (at a concentration around 6-10 μM) were dispensed into 96-well, half-area, black polystyrene plates (Corning) using a Microlab ML510B dispenser (Hamilton). Each well contained increasing concentrations of guanidinium hydrochloride (GdmHCl) between 0 M and 6 M in increments of 0.1 M per well. Buffer was added into each well to reach a final volume of 150 μL per well. Plates were incubated at 25ºC for 1 h, covered with 96-well Microplate Aluminium Sealing Tape (Corning) to avoid evaporation. Plate measurements were carried out on a CLARIOstar Plate Reader (BMG LABTECH) with a tryptophan-detection filter set consisting

of an excitation filter (275-285 nm), a dichroic PL325 nm, and an emission filter (350-370 nm) at 25ºC. Each measurement was performed in triplicate.

The plots of fluorescence intensity versus denaturant concentration were fitted to a two-state model, in which only two states are populated during unfolding (the native [N] and fully unfolded [U] states). The change in Gibbs free energy that describes the reversible equilibrium [N] ⇄ [U] can be defined as

$$\Delta G_{N-U} = -RTlnK_{eq} = -RTln\frac{[U]}{[N]}$$

Eq. 2.5

where R is the gas constant and T is the temperature in Kelvin. K_{eq} can be expressed as signal intensity change depending on the concentration of both states, according to the relation

$$\frac{[U]}{[N]} = \frac{S_N - S}{S - S_U}$$

Eq. 2.6

where S_N is the signal at the native state, S_U is the signal at the unfolded state and S is the signal at a particular concentration of denaturant. Combining the previous two equations, we obtain

$$\exp\left(-\frac{\Delta G_{N-U}}{RT}\right) = \frac{S_N - S}{S - S_U}$$

Eq. 2.7

And can be rearranged to obtain

$$S = \frac{S_N + S_U \cdot \exp\left(-\frac{\Delta G_{N-U}}{RT}\right)}{1 + \exp\left(-\frac{\Delta G_{N-U}}{RT}\right)}$$

Eq. 2.8

A linear extrapolation method is commonly used to describe the linear dependence of ΔG_{N-U} on the concentration of denaturant.

$$\Delta G_{N-U} = \Delta G_{N-U}^{H2O} - m[D]$$

Eq. 2.9

where ΔG_{N-U}^{H2O} is the free energy change in the absence of denaturant, [D] is the concentration of denaturant and *m* is a constant proportional to the change in solvent accessible surface area upon denaturation.

We know from Eq. 2.5 that when [U]=[N], ΔG_{N-U} equals zero. Therefore,

$$\Delta G_{N-U}^{H2O} = m[D_{50\%}]$$

Eq. 2.10

where $D_{50\%}$ is the denaturant concentration at the midpoint of the unfolding transition. Therefore,

$$\Delta G_{N-U} = m[D_{50\%}] - m[D] = m([D_{50\%}] - [D])$$

Eq. 2.11

The new relation can be substituted into Eq. 2.8 to get

$$S = \frac{S_N + S_U \cdot \exp\left(-\dfrac{m([D_{50\%}] - [D])}{RT}\right)}{1 + \exp\left(-\dfrac{m([D_{50\%}] - [D])}{RT}\right)}$$

Eq. 2.12

This equation describes an idealised signal trace. The measured signal intensity, however, varies linearly with the concentration of denaturant, independently of any state transition, and this is visualised as sloping baselines. The correction introduces denaturant dependent linear descriptions to S_N and S_U, giving the following equation:[154]

$$S = \frac{\alpha_N + \beta_N[D] + (\alpha_U + \beta_U[D]) \cdot \exp\left(-\dfrac{m([D_{50\%}] - [D])}{RT}\right)}{1 + \exp\left(-\dfrac{m([D_{50\%}] - [D])}{RT}\right)}$$

Eq. 2.13

where α_N and α_U are the state signal at the lowest and highest concentration of denaturant, respectively, and β_N and β_U are the slopes of the native state and unfolded state baseline, respectively.

Assuming that the protein is fully folded in the absence of denaturant and fully unfolded at the highest concentration of denaturant, the signal can be converted into fraction folded, λ, or fraction unfolded, 1-λ, using

$$S = (\alpha_N + \beta_N[D])\lambda + (\alpha_U + \beta_U[D])(1 - \lambda)$$

Eq. 2.14

which can be rearranged into

$$\lambda = \frac{S - \alpha_U - \beta_U[D]}{\alpha_N - \alpha_U + [D](\beta_N - \beta_U)} \qquad \text{Eq. 2.15}$$

or

$$1 - \lambda = 1 - \frac{S - \alpha_U - \beta_U[D]}{\alpha_N - \alpha_U + [D](\beta_N - \beta_U)} = \frac{-S + \alpha_N + \beta_N[D]}{\alpha_N - \alpha_U + [D](\beta_N - \beta_U)} \qquad \text{Eq. 2.16}$$

2.9   Isothermal titration calorimetry (ITC)

ITC was performed at 25°C using a MicroCal VP-ITC (Malvern Panalytical). Proteins were dialysed into ITC Buffer (500 mL of 10 mM sodium phosphate buffer, 150 mM NaCl, 0.5 mM TCEP, pH 7.4) for a total of 3 h. TNKS2 ARC4 was titrated into the sample cell containing one of the $_n$TBP-CTPR$_{2n}$ or control proteins. Injections of TNKS2 ARC4 into the cell were initiated with one injection of 5 µL over 6 seconds, followed by 29 injections of 10 µL over 12 seconds. Raw data were first subjected to baseline determination using NITPIC software[155] and were fitted using the OneSite model within Origin 7.0 software to a non-linear regression. The values of -TΔS were obtained multiplying the change in entropy (ΔS, directly measured by the instrument) by the temperature (T, in Kelvin). The values of Gibbs free energy (ΔG) were then calculated using

$$\Delta G = \Delta H - T\Delta S \qquad \text{Eq. 2.17}$$

where ΔH is the change in enthalpy, directly measured by the instrument.

2.10   Tissue culture

HEK293T cells were cultured in 15 mL of cell growth medium consisting of Dulbecco's Modified Eagles Medium (DMEM), high glucose, pyruvate, (Thermo Fisher Scientific) supplemented with 10% Fetal Bovine Serum (Sigma-Aldrich) and 1X Penicillin/Streptomycin (Gibco) in adherence in a T75 flask (Sarstedt). Cells were incubated at 37°C, 5% CO$_2$ and upon reaching confluency, cells were subcultured at a 1:6 ratio. For splitting, cells were washed with 1X PBS (Thermo Fisher Scientific), detached from the surface using trypsin-EDTA (Gibco) and diluted in growth medium.

2.11 HiBiT-qIP

2.5×10^6 HEK293T cells were seeded in a 10 cm dish and allowed to adhere overnight. The following day, cells were transfected with 3.5 μg of pcDNA3.1(-) vector encoding HiBiT-tagged full-length TNKS2 and 3.5 μg of pcDNA3.1(-) vector encoding HA-tagged 3TBP-CTPR6, CTPR6 or empty vector. Transfection was performed using Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Following 48 h from transfection, cells were washed twice in 1X PBS and lysed for 30 minutes in 1 mL cold lysis buffer (50 mM Tris-HCl, 225 mM KCl, 1% NP-40, pH 7.5) including protease inhibitor cocktail (Sigma-Aldrich), 10 mM NaF, 1 mM Na_3VO_4, 1 mM PMSF. Cell debris was removed by centrifugation at 16000 xg for 20 minutes. The supernatant was incubated with 20 μL monoclonal anti-HA Agarose (Sigma-Aldrich), preequilibrated in lysis buffer, with gentle rotation for 4 h. Anti-HA agarose resin was washed 4 times with 500 μL lysis buffer. Elution was performed by adding 20 μL 2X Loading dye containing SDS to the settled resin. All steps following cell washing were performed at 4°C. The presence of HiBiT-tagged TNKS2 throughout the IP process was detected by mixing 5 μL of each sample with the Nano-Glo HiBiT Lytic Detection System according to the manufacturer's instructions (Promega). The amount of HiBiT-tagged TNKS2 bound to the resin before elution was instead quantified by mixing 5 μL of 50% anti-HA Agarose resin slurry with the Nano-Glo HiBiT Lytic Detection System. Data were obtained from two biological replicates.

2.12 Western Blot and HiBiT Blot on HiBiT-qIP samples

HiBiT-qIP samples to be analysed by Western Blot were run on a 10% polyacrylamide gel and transferred into a nitrocellulose membrane (pore size 0.20 μm, Pharmacia Biotech) by semi-dry transfer in a three-buffer system[156] comprising Anode I buffer (0.3M Tris pH 10.4, 10% methanol), Anode II buffer (25mM Tris pH 10.4, 10% methanol) and Cathode buffer (25mM Tris pH9.4, 192mM 6-amino-n-caproic acid, 10% methanol). Transfer was performed on a Pierce[TM] Power Blot Cassette (Thermo Scientific) with voltage set at 15 V and current at 2.5 A for 25 minutes. The membrane was cut into three parts to allow multiple detections on the same samples. HiBiT-tagged TNKS2 was visualised using the Nano-Glo HiBiT Blotting System according to the manufacturer's instructions (Promega), which does not require membrane blocking. The other parts of the membrane were blocked in 5% milk (Marvel), shaking at room temperature for 1 h. HA-tagged 3TBP-CTPR6 and CTPR6 were detected using anti-HA-Tag (C29F4, 1:1000 dilution) rabbit monoclonal antibody (Cell Signaling

Technology). Tubulin was identified using anti-alpha Tubulin antibody (ab7291, 1:10000 dilution) mouse monoclonal antibody (Abcam). HRP-conjugated goat anti-mouse (P0447, 1:10000 dilution) and swine anti-rabbit (P0399, 1:10000 dilution) were used as secondary antibodies (Dako). The membrane was developed using Amersham ECL Western Blotting Detection Reagent and ECL Select Western Blotting Detection Reagent (Cytiva) on a LI-COR Odyssey Fc Imaging System.

2.13 Co-precipitation assay

20 μL of 10 μM TNKS2 ARC1-3 was mixed with an equal volume of 1TBP-CTPR2, 3TBP-CTPR6, 3TBP-CTPR6-foldon or CTPR6 at the indicated concentration in co-precipitation buffer (50 mM Tris-HCl, 150 mM NaCl, 0.5 mM TCEP, pH 7.3). Samples were incubated at room temperature for 1 h and centrifuged at 20000 xg for 30 minutes. The supernatant was collected and the pellet resuspended in an equal volume of co-precipitation buffer. Gel samples were prepared by adding the adequate amount of 4X Loading Dye and run on a 12% or 16% polyacrylamide gel. Gels were imaged using the Li-COR Odyssey Fc Imaging System and protein band densities were analysed using the Image Studio Lite 5.2 software.

2.14 Transmission Electron Microscopy (TEM)

Purified TNKS2 ARC1-3 was mixed with an equal volume and molar concentration (5 μM) of 1TBP-CTPR2, 3TBP-CTPR6, 3TBP-CTPR6-foldon, CTPR6 or co-precipitation buffer. Samples were incubated at room temperature for 1 h. Few μL were then dispensed on carbon support film, 400 mesh, 3 mm nickel grids (EM Resolutions Ltd., Saffron Walden, UK) and stained with 2% uranyl acetate (w/v). The samples were imaged on a FEI Tecnai G2 transmission electron microscope in the Cambridge Advanced Imaging Centre (CAIC, University of Cambridge, Cambridge, UK). Images were analysed using the SIS Megaview II Image Capture system. Grid preparation and imaging was performed by Dr. Janet R. Kumita.

2.15 Fluorescence microscopy and FRAP

$1x10^5$ HEK293T cells per well were seeded overnight in 700 μL of cell growth medium in a μ-Slide chamber (Ibidi). The following day, cells were transfected with 400 ng of m-Cherry tagged 1TBP-CTPR2, 3TBP-CTPR6, 3TBP-CTPR6-foldon or CTPR6 in combination with 600 ng of eGFP-tagged TNKS2 ARC1-5 or 100 ng of TNKS2 ARC1, using 1.75 μL Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo

Fisher Scientific). Following 48 h from transfection, cells were imaged on a Leica TCS SP5 confocal microscope with a 100X oil-immersion objective lens (1.4 numerical aperture). Excitation and filters were as follows: eGFP, excitation at 488 nm, emission from 500-540 nm; mCherry, excitation at 543 nm, emission from 600-630 nm. For FRAP, individual circular regions of interest (ROI) similar in size were selected within the fluorescent agglomerates and bleached using the 488 nm laser at 100% power for 5 seconds. Fluorescence intensity changes were recorded comparing 5 pre-bleaching frames with 60 post-bleaching frames (1.3 seconds/frame). Data were analysed using Leica LAS AF suite software and data were normalised.

2.16 TopFlash dual-luciferase reporter assay

2.16.1 Wnt conditioned medium preparation

Wnt pathway activity was induced by treating cells with conditioned medium obtained from L-cells expressing Wnt3A (ATCC CRL-2647) for 8 days, according to ATCC guidelines. L-cells were kindly provided by Dr Marc de la Roche. Briefly, confluent L-cells were subcultured 1:5 in a T75 flask with 10 mL of cell growth medium. After five days, the medium was collected and sterilised using a 0.22 μm filter. Cells were supplemented with 10 mL of fresh cell growth medium and allowed to grow for another three days, at the end of which the medium was collected, sterilised and combined with the previous batch. Wnt conditioned medium was aliquoted, snap frozen and stored at -80°C for future use.

2.16.2 TopFlash assay

For the TopFlash assay, 1×10^5 HEK293T cells per well in 24-well plates (NUNC) were transfected with 100 ng of a modified pGL3 vector (TCF/LEF-firefly), 10 ng of pRL-CMV vector (CMV-*Renilla*) and 100 ng of pcDNA3.1(-) or HiBiT-pcDNA3.1(-) plasmid encoding $_n$TBP-CTPR$_{2n}$, $_n$TBP-CTPR$_{2n}$-foldon or the control CTPR constructs using 0.5 μL Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). pGL3 and pRL-CMV vectors were kindly provided by Dr Marc de la Roche. Transfected cells were allowed to recover in cell growth medium for 8 h and were then treated with Wnt-conditioned medium in a 1:1 ratio for a further 16 h. After washing the wells, the cells were imaged on an EVOS Floid Cell Imaging Station. TopFlash assay was then completed using the Dual-Luciferase® Reporter Assay System (Promega), as previously described.[157] The luminescence activities of firefly and *Renilla* luciferases were measured sequentially from the

same sample, using the CLARIOstar microplate reader (BMG Labtech). Relative luciferase values were obtained dividing the averaged firefly luminescence value by the averaged *Renilla* luminescence values for each sample. Data were obtained from three biological replicates.

When cells were transfected with HiBiT-pcDNA3.1(-) plasmids, the luminescence intensity of each HiBiT-tagged protein was also measured, by mixing 20 μL of the same cell lysate with the Nano-Glo HiBiT Lytic Detection System according to the manufacturer's instructions (Promega). Luminescence was measured at the CLARIOstar microplate reader (BMG Labtech) with emission filter from 380-540 nm, gain set at 2200 and focal height at 10.5.

2.16.3 Western Blot on TopFlash samples

TopFlash cell lysates to be analysed by Western Blot were run on a 12% or 15% polyacrylamide gels and transferred into a Immobilon-P PVDF membrane (pore size 0.45 μm, Merck Millipore) by semi-dry transfer using Tris/Glycine Transfer Buffer (National Diagnostic) on a Pierce^{TM} Power Blot Cassette (Thermo Scientific) with voltage set at 15 V and current at 2.5 A for 30 minutes. The membrane was cut into two parts to allow multiple detections on the same samples. Membranes were blocked in 5% milk (Marvel), shaking at room temperature for 1 h. HA-tagged CTPR proteins were detected using anti-HA-Tag (sc-7392, 1:1000 dilution) mouse monoclonal antibody (Santa Cruz Biotechnology). Tubulin was identified using anti-alpha Tubulin antibody (ab7291, 1:10000 dilution) mouse monoclonal antibody (Abcam). HRP-conjugated goat anti-mouse (P0447, 1:10000 dilution) was used as secondary antibody (Dako). The membrane was developed using Amersham ECL Western Blotting Detection Reagent and ECL Select Western Blotting Detection Reagent (Cytiva) on a Fuji Medical X-Ray film (FujiFilm). The film was then developed in the dark room.

2.17 Proteasomal-induced degradation assay

1.5×10^4 HEK293T cells per well were seeded in a 96-well plate and transfected the following day in duplicates with 30 ng of HiBiT-pcDNA3.1(-) vector encoding a CTPR construct using 0.15 μL of Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Following 19 h incubation, cells were treated or not with 10 μM MG132 (Calbiochem) and incubated for an additional 5 h. The number of viable cells was measured using the CellTiter-Fluor Cell Viability Assay (Promega) according to the manufacturer's instructions. HiBiT-tagged proteins were quantified using the Nano-Glo HiBiT

Lytic Detection System (Promega) as above. HiBiT values obtained from three independent replicates were normalised using the corresponding cell viability measurements and the ratio +MG132/-MG132 for each replicate was calculated and plotted with standard deviation. Each plate also included a cell titration, to convert the CellTiter-Fluor fluorescence values into cell number.

2.18 Liposomal delivery of $_n$TBP-CTPR$_{2n}$ proteins

2.18.1 Liposomal formulation of 3TBP-CTPR6

Fusogenic Liposomes (FL) containing 3TBP-CTPR6 (FL-3TBP-CTPR6) were prepared using a previously reported method.[158] In a typical procedure, 250 µg of 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), 250 µg of 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), and 25 µg of 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindotricarbocyanine iodide (DiR) (Avanti Polar lipids) were dissolved in 250 µL of chloroform (Merck). The solution was dried overnight inside a vacuum desiccator and the resulting lipid film was hydrated with 125 µL of 25 µM 3TBP-CTPR6 (in 10 mM HEPES, pH 7.4). This dispersion was first vortexed for 2 minutes and then ultrasonicated for 20 minutes at room temperature. Control FL were prepared similarly, by hydrating the lipid film with 125 µL of 10 mM HEPES, pH 7.4 with no protein. The liposomes were stored at 4°C until further use. The zeta potential and hydrodynamic size of the FL were measured at 25°C using a Zetasizer Nano ZS (Malvern Instruments). Liposomes preparation and characterisation were performed by Dr. Piyush K. Chaturbedy.

2.18.2 Cell viability upon liposomal delivery

4×10^4 HEK293T cells per well were seeded in 100 µL of cell growth medium in a 96-well plate and incubated for 24 h. Medium was replaced with 100 µL cell growth medium without FBS, containing different volumes (1-12 µL) of FL or FL-3TBP-CTPR6 and incubated for 15 minutes at 37°C. After washing the wells with 1X PBS twice, 100 µL of CellTiter-Glo Reagent (Promega) was added, and luminescence was measured using a CLARIOstar microplate reader (BMG Labtech) according to the manufacture's protocol. Untreated cells were used as control. Experiments were performed in triplicates, and the standard deviations were calculated from two independent experiments. This assay was performed by Dr. Piyush K. Chaturbedy.

2.18.3 Intracellular delivery of 3TBP-CTPR6

To visualise the liposome-mediated intracellular protein delivery, 3TBP-CTPR6 was labelled with rhodamine B isothiocyanate (RITC) (Sigma). In a typical procedure, 50 µL of RITC (1 mg/mL in DMSO) was slowly added to a 1 mL solution of 3TBP-CTPR6 (2 mg/mL) in 0.1 M sodium carbonate buffer pH 9.0, with stirring. The solution was stirred at 4°C for 8 h. Ammonium chloride (Sigma) was then added to a final concentration of 50 mM and stirring was continued for a further 2 h. RITC labelled 3TBP-CTPR6 (3TBP-CTPR6-RITC) was isolated from unreacted RITC using a PD10 desalting column (Cytiva). Liposomal formulation of 3TBP-CTPR6-RITC (FL-3TBP-CTPR6-RITC) was performed as described above. 1.4×10^5 HEK293T cells per well were seeded overnight in 700 µL of cell growth medium in a µ-Slide chamber (Ibidi). The following day, medium was replaced with 700 µL cell growth medium without FBS and containing 28 µL of FL-3TBP-CTPR6-RITC. Samples were incubated for 15 minutes at 37°C and then washed twice with 1X PBS. Confocal images were acquired using a Leica TCS SP5 confocal microscope. This assay was performed by Dr. Piyush K. Chaturbedy.

2.18.4 Cellular activity of liposome-encapsulated 3TBP-CTPR6 protein in the TopFlash assay

HEK293T cells were seeded in 24-well plates as described above. Cells were transfected with 100 ng of a modified pGL3 vector (TCF/LEF-firefly), 10 ng of pRL-CMV vector (CMV-*Renilla*) per well using 0.25 µL of Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Transfected cells were allowed to recover in cell growth medium for 8 h and were then treated with Wnt3A-conditioned medium, as above. 16 h post Wnt pathway activation, cells were treated with cell growth medium without FBS, containing 20 µL of liposomes and incubated for 15 minutes at 37°C. Following one wash in 1X PBS, Wnt3A conditioned medium was added and cells were incubated for 6 h. The TopFlash assay was performed, as described above, in triplicate in two independent experiments.

For the dose-dependence analysis, the volume of liposome added was kept constant (20 µL), while the protein concentration was varied. The FL-3TBP-CTPR6 samples were prepared in the following way: lipid cakes were hydrated with 10 mM HEPES pH 7.4 and 3.125 µM, 6.25 µM, 12.5 µM, and 25 µM 3TBP-CTPR6. 20 µL of these liposomes in 500 µL cell growth medium resulted in 3TBP-CTPR6 concentrations of 0.125 µM, 0.25 µM, 0.5 µM and 1 µM, respectively. Samples of the liposome-encapsulated CTPR6 control protein, referred to as FL-

CTPR6, were prepared by hydrating lipid cake with 25 µM CTPR6 in 10 mM HEPES pH 7.4. This assay was performed in collaboration with Dr. Piyush K. Chaturbedy.

2.18.5 TopFlash assay using small molecule TNKSi

HEK293T cells were seeded in 24-well plates as described above. Cells were transfected with 100 ng of a modified pGL3 vector (TCF/LEF-firefly), 10 ng of pRL-CMV vector (CMV-*Renilla*) per well using 0.25 µL Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). The activity of the small molecule TNKS inhibitors XAV939, IWR-1 and G007 was tested prophylactically or interventionally, and the TopFlash assay was performed with small alterations, accordingly. When tested prophylactically, inhibitors were mixed with Wnt-conditioned medium and added to the well for 16 h. When tested interventionally, instead, inhibitors were added after the overnight treatment with Wnt-conditioned medium and incubated for another 6 h in Wnt-conditioned medium. All small molecule TNKS inhibitors were used at a final concentration of 1 µM in 0.5% DMSO. The TopFlash assay was then completed as described above.

2.19 TNKS2 degradation mediated by single- and hetero-bifunctional CTPR proteins

1.5×10^4 HEK293T cells per well in 96-well plates were co-transfected with 20 ng of HiBiT-pcDNA3.1(-) vector encoding full-length HiBiT-TNKS2 and 100 ng of vector encoding a single- or hetero-bifunctional CTPR construct, with or without the HiBiT tag, using 0.25 µL of Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Following a 24 h incubation, the number of viable cells was measured using the CellTiter-Fluor Cell Viability Assay (Promega) according to the manufacturer's instructions. HiBiT-tagged proteins were quantified using the Nano-Glo HiBiT Lytic Detection System (Promega) as above. HiBiT values obtained from three independent replicates were normalised using the corresponding cell viability measurements.

Chapter 3

Different modalities of peptide grafting onto the CTPR scaffold

3.1 Introduction

The relationship between protein structure and function has been a cornerstone of biology for decades. The tertiary structure allows a protein to perform its function by mediating the interaction with other proteins or small molecules. Protein-protein interactions (PPI) affect most biological processes and are themselves tightly regulated by factors such as sequence specificity, affinity, post-translational modifications and cellular localisation. Extensive efforts have been put into understanding biological protein interactions, and in particular the structural motifs involved in homo- and heterodimeric protein-protein interfaces. A secondary structure-based analysis has identified and classified the different binding motifs within these two interface categories.[159] Importantly, not only regular secondary structure elements (α-helices and β-strands) are involved in binding interfaces, but also non-regular structures, such as turns and loops, contribute to PPI. In particular, the study demonstrates that the former are more abundant in homo-dimer interfaces, whereas the latter frequently mediate heterodimeric interactions.[159] Among all different protein's structural motifs, this chapter focuses on unstructured and α-helical binding motifs.

3.1.1 The world of short linear binding motifs

In recent years, the unstructured or intrinsically disordered regions of the eukaryotic proteome have gained increasing interest, and disordered segments have been identified in ~44% of the human protein-coding genes.[160] Firstly, protein folding studies have revealed that a broad range

of intermediate folding states exist between intrinsically disordered proteins (IDPs) and folded proteins. The versatility of protein folding entails that the majority of proteins include some structured domains as well as some intrinsically disordered regions (IDRs).[161] Secondly, it is now well understood that polypeptide chains that do not fold into a defined three-dimensional structure can also be functional.[161] This concept has disrupted the traditional structure-function paradigm, according to which protein function is critically dependent on the folded structure of the protein and unstructured segments are passive segments that link structured domains.[162] Further investigations revealed that IDRs can include protein binding sequences, promoting protein-protein interactions or higher-order protein complexes assembly, and are frequently subject to post-translational modifications (PTMs), further expanding their functional features. IDRs can also function as entropic chains, such as flexible linkers, which allow movement between domains, or spacers, that regulate the distance between them.[161]

Despite lacking a persistent secondary structure, the conformational ensembles of IDPs are extremely heterogeneous. Their amino acid sequences dictate the fold of the IDP. According to Pappu's molecular dynamics simulations, the overall conformation of an IDP depends on the fraction of positively, f_+, and negatively charged residues, f_-, and falls within one of the five conformational regions (R1-R5) (Figure 3.1a).[163] Region 1 corresponds to either weak polyampholytes[1] or weak polyelectrolytes[2] that fold into globule or tadpole-like conformations. Region 3 corresponds to strong polyampholytes that form non-globular conformations. The boundary region 2 represents a continuum of conformational possibilities between those adopted in regions 1 and 3. Regions 4 and 5 correspond to strong polyelectrolytes with one charge prevailing over the other, that are expected to fold into coil-like conformations. It is not only the frequency but also the distribution of charged residues over the IDR that have an impact on the conformation. For this reason, Das and Pappu introduced the parameter κ to describe the patterning of opposite charges in strong polyampholites.[163] When opposite charges are all evenly distributed, κ approaches zero, as interchain electrostatic repulsions and attractions are counterbalanced. When opposite charges are segregates, instead, κ is closer to one. With low κ the IDRs form swollen coils; with high κ they fold into hairpin-like conformations with long-range interactions between the two ends being formed by the electrostatic attraction of opposite charges (Figure 3.1a, b). According to

---

[1] Polyampholyte: molecule containing both positive and negative charges
[2] Polyelectrolyte: molecule with prevailing positives or negative charges

the same study, in nature most IDPs are strong polyampholytes (class R3) with low values of κ, further driving their random coil conformation.[163]
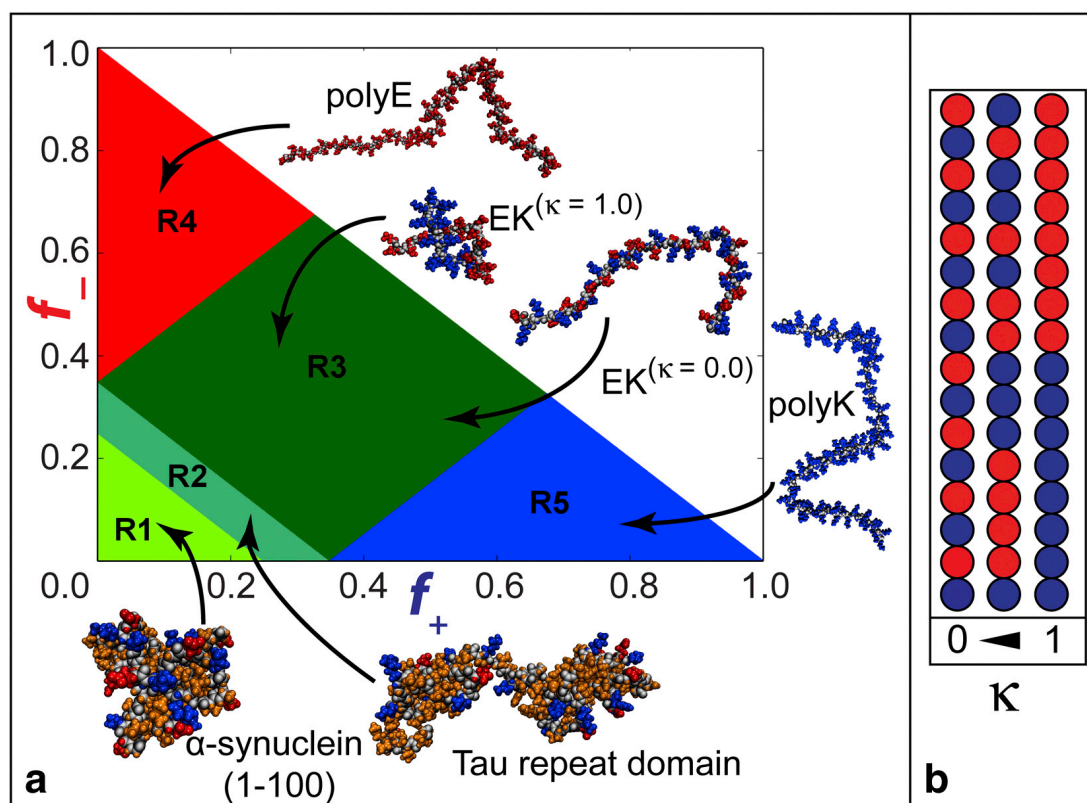
Figure 3.1: Das-Pappu conformational ensembles of IDPs. (a) Das-Pappu diagram of states with representative examples for each of the five conformational regions. (b) Schematic representation of a series of peptides having the same amount of positive (blue) and negative (red) residues, but different values of κ. Figure obtained from Holehouse *et al.*[164]

Moreover, IDRs often lack bulky hydrophobic amino acids, explaining why they are unable to fold into a hydrophobic core that characterises a structured domain. IDRs, instead, are rich in uncompensated charged groups (often negative), leading to a large net charge at neutral pH.[165]

The rising interest in IDRs is due in part to the concentration in these regions of short independently functioning binding motifs known as MoRFs (molecular recognition features) or SLiMs (short linear motifs). These motifs are short peptides of about 3-10 amino acids, and it has been estimated that there are over 100,000 such SLiMs in the human proteome.[166] According to their sequence specificity, they allow protein-protein as well as protein-nucleic acid interactions, generally with relatively low affinities (with dissociation constants in the 1-10 μM range) due to the limited number of residues involved in the binding.[167] Apart from

mediating interactions, SLiMs facilitate a variety of other functions, for example by acting as recognition sites for post-translational modifications or proteolytic cleavage, or by altering the protein's specific subcellular locations or its context-dependent activity.[167] SLiMs preferentially occur in disordered regions and are themselves unstructured but often acquire a secondary structure upon binding to their interaction partner.[168]

As PPI are disrupted in a significant number of diseases, IDRs have recently been explored in drug discovery, mostly as potential inhibitors of disease-associated targets. One approach to exploit these motifs is to chemically synthesise them while trying to mimic their bioactive conformation. For this purpose, peptides are often combined with modifications such as cross-linking and macrocyclisation in order to improve their affinity, half-life and cell penetration.[169] Alternatively, rational design can be used to graft SLiMs as solvent-exposed loops onto stable protein scaffolds, as discussed in Section 1.1.3.

3.1.2 Grafting loops onto CTPRs

Despite the fact that hetero-dimeric protein-protein interfaces containing loops are frequent in nature and loops frequently mediate biological interactions,[159] loop grafting has been challenging for the following reasons: loop conformations can be hard to predict computationally and experimentally, loop flexibility comes with an entropic penalty upon binding and an appropriate protein scaffold is required.[170]

In the past few years, the Itzhaki group has explored the possibility of functionalising the CTPR repeats by grafting non-helical SLiMs onto a CTPR loop using a simple "cut-and-paste" design. Studies from previous members of the Itzhaki group demonstrated that the CTPR scaffold can tolerate the extension of the existing 4-residue loop connecting adjacent repeats (DPNN/DPRS), to form a longer, solvent-exposed loop without compromising the correct folding of the CTPR scaffold (Figure 3.2).[80] The scaffold can accept the grafting of a single or multiple non-helical sequences up to 25 residues in length onto individual loops, with a small loss in stability.[80] More recently, even longer sequences (up to 50 residues in length) have been grafted as solvent-exposed loops in the Itzhaki lab, without disrupting the overall fold.[81] For the stability of the protein, however, a minimum of two CTPR repeats is required. An additional study performed in the Itzhaki group explored the effect of extending multiple loops within the same CTPR protein, with variable length, frequency and localisation.[146] The authors demonstrated how multiple loop extension leads to an increased dynamics of the internal repeats, compared the unmodified CTPR scaffold.[146]
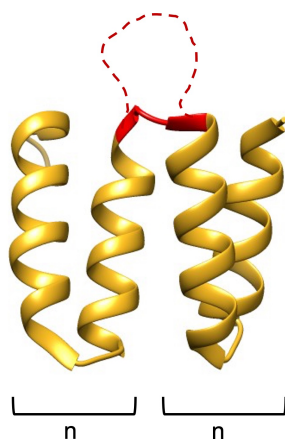
Figure 3.2: Loop grafting onto the CTPR scaffold. The residues forming the loop (DPRS) connecting adjacent repeats (indicated with n) are coloured in red. Peptide grafting within this loop forms an extended, solvent-exposed loop, schematically represented by a red dashed line, variably extended depending on the number of residues being introduced. (PDB: 2FO7) Image of the CTPR2 protein was generated using the software UCSF Chimera.

Additionally, it was shown that functionalising the CTPR scaffold by loop grafting was possible, as exemplified by introducing a short binding peptide derived from Nrf2 (Nuclear factor erythroid-2-related factor 2). This sequence, with known binding specificity for the E3 ubiquitin ligase Keap1 (Kelch-like ECH-associated protein 1), is an appropriate candidate, given its turn-like native conformation.[82] The authors demonstrated the ability of the functionalised scaffold to bind the target *in vitro*, confirming that the peptide maintains its binding properties when grafted onto the scaffold, opening the possibility for designing target-specific binders using the CTPR scaffold.[82]

Here, the same approach was adopted to design TNKS-specific inhibitors by grafting the consensus sequence recognised by the ARC subdomains of TNKS (referred to subsequently as the tankyrase-binding peptide or TBP)[46] onto the loop between consecutive repeats. Previous studies in the Itzhaki laboratory have also shown that the macrocyclic TBP peptide is able to bind TNKS with a slightly enhanced affinity than the unstapled sequence, can compete with Axin for the binding to TNKS and inhibit Wnt signalling in the cell.[85]

3.1.3 α-helices as structured binding motifs

Apart from unstructured loops, α-helices also mediate many protein-protein interactions. Helices are key building blocks in protein folding[171] and are characterised by a specific pattern of 3.6 residues per helical turn.[172] A common motif within α-helices is called the heptad repeat

and consists of seven residues folded into two helical turns.[172] Helices can vary in length and are stabilised by hydrogen bonds formed between every backbone N-H group and the backbone C=O group of the amino acid located three or four residues earlier along the helix. In nature, α-helices within tertiary structures often exhibit two "faces" and are therefore called amphipathic: hydrophobic and polar residues are segregated on opposite faces, each responsible for specific functions. While the former are oriented towards the interior of the protein and contribute to stabilising hydrophobic interactions and folding, the latter are solvent exposed. These features can be easily visualised on a helical wheel plot, a representation that illustrates the orientation of each residue along the helical axis. The same amphipathic nature is also observed in the helices of the CTPR repeats.[61]

Due to their involvement in many PPIs, α-helices are attractive therapeutic tools as inhibitors of disease-related interactions. Binding helices have therefore been extrapolated from their folded protein and used as peptide inhibitors. However, they come with major limitations: peptides are generally unstable in the cell and subject to proteolysis; they are not cell permeable and they mostly remain unfolded without the remaining parts of the protein and hence have low binding affinities. To overcome these issues, helices have been stabilised by helix stapling or helix grafting, both of which provide a solution to the problems mentioned above.

For helix stapling, or cyclization, peptides are chemically synthesised incorporating two unnatural amino acids at a distance of one (*i, i*+4) or two (*i, i*+7) or three helical turns (*i, i*+11), so that their side chain are on the same side of the helix. Moreover, unnatural residues are placed on the opposite face than binding residues, to avoid disrupting the binding interface. Helix stapling is then performed by covalently linking the two unnatural amino acid side chains according to various macrocyclisation chemistries. Upon stapling, the population of unfolded conformations decreases, constraining the peptide into a helical fold. The entropic penalty for binding is therefore decreased.[173] Apart from inducing the binding-competent helical conformation, stapling has also been shown to increase the stability and the plasma half-life of the peptide, as well as its cell penetration ability, leading to improved pharmacokinetic as well as pharmacodynamic properties.[174]

Helix grafting, instead, is performed by resurfacing a helical domain within a structured, stable protein scaffold with the key contact residues involved in PPI of a different helix. As all helices share the same heptad conformation, helix grafting is performed by replacing some residues of the "host" helix with the corresponding amino acid of the "guest" binding helix, located in the same heptad position.[83,84,175–177] Initially, the two helices are

compared to identify the residues in the same heptad position. This allows residues to still be displayed in the same conformation after the transfer of a residue from one helix to the other. Secondly, residues interacting with the target need to be identified and located within the "guest" helix, whereas residues necessary for the stability of the protein scaffold need to the identified on the other. The transfer of the first set of residues into the second while maintaining the same heptad position generates the grafted helix. In this way, a helical recognition epitope is transferred from its native α-helical context onto the helix of a stable protein scaffold (Figure 3.3).[176]

3.1.4 Grafting α-helices onto CTPRs

Given the high α-helical content of the CTPR protein, the Itzhaki lab has also explored the possibility of grafting helical sequences onto this scaffold. The group has attempted grafting of binding helices at both termini of the CTPR scaffold (unpublished data). These studies have demonstrated how the CTPRs can tolerate the grafting of α-helical motifs at either end without affecting the overall stability and folding of the scaffold. The procedure adopted for helix grafting is detailed in the Results section of this Chapter.

Here, helix grafting was exploited to introduce the coiled-coil foldon (CC-foldon) motif at the C-terminal end of each of the TNKS-binding CTPR constructs (Figure 3.3). Rather than a new binding functionality, the grafted foldon domain induces the CTPR protein to form a trimer, thereby increasing the valency.
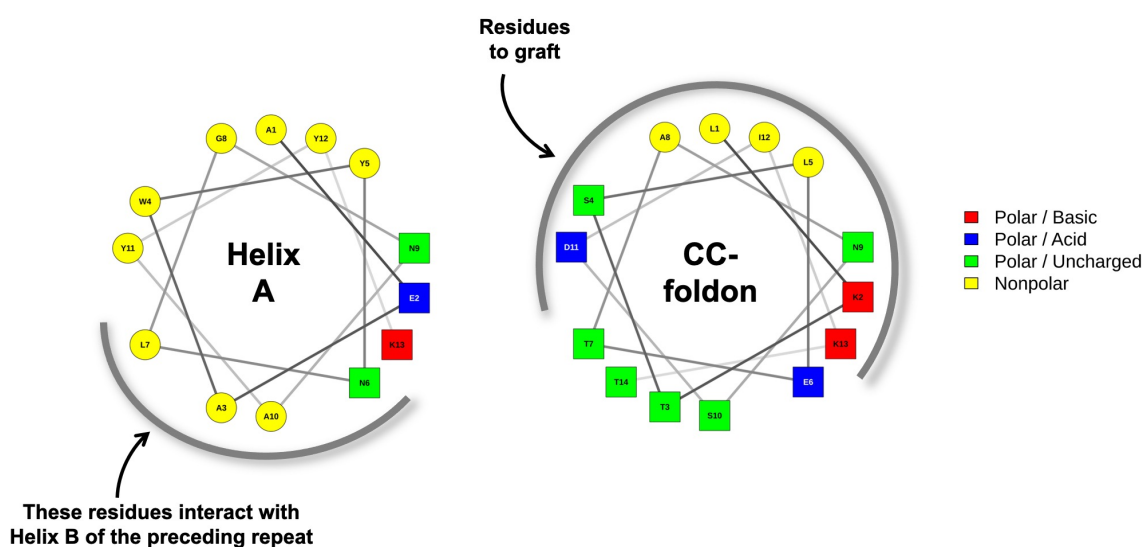
Figure 3.3: Helical wheel representation of a CTPR Helix A (Left) and the coiled-coil (CC) helix of the foldon motif (Right). For helix A, the residues involved in interactions with the preceding repeat are highlighted and need to be maintained. All the other residues are replaced with those of the CC-

foldon helix, located at the corresponding heptad position. Polar residues are shown as boxes, nonpolar residues as circles. Figures were obtained from the web-based application NetWheels.[62]

3.1.5 The foldon domain

The foldon is a trimerisation motif and corresponds to the C-terminal residues of fibritin, a viral protein encoded by the bacteriophage T4.[178] Fibritin is the main component of the filaments protruding from the "neck" of the virion and it plays essential roles in virus assembly and as environment-sensing device. Its structure consists of a N-terminal domain responsible for attachment to the phage particle; a long, coiled-coil central region divided into 13 helical segments, and a C-terminal domain containing the foldon motif. The foldon motif is essential for the correct trimerisation of fibritin, leading to coiled-coil formation.[179] The three-dimensional structure of the foldon domain was firstly solved by Tao *et al.* in 1997, when the final, C-terminal 119 residues of fibritin (residues 368-487) were crystallised in a trimer configuration (Figure 3.4C).[178] The structure reveals the conformation of the last three segments of the coiled-coil region (residues 368-457), continuing into the 11-amino acid foldon motif (residues 458-487) (Figure 3.4B). The foldon domain folds into a β-hairpin motif and an extended loop connects the β-hairpin to the end of the last α-helical coiled-coil segment (Figure 3.4A). The foldon is rich in hydrophobic residues, forming stabilising hydrogen bonds within the same chain (Figure 3.4A). Foldon trimerisation is mediated by two additional hydrogen bonds formed between each pair of neighboring β-hairpins.

The structure of the foldon domain was also confirmed by Boudko *et al.* in 2004.[180] The authors obtained the structure of a designed fibritin fragment referred to as 'NCCF' containing the N-terminal region, the first coiled-coil segment and the foldon motif. [180] In the same year, a 27-amino acid peptide corresponding to the foldon motif was proven to fold into a trimer independently of the structural context of fibritin, making it the ideal candidate for protein engineering efforts.[181] The authors also concluded that the foldon is evolutionarily optimised for rapid folding and trimeric fibritin assembly.[181] Given its intrinsic stability and fast folding, the foldon domain has already been used as self-assembling building blocks for nanomaterials or as scaffold of trimeric peptide and protein structures.[182–187]
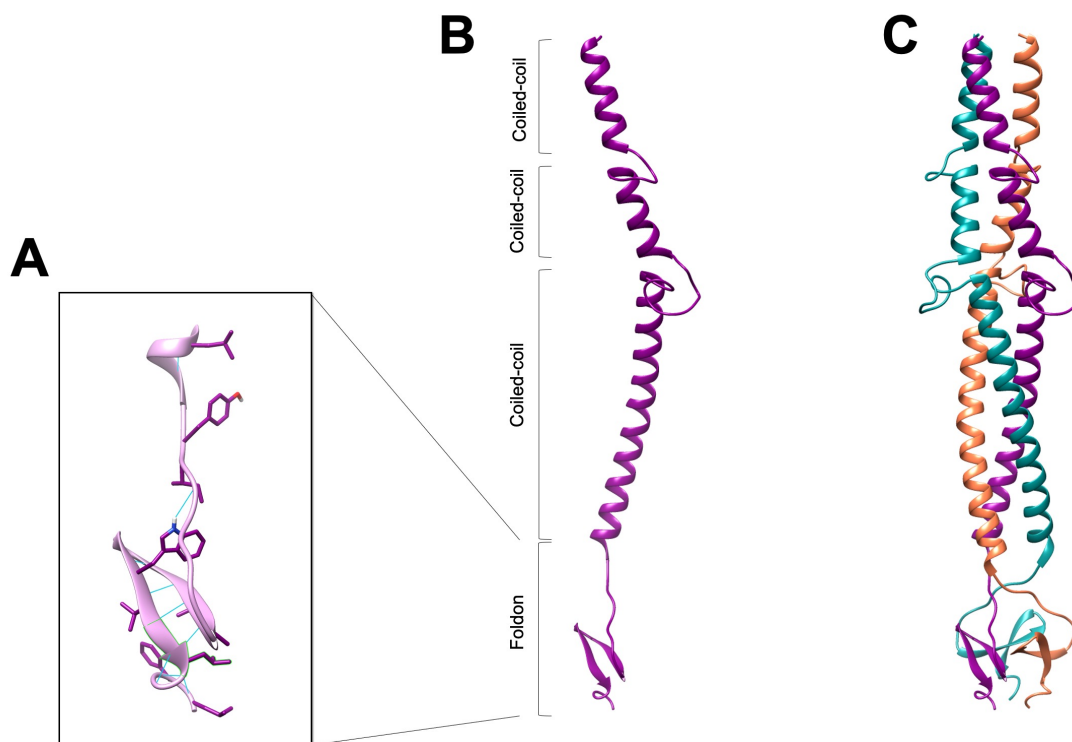
Figure 3.4: Crystal structure of fibritin. (A) Ribbon representation of the C-terminal foldon domain, with the sidechains of the hydrophobic residues involved in stabilising intra-molecule and inter-molecules hydrogen bonds shown and coloured by heteroatom. Hydrogen bonds are represented with light blue lines (PDB 1AA0). (B and C) Ribbon representation of the C-terminal end of fibritin, consisting of three coiled-coil segments followed by the foldon domain. The monomer (B) and the trimer (C) of fibritin are shown (PDB 1AA0). Images were generated using the software UCSF Chimera.

3.2   Results

3.2.1   Protein design and grafted peptide sequence

The CTPR sequence was developed by Main and co-workers in 2003 using a statistical approach.[61] In 2005, Kajander *et al.* mutated the CTPR sequence to facilitate the cloning method, generating a new CTPR variant named CTPRa. The CTPRa sequence was obtained by replacing the two final residues of the DP**NN** loop (two asparagine amino acids at positions 33 and 34 in the TPR repeat, connecting helix B and helix A of the following repeat) with a DP**RS** loop.[149] This difference arose from the need for a simple cloning strategy that could allow the concatemerisation of tandem repeats at the DNA level. The final residues RS, indeed, correspond to the BglII restriction site where the ligation between consecutive CTPR units occurs, as explained in the method Section 2.3.3.[149]

Based on previous studies and in-house results, the Itzhaki group has adopted the alterations listed below, generating the CTPR sequences described in this study.

1. For all the constructs made, the N-terminal capping sequence was not included.

2. Except for one construct, the CTPR proteins in this study do not have the extra C-terminal solvating helix that has been used by some groups to improve solubility.[41,61,188,189] The Solvating helix was introduced only in one construct, named 1TBP-CTPR4-Solvating Helix, that was designed for crystallographic studies. In all the other constructs, the CTPR scaffold consists of the 34 amino acid residues denoted with n (Figure 3.5).

3. The original residues at position 18 and 19, aspartic acid (D18) and glutamic acid (E19), were mutated to glutamine (Q) and lysine (K), respectively,[67,190] to further enhance the intrinsic stability of the CTPR scaffold by introducing novel charge-charge interactions. Q at position 18 corresponds to the amino acid with the third highest global propensity value, while K is the second most frequent residue at position 19.[61]

4. To facilitate the concatemerisation of tandem 1TBP-CTPR2 units, the two final resides within the DPNN loop were replaced by DPRS residues, containing the BglII recognition site at the DNA level.[149] Due to this cloning strategy, two CTPR sequences were therefore generated in the $_n$TBP-CTPR$_{2n}$ array, one preceding (1) and one following (2) the grafted loop (Figure 3.5). The DPNN-to-DPRS mutation, however, causes a stability loss, as shown in previous studies.[80,191]
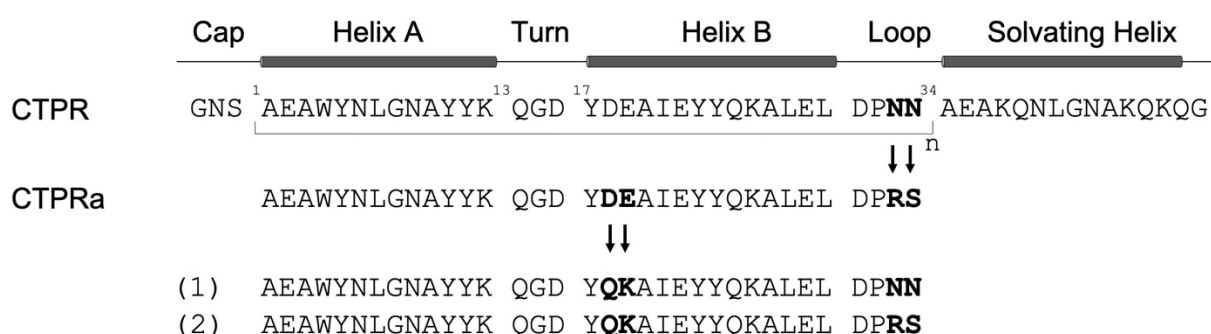
Figure 3.5: CTPR sequence development. Sequences of the CTPR and CTPRa motifs designed by the Regan laboratory are shown.[61,149] (1) and (2) denote the CTPR sequences used for this study, with the first preceding the loop, and the second following it. Single point mutations are indicated with individual arrows. Secondary structural elements are shown at the top of the figure. Helices are represented as rods.

The CTPR scaffold used in this study, therefore, consists of alternating sequence (1) and sequence (2), provided in Figure 3.5. This scaffold was functionalised by grafting an 8-residue TNKS-binding peptide (TBP), REAGDGEE, onto the loop between two adjacent repeats, to form an extended, solvent-exposed loop. The TBP was identified by Guettler *et al.* from a mutational analysis of the TNKS binding sequences of 3BP2, RSPPDGQS (K_d 4.9 ± 0.4 µM), and is known to bind more tightly to TNKS2 ARC4 (K_d 0.6 ± 0.04 µM).[46] According to the crystal structure of ARC4 in complex with the 3BP2 binding peptide, the peptide binds ARC4 in an extended conformation. The TBP peptide is also expected to be unstructured: a prerequisite for successfully grafting it into a solvent-exposed loop.

The TBP was inserted immediately after the DPNN loop of the first CTPR repeat. A DPRS sequence was also introduced at the C-terminus of the TBP sequence, to terminate the extended loop and induce the correct folding of the following repeat. The flanking DPXX sequences also place the TBP sequence in the middle of the loop, at an equal distance between the two adjacent repeats. Moreover, the presence of a proline residue in each flanking DPXX sequence ensures a conformational turn, maintaining helices A and B correctly folded and the SLiM unstructured. Therefore, the final loop sequence is N'-DPNNREAGDGEEDPRS- C'.

An analysis using the CIDER software was performed to predict the physical properties of the DPNN-TBP-DPRS sequence.[164] CIDER calculates various sequence parameters, mostly related to the number and distribution of charged amino acids, across a disordered protein sequence, and plots a Das-Pappu phase diagram with the fraction of positively (x-axis) and negatively charged residues (y-axis). The sequence lies in the strong polyampholyte region, as it contains both positive and negative charges in similar proportions, and strong polyampholytes are expected to fold into nonglobular conformations that are coil-like, hairpin-like or chimeras (Figure 3.6). The CIDER software also revealed that the κ value for the DPNN-TBP-DPRS loop is equal to 0.34, in accordance with the finding that naturally occurring strong polyampholytes have low values of κ.[192]
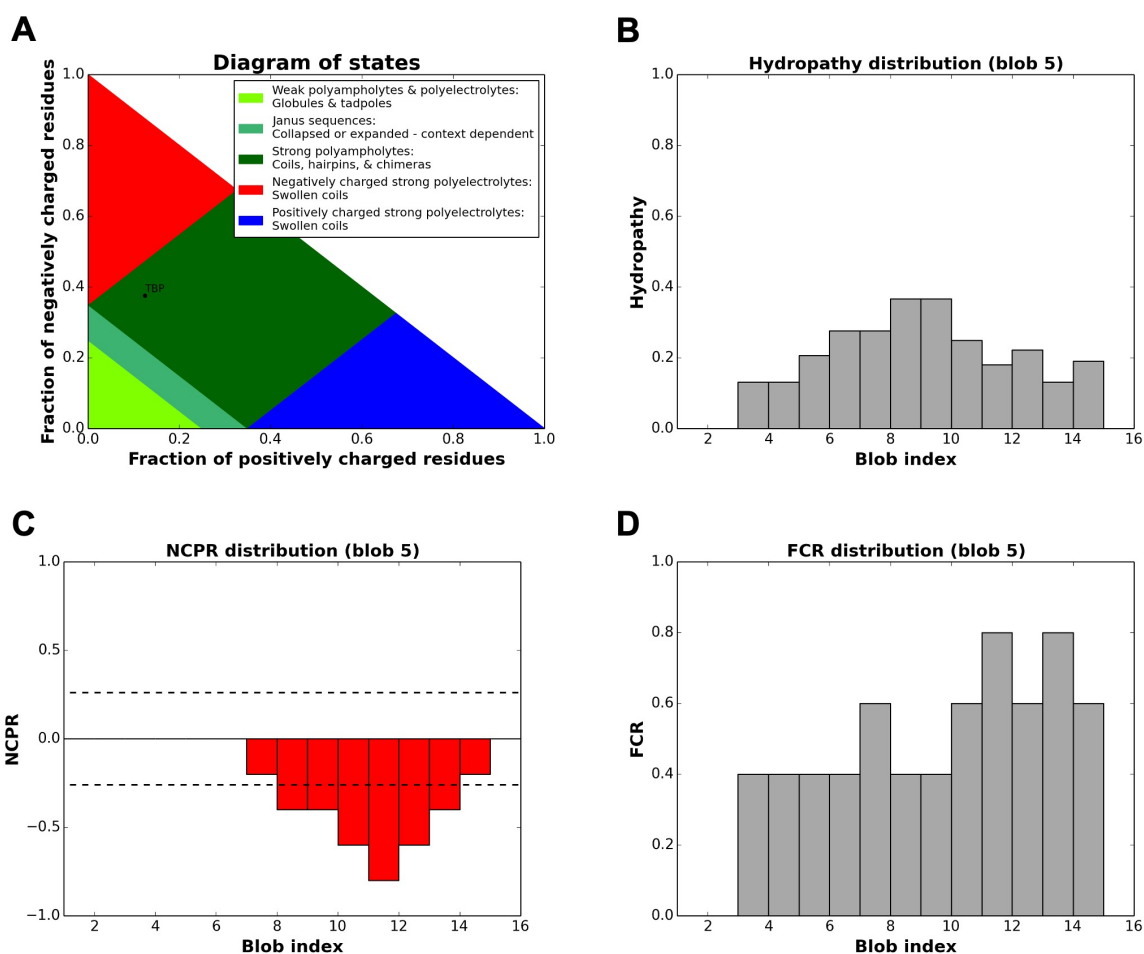
Figure 3.6: CIDER calculations on the DPNN-TBP-DPRS sequence (16 residues). (A) Das-Pappu phase diagram of the TBP loop in the context of the CTPR scaffold. (B) Linear Hydropathy diagram using a sliding window (blob) of five residues, with values scaled to lie between 0 (hydrophilic) and 1 (hydropathy). It indicates that the A and G residues in the central region are the most hydrophobic ones. (C) NCPR (net charge per residue) diagram using a sliding window (blob) of five residues. Values are calculated as $f_+ - f_-$, where f_+ and f_- denote the fraction of positive and negative charges within the five-residue window. (D) Linear fraction of charged residues diagram using a sliding window (blob) of five-residues. Values are calculated as $f_+ + f_-$, with f measured within the five-residue window. C and D indicate that charged residues are more frequent on the second half of the DPNN-TBP-DPRS sequence. Diagrams obtained using the web-software CIDER.[192]

The construct generated by loop grafting was named 1TBP-CTPR2, as it consists of one TBP loop between two CTPR motifs (Figure 3.7). By grafting the TBP into an inter-repeat loop, the TBP is expected to be solvent-exposed and, therefore, able to bind the target.
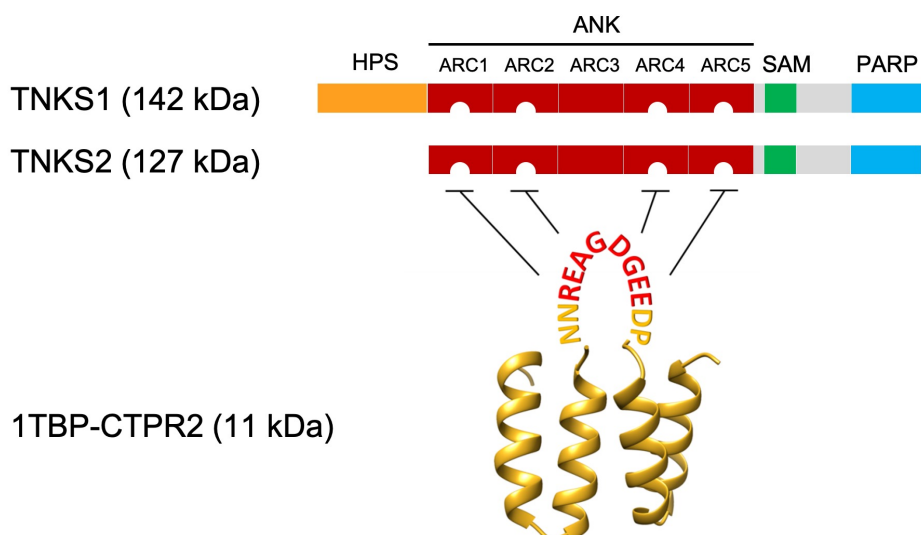
Figure 3.7: Top: Domain architecture of TNKS1 and TNKS2, comprising a histidine, proline, serine-rich (HPS) domain, a substrate-binding ankyrin-repeat (ANK) domain consisting of five ankyrin-repeat clusters (ARC), a sterile alpha motif (SAM) and the catalytic PARP domain. Semi-circles indicate TBP-binding sites within ARC1, ARC2, ARC4 and ARC5. Bottom: Schematic representation of the 1TBP-CTPR2 construct, showing the TNKS-binding peptide (TBP, coloured in red, sequence REAGDGEE) grafted onto the loop between adjacent CTPR repeats (PDB 2HYZ).

3.2.2 $_n$TBP-CTPR$_{2n}$: a multivalent linear array

Taking advantage of the repetitive nature of the CTPR scaffold, as well as the presence of multiple ARC subdomains within TNKS, a series of multivalent constructs were generated to study the effect of multivalency on the binding to TNKS. To this aim, the minimal TNKS-binding unit was tandemly repeated and joined together at the DNA level by concatemerisation, to generate a series of multivalent, linear CTPR molecules with increasing valency and molecular weight. The constructs were named $_n$TBP-CTPR$_{2n}$, with n between one and four, according to the number of TBP loops and CTPR repeats present in each. Using this method, CTPR constructs with identical sequence repeats at both the DNA and protein level were generated. As a control, a CTPR6 construct with no loops was used. A schematic representation of the linear array is shown in Figure 3.8.

This approach resulted in the formation of multivalent constructs having TBP loops located in alternating inter-repeat interfaces. As loop grafting causes a stability loss at the interface in which it is located,[80] this design allows to maintain a stabilising interface in between two TBP loops.

According to the crystal structure of CTPR8 and CTPR20 proteins, long arrays of CTPR repeats adopt a rigid, superhelical conformation, with eight repeats required to complete a superhelical turn.[64] Thus, in the $_n$TBP-CTPR$_{2n}$ multivalent proteins, the TBP loops are expected to point towards different directions and be offset from each other by approximately 90° (Figure 3.9), as loop grafting did not affect the overall folding of the proteins previously tested.[80,82,146] However, it remains difficult to judge the effect of spacing and orientation of the TBP loops within these multivalent TBP-CTPR constructs on the binding with TNKS.

Moreover, due to the absence of the crystal structure of the full-length ANK domain of TNKS and its dynamic nature,[89] a structure-based design approach could not be applied to drive the development of multivalent TBP-CTPR binders, further explaining the rationale behind the construct design adopted here (further details provided in Section 6.1.1).
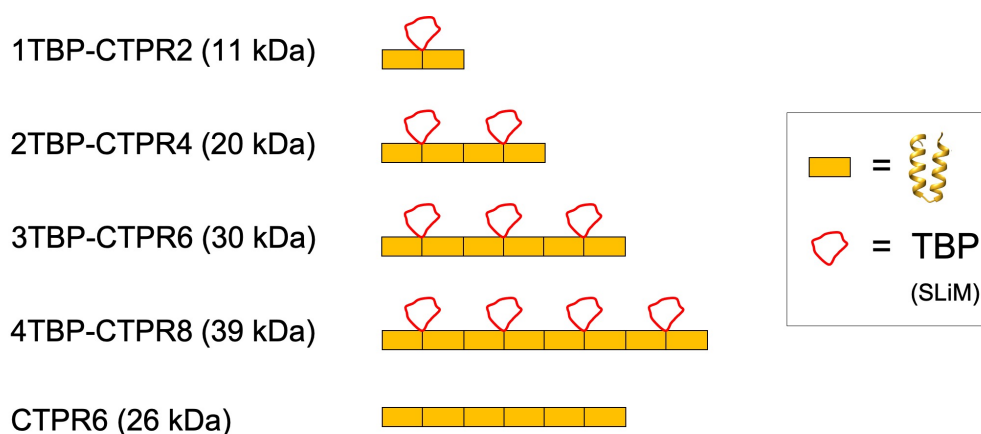
Figure 3.8: Schematic representation of the $_n$TBP-CTPR$_{2n}$ and control constructs of the linear array. Each CTPR is shown as a yellow rectangle, and the TBP grafted onto the inter-repeat loops are in red. The amino acid sequence of the generated constructs is provided in Appendix B.
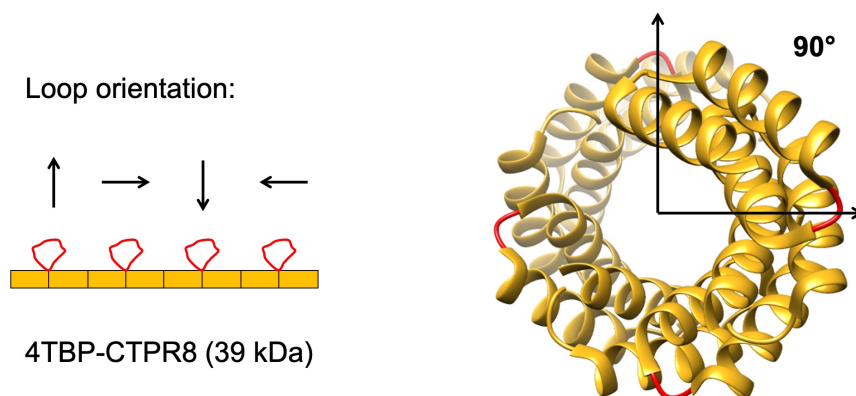
Figure 3.9: The superhelix conformation affects the orientation of the TBP loops in the linear array. Left: schematic representation of 4TBP-CTPR8 (chosen as representative), with arrows indicating the loop orientation within the construct. Right: crystal structure of CTPR8 (PDB 2HYZ) viewed along the superhelical axis. CTPR helices are in yellow and the inter-repeats loops, onto which the TBP sequence is grafted, are in red. The image was generated using the software UCSF Chimera.

3.2.3 Linear protein array expression and purification

All His$_6$ tagged $_n$TBP-CTPR$_{2n}$ were recombinantly expressed in the C41 *E. coli* strain as described in Section 2.4. A small aliquot of the overnight culture was used to perform a solubilisation test using BugBuster® Master Mix, which confirmed that all proteins were expressed as soluble proteins in solution. All $_n$TBP-CTPR$_{2n}$ constructs could be purified in high yields (~20 mg per litre of culture). All proteins were purified by affinity chromatography first, followed by size exclusion chromatography. Affinity chromatography relies on the presence of an affinity tag, such as the His$_6$, whereas size-exclusion chromatography separates molecules depending on their size and 3D conformation. $_n$TBP-CTPR$_{2n}$ could elute from the HisTrap affinity chromatography column as highly pure proteins and at high concentration. Higher order constructs, in particular 4TBP-CTPR8, showed the presence of lower and higher molecular weight contaminants, probably corresponding to recombination products generated at the DNA or the protein level. To further improve their purity, proteins were run on a Superdex 75 16/600 size-exclusion chromatography column (Figure 3.10A, left). As expected, the retention volume of the $_n$TBP-CTPR$_{2n}$ proteins lowers with their increasing molecular weight. All chromatograms share a similar shape, characterised by an asymmetric peak with a long trailing tail that follows the maximum. This tail is recognizable on the size-exclusion chromatograms of all the CTPR proteins tested in the Itzhaki lab (with variable length, number of loops and loop sequences) and we believe is most likely due to their non-specific binding of the CTPR repeats to the column matrix. It is also worth noting the presence of a second peak

in the chromatogram of most proteins, that elutes after one column volume (120 mL). It was however observed that the 1:5 dilution of the same 2TBP-CTPR4 protein sample run under identical experimental conditions abolished the second elution peak (Figure 3.10A, right). Therefore, I speculate that the combination of high protein concentration and non-specific binding to the column matrix might slow down a fraction of the protein sample, resulting in the observed second peak. When comparing the molecular weights of the purified samples with their size-exclusion chromatography traces, one can see that the CTPR6 protein elutes earlier than expected and with a broad peak, probably indicating the existence of fast-exchanging oligomerisation species (further evidence are provided in Chapter 4). SDS-PAGE electrophoresis confirmed the high purity of all purified samples (Figure 3.10B). Only in the 4TBP-CTPR8 sample, some contaminants were still present even after size-exclusion chromatography. Finally, mass spectrometry confirmed the correct identity of the purified proteins (Appendix C).
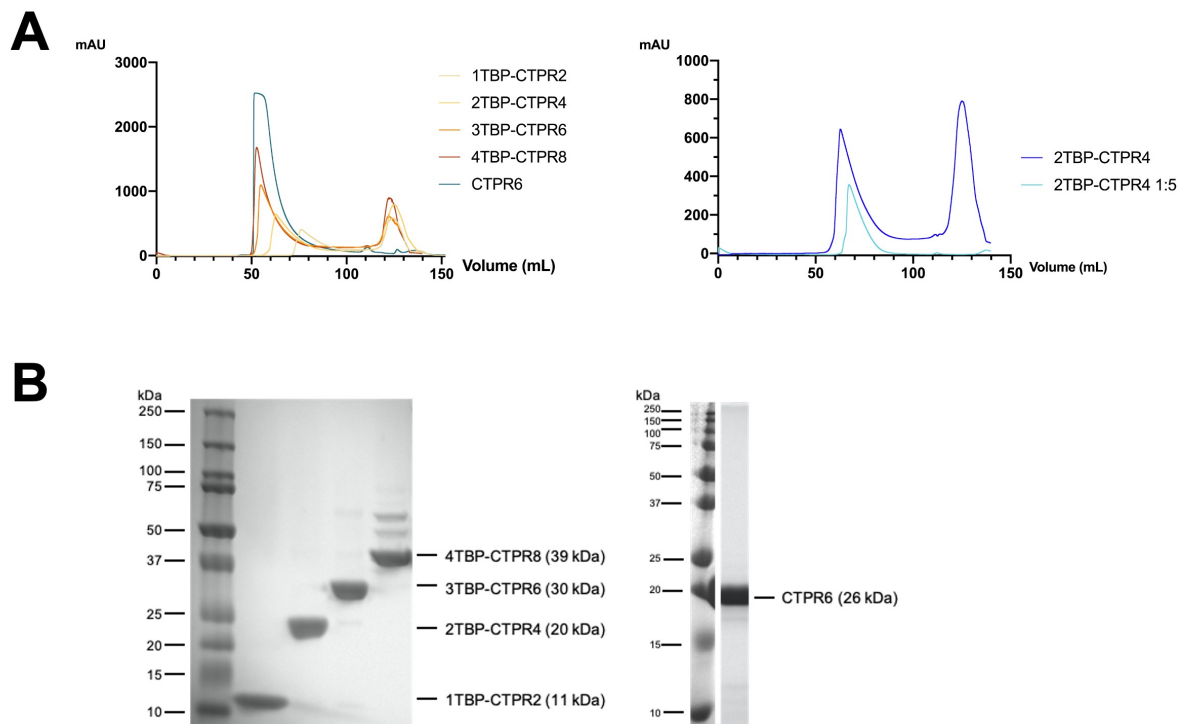
Figure 3.10: protein purification. (A) Superdex 75 16/600 size-exclusion chromatography traces of the indicated proteins. On the right, the traces obtained with highly concentrated and 1:5 diluted 2TBP-CTPR4 sample are compared. (B) SDS-PAGE gels with lanes corresponding to the purified sample of the indicated proteins.

3.2.4 $_n$TBP-CTPR$_{2n}$-foldon: a multivalent trimeric array

An additional, trimeric array of $_n$TBP-CTPR$_{2n}$ constructs was generated alongside the linear one. Trimerisation of the $_n$TBP-CTPR$_{2n}$ proteins was achieved by grafting the 'foldon' motif at one end of each linear construct (now referred to as $_n$TBP-CTPR$_{2n}$-foldon), generating their homo-trimeric counterparts having three, six, nine or twelve TBP loops. The rationale for making the trimeric $_n$TBP-CTPR$_{2n}$-foldon proteins was to further extend their multivalent capabilities and to produce additional and more complex display geometries beyond what is possible with the linearly arrayed $_n$TBP-CTPR$_{2n}$ format.

The foldon motif was grafted at the C-terminus of each $_n$TBP-CTPR$_{2n}$ construct as described in Sections 3.1.3 and 3.1.4. More precisely, helical grafting onto the C-terminal end of the CTPR scaffold requires the introduction of an additional helix A following helix B of the last repeat. This helix A corresponds to the "host" helix on which we grafted a portion of the coiled-coil that precedes the foldon motif, continuing into the foldon motif itself. When grafting a helical region onto the final CTPR helix, the residues at the interface between consecutive CTPR repeats need to be preserved in order to maintain the stability of the CTPR scaffold. The residues were identified in alanine 3 (A3), asparagine (N6), leucine 7 (L7), and alanine 10 (A10). These residues are located on the same face of the helix, as represented in Figure 3.3, and their side chains are involved in the CTPR inter-repeat interactions, therefore they need to be conserved. Residue A1 was also maintained to provide stability on the helical conformation at a position immediately after the loop connecting adjacent repeats. Residues asparagine 9 (N9) and lysine 13 (K13) of helix A were also maintained, as they match the CC residues at the same heptad position. The remaining residues of the terminal helix A were substituted with those of the fibritin NCCF coiled-coil construct[180] (residues 65 to 77) at the corresponding heptad position. These were followed by the remaining helical residues (NCCF residues 78 to 82) and the foldon motif (NCCF residues 83 to 109) (Figure 3.11A). The resulting constructs are referred to as $_n$TBP-CTPR$_{2n}$-foldon, and their protein sequence is given in Appendix B. A model of the trimeric CTPR2-foldon protein is shown in Figure 3.11B, with the loops onto which the TBP is grafted highlighted in red. Apart from increasing the number of TBP loops, this trimeric configuration is expected to display the loops in a different geometry, being offset from each other by 120°, rather than 90° as for the monomeric constructs.

3.2.5 Trimeric protein array expression and purification

The $_n$TBP-CTPR$_{2n}$-foldon constructs were expressed as soluble proteins and were purified using Ni-NTA affinity resin. In this batch purification method, beads agglomeration was observed, possibly caused by the fact that for each trimer there are three His$_6$ tags, therefore causing a cross-linking effect between beads. As this effect has never been observed with any other His$_6$ tagged-CTPR protein in the Itzhaki lab, this evidence represents a first indication that the grafted foldon causes trimerisation of the CTPR constructs in solution. Cross-linking did not affect the efficiency of elution and purification. Trimeric $_n$TBP-CTPR$_{2n}$-foldon proteins could be further purified by size-exclusion chromatography.
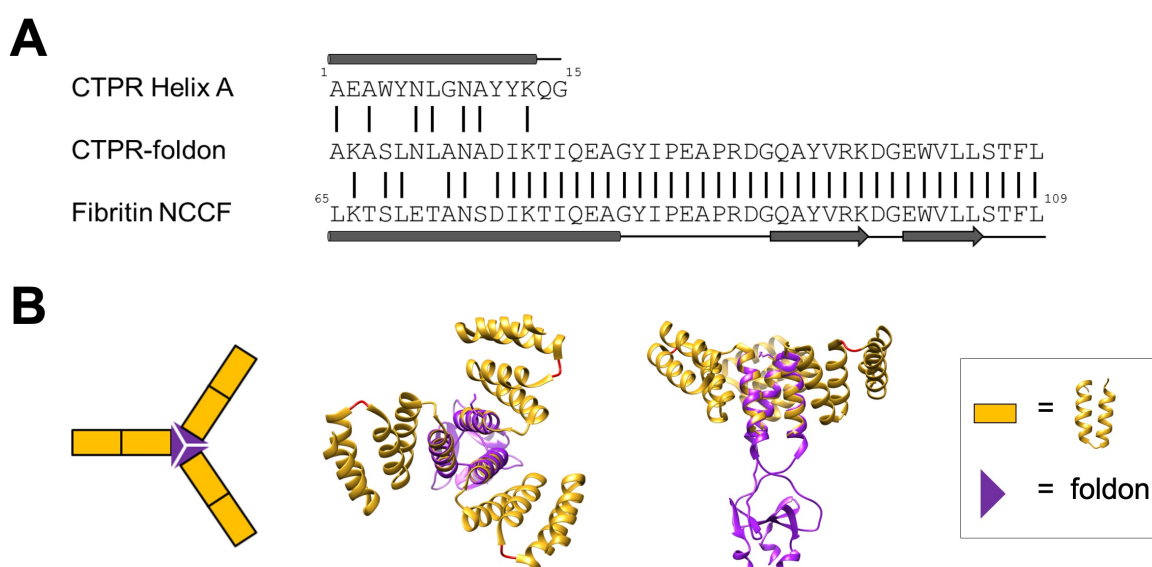
Figure 3.11: Foldon motif. (A) Sequence alignment between CTPR helix A (residues 1 to 15), the grafted CTPR-foldon construct, and fibritin NCCF (residues 65 to 109). Rods represent α helices, arrows represent β strands, thin lines indicate unstructured regions. (B) Left: schematic representation of CTPR2-foldon. Each CTPR is shown as a yellow rectangle and the foldon domain as a purple triangle, as indicated in the legend. Middle and Right: Different views of the modelled structure of CTPR2-foldon. The model was generated by grafting the foldon helix (PDB: 1OX3) onto the CTPR helix (PDB: 2HYZ) using the software UCSF Chimera. CTPR helices are shown in yellow, the inter-repeats loops, onto which the TBP is grafted, are in red.

Trimerisation of the $_n$TBP-CTPR$_{2n}$-foldon constructs was first verified by native gel electrophoresis. Trimeric $_n$TBP-CTPR$_{2n}$-foldon proteins were run alongside their linear counterparts for comparison (Figure 3.12). The series lacking the foldon domain (1TBP-CTPR2, 2TBP-CTPR4, 4TBP-CTPR8, calculated molecular weights 11.3 kDa, 20.6 kDa and

39.4 kDa, respectively) run at increasing molecular weights between the 20 kDa and 66 kDa molecular weight markers. In contrast, those including the foldon domain (1TBP-CTPR2-foldon, 2TBP-CTPR4-foldon, 4TBP-CTPR8-foldon, calculated molecular weights as trimers 49.2 kDa, 77.4 kDa and 133.5 kDa, respectively) run at higher molecular weights between the 66 kDa and 146 kDa markers. Native gel electrophoresis, therefore, indicates that all the $_n$TBP-CTPR$_{2n}$-foldon proteins tested run as trimers.
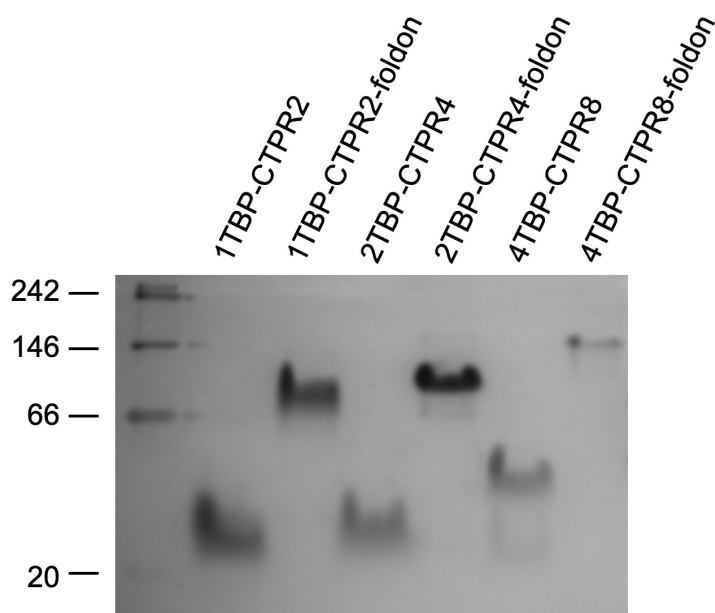
Figure 3.12: Native gel electrophoresis analysis. Each lane corresponds to the indicated $_n$TBP-CTPR$_{2n}$ or $_n$TBP-CTPR$_{2n}$-foldon protein. A polyacrylamide gel not containing SDS and NativeMark^{TM} Protein Std (Invitrogen) were used. The gel was cast and run by a previous member of the Itzhaki group, Dr Albert Perez-Riba.

3.3   Discussion

In this chapter, the potential of the CTPR as a protein scaffold, as well as the different grafting methods that can be applied to it, has been explored and exemplified.

One way of functionalising the CTPR scaffold is by grafting a target-binding SLiM onto the loop connecting consecutive repeats, forming an extended, solvent-exposed loop. This method can be potentially applied to a great variety of non-helical SLiMs. Here, loop grafting was applied to introduce the TNKS binding sequence, named TBP, onto the scaffold. The specific reasons for choosing this target are: first, TNKS is an important therapeutic target

involved in several diseases, against which an effective and specific inhibitor is still lacking, and second, the interaction between TNKS and the TBP peptide has been well studied by the Guettler laboratory, and further investigation has been performed by our group using chemically constrained TNKS-binding peptides.[85]

Another way of functionalising the CTPR scaffold is by helical grafting at either termini of the CTPR scaffold, by adding an additional "host" helix (helix A if added at the C-terminal end or helix B if added at the N-terminal end) onto which to graft the residues of a "guest" helix. The method has been performed here to introduce the trimerisation foldon and generate trimers of the $_n$TBP-CTPR$_{2n}$ constructs. Helical grafting on the CTPR scaffold is particularly convenient, as helices are stacked in a linear array and only few residues have to be conserved within the "host" helix, leaving many heptad positions available for grafting. Moreover, there is no residue clashing on any of the free heptad positions of the "host" helix, therefore any amino acid can be accommodated.

The CTPR unit has many other favourable properties. It is very stable, it can be expressed in *E. coli* as soluble protein, in high yields and can be easily engineered for different purposes. The repeating nature of CTPRs also allowed us to create a multivalent system by simply combining consecutive TNKS-binding units. Unlike other artificial binding proteins such as bi-specific antibodies and DARPins[9,53,193] where multivalency is generated by connecting multiple repeat units like "beads-on-a-string" through flexible linkers, multivalency in CTPRs is induced by simply introducing more CTPR binding units in the same DNA open reading frame, therefore producing an extended, higher-molecular weight protein. Moreover, given the superhelical conformation of CTPR repeats, the geometric arrangement with which the multiple functions are displayed in the CTPR scaffold is intrinsic to its repeating architecture and can be pre-programmed in a precise and predictable way. By grafting the TBP loop in alternating inter-repeat interfaces, the extended loops are expected to be displayed offset from each other by 90°, therefore avoiding steric hindrance between them.

Multivalent $_n$TBP-CTPR$_{2n}$ constructs were obtained by concatemerisation of the 1TBP-CTPR2 unit at the DNA level. This corresponds to the simplest and most efficient way to generate higher order tandem-repeat constructs, mostly because it is not possible to order DNA fragments with a highly repetitive sequence. By concatemerisation, therefore, repeats having identical DNA and protein sequences are therefore repeated in tandem as desired. Having a long $_n$TBP-CTPR$_{2n}$ construct encoded by the same repeating DNA sequence can, however, be a downside. As observed in Figure 3.10B, 4TBP-CTPR8 was purified together with lower and higher molecular weight contaminants, probably corresponding to recombination products.

This observation led us to the conclusion that any further CTPR construct designed subsequently should be encoded by a maximally diversified DNA sequence, with all repeats being different, where possible (as in Chapter 5). By grafting the foldon domain, the number of TBP loops was further increased by trimer formation rather than extending the linear $_n$TBP-CTPR$_{2n}$ constructs with additional TNKS-binding units. In this way, not only further concatemerisation is avoided, but also the downsides connected to it are circumvented.

Loop and helical grafting can be applied on the same CTPR construct with endless combinations. When different binding sequences are grafted on the same protein, hetero-bifunctional constructs are generated. Hetero-bifunctional proteins can be exploited for several applications, further expanding the potentiality of the CTPR scaffold. An example of hetero-bifunctional $_n$TBP-CTPR$_{2n}$ constructs is provided in Chapter 9.

For simplicity, from now onwards the CTPR constructs generated in this study will be named $_n$TBP-CTPR$_{2n}$ when referring to the linear array, $_n$TBP-CTPR$_{2n}$-foldon when referring to the trimeric array, CTPR$_n$ when referring to the control constructs with no binding loops or TBP-CTPR constructs when referring to all TNKS-binding CTPR proteins, indiscriminately.

Chapter 4

Biophysical characterisation of the CTPR constructs

4.1 Introduction

Early characterisation and development of protein biotherapeutics rely on the analysis of their biophysical properties, such as solubility, folding, target-binding specificity and monodispersity in solution to prevent the formation of protein aggregates. Continuous efforts in the biopharmaceutical industry are indeed spent to improve the protein's stability, as it has an effect on its efficacy, manufacturing, formulation and potentially safety too. The methods and techniques adopted in this chapter aim to evaluate the biophysical properties of the $_n$TBP-CTPR$_{2n}$ proteins of the linear array.

4.1.1 Biophysical properties for the development of biotherapeutics

Arrays of CTPR modules generate highly stable proteins.[61] The formation of the hydrophobic core at the interface between adjacent repeats guarantees high protein stability and drives the overall folding of the repeat protein.[61,65]

In this Chapter, particular attention is therefore put on the effect of inter-repeat loop extension on the native structure of the CTPR array and the overall thermodynamic stability. Previous studies performed in the Itzhaki lab have highlighted how the extension of a single or multiple loops within the same CTPR construct does not alter the secondary structure of the native state.[80,82,146] However, loop grafting in the CTPR scaffold is expected to cause a context-dependent stability loss due to the entropic penalty of closing the loop. This behaviour has been previously described by the Fersht and Regan labs, whereby increasing the length of an

unstructured loop in a globular protein led to reduced protein stability.[194,195] Moreover, a correlation between longer loop length and decreased protein stability was observed.[195] In agreement with these studies, evidence from our own lab also indicate that loop extension destabilises the CTPR constructs through both the entropic cost of loop closure and the consequent weakened interactions at the inter-repeat interface.[80] The effect of loop length (10 to 25 amino acids) on the stability loss was however variable, depending on the total number of repeats in the CTPR array.[80] Moreover, a subsequent study concluded that the frequency of loops also affects the stability, as loops extended at each inter-repeat interface cause adjacent repeats to be only weakly coupled.[146] This extreme configuration was therefore avoided in the $_n$TBP-CTPR$_{2n}$ constructs design, where loops were extended only at alternating inter-repeat interfaces and the other native, stabilising interfaces were maintained.

4.2 Results

Biophysical characterisation of the linear $_n$TBP-CTPR$_{2n}$ constructs was performed using a number of different methods. Far-UV circular dichroism (CD) was used to evaluate the correct folding of the $_n$TBP-CTPR$_{2n}$ proteins. Denaturation studies induced by temperature ramp or increasing concentrations of a chemical denaturant were performed to investigate the stability of the proteins. The TBP was introduced into the inter-repeat loops to functionalise the CTPR scaffold and provide TNKS-binding specificity. The $_n$TBP-CTPR$_{2n}$ constructs' ability to bind the target of interest was therefore investigated using isothermal titration calorimetry (ITC), which represents the "gold-standard" method for the analysis of thermodynamics of binding in solution. The dispersity of the samples was tested by size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS), which is a relatively straightforward and sensitive technique. Finally, we attempted X-ray crystallography to determine the structures of the TBP-grafted CTPR proteins.

4.2.1 Effect of TBP insertions on CTPR folding and stability measured by far-UV circular dichroism

We first investigated the effects of the TBP grafting on protein folding using far-UV circular dichroism (CD). CD refers to the differential absorption of left and right circularly polarized light (LCP and RCP, respectively) by optically active chiral molecules and provides an indication of the secondary structure content of a protein in solution. In particular, proteins with high α-helical content, such as CTPRs, generate CD spectra with double minima at 208

nm and 222 nm.[61,188,189] All $_n$TBP-CTPR$_{2n}$ constructs showed the expected spectrum for CTPR repeats, confirming that all $_n$TBP-CTPR$_{2n}$ proteins are correctly folded (Figure 4.1A). A similar CD profile was previously observed in the Itzhaki lab for CTPR proteins with different grafted loops, confirming that loop extension does not radically alter the secondary structure of the CTPR repeats.[80,82,146] In all the CD spectra of CTPR proteins previously published, including those shown here, the minimum at 208 nm is not pronounced and might represent a characteristic feature of this protein system.[61,80,82,146,188,189] Of note is the linear increase in the molar ellipticity at 222 nm observed with increasing number of the CTPR repeats, which further proves that the proteins are natively folded (Figure 4.1B).
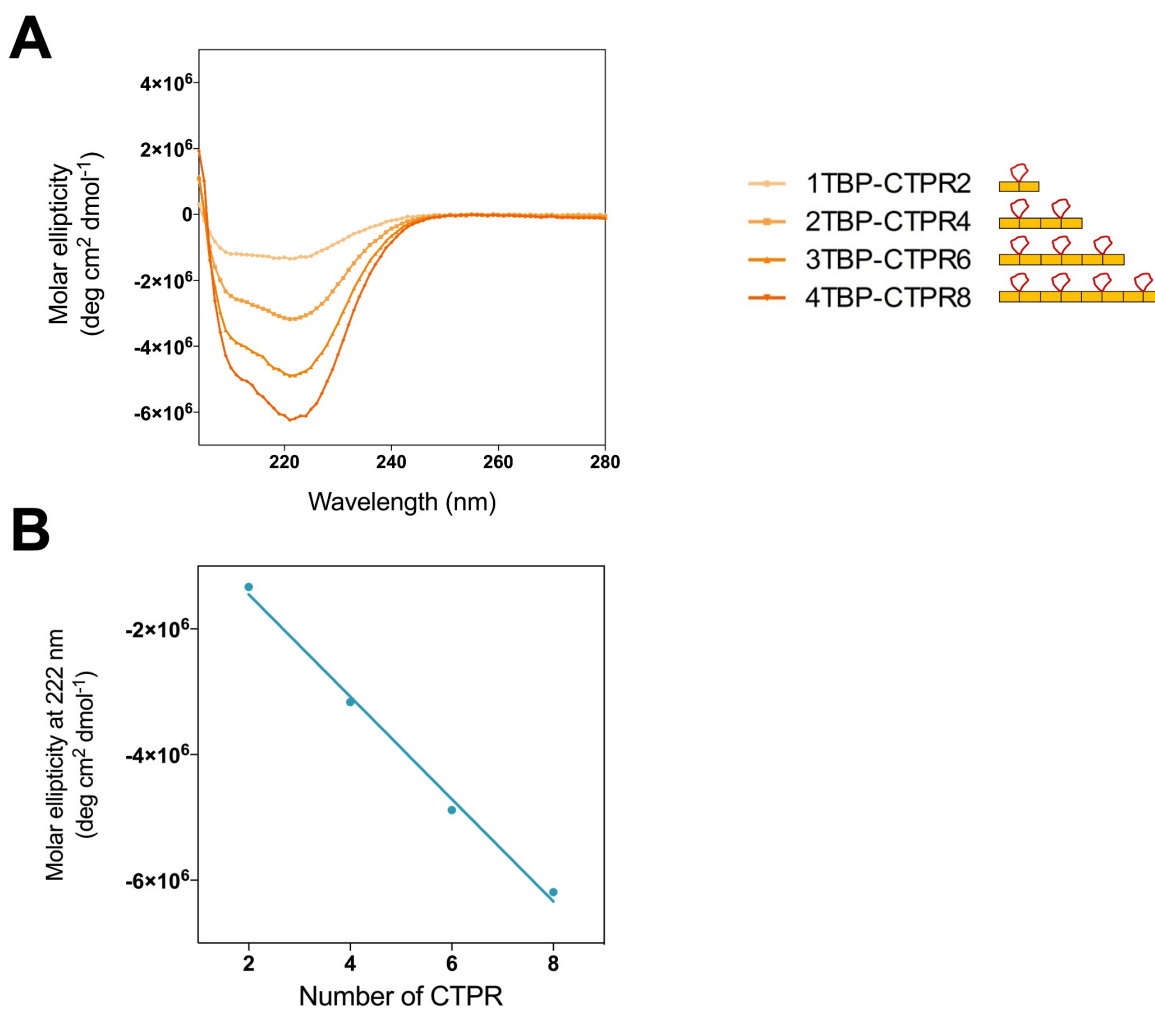
Figure 4.1: CD spectroscopy of $_n$TBP-CTPR$_{2n}$ proteins. (A) Far-UV CD spectra of 1TBP-CTPR2, 2TBP-CTPR4, 3TBP-CTPR6 and 4TBP-CTPR8. Protein concentration was 20 µM in 10 mM sodium phosphate pH 7.4, 150 mM NaCl at 20°C. On the right, the figure legend with a schematic representation for each construct. (B) Plot of molar ellipticity at 222 nm, obtained from the experiment in A, against the number of CTPRs in the protein.

Next, thermal denaturation was performed to investigate the stability and folding of the $_n$TBP-CTPR$_{2n}$ constructs. As ellipticity at 222 nm is a good measure of the α-helical content, this wavelength was chosen to follow the denaturation of the protein.[196] According to the classical cooperative folding model, protein denaturation is visualised as a two-state transition from the folded to the unfolded state without stable intermediate states. When a complete two-state transition is recorded, it is possible to measure the apparent midpoint of the unfolding transition, indicated with T_m, which corresponds to the temperature at which the folded and unfolded populations are equal.[196] The value of T_m is a simple way to compare the stability of different constructs.

All $_n$TBP-CTPR$_{2n}$ proteins were found to be extremely thermostable, with only the smallest protein, 1TBP-CTPR2, undergoing complete denaturation. The larger $_n$TBP-CTPR$_{2n}$ proteins remained partly folded even at the highest temperature. The apparent midpoint of the unfolding transition is therefore provided only for 1TBP-CTPR2 and corresponds to 75°C (Figure 4.2). This result is in agreement with previous thermostability studies using CTPR proteins of various lengths, which concluded that the stability increases with increasing number of repeats,[61] but the apparent T_m value was not measured beyond CTPR3. Previous members of the Itzhaki group also investigated the apparent T_m values of various CTPR2 constructs, with and without loop extension.[82]

It is possible that for the larger proteins approximately two repeats undergo unfolding at the highest temperatures, since the molar ellipticity of $_n$TBP-CTPR$_{2n}$ at these temperatures is similar to the molar ellipticity measured for $_{n-1}$TBP-CTPR$_{2n-2}$ at 20°C. These are most likely to be the terminal repeats, as they have only one neighbouring repeat to form stabilising inter-repeat interactions and are therefore less stable than the internal repeats.[65,197]
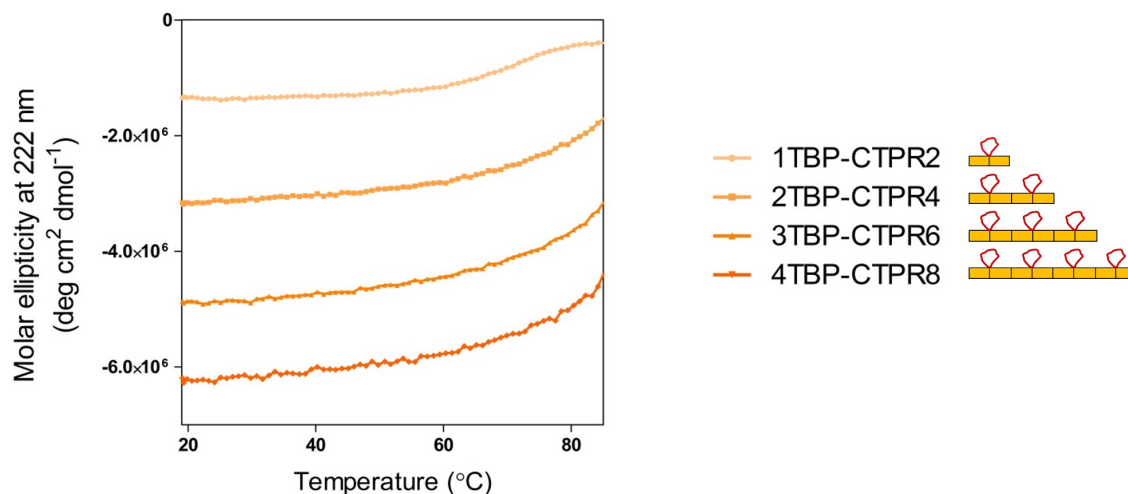
Figure 4.2: Thermal denaturation curves of the samples in Figure 4.1, monitored by CD. On the right, the figure legend with a schematic representation for each construct.

Following thermal denaturation, the samples were allowed to return to 20°C, and the CD spectrum of each construct was re-measured to evaluate the reversibility of the unfolding reaction. For all of the proteins, there was no significant difference between the CD spectrum recorded before and after thermal denaturation, indicating that unfolding is reversible (Figure 4.3).
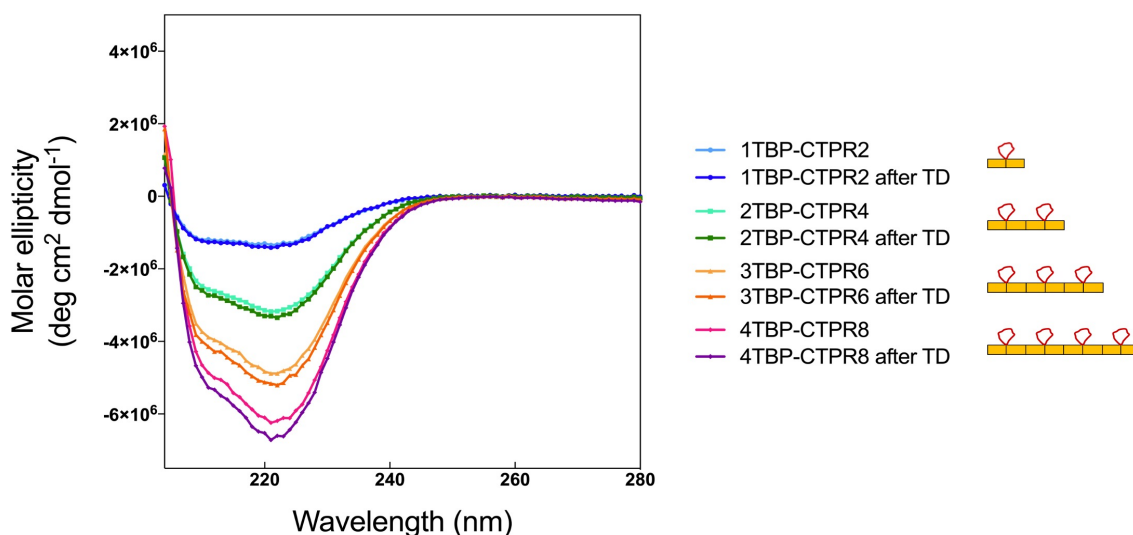
Figure 4.3: CD spectra of the proteins before and after thermal denaturation. Far-UV CD spectra of 1TBP-CTPR2, 2TBP-CTPR4, 3TBP-CTPR6 and 4TBP-CTPR8. All proteins were at 20 μM concentration in 10 mM sodium phosphate pH 7.4, 150 mM NaCl, and measurements were made at 20°C, before (light colour) and after (darker colour) thermal denaturation (TD).

4.2.3 Effect of TBP insertions on CTPR folding and stability measured by chemical denaturation

Protein stability was also assessed by chemical-induced denaturation. The aim was to investigate the destabilising effect of loop extension and provide a comparison with the CTPR counterpart without loops. Although loops have been extended in alternating rather than in all inter-repeat interfaces, a smaller, but still measurable destabilising effect is expected, as demonstrated in a previous study.[146] On the other hand, this assay should also confirm the improved stability with increasing numbers of repeats, as observed in the thermal unfolding experiment. Chemical-induced denaturation is achieved by incubating proteins with increasing concentrations of guanidinium hydrochloride (GdmHCl). GdmHCl is a chaotropic agent and one of the strongest protein denaturants, although its mechanism of action is still not fully understood. A high-throughput method developed by a previous member of the Itzhaki lab was used, which relies on the fast and accurate monitoring of the fluorescence emitted by the tryptophan residues (one per repeat) using a microplate reader.[153] Tryptophan is an intrinsic fluorophore due to its aromatic nature, and its fluorescence emission spectrum is highly affected by changes in the microenvironment induced by chemical denaturation. Given its simplicity, chemical denaturation is widely used as a highly sensitive technique to assess protein stability and thereby to optimise the appropriate formulation conditions for biologics.[198] This method has also been used extensively to investigate the stability of CTPR arrays of different lengths.[80,146,191]

Similar to the thermal denaturation studies, a single transition between the folded and unfolded states was observed for all of the constructs, both the CTPR series and the $_n$TBP-CTPR$_{2n}$ series. This observation is in agreement with previous chemical denaturation studies on CTPR constructs of increasing length.[80,146,191] The data were fitted to a two-state model with sloping baselines to obtain the midpoints of unfolding (D$_{50\%}$), the slope of the transition (m value, a parameter related to the change in solvent-exposure upon unfolding) and the free energy of unfolding (ΔG_{N-U}^{H2O}) (Figure 4.4 and Table 4.1). First, the $_n$TBP-CTPR$_{2n}$ showed an expected decrease in the stability compared with their respective counterparts without the TBP loops. As shown previously,[80] loop insertion has a destabilising effect through both the entropic cost of loop closure and a consequent weakened coupling of adjacent repeats. Second, there is a significant increase in D$_{50\%}$, m value and ΔG_{N-U}^{H2O} with increasing number of CTPR units. The largest increase in stability is between 1TBP-CTPR2 and 2TBP-CTPR4, whereas for all the other $_n$TBP-CTPR$_{2n}$ constructs the increases are much smaller. We do not see the incremental

stability increase that characterises the CTPR2, CTPR3, CTPR4 and CTPR6 series. The explanation is that 1TBP-CTPR2 has no native inter-repeat interface and the three larger proteins have a native CTPR2 interface alternating with the loop-inserted interface and therefore the stability does not build up.
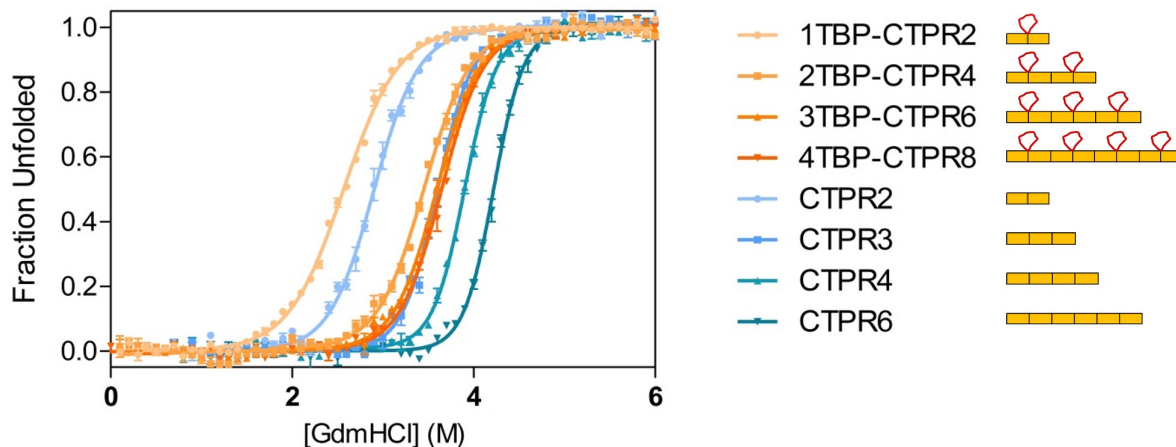
Figure 4.4: Equilibrium denaturation curves of CTPR and $_n$TBP-CTPR$_{2n}$ proteins monitored by fluorescence. Protein concentration was 6-10 µM in 50 mM sodium phosphate pH 6.8, 150 mM NaCl at 25ºC. The CTPR proteins with no extended loops were purified and analysed by Dr. Albert Perez-Riba, a member of the Itzhaki group at the time. The data are fitted to a two-state model with sloping baselines.

Protein	$D_{50\%}$ (M)	m (kcal mol^{-1} M^{-1})	ΔG_{N-U}^{H2O} (kcal mol^{-1})
1TBP-CTPR2	2.57 ± 0.01	1.94 ± 0.05	-4.98 ± 0.13
2TBP-CTPR4	3.46 ± 0.01	2.33 ± 0.09	-8.06 ± 0.31
3TBP-CTPR6	3.60 ± 0.01	2.49 ± 0.06	-8.96 ± 0.22
4TBP-CTPR8	3.65 ± 0.01	2.55 ± 0.07	-9.31 ± 0.26
CTPR2	2.89 ± 0.02	2.42 ± 0.12	-6.99 ± 0.35
CTPR3	3.60 ± 0.01	3.06 ± 0.15	-11.02 ± 0.54
CTPR4	3.90 ± 0.01	3.37 ± 0.13	-13.14 ± 0.51
CTPR6	4.22 ± 0.01	3.54 ± 0.15	-14.94 ± 0.63

Table 4.1: Values of $D_{50\%}$, m and ΔG_{D-N}^{H2O} obtained from a two-state fit with sloping baselines of the equilibrium denaturation data. All measurements were performed in triplicate, and the errors listed are the SE of the mean.

4.2.4 Binding studies between $_n$TBP-CTPR$_{2n}$ constructs and TNKS2 ARC4 measured by isothermal titration calorimetry (ITC)

The TBP loop sequence is a consensus TNKS-binding motif identified by Guettler *et al.* using a peptide library screen.[46] The binding affinity of the TBP to TNKS2 ARC4 was measured by the same authors and was reported to be 0.6 ± 0.04 μM, and it is expected to bind to the other ARC subdomains with a similar affinity.[46] The crystal structures of TNKS2 ARC4 in complex with diverse TNKS-binding motifs, including 3BP2, show how they all employ an identical ARC4:peptide binding mode (Figure 4.5A).[46] The same extended and unstructured conformation characterises all TNKS-binding peptides, when bound to ARC4 or other TNKS ARC subdomains.[89,144]

A previous Itzhaki lab member generated a series of macrocyclic variants of the TBP peptide using a number of different chemical cross-linkers.[85] Macrocyclisation serves to constrain the peptide in its bioactive conformation and to make it protease-resistant, for use as an inhibitor in the cell. The crystal structure of a macrocyclic peptide, cp4n2m3, in complex with TNKS2 ARC4 shows that the peptide does assume an elongated conformation and interacts with TNKS2 ARC4 in a similar way to that of the natural peptides (Figure 4.5B).[85]

Here we took the same TBP sequence identified by Guettler *et al.* and grafted onto the loop between consecutive CTPR repeats with DPXX linker residues on each side (DPNN preceding and DPRS following the TBP loop). To investigate whether the TBP sequence constrained in this way is able to maintain its binding properties to TNKS, we wanted to measure the binding affinity of each $_n$TBP-CTPR$_{2n}$ construct for TNKS and establish whether all of the TBP loops in the multivalent constructs are accessible for binding. To simplify the analysis, the binding of each $_n$TBP-CTPR$_{2n}$ construct was measured in the presence of TNKS2 ARC4 (the fourth ankyrin-repeat cluster of TNKS2), which contains a single binding site for the TBP (Figure 4.6A).

The analysis required first the recombinantly expression and purification of the TNKS2 ARC4 construct, following a protocol previously established in the lab and described in Section 2.4.3. The TNKS2 ARC4 construct was expressed as a fusion protein to a N-terminal glutathione S-transferase (GST) tag immediately followed by a thrombin cleavage site. The GST tag is a quite bulky affinity tag of 211 amino acids (26 kDa) that often promotes greater expression and higher solubility of recombinant proteins than expression without a tag. Moreover, when adopted as an affinity tag, it enables to reach higher purity rates from affinity chromatography compared to less specific tags, such as the His$_6$ tag.
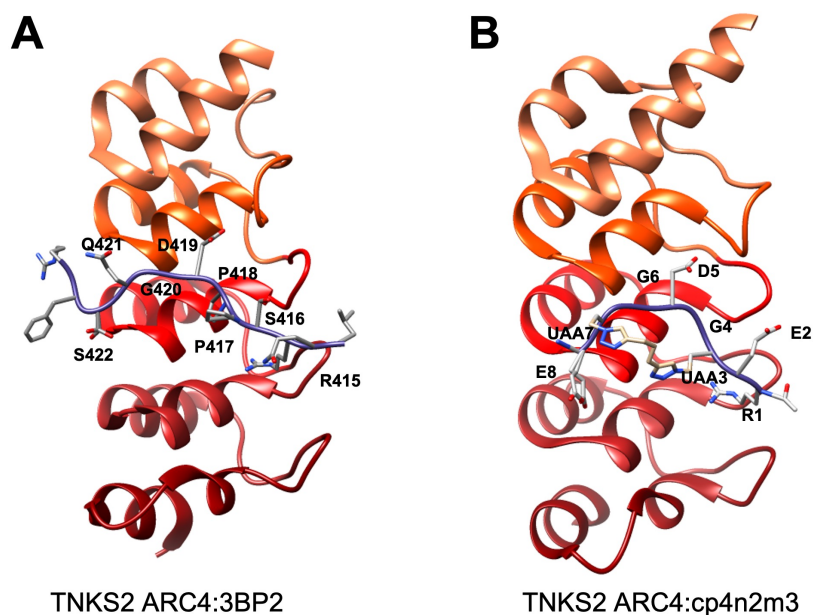
Figure 4.5: Crystal structures of TNKS2 ARC4 binding moieties. (A) Ribbon representation of TNKS2 ARC4 binding the 3BP2 peptide, with the peptide's side chains shown in stick representation (PDB: 3TWR). The five ANK repeats of TNKS2 ARC4 are coloured in shades of red, while the 3BP2 peptide is in purple with the side chains coloured by heteroatoms (oxygen in red, nitrogen in blue) and the eight residues that constitute the TNKS-binding motif (RSPPDGQS) are labelled. (B) Ribbon representation of TNKS2 ARC4 binding the cp4n2m3 macrocyclic peptide, with the peptide's side chains shown in stick representation (PDB: 5BXO). The five ANK repeats of TNKS2 ARC4 are coloured in shades of red, and the peptide is in purple with the side chains coloured by heteroatoms (oxygen in red, nitrogen in blue) and the eight resides that constitute the TNKS-binding motif (REXGDGXE) are labelled. UAA3 and UAA7 correspond to unnatural amino acids required for the cross-linking click reaction. Images were created using the software UCSF Chimera.

Batch purification of GST-tagged TNKS2 ARC4 was performed using the Glutathione Sepharose 4B resin (Cytiva). This method also allows on beads cleavage of the GST tagged protein in the presence of bovine thrombin. Following overnight cleavage, highly pure and untagged TNKS2 ARC4 protein could be eluted from the beads, while the GST tag remained bound to the Glutathione resin (Figure 4.6B). Experimental evidence from the Itzhaki lab showed that the same approach cannot be applied to Ni-NTA beads, further validating the choice of the GST tag over the His$_6$ tag for the purification of TNKS2 ARC4. Following affinity purification, the protein was further purified by size-exclusion chromatography, with the resulting highly purity protein shown in Figure 4.6C.
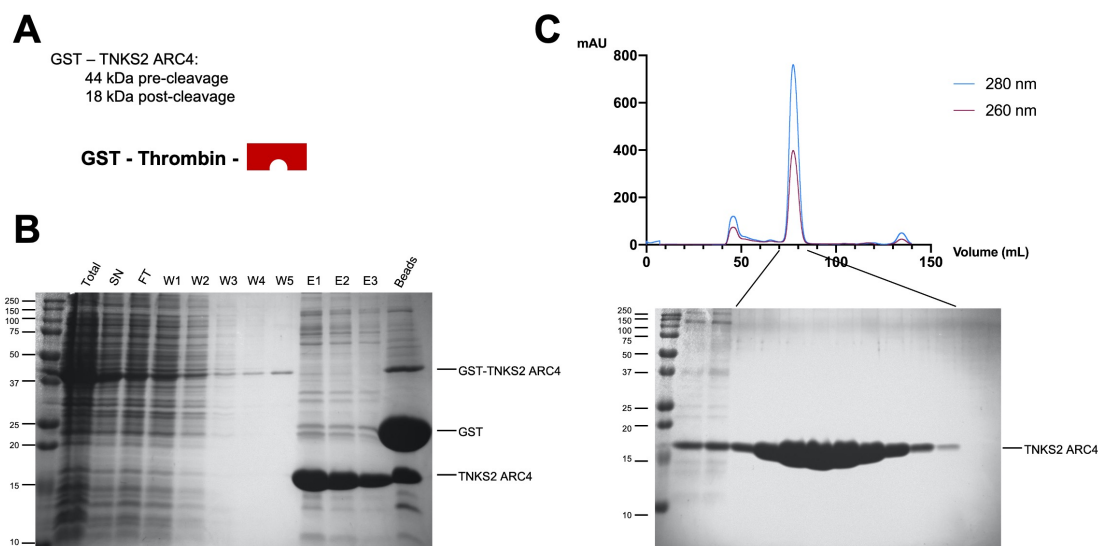
Figure 4.6: GST-tagged TNKS2 ARC4 purification. (A) Schematic representation of the GST-tagged TNKS2 ARC4 construct. The molecular weight of the construct before and after Thrombin cleavage is provided. (B) SDS-PAGE gels with lanes corresponding to the fractions collected during affinity chromatography (SN: supernatant, FT: flow-through, W: wash, E: elution) (C) Top: gel filtration chromatogram obtained when running the sample eluted in A on a HiLoad 16/600 Superdex 75 pg column (Cytiva). Bottom: SDS-PAGE gels with lanes corresponding to the eluted fractions of the main peak observed in the chromatogram above.

The binding between $_n$TBP-CTPR$_{2n}$ constructs and TNKS2 ARC4 was studied by ITC. This technique provides the thermodynamic parameters of interaction in solution, as well as details on the stoichiometry of binding, an important consideration in our analysis.[199] Depending on the nature of the interaction, the instrument measures the amount of heat generated or taken up following each injection of ligand into the sample cell containing the binding partner. For these experiments, TNKS2 ARC4 was titrated into the cell, containing either one of the $_n$TBP-CTPR$_{2n}$ constructs or a control (CTPR6 or ITC buffer). A representative dataset obtained by ITC is shown in Figure 4.7.

All $_n$TBP-CTPR$_{2n}$ constructs showed similar low-micromolar affinities (average K_d of 14 μM), confirming that the binding of the TBP for the target protein is maintained also when grafted in the CTPR scaffold. Importantly, the stoichiometry of binding, indicated with N, was found to increase in proportion to the number of TBP loops, from one to three, confirming that all sites are solvent-exposed and available for binding (Table 4.2). We speculate that the stoichiometry for 4TBP-CTPR8 is lower than 4, possibly due to steric hindrance effects between multiple TNKS2 ARC4 molecules binding to the same 4TBP-CTPR8 molecule. The presence of impurities in the 4TBP-CTPR8 sample might also cause a shift in the value of N,

due to the impossibility of accurately measuring the protein concentration. No binding could be detected for the control samples, where TNKS2 ARC4 was titrated into the sample cell containing the CTPR6 protein with no TBP loops, or simply ITC Buffer.
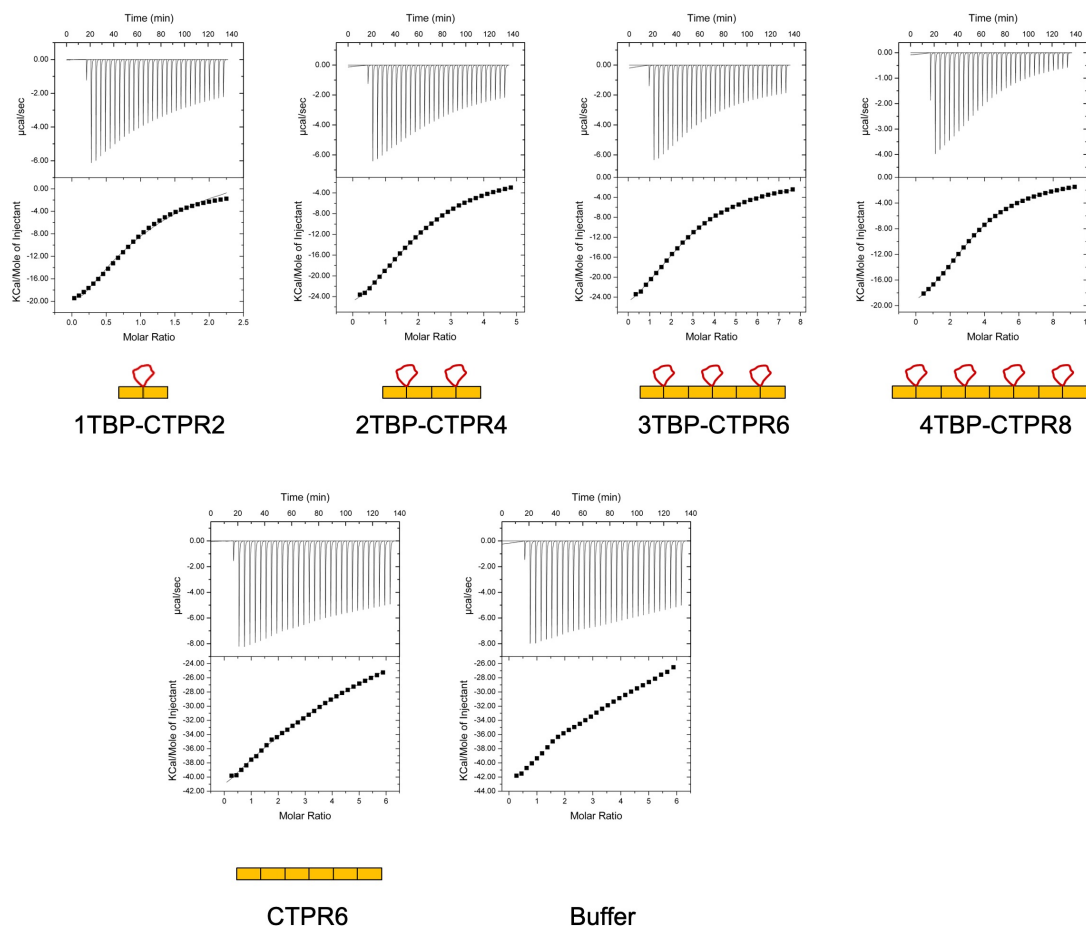
Figure 4.7: ITC measurement, with TNKS2 ARC4 titrated into the main cell containing the indicated CTPR constructs or ITC Buffer. Top from left to right: TNKS2 ARC4 (500 µM) into 1TBP-CTPR2 (50 µM), TNKS2 ARC4 (500 µM) into 2TBP-CTPR4 (25 µM), TNKS2 ARC4 (500 µM) into 3TBP-CTPR6 (16.6 µM), TNKS2 ARC4 (500 µM) into 4TBP-CTPR8 (12.5 µM). The concentration of $_n$TBP-CTPR$_{2n}$ was calculated as 50/n, so that the molar ratio between the number of TBP loops and TNKS2 ARC4 remains constant across the runs. Bottom from left to right: control runs where TNKS2 ARC4 (500 µM) was titrated into CTPR6 (16.6 µM) or ITC buffer. All experiments were performed at 25ºC. A representative ITC dataset is shown.

The values of ΔG (Gibbs free energy change), ΔH (enthalpy change) and $-T\Delta S$ (entropy change) obtained for the four $_n$TBP-CTPR$_{2n}$ proteins are listed in Table 4.3 and plotted on the right of it. The results show that in all cases the interaction is exothermic (negative ΔH) and the negative ΔG is indicative of a spontaneous process, enthalpically driven. It is also possible to notice a slight tendency for both the enthalpic gain and entropic cost of binding to increase

with increasing number of TNKS2-binding loops. The increase in enthalpy with increasing number of binding loops is to be expected, as having multiple binding sites in close proximity should make it more likely for the protein to bind a target again after dissociation. The greater entropic cost of ordering a greater number of loops upon binding is also as expected, as they are all available for binding. The greater entropic cost is offset by the greater enthalpic gain, and hence the value of ΔG is similar for all four $_n$TBP-CTPR$_{2n}$ proteins.

Protein	Stoichiometry (N)	K_d (µM)
1TBP-CTPR2	1.01 ± 0.03	14.68 ± 0.99
	0.90 ± 0.02	12.80 ± 1.10
2TBP-CTPR4	2.18 ± 0.02	11.89 ± 0.26
	2.05 ± 0.01	18.59 ± 0.39
3TBP-CTPR6	2.72 ± 0.03	5.62 ± 0.21
	2.77 ± 0.02	20.12 ± 0.39
4TBP-CTPR8	3.38 ± 0.01	13.61 ± 0.19
	2.91 ± 0.03	15.48 ± 0.43

Table 4.2: Stoichiometry (N) and dissociation constant (K_d) obtained by ITC measurements, characterising the binding between $_n$TBP-CTPR$_{2n}$ proteins and TNKS2 ARC4. For each protein, the values from two independent experiments are listed.

Protein	ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
1TBP-CTPR2	-6.60	-15.81 ± 0.73	9.21
	-6.69	-24.87 ± 0.87	18.18
2TBP-CTPR4	-6.73	-18.35 ± 0.26	11.62
	-6.47	-34.12 ± 0.32	27.65
3TBP-CTPR6	-7.18	-12.39 ± 0.21	5.21
	-6.44	-35.85 ± 0.37	29.41
4TBP-CTPR8	-6.65	-25.04 ± 0.21	18.39
	-6.59	-31.44 ± 0.48	24.85
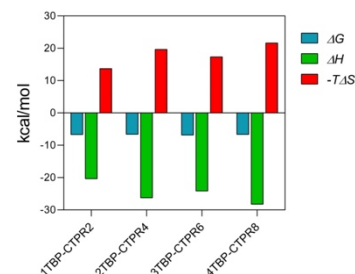
Table 4.3: Values of the changes in Gibbs free energy (ΔG), enthalpy (ΔH) and entropy ($-T\Delta S$) for the interactions of the $_n$TBP-CTPR$_{2n}$ proteins with TNKS2 ARC4 obtained by ITC. The values listed are from two independent experiments. For each parameter, the average from two independent experiments is plotted on the right.

4.2.5 Assessment of the oligomerisation states of the CTPR proteins

The oligomerisation state of some representative CTPR proteins was analysed by size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS). SEC-MALS relies on the separation of proteins according to their size through SEC, followed by the oligomerisation analysis of the eluted fractions according to their light scattering potential. This method also allowed us to further confirm the protein trimerisation mediated by the foldon domain. The analysis was performed on two sets of proteins: 3TBP-CTPR6 and 3TBP-CTPR6-foldon, CTPR6 and CTPR6-foldon. The four proteins were purified by His affinity chromatography, followed by size-exclusion on a Superdex 200 10/300 column. The retention volume for each eluted protein on the size-exclusion column can already provide insights into the monomeric versus trimeric conformations, due to the difference in molecular weight. Given their smaller size, the monomeric 3TBP-CTPR6 and CTPR6 eluted at much higher retention volumes than their trimeric counterparts, confirming the trimerisation of their counterparts containing the foldon motif. The elution volume of the trimeric CTPR6-foldon remains questionable, as it elutes earlier than 3TBP-CTPR6-foldon, despite its smaller molecular weight (Figure 4.8).
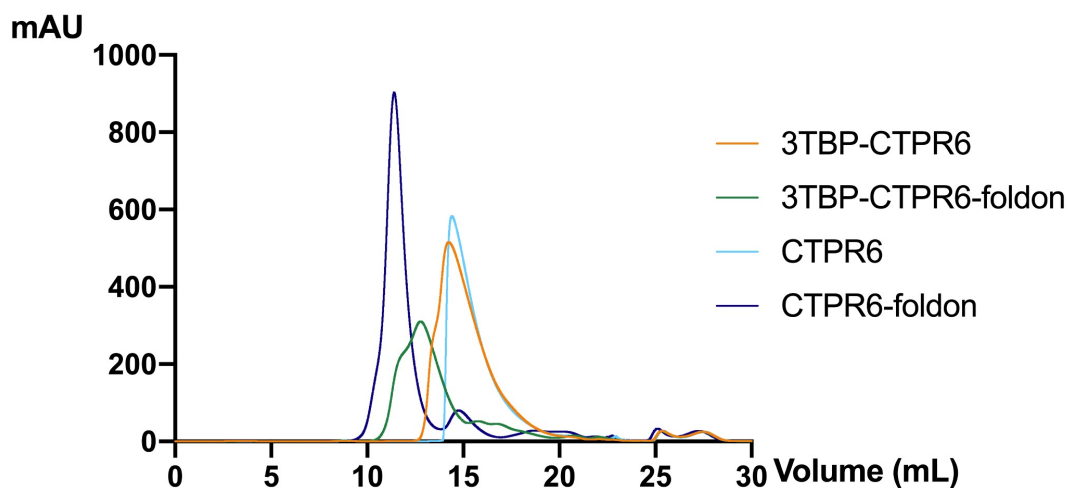
Figure 4.8: Size-exclusion chromatograms. 3TBP-CTPR6 (orange), 3TBP-CTPR6-foldon (green), CTPR6 (light blue) and CTPR6-foldon (dark blue) were separately purified by size-exclusion chromatography on a S200 10/300 column.

The purified proteins were run over SEC-MALS with the help of Dr. Stephen McLaughlin (MRC, LMB, Cambridge). SEC was performed on the same S200 10/300 column and proteins

eluted with a similar profile. The column was connected to a MALS detector for direct analysis of each eluted fraction. The SEC-MALS results demonstrate that 3TBP-CTPR6 and 3TBP-CTPR6-foldon are monodispersed in solution (Figure 4.9 top). The mass for 3TBP-CTPR6 (30.0 kDa calculated molecular weight) is 40.2 kDa at the beginning of the peak, with the trailing edge having a mass of 31.6 kDa, resulting in a peak average molecular weight of 38.4 kDa, consistent with the protein being monomeric. The vast majority of the 3TBP-CTPR6-foldon sample (105.6 kDa calculated molecular weight as a trimer) elutes at an average molecular weight of 130.6 kDa, consistent with it forming a trimer with some less-populated oligomerisation species at lower and higher molecular weight also present (99.3, 203.5 and 320.8 kDa, respectively). The results for CTPR6 and CTPR6-foldon constructs, instead, indicate dimerisation for both proteins under the same experimental conditions (Figure 4.9 bottom). The molecular weight observed for CTPR6 (expected molecular weight 26.2 kDa) at the beginning of the peak is 49.6 kDa and 30.7 kDa at the trailing edge, leading to an average molecular weight of 44.3 kDa. The result might indicate a fast-exchanging equilibrium between the monomer and the dimer form under these experimental conditions. The vast majority of the CTPR6-foldon sample (94.0 kDa calculated molecular weight as a trimer) elutes at an average molecular weight of 370.2 kDa, corresponding to the mass of four CTPR6-foldon trimers (94.0 x 4 = 376.0 kDa). This oligomerisation product is achievable when each CTPR6 of a trimer molecule interacts with a CTPR6 of another trimer, consistent with the monomeric CTPR6 being instead mostly dimeric. Less-populated oligomerisation species for the CTPR6-foldon sample elute with a mass of 543.0, 234.8 and 132.8 kDa. The gel filtration was repeated at a 10-fold lower concentration of CTPR6-foldon, and the distribution of masses was found to be the same (Figure 4.9 bottom).
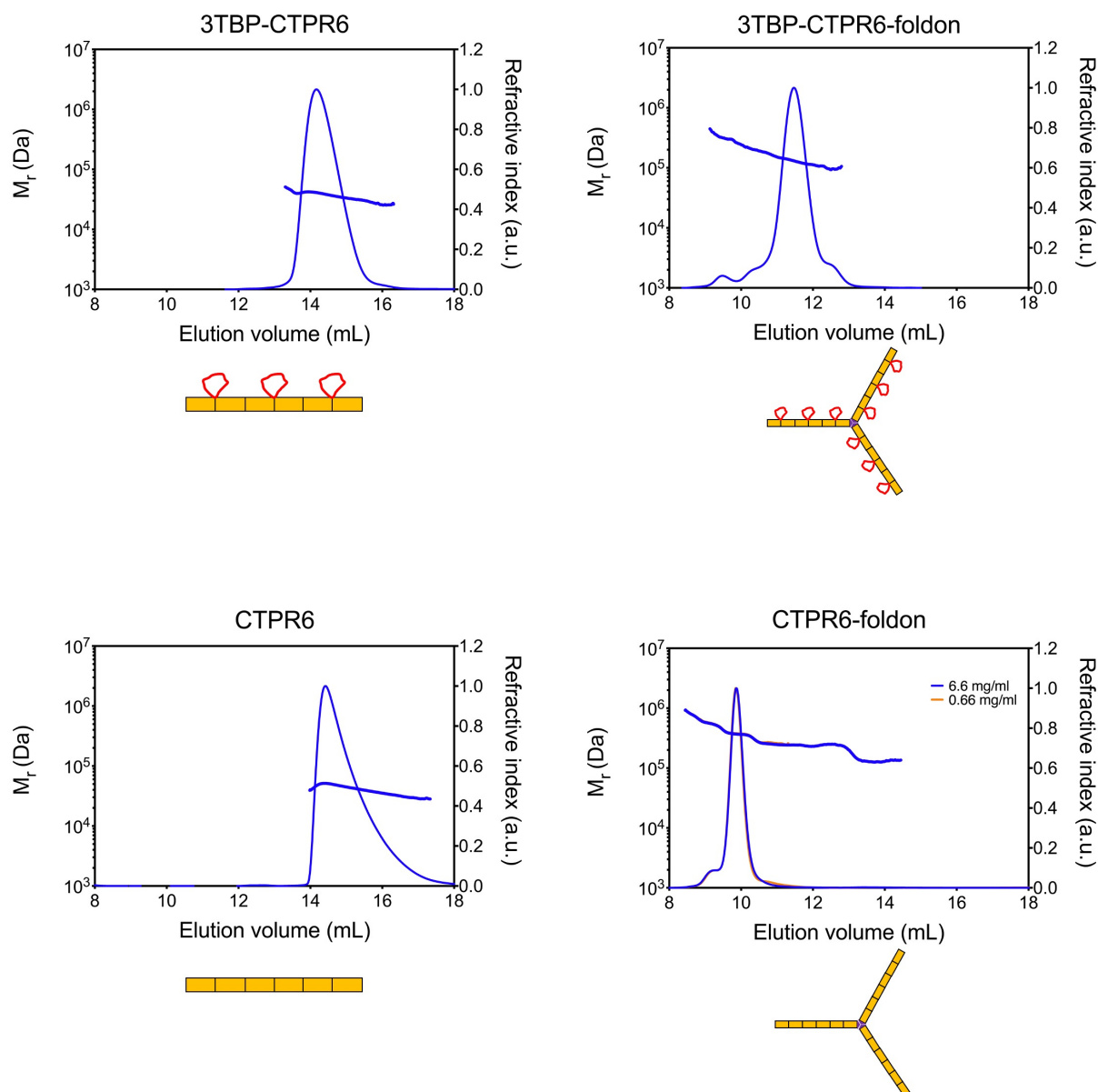
Figure 4.9: SEC-MALS analysis of 3TBP-CTPR6, CTPR6 and their trimeric counterparts. The total protein concentration was 2.2 mg/mL for 3TBP-CTPR6, 1.7 mg/mL for 3TBP-CTPR6-foldon, 2.3 mg/mL for CTPR6 and 6.6 mg/mL for CTPR6-foldon. A schematic representation of the constructs used is provided below the corresponding graph.

4.2.6 X-ray crystallography studies

Interest in determining the structure of a representative $_n$TBP-CTPR$_{2n}$ construct led us to investigate the possibility of crystallising these constructs. X-ray crystallography is a powerful technique that allows protein structure determination at high resolution. Protein crystallisation remains a challenging method, as not all proteins can form crystals, and it is impossible to predict whether or not a protein will crystallise and the crystallisation conditions required. If

successful, the 3D structure of a $_n$TBP-CTPR$_{2n}$ protein, on its own or in complex with TNKS2 ARC4, would provide insights into the conformation and orientation of the TBP loop under the two sample conditions. Moreover, it would allow us to identify any conformational changes in the inter-repeat interface induced by the loop extension. The feasibility of x-ray crystallography for this project was supported by the fact that multiple crystal structures of CTPR molecules have already been published, as well as those of the TNKS2 ARC4 subdomain in complex with the TBP peptide, including the one obtained by our lab.[46,61,64,66,85] Moreover, the high expression levels and high purity obtained for the $_n$TBP-CTPR$_{2n}$ proteins and TNKS2 ARC4 are ideal starting points for crystallography. However, the presence of an unstructured loop might represent an element of instability for the initiation of crystal packing, thereby preventing crystallisation. Initial unsuccessful attempts at crystallisation using His-tagged 2TBP-CTPR4 in the presence or absence of TNKS2 ARC4 led us to re-evaluate the optimal construct design for crystallisation.

Several factors were considered with the aim of maximising the chances of success:

1. Given the loss in stability observed due to the grafted loop, as well as its unstructured nature, the number of TBP loops was reduced to one.

2. Stability can be enhanced by adding more CTPRs, with a big increase observed when the number of repeats is increased from two to four, and the single TBP loop should be introduced at an inter-repeats interface located in the middle of the CTPR protein rather than at a terminal repeat.

3. Protein tags need to be removed.

4. Most of the previously crystallised CTPR constructs include a C-terminal "solvating" helix.[61,64] The solvating helix is a variant of helix A, with four solvent-exposed hydrophobic residues mutated into polar or charged amino acids (two lysine and two glutamine residues), and it has been shown to enhance the solubility of the CTPR proteins (Figure 4.10).[61]

5. The TBP loop represents a flexible element that consequently might not be visible in the crystal structure. In complex with TNKS2 ARC4, the TBP loop is likely to become more rigid and would be resolved, and therefore it was decided to attempt crystallography of a TBP-CTPR protein in complex with TNKS2 ARC4.

6. High protein purity and high protein concentration are required. To minimise the formation of recombination products during cloning and protein expression, the DNA sequence encoding each repeat was maximally differentiated, where possible.

A

CTPR Helix A   AEAWYNLGNAYYK QGD
                         ↓↓   ↓↓
Solvating Helix   AEAKQNLGNAKQK QG
                         4 5   11 12

B
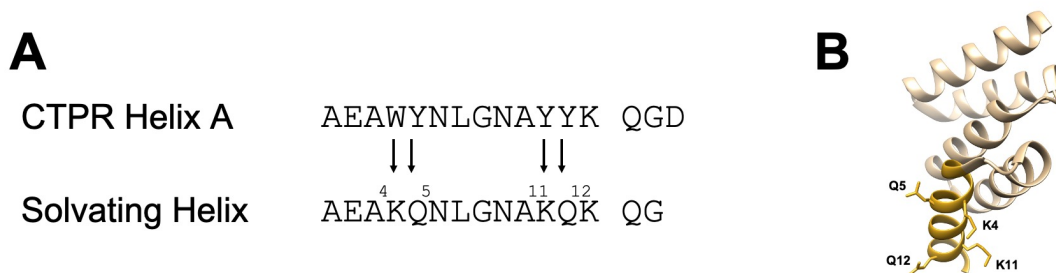
Figure 4.10: Solvating helix. (A) Protein sequence alignment between a CTPR helix A and the solvating helix. Each mutated residue is indicated by an arrow and numbered. (B) Ribbon structure of the CTPR2-Solvating Helix protein obtained by Main *et al.*[61] The CTPR repeats are coloured in pale yellow, the solvating helix is in goldrod. The side chains of the mutated residues are shown and labelled (PDB: 1NA3). The image was generated using the software UCSF Chimera.

With these considerations in mind, a new protein construct was designed to fulfil the requirements detailed above and was named 1TBP-CTPR4-SolvH (protein sequence available in the Appendix B). It consists of four CTPRs (with identical amino acid sequence, but diversified at the DNA level), a TBP loop located in the central inter-repeat interface and a C-terminal solvating helix, as graphically represented in Figure 4.11A. The construct was fused to a N-terminal GST tag followed by a thrombin recognition sequence to allow on-beads cleavage of the protein and achieve a higher sample purity by GST affinity purification compared to His affinity chromatography (Figure 4.11A). GST-tagged 1TBP-CTPR4-SolvH was purified by affinity chromatography on GST beads and eluted by O/N thrombin cleavage on beads (Figure 4.11B). Despite the abundance of uncleaved product still bound to the resin, the amount of untagged, cleaved protein collected in the eluted fractions was sufficient to proceed with the purification. The protein was therefore further purified by gel filtration on a Superdex 75 16/600 column, which successfully separated the protein from all the higher molecular weight contaminants (Figure 4.11C). Figure 4.11C also provides a comparison between the chromatograms of 1TBP-CTPR4-SolvH and 2TBP-CTPR4, a CTPR construct of similar size (therefore eluting at a similar retention volume). It is worth noting how the peak shape differs in the two chromatograms: the presence of the solvating helix improves the elution profile of the protein, probably by preventing unspecific binding to the column. The elution peak of 1TBP-CTPR4-SolvH is indeed more symmetrical and does not show the trailing tail that instead characterises the 2TBP-CTPR4 chromatogram.
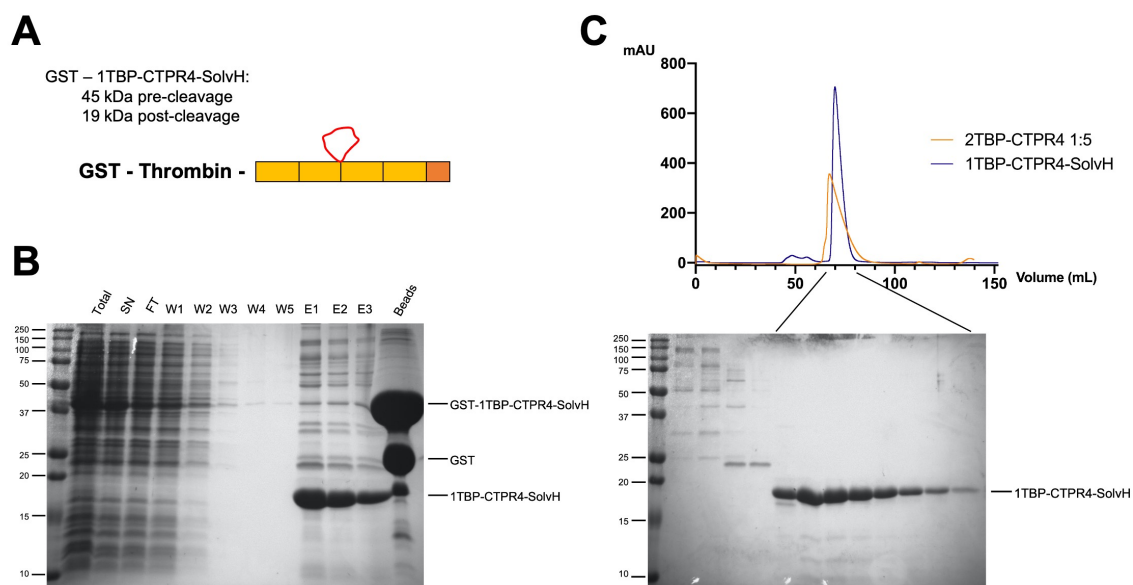
Figure 4.11: 1TBP-CTPR4-Solvating Helix purification. (A) Schematic representation of the GST-tagged 1TBP-CTPR4-SolvH construct. The molecular weight of the construct before and after Thrombin cleavage is provided. (B) SDS-PAGE protein gel with lanes corresponding to different fractions collected throughout the GST affinity chromatography (Total: total lysate, SN: supernatant, FT: flowthrough, W: wash, E: elution, Beads: bead sample after O/N thrombin cleavage). (C) Size-exclusion chromatogram of the untagged 1TBP-CTPR4-SolvH (blue) over a Superdex 75 16/600 column and SDS-PAGE protein gel with lanes corresponding to fractions of the eluted peaks. A comparison with the size-exclusion chromatogram of 2TBP-CTPR4 diluted 1:5 (orange, same as in Figure 3.9A) is provided.

The protein was concentrated to 35.7 mg/mL, and crystal screening was performed at the Crystallographic X-ray Facility (Department of Biochemistry, University of Cambridge). Protein sample was dispensed at 1:1 and 2:1 protein:buffer dilutions onto a variety of crystallisation plates covering a broad range of buffers and precipitants. Despite the very high protein concentration, the high solubility of the CTPR scaffold, further enhanced by the presence of the TBP loop and the solvating helix, might have prevented crystal formation, as all drops look clear with no evidence of precipitation even after months.

In order to determine the crystal structure of the protein complex, the latter was induced by incubating 1TBP-CTPR4-SolvH in the presence of a molar excess of untagged TNKS2 ARC4 (1:3). The sample was run on a Superdex 75 gel filtration column, which successfully separated the complex (corresponding to the first eluted peak, pink chromatogram) from the TNKS2 ARC4 molecules in excess (corresponding to the second peak, pink chromatogram Figure 4.12). The protein complex, given its higher molecular weight, eluted earlier than the single components (Figure 4.12 top). It is worth noting how, despite the weak binding affinity

between the TBP loop and TNKS2 ARC4 (K_d of 14 µM), the complex remained intact through the gel filtration column without the need of chemical cross-linking.
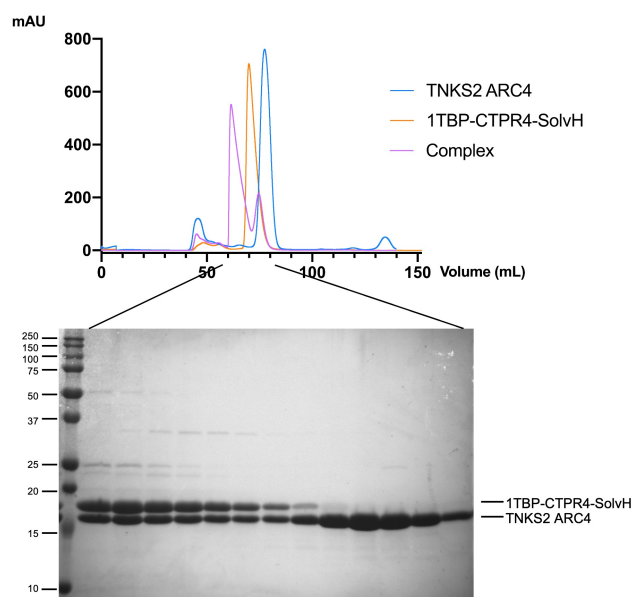
Figure 4.12: TNKS2 ARC4:1TBP-CTPR4-SolvH complex purification. (Top) Comparison between the size-exclusion chromatogram traces of TNKS2 ARC4 (blue), 1TBP-CTPR4-SolvH (orange) and the protein complex formed between the two (purple) run over a Superdex 75 16/600 column. (Bottom) SDS-PAGE protein gel with lanes corresponding to the peak fractions eluted from the protein complex run. While the fractions from the first peak contain both proteins (left-hand side of the protein gel), the fractions from the second peak comprise the TNKS2 ARC4 in excess (right-hand side of the protein gel).

The complex was concentrated to 28.0 mg/mL and crystal screening was performed by dispensing the protein sample at 1:1 and 2:1 protein:buffer dilutions on the same set of crystallisation plates as before. Compared to the 1TBP-CTPR4-SolvH crystal plates, a higher rate of precipitation was observed in these wells; however no crystals could be detected.

4.3 Discussion

In this Chapter, the biophysical properties of the engineered CTPR proteins were characterised. As previously observed with other grafted loop peptides,[80,146] the grafted TBP did not affect the overall fold of the CTPR repeats, and by increasing the number of repeats a stabilising effect was obtained. To assess their binding properties, ITC experiments were performed. To simplify the analysis, a TNKS2 construct with only one binding site for TBP was used (TNKS2 ARC4). The study had indeed the purpose to evaluate the binding capacity of the grafted loop,

when present in one or more copies. The results obtained from the ITC studies confirmed our hypothesis that grafting the TNKS-binding sequence between consecutive repeats with sufficient flanking "linker" residues enables this peptide to maintain its binding-competent conformation. This is a noteworthy result, considering that the TBP loop binds to the ARC4 subdomain in an extended conformation,[85] and loop grafting might impose a structural constriction on the TBP that would weaken its binding affinity. This constriction, however, might explain the lower affinity of binding observed for the grafted TBP, compared to its peptide counterpart, to the same target. Altogether, this result consolidates our confidence in the broad applicability of this grafting approach, as we believe it can be applied to display a great variety of non-helical SLiMs in very different binding conformations. Thus, the Keap1-binding peptide from Nrf2 adopts a turn-like conformation that appears ideal for grafting onto the CTPR inter-repeat loop,[82] but peptides binding in an extended conformation, such as the TBP, can also be successfully grafted.

The ITC studies in the presence of a multivalent $_n$TBP-CTPR$_{2n}$ construct where $n>1$ showed that all the TBP loops are accessible for binding to TNKS2 ARC4. This result lays the foundation for the development of multivalent or multispecific CTPR arrays. This result has important implications: not only can multivalency be achieved by displaying multiple copies of a binding peptide for the same target molecule, but also different targets could potentially be recruited on a hetero-bifunctional CTPR construct displaying different binding peptides (Chapter 9). A deeper insight into the binding properties of the TBP-CTPR constructs is therefore explored in Chapters 5, 6 and 9.

Lastly, we aimed to determine the structures at atomic resolution using X-ray crystallography. Unfortunately, no crystals were formed. As it has been observed in the crystal lattice of previously crystallised CTPR constructs, crystal formation requires the tight packing of CTPR superhelices in all directions.[64,66] The presence of the TBP loop might disrupt the lattice formation, thereby preventing crystal formation.

Chapter 5

Intracellular interaction between $_n$TBP-CTPR$_{2n}$ and TNKS2

5.1 Introduction

The ITC experiments described in Chapter 4 confirmed the binding interaction between the $_n$TBP-CTPR$_{2n}$ constructs and TNKS2 ARC4. The first step towards functional cellular-based assay is to determine whether this interaction is maintained in a cellular environment, where many other factors, such as the presence of TNKS substrates with which the grafted loop needs to effectively compete, TNKS distribution in the cell and different microenvironments come into play. To answer this question, a cell-based assay was established, as described in this Chapter.

5.1.1 Experimental design

The simplest and fastest method to verify intracellular protein-protein interactions is by immunoprecipitation (IP) under native conditions. IP relies on the isolation of an antigen (here TNKS) from the cellular lysate through its binding to a protein bait ($_n$TBP-CTPR$_{2n}$), itself bound to an affinity resin. In our system, the $_n$TBP-CTPR$_{2n}$ bait is expressed with a C-terminal HA tag, which allows sedimentation of the protein-protein complex of interest upon recognition by an anti-HA antibody coupled to a matrix. Following a series of sample washes to reduce unspecific protein-protein interactions, the composition of the protein complex on the surface of the matrix can be analysed by SDS-PAGE electrophoresis or, more specifically, by Western Blot using antibodies recognising the bait and the target of interest.

A series of considerations were assessed in order to apply this assay to our system. First, the method requires the presence of both the protein bait (HA-tagged $_n$TBP-CTPR$_{2n}$) and

its substrate (TNKS1 and TNKS2) within a cellular environment. Following cell lysis, the protein complex can be pulled-down by binding to the anti-HA resin. HA-tagged $_n$TBP-CTPR$_{2n}$ can be expressed in the cell by transfection of the mammalian expression vector. Given the high transfection efficiency of HEK293T cells and their greater ease of *in vitro* culturing compared with other cell lines, these cells were selected for this experiment. Moreover, as the amount of endogenous, intracellular, soluble TNKS within HEK293T cells is unknown and might be insufficient for detection by Western Blot, a plasmid encoding the full-length TNKS2 was also transfected to enhance the intracellular TNKS2 protein levels and thereby facilitate its detection. To further facilitate TNKS2 detection following pull-down, the TNKS2 protein was expressed as a fusion protein with a different tag than the HA tag, which is already used for the bait protein. The need of a tag for TNKS2 is further motivated by the lack of highly specific anti-TNKS antibodies, as proven by our early attempts at Western Blot using a number of different commercial primary antibodies that failed to detect TNKS1/2 specifically (data not shown). To overcome this problem, the full-length TNKS2 gene was cloned in frame with an N-terminal HiBiT-tag in the pcDNA3.1(-) mammalian expression vector, for specific recognition.

5.1.2   The HiBiT tag

The HiBiT tag is an innovative, proprietary reporter tag developed by Promega. In 2012, the company published the development of a novel luciferase, named NanoLuc, smaller in size (19 kDa), more stable, ATP independent and >150-times brighter than traditional firefly and *Renilla* luciferases, the two most frequently adopted luciferases, among others. NanoLuc was derived from the catalytic domain of a deep-sea shrimp *Oplophorus gracilirostris* luciferase and engineered by structural optimisation with three mutagenesis rounds.[200,201] The NanoLuc substrate, coelenterazine, was also optimised into an improved analogue, called furimazine, which is more stable and generates a higher and longer-lasting luminescent signal.[200] As a result of these favourable properties, NanoLuc has become quickly the preferred tool in several luminescent-based biomedical research applications.[201]

The structure of NanoLuc consists of a 10-stranded ß-barrel and a "lid" domain composed of three α-helices (PDB 5IBO, Figure 5.1). Taking advantage of its small size, Promega proceeded with the development of the NanoLuc® Binary Technology (NanoBiT), a structural complementation reporter system relying on the split of the NanoLuc luciferase between the 9th and 10th strand of the ß-barrel. NanoLuc is therefore split into a larger (LgBiT)

and a smaller (SmBiT) fragment, that bind to each other with very low binding affinity.[202] The two complementing fragments were further evolved for improved solubility, function and binding affinity. HiBiT corresponds to the evolved, smaller fragment that binds the larger LgBiT with much higher affinity ($K_d = 0.7 \times 10^{-9}$ M) than the natural peptide ($K_d = 0.9 \times 10^{-6}$ M).[203] Upon binding to LgBiT, HiBiT complements the ß-barrel to reconstitute the full-length and functional luciferase (Figure 5.1).
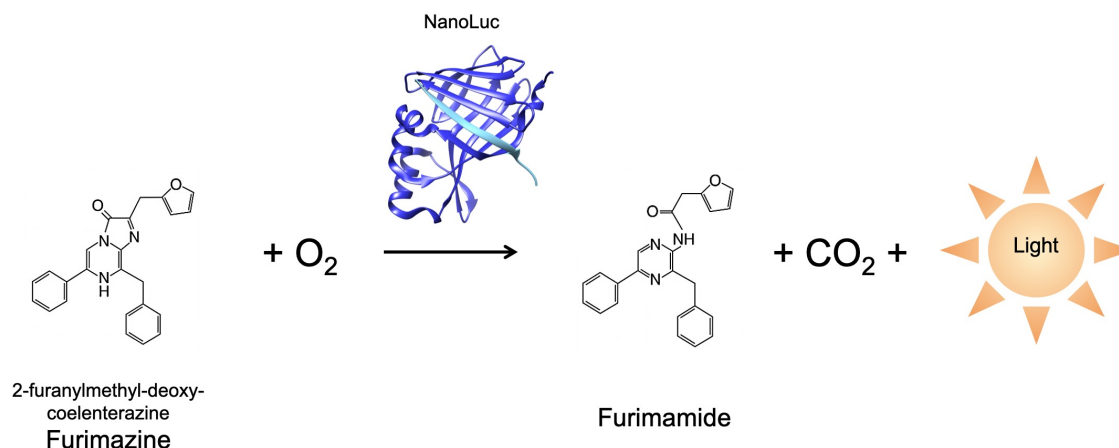
Figure 5.1: Schematic representation of the NanoLuc enzymatic reaction. NanoLuc converts furimazine in the presence of O_2 into furimamide and releases CO_2 and light. The structure of the NanoLuc enzyme is shown (PDB 5IBO), with the light blue ß-strand corresponding to the HiBiT sequence. The NanoLuc image was generated using the software UCSF Chimera.

The HiBiT tag can therefore be used as a small, 11-amino acid (VSGWRLFKKIS) peptide tag, that is unstructured in solution. Given its small size, the risk of its interference with a protein's normal function and structure is minimised, unlike for other bulky reporter tags. For the same reason, the HiBiT tag is the ideal candidate for CRISPR/Cas9 genome editing of endogenous proteins, as it can provide a higher insertion rate than other larger reporters.[203] When HiBiT, the large LgBiT subunit and the furimazine substrate are present, the reconstituted NanoLuc emits a luminescence signal proportional to its protein amount.[203] The HiBiT tag, therefore, provides an unambiguous and quantitative method of correlating the luminescent signal with the concentration of the HiBiT fusion. Moreover, bioluminescence is more sensitive than fluorophore-based methods, therefore allowing the detection of extremely low protein levels (up to 1 amol).[203] Therefore, the split-NanoLuc luciferase technology represents an easy and sensitive alternative to detect and quantify the amount of HiBiT-tagged protein.[203] Once the

NanoLuc is reconstituted, the luminescence readout can be measured using a microplate luminometer (under the appropriate settings) or by HiBiT blotting. The microplate reader allows extremely fast, quantitative and high-throughput measurements, as multiple samples can be processed within a few minutes. HiBiT blotting is, instead, an innovative Promega technology similar to Western Blot, where the large NanoLuc subunit replaces the primary and the secondary antibodies for the HiBiT tag recognition on a transfer membrane and generates a luminescent signal once bound to it.

5.1.3 HiBiT-qIP method

By combining the IP method with the split-NanoLuc technology, I developed a new method, named "HiBiT-based quantitative immunoprecipitation (HiBiT-qIP)", which allows a fast, high-throughput, quantitative (q) and highly specific detection of any HiBiT-tagged protein of interest throughout each stage of the IP protocol, without the need of antibodies and laborious Western Blot procedures. Here, the method allowed the real-time detection of the HiBiT-tagged TNKS2 construct throughout each step of the IP. As shown in Section 5.2.3, the HiBiT-tagged protein could be detected in the cell lysate, in the unbound and washes samples, as well as onto the resin, therefore circumventing the need of eluting the protein complex from the beads. A schematic representation of the constructs designed for this assay and the HiBiT-qIP method is provided in Figure 5.2A and B, respectively. A similar HiBiT-qIP method has been also developed by Ranawakage *et al.*, but for a different application, as it is used for antibody affinity determination in the presence of purified, HiBiT-tagged epitope.[204]
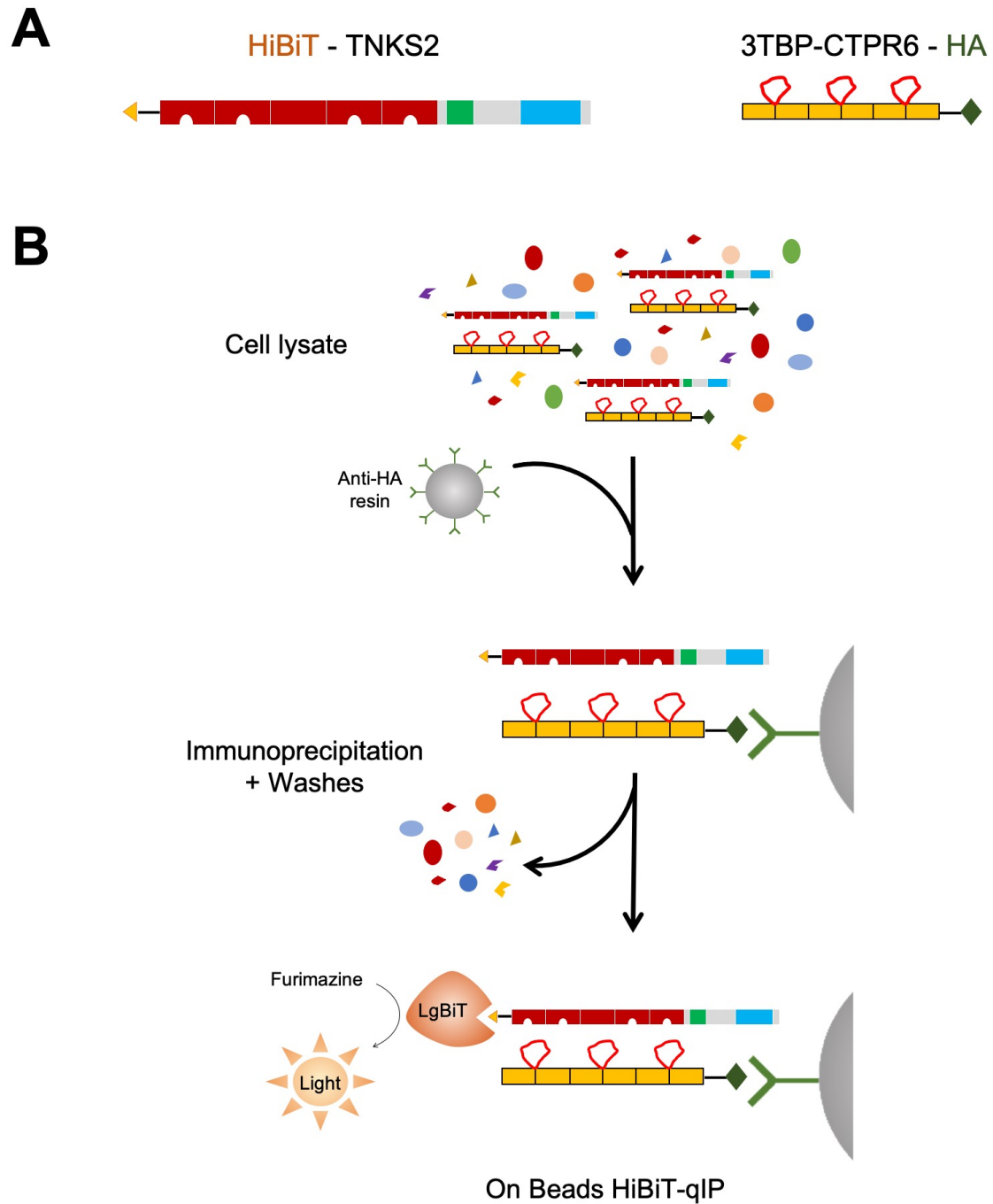
Figure 5.2: Schematic representation of the HiBiT-qIP method. (A) Schematic representation of the two constructs transfected into HEK293T cells, N-terminal HiBiT-tagged TNKS2 and C-terminal HA-tagged 3TBP-CTPR6. (B) In the HiBiT-qIP method, the HEK293T cell lysate is incubated with anti-HA resin to immunoprecipitate the HA-tagged bait. Washes allow the clearance of all the other proteins not binding to the resin. The presence of HiBiT-tagged TNKS2 is detected directly on beads by measuring the luminescence generated by the reconstituted NanoLuc luciferase.

5.2 Results

5.2.1 Molecular biology using the pcDNA3.1(-) vector

The immunoprecipitation method described above requires transfection into HEK293T cells of a mammalian expression vector encoding for the protein baits of interest fused with the appropriate epitope tag. For this purpose, the in-house variant of the commercially available pcDNA3.1(-) vector (Thermo Fisher) was chosen, which allows C-terminal HA tagging of recombinant proteins. For the immunoprecipitation assay, the 3TBP-CTPR6 and CTPR6 constructs were subcloned into this vector.

5.2.2 Molecular biology using the HiBiT-pcDNA3.1(-) vector

A variant of the mammalian expression vector pcDNA3.1(-) including a N-terminal HiBiT tag, named HiBiT-pcDNA3.1(-), was initially generated to facilitate the production of HiBiT-tagged TNKS2 gene constructs. Round-the-Horn site-directed mutagenesis (as described in Section 2.3.4) was, therefore, applied to insert the HiBiT encoding DNA sequence immediately downstream the ATG start codon. Moreover, as recommended in the Promega Nano-Glo HiBiT Lytic Detection System Technical Manual,[145] the N-terminal HiBiT tag was designed to be followed by a 6-amino acid linker region (GSSGGS), providing flexibility and distance between the HiBiT tag and the protein of interest, to facilitate accessibility and recognition between HiBiT and LgBiT. The newly generated vector was named HiBiT-pcDNA3.1(-). As for the parental pcDNA3.1(-) vector, conventional restriction digestion and ligation was adopted using the unique restriction sites BamHI and EcoRI. Due to the size of the full-length TNKS2-encoding gene (3507 bp) and the presence of additional EcoRI sites within the DNA sequence, cloning was particularly challenging. After few attempts, the TNKS2-encoding gene was successfully fused in frame with the HiBiT tag and the linker region. Finally, as the HiBiT-pcDNA3.1(-) vector also induces C-terminal HA tagging, an additional Round-the-Horn site-directed mutagenesis step was required to introduce a stop codon immediately upstream of the HA tag sequence. The presence of the HA tag on TNKS2 would otherwise compromise the correct outcome of the immunoprecipitation procedure, which relies on the HA-tagged bait for pull-down.

5.2.3 nTBP-CTPR$_{2n}$ proteins and TNKS2 interact in the cell

Intracellular complex formation between HiBiT-tagged TNKS2 and nTBP-CTPR$_{2n}$ was confirmed using the HiBiT-qIP method described above. HA-tagged 3TBP-CTPR6 (chosen as representative of the linear nTBP-CTPR$_{2n}$ array) was used as bait to pull-down the HiBiT-

tagged TNKS2 using an Anti-HA resin. In the control samples, CTPR6 or the empty pcDNA3.1(-) vector (expressing the HA tag only) were used instead. HEK293T cells were co-transfected with two plasmids, encoding the HiBiT-tagged TNKS2 and one of the bait proteins. After 48 h incubation, the presence of HiBiT-tagged TNKS2 was followed and quantified throughout the IP process, by measuring the luminescence signal generated by the reconstituted NanoLuc luciferase. At each stage of the IP protocol, a small aliquot was saved for measurement on the microplate reader, allowing the real-time detection of the pulled-down target. With a conventional IP method, in contrast, the user has to wait until the Western Blot membrane is developed (1-2 days) before knowing the outcome of the assay. The split-NanoLuc technology, therefore, not only simplifies the analysis process, but also allows a faster and more quantitative measurement than Western Blot.

Similar levels of HiBiT-tagged TNKS2 were observed in the cell lysate, as well as in the unbound fraction of all the three samples. The amount of HiBiT-tagged TNKS2 then progressively reduced at each washing step and a luminesce value corresponding to background noise could be measured in wash 4 and wash 5, confirming that the resin had been thoroughly washed. Elution of the HA-tagged bait from the resin failed after incubation with the HA peptide. As an alternative, a small amount of resin was incubated with the Nano-Glo HiBiT Lytic Detection System buffer containing LgBiT and the substrate. By doing so, the measured luminesce readout is generated by the HiBiT-tagged TNKS2 still bound to the resin through the bait. The luminescent signal obtained from the resin samples confirmed that HiBiT-tagged TNKS2 was pulled-down only by 3TBP-CTPR6, and not by the control CTPR6 construct and the HA peptide expressed by the empty vector. This evidence is in agreement with the previous ITC binding studies and, moreover, it confirms that the grafted TBP loop mediates the binding to full-length TNKS2 within the cellular environment (Figure 5.3). Moreover, non-specific binding of HiBiT-tagged TNKS2 to the Anti-HA Agarose resin was not detectable, as the luminesce readout of the resin for the two control samples is comparable to background noise (Figure 5.3). This result confirms that the pulled-down protein complex can be detected directly onto the resin when using the new HiBiT-qIP method.
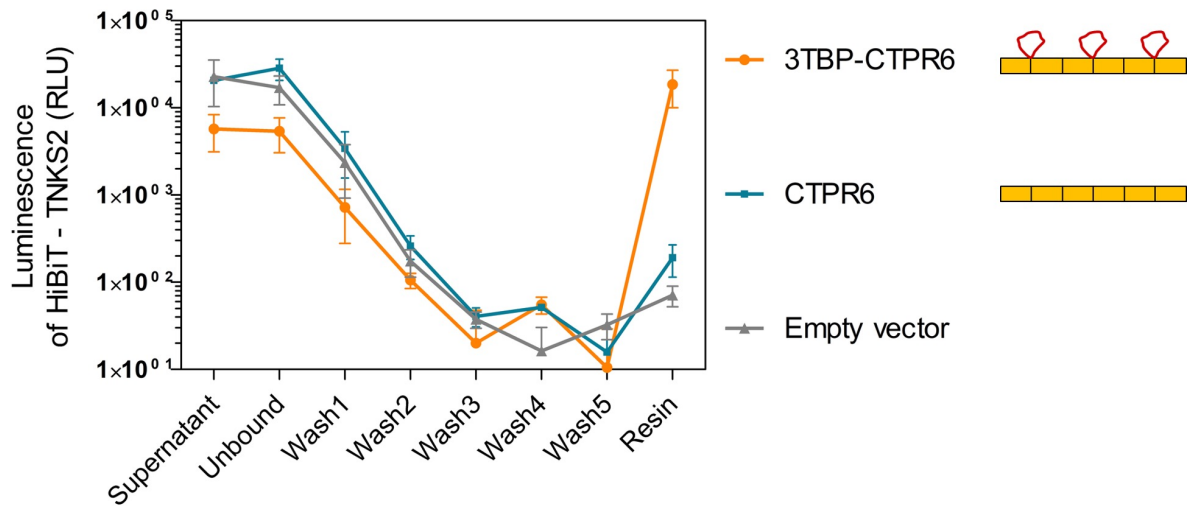
Figure 5.3: HiBiT-tagged full-length TNKS2 protein levels measured by luminescence intensity throughout each step of the HiBiT-qIP method. IP was performed in the presence of HA-tagged 3TBP-CTPR6, CTPR6 or an empty vector expressing the HA tag only. Averaged values and error bars corresponding to the standard deviation from two independent biological replicates are shown.

Elution of the proteins bound to the resin was then achieved by adding loading dye containing SDS. The HiBiT-qIP results were further validated by Western Blot and Nano-Glo HiBiT blotting (Figure 5.4). The transfer membrane was cut into three parts to evaluate the presence of tubulin and HA-tagged baits by traditional Western Blot and HiBiT-tagged TNKS2 by Nano-Glo HiBiT blotting. The detection of tubulin was used to confirm that the initial cellular content (SN of the cell lysate) is comparable between the treatments. Unexpectedly, tubulin was also present in the elution samples, suggesting that it binds non-specifically to the resin, and mostly to the one incubated with "Empty vector" cell lysate having no bait protein occupying the surface. An anti-HA primary antibody was instead used to verify the correct expression of the bait protein and its binding to the Anti-HA resin. The anti-HA blot shows that the recombinant protein amounts of HA-tagged 3TBP-CTPR6 are much higher than those of the HA-tagged CTPR6 construct (consistent with other results shown in Chapter 7). Upon binding to the resin, both proteins get concentrated on its surface, explaining the higher intensity of the two bands in the "Elution" samples. Finally, Nano-Glo HiBiT blotting proved that HiBiT-tagged TNKS2 was correctly expressed as a fusion protein of 129 kDa in size. It was also expressed in similar amounts in all the samples tested, less in 3TBP-CTPR6 as observed also in Figure 5.3. Nano-Glo HiBiT blotting further confirmed that HiBiT-tagged TNKS2 was pulled-down only when 3TBP-CTPR6 was used as bait (Figure 5.4).
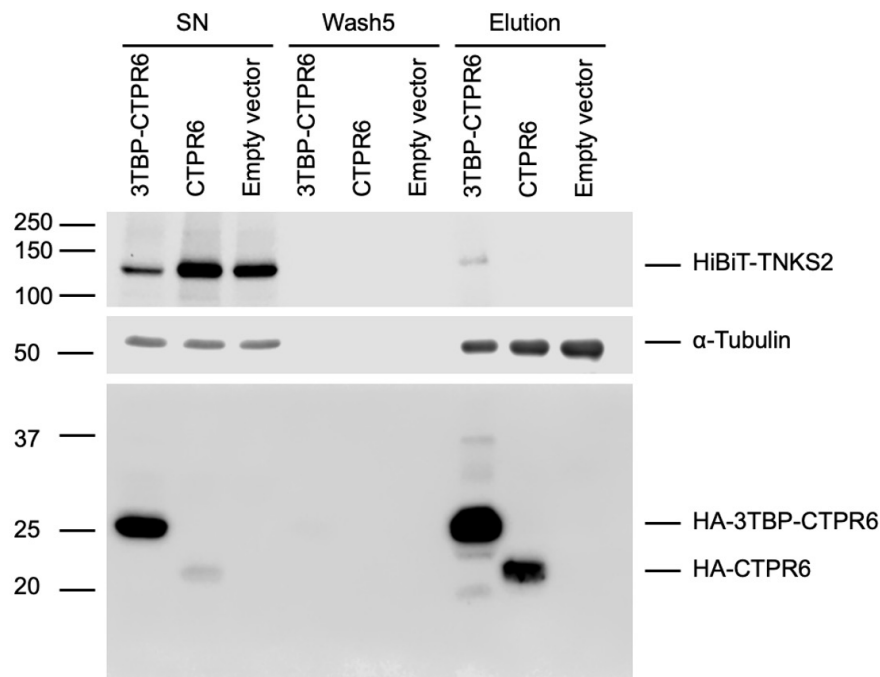
Figure 5.4: Complementary HiBiT-qIP analysis performed by Nano-Glo HiBiT blotting and conventional Western Blot on the samples in Figure 5.3. For each treatment, the supernatant (SN), the final wash (Wash5) and the Elution samples are shown. All blots were obtained from the same transfer membrane, divided into three parts as indicated.

5.3   Discussion

In this Chapter, the development of an innovative and antibody-free HiBiT-qIP method was described. The method combines for the first time the IP process with the HiBiT split-NanoLuc technology. This technique is particularly useful to detect any protein target for which a specific antibody is still unavailable, as is the case for TNKS1 and TNKS2. The only requirement is the addition of a HiBiT tag at one or the other end of the protein of interest, through conventional molecular biology applications or by CRISPR/Cas9. In the latter, the endogenous protein is tagged, to achieve a more accurate measurement and preventing transfection in difficult-to-work cell lines. The HiBiT-qIP technique allowed the real-time tracking of the HiBiT-tagged TNKS2 construct throughout the IP process, when it was present in solution, as well as bound to the Anti-HA resin. The method confirmed that the TBP loop within the CTPR scaffold can bind to full-length TNKS2 in the cellular environment. Until now and in the literature, the binding of the TBP peptide to TNKS proteins had only been assessed using purified proteins and with a small fragment of TNKS2 (TNKS2 ARC4).[46,85] The

results obtained here, therefore, extend the TBP binding properties to the full-length TNKS and in a cellular environment. Moreover, given the high sequence homology between TNKS1 and TNKS2, it is reasonable to assume that the $_n$TBP-CTPR$_{2n}$ constructs can interact with both TNKS homologue proteins.

Western Blot and Nano-Glo HiBiT-blotting were also performed in support of the HiBiT-qIP results. The Nano-Glo HiBiT-blotting allows the sensitive detection of HiBiT-tagged proteins separated according to their molecular weight. It therefore confirmed that the complex was formed between 3TBP-CTPR6 and full-length HiBiT-tagged TNKS2, and that no TNKS2 degradation products are present. Despite transfecting cells with identical amounts of plasmid DNA, different levels of 3TBP-CTPR6 and CTPR6 protein were observed in the samples, according to the anti-HA blot. The same variability was also observed in other experiments described in Chapter 7, and a potential explanation for this observation is provided.

To conclude, the results obtained by HiBiT-qIP set the basis for investigating in greater detail the binding of the engineered TBP-CTPR proteins to TNKS in the cell (Chapter 6), as well as the functional effects of the TBP-CTPR proteins on endogenous TNKS activity in the cell, as described in Chapter 7.

Chapter 6

Multivalency induces the formation of large macromolecular assemblies

6.1   Introduction

In biology, multivalency plays a great role in diverse processes: from extracellular carbohydrate-lectin and antibody-receptor binding, to various intracellular signalling event, RNA metabolism and chromatin organisation, multivalency orchestrate the formation of complexes with specific properties and functional outcomes.[4] Multivalency has been mostly studied in the contest of extracellular ligands. More recently, the effects of multivalency on intracellular proteins has been explored, where interacting systems are often characterised by modest affinity but high valency, with binding sites connected by long and flexible linkers to facilitate the interaction. Of particular interest in recent years has been the role of multivalent interactions in the formation of cross-linked networks and liquid-liquid phase separation, as observed in so-called membraneless organelles or biomolecular condensates.[205–207] TNKS is itself a multivalent protein due to its multiple substrate-binding sites (the ARC domains) and indeed one of its substrates, Axin, engages with TNKS through two TNKS-binding peptides connected by a flexible linker.[144]

Looking more closely into the mechanism of binding, when two multivalent protein binders, both having more than one binding site, interact with each other in solution, they have the potential to engage according to the two modalities shown schematically in Figure 6.1. Scenario A represents avidity of binding, which can be achieved when multiple binding events

occur between the same two protein molecules. Avidity requires complementarity in terms of spacing and orientation of the multiple binding sites in the two interacting molecules. This binding mechanism is what characterises the 1:1 interaction between TNKS and Axin1[89] and in nature, as well as in drug development, is what allows to convert a low-affinity ligand into a high-avidity one. Scenario B represents the formation of a macromolecular complex being generated when both proteins interact with more than one binding partner and are both present in high concentration, leading to the formation of an extended, clustered network of molecules. This scenario has been found to be extremely frequent among biological interactions, in particular those involving multivalent proteins with intrinsically disordered regions.[208,209] These macromolecular structures have also been shown to undergo a collapse transition, where the interacting partners precipitate into a globule and convert into a membraneless organelle. This process is also known as liquid-liquid phase separation and can occur under specific conditions and concentrations.[210,211] This type of high-order assembly formation is not an undesired consequence of the failure to form 1:1 interactions as in Scenario A, but rather it enables proper functioning by cellular sub-compartmentalisation. New examples of functional liquid-liquid phase separations are discovered every year, but the problem of how to drug them remains unsolved.

6.1.1 Multivalent interactions between the CTPR constructs and TNKS

In our system, the linear and trimeric TNKS-binding CTPR constructs, as well as the target, are both multivalent (except for 1TBP-CTPR2) and they have the possibility to engage according to the two modalities of binding described above (Figure 6.1).
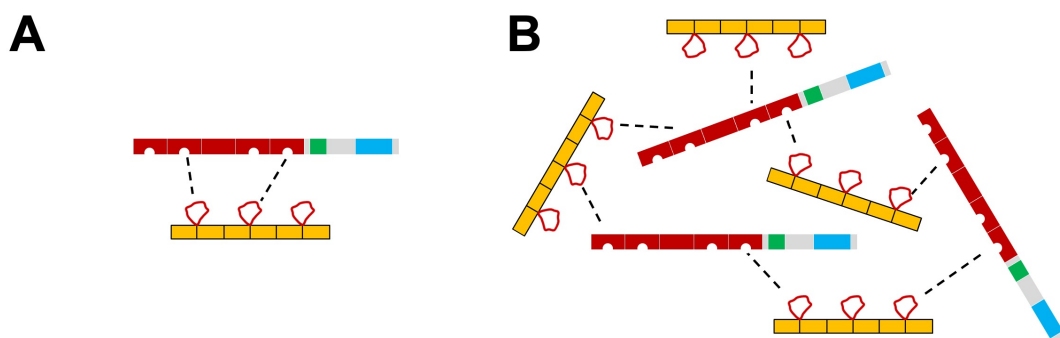
Figure 6.1: Different modalities of multivalent binding between full-length, multivalent TNKS2 and 3TBP-CTPR6 (chosen as representative construct of our multivalent array). (A) Schematic representation of avidity, with multiple binding events (dashed lines) occurring between the same two molecules (B) Schematic representation of a macromolecular complex being formed by the two proteins engaging simultaneously with multiple partner molecules.

The multivalent linear and trimeric CTPR constructs studied here were not designed to accurately match the multiple substrate-binding sites present on the ARC1-5 subdomains of TNKS. This is due to the fact that there is no high-resolution structure of the entire ARC1-5 domain of TNKS1 or TNKS2 to guide a structure-based design of a matching multivalent CTPR construct. Moreover, the conformation of the entire ARC1-5 domain is likely to be quite flexible, as proven by the SAXS data[89] and the impossibility of forming protein crystals. For this reason, the multivalent array described here was designed by simply joining multiple TBP-CTPR units to each other, rather than spacing the TBP loops according to the distance between TNKS ARC subdomains. Overall, it was thought as an initial proof-of-concept study for the design of multivalent CTPR constructs against a target that is itself multivalent. Although avidity of binding remains a possibility especially in the TBP-CTPR-foldon constructs, where several binding combinations are possible, we believe that the multivalent constructs are more likely to interact according to scenario in Figure 6.1B. We tested this hypothesis by using a number of different approaches, applied both in the test tube and in cell-based assays, and compared the behaviour of single- and multivalent CTPR constructs in the presence of multivalent TNKS constructs.

6.2   Results

6.2.1   Multivalent TNKS2 construct design and purification

The ITC experiments described in Chapter 4 confirmed the binding interaction between the $_n$TBP-CTPR$_{2n}$ constructs and the single-valent TNKS2 ARC4 subdomain. However, to study the effect of multivalency on the complex formation, a TNKS2 construct comprising more than one ARC subdomain had to be generated. For this purpose, two TNKS2 ARC1-5 constructs comprising all five consecutive ARC subdomains (residues 2-797) were cloned, with either a GST tag or a His$_6$ tag to aid protein purification. A His$_6$-tagged TNKS2 ARC1-5 construct has been successfully purified previously by Eisemann *et al.*[89] According to the large-scale protein expression test and in agreement with previous findings,[89] the His$_6$-tagged TNKS2 ARC1-5 construct was expressed in higher amounts compared to the GST-tagged counterpart (Figure 6.2). It was therefore decided to proceed with the purification of the His-tagged TNKS2 ARC1-5 construct, which has four binding interfaces (ARC1, ARC2, ARC4 and ARC5) for TBP (Figure 6.3A).
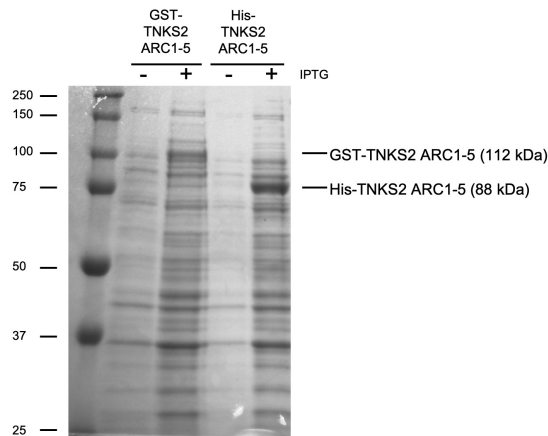
Figure 6.2: Large scale expression test of the GST- and His-tagged TNKS2 ARC1-5 constructs. SDS-PAGE protein gel with lanes corresponding to bacterial lysates before (-) and after (+) induction of protein expression by IPTG. The expression levels of the two indicated constructs are compared.

The first step of purification by HisTrap chromatography did not effectively purify TNKS2 ARC1-5, and many contaminants were still present in the eluted fractions (Figure 6.3B). Further purification steps by ion-exchange chromatography (unsuccessful, not shown) and gel filtration (Figure 6.3C) also failed to separate the protein of interest from the contaminants. In particular, two bands with a slightly lower molecular weight than His_6-TNKS2 ARC1-5 (about 55 kDa and 65 kDa) are visible throughout the purification steps and in the final samples eluted from gel filtration. Given the high band intensity of these proteins, it is likely that they correspond to degradation products of His_6-TNKS2 ARC1-5.
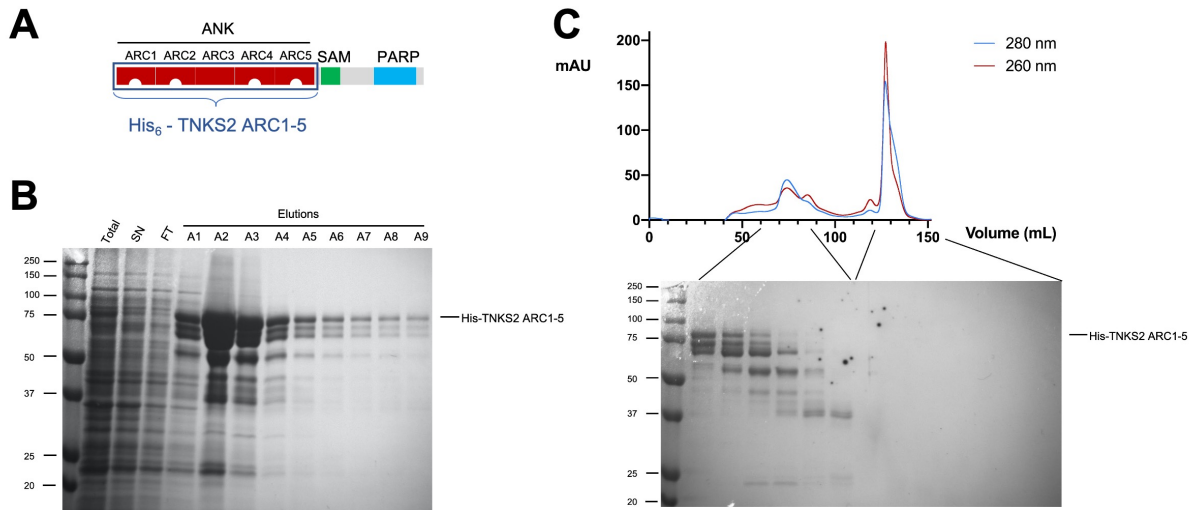
Figure 6.3: Purification of His$_6$-tagged TNKS2 ARC1-5. (A) Schematic representation of the His$_6$-tagged TNKS2 ARC1-5 construct compared to the full-length TNKS2 protein. (B) SDS-PAGE protein gel with lanes corresponding to different fractions collected throughout the HisTrap affinity chromatography. Total: total lysate, SN: supernatant, FT: flow-through. (C) Size-exclusion chromatogram of the His$_6$-tagged TNKS2 ARC1-5 over a Superdex 200 16/600 column and SDS-PAGE protein gel with lanes corresponding to fractions of the two eluted peaks.

To improve the purity and stability of the TNKS2 protein sample, a new multivalent TNKS2 protein construct was generated, corresponding to the first three ARC (ARC1-3) subdomains of TNKS2 (residues 2-485, 54.8 kDa) (Figure 6.4A). Although this construct has only two TBP binding interfaces (ARC1 and ARC2), which risks reducing the effect of multivalency in the complex formation, TNKS2 ARC1-3 has a more rigid overall structure than TNKS2 ARC1-5, as shown in the crystal structure and may be therefore less prone to proteolysis.[89] The higher flexibility of the ARC4-5 subdomains might explain why TNSK2 ARC1-5 appeared to be susceptible to proteolytic degradation resulting in the formation of degradation products, one of them with a molecular weight of about 55 kDa (Figure 6.3C) that is approximately the size of the ARC1-3 subdomain. The TNKS2 ARC1-3 construct was cloned with a N-terminal His$_6$ tag and purified by HisTrap chromatography and gel filtration (Figure 6.4B, C). The construct was expressed at a much higher yield than TNKS2 ARC1-5 and, at the end of the purification, pure protein was obtained, with only a lower-molecular weight contaminant present in some of the fractions. A multivalent TNKS2 protein construct was therefore successfully generated, allowing us to proceed with the binding analysis.
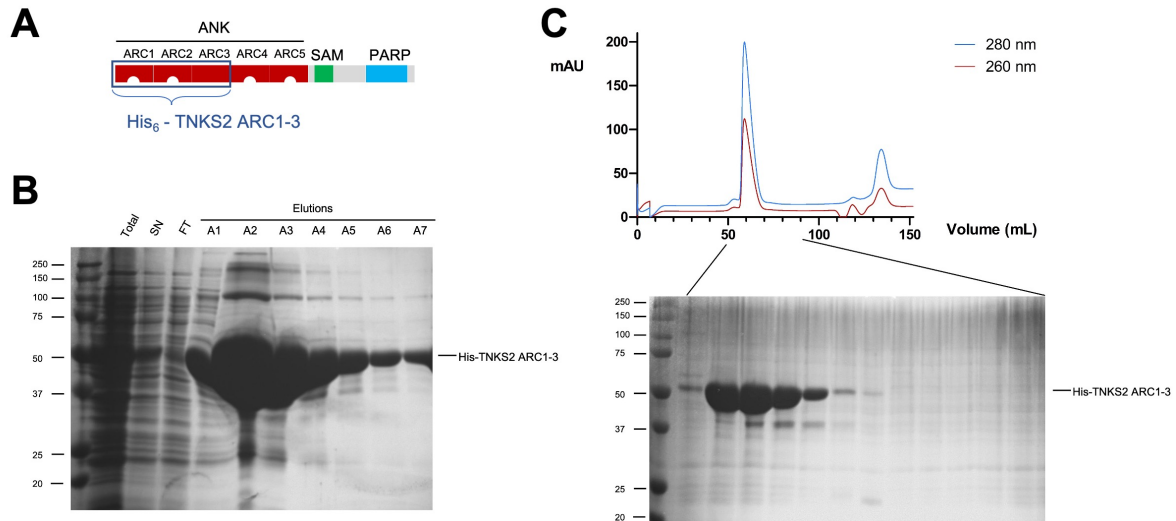
Figure 6.4: Purification of His$_6$-tagged TNKS2 ARC1-3. (A) Schematic representation of the His$_6$-tagged TNKS2 ARC1-3 construct compared to the full-length TNKS2 protein. (B) SDS-PAGE protein gel with lanes corresponding to different fractions collected throughout the HisTrap affinity chromatography. Total: total lysate, SN: supernatant, FT: flow-through. (C) Size-exclusion chromatogram of the His$_6$-tagged TNKS2 ARC1-3 over a Superdex 75 16/600 column and SDS-PAGE protein gel with lanes corresponding to fractions of the eluted peak.

6.2.2 Multivalent 3TBP-CTPR6 and 3TB-CTPR6-foldon induce the formation of large macromolecular complexes, studied by co-precipitation

To study the effect of multivalency on our protein system and verify whether the scenario in Figure 6.1B is favoured, a co-precipitation assay was established. The assay adopts a simple and straightforward method, relying on the co-precipitation induced by high gravitational force of the protein complex formed by the two interacting proteins. For this assay, the multivalent 3TBP-CTPR6 and 3TBP-CTPR6-foldon constructs, chosen as representatives of the linear and trimeric arrays, respectively, were incubated in the presence of TNKS2 ARC1-3. 1TBP-CTPR2 and CTPR6 were chosen as single-valent and non-binding controls for this assay. Following a 1 h incubation, samples were centrifuged at 20000 xg for 30 minutes using a tabletop centrifuge. Although much higher speeds are often applied by ultracentrifugation and analytical ultracentrifugation, a pellet was visibly formed using our experimental settings, avoiding the need of more sophisticated devices. The supernatant (S) and pellet (P) of each sample were then collected and run on a polyacrylamide gel to verify their composition (Figure 6.5).

Whereas all the CTPR constructs remained in the supernatant when incubated with co-precipitation buffer only, 3TBP-CTPR6 and 3TBP-CTPR6-foldon precipitated when mixed with TNKS2 ARC1-3 (Figure 6.5 top). Despite TNKS2 ARC1-3 self-precipitated to some

extent (about 50-70% of the total protein precipitates in buffer), the amount of precipitant increases in the presence of the multivalent 3TBP-CTPR6 and 3TBP-CTPR6-foldon, as the protein band intensity indicates. This result is the first evidence of macromolecular complex formation in solution mediated by multivalency. In contrast, the mono-valent 1TBP-CTPR2 construct and the control CTPR6 construct remained in the supernatant even when incubated with TNKS2 ARC1-3 (Figure 6.5 bottom and top, respectively), confirming that the multivalency of the TBP loops is required for the formation of a large supramolecular complex with multivalent TNKS2. It is also worth noting that, despite 3TBP-CTPR6 and 3TBP-CTPR6-foldon being incubated at the same TBP molar concentration, the trimeric 3TBP-CTPR6-foldon construct precipitated to a higher extent and led to an increased precipitation rate of TNKS2 ARC1-3, as the high protein band intensity in the pellet (P) of the 3TBP-CTPR6-foldon+TNKS2 ARC1-3 sample indicates (Figure 6.5 top).

Co-precipitation of 3TBP-CTPR6 was also observed in the presence of the higher-valency construct TNKS2 ARC1-5, although the many degradation products of this construct ruined the quality of the gel (data not shown).
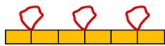
Additionally, co-precipitation assay was also performed by titrating increasing concentrations of 3TBP-CTPR6 or 3TBP-CTPR6-foldon proteins (used at the same TBP molar concentration) into TNKS2 ARC1-3. This experiment showed that the extent of co-precipitation was dependent on the molar ratio of the two interacting molecules, until saturation is reached. A quantitative analysis was performed by measuring the protein band intensities, which allowed to calculate the molar protein concentration of the species in each band. As the proteins in the pellet correspond to those forming the precipitated complex, the calculated molar protein concentration in the pellet provides an estimate of the molar composition of the macromolecular complex that has been first formed and then precipitated. For accurate analysis, the molar concentration of precipitated TNKS2 ARC1-3 was subtracted by the amount that self-precipitates. The values indicate a consistent 1:1 molar composition for the 3TBP-CTPR6:TNKS2 ARC1-3 complex, and a composition of 1:2 for the 3TBP-CTPR6-foldon:TNKS2 ARC1-3 complex, explaining why the trimeric 3TBP-CTPR6-foldon can induce a greater extent of TNKS2 ARC1-3 co-precipitation than the linear 3TBP-CTPR6 (Figure 6.6 and Table 6.1). Overall, it is clear from this result that not all TBP loops are involved in binding, as the scenario in Figure 6.1B indicates.
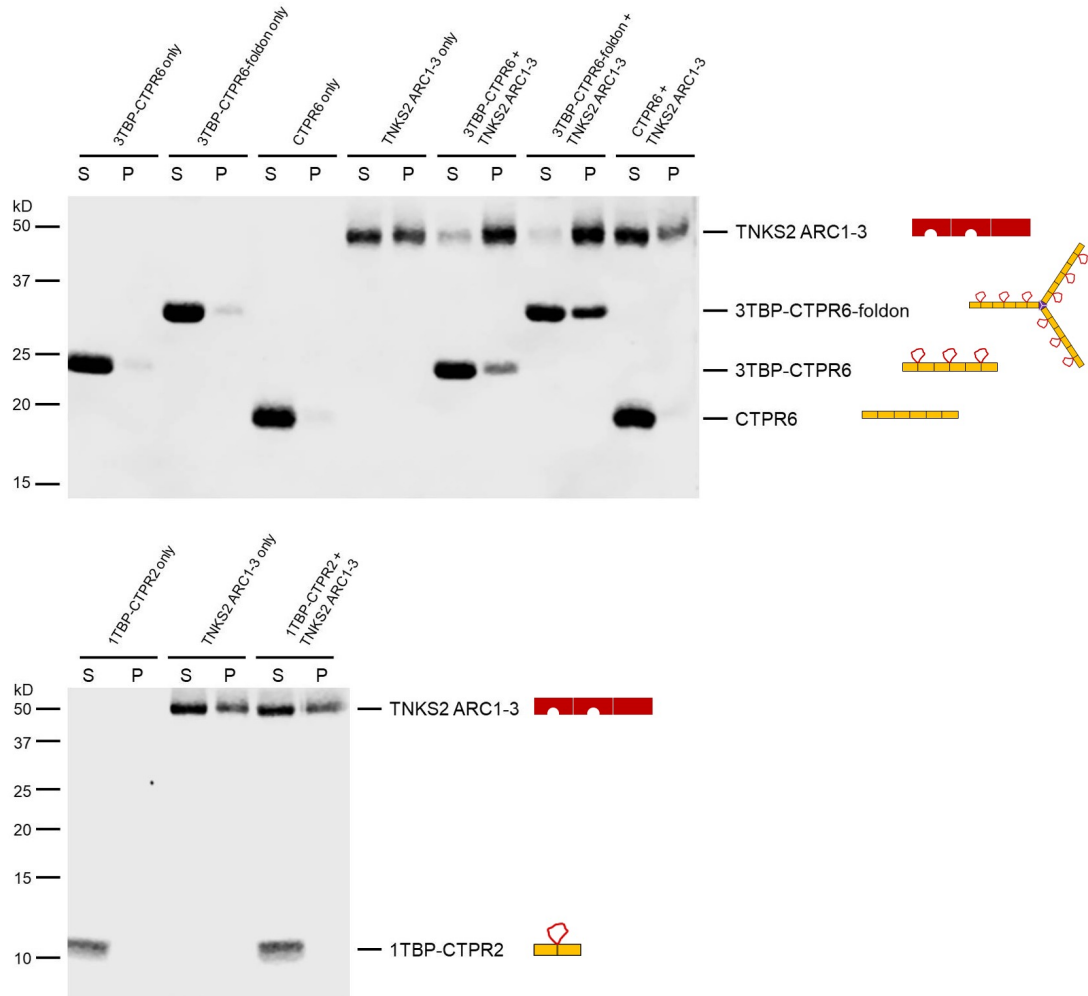
Figure 6.5: Multivalent interactions between TBP-CTPR proteins and TNKS2 ARC1-3 analysed by co-precipitation assay. After centrifugation, supernatant (S) and pellet (P) were separated and run on a 12% (top) or 16% (bottom) SDS polyacrylamide gel. For co-precipitation, proteins were mixed in equal volumes at the following concentration: 10 μM 1TBP-CTP2 (10 μM TBP loop concentration), 10 μM 3TBP-CTPR6 (30 μM TBP loop concentration), 3.3 μM 3TBP-CTPR6-foldon (30 μM TBP loop concentration), 10 μM CTPR6 and 10 μM TNKS2 ARC1-3 (20 μM TBP-binding sites). Gel images were obtained using Li-COR Odyssey Fc Imaging System.
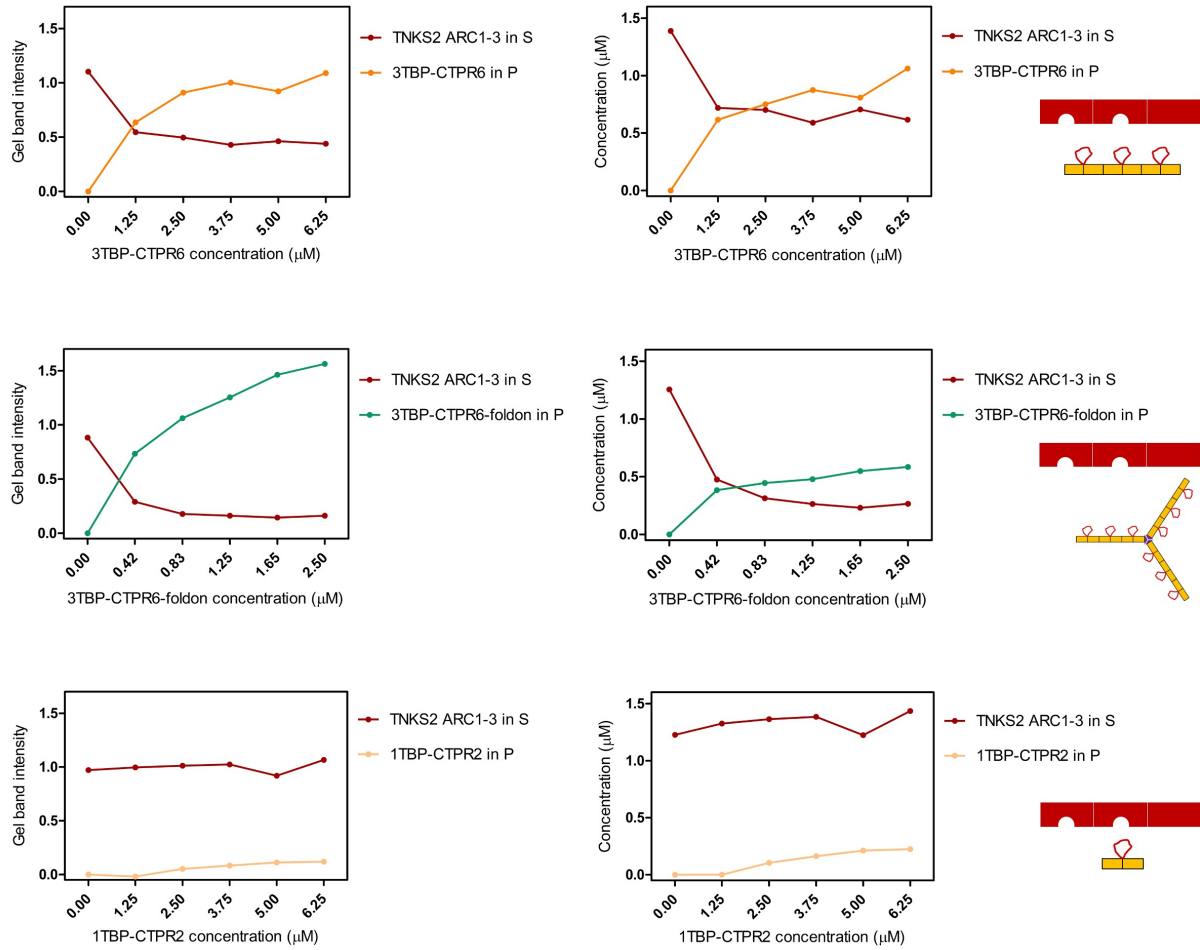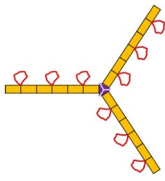
Figure 6.6: Co-precipitation experiments as a function of increasing CTPR protein concentration. Samples were prepared by mixing a constant amount of TNKS2 ARC1-3 (10 µM) with the indicated CTPR proteins at increasing concentrations (from 0 to 12.5 µM), in equal volumes. Therefore, the final protein concentration is halved. The trimeric 3TBP-CTPR6-foldon was prepared at a concentration three times lower the one of the linear 3TBP-CTPR6, for comparison purposes. After centrifugation, supernatants (S) and pellets (P) were separated and run on polyacrylamide gels. Gel band intensities were quantified and plotted on the left-hand side graphs. Knowing the initial protein concentration, the gel band intensities were converted into the corresponding protein concentration and plotted on the right-hand side graphs. Shown is a representative analysis of two independent experiments.

Initial 3TBP-CTPR6 (µM)	Initial TNKS2 ARC1-3 (µM)	3TBP-CTPR6 in P (µM)	TNKS2 ARC1-3 in P (µM)*	Ratio	
0.00	5.00	0.00	0.00	-	
1.25	5.00	0.62	0.67	0.92	
2.50	5.00	0.75	0.69	1.09	
3.75	5.00	0.88	0.80	1.10	**1:1**
5.00	5.00	0.81	0.68	1.18	
6.25	5.00	1.06	0.77	1.37	

Initial 3TBP-CTPR6-foldon (µM)	Initial TNKS2 ARC1-3 (µM)	3TBP-CTPR6-foldon in P (µM)	TNKS2 ARC1-3 in P (µM)*	Ratio	
0.00	5.00	0.00	0.00	-	
0.40	5.00	0.38	0.78	0.49	
0.83	5.00	0.45	0.94	0.47	
1.25	5.00	0.48	0.99	0.48	
1.65	5.00	0.55	1.02	0.54	**1:2**
2.50	5.00	0.58	0.99	0.59	

Table 6.1: Estimated molar composition of the assemblies formed between the multivalent CTPR constructs and the multivalent TNKS2 ARC1-3 protein. As the TNKS2 ARC1-3 protein showed some precipitation itself, the molar concentration of TNKS2 ARC1-3 in the pellet was corrected accordingly. The values obtained thus correspond to the amount of TBP-induced TNKS2 ARC1-3 precipitation. The stoichiometry (ratio) is calculated by dividing the concentration of the CTPR protein by the concentration of the TNKS2 ARC1-3 protein. *Corrected for the self-precipitation of TNKS2 ARC1-3.

6.2.3 Macromolecular assemblies visualised by negative stain-EM

Further characterisation of the macromolecular assemblies was performed by negative stain transmission electron microscopy (TEM). In agreement with the previous results, large assemblies, microns in size, could be observed only when both interacting partners are multivalent (TNKS ARC1-3 in combination with 3TBP-CTPR6 or 3TBP-CTPR6-foldon), and not in samples with 1TBP-CTPR2 or CTPR6 having no TNKS binding loops. Qualitatively, the clusters did not appear to have any particular structural organisation, but they display quite rounded edges. 3TBP-CTPR6 resulted in the formation of variable sized clusters throughout the TEM grid, ranging from small (1-2 µm clusters) to large networks many microns in size, whereas clusters formed by the trimeric 3TBP-CTPR6-foldon appeared to be predominantly

large, with smaller assemblies only occasionally visible. In contrast, samples of TNKS2 ARC1-3 in combination with 1TBP-CTPR2, CTPR6 or alone showed only some small amorphous aggregates (Figure 6.7).
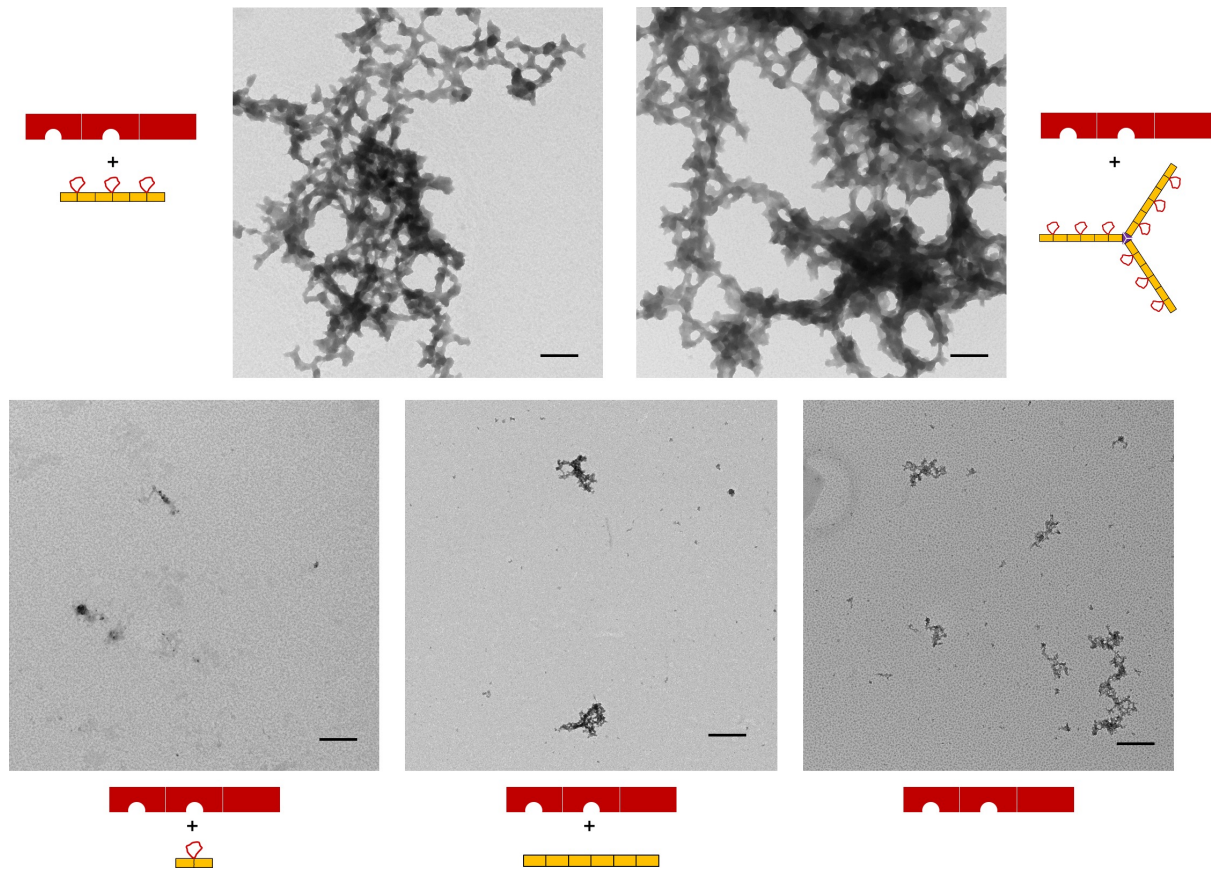
Figure 6.7: Negative stain TEM. TNKS2 ARC1-3 was incubated with 3TBP-CTPR6, 3TBP-CTPR6-foldon (Top), 1TBP-CTPR2, CTPR6 or buffer (Bottom) and imaged by negative stain TEM following a 1 h incubation. Proteins were mixed in equal volumes at the following concentrations: 5 μM 1TBP-CTPR2 (5 μM TBP loop concentration), 5 μM 3TBP-CTPR6 (15 μM TBP loop concentration), 1.7 μM 3TBP-CTPR6-foldon (15 μM TBP loop concentration), 5 μM CTPR6 and 5 μM TNKS2 ARC1-3 (10 μM TBP-binding sites). Scale bars for all images are 500 nm. Grid preparation and imaging was performed by Dr Janet R. Kumita, a post-doc in the Itzhaki lab.

6.2.4 Multivalent TBP-CTPR and TNKS constructs co-localise in the cell

Once the formation of a macromolecular complex was confirmed, we then wanted to investigate the effect of multivalency inside the cell. Fluorescence microscopy was therefore applied in a co-localisation assay. To this end, single- and multi-valent eGFP-tagged TNKS2 constructs were cloned in a mammalian expression vector, while the CTPR constructs were tagged with mCherry. These two fluorescent proteins have non-overlapping excitation and

emission spectra, allowing us to localise both signals without interference. For this assay, HEK293T cells were co-transfected with eGFP-tagged TNKS2 ARC1-5 in combination with a mCherry-tagged CTPR construct having one, multiple or no TNKS-binding loops. Although the eGFP-tagged full-length TNKS2 construct was also cloned, the TNKS2 ARC1-5 construct was chosen instead, as it was expressed in higher amounts, resulting in a brighter intracellular fluorescence signal (data not shown). After 48 h from transfection, the presence of large, polymeric species was visible in the cytoplasm of cells co-transfected with the multivalent TNKS2 ARC1-5 and the multivalent 3TBP-CTPR6 or 3TBP-CTPR6-foldon construct, where the fluorescent signal of eGFP and mCherry perfectly co-localised (Figure 6.8). In agreement with the previous results, formation of these large assemblies depends on the interaction between two high-valency molecules: when TNKS2 ARC1-5 was co-expressed with the mono-valent 1TBP-CTPR2 or CTPR6 instead, no such assemblies were produced and observed (Figure 6.8).

In addition, Figure 6.8 also indicates that 1TBP-CTPR2 and CTPR6 are present in the cytoplasm as well as in the nucleus of the cells, whereas 3TBP-CTPR6 and 3TBP-CTPR6-foldon are retained in the cytoplasm. This might be a result of their larger molecular weight making them unable to translocate into the nucleus or, more likely, the effect of their recruitment into large macromolecular assemblies.

Despite the large size of these macromolecular assemblies (up to 1-2 μm), any particular effect on the cell morphology and cell viability could be visibly observed under the microscope in cells containing the agglomerates.
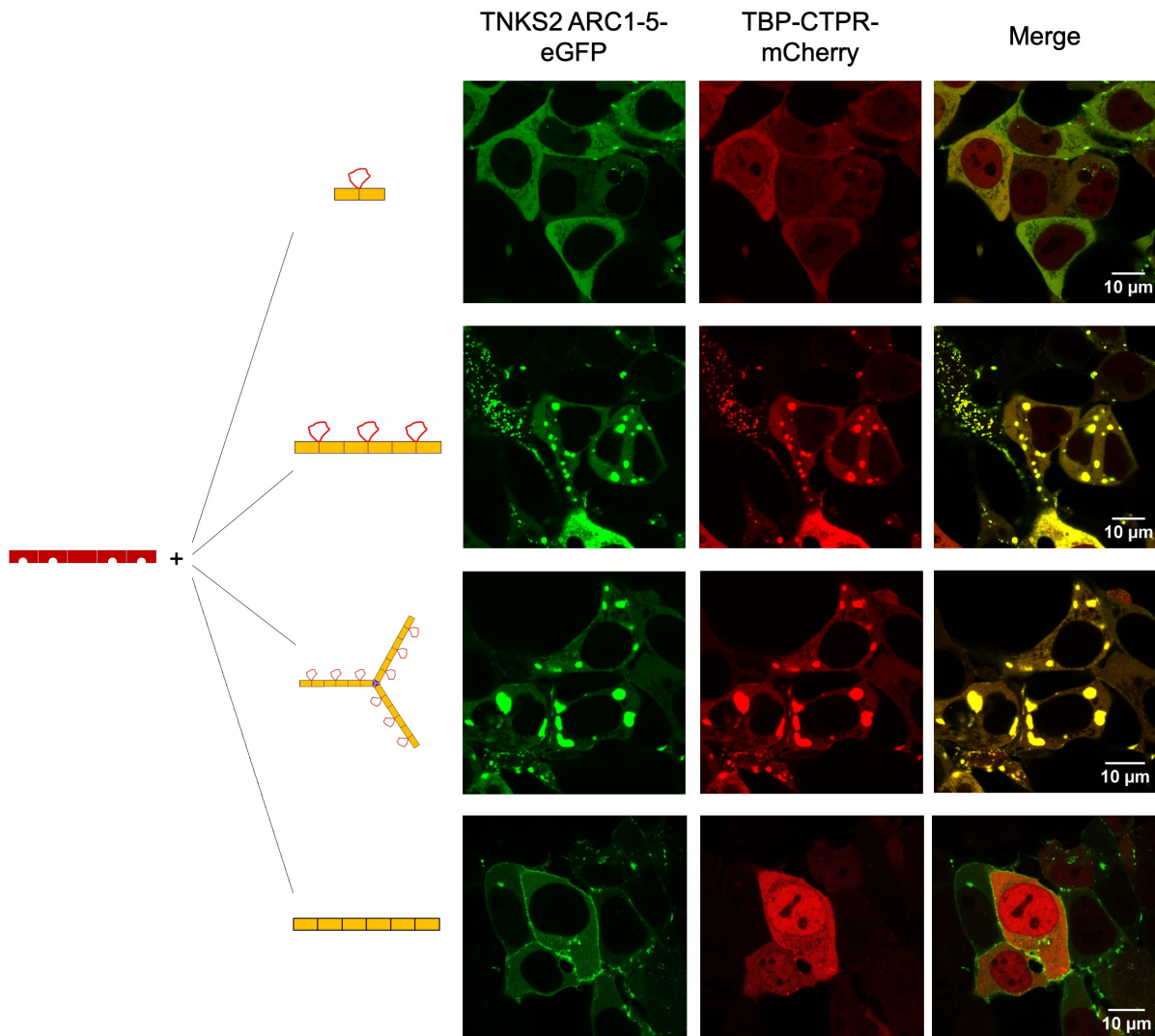
Figure 6.8: Fluorescence microscopy of HEK293T cells co-transfected with mCherry-tagged CTPR proteins and eGFP-tagged TNKS2 ARC1-5. Co-localisation in large macromolecular clusters is observed for eGFP-tagged TNKS2 ARC1-5 in combination with mCherry-tagged 3TBP-CTPR6 or with mCherry-tagged 3TBP-CTPR6-foldon. These clusters are not observed for the mono-valent mCherry-tagged 1TBP-CTPR2 protein or the control mCherry-tagged CTPR6 protein. Scale bars for all images are 10 μm.

Likewise, when the same CTPR proteins were co-expressed with a monovalent TNKS2 construct comprising only the first ARC domain (eGFP-TNKS2 ARC1), no polymeric species were visible in any of the combinations tested (Figure 6.9). With the macromolecular assemblies not being formed, 3TBP-CTPR6 was also localised inside the nucleus of the cells, similarly to 1TBP-CTPR2 and CTPR6. 3TBP-CTPR6-foldon was instead only present in the cytoplasm.
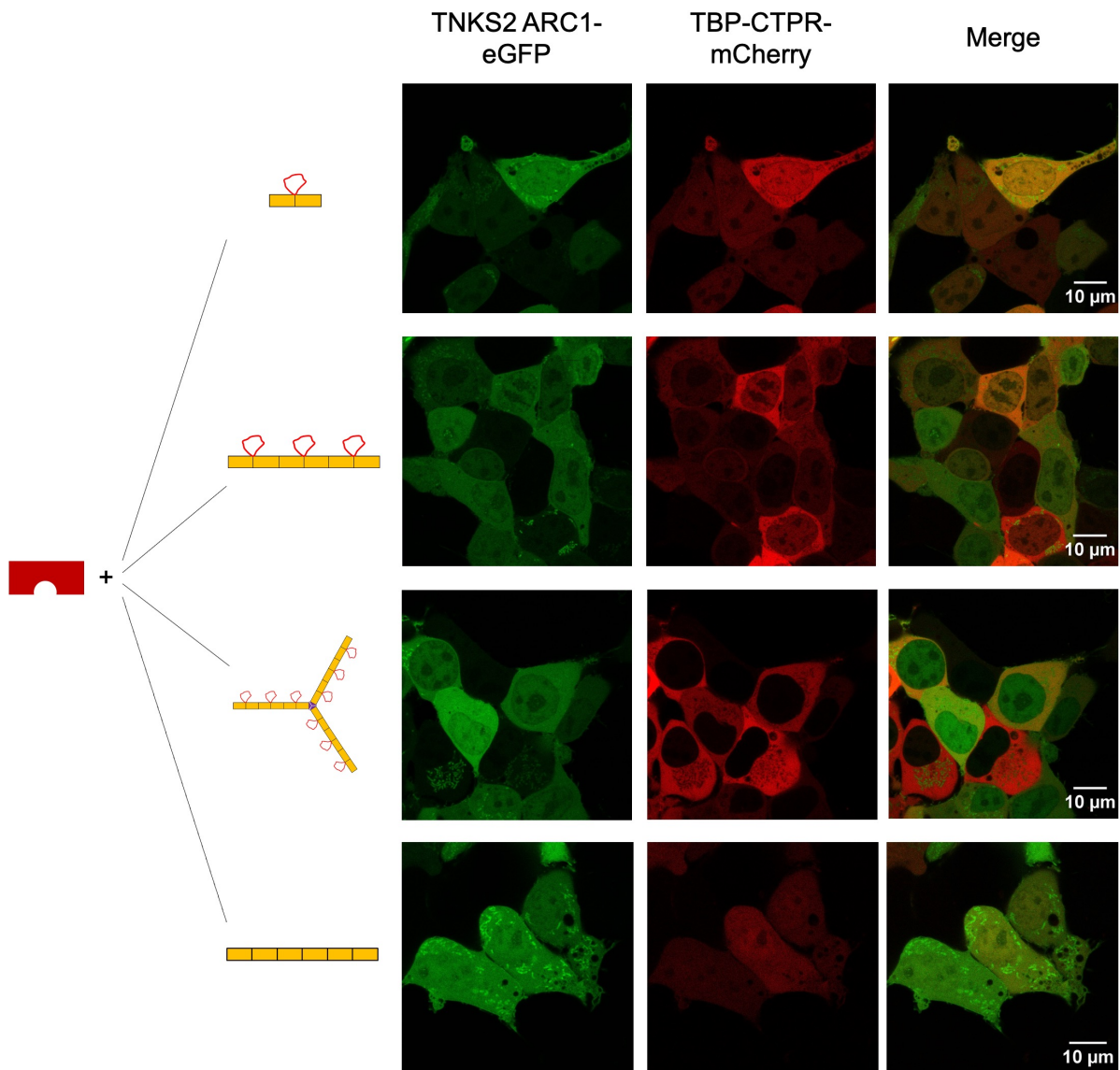
Figure 6.9: Fluorescence microscopy of HEK293T cells co-transfected with mCherry-tagged CTPR proteins and eGFP-tagged TNKS2 ARC1. The monovalent nature of the TNKS2 ARC1 construct prevents the formation of large macromolecular assemblies. Scale bars for all images are 10 μm.

6.2.5 Macromolecular assemblies have different internal dynamics

Having observed the formation of large assemblies inside the cell, Fluorescence Recovery After Photobleaching (FRAP) was used to assess their internal dynamics. FRAP relies on the observation of the fluorescent signal over time, before, during and after photobleaching a small area. The extent of fluorescence recovery provides an indication on the kinetics of diffusion of fluorescent molecules in living cells, using fluorescent microscopy.[212] FRAP was performed on those fluorescent macromolecular assemblies generated by eGFP-TNKS2 ARC1-5 in

complex with the mCherry-tagged linear 3TBP-CTPR6 or the trimeric 3TBP-CTPR6-foldon. For each sample, small Regions of Interest (ROI) were selected within the assemblies and in these the recovery rate of the eGFP signal was followed after its localised photobleaching (Figure 6.10). Macromolecular assemblies induced by the linear 3TBP-CTPR6 showed a rapid recovery rate, with signal reaching a plateau within 30 seconds after photobleaching. This evidence suggests that the proteins within this assembly are highly dynamic, with binding partners in rapid exchange. Fluctuations in the recovery rate were also observed due to the slight movement of the assembly around the selected ROI, providing further evidence of the dynamic nature of these assemblies over a short window of time. In contrast, the assemblies generated by the trimeric 3TBP-CTPR6-foldon did not show any sign of fluorescence recovery within the same experimental settings, even after one minute from photobleaching. In agreement with the co-precipitation assay and stoichiometry analysis, we believe that the trimeric conformation of the 3TBP-CTPR6-foldon molecule, together with its increased valency, allows this construct to engage with more target molecules and establish a more interconnected, rigid network which abolishes the internal dynamic of the complex.

6.3   Discussion

The TNKS-binding CTPR constructs developed here combine target specificity with multivalency, two features that have not yet been explored in previous drug development efforts against TNKS. Having previously confirmed the target-binding properties of the CTPR constructs, the effect of multivalency was then investigated using a variety of *in-vitro* and cell-based assays. The presence of multiple binding sites in both the interacting partners, TNKS and the designed CTPR proteins, manifests in the formation of large, intracellular macromolecular assemblies, both in the cell and in the test tube. The co-precipitation assay provided the first evidence of macromolecular complex formation only when multivalent TNKS2 ARC1-3 was incubated with a multivalent TNKS-binding CTPR construct. The same protein assemblies were also observed by negative stain TEM. The results from both assays indicate the formation of larger assemblies when 3TBP-CTPR6-foldon is used, compared to those induced by the linear 3TBP-CTPR6 construct.
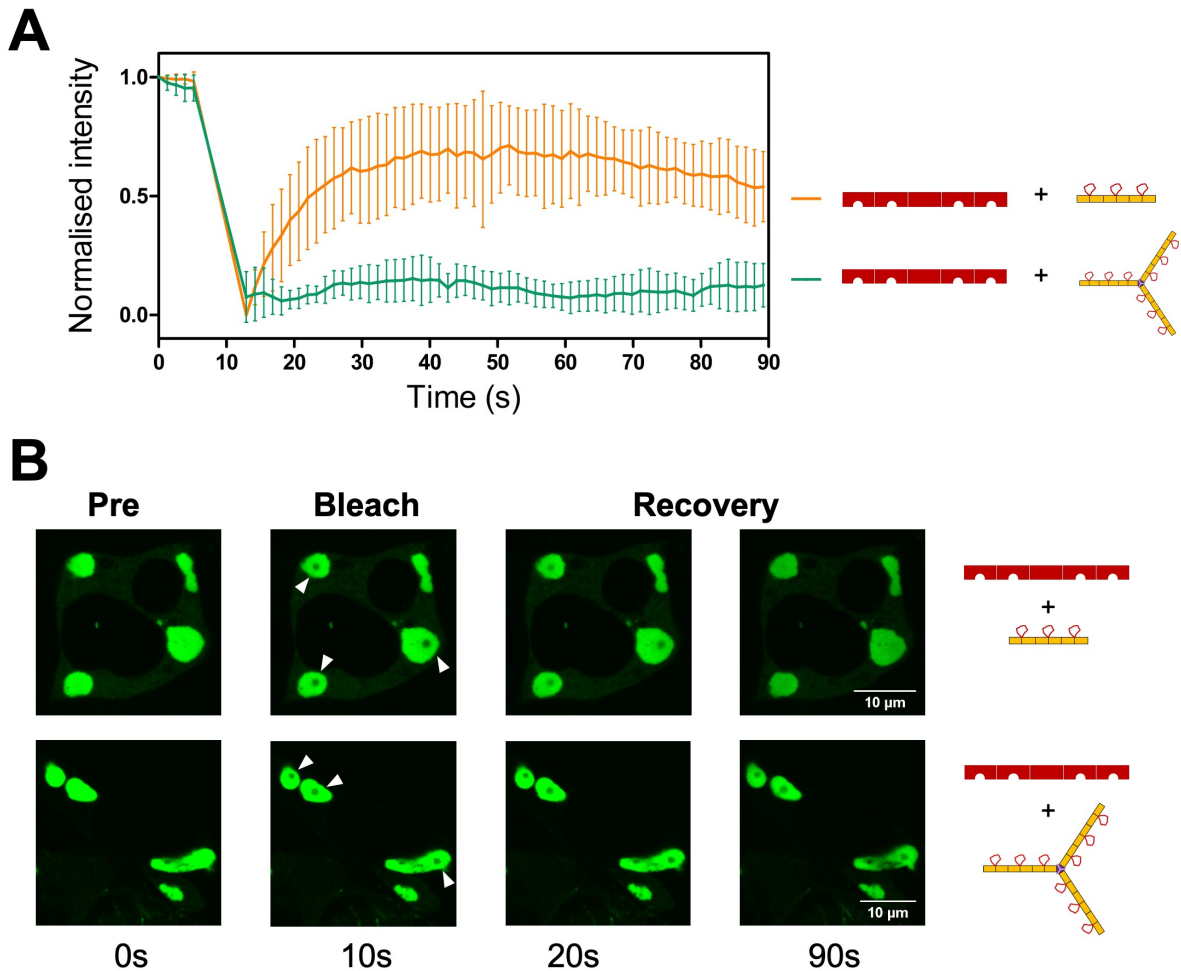
Figure 6.10: FRAP analysis of the assemblies formed between TNKS2 ARC1-5 and the indicated CTPR proteins. (A) FRAP analysis was performed on seven individual Regions of Interest (ROI) selected within macromolecular assemblies induced by eGFP-tagged TNKS2 ARC1-5 in complex with 3TBP-CTPR6 (orange) or 3TBP-CTPR6-foldon (green). HEK293T cells were bleached in the ROI and fluorescence recovery of eGFP-tagged TNKS2 ARC1-5 was monitored over 90 seconds. (B) Representative ROI selected within three independent macromolecular assemblies are shown before, during and post bleaching for both protein complexes.

A co-localisation assay demonstrated the formation of these macromolecular assemblies also in a cellular environment. Moreover, these assemblies showed different internal dynamics depending on the configuration of the CTPR construct (i.e. monomeric versus trimeric). The FRAP assay demonstrated how the macromolecular assemblies induced by the 3TBP-CTPR6-foldon construct are very rigid and interconnected, unlike those produced by the linear 3TBP-CTPR6. Interestingly, all the assays shown in this Chapter, together with the stoichiometry evaluation, point to the same conclusion: the 3TBP-CTPR6-foldon, due to its increased valency compared to 3TBP-CTPR6, is able to engage with several TNKS molecules, resulting in the formation of a larger and more interconnected network of proteins. Although these assays were

performed in the presence of TNKS2 constructs that are shorter than the full-length protein, the results shown here should still be representative of the binding interaction between the TBP-CTPR constructs and the endogenous TNKS. One final caveat is that here we over-expressed both the CTPR constructs and the TNKS construct. This will lead to higher protein levels that may enhance the formation of higher-order species. The functional effects that the macromolecular assemblies might have on TNKS activity should be further explored, and an initial investigation is provided in the following Chapter. It remains to be determined whether the observed intracellular macromolecular assemblies have undergone liquid-liquid phase separation and resemble biomolecular condensates. High valency and low binding affinities is what characterises most biomolecular condensates as well as the multivalent proteins under investigation here, leaving the door open to future studies.

Chapter 7

$_n$TBP-CTPR$_{2n}$ and $_n$TBP-CTPR$_{2n}$-foldon proteins inhibit Wnt signalling

7.1 Introduction

Most biologics and small molecule drugs available in the market are able to inhibit a protein target by simply binding to it. In most cases, the target:drug binding causes the inhibition of the target's catalytic activity or, less commonly, prevents the recognition of its natural substrate. As TNKS are clinically relevant targets, several small-molecule inhibitors of the TNKS PARP domain's active site have been developed and tested *in vitro* and, in some cases, in xenografts and/or genetically engineered mouse models of cancer.[135,138] Given the involvement of TNKS in the Wnt signalling pathway, one can measure the efficacy of TNKS inhibitors (TNKSi) by observing the alterations induced in this pathway. The most common method used to measure Wnt signalling is by TopFlash,[157] a Wnt-responsive luciferase reporter assay. The same assay was extensively adopted to screen for Wnt inhibitors[109,134–137] and to investigate the functional effect of small molecule TNKSi on this pathway.[138] In the Itzhaki group, the inhibitory effects of a series of macrocyclic, cell-penetrating TBP peptides was demonstrated using the same assay.[85] Inhibition of TNKS PARP activity leads to stabilisation of Axin. As Axin is a component of the β-catenin destruction complex, Axin stabilisation leads to the reduction of the cytoplasmic levels of β-catenin. This results in a reduced translocation of β-catenin into the nucleus and, therefore, in a decreased transcriptional activation of Wnt-controlled genes.

7.1.1 The TopFlash dual-luciferase reporter assay

The TopFlash assay relies on the luciferase activity of two different luciferases, expressed in cultured cells. Here, cells are co-transfected with a plasmid encoding luc+ firefly luciferase under the control of the TCF/LEF Wnt-dependent promoter and a second plasmid encoding the *Renilla* luciferase constitutively expressed from the CVM promoter. The *luc+* gene encodes a modified firefly (*Photinus pyralis*) luciferase, that has been optimised for improved expression and monitoring in transfected eukaryotic cells. *Renilla*, instead, is the wild-type luciferase from *Renilla reniformis* and is used as internal control to allow more accurate results. Wnt signalling is enhanced by adding purified Wnt protein or Wnt conditioned medium to the cell growth medium. Wnt conditioned medium is obtained by culturing cells expressing and releasing Wnt protein into their growth medium. Upon binding of the Wnt molecule to the Wnt-Frizzled receptor on the cellular surface, the canonical Wnt signalling pathway is activated, resulting in the accumulation of β-catenin and its translocation into the nucleus, where it induces the transcription of Wnt-dependent genes. As a consequence, an increase in firefly luciferase expression is measured in the reporter assay, compared to the untreated negative control sample. The expression levels of the *Renilla* luciferase, instead, remain unaltered. The introduction of TNKS inhibitors in the form of small molecules or peptides, instead, has an inhibitory effect on the Wnt signalling pathway, measurable as a reduction in firefly luciferase expression and, therefore, reduced firefly luminescence signal.

Here, given the high transfection efficacy of HEK293T cells and their greater ease of *in vitro* culturing compared to other cell lines, these cells were selected for the TopFlash experiments. Activation of Wnt signalling was achieved by culturing transfected HEK293T cells in the presence of the Wnt conditioned medium, previously produced by culturing L Wnt-3A cells (ATCC), a cell line engineered to produce and secrete a non-tagged form of the biologically active Wnt3A glycoprotein in the medium.[213] Small molecule TNKSi or a macrocyclic TBP peptide can then be simply added to the culturing medium, as they are cell permeable.[85] In contrast, the nTBP-CTPR2n and the nTBP-CTPR2n-foldon proteins are unable to cross the plasma membrane on their own, due to their bulky size. Instead, a mammalian expression vector encoding the TBP-CTPR constructs was transfected into the cells, meaning that co-transfection of three plasmids was required for this TopFlash assay (Figure 7.1). As an alternative approach, purified TBP-CTPR protein can be encapsulated into fusogenic nanocarriers for their delivery into cells (and also for their protection against digestion by extracellular proteases). Both strategies were used and the results are described in Chapter 7 and Chapter 8, respectively. A dual-luciferase reporter assay (DLR, Promega) was then

performed to provide the corresponding substrate to each luciferase and measure the luminescence levels generated by both (Figure 7.1).

The stage in the experimental setting at which the TNKSi or CTPR constructs are introduced into each well is also critical for the analysis, as mentioned in the following section.
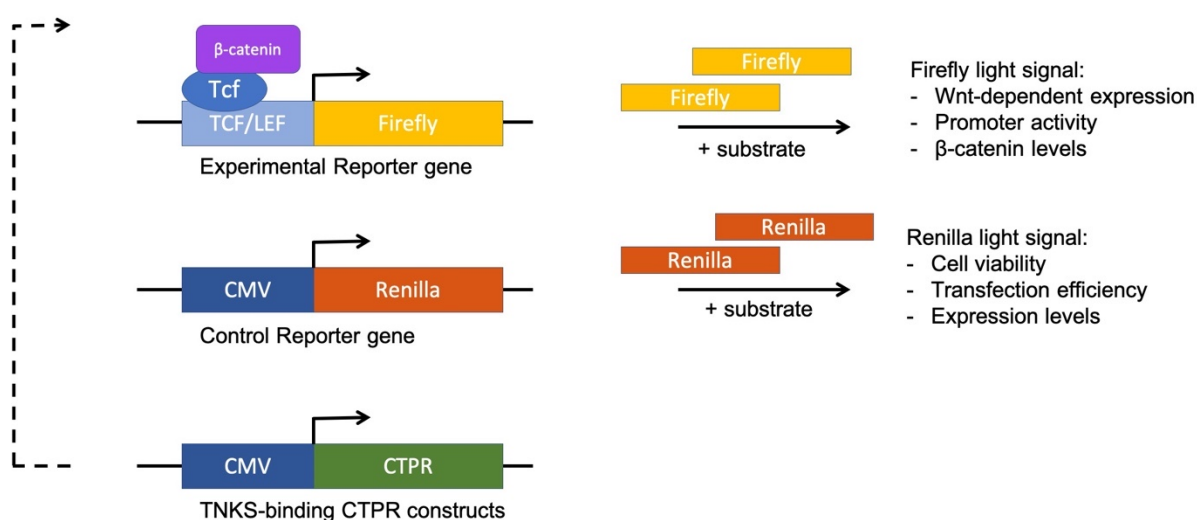
Figure 7.1: Schematic representation of the transfected plasmids required for the dual-luciferase reporter assay. The expression levels of firefly and *Renilla* are indicative of the events listed on the right. A dashed arrow corresponds to the inhibitory effect that the TNKS-binding CTPR constructs can have on the firefly expression levels.

7.1.2   Prophylactic vs interventional administration of drugs

Depending on their mechanism of action, drugs can exploit their pharmacological function prophylactically or interventionally. Agents with a pharmacological prophylactic activity prevent a disease from occurring and need to be used before a potential issue might arise. Drugs administered interventionally, instead, act by interfering with the biological target and pathway causing the disease. To be of clinical use in the treatment of Wnt-dependent cancers, drugs need to be administered and be active interventionally (i.e. when Wnt signalling is already upregulated). Evidence from my own work and from our collaborator Dr Marc de la Roche has revealed the prophylactic effect of the small molecule TNKSi in preventing activation of the Wnt signalling. These TNKSi, however, are not as potent when used interventionally.[139] To explore whether the linear nTBP-CTPR$_{2n}$ and trimeric nTBP-CTPR$_{2n}$-foldon constructs have an inhibitory effect on the Wnt signalling pathway, and to what extent, the Topflash assay was performed and the results compared to three well-characterised small molecule TNKSi. When

used prophylactically (as shown in this Chapter), the TNKS inhibitors, in the form of small molecules or CTPR constructs, are introduced before or at the same time as the Wnt conditioned medium and their activity is to prevent the activation of the Wnt signalling pathway. When used interventionally (as shown in the following Chapter), instead, the TNKS inhibitors are introduced after the pathway has already been activated by the Wnt molecule and their activity is to reduce the Wnt signalling.

7.2   Results

7.2.1   Subcloning into pcDNA3.1(-) and HiBiT-pcDNA3.1(-)

The TopFlash assay requires cell transfection with a mammalian expression vector encoding each of the linear $_n$TBP-CTPR$_{2n}$ and trimeric $_n$TBP-CTPR$_{2n}$-foldon constructs, as well as the control constructs. As described in Chapter 2, all the linear and trimeric CTPR constructs were therefore subcloned into our lab's variant of the commercially available pcDNA3.1(-) vector (Thermo Fisher), providing C-terminal HA tagging of the recombinant proteins. All the constructs were subsequently also cloned into the HiBiT-pcDNA3.1(-), an additional variant of the pcDNA3.1(-) vector that I generated, for the reasons mentioned below. Genes in this vector are fused to a N-terminal HiBiT tag and a C-terminal HA tag.

7.2.2   TopFlash dual-luciferase reporter assay - Prophylactic

The TopFlash reporter assay was used to evaluate the prophylactic inhibitory effect of the monomeric $_n$TBP-CTPR$_{2n}$ and trimeric $_n$TBP-CTPR$_{2n}$-foldon constructs on Wnt signalling. As controls, an empty vector (causing maximum firefly luciferase expression, set at 100% for normalisation); CTPR2, CTPR6, CTPR6-foldon and a 3RL-CTPR6 construct containing a non-binding peptide named "Random Loop" (RL) were tested in parallel. The cells that were not treated with Wnt conditioned medium (No WNT), instead, maintain minimal levels of Wnt signalling.

Several parameters of the TopFlash assay were initially optimised, such as the number of seeded cells, the quantity of plasmids to transfect, as well as the ratio among them, and the type of Wnt signalling activator (Wnt conditioned medium instead of LiCl). To test the prophylactic activity of the CTPR constructs, HEK293T cells were co-transfected with plasmids encoding the TCF/LEF-dependent firefly, CMV-dependent *Renilla* and a HA-tagged CTPR construct. Cells were allowed to recover for eight hours before activating the Wnt signalling pathway by adding the Wnt conditioned medium to the transfected cells. Following

a 16 h incubation, cells were observed under the microscope to identify any signs of toxicity induced by the Wnt activation, the transfection or the expression of the recombinant constructs. Figure 7.2 provides a representative comparison between cells transfected with an empty vector with and without the addition of Wnt conditioned medium, or the 3TBP-CTPR6 expression vector in the presence of Wnt conditioned medium. All wells were observed under the white light microscope before completing the assay, and no effects on cell viability induced by the Wnt conditioned medium or the expression of CTPR constructs could be observed.
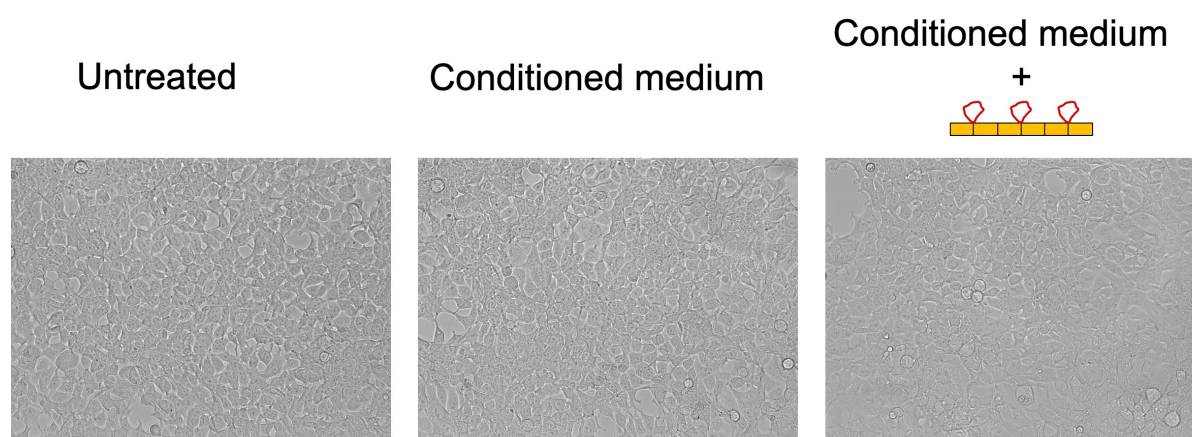
Figure 7.2: Living HEK293T cells imaged under bright field, using the EVOS Floid Cell Imaging Station, just before completion of the TopFlash assay. Representative figures are shown. Left: cells transfected with empty pcDNA3.1(-) vector; Middle: cells transfected with empty pcDNA3.1(-) vector and treated with Wnt conditioned medium; Right: cells transfected with pcDNA3.1(-) vector encoding 3TBP-CTPR6 and treated with Wnt conditioned medium.

The TopFlash assay was then completed using the Dual-Luciferase® Reporter Assay System that, following cell lysis, allows the consecutive measurement of the luminescent signal generated by firefly and *Renilla* on the same sample. To calculate the percentage luciferase activity, the firefly averaged luminescence signal was divided by the *Renilla* averaged luminescence signal in each sample and the ratio was normalised to the empty vector control sample set at 100%. The results indicate that treatment of HEK293T cells with all the $_n$TBP-CTPR$_{2n}$ and $_n$TBP-CTPR$_{2n}$-foldon constructs, with the exception of 1TBP-CTPR2, led to a significant reduction in Wnt pathway activity compared to the control samples having no TBP loop (Figure 7.3). The control constructs CTPR2, CTPR6, 3RL-CTPR6 (sequence provided in Appendix B) and CTPR6-foldon were unable to inhibit Wnt signalling, as expected. Notably, the resulting Wnt signalling inhibition observed here is an indirect measure of the effective

inhibition of the endogenous TNKS (not overexpressed TNKS as in Chapter 5) induced by the TBP-CTPR constructs.
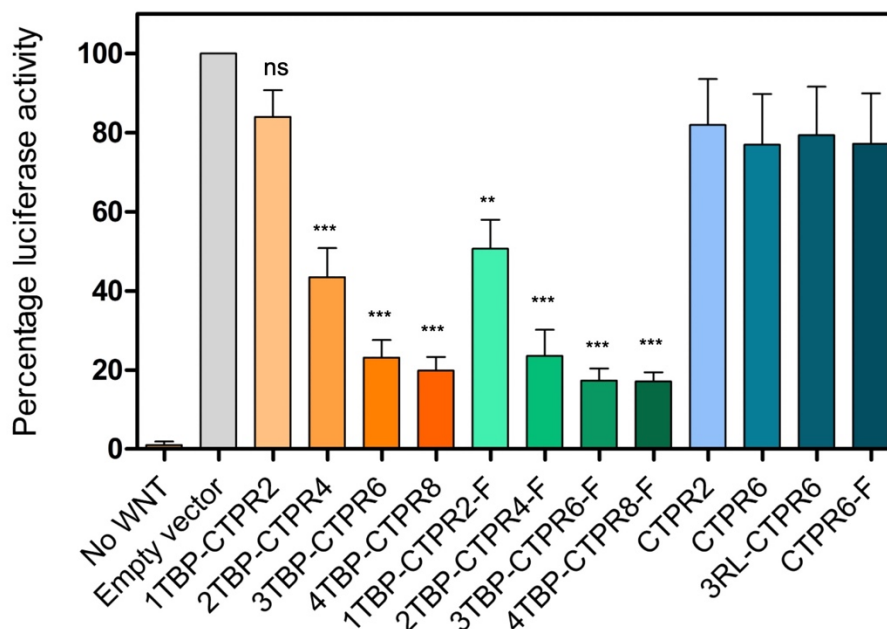
Figure 7.3: TopFlash reporter assay for HEK293T cells transfected with TCF-firefly, CMV-*Renilla* reporter gene vectors and an expression vector encoding the constructs listed. For each sample, firefly activity was normalised with the corresponding *Renilla* signal and the ratio was normalised to the control well transfected with the empty vector set at 100%. The monomeric constructs, trimeric constructs and controls are shown in different shades of orange, green and blue, respectively. F indicates the foldon motif. Standard deviation was calculated from six independent sample measurements. The significance of the difference between samples (ns: non-significant, $p>0.05$, *$p\leq0.05$, **$p\leq0.01$, ***$p\leq0.001$) was assessed using One-way ANOVA coupled with Dunnett's Multiple Correction test. 1TBP-CTPR2 and 1TBP-CTPR2-F were compared to CTPR2. Multivalent linear and trimeric constructs were compared to 1TBP-CTPR2 and 1TBP-CTPR2-F, respectively.

To assess the levels of HA-tagged CTPR proteins present in each sample during the TopFlash assay, Western Blot was performed on the same cell lysates using an anti-HA primary antibody. Despite transfecting cells with an equal amount of pcDNA3.1(-) plasmid encoding the various constructs, it was surprising to see a noteworthy difference in protein levels after 24 h from transfection (Figure 7.4). According to the Western Blot bands intensity, the trimeric $_n$TBP-CTPR$_{2n}$-foldon proteins appeared to be present at higher levels than their linear counterparts. Moreover, the membrane had to be overexposed in order to detect protein bands corresponding to shorter CTPR constructs, with 1TBP-CTPR2 still barely detectable.
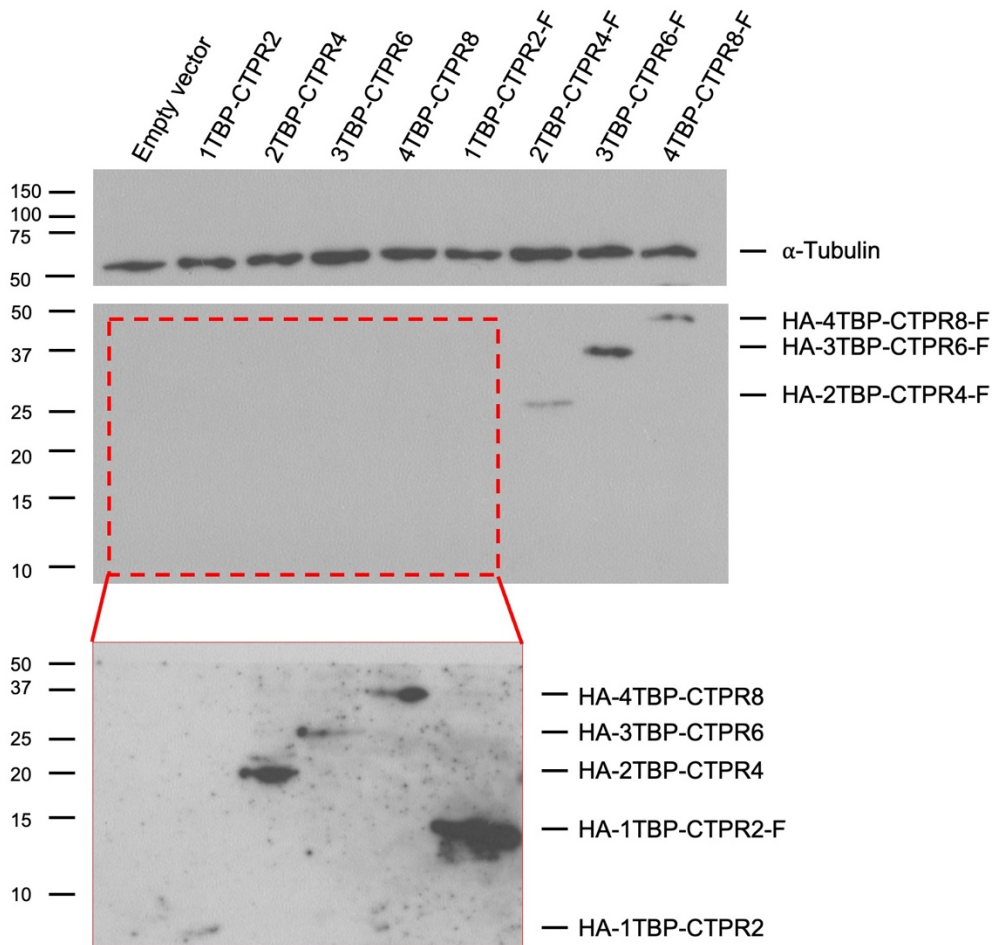
Figure 7.4: Western Blot analysis performed on some of the TopFlash cell lysate samples from Figure 7.3. Membranes were probed for α-tubulin and HA-tagged CTPR constructs, as indicated.

To address this problem, a more sensitive and quantitative technique than Western Blot was therefore necessary and was identified in the Nano-Glo HiBiT Lytic Detection System. As shown in Chapter 5, the Nano-Glo HiBiT technology allows the detection of proteins with high sensitivity directly on the cell lysate without the need of performing Western Blot. This technology, however, requires all of the CTPR constructs to be tagged with the HiBiT sequence. For this purpose, all CTPR-encoding genes were subcloned into the HiBiT-pcDNA3.1(-) expression vector. The TopFlash assay was therefore repeated with the newly generated HiBiT-tagged constructs under the same experimental settings. Results consistent to those in Figure 7.3 were obtained, with the same extent of inhibition being observed for each protein construct (Figure 7.5A).

The cell lysates obtained by TopFlash were then incubated with the Nano-Glo HiBiT Lytic Detection System. Given the high sensitivity of the split-luciferase Nano-Glo technology, the intracellular protein level of each HiBiT-tagged CTPR construct could be successfully measured and quantified (Figure 7.5C). The measured values fall within the broad linearity range of luminescence generated by HiBiT, according to published and my own evidence (data not shown).[203] The result confirmed that the trimeric TNKS-binding constructs were present in higher amounts than their linear counterparts, as observed previously by Western Blot. Moreover, in both the linear and the trimeric geometries, the protein levels generally increased with longer CTPR constructs, with the 3TBP-CTPR6 construct showing the highest protein amount. This trend, however, was not observed for the control constructs, that were all present in very low amounts, independently of their molecular weight. When comparing Figure 7.5A with Figure 7.5C, it is possible to detect a complementary trend, suggesting that a lower protein concentration might lead to a lower Wnt signalling inhibition. Therefore, the extent of Wnt inhibition appears to be proportional to the CTPR protein levels as quantified 24 h after transfection.

Finally, the effects of the CTPR proteins were compared to those obtained using three well characterised small molecule TNKSi, tested prophylactically. Each of the three TNKSi was added with the Wnt conditioned medium and at a final concentration of 1 μM (Figure 7.5B). The comparison shows that some of the $_n$TBP-CTPR$_{2n}$ and $_n$TBP-CTPR$_{2n}$-foldon constructs can inhibit the Wnt signalling pathway to a similar extent as the small molecule TNKSi.
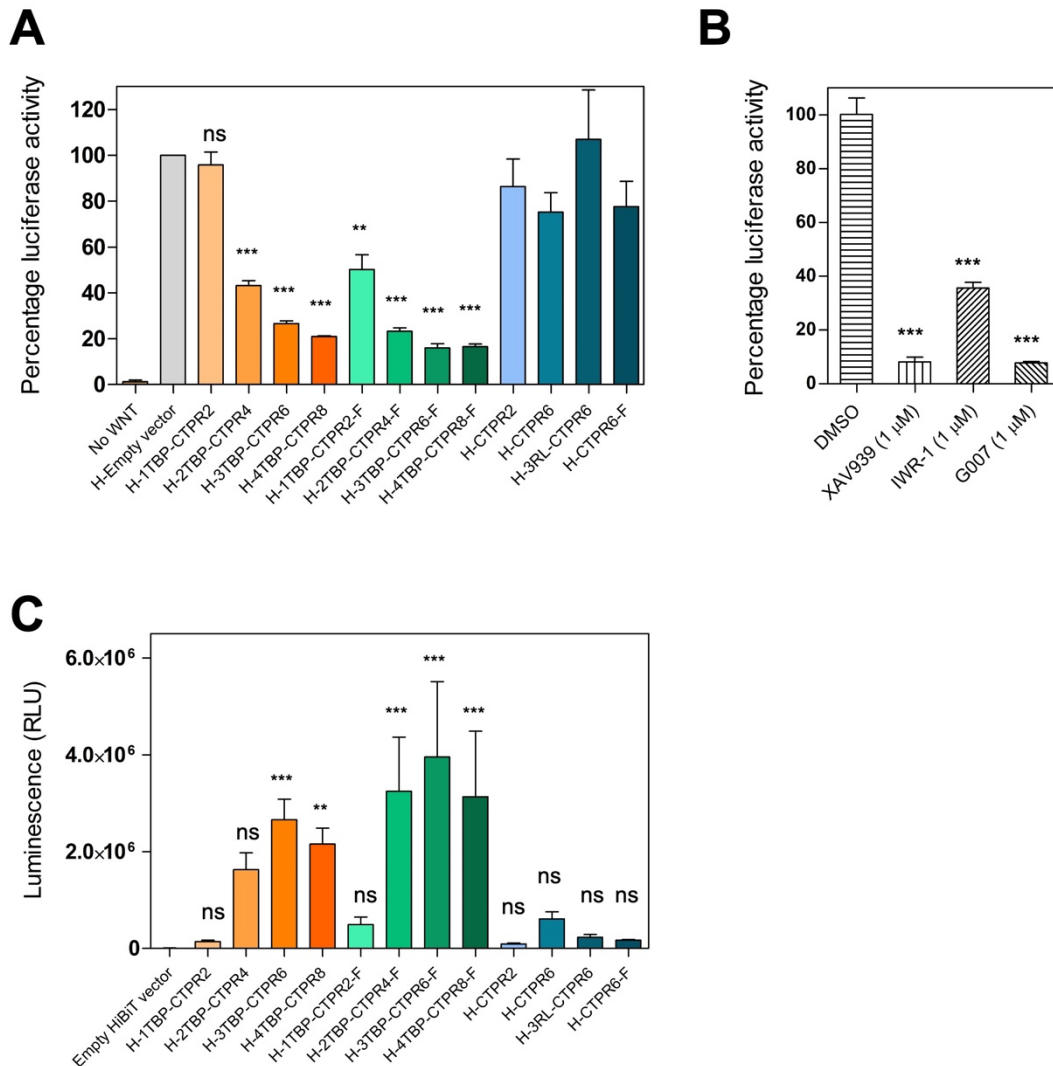
Figure 7.5: Prophylactic activity of the CTPR constructs and TNKSi in inhibiting the Wnt signalling. (A) TopFlash reporter assay for HEK293T cells transfected with TCF-firefly, *Renilla* reporter gene vectors and an expression vector encoding the constructs listed. For each sample, firefly activities were normalised with the corresponding *Renilla* values and the ratio was expressed as relative luciferase activity to the control well transfected with the empty HiBiT vector set at 100%. The monomeric constructs, trimeric constructs and controls are shown in different shades of orange, green and blue, respectively. F indicates the foldon motif, H indicates the N-terminal HiBiT tag. Standard deviation was calculated from triplicate sample measurements. The significance of the difference between samples (ns: non-significant, p>0.05, *p≤0.05, **p≤0.01, ***p≤0.001) was assessed using One-way ANOVA coupled with Dunnett's Multiple Correction test. 1TBP-CTPR2 and 1TBP-CTPR2-F were compared to CTPR2. Multivalent linear and trimeric constructs were compared to 1TBP-CTPR2 and 1TBP-CTPR2-F, respectively. (B) Effect of small molecule TNKSi tested prophylactically in the TopFlash assay. Cells were treated with inhibitors and Wnt-conditioned medium simultaneously. Final concentration was 1 µM for all the small molecule TNKSi tested. Data were normalised by the untreated control well, set at 100% (not shown in the graph). Error bars were determined from two independent sample measurements. The statistical analysis was performed as in A, with samples compared to DMSO. (C) Luminescence readings corresponding to the intracellular HiBiT-tagged CTPR protein levels from samples in (A). Data were averaged and standard deviation was calculated from triplicate sample measurements. The significance of the difference between each sample and Empty HiBiT vector (ns: non-significant, p>0.05, *p≤0.05, **p≤0.01, ***p≤0.001) was assessed using One-way ANOVA coupled with Dunnett's Multiple Correction test.

7.2.3   Proteasome-induced degradation assay

Data in Figure 7.5C reveal that all the CTPR2 proteins, in the form of 1TBP-CTPR2, 1TBP-CTPR2-foldon and CTPR2, are present at much lower levels than the larger proteins. Differences in protein amounts might arise from variable expression rates at the transcription and/or translation level or might be due to differential resistance to degradation. Due to their intrinsic lower stability (Chapter 4), we hypothesised that the smaller, less stable CTPR2 constructs might undergo proteasome-induced degradation to a greater extent than the other CTPR constructs. To address this hypothesis, a proteasomal-induced degradation assay, in combination with the Nano-Glo HiBiT Lytic Detection System, was established. In this assay, wells transfected in duplicates with a HiBiT-tagged construct were incubated in the presence or absence of the proteasome inhibitor MG132 for 5 h. MG132 is a potent, cell-permeable and synthetic peptide aldehyde, able to inhibit multiple peptidase activities within the proteasome. MG132 and other proteasome inhibitors are known to be toxic to cells, leading to cell death when used at high concentrations or for prolonged incubation times. Therefore, MG132 treatment needs to be complemented with a cell viability assay for normalisation. The appropriate cell viability assay has to fulfil the following requirements, in order to be multiplexed with the Nano-Glo luminescence assay:

- Cell viability and HiBiT-mediated luminescence should be measured sequentially on the same cell sample and in this order, as the Nano-Glo System is lytic. The cell viability assay must, therefore, adopt a non-lytic reagent.

- The cell viability measurement must not interfere with the consecutive HiBiT-mediated luminescence and must, therefore, exploit a physical property different than luminescence.

The compatible cell viability assay was identified to be the CellTiter-Fluor Cell Viability Assay (Promega). The assay relies on a fluorogenic, cell-permeable peptide substrate (Gly-Phe-AFC), that produces a fluorescent signal upon cleavage by the conserved and constitutive live-cell protease activity. The live-cell protease activity is restricted to intact viable cells, providing a measurable signal proportional to the number of living cells.

Following MG132 treatment for the indicated incubation time, cell viability was measured, and a significant decrease in the number of viable cells was observed in wells treated with MG132, further validating the need for a sensitive cell viability measurement to be multiplexed in this assay (Figure 7.6A). Cells were then analysed for their CTPR protein levels and the HiBiT-induced luminescence was measured. The luminescent signal was then

normalised by the corresponding cell viability to account for the variability in cell numbers between the two experimental conditions (+MG132 and -MG132). The ratio +MG132/-MG132 was then calculated and plotted (Figure 7.6B). The experiment shows that, in the presence of MG132, there is an approximate 3-fold increase in protein levels of all the constructs, with the exception of the two smallest proteins 1TBP-CTPR2 and CTPR2, whose stability is more affected. In other words, the results indicate that 1TBP-CTPR2 and CTPR2 undergo proteasome-mediated degradation to a much greater extent than the other, larger protein constructs. Moreover, when comparing the linear with the trimeric constructs, the foldon domain appears to provide the CTPR2 construct with protection against proteasome-mediated degradation, whereas the foldon motif does not appear to provide protection to the larger CTPR constructs presumably because these proteins are already reasonably stable (Figure 7.6B).

The cell viability measurement shown here provided additional evidence regarding the potential cytotoxicity effect induced by the recombinant CTPR constructs. In agreement with the microscopy images in Figure 7.2, the cell viability measurements in the absence of MG132 are comparable across all the samples tested, confirming that none of the constructs showed cytotoxicity (Figure 7.6A).
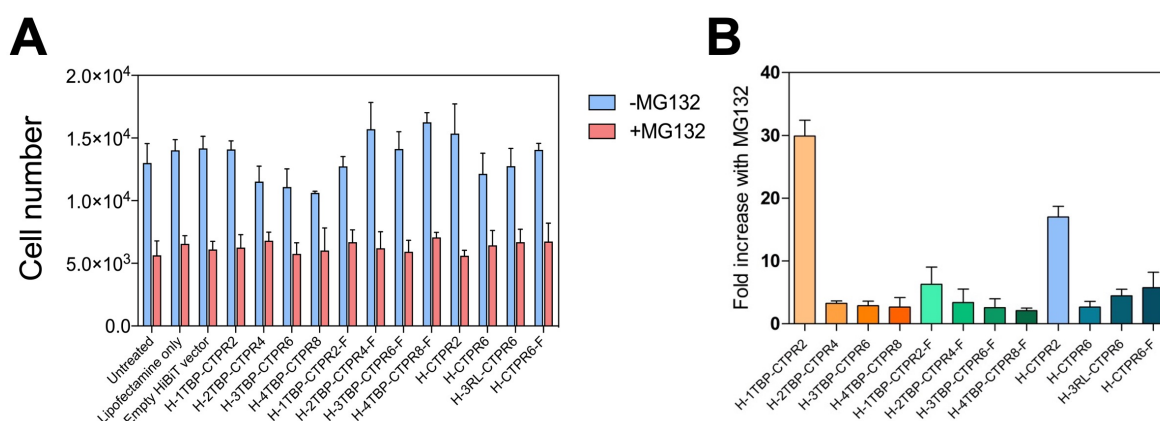
Figure 7.6: Proteasome-induced degradation assay. (A) Cell viability assay showing the viable cell number in the presence or absence of MG132, following transfection with the indicated constructs. (B) Fold increase in HiBiT-tagged CTPR protein levels in the presence of the proteasome inhibitor MG132. Data were averaged, and standard deviation was calculated from triplicate samples.

131

7.3 Discussion

The results in this Chapter provide interesting insights into the inhibitory potential of the linear $_n$TBP-CTPR$_{2n}$ and trimeric $_n$TBP-CTPR$_{2n}$-foldon constructs when tested prophylactically. To verify whether the proven intracellular interaction between the TBP loop and TNKS (demonstrated in Chapters 5 and 6) can lead to the desired Wnt signalling inhibition, the TopFlash assay was adopted. The TopFlash assay is a well-established and easy method to measure the effect of Wnt inhibitors on the Wnt signalling pathway. However, it requires cell transfection with at least two plasmids (Wnt-dependent firefly and a control luciferase reporter) and the activation of Wnt signalling. When small molecule Wnt-antagonists are tested with the TopFlash assay, they can be added directly into the culture medium as they are cell-permeable. In the TopFlash assay described here, instead, the CTPR proteins were introduced by transfecting cells with an additional plasmid encoding them, making a total of three plasmids to transfect. The same multi-plasmid transfection has been previously tested to compare the activity of transfected Wnt ligands or protein-based antagonists such as Dickkopf-related protein 1 (Dkk-1) and the Secreted Frizzled-related Proteins (SFRPs), validating the approach used here.[214,215] The CTPR-encoding plasmid was transfected in combination with the firefly- and *Renilla*-expression vectors, eight hours before Wnt activation. By doing so, the CTPR constructs are tested for their prophylactic activity. However, CTPR expression under the CMV promoter within the pcDNA3.1(-) plasmid is expected to occur over a long period of time, causing a continuous TNKS inhibition both before and after the introduction of the Wnt conditioned medium.

Significantly different CTPR protein levels were measured after 24 hours, making it difficult to compare the inhibitory effects of the different linear $_n$TBP-CTPR$_{2n}$ and trimeric $_n$TBP-CTPR$_{2n}$-foldon constructs and establish their prophylactic activity. Differences in protein levels can be partially explained by their different stability and susceptibility to proteasome-mediated degradation. Shorter CTPR constructs, such as 1TBP-CTPR2 and CTPR2, undergo proteasome-mediated degradation to a greater extent than do the larger CTPR constructs. Their intrinsic lower stability might contribute to their enhanced degradation. As proteasomal degradation equally affects longer TNKS-binding and control CTPR constructs, we also propose that the multivalency-induced macromolecular assemblies (as described in Chapter 6) might help protect the multivalent TBP-CTPR constructs from degradation. This effect would explain why the levels of the multivalent TBP-CTPR proteins are much higher than those of 1TBP-CTPR2 and the non-binding control constructs. On a similar note, as Wnt

inhibition is observed in cells transfected with the multivalent TBP-CTPR constructs, it is also possible that Wnt inhibition is enhanced by clustering TNKS within these macromolecular structures. Unlike the previously observed macromolecular complexes formed by overexpressed eGFP-tagged TNKS2 (Chapter 6), the macromolecular complexes suggested here would contain the endogenous TNKS1 and TNKS2 proteins. Further investigation is required to validate these hypotheses.

Three well-characterised small molecule TNKSi (XAV939, IWR-1 and G007) were also tested for comparison and, in agreement with previous studies,[109,134,138] around 90% inhibition of Wnt pathway activity was observed for XAV939 and G007 when used prophylactically. It was noteworthy to see that some of the linear nTBP-CTPR$_{2n}$ and trimeric nTBP-CTPR$_{2n}$-foldon constructs were able to induce a similar extent of Wnt signalling inhibition in this setting. Lastly, no cytotoxicity effects could be observed upon over-expression of the CTPR constructs, a desired feature for any further development of these proteins as biotherapeutic molecules. Altogether, the Wnt inhibition observed here is an encouraging starting point from which to explore further the inhibitory potential of the TBP-CTPR proteins. In particular, the possibility of using them interventionally needs to be tested for them to be of clinical use. To this end, the TopFlash assay was performed in a different format, as described in the next Chapter.

Chapter 8

Intracellular delivery of $_n$TBP-CTPR$_{2n}$ proteins

The results presented in this Chapter were obtained in collaboration with Dr Piyush K. Chaturbedy, a post-doc in the Itzhaki lab with experience in nanoparticle synthesis and their application as intracellular drug delivery methods.

8.1   Introduction

Intracellular delivery of biotherapeutics represents a major challenge today. To avoid this problem, the vast majority of biotherapeutics, such as antibody-based drugs and alternative scaffolds-based inhibitors, are developed to bind and inhibit extracellular epitopes, as their large molecular weight size impedes their diffusion through the plasma membrane. Delivery methods are required to reach an intracellular target with a pharmacological molecule that is unable to cross the plasma membrane on its own.[216] Various delivery methods have been developed in recent years, and some are in clinical use.[217] The results described in the previous Chapter showed the potential of the CTPR proteins in inhibiting TNKS using the TopFlash Wnt signalling assay. In those experiments, the cells were transfected with DNA to express the CTPR proteins; in this Chapter, instead, we attempted to mimic more closely what would be required for a drug, and used so-called "fusogenic liposomes" to deliver the proteins into the cell.

8.1.1 Fusogenic liposomes

Fusogenic liposomes are a class of non-viral delivery systems consisting of lipid vesicles. They can encapsulate large water-soluble cargos, such as proteins, and can deliver them into the cellular cytoplasm by membrane fusion. They are easy to prepare, safe and can be optimised to the desired structure, size and surface charge by varying their physio-chemical properties. Additional pharmacological properties include their scarce immunogenicity and their easy manipulation. Moreover, they protect the cargo, preventing its premature degradation and its recognition by the immune system.[218] Fusogenic liposomes based on cationic lipids have gained interest in biomedical applications. The lipid's positive charge, indeed, allows encapsulation of large amounts of DNA used for gene therapy applications, and promotes the interaction with the plasma membrane, therefore favouring the fusion and uptake by the cell. Liposomes are generally produced by the combination of a cationic lipid with a helper lipid. The fusogenic liposomes used in this study are formed by a 10:10:1 mixture of 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) and 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindotricarbocyanine iodide (DiR).[158] DOTAP is a widely used cationic lipid, known for its ability to generate stable and fluid bilayers. DOPE is a common helper lipid, which is used to improve structural stability and decrease toxicity. Both DOTAP and DOPE have unsaturated fatty acid lipid chains, and this characteristic provides higher flexibility to the vector surface and also allows a better incorporation of the cargo. DiR is a lipophilic dye used to trace the presence of the liposomes in an experimental setting.

Preparation of liposomes is usually a three-step process:

1. Lipids and helpers dissolved in volatile organic solvents are mixed and dried to form thin films named "lipid cakes";
2. Lipid cakes are rehydrated in the desired buffer containing the cargo to be encapsulated;
3. Sonication with high-power ultrasounds generates unilamellar liposomes, bounded by a single bilayer of lipids.

Liposomes are stable at room temperature and their surface positive charge attenuates their tendency to fuse to each other, as a result of charge repulsion among liposomes. Fusogenic liposomes have also been shown to merge with the cell membrane and deliver their cargo directly into the cell cytoplasm, bypassing endocytosis.[158]

8.1.2 Experimental design

In this Chapter, we tested the possibility of delivering the CTPR proteins into the cell by encapsulating them using fusogenic liposomes. Among our set of linear and trimeric CTPR proteins, 3TBP-CTPR6 was chosen as representative of the linear array. As control, the construct of equivalent CTPR units but lacking the TBP loops, CTPR6, was used. As the overall aim of these experiments was to demonstrate intracellular delivery, only those two proteins were used. The negative charge of the CTPR proteins is a favourable property for their encapsulation into cationic fusogenic liposomes. CTPR constructs were loaded into fusogenic liposomes by rehydrating lipid cakes in the presence of the purified protein as described in the previous section.

Encapsulated and empty liposomal formulations were first characterised in terms of their biophysical properties and for their potential cytotoxicity in a cell-based assay. Then, the ability to successfully deliver the cargo of interest was assessed using the TopFlash assay (Figure 8.1).
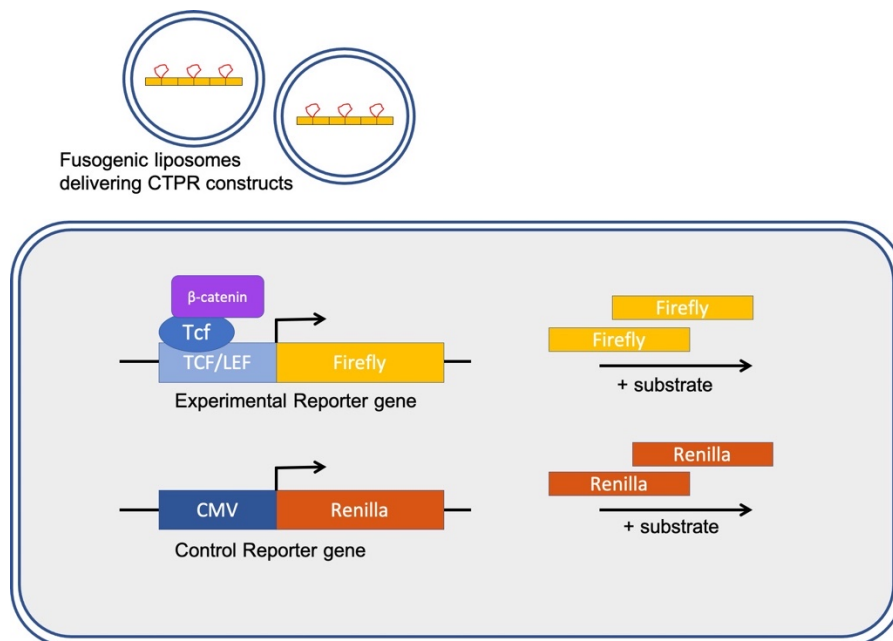
Figure 8.1: Schematic representation of the dual-luciferase reporter assay, performed here by transfecting the firefly and *Renilla* plasmids and delivering the CTPR constructs using fusogenic liposomes. The lipid bilayer of the plasma membrane and the liposomes is represented as two continuous lines close to each other. 3TBP-CTPR6 was chosen as representative CTPR construct to be encapsulated into fusogenic liposomes.

8.2 Results

8.2.1 Liposome characterisation

Fusogenic liposomes were prepared as mentioned above and initially characterised for their biophysical properties. The surface charge of fusogenic liposomes containing 3TBP-CTPR6 (FL-3TBP-CTPR6) and control liposomes without protein (FL) was therefore measured. Given the positive charge of the lipids, both FL and FL-3TBP-CTPR6 had highly positive surface with zeta potential (ZP) values of +126 mV and +75.7 mV, respectively, at pH 7.4 (Figure 8.2A, B). The lower ZP of FL-3TBP-CTPR6 is expected, due to encapsulation of the negatively charged 3TBP-CTPR6 (pI ~ 4.8) within the liposomes. The hydrodynamic sizes of FL and FL-3TBP-CTPR6 were similar, at ~106 nm (Figure 8.2 C, D).
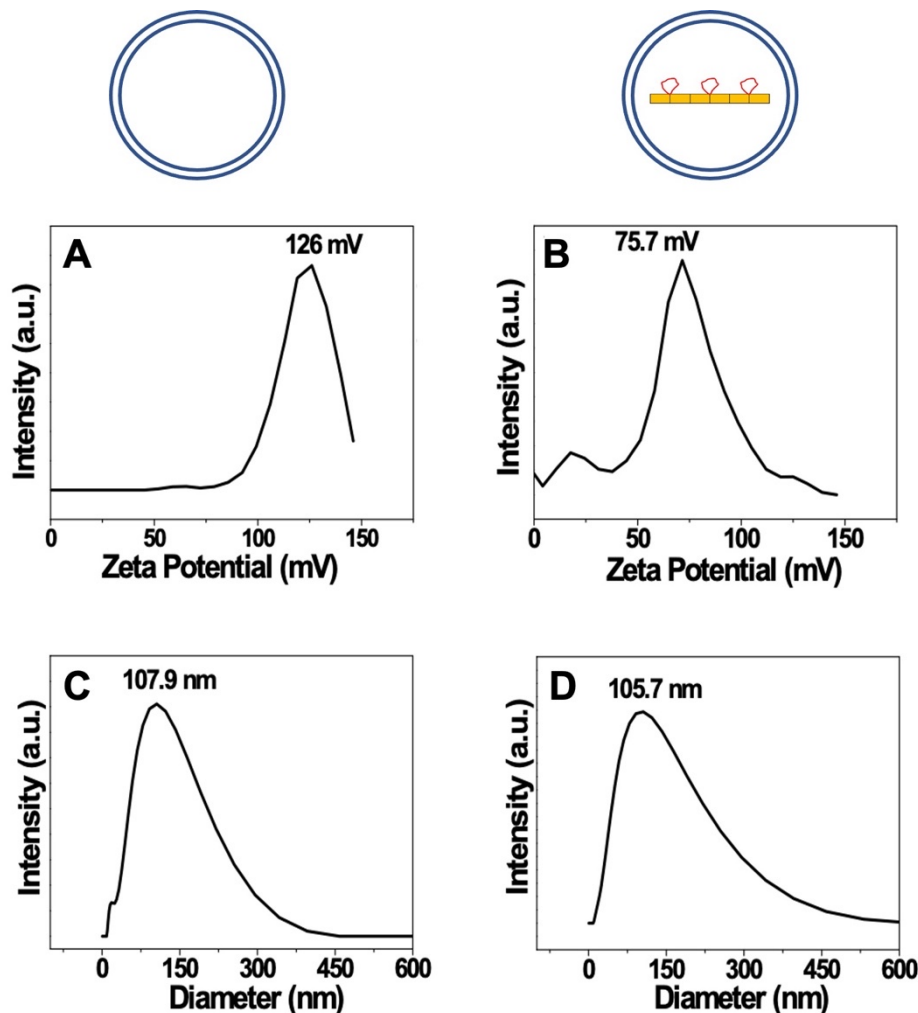
Figure 8.2: Characterisation of fusogenic liposomes. Schematic representation of FL (left) and FL-3TBP-CTPR6. The surface charge and the hydrodynamic size of FL (A, C) and FL-3TBP-CTPR6 (B, D) were measured. This experiment was performed by Dr. Piyush K. Chaturbedy.

We next wanted to assess the cytotoxicity potential of fusogenic liposomes when incubated with living cells. To this purpose, HEK293T cells were treated with increasing amounts of FL and FL-3TBP-CTPR6 for 15 minutes and the CellTiter-Glo viability assay was then performed. FL and FL-3TBP-CTPR6 did not show any significant cytotoxicity even for amounts higher than those used in the subsequent experiments, relative to the number of seeded cells (Figure 8.3).
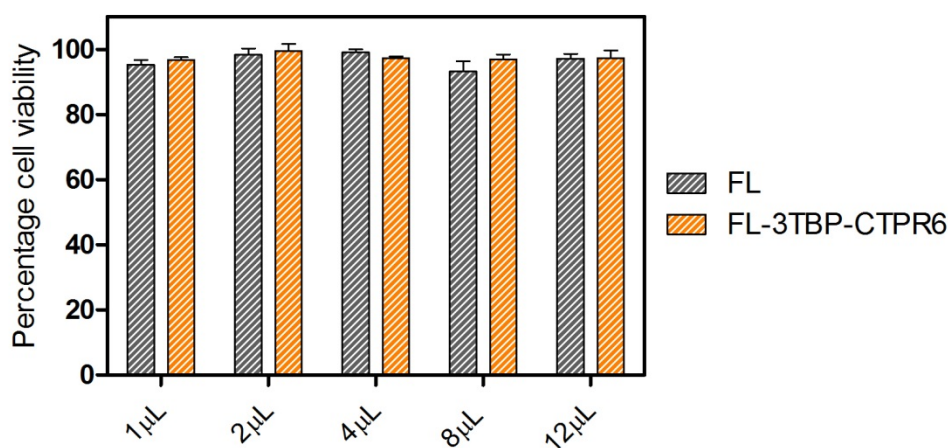
Figure 8.3: Cell viability assays of empty liposomes (FL, black bars) and liposome-encapsulated 3TBP-CTPR6 protein (FL-3TBP-CTPR6; grey bars). Untreated cells were taken as control for the experiment. Data were normalised relative to untreated cells set at 100% (not shown). Error bars were obtained from triplicate sample measurements from two independent experiments. This experiment was performed by Dr. Piyush K. Chaturbedy.

We then sought to visualise the liposome-mediated intracellular protein delivery under the confocal microscope. To this aim, 3TBP-CTPR6 was fluorescently labelled with rhodamine B isothiocyanate (RITC), a fluorescent dye used as tracer. Liposomes were prepared in the presence of the labelled protein and the new formulation was named FL-3TBP-CTPR6-RITC. The liposome formulation includes DiR, a lipophilic dye which allows us to localise where the liposomes have fused to the plasma membrane. Confocal images of HEK293T cells treated with FL-3TBP-CTPR6-RITC (Figure 8.4) clearly indicate that liposomes have fused to the plasma membrane and the protein has been delivered inside the cells and is distributed throughout the cytoplasm. Cells treated with empty FL did not show any signal for 3TBP-CTPR6-RITC, as expected (Figure 8.5).
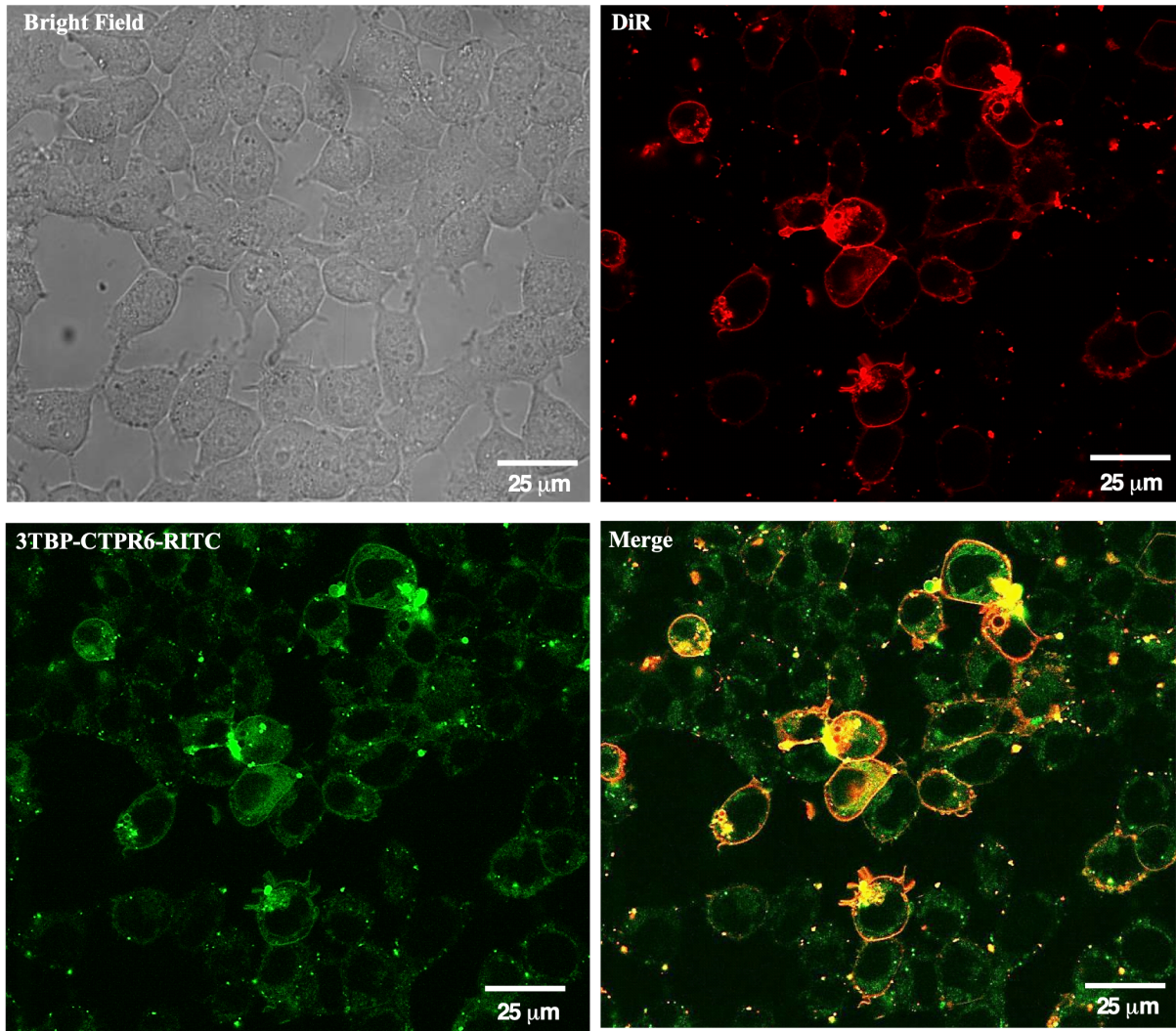
Figure 8.4: Confocal microscope images of HEK293T cells treated with liposome-encapsulated protein (FL-3TBP-CTPR6-RITC). DiR (λ_{ex} = 633 nm, λ_{em} = 720-800 nm) stains the cellular membrane. 3TBP-CTPR6-RITC (λ_{ex} = 514 nm, λ_{em} = 530-650 nm) is distributed throughout cell cytoplasm. The merge of red and the green channel clearly shows that protein has been delivered inside the cells. This experiment was performed by Dr. Piyush K. Chaturbedy.
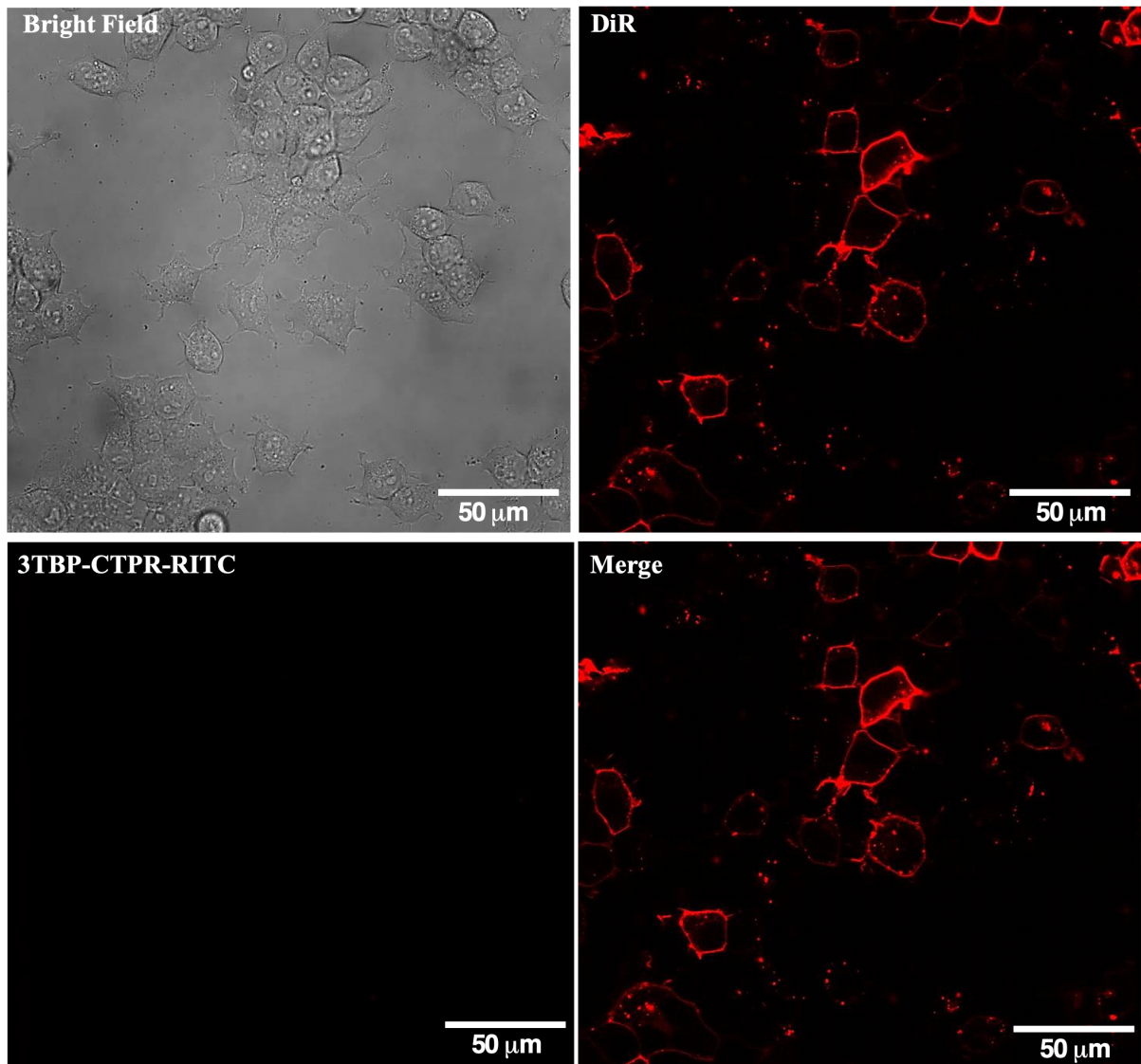
Figure 8.5: Confocal microscopy of HEK293T cells treated with empty FL. DiR ($\lambda_{ex} = 633$ nm, $\lambda_{em} = 720\text{-}800$ nm) stains the cellular membrane. This experiment was performed by Dr. Piyush K. Chaturbedy.

8.2.2   TopFlash dual-luciferase reporter assay – Interventional

The effect of liposome-delivered 3TBP-CTPR6 on Wnt signalling was then tested using the TopFlash assay. For this assay, fusogenic liposomes containing the control construct CTPR6 were also prepared, for comparison. Liposomal delivery of the CTPR proteins allows us to measure their inhibitory effect on the Wnt signalling pathway interventionally. Unlike DNA transfection, which induces constitutive protein expression throughout the incubation period, liposome-mediated delivery allows the dosage of the protein in a much more accurate way and with a specific timing. To test their interventional activity, cells were treated with liposome-

encapsulated CTPR constructs 16 hours after Wnt signalling had been activated using Wnt conditioned medium. This experimental set-up more closely correlates to the clinical setting, as Wnt-dependent cancer cells have high activation of the Wnt signalling pathway at the time of pharmacological treatment, and the drug aims to reduce this elevated signalling.[219] Wnt-activated HEK293T cells were therefore treated with FL-3TBP-CTPR6 and FL-CTPR6 for six hours and the TopFlash assay was then completed as indicated in the previous Chapter. A decrease in Wnt signalling was observed in cells treated with FL-3TBP-CTPR6, with the percentage luciferase activity lowered by about 50% after this short incubation time (Figure 8.6A). No effect on the luciferase levels was observed in cells treated with FL-CTPR6, as expected. The assay was also performed in the presence of fusogenic liposomes containing progressively lower concentrations of 3TBP-CTPR6. Importantly, we could observe a correlation between the TopFlash activity and the concentration of FL-3TBP-CTPR6 (Figure 8.6A), indicating a dose-dependent inhibition of endogenous TNKS. As additional control, treatment with free 3TBP-CTPR6 did not affect the luciferase expression levels as the protein cannot enter cells on its own. Similarly, FL did not alter luciferase levels, indicating that membrane fusion of liposomes does not interfere with the Wnt signalling. Importantly, the TopFlash assay was also used to evaluate the interventional activity of the three small molecule TNKSi under identical experimental conditions and at the highest concentration of FL-3TBP-CTPR6 (Figure 8.6B). It is noteworthy that FL-3TBP-CTPR6 resulted in a greater extent of Wnt signalling inhibition compared to all three small molecule TNKSi, further underlying the potential of our molecules for future applications.
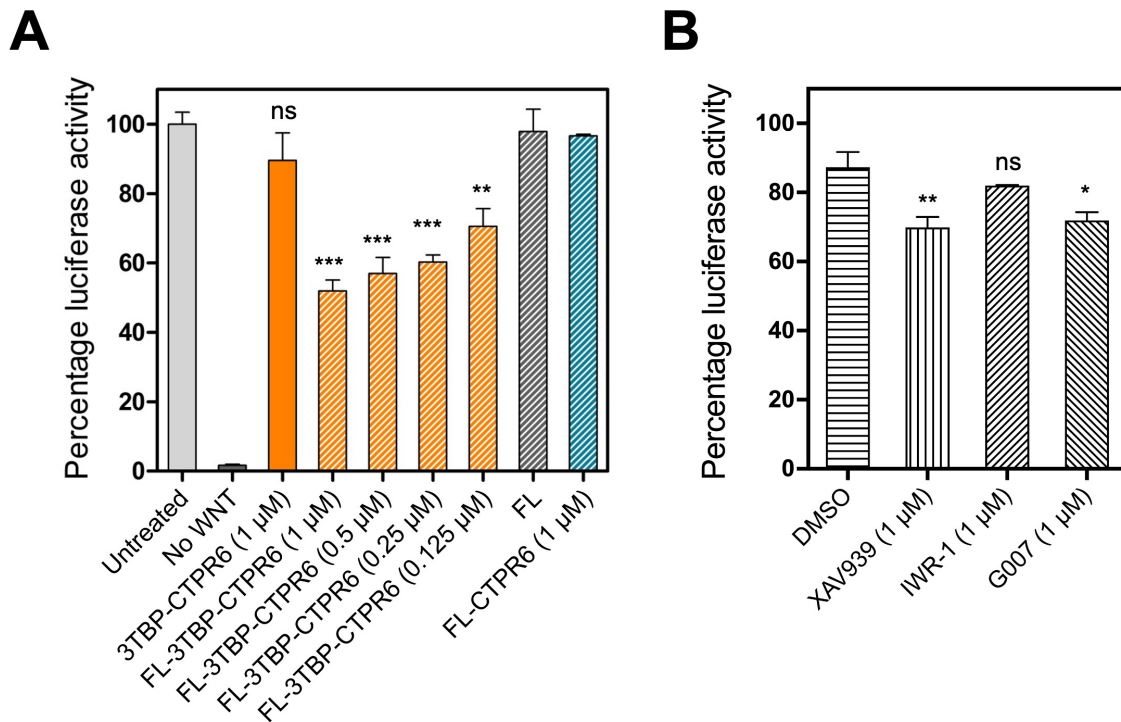
Figure 8.6: TopFlash assay with CTPR constructs and TNKSi tested interventionally. (A) Inhibition of Wnt signalling in HEK293T cells by fusogenic liposome-encapsulated 3TBP-CTPR6. Each treatment was with 20 µL of liposomes. In brackets are the concentrations of the proteins used. For each run, data were normalised by the untreated control well, set at 100%. Bars with diagonal stripes correspond to samples treated with liposomes. No Wnt: cells without Wnt pathway activation and not treated with liposomes. Untreated cell: cells not treated with liposomes. FL: empty liposomes. Error bars were obtained from triplicate sample measurements from two independent experiments. The significance of the difference between samples (ns: non-significant, p>0.05, *p≤0.05 **p≤0.01, ***p≤0.001) was assessed using One-way ANOVA coupled with Dunnett's Multiple Correction test. Samples having an inhibitory effect were compared to the control sample FL-CTPR6. The results in A were obtained in collaboration with Dr. Piyush K. Chaturbedy. (B) Effect of small molecule TNKSi tested interventionally. Cells were first treated overnight with Wnt conditioned medium only and then TNKSi were added and incubated for an additional 6 h. Final concentration was 1 µM for all the small molecule TNKSi tested, in 0.5% DMSO, equal to the highest 3TBP-CTPR6 concentration used in A. Data were normalised by the untreated control well, set at 100% (not shown). Error bars were determined from two independent sample measurements. The same statistical analysis was performed as in (A), comparing the small molecule samples to DMSO.

8.3 Discussion

In the previous Chapter, the linear $_n$TBP-CTPR$_{2n}$ and trimeric $_n$TBP-CTPR$_{2n}$-foldon constructs were tested for their prophylactic effects on Wnt signalling using the TopFlash assay. However, to be of clinical use in a Wnt-activated disease, the drug needs to be used interventionally. Moreover, given their large size of the CTPR proteins and their inability to cross the plasma membrane, a delivery method is also required for therapeutic purposes and to allow them to

reach their intracellular target. For these reasons, the purified CTPR proteins were encapsulated within fusogenic liposomes and were tested for their interventional potential on Wnt-activated HEK293T cells. Liposomal formulations were stable at room temperature and easy to prepare, a prerequisite for pharmacological development. Encapsulation within fusogenic liposomes not only favours intracellular delivery bypassing endocytosis but will also protects the CTPR constructs from extracellular proteolytic degradation and adverse immune response, if any, when used therapeutically.[218]

The results showed that fusogenic liposomes could effectively deliver 3TBP-CTPR6 into the cytoplasm of cells, without inducing cytotoxicity effects. A variation of the previously adopted TopFlash assay allowed us to measure the interventional activity of 3TBP-CTPR6, delivered after Wnt activation. The protein construct elicited very significant inhibition of Wnt signalling, in a dose-dependent manner. This is a noteworthy result that demonstrates interventional inhibition of the pathway, the scenario for targeting tumours dependent on deregulated Wnt pathway activity *in vivo*. Moreover, the luciferase protein reporter of Wnt pathway activity has a relatively long half-life of approximately 12 hours, and therefore inhibition after 6 hours treatment likely reflects a much higher attenuation of Wnt pathway activity beyond the measured 50% inhibition. In other words, 3TBP-CTPR6 affects the firefly transcription rate only after its liposomal delivery, but does not interfere with those firefly molecules expressed beforehand that are still present within the cell and contribute to the overall luminescence signal. Therefore, in these experimental conditions, the luminescence levels cannot be completely abolished, but only reduced. Although this experimental set-up represents an improvement to the previously described TopFlash assay (Chapter 7), it still relies on cell transfection with two luciferase-expressing plasmids. Transfection efficiency can vary across experimental replicates and cell lines, can be affected by several factors (such as passage number and cell confluency) and might require selection of transfected cells using the appropriate antibiotic. Ideally, a stable cell line, which expresses a luciferase reporter under the control of a Wnt-responsive promoter, should be used instead. Apart from providing more accurate results, this Wnt reporter cell line would also prevent the need of transfecting cells with the control *Renilla* luciferase.

According to my own (Figure 8.6B) and previously published results,[139] the interventional activities of the small molecule TNKSi is much less effective than their prophylactic activities, due to a cell-intrinsic feed-forward mechanism preserving high Wnt pathway activity once over-activated.[139] Importantly, 3TBP-CTPR6 delivered by fusogenic liposomes induced a higher extent of inhibition compared to all three small molecule TNKSi

tested interventionally at the same concentration. This striking result leads to the following conclusions. First, it shows that targeting the non-catalytic activity of TNKS can be an effective alternative to inhibiting PARP activity. Indeed, the former has multiple advantages over the latter: targeting the substrate-binding domain will enable inhibition of both catalytic and non-catalytic (scaffolding) functions, and it will overcome the off-target effects of TNKSi on other PARP proteins. Second, the presence of multiple binding sites on 3TBP-CTPR6 might contribute to its enhanced interventional activity compared to small molecule inhibitors, through the formation of large macromolecular assemblies. Finally, it shows that Wnt signalling was inhibited irrespective of the format by which the protein binders were introduced, whether by DNA transfection or by protein delivery. As our aim was to test the possibility of delivering the TBP-CTPR proteins intracellularly, we looked at one protein only. However, in the future it would be worth testing the effect induced by other TNKS-binding CTPR constructs using this delivery approach, to efficiently compare their inhibitory potential (in particular compare the effects of single- and multi-valency) and further validate our conclusions. There is currently great interest in developing mRNA drugs, and a number of liposome encapsulated molecules are already in clinical trials and in the clinic.[217,220] Having demonstrated that the CTPR proteins can be delivered into cells and are still active, future studies might consider delivering CTPRs as mRNA, as there are a number of advantages in terms of formulation, costs and ease of production of mRNA versus protein drugs.

Chapter 9

Design of hetero-bifunctional CTPR proteins

9.1 Introduction

Viruses, as obligate intracellular parasites, have evolved clever ways to manipulate the host proteome, with the aim to maximise the viability of viral proteins and allow efficient viral replication. One striking class of example is viral proteins that are able to simultaneously recruit a host E3 ubiquitin ligase to a host target protein, leading to the poly-ubiquitination of the host protein and its subsequent proteasomal degradation.[221] Copying nature, for the past few years, one of the new frontiers in drug discovery has been the production of hetero-bifunctional compounds promoting selective degradation of a target protein, rather than simply inhibiting the target by binding to it.[222] This approach is particularly applicable to those diseases caused by protein overexpression and accumulation, such as cancer and neurodegenerative disorders.

9.1.1 Targeted protein degradation

Degraders are hetero-bifunctional constructs designed to induce protein degradation of a target of interest. Two major types of degraders have been developed to date: proteolysis targeting chimeras (PROTACs) and molecular glues. A PROTAC is a compound consisting of two small molecule binders connected by a short linker. PROTACs are designed to recruit and bring into close proximity a target of interest and an E3 ubiquitin ligase, by forming a ternary complex.[223] Molecular glues, instead, are a smaller version of PROTAC comprising a single small molecule

that binds both the target and the E3 ligase "gluing" them together.[224] The efficacy of degraders has already been proven *in vivo* against a variety of protein targets involved in several therapeutic areas and recently some PROTACs have entered clinical trials. These targets are all proteins for which there is already a small molecule ligand. Likewise, only a small number (<10) of the over 600 E3 ligases in the human proteome have so far been exploited, namely those with small molecule ligands. As it is very unlikely that any one E3 will be able to effectively degrade every target, access to as many E3s as possible is desirable. The advantages of degradation versus inhibition are numerous, including the fact that degradation is irreversible, PROTACs have been shown to act catalytically and they are likely to be more potent than the inhibitors from which they are built, and lastly the small molecule does not need to bind to a functional site on the protein but can bind anywhere meaning that more proteins can potentially be targeted with PROTACs.

Taking advantage of the modularity and combinatorial assembly of the CTPR repeats, the CTPR platform was exploited to generate modular, hetero-bifunctional constructs having two binding moieties: the TBP loop to engage with TNKS, and one of the binding peptides, known as degrons, to recruit E3 ubiquitin ligases. Like for PROTACs and molecular glues, the proximity between TNKS and the E3 enzyme induced by the ternary complex formation is expected to facilitate the transfer of ubiquitin molecules from the E3 ligase to TNKS, followed by TNKS degradation via the ubiquitin-proteasome system (UPS) system.

9.1.2   The UPS

The ubiquitin-proteasome system (UPS) is a highly regulated process that coordinates intracellular protein turnover. The activity of the UPS is mediated by three enzymes acting sequentially and ubiquitin - a 76-amino acid co-factor that acts as a molecular recognition label. Once a substrate is conjugated with multiple ubiquitin molecules, it is recognised and degraded by the proteasome.[225] The UPS enzymes consists of the E1 (ubiquitin activating enzyme), E2 (ubiquitin conjugating enzyme) and E3 (ubiquitin ligases). In the first step, a cysteine residue in the E1 binds and activates a ubiquitin molecule at its C-terminal glycine in an ATP-dependent manner. Ubiquitin is then transferred to the E2 enzyme and finally the E3 catalyses the transfer of the ubiquitin molecule from the E2 to an exposed residue within the protein substrate. Ubiquitin is mostly transferred onto lysine residues, and less frequently onto cysteine, serine and threonine residues or the amino group at the N-terminal end of the protein substrate. Ubiquitin itself has seven lysine residues that allow linkage of multiple ubiquitin

molecules onto the same substrate to form a polyubiquitin chain. The fate of the ubiquitinated substrate depends on the pattern and distribution of the ubiquitin chain. Chains with ubiquitin molecules linked through lysine at position 48 (K48) is associated with protein degradation, as it is recognised by the eukaryotic 26S proteasome. Subunits of the proteasome bind to the substrate, induce its de-ubiquitination, unfold it and feed it through the central pore of the proteasome where it is cleaved by proteases into small peptide fragments of 3-15 amino acids (Figure 9.1).[226]

The human genome encodes for just a couple of E1 enzymes, ~40 E2s and over 600 E3s. E3 ligases are the most heterogeneous class of enzymes in the UPS, as they are responsible for substrate recognition and specificity. Each E3 ligases recognises a short amino acid sequence, called a degron, on their protein substrates.
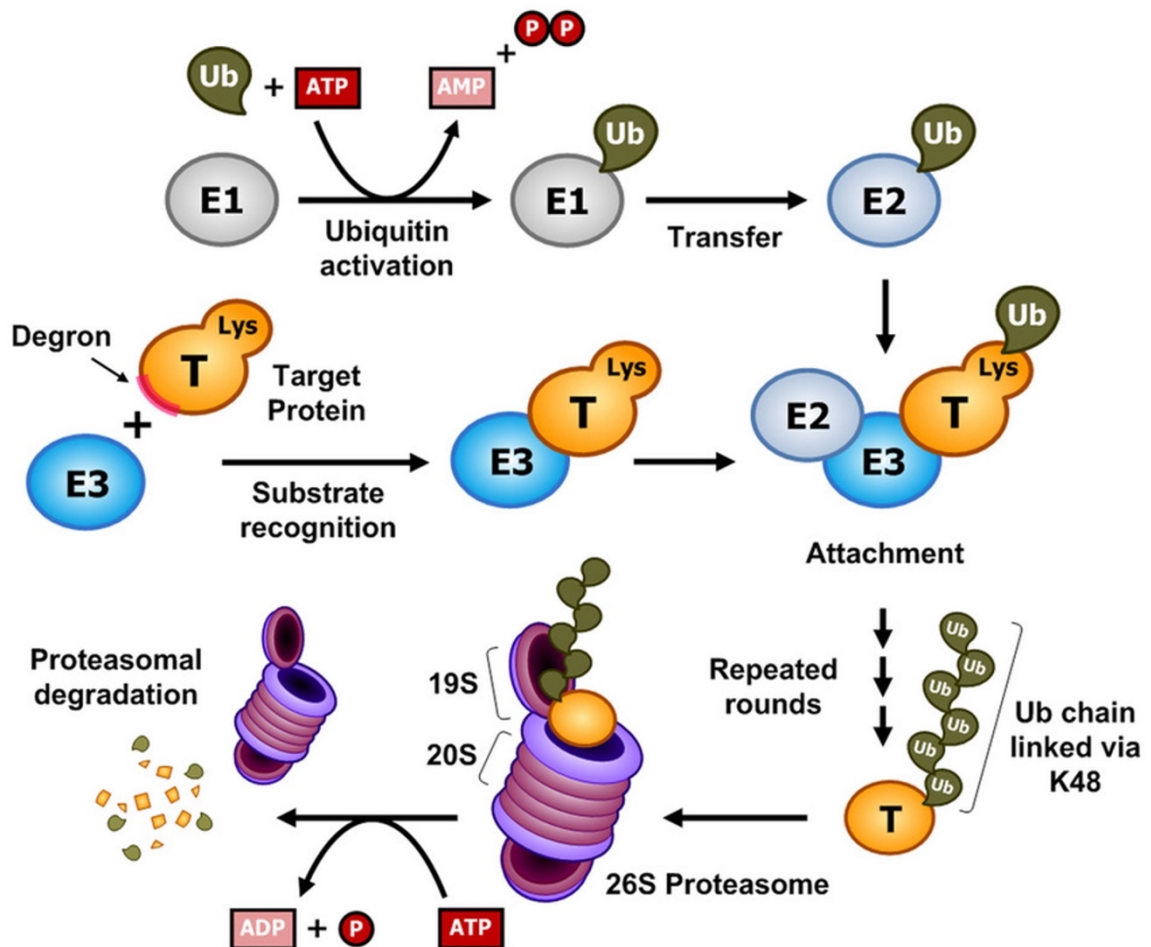
Figure 9.1: The UPS system. Schematic representation of protein degradation mediated by the UPS. Figure adapted from Lee *et al.*[227]

9.1.3   Degrons

Each E3 engages with a specific set of substrates by recognising peptide sequences called degrons. Degrons are short sequences, typically of 5-15 amino-acids, and several have already been identified and structurally characterised in complex with their E3s. Many degrons are regulated by post-translational modifications, most commonly phosphorylation. Degrons can be located anywhere on the surface of the substrate protein, and some proteins contain more than one degron. The presence and type of degron is a factor contributing to the timing and rate of a protein's degradation.[228]

9.1.4   Hetero-bifunctional CTPR/RTPR constructs and experimental design

The modular TPR platform allows the incorporation of one or more functions, such as target binding and E3 binding, in a combinatorial fashion by simply joining multiple binding modules in a specific order. Hetero-bifunctional constructs were therefore generated by introducing a degron at the C-terminal end of the linear 2TBP-CTPR4 construct. Control constructs having no TBP loop or no degron sequence were also made. The gene fragments encoding the degrons of interest were already available in the Itzhaki lab and consist of a degron sequence grafted as a solvent-exposed loop or as a C-terminal α-helix on a RTPR scaffold. The RTPR is a variant of the CTPR scaffold designed in the Itzhaki lab, where all lysine residues have been replaced by arginine residues (R, hence the name RTPR) (Figure 9.2).
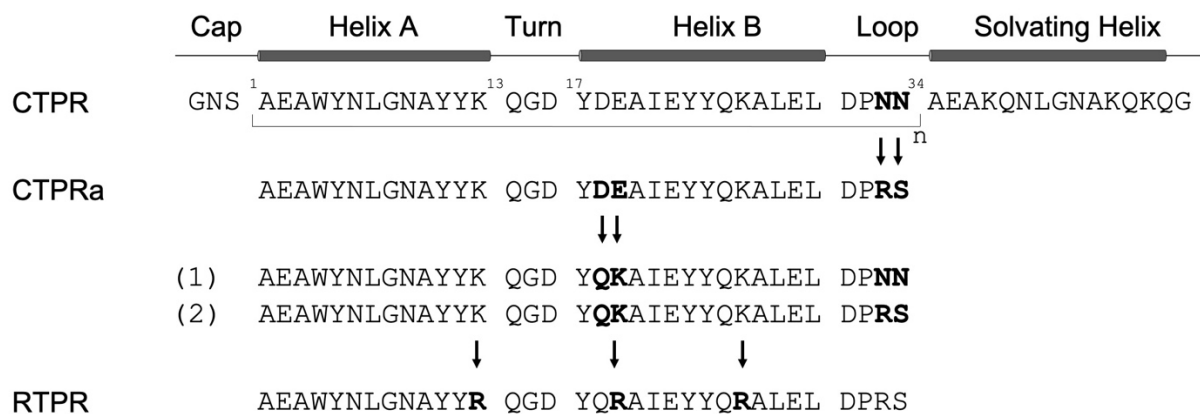
Figure 9.2: RTPR sequence design. The RTPR sequence was generated by mutating the three lysine (K) residues into arginine (R) from the CTPR sequence (2) optimised in the Itzhaki lab.

The rationale for grafting degrons onto the RTPR rather than the CTPR scaffold was to minimise the likelihood of lysine-mediated ubiquitination of the scaffold itself upon E3 recruitment. In this way, the RTPR hetero-bifunctional constructs could potentially act catalytically, meaning that they could be recycled to promote ubiquitination of several target molecules in a consecutive manner, as it was proven for the PROTACs molecules (Figure 9.3).
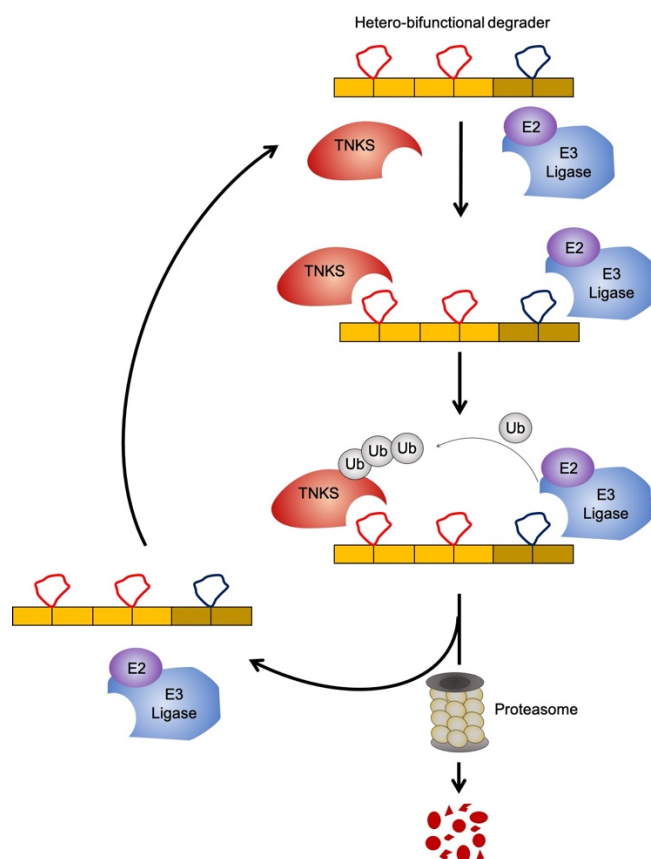
Figure 9.3: Schematic representation of a hetero-bifunctional CTPR/RTPR degrader and its intended mechanism of action. The TNKS-binding degrader is represented as a CTPR/RTPR chimera, by colouring the scaffold in different shades of yellow. Upon ternary complex formation between the degrader, TNKS and the E3-E2 complex, ubiquitin molecules are transferred onto TNKS. TNKS is then recognised and degraded via the proteasome, and the degrader can be recycled.

Preliminary results were obtained by co-transfecting HEK293T cells with a plasmid encoding HiBiT-tagged TNKS2 in combination with a single- or hetero-bifunctional CTPR/RTPR construct. The HiBiT-mediated luminescence signal was then compared across samples. In this experimental set-up, successful TNKS2 degradation is expected to lead to a reduction in the luminescence signal.

9.2 Results

9.2.1 Molecular biology for hetero-bifunctional constructs generation

Hetero-bifunctional constructs were obtained by introducing an RTPR-degron encoding cassette between the 2TBP-CTPR4 gene and the HA tag in the pcDNA3.1(-) plasmid (Thermo Fisher). Given the diversity of E3 ligases, varying in localisation and concentration in different cell lines and throughout the cell cycle, a selection of degrons, and therefore E3 ligases, was adopted to enhance the chances of successful TNKS2 degradation. The degrons used in this study are listed in Table 9.1 and named according to the substrate in which they have been found. The corresponding E3 ubiquitin ligase recruited by each degron is also listed. As controls, constructs having no TBP loops (CTPR2) or no degron sequence were also generated for comparison. For simplicity, hetero-bifunctional constructs were named 2TBP-CTPR4-"degron" and the control constructs CTPR2-"degron". The protein sequence of the hetero-bifunctional and control constructs is provided in Appendix B.

Degron	Sequence	E3 ligase	E3 cellular localisation	Site of grafting
ABBA	SLSSAFHVFEDGNKEN	APC/C	dynamic	Loop
DBOX	PRLPLGDVSNN	APC/C	dynamic	Loop
KEN	SEDKENVPP	APC/C	dynamic	Loop
Nrf2	DPETGEL	Keap1	Cytoplasm, nucleus	Loop
p27	AGSNEQEPKKRS	Skp2	Nucleus, cytoplasm	Loop
p53	FxxxWxxL	MDM2	Nucleus	α-helix
PHYL	LRPVAMVRPTV	SIAH	Cytoplasm, nucleus	Loop
Puc	LACDEVTSTTSSSTA	SPOP	Nucleus	Loop

Table 9.1: List of degrons (names after the substrates from which they were derived) and their corresponding E3 ubiquitin ligases used in this study. The degron sequences and an indication of the cellular localisation of the E3 ligases (according to the UniProt database) are provided. The site of grafting (loop or α-helix) within the RTPR scaffold is also listed.

The hetero-bifunctional constructs therefore obtained are CTPR/RTPR chimeras, as they consist of four CTPRs including the two TBP loops, followed by the RTPRs onto which the degron is grafted (Figure 9.3).

9.2.2   HiBiT-tagged TNKS2 degradation assay

A simple assay involving co-transfection of HEK293T cells with a HiBiT-tagged TNKS2 and a hetero-bifunctional construct was first attempted. The HiBiT-mediated luminesce readout allows us to quantify the levels of the HiBiT-TNKS2 construct. The assay was multiplexed with the CellTiter-Fluor Cell Viability assay to evaluate in parallel any cytotoxicity effect of the constructs, and the luminescence values were then normalised by the cell viability. Although it does not correspond to the ideal experimental set-up, this assay was designed to provide a quick evaluation over the possibility of degrading TNKS2 using hetero-bifunctional constructs, in order to identify the degron with the highest activity for further studies. Hetero-bifunctional constructs were compared to those having no degron peptide (RTPR2) or no TBP loop (CTPR2). No cytotoxicity effects were observed for any of the constructs tested (data not shown). The normalised data were plotted as a percentage relative to the control well treated with empty vector, as shown Figure 9.4. Although all hetero-bifunctional constructs (except for 2TBP-CTPR4-p27) appear to cause a reduction in the HiBiT luminescence signal, a similar reduction was also observed with the control construct 2TBP-CTPR4 having no degron (in plain orange in Figure 9.4). The control CTPR2-degron constructs (containing no TBP) do not appear to cause a reduction in HiBiT-TNKS2 luminescence, but the data were highly variable.
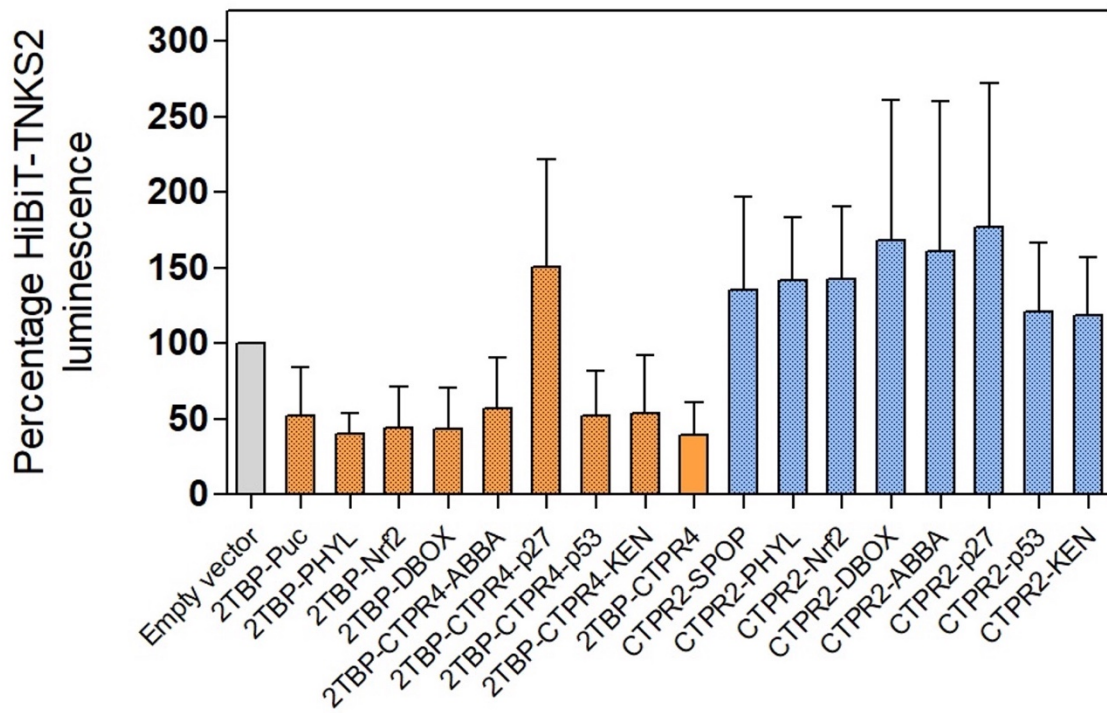
Figure 9.4: Effect of hetero-bifunctional and control CTPR/RTPR constructs on HiBiT-TNKS2 luminescence. Dotted bars correspond to CTPR/RTPR constructs containing a grafted degron. Data were normalised to cell viability and plotted as a percentage relative to the control well treated with empty vector, set at 100%. Error bars corresponding to the standard deviation were determined from three independent sample measurements.

Unfortunately, the results obtained from this assay indicated that none of the degrons when incorporated into the TNKS-binding TPR protein is able to induce a reduction in HiBiT signal above that of the control construct 2TBP-CTPR4. This observation, however, raised the possibility that the TBP loop itself might, somehow, induce lower HiBiT-TNKS2 levels by an unknown mechanism, possibly correlated with the macromolecular assembly formation (as observed in Chapter 6). With this hypothesis in mind, we tested the effect of the different linear $_n$TBP-CTPR$_{2n}$ and trimeric $_n$TBP-CTPR$_{2n}$-foldon constructs (used in Chapter 5-7) on the HiBiT-TNKS2 protein levels. Here, cells were co-transfected with the HiBiT-TNKS2 encoding plasmid in combination with one of the CTPR constructs. In addition, the assay was modified to have the following plasmid combinations both tested on the same plate:

1) HiBiT-TNKS2 + CTPR
2) HiBiT-TNKS2 + HiBiT-CTPR

The higher luminescence intensity observed for the second set of transfections is the sum of HiBiT-tagged TNKS2 and HiBiT-tagged CTPR construct. The assay was designed so that the difference in luminescence between the two sets of transfections should provide a measure of the HiBiT-tagged CTPR proteins levels, assuming that the 11 amino acid HiBiT tag does not alter their expression levels. Results from the first set of transfections are shown in Figure 9.5A. They show reduced levels of HiBiT-tagged TNKS2 in the presence of all of the linear $_n$TBP-CTPR$_{2n}$ or trimeric $_n$TBP-CTPR$_{2n}$-foldon constructs. This result is in agreement with the previously observed effect of 2TBP-CTPR4 on the HiBiT-TNKS2 signal (Figure 9.4), but it does not provide any additional insights into the underlying mechanism. When the difference in luminescence readout between the two identical sets of transfections was calculated, the signal corresponding to the amount of each HiBiT-tagged CTPR construct is obtained (Figure 9.5B). The CTPR protein levels obtained are consistent with those observed in the TopFlash assay (i.e. levels were higher for the larger and trimeric TBP-CTPR constructs, as shown in Chapter 7, Figure 7.5C). Overall, with this assay it is not possible to conclude whether any degron sequence or the TBP itself have an effect on TNKS2 protein levels. It is also possible that TNKS2 expression is differentially affected by co-transfection with the different TPR constructs.
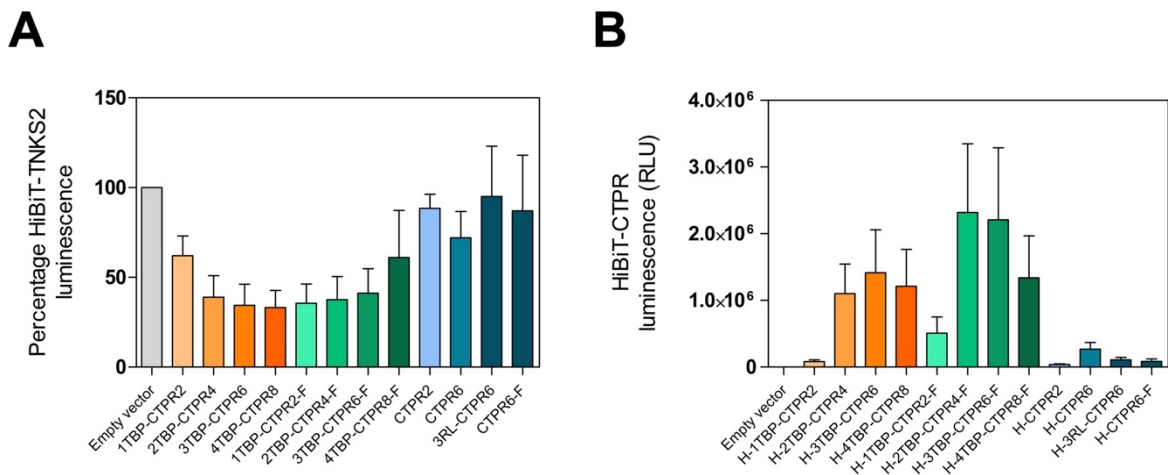
Figure 9.5: Effect of single-functional linear $_n$TBP-CTPR$_{2n}$ and trimeric $_n$TBP-CTPR$_{2n}$-foldon constructs on HiBiT-TNKS2 protein levels. (A) Effect of CTPR constructs on the HiBiT-TNKS2 mediated luminescence. Data were normalised to cell viability and plotted as a percentage relative to the control well treated with empty vector, set at 100%. Error bars corresponding to the standard deviation were determined from three independent sample measurements. (B) Luminescence difference between the two sets of transfections (HiBiT-CTPR vs CTPR). Error bars corresponding to the standard deviation were determined from three independent sample measurements.

9.3 Discussion

In this Chapter, the design of hetero-bifunctional TPR proteins was described. This approach, as with PROTACs, has tremendous potential because of its fundamental advantage in inducing target degradation rather than just inhibition. This approach should provide a more robust and durable pharmacological effect compared to target inhibition, that more often leads to a transient response and is unable to prevent the scaffolding functions of a target protein. Moreover, as degraders can act catalytically, they can be administered at a lower dose, therefore causing fewer side effects, if any.

There are several important differences between PROTACs and our hetero-bifunctional proteins. For PROTACs, fewer than ten of the over 600 E3s in the human proteome have small molecule ligands available, meaning that the E3 toolbox is extremely limited. It is very unlikely that a single E3 will be able to effectively degrade every target, and therefore access to as many E3s as possible is desirable. Unlike small molecules, there are naturally occurring and structurally characterised degrons for many E3s that can be grafted onto our platform, expanding its E3 repertoire compared to PROTACs. Additionally, PROTACs can be designed only against those protein targets for which a selective small molecule is already available. Our approach instead has no such limitations, because we can exploit the much larger number of known short binding peptides to many targets identified from naturally occurring protein-protein interactions. This is particularly relevant for those target proteins considered undruggable by conventional small molecules.

An alternative TPR scaffold, named RTPR, in which lysine residues were replaced with arginine (R), was designed in the Itzhaki group to prevent poly-ubiquitination and degradation of the protein scaffold. This should prevent degradation of the hetero-bifunctional RTPR protein and potentially allow it to act catalytically. In order to do so, the construct will need to disengage from one target molecule to engage with another. For this reason, it has been proposed that the binding affinity between the hetero-bifunctional construct and its targets should not be too high. Indeed, a recent study showed that there was little correlation between target engagement and potency of degradation.[229] The TBP loop, with its TNKS-binding affinity of 14 μM, might fulfil this requirement. Overall, it may be challenging to achieve ternary complex formation between the degrader, the target and the E3 ligase that is productive for target degradation. Several factors need to be considered:

- Each component of the complex needs to be positioned in such a way as to avoid steric hindrance but at the same time facilitate the efficient transfer of ubiquitin onto the

target. To date, researchers in the PROTAC field cannot predict the best geometrical arrangement (which will be different for each target and for each E3) for achieving ubiquitination. Moreover, not all ubiquitination will lead to degradation – the site of ubiquitination on the target will affect how the target is presented to the proteasome and thereby how easily it is unfolded and degraded. Thus, we currently need to find the best combination of target binder, E3-binder and 'linker' empirically.

- The presence of the E3 ligase in the correct cellular localisation and at levels required to elicit successful degradation need to be confirmed. The appropriate degron needs to be identified, grafted onto the RTPR scaffold, and characterised biophysically. Further work may be needed to also identify the optimal degron-E3 binding affinity.

- Finally, the choice of the protein target might also influence the effectiveness of the strategy. TNKS2 is a very large (127 kDa) protein and its degradation via the proteasome might be more challenging to achieve through an artificial system than for degradation of smaller proteins.

Unfortunately, due to the short time available for this part of the project, it was not possible to evaluate in detail each of the factors mentioned above. From an experimental perspective, cellular co-transfection in combination with the HiBiT technology was developed as a high-throughput screening method to identify successful degron sequences and provide a quick indication on the possibility of targeting TNKS to the proteasome. However, plasmids co-transfection does not represent the ideal scenario for testing protein degradation. Although using the HiBiT tag provides sensitive, quantitative and highly specific measurements (unlike anti-TNKS antibodies), TNKS2 protein produced via the transfected pcDNA3.1(-) plasmid is expressed throughout the incubation time, potentially masking any degradation. A better experimental set-up would take advantage of the CRISPR/Cas9 technology to generate a genetically modified cell line expressing the target of interest fused to the HiBiT tag, as has been done by Promega.[203] Alternatively, an inducible system for target expression could be used. In addition, the dose- and time-dependent effects of the degraders will need to be assessed. Thus, the data shown in this Chapter are preliminary results that need to be further validated in combination with assay development, before drawing any conclusions. All these considerations taken together laid the foundation for a spin-out company, PolyProx Therapeutics, focusing on discovering biotherapeutic molecules against hard-to-drug targets that have no known small molecule ligands and are therefore not accessible to small molecule PROTACs.

Chapter 10

Final discussion and future work

10.1 General discussion

Engineered tandem repeat proteins have been exploited recently by several groups for biotechnological purposes ranging from biomaterials to biotherapeutics development.[38,44,79] Despite their advantageous properties, CTPR repeats have been extensively used as building blocks for novel biomaterial applications but very little as scaffolds for biotherapeutic development. In this proof-of-concept study, we demonstrated the tremendous potential of the CTPR as a novel biotherapeutic scaffold. Using a simple "cut-and-paste" approach, CTPRs can be engineered to display single or multiple copies of short peptide motifs of known binding activities. Moreover, the simple loop and helix grafting approaches adopted by the Itzhaki group may require little or no in silico design procedures and should be generalisable to any of the 100,000+ motifs present in the human proteome. Thus, the platform has the potential to provide a useful tool for various cellular applications ranging from target validation, pathway modulation and ultimately molecular therapeutics.

Grafting binding motifs onto a structural scaffold has several uses: it increases the peptides' resistance to proteolysis and allows one to control their conformation and, when multiple peptides are grafted, their orientation relative to each other. The CTPR scaffold also provides additional advantages over other scaffolds:

- CTPR proteins are small and ultra-stable. As a result, they can accommodate small or even large sequence insertions yet remain folded and stable. As they do not require disulphide bonds for stability, they are not affected by the redox environment inside and outside the cells.

- As a result of their high stability and absence of disulphide bonds, they can be easily and cheaply produced recombinantly in *E. coli* in very high yields.

- The repetitive, modular nature of the scaffold is a particularly unique selling point, as each repeat can be engineered to display the same or different binding motifs and then combined to create multivalent or multi-functional molecules, or repeats can act as a spacer between the different binding motifs, all without compromising the overall fold. The rigidity of the CTPR scaffold ultimately works as a molecular ruler of predefined length and pitch, providing the possibility to control the geometrical arrangement in which the binding moieties are displayed.

- Lastly, unlike other artificial binding proteins, the modular CTPR scaffold allows multi-valency and multi-functionality to be achieved without having to join two separate domains together (which can create liabilities in terms of protease sensitivity of the linker between the domains).

Here we have shown that the multivalent capability of the CTPR scaffold can be achieved by grafting multiple TBPs (TNKS binding peptides) onto a CTPR array and can be further enhanced by trimerisation through the foldon motif. As detailed in Chapter 3, CTPRs can be functionalised by grafting binding motifs onto the loops between adjacent repeats or onto the terminal helices. Examples of both grafting methods were provided: the TBP was grafted onto alternating inter-repeat loops, and the foldon domain was grafted onto a C-terminal helix. In Chapters 4 and 5, we have shown that CTPR proteins with grafted TBP motifs are stable, correctly folded and can bind TNKS *in vitro* and in the cell. Interestingly, the binding between multivalent TBP-CTPR proteins and full-length TNKS resulted in the formation of large macromolecular assemblies with diverse internal dynamics depending on the linear or trimeric configuration of the TBP-CTPR proteins (Chapter 6). In Chapter 7 these engineered TBP-CTPR proteins were shown to inhibit TNKS activity and thereby the Wnt signalling pathway by blocking the substrate-binding ARC subdomains rather than inhibiting the catalytic activity of TNKS. Moreover, experiments using fusogenic liposomes to deliver the multivalent 3TBP-CTPR6 protein (Chapter 8) demonstrated rapid inhibition of Wnt pathway activity within six hours of treatment. The greater effectiveness of liposome-delivered 3TBP-CTPR6 for interventional inhibition of Wnt pathway activity compared to the small molecules TNKSi further highlights the potential of this strategy for targeting TNKS. It is likely that the high level of TNKS inhibition observed is a result of the combination of the multivalency of the constructs and the sequestering of TNKS within macromolecular assemblies. Such a mechanism of action warrants further investigation and should be explored for future

therapeutic efforts against TNKS and other multivalent targets. Only recently, in 2019, the Guettler group published a study aimed at identifying small molecule TNKS antagonists targeting the substrate-binding ARC subdomains using fragment-based screening.[230] The small molecule inhibitors identified bind the ARCs at the same site as the TBP but with a very low affinity (K_d values higher than 1 mM), and no functional data were provided.[230] The study further underlines the need for a peptide- or protein-based biotherapeutic to enable extended and higher-affinity binding interactions with the ARC subdomains and thereby inhibit TNKS function. Last, in Chapter 9, the ability to graft multiple peptides into the same CTPR protein was exploited to generate hetero-bifunctional constructs. Hetero-bifunctional biotherapeutics aiming to bridge a TNKS molecule to an E3 ligase were designed as degraders to enhance their inhibitory effect relative to the TBP-CTPR proteins. Preliminary results did not show degradation of TNKS, and further work will be needed to pursue this aim.

10.2 Future work

Future work should be focused on further optimisation of the TNKS-binding CTPR protein design and on improving TNKS inhibition via multivalency.

Such optimisation could be achieved from multiple approaches:

1. Computational modelling of the multivalent TBP-CTPR constructs to redesign them so that the spacing of the TBPs matches the spacing and orientation of the consecutive ARCs of TNKS, with the aim to increase the binding affinity and achieve avidity of binding. However, it will be challenging to determine the exact arrangement required because of the lack of a crystal structure of the full-length TNKS ANK domain and its known high degree of flexibility. Eisemann *et al*. recently tested the functional cooperation between different ARC combinations when binding the multivalent Axin1 substrate.[89] A similar analysis of the interaction of the current and redesigned TBP-CTPR proteins could help us to understand how many TBP loops can engage with the target simultaneously and whether the multivalency results in avidity of binding.

2. It is clear from the crystal structure of TNKS2 ARC4 in complex with TNKS-binding peptides that the peptides bind in an extended conformation. The CTPR scaffold might be constraining the TBP sequence in a sub-optimal, turn-like conformation, possibly explaining the weaker TNKS-binding affinity observed for the CTPR-grafted TBP compared with the isolated peptide and the chemically constrained peptides generated

in our previous study.[46,85] Therefore, it is possible that the TBP loop in the CTPR scaffold should be lengthened by introducing additional flanking residues at either side of the 8-residue TBP so that it can adopt a more extended, elongated conformation. This might result in an increase in binding affinity. Ideally, a computational protein design algorithm like Rosetta should also be used to find the optimal grafting of the TBP onto the CTPR scaffold.[231,232]

Future work investigating the effect of multivalency on TNKS inhibition should, instead, look at the following:

1. The formation of the large, intracellular assemblies observed for the overexpressed TNKS in the presence of the multivalent TBP-CTPR proteins should be characterised further. In particular, it will be important to determine whether these assemblies also form with the endogenous rather than the overexpressed TNKS. This would validate our hypothesis that the potent TNKS inhibition observed with the multivalent TBP-CTPRs occurs by sequestering TNKS into these assemblies. In addition, their biophysical properties should be studied further, in particular to determine whether they undergo liquid-liquid phase separation and form membraneless organelles. A recent paper showed that small molecule TNKSi (G007-LK) result in the formation of cytoplasmic puncta (degradasomes) comprising the inactivated TNKS and components of the ßDC destruction complex (ßDC), thereby leading to enhanced degradation of ß-catenin. Degradasomes are highly dynamic, mobile and membrane-free structures, in which ß-catenin is rapidly turned over.[233] It is possible that the large assemblies that we have observed also contain components of the ßDC. Alternatively, our TNKS inhibitors could be acting by sequestering TNKS in macromolecular assemblies distinct from the ßDC, and thus the composition of our assemblies warrants further investigation. These studies may also help us to determine how our TNKS inhibitors can act interventionally, whereas colorectal cancer cells and cells after prolonged Wnt stimulation are unresponsive to small molecule TNKSi.

2. Additional hetero-bifunctional TBP-CTPRs should be designed to inhibit TNKS by inducing its degradation. First, it would be useful to identify the most appropriate E3 ligase for TNKS, according to expression levels in cancer versus normal cells for example. Second, structural modelling would be useful to predict the conformation of the ternary complex and thereby help to design the hetero-bifunctional CTPR proteins. However, in the case of TNKS, its very large size and the absence of the full-length

crystal structure would make this approach impossible. Nevertheless, modelling could be used to understand how the binary complexes (CTPR and E3, CTPR and TNKS ARC subdomain) also form. Third, if we verify that the multivalent TBP-CTPR proteins form biomolecular condensates in the presence of endogenous TNKS, then hetero-bifunctional degraders might have enhanced activity within these condensates as the enzyme (the E3) and the substrate (TNKS) are present at higher concentrations. This approach is currently being explored in the Itzhaki lab. Lastly, an improved experimental method will be required to evaluate the proteasomal degradation of TNKS. One option would be to generate a clinically relevant cell line stably expressing HiBiT-tagged TNKS in which hetero-bifunctional constructs in variable concentrations could be tested. If successful, constructs could then be tested in higher complexity models, such as organoids, and animal models.

In the long term, several other properties of the CTPR scaffold need to be investigated. First, the consensus CTPR sequence and the RTPR might require optimisation to avoid the possible activation of the human immune response. Second, the delivery method is most likely to rely on encapsulation of mRNA rather than protein (preliminary findings from PolyProx Therapeutics). Third, pharmacokinetics and biodistribution of these biotherapeutics will also need to be tackled using the most up-to-date technologies. Finally, resistance to intracellular proteolytic and proteasomal degradation should be evaluated.

To conclude, the exciting results obtained in this thesis pave the way to further investigations into the potential of the CTPR array as a scaffold for biotherapeutic applications.

References

1.   Chiu, M. L., Goulet, D. R., Teplyakov, A. & Gilliland, G. L. Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies* **8**, 55 (2019).

2.   Saeed, A. F. U. H., Wang, R., Ling, S. & Wang, S. Antibody engineering for pursuing a healthier future. *Frontiers in Microbiology* vol. 8 (2017).

3.   Topin, I. Monoclonal antibodies – all you need to know about antibody generation | tebu-bio's blog. https://www.tebu-bio.com/blog/2018/11/13/monoclonal-antibodies-all-you-need-to-know-about-antibody-generation/.

4.   Mammen, M., Choi, S. K. & Whitesides, G. M. Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew. Chemie - Int. Ed.* **37**, 2754–2794 (1998).

5.   Olsen, J. G., Teilum, K. & Kragelund, B. B. Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on fuzziness. *Cellular and Molecular Life Sciences* vol. 74 3175–3183 (2017).

6.   Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).

7.   Fuxreiter, M. Fuzziness: Linking regulation to protein dynamics. *Molecular BioSystems* vol. 8 168–177 (2012).

8.   Cuesta, A. M., Sainz-Pastor, N., Bonet, J., Oliva, B. & Alvarez-Vallina, L. A. Multivalent antibodies: when design surpasses evolution. *Trends Biotechnol.* **28**, 355–362 (2010).

9.   Labrijn, A. F., Janmaat, M. L., Reichert, J. M. & Parren, P. W. H. I. Bispecific antibodies: a mechanistic review of the pipeline. *Nat. Rev. Drug Discov.* **18**, 585–608 (2019).

10.   Groß, A., Hashimoto, C., Sticht, H. & Eichler, J. Synthetic peptides as protein mimics. *Front. Bioeng. Biotechnol.* **3**, (2016).

11.   Dirksen, A., Meijer, E. W., Adriaens, W. & Hackeng, T. M. Strategy for the synthesis of multivalent peptide-based nonsymmetric dendrimers by native chemical ligation. *Chem. Commun.* 1667–1669 (2006) doi:10.1039/b600286b.

12.   Englund, E. A. *et al.* Programmable multivalent display of receptor ligands using peptide nucleic acid nanoscaffolds. *Nat. Commun.* **3**, (2012).

13. Stephanopoulos, N. Peptide-Oligonucleotide Hybrid Molecules for Bioactive Nanomaterials. *Bioconjug. Chem.* **30**, 1915–1922 (2019).

14. Tiede, C. *et al.* Adhiron: A stable and versatile peptide display scaffold for molecular recognition applications. *Protein Eng. Des. Sel.* **27**, 145–155 (2014).

15. Reverdatto, S., Burz, D. S. & Shekhtman, A. Peptide Aptamers: Development and Applications. *Curr Top Med Chem* **15**, 1082–1101 (2015).

16. Schlatter, D. *et al.* Generation, characterization and structural data of chymase binding proteins based on the human Fyn kinase SH3 domain. *MAbs* **4**, 497–508 (2012).

17. Sonntag, M. H., Schill, J. & Brunsveld, L. Integrin-Targeting Fluorescent Proteins: Exploration of RGD Insertion Sites. *ChemBioChem* **18**, 441–443 (2017).

18. Pacheco, S. *et al.* Affinity transfer to the archaeal extremophilic Sac7d protein by insertion of a CDR. *Protein Eng. Des. Sel.* **27**, 431–438 (2014).

19. Nielsen, F. S. *et al.* Insertion of foreign T cell epitopes in human tumor necrosis factor α with minimal effect on protein structure and biological activity. *J. Biol. Chem.* **279**, 33593–33600 (2004).

20. Rossmann, M., Greive, S. J., Moschetti, T., Dinan, M. & Hyvönen, M. Development of a multipurpose scaffold for the display of peptide loops. *Protein Eng. Des. Sel.* **30**, 419–430 (2017).

21. Crook, Z. R., Nairn, N. W. & Olson, J. M. Miniproteins as a Powerful Modality in Drug Development. *Trends in Biochemical Sciences* vol. 45 332–346 (2020).

22. Binz, H. K., Amstutz, P. & Plückthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nature Biotechnology* vol. 23 1257–1268 (2005).

23. Nicaise, M., Valerio-Lepiniec, M., Minard, P. & Desmadril, M. Affinity transfer by CDR grafting on a nonimmunoglobulin scaffold. *Protein Sci.* **13**, 1882–1891 (2004).

24. Chevalier, A. *et al.* Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).

25. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* vol. 537 320–327 (2016).

26. Baker, E. G., Bartlett, G. J., Porter Goff, K. L. & Woolfson, D. N. Miniprotein Design: Past, Present, and Prospects. *Acc. Chem. Res.* **50**, 2085–2092 (2017).

27. Silva, D. A. *et al.* De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).

28. Hober, S., Lindbo, S. & Nilvebrant, J. Bispecific applications of non-immunoglobulin scaffold binders. *Methods* **154**, 143–152 (2019).

29.	Boersma, Y. L., Chao, G., Steiner, D., Dane Wittrup, K. & Plückthun, A. Bispecific Designed Ankyrin Repeat Proteins (DARPins) Targeting Epidermal Growth Factor Receptor Inhibit A431 Cell Proliferation and Receptor Recycling. *J. Biol. Chem.* **286**, 41273–85 (2011).

30.	Rönnmark, J. *et al.* Construction and characterization of affibody-Fc chimeras produced in Escherichia coli. *J. Immunol. Methods* **261**, 199–211 (2002).

31.	Bass, T. Z. *et al.* In vivo evaluation of a novel format of a bivalent HER3-targeting and albumin-binding therapeutic affibody construct. *Sci. Rep.* **7**, (2017).

32.	Fleetwood, F. *et al.* Simultaneous targeting of two ligand-binding sites on VEGFR2 using biparatopic Affibody molecules results in dramatically improved affinity. *Sci. Rep.* **4**, (2014).

33.	Willuda, J. *et al.* Tumor Targeting of Mono-, Di-, and Tetravalent Anti-p185HER-2 Miniantibodies Multimerized by Self-associating Peptides. *J. Biol. Chem.* **276**, 14385–14392 (2001).

34.	Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).

35.	Pellegrini, M., Marcotte, E. M. & Yeates, T. O. A fast algorithm for genome-wide analysis of proteins with repeated sequences. *Proteins Struct. Funct. Genet.* **35**, 440–446 (1999).

36.	Kajava, A. V. Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol.* **179**, 279–288 (2012).

37.	Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. Protein repeats: Structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131 (2001).

38.	Javadi, Y. & Itzhaki, L. S. Tandem-repeat proteins: Regularity plus modularity equals design-ability. *Current Opinion in Structural Biology* vol. 23 622–631 (2013).

39.	Cortajarena, A. L., Yi, F. & Regan, L. Designed TPR modules as novel anticancer agents. *ACS Chem. Biol.* **3**, 161–166 (2008).

40.	Grove, T. Z., Hands, M. & Regan, L. Creating novel proteins by combining design and selection. *Protein Eng. Des. Sel.* **23**, 449–455 (2010).

41.	Phillips, J. J., Millership, C. & Main, E. R. G. Fibrous nanostructures from the self-assembly of designed repeat protein modules. *Angew. Chemie - Int. Ed.* **51**, 13132–13135 (2012).

42.	Sanchez-deAlcazar, D., Mejias, S. H., Erazo, K., Sot, B. & Cortajarena, A. L. Self-assembly of repeat proteins: Concepts and design of new interfaces. *J. Struct. Biol.*

201, 118–129 (2018).

43. Mejias, S. H., Aires, A., Couleaud, P. & Cortajarena, A. L. Designed repeat proteins as building blocks for nanofabrication. *Adv. Exp. Med. Biol.* **940**, 61–81 (2016).

44. Plückthun, A. Designed ankyrin repeat proteins (DARPins): Binding proteins for research, diagnostics, and therapy. *Annual Review of Pharmacology and Toxicology* vol. 55 489–511 (2015).

45. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).

46. Guettler, S. *et al.* Structural basis and sequence rules for substrate recognition by tankyrase explain the basis for cherubism disease. *Cell* **147**, 1340–1354 (2011).

47. Xing, Y. *et al.* Crystal Structure of a Full-Length β-Catenin. *Structure* **16**, 478–487 (2008).

48. Parashar, V., Jeffrey, P. D. & Neiditch, M. B. Conformational Change-Induced Repeat Domain Expansion Regulates Rap Phosphatase Quorum-Sensing Signal Receptors. *PLoS Biol.* **11**, (2013).

49. Cho, U. S. & Xu, W. Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature* **445**, 53–57 (2007).

50. Kobe, B. & Deisenhofer, J. Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. *J. Mol. Biol.* **264**, 1028–1043 (1996).

51. Leinala, E. K., Davies, P. L. & Jia, Z. Crystal structure of β-Helical antifreeze protein points to a general ice binding model. *Structure* **10**, 619–627 (2002).

52. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

53. Wu, Y. *et al.* Rigidly connected multispecific artificial binders with adjustable geometries. *Sci. Rep.* **7**, 11217 (2017).

54. Das, A. K., Cohen, P. T. W. & Barford, D. The structure of the tetratricopeptide repeats of protein phosphatase 5: Implications for TPR-mediated protein-protein interactions. *EMBO J.* **17**, 1192–1199 (1998).

55. Sikorski, R. S., Boguski, M. S., Goebl, M. & Hieter, P. A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell* **60**, 307–317 (1990).

56. Hirano, T., Kinoshita, N., Morikawa, K. & Yanagida, M. Snap helix with knob and hole: Essential repeats in S. pombe nuclear protein nuc2 +. *Cell* **60**, 319–328 (1990).

57. Zeytuni, N. & Zarivach, R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* vol. 20 397–405 (2012).

58. Blatch, G. L. & Lässle, M. The tetratricopeptide repeat: A structural motif mediating protein-protein interactions. *BioEssays* vol. 21 932–939 (1999).

59. D'Andrea, L. D. & Regan, L. TPR proteins: The versatile helix. *Trends in Biochemical Sciences* vol. 28 655–662 (2003).

60. Perez-Riba, A. & Itzhaki, L. S. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Current Opinion in Structural Biology* vol. 54 43–49 (2019).

61. Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. Design of stable α-helical arrays from an idealized TPR motif. *Structure* **11**, 497–508 (2003).

62. Mól, A. R., Fontes, W. & Castro, M. S. http://lbqp.unb.br/NetWheels/.

63. Jínek, M. *et al.* The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. *Nat. Struct. Mol. Biol.* **11**, 1001–1007 (2004).

64. Kajander, T., Cortajarena, A. L., Mochrie, S. & Regan, L. Structure and stability of designed TPR protein superhelices: Unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 800–811 (2007).

65. Main, E. R. G., Stott, K., Jackson, S. E. & Regan, L. Local and long-range stability in tandemly arrayed tetratricopeptide repeats. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5721–5726 (2005).

66. Cortajarena, A. L., Wang, J. & Regan, L. Crystal structure of a designed tetratricopeptide repeat module in complex with its peptide ligand. *FEBS J.* **277**, 1058–1066 (2010).

67. Cortajarena, A. L., Kajander, T., Pan, W., Cocco, M. J. & Regan, L. Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins. *Protein Eng. Des. Sel.* **17**, 399–409 (2004).

68. Jackrel, M. E., Valverde, R. & Regan, L. Redesign of a protein-peptide interaction; Characterization and applications. *Protein Sci.* **18**, 762–774 (2009).

69. Cortajarena, A. L., Liu, T. Y., Hochstrasser, M. & Regan, L. Designed proteins to modulate cellular networks. *ACS Chem. Biol.* **5**, 545–552 (2010).

70. Mejías, S. H., Sot, B., Guantes, R. & Cortajarena, A. L. Controlled nanometric fibers of self-assembled designed protein scaffolds. *Nanoscale* **6**, 10982–10988 (2014).

71. Grove, T. Z., Osuji, C. O., Forster, J. D., Dufresne, E. R. & Regan, L. Stimuli-

responsive smart gels realized via modular protein design. *J. Am. Chem. Soc.* **132**, 14024–14026 (2010).

72. Grove, T. Z., Regan, L. & Cortajarena, A. L. Nanostructured functional films from engineered repeat proteins. *J. R. Soc. Interface* **10**, 20130051 (2013).

73. Mejías, S. H. *et al.* Repeat protein scaffolds: Ordering photo- and electroactive molecules in solution and solid state. *Chem. Sci.* **7**, 4842–4847 (2016).

74. Aires, A., Lopez-Martinez, E. & Cortajarena, A. L. Sensors based on metal nanoclusters stabilized on designed proteins. *Biosensors* **8**, (2018).

75. Liutkus, M. *et al.* Protein-directed crystalline 2D fullerene assemblies. *Nanoscale* **12**, 3614–3622 (2020).

76. Mejias, S. H. *et al.* Repeat proteins as versatile scaffolds for arrays of redox-active FeS clusters. *Chem. Commun.* **55**, 3319–3322 (2019).

77. Aires, A. *et al.* A Simple Approach to Design Proteins for the Sustainable Synthesis of Metal Nanoclusters. *Angew. Chemie - Int. Ed.* **58**, 6214–6219 (2019).

78. Sánchez-deAlcázar, D., Velasco-Lozano, S., Zeballos, N., López-Gallego, F. & Cortajarena, A. L. Biocatalytic Protein-Based Materials for Integration into Energy Devices. *ChemBioChem* **20**, 1977–1985 (2019).

79. Romera, D., Couleaud, P., Mejias, S. H., Aires, A. & Cortajarena, A. L. Biomolecular templating of functional hybrid nanostructures using repeat protein scaffolds. *Biochem. Soc. Trans.* **43**, 825–831 (2015).

80. Perez-Riba, A., Lowe, A. R., Main, E. R. G. & Itzhaki, L. S. Context-Dependent Energetics of Loop Extensions in a Family of Tandem-Repeat Proteins. *Biophys. J.* **114**, 2552–2562 (2018).

81. Ripka, J., Perez-Riba, A., Chaturbedy, P. K. & Itzhaki, L. Testing the length limit of loop grafting in a helical repeat protein. *Curr. Res. Struct. Biol.* (2020) doi:10.1016/j.crstbi.2020.12.002.

82. Madden, S. K., Perez-Riba, A. & Itzhaki, L. S. Exploring new strategies for grafting binding peptides onto protein loops using a consensus-designed tetratricopeptide repeat scaffold. *Protein Sci.* **28**, 738–745 (2019).

83. Sia, S. K. & Kim, P. S. Protein grafting of an HIV-1-inhibiting epitope. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9756–9761 (2003).

84. Chin, J. W. & Schepartz, A. Design and Evolution of a Miniature Bcl-2 Binding Protein. *Angew. Chemie Int. Ed.* **40**, 3806–3809 (2001).

85. Xu, W. *et al.* Macrocyclized Extended Peptides: Inhibiting the Substrate-Recognition

Domain of Tankyrase. *J. Am. Chem. Soc.* **139**, 2245–2256 (2017).

86.   Gibson, B. A. & Kraus, W. L. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nature Reviews Molecular Cell Biology* vol. 13 411–424 (2012).

87.   Smith, S., Giriat, I., Schmitt, A. & De Lange, T. Tankyrase, a poly(ADP-ribose) polymerase at human telomeres. *Science (80-.).* **282**, 1484–1487 (1998).

88.   Lyons, R. J. *et al.* Identification of a Novel Human Tankyrase through Its Interaction with the Adaptor Protein Grb14. *J. Biol. Chem.* **276**, 17172–17180 (2001).

89.   Eisemann, T. *et al.* Tankyrase-1 Ankyrin Repeats Form an Adaptable Binding Platform for Targets of ADP-Ribose Modification. *Structure* **24**, 1679–1692 (2016).

90.   Seimiya, H. & Smith, S. The telomeric poly(ADP-ribose) polymerase, tankyrase 1, contains multiple binding sites for telomeric repeat binding factor 1 (TRF1) and a novel acceptor, 182-kDa tankyrase-binding protein (TAB182). *J. Biol. Chem.* **277**, 14116–14126 (2002).

91.   Seimiya, H., Muramatsu, Y., Smith, S. & Tsuruo, T. Functional Subdomain in the Ankyrin Domain of Tankyrase 1 Required for Poly(ADP-Ribosyl)ation of TRF1 and Telomere Elongation. *Mol. Cell. Biol.* **24**, 1944–1955 (2004).

92.   DaRosa, P. A., Klevit, R. E. & Xu, W. Structural basis for tankyrase-RNF146 interaction reveals noncanonical tankyrase-binding motifs. *Protein Sci.* **27**, 1057–1067 (2018).

93.   Mariotti, L. *et al.* Tankyrase Requires SAM Domain-Dependent Polymerization to Support Wnt-β-Catenin Signaling. *Mol. Cell* **63**, 498–513 (2016).

94.   Riccio, A. A., McCauley, M., Langelier, M. F. & Pascal, J. M. Tankyrase Sterile α Motif Domain Polymerization Is Required for Its Role in Wnt Signaling. *Structure* **24**, 1573–1581 (2016).

95.   Sbodio, J. I., Lodish, H. F. & Chi, N. W. Tankyrase-2 oligomerizes with tankyrase-1 and binds to both TRF1 (telomere-repeat-binding factor 1) and IRAP (insulin-responsive aminopeptidase). *Biochem. J.* **361**, 451–459 (2002).

96.   Xu, D. *et al.* USP25 regulates WNT signalling by controlling the stability of tankyrases. *Genes Dev.* **31**, 1024–1035 (2017).

97.   Lehtiö, L. *et al.* Zinc Binding Catalytic Domain of Human Tankyrase 1. *J. Mol. Biol.* **379**, 136–145 (2008).

98.   Karlberg, T. *et al.* Structural basis for the interaction between tankyrase-2 and a potent Wnt-signaling inhibitor. *J. Med. Chem.* **53**, 5352–5355 (2010).

99. Guettler, S. AXIN Shapes Tankyrase ARChitecture. *Structure* vol. 24 1625–1627 (2016).

100. Gelmini, S. *et al.* Tankyrase, a positive regulator of telomere elongation, is over expressed in human breast cancer. *Cancer Lett.* **216**, 81–87 (2004).

101. Gelmini, S. *et al.* Distribution of Tankyrase-1 mRNA expression in colon cancer and its prospective correlation with progression stage. *Oncol. Rep.* **16**, 1261–1266 (2006).

102. Gelmini, S. *et al.* Tankyrase-1 mRNA expression in bladder cancer and paired urine sediment: Preliminary experience. *Clin. Chem. Lab. Med.* **45**, 862–866 (2007).

103. Shebzukhov, Y. V. *et al.* Human tankyrases are aberrantly expressed in colon tumors and contain multiple epitopes that induce humoral and cellular immune responses in cancer patients. *Cancer Immunol. Immunother.* **57**, 871–881 (2008).

104. Li, X. *et al.* Proteomic Analysis of the Human Tankyrase Protein Interaction Network Reveals Its Role in Pexophagy. *Cell Rep.* **20**, 737–749 (2017).

105. Zhang, Y. *et al.* RNF146 is a poly(ADP-ribose)-directed E3 ligase that regulates axin degradation and Wnt signalling. *Nat. Cell Biol.* **13**, 623–629 (2011).

106. DaRosa, P. A. *et al.* Allosteric activation of the RNF146 ubiquitin ligase by a poly(ADP-ribosyl)ation signal. *Nature* **517**, 223–226 (2014).

107. Cook, B. D., Dynek, J. N., Chang, W., Shostak, G. & Smith, S. Role for the Related Poly(ADP-Ribose) Polymerases Tankyrase 1 and 2 at Human Telomeres. *Mol. Cell. Biol.* **22**, 332–342 (2002).

108. Smith, S. & De Lange, T. Tankyrase promotes telomere elongation in human cells. *Curr. Biol.* **10**, 1299–1302 (2000).

109. Huang, S. M. A. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009).

110. Chi, N. W. & Lodish, H. F. Tankyrase is a Golgi-associated mitogen-activated protein kinase substrate that interacts with IRAP in GLUT4 vesicles. *J. Biol. Chem.* **275**, 38437–38444 (2000).

111. Troilo, A. *et al.* Angiomotin stabilization by tankyrase inhibitors antagonizes constitutive TEAD-dependent transcription and proliferation of human tumor cells with Hippo pathway core component mutations. *Oncotarget* **7**, 28765–28782 (2016).

112. Wang, W. *et al.* Tankyrase Inhibitors Target YAP by Stabilizing Angiomotin Family Proteins. *Cell Rep.* **13**, 524–532 (2015).

113. Li, N. *et al.* Tankyrase disrupts metabolic homeostasis and promotes tumorigenesis by inhibiting LKB1-AMPK signalling. *Nat. Commun.* **10**, (2019).

114. Hsiao, S. J. & Smith, S. Tankyrase function at telomeres, spindle poles, and beyond. *Biochimie* **90**, 83–92 (2008).

115. Chang, W., Dynek, J. N. & Smith, S. NuMA is a major acceptor of poly(ADP-ribosyl)ation by tankyrase 1 in mitosis. *Biochem. J.* **391**, 177–184 (2005).

116. Dynek, J. N. & Smith, S. Resolution of Sister Telomere Association Is Required for Progression Through Mitosis. *Science (80-.).* **304**, 97–100 (2004).

117. Dregalla, R. C. *et al.* Regulatory roles of tankyrase 1 at telomeres and in DNA repair: Suppression of T-SCE and stabilization of DNA-pkcs. *Aging (Albany. NY).* **2**, 691–708 (2010).

118. Levaot, N. *et al.* Loss of Tankyrase-mediated destruction of 3BP2 is the underlying pathogenic mechanism of cherubism. *Cell* **147**, 1324–1339 (2011).

119. Kim, M. K., Dudognon, C. & Smith, S. Tankyrase 1 regulates centrosome function by controlling CPAP stability. *EMBO Rep.* **13**, 724–732 (2012).

120. Li, N. *et al.* Poly-ADP ribosylation of PTEN by tankyrases promotes PTEN degradation and tumor growth. *Genes Dev.* **29**, 157–170 (2015).

121. MacDonald, B. T., Tamai, K. & He, X. Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell* vol. 17 9–26 (2009).

122. Aberle, H., Bauer, A., Stappert, J., Kispert, A. & Kemler, R. β-catenin is a target for the ubiquitin-proteasome pathway. *EMBO J.* **16**, 3797–3804 (1997).

123. Yang, E. *et al.* Wnt pathway activation by ADP-ribosylation. *Nat. Commun.* **7**, 1–14 (2016).

124. Wang, Z., Tacchelly-Benites, O., Yang, E. & Ahmed, Y. Dual Roles for Membrane Association of Drosophila Axin in Wnt Signaling. *PLoS Genet.* **12**, (2016).

125. Behrens, J. *et al.* Functional interaction of β-catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).

126. Huber, O. *et al.* Nuclear localization of β-catenin by interaction with transcription factor LEF-1. *Mech. Dev.* **59**, 3–10 (1996).

127. Molenaar, M. *et al.* XTcf-3 transcription factor mediates β-catenin-induced axis formation in xenopus embryos. *Cell* **86**, 391–399 (1996).

128. Powell, S. M. *et al.* APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235–237 (1992).

129. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

130. Mariotti, L., Pollock, K. & Guettler, S. Regulation of Wnt/β-catenin signalling by

tankyrase-dependent poly(ADP-ribosyl)ation and scaffolding. *Br. J. Pharmacol.* **174**, 4611–4636 (2017).

131. Bae, J., Donigian, J. R. & Hsueh, A. J. W. Tankyrase 1 interacts with Mcl-1 proteins and inhibits their regulation of apoptosis. *J. Biol. Chem.* **278**, 5195–5204 (2003).

132. Riffell, J. L., Lord, C. J. & Ashworth, A. Tankyrase-targeted therapeutics: Expanding opportunities in the PARP family. *Nat. Rev. Drug Discov.* **11**, 923–936 (2012).

133. Seimiya, H. The telomeric PARP, tankyrases, as targets for cancer therapy. *British Journal of Cancer* vol. 94 341–345 (2006).

134. Chen, B. *et al.* Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat. Chem. Biol.* **5**, 100–107 (2009).

135. Waaler, J. *et al.* A novel tankyrase inhibitor decreases canonical Wnt signaling in colon carcinoma cells and reduces tumor growth in conditional APC mutant mice. *Cancer Res.* **72**, 2822–2832 (2012).

136. Waaler, J. *et al.* Novel synthetic antagonists of canonical Wnt signaling inhibit colorectal cancer cell growth. *Cancer Res.* **71**, 197–205 (2011).

137. James, R. G. *et al.* WIKI4, a Novel Inhibitor of Tankyrase and Wnt/ß-Catenin Signaling. *PLoS One* **7**, e50457 (2012).

138. Lau, T. *et al.* A novel tankyrase small-molecule inhibitor suppresses APC mutation-driven colorectal tumor growth. *Cancer Res.* **73**, 3132–3144 (2013).

139. De La Roche, M., Ibrahim, A. E. K., Mieszczanek, J. & Bienz, M. LEF1 and B9L shield b-catenin from inactivation by axin, desensitizing colorectal cancer cells to tankyrase inhibitors. *Cancer Res.* **74**, 1495–1505 (2014).

140. Ulsamer, A. *et al.* Axin pathway activity regulates in vivo pY654-β-catenin accumulation and pulmonary fibrosis. *J. Biol. Chem.* **287**, 5164–5172 (2012).

141. Zhong, Y. *et al.* Tankyrase Inhibition Causes Reversible Intestinal Toxicity in Mice with a Therapeutic Index < 1. *Toxicol. Pathol.* **44**, 267–278 (2016).

142. https://clinicaltrials.gov.

143. Rouleau, M., Patel, A., Hendzel, M. J., Kaufmann, S. H. & Poirier, G. G. PARP inhibition: PARP1 and beyond. *Nat. Rev. Cancer* **10**, 293–301 (2010).

144. Morrone, S., Cheng, Z., Moon, R. T., Cong, F. & Xu, W. Crystal structure of a tankyrase-axin complex and its implications for axin turnover and tankyrase substrate recruitment. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1500–1505 (2012).

145. Corporation, P. *Nano-Glo ® HiBiT Lytic Detection System Instructions for Use of Products N3030, N3040 and N3050.* www.promega.com.

146. Perez-Riba, A., Komives, E., Main, E. R. G. & Itzhaki, L. S. Decoupling a tandem-repeat protein: Impact of multiple loop insertions on a modular scaffold. *Sci. Rep.* **9**, (2019).

147. NEB. URL https://tmcalculator.neb.com/#!/main.

148. Kempe, T. *et al.* Multiple-copy genes: production and modification of monomeric peptides from large multimeric fusion proteins. *Gene* **39**, 239–245 (1985).

149. Kajander, T., Cortajarena, A. L., Main, E. R. G., Mochrie, S. G. J. & Regan, L. A new folding paradigm for repeat proteins. *J. Am. Chem. Soc.* **127**, 10188–10190 (2005).

150. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. & Galas, D. J. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* **17**, 6545–6551 (1989).

151. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* 571–607 (Humana Press, 2005). doi:10.1385/1-59259-890-0:571.

152. Circular Dichroism Units and Conversions. URL https://www.photophysics.com/circular-dichroism/chirascan-technology/circular-dichroism-spectroscopy-units-conversions/.

153. Perez-Riba, A. & Itzhaki, L. S. A method for rapid high-throughput biophysical analysis of proteins. *Sci. Rep.* **7**, 9071 (2017).

154. Jackson, S. E. & Fersht, A. R. Folding of Chymotrypsin Inhibitor 2.1. Evidence for a Two-State Transition. *Biochemistry* **30**, 10428–10435 (1991).

155. Scheuermann, T. H. & Brautigam, C. A. High-precision, automated integration of multiple isothermal titration calorimetric thermograms: New features of NITPIC. *Methods* **76**, 87–98 (2015).

156. Kyhse-Andersen, J. Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polycrylamide to nitrocellulose. *J. Biochem. Biophys. Methods* **10**, 203–209 (1984).

157. Korinek, V. *et al.* Constitutive Transcriptional Activation by a b-Catenin–Tcf Complex in APC-/- Colon Carcinoma. *Science (80-.).* **275**, 1784–1787 (1997).

158. Kube, S. *et al.* Fusogenic liposomes as nanocarriers for the delivery of intracellular proteins. *Langmuir* **33**, 1051–1059 (2017).

159. Guharoy, M. & Chakrabarti, P. Secondary structure based analysis and classification of biological interfaces: Identification of binding motifs in protein-protein interactions. *Bioinformatics* **23**, 1909–1918 (2007).

160. Oates, M. E. *et al.* D2P2: Database of disordered protein predictions. *Nucleic Acids Res.* **41**, (2013).

161. Van Der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chemical Reviews* vol. 114 6589–6631 (2014).

162. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).

163. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13392–13397 (2013).

164. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. & Pappu, R. V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* **112**, 16–21 (2017).

165. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Genet.* **41**, 415–427 (2000).

166. Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A Million peptide motifs for the molecular biologist. *Molecular Cell* vol. 55 161–169 (2014).

167. Van Roey, K. *et al.* Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chemical Reviews* vol. 114 6733–6778 (2014).

168. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).

169. Pelay-Gimeno, M., Glas, A., Koch, O. & Grossmann, T. N. Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angew. Chemie - Int. Ed.* **54**, 8896–8927 (2015).

170. Der, B. S. & Kuhlman, B. Strategies to control the binding mode of de novo designed protein interactions. *Current Opinion in Structural Biology* vol. 23 639–646 (2013).

171. Eisenberg, D. The discovery of the α-helix and β-sheet, the principal structural features of proteins. *Proceedings of the National Academy of Sciences of the United States of America* vol. 100 11207–11210 (2003).

172. PAULING, L., COREY, R. B. & BRANSON, H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **37**, 205–211 (1951).

173. Tan, Y. S., Lane, D. P. & Verma, C. S. Stapled peptide design: principles and roles of computation. *Drug Discovery Today* vol. 21 1642–1653 (2016).

174. Fosgerau, K. & Hoffmann, T. Peptide therapeutics: Current status and future directions. *Drug Discovery Today* vol. 20 122–128 (2015).

175. Kritzer, J. A. *et al.* Miniature protein inhibitors of the p53-hDM2 interaction. *ChemBioChem* **7**, 29–31 (2006).

176. Zondlo, N. J. & Schepartz, A. Highly specific DNA recognition by a designed miniature protein. *Journal of the American Chemical Society* vol. 121 6938–6939 (1999).

177. Chin, J. W. & Schepartz, A. Concerted evolution of structure and function in a miniature protein. *Journal of the American Chemical Society* vol. 123 2929–2930 (2001).

178. Tao, Y., Strelkov, S. V, Mesyanzhinov, V. V & Rossmann, M. G. Structure of bacteriophage T4 fibritin: A segmented coiled coil and the role of the C-terminal domain. *Structure* **5**, 789–798 (1997).

179. Letarov, A. V, Londer, Y. Y., Boudko, S. P. & Mesyanzhinov, V. V. The Carboxy-Terminal Domain Initiates Trimerization of Bacteriophage T4 Fibritin. *Biochem.* **64**, 817–823 (1999).

180. Boudko, S. P., Strelkov, S. V., Engel, J. & Stetefeld, J. Design and crystal structure of bacteriophage T4 mini-fibritin NCCF. *J. Mol. Biol.* **339**, 927–935 (2004).

181. Güthe, S. *et al.* Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *J. Mol. Biol.* **337**, 905–915 (2004).

182. Gómez-González, J. *et al.* Directed self-assembly of trimeric DNA-bindingchiral miniprotein helicates. *Front. Chem.* **6**, (2018).

183. Kobayashi, N. *et al.* Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric de Novo Protein. *J. Am. Chem. Soc.* **137**, 11285–11293 (2015).

184. Glover, D. J., Giger, L., Kim, S. S., Naik, R. R. & Clark, D. S. Geometrical assembly of ultrastable protein templates for nanomaterials. *Nat. Commun.* **7**, 11771 (2016).

185. McLellan, J. S. *et al.* Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science (80-.).* **342**, 592–598 (2013).

186. Stetefeld, J. *et al.* Collagen stabilization at atomic level: Crystal structure of designed (GlyProPro)10foldon. *Structure* **11**, 339–346 (2003).

187. Du, L. *et al.* A recombinant vaccine of H5N1 HA1 fused with foldon and human IgG Fc induced complete cross-clade protection against divergent H5N1 viruses. *PLoS One* **6**, (2011).

188. Javadi, Y. & Main, E. R. G. Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17383–17388 (2009).

189. Millership, C., Phillips, J. J. & Main, E. R. G. Ising Model Reprogramming of a Repeat Protein's Equilibrium Unfolding Pathway. *J. Mol. Biol.* **428**, 1804–1817 (2016).

190. Cortajarena, A. L., Mochrie, S. G. J. & Regan, L. Modulating repeat protein stability: The effect of individual helix stability on the collective behavior of the ensemble. *Protein Sci.* **20**, 1042–1047 (2011).

191. Phillips, J. J., Javadi, Y., Millership, C. & Main, E. R. G. Modulation of the multistate folding of designed TPR proteins through intrinsic and extrinsic factors. *Protein Sci.* **21**, 327–338 (2012).

192. Pappu, R. V. http://pappulab.wustl.edu/CIDER/analysis/.

193. Simeon, R. & Chen, Z. In vitro-engineered non-antibody protein therapeutics. *Protein Cell* **9**, 3–14 (2018).

194. Ladurner, A. G. & Fersht, A. R. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* **273**, 330–337 (1997).

195. Nagi, A. D. & Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold. Des.* **2**, 67–75 (1997).

196. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527–2535 (2007).

197. Cortajarena, A. L., Mochrie, S. G. J. & Regan, L. Mapping the Energy Landscape of Repeat Proteins using NMR-detected Hydrogen Exchange. *J. Mol. Biol.* **379**, 617–626 (2008).

198. Freire, E., Schön, A., Hutchins, B. M. & Brown, R. K. Chemical denaturation as a tool in the formulation optimization of biologics. *Drug Discovery Today* vol. 18 1007–1013 (2013).

199. Garbett, N. C. & Chaires, J. B. Thermodynamic studies for drug design and screening. *Expert Opinion on Drug Discovery* vol. 7 299–314 (2012).

200. Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7**, 1848–1857 (2012).

201. England, C. G., Ehlerding, E. B. & Cai, W. NanoLuc: A Small Luciferase Is

Brightening Up the Field of Bioluminescence. *Bioconjugate Chemistry* vol. 27 1175–1187 (2016).

202. Dixon, A. S. *et al.* NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **11**, 400–408 (2016).

203. Schwinn, M. K. *et al.* CRISPR-Mediated Tagging of Endogenous Proteins with a Luminescent Peptide. *ACS Chem. Biol.* **13**, 467–474 (2018).

204. Ranawakage, D. C., Takada, T. & Kamachi, Y. HiBiT-qIP, HiBiT-based quantitative immunoprecipitation, facilitates the determination of antibody affinity under immunoprecipitation conditions. *Sci. Rep.* **9**, (2019).

205. Banjade, S. *et al.* Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6426–E6435 (2015).

206. Su, X. *et al.* Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science (80-.).* **352**, 595–599 (2016).

207. Schaefer, K. N. & Peifer, M. Wnt/Beta-Catenin Signaling Regulation and a Role for Biomolecular Condensates. *Developmental Cell* vol. 48 429–444 (2019).

208. Fung, H. Y. J., Birol, M. & Rhoades, E. IDPs in macromolecular complexes: the roles of multivalent interactions in diverse assemblies. *Current Opinion in Structural Biology* vol. 49 36–43 (2018).

209. Weng, J. & Wang, W. Dynamic multivalent interactions of intrinsically disordered proteins. *Current Opinion in Structural Biology* vol. 62 9–13 (2020).

210. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).

211. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology* vol. 18 285–298 (2017).

212. Carnell, M., Macmillan, A. & Whan, R. Fluorescence recovery after photobleaching (FRAP): Acquisition, analysis, and applications. *Methods Mol. Biol.* **1232**, 255–271 (2015).

213. He, S. *et al.* Wnt3a: functions and implications in cancer. *Chinese journal of cancer* vol. 34 554–562 (2015).

214. Najdi, R. *et al.* A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities. *Differentiation* **84**, 203–213 (2012).

215. Alok, A. *et al.* Wnt proteins synergize to activate β-catenin signaling. *J. Cell Sci.* **130**,

1532–1544 (2017).

216. Ray, M., Lee, Y. W., Scaletti, F., Yu, R. & Rotello, V. M. Intracellular delivery of proteins by nanocarriers. *Nanomedicine* **12**, 941–952 (2017).

217. Beltrán-Gracia, E., López-Camacho, A., Higuera-Ciapara, I., Velázquez-Fernández, J. B. & Vallejo-Cardona, A. A. Nanomedicine review: Clinical developments in liposomal applications. *Cancer Nanotechnology* vol. 10 1–40 (2019).

218. Gu, Z., Biswas, A., Zhao, M. & Tang, Y. Tailoring nanocarriers for intracellular protein delivery. *Chem. Soc. Rev.* **40**, 3638–3655 (2011).

219. Kahn, M. Can we safely target the WNT pathway? *Nature Reviews Drug Discovery* vol. 13 513–532 (2014).

220. Kim, Y.-K. RNA Therapy: Current Status and Future Potential. *Chonnam Med. J.* **56**, 87 (2020).

221. Davey, N. E., Travé, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends in Biochemical Sciences* vol. 36 159–169 (2011).

222. Röth, S., Fulcher, L. J. & Sapkota, G. P. Advances in targeted degradation of endogenous proteins. *Cellular and Molecular Life Sciences* vol. 76 2761–2777 (2019).

223. An, S. & Fu, L. Small-molecule PROTACs: An emerging and promising approach for the development of targeted therapy drugs. *EBioMedicine* vol. 36 553–562 (2018).

224. Che, Y., Gilbert, A. M., Shanmugasundaram, V. & Noe, M. C. Inducing protein-protein interactions with molecular glues. *Bioorganic and Medicinal Chemistry Letters* vol. 28 2585–2592 (2018).

225. Amm, I., Sommer, T. & Wolf, D. H. Protein quality control and elimination of protein waste: The role of the ubiquitin-proteasome system. *Biochimica et Biophysica Acta - Molecular Cell Research* vol. 1843 182–196 (2014).

226. Tanaka, K. The proteasome: Overview of structure and functions. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* vol. 85 12–36 (2009).

227. Lee, D., Kim, M. & Cho, K. H. A design principle underlying the paradoxical roles of E3 ubiquitin ligases. *Sci. Rep.* **4**, 1–13 (2014).

228. Lucas, X. & Ciulli, A. Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies. *Current Opinion in Structural Biology* vol. 44 101–110 (2017).

229. Donovan, K. A. *et al.* Mapping the Degradable Kinome Provides a Resource for Expedited Degrader Development. *Cell* **183**, 1714-1731.e10 (2020).

230. Pollock, K. *et al.* Fragment-based screening identifies molecules targeting the

substrate-binding ankyrin repeat domains of tankyrase. *Sci. Rep.* **9**, (2019).

231. Bonet, J. *et al.* Rosetta FunFolDes – A general framework for the computational design of functional proteins. *PLoS Comput. Biol.* **14**, e1006623 (2018).

232. Sesterhenn, F. *et al.* De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science (80-.).* **368**, (2020).

233. Thorvaldsen, T. E. *et al.* Structure, dynamics, and functionality of tankyrase inhibitor-induced degradasomes. *Mol. Cancer Res.* **13**, 1487–1501 (2015).
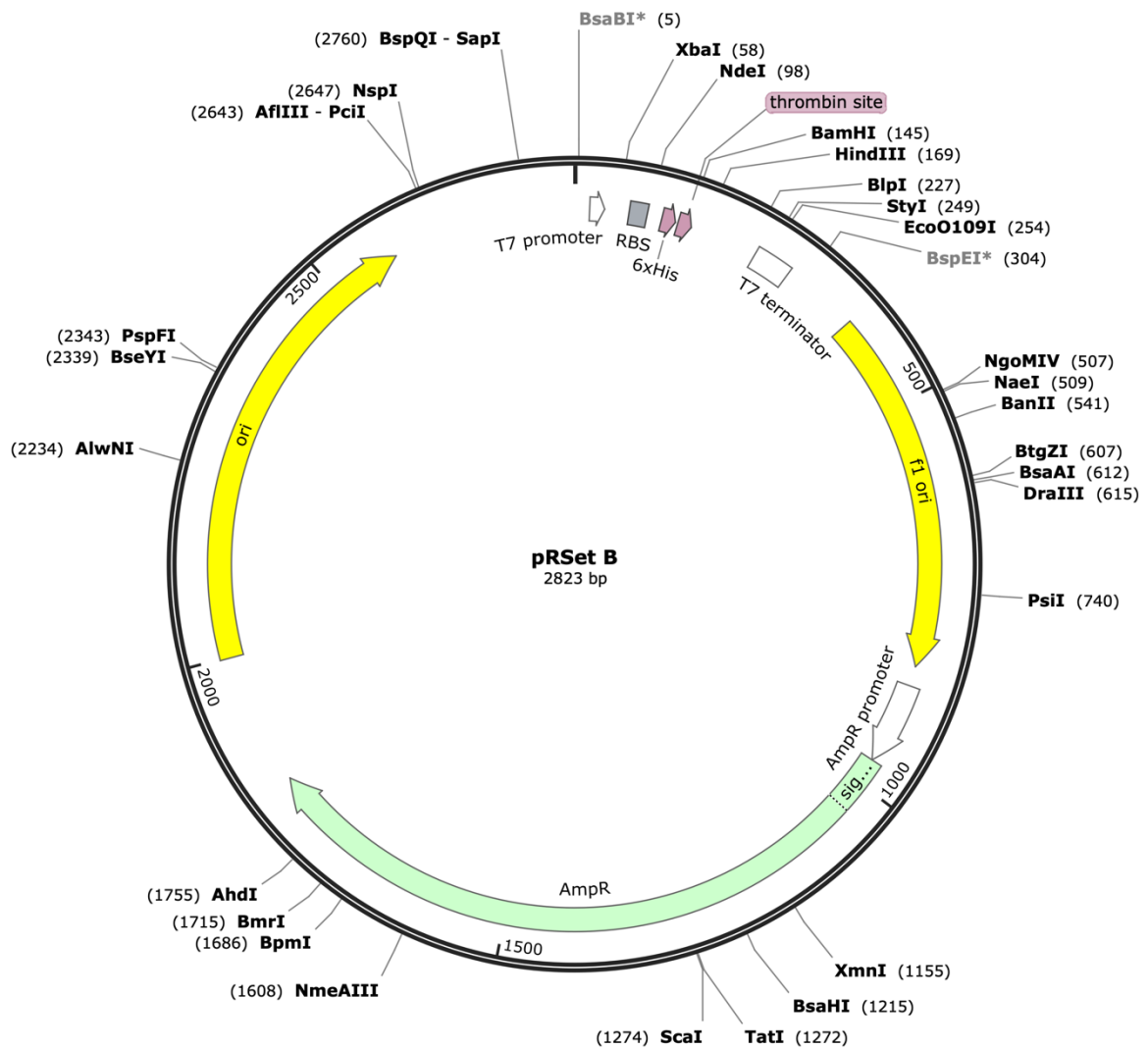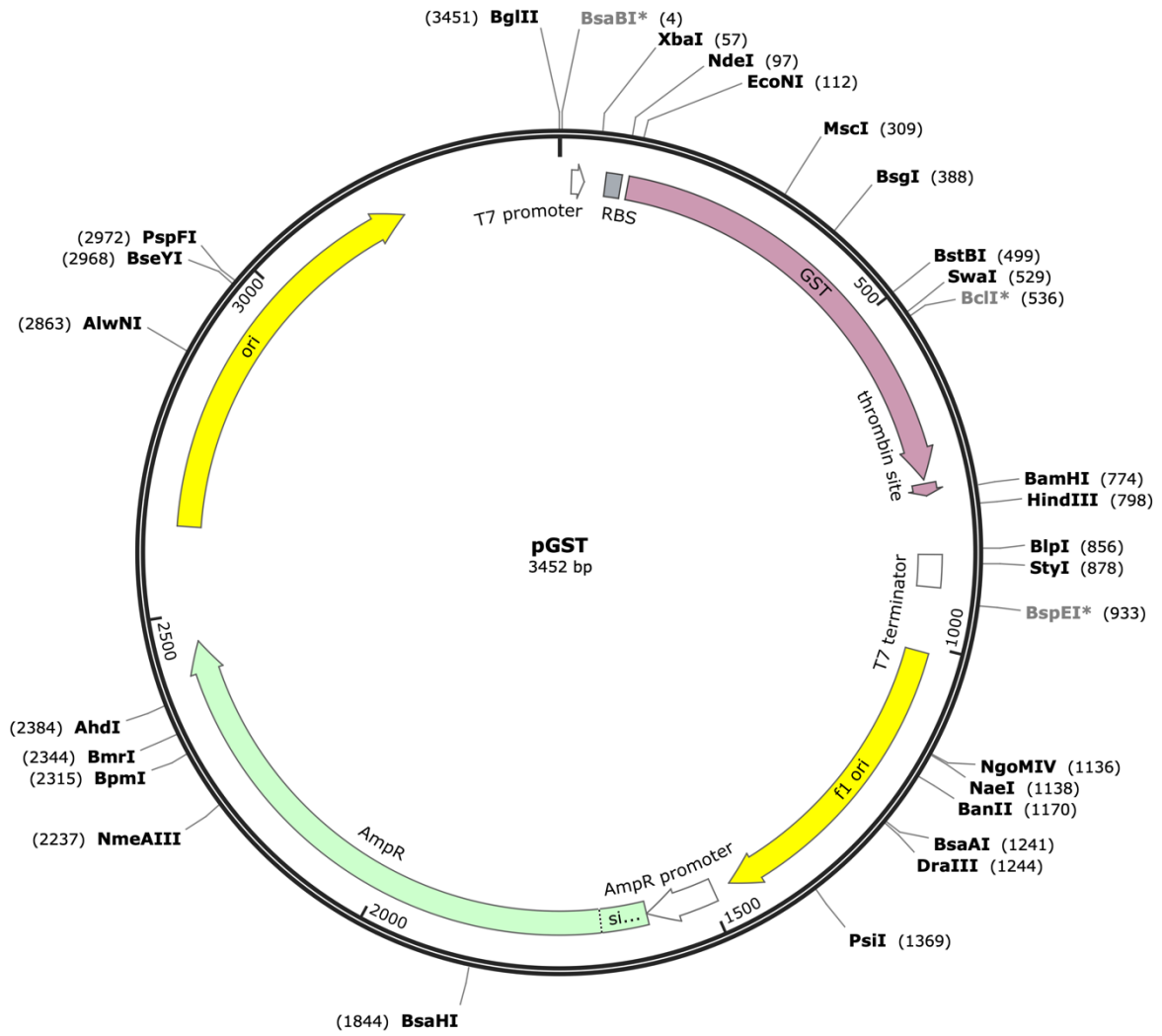
Appendix A

Vector maps
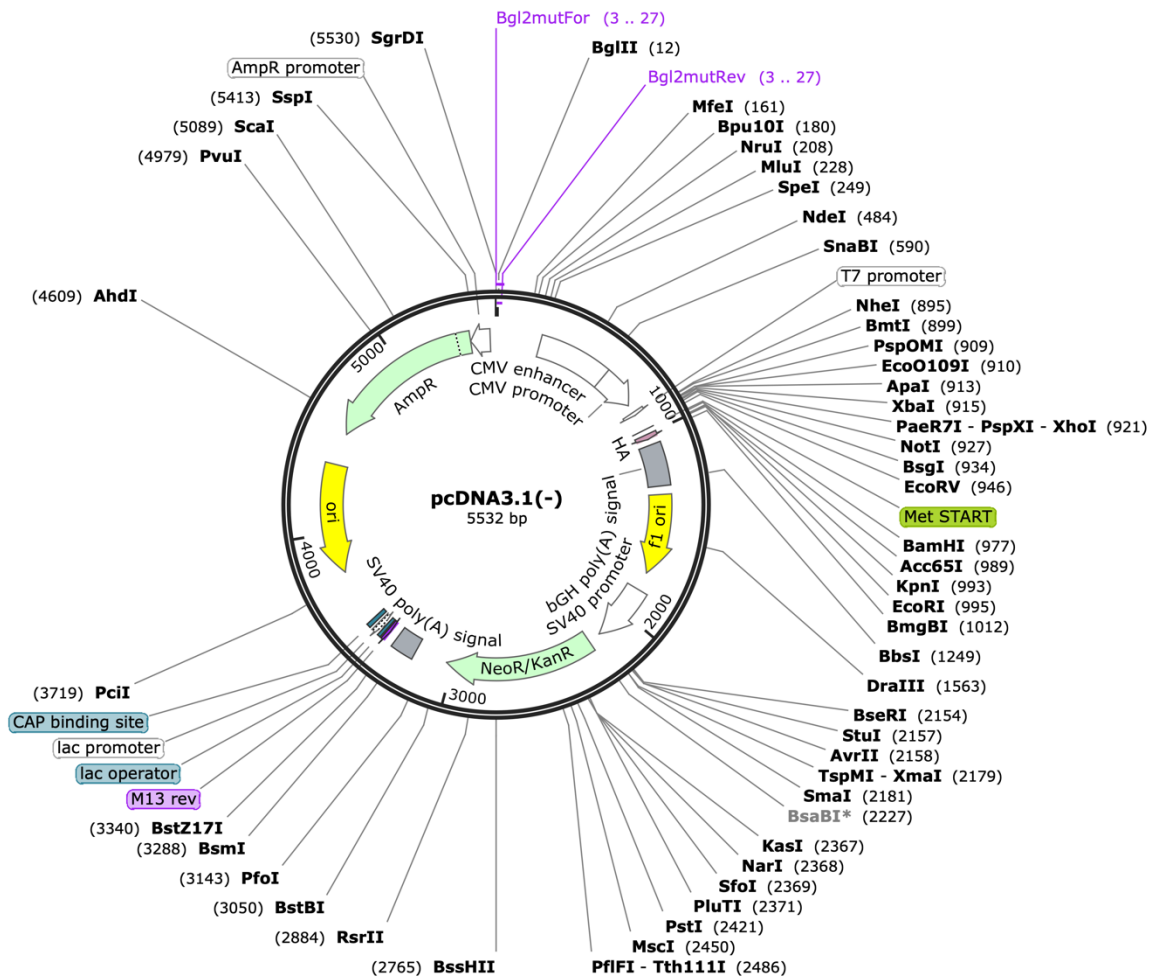
Vector maps were created with SnapGene Viewer

pRSet B

pRSet B
2823 bp

BsaBI* (5)
XbaI (58)
NdeI (98)
thrombin site
BamHI (145)
HindIII (169)
BlpI (227)
StyI (249)
EcoO109I (254)
BspEI* (304)
NgoMIV (507)
NaeI (509)
BanII (541)
BtgZI (607)
BsaAI (612)
DraIII (615)
PsiI (740)
XmnI (1155)
BsaHI (1215)
TatI (1272)
(1274) ScaI
NmeAIII (1608)
(1686) BpmI
(1715) BmrI
(1755) AhdI
(2234) AlwNI
(2339) BseYI
(2343) PspFI
(2643) AflIII - PciI
(2647) NspI
(2760) BspQI - SapI

T7 promoter RBS 6xHis T7 terminator f1 ori AmpR promoter sig... AmpR ori

pGST

pGST
3452 bp

(3451) **BglII**
BsaBI* (4)
XbaI (57)
NdeI (97)
EcoNI (112)
MscI (309)
BsgI (388)
BstBI (499)
SwaI (529)
BclI* (536)
BamHI (774)
HindIII (798)
BlpI (856)
StyI (878)
BspEI* (933)
NgoMIV (1136)
NaeI (1138)
BanII (1170)
BsaAI (1241)
DraIII (1244)
PsiI (1369)
(1844) **BsaHI**
(2237) **NmeAIII**
(2315) **BpmI**
(2344) **BmrI**
(2384) **AhdI**
(2863) **AlwNI**
(2972) **PspFI**
(2968) **BseYI**

T7 promoter
RBS
GST
thrombin site
T7 terminator
f1 ori
AmpR promoter
si...
AmpR
ori

179

pcDNA3.1(-)

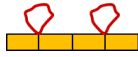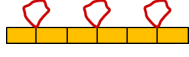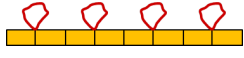The forward and revers primers used to delete the BglII restriction site are highlighted in purple.

pcDNA3.1(-)
5532 bp

HiBiT-pcDNA3.1(-)

(3 .. 27) Bgl2mutFor
BglII (12)
(5575) **SgrDI**
Bgl2mutRev (3 .. 27)
AmpR promoter
MfeI (161)
(5458) **SspI**
NruI (208)
MluI (228)
(5134) **ScaI**
SpeI (249)
(5024) **PvuI**
NdeI (484)
SnaBI (590)
T7 promoter
(4654) **AhdI**
NheI (895)
BmtI (899)
PspOMI (909)
EcoO109I (910)
ApaI (913)
XbaI (915)
PaeR7I - PspXI - XhoI (921)
NotI (927)
BsgI (934)
EcoRV (946)
Met START
HiBiT Rv (960 .. 1000)
HiBiT GSSG Fw (1001 .. 1039)
BamHI (1022)
Acc65I (1034)
KpnI (1038)
EcoRI (1040)
BmgBI (1057)
HA
BbsI (1294)
DraIII (1608)
BseRI (2199)
StuI (2202)
AvrII (2203)
TspMI - XmaI (2224)
SmaI (2226)
BsaBI* (2272)
KasI (2412)
NarI (2413)
SfoI (2414)
PluTI (2416)
PstI (2466)
MscI (2495)
(2531) **PflFI - Tth111I**
(2810) **BssHII**
(2929) **RsrII**
(3095) **BstBI**
(3188) **PfoI**
(3333) **BsmI**
(3385) **BstZ17I**
M13 rev
lac operator
lac promoter
CAP binding site
(3764) **PciI**

NeoR/KanR

SV40 promoter
f1 ori
bGH poly(A) signal

SV40 poly(A) signal

ori

AmpR

CMV enhancer
CMV promoter

HiBiT

5000
1000
2000
3000
4000

HiBiT pcDNA3.1(-)
5577 bp

181

Appendix B

Protein sequences

Amino acid sequences of the CTPR constructs generated in this study are listed. The TBP is in red, the RL is in green, the solvating helix is in orange and the foldon domain is in purple. Schematic representations of each construct are also shown, with the CTPR repeat shown as a yellow rectangle.

	Sequence	
1TBP-CTPR2	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRS	
2TBP-CTPR4	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPRS	
3TBP-CTPR6	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSA EAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS	
4TBP-CTPR8	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSA EAWYNLGNAYYKQGDYQKAIEYYQKALELDPRSAE AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNRE AGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYY QKALELDPRS	
1TBP-CTPR4-SolvH	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN REAGDGEEDPNNAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAWYNLGNAYYKQGDYQKAIE YYQKALELDPNNAEAKQNLGNAKQKQG	

1TBP-CTPR2-foldon AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPRSAKASLNLANADIKTIQEAGYIPEAPR
DGQAYVRKDGEWVLLSTFLRS

2TBP-CTPR4-foldon AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPRSAKASLNLANADIKT
IQEAGYIPEAPRDGQAYVRKDGEWVLLSTFLRS

3TBP-CTPR6-foldon AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSA
EAWYNLGNAYYKQGDYQKAIEYYQKALELDPRSA
KASLNLANADIKTIQEAGYIPEAPRDGQAYVRKDGE
WVLLSTFLRS

4TBP-CTPR8-foldon AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSA
EAWYNLGNAYYKQGDYQKAIEYYQKALELDPRSAE
AWYNLGNAYYKQGDYQKAIEYYQKALELDPNNRE
AGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEYY
QKALELDPRSAKASLNLANADIKTIQEAGYIPEAPRD
GQAYVRKDGEWVLLSTFLRS

CTPR2 AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS

CTPR3 AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS

CTPR4 AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS

CTPR6 AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS

3RL-CTPR6

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
GSDDPRGSRSAEAWYNLGNAYYKQGDYQKAIEYYQ
KALELDPRSAEAWYNLGNAYYKQGDYQKAIEYYQ
KALELDPNNGSDDPRGSRSAEAWYNLGNAYYKQGD
YQKAIEYYQKALELDPRSAEAWYNLGNAYYKQGD
YQKAIEYYQKALELDPNNGSDDPRGSRSAEAWYNL
GNAYYKQGDYQKAIEYYQKALELDPRS

CTPR6-foldon

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
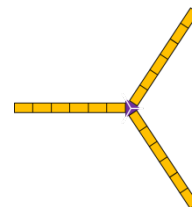AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AKASLNLANADIKTIQEAGYIPEAPRDGQAYVRKDG
EWVLLSTFLRS

Amino acid sequences of the hetero-bifunctional CTPR constructs in this study are listed. Sequence corresponding to the TBP is in red, the degron in shades of blue. Schematic representations of each construct are also shown, with the CTPR repeat shown as a yellow rectangle and the RTPR repeat as a brown rectangle.

Name	Sequence	
2TBP-CTPR4-ABBA	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR QGDYQRAIEYYQRALELDPGGSLSSAFHVFEDGNKE NGGPNAEAWYNLGNAYYRQGDYQRAIEYYQRALE LDPRS	
2TBP-CTPR4-DBOX	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR QGDYQRAIEYYQRALELDPGGPRLPLGDVSNNGGPN AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPRS	
2TBP-CTPR4-KEN	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR QGDYQRAIEYYQRALELDPGGSEDKENVPPGGPNAE AWYNLGNAYYRQGDYQRAIEYYQRALELDPRS	
2TBP-CTPR4-Nrf2	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR QGDYQRAIEYYQRALELDPGGDPETGELGGPNAEA WYNLGNAYYRQGDYQRAIEYYQRALELDPRS	
2TBP-CTPR4-p27	AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR QGDYQRAIEYYQRALELDPNNAGSNEQEPKKRSPDA EAWYNLGNAYYRQGDYQRAIEYYQRALELDPRS	

2TBP-CTPR4-p53

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR
QGDYQRAIEYYQRALELDPNNFAAYWNLLSAYG

2TBP-CTPR4-PHYL

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR
QGDYQRAIEYYQRALELDPGGLRPVAMVRPTVGGP
NAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPRS

2TBP-CTPR4-Puc

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR
QGDYQRAIEYYQRALELDPGGLACDEVTSTTSSSTA
GGPNAEAWYNLGNAYYRQGDYQRAIEYYQRALEL
DPRS

2TBP-CTPR4-RTPR2

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN
REAGDGEEDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPRSAEAWYNLGNAYYKQGDYQKAIEY
YQKALELDPNNREAGDGEEDPRSAEAWYNLGNAYY
KQGDYQKAIEYYQKALELDPRSAEAWYNLGNAYYR
QGDYQRAIEYYQRALELDPNNAEAWYNLGNAYYR
QGDYQRAIEYYQRALELDPRS

CTPR2-ABBA

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGS
LSSAFHVFEDGNKENGGPNAEAWYNLGNAYYRQG
DYQRAIEYYQRALELDPRS

CTPR2-DBOX

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGP
RLPLGDVSNNGGPNAEAWYNLGNAYYRQGDYQRAI
EYYQRALELDPRS

CTPR2-KEN

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGS
EDKENVPPGGPNAEAWYNLGNAYYRQGDYQRAIEY
YQRALELDPRS

CTPR2-Nrf2

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGG
DPETGELGGPNAEAWYNLGNAYYRQGDYQRAIEYY
QRALELDPRS

CTPR2-p27

MGSAEAWYNLGNAYYKQGDYQKAIEYYQKALELD
PRSAEAWYNLGNAYYKQGDYQKAIEYYQKALELDP
RSAEAWYNLGNAYYRQGDYQRAIEYYQRALELDPN

NAGSNEQEPKKRSPDAEAWYNLGNAYYRQGDYQR
AIEYYQRALELDPRS

CTPR2-p53

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPNNF
AAYWNLLSAYG

CTPR2-PHYL

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGL
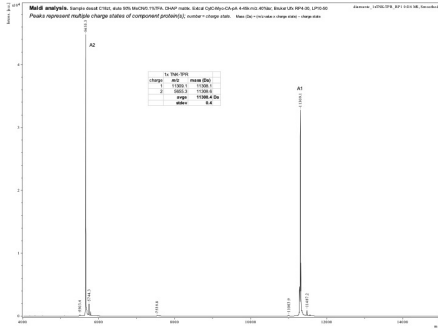RPVAMVRPTVGGPNAEAWYNLGNAYYRQGDYQRA
IEYYQRALELDPRS

CTPR2-Puc

AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPRS
AEAWYNLGNAYYRQGDYQRAIEYYQRALELDPGGL
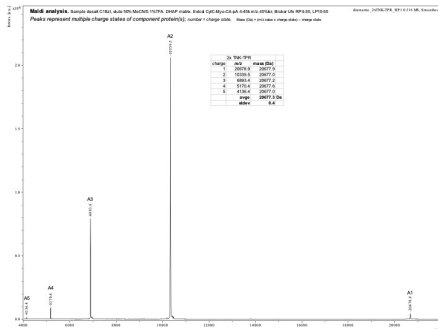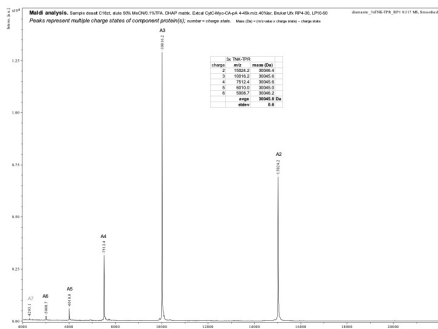ACDEVTSTTSSSTAGGPNAEAWYNLGNAYYRQGDY
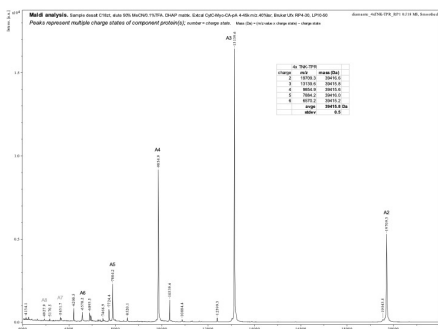QRAIEYYQRALELDPRS

Appendix C

Mass Spectrometry

1TBP-CTPR2
Theoretical MW: 11308 Da
Measured MW: 11308 Da

2TBP-CTPR4
Theoretical MW: 20677 Da
Measured MW: 20677 Da

3TBP-CTPR6
Theoretical MW: 30046 Da
Measured MW: 30046 Da

4TBP-CTPR8
Theoretical MW: 39415 Da
Measured MW: 39416 Da
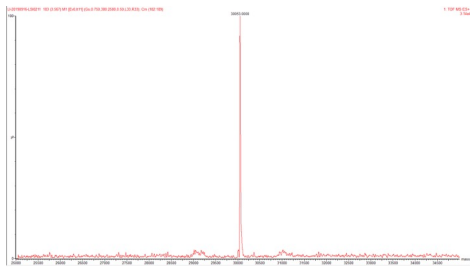
3TBP-CTPR6
Theoretical MW: 30046 Da
Measured MW: 30053 Da
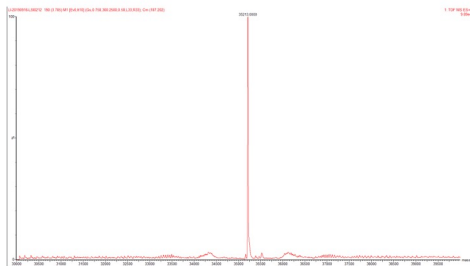
3TBP-CTPR6-foldon (single molecule)
Theoretical MW: 35206 Da
Measured MW: 35213 Da
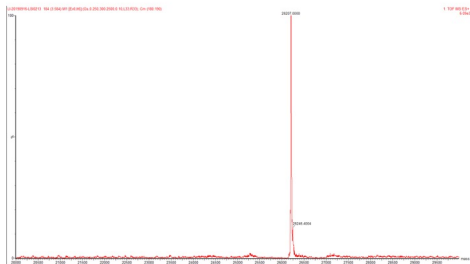
CTPR6
Theoretical MW: 26194 Da
Measured MW: 26207 Da
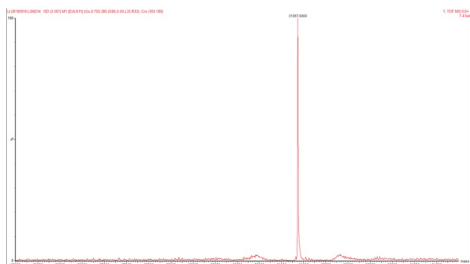
CTPR6-foldon (single molecule)
Theoretical MW: 31354 Da
Measured MW: 31367 Da

Appendix D

Publication

Engineering mono- and multi-valent inhibitors on a modular scaffold

Aurora Diamante, Piyush K. Chaturbedy, Pamela J. E. Rowling, Janet R. Kumita,

Rohan S. Eapen, Stephen H. McLaughlin, Marc de la Roche,

Albert Perez-Riba and Laura S. Itzhaki

Showcasing research from Professor Itzhaki's laboratory, Department of Pharmacology, University of Cambridge, UK.

Engineering mono- and multi-valent inhibitors on a modular scaffold

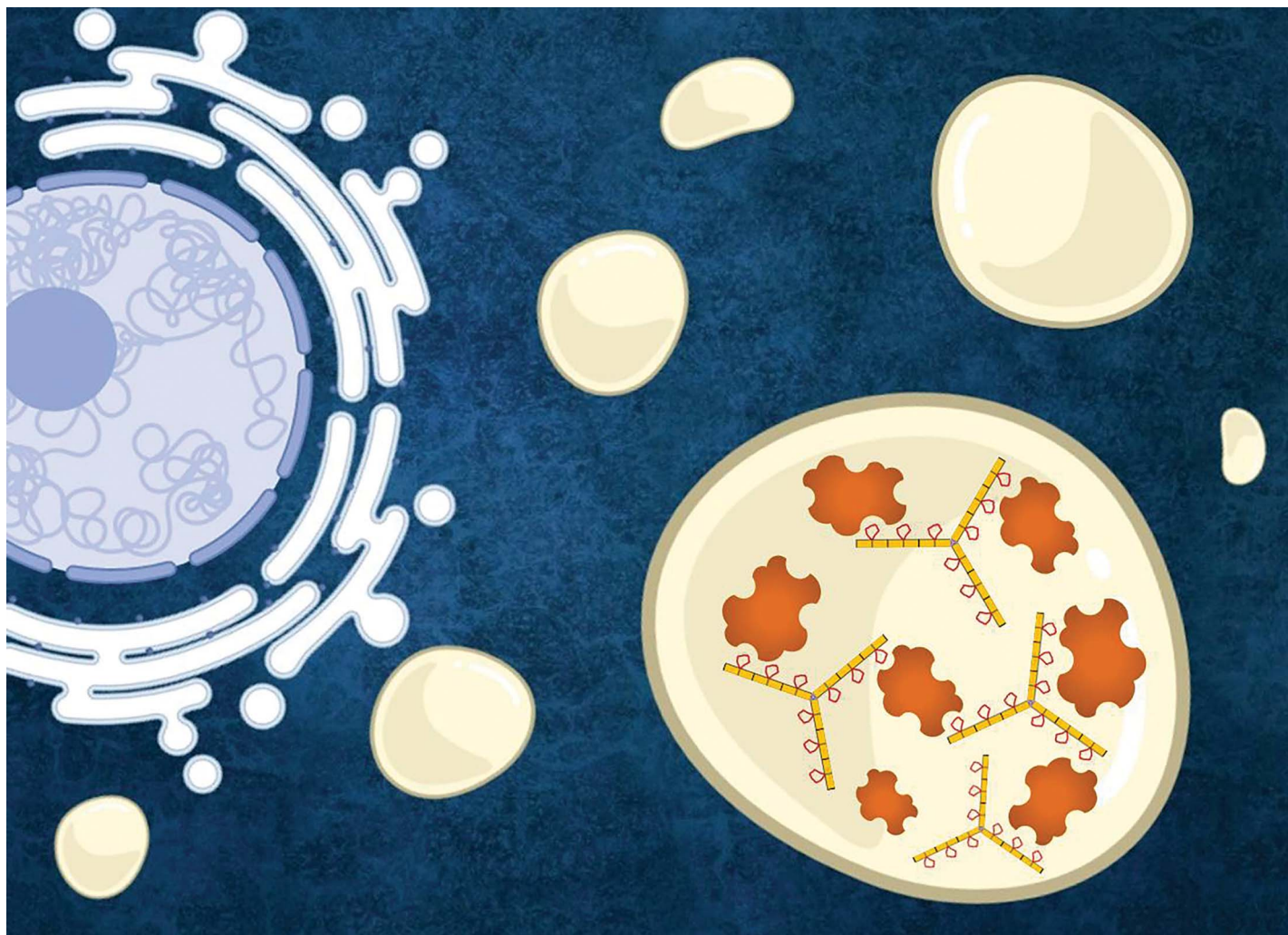We exploit the simple, ultra-stable and modular architecture of repeat proteins to create a platform capable of displaying both single as well as multiple functions in diverse configurations. We demonstrate proof of concept by building potent mulitvalent molecules to inhibit the multivalent tankyrase proteins, which play a key role in Wnt signaling and are upregulated in cancer. Our results highlight the potential of the repeat-protein platform with its built-in multivalent capabilities and precise, pre-programmed geometries. Multivalent inhibitors are key to targeting complex protein-protein interactions, where simple monovalent molecules are ineffective.

As featured in:

See Albert Perez-Riba,
Laura S. Itzhaki *et al.*, *Chem. Sci.*,
2021, **12**, 880.

ROYAL SOCIETY
OF **CHEMISTRY**

rsc.li/chemical-science

EDGE ARTICLE

Check for updates

Engineering mono- and multi-valent inhibitors on a modular scaffold†

Aurora Diamante,[a] Piyush K. Chaturbedy,[a] Pamela J. E. Rowling,[a] Janet R. Kumita, [ID] [a] Rohan S. Eapen, [ID] [a] Stephen H. McLaughlin,[b] Marc de la Roche,[c] Albert Perez-Riba [ID] ‡*[a] and Laura S. Itzhaki [ID] *[a]

Here we exploit the simple, ultra-stable, modular architecture of consensus-designed tetratricopeptide repeat proteins (CTPRs) to create a platform capable of displaying both single as well as multiple functions and with diverse programmable geometrical arrangements by grafting non-helical short linear binding motifs (SLiMs) onto the loops between adjacent repeats. As proof of concept, we built synthetic CTPRs to bind and inhibit the human tankyrase proteins (hTNKS), which play a key role in Wnt signaling and are upregulated in cancer. A series of mono-valent and multi-valent hTNKS binders was assembled. To fully exploit the modular scaffold and to further diversify the multi-valent geometry, we engineered the binding modules with two different formats, one monomeric and the other trimeric. We show that the designed proteins are stable, correctly folded and capable of binding to and inhibiting the cellular activity of hTNKS leading to downregulation of the Wnt pathway. Multivalency in both the CTPR protein arrays and the hTNKS target results in the formation of large macromolecular assemblies, which can be visualized both *in vitro* and in the cell. When delivered into the cell by nanoparticle encapsulation, the multivalent CTPR proteins displayed exceptional activity. They are able to inhibit Wnt signaling where small molecule inhibitors have failed to date. Our results point to the tremendous potential of the CTPR platform to exploit a range of SLiMs and assemble synthetic binding molecules with built-in multivalent capabilities and precise, pre-programmed geometries.

Introduction

The relationship between protein structure and function has been a cornerstone of biology for decades. However, in recent years, the unstructured or intrinsically disordered regions of the eukaryotic proteome (40% in humans) have gained increasing interest. This is due in part to the abundance in these regions of short independently functioning binding modules known as MoRFs (molecular recognition features) or SLiMs (short linear motifs).[1] One approach to exploit these motifs for inhibiting protein–protein interactions (PPIs) is to chemically synthesise them in combination with modifications such as cross-linking and macrocyclisation designed to improve affinity, half-life and cell penetration.[2] However, in nature high-affinity and

high-specificity interactions and more complex regulatory mechanisms are achieved through multivalency and avidity, neither of which are straightforward to realise with conventional peptide technologies. It is also uncertain whether small molecules could ever act as effective inhibitors of such complex networks of multivalent contacts.[3,4] Consequently, in order to interrogate and ultimately to drug such intricate networks of inter-linked motifs, new technologies are needed. There are many multivalent antibody technologies leveraging the natural modularity (*i.e.* multi-domain nature) of immunoglobulins.[5,6] However, multivalency has been less successful in antibody-like domains, where it has only been achieved by connecting monovalent units in "beads-on-a-string" or by directly assembling them on antibody scaffolds.[7–10] Likewise, functional peptide motifs have been assembled on synthetic chemical scaffolds,[11,12] DNA[13,14] and protein scaffolds,[15–21] but no modular platform has been developed to date that is capable of combinatorial incorporation of multiple SLiMs in a single stable and reliable scaffold as well as presenting them with varied, precise and programmable geometries. Here we show that tandem-repeat proteins possess all the necessary features with which to build such a platform.

Tandem-repeat proteins comprise tandem arrays of small structural motifs that pack in a linear fashion to produce

[a]Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1PD, UK. E-mail: lsi10@cam.ac.uk; albert.perezriba@utoronto.ca

[b]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0QH, UK

[c]Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc03175e

‡ Current address: Donnelly Centre for Cellular & Biomolecular Research, University of Toronto, Toronto, Canada.

regular, elongated, quasi-one-dimensional architectures and function in binding to other proteins, small molecules or nucleic acids. The repetitive modular organisation of this architecture makes it straightforward both to dissect and to redesign the biophysical properties.[22–27] One of the simplest repeat structures is the tetratricopeptide repeat (TPR),[28,29] a 34-residue motif comprising two antiparallel α-helices with

adjacent repeats connected by a short turn (Fig. 1A, bottom).[30] Artificial proteins comprising multiple copies of a consensus-designed TPR (CTPR) sequence have been shown to be extraordinarily stable.[31] Moreover, the modular nature of the architecture means that consensus repeats are self-compatible and can be individually designed and put together in any order. We recently showed that the CTPR scaffold can
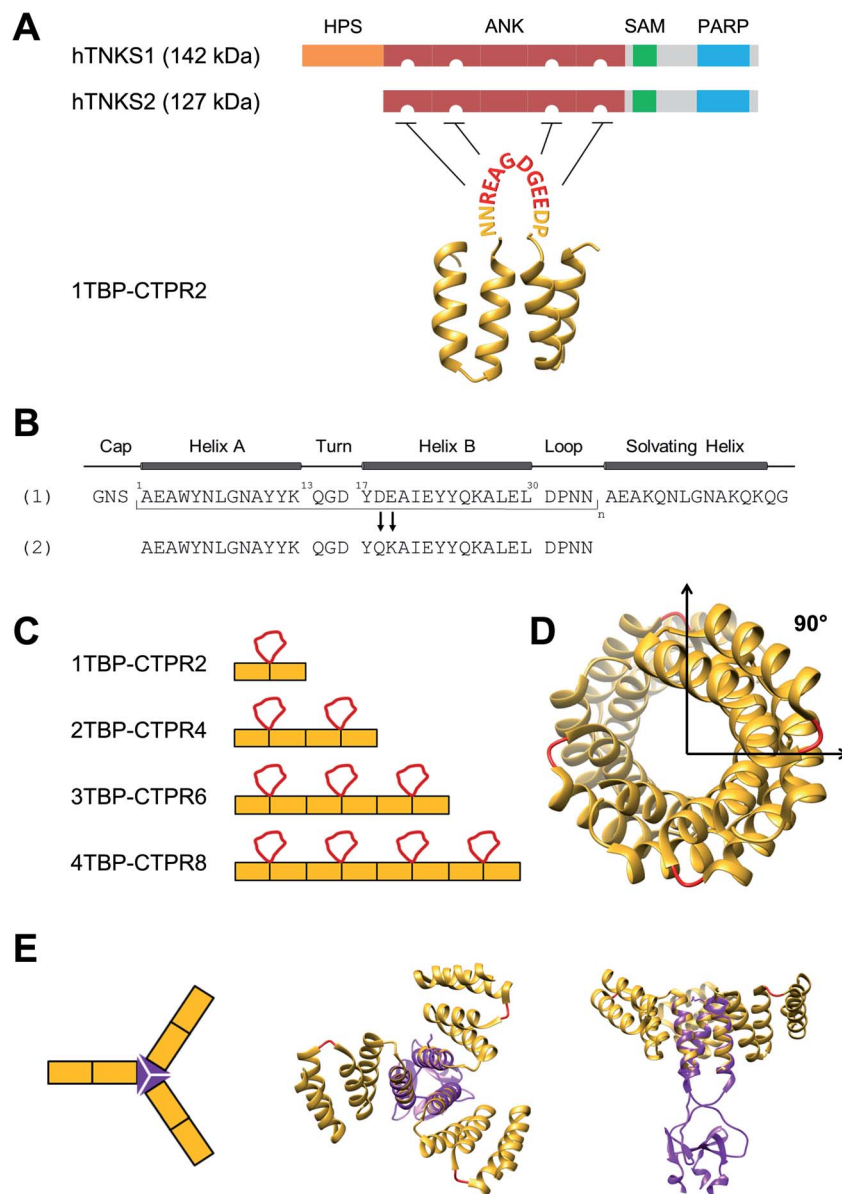
Fig. 1 Design of hTNKS-binding CTPR proteins in different formats and valencies. (A) (Top) Domain architecture of hTNKS1 and hTNKS2, comprising a histidine, proline, serine-rich (HPS) domain, a substrate-binding ankyrin-repeat (ANK) domain made up of five ankyrin-repeat clusters (ARC), a sterile alpha motif (SAM) and the catalytic PARP domain. (Bottom) Schematic representation of 1TBP-CTPR2 construct, showing the hTNKS-binding peptide (TBP) grafted onto the loop between adjacent CTPR repeats (PDB 2HYZ) in order to inhibit hTNKS activity. As denoted by the semi-circular cut-outs, four of the five ARCs can bind to the TBP. (B) Sequence of the CTPR motif designed by Main and co-workers (1) and that used for this study (2).[31,63,64] Secondary structural elements are shown. (C) Schematic representation of the nTBP-CTPR2n constructs. Each CTPR is shown as a yellow rectangle, and the TBP grafted onto the inter-repeat loops are in red. The amino-acid sequence of each construct is provided in Table S1.† (D) Crystal structure of CTPR8 (PDB 2HYZ) viewed along the superhelical axis. CTPR helices are in yellow and the inter-repeat loops, onto which the TBP sequence is grafted, are in red. (E) (Left) Schematic representation of CTPR2-foldon. Each CTPR is represented by a yellow rectangle and the foldon domain by a purple triangle. (Middle and Right) Different views of the modelled structure of CTPR2-foldon. The model was generated by grafting the foldon helix (PDB: 1OX3) onto the CTPR helix (PDB: 2HYZ) using the UCSF Chimera software 1.8.1.[65] CTPR helices are shown in yellow, the inter-repeat loops, onto which the TBP is grafted, are in red.

accommodate peptide extensions in the loop between adjacent repeats up to as many as 25 amino acids without compromising the native structure.[32,33] We then demonstrated that we could graft a SLiM from the protein Nrf2 that recognises the oncogenic protein Keap1 onto the scaffold and that we could not only recapitulate the native binding affinity but also improve it without need of sophisticated computational modelling. Thus, by combining this modular scaffold with SLiM grafting we potentially have the capacity for diverse functionalization.[33]

As proof of concept, we herein construct a set of monovalent and multivalent CTPR proteins to bind and inhibit the human poly (ADP) ribose polymerase (PARP) proteins tankyrase 1 and tankyrase 2 (referred to subsequently as hTNKS). hTNKS are unique in the PARP family in having a series of ankyrin-repeat cluster (ARC) subdomains (Fig. 1A).[34,35] Four of the five ARCs (ARCI, ARCII, ARCIV and ARCV) mediate protein–protein interactions including substrate recognition by binding short (8 residue) peptide motifs.[36–39] hTNKS has been implicated in the regulation of a number of cellular processes,[40–43] but more recently the role of hTNKS in the regulation of Wnt pathway activity has received particular attention.[44,45] The Wnt pathway is the principle driver of colon cancer, and a number of studies have reported the development of small molecule inhibitors of the hTNKS catalytic activity as a means of modulating Wnt pathway activity.[44,46–49] However, by targeting the catalytic PARP domain these small molecules may lack specificity for hTNKS over other PARP family members.[50–52] Moreover, they can only inhibit the catalytic function of hTNKS but not the well-documented non-catalytic functions.[53,54] To overcome these limitations, our aim was to target the unique substrate-binding ARC subdomains and thereby block both catalytic and non-catalytic functions. Guettler *et al.* previously determined a consensus sequence that is recognised by the ARC subdomains (referred to subsequently as the tankyrase-binding peptide or TBP)[36] (Fig. 1A), and we have shown that macrocyclised TBPs are able to bind hTNKS and inhibit Wnt signalling.[55]

We grafted the TBP onto the inter-repeat loop of the CTPR scaffold to generate a series of mono- and multi-valent binding proteins (nTBP-CTPR2n). Our aim was to explore the effects of multivalency on the biophysical properties of the proteins and on their interactions with and cellular inhibition of hTNKS, which is itself multivalent. Further motivation came from the fact that several hTNKS substrates are themselves multivalent as they contain multiple TBPs.[56] In order to further extend the multivalent capabilities and to produce more complex binding geometries beyond what is possible with the linearly arrayed nTBP-CTPR2n format, we engineered a trimeric 'foldon' motif into the constructs (nTBP-CTPR2n-foldon).[57–61] Our results highlight the power of multivalent inhibitors for targeting complex protein–protein interactions, where occupancy-driven inhibition using simple monovalent molecules is unlikely to be effective.

Results

nTBP-CTPR2n proteins: design of mono- and multi-valent tankyrase inhibitors

We functionalised the CTPR scaffold by grafting an 8-residue hTNKS-binding consensus peptide (TBP), REAGDGEE,

identified from a mutational analysis of hTNKS substrates,[36] onto the loop between two adjacent repeats (Fig. 1A). This minimal hTNKS-binding unit was then tandemly repeated to generate a series of mono- and multi-valent molecules named nTBP-CTPR2n, with n between one and four (Fig. 1C). The CTPR proteins adopt a superhelical conformation with eight repeats required to complete a superhelical turn.[62] Thus, in the nTBP-CTPR2n construct design, the TBP loops will be offset from each other by approximately 90° (Fig. 1D). All nTBP-CTPR2n constructs could be expressed and purified in high yields in *E. coli* (~20 mg per litre of culture). All proteins were soluble in solution.

Effect of TBP insertions on CTPR folding and stability

We first investigated the effects on protein stability of grafting the TBP sequence onto the inter-repeat loop using far-UV circular dichroism (CD). Given the high α-helical content of the nTBP-CTPR2n constructs, all CD spectra have the expected double minima at 208 nm and 222 nm similar to those observed previously for the CTPR proteins (Fig. S1A,† left),[32] confirming that all nTBP-CTPR2n constructs were correctly folded. There was a linear increase in the molar ellipticity at 222 nm with increasing size of the CTPR construct, which further confirms that the proteins were natively folded (Fig. S1A,† right).

Next, thermal unfolding was performed to investigate the stability of the nTBP-CTPR2n constructs. We used the ellipticity at 222 nm as a measure of α-helical content. All proteins were found to be extremely thermostable, with only the smallest protein, 1TBP-CTPR2, undergoing complete denaturation. The larger nTBP-CTPR2n proteins remained partly folded even at the highest temperature, confirming that the stability increases with increasing number of repeats. The midpoint of the unfolding transition (T_M) is therefore provided only for 1TBP-CTPR2 and corresponds to 75 °C (Fig. S1B†). Following thermal denaturation, the samples were allowed to return to 20 °C, and the CD spectra re-measured to evaluate the reversibility of the reaction. For all of the proteins, there was no significant difference between the CD spectrum before and after thermal denaturation, indicating that unfolding is reversible (Fig. S1C†).

Chemical-induced denaturation experiments were also performed, monitoring the fluorescence of the tryptophan residues (one per repeat) (Fig. S2†). A single transition was observed for all of the constructs, and the data were fitted to a two-state model to give the midpoints of unfolding ($D_{50\%}$), m-values ($m_{D–N}$, a parameter related to the change in solvent-exposure upon unfolding) and the free energies of unfolding ($\Delta G_{D–N}^{H_2O}$) (Table S2†). There was a significant increase in $D_{50\%}$ and m-value between 1TBP-CTPR2 and 2TBP-CTPR4 and smaller increases for the larger proteins. Lastly, the nTBP-CTPR2n showed a decrease in the stability compared with their respective counterparts without the TBP loops. As previously shown,[32] loop insertion has a destabilizing effect through both the entropic cost of loop closure and a consequent weakened coupling of adjacent repeats.

We next investigated whether all of the TBP loops in the multi-valent constructs are accessible for binding by measuring the
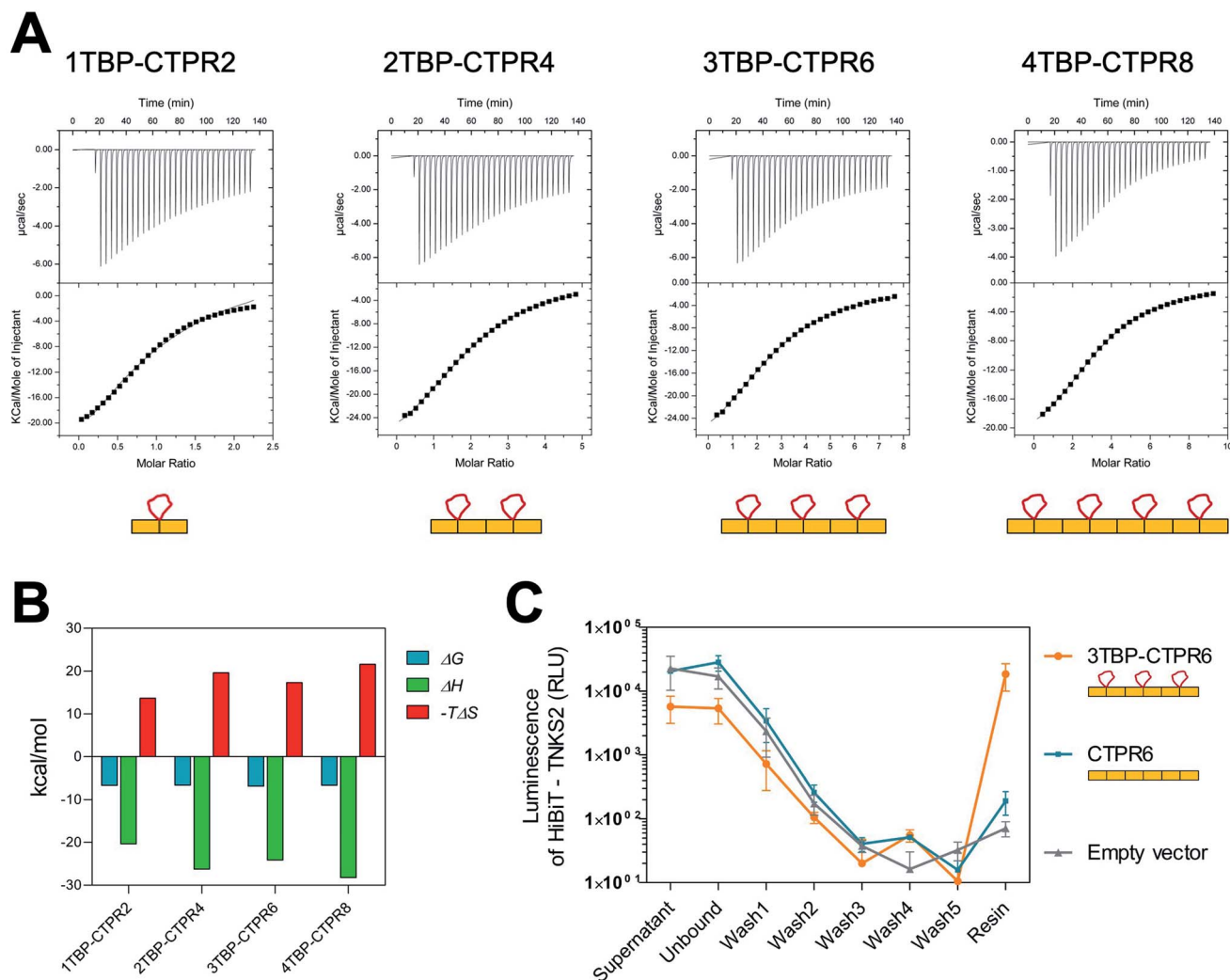
Fig. 2 Binding of TBP-CTPR proteins to hTNKS2 *in vitro* and in the cell. (A) Representative ITC traces from left to right: hTNKS2 ARC4 (500 μM) into 1TBP-CTPR2 (50 μM), hTNKS2 ARC4 (500 μM) into 2TBP-CTPR4 (25 μM), hTNKS2 ARC4 (500 μM) into 3TBP-CTPR6 (16.6 μM), hTNKS2 ARC4 (500 μM) into 4TBP-CTPR8 (12.5 μM); experiments were performed at 25 °C. The concentrations of nTBP-CTPR2n used were chosen so that the molar ratio between the number of TBP loops and TNKS2 ARC4 protein was constant. (B) Thermodynamic parameters (ΔG, ΔH and $-T\Delta S$) obtained by ITC for the binding of the four nTBP-CTPR2n proteins to hTNKS2 ARC4. For each parameter, the average from two independent experiments is plotted. (C) HiBiT-tagged hTNKS2 protein levels measured by luminescence intensity throughout each step of the HiBiT-qIP method. IP was performed in the presence of HA-tagged 3TBP-CTPR6, CTPR6 and an empty vector. Average values and standard deviation from two independent biological replicates are shown.

affinity and stoichiometry for each nTBP-CTPR2n construct binding to hTNKS2 ARC4 (the fourth ankyrin-repeat cluster of hTNKS2), which contains a single binding site for the TBP (Fig. 2A). All nTBP-CTPR2n constructs showed similar low-micromolar affinities, and the stoichiometry was found to increase from 1TBP-CTPR2 to 3TBP-CTPR6 in proportion to the number of TBP loops, confirming that all sites are available for binding (Table 1). For 4TBP-CTPR8, the stoichiometry is lower than 4, likely because steric hindrance precludes binding of four hTNKS2 ARC4 molecules to one 4TBP-CTPR8 molecule. No binding could be detected for a control CTPR6 protein containing no TBP (Fig. S3†). The changes in the free energy (ΔG), enthalpy (ΔH) and entropy ($-T\Delta S$) for the binding of the four nTBP-CTPR2n proteins to hTNKS2 ARC4 are plotted in Fig. 2B and

Table 1 Stoichiometry (*N*) and dissociation constant (K_d) obtained by ITC for the binding of nTBP-CTPR2n proteins to hTNKS2 ARC4. For each protein, the values listed are from two independent experiments

Protein	Stoichiometry (*N*)	K_d (μM)
1TBP-CTPR2	1.01 ± 0.03	14.68 ± 0.99
	0.90 ± 0.02	12.80 ± 1.10
2TBP-CTPR4	2.18 ± 0.02	11.89 ± 0.26
	2.05 ± 0.01	18.59 ± 0.39
3TBP-CTPR6	2.72 ± 0.03	5.62 ± 0.21
	2.77 ± 0.02	20.12 ± 0.39
4TBP-CTPR8	3.38 ± 0.01	13.61 ± 0.19
	2.91 ± 0.03	15.48 ± 0.43

listed in Table S3.† The results show that in all cases the interaction is enthalpy-driven. There is a slight tendency for both the enthalpic gain and entropic cost of binding to increase with increasing number of hTNKS2-binding loops. The increase in enthalpy with increasing number of binding loops is to be expected, as having multiple binding sites in close proximity should make it more likely for the protein to bind a target again after dissociation. The greater entropic cost of ordering a greater number of loops upon binding is also as expected. The greater entropic cost is offset by the greater enthalpic gain, and hence the values of ΔG are similar for all four nTBP-CTPR2n proteins.

nTBP-CTPR2n proteins and hTNKS interact in the cell

Binding between hTNKS and nTBP-CTPR2n in the cell was confirmed using a modified version of the HiBiT-based quantitative immunoprecipitation (HiBiT-qIP) method, that provides a much faster and more sensitive read-out than conventional IP by western blot.[66] The method relies on the split luciferase complementation of the two NanoLuc fragments, HiBiT and LgBiT, providing a sensitive and quantitative method to track any HiBiT-tagged protein.[67] HA-tagged nTBP-CTPR2n was used as bait to pull-down HiBiT-tagged hTNKS2 using an anti-HA resin. The presence of HiBiT-tagged hTNKS2 was followed and quantified through the luminescence signal generated by the HiBiT technology, allowing a real-time measurement of the pulled-down target throughout the process. The results show that HiBiT-tagged hTNKS2 is pulled down by 3TBP-CTPR6 but not by the control CTPR6 construct, confirming that the grafted TBP loop mediates the binding to hTNKS2 within the cellular environment also. Moreover, no non-specific binding of HiBiT-tagged hTNKS2 to the Anti-HA Agarose resin was detected (Fig. 2C). Nano-Glo HiBiT blotting further confirmed the presence of HiBiT-tagged hTNKS2 in the eluted samples (Fig. S4†). Likewise, the bait constructs and tubulin were detected by standard western blot methods (Fig. S4†).

Design of trimeric nTBP-CTPR2n-foldon constructs

We introduced the foldon motif at the C-terminus of each nTBP-CTPR2n construct to induce trimerisation. In this way, each homo-trimeric construct will display up to twelve TBP loops and in a different geometry compared to the corresponding monomeric constructs. The foldon is a short (30 residue) but highly stable motif of the protein fibritin from the T4 bacteriophage.[57] The so-called 'NCCF' construct of fibritin contains an N-terminal trimeric coiled-coil region followed by the C-terminal foldon.[68] Here, we introduced the foldon and part of its preceding coiled-coil domain into the nTBP-CTPR2n constructs by grafting the helical coiled-coil onto a newly introduced C-terminal CTPR helix (Fig. S5,† top). This helix corresponds to an extra helix A (*i.e.* half of one CTPR). As all α-helices share the same heptad conformation, it is possible to graft residues of a particular heptad position from a 'guest' helix onto a 'host' helix. When grafting a new site onto a CTPR helix, the native residues at the interface between the helix and the neighbouring repeat need to be preserved in order for the grafted helix to remain folded onto the rest of the CTPR scaffold. These residues

were identified as Ala3, Leu7 and Ala10 of helix A, which are involved in the inter-repeat interface with the helix B of an adjacent repeat. Part of the remaining residues of helix A were substituted with the residues of the fibritin NCCF coiled-coil domain (residues 65 to 77) involved in the trimeric interface. This was followed by an insertion of the remaining C-terminal end of the coiled-coil domain and the foldon (NCCF residues 78 to 109) (Fig. S5,† top). The resulting constructs are referred as nTBP-CTPR2n-foldon, and the protein sequences are given in Table S1.† A model of the trimeric CTPR2-foldon protein is shown in Fig. 1E. This protein is expected to display the TBP loops at an angle of 120° to one another. The nTBP-CTPR2n-foldon proteins were expressed and purified as for their monomeric counterparts. Analysis by native gel electrophoresis is consistent with trimerisation of those proteins that contain the foldon domain (Fig. S5,† bottom). The series lacking the foldon domain (1TBP-CTPR2, 2TBP-CTPR4, 4TBP-CTPR8, calculated molecular weights 11.3 kDa, 20.6 kDa and 39.4 kDa, respectively) run at increasing molecular weights between the 20 kDa and 66 kDa molecular weight markers. In contrast, those with the foldon domain (1TBP-CTPR2-foldon, 2TBP-CTPR4-foldon, 4TBP-CTPR8-foldon, calculated molecular weights as trimers 49.2 kDa, 77.4 kDa and 133.5 kDa, respectively) run at higher molecular weights between the 66 kDa and 146 kDa markers.

The oligomerization was further analysed by size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS). The results show that 3TBP-CTPR6 and 3TBP-CTPR6-foldon (representative of the two types of nTBP-CTPR2n constructs) are, respectively, monomeric and trimeric in solution (Fig. S6†). The average molecular weight obtained for 3TBP-CTPR6 (30 kDa calculated molecular weight) is 38 kDa, consistent with it being monomeric. The majority of the 3TBP-CTPR6-foldon sample (105 kDa calculated molecular weight as a trimer) elutes at an average molecular weight of 131 kDa, consistent with it forming a trimer with some less-populated oligomeric species at lower and higher molecular weight also present (99, 204 and 321 kDa, respectively).

Multivalent interactions induce the formation of large macromolecular assemblies with variable internal dynamics

When two multivalent proteins interact, they have the potential to engage with multiple partners simultaneously, leading to the formation of large supramolecular assemblies (see for example work by Rosen and colleagues[69]). This is indeed what we observe with hTNKS2 and the multivalent tankyrase-binding CTPR proteins (schematically represented in Fig. 3A). These species could be detected using a co-precipitation assay, in which the multivalent 3TBP-CTPR6 and 3TBP-CTPR6-foldon constructs, chosen as representatives of the linear and trimeric arrays, respectively, co-precipitate in the presence of the multivalent hTNKS2 construct corresponding to the first three ARC sub-domains (hTNKS2 ARC1–3). It is worth noting that, despite 3TBP-CTPR6 and 3TBP-CTPR6-foldon being incubated at the same TBP molar concentration, the trimeric 3TBP-CTPR6-foldon construct precipitated to a higher extent and lead to an
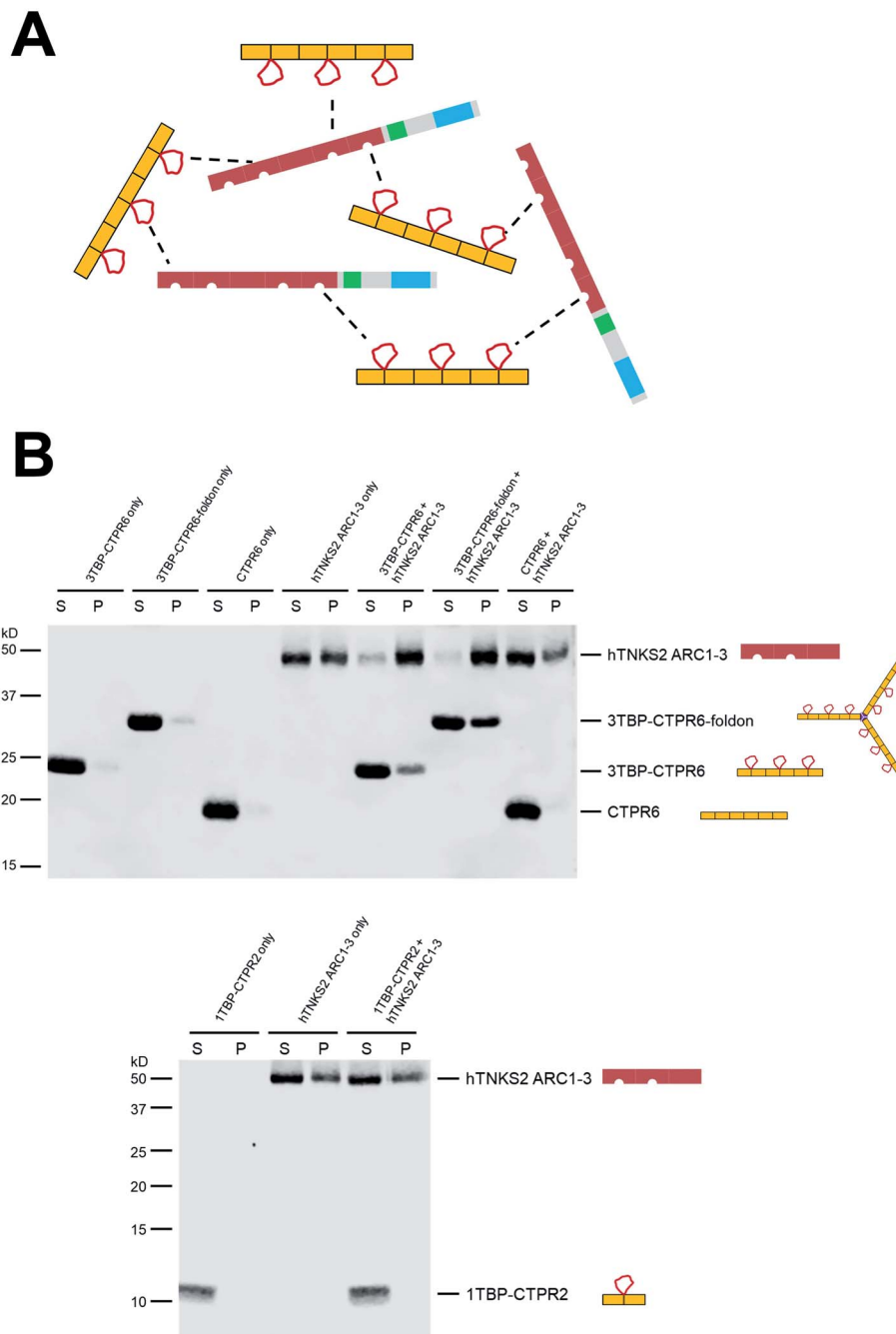
Fig. 3 Multivalent interactions between TBP-CTPR proteins and hTNKS2 ARC1–3 analysed using a co-precipitation assay. (A) Schematic representation of the macromolecular assemblies formed by the interaction of the multivalent hTNKS2 protein and the 3TBP-CTPR6 protein (used as representatives of the multivalent CTPR arrays). (B) After centrifugation, supernatant (S) and pellet (P) were separated and run on a 12% (top) or 16% (bottom) SDS polyacrylamide gel. For co-sedimentation, proteins were mixed in equal volumes at the following concentration: 10 μM 3TBP-CTPR6 (30 μM TBP loop concentration), 3.3 μM 3TBP-CTPR6-foldon (30 μM TBP loop concentration), 10 μM CTPR6 and 10 μM hTNKS2 ARC1–3 (20 μM TBP-binding sites). Gel images were obtained using Li-COR Odyssey Fc Imaging System.

increased precipitation rate of hTNKS2 ARC1–3. In contrast, the mono-valent 1TBP-CTPR2 construct and the control CTPR6 construct remained in the supernatant, confirming that multi-valency of TBP loops is required for the formation of a large supramolecular assembly with multivalent hTNKS2 (Fig. 3B).

Titration of increasing concentrations of 3TBP-CTPR6 or 3TBP-CTPR6-foldon proteins into hTNKS2 ARC1–3 showed that

the extent of co-precipitation was dependent on the molar ratio of the two interacting molecules, until saturation is reached. Analysis of the gel band intensities allowed us to calculate the molar concentration of each species in the pellet, providing an estimation of the stoichiometries of the macromolecular complexes that had precipitated (Fig. S7 and Table S4†). The results indicate a 1 : 1 stoichiometry for the 3TBP-
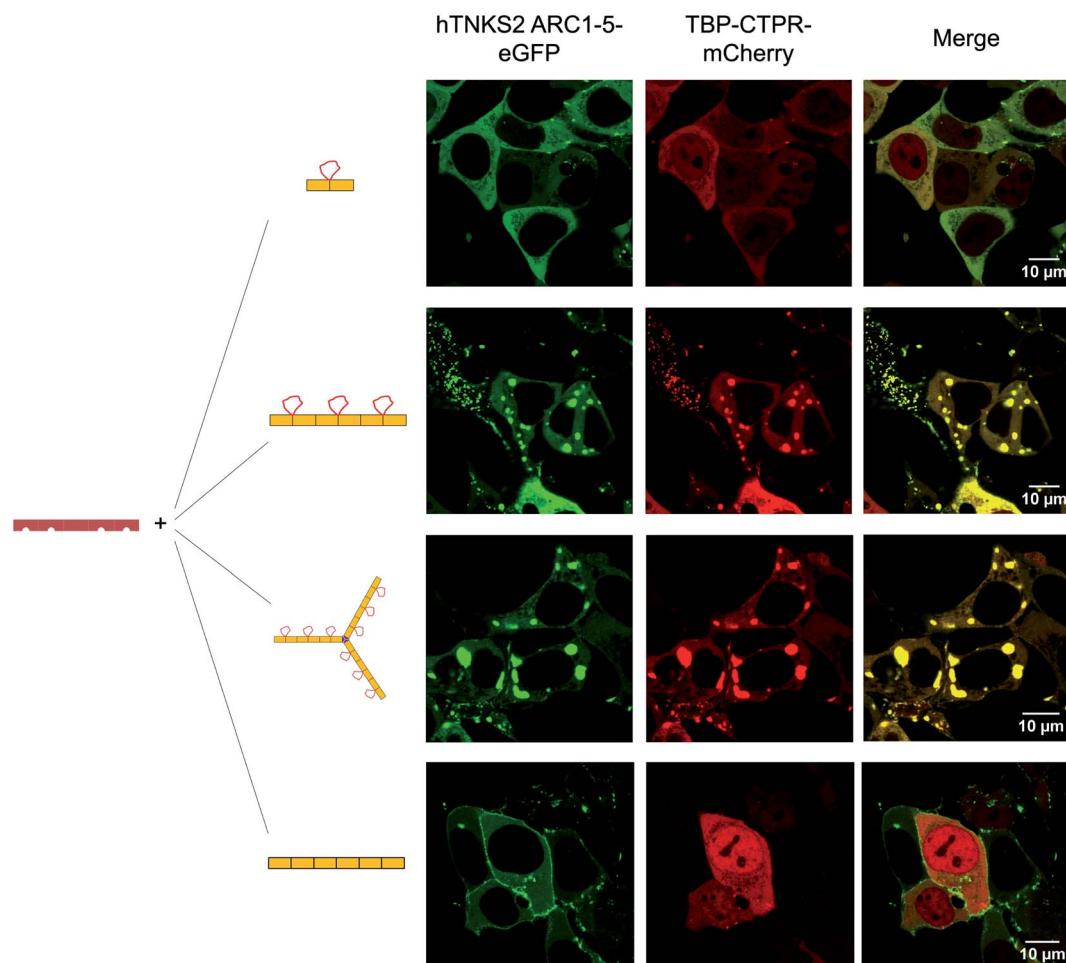
Fig. 4 Fluorescence microscopy of HEK293T cells co-transfected with mCherry-tagged CTPR proteins and eGFP-tagged hTNKS2 ARC1–5. Co-localisation in large macromolecular clusters is observed for eGFP-tagged hTNKS2 ARC1–5 in combination with mCherry-tagged 3TBP-CTPR6 or with mCherry-tagged 3TBP-CTPR6-foldon. These clusters are not observed for the mono-valent mCherry-tagged 1TBP-CTPR2 protein or the control mCherry-tagged CTPR6 protein. Scale bars for all images are 10 μm.

CTPR6 : hTNKS2 ARC1–3 complex, and a higher stoichiometry of 1 : 2 for the 3TBP-CTPR6-foldon : hTNKS2 ARC1–3 complex, explaining why the trimeric 3TBP-CTPR6-foldon can induce a greater extent of hTNKS2 ARC1–3 co-precipitation than the linear 3TBP-CTPR6.

Next, these protein assemblies were examined using negative stain Transmission Electron Microscopy (TEM). The results shown in Fig. S8† confirm that the large, macromolecular clusters are formed only when both interacting partners are multivalent.

To visualise the effect of multivalency inside the cell, HEK293T cells were co-transfected with an hTNKS2 construct encoding the ARC1–5 subdomains fused to eGFP (referred to as hTNKS2 ARC1–5–eGFP) and the multivalent 3TBP-CTPR6 or 3TBP-CTPR6-foldon constructs fused to mCherry. Large, polymeric species were visible in the cytoplasm, where the signals of eGFP and mCherry perfectly co-localised (Fig. 4). In agreement with the findings of the co-precipitation assays, these large assemblies were formed only when the multivalent 3TBP-CTPR6 and 3TBP-CTPR6-

foldon proteins were co-expressed with hTNKS2 ARC1–5, which is also multivalent. In contrast, when the same CTPR proteins were co-expressed with a monovalent hTNKS2 construct comprising only the first ARC domain (hTNKS2 ARC1–eGFP), no polymeric assemblies were observed (Fig. S9†). Likewise, when the multivalent hTNKS2 was co-expressed with the monovalent 1TBP-CTRP2, no polymeric species were observed (Fig. 4). To further compare the properties of the macromolecular assemblies generated in the cell by 3TBP-CTPR6 and 3TBP-CTPR6-foldon bound to hTNKS2 ARC1–5, Fluorescence Recovery After Photo-bleaching (FRAP) was used to assess the dynamics by measuring the recovery rate of the eGFP signal (Fig. 5). The assemblies induced by 3TBP-CTPR6 showed fluorescence recovery within 30 seconds, suggesting that the proteins within the assembly are dynamic in nature. In contrast, the assemblies generated by the 3TBP-CTPR6-foldon showed no fluorescence recovery within the same timeframe, suggesting that the presence of nine hTNKS binding sites on 3TBP-
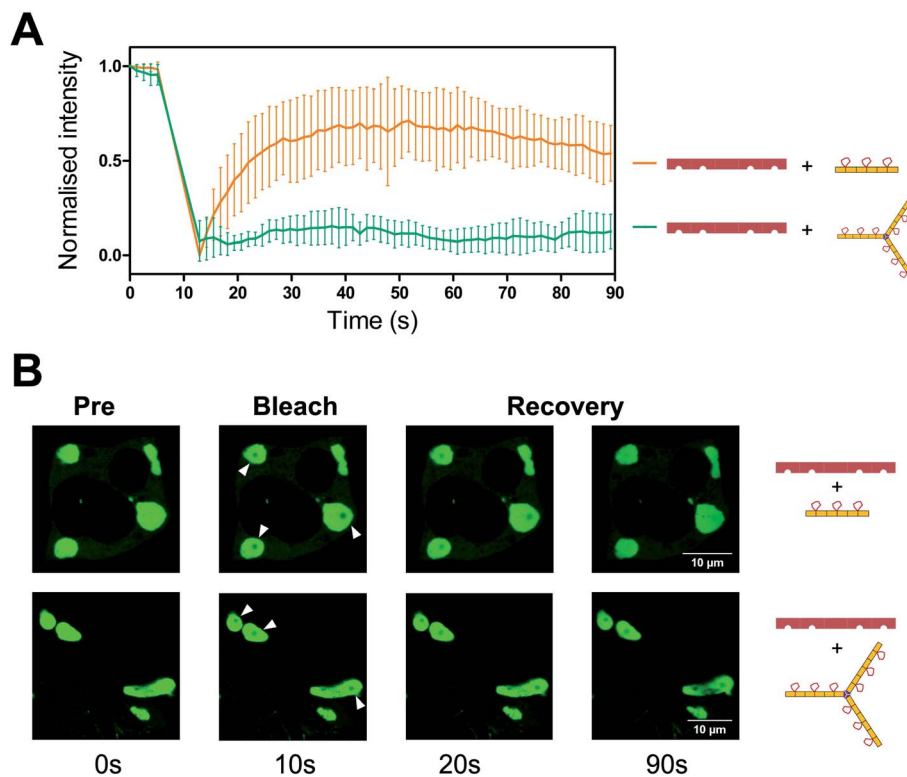
Fig. 5 FRAP analysis of the assemblies formed between hTNKS2 ARC1–5 and TBP-CTPR proteins. (A) FRAP analysis was performed on seven individual Regions of Interest (ROI) selected within macromolecular assemblies induced by eGFP-tagged hTNKS2 ARC1–5 in complex with 3TBP-CTPR6 (orange) or 3TBP-CTPR6-foldon (green). HEK293T cells were bleached in the ROI and fluorescence recovery of eGFP-tagged hTNKS2 ARC1–5 was monitored over 90 seconds. (B) Representative ROI selected within three independent macromolecular assemblies are shown before, during and post bleaching.

CTPR6-foldon leads to the formation of a more interconnected, rigid assembly.

### *n*TBP-CTPR2*n* and *n*TBP-CTPR2*n*-foldon proteins inhibit Wnt signalling

We next evaluated the inhibitory capabilities of the monomeric and trimeric *n*TBP-CTPR2*n* constructs on Wnt signalling using the TOPFLASH reporter assay. Treatment of HEK293T cells with all the *n*TBP-CTPR2*n* constructs, with the exception of 1TBP-CTPR2, led to a reduction in Wnt pathway activity compared with the controls (empty vector, CTPR2 and CTPR6 and 3RL-CTPR6 containing a non-binding peptide named 'RL', random loop) (Fig. 6A). To assess the levels of *n*TBP-CTPR2*n* proteins in the cell during the TOPFLASH assay, we used N-terminally HiBiT-tagged *n*TBP-CTPR2*n* proteins. The 11-amino-acid HiBiT tag allows a more sensitive and quantitative measurement than western blot *via* the antibody-free split NanoLuc® luciferase technology (Promega). The results show that the extent of Wnt inhibition is proportional to the *n*TBP-CTPR2*n* protein levels as quantified 24 h after transfection (Fig. S10†). In particular, it is clear that the CTPR2 protein, whether in the monomeric or the trimeric format, is present at much lower levels than the larger proteins. To test whether the CTPR2 constructs are subject to proteasome-induced degradation, transfected wells were incubated with the proteasome inhibitor

MG132 for 5 hours. The results show that there is an approximate 3-fold increase in protein levels in the presence of MG132, the exceptions being the two smallest constructs 1TBP-CTPR2 and CTPR2, whose levels are affected to a greater extent (Fig. S11†). Thus, the smallest, least thermodynamically stable CTPR protein is the most susceptible to proteasome-mediated degradation, and the foldon domain enhances its intracellular stability. In addition, the formation of the large macromolecular assemblies is likely to protect the multivalent hTNKS-binding CTPR proteins from proteolysis, further explaining why the monovalent 1TBP-CTPR2 and the non-binding control CTPRs do not accumulate in the cell. Lastly, we confirmed that none of the constructs showed cytotoxicity (Fig. S12†).

Intracellular delivery of 3TBP-CTPR6 by encapsulation with fusogenic liposomes induces high levels of Wnt inhibition

Intracellular delivery of proteins is challenging.[70] We used an encapsulation approach in which purified protein (3TBP-CTPR6) was loaded into fusogenic liposomes. Fusogenic liposomes have been shown to merge with the cell membrane and deliver their cargo directly into the cell cytoplasm.[71] Encapsulation of 3TBP-CTPR6 in liposomes could minimise their premature degradation and adverse immune response, if any.[72] Liposomal formulation of 3TBP-CTPR6 (FL-3TBP-CTPR6) and control liposomes without protein (FL) were prepared as
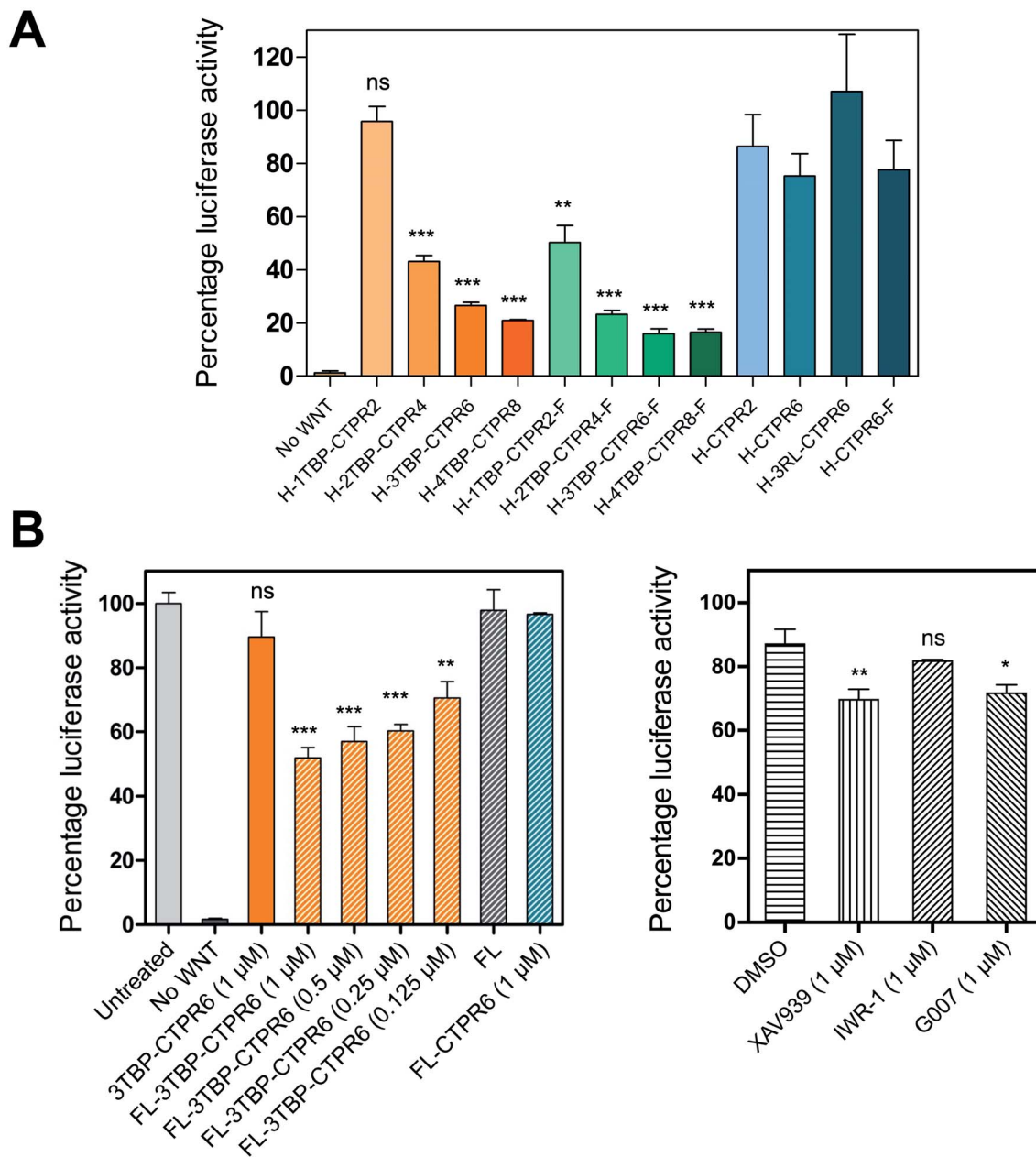
Fig. 6 The effect on Wnt-activated HEK293T cells following treatment with the indicated HiBiT-tagged nTBP-CTPR2n constructs. (A) TOP-FLASH reporter assay for HEK293T cells transfected with TCF-Firefly and Renilla reporter gene vectors and an expression vector encoding the constructs listed. For each sample, Luciferase activities were normalised with the Renilla values and the ratio was expressed as relative luciferase activity to the control well transfected with the empty HiBiT vector set at 100% (not shown in the graph). The monomeric constructs, trimeric constructs and controls are shown in different shades of orange, green and blue, respectively. F indicates the foldon motif, H indicates the N-terminal HiBiT tag. Standard deviation was calculated from triplicate sample measurements. The significance of the difference between samples (ns: non-significant, $p > 0.05$, $*p \leq 0.05$ $**p \leq 0.01$, $***p \leq 0.001$) was assessed using One-way ANOVA coupled with Dunnett's Multiple Correction test. 1TBP-CTPR2 and 1TBP-CTPR2-F were compared to CTPR2. Multivalent linear and trimeric constructs were compared to 1TBP-CTPR2 and 1TBP-CTPR2-F, respectively. (B) (left) The figure on the left shows the effect on Wnt signaling of treatment with fusogenic liposome-encapsulated 3TBP-CTPR6. Each treatment was with 20 μL of liposomes. In brackets are the concentrations of the proteins used. For each run, data were normalised by the untreated control well, set at 100%. Bars with diagonal stripes correspond to samples treated with liposomes. No Wnt: cells without Wnt pathway activation and not treated with liposomes. Untreated cell: cells not treated with liposomes. FL: empty liposomes. Error bars were obtained from triplicate sample measurements from two independent experiments. The same statistical analysis was performed as in (A), comparing the samples to FL-CTPR6. (B) (right) For comparison, the figure on the right shows the effect on Wnt signaling of hTNKS small molecule inhibitors used interventionally at the indicated concentration. Data were normalised relative to the untreated control well, which was set at 100% (not shown in the graph). Error bars were determined from two independent sample measurements. The same statistical analysis was performed as in (A), comparing the small molecules to DMSO.

described in the experimental section. Both FL and FL-3TBP-CTPR6 had highly positive surface with zeta potential (ZP) values of +126 mV and +75.7 mV, respectively, at pH 7.4 (Fig. S13†). The lower ZP of FL-3TBP-CTPR6 is due to encapsulation/association of the negatively charged 3TBP-CTPR6 (pI $\sim$ 4.8) with liposomes. The hydrodynamic sizes of FL and FL-3TBP-CTPR6 were similar, at $\sim$106 nm (Fig. S13†). FL and FL-3TBP-CTPR6 did not show any significant cytotoxicity even for amounts three times higher than those used in our experiments (Fig. S14†). To visualize intracellular delivery under the confocal microscope, 3TBP-CTPR6 was fluorescently labelled with rhodamine. Liposomes were prepared using the labelled protein and designated as FL-3TBP-CTPR6-RITC. Confocal images of HEK293T cells treated with FL-3TBP-CTPR6-RITC (Fig. S15†) clearly suggest that protein is inside the cells and is distributed throughout the cytoplasm. Cells treated with FL did not show any signal for 3TBP-CTPR6-RITC (Fig. S16†).

We next tested the effect of liposome-delivered 3TBP-CTPR6 on Wnt signaling using the TOPFLASH assay. A decrease in the TOPFLASH activity was observed, reaching around 50% after 6 hours incubation. Importantly, decreasing TOPFLASH activity correlated with increasing concentration of 3TBP-CTPR6 (Fig. 6B, left), indicating a dose-dependent inhibition of hTNKS. Treatment with free 3TBP-CTPR6 did not affect the luciferase expression levels as the protein cannot enter cells on its own. Also, FL did not alter luciferase expression, indicating that membrane fusion of liposomes does not interfere with the intracellular Wnt signalling. Further, control protein, CTPR6, lacking hTNKS-binding units and delivered using liposomes, did not change the intracellular luciferase levels.

For comparison, we also performed the TOPFLASH assay in the presence of three well-characterized small molecule hTNKS inhibitors (XAV939, IWR-1 and G007) that bind the catalytic domain of hTNKS. Critically, whereas the small molecule inhibitors are effective when used prophylactically (i.e. before Wnt pathway activation) (Fig. S17†), they are ineffective when used interventionally, i.e. after Wnt pathway activation (as would be the case in the clinic, since patients will have elevated Wnt signaling) (Fig. 6B, right). Strikingly, FL-3TBP-CTPR6 at the highest dose had a significantly greater inhibitory effect than any of the small molecule hTNKS inhibitors (Fig. 6B). This result shows that targeting the non-catalytic activity of hTNKS could be an effective treatment where small-molecule inhibitors have so far failed.

Discussion

In this study, we created potent hTNKS inhibitors by combining target specificity with multivalency, two features that have not been explored in previous drug development efforts against this target. We demonstrate that modular CTPR proteins can be engineered to display both single as well as multiple copies of short hTNKS binding motifs by grafting the TBP onto the loops between adjacent repeats. Helical grafting was also applied to engineer trimeric CTPR constructs with enhanced multivalent capabilities. We show that CTPR proteins with one or more grafted hTNKS-binding motifs are stable, correctly folded and exceptionally active in inhibiting Wnt

signalling. The effect of multivalency in both partners of the interaction – hTNKS and the designed nTBP-CTPR2n proteins – manifests in the formation of large, intracellular macromolecular assemblies with different dynamics dependent on the configuration (i.e. monomeric versus trimeric). Thus, hTNKS inhibition by these multivalent molecules is enhanced by clustering of the protein within these structures. Such a mechanism of action warrants further investigation and should be explored for future therapeutic efforts against hTNKS and other multivalent targets. Our experiments using fusogenic liposome delivery of 3TBP-CTPR6, one of our most potent molecules, further demonstrated rapid inhibition of Wnt pathway activity within 6 hours of treatment of HEK293T cells pre-stimulated with Wnt3A ligand. This is a noteworthy result that demonstrates interventional inhibition of the pathway, the scenario required for targeting tumours dependent on deregulated Wnt pathway activity in vivo. Moreover, the luciferase protein reporter of Wnt pathway activity has a relatively long half-life of approximately 12 hours, and therefore inhibition after 6 hours treatment likely reflects a much higher attenuation of Wnt pathway activity beyond the measured 50% inhibition.

We compared our molecules with three well-characterised small molecule hTNKS inhibitors and, in agreement with previous studies, we observed around 90% inhibition of Wnt pathway activity for two of them when used prophylactically: i.e. when they are applied either before Wnt pathway activation by Wnt3A ligand or in combination with it. Critically, however, to be of clinical use, such drugs will need to work interventionally; interventional treatment with these small molecule hTNKS inhibitors has been shown by our group and others to be much less effective due to a cell-intrinsic feed-forward mechanism preserving Wnt pathway activity.[73] Thus, the effectiveness of liposome-delivered 3TBP-CTPR6 for interventional inhibition of Wnt pathway activity demonstrated here further highlights the power of our approach.

The work presented here points to the tremendous potential of the repeat-protein scaffold for the rational design of multivalent – and thereby potentially multi-specific – molecules for inhibiting PPIs: (1) First, CTPR proteins are small and ultra-stable without need of disulphide bonds. As a result, they can accommodate small or even large sequence insertions yet remain folded and stable.[32,74] Moreover, multiple such insertions are also possible without compromising the fold; it is doubtful whether other small single-domain scaffolds based on globular structures could do the same. (2) The second key consequence of the repeat-protein stability is that they can be produced recombinantly in very high yields. (3) Lastly, the repetitive, modular nature of the scaffold is a particularly unique selling point; we have shown here that we can exploit these characteristics to display one or multiple binding motifs in a precise and programmable manner. It has been estimated that there are around 100 000 such SLiMs in the human proteome,[75] each one of which provides a potential starting point for drug discovery. The simple cut-and-paste approach used here, requiring little or no in silico design procedures, could be applied to harness some of these SLiMs, and the platform thus has potential to be used as a synthetic tool in diverse

applications ranging from target validation to PPI inhibition, pathway modulation and ultimately molecular therapeutics.

Our confidence in the broad applicability of the scaffold to the display of diverse SLiMs arises from the finding that we could effectively graft two SLiMs with very different binding conformations onto the CTPR loops. The Keap1-binding peptide from Nrf2 adopts a tight turn-like conformation,[33] whereas the TBP in this work binds to ARC domains in a highly extended conformation.[36,55]

It might appear that a limiting factor in this multivalent platform would be the potential for steric clashes between target and scaffold as well as between targets in a multivalent display. The first challenge should be approached on a case-by-case basis, as it is common to all peptide display methodology. The second problem can be mitigated by exploiting the modular nature of this technology. For example, the ITC data show that both 3TBP-CTPR6 and 4TBP-CTPR8 bind to approximately three hTNKS2 ARC4 molecules, suggesting that a maximum has been reached. We limited our grafting to one peptide for every two repeats, but peptides could be set further apart by increasing the number of intervening repeats. The rigidity of the CTPR scaffold ultimately works as a molecular ruler of predefined length and pitch. Furthermore, we have shown that trimerisation by means of the small foldon domain enhances the multivalent capability of the CTPR scaffold, presumably by providing more degrees of freedom for the formation of macromolecular complexes. Future work will focus on building bi- and multi-functional CTPR arrays using multiple SLiMs to bring two or more different targets together in a predefined geometry, for example to redirect enzyme activity, alter sub-cellular localisation or reprogramme signaling pathways.

Experimental section

Molecular biology

The gene encoding 1TBP-CTPR2 was synthesised by Integrated DNA Technologies. It was designed with appropriate restriction sites at each terminus to allow ligation and the generation of multivalent constructs by concatemerisation.[76] We used a variant of the original CTPR sequence developed by Regan and coworkers modified at two positions (D18Q and E19K)[63,64] to further enhance the stability of the CTPR scaffold. Our CTPR proteins do not have the extra C-terminal helix that has been used by some groups to improve solubility (Fig. 1B).[25,31,77,78] Likewise, the gene encoding the foldon motif was synthesized and ligated at the 3′ end of each nTBP-CTPR2n construct using the appropriate restriction enzymes (Thermo Fisher Scientific). The genes encoding hTNKS2 fragments were also amplified and ligated using the same method. Different combinations of tags (His, HA, HiBiT, eGFP, mCherry) were fused at the N- or C-terminal end of each construct. Genes were ligated into a modified pRSRET B vector (Thermo Fisher Scientific) for bacterial expression, or into the pcDNA3.1(−) vector (Thermo Fisher Scientific) for mammalian expression. Correct clones were confirmed by DNA sequencing (Eurofins Genomics). Protein sequences of all of the generated CTPR constructs are listed in Table S1.† An *E. coli* expression plasmid encoding TNKS2 ARC4 was previously generated.[55]

Protein purification

The pRSET B vectors encoding the His-tagged nTBP-CTPR2n constructs were transformed into chemically competent *Escherichia coli* C41 cells. Colonies were grown at 37 °C in 2xYT media (Formedium) containing ampicillin (50 μg mL^{-1}), shaking at 220 rpm until the optical density at 600 nm was ~0.6. Protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16–20 hours at 20 °C. Cells were harvested and resuspended in buffer A (50 mM Tris–HCl, 150 mM NaCl, pH 8.0) including EDTA-free SIGMAFAST protease inhibitors (Sigma-Aldrich), Lysozyme and DNase I. Cells were lysed by high-pressure homogenization using an Emulsiflex C5 homogenizer (Avestin) at 15 000 psi, and cell debris was removed by centrifugation steps at 40 000g for 40 minutes. nTBP-CTPR2n constructs were purified by immobilised metal ion affinity chromatography (IMAC) on a 5 mL HisTrap Excel column according to the manufacturer's instructions (Cytiva). The column was washed with 20 CV of buffer A containing 20 mM imidazole to prevent nonspecific interaction of lysate proteins to the beads. Proteins were eluted with buffer B (50 mM Tris–HCl, 150 mM NaCl, 500 mM imidazole, pH 8.0). All proteins were subsequently purified by size-exclusion using a HiLoad 16/600 Superdex 75 pg column (Cytiva) pre-equilibrated in buffer A. Purified protein was flash-frozen and stored at −80 °C until further use.

Proteins to be analysed by size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) were further purified by SEC using a Superdex 200 10/300 GL (Cytiva) pre-equilibrated in buffer A. Purified protein samples (100 μL) were subjected to SEC-MALS, performed on a SEC Superdex 200 increase 10/300 GL column (Cytiva) preequilibrated in buffer A, in line with a multi-angle light scattering module (DAWN-8+; Wyatt Technologies) and a differential refractometer (Optilab T-rEX; Wyatt Technologies). The light scattering and protein concentration at each point across the peaks in the chromatograph were used to determine the absolute molecular mass from the intercept of the Debye plot using Zimm's model as implemented in the ASTRA v7.3.0.11 software (Wyatt Technologies). To determine interdetector delay volumes, band-broadening constants and detector intensity normalization constants for the instrument, BSA (Thermo Fisher Scientific) was used as a standard prior-to sample measurement. Data were plotted with the program PRISM 8 (GraphPad Software Inc.).

His-tagged Tankyrase-2 Ankyrin Repeat Cluster 4 (residues 488-649, referred to as hTNKS2 ARC4) and Tankyrase-2 Ankyrin Repeat Clusters 1–3 (residues 2-485, referred to as hTNKS2 ARC1–3) were expressed and purified as reported above and previously,[55] with all buffers containing 2 mM DTT. To remove the His tag, proteins were incubated with 125 U of thrombin (Sigma-Aldrich) overnight at RT on a rotating mixer. Cleaved protein was further run over IMAC, and the flow-through was collected. The final purity and identity of all the proteins were determined by SDS-PAGE and MALDI mass spectrometry (PNAC, Department of Biochemistry,

Cambridge, UK), respectively. Concentrations were determined by absorbance at 280 nm using the calculated extinction coefficient (ExPaSy ProtParam)[79] for each protein.

Circular dichroism spectroscopy (CD) and thermal denaturation experiments

All protein samples were prepared in 10 mM sodium phosphate buffer, 150 mM NaCl, pH 7.4, at a final concentration of 20 μM. Circular dichroism (CD) measurements were performed on a Chirascan CD spectrophotometer (Applied Photophysics) at 20 °C. CD spectra were acquired at wavelengths between 200 nm and 280 nm using a 1 nm bandwidth at a scan speed of 120 nm per minute. Readings were repeated in triplicate and the data averaged. Thermal denaturation experiments were carried out by increasing the temperature of the protein samples from 20 °C to 92 °C in 1 °C steps, and the ellipticity at 222 nm was monitored. Readings were repeated five times, averaged and fitted to a sigmoidal Boltzmann equation including a sloping baseline term using PRISM (GraphPad Software Inc.). Subsequently, the sample was allowed to cool back to 20 °C and the CD spectrum was re-acquired.

Equilibrium denaturation monitored by fluorescence spectroscopy

High-throughput equilibrium denaturation experiments were performed as previously described.[74] All protein samples (at concentrations between 6–10 μM) were prepared in 50 mM sodium phosphate buffer, 150 mM NaCl, pH 6.8. Protein samples were dispensed into 96-well, half-area, black polystyrene plates (Corning) using a Microlab ML510B dispenser (Hamilton) containing different concentrations of guanidinium hydrochloride (GdmHCl) between 0 M and 6 M in increments of 0.1 M per well. Plates were covered with 96-well Microplate Aluminium Sealing Tape (Corning) to avoid evaporation and incubated at 25 °C for 1 hour. Plate measurements were carried out on a CLARIOstar Plate Reader (BMG LABTECH) with a tryptophan-detection filter set consisting of an excitation filter (275–285 nm), a dichroic PL325 nm, and an emission filter (350–370 nm) at 25 °C. Each measurement was performed in triplicate. The plots of fluorescence intensity *versus* denaturant concentration were fitted to a two-state model with sloping baselines, in which only fully folded and fully unfolded states are populated, using the following equation:[80]

$$\lambda_{obs} = \frac{\alpha_N + \beta_N[D] + (\alpha_D + \beta_D[D]) \cdot \exp\left[m_{D-N}([D] - [D]_{50\%})/RT\right]}{1 + \exp[m_{D-N}([D] - [D]_{50\%})]}$$

where a_N and a_D are the state signal at the lowest and highest concentration of denaturant, respectively, β_N and β_D are the slopes of the native and unfolded state baseline, respectively, m_{D-N} is a constant related to the change in solvent accessible surface area upon unfolding, $[D]$ is the denaturant concentration and $[D]_{50\%}$ is the midpoint of the unfolding transition. The free energy of unfolding, $\Delta G_{D-N}^{H_2O}$, was calculated as the product of the midpoint of unfolding ($[D]_{50\%}$) and the m-value (m_{D-N}), a constant proportional to the surface area exposed upon unfolding.

Isothermal titration calorimetry (ITC)

ITC was performed at 25 °C using a MicroCal VP-ITC (Malvern Panalytical). Proteins were dialysed into 10 mM sodium phosphate buffer, 150 mM NaCl, 0.5 mM TCEP, pH 7.4. TNKS2 ARC4 was titrated into the sample cell containing one of the nTBP-CTPR2n constructs. Injections of TNKS2 ARC4 into the cell were initiated with one injection of 5 μL over 6 seconds, followed by 29 injections of 10 μL over 12 seconds. Raw data were first subjected to baseline determination using NITPIC software[81] and were fitted using the OneSite model within Origin 7.0 software to a non-linear regression.

Immunoprecipitation (IP)

HEK293T cells were cultured in 15 mL cell growth media (Dulbecco's Modified Eagles Medium (DMEM), high glucose, pyruvate, (Thermo Fisher Scientific) supplemented with 10% Fetal Bovine Serum (Sigma-Aldrich) and 1× Penicillin/ Streptomycin (Gibco). HEK293T cells in 10 cm dishes were transfected with 3.5 μg of pcDNA3.1(−) vector encoding HiBiT-tagged hTNKS2 and 3.5 μg of pcDNA3.1(−) vector encoding a HA-tagged 3TBP-CTPR6, CTPR6 or empty vector. Transfection was performed using Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Following 48 hours from transfection, cells were washed twice in PBS and lysed for 30 minutes in 1 mL cold lysis buffer (50 mM Tris–HCl, 225 mM KCl, 1% NP-40, pH 7.5) including protease inhibitor cocktail (Sigma-Aldrich), 10 mM NaF, 1 mM Na$_3$VO$_4$, 1 mM PMSF. Cell debris was removed by centrifugation at 16 000g for 20 minutes. IP of HiBiT-tagged hTNKS2 bound to HA-tagged 3TBP-CTPR6 was performed by incubating the supernatant for 4 hours with 20 μL monoclonal anti-HA Agarose (Sigma-Aldrich), pre-equilibrated in lysis buffer. Anti-HA agarose resin was washed 4 times with 500 μL lysis buffer. Elution was performed by adding 20 μL 2× loading dye containing denaturant SDS to the settled resin. All steps following cell washing were performed at 4 °C. The presence of HiBiT-tagged hTNKS2 throughout the IP process was detected by mixing 5 μL of each sample with the Nano-Glo HiBiT Lytic Detection System according to the manufacturer's instructions (Promega). The amount of HiBiT-tagged hTNKS2 bound to the beads was quantified by mixing 5 μL of 50% anti-HA Agarose resin slurry with the Nano-Glo HiBiT Lytic Detection System before elution. Data were obtained from two biological replicates.

Samples for western blot were transferred from a 10% polyacrylamide gel to a nitrocellulose membrane (pore size 0.20 μm, Pharmacia Biotech). HiBiT-tagged hTNKS2 was visualised using the Nano-Glo HiBiT Blotting System according to the manufacturer's instructions (Promega). HA-tagged 3TBP-CTPR6 and CTPR6 were detected using anti-HA-Tag (C29F4, 1 : 1000 dilution) rabbit monoclonal antibody (cell signaling). Tubulin was identified using anti-alpha Tubulin antibody (ab7291, 1 : 10 000 dilution) mouse monoclonal antibody (Abcam). HRP-conjugated goat anti-mouse (P0447, 1 : 10 000 dilution) and swine anti-rabbit (P0399, 1 : 10 000 dilution) were used as secondary antibodies (Dako). The membrane was developed

using Amersham ECL western blotting Detection Reagent and ECL select western blotting Detection Reagent (Cytiva) on a LI-COR Odyssey Fc Imaging System.

Co-precipitation assay

20 μL of 10 μM hTNKS2 ARC1–3 was mixed with an equal volume of 1TBP-CTPR2, 3TBP-CTPR6, 3TBP-CTPR6-foldon or CTPR6 at the indicated concentration in co-precipitation buffer (50 mM Tris–HCl, 150 mM NaCl, 0.5 mM TCEP, pH 7.3). Samples were incubated at room temperature for 1 hour and centrifuged at 20 000g for 30 minutes at room temperature. The supernatant was collected and the pellet resuspended in an equal volume of buffer. Samples were run on a polyacrylamide gel, and the gels were imaged using the Li-COR Odyssey Fc Imaging System and protein band densities were analysed using the Image Studio Lite 5.2 software.

Fluorescence microscopy and FRAP

1×10^5 HEK293T cells were seeded overnight in 700 μL of cell growth media in a μ-Slide (Ibidi). The following day, cells were transfected with 400 ng of m-Cherry tagged nTBP-CTPR2n, nTBP-CTPR2n-foldon or the control CTPR construct in combination with 600 ng of eGFP-tagged TNKS2 ARC1–5 or 100 ng of TNKS2 ARC1, using 1.75 μL lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Following 48 hours from transfection, cells were imaged on a Leica TCS SP5 confocal microscope with a 100× oil-immersion objective lens (1.4 numerical aperture). Excitation and filters were as follows: eGFP, excitation at 488 nm, emission from 500–540 nm; mCherry, excitation at 543 nm, emission from 600–630 nm. For FRAP, individual circular regions of interest (ROI) were bleached using the 488 laser at 100% power for 5 seconds. Fluorescence intensity changes were recorded comparing 5 pre-bleaching frames with 60 post-bleaching frames (1.3 seconds per frame). Data were analysed using Leica LAS AF suite software and data were normalised.

Transmission electron microscopy (TEM)

hTNKS2 ARC1–3 was mixed with an equal volume and concentration (5 μM) of 1TBP-CTPR2, 3TBP-CTPR6, 3TBP-CTPR6-foldon, CTPR6 or co-precipitation buffer. Samples were incubated at room temperature for 1 hour. They were then dispensed on carbon support film, 400 mesh, 3 mm nickel grids (EM Resolutions Ltd, Saffron Walden, UK) and stained with 2% uranyl acetate (w/v). The samples were imaged on a FEI Tecnai G2 transmission electron microscope in the Cambridge Advanced Imaging Centre (CAIC, University of Cambridge, Cambridge, UK). Images were analysed using the SIS Megaview II Image Capture system.

TOPFLASH dual-luciferase reporter assay of activity of transfected CTPR constructs

Wnt pathway activity was induced by treating cells with conditioned media obtained from L-cells expressing Wnt3A (ATCC CRL-2647) for 8 days, according to ATCC guidelines. For the TOPFLASH assay, HEK293T cells in 24-well plates were transfected with 100 ng of TCF7L2-Firefly plasmid, 10 ng of CMV-Renilla plasmid and 100 ng of pcDNA3.1(−) vector encoding HiBiT-tagged nTBP-CTPR2n, nTBP-CTPR2n-foldon or the control CTPR constructs using 0.5 μL lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Transfected cells were allowed to recover in cell growth media for 8 hours, followed by treatment with Wnt-conditioned media (derived from L-cells expressing Wnt3A; ATCC CRL-2647) at a 1 : 1 ratio for a further 16 hours. The TOPFLASH assay was performed using the Dual-Luciferase Reporter Assay System (Promega), as previously described.[82] Relative luciferase values were obtained from three independent experiments by dividing the Firefly luciferase values by the Renilla luciferase values. The luminescence intensity of each HiBiT-tagged construct was measured by mixing 20 μL of the same cell lysate with the Nano-Glo HiBiT Lytic Detection System according to the manufacturer's instructions (Promega).

Measurements of cellular levels of HiBiT-tagged CTPR proteins

HEK293T cells in 96-well plates were transfected with 30 ng of pcDNA3.1(−) vector encoding HiBiT-tagged nTBP-CTPR2n, nTBP-CTPR2n-foldon or the control CTPR constructs using Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Following 19 hours incubation, cells were treated with 10 μM MG132 (Calbiochem) for 5 hours. The number of viable cells was measured using the CellTiter-Fluor Cell Viability Assay (Promega) according to the manufacturer's instructions. HiBiT-tagged constructs were quantified using the Nano-Glo HiBiT Lytic Detection System (Promega) as above. HiBiT values obtained from three independent replicates were normalised using the corresponding cell viability measurements.

Liposomal formulation of 3TBP-CTPR6 protein

Fusogenic Liposomes (FL) containing 3TBP-CTPR6 (FL-3TBP-CTPR6) were prepared using a previously reported method.[71] In a typical procedure, 250 μg of 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), 250 μg of 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), and 25 μg of 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindotricarbocyanine iodide (DiR) (Avanti Polar lipids) were dissolved in 250 μL of chloroform (Merck). The solution was dried overnight inside a vacuum desiccator and the resulting lipid film was hydrated with 125 μL of 25 μM 3TBP-CTPR6 (in 10 mM HEPES, pH 7.4). This dispersion was first vortexed for 2 minutes and then ultra-sonicated for 20 minutes at room temperature. Control FL were prepared similarly by hydrating the lipid film with 125 μL of 10 mM HEPES, pH 7.4. The liposomes were stored at 4 °C until further use. The zeta potential and hydrodynamic size of the FL were measured at 25 °C and pH 7.4 in 10 mM HEPES buffer using a Zetasizer Nano ZS (Malvern Instruments).

Intracellular delivery of liposome-encapsulated 3TBP-CTPR6 protein

To directly visualise intracellular protein delivery using confocal microscopy, 3TBP-CTPR6 was labelled with rhodamine B isothiocyanate (RITC) (Sigma). In a typical procedure, 50 μL of RITC (1 mg mL^{-1} in DMSO) was slowly added to a 1 mL solution of 3TBP-CTPR6 (2 mg mL^{-1}) in 0.1 M sodium carbonate buffer pH 9.0, with stirring. The solution was stirred at 4 °C for 8 hours and ammonium chloride (Sigma) was added to a final concentration of 50 mM and stirring was continued for a further 2 hours. RITC labelled 3TBP-CTPR6 (3TBP-CTPR6-RITC) was isolated from unreacted RITC using a PD10 desalting column (Cytiva). Liposomal formulation of 3TBP-CTPR6-RITC (FL-3TBP-CTPR6-RITC) was performed as described above. 1.4 × 10^5 HEK293T cells were seeded overnight in 700 μL of cell growth media in a μ-Slide (Ibidi). The following day, medium was replaced with 700 μL cell growth media without FBS and containing 28 μL of FL-3TBP-CTPR6-RITC. Samples were incubated for 15 minutes at 37 °C and then washed twice with 1× PBS. Confocal images were acquired using a Leica TCS SP5 confocal microscope.

Cell viability assay

4 × 10^4 HEK293T cells were seeded for 24 hours in 100 μL of cell growth media in a 96-well plate. Cells were incubated with 100 μL liposomes in cell growth media without FBS and containing different volumes (1–12 μL) of FL and FL-3TBP-CTPR6 for 15 minutes at 37 °C. After washing twice with 1× PBS, 100 μL of CellTiter-Glo Reagent (Promega) was added, and luminescence was measured using a Clariostar microplate reader (BMG LAB-TECH) according to the manufacture's protocol. Untreated cells were used as control. Data were obtained from triplicate samples, and the standard deviation was calculated from two independent experiments.

TOPFLASH assay of cellular activity of liposome-encapsulated 3TBP-CTPR6 protein and small molecule hTNKS inhibitors

HEK293T cells were seeded in 24-well plates as described above. Cells were transfected with 100 ng of TCF7L2-Firefly plasmid, 10 ng of CMV-Renilla plasmid per well using Lipofectamine 2000 transfection reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Transfected cells were allowed to recover in cell growth media for 8 hours, and treated with Wnt3A-conditioned media, as above. 16 hours post Wnt pathway activation, proteins were delivered into the cells by liposomal treatment. Cells were incubated with cell growth media minus FBS containing 20 μL of liposomes, for 15 minutes at 37 °C. Following one wash in 1× PBS, Wnt3A conditioned media was added and cells were incubated for 6 hours. The TOPFLASH assay was performed, as described above, in triplicate (from two independent experiments).

For the dose-dependence analysis, the volume of liposome added was kept constant (20 μL) and the protein concentration was varied. The FL-3TBP-CTPR6 samples were prepared in the following way: lipid cakes were hydrated with 10 mM HEPES pH

7.4 and 3.125 μM, 6.25 μM, 12.5 μM, and 25 μM 3TBP-CTPR6. 20 μL of these liposomes in 500 μL cell growth media resulted in 3TBP-CTPR6 concentrations of 0.125 μM, 0.25 μM, 0.5 μM and 1 μM respectively. Samples of the liposome-encapsulated unfunctionalised control protein, referred to as FL-CTPR6, were prepared by hydrating lipid cake with 25 μM CTPR6 in 10 mM HEPES pH 7.4.

For testing the small molecule hTNKS inhibitors XAV939, IWR-1 and G007 in the TOPFLASH assay, cells were treated prophylactically with inhibitors mixed with Wnt-conditioned media and incubated for 16 hours, or interventionally with inhibitors added after overnight treatment with Wnt-conditioned media and incubated for another 6 hours. All small molecule hTNKS inhibitors were used at a final concentration of 1 μM in 0.5% DMSO. The TOPFLASH assay was then completed as described above.

Conflicts of interest

Authors APR, LSI and PJER are inventors on patent application PCT/EP2018/068580. Authors APR, LSI and PJER are founders of PolyProx Therapeutics Limited (company number 11664980).

Acknowledgements

References

1 K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson and N. E. Davey, *Chem. Rev.*, 2014, **114**, 6733–6778.

2 M. Pelay-Gimeno, A. Glas, O. Koch and T. N. Grossmann, *Angew. Chem., Int. Ed.*, 2015, **54**, 8896–8927.

3 H. Y. J. Fung, M. Birol and E. Rhoades, *Curr. Opin. Struct. Biol.*, 2018, **49**, 36–43.

4 J. Weng and W. Wang, *Curr. Opin. Struct. Biol.*, 2020, **62**, 9–13.

5 A. M. Cuesta, N. Sainz-Pastor, J. Bonet, B. Oliva and L. A. Alvarez-Vallina, *Trends Biotechnol.*, 2010, **28**, 355–362.

6 A. F. Labrijn, M. L. Janmaat, J. M. Reichert and P. W. H. I. Parren, *Nat. Rev. Drug Discovery*, 2019, **18**, 585–608.

7 Y. L. Boersma, G. Chao, D. Steiner, K. Dane Wittrup and A. Plückthun, *J. Biol. Chem.*, 2011, **286**, 41273–41285.

8 T. Z. Bass, M. Rosestedt, B. Mitran, F. Y. Frejd, J. Löfblom, V. Tolmachev, S. Ståhl and A. Orlova, *Sci. Rep.*, 2017, 7, 43118.

9 F. Fleetwood, S. Klint, M. Hanze, E. Gunneriusson, F. Y. Frejd, S. Ståhl and J. Löfblom, *Sci. Rep.*, 2014, **4**, 7518.

10 S. Hober, S. Lindbo and J. Nilvebrant, *Methods*, 2019, **154**, 143–152.

11 A. Groß, C. Hashimoto, H. Sticht and J. Eichler, *Front. Bioeng. Biotechnol.*, 2016, **3**, 211.

12 A. Dirksen, E. W. Meijer, W. Adriaens and T. M. Hackeng, *Chem. Commun.*, 2006, 1667–1669.

13 E. A. Englund, D. Wang, H. Fujigaki, H. Sakai, C. M. Micklitsch, R. Ghirlando, G. Martin-Manso, M. L. Pendrak, D. D. Roberts, S. R. Durell and D. H. Appella, *Nat. Commun.*, 2012, **3**, 614.

14 N. Stephanopoulos, *Bioconjugate Chem.*, 2019, **30**, 1915–1922.

15 C. Tiede, A. A. S. Tang, S. E. Deacon, U. Mandal, J. E. Nettleship, R. L. Owen, S. E. George, D. J. Harrison, R. J. Owens, D. C. Tomlinson and M. J. McPherson, *Protein Eng., Des. Sel.*, 2014, **27**, 145–155.

16 S. Reverdatto, D. S. Burz and A. Shekhtman, *Curr. Top. Med. Chem.*, 2015, **15**, 1082–1101.

17 D. Schlatter, S. Brack, D. W. Banner, S. Batey, J. Benz, J. Bertschinger, W. Huber, C. Joseph, A. C. Rufer, A. Van Der Klooster, M. Weber, D. Grabulovski and M. Hennig, *mAbs*, 2012, **4**, 497–508.

18 M. H. Sonntag, J. Schill and L. Brunsveld, *ChemBioChem*, 2017, **18**, 441–443.

19 S. Pacheco, G. Béhar, M. Maillasson, B. Mouratou, F. Pecorari and J. Osbourn, *Protein Eng., Des. Sel.*, 2014, **27**, 431–438.

20 F. S. Nielsen, J. Sauer, J. Bäcklund, B. Voldborg, K. Gregorius, S. Mouritsen and T. Bratt, *J. Biol. Chem.*, 2004, **279**, 33593–33600.

21 M. Rossmann, S. J. Greive, T. Moschetti, M. Dinan and M. Hyvönen, *Protein Eng., Des. Sel.*, 2017, **30**, 419–430.

22 R. Tamaskovic, M. Simon, N. Stefan, M. Schwill and A. Plückthun, in *Methods in Enzymology*, Academic Press Inc., 2012, vol. 503, pp. 101–134.

23 A. L. Cortajarena, F. Yi and L. Regan, *ACS Chem. Biol.*, 2008, **3**, 161–166.

24 T. Z. Grove, M. Hands and L. Regan, *Protein Eng., Des. Sel.*, 2010, **23**, 449.

25 J. J. Phillips, C. Millership and E. R. G. Main, *Angew. Chem., Int. Ed.*, 2012, **51**, 13132–13135.

26 D. Sanchez-deAlcazar, S. H. Mejias, K. Erazo, B. Sot and A. L. Cortajarena, *J. Struct. Biol.*, 2018, **201**, 118–129.

27 S. H. Mejias, A. Aires, P. Couleaud and A. L. Cortajarena, *Adv. Exp. Med. Biol.*, 2016, **940**, 61–81.

28 R. S. Sikorski, M. S. Boguski, M. Goebl and P. Hieter, *Cell*, 1990, **60**, 307–317.

29 T. Hirano, N. Kinoshita, K. Morikawa and M. Yanagida, *Cell*, 1990, **60**, 319–328.

30 A. K. Das, P. W. Cohen and D. Barford, *The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions*, 1998, vol. 17.

31 E. R. G. Main, Y. Xiong, M. J. Cocco, L. D'Andrea and L. Regan, *Structure*, 2003, **11**, 497–508.

32 A. Perez-Riba, A. R. Lowe, E. R. G. Main and L. S. Itzhaki, *Biophys. J.*, 2018, **114**, 2552–2562.

33 S. K. Madden, A. Perez-Riba and L. S. Itzhaki, *Protein Sci.*, 2019, **28**, 738–745.

34 S. Smith, I. Giriat, A. Schmitt and T. De Lange, *Science*, 1998, **282**, 1484–1487.

35 S. J. Hsiao and S. Smith, *Biochimie*, 2008, **90**, 83–92.

36 S. Guettler, J. Larose, E. Petsalaki, G. Gish, A. Scotter, T. Pawson, R. Rottapel and F. Sicheri, *Cell*, 2011, **147**, 1340–1354.

37 T. Eisemann, M. McCauley, M. F. Langelier, K. Gupta, S. Roy, G. D. Van Duyne and J. M. Pascal, *Structure*, 2016, **24**, 1679–1692.

38 H. Seimiya and S. Smith, *J. Biol. Chem.*, 2002, **277**, 14116–14126.

39 H. Seimiya, Y. Muramatsu, S. Smith and T. Tsuruo, *Mol. Cell. Biol.*, 2004, **24**, 1944–1955.

40 B. D. Cook, J. N. Dynek, W. Chang, G. Shostak and S. Smith, *Mol. Cell. Biol.*, 2002, **22**, 332–342.

41 S. Smith and T. De Lange, *Curr. Biol.*, 2000, **10**, 1299–1302.

42 J. N. Dynek and S. Smith, *Science*, 2004, **304**, 97–100.

43 N. W. Chi and H. F. Lodish, *J. Biol. Chem.*, 2000, **275**, 38437–38444.

44 S. M. A. Huang, Y. M. Mishina, S. Liu, A. Cheung, F. Stegmeier, G. A. Michaud, O. Charlat, E. Wiellette, Y. Zhang, S. Wiessner, M. Hild, X. Shi, C. J. Wilson, C. Mickanin, V. Myer, A. Fazal, R. Tomlinson, F. Serluca, W. Shao, H. Cheng, M. Shultz, C. Rau, M. Schirle, J. Schlegl, S. Ghidelli, S. Fawell, C. Lu, D. Curtis, M. W. Kirschner, C. Lengauer, P. M. Finan, J. A. Tallarico, T. Bouwmeester, J. A. Porter, A. Bauer and F. Cong, *Nature*, 2009, **461**, 614–620.

45 Y. Zhang, S. Liu, C. Mickanin, Y. Feng, O. Charlat, G. A. Michaud, M. Schirle, X. Shi, M. Hild, A. Bauer, V. E. Myer, P. M. Finan, J. A. Porter, S. M. A. Huang and F. Cong, *Nat. Cell Biol.*, 2011, **13**, 623–629.

46 L. Mariotti, K. Pollock and S. Guettler, *Br. J. Pharmacol.*, 2017, **174**, 4611–4636.

47 T. Lau, E. Chan, M. Callow, J. Waaler, J. Boggs, R. A. Blake, S. Magnuson, A. Sambrone, M. Schutten, R. Firestein, O. Machon, V. Korinek, E. Choo, D. Diaz, M. Merchant, P. Polakis, D. D. Holsworth, S. Krauss and M. Costa, *Cancer Res.*, 2013, **73**, 3132–3144.

48 R. G. James, K. C. Davidson, K. A. Bosch, T. L. Biechele, N. C. Robin, R. J. Taylor, M. B. Major, N. D. Camp, K. Fowler, T. J. Martins and R. T. Moon, *PLoS One*, 2012, **7**, e50457.

49 J. W. Johannes, L. Almeida, B. Barlaam, P. A. Boriack-Sjodin, R. Casella, R. A. Croft, A. P. Dishington, L. Gingipalli, C. Gu, J. L. Hawkins, J. L. Holmes, T. Howard, J. Huang, S. Ioannidis, S. Kazmirski, M. L. Lamb, T. M. McGuire, J. E. Moore, D. Ogg, A. Patel, K. G. Pike, T. Pontz, G. R. Robb, N. Su, H. Wang, X. Wu, H. J. Zhang, Y. Zhang, X. Zheng and T. Wang, *ACS Med. Chem. Lett.*, 2015, **6**, 254–259.

50 P. O. Hassa and M. O. Hottiger, *Bioscience*, 2008, **13**, 3046–3082.

51 M. Rouleau, A. Patel, M. J. Hendzel, S. H. Kaufmann and G. G. Poirier, *Nat. Rev. Cancer*, 2010, **10**, 293–301.

52 J. L. Riffell, C. J. Lord and A. Ashworth, *Nat. Rev. Drug Discovery*, 2012, **11**, 923–936.

53 L. Mariotti, C. M. Templeton, M. Ranes, P. Paracuellos, N. Cronin, F. Beuron, E. Morris and S. Guettler, *Mol. Cell*, 2016, **63**, 498–513.

54 A. A. Riccio, M. McCauley, M. F. Langelier and J. M. Pascal, *Structure*, 2016, **24**, 1573–1581.

55 W. Xu, Y. H. Lau, G. Fischer, Y. S. Tan, A. Chattopadhyay, M. De La Roche, M. Hyvönen, C. Verma, D. R. Spring and L. S. Itzhaki, *J. Am. Chem. Soc.*, 2017, **139**, 2245–2256.

56 S. Morrone, Z. Cheng, R. T. Moon, F. Cong and W. Xu, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 1500–1505.

57 Y. Tao, S. V Strelkov, V. V Mesyanzhinov and M. G. Rossmann, *Structure*, 1997, **5**, 789–798.

58 J. Gómez-González, D. G. Peña, G. Barka, G. Sciortino, J. D. Maréchal, M. V. López and M. E. Vázquez, *Front. Chem.*, 2018, **6**, 520.

59 N. Kobayashi, K. Yanase, T. Sato, S. Unzai, M. H. Hecht and R. Arai, *J. Am. Chem. Soc.*, 2015, **137**, 11285–11293.

60 D. J. Glover, L. Giger, S. S. Kim, R. R. Naik and D. S. Clark, *Nat. Commun.*, 2016, **7**, 11771.

61 J. S. McLellan, M. Chen, M. G. Joyce, M. Sastry, G. B. E. Stewart-Jones, Y. Yang, B. Zhang, L. Chen, S. Srivatsan, A. Zheng, T. Zhou, K. W. Graepel, A. Kumar, S. Moin, J. C. Boyington, G. Y. Chuang, C. Soto, U. Baxa, A. Q. Bakker, H. Spits, T. Beaumont, Z. Zheng, N. Xia, S. Y. Ko, J. P. Todd, S. Rao, B. S. Graham and P. D. Kwong, *Science*, 2013, **342**, 592–598.

62 T. Kajander, A. L. Cortajarena, S. Mochrie and L. Regan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2007, **63**, 800–811.

63 A. L. Cortajarena, S. G. J. Mochrie and L. Regan, *Protein Sci.*, 2011, **20**, 1042–1047.

64 A. L. Cortajarena, T. Kajander, W. Pan, M. J. Cocco and L. Regan, *Protein Eng., Des. Sel.*, 2004, **17**, 399–409.

65 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

66 D. C. Ranawakage, T. Takada and Y. Kamachi, *Sci. Rep.*, 2019, **9**, 6895.

67 M. K. Schwinn, T. Machleidt, K. Zimmerman, C. T. Eggers, A. S. Dixon, R. Hurst, M. P. Hall, L. P. Encell, B. F. Binkowski and K. V Wood, *ACS Chem. Biol.*, 2018, **13**, 467–474.

68 S. P. Boudko, S. V. Strelkov, J. Engel and J. Stetefeld, *J. Mol. Biol.*, 2004, **339**, 927–935.

69 P. Li, S. Banjade, H. C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q. X. Jiang, B. T. Nixon and M. K. Rosen, *Nature*, 2012, **483**, 336–340.

70 M. Ray, Y. W. Lee, F. Scaletti, R. Yu and V. M. Rotello, *Nanomedicine*, 2017, **12**, 941–952.

71 S. Kube, N. Hersch, E. Naumovska, T. Gensch, J. Hendriks, A. Franzen, L. Landvogt, J. P. Siebrasse, U. Kubitscheck, B. Hoffmann, R. Merkel and A. Csiszár, *Langmuir*, 2017, **33**, 1051–1059.

72 Z. Gu, A. Biswas, M. Zhao and Y. Tang, *Chem. Soc. Rev.*, 2011, **40**, 3638–3655.

73 M. De La Roche, A. E. K. Ibrahim, J. Mieszczanek and M. Bienz, *Cancer Res.*, 2014, **74**, 1495–1505.

74 A. Perez-Riba and L. S. Itzhaki, *Sci. Rep.*, 2017, 7, 9071.

75 P. Tompa, N. E. Davey, T. J. Gibson and M. M. Babu, *Mol. Cell*, 2014, **55**, 161–169.

76 T. Kajander, A. L. Cortajarena, E. R. G. Main, S. G. J. Mochrie and L. Regan, *J. Am. Chem. Soc.*, 2005, **127**, 10188–10190.

77 Y. Javadi and E. R. G. Main, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 17383–17388.

78 C. Millership, J. J. Phillips and E. R. G. Main, *J. Mol. Biol.*, 2016, **428**, 1804–1817.

79 E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch, in *The Proteomics Protocols Handbook*, Humana Press, Totowa, NJ, 2005, pp. 571–607.

80 S. E. Jackson and A. R. Fersht, *Biochemistry*, 1991, **30**, 10428–10435.

81 T. H. Scheuermann and C. A. Brautigam, *Methods*, 2015, **76**, 87–98.

82 V. Korinek, N. Barker, P. J. Morin, D. van Wichen, R. de Weger, K. W. Kinzler, B. Vogelstein and H. Clevers, *Science*, 1997, **275**, 1784–1787.