



The University of
Nottingham

Inhibiting Protein-Protein Interactions in Telomeres
as an Approach to Cancer Chemotherapy

Twana Salih

BSc. Pharmacy, MSc. Pharmaceutical Chemistry

Thesis submitted to the
University of Nottingham for the degree of
Doctor of Philosophy

December 2015

Abstract

Stable telomeres play a key role to the survival of cancer cells; therefore, different cancer chemotherapeutic approaches have been developed in order to disrupt or destabilise telomeres or telomerase. One of the newest methods is the disruption of vital protein–protein interactions in the telomere, such as that between shelterin components TRF1 and TIN2. The principal aim of this project was to obtain a novel peptide-like molecule, an analogue of a key interacting region of TIN2 that could compete effectively for the binding sites on TRF1 and so lead to the destabilisation of telomere structure. Molecular modelling and simulations were undertaken as the starting point of the project. Structure-based drug design was applied, starting from the available crystal structure data. A library of peptide analogues of the TRF1-binding motif in TIN2 was designed using the MM-GBSA simulation method to predict binding affinities. Then, a number of the peptide analogues were selected from the library for further investigations.

The secondary goal was to investigate the accuracy of the predicted $\Delta G_{\text{binding}}$ values and try to optimise them; the latter aim was set out after finding a significant difference in the predicted binding free energy values after repeating the identical protocol for the same complex system. Therefore, different approaches were applied to optimise the predicted $\Delta G_{\text{binding}}$ values. Subsequently, selected TIN2 peptide analogues were synthesised in the laboratory using Fmoc solid-phase peptide synthesis. Then, the hTRF1 protein

was expressed and purified in preparation for the development of the *in vitro* assay. Finally, biophysical evaluations and screening of the peptide analogues were performed using fluorescence polarisation assay.

One of the peptide analogues developed in this study was identified as an early lead compound. In addition, the findings of this research showed that the $\Delta G_{\text{binding}}$ values of the peptide analogues have significantly improved accuracy after optimisation. As a result of these investigations, suggestions were identified for future research.

Acknowledgments

Firstly, I would like to express my heartfelt gratitude to Dr Charles Laughton, for providing me the opportunity to work within his group, also for his extremely supportive nature and continuous enthusiasm throughout the PhD process. I would also like to thank Dr Weng Chan for his teaching and guidance throughout my PhD. Thanks must be extended to Dr Lodewijk Dekker for providing me the opportunity to work in his laboratory and his frequent and helpful discussions relating to the field of biophysical assays

I wish to express my warmest thanks to all of my colleagues in the Molecular Recognition group and C22 laboratory, in particular Dr Sivaneswary Genapathy, Dr Fabio Rui and Mohammed Khaled Tumbi for their frequent support and sharing their knowledge and expertise. I would also like to give a special thanks to Prof. Panos Soultanas and Dr Olivier Rannou who assisted and supervised my protein expression. Furthermore, I wish to add my thanks to Lee Hibbett and the rest of support staff in the Centre of Biomolecular Sciences, who helped me in numerous ways

Finally, I wish to express my deepest thanks and sincere gratitude to my family, especially my wife Rozha, as her unwavering encouragement and absolute faith in my abilities has inspired me to realise my dreams.

Contents

Abstract.....	I
Acknowledgments.....	III
Abbreviations.....	XII
1. Introduction.....	1
1.1 Telomeres and Telomerase Enzyme.....	1
1.2 History of Telomere Discovery	3
1.3 Telomeric Proteins.....	4
1.3.1 TRF1 and TRF2 Proteins.....	6
1.3.2 TIN2 Protein	9
1.3.3 RAP1 Protein	9
1.3.4 TPP1 Protein.....	10
1.3.5 POT1 Protein	10
1.4 Tumourigenesis and Telomeres.....	11
1.5 Telomere Disruption and Destabilisation for Cancer Treatment	13
1.5.1 Targeting Telomerase Enzyme	13
1.5.2 Targeting Telomeric DNA (G-quadruplexes).....	14
1.5.3 Targeting Shelterin Proteins	16
1.6 Drug Discovery Process.....	19
1.7 Lead Compound Identification.....	21
1.7.1 Phenotypic Screening	22
1.7.2 High Throughput Screening.....	22
1.7.3 Fragment-Based Drug Discovery	23
1.7.4 Computer-Aided Drug Discovery.....	24

1.8	Molecular Recognition	28
1.8.1	Hydrophobic Contacts	28
1.8.2	Van der Waals Interactions.....	30
1.8.3	Hydrogen Bonds	31
1.8.4	π effects.....	31
1.8.5	Salt Bridges.....	32
1.9	Therapeutic Molecule Sizes.....	33
1.9.1	Small Molecules	33
1.9.2	Protein-Based Drugs	34
1.9.3	Peptide-Based Drugs.....	34
1.10	Inhibiting PPIs as a Therapeutic Approach	36
1.11	Aims and Objectives.....	39
2.	The Application of Molecular Modelling to Ligand Design.....	41
2.1	Introduction.....	41
2.2	Molecular Dynamic Simulation.....	42
2.3	<i>In silico</i> Calculations of Binding Free Energy.....	44
2.4	MM-PBSA/GBSA Method	45
2.5	The Balance Between Enthalpy and Entropy	50
2.6	Molecular Recognition in the TRF1–TIN2 and TRF2–Apollo Crystal Structures	51
2.7	<i>In silico</i> Design of TIN2 Peptide Inhibitors	57
2.7.1	TIN2–Apollo Cross-Matching Mutations.....	57
2.7.2	Mutation of Selected TIN2 Residues Based on the Crystal Structure of TRF1–TIN2	60
2.8	MD Simulations of the Protein–Peptide Complexes.....	63
2.9	$\Delta G_{\text{binding}}$ Prediction of TRF1–TIN2 Peptide Analogues.....	63
2.10	Replicating MD Simulations.....	69

2.11	Optimising the Accuracy of Predicted Results	72
2.11.1	Optimising the Replicate Number of MD Simulations	73
2.11.2	Optimising MD Simulation Time	79
2.12	Analysis of Key Interactions.....	84
2.12.1	TS01–TRF1 Binding Interactions	86
2.12.2	TS02–TRF1 Binding Interactions	88
2.12.3	TS03–TRF1 Binding Interactions	90
2.12.4	TS04–TRF1 Binding Interactions	92
2.12.5	TS05–TRF1 Binding Interactions	94
2.12.6	TS06–TRF1 Binding Interactions	95
2.12.7	TS07–TRF1 Binding Interactions	96
2.12.8	TS08–TRF1 Binding Interactions	98
2.12.9	TS09–TRF1 Binding Interactions	99
2.12.10	TS10–TRF1 Binding Interactions	100
2.12.11	TS11–TRF1 Binding Interactions	101
2.12.12	TS12–TRF1 Binding Interactions	103
2.12.13	TS13–TRF1 Binding Interactions	105
2.13	Fluorescent Probe Design	106
3.	Total Chemical Synthesis of the Peptide Ligands.....	108
3.1	Introduction.....	108
3.2	Peptide Synthesis.....	109
3.3	Solid-Phase Peptide Synthesis.....	110
3.3.1	Resin Materials and Linker Types for SPPS	114
3.3.2	Coupling Reagents in SPPS.....	116
3.4	Synthetic Strategy of the TIN2 Peptide Analogues.....	122
3.4.1	Condensation of Fmoc-Amino Acids with Rink Amide Novagel	124
3.4.2	Peptide Assembly	125

3.4.3	Cleavage of the Linear Peptides	126
3.5	Purification and Analyses of the Linear Peptides.....	127
4.	Biophysical Evaluations for the Protein–Peptide Complexes.....	135
4.1	Introduction.....	135
4.2	Introduction to the FP Assay	136
4.3	Principles of the FP Assay	139
4.4	Characteristics of the Fluorophore Molecule.....	141
4.5	Advantages and Disadvantages of the FP Assay	141
4.6	FP Assay of the TRF1–TIN2 Peptide Analogues	142
4.6.1	Fluorophore Selection for the FP Assay	143
4.6.2	Selecting the Fluorophore Tagging Site	144
4.6.3	Detection Sensitivity of the Instrument for the Labelled TIN2 Peptide	144
4.6.4	Determining the TRF1 Concentration	146
4.6.5	Competitive Binding of TS01 and TS01-L with the TRF1	148
4.6.6	Competitive Binding Assay of the Labelled TIN2-TRF1 with the Peptide Analogues	149
5.	Conclusions and Future Work.....	154
6.	Experimental	171
6.1	Molecular Modeling.....	171
6.1.1	Systems Preparation.....	171
6.1.2	Systems Minimisation.....	171
6.1.3	Molecular Dynamic Simulations	172
6.1.4	Trajectory Analysis Techniques	173
6.1.5	Visualisation Software	174
6.2	Peptide Synthesis.....	175
6.2.1	Materials	175
6.2.2	Instrumentation	175

6.2.3	Procedure of the Peptides Assembly	176
6.2.4	Produced Peptides.....	178
6.3	Fluorescence Polarisation Assay.....	185
6.3.1	Materials and Instrumentation	185
6.3.2	Detection Sensitivity of the Instrument for TS01-L	185
6.3.3	Determining the Concentration of TRF1 for the Assays	186
6.3.4	Competitive Binding Assay	187
7.	References.....	189
	Appendix	206
	A1 TRF1 Expression and Purification.....	206
	A1.1 TRF1 Expression.....	206
	A1.2 TRF1 Protein Purification	208
	A1.2.1 Affinity Chromatography	208
	A1.2.2 Size Exclusion Chromatography	209
	A1.3 Dialysis	210
	A1.4 Calculating hTRF1 Concentration	210
	A2. Amino acid Abbreviations (IUPAC)	212
	A3. Principal Component Analysis of the TIN2 peptide analogues	213

List of Figures

Figure 1.1 Telomeres location and structure in the cells.....	2
Figure 1.2 Schematic representation of the shelterin proteins in the telomere.....	7
Figure 1.3 TRF1 and TRF2 binding proteins.. ..	8
Figure 1.4 Consequences of the telomere attrition in the eukaryotic cells.....	12
Figure 1.5 Structures of G-tetrad and G-quadruplexes.....	16
Figure 1.6 Different approaches for inhibiting telomeres.....	17
Figure 1.7 Main steps of drug discovery and development process.....	20
Figure 1.8 The role of CADD in the drug discovery processes.....	25
Figure 1.9 Structure-based drug discovery process to find a lead compound.....	27
Figure 1.10 Stacking conformations between aromatic groups.....	32
Figure 2.1 A scheme representing the process of molecular dynamic simulations.....	43
Figure 2.2 Calculating binding free energy of the receptor–ligand complex.....	48
Figure 2.3 Structure of TRF1 and TRF2 dimerisation domains and TRF1 helices.....	53
Figure 2.4 The TRF1–TIN2 crystal structure.....	54
Figure 2.5 Interactions between TRF1 _{TRFH} protein and TIN2 _{TBM} peptide.....	55
Figure 2.6 A scheme denoting TRF2 _{TRFH} -Apollo _{TBM} interactions.....	57
Figure 2.7 Overlapping 3D structures of the TIN2 and Apollo peptides.....	58
Figure 2.8 The shared sequence of Apollo peptide and TIN2 peptide.....	59
Figure 2.9 The interactions of TIN2 _{TBM} peptide with TRF1 _{TRFH} protein.....	62
Figure 2.10 Per-residue decomposition of the TRF1–TIN2 complex system,.....	65
Figure 2.11 Histograms showing the distribution frequency of $\Delta G_{\text{binding}}$	77
Figure 2.12 Distribution of replicate clusters for each TRF1-peptide analogue.....	79
Figure 2.13 Mean $\Delta G_{\text{binding}}$ values calculated across 50 replicates of TS01–TRF1.....	80
Figure 2.15 Average RMSDs of 50 replicates for each snapshot of the complex systems.....	84
Figure 2.17 H-bond interaction between H257 and Q127 in TS01–TRF1 complex.....	87
Figure 2.18 Overlapping TRF1–TIN2 crystal structure and equilibrated MD snapshot.....	88
Figure 2.19 Binding interactions of Y258 in TS02 peptide with TRF1 protein.....	89

Figure 2.20 Overlapping the entire structures of TRF1–TS01 and TRF1–TS02	90
Figure 2.21 N259 of TS01 and L259 of TS03 interactions with TRF1.....	91
Figure 2.22 Comparing the binding interactions of T261–E106 with A261–E106.....	93
Figure 2.23 Overlay of TS01–TRF1 and TS05–TRF1 complexes	94
Figure 2.24 Binding interactions of L260 in TS01 and G260 in TS06, with TRF1 protein.	95
Figure 2.25 Comparison of the binding interactions of TS07 and TS01 with TRF1	97
Figure 2.26 Interactions between the TS08 mutant residues with the TRF1 protein.	98
Figure 2.27 Binding interactions between the mutated residues of TS09 and TRF1 _{TRFH} .	100
Figure 2.28 Binding interactions between TS10 residues and TRF1 _{TRFH}	101
Figure 2.29 Binding interactions between R261 of TS11 and TRF1 _{TRFH}	102
Figure 2.30 Overlapping the crystal structure of TRF1–TIN2 and TRF1–TS11.....	103
Figure 2.31 Binding interactions between TS12 analogue and TRF1 protein.....	104
Figure 2.33 Fluorescein molecule conformation during MD simulations	107
Figure 3.1 Structures of DCC and DIC activating reagents.	116
Figure 3.2 The structures of a) Oxyma and b) Pyoxim activating reagents.....	119
Figure 3.3 2D 1H-1H COSY NMR spectrum of the TS01 peptide.	130
Figure 3.4 1H-1H COSY NMR spectrum of the TS11 peptide analogue.	132
Figure 4.1 The difference of emitted light during the FP assay.....	140
Figure 4.2 The polarisation signals (mP) for the different concentrations of TS01-L.	145
Figure 4.3 Direct binding of TS01-L with different concentrations of TRF1 protein..	147
Figure 4.4 Competitive binding assay between the labelled and unlabelled TIN2.....	149
Figure 4.5 Competitive binding assay of TS01 and the 12 peptide analogues	150
Figure 4.6 Competitive binding assay of the peptides TS01, TS02 and TS11 with the mixture of TS01-L and TRF1 protein.	153
Figure 5.1 Correlation between $\Delta G_{\text{binding}}$ results and experimental results.....	162
Figure 5.2 Conformations of the P262 and F141 in the TRF1–TIN2 crystal structure.....	166
Figure 5.3 The TRF1–TIN2 crystal structure showing the conformations and the distance between the side-chains of S256 and L263.....	168

List of Tables

Table 1-1 The full name, main binding site, function and dysfunction of shelterin and shelterin-associated proteins that regulate telomeric length.....	5
Table 2-1 Predicted binding free energies of TIN2 peptide and four analogues derived from cross mutation with the Apollo peptide.	64
Table 2-2 The $\Delta G_{\text{binding}}$ values of 49 TIN2 peptide analogues.	66
Table 2-3 Binding free energies of the single MD simulation and average of 50 replicates MD simulations.	71
Table 2-4 Standard deviation and calculated sample size for the complex systems with different margins of error.	74
Table 3-1 Commonly used linkers that produce peptides.....	115
Table 3-2 Summary of the yield, RP-HPLC retention time (t_R), crude purity and both calculated and measured MS for the peptide analogues.....	128
Table 3-3 Proton chemical shifts for the amino acids of the TS01 peptide.....	131
Table 3-4 Proton chemical shifts for each amino acid of the TS11 peptide.	133
Table 4-1 Results of IC_{50} for the TIN2 analogues obtained from the FP assays.	152
Table 5-1 Predicted and experimental affinity of the TIN2 peptide analogues.	160

Abbreviations

5-FAM	5-Carboxyfluorescein
ALT	Alternative lengthening of telomere
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia related
Boc	Butoxycarbonyl
BRCT	BRCA 1 C terminus
C-terminal	Carboxyl terminal
CADD	Computer-aided drug design
COSY	Correlation spectroscopy
D-loop	Displacement loop
Da	Dalton

DCC	1,3-Dicyclohexylcarbodiimide
DCM	Dichloromethane
DDR	DNA Damage response
DIC	1,3-Diisopropylcarbodiimide
DIPEA	<i>N,N</i> -Diisopropylethylamine
DMF	<i>N,N</i> -Dimethylformamide
DMSO	Dimethylsulfonate
DNA	Deoxyribonucleic acid
Ds	Double strand
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FBDD	Fragment-based drug discovery
FEP	Free-energy perturbation

FF	Force field
Fmoc	Fluorenyl-9-methoxycarbonyl
FP	Fluorescence polarisation
fs	Femtosecond
GBSA	Generalized-Born surface area
G-rich	Guanine-rich
G4	G-Quadruplex
H-bond	Hydrogen bond
His ₆ -tag	Affinity tag of six histidine residues
HOAt	1-Hydroxy-7-azabenzotriazole
HOBt	1-Hydroxybenzotriazole
hTRF1	Human telomere repeat binding factor 1

HTS	High throughput screening
IC ₅₀	Inhibitory concentration 50%
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kb	Kilobase
Kcal	Kilocalories
Kd	Dissociation constant
kDa	Kilodalton
LB	Luria broth
LBDD	Ligand-based drug design
LIE	Linear interaction energy
M	Molar

MC	Monte Carlo
MD	Molecular dynamics
Min	Minute
mM	Millimolar
MM	Molecular mechanics
mP	Milli-polarisation
MS	Mass spectrometry
MW	Molecular weight
<i>N</i> -terminal	Amino terminal
nm	Nanometre
nM	Nanomolar
NMR	Nuclear magnetic resonance

ns	Nanoseconds
O.D	Optical density
OB-fold	Oligonucleotide/oligosaccharide binding-fold
ORC1	Origin recognition complex 1
PCA	Principal Component Analysis
PBSA	Poisson-Boltzmann surface area
PDB	Protein Data Bank
PEG	Polyethylene glycol
POT1	Protection of telomeres 1
PPI	Protein-protein interaction
ppm	Parts per million
ps	Picosecond
QM	Quantum mechanics

r	Interatomic distance
R & D	Research and development
R^2	Coefficient of determination
RAP1	Repressor activator protein 1
RMSD	Root mean square deviation
RP-HPLC	Reversed phase- high performance liquid chromatography
RPA	Replication protein A
rpm	Revolution per minute
SBDD	Structure-based drug design
SDS-PAGE	Sodium dodecyl sulfate- polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SPPS	Solid-phase peptide synthesis

SPS	Solution-phase synthesis
Ss	Single strand
T-loop	Telomeric loop
tBu	Tert-butyl
TFA	Trifluoroacetic acid
TI	Thermodynamic integration
TIN2	TRF1 interacting nuclear factor 2
TIP3P	Transferable intermolecular potential 3P
TIPS	Triisopropylsilane
TL	Telomere length
TMS	Tetramethylsilane
TPP1	Tripeptidyl peptidase 1
t_R	Retention time

TRF1	Telomere repeat binding factor 1
TRF2	Telomere repeat binding factor 2
TRFH	Telomere related factor homology
Trt	Trityl
uHTS	Ultra high throughput screening
US	Umbrella sampling
UV	Ultraviolet
v/v	Volume per volume
VMD	Visual molecular dynamics
WRN	Werner protein
$\Delta G_{\text{binding}}$	Gibbs free energy of binding
δH	Chemical shift

μM

Micromolar

1. Introduction

1.1 Telomeres and Telomerase Enzyme

Eukaryotic chromosomes are composed of linear DNA molecules, condensed around histone proteins in the nucleus of the cells. Generally, chromosome ends face two specific challenges: the first is how to ensure that they are not mistaken for DNA double strand (ds) breaks and processed by a cell's DNA damage repair systems, and the second is how to preserve genomic integrity during the DNA replication process of cell proliferations ¹. As will be discussed in more detail below, the existence of telomeres at the end of the linear DNA could answer both of the mysteries ².

Telomeres can be defined as specialised nucleoprotein structures at the ends of chromosomes (Figure 1.1) to protect both ends of a chromosome from all events that can cause cell death ³. The nucleic acid part is repetitive sequence tract of TTAGGG/AATCCC double stranded sequences ending with a single-stranded (ss) G-rich 3' overhang. The G-rich strand of telomeric DNA is oriented 5'-3' towards the terminal part of the chromosome ⁴. The 3' single-strand overhang is approximately 200 nucleotides, bending backward to create a large loop structure known as a "T-loop" that sequesters the chromosome terminus ⁵. The overhang invades the double-stranded telomeric region and replaces one of the telomeric strands to generate a small loop called

displacement loop (D-loop) (Figure 1.1B). The structure produced seals the chromosome end and prevents it being recognised as a double-strand break ⁶.

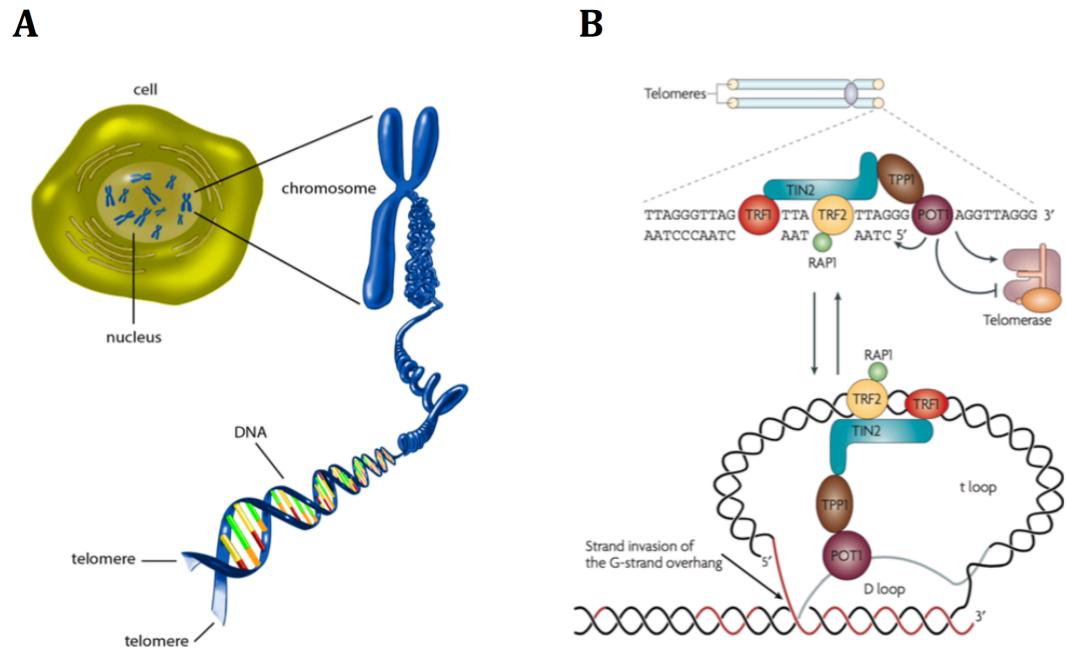


Figure 1.1 Telomeres location and structure in the cells. (A) Telomeres at the end of chromosomal DNA in the nucleus of the cell (taken from Hubbart 2014) ⁷. (B) The produced T-loop and D-loop structure in the telomeres (taken from Deng *et al.*, 2008) ⁸.

Human and other eukaryotes telomeres have virtually the same structure ⁹ and function ⁵. Telomere shortening can induce cancer and aging issues ¹⁰. Loss of the telomeric regions occurs slowly with each cell division ⁴. In humans, (TTAGGG)-n repeats at birth are about 15–20 kb (kilo base pairs) length, whereas they become 8–10 kb in adults because each cell division leads to the loss of some of the (TTAGGG)-n tracts, and occasionally the tracts loss are arising from exonuclease activities or deletion ⁵. Conversely, tract lengthening may occur either by the telomerase enzyme or by recombination mechanism.

The essential pathway of telomere length maintenance is through the telomerase enzyme, a reverse transcriptase, which adds further copies of the G-rich repetitive sequences to the 3' end of telomeres, after which DNA polymerase finalises the complementary strand. This enzyme is generally only activated in stem cells and gametes and permits them to proliferate indefinitely. As a result, in human somatic cells, replication potential is limited to 50–70 cell divisions because without active telomerase, telomeres shortening gradually lead to a permanent proliferation arrest (senescence), proliferating cells beyond senescence limits (crisis) and cell death subsequently. In contrast, the proliferation potential of most cancer cells is unlimited because the telomerase enzyme is also activated to maintain the stability of the short individual telomeres; hence, any approach that could inhibit the enzyme and thus promote telomeric attrition has the potential for the selective killing of cancer cells ^{11,12}.

1.2 History of Telomere Discovery

The importance of chromosome ends was first recognised by Muller (1938) and McClintock (1941) who established that the terminus of eukaryotic chromosomes has a different structure ¹³. Then, Leonard Hayflick discovered that cultured normal human cells have limited capacity for division; afterward, they reach the state of senescence; a phenomenon now defined by the 'Hayflick limit'. This achievement has enabled other researchers to obviously progress in understanding the molecular mechanisms of ageing ¹⁴. Later on, Alexey Olovnikov hypothesised that some mechanism had to be available to

repair and maintain the ends of chromosomes or the continuing replication of linear DNA by polymerase enzymes would gradually cause the loss of terminal sequences; and, in 1971, Olovnikov described the existence of the DNA end-under replication problem ¹³. Finally, the Nobel Prize in medicine or physiology 2009 was awarded to Elizabeth Blackburn, Carol Greider and Jack Szostak because of their role in the discovery of telomere and the telomerase enzyme ¹⁵.

1.3 Telomeric Proteins

A telomere contains many copies of each of six key protein components, known collectively as “shelterin” or the “telosome”, as shown in Figure 1.1B ¹⁶. Shelterin has at least three key functions in the telomeres, which are: regulating the structure of the telomeric end, contributing to the creation of T-loops for protecting telomeres and restraining the synthesis of telomeric DNA by telomerase ¹⁷. All of the shelterin proteins can perform their functions in conjunction with associated proteins (Table 1-1).

Table 1-1 The full name, main binding site, function and dysfunction of shelterin and shelterin-associated proteins that regulate telomeric length ^{16,18,19}.

Protein	Full name	Binds to	Normal function	Dysfunction
Binding directly with telomeres				
TRF1	Telomere repeat binding factor 1	Ds DNA	Negative regulator of TL: telomerase repressor	Inhibition (↓) lead to TL increase (↑)
TRF2	Telomere repeat binding factor 2	DsDNA	Negative regulator of TL and ATM inhibitor	↓ lead to G-overhang ↓; hRAP inhibition; DNA damage factors ↑
POT1	Protection of telomeres 1	Ss DNA	Connecting ss to dsDNA	POT 1 ↓ cause G-overhang ↓ and TL↑
		TPP1	Stimulating the activity of helicase, and preventing re-annealing	
Indirect binding through TRF1 or TRF2				
TPP1	Tripeptidyl peptidase 1	POT1	Binding POT1 to TRF1	↓ lead to TL ↑
		TIN2		
TIN2	TRF1 interacting nuclear factor 2	TRF1	Control poly (ADP-ribose) polymerase activity tankyrase	↓ lead to TRF1/2 ↓, which cause hRAP1 ↓ and TL ↑
		TRF2	Connecting TRF1 and TRF2	
		TPP1	Tethering TRF1 complex to TPP1	
RAP1	Repressor activator protein 1	TRF2	Negative regulator TL and component of the DNA repair	↓ lead to TL ↑

		response		
Associated proteins				
Apollo		TRF2	Protection of telomere during and after replication	↓ lead to DNA damage signal
pinX1		TRF1	↓ of telomerase activity	↑ lead to telomerase activity ↓ and TL ↓
pinX3				↑ lead to TL ↑
Tankyrase 1	TRF1-interacting ankyrin-related ADP-ribose polymerase 1 and 2	TRF1	TRF1 down-regulation	Tankyrase ↑ lead to TRF1 ↓ and cause TL ↑
Tankyrase 2			Cell division	

1.3.1 TRF1 and TRF2 Proteins

Both of the TRF1 and TRF2 proteins are fundamental, directly binding double-stranded telomeric DNA ²⁰ and then recruiting most of the other telomere-associated proteins. However, their molecular mechanism remains unknown. The two proteins share the same molecular architecture, characterised by a C-terminal Myb/SANT DNA binding domain, which is responsible on binding with telomeric DNA, and an *N*-terminal TRFH (Telomere Related Factor Homology) domain, which mediates homodimerisation and recruits other telomeric proteins ²¹. The N terminus of TRF2 contains a glycine and arginine (GAR) domain, while the N terminus of TRF1 encompasses aspartic acid and glutamic acid (DE) domain (Figure 1.2a).

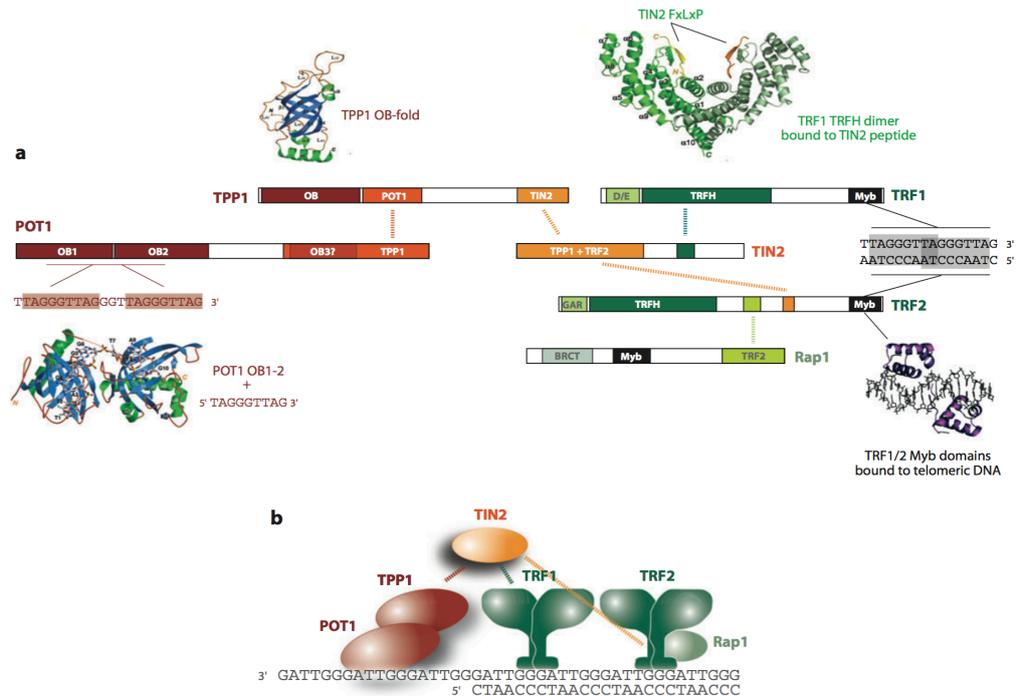


Figure 1.2 Schematic representation of the shelterin proteins in the telomere, their binding domains and their interactions. (a) Interactions of the components of human shelterin (b) Shelterin position in the telomeric DNA (taken from Palm and de Lange, 2008)¹⁹.

As shown in Figure 1.3, TRFH domains of both TRF1 and TRF2 have virtually the same protein docking sites (F142 in TRF1 and F120 in TRF2); however, *in vivo*, TRF1 interacts with TIN2 through TRFH domain, while TRF2 interacts with TIN2 in a region outside of the TRFH domain, through a C-terminal domain, and TRF2_{TRFH} recruits Apollo protein (shelterin-associated protein), instead; it is clear that the differences in recruiting proteins by TRF2_{TRFH} and TRF1_{TRFH} result from some structural differences²¹. Structural analysis shows that sequences FxLxP and YxLxP are the target sites for the TRF1 F142 and TRF2 F120 docking sites, respectively (x is any amino acid). In addition to

TIN2 and Apollo, other TRF1 and TRF2 interacting proteins might also use the F/YxLxP motif to interact with the TRFH docking site. Indeed, a number of non-shelterin proteins that are suggested to interact with TRF1 and TRF2 have a conserved F/YxLxP motif; for example, Ataxia Telangiectasia Mutated (ATM) protein. Nevertheless, several accessory proteins make interactions outside of the TRFH domains. For example, tankyrase1 interacts with the *N*-terminal domain of TRF1 (D/E rich area) (Figure 1.3a); also, WRN and ORC1 bind with the *N*-terminus of TRF2 (Figure 1.3b) ¹⁹.

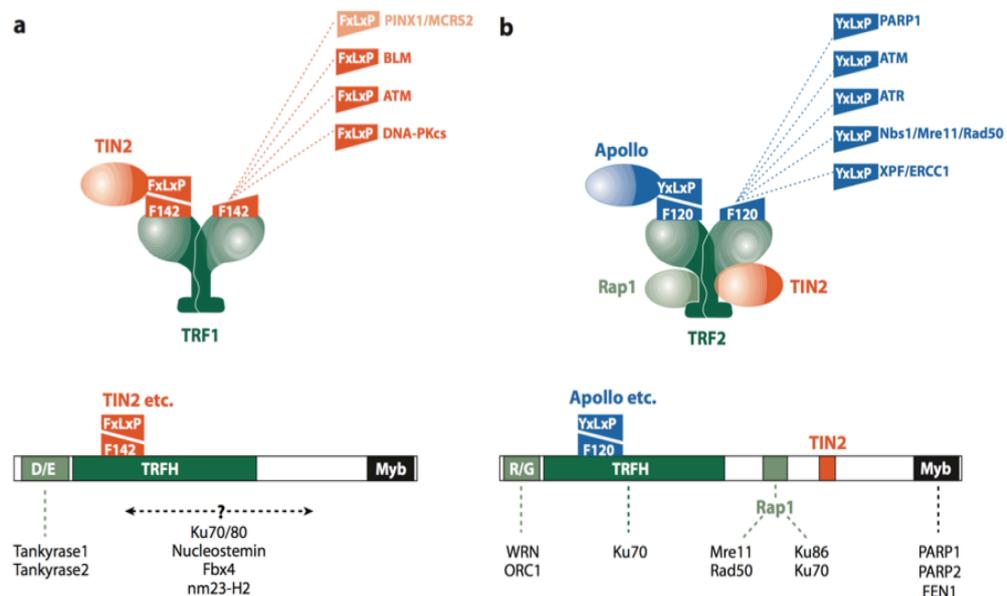


Figure 1.3 TRF1 and TRF2 binding proteins. (a) Proteins that are associated with TRF1. (b) Proteins associated with TRF2 (taken from Palm and de Lange 2008) ¹⁹.

Regarding the roles of the two proteins, TRF1 has a crucial role to maintain telomeric length and shelterin component stability in the telomeres of the chromosomal DNA ²², and TRF2 protects telomeres through keeping T-loops stable and inhibiting DNA Damage Response (DDR) ²³; any disruption of these

proteins causes telomeric instability and may lead to direct apoptosis of the cells²⁴.

1.3.2 TIN2 Protein

TIN2 is one of the shelterin proteins; it creates a bridge and assembles shelterin proteins by directly interacting with TRF1, TRF2 and TPP1/POT1; moreover, it tethers double stranded and single stranded telomeric DNA, stabilise TRF2 protein through attaching TRF1 and it is essential for the stability and function of the complex^{25,26} (Figure 1.2). As discussed above, TIN2 binds to the TRF1_{TRFH} through the C-terminal motif, whereas, it connects to the TRF2 and TPP1 through the *N*-terminal domain. Depletion or the release of mutant variants of this protein has a destabilising effect on shelterin^{27,28}.

1.3.3 RAP1 Protein

Another key shelterin proteins is RAP1; it's a crucial binding partner of TRF2; each RAP1 interacts with one TRF2, therefore, this protein's stability is related to the TRF2. It has three detectable domains; a Myb domain that mediates interactions of the protein with an unknown partner, an *N*-terminal BRCT motif, which binds with a phosphorylated peptide, and a C-terminal domain that confers the interaction with short helical region of the TRF2 hinge domain (Figure 1.2)¹⁹.

1.3.4 TPP1 Protein

TPP1 protein links the POT1 protein with TIN2 through its centrally located POT1 interaction domain and C-terminal TIN2 interaction domain (Figure 1.2), thus recruiting POT1 to the shelterin proteins TRF1 and TRF2. It has a Ser-rich region with an unknown function. It is presumed that TPP1 interacts with the telomerase enzyme due to the presence of an OB-fold domain at the N terminus (Figure 1.2a). Mutants of TPP1 protein, chiefly those with disturbed binding to POT1, lead to the removal of POT1 protein from telomeres, telomere deprotection and the triggering of a DNA damage response^{2,19}.

1.3.5 POT1 Protein

POT1 is one of the shelterin-complex proteins that bind directly to the telomeric DNA, connecting the single-stranded 3' extension at the end of chromosomes and TPP1 protein, as shown in Figure 1.2b. The N-terminus of this protein contains two OB folds, which have the ability to recognise the G-strand telomeric sequence *in vitro*. The OB-folds are composed of a common protein domain, originally identified as an oligonucleotide or oligosaccharide-binding domain, and can recognise single-stranded telomeric overhangs. Moreover, according to sequence analysis, a third OB-fold is located in the C-terminus of POT1 (Figure 1.2a)¹⁹. Genetic studies in humans, mice, plants, yeast and Tetrahymena have documented that POT1 plays a basic role in telomere integrity, since POT1 knockout stimulates damaging of the DNA, initiated by the ataxia telangiectasia related protein kinase (ATR). Inhibition of

the activity of ATR by POT1 is probably due to POT1 binding to telomeric ssDNA and inhibiting ATR activation by blocking access of the ss binding protein RPA (replication protein A), by which ATR is recruited to the telomere²⁸. In addition, repression of POT1 may lead to increase in the telomere length¹⁶.

1.4 Tumourigenesis and Telomeres

A set of genetic modifications, which lead to disturbed growth and differentiation of normal cells, define cancers. As described in the literature, tumourigenesis can arise from various cornerstones and is characterised by distinctive hallmarks. The alterations can be rationalised as inhibitions of tumour suppressor genes and improperly activated normal cellular genes²⁹. According to Hanahan and Weinberg (2011), six distinctive signs characterise cancer from normal cells, these are: escaping growth suppressor signals, avoiding DNA damage responses or resisting apoptosis (automatic cell death), incorrect signaling of cells to proliferate, production of angiogenesis, infinite cell replicative potential and invasive character³⁰⁻³².

Intact telomeres have a profound impact on the division and survival of cells. Telomeric shortening or attrition contributes to chromosomal instability that may promote tumourigenesis^{33,34}. Shamas has suggested telomeric dysfunctions are highly associated with head, neck, gastrointestinal, renal cell, bladder and lung cancers due to a sequence of biochemical reactions *in vivo*³⁵.

Normally, any critical shortening or erosions of the telomeres in stem, somatic and epigenetic cells can deprotect the telomeric cap and lead to a DNA damage response, which activates the tumour suppressor p53 protein, resulting in replicative senescence and cell apoptosis. In contrast, deficiencies related to the lack of p53 protein or incompetency of the cells' checkpoint mechanisms can trigger continuous telomere shortening until the cells reach crisis and genomic instability; the consequence will be chromosomal end-to-end fusions that either produce cell apoptosis or developing oncogenes (Figure 1.4) ^{36,37}.

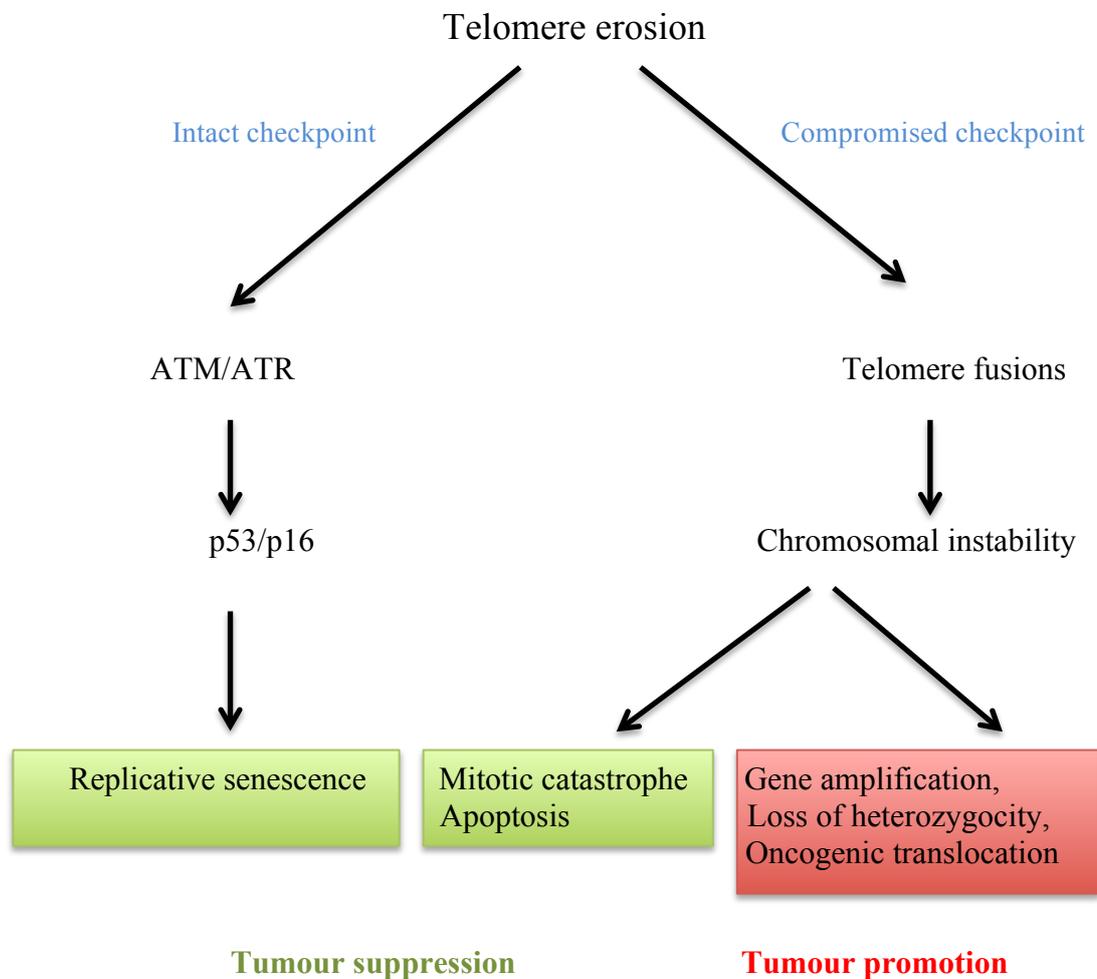


Figure 1.4 Consequences of the telomere attrition in the eukaryotic cells (adapted from Xu *et al.*, 2013) ³⁷.

1.5 Telomere Disruption and Destabilisation for Cancer Treatment

Cancer cells prevent their telomere shortening and cell death through two mechanisms. The first pathway involves activating the telomerase enzyme as a maintenance factor to prohibit further telomere shortening because it adds multiple copies of the 5'-GGTTAG-3' to the telomeric end (G-strand of the telomere). Telomerase activity is overexpressed in about 85–90% of tumours and it is absent in only 10–15% of those cells. The second pathway of lengthening telomeres in cancer cells lacking telomerase is sustaining telomeric length through recombination-based mechanism, known as alternative lengthening of telomere (ALT)^{38,39}. Since stable telomeres are essential to the survival of cancer cells, different chemotherapeutic approaches have been developed in order to disrupt or destabilise telomeres or telomerase⁴⁰. Most of the research has been focused on targeting the activity of the telomerase enzyme by directly inhibiting telomerase or by preventing telomerase access to the telomeres; however, more recently the approach of directly targeting telomeric proteins has developed as a pathway to cancer chemotherapy⁴¹⁻⁴³.

1.5.1 Targeting Telomerase Enzyme

The strategy of targeting telomerase enzyme for cancer treatment has been the focus of most researchers because telomerase overexpression is the most obvious characteristic of most cancer cells⁴⁴ and it could be an attractive target

for cancer therapy as a consequence of the persuasive link between cellular immortalisation and reactivation of telomerase. Any compound able to inhibit the telomerase enzyme would theoretically return cancerous cells to the normal situation of a (finite) potential for replication. Furthermore, molecules targeting telomerase enzyme are characterised by selectivity for cancer cells over normal somatic cells. However, therapeutic strategies for shortening telomeres through administration of telomerase inhibitors are practically challenging because significant telomere shortening would require continuous treatment of the patient for multiple population doublings of tumour cells and the population doubling time of most solid tumours is several days to weeks; thus, telomerase inhibitors may take months to have an effect on the patient. These obstacles have hampered the release of these drug candidates into the market and clinic.

40,41,45

1.5.2 Targeting Telomeric DNA (G-quadruplexes)

This approach directly targets the guanine rich sequences of the telomeric ends and modifies telomeric structures⁴⁶ by guanine–guanine base pairing of the 3' sequences to form a particular structure called G-quadruplex or G4 DNA that makes a protective cap which is no longer recognised by the telomerase enzyme. The new G-quadruplex structures consist of two or more G-tetrads (quartets), and each quartet comprises four guanine bases, linked together by 8 hydrogen bonds in a cycle (Figure 1.5A), when a central cation can improve the stability of the structure. G-quadruplexes can assemble either through intermolecular bonding (Figure 1.5B) or intramolecular bonding (Figure 1.5C),

but human telomeric DNA can fold to generate an intramolecular G-quadruplex structure, depicted as a basket, having three G-tetrads, each stabilised by Hoogsteen base pairing (Figure 1.5D). The stabilised G4 DNA in cancerous cells prevents the ability of telomeres to lengthen after cell division, and eventually cancer cells will die ⁴⁷.

Primarily, the aim of targeting the 3' DNA sequence through G-quadruplex inducing ligands was to mitigate telomerase enzyme effects on the telomeres ^{48,49}, but a series of studies on a variety of diverse G-quadruplex ligands, confirmed unexpected outcomes, specifically the apoptosis of the cells after just a few days of ligand exposure, before significant telomere shortening could have taken place. These results have been interpreted as due to ligand induced telomeric cap opening and structural disruptions, which finally produce chromosomal end-to-end fusions ⁵⁰⁻⁵². These arguments have support from a number of observations and the conclusion is that the target of G-quadruplex ligands is in effect telomeres rather than telomerase ⁴¹. Hence, based on these results, the formation of G-quadruplex promoting ligands could be a novel therapy for cancer treatment ⁵³.

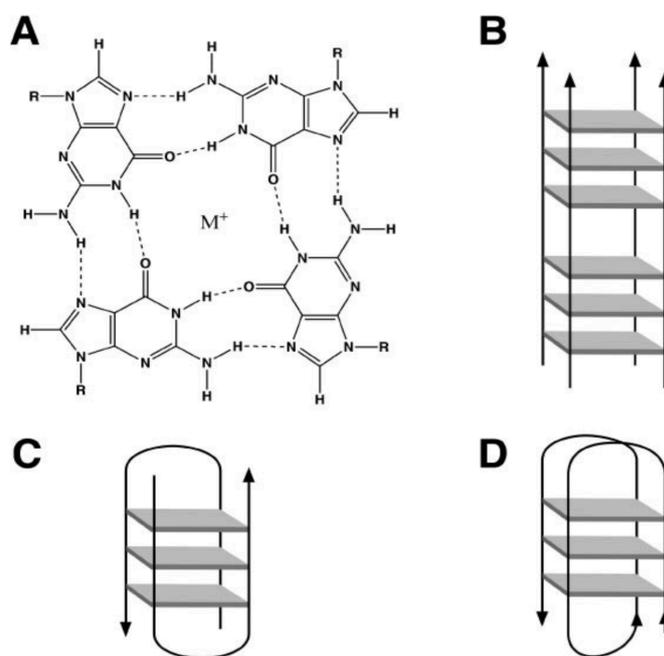


Figure 1.5 Structures of G-tetrad and G-quadruplexes. (A) G-tetrads are 4 guanine residues forming a planar structure (B) A parallel model of G-quadruplexes. (C) A model of intermolecular G-quadruplexes (D) A model of intramolecular basket of G-quadruplexes (taken from Rezler *et al.*, 2003)^{47,54}.

1.5.3 Targeting Shelterin Proteins

The shelterin targeting approach is the strategy of targeting compounds that have destabilising, disrupting and/or inhibiting effects on the shelterin proteins, on the basis that they may have antitumour effects through disruption and apoptosis of the cells (Figure 1.6); for example, inhibition of TRF2 protein can induce massive telomere dysfunction and apoptosis of tumour cells by uncapping of the telomere ends, end-to-end fusions and activation of ATM/p53 DDR pathway⁵⁵; another example is the telomeric damage induced by TPP1 knock down due to inhibiting the associations between POT1 with TRF1 and

TRF2 proteins⁵⁶. Furthermore, Pal *et al.* (2015) explained that inhibiting TRF1 and TRF2 proteins can produce a telomere disruption and cell apoptosis in renal cell carcinoma through arresting the progression of cell cycle in Phase S and G1/S, respectively⁵⁷.

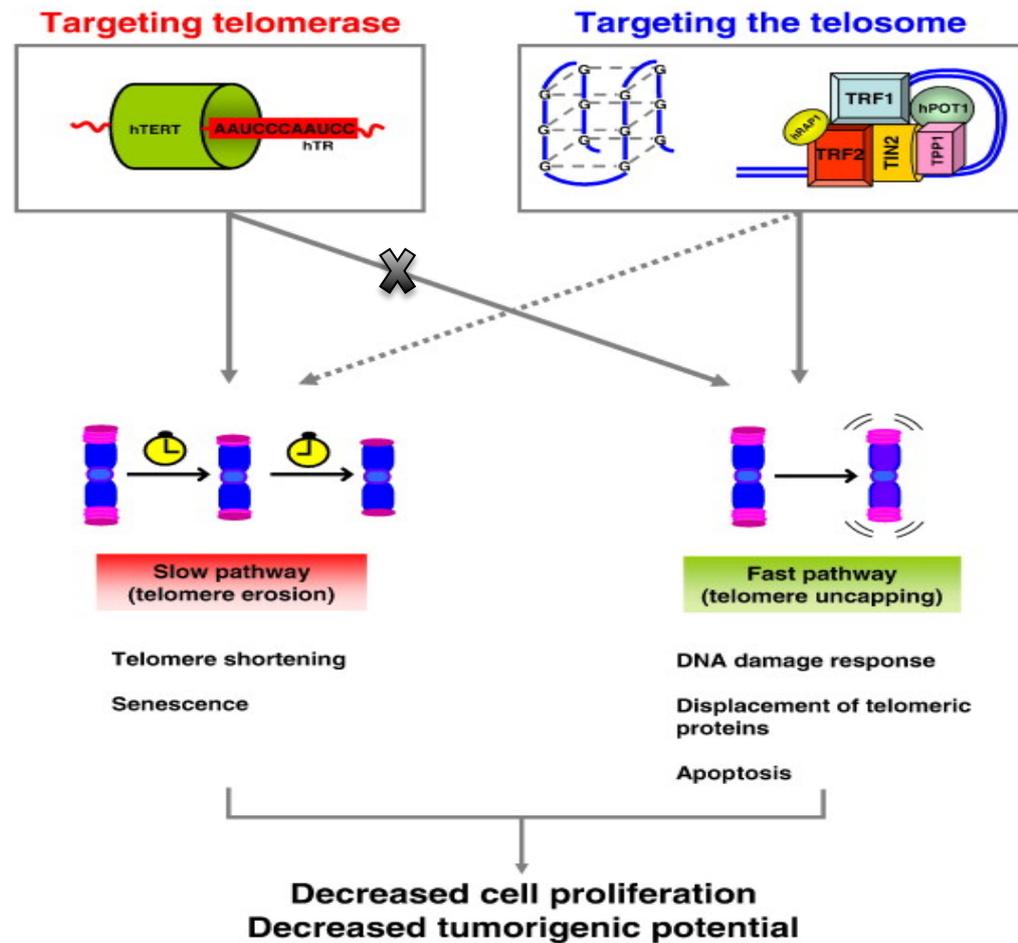


Figure 1.6 Different approaches for inhibiting telomeres (adapted from Folini *et al.*, 2009)⁵⁸.

The latest publication on mouse TRF1 ablation by Garcia-Beccaria *et al.* (2015) has examined the effect of TRF1 abrogation in lung cancer cells with inactivated p53 suppressor proteins; significant positive results have been

obtained on the number and size of the malignant cells in the first generation, disregarding the length of the telomeres; this reveals that disturbing TRF1 could potentially stop cancer progression and metastasis in both mouse models and human. In addition, this approach may retain selectivity for cancer over normal cells (a worry when targeting telomeres rather than telomerase), as it has proved to have a minimal effect on the growth and activity of other organs when used long-term, just a slight reversible inhibition of bone marrow and blood cells was noticed. The promising results could become a possible new chemotherapy for lung cancer. Furthermore, this strategy could be effective for diverse types of cancer because it follows a universal mechanism of telomere inhibition (telomeres uncapping) ⁵⁹.

In conclusion, integrity of the shelterin proteins and protein–protein interactions (PPIs) has a significant effect in telomere maintenance ⁶⁰. In addition, the structure and accessibility of telomeres in cancer cells seems to vary considerably from that of normal cell telomeres, which might lead to specific cytotoxicity ⁴¹. Also, the promising results of the G-quadruplex compounds (as discussed in Section 1.5.2) might increase the chance of cytoselectivity in shelterin inhibitors. However, the individual shelterin components have not so far been studied greatly as targets for small-molecule drugs, and the challenge remains to discover novel compounds, which selectively interfere with protein–protein and protein-DNA interactions to destabilise shelterin components.

1.6 Drug Discovery Process

The sequence of steps from candidate molecule identification to market release of a new medication is termed the drug discovery process. In medicinal chemistry the term “hit” describes a molecule able to modulate and affect the specific target with sufficient activity⁶¹; while a “lead compound” is a molecule that has therapeutic and pharmacological activities, but requiring chemical optimisation to improve physicochemical characteristics and so be suitable for use as a drug⁶². Normally, the drug discovery process is an expensive and time consuming (12–15 years). It starts with a hit and lead identification; after that, lead optimisation is achieved. Then, the expensive processes of preclinical and clinical studies follow (Figure 1.7). Target identification and validation is realised either before or after a ‘hit’ identification. In classical drug discovery, lead identification comes before target validation⁶³⁻⁶⁵, while, in most more contemporary approaches, target identification and validation precedes hit and lead finding⁶⁶.

Significantly, 90% of molecules identified at the start of the drug discovery process fail somewhere along the pipeline. Within the clinical phases, on average, 38% of candidates are abandoned in phase I because of their toxicity, lack of efficacy and bioavailability hurdles. Next, phase II causes 63% of the remaining to leave the process, and only 45% of the molecules that have succeeded in previous phases, can pass phase III clinical trials. On the basis of this data, the most obvious difficulties in developing new drugs are clearly issues of efficacy, bioavailability and toxicity⁶⁷. Two main factors are

contributing to increase costs of research and development (R&D) in the pharmaceutical industries; the first is the new regulation by the authorities to confirm a drug with high efficacy, safety and quality; the second is the increased efforts of governments to decrease medicine costs and compensating brands by generic companies ⁶⁸.

It is clear therefore that both of the high failure rate of new drug development projects and high costs of R&D cause risen costs of new medicine in the market, which may not be affordable for the consumers and the companies ⁶⁹. Therefore, pharmaceutical companies to be profitable and competitive, have to reduce costs through improving efficacy and bioavailability and decrease toxicity of the candidate molecules.

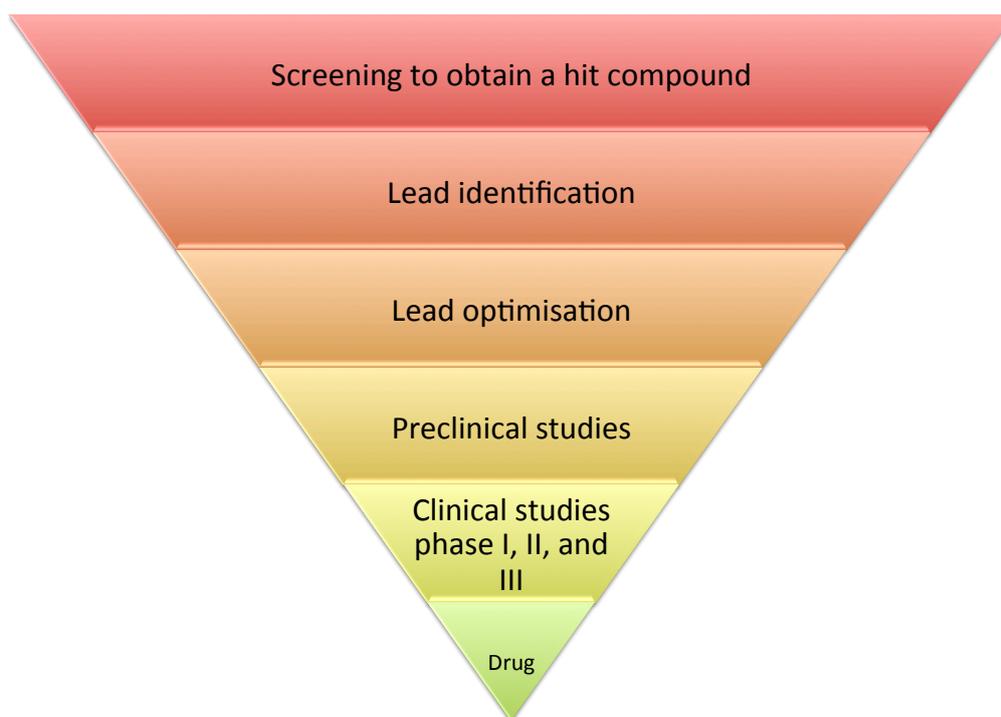


Figure 1.7 Main steps of drug discovery and development process.

1.7 Lead Compound Identification

Different approaches are available to first identify a hit or lead compound ⁷⁰. Prior knowledge on the receptor or ligand is crucial for developing and optimising a lead compound. In the modern drug discovery processes, structural information on the protein or any targets plus familiarity with the ligand's structural, physicochemical and pharmacokinetic properties (e.g. from identified hits) are important factors to improve the success of the process ⁷¹. It can be used with the aid of *in silico* approaches to develop a lead molecule ^{72,73}.

Most of the methods are based on target molecules. After target identification or validation, various experiments are performed to find a hit compound, perform hit-to-lead development and lead optimisations ⁶⁵ and usually the target is a protein, gene, gene product, or the modulation of a molecular mechanism ⁷⁴. The advantage of this approach is the direct identification of an effective molecule against the specific well-known target, whose role in the disease is presumed well established; however, in reality the target may not provide a promising result due to a poor correlation with the disease state or an insufficient therapeutic window ⁷⁵. On the other hand, sometimes the 3D structure of the target molecule is not identified or the approach is not dependent on the target structure, such as phenotypic screening. Most of the approaches are illustrated below:

1.7.1 Phenotypic Screening

The traditional method of hit and lead compound discovery in classical pharmacology is the phenotypic screening approach. All of the serendipitous discoveries are related to this method. The technique is based on testing compounds *in vivo*, consequences on the cells, tissues, organs and the whole body system can be seen through physiological changes and modifications to the disease state. After confirming the results, investigations are undertaken to identify the molecular target ⁷⁶. The advantage of this method is that one obtains directly disease responses to the tested compounds, which are closer to reality than proxy readout and there is no need for the prior knowledge of the molecular mechanism of the disease ⁶⁵. However, the limitations are low throughput screening and potentially significant challenges to optimise the candidate molecules without the target information ⁷⁵.

1.7.2 High Throughput Screening

High throughput screening (HTS) involves a robotic assay process for mass screening of compound libraries in the early stages of drug discovery to identify hit molecules, development to a lead molecule and optimisation of a lead ⁷⁷. In the early days of HTS, thousands of compound were screened per day, after several years this evolved to 100,000 per day in ultra HTS (uHTS) ⁷⁸, and recently, a paper published in 2010 by Agresti *et al.*, has described a new uHTS technique that can screen 100,000 assays per only 10 hours ⁷⁹. The first published work on HTS dates from the 1990s; now, it is an essential part of

most drug discovery projects in the pharmaceutical realm⁸⁰ because of screening large compound libraries during a limited time and miniaturised format of the assays through advances of liquid handling, microplate reader sensitivity and wells availability as 96, 384 and 3456^{78,81}. However, more recently the use of this technique has been cut back by several companies, due to the low rate of the molecules success to be a drug⁸². Typically, HTS is achieved by big Pharma companies; in addition, it is of interest to academic, government and non profit institutions with a smaller scale of running⁸³.

1.7.3 Fragment-Based Drug Discovery

Fragment-based drug design (FBDD) is considered to be a promising new strategy for screening compounds to find a lead compound⁸⁴. The principle of this technique is the screening of small fragments, following the “rule of three”: the fragments molecular weight (MW) < 300 Da, calculated log P (ClogP) ≤ 3 and both hydrogen (H) bond donors and acceptors should be ≤ 3 ^{85,86}. Individually the small fragments are low affinity molecules, but after combining several fragments, high-affinity ligands are produced. This approach require a dramatically smaller library than HTS method and it can be used as a complementary method of drug discovery⁸⁷.

1.7.4 Computer-Aided Drug Discovery

Computer-aided drug design (CADD) is also called *in silico* drug design. It is based on computationally synthesising molecules and analysing molecular interactions to assist and speed up hit molecule identification, lead selection and optimisation and also to predict potential pharmacokinetic difficulties⁸⁸. CADD can considerably reduce the time, cost and workload in drug discovery projects; for example, selecting compounds from a library of molecules through virtual screening before experimental tests could produce the same level of lead identification at much reduced time and cost. Furthermore, it has the potential to improve efficacy, reduce toxicity and optimise pharmacokinetic activities of the selected molecules. In addition, it may be applied to design new molecules or modify existing structures to produce a novel compound^{82,89}. Figure 1.8 outlines the role of CADD in drug discovery process, which the first step is identifying a target. Then, according to the availability of 3D receptor structure, choosing between ligand-based or structure-based drug design strategies. After *in silico* screening of the compounds to find a lead compound, lead optimisations step is achieved, followed by *in vivo* assays to confirm that if the compound can be a drug candidate and ready for the new steps of preclinical tests⁸².

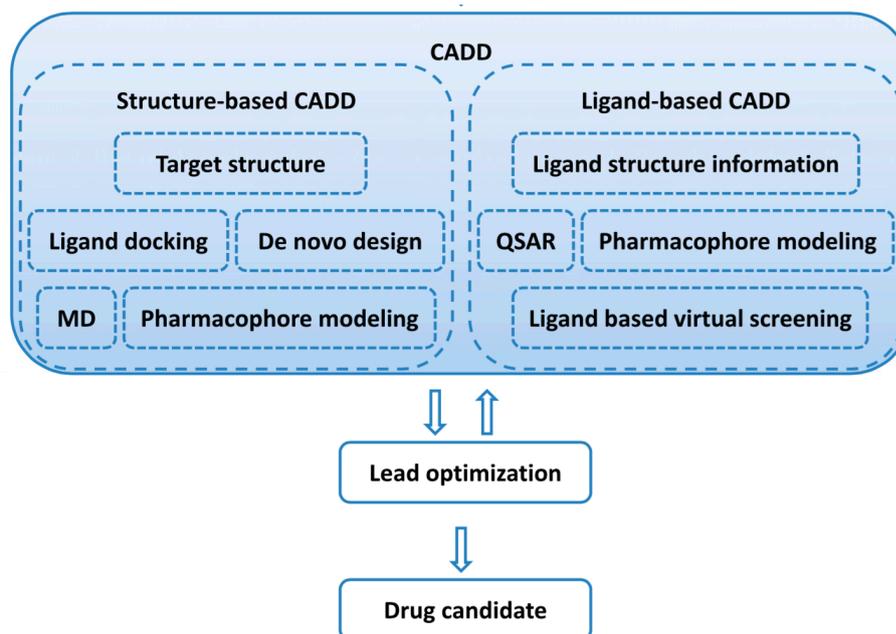


Figure 1.8 The role of CADD in the drug discovery processes ⁸².

1.7.4.1 Structure-Based Drug Design

Structure-based drug design (SBDD), which can also be described as a receptor-based drug design, encompasses identifying a lead compound and optimising it on the basis of knowledge of the 3D crystal or nuclear magnetic resonance (NMR) structure of the potential target. Docking and screening are applied to a library of selected molecules to find the most appropriate molecule as a novel compound to activate or inhibit (as required) the target receptor through predicting interaction energies between them. Detailed structural information on the target is the basis for designing ligand molecules and then docking. A major factor driving the development of SBDD has been the proteomic and genomic evolution, which has resulted in the production of

hundreds of new proteins and availability of their high-resolution X-ray crystal structures that can be used as potential drug targets⁹⁰⁻⁹³. Most of the high-resolution crystal structures of approximately one hundred thousand proteins are published in the Protein Data Bank (PDB) and Cambridge Crystallographic Data Centre that can be used as potential sources for SBDD^{94,95}.

Formerly, static structures of the proteins were used in SBDD, but in reality, protein samples are ensembles of various conformation states, each conformation has a different free energy. Usually, proteins occupy a low energy state that their structures may be significantly different from the crystal structure; consequently, docking ligands to a single, specific conformation of the receptor may well lead to incorrect predictions, as it may not be the ligand bound conformation. However, now molecular dynamics and molecular modelling computational tools are helping solve these issues, as they can simulate and estimate a protein's conformational space as a collection of snapshots for each protein structure describing their fluctuations due to conformational sampling^{96,97}. The essential steps of SBDD are illustrated in Figure 1.9.

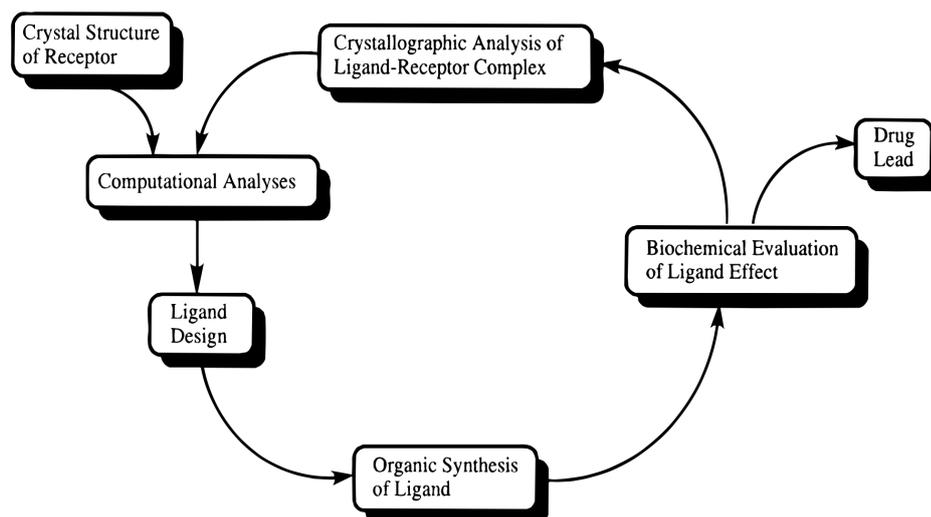


Figure 1.9 Structure-based drug discovery process to find a lead compound (taken from Babine and Bender 1997) ⁹⁸.

1.7.4.2 Ligand-Based Drug Design

Ligand-based drug design (LBDD) is one of the approaches of *in silico* drug design to find a lead molecule and optimise it when the 3D structure of the potential target is typically unknown; therefore, structural investigations of the ligand molecule and its pharmacophore are performed ⁹⁹. Because LBDD is a ligand knowledge-based technique, previous information from active ligands is used to build up a picture of the similarity of the active sites and functional groups ⁸².

1.8 Molecular Recognition

Molecular recognition deals with the investigations and analyses of the non-covalent binding interactions between two or more molecules in the biological systems. In addition, it is essential to investigate various receptor–ligand binding interactions¹⁰⁰. Advances in biological, chemical and technical sciences have caused the evolution of this subject, especially those techniques that reveal 3D structural information of the molecules. In the drug discovery arena, all of the interactions and contacts between cell–cell, antigen–antibody, enzyme–substrate, hormone–receptor and drug–receptor are realised and analysed in terms of molecular recognition¹⁰¹. In addition, it provides a theoretical framework to evaluate and optimise specificity, potency and stability of the ligands^{61,102}. Most of the non-covalent binding interactions are elucidated below.

1.8.1 Hydrophobic Contacts

Hydrophobic or lipophilic molecules are nonpolar molecules that do not optionally interact with water. In aqueous solution, these molecules tend to self-associate and form clusters to decrease their surface area and exclude water molecules¹⁰³. The hydrophobic molecules in the aqueous medium form a ‘hole’, which is surrounded by water molecules to produce a cage-like structure. The cages decrease the entropy of the systems due to aggregating the hydrophobic groups and producing a cluster to decrease the contact surface

area with the aqueous medium. Contacts between hydrophobic molecules are known as hydrophobic interactions; they are relatively stronger than other types of non-covalent interactions, such as hydrogen bonds and van der Waals forces¹⁰⁴. The strength of hydrophobic interactions is affected by several factors; for instance, aliphatic organic molecules have a stronger interaction than those of aromatic organic molecules. Additionally, a stronger interaction is produced through linear carbon chains rather than branched chains because the latter produce steric hindrance and cannot repel water molecules so effectively. Furthermore, increasing numbers of carbons in the molecule and temperature are directly correlated with interaction strength¹⁰⁵. Importantly, hydrophobic effects have a dynamic function in folding and energy-minimising protein structures to keep them biologically active, decreasing their surface area and so protecting them from undesirable reactions with water.

Globular proteins in aqueous solutions are arranged in a such way that hydrophobic amino acids are positioned in the interior, which might be a ligand-binding site in a druggable protein, whereas hydrophilic amino acids often cover the surface of proteins¹⁰³. Hydrophobic contacts between the ligands and proteins can be rationalised as entropic or enthalpic gain because water molecules are no more positionally restricted (free) after establishing hydrophobic interactions between the ligands and the proteins; also, the displaced water molecules allow an established hydrophobic contact between the ligands and the receptors; in addition, the liberated water molecules, which previously could not interact with hydrophobic groups, can form hydrogen bonds with the bulk of solvent molecules and lead to enthalpic gain^{106,107}.

1.8.2 Van der Waals Interactions

The intermolecular attractive and repulsive forces are known as van der Waals forces; the attractive forces are owing to the favourable interactions between electronic multipoles, while the repulsive forces are due to spatial overlap of the electron orbitals^{108,109}. There are three types of van der Waals forces. The first is the attraction between two molecules with induced dipoles (induced dipole–induced dipole). The second and stronger type is between a permanently polar pole and a momentarily polar pole (dipole–induced dipole). The last van der Waals force is dipole–dipole interactions, which are produced between two molecules with permanently polar poles¹¹⁰. Van der Waals forces are generally considered as a weak non-covalent interactions between molecules, while summation of them can produce strong binding interactions and it is one of the most significant interactions in protein molecules¹¹¹.

Van der Waals attractive forces between a protein and a ligand occur over a very short distance range and have a distance dependence of $1/r^6$ (r is a distance between two nuclei). Therefore, optimised van der Waals attraction of a protein–ligand complex can occur when they have shape complementarity, which occasionally happens; however, the ligand or the protein might change their conformation to provide a complementary structure. Nevertheless, van der Waals attraction remains favourable and can stabilise a complex if the ligand atoms are not too close to the protein atoms⁹⁸.

1.8.3 Hydrogen Bonds

Aruna *et al.* in 2011 states that “a hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation”¹¹². Typically, hydrogen bonds exert a fundamental role in various biological interactions like proteins interactions¹¹³. Hydrogen bond strength is typically between that of covalent and van der Waals interactions; the energy of hydrogen bonds is directly related to donor–acceptor separation and the linearity of the donor atom and acceptor lone pairs¹¹⁴. Universally, hydrogen bonds favour specific distances between heteroatoms (donor–acceptor atoms) that lead to different energy outcomes, in such a way that strong bonds distance range are 2.2–2.5 Å with energy in the range 40–14 kcal/mol, moderate strength bonds are 2.5–3.2 Å with 15–4.0 kcal/mol energy and weak bonds are 3.2–4.0 Å with energy < 4.0 kcal/mol¹¹⁵.

1.8.4 π effects

The interaction of π systems of aromatic rings in different molecules can have energetic consequences. Three types of π effects are defined: π – π interactions, cation- π interactions and anion- π interactions. Regarding the π – π clouds of two aromatic rings, three interaction arrangements may be defined, displaced, T-shaped edge-to-face and sandwich bindings (Figure 1.10)¹¹⁶. In proteins, these interactions occur between the aromatic side chains of adjacent amino acids

like Phe and Tyr, as well as between the side chain aromatic rings of proteins and ligands possessing aromatic groups¹¹⁷.

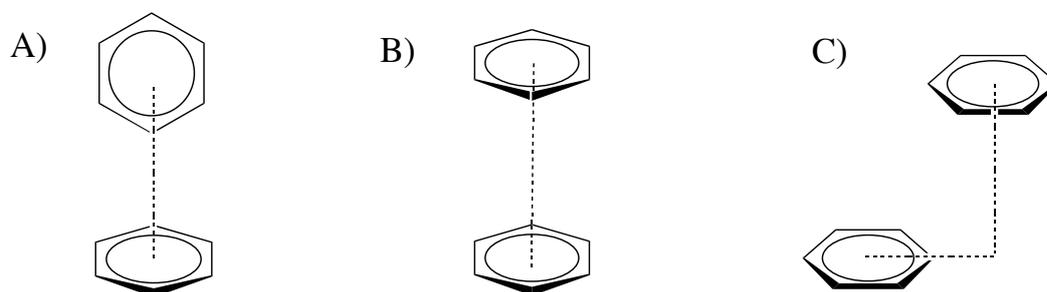


Figure 1.10 Stacking conformations between aromatic groups. (A) T-shaped $\pi-\pi$ interaction. (B) Sandwich shaped $\pi-\pi$ interaction. (C) Staggered or parallel-displaced $\pi-\pi$ interactions.

Another class of π interactions is those with cations or anions. Various investigations on proteins and proteins with ligands have proved this type of contact as very common between charged side chain amino acids and π -electron clouds of the aromatic rings of Phe, Trp and Tyr. Almost always, this interaction type is considered as strong¹¹⁸⁻¹²⁰.

1.8.5 Salt Bridges

Adjacent molecules of opposite charges may interact non-covalently through salt bridges, in which the interacting atoms have separation in the hydrogen bond range¹²¹. They can have an essential function in protein stability and protein-ligand interactions¹²²⁻¹²⁴ because they are a combination of

electrostatic and hydrogen bonds. Ordinarily, in protein structures, the bridges are created between the side chain carboxylate anions (RCOO^-) of Glu or Asp and the positive ammonium (RNH_3^+) or guanidinium charges of Lys and Arg, respectively ¹¹⁶. In addition, other amino acids like His, Ser and Tyr may produce salt bridges depending on their ionisation state, related to the environmental pH. However, if the distance between the two interacting atom is greater than 4.0 Å, it can not be consider as a salt-bridge ¹²⁴.

1.9 Therapeutic Molecule Sizes

1.9.1 Small Molecules

Drug molecules with a size < 900 Dalton (Da) are considered as small molecules. Generally, they can pass biological barriers such as cell membranes because of their tiny sizes and they can easily reach the required destinations ^{125,126}. Most of these drugs were discovered by different methods such as rational drug design, screening tests or serendipity and they were released to the market during the last century ^{65,127}. The molecules are typically synthesised in chemical laboratories ¹²⁸. To improve permeability and absorption of the small molecules, they can follow Lipinski's rule of five: the molecule should have less than 10 hydrogen bond acceptors and 5 hydrogen bond donors, $\text{MW} < 500$ Da and $\text{ClogP} < 5$ ¹²⁹.

1.9.2 Protein-Based Drugs

Recently, we see more compounds derived from biological sources coming to the market, biosynthesised in molecular biology laboratories through a protein expression and purification processes; they are termed biologics¹³⁰. These medicines may contain protein, nucleic acids and sugars. Unlike small molecules, they have a massive size (roughly > 5000 Da) and are characterised by increased specificity and less toxicity because size of biologics can significantly inhibit “off-target” issues. Nevertheless, the biologics huge size have a serious disadvantages, such as metabolic instability, poor membrane permeability, significantly low bioavailability and highly prone to immune response¹³¹.

1.9.3 Peptide-Based Drugs

Therapeutic peptides are sequences of 2–50 amino acids or similar peptidomimetics with agonistic or antagonistic activities on receptors^{132,133}. Peptides can be produced through either chemical synthesis or biosynthetic methods^{134,135}. The size of the desired peptide is the major factor that determines the method of production; small and medium sized peptide (2–50 amino acids), and peptides involving unnatural amino acids are constructed by chemical synthesis¹³⁶, while the recommended method for producing large and complicated peptides is biosynthesis because chemical synthesis is expensive and provides low yields.

Peptide drugs have several advantages over small molecule ligands; the first is the ability of peptides to inhibit PPIs because small molecules to a large extent could not inhibit them due to a wide interacting area of most proteins. Moreover, peptides typically have high target specificity and efficacy because of the significantly higher biological and chemical diversity compared with the small molecules^{133,137}. Furthermore, *in vivo*, peptides are metabolised into amino acids, which decrease the probabilities of drug–drug interactions and systemic toxicity¹³⁸. Finally, the short duration of action of peptides may decrease the risk of accumulating metabolites in the various tissues and producing complications¹³³. Peptides can also have potential benefits over biologics like antibodies and recombinant DNA therapies because due to their smaller sizes, they are less capable of producing immunogenic reactions; additionally, they have a significantly better cell penetration, lower manufacturing costs and greater stability^{136,139,140}.

However, low bioavailability and pharmacokinetic issues, especially in the linear peptides are considered as the two inevitable difficulties associated with developing peptide therapeutics. The low bioavailability problem is due to the peptides easily degradation by protease enzymes; and pharmacokinetic problems are due to the absorption limitations, easily metabolism and their low plasma residence time. Therefore, various approaches have been applied to overcome the mentioned hurdles, such as choosing alternative route of drug administration, applying new strategies of chemical synthesis and chemical modifications to circumvent pharmacokinetic and bioavailability issues of the peptides¹³⁶.

1.10 Inhibiting PPIs as a Therapeutic Approach

PPIs play vital functions in the body, as they are responsible for regulating an enormous variety of biological processes and cellular activities. The binding partners interact via a patch on the accessible surface area of each protein, termed the protein interface area, which repels solvent molecules¹⁴¹. Interestingly, improper or disrupted interactions between the two proteins at the interface can result in abnormality and diseases¹⁴²; therefore, it is not surprising that intervening at these interfaces and grooves may offer an attractive therapeutic target¹⁴³. Almost invariably, proteins interact with each other through a large interface area, but most of the binding affinity values are associated with a limited number of residues in the critical region of PPIs interface, which is called the “hot spot”¹⁴⁴.

The aim of PPI inhibitors is modulating the functions of interacting proteins to bring about a therapeutic effect; the modulations are typically achieved through molecules targeting the proteins hot spot¹⁴⁵; and importantly, small size hot spots are favourable targets for inhibitor molecules, as they are commensurate with the sizes of typical small molecule drug structures^{143,144,146}.

In the past, some of the pharmaceutical industries avoided inhibiting PPIs projects, concerning about technological difficulties and the immaturity of the area. Generally, inhibiting PPIs is a challenging process compared with the traditional drug targets; the essential hurdle of these targets is the large surface area of the typical protein–protein interface, which makes targeting PPIs

through a small molecule hit compound is a difficult process; it could be solved through increasing the size of the hit molecules in order to cover the target surface area of the proteins; however, it is at the cost of pharmacokinetic properties¹⁴³.

Despite the difficulties of targeting PPIs, recently several factors have encouraged pharmaceutical industries to get involved and invest intensively in project related to PPIs and their inhibitors. The first is advances in the recognition and crystallisation of interacting proteins, thus revealing and defining hot spots and druggable pockets that biophysicists, biologists and chemists can evolve novel hits against¹⁴⁷. In addition, there are documented successes with emerging peptide-based inhibitors. Also, the genomics and proteomics revolution has resulted in the identification of a huge number of new targets. Finally, the growth in computational methods and combinatorial synthesis approaches have supplied new tools to apply to this type of project¹⁴⁸.

Without doubt, as stated in Section 1.6, the major factors in increasing drug discovery costs and failure rates are the problems of efficacy, and toxicity of the candidate molecules due to off-target issues with the small molecules. To overcome these obstacles, larger and more target-specific molecules may require, which are either biologics or peptides¹³¹. In the specific case of PPI inhibitors, likewise most of the hurdles are linked to small molecules because the protein interface may require a considerably larger ligand with more binding interactions to cover the interface. Small molecules typically can not

cover the flat and extended surface area of the “undruggable” protein interface area because they are characterised by restricted size, physicochemical properties and surface area ¹⁴⁹. Given this, the contrasting properties of peptide-based molecules make them an attractive choice for the targeting of protein interfaces.

1.11 Aims and Objectives

This project aims to find novel peptide-like molecules, analogues of a key interacting region of TIN2, that can compete effectively for the binding sites for TIN2 on TRF1 and so lead to the destabilisation of telomere structure; along the way, a subsidiary aim is the investigation and optimisation of the accuracy of $\Delta G_{\text{binding}}$ predictions, made using the MM-GBSA molecular modelling method for the TIN2 peptide analogues. The aims were achieved through the following objectives:

1. Design of a library of peptides, analogues of the TRF1-binding motif in TIN2, using the X-ray crystal structure of the TRF1–TIN2 complex, and prediction of their likely activity as competitive inhibitors of the TRF1–TIN2 PPI through the calculation $\Delta G_{\text{binding}}$ for the peptide analogues.
2. Extensive investigation and optimisation of molecular modelling strategies to predict the $\Delta G_{\text{binding}}$ values of the TIN2 peptide analogues.
3. Chemical synthesis of the TIN2 peptide and selected peptide analogues using solid-phase peptide synthesis methods and preparation of them for experimental tests.
4. Biophysical evaluation of the peptide analogues as inhibitors of TRF1–TIN2 interaction (IC_{50} values), through the development and application of a Fluorescent Polarisation (FP) assay.

5. Correlation of $\Delta G_{\text{binding}}$ results (before and after optimisations) with IC_{50} values to confirm the validity and accuracy of the computational methods and results.

Detailed background description for each method is provided at the start of each chapter.

2. The Application of Molecular Modelling to Ligand Design

2.1 Introduction

Molecular modelling is a term associated with molecular structure and can be defined as a computational technique to simplify the process of mathematical calculation. It is crucial to model, explain, recognise and evaluate the behaviour of molecules or molecular systems, thus significantly facilitating teaching and research¹⁵⁰. Generally, molecular models can show molecules as three-dimensional structures and permit molecular manipulations, not only of small molecules, but also of large biomolecules. Calculations in molecular modelling are used to predict molecular energy, vibrational frequencies and the behaviour of the molecule in the presence of other molecules¹⁵¹.

Molecular modelling calculations can be performed by either quantum mechanical or molecular mechanical methods. Quantum mechanics (QM) approaches are extensively used due to the relative accuracy with which they represent structural and thermodynamic changes¹⁵². QM energy calculations rely on the Born-Oppenheimer approximation, when nuclear motions are separated from the electronic motions and assume that the nuclei do not move because nuclei motion is significantly slower compared with the speed of electrons. This approximation is to simplify the Schrodinger equation; therefore, it is known as “electronic Schrodinger equation”¹⁵³. This method is

very computationally expensive and is usually limited to small systems with hundreds of atoms¹⁵⁴. Molecular mechanics (MM) calculations, on the other hand, rely on a simple model of interactions relative to QM, and encompass four key forces or components, which are bond stretching, angle opening and closing, bond rotations and non-bonded interactions, such as van der Waals forces and electrostatic interactions (Equation 2.5). This approach uses force field parameters to model molecular systems, and typically considers an individual atom as a point charge with an associated mass while neglecting the electronic motions. MM calculations are applicable to small as well as large biological systems¹⁵¹.

2.2 Molecular Dynamic Simulation

Investigating molecular and atomic motions and interactions, specifically in biomolecules, can be achieved by a Molecular Dynamics (MD) simulation technique. MD simulations, a rapidly developing field with diverse applications such as providing insights into natural dynamics on different time scales of biomolecules in solution, can provide the thermal averages of molecular properties and explores the thermally accessible conformations of the molecule or complex (Figure 2.1).

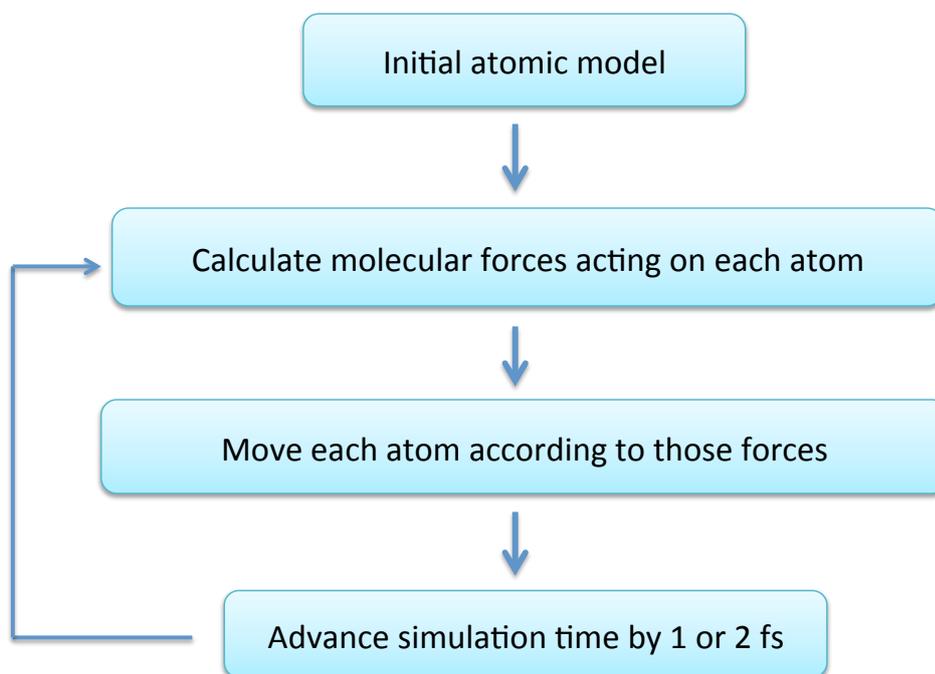


Figure 2.1 A scheme representing the process of molecular dynamic simulations⁹⁷.

This technique can be applied to the exploration of the conformational spaces and dynamical evolution of the molecular systems¹⁵⁵ using Newtonian dynamics (Newton's second law of motion) to simulate the movement of interacting atoms and molecules (Equation 2.1).

$$F = ma$$

Equation 2.1

where F is the external force on the particle, m is the mass of the particle and a is the acceleration of the particle. Before MD simulations can be initiated, energy minimisation of the initially-constructed molecular model is frequently required to obtain a molecular geometry with low potential energy, thus protecting the molecule from fracture due to large initial forces in the system

during MD simulation ¹⁵⁶. Two important issues with MD simulations are the sometimes uncertain reliability of the force field and the high computational demands for execution of the extensive calculations ⁹⁷.

2.3 *In silico* Calculations of Binding Free Energy

The application of molecular modelling in the field of drug design mostly involves the realisation of 3D structures and the calculation of attraction and repulsion forces between ligands and receptors or within biological systems. Typically, the calculation of $\Delta G_{\text{binding}}$ value is a valuable tool across different areas of computational biochemical research, for example, in protein structure determination and drug design processes through the estimation of binding affinities of drug-candidate compounds ¹⁵⁷. Many computational approaches are available to estimate free energies, ranging from rapid but relatively inaccurate techniques such as linear interaction energy analysis (LIE) ¹⁵⁸ and Molecular Mechanics-Poisson Boltzmann Surface Area/Generalized Born Surface Area (MM-PBSA/GBSA) calculations ¹⁵⁹, to more accurate but extremely slow methods including free-energy perturbation (FEP) ^{160,161}, umbrella sampling (US) ¹⁶² and thermodynamic integration (TI) ^{163,164}. These methods can analyse trajectories produced by MD or Monte Carlo (MC) simulations to calculate free energies.

2.4 MM-PBSA/GBSA Method

The MM-PBSA method is the well-known end-point implicit free energy calculation approach, which was applied for the first time in 1998¹⁶⁵. It is a consolidation of molecular and continuum solvent models¹⁶⁶ that has been used widely in many drug discovery projects for calculating the ligand–receptor binding free energy, determining the structural stability and evaluating the docking poses¹⁵⁷. It is a particularly popular method for evaluating differences in free energy between the bound and unbound states of solvated molecules, or for comparing the free energy of the same solvated molecule in different conformations¹⁶⁷.

This method is of increasing interest to the researchers in drug discovery projects for estimating relative binding free energies, due to its acceptable accuracy compared to experimental data¹⁶⁸, broad applicability to diverse systems, efficient computation and calculation of free energy by a more refined mechanism when compared to common scoring functions¹⁵⁹. However, diverse limitations are a feature of this computational method; the first is force field accuracy and partial charge models¹⁶⁹ followed by inadequacy of sampling within a large conformational space¹⁷⁰. The sampling problem can be exacerbated if it traps systems in a local minima for a prolonged period of simulated time¹⁷¹. Ultimately, the limitations can affect the accuracy of MM-PBSA/GBSA results.

The following equation (Equation 2.2) can be used to illustrate the binding free energy of a protein–ligand complex applying MM-PBSA/GBSA methods:

$$\Delta G_{\text{binding}} = G_{\text{com}} - G_{\text{rec}} - G_{\text{lig}} \quad \text{Equation 2.2}$$

where $\Delta G_{\text{binding}}$ is the binding free energy of protein–ligand, G_{com} is complex energy, G_{rec} is receptor energy and G_{lig} is ligand energy^{172,173}.

The free energy of each G from Equation 2.2 is predicted by Equation 2.3:

$$G = H - TS \quad \text{Equation 2.3}$$

where H denotes enthalpy, T is the absolute temperature and S is the entropy of the molecule. Equation 2.4 explains H as follows:

$$H = E_{\text{MM}} + G_{\text{solv}} \quad \text{Equation 2.4}$$

where E_{MM} represents the molecular mechanics energy of the molecule and G_{solv} is the free energy of solvation in MM-PBSA method. Both Equations 2.5 and 2.6 clarify E_{MM} and G_{solv} separately:

$$E_{\text{MM}} = E_{\text{internal}} + E_{\text{electrostatic}} + E_{\text{vdw}} \quad \text{Equation 2.5}$$

where E_{internal} is the summation of all internal energies, which are bond, angle and dihedral energies. $E_{\text{electrostatic}}$ and E_{vdw} are used respectively to represent the electrostatic and van der Waals interaction energies ¹⁷².

$$G_{\text{solv}} = G_{\text{PB/GB}} + G_{\text{SA}} \quad \text{Equation 2.6}$$

where G_{solv} is the summation of the polar ($G_{\text{PB/GB}}$, electrostatic) and nonpolar (G_{SA} , nonelectrostatic) components, $G_{\text{PB/GB}}$ is the polar contribution of solvation energy of the molecule, and G_{SA} is the non-polar solvation free energy. Poisson-Boltzmann (PB) and Generalized Born (GB) models are implicit solvation models used to compute the polar contribution of solvation ¹⁶⁹.

As illustrated in Figure 2.2 and Equation 2.7, $\Delta G_{\text{binding}}$ of the molecular systems in the MM-PBSA/GBSA approach are calculated through the summation of their gas-phase energy, G_{solv} and a configurational entropy of the solute ($-T\Delta S$) ¹⁷⁴. The gas-phase energy is the MM energy of the molecules (E_{MM}). Furthermore, as shown in Equation 2.6, G_{solv} is the polar and nonpolar contribution of the solvation free energy; the polar contribution ($G_{\text{PB/GB}}$) is dependent on the transfer of a charged molecules from a homogenous medium (gas-phase) with dielectric constant = 1 to the solvent with dielectric constant = 78–80, which is estimated by the implicit solvent models ¹⁷⁵. The PB equation was initially used (and still is still widely applied) to calculate the polar contribution energy of G_{solv} ¹⁷⁶; however, an alternative implicit solvent model is mostly used at present, based on the GB theory and a computationally more

efficient approximation compared with PB approximation^{177,178}. On the other hand, the nonpolar solvation free energy (G_{SA}) is the energy required to form a cavity for the solute by the solvent and is proportional to the surface area or solvent accessible surface area (SASA) of the solute. The final term is the configurational entropy (S) of the solute in the gas-phase, which is calculated using either quasi-harmonic¹⁷⁹ or normal mode analysis¹⁸⁰. In conclusion, the $\Delta G_{\text{binding}}$ value in the implicit solvent is equivalent to the summation of E_{MM} in the gas-phase, the difference of G_{solv} between the complex and the nonbonded molecules and the configurational entropy related to the complex formation in the gas-phase at a specific temperature ($T\Delta S$).

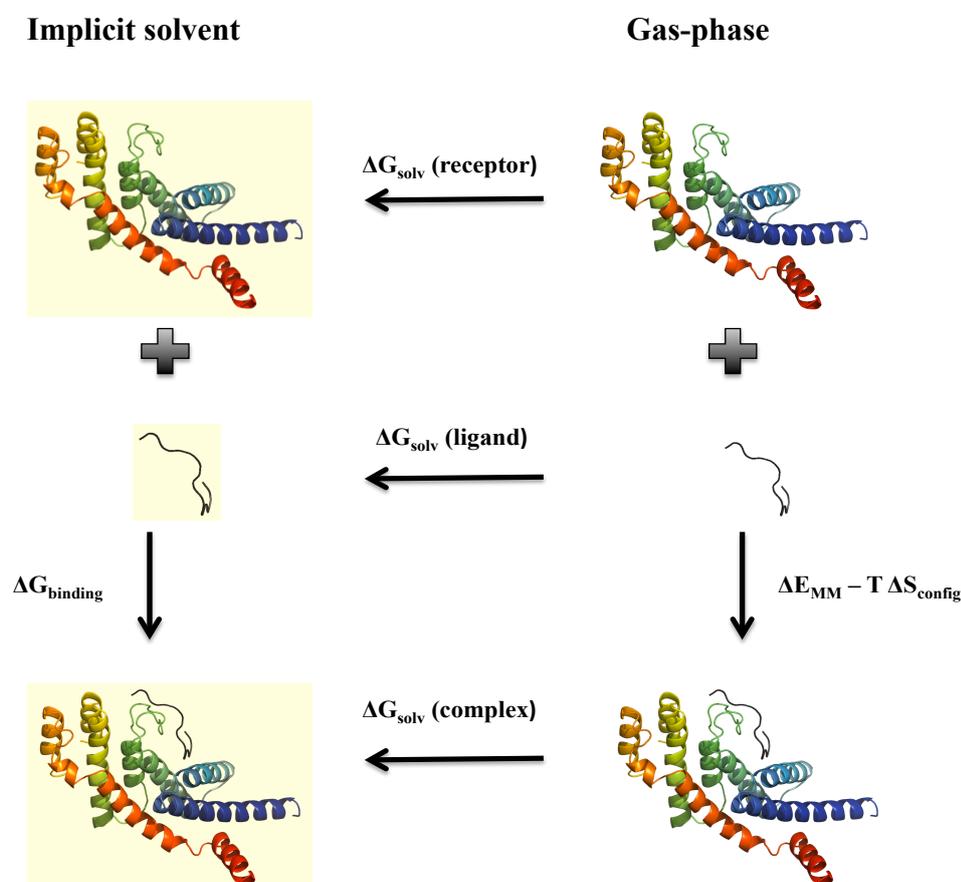


Figure 2.2 Calculating binding free energy of the receptor–ligand complex, using MM-PBSA/GBSA method¹⁵⁷.

$$\Delta G_{\text{binding}} = \Delta E_{\text{MM}} + (\Delta G_{\text{solv}} \text{ complex} - \Delta G_{\text{solv}} \text{ receptor} - \Delta G_{\text{solv}} \text{ ligand}) - T\Delta S$$

Equation 2.7

Calculation of MM-PBSA/GBSA between a receptor and a ligand can be achieved using either multiple or single trajectory approaches. In terms of multiple (separate) trajectory approach, MD simulations are carried out for unbound receptor, unbound ligand and the complex, separately; then, the energy terms are calculated from the snapshots of the individual trajectories¹⁸¹. On the other hand, in the single trajectory approach, MD simulations are performed only for the complex system, after which snapshots of the single trajectory are taken to create conformations of the receptor and ligand and calculate their free energies^{182,183}. In the single trajectory approach, cancelling the E_{internal} of the ligand, receptor and complex has a significant effect on decreasing the noise^{181,184}. Using the single trajectory approach is more common compared with the separate trajectory approach because it requires less effort and it is not as expensive as the separate approach. In addition, it can provide relatively accurate results for the greatest applications of protein–ligand interactions¹⁷⁰. However, in such cases, the single trajectory approach results are less accurate than the separate trajectory approach results owing to neglecting the explicit structural relaxation of the protein and ligand before binding¹⁸⁵.

The starting structure for the calculations is usually a crystal or minimised structure; the conformational ensembles are compared to the reference structure in order to calculate the energy of the conformers and average the

output. Therefore, conformational changes in the system components determine the final energy¹⁵⁷.

2.5 The Balance Between Enthalpy and Entropy

As explained in Section 2.4 (above), the energy of the molecule (G) can be defined as a summation of enthalpy (H) and entropy (S) at a constant temperature (T); hence, $\Delta G_{\text{binding}}$ is a function of changing enthalpy (ΔH) and entropy (ΔS) and is the determinant of the binding affinity between a receptor and a ligand. Therefore, the potency of lead compounds can be improved through favourable improvement of the enthalpy and entropy¹⁸⁶ (Equation 2.8).

$$\Delta G_{\text{binding}} = \Delta H - T\Delta S \quad \text{Equation 2.8}$$

In early drug discovery process and medicinal chemistry, to increase the binding affinity of ligands, the functional groups are typically modified to establish more and stronger favourable binding interactions between the ligand and the receptor (increase $-\Delta H$); however, the greatest modifications could not improve the potency of the ligand because the enthalpic gain may be cancelled totally or further inhibit a favourable $\Delta G_{\text{binding}}$ value due to an unfavourable entropy (decrease $-T\Delta S$). The causes of unfavourable entropy are conformational entropy loss and/or inhibiting desolvation¹⁸⁷. Therefore, a combination of favourable enthalpy and entropy can produce a highly potent lead compound¹⁸⁸.

2.6 Molecular Recognition in the TRF1–TIN2 and TRF2–Apollo Crystal Structures

X-ray crystallography, the most important technique for the determination of protein structure, provides atomic-level clarification of biological structure and function across multiple fields of research. This valuable tool can be applied in different areas such as SBDD, the elucidation of enzyme mechanisms, the design and interpretation of site directed mutagenesis studies and the elucidation of the specificity of protein–ligand interactions ¹⁸⁹.

In terms of the sizes of the different proteins investigated in this study, mature TRF1 and TRF2 proteins are comprised of 438 and 542 residues, respectively. Furthermore, the full chain of the TIN2 protein is 450 amino acids in length, while the Apollo consists of 532 residues ¹⁹⁰. Nevertheless, crystallising the full-length structures of these proteins is difficult as they include structural and functional domains which are flexibly linked together ¹⁹¹. Hence, the crystal structure of the TRF1–TIN2 complex (PDB code 3BQO) was determined at 2 Å resolution and composed of residues 62–268 and 256–268 for the TRF1 and TIN2 respectively; the crystal structure of TRF2–Apollo (PDB code 3BUA) was determined at 2.5 Å resolution and shown to include only residues at position 44–245 of TRF2 and 499–510 of the Apollo peptide.

As illustrated in Figure 2.3, both TRF1 and TRF2 proteins form a homodimer in solution, with each dimer comprised of two monomers in an antiparallel arrangement, thus forming a symmetrical complex whose overall structure

resembles a twisted horseshoe¹⁹². The proteins share virtually the same architecture of the C-terminal Myb/SANT domains. Moreover, the TRFH domains of both proteins are highly superimposable; they produce almost identical 3D structures, which are essential for binding telomeric DNA^{192,193}.

The TRF1 and TRF2 monomers are structurally composed of nine α helices in an elongated helix bundle formation. The helices are divided into two sections, the first of which includes α helices 1, 2 and 9, arranged into what is known as the dimer interface. The second section includes α helices 3 and 8, forming the ‘arms’ of the horseshoe. The remaining helices 4, 5, 6 and 7 are located adjacent to helices 3 and 8. Extended protein–protein interaction surfaces are provided by a six- α helix bundle dimerisation interface (helices 1, 2 and 9 from each monomer) (Figure 2.3B). In addition to dimerisation, TRF1 and TRF2 proteins also bind with their partner proteins through their TRFH domains (Section 1.3.1). Despite the observation that the dimerisation domains of both shelterin proteins feature a conserved common architecture, several different residues are located at the surfaces, thus enabling the recognition of different telomere-associated proteins and preventing the formation of TRF1/TRF2 heterodimers¹⁹⁴.

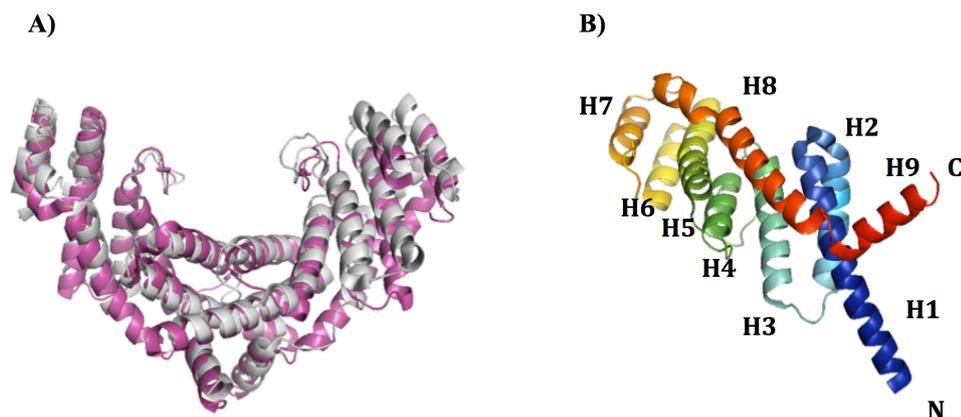


Figure 2.3 Structure of TRF1 and TRF2 dimerisation domains and TRF1 helices. A ribbon representation is used to display structures. (A) Superposition of TRF1 and TRF2 dimerisation domain. The grey colour is characterising TRF1 and the pink is depicting TRF2 protein. (B) The TRFH monomer structure of TRF1. Number and colour are used to mark all of the α helices (adapted from Fairall *et al.*, 2001)¹⁹⁴.

Regarding the interactions between the TRF1 and TIN2 proteins, the C-terminus of the TIN2 protein, termed the TIN2-TRFH binding motif (TBM) or TIN2_{TBM}, is recognised by TRF1_{TRFH} domains¹⁹⁵. TRF1_{TRFH} exists *in vivo* as a homodimer, with each TRF1_{TRFH} protein interacting with one TIN2_{TBM} peptide.

As demonstrated by Chen *et al.* (2008) and illustrated in Figure 2.4 and 2.5, the crystal structure of the TRF1–TIN2 complex reveals the interactions of amino acids between TRF1_{TRFH} and TIN2_{TBM}. The electron density map demonstrates that residues 257 to 268 of TIN2_{TBM} assume a well-defined conformation. The TIN2_{TBM} peptide is composed of 13 residues (S256-H-F-N-L-A-P-L-G-R-R-R-V268); it interacts with TRF1_{TRFH} through its C and N-termini in unison. The

C-terminus of TIN2_{TBM} consists of L263-G-R-R-R-V268, forming an antiparallel β sheet on the surface of loop L34. Hydrogen bond (H-bonds) can occur between Q141 of TRF1 and both L263 and G264 of TIN2. Furthermore, R265 of TIN2 interacts with D139 of TRF1 and R266 of TIN2 peptide interacts with L138, D139, R147 and E192 through a combination of H-bonds and salt bridge interactions. Lastly, R267 of TIN2 attracts the E146 side chain of TRF1. In contrast, the *N*-terminus of TIN2_{TBM}, which is composed of H257-F-N-L-A-P262 and has an extended conformation, stabilised by intermolecular hydrogen bonding and hydrophobic contact with TRF1_{TRFH} residues¹⁹².

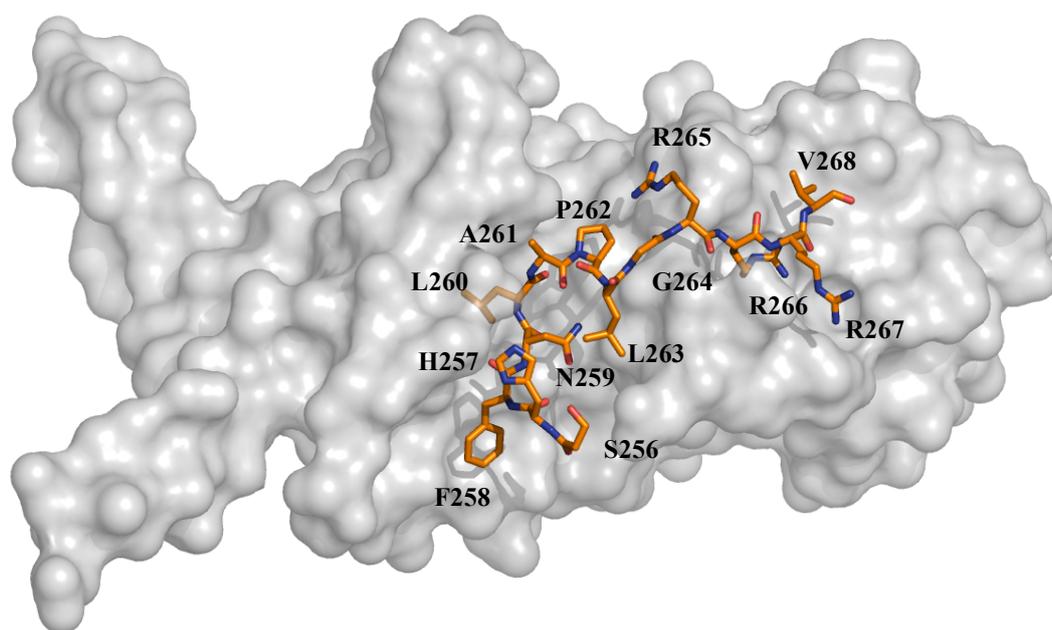


Figure 2.4 The TRF1–TIN2 crystal structure. The grey color is TRF1 and the brown sequence is TIN2.

The hydrophobic groove of TRF1_{TRFH}, consisting of I109, L115, L120, I123 and Y124, recruits F258 of TIN2_{TBM}. In addition, N259 undergoes an

intermolecular hydrophilic interaction with F142 and an intramolecular hydrophilic interaction with A261. The oxygen atom of the N259 main-chain attracts both Q127 and R131 of TRF1. Furthermore, L260 side chain inserts the concave hydrophobic pocket of TRF1_{TRFH} and its backbone N-H group has a H-bond interaction with the E106 carboxylic acid side chain. The next residue of TIN2 is A261; it undergoes only a H-bond interaction through its backbone N-H group with the side chain carboxylic acid group of E106 (Figure 2.5)¹⁹². The last residue of the *N*-terminus is P262; according to the crystal structure analysis, it has not any contact with the TRF1 protein.

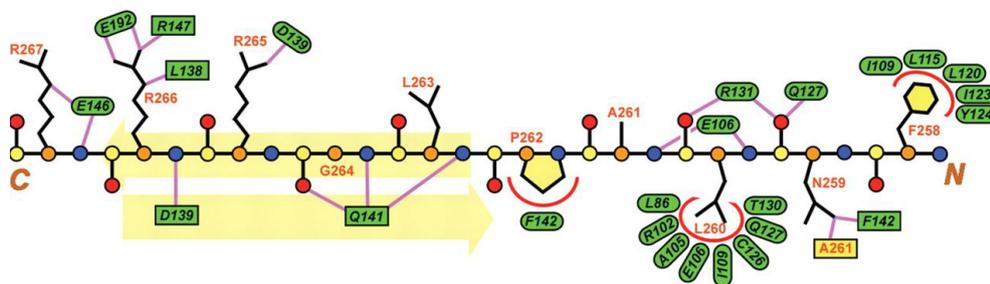


Figure 2.5 Interactions between TRF1_{TRFH} protein and TIN2_{TBM} peptide. Green ovals and square boxes symbolise the side chain and main-chain portions of TRF1_{TRFH} amino acids, respectively. Circles denote the primary chain atoms of TIN2_{TBM}: blue circles represent nitrogen atoms, orange circles represent α carbon atoms, β carbon atoms are yellow and oxygen atoms are represented by red circles. Straight magenta lines represent hydrophilic interactions, curved red lines show hydrophobic contacts and pale yellow arrows denote intermolecular β sheets (taken from Chen *et al.*, 2008)

As stated in Section 1.3, one of the telomere-associated proteins is known as Apollo¹⁹⁶; it is recruited by TRF2 and contributes to the protection of chromosomes¹⁹⁷. The region of TRF2 responsible for attracting Apollo protein is called TRF2_{TRFH} and the TRF2 interacting segment of Apollo is termed the Apollo–TRFH binding motif (Apollo_{TBM}), consisting of amino acids 498–509. The C-terminal region of Apollo_{TBM} (Y504-L-L-T-P-V509) is essential for interacting with TRF2_{TRFH} (Figure 2.6)¹⁹². Structural similarities exist between the TRF1_{TRFH}-TIN2_{TBM} and the TRF2_{TRFH}-Apollo_{TBM} complexes, in that both segments share an identical conformation and preserve virtually the same hydrogen bonds. Furthermore, the TBMs of TIN2 and Apollo share the sequence Y/F-X-L-X-P (where X is a non-conserved amino acid). However, several differences exist between TIN2_{TBM} and Apollo_{TBM}; the highly interacting region of TIN2_{TBM} localises at the *N*-terminal, but is located at the C-terminal of Apollo_{TBM}¹⁹. Additionally, the C-terminus of TIN2_{TBM} consists of a polyarginine residue extension and creates a β sheet, while the *N*-terminus of Apollo_{TBM} forms a short helix and is an extension of six amino acids (R498-G499-L500-A501-L502-K503). Finally, to allow efficient binding with TRF2_{TRFH}, a more hydrophilic amino acid such as Tyr is preferred in the Y/F-X-L-X-P motif of Apollo_{TBM}, whereas, for binding with TRF1_{TRFH}, a more hydrophobic residue such as Phe is preferred in the Y/F-X-L-X-P motif of TIN2_{TBM} (Figures 2.5 and 2.6)¹⁹².

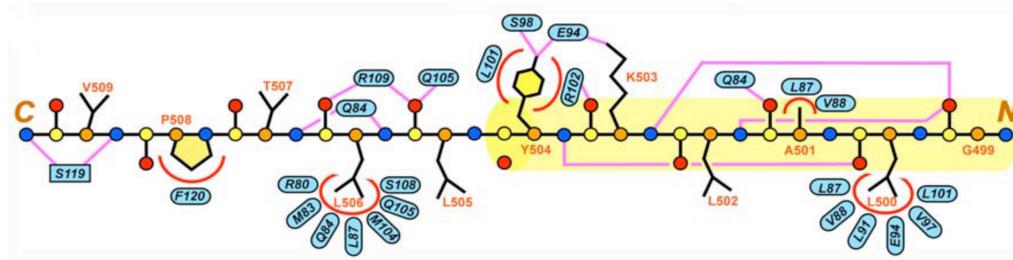


Figure 2.6 A scheme denoting TRF2_{TRFH}-Apollo_{TBM} interactions. All symbols and colours are as for **Error! Reference source not found.**, except for the yellow lines denoting the *N*-terminal helix of Apollo (taken from Chen *et al.*, 2008)¹⁹².

2.7 *In silico* Design of TIN2 Peptide Inhibitors

Specific amino acid residues of TIN2 peptide were computationally mutated in an attempt to identify the optimal sequence of a synthetic peptide that might function as a competitive inhibitor of the TRF1–TIN2 protein–protein interaction. Two different strategies were used to obtain a novel TIN2 peptide analogue with a significantly higher binding free energy and potential as an early lead compound.

2.7.1 TIN2–Apollo Cross-Matching Mutations

The first strategy used to design peptide analogues was the mutation of selected TIN2 peptide amino acid residues to the corresponding amino acid in the Apollo peptide within the shared sequence (F/Y-X-L-X-P). This approach was selected due to the structural similarities between the host TIN2 and

Apollo molecules and the close resemblance of the two-peptide TBM sequences, especially the *N*-terminal of TIN2 and the *C*-terminal of Apollo. As stated in Section 2.6, the peptides share the same conformation and virtually the same H-bonds (Figures 2.5, 2.6 & 2.7) ¹⁹².

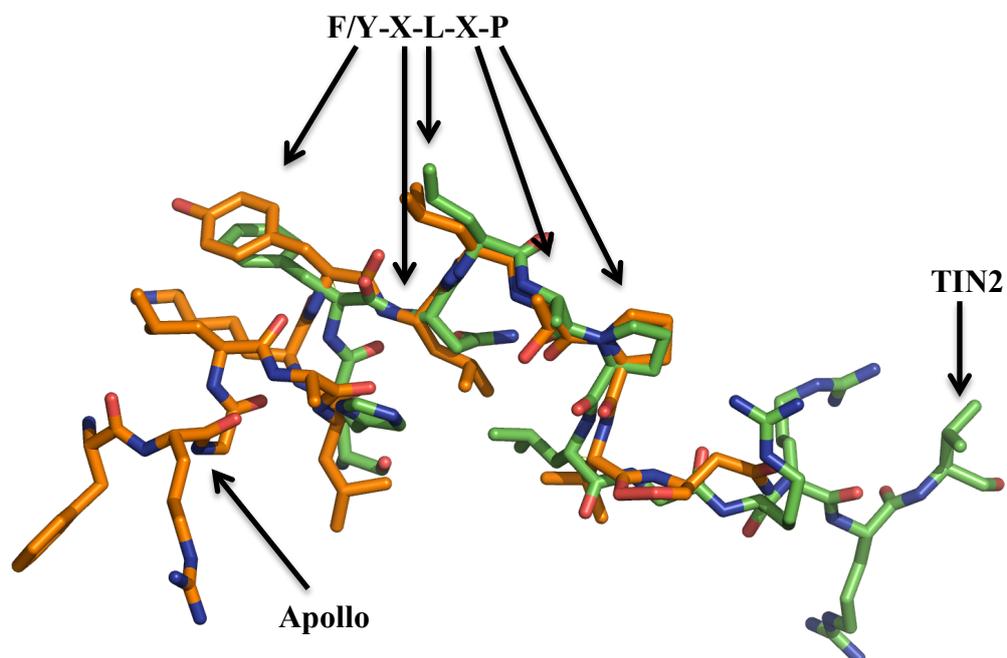


Figure 2.7 Overlapping 3D structures of the TIN2 and Apollo peptides.

As shown in Figure 2.8, three amino acids in the shared sequence of TIN2 peptide differ from their matched Apollo peptide structure and were thus selected for mutation. An additional mutation outside of the shared amino acid residues was selected as a negative control. Hence, the first peptide analogue was a mutation of F258 to Y amino acid (F258Y) in TIN2 peptide; this peptide analogue was identical to the TIN2 structure with a hydroxyl (-OH) group

added to the side chain phenyl ring of F258. The second and third peptide analogues were generated through mutating N259 to L and A261 to T, respectively. R266 to P was selected as a negative control as the crystal structure¹⁹² demonstrates the critical role of R266 in TIN2 interactions with TRF1. All four TIN2 peptide analogues and TIN2 peptide were named in order to facilitate their definition (Table 2-1).

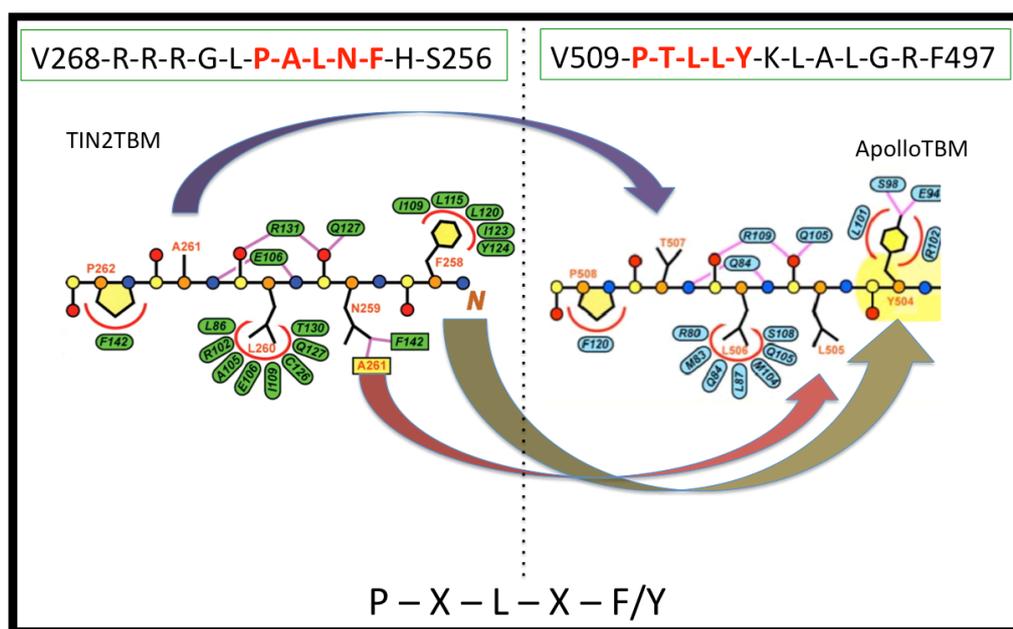


Figure 2.8 The shared sequence of the C-terminal of Apollo peptide and the N-terminal of TIN2 peptide. Arrows denote the amino acid residues of TIN2 selected for mutation to the corresponding amino acids in the Apollo peptide.

2.7.2 Mutation of Selected TIN2 Residues Based on the Crystal Structure of TRF1–TIN2

The second strategy for the design of TIN2 peptide analogues utilised SBDD¹⁹⁸. The structures of TRF1_{TRFH} and the TRF1–TIN2 interactions are well characterised through crystal structure records¹⁹⁴ (Figures 2.5 & 2.9)¹⁹². Through careful analysis of these structures, mutations of TIN2 were predicted that might increase binding interactions between the peptides and the TRF1 protein.

Mutations of the TIN2 residues were selected based on their role in interactions with TRF1. One of the determining factors for recognising the role of each residue was the per-residue decomposition result (Figure 2.10), which shows the contribution of each residue in binding with TRF1. According to these results, S256 has an unfavourable effect and inhibit the binding free energy value ($\Delta G_{\text{binding}} = 8.56$ kcal/mol); in addition, both H257 and V268 have the lowest binding free energy values (Figure 2.10). Therefore, these three residues were deleted or modified to an alternative amino acid residues in most of the designed peptide analogues to improve favourable binding free energy and decrease molecular weight of the peptide analogues.

Mutation of the TIN2 residues was not only dependent on the per-residue result, but also on the 3D visualisation of the TRF1–TIN2 crystal structure and the equilibrated MD trajectories as this can show the conformation of each residue and its distance from the nearby TRF1 residue functional groups;

hence, an expected better amino acid alternatives were selected to improve $\Delta G_{\text{binding}}$. For example, per-residue decomposition results show the unfavourable effect of S256 residue on $\Delta G_{\text{binding}}$ and, according to the MD simulations, the S256 side-chain cannot form interactions with TRF1 residues. S256 was therefore truncated in most of the peptide analogues or mutated to Q256 to have a longer side-chain for creation of a H-bond with the N144 side-chain of TRF1 because, according to the MD simulations, the N144 side-chain amide group distance from the side chain hydroxyl group of S256 is approximately 6 Å, which cannot form a stable H-bond.

As shown in Table 2.2, A261 and L263 of the TIN2 peptide were mutated in most of the peptide analogues, while the per-residue decomposition data shows that the $\Delta G_{\text{binding}}$ contribution of A261 is -2.17 kcal/mol and -3.81 kcal/mol for L263. The purpose of these mutations was that the side chains do not participate in the binding interactions with the TRF1 residues. Therefore, mutations were performed to such residues that may form favourable binding interactions with the TRF1 residues. Another mutation, which was achieved in 20 of the peptide analogues, is F258Y; this was performed to preserve the hydrophobic contact between the Y258 side-chain and the residues of TRF1 and to create a H-bond between the side-chain hydroxyl groups of Y258 and Y124.

Despite the previously mentioned factors used for selecting the TIN2 mutations, data from the literature was also considered. For instance, according to the crystal structure and MD trajectories, P262 cannot form a favourable

binding interaction with the TRF1 residues, although all proteins that bind with TRF1 through the TRFH domain have a preserved P262 residue²¹. Therefore, this residue was also preserved in the peptide analogues in this project, except in one (P03) which was mutated (P262F) to form a π - π stacking between F262 and F142 side chain of TIN2 and TRF1, respectively; however, the result was the inhibition of binding free energy. Thus, a library of peptide analogues was designed, comprised of 49 analogues of TIN2 peptide; $\Delta G_{\text{binding}}$ energies were subsequently predicted for each peptide analogue bound to TRF1 protein (Table 2-2).

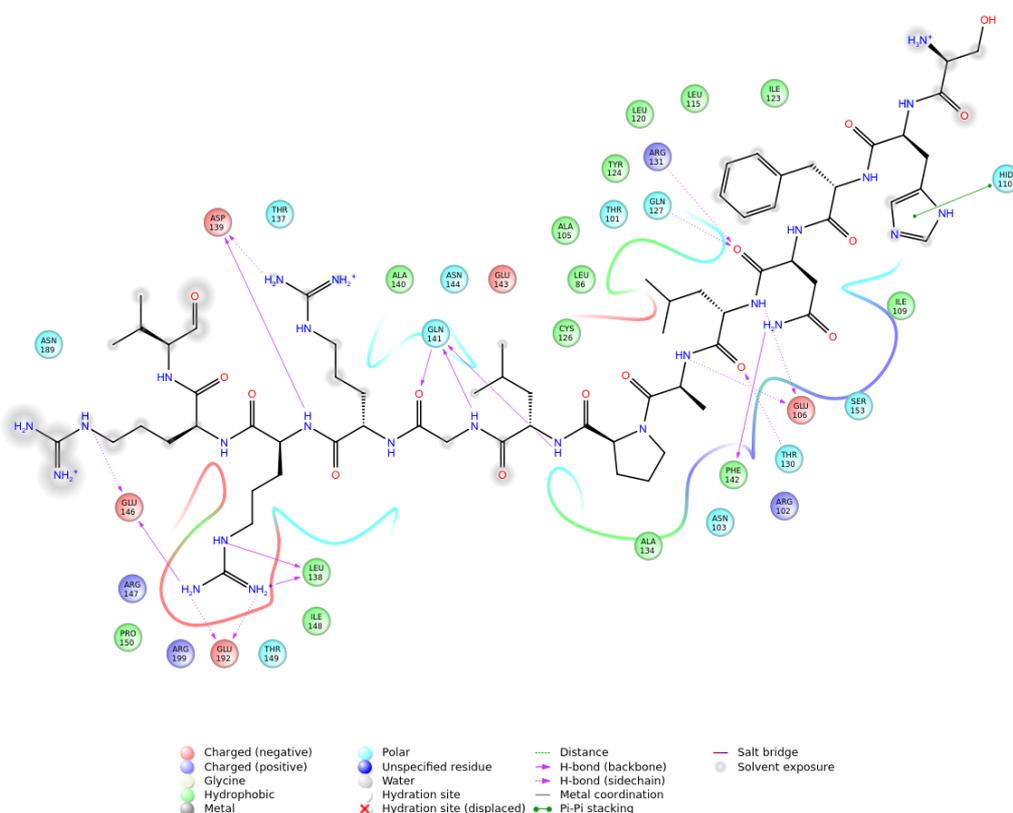


Figure 2.9 The interactions of TIN2_{TBM} peptide with TRF1_{TRFH} protein as determined from the crystal structure.

2.8 MD Simulations of the Protein–Peptide Complexes

As described above, through mutation of selected amino acid residues of TIN2 peptide from the crystal structure of the TRF1–TIN2 complex (3BQO), novel analogues were designed. Each TRF1–TIN2 analogue was then parameterised and prepared for MD simulation by producing topology and coordinate files; subsequently, the complex systems were minimised and 10 nanoseconds (ns) MD simulations were performed to produce trajectory files. Finally, 500 snapshots of the trajectory files were used to determine $\Delta G_{\text{binding}}$ energy and investigate the molecular motions and conformations of the peptide analogues (See Section 6.1).

2.9 $\Delta G_{\text{binding}}$ Prediction of TRF1–TIN2 Peptide

Analogues

Binding free energies were calculated for all complex systems using the MM-GBSA method and the single trajectory approach was applied for the calculations. The aim of $\Delta G_{\text{binding}}$ calculations in this study was to determine and compare the relative binding affinity between TRF1 protein and TIN2 peptide analogues. As the relative binding free energies of similar ligands are required in this project, the entropy term was disregarded in order to decrease the noise and error associated with entropy changes^{157,199}.

The reason of selecting MM-GBSA method rather than MM-PBSA was due to that MM-GBSA could display a competitive computational efficiency²⁰⁰ and better performance in predicting relative binding free energy when compared to the MM-PBSA approach¹⁸³. Theoretically, however, the MM-PBSA method demonstrates a higher accuracy in estimating absolute binding free energy¹⁸⁴. Data from the first strategy, the TIN2–Apollo cross matching mutations, are shown in Table 2-1.

Table 2-1 Predicted binding free energies of TIN2 peptide and four analogues derived from cross mutation with the Apollo peptide.

TIN2 peptide mutation	Number of amino acids	Name of TIN2 analogue	Calculated $\Delta G_{\text{binding}}$ (kcal/mol)
TIN2	13	TS01	-75.01 ± 0.49
TIN2–F258Y	13	TS02	-89.83 ± 0.20
TIN2–N259L	13	TS03	-85.42 ± 0.19
TIN2–A261T	13	TS04	-82.55 ± 0.25
TIN2–R266P	13	TS05	-59.49 ± 0.37

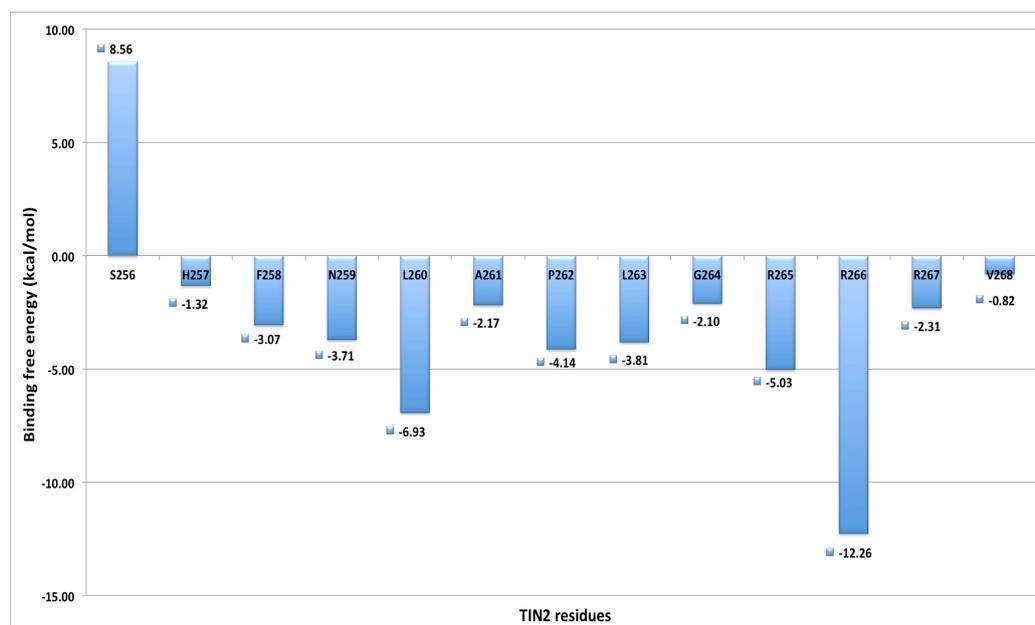


Figure 2.10 Per-residue decomposition of the TRF1–TIN2 complex system, which shows the binding free energy contribution of each TIN2 residue with TRF1.

The second strategy used to design the peptide analogues was based on the protein and the peptide structures. As shown in Table 2-2, $\Delta G_{\text{binding}}$ values for 49 protein-peptide analogue complexes were determined and eight of the peptide analogues were selected according to their competitive $\Delta G_{\text{binding}}$ values (> -90 kcal/mol) because investigation of all analogues was prohibitively expensive. Two estimated negative controls were designed; the estimated negative control used in the first strategy was TS05 (R266P) and the estimated negative control, which designed in the second strategy, was the P25 (L260G–R266P) molecule.

Table 2-2 The $\Delta G_{\text{binding}}$ values of 49 TIN2 peptide analogues. Blue rows represent the selected molecules to be chemically synthesised.

Name	Sequence	Residue No.	$\Delta G_{\text{binding}}$ (kcal/mol)
P01	S256-H-F-N-L-A-P-L-G-R-R-R-V	13	-75.01 \pm 0.49
P02	Q256-H-F-N-L-A-P-L-G-R-R-R-V	13	-79.00 \pm 0.57
P03	S256-H-F-N-L-A-F-L-G-R-R-R-V	13	-71.32 \pm 0.47
P04	S256-H-F-N-L-R-P-L-G-R-R-R-V	13	-102.31 \pm 0.68
P05	S256-H-F-N-L-A-P-L-G-R-R-R-L	13	-92.32 \pm 0.35
P06	H257-F-N-L-A-P-L-G-R-R-R-V	12	-86.90 \pm 0.41
P07	H257-F-N-L-A-P-N-G-R-R-R-V	12	-83.84 \pm 0.80
P08	H257-F-N-L-A-P-L-G-K-R-R-R-V	12	-87.04 \pm 0.61
P09	H257-F-N-L-A-P-N-G-K-R-R-R-V	12	-66.97 \pm 0.85
P10	H257-Y-N-L-A-P-N-G-K-R-R-R-V	12	-67.66 \pm 1.00
P11	H257-F-N-L-A-P-N-G-K-R-S-V	12	-70.52 \pm 0.76
P12	H257-F-N-L-N-P-N-G-K-R-S-V	12	-101.75 \pm 0.53
P13	H257-F-N-I-A-P-L-G-R-R-R-V	12	-84.33 \pm 0.57
P14	H257-F-N-L-N-P-L-G-R-R-R-V	12	-86.01 \pm 0.80
P15	H257-F-N-L-Q-P-L-G-R-R-R-V	12	-83.47 \pm 0.55
P16	H257-Y-N-L-N-P-N-G-K-R-S-V	12	-91.3 \pm 0.70
P17	H257-Y-N-I-N-P-N-G-K-R-S-V	12	-70.26 \pm 0.66
P18	H257-F-N-L-A-P-L-G-K-R-S-V	12	-67.14 \pm 0.70
P19	H257-F-N-L-A-P-L-G-R-R-S-V	12	-83.4 \pm 0.69
P20	H257-F-N-L-N-P-L-G-K-R-S-L	12	-90.99 \pm 0.49

P21	H257-F-N-L-N-P-N-G-R-R-R-V	12	-100.12 ± 0.86
P22	H257-F-N-L-N-P-N-G-R-R-S-V	12	-86.32 ± 0.96
P23	H257-F-N-L-N-P-N-G-K-R-S-L	12	-98.40 ± 0.76
P24	H257-Y-N-L-I-P-N-G-G-R-R-L	12	-62.47 ± 0.40
P25	H257-F-N-G-A-P-L-G-R-P-R-V	12	-50.37 ± 0.49
P26	H257-F-N-L-A-P-L-G-R-R-R	11	-81.64 ± 0.35
P27	H257-F-N-L-A-P-N-G-R-R-R	11	-80.01 ± 0.43
P28	H257-F-N-L-A-P-N-G-K-R-R	11	-70.89 ± 0.46
P29	H257-Y-N-L-A-P-N-G-K-R-R	11	-74.04 ± 0.62
P30	H257-F-N-L-N-P-N-G-K-R-S	11	-81.69 ± 0.46
P31	H257-Y-N-L-N-P-N-G-K-R-S	11	-70.58 ± 0.59
P32	H257-F-N-L-N-P-N-G-K-R-K	11	-67.40 ± 0.32
P33	F258-N-L-N-P-N-G-K-R-S-V	11	-77.12 ± 0.48
P34	Y258-N-L-N-P-N-G-K-R-R	10	-80.18 ± 0.48
P35	Y258-N-L-N-P-N-G-K-R-S	10	-80.04 ± 0.43
P36	Y258-N-L-N-P-G-N-K-R-S	10	-58.90 ± 0.50
P37	F258-N-L-N-P-N-G-K-R-S	10	-71.21 ± 0.45
P38	F258-N-L-A-P-L-G-R-R	9	-64.95 ± 0.35
P39	Y258-N-L-A-P-L-G-R-R	9	-55.07 ± 0.37
P40	Y258-N-L-D-P-L-G-R-R	9	-59.82 ± 0.39
P41	Y258-N-L-D-P-N-G-R-R	9	-58.13 ± 0.38
P42	Y258-N-L-D-P-N-N-R-R	9	-59.19 ± 0.46
P43	Y258-N-L-D-P-N-N-K-R	9	-62.40 ± 0.34
P44	Y258-N-L-D-P-N-N-G-R	9	-55.46 ± 0.32

P45	L258-N-L-D-P-N-N-G-R	9	-53.85 ± 0.40
P46	Y258-N-L-N-P-Q-N-G-R	9	-52.54 ± 0.34
P47	F258-N-L-N-P-N-G-K-R	9	-72.80 ± 0.43
P48	Y258-N-L-N-P-N-G-K-R	9	-67.57 ± 0.47
P49	Y258-N-L-Y-P-N-G-K-R	9	-65.66 ± 0.36
P50	Y258-N-L-A-P-N-G-K-R	9	-55.65 ± 0.38

As shown in Table 2-2, P01 is the wild-type (TIN2) peptide, and the subsequent four analogues (P02–P05) were designed with an identical number of TIN2 residue numbers, two of which (P04 and P05) had a significant higher predicted binding free energy relative to the TIN2 peptide. Starting from P06, the analogues designed were truncated TIN2 molecules; residue numbers were decreased in order to improve the likely pharmacokinetic characteristics of the ligand molecules and reduce manufacturing costs¹³³. P06 is a TIN2 peptide, truncated at S256 to produce 12 amino acid residues. The next 19 peptide analogues (P07–P25) are composed of 12 amino acid sequences, only five of which (P12, P16, P20, P21 and P23) produced $\Delta G_{\text{binding}}$ value > -90 kcal/mol. Numbered P26 to P33, the next eight peptides were sequences of 11 amino acid residues, with P26 produced by truncating S256 and V268 in the wild-type and the remaining seven analogues as mutants of P26.

According to the data, none of the analogues were observed to have a binding free energy > -90 kcal/mol. The four peptides denoted by P34 to P37 were 10 residue sequences, truncated at S256, H257 and V268 of TIN2, respectively. Finally, P38-P50 was a set of 9-mer analogues in which the truncated amino

acid residues were S256, H257, R267 and V268. Upon shortening of the amino acid residues to 9-mer, the observed $\Delta G_{\text{binding}}$ values were significantly lowered. Peptide analogues with a significantly higher binding free energy, relative to the wild-type TIN2 peptide ($\Delta G_{\text{binding}} > -90$ kcal/mol), were selected for the chemical synthesis. These peptides are represented by blue rows in the Table 2-2 and are peptide analogues with higher predicted binding free energies relative to P01, except for P25, which was considered a negative control.

In summary, 13 peptides were selected for chemical synthesis. These were TIN2, the four peptide analogues designed by the first mutation strategy (TIN2–Apollo cross-matching mutations) and the eight peptide analogues from the second mutation strategy (P04, P05, P12, P16, P20, P21, P23 and P25) based on their binding free energies. The chosen peptide analogues were named according to the first strategy format, as shown in Table 2-3.

2.10 Replicating MD Simulations

It was initially assumed that single MD simulations for each TRF1-peptide complex would suffice. Following repeated MD simulations of the identical complex (TRF1–TIN2 complex), trajectories were yielded that individually seemed to satisfy the standard equilibration and sampling metrics but differed markedly in predicted MM-GBSA binding affinities (Table 2-3)¹⁷¹. Although, all replicates were simulated from the same starting structure and using the

same protocols (except for the choice of initial velocities), the predicted energies of each replicate were significantly different (Figure 2.11).

Regarding the accuracy of the MM-PBSA approach and the running single and multiple replicates, Sadiq *et al.* (2010), investigated the predicted relative binding free energy between the HIV-1 proteases and their inhibitors, running single MD simulations for 50 ns; after that, 50 replicates of MD simulations were run for 1 ns and 4 ns using MM-PBSA approach for calculating binding free energies. After plotting the predicted and experimental values, the results observed that the long MD simulation correlation coefficient value = 0.62, while the correlation coefficient of the 50 x 1 and 50 x 4 ns simulations = 0.98. These results demonstrated that the ensemble MD simulations could significantly improve the accuracy of the predicted values¹⁷⁰. A more recent study by Marc Adler and Paul Beroza (2013) on the polo-like kinase-2 receptors and their inhibitors reported that, when single MD simulation was achieved for each system; then, multiple MD simulations (10 replicates) were used, the results of the replicate MD trajectories were more accurate than the single instances, as the single MD trajectory coefficient of determination (r^2) was 0.36, while r^2 for the average of 10 replicates was improved to 0.47¹⁶⁷. According to this method, the same starting velocity was used but the coordinates were randomly perturbed by 0.001 Å. MM-PBSA was used as a post-processing method and the predicted results compared with the IC_{50} values.

It is likely that individual, single MD simulations produce inaccurate results due to inadequate sampling of the conformational space¹⁷⁰ and long periods of simulated time spent trapped in local minima. These observations suggest that computational optimisation of $\Delta G_{\text{binding}}$ values through multiple MD simulations could be used to address this limitation (Table 2-3).

Table 2-3 Binding free energies of the single MD simulation and average of 50 replicates MD simulations. Within the sequences, mutated amino acid residues are highlighted in red.

Name	Sequence	Single predicted	Average predicted
		$\Delta G_{\text{binding}}$ (kcal/mol)	$\Delta G_{\text{binding}}$ (kcal/mol)
TS01	S256-H-F-N-L-A-P-L-G- R-R-R- V268	-75.01 \pm 0.49	-85.07 \pm 0.72
TS02	S256-H- Y -N-L-A-P-L-G- R-R-R- V268	-89.83 \pm 0.20	-84.61 \pm 0.81
TS03	S256-H-F- L -L-A-P-L-G- R-R-R- V268	-85.42 \pm 0.19	-82.69 \pm 0.77
TS04	S256-H-F-N-L- T -P-L-G- R-R-R- V268	-82.55 \pm 0.25	-87.98 \pm 0.95
TS05	S256-H-F-N-L-A-P-L-G- R- P -R- V268	-59.49 \pm 0.37	-60.99 \pm 0.77
TS06	H257-F-N- G -A-P-L-G-R- P -R-V268	-50.34 \pm 0.49	-51.42 \pm 0.68
TS07	H257-F-N-L- N -P- N -G- K -	-98.4 \pm 0.76	-81.95 \pm 1.03

	R-S-L268		
TS08	H257-F-N-L-N-P-N-G-R-	-100.12 ± 0.86	-94.94 ± 1.11
	R-R-V268		
TS09	H257-F-N-L-N-P-L-G-K-	-90.99 ± 0.49	-88.92 ± 0.86
	R-S-L268		
TS10	H257-F-N-L-N-P-N-G-K-	-101.75 ± 0.53	-82.96 ± 1.26
	R-S-V268		
TS11	S256-H-F-N-L-R-P-L-G-	-102.31 ± 0.68	-95.49 ± 1.21
	R-R-R-V268		
TS12	H257-Y-N-L-N-P-N-G-K-	-91.30 ± 0.70	-81.99 ± 1.20
	R-S-V268		
TS13	S256-H-F-N-L-A-P-L-G-	-92.32 ± 0.35	-89.00 ± 0.70
	R-R-R- L268		

2.11 Optimising the Accuracy of Predicted Results

Different approaches were implemented in order to improve the accuracy of the predicted $\Delta G_{\text{binding}}$ values. The first was optimising the replicate number of MD simulations and the second was optimising the time required for execution of MD simulations.

2.11.1 Optimising the Replicate Number of MD Simulations

In this section, data from an in-depth analysis of the convergence of such systems are presented. The output from up to 50 replicates and 10 ns simulations of the protein–peptide systems was subjected to a rigorous statistical analysis of the replicate values in order to identify the quantity of replicate runs required. In order to reach a defined level of accuracy and precision²⁰¹, the optimum number of independent simulations should be performed. Determining the sample size is critical, as samples that are too large may waste time, resources and money, while samples that are too small may lead to inaccurate results. Analysis of sample distribution is required to confirm that the assumptions of the parametric test are met prior to its application.

Histograms of binding free energies calculated from the replicate simulations of the protein–peptide complexes do not show a typical bell shape (Figure 2.11) but might still be applicable if it shown to be likely produced from an underlying normal distribution. A variety of tests were performed to test for normality, such as the Kolmogorov-Smirnov, D’Agostino and Pearson omnibus and Shapiro-Wilk tests, with the variables shown to be that normally distributed for the replicates as a consequence (p -value > 0.05). In addition to considering this indicator when determining the number of replicates, the application of standard statistical methods was necessary to estimate the number of replicates required to obtain a value of $\Delta G_{\text{binding}}$ to within defined confidence limits with acceptable probability. A method for determining the

sample size of the TRF1–TIN2 complex and its analogues was achieved using the sample size equation (Equation 2.9):

$$\mathbf{n} = \left[\frac{Z_{\alpha/2} \delta}{E} \right]^2 \quad \text{Equation 2.9}$$

where \mathbf{n} is the minimum sample size, E is margin of error (kcal/mol), $Z_{\alpha/2}$ is a critical value, which is 1.96 in the 95% confidence calculation and δ is the population standard deviation²⁰². The outcomes of using Equation 2.6 to statistically calculate the sample size for the complex systems are detailed in Table 2-4, with the standard deviations taken from Figure 2.11.

Table 2-4 Standard deviation and calculated sample size for the complex systems with different margins of error.

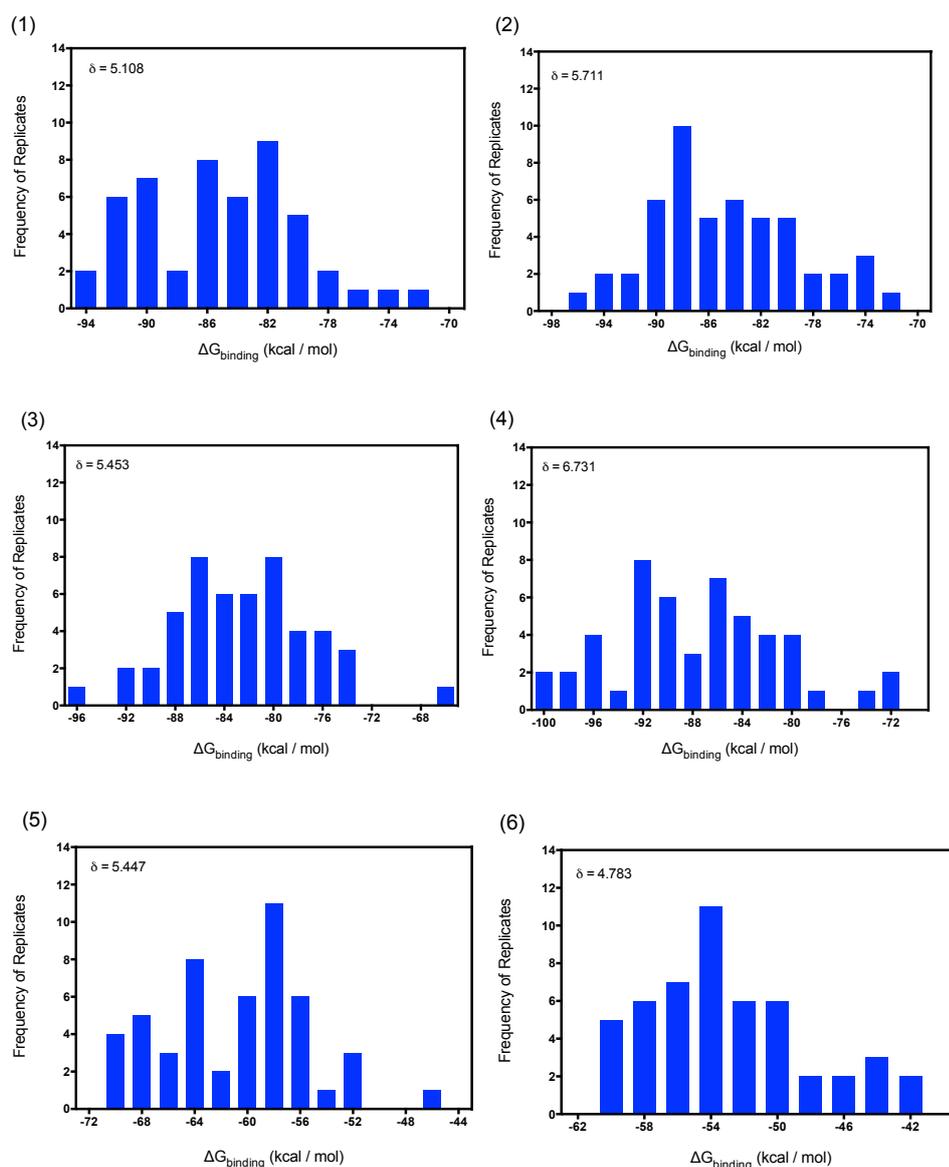
Complex systems	Standard deviation	N (E=1 kcal/mol)	N (E=2 kcal/mol)	N (E=3 kcal/mol)	N (E=4 kcal/mol)
TS01–TRF1	5.108	100	25	11	6
TS02–TRF1	5.711	125	31	14	8
TS03–TRF1	5.453	114	29	13	7
TS04–TRF1	6.731	174	44	19	11
TS05–TRF1	5.447	114	28	13	7
TS06–TRF1	4.783	88	22	10	5
TS07–TRF1	7.315	206	51	23	13
TS08–TRF1	7.868	238	59	26	15

TS09–TRF1	6.072	142	35	16	9
TS10–TRF1	8.899	304	76	34	19
TS11–TRF1	8.579	283	71	31	18
TS12–TRF1	7.919	241	60	27	15
TS13–TRF1	4.924	93	23	10	6

As illustrated in Table 2-4, two factors affect the sample size (i.e., number of replicates). Firstly are different margins of error; for example, calculating the sample size using 1, 2, 3, or 4 kcal/mol as a margin of error resulted in diverse outcomes. In TS01–TRF1 complex, for instance, the minimum sample size was approximately 100 replicates when 1 kcal/mol was selected as the margin of error, with 25 replicates for 2 kcal/mol, 11 replicates for 3 kcal/mol and 6 replicates for 4 kcal/mol margins of error. Secondly, minimum sample sizes differ among the diverse complex systems, as each system has a specific standard deviation, which directly affects the n value. Different standard deviation values relate to the distribution of $\Delta G_{\text{binding}}$ values for different replicates of the system. Therefore, increased standard deviations lead to greater numbers of replicates required.

As shown by Equation 2.9 and Table 2-4, the standard deviation value is directly proportional to the number of replicates. For example, when $E = 1$ kcal/mol, the highest number of replicates was required for TS10–TRF1 complex (304 replicates), as it had the highest standard deviation (8.899). Hence, if all systems applied 1 kcal/mol as the margin of error, 304 replicates would be required for all systems in order to obtain reliable and consistent

results, a highly expensive computational undertaking. Due to the limited computing capacity in this study, 3 kcal/mol was selected as the margin of error, with TS10–TRF1 complex subsequently requiring the highest number of replicates at 34. Thus, the minimum replicate number for each system was 34, and it was presumed that the selection of 50 replicate calculations for each system (a figure significantly above the minimum replicate number) would provide consistency across all complex systems.



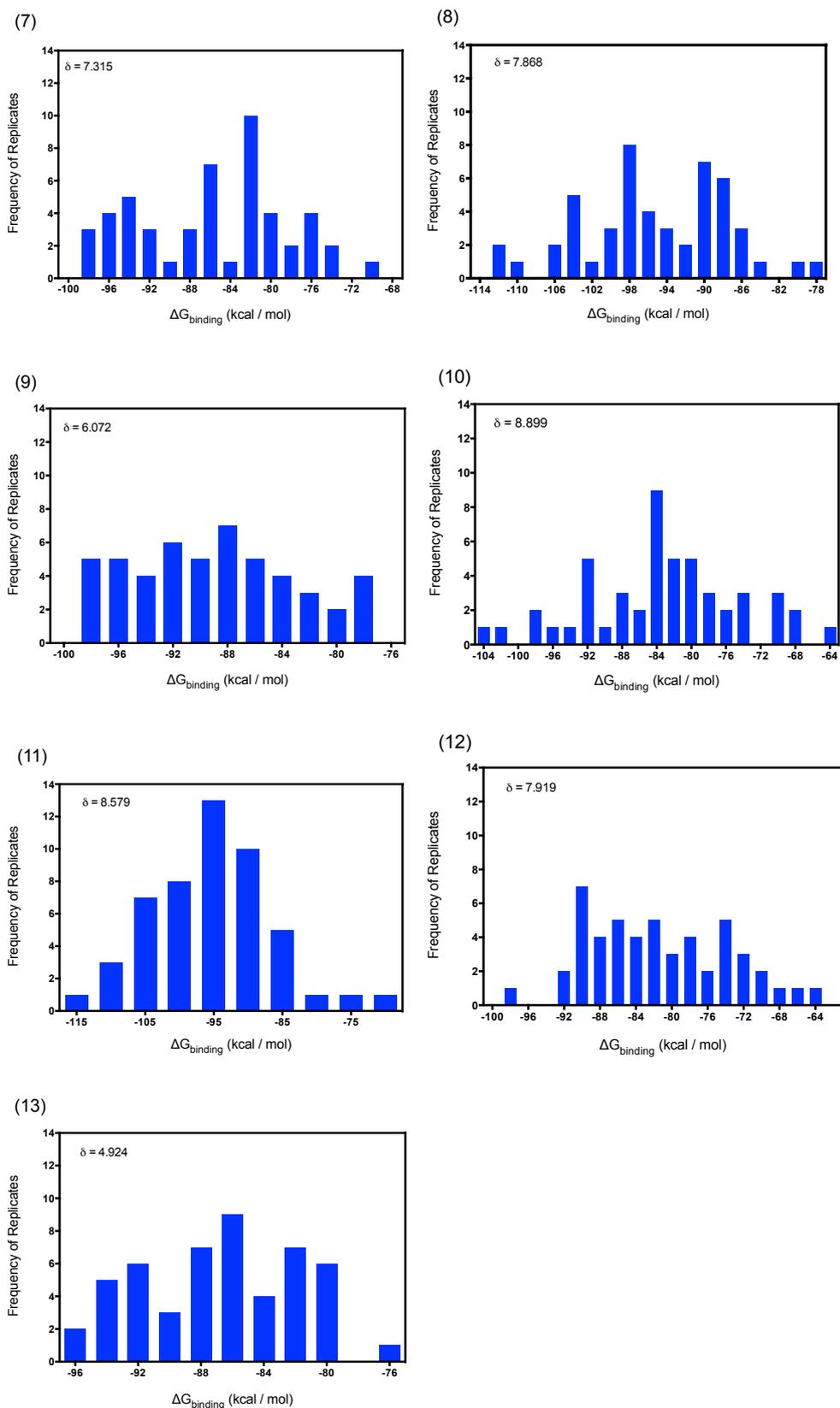


Figure 2.11 Histograms showing the distribution frequency of $\Delta G_{\text{binding}}$ of the protein–peptide complex replicates. The numbers on the histograms represent the numbers of the complex systems.

After running 50 replicates for each complex system, a relative comparison of the replicate clusters for each system was performed. Using this approach, comparisons were made between the wild-type and all mutated systems to determine the predictive binding affinity of each, relative to the wild-type complex system. One-way ANOVA was performed using GraphPad Prism V6.0 (GraphPad Software Inc., San Diego, CA; www.graphpad.com). As shown in Figure 2.12, the $\Delta G_{\text{binding}}$ values of the replicates for each complex system are distributed broadly, with small lines representing the mean $\Delta G_{\text{binding}}$ value of each system. The lines can be relatively compared, with each system coloured differently. TS01 is represented by blue line; TS05 and TS06 (orange and black, respectively) are negative controls with a markedly lower $\Delta G_{\text{binding}}$ value compared to the blue line. All remaining lines show relatively close values of binding free energy relative to TS01, except for TS08 (dark blue) and TS11 (dark green), displaying significantly higher binding free energies. Hence, TS08 and TS11 were proposed as an early lead compounds, particularly TS11 with the highest $\Delta G_{\text{binding}}$ value.

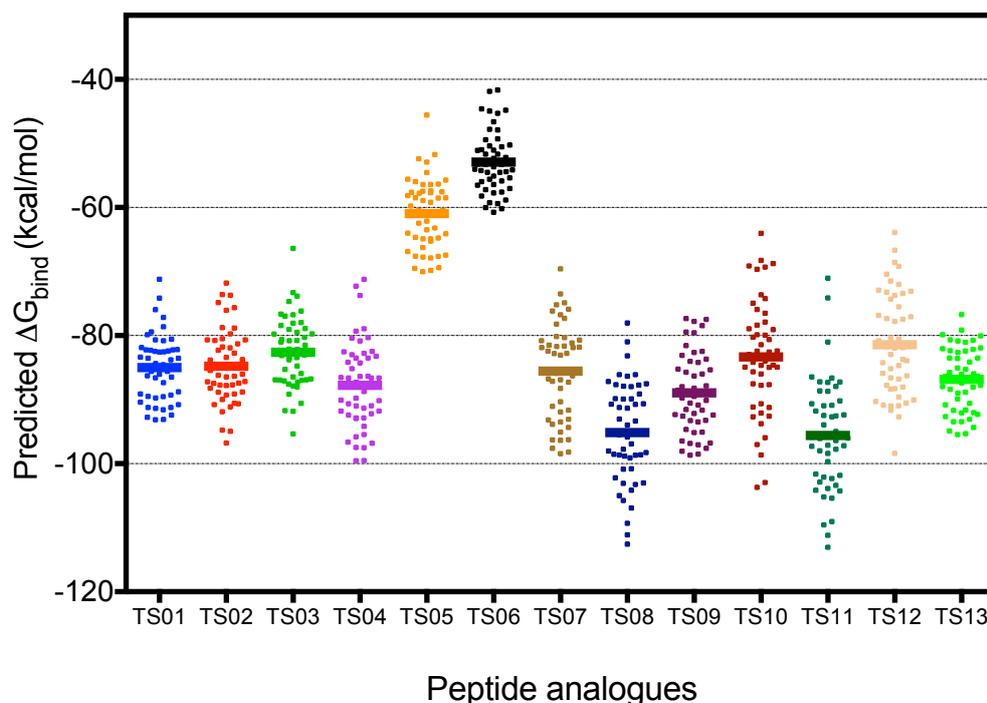


Figure 2.12 Distribution of replicate clusters for each TRF1-peptide analogue complex system. The lines represent mean value.

2.11.2 Optimising MD Simulation Time

An essential step in any molecular modelling study is the determination of the optimum time for MD simulations (the equilibration time after which no further simulation of the system is required). Two approaches were applied to obtain the optimum time for MD simulations. The first determined the equilibration time for the wild-type system by calculating $\Delta G_{\text{binding}}$ values of the set of 50 replicates at each time point over the course of the simulations. As demonstrated in Figure 2.13, the curve plateaus after about 4 ns, suggesting that the 10 ns simulations applied in this study should be of sufficient duration.

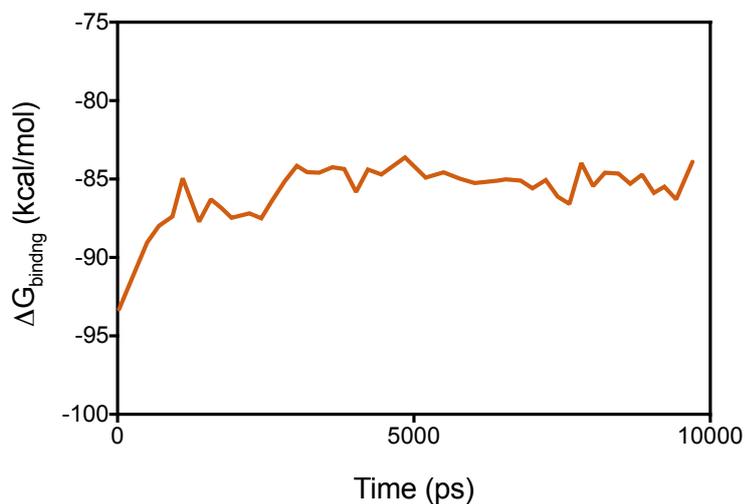


Figure 2.13 Mean $\Delta G_{\text{binding}}$ values calculated across 50 replicates of TS01–TRF1 complex at each MD simulation time point.

The second approach is to measure root mean square deviation (RMSD) with respect to MD simulation time. RMSD is the metric applied in drug design to compare the geometry of different models and/or the deviation of the predicted models from the ideal structure. This approach can be used in MD analysis as a basic tool to monitor the equilibration process, estimate the quality of the simulation, the extent of sampling of different conformations, and the variations between structural conformations²⁰³.

In this project, RMSDs were calculated for all 50 replicate of the 13 protein–peptide complexes, for the ligands and TRF1 atoms or residues at a distance of $\leq 5 \text{ \AA}$ from the peptide ligands. Computed RMSDs were produced for all 500 snapshots, calculated at each 20 ps over 10 ns of MD simulations with respect to the starting conformation (first frame), the closest to the crystal structure. RMSD plots were then used to display structural conformation changes in each trajectory over time from the initial MD snapshot, the reference structure.

RMSDs of each replicate of the complex system versus the time of MD simulations were recorded. The goal of this process was to identify system equilibrations. As illustrated in Figure 2.14, equilibration of the TRF1–TIN2 complex was assumed after approximately 3–5 ns for almost all 50 replicates.

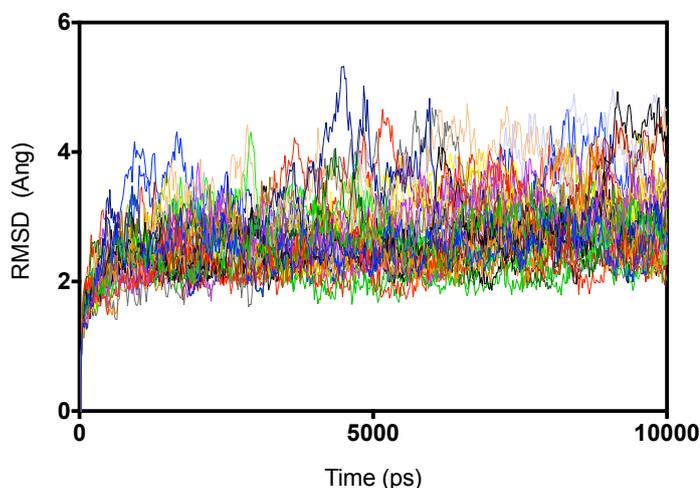
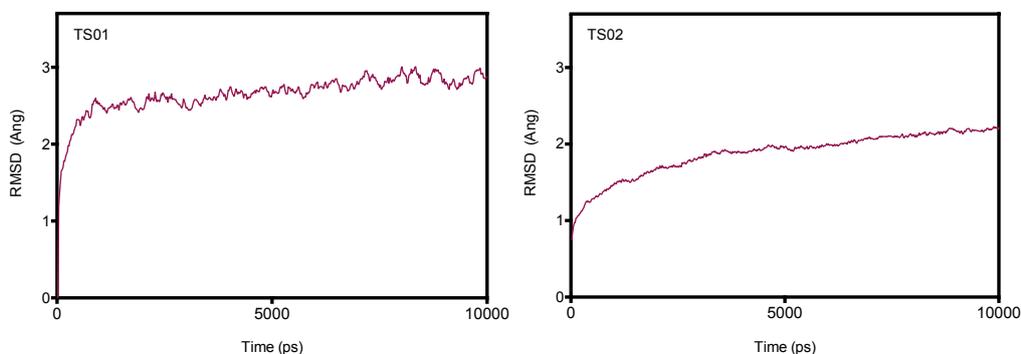


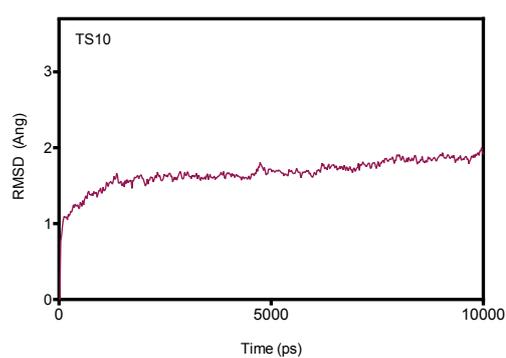
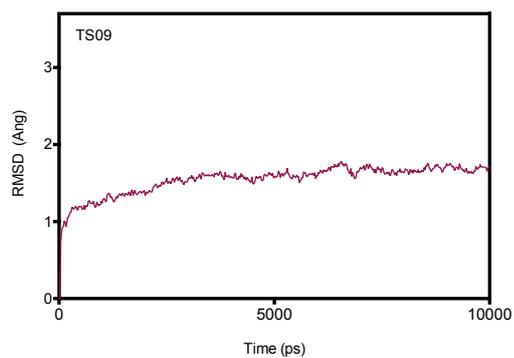
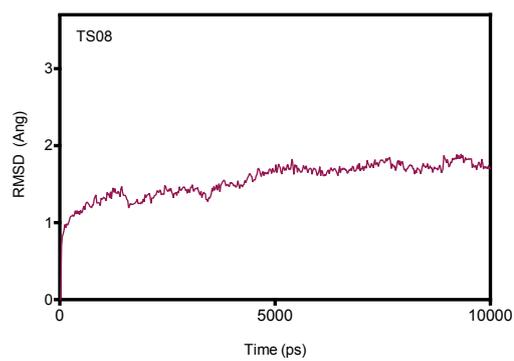
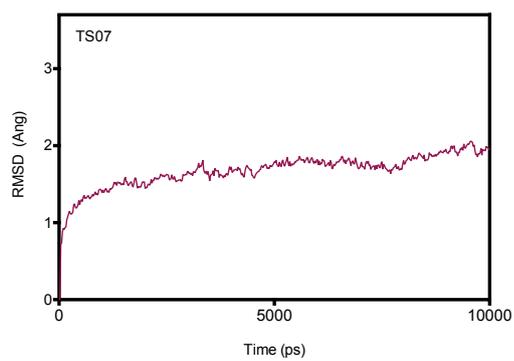
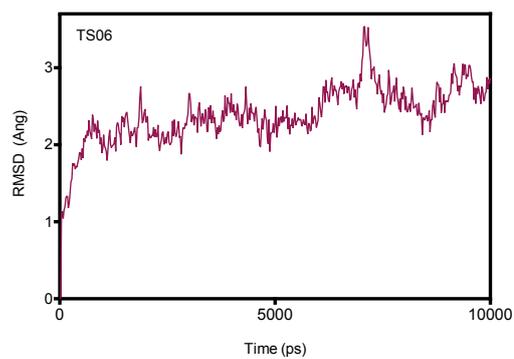
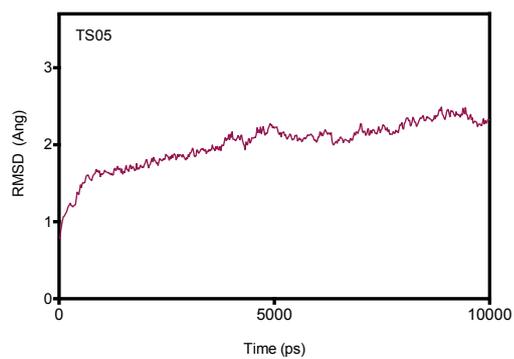
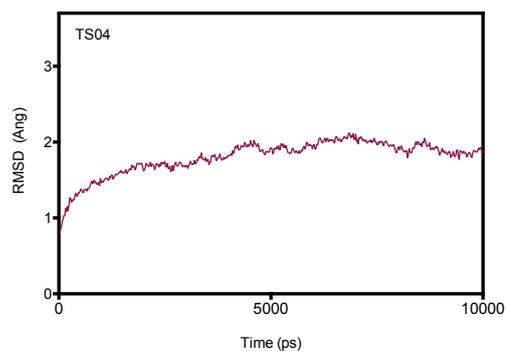
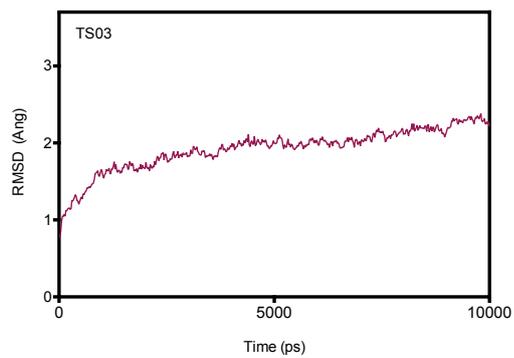
Figure 2.14 RMSDs plot of 50 replicates versus time for TS01-TRF1 complex during a simulation time of 10 ns.

This approach is not entirely solid, however, as each replicate conserves a distinctive RMSD equilibration value in the range 3–5 Å. Therefore, the average RMSDs values of each snapshot for all replicates of a complex system were calculated and plotted against time to denote equilibration time (Figure 2.15). When compared between Figure 2.14 and TS01 of Figure 2.15, both graphs represent the RMSD of the same complex system with different approaches. Nevertheless, the trends in both graphs are virtually identical; however, significant RMSD fluctuations were observed in replicates of Figure 2.14, while minor RMSD fluctuations were shown in Figure 2.15 due to the

average RMSD calculated for all replicates in each time point. Most of the complex systems in Figure 2.15 equilibrated after running approximately 5 ns MD simulations; however some of the complexes trend gradually increasing RMSDs values with respect to MD simulation time, i.e., increased conformational changes with increasing time. In addition, TS06–TRF1 (negative control) was significantly different from the other systems, as it was not equilibrated, which could be due to the inhibition of the binding affinity.

It is challenging, however, to analyse the trajectories of the complex system solely by using RMSDs analysis²⁰⁴; it is generally considered as a crude tool for comparing conformations and monitoring system convergence. Additional analysis of the systems is therefore required²⁰⁵. Factors typically related to the poor convergence of MD simulations of the complex systems are incorrect selection of force field parameters, which may lead to inaccuracies in the comparison to the crystal structure¹⁶⁹, insufficient sampling of the conformational spaces and deficiencies in energy functions.²⁰⁶





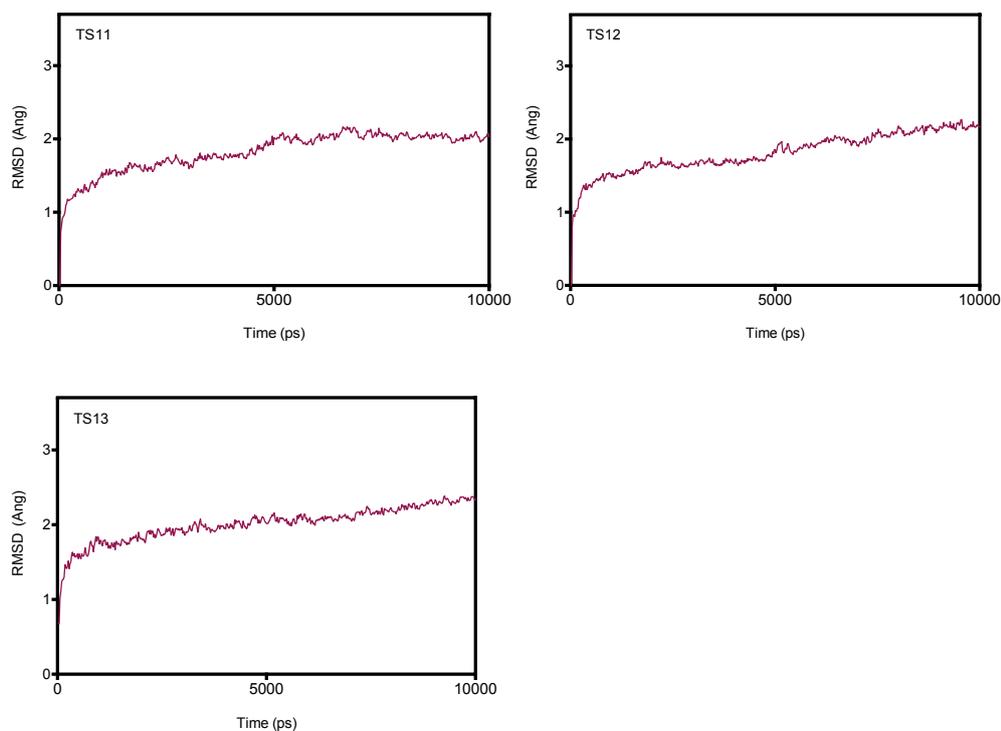


Figure 2.15 Average RMSDs of 50 replicates for each snapshot of the complex systems. The graph number represents the complex system name.

2.12 Analysis of Key Interactions

Chen *et al.* in 2008¹⁹² identified the essential binding interactions between TRF1 protein and TIN2 peptide, on the basis of the crystal structure of the TRF1–TIN2 complex (3BQO.pdb). Through our MD simulations, several interactions between TRF1 and TIN2 were identified that they were not obvious in the crystal structure. In addition, MD simulations of TRF1–TIN2 analogues were shown to either promote or abrogate the original binding interactions between the ‘hot spot’ of TRF1_{TRFH} and the peptide analogues, as discussed below.

In this study, the most significant conformational changes were determined for each protein–peptide complex system using principal component analysis (PCA). This method is used for reducing a large number of independent variables to a smaller number of collective principal components. The PCA technique was implemented in this project to analyse the snapshots of the MD trajectories for all replicates of each complex system ²⁰⁷ using an in-house produced tool known as pyPcazipgui, developed by Dr Gareth Shannon in 2015 as a derivative and developed version of PCAZIP software ²⁰⁸. The results of PCA analysis illustrate the cluster analysis and 2D histogram of the snapshots of 50 replicates in each complex system to reveal the distribution of the molecules in the conformational space, depending on projections 0 and 1. In addition, this analysis shows a graph representing the percentage of occupancy of each cluster versus time-frame (500 frames) for the complex systems revealed the presence of more than one cluster.

Interactions between the peptide analogues and the TRF1 protein in each of the complex systems are clarified below. The complex systems that are located at one region of the conformational space are illustrated by a single figure, while the systems located at a different region of the conformational space are shown as a figure either representing conformation of the molecule at the highest occupancy percentage at the equilibrium state or overlapping different conformations of the molecule (appendix A3 shows the cluster analysis, 2D histogram and the percentage occupancy for each of the peptide analogues).

2.12.1 TS01–TRF1 Binding Interactions

In addition to the binding interactions in the crystal structure, the MD simulations identified additional interactions between TIN2 and TRF1 proteins. TIN2_{TBM} peptide preserves three salt-bridges with TRF1 protein, which were confirmed through MD simulations; the first bridge is between the side chains of R265 and D139, the second is between the side chains of R266 and E192 and the last is the interaction of the R267 and E146 side chains. These salt-bridges are described as the hydrophilic interactions by Chen *et al.* (2008). We observed additional interactions of S256 and H257 of TIN2_{TBM} with TRF1_{TRFH} that were not previously discussed¹⁹²; the first is an unstable binding interaction between the side chain hydroxyl group of S256 and the side chain amide group of N144 (Figure 2.16).

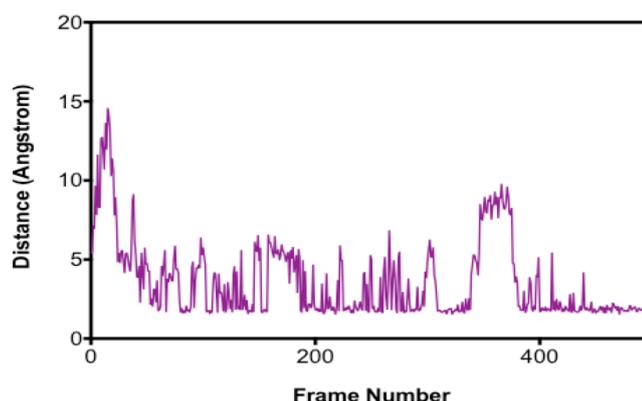


Figure 2.16 The distance between the oxygen atom of the S256 side-chain and the nitrogen atom of the amide side-chain of N144 during 10 ns MD simulations.

The next residue from the *N*-terminus of TIN2 peptide is H257. Following system equilibration during the MD simulations, it was shown to produce a strong and stable H-bonding interaction (through its backbone carbonyl group)

with the amide side chain of Q127 in TRF1_{TRFH} (Figure 2.17). According to the PCA analysis, all of the replicates are distributed at one region of the conformational space (See A3.1).

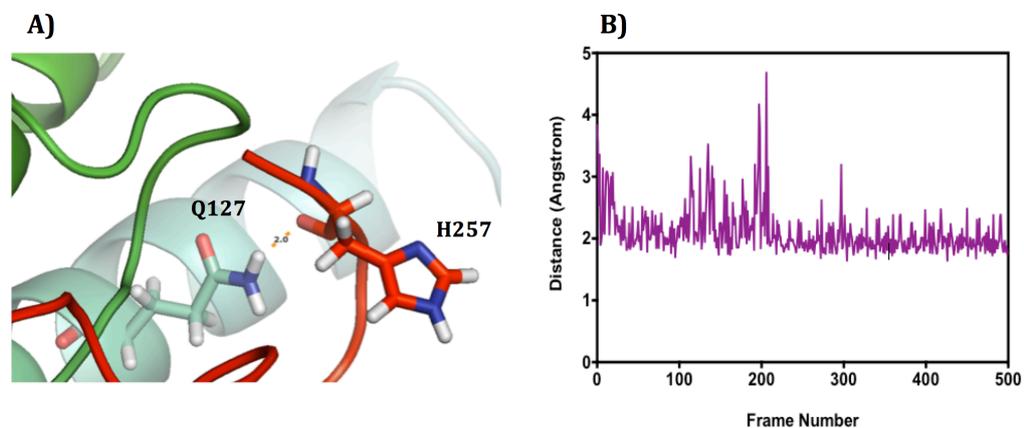


Figure 2.17 H-bond interaction between H257 and Q127 in TS01-TRF1 complex. A) A snapshot showing interaction of H257 of TIN2 with Q127 of TRF1 during the MD simulations. B) H-bond distance between the two residue atoms with respect to MD simulation time.

As shown in Figure 2.18, the conformations of S256 and H257 during the MD simulations are significantly different from the crystal structure, which result in the new H-bonds.

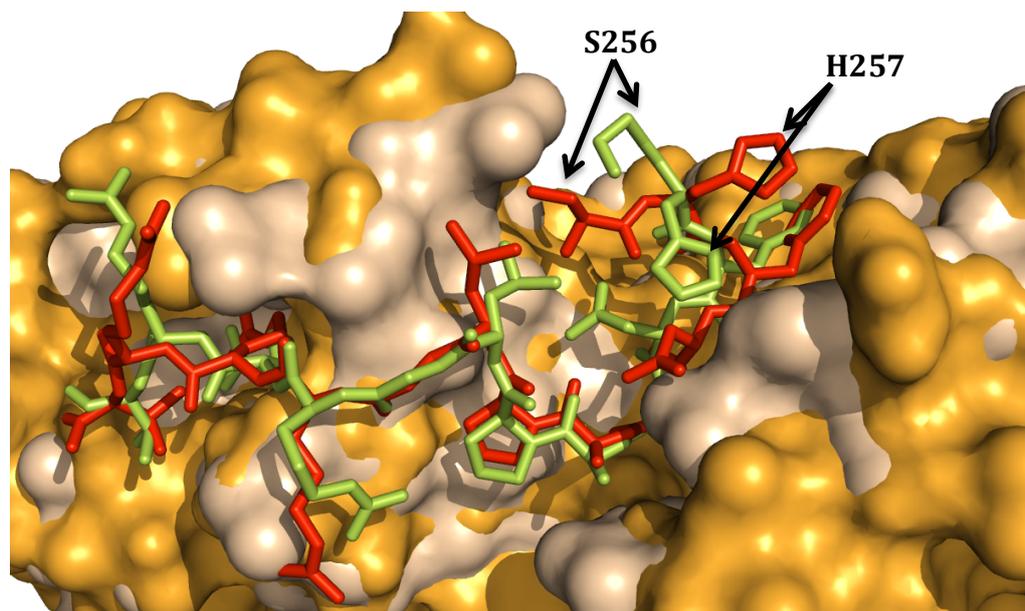


Figure 2.18 Overlapping the TRF1–TIN2 crystal structure and an equilibrated MD snapshot. The red and green colours are TIN2 of the MD snapshot and the crystal structure, respectively. The wheat color is TRF1 of the crystal structure and the orange is TRF1 of the MD snapshot.

2.12.2 TS02–TRF1 Binding Interactions

TS02 is the F258Y mutation of the wild-type peptide. As illustrated by the crystal structure, F258 has a crucial role in binding the TIN2_{TBM} molecule with TRF1_{TRFH}. F258 located on the concave surface of TRF1_{TRFH}, it forms a hydrophobic interaction with a number of TRF1_{TRFH} residues, in particular I109, L115, L120, I123 and Y124. Following mutation of this residue to a Tyr amino acid, hydrophobic contact with a group of protein residues is retained, but conformational change compared to F258 is observed to form a new favourable H-bond between the side-chain phenoxyl group of Y258 and the side-chain amide group of Q127 (Figure 2.19A). This H-bond is a strong

bonding interaction after 5 ns MD simulations as the distance between the interacting atoms is approximately 2 Å (Figure 2.19B).

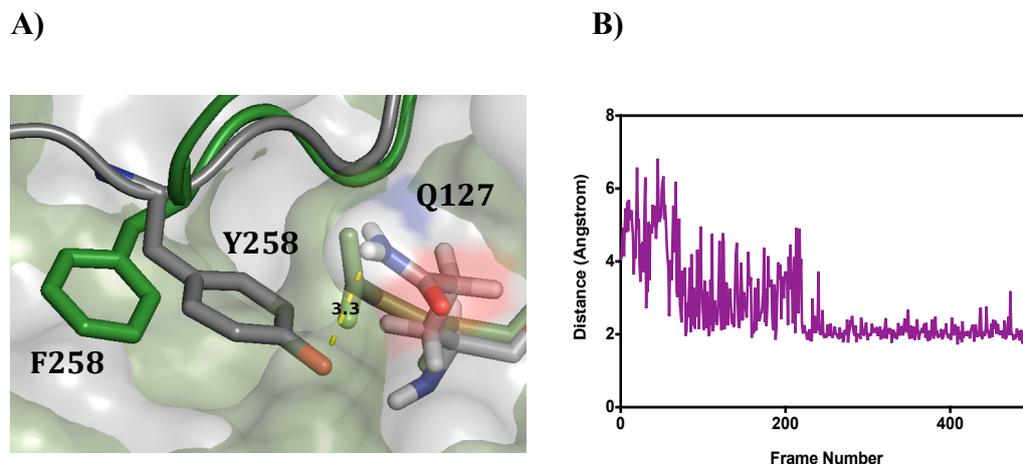


Figure 2.19 Binding interactions of Y258 in TS02 peptide with TRF1 protein. A) Overlapping TS01 and TS02, which show conformations and the hydrophobic interactions of F258 and Y258, in TS01 and TS02, respectively. In addition, it indicates H-bonding between the side-chain hydroxyl group of Y258 and the side-chain amide group of Q127. B) A graph illustrating the distance of the H-bond between the Y258 hydroxyl group side-chain and the Q127 amide side-chain during a 10 ns MD simulation.

TS02–TRF1 preserves the three salt-bridges as in TS01–TRF1. Despite generating a new H-bond after mutating F258Y, the $\Delta G_{\text{binding}}$ value is not significantly higher than that of TS01–TRF1 (Table 2-3) because the *N*-terminus of the peptide changes its conformation and cancels the hydrogen bond between H256 and Q127. In addition, changing the conformation (Figure 2.20) may affect the van der Waals and electrostatic interactions between the

protein and the ligand. PCA analysis shows distribution of the snapshots of all replicates in the one conformational region (See A3.2).

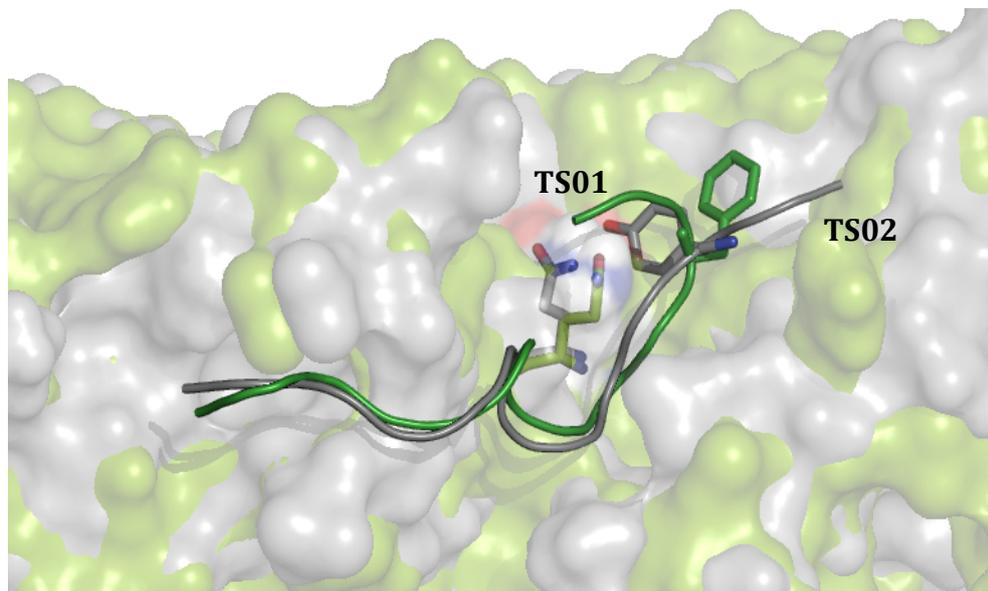


Figure 2.20 Overlapping the entire structures of TRF1–TS01 and TRF1–TS02 to show the difference in the conformations of TS01 and TS02.

2.12.3 TS03–TRF1 Binding Interactions

TS03 is the sequence of TS01 with mutating N259L. PCA analysis indicates four clusters of the conformational space for all of the 50 replicates (See A3.3); however, only two of the conformations were taken because the percentage of occupancy of the selected two clusters is 80% during the equilibrium state of the system. As illustrated by the crystal structure and MD simulations, N259 of TIN₂_{TBM} has three binding interactions with TRF1 protein (Figure 2.21A). MD simulations revealed that L259 of TS03 abolishes two of these binding interactions, with only one H-bond remaining. In one of the TS03 conformations (red), H-bond exists between the main-chain carbonyl group of

L259 and the side-chain guanidinium group of R131, but in the second conformation cluster of TS03 (green), H-bond exist between the main-chain carbonyl group of L259 and the side-chain amide group of Q127 (Figure 2.21B). Most other interactions of TS03 with the residues of TRF1_{TRFH} remain, but $\Delta G_{\text{binding}}$ inhibits at around 2.5 kcal/mol less than the wild-type complex, an observation presumably related to the elimination of the H-bonds.

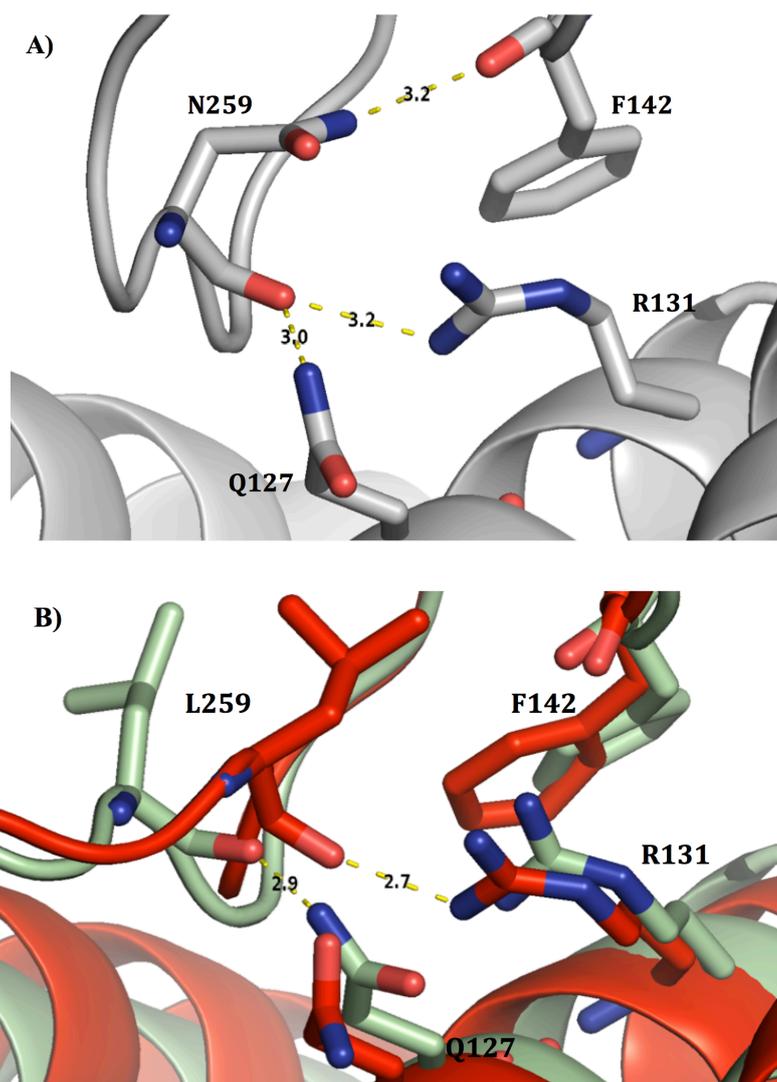
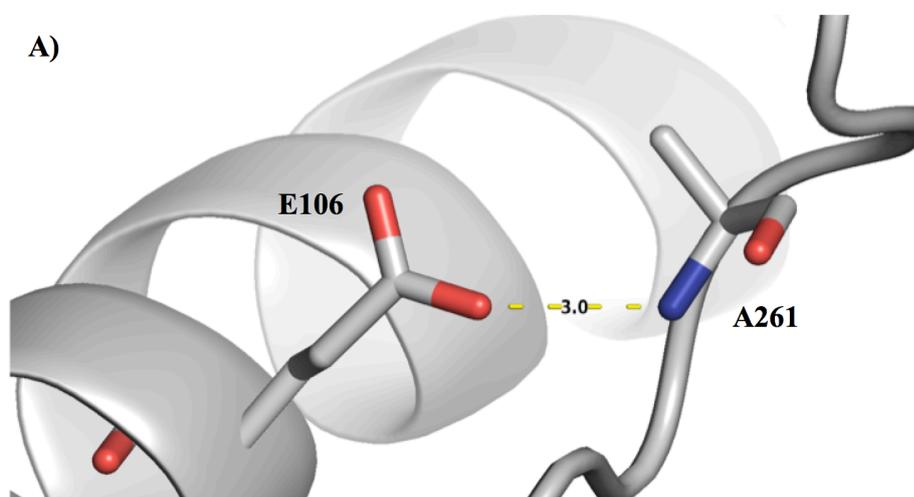


Figure 2.21 N259 of TS01 and L259 of TS03 interactions with TRF1. A) Interactions between N259 of TS01 with TRF1. B) Interactions of L259 of TS03 with TRF1.

2.12.4 TS04–TRF1 Binding Interactions

PCA analysis of TS04 reveals four clusters, meaning that the snapshots of the 50 replicates are distributed in the four major regions of the conformational space. The percentage of occupancy is around 30% for three of the clusters and 10% for one of them, in the equilibrium state (See A3.4). Therefore, the interactions between TS04 and TRF1 are represented as overlapping four different figures.

A261 of TIN2 typically has a single H-bond with the carboxylic acid side-chain of E106 through its main-chain N-H group. Following mutation of A261 to Thr in compound TS04, the bonding interaction between A261 and E106 changed to T261–E106; however, T261 binds with E106 carboxylic acid side-chains not only through the main-chain N-H group, but also through the hydroxyl group side-chain and establishes a new favourable H-bond (Figure 2.22B).



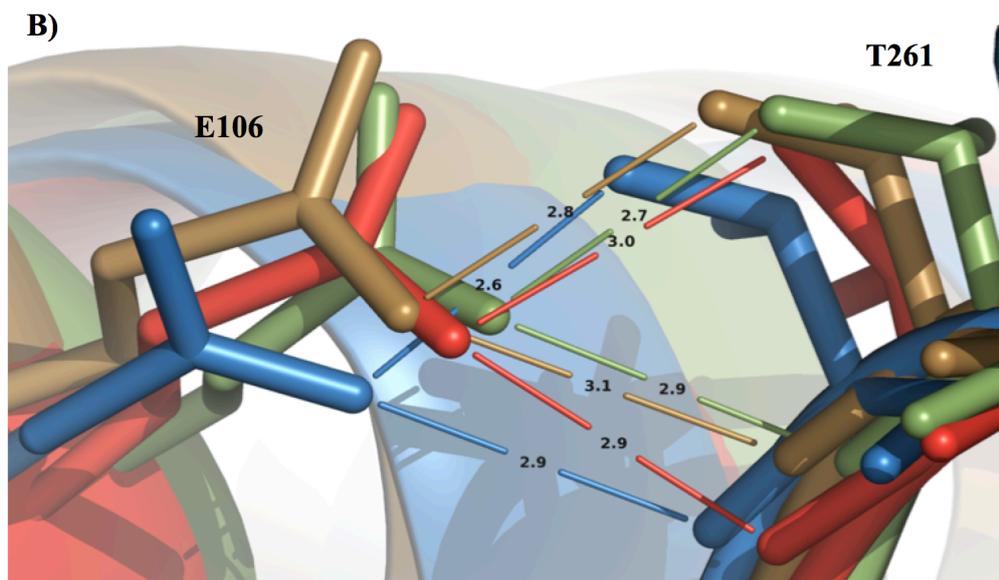


Figure 2.22 Comparing the binding interactions of T261–E106 in TS04–TRF1 complex with A261–E106 of TS01–TRF1 complex. A) Interactions between the A261 of TS01 with the E106 of TRF1 B) Overlapping the TS04 conformations to illustrate bonding interactions between T261 and E106.

The difference in the predicted binding free energy of TS01–TRF1 and TS04–TRF1 is less than 3 kcal/mol, with TS04–TRF1 complex displaying a higher binding free energy (Table 2-3). This difference contributes to the calculated margin of error; therefore, TS04–TRF1 complex cannot be considered as having a higher binding free energy. MD simulations predicted that a salt-bridge between R265 and D139 is eliminated in TS04–TRF1 complex and that only two of the salt-bridges are conserved (R266–E192 and R267–E146).

2.12.5 TS05-TRF1 Binding Interactions

The aim of this mutation was the significant inhibition of $\Delta G_{\text{binding}}$, given that R266 of TIN2 peptide has a critical role in binding with TRF1 protein; it interacts with four residues of the TRF1_{TRFH} domain (L138, D139, R147 and E192) via its side and primary chains. Following mutation of R266P, all four interactions abolish (Figure 2.23) and the salt-bridge between R266 and E192 eliminates. This process leads to a reduction in the calculated binding free energy from -85 to -61 kcal/mol. PCA analysis shows that virtually all of the snapshots are located at one region of the conformational space (See A3.5).

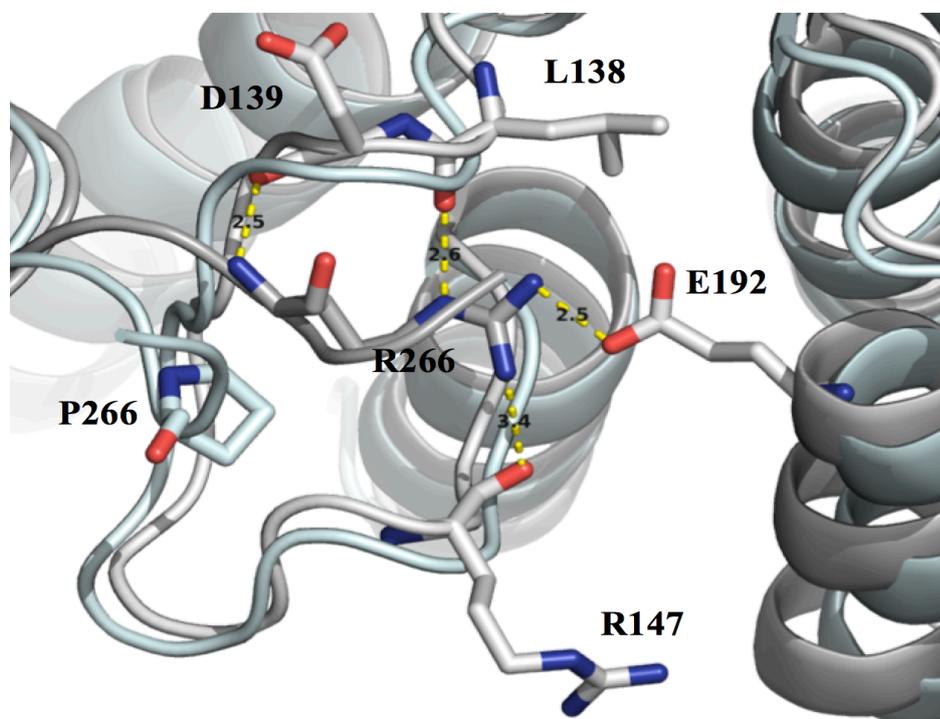


Figure 2.23 Overlay of TS01–TRF1 and TS05–TRF1 complexes to demonstrate the effects of R266P mutation on the binding interactions in the TS05 peptide.

2.12.6 TS06–TRF1 Binding Interactions

Compound TS06 was designed to be a second negative control peptide; it is related to the TS01 peptide with two amino acid mutations, the first being L260G and the second R266P. As shown for TS05–TRF1, TS05 peptide significantly inhibits the predicted binding free energy through the mutation of R266P and the $\Delta G_{\text{binding}}$ value further reduces in TS06–TRF1 (-51.42 kcal/mol) due to more disruption of the interactions between the peptide and TRF1 protein by the additional mutation of L260G. The side-chain of the L260 residue is situated in a deep hydrophobic pocket of the TRF1_{TRFH} domain, as illustrated by the crystal structure and MD simulations. Following the mutation of L260G, the hydrophobic contact abolishes, most likely contributing to the further decrease in the $\Delta G_{\text{binding}}$ value (Figure 2.24). Moreover, the R266–E192 salt-bridge abrogates in this complex compared to TS01–TRF1 complex. All of the replicates snapshots are presented as one cluster in the conformational space (See A3.6).

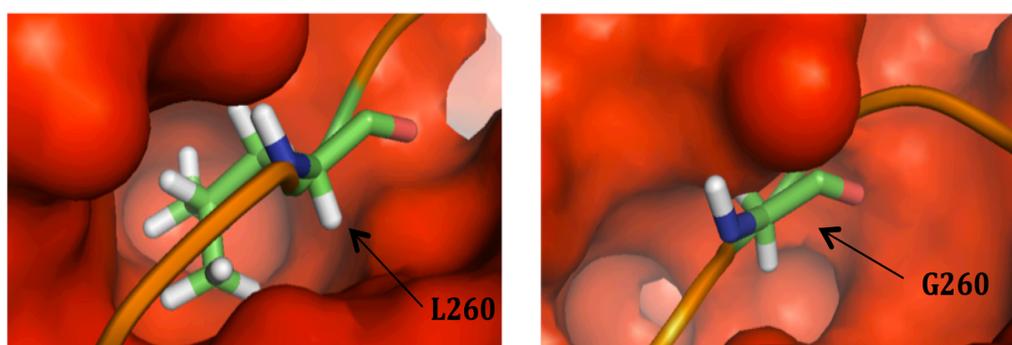


Figure 2.24 Binding interactions of L260 in TS01 and G260 in TS06, with TRF1 protein. A) L260 of TIN2 peptide located in a hydrophobic pocket of the TRF1_{TRFH} domain. B) G260 of TS06 peptide inhibits hydrophobic interactions with the TRF1 protein.

2.12.7 TS07–TRF1 Binding Interactions

The structure of TS07 is significantly different from TS01, since S256 is deleted and five amino acid residues are mutated, namely A261N, L263N, R265K, R267S and V268L. As illustrated in Figure 2.19, in terms of interactions between TIN2 and TRF1, the backbone N-H group of A261 has a H-bond interaction with E106, but N261 of the TS07 peptide produces a weaker binding interaction. Furthermore, the mutation of L263N leads to the elimination of the hydrophilic interaction between the main-chain N-H group of L263 and the main-chain carbonyl group of Q141. A new weaker, but favourable H-bond creates, however, between the amide group side-chain of N263 and the main-chain carbonyl group of E143.

Moreover, the R265K substitution causes the abolishment of a H-bond between R265 and D139 side-chains and K265 does not make contact with TRF1 residues. Further along the peptide chain, the R267S mutation results in a new binding interaction between the side-chain hydroxyl group of S267 and E146, whereas in the wild-type peptide, R267 maintains two hydrophilic interactions with E146. Also, V268L does not appear to produce any new bonding interactions. Finally, after determining the salt-bridges, only the R266–E192 salt-bridge shows to be conserved for TS07–TRF1. PCA analysis indicates two clusters, but the occupancy of one of the clusters is 60% (See A3.7), which is presented in Figure (2.25B).

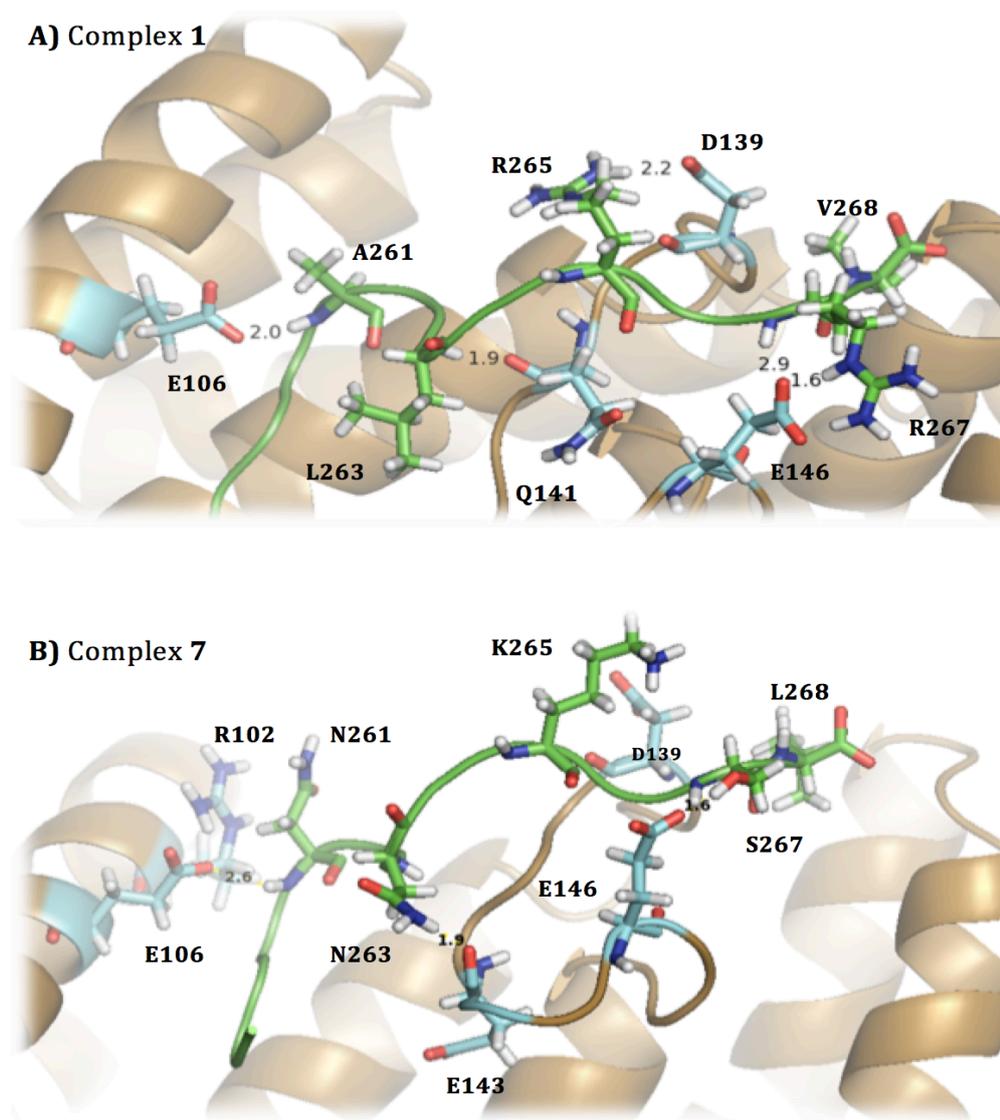


Figure 2.25 Comparison of the binding interactions of TS07 and TS01 with TRF1 protein. A) Interactions between selected residues of the TS01 molecule with TRF1 protein. B) Interactions of mutated residues of TS07 peptide with TRF1 protein.

2.12.8 TS08–TRF1 Binding Interactions

TS08–TRF1 complex has a significantly higher estimated binding free energy (-94.94 kcal/mol) than TS01–TRF1 complex. This peptide analogue has A261N and L263N mutations. MD simulations demonstrated new interactions between N261 and both R102 and E106, as well as H-bonding between the main-chain N-H group of N263 and the main-chain carbonyl group of Q141 (Figure 2.26). Moreover, the three salt-bridges of TS01-TRF1 are conserved. According to the PCA analysis, two conformation clusters are available. Figure 2.26 represents one of the clusters, which has 60% occupancy in the equilibrium state (See A3.8).

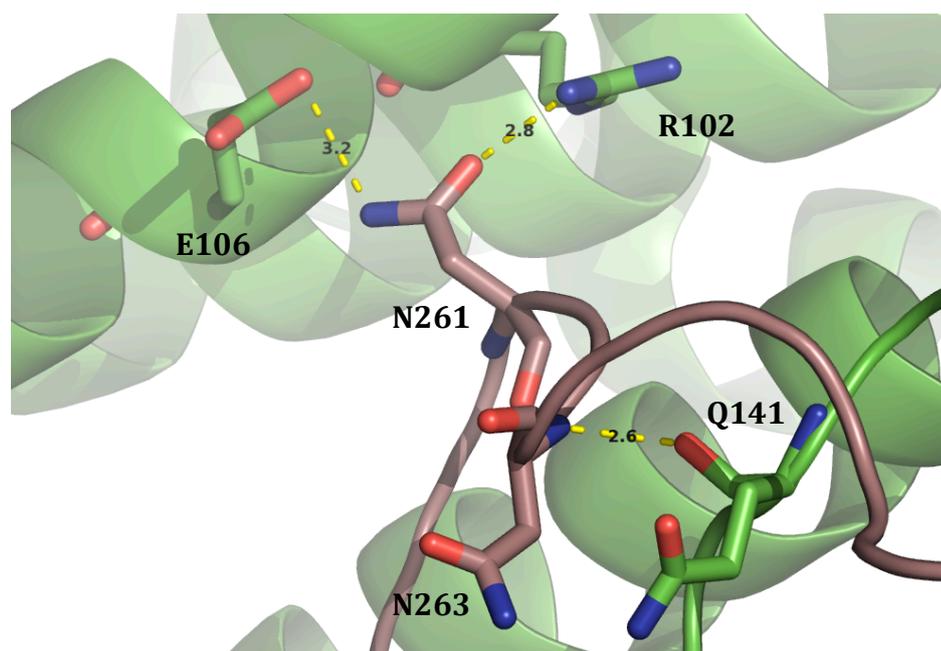


Figure 2.26 Interactions between the TS08 mutant residues (N261 and N263) with the TRF1 protein.

2.12.9 TS09–TRF1 Binding Interactions

The mutations of TS09 are A261N, R265K, R267S and V268L. Amino acid modifications in TS09 are identical to those of TS07, except that L263 does not mutate. The calculated $\Delta G_{\text{binding}}$ value increases slightly to -88.92 kcal/mol. With respect to the mutant residues of TS09 peptide, N261 produces H-bonding interactions through the side-chain amide group with the side-chain carboxylic acid group of E106 and the side-chain guanidinium group of R102. Unlike the TS07 peptide and like R265 of the TIN2 peptide, the amino group side-chain of K265 creates a H-bond with the D139 side-chain carboxyl group. S267 and L268 are free from any bonding interactions with TRF1 protein, while the C-terminal carboxylic acid produces an unstable H-bond with the N189 side-chain of the TRF1 molecule (Figure 2.27). This complex system displays a single salt-bridge between R266 and E192. PCA analysis reveals that the snapshots are distributed into two clusters. Figure 2.27 represents 65% occupancy of the conformational space (See A3.8).

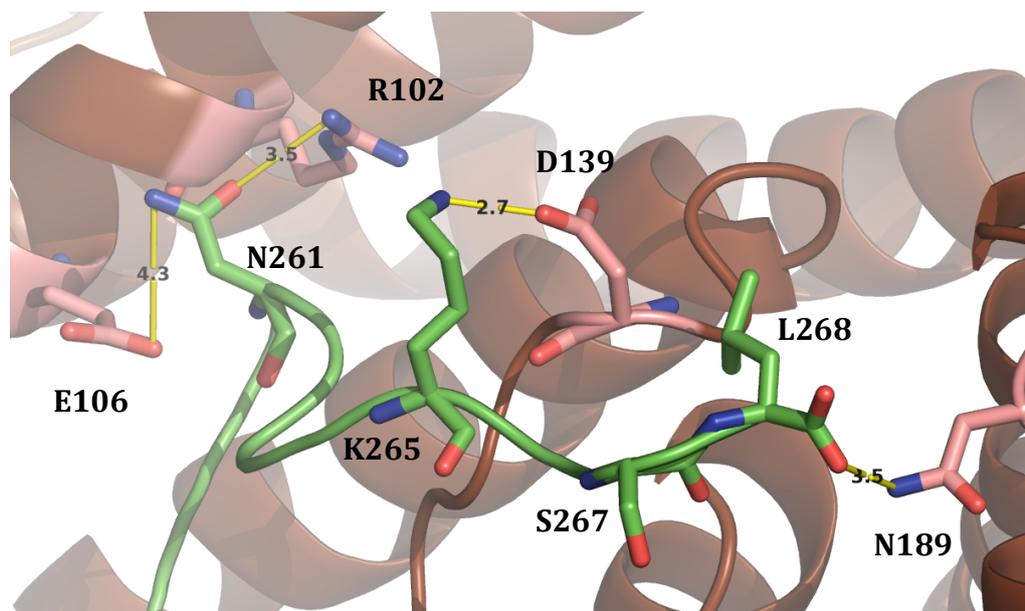


Figure 2.27 Binding interactions between the mutated residues of TS09 peptide and the TRF1_{TRFH} protein.

2.12.10 TS10–TRF1 Binding Interactions

TS10 is similar to TS07, excluding the V268L modification. The amino acid residue mutations are A261N, L263N, R265K and R267S. The binding free energy (-82 kcal/mol) is slightly less than that of TS01–TRF1 complex. R266–E192 and K265–D139 are the salt-bridges between the ligand and the protein. As shown in Figure 2.28, N261 is bound to the side-chain carboxylic acid group of E106 through both the backbone N-H and the side-chain amide). During MD simulations, N263 and S267 are shown to be free from any contact with TRF1 protein. PCA analysis provides one cluster for all of the snapshots of the 50 replicates (See A3.10).

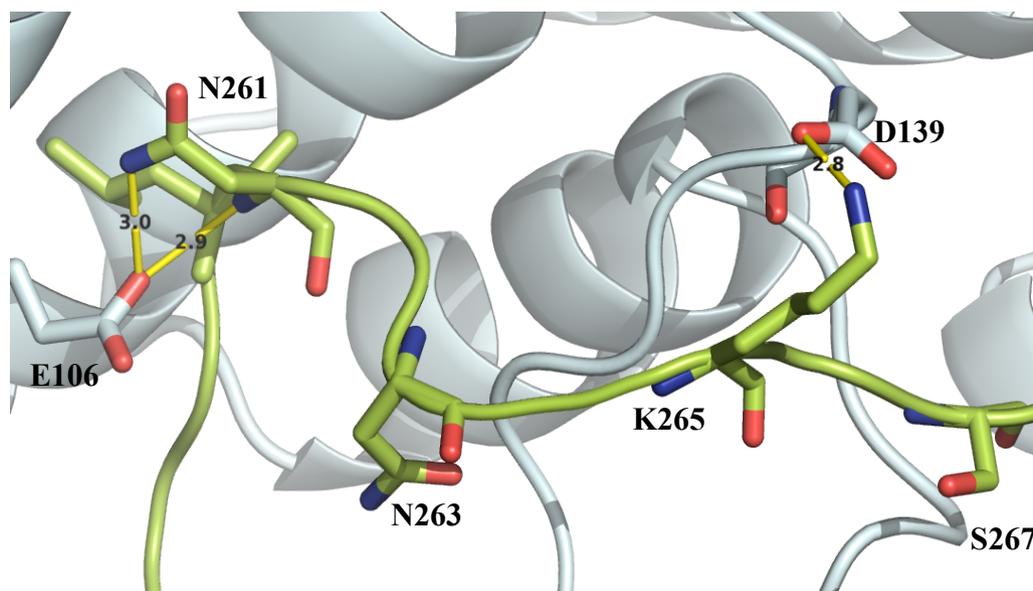


Figure 2.28 Binding interactions between TS10 residues and TRF1_{TRFH}.

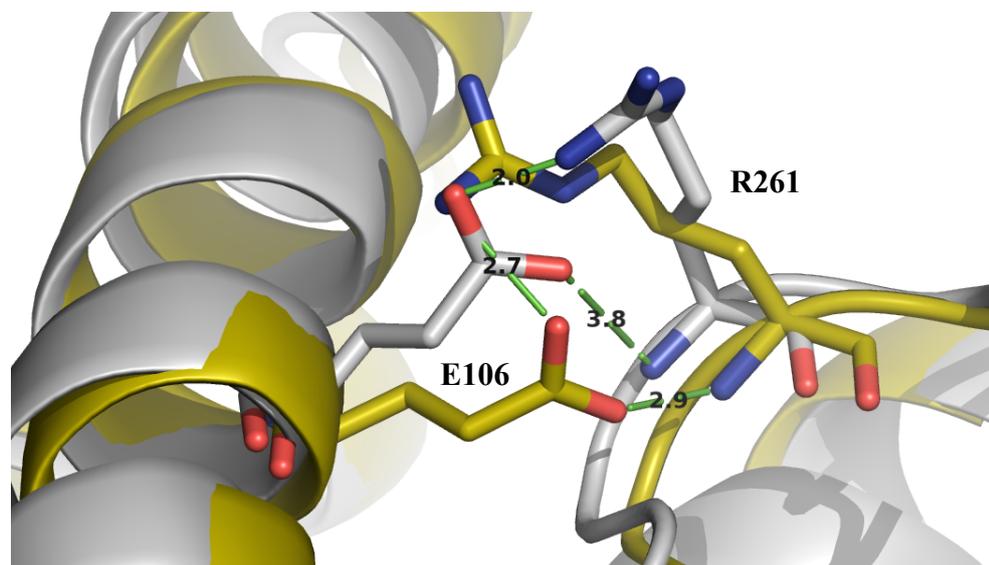
2.12.11 TS11–TRF1 Binding Interactions

Following PCA analysis of the TS11 peptide, two clusters were identified at the equilibrium state. Both have approximately the same percentage of occupancy (50% for each of the clusters), meaning that the snapshots of the 50 replicates at the equilibrium state have two main conformations (See A3.11). Therefore, the figures shown in this section represent both of the conformations, overlapped.

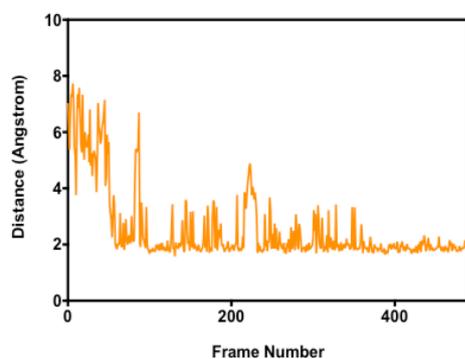
Mutating A261 of TIN2 into Arg in TS11 produces a significant change in $\Delta G_{\text{binding}}$ value of TS11–TRF1 complex, with an increase of -10 kcal/mol than that of TS01–TRF1 complex. A search for salt-bridges between TRF1 protein and TS11 peptide revealed four, which are R261–E106, R265–D139, R266–

E192 and R267–E146. In addition, a strong H-bonding observes between the backbone N-H group of R261 and the side-chain carboxylic acid group of E106 (Figure 2.29).

A)



B)



C)

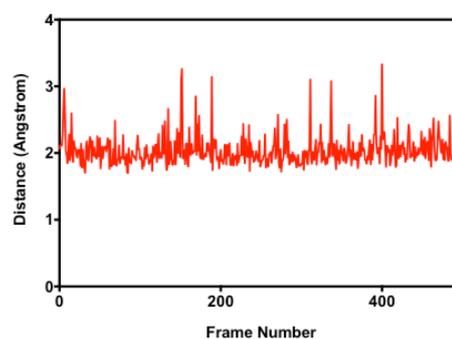


Figure 2.29 Binding interactions between R261 of the TS11 peptide and the TRF1_{TRFH} protein. A) A snapshot showing the interactions of the R261 residue of TS11 analogue with TRF1_{TRFH} protein. B) Salt-bridge between the guanidinium side-chain of R261 and the carboxylic acid group side-chain of E106. C) H-bonding between the main-chain N-H groups of R261 and the carboxylic acid group side-chain of E106.

As illustrated in Figure 2.30, both TS11 clusters have virtually the same conformation and binding interactions as the crystal structure; however, the F259 side-chain in the B conformation (yellow) changes conformation and is not close to the hydrophobic groove and the side-chain of R268 in the A conformation (red) has changed position and lost the H-bond contact with E146 of TRF1.

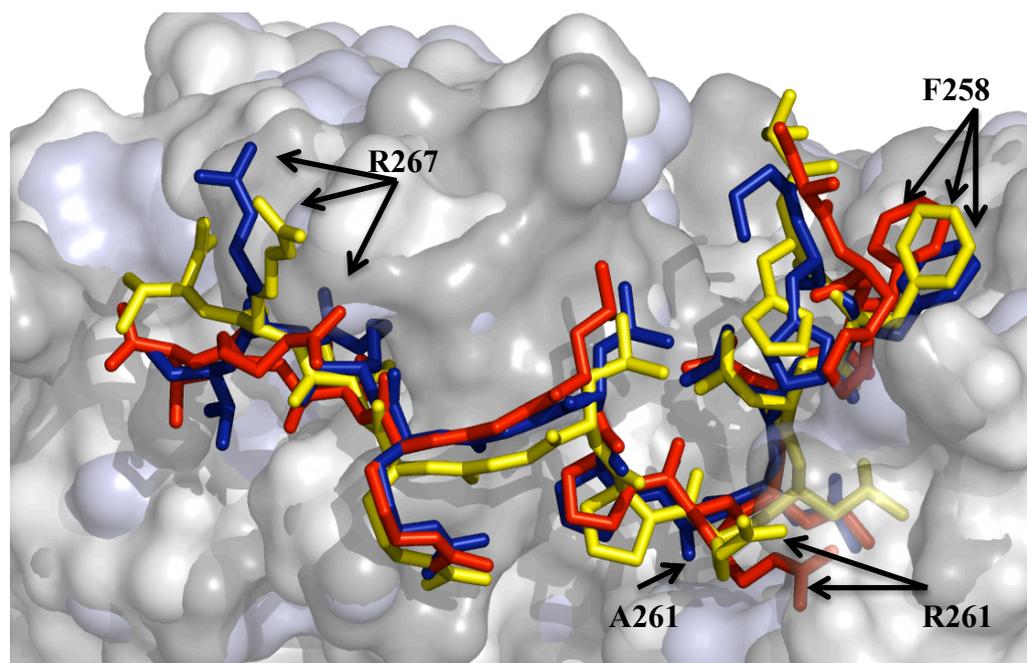


Figure 2.30 Overlapping the crystal structure of TRF1–TIN2 and the 2 conformations of TRF1–TS11. The dark blue sequence is the TIN2 crystal structure. The red (A) and yellow (B) sequences represent the 2 cluster conformations of TS11.

2.12.12 TS12–TRF1 Binding Interactions

The TS12 peptide differs from TS10 only in the F258Y mutation. $\Delta G_{\text{binding}}$ of this complex system decreases by around 4 kcal/mol compared to TS01–TRF1 complex and 1 kcal/mol relative to TS10–TRF1 complex. Despite preserving

all TS10–TRF1 complex interactions in the TS12–TRF1 complex and producing salt-bridges (K265–D139 and R266–E192), this system displays lower binding free energy. This observation is most likely related to F258Y, as F258 of TS10 entirely preserves a hydrophobic contact but Y258 of TS12 disrupted the hydrophobic interactions. Furthermore, and unlike TS02–TRF1, a slight change in the conformation results in the abolishment of H-bonding between the phenoxy group of Y258 and the side-chain amide group of Q127 (Figure 2.31). PCA analysis shows virtually one cluster for the snapshots (See A3.12).

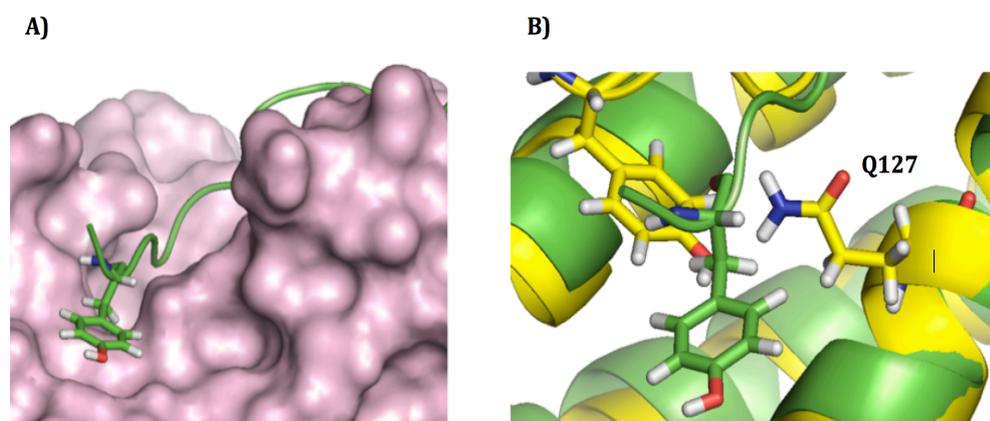


Figure 2.31 Binding interactions between TS12 analogue and TRF1 protein. A) Hydrophobic contact of Y258 of TS12 peptide with the TRF1 residues. B) Superimposed TS12–TRF1 complex (green) and TS02–TRF1 complex (yellow) shows the different conformations of Y258 in each peptide, dispelling the H-bond in TS12–TRF1 complex.

2.12.13 TS13–TRF1 Binding Interactions

Mutating V268 into Leu in the TS13 peptide results in increased $\Delta G_{\text{binding}}$ value to -89 kcal/mol in TS13–TRF1 complex. During MD simulations, TS13 peptide was shown to maintain three salt-bridges with the TRF1 protein (R265–D139, R266–E192 and R267–E146), identical to TS01–TRF1 salt-bridge elements. As shown in Figure 2.32, the conformations of V268 in TS01 and L268 in TS13 molecule are quite similar. The cause of increased binding free energy is presumably related to non-specific improvements in the van der Waals and electrostatic forces between the TS13 peptide and the TRF1 protein. PCA analysis indicates four clusters, but Figure 2.32 represents one of the clusters, which has 40% occupancy in the equilibrium state (See A3.13).

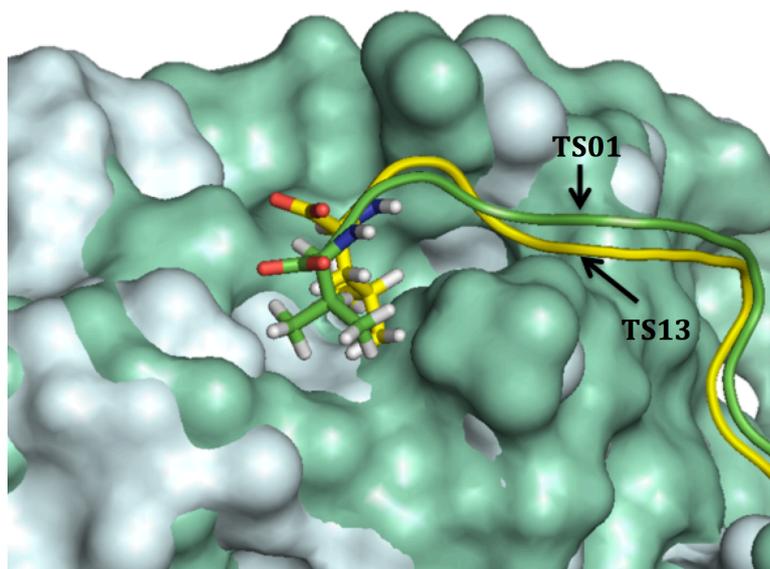


Figure 2.32 Superimposition of TS01–TRF1 and TS13–TRF1 to determine the differences in conformation of V268 and L268.

2.13 Fluorescent Probe Design

Prior to synthesis of the TIN2 peptide, the optimal site for attachment of the fluorescein molecule (See Chapter 4) was investigated through computational modelling. Typically, peptides are tagged at defined positions such as the C- or N-terminus^{145,209}. In this project, because peptide synthesis commenced from the C-terminus, the N-terminus was selected as the most straightforward conjugation position. MD simulations were undertaken to determine whether 5-carboxyfluorescein (5-FAM) caused any interference in peptide–protein interactions, but the tagged molecule was shown to be outside of the interaction region (Figure 2.33). Moreover, $\Delta G_{\text{binding}}$ of TRF1-labelled TIN2 was calculated using MM-GBSA method to realise any changes happen in the free energy of binding between the peptide and the TRF1 protein; it was -84.7 kcal/mol, which was approximately the same value as TRF1–TIN2 (-85.07 kcal/mol).

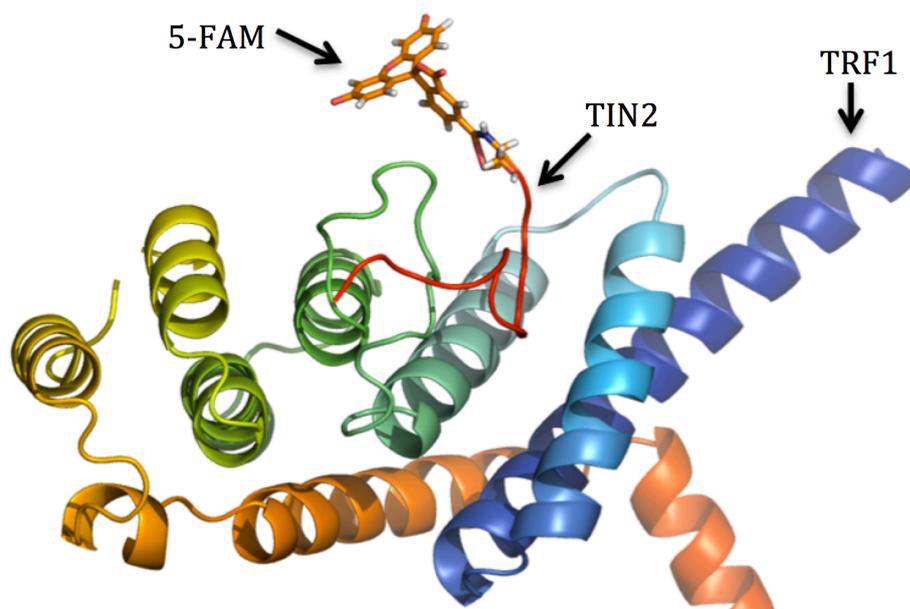


Figure 2.33 Fluorescein molecule conformation during MD simulations after tagging of the *N*-terminus part of TIN2 peptide in the TRF1–TIN2 system.

3. Total Chemical Synthesis of the Peptide Ligands

3.1 Introduction

In chapter two, the design and investigation of TIN2 and 53 of the peptide analogues, and the prediction of their binding free energies have been described. Based on the *in silico* results, a selection of TIN2 and 12 of the analogues (13 compounds in total) were synthesised. The aim of synthesising TIN2 analogues is to perform competitive binding assays (biophysical assays) with TRF1 protein.

In this chapter, the principles and methods of peptide synthesis are summarised in Section 3.2. A consideration of the solid-phase synthesis, solid supports, linker resins, activating reagents and factor(s) behind selecting a specific linker resin and activating reagent in this project is described in Section 3.3. Section 3.4 is focused on establishing the solid-phase peptide synthesis of the TIN2 peptide and its analogues, and on a chemical method to label TIN2 with a fluorescent probe. Finally, in Section 3.5, the purification method, yields and analyses of the produced peptides are presented.

3.2 Peptide Synthesis

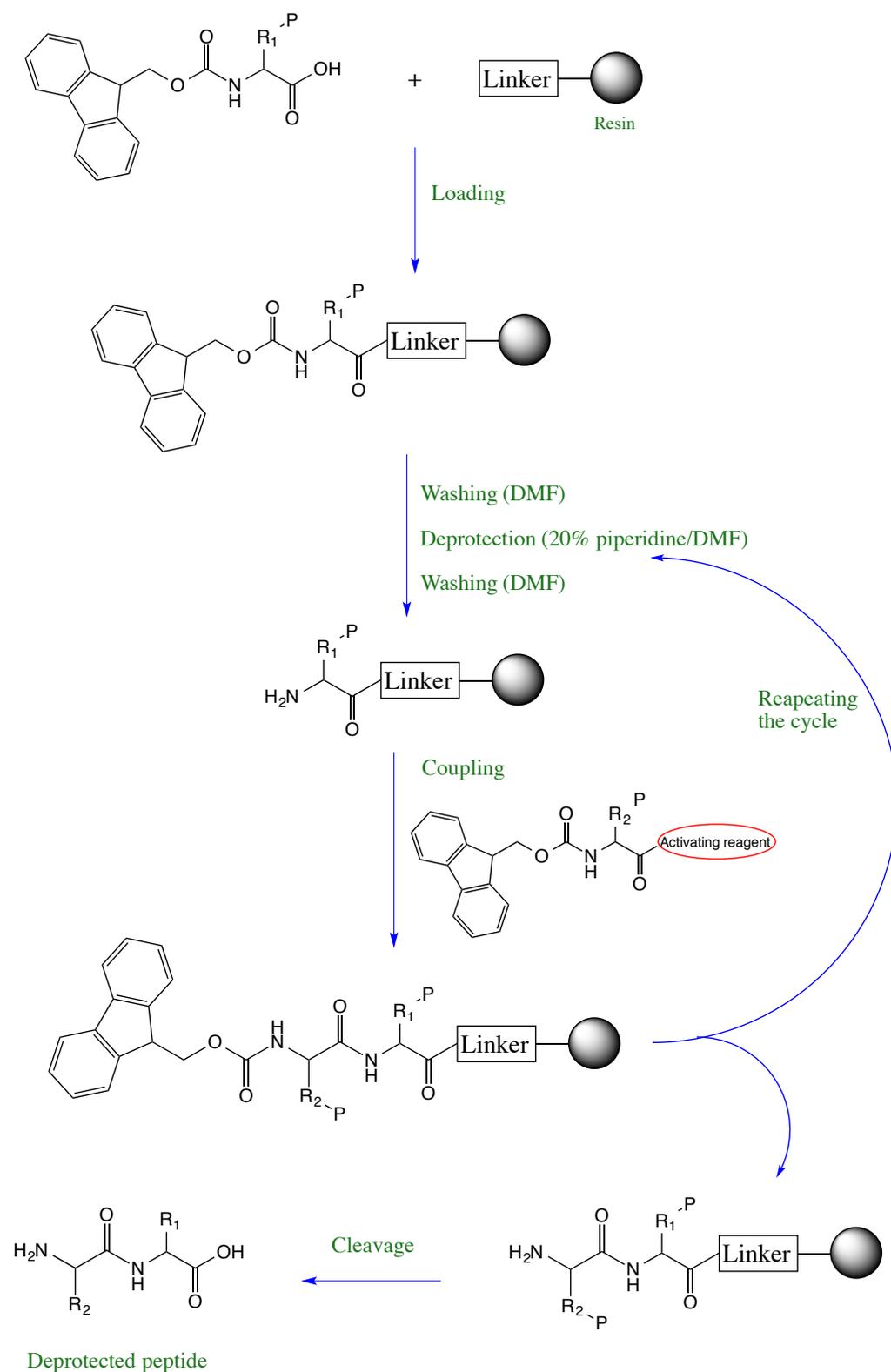
Peptides are produced by making amide bonds between two or more amino acids. The peptide or amide bond is formed through linking the α -carboxyl group and the α -amino group of two adjacent amino acids²¹⁰. There are two essential synthetic methods for assembling peptides chemically: the first is the classical or solution-phase synthesis (SPS) and the second is solid-phase peptide synthesis (SPPS). Both methods have advantages and disadvantages relative to each other. Drawbacks of the SPS technique compared with SPPS are: it is labour-intensive because there is no automation to remove the product from the solution mixture; it takes longer because of the complexity of the assembly procedure and the difficulty of the purifications; there are frequent solubility problems, particularly with increasing size of the peptides²¹¹; and the yields are lower²¹². On the other hand, the advantages of SPS over SPPS are: lower production costs for large-scale manufacturing; and a higher chance of producing a pure compound because the intermediate products are purified after each bonding step and any side reactions can be easily identified^{213,214}. To a large extent, SPS has been replaced by SPPS in the majority of laboratories. Nevertheless, manufacturers still find SPS useful for the assembly of peptide in large quantities²¹⁵.

3.3 Solid-Phase Peptide Synthesis

Between 1950 and 1960, Bruce Merrifield developed SPPS, a fast and simple method for assembling peptides. Inclusion of the term ‘solid phase’ embodies the use of a supporting polymer attached to the growing peptide, which allows for washing, deprotection and the removal of byproducts and reagents. Therefore, this method does not require purification, such as recrystallisation of intermediates to remove impurities²¹⁶.

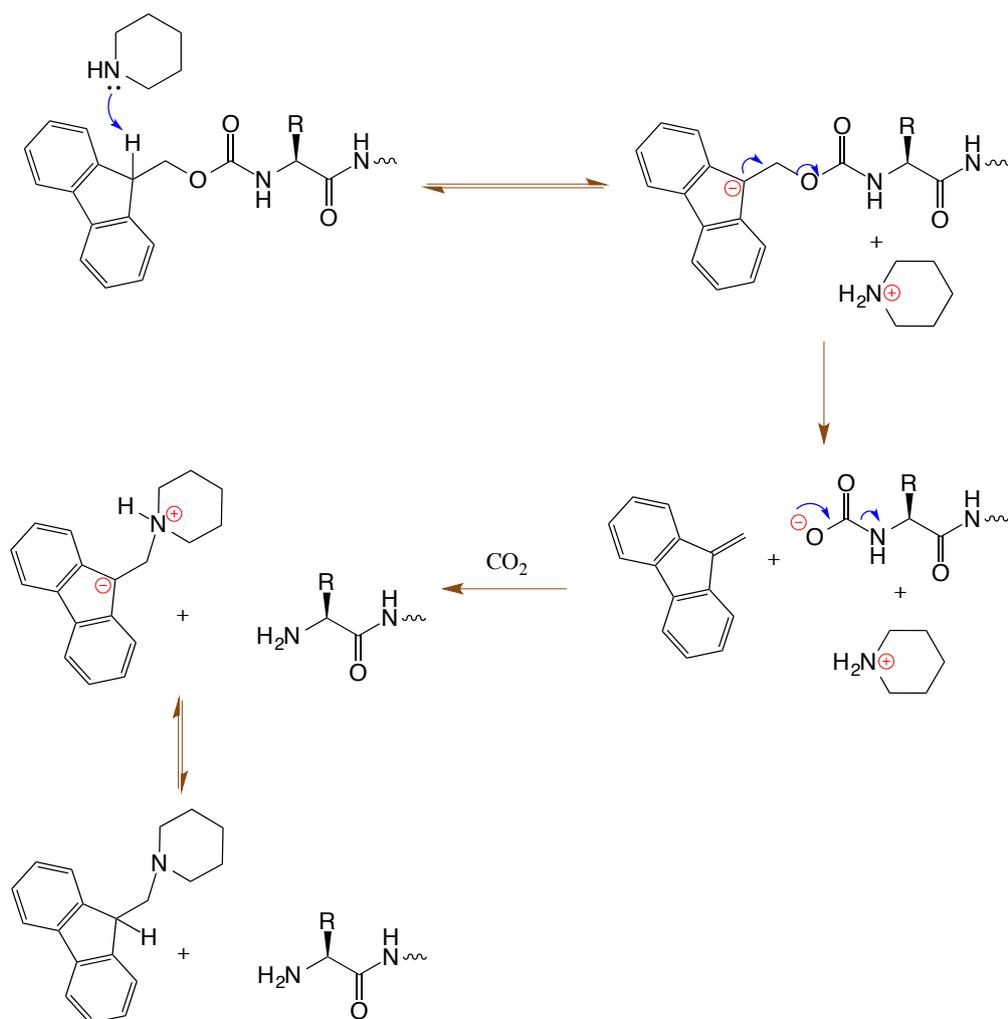
Generally, SPPS is classified into two types according to the chemical groups protecting the α -amino group of the amino acids. The first method is Boc/benzyl SPPS; here, the α -amino group of the amino acid is protected by the tert-butoxycarbonyl (Boc) group, which is sensitive to acids²¹⁷. The second method is the Fmoc/tBu approach during which the fluorenyl-9-methyloxycarbonyl (Fmoc) group protects the α -amino group and is sensitive to bases, such as piperidine²¹⁸. In the Fmoc approach, 20% v/v piperidine in dimethylformamide (DMF) can be used to deprotect the α -amino group of the amino acids. In contrast, in the Boc technique, acids, like trifluoroacetic acid (TFA) are required to deprotect the α -amino groups. The use of milder chemicals in the Fmoc method of SPPS makes it safer than, and therefore preferable to, the Boc approach. Another advantage of the Fmoc method over that of Boc is the orthogonal nature of Fmoc²¹⁹: the Fmoc temporary protecting group is deprotected under basic conditions whereas the side-chain protecting groups and the peptide–resin linkage are deprotected by an acid²²⁰. For these reasons, Fmoc SPPS was used in this study.

The process of peptide building starts from the C-terminal, which grows to the N-terminal of the peptide sequence²²¹. As illustrated in Scheme 3-1, the starting point of this technique is the linker resin; it possesses a functional group that can make an ester or amide covalent bond with the α -carboxyl group of the first amino acid introduced in the synthesis (See Section 3.3.1 for the chemical composition of the linker). Coupling of the first amino acid is crucial for the success of SPPS because it can have a big impact on the peptide yield. An activating reagent is required to activate the α -carboxyl group of the amino acid, which accelerates the covalent bond construction and prevents side reactions^{222,223}. After coupling, DMF is used to wash and remove excess amino acid, byproducts and reagents from the resin. Next, the Fmoc-protected α -amino group of the amino acid that is now attached to the linker resin is deprotected using 20% *v/v* piperidine/DMF. Subsequently, the next carboxy-activated amino acid is attached to the growing sequence. This cycle is repeated until the desired peptide is produced. After producing the peptide sequence, the α -amino protecting group is removed by 20% *v/v* piperidine/DMF, while the side-chain protecting groups and the linker resin are cleaved by an acid, such as TFA^{220,221}.



Scheme 3-1: The steps of synthesising peptides using Fmoc SPPS method.

Scheme 3-2 represents the steps of Fmoc cleavage from the α -amino group: the first step is the deprotonation by piperidine and the formation of a cyclopentadiene ring; dibenzofulvene is then separated from the amino acid and then forms a dibenzofulvene–piperidine adduct. The produced adduct absorbs ultraviolet (UV) light at 350 nm and is therefore a valuable tool for monitoring α -amino group deprotection of amino acids by simple UV spectrophotometry²²⁴.



Scheme 3-2: Mechanism of Fmoc deprotection and dibenzofulvene–piperidine adduct production from the α -amino group of the amino acids in the SPPS method.

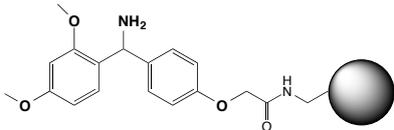
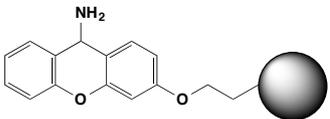
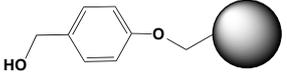
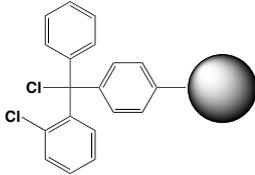
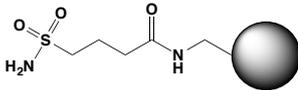
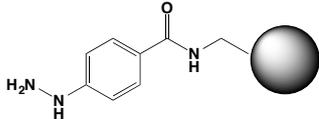
3.3.1 Resin Materials and Linker Types for SPPS

Selecting an appropriate resin and linker is fundamental for successful peptide synthesis. Resin is composed of insoluble polymer, such as polystyrene. The resin swells extensively in solvents such as dichloromethane (DCM) and DMF, allowing penetration of the reagents and construction of a peptide sequence inside or on the particles²²⁵. Resins are connected to the peptides through linkers; they should be chemically stable and must not react with different solvents and reagents^{217,226}. The most common resin comprises polystyrene supports; these are spherical beads of two size ranges, 35–75 microns and 75–150 microns²²¹. Other types of resins are polyamines and polyethylene glycols (PEG) resin²²⁷. A PEG resin has been developed recently, known as ChemMatrix[®], which can be used for the production of many long and complex peptides²²⁸.

Linkers are the connectors between the resin, i.e. polymeric support and the peptide chains. Their primary function is to anchor the C-terminal part of the peptide structure reversibly through a reaction with the α -carbonyl group of the C-terminal amino acid. The second function of the linkers is to protect the C-terminus of the peptide during the assembly process. Various types of linkers are commercially available; the type can determine not only the functional group of the peptide C-terminal, but also the chemicals used to cleave the peptide from the resin. As illustrated in Table 3-1, different types of linkers can provide different C-terminal functional groups of the peptides, such as carboxylic acid, or amide, ester²²⁹ and thioester groups²³⁰. Most of the peptide linkers release a peptide with the C-terminal carboxylic acid (e.g. Wang resin)

or amide group (e.g. Rink amide resin) (Scheme 3-8), both of which require acidic conditions for their cleavage^{220,221,229,230}.

Table 3-1 Commonly used linkers that produce peptides with different C-terminal functional groups.

Linker name	C-terminal functionality	Structure
Rink amide linker	Peptide amide	
Siber linker	Peptide amide	
Wang linker	Peptide acid	
2-chlorotrityl chloride linker	Peptide acid	
Safety-catch linker	Peptide thioesters	
Aryl hydrazide linker	Peptide amines	

3.3.2 Coupling Reagents in SPPS

Coupling or activating reagents are chemicals used to activate the α -carboxylic acid of the amino acids in SPPS in order they could react with the α -amino group of the growing peptide sequence. The coupling reaction to form a peptide bond is through increasing electrophilicity of the α -carbonyl group. The reaction can be achieved by replacing the hydroxyl group of the α -carbonyl group with an electron-withdrawing group. There are different chemical classes of activating reagents, such as carbodiimides and onium (aminium/phosphonium)-based activating reagents^{220,231}. Choosing an appropriate coupling reagent is one of the critical factors for succeeding with SPPS²³². Carbodiimide reagents such as dicyclohexylcarbodiimide (DCC) and diisopropylcarbodiimide (DIC) are the first class of reagents applied in peptide chemistry (Figure 3.1). When they react with the α -carboxyl group of amino acids, a highly reactive O-acylisourea intermediate compound is produced, which can easily be attacked by a nucleophilic α -amino group of the amino acids.

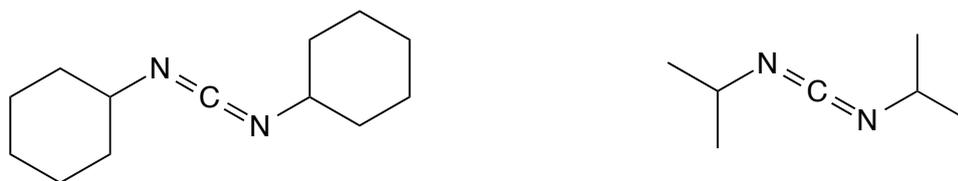
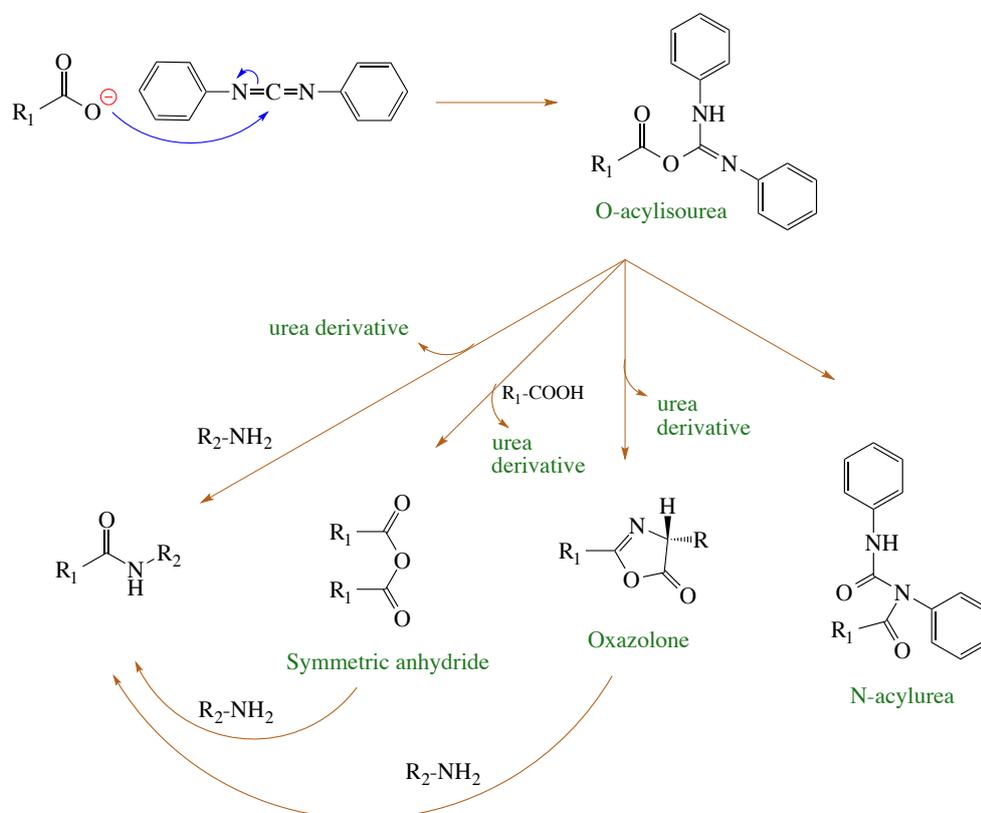


Figure 3.1 Structures of DCC and DIC activating reagents.

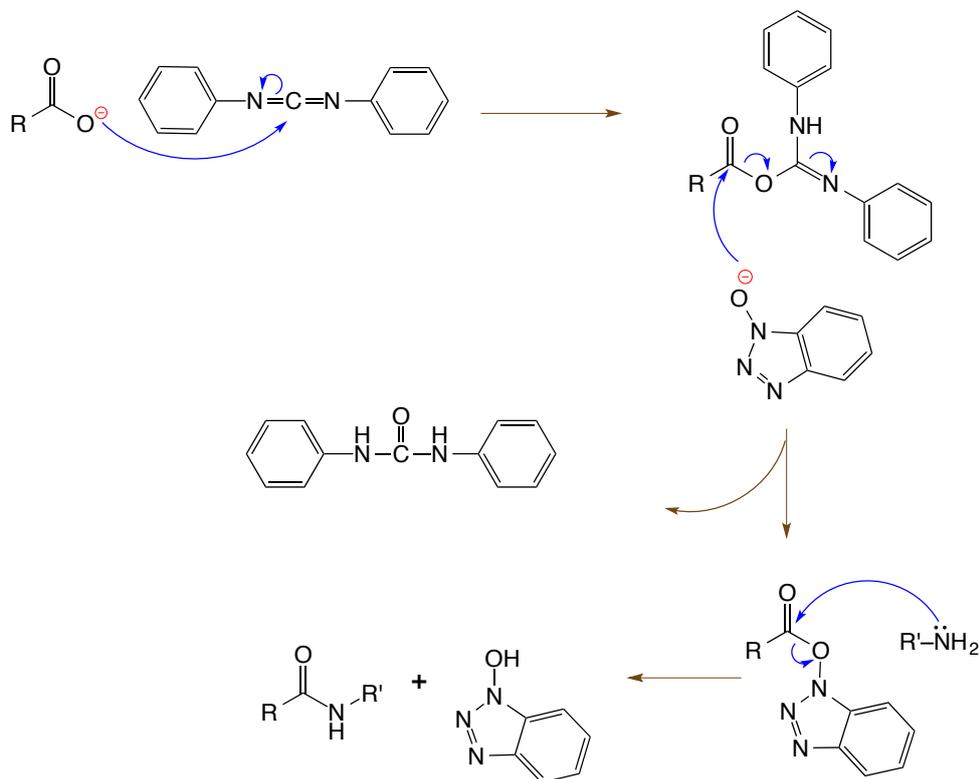
The major drawback of carbodiimides is the production of the highly reactive O-acylisourea, which leads to racemisation and the yield of side-products, such as N-acylurea (Scheme 3-3). In addition, DCC produces dicyclohexylurea that has poor solubility in organic solvents^{221,233}.



Scheme 3-3: Coupling reaction when using only Carbodiimides activating reagent (adapted from El-Faham and Albericio 2011).

Hence, *N*-hydroxy triazole reagents were introduced as additive compounds with carbodiimides to improve the efficiency of reactions, as they produce corresponding active esters^{231,234}. Tertiary amine groups of the triazole activating reagents can facilitate formation of the active ester²³⁵. One of the most commonly used triazoles is 1-hydroxy-benzotriazole (HOBt), which

produce an active ester (Scheme 3-4). However, a main disadvantage of these reagents is their explosive nature^{236,237}.



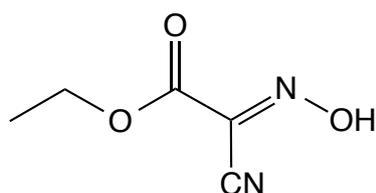
Scheme 3-4: Coupling reaction using DCC and HOBt as activating reagents.

The newer and most widely used coupling reagents are onium (aminium/phosphonium) reagents. They are more powerful and safer reagents compared with previous reagents^{238,239}. The benzotriazole moiety is the essential part in the structure of the most aminium and phosphonium activating reagents. The potent reactivity of aminium salts leads to side reactions, especially when excess reagent is used. The excess aminium reagent has the potential to react with the free *N*-terminal of the amino acid to afford a guanidine moiety. Usually, phosphonium salts are preferred to aminium salts,

as they retain the same reactivity; in addition, their positively charged phosphorus centre can prevent peptide termination and cyclisation²³⁸.

Recently, the Oxyma activating reagent, ethyl 2-cyano-2-(hydroxyimino) acetate has been developed as a non-explosive alternative of to the *N*-hydroxy triazoles (HOBt and HOAt), which can be used as an additive with carbodiimides (Figure 3.2)²³⁷. Oxyma can be considered as a powerful replacement for benzotriazole-based additives due to a number of desirable characteristics: first, it is highly soluble in widely different solvents; second, it poses significantly lower thermal risks compared with benzotriazole compounds such as HOBt. Finally, it can significantly inhibit racemisation^{237,240}.

a)



b)

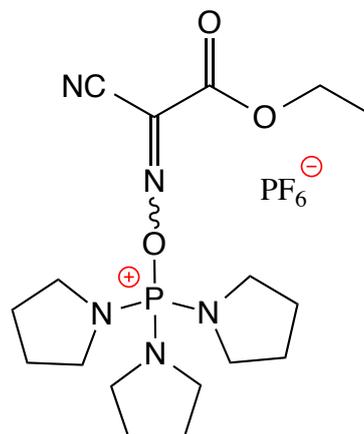
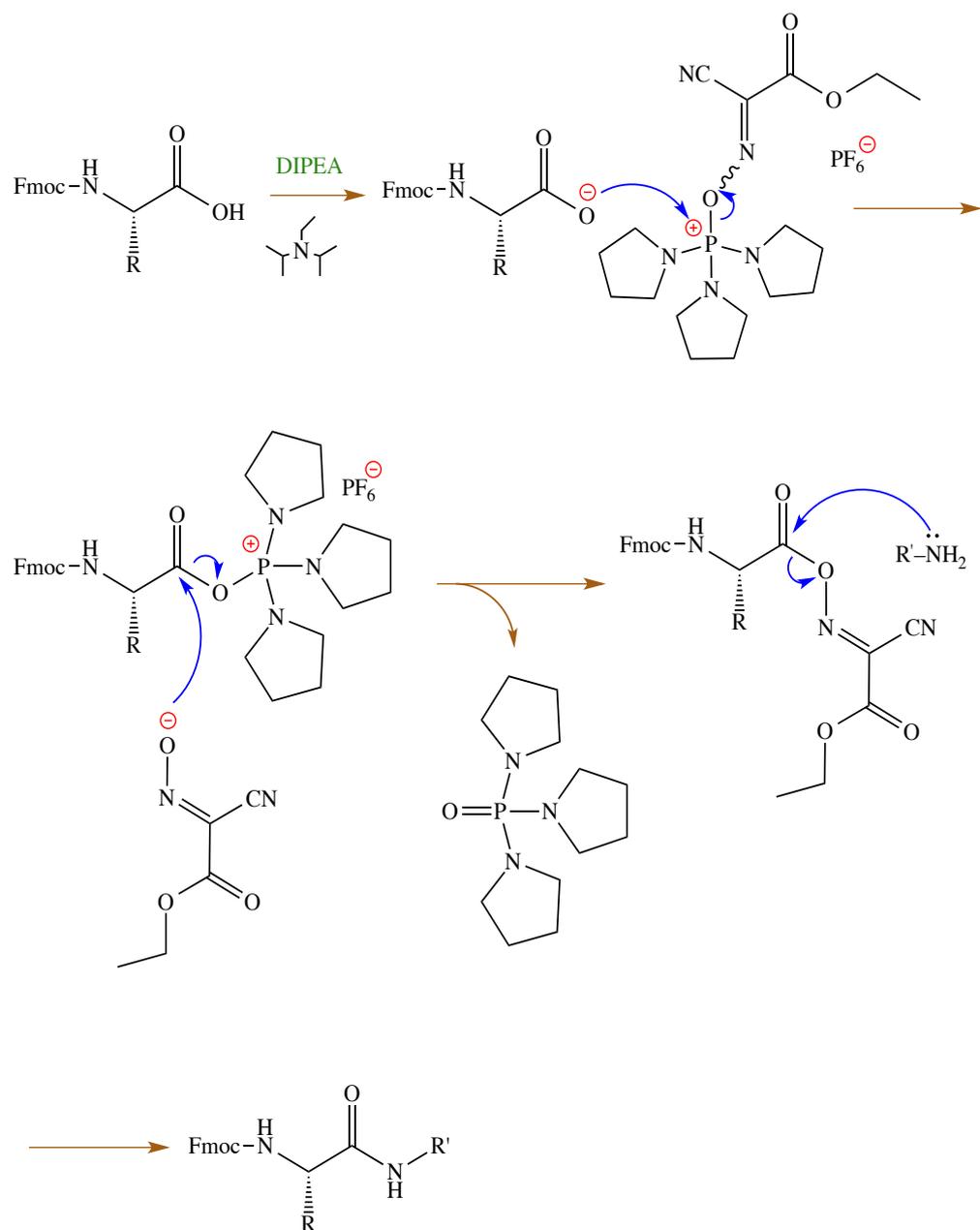


Figure 3.2 The structures of a) Oxyma and b) Pyoxim activating reagents.

In 2010, Subiros-Funosas *et al.* introduced an Oxyma-based phosphonium salt activating reagent known as PyOxim: O-[(cyano (ethoxycarbonyl)methylidene)-amino]-yloxytri(pyrrolidino)phosphonium hexafluorophosphate ²³⁴. This compound is a good choice for the SPPS: first, it combines the elegant characteristics of Oxyma based reagents and phosphonium-based reagents; second, it produces peptides with a significantly lower probability of epimerisation or racemisation; third, it is more reactive than the earlier reagents ²³¹; fourth, it is safer and has a lower thermal response compared to the benzotriazole-based compounds ²³⁷; fifth, it has greater solubility in organic solvents such as DMF and DCM than benzotriazole activating reagents; and, finally, its allergenic potential is minimal ²³⁴. Therefore, PyOxim was selected as an activating reagent for the peptide assembly in this project.

The mechanism of action of PyOxim is illustrated in Scheme 3-5. The carboxylate anion reacts with the phosphonium cation of PyOxim to produce the acyloxyphosphonium intermediate, which in turn is attacked by the released Oxyma anion to yield an Oxyma amino acid ester and the byproduct, phosphonium oxide. Finally, aminolysis of the Oxyma active ester results in producing the desired peptide bond ²⁴⁰.

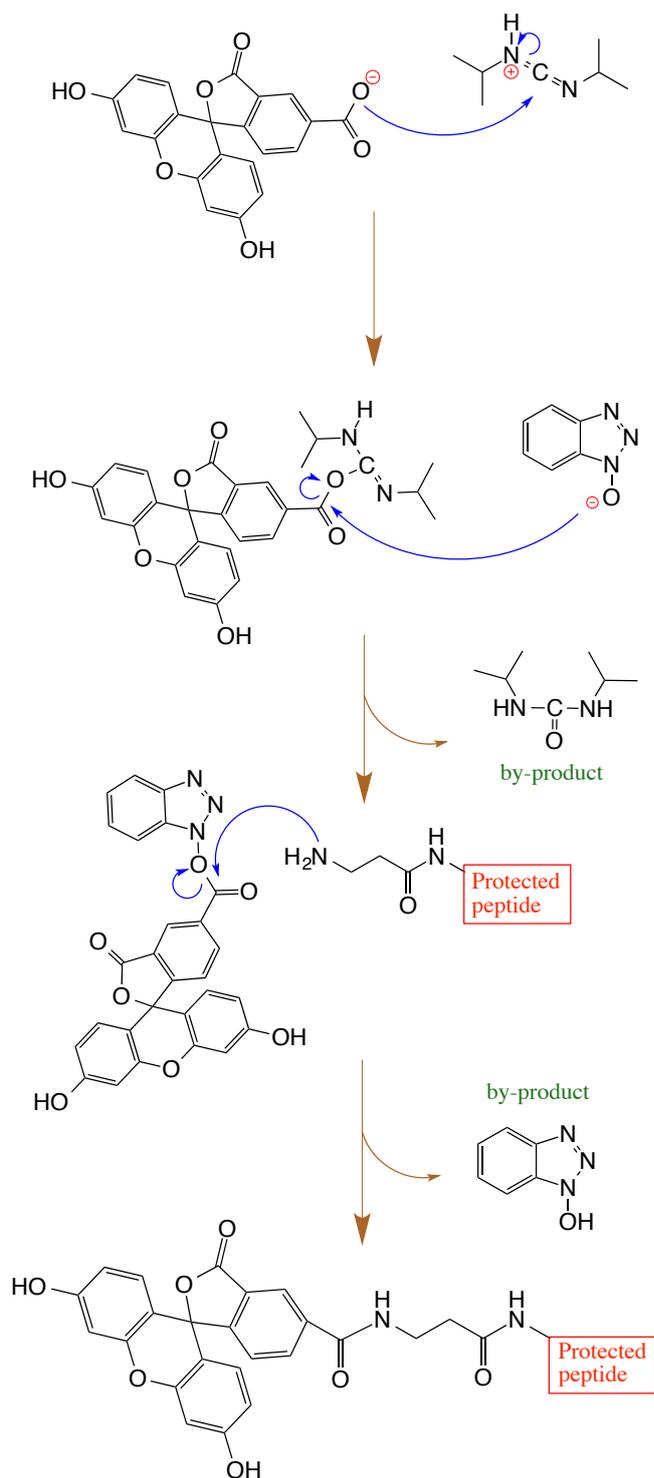


Scheme 3-5: Mechanism of the coupling reaction between two amino acids, catalyzed by the PyOxim activating reagent.

3.4 Synthetic Strategy of the TIN2 Peptide Analogues

The chemical assembly of the TIN2_{TBM} peptide and the selected 12 peptide analogues was based on the computational results. Using the molecular modelling calculations, the $\Delta G_{\text{binding}}$ values between the peptide analogues and TRF1 protein were predicted and compared with the wild-type TRF1–TIN2 complex system. Details of the peptide analogues selection and the binding free energies were discussed in Sections 2.5 and 2.6. The TIN2 peptide and its analogues, which are linear peptides, were produced through applying the Fmoc/tBu SPPS method. The method depends on using an orthogonal-protecting-group strategy in which an acid removes the side-chain protecting groups and the linker support, while a base releases the α -amino-protecting groups.

All the synthesised peptides consisted of either 12 or 13 amino acid residues, except the labelled wild-type (TS01-L), which was composed of 13 amino acid residues with β -alanine (β -Ala) and the 5-FAM molecule (Scheme 3-6): β -Ala used as a linker between the *N*-terminal residue (Ser256) and the fluorescein molecule to separate 5-FAM from the interaction region of the peptide and the protein (for more details, see Section 2.13). The fluorescein molecule is a suitable reagent in the fluorescence-based assays. After assembly of the peptide sequence with β -Ala, the *N*-terminal of TIN2 was labelled using 5-FAM. The purpose of tagging the peptide with the 5-FAM molecule was to monitor the protein–peptide interactions²⁴¹. The mechanism of labelling TIN2 by the 5-FAM molecule is illustrated in Scheme 3-6.

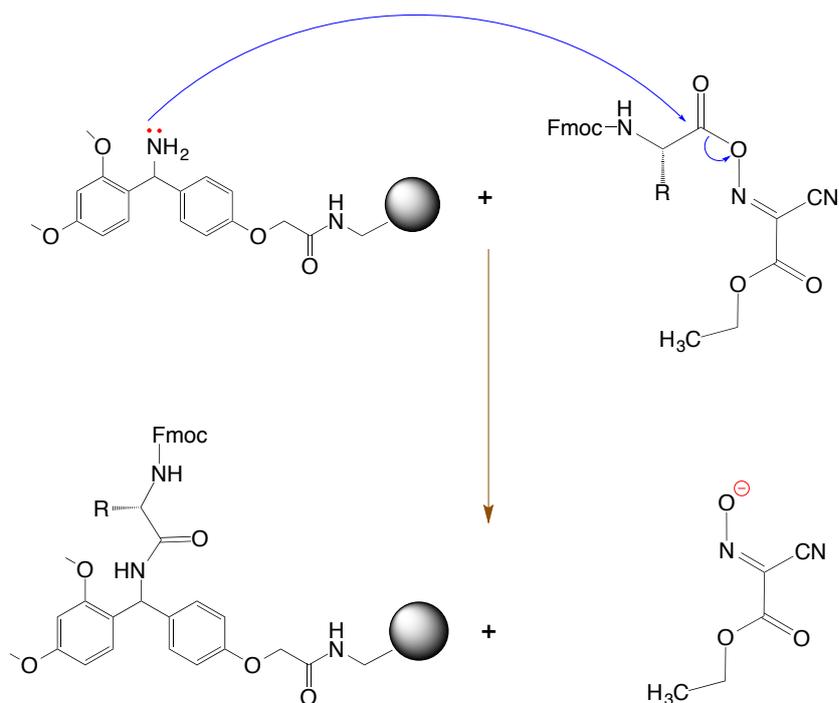


Scheme 3-6: Mechanism of tagging the *N*-terminal part of the TIN2 peptide by the 5-FAM molecule, using DIC and HOBt as coupling reagents.

3.4.1 Condensation of Fmoc-Amino Acids with Rink Amide

Novagel

The first step of assembling the TIN2 peptide and its analogues was the reaction between the linker and the α -carboxyl group of the C-terminal amino acid residue. Rink amide Novagel was selected as a resin linker in order to produce the peptide amide. The structure and coupling mechanism of the linker with an amino acid is explained in Scheme 3-7, which shows that the reaction is identical to the coupling of amino acids. After activation of the α -carboxyl group of the amino acid by the PyOxim reagent (Section 3.3.2), the nucleophilic functional group of the Rink amide linker (amino group) attaches the α -carbonyl of the activated amino acid to produce an amide bond.



Scheme 3-7: Coupling reaction of Rink amide Novagel with α -carbonyl group of an amino acid.

Hence, prior to coupling of the resin linker with the amino acid, the resin was swollen in the DMF overnight. The first Fmoc-amino acid and PyOxim were then dissolved in 0.6 ml DMF, followed by the addition of 2 equivalent (eq.) of *N,N*-diisopropylethylamine (DIPEA). The mixtures were loaded onto the swollen Rink amide Novagel in the reaction column and stirred gently for 3 hours at room temperature.

All the amino acids used were protected: α -amino groups were masked by the Fmoc group as a temporary protecting group, and side-chain functional groups were masked by semi-permanent protecting groups, such as tert-butyl (tBu) and trityl (Trt) that were not affected by reaction conditions during assembly of the peptide chains.

3.4.2 Peptide Assembly

After coupling the first amino acid with the resin, the temporary α -amino protecting group (Fmoc) was removed (deprotection) using 20% *v/v* piperidine in DMF for 7 min (3 ml min^{-1}) in a semi-automated peptide synthesiser. As illustrated in Scheme 3-2, monitoring the Fmoc deprotection semi-quantitatively was achieved through UV absorption of the dibenzofulvene–piperidine adduct at 350 nm.

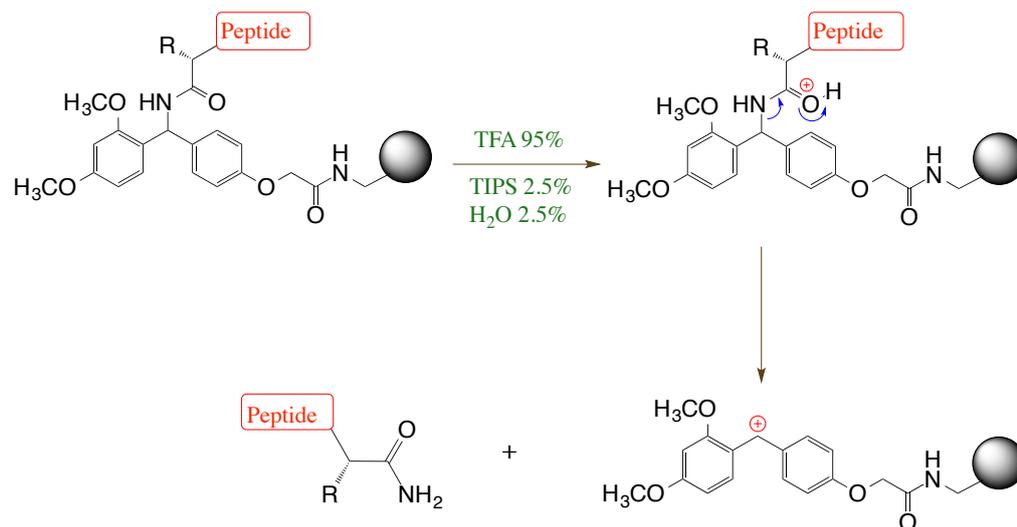
Subsequently, an excess of the next carboxy-activated amino acid was added to the growing peptide sequence (Scheme 3-5). After 3 hours, *via* the use of the semi-automated peptide synthesiser, washing and deprotection were performed

to remove the excess reagents and to cleave the Fmoc group, respectively. Each washing was achieved using 3 ml min⁻¹ DMF and lasted for 7 min.

For each coupling reaction, before loading the next amino acid, a solution of the carboxy-activated Fmoc-amino acid was prepared, which consisted of 4 eq. Fmoc-amino acid, 4 eq. PyOxim, 8 eq. DIPEA and 0.6 ml DMF. All of the steps discussed above were repeated until the desired sequences were produced. After assembling the desired peptides, the resin-peptides were washed with DMF, filtered, and washed with DCM and hexane, and finally, dried *in vacuo*.

3.4.3 Cleavage of the Linear Peptides

In order to deprotect and release the linear peptide analogues, the peptide-linker solid support and the semi-permanent protecting groups were cleaved using a TFA cocktail. Thus, the cocktail was composed of 95% TFA, 2.5% TIPS and 2.5% water²²⁰. The reaction was allowed to proceed for 3 hours at ambient temperature. The mechanism of cleaving the produced peptides from the Rink amide linker is shown in Scheme 3-8. The peptide products were filtered, evaporated using a rotary evaporator and triturated using diethyl ether, which produced a precipitated thin film of the peptide.



Scheme 3-8: Rink amide Novagel cleavage from the produced peptide through acidolysis.

3.5 Purification and Analyses of the Linear Peptides

The crude and purified peptides were analysed using reverse-phase high performance liquid chromatography (RP-HPLC) and mass spectrometry (MS). Analysis of the produced peptides using MS showed that all the m/z values corresponded to the $(M+3H^+)$ species (Table 3-2).

Table 3-2 Summary of the yield, RP-HPLC retention time (t_R), crude purity and both calculated and measured MS for the peptide analogues.

Name	Sequence	Yield (%)	t_R (min)	Crude purity (%)	MS	
					Calculated (M+3H ⁺)	Observed
TS01	S256-H-F-N-L-A-P-L-G-R-R-R-V268	96	5.9	78	507.9333	507.9683
TS01-L	5FAM-βA-S256-H-F-N-L-A-P-L-G-R-R-R-V268	93	7.4	78	651.3233	651.3334
TS02	S256-H- Y -N-L-A-P-L-G-R-R-R-V268	94	5.4	83	513.2900	513.2959
TS03	S256-H-F- L -L-A-P-L-G-R-R-R-V268	98	6.4	82	507.6400	507.6575
TS04	S256-H-F-N-L- T -P-L-G-R-R-R-V268	96	5.9	91	517.9633	517.9702
TS05	S256-H-F-N-L-A-P-L-G-R- P -R-V268	98	6.3	89	488.2766	488.2768
TS06	H257-F-N- G -A-P-L-G-R- P -R-V268	96	5.0	95	440.5800	440.5880
TS07	H257-F-N-L- N -P- N -G- K -R- S - L268	94	5.1	81	466.9166	466.0143
TS08	H257-F-N-L- N -P- N -G-R-R-R-V268	86	5.0	83	493.6033	493.6815
TS09	H257-F-N-L- N -P-L-G- K -R- S - L268	96	6.3	78	465.5966	465.6921

TS10	H257-F-N- L-N-P-N-G- K-R-S-V268	98	4.9	93	461.2466	461.3439
TS11	S256-H-F- N-L-R-P-L- G-R-R-R- V268	91	5.6	71	536.3133	536.3444
TS12	H257-Y-N- L-N-P-N-G- K-R-S-V268	96	4.5	80	466.5766	466.6737
TS13	S256-H-F- N-L-A-P-L- G-R-R-R- L268	94	6.0	68	512.6333	512.6854

Furthermore, 2D ^1H - ^1H correlation spectroscopy (COSY) nuclear magnetic resonance (NMR) analyses were used to characterise the wild-type TS01 and the analogue TS11. In this study, COSY NMR was chosen because of the overlapping multiplet signals. Figure 3.3 shows COSY NMR for TS01, which confirmed the presence of all of the secondary amide protons and their correlation with the adjacent α protons.

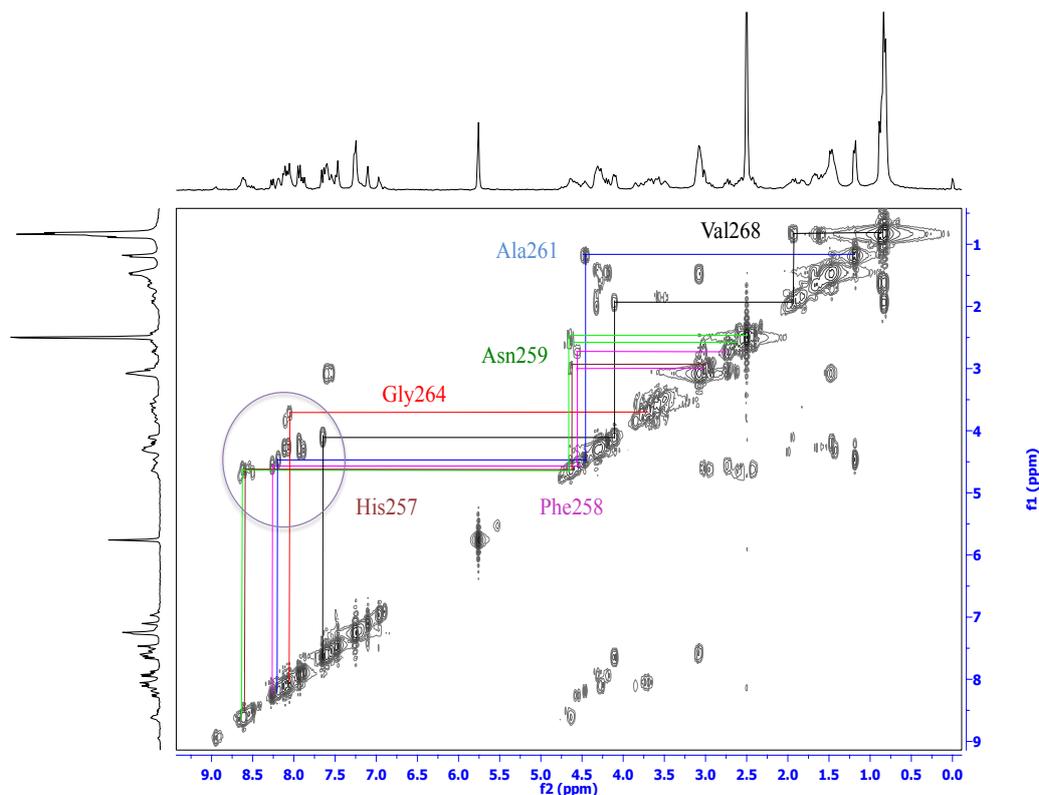


Figure 3.3 2D ^1H - ^1H COSY NMR spectrum of the TS01 peptide. All of the NH signals of the secondary amides are in the circle and the different coloured lines demonstrate the correlation of the proton signals in each amino acid residue.

COSY NMR also confirmed the amino acid residues of TS01 from the observations of the cross peaks of the NH signals for each residue and their correlated proton signals of the residue (Table 3-3). However, the method could not confirm Pro262 and Ser256 because Pro possesses a tertiary amide bond and Ser was at the *N*-terminal of the peptide. In addition, it could not precisely characterise the Arg residues due to the presence of three Arg residues in the peptide molecule.

Table 3-3 Proton chemical shifts for the amino acids of the TS01 peptide having the secondary amide signals.

Residue	NH	α H	β H	γ H	δ H
Val268	7.64	4.11 (1H)	1.95 (1H)	0.92 (6H)	
Arg	7.88	4.32 (1H)	1.41, 2.00 (2H)	1.94, 1.96 (2H)	4.00, 4.08 (2H)
Arg	7.93	4.31 (1H)	1.50 (2H)	1.5, 1.49 (2H)	3.11 (2H)
Arg	7.94	4.18 (1H)	1.46 (2H)	1.48 (2H)	3.07 (2H)
Gly264	8.04	3.70 (2H)			
Leu	8.07	4.26 (1H)	1.57 (2H)	1.56 (1H)	0.82 (6H)
Ala261	8.19	4.47 (1H)	1.19 (3H)		
Leu	8.11	4.28 (1H)	1.68 (2H)	1.70 (1H)	0.84 6H)
Asn259	8.65	4.64 (1H)	2.46, 2.58 (2H)		
Phe258	8.26	4.57 (1H)	2.72, 2.98 (2H)		
His257	8.59	4.62 (1H)	3.08 (2H)		

Moreover, COSY NMR was performed to characterise the structure of the TS11 peptide. According to the correlation map of the 2D COSY spectrum

(Figure 3.4), TS11 analogue signals were virtually identical to TS01 signals; all of the secondary amides were recognised from their α -proton cross signals. The proton chemical shifts for the different amino acids of the compound are given in Table 3-4.

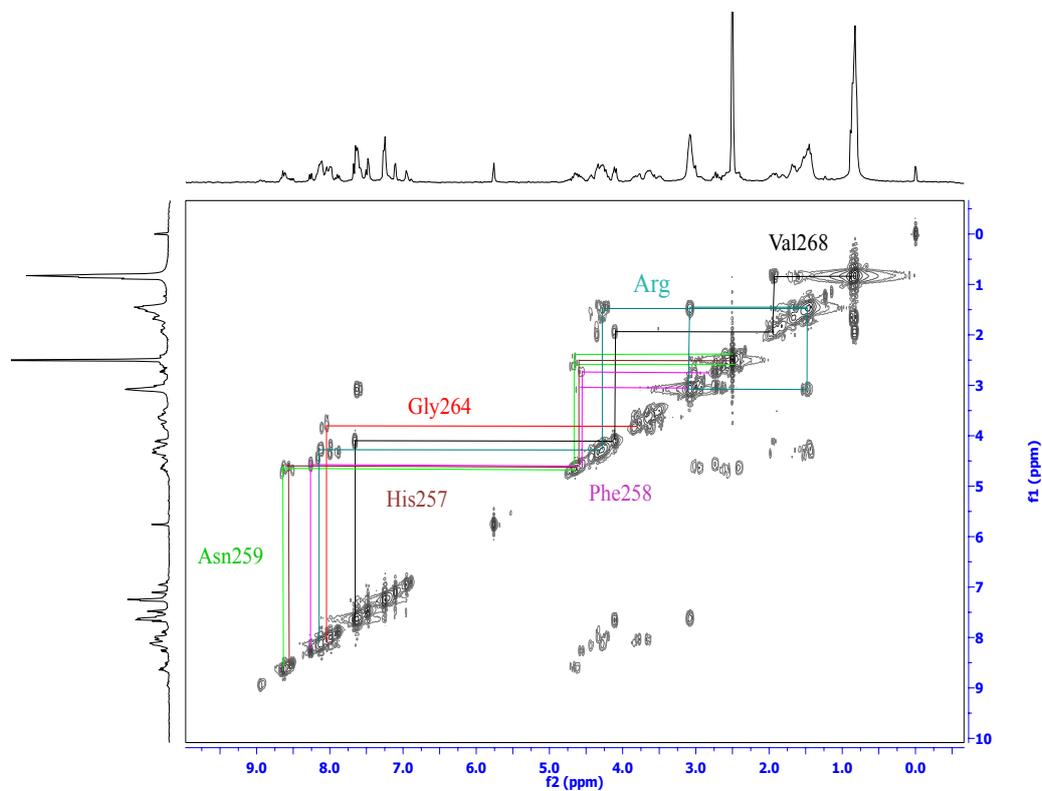


Figure 3.4 ^1H - ^1H COSY NMR spectrum of the TS11 peptide analogue. Coloured lines represent cross peaks of each amino acid.

Table 3-4 Proton chemical shifts for each amino acid of the TS11 peptide.

Residue	NH	α H	β H	γ H	δ H
Val268	7.65	4.10 (1H)	1.93 (1H)	0.84 (6H)	
Arg	7.88	4.34 (1H)	1.44, 1.99 (2H)	1.44 (2H)	4.32 (2H)
Arg	7.99	4.19 (1H)	1.43 (2H)	1.43 (2H)	4.19 (2H)
Arg	8.00	4.36 (1H)	1.40, 1.69 (2H)	1.51 (2H)	4.33, 4.31 (2H)
G264	8.04	3.80 (2H)			
L263	8.14	4.27 (1H)	1.47 (2H)	1.47 (1H)	0.82 (6H)
Arg	8.11	4.26 (1H))	1.48 (2H)	1.47 (2H)	3.08 (2H)
Leu	8.16	4.44 (1H)	1.53 (2H)	1.53 (1H)	0.82 (6H)
Asn259	8.64	4.66 (1H)	2.39, 2.58 (2H)		
Phe258	8.26	4.56 (1H)	2.75, 3.04 (2H)		
His257	8.56	4.60 (1H)	2.51 (2H)		

Specifically, the secondary amide proton signal at $\delta_{\text{H}} = 8.19$ ppm and its cross signals at $\delta_{\text{H}} = 4.47$ and 1.19 ppm for α and β protons, respectively, were assigned to Ala261 in TS01. In contrast, the secondary amide signal was

shown at $\delta_{\text{H}} = 8.11$ ppm and its cross signals at $\delta_{\text{H}} = 4.26, 1.48, 1.47$ and 3.08 ppm, respectively (Table 3-4) was assigned to Arg261 residues, which is specifically present in the analogue TS11.

4. Biophysical Evaluations for the Protein–Peptide Complexes

4.1 Introduction

Direct detection of receptor–ligand interactions plays a vital role in the study of biological systems and in new drug development processes because it can be used to determine the distribution of the receptors and identify their subtypes, to discover endogenous ligands, and to screen for and find new chemical entities as drug candidates ²⁴². The aim of using the receptor–ligand binding assay in this research has been to find a new molecular entity. Diverse biophysical approaches are available to investigate binding interactions and affinities between ligands and receptor proteins, such as the radioligand binding assay ²⁴³, isothermal titration calorimetry (ITC) ²⁴⁴ and fluorescence polarisation (FP) assay ^{245,246}.

Each type of the receptor–ligand binding assay is characterised by several advantages and drawbacks. For example, ITC is a reliable technique for determining the binding thermodynamics between receptors and ligands, such as peptide–proteins interactions. This method does not require the ligand’s labelling; however, it does consume a considerable amount of the ligand and has a low throughput capability ^{244,247}. Regarding radioligand binding assay, the most significant advantages of these binding assays are their

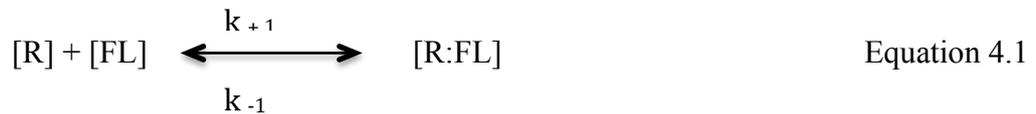
reproducibility, ease of use and the precision of the receptor-ligand binding affinities which is achieved because the labelling of the ligands by radioactive compounds does not interfere with the receptor-ligand binding interactions²⁴⁸. However, the disadvantages of radioligand assays are the requirement to separate bound and unbound ligands, the use of extremely hazardous radioactive chemicals, the associated waste disposal problems and the requirement for special laboratories. Additionally, this method is expensive and relatively slow²⁴². Due to the disadvantages of radioactive assays, non-radioactive assays have been developed, which are based on optical methods. These methods are usually preferred to the radioactive assays because they are less health hazardous, less expensive and create less environmental pollution. The FP assay is an example of a non-radioactive biophotonic technique, which is widely applied to the identification of the ligand-receptor bindings²⁴². In this project, FP assays were performed to recognise binding interactions between the peptide analogues and the TRF1 protein.

4.2 Introduction to the FP Assay

The term fluorescence polarisation refers to a process of monitoring molecular rotations and the factors that modulate their orientations when small molecules interact with macromolecules like protein or DNA molecule. The method can be applied in different fields, for example, for detecting receptor–ligand interactions, DNA–protein interactions, enzyme assays and competitive immunoassays^{249,250}. The FP assay is used essentially as a binding assay between a receptor and ligand in order to identify the binding affinity of the

ligand analogues to the receptor (competitive binding assay) or infrequently to obtain the dissociation constant (K_d) of the ligands. Our goal of using the FP assay is to identify the binding affinity between the TRF1 protein and the TIN2 peptide, and then the competitive bindings of the peptide analogues with the TRF1–TIN2 complex. Theoretical aspects and calculations of the receptor–ligand binding affinities are explained in the following equations (a colon represents a non-covalent bond).

Reversible interactions between two molecules are illustrated in Equation 4.1, which is a binding between the receptor [R] and the labelled ligand [FL] to form a complex.

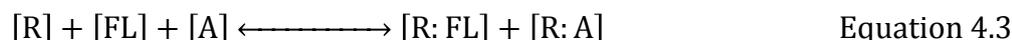


where k_{+1} and k_{-1} are the rate constants. The forward rate is $k_{+1}[R][FL]$, and the backward rate is $k_{-1}[R:FL]$. At equilibrium condition, both forward and backward rates are constant. K_d demonstrates the quantity of the ligand that saturates half of the binding sites, and it can be considered as an affinity of the interactions. At equilibrium, K_d is the ratio k_{-1}/k_{+1} , as shown in Equation 4.2.

$$K_d = \frac{k_{-1}}{k_{+1}} = \frac{[FL][R]}{[R:FL]} \quad \text{Equation 4.2}$$

When the ligand analogue molecule [A] is added to the complex [R:FL], two different receptor ligand complexes are produced, as the analogue [A]

displaces a number of the labelled ligands [FL]; the ratio of displacement depends on the affinity and concentration of [A] (Equation 4.3).



An inhibition curve can be generated when keeping the concentrations of [R] and [FL] constant, while varying the concentration of [A]. Through the inhibition curve, the inhibitory concentration 50% value (IC₅₀) can be obtained, which represents the concentration of [A] that displaces 50% of [FL] bound with [R]^{242,251}.

For calculating the IC₅₀ of the peptide ligands in the competitive binding assay, GraphPad Prism was applied. The titration curve was fitted using the log of the concentrations of the inhibitors versus response-variable slope (using four parameters) (Equation 4.4).

$$P = \frac{\text{min} + (\text{max} - \text{min})}{(1 + 10^{(r - cp) \times H})} \quad \text{Equation 4.4}$$

where cp is the log of concentration of the peptide analogues, r is the log of IC₅₀, P is the FP reading results, min is the minimum value of the sigmoidal curve, which represents the FP of the free-labelled peptide, max is the maximum value of the sigmoidal curve, which represents the FP of the labelled peptide bound to the TRF1 protein and H is the Hill slope, which is the slope factor.

4.3 Principles of the FP Assay

The FP assay starts with the attachment of a fluorophore to a small molecule (typically < 1500 Da). When a fluorophore-labelled molecule is free and excited by plane-polarised light, it will emit a long a different plane from the excitation light during the excitation lifetime of the fluorophore. In contrast, when a fluorophore labelled molecule binds with a high MW protein (typically > 10 kDa) and is excited by a plane-polarised light, it will to a large extent emit the identical polarised light. The emitted light is measured in two different planes: the first is the same plane of the excitation light; the second is perpendicular to the first plane. According to these observations, the cause of a depolarising free-labelled molecule is the rapid Brownian molecular reorientation during the excitation lifetime, whereas after it has bound with a protein, there is a decrease in the mobility of the labelled molecule during the excitation lifetime, which causes the emission of polarised light. In conclusion, the rate of the emitted polarised signal is inversely proportional to the rate of molecular rotation (Figure 4.1)^{241 252}.

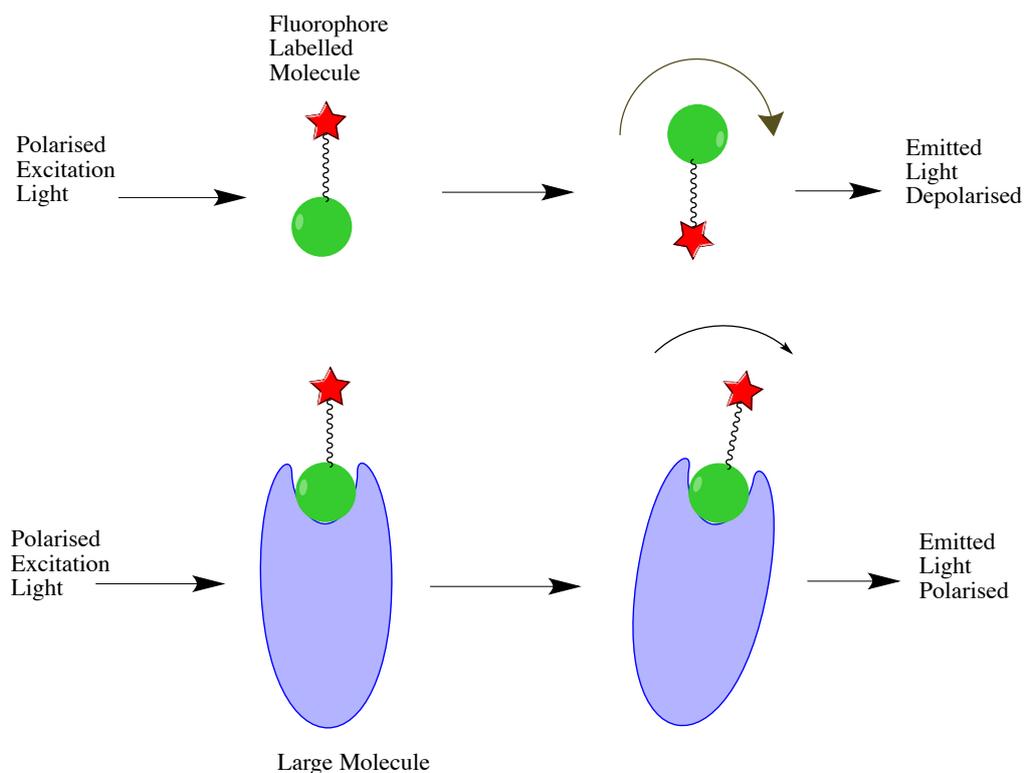


Figure 4.1 The difference of emitted light between an unbound labelled ligand and a bound labelled ligand with a receptor during the FP assay.

FP assays can be explained through Equation 4.6, which calculates polarisations (P) through vertically and horizontally polarised emissions:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \quad \text{Equation 4.6}$$

where, I_{\parallel} is intensity with the polariser parallel and I_{\perp} is intensity with the polariser perpendicular²⁵³. Usually, measured P is in the range of 0.01-0.3; therefore, it is multiplied by 1000, which is a millipolarisation (mP) unit.

4.4 Characteristics of the Fluorophore Molecule

Currently, the FP technique is widely applied and many studies has been focused on developing new chemical classes of fluorophores to increase their solubility, improve their physical properties, optimise their brightness (intensity of fluorescence = molar extinction coefficient x quantum yield) and enhance their photostability⁷⁸. Therefore, to achieve a successful FP assay, a fluorophore molecule should have: a high quantum yield, a high extinction coefficient (ϵ) to increase sensitivity in aqueous media for detection, good photostability, a high excitation wavelength to inhibit autofluorescence and suitable molecular size. Furthermore, using a spacer between the ligand and the fluorophore molecule is essential to reduce the likelihood of inducing steric hindrance caused by the attachment of the bulky fluorophore molecule. Finally, the position of tagging fluorophores needs to be in a location that does not interfere with the receptor–ligand interaction²⁴².

4.5 Advantages and Disadvantages of the FP Assay

The FP assay has a number of advantages over other methods when it is used to identify receptor–ligand interactions. First, it can be easily automated for use in ultra-high-throughput screening (uHTS)^{250,254}. Second, the assay is relatively inexpensive because the plate reader can be used for different types of HTS assays^{255,256} and miniaturisation can decrease the cost of the assay²⁵⁷. Third, it is unlike other fluorescence assays in only needing one labelling step. Fourth, it is a homogenous assay (mix-and-measure) because it does not

require separation of bound and free ligands, while radiological assays do require this separation and therefore entail filtration, centrifugation and dialysis before the samples can be read, which is time-consuming and may disturb the reaction equilibrium²⁴². Fifth, the FP assay has significantly lower health risks and fewer safety issues. Finally, the chemicals can be kept for a long period of time²⁵².

However, the FP assay does have several drawbacks. The first concerns the size of interacting molecules. Typically, the size of the ligand is < 1500 Da and the size of the receptor > 10 kDa; nevertheless, a ligand size of up to 5000 Da could be used if the receptor has a very large MW^{241,257}. The second drawback is that the labelled fluorophores may involve interactions between the ligand and the receptor^{242,257}. Finally, the demand for protein is higher in an FP assay relative to a radioligand assay because of the need to titrate proteins against a fixed concentration of the ligand; however, it still requires less amount of protein compared with the ITC technique²⁵¹.

4.6 FP Assay of the TRF1–TIN2 Peptide Analogues

Before performing the FP competitive binding assays between the TRF1 protein and the TIN2 peptide analogues, the method was developed and optimised by: choosing a suitable fluorophore molecule as well as a site of labelling fluorophore to the TIN2 peptide; investigating the sensitivity of the plate reader to the fluorophore-labelled TIN2; and identifying the required concentration of TRF1 protein for the assays.

4.6.1 Fluorophore Selection for the FP Assay

Different types of fluorophore molecules have different excited state lifetimes, and choosing a fluorophore with an appropriate excitation lifetime for a specific type of FP assay can be considered to be one of the critical factors for a successful FP assay²⁵⁰. Theoretically, to achieve FP measurements between the ligand and the receptor, the binding should be between a low MW ligand and a much higher MW receptor (substrate), while in practice, as discussed above, the bindings are typically between ligands of approximately ≤ 1500 Da and receptors of approximately ≥ 10 kDa; this limitation is due to the short excited-state half life of the fluorophore molecules, which need a consistency between the fluorescence lifetime of the fluorophore and the rotation of the labelled ligand-receptor complex²⁵⁸. Fluorophore molecules such as Fluorescein and rhodamines have lifetimes of approximately 4 ns which can allow an obvious FP separation between the free-labelled ligand and the labelled ligand-receptor complex when the size of both molecules are within a range discussed above. In this study, the MW of the peptide analogues was approximately ≤ 1500 Da and the MW of the TRF1 protein was approximately 50 kD. The selected fluorophore was 5-carboxyfluorescein (5-FAM) because its excitation lifetime is 4 ns, which was consistent with the MW of the ligands and receptor. In addition, the maximum absorption and emission of the most plate-reader filters are at 492 nm and 518 nm, respectively; the EnVision[®] plate reader, which was used in this FP assays has an excitation wavelength at 480 nm and an emission wavelength at 535 nm. Finally, 5-FAM is inexpensive compared with the other fluorophores^{259,260}.

4.6.2 Selecting the Fluorophore Tagging Site

The 5-FAM molecule can be labelled at the *N*-terminal, C-terminal or inside the peptide; however, the label should be in a position such that it cannot interfere with the peptide-protein binding site. Furthermore, tagging a flexible region of the peptide or having a long linker between the peptide and 5-FAM may affect the FP measurement²⁴¹. Therefore, based on the structural data, the *N*-terminal of the TIN2 peptide was selected, as it is outside the binding area of the TRF1–TIN2 complex¹⁹². Also, β -Ala was chosen as a linker between 5-FAM and TIN2 to ensure that 5-FAM was positioned safely outside of the binding region. MD simulations were then carried out to confirm that the labelled compound could not interfere with the binding interactions between the TRF1 protein and the TIN2 peptide (details in Section 2.13).

4.6.3 Detection Sensitivity of the Instrument for the Labelled TIN2 Peptide

One of the vital steps before starting the FP assays was to determine the sensitivity of the plate reader for the synthesised fluorescent tracer (TS01-L) in order to select the minimum concentration of it, which could be detected by the plate reader^{241,261}. This step was achieved by using different concentrations of TS01-L, and identifying the concentration above which there was no further change in the polarisation signal and which was above the sensitivity of the instrument. As demonstrated in Figure 4.2, different concentrations of TS01-L were applied, starting at 10 μ M and going down to 10 nM. The signals of

concentrations below 100 nM (30 nM and 10 nM) different from the other concentrations: the mP readings increased with decreasing concentration, which indicated these concentrations were below the detection limit of the instrument. Thus, the lowest fluorescent detection limit of the multi-well plate reader for the TS01-L compound was 100 nM and therefore, this concentration of the labelled tracer was identified as a fixed concentration for the future FP assay. The selected concentration should not be higher than twice the K_d value ($K_d = 0.31 \mu\text{M}$) because it may lead to stoichiometric titration and decrease the sensitivity of the assay^{192,241}. The selected concentration was close to the K_d value.

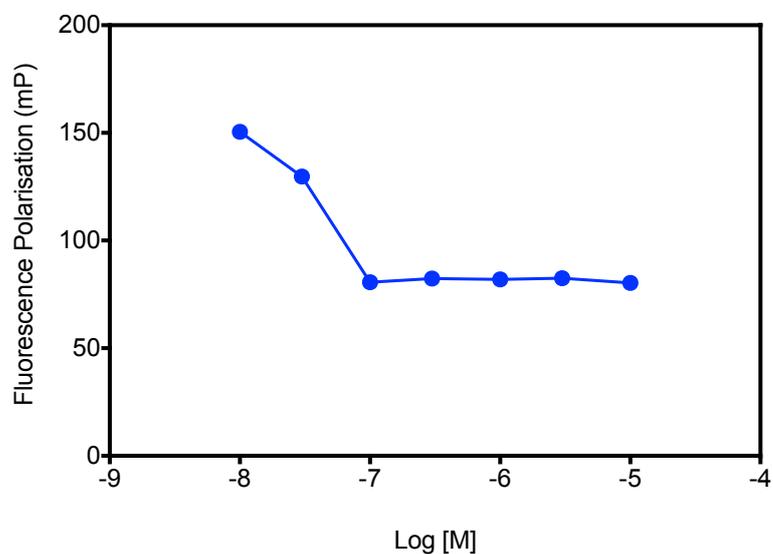


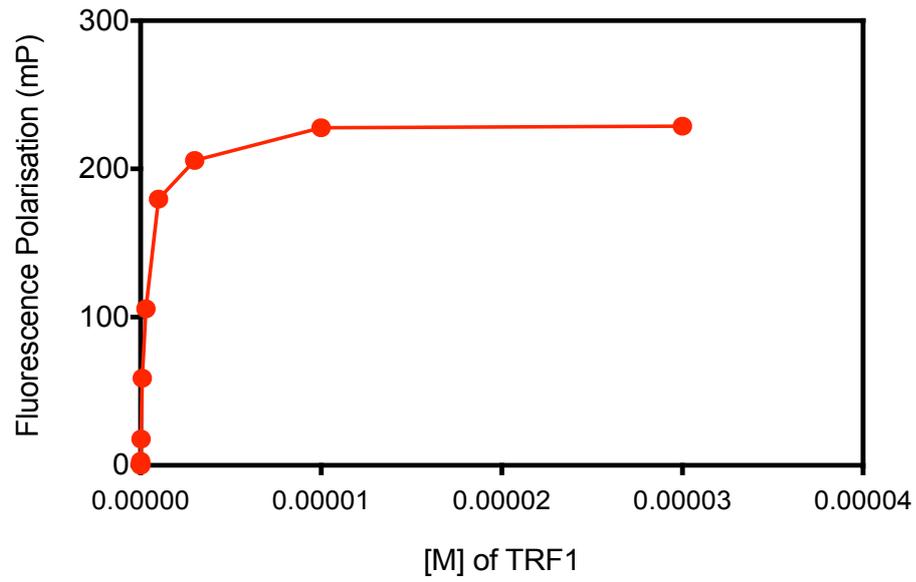
Figure 4.2 The polarisation signals (mP) for the different concentrations of TS01-L.

4.6.4 Determining the TRF1 Concentration

After confirming the TS01-L concentration, the next step in developing the assay was to determine the lowest possible concentration of TRF1 that should be used in the experiment to avoid wasting valuable material, nevertheless, high enough to ensure that most of TS01-L is bound before adding competitors. This step was achieved by maintaining a fixed concentration of the labelled probe (TS01-L) and measuring changes in the polarisation signals whilst titrating with different concentration of the TRF1 protein^{241,261}. The aims of this measurement were not only to find the optimum protein concentration for the future assays, but also to demonstrate the protein-peptide binding interactions. In this experiment, the different concentrations of TRF1 protein (1 nM–30 μ M) were titrated against a fixed concentration of TS01-L (100 nM); and the polarisation readings were plotted against the TRF1 concentrations to determine at what point the curve reached a plateau, this point being due to a direct binding of the all labelled ligand with the protein (Figure 4.3A).

The data were then replotted using the log of the TRF1 concentrations, and were expressed as a sigmoidal curve (Figure 4.3B). The baseline of the curve represents the free TS01-L; the plateau (220 mP) represents the total binding of TS01-L with TRF1; between these points there is increasing polarisation due to increased binding between TS01-L and TRF1.

A)



B)

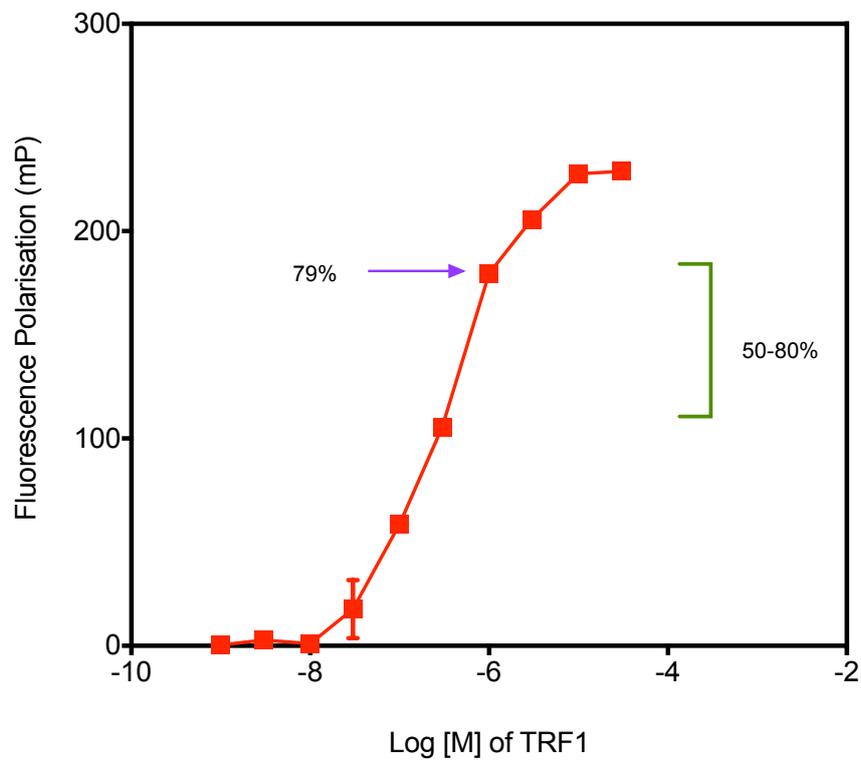


Figure 4.3 Direct binding of 100 nM TS01-L with different concentrations of TRF1 protein. (A) FP versus TRF1 concentrations. (B) FP versus log of TRF1 concentrations.

The sigmoidal binding curve shown in Figure 4.3B was used to choose the optimum concentration of the TRF1 protein. The selected concentration needed to be within a range of 50–80% of increasing polarisation measurements from the free-labelled peptide to the completely bound state (plateau) because, on the one hand, selecting protein concentrations $< 50\%$ would not show an obvious assay window when the labelled peptide was displaced by the unlabelled competitor compounds, and on the other hand, concentrations $> 80\%$ would not be sensitive to competitors²⁴¹. Therefore, $1\ \mu\text{M}$ of the TRF1 protein was determined for the assays, which corresponded to 79% of increasing polarisation from the baseline to the plateau.

4.6.5 Competitive Binding of TS01 and TS01-L with the TRF1

After determining the concentrations of the fluorescent tracer and the protein, a standard competitive binding assay was undertaken between the labelled and the unlabelled TIN2 peptide with the TRF1 protein. This assay was applied to confirm that the TIN2 peptide was interacting with a specific TRFH domain of TRF1 and that the FP assay was working properly. The assay is based on measuring the decrease of mP directly as the concentration of the unlabelled peptide – which is added to the mixture of the labelled peptide and the protein – is increased. The reason of decreasing mP is that increasing the concentration of the unlabelled peptide leads to increase displacement of the bound labelled peptide from the protein. To achieve this assay, different concentrations of TS01 (10 nM–300 μM) were titrated against a fixed concentration of TS01-L and TRF1 protein (100 nM and $1\ \mu\text{M}$, respectively) in the plate wells. By

increasing the concentration of TS01 added to the complex of TS01-L and TRF1, the polarisation signals were decreased due to displacement of TS01-L by the TS01 peptide. The sigmoidal curve shows the polarisation signal starts to inhibit from increasing the TS01 concentration at 100 nM and it displaces all of the TS01-L at around 300 mM (Figure 4.4).

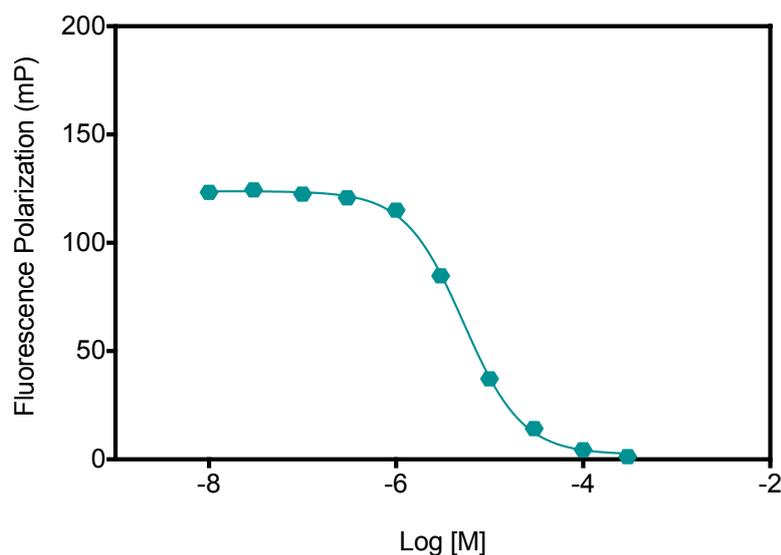


Figure 4.4 Competitive binding assay between the labelled and unlabelled TIN2 peptide with TRF1 protein.

4.6.6 Competitive Binding Assay of the Labelled TIN2-TRF1 with the Peptide Analogues

After the successful competitive FP assay of TS01-L and TS01 with the TRF1 protein, all of the peptide analogues (12 peptides) were screened to obtain their binding affinities. Each peptide analogue was titrated against the mixture of TS01-L (100 nM) and TRF1 protein (1 μ M). Decreasing mP signals for each

peptide assay were observed as the concentration of the peptide analogue increased because TS01-L was gradually being displaced by the peptide analogues. However, different concentrations of the various peptide analogues were required to displace TS01. As shown in Figure 4.5, reducing FP intensities were different for the different peptide analogues while using the same concentration scale; it is reasonably certain that these results were due to the diverse potencies of the peptide analogues.

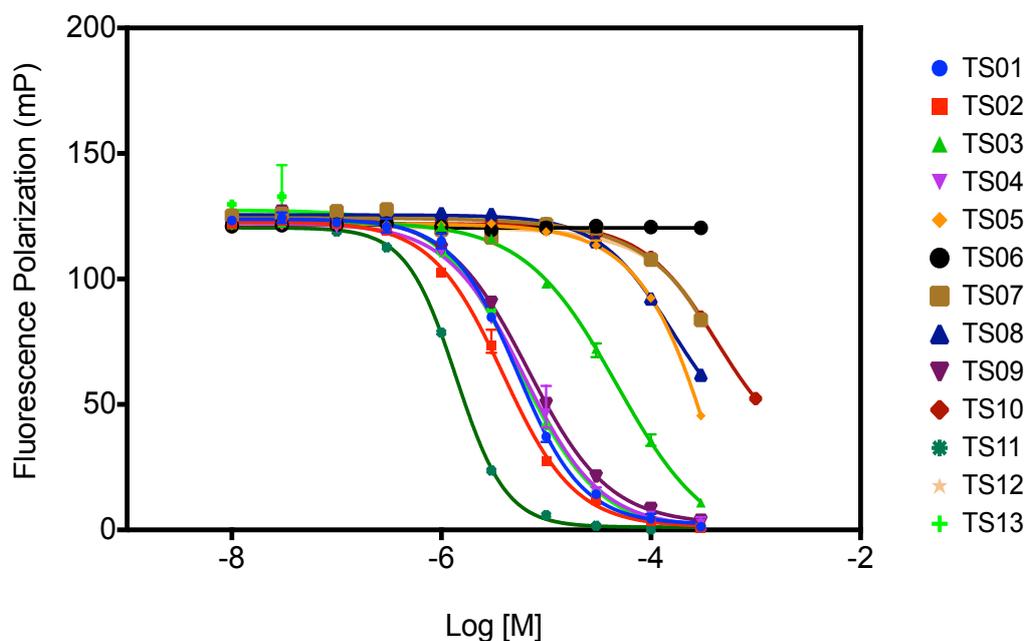


Figure 4.5 Competitive binding assay of TS01 and the 12 peptide analogues with the complex of labelled TIN2 and TRF1 protein.

The results of the FP assays for each competitive binding assay were demonstrated by a sigmoidal curve: the top of the curve denoted the total binding of TS01-L with TRF1; the bottom of the curve showed the total displacement of TS01-L by the competitor peptide; the mid-point of the curve

showed the concentration at which half of the TS01-L was displaced by the competitor ligand, which was IC_{50} . In pharmacology and medicinal chemistry, IC_{50} is commonly used to indicate the potency of a compound or drug antagonist *in vitro*, as it is a quantitative measurement, i.e. the concentration of a compound that will inhibit 50% of the biological process being considered. Hence, IC_{50} was obtained for each of the peptide analogues to investigate the inhibitory activity of the analogues and obtain an early lead compound. All the IC_{50} results are presented in Table 4-1. However, the assays were performed in triplicate (Section 6.3.4), the standard errors of mean for the peptide analogues TS07 and TS12 are not exist because IC_{50} of these analogues was obtained in one of the experiments owing to the uncompleted sigmoidal curves in the remaining two assays.

The IC_{50} of the peptide analogues were compared with TS01 because the latter is a wild-type compound. According to the modelling results, both of the compounds TS05 and TS06 were predicted to be negative controls, as they had significantly lower $\Delta G_{\text{binding}}$ values compared with TS01; here also, it has been clearly demonstrated that they had dramatically higher IC_{50} values (625 μM and 928 μM , respectively).

Table 4-1 Results of IC_{50} for the TIN2 analogues obtained from the FP assays.

Compound Name	IC_{50} (μ M)	Log IC_{50} (M)
TS01	5.04 ± 0.3	-5.30 ± 0.03
TS02	3.97 ± 0.1	-5.40 ± 0.01
TS03	41.41 ± 3.8	-4.39 ± 0.04
TS04	5.81 ± 0.3	-5.24 ± 0.02
TS05	625.00 ± 27.8	-3.15 ± 0.06
TS06	948.20 ± 23.5	-3.02 ± 0.01
TS07	428.30	-3.368
TS08	268.70 ± 10.5	-3.57 ± 0.02
TS09	6.24 ± 0.6	-5.21 ± 0.05
TS10	681.95 ± 9	-3.17 ± 0.01
TS11	1.40 ± 0.1	-5.85 ± 0.03
TS12	118.00	-3.93
TS13	5.04 ± 0.7	-5.436 ± 0.07

Only two peptide analogues had lower IC_{50} values compared with TS01, these being TS02 ($IC_{50} = 3.97 \mu$ M) and TS11 ($IC_{50} = 1.4 \mu$ M). TS02 had a slightly higher binding affinity than TS01, while TS11 was significantly more potent than the wild-type, as obviously verified in the sigmoidal curve (Figure 4.6). All of the other peptide analogues had either approximately similar or lower binding affinities compared with TS01. Therefore, we can conclude that according to the FP results, TS11 is the only peptide analogue with a

significantly higher binding affinity relative to TS01 and can be considered as an early lead compound.

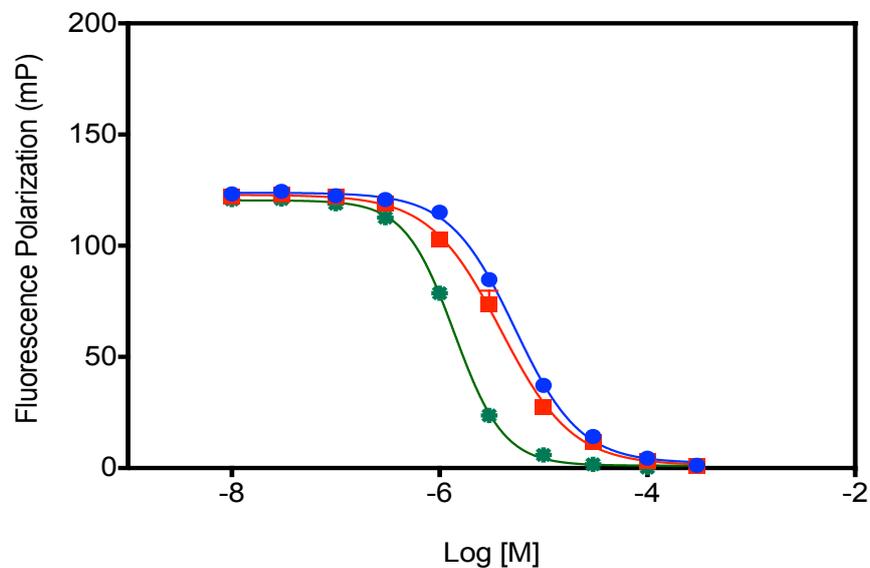


Figure 4.6 Competitive binding assay of the peptides TS01, TS02 and TS11 with the mixture of TS01-L and TRF1 protein. (blue curve = TS01; red curve = TS02; green curve = TS11).

5. Conclusions and Future Work

The essential aim of this study was destabilising PPIs in the telomeres of the cancer cells. Choosing this target rather than telomerase or telomeric DNA (G-quadruplexes) was based on the literature data. Telomerase inhibitors have a delayed onset of action and they could produce senescence only in p53 active cells^{40,41,262}. In contrast, G-quadruplex ligands could overcome the problem of delayed onset of action of telomerase inhibitors because they may directly disrupt and destabilise the shelterin proteins of telomeres^{41,50,51,263}. In addition, *in vivo* results demonstrated the significant selectivity and sensitivity of the G-quadruplexes for the cancer cells⁴². Hence, directly inhibiting the PPIs of the shelterin components may be an appropriate choice and a new method of targeting cancer cells.

Inhibiting PPIs between TRF1 and TIN2 proteins was the target chosen in this work because they are crucial for the stability of telomeres. TRF1 is essential to maintain the consistency and chromosomal stability of the shelterin components²². Furthermore, according to Pal *et al.* (2015), overexpressing TRF1 and TRF2 have been recognised in the renal cancer cells; therefore, silencing their expression have been achieved, which led to reduce the number and proliferation of the treated cells significantly compared with the control cells. This observation was explained to be related to the cell cycle arrest, consequently, telomere destabilisation and apoptosis of the cancerous renal cells⁵⁷. A recently published article by Garcia-Beccaria *et al.* (2015) reported

the potential effect of TRF1 inhibition on lung cancer cells when p53 was inactivated and the high degree of selectivity for cancer cells⁵⁹.

Peptide molecules have been selected to target the PPIs between the TRF1–TIN2 complex. A library of TIN2 peptide analogues (53 compounds) has been designed *in silico* to compete against TIN2 protein efficiently on TRF1, inhibit PPIs and destabilise the telomeres. Then, 10 ns MD simulations were carried out and the MM-GBSA method was used to calculate the binding free energy of each peptide analogue–TRF1 complex system. MM-PBSA/GBSA is one of the most common methods for the prediction of drug–target binding affinities in drug design projects; however, its main weakness is calculating absolute binding affinity values because of the difficulty in predicting entropy changes.

Typically, entropy can produce $\Delta G_{\text{binding}}$ values closer to the experimental values, but when calculated in MM-PBSA/GBSA method, it may dramatically overestimate the $\Delta G_{\text{binding}}$ values^{264,265}. Hence, predicting the absolute values of the binding affinity may be unrealistic when utilising MM-PBSA/GBSA methods, whereas their accuracy can be significantly improved when neglecting entropy issues and determining relative affinities of similar ligands or conformers^{157,183,266}. In addition, for relative binding free energy predictions, MM-GBSA seems often to be more accurate than MM-PBSA¹⁸¹.

Predicted values of $\Delta G_{\text{binding}}$ for a wide range of candidate peptide–protein complex systems varied over a broad range of over 50 kcal/mol ($\Delta G_{\text{binding}}$ value for P04 = -102.31 kcal/mol while for P25 = -50.37 kcal/mol) (Table 2-2).

It was impracticable to investigate and test all of the peptide analogues. Therefore, 13 peptides were selected from the library of peptide analogues to perform extensive computational analysis and optimisation of their binding free energies, chemical synthesis and biophysical assays. The selected peptides consisted of the wild-type TIN2 peptide as a reference compound, two peptide analogues as estimated negative controls, three peptide analogues produced from TIN2–Apollo cross-matching mutations, and seven peptide analogues that have been predicted to have significantly higher binding free energy compared with the wild-type and might compete effectively with TIN2 peptide.

Replicate MD simulations for the wild-type system led to an interesting observation, which was significantly different $\Delta G_{\text{binding}}$ value between replicates (the difference was 10 kcal/mol). After that, one question that needed to be addressed was whether the $\Delta G_{\text{binding}}$ values predicted from the single MD simulations are sufficiently accurate to depend on? There are two likely causes for the poor convergence of the MD simulations; it might be related to the trapping of the system in local minima for a long period of simulation time and/or insufficient sampling of the conformational space.

In order to improve the accuracy of the results, MD simulations were repeated 50 times for all of the complex systems (using the same protocol and starting structure, but different initial velocities) and the average $\Delta G_{\text{binding}}$ values for the replicates were calculated. Then, the protocol for the calculation of binding free energies of the complex systems were optimised, which was started by testing the statistical distribution of the replicate values for each system (Figure

2.11); subsequently, the required number of replicates for each system was calculated through using the sample size equation, which were different according to their margin of error (Table 2-4).

After calculating average $\Delta G_{\text{binding}}$ value for the complex systems, they were compared with the wild-type value as a reference values; most of the $\Delta G_{\text{binding}}$ values were within the margin of error for the wild-type peptide, except compounds TS05, TS06, TS08 and TS11, which had significantly different binding affinity values. Both TS05 and TS06 had significantly lower values, which were predicted to be negative controls, while TS08 and TS11 were expected to be lead compounds.

Then, the optimum time of MD simulations was determined through two different approaches. The first was calculating $\Delta G_{\text{binding}}$ at each time point of the MD simulations for the 50 replicates; this strategy was tested just for the wild-type system because it was expensive computationally. According to this approach, the system was equilibrated after approximately 4 ns of MD simulations. The second was calculating RMSD for all of the replicates of the complex systems; for each peptide–protein system, the average RMSD value of the replicates at each time point was taken. The RMSD was calculated for all peptides atoms and all TRF1 residues within 5 Å of the ligand, because the greatest binding interactions were in that distance range. As shown in Figure 2.11, the complex systems approached equilibration after approximately 4–5 ns MD simulations; however, a number of the systems trend with respect to time were to increase RMSD gradually, which means the systems were not yet completely equilibrated.

Subsequently, the 12 designed 13-mer peptide analogues and the TIN2 peptide were synthesised using Fmoc SPPS. Rink amide Novagel was used as a resin linker to produce the peptide amide, which can decrease the charge of the peptides, increase cell permeability and increase resistance against enzymatic degradations. PyOxim was used as an activating reagent, as it is a great choice for SPPS because it can significantly inhibit racemisation, improve reactivity, has less explosive capability and is highly soluble in organic solvents.

The last stage of this project was applying FP as a biophysical assay to obtain the IC_{50} of the peptide analogues. The reasons for selecting this assay, as discussed in Chapter 4, were utility to identify protein–peptide interactions, easy automation to use in HTS, relative economy because a plate reader instrument can be used for different tests and miniaturisation of the assay, homogenous assay and significantly fewer safety problems comparing with radiolabelled assays. Furthermore, we could overcome the main limitations related to this assay (the size difference between the ligand and the receptor) because of the huge gap between the size of the peptides and the protein: the size of the peptide analogues did not exceed 1600 Da and the size of TRF1 is approximately 50 kDa. Moreover, TIN2 was suitable to be labelled by 5-FAM and according to the molecular modelling results, it would not interfere with the interacting regions of TRF1–TIN2. Finally, abundant and pure TRF1 protein was prepared for the assay.

This assay was developed to screen for peptide analogues against TRF1_{TRFH}–fluorescent tracer (TS01-L) mixture, which was the first demonstration of an

FP assay for TRF1 protein and then it can be used in high-throughput (HT) format. The assay developed here was applied to identify a lead compound. The initial development of the assay was time-consuming as usual; however, the developed methodology of this assay could be used for the future screening of chemical libraries against TRF1–TIN2 interactions and easily automated.

After optimising the methodology for the assay and before screening the peptide analogues, the unlabelled wild-type peptide (TS01) was titrated against the mixture of TRF1 and the labelled wild-type peptide (TS01-L) to confirm binding of TS01 with the TRFH domain of the TRF1 protein and that the assay was working properly; this titration is termed a “Gold-standard assay”²⁴¹. Subsequently, a competitive binding assay for the peptide analogues against the mixture of TRF1 and TS01-L was done, which led to obtaining IC₅₀ of the peptide analogues. The lowest IC₅₀ values were obtained from two peptide analogues with a single residue mutation, which were TS02 (TIN2-F258Y) and TS11 (TIN2-A261R) with 3.97 μM and 1.4 μM, respectively, and they had lower IC₅₀ compared with TS01 (IC₅₀ = 5.04 μM). However, TS11 had a significantly higher binding affinity than TS01, as it is approximately five-fold more potent than TS01 (Table 5-1); hence, it can be considered as an early lead molecule.

Table 5-1 Predicted and experimental affinity of the TIN2 peptide analogues.

Compound	Sequence	IC ₅₀ (μM)	Average predicted ΔG _{binding} (kcal/mol)
TS01	S256-H-F-N-L-A-P-L-G- R-R-R- V268	5.04 ± 0.3	-85.07 ± 0.72
TS02	S256-H- Y -N-L-A-P-L- G-R-R-R- V268	3.97 ± 0.1	-84.61 ± 0.81
TS03	S256-H-F- L -L-A-P-L-G- R-R-R- V268	41.41 ± 3.8	-82.69 ± 0.77
TS04	S256-H-F-N-L- T -P-L-G- R-R-R- V268	5.81 ± 0.3	-87.98 ± 0.95
TS05	S256-H-F-N-L-A-P-L-G- R- P -R- V268	625.00 ± 27.8	-60.99 ± 0.77
TS06	H257-F-N- G -A-P-L-G- R- P -R-V268	948.20 ± 23.5	-51.42 ± 0.68
TS07	H257-F-N-L- N -P- N -G- K-R-S-L268	428.30	-81.95 ± 1.03
TS08	H257-F-N-L- N -P- N -G- R-R-R-V268	268.70 ± 10.5	-94.94 ± 1.11
TS09	H257-F-N-L- N -P-L-G- K-R-S-L268	6.24 ± 0.6	-88.92 ± 0.86
TS10	H257-F-N-L- N -P- N -G- K-R-S-V268	681.95 ± 9	-82.96 ± 1.26
TS11	S256-H-F-N-L- R -P-L-G- R-R-R-V268	1.40 ± 0.1	-95.49 ± 1.21
TS12	H257- Y -N-L- N -P- N -G- K-R-S-V268	118.00	-81.99 ± 1.2
TS13	S256-H-F-N-L-A-P-L-G- R-R-R- L268	5.04 ± 0.7	-89.00 ± 0.7

As illustrated in Table 5-1, the *in silico* and experimental results show that most of the amino acid residue mutations lead to inhibition of the binding affinity between the ligands and the TRF1 protein because of changes in the conformations of the peptides that affect other binding interactions, or as a result of the entropy penalties. The only two mutations that produced positive results are the F258Y and A261R in TS02 and TS11, respectively.

The second aim of this study was to investigate and optimise the accuracy of $\Delta G_{\text{binding}}$ by implementing MM-GBSA methods. The results of single and average (optimised) predicted binding free energy for the peptide analogues were reported in Chapter 2 and the experimental results were reported in Chapter 4. To reveal the accuracy of the single and average predicted results, the correlation between each set of the predicted values and the experimental values was tested. The results of the correlational analysis are displayed in

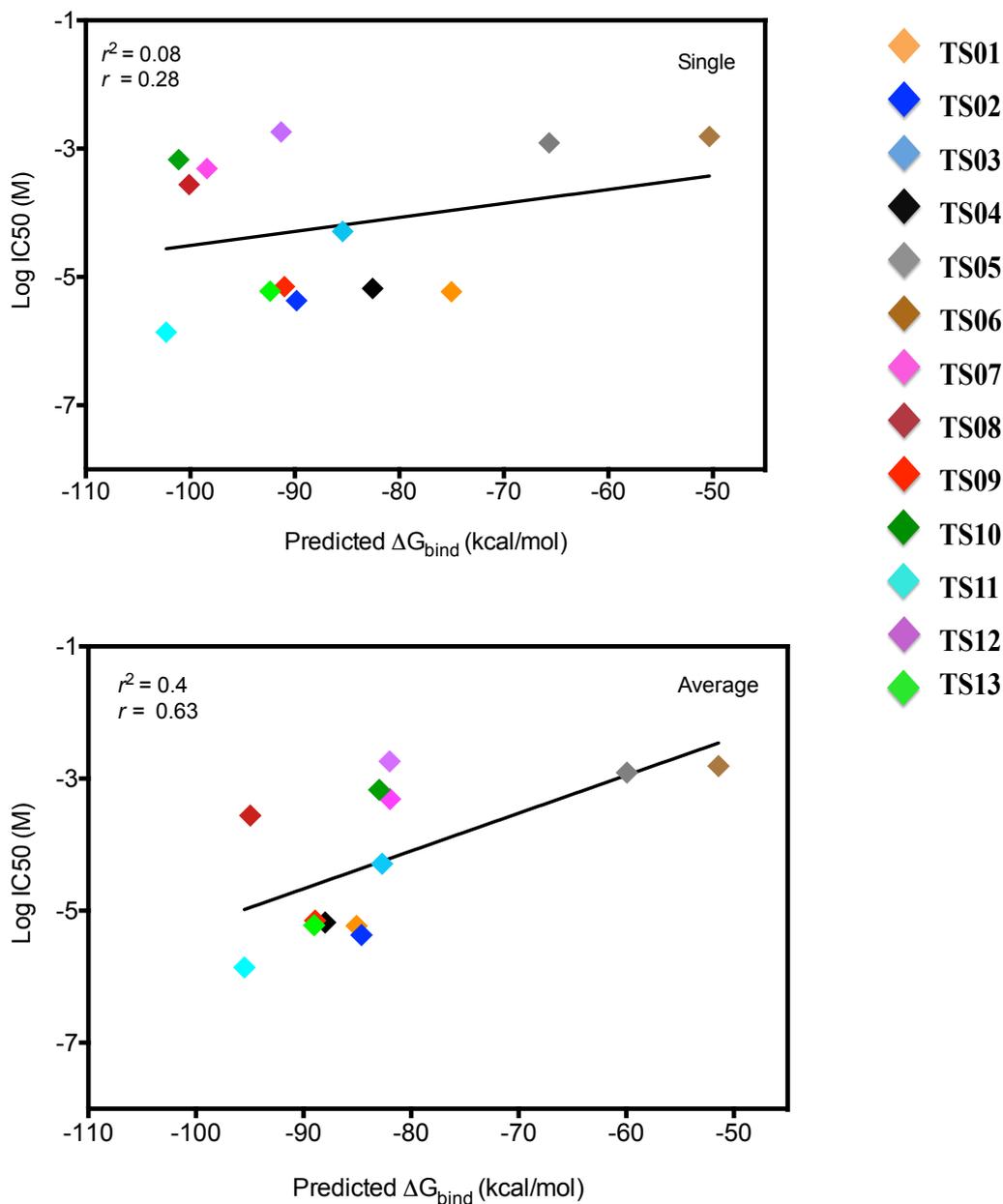


Figure 5.1 Correlation between $\Delta G_{\text{binding}}$ results and experimental results. (A) Correlation between single $\Delta G_{\text{binding}}$ value of the peptide analogues and their IC₅₀. (B) Correlation between average $\Delta G_{\text{binding}}$ results of the peptide analogues and their IC₅₀ results.

What stands out in the Figure 5.1 is the difference between the accuracy of the single predicted result and the average predicted results, when after optimising

the predicted binding free energy values, the coefficient of determination (r^2) significantly increase and improve correlation from 0.08 to 0.40 between the predicted $\Delta G_{\text{binding}}$ values and the experimental data in the form of IC_{50} values. The above can explain the general accuracy of the predicted values; however, after dissecting the predicted and experimental value of each peptide, various relative relationships can be found. For example, both of the negative controls peptide analogues (TS05 and TS06) have significantly lower predicted and experimental binding affinities compared with those for TS01 (Table 5-1), which is due to mutating the critical residues that could highly interact with TRF1 residues. In addition, predicted and experimental binding affinity values in TS06 are significantly lower than those for TS05 because, as well as mutating R266P, it has a second mutation: L260G. Another example is TS11; according to the molecular modelling results, it has the highest predicted binding affinity (-95.49 kcal/mol) and it is identical in the experimental result, as it has the highest binding affinity ($IC_{50} = 1.4 \mu\text{M}$). These examples show the significantly high accuracy of the computational results.

In contrast to the above examples, some of the peptide analogues' computational and experimental results are significantly different. The most obvious example is TS08, for which the results are not very encouraging, as the predicted $\Delta G_{\text{binding}}$ value is -94.94 kcal/mol, which is significantly higher than TS01, but the experimental binding affinity is significantly lower than TS01, which mean that there is not any correlation between the predicted and experimental results (Table 5-1).

Overall, the findings of this research support what is stated in the literature, that both computational and experimental investigations in drug design projects could be used in parallel to enhance the success rate of the projects^{82,267,268}. However, the predicted binding free energy between the receptor and ligand was not such accurate to match the experimental data and still may not be a reality. In this study, it is clear that the *in silico* results can be relatively correlated with the experimental values because all of the computational values are overestimated. In addition, the computational and experimental values of the several analogues like TS07, TS08 and TS10 are not correlated (Figure 5.1).

Nevertheless, it is apparent that the prediction of binding affinity in most projects is within the acceptable range, and that the continuous development of the computational programme and capacity are also required to correct the penalties and improve the accuracy through improving the force field parameters and undertaking higher sampling of the conformational space. As demonstrated in this project, accuracy of the predicted $\Delta G_{\text{binding}}$ values was significantly improved compared with the experimental results, after optimising the computational $\Delta G_{\text{binding}}$ values through 50 replications of the MD simulations to obtain average values (Figure 5.1). However, multiple MD simulations are expensive at present and require the use of specific supercomputer centres: this may improve in the future, however, if more powerful computers become available.

Further studies

The promising results of the work presented in this thesis have the potential for further exploration and investigation. The essential goal of the project was obtaining a peptide analogue of the TIN2 peptide that could inhibit PPIs between TRF1–TIN2. The compound TS11 was identified, which is a peptide analogue of TIN2; it has significantly higher binding affinity than TIN2 and could be considered as an early lead compound. However, the peptide analogue requires chemical optimisation of the potency, pharmacokinetics and physicochemical properties.

Optimisation could be achieved through various ways, such as mutating amino acid residues to improve binding affinity, trying to truncate unnecessary group or residues in order to decrease the size of the molecule, adding unnatural amino acids and cyclising the peptide analogues so as to increase the stability and cell permeability.

In our investigations, a library of peptide analogues was designed through mutation of the TIN2 sequence residues. However, only the peptide analogues TS02 and TS11 were capable of improving $\Delta G_{\text{binding}}$. Hence, future experiments could mutate F258Y and A261R to combine the favourable interactions of both residues and significantly increase $\Delta G_{\text{binding}}$. Furthermore, mutating P262 of the TIN2 peptide to Phe (P262F) may improve free binding energy as according to the crystal structure and MD simulations, P262 has not binding interaction with the TRF1 residues and its side-chain fairly close to the

F142 of TRF1 (Figure 5.2). However, P262F was designed (P03) and $\Delta G_{\text{binding}}$ was inhibited -3.5 kcal/mol, but it was applied for the single MD simulations and experimental assays were not performed.

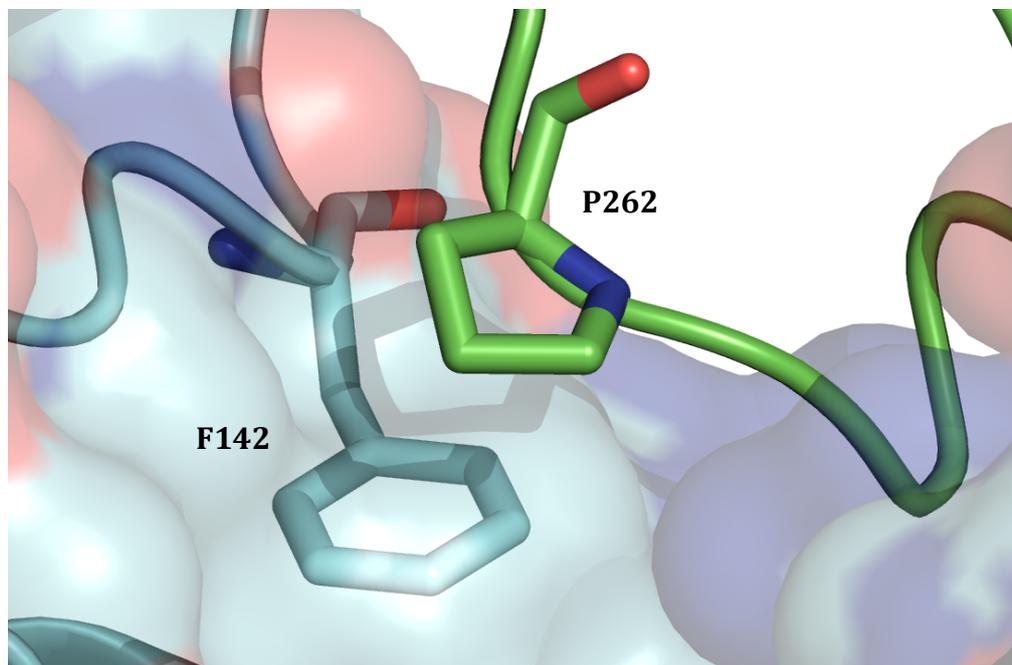


Figure 5.2 Conformations of the P262 and F141 in the TRF1–TIN2 crystal structure.

Amino acid residues S256, H257 and V268 can be deleted, as S256 has an unfavourable effect on the binding affinity with TRF1. In addition, per-residue decomposition results show that H257 and V268 contributions to the binding free energy are minimal (-1.32 and -0.82 kcal/mol, respectively) (Figure 2.10). The $\Delta G_{\text{binding}}$ value of truncated S256 (P06) was determined, which improved the binding free energy by approximately -11 kcal/mol (Table 2-2). In addition, deletion of both S256 and V268 in compound P26 improved $\Delta G_{\text{binding}}$ to -6.5 kcal/mol. However, these are single MD simulation results and their accuracy is unspecified. Truncating the residues leads to significant decrease in the sizes

of the peptide analogues (approximately 300 Da), crucial to improve the pharmacokinetic properties and reduce manufacturing cost.

The major limitation of peptide drugs, particularly the natural peptides, is their low residence time in the blood and rapid metabolism by the protease enzymes. This could be overcome by using various strategies that do not affect the biological activity of the molecule. For example, changing the entire peptide residues from L to D configuration, which is known as retro-inverso strategy²⁶⁹. This technique involves inversion of all of peptide bonds to form a retro-inverso peptide. It may be one of the approaches to improve stability of the peptide analogues, as the resulting peptide cannot be recognised by proteases and thus maintains biological activity.

In addition, cyclising of the TIN2 peptide analogues could be a promising strategy as it constrains flexibility of the peptide analogues, reduces the entropic penalties and improves the peptide's stability by decreasing susceptibility to proteolytic enzymes¹³³. As shown in Figure 5.3, the S256 and L263 side-chains are virtually free from interactions and the distance between them is 4.7 Å; closing this ring could constrain the main interacting region of TIN2 (F-X-L-X-P), which could be a promising approach for optimising the molecule. This can be achieved through mutating L263K and S256D to form a covalent bond between the side-chains of K263 and D256.

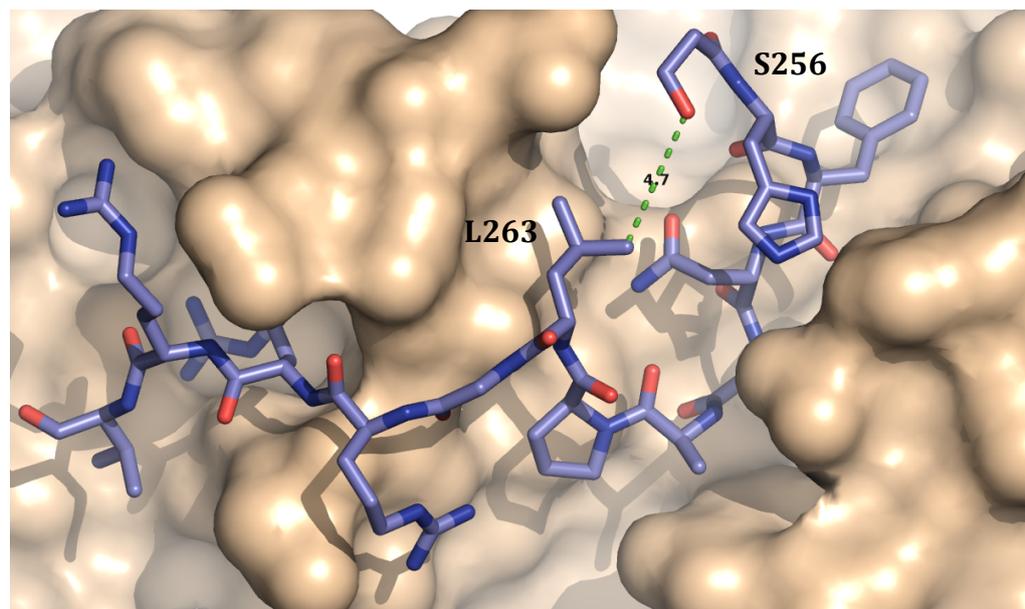


Figure 5.3 The TRF1–TIN2 crystal structure showing the conformations and the distance between the side-chains of S256 and L263.

The second goal of the project was optimising predicted $\Delta G_{\text{binding}}$ values calculated from the MM-GBSA method. After optimising the values and improving the correlation significantly, future work would be repeating the same approach on different systems to confirm the validity and reliability of the approach.

Recommendations for further research work

Optimising binding affinity in early drug development is a fundamental and challenging step. Improving potency is increasing $\Delta G_{\text{binding}}$ of the ligands through favourable increase in enthalpy and entropy. It is normal when introducing new functional groups or amino acid residues during lead optimisation to get unwanted enthalpic or entropic penalties, but recognising the causes of those penalties may accelerated the process of optimisation¹⁸⁷.

More importantly, before mutating any residues of the peptides or introducing new functional groups, it is essential to consider that the interacting atoms of the peptides are located at the correct angle and distance from the corresponding atoms of the amino acid residues in TRF1 protein in order to be sure that any enthalpic gain could overcome enthalpic penalty owing to desolvation of polar groups. However, it is apparent that improving enthalpy through mutating residues to construct a new H-bond or salt-bridges may not be sufficient to increase affinity because improving favourable enthalpy is always at the cost of unfavourable entropy, and large ligands such as peptides are usually entropy dominated²⁷⁰. Therefore, it will be important to consider conformational changes when wanting to establish a new or stronger binding interaction between the peptide analogues and the TRF1 protein and minimising structural effects by performing the peptides mutation in a well-structured region of the peptides and TRF1 protein.

In this study, almost of the mutations were made to increase hydrophilicity, but always adding hydrophilic residues may lead to unfavourable entropy¹⁸⁶. Therefore, if mutations are to increase the peptides hydrophobicity in the future, it may improve binding affinity significantly because of providing favourable entropy¹⁸⁷.

6. Experimental

6.1 Molecular Modeling

6.1.1 Systems Preparation

The starting coordinates of the TRF1–TIN2 complex (solved at 2.0 Å resolution) was downloaded from the Protein Data Bank (PDB) website (<http://www.rcsb.org>, access code 3BQO¹⁹²). The file was edited to retain a monomer of the complex structure. Then, mutations were carried out for the selected amino acids of the TIN2 peptide to produce the peptide analogues. Mutations were performed using the UCSF Chimera 1.6 package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco²⁷¹. Subsequently, the monomer complexes (Protein–Peptide) were prepared for MD simulations with AMBER suite version 12²⁷² and the ff99SB force field²⁷³. Counter ions (Na⁺) were added to neutralise the complex systems. The next step was immersing the Protein–Peptide complexes in an octahedral box of TIP3P water molecules that extended 10 Å outside of the complex in all dimensions²⁷⁴.

6.1.2 Systems Minimisation

After preparing and parameterising the complex systems, minimisations were achieved through two steps, and for each step, two different minimisation algorithms were applied. The first and the faster one was steepest descent

minimisation, which was used to quickly reduce unfavourable clashes. The next minimisation algorithm was conjugate gradient, which was performed to minimise systems energy after the steepest descent as its more efficient. In terms of the minimisation steps, restrained and the unrestrained minimisation were performed to relax the complex structures. Firstly, all of the water and ions were relaxed, while the protein–peptide coordinates were restrained to their original values, and then, unrestrained minimisation was applied to all of the system. The restrained minimisation procedure was 2000 cycles (500 cycles of steepest descent and 1500 cycles of conjugate gradient minimisation). The whole system minimisation was 3000 cycles (1000 cycles of steepest descent and 2000 cycles of conjugate gradient minimisation).

6.1.3 Molecular Dynamic Simulations

After minimisations, equilibrating MD simulations were performed in two stages: solute-restrained and then unrestrained. All simulations were performed using the pmemd module in Amber 12²⁷². For all simulations, the SHAKE algorithm was used to constrain bond lengths involving hydrogen atoms²⁷⁵. The time step was 2 femtoseconds (fs), and Langevin dynamics were carried out to control the temperature with a collision frequency of 1.0 ps⁻¹²⁷⁶. For the restrained 2 ns MD simulations, a constant volume periodic boundary and a 10 Å cutoff were used. Furthermore, the systems were gradually heated from 0 to 300 K over 200 picoseconds (ps) as a pre-equilibration period. Random initial velocities were generated from a Boltzmann distribution.

For the subsequent unrestrained equilibration MD simulations, the temperature was kept at 300 K and a 10 Å cutoff was used. Furthermore, a constant pressure periodic boundary was applied (1 atmosphere) with isotropic position scaling to maintain constant pressure dynamics. The relaxation time of the pressure was 2 ps. Finally, snapshots were taken every 10,000 steps (every 20 ps) of 10 ns MD simulations, thus producing 500 frames for each MD simulation²⁷⁷. While preliminary MD simulations were run using local facilities, the UK national supercomputing services HECToR (<http://www.hector.ac.uk>) and ARCHER (<http://www.archer.ac.uk>) were used for the later replicate MD simulations.

6.1.4 Trajectory Analysis Techniques

6.1.4.1 Predicted Binding Free Energies

The MM-PBSA/GBSA method was used to estimate $\Delta G_{\text{binding}}$ values for all of the 13 Protein–Peptide complexes using the standard AMBER python scripts (MMPBSA.py)²⁷⁸. Calculations were performed using all of the frames in each system trajectory file (10 ns, 500 frames). Water molecules were stripped off with an infinite cut-off. Four compatible topology files were created for each complex system, which were: solvated complex, non-solvated complex, receptor and ligand file²⁷⁷. To predict binding free energy using MM-GBSA method, water and counterions were stripped, the generalized Born method (igb =5) was used²⁷⁹, and 0.1 M salt concentration was assumed. Data analysis and visualisation were performed using GraphPad Prism V6.0 (GraphPad Software Inc., San Diego, CA; www.graphpad.com).

6.1.4.2 Molecular Similarity

Molecular similarity was measured by using Root Mean Square Deviation (RMSD) to describe the difference or the distance between the reference (native state) and the equilibrated structures of superimposed molecules, and also to study the equilibration of the system replicates (Equation 7-1)²⁸⁰.

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N d_i^2}{N}} \quad \text{Equation 6-1}$$

where N is the number of atoms and d_i is the distance between coordinates of equivalent atoms in the two superimposed structures. RMSD was performed to analyse the trajectory files by applying the ptraj utility of AmberTools²⁷⁷. In addition, ptraj was used to quantify conformational changes versus time and to measure the distance between atoms of superimposed structures.

6.1.5 Visualisation Software

Visual Molecular Dynamics (VMD) software from the Theoretical and Computational Biophysics Group, NIH Resource for Macromolecular Modeling and Bioinformatics at the University of Illinois at Urbana-Champaign was used to visualise structures, trajectories, measure hydrogen bonds and salt bridges²⁸¹. Furthermore, the PyMOL Molecular Graphics System²⁸² was used to visualise interactions, distances between atoms and conformational changes of the Protein–Peptide complex systems.

6.2 Peptide Synthesis

6.2.1 Materials

Chemicals for the experiments were purchased from Novabiochem, Sigma Aldrich, and MP Biomedicals, and directly used without further purification. The solvents, which were either reagent or HPLC grade, were bought from Fisher Scientific and Sigma Aldrich. All of the laboratory glassware was oven-dried overnight prior to use.

6.2.2 Instrumentation

Mass spectra were recorded by positive electrospray ionisation using a Waters 2759 spectrometer. Lyophilisation was carried out on an Edwards Modulyo freeze drier. Analytical RP-HPLC was performed using a Waters 510 twin pump using an OnyxTM monolithic C₁₈ analytical column (100 x 4.6 mm) at a flow rate of 3.0 ml min⁻¹. Eluent detection was monitored by UV absorbance using a Waters 486 Tunable Absorbance Detector at 214 nm. Solvent **A** was 100% water + 0.06% TFA and solvent **B** was 90% aqueous acetonitrile + 10% water + 0.06% TFA. The linear elution gradient was 10 to 60 %**B** over 18 min, at 3.0 ml min⁻¹. In addition, preparative RP-HPLC was performed on an OnyxTM monolithic-C₁₈ Semi-PREP column (100 x 10 mm) at a flow rate of 4 ml min⁻¹ with an elution gradient of 10 to 60 %**B** over 19 min. Peptide synthesis was carried out using a NOVASYN[®] GEM manual peptide synthesizer. UV absorbance was measured at 355 nm using a Gilson 115 UV detector.

^1H -NMR was performed at room temperature using a Bruker 400 Ultrashield operating at 400 MHz. Chemical shifts, denoted by δ were recorded in parts per million (ppm) relative to tetramethylsilane (TMS) and coupling constants (J) were measured in hertz (Hz). Deuterated solvent used was DMSO- d_6 ($\delta\text{H} = 2.50$ ppm). Finally, COSY was applied for ^1H assignments.

6.2.3 Procedure of the Peptides Assembly

The procedure for all peptides assembly was started by placing 0.1 mmol (0.156 g) of the Rink amide Novagel in the reaction column, which was then swollen with 1.5 ml DMF for 15 hours. The resin was then washed using DMF (3 ml min^{-1} , 7 min). The next amino acid was prepared for the acylation reaction as follows: 4 equivalent (eq.) of the N-protected amino acid was mixed with PyOxim (4 eq.) as an activating reagent; the mixture was then dissolved in the minimum amount of DMF (0.6 ml) and DIPEA (8 eq.) added. Subsequently, the mixture was added to the washed resin in the reaction column. Washing, Fmoc deprotection and a second washing were performed after each acylation. Fmoc deprotection was executed using a continuous flow of 20% v/v piperidine in DMF (3 ml min^{-1} , 7 min), while the washing processes were executed by flowing only DMF (3 ml min^{-1} , 7 min). To perform all the washing and deprotection processes, a NOVASYN[®] GEM peptide synthesizer was used and the reactions were monitored thorough UV absorption at 355 nm. Each of the consecutive acylation reactions, which lasted 3 hours, was performed in the reaction column at ambient temperature.

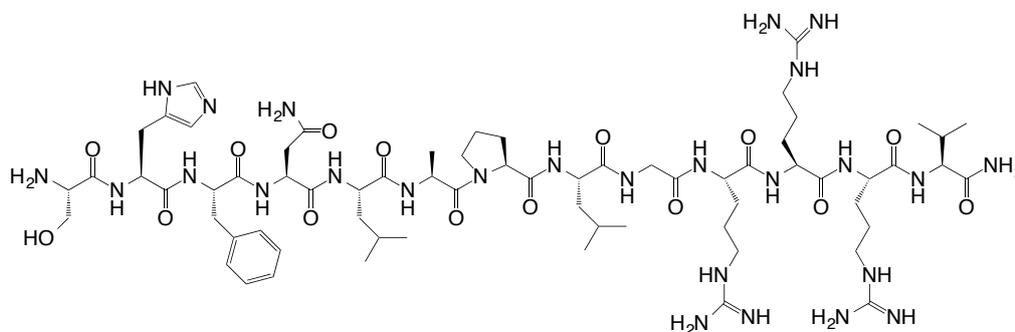
After the final acylation, the peptide–resin was washed using DMF, then filtered and washed again using DCM (5 ml) and hexane (5 ml) and dried overnight *in vacuo*. The next step was the cleavage of the resin and the semi-permanent protecting groups. This was carried out by suspending the peptide–resin in a mixture of 95% TFA, 2.5% H₂O, and 2.5% TIPS (10 ml) for 3 hours, and then filtering and evaporation *in vacuo* to dryness and triturating with 4 ml diethyl ether. Finally, the off-white solids were analysed by RP-HPLC and MS.

Regarding TIN2 peptide synthesis, after completion of all the amino acid coupling reactions, half of the peptide-resin was separated and β-Ala was coupled to the Ser256 residue of the one half peptide sequence; then, the *N*-termini of the peptide sequence labelled by the fluorescein molecule (5-FAM). In the fluorescein tagging TIN2 reaction step, HOBt (1 eq.) and DIC (1 eq.) were used as activating reagents.

6.2.4 Produced Peptides

TS01

Ser256-His-Phe-Asn-Leu-Ala-Pro-Leu-Gly-Arg-Arg-Arg- Val268-NH₂

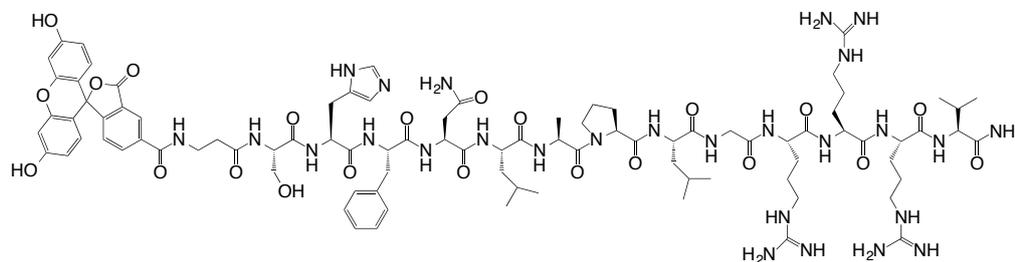


MS: m/z (+ESI) calculated for $C_{67}H_{112}N_{26}O_{15}$ is 1520.88, found 761.4521 ($M+2H^+$) and 507.9683 ($M+3H^+$). RP-HPLC 10-60 % B, in 18 min, t_R 5.9 min.

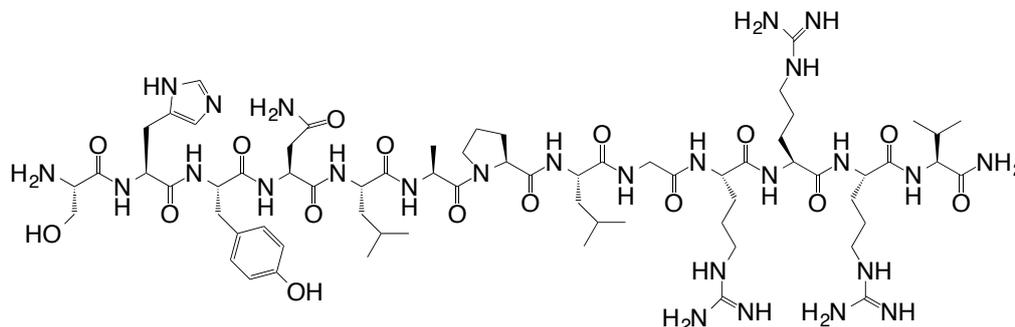
TS01-L

5-FAM- β -Ala-Ser256-His-Phe-Asn-Leu-Ala-Pro-Leu-Gly-Arg-Arg-Arg-

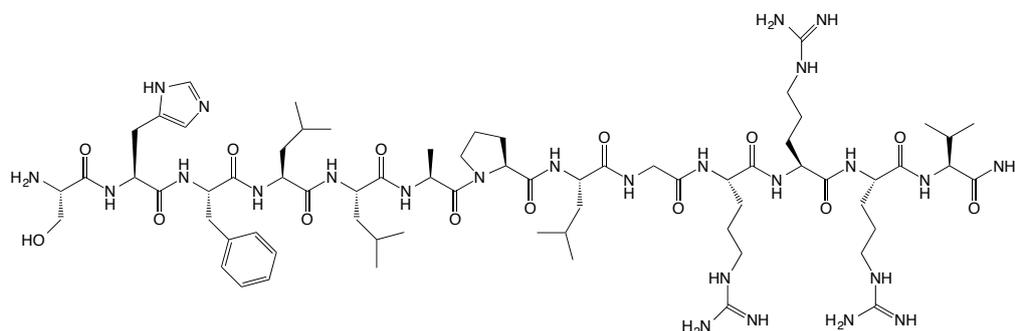
Val268-NH₂



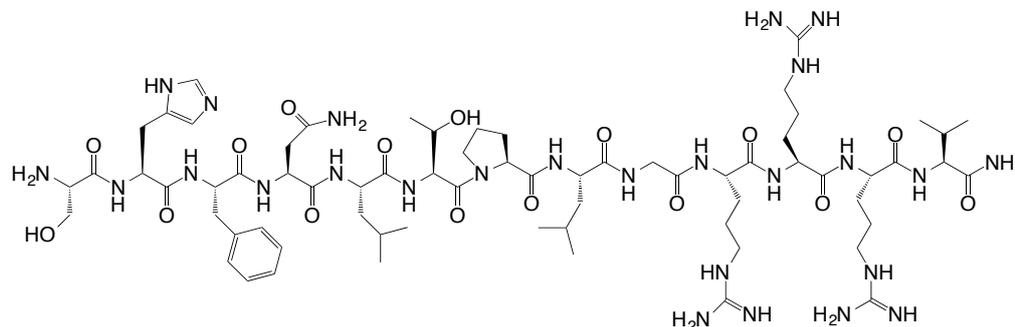
MS: m/z (+ESI) calculated for $C_{91}H_{127}N_{27}O_{22}$ is 1950.97; found 976.4952 ($M+2H^+$), 651.3334 ($M+3H^+$), and 488.7458 ($M+4H^+$). RP-HPLC 10-60 % B, in 18 min, t_R 7.4 min.

TS02Ser256-His-Tyr258-Asn-Leu-Ala-Pro-Leu-Gly-Arg-Arg-Arg- Val268-NH₂

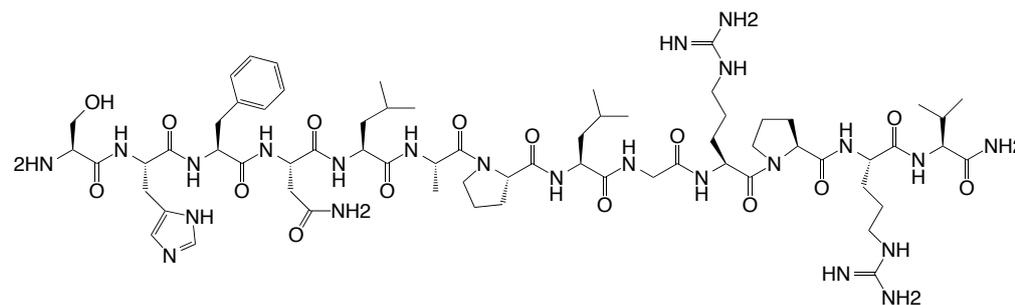
MS: m/z (+ESI) calculated for C₆₇H₁₁₂N₂₆O₁₆ is 1536.87, found 769.4514 (M+2H⁺) and 513.2959 (M+3H⁺). RP-HPLC 10-60 % B, in 18 min, *t_R* 5.4 min.

TS03Ser256-His-Phe-Leu259-Leu-Ala-Pro-Leu-Gly-Arg-Arg-Arg- Val268-NH₂

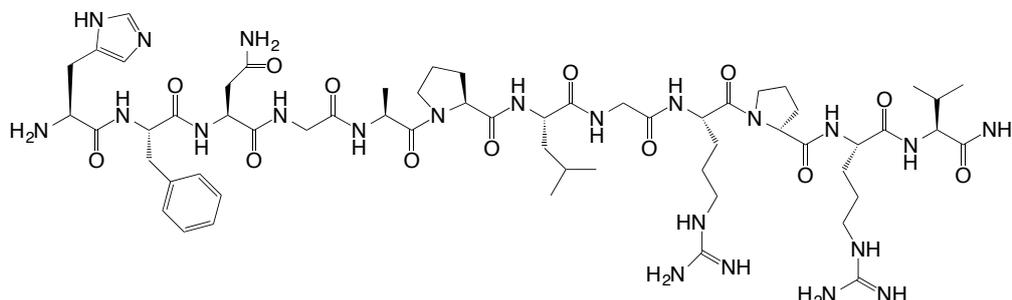
MS: m/z (+ESI) calculated for C₆₉H₁₁₇N₂₅O₁₄ is 1519.92, found 760.9876 (M+2H⁺) and 507.6575 (M+3H⁺). RP-HPLC 10-60 % B, in 18 min, *t_R* 6.4 min.

TS04Ser256-His-Phe-Asn-Leu-**Thr261**-Pro-Leu-Gly-Arg-Arg-Arg- Val268-NH₂

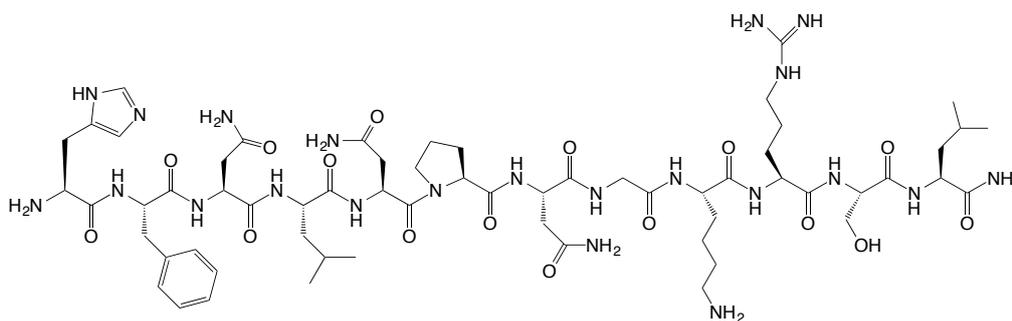
MS: m/z (+ESI) calculated for C₆₈H₁₁₄N₂₆O₁₆ is 1550.89, found 776.4598 (M+2H⁺) and 517.9702 (M+3H⁺). RP-HPLC 10-60 % B, in 18 min, *t_R* 5.9 min.

TS05Ser256-His-Phe-Asn-Leu-Ala-Pro-Leu-Gly-Arg-**Pro266**-Arg- Val268-NH₂

MS: m/z (+ESI) calculated for C₆₆H₁₀₇N₂₃O₁₅ is 1461.83, found 731.9208 (M+2H⁺) and 488.2768 (M+3H⁺). RP-HPLC 10-60 % B, in 18 min, *t_R* 6.3 min.

TS06Ser256-His-Phe-Asn-Gly260-Ala-Pro-Leu-Gly-Arg-**Pro266**-Arg- Val268-NH2

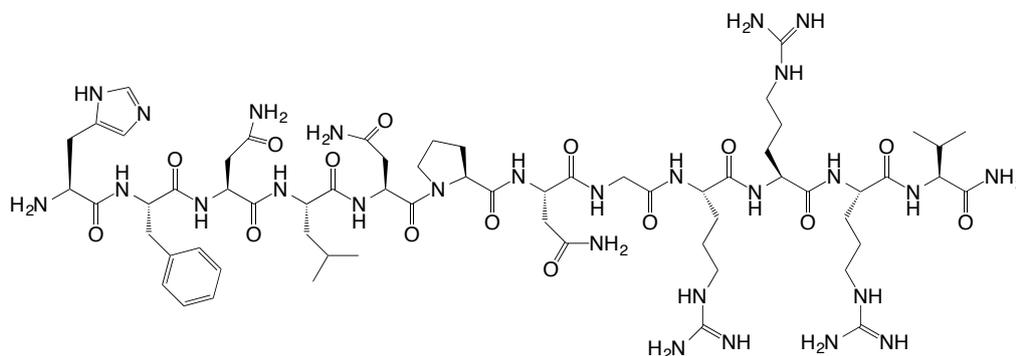
MS: m/z (+ESI) calculated for $C_{59}H_{94}N_{22}O_{13}$ is 1318.74, found 660.3820 ($M+2H^+$) and 440.5880 ($M+3H^+$). RP-HPLC 10-60 % B, in 18 min, t_R 5.0 min.

TS07Ser256-His-Phe-Asn-Leu-**Asn261**-Pro-**Asn263**-Gly-**Lys265**-Arg-**Ser267**-**Leu268**-NH2

MS: m/z (+ESI) calculated for $C_{61}H_{98}N_{22}O_{16}$ is 1394.75, found 698.2541 ($M+2H^+$) and 466.0143 ($M+3H^+$). RP-HPLC 10-60 % B, in 18 min, t_R 5.1 min.

TS08

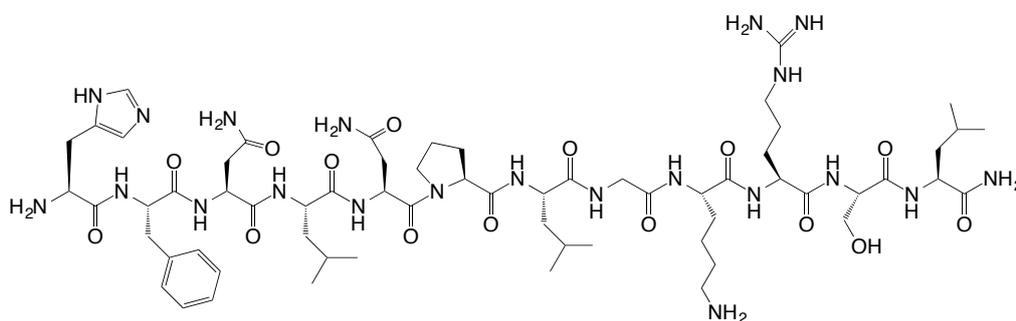
Ser256-His-Phe-Asn-Leu-**Asn261**-Pro-**Asn263**-Gly-Arg-Arg-Arg-Val268-NH₂



MS: m/z (+ESI) calculated for C₆₃H₁₀₃N₂₇O₁₅ is 1477.81, found 493.6815 (M+2H⁺) and 370.6317 (M+3H⁺). RP-HPLC 10-60 % B, in 18 min, *t_R* 5.0 min.

TS09

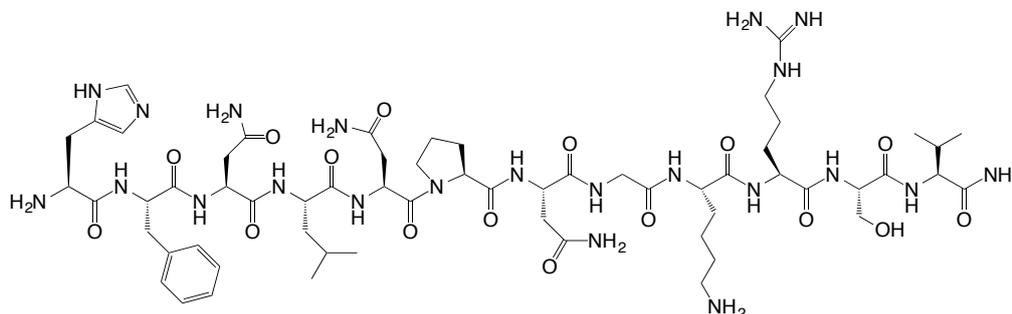
Ser256-His-Phe-Asn-Leu-**Asn261**-Pro-Leu-Gly-**Lys265**-Arg-**Ser267**-**Leu268**-NH₂



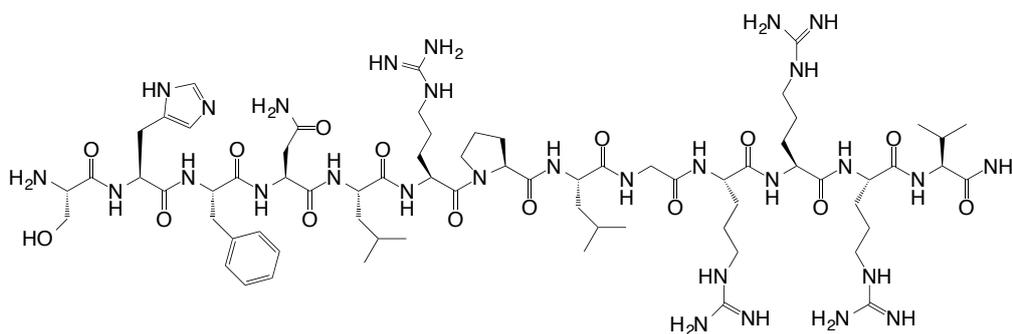
MS: m/z (+ESI) calculated for C₆₃H₁₀₃N₂₁O₁₅ is 1393.79, found 697.7742 (M+2H⁺) and 465.6921 (M+3H⁺). RP-HPLC 10-60 % B, in 18 min, *t_R* 6.3 min.

TS10

Ser256-His-Phe-Asn-Leu-Asn261-Pro-Asn263-Gly-Lys265-Arg-Ser267-

Val268-NH₂

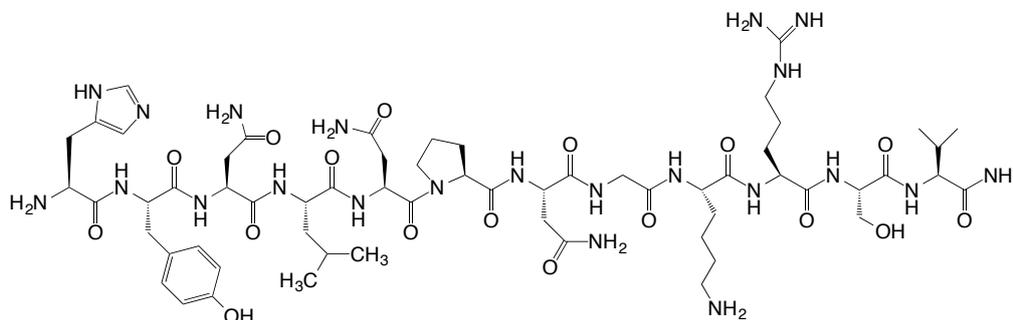
MS: m/z (+ESI) calculated for C₆₀H₉₆N₂₂O₁₆ is 1380.74, found 691.2517 (M+2H⁺) and 461.3439 (M+3H⁺). RP-HPLC 10-60 % B, in 18 min, *t_R* 4.9 min.

TS11Ser256-His-Phe-Asn-Leu-Arg261-Pro-Leu-Gly-Arg-Arg-Arg- Val268-NH₂

MS: m/z (+ESI) calculated for C₇₀H₁₁₉N₂₉O₁₅ is 1605.94; found 803.7291 (M+2H⁺), 536.3444 (M+3H⁺) and 402.6352 (M+4H⁺). RP-HPLC 10-60 % B, in 18 min, *t_R* 5.6 min.

TS12

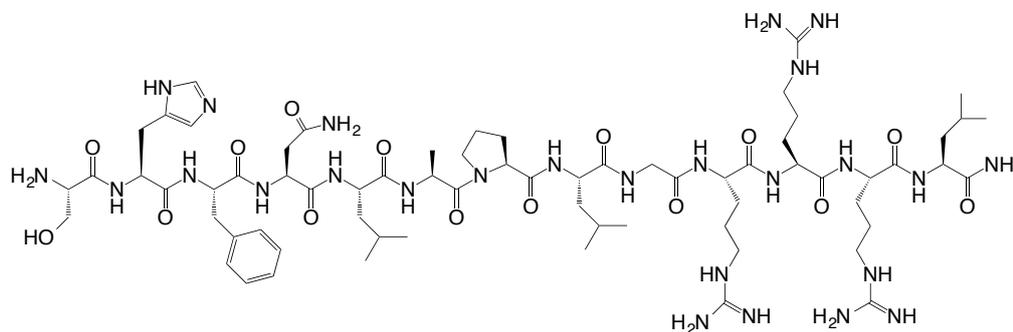
Ser256-His-Tyr258-Asn-Leu-Asn261-Pro-Asn263-Gly-Lys265-Arg-Ser267-Val268-NH₂



MS: m/z (+ESI) calculated for $C_{60}H_{96}N_{22}O_{17}$ is 1396.73, found 699.2429 ($M+2H^+$) and 466.6737 ($M+3H^+$). RP-HPLC 10-60 % B, in 18 min, t_R 4.5 min.

TS13

Ser256-His-Phe-Asn-Leu-Ala-Pro-Leu-Gly-Arg-Arg-Arg-Leu268-NH₂



MS: m/z (+ESI) calculated for $C_{68}H_{114}N_{26}O_{15}$ is 1534.90, found 768.2478 ($M+2H^+$) and 512.6854 ($M+3H^+$). RP-HPLC 10-60 % B, in 18 min, t_R 6.0 min.

6.3 Fluorescence Polarisation Assay

6.3.1 Materials and Instrumentation

The essential materials used in the FP assays were: TRF1 protein stock solution (41.7 μM in 50% TED₃₅₀ protein dilution buffer and 50% v/v glycerol pH 8.5); TS01-L stock solution (3 mM in water); and stock solutions of each of the peptide analogues (3 mM in water). An EnVision[®] Multilabel Plate reader 2104 (Perkin Elmer) was used for reading the plates, these being black opaque, flat-bottomed, 96-well microplates (Thermo scientific Sterilin, Catalogue No. 611F96BK). The reader used EnVision[®] version 2.12 data-capturing software (excitation wavelength was set on 480 nm and the emission wavelength was set on 535 nm); and GraphPad Prism Software was applied for analysing the results.

6.3.2 Detection Sensitivity of the Instrument for TS01-L

Increasingly diluted concentrations of TS01-L were prepared from the 3 mM stock solution, starting from 10 μM with subsequent dilutions down to 10 nM. Dilutions were performed in 0.5 ml eppendorf tubes and 100 μl was aliquoted to each well of the 96-well plate. In addition, 100 μl from both the 5 nM 5-FAM (the solvent was water) and the TED₃₅₀ buffer were aliquoted to each of the wells to act as gain detector and blank (to correct background fluorescence), respectively. Subsequently, the plate was loaded onto the EnVision[®] plate reader; the software automatically measured the plate

dimensions, but the gain and blank were manually adjusted. The plate was then read and FP measured for the samples. All of the saved data were exported to Graphpad software, which plotted the average of the measured FP of the wells for each point as a function of the peptide concentration.

6.3.3 Determining the Concentration of TRF1 for the Assays

The concentration of TRF1 protein was optimised for the FP assays through measurement of its direct binding with TS01-L. Concentrating of the TRF1 protein was the first step of this process because normally the highest concentration of the protein for the assay needs to be approximately 30–100 μM so as to determine a plateau region and produce a complete binding curve²⁴¹. For this step, 700 μl (7 aliquots each containing 100 μl) of TRF1 protein were put in Falcon™ 15 ml conical centrifuge tubes (Fischer Scientific). The glycerol concentration was diluted from 50% to 5.5% by adding 5.664 ml of TED₃₅₀ buffer pH 8.5. After 5 hours centrifugation of the protein, the concentration had increased to 86 μM . NanoDrop was used to measure TRF1 concentration. For the TRF1 centrifugation, a Vivaspin 20 ml spin concentrator (5,000 Da MW cut-off, Sartorius) in an Allegra® 25R centrifuge (Beckman Coulter) was used; it was operated at 2 °C and 4100 revolutions per minute (rpm).

From the concentrated TRF1 protein solution, dilutions to 33.3 μM and 11.1 μM (200 μl from each) were prepared. Eleven eppendorf tubes (0.5 ml) were then prepared as follows. The first tube received 180 μl of the 33.3 μM TRF1 solution. The second tube received 180 μl of the 11.1 μM TRF1 solution. Eight

of the remaining eppendorf tubes each received 160 μl of TED₃₅₀ buffer pH 8.5, followed by each of them receiving 20 μl of different TRF1 concentrations (50% glycerol) as follows: 30 μM , 10 μM , 3 μM , 1 μM , 300 nM, 100 nM, 30 nM and 10 nM. The last eppendorf tube received 180 μl of TED₃₅₀ buffer containing 5.5% glycerol pH 8.5 (negative control). Subsequently, 20 μl of 1 μM TS01-L solution were aliquoted to each of the eleven-eppendorf tubes, which led change glycerol concentration in each of the tubes to a 5%. All of the tubes were mixed through pipetting. The next step was to aliquot 200 μl from each sample into two wells (100 μl /well) of the 96-well plate. The samples were incubated for 30 min at room temperature and protected from light. The plates were then loaded onto the plate reader and measured as described in Section 6.3.2. The FP values were corrected through subtracting the measured control values.

6.3.4 Competitive Binding Assay

These assays were achieved in two steps. First, a competitive assay between the labelled (TS01-L) and unlabelled TIN2 (TS01) with TRF1 was performed, and second, competitive binding assays between the mixture of TS01-L and TRF1 with all of the 13 peptides were performed. The stock solution of TRF1 protein was diluted from 41.7 μM to 10 μM through collecting 12 aliquots of the protein (1.2 ml) in Falcon™ 15 ml conical centrifuge tubes (Fischer Scientific) and adding 3.804 ml of the buffer (50% TED₃₅₀ and 50% glycerol) pH 8.5; the final volume of the protein solution was 5.004 ml. All of the dilutions were undertaken in an icebox, after which 5 ml of 1 μM TS01-L were

prepared as a stock solution using water as a solvent. The next step was to make up 5 ml of 5 nM 5-FAM, again using water as a solvent.

For each of the peptide analogues, starting with the 3 mM stock solution, increasingly diluted solutions (down to 100 nM) were prepared; afterwards, 40 μ l of each analogue concentration was aliquoted into separate 0.5 ml eppendorf tubes corresponding to the analogue concentration. Next, 40 μ l of each of the TRF1 and TS01-L solutions and 280 μ l of TED₃₅₀ buffer pH 8.5 were added to each 0.5 ml eppendorf tube. Three extra eppendorf tubes had been prepared: the first contained only 360 μ l TED₃₅₀ buffer and 40 μ l TS01-L (to act as a positive control); the second contained only the prepared 5 nM of 5-FAM (to act as the gain); and the third contained the TED₃₅₀ buffer (to act as a blank to correct background fluorescence). The contents of each 0.5 ml eppendorf tube were mixed gently through pipetting. Finally, 300 μ l from each eppendorf tube were aliquoted to three wells (100 μ l/well) of the 96-well plate. The microplates were incubated for 30 min at room temperature and protected from light. After loading the plates to the EnVision[®] plate reader, the software automatically measured the plate dimensions, but the gain and blank were manually adjusted. The plate was then read and the FP measured for the samples. All of the saved data was exported to Graphpad software, which plotted the average of the measured FP of the three wells for each point as a function of the peptide concentration. The FP values were corrected through subtracting measured control values. Experiments were performed in triplicate.

7. References

1. Zhdanova, N.S., Minina, J.M. & Rubtsov, N.B. Mammalian telomere biology. *Molecular Biology* **46**, 481-495 (2012).
2. Cookson, J.C. & Laughton, C.A. The levels of telomere-binding proteins in human tumours and therapeutic implications. *European Journal of Cancer* **45**, 536-550 (2009).
3. Frias, C., Pampalona, J., Genesca, A. & Tusell, L. Telomere dysfunction and genome instability. *Frontiers in Bioscience* **17**, 2181-2196 (2012).
4. Riethman, H. Human telomere structure and biology. *Annual Review of Genomics and Human Genetics* **9**, 1-19 (2008).
5. Hiyama, K., Hiyama, E. & Shay, J.W. Telomeres and Telomerase in Humans. in *Telomeres and Telomerase in Cancer* (ed. Hiyama, K.) 3-21 (Humana Press 2009).
6. Gomez, D.E. et al. Telomere structure and telomerase in health and disease (Review). *International Journal of Oncology* **41**, 1561-1569 (2012).
7. Hubbard, B. Let's talk about telomeres. in *Helix magazine* (eds Daugherty, R. & Kennedy, M.) (Science in Society, 2014).
8. Deng, Y., Chan, S.S. & Chang, S. Telomere dysfunction and tumour suppression: the senescence connection. *Nature Reviews Cancer* **8**, 450-458 (2008).
9. de Lange, T. T-loops and the origin of telomeres. *Nature Reviews Molecular Cell Biology* **5**, 323-329 (2004).
10. Shay, J.W. & Wright, W.E. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* **26**, 867-874 (2005).
11. Zvereva, M.I., Shcherbakova, D.M. & Dontsova, O.A. Telomerase: Structure, functions, and activity regulation. *Biochemistry-Moscow* **75**, 1563-1583 (2010).
12. Zimmermann, S. & Martens, U.M. Telomeres and telomerase as targets for cancer therapy. *Cellular and Molecular Life Sciences* **64**, 906-921 (2007).
13. Olovnikov, A.M. Telomeres, telomerase, and aging: Origin of the theory. *Experimental Gerontology* **31**, 443-448 (1996).
14. Shay, J.W. & Wright, W.E. Hayflick, his limit, and cellular ageing. *Nature Reviews Molecular Cell Biology* **1**, 72-76 (2000).
15. Rajpar, S., Guittat, L. & Mergny, J.L. Telomeres: a Nobel Prize at the beginning ... of the end. *Bulletin Du Cancer* **98**, 999-1009 (2011).
16. De Boeck, G., Forsyth, R.G., Praet, M. & Hogendoorn, P.C.W. Telomere-associated proteins: cross-talk between telomere maintenance and telomere-lengthening mechanisms. *Journal of Pathology* **217**, 327-344 (2009).
17. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & Development* **19**, 2100-2110 (2005).

18. Giraud-Panis, M.J., Pisano, S., Poulet, A., Le Du, M.H. & Gilson, E. Structural identity of telomeric complexes. *FEBS Letters* **584**, 3785-3799 (2010).
19. Palm, W. & de Lange, T. How Shelterin Protects Mammalian Telomeres. *Annual Review of Genetics* **42**, 301-334 (2008).
20. Wang, F. et al. The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506-510 (2007).
21. Chen, Y. et al. A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* **319**(2008).
22. Okamoto, K., Iwano, T., Tachibana, M. & Shinkai, Y. Distinct roles of TRF1 in the regulation of telomere structure and lengthening. *Journal of Biological Chemistry* **283**, 23981-23988 (2008).
23. Di Maro, S. et al. Shading the TRF2 recruiting function: a new horizon in drug development. *Journal of The American Chemical Society* **136**, 16708-16711 (2014).
24. Karlseder, J., Broccoli, D., Dai, Y.M., Hardy, S. & de Lange, T. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **283**, 1321-1325 (1999).
25. Liu, D., O'Connor, M.S., Qin, J. & Songyang, Z. Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *Journal of Biological Chemistry* **279**, 51338-51342 (2004).
26. Ye, J.Z. et al. TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *Journal of Biological Chemistry* **279**, 47264-47271 (2004).
27. Bhanot, M. & Smith, S. TIN2 Stability Is Regulated by the E3 Ligase Siah2. *Molecular and Cellular Biology* **32**(2012).
28. Xin, H., Liu, D. & Zhou, S. The telosome/shelterin complex and its functions. *Genome Biology* **9**, 232.1-232.9 (2008).
29. Chang, F., Syrjanen, S., Tervahauta, A. & Syrjanen, K. Tumourigenesis associated with the p53 tumour suppressor gene. *British Journal of Cancer* **68**, 653-661 (1993).
30. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
31. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
32. Hickman, J.A. Apoptosis and tumourigenesis. *Current Opinion in Genetic & Development* **12**, 67-72 (2002).
33. Campbell, P.J. Telomeres and Cancer: From Crisis to Stability to Crisis to Stability. *Cell* **148**, 633-635 (2012).
34. Wu, X. et al. Telomere Dysfunction: A Potential Cancer Predisposition Factor. *JNCI Journal of the National Cancer Institute* **95**, 1211-1218 (2003).
35. Shamma, M.A. Telomeres, lifestyle, cancer, and aging. *Current Opinion in Clinical Nutrition and Metabolic Care* **14**, 28-34 (2011).
36. Artandi, S.E. & DePinho, R.A. Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9-18 (2010).
37. Xu, L., Li, S. & Stohr, B.A. The role of telomere biology in cancer. *Annual Review of Pathology Mechanisms of Disease* **8**, 49-78 (2013).

38. Dunham, M.A., Neumann, A.A., Fasching, C.L. & Reddel, R.R. Telomere maintenance by recombination in human cells. *Nature Genetics* **26**, 447-450 (2000).
39. Shammass, M.A. et al. Telomere maintenance in laser capture microdissection-purified Barrett's adenocarcinoma cells and effect of telomerase inhibition in vivo. *Clinical Cancer Research* **14**, 4971-4980 (2008).
40. Harley, C.B. Telomerase and cancer therapeutics. *Nature Reviews Cancer* **8**, 167-179 (2008).
41. De Cian, A. et al. Targeting telomeres and telomerase. *Biochimie* **90**, 131-155 (2008).
42. Bilsland, A.E., Cairney, C.J. & Keith, W.N. Targeting the telomere and shelterin complex for cancer therapy: current views and future perspectives. *Journal of Cellular and Molecular Medicine* **15**, 179-186 (2011).
43. Bollmann, F.M. Targeting ALT: The role of alternative lengthening of telomeres in pathogenesis and prevention of cancer. *Cancer Treatment Reviews* **33**, 704-709 (2007).
44. Bisoffi, M., Heaphy, C.M. & Griffith, J.K. Telomeres: Prognostic markers for solid tumors. *International Journal of Cancer* **119**, 2255-2260 (2006).
45. Shammass, M.A. et al. Telomerase inhibitor GRN163L inhibits myeloma cell growth in vitro and in vivo. *Leukemia* **22**, 1410-1418 (2008).
46. Neidle, S. Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *Federation of European Biochemical Societies Journal* **277**, 1118-1125 (2010).
47. Rezler, E.M., Bearss, D.J. & Hurley, L.H. Telomere inhibition and telomere disruption as processes for drug targeting. *Annual Review of Pharmacology and Toxicology* **43**, 359-379 (2003).
48. Sun, D. et al. Inhibition of human telomerase by a G-quadruplex-interactive compound. *Journal of Medicinal Chemistry* **40**, 2113-2116 (1997).
49. Mergny, J.L. et al. The development of telomerase inhibitors: the G-quartet approach. *Anti-Cancer Drug Design* **14**, 327-339 (1999).
50. Gowan, S.M. et al. A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity. *Molecular Pharmacology* **61**, 1154-1162 (2002).
51. Pennarun, G. et al. Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene* **24**, 2917-2928 (2005).
52. Gowan, S.M., Heald, R., Stevens, M.F. & Kelland, L.R. Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with G-quadruplexes. *Molecular Pharmacology* **60**, 981-988 (2001).
53. Balasubramanian, S., Hurley, L.H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature Reviews Drug Discovery* **10**, 261-275 (2011).

54. Seimiya, H. The telomeric PARP, tankyrases, as targets for cancer therapy. *British Journal of Cancer* **94**, 341-345 (2006).
55. Boukamp, P. & Mirancea, N. Telomeres rather than telomerase a key target for anti-cancer therapy? *Experimental Dermatology* **16**, 71-79 (2007).
56. O'Connor, M.S., Safari, A., Xin, H.W., Liu, D. & Songyang, Z. A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. *Proceedings of the National Academy of Sciences* **103**, 11874-11879 (2006).
57. Pal, D., Sharma, U., Singh, S.K., Kakkar, N. & Prasad, R. Over-expression of telomere binding factors (TRF1 & TRF2) in renal cell carcinoma and their inhibition by using SiRNA induce apoptosis, reduce cell proliferation and migration invitro. *PLoS One* **10**, e0115651 (2015).
58. Folini, M., Gandellini, P. & Zaffaroni, N. Targeting the telosome: Therapeutic implications. *Biochimica Et Biophysica Acta-Molecular Basis of Disease* **1792**, 309-316 (2009).
59. Garcia-Beccaria, M. et al. Therapeutic inhibition of TRF1 impairs the growth of p53-deficient K-RasG12V-induced lung cancer by induction of telomeric DNA damage. *EMBO Molecular Medicine* **7**, 930-949 (2015).
60. Chen, L.-Y., Liu, D. & Songyang, Z. Telomere maintenance through spatial control of telomeric proteins. *Molecular and Cellular Biology* **27**, 5898-5909 (2007).
61. Bleicher, K.H., Bohm, H.-J., Muller, K. & Alanine, A.I. Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery* **2**, 369-378 (2003).
62. Hopkins, A.L., Groom, C.R. & Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **9**, 430-431 (2004).
63. Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C. & Greyson, D. The cost of drug development: a systematic review. *Health Policy* **100**, 4-17 (2011).
64. Samsdodd, F. Target-based drug discovery: is something wrong? *Drug Discovery Today* **10**, 139-147 (2005).
65. Hughes, J.P., Rees, S., Kalindjian, S.B. & Philpott, K.L. Principles of early drug discovery. *British Journal of Pharmacology* **162**, 1239-1249 (2011).
66. Comess, K. & Schurdak, M. Affinity-based screening techniques for enhancing lead discovery. *Current Opinion in Drug Discovery & Development* **7**, 411-416 (2004).
67. DiMasi, J.A. & Grabowski, H.G. The cost of biopharmaceutical R&D: is biotech different? *Managerial and Decision Economics* **28**, 469-479 (2007).
68. Schmid, E.F. & Smith, D.A. Keynote review: Is declining innovation in the pharmaceutical industry a myth? *Drug Discovery Today* **10**, 1031-1039 (2005).
69. Rawlins, M.D. Cutting the cost of drug development? *Nature Reviews Drug Discovery* **3**, 360-364 (2004).
70. Alanine, A., Nettekoven, M., Roberts, E. & Thomas, A.W. Lead Generation–Enhancing the Success of Drug Discovery by Investing in

- the Hit to Lead Process. *Combinatorial Chemistry & High Throughput Screening* **6**, 51-66 (2003).
71. Ghose, A.K., Herbertz, T., Salvino, J.M. & Mallamo, J.P. Knowledge-based chemoinformatic approaches to drug discovery. *Drug Discovery Today* **11**, 1107-1114 (2006).
 72. Boppana, K., Dubey, P.K., Jagarlapudi, S.A.R.P., Vadivelan, S. & Rambabu, G. Knowledge based identification of MAO-B selective inhibitors using pharmacophore and structure based virtual screening models. *European Journal of Medicinal Chemistry* **44**, 3584-3590 (2009).
 73. McInnes, C. Virtual screening strategies in drug discovery. *Current Opinion in Chemical Biology* **11**, 494-502 (2007).
 74. Knowles, J. & Gromo, G. A guide to drug discovery: Target selection in drug discovery. *Nature Reviews of Drug Discovery* **2**, 63-69 (2003).
 75. Swinney, D.C. & Anthony, J. How were new medicines discovered? *Nature Reviews Drug Discovery* **10**, 507-519 (2011).
 76. Kotz, J. Phenotypic screening, take two. *Science-Business eXchange* **5**, 1-3 (2012).
 77. Broach, J.R. & Thorner, J. High-throughput screening for drug discovery. *Nature* **384**, 14-16 (1996).
 78. Hertzberg, R. & Pope, A. High-throughput screening: new technology for the 21st century. *Current Opinion in Chemical Biology* **4**, 445-451 (2000).
 79. Agresti, J.J. et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences* **107**, 4004-4009 (2010).
 80. Macarron, R. et al. Impact of high-throughput screening in biomedical research. *Nature Reviews Drug Discovery* **10**, 188-195 (2011).
 81. Szymanski, P., Markowicz, M. & Mikiciuk-Olasik, E. Adaptation of high-throughput screening in drug discovery-toxicological screening tests. *International Journal of Molecular Sciences* **13**, 427-452 (2012).
 82. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E.W., Jr. Computational methods in drug discovery. *Pharmacological Reviews* **66**, 334-395 (2014).
 83. Fox, S. et al. High-throughput screening: update on practices and success. *Journal of Biomolecular Screening* **11**, 864-869 (2006).
 84. Murray, C.W., Verdonk, M.L. & Rees, D.C. Experiences in fragment-based drug discovery. *Trends in Pharmacological Sciences* **33**, 224-232 (2012).
 85. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discovery Today* **8**, 876-877 (2003).
 86. Jhoti, H., Williams, G., Rees, D.C. & Murray, C.W. The 'rule of three' for fragment-based drug discovery: where are we now? *Nature Reviews Drug Discovery* **12**, 644-645 (2013).
 87. Erlanson, D.A., McDowell, R.S. & O'Brien, T. Fragment-based drug discovery. *Journal of Medicinal Chemistry* **47**, 3463-3482 (2004).
 88. Kapetanovic, I.M. Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach. *Chemico-Biological Interactions* **171**, 165-176 (2008).

89. Horvath, D. A virtual screening approach applied to the search for trypanothione reductase inhibitors. *Journal of Medicinal Chemistry* **40**, 2412-2423 (1997).
90. Gane, P. & Dean, P. Recent advances in structure-based rational drug design. *Current Opinion in Structural Biology* **10**, 401-404 (2000).
91. Anderson, A.C. The Process of Structure-Based Drug Design. *Chemistry & Biology* **10**, 787-797 (2003).
92. Bambini, S. & Rappuoli, R. The use of genomics in microbial vaccine development. *Drug Discovery Today* **14**, 252-260 (2009).
93. Lundstrom, K. Genomics and drug discovery. *Future Medicinal Chemistry* **3**, 1855-1858 (2011).
94. Berman, H.M. et al. The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).
95. International Union of Crystallography. Acta crystallographica. Section B, Structural crystallography and crystal chemistry. 15 v. (Published for the International Union of Crystallography by Munksgaard,, Copenhagen, 1967).
96. Durrant, J.D. & McCammon, J.A. Computer-aided drug-discovery techniques that account for receptor flexibility. *Current Opinion in Pharmacology* **10**, 770-774 (2010).
97. Durrant, J.D. & McCammon, J.A. Molecular dynamics simulations and drug discovery. *Bmc Biology* **9**(2011).
98. Babine, R.E. & Bender, S.L. Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. *Chemical Reviews* **97**, 1359-1472 (1997).
99. Acharya, C., Coop, A., Polli, J. & MacKerell Jr, A. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Current Computer-Aided Drug Design* **7**, 10-22 (2011).
100. Cerny, J. & Hobza, P. Non-covalent interactions in biomacromolecules. *Physical Chemistry Chemical Physics* **9**, 5291-5303 (2007).
101. Breckenridge, R.J. Molecular recognition: Models for drug design. *Cellular and Molecular Life Sciences* **47**, 1148-1161 (1991).
102. Gellman, S.H. Introduction: Molecular Recognition. *Chemical Reviews* **97**, 1231-1232 (1997).
103. Blokzijl, W. & Englbarts, J. Hydrophobic Effects. Opinions and Facts. *Angewandte Chemie International Edition* **32**, 1545-1579 (1993).
104. Grisham, C. & Garrett, R. the Facts of Life: Chemistry is the Logic of Biological Phenomena. in *Molecular Components of Cells* 5th edn (ed. McGahey, P.) (Finch, M., Cengage Learning, Canada, 2012).
105. Chang, R. *Physical Chemistry for the Biosciences*, (USA, 2005).
106. Bohm, H.J. & Klebe, G. What Can We Learn from Molecular Recognition in Protein-Ligand Complexes for the Design of New Drugs? *Angewandte Chemie International Edition* **35**, 2588-2614 (1996).
107. Searle, M.S. & Williams, D.H. The cost of conformational order: entropy changes in molecular associations. *Journal of the American Chemical Society* **114**, 10690-10697 (1992).
108. Burley, S. & Petsko, G. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**, 23-28 (1985).

109. Van Oss, C., Absolom, D. & Neumann, A. Applications of net repulsive van der Waals forces between different particles, macromolecules, or biological cells in liquids. *Colloids and Surfaces* **1**, 45-56 (1980).
110. Petrucci, R. *General chemistry: Principles and modern applications* 764 (Collier Macmillian, London, 1982).
111. Roth, C.M., Neal, B.L. & Lenhoff, A.M. Van der Waals interactions involving proteins. *Biophysical Journal* **70**, 977-987 (1996).
112. Arunan, E. et al. Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure and Applied Chemistry* **83**, 1637-1641 (2011).
113. Martin, T.W. & Derewenda, Z.S. The name is bond - H bond. *Nature Structural Biology* **6**, 403-406 (1999).
114. Meyer, M., Wilson, P. & Schomburg, D. Hydrogen Bonding and Molecular Surface Shape Complementarity as a Basis for Protein Docking. *Journal of Molecular Biology* **264**, 199-210 (1996).
115. Jeffrey A. George. *An Introduction to Hydrogen Bonding*, (Oxford University Press, 1997).
116. Anslyn, E.V. & Dougherty, D.A. *Modern physical organic chemistry*, 1095 (University Science, Sausalito, CA, 2006).
117. Burley, S.K. & Petsko, G.A. Weakly polar interactions in proteins. *Advances in Protein Chemistry* **39**, 125-189 (1988).
118. Ma, J.C. & Dougherty, D.A. The cation- π interaction. *Chemical Reviews* **97**, 1303-1324 (1997).
119. Gallivan, J.P. & Dougherty, D.A. Cation- π interactions in structural biology. *Proceedings of the National Academy of Sciences* **96**, 9459-9464 (1999).
120. Dougherty, D.A. Cation- π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **271**, 163-168 (1996).
121. Donald, J.E., Kulp, D.W. & DeGrado, W.F. Salt bridges: geometrically specific, designable interactions. *Proteins* **79**, 898-915 (2011).
122. Hendsch, Z.S. & Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Science* **3**, 211-226 (1994).
123. Sindelar, C.V., Hendsch, Z.S. & Tidor, B. Effects of salt bridges on protein structure and design. *Protein Science* **7**, 1898-1914 (1998).
124. Kumar, S. & Nussinov, R. Close-range electrostatic interactions in proteins. *European Journal of Chemical Biology* **3**, 604-617 (2002).
125. Thierauch, K.-H. Small Molecule Drugs. in *Encyclopedia of Cancer* (ed. Schwab, M.) 3448-3451 (Springer Berlin Heidelberg, 2012).
126. Cho, M.J. & Juliano, R. Macromolecular versus small-molecule therapeutics: drug discovery, development and clinical considerations. *Trends in Biotechnology* **14**, 153-158 (1996).
127. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *Journal Natural Products* **75**, 311-335 (2012).
128. Schreiber, S.L. Organic synthesis toward small-molecule probes and drugs. *Proceedings of the National Academy of Sciences* **108**, 6699-6702 (2011).
129. Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and

- permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **64**, 4-17 (2012).
130. Baumann, A. Early Development of Therapeutic Biologics - Pharmacokinetics. *Current Drug Metabolism* **7**, 15-21 (2006).
 131. Craik, D.J., Fairlie, D.P., Liras, S. & Price, D. The future of peptide-based drugs. *Chemical Biology & Drug Design* **81**, 136-147 (2013).
 132. Kaspar, A.A. & Reichert, J.M. Future directions for peptide therapeutics development. *Drug Discovery Today* **18**, 807-817 (2013).
 133. Mason, J.M. design and development of peptides as a therapy intervention_Future_Med_Chem_2010.pdf. *Future Medicinal Chemistry* **2**, 1813-1822 (2010).
 134. Latham, P. Therapeutic peptides revisited. *Nature Biotechnology* **17**, 755-757 (1999).
 135. Sato, A.K., Viswanathan, M., Kent, R.B. & Wood, C.R. Therapeutic peptides: technological advances driving peptides into development. *Current Opinion in Biotechnology* **17**, 638-642 (2006).
 136. Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatsky, M. Synthetic therapeutic peptides: science and market. *Drug Discovery Today* **15**, 40-56 (2010).
 137. Hummel, G., Reineke, U. & Reimer, U. Translating peptides into small molecules. *Molecular BioSystems* **2**, 499-508 (2006).
 138. Loffet, A. Peptides as drugs: is there a market? *Journal of Peptide Science* **8**, 1-7 (2002).
 139. Ladner, R.C., Sato, A.K., Gorzelany, J. & de Souza, M. Phage display-derived peptides as therapeutic alternatives to antibodies. *Drug Discovery Today* **9**, 525-529 (2004).
 140. McGregor, D.P. Discovering and improving novel peptide therapeutics. *Current Opinion in Pharmacology* **8**, 616-619 (2008).
 141. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology* **285**, 2177-2198 (1999).
 142. Chen, J., Sawyer, N. & Regan, L. Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Science* **22**, 510-515 (2013).
 143. Arkin, M.R. & Wells, J.A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Natures Reviews Drug Discovery* **3**, 301-317 (2004).
 144. Clackson, T. & Wells, J.A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-386 (1995).
 145. Moerke, N.J. et al. Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G. *Cell* **128**, 257-267 (2007).
 146. Bogan, A.A. & Thorn, K.S. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* **280**, 1-9 (1998).
 147. Mullard, A. Protein-protein interaction inhibitors get into the groove. *Nature Reviews Drug Discovery* **11**, 173-175 (2012).
 148. Zhao, L. & Chmielewski, J. Inhibiting protein-protein interactions using designed molecules. *Current Opinion in Structural Biology* **15**, 31-34 (2005).

149. Wilson, A.J. Inhibition of protein-protein interactions using designed molecules. *Chemical Society Reviews* **38**, 3289-3300 (2009).
150. Pensak, D. Molecular modelling: scientific and technological boundaries. *Pure and Applied Chemistry* **61**, 601-603 (1989).
151. Leach, A.R. *Molecular Modelling: Principles and Applications*, 720 pp (Prentice Hall, 2000).
152. Raha, K. et al. The role of quantum mechanics in structure-based drug design. *Drug Discovery Today* **12**, 725-731 (2007).
153. McQuarrie, D.A. & Simon, J.D. *Physical Chemistry: A Molecular Approach*, (University Science Books, 1997).
154. Bennett, C.H., Bernstein, E., Brassard, G. & Vazirani, U. Strengths and Weaknesses of Quantum Computing. *SIAM Journal on Computing* **26**, 1510-1523 (1997).
155. Hansson, T., Oostenbrink, C. & van Gunsteren, W.F. Molecular dynamics simulations. *Current Opinion in Structural Biology* **12**, 190-196 (2002).
156. van Gunsteren, W.F. & Berendsen, H.J.C. Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Physics* **34**, 1311-1327 (1977).
157. Homeyer, N. & Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Molecular Informatics* **31**, 114-122 (2012).
158. Aqvist, J., Medina, C. & Samuelsson, J. A new method for predicting binding affinity in computer-aided drug design. *Protein Engineering* **7**, 385-391 (1994).
159. Kollman, P.A. et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research* **33**, 889-897 (2000).
160. Lu, N.D. & Kofke, D.A. Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling. *Journal of Chemical Physics* **114**, 7303-7311 (2001).
161. Kollman, P.A. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chemical Reviews* **93**, 2395-2417 (1993).
162. Virnau, P. & Muller, M. Calculation of free energy through successive umbrella sampling. *Journal of Chemical Physics* **120**, 10925-10930 (2004).
163. Gilson, M.K., Given, J.A., Bush, B.L. & McCammon, J.A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophysical Journal* **72**, 1047-1069 (1997).
164. Lybrand, T.P., McCammon, J.A. & Wipff, G. Theoretical calculation of relative binding affinity in host-guest systems. *chemistry* **83**, 833-835 (1986).
165. Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A. & Case, D.A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **119**, 7095-7104 (1998).
166. Pearlman, D.A. Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using Congeneric Series

- of Ligands to p38 MAP Kinase. *Journal of Medicinal Chemistry* **48**, 7796-7807 (2005).
167. Adler, M. & Beroza, P. Improved Ligand Binding Energies Derived from Molecular Dynamics: Replicate Sampling Enhances the Search of Conformational Space. *Journal of Chemical Information and Modeling* **53**, 2065-2072 (2013).
168. Lyne, D.P., Lamb, L.M. & Saeh, C.J. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring. *Journal of Medicinal Chemistry* **49**, 4805-4808 (2006).
169. Xu, L., Sun, H., Li, Y., Wang, J. & Hou, T. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *Journal of Physical Chemistry B* **117**, 8408-8421 (2013).
170. Sadiq, S.K., Wright, D.W., Kenway, O.A. & Coveney, P.V. Accurate Ensemble Molecular Dynamics Binding Free Energy Ranking of Multidrug-Resistant HIV-1 Proteases. *Journal of Chemical Information and Modeling* **50**, 890-905 (2010).
171. Caves, L.S.D., Evanseck, J.D. & Karplus, M. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science* **7**, 649-666 (1998).
172. Huang, N., Kalyanaraman, C., Bernacki, K. & Jacobson, M.P. Molecular mechanics methods for predicting protein-ligand binding. *Physical Chemistry Chemical Physics* **8**, 5166-5177 (2006).
173. Genheden, S. & Ryde, U. How to obtain statistically converged MM/GBSA results. *Journal of Computational Chemistry* **31**, 837-846 (2010).
174. Moll, A., Hildebrandt, A., Lenhof, H.-P. & Kohlbacher, O. BALLView: An object-oriented molecular visualization and modeling framework. *Journal of Computer-Aided Molecular Design* **19**, 791-800 (2006).
175. Wang, J., Hou, T. & Xu, X. Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. *Current Computer-Aided Drug Design* **2**, 287-306 (2006).
176. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144-1149 (1995).
177. Gohlke, H., Kiel, C. & Case, D.A. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. *Journal of Molecular Biology* **330**, 891-913 (2003).
178. Gohlke, H. & Case, D.A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem* **25**, 238-250 (2004).
179. Tan, C., Tan, Y.-H. & Luo, R. Implicit nonpolar solvent models. *The Journal of Physical Chemistry B* **111**, 12263-12274 (2007).
180. Gouda, H., Kuntz, I.D., Case, D.A. & Kollman, P.A. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers* **68**, 16-34 (2003).

181. Hou, T.J., Wang, J.M., Li, Y.Y. & Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **51**, 69-82 (2011).
182. Maffucci, I. & Contini, A. Explicit Ligand Hydration Shells Improve the Correlation between MM-PB/GBSA Binding Energies and Experimental Activities. *Journal of Chemical Theory and Computation* **9**, 2706-2717 (2013).
183. Hou, T. & Wang, J. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **51**, 69-82 (2011).
184. Srivastava, H.K. & Sastry, G.N. Molecular Dynamics Investigation on a Series of HIV Protease Inhibitors: Assessing the Performance of MM-PBSA and MM-GBSA Approaches. *Journal of Chemical Information and Modeling* **52**, 3088-3098 (2012).
185. Hayes, J. & Archontis, G. MM-GB(PB)SA Calculations of Protein-Ligand Binding Free Energies. in *Molecular Dynamics-Studies of Synthetic and Biological Macromolecules* (ed. Wang, L.) (2012).
186. Velazquez-Campoy, A., Todd, M.J. & Freire, E. HIV-1 protease inhibitors: Enthalpic versus entropic optimization of the binding affinity. *Biochemistry* **39**, 2201-2207 (2000).
187. Lafont, V. et al. Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chemical Biology & Drug Design* **69**, 413-422 (2007).
188. Ohtaka, H., Velazquez-Campoy, A., Xie, D. & Freire, E. Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. *Protein Science* **11**, 1908-1916 (2002).
189. Smyth, M.S. & Martin, J.H.J. x Ray crystallography. *Journal of Clinical Pathology-Molecular Pathology* **53**, 8-14 (2000).
190. UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Research* **43**, D204-D212 (2015).
191. Halle, B. Flexibility and packing in proteins. *Proceedings of the National Academy of Sciences* **99**, 1274-1279 (2002).
192. Chen, Y. et al. A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* **319**, 1092-1096 (2008).
193. Nishikawa, T. et al. Solution Structure of a Telomeric DNA Complex of Human TRF1. *Structure* **9**, 1237-1251 (2001).
194. Fairall, L., Chapman, L., Moss, H., de Lange, T. & Rhodes, D. Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Molecular Cell* **8**, 351-361 (2001).
195. Kim, S.H., Kaminker, P. & Campisi, J. TIN2, a new regulator of telomere length in human cells. *Nature Genetics* **23**, 405-412 (1999).
196. Sarthy, J.F. & Baumann, P. Apollo - Taking the Lead in Telomere Protection. *Molecular Cell* **39**, 489-491 (2010).

197. Lenain, C. et al. The Apollo 5' exonuclease functions together with TRF2 to protect telomeres from DNA repair. *Current Biology* **16**, 1303-1310 (2006).
198. Huang, H.-J. et al. Current developments of computer-aided drug design. *Journal of the Taiwan Institute of Chemical Engineers* **41**, 623-635 (2010).
199. Walker, R. & Steinbrecher, T. Amber Advanced Tutorial: MM-PBSA. Vol. 2012 (2006).
200. Hou, T.J., Wang, J.M., Li, Y.Y. & Wang, W. Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated From Docking. *Journal of Computational Chemistry* **32**, 866-877 (2011).
201. Genheden, S. & Ryde, U. Comparison of the Efficiency of the LIE and MM/GBSA Methods to Calculate Ligand-Binding Energies. *Journal of Chemical Theory and Computation* **7**, 3768-3778 (2011).
202. Peat, J.K., Barton, B. & ebrary, I. *Medical statistics : a guide to data analysis and critical appraisal*, 324 (Blackwell Pub., Malden, Mass., 2005).
203. Garton, M. Thesis, University of Nottingham (2012).
204. Godschalk, F., Genheden, S., Soderhjelm, P. & Ryde, U. Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations. *Physical Chemistry Chemical Physics* **15**, 7731-7739 (2013).
205. Garton, M. & Laughton, C. A comprehensive model for the recognition of human telomeres by TRF1. *Journal of Molecular Biology* **425**, 2910-2921 (2013).
206. Merz, J.K.M., Ringe, D. & Reynolds, C.H. *Drug Design [Electronic book] : Structure- and Ligand-Based Approaches*, 1 online resource (286 p.) (Cambridge University Press, Cambridge, 2010).
207. Wolf, A. & Kirschner, K.N. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11 center dot 23S subdomain. *Journal of Molecular Modeling* **19**, 539-549 (2013).
208. Meyer, T. et al. Essential Dynamics: A Tool for Efficient Trajectory Compression and Management. *J. Chem. Theory Comput.* **2**, 251-258 (2006).
209. Ricci, L. & Williams, K.P. Development of fluorescence polarization assays for the molecular chaperone Hsp70 family members: Hsp72 and DnaK. *Current Chmical Genomics* **2**, 90-95 (2008).
210. Merrifield, B. Concept and early development of solid-phase peptide synthesis. *Methods in Enzymology* **289**, 3-13 (1997).
211. Chandrudu, S., Simerska, P. & Toth, I. Chemical methods for peptide and protein production. *Molecules* **18**, 4373-4388 (2013).
212. Mitchell, A.R. Bruce Merrifield and solid-phase peptide synthesis: a historical assessment. *Biopolymers* **90**, 175-184 (2008).
213. Kent, S.B. Chemical synthesis of peptides and proteins. *Annual Review of Biochemistry* **57**, 957-989 (1988).
214. Carpino, L.A. et al. Complex polyfluoride additives in Fmoc-amino acid fluoride coupling processes. Enhanced reactivity and avoidance of stereomutation. *Organic Letters* **5**, 975-977 (2003).

215. Stawikowski, M. & Fields, G.B. Introduction to Peptide Synthesis. in *Current Protocols in Protein Science* (John Wiley & Sons, Inc., 2001).
216. Merrifield, B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85**, 2149-2154 (1963).
217. Merrifield, B. Solid phase synthesis. *Science* **232**, 341-347 (1986).
218. Carpino, L.A. & Han, G.Y. 9-Fluorenylmethoxycarbonyl amino-protecting group. *The Journal of Organic Chemistry* **37**, 3404-3409 (1972).
219. Barany, G. & Merrifield, R.B. A new amino protecting group removable by reduction. Chemistry of the dithiasuccinoyl (Dts) function. *Journal of The American Chemical Society* **99**, 7363-7365 (1977).
220. Chan, W.C. & White, P.D. *Fmoc solid phase peptide synthesis : a practical approach*, 346 (Oxford University Press, New York, 2000).
221. Palomo, J.M. Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides. *Royal Society of Chemistry Advances* **4**, 32658-32672 (2014).
222. Sheehan, J.C., Boshart, G.L. & Cruickshank, P.A. A Convenient Synthesis of Water-Soluble Cabodiimides. *Journal of Organic Chemistry* **26**, 2525-2528 (1961).
223. Al-Warhi, T.I., Al-Hazimi, H.M.A. & El-Faham, A. Recent development in peptide coupling reagents. *Journal of Saudi Chemical Society* **16**, 97-116 (2012).
224. Coin, I., Beyermann, M. & Bienert, M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nature Protocols* **2**, 3247-3256 (2007).
225. Forns, P. & Fields, G.B. The Solid Support. in *Solid-Phase Synthesis: A practical guide* 848 (CRC Press, 2000).
226. Fields, G.B. & Forns, P. The Solid Support. in *Solid-Phase Synthesis A Practical Guide* (eds. Kates, S. & Albericio, F.) (2000).
227. Kim, M., Park, Y.-S., Shin, D.-S., Lee, S. & Lee, Y.-S. PEG grafted AM SURE™ resin and its application to solid-phase peptide synthesis. *Tetrahedron Letters* **53**, 4576-4579 (2012).
228. Garcia-Ramos, Y., Paradis-Bas, M., Tulla-Puche, J. & Albericio, F. ChemMatrix((R)) for complex peptides and combinatorial chemistry. *Journal of Peptide Science* **16**, 675-678 (2010).
229. Woo, Y.-H., Mitchell, A. & Camarero, J. The Use of Aryl Hydrazide Linkers for the Solid Phase Synthesis of Chemically Modified Peptides. *International Journal of Peptide Research and Therapeutics* **13**, 181-190 (2007).
230. Camarero, J.A., Hackel, B.J., de Yoreo, J.J. & Mitchell, A.R. Fmoc-Based Synthesis of Peptide α -Thioesters Using an Aryl Hydrazine Support†. *The Journal of Organic Chemistry* **69**, 4145-4151 (2004).
231. El-Faham, A. & Albericio, F. Peptide coupling reagents, more than a letter soup. *chemical Reviews* **111**, 6557-6602 (2011).
232. Chantell, C.A., Onaiyekan, M.A. & Menakuru, M. Fast conventional Fmoc solid-phase peptide synthesis: a comparative study of different activators. *Journal of Peptide Science* **18**, 88-91 (2012).

233. Merrifield, B. Solid phase synthesis. Nobel lecture, 8 December 1984. *Bioscience Reports* **5**, 353-376 (1985).
234. Subiros-Funosas, R., El-Faham, A. & Albericio, F. PyOxP and PyOxB: the Oxyma-based novel family of phosphonium salts. *Organic & Biomolecular Chemistry* **8**, 3665-3673 (2010).
235. Carpino, L.A. & El-Faham, A. The diisopropylcarbodiimide/ 1-hydroxy-7-azabenzotriazole system: Segment coupling and stepwise peptide assembly. *Tetrahedron* **55**, 6813-6830 (1999).
236. Wehrstedt, K.D., Wandrey, P.A. & Heitkamp, D. Explosive properties of 1-hydroxybenzotriazoles. *Journal of Hazardous Materials* **126**, 1-7 (2005).
237. Subiros-Funosas, R., Prohens, R., Barbas, R., El-Faham, A. & Albericio, F. Oxyma: an efficient additive for peptide synthesis to replace the benzotriazole-based HOBt and HOAt with a lower risk of explosion. *Chemistry* **15**, 9394-9403 (2009).
238. Albericio, F., Bofill, J.M., El-Faham, A. & Kates, S.A. Use of Onium Salt-Based Coupling Reagents in Peptide Synthesis. *The Journal of Organic Chemistry* **63**, 9678-9683 (1998).
239. Albeiiicio, F., Chinchilla, R., Dodsworth, D.J. & Nájera, C. New trends in peptide coupling reagents. *Organic Preparations and Procedures International* **33**, 203-303 (2001).
240. Subiros-Funosas, R., Khattab, S.N., Nieto-Rodriguez, L., El-Faham, A. & Albericio, F. Advances in Acylation Methodologies Enabled by Oxyma-Based Reagents. *Aldrichimica Acta* **46**, 21-40 (2013).
241. Moerke, N.J. Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. *Current Protocols in Chemical Biology* **1**, 1-15 (2009).
242. de Jong, L.A., Uges, D.R., Franke, J.P. & Bischoff, R. Receptor-ligand binding assays: Technologies and Applications. *Journal of Chromatography B* **829**, 1-25 (2005).
243. Maguire, J.J., Kuc, R.E. & Davenport, A.P. Radioligand binding assays and their analysis. *Methods in Molecular Biology* **897**, 31-77 (2012).
244. Rajarathnam, K. & Rosgen, J. Isothermal titration calorimetry of membrane proteins - progress and challenges. *Biochimica et Biophysica Acta* **1838**, 69-77 (2013).
245. Langelaan, D.N., Ngweniform, P. & Rainey, J.K. Biophysical characterization of G-protein coupled receptor-peptide ligand binding. *Biochemistry and Cell Biology* **89**, 98-105 (2011).
246. Lynch, B.A., Loiacono, K.A., Tiong, C.L., Adams, S.E. & MacNeil, I.A. A fluorescence polarization based Src-SH2 binding assay. *Analytical Biochemistry* **247**, 77-82 (1997).
247. Torres, F.E., Recht, M.I., Coyle, J.E., Bruce, R.H. & Williams, G. Higher throughput calorimetry: opportunities, approaches and challenges. *Current Opinion Structural Biology* **20**, 598-605 (2010).
248. Smisterova, J., Ensing, K. & De Zeeuw, R.A. Methodological aspects of quantitative receptor assays. *Journal of Pharmaceutical and Biomedical Analysis* **12**, 723-745 (1994).
249. Huang, X. Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *Journal of Biomolecular Screening* **8**, 34-38 (2003).

250. Owicki, J.C. Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer. *Journal of Biomolecular Screening* **5**, 297-306 (2000).
251. Rossi, A.M. & Taylor, C.W. Analysis of protein-ligand interactions by fluorescence polarization. *Nature Protocols* **6**, 365-387 (2011).
252. Steel, R. Research Thesis, University of East Anglia (2014).
253. Serdyuk, I.N., Zaccai, N.R. & Zaccai, G. *Methods in molecular biophysics : structure, dynamics, function*, 1120 (Cambridge University Press, Cambridge, 2007).
254. Pope, A.J., Haupts, U.M. & Moore, K.J. Homogeneous fluorescence readouts for miniaturized high-throughput screening: theory and practice. *Drug Discovery Today* **4**, 350-362 (1999).
255. Sittampalam, G.S., Kahl, S.D. & Janzen, W.P. High-throughput screening: advances in assay technologies. *Current Opinion in Chemical Biology* **1**, 384-391 (1997).
256. Eggeling, C., Brand, L., Ullmann, D. & Jäger, S. Highly sensitive fluorescence detection technology currently available for HTS. *Drug Discovery Today* **8**, 632-641 (2003).
257. Sportsman, J. & Leytes, L. Miniaturization of homogenous assays using fluorescence polarization. *Drug Discovery Today* **1**, 27-32 (2000).
258. Kimple, A.J. et al. A High-Throughput Fluorescence Polarization Assay for Inhibitors of the GoLoco Motif/G-alpha Interaction. *Combinatorial Chemistry & High Throughput Screening* **11**, 396-409 (2008).
259. Weber, P.J., Bader, J.E., Folkers, G. & Beck-Sickinger, A.G. A fast and inexpensive method for N-terminal fluorescein-labeling of peptides. *Bioorganic & Medicinal Chemistry* **8**, 597-600 (1998).
260. Fischer, R., Mader, O., Jung, G. & Brock, R. Extending the applicability of carboxyfluorescein in solid-phase synthesis. *Bioconjugate Chemistry* **14**, 653-660 (2003).
261. Sittampalam, G.S. et al. *Receptor Binding Assays for HTS and Drug Discovery*, (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2012).
262. Feldser, D.M. & Greider, C.W. Short telomeres limit tumor progression in vivo by inducing senescence. *Cancer Cell* **11**, 461-469 (2007).
263. Lipps, H.J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends in Cell Biology* **19**, 414-422 (2009).
264. Singh, N. & Warshel, A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins-Structure Function and Bioinformatics* **78**, 1705-1723 (2010).
265. Hayes, J.M. et al. Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: The role of water molecules examined. *Proteins-Structure Function and Bioinformatics* **79**, 703-719 (2011).
266. Genheden, S., Kuhn, O., Mikulskis, P., Hoffmann, D. & Ryde, U. The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant. *Journal of Chemical Information and Modeling* **52**, 2079-2088 (2012).

267. Sawyer, J.S. et al. Synthesis and activity of new aryl- and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor-beta type I receptor kinase domain. *Journal of Medicinal Chemistry* **46**, 3953-3956 (2003).
268. Shekhar, C. In silico pharmacology: computer-aided methods could transform drug development. *Chemical Biology* **15**, 413-414 (2008).
269. Fletcher, M.D. & Campbell, M.M. Partially modified retro-inverso peptides: development, synthesis, and conformational behavior. *Chemical reviews* **98**, 763-796 (1998).
270. Ferenczy, G.G. & Keseru, M.G. Enthalpic Efficiency of Ligand Binding. *Journal of Chemical Information and Modeling* **50**, 1536-1541 (2010).
271. Pettersen, E.F. et al. UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612 (2004).
272. Case, D.A. et al. AMBER 12. *University of California, San Francisco* (2012).
273. Hornak, V. et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712-725 (2006).
274. Aqvist, J. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *The journal of Physical Chemistry* **94**, 8021-8024 (1990).
275. Ryckaert, J.P., Ciccotti, G. & Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).
276. Wu, X. & Brooks, B.R. Self-guided Langevin dynamics simulation method. *Chemical Physics Letters* **381**, 512-518 (2003).
277. Case, D.A. et al. AmberTools 12. *University of California, San Francisco* (2012).
278. Miller, B.R. et al. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **8**, 3314-3321 (2012).
279. Onufriev, A., Bashford, D. & Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**, 383-394 (2004).
280. Coutsias, E.A., Seok, C. & Dill, K.A. Using quaternions to calculate RMSD. *Journal of Computational Chemistry* **25**, 1849-1857 (2004).
281. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **14**, 33-38 (1996).
282. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. (2010).
283. Golding, E. Thesis, University of Nottingham (2012).
284. Magrane, M. & Consortium, U. UniProt Knowledgebase: a hub of integrated protein data. *The Journal of Biological Database and Curation* **2011**, bar009 (2011).

285. Leinonen, R., Nardone, F., Zhu, W. & Apweiler, R. UniSave: the UniProtKB sequence/annotation version database. *Bioinformatics* **22**, 1284-1285 (2006).
286. Gasteiger, E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **31**, 3784-3788 (2003).

Appendix

A1 TRF1 Expression and Purification

The hTRF1 expression and purification were performed to collect a pure TRF1 protein for the experimental assays. The protocol was taken from the PhD thesis of Emily Golding²⁸³, and the work was undertaken by me with the assistance and supervision of Dr Olivier Rannou. The construct (hTRF1 pET28aHis₆), which was prepared before, used for the protein expression. Then, to purify the protein, immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) were carried out. Finally, dialysis was performed.

A1.1 TRF1 Expression

10 µl of XL1 blue competent cells, which contained the construct, was transformed into a 20 ml plastic tube contained 10 ml Luria Broth (16 g/l Casein Digest Peptone, 5 g/l NaCl, 10 g/l yeast extract) and 10 µl (10 mg/ml) kanamycin antibiotic. It incubated overnight at 30 °C and 180 rpm to grow the cells. Then, 2 ml was withdrawn and centrifuged at 12000 rpm for 1 min; the supernatant layer was removed. This step was repeated 7 times. Next, the plasmid DNA was separated from the precipitated cells using GeneJET

plasmid miniprep kit (Thermo Scientific). Subsequently, 10 μ l of the plasmid DNA was transferred into 200 μ l of BL21 (DE3) strain of competent *E. coli* cells, mixed gently, incubated for 30 min in the ice, put in the water bath for 45 seconds at 42 °C and then to the ice for 2 min. After that, the cells were smeared onto Luria broth (LB)-agar plates contained chloramphenicol 34 μ l/ml and kanamycin 30 μ l/ml; the incubation was carried out for 24 hours at 37 °C. The colonies appeared were selected to inoculate 4 plastic tube cultures; each of the cultures contained 10 ml LB, chloramphenicol and kanamycin. The cultures were grown overnight in a shaking incubator at 37 °C. Add 10 ml from the overnight cultures to each of the 4 flasks (starter culture), which were prepared before and consisted of 500 ml of fresh LB salt (15.5 g broth in water); then, 0.5 ml chloramphenicol, 0.5 ml kanamycin and 1 drop of antifoam agent were added. The starter cultures were grown to optical density (O.D.)₅₉₅ 0.66, after which 0.5 ml of 1 M isopropyl β -D-1-thiogalactopyranoside (IPTG) was added. Next, the flasks were incubated for 5 hours at 30 °C and 180 rpm, followed by centrifugation at 4000 rpm and 4 °C for 20 min.

The pellet was collected and the supernatant was discarded. Thereafter, 150 μ l of the cocktail protease inhibitor and 40 ml of TED₃₅₀ pH 8.5 buffer (50 mM Tris-HCl pH 8.5, 1 mM EDTA, 1 mM DTT, 300 mM NaCl) were added to the pellet. The sample was sonicated at 20 amplitude microns for 30 seconds using a MSE Soniprep 150 sonicator. The sonication was repeated for 6 times and after each sonication put the sample in the ice. After lysis of the cells, it was centrifuged at 4000 rpm and 4 °C for 40 min. The supernatant layer was

separated and filtered by syringe and filter 0.2 μ M (Sartorius) in order to remove impurities. The expressed TRF1 protein was identified through sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure A.1).

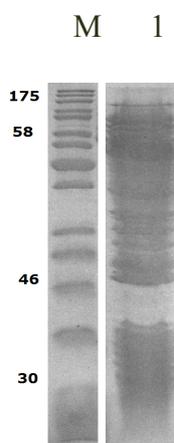


Figure A.1. SDS-PAGE of hTRF1 before purification. 1 is the protein and M is the marker.

A1.2 TRF1 Protein Purification

A1.2.1 Affinity Chromatography

Affinity chromatography also termed IMAC was performed for His₆-tag purification using an AKTA prime purification system. 5 ml HiTrapTM HP Chelating columns (GE Healthcare) was (re) charged with Ni₂⁺ using 0.1 M NiSO₄, following the manufacturers instructions. The column was first washed with 0.1 M EDTA, water and 0.1 M nickel. Two different types of buffer were used, buffer A (20 mM sodium phosphate buffer pH 8, 500 mM NaCl) and buffer B (20 mM sodium phosphate buffer pH 8, 500 mM NaCl and 500 mM imidazole). The column was equilibrated with IMAC binding buffer (buffer B). Loading clarified supernatant was started with buffer A, which resulted in

binding His-tag TRF1 to the nickel of the column, while other impurities were eluted. Then, buffer B washed the column and the protein eluted with an imidazole gradient 5-500 mM over 120 ml. The process was monitored spectroscopically measuring absorption at a wavelength of 280 nm. 4 ml fraction collected and was analysed by SDS PAGE (Figure A.2). Samples containing purified protein were pooled and stored at 4 °C.

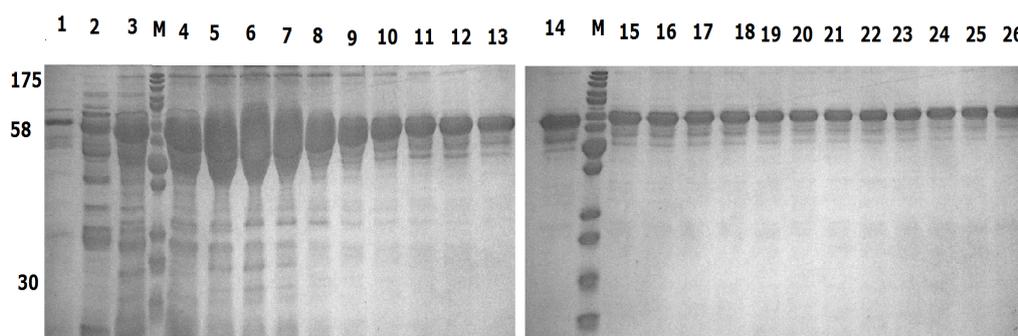


Figure A.2. SDS-PAGE monitoring hTRF1 (58 KDa) elution from HiTrap TM HP Chelating columns.

A1.2.2 Size Exclusion Chromatography

Size exclusion chromatography is also known as gel filtration, was performed using a Superdex[®] 75 (HiLoad 26/60 prep grade, GE healthcare) at 2 ml/min in TED₃₅₀ buffer. S75 column was connected to the AKTA prime system to carry out gel filtration. TRF1 elution was noted after running 124 ml and collected. SDS PAGE was performed for 21 fractions using two gels; the first gel was for fractions 1-13 and the second gel was for fractions 14-21 (Figure A.3).

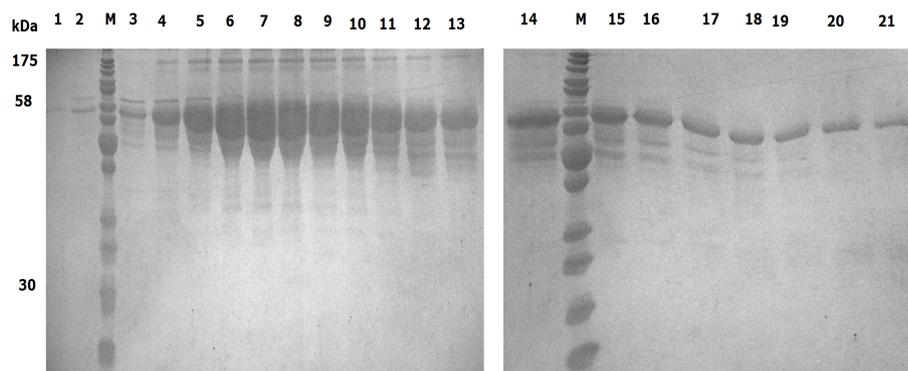


Figure A.3. SDS-PAGE monitoring hTRF1 (58 KDa) elution from the HiLoad 26/60 Superdex 75 prep grade gel filtration column.

A1.3 Dialysis

The protein sample in the dialysis tube was placed in a beaker contained 500 ml TED₃₅₀ buffer and 50% *v/v* glycerol in a cold room. The buffer was changed 3 times during 36 hours. The final concentrated volume of the protein was 10 ml, which was snap frozen in liquid nitrogen as 0.1 ml aliquots and stored at -80 °C.

A1.4 Calculating hTRF1 Concentration

The concentration of the purified hTRF1 protein was determined by measuring an absorbance at 280 nm using the NanoDrop spectrophotometer (Thermo Scientific). Sequence of the protein is composed of 449 amino acids and the MW is 51327 Da. For calculating molar absorptivity of hTRF1 stock solution, ProtParam tool of ExPasy (<http://web.expasy.org/protparam/>)²⁸⁴⁻²⁸⁶ was used and the Beer-Lambert law calculated the concentration of it (Equation A.1).

$$A_{280} = \sum b c \quad (\text{Beer-Lambert Law}) \quad \text{Equation A.1}$$

A_{280} is an absorbance at 280 nm, \sum is the molar absorptivity in $\text{L mol}^{-1} \text{cm}^{-1}$, b is the length of the light path in cm and c is the concentration in molarity.

When, $A_{280} = 1.77$, $\sum = 42440 \text{ L mol}^{-1} \text{cm}^{-1}$ and $b = 1 \text{ cm}$; hence, $c = 41.7 \text{ }\mu\text{M}$, or 2.14 mg/ml .¹⁹²

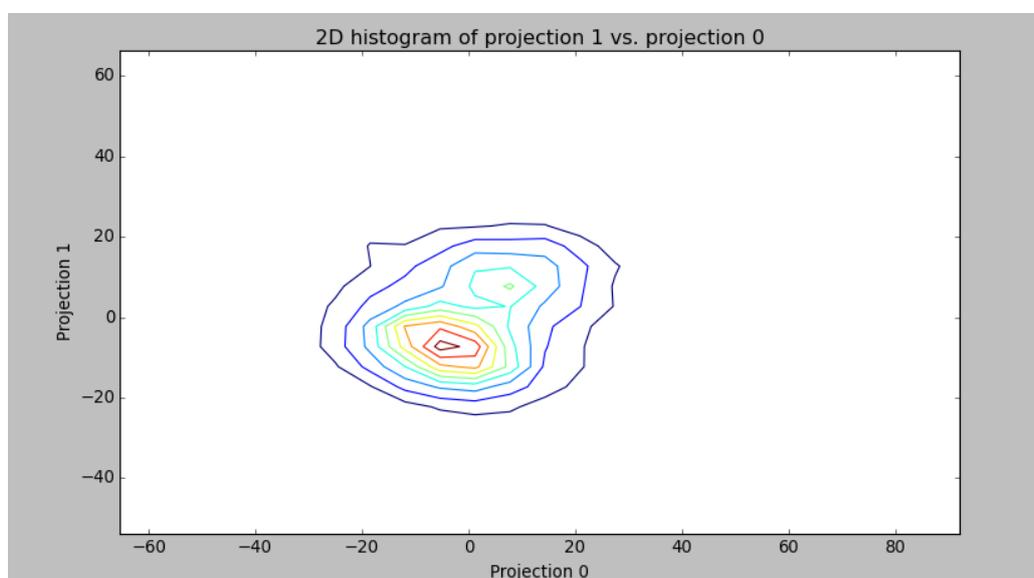
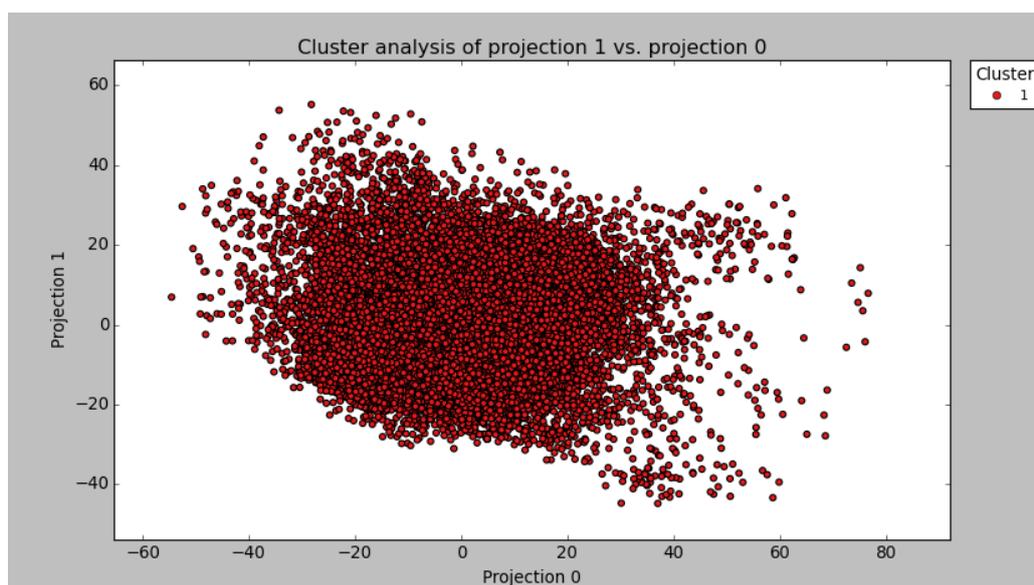
A2. Amino acid Abbreviations (IUPAC)

Abbreviation		Amino acid name
Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamic acid
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine
β-Ala	β-A	β-Alanine

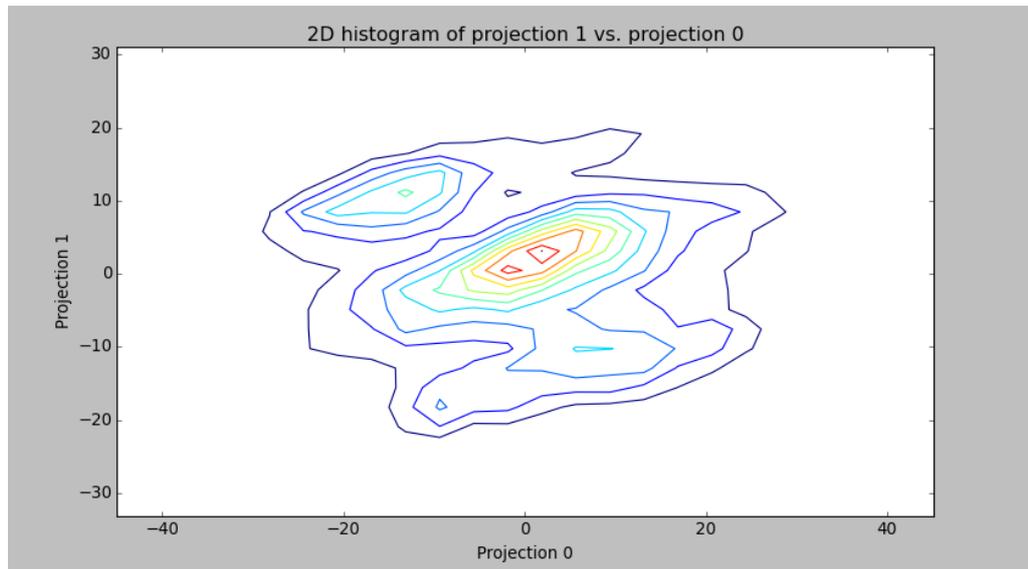
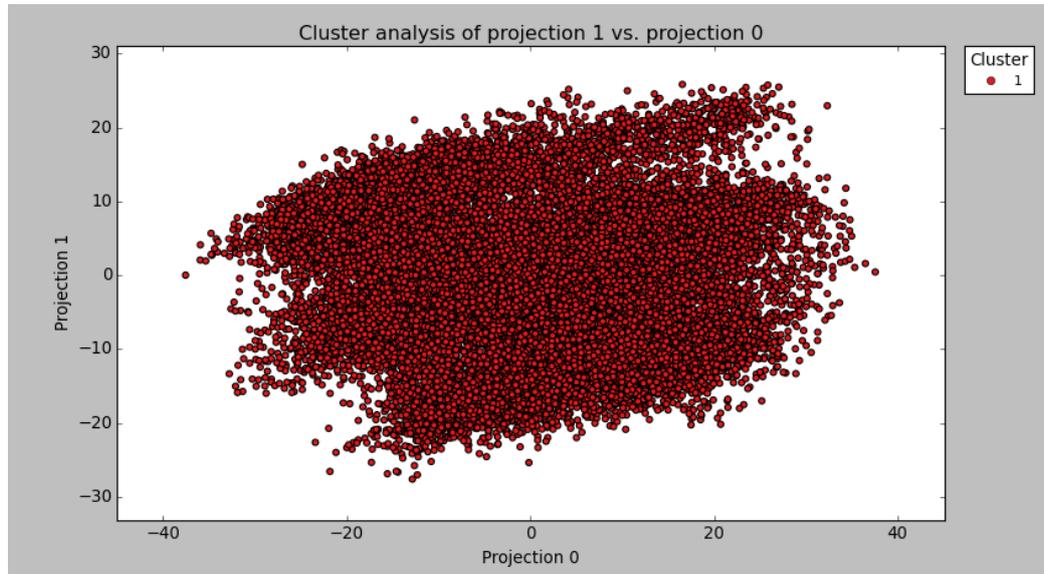
A3. Principal Component Analysis of the TIN2 Peptide

Analogues

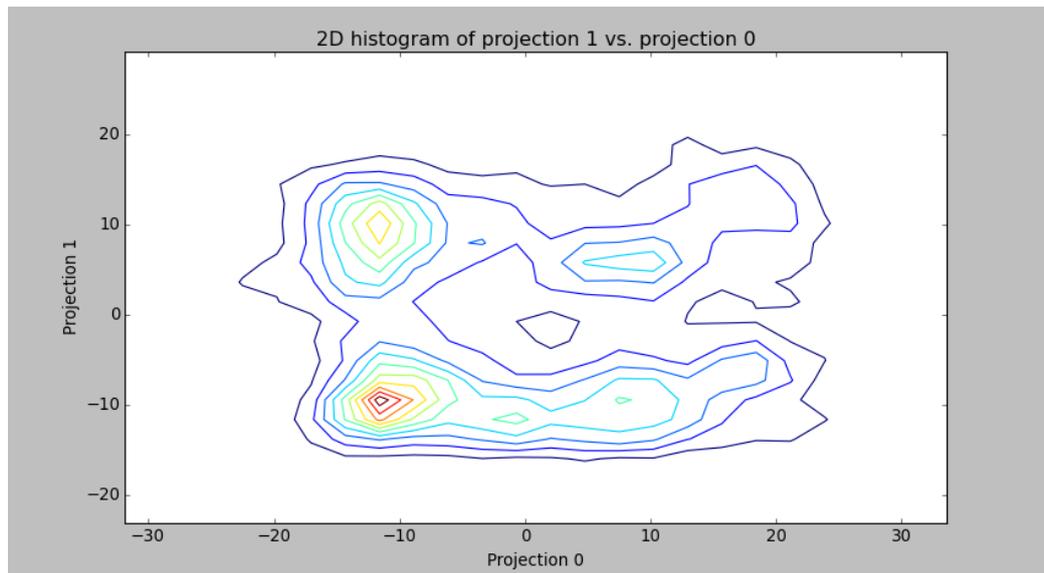
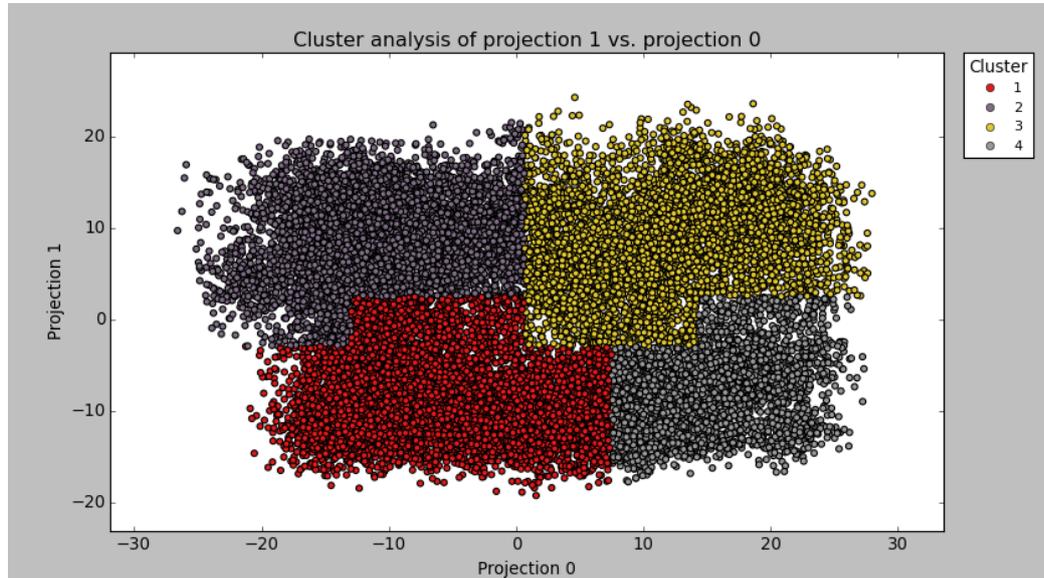
A3.1 TS01

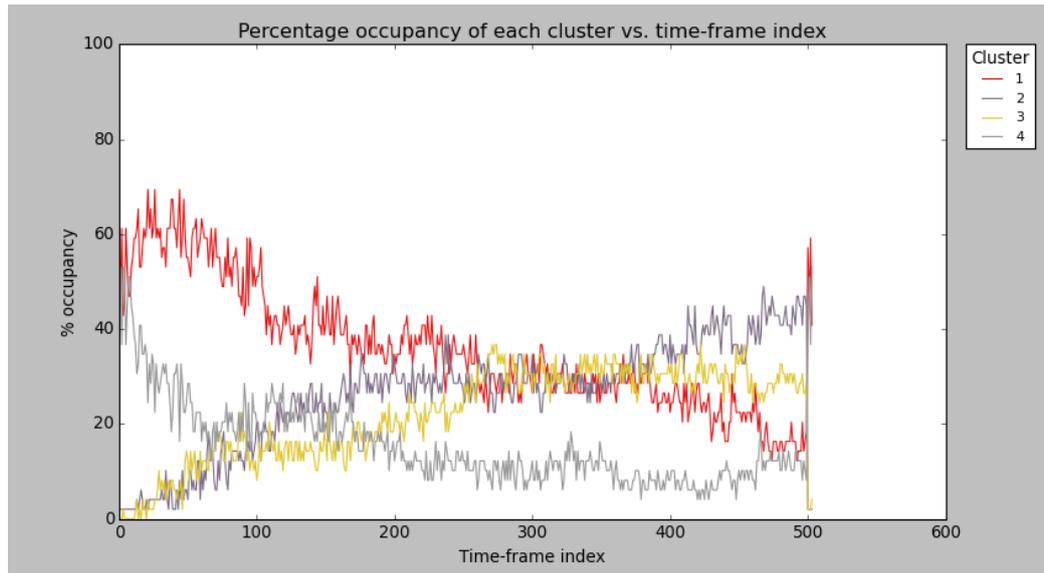


A3.2 TS02

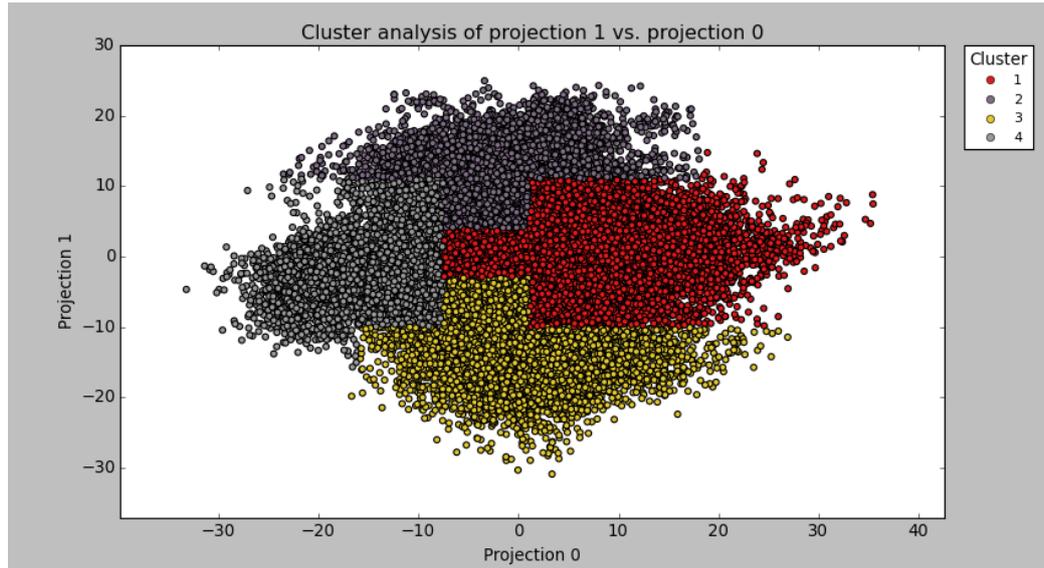


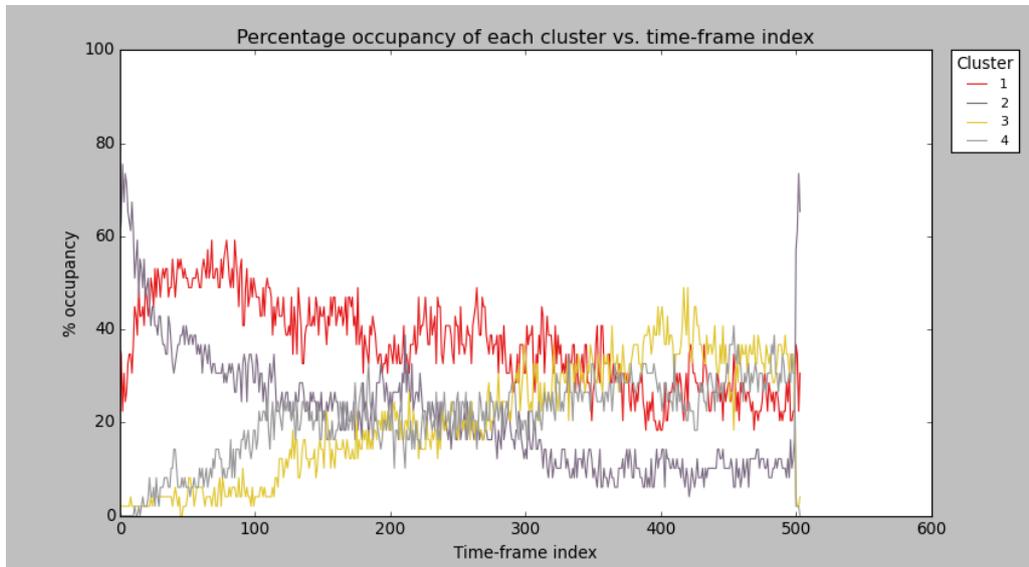
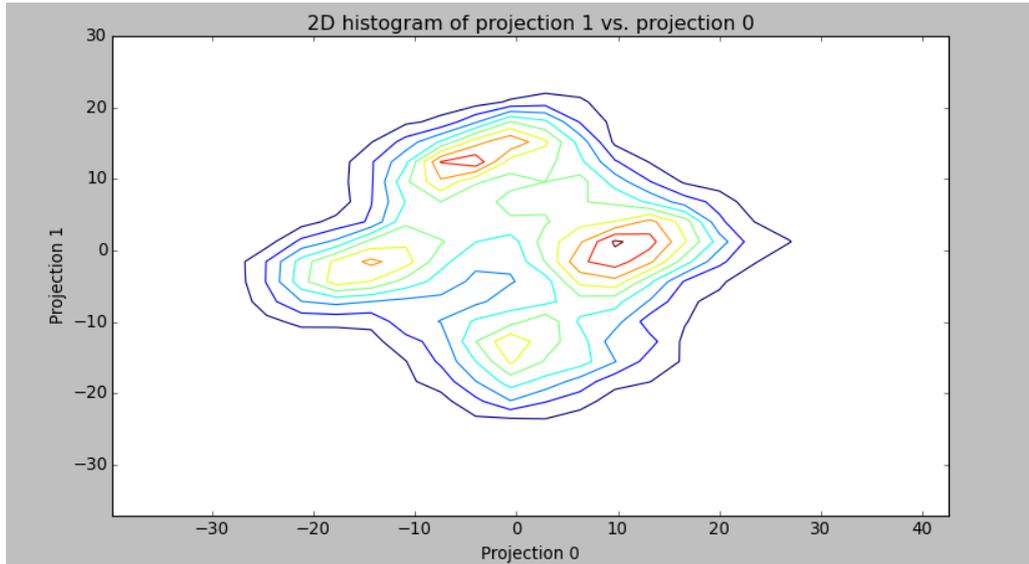
A3.3 TS03



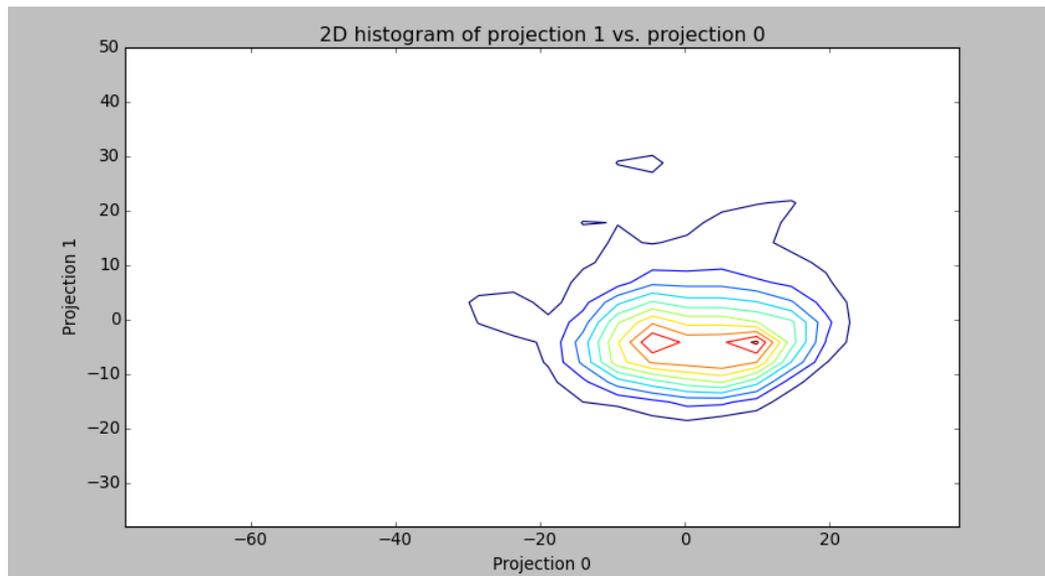
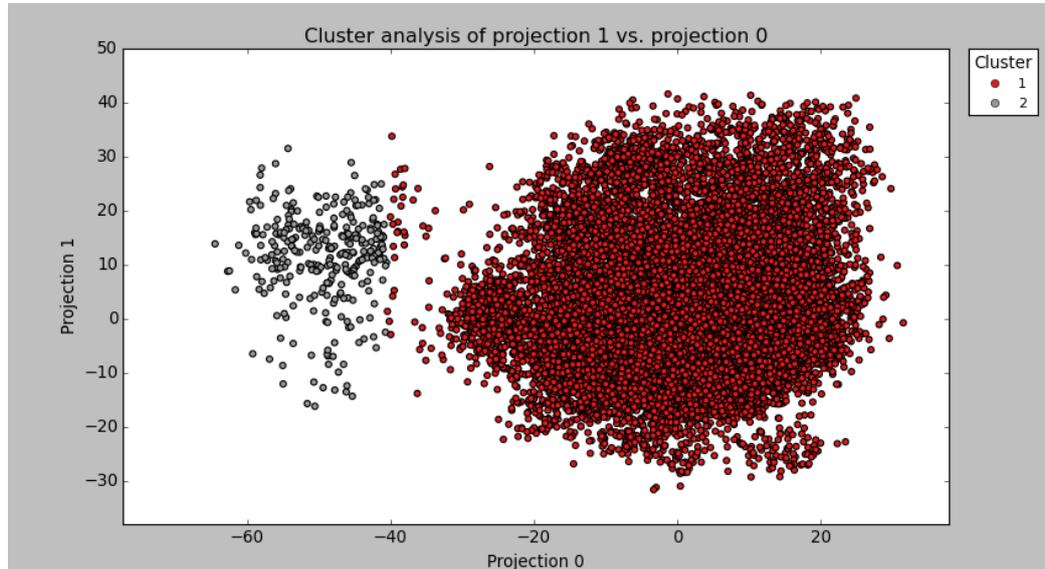


A3.4 TS04

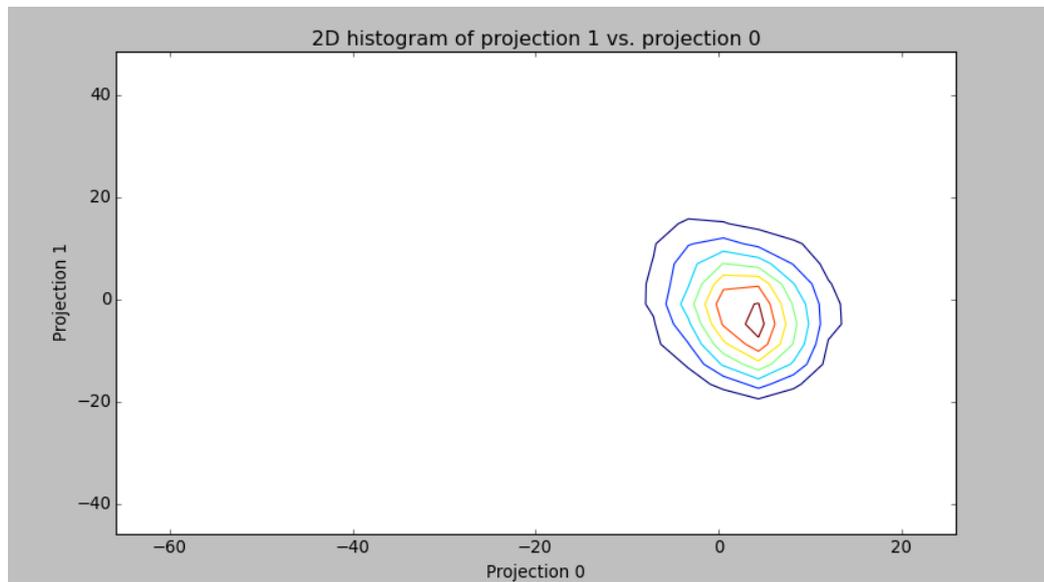
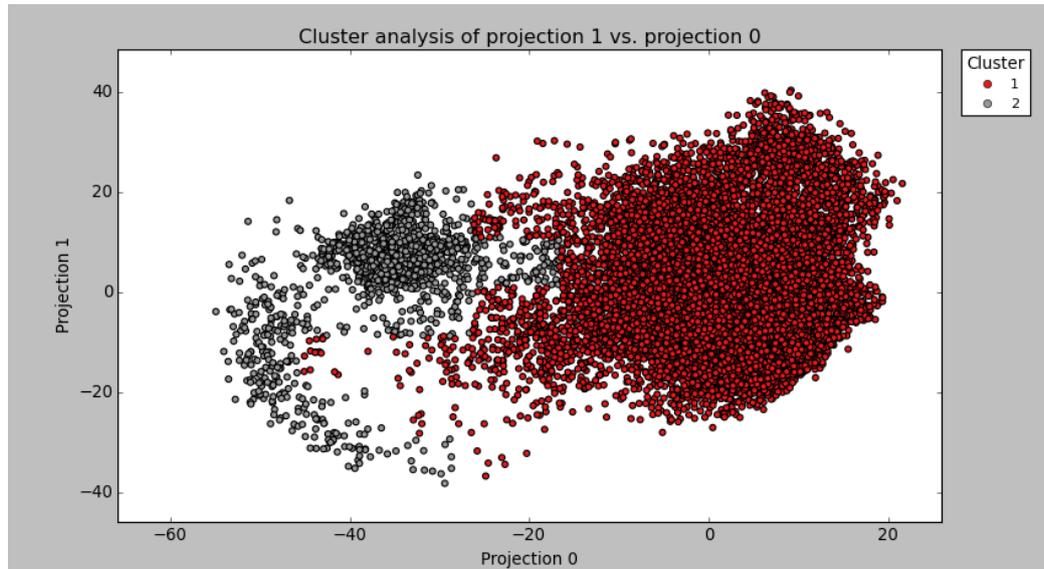


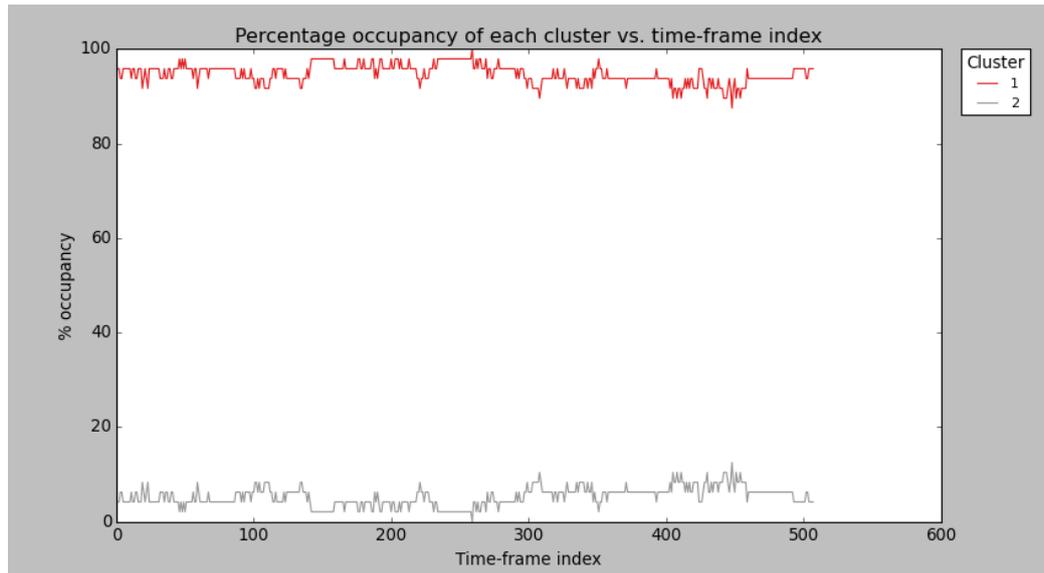


A3.5 TS05

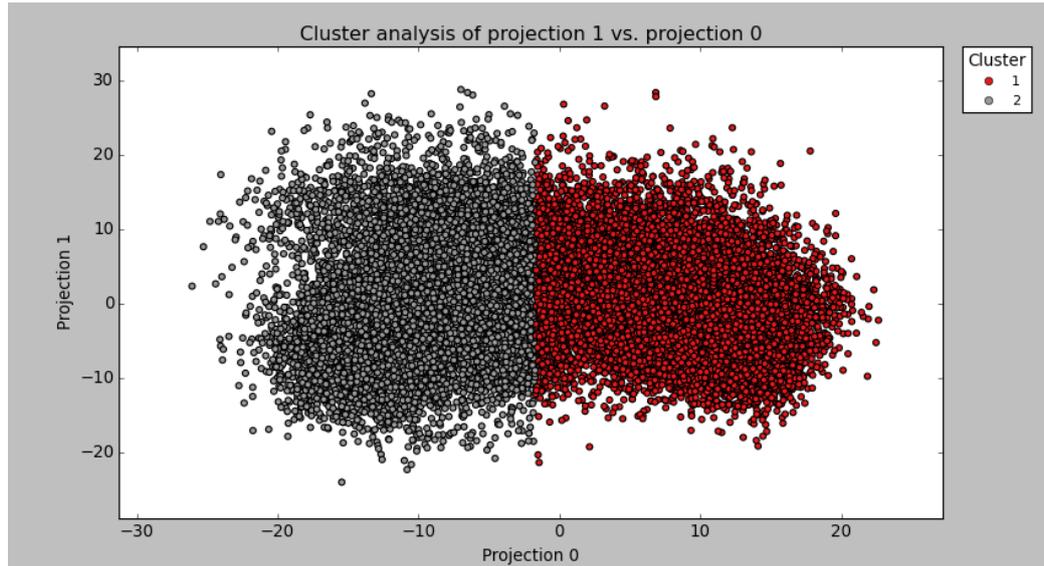


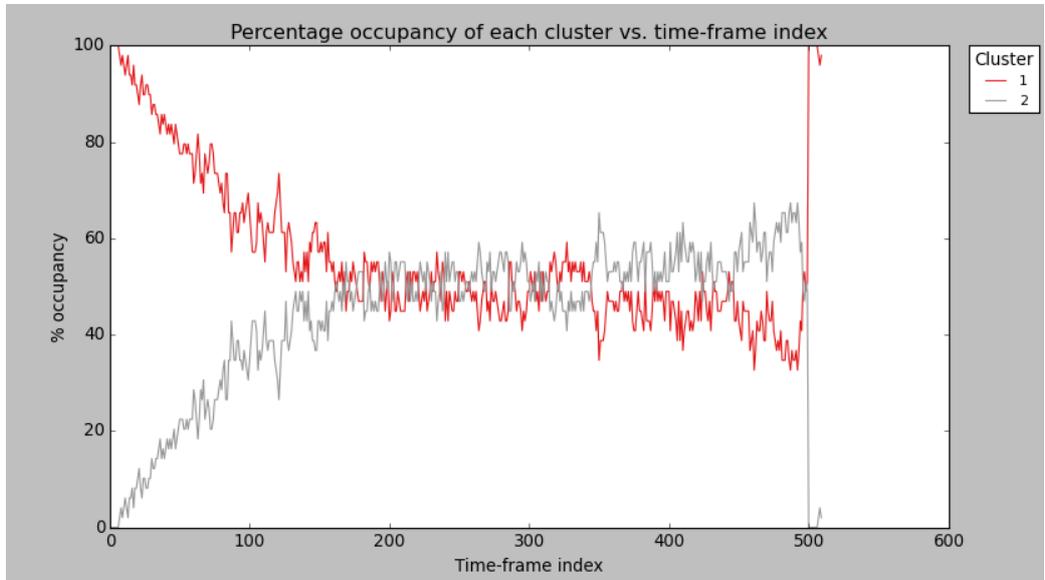
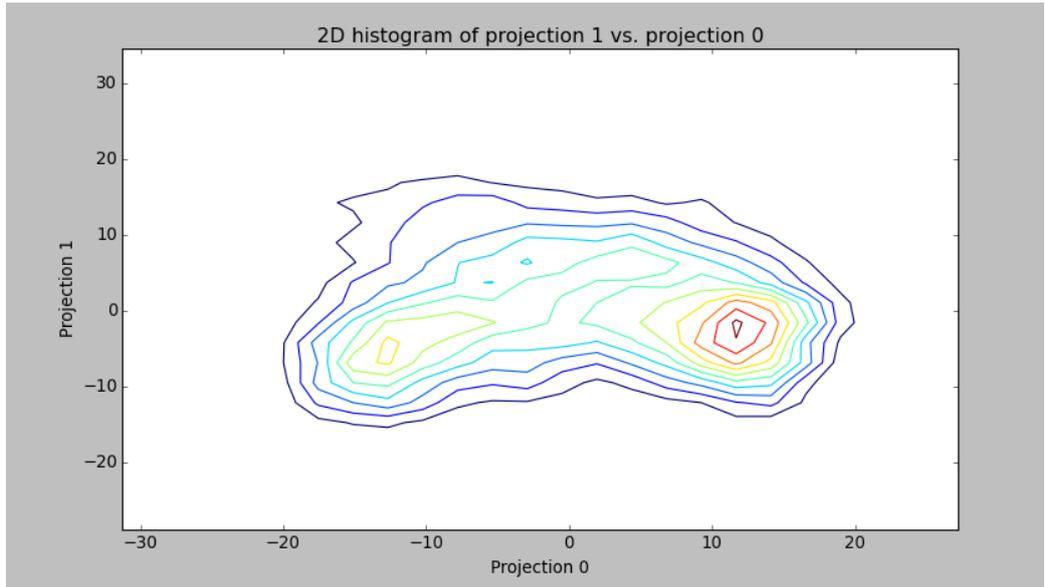
A3.6 TS06



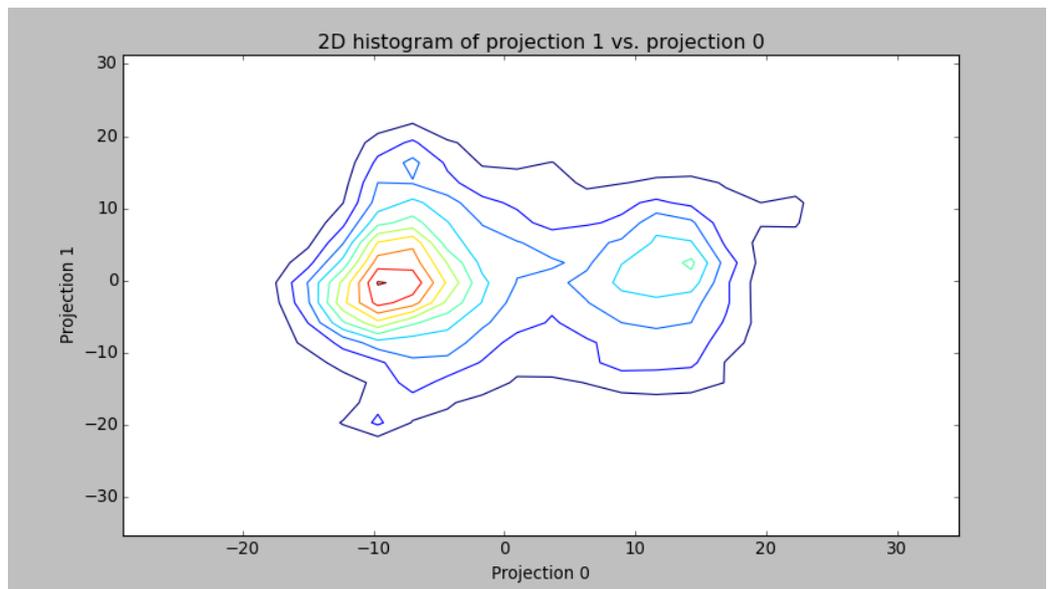
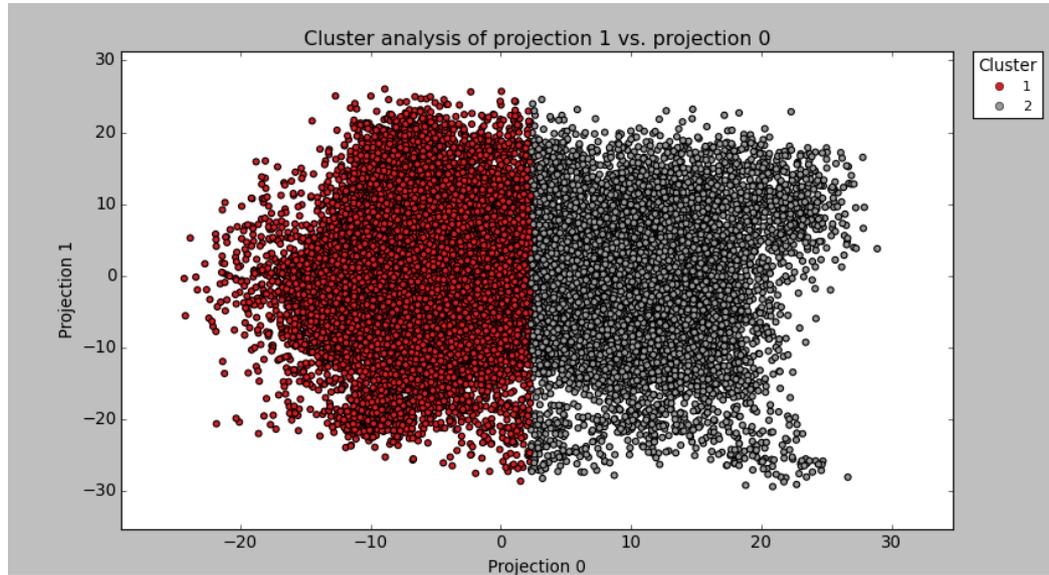


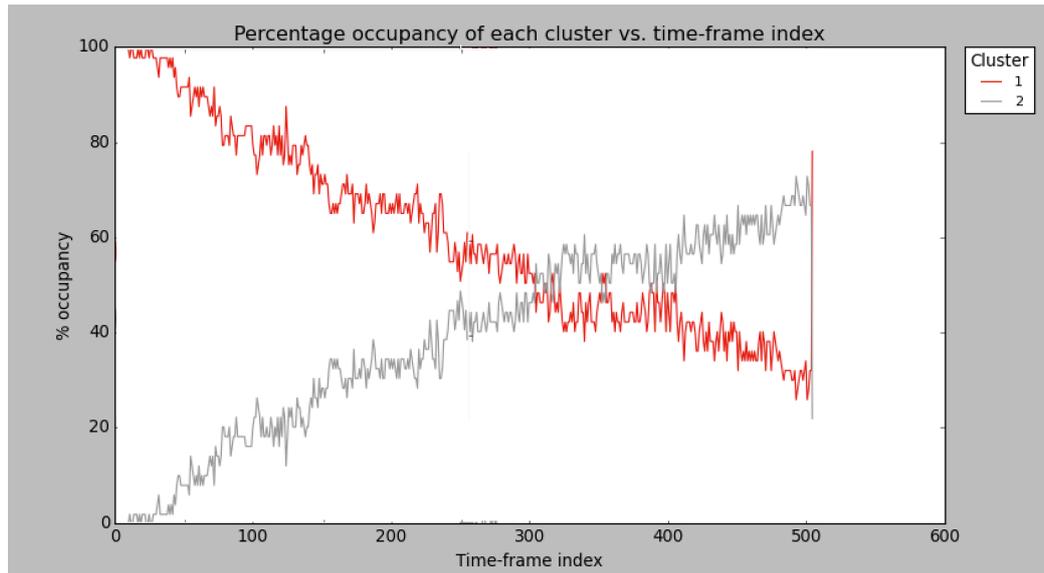
A3.7 TS07



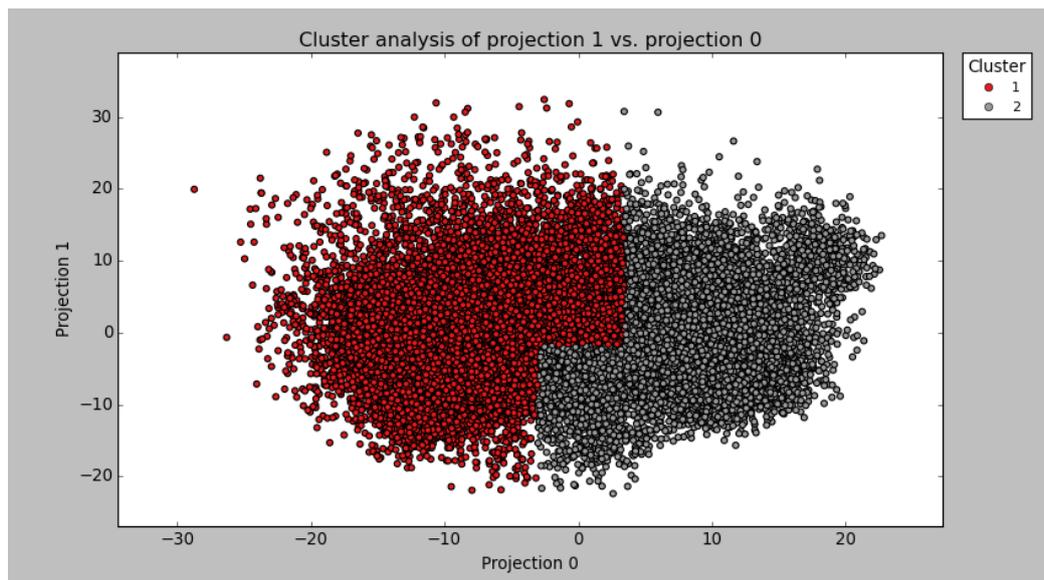


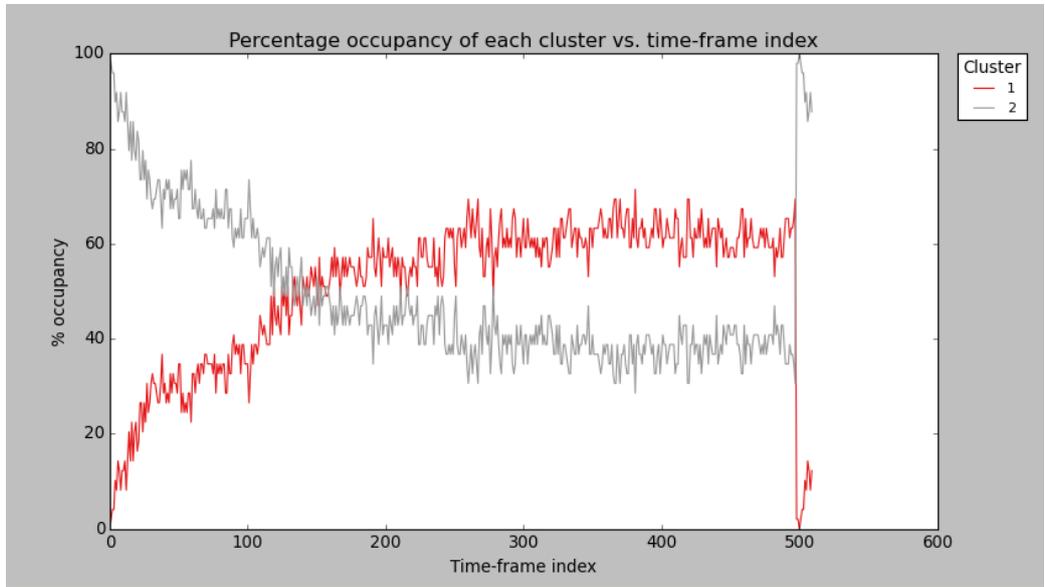
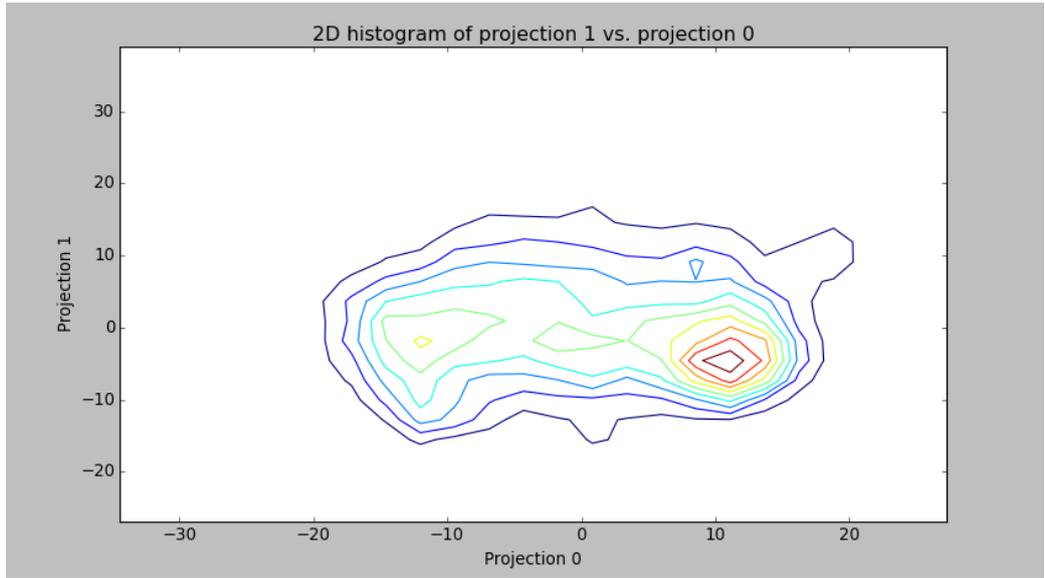
A3.8 TS08



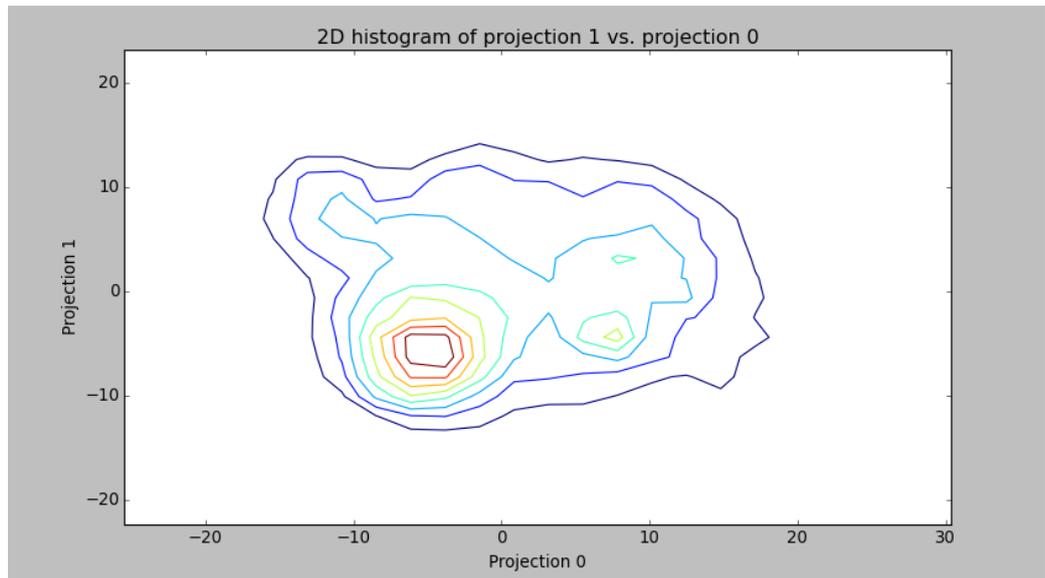
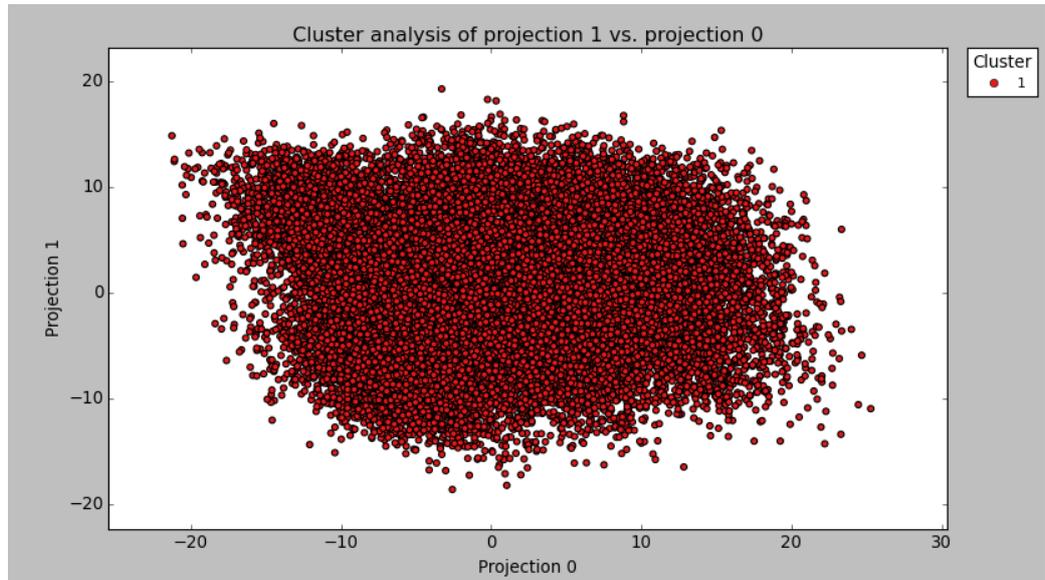


A3.9 TS09

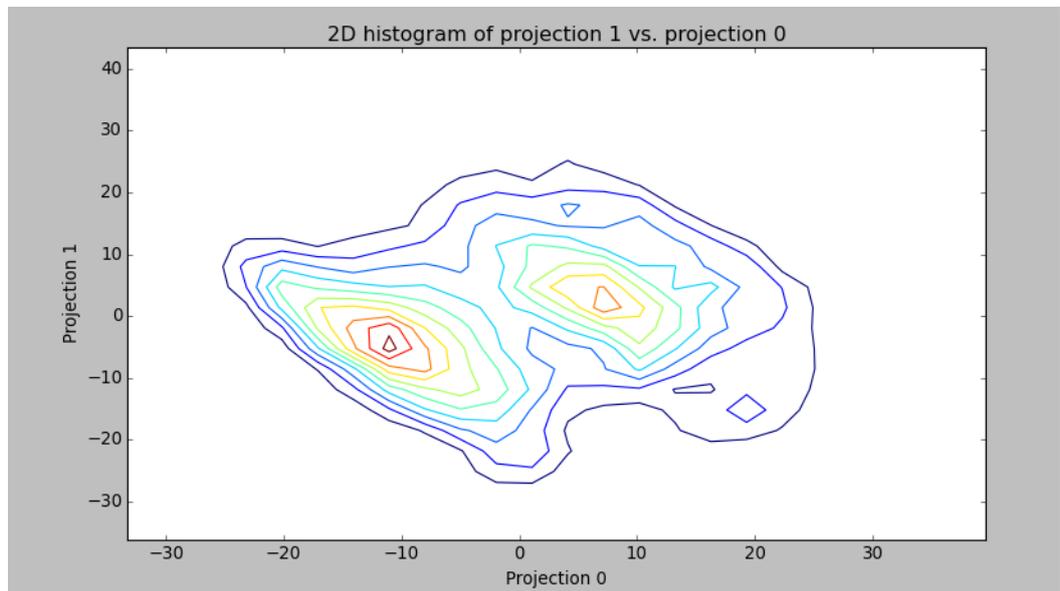
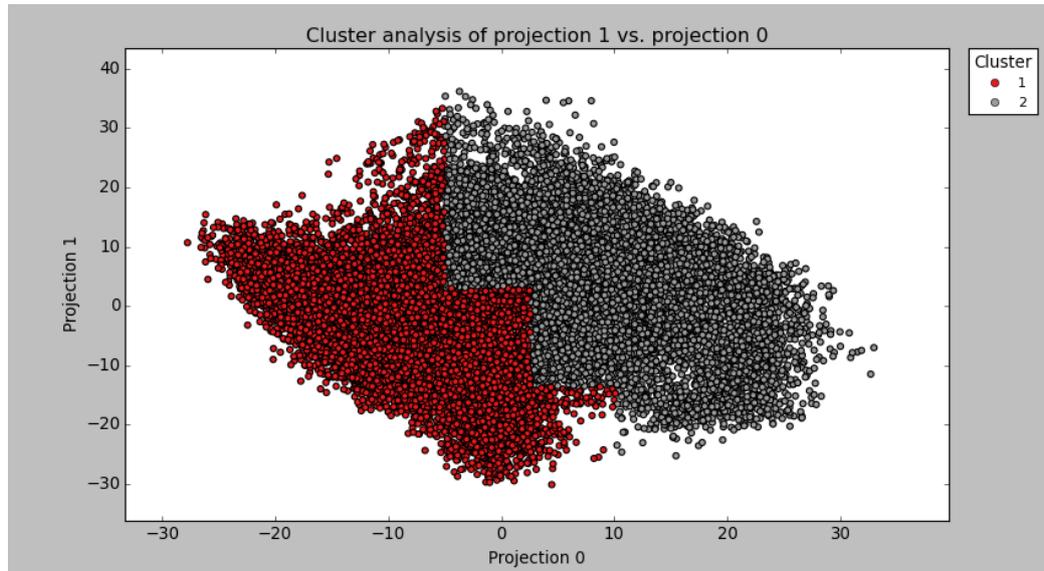


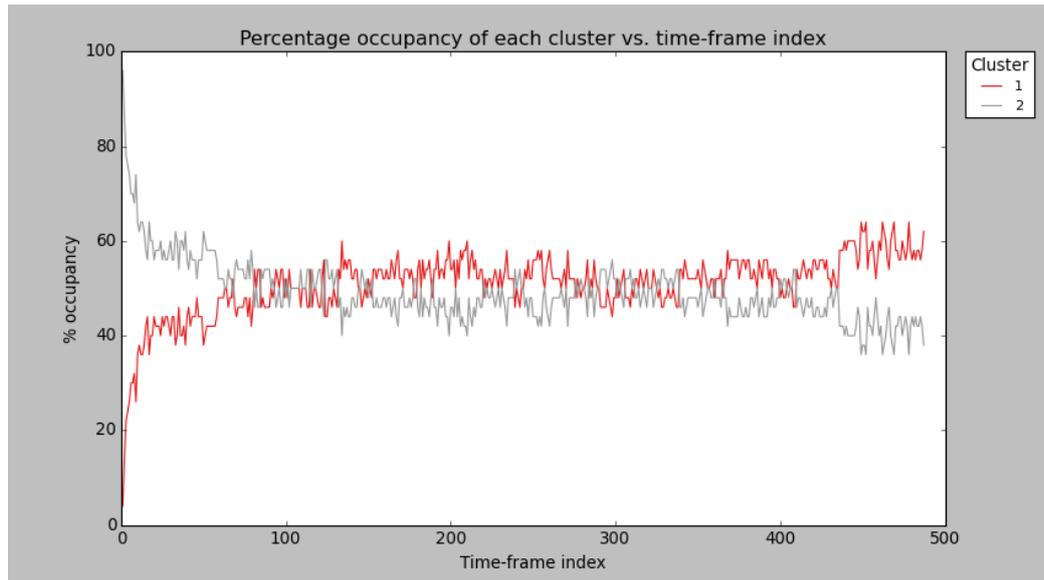


A3.10 TS10

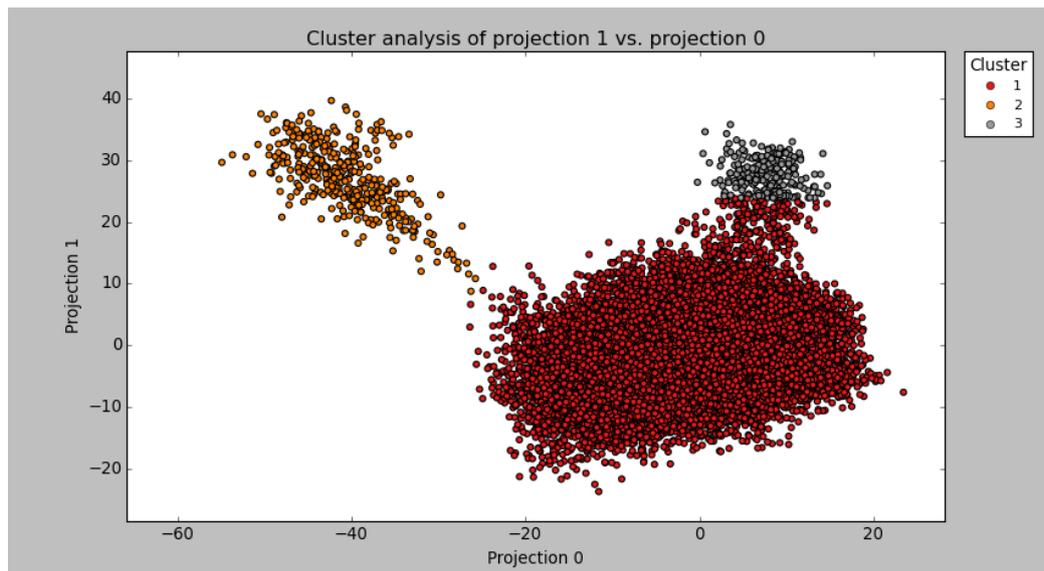


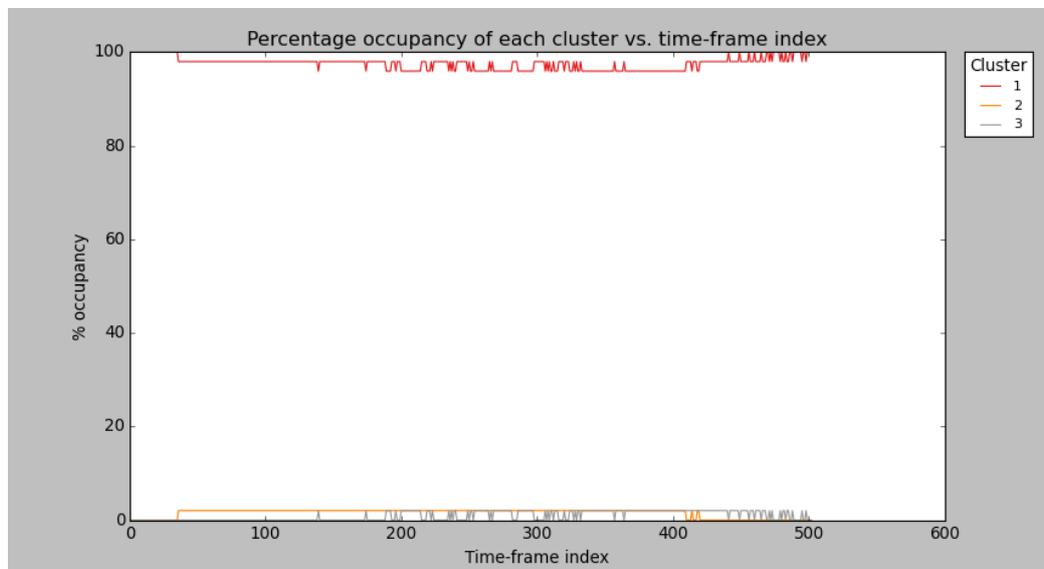
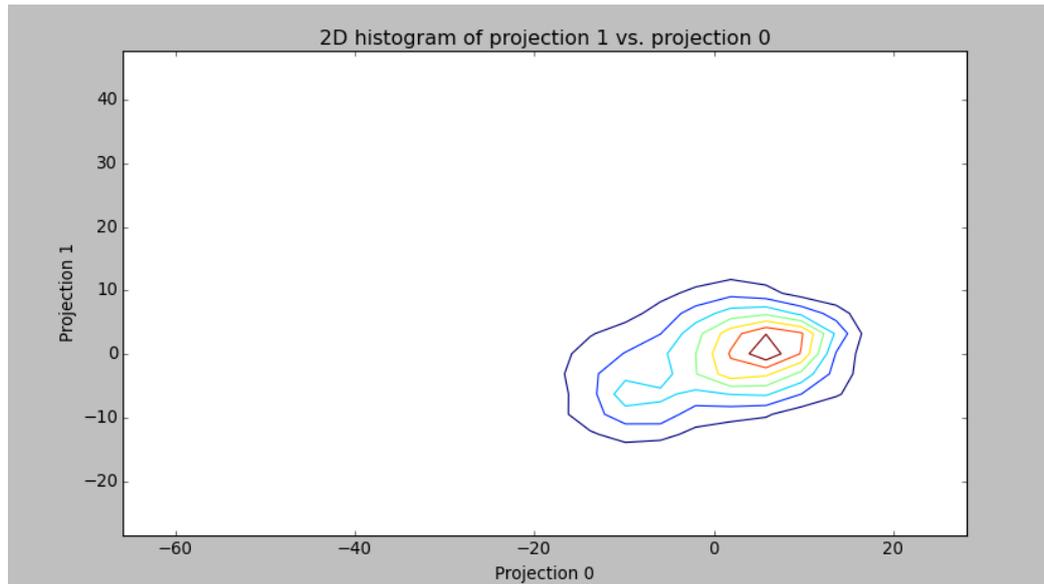
A3.11 TS11





A3.12 TS12





A3.13 TS13

