

Department Chemie,
Lehrstuhl für Biomolekulare NMR-Spektroskopie

Molecular recognition of splicing factors involved in Fas alternative splicing

Pravin Kumar Ankush Jagtap

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Bernd Reif

Prüfer der Dissertation:

1. Prof. Dr. Michael Sattler
2. Prof. Dr. Johannes Buchner
3. Prof. Dr. Dierk Niessing

Die Dissertation wurde am 12.07.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 15.09.2016 angenommen.

DECLARATION

I hereby declare that parts of this thesis have already been published in the following scientific journals:

Wang I, Hennig J, **Jagtap PKA**, Sonntag M, Valcarcel J, Sattler M. 2014. Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1. *Nucleic acids research* 42: 5949-5966.

Table of content

Abstract	7
Chapter 1 Introduction I: Biological background	11
1.1 Splicing and spliceosome assembly	12
1.1.1 Pre-mRNA splicing	12
1.1.2 Alternative splicing	14
1.2 Regulation of Fas alternative splicing.....	16
1.2.1 Role of TIA-1 and U1C proteins in Fas alternative splicing	17
1.2.2 Role of UHM-ULM interactions in Fas alternative splicing	22
1.2.3 Targeting spliceosome assembly with inhibitors.....	25
Chapter 2 Introduction II: Techniques used for integrated structural biology	29
2.1 NMR spectroscopy.....	30
2.1.1 Principle of NMR spectroscopy	31
2.1.2 Larmor precession	31
2.1.3 Vector formalism.....	32
2.1.4 Product operator formalism.	32
2.1.5 NMR experiments for protein assignment.....	34
2.1.6 Structure calculations using NMR assignments	37
2.1.7 Protein dynamics by NMR	38
2.2 X-ray crystallography	42
2.2.1 Protein crystallization.....	42
2.2.2 Principle of X-ray crystallography	42
2.2.3 Braggs Law.....	43
2.2.4 Molecular replacement	45
2.3 Small Angle X-ray Scattering	46
Scope of the Thesis	49
Chapter 3 Materials and Methods	51
3.1 Materials	52
3.1.1 Buffers	52
3.1.2 Media.....	52
3.1.3 ¹⁵ N labelled M9 salts	53
3.1.4 Trace elements solution.....	53
3.2 Methods.....	54
3.2.1 Protein expression and purification	54
3.2.2 NMR titrations.....	56

3.2.3	NMR structure calculation and validation of TIA-1 RRM1	56
3.2.4	Assignment of backbone and side-chain resonances of TIA-1 RRM1.....	57
3.2.5	NMR relaxation measurements	57
3.2.6	Small angle X-ray scattering experiments.....	58
3.2.7	Crystallization of TIA-1 RRM1-GS15-U1C30-61	59
3.2.8	SPF45 UHM-cyclic peptide crystallization and data processing	59
3.2.9	Puf60-small molecules crystallization and data processing	60
3.2.10	Isothermal Titration Calorimetry (ITC).....	61
3.2.11	Fluorescence Polarization Assay	61
3.2.12	High-throughput screening.....	63
3.2.13	AlphaScreen assay.....	64
Chapter 4 Structural insights into the interaction of TIA-1 with RNA and U1C		67
4.1	RRM1, 2, 3 forms a compact shape in the presence of RNA	68
4.1.1	NMR relaxation studies of TIA-1 RRM1,2,3-RNA complex	68
4.1.2	SAXS analysis of TIA-1 RRM1,2,3-RNA complex	70
4.2	NMR structure of TIA-1 RRM1 domain	71
4.3	Concentration dependent dimerization of U1C.....	74
4.3.1	Backbone assignment of U1C (1-61)	74
4.3.2	SAXS analysis of U1C (1-61).....	77
4.3.3	ITC experiments to study U1C dimerization.....	81
4.4	Interaction between U1C and RRM1	82
4.4.1	Backbone assignment of U1C 30-61	85
4.4.2	Interaction of U1C 30-61 and TIA-1 RRM1	87
4.5	Structure of RRM1-U1C complex	90
4.5.1	Linking TIA-1 RRM1 and U1C 30-61 peptide with GS linker for structural studies	90
4.5.2	Crystal structure of RRM1-GS15-U1C30-61	91
4.6	Discussion.....	94
4.6.1	Current understanding of different roles of U1C domains	94
4.6.2	Structural model for TIA-1 U1 snRNP interaction.....	96
Chapter 5 Rational design of cyclic peptide inhibitors of SPF45 UHM domain		99
5.1	Crystal structure of SPF45 UHM-cyclic peptide complex	101
5.2	Structure based design of new peptides	103
5.3	<i>In vitro</i> splicing activity of P10.....	106
Chapter 6 Targeting UHM domains with small molecules to modulate pre-mRNA splicing		111
6.1	High throughput screening for hit identification.....	112

6.1.1	Development of fluorescence polarization assay for high throughput screening	112
6.1.2	Results of high throughput screening	114
6.2	Hit validation	115
6.2.1	FP assays titrations	115
6.2.2	NMR titrations.....	116
6.3	Hit optimization	118
6.3.1	Medicinal chemistry based approach for hit optimization	118
6.3.2	Crystallization of positive hits with UHM domain.....	120
6.3.3	Analysis of Thx-Puf60 UHM-TOK116 crystal structure	122
6.3.4	Structure based hit optimization	124
6.4	UHM inhibitors stall spliceosome assembly.....	126
6.5	UHM inhibitors target all UHM domains	128
Conclusions and Outlook		131
Appendix		135
	Protein sequences.....	136
	NMR chemical shift assignments of TIA-1 RRM1	138
	NMR backbone chemical shifts of U1C 30-61	145
	Chemical structures of the compounds	146
Abbreviations		153
List of Figures		155
List of Tables		157
Acknowledgements		159
References		161

Molecular recognition of splicing factors involved in Fas alternative splicing

Abstract

Alternative splicing (AS) is an essential cellular process that greatly expands the coding capacity of eukaryotic genomes by generating multiple protein isoforms from a single primary transcript. The regulation of AS involves the recognition of *cis* regulatory elements, i.e. short RNA sequence motifs, by *trans* acting factors, i.e. RNA binding proteins. Aberrant splicing has been implicated in human disease including many aspects of cancer progression.

Fas is a cell surface receptor involved in apoptotic signaling. It can be alternatively spliced to produce either membrane bound pro-apoptotic form or a soluble anti-apoptotic form. Regulation of alternative splicing of the *Fas* pre-mRNA is mediated, amongst others, by T-cell intracellular antigen 1 (TIA-1) and splicing factor 45 (SPF45) proteins, which promote the formation of pro- and anti-apoptotic forms of Fas, respectively. TIA-1 binds to poly-pyrimidine tracts downstream of the 5' splice site (ss) and recruits the U1 snRNP complex to the 5'ss by interacting with the U1 snRNP specific protein U1C. The interaction of TIA-1 and U1C involves the RRM1 and Q-rich domains of TIA-1.

U2AF homology motifs (UHMs) are atypical RNA Recognition Motif (RRM) domains that mediate critical protein-protein interactions during the regulation of alternative pre-mRNA splicing and other processes. The recognition of UHM domains by UHM Ligand Motif (ULM) peptide sequences plays important roles during early steps of spliceosome assembly. SPF45 is an alternative splicing factor implicated in breast and lung cancer and splicing regulation of apoptosis-linked *Fas* pre-mRNA by SPF45 was shown to depend on interactions of its UHM domain with ULM motifs in constitutive splicing factors.

The aim of this thesis is to decipher the structural mechanisms for the function of TIA-1 in alternative splicing regulation, and its interaction with U1C using an approach of integrated structural biology. Further, cyclic peptide and small molecule inhibitors are developed to inhibit UHM-ULM interactions in splicing factors and thereby provide novel tools to modulate splicing and study early spliceosome assembly.

Chapter 1 of this thesis provides a biological background of the pre-mRNA splicing along with the role of TIA-1 and SPF45 proteins in *Fas* alternative splicing. Current state of

the art is presented for the role of different domains of these proteins in splicing regulation. Chapter 2 gives an overview of the integrated structural biology methods used for the study of these proteins. In Chapter 3, materials and methods used for the biochemical and structural analysis of these proteins is described. Chapter 4 describes the structural aspects of protein-RNA and protein-protein interactions mediated by the TIA-1 protein and their contribution to the activity of TIA-1 in splicing regulation. The NMR-derived solution structure of the RRM1 domain of TIA-1 is presented and a crystal structure of TIA-1 RRM1 bound to a peptide derived from the C-terminal region of the U1C protein is reported. In addition, RNA binding contributions by the three RNA recognition motif (RRM) domains of TIA-1 are studied. NMR and SAXS data of the three RRM domains of TIA-1 in the presence of *Fas* and poly-U RNAs show that the three RRM domains of TIA-1 tumble together in solution with the formation of a compact shape. Based on the results obtained a structural model for the recognition of intron RNA by the U1 snRNP and TIA-1 is provided that suggests how TIA-1 can aid this process.

In Chapter 5 and Chapter 6, the structure-based development of peptide and small molecule inhibitors that interfere with UHM-ULM interactions are reported. Cyclic peptides are developed by cyclizing the native ULM peptide sequence to obtain a specific inhibitor of the SPF45 UHM domain which is 4-fold more potent than the native ULM and discriminates between the UHM domains of constitutive and alternative splicing factors with 270-fold selectivity. In addition, a fluorescently labelled cyclic peptide was developed as a probe to screen ~42000 compounds using fluorescence polarization assay. This assay identified small molecules containing phenothiazine moiety as a general inhibitor of the UHM-ULM interaction. The small molecules discovered are further optimized by structure-based approaches wherein the structure of the small molecule was determined by X-ray crystallography in complex with the PUF60 UHM domain. Both, the cyclic peptide and the small molecule inhibitors modulate the pre-mRNA splicing of IgM and MINX pre-mRNAs and stalled the spliceosome assembly at complex A formation *in vitro* and thus provide novel molecular tools to study and modulate alternative splicing.

The results presented in this thesis provide novel structural insights into molecular mechanisms of splicing regulation by alternative splicing factors. The structural basis for interactions between TIA-1 and U1C represents one of the first examples demonstrating how a *trans*-activating splicing factor can interact with the core splicing machinery. The data allow to propose a model for the TIA-1 U1 snRNP interactions that shows how a combination of

protein-RNA and protein-protein interactions establishes a unique spatial arrangement of these factors.

The development of novel UHM inhibitors will be important for studying mechanisms of alternative splicing and for structural studies using stalled spliceosome complexes. These inhibitors are unique in that they can interfere with the early stages of spliceosome assembly, where so far no inhibitors have been reported and also are a first proof of principle that spliceosome assembly can be inhibited by targeting UHM-ULM interactions.

Chapter 1
Introduction I:
Biological background

1.1 Splicing and spliceosome assembly

1.1.1 Pre-mRNA splicing

Pre-mRNA splicing is the process of removing non-coding intervening sequences (introns) from the pre-mRNA to produce mature mRNA ((Berget et al. 1977; Chow et al. 1977) reviewed in (Black 2003)). The process itself is fundamental for the expression of most metazoan genes and occurs before nuclear export and translation of the mRNA.

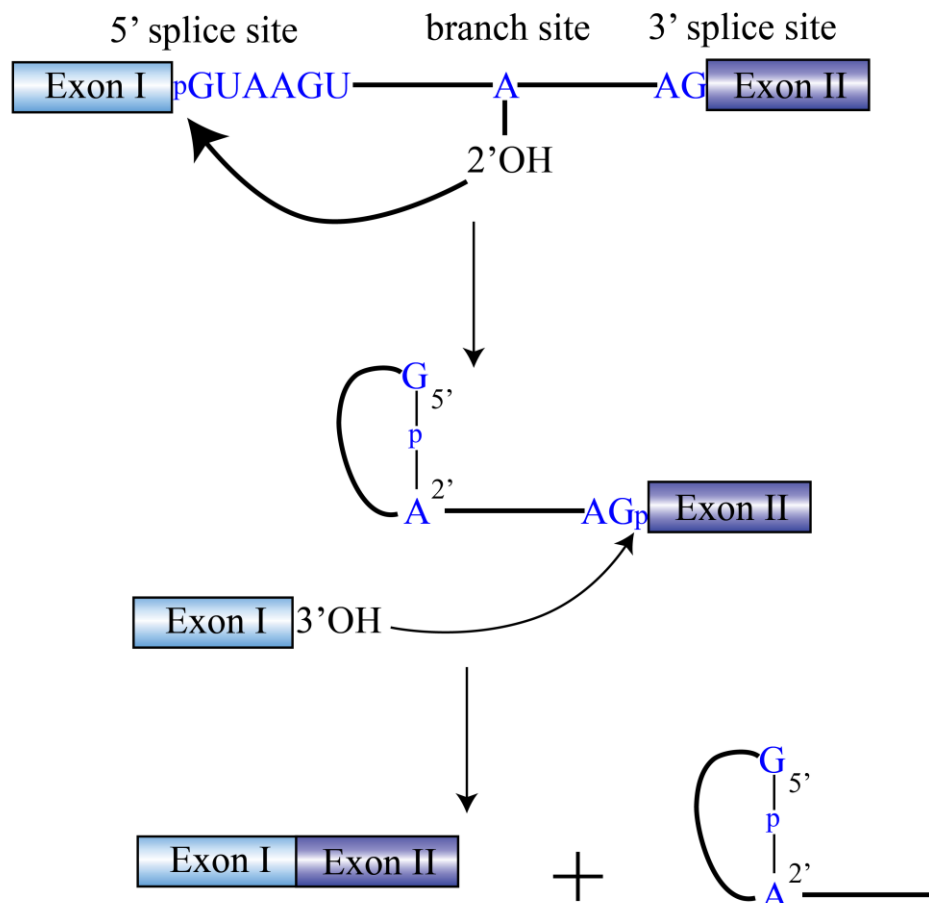


Figure 1. Schematic overview of pre-mRNA splicing

The chemistry of splicing reaction is depicted in this schematic. In the first step, the 2' OH of branch point adenosine attacks the 5' ss of exon I while in the second step, the 3' OH of exon I attacks the phosphodiester bond of exon II, thus forming the mature mRNA and the intron lariat.

The splicing reaction is chemically a simple two-step transesterification reaction occurring between RNA nucleotides (**Figure 1**). In the first step, the 2' hydroxyl of the branch point adenosine attacks the phosphodiester bond at the 5' splice site (5' ss) and displaces the 5' exon. In the second step, the 3' hydroxyl of the first exon attacks the phosphodiester bond

at the 3' splice site of the second exon, thus displacing the intron. Next, the two exons are ligated while the intron is displaced as an intron lariat.

Inside the cell, the process of splicing is carried out in two steps by a highly dynamic machinery called the spliceosome, consisting of ribonucleoprotein (RNP) complexes (Lerner et al. 1980). During the process, complex steps of assembly and disassembly of splicing factors at the splice site occur with large amount of ATP, the primary source of energy to drive the process, being hydrolyzed (Will and Luhrmann 2011).

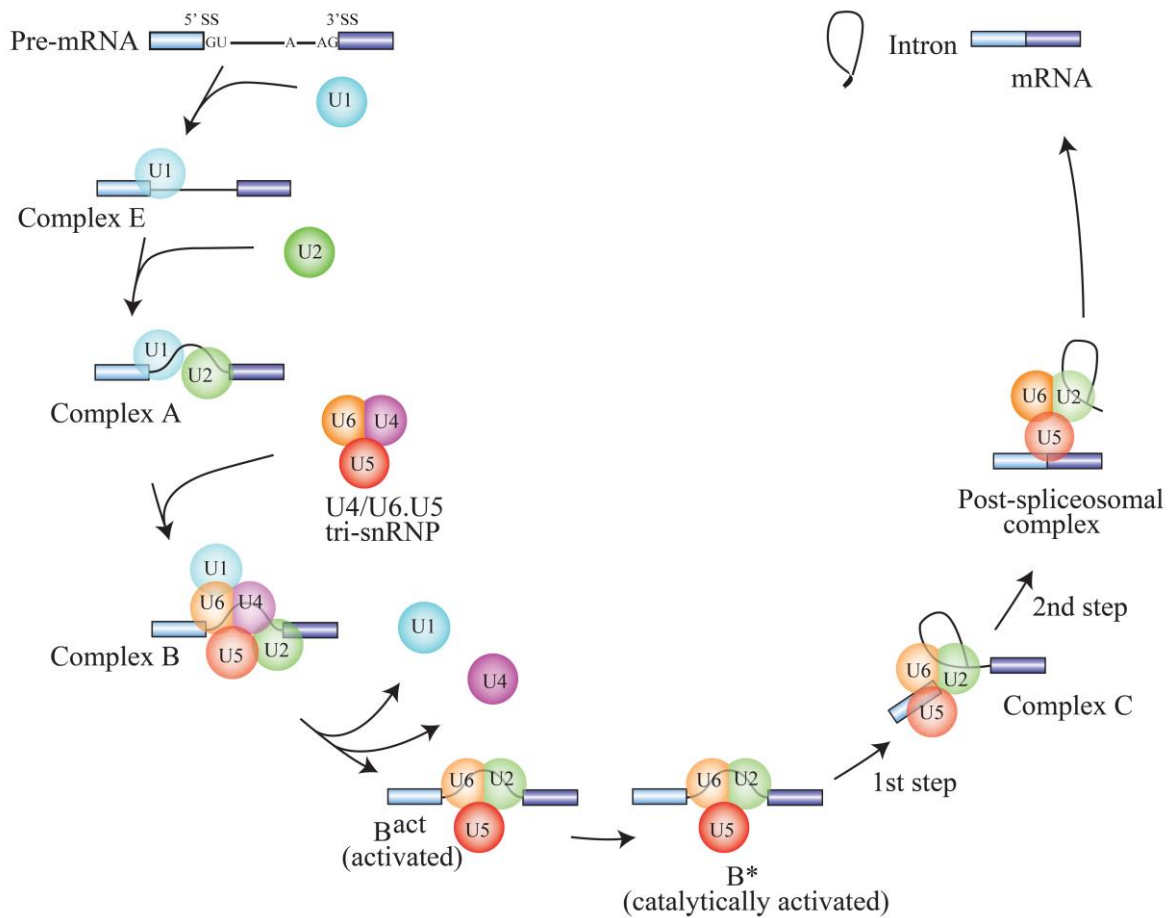


Figure 2. Spliceosome assembly and pre-mRNA splicing

Steps in the spliceosome assembly and pre-mRNA splicing are shown. (Adapted from (Will and Luhrmann 2011))

The spliceosome consists of five different RNP subunits in addition to various associated protein cofactors (Jurica and Moore 2003; Will and Luhrmann 2011). The spliceosome subunits are known as small nuclear ribonucleoproteins (snRNPs) to distinguish them from the ribonucleoprotein machinery involved in other cellular processes such as those

from ribosomal subunits. The components of the spliceosome assemble on the pre-mRNA during transcription wherein the RNA component of the snRNPs interact with the intron of the pre-mRNA.

During splicing reaction, the spliceosome forms different complexes as the reaction proceeds with a stepwise assembly of various snRNP particles on the pre-mRNA substrate (**Figure 2**). The early spliceosome complex also called complex E is formed when the U1 snRNP binds to the GU sequence at the 5'ss of the intron along with the binding of the splicing factor 1 (SF1) to the branch point sequence (Seraphin and Rosbash 1989; Jamison et al. 1992). At the 3'ss of the intron, U2AF1 in complex with U2AF2 binds to the polypyrimidine tract (Py tract). Complex E formation does not require ATP and is called commitment complex as formation of E complex commits the pre-mRNA for splicing (Legrain et al. 1988).

In the next step, the U2snRNP replaces SF1 in ATP dependent manner and binds to the branch point sequence. The so formed complex is called as complex A or pre-spliceosomal complex. Once complex A is formed, U4/U5/U6 tri-snRNPs can dock onto it to form complex B, which after undergoing several rearrangements (and formation of intermediate B^{act} and B* complexes) forms complex C. The catalytically activated B* complex catalyzes the first step of the splicing reaction followed by second reaction catalyzed by complex C.

Once the two exons are ligated, the spliced RNA is released from the complex and the lariat is degraded (Cheng and Menees 2011). The snRNPs are recycled for catalyzing the next round of splicing reaction.

1.1.2 Alternative splicing

Alternative splicing (AS) is the pre-mRNA splicing process where multiple mature mRNA transcripts can be produced from a single pre-mRNA by varying the exon composition. This process allows the cell to increase its repertoire of mRNA isoforms starting from the same pre-mRNA thus greatly expanding the protein coding capacity of the eukaryotic genome. AS of genes seems to play a crucial role in the organismal complexity with higher eukaryotic organisms showing higher percentage of genes undergoing AS. Therefore, it is not surprising that in humans, the most complex organism of all; around 95% of the genes are alternatively spliced (Pan et al. 2008).

The exons always included in the mature mRNA are called constitutive exons whereas the ones, which could either be skipped or included to produce different isoforms of the mRNA

are known as alternative exons. AS can be categorized into following seven major types (Roy et al. 2013): exon skipping; alternative 5'ss; alternative 3'ss; intron retention; mutually exclusive alternative exons; alternative promoter and first exon; and alternative poly A site and terminal exon. Amongst these, exon skipping is the most widely observed mode in the mammalian pre-mRNA splicing where the exon may be spliced out or retained as required (Sammeth et al. 2008).

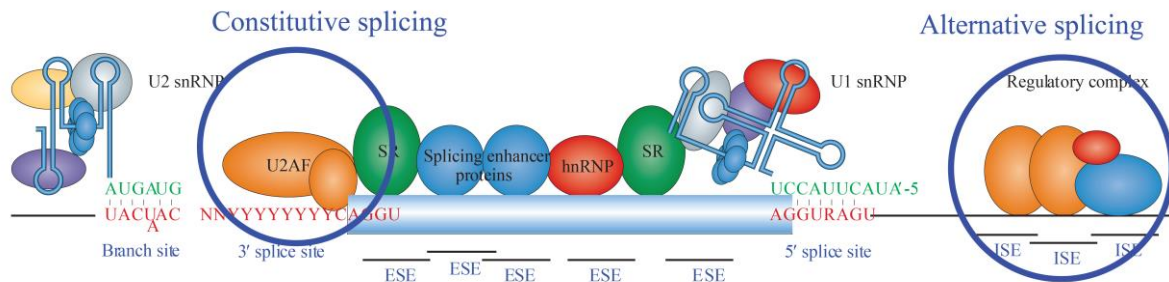


Figure 3. Splicing regulation

The *cis* RNA elements and *trans* protein factors involved in the splicing regulation are shown (adapted from (Wang and Cooper 2007)).

Selection of 5'ss by U1 snRNP during the early E complex formation is fundamental to the process of pre-mRNA splicing and dictates whether an exon will be included in the mature mRNA or not. Thousands of 5'ss are known which act as bona fide 5'ss in human transcriptome thereby increasing the complexity of the alternative splicing process. In humans more than 9000 sequence variants of the consensus 5'ss are known (Roca et al. 2012). Most of these splice sites are present interspersed throughout the intronic sequences and resemble closely to the authentic splice sites in terms of sequence similarity and length. Such splice sites are called as pseudo-5'ss. As the splicing process occurs with high fidelity with single nucleotide precision, this suggests that the sequence at the 5'ss cannot be the only determinant of the 5'ss selection.

Therefore, in addition to the splice site consensus sequences, AS is highly regulated by trans-acting proteins, which bind to the cis-acting elements on the pre-mRNA. The trans-acting proteins include activators and repressors whereas the cis-acting elements consists of silencers and enhancers, which could up or down regulate the splicing process (Matlin et al. 2005; Wang and Burge 2008). Based on the location of the sequence in the pre-mRNA where the trans-factors bind, cis-acting elements can be classified as exonic splicing enhancers (ESEs), intronic

splicing enhancers (ISEs), exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs).

Most of the trans-acting factors that regulate the AS are RNA binding proteins (RBPs) which bind to cis- regulatory elements and thus guide the spliceosome to the correct splice site. The RBPs bind to the RNA sequences in the cis-elements with varying degrees of sequence specificities and thus dictate the fate of the pre-mRNA (Chen and Manley 2009; Nilsen and Graveley 2010). The classical RBPs that are involved in the splicing regulation by binding to the cis-regulatory elements include serine/arginine-rich proteins (SR proteins) and heterogeneous ribonucleoproteins (hnRNPs). SR proteins, when bound to ESEs, tend to promote exon inclusion whereas hnRNPs promote exon exclusion when bound to ESSs and/or ISSs (**Figure 3**).

1.2 Regulation of Fas alternative splicing

Given the crucial role of RBPs in alternative splicing regulation, it is not surprising that their aberrant expression and regulation results in various diseases. One of the disease where the deregulation of splicing is widely observed is cancer. Cancer cells are known to evade apoptosis (Letai 2008), which occurs through activation of one of the several pathways present in normal cells. Many of the mRNA transcripts from apoptotic genes are known to be alternatively spliced thereby producing proteins of opposite functions which either promote or prevent apoptosis (Schwerk and Schulze-Osthoff 2005).

The Fas receptor is a death receptor present on the cell surface and is involved in apoptotic signaling. It can be spliced either as a single-pass transmembrane form that is a fully functional Fas receptor hence being pro-apoptotic or as a soluble protein that lacks the transmembrane region also called the anti-apoptotic form. There are eight splice variants of *Fas* mRNA known, producing seven different but related proteins. Amongst this, the Fas receptor is one of the two major isoforms encoded by the isoform 1, which has a transmembrane region encoded by exon 6 of the mRNA. The membrane bound Fas receptor binds to the Fas ligand and activates the caspase cascades. On the other hand, the soluble Fas receptor is secreted out of the cell and is known to induce autoimmune phenotypes in mice.

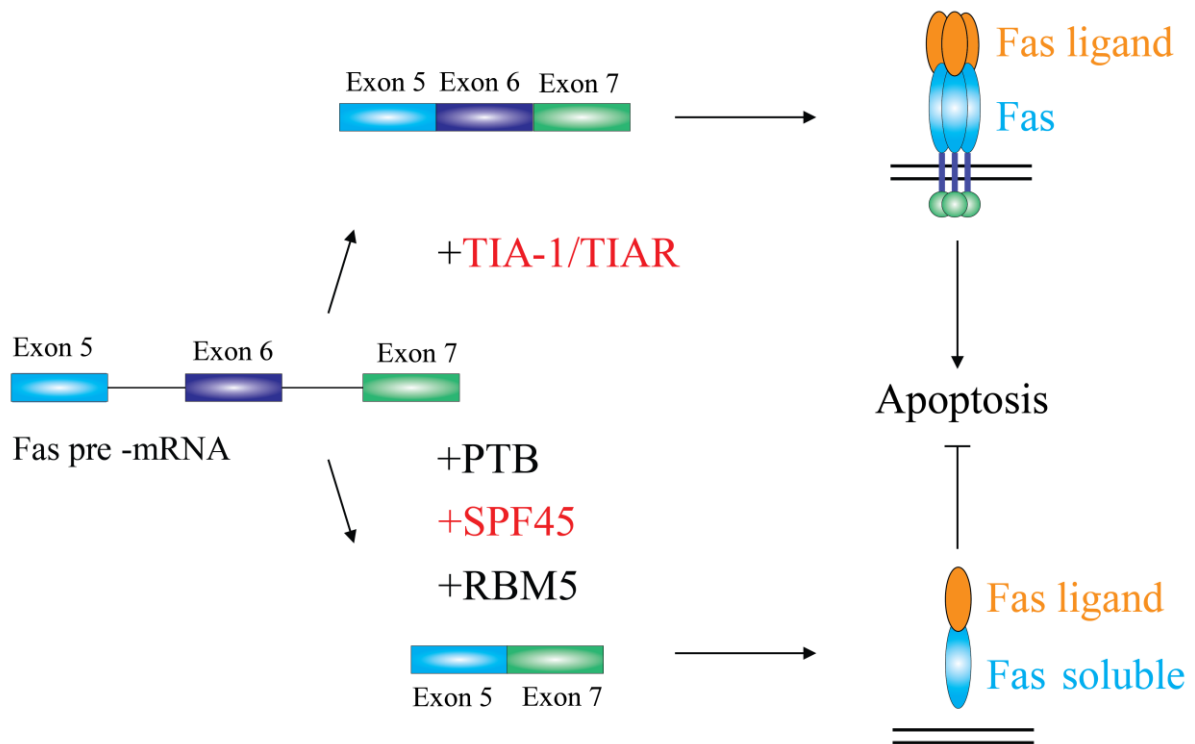


Figure 4. Schematic of Fas alternative splicing

Proteins promoting the two isoforms of *Fas* pre-mRNA are shown. The membrane bound form of Fas promotes apoptosis whereas the soluble Fas receptor inhibits apoptosis by titrating away Fas ligand.

The regulation of the alternative splicing of *Fas* pre-mRNA to produce membrane bound and soluble Fas receptor is mediated by TIA-1/TIA-R (Tian et al. 1995), PTB (Izquierdo et al. 2005), SPF45 (Corsini et al. 2007; Liu et al. 2013) and RBM5 (Bonnal et al. 2008) proteins (Figure 4). TIA-1 and PTB regulate the alternative splicing of *Fas* antagonistically. TIA-1 binds to an intronic polyU sequence to include exon 6 in the mRNA encoding Fas receptor whereas PTB binds to an exonic splicing silencer and promotes exon skipping (Izquierdo et al. 2005).

1.2.1 Role of TIA-1 and U1C proteins in Fas alternative splicing

T-cell intracellular antigen-1 (TIA-1) is a multi-domain RNA binding protein. It consists of three RRM domains (RRM1, RRM2 and RRM3) and a C-terminal Q-rich domain (Figure 5A) and is involved in the alternative splicing of many pre-mRNA transcripts including *Fas* (Forch et al. 2000; Zuccato et al. 2004; Singh et al. 2011). In addition to its role in alternative splicing, TIA-1 also mediates and suppresses mRNA translation under environmental stress by binding to the AU rich elements at the 3' untranslated region of the

TIA-1 along with its close homologue TIA-1 related protein TIA-R, are widely expressed in cells and mediate splicing of their own pre-mRNAs (Izquierdo and Valcarcel 2007b; Reyes et al. 2009). However, the two proteins have different expression patterns at cellular and tissue level (Dember et al. 1996; Forch and Valcarcel 2001; Izquierdo and Valcarcel 2007b). TIA-1 has two isoforms: TIA-1a and TIA-1b. The two isoforms differ in the linker between RRM1 and RRM2 domains where TIA-1a has an eleven amino acid insertion. Nevertheless, the two isoforms of TIA-1 show similar cellular distribution and RNA binding activity. In spite of the similar RNA binding activity, TIA-1b, which lacks the eleven amino acid insertion, displays enhanced splicing activity both *in vitro* and *in vivo* (Izquierdo and Valcarcel 2007b).

The three RRM domains of TIA-1 bind to RNA with different affinities, with RRM2 displaying highest affinity followed by RRM3 and then RRM1 displaying least affinity (Dember et al. 1996; Forch et al. 2002; Bauer et al. 2012; Cruz-Gallardo et al. 2013). RRM3 has recently been shown to bind AU rich RNA sequences in a pH-dependent manner (Cruz-Gallardo et al. 2013).

From the iCLIP experiments, TIA-1 binding sites on pre-mRNA transcripts have been mapped on to 10-28 nucleotides downstream of exon/intron boundaries (Wang et al. 2010). Therefore, corresponding region from the intron 5 of *Fas* pre-mRNA provides a good candidate for structural biology studies of TIA-1 RNA complex. It should however be noted that polyU stretches upstream of an exon are also involved in TIA-1 binding (Zuccato et al. 2004). Therefore, polyU stretches also provide suitable RNA candidates to study interaction of TIA-1 RRM domains with RNA.

The three RRM domains of TIA-1 are connected by 10-12 residue long flexible amino acid linkers. Currently there is no structural information available on how the three RRM domains of TIA-1 interact with the RNA at molecular level. As for the individual domains, only the crystal structure of RRM2 domain (Kumar et al. 2008; Kuwasako et al. 2008) and NMR structure of tandem RRM2, RRM3 domain is available (**Figure 6C**) (Wang et al. 2014). Both, RRM2 and RRM3 domains adopt canonical RRM fold. From the NMR structure, it was clear that in the absence of RNA, the two domains tumble independently of each other with no inter-domain contacts. In the absence of the structures of TIA-1 bound to RNA, the role of linkers in molecular recognition of the RNA and the conformation of TIA-1 when bound to the RNA sequences remains unknown.

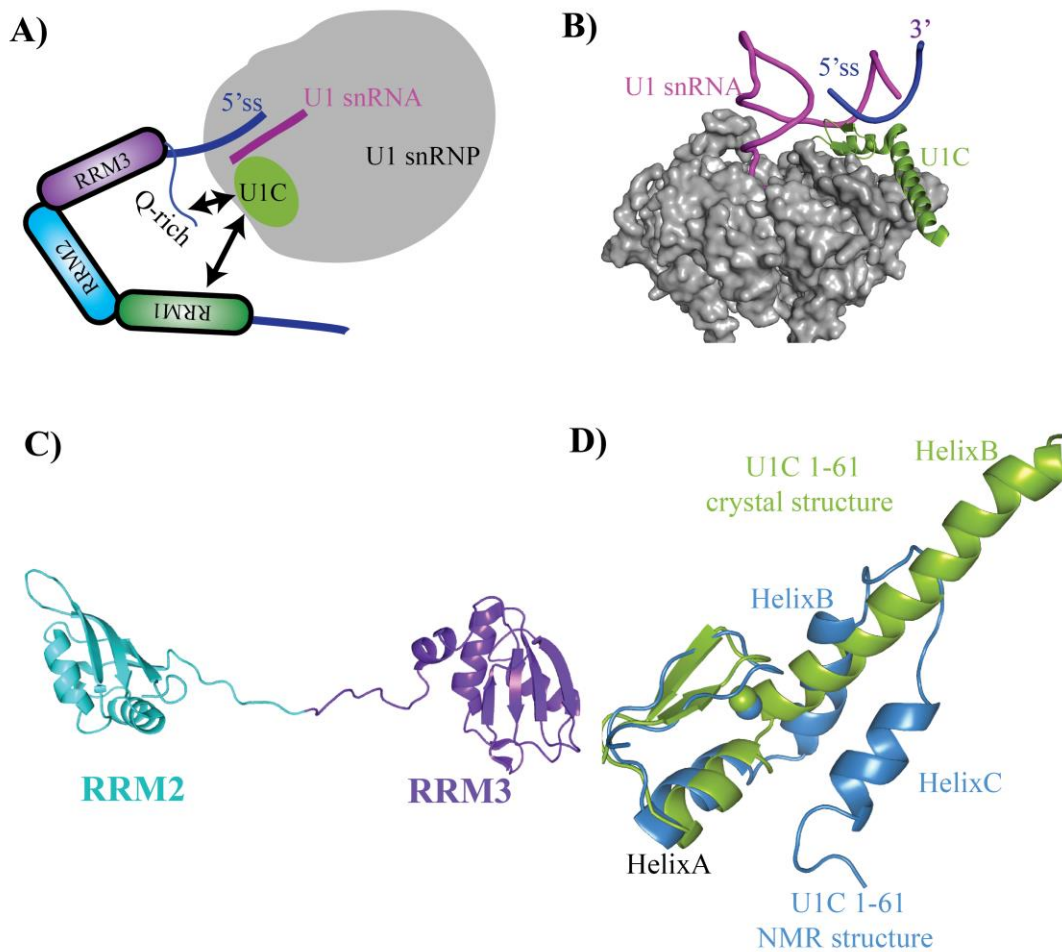


Figure 6. Available biochemical and structural information of TIA-1 and U1C proteins

A) Schematic of the role of TIA-1 in the recruitment of U1 snRNP at the 5'ss. RRM1 and Q-rich domains of TIA-1 interacts with the U1C protein that is a part of U1 snRNP complex. B) Crystal structure of U1 snRNP complex (PDB id: 4PJO). All the protein components of U1 snRNP except U1C are colored gray for clarity. U1C is shown in green. C) Available NMR structure of TIA-1 RRM2, 3 domains (PDB id: 2MJN). D) Structure comparison between NMR (PDB id: 2VRD) and crystal structure (PDB id: 2PJO) of U1C.

Although the role of individual RRM domains in isolation for RNA binding have been studied in detail, the details of how the three RRM domains contribute to RNA binding remains unknown. It was shown recently that *Fas* intron 5 exhibits two binding sites with comparable affinities when binding to TIA-1 and all the three RRM domains contribute to binding to a polyU 20 RNA (Bauer et al. 2012). However, the extent of contribution of each domain in context of full length TIA-1 and whether all the three RRM domains tumble together in solution when bound to RNA remains to be determined.

The U1 snRNP recruiting activity of TIA-1 at the 5'ss depends upon the interaction between its Q-rich domain and the U1 snRNP specific protein U1C (Forch et al. 2002) (**Figure 6A**). The RRM1 domain of TIA-1 augments the interaction between Q-rich

domain and U1C. The interaction between the Q-rich domain and U1C is independent of the presence of RNA. TIA-1 binds to IAS1, an activating sequence in the exon K-SAM of fibroblast growth factor receptor-2 (FGFR2), in U1 snRNP dependent manner (Del Gatto-Konczak et al. 2000).

The human U1 snRNP is composed of U1 snRNA, seven Sm proteins (SmB/SmB', SmD1, SmD2, SmD3, SmE, SmF and SmG) and three U1-specific proteins (U1-70K, U1-A and U1-C). The crystal structure of the human U1 snRNP has been determined previously which shows that the U1C protein interacts with the phosphate backbone of U1 snRNA-5' ss duplex but does not make any base specific interaction with the duplex (**Figure 6B**) (Kondo et al. 2015).

U1C is a 156 amino acid long protein with a molecular weight of 17 kDa. It consists of an N-terminal zinc finger region and a C-terminal region rich in proline and methionine sequences (Sillekens et al. 1988). The first 40 residues from the N-terminal zinc finger region are highly conserved from *Saccharomyces cerevisiae* to humans. However, the C-terminal low complexity region has diverged considerably in different organisms. This region is absent in the U1C orthologues from *S. cerevisiae* and *Schizosaccharomyces pombe* thus suggesting its function has only evolved in higher organisms. The zinc finger region of U1C contains three cysteine and three histidine residues amongst which Cys6, Cys9, His24 and His30 are required for its association with U1 snRNP (Nelissen et al. 1991). These residues also coordinate a Zinc ion and thus mutation of any of these residues is expected to destroy the zinc finger fold of the protein. The zinc finger region of the U1C has also been shown to be sufficient for its incorporation into U1 snRNP lacking the U1C protein (Nelissen et al. 1991).

U1C facilitates the association of U1 snRNP to the 5' ss which is reduced substantially by the deletion of 5' end of U1 snRNA (Heinrichs et al. 1990). It was suggested that U1C protein enables the base-pairing of the 5' end of U1 snRNA and 5' ss. This was confirmed when a high-resolution structure of the U1 snRNP complex was solved by X-ray crystallography. It showed the direct interaction of the zinc finger of U1C with the RNA double helix formed by the U1 snRNA and the 5' ss (Pomeranz Krummel et al. 2009; Kondo et al. 2015). Also the N-terminal zinc finger of U1C has been shown to be essential and sufficient for the formation of complex E in *in vitro* assays where reconstituted U1 snRNP complexes were tested for the formation of complex E and restoring of splicing activity (Will et al. 1996).

Currently there are two structures of U1C protein available: one is a solution structure determined by NMR of the first 61 residues (U1C 1-61), the other is the structure of U1C 1-61

in complex with the U1 snRNP complex. The residues 1-30 form the zinc finger and show acceptable superposition between the NMR and the crystal structure (RMSD 0.70 Å). However, both the structures differ significantly in the conformation of the C-terminal helix that is formed by residues 31-61. In the NMR structure, helix B folds back onto the helix C whereas in U1C bound to U1 snRNP structure, helix B and helix C form a long continuous helix (termed as helix B) (**Figure 6D**). Therefore, which structure amongst the two represents the true solution structure of the U1C protein or if the conformation of U1C changes upon binding to U1 snRNP remains to be seen.

1.2.2 Role of UHM-ULM interactions in Fas alternative splicing

Although the actual pre-mRNA splicing reaction is carried out by the RNA bases and the RNA core of the spliceosome has been highly conserved for >1 billion years, the role of various protein factors involved in the recognition of the splice site during the spliceosome assembly cannot be underestimated. During the pre-mRNA splicing and spliceosome assembly, three important types of interactions take place: RNA-RNA interactions between the 5' splice site and the U1 snRNA; protein-RNA interactions between various splicing factors and the U1 snRNA and cis-regulatory elements at the splice site; and protein-protein interactions between the various protein factors involved in splicing.

During complex E formation in the early stages of pre-mRNA splicing, the U2AF65-SF1 complex recognizes the consensus sequences on the pre-mRNA near the 3' splice site (Zamore et al. 1992; Berglund et al. 1997; Berglund et al. 1998). This protein-protein interaction is mediated by the C-terminal U2AF homology motif (UHM) domain of U2AF65 and the N-terminal of SF1 protein called as UHM ligand motif (ULM).

UHM domains are non-canonical RRM domains with $\beta\alpha\beta\beta\alpha\beta$ topology (Kielkopf et al. 2001; Selenko et al. 2003). Unlike RRM domains, UHM domains have degenerate RNP1 and RNP2 motifs; thus, they are unable to bind RNA. They contain aliphatic amino acids at the first position of RNP1 and second position of RNP2 motifs instead of the aromatic amino acids in the classical RRM domains. Besides this, they have an Arg-X-Phe amino acid sequence (where X can be any amino acid). This sequence is present in the loop connecting the α -helix B and the β -strand of the UHM domain. Also, the α -helix A has more acidic character than the canonical RRM domains (Kielkopf et al. 2004).

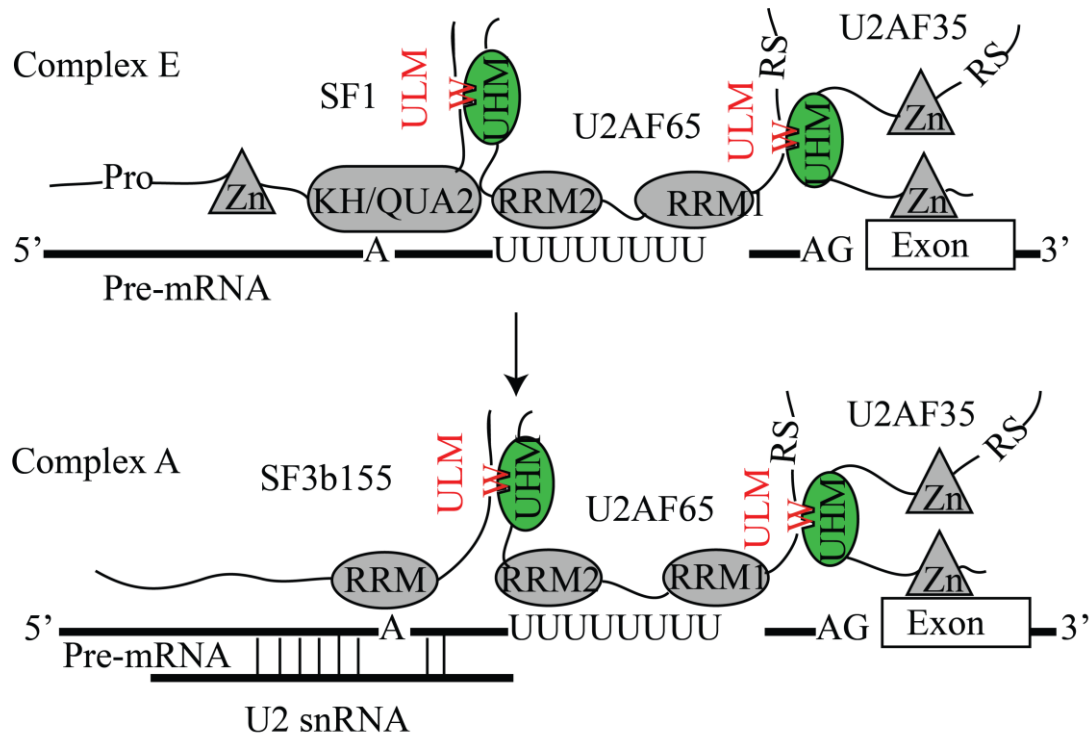


Figure 7. Early stages of spliceosome assembly

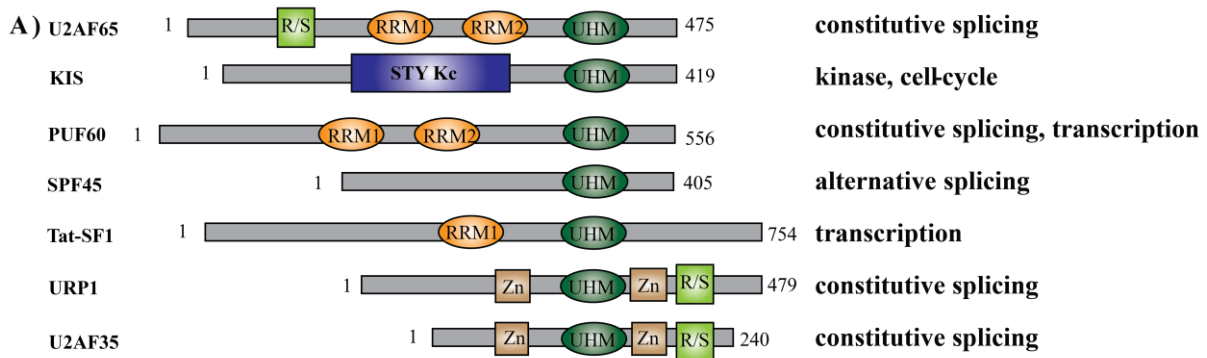
Schematic overview of UHM-ULM interactions during spliceosome assembly. UHM domains and ULM peptide motifs are shown in green and red colors respectively.

The UHM domains recognize the tryptophan containing ULM peptide sequences. The ULM peptide sequences consist of a highly conserved tryptophan residue flanked by basic and acidic residues. The tryptophan from the ULMs inserts into the hydrophobic pocket formed by the aliphatic amino acids of the RNP motifs and the Arg-X-Phe motif, whereas the acidic and basic amino acids flanking the tryptophan make charged interactions with the UHM domain.

UHM domains were first identified in both the subunits of U2AF heterodimer (U2AF65 and U2AF35). In the recent years, UHM-ULM interactions have been identified in several other proteins including SPF45, PUF60, KIS kinase, Caper- α /HCC1 which mediate diverse biological functions (Maucuer et al. 1997; Kielkopf et al. 2004; Corsini et al. 2007; Manceau et al. 2008; Corsini et al. 2009; Loerch et al. 2014) However, the role of UHM-ULM interactions in these proteins has not been completely understood.

Although, the various identified UHM domains share little sequence identity, the mode of recognition of the ULM peptides by these domains remains highly conserved (Kielkopf et al. 2004). One such UHM domain containing protein, splicing factor 45 (SPF45) was first

identified as a component of the spliceosome by mass spectroscopy (Neubauer et al. 1998). It has been shown to activate the cryptic 3' splice site in β -thalassemia (Lallena et al. 2002). Besides, it regulates the splicing of sex lethal (Sxl) protein in *Drosophila melanogaster* (Chaouki and Salz 2006). Sxl is known to be present exclusively in female flies where it regulates the splicing of exon3 of its own pre-mRNA. SPF45 binds to Sxl and inhibits the ligation of exon3 thus producing the functional Sxl protein (Bell et al. 1991; Lallena et al. 2002).



B)

SF3b155	ULM1 (192-204)	: PPSKRKR RWDQ TA
SF3b155	ULM2 (210-222)	: ATPKLLSS WDQ AE
SF3b155	ULM3 (224-236)	: PGHTPSL RWDE TP
SF3b155	ULM4 (285-297)	: TSSARKN RWDE TP
SF3b155	ULM5 (330-342)	: GASKRKS RWDE TP
U2AF65	ULM (84-96)	: KKKKVRKY WDV PP
SF1	ULM (14-26)	: SKKRKRS RWNQ DT

Figure 8. UHM domains in various proteins

A) Domain organization of various UHM domain containing proteins. B) ULM sequence alignment of various ULMs.

SPF45 consists of an N-terminal unstructured region, a G-patch motif (Aravind and Koonin 1999) and the C-terminal UHM domain. The G-patch region is a \sim 40 residue long motif and is predicted to adopt a α -helical conformation. It has seven highly conserved glycines and it mediates protein-protein (Silverman et al. 2004) and protein-nucleic interactions (Svec et al. 2004; Frenal et al. 2006).

In addition to regulating the splicing of β -thalassemia and Sxl pre-mRNA, SPF45 has also been shown to regulate the alternative splicing of *Fas* pre-mRNA (Corsini et al. 2007). It induces skipping of exon 6 in *Fas* pre-mRNA thereby producing an inactive Fas receptor. This

activity of SPF45 depends on its UHM domain, which binds to ULM sequences from splicing factors SF1, SF3b155 and U2AF65. This UHM-ULM interaction between the SPF45 UHM domain and ULMs from the constitutive splicing factors established a general role for these interactions in alternative splicing (Corsini et al. 2007).

SPF45 has been shown to be overexpressed in many cancers including breast, lung, colon and ovarian cancers (Sampath et al. 2003). It also confers broad multi-drug resistance against anticancer drugs (Sampath et al. 2003; Perry et al. 2005). The switch in expression of pro- and anti-apoptotic isoforms of Fas is tightly regulated (Izquierdo et al. 2005) (**Figure 4**). Therefore, it is not surprising that an imbalance in the Fas isoforms by overexpression of SPF45 could provide a means for tumor cells to escape apoptosis.

Besides SPF45, many other UHM domain harboring proteins have been associated with various diseases in humans (**Table 1**). This is not surprising given the important roles these proteins play in the cell. Many of these proteins are not only involved in constitutive and alternative splicing, but recently some have also been found to be involved in regulating transcription, cell signaling and cell cycle. However, except in the case of SPF45, it is not clear whether the UHM domains or the other regions of these proteins are responsible for their respective roles in the human diseases.

Table 1. Diseases associated with UHM domains

UHM proteins	Relevance to human diseases
SPF45	Overexpressed in many human cancers
KIS	Neurological tumors
PUF60	Xeroderma pigmentosum
URP	Developmental defects
TAT-SF1	Involved in HIV-1 pathogenesis
MAN1	Vascular diseases and cancers
HCC1	Nuclear autoantigen

1.2.3 Targeting spliceosome assembly with inhibitors

Given the importance of pre-mRNA splicing in producing functional protein isoforms to regulate various cellular functions that are many times antagonistic in nature, maintaining high fidelity in the process is indispensable. Hence, it is not surprising that ~15% of the

inherited human diseases are caused by point mutations affecting the pre-mRNA elements involved in splice site recognition (Krawczak et al. 2007; Lim et al. 2011).

In addition to the inherited mutations, misregulation of alternative splicing is known to contribute to many aspects of cancer progression such as programmed cell death, cancer cell metabolism, cell proliferation, angiogenesis and metastasis (David and Manley 2010; Bonnal et al. 2012; Kaida et al. 2012). Therefore, spliceosome presents a novel target for antitumor drugs.

Small molecules provide a powerful tool for studying complex biological processes. Small molecules have been used to study cellular transcriptional and translation machinery previously. However, mostly antibiotics were used to study translation in prokaryotes. In the recent years, several natural compounds have been discovered which target splicing.

These compounds include Spliceostatin A, FR901464 (Kaida et al. 2007), Plandienolide B (Kotake et al. 2007), Herboxidiene (Hasegawa et al. 2011), Sudamycin (Fan et al. 2011) and Isoginkgetin (Fan et al. 2011). All these molecules target the SF3b subunit of U2 snRNP except Isoginkgetin whose target is unknown. Besides these natural compounds, several other attempts have been made to identify small molecules to target splicing. However most of these small molecules target splicing either only *in vitro* such as Flunarizine, Chlorhexidine and Clotrimazol (Younis et al. 2010) or *in vivo* such as Napthoquinine and Tetrocarcin (Effenberger et al. 2013). In addition, the molecular targets of most of these small molecules remain to be determined.

All the inhibitors mentioned above inhibit the later stages of spliceosome assembly i.e. during or after complex A to complex B transition. Interestingly, most of the splicing regulation takes place during the early stages of spliceosome assembly i.e. during complex E and complex A formation. Therefore, it is highly desirable to develop inhibitors which could inhibit the early stages of spliceosome assembly and hence the pre-mRNA splicing.

As mentioned before, UHM-ULM interactions play a crucial role in constitutive and alternative splicing. Besides, these interactions are important during complex E and complex A formation i.e. during early stages of spliceosome assembly. UHM domains are also structurally very well characterized. Therefore, UHM-ULM interactions present a tempting target against which inhibitors could be developed. If successful, these inhibitors could not

only be used as precursors to develop lead candidates for drugs to inhibit splicing but also for studying spliceosome assembly during early stages of spliceosome assembly.

**Chapter 2 Introduction II:
Techniques used for integrated structural biology**

Structural biology is the study of macromolecular structures such as proteins and nucleic acids in order to obtain insights into the functioning of biological systems. For long, the focus of the field had been to determine the atomic level structures of these macromolecules primarily using X-ray crystallography. However, in recent years, this notion has changed owing to the understanding that biological macromolecules are dynamic in nature at different resolution range and both *in vitro* and *in vivo*. In addition, multi-domain proteins connected by long flexible linkers present a significant challenge to study by a single structural biology technique. The structural information obtained from the individual domains of these protein remains incomplete in the absence of structural information about the inter-domain interactions in the full length protein.

Therefore, the need to understand the atomic details of the biological macromolecules along with their dynamics at different time scales and spatial resolution necessitates combining a wide range of structural biology methods such as X-ray crystallography, NMR, electron microscopy, small angle X-ray and neutron scattering, mass spectroscopy and advanced light microscopy techniques. An integrated structural biology approach which utilizes various structural biology techniques to obtain complimentary structural information provides a way forward to study biological macromolecules. In this thesis, with two specific aims at hand: to study the multi-domain TIA-1 protein and understand the protein-protein and protein-RNA interactions mediated by it and to design small molecule inhibitors of UHM domains to target spliceosome assembly, the following structural biology techniques were used.

2.1 NMR spectroscopy

After protein crystallography, Nuclear Magnetic Resonance spectroscopy is the principle technique used to determine the atomic structure of proteins and nucleic acids. It provides a powerful tool not only to study the structure of the biomolecules, but also to understand the dynamics and study biomolecule-ligand interaction in the solution amongst many possible uses of the NMR spectroscopy. Use of NMR to study protein structures is a relatively new field and is often limited by the size of the proteins. However, recent advances in the NMR methods and the modern isotope enrichment schemes including selective labelling of amino acids and deuteration of the proteins have made it possible to study significantly larger macromolecules by NMR.

2.1.1 Principle of NMR spectroscopy

NMR spectroscopy relies on the principle that atomic nuclei with odd mass have a property called as spin whereas the nuclei with even mass may or may not have this property. The rotation of these nuclei around a given axis is characterized by spin angular momentum I . As the nuclei are charged particles, the rotation of the nuclei in a magnetic field creates a magnetic dipole which corresponds to magnetic moment μ . As shown in the following equation, the magnetic moment of the nuclei is directly proportional to the spin angular momentum with a proportionality constant which is also called as gyromagnetic ratio (γ)

$$\mu = \gamma I \quad \text{Eq. 1}$$

2.1.2 Larmor precession

In the absence of any magnetic field, the magnetic moment of the nuclei is expected to be randomly oriented. However, in the presence of an external magnetic field B_0 , the magnetic moment μ does not simply aligns along the B_0 field but precesses around it with the μ vector tracing a cone around B_0 . This is analogous to the precession of a gyroscope under the influence of the earth's magnetic field. The motion is called Larmor precession and is depicted by Larmor frequency ν_0 . The Larmor frequency is given by:

$$\nu_0 = \frac{|\gamma| B_0}{2\pi} \quad \text{Eq. 2}$$

Therefore, the Larmor frequency is directly dependent on the strength of the externally applied magnetic field. The higher the external field, the more the precession frequency. When another weak field B_1 is applied perpendicular to B_0 , B_1 will exert a torque on the magnetic moment of the nuclei to change its precession angle θ around the B_0 . Thus the resulting motion of μ can be described as caused by the resultant field B_0+B_1 . If B_1 is static, the θ would increase and decrease with the precession of μ . However, if B_1 is rotating with the same frequency as that of μ , the relative orientation of μ with respect to B_1 would stay constant. Therefore, if B_1 is perpendicular to B_0 and μ , then the torque exerted by B_1 on μ would be away from B_0 .

2.1.3 Vector formalism

The above rotation of μ away from the B_0 field can be formalized in a vector form. In the vector formalism, the bulk magnetization can be represented as a vector quantity. In the presence of B_0 , the bulk magnetization experiences a torque. This can be written in the mathematical form as:

$$\frac{dM(t)}{dt} = M(t) \times \gamma B(t) \quad \text{Eq. 3}$$

where $M(t)$ is the bulk magnetization, $B(t)$ is the magnetic field strength and γ is the gyromagnetic ratio of the nuclei.

As stated before, the magnetization would rotate away from B_0 under the influence of B_1 if B_1 is perpendicular to both B_0 and the magnetization vector. If M is aligned to the z-axis and parallel to B_0 , a short radio pulse applied along the x-axis will turn the magnetization vector towards $-y$ -axis. The angle of rotation θ will depend on the length of the applied pulse. The direction of the rotation is determined by the right hand rule, known from the physics of electromagnetism. The magnetization will start precessing around the z-axis or the external magnetic field B_0 with Larmor frequency ω_0 , generating the signal in NMR detection coil.

When the transverse radio pulse applied along the x-axis is switched off, the bulk magnetization will return to the ground state due to relaxation effects including loss of spin alignment (transverse relaxation T_2) and return of the system to the thermodynamic equilibrium state (longitudinal relaxation T_1).

2.1.4 Product operator formalism.

Vector formalism is able to explain simple NMR experiments performed on isolated spins. However, in order to explain more complex phenomena in NMR, product operators were introduced. Product operators provide a complete quantum mechanical description of the NMR experiments and their expected outcome.

In product operator formalism, the components of the spin angular momentum I along x, y, and z axis are represented as I_x , I_y and I_z respectively. The entire set of spins are described by a wave function $\psi(t)$ or density operator $\sigma(t)$. If we neglect the relaxation of the spins, the evolution of density operator with time is given by the Liouville-von Neumann equation as:

$$\frac{d\sigma(t)}{dt} = -i[H(t), \sigma(t)] \quad \text{Eq. 4}$$

where $H(t)$ is the Hamiltonian operator. The density operator for a single spin 1/2 can be described in Cartesian coordinate system as a sum of three product operators:

$$\sigma(t) = a(t)I_x + b(t)I_y + c(t)I_z \quad \text{Eq. 5}$$

At equilibrium, the x and y components are zero and the density operator is proportional to I_z . During the NMR experiments I_z evolves sequentially with time. For example, the evolution of I_z with a 90° pulse can be described as:

$$I_z \xrightarrow{90^\circ I_x} -I_y \quad I_z \xrightarrow{90^\circ I_y} I_x \quad I_z \xrightarrow{90^\circ I_z} I_z \quad \text{Eq. 6}$$

The evolution which can be perceived as a rotation along an axis and corresponds well with the vector model can be calculated for any degree of rotation.

The difference in the energy states of two spin states is dependent on the externally applied magnetic field and the local magnetic field. Therefore, the resonance frequency for each nuclei is different as the local magnetic field experienced by each nuclei is different. This difference in the local magnetic field experienced by each nucleus will manifest itself as different spin resonance frequency and is called the chemical shift of the nuclei.

The chemical shift evolves with the offset Ω which is the difference between a signal and a reference value, during the time t of precession.

$$\begin{aligned} I_x &\xrightarrow{\Omega t I_z} I_x \cos \Omega t + I_y \sin \Omega t \\ I_y &\xrightarrow{\Omega t I_z} I_y \cos \Omega t - I_x \sin \Omega t \\ I_z &\xrightarrow{\Omega t I_z} I_z \end{aligned} \quad \text{Eq. 7}$$

The product operator approach is not only useful for the uncoupled spins but also for the coupled spin systems. To explain the evolution of the spins with J-coupling, a second spin S is introduced which is described by product operators S_x , S_y and S_z . Due to the J-coupling the states of I and S spins will mix. The result is a product operator for two spins $2IS$ (the factor

2 is needed for normalization purposes). The operators for two spins evolve under offsets and pulses the same way as operators for a single spin. The rotations, however, have to be applied separately for each spin and the spins do not affect each other. Operators I_x , I_y , S_x and S_y evolve under coupling, whereas I_z and S_z do not.

2.1.5 NMR experiments for protein assignment

The basic experiment used to assess the quality of the protein is the one-dimensional hydrogen spectrum which gives an idea whether a protein is folded and well behaved in the NMR buffer. The 1D spectrum is unique for each protein but is too complex to analyse as most of the signals overlap with one another. 1D proton spectrum for a well folded protein shows high dispersion of peaks from around -0.5 to 12 ppm.

Once the protein looks folded and well behaved from the 1D spectrum, a two dimensional spectrum could be acquired which is useful for assignment purpose. If the correlation is recorded between the nuclei of same isotope, then it is called homonuclear spectrum such as ^1H - ^1H correlation spectrum or else it is called heteronuclear spectrum such as ^1H - ^{15}N or ^1H - ^{13}C spectrum. The ^1H and ^{15}N dimensions for a protein could be acquired by recording a ^1H - ^{15}N HSQC (Heteronuclear Single Quantum Coherence) spectrum where the amide protons are correlated with the amide nitrogens. This spectrum, besides containing the correlations for the protein backbone amides, also shows peaks for the Asn and Gln side chain residues and the aromatic HNe protons for Trp and His residues of the protein. The spectrum is often called as the fingerprint spectrum of a protein as it is unique for a given protein and reflects the folding state of the protein.

As an individual peak in the ^1H - ^{15}N HSQC spectrum shows the backbone amides of one amino acid in the protein and is highly sensitive to the changes in the local environment (pH, temperature, changes in protein structure etc.), this can be used to study protein-ligand interactions. As the ligand binds, the peak positions and intensities near the binding site of the ligand are expected to change as the local chemical environment of the protein changes. Such ^1H - ^{15}N HSQC titrations could be used to map the binding site of the ligand and determine the affinity of the ligand to the protein.

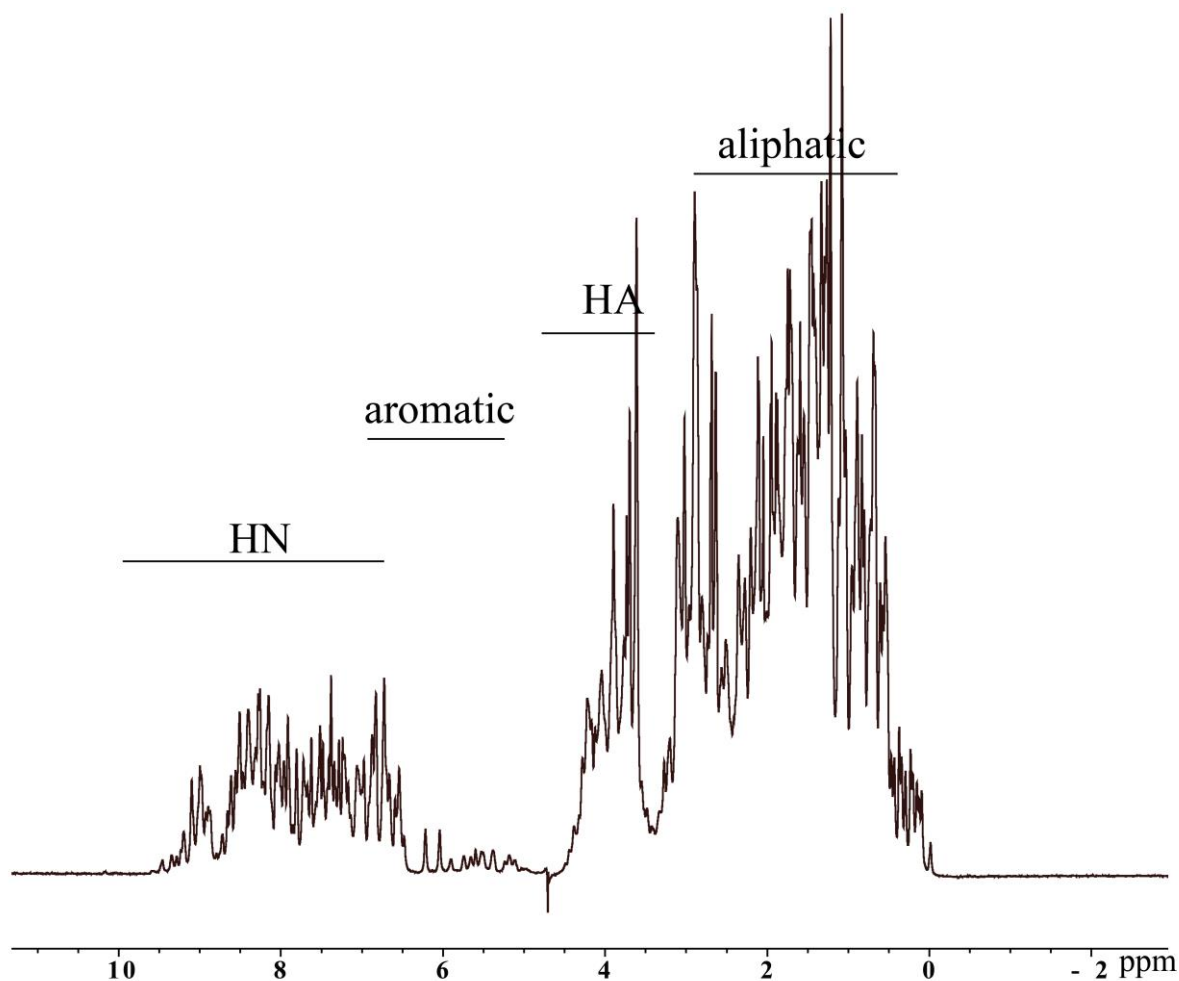


Figure 9 ^1H spectrum of TIA-1 RRM1 domain

Information content in the different regions of one dimensional proton spectrum is shown.

Large proteins show significant overlap in the 7-8 ppm region of the ^1H - ^{15}N HSQC spectrum and also high T_2 relaxation rates (transverse relaxation) due to the presence of more hydrogen atoms in the protein as the molecular size of the protein increases. This could be overcome by deuteration of the protein (Gardner and Kay 1998). Also TROSY (Transverse Relaxation Optimized SpectroscopY) (Pervushin et al. 1997; Salzman et al. 1998) spectrum could be recorded for such proteins which gives same correlation as that of ^1H - ^{15}N HSQC experiment but it reduces the relaxation effects such that better line shapes are obtained. TROSY experiment thus extends the protein size limitation which could be studied by NMR (Fernandez and Wider 2003).

To assign the correlations observed in the ^1H - ^{15}N HSQC spectrum to the primary sequence of the protein a sequential chemical shift assignments of the backbone residues are

obtained based on triple resonance experiments. These experiments include HNCA, HNCACB and HN(CO)CACB triple resonance experiments (Shan et al. 1996; Sattler et al. 1999b). The experiments give correlations between the backbone amides of protein to the side-chain $C\alpha$ and $C\beta$ carbon atoms of the self and the previous amino acids in the protein sequence. The self $C\alpha$ and $C\beta$ chemical shifts of one peak in the ^1H - ^{15}N HSQC spectrum can thus be matched and connected to peaks corresponding to $C\alpha$ and $C\beta$ chemical shifts of the previous residue thus building the sequential connectivity of the amino acids in a protein (**Figure 10**).

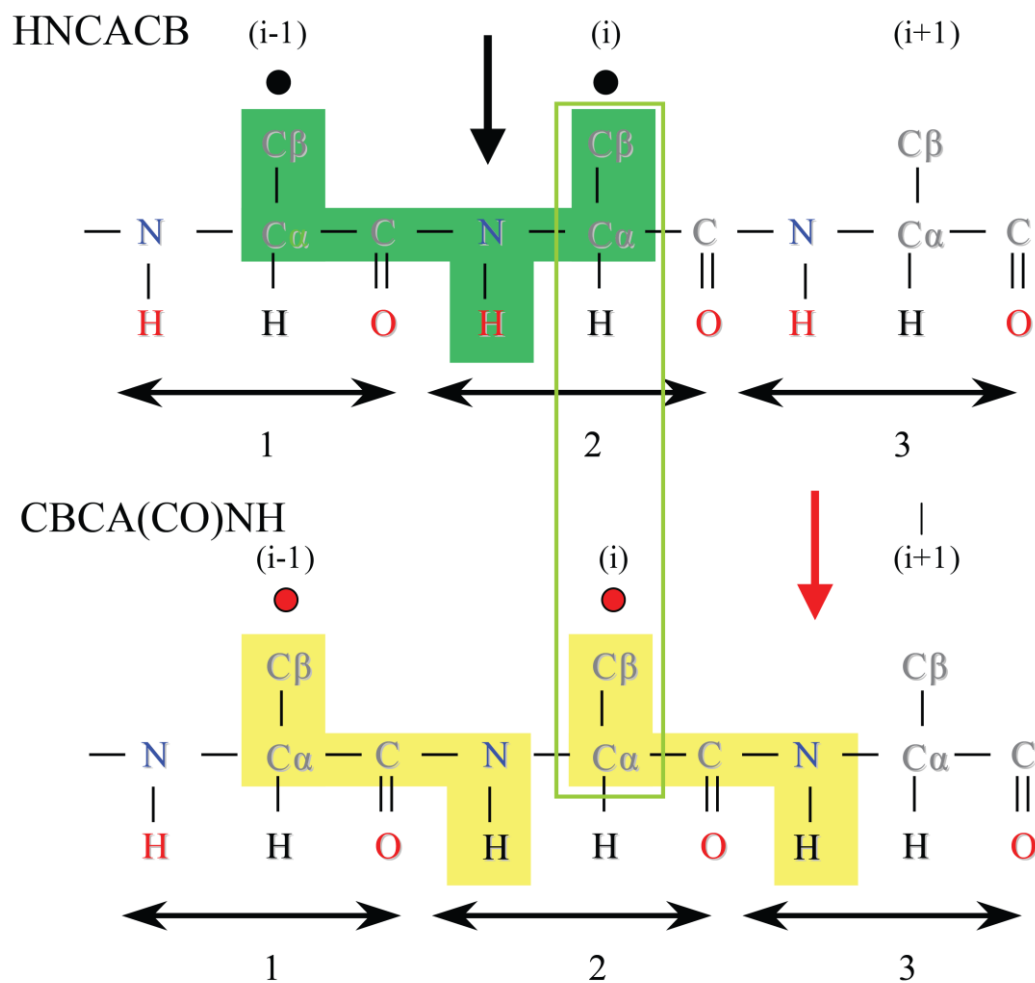


Figure 10 Schematic for protein backbone assignment

Schematic overview of the assignment of protein backbone using HNCACB and CBCA(CO)NH spectrum is shown. These experiments help to link the neighboring amides thus forming a linear chain of amides that can then be assigned to a specific fragment in the protein sequence.

The HNCA, HNCACB and HN(CO)CACB experiments in principle provide unambiguous correlations for the $C\alpha$ and $C\beta$ backbone carbon frequencies with the HN

resonances of the same and previous amino acid. As the $C\alpha$ and $C\beta$ chemical shifts of the amino acids are related to the identity of the side chains, different amino acids can be easily distinguished and, in the context of the neighbouring residues, can be used to unambiguously sequentially assign the $C\alpha$, $C\beta$, amide N and HN chemical shifts to the corresponding residue in the protein sequence.

Preliminary information about the secondary structure of the protein can be extracted from the backbone assignments of the protein. The $^{13}C\alpha$ and $^{13}C\beta$ secondary chemical shifts are sensitive indicators for the secondary structure elements in the protein i.e. α -helix, β -sheet and the loops (Spera and Bax 1991). For this the random coil shifts for each amino acid are subtracted from the actual chemical shift. Several random coil data sets based on either a database (Wishart et al. 1992) or model peptides under a variety of experimental conditions (Wishart et al. 1995) have been published. Observed secondary chemical shifts in the structural parts of a folded protein will differ significantly from the random coil chemical shifts: positively in the α -helical regions and negatively in the regions of β -sheets.

After the backbone assignments, the side-chain assignments are performed using HcCH-TOCSY and hCCH-TOCSY experiments which correlate all the side-chain carbons and hydrogens and CCCC(O)NH experiments which correlate the backbone amide with the side chain carbon atoms. These assignments can then be used for further structure calculations.

2.1.6 Structure calculations using NMR assignments

Structure calculation by NMR utilizes simulated folding of the biomolecule using the structural restraints obtained by NMR experiments. The backbone and side chain assignments obtained are used to assign the NOE (Nuclear Overhauser Effect) spectra, peak volume of which gives the distance restraints required for the calculation of the NMR structure.

The first structure using NOE-derived interatomic distances and scalar coupling constants was calculated for protease inhibitor IIA (Williamson et al. 1985). Dihedral angle restraints from the backbone (Φ and Ψ) and sometimes from side chains (χ_1 and χ_2) are also used for the structure calculation which are usually predicted by the bioinformatics programs, such as TALOS+ (Shen et al. 2009). Additional restraints obtained from residual dipolar couplings (RDCs) and paramagnetic relaxation enhancement (PRE) measurements can also be used to determine the relative position of structural elements within the molecule.

Besides using the experimental restraints, restraints derived from the proper geometry of the molecule, like bond length, chirality or planarity of the aromatic rings and peptide units are used during structure calculations. A simulated annealing protocol is used to carry out structure calculation where the system is virtually heated and then slowly cooled down. The program used for the structure calculation tries then to find coordinates for each atom that would best satisfy the given restraints. The structure calculation protocol is repeated several times to determine an ensemble of lowest energy structures which are consistent with the NMR input data. The quality of lowest energy ensemble is checked by determining how well the calculated structure fulfils the experimental data and how many restraints are violated by the calculated structure. The stereochemical quality of the structure is usually judged by quantifying the distributions of backbone and side chain dihedral angles, the number of van der Waals steric clashes etc. using NMR software programs like iCING (Doreleijers et al. 2012a).

2.1.7 Protein dynamics by NMR

NMR can also be used to study the protein dynamics occurring at the atomic level and ranging from picosecond-nanosecond to milliseconds-seconds time scale. Since in this thesis only ps-ns timescale dynamics is studied for the proteins under consideration, only these experiments are briefly discussed here.

Application of the radio frequency pulse moves the spins away from their thermal equilibrium. Relaxation refers to the phenomenon where the spins come back to their original thermal equilibrium state. T_1 (longitudinal relaxation along z-magnetization) and T_2 (transverse relaxation along x,y magnetization) represents the time constants for the spins to return to the thermal equilibrium state. The thermal equilibrium in the spins is usually induced by local fluctuating magnetic fields that are caused by tumbling of a molecule in solution based on the following internal interactions:

- 1) Dipole-dipole couplings between spins,
- 2) Different orientations of the molecules in the solution leading to different shielding (chemical shift anisotropy).
- 3) Electric quadrupolar interactions of the nucleus with the non-constant electric field produced by the electrons.

The longitudinal T_1 relaxation rate, also called as spin-lattice relaxation is induced by the interaction of the protein spins with the surrounding lattice. The lattice is assumed to be in

thermal equilibrium and have infinite heat capacity. Random Brownian motion causes the local fluctuations in the magnetic fields thus inducing the transition between spin states. This causes the recovery of the z-component of the magnetization to the equilibrium state. The recovery or decay is described by the time constant T_1 or the relaxation rate $R_1=1/T_1$.

The relaxation rates depend on the spectral density function, which is the Fourier transform of the autocorrelation function of the fluctuating magnetic field.

$$J(\omega) = \frac{2}{5} \left[\frac{\tau_c}{1 + \omega^2 \tau_c^2} \right] \quad \text{Eq. 8}$$

It can be shown that for dipolar relaxation, the T_1 relaxation rate is proportional to the square of the dipole field strength times the spectral density of the field fluctuation at frequency ω_0 . The spectral density has the appearance as shown in **Figure 11**:

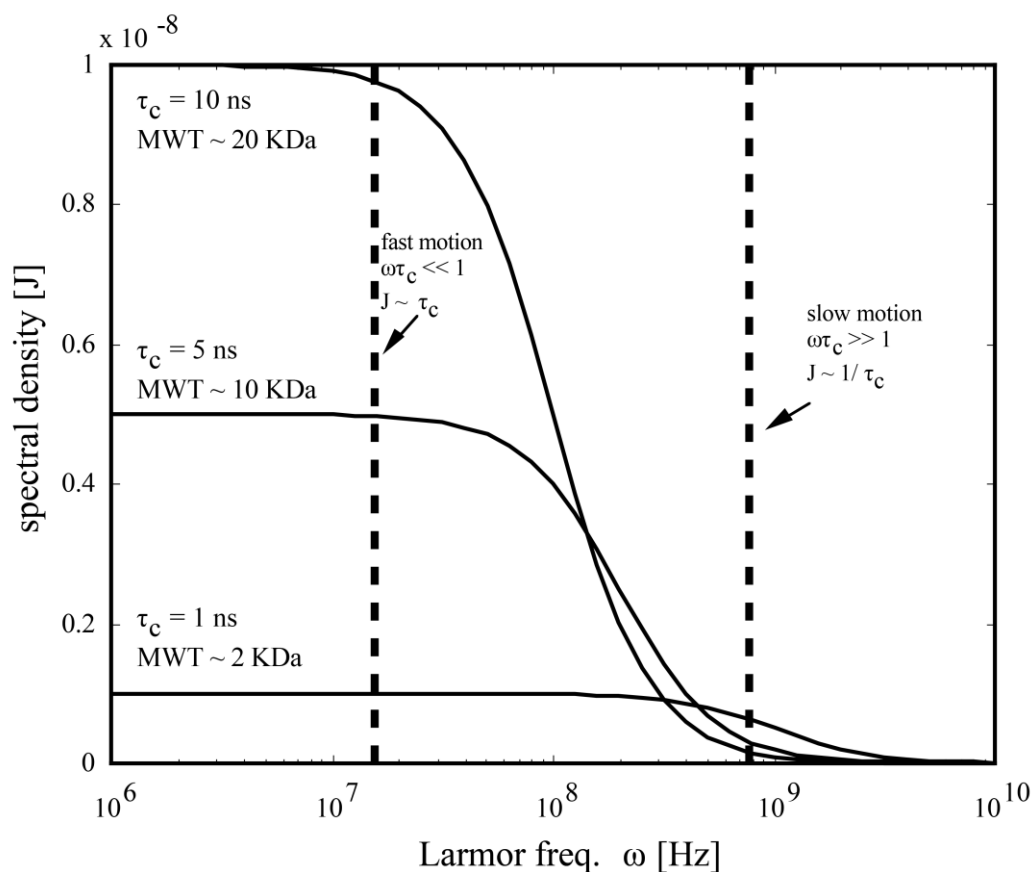


Figure 11 Spectral density for different Larmor frequencies and rotational correlation times (adapted from Understanding NMR spectroscopy, James Keeler; 2002)

The transverse relaxation or spin-spin relaxation (T_2 relaxation) on the other hand is caused by the interaction between nuclear spins leading to the loss of the coherence between them. This is manifested as the loss of x and y magnetization. The time constant T_2 or relaxation rate $R_2=1/T_2$ describes the exponential decay of the magnetization caused by the spin-spin relaxation. The ratio of T_1/T_2 describes the rotational correlation time τ_c or the molecular tumbling of the protein in solution. τ_c gives information about the molecular size and the flexibility of each amino acid in the protein sequence (Kay et al. 1989). The dependency of T_1 and T_2 as a function of τ_c is shown in **Figure 12**.

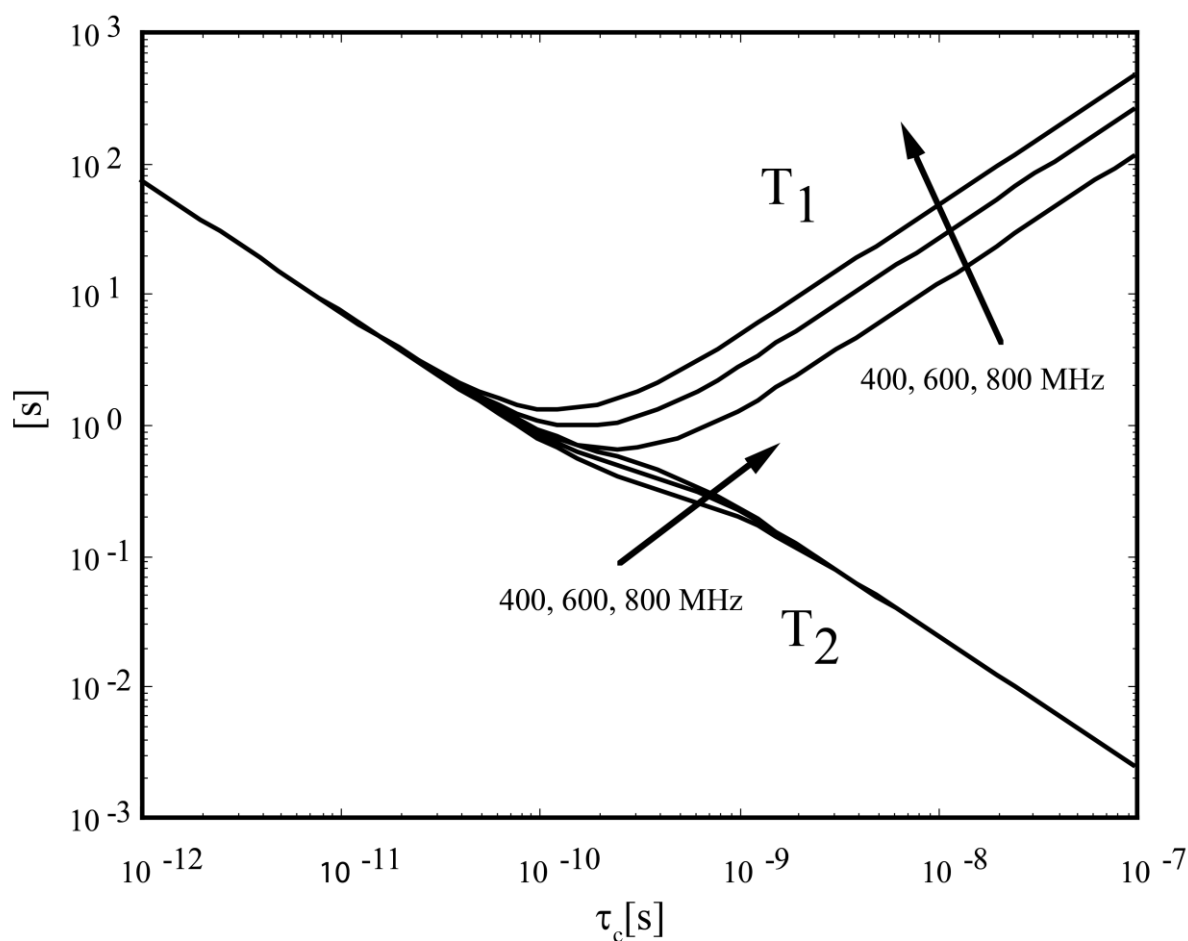


Figure 12 Behavior of T_1 and T_2 as a function of τ_c

T_1 and T_2 for two-spin system consisting of two protons with identical Larmor frequencies (400, 600 or 800 MHz) at a distance of 2 Å as a function of the correlation time is shown

The $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE experiment gives information about internal motion of individual H-N bond at sub nanosecond time scales. This is measured by saturating the

proton (^1H) signal and observing changes in the ^{15}N signal. The rate at which this occurs is the heteronuclear cross relaxation rate. The proton spin and heteronuclear spins are often called as I and S, respectively. The steady state NOE enhancement compares the z-magnetization of the S-spin in thermal equilibrium to the z-magnetization of the S-spin at equilibrium when the I-spin is saturated:

$$NOE(\{I\} - S) = \frac{\langle\langle S_z \rangle\rangle_{I_{sat}}}{\langle\langle S_z \rangle\rangle_{eq}} \quad \text{Eq. 9}$$

Flexible regions of the proteins show faster overall tumbling and decreased NOE intensity compared to the average observed. The ^1H - ^{15}N heteronuclear NOE has an average intensity of 0.77 and values lower than this indicates flexible regions of the protein (Kay et al. 1989).

2.2 X-ray crystallography

As of May 31, 2016, there are 118,949 structures deposited in PDB of which 106,462 are crystal structures, 11,430 are NMR structures and 1,057 are EM structures. As 89.5% of the structures present in PDB are determined by X-ray crystallography, it is one of the widely and primarily used technique for protein structure determination followed by NMR and EM. Although, crystallography gives a static picture of the macromolecular structure, no size limitations for studying protein molecules and ease of use are the primary reasons for its method of choice for studying macromolecular complexes.

2.2.1 Protein crystallization

For obtaining a three-dimensional crystal structure, well diffracting protein crystals are required. Prerequisites for obtaining such crystals are homogeneous and highly pure protein samples. Usually, the primary condition for the crystallization of the protein is found by setting up several sparse matrix screens. If the obtained crystals from the screening are not suitable for the diffraction experiments, they are optimized by grid screening around the parent condition to obtain well diffracting crystals.

Crystallization of proteins is a multi-parametric process involving crystal nucleation and growth. There are several methods to crystallize proteins. Aim of all these methods is to bring the protein to a super-saturation state where usually there is a high probability of crystal nucleation and growth. Two of the primary methods used for achieving the super-saturation phase of the protein crystallizations are vapor diffusion and dialysis. Here only vapor diffusion method would be discussed, as this was the technique chosen for crystallizing protein crystals in the present work.

The vapor diffusion method could be carried out by using sitting or hanging drop methods wherein the crystallization drop is set by mixing protein and the crystallization buffer and the drop is then equilibrated against a reservoir solution of crystallization buffer. Drop equilibration is carried out due to differences in the vapor pressure of the reservoir and the crystallization drop. During this equilibration, the precipitant concentration slowly increases in the crystal drop leading to the protein reaching super saturation.

2.2.2 Principle of X-ray crystallography

Diffraction is the phenomenon of the slight bending of light as it passes around the edge of an object. The amount of bending is dependent on the relative size of the wavelength of light

to the size of the opening. The diffraction pattern observed when the light passes through a slit would show constructive and destructive interference arising due to in phase and out of phase interaction of light waves, respectively.

In a protein crystal, the protein molecules are arranged in an ordered manner. It can be considered as a grid defined by three axes and the angles between them. Each repetitive unit in the crystal is called a unit cell. When X-rays pass through the crystal, they are diffracted due to their interaction with the electron cloud surrounding the atoms of the crystals. In the unit cell, the unit cell constants are represented by the axes and the angles between them, denoted as a , b , c and α , β , γ , respectively. Each atom in the crystal could be represented by a point to obtain a crystal lattice. Within this crystal lattice, infinite numbers of planes could be drawn through the lattice points and the lattice could be represented by Millers indices (hkl) . The index h represents the number of times the ‘ a ’ axis is cut by these planes and so on.

2.2.3 Bragg's Law

The diffraction from the single crystal can be mathematically treated as a reflection from a set of equivalent parallel planes. According to the Bragg's law, these set of planes will produce a constructive interference pattern when the following equation is satisfied:

$$n\lambda = 2d\sin\theta \quad \text{Eq. 10}$$

where n is a positive integer, λ is the wavelength of the radiation, d is the spacing between the Millers planes and θ is the scattering angle.

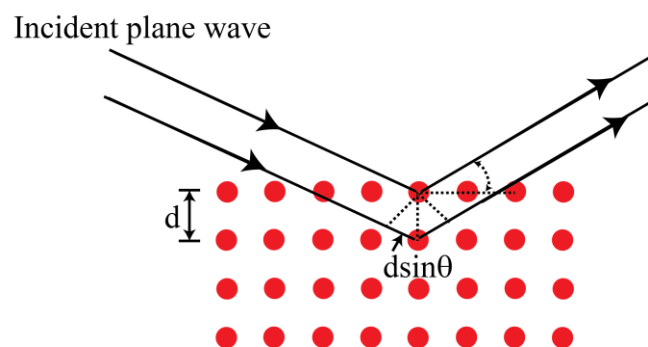


Figure 13 Schematic to derive Bragg's Law

The Miller plane formed by the lattice points (atoms in protein crystals in real space) are shown by red dots. The light is diffracted only by the Miller planes, which satisfies the Bragg's equation.

The diffraction from the protein crystals can be interpreted by the Ewald's sphere; a geometric construction proposed by Ewald in 1921 (Ewald, 1921). The sphere centered on the crystal M has a radius of $1/\lambda$. As the beam s_0 is scattered by the crystal M, a reflection hkl occurs in the direction of MP (s) when reciprocal lattice point P_{hkl} meets this sphere. hkl is the result of the reflection from the set of equivalent real-space planes hkl . As the crystals rotates, other lattice points come into the contact with this sphere thus producing new reflections.

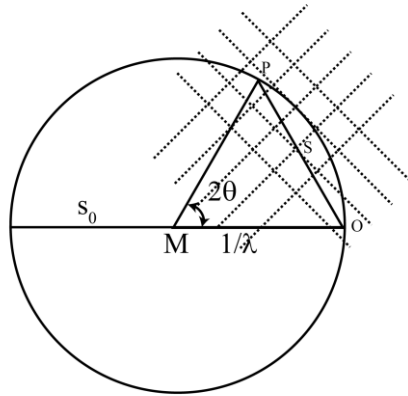


Figure 14 Schematic of the Ewald sphere

The Ewald's sphere provides a convenient tool to explain the diffraction produced by crystals. Only those lattice points that come in the contact with the Ewald's sphere are observed in the diffraction pattern.

The diffraction pattern produce by a crystal lattice is also a lattice but the dimensions of the unit cell of the diffraction lattice in the real space are inversely proportional to the lattice in the reciprocal space. The intensity of a reflection with Miller indices hkl is proportional to $F_{(hkl)}^2$ where $F_{(hkl)}$ is given by:

$$F(hkl) = \sum_j f_j \exp[2\pi i(hx_j + ky_j + lz_j)] \quad \text{Eq. 11}$$

and Intensity $I_{(hkl)}$ is given by

$$I_{(hkl)} = F_{(hkl)}^2 \quad \text{Eq. 12}$$

where f_j in above structure factor equation is the atomic scattering factor for the X-ray for the j^{th} atom of the co-ordinate (x_j, y_j, z_j) expressed as a fraction of the unit cell constants a, b, c . The electron density $\sigma_{(x,y,z)}$ of the unit cell is the Fourier transform of the structure factor equation and it relates the electron density with the structure factor $F_{(hkl)}$ and is given by:

$$\sigma_{(x,y,z)} = (1/V) \sum_{hkl} F_{(hkl)} e^{-i\alpha_{hkl}} e^{-2\pi i(hx+ky+lz)} \quad \text{Eq. 13}$$

If the phases α_{hkl} and the amplitude of all the hkl planes are known, then the electron density can be calculated for all the points (x, y, z) in the unit cell and the crystal structure can be solved. To determine the phases in protein crystallography, three methods can be used:

- 1) Molecular replacement (MR)
- 2) Multiple isomorphous replacement (MIR) and,
- 3) Multi-wavelength anomalous diffraction (MAD).

As only MR was used in this thesis to solve the structures, it is described below in brief.

2.2.4 Molecular replacement

Molecular replacement is the method to obtain the first model of a protein using the structure of a homologous protein. With the structure of the homologous protein, the starting set of phases are calculated with the amplitude of the unknown structure and then the phases are refined iteratively to build the final model. In order to calculate the initial phases from the homologous protein, the protein must be oriented and positioned in the unit cell of the target molecule in such a way that maximizes the overlap of the diffraction pattern of the search model and the target protein.

The Phaser (McCoy et al. 2007) and the Molrep (Vagin and Teplyakov 2010) software, which are usually used for the molecular replacement phasing, first do a rotation search of the protein structure to determine the spatial orientation of the known and unknown molecules with respect to each other. Once this is done; the software then does a translational search to superimpose the now correctly oriented molecule onto the other one.

It is not always straightforward to calculate the phases using molecular replacement as the flexible regions in the known structure of the homologous protein may not necessarily superimpose on the unknown structure. Thus, the model may need extensive modification such as deletion of the flexible regions, side chains or change of the resolution range of the X-ray data used for the search. Given the substantial number of the protein structures in the protein data bank, MR has become an extremely useful technique for the structure determination of the proteins.

2.3 Small Angle X-ray Scattering

Small-angle X-ray scattering (SAXS) is a powerful technique to study biological macromolecules in solution. Unlike X-ray crystallography and NMR which yield atomic level information about the protein structure, SAXS is more suitable for acquiring low resolution information about the macromolecules, usually yielding structural details at about 20Å resolution. Nevertheless, the technique has become popular and widely successful for its application in structural biology in the recent years due to its ease of use and accessibility of high intensity X-rays owing to the easy access of synchrotron radiation. Besides this, it is also a rapid technique with time for data collection at the synchrotrons being few seconds to few hours at the home source.

For SAXS measurements, around 30-100 µl of sample is required per measurement and each protein is usually measured at different concentrations (1-10 mg/ml). Therefore, overall, the sample requirement for the SAXS experiment is quite low requiring 1-2 mg of protein in total. From the SAXS analysis of the macromolecules, in addition to obtaining few simple geometric parameters such as the radius of gyration (R_g) and the maximum dimension (D_{max}) of the macromolecule, it is possible to extract the overall shape of the molecule in the form of molecular envelopes and thus also the probable conformation of the macromolecule in solution.

During the experiment, the solution of the macromolecule that is usually present in a capillary tube is exposed to the X-rays. As the macromolecule in the solution is present in all the possible orientations, the resulting diffraction pattern is radially averaged. This spatial averaging of the data due to the random orientation of the particles in the sample leads to the low resolution obtained in the SAXS experiments and hence is an inherent property of the experiment

The intensity of the diffracted X-rays in SAXS is expressed as the scattering vector q , which is inversely proportional to the wavelength λ and directly proportional to the scattering angle θ (Equation Eq. 14). As the buffer itself, in which the macromolecule is dissolved, also diffracts substantially, the SAXS 1D curve obtained from the sample has to be subtracted from the buffer curve to obtain the SAXS 1D profile of the protein.

$$q = \frac{4\pi \sin(\theta)}{\lambda} \quad \text{Eq. 14}$$

The information content of a SAXS curve is illustrated in **Figure 15**. At low q value ($0-5 \text{ nm}^{-1}$), the curve decays rapidly and essentially depict the shape of the particle. As different proteins differ significantly in shape and size, this region differs clearly for different proteins or different shapes of the same protein. At medium resolution ($5-10 \text{ nm}^{-1}$), the differences in the curves for different proteins start vanishing and at bigger q values, all curves look essentially the same. As diffraction intensity from water overlays the protein signal at $q > 20 \text{ nm}^{-1}$, the SAXS data is not recorded beyond this region. Therefore, SAXS data indeed contains information about the shape, quaternary and tertiary structure of the macromolecule but is not suitable for analyzing the atomic structure of the macromolecule.

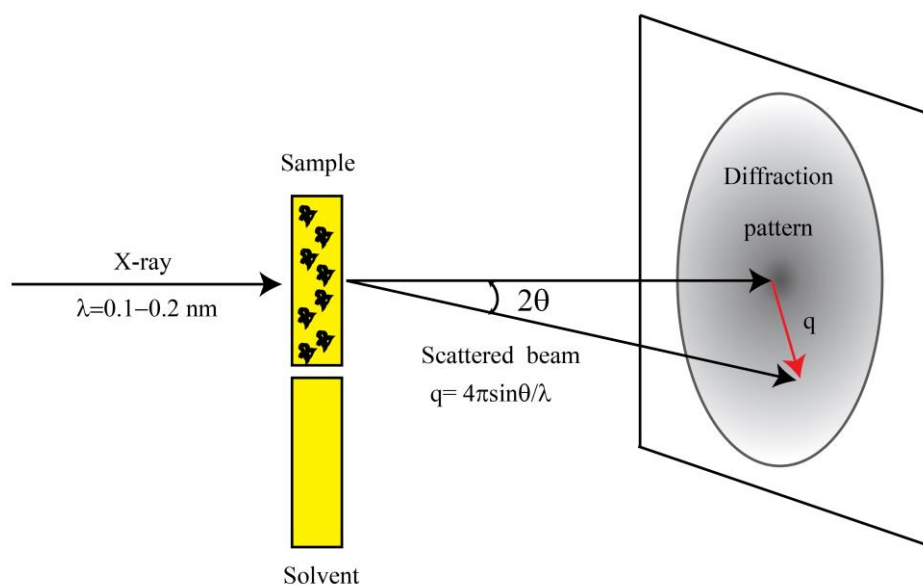


Figure 15 Schematic of the SAXS experiment setup

X-rays diffracted from the protein sample are subtracted from the diffraction produced by the solvent alone and the data is circularly averaged to obtain the experimental 1D SAXS curve for a protein.

The experimental SAXS data has three different regions namely Guinier, Fourier and Porod from which information could be extracted. The determination of the experimental R_g is based on the Guinier approximation and is only true for the Guinier region where no intramolecular interference is observed. In a Guinier plot ($\ln I(s)$ vs s^2), the R_g is determined from the slope of the linear part which satisfies the condition $s \times R_g < 1.3$. As the R_g is highly affected by the polydispersity and aggregation of the macromolecules, the Guinier plot already gives insights into the oligomeric states of the protein.

As the X-ray scattering for proteins is a function of their electron density, the scattering profile could be written as the Fourier transform of the spatially averaged autocorrelation function of the electron density (Patterson function) (Equation Eq. 15).

$$I(s) = 4\pi \int_0^{D_{max}} \langle \rho(r) \times \rho(-r) \rangle \frac{\sin(qr)}{qr} r^2 dr \quad \text{Eq. 15}$$

The Patterson function multiplied by r^2 is called pair distribution function ($p(r)$ function; Equation Eq. 16) which gives the distribution of the electrons that are within distance r of each other. This thus yields the maximum intra-particle distance (D_{max}) and the R_g . The shape of the $p(r)$ curve also tells about the overall shape of the particle. From the Porod plot ($q^4 I(q)$ vs q), the information about the Porod volume and molecular weights of the macromolecules could be obtained.

$$p(r) = r^2 \langle \rho(r) \times \rho(-r) \rangle \quad \text{Eq. 16}$$

The SAXS data could be used to calculate the ab-initio model of the protein. Here, usually a search volume big enough to represent the protein is filled with dummy beads where each amino acid is represented by a bead. A theoretical SAXS curve from this bead model is determined and compared with the experimental SAXS curve. Thereafter the positions of the beads are varied with trial and error method until the χ^2 of the superposition of the theoretical and experimental SAXS curve is minimum.

In case the high-resolution structure of the protein is already known, the theoretical SAXS curve of this structure could also be compared to the experimental curve. The χ^2 of this comparison provides information about the agreement of the high resolution structure with the SAXS data.

Overall, SAXS provides a quick method to validate the crystal or NMR structures of proteins. In addition, useful information about the polydispersity of the protein, its oligomeric state at different concentrations and changes in the shape of the protein in the presence and absence of the ligand can be deduced from the SAXS data analysis.

Scope of the Thesis

Splice site recognition by the splicing machinery is fundamental to the pre-mRNA splicing and is initiated by the binding of U1 snRNP to the 5'ss and the U2 snRNP auxiliary factor U2AF to the polypyrimidine tract. U1 snRNP recognizes 5'ss by base-pairing to the 5' end of the U1 snRNA. Given the essential nature of this process, the splice site selection during the early spliceosome assembly is highly regulated by *cis* and *trans* regulatory elements. It is not surprising that misregulation of the splicing process has been associated with many diseases.

TIA-1 is a *trans* acting splicing factor which regulates the U1 snRNP recruitment to the 5' alternative splice sites of several pre-mRNAs including apoptosis-linked *Fas* pre-mRNA. This role of TIA-1 is attributed to the binding of its RRM2,3 domains to uridine rich sequences downstream of the 5'ss and the interaction of its RRM1 and Q-rich domain with the U1 snRNP specific protein U1C. However, the structural basis of these interactions remain unknown.

Splicing regulation of the *Fas* pre-mRNA at the 3'ss by *trans* acting factor SPF45 was shown to depend on interactions of its U2AF homology motifs (UHM) domain with UHM Ligand Motif (ULM) in constitutive splicing factors. UHMs are atypical RNA recognition motif domains that mediate critical protein-protein interactions during the regulation of alternative pre-mRNA splicing and other processes. Various UHM containing proteins have been associated with different diseases in general and SPF45 has been implicated in breast and lung cancer in particular. Therefore, targeting the UHM-ULM interactions with inhibitors could provide a viable tool to study the role of these interactions in spliceosome assembly and the role of UHM domains in diseases. In addition, such inhibitors could also provide a means to stall spliceosome assembly at early stage making its biochemical and structural analysis feasible.

The aim of this thesis is to understand the structural basis underlying the U1 snRNP recruitment by TIA-1 at the 5'ss using an integrated approach of Nuclear magnetic resonance (NMR), X-ray crystallography and small angle X-ray scattering (SAXS) and to modulate the early spliceosome assembly by developing inhibitors of the UHM-ULM interaction using structure based drug discovery approach.

Chapter 3 Materials and Methods

3.1 Materials

3.1.1 Buffers

Buffer	Components
Lysis Buffer	20 mM Tris pH 7.5, 500 mM NaCl, 10 mM Imidazol, 0.002% NaN ₃ , 1 mM β-mercaptoethanol
Elution Buffer	20 mM Tris pH 7.5, 250 mM NaCl, 500 mM Imidazol, 0.002% NaN ₃ , 1 mM β-mercaptoethanol
TEV cleavage Buffer	20 mM Tris pH 7.5, 150 mM NaCl, 0.002% NaN ₃ , 1 mM β-mercaptoethanol
U1C Lysis buffer	20 mM HEPES-Na, pH 7.5, 500 mM NaCl, 1 M urea, 5 mM β-mercaptoethanol
SP-A buffer	20 mM HEPES-Na, pH 7.5, 100 mM NaCl, 1 M urea, 1 mM PMSF, 5 mM β-mercaptoethanol
SP-B buffer	20 mM HEPE-Na, pH 7.5, 2 M NaCl, 1 M urea, 1 mM PMSF, 5 mM β-mercaptoethanol
HA-Dilution buffer	10 mM KPi, pH 7.4, 5 mM β-mercaptoethanol, 1 mM PMSF
HA-A Buffer	10 mM KPi, pH 7.4, 75 mM NaCl, 5 mM β-mercaptoethanol
HA-B Buffer	10 mM KPi, pH 7.4, 75 mM NaCl, 5 mM β-mercaptoethanol, 12% w/v (NH ₄) ₂ SO ₄
NMR/ITC Buffers	
SPF45	50 mM KPi, 150 mM NaCl, 1 mM DTT, pH 6.8
RRM1/TIA-1	50 mM KPi, 100 mM NaCl, 1 mM DTT, pH 6
U1C	20 mM, MES.NaOH, pH 6.5, 100 mM MgSO ₄ •7 H ₂ O, 200 mM NaCl
Crystallization Buffers	
SPF45+cyclic peptide	20 mM Tris pH 7, 150 mM NaCl, 1 mM dithiothreitol
PUF60	20 mM Tris pH 7, 150 mM NaCl, 5 mM β-mercaptoethanol
RRM1_GS15_U1C30-61	10 mM Tris pH 7.5, 100 mM NaCl, 1 mM dithiothreitol
FP Buffer	
SPF45 FP buffer	20 mM Tris pH 7.5, 300 mM NaCl, 1 mM DTT

3.1.2 Media

Medium	Components/Litre
Lysogeny broth (LB) medium	1% tryptone, 0.5% yeast extract, 0.5% NaCl
Terrific Broth (TB) medium	1.2% tryptone, 2.4% yeast extract, 0.5% glycerol, 100 mL TB salts (0.17 M KH ₂ PO ₄ , 0.72 M K ₂ HPO ₄)
¹⁵ N Labelled M9 minimal medium	100 ml M9 salt solution (10x), 20 ml 20% (w/v) glucose, 1 ml 1 M MgSO ₄ , 0.3 ml 1 M CaCl ₂ , 1 ml biotin (1mg/ml), 1 ml Thiamin (1 mg/ml), 10 ml trace elements solution (100x)
¹⁵ N, ¹³ C Labelled M9 minimal medium	100 ml M9 salt solution (10x), 2 g ¹³ C labelled glucose, 1 ml 1 M MgSO ₄ , 0.3 ml 1 M CaCl ₂ , 1 ml biotin (1 mg/ml), 1 ml Thiamin (1 mg/ml), 10 ml trace elements solution (100x)

3.1.3 ¹⁵N labelled M9 salts

¹⁵ N labelled M9 salts (10x)	Mass/Litre
Na ₂ HPO ₄ ·2H ₂ O	75.2 g/L
KH ₂ PO ₄	30 g/L
NaCl	5 g/L
¹⁵ NH ₄ Cl	5 g/L

3.1.4 Trace elements solution

Trace Elements solution (100x)	Mass/Litre
EDTA	5 g/L
FeCl ₃ ·6H ₂ O	0.83 g/L
ZnCl ₂	84 mg/L
CuCl ₂ ·2H ₂ O	13 mg/L
CoCl ₂ ·2H ₂ O	10 mg/L
H ₃ BO ₃	10 mg/L
MnCl ₂ ·4H ₂ O	1.6 mg/L

3.2 Methods

3.2.1 Protein expression and purification

All proteins were expressed in BL21 (DE3) *Escherichia coli* strain. Plasmid containing gene for the respective protein was transformed into chemically competent *E.coli* cells and grown overnight on kanamycin (50 µg/ml) agar plates at 37 °C. An overnight 20 ml starter culture in LB media was inoculated with a single colony from the plate. Next day, 1 litre of LB media with kanamycin (50 µg/ml) was inoculated with the starter culture. Cells were grown at 37 °C till the O.D. 600 reached 0.6 and the protein expression was induced by adding 0.5 mM IPTG solution after which the cultures were grown overnight at 20 °C. Next day, the culture was centrifuged at 5000 g for 20 min to pellet the cells. The cell pellet was resuspended in 30 ml lysis buffer per litre of culture along with 0.1 mg/ml lysozyme and 1 mM AEBSF. The cells were stored at -20°C until further use.

For ¹⁵N labelling or ¹³C, ¹⁵N labelling of the protein, 20 ml starter culture was centrifuged in a 50 ml falcon tube at 4000 g for 20 min. For uniform labelling, the pellet was washed with 5 ml of minimal media and was resuspended in another 10 ml of minimal media. This resuspended pellet was used to inoculate 1 litre of ¹⁵N labelled minimal medium for further protein expression and purification.

For protein purification, the suspended pellet was thawed and further incubated on ice for 20 min. The cell wall of the bacteria was disrupted by sonication on ice and the cell debris was separated from the cell lysate by centrifuging the cell lysate at 42000 g for 45 min at 4 °C. The supernatant was loaded on a 3 ml nickel column pre-equilibrated with lysis buffer. The column was then washed with 100 ml of lysis buffer and the protein was eluted in 10 ml of elution buffer. In case where it was not required to cleave the expression tag, the eluted protein was concentrated to 5 ml and was directly loaded on the size-exclusion chromatography column (Hiload 16/60 Superdex75 column, GE Healthcare) to further purify the protein. To cleave off the expression tag, 1 mg of TEV protease was added to the eluted protein and the protein was dialyzed against TEV cleavage buffer overnight at 4 °C. Next day, the TEV cleaved protein was loaded on to the 2nd nickel column to separate the expression tag/uncleaved protein from the cleaved protein. The flow-through was collected and concentrated to 5 mg/ml using Amicon® Ultra centrifugal filter units with MWCO 3.5 kDa. Finally, the concentrated protein was loaded on to the size-exclusion chromatography column pre-equilibrated with respective NMR or crystallization buffer. The peak fractions from the Superdex75 column were

concentrated, the proteins were aliquoted in 50 μ L aliquots in 1.5 ml Eppendorf tubes, and then flash-frozen in liquid nitrogen at -80 $^{\circ}$ C until further use.

For purifying SPF45 UHM domain for crystallization, an extra step of Ion exchange chromatography was introduced between the 2nd nickel column and the size exclusion chromatography. Briefly, the protein was diluted with 20 mM Tris buffer pH 7 after the 2nd nickel column until the final NaCl concentration was 50 mM and then loaded on to Mono Q anion exchange column. The protein was eluted with a linear gradient of NaCl from 0 to 500 mM and the peak fractions were concentrated and loaded onto the Superdex75 size-exclusion chromatography column pre-equilibrated with crystallization buffer.

For TIA-1 constructs containing the Q-rich domain, the protein was eluted in elution buffer with 1 M urea to prevent precipitation and fibril formation by TIA-1 at high concentrations. TEV cleavage was carried out in TEV cleavage buffer with 0.5 M urea overnight before loading on to 2nd nickel column. The protein was then concentrated with Amicon® Ultra centrifugal filter units with MWCO 30 kDa before loading on to the size-exclusion chromatography column. The peak fractions were pooled and stored at 4 $^{\circ}$ C until further use.

U1C protein was purified using two steps of ion exchange chromatography followed by size-exclusion chromatography. The plasmid containing the gene for the expression of U1C was transformed in BL21 cells and grown overnight on kanamycin plates (50 μ g/ml). Further, a single colony was streaked on a kanamycin plate and incubated overnight. Next day, 1 litre of TB medium supplemented with kanamycin (50 μ g/ml) was inoculated with cell mass from two such plates and grown at 37 $^{\circ}$ C till the O.D. 600 of the cells reached 1-1.2. The protein expression was induced with 0.5 mM IPTG at 37 $^{\circ}$ C. The cells were harvested after 4 h. For the production of ¹⁵N labelled U1C, 1 litre of LB medium supplemented with kanamycin (50 μ g/ml) was inoculated with two plates of cell mass from the restreaked U1C agar plates. The cells were grown until O.D. 600 reached 1-1.2 after which they were harvested by centrifugation. The cell pellet was resuspended in ¹⁵NH₄Cl containing M9 minimal medium and grown further for 1 h. The protein expression was induced with 0.5 mM IPTG at 37 $^{\circ}$ C and cells were harvested after 4 h. The cells were resuspended in 30 ml of U1C lysis buffer and frozen at -20 $^{\circ}$ C until further use.

For protein purification, the frozen cells were thawed and lysed using sonication. The lysate was centrifuged at 42000 g for 30 min at 4 °C and the supernatant was diluted 5 fold with the lysis buffer without NaCl. The lysate was loaded on a 20 ml bed volume SP-Sepharose column at a flow rate of 2 ml/min. The column was washed with SP-A buffer and the protein was eluted with a linear gradient of SP-B buffer from 100 mM to 2 M NaCl. The eluted peak fractions were checked on gel and U1C fractions were pooled and diluted three fold with HA dilution buffer. These diluted SP-Sepharose fractions were loaded on a 15 ml bed volume hydroxyapatite (HA) column equilibrated with HA-A buffer and the protein was eluted with two step gradient of $(\text{NH}_4)_2\text{SO}_4$ using HA-B buffer.

The peak fractions were analyzed on gel and the U1C fractions were pooled and concentrated with Amicon® Ultra centrifugal filter unit with MWCO 3.5 kDa and loaded on Superdex75 column equilibrated with U1C size exclusion buffer. The peak fractions were concentrated, aliquoted in 50 ml volume in 1.5 ml Eppendorf tubes, snap frozen in liquid nitrogen and stored in -80 °C for further use.

3.2.2 NMR titrations

For SPF45 UHM-peptide titrations ^1H , ^{15}N Heteronuclear single quantum correlation (HSQC) NMR spectra were acquired at 298 K using a AVIII600 Bruker NMR spectrometer equipped with a cryogenic probe. 50 μM of ^{15}N labelled SPF45 UHM domain in NMR buffer and 10% D_2O (for lock) was titrated with two-fold excess of cyclic peptide. Spectra were processed with NMRPipe/Draw (Delaglio et al. 1995) software and analyzed using CCPN (Vranken et al. 2005) analysis software.

To study SPF45 UHM domain-small molecule interaction, the compounds were dissolved in 100% deuterated DMSO. ^1H , ^{15}N HSQC titrations were recorded by titrating the compounds in 50 μM ^{15}N labelled SPF45 UHM domain as mentioned above.

3.2.3 NMR structure calculation and validation of TIA-1 RRM1

The NMR structure of TIA-1 RRM1 domain was calculated using CYANA 3.0 (Guntert 2004). The cross-peaks of ^{15}N - and ^{13}C -edited NOESY-HSQC spectra were assigned in an automated way using CYANA 3.0 and the dihedral angle restraints were predicted using TALOS+ (Shen et al. 2009). 200 structures were calculated using these restraints and the structures were further refined by water-refinement in ARIA 1.2 (Linge et al. 2003a; Linge et al. 2003b). An ensemble of 20 lowest energy structures were selected and further used for

structure validation by iCing (Doreleijers et al. 2012b), PROCHECK (Laskowski et al. 1993) and WHATCHECK (Vriend and Sander 1993).

3.2.4 Assignment of backbone and side-chain resonances of TIA-1 RRM1

All NMR spectra were acquired at 298 K using a AVIII500, AVIII600, AVIII750, AVIII800 and a AVI900 Bruker NMR spectrometer, equipped with cryogenic or room temperature (750 MHz) triple resonance gradient probes. Sample contained ~0.5 mM TIA-1 RRM1 protein in 50 mM Potassium phosphate (pH 6.0), 100 mM NaCl, 1 mM DTT with 10 % D₂O added for the lock. All spectra were processed using NMRPipe/Draw (Delaglio et al. 1995) and analyzed using NMRView (Johnson and Blevins 1994) software. Protein backbone assignments for ¹⁵N, ¹H_N, ¹³C_α, ¹³C_β, and ¹³C' chemical shifts were obtained from HNCA, HNCACB, CBCA(CO)NH and HNCO experiments (Sattler et al. 1999a) and assignments were made manually in CARA (Keller 2004) software.

Three-dimensional total correlation spectroscopy (TOCSY) experiments were performed to assign carbon and proton resonances of the RRM1 side chains. Two HCCH-TOCSY experiments with ¹³C and ¹H evolution were recorded for this, along with CC(CO)NH-TOCSY and HBHA(CO)NH (Grzesiek and Bax 1993) experiments to correlate the amide group resonances with the side-chain residues. Aromatic resonances were assigned using 2-D ¹H-¹³C HSQC, HBCBCGCDHD, HBCBCGCDCEHE (Yamazaki et al. 1993) and ¹³C edited NOESY-HSQC spectra. ¹⁵N- and ¹³C-edited NOESY-HSQC experiments were recorded with 70 ms mixing time. Assignment of side-chain residues and picking of NOESY cross-peaks was carried out in CCPN analysis (Vranken et al. 2005) software.

3.2.5 NMR relaxation measurements

To study whether the three RRM domains of TIA-1 tumble together in solution when bound to RNA and hence to study the dynamics of the protein-RNA complex, NMR relaxation data were recorded for TIA-1 RRM1,2,3 in the presence of U15 RNA. The data were recorded at 298 K for 200 μM of TIA-1 RRM1,2,3 in the presence of 1.2 fold excess of RNA on a 800 MHz Bruker NMR spectrometer. Amide ¹⁵N relaxation data of R_1 , $R_{1\rho}$, and steady-state heteronuclear {¹H}-¹⁵N NOE experiments were performed as described (Tjandra et al. 1995; Massi et al. 2004). R_1 data were measured with thirteen relaxation points with three duplicate delays, 21.6/21.6, 43.2, 86.4, 172.8, 259.2, 345.6/345.6, 518.4, 669.6/669.6, 885.6, 1080, 1296, 1512, and 1728 ms. $R_{1\rho}$ data were recorded using ten different delay points together with two duplicate delays of 5/5, 10, 15, 20, 30, 40, 50, 60, 80/80, and 100 ms. Error was estimated from

duplicate time points. The transverse relaxation rate R_2 for each residue was estimated by correction of the observed relaxation rate $R_{1\rho}$ with the offset $\Delta\nu$ of the radio-frequency field to the resonance using the relation $R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta$, where $\theta = \tan^{-1}(\nu_1/\Delta\nu)$. The correlation time (τ_c) of the protein was then estimated using the ratio of averaged R_2/R_1 values (Daragan and Mayo 1997). Steady-state heteronuclear $\{^1\text{H}\}$ - ^{15}N NOE spectra were recorded with and without 3 s of ^1H saturation. All relaxation experiments were acquired as pseudo-3D experiments and converted to 2D data sets during processing. The peak intensity, relaxation rates and errors were calculated using PINT (Ahlner et al. 2013) software. The relaxation data of the resonances of Phe⁹⁵-Leu¹⁰² were not analyzed because their amide signals could not be observed.

3.2.6 Small angle X-ray scattering experiments

SAXS measurements for TIA-1 RRM1,2,3 (free and bound to RNA) were done on a Rigaku BIOSAXS1000 instrument with a HF007 micro-focus generator equipped with a Cu-target at 40 kV and 30 mA. Transmission measurements were done with a photodiode beamstop. q-calibration was made by a silver-behenate measurement. Sample measurements were done in multiples of 900 second frames checked for beam damage and averaged. Circular averaging and background subtraction was done with the Rigaku SAXSLab software v 3.0.1r1. The data were collected at 25 °C.

For SAXS measurements on U1C protein, 50 μl of sample and buffer were measured at 25 °C at the BioSAXS beamline BM29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, using a 2D Pilatus detector. Ten frames with 2s exposure time per frame were recorded for each complex and buffer sample, using an X-ray wavelength of $\lambda = 1.008 \text{ \AA}$. Measurements were performed in flow mode where samples were pushed through the capillary at a constant flow rate to minimize radiation damage. Frames showing radiation damage were removed prior to data analysis.

The intensities of circularly averaged images of TIA-RRM1,2,3 protein samples were further processed for buffer subtraction in PRIMUS. In case of U1C protein, the dedicated beamline software BsxCuBE was used in an automated manner. The one-dimensional scattering intensities of samples and buffers were expressed as a function of the modulus of the scattering vector $Q = (4\pi/\lambda)\sin\theta$ with 2θ being the scattering angle and λ the X-ray wavelength. After buffer subtraction, R_g of all the samples were determined using the same program using Guinier approximation. The validity of the Guinier approximation, R_g for $Q < 1.3$, was checked

and fulfilled in each case. R_g and D_{max} were calculated from pairwise distribution functions using GNOM (Svergun 1992).

3.2.7 Crystallization of TIA-1 RRM1-GS15-U1C30-61

For the crystallization of TIA-1 RRM1-GS15-U1C30-61 construct, sparse matrix crystallization screens were set up at room temperature and 4 °C. The plates were set up with two different concentrations of protein (4.2 mg/ml and 8.4 mg/ml) as the protein showed concentration dependent dimerization as observed with dynamic light scattering. Crystals were observed in 2 different conditions in 8.4 mg/ml drops. Hexagonal crystals appeared in 3 days in 0.1 M Na-Cacodylate, 1 M Na-citrate tribasic pH 6.5 and rod shaped crystals appeared in 12 days in 0.1 M HEPES pH 7.5, 10% PEG6000 and 5% MPD. The Na-Cacodylate condition was not reproducible even after repeated attempts (grid screening around parent condition and Hampton Additive screen). Contrastingly, the HEPES condition readily reproduced producing diffraction suitable crystals in 1 day when set up with optimized condition obtained using grid screening approach. The final crystallization condition contained 0.1 M HEPES pH 7.5, 10% PEG6000 and 10% MPD. The crystals were cryo-protected in 0.1 M HEPES pH 7.5, 10% PEG6000 and 30% MPD and were flash frozen in liquid nitrogen.

Datasets for the crystals were collected at ID29 beamline equipped with PILATUS3 6M detector at ESRF Grenoble. Datasets from best diffracting crystals were processed with XDS (Kabsch 2010) software package and the structure was solved by molecular replacement using Phaser (McCoy et al. 2007). The NMR structure of TIA-1 RRM1 and the PDB coordinates of U1C (30-56) present in the U1snRNP crystal structure (PDB id: 4PJO) were used as the search models. The missing residues were built using Coot (Emsley and Cowtan 2004) model building software with multiple rounds of model building and refinement using Refmac (Murshudov et al. 1997) software from CCP4 (Winn et al. 2011) suite.

3.2.8 SPF45 UHM-cyclic peptide crystallization and data processing

For crystallization of SPF45 UHM domain-cyclic peptide complex (SPF45 UHM domain and cyclic peptide mixed in 1:1.5 molar ratio), sparse matrix crystal screens were set up at room temperature and 4 °C. A potential condition was identified in the room temperature screen and the condition was further optimized by grid screening around the parent condition. The refined crystallization condition contained 50 mM MES pH 6.0 and 70% MPD. Crystals were obtained by mixing 2 µl protein (at 10 mg/ml concentration in 20 mM Tris pH 7, 150 mM NaCl and 1 mM DTT) and 2 µl reservoir solution using hanging drop method. Thin plate

crystals suitable for diffraction and data collection were obtained in 5-7 days. The crystals were directly flash frozen in liquid nitrogen for data collection as MPD provides a good cryo-preserved for crystallization. Several datasets for the crystals were collected at the PXIII beam line at Swiss light source. The datasets from the best diffracting crystals were integrated and scaled with the XDS (Kabsch 2010) software package. The structure was solved via molecular replacement using the native structure of SPF45 as a search model (PDB id: 2PE8) using Phaser (McCoy et al. 2007). The cyclic peptide was built in the visible electron density using Coot (Emsley and Cowtan 2004) model building software and the model was further refined in Refmac (Murshudov et al. 1997) from the CCP4 suite (Winn et al. 2011).

3.2.9 Puf60-small molecules crystallization and data processing

Crystals of thioredoxin tagged Puf60 was reproduced from the already published condition (Corsini et al. 2008). Briefly the crystals were obtained by vapor diffusion hanging drop method containing 1 µl protein (at 75 mg/ml concentration in 20 mM Tris pH 7.0, 50 mM NaCl, 1 mM BME) and 1 µl of well solution (1.4 M (NH₄)₂SO₄ and 50 mM K-formate). Crystals suitable for diffraction grew in 4-5 days. Crystals were soaked overnight in 2 µl fresh solution containing 1.5M (NH₄)₂SO₄, 50mM K-formate and 1 mM of small molecule inhibitor. Next day the crystals were cryo-protected by serial transfer into a solution of 1.5M (NH₄)₂SO₄, 50 mM K-formate, 1 mM small molecule inhibitor and 20% ethylene glycol. Several datasets of crystals soaked in different inhibitors were collected at the beamlines available at European synchrotron research facility (ESRF), Grenoble, France.

Datasets for best diffracting crystals for each inhibitor were integrated and scaled with XDS (Kabsch 2010) software package. All the structures were solved with molecular replacement method using the crystal structure of thioredoxin tagged Puf60 (PDB id: 3DXB) as the search model. Solutions could be found for crystals soaked in TOK116, TOK196, TOK211, TOK246, TOK263 and dimethoxy-chlorpromazine inhibitors whereas other crystals showed severe twinning with no solution after molecular replacement in Phaser (McCoy et al. 2007). The inhibitors and the missing residues were built in the visible electron density after molecular replacement using Coot (Emsley and Cowtan 2004) model building software. The coordinates and the restraints files for the inhibitors were obtained from the PRODRG server (Schuttelkopf and van Aalten 2004). The built models were refined with Refmac (Murshudov et al. 1997), analyzed in Chimera software (Pettersen et al. 2004) and images suitable for publication were made in PYMOL (Schrodinger 2015) software.

3.2.10 Isothermal Titration Calorimetry (ITC)

Prior to ITC experiments, all proteins were dialyzed overnight into the ITC buffer. ITC experiments were performed by titrating cyclic peptides into 5-20 μM of SPF45 UHM domain. A 10-fold concentration of cyclic peptides was used in the syringe and a series of 1.5 μl injections were made into the cell. The experiments were performed with ITC200 Microcal system and the data were fit with the Origin software provided with the instrument using a one-site binding model.

3.2.11 Fluorescence Polarization Assay

Fluorescence polarization is a solution-based technique that could be used to study interaction between two molecules. The technique provides information on the molecular orientation and mobility of the molecules and is based on the change in the degree of polarization of the emitted light. If the molecular weight of the fluorescently labelled molecule is small, it tumbles fast in solution due to the Brownian motion thus depolarizing the light whereas fluorescent molecules with large molecular weight tumble relatively slowly thereby producing a small change in the degree of polarization.

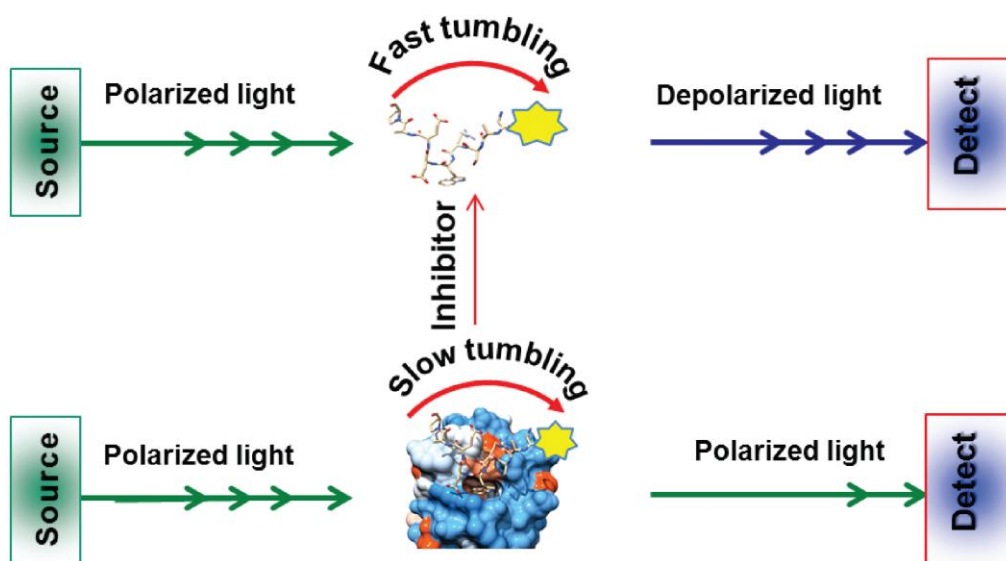


Figure 16. Principle of Fluorescence polarization assay

The basic principle of the FP assay is shown. Fluorescein was used as a fluorescent tag to label the cyclic peptide probe.

To develop the FP assay for studying UHM-ULM interaction, a combination of Fluorescein labelled cyclic peptide derived from the native SF3b155 ULM sequence was used

as a probe and Z-tagged SPF45 UHM domain was used as its binding partner. The labelled cyclic peptide tumbles freely in solution when in isolation, thus depolarizing the polarized light whereas when it interacts with the UHM domain, the whole complex tumbles slowly due to increase in the molecular weight leading to a decrease in the amount of depolarized light.

Fluorescence polarization (FP) assay was carried out in a 384 well plate. The buffer conditions used for the FP assay were 20 mM Tris pH 7.5, 300 mM NaCl, 1 mM DTT, 100 nM tracer and 1.6% DMSO that was decided based on the maximal signal gained in multiple rounds of assay optimization. During assay optimization, several pH, NaCl and tracer concentrations were tested. For obtaining the binding curve, a constant amount of tracer (100 nM) was titrated into serial dilutions of protein. The assay was carried out in 40 μ l volume and the protein tracer mixture was incubated for 1 h before reading the plate. The polarization was measured in Envision plate reader (Perkin Elmer, Waltham, MA) using FP480 (excitation) and FP533 (emission filters). The millipolarization (mP) values were calculated for each data point using:

$$mp = \frac{1000 \times (S - G \times P)}{(S + G \times P)}$$

where S and P are the fluorescence counts rated on the planes parallel (S) or perpendicular (P) to the excitation filter and G is the grating factor which is dependent on the factors of the instrument.

The data was plotted against the log protein concentration and was fitted with Sigmaplot software using the following 4-parameter logistic nonlinear regression model to obtain the binding curve:

$$F(x) = \frac{(A - D)}{\left(1 + \left(\frac{x}{c}\right)^B\right)} + D$$

where x is protein concentration, A is minimum asymptote, B is Hill slope, C is inflection point and D is maximum asymptote.

The binding curve was used to define the EC₈₀ concentration (concentration of protein required to achieve 80% of maximal response units) which was further used in the high-throughput screening.

3.2.12 High-throughput screening

A high-throughput screening (HTS) by means of single point titration in the FP assay was carried out. For this, the FP assay was adopted to work with a robotic set-up. The protein (28 μ l of 1.4-fold EC80 concentration calculated from the binding curve) was added to the assay plate by the Multidrop Combi reagent dispenser and the buffer tracer mix (12 μ l of 333.3 nM tracer + 5.3% DMSO stock solution in FP buffer) was added by the robot.

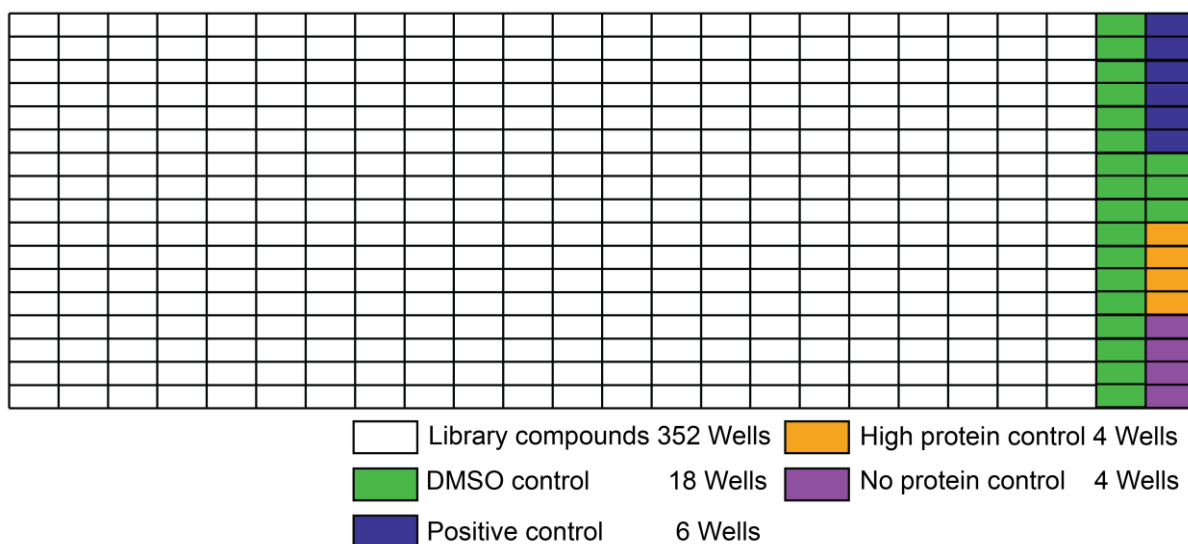


Figure 17. Schematic overview of the 384 plate used for the high-throughput screening

The design of the plate used in HTS is shown. Last two columns were used for various controls as mentioned in the figure legend.

The protein concentration used in the HTS was taken as that required for achieving 80% of maximum response in FP units for tracer binding. It was determined from the binding curve obtained by titrating a series of protein concentrations in 100 nM of tracer. The protein concentration used in the assay varied from batch to batch (of purification) and was determined afresh for each batch. Last two columns of the assay plate were used for controls in each plate. The controls in the plates included DMSO control (negative control with no compound), positive control (100 μ M of cyclic peptide inhibitor), high protein and no protein controls (to obtain the assay window). The diagrammatic format of the plate is shown **Figure 17**.

The compound library was also provided in a 384 well format. The compounds were pipetted directly in the assay plate to a final concentration of 125 μ M. After addition of the compounds, the plates were sealed and incubated for 1 h in dark. The compound libraries used

for HTS were obtained from Prestwick, Chemdiv/Berg, MayBridge, Chemdiv/CBN, LOPAC, Tocris, BioMOL and a custom library developed in Dr. Felix Hausch's group at the Max Planck institute of Psychiatry.

3.2.13 AlphaScreen assay

AlphaScreen (Amplified Luminescence Proximity Homogeneous Assay) is a bead-based proximity assay. It is based on the principle of singlet oxygen transfer from a donor bead to the acceptor bead when the two are in close proximity to each other (< 200nm).

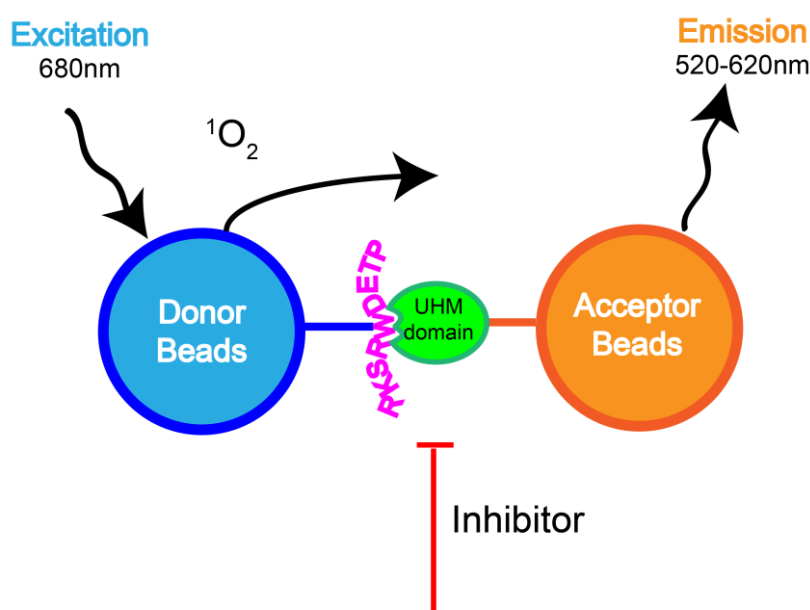


Figure 18. Schematic overview of the AlphaScreen technology

The assay setup used for the AlphaScreen is shown. The ULM peptide was tagged with biotin that could then bind to the streptavidin donor beads whereas the his-tagged UHM domain could bind to the Ni²⁺-NTA acceptor beads.

For our purpose, we used streptavidin donor beads, which could bind biotinylated peptide, and Ni²⁺ acceptor beads, which could bind His-tagged protein. If the peptide and the protein interact, the beads are expected to come in close proximity favoring the transfer of singlet oxygen species when excited with a laser of 680 nm. The principle is shown in **Figure 18**.

The assay was carried out by incubating 10 nM protein with 10 nM biotinylated peptide (biotinylated-RKSRWDETP) in 20 mM potassium phosphate pH 7.5, 150 mM NaCl, 0.5% bovine serum albumin, 0.05% NP40 detergent, 1 mM BME and 1% DMSO for 30 min. In case of the competition assay, the desired small molecule inhibitor was added and the mixture was

incubated in the dark for 30 min. The data were plotted in OriginPro9.0 software using a nonlinear dose response curve fitting function.

Chapter 4
Structural insights into the interaction of TIA-1 with RNA and U1C

In line with the aim of the thesis, the effect of RNA binding to TIA-1 RRM1,2,3 and interaction between TIA-1 and U1C was studied using integrated structural biology approach.

4.1 RRM1, 2, 3 forms a compact shape in the presence of RNA

4.1.1 NMR relaxation studies of TIA-1 RRM1,2,3-RNA complex

It is now well established that protein flexibility plays a crucial role in the functioning of the protein molecules. Protein flexibility can greatly vary in the presence and absence of the ligand. At least in case of multi-domain proteins connected by long flexible linkers, the change in the protein flexibility upon ligand binding is usually accompanied by large changes in the overall protein structure. NMR combined with SAXS provides a good tool to study the protein flexibility in the presence and absence of the ligand along with the changes in the shape of the protein. Here, NMR spectroscopy was employed to study the dynamics of individual residues at ns timescale of TIA-1 RRM1,2,3 in the presence and absence of 15mer poly U RNA (U15). The changes in the shape of the protein in the presence and absence of RNA were studied by SAXS.

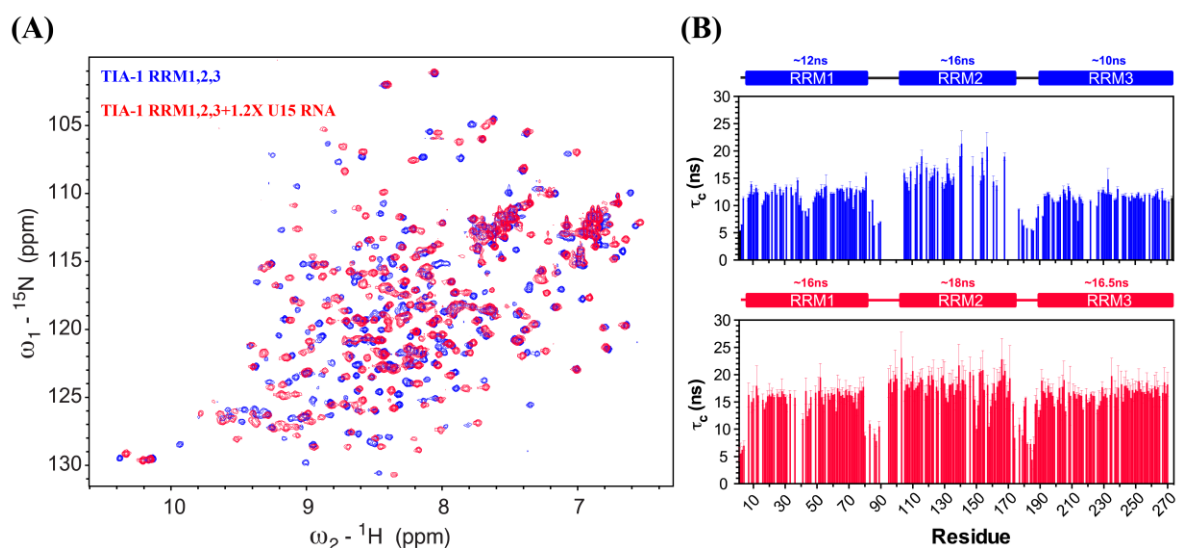


Figure 19. TIA-1 RRM1,2,3 tumble together in presence of RNA

^{15}N relaxation data were recorded to determine the τ_c values of TIA-1 RRM1,2,3 in the presence and absence of RNA. A) ^1H - ^{15}N HSQC spectrum of TIA-1 RRM1,2,3 in the presence and absence of U15 RNA (Data recorded and analyzed by Dr. Iren Wang). B) The three RRM domains of TIA-1 act as independent modules in the absence of RNA (blue) showing different τ_c values. However, in the presence of U15 RNA (red), the τ_c value for individual domains is more close to each other signifying that the three domains tumble together in solution. (Data already published in (Wang et al. 2014))

Previously, it was shown that TIA-1 RRM1 binds very weakly to RNA. This conclusion was made based on the interaction of individual domains with RNA. However, the contribution

of RRM1 domain for RNA binding remains unclear in context of the full-length protein. In order to understand whether RRM1 contributes to RNA binding and also to determine the overall domain dynamics of the three RRM domains in the presence of RNA, ^{15}N relaxation data for TIA-1 RRM1,2,3 was recorded in the presence and absence of U15 RNA.

The average tumbling time for the residues in RRM1, RRM2 and RRM3 in RRM1,2,3 protein is 12 ns, 16 ns and 10 ns respectively as shown in **Figure 19B**. However, the expected τ_c value for individual RRM1, RRM2 and RRM3 domains in isolation is expected to be 8.6 ns, 6.1 ns and 5.9 ns respectively as calculated with HYDRONMR (Garcia de la Torre et al. 2000; Bernado et al. 2002). The greater than expected rotational correlation time for the RRM domains in the RRM1,2,3 protein could be explained by the motional coupling of the three domains with each other which seems to be present even in the presence of long flexible linkers between RRM1, RRM2 and RRM3.

Therefore, although the three domains show higher rotational correlation time, the τ_c values for each domain differs significantly from one another. This suggests that these three domains tumble individually, independent of each other in solution and do not interact with each other in the absence of RNA and there are no inter-domain contacts between the three domains. Nevertheless, transient and weak inter-domain contacts cannot be ruled out as it has been suggested before that the RRM2 domain thermodynamically stabilizes RRM3 in thermal unfolding experiments (Aroca et al. 2011).

In TIA-1, both RRM2 and RRM3 domains have been shown to bind to RNA and form a compact shape. However, as RRM1 in isolation interacts weakly with RNA, it was not clear whether RRM1 contributes to RNA binding in context of RRM1,2,3. To understand this, we determined the average tumbling correlation time of TIA-1 RRM1,2,3 in the presence of U15 RNA. In the presence of U15 RNA, the three domains show increased average tumbling correlation times of 16 ns, 18 ns and 16.5 ns respectively. This agrees well with the binding of RNA to the RRM domains and thus justifies the increase in the tumbling time. The presence of RNA also decreased the difference between tumbling times of individual domains compared to that in the absence of RNA suggesting that RRM1,2,3 tumble together in the presence of RNA.

4.1.2 SAXS analysis of TIA-1 RRM1,2,3-RNA complex

To analyze the effect of RNA binding on the overall shape of RRM1,2,3 in the presence of RNA, SAXS data was recorded for RRM1,2,3 in the presence and absence of U15 and *Fas* intron6 15mer RNA (GUUCUUGC UUUGUUC). *Fas* intron6 was also chosen in this study as it is the known natural binding target of TIA-1. SAXS data was recorded at three different concentrations: 2.5mg/ml, 5mg/ml and 10mg/ml for RRM1,2,3 free and in the presence of either 1.2X U15 RNA or 1.2X *Fas* intron6 RNA. All the three concentrations of the protein-RNA complex do not show any concentration dependent aggregation behavior as shown by the intensity curve at low q range (**Figure 20A**). Therefore, SAXS curves recorded at 10 mg/ml were chosen for further analysis as they have the least noise.

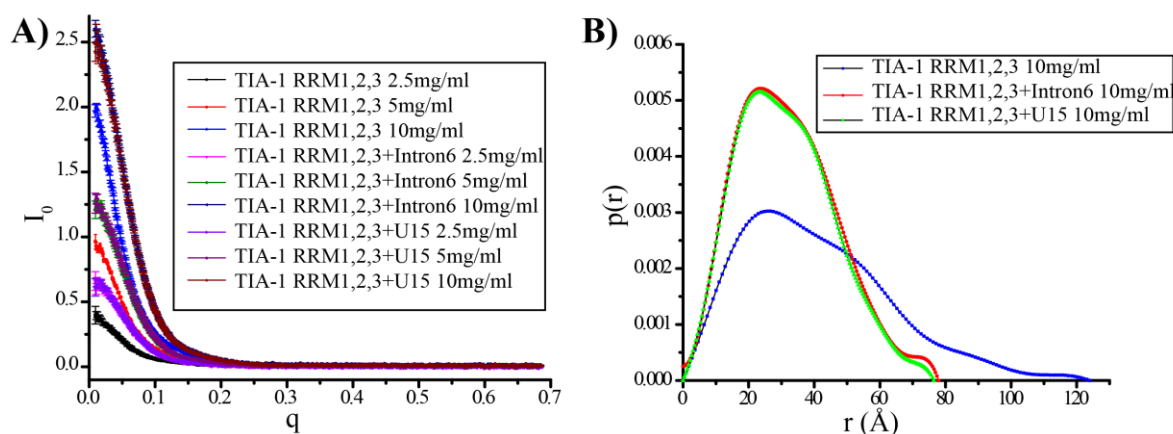


Figure 20. TIA-1 RRM1,2,3 form a compact shape in the presence of RNA

A) Intensity vs. q curve for the RRM1,2,3 with and without RNA at three different concentrations.
 B) $p(r)$ distribution of the RRM1,2,3 with and without RNA. The protein obtains a compact shape in the presence of RNA.

Upon RNA binding the radius of gyration, R_g of the free RRM1,2,3 domain decreases from $31.46 \pm 0.46 \text{ \AA}$ to $24.8 \pm 0.28 \text{ \AA}$ and $24.4 \pm 0.79 \text{ \AA}$ respectively for U15 and *Fas* intron6 bound protein based on the Guinier approximation. The maximum pairwise distance, D_{max} , decreases from 123.9 \AA in the absence of RNA to 77.8 \AA and 73.4 \AA upon RNA binding to TIA-1 RRM1,2,3 (**Figure 20B**) (SAXS statistics according to Jacques et al. 2012) are listed in **Table 2**. The larger R_g for TIA-1 RRM1,2,3 in the absence of RNA confirms that the three RRM domains tumble freely in solution and have no fixed orientation towards each other. The averaged ensemble is therefore larger in diameter than in the presence of RNA, where the R_g decreases substantially.

Table 2. SAXS data collection and processing statistics for RRM1,2,3 free and RNA complexes

	TIA-1 RRM1,2,3 free	TIA-1 RRM2,3 + U15	TIA-1RRM2,3 + <i>Fas</i> intron6
Data-collection			
Instrument	Rigaku BIOSAXS1000	Rigaku BIOSAXS1000	Rigaku BIOSAXS1000
Beam geometry	10 mm slit	10 mm slit	10 mm slit
Wavelength (Å)	1.5	1.5	1.5
q range (Å ⁻¹)	0.004-0.45	0.004-0.45	0.004-0.45
Exposure time (s) ^a	900	900	900
Concentration range (mg ml ⁻¹)	2.5-10	2.5-10	2.5-10
Temperature (K)	298	298	298
Structural parameters^b			
$I_{(0)}$ (cm ⁻¹) [from P(r)]	2.05 ± 0.0	2.6 ± 0.0	2.5 ± 0.0
R_g (Å) [from P(r)]	32.91 ± 0.01	24.8 ± 0.01	24.2 ± 0.00
$I_{(0)}$ (cm ⁻¹) [from Guinier]	2.03 ± 0.007	2.61 ± 0.007	2.53 ± 0.01
R_g (Å) [from Guinier]	31.46 ± 0.46	24.8 ± 0.28	24.4 ± 0.79
D_{max} (Å)	123.9	77.8	73.4
Porod volume estimate (Å ³)	58232.6	41735.6	40623.1
Software employed			
Primary data reduction	Rigaku SAXSLab v 3.0.1r1	Rigaku SAXSLab v 3.0.1r1	Rigaku SAXSLab v 3.0.1r1
Data processing	PRIMUS	PRIMUS	PRIMUS

^a 8 frames were recorded for each sample, ^b reported for a 10 mg/ml measurement

4.2 NMR structure of TIA-1 RRM1 domain

In order to study the interaction between two proteins by NMR spectroscopy, it is highly desirable to have the atomic structure of at least one of the interacting partner. This enables mapping of the interaction interface on the surface of the protein and thus further enhances the understanding of the interaction at atomic detail.

For determining the NMR structure of TIA-1 RRM1 domain (1-92), the protein was expressed in ¹³C, ¹⁵N labelled M9 minimal medium as TEV cleavable thioredoxin tagged protein in *E.coli* BL21 (DE3) cells in high yields. The protein was purified as mentioned in Methods section. For structure determination, various NMR experiments including backbone and side-chain assignment experiments were recorded. 95.3% complete assignment for the amino acids could be obtained. Additionally, an expanded network of ¹H-¹H-NOE and significant number of long range NOEs were obtained which were assigned automatically using CYANA3.0. Together with TALOS+ derived torsion angle restraints, a good quality

structure was obtained which converged well during the structure calculation cycles. The structural statistics is shown is **Table 3**.

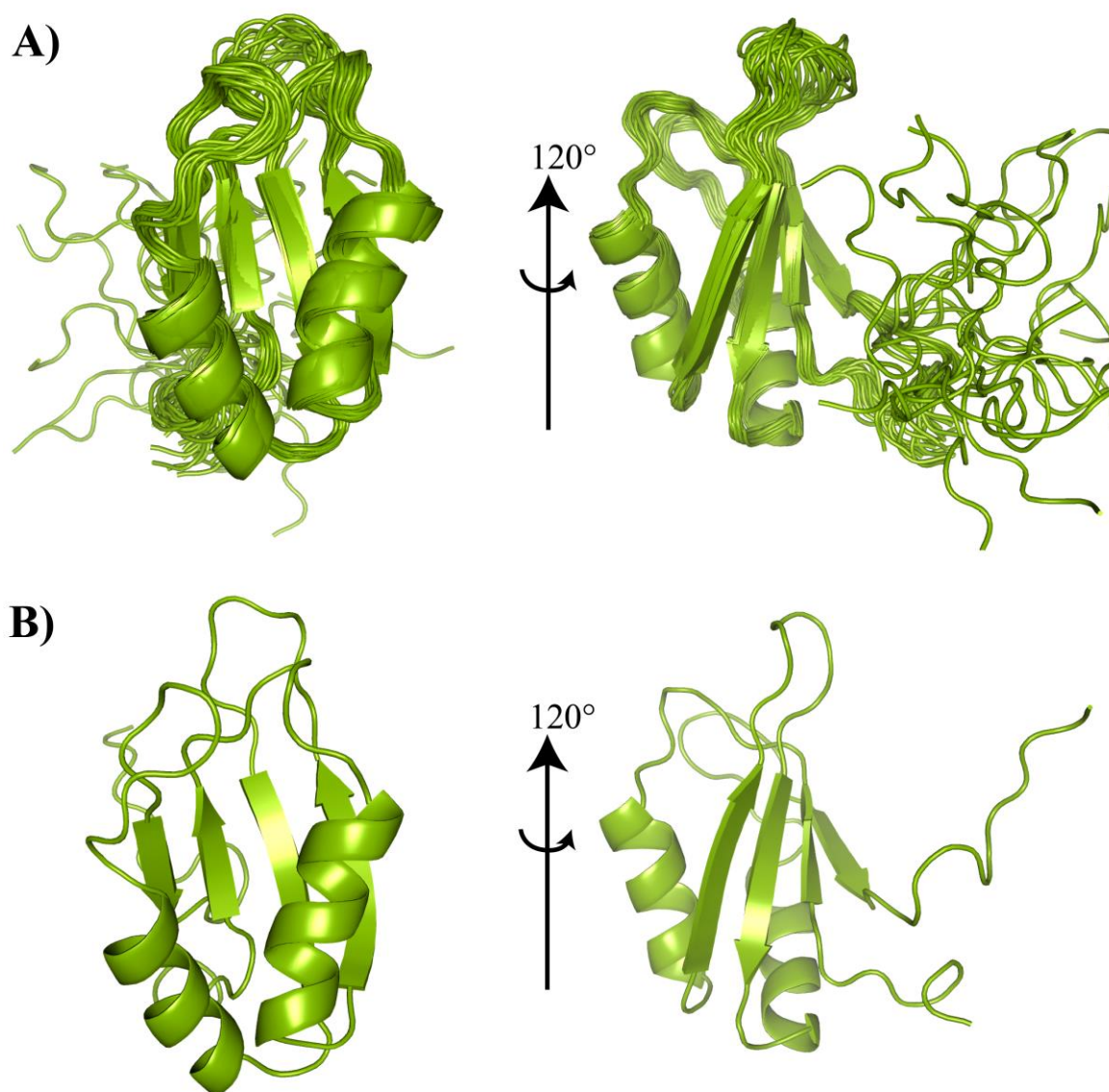


Figure 21. NMR structure of TIA-1 RRM1

Solution NMR structure of TIA-1 RRM1. A) Twenty lowest energy structures after water refinement are shown here. The structure is well converged with the formation of secondary structures except for N and C terminal regions which are flexible. B) A single lowest energy structure after water refinement is shown for clarity.

The ensemble of twenty lowest energy structures of TIA-1 RRM1 after water refinement with ARIA (Linge et al. 2003a) is shown in **Figure 21A** and for clarity, a single lowest energy structure is shown is **Figure 21B**. The whole structure converges well with an RMSD of 0.6 Å for backbone residues except the N and C terminal flexible regions (residue 1-

5 and 83-92) due to the absence of inter-residue NOEs. The RRM1 domain adopts the canonical RRM fold with $\beta\alpha\beta\beta\alpha\beta$ topology. It is composed of four antiparallel β -sheets that are covered on one side by two α -helices.

Table 3 Structural statistics for TIA-1 RRM1

Structural statistics for TIA1-RRM1^a	
<i>Structure calculation restraints</i>	
Distance restraints	
Total NOEs	981
Sequential ($ i-j = 1$)	564
Medium-range ($ i-j \leq 4$)	125
Long-range ($ i-j > 4$)	292
Hydrogen bonds	29
Dihedral restraints ($\phi+\psi$)	138
<i>Quality analysis</i>	
Restraints violations (mean \pm s.d.)	
Distance restraints (\AA)	0.054 \pm 0.021
Dihedral angle restraints ($^\circ$)	0.38 \pm 0.00
Deviation from idealized geometry	
Bond length (\AA)	1.024 \pm 0.001
Bond angles ($^\circ$)	0.329 \pm 0.006
Improper dihedral distribution ($^\circ$)	0.412 \pm 0.017
Average pairwise r.m.s. deviation (\AA) ^a	
Heavy	1.13 \pm 0.1
Backbone	0.49 \pm 0.08
Ramachandran values (%) ^{a,b}	
Most favored regions	91.6
Allowed regions	7.4
Generously allowed regions	0.2
Disallowed regions	0.8
WhatIf analysis ^{a,c}	
First generation packing	2.278 \pm 0.0.889
Second generation packing	5.655 \pm 1.711
Ramachandran plot appearance	-2.341 \pm 0.589
Chi-1/Chi-2 rotamer normality	-2.025 \pm 0.744
Backbone conformation	0.653 \pm 0.421

^a For residues 9–40, 47-81, ^b With Procheck., ^c Analyzed by iCING. Structure Z-scores, a positive number is better than average.

Nucleic acid binding proteins usually have basic isoelectric potential (pI) and positive surface charge potential. This is due to the presence of many surface exposed positively charged amino acids which could interact with the negatively charged nucleic acid bases. This is also true for RRM domains that bind to RNA. However, TIA-1 RRM1 shows a slightly acidic pI of 6.81 as calculated from ProtParam server (Wilkins et al. 1999).

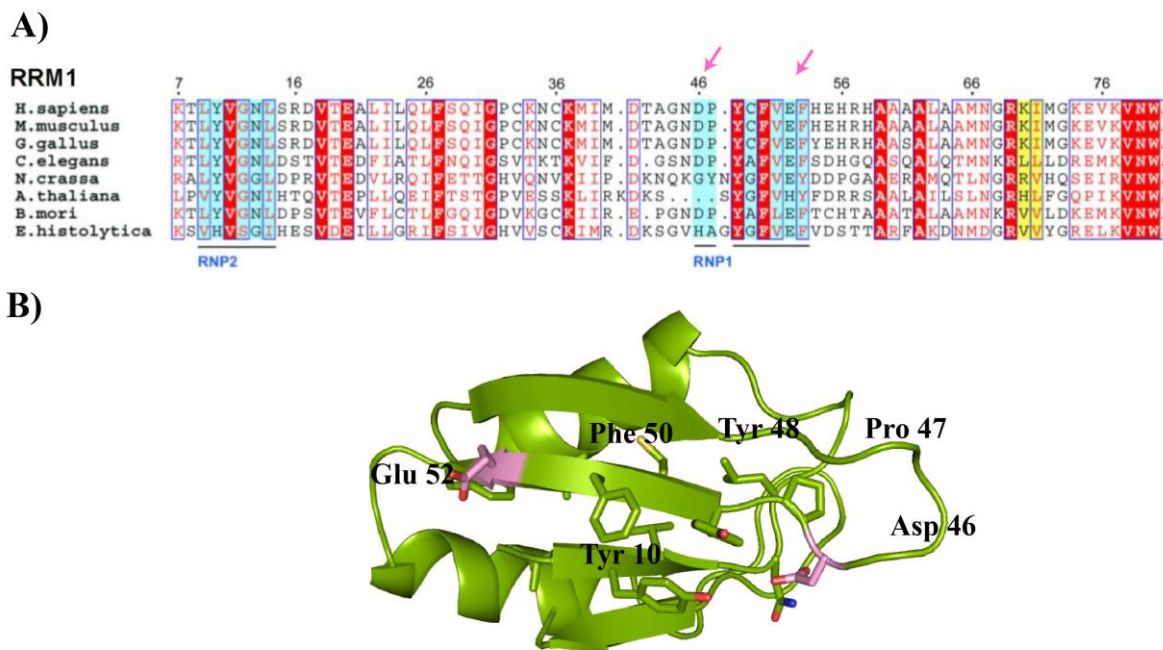


Figure 22. Structure analysis of TIA-1 RRM1

A) Sequence alignment of the TIA-1 RRM1 domain from different organisms is shown. Residues from the RNP1 and RNP2 are underlined and the negatively charged residues in the RNP1 are marked by pink arrows. B) The RNP1 and RNP2 residues are shown on the NMR structure of RRM1. The negatively charged residues are shown in pink.

The RRM1 structure clearly shows the position of the RNP residues on β 1- β 3 sheets. Tyr10 on the β 1 and Phe50 and Tyr48 on β 3 sheet maintain proper orientation for RNA interaction. However, the RNP1 motif on the β 3 sheet is interrupted by the presence of negatively charged residues Glu52 and Asp45, which are not expected to interact with the RNA. These negatively charged residues are conserved in the RRM1 domain of TIA-1 from different organisms (**Figure 22**). The presence of negatively charged residues in TIA-1 RRM1 is in agreement with the previous results that show that the RRM1 domain of TIA-1 alone has negligible affinity for RNA.

4.3 Concentration dependent dimerization of U1C

4.3.1 Backbone assignment of U1C (1-61)

Currently there are two structures of U1C protein in the PDB. One is determined by NMR spectroscopy (PDB id: 2VRD) and the other is a crystal structure as a part of U1 snRNP complex (PDB id: 4PJO). The NMR structure was determined by recording the NMR spectra

at natural abundance where the ^1H , ^1H TOCSY/COSY and ^1H , ^1H NOESY experiments were recorded at 1mM protein concentration. However, these experiments are not suitable for NMR titration experiments to study protein-protein interactions, as these spectra are very crowded with significant peak overlap.

The NMR structure of U1C was determined using homo-nuclear experiments, as the authors were unable to express the proteins in the labelled M9 medium (Muto et al. 2004). This was attributed to the toxicity of the protein. Hence, the protein was only expressed in the Terrific-broth rich medium. Accordingly, we also faced several challenges to express the protein in the ^{13}C , ^{15}N labelled medium. Nevertheless, we developed a protocol in close collaboration with the same authors (Dr. Kiyoshi Nagai's lab at LMB Cambridge) wherein the cells were first grown to high density in the LB medium and then transferred to the minimal ^{13}C , ^{15}N labelled M9 medium for protein expression. The protein was purified according to the previously published protocol (Muto et al. 2004). This resulted in protein sample suitable for acquiring NMR spectra for hetero-nuclear experiments.

The ^1H , ^{15}N HSQC spectrum showed uneven peak intensity signifying that the protein residues corresponding to these peaks are undergoing exchange broadening. This made it difficult to assign the backbone amides of all peaks in the spectrum as most of the peaks with weak intensities had the $\text{C}\alpha$, $\text{C}\beta$ carbons missing in the three dimensional heteronuclear spectra, making the sequential assignment difficult. Nevertheless, 39 backbone amide peaks could be unambiguously assigned whereas large stretch of residues could not be assigned unambiguously (**Figure 24**).

The residues missing in the assignment are present in proximity to the zinc finger region of the protein. Besides this, it was observed that the distribution of the peak intensity improved significantly with the decrease in the concentration of the protein. Therefore, this concentration dependent line broadening was attributed to the multimerization of the protein with increasing concentration. Attempts to overcome this effect with changing salt concentrations or pH were unsuccessful.

To confirm that the protein is not aggregated in the NMR buffer and has a proper molecular weight, static light scattering experiment (SLS) was carried out with the SLS detector attached to the size-exclusion chromatography column. The experiment showed a molecular weight of 8.8 kDa (**Figure 23**). This differs significantly from the calculated molecular weight of 7.4 kDa although it is quite away from the molecular weight of the dimer

protein. However, it should be noted that during size-exclusion chromatography, the protein is diluted and therefore, it seems that the slightly higher molecular weight of the protein is due to presence of minor concentration dependent aggregation.

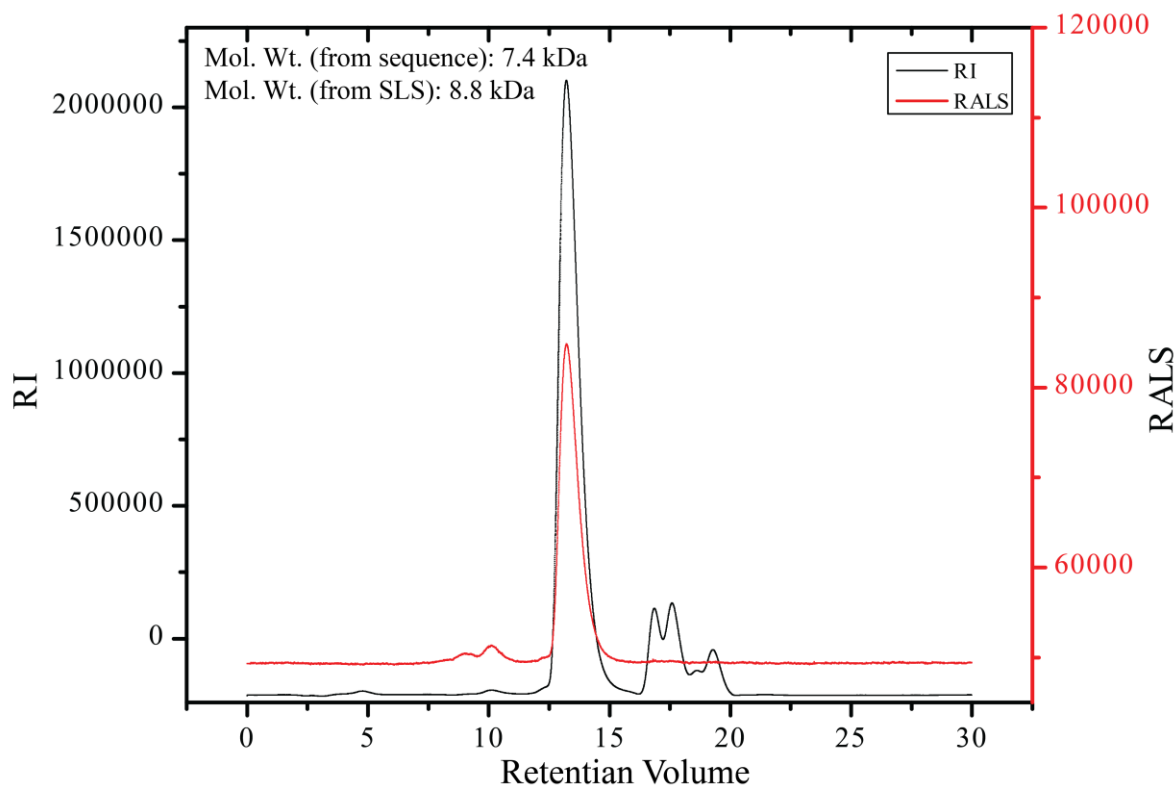


Figure 23. Static light scattering of U1C protein

Refractive Index (RI) and Right Angle Light Scattering (RALS) was recorded for U1C using static light experiment. The experimental molecular weight determined for the protein is 8.8 kDa whereas the calculated molecular weight from the protein sequence is 7.4 kDa

To understand the rigidity of the secondary structure of the protein, we recorded $^1\text{H},^{15}\text{N}$ heteronuclear NOE experiment. This experiment provides the information about the motion of individual N-H bond vectors. The amide bonds undergoing motion faster than the overall tumbling of the molecule show decreased NOE intensity relative to the average observed for the majority of the residues. It was clear from the experiment that the residues from the zinc finger region (1-30) along with the rest of the protein are not flexible except the last two residues from the C-terminal region that show negative NOE value and thus are flexible.

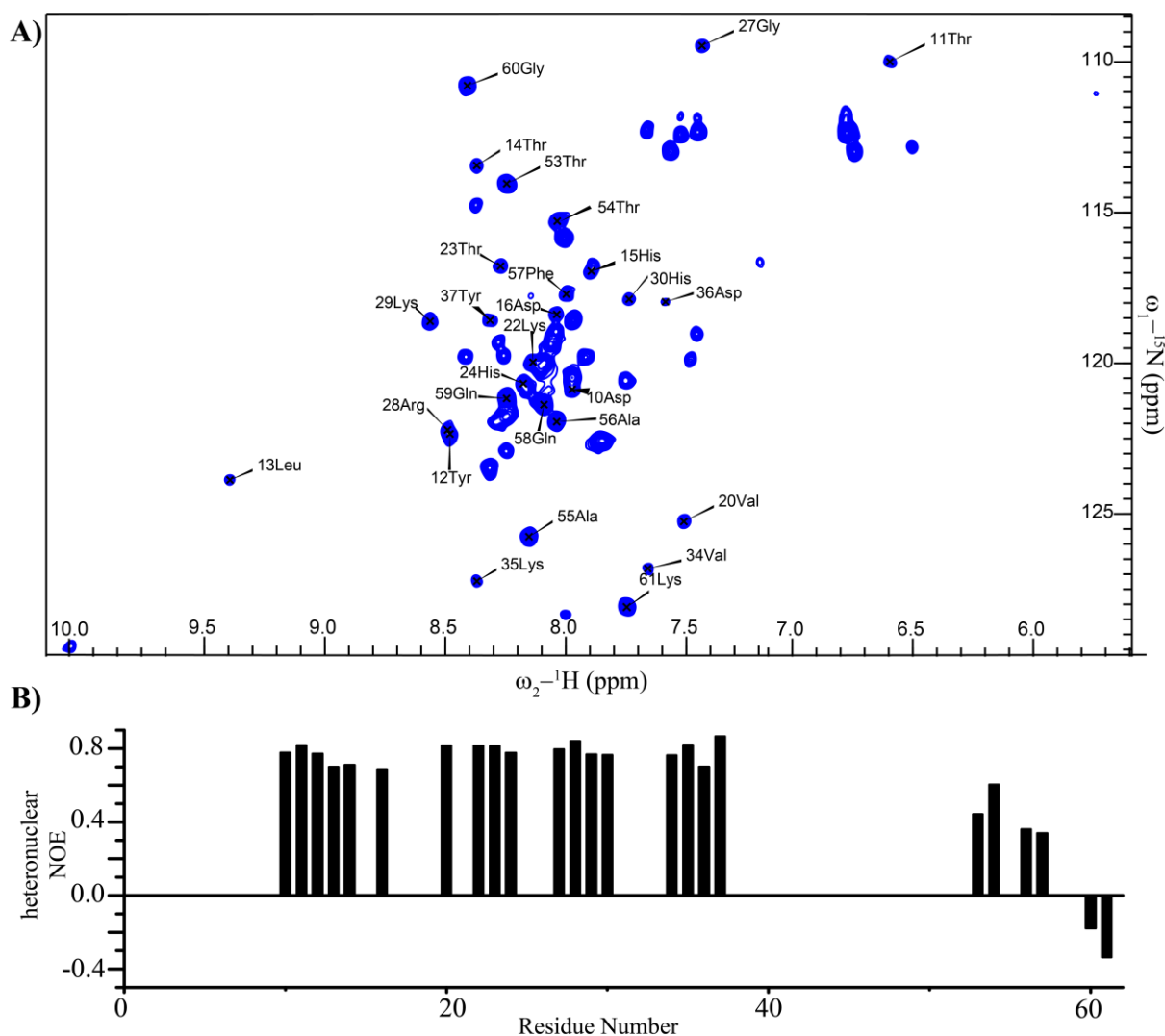


Figure 24. Backbone assignment of U1C (1-61).

A) ^1H - ^{15}N HSQC spectra for U1C (1-61) protein with the partially assigned backbone. B) Heteronuclear NOE data for the backbone amides of U1C showing the rigidity of the protein backbone.

Given that the NMR spectra for structure calculations were recorded at high concentrations of the protein (1mM; 7.4 mg/ml), the protein is expected to partially dimerize according to the results presented in this thesis. However, it is surprising that the NMR structure is monomeric in which helix B observed in the crystal structure is split into helix B and helix C and helix C folds back onto the helix B (**Figure 6D**).

4.3.2 SAXS analysis of U1C (1-61)

As the U1C protein showed concentration dependent changes in ^1H , ^{15}N HSQC spectrum indicative of multimerization of the protein, we decided to probe this further by

performing SAXS experiments. SAXS data were recorded for various concentrations of the protein ranging from 0.5 mg/ml to 9.6 mg/ml. The protein does not show concentration dependent aggregation behavior as deduced by the Intensity vs. angle of diffraction curve (**Figure 25A**) at low q range.

Table 4. SAXS data collection and data processing statistics for U1C (1-61)

Parameters	U1C (1-61);0.5mg/ml	U1C (1-61);9.6mg/ml
Data-collection		
Instrument	BioSAXS BM29 ESRF	BioSAXS BM29 ESRF
Beam geometry	10 mm slit	10 mm skit
Wavelength (Å)	1.008	1.008
q range (Å ⁻¹)	0.004-0.45	0.004-0.45
Exposure time (s) ^a	2	2
Temperature (K)	298	298
Structural parameters		
$I_{(0)}$ (cm ⁻¹) [from P(r)]	12.82 ± 0.0	24.43 ± 0.00
R_g (Å) [from P(r)]	19.4 ± 0.00	25.3 ± 0.02
$I_{(0)}$ (cm ⁻¹) [from Guinier]	12.73 ± 0.023	24.21 ± 0.022
R_g (Å) [from Guinier]	18.5 ± 0.06	24.0 ± 0.03
D_{max} (nm)	7.47	11.47
Porod volume estimate (Å ³)	12240	27410
Software employed		
Primary data reduction	BsxCuBE	BsxCuBE
Data processing	PRIMUS	PRIMUS

^a 20 frames were recorded for each sample

The pair-distance distribution function (P(r)) describes the paired-set of all the distances between points within the same object. A P(r) function in SAXS describes the paired set of distances between all the electrons in the macromolecular structure. Therefore, changes in the position of few residues can be easily visualized by the changes in the P(r) function of the protein. The P(r) function plotted at various concentrations for the U1C protein showed a concentration dependent increase in D_{max} of the protein. D_{max} value describes the maximum dimension of the protein. The D_{max} values for the 0.5, 1, 2, 3, 5, 7 and 9.6 mg/ml of the protein concentrations were 7.47, 8.34, 8.28, 10.5, 10.9, 11.33 and 11.9 nm respectively. This increase in dimension suggests an increase in the size of the protein molecule most probably due to the formation of multimeric states.

By plotting the I_0 values (derived from the intensity vs. q curve) and radius of gyration (R_g) vs. the protein concentrations, a sigmoidal curve could be obtained for the U1C protein,

which became saturated at high protein concentrations used in the study. The curve could be fitted with the 4-parameter logistic nonlinear regression model thus providing the dissociation constant. The K_D values thus derived from the two curves were respectively 4.3 mg/ml (581 μ M) and 4.4 mg/ml (594 μ M).

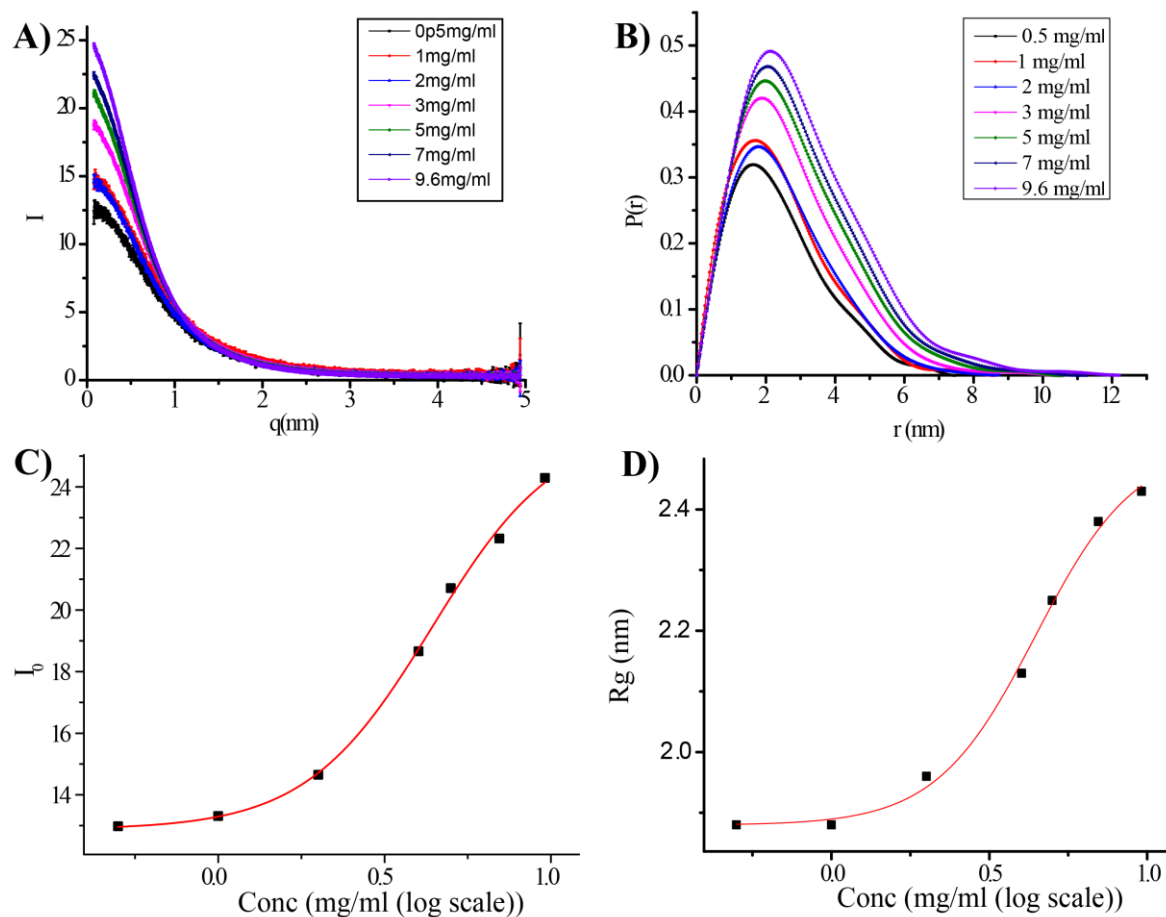


Figure 25. Concentration dependent dimerization of U1C

SAXS data analysis of U1C (1-61) protein is shown. A) Intensity vs. q curve at various concentrations of the U1C protein. B) $P(r)$ distribution function showing the changes in D_{max} with increase in the concentration of the protein. (C) and (D) show the sigmoidal fitting of the I_0 and R_g values vs. protein concentration to calculate the dimerization constant of the protein.

It has been shown before that U1C forms homodimers *in vitro* as well as in the yeast two hybrid system (Gunnawiek et al. 1995). It was demonstrated that the first 30 residues comprising the zinc finger in the U1C protein are required for protein dimerization. In addition, the dimers were shown to be due to non-covalent interactions and were not cysteine mediated. Therefore, it can be concluded that the higher oligomeric entity observed from the SAXS data

as well as concentration dependent changes in the NMR spectrum shows the concentration dependent dimer formation of the U1C protein.

There are two atomic structures of the U1C (1-61) protein derived from two different methods, NMR (PDB id:2VRD) and crystallography (PDB id:4PJO; as part of U1 snRNP complex) and both the structures differ significantly (**Figure 6D**). To confirm which structure agrees well with the SAXS data and thus represents the solution state more closely, we compared the SAXS curves obtained from each of the structures to the experimental SAXS curve. For this purpose, the SAXS curve obtained at 0.5 mg/ml was used, as this is the lowest concentration data that was possible to record with the synchrotron radiation with acceptable noise.

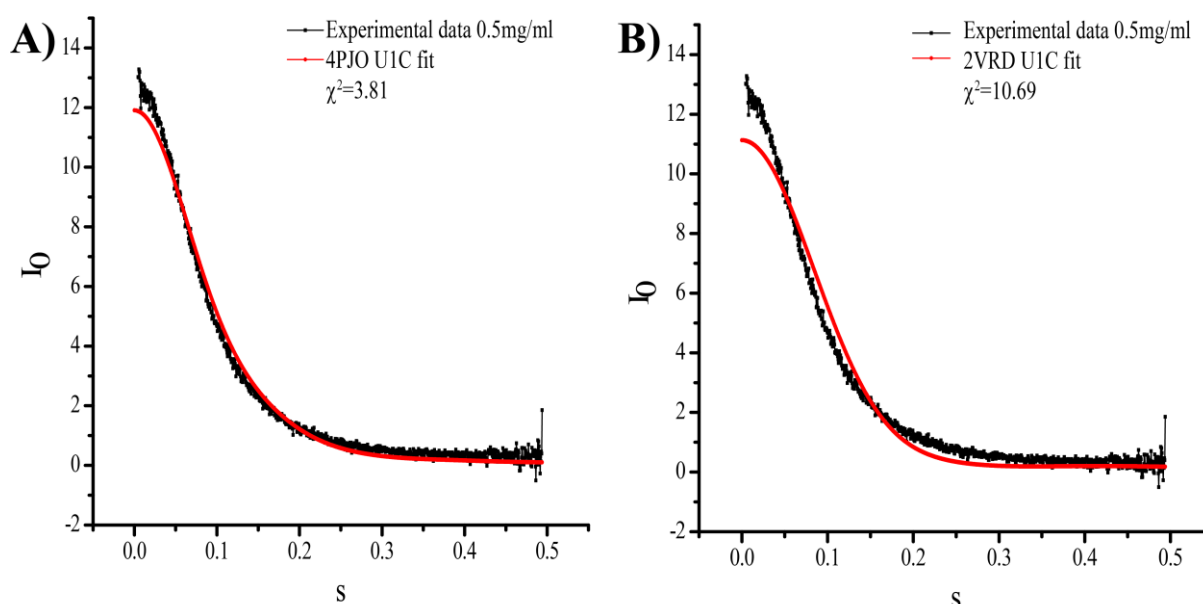


Figure 26. U1C is elongated in solution

SAXS curve superposition of U1C experimental SAXS data and curve calculated from the structures in PDB. A) The calculated SAXS curve of U1C crystal structure (Residue 1-54) (PDB id: 4PJO) superposes well with the experimental SAXS data for U1C (1-61) whereas B) the fit for the NMR structure (PDB id: 2VRD) is quite poor.

As it is clear from **Figure 26A**, structure of U1C from the crystal structure of U1 snRNP complex agrees quite well with the SAXS data with a χ^2 of 3.81. It should be noted that in the crystal structure of U1 snRNP, only 1-54 residues are visible for the U1C protein whereas the SAXS data is recorded for U1C (1-61) protein. Thus, the differences in the SAXS curves observed at low q range most probably arise due to the missing seven residues in the crystal structure of U1C and not due to a different shape of U1C in solution and the crystal structure.

On the other hand, the SAXS curve generated from the NMR structure seems to deviate substantially from the experimental SAXS curve ($\chi^2=10.69$). In NMR structure, the C-terminal helix of the protein folds back on the zinc finger however, in the crystal structure it is extended. Therefore, from the SAXS data, the crystal structure and not the NMR structure seems to represent the real solution structure of U1C.

4.3.3 ITC experiments to study U1C dimerization

In order to understand the thermodynamics of the U1C dimerization, dilution ITC experiments were carried out by titrating concentrated U1C protein (1.3 mM) in the syringe into the ITC cell containing the protein buffer (20 mM MES pH 6.5, 100 mM NaCl, 50 mM MgSO_4). This gave rise to endothermic heat pulses because of dimer dissociation (**Figure 27**).

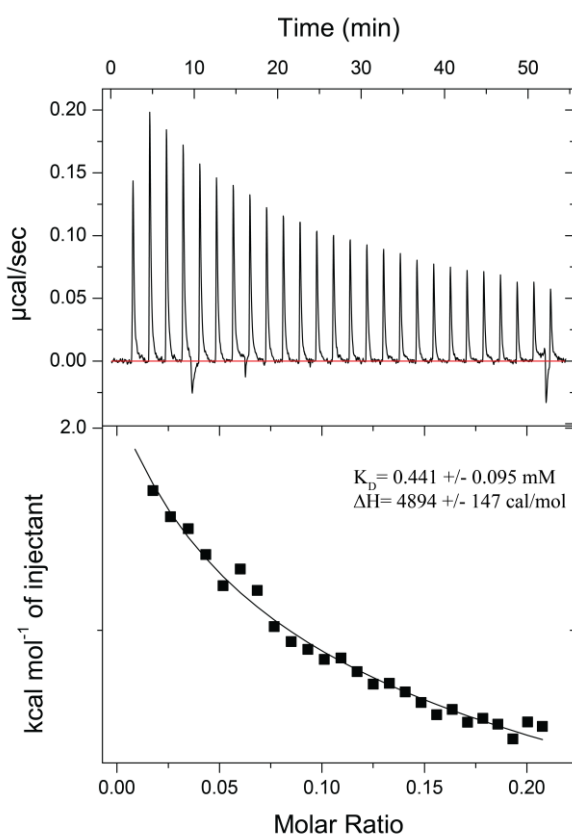


Figure 27. U1C dimerization constant determined by ITC

The dimerization constant was determined by titrating concentrated solution of U1C into the buffer. The dimerization K_D is 441 μM agreeing well with the dissociation constant determined from SAXS.

As the data gave a good fit with a dimer dissociation model giving a hyperbolic curve, it further agrees with the concentration dependent dimerization of U1C. The calculated dimerization constant from ITC experiments (441 μM) fits well with that determined from

SAXS experiments ($\sim 581 \mu\text{M}$). As the enthalpy of dissociation is positive ($4894 \pm 147 \text{ cal/mol}$), it seems that the dimer interaction is mostly mediated by hydrophobic interactions.

4.4 Interaction between U1C and RRM1

TIA-1 was shown before to interact with the U1C protein of U1 snRNP complex (Forch et al. 2002). The interaction itself was deduced from the presence of Nam8p as a stable component of U1 snRNP in yeast (Gottschalk et al. 1998) and Nam8p was shown to bind the sequences downstream of U1 snRNP binding site and also modulate U1 snRNP binding at the 5' ss (Puig et al. 1999). Nam8p is a close homolog of the protein TIA-1 in humans.

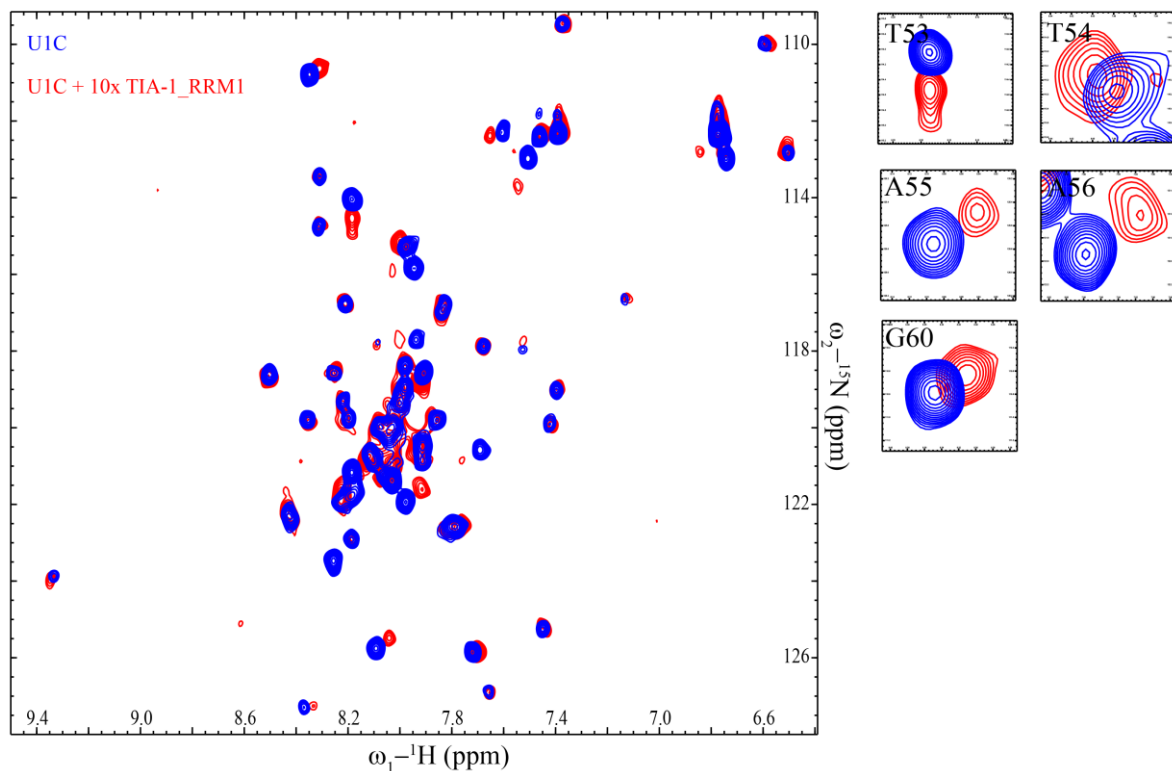


Figure 28. U1C interacts with TIA-1 RRM1

^1H - ^{15}N HSQC titration of U1C with RRM1 domain. Zoom in of residues showing significant chemical shift changes are shown on the right.

The interaction of TIA-1 and U1C was shown to be mediated by Q-rich domain of TIA-1 and first 61 residues of U1C that also consists of a zinc finger. This interaction was independent of the ability of TIA-1 to bind RNA and hence was inferred as the result of protein-protein interaction. In addition to the interaction of Q-rich domain with U1C, it was shown that the interaction is strengthened by RRM1 domain.

As the U1C protein was successfully isotope labelled and a partial backbone assignment was obtained, we next studied the interaction of the U1C protein with the RRM1 domain of TIA-1. The concentration of U1C used to record the NMR spectrum was 25 μM (0.185 mg/ml) to avoid formation of U1C dimers in the NMR sample. RRM1 was titrated at high concentrations (10-fold excess) as the interaction is expected to be weak.

Titration of RRM1 in the U1C at tenfold excess concentration showed significant shifts in the ^1H - ^{15}N HSQC spectrum (**Figure 28**). Besides this, the peak shifts were in fast exchange, an observation consistent with the expected weak interaction of the two proteins in the absence of full-length TIA-1 protein and RNA. Titration of more than 10-fold excess of RRM1 in the U1C protein showed excessive exchange line broadening probably due to the unfavorable rates of association/dissociation on the NMR time scale. Upon titration of RRM1 in U1C, only residues from C-terminal helix showed chemical shift perturbations. There were no significant changes in the zinc finger region of the protein.

In order to confirm that this interaction is conserved in the TIA-1 protein which includes Q-rich domain, a construct of TIA-1 (TIA1 (1-319)) was titrated into the U1C protein at 4.5-fold excess (**Figure 29**). This was the highest concentration of TIA-1 protein that could be titrated into U1C as TIA-1 forms protein fibrils at high protein concentrations due to the presence of Q-rich domain. As for the U1C-RRM1 NMR titrations, titration of TIA-1 in U1C also showed similar shifts limited to the residues from the C-terminal helix of the U1C protein, thus confirming that this interaction is conserved in the full-length TIA-1 protein.

From the NMR titrations, it was clear that the residues Thr53, Thr54, Ala55, Ala56 and Gly60 from the C-terminal helix interact with the RRM1 domain. Based on these results, a chemically synthesized U1C peptide from 47-61 (QSLIDKTTAAFQQGK) region was titrated into the TIA-1 RRM1 domain. However, there were no significant CSPs even at high concentrations of the peptide (20-fold excess titration; data not shown). This was surprising given that the region from 53-61 residues in U1C protein show clear CSPs on titration with RRM1/TIA-1 domain.

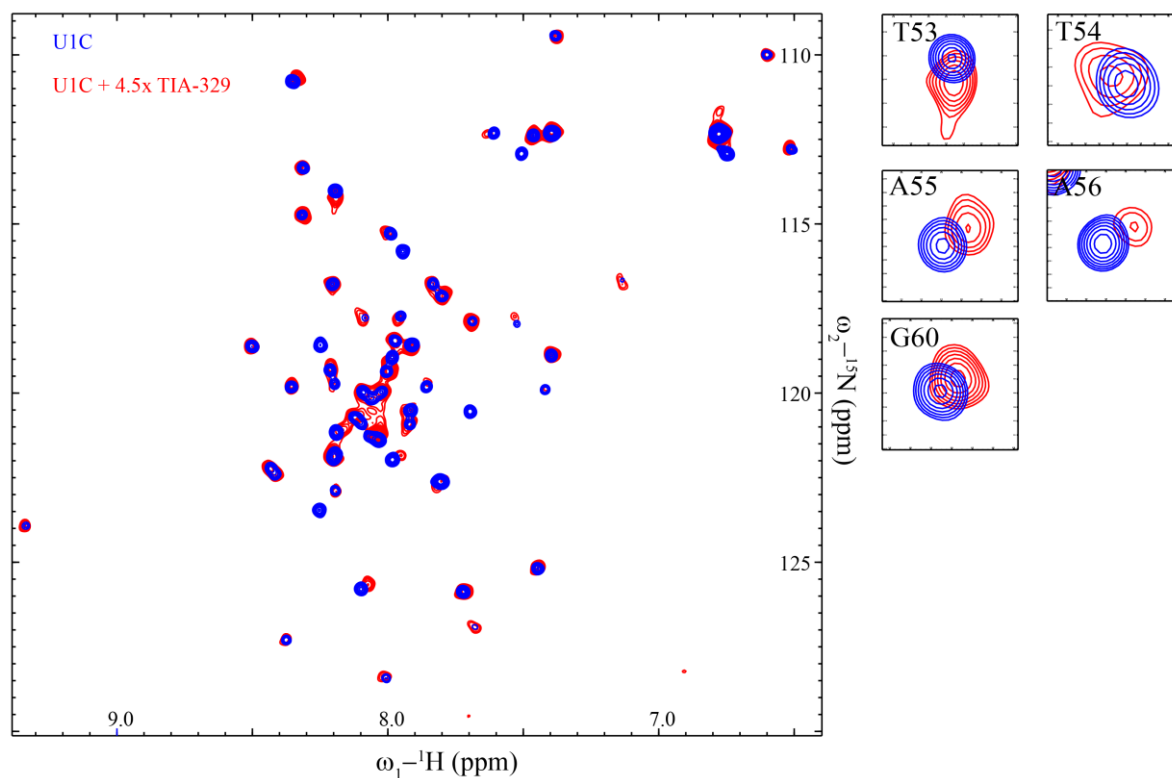


Figure 29. U1C interacts with TIA-1

^1H - ^{15}N HSQC titration of U1C with TIA-1 (1-329). Zoom in of residues showing significant chemical shift changes are shown on the right.

However, as the residues preceding Thr53 are not assigned, it was not possible to track the chemical shifts changes occurring in the helical region preceding Thr53. Therefore, it cannot be negated that the whole of the C-terminal helix region of U1C (i.e. from 30-61) could participate in the interaction of U1C with RRM1/TIA-1.

It is also possible that the interaction will be much stronger in the presence of RNA when U1 snRNP binds at the 5' ss along with TIA-1 binding downstream at the uridine-rich sequences. In this scenario, the two proteins will already be in close proximity and this will further enhance the protein-protein interaction between U1C and TIA-1 due to increase in the local concentration of the two proteins.

During the NMR titrations of the two proteins, it was noticed that significant number of peaks showed line broadening at very high concentration of the other binding partner. Therefore, it was not deemed possible to get inter molecular NOEs between RRM1 and U1C for structure calculation attempts. In addition, it was not possible to track the shifts of the peaks during titration at high concentrations of the protein as the peaks disappeared thus making it

difficult to plot the CSPs for individual residues against protein concentration and calculate the binding affinity.

4.4.1 Backbone assignment of U1C 30-61

To test the hypothesis that the whole C-terminal helix of U1C (residues 30-61) interacts with TIA-1, the region was cloned into a modified pET 24d vector with cleavable His-thioredoxin expression tag. The protein was expressed in high levels in *E.coli* BL21 (DE3) cells and purified by standard protein purification method with Nickel affinity chromatography followed by TEV cleavage of the expression tag and then size exclusion chromatography.

For backbone assignment, the protein was expressed in ^{13}C , ^{15}N labelled M9 medium. The ^1H , ^{15}N HSQC spectrum of the peptide shows all the backbone resonances spread between 7.5 to 8.5 ppm consistent with the shifts for residues in α -helical/ unstructured region (**Figure 30**). Besides this, the spectrum shows even distribution of peak intensities depicting that there is no aggregation of the peptide. The backbone was assigned with standard NMR backbone assignment experiments. All backbone amides could be assigned unambiguously except for Lys40.

As this region of U1C is present in the crystal structure as a rigid long helix, to confirm that in isolation also it forms a stable rigid structure and is not completely flexible, heteronuclear NOE experiments were recorded. This experiment shows that in isolation also much of the peptide maintains its rigidity except at the N and C termini.

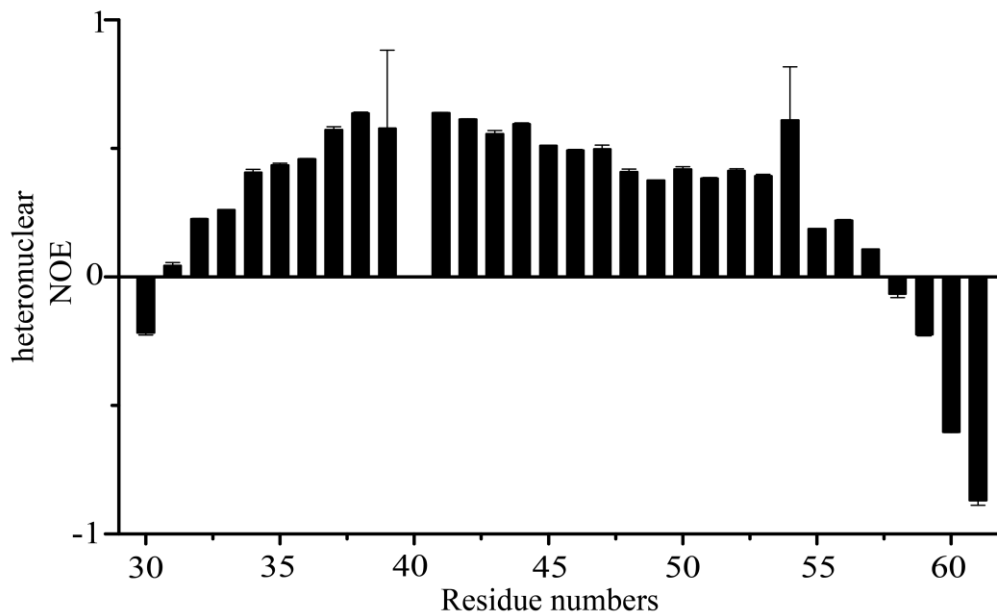
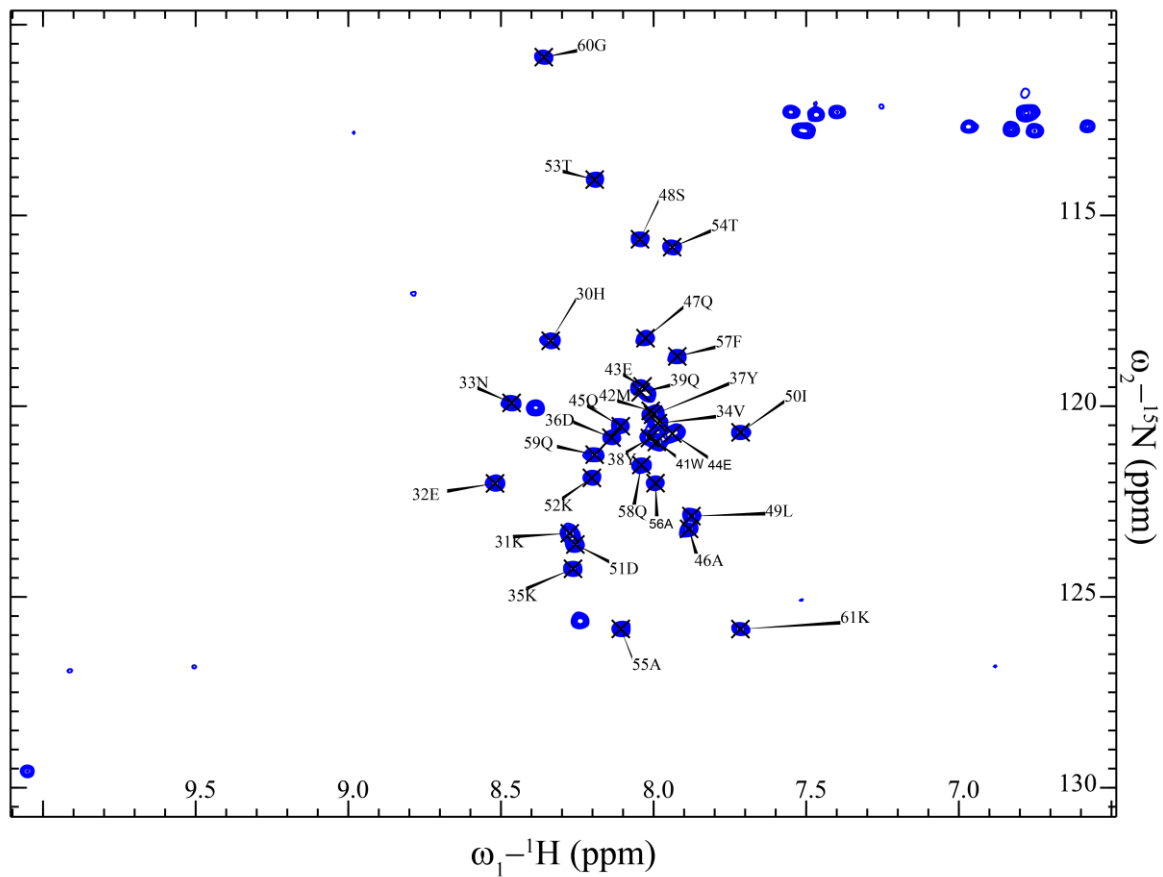


Figure 30. Backbone assignment of U1C 30-61 peptide

A) ^1H - ^{15}N HSQC spectrum of U1C 30-61 peptide long with the assigned backbone. B) ^1H - ^{15}N heteronuclear NOE data depicting only the N and C terminal of the peptide are flexible.

4.4.2 Interaction of U1C 30-61 and TIA-1 RRM1

After the assignment of the U1C 30-61 peptide, to study its interaction with TIA-1 RRM1 domain, the RRM1 domain was titrated at 16-fold excess concentration into the ^1H - ^{15}N labelled U1C 30-61 peptide. The peptide showed significant chemical shift perturbations upon addition of the RRM1 domain confirming that indeed the whole C-terminal helix region is required for interaction of U1C with the TIA-1 RRM1 domain.

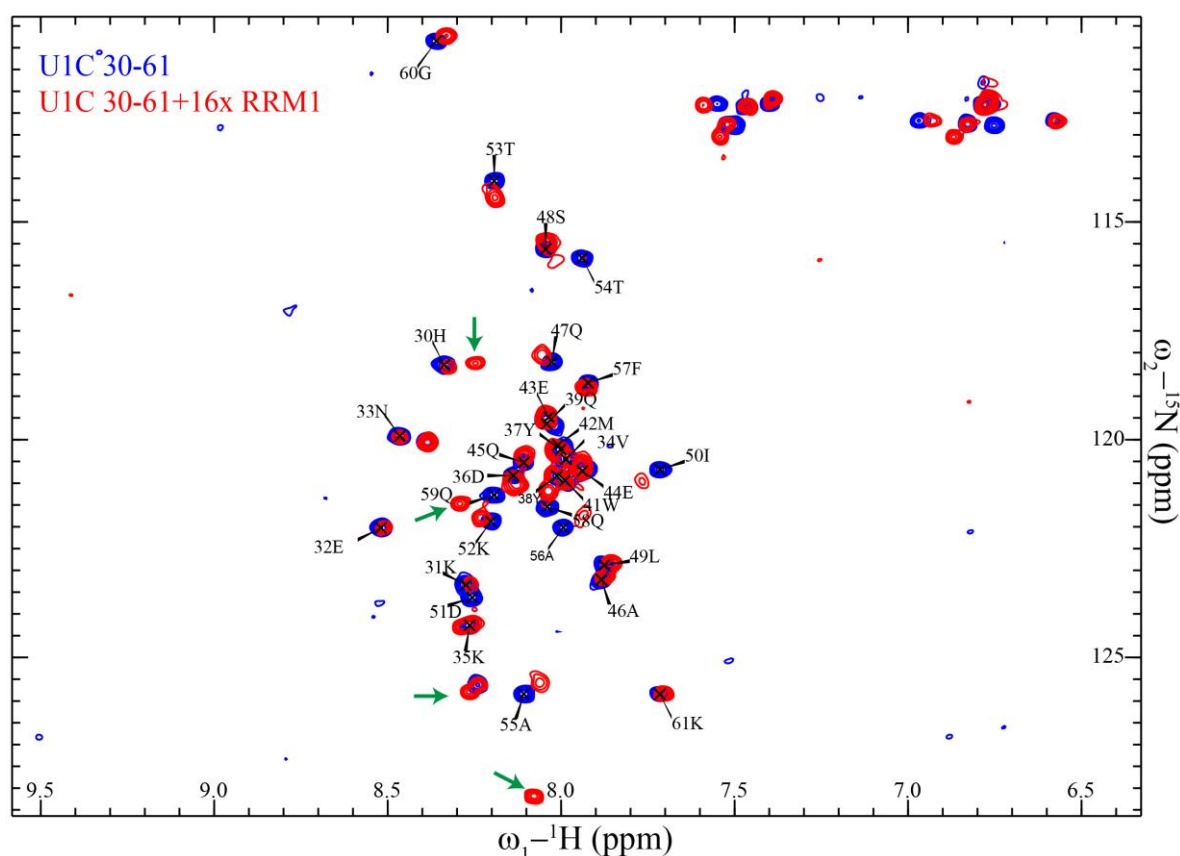


Figure 31. Interaction between U1C 30-61 peptide and TIA-1 RRM1

A single point titration of 16-fold excess of unlabeled RRM1 into ^1H , ^{15}N labelled U1C 30-61 peptide is shown. The newly appearing peak resonances are shown with green arrows.

Besides the CSPs, appearance of many new peak resonances was also observed upon titration of the RRM1 domain into the U1C peptide (**Figure 31**). The appearance of new peaks corresponded with the disappearance of the peaks for His30, Gln39, Ile50, Asp51 and Thr54. As the new peak resonances probably represent the disappearing resonances at other position, this suggests that these residues probably undergo significant structural changes on

binding of U1C 30-61 peptide to RRM1 domain. However, in the absence of the assignment for the newly appeared resonances in the RRM1 bound form of U1C peptide, it cannot be concluded unambiguously which disappeared peak corresponds to which of the new resonances.

NMR chemical shift perturbations provide a very useful tool to understand the residues in the protein that undergo changes in the chemical environment upon titration with the binding partner. In most of the cases these CSPs could not only be used to map the surface on the protein which is involved in the interaction, but also could be used to determine the affinity of the interaction by following the chemical shift/peak intensity changes with the titrated ligand ratio. However, the caveat of the method lies in the fact that the residues that are affected allosterically upon binding to the ligand partner will also show CSPs. These CSPs could very well be present remote from the actual binding site due to allosteric effects and thus could be misleading in the absence of convergence of the binding surface upon mapping of the CSPs.

In the case of U1C 30-61/RRM1 titrations where unlabeled RRM1 domain was titrated into the ^1H - ^{15}N labelled U1C peptide, mapping of the CSPs on the surface of the helix is not likely to provide much useful data. This is because the U1C peptide is expected to be a 30 residue long α -helix and almost all of the residues from the peptide show significant structural changes upon titration with RRM1 domain. Therefore, mapping of these CSPs on the helix surface would highlight whole of the helix instead of converging on a single small region.

Therefore, to determine the affinity of the interaction and to be able to map the binding interface on one of the protein partner, we titrated the unlabeled U1C 30-61 peptide into the ^1H , ^{15}N labelled RRM1 domain during NMR ^1H , ^{15}N HSQC titrations. NMR titrations showed shifting of various peaks with the saturation reaching at 20-fold excess of the U1C peptide. This reflected the low affinity of the interaction as has been observed in the previous experiments. Besides this, the spectral quality also deteriorated significantly at high concentrations of the peptide probably owing to the unfavorable exchange of the interaction on the NMR time scale.

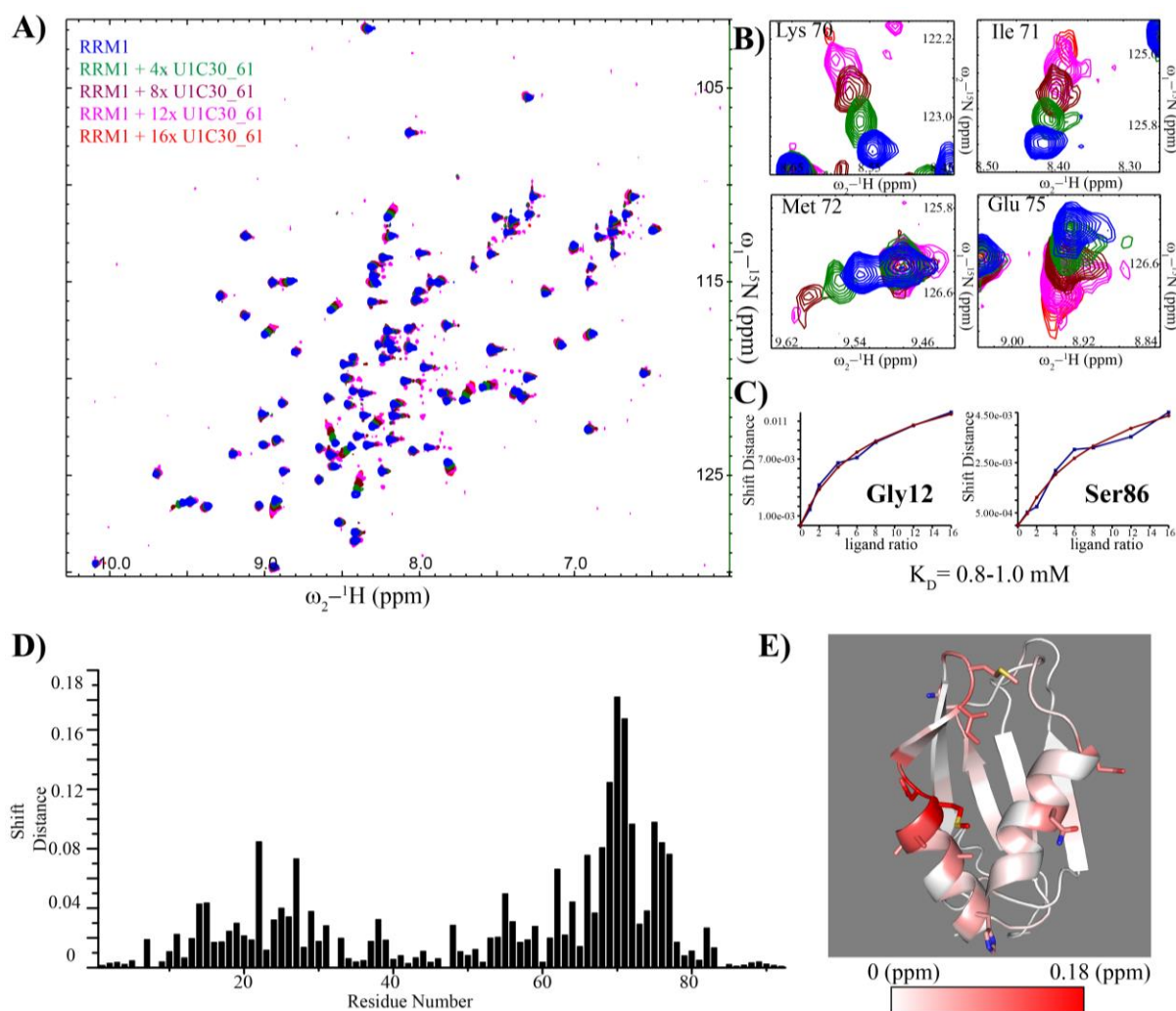


Figure 32. Binding affinity of RRM1-U1C interaction

Binding affinity of RRM1-U1C interaction was determined. A) ^1H , ^{15}N labelled RRM1 was titrated with 2x, 4x, 8x, 12x, 16x and 20x unlabeled U1C 30-61 peptide. For clarity, only four titration points are shown in the ^1H , ^{15}N HSQC spectrum. B) Peaks for Lys70, Ile71, Met72 and Glu75, which show significant shifts during the titration, are shown. C) Fitting of the NMR shift differences to determine the binding affinity of the peptide. The affinity from NMR titration is around 0.8 mM to 1 mM. D) Chemical shift perturbations are shown for each residue of the RRM1 domain on titration with the highest tested concentration of U1C 30-61 peptide (20x). E) The CSPs were mapped on the NMR structure of RRM1 domain. They concentrate on the two α -helices of the RRM1 domain on the surface opposite to the β -sheets.

The shifts in the peak positions were plotted as a function of the ligand ratio and the data was fitted using nonlinear curve fitting to obtain the binding constant for the interaction. The binding affinity of the interaction was around 0.8-1.0 mM, which is very weak as expected from the previous experiments mentioned above. The shifts could also be plotted against the residue number and on the NMR structure of RRM1 domain to visualize the binding interface

of this interaction. Majority of the CSPs map to the two α -helices of the RRM1 domain with Phe27, Lys70, Ile71, Met72 and Glu75 showing the highest shifts. Therefore, these residues seem to be involved in direct interaction with the U1C protein.

4.5 Structure of RRM1-U1C complex

4.5.1 Linking TIA-1 RRM1 and U1C 30-61 peptide with GS linker for structural studies

As the affinity of the RRM1-U1C peptide interaction is very weak, the two proteins are not expected to form a stable complex making it unsuitable for structural studies using crystallography. We therefore attempted to record NMR data to determine the NMR structure of the complex. In order to achieve this, it is very important to obtain NOEs between the two proteins (intermolecular NOEs). For this purpose, RRM1 was ^{15}N , ^{13}C labelled whereas unlabeled peptide at a concentration of 20-fold excess was titrated into it to record ^{15}N , ^{13}C edited NOE experiment with $\omega 1$ filter. This experiment allows studying intermolecular NOEs between the protein and the ligand. To obtain sufficient signal to noise in this experiment, the concentration of the RRM1 domain was kept at 300 μM and the U1C 30-61 peptide was kept at 4.8 mM (16-fold excess). However, the ^1H , ^{15}N HSQC spectrum obtained with this sample was not superimposable with that obtained with a 10-fold lower concentrated sample (50 μM RRM1 and 1 mM U1C 30-61 peptide). This could very well be attributed to the dimerization/aggregation of the U1C peptide at high concentrations (at 4.8 mM). Therefore, this approach was not pursued further in order to avoid complexities with data analysis.

In absence of any suitable alternative to obtain structural information at atomic level for the complex of RRM1-U1C peptide, the C-terminal of RRM1 domain was fused to the N terminal of the U1C 30-61 peptide to artificially increase the affinity of the proteins for each other. The proteins were linked with 10 and 15 residue long Gly-Ser linkers. A rough model of RRM1-U1C 30-61 peptide was made in Chimera software (Pettersen et al. 2004) by moving the two proteins manually and bringing them in proximity based on the CSPs obtained in the NMR titrations earlier. The minimum length of 10 residues long GS linker was decided based on the maximum possible distance between the C-termini of RRM1 and N-termini of U1C peptide. Each amino acid in the GS linker was assumed to cover a distance of 3.6Å considering the unstructured nature of the GS linkers.

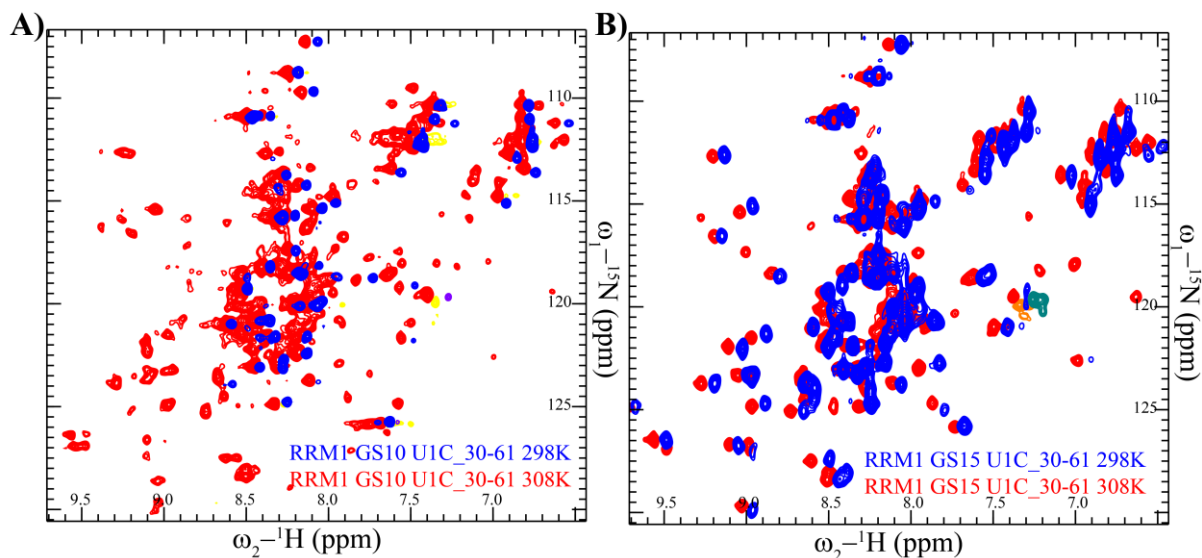


Figure 33. Optimization of GS linker length between RRM1 and U1C peptide

A) ^1H , ^{15}N HSQC for RRM1-GS10-U1C 30-61 construct and B) RRM1-GS15-U1C 30-61 construct at 295K (blue) and 308K (red).

The two proteins connected with the 10/15 residue long GS linkers were cloned by overlapping PCR with the required GS linker codons introduced in the primers and were expressed in BL21 (DE3) cells in ^1H , ^{15}N labelled M9 medium. ^1H , ^{15}N HSQC spectra for the two proteins were recorded at 298 K and 308 K. RRM1-U1C 30-61 protein fused with 10-residue long GS linker showed very few peaks at 298 K with many new peaks appearing at 308 K. However, a number of the newly appeared peaks at 308 K showed significant exchange broadening. This was probably due to insufficient GS linker length in the 10-residue long linker construct where in the U1C peptide is not able to make optimal contacts with the RRM1 domain. On the other hand, RRM-U1C 30-61 protein construct fused with 15-residue long GS linker showed ^1H , ^{15}N HSQC spectrum expected for a properly folded protein. However, the number of peaks observed in this spectrum at 298 K were still less compared to the number of backbone amides in the protein. At 308 K, many new peaks appeared thus signifying that the 15-residue long linker is long enough so that the U1C 30-61 peptide could make optimal contacts with the RRM1 domain.

4.5.2 Crystal structure of RRM1-GS15-U1C30-61

Since the ^1H , ^{15}N HSQC spectrum of RRM1-GS15-U1C30-61 protein indicated that the protein is well folded (with optimal Gly-Ser linker length), it was chosen for setting up sparse

matrix crystallization screens at room temperature and 4 °C. Rod shaped crystals were observed after 12 days in 0.1 M HEPES pH 7.5, 10% PEG 6000 and 5% MPD at room temperature. These crystals were further optimized by grid screening around the parent condition for obtaining bigger crystals suitable for diffraction (**Figure 34A**). The final crystallization condition contained 0.1 M HEPES pH 7.5, 16% PEG 6000, and 10% MPD.

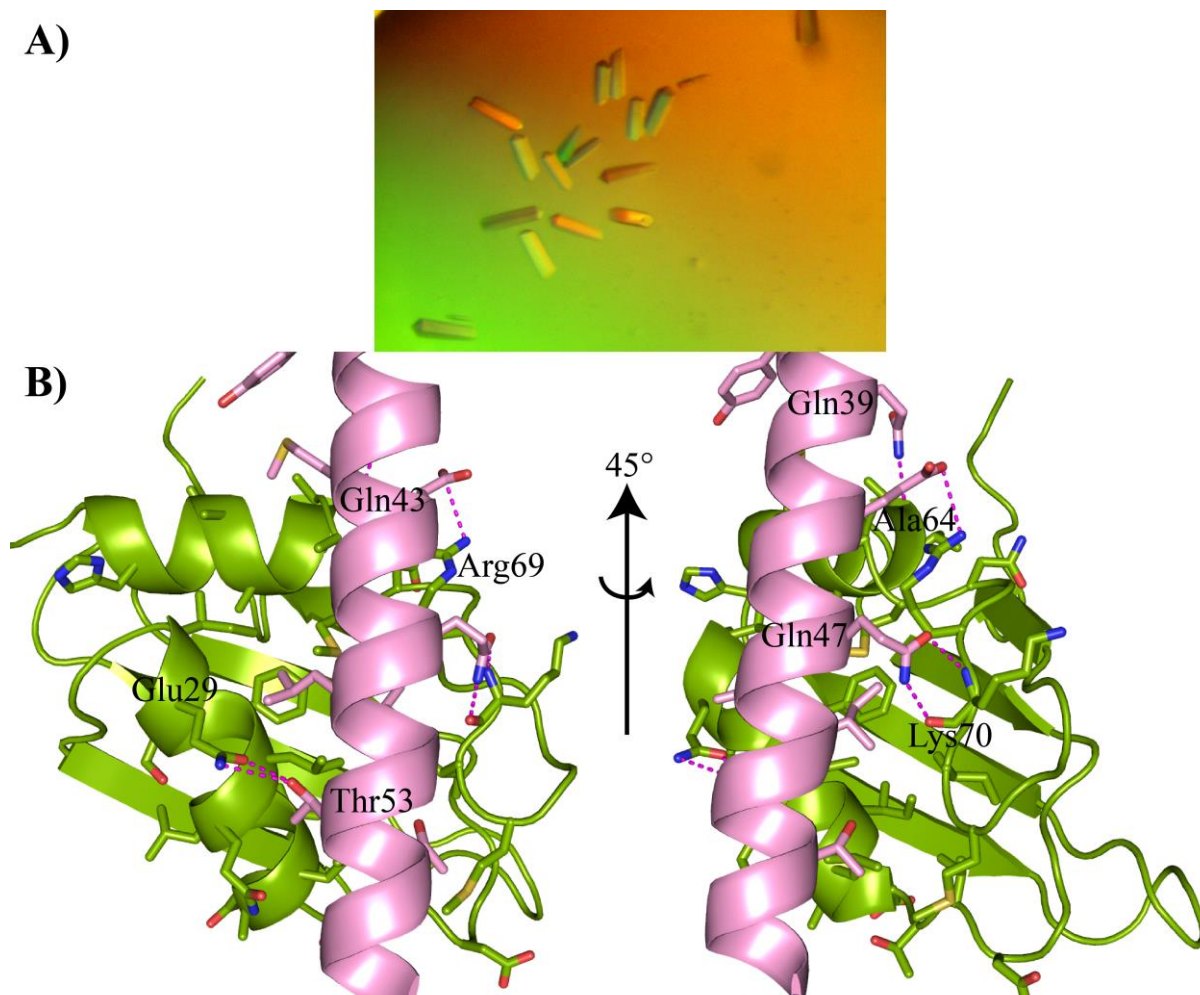


Figure 34. Crystal structure of RRM1-U1C peptide complex

A) Rod shaped crystals used for data collection are shown. B) Crystal structure of the RRM1 domain and the U1C30-61 peptide is shown. The electron density for the GS linker is not visible and thus it is not modeled. The hydrogen bonds between U1C and the RRM domain are shown with dashed lines.

The crystals were set up by handing drop method with 2 μ l of 8.4 mg/ml protein and 2 μ l crystallization buffer. Diffraction quality crystals in the optimized condition appeared in 1 day. The crystals were then transferred into the crystallization buffer containing 20 % MPD for few minutes after which they were flash frozen in liquid nitrogen. The crystals diffracted to 2.4

Å-3.0Å at the synchrotron and the dataset for best crystal was processed further. The phases were obtained by molecular replacement with the NMR structure of RRM1 domain as the search model. Clear and unambiguous electron density could be seen after the molecular replacement which was modeled with the U1C peptide 30-61 with coot software (Emsley and Cowtan 2004). The statistics for the data processing and model refinement is shown in **Table 5**.

Table 5. Data processing and refinement statistics for RRM1-GS15-U1C 30-61 crystals

Parameter	RRM1-GS15-U1C 30-61
Wavelength	0.9193 Å
Resolution range	43.03 - 2.66 (2.755 - 2.66)
Space group	P 64
Unit cell	70.82 70.82 60.39 90 90 120
Total reflections	53294 (3662)
Unique reflections	4963 (433)
Multiplicity	10.7 (8.5)
Completeness (%)	0.99 (0.88)
Mean I/sigma(I)	20.16 (2.41)
Wilson B-factor	65.29
R-merge	0.08015 (0.6728)
R-meas	0.0841 (0.7139)
Reflections used in refinement	4958 (432)
Reflections used for R-free	248 (21)
R _{work}	0.2044 (0.3293)
R _{free}	0.2443 (0.4016)
Number of non-hydrogen atoms	928
macromolecules	928
Protein residues	118
RMS(bonds)	0.015
RMS(angles)	1.95
Ramachandran favored (%)	79
Ramachandran allowed (%)	15
Ramachandran outliers (%)	6.1
Rotamer outliers (%)	8.2
Average B-factor (macromolecules)	82.63

Statistics for the highest-resolution shell are shown in parentheses.

From the crystal structure, it could be seen that the U1C peptide adopts an α -helical structure. The peptide binds in a very shallow pocket of the RRM1 domain between the two α -helices of the protein. As the protein-protein interaction pocket is very shallow and there are no specific residues that anchor the U1C peptide to the RRM1 domain, it is not surprising that the interaction between the U1C peptide and the RRM1 domain is very weak.

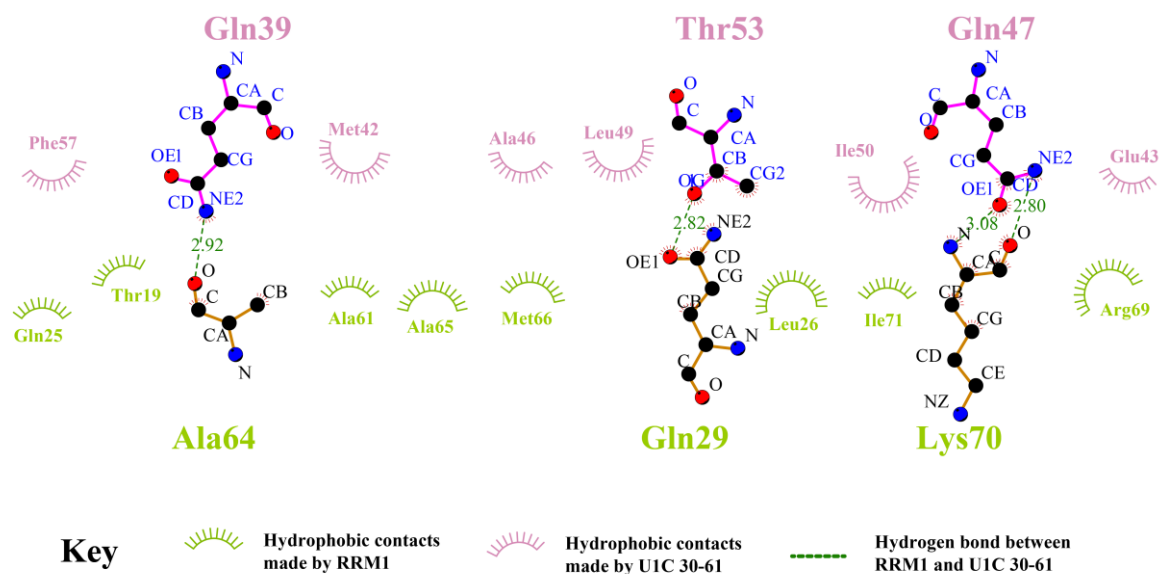


Figure 35. Interaction between RRM1 and U1C 30-61 peptide

Hydrophobic and hydrogen bond interactions between RRM1 (green) and U1C 30-61 peptide (pink) are shown. The numbers on the dotted lines represent the distance between atoms forming hydrogen bonds. The interaction was plotted with DIMPLOT from Ligplot⁺ suite (Wallace et al. 1995; Laskowski and Swindells 2011).

As shown in **Figure 35** and **Figure 34**, Gln39, Gln47, and Thr53 of the U1C peptide form hydrogen bonds with the backbone carbonyl of Ala64, amide and carbonyl of Lys70, and carbonyl of Gln29 respectively. Besides this, Met42, Ala46 and Ile50 form hydrophobic contacts with Leu26, Met66 and Ile71 of the RRM1 domain. It should be noted that Lys70 and Ile71 show highest CSPs when RRM1 is titrated with U1C 30-61 peptide in ¹H, ¹⁵N HSQC titrations. Thus, the crystal structure agrees quite well with the NMR data.

4.6 Discussion

4.6.1 Current understanding of different roles of U1C domains

From the current work, it is clear that the previous understanding of the role of different regions/domains of the U1C protein needs to be updated. Due to the partially inaccurate NMR structure of the U1C (1-61) protein (PDB id: 2VRD) where helix B in the crystal structure is split into helix B and helix C in the NMR structure and folds back on the zinc finger domain, the whole of U1C (1-61) was called as zinc finger domain (**Figure 36**). However, from the current work and the crystal structure of U1C in complex with the U1 snRNP complex (PDB id: 4PJO), it is clear that only region 1-30 of U1C protein forms the zinc finger and residues 31-61 are involved in the protein-protein interaction with the TIA-1 RRM1 domain.

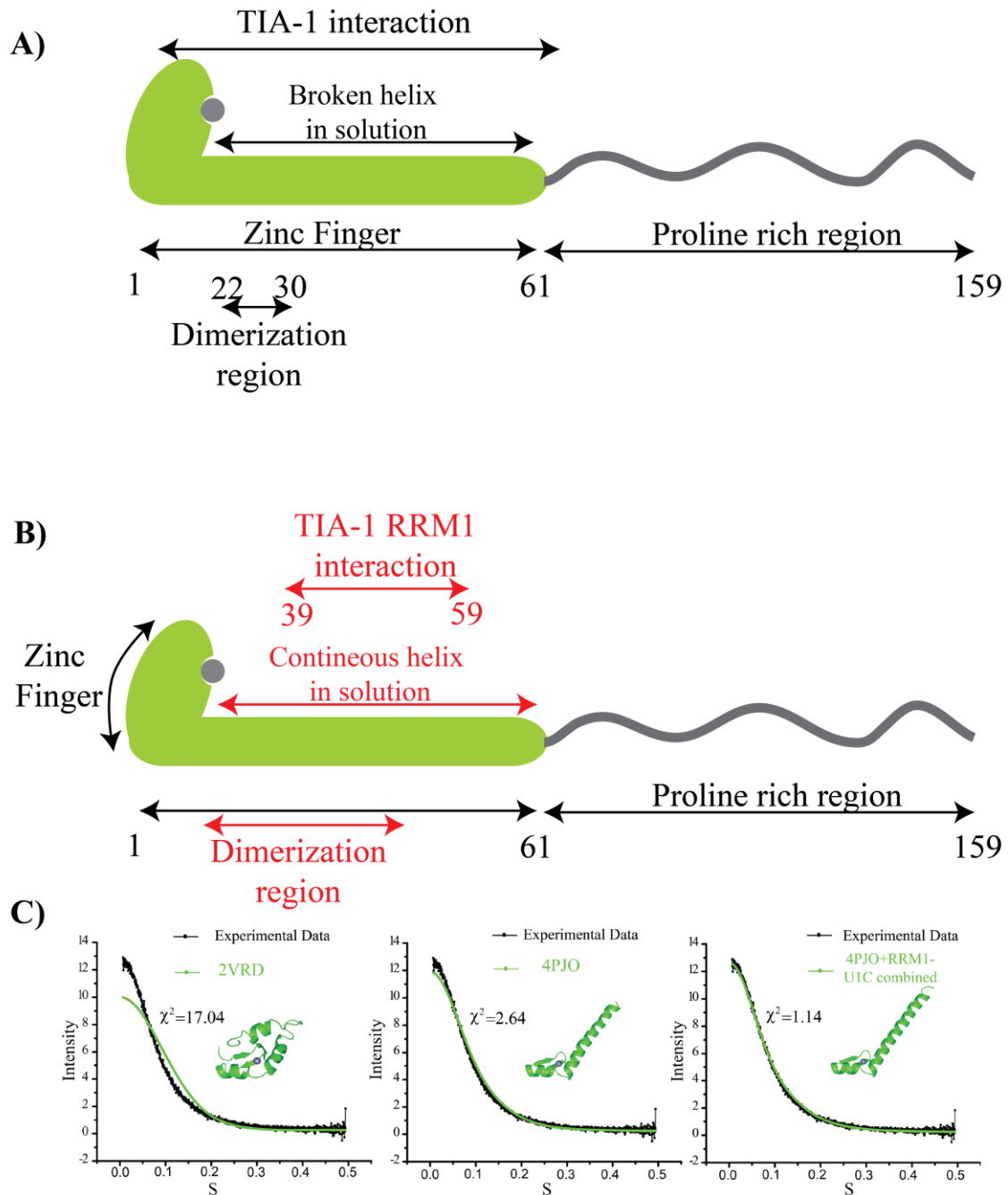


Figure 36. Current understanding of of U1C protein

A) An overview of the earlier understanding of the role of different regions/domains of U1C protein. B) Updated overview based on the results of this thesis. The understanding on new results is shown in red. C) SAXS curve calculated from the combined model of U1C crystal structure and U1C 30-61 helix present in RRM1-U1C (30-61) crystal structure superimposes with the experimental data with least χ^2 .

Also, the dimerization region of U1C protein seems to be limited to somewhere between residue 25-35 as the peptide 30-61 still shows slight tendency to dimerize at 1 mM concentration as compared to the U1C (1-61). The SAXS data clearly proved that the helix B is elongated in solution agreeing more with the crystal structure of U1C.

From the SAXS data, it can be seen that the dimeric U1C protein has a D_{\max} of 11.47 nm whereas the monomeric U1C has a D_{\max} of 7.47 nm (**Table 4**). This 1.5-fold increase in the D_{\max} value from monomer to dimer hints that the dimer is significantly more elongated than the monomer. In the NMR structures, authors note that the NOE's between the helix C and helix B are weak. The weak NOE's were attributed to the loose packing of helix B against helix C. The folding of the helix C on helix B in the NMR structure could be rationalized if the structure was calculated considering U1C as a strict monomer. From the partially wrong NMR structure and the SAXS data, it cannot be negated that the weak NOE's observed between the helix B and C in the NMR structure arises probably from the U1C dimer and their weak intensity indicates the exchange arising due to monomer-dimer equilibrium.

Although the crystal structure of U1C agrees quite well with the SAXS data of the U1C recorded at 0.5 mg/ml, the χ^2 of the fitting of experimental and calculated SAXS curves was still slightly higher ($\chi^2=2.3$). This we attributed to the absence of seven C-terminal residues in the PDB model. As these residues are visible in the RRM1-U1C (30-61) complex structure, SAXS curve calculated from the structural model of U1C obtained by combining the U1C crystal structure in complex with the U1 snRNP and the U1C (30-61) helix structure present in the RRM1-U1C (30-61) complex structure is expected to fit better to the experimentally determined SAXS curve of U1C. This is shown in **Figure 36C** where the χ^2 is now 1.4. This further supports the crystal structure of U1C.

4.6.2 Structural model for TIA-1 U1 snRNP interaction

As the crystal structure of U1C in complex with U1 snRNP is known which also depicts a short 5'ss RNA, it was possible to superimpose the U1C 30-61 helix with the U1C protein structure from the U1 snRNP complex. This superposition can be used to understand the binding of TIA-1 in context with U1 snRNP at 5'ss.

From the superposition, it is clear that the RRM1 domain binds to the solvent exposed part of U1C protein. In addition, there are no steric clash of RRM1 with any other component of the U1 snRNP complex. The C-terminal of RRM1 domain extends towards the expected

position of the extension of the 5' ss. Therefore, it can be envisaged that RRM2 and RRM3 would bind to this region of RNA and hence TIA-1 is thus expected to bind in the 3' to 5' direction of 5' ss RNA sequence (**Figure 37**).

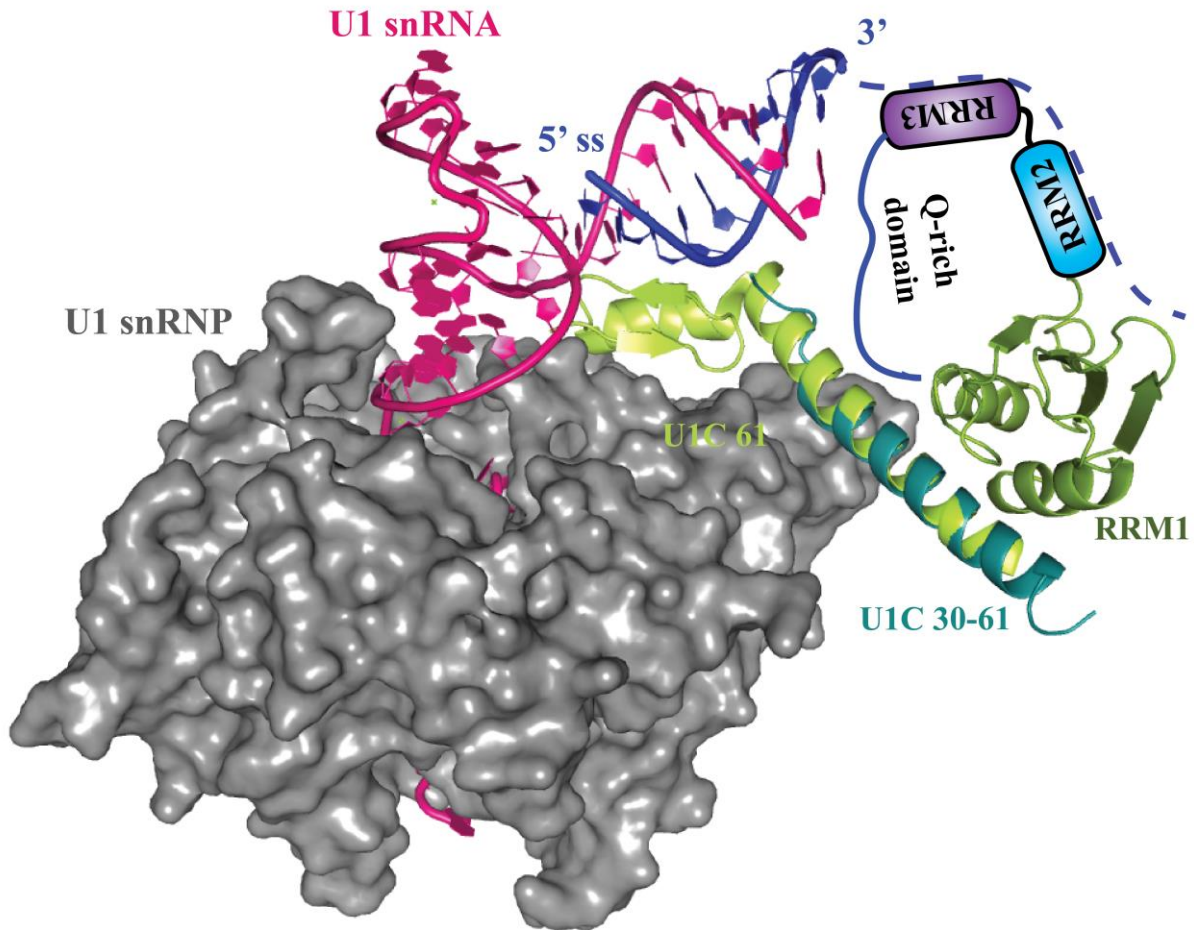


Figure 37. Model for TIA-1 U1 snRNP interaction

A model derived from the current understanding of TIA-1-U1 snRNP interaction is shown. The U1 snRNP structure (U1 snRNP core proteins (grey), U1 snRNA (magenta), 5' ss RNA (blue) and UIC (green)) is taken from PDB (PDB id: 4PJO). The 5' ss RNA is shown in blue with the expected TIA-1 binding site shown by blue dashed lines as continuation of the 5' ss RNA. The RRM1 (split pea green)-UIC 30-61 peptide (teal) complex structure is structurally aligned to the UIC (1-61) (green) in the U1 snRNP structure. The RRM2 and RRM3 would probably bind RNA from 3' to 5' direction as there is substantial distance between the 3' end of 5' ss RNA and the RRM1 bound to UIC which would be occupied by RRM2, RRM3 and the Q-rich domain. As the Q-rich domain is known to interact with the UIC protein, it would perhaps fold back as shown in the model to interact with the UIC and probably RRM1.

The structure also highlights the fact that the β -sheet surface of RRM1 in the complex structure is free to bind RNA. Therefore, although RRM1 binds to RNA with negligible affinity

in isolation, it cannot be ruled out that RRM1 could contribute to RNA binding in context of the full length TIA-1 and in context of TIA-1 U1 snRNP complex.

In addition to this, the structural model also highlights the close proximity of the TIA-1 to the U1 snRNP and hence also U1C protein. Therefore, although the affinity of the U1C protein to the TIA-1 RRM1 domain is weak and is in high μM range, once RRM2 and RRM3 are bound to the 5'ss, the proximity of U1C to the RRM1 domain would likely enhance this interaction.

It was shown before that the Q-rich domain interacts directly with the U1C protein and RRM1 enhances this interaction. From these results, it could be possible that the Q-rich domain also interacts with the RRM1 and U1C protein together to form a ternary protein-protein interaction. This is also supported by the fact that TIA-1 RRM1,2,3 domains form a compact V-shaped structure in solution. This V-shaped structure is expected to bring the RRM1 and the Q-rich domain in proximity thereby promoting the interaction between the two domains and thus probably augmenting the interaction of Q-rich domain with U1C protein. However, due to difficulties in working with the Q-rich domain for NMR as it forms protein fibrils in very short period, the interaction of Q-rich domain with U1C and RRM1 was not studied here. So it remains to be seen how Q-rich domain interacts with U1C and RRM1 at the atomic level.

Chapter 5
Rational design of cyclic peptide inhibitors of SPF45 UHM
domain

As the role of several UHM-ULM interactions has been discovered in splicing regulation in the recent past, it becomes tempting to modulate these regulatory steps to understand the role of these interactions in pre-mRNA splicing. In addition, given that many of the UHM containing proteins are associated with various diseases where the role of their UHM domains in the disease remain to be studied, inhibitors which could modulate these interactions will provide useful tools to understand the role of UHM domains in these diseases.

Cyclic peptides as inhibitors present several advantages over linear peptides such as their decreased sensitivity to degradation by proteases inside the cell, stability of the secondary structure of the peptide due to rigidification of the cyclization bond and increased bioavailability and half-life. In cases where enthalpy-entropy compensation effects are negligible, cyclization of the peptide reduces the flexibility and increases the pre-organized nature of the peptide thus increasing their affinity.

Previous attempts in the group (by Dr. Divita Garg) to identify the minimal motif of the linear SF3b ULM peptide and ideal cyclization points in the minimal peptide identified peptide P3 (cyclo KSRWDE) which had decent affinity (1.4 μ M) for the SPF45 UHM domain and was suitable for further modifications. Therefore, the work performed in this project was started with P3 as the starting point.

In line with the aims of this thesis, cyclic peptide and small molecule inhibitors that could modulate the splicing were developed. The developed cyclic peptides could selectively bind to the SPF45 UHM domain whereas the small molecule inhibitors could target different UHM domains with similar affinity by targeting the highly conserved ULM binding surface of the UHMs.

The optimization of both, cyclic peptide and small molecule inhibitors was guided by structure based drug design approach for which several UHM-inhibitor complex crystal structures were. Structure based approach lead to the identification of regions in UHM domains that could be targeted for selectivity such as the less conserved region after the RXF motif and pockets for growing the small molecules to increase their affinity.

5.1 Crystal structure of SPF45 UHM-cyclic peptide complex

In order to investigate the structural basis of the interaction between SPF45 UHM domain and the P3 cyclic peptide, the crystal structure of the SPF45 UHM domain - P3 cyclic peptide complex was determined. Thin plate shaped crystals were obtained in the crystallization condition (50 mM MES pH 6, 70% MPD) in 5-7 days. The crystals were highly stacked over each other thereby making it difficult to pick up single plates in the cryo loops. Data from multiple crystals were collected and the best dataset, which showed presence of single crystal in the loop, was further chosen for solving the structure.

Table 6. Crystallographic data collection and refinement statistics for the SPF45-UHM/P3 complex

Parameter	Value
Wavelength	1.0 Å
Resolution range	37.24 - 2.22 (2.30 - 2.22)
Space group	P 1 21 1
Unit cell	37.6 73.59 45.01 90 97.89 90
Total reflections	60259 (4142)
Unique reflections	11806 (1033)
Multiplicity	5.1 (4.0)
Completeness (%)	98 (88)
Mean I/ σ (I)	14.61 (5.29)
Rmerge	0.07603 (0.2116)
Rwork	0.1812 (0.2704)
Rfree	0.2444 (0.3842)
RMS(bonds)	0.019
RMS(angles)	1.87
Ramachandran favoured (%)	99
Ramachandran allowed (%)	0.95
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.7
Average B-factor	21.04
macromolecules	20.32
solvent	27.38

Statistics for the highest-resolution shell are shown in parentheses.

The structure of the complex was solved by molecular replacement using the crystal structure of SPF45 UHM domain in complex with SF3b155-ULM as the search model. This gave a single solution with two molecules in the asymmetric unit in Phaser. The refinement statistics are given in **Table 6**.

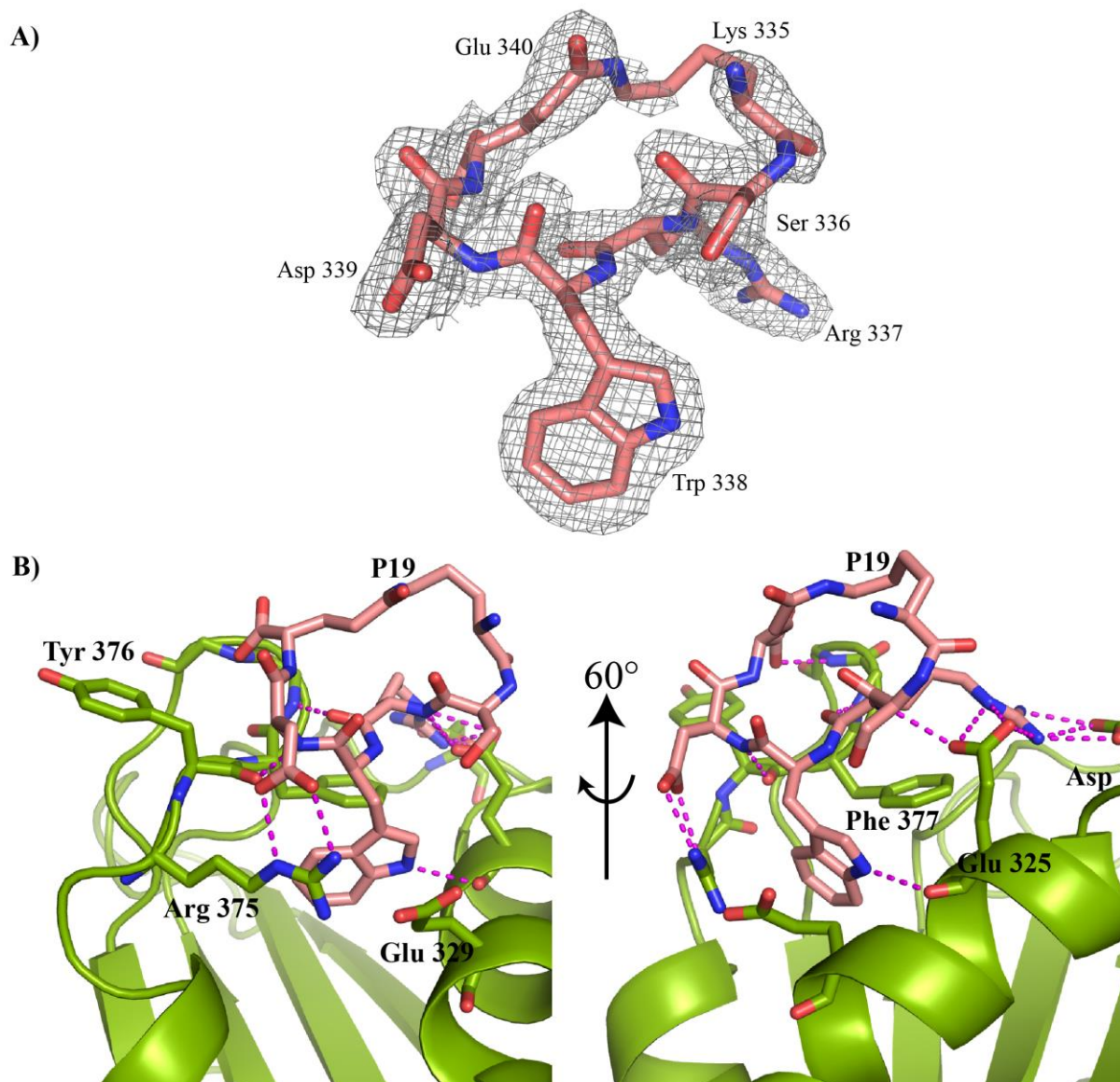


Figure 38. Crystal structure of SPF45 UHM-P3 cyclic peptide complex.

The crystal structure of SPF45 UHM in complex with the cyclic peptide is shown. The cyclic peptide is shown in pink color whereas the UHM domain is shown in green color. A) Electron density of the cyclic peptide contoured at 1σ level. All atoms of the peptide are visible except the $C\beta$ and $C\gamma$ of the Lys335. B) Hydrogen bonds between the cyclic peptide and the UHM domain are shown as dashed lines in magenta color.

The structure shows unambiguous electron density for the cyclic peptide in the ULM binding pocket of the UHM domain (**Figure 38**). Overall, the UHM domain shows minor structural differences in the cyclic peptide bound form when compared to the structure of the complex in the presence of the linear peptide. The Trp338 side-chain of the cyclic peptide inserts into the hydrophobic tryptophan-binding pocket of the UHM domain and anchors the

peptide to the UHM domain. It forms parallel and T-stacking interactions with the side chains of Arg375 and Phe377 in the UHM domain, respectively. Asp339 and Arg337 of the peptide form salt bridges with Arg375 and Asp319 of the UHM domain. The backbone amides and carbonyls of the cyclic peptide are extensively hydrogen bonded to the backbone amides of the UHM domain. The cyclic peptide therefore preserves the critical interactions with the UHM domain compared to the SPF45 UHM-SF3bULM complex structure. The conformation of the cyclic peptide nevertheless differs significantly compared to the linear peptide when bound to the SPF45 UHM domain (backbone coordinates RMSD 0.97 Å). Most of the structural differences are limited to the Lys335 and Glu340 side-chains of the peptide, which mediate the cyclization. The terminal atoms of Lys335 and Glu340 of the cyclic peptide are clearly visible in the electron density. These residues in the linear peptide do not show any electron density when bound to the SPF45 UHM domain (PDB accession: 2PEH) thus indicating the significant rigidification of the peptide induced by the cyclization.

5.2 Structure based design of new peptides

Based on the structure of SPF45 UHM-P3 complex, seven new cyclic peptides were designed to improve the affinity. Given that the carboxyl terminal of Glu340 gives easy access for the modification of the peptides, these peptides were designed with the substitution of different amino acids at the carboxyl terminus of Glu340 to enable the interaction of the substituted amino acid in the cyclic peptide with Tyr376 of SPF45 UHM domain (position R1 in **Figure 39A**). The substituents at R1 position were chosen such that they would make π - π and cation- π interactions with Tyr376. Additionally, as the Tyr376 is solvent exposed, stacking interactions with the substituted amino acids at R1 position are expected to free the water molecules bound around Tyr376, thus positively contributing to the binding entropy. Peptide P8 was synthesized with the substitution of methyl group at position R2 to restrict the flexibility of the cyclic peptide by further rigidifying the cyclization bond between Lys335 and Glu340. This peptide was also expected to contribute positively towards binding entropy by further restricting the conformation of the cyclic peptide to the one observed in the UHM-P3 complex structure.

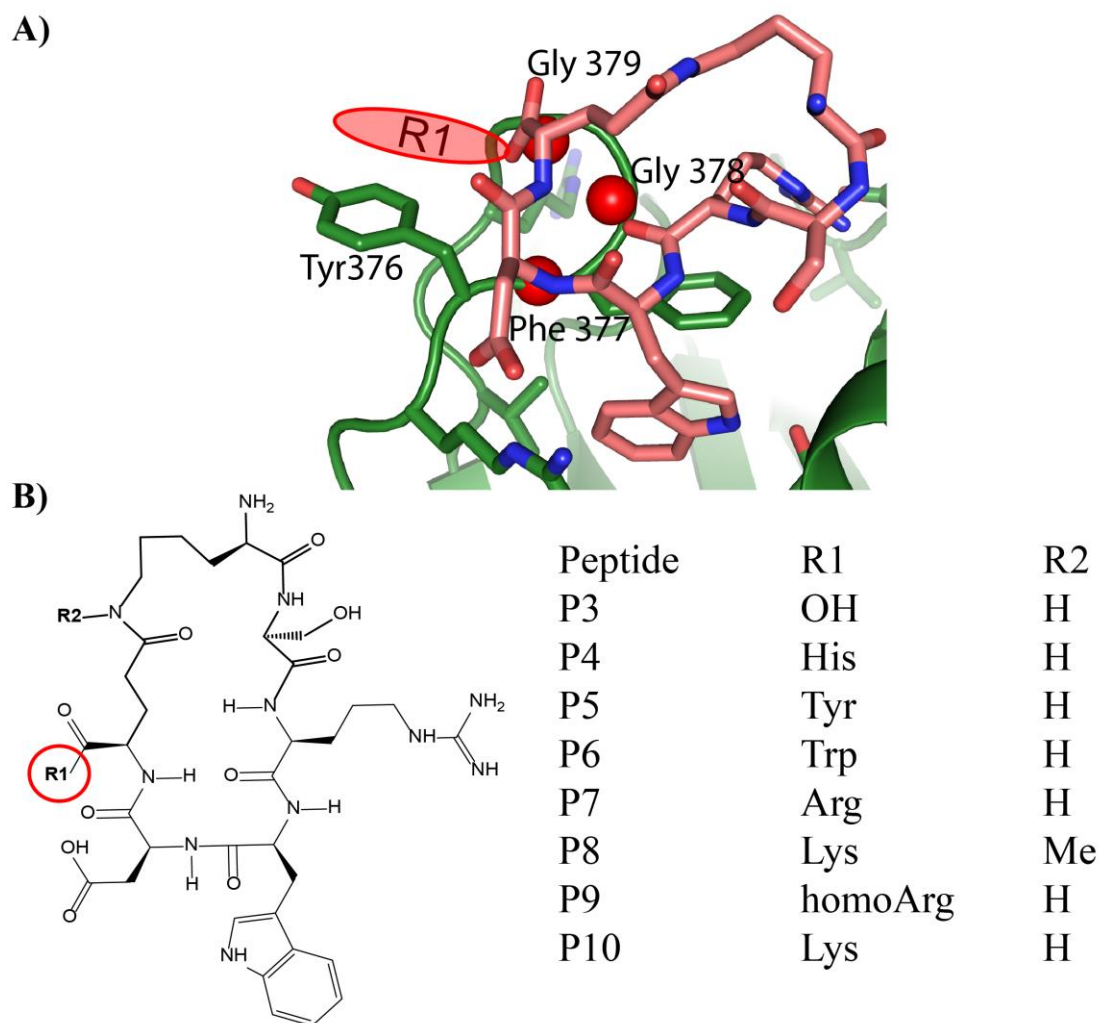


Figure 39. Overview of the synthesized focused library of the cyclic peptides

A) The position of amino acid substituents in the cyclic peptides is shown with R1. The red oblong shape represents the expected direction of the extension of the new substituents at R1 position. The red spheres represent the amides of the UHM domain showing chemical shift perturbation in NMR titrations compared to the titration of the linear peptide. B) Chemical structure of the P3-cyclic peptide. The position of the R1 substitution is shown with red circle.

The binding affinity of the synthesized peptides was determined by ITC (**Table 7**). Substitution of aromatic amino acids at position R1 (peptide P4, P5 and P6) did not improve the overall binding affinity compared to the native linear ULM or P3 peptide. Contrastingly, the binding entropy of these peptides significantly improved compared to native ULM and P3 peptide. This increase in favorable entropy was probably due to the hydrophobic effect (Chandler 2005) as hypothesized before. However, the increase in binding entropy was compensated by a drastic decrease in binding enthalpy. This enthalpy-entropy compensation effect (Gilli et al. 1994) clearly signifies the importance of peptide flexibility wherein sufficient ligand flexibility is required to make proper interactions with the UHM domain. Substitution

of positively charged amino acids (Arg and Lys) at R1 position showed an increase in binding affinity. This effect is attributed to a significant increase in favorable binding entropy and a slight decrease in binding enthalpy compared to the linear ULM and P3 peptide thereby leading to an overall increase in binding energy and binding affinity. Therefore, substitution of R1 with positively charged flexible amino acids seems to maintain sufficient peptide flexibility to make optimal interactions with the UHM domain without affecting the binding enthalpy.

As mentioned above, the requirement of the peptide flexibility for binding to the SPF45 UHM domains was exemplified during the optimization of the cyclic peptides. Those peptides that showed substantial increase in the binding affinity achieved either due to highly pre-organized β -turn structure or due to substitutions of amino acids at the carboxyl terminus of Glu340 that made tight interactions with the UHM domain led to a significant drop in favorable binding entropy. This presented a classic case of enthalpy-entropy compensation. Therefore, it was highly important in case of designing the cyclic peptides against SPF45 UHM domain to take into account the requirement of maintaining sufficient peptide flexibility in the cyclic peptide so that it makes optimal contacts with the SPF45 UHM domain.

Table 7. Binding affinity of different cyclic peptides to the SPF45 UHM domain determined by isothermal titration calorimetry.

	Peptide ^[a]	K_D (μ M)	ΔH ^[b]	$-T\Delta S$ ^[b]
P0	KSRWDE	15.7 \pm 0.98	-10.4 \pm 36	3.8 \pm .17
P1 ^[c]	[sc,bb(KSRWDE)]	122.3 \pm 3.9	-7.8 \pm .23	2.5 \pm .21
P2 ^[c]	[sc,sc(OrnSRWDE)]	2.0 \pm .3	-16.5 \pm .55	8.7 \pm .55
P3 ^[c]	[sc,sc(KSRWDE)]	1.4 \pm .27	-18.3 \pm 1.3	10.3 \pm 1.4
P4	[sc,sc(KSRWDE)]-H	0.96 \pm .08	-10.0 \pm .19	1.8 \pm .19
P5	[sc,sc(KSRWDE)]-Y	1.85 \pm .21	-9.2 \pm .31	1.4 \pm .32
P6	[sc,sc(KSRWDE)]-W	0.80 \pm .14	-11.0 \pm .54	2.6 \pm .55
P7	[sc,sc(KSRWDE)]-R	0.51 \pm .12	-13.7 \pm .76	5.2 \pm .77
P8	[sc,sc((N ϵ Me)KSRWDE)]-K	1.26 \pm .11	-9.2 \pm .20	1.1 \pm .21
P9	[sc,sc(KSRWDE)]-homoR	0.75 \pm .05	-20.2 \pm .19	11.9 \pm .19
P10 ^[d]	[sc,sc(KSRWDE)]-K	0.18 \pm .02	-16.4 \pm .18	7.2 \pm .19

^[a] Errors represent standard deviation of the fitting errors calculated by error propagation, ^[b] kcal/mol, ^[c] Peptide designed/ITC data from Dr. Divita Garg, ^[d] Peptide designed by Dr. Divita Garg.

As substitution of lysine at R1 position showed the highest increase in binding affinity, P10 was chosen for further analysis. To test our earlier hypothesis that amino acid substitution at position R1 indeed makes contacts to Tyr376, we monitored chemical shift perturbations

(CSPs) in $^1\text{H},^{15}\text{N}$ HSQC experiments comparing addition of P3 and P10 to ^{15}N -labeled SPF45 UHM domain. Overlay of the two spectra at saturating concentrations of the cyclic peptides shows that most of the signals exhibit similar chemical shifts in the bound state. Thus, it signifies that P3 and P10 peptides bind to the SPF45 UHM domain in a similar manner. However, significant chemical shift differences are found for the backbone amides of residues Phe377, Gly378 and Gly379 (**Figure 40**). These residues are in the immediate proximity of Tyr376 (**Figure 39A**). This difference in CSPs may reflect ring current effects induced by a conformational rearrangement of the Tyr376 side chain arising due to the stacking of attached lysine at position R1 in the P10 peptide. Given that there is a downfield shift in the peak positions in the hydrogen dimension, it also seems that the hydrogen bond strengths of the amide groups of these residues is weakened due to the substitution of lysine at position R1. This is in agreement with the ITC data as the binding enthalpy of P10 peptide is 2 kcal/mol lower than that of P3 (-16.4 vs -18.3 for P10 and P3 respectively).

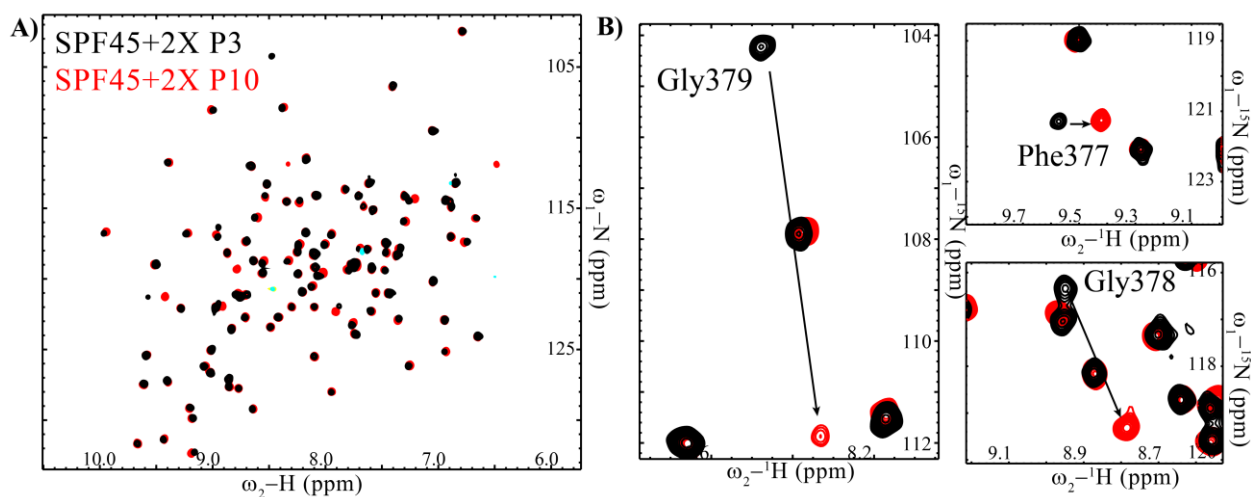


Figure 40. NMR titration of P3 and P10 cyclic peptide into SPF45 UHM domain

A) Chemical shifts upon titration of the P3 and P10 peptides are shown. B) Zoom in views of the residues that show significant chemical shifts upon P3 binding. These residues are shown as red spheres in **Figure 39A**.

5.3 *In vitro* splicing activity of P10

To test the functional activity of P10 peptide, *in vitro* splicing assays were carried out. Assembly of the early spliceosome complex E depends upon UHM-ULM interactions involving U2AF35, U2AF65, SF1 and SF3b155 proteins and it precedes formation of the spliceosome complex A (Spadaccini et al. 2006; Thickman et al. 2006). The SF3b155-ULM5

binds to both, SPF45 (Corsini et al. 2007) and U2AF65 (Spadaccini et al. 2006; Thickman et al. 2006) UHM domains with $K_D = 0.7 \mu\text{M}$ and $6.7 \mu\text{M}$, respectively. On the other hand, the U2AF35-ULM binds to U2AF65-UHM with nanomolar affinity ($K_D = 135 \text{ nM}$) (Kielkopf et al. 2001; Corsini et al. 2007). Therefore, it seems logical to assume that the cyclic peptide inhibitor will modulate splicing by interfering with the UHM-ULM interactions involving SF3b155 ULM only as the interaction between U2AF35-ULM and U2AF65-UHM is too strong.

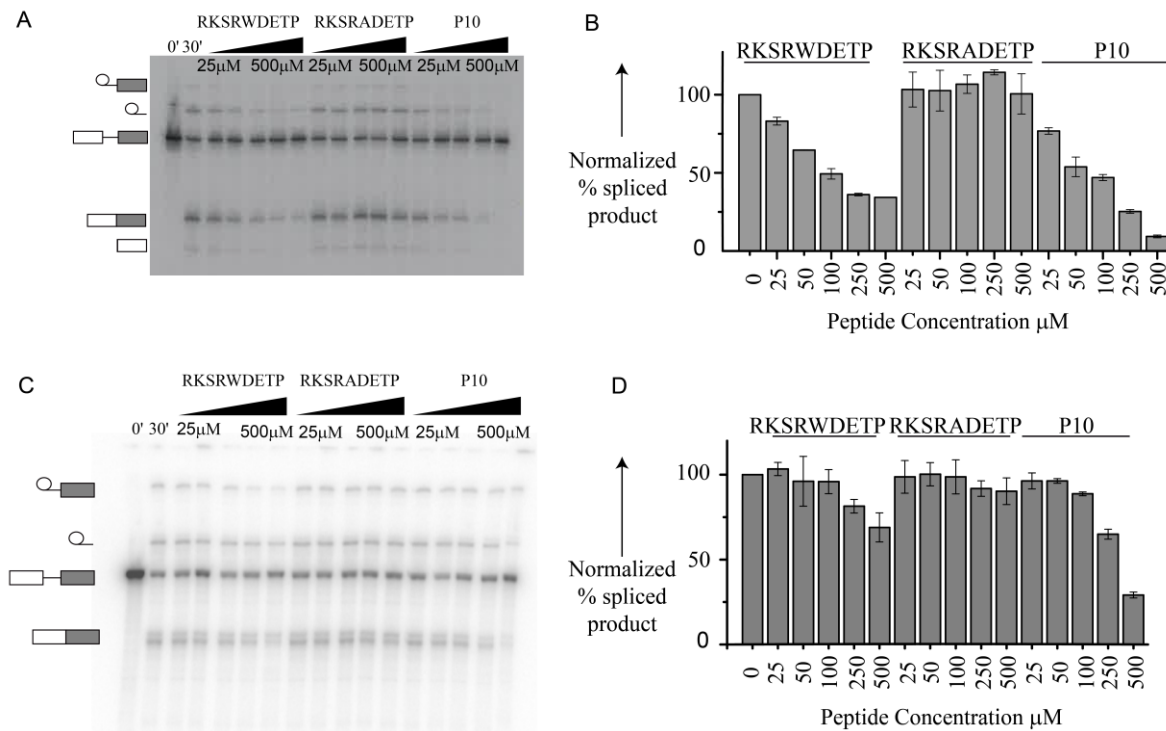


Figure 41. In vitro splicing assays with P10 peptide

Splicing assay to monitor the inhibition of splicing due to P10 inhibitor. A) Inhibition of IgM pre-mRNA splicing by P10. B) Quantification of the spliced product as observed in (A). Nearly complete inhibition of splicing is observed at the highest tested concentration of P10 (500 μM). C) Inhibition of MINX pre-mRNA splicing by P10. D) Quantification of the spliced product as observed in (C). The splicing inhibition by P10 is less efficient compared to IgM pre-mRNA.

To test this hypothesis, we monitored splicing reaction and splicing complex formation *in vitro* with IgM and MINX pre-mRNAs as splicing substrates (Guth et al. 1999) in the presence of P10 peptide. The native ULM5 of SF3b155 (RKSRAWDETP) and its W338A (RKSRADETP) variant were used as positive and negative controls respectively as shown in **Figure 41**. Splicing was nearly abolished at the highest concentration of P10 tested in case of IgM pre-mRNA. Similar results were observed with the MINX pre-mRNA, which possesses

a stronger 3' splice site compared to IgM pre-mRNA. However, the extent of splicing inhibition for MINX pre-mRNA is significantly different from IgM pre-mRNA, i.e. 29% spliced product formed relative to the no peptide control vs 9%, for MINX and IgM pre-mRNA respectively, at the highest tested concentrations of P10. This is consistent with a differential requirement of UHM-ULM interactions for splicing of substrates with weak versus strong 3' splice sites.

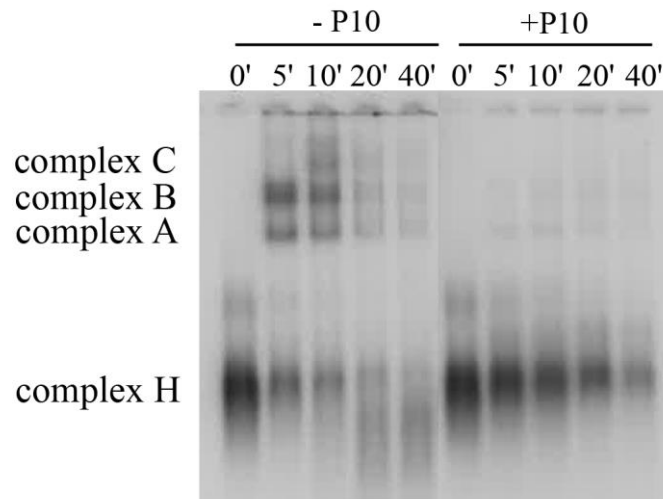


Figure 42. Separation of spliceosome complexes on agarose gel.

Splicing complexes were separated on an agarose gel. The reaction was monitored for a total of 40 min with samples being taken out for analysis at 0, 5, 10, 20 and 40 minutes. The reaction proceeds normally in the absence of P10, however the assembly is stalled at complex A stage in the presence of 500 μ M P10.

In order to rule out that the observed splicing inhibition is due to other factors and to confirm that P10 indeed modulates UHM-ULM interactions during spliceosome assembly, splicing complex assembly was monitored on an agarose gel (**Figure 42**). In the absence of the peptide inhibitor, splicing complex formation proceeds normally and complex A, B and C are formed during the course of the reaction. In the presence of P10, no complex A is formed and the spliceosome assembly is already stalled at complex E stage. During complex E to complex A transition, SF1 is replaced by SF3b155 (Gozani et al. 1998; Das et al. 2000) (**Figure 7**), indicating that the cyclic peptide indeed inhibits UHM-ULM interactions, consistent with previous results (Guth et al. 1999). These data confirm that our novel inhibitors are able to modulate UHM-ULM interactions and stall the spliceosome assembly during or prior to complex A formation. Our data are consistent with P10 interfering with early spliceosome assembly by targeting the U2AF65 UHM-SF3b155 ULM interaction.

Table 8. Dissociation constants (K_D) of the U2AF65 UHM domain determined by ITC

Peptide	K_D (μ M) ^[a]	Selectivity ^[b]
RKSRWDETP	6.46±0.29	9.1
(P3) [sc,sc(KSRWDE)]	7.2±0.36	5.1
(P10) [sc,sc(KSRWDE)]-K	49±1.58	270.4

^[a] Errors represent standard deviation of the fitting errors calculated by error propagation, ^[b] $K_D^{U2AF65-UHM}/K_D^{SPF45-UHM}$

As the cyclic peptides were derived from the SF3bULM5 and were specifically developed to target SPF45 UHM domain, in order to determine the specificity and selectivity of these peptides against other UHM domains, their binding affinity was determined for the constitutive splicing factor U2AF65 UHM domain (**Table 8**). The linear and the P3 peptide showed minor selectivity towards the U2AF65 UHM domain. Therefore, mere cyclization of the linear peptide does not confer any selectivity in the cyclic peptide.

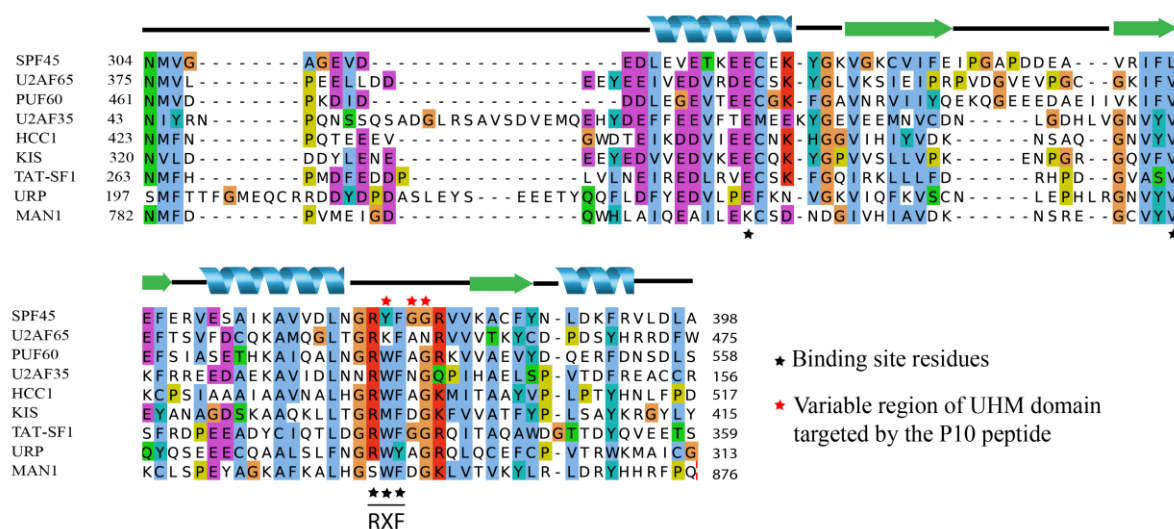


Figure 43. Sequence alignment of UHM domains from different proteins

Binding site residues are shown with a * symbol. The sequence alignment shows the conservation of the binding site residues across various UHM domain. The variable RXF motif is underlined.

This is expected as the cyclic peptide P3 does not make any new contacts with the SPF45 UHM domain and hence behaves similar to the linear peptide. In addition to this, the ULM peptides are highly homologous with the stringent conservation of the signature motif. However, the lysine substituted cyclic peptide P10 shows a 270-fold increase in selectivity for U2AF65 UHM domain. From NMR titration experiments (**Figure 40**), it is clear that P10 shows strong CSPs in residues Phe377, Gly378 and Gly379. Gly378 and Gly379 follow the

RXF motif and the sequence alignment of the UHM domains (**Figure 43**) from multiple proteins showed that the residues following the signature RXF motif of the UHM domains show considerable variation. Therefore, it seems that the significantly enhanced selectivity of P10 arises due to the targeting of this region by the substituted lysine in P10.

Given that the binding pocket of the UHM domains are highly conserved and thus different ULMs bind to a given UHM domain with similar affinities, the prospect of exploiting this variable region of UHM domains for gaining selectivity for the cyclic peptides is very interesting. Thus, for targeting other UHMs with cyclic peptides, their native ULMs could be cyclized with different substituents in the cyclic peptide targeting this variable region in order to gain selectivity against other UHMs. Selective UHM inhibitors can thus be used to delineate the functions of individual UHM domains without interfering with the roles of other domains in multi-domain UHM proteins.

The splicing inhibition shown by the cyclic peptide for both substrates is stronger than the linear peptide signifying that the peptide is indeed more potent than the linear one also in the cell extracts. However, it is surprising that the peptide is around 1000 fold less active in the cell extracts where it inhibits the UHM-ULM interaction with high micro-molar affinity whereas in the ITC experiments it binds to the UHM domain with low nano-molar affinity. This could be rationalized based on the fact that in the cell extract, several other UHMs are present which could potentially titrate away the ULM peptide.

The separation of the inhibited splicing complexes on the agarose gel confirms that the peptide inhibits E to A complex transition where SF1 is replaced by SF3b155. It cannot be negated here that the peptide actually inhibits the UHM-ULM interaction mediated by U2AF65 UHM domain with SF3b155 ULM. This would also agree with the fact that the cyclic peptide inhibitor is only active at high concentrations in the cell extracts at which it is also bind to the U2AF65 UHM domain as shown by ITC experiments. However, as the peptide is highly selective for SPF45 UHM domain, the functional data could also point towards its role in IgM and MINX pre-mRNA splicing.

In conclusion, a cyclic peptide that selectively binds the UHM domain of the alternative splicing factor SPF45 and discriminates against the UHM domain of the essential splicing factor U2AF65 (Banerjee et al. 2004) was designed to modulate the alternative splicing. The developed cyclic peptide inhibitor not only shows better selectivity but also exhibits increased affinity for the SPF45 UHM domain compared to linear peptides.

Chapter 6
**Targeting UHM domains with small molecules to modulate
pre-mRNA splicing**

6.1 High throughput screening for hit identification

6.1.1 Development of fluorescence polarization assay for high throughput screening

Cyclic peptides offer tremendous advantage over the linear peptides for their usage as a tool to study the biological processes or as drug candidates. They are more stable to the enzymatic degradation in *in vivo* conditions or cell based assays. Nevertheless, they suffer from low cell membrane permeability and bioavailability compared to conventional small molecule inhibitors due to their large size and presence of negatively charged residues/backbone atoms. Given that our cyclic peptides could successfully modulate the pre-mRNA splicing in *in vitro* assays by selectively targeting the UHM domains, we decided to develop small molecules, which could target the pre-mRNA splicing by inhibiting early stages of spliceosome assembly through inhibition of the UHM domains.

In order to find potential lead molecules to be developed further into the small molecule inhibitors, classical method of high throughput screening of the small molecule chemical libraries is an attractive approach. However, this approach requires the availability of a viable screening assay. Here we developed a fluorescence polarization based assay to screen a library of 42,000 compounds to identify small molecules to target the UHM-ULM interactions involved in splicing regulation. The assay was developed and the compound screening was carried out in collaboration with the lab of Dr. Felix Hausch at Max-Planck institute of Psychiatry.

Fluorescence polarization assay is based on the fact that polarization of a fluorophore in solution is inversely proportional to its molecular tumbling. Therefore, if a protein domain interacts with a fluorescently labelled peptide/small molecule, the molecular tumbling time increases thus decreasing the polarization of the emitted light. If this interaction is then again inhibited by an inhibitor, the fluorophore is free to tumble in the solution consequently leading to an increase in polarization.

In cases where the fluorescence polarization assay is developed for studying protein-protein interactions, usually one of the partner protein/peptide is tagged with a synthetic organic dye such as fluorescein. The choice of the partner molecule to be tagged with the fluorescent dye depends on many factors including molecular weight and the ease of chemical synthesizability. Therefore, small peptides provide the most suitable and viable option to be used as fluorescently tagged probes in fluorescence polarization assays.

In order to develop fluorescence polarization assay for UHM-ULM interaction, multiple probes were synthesized by tagging the linear ULM peptide with the fluorescein dye at various positions. However, this led to a significant loss in binding affinity of the probe to the UHM domain (data not shown). As the developed cyclic peptides showed improved affinity for the SPF45 UHM domain, a cyclic peptide derived from P3 was tagged with the fluorescein dye as shown in **Figure 44**. This peptide maintained the binding affinity of 2.4 μM as measured by ITC. It also showed a concentration dependent binding to the SPF45 UHM domain and hence the depolarization of light with an effective concentration reaching to 80% (EC_{80}) at 2.3 μM of protein concentration.

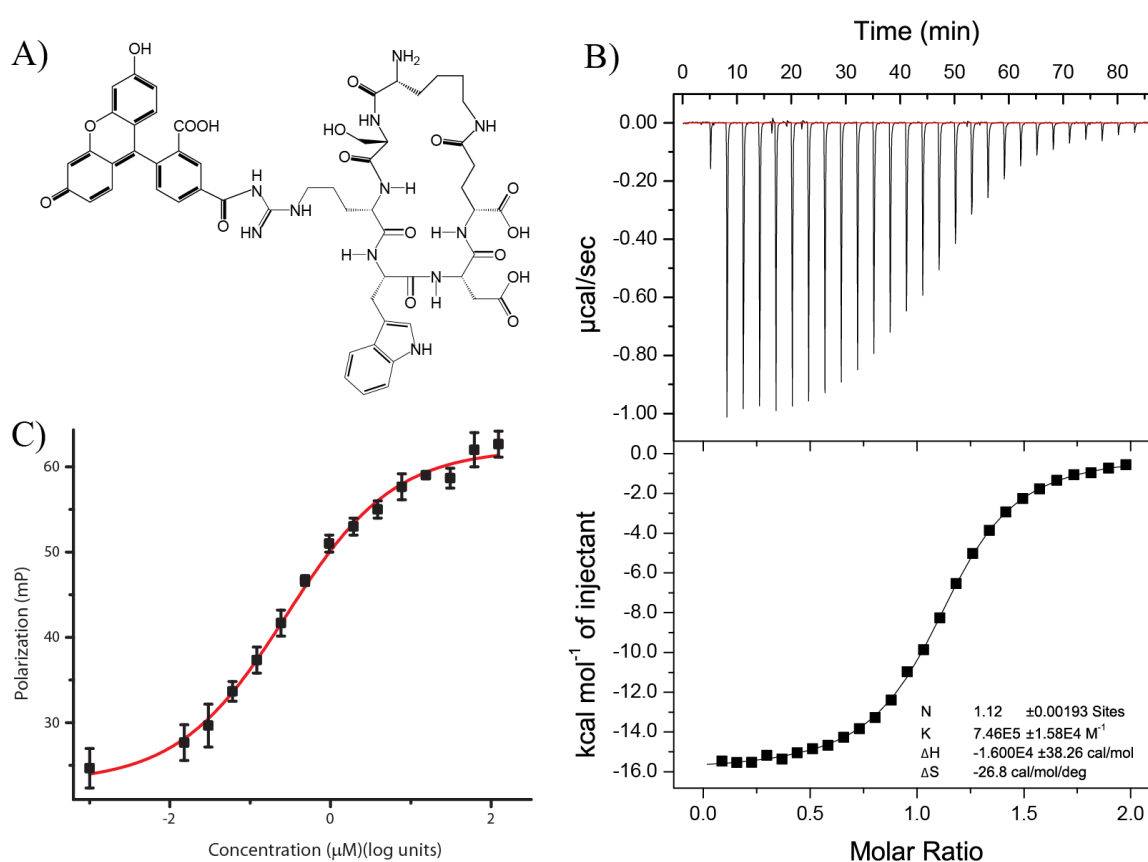


Figure 44. Development of Fluorescence polarization assay

A) Chemical structure of the fluorescently labelled peptide probe used in the fluorescence polarization assay. B) The peptide maintains its binding affinity to the SPF45 UHM domain after the attachment of the fluorophore and binds to the UHM domain with an affinity of $K_D=2.4 \mu\text{M}$. C) The peptide binds to the SPF45 UHM domain and shows a concentration dependent depolarization of the polarized light. The tracer binds to the UHM domain with an EC_{80} of 2.3 μM . The total assay window is $\sim 35 \text{ mP}$.

Given that the cyclic peptide probe shows efficient binding to the SPF45 UHM domain, the assay was further optimized for buffer conditions by changing the salt, tracer and DMSO

concentrations and the buffer pH. The buffer condition with 20 mM Tris pH 7.5, 300 mM NaCl, 1 mM DTT, 100 nM tracer and 1.6% DMSO, that gave the highest assay window of 35 mP was chosen as the final condition for the high throughput screening.

6.1.2 Results of high throughput screening

The developed fluorescence polarization assay was used to screen a library of 42,000 compounds using a single point titration assay. The screening of the compounds was carried out in duplicates to avoid false positives. As the whole assay was carried out in a total of 240 plates and over a period of 3 weeks, the controls in each plate showed slight variations in plate dependent manner. Therefore, all the values from control wells were normalized against each other and data from each plate were normalized against the controls from the respective plates.

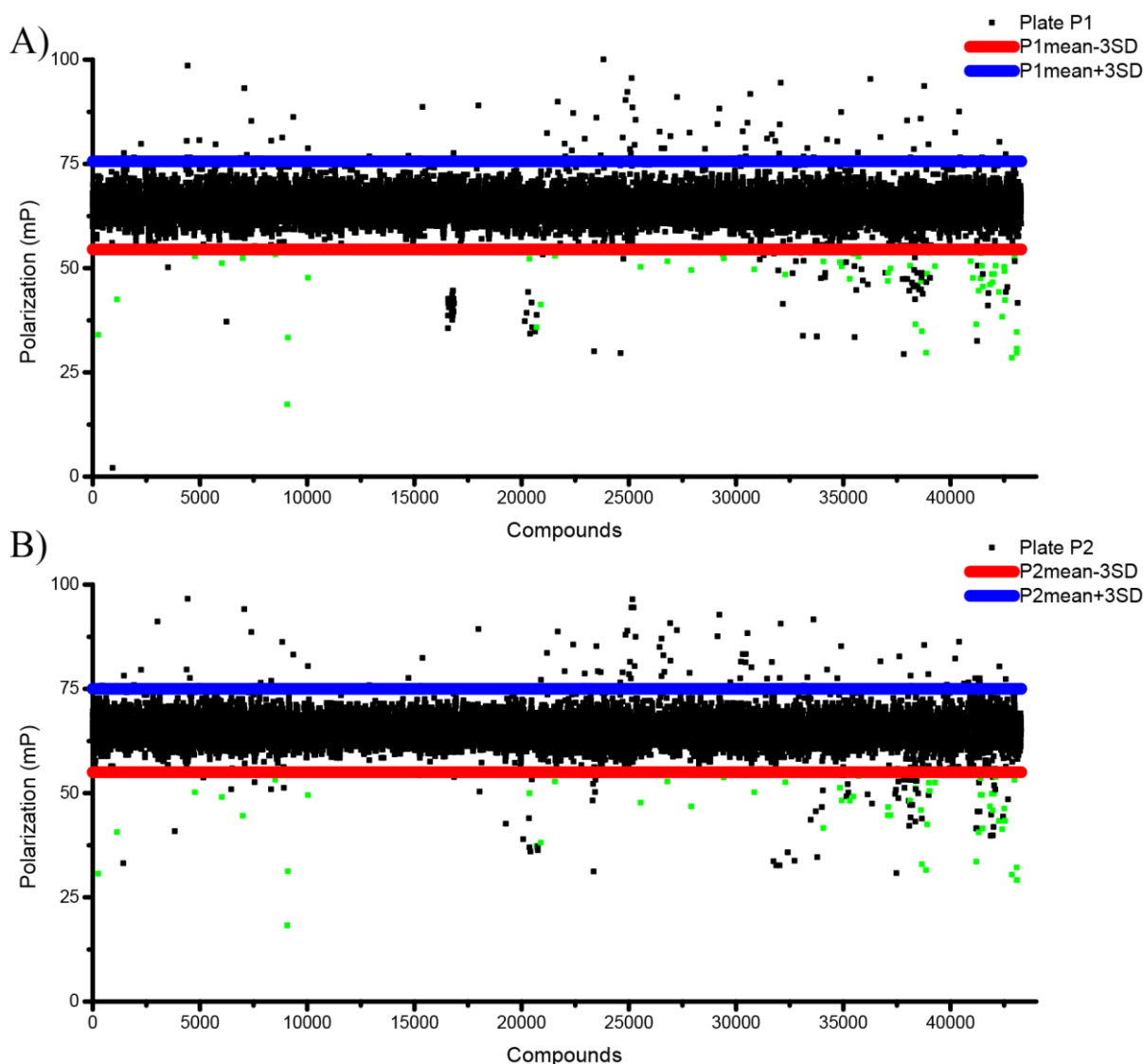


Figure 45. Results of high throughput screening

The results of the HTS plotted as the polarization value of the well against the compound number in the assay plate. The assay was carried out in duplicates as shown in (A) and (B). Green dots represent the compounds which showed depolarization of the light below the 3×standard deviation value and were reproducible in the duplicate assay.

The normalized polarization values were plotted against the compound number in the assay plate as shown in **Figure 45**. Positive hits were defined as the wells which show decrease in the polarization value by three standard deviations from the mean plate value and were reproducible in duplicates. With these criteria, 57 hits were identified as positive hits and these were further subjected to rescreening.

6.2 Hit validation

6.2.1 FP assays titrations

As the fluorescence polarization assay is very sensitive to the false positives arising due to the colored compounds as the fluorescence from these compounds interfere with the assay itself, all the 57 positive hits from the screening campaign were further subjected to rescreening by multipoint fluorescence polarization assay. The screening was again carried out in duplicates with 16 titration points for each compound. The 57 compounds were picked up from the master plates and were 2 fold serially diluted 16 times starting with the concentration of 125 μ M. Cyclic peptide P10 was included as a positive control in each plate.

From the set of 57 hits, only eight hits (seven non-redundant compounds) showed a dose dependent decrease in polarization showing that only these eight hits (seven compounds) thus represent true hits obtained via the FP screening and the rest of the 49 hits were false positives. The dose response curve for these seven hits are shown in **Figure 46**. Out of the 8 hits, two were duplicates (C and D in **Figure 46**) and three compounds share the same scaffold (compound F, G and H). This high rate of false positives in the rescreening assay was expected as almost all of these hits were colored in nature and were thus expected to interfere with the assay. Nevertheless, presence of three compounds with the same scaffold and showing similar EC_{50} concentration values identifies them as positive hits with very high confidence.

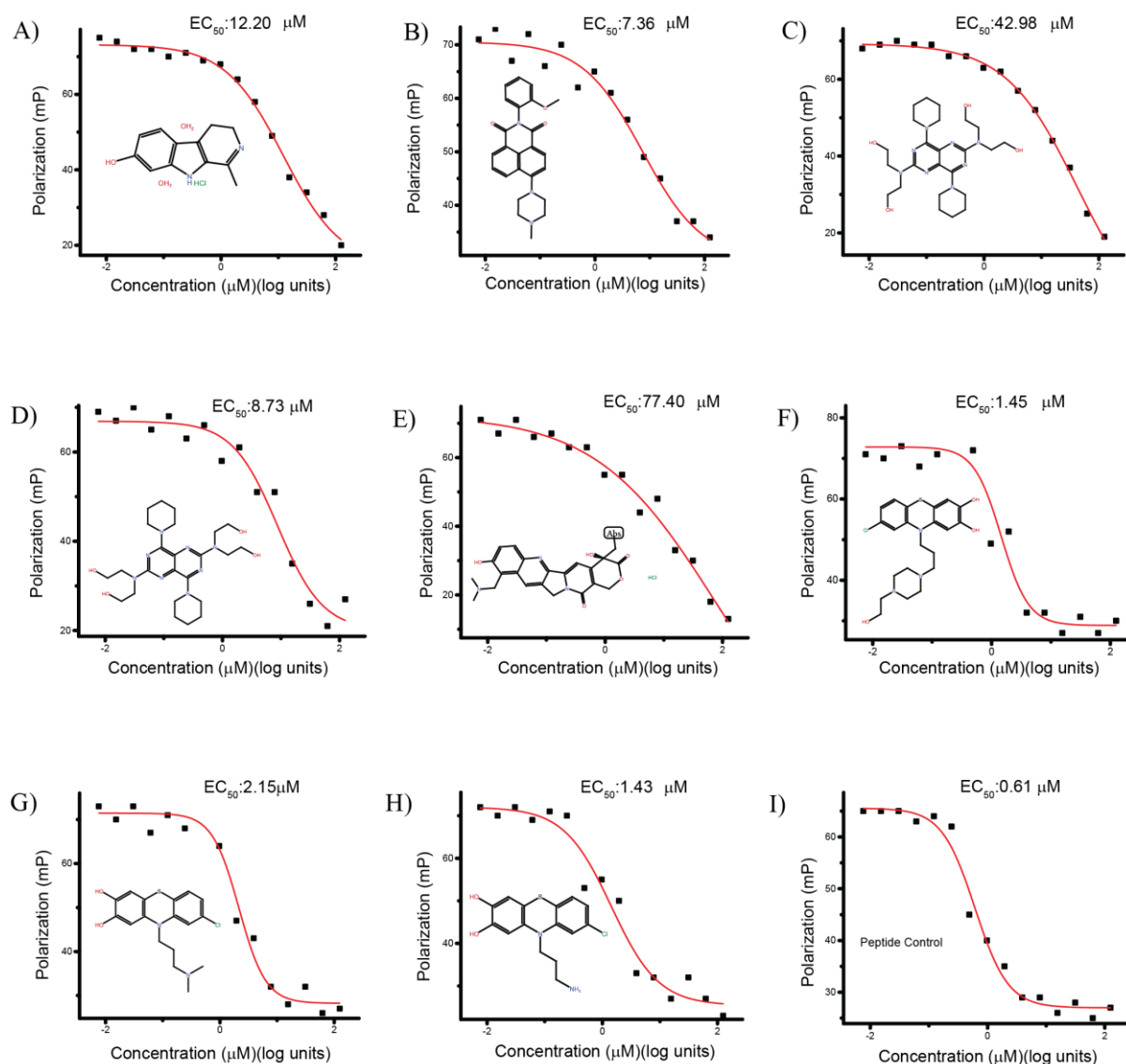


Figure 46. Validation of HTS results by FP

Initial validation of the results from the HTS screen was done using multipoint titration of the compounds using FP assay. Compounds that show positive results are shown here along with their EC₅₀ values. Compounds (C) and (D) were present in two different assay plates and showed inhibition in both instances. Compounds (F), (G) and (H) share the same scaffold. Peptide P10 (I) was used as a positive control.

6.2.2 NMR titrations

The validated hits in multipoint fluorescence polarization assay were further subjected to a second round of validation by titration in ¹H-¹⁵N HSQC NMR experiments. Hits are expected to cause significant chemical shift perturbations upon titration with the ¹⁵N labelled SPF45 UHM domain. Out of the seven compounds, five were commercially available and were purchased for NMR experiments.

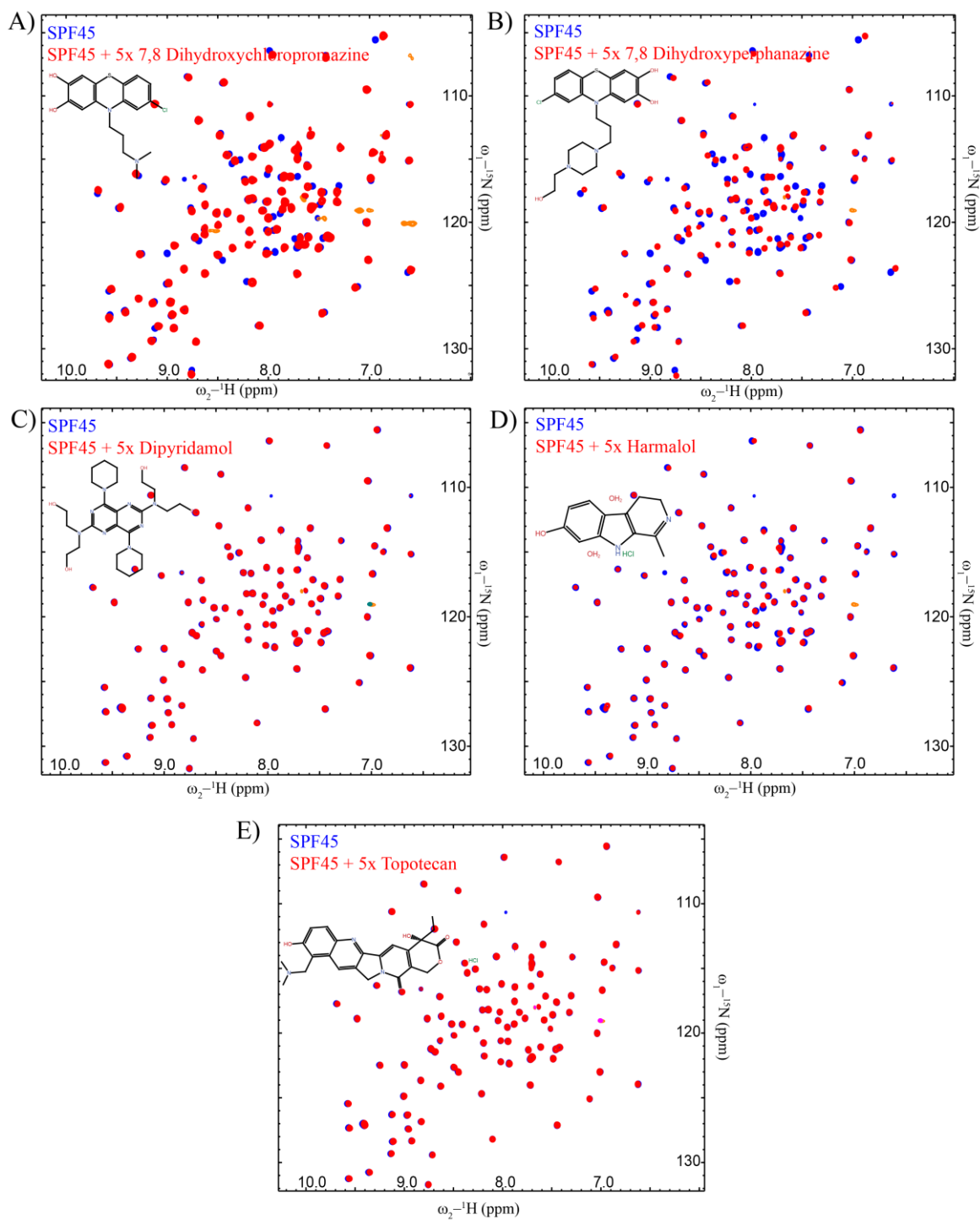


Figure 47. Validation of HTS by NMR titrations

Further validation of the hits from HTS by ^1H - ^{15}N HSQC NMR titrations of the compounds showing inhibition in FP multipoint titration assay. Compounds were titrated in 5-fold excess of the protein in a single point titration assay and chemical shift changes were monitored. Compound (A) and (B) show significant chemical shift changes whereas compound (C) shows minor changes.

The compounds were titrated at fivefold excess concentrations as shown in **Figure 47**. Out of these five compounds, only two (compound A and B) showed significant chemical shift perturbations. Compound D showed minor CSPs whereas compound C and E did not show any CSPs.

Given that only compounds A and B showed significant CSPs, these were then selected as true hits and were chosen for further optimization. As the two compounds share a similar phenothiazine (PTZ) scaffold, they were expected to bind at the same binding site on the SPF45 UHM domain. The two compounds, 7,8 dihydroxy chlorpromazine and 7,8 dihydroxy perphenazine are metabolites of chlorpromazine and perphenazine, respectively. Chlorpromazine and perphenazine are typical antipsychotic drugs primarily used to treat psychotic disorders such as schizophrenia and bipolar disorder.

6.3 Hit optimization

6.3.1 Medicinal chemistry based approach for hit optimization

In order to optimize the potency of the two positive hits further, several new compounds were synthesized for quantitative structure-activity relationship studies (QSAR). In the absence of any structural information regarding the binding mode of the positive hit to the SPF45 UHM domain, the primary hit was broadly divided into four R groups (R1, R2, R3 and R4) based on the synthetic feasibility of new compounds. In the initial phase, new compounds based on R1, R2 and R3 group substitutions were designed and synthesized. All the compounds were designed in collaboration with and synthesized by Dr. Tomas Kubelka and Prof. Dr. Thorsten Bach at Technical University Munich. The chemical structures of these compounds are shown in **Figure 48** (also see Appendix: Chemical structures of the compounds).

From the QSAR studies, the triple ring system of the phenothiazine proved crucial for binding to the UHM domain as new compounds designed based on altering the R1 group showed either no binding or very weak binding to the SPF45 domain as determined by ITC. Substitution of sulphur in the phenothiazine ring to an oxygen to obtain phenoxazine also led to the decrease in the binding affinity. Sulphur in the phenothiazine ring system is expected to provide a bend to the whole three-ring system due to the larger radius of sulphur atom whereas the phenoxazine will be more flat. Therefore, the curvature of the ring system seems to be important for the binding of the compounds to the UHM domain. Exchanging the chlorine (R2 group) atom with bulkier methanolate ($\text{CH}_3\text{CO}-$) and methanethiolate ($\text{CH}_3\text{S}-$) groups also led to a drastic loss in binding affinity of the compounds. This probably shows that the chlorine

atom snugly fits into the binding pocket without any extra space for accommodating other atoms into the pocket.

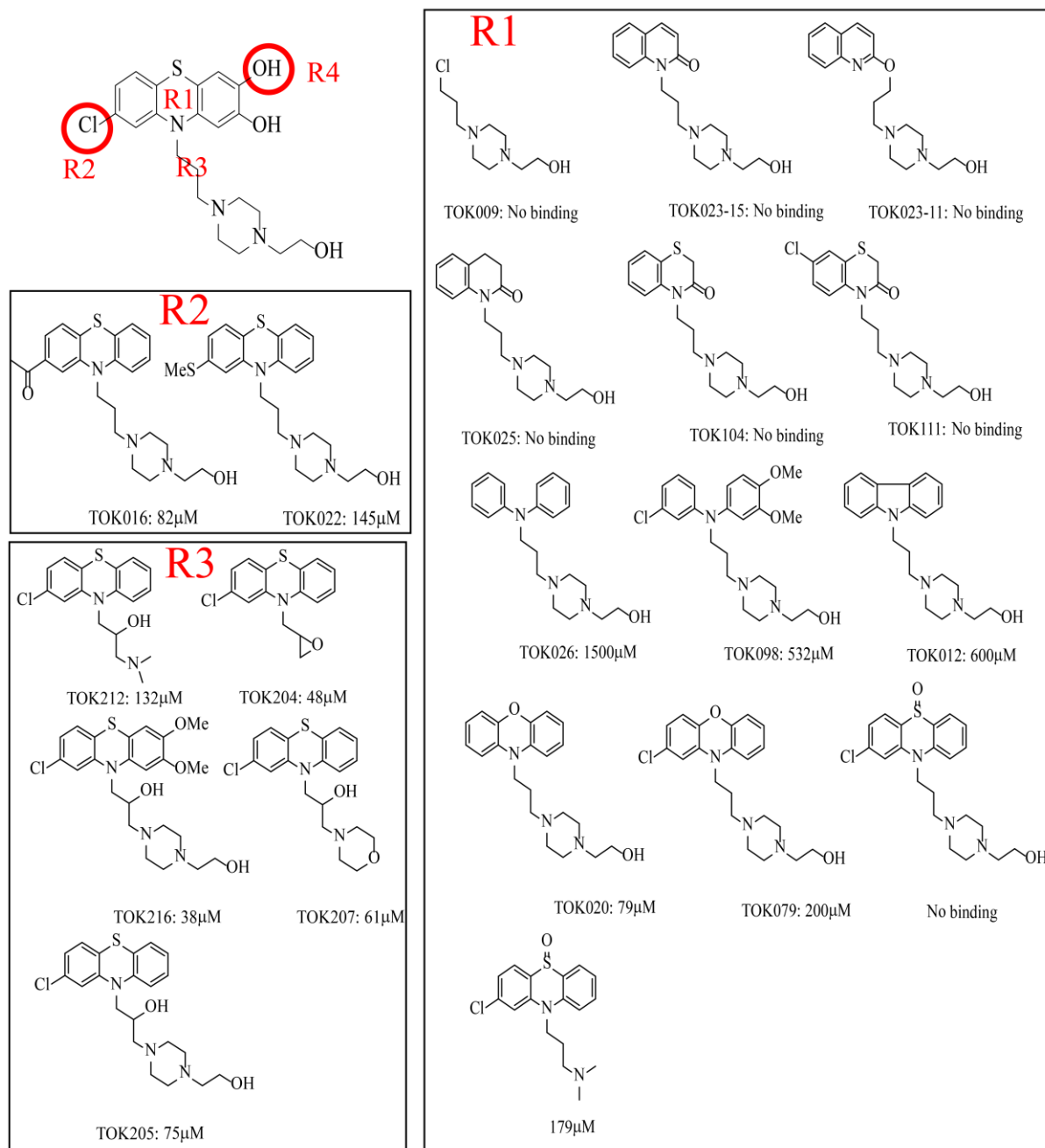


Figure 48. Substitutions at R1, R2 and R3 groups

For QSAR, 7,8 dihydroxy perphenazine was divided into R1, R2, R3 and R4 groups. The compounds synthesized based on altering the R1, R2 and R3 groups are shown here. All three groups proved crucial for binding and any changes in them led to a drastic loss of binding affinity. All binding affinities reported here were determined by ITC.

Altering the piperazinyl moiety (R3 group) also led to a loss in binding affinity. Additionally, the compounds with hydroxyl substitution in the piperazinyl moiety proved to be highly unstable and prone to degradation thus proving them unsuitable for further optimization. Amongst the R3 groups from the parent compounds (N, N dimethyl in 7,8 dihydroxy chlorpromazine and piperazine in 7,8 dihydroxy perphenazine), the latter shows more CSPs in the NMR titrations. Therefore, it seems obvious that the piperazine moiety in 7,8 dihydroxy perphenazine makes extra contacts with the SPF45 UHM domain as compared to N,N dimethyl moiety in 7,8 dihydroxy chlorpromazine.

Phenothiazine and its derivatives have historically been used as sedatives, insecticide, drugs to treat infections with parasitic worms in livestock, treatment for malaria and as antipsychotic drugs. The phenothiazine compounds discovered in our screening had hydroxyl groups at position 7 and 8 of the scaffold leading to the degradation of the scaffold by oxidative degradation. As the attempts to replace the phenothiazine ring led to abolishment of the binding of the compound to the UHM domain, it suggested that this part of the compound makes extensive contacts with the UHM domain. Given that the phenothiazine is hydrophobic in nature, the results were not surprising as the tryptophan-binding pocket is hydrophobic too and the phenothiazine ring probably interacts with this pocket.

6.3.2 Crystallization of positive hits with UHM domain

As there was no significant improvement in the binding affinity of the compounds by changing the R1, R2 and R3 groups, we decided to crystallize the primary hits obtained from NMR titration experiments in complex with the UHM domain. However, the primary hits showed severe oxidative degradation with time accompanied with change in the color of the compound from colorless to pink. In order to overcome this degradation problem, the hydroxyl moieties attached to the phenothiazine ring were substituted with methoxy groups (**Figure 49A**). This substitution prevented further oxidation of the compound thus leading to a stable compound suitable for structural studies.

Our multiple attempts to either co-crystallize the stable methoxy derivatives of the primary hits (7,8 dimethoxy chlorpromazine and 7,8 dimethoxy perphenazine (called TOK116 hereafter)) with SPF45 UHM domain or to soak them in the SPF45 UHM domain crystals obtained in previously published condition (Corsini et al. 2007) failed. In the case of soaking experiments, there was no visible electron density of the compounds at the expected tryptophan-binding pocket even after prolonged soaking of 2 days of the crystals in high

concentrations of compound solution. This was not surprising as the ULM binding site in this crystal form of the SPF45 UHM domain was partially blocked by the neighboring symmetry molecule. Thus, the binding pocket was probably not completely accessible for the compounds.

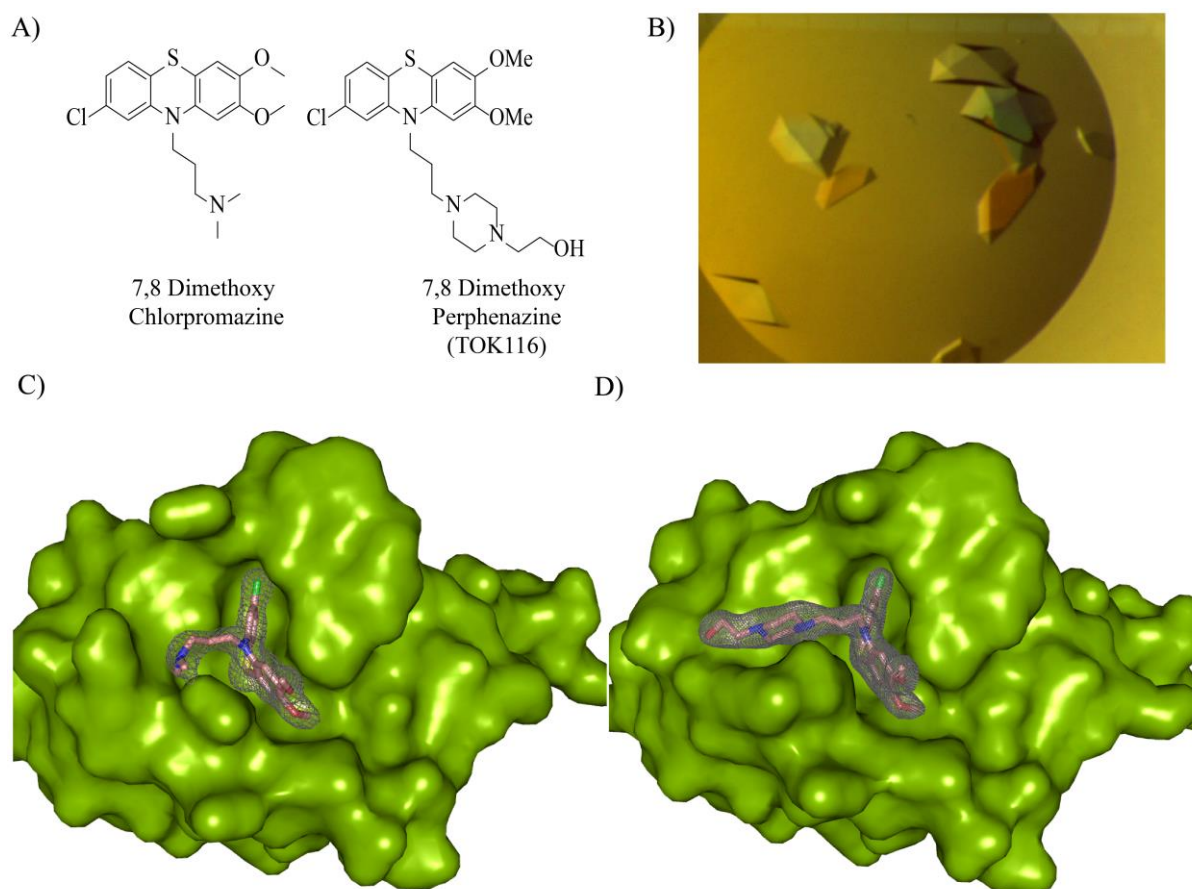


Figure 49. Crystallization of HTS hits with Puf60 UHM

(A) Structures of the compounds soaked in the Puf60 UHM crystals. (B) Crystals of the thioredoxin tagged Puf60 UHM domain as obtained using the crystallization conditions published before (Corsini et al. 2008). (C), (D) the crystal structures of Puf60 UHM in complex with these compounds. The electron density of the compounds in the crystal structure is shown at a contour level of 1σ .

As the ULM binding pockets in UHM domains are highly conserved amongst different UHMs, we hypothesized that our inhibitors could also bind to other UHM domains. Therefore, soaking experiments were carried out with Thioredoxin-Puf60 crystals reproduced in conditions as published earlier (Corsini et al. 2008) (**Figure 49B**). Crystals diffracted to $\sim 2 \text{ \AA}$ resolution (**Table 9**). However, soaking of the Puf60 crystals into the compounds lead to a decrease in the length of the c-axis of the unit cell from 299.39 \AA to 149.4 \AA . Besides this, the symmetry also changed from P212121 to P21212 with loss of one screw axis in the c-axis and so the consequent number of Thx-Puf60 molecules in the asymmetric from eight to four.

Nevertheless, the structure was solved by molecular replacement using the structure of native Puf60 as search model.

Clear and unambiguous electron density of the compound was found in two out of four molecules in the asymmetric unit. The absence of electron density of the ligand in other two molecules of the asymmetric unit was attributed to the blocking of binding site by the symmetry-mate of Puf60. Dimethoxy chlorpromazine shows high B-factors (90.67) in the complex structure and weak density for the *N, N*-dimethyl propan-1-amine moiety signifying that this is quite flexible in the structure compared to other atoms of the ligand. On the other hand, TOK116 shows acceptable B-factors for the ligand (52.86). Due to this reason, TOK116 was chosen for further optimization.

Table 9. Data collection and refinement statistics for Thioredoxin-Puf60 UHM crystals in complex with 7,8 Dimethoxy chlorpromazine and 7,8 Dimethoxy perphenazine

Parameter	Dimethoxy Chlorpromazine	Dimethoxy perphenazine (TOK116)
Resolution range	19.9 - 2.0 (2.07 - 2.0)	19.87 - 1.9 (1.97 - 1.9)
Space group	P 21 21 2	P 21 21 2
Unit cell	81.43 89.48 149.4 90 90	81.53 89.4 149.2 90 90
Unique reflections	73889 (7169)	86090 (8254)
Completeness (%)	99	99
Wilson B-factor	32.14	31.19
R _{work}	0.1911 (0.2399)	0.2024 (0.2900)
R _{free}	0.2400 (0.2447)	0.2463 (0.3910)
RMS(bonds)	0.023	0.020
RMS(angles)	1.93	1.89
Ramachandran favored (%)	97	98
Ramachandran allowed (%)	2.6	1.6
Ramachandran outliers (%)	0.12	0.48
Average B-factor	40.23	39.76
macromolecules	39.16	39.54
ligands	90.67	52.86

Statistics for the highest-resolution shell are shown in parentheses.

6.3.3 Analysis of Thx-Puf60 UHM-TOK116 crystal structure

It was shown before that Puf60 UHM domain binds to various ULMs including SF3b155 ULM with low micro-molar affinity (Corsini et al. 2009). However, there is no available structure of Puf60 UHM in complex with SPF3b155 ULM. So, the crystal structure of Puf60 UHM-TOK116 complex structure was structurally superimposed with the SPF45 UHM-SF3b155 ULM5 complex structure to compare the binding mode of TOK116 with

respect to the peptide-binding mode. The two UHM domains superimpose well with an RMSD of 0.8 Å.

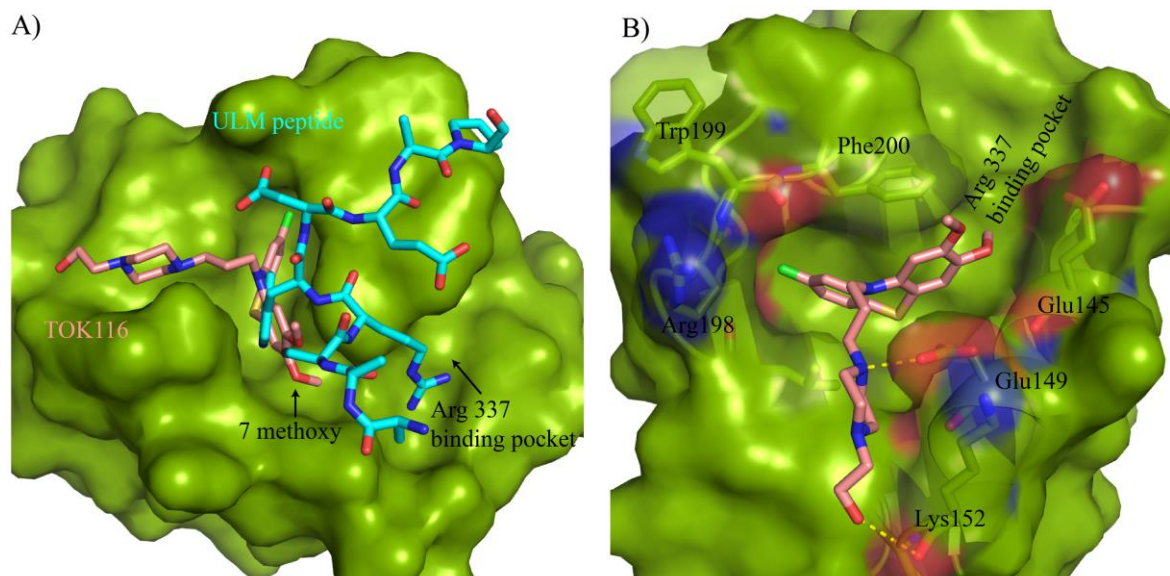


Figure 50. Structure analysis of Puf60 UHM-TOK116 complex

A) Structural comparison of Thx-Puf60 UHM (shown as surface representation in green) in complex with TOK116 (salmon) and SPF45 UHM bound to ULM peptide (cyan) (PDB id: 2PEH). The phenothiazine ring of TOK116 occupies the tryptophan-binding pocket. B) Binding pocket residues of Puf60 UHM-TOK116 complex are shown. The hydrogen bonds are shown as yellow dotted lines.

The phenothiazine moiety of TOK116 occupies the tryptophan-binding pocket with the slight expansion of the binding pocket compared to the expected peptide bound form in order to accommodate the three rings of the phenothiazine. The tricyclic ring system adopts a non-planar shape due to the presence of sulphur heteroatom. The hexagonal ring attached to the 7,8-dimethoxy position forms T-stacking interaction with Phe200 of the UHM domain. The chlorine atom snugly fits into the hydrophobic tryptophan-binding pocket of the UHM domain. The 7-methoxy group of the compound points to the pocket that is occupied by Arg337 of the ULM peptide whereas the 8-methoxy group is solvent exposed. The piperazine moiety extends in the direction opposite to the peptide-binding region. The amine group from the piperazine moiety and the terminal -OH group interacts with the Glu149 and backbone carbonyl of Lys152 via a hydrogen bond. In UHM-ULM peptide complex structures, the Arg198 of the UHM forms salt bridge with Glu149 and forms parallel stacking interaction with the tryptophan of ULM peptide. Thus, Arg198 and Glu149 forms the tryptophan-binding pocket along with Phe200. However, as the piperazine moiety extends in the direction opposite

to the peptide binding region, the salt-bridge between Arg198 and Glu149 is not formed and thus the tryptophan-binding pocket is not completely formed.

6.3.4 Structure based hit optimization

Based on the crystal structure of the Puf60-UHM in complex with TOK116, it was clear that the R1, R2 and R3 groups make crucial contacts with the protein. Therefore, it was not surprising that any changes made to these groups led to decrease in binding affinity. However, the methoxy group at position 7 of the compound points towards the arginine-binding pocket of the UHM. Therefore, new compounds were designed with substitutions at the position 7 methoxy group (Position R4 in **Figure 48**) of TOK116.

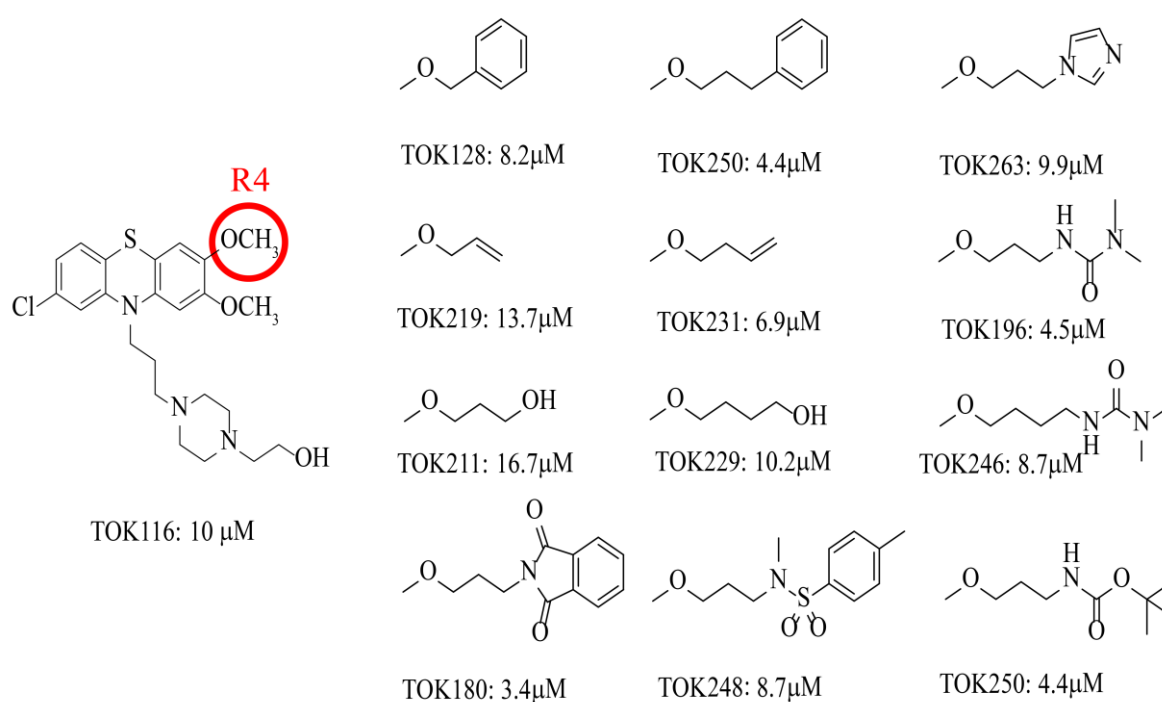


Figure 51. R4 group substituents of TOK116

Substituents of the R4 group are listed along with the chemical formula and the affinity of the corresponding inhibitor. The affinities were determined by AlphaScreen assay.

The groups attached at R4 ranged from bulky unsaturated ring systems to linear saturated and unsaturated alkyl chains with charged groups (**Figure 51**). The purpose of designing these substituents was to fill up the arginine-binding pocket completely to make hydrophobic contacts in the pocket along with salt-bridges and charged interactions with the acidic residues present in the pocket.

Addition of benzyl group (TOK128) or imidazole connected via a propyl linker (TOK263) did not lead to any significant increase in the affinity of the inhibitors. However, addition with benzene connected with a propyl linker led to a ~ 2 -fold increase in binding affinity. Substitution of R4 with unsaturated propyl (TOK219) or butyl (TOK231) groups and saturated linkers such as propanol (TOK211) or butanol (TOK229) did not contribute positively to the binding affinity.

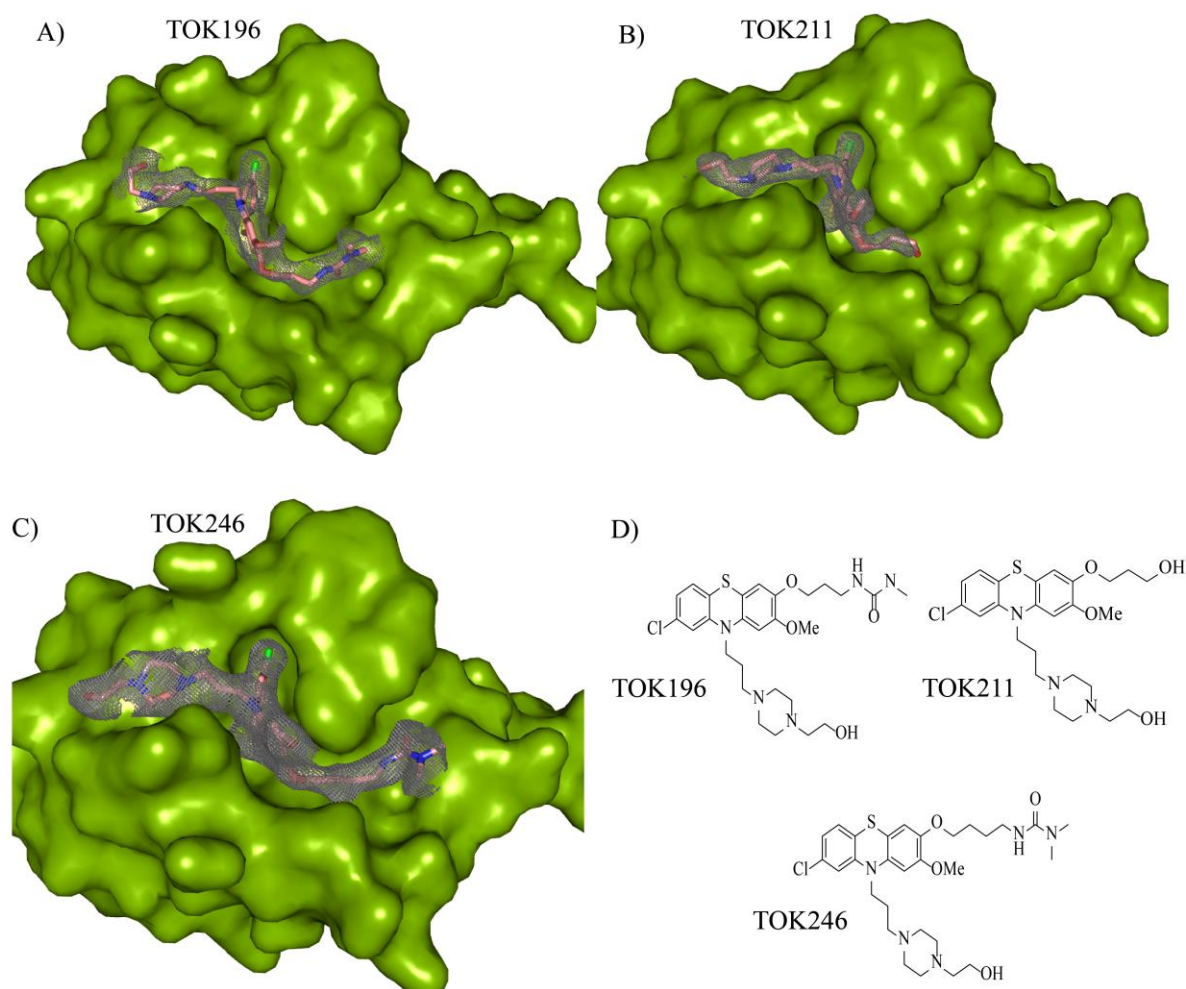


Figure 52 Crystal structure of Thx-PUF60 in complex with inhibitors having R4 substituents.

The electron density for A) TOK196 and B) TOK211 inhibitor is contoured at 1σ whereas for C) TOK246 is contoured at 0.5σ . D) The chemical structures of these compounds are shown. The R4 substituents from all the three compounds occupy the arginine-binding pocket.

In order to confirm that the R4 substituents indeed occupy the arginine-binding pocket, crystal structures of PUF60 in complex with TOK196, TOK211 and TOK246 were solved (**Figure 52**) (**Table 10**). These structures showed clear and unambiguous electron density for

the inhibitors along with the electron density of the R4 substituents in the arginine-binding pocket. The substituted R4 group occupied the arginine-binding pocket. The most crucial contact made by the R4 group was the hydrophobic interaction between alkyl part of the R4 group and the hydrophobic cavity of the arginine-binding pocket. No charged interactions were made between the charged part of the R4 group and the charged residues of the pocket. Therefore, it was not surprising that introducing the R4 group to occupy the arginine-binding pocket did not substantially increase the affinity of the new compounds in the absence of strong charged interactions.

Table 10. Data collection and refinement statistics for Thx PUF60-small molecule inhibitor complexes.

	TOK196	TOK211	TOK246
Resolution range	43.17 - 2.42 (2.51 - 2.42)	42.68 - 1.89 (1.96 - 1.89)	39.04 - 2.4 (2.49 - 2.4)
Space group	P 21 21 2	P 21 21 2	P 21 21 2
Unit cell	80.11 88.71 148.2 90 90 90	80.76 89.14 147.9 90 90 90	80.86 89.14 150.1 90 90 90
Unique reflections	40240 (3719)	78922 (5264)	43150 (4235)
Completeness (%)	0.98	0.91	1.00
Wilson B-factor	46.82	12.52	36.27
Rwork	0.2522 (0.4237)	0.3219 (0.7003)	0.2423 (0.3324)
Rfree	0.2852 (0.4354)	0.3416 (0.7459)	0.2794 (0.3393)
RMS(bonds)	0.023	0.025	0.022
RMS(angles)	1.57	1.75	1.59
Ramachandran favored (%)	98	98	98
Ramachandran allowed (%)	1.4	1.1	1.2
Ramachandran outliers (%)	0.59	0.72	0.36
Average B-factor	37.00	35.39	43.56
macromolecules	36.67	35.31	43.33
ligands	90.77	51.97	80.64

Statistics for the highest-resolution shell are shown in parentheses.

6.4 UHM inhibitors stall spliceosome assembly

To confirm the *in vitro* activity of the developed inhibitors, IgM pre-mRNA splicing assays were carried out for TOK116, TOK180, TOK217, TOK231 and TOK248. These inhibitors were chosen as they had the highest affinity in all the compounds developed in this study. In addition, they also represent different classes of substituents at R4 position. TOK116

has methoxy at R4 position whereas TOK180 and TOK248 have bulky groups attached to the R4 position. TOK217 and TOK231 have linear saturated and unsaturated alkyl chains at the R4 position, respectively.

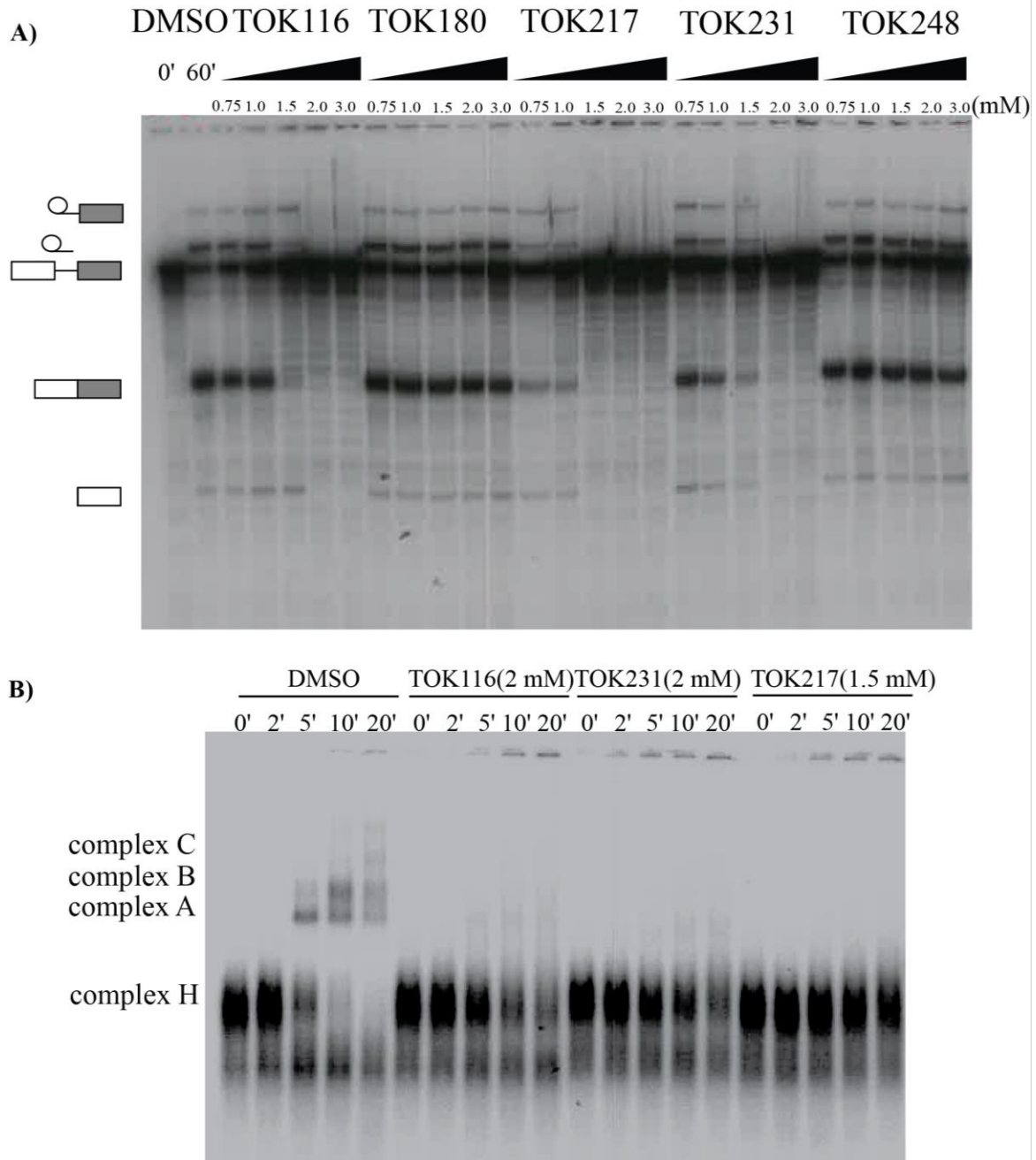


Figure 53. *In vitro* splicing assays with small molecule inhibitors

In vitro splicing assays were carried out on IgM pre-mRNA with five different inhibitors showing highest affinity for SPF45 UHM domain in AlphaScreen assay. A) Of the five inhibitors tested, TOK116, TOK217 and TOK 231 showed complete inhibition of splicing reaction at 1.5 mM, 1.5 mM and 2 mM inhibitor concentration respectively, whereas TOK180 and TOK248 did not show any inhibition. B) The splicing complexes were separated on a 2% agarose gel. The TOK116, TOK231 and TOK217 inhibited the Complex A formation.

The different classes of the substituents were chosen as although these compounds show binding to the purified UHM domains it is not necessary that they bind the UHM domain and inhibit splicing in the cell extracts. From the splicing assays, it was clear that only compounds TOK116, TOK217 and TOK231 showed inhibition of the IgM pre-mRNA whereas TOK180 and TOK248 did not show any inhibition even at 3 mM concentration (**Figure 53A**). TOK116 and TOK231 showed complete inhibition at 2 mM compound concentration whereas TOK217 inhibited splicing completely at 1.5 mM. This was surprising given that TOK116, TOK231 and TOK217 show binding affinity to the UHM domains in the similar range.

To further confirm that the splicing inhibition observed in the splicing assays is not due to any other factors such as non-specific precipitation of the proteins in the cell extract, the splicing complexes were separated on a 2% agarose gel. All the three inhibitors showing inhibition of pre-mRNA splicing inhibited the complex A formation. As mentioned previously, during complex E/H to complex A transition, SF1 which interacts with the UHM domain of U2AF65 is replaced by SF3b155. Therefore, the compounds indeed seem to inhibit this interaction confirming that these inhibitors inhibit the splicing by inhibiting the UHM-ULM interactions in the cell extracts.

In addition, it should be noted that TOK116 and TOK231 show very faint bands for complex A whereas TOK217 shows a complete inhibition of the complex A. Therefore, it can be concluded that TOK217, containing substitution of unsaturated propyl linker at position 7 in 7,8 dimethoxy perphenazine, is the most potent compound amongst all the developed compounds in this project.

6.5 UHM inhibitors target all UHM domains

As the crystal structure of PUF60 UHM domain with the small molecule inhibitors showed that the inhibitors bind to the highly conserved region in the tryptophan-binding pocket, it was probable that the inhibitors will bind to other UHM domains as well. To understand the selectivity of these developed inhibitors, the affinity of all the inhibitors was determined against three different UHM domains by AlphaScreen assay.

The EC₅₀ values of all the inhibitors against PUF60, U2AF65 and SPF45 UHM domains are shown in **Figure 54**. Each inhibitor binds to these three UHMs with similar affinity. Therefore, it can be concluded that none of the inhibitor selectively targets the UHM domains.

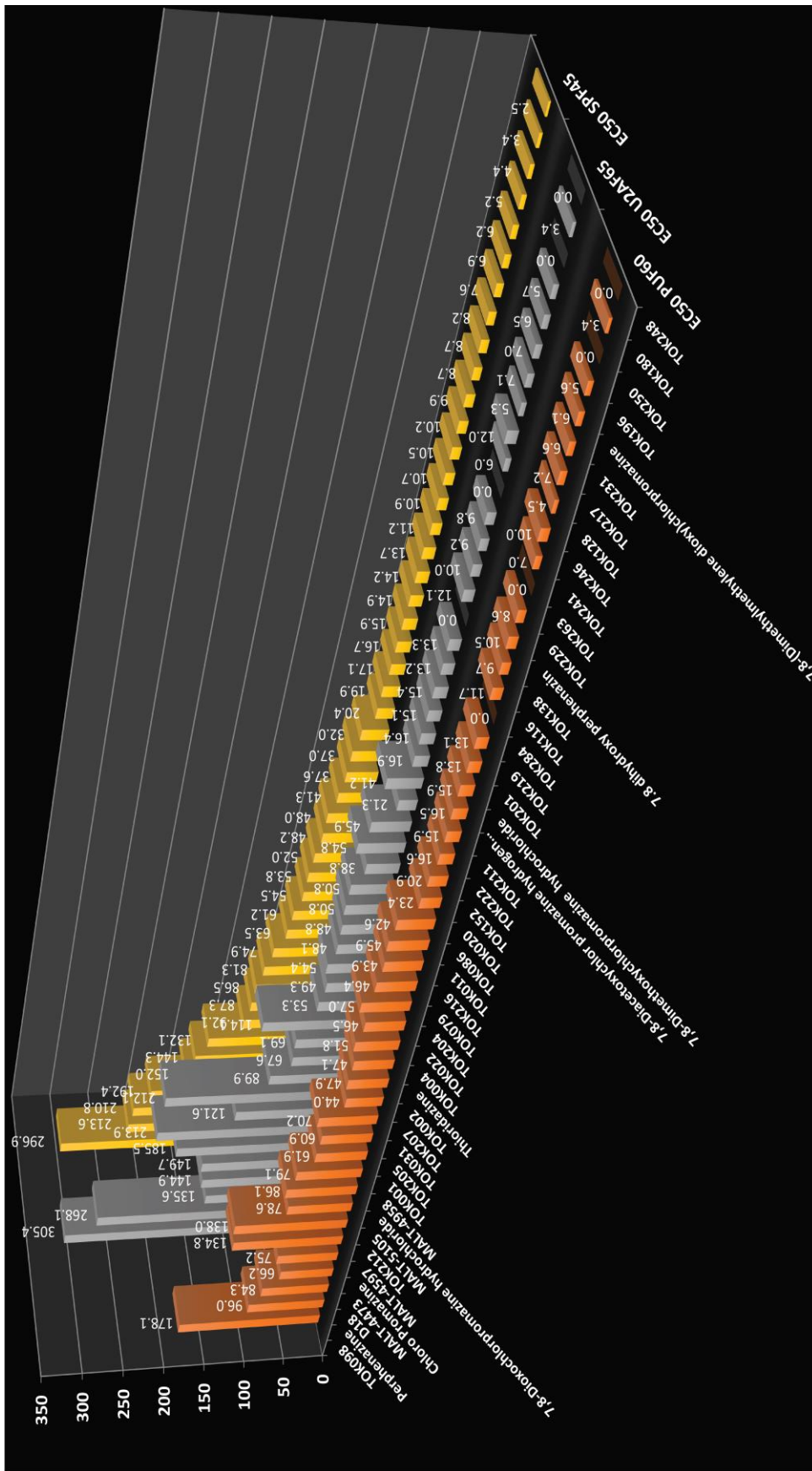


Figure 54. Selectivity of UHM inhibitors for PUF60, U2AF65 and SPF45 UHM domains

EC₅₀ values determined by AlphaScreen for different UHM inhibitors against PUF60, U2AF65 and SPF45 UHM domains are plotted. All the values are in μM. Values which have not been determined are shown as 0.0

This was expected as the inhibitor binds to the conserved pocket of the UHM domain and there are no contacts from the inhibitor to the variable regions of the UHM domain.

As these inhibitors stall the spliceosome assembly at complex A formation stage, based on the results of the selectivity screening it cannot be concluded which UHM-ULM interaction is inhibited by the small molecule inhibitors. For complex E formation, UHM-ULM interactions between U2AF65 UHM-SF1 ULM and U2AF35 UHM- U2AF65 ULM are required. During complex E to complex A transition, the SF1 ULM is replaced with SF3b155 ULM involved in the interaction with the U2AF65 UHM. Therefore, it is possible that the inhibitors target both U2AF65 and U2AF35 UHMs. However, since U2AF35 UHM-U2AF65 ULM interaction has low nanomolar binding affinity, it is unlikely that this interaction is affected by these inhibitors, suggesting that U2AF65 UHM-SF1 ULM interaction is the more likely candidate for inhibition in the splicing assays.

These results are also a first proof of principle showing that by targeting UHM-ULM interactions, spliceosome assembly can be stalled at initial stages where most of the splicing regulation takes place. A number of splicing inhibitors have been reported in recent years (Bonnal et al. 2012; Zaharieva et al. 2012). However, the molecular targets and mode of action of these inhibitors are poorly understood. Notably, most of these inhibitors are natural products, which renders their chemical modification and optimization difficult. In addition, most of these inhibitors interfere with the splicing reaction at later stages. In this respect, the UHM inhibitors developed in this thesis are distinct in two aspects: 1) the structural and rational design approach is based on detailed knowledge of the target, and the mode of action involving UHM inhibition, which is confirmed by our splicing assays, 2) The inhibitor opens novel ways to modulate splicing and interfere with spliceosome assembly at early stages, where alternative splicing is typically regulated by alternative splicing factors. Importantly, the fact that the UHM inhibitor is able to stall spliceosome assembly at an early stage opens the possibility for biochemical and structural studies of very early splicing complexes.

Conclusions and Outlook

Alternative splicing of the pre-mRNA transcripts by the spliceosome machinery is a crucial process. Misregulation of this process has been associated with several diseases. The process is highly regulated by several regulatory proteins for proper functioning of the cell. Here, structural insights into the interaction of TIA-1 with the *Fas* intronic RNA and the protein-protein interaction mediated by the TIA-1 RRM1 domain with the U1 snRNP specific protein U1C were obtained using an integrated structural biology approach.

In the absence of RNA, the three RRM domains of TIA-1 tumble independently as shown by the NMR relaxation data. However, transient inter-domain interactions between the three domains cannot be ruled out from the current results as a recent study has shown the thermodynamic stabilization of RRM3 in the presence of RRM2 (Aroca et al. 2011). From the NMR structure of RRM1 domain, it is clear that the RNA binding interface of RRM1 domain harbors negatively charged residues. This explains the negligible binding of RNA to the RRM1 domain in isolation. However, all the three RRM domains of TIA-1 tumble together in the presence of RNA suggesting a synergistic effect of the three RRM domains for RNA binding. This is corroborated by the SAXS data, where the radius of gyration of the TIA-1 RRM1,2,3 domains decreases by 38% in the presence of RNA compared with free RRM1,2,3.

The SAXS data of the U1C protein shows a concentration dependent oligomerization of the protein with a very weak dissociation constant. U1C has been shown to form dimers both *in vivo* and *in vitro* (Gunnewiek et al. 1995). Therefore, the oligomeric species seen in the SAXS experiments probably represents the dimerization of the protein. The SAXS data also agrees more with the crystal structure of the U1C suggesting that the NMR structure of the U1C protein is partially wrong. The zinc finger of the U1C superimposes well in the crystal and NMR structure. However, the helix B in NMR structure is split into helix B and C and the helix C folds back onto the helix B. The inaccuracies in the NMR structure could be justified if the NMR structure was calculated assuming a monomer, preventing the structure calculation program from converge onto a dimer.

NMR titration experiments with the RRM1 and U1C protein identified the minimal region from U1C (helix B) which is required for its interaction with the RRM1 domain. The crystal structure of the RRM1/U1C (30-61) fusion protein identified the molecular details of this weak interaction. This structure gives first structural insights into the interaction of TIA-1 and U1C and the model obtained by superposition of the U1C 30-61 helix structure in the

fusion protein and the U1C from U1 snRNP structure provides the expected mode of binding of TIA-1 to U1 snRNP.

Although these data explain the molecular details of TIA-1 RRM1-U1C interaction, in the absence of the U1 snRNP-TIA-1 complex structure, it remains to be seen at the structural level how the full-length TIA-1 protein interacts with the U1 snRNP protein and the intronic RNA. The primary interaction between TIA-1 and U1C is mediated by the Q-rich domain of the TIA-1 protein. Therefore, it would be crucial to understand the structural details of the TIA-1 Q-rich domain-U1C interaction.

The data presented in this thesis will pave the way for designing the approaches for stabilization of the ternary complex between TIA-1-U1 snRNP complex- intronic RNA and thus structural characterization of the whole complex using structural biology techniques. Several ways could be envisaged for stabilization of the TIA-1, U1 snRNP complex *in vitro* based on the crystal structure of the RRM1-U1C 30-61. These could be fusion protein of the U1C C-terminal linked to the N-terminal of the TIA-1 using a flexible GS linker or cysteine engineering to link the RRM1 of TIA-1 with the U1C protein using a cystine bond.

UHM domains present in many splicing factors play a crucial role in the regulation of the pre-mRNA splicing by mediating UHM-ULM interactions. In addition, UHM containing proteins have been implicated in many human disorders. However, whether the UHM domain itself is responsible for the disorders in these proteins remains to be determined.

The cyclization of the native ULM peptide motif presents an ideal strategy to develop peptide inhibitors of these domains as the β -turn formed by the ULM peptide on binding to the UHM domain is stabilized upon cyclization. This is confirmed by the crystal structure of the SPF45 UHM domain-P3 cyclic peptide complex. In addition, targeting the RXF motif in the UHM domains provides the required specificity and selectivity to the cyclic peptide inhibitor making it useful for distinguishing between different UHM domains.

The small molecule library screening identified Phenothiazine as the inhibitor of the UHM domains. Phenothiazine is also a suitable candidate for further optimization as many of its derivatives are approved drugs. Crystal structure of the PUF60 UHM domain with the inhibitors identified the arginine binding pocket in the UHM domain as the site for further optimization of the small molecules. Optimization of the inhibitors to explore the arginine binding pocket lead to ~ 4 fold increase in the affinity compared to the parent compound.

The cyclic peptide and small molecule inhibitors developed here are first examples of the UHM domain inhibitors. These inhibitors inhibited the complex A formation of the spliceosome assembly in line with the role of the UHM-ULM interaction in the early spliceosome assembly. These results are also a first proof of principle that the spliceosome assembly can be inhibited by targeting the UHM-ULM interactions.

The developed inhibitors however showed decrease in the binding affinity in cellular extract, probably either due to their interaction with other UHM domains or non-specific interaction with other proteins. Therefore, the inhibitors could be still optimized for better activity in the cell extracts or inside the cell. Although the cyclic peptide showed high selectivity for the SPF45 UHM domain, the small molecules lack the selectivity and inhibit the UHM-ULM interaction in several UHM domains. This is because the small molecules target the highly conserved regions of the UHM domain. Nevertheless, the developed small molecules provide good starting point/precursor molecules that could be further developed into selective and potent UHM domain inhibitors.

Appendix

Protein sequences

Extra residues left after TEV cleavage are shown in red.

SPF45 UHM domain (301-400):

GAMGKCPTKVLLLRNMVGAGEVDEDLEVETKEECEKYGKVGKCVIFEIPGAPDDEA
VRIFLEFERVESAIKAVVDLNGRYFGGRVVKACFYNLDKFRVLDLAEQ

U2AF65 UHM domain (371-471):

GAMGPLGSTEVLCMLNMVLPEELLDDDEEYEEIVEDVRDECSKYGLVKSIEIPRPVDG
VEVPGCGKIFVEFTSVFDCQKAMQGLTGRKFANRVVVTKYCDPDSYHRRDFW

Thioredoxin-Puf60 UHM domain:

MKHHHHHHHPMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIAD
EYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSCKGQLKEFLD
ANLAGSAMESTVMVLRNMVDPKDIDDDLEGEVTEECGKFGAVNRVIIYQEKQGE
DAEIIVKIFVEFSIASETHKAIQALNGRWFAGRKVVAEVYDQERFDNSDLA

TIA-1 Full length (1-319):

GAMEDEMPKTLVVGNSRDVTEALILQLFSQIGPCKNCKMIMDTAGNDPYCFVEFH
EHRHAAAALAAMNGRKIMGKEVKVNWATTPSSQKKDTSNHFHVFGDLSPEITTED
IKAAFAPFGRISDARVVKDMATGKSKGYGFVSFFNKWDAENAIQQMGGQWLGGRRQ
IRTNWATRKPAPKSTYESNTKQLSYDEVVNQSSPSNCTVYCGGVTSGLTEQLMRQT
FSPFGQIMEIRVFPDKGYSFVRFNSHESAAHAIVSVNGTTIEGHVVKCYWGKETLDMI
NPVQQQNQIGYPQPYGQWGQWYGNAQQIGQYMPNGWQVPAYGMYGQAWNQQG
FNQTQSSAPWMGPNYGVQPPQGQNGSMLPNQPSGYRVAGYETQ

TIA-1 RRM1 (1-92):

GAMEDEMPKTLVVGNSRDVTEALILQLFSQIGPCKNCKMIMDTAGNDPYCFVEFH
EHRHAAAALAAMNGRKIMGKEVKVNWATTPSSQKKDTS

TIA-1 Q-rich domain (274-319)

GAMAKETLDMINPVQQQNQIGYPQPYGQWGQWYGNAQQIGQYMPNGWQVP

TIA-1 RRM1+Q-rich domain:

GAMEDEMPKTLYVGNLSRDVTEALILQLFSQIGPCKNCKMIMDTAGNDPYCFVEFH
EHRHAAAALAAMNGRKIMGKEVKVNWATTPSSQKKDTSKETLDMINPVQQNQIG
YPQPYGQWGQWYGNAQQIGQYMPNGWQVPAYGMYGQAWNQQGFNQTQSSAPW
MGPNYGVQPPQGQNGSMLPNQPSGYRVAGYETQ

TIA-1 RRM1,2,3 (1-274):

GAMEDEMPKTLYVGNLSRDVTEALILQLFSQIGPCKNCKMIMDTAGNDPYCFVEFH
EHRHAAAALAAMNGRKIMGKEVKVNWATTPSSQKKDTSNHFHVFGDLSPEITTED
IKAAFAPFGRISDARVVKDMATGKSKGYGFVFFNKWDAENAIQQMGGQWLGGGRQ
IRTNWATRKPAPKSTYESNTKQLSYDEVVNQSSPSNCTVYCGGVTSGLTEQLMRQT
FSPFGQIMEIRVFPDKGYSFVRFNSHESAAHAIVSVNGTTIEGHVVKCYWGK

TIA-1 RRM1-GS15-U1C30-61

GAMEDEMPKTLYVGNLSRDVTEALILQLFSQIGPCKNCKMIMDTAGNDPYCFVEFH
EHRHAAAALAAMNGRKIMGKEVKVNWATTPSSQKKDTSGSGGSGGSGGSGGSGH
KENVKDYYQKWMEEQAQSLIDKTAAAFQQGK

TIA-1 RRM1-GS10-U1C30-61

GAMEDEMPKTLYVGNLSRDVTEALILQLFSQIGPCKNCKMIMDTAGNDPYCFVEFH
EHRHAAAALAAMNGRKIMGKEVKVNWATTPSSQKKDTSGSGGSGGSGSMPKFYCD
YCDTYLTHDSPSVRKTHCSGRKHKENVKDYYQKWMEEQAQSLIDKTAAAFQQGK

NMR chemical shift assignments of TIA-1 RRM1

Res. No	Atom Name	Assignment (ppm)
1	C	177.099
1	N	119.199
1	CA	56.268
1	CB	32.451
1	HA	4.335
1	HB2	2.021
1	HB3	2.02
2	C	177.025
2	N	120.778
2	CA	57.593
2	CG	36.479
2	CB	29.899
2	H	8.444
2	HA	4.139
2	HG3	2.291
2	HG2	2.223
2	HB3	1.983
2	HB2	1.976
3	C	176.434
3	CG	176.368
3	N	120.138
3	CA	55.048
3	CB	41.096
3	H	8.167
3	HA	4.473
3	QB	2.628
4	N	119.462
4	CA	56.136
4	CG	36.284
4	CB	30.209
4	H	8.042
4	HA	4.161
4	QG	2.156
4	HB2	2.021
4	HB3	1.837
5	N	120.762
5	CA	53.719
5	CG	32.474
5	CB	31.346
5	H	7.94
5	HA	4.263
5	HG2	2.423
5	HG3	2.268
5	QB	1.737
6	C	177.884
6	CA	63.89
6	CD	50.417
6	CB	32.59
6	CG	27.079
6	HA	4.39
6	HD2	3.423
6	HD3	3.005
6	HB2	2.171
6	HB3	1.779
6	HG2	1.395
6	HG3	0.983
7	C	174.799
7	N	118.504
7	CA	56.773
7	CE	42.553
7	CB	34.9
7	CD	29.644
7	CG	24.98
7	H	8.863
7	HA	4.808
7	QE	3.154
7	QB	2.509
7	QD	1.666
7	QG	1.201
8	C	173.883
8	N	115.49
8	CB	69.526
8	CA	61.818
8	CG2	21.717
8	H	8.055
8	HA	5.608
8	HB	4.142
8	QG2	1.175
9	C	175.836
9	N	126.3
9	CA	53.123
9	CB	44.205
9	CG	26.569
9	CD1	23.12
9	CD2	23.119
9	H	9.549
9	HA	5.067
9	HB2	1.579
9	HB3	1.301
9	HG	0.851
9	QD1	0.07
9	QD2	0.067
10	CD1	130.576
10	CD2	130.576
10	N	123.954
10	CE1	115.096
10	CE2	115.096
10	CA	57.152
10	CB	40.455
10	H	9.269
10	QD	6.783
10	QE	6.153
10	HA	4.864
10	HB2	2.887
10	HB3	2.753
11	C	174.126
11	N	127.988
11	CA	60.843
11	CB	32.51
11	CG1	21.289
11	CG2	21.25
11	H	8.479
11	HA	4.551
11	HB	1.965
11	QG1	0.754
11	QG2	0.518
12	C	173.684
12	N	112.642
12	CA	43.266
12	H	9.192
12	HA2	4.639
12	HA3	3.626
13	C	173.992
13	N	116.787
13	CA	54.094

13	CB	38.62	17	HB2	2.743	22	CD1	26.187
13	H	9.187	17	HB3	2.451	22	CD2	26.187
13	HA	4.292	18	C	174.021	22	H	7.662
13	HB2	3.716	18	N	119.989	22	HA	4.144
13	HB3	2.515	18	CA	63.9	22	HB2	1.804
14	C	178.025	18	CB	31.846	22	HB3	1.672
14	N	113.129	18	CG1	22.959	22	HG	1.581
14	CA	53.73	18	CG2	22.959	22	QQD	0.974
14	CB	42.836	18	H	7.382	23	C	178.34
14	CG	26.562	18	HA	3.739	23	N	118.562
14	CD1	23.375	18	HB	2.089	23	CA	62.316
14	CD2	23.375	18	QG1	1.012	23	CB	35.059
14	H	7.073	18	QG2	0.92	23	CG1	26.355
14	HA	4.219	19	C	174.679	23	CG2	18.5
14	HB2	1.502	19	N	112.365	23	CD1	9.293
14	HB3	1.172	19	CB	72.192	23	H	7.611
14	QQD	0.663	19	CA	58.821	23	HA	3.732
14	HG	0.555	19	CG2	21.699	23	HB	2.117
15	C	175.744	19	H	6.563	23	HG12	1.467
15	N	115.085	19	HA	4.614	23	HG13	1.279
15	CB	63.929	19	HB	4.605	23	QG2	0.619
15	CA	57.378	19	QG2	1.19	23	QD1	0.616
15	H	8.352	20	C	177.671	24	C	180.155
15	HA	4.369	20	N	121.818	24	N	119.959
15	HB2	4.053	20	CA	59.881	24	CA	58.545
15	HB3	3.702	20	CG	37.088	24	CB	41.653
16	C	175.972	20	CB	30.29	24	CG	27.031
16	N	123.567	20	H	9.091	24	CD1	24.25
16	CA	58.473	20	HA	3.719	24	CD2	24.25
16	CD	43.331	20	QG	2.193	24	H	8.556
16	CB	29.699	20	HB2	2.025	24	HA	3.976
16	CG	27.849	20	HB3	1.955	24	QB	1.68
16	H	8.723	21	C	180.393	24	HG	1.532
16	HA	4.009	21	N	118.415	24	QD1	0.781
16	QD	3.212	21	CA	55.096	24	QD2	0.753
16	QG	1.904	21	CB	18.326	25	N	121.315
16	HB2	1.897	21	H	8.132	25	CA	59.57
16	HB3	1.812	21	HA	4.013	25	CG	33.857
17	C	176.091	21	QB	1.357	25	CB	28.476
17	N	115.928	22	C	178.752	25	H	8.237
17	CA	54.432	22	N	120.447	25	HA	3.923
17	CB	41.461	22	CA	57.918	25	QG	2.459
17	H	8.092	22	CB	42.142	25	QB	2.25
17	HA	4.585	22	CG	27.359	26	C	180.169

26	N	118.18	30	C	176.112	34	CG	24.453
26	CA	57.686	30	N	118.563	34	H	8.951
26	CB	42.822	30	CA	61.863	34	HA	3.981
26	CG	25.96	30	CB	37.062	34	QD	1.349
26	CD1	23.116	30	CG1	28.25	34	HB2	1.136
26	CD2	23.116	30	CG2	18.458	34	QG	1.01
26	H	8.272	30	CD1	11.171	34	HB3	0.57
26	HA	4.093	30	H	7.61	35	C	172.802
26	HB2	1.903	30	HA	3.886	35	N	111.653
26	HB3	1.316	30	HB	2.215	35	CA	53.112
26	HG	0.812	30	HG12	1.364	35	CB	42.099
26	QD1	0.788	30	HG13	1.216	35	H	7.575
26	QD2	0.785	30	QG2	0.828	35	HA	4.671
27	C	177.265	30	QD1	0.64	35	QB	2.631
27	CD1	129.087	31	N	105.528	36	C	174.736
27	CD2	129.087	31	CA	44.695	36	N	120.663
27	CE1	127.491	31	H	7.365	36	CA	57.505
27	CE2	127.491	31	HA2	3.768	36	CB	28.861
27	N	117.715	31	HA3	3.133	36	H	8.497
27	CA	62.792	32	C	177.815	36	HA	5.047
27	CB	37.946	32	CA	63.971	36	QB	2.574
27	H	9.07	32	CD	49.414	37	C	174.431
27	QD	7.66	32	CB	32.453	37	N	127.468
27	QE	6.999	32	CG	27.637	37	CA	54.592
27	HA	4.1	32	HA	4.418	37	CE	42.345
27	HB2	3.105	32	HD2	3.435	37	CB	35.464
27	HB3	2.997	32	HD3	3.372	37	CD	29.237
28	C	175.309	32	HB2	2.306	37	CG	24.768
28	N	114.752	32	QG	1.966	37	H	8.595
28	CB	63.212	32	HB3	1.889	37	HA	4.701
28	CA	60.742	33	C	173.649	37	QE	2.925
28	H	8.37	33	N	123.116	37	HB2	1.842
28	HA	5.13	33	CA	58.889	37	HB3	1.689
28	HB2	4.031	33	CB	29.325	37	QD	1.665
28	HB3	3.946	33	H	8.987	37	QG	1.388
29	C	177.346	33	HA	4.355	38	C	175.065
29	N	118.301	33	HB2	2.379	38	N	126.601
29	CA	57.7	33	HB3	2.024	38	CA	54.647
29	CG	34.183	34	C	175.812	38	CB	33.438
29	CB	28.581	34	N	124.83	38	CG	32.244
29	H	7.152	34	CA	57.613	38	H	9.434
29	HA	4.024	34	CE	41.999	38	HA	4.607
29	QG	2.675	34	CB	34.35	38	HG2	2.589
29	QB	2.066	34	CD	29.51	38	HG3	2.513

38	HB2	2.001	43	C	177.977	48	H	7.947
38	HB3	1.732	43	N	124.798	48	QD	6.664
39	C	174.381	43	CA	52.879	48	QE	6.654
39	N	129.715	43	CB	18.889	48	HA	5.037
39	CA	60.121	43	H	8.313	48	HB2	3.008
39	CB	37.005	43	HA	4.325	48	HB3	2.471
39	CG1	27.654	43	QB	1.402	49	C	172.304
39	CG2	18.258	44	C	174.437	49	N	115.037
39	CD1	12.084	44	N	107.336	49	CA	55.941
39	H	9.004	44	CA	45.62	49	CB	32.368
39	HA	4.006	44	H	8.138	49	H	9.014
39	HG12	1.132	44	HA2	4.059	49	HA	5.332
39	HG13	1.039	44	HA3	3.802	49	HB2	2.588
39	QD1	0.689	45	C	174.745	49	HB3	2.513
39	HB	0.635	45	N	118.482	50	C	176.271
39	QG2	0.631	45	CA	53.17	50	CD1	128.964
40	C	175.967	45	CB	39.798	50	CD2	128.964
40	N	123.698	45	H	8.25	50	CE1	128.835
40	CA	54.348	45	HA	4.803	50	CE2	128.835
40	CB	33.436	45	QB	2.847	50	N	118.944
40	CG	32.189	46	N	121.023	50	CA	55.908
40	H	8.109	46	CA	53.986	50	CB	41.831
40	HA	4.597	46	CB	40.599	50	H	8.5
40	HG2	2.589	46	H	8.663	50	QD	7.624
40	HG3	2.502	46	HA	4.651	50	QE	7.251
40	HB2	2.009	46	QB	2.721	50	HA	5.845
40	HB3	1.87	47	C	174.506	50	HB2	2.864
41	C	177.007	47	CA	63.178	50	HB3	2.816
41	N	122.969	47	CD	50.383	51	C	173.125
41	CA	53.687	47	CB	33.036	51	N	123.409
41	CB	41.696	47	CG	27.604	51	CA	61.308
41	H	8.467	47	HA	4.742	51	CB	34.443
41	HA	4.739	47	QD	3.752	51	CG1	21.992
41	HB2	2.82	47	HB2	2.297	51	CG2	21.992
41	HB3	2.752	47	QG	1.82	51	H	9.079
42	C	175.038	47	HB3	1.776	51	HA	4.372
42	N	113.754	48	C	172.55	51	HB	1.47
42	CB	69.154	48	CD1	131.545	51	QG1	0.29
42	CA	62.79	48	CD2	131.545	51	QG2	0.265
42	CG2	21.725	48	CE1	115.103	52	C	176.157
42	H	8.301	48	CE2	115.103	52	N	128.406
42	HB	4.333	48	N	115.051	52	CA	54.379
42	HA	4.266	48	CA	55.614	52	CG	36.071
42	QG2	1.218	48	CB	40.635	52	CB	32.457

52	H	8.486	56	HE1	7.998	61	CB	17.976
52	HA	5.124	56	HD2	7.22	61	H	7.417
52	QG	2.015	56	HA	4.406	61	HA	4.103
52	QB	1.811	56	HB2	3.31	61	QB	1.424
53	C	175.38	56	HB3	3.072	62	N	122.481
53	CD1	129.224	57	C	178.611	62	CA	55.093
53	CD2	129.224	57	N	115.699	62	CB	19.502
53	CE1	128.671	57	CA	58.625	62	H	8.374
53	CE2	128.671	57	CD	43.346	62	HA	3.818
53	N	126.518	57	CB	28.47	62	QB	1.357
53	CA	58.778	57	CG	26.546	63	N	121.18
53	CB	41.203	57	H	9.353	63	CA	59.224
53	H	9.101	57	HA	3.917	63	CB	41.626
53	QD	7.328	57	QD	3.05	63	CG	27.34
53	QE	7.15	57	QB	1.768	63	CD2	25
53	HA	4.649	57	QG	1.276	63	CD1	24.943
53	QB	3.009	58	CE1	135.504	63	H	7.894
54	CE1	135.802	58	N	119.745	63	HA	3.58
54	N	117.424	58	CD2	113.622	63	HB2	1.512
54	CD2	116.04	58	CA	58.147	63	HG	1.183
54	CA	60.29	58	CB	32.001	63	HB3	1.054
54	CB	31.179	58	HE1	7.823	63	QD1	0.466
54	H	8.143	58	HD2	7.004	63	QD2	0.189
54	HE1	7.947	58	H	6.63	64	C	179.724
54	HD2	6.675	58	HA	4.28	64	N	117.714
54	HA	4.379	58	HB2	3.247	64	CA	54.165
54	HB2	3.39	58	HB3	2.741	64	CB	18.534
54	HB3	3.3	59	C	178.174	64	H	6.952
55	C	177.017	59	N	122.638	64	HA	4.062
55	N	114.979	59	CA	54.801	64	QB	1.413
55	CA	54.724	59	CB	18.36	65	C	179.662
55	CG	36.288	59	H	6.984	65	N	117.264
55	CB	32.92	59	HA	2.925	65	CA	54.068
55	H	8.895	59	QB	1.479	65	CB	19.938
55	HA	4.928	60	C	180.55	65	H	7.893
55	QG	2.291	60	N	118.28	65	HA	4.212
55	HB2	2.104	60	CA	54.803	65	QB	1.315
55	HB3	1.905	60	CB	17.718	66	N	111.672
56	CE1	134.809	60	H	8.34	66	CA	54.118
56	N	124.932	60	HA	3.813	66	CG	32.595
56	CD2	122.088	60	QB	1.289	66	CB	32.561
56	CA	60.214	61	C	180.511	66	H	8.274
56	CB	28.373	61	N	120.95	66	HA	4.491
56	H	9.766	61	CA	54.539	66	HG3	2.589

66	HG2	2.505	71	N	125.97	75	CG	36.733
66	HB2	1.925	71	CA	61.153	75	CB	29.601
66	HB3	1.618	71	CB	39.548	75	H	9
67	C	176.629	71	CG1	26.411	75	HA	4.143
67	N	115.636	71	CG2	17.465	75	QG	2.106
67	CA	56.334	71	CD1	13.277	75	HB2	1.911
67	CB	38.532	71	H	8.48	75	HB3	1.673
67	H	7.258	71	HA	4.13	76	C	175.316
67	HA	4.306	71	HB	1.728	76	N	123.519
67	HB2	2.971	71	HG12	1.449	76	CA	61.47
67	HB3	2.883	71	HG13	0.901	76	CB	33.042
68	C	173.639	71	QG2	0.746	76	CG1	23.218
68	N	116.467	71	QD1	0.708	76	CG2	20.213
68	CA	45.845	72	C	176.365	76	H	8.512
68	H	8.641	72	N	126.407	76	HA	4.59
68	HA2	4.204	72	CA	56.303	76	HB	2.229
68	HA3	3.712	72	CG	33.242	76	QG2	1.049
69	C	175.742	72	CB	29.676	76	QG1	0.989
69	N	121.11	72	H	9.599	77	C	174.747
69	CA	56.047	72	HA	3.982	77	N	124.33
69	CD	43.146	72	QG	2.39	77	CA	54.428
69	CB	30.613	72	HB2	2.282	77	CE	42.309
69	CG	27.458	72	HB3	1.919	77	CB	34.401
69	H	7.787	73	C	173.754	77	CD	29.346
69	HA	4.15	73	N	101.918	77	CG	24.842
69	QD	3.167	73	CA	45.606	77	H	7.885
69	QB	1.772	73	H	8.399	77	HA	4.85
69	HG2	1.691	73	HA2	4.207	77	HE2	2.963
69	HG3	1.536	73	HA3	3.509	77	HE3	2.924
70	C	176.354	74	C	175.545	77	HB2	1.73
70	N	123.334	74	N	120.651	77	QD	1.729
70	CA	56.521	74	CA	55.502	77	HB3	1.643
70	CE	42.116	74	CE	41.491	77	QG	1.395
70	CB	32.907	74	CB	35.905	78	C	174.045
70	CD	29.642	74	CD	29.263	78	N	124.535
70	CG	26.176	74	CG	25.019	78	CA	59.848
70	H	8.617	74	H	7.448	78	CB	34.36
70	HA	4.361	74	HA	4.599	78	CG1	24.342
70	HE2	2.963	74	HB2	1.855	78	CG2	22.878
70	HB2	1.697	74	HB3	1.653	78	H	8.615
70	QD	1.561	74	QG	1.399	78	HA	5.478
70	HB3	1.472	75	C	175.964	78	HB	1.782
70	QG	1.061	75	N	126.243	78	QG2	1.033
71	C	175.807	75	CA	56.679	78	QG1	1.013

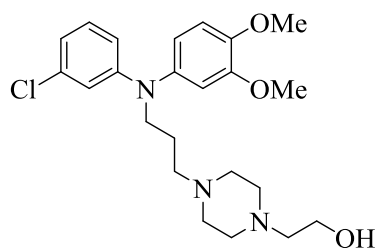
79	C	174.279	84	CA	63.791	88	QE	2.961
79	N	121.284	84	CD	51.065	88	HB2	1.798
79	CA	51.663	84	CB	32.15	88	HB3	1.726
79	CB	42.884	84	CG	27.559	88	QD	1.643
79	H	8.958	84	HA	4.368	88	QG	1.396
79	HA	4.821	84	HD2	3.787	89	C	176.232
79	HB3	2.891	84	HD3	3.652	89	N	123.472
79	HB2	2.704	84	HB2	2.23	89	CA	56.14
80	C	177.304	84	HG2	1.977	89	CE	42.297
80	CD1	124.843	84	HG3	1.877	89	CB	33.317
80	N	121.348	84	HB3	1.873	89	CD	29.212
80	CA	58.817	85	C	174.878	89	CG	24.699
80	CB	30.185	85	N	116.057	89	H	8.418
80	HE1	10.162	85	CB	63.742	89	HA	4.325
80	H	8.533	85	CA	58.716	89	QE	2.961
80	HD1	7.358	85	H	8.384	89	HB2	1.81
80	HA	4.802	85	HA	4.349	89	HB3	1.721
80	HB2	3.48	85	QB	3.827	89	QD	1.641
80	HB3	3.038	86	C	174.473	89	QG	1.398
81	C	177.268	86	N	117.533	90	C	176.348
81	N	125.215	86	CB	63.882	90	N	121.907
81	CA	52.505	86	CA	58.581	90	CA	54.497
81	CB	19.981	86	H	8.278	90	CB	41.199
81	H	8.717	86	HA	4.387	90	H	8.432
81	HA	4.587	86	QB	3.843	90	HA	4.645
81	QB	1.426	87	C	175.758	90	HB2	2.708
82	C	174.362	87	N	121.948	90	HB3	2.64
82	N	114.2	87	CA	55.797	91	C	173.934
82	CB	69.98	87	CG	33.859	91	N	113.556
82	CA	61.375	87	CB	29.48	91	CB	69.877
82	CG2	21.713	87	H	8.244	91	CA	61.431
82	H	8.375	87	HA	4.29	91	CG2	21.574
82	HA	4.43	87	QG	2.299	91	H	8.071
82	HB	4.162	87	HB2	2.061	91	HA	4.384
82	QG2	1.165	87	HB3	1.907	91	HB	4.33
83	N	118.979	88	C	176.355	91	QG2	1.17
83	CB	69.833	88	N	123.19	92	N	123.566
83	CA	59.852	88	CA	56.213	92	CB	64.766
83	CG2	21.29	88	CE	42.247	92	CA	60.3
83	H	8.307	88	CB	33.011	92	H	8.018
83	HA	4.547	88	CD	29.284	92	HA	4.225
83	HB	4.326	88	CG	24.734	92	QB	3.839
83	QG2	1.176	88	H	8.261			
84	C	177.178	88	HA	4.262			

NMR backbone chemical shifts of U1C 30-61

Residue Number	Atom Name	Assignment (ppm)						
30	CA	52.41	41	CA	56.822	51	H	8.256
30	CB	26.607	41	CB	34.918	51	N	123.516
30	H	8.332	41	H	7.985	52	CA	53.67
30	N	118.172	41	N	120.794	52	CB	29.009
31	CA	53.276	42	CA	54.537	52	H	8.191
31	CB	30.033	42	CB	25.542	52	N	121.734
31	H	8.268	42	H	8.005	53	CA	60.21
31	N	123.234	42	N	120.036	53	CB	66.277
32	CA	53.387	43	CA	53.749	53	H	8.186
32	CB	26.724	43	CB	28.851	53	N	113.953
32	H	8.514	43	H	8.033	54	CA	59.848
32	N	121.922	43	N	119.391	54	CB	66.304
33	CA	49.967	44	CA	54.616	54	H	7.933
33	CB	35.863	44	CB	26.645	54	N	115.734
33	H	8.461	44	H	7.921	55	CA	50.046
33	N	119.766	44	N	120.516	55	CB	15.929
34	CA	59.343	45	CA	54.695	55	H	8.098
34	CB	29.087	45	CB	26.172	55	N	125.672
34	H	7.985	45	H	8.092	56	CA	49.888
34	N	120.406	45	N	120.328	56	CB	16.087
35	CA	53.434	46	CA	50.204	56	H	7.986
35	CB	29.56	46	CB	14.905	56	N	121.922
35	H	8.25	46	H	7.867	57	CA	54.695
35	N	124.078	46	N	123.019	57	CB	35.785
36	CA	51.228	47	CA	53.592	57	H	7.916
36	CB	38.07	47	CB	25.463	57	N	118.547
36	H	8.127	47	H	8.009	58	CA	52.489
36	N	120.703	47	N	117.984	58	CB	26.408
37	CA	55.01	48	CA	55.877	58	H	8.033
37	CB	28.772	48	CB	60.604	58	N	121.453
37	H	7.986	48	H	8.033	59	CA	52.882
37	N	120.201	48	N	115.453	59	CB	26.408
38	CA	55.719	49	CA	52.489	59	H	8.186
38	CB	25.227	49	CB	38.779	59	N	121.172
38	H	8.015	49	H	7.859	60	CA	41.694
38	N	120.703	49	N	122.738	60	H	8.356
39	CA	53.749	50	CA	58.319	60	N	110.766
39	CB	28.693	50	CB	35.154	61	CA	54.695
39	H	8.042	50	H	7.704	61	CB	30.19
39	N	119.633	50	N	120.516	61	H	7.71
			51	CA	51.385	61	N	125.672
			51	CB	37.754			

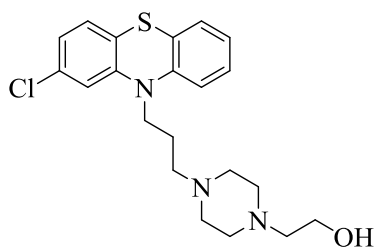
Chemical structures of the compounds

TOK098



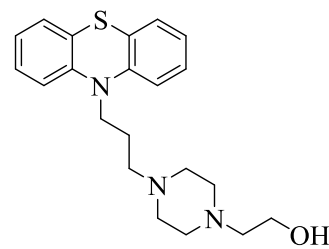
Chemical Formula: $C_{23}H_{32}ClN_3O_3$
Molecular Weight: 433.98

Perphenazine



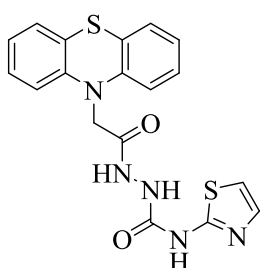
Chemical Formula: $C_{21}H_{26}ClN_3OS$
Molecular Weight: 403.97

D18



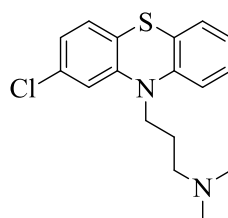
Chemical Formula: $C_{21}H_{27}N_3OS$
Molecular Weight: 369.53

MALT-4473



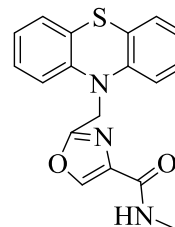
Chemical Formula: $C_{18}H_{15}N_5O_2S_2$
Molecular Weight: 397.47

Chlorpromazine



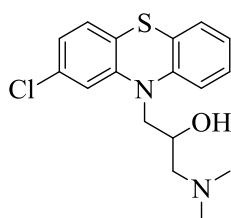
Chemical Formula: $C_{17}H_{19}ClN_2S$
Molecular Weight: 318.86

MALT-4597



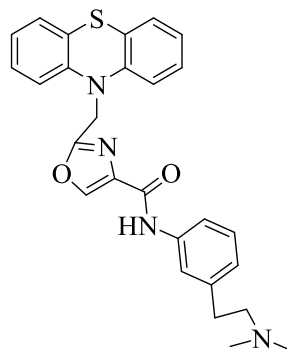
Chemical Formula: $C_{18}H_{15}N_3O_2S$
Molecular Weight: 337.40

TOK212



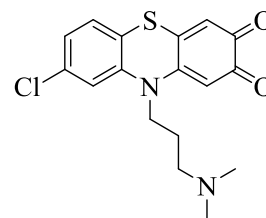
Chemical Formula: $C_{17}H_{19}ClN_2OS$
Molecular Weight: 334.86

MALT-5105



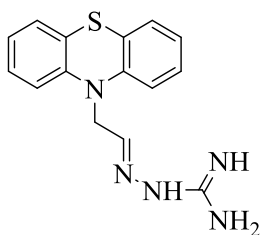
Chemical Formula: $C_{27}H_{26}N_4O_2S$
Molecular Weight: 470.59

7,8 Dioxochlorpromazine



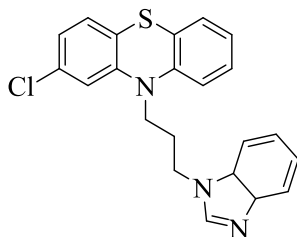
Chemical Formula: $C_{17}H_{17}ClN_2O_2S$
Molecular Weight: 348.85

MALT-4958



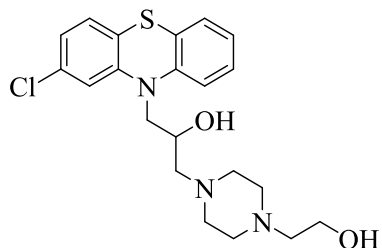
Chemical Formula: $C_{15}H_{15}N_5S$
Molecular Weight: 297.38

TOK001



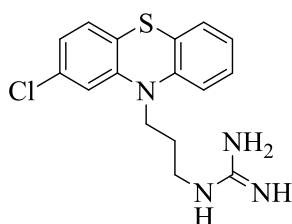
Chemical Formula: $C_{22}H_{20}ClN_3S$
Molecular Weight: 393.93

TOK205



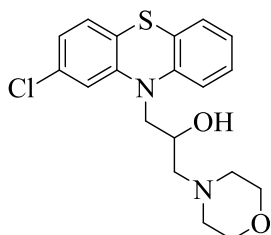
Chemical Formula: $C_{21}H_{26}ClN_3O_2S$
Molecular Weight: 419.97

TOK031



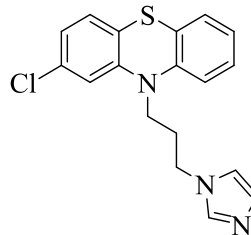
Chemical Formula: $C_{16}H_{17}ClN_4S$
Molecular Weight: 332.85

TOK207



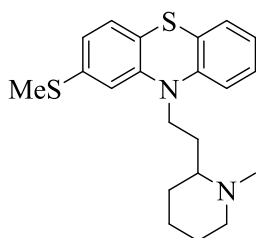
Chemical Formula: $C_{19}H_{21}ClN_2O_2S$
Molecular Weight: 376.90

TOK002



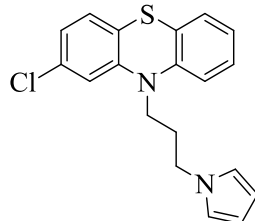
Chemical Formula: $C_{18}H_{16}ClN_3S$
Molecular Weight: 341.86

Thioridazine



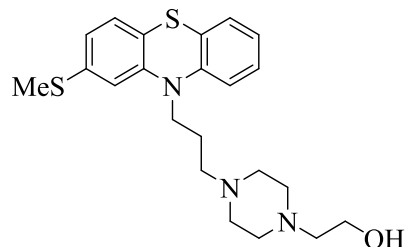
Chemical Formula: $C_{21}H_{26}N_2S_2$
Molecular Weight: 370.57

TOK004



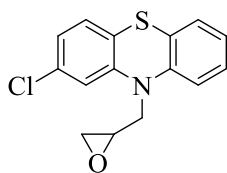
Chemical Formula: $C_{19}H_{17}ClN_2S$
Molecular Weight: 340.87

TOK022



Chemical Formula: $C_{22}H_{29}N_3OS_2$
Molecular Weight: 415.61

TOK204

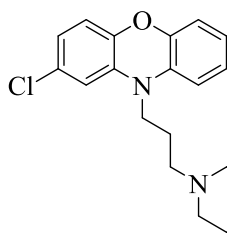


Chemical Formula:

 $C_{15}H_{12}ClNOS$

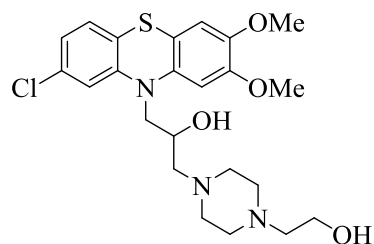
Molecular Weight: 289.78

TOK079

Chemical Formula: $C_{21}H_{26}ClN_3O_2$

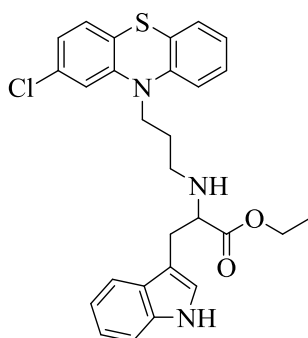
Molecular Weight: 387.91

TOK216

Chemical Formula: $C_{23}H_{30}ClN_3O_4S$

Molecular Weight: 480.02

TOK011

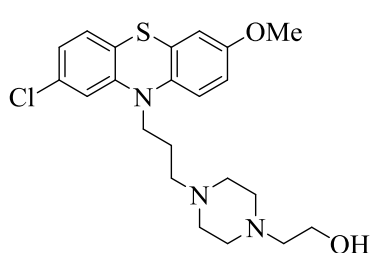


Chemical Formula:

 $C_{28}H_{28}ClN_3O_2S$

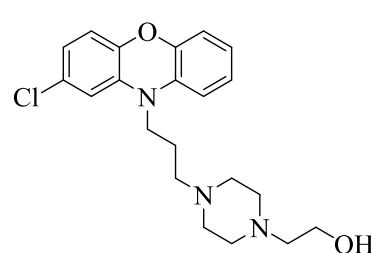
Molecular Weight: 506.06

TOK086



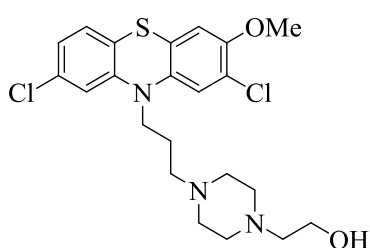
Molecular Weight: 433.99

TOK020

Chemical Formula: $C_{21}H_{26}ClN_3O_2$

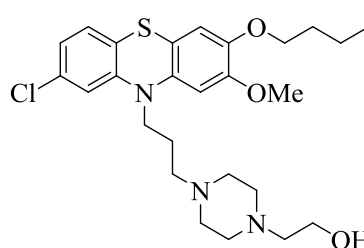
Molecular Weight: 387.91

TOK152

Chemical Formula: $C_{22}H_{27}Cl_2N_3O_2S$

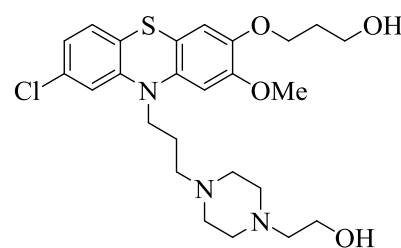
Molecular Weight: 468.44

TOK222

Chemical Formula: $C_{26}H_{36}ClN_3O_3S$

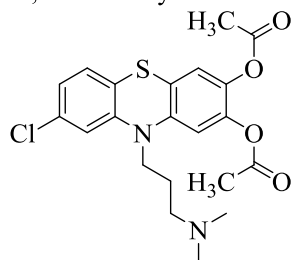
Molecular Weight: 506.10

TOK211

Chemical Formula: $C_{25}H_{34}ClN_3O_4S$

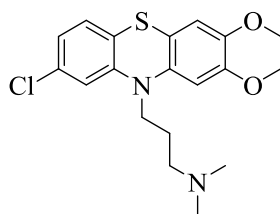
Molecular Weight: 508.07

7,8 Diacetoxy chlorPromazine



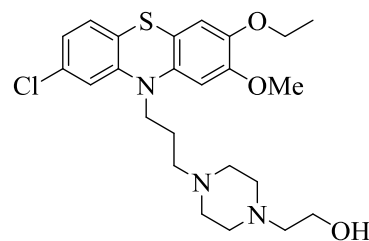
Chemical Formula: C₂₁H₂₃ClN₂O₄S
Molecular Weight: 434.94

7,8 Dimethoxy chlorPromazine



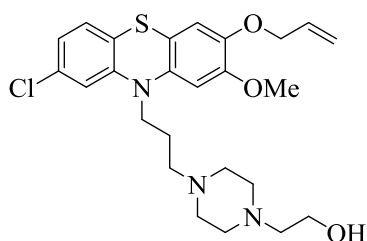
Chemical Formula: C₁₉H₂₃ClN₂O₂S
Molecular Weight: 378.92

TOK201



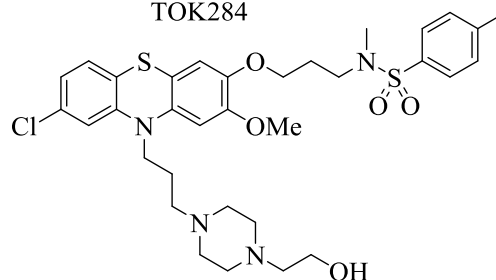
Chemical Formula: C₂₄H₃₂ClN₃O₃S
Molecular Weight: 478.05

TOK219



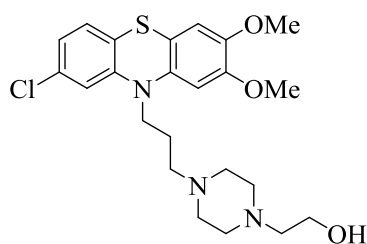
Chemical Formula: C₂₅H₃₂ClN₃O₃S
Molecular Weight: 490.06

TOK284



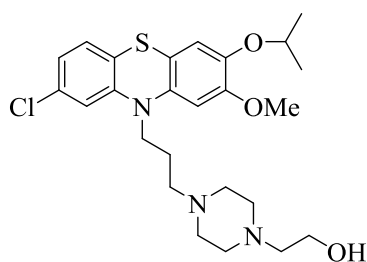
Chemical Formula: C₃₃H₄₃ClN₄O₅S₂
Molecular Weight: 675.30

TOK116



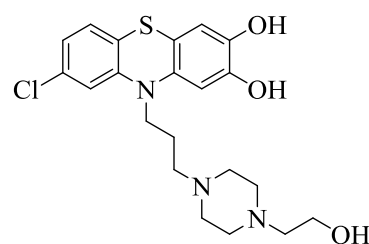
Chemical Formula: C₂₃H₃₀ClN₃O₃S
Molecular Weight: 464.02

TOK138

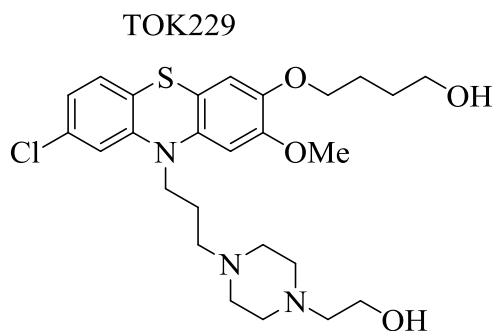


Chemical Formula: C₂₅H₃₄ClN₃O₃S
Molecular Weight: 492.07

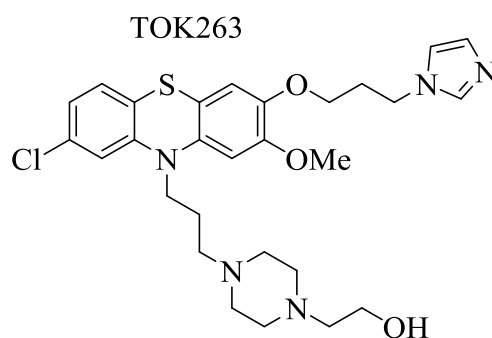
7,8 dihydroxy perphenazine



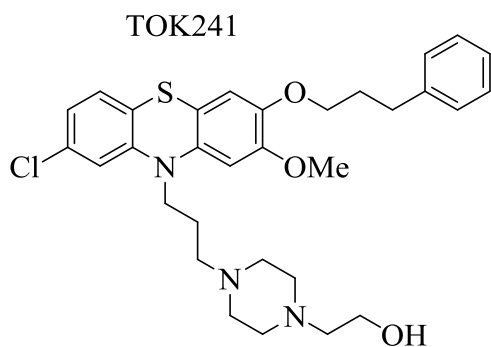
Chemical Formula:
C₂₁H₂₆ClN₃O₃S
Molecular Weight: 435.97



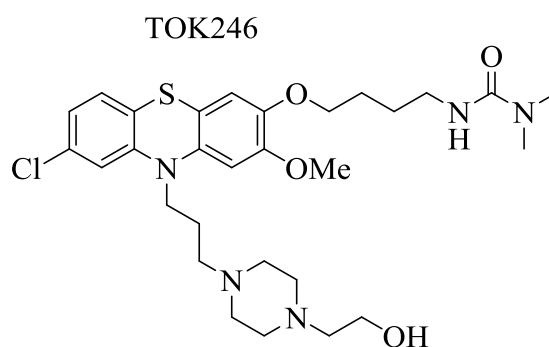
Chemical Formula: $C_{26}H_{36}ClN_3O_4S$
Molecular Weight: 522.10



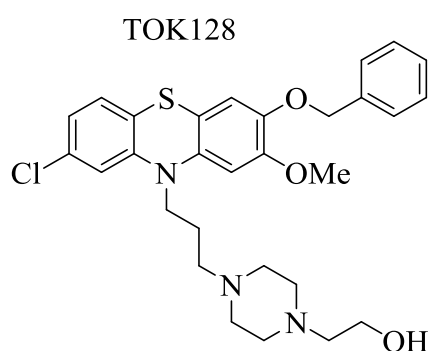
Chemical Formula: $C_{28}H_{36}ClN_5O_3S$
Molecular Weight: 558.14



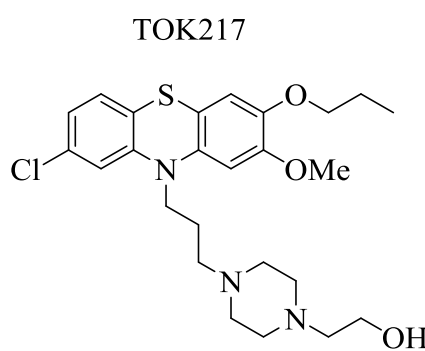
Chemical Formula: $C_{31}H_{38}ClN_3O_3S$
Molecular Weight: 568.17



Chemical Formula: $C_{29}H_{42}ClN_5O_4S$
Molecular Weight: 592.20

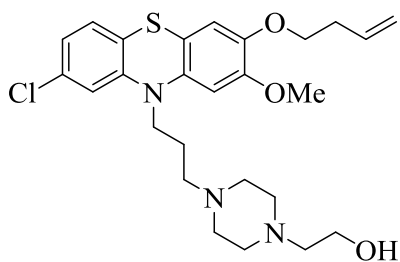


Chemical Formula: $C_{29}H_{34}ClN_3O_3S$
Molecular Weight: 540.12



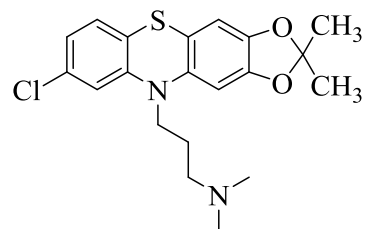
Chemical Formula: $C_{25}H_{34}ClN_3O_3S$
Molecular Weight: 492.07

TOK231



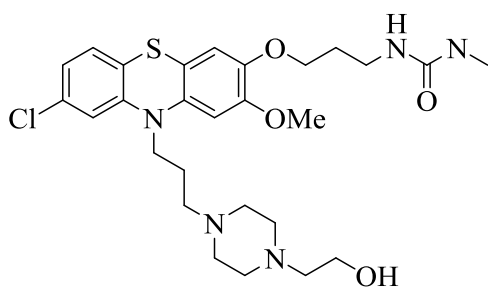
Chemical Formula: $C_{26}H_{34}ClN_3O_3S$
 Molecular Weight: 504.09

7,8(Dimethylmethylenedioxy)
 ChloroPromazine



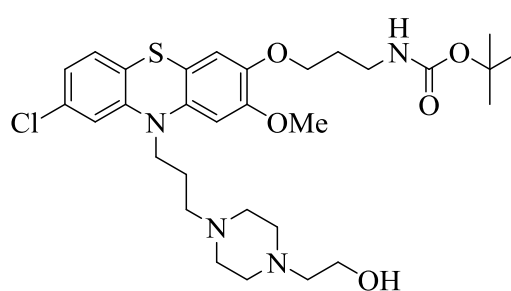
Chemical Formula: $C_{20}H_{23}ClN_2O_2S$
 Molecular Weight: 390.93

TOK196



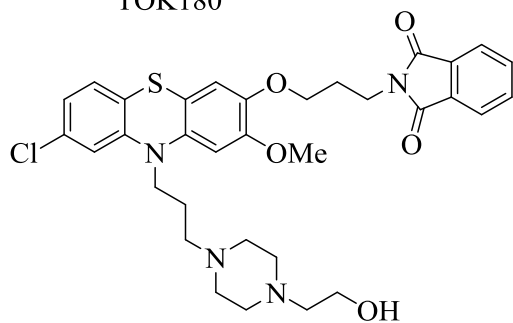
Chemical Formula: $C_{28}H_{40}ClN_5O_4S$
 Molecular Weight: 578.17

TOK250



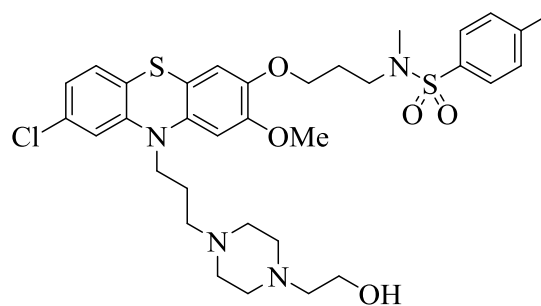
Chemical Formula: $C_{30}H_{43}ClN_4O_5S$
 Molecular Weight: 607.21

TOK180



Chemical Formula: $C_{33}H_{37}ClN_4O_5S$
 Molecular Weight: 637.19

TOK248



Chemical Formula: $C_{33}H_{43}ClN_4O_5S_2$
 Molecular Weight: 675.30

Abbreviations

1D, 2D, 3D	One-, Two-, Three-Dimensional
AEBSF	4- Benzenesulfonyl fluoride hydrochloride
BME	β -mercaptoethanol
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
FP	Fluorescence polarization
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear single quantum coherence spectroscopy
HTS	High throughput screening
IPTG	Isopropyl β -D-1-thiogalactopyranoside
K_D	Equilibrium Dissociation Constant
kDa	Kilo Dalton
LB	Lysogeny Broth Medium
LOPAC	Library of Pharmacologically Active Compounds
MWCO	Molecular weight cut-off
Ni^{2+}	Nickle
OD	Optical Density at 600nm Wavelength
PMSF	Phenylmethylsulfonyl fluoride
pre-mRNA	precursor messenger RNA
PTZ	Phenothiazine
QSAR	Quantitative structure-activity relationship
R_g	Radius of gyration
RMSD	Root mean square deviation
SAXS	Small angle X-ray scattering
TEV	Tobacco Etch Virus
TOCSY	Total correlation spectroscopy
TOKXXX	Name of the compounds synthesized along with the compound number
TROSY	Transverse relaxation optimized spectroscopy
UHM	U2AF homology motifs

List of Figures

Figure 1. Schematic overview of pre-mRNA splicing.....	12
Figure 2. Spliceosome assembly and pre-mRNA splicing	13
Figure 3. Splicing regulation.....	15
Figure 4. Schematic of Fas alternative splicing.....	17
Figure 5. Domain arrangement of TIA-1 and its role in Fas splicing.....	18
Figure 6. Available biochemical and structural information of TIA-1 and U1C proteins.....	20
Figure 7. Early stages of spliceosome assembly.....	23
Figure 8. UHM domains in various proteins	24
Figure 9 ¹ H spectrum of TIA-1 RRM1 domain.....	35
Figure 10 Schematic for protein backbone assignment	36
Figure 11 Spectral density for different Larmor frequencies and rotational correlation times (adapted from Understanding NMR spectroscopy, James Keeler; 2002).....	39
Figure 12 Behavior of T_1 and T_2 as a function of τ_c	40
Figure 13 Schematic to derive Bragg's Law.....	43
Figure 14 Schematic of the Ewald sphere.....	44
Figure 15 Schematic of the SAXS experiment setup.....	47
Figure 16. Principle of Fluorescence polarization assay	61
Figure 17. Schematic overview of the 384 plate used for the high-throughput screening	63
Figure 18. Schematic overview of the AlphaScreen technology.....	64
Figure 19. TIA-1 RRM1,2,3 tumble together in presence of RNA.....	68
Figure 20. TIA-1 RRM1,2,3 form a compact shape in the presence of RNA	70
Figure 21. NMR structure of TIA-1 RRM1.....	72
Figure 22. Structure analysis of TIA-1 RRM1	74
Figure 23. Static light scattering of U1C protein.....	76
Figure 24. Backbone assignment of U1C (1-61).	77
Figure 25. Concentration dependent dimerization of U1C.....	79
Figure 26. U1C is elongated in solution	80
Figure 27. U1C dimerization constant determined by ITC.....	81
Figure 28. U1C interacts with TIA-1 RRM1	82
Figure 29. U1C interacts with TIA-1	84
Figure 30. Backbone assignment of U1C 30-61 peptide.....	86
Figure 31. Interaction between U1C 30-61 peptide and TIA-1 RRM1	87

Figure 32. Binding affinity of RRM1-U1C interaction	89
Figure 33. Optimization of GS linker length between RRM1 and U1C peptide	91
Figure 34. Crystal structure of RRM1-U1C peptide complex	92
Figure 35. Interaction between RRM1 and U1C 30-61 peptide	94
Figure 36. Current understanding of of U1C protein	95
Figure 37. Model for TIA-1 U1 snRNP interaction	97
Figure 38. Crystal structure of SPF45 UHM-P3 cyclic peptide complex	102
Figure 39. Overview of the synthesized focused library of the cyclic peptides	104
Figure 40. NMR titration of P3 and P10 cyclic peptide into SPF45 UHM domain	106
Figure 41. <i>In vitro</i> splicing assays with P10 peptide	107
Figure 42. Separation of spliceosome complexes on agarose gel	108
Figure 43. Sequence alignment of UHM domains from different proteins	109
Figure 44. Development of Fluorescence polarization assay	113
Figure 45. Results of high throughput screening	114
Figure 46. Validation of HTS results by FP	116
Figure 47. Validation of HTS by NMR titrations	117
Figure 48. Substitutions at R1, R2 and R3 groups	119
Figure 49. Crystallization of HTS hits with Puf60 UHM	121
Figure 50. Structure analysis of Puf60 UHM-TOK116 complex	123
Figure 51. R4 group substituents of TOK116	124
Figure 52. Crystal structure Thx-PUF60 in complex with inhibitors having R4 substituents	125
Figure 53. <i>In vitro</i> splicing assays with small molecule inhibitors	127
Figure 54. Selectivity of UHM inhibitors for PUF60, U2AF65 and SPF45 UHM domains	129

List of Tables

Table 1. Diseases associated with UHM domains	25
Table 2. SAXS data collection and processing statistics for RRM1,2,3 free and RNA complexes	71
Table 3 Structural statistics for TIA-1 RRM1	73
Table 4. SAXS data collection and data processing statistics for U1C (1-61)	78
Table 5. Data processing and refinement statistics for RRM1_GS15_U1C 30-61 crystals	93
Table 6. Crystallographic data collection and refinement statistics for the SPF45-UHM/P3 complex.....	101
Table 7. Binding affinity of different cyclic peptides to the SPF45 UHM domain determined by isothermal titration calorimetry.....	105
Table 8. Dissociation constants (K_D) of the U2AF65 UHM domain determined by ITC	109
Table 9. Data collection and refinement statistics for Thioredoxin-Puf60 UHM crystals in complex with 7,8 Dimethoxy chlorpromazine and 7,8 Dimethoxy perphenazine	122
Table 10. Data collection and refinement statistics for Thx PUF60-small molecule inhibitor complexes.	126

Acknowledgements

First, I would like to thank my supervisor Prof. Dr. Michael Sattler for giving me an opportunity to work in his lab. During this thesis, I worked on several projects and the work done during this thesis would not have been possible without freedom provided by him for doing science. Working in his lab has benefited me at professional as well as personal level.

I cannot thank enough to Boehringer Ingelheim Fonds for the doctoral fellowship. Besides the generous monetary support, all the workshops, meetings and seminars organized by Boehringer Ingelheim Fonds have been immensely useful for my personal as well as scientific development. Being associated with the fellowship program not only exposed me to the exciting research being done by fellow researchers from wide range of fields but also opened up the way for being associated with its strong alumni network.

My sincere regards go to the members of my Thesis Advisory Committee, Prof. Dr. Dierk Niessing from Helmholtz Zentrum, Munich and Prof. Dr. Johannes Buchner from Technical University, Munich for their support and guidance during my doctoral studies.

My special thanks go to my various collaborators on different projects. First to Dr. Divita Garg and Dr. Iren Wang from whom I took over the SPF45 and TIA-1 projects respectively. Dr. Tomas Kubelka and Dr. Thorsten Bach at Technical University, Munich for their help in designing and synthesizing hundreds of compounds for the SPF45 small molecule project, to Dr. Tobias Kapp and Prof. Dr. Horst Kessler for their help in designing and synthesizing numerous cyclic peptides for the SPF45 cyclic peptide project. SPF45 could not have been a success without the help of Prof. Dr. Felix Hausch and his lab members specially Claudia Sippel during the small molecule library screening and to Dr. Cindy Will from Prof. Dr. Reinhard Lührmann's lab at Max Planck Institute for Biophysical Chemistry, Göttingen for all the *in vitro* assays done in this project. The SPF45 project also benefited immensely from the help of Dr. Kenji Schropp from Dr. Kamyar Hadian's lab at Helmholtz Zentrum, Munich during the AlphaScreen development assay.

Also, I cannot thank enough to Dr. Chris Oubridge and Prof. Dr. Kiyoshi Nagai from MRC Laboratory of Molecular Biology, United Kingdom for all the help in the TIA-1 project and generously sharing their previous data on this projects.

I would like to acknowledge Sam, Ralf, Gerd and Rainer for their help and for managing and maintaining the wonderful NMR and SAXS facilities and Waltraud for taking care of the administration, and keeping order in our scientifically organized small world.

With particular gratitude, I would like to thank my colleagues and friends in Munich specially Leo, Eleni, Carolina, Martin, Diana, Ashish, Johannes, Miriam, Arie, Ana, Anna, and Grzegorz and to all the Sattler group members for their friendly discussions about everything including science and providing a wonderful environment to work at HMGU and TUM. I would also like to thank Dr. Andre Mourao and Dr. Janosch Hennig for teaching me NMR during the early days and being patient with me.

I am greatly indebted to Arie, Ana and Astrid for taking care of the HMGU labs and maintaining everything at the proper place. I can surely say that they have saved countless hours for everyone working at HMGU by keeping things organized at the PEPF and preventing things going to chaos from time to time. Besides, many a times, a brief casual discussion with Arie regarding the protein purification problems and molecular biology have saved me several hours of literature digging and troubleshooting's.

Last but not the least, I owe my deepest gratitude to my mother and sister, who understood my need to leave India for higher education at a time when they need me the most and supported all my decisions that I took in my life. And to the love of my life, my friend and my life partner Komal, who have added several new dimensions to my life and without whom it has become difficult to imagine life.

References

- Ahlner A, Carlsson M, Jonsson BH, Lundstrom P. 2013. PINT: a software for integration of peak volumes and extraction of relaxation rates. *Journal of biomolecular NMR* **56**: 191-202.
- Aravind L, Koonin EV. 1999. G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends in biochemical sciences* **24**: 342-344.
- Aroca A, Diaz-Quintana A, Diaz-Moreno I. 2011. A structural insight into the C-terminal RNA recognition motifs of T-cell intracellular antigen-1 protein. *FEBS letters* **585**: 2958-2964.
- Aznarez I, Barash Y, Shai O, He D, Zielenski J, Tsui LC, Parkinson J, Frey BJ, Rommens JM, Blencowe BJ. 2008. A systematic analysis of intronic sequences downstream of 5' splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation. *Genome Res* **18**: 1247-1258.
- Banerjee H, Rahn A, Gawande B, Guth S, Valcarcel J, Singh R. 2004. The conserved RNA recognition motif 3 of U2 snRNA auxiliary factor (U2AF 65) is essential in vivo but dispensable for activity in vitro. *Rna* **10**: 240-253.
- Bauer WJ, Heath J, Jenkins JL, Kielkopf CL. 2012. Three RNA recognition motifs participate in RNA recognition and structural organization by the pro-apoptotic factor TIA-1. *J Mol Biol* **415**: 727-740.
- Bell LR, Horabin JI, Schedl P, Cline TW. 1991. Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in Drosophila. *Cell* **65**: 229-239.
- Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 3171-3175.
- Berglund JA, Abovich N, Rosbash M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes & development* **12**: 858-867.
- Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**: 781-787.
- Bernado P, Garcia de la Torre J, Pons M. 2002. Interpretation of ¹⁵N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *Journal of biomolecular NMR* **23**: 139-150.

- Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* **72**: 291-336.
- Bonnal S, Martinez C, Forch P, Bachi A, Wilm M, Valcarcel J. 2008. RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Molecular cell* **32**: 81-95.
- Bonnal S, Vigevani L, Valcarcel J. 2012. The spliceosome as a target of novel antitumour drugs. *Nature reviews Drug discovery* **11**: 847-859.
- Chandler D. 2005. Interfaces and the driving force of hydrophobic assembly. *Nature* **437**: 640-647.
- Chaouki AS, Salz HK. 2006. Drosophila SPF45: a bifunctional protein with roles in both splicing and DNA repair. *PLoS genetics* **2**: e178.
- Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews Molecular cell biology* **10**: 741-754.
- Cheng Z, Menees TM. 2011. RNA splicing and debranching viewed through analysis of RNA lariats. *Molecular genetics and genomics : MGG* **286**: 395-410.
- Chow LT, Gelinis RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**: 1-8.
- Corsini L, Bonnal S, Basquin J, Hothorn M, Scheffzek K, Valcarcel J, Sattler M. 2007. U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nature structural & molecular biology* **14**: 620-629.
- Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G. 2008. Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein science : a publication of the Protein Society* **17**: 2070-2079.
- Corsini L, Hothorn M, Stier G, Rybin V, Scheffzek K, Gibson TJ, Sattler M. 2009. Dimerization and protein binding specificity of the U2AF homology motif of the splicing factor Puf60. *The Journal of biological chemistry* **284**: 630-639.
- Cruz-Gallardo I, Aroca A, Persson C, Karlsson BG, Diaz-Moreno I. 2013. RNA binding of T-cell intracellular antigen-1 (TIA-1) C-terminal RNA recognition motif is modified by pH conditions. *J Biol Chem* **288**: 25986-25994.
- Damgaard CK, Lykke-Andersen J. 2011. Translational coregulation of 5'TOP mRNAs by TIA-1 and TIAR. *Genes Dev* **25**: 2057-2068.

- Daragan VA, Mayo KH. 1997. Motional model analyses of protein and peptide dynamics using ^{13}C and ^{15}N NMR relaxation. *Progress in Nuclear Magnetic Resonance Spectroscopy* **31**: 63-105.
- Das R, Zhou Z, Reed R. 2000. Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Molecular cell* **5**: 779-787.
- David CJ, Manley JL. 2010. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & development* **24**: 2343-2364.
- Del Gatto-Konczak F, Bourgeois CF, Le Guiner C, Kister L, Gesnel MC, Stevenin J, Breathnach R. 2000. The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. *Molecular and cellular biology* **20**: 6287-6299.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **6**: 277-293.
- Dember LM, Kim ND, Liu KQ, Anderson P. 1996. Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. *J Biol Chem* **271**: 2783-2788.
- Doreleijers JF, Sousa da Silva AW, Krieger E, Nabuurs SB, Spronk CA, Stevens TJ, Vranken WF, Vriend G, Vuister GW. 2012a. CING: an integrated residue-based structure validation program suite. *Journal of biomolecular NMR* **54**: 267-283.
- Doreleijers JF, Vranken WF, Schulte C, Markley JL, Ulrich EL, Vriend G, Vuister GW. 2012b. NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. *Nucleic acids research* **40**: D519-524.
- Effenberger KA, Perriman RJ, Bray WM, Lokey RS, Ares M, Jr., Jurica MS. 2013. A high-throughput splicing assay identifies new classes of inhibitors of human and yeast spliceosomes. *Journal of biomolecular screening* **18**: 1110-1120.
- Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica Section D, Biological crystallography* **60**: 2126-2132.
- Fan L, Lagisetty C, Edwards CC, Webb TR, Potter PM. 2011. Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. *ACS chemical biology* **6**: 582-589.
- Fernandez C, Wider G. 2003. TROSY in NMR studies of the structure and function of large biological macromolecules. *Current opinion in structural biology* **13**: 570-580.

- Forch P, Puig O, Kedersha N, Martinez C, Granneman S, Seraphin B, Anderson P, Valcarcel J. 2000. The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Molecular cell* **6**: 1089-1098.
- Forch P, Puig O, Martinez C, Seraphin B, Valcarcel J. 2002. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *The EMBO journal* **21**: 6882-6892.
- Forch P, Valcarcel J. 2001. Molecular mechanisms of gene expression regulation by the apoptosis-promoting protein TIA-1. *Apoptosis* **6**: 463-468.
- Frenal K, Callebaut I, Wecker K, Prochnicka-Chalufour A, Dendouga N, Zinn-Justin S, Delepierre M, Tomavo S, Wolff N. 2006. Structural and functional characterization of the TgDRE multidomain protein, a DNA repair enzyme from *Toxoplasma gondii*. *Biochemistry* **45**: 4867-4874.
- Garcia de la Torre J, Huertas ML, Carrasco B. 2000. HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *Journal of magnetic resonance* **147**: 138-146.
- Gardner KH, Kay LE. 1998. The use of ²H, ¹³C, ¹⁵N multidimensional NMR to study the structure and dynamics of proteins. *Annual review of biophysics and biomolecular structure* **27**: 357-406.
- Gesnel MC, Theoleyre S, Del Gatto-Konczak F, Breathnach R. 2007. Cooperative binding of TIA-1 and U1 snRNP in K-SAM exon splicing activation. *Biochem Biophys Res Commun* **358**: 1065-1070.
- Gilli P, Ferretti V, Gilli G, Borea PA. 1994. Enthalpy-entropy compensation in drug-receptor binding. *The Journal of Physical Chemistry* **98**: 1515-1518.
- Gottschalk A, Tang J, Puig O, Salgado J, Neubauer G, Colot HV, Mann M, Seraphin B, Rosbash M, Luhrmann R et al. 1998. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *Rna* **4**: 374-393.
- Gozani O, Potashkin J, Reed R. 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Molecular and cellular biology* **18**: 4752-4760.
- Grzesiek S, Bax A. 1993. Amino acid type determination in the sequential assignment procedure of uniformly ¹³C/¹⁵N-enriched proteins. *Journal of biomolecular NMR* **3**: 185-204.

- Gunnewiek JM, van Aarssen Y, Wassenaar R, Legrain P, van Venrooij WJ, Nelissen RL. 1995. Homodimerization of the human U1 snRNP-specific protein C. *Nucleic acids research* **23**: 4864-4871.
- Guntert P. 2004. Automated NMR structure calculation with CYANA. *Methods in molecular biology* **278**: 353-378.
- Guth S, Martinez C, Gaur RK, Valcarcel J. 1999. Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Molecular and cellular biology* **19**: 8263-8271.
- Hasegawa M, Miura T, Kuzuya K, Inoue A, Won Ki S, Horinouchi S, Yoshida T, Kunoh T, Koseki K, Mino K et al. 2011. Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product. *ACS chemical biology* **6**: 229-233.
- Heinrichs V, Bach M, Winkelmann G, Luhrmann R. 1990. U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* **247**: 69-72.
- Ivanov P, Kedersha N, Anderson P. 2011. Stress puts TIA on TOP. *Genes Dev* **25**: 2119-2124.
- Izquierdo JM, Majos N, Bonnal S, Martinez C, Castelo R, Guigo R, Bilbao D, Valcarcel J. 2005. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Molecular cell* **19**: 475-484.
- Izquierdo JM, Valcarcel J. 2007a. Fas-activated serine/threonine kinase (FAST K) synergizes with TIA-1/TIAR proteins to regulate Fas alternative splicing. *The Journal of biological chemistry* **282**: 1539-1543.
- . 2007b. Two isoforms of the T-cell intracellular antigen 1 (TIA-1) splicing factor display distinct splicing regulation activities. Control of TIA-1 isoform ratio by TIA-1-related protein. *J Biol Chem* **282**: 19410-19417.
- Jacques DA, Guss JM, Svergun DI, Trewhella J. 2012. Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta crystallographica Section D, Biological crystallography* **68**: 620-626.
- Jamison SF, Crow A, Garcia-Blanco MA. 1992. The spliceosome assembly pathway in mammalian extracts. *Molecular and cellular biology* **12**: 4279-4287.
- Johnson BA, Blevins RA. 1994. NMR View: A computer program for the visualization and analysis of NMR data. *Journal of biomolecular NMR* **4**: 603-614.
- Jurica MS, Moore MJ. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* **12**: 5-14.

- Kabsch W. 2010. Xds. *Acta crystallographica Section D, Biological crystallography* **66**: 125-132.
- Kaida D, Motoyoshi H, Tashiro E, Nojima T, Hagiwara M, Ishigami K, Watanabe H, Kitahara T, Yoshida T, Nakajima H et al. 2007. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nature chemical biology* **3**: 576-583.
- Kaida D, Schneider-Poetsch T, Yoshida M. 2012. Splicing in oncogenesis and tumor suppression. *Cancer science* **103**: 1611-1616.
- Kawai T, Lal A, Yang X, Galban S, Mazan-Mameczarz K, Gorospe M. 2006. Translational control of cytochrome c by RNA-binding proteins TIA-1 and HuR. *Mol Cell Biol* **26**: 3295-3307.
- Kay LE, Torchia DA, Bax A. 1989. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**: 8972-8979.
- Keller RLJ. 2004. *The Computer Aided Resonance Assignment Tutorial*. CANTINA Verlag.
- Kielkopf CL, Lucke S, Green MR. 2004. U2AF homology motifs: protein recognition in the RRM world. *Genes & development* **18**: 1513-1526.
- Kielkopf CL, Rodionova NA, Green MR, Burley SK. 2001. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **106**: 595-605.
- Kondo Y, Oubridge C, van Roon AM, Nagai K. 2015. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *eLife* **4**.
- Kotake Y, Sagane K, Owa T, Mimori-Kiyosue Y, Shimizu H, Uesugi M, Ishihama Y, Iwata M, Mizui Y. 2007. Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nature chemical biology* **3**: 570-575.
- Krawczak M, Thomas NS, Hundrieser B, Mort M, Wittig M, Hampe J, Cooper DN. 2007. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Human mutation* **28**: 150-158.
- Kumar AO, Swenson MC, Benning MM, Kielkopf CL. 2008. Structure of the central RNA recognition motif of human TIA-1 at 1.95Å resolution. *Biochemical and biophysical research communications* **367**: 813-819.
- Kuwasako K, Takahashi M, Tochio N, Abe C, Tsuda K, Inoue M, Terada T, Shirouzu M, Kobayashi N, Kigawa T et al. 2008. Solution structure of the second RNA recognition

- motif (RRM) domain of murine T cell intracellular antigen-1 (TIA-1) and its RNA recognition mode. *Biochemistry* **47**: 6437-6450.
- Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcarcel J. 2002. Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* **109**: 285-296.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **26**: 283-291.
- Laskowski RA, Swindells MB. 2011. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of chemical information and modeling* **51**: 2778-2786.
- Legrain P, Seraphin B, Rosbash M. 1988. Early commitment of yeast pre-mRNA to the spliceosome pathway. *Molecular and cellular biology* **8**: 3755-3760.
- Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. 1980. Are snRNPs involved in splicing? *Nature* **283**: 220-224.
- Letai AG. 2008. Diagnosing and exploiting cancer's addiction to blocks in apoptosis. *Nature reviews Cancer* **8**: 121-132.
- Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. 2011. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 11093-11098.
- Linge JP, Habeck M, Rieping W, Nilges M. 2003a. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**: 315-316.
- Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M. 2003b. Refinement of protein structures in explicit solvent. *Proteins: Structure, Function, and Bioinformatics* **50**: 496-506.
- Liu Y, Conaway L, Rutherford Bethard J, Al-Ayoubi AM, Thompson Bradley A, Zheng H, Weed SA, Eblen ST. 2013. Phosphorylation of the alternative mRNA splicing factor 45 (SPF45) by Clk1 regulates its splice site utilization, cell migration and invasion. *Nucleic acids research* **41**: 4949-4962.
- Loerch S, Maucuer A, Manceau V, Green MR, Kielkopf CL. 2014. Cancer-relevant splicing factor CAPERalpha engages the essential splicing factor SF3b155 in a specific ternary complex. *The Journal of biological chemistry* **289**: 17325-17337.

- Lopez de Silanes I, Galban S, Martindale JL, Yang X, Mazan-Mamczarz K, Indig FE, Falco G, Zhan M, Gorospe M. 2005. Identification and functional outcome of mRNAs associated with RNA-binding protein TIA-1. *Molecular and cellular biology* **25**: 9520-9531.
- Manceau V, Kielkopf CL, Sobel A, Maucuer A. 2008. Different requirements of the kinase and UHM domains of KIS for its nuclear localization and binding to splicing factors. *Journal of molecular biology* **381**: 748-762.
- Massi F, Johnson E, Wang C, Rance M, Palmer AG, 3rd. 2004. NMR *R1* rho rotating-frame relaxation with weak radio frequency fields. *J Am Chem Soc* **126**: 2247-2256.
- Matlin AJ, Clark F, Smith CW. 2005. Understanding alternative splicing: towards a cellular code. *Nature reviews Molecular cell biology* **6**: 386-398.
- Maucuer A, Ozon S, Manceau V, Gavet O, Lawler S, Curmi P, Sobel A. 1997. KIS is a protein kinase with an RNA recognition motif. *The Journal of biological chemistry* **272**: 23151-23156.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *Journal of applied crystallography* **40**: 658-674.
- Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica Section D, Biological crystallography* **53**: 240-255.
- Muto Y, Pomeranz Krummel D, Oubridge C, Hernandez H, Robinson CV, Neuhaus D, Nagai K. 2004. The structure and biochemical properties of the human spliceosomal protein U1C. *Journal of molecular biology* **341**: 185-198.
- Nelissen RL, Heinrichs V, Habets WJ, Simons F, Luhrmann R, van Venrooij WJ. 1991. Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic acids research* **19**: 449-454.
- Neubauer G, King A, Rappsilber J, Calvio C, Watson M, Ajuh P, Sleeman J, Lamond A, Mann M. 1998. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature genetics* **20**: 46-50.
- Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**: 457-463.
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* **40**: 1413-1415.

- Perry WL, 3rd, Shepard RL, Sampath J, Yaden B, Chin WW, Iversen PW, Jin S, Lesoon A, O'Brien KA, Peek VL et al. 2005. Human splicing factor SPF45 (RBM17) confers broad multidrug resistance to anticancer drugs when overexpressed--a phenotype partially reversed by selective estrogen receptor modulators. *Cancer research* **65**: 6593-6600.
- Pervushin K, Riek R, Wider G, Wuthrich K. 1997. Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 12366-12371.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**: 1605-1612.
- Piecyk M, Wax S, Beck AR, Kedersha N, Gupta M, Maritim B, Chen S, Gueydan C, Krusys V, Streuli M et al. 2000. TIA-1 is a translational silencer that selectively regulates the expression of TNF-alpha. *Embo J* **19**: 4154-4163.
- Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K. 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**: 475-480.
- Puig O, Gottschalk A, Fabrizio P, Seraphin B. 1999. Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes & development* **13**: 569-580.
- Reyes R, Alcalde J, Izquierdo JM. 2009. Depletion of T-cell intracellular antigen proteins promotes cell proliferation. *Genome Biol* **10**: R87.
- Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR. 2012. Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes & development* **26**: 1098-1109.
- Roy B, Haupt LM, Griffiths LR. 2013. Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity. *Current genomics* **14**: 182-194.
- Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. 1998. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 13585-13590.

- Sammeth M, Foissac S, Guigo R. 2008. A general definition and nomenclature for alternative splicing events. *PLoS computational biology* **4**: e1000147.
- Sampath J, Long PR, Shepard RL, Xia X, Devanarayan V, Sandusky GE, Perry WL, 3rd, Dantzig AH, Williamson M, Rolfe M et al. 2003. Human SPF45, a splicing factor, has limited expression in normal tissues, is overexpressed in many tumors, and can confer a multidrug-resistant phenotype to cells. *The American journal of pathology* **163**: 1781-1790.
- Sattler M, Schleucher J, Griesinger C. 1999a. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**: 93-158.
- Sattler M, Schleucher J, Griesinger C. 1999b. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**: 93-158.
- Schrodinger, LLC. 2015. The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8.
- Schuttelkopf AW, van Aalten DM. 2004. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta crystallographica Section D, Biological crystallography* **60**: 1355-1363.
- Schwerk C, Schulze-Osthoff K. 2005. Regulation of apoptosis by alternative pre-mRNA splicing. *Molecular cell* **19**: 1-13.
- Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M. 2003. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Molecular cell* **11**: 965-976.
- Seraphin B, Rosbash M. 1989. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**: 349-358.
- Shan, X., Gardner, H. K, Muhandiram, R. D, RAO, S. N, Arrowsmith, H. C et al. 1996. ASSIGNMENT OF ¹⁵N, ¹³CALPHA, ¹³CBETA, AND HN RESONANCES IN AN ¹⁵N,¹³C, ²H LABELED 64 KDA TRP REPRESSOR : OPERATOR COMPLEX USING TRIPLE-RESONANCE NMR SPECTROSCOPY AND ²H-DECOUPLING. American Chemical Society, Washington, DC, ETATS-UNIS.
- Shen Y, Delaglio F, Cornilescu G, Bax A. 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR* **44**: 213-223.

- Sillekens PT, Beijer RP, Habets WJ, van Venrooij WJ. 1988. Human U1 snRNP-specific C protein: complete cDNA and protein sequence and identification of a multigene family in mammals. *Nucleic acids research* **16**: 8307-8321.
- Silverman EJ, Maeda A, Wei J, Smith P, Beggs JD, Lin RJ. 2004. Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing. *Molecular and cellular biology* **24**: 10101-10110.
- Singh NN, Seo J, Ottesen EW, Shishimorova M, Bhattacharya D, Singh RN. 2011. TIA1 prevents skipping of a critical exon associated with spinal muscular atrophy. *Mol Cell Biol* **31**: 935-954.
- Spadaccini R, Reidt U, Dybkov O, Will C, Frank R, Stier G, Corsini L, Wahl MC, Luhrmann R, Sattler M. 2006. Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. *Rna* **12**: 410-425.
- Spera S, Bax A. 1991. Empirical correlation between protein backbone conformation and C.alpha. and C.beta. ¹³C nuclear magnetic resonance chemical shifts. *Journal of the American Chemical Society* **113**: 5490-5492.
- Svec M, Bauerova H, Pichova I, Konvalinka J, Strisovsky K. 2004. Proteinases of betaretroviruses bind single-stranded nucleic acids through a novel interaction module, the G-patch. *FEBS letters* **576**: 271-276.
- Svergun D. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* **25**: 495-503.
- Thickman KR, Swenson MC, Kabogo JM, Gryczynski Z, Kielkopf CL. 2006. Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. *Journal of molecular biology* **356**: 664-683.
- Tian Q, Taupin J, Elledge S, Robertson M, Anderson P. 1995. Fas-activated serine/threonine kinase (FAST) phosphorylates TIA-1 during Fas-mediated apoptosis. *The Journal of experimental medicine* **182**: 865-874.
- Tjandra N, Kuboniwa H, Ren H, Bax A. 1995. Rotational dynamics of calcium-free calmodulin studied by ¹⁵N-NMR relaxation measurements. *Eur J Biochem* **230**: 1014-1024.
- Vagin A, Teplyakov A. 2010. Molecular replacement with MOLREP. *Acta crystallographica Section D, Biological crystallography* **66**: 22-25.

- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**: 687-696.
- Vriend G, Sander C. 1993. Quality control of protein models: directional atomic contact analysis. *Journal of applied crystallography* **26**: 47-60.
- Wallace AC, Laskowski RA, Thornton JM. 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering* **8**: 127-134.
- Wang GS, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature reviews Genetics* **8**: 749-761.
- Wang I, Hennig J, Jagtap PK, Sonntag M, Valcarcel J, Sattler M. 2014. Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1. *Nucleic acids research* **42**: 5949-5966.
- Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna* **14**: 802-813.
- Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, Rot G, Zupan B, Curk T, Ule J. 2010. iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol* **8**: e1000530.
- Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. 1999. Protein identification and analysis tools in the ExPASy server. *Methods in molecular biology* **112**: 531-552.
- Will CL, Luhrmann R. 2011. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* **3**.
- Will CL, Rumpler S, Klein Gunnewiek J, van Venrooij WJ, Luhrmann R. 1996. In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic acids research* **24**: 4614-4623.
- Williamson MP, Havel TF, Wuthrich K. 1985. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ¹H nuclear magnetic resonance and distance geometry. *Journal of molecular biology* **182**: 295-315.
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A et al. 2011. Overview of the CCP4 suite and current developments. *Acta crystallographica Section D, Biological crystallography* **67**: 235-242.

- Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1995. ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of biomolecular NMR* **5**: 67-81.
- Wishart DS, Sykes BD, Richards FM. 1992. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**: 1647-1651.
- Yamazaki T, Forman-Kay JD, Kay LE. 1993. Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./epsilon. chemical shifts of aromatic residues in ^{13}C -labeled proteins via scalar couplings. *Journal of the American Chemical Society* **115**: 11054-11055.
- Younis I, Berg M, Kaida D, Dittmar K, Wang C, Dreyfuss G. 2010. Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Molecular and cellular biology* **30**: 1718-1728.
- Zaharieva E, Chipman JK, Soller M. 2012. Alternative splicing interference by xenobiotics. *Toxicology* **296**: 1-12.
- Zamore PD, Patton JG, Green MR. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**: 609-614.
- Zuccato E, Buratti E, Stuani C, Baralle FE, Pagani F. 2004. An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing. *J Biol Chem* **279**: 16980-16988.

Curriculum Vitae

Pravin Kumar Ankush Jagtap

Personal Information

Sex Male
Nationality Indian
Address (work) Helmholtz Zentrum München
Institute for Structural Biology (STB)
Ingolstädter Landstr. 1 85764 Neuherberg
Date of Birth June 9, 1988
E-mail: pravin.jagtap@tum.de; pravinjaggu@gmail.com

EDUCATION

Technische Universität München, Munich, Germany

- PhD in Structural Biology, November 2011-present

Indian Institute of Technology Kanpur, Kanpur, India

- Masters of Technology (Biological Sciences and Bioengineering), 2009-2011, 9.67 (CGPA)

Tatyasaheb Kore Institute Of Engg. & Technology, Kolhapur, India

- Bachelor of Engineering (Biotechnology), 2005-2009, 79.43%

FELLOWSHIPS AND AWARDS

- PhD fellowship, Boehringer Ingelheim Fonds (Jul 2011-Oct 2014)
- Cadence Gold Medal for “Best M.Tech Thesis” work for the year 2011
- Academic Excellence Award for highest CGPA in M.Tech for the year 2009-2010
- GATE fellowship (All India 2nd Rank, 99.99 percentile), Ministry of Human Resource Development, India (2009-2011)
- Best Outgoing Student Award for the year 2009, Tatyasaheb Kore institute of engineering and technology, Warnanagar, Kolhapur, India,
- Merit Scholarship for securing highest grades for three consecutive years of engineering, Shivaji University, India (2007-2009)

PUBLICATIONS

- Guaitoli G, Raimondi F, Gilsbach B, Gómez-Llorente Y, Deyaert E, Renzi F, Li X, Schaffner A, **Jagtap PKA**, Boldt K, Zweydorf F, Gotthardt K, Lorimer D, Yue Z, Burgin A, Janjic N, Sattler M, Versées W, Ueffing M, Beldandia I, Kortholt A, Gloeckner CJ. A structural model of the dimeric Parkinson's Protein LRRK2 reveals a compact architecture involving distant inter-domain contacts. (Accepted in PNAS)
- Piccoli G, Onofri F, Cirnar M, Kaiser C, **Jagtap PKA**, Kastenmüller A, Pischedda F, Marte A, Zweydorf F, Vogt A, Giesert F, Pan L, Antonucci F, Kiel C, Zhang M, Weinkauff S, Sattler M, Sala C, Matteoli, Ueffing M. LRRK2 binds to neuronal vesicles through protein interactions mediated by its C-terminal WD40 domain. *Mol.Cell.Biol.* 2014 Jun;34(12):2147-61
- Wang I, Hennig J, **Jagtap PKA**, Sonntag M, Valcárcel J, Sattler M. Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1. *Nucleic Acids Res.* 2014 May;42(9):5949-66

- **Jagtap PKA***, Verma SK*, Vithani N*, Bais VS*, Prakash B. Crystal structures identify an atypical two-metal ion mechanism for uridyl transfer in GlmU: Its significance to sugar nucleotidyl transferases. *J Mol Biol.* 2013 May 27;425(10):1745-59
- **Jagtap PKA***, Soni V*, Vithani N, Jhingan GD, Bais VS, Nandicoori VK, Prakash B. Substrate-bound crystal structures reveal features unique to Mycobacterium tuberculosis N-acetyl-glucosamine 1-phosphate uridyltransferase and a catalytic mechanism for acetyl transfer. *J Biol Chem.* 2012 Nov 16;287(47):39524-37

MANUSCRIPTS UNDER PREPARATION

- **Jagtap PKA***, Garg D*, Kapp TG, Will C, Demmer O, Lührmann R, Kessler H, Sattler M. Rational design of cyclic peptide inhibitors of UHM domains to modulate pre-mRNA splicing.
- **Jagtap PKA**, Asami S, Sippel C, Hausch F, Sattler M. Implications of conformational dynamics in designing FKBP51 selective inhibitors.
- Sonntag M*, **Jagtap PKA***, Simon Bernd, Appavou MS, Gabel F, Hennig J*, Sattler M. Small angle neutron scattering of segmentally deuterated multi-domain protein TIA-1 in complex with RNA.
- **Jagtap PKA***, Kubelka T*, Garg D, Schropp K, Will C, Sippel C, Potukuchi H, Hausch F, Hadian K, Lührmann R, Bach T, Sattler M. Rational design of small molecule inhibitors of UHM domains to modulate pre-mRNA splicing.

RESEARCH EXPERIENCE

Technische Universität München, Munich, Germany

Graduate student with Prof. Dr. Michael Sattler, 2011-present

Molecular recognition of splicing factors involved in Fas alternative splicing.

- Characterization of TIA-1-U1C interaction using an integrated approach of NMR, X-ray crystallography and SAXS
- Development of cyclic peptide and small molecule inhibitors of SPF45 UHM domain involved in Fas alternative splicing.

Indian Institute of Technology Kanpur, India

Master's thesis Jan 2010-Jun 2011 and Research Assistant Jul 2011-Sep 2011 with Dr. Balaji Prakash
Structural and biochemical studies on the Acetyltransferase activity of GlmU from *M. tuberculosis*

- Crystallized GlmU protein from *M. tuberculosis* along with its Acetyltransferase substrates
- Mutational analysis of the acetyltransferase active site based on the obtained structural data

Tatyasaheb Kore Institute Of Engg. & Technology, Kolhapur, India

Bachelor's thesis with Dr Rajshree Kamble, 2009

Comparative studies on the removal of lead and zinc by *B. subtilis* and *P. desmolyticum*

- Immobilized bacterial species in cross-linked porous agarose gels
- Utilized these gels for the removal of heavy metal contaminants from industrial waste

CONFERENCES AND WORKSHOPS

- 3rd workshop of Cross Program Activity (CPA) Structural Biology, 5th-6th November 2015, Berlin
Oral presentation: Structure based design of small molecules and cyclic peptides targeting alternative splicing factors
- G-NMR school, 13th-15th Oct 2014, Helmholtz Zentrum München
- 2nd HFG Structural Biology Workshop, 28th-29th November 2013, Helmholtz Zentrum München
- EUROMAR 2013, 30th Jun. -5th Jul. 2013, Hersonissos, Crete, Greece
Poster Presentation: Molecular recognition of alternative splicing factors that modulate *Fas* alternative splicing
- Electron Microscopy Workshop, 18th -22nd Mar. 2013, Gene center, München
- Small angle X-ray scattering Workshop, 14th -15th Mar. 2013, Gene center, München
- International Conference on Tissue Engg. and Stem Cell research using Nano materials, held in February 2009 at Amrita Center for Nano sciences, Kerala
- State level workshop on Intellectual Property Rights held in Feb 2009 at TKCP, Kolhapur
- Presented a poster and paper in National Level Symposium “Synapse-2007” at KLE Belgaum

