

Development and application of methods for structural analysis of the splicing factor TIA-1

(Entwicklung und Anwendung von Methoden für die Strukturuntersuchung des Spleißfaktors TIA-1)

Dissertation

zum Erlangen des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von

Miriam Elisabeth Sonntag

geboren am 10.02.1987 in München



Technische Universität München

Fakultät für Chemie

Bayerisches NMR-Zentrum

Lehrstuhl für biomolekulare NMR Spektroskopie

Garching bei München, 2017



TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie
Bayerisches NMR-Zentrum
Lehrstuhl für biomolekulare NMR Spektroskopie

Development and application of methods for structural analysis of the splicing factor TIA-1

MIRIAM ELISABETH SONNTAG

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. Franz Hagn

Prüfer der Dissertation:

1. Prof. Dr. Michael Sattler
2. Prof. Dr. Dierk Niessing

Die Dissertation wurde am 22.08.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 17.10.2017 angenommen.

Contents

Abstract/Zusammenfassung	1
1. Introduction	5
2. Methods	11
2.1. Recombinant protein expression and purification	11
2.2. Segmental isotope labeling	11
2.3. Small angle scattering experiments	15
2.3.1. Small angle X-ray scattering	17
2.3.2. Small angle neutron scattering	18
2.4. Biophysical characterization of TIA-1	18
2.4.1. Static light scattering	19
2.4.2. Isothermal titration calorimetry	20
2.5. NMR experiments	21
2.5.1. Resonance assignment	22
2.5.2. Relaxation measurements	24
2.6. Crystallization	26
3. Publications	27
3.1. Efficient segmental isotope labeling of multi-domain proteins using Sortase A	27
3.2. Segmental, domain-selective perdeuteration and small angle neutron scattering for structural analysis of multi-domain proteins	28
4. Conclusions	31
A. Appendix	35
Abbreviations	39
Bibliography	41

Abstract/Zusammenfassung

Abstract

Posttranscriptional processing of the pre-mRNA provides essential regulation of gene expression and is important for biological function and human diseases. Alternative splicing greatly expands protein diversity and complexity in higher eukaryotes. Comprising multiple non-coding intronic sequences and gene-coding exonic sequences, the eukaryotic pre-mRNA needs to be processed prior to translation. In a catalytic process, termed splicing, the intronic sequences are excised from the pre-mRNA. Alternative patterns of intron removal generate various mRNA-encoding gene products that function in diverse cellular processes including cell growth, differentiation and death. A complex network of protein-RNA interactions modulates and controls intron excision, which is essential for human and metazoan development. Disruption or misregulation of this tightly regulated interplay between positive and negative *cis*-regulatory RNA elements that are recognized by *trans*-acting RNA binding proteins underlies various diseases including cancer and autoimmune disorders.

One key regulatory protein involved in alternative splicing regulation of the cancer-related *fas* pre-mRNA is the human multi-domain protein T-cell restricted intracellular antigen-1 (TIA-1). TIA-1 stimulates U1 snRNP recruitments to the 5' splice site by binding to uridine-rich (U-rich) intronic splicing enhancer sequences immediately downstream of alternative exons.

TIA-1 harbours three canonical RNA recognition motifs (RRMs) and a C-terminal glutamine-rich domain all connected by flexible linkers. Although it has been shown that RRM2 and RRM3 associated with U-rich intronic pre-mRNA sequences the role of RRM1 remains elusive. Moreover, the underlying molecular mechanisms and dynamics of splice site recognition and activation are poorly understood.

To elucidate the complex network of protein-RNA and protein-protein interactions involved in alternative splicing regulation of the Fas receptor gene an integrated structural biology approach was employed. Small-angle X-ray scattering (SAXS) and contrast-matching small angle neutron scattering (SANS) of domain-selective perdeuterated TIA-1 RRM123 were combined with computational modeling, nuclear magnetic resonance spectroscopy and X-ray crystallography. Domain-selective perdeuterated TIA-1 samples for SANS measurements were generated using LPXTG-specific Sortase A-mediated protein ligation. Existing protocols were optimized in two steps to ensure high yields of pure, highly domain-selective labeled protein samples in the milligram range. Contrast-matching SANS experiments based on segmental perdeuteration integrated as a filtering step into rigid body modeling provided unique insights into the domain-architecture of the RRM123/RNA complex. Upon RNA binding, TIA-1 adopts an elongated L-shape in which RRM1 is extended from the closely packed RRM23 domains. Consistent with a role in protein-protein rather than protein-RNA interaction, RRM1 is detached from RRM23 and does not associate with the RNA.

Taken together, the methodology developed and employed in this thesis provides an efficient tool to dissect the structure and dynamics of multi-domain proteins. For TIA-1 unique insights into how multiple domains cooperate in alternative splice site recognition was gained. Upon binding to U-rich intronic splicing enhancers all three domains rearrange in such a way that the recruitment of U1 snRNP to the 5' splice site is facilitated.

Zusammenfassung (Abstract in German)

Posttranslationale prä-mRNA Prozessierung leistet einen wichtigen Beitrag zur Regulation der Genexpression und ist entscheidend für die biologische Funktion und Entstehung von Krankheiten. In höheren Eukaryoten erweitert alternatives Spleißen die Proteinviefalt und –Komplexität. Bestehend aus nicht-kodierenden, intronischen und gen-kodierenden, exonischen Sequenzen muss die prä-mRNA erst prozessiert werden, bevor sie translatiert werden kann. In einem katalytischen Prozess, dem sogenannten Spleißen, werden die intronischen Sequenzen aus der prä-mRNA herausgeschnitten. Alternative Muster der Intron-Entfernung generieren verschiedenste mRNA-kodierende Genprodukte, die an den unterschiedlichsten zellulären Prozessen einschließlich dem Zellwachstum, der Differenzierung und dem Tod, beteiligt sind. Ein komplexes Netzwerk aus Protein-RNA Interaktionen kontrolliert und moduliert das für die menschliche und metazoische Entwicklung entscheidende Herausschneiden der intronischen Sequenzen. Wird dieses eng regulierte Zusammenspiel zwischen positiven und negativen *cis*-regulatorischen RNA Sequenzen, die durch *trans*-agierende RNA-bindende Proteine erkannt werden, gestört, können verschiedene Krankheiten einschließlich Krebs und Autoimmunerkrankungen entstehen.

Ein wichtiges regulatorisches Protein, das an der Regulierung des alternativen Spleißens der krebs-relevanten *fas* prä-mRNA beteiligt ist, ist das Multidomänen-Protein T-cell restricted intracellular antigen-1 (TIA-1). TIA-1 erkennt und bindet U-reiche intronische Spleißverstärker-Sequenzen direkt nach alternativen Exons und erleichtert somit die Rekrutierung von U1 snRNP zur 5' Spleißstelle.

TIA-1 besitzt drei, über flexible Verbindungen verknüpfte, kanonische RNA-Bindemotive (RRMs) und eine C-terminale Glutamin-reiche Domäne. Bisher sind nur die Funktion der zentralen RRM2 und RRM3 Domäne bekannt. Beide Domänen erkennen und binden kooperativ U-reiche intronische prä-mRNA Sequenzen, jedoch sind die zugrundeliegenden molekularen Mechanismen und Dynamiken der alternativen Spleißstellen-Erkennung und –Aktivierung noch weitestgehend unbekannt.

Um Einblicke in das komplexe Netzwerk der an der Regulierung des alternativen Spleißens des Fas Rezeptor-Gens beteiligten Protein-Protein und Protein-RNA Interaktion, zu gewinnen, wurde ein integrierter strukturbioologischer Ansatz verfolgt. Kleinwinkelröntgenstreuung (SAXS) und kontrastangepasste -Neutronenstreuung (SANS) von domänenspezifisch perdeutertem TIA-1 RRM123 wurden dabei mit computerbasierter Modellierung, Kernspinresonanz-Spektroskopie und Röntgenkristallographie kombiniert. Selektive Perdeuterierung wurde mittels LPXTG-spezifischer Sortase A-vermittelter Proteinligation erreicht. Die Einführung von zwei Optimierungsschritten in bestehende Protokolle garantierte eine hohe Ausbeute im Milligramm-Maßstab von reinen, im hohen Maß domänenspezifisch markierten Proteinproben. Die Integration von kontrastangepassten, auf segmentweiser Perdeuterierung basierenden SANS Messungen als Filter bei der Starrkörper Modellierung bot einmalige Einblicke in die Domänenarchitektur des RRM123/RNA-Komplexes. In der RNA gebundenen Form nimmt TIA-1 eine langgestreckte, L-förmige Konformation ein, in der RRM1 räumlich getrennt von den dicht gepackten RRM2- und RRM3-Domänen und der RNA vorliegt. Diese Beobachtungen zeigen, dass RRM1 eher in die Protein-Protein-Interaktion anstatt der Protein-RNA-Interaktion eingebunden ist.

Abschließend kann festgehalten werden, dass der in dieser Arbeit entwickelte und angewendete Ansatz ein effizientes Hilfsmittel liefert, um die Rolle und Dynamiken von Multidomänen-Proteinen eingehend zu analysieren. In Bezug auf TIA-1 konnten einmalige Einblicke in das Zusammenspiel von mehreren Domänen bei der alternativen Spleißstellen-Erkennung gewonnen werden. Kommt es zur Bindung von U-reichen intronischen Spleißverstärker-Sequenzen, richten sich alle drei Domänen so aus, dass die 5' Spleißstellen-Erkennung durch U1 snRNP erleichtert wird.

1. Introduction

Proteins are synthesised from messenger RNA (mRNA). In eukaryotes, gene-coding DNA is transcribed into pre-mRNA, the non-translatable precursor of the mRNA comprising gene-coding exonic and non-coding intronic sequences (Figure 1.1 upper panel). To generate mature mRNA, which can be further translated to proteins, the intronic sequences need to be excised from the pre-mRNA (Figure 1.1 lower panel). Alternative patterns of intron excision generates diverse mRNA encoding gene products with distinct, sometimes opposing functions, thus expanding protein complexity and diversity in higher eukaryotes.¹⁻³

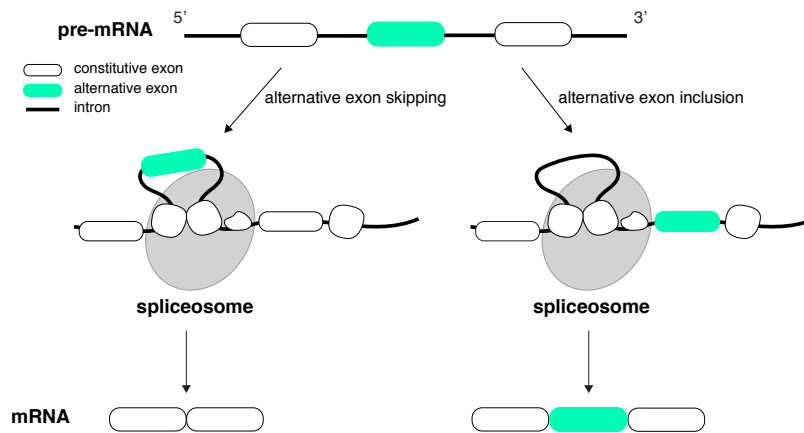


Figure 1.1.: Schematic representation of pre-mRNA processing. To generate functional mRNA, intronic sequences need to be removed from the pre-mRNA. Whereas constitutive exons (white) are always included, alternative exons (green) can be either excised as part of the intronic sequence or included in the functional mRNA. This process, known as alternative splicing, generates various mRNAs with distinct functions.

Intron removal and thus exon splicing is carried out by the spliceosome, a massive complex composed of five small nuclear ribonucleoprotein particles (snRNPs) and a large number of proteins^{4,5} that cooperate to define the gene-coding and non-coding stretches of the pre-mRNA. Spliceosomal assembly, a stepwise procedure, is initiated by the recognition of RNA sequences at the intron-exon junction termed splice sites (ss) (Figure 1.2 step 1). Binding of U1 and U2 snRNP to the 5' and 3' ss, respectively, is followed by the incorporation of U4/U6-U5 tri-snRNP (Figure 1.2 step 2). After extensive rearrangements and remodelling of the whole complex, the catalytically active spliceosome is formed.⁵

Splice site recognition defines which exon is removed or included in the mature mRNA and involves exon-centered interactions between U1 and U2 snRNP (Figure 1.2 exon definition). The choice of a splice site is further regulated by the combinatorial and competitive interplay between positive and negative RNA sequences within introns and exons. Divided into four categories, exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE) and intronic splicing silencer (ISS), these RNA sequences, termed *cis*-regulatory elements, are recognized by non-snRNP protein components including members of the heterogeneous nuclear RNPs (hnRNPs) and SR (Ser-Arg) protein family (Figure 1.3).^{4,6-8} While SR proteins bind to ESE elements and act as splicing activators by facilitating splice site recognition, hnRNPs function as splicing inhibitors by binding to ESS and ISS elements which sterically blocks

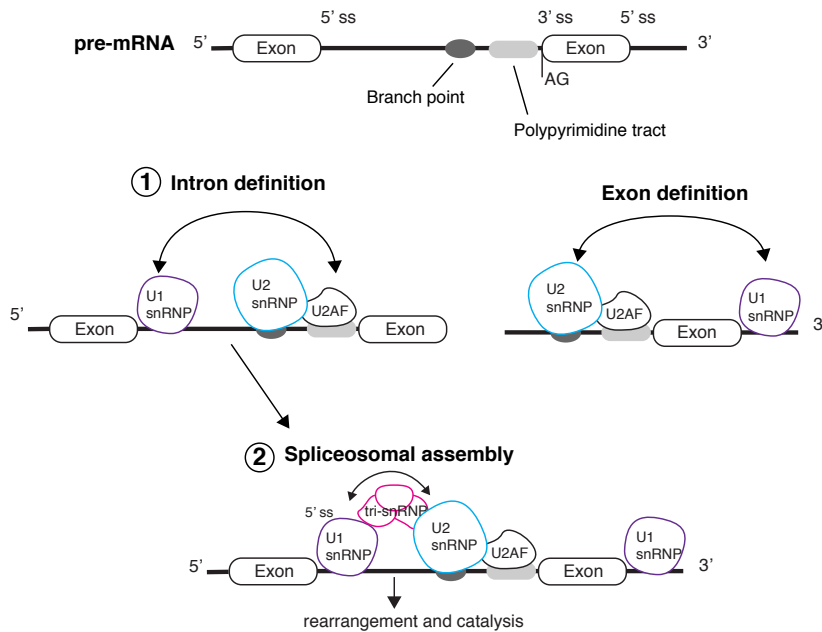


Figure 1.2.: Schematic representation of spliceosomal assembly. In a first step (1), U1 snRNP and U2AF bind to the 5' and 3' ss respectively. Next, U2 snRNP is recruited to the branch point. Bridging events between U1 and U2 snRNP occur either across the intron or exon, defining these regions. After defining the intron (1) the last component of the spliceosomal complex (U4/U6–U5 tri-snRNP) (2) is recruited to the ss, which leads to extensive rearrangement of all sub-units and thus the formation of the catalytically active spliceosome.

the access of either snRNPs to the regulated exon or splicing activators to enhancer sites (Figure 1.3).⁶ Regulatory proteins that stimulate splicing by binding to ISE elements, the least characterized *cis*-regulatory element, include the T-cell intracellular antigen-1 (TIA-1).⁹

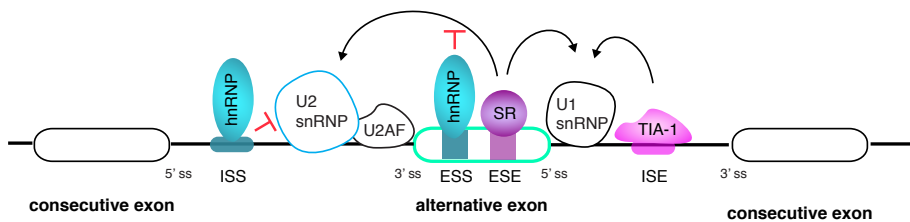


Figure 1.3.: Schematic model of splice site selection. Intron removal is mediated by the complex interplay of specific RNA sequences within introns and exons. Binding of non-snRNP proteins to exonic splicing silencers and enhancers (ESS and ESE) as well as intronic splicing silencers and enhancers (ISS and ISE) controls early steps in spliceosomal assembly. Activating regulatory proteins such as SR proteins bind to ESE elements and stimulate U2 and U1 snRNP association with the 3' and 5' ss, respectively. Contrary, members of the hnRNP protein family inhibit splicing by associating with splicing silencers within introns and exons (ISS and ESS).

TIA-1 is an RNA-binding protein that shuttles between the cytoplasm and the nucleus to control diverse aspects of mRNA metabolism.¹⁰ In the cytoplasm TIA-1 functions as translational silencer by binding to A/C or C/U-rich regions in the 3' untranslated regions (3' UTR) of target mRNAs.^{11–15} Under environmental stress TIA-1 isolates mRNA targets into stress granules and suspends their translation until the stress is resolved.^{16–18} In the nucleus

TIA-1 modulates alternative splicing of various pre-mRNAs by binding to uridine-rich (U-rich) RNA sequences downstream of weak 5' splice sites.^{19–23} Its most prominent role has been described as regulating splicing of cancer-related *fas* mRNA.^{9,24,25}

Balanced fine-tuning between cell growth and death is essential for the proper development and function of the immune system. Derailment of the homeostatic control of immune cells is linked to cancer, autoimmunity or acquired immunodeficiency syndrome (AIDS).²⁶ One way to regulate the number of cells is programmed cell death or apoptosis.²⁶ Auto-reactive, inoperable or no longer used immune cells are typically eliminated by one of the two distinct pathways of apoptosis.²⁷ The extrinsic pathway involves the activation of death receptors by extracellular signals, whereas the intrinsic one is activated by noxious stimuli (e.g. ultraviolet radiation, chemotherapeutic drugs, starvation, or lack of the growth factors required for survival).^{27,28} Both pathways result in the activation of caspases, a family of catalytic enzymes necessary for the activation of chromosomal DNA fragmentation by endonucleases.²⁸

Identified as an effector of programmed cell death,^{9,24} TIA-1 modulates alternative splicing of the widely cell-surface expressed²⁹ member of the tumor necrosis factor receptor (TNF-R) superfamily, Fas death receptor^{26,30} gene (Figure 1.4 a). Alternative splicing of the Fas pre-mRNA generates two receptor isoforms with antagonistic effects in apoptosis. Functional *fas* mRNA lacking the transmembrane domain-encoding exon 6 generates the soluble-anti apoptotic form of the receptor whereas exon 6 inclusion results in the membrane-bound pro-apoptotic form of the receptor (Figure 1.4 b).^{9,31}

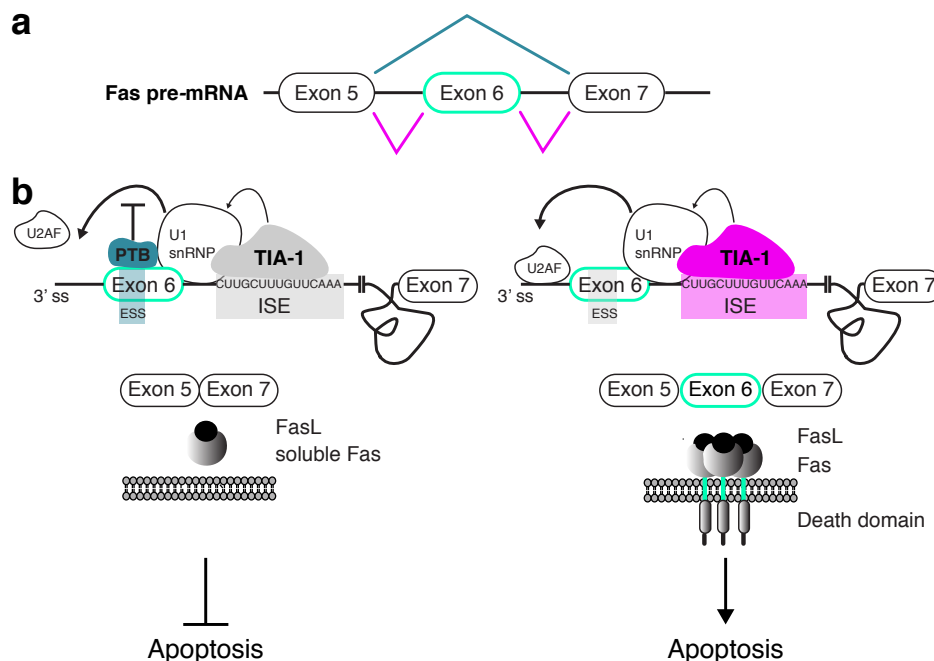


Figure 1.4.: Alternative splicing of the Fas death receptor gene. a) Fas pre-mRNA can be spliced into two antagonistic isoforms of the receptor. The transmembrane domain-encoding exon 6 (green) can either be skipped (blue lines) or included (magenta lines) in the functional mRNA. b) Alternative splicing of Fas pre-mRNA is modulated by regulatory factors like PTB or TIA-1. Left panel: PTB binding to an ESS motif inhibits exon 6 inclusion. Right panel: TIA-1 binding to an ISE motif promotes exon 6 inclusion. Binding of the physiological ligand, FasL, activates the extrinsic death signaling pathway of the receptor bearing cell.^{26,30} Essential for the transduction of the death signal is the intracellular death domain of the membrane-bound Fas death receptor.

TIA-1 binds to an ISE site at intron 6 and recruits U1 snRNP to initiate splicing (Figure 1.4 b).^{19,21,24,32–35} As part of a positive feed-back loop, TIA-1 directly interacts with the Fas-activated serine/threonine kinase (FASTK) to regulate Fas alternative splicing.^{36,37} Binding of the physiological ligand FasL rapidly dephosphorylates and activates FASTK. The active kinase phosphorylates TIA-1 which enhances U1 snRNP recruitment to the 5' ss of exon 6.³⁶ The subsequent enhanced exon 6 inclusion can lead to higher levels of membrane-bound Fas receptor expression and thus amplified Fas response.

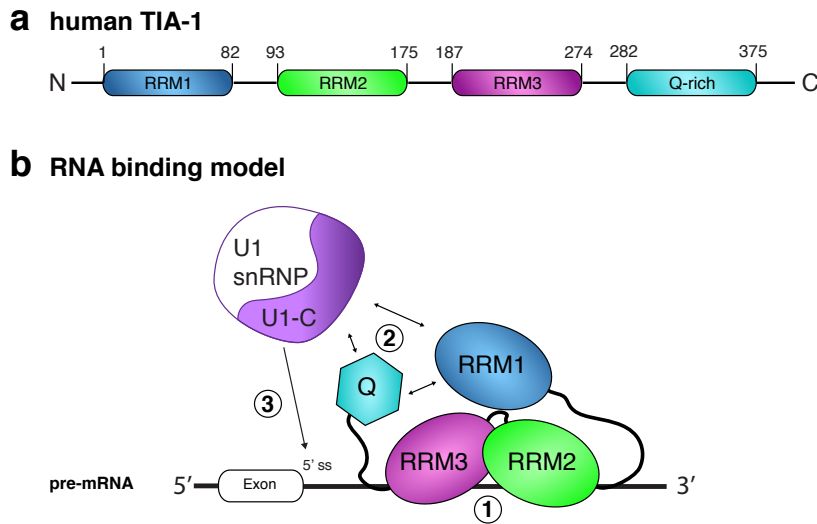


Figure 1.5.: Potential mechanism of RNA interaction by TIA-1. a) Schematic representation of human TIA-1 domains. b) RNA binding model. First step: The central RRM2 and RRM3 domain cooperatively bind to U-rich intronic stretches downstream of weak 5' ss. Second step: RRM1 and the Q-rich domain cooperate in U1 snRNP recruitment. U1 snRNP binding to the 5' ss initiates splicing (third step).

TIA-1 features three consecutive RRM domains followed by a C-terminal unstructured glutamine-rich (Q-rich) domain (Figure 1.5 a). All domains are connected by flexible linkers and possess distinct RNA or protein binding properties.^{21,34,38–40} High affinity binding to U-rich RNA stretches downstream of weak 5' ss is predominantly mediated by the central RNA recognition motif (RRM2).^{21,34} Although RRM3 does not cross-link with U-rich RNA, it increases the RNA binding affinity of RRM2.^{34,41} In contrast, RRM1 binds only weakly to single stranded RNA without any sequence specificity probably due to the presence of negatively charged residues in its ribonucleoprotein consensus motif (RNP-1).^{34,38} The current model suggests that RRM2 and RRM3 bind to pre-mRNA, whereas RRM1 and the Q-rich domain interact with U1C protein - a component of the U1 snRNP complex (Figure 1.5 b).²¹

To date, efforts to obtain high resolution structures of a tandem RRM23 or triple RRM123 construct of TIA-1 bound to RNA have been unsuccessful. A recently published crystal structure of RRM2 bound to DNA (PDB ID: 5ITH) showed unique structural insights into the molecular mechanisms of target recognition and interaction.⁴² As for most canonical RRM domains, stacking interactions on conserved aromatic rings (Phe98 and Phe140), hydrophobic interactions with sugar rings or phosphate as well as base-specific hydrogen bonds (His96, Trp170, Asn169, Asn101, and Asp100) are observed.⁴² Several crystallization trials were set up for the tandem RRM23 bound to thymine-rich (T-rich) 10 mer DNA oligonucleotides.⁴² In

each diffracting crystal, however, only RRM2 bound to three Ts were present whereas RRM3 and the majority of DNA nucleobases were absent.⁴²

Typically, RNA binding proteins (RBPs) regulate ligand interaction by either dynamic population shifts of the protein ensemble or cooperative assembly of multiple domains.^{43–46} The structure and underlying conformational dynamics of such multi-domain proteins are characterized by the combination of complementary solution techniques including nuclear magnetic resonance (NMR) spectroscopy and small angle scattering of X-rays (SAXS) and neutrons (SANS). NMR spectroscopy determines binding interfaces, domain conformations and dynamics whereas SAXS and SANS provide information on the overall shape. Up to now, combining NMR with SAXS data revealed that upon RNA interaction, all RRM domains of either the tandem RRM23 or triple RRM123 tumble together and adopt a more compacted domain arrangement in which the linkers keep their flexibility.^{38,40,42} However, no further information about the domain orientation relative to each other is available. In order to fully understand the role of TIA-1 in alternative splicing and thus regulating Fas signal transduction, the three-dimensional structure of TIA-1 bound to a *fas* derived mRNA is of high interest.

This thesis aimed to shed light on TIA-1-mediated alternative splicing of the Fas receptor gene. NMR spectroscopy, SAXS, SANS and X-ray crystallography were integrated to gain insight into the molecular mechanisms and conformational dynamics of TIA-1 upon RNA interaction. Different TIA-1 constructs summarized in Figure 2.1 were structurally and functionally characterized in the presence of either U-rich *fas* intron 6 derived or polyU oligonucleotides. Exchange processes such as multiple-register binding of the tested RNA oligonucleotides, however, impeded crystallization and led to severe line-broadening and disappearance of NMR signals. To overcome challenges associated with fast exchanging systems and highlight the specific role of RRM1 upon RNA binding, contrast-matching⁴⁷ SANS experiments on domain-selective perdeuterated TIA-1 RRM123 constructs (Figure 2.8) were combined with SAXS and computational modeling (see section 3.2 and ref. 48 for details). Selective perdeuteration of individual domains within the triple domain protein was achieved by sortase-mediated protein ligation,⁴⁹ a recently introduced enzyme-based segmental labeling technique. Two critical improvements compared to existing protocols (described in detail in section 3.1 and ref. 50) enabled the production of domain-selective perdeuterated TIA-1 RRM123 in the milligram range required for SANS measurements. To generate a three-dimensional model of Fas alternative splicing regulation, available NMR and X-ray structures of the individual domains were used for rigid body modeling of the multi-domain protein after confirming that the tertiary structures of each domain is conserved in the full-length protein. Integrating contrast-matching SANS data of segmental perdeuterated TIA-1 RRM123 with SAXS data as filtering steps into the restrained molecular dynamics protocol provided unique structural insights into the domain architecture of the protein-RNA complex. Due to the high discriminative power of the SANS data the relative domain positions could be defined and the question whether RRM1 contributes to RNA binding or not could be addressed.

In conclusion, the approach employed in this thesis not only revealed valuable information on how TIA-1 interacts with RNA but contributes to elucidate the molecular mechanisms of alternative splicing regulation.

2. Methods

This chapter reviews the conducted experiments to gain insight into the molecular mechanisms of how RNA interacts with TIA-1. Each section provides basic concepts of the applied method followed by a brief description of the experimental setup.

2.1. Recombinant protein expression and purification

All prepared TIA-1 constructs are summarized in Figure 2.1. The amplified TIA-1 DNA was sub-cloned into pET trxA_T_1a (European Molecular Biology Laboratory, Heidelberg, Germany), a modified pET 24d(+) vector. TIA-1 proteins contained a N-terminal His-tag fused to thioredoxin and a Tobacco Etch virus (TEV) protease cleavage site. Both RRM1 constructs, designed for sortase-mediated protein ligation (Figure 2.1 b), comprise an extra non-cleavable His-tag at its C-terminus. For segmental isotope labeling via sortase-mediated protein ligation, described in section 2.2, two different LPXTG specific sortase A (srtA) constructs (Figure 2.2) were tested. The amplified srtA DNA was sub-cloned into pET GB (provided by Peijian Zou), a modified pET 24d(+) vector.

Transformed into *E. coli* strain BL21 (DE3), the recombinant proteins were expressed in either LB media or M9 minimal media at 20 °C after induction at an optical density at 600 nm (OD₆₀₀) of 0.6–0.8 with 0.5 mM IPTG. All srtA as well as TIA-1 samples for crystallization, SAXS, isothermal titration calorimetry (ITC) or static light scattering (SLS) experiments were produced by growing cells in LB media. For specific isotope labeling, required for SANS and NMR, TIA-1 containing cells were grown in M9 media supplemented with different M9 salts. The M9 media was supplemented with ¹⁵NH₄Cl and/or ¹³C glucose to uniformly ¹⁵N and/or ¹³C label proteins for NMR experiments. Perdeuterated proteins, necessary for contrast-matching SANS experiments, were expressed in 99.8 % D₂O using ²H glucose as sole carbon source. To ensure near-complete perdeuteration of the expressed proteins, water-free M9 salts, vitamins, trace elements, CaCl₂ and MgSO₄ were dissolved in 99.8 % D₂O. The expressed proteins were purified using Ni-NTA affinity chromatography, TEV cleavage followed by a second Ni-NTA step, ion exchange, and a final size exclusion chromatography, described in detail in ref. 50. A 1 M NaCl washing step in the initial Ni-NTA affinity chromatography in combination with the two mentioned chromatographic methods ensured the removal of nucleic acid and RNase contamination, crucial for studying specific protein-RNA interactions. The absence of nucleic acid was further confirmed by measuring the absorbance ratio of 260 nm to 280 nm, which should be between 0.5 and 0.6 for solely protein containing samples. The purity of the final protein samples was assessed by Coomassie stained sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

2.2. Segmental isotope labeling

Segmental isotope labeling is a technique to label only defined regions within the intact protein, like individual domains of a multi-domain protein (Figure 2.3 a). NMR studies of molecules larger than 35 kDa benefit from segmental isotope labeling as it drastically reduces the number of peaks in the spectra and thus signal overlap.⁵⁴ Several different techniques like native

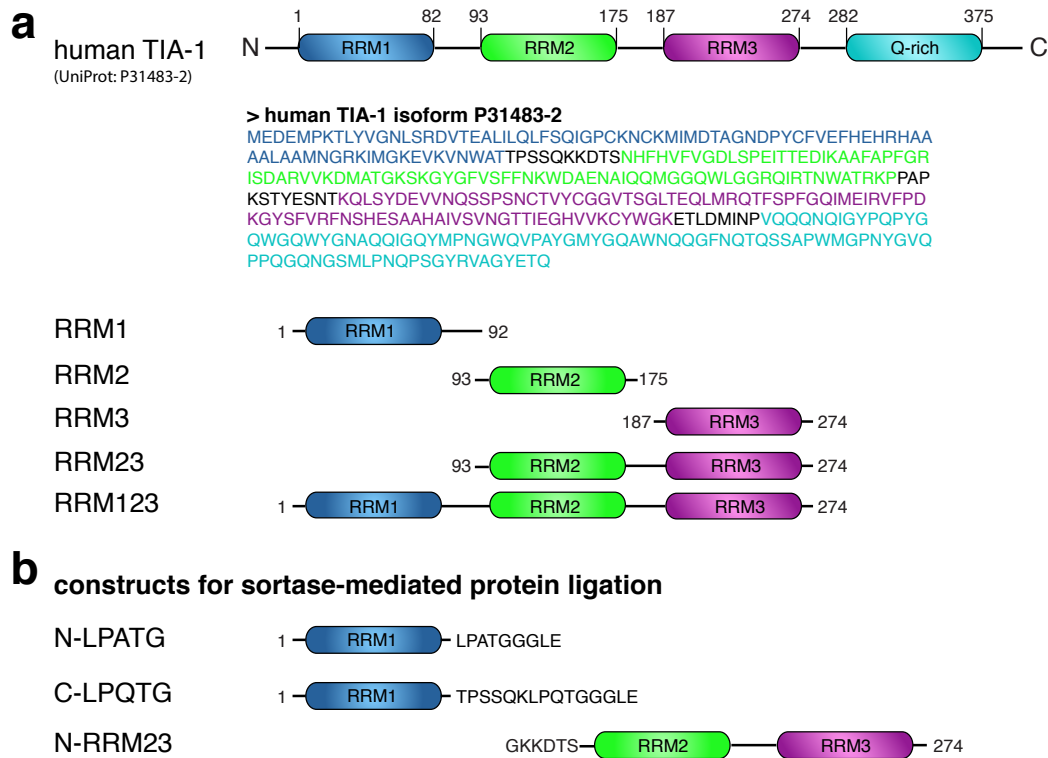


Figure 2.1.: TIA-1 constructs used in this thesis. a) All constructs originated from the human Nucleolysin TIA-1 isoform p40 (UniProt: P31483-2). TIA-1 harbours three canonical RRM domains (blue, green and purple), followed by an unstructured Q-rich domain (cyan). The sequence of the full-length protein is provided below the cartoon representation in the FASTA format. The sequence of each domain is highlighted according to the cartoon representation. Individual domains (RRM1, RRM2, RRM3), a tandem (RRM23) and a triple (RRM123) construct were used in this thesis. Note that the RRM3 domain contained an additional helix (α_0) preceding the canonical RRM fold (residues 191-196).^{38,51} b) Constructs for sortase-mediated protein ligation. The sortase recognition site (LPXTG), where X can be D, E, A, N, Q or K,⁵² was placed in the RRM1-RRM2 linker. In the N-LPATG construct, the first five amino acids (aa) of the linker were mutated to LPATG. In the C-LPQTG construct, the last five amino acids of the linker were mutated to LPQTG. In each case, the recognition site was protected by four additional aa (GGLE). In order to keep the linker length constant (10–11 aa), a new RRM23 domain (N-RRM23), containing the last five aa of the original RRM1-2 linker was designed to be ligated with the N-LPATG construct.

Two different LPXTG-specific sortase A variants from *S. aureus*

a

	His-tag	GB1 fusion tag
srt A wild type	MKHHHHHPMKQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWYDDATK	
srt A mutant	MKHHHHHPMKQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWYDDATK	
	TEV cleavage site	N-terminal unstructured part
srt A wild type	TFTVTEGSGSGSENLYFQGSAM	AYLFAKPHIDNYLHDKDKDEKIEQYDKNVKEQASKDKKQQ
srt A mutant	TFTVTEGSGSGSENLYFQGSAM
	60	80
srt A wild type	AKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSF AEENESLDDQNI SIAGHTF	
srt A mutant	AKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATSEQLNRGVSF AEENESLDDQNI SIAGHTF	
	140	160
srt A wild type	IDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVGVLDEQK GKDKQLTLI	
srt A mutant	IDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTAVGV LDEQK GKDKQLTLI	
	200	180
srt A wild type	TCDDYNEKTGVWEKRKIFVATEVK	
srt A mutant	TCDDYNEKTGVWETR KIFVATEVK	

b

> sortase wild type GS variant

```
MKHHHHHPMKQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEW
TYDDATKTFTVTEGSGSGSENLYFQGSAMAYLFAKPHIDNYLHDKDKDEKIEQYD
KNVKEQASKDKKQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRG
VSFAEENESLDDQNI SIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYK
MTSIRDVKPTDVGVLDEQK GKDKQLTLITCDDYNEKTGVWXRKIFVATEVK
```

> sortase mutant GS variant

```
MKHHHHHPMKQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEW
TYDDATKTFTVTEGSGSGSENLYFQGSAMAKPQIPKDKSKVAGYIEIPDADIKEPV
YGPATSEQLNRGVSF AEENESLDDQNI SIAGHTFIDRPNYQFTNLKAAKKGSM
VYFKVGNETRKYKMTSIRNVKPTAVGV LDEQK GKDKQLTLITCDDYNEKTGVW
ETRKIFVATEVK
```

Figure 2.2.: LPXTG-specific srtA variants from *S. aureus*. a) Sequence alignment of the two srtA constructs (derived from UniProt: Q9S446). Each construct contained a N-terminal His-tag fused to GB1, and a modified TEV cleavage site. The ENLYFQG motif was mutated to ENLYFQS and is highlighted in cyan. The mutant contained four point mutations (P94S, D160N, D165A and K196T),⁵³ indicated by magenta stars and lacked the unstructured N-terminal domain. The N-terminal domain is required for the activity of the wild type srtA construct. b) The sequences for each sortase construct are provided in the FASTA format. As the G was mutated to an S in the ENLYFQG TEV recognition site, both constructs contain 'GS variant' in their name.

chemical ligation (NCL), protein *trans*-splicing (PTS), expressed protein ligation (EPL) and sortase-mediated ligation are employed for protein NMR.^{49,55–59}

Of all established labeling techniques, the recently introduced sortase-mediated ligation provides high specificity and easy implementation. It only requires the presence of the enzyme recognition motif at the C-terminus of one of the two proteins to be ligated (Figure 2.3 b). SrtA from *S. aureus*, is a thiol transpeptidase which covalently attaches LPXTG (X=D, E, A, N, Q or K)⁵² containing peptides to the pentaglycin cross-bridge of the cell wall peptidoglycan. The enzyme cleaves between T and G of the LPXTG recognition motif, forming an acyl-enzyme intermediate (Figure 2.3 b). In the presence of a nucleophile like pentaglycine and after the formation of the intermediate, the carboxyl group of T is covalently bound to the amide group of the N-terminal glycine.^{49,59,60} In the absence of a nucleophile, however, srtA catalyses the hydrolysis rather than the transpeptidation reaction of the LPXTG containing peptide.⁶¹

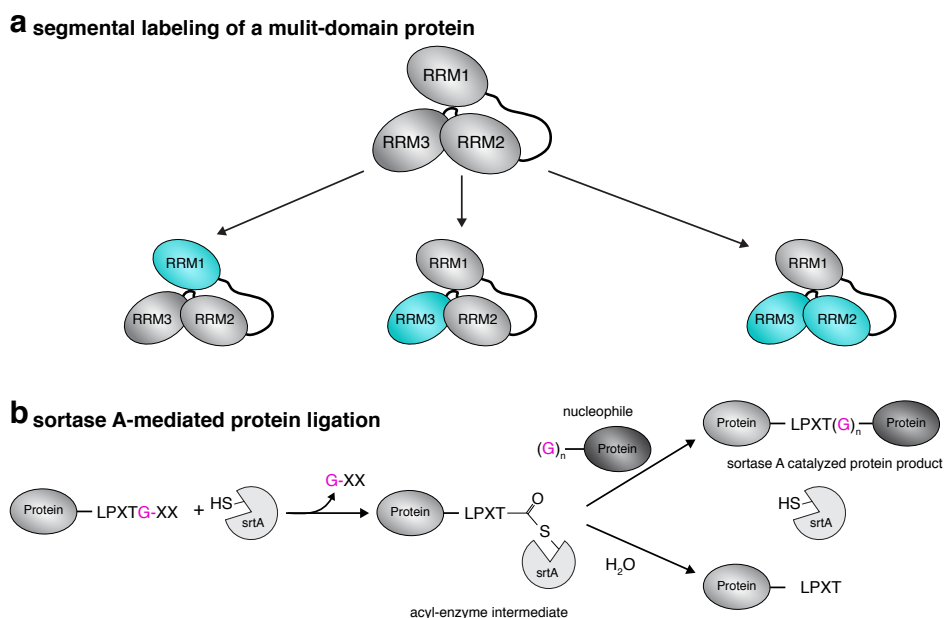


Figure 2.3.: Segmental isotope labeling of multi-domain proteins. a) Individual domains within the full-length protein can be isotopically labeled. Labeled domains within the intact multi-domain protein are highlighted in cyan. b) Schematic representation of sortase A-mediated protein ligation. SrtA recognizes LPXTG-containing proteins and cleaves between the T and G (magenta font) of the recognition motif. An acyl-enzyme intermediate is formed. Two different reactions, hydrolysis and transpeptidation, are catalysed by srtA. In the presence of a nucleophile the transpeptidation reaction occurs, linking the nucleophile to the LPXTG-containing protein. In the absence of a nucleophile the acyl-enzyme intermediate is hydrolysed into LPXT-containing proteins and free srtA. Note that the enzyme is recovered after each catalyzed reaction.

In this thesis sortase-mediated protein ligation was applied to generate domain-selective labeled TIA-1 RRM123 (~ 30.5 kDa) samples. In order to analyze the role of RRM1 upon RNA binding, the enzyme recognition motif was introduced in the RRM1–RRM2 linker (Figure 2.1 b). To ensure that the linker modifications do not affect the overall conformation, two different LPXTG-containing RRM1 domains were designed (Figure 2.1 b). To obtain domain-selective labeled samples only one of the two precursors was expressed in isotope enriched M9 minimal media whereas the other one was expressed in LB (Figure 2.4 a and b). In the subsequent

ligation reaction isotope enriched and unlabeled precursors were mixed in the presence of *srtA* (Figure 2.4 c and d). A detailed description of sample preparation and the ligation reaction setup can be found in ref. 50.

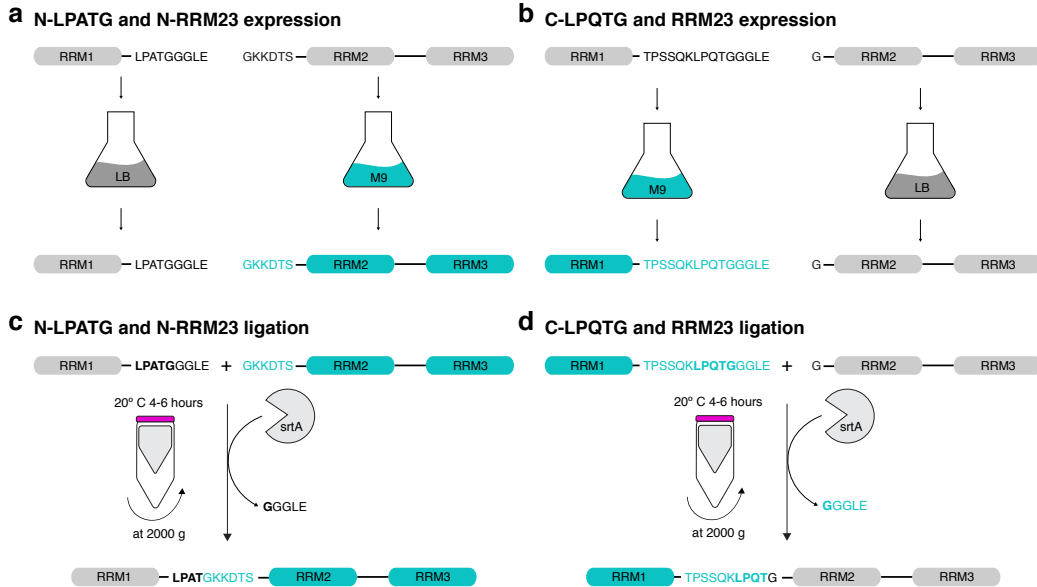


Figure 2.4.: Sortase-mediated protein ligation for two different TIA-1 RRM123 constructs. a) N-LPATG and N-RRM23 expression. N-terminal (N-LPATG) precursor expression in LB media generates unlabeled RRM1 domains. C-terminal (N-RRM23) precursor expression in isotope enriched M9 minimal media generates isotope labeled RRM23 domains. b) C-LPQTG and RRM23 expression. N-terminal (C-LPQTG) precursor expression in isotope enriched M9 minimal media generates isotope labeled RRM1 domains. C-terminal (RRM23) precursors expression in LB media generates unlabeled RRM23 domains. c) and d) Sortase ligation reaction. The purified N- and C-terminal domains were mixed with sortase in a 1:2:2 molar ratio. The ligation reaction was performed in a centrifugal unit while constantly spinning for 4–6 h at 20 °C.⁵⁰ The enzyme recognition motif is highlighted in bold.

2.3. Small angle scattering experiments

Small angle scattering (SAS) of neutrons or X-rays is a powerful technique to observe large-scale conformational or structural changes, as well as interactions of different biological macromolecules ranging from one kDa up to several MDa under near-native conditions.⁶² SAS focusses on coherent or elastic scattering meaning that only the direction but not the magnitude of the incident X-ray or neutron beam changes upon interaction with the molecules in solution. While X-rays interact with the electrons surrounding the atomic nuclei, neutrons directly interact with the nucleus itself. In each case, the incident wave vector is scattered at the molecules at an angle of 2θ (Figure 2.5).⁶³ As molecules typically consist of an ensemble of scattering centers, an interference pattern characteristic for their internal structure is produced at the detector. The intensity of the interference pattern depends on the relative orientation and the distance of the emitting atoms to each other and is recorded as function of the momentum transfer q .⁶⁴

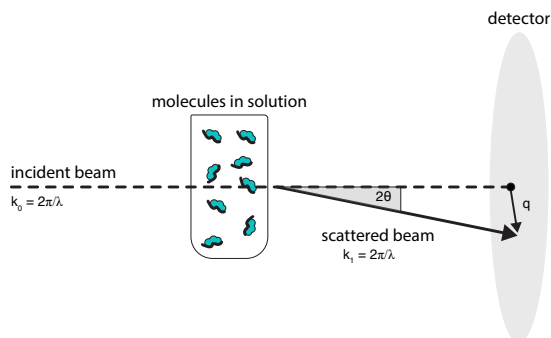


Figure 2.5.: Basic SAS setup. The incident, monochromatic X-ray or neutron beam is elastically scattered by the molecules in solution at an angle of 2θ . The scattering vector q describes the change in direction of the incident wave vector.

$$q = \frac{4\pi}{\lambda} \sin \theta \quad [q] = \text{nm}^{-1} \quad (2.1)$$

In the ideal case of a monodisperse, dilute sample, all particles are randomly orientated without any inter-particle interactions. Thus the observed scattering intensity $I(q)$ of the entire ensemble is proportional to the scattering of one particle averaged over all possible orientations.⁶³ Typically, SAS patterns are recorded at very small angles, for q values ranging between 0.01 – 5 nm^{-1} (Figure 2.6) and presented as radially averaged one-dimensional curves.^{62, 63} Within that range, the size and shape of the molecules determines the scattering curve. This effect, however, decreases as q increases. Without any further structural information, SAS patterns and *ab-initio* modeling using Monte Carlo-based minimization methods and simple constraints provide information on the overall shape, size and oligomeric state of the molecule. Simple constraints such as compactness and connectivity only report on the protein's envelope (Figure 2.6) but fail to dissect structural details or local structures of individual domains within a multi-domain complex. When the atomic structure of the molecule is known, however, SAS data can be used for structure validation and identification of biological active oligomers in solution. Theoretical SAS curves, back-calculated from the atomic structure, can be scored against the experimental curves. Assessed by χ^2 , SAS data can help to discriminate between different structural models.^{65–67}

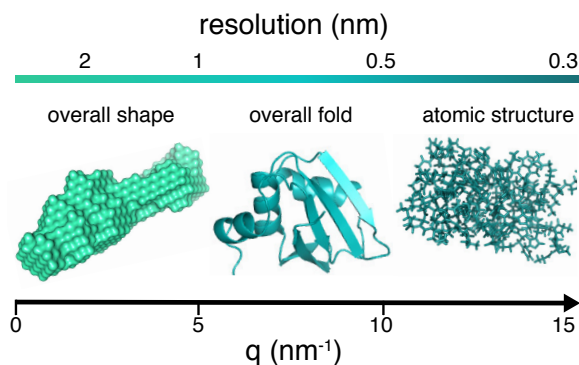


Figure 2.6.: Extent of structural information depending on the measured q range. The corresponding X-ray resolution ($\Delta = \frac{2\pi}{q}$) is displayed on top.

The scattering intensity depends further on the contrast $\Delta\rho$.⁶⁴ When the average scattering length density of the solvent ($\Delta\rho_s$) matches the average scattering length density of the particle ($\Delta\rho_p$), the particle becomes invisible and no SAS pattern is observed. For X-rays the average scattering length density is directly proportional to the atomic number,^{63,64} the higher the electron density the larger $\Delta\rho$. Consequently, the particle becomes more visible as $\Delta\rho$ and $I(q)$ increase. For neutrons, however, there is an irregular relationship between the average scattering length density and the atomic weight.^{63,68} The most prominent example is the neutron scattering length density difference between hydrogen (^1H) and deuterium (D). For ^1H the neutron scattering length density is negative whereas for D it is positive. Therefore, by adjusting the D_2O concentration in aqueous solutions (Figure 2.7) components with different scattering densities can be rendered invisible.⁶⁹ Combined with selective perdeuteration, contrast-matching SANS experiments can provide information on the location and orientation of individual subunits or domains within a segmentally perdeuterated protein complexes (Figure 2.8).⁴⁷

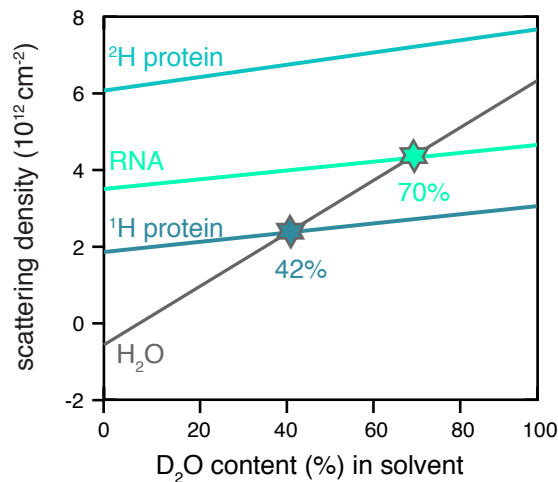


Figure 2.7.: Scattering densities as a function of D_2O concentration in aqueous solvents for protonated and perdeuterated proteins as well as RNA (according to Jacrot⁶⁹). Contrast matching points are highlighted by stars and the corresponding D_2O concentration is given below. At 42% D_2O in the solvent protonated proteins are rendered invisible. At 70% D_2O in the solvent RNA becomes invisible.

2.3.1. Small angle X-ray scattering

Small angle X-ray scattering (SAXS) experiments were recorded on TIA-1 RRM23, wild type RRM123 as well as on both sortase ligated RRM123 samples. Each samples was measured in the free and RNA bound state with at least three concentrations ranging between 1–10 mg mL⁻¹. For all RNA bound samples, an excess of RNA (~ 1.2 equivalents) was added to the protein. After incubation of at least 30 min, the samples were subjected to size exclusion chromatography to remove unbound RNA. TIA-1 samples were measured in 10 mM potassium phosphate pH 6.0, 50 mM NaCl and 10 mM dithiothreitol (DTT). 80 μL of each TIA-1 RRM123 sample (free and bound to U15 (synthesized, IBA GmbH)) were measured on the home source beamline Rigaku BIOSAXS1000 at TUM, Garching at 20 °C. 40 μL of TIA-1 RRM23 free and bound to *fas*

intron 6 10mer RNA (UGCUUUGUUC) (synthesized, IBA GmbH) were measured on BM29 at ESRF, Grenoble⁷⁰ at 25 °C. Background subtraction and circular averaging for scattering profiles recorded on the Rigaku BIOSAXS1000 beamline were performed with the Rigaku SAXSLab software v 3.0.1 r 1. BM29 derived scattering data was processed with the BsxCuBE software. Each one-dimensional scattering curve, expressed as a function of the momentum transfer q (eq. 2.1), was further analysed using the ATSAS package v.3.0.2.⁷¹ The radii of gyration (R_g) of all samples were extracted using the Guiner approximation in PRIMUS.⁷² In each case the validity of the Guinier approximation, R_g for $Q < 1.3$, was verified and fulfilled. The pairwise distribution functions and consequently D_{\max} values were calculated using GNOM.⁷³

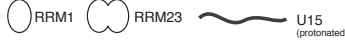
2.3.2. Small angle neutron scattering

Small angle neutron scattering (SANS) experiments were performed at the large dynamic range diffractometers KWS-1 and KWS-2 at the Heinz Maier-Leibnitz Zentrum⁷⁴ and at the D22 beamline at the ILL, Grenoble.⁷⁵ Contrast-matching SANS experiments were recorded on domain-selective perdeuterated TIA-1 RRM123 samples. As the aforementioned SAXS experiments revealed no difference between both sortase ligated samples and the wild type, the C-LPQTG construct (Figure 2.4 b and d) was arbitrarily chosen for SANS measurements. In total 12 different samples at three different D_2O contents (0 %, 42 % and 70 %) in the solvent, summarized in Figure 2.8, were measured at a single concentration of 5 mg mL⁻¹. For RNA bound samples, 1.05 equivalents of U15 were added to the protein. 200 μ L of each sample in 10 mM potassium phosphate pH 6.0, 50 mM NaCl and 10 mM DTT were measured at 25 °C. As a control, wild type TIA-1 RRM123 was measured under the same conditions. Instrumental details and SANS data reduction are described in ref. 48.

2.4. Biophysical characterization of TIA-1

In order to increase RNA binding affinity, multi-domain proteins specifically arrange their RRM domains to provide an extended binding platform.^{76,77} The structural arrangement of domains within a multi-domain protein and their mode of interaction upon ligand binding can be quite diverse and lead to changes in size and shape of the protein. RNA binding involves the formation and breakage of non-covalent interactions like hydrogen bonds, electrostatic, hydrophobic or aromatic stacking interactions^{76,77} which leads to the release of solvent molecules. In order to fully understand the mode of RNA recognition and relate the structure to its biological function, it is necessary to determine the underlying thermodynamics of a binding event and study the conformational changes induced upon recognition.

In this thesis, SLS and ITC experiments were combined to determine the stoichiometry of TIA-1 RRM23/UGCUUUGUUC and RRM123/U15 complexes and compare binding affinities of single RRM2 and RRM3 with the tandem RRM23 domains.



Buffer	free	+ RNA	Labeling
0 % D ₂ O			¹ H RRM1 - ² H RRM23
			² H RRM1 - ¹ H RRM23
42 % D ₂ O			² H RRM1 - ¹ H RRM23
			¹ H RRM1 - ² H RRM23
70 % D ₂ O			² H RRM1 - ¹ H RRM23
			¹ H RRM1 - ² H RRM23

Figure 2.8.: Contrast matching and the domain-selective perdeuterated TIA-1 proteins for SANS measurements. Schematic representations of each component in the sample are given above the table. Perdeuterated domains are coloured dark gray, whereas protonated domains are coloured light gray. At 42 % D₂O in the buffer, the scattering density of the solvent (gray box) matches the scattering density of the protonated domain. At 70 % D₂O in the buffer, the scattering density of the solvent (dark gray box) matches the scattering density of the RNA. Consequently at 42 % the protonated domain becomes invisible whereas at 70 % D₂O the RNA is rendered invisible.

2.4.1. Static light scattering

Static light scattering (SLS) also known as Rayleigh scattering is an optical method to determine the absolute molecular weight (M_W) and radius of gyration (R_g) of molecules in solution.^{78,79} It further assesses the oligomeric state and stoichiometry of complexes, e.g. protein–RNA complexes.⁸⁰ If combined with gel permeation (GPC) or size exclusion (SEC) chromatography, the composition of a macromolecular complex can be defined by separating the different species of the complex and determining their corresponding M_W .⁸¹ Whenever light hits a molecule, the photons of the incident beam initiate an oscillation of the electrons. Upon interaction, the incoming light is scattered in all directions by the molecule in solution. SLS measures the excess light scattering (LS) at a constant angle of θ .⁸² Typically, proteins are smaller than $\frac{1}{20}$ of the incident wavelength, thus the Rayleigh equation simplifies to:⁸³

$$\frac{Kc}{R_{(\theta,c)}} = \frac{1}{M_W} + 2A_2c \quad (2.2)$$

where K is an optical constant, c the concentration, θ the measured angle, R the Rayleigh ratio (ratio of scattered light intensity to incoming light intensity) measured as a function of θ , M_W the weight averaged molecular weight and A_2 the second virial coefficient. From equation 2.2 it can be seen, that the intensity of scattered light at a given angle θ is directly proportional to the M_W and the concentration of the molecule.

Protein–RNA complex formation for TIA-1 RRM23 as well as RRM123 were characterized using SLS. Moreover, SLS was used to address differences between both sortase ligated samples and wild type RRM123. TIA-1 RRM23/UGC UUUGUUC complex was measured in 10 mM

potassium phosphate pH 6.0, 50 mM NaCl and 10 mM DTT whereas all RRM123 samples bound to U15 were measured in 50 mM potassium phosphate pH 6.0, 100 mM NaCl and 1 mM 2-mercaptoethanol. The TIA-1 RRM123 samples were directly recovered from ITC experiments. Higher salt and buffer concentrations improved the ITC data. In each case, the SLS instrument (Viscotek TDA 305, Malvern Instruments) was connected to an analytical size exclusion column (Superdex 75 or 200, 10/300 GL, GE Healthcare). The detector was calibrated with a 4 mg mL⁻¹ bovine serum albumin (BSA) and the monomeric molecular weight of 64 kDa. Typical $\frac{dn}{dc}$ values of 0.185 mL g⁻¹ for proteins and 0.165 mL g⁻¹ for RNA were used for M_W calculations from right angle scattering (RALS).⁴⁸

2.4.2. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC), is a biophysical technique to study binding affinities and thermodynamics of interacting proteins. The instrument is equipped with two identical cells, a sample and a reference cell (Figure 2.9 a). Both cells are surrounded by a nearly adiabatic jacket and kept at thermal equilibrium. During an experiment, a precise volume of ligand is titrated into the sample cell (Figure 2.9 a). Upon binding heat is either released or taken up. The quantity of heat is directly related to the extend of binding. During each injection the instrument measures the electrical power necessary to keep both cells at the same temperature ($\Delta T = 0$).^{84,85} Over time, the binding sites of the molecule become saturated and the heat changes decay to zero. If the exact concentration of both ligand and target is known the heat changes can be integrated over time and plotted against the molar ratio of ligand to target. The reaction enthalpy ΔH can be directly derived from the resulting sigmoidal curve. Furthermore, binding affinities (K_a) and the stoichiometry (n) of a reaction can be obtained after fitting the curve to a binding model.⁸⁴ Knowing K_a and ΔH the binding entropy (ΔS) can be calculated according to Gibb's equation:

$$\Delta S = \frac{(\Delta H - \Delta G)}{T} \quad (2.3)$$

with

$$\Delta G = \Delta H - T\Delta S \quad (2.4)$$

and

$$\Delta G = -RT \ln K_a \quad (2.5)$$

where T is the measurement temperature in Kelvin and R the gas constant.

Binding affinities of different TIA-1 constructs (RRM2, RRM3 and RRM23) to U-rich RNA sequences differing in length and composition were characterized (Figure 2.9 b). 400 μ M of U4, U5, U6 or C-rich 5 mer oligonucleotides were titrated to 30 μ M of either RRM2 or RRM3 in the cell. 200–250 μ M intron 6 derived 10 mer RNA (UGCUUUGUUC) were injected to 20 μ M RRM23. The individual domains were dialysed against 50 mM potassium phosphate pH 6.0, 100 mM NaCl and 1 mM 2-mercaptoethanol. To be consistent with NMR measurements, RRM23 was dialysed against 10 mM potassium phosphate pH 6.0, 50 mM NaCl. The dialysis was used to prepare the RNA solution and to provide a baseline. All measurements were carried out at 25 °C on a MicroCal iTC200 calorimeter (Microcal, Northhampton, USA). 40 μ L

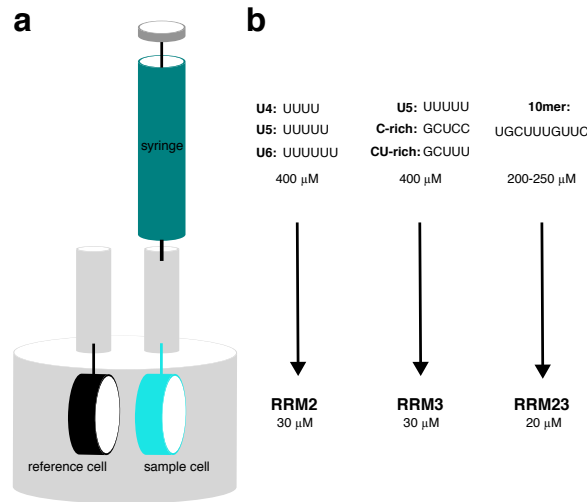


Figure 2.9.: Isothermal titration calorimetry of different TIA-1 constructs. a) Standard ITC setup. b) In each case the RNA was titrated into the protein in the sample cell. See text for details.

of RNA were injected to the protein in 180 s intervals of 1 μ L, while stirring at 750 revolutions per minute (rpm). The raw data was integrated and normalized using Origin ITC Analysis software provided by Microcal. The heat change per injection was plotted against the molar ratio of ligand to target. The resulting curves were fit to an appropriate binding model (one-site or two-sites binding).

2.5. NMR experiments

NMR spectroscopy is a powerful tool to understand biological processes at atomic level. In contrast to X-ray crystallography and electron microscopy, solution state NMR enables the study of not only the structure, but also the inherent motions and dynamics of the molecule. The basis for an NMR signal is the magnetic moment μ of an atomic nucleus (Figure 2.10). The intrinsic angular momentum of a nucleus is determined by the spin quantum number (I) which depends mainly on the number of unpaired protons and neutrons in the nucleus. The magnitude of the magnetic moment and angular momentum are directly proportional to a nucleus-specific factor, termed gyromagnetic ratio γ .^{86,87}

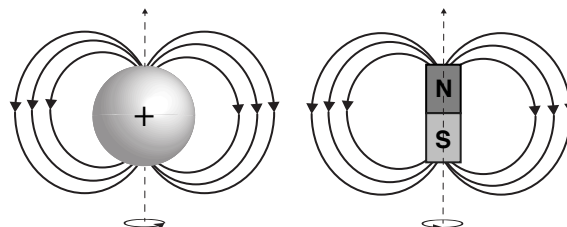


Figure 2.10.: Nuclear magnetic moment. Nuclei with a non-zero quantum number I possess an intrinsic angular momentum. The precession of the spins generates a magnetic moment similar to a rotating bar magnet.

$$\vec{\mu} \propto \vec{I} \qquad \vec{\mu} = \gamma \vec{I} \qquad (2.6)$$

Bio-NMR spectroscopy focuses on spin $\frac{1}{2}$ nuclei, which includes protons, ^{15}N or ^{13}C isotopes. When placed in an external magnetic field B_0 , the magnetic moment of a spin $\frac{1}{2}$ can adopt only two allowed states, $+\frac{1}{2}$ and $-\frac{1}{2}$, of different energies. The spins are pointing along or in the opposite direction of the external magnetic field, respectively (Figure 2.11) and rotating about the magnetic field with an intrinsic frequency, termed Larmor frequency ω_0 . The energy difference between these two states depends on B_0 with:^{86,87}

$$\Delta E = \gamma \hbar B_0 \quad (2.7)$$

where \hbar is the reduced Planck constant. The spins are distributed among the two energy states according to the Boltzmann distribution. The higher the applied magnetic field strength is, the larger ΔE (Figure 2.11) and thus the population difference between the two states becomes.^{86,87}

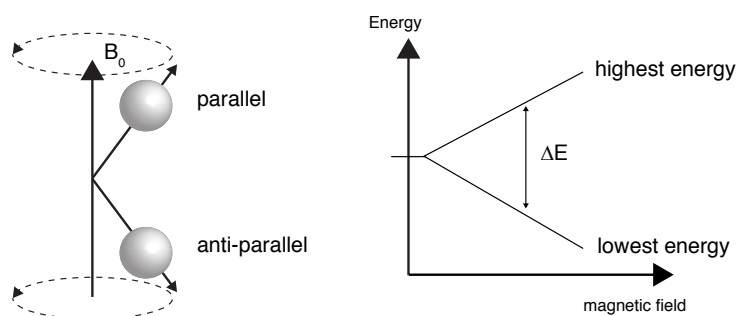


Figure 2.11.: Effect of an external magnetic field on a spin $\frac{1}{2}$ particle. Spin $\frac{1}{2}$ align either along or opposed to B_0 . Each spin state has a distinct energy level.

To observe magnetic resonance, a short radio frequency radiation is applied to the sample. When the frequency ν of the radiation matches the Larmor frequency of the nuclei ($\nu = \omega_0$), with

$$\omega_0 = -\gamma B_0 \quad (2.8)$$

resonance is reached. This causes spin flips from the low energy state $m = +\frac{1}{2}$ to the high energy state $m = -\frac{1}{2}$ (for nuclei with positive γ). This energy transition gives rise to a detectable NMR signal. The signal intensity depends on the population difference of the two states. The higher the population difference is, the higher the net absorption of energy and thus the detectable signal becomes. Each nucleus within a molecule resonates with a slightly different frequency in B_0 , depending on the local electron distribution. This effect is called chemical shift and allows to distinguish between all protons within a protein.

2.5.1. Resonance assignment

NMR spectroscopy provides information not only on chemical shifts but also the magnetic couplings of nuclei through bonds or space. These interactions can be used to characterize the structure of biological macromolecules at atomic resolution under conditions close to their physiological state.

Spin-spin interactions mediated by the electrons in chemical bonds are called scalar or J couplings. These interactions will cause NMR lines to split into characteristic multiplets. The spacing between splitted peaks is referred to as the coupling constant J and correlates with the strength of the spin-spin interaction.^{86,87} Couplings are typically observed for nuclei separated by 1–3 bonds. In 2D heteronuclear NMR experiments the magnetization is transferred from e.g. ^1H to the attached ^{15}N nuclei (Figure 2.12 left panel), providing a distinct ^1H – ^{15}N correlation, characteristic for each amino acid. In 3D heteronuclear experiments the magnetization is for instance transferred via INEPT (insensitive nuclei enhanced by polarization transfer)⁸⁸ from ^1H to ^{15}N and subsequently to ^{13}C (Figure 2.12 right panel). A series of 2D and 3D heteronuclear experiments is recorded to sequentially assign the backbone and side-chain resonances of the protein of interest.

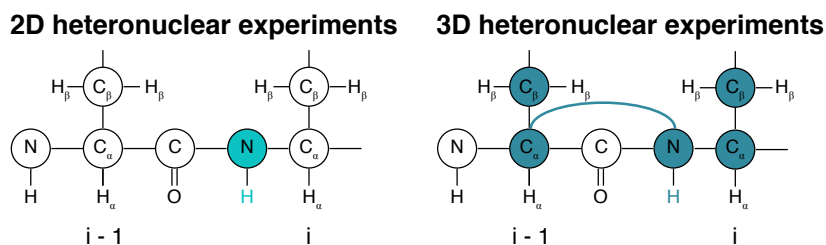


Figure 2.12.: 2D and 3D heteronuclear NMR experiments for protein backbone assignment. In a standard 2D ^1H , ^{15}N HSQC (heteronuclear single quantum coherence) experiment, the magnetization is transferred across the ^1H – ^{15}N bond. Whereas in an 3D HNCACB experiment, the magnetization is transferred from the ^1H to ^{15}N and subsequently to ^{13}C . Consequently, the C_{α} and C_{β} of residue (i) can be linked to the preceding amino acid (i-1).

Besides J-coupling, two spins can interact through space via dipolar cross-relaxation (Figure 2.13).^{86,87} Imaging two spins I and S separated by less than 5–6 Å. Applying a strong selective radio frequency pulse on spin I saturates both I transitions but does not have an effect on the population difference of spin S. Cross-relaxation restores the equilibrium distribution of spin I by transferring z magnetization to its dipolar coupled spin S. This energy transfer either decreases or increases the peak intensity of spin S. This effect termed Nuclear Overhauser effect (NOE) provides information on inter-atomic distances and is valuable for refining the secondary and tertiary structure of the studied protein.

Assignment experiments were acquired on AVIII600, AVII750 and AVIII800 Bruker spectrometers equipped with cryogenic or room temperature (750 MHz) triple resonance gradient probes at 298 K. The sample contained 0.7 mM double labeled TIA-1 RRM23 bound to 1.5 molar excess of UGCUUUGUUC in 10 mM potassium phosphate adjusted to pH 6.0, 50 mM NaCl, 10 mM DTT, 0.02 % NaN_3 with 10 % D_2O added for the lock. All spectra were processed with the NMRPipe/Draw⁸⁹ software. Protein backbone assignments for ^1HN , ^{15}N , $^{13}\text{C}_{\alpha}$ and $^{13}\text{C}_{\beta}$ chemical shifts were obtained from HNCA and HNCACB spectra (Figure 2.12). Side-chain assignments were received from HAHB(CO)NH, HCCH-TOCSY (total correlated spectroscopy) and ^{13}C -HSQC spectra. The assignment was done using CARA (Computer-aided resonance assignment by Kurt Wüthrich).⁹⁰

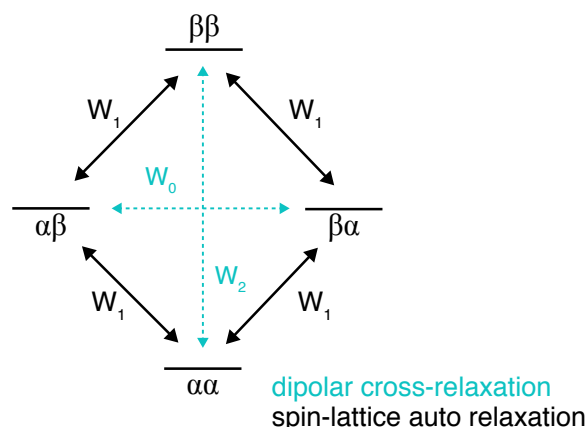


Figure 2.13.: The NOE effect. Energy levels for two spin $\frac{1}{2}$ nuclei, I and S are depicted. Besides 4 single quantum transitions, one double quantum (W_2) and one zero (W_0) transitions are observed. W_0 and W_2 are termed dipolar cross-relaxation.

2.5.2. Relaxation measurements

Relaxation in NMR describes the process which allows the spins to return to equilibrium magnetization. At thermal equilibrium, if the external B_0 points along the z -axis, the z -component of the magnetic moment of each nucleus will align to B_0 (Figure 2.14 a middle panel), whereas the x - and y -components are randomly distributed. This results, summed over all spins in the sample, in a net or bulk magnetization (M_0) along z (Figure 2.14 a right panel) but no transverse magnetization along x and y . The time to reach the Boltzmann distribution is called T_1 or spin-lattice relaxation time (Figure 2.14 b).

When a short 90° radio frequency pulse is applied to M_0 (Figure 2.14 c left panel), the magnetization M precesses about the z -axis with the Larmor frequency ω_0 (Figure 2.14 c right panel). Each individual spin possesses a slightly different precession frequency as it experiences not only the static but also local magnetic fields generated by nearby spins. Over time, the precession about z will lose coherence (Figure 2.14 d). As a consequence, transverse magnetization decays to zero with a characteristic spin-spin relaxation time T_2 (Figure 2.14 d) whereas T_1 relaxation restores the equilibrium magnetization along z (Figure 2.14 b). Nuclear spin relaxation is mediated by dipole-dipole interactions of neighbouring spins and chemical shift anisotropy. Owing to the thermal motions of molecules in solution, the magnetic interaction between two nuclei fluctuates rapidly, generating time-dependant local magnetic fields which can induce relaxation.

In solution, molecules collide more frequently than in gases. Each collision between molecules will randomly deflect the axis and angle of rotation in small steps. This chaotic motion is called rotational diffusion. The overall rotational correlation time τ_c describes the average time a molecule needs to tumble through an angle of 1 rad. Larger molecules tumble in average slower, whereas small molecules rotate faster which results in shorter correlation times.

In order to determine the overall rotational correlation time of TIA-1 RRM23 bound to 10 mer intron 6 derived RNA (UGCUUUGUUC) ^{15}N T_1 , T_2 relaxation as well as steady state heteronuclear ^{15}N NOE data was collected at a 750 MHz AVII750 Bruker spectrometer, equipped with a room temperature triple resonance gradient probe, at 298 K. The sample

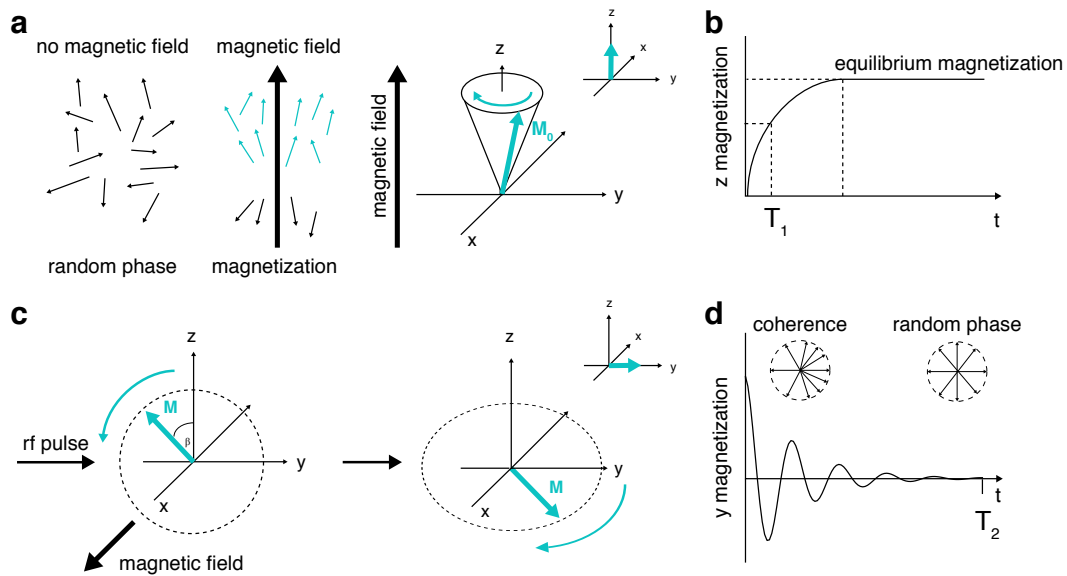


Figure 2.14.: Schematic representation of T_1 and T_2 relaxation. a) Effect of an external magnetic field on the magnetic moments of the individual spin $\frac{1}{2}$ particles. Left panel: In the absence of a strong magnetic field, the magnetic moments of the individual spins are randomly distributed. Middle panel: When a strong magnetic field B_0 pointing along the z -axis is applied, the magnetic moments of the individual spins are distributed between the two allowed energy states, along or opposed to the direction of B_0 . Right panel: At equilibrium, the low energy state (along the field direction) is slightly more populated which results in a net or bulk magnetization M_0 (cyan arrow) along z . M_0 is tilted away from the z -axis and precesses about B_0 with the Larmor frequency ω_0 . The rotational motion around B_0 describes a cone of constant angle to the z -axis. b) The time to build up the equilibrium magnetization along z is called T_1 relaxation time. c) Effect of a 90° radio frequency pulse along the x -axis. Left panel: When a pulse along the x -axis is applied, the magnetization M rotates into the yz -plane at an angle of β , generating a net y magnetization (right panel). Right panel: M precesses in the xy -plane around B_0 with ω_0 . c) Over time, the magnetic moments of the individual spins will lose coherence in the xy -plane. Consequently, the y - or transverse magnetization decays to zero with a characteristic time T_2 .

contained 0.7 mM double labeled TIA-1 RRM23 bound to 1.5 molar excess of UGCUUUGUUC in 10 mM potassium phosphate adjusted to pH 6.0, 50 mM NaCl, 10 mM DTT, 0.02 % NaN₃ with 10 % D₂O added for the lock. The relaxation rates and error calculation were determined using Pint.⁹¹ The overall correlation time τ_c was calculated as a function of the T_2 ¹⁵N-relaxation time, assuming that the protein-RNA complex rotates isotropically similar to a sphere:⁹²

$$\tau_c \approx \frac{1}{4\pi\nu_N} \sqrt{6\frac{T_1}{T_2} - 7} \quad (2.9)$$

where ν_N is the ¹⁵N resonance frequency. The overall tumbling time was obtained by averaging the individual tumbling times. Residues with negative heteronuclear NOEs (hetNOEs) or hetNOEs less than 0.65 as well as residues with poor resolution and low intensity were excluded from the calculation.

2.6. Crystallization

Proteins at high concentration can crystallize under specific conditions. Upon addition of ionic salts or water-binding molecules like polyethylene glycol (PEG), proteins slowly precipitate without denaturing. The molecules become ordered and form a three-dimensional lattice held together mainly by hydrogen bonds between their hydrated surfaces.⁹³ The crystalline form is characterized by a very high degree of regularity. If an X-ray beam is directed towards the crystal, the incoming beam is elastically scattered by the electrons of the molecules in the crystal unit in accordance to Bragg's Law. In contrast to the liquid state, the phase relations of the molecules to each other are fixed and repetitive, thus a scattering pattern with very well defined and sharp peaks is collected at the detector.⁹³ This effect is called diffraction. The electron distribution within the unit cell can be calculated from the diffraction pattern. The amino acid sequence can then be modelled into the electron density to reconstruct the three-dimensional atomic model of the protein.

Different sparse-matrix crystallization screens for structure determination of TIA-1 RRM23 bound to a *fas* intron 6 derived 10 mer oligonucleotide (UGCUUUGUUC) (synthesized, IBA GmbH) were set up at room temperature and 4 °C. Around 8 mg mL⁻¹ of RRM23 was incubated with 1.2 molar equivalents of RNA for 30 min on ice prior to crystallization. Typically, 100 nL of sample and 100 nL of reservoir solution were dispensed as sitting drop into each well of the screening plates. Crystals were obtained in 50 mM MES (2-morpholin-4-ylethanesulfonic acid) buffer pH 6.0, 5 mM magnesium sulfate and 5 % PEG 4000 after two months. Flash frozen in 30 % glycerol and mixed with reservoir solution, diffraction patterns were collected at beamline ID29 at ESRF, Grenoble⁹⁴ and processed with XDS.⁹⁵ The structure was solved by molecular replacement using Phaser⁹⁶ from the CCP4⁹⁷ and completed with several rounds of manual building in Coot⁹⁸ and refinement in Refmac.⁹⁹

3. Publications

This thesis is a publication based dissertation. Two original studies were published in international scientific journals are summarized in the following sections.

3.1. Efficient segmental isotope labeling of multi-domain proteins using Sortase A

The article, Efficient segmental isotope labeling of multi-domain proteins using Sortase A⁵⁰ appeared in the Journal of Biomolecular NMR and may be found at <http://dx.doi.org/10.1007/s10858-015-9981-0>. The author of this thesis, Miriam Sonntag, and Lee Freiburger contributed equally to the original work and writing of this article. In the following section citations of ref. 50 are omitted since it is a summary of this article.

Sortase A (srtA) mediated protein ligation – a recently introduced method^{49,100–105} – is a valuable tool for segmental isotope labeling of multi-domain proteins. Other than established methods, this enzyme based approach is highly specific and requires only the presence of an LPXTG (X=D, E, A, N, Q or K)⁵² recognition site at the C-terminus of the N-terminal domain and a glycine residue at the N-terminus of the C-terminal domain.^{59,60}

The article “Efficient segmental isotope labeling of multi-domain proteins using Sortase A” presents a robust, efficient and versatile protocol for high-yield domain-selective labeling of single-chain multi-domain proteins. Lee Freiburger and myself optimized the existing srtA protocol in two steps. Firstly, the ligation reaction was performed in a centrifugal concentrator with a molecular weight cut-off allowing the removal of the cleaved fragment (Figure 2.4 c and d). Secondly, we designed the N-terminal domain with a cleavable His-tag at the N-terminus and a non-cleavable His-tag at the C-terminus. This strategic use of His-tags eases the purification of the ligated product.

Two representative multi-domain proteins, the heat shock protein Hsp90 and the T-cell intracellular antigen-1 (TIA-1), were used to test our new protocol. I designed, expressed and purified two different TIA-1 constructs, an N-terminal domain including the srtA recognition site and the two His-tags and a C-terminal domain containing a glycine residue at the N-terminus. Moreover, I set up multiple srtA ligation reactions to screen for the optimum precursor concentrations, ligation temperature and time. For TIA-1 the ligation works best in a 1:2:2 molar ratio of the N-, C-terminal domain and srtA. Quenching the ligation reaction after 5 to 6 h at 20 °C yields sufficient amounts of ligated product but prevents degradation of the precursors. Furthermore, I directly compared two different ligation techniques, dialysis bag and centrifugal concentrator, in terms of final ligation yield. Active removal of the cleaved fragment by constant centrifugation significantly increased the final yield while shortening the ligation time. Due to the combination of cleavable and non-cleavable His-tags in the N-terminal domain, I was able to obtain pure, highly domain-selectively labeled TIA-1 samples with a TEV protease cleavage step in between two Ni-NTA purifications. Lee Freiburger conducted the same experiments with Hsp90.

The two optimizations of the reaction protocol are straightforward to implement and routinely yield milligram amounts of high quality samples, thereby expanding its application to study more complex systems by NMR and other biophysical methods.

3.2. Segmental, domain-selective perdeuteration and small angle neutron scattering for structural analysis of multi-domain proteins

The article, Segmental, domain-selective perdeuteration and small angle neutron scattering for structural analysis of multi-domain proteins⁴⁸ appeared in Angewandte Chemie and may be found at <http://dx.doi.org/10.1002/anie.201702904>. The author of this thesis, Miriam Sonntag, and Pravin Kumar Ankush Jagtap contributed equally to the original work and writing of this article. In the following section citations of ref. 48 are omitted since it is a summary of this article.

Multi-domain proteins, essential for most cellular processes, contain multiple globular domains connected by flexible intrinsically disordered linkers.^{106,107} Specific binding to proteins, DNA or RNA ligands are fundamental for their biological function. These interactions are tightly regulated by dynamic population shifts or cooperative assembly of the domains.^{44–46} To characterize the underlying conformational dynamics, different solution techniques as NMR and SAS experiments need to be combined.^{65,67,108–114}

The article “Segmental, domain-selective perdeuteration and small angle neutron scattering for structural analysis of multi-domain proteins” focuses on the use of domain-selective perdeuteration combined with contrast-matching⁴⁷ SANS experiments. TIA-1, an alternative splicing factor, harbours three RRM domains all connected by flexible linkers.³⁸ Efforts to obtain high resolution structures by crystallography or NOE-based NMR methods have so far been unsuccessful. To, nevertheless, gain insight into the molecular mechanism of target recognition by TIA-1 RRM123 and establish the role of RRM1 upon ligand interaction, I prepared different domain-selective perdeuterated TIA-1 RRM123 samples, in the presence and absence of U15 (Figure 2.8), using sortase mediated protein ligation.⁵⁰ For each sample I collected SANS data at the large dynamic range diffractometers KWS-1 and KWS-2 at the Heinz Maier-Leibnitz Zentrum.⁷⁴

To define the three-domain arrangement of the protein-RNA complex and gain insight into the mode of RNA recognition, Pravin Kumar Ankush Jagtap and myself determined the NMR solution structure of RRM1 (PDB ID: 5O2V) and the crystal structure of RRM2 bound to a UU dinucleotide (PDB ID: 5O3J), respectively. Based on both newly identified structures and the previously published NMR solution structure of α RRM3,³⁸ Bernd Simon and Jansoch Hennig employed a restrained molecular dynamics protocol. Consistent with the crystal structure, two out of five nucleotides (U6–U10), fixed to RRM2, were strictly restrained to the central RRM domain. The remaining nucleotides as well as all connecting linkers were randomized, generating an initial pool of 5000 structures. To refine the allowed conformational space of the three domains, the primary ensemble was scored against the experimental SAXS and SANS data. We clearly show, that the SANS data is highly discriminative, as it drastically reduces the conformational space to five models after the SAXS filter. Furthermore, the SANS

filter provides unique structural insight into the domain architecture of the protein-RNA complex, by precisely defining the domain positions with respect to each other.

The protocol presented, demonstrates the use and power of domain-selective perdeuteration combined with contrast matching SANS experiments for rigid body refinement of multi-domain proteins. It is applicable to a vast variety of multi-domain proteins and enables the study of conformational arrangements of individual domains and changes induced by ligand interaction.

4. Conclusions

Sortase-mediated protein ligation is a powerful technique to segmentally label multi-domain proteins for structural characterization by NMR or other biophysical methods. It has been shown that it only requires the presence of the enzyme recognition site (LPXTG) on the N-terminal domain and a Gly-residue in the C-terminal domain.^{49,59,60} The protocol optimized in this thesis⁵⁰ ensures the production of milligram amounts of segmentally labeled proteins of high quality. The protocol is applicable to a wide variety of proteins without requirement for extensive optimization of the ligation step itself. However, the recognition site can only be introduced in one segment of the protein. In case of a multi-domain protein containing three RRM domains, only one of the two connecting linkers can harbour the LPXTG-motif. Simultaneous insertion into both linkers would generate proteins differing in length and domain organisation, as the enzyme *srtA* does not distinguish between multiple recognition sites. Consequently, it is not possible to exclusively label the middle domain of multi-domain proteins. Over the last decade, however, the sortase construct has been optimized in terms of ligation efficiency, substrate specificity and reaction reversibility.^{59,60} Recently, it has been shown, that the combination of sortase variants with distinct recognition sites enables the construction of complex protein conjugates.¹¹⁵ It needs to be tested whether this approach can be applied to exclusively label the middle domain within multi-domain proteins for structural characterization by solution techniques such as NMR or contrast-matching SANS experiments.

Here, sortase-mediated protein ligation was used to domain-selectively perdeuterate TIA-1 RRM123. The sortase recognition site was placed in the RRM1-RRM2 linker to characterize the structure of the RNA bound form of RRM123 and assess the role of RRM1 upon binding, which is quite controversially discussed in literature.^{34,38,40} The solution structure of the single RRM1 domain (PDB ID: 5O2V) shows a typical canonical RRM fold, with two α -helices packed against four anti-parallel β -sheets.⁴⁸ The RNP-1 motif, however, contains two negatively charged residues Asp and Glu which could explain the lack of intrinsic RNA binding affinity of the single RRM1 domain. Furthermore, the presence of these two residues could account for the absence of an effect on the binding affinities of RRM12 and RRM123 to cellular RNAs, when compared to RRM2 and RRM23.^{21,34,38} Accordingly, a role in protein-protein rather than protein-RNA recognition has been discussed. Upon binding of TIA-1 to U-rich sequences downstream of alternative exons, U1 snRNP is recruited to the 5' splice site and splicing is initiated.^{32,33,35} Förch et al.²¹ demonstrated that U1C, a subunit of U1 snRNP, directly interacts with the Q-rich domain of TIA-1 and that this interaction is enhanced in the presence of RRM1. However, this study did not reveal any structural information on how RRM1 assists the Q-rich domain. Assuming, that only the α -helices of RRM1 are involved in protein-protein interaction, the β -sheets would still be able to bind RNA.^{76,77} Indeed, another study suggested a role of RRM1 in selective recognition of polyU RNA.⁴⁰ In order to shed light on the specific role of RRM1 upon RNA binding in context of the full-length protein, contrast-matching SANS experiments were recorded on domain-selective perdeuterated TIA-1 RRM123 (Figure 2.8). By adjusting the D₂O ratio in the solvent positive scattering contrast of protonated and perdeuterated proteins as well as RNA can be matched and thus rendered invisible (Figure 2.7 and Figure 2.8).^{47,69} To fully exploit the information provided by the one-dimensional scattering curves and reduce the ambiguity of SAS data interpretation,¹¹³ all

available structural information obtained by NMR and X-ray crystallography were combined and incorporated into rigid body modeling.⁴⁸

High resolution NMR structures of RRM1 (PDB ID: 5O2V)⁴⁸ and RRM23 (PDB ID: 2MJN)³⁸ in their free form have been solved in our lab. TIA-1, however, is an RNA binding protein involved in alternative splicing of various mRNAs.³² Structural investigation of the RNA bound form would consequently provide valuable information on the molecular mechanisms involved in target recognition and thus alternative splicing regulation. Different studies revealed that TIA-1 specifically binds to U-rich stretches \approx 10–28 nucleotides downstream of the exon/intron junction.^{19,21,32–35} A direct correlation with the uridine content of the intronic sequence and the effect of exon skipping upon TIA-1 depletion was thereby established.³² The central RRM2 domain including residues from the RRM2-RRM3 linker and the extra N-terminal helix (α_0) of the RRM3 domain mediate RNA recognition and determine the sequence specificity.^{34,41} Accordingly, an intronic U-rich sequence (UGCUUUGUUC) downstream of the alternative exon 6 of the *fas* mRNA was chosen to characterize the underlying mechanism of RNA recognition by the tandem TIA-1 RRM23 construct. ITC experiments showed that the single RRM2 domain (residue 93–175) binds short polyU stretches (U4–U6) in the micromolar range, whereas the single RRM3 domain (residue 187–274) lacks detectable RNA binding, even in the presence of C-rich pentamers.³⁹ When both RRMs are present U-rich RNA sequences like U9 and the *fas* intron 6 derived 10 mer oligonucleotide are bound in the nanomolar range. This drastic increase in affinity could be explained by the formation of a larger binding platform, which requires the association of two or more consecutive RRM domains.^{76,77} Indeed, NMR relaxation experiments (Figure A.1 c and d) together with SAXS data indicate that in the presence of RNA, both domains tumble together and adopt a more compact domain arrangement in which the linker keeps its flexibility. SLS and NMR titration experiments (Figure A.1 a and b) further suggest that the tested 10 mer oligonucleotide is bound with a 1:1 stoichiometry, involving all β -sheets, loops 3 and 4 of RRM2, the connecting linker as well as the extra helix (α_0) of RRM3. Similar to Sex-lethal,¹¹⁶ nucleolin,^{117,118} PABP¹¹⁹ or HuD¹²⁰ the connecting linker could arrange each domain in such a way that both can cooperatively bind to RNA, providing high affinity. The extra N-terminal helix (α_0) of RRM3 (Figure A.1 b) could thereby support the correct positioning of the two domains with respect to each other, enabling stable RNA interaction.¹²¹

Besides NMR, different crystallization conditions for the RRM23/UGCUUUGUUC complex were screened. Only in one of the tested conditions diffracting crystals were achieved and the structure could be solved. The electron density of the diffracting crystals contained only RRM2 bound to two uridines (Figure A.2 b). SDS-PAGE of the remaining crystal debris in the reservoir solution proved, however, that no proteolysis of the tandem construct occurred upon crystallization. This phenomenon was observed earlier by Waris et al.⁴² Similar to the RRM2/TTT crystal structure (PDB ID: 5ITH),⁴² aromatic-stacking interactions of the RNP-2 motif position 2 (Phe98) as well as base-specific hydrogen bonds between the amide nitrogen of Asn169 and U1 (Figure A.2 b upper right panel) are observed. Moreover, Asp101 and Lys136 form hydrogen bonds with the phosphate or sugar ring, respectively. Each interaction accounts for affinity rather than specificity. Due to the lack of information on the preceding and following nucleotide in the electron density and the presence of three UU repeats in the tested RNA sequence, the RRM2/UU crystal structure (PDB ID: 5O3J), however, fails to report on the

directionality of the *fas* intron 6 10 mer RNA. The absence of a unique set of residues bound by RRM2 suggests that RRM2 as well as RRM3 bind the intronic U-rich sequence in multiple registers, similar to RRM1 and RRM2 of the U2AF 65 kDa subunit.¹²² Different protein-RNA complexes are formed by either sliding (Figure 4.1 b) or dissociation and re-association at different positions on the RNA by RRM2 with respect to RRM3 (Figure 4.1 a). The different conformations are presumably in rapid exchange which could account for the absence of 8 RNA nucleotides and RRM3 in the crystals. The fast exchange between different conformations could further explain the lack of intermolecular (RNA to protein) NOE signals.

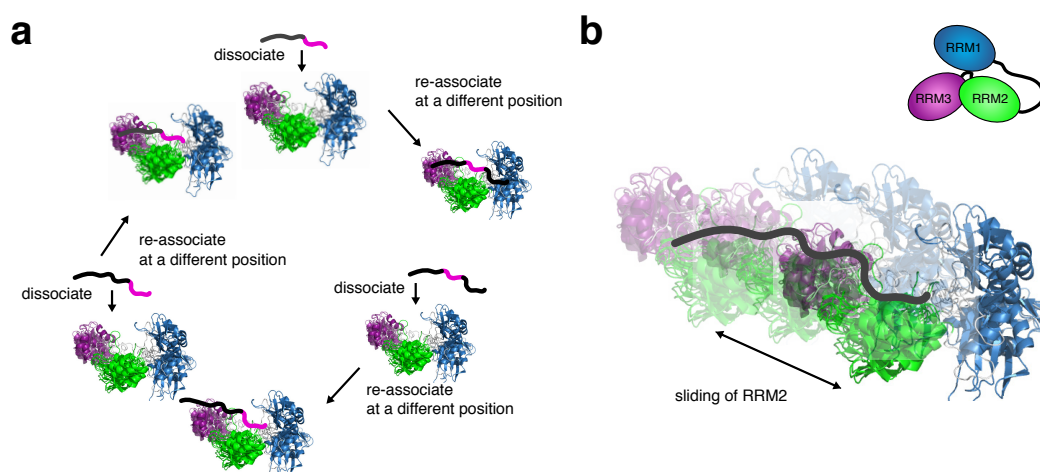


Figure 4.1.: Multiple register binding by TIA-1 RRM23. a) RRM2 dissociates from the mRNA and re-associate at a different U-rich location. b) Sliding of RRM2 on the mRNA. Both pathways create different protein-RNA complexes which are in rapid exchange. The final ensemble (top 5 structures) of the RRM123/U15 complex are shown. The ensemble was generated by combining SAXS and SANS data (see text and ref. 48 for details).

The structural information described above was used for rigid body modelling of the RRM123/U15 complex.⁴⁸ Based on the crystal structure and the fact that most RRM domains recognize 2–4 nucleotides, two nucleotides within U6–U10 were strictly restrained to bind to RRM2. All remaining nucleotides as well as the connecting linkers and the N-terminus were kept flexible. The generated pool of 5000 structures was scored against the experimental SAXS and SANS data which drastically reduced the number of structures down to five.⁴⁸ This final ensemble of five structures reveals valuable insight into the architecture of TIA-1 RRM123 bound to U15. In the presence of RNA RRM123 adopts a compact elongated L-shaped conformation in which RRM2 and RRM3 are closely packed against each other, whereas RRM1 is spatially more separated. Furthermore, RRM1 is detached from RRM2 and the RNA which supports a minor role in RNA binding rather than mediating high affinity binding of long poly U sequences.⁴⁰ However, to refine the RRM-RNA and RRM-RRM interfaces, further experimental data is required.

In conclusion, contrast-matching SANS experiments on domain-selective perdeuterated TIA-1 RRM123 provide unique insights into the molecular mechanisms involved in splicing regulation. Harboring three RRM domains and a Q-rich domain, TIA-1 facilitates splice site recognition and thus exon inclusion by recruiting U1 snRNP to the 5'ss. The central RRM2 and RRM3 domain screen the pre-mRNA downstream of alternative exons for stretches with

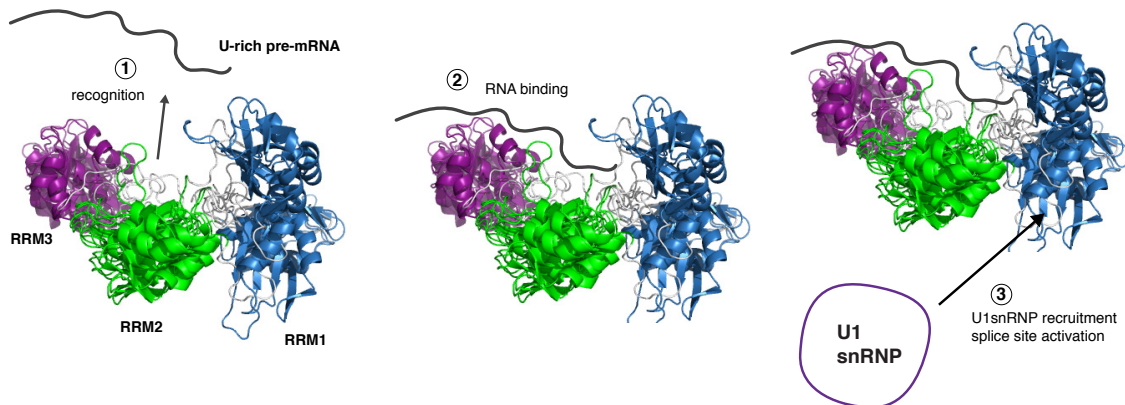


Figure 4.2.: Mode of RNA recognition and splice site activation. RRM2 and RRM3 cooperatively recognize (1) and bind (2) to U-rich pre-mRNA sequences. In contrast, RRM1 assists the Q-rich domain (not drawn here) in U1 snRNP recruitment (3) and thus ss activation. The final ensemble (top 5 structures) of the RRM123/U15 complex is shown. The ensemble was generated by combining SAXS and SANS data (see text and ref. 48 for details).

a high uridine content (Figure 4.2 first two steps) whereas RRM1 assists the Q-rich domain in direct protein-protein interaction with the U1-C subunit of U1 snRNP (Figure 4.2 third step). It still remains elusive whether RRM1 and the Q-rich domain tether U1 snRNP to assist TIA-1 binding¹²³ to 5' ss or if the cooperative binding of RRM2 and RRM3 ensures the correct positioning of RRM1 and the Q-rich domain to attract and recruit U1 snRNP to the 5' ss.

A. Appendix

This appendix shows unpublished NMR data of the tandem TIA-1 RRM23 as well as a close up look into the molecular mechanism of how TIA-1 RRM2 recognizes RNA. As TIA-1 adopts rapidly exchanging conformations even in the RNA bound form NOE-based structure calculation becomes challenging and is still ongoing. Further experiments including paramagnetic relaxation enhancement (PRE) and residual dipolar coupling (RDC) experiments are required to define the RRM-RRM and RRM-RNA interfaces and further refine the final ensemble of five structures (Figure 4.2)⁴⁸ obtained by combining contrast-matching SANS and SAXS experiments.^{124, 125} Figure A.1 presents NMR chemical shift perturbation and relaxation data of the tandem RRM23 construct in the presence of a *fas* intron 6 derived 10 mer oligonucleotide. Figure A.2 provides structural insights into how TIA-1 RRM2 recognizes U-rich RNA stretches.

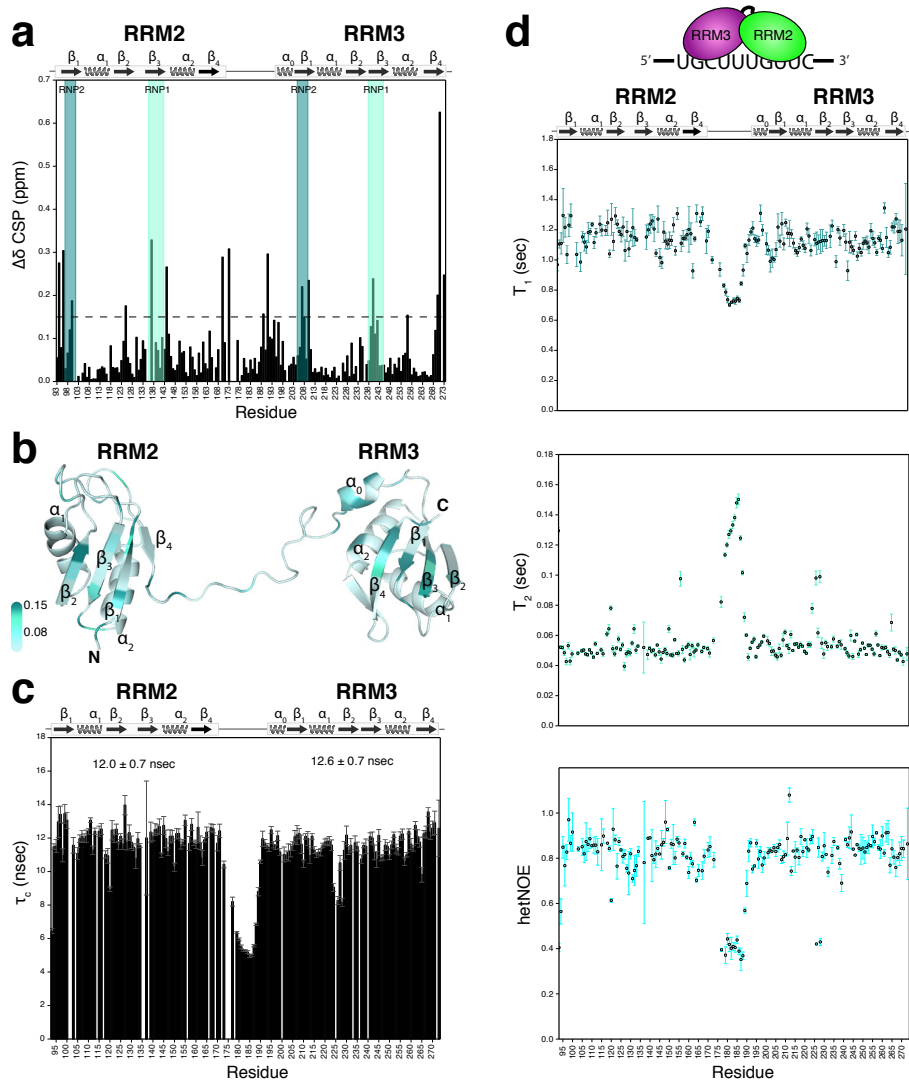


Figure A.1.: Chemical shift perturbation and relaxation data of TIA-1 RRM23 bound to UGCUUUGUUC a) ^1H and ^{15}N chemical shift perturbations upon RNA binding were calculated using the equation: $\Delta\delta = \sqrt{(\Delta\delta_H N)^2 + 0.1(\Delta\delta_N)^2}$. The dashed line in black shows a cut-off $\Delta\delta > 0.15$ (average chemical shift perturbation over all residues + standard deviation). RNP motifs are indicated by vertical boxes. Secondary structure elements are given above the plot. b) The shifts are plotted on the tandem RRM23 structure in the free form (PDB ID: 2MJN).³⁸ Note that an extra N-terminal helix (α_0) precedes the canonical RRM3 fold. c) average rotational correlation times (τ_c) for RRM23 bound to *fas* 10 mer RNA. Average correlation times for each domain are given (average τ_c values over all residues within each domain + standard deviation). d) Backbone relaxation parameters for TIA-1 RRM23 in the RNA bound form were measured at 289 K and pH 6.0. See sec. 2.5.2 for details.

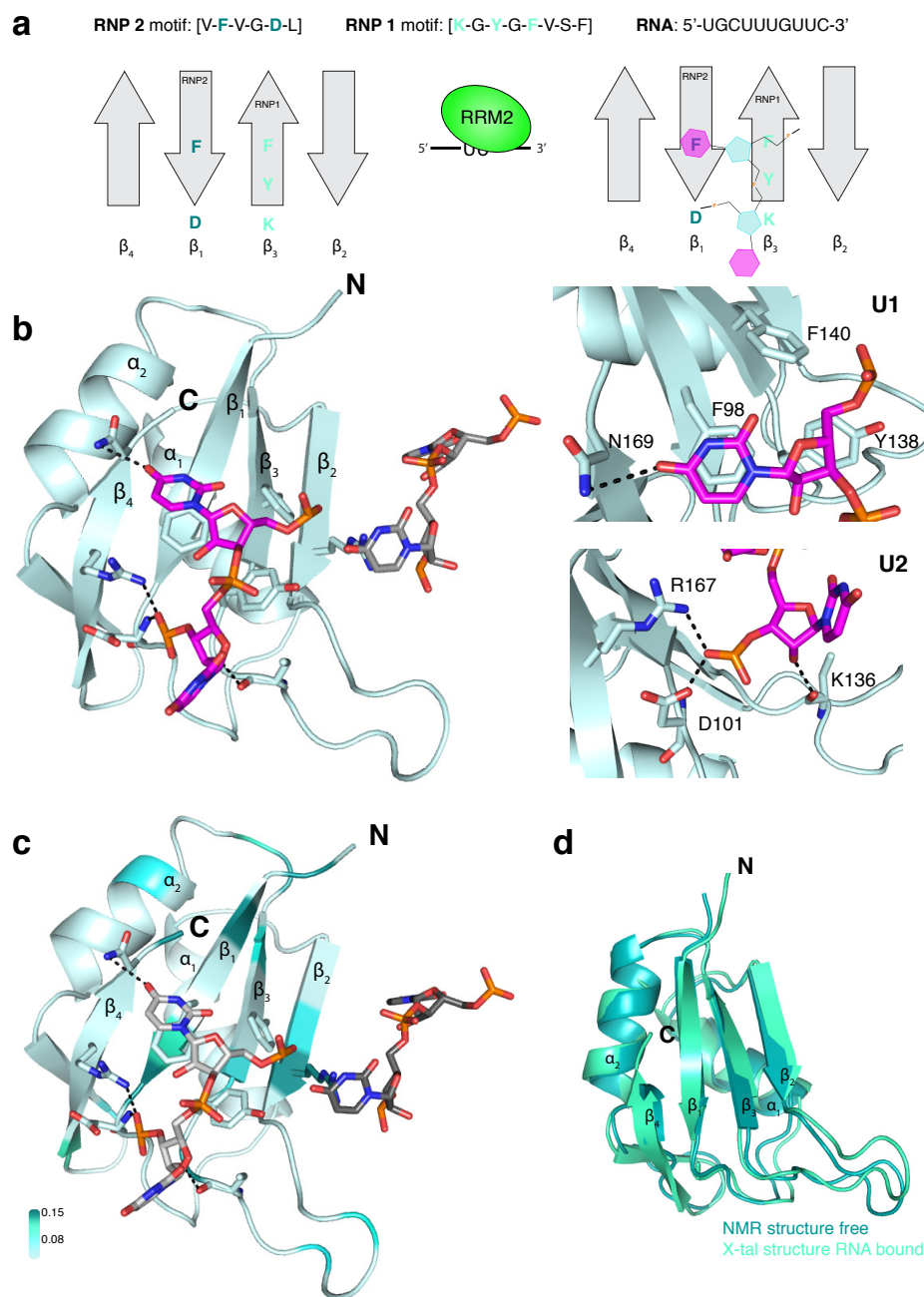


Figure A.2.: Structure of TIA-1 RRM2 bound to two uridine bases (PDB ID: 5O3J). a) Schematic representation of the β sheets and RNP motifs involved in RNA binding (left panel). The right panel schematically illustrates the position of the dinucleotide across the β -sheet surface of RRM2 b) Cartoon representation of TIA-1 RRM2. The RNA is shown as sticks, colored by atom type. The backbone and side-chain of amino acids contributing in RNA recognition are shown as sticks, and oxygens and nitrogens are colored red and blue, respectively. On the right hand side details of TIA-1 RRM2 RNA interaction are given. Hydrogen bonds are indicated by black dashed lines. c) Chemical shift perturbations upon RNA binding determined by NMR (compare with Figure A.1b) are plotted on the crystal structure with the same representation as in (b). The darker the blue colour is, the higher the amino acid shifted upon RNA binding. d) Overlay of the NMR structure of the isolated RRM2 domain (taken from the tandem TIA-1 RRM23 PDB ID: 2MJN³⁸) with the TIA-1 RRM2/UU complex.

Abbreviations

AIDS	acquired immunodeficiency syndrome	MES	1-morpholin-4-ylethanesulfonic acid
BSA	bovine serum albumin	mRNA	messenger RNA
C-terminus	carboxy-terminus	M_w	molecular weight
DNA	desoxyribonucleic acid	NCL	native chemical ligation
DTT	dithiothreitol	Ni-NTA	nickel nitrilotriacetic acid
<i>E. coli</i>	<i>Escherichia coli</i>	NMR	nuclear magnetic resonance
EPL	expressed protein ligation	N-terminus	amino-terminus
ESS	exonic splicing silencer	OD	optical density
ESE	exonic splicing enhancer	PABP	polyadenylate-binding protein
Fas	tumor necrosis factor receptor superfamily member 6	PRE	paramagnetic relaxation enhancement
FASTK	Fas-activated serine / threonine kinase	pre-mRNA	precursor messenger RNA
GB1	B1 domain of streptococcal protein G	PTB	polypyrimidine tract-binding protein 1
GPC	gel permeation chromatography	PTS	protein <i>trans</i> -splicing
hnRNPS	heterogenous nuclear RNP	Q-rich	glutamine-rich
HuD	Hu-antigen D	RALS	right angle scattering
Hsp90	heat shock protein	RBP	RNA binding protein
HSQC	heteronuclear single quantum coherence	RDC	residual dipolar coupling
IPTG	isopropyl- β -D-thiogalactopyranosid	R_g	radius of gyration
ISE	intronic splicing enhancer	RNA	ribonucleic acid
ISS	intronic splicing silencer	RNP-1	ribonucleoprotein consensus motif 1
ITC	isothermal titration calorimetry	RNP-2	ribonucleoprotein consensus motif 2
LB	lysogeny broth medium	rpm	revolutions per minute
LPXTG	Leu-Pro-any-Thr-Gly sortase recognition motif	RRM	RNA recognition motif
LS	light scattering	SAS	small angle scattering

SANS	small angle Neutron scattering	TEV	Tobacco Etch Virus
<i>S. aureus</i>	<i>Staphylococcus aureus</i>	TIA-1	T-cell-restricted intracellular antigen-1
SAXS	small angle X-ray scattering	TNF	tumor necrosis factor receptor superfamily
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis	TOCSY	total correlation spectroscopy
SLS	static light scattering	U1C	U1 snRNP specific protein C
snRNP	small nuclear ribonucleoprotein	U-rich	uridine rich
SrtA	sortase A	U1 snRNP	U1 small nuclear ribonucleoprotein
SR proteins	Ser-Arg rich proteins	UTR	untranslated region
ss	splice site		

Bibliography

- [1] T. W. Nilsen and B. R. Graveley, “Expansion of the eukaryotic proteome by alternative splicing,” Nature, vol. 463, no. 7280, pp. 457–463, 2010. (cited p. 5)
- [2] E. T. Wang, R. Sandberg, S. Luo, I. Khrebtkova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, and C. B. Burge, “Alternative isoform regulation in human tissue transcriptomes,” Nature, vol. 456, no. 7221, pp. 470–476, 2008. (cited p. 5)
- [3] Q. Pan, O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe, “Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing,” Nature Genetics, vol. 40, no. 12, pp. 1413–1415, 2008. (cited p. 5)
- [4] D. L. Black, “Mechanisms of alternative pre-messenger RNA splicing,” Annual Review of Biochemistry, vol. 72, no. 1, pp. 291–336, 2003. (cited p. 5)
- [5] M. C. Wahl, C. L. Will, and R. Lührmann, “The spliceosome: Design principles of a dynamic RNP machine,” Cell, vol. 136, no. 4, pp. 701–718, 2009. (cited p. 5)
- [6] M. Chen and J. L. Manley, “Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches,” Nature Reviews Molecular Cell Biology, vol. 10, no. 11, pp. 741–754, 2009. (cited p. 5 and 6)
- [7] U. Braunschweig, S. Gueroussov, A. M. Plocik, B. R. Graveley, and B. J. Blencowe, “Dynamic integration of splicing within gene regulatory pathways,” Cell, vol. 152, no. 6, pp. 1252–1269, 2013. (cited p. 5)
- [8] X.-D. Fu and M. Ares Jr, “Context-dependent control of alternative splicing by RNA-binding proteins,” Nature Reviews Genetics, vol. 15, no. 10, pp. 689–701, 2014. (cited p. 5)
- [9] J. M. Izquierdo, N. Majós, S. Bonnal, C. Martínez, R. Castelo, R. Guigó, D. Bilbao, and J. Valcárcel, “Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition,” Molecular Cell, vol. 19, no. 4, pp. 475–484, 2005. (cited p. 6 and 7)
- [10] T. Zhang, N. Delestienne, G. Huez, V. Kruys, and C. Gueydan, “Identification of the sequence determinants mediating the nucleo-cytoplasmic shuttling of TIAR and TIA-1 RNA-binding proteins,” Journal of Cell Science, vol. 118, no. 23, pp. 5453–5463, 2005. (cited p. 6)
- [11] M. Piecyk, S. Wax, A. R. Beck, N. Kedersha, M. Gupta, B. Maritim, S. Chen, C. Gueydan, V. Kruys, M. Streuli, and others, “TIA-1 is a translational silencer that selectively regulates the expression of TNF- α ,” The EMBO Journal, vol. 19, no. 15, pp. 4154–4163, 2000. (cited p. 6)
- [12] D. A. Dixon, G. C. Balch, N. Kedersha, P. Anderson, G. A. Zimmerman, R. D. Beauchamp, and S. M. Prescott, “Regulation of cyclooxygenase-2 expression by the translational silencer TIA-1,” Journal of Experimental Medicine, vol. 198, no. 3, pp. 475–481, 2003. (cited p. 6)

- [13] C. K. Damgaard and J. Lykke-Andersen, "Translational coregulation of 5' TOP mRNAs by TIA-1 and TIAR," Genes & Development, vol. 25, no. 19, pp. 2057–2068, 2011. (cited p. 6)
- [14] I. L. De Silanes, S. Galbán, J. L. Martindale, X. Yang, K. Mazan-Mamczarz, F. E. Indig, G. Falco, M. Zhan, and M. Gorospe, "Identification and functional outcome of mRNAs associated with RNA-binding protein TIA-1," Molecular and Cellular Biology, vol. 25, no. 21, pp. 9520–9531, 2005. (cited p. 6)
- [15] T. Kawai, A. Lal, X. Yang, S. Galban, K. Mazan-Mamczarz, and M. Gorospe, "Translational control of cytochrome c by RNA-binding proteins TIA-1 and HuR," Molecular and Cellular Biology, vol. 26, no. 8, pp. 3295–3307, 2006. (cited p. 6)
- [16] S. Waris, M. C. J. Wilce, and J. A. Wilce, "RNA recognition and stress granule formation by TIA proteins," International Journal of Molecular Sciences, vol. 15, no. 12, pp. 23377–23388, 2014. (cited p. 6)
- [17] N. L. Kedersha, M. Gupta, W. Li, I. Miller, and P. Anderson, "RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 α to the assembly of mammalian stress granules," The Journal of Cell Biology, vol. 147, no. 7, pp. 1431–1442, 1999. (cited p. 6)
- [18] P. Anderson and N. Kedersha, "Stressful initiations," Journal of Cell Science, vol. 115, no. 16, pp. 3227–3234, 2002. (cited p. 6)
- [19] F. Del Gatto-Konczak, C. F. Bourgeois, C. Le Guiner, L. Kister, M.-C. Gesnel, J. Stévenin, and R. Breathnach, "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, vol. 20, no. 17, pp. 6287–6299, 2000. (cited p. 7, 8, and 32)
- [20] E. Zuccato, E. Buratti, C. Stuani, F. E. Baralle, and F. Pagani, "An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing," Journal of Biological Chemistry, vol. 279, no. 17, pp. 16980–16988, 2004. (cited p. 7)
- [21] P. Förch, O. Puig, C. Martínez, B. Séraphin, and J. Valcárcel, "The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites," The EMBO Journal, vol. 21, pp. 6882–92, Dec. 2002. (cited p. 7, 8, 31, and 32)
- [22] A. McAlinden, L. Liang, Y. Mukudai, T. Imamura, and L. J. Sandell, "Nuclear protein TIA-1 regulates COL2A1 alternative splicing and interacts with precursor mRNA and genomic DNA," Journal of Biological Chemistry, vol. 282, no. 33, pp. 24444–24454, 2007. (cited p. 7)
- [23] H. Zhu, R. A. Hasman, K. M. Young, N. L. Kedersha, and H. Lou, "U1 snRNP-dependent function of TIAR in the regulation of alternative RNA processing of the human calcitonin/CGRP pre-mRNA," Molecular and Cellular Biology, vol. 23, no. 17, pp. 5959–5971, 2003. (cited p. 7)

- [24] P. Förch, O. Puig, N. Kedersha, C. Martínez, S. Granneman, B. Séraphin, P. Anderson, and J. Valcárcel, “The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing,” *Molecular Cell*, vol. 6, no. 5, pp. 1089–1098, 2000. (cited p. 7 and 8)
- [25] P. Förch and J. Valcárcel, “Molecular mechanisms of gene expression regulation by the apoptosis-promoting protein TIA-1,” *Apoptosis*, vol. 6, no. 6, pp. 463–468, 2001. (cited p. 7)
- [26] P. H. Krammer, “CD95’s deadly mission in the immune system,” *Nature*, vol. 407, pp. 789–95, Oct. 2000. (cited p. 7)
- [27] J. M. Adams, “Ways of dying: Multiple pathways to apoptosis,” *Genes & Development*, vol. 17, no. 20, pp. 2481–2495, 2003. (cited p. 7)
- [28] C. Schwerk and K. Schulze-Osthoff, “Regulation of apoptosis by alternative pre-mRNA splicing,” *Molecular Cell*, vol. 19, no. 1, pp. 1–13, 2005. (cited p. 7)
- [29] A. Strasser, P. J. Jost, and S. Nagata, “The many roles of FAS receptor signaling in the immune system,” *Immunity*, vol. 30, no. 2, pp. 180–192, 2009. (cited p. 7)
- [30] T. Kaufmann, A. Strasser, and P. J. Jost, “Fas death receptor signalling: Roles of Bid and XIAP,” *Cell Death & Differentiation*, vol. 19, no. 1, pp. 42–50, 2012. (cited p. 7)
- [31] J. Cheng, T. Zhou, C. Liu, J. P. Shapiro, M. J. Brauer, M. C. Kiefer, P. J. Barr, and J. D. Mountz, “Protection from Fas-mediated apoptosis by a soluble form of the Fas molecule,” *Science*, vol. 263, no. 5154, pp. 1759–1763, 1994. (cited p. 7)
- [32] I. Aznarez, Y. Barash, O. Shai, D. He, J. Zielenski, L.-C. Tsui, J. Parkinson, B. J. Frey, J. M. Rommens, and B. J. Blencowe, “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 Research*, vol. 18, pp. 1247–58, Aug. 2008. (cited p. 8, 31, and 32)
- [33] C. Le Guiner, F. Lejeune, D. Galiana, L. Kister, R. Breathnach, J. Stévenin, and F. Del Gatto-Konczak, “TIA-1 and TIAR activate splicing of alternative exons with weak 5’ splice sites followed by a U-rich stretch on their own pre-mRNAs,” *Journal of Biological Chemistry*, vol. 276, pp. 40638–46, Nov. 2001. (cited p. 8, 31, and 32)
- [34] L. M. Dember, N. D. Kim, K. Q. Liu, and P. Anderson, “Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities,” *Journal of Biological Chemistry*, vol. 271, pp. 2783–8, Feb. 1996. (cited p. 8, 31, and 32)
- [35] Z. Wang, M. Kayikci, M. Briese, K. Zarnack, N. M. Luscombe, G. Rot, B. Zupan, T. Curk, and J. Ule, “iCLIP predicts the dual splicing effects of TIA-RNA interactions,” *PLoS Biology*, vol. 8, p. e1000530, Oct. 2010. (cited p. 8, 31, and 32)
- [36] J. M. Izquierdo and J. Valcárcel, “Fas-activated serine/threonine kinase (FAST K) synergizes with TIA-1/TIAR proteins to regulate Fas alternative splicing,” *Journal of Biological Chemistry*, vol. 282, no. 3, pp. 1539–1543, 2007. (cited p. 8)

- [37] Q. Tian, J. Taupin, S. Elledge, M. Robertson, and P. Anderson, "Fas-activated serine/threonine kinase (FAST) phosphorylates TIA-1 during Fas-mediated apoptosis.," Journal of Experimental Medicine, vol. 182, no. 3, pp. 865–874, 1995. (cited p. 8)
- [38] I. Wang, J. Hennig, P. K. A. Jagtap, M. Sonntag, J. Valcárcel, and M. Sattler, "Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1," Nucleic Acids Research, vol. 42, pp. 5949–66, May 2014. (cited p. 8, 9, 12, 28, 31, 32, 36, and 37)
- [39] I. Cruz-Gallardo, Á. Aroca, M. J. Gunzburg, A. Sivakumaran, J.-H. Yoon, J. Angulo, C. Persson, M. Gorospe, B. G. Karlsson, J. A. Wilce, and I. Díaz-Moreno, "The binding of TIA-1 to RNA C-rich sequences is driven by its C-terminal RRM domain," RNA Biology, vol. 11, no. 6, pp. 766–76, 2014. (cited p. 8 and 32)
- [40] W. J. Bauer, J. Heath, J. L. Jenkins, and C. L. Kielkopf, "Three RNA recognition motifs participate in RNA recognition and structural organization by the pro-apoptotic factor TIA-1," Journal of Molecular Biology, vol. 415, pp. 727–40, Jan. 2012. (cited p. 8, 9, 31, and 33)
- [41] H. S. Kim, S. J. Headey, Y. M. K. Yoga, M. J. Scanlon, M. Gorospe, M. C. J. Wilce, and J. A. Wilce, "Distinct binding properties of TIAR RRMs and linker region," RNA Biology, vol. 10, pp. 579–89, Apr. 2013. (cited p. 8 and 32)
- [42] S. Waris, S. M. García-Mauriño, A. Sivakumaran, S. A. Beckham, F. E. Loughlin, M. Gorospe, I. Díaz-Moreno, M. C. J. Wilce, and J. A. Wilce, "TIA-1 RRM23 binding and recognition of target oligonucleotides," Nucleic Acids Research, vol. 45, pp. 4944–4957, May 2017. (cited p. 8, 9, and 32)
- [43] D. D. Boehr, R. Nussinov, and P. E. Wright, "The role of dynamic conformational ensembles in biomolecular recognition," Nature Chemical Biology, vol. 5, no. 11, pp. 789–796, 2009. (cited p. 9)
- [44] P. Bernadó and M. Blackledge, "Structural biology: Proteins in dynamic equilibrium," Nature, vol. 468, no. 7327, pp. 1046–1048, 2010. (cited p. 9 and 28)
- [45] E. Delaforge, S. Milles, J.-r. Huang, D. Bouvier, M. R. Jensen, M. Sattler, D. J. Hart, and M. Blackledge, "Investigating the Role of Large-Scale Domain Dynamics in Protein-Protein Interactions," Frontiers in Molecular Biosciences, vol. 3, p. 54, 2016. (cited p. 9 and 28)
- [46] C. D. Mackereth, T. Madl, S. Bonnal, B. Simon, K. Zanier, A. Gasch, V. Rybin, J. Valcárcel, and M. Sattler, "Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF," Nature, vol. 475, no. 7356, pp. 408–411, 2011. (cited p. 9 and 28)
- [47] G. Zaccai and B. Jacrot, "Small angle neutron scattering," Annual Review of Biophysics and Bioengineering, vol. 12, no. 1, pp. 139–157, 1983. (cited p. 9, 17, 28, and 31)
- [48] M. Sonntag, P. Jagtap, B. Simon, M.-S. Appavou, A. Geerlof, R. Stehle, F. Gabel, J. Hennig, and M. Sattler, "Segmental, domain-selective perdeuteration and small angle

- neutron scattering for structural analysis of multi-domain proteins,” Angewandte Chemie International Edition, June 2017. (cited p. 9, 18, 20, 28, 31, 32, 33, 34, and 35)
- [49] H. Mao, S. A. Hart, A. Schink, and B. A. Pollok, “Sortase-mediated protein ligation: A new method for protein engineering,” Journal of the American Chemical Society, vol. 126, pp. 2670–1, Mar. 2004. (cited p. 9, 14, 27, and 31)
- [50] L. Freiburger, M. Sonntag, J. Hennig, J. Li, P. Zou, and M. Sattler, “Efficient segmental isotope labeling of multi-domain proteins using Sortase A,” Journal of Biomolecular NMR, vol. 63, pp. 1–8, Sept. 2015. (cited p. 9, 11, 15, 27, 28, and 31)
- [51] Á. Aroca, A. Díaz-Quintana, and I. Díaz-Moreno, “A structural insight into the C-terminal RNA recognition motifs of T-cell intracellular antigen-1 protein,” FEBS letters, vol. 585, no. 19, pp. 2958–2964, 2011. (cited p. 12)
- [52] S. K. Mazmanian, H. Ton-That, and O. Schneewind, “Sortase-catalysed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*,” Molecular Microbiology, vol. 40, pp. 1049–1057, June 2001. (cited p. 12, 14, and 27)
- [53] I. Chen, B. M. Dorr, and D. R. Liu, “A general strategy for the evolution of bond-forming enzymes using yeast display,” Proceedings of the National Academy of Sciences, vol. 108, no. 28, pp. 11399–11404, 2011. (cited p. 13)
- [54] L. Skrisovska, M. Schubert, and F. H.-T. Allain, “Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins,” Journal of Biomolecular NMR, vol. 46, pp. 51–65, Jan. 2010. (cited p. 11)
- [55] D. Cowburn, A. Shekhtman, R. Xu, J. J. Ottesen, and T. W. Muir, “Segmental isotopic labeling for structural biological applications of NMR,” Protein NMR Techniques, pp. 47–56, 2004. (cited p. 14)
- [56] M. Muona, A. S. Aranko, V. Raulinaitis, and H. Iwai, “Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro,” Nature Protocols, vol. 5, no. 3, pp. 574–587, 2010. (cited p. 14)
- [57] R. Xu, B. Ayers, D. Cowburn, and T. W. Muir, “Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies,” Proceedings of the National Academy of Sciences, vol. 96, no. 2, pp. 388–393, 1999. (cited p. 14)
- [58] S. Züger and H. Iwai, “Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies,” Nature Biotechnology, vol. 23, no. 6, pp. 736–740, 2005. (cited p. 14)
- [59] J. M. Antos, M. C. Truttman, and H. L. Ploegh, “Recent advances in sortase-catalyzed ligation methodology,” Current Opinion in Structural Biology, vol. 38, pp. 111–8, June 2016. (cited p. 14, 27, and 31)
- [60] T. Proft, “Sortase-mediated protein ligation: An emerging biotechnology tool for protein modification and immobilisation,” Biotechnology Letters, vol. 32, pp. 1–10, Jan. 2010. (cited p. 14, 27, and 31)

- [61] X. Huang, A. Aulabaugh, W. Ding, B. Kapoor, L. Alksne, K. Tabei, and G. Ellestad, “Kinetic mechanism of *Staphylococcus aureus* sortase SrtA,” Biochemistry, vol. 42, pp. 11307–15, Sept. 2003. (cited p. 14)
- [62] D. I. Svergun and M. H. J. Koch, “Advances in structure analysis using small-angle scattering in solution,” Current Opinion in Structural Biology, vol. 12, pp. 654–60, Oct. 2002. (cited p. 15 and 16)
- [63] M. H. Koch, P. Vachette, and D. I. Svergun, “Small-angle scattering: A view on the properties, structures and structural changes of biological macromolecules in solution,” Quarterly Reviews of Biophysics, vol. 36, pp. 147–227, May 2003. (cited p. 15, 16, and 17)
- [64] Y. S. Heimo Schnablegger, The SAXS Guide. Anton Paar GmbH, Austria, 2013. (cited p. 15 and 17)
- [65] J. Hennig, I. Wang, M. Sonntag, F. Gabel, and M. Sattler, “Combining NMR and small angle X-ray and neutron scattering in the structural analysis of a ternary protein-RNA complex,” Journal of Biomolecular NMR, vol. 56, pp. 17–30, May 2013. (cited p. 16 and 28)
- [66] A. Grishaev, J. Wu, J. Trehwella, and A. Bax, “Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data,” Journal of the American Chemical Society, vol. 127, no. 47, pp. 16621–16628, 2005. (cited p. 16)
- [67] T. Madl, F. Gabel, and M. Sattler, “NMR and small-angle scattering-based structural analysis of protein complexes in solution,” Journal of Structural Biology, vol. 173, no. 3, pp. 472–482, 2011. (cited p. 16 and 28)
- [68] A. Whitten and J. Trehwella, “Small-Angle Scattering and Neutron Contrast Variation for Studying Bio-Molecular Complexes,” Methods in Molecular Biology, vol. 544, pp. 307–323, 2009. (cited p. 17)
- [69] B. Jacrot, “The study of biological structures by neutron scattering from solution,” Reports on Progress in Physics, vol. 39, no. 10, p. 911, 1976. (cited p. 17 and 31)
- [70] “BM29 BioSAXS.” http://www.esrf.eu/cms/live/live/en/sites/www/home/UsersAndScience/Experiments/MX/About_our_beamlines/bm29.html. (cited p. 18)
- [71] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. Mertens, P. V. Konarev, and D. I. Svergun, “New developments in the ATSAS program package for small-angle scattering data analysis,” Journal of applied crystallography, vol. 45, no. 2, pp. 342–350, 2012. (cited p. 18)
- [72] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. Koch, and D. I. Svergun, “PRIMUS: A Windows PC-based system for small-angle scattering data analysis,” Journal of applied crystallography, vol. 36, no. 5, pp. 1277–1282, 2003. (cited p. 18)

- [73] D. Svergun, "Determination of the regularization parameter in indirect-transform methods using perceptual criteria," Journal of applied crystallography, vol. 25, no. 4, pp. 495–503, 1992. (cited p. 18)
- [74] H. M.-L. Zentrum, "KWS-1: Small-angle scattering diffractometer," Journal of large-scale research facilities, vol. 1, p. A28, 2015. (cited p. 18 and 28)
- [75] "D22 - Large dynamic range small-angle diffractometer." <https://www.i11.eu/instruments-support/instruments-groups/instruments/d22/characteristics/?L=0>. (cited p. 18)
- [76] C. Maris, C. Dominguez, and F. H.-T. Allain, "The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression," FEBS Journal, vol. 272, pp. 2118–31, May 2005. (cited p. 18, 31, and 32)
- [77] T. Afroz, Z. Cienikova, A. Cléry, and F. H. T. Allain, "One, Two, Three, Four! How Multiple RRM's Read the Genome Sequence," Methods in Enzymology, vol. 558, pp. 235–78, 2015. (cited p. 18, 31, and 32)
- [78] R. M. Murphy, "Static and dynamic light scattering of biological macromolecules: What can we learn?," Current Opinion in Biotechnology, vol. 8, pp. 25–30, Feb. 1997. (cited p. 19)
- [79] U. Nobbmann, M. Connah, B. Fish, P. Varley, C. Gee, S. Mulot, J. Chen, L. Zhou, Y. Lu, F. Shen, J. Yi, and S. E. Harding, "Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies," Biotechnology & Genetic Engineering Reviews, vol. 24, pp. 117–128, 2007. (cited p. 19)
- [80] J. Mogridge, "Using light scattering to determine the stoichiometry of protein complexes," Methods in Molecular Biology, vol. 261, pp. 113–118, 2004. (cited p. 19)
- [81] P. J. Wyatt, "Light scattering and the absolute characterization of macromolecules," Analytica Chimica Acta, vol. 272, no. 1, pp. 1–40, 1993. (cited p. 19)
- [82] N. C. Santos and M. A. Castanho, "Teaching light scattering spectroscopy: The dimension and shape of tobacco mosaic virus," Biophysical Journal, vol. 71, pp. 1641–1650, Sept. 1996. (cited p. 19)
- [83] B. H. Zimm, "Apparatus and Methods for Measurement and Interpretation of the Angular Variation of Light Scattering; Preliminary Results on Polystyrene Solutions," The Journal of Chemical Physics, vol. 16, pp. 1099–1116, Dec. 1948. (cited p. 19)
- [84] M. I. Recht, S. P. Ryder, and J. R. Williamson, "Monitoring assembly of ribonucleoprotein complexes by isothermal titration calorimetry," Methods in Molecular Biology, vol. 488, pp. 117–127, 2008. (cited p. 20)
- [85] C. Crane-Robinson, A. I. Dragan, and C. M. Read, "Defining the thermodynamics of protein/DNA complexes and their components using micro-calorimetry," Methods in Molecular Biology, vol. 543, pp. 625–651, 2009. (cited p. 20)

- [86] J. Keeler, Understanding NMR Spectroscopy. John Wiley & Sons, 2010. (cited p. 21, 22, and 23)
- [87] P. Hore, Nuclear Magnetic Resonance. Oxford University Press, USA, 2011. (cited p. 21, 22, and 23)
- [88] M. Sattler, J. Schleucher, and C. Griesinger, “Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution,” Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 34, pp. 93–158, 1999. (cited p. 23)
- [89] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, “NMRPipe: A multidimensional spectral processing system based on UNIX pipes,” Journal of Biomolecular NMR, vol. 6, no. 3, pp. 277–293, 1995. (cited p. 23)
- [90] R. Keller, The Computer Aided Resonance Assignment Tutorial. Cantina Verlag, 2004. (cited p. 23)
- [91] A. Ahlner, M. Carlsson, B.-H. Jonsson, and P. Lundström, “PINT: A software for integration of peak volumes and extraction of relaxation rates,” Journal of Biomolecular NMR, vol. 56, pp. 191–202, July 2013. (cited p. 26)
- [92] L. E. Kay, D. A. Torchia, and A. Bax, “Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease,” Biochemistry, vol. 28, pp. 8972–9, Nov. 1989. (cited p. 26)
- [93] G. Rhodes, Crystallography Made Crystal Clear. Academic Press, Elsevier Inc., 2006. (cited p. 26)
- [94] “ID29 MX tunable microbeam.” <http://www.esrf.eu/id29>. (cited p. 26)
- [95] W. Kabsch, “Xds,” Acta Crystallographica Section D: Biological Crystallography, vol. 66, no. 2, pp. 125–132, 2010. (cited p. 26)
- [96] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, “Phaser crystallographic software,” Journal of Applied Crystallography, vol. 40, no. 4, pp. 658–674, 2007. (cited p. 26)
- [97] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, and others, “Overview of the CCP4 suite and current developments,” Acta Crystallographica Section D, vol. 67, no. 4, pp. 235–242, 2011. (cited p. 26)
- [98] P. Emsley and K. Cowtan, “Coot: Model-building tools for molecular graphics,” Acta Crystallographica Section D, vol. 60, no. 12, pp. 2126–2132, 2004. (cited p. 26)
- [99] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, “Refinement of macromolecular structures by the maximum-likelihood method,” Acta Crystallographica Section D, vol. 53, no. 3, pp. 240–255, 1997. (cited p. 26)

- [100] F. Clow, J. D. Fraser, and T. Proft, "Immobilization of proteins to biacore sensor chips using *Staphylococcus aureus* sortase A," *Biotechnology Letters*, vol. 30, no. 9, pp. 1603–1607, 2008. (cited p. 27)
- [101] C. P. Guimaraes, M. D. Witte, C. S. Theile, G. Bozkurt, L. Kundrat, A. E. Blom, and H. L. Ploegh, "Site-specific C-terminal internal loop labeling of proteins using sortase-mediated reactions," *Nature Protocols*, vol. 8, no. 9, p. 1787, 2013. (cited p. 27)
- [102] Y. Kobashigawa, H. Kumeta, K. Ogura, and F. Inagaki, "Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method," *Journal of Biomolecular NMR*, vol. 43, no. 3, p. 145, 2009. (cited p. 27)
- [103] D. A. Levary, R. Parthasarathy, E. T. Boder, and M. E. Ackerman, "Protein-protein fusion catalyzed by sortase A," *PloS One*, vol. 6, no. 4, p. e18342, 2011. (cited p. 27)
- [104] S. K. Mazmanian, G. Liu, H. Ton-That, and O. Schneewind, "*Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall," *Science*, vol. 285, no. 5428, pp. 760–763, 1999. (cited p. 27)
- [105] M. A. Refaei, A. Combs, D. J. Kojetin, J. Cavanagh, C. Caperelli, M. Rance, J. Sapiro, and P. Tsang, "Observing selected domains in multi-domain proteins via sortase-mediated ligation and NMR spectroscopy," *Journal of Biomolecular NMR*, vol. 49, no. 1, pp. 3–7, 2011. (cited p. 27)
- [106] B. Ma, C.-J. Tsai, T. Haliloğlu, and R. Nussinov, "Dynamic allostery: Linkers are not merely flexible," *Structure*, vol. 19, no. 7, pp. 907–917, 2011. (cited p. 28)
- [107] Y. Shamoo, N. Abdul-Manan, and K. R. Williams, "Multiple RNA binding domains (RBDs) just don't add up," *Nucleic Acids Research*, vol. 23, no. 5, pp. 725–728, 1995. (cited p. 28)
- [108] B. Antonsson, R. Leberman, B. Jacrot, and G. Zaccai, "Small-angle neutron scattering study of the ternary complex formed between bacterial elongation factor Tu, guanosine 5'-triphosphate, and valyl-tRNA^{Val}," *Biochemistry*, vol. 25, no. 12, pp. 3655–3659, 1986. (cited p. 28)
- [109] M. Falb, I. Amata, F. Gabel, B. Simon, and T. Carlomagno, "Structure of the K-turn U4 RNA: A combined NMR and SANS study," *Nucleic Acids Research*, p. gkq380, 2010. (cited p. 28)
- [110] J. Hennig, C. Militti, G. M. Popowicz, I. Wang, M. Sonntag, A. Geerlof, F. Gabel, F. Gebauer, and M. Sattler, "Structural basis for the assembly of the Sxl-Unr translation regulatory complex," *Nature*, vol. 515, pp. 287–90, Nov. 2014. (cited p. 28)
- [111] A. Lapinaite, B. Simon, L. Skjaerven, M. Rakwalska-Bange, F. Gabel, and T. Carlomagno, "The structure of the box C/D enzyme reveals regulation of RNA methylation," *Nature*, vol. 502, no. 7472, pp. 519–523, 2013. (cited p. 28)

- [112] A. G. Kikhney and D. I. Svergun, “A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins,” FEBS Letters, vol. 589, no. 19PartA, pp. 2570–2577, 2015. (cited p. 28)
- [113] D. A. Jacques and J. Trewhella, “Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls,” Protein Science, vol. 19, no. 4, pp. 642–657, 2010. (cited p. 28 and 31)
- [114] T. Carlomagno, “Present and future of NMR for RNA–protein complexes: A perspective of integrated structural biology,” Journal of Magnetic Resonance, vol. 241, pp. 126–136, 2014. (cited p. 28)
- [115] B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof, and D. R. Liu, “Reprogramming the specificity of sortase enzymes,” Proceedings of the National Academy of Sciences, vol. 111, no. 37, pp. 13343–13348, 2014. (cited p. 31)
- [116] N. Handa, O. Nureki, K. Kurimoto, I. Kim, H. Sakamoto, Y. Shimura, Y. Muto, and S. Yokoyama, “Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein,” Nature, vol. 398, pp. 579–85, Apr. 1999. (cited p. 32)
- [117] C. Johansson, L. D. Finger, L. Trantirek, T. D. Mueller, S. Kim, I. A. Laird-Offringa, and J. Feigon, “Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target,” Journal of Molecular Biology, vol. 337, pp. 799–816, Apr. 2004. (cited p. 32)
- [118] F. H. Allain, P. Bouvet, T. Dieckmann, and J. Feigon, “Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin,” The EMBO Journal, vol. 19, pp. 6870–81, Dec. 2000. (cited p. 32)
- [119] R. C. Deo, J. B. Bonanno, N. Sonenberg, and S. K. Burley, “Recognition of polyadenylate RNA by the poly(A)-binding protein,” Cell, vol. 98, pp. 835–45, Sept. 1999. (cited p. 32)
- [120] X. Wang and T. M. Tanaka Hall, “Structural basis for recognition of AU-rich element RNA by the HuD protein,” Nature Structural Biology, vol. 8, pp. 141–5, Feb. 2001. (cited p. 32)
- [121] C. Netter, G. Weber, H. Benecke, and M. C. Wahl, “Functional stabilization of an RNA recognition motif by a noncanonical N-terminal expansion,” RNA, vol. 15, pp. 1305–13, July 2009. (cited p. 32)
- [122] H. Banerjee, A. Rahn, W. Davis, and R. Singh, “Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding,” RNA, vol. 9, pp. 88–99, Jan. 2003. (cited p. 33)
- [123] M.-C. Gesnel, S. Theoleyre, F. Del Gatto-Konczak, and R. Breathnach, “Cooperative binding of TIA-1 and U1 snRNP in K-SAM exon splicing activation,” Biochemical and Biophysical Research Communications, vol. 358, pp. 1065–70, July 2007. (cited p. 34)

-
- [124] C. Göbl, T. Madl, B. Simon, and M. Sattler, “NMR approaches for structural analysis of multidomain proteins and complexes in solution,” Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 80, pp. 26–63, 2014. (cited p. 35)
- [125] J. Hennig and M. Sattler, “The dynamic duo: Combining NMR and small angle scattering in structural biology,” Protein Science, vol. 23, no. 6, pp. 669–682, 2014. (cited p. 35)