

TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie,
Lehrstuhl für biomolekulare NMR-Spektroskopie

**Structural studies on the Cytoplasmic
Polyadenylation Element Binding (CPEB) protein
family**

Constanze Schelhorn

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

Vorsitzender: Univ.-Prof. Dr. Bernd Reif

Prüfer der Dissertation: 1. Univ.-Prof. Dr. Michael Sattler
2. Priv.-Doz. Dr. Dierk Niessing
Ludwig-Maximilians-Universität München

Die Dissertation wurde am 26.11.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 17.12.2014 angenommen.

TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie,
Lehrstuhl für biomolekulare NMR-Spektroskopie

Structural studies on the Cytoplasmic
Polyadenylation Element Binding (CPEB) protein
family

This work has been carried out at the **Institute for Research in Biomedicine (IRB Barcelona)**, Structural & Computational Biology Programme in the Laboratory of *Structural characterization of macromolecular assemblies* under the supervision of **Dr. Maria J. Macias** from September 2010 to October 2014.

Constanze Schelhorn, 2014



“Your assumptions are your windows on the world. Scrub them off every once in a while, or the light won’t come in.”
– Isaac Asimov

Table of Contents

1	Introduction	1
1.1	CPEB family: Cytoplasmic polyadenylation and translational control . . .	1
1.2	Peptidyl-prolyl isomerase Pin1	8
2	Aims of the thesis	13
3	Material and Methods	17
3.1	Cloning of a target protein	17
3.2	Fmoc Solid-Phase Peptide Synthesis - Basic principles	19
3.3	Biophysical methods	23
3.3.1	Ion Mobility -Mass Spectrometry	23
3.3.2	Isothermal Titration Calorimetry	25
3.4	X-ray Crystallography	27
3.4.1	Protein crystallization	27
3.4.2	Data collection	27
3.4.3	Data analysis	28
3.5	NMR Spectroscopy in Structural Biology	31
3.5.1	Basic principles of NMR	31
3.5.2	Chemical Shift and J-coupling	34
3.5.3	Two Dimensional NMR Experiments	42
3.5.4	Three Dimensional NMR Experiments	45
3.5.5	Ligand binding studies by NMR	47
3.5.6	Relaxation Studies in NMR	48
3.5.7	Structure calculation and quality control	49
3.6	Experimental Procedures	51
3.6.1	Cloning of CPEB and Pin1 constructs	51
3.6.2	Protein Expression and Purification	51
3.6.3	Manual SPPS of phosphorylated and unphosphorylated peptides	53
3.6.4	Peptide purification and analysis	55
3.6.5	Ion Mobility - Mass Spectrometry	55
3.6.6	ITC	56
3.6.7	NMR Chemical Shift Perturbation Experiments	56
3.6.8	NMR Assignments	57
3.6.9	Relaxation Measurements	57
3.7	Buffers and solutions	59

4	Results	63
4.1	RNA Recognition and self-association of CPEB4 by its tandem RRM domains	63
4.1.1	Interactions of the RRM domains in the absence of RNA	63
4.1.2	RRM domains are tumbling as a single unit	67
4.1.3	CPEB4 uses both RRM domains to bind CPE motifs	70
4.1.4	RRM1 and RRM1-2 have the ability to dimerize	73
4.1.5	Dimerization of RRM domains does not disrupt RNA binding	75
4.1.6	Identifying the RNA binding site of RRM1-RRM2 via NMR	78
4.1.7	Crystallization Trials CPEB4 RRM1-RRM2	80
4.2	Pin1 a key factor in CPEB1 degradation	83
4.2.1	Pin1 WW domain interacts with CPEB prior to phosphorylation	83
4.2.2	Phosphorylation enhances binding affinity of Pin1 to CPEB1	87
4.2.3	Identifying the binding site for CPEB1 phosphopeptides within Pin1	91
4.2.4	Phosphate group directly involved in binding	95
4.2.5	Dynamics of Pin1 upon binding of CPEB1 pS210	97
4.2.6	Structure of Pin1 WW CPEB1 pS210 complex	99
5	Discussion	105
5.1	RNA recognition and self-association of CPEB4 is mediated by its tandem RRM domains	105
5.2	Pin1 mediated CPEB destruction - a two-step interaction	111
6	Summary	117
7	Appendix	121
7.1	CPEB4 RRM domains	121
7.2	Glossary	127
7.3	Publication	129
	Curriculum Vitae	140
	Bibliography	145
	Acknowledgements	154

1 Introduction

1.1 CPEB family: Cytoplasmic polyadenylation and translational control

Regulation of gene expression

Gene expression is the process by which the nucleotide sequence of a gene is used to synthesize a functional gene product. These products are proteins or functional RNAs. In eukaryotic cells, the DNA containing the genetic information is located in the cell nucleus, where messenger RNA (mRNA) is transcribed from it. Subsequently the mRNA is transported to the cytoplasm where proteins are synthesized. The mRNA is a single-stranded copy of the gene transcribed and needs to be translated into a protein molecule.

Almost every step of gene expression from transcriptional initiation to post-translational modification of a protein is subjected to regulation. At the stage of translation, an important control mechanism are cis-acting elements within the mRNA, which are frequently located in a portion of mRNA referred to as untranslated region (UTR). The UTR is located before the start of the coding region and does not affect the protein sequence produced from the mRNA. These cis-regulatory elements serve as binding sites for trans-acting factors, e.g. RNA binding proteins. A combination of one or more cis-acting elements and of one or more trans-acting factors provide mechanisms allowing for an effective regulation of translation in a spatial, temporal or context-specific manner [1].

Translational Control

In early development many maternal mRNAs are stored translationally repressed in oocytes until their activation during the late stage of meiosis or after fertilization [2], [3]. This group of transcripts is stored with a short poly(A)-tail and is activated by cytoplasmic polyadenylation as a response to hormonal stimulation by progesterone. Translational control by cytoplasmic elongation of the poly(A) tails was firstly described as a key mechanism that regulates gene expression during meiotic progression in transcriptionally silent *Xenopus* oocytes. Basically the translation of a mRNA is

divided into three steps: initiation, elongation and termination. Although each single step of this mechanism is a target for regulation, the first step, translation initiation, is the most complex and rate-limiting one and therefore an important target for control mechanisms [4]. The 5' end of all eukaryotic nuclear transcribed mRNAs contain the 5' cap. The 3' end of the transcript is characterized by a long (200-500 nt) adenosine stretch. Eukaryotic translation initiation factors (eIFs) recognize these two elements and subsequently recruit various cofactors for the assembly of the translation initiation complex resulting in a pseudo-circularization of the mRNA molecules [5]. As a matter of fact, this complex is the target for many factors stimulating or inhibiting the translation of specific mRNAs. Interestingly the most common scenario for regulation are RNA-binding proteins interacting with specific motifs in the 3' UTR of their targets [6], [7], [8], [9], [10]. This case also applies to the Cytoplasmic Polyadenylation Element Binding protein (CPEB) family of proteins.

The CPEB family of proteins

The CPEB family of proteins comprises of four paralogs (CPEB1-4). CPEBs 2-4 are closely related and CPEB1 is the most distant relative [11]. The proteins are widely expressed in a variety of tissues, cell types and tumors with partially overlapping patterns. Additional functions to that of oocyte maturation have been identified including roles in synaptic plasticity and cellular senescence [12]. All members of the family comprise of a largely conserved C-terminal region that contain two RNA recognition motifs (RRM1 and RRM2), which have been shown to be essential for the CPE interaction, and a ZZ-type domain which is postulated to be a protein-protein interaction module [13], [14]. The N-terminal half of the protein is more variable between the different members. In CPEB1 it contains a PEST-degradation motif, the site for Aurora A mediated phosphorylation [15] and several Cdc2 [16] and Plk1 [17] phosphorylation sites that trigger its ubiquitination and degradation through the PEST box [18].



Figure 1.1: Schematic representation of the topology of the members of the CPEB family

Although the paralogs CPEB2-4 are less extensively studied, their N-terminal part also seems to contain nuclear import and export signals as well as regulatory phosphorylation sites. CPEB4, for instance, lacks a consensus Aurora A kinase sequence,

nevertheless it has been reported to become phosphorylated in response to progesterone [19], [20].

CPEB mediated translational repression and activation

The CPEB protein family is involved in the translational regulation of cytoplasmic polyadenylation element (CPE) containing mRNAs. The mechanism by which cytoplasmic polyadenylation controls the translation of many key mRNAs has been elucidated in meiotic maturation of *Xenopus* oocytes. Many maternal mRNAs are translationally repressed in oocytes arrested at the end of meiosis prophase I. This repression is achieved *via* the interaction of CPEB with the *cis* acting element CPE, a U-rich sequence (U₄₋₆A₁₋₂U₁₋₂) [21],[22] in the 3'UTR of the specific mRNAs and the recruitment of various protein factors by CPEB that facilitate translational regulation. The poly(A)-specific ribonuclease PARN and the noncanonical poly(A) polymerase Gld2 are recruited by CPEB. Although both protein factors are active, their antagonistic behavior results in mRNAs with a short poly(A) tail, which are therefore transitionally repressed [23]. The short poly(A) tail furthermore prevents its interaction with ePAB, which in turn associates with CPEB [24]. Other factors that interact with CPEB include members of the mRNA 3' end processing machinery such as symplekin and CPSF (cleavage and polyadenylation specific factor) [25] (see Figure 1.2 *upper left*).

An alternative model of the repression complex involves the interaction of CPEB with the protein Maskin and its interaction with the translational initiation factor eIF4E [26], [27]. Therefore the recruitment of eIF4G by eIF4E is blocked and inhibits the assembly of the translation initiation complex (see Figure 1.2 *lower left*). CPEB1 and CPEB4 have been shown to have a dual function as a translational repressor as well as activator. CPEB2 and CPEB3, however, seem to solely act as translational repressor.

Upon reentry into meiosis, a progesterone stimulated signaling cascade leads to the activation of kinase Aurora A which in turn phosphorylates CPEB1 [19]. The phosphorylation of CPEB1 results in the expulsion of PARN from the ribonucleo-protein complex and in the enhancement of the CPEB-CPSF interaction leading to Gld2-catalyzed poly(A) tail extension of the mRNA [24]. The poly(A)-tail recruits further protein factors important for the subsequent translation of the transcript (see Figure 1.2 *right*). It has been demonstrated that although CPEB1 and CPEB4 are functionally interchangeable, during meiotic progression they are differentially regulated by phase specific kinases.

In general, the activation of CPE containing mRNAs is not occurring en masse during meiotic progression. Instead, the polyadenylation of mRNAs is temporarily regulated by CPEBs [19]. At Prophase I (PI) the phosphorylation of CPEB1 by Aurora A kinase is required for the first wave of polyadenylation as well as the PI-MI (Metaphase I) transition [19]. Subsequently at MI, Cdc2 and Plx1 phosphorylate CPEB1, which

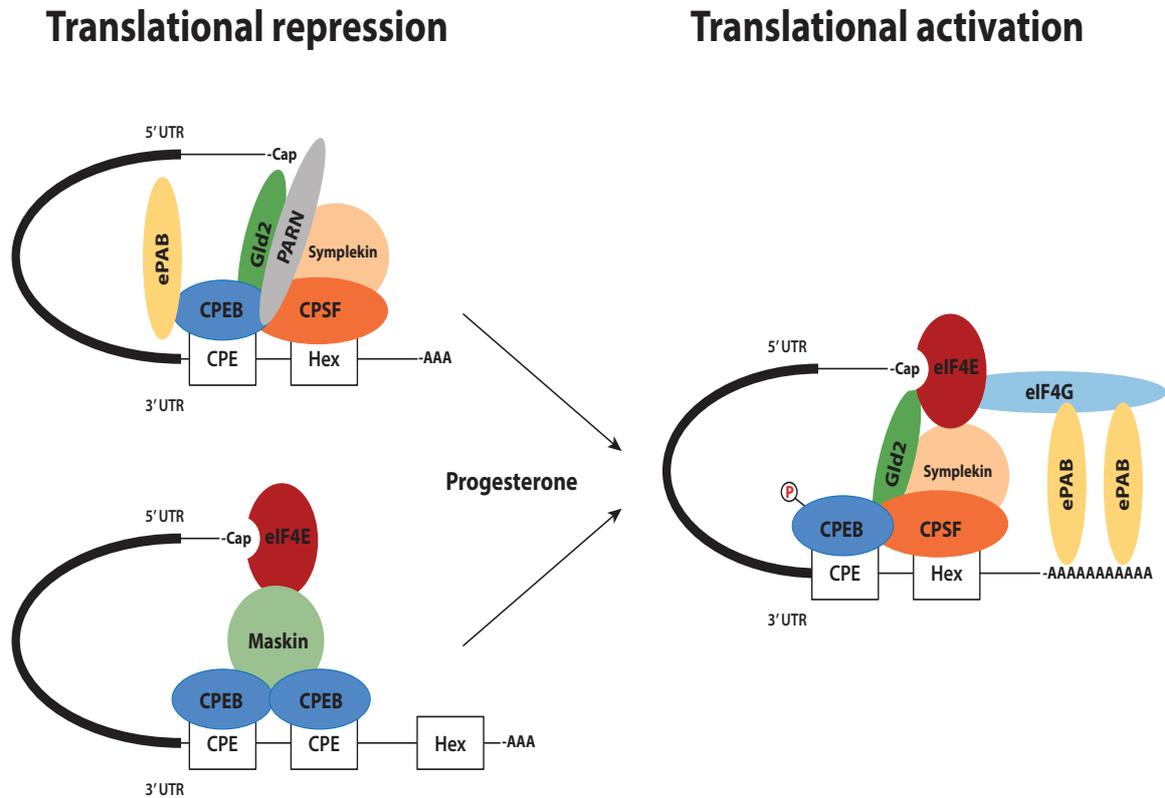


Figure 1.2: CPEB mediated translational control mechanisms. Two schematic representations for repression complexes (adapted from [4] and [2]). *upper left:* Repression complex mediated by CPEB interaction with PARN and Gld2 resulting in a shortened poly(A) tail. *lower left:* Repression complex mediated by CPEB interaction with Maskin blocking the formation of the cap-binding complex. *right:* Upon progesterone stimulation CPEB is phosphorylated and the activation complex is formed.

targets it for degradation. The low level of CPEB1 at this time is necessary to activate the second wave of polyadenylation and the Metaphase I-Metaphase II transition. However, the consequences are very low levels CPEB1 in the cell for the second meiotic division, when a third wave of cytoplasmic polyadenylation is essential for MII entry and cytostatic factor (CSF) arrest. In addition, Aurora A kinase is inactivated during interkinesis simultaneously with increased levels of Protein Phosphatase 1 (PP1), which dephosphorylates CPEB1 on the Aurora A phosphorylation site and renders it inactive regarding translation [28]. For the second meiotic division CPEB4 replaces CPEB1 and recruits the polyadenylation machinery necessary to activate the translation of mRNAs encoding proteins needed. CPEB4, which is encoded by a maternal mRNA activated by CPEB1 during the first wave of cytoplasmic polyadenylation, is subsequently partially inactivated by C3H-4-mediated deadenylation. This translational regulation leads to the gradual accumulation of CPEB4 from MI to reach

1.1 CPEB family: Cytoplasmic polyadenylation and translational control

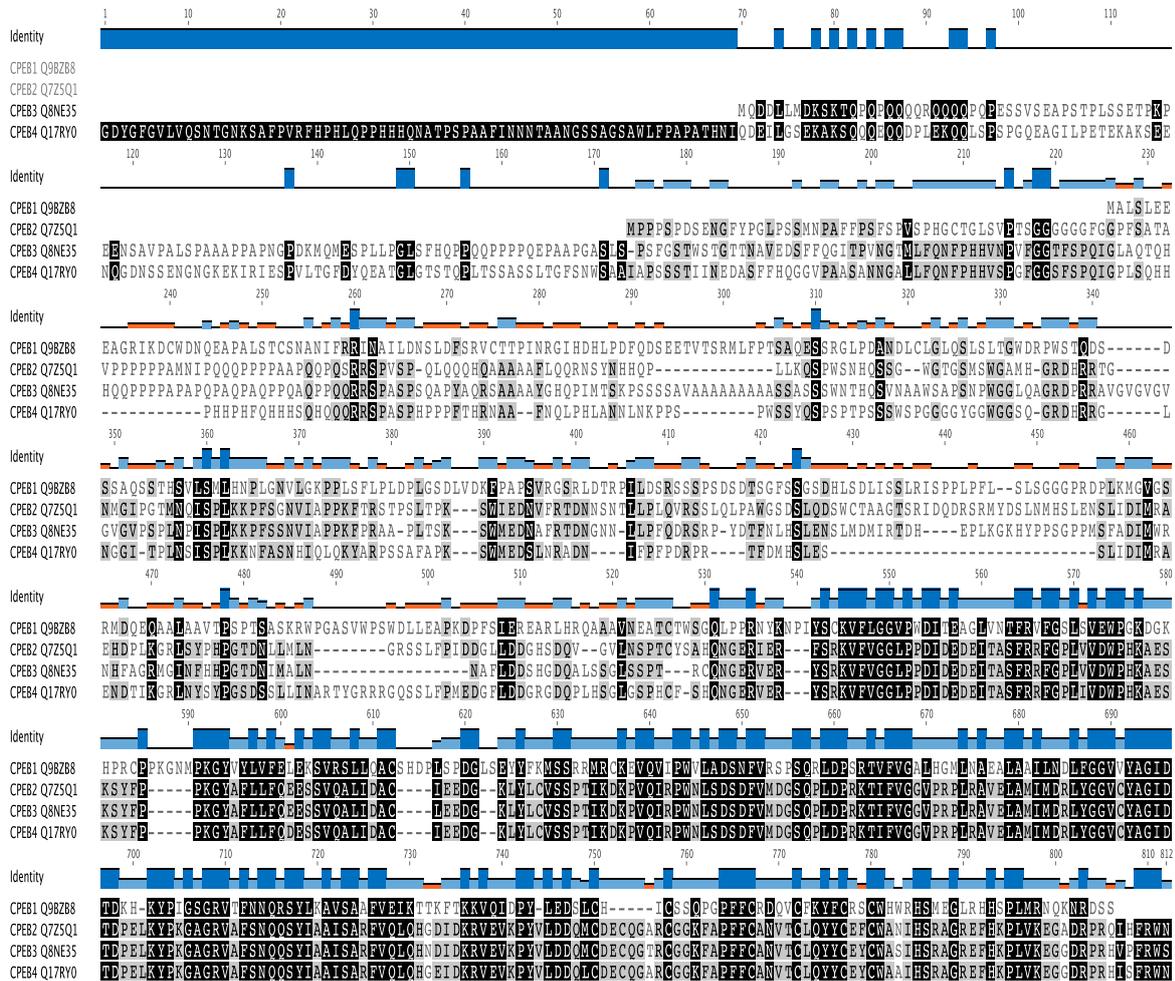


Figure 1.3: Sequence alignment displaying the high degree of conservation of the C-terminal RRM domains and ZZ-domains within the CPEB-family. The RRMs of CPEB2-4 share 97% sequence identity between them. In contrast, the RRMs of CPEB1 share 48% pairwise sequence identity with those of CPEB2-4. The N-terminal part shows large variation between the single family members. Alignments were prepared using the Geneious R6 software (Biomatters, <http://www.geneious.com>).

maximal levels in the MII arrest. Reflecting the relative levels of CPEB1 and CPEB4, CPEB1 mediates cytoplasmic polyadenylation in the first meiotic division and CPEB4 in the second [19]. This process is regulated by different signal transduction pathways to ensure that they stay in their active forms at the appropriate meiotic phases and the polyadenylation machinery is active during the whole meiotic progression. The self-sustainability of the three waves of cytoplasmic polyadenylation, once the initial stimulation by progesterone takes place, is accomplished by the translational control of CPEB4 by CPEB1 before the degradation of the latest by the APC (Anaphase-Promoting Complex)[19].

RNA recognition motifs

The RNA recognition motif (RRM) was first identified when it was demonstrated that mRNA precursors (pre-mRNA) and heterogenous nuclear RNA (hnRNA) are always found in complex with proteins (reviewed in [29]). It is also known as RNA-binding protein (RBD) or ribonucleoprotein domain (RNP). The RRM domain typically comprises ≈ 90 amino acids and has been identified in all life kingdoms with its highest abundance in eukaryotes. Up to now a total of 49120 RRM domains in 32653 different proteins have been identified. In humans, there are 829 known proteins containing at least one RRM (<http://pfam.xfam.org/family/PF00076tabview=tab7>). RRMs are frequently found as multiple copies within one protein (44 % contain two to six domains) and/or adjacent to other protein domains.

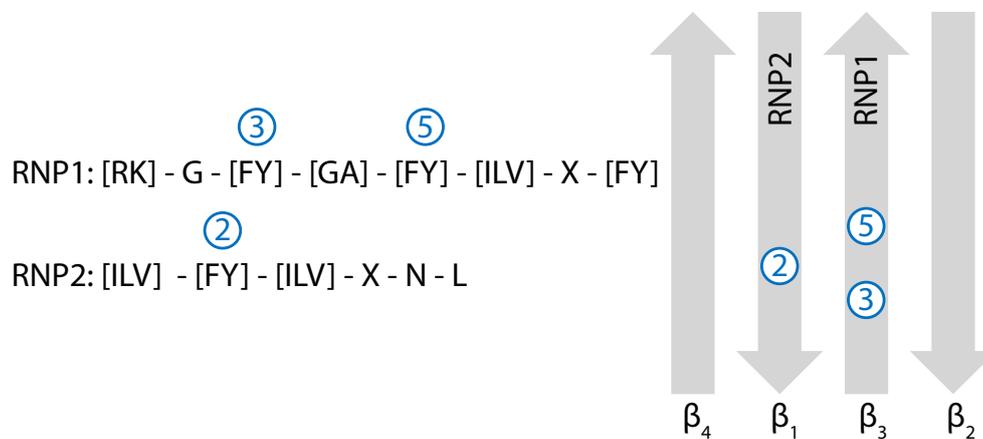


Figure 1.4: Scheme of the four-stranded antiparallel β -sheet indicating the conserved RNP1 on β_3 -strand and RNP2 on β_1 -strand in blue (adapted from [30]). RNP1 and RNP2 consensus sequences of RRMs are displayed. X stands for any amino acid.

The canonical RRM domain fold is a $\alpha\beta$ -sandwich with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology [31]. The fold is composed of a four-stranded antiparallel β -sheet, which is spatially arranged in the order $\beta_4\beta_1\beta_3\beta_2$ when facing the sheet, and two α -helices packed against the β -sheet. Extensions to the canonical fold such as an additional α -helix or β -strand formed by loop regions or the usually poorly ordered C- or N-terminal regions have been reported for various structures solved. The β -sheet has been identified to be the conserved binding platform for RNA interactions. In the central strands of the RRM β -sheet lie two conserved motifs, known as the RNP consensus sequence, that are generally involved in RRM-RNA interactions: RNP1 and RNP2. The consensus sequence of RNP1 is: (RK)-G-(FY)-(GA)-(FY)-(ILV)-X-(FY) while that of RNP2 is: (ILV)-(FY)-(ILV)-X-N-L, where X can be any amino acid.

Regarding the CPEB proteins, the first RRM domain (RRM1) contains, for the most part, both RNP sequences. However, the second (RRM2), more C-terminal domain

lacks the consensus sequence of an RNP1 motif. This characteristic is shared among all CPEB family members (cf. Figure 1.5).

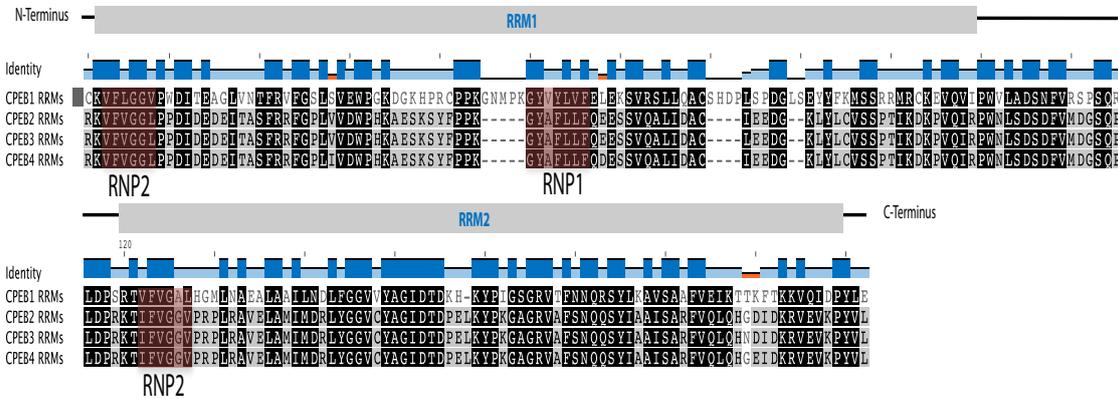


Figure 1.5: Sequence alignment of the C-terminal RRM domains within the CPEB-family. The RNP motifs of both RRM1 and RRM2 are indicated by a red box. Alignments were prepared using the Geneious R6 software (Biomatters, <http://www.geneious.com>)

Since RRM domains are occurring frequently as multiple copies in one protein, the interplay between the domains provides a high versatility of the modes of nucleic acid binding possible. The relative orientation of the RRM domains and in particular whether these RRMs are interacting with each other and adopt a single defined relative orientation are important factors characterizing the mode of RNA recognition (reviewed in [32] and [33]). For tandem RRM domains, as in CPEB, the following domain arrangements in the free as well as in the bound state have been identified so far in various structures. In the free state the domains can either be independent, with pre-formed domain contacts or adopt a closed, autoinhibited form. Upon ligand binding RRMs can, for example, independently bind separate RNA motifs or form a continuous RNA-binding platform, stabilized by inter-domain contacts or increased linker rigidity. Another mechanism to stabilize the RNA binding is mediated by protein-protein interactions with an auxiliary protein. The combination of two RRM domains provides a range of possible conformations and allows the continuous recognition of longer nucleic acids sequences resulting in a modulation of RNA binding affinity and specificity compared to the isolated RRM domains.

1.2 Peptidyl-prolyl isomerase Pin1

One important and universal regulatory mechanisms in the cell is the reversible phosphorylation of proteins. The ability to define the regulatory components of the phosphorylation/dephosphorylation cascades is essential to understand various biological processes and their molecular mechanisms. The phosphorylation of proteins on serine or threonine residues preceding a proline represents a key signaling mechanism for controlling the correct function of signaling molecules in various cellular processes. Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (Pin1) is a member of the peptidyl-prolyl isomerase (PPIase) family that recognizes p(Ser/Thr)-Pro motifs within phosphorylated proteins [34]. Pin1 is the only phosphorylation dependent cis-trans isomerase and catalyses the prolyl isomerization of the specific phosphorylated Ser/Thr-Pro bonds.

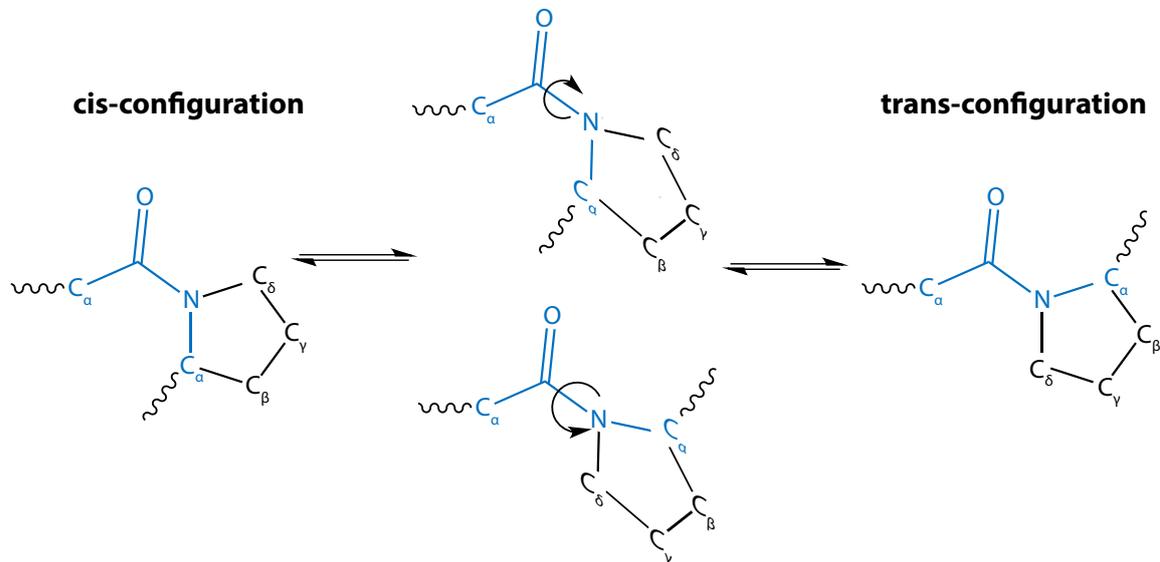


Figure 1.6: Isomerization of an X-Pro peptide bond. The partial double bond has already been broken and therefore the energy needed for configuration conversion is decreased. Transition states are indicated in the middle.

Pin1 is a relatively small, ~160 amino acids, enzyme that contains an N-terminal WW domain acting as a substrate binding module and a C-terminal catalytic domain, which are connected through a thirteen amino acids glycin-rich flexible linker. WW domains are 38-40 amino acids domains characterized by two highly conserved tryptophanes. Their canonical fold is a three stranded antiparallel β -sheet, which depending on its classification recognizes PY motifs (group I), PPLP motifs (group II), poly PR motifs (group III) or p(S/T)P-motifs (group IV). Pin1 WW domain belongs to the group IV WW domains [35], [36].

Isomerization of Ser/Thr-Pro motifs is crucial since kinases and phosphatases specif-

ically recognize the cis or the trans configuration of their substrate's prolyl peptide bond. In the folded state of a protein, generally only a single isomer is adopted at each position. The trans form is preferred in most peptide bonds (roughly a 1000:1 ratio in trans:cis populations) [37]. For X-Pro peptide groups the ratio tends to be 3:1, presumably because the symmetry between the C_α and C_δ atoms of proline renders the cis and trans isomers nearly equal in energy. Nonetheless, the configuration conversion only occurs very rarely, resulting in a predominant trans-configuration. Moreover, phosphorylation of the adjacent serine or threonine further slows down the isomerization rate of prolines. The transition from cis to trans requires the partial double bond to be broken, therefore an activation energy of ~30 kcal/mol is needed. This activation energy can be lowered by changes that favor the single-bond form. For instance, placing the peptide group in an hydrophobic environment or donating a hydrogen bond to the nitrogen of an X-Pro peptide group would catalyze the isomerization. These mechanisms for lowering the activation energy have been observed in PPIases such as Pin1 [38], [39]. Due to its wide range of target proteins, Pin1 activity controls various protein functions in diverse cellular processes, such as cell cycle control, transcription, splicing regulation, DNA replication, and DNA damage response [40], [41]. Importantly, Pin1 is tightly regulated by several mechanisms, and its deregulation can contribute to an increasing number of human diseases like cancer, neurological disorders or autoimmune diseases [42], [43], [44], [45].

Pin1 mediated degradation by the ubiquitin-proteasome system

Pin1-mediated post-phosphorylation modifications can have profound effects on phosphorylation dependent signaling by regulating a spectrum of target activities, with changes in protein stability being the most common consequence [46]. Pin1-catalyzed prolyl cis/trans isomerization could be a crucial step in controlling the phosphorylation-dependent ubiquitylation pathway. Ubiquitin-mediated protein degradation by the ubiquitin-proteasome system (UPS) has been found to be responsible for the regulated degradation of the majority of intracellular proteins through the 26S proteasome [47]. The catalytic processes of phosphorylation and ubiquitylation occur very rapidly under normal conditions [48]. The cis/trans configuration conversion, however, which is crucial for E3 ligase recognition, is the rate limiting process. The presence of Pin1 enhances the cis/trans configurational changes by reducing the free energy barrier, resulting in an increased conversion rate up to 100-1000 -fold, indicating that Pin1 plays an important role in regulating protein degradation [46].

The first structure of full-length Pin1 was solved in 1997 by X-Ray crystallography [49], in 2003 the first solution structure of the full-length protein was published [50]. Structural studies on the Pin1 WW domain free and in complex with phosphorylated substrate peptides have been accomplished and the binding mode for Pin1 WW substrate recognition has been described [51], [52]. However, the exact mechanism and

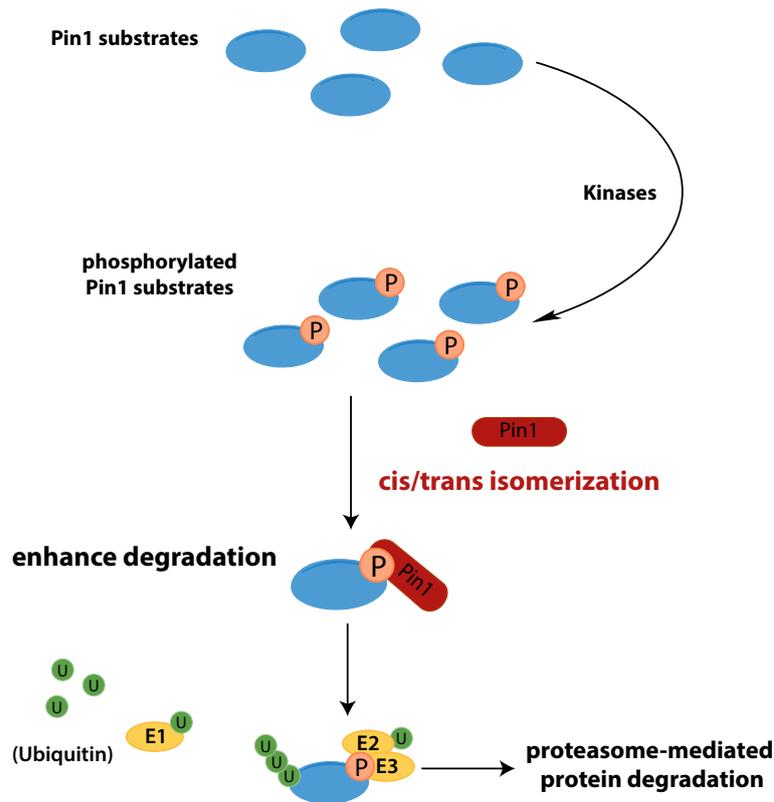


Figure 1.7: Schematic representation by Pin1 mediated substrate degradation (adapted from [46]). Pin1 boosts cis/trans isomerization rate to provide the right configuration for subsequent ubiquitylation of phosphoproteins by a E3 ligase and to activate proteasome-mediated protein degradation.

dynamics of Pin1 binding and isomerization has not been elucidated at molecular level.

Pin1 mediated destruction of CPEB1

It has been shown that during oocyte maturation, a partial destruction of CPEB1 is necessary for the transition from MI to MII and its replacement by CPEB4 [20]. This mechanism guarantees the hierarchal, time-specific manner of the polyadenylation of different CPE-containing mRNAs regulated by the level of CPEBs. CPEB1 degradation during maturation requires a Cdc2-mediated phosphorylation of six Ser residues (138, 144, 184, 210, 248, and 423) in Ser-Pro motifs, although it has been demonstrated that phosphorylation of serine 210 is essential for subsequent degradation [16]. Phospho-CPEB1 has been recently identified as a target for Pin1 and interestingly it has been found to associate with CPEB also prior to detectable CPEB phosphorylation [53]. It has been demonstrated that both interactions at the first phase depend

on the WW domain of Pin1 [53]. Nonetheless, the presence of the PPIase domain is indispensable for full Pin1 activity to promote CPEB destruction [53].

The cis/trans isomerization and subsequent configurational change imposed by Pin1 helps establishing a proper three-dimensional architecture for E3 ligase recognition, resulting in the ubiquitination and destruction of target proteins [46], [54], [55]. It has been reported that the $SCF^{\beta-TrCP}$ ubiquitin ligase is involved in the degradation of CPEB [17] by its interaction with the conserved sequence ${}_{190}TSGFSS_{195}$ when T_{190} , S_{191} and S_{195} are phosphorylated by Plx1. In order to create a docking site for Plx1 T_{125} , forming part of the ${}_{124}STP$ Plx1-docking motif, needs to be phosphorylated. It has been shown *in vitro* that apart from the six serine residues, Cdc2 also phosphorylates T_{125} [17].

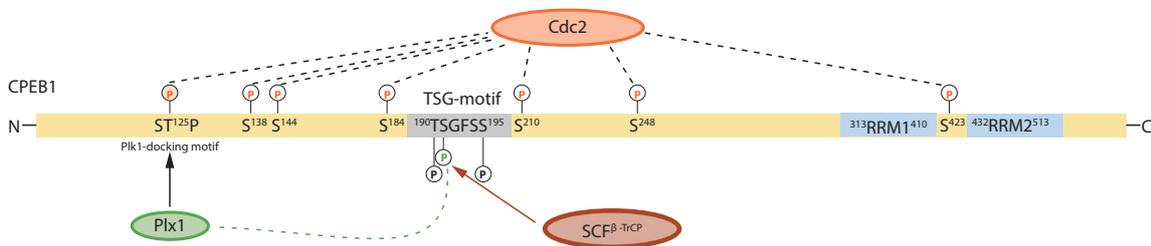


Figure 1.8: Schematic model for the molecular mechanisms of CPEB degradation adapted from [17]. Cdc2 phosphorylates CPEB on T_{125} and six further SP motifs. Subsequent binding of Plx1 to T_{125} is required for phosphorylation of S_{191} in the TAG-motif by Plx1. The phosphorylated TSG-motif is recognized by $SCF^{\beta-TrCP}$ ubiquitin ligase targeting CPEB for degradation.

It was suggested that the six serine phosphorylations by Cdc2 are probably not only required for phosphorylations of the TSG motif, but also play an important role in another event involved in efficient $SCF^{\beta-TrCP}$ binding. Taking into account their rather dispersed locations (138, 144, 184, 210, 248, and 423) and their requirement for subsequent T_{125} phosphorylation by Cdc2 itself, the multiple serine phosphorylations by Cdc2 could be the key to allow efficient $SCF^{\beta-TrCP}$ binding as well as T_{125} phosphorylation [17].

All of the six serine phosphorylation sites within CPEB1 are followed by a proline and therefore are potential binding sites for the Pin1 WW domain. Up to now, a detailed analysis of potential interactions of Pin1 with the single phosphorylation sites has not been performed. It has only been demonstrated that upon deletion of regions 48-183 and 211-290 in CPEB1 the CPEB1-Pin1 interaction is abrogated [53].

2 Aims of the thesis

The Cytoplasmic Polyadenylation Element Binding (CPEB) proteins are RNA binding proteins, which are found in a wide range of vertebrates and invertebrates. This family of proteins mediates many biological processes such as germ-cell development, synaptic plasticity and cellular senescence. Furthermore CPEB regulates polyadenylation-induced translational control. Processing of primary transcripts by RNA splicing and polyadenylation, as well as translational control by RNA structures, RNA binding proteins and microRNAs, opens a myriad of opportunities for gene regulation which are extensively used in virtually every human gene. Alterations of these finely tuned processes have been reported as the basis of pathological processes including oncogenic transformation and tumor progression.

The mechanism by which cytoplasmic polyadenylation controls the translation of many key mRNAs has been elucidated in meiotic maturation of *Xenopus* oocytes. Many maternal mRNAs are translationally repressed in oocytes arrested at the end of meiosis prophase I. This repression is achieved via the interaction of CPEB to the CPE U-rich sequence ($U_{4-6}A_{1-2}U_{1-2}$) in the 3'UTR of the specific mRNAs. The CPEB family of proteins comprise of four paralogs (CPEB1-4). CPEB's 2-4 are closely related and CPEB1 is the most distant relative in terms of sequence conservation. The functional differences between the two groups are yet to be described in detail. Interestingly, despite the fact that CPEB 2-4 share 97% pairwise sequence identity regarding their RNA Recognition Motifs, only CPEB1 and CPEB4 have been found to have a dual function as translational activator as well as repressor.

Unlike the RNA-binding domains, the N-terminal half of the protein is more variable between the different members and only in CPEB1 contains a PEST-degradation motif and the site for Aurora A mediated phosphorylation. However, a common feature is the occurrence of phosphorylation sites, several of which are uncharacterized. Among various protein interactions, Pin1 was recently identified to be a new and essential factor regulating CPEB degradation upon its phosphorylation. It has also been shown that even prior to CPEB phosphorylation Pin1 interacts with CPEB1. Phosphorylation dependent CPEB1-Pin1 interaction is based on Cdc2-catalyzed CPEB1 phosphorylation on SP motifs.

1. Characterization of CPEB4 RRM domains

The first aim of this thesis was to characterize the interaction of the two RRM domains of CPEB4 with various cytoplasmic polyadenylation element (CPE) containing RNA fragments. Of special interest was whether both domains would contribute to the binding, since the second RRM domain doesn't feature the consensus binding motif found in RRMs. Furthermore, the study of the dynamic properties of both the free domains and in complex with RNA was one of the objectives. To achieve these goals, isothermal titration calorimetry, ion mobility mass spectrometry and nuclear magnetic resonance were applied as the mayor biophysical techniques, since they provide complementary information. In this regard, since a possible application of our study is the design of inhibitor molecules, crystallization trials were set-up. Preliminary results are presented.

2. Characterization of CPEB1-Pin1 interaction

The second part of the thesis was focused on studying the recently discovered interaction of the N-terminal part of CPEB1 with the prolyl *cis/trans* isomerase Pin1 in order to understand its role within the CPEB1 degradation pathway. The interaction has been described as a two-step mechanism where an interaction occurs already prior to CPEB phosphorylation. In particular, we have been interested to study which N-terminal regions are involved in Pin1 binding prior to phosphorylation of CPEB and whether this interaction would resemble the ones described of Pin1 with other p(S/T)P motif such as Smad3 pT179 or pS213 [56]. Within CPEB1 there are six identified Cdc2 phosphorylation sites, which are all potential binding partners for Pin1. We therefore focused on understanding the differences and similarities in binding of all six potential binding partners, particularly with regard to the contributions of Pin1's substrate binding domain and catalytic domain. Once identified the special properties of CPEB1 pS210, the dynamic characteristics, especially with regard to the changes in Pin1 inter domain interaction were investigated. In particular the understanding of the binding mode of Pin1's WW domain with its CPEB1 pS210 substrate at an atomic level and comparison with other known substrate binding modes was another objective of the work presented here.

3 Material and Methods

3.1 Cloning of a target protein

For the majority of techniques used in structural biology the protein sample needs to be of high purity and concentration. In general recombinant proteins are expressed in *Escherichia coli* (*E.coli*) bacteria which priorly have been transfected with a bacterial plasmid as a vector including the DNA of the target protein.

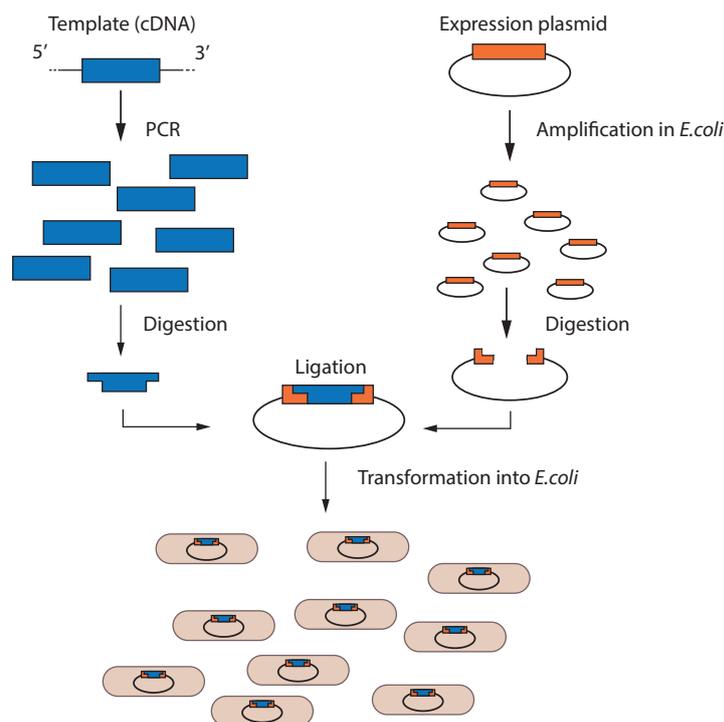


Figure 3.1: Basic workflow of ligation-based cloning for expression in *E.coli*

The construct design is a crucial step when cloning only a fragment of the full-length protein. For the classic ligation-based cloning the design of primers (short DNA fragments) for the polymerase chain reaction (PCR) is the initial step. Primers containing sequences complementary to the target region along with a DNA polymerase are key components to enable selective and repeated amplification. The PCR product is con-

sequently digested by highly specific restriction enzymes, which leave an overhang at each DNA end, also known as “sticky ends“. The vector is then digested using the same enzymes, so that the PCR fragment can be inserted into it by the help of a DNA ligase. For the insertion of a DNA fragment into a plasmid vector, it is preferable to use two different restriction enzymes to digest the DNA in order to generate different cloning site ends. Therefore religation of the vector without any insert can be prevented and the fragment is inserted in a directional manner.

Bioinformatic tools, which are open access online, provide helpful information about conserved regions, expected secondary structures or functional residues, based on which the boundaries of the construct of interest can be decided [57]. Various chemical and physical parameters (e.g. molecular weight, theoretical pI, amino acid composition, extinction coefficient, etc.) of a given construct can be calculated using the *ProtParam tool* of the *ExPASy* server (<http://web.expasy.org/protparam/>).

For protocols and details regarding expression and purification of the cloned proteins, please refer to section 3.6.

3.2 Fmoc Solid-Phase Peptide Synthesis - Basic principles

Solid-Phase Peptide Synthesis (SPPS), developed by Robert Bruce Merrifield in the 1960s [59], allows the synthesis of natural peptides as well as the incorporation of modified amino acids, for example phospho-amino acids in an easy and efficient manner. The basic principle is the growth of the target peptide with its C-terminal residue attached to an insoluble support via its carboxyl group. Any functional group within amino acid side-chains must be masked with protecting groups, that remain unaffected by the reaction conditions employed during the peptide chain growth. In addition *temporary* protecting groups shielding the α -amino group during the coupling reaction are applied. These groups are removed revealing the new N-terminal amine, to which another residue can be coupled. An excess of the $(n + 1)^{th}$ amino acid is introduced while activating the carboxy group of the n^{th} , already attached amino acid, by reaction with a coupling agent. The great advantage of this approach for peptide synthesis lies in the ability to perform wash cycles, removing excess reagents, after each reaction step while the intermediate peptide of interest remains covalently attached to the resin support. After the $(n + 1)^{th}$ coupling the protecting group of the N-terminus of the $(n + 1)$ -peptide is removed, prior to the addition of the $(n + 2)^{th}$ residue. This process is repeated until the desired peptide sequence is synthesised. Consequently the side-chain protecting groups are removed and the peptide is released from the support resin. When possible, conditions for the side-chain removal and resin cleavage of the peptide are chosen such that these reactions are carried out in the same conditions.

Nowadays, the Fmoc-approach proposed by Carpino in 1972 [60] is commonly used. This method has been applied to prepare the peptides used in this work. Unlike the Merrifield approach of SPPS where a regime of graduated acidolysis is applied to achieve selectivity in the removal of temporary and permanent protection, the Fmoc (9-fluorenylmethoxycarbonyl)-based SPPS relies on an orthogonal protecting group strategy. The base-labile Fmoc-group is used for protection of the α -amino group, whereas acid-labile side-chain protecting groups and resin linkage are chosen. Owing to the completely different reaction mechanisms applied for the removal of temporary and permanent protection, side-chain protecting groups and linkage agents can be employed, that are removable under appreciably milder conditions compared to the ones applied in the Merrifield approach. The Fmoc protecting group is readily cleaved under mild basic conditions such as 20 - 40 % v/v piperidine in dimethylformamide (DMF). Since the majority of the amino acids have side-chains featuring reactive functional groups, those potentially reactive groups are normally protected against the rather harsh conditions applied during synthesis. Permanent protecting groups commonly employed are 2,2,4,6,7-pentamethyldihydro-benzofuran-5-sulfonyl (Pbf), tert-Butyl (tBu) or trityl-based side-chain protecting groups. These protecting groups are

stable under basic conditions, but easily released at low pH allowing to globally de-protect the peptide at the same time as the cleavage of the support is performed.

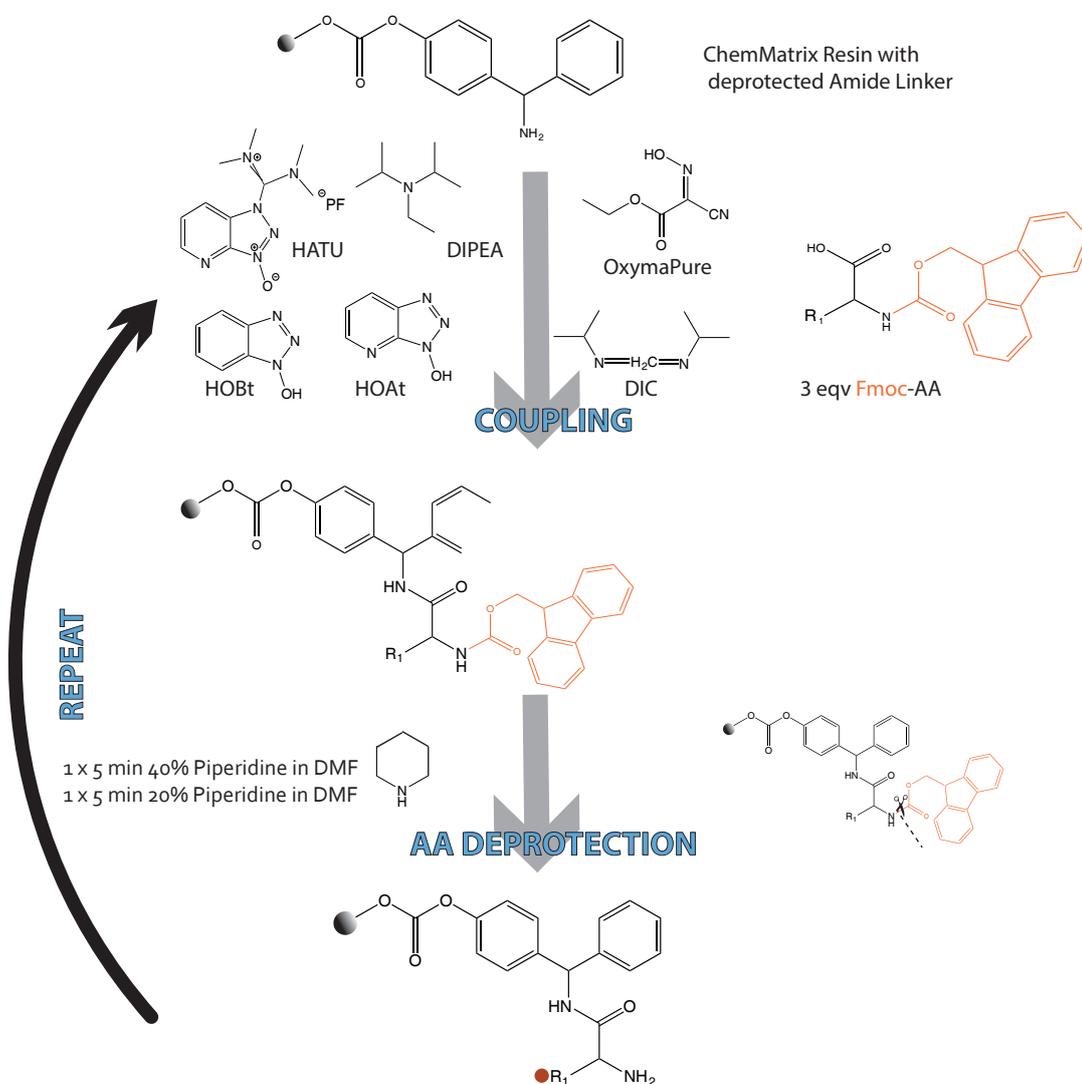


Figure 3.2: Schematic workflow of Fmoc solid phase peptide synthesis

The coupling of the $(n + 1)^{th}$ Fmoc-amino acid is driven by a nucleophilic attack of the deprotected amine group of the n^{th} residue at the carbonyl group of amino acid n . The increase of the nucleophilic substitution rate is achieved by the addition of substances like carbodiimides or triazoles, which produce highly reactive esters. However, due to the high reactivity of the carbodiimide esters racemization can occur. Triazoles like 1-hydroxy-benzotriazole (HOBT) or 1-hydroxy-7-aza-benzotriazole (HOAt) form an active ester which is less reactive and thus less prone to racemize. An alternative

to HOBt is ethyl 2-cyano-2-(hydroxyimino)acetate (trade name Oxyma Pure), which in contrast to HOBt is not explosive, when fully dehydrated. Further developments led to activating groups like 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) omitting carbodiimides completely introducing the active ester as a phosphonium salt of a non-nucleophilic anion (e.g. hexafluorophosphate).

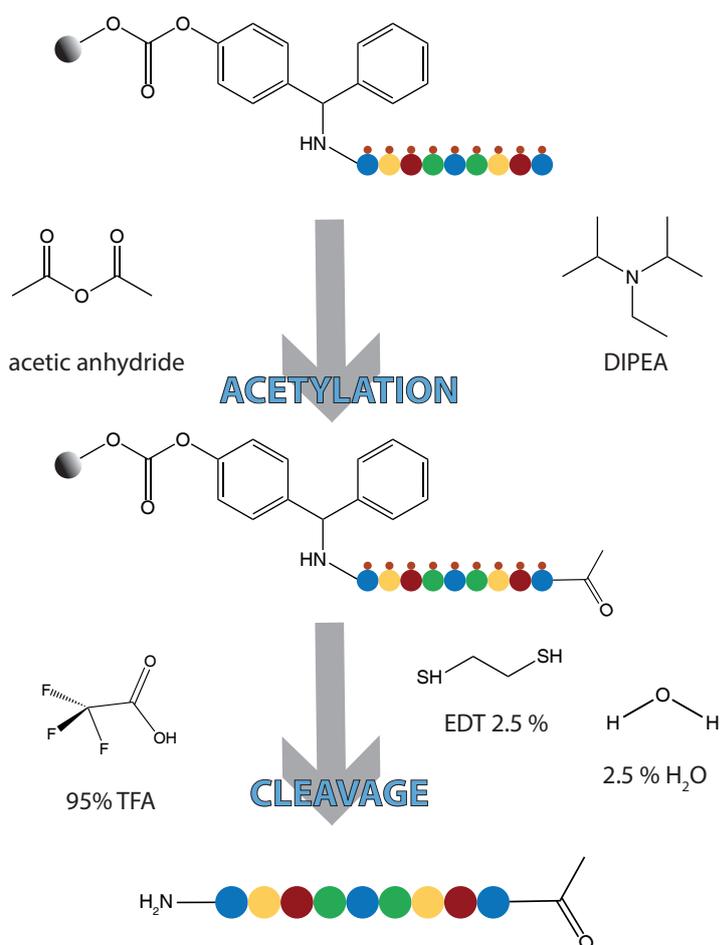


Figure 3.3: Schematic workflow of Fmoc solid phase peptide synthesis

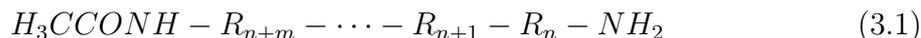
During the assembly of the peptide chain, each coupling step can be qualitative monitored with colorimetric tests for the presence or absence of free amino groups. The ninhydrin test, designed by Kaiser, is the one most commonly applied. It is based on the reaction of ninhydrin with primary amino groups forming a blue adduct. For the detection of secondary amines the chloranil test was applied.

The cleavage of peptides from acid-sensitive linkages is commonly achieved by the

treatment with Trifluoroacetic acid (TFA). Under this condition side-chain protecting groups are also cleaved resulting in a fully side-chain deprotected peptide. During the cleavage process highly reactive cationic species are generated from the protecting groups and resin linkers which are prone to react with amino acids containing electron-rich functional groups, for example tyrosine, tryptophan, methionine or cysteine. Therefore nucleophilic reagents, so called scavengers, are added to the TFA in order to quench reactive carbocations. Commonly used scavenger molecules are water, triisopropylsilane (TIS), triethylsilane (TES) or 1,2-Ethanedithiol (EDT).

Following the cleavage process, the peptide is precipitated by the addition of cold diethyl ether. The precipitation can be performed directly from the TFA-solution or after evaporating the majority of the TFA and volatile scavengers. The latter approach usually increases the yield, especially for short peptide sequences.

All peptides synthesized for this work have been acetylated before cleavage and therefore are following this schematic composition:



Protein phosphorylation by protein kinases is an important mechanism by which eukaryotic cellular processes are modulated. The development of methods for the synthesis of phosphopeptides has been of great interest since phosphopeptides are essential tools for the study of protein phosphorylation. Currently there are two strategies for the preparation of phosphopeptides: the global phosphorylation method, which involves post-synthetic phosphorylation and the building block approach, which has been applied for all phosphopeptides synthesised in this work. The building block approach is the more straightforward method and relies on the incorporation of pre-formed protected phosphoamino acids, such as Fmoc-Ser(PO(OBzl)OH)-OH and Fmoc-Thr(PO(OBzl)OH)-OH. Incorporation of phosphoamino acids can thoroughly affect the ease of the synthesis of a given sequence, especially the coupling of the residue following the phosphoamino acid. Repetition of couplings and profound monitoring of the couplings are therefore advisable.

3.3 Biophysical methods

3.3.1 Ion Mobility -Mass Spectrometry

An ion mobility mass spectrometer features an ion mobility cell before the mass analyzer [63]. The gaseous charged analyte ions are formed using nano flow electrospray ionisation. Gaseous ions are injected into the cell and accelerated by a weak electric field. The application of soft ionisation techniques is crucial for biomolecular ions as these ionisation methods preserve tertiary structures and non-covalent interactions of complexes or aggregates. Due to the presence of buffer gas in this cell, low energy collisions with the buffer gas occur. The higher the collision cross sections (CCS) of the ion, the greater the number of collisions with the buffer gas. As collisions increase an energy loss of the ions occur and accordingly ions take a longer to cross the IM cell (the drift time). Consequently ions are injected into the mass analyzer, achieving a simultaneous separation on the basis of the CCS to charge ratio and the m/z ratio. The three-dimensional spectrum obtained consists of mass, drift time and relative intensity.

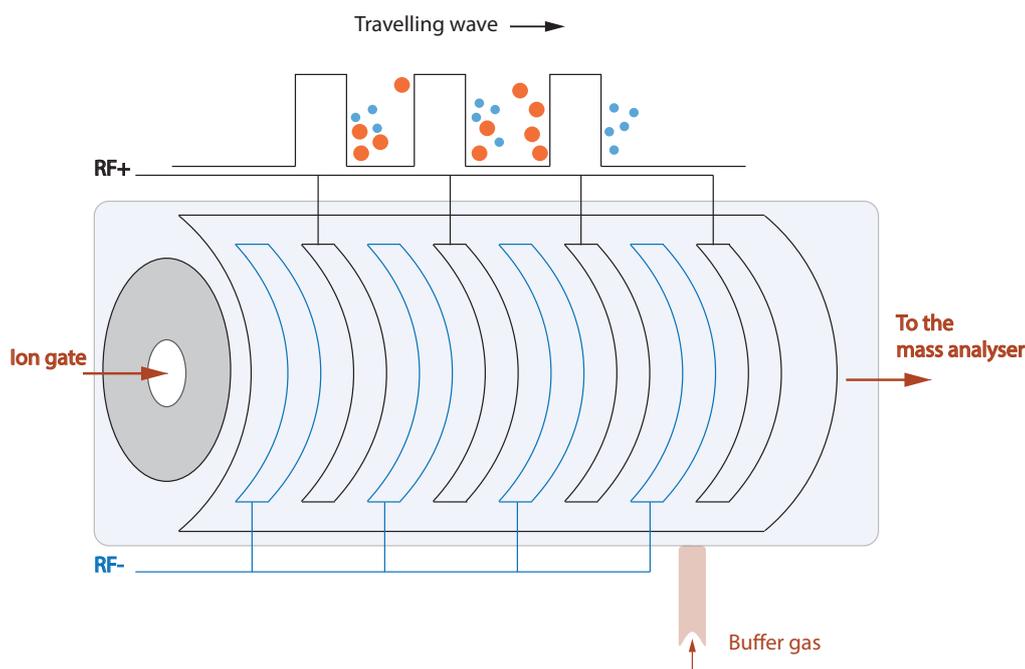


Figure 3.4: Separation of gaseous ions on the basis of ion mobility: Schematic representation of traveling wave IM, in which ions are pushed by consecutive voltage pulses. The ions that show more resistance with the buffer gas are more likely to surf the wave

Travelling Wave IM (TWIM), in comparison to drift tube or high field asymmetric

waveform IM, is the prevalent IM form nowadays [64]. The TWIM cell consists of successive stacked ring electrodes. Adjacent rings have opposite phases of radio frequency voltage applied, thus confining the analyte ions radially while allowing them to pass unhindered along the axis. A Direct Current (DC) voltage can be applied to a pair of adjacent rings, to produce a potential barrier that the ions cannot cross. As the DC potential is shifted to an adjacent set of rings the ion barrier moves forward, causing any ions in front of it to be propelled forwards. Stepping the ion barrier sequentially along the rings of the device to the other creates a travelling wave that drives the ions through the device.

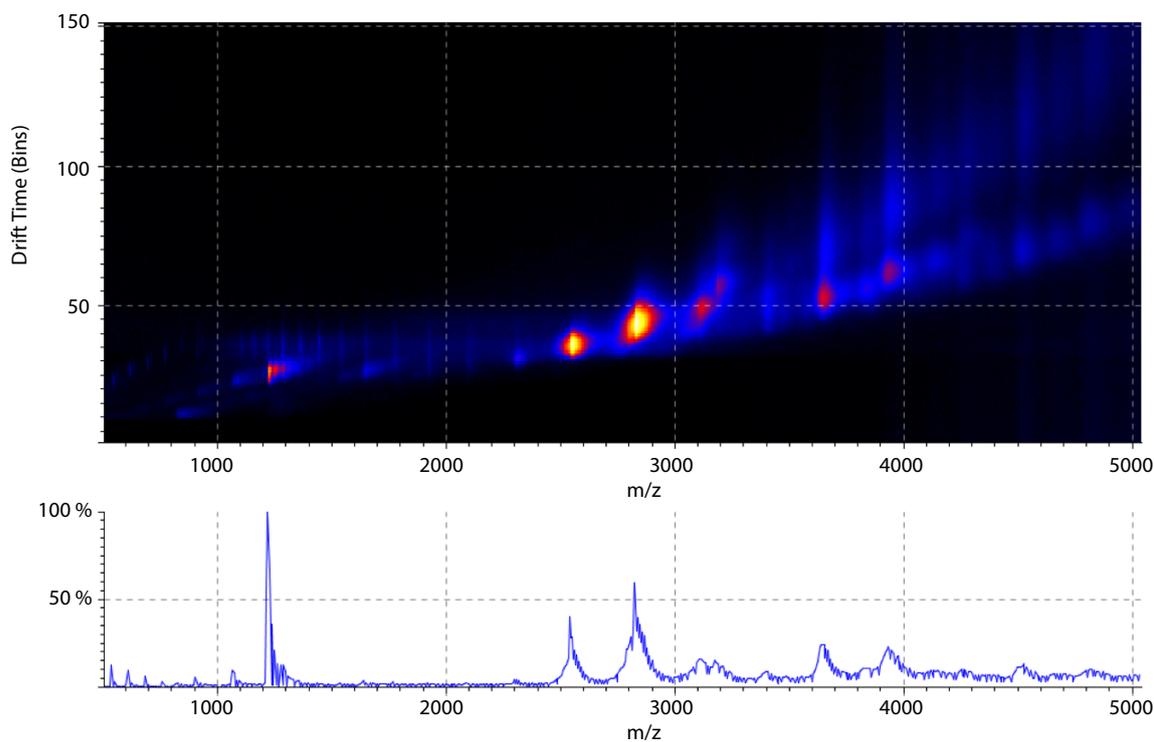


Figure 3.5: Example of a three dimensional driftscope representation (*upper panel*) showing the m/z ratio versus the drift-time. The intensity of the ions in color-encoded. Based on the m/z -ratio ions can be identified (*lower panel*). Different drift times for ions from different species but with identical mass can be distinguished in the driftscope representation

3.3.2 Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) is a powerful and fast method to study the physical basis of molecular interactions. The ability to obtain a complete thermodynamic characterisation from an interaction from a single ITC experiment makes this technique very popular for exploring interactions of biological systems. Accessible affinities range from the low millimolar range for weak binders to sub-nanomolar interactions.

A typical experiment involves the addition of one binding partner (titrant) into the other (titrate) over time applying several individual injections. When substances interact the binding energy in form of heat is either absorbed (endothermic reaction) or released (exothermic reaction). The heat produced during each injection is proportional to the amount of complex formed. Therefore maximal enthalpies are expected at the beginning of a titration decreasing in intensity as the free titrate is consumed. The heat absorbed/emitted of each injection (q_i) can be obtained by integrating the peak and relating it to the fractional saturation of the binding reaction (F) based on the stoichiometry (n), the total concentration of the titrate (M_T), the binding enthalpy (ΔH) and the cell volume (V) with known concentration of the titrant (B_T).

$$q_i = nFM_T\Delta HV \quad (3.2)$$

$$F^2 - F\left\{1 + \frac{B_T}{nM_T} + \frac{1}{nK_B M_T}\right\} + \left\{\frac{B_T}{nM_T}\right\} = 0 \quad (3.3)$$

From equation (3.2) we can solve for F and substitute into equation (3.3) to obtain a relationship which can be solved for q_i . With the knowledge of B_T one can fit the solution to a nonlinear least squares analysis to solve for the thermodynamic parameters like ΔH , K_B and n , with K_B being the binding constant of the interaction. K_B is used to obtain the Gibbs free energy (eq. 3.4) and the entropy ΔS can be calculated from eq. 3.5 with the enthalpy ΔH which is measured directly:

$$\Delta G = -RT\ln(K_B) \quad (3.4)$$

$$\Delta G = \Delta H - T\Delta S \quad (3.5)$$

A schematic diagram of a typical calorimeter is shown in Figure 4.7. Two equivalent lollipop shaped cells are suspended in an adiabatic jacket. The temperature of each cell is monitored and maintained at a constant temperature through an electronic feedback loop that controls thermoelectric heaters located adjacent to each cell. During the experiment the titrant, which is in the injection syringe, is added to the sample cell. If the reaction is exothermic, heat evolves and ΔT_1 increases. The feedback loop

responds by reducing the power to the resistive heater around the sample cell to restore ΔT_1 to zero. Thus, in a power versus time curve, the injection results in a negative deflection from the baseline, the integrated area of which corresponds to the total enthalpy released as a function of the injection. An endothermic reaction would result in a corresponding positive peak. At saturation when no reaction is taking place, the baseline reflects a constant power consumption proportional to the power needed to maintain the temperature differential between the cell and the adiabatic jacket (ΔT_2).

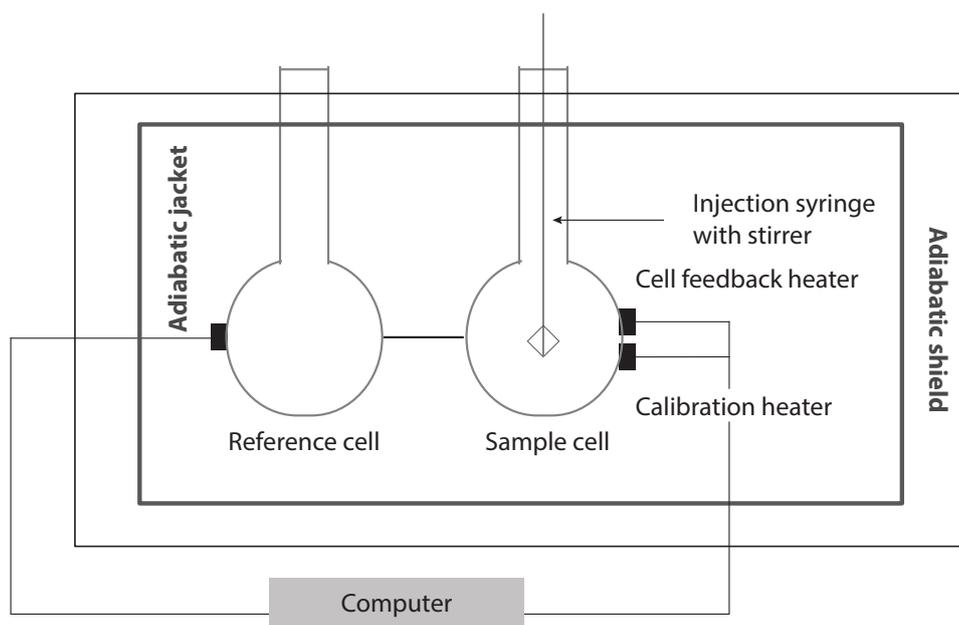


Figure 3.6: A schematic diagram of a typical energy compensation isothermal titration calorimeter

When planning an ITC experiment, one very important step is to decide which ratio titrant to titrate concentration will be used. The c value, a measure of the sigmoidicity of the curve can be calculated as $n \times K_B \times M_T$ [65]. Ideally c should be $10 < c < 100$. While large c values result in very sharp transitions with a possible saturation within one single titration point, low c values yield very broad transitions that approach linearity [66].

3.4 X-ray Crystallography

3.4.1 Protein crystallization

Protein crystallization is the process of formation of protein crystals. Proteins, like many other molecules, can be prompted to form crystals when their solution becomes supersaturated. Under these conditions, individual protein molecules can pack in a repeating array, held together by non covalent interactions. The goal of crystallization is to produce a well-ordered crystal. Many factors influence the likelihood of crystallization of a protein, for instance: protein purity, pH of the buffer solution, concentration of the protein, temperature, precipitants and additives in the crystallization condition solution. The more homogeneous the protein solution is, the more likely crystallization will occur. Typically, protein samples above 97% purity are considered suitable for crystallization, although high purity is not a guarantee for crystallization. The pH of the solution is very important and depending on the pH different packing orientations of the protein may result. Precipitants in the crystallization condition, such as ammonium sulfate or polyethylene glycol, are generally used to promote the formation of protein crystals. The most commonly employed method for protein crystallization is vapor diffusion. For initial screening trials a droplet of ~ 100 nL containing purified protein is mixed with the crystallization condition and are allowed to equilibrate with a larger reservoir containing only the crystallization condition. Initially, the droplet with the protein solution contains, compared to the reservoir, a lower precipitant concentration, but as the drop and the reservoir equilibrate, the precipitant concentration increases in the drop. If the appropriate crystallization condition is used for a given protein, crystallization will occur in the drop. Since this method allows for gentle and gradual changes in concentrations, it aids in the growth of large and well-ordered crystals.

3.4.2 Data collection

Protein crystals are mounted on a loop and subsequently flash-frozen with liquid nitrogen. This freezing reduces radiation damage of the X-rays, as well as noise in the Bragg peaks due to thermal motion (Debye-Waller effect). Since untreated protein crystals often crack when flash-freezing, pre-soaking in a cryoprotectant solution before freezing is performed. The pre-soaking, however, may also cause the crystal to crack. Therefore successful cryo-conditions are identified by trial and error.

The loop with the frozen crystal is mounted on a goniometer, which allows to position and rotate the crystal accurately in the X-ray beam. With the aid of a camera focused on the crystal, it can be centered in the beam with high accuracy. The mounted crystal is irradiated with a beam of monochromatic X-rays. Using a synchrotron as an X-ray source provides high brilliance, polarization and collimation. Furthermore the wide

tunability of the X-ray wavelength is useful for multi-wavelength anomalous dispersion phasing.

A diffraction pattern of the crystal is obtained by measuring the intensity of scattered waves as a function of the scattering angle. Constructive interference occurs when by irradiation with X-rays Bragg's law (Equation 3.6) is satisfied:

$$n\lambda = 2d_{hkl} \sin \Theta \quad (3.6)$$

where n is the order of reflection, λ the wavelength of the incident X-ray beam, d_{hkl} is the spacing between the crystal planes and Θ is the angle between the incident ray and the scattering planes. The relative intensities of these spots provide the information to determine the arrangement of molecules within the crystal in atomic detail. The spots at small angles correspond to low-resolution data, those at high angles represent high-resolution data. Therefore an upper limit off the resolution of the data can be estimated from the first few images. One image of reflections is insufficient to reconstruct the energy density as it represents only one cross-section of the crystal. To collect all the necessary information the crystal must be rotated incrementally acquiring an image at each step. Depending on the space group and therefore on the symmetry of the crystal, a minimum angular range needs to be recorded in order to guarantee completeness of the dataset.

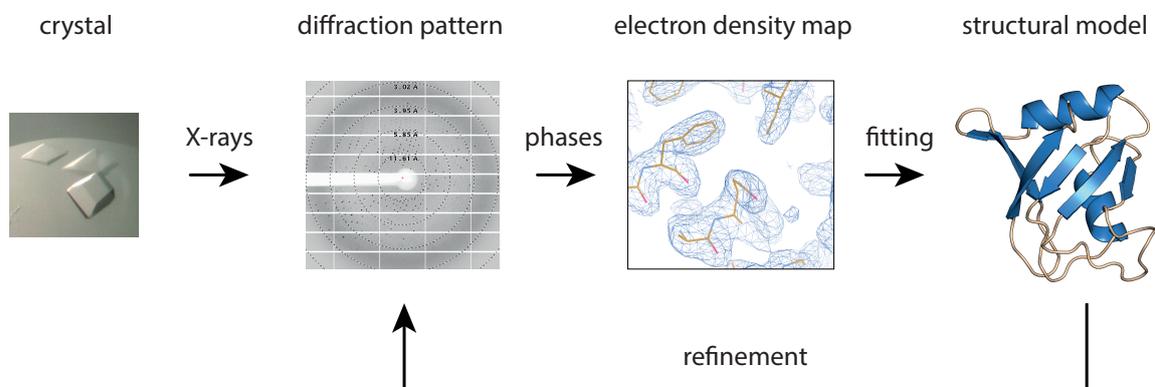


Figure 3.7: Basic workflow for solving a protein structure by X-ray crystallography.

3.4.3 Data analysis

The data set which consists of two-dimensional diffraction patterns, corresponding different crystal orientations, can be converted into a three-dimensional model of the electron density. For this purpose Fourier transformation is applied. The first step of data processing is indexing the reflections, i.e. identifying the size of the unit cell and the correlation of each image peak to its position in reciprocal space. Furthermore

the symmetry of the crystal, its space group, is determined. Indexing is generally accomplished using an auto indexing routine. Once assigned the symmetry, a single file containing the information of the Miller index and intensity of each reflection is generated. Since a full data set consist of hundreds of images taken at different orientations of the crystal, in order to combine the single images they need to be first merged, i.e. to identify which reflections are detected in two or more images, and scaled in order to obtain a overall consistent intensity scale. The quantities R_{sym} and R_{merge} are used to describe the internal agreement of measurements in a crystallographic data set.

The data acquired only contains information of the amplitude of the incident wave, information about its phase is systematically lost. In X-ray crystallography, the information about amplitude of the 3D Fourier transform of the molecule's electron density in the unit cell is provided by the acquired data. If the phases were known, the electron density could be calculated by Fourier synthesis. The bottleneck of determining a structure from X-ray data is solving the phase problem. There are several ways of recovering the lost phase information. Initial phase information can be obtained by : ***ab initio* phasing or direct methods** are usually only applicable for high resolution data and/or small molecules; **anomalous dispersion** (MAD (*Multi-wavelength anomalous dispersion*) or SAD(*Single wavelength anomalous dispersion*) phasing) requires the incorporation of anomalous scattering atoms in the protein, such as seleno-methionine derivatives; anomalous scattering from heavy atoms results in effects on the diffraction pattern, i.e. the reflections are not centrosymmetric anymore resulting in different intensities of Friedel pairs (h,k,l) and $(\bar{h},\bar{k},\bar{l})$. Therefore they can be distinguished from the rest. The positions of the few heavy atoms is determined with direct methods. Based on the heavy atom substructure, the phase of the the whole structure can be determined; **isomorphus replacement**, like for anomalous dispersion in order to apply SIR(*single isomorpous replacement*) or MIR (*multiple isomorpous replacement*) heavy atoms are introduced in the crystal either by soaking or co-crystallization. The reflections from these electron-dense metal atoms can be distinguished and used to determine their location and to obtain initial phases by direct methods. In contrast to anomalous dispersion the diffraction patterns of the native and one (SIR) or several (MIR) derivative crystal are compared; **molecular replacement** is the method of choice when a homologous structure is known, which is used as a search model. Here the Patterson map, an interatomic vector map, of the unknown structure and the search model are compared. By rotation and translation the correlation between the two maps is maximized. This optimized search model is used to derive initial phases from which an electron density map can be calculated.

Based on the initial phases an initial model can be built. This model is refined to fit the observed diffraction data, ideally yielding a better set of phases. The refinement cycle continues until no further improvements of the correlation between the diffraction data and the model can be achieved. The agreement between the crystallographic model

and the experimental X-ray diffraction data is judged by the R-factor:

$$R = \frac{\sum |F_{obs} - F_{calc}|}{\sum |F_{obs}|} \quad (3.7)$$

where F is the structure factor, which is proportional to the intensity of a reflection $I_{hkl} \propto |F(hkl)|^2$. The range of values for R is zero to one, with a large R -factor indicating a poorly modeled solution. A similar quality criterion is R_{free} , which is calculated the same way as the conventional R factor but only from a subset, $\sim 5\%$ of reflections that were not included in the structure refinement. The purpose is to validate the refinement process and to check that the R factor is not artificially reduced by the introduction of too many parameters. For a good model which is not over-interpreting the data, the R -free will be similar to the R -value.

3.5 NMR Spectroscopy in Structural Biology

The principle of Nuclear Magnetic Resonance (NMR) is based on the interaction of a nuclear Spin with an external magnetic field. The first experiment detecting magnetic resonance was carried out by I.I.Rabi in 1937 at the Columbia University (USA). Based on the development of quantum mechanics at the beginning of the 20th century, the discovery of the spin angular momentum property of the nucleus and the quantization of energy proposed by Max Planck (1900), Rabi designed *The molecular-beam magnetic-resonance detection method* [67]. He predicted that the magnetic moments of nuclei could be induced to flip their magnetic orientation if they absorbed energy from an electromagnetic wave of a specific frequency emitting the same amount of energy by the transition to the lower energy state. The experiments involved passing a beam of lithium chloride molecules through a vacuum chamber and diverting the beam with two oppositely poled magnetic fields. In an unperturbed system, the diversion of the beam would be cancelled. To provoke the magnetic moments of the nuclei to flip, the molecular beam is irradiated with radio waves. By adjusting the external magnetic field and the radio frequency, magnetic resonance absorption is observed. Since each atom or molecule has a characteristic pattern of resonance frequencies, Rabi realized he could detect a series of resonances in different molecules that could be used to identify the type of atom or molecule and ultimately giving more detail into molecular structure. The foundation of modern NMR spectroscopy was laid in 1946 when Edward Purcell [68] and Felix Bloch [69] and co-workers independently studied the magnetic resonance properties of atoms and molecules in solids and liquids, instead of individual atoms or molecules as in Rabi's molecular beam method. In the following chapter the basic theory of NMR is presented. Besides the explanation of the NMR phenomenon a mathematical description by the classical vector model as well as the product operator formalism is provided.

3.5.1 Basic principles of NMR

The magnetic resonance phenomenon occurs as a result of the magnetic angular momentum μ intrinsic to a number of different nuclei depending on their composition of proton and neutrons.

$$\mu = \gamma I \hbar \quad (3.8)$$

The gyromagnetic ratio γ is the proportionality constant which determines the resonance frequency of a nucleus for a given external magnetic field. The nuclear spin I can be zero or have integral or fractional values. Only spins with non-zero nuclear spin can be detected by NMR spectroscopy. Note that common biological nuclei such as ¹²C or ¹⁴N are not NMR active. Labeling samples with stable isotopes like ¹³C or ¹⁵N renders the detection by NMR of these nuclei accessible. ¹H, ¹³C, ¹⁵N are the most common NMR active nuclei studied by biomolecular NMR, but also other atoms like ³¹P

and ^{19}F can be observed. Magnetic resonance is occurring in these isotopes with spin $I = \frac{1}{2}$ when introduced in an external magnetic field B_0 . The so called *Zeeman-Effect* describes the splitting of energy levels of the particles according to the orientation of their magnetic angular momentum parallel or antiparallel to the B_0 -field. The energy difference between the Zeeman levels is :

$$\Delta E = -\gamma\hbar B_0 \quad (3.9)$$

Using Planck's law $E = h\nu$ the resonance frequency or *Larmor*-frequency ω_L for the transition between the two Zeeman-levels is:

$$\omega_L = -2\pi\nu = -\gamma B_0 \quad (3.10)$$

In thermal equilibrium the difference in population of the two energy levels $|\alpha\rangle$ (spin parallel to the external magnetic field) and $|\beta\rangle$ (spin antiparallel to the external magnetic field) is described by the *Boltzmann* distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{k_B T}} = e^{-\frac{\gamma\hbar B_0}{k_B T}} \quad (3.11)$$

At room temperature the deviation from equally populated energy levels is very small ($\frac{N_\beta}{N_\alpha} \approx 0.99993$ for ^1H at $B_0 = 14.1$ T). Thus this surplus results in a net macroscopically observable magnetization of the sample, which is usually referred to as \vec{M} . Its z-component M_z parallel to the external magnetic field B_0 .

The vector model

In thermal equilibrium \vec{B}_0 and \vec{M} are parallel, e.g. $\vec{M} = (0, 0, M_z)$. In order to obtain an NMR signal, an additional magnetic field \vec{B}_1 oscillating with the resonance frequency of the system is applied perpendicular to the external field \vec{B}_0 . Due to this perturbation the magnetization \vec{M}_0 is displaced from the thermal equilibrium. Using a high frequency resonator a radio frequency (RF) pulse is emitted generating the magnetic field \vec{B}_1 at the sample. The resulting magnetic field is the superimposition of $\vec{B} = \vec{B}_0 + \vec{B}_1$. The bulk magnetization \vec{M} of the sample is experiencing a torque proportional to $\vec{M} \times \vec{B}$, the magnetization is turned towards the xy-plane and is subsequently precessing around \vec{B}_0 . The angle is adjusted by the the duration and power of the RF-pulse.

$$\alpha = \gamma \int_t \vec{B}_1 dt \quad (3.12)$$

The direction of the rotation follows the right-hand-rule known from basic physics of electromagnetism. A pulse turning \vec{M} into the xy-plane is called 90° -pulse. After the pulse the transversal magnetization is oscillating with the resonance frequency

inducing an electric current in the receiver coils. The signal detected is called the *Free Induction Decay (FID)*, which is a time-dependent exponentially decaying function. As the system is out of equilibrium due to its excitation with the RF pulse, it will seek to relax back to its low energy state.

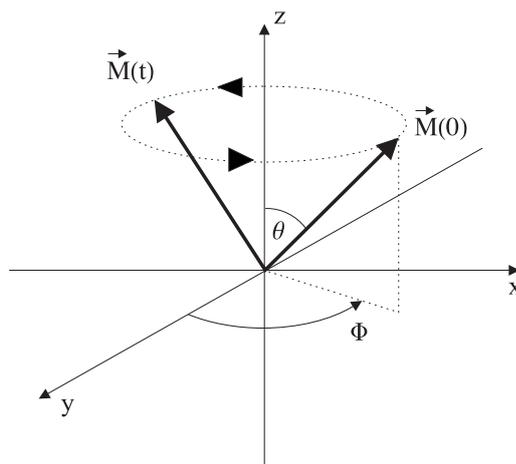


Figure 3.8: Precession of macroscopic magnetization vector in a constant magnetic field

Fourier Transform and NMR

The recorded time dependent FID is related to the frequency domain spectrum through *Fourier Transformation (FT)* as follows:

$$f(\omega) = \int_{-\infty}^{+\infty} f(t)e^{i\omega t} dt \quad (3.13)$$

With this mathematical transformation the convoluted FID is transformed to a frequency spectrum. Fourier transformation is based on the possibility to describe any periodic function as a sum of sine and cosine functions. By applying FT the unknown frequencies included in the time dependent oscillating FID can be extracted and further evaluated. FT NMR allows the addition of consecutive scans taking advantage of the fact that whilst the NMR signal is directly proportional to the number of scans N , the noise only proportional to \sqrt{N} . Therefore adding several scans will result in an increase of the signal-to-noise ratio.

3.5.2 Chemical Shift and J-coupling

As already previously mentioned the gyromagnetic ratio of a certain nucleus is determining its *Larmor*-frequency, for example in a 14.1 Tesla magnetic field ^1H precesses with $\omega_L = 600$ MHz, ^{13}C with 150MHz and ^{15}N with 60MHz. The energy difference between the two spin states of a certain nucleus is furthermore slightly altered depending on the local magnetic field. Moving electrons, which are surrounding the nuclei, are inducing a local magnetic field at the nucleus, which opposes the applied external field. Hence the effective field at the nucleus is:

$$B_{eff} = B_0(1 - \sigma) \quad (3.14)$$

The nucleus is said to be shielded by the moving electric charges of the electrons. The extent of shielding is given by the chemical shift tensor σ , which can be anisotropic. The *Larmor*-frequency of a certain nucleus shielded by a electron cloud is therefore:

$$\omega = \frac{\gamma}{2\pi} B_0(1 - \sigma) \quad (3.15)$$

The variation of the spin resonance frequency of the same type of nucleus due to the variations in the local chemical environment is called the chemical shift (CS). The chemical shift is generally relatively small and defined as δ in parts per million:

$$\delta = \frac{\omega - \omega_L}{\omega_L} \cdot 10^6 \quad (3.16)$$

where ω_L is the *Larmor*-frequency and ω is the resonant frequency of the nucleus of interest.

Another basic concept of NMR spectroscopy is called *J-coupling*, *spin-spin-coupling* or *scalar coupling*. It occurs in nuclear spins, which are connected via one or several chemical bonds. The coupling results from the interactions of the involved nuclear spins with the electron spins, which form the chemical bond. Below the basic concept is illustrated by a simple example:

Two nuclear spins which are connected through a chemical bond formed by two electrons are coupled. Assuming the electron spins obey the *Pauli exclusion principle* they will adopt an “up-down“ or “down-up“ state. Due to the magnetic hyperfine interaction between nuclear spins with their adjacent electron, the nuclear spins can adopt two different states relative to each other, either parallel or antiparallel. At parallel orientation of the nuclear spins, also one of the electron spins is always parallel aligned leading to a slight energy gain of this state. Antiparallel nuclear spin orientation implies antiparallel electron spin alignment resulting in a slight lowering of the energy level. This small energy difference of the above described states is detected as multiplet splitting of the J-coupled nuclear spins (order of magnitude: 1-100 Hz). In general

the multiplicity is given by $m = 2nI + 1$, where I is the spin quantum number of the nuclear spin and n is the number of identical nuclei. The spacing between the lines in a multiplet equals the coupling constant. The intensity of each line in the multiplet is depending of the number of possible combinations of individual spins and is following a binomial distribution which can be calculated by Pascal's triangle. For the two spin example we would obtain a doublet with two equal intensity signals. The magnitude of J provides information on the proximity of the coupling partners. Spin-spin coupling does not only occur in homonuclear spin systems, but also in heteronuclear ones, thus allowing to transfer magnetization between coupled nuclei.

Product operator formalism

The above presented vector model describes processes during an NMR experiment macroscopically. There are intrinsic limitations in representing RF pulses using vectors, because in high resolution NMR spectroscopy effects, like J-coupling, are visible, which cannot be described via the classical vector model. For this purpose the product operator formalism was developed.

Density operator and macroscopic magnetization

As a good approximation a sample in an NMR experiment can be described as an ensemble of Spin- $\frac{1}{2}$ -particles, which do not interact. The state of each spin $|\Psi\rangle = c_\alpha |\alpha\rangle + c_\beta |\beta\rangle$ is generally represented by the state vector

$$|\Psi\rangle = \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix} \quad (3.17)$$

where c_α and c_β are superposition coefficients. The two *Zeeman* eigenstates are represented as

$$|\alpha\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |\beta\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (3.18)$$

The expectation value of an operator \hat{Q} can be written as

$$\langle \hat{Q} \rangle = \langle \Psi | \hat{Q} | \Psi \rangle = \text{Tr}\{ |\Psi\rangle \langle \Psi | \hat{Q} \} \quad (3.19)$$

This representation offers various advantages, especially when observing a system with multiple spins. The states of those spins are $|\Psi_1\rangle, |\Psi_2\rangle, \dots, |\Psi_n\rangle$. The result of a single measurement of the expectation value cannot be predicted due to the uncertainty relation of quantum mechanics, but the average of various measurements:

$$Q_{obs}(\text{average}) = \langle \Psi_1 | \hat{Q} | \Psi_1 \rangle + \dots + \langle \Psi_n | \hat{Q} | \Psi_n \rangle \quad (3.20)$$

$$Q_{obs}(\text{average}) = \text{Tr}\{(|\Psi_1\rangle\langle\Psi_1| + \dots + |\Psi_n\rangle\langle\Psi_n|)\hat{Q}\}$$

Another operator $\hat{\rho}$, which is called density operator, is defined as following:

$$\hat{\rho} = \frac{1}{N}(|\Psi_1\rangle\langle\Psi_1| + \dots + |\Psi_n\rangle\langle\Psi_n|) = \overline{|\Psi\rangle\langle\Psi|} \quad (3.21)$$

where N is the number of spins in the ensemble. With this the macroscopic measurement for Q for the entire ensemble as a good approximation results in

$$\frac{1}{N}Q_{\text{macro}} \cong \text{Tr}\{\hat{\rho}\hat{Q}\} = \langle Q \rangle. \quad (3.22)$$

This equation means that the result of each macroscopic measurement on the described spin-system can be calculated from two operators ($\hat{\rho}$ and \hat{Q}) and namely independent from the actual number of spins. This greatly simplifies the problem; instead of the summation of 10^{22} spin states, the indication of the density operator $\hat{\rho}$ is sufficient. The result of the measurement equals the expectation value of \hat{Q} . For an ensemble of non interacting Spin- $\frac{1}{2}$ -particles the matrix representation of the density operator reads as follows:

$$\hat{\rho} = \begin{pmatrix} \rho_{\alpha\alpha} & \rho_{\alpha\beta} \\ \rho_{\beta\alpha} & \rho_{\beta\beta} \end{pmatrix} = \begin{pmatrix} \overline{c_\alpha c_\alpha^*} & \overline{c_\alpha c_\beta^*} \\ \overline{c_\beta c_\alpha^*} & \overline{c_\beta c_\beta^*} \end{pmatrix} = \begin{pmatrix} \rho_\alpha & \rho_+ \\ \rho_- & \rho_\beta \end{pmatrix} \quad (3.23)$$

The diagonal entries ρ_α and ρ_β are the *populations* of the states $|\alpha\rangle$ and $|\beta\rangle$. The anti diagonal entries ρ_+ and ρ_- are the *coherences* between the states $|\alpha\rangle$ and $|\beta\rangle$. Only the coherences (more precisely the single-quantum-coherences) give rise to detectable NMR signal.

As previously mentioned, the macroscopic magnetization of the sample arises from the difference in population of the state $|\alpha\rangle$ and $|\beta\rangle$ due to the Boltzmann distribution. The magnetization vector $\vec{M} = (M_x, M_y, M_z)$ can be determined by the components of the density matrix $\hat{\rho}$:

$$\begin{aligned} M_z &= \frac{2}{\mathfrak{B}}(\rho_\alpha - \rho_\beta) \\ M_x &= \frac{4}{\mathfrak{B}}\text{Re}\{\rho_-\} \\ M_y &= \frac{4}{\mathfrak{B}}\text{Im}\{\rho_-\} \end{aligned} \quad (3.24)$$

with the Boltzmann-factor $\mathfrak{B} = \frac{\hbar\gamma B_0}{k_B T}$.

Vice versa the populations and coherences (and therefore $\hat{\rho}$) can be calculated from the magnetization vector \vec{M} :

$$\rho_\alpha = \frac{1}{2} + \frac{1}{4}\mathfrak{B}M_z \quad (3.25)$$

$$\begin{aligned}\rho_{\beta} &= \frac{1}{2} - \frac{1}{4}\mathfrak{B}M_z \\ \rho_{+} &= \frac{1}{4}\mathfrak{B}(M_x - iM_y) \\ \rho_{-} &= \frac{1}{4}\mathfrak{B}(M_x + iM_y)\end{aligned}$$

Calculation rules for the product operator formalism

Using the density operator it is possible to completely describe the evolution of a spin system. For this purpose matrix calculations need to be performed which are rather extensive and complex. For the practical design of new NMR experiments and the prediction of their results, the product operator formalism has been developed. It applies to spin- $\frac{1}{2}$ -particles.

In the product operator formalism the spin system's density operator is expressed as a linear combination of products of cartesian basis set of operators, the identity matrix \hat{E} and the Pauli spin matrices \hat{I}_x, \hat{I}_y and \hat{I}_z :

$$\hat{\rho}(t) = a(t)\hat{E} + b_x(t)\hat{I}_x + b_y(t)\hat{I}_y + b_z(t)\hat{I}_z \quad (3.26)$$

with

$$\begin{aligned}\hat{E} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \hat{I}_x &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \hat{I}_y = \frac{i}{2} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \hat{I}_z = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\end{aligned}$$

The time evolution of the density operator disregarding relaxation is described by the *Liouville-von Neumann* equation:

$$\frac{d\hat{\rho}(t)}{dt} = -i[\mathcal{H}, \hat{\rho}(t)] \quad (3.27)$$

\mathcal{H} is the Hamiltonian operator, which includes coupling terms, chemical shift terms, etc. Assuming its piecewise time independence equation 3.27 can be written as

$$\hat{\rho}(t) = e^{-i\mathcal{H}t}\hat{\rho}(0)e^{i\mathcal{H}t} \quad (3.28)$$

or

$$\hat{\rho}^1 \xrightarrow{\mathcal{H}t} \hat{\rho}^2 \quad (3.29)$$

Here the transformation of the initial density operator $\hat{\rho}^1$ in a new operator $\hat{\rho}^2$ is described as a rotation under the influence of the respective Hamilton operator \mathcal{H} .

In NMR spectroscopy the Hamiltonian comprises three relevant components: RF

pulse, chemical shift and J-coupling. The spin-system's relaxation processes towards the thermal equilibrium are neglected for the time being. Note that the contributions of chemical shift and j-coupling to the Hamilton operator commute. For each of the three contributions to the Hamilton operator simple calculation rules can be established in order to solve the *Liouville-von Neumann* equation :

RF pulse with angle α

On resonant RF pulses along a polarization direction x or y result in a rotation of the spin system around the x- or y-axis, respectively. The corresponding Hamilton operator denotes as $\mathcal{H}t = \alpha I_x$ for a pulse with polarization direction x and torsion angle α .

$$\begin{aligned} I_x &\xrightarrow{\alpha I_{\pm x}} I_x \\ I_y &\xrightarrow{\alpha I_{\pm x}} I_y \cos \alpha \pm I_z \sin \alpha \\ I_z &\xrightarrow{\alpha I_{\pm x}} I_z \cos \alpha \pm I_y \sin \alpha \end{aligned}$$

$$\begin{aligned} I_x &\xrightarrow{\alpha I_{\pm y}} I_x \cos \alpha \pm I_z \sin \alpha \\ I_y &\xrightarrow{\alpha I_{\pm y}} I_y \\ I_z &\xrightarrow{\alpha I_{\pm y}} I_z \cos \alpha \pm I_x \sin \alpha \end{aligned}$$

Chemical shift

During free precession of spin I the Hamiltonian of the chemical shift is denoted as $\mathcal{H} = \Omega_I I_z$, where Ω_I is the offset of spin I in the spectrum. The evolution of the chemical shift during time t is described as follows:

$$\begin{aligned} I_x &\xrightarrow{\Omega_I I_z t} I_x \cos(\Omega_I t) + I_y \sin(\Omega_I t) \\ I_y &\xrightarrow{\Omega_I I_z t} I_y \cos(\Omega_I t) + I_x \sin(\Omega_I t) \\ I_z &\xrightarrow{\Omega_I I_z t} I_z \end{aligned}$$

J-coupling

In a weakly coupled two-spin-system with spins I and S and the scalar coupling constant J_{IS} the Hamilton operator is adopting the following form: $\mathcal{H} = 2\pi J_{IS} I_z S_z$. The evolution of the J-coupling during time t is described

as follows:

$$\begin{aligned}
 I_x &\xrightarrow{2\pi J_{IS}I_z S_z t} I_x \cos(2\pi J_{IS}t) + 2I_y S_z \sin(2\pi J_{IS}t) \\
 I_y &\xrightarrow{2\pi J_{IS}I_z S_z t} I_y \cos(2\pi J_{IS}t) - 2I_x S_z \sin(2\pi J_{IS}t) \\
 I_z &\xrightarrow{2\pi J_{IS}I_z S_z t} I_z \\
 \\
 2I_x S_z &\xrightarrow{2\pi J_{IS}I_z S_z t} 2I_x S_z \cos(2\pi J_{IS}t) + I_y \sin(2\pi J_{IS}t) \\
 2I_y S_z &\xrightarrow{2\pi J_{IS}I_z S_z t} 2I_y S_z \cos(2\pi J_{IS}t) - I_x \sin(2\pi J_{IS}t) \\
 2I_z S_z &\xrightarrow{2\pi J_{IS}I_z S_z t} 2I_z S_z
 \end{aligned}$$

Since the Hamilton operators of the chemical shift and the J-coupling commute, the calculation of complex experiments can be divided in several sections. Below calculations for some basic experiments are shown.

Spin-Echo

Spin-echoes are one of the fundamental building blocks for pulse sequences in modern NMR spectroscopy. Polarisationtransfer in heteronuclear spin-systems as well as many multidimensional experiments are derived from this concept.

One-spin-system

After applying a 90_x° pulse the density operator is in the state $-I_y$. During the time τ the chemical shift ΩI_z is acting:

$$-I_y \xrightarrow{\Omega I_z \tau} -I_y \cos(\Omega\tau) + I_x \sin(\Omega\tau) \quad (3.30)$$

Subsequently a 180_x° pulse is applied:

$$\begin{aligned}
 -I_y \cos(\Omega\tau) + I_x \sin(\Omega\tau) &\xrightarrow{\pi I_x} -I_y \cos(\Omega\tau) \cos \pi - I_z \cos(\Omega\tau) \sin \pi + I_x \sin(\Omega\tau) \\
 &\xrightarrow{\pi I_x} I_y \cos(\Omega\tau) + I_x \sin(\Omega\tau)
 \end{aligned} \quad (3.31)$$

During the second waiting time τ the chemical shift ΩI_z is acting again:

$$\begin{aligned}
 I_y \cos(\Omega\tau) + I_x \sin(\Omega\tau) &\xrightarrow{\Omega I_z \tau} I_y \cos(\Omega\tau) \cos(\Omega\tau) - I_x \sin(\Omega\tau) \cos(\Omega\tau) \\
 &\quad + I_x \sin(\Omega\tau) \cos(\Omega\tau) + I_y \sin(\Omega\tau) \sin(\Omega\tau) \\
 &\xrightarrow{\Omega I_z \tau} I_y
 \end{aligned} \quad (3.32)$$

with $\cos^2 \Theta + \sin^2 \Theta = 1$. For the entire pulse sequence we obtain accord-

ingly:

$$I_z \xrightarrow{90_x^\circ - \tau - 180_x^\circ - \tau} I_y \quad (3.33)$$

The result is independent of the chemical shift Ω and the waiting time τ . The chemical shift is refocused by this pulse sequence.

Homonuclear two-spin-system

Given a two-spin-system I, S whose initial state is I_x , it is known from the previous section that a spin echo is refocussing the chemical shift of a one-spin system. Hence it is sufficient to consider only the term of J-coupling. During the first waiting time it evolves as follows:

$$I_x \xrightarrow{2\pi J_{IS} I_z S_z \tau} I_x \cos(\pi J_{IS} \tau) + 2I_y S_z \cos(\pi J_{IS} \tau)$$

The 180_x° pulse is acting on both spins:

$$\begin{aligned} I_x \cos(\pi J_{IS} \tau) + 2I_y S_z \cos(\pi J_{IS} \tau) &\xrightarrow{\pi I_x} I_x \cos(\pi J_{IS} \tau) - 2I_y S_z \cos(\pi J_{IS} \tau) \\ I_x \cos(\pi J_{IS} \tau) - 2I_y S_z \cos(\pi J_{IS} \tau) &\xrightarrow{\pi S_x} I_x \cos(\pi J_{IS} \tau) + 2I_y S_z \cos(\pi J_{IS} \tau) \end{aligned}$$

The 180_x° pulse didn't effect the spin system and can be ergo neglected for further calculations. Consequently only the evolution of the J-coupling during the time 2τ is relevant:

$$I_x \xrightarrow{\tau - 180_x^\circ - \tau} I_x \cos(\pi J_{IS} \tau) + 2I_y S_z \cos(\pi J_{IS} \tau)$$

Using $\tau = \frac{1}{4J_{IS}}$ a complete interconversion of in-phase magnetization to anti-phase magnetization is achieved. Note that the pulse sequence $-\tau - 180_x^\circ(I, S) - \tau$ is equivalent to $-2\tau - 180_x^\circ(I, S) -$. The chemical shift is refocussed, the J-coupling is acting during the time 2τ .

Heteronuclear two-spin-system

For heteronuclear two-spin systems the 180° pulse can be applied on both spins simultaneously or only on one spin. The results obtained are summed up below:

Polarisationtransfer

An important application of the spin-echo in a heteronuclear two-spin-system is the magnetization transfer from one spin-system to another. The respective experiment is called *Insensitive Nuclei Enhanced by Polarization Transfer (INEPT)* and is an essential building block in NMR spectroscopy experiments. The objective of the *INEPT*-sequence is to transfer magnetization from a nuclear spin with high sensitivity (high

180° pulse applied on	chemical shift Ω_I	J-coupling J_{IS}
I,S	refocussed	active over 2τ
I	refocussed	refocussed
S	active over 2τ	refocussed

gyromagnetic ratio γ_I , usually protons) to a nuclear spin with lower sensitivity (lower gyromagnetic ratio γ_S , e.g. ^{13}C or ^{15}N , taking advantage of the J-coupling between the two spins. The detectable signal from ^{13}C or ^{15}N spins is enhanced by the magnetization transfer. The *INEPT* pulse sequence resembles the heteronuclear spin-echo:

$$\begin{array}{l}
 I \text{ spin: } 90_x^\circ - \tau \quad 180_x^\circ - \tau - 90_y^\circ \\
 S \text{ spin: } \quad \quad \quad 180_x^\circ \quad \quad 90_x^\circ - FID
 \end{array}$$

In addition to the previous calculations, the Boltzmann factor $K_I = \frac{\hbar\omega_I}{4k_B T}$ for the initial magnetization in the thermal equilibrium $K_I I_z$ is considered. Along the lines of the above calculations the state of the spin-systems after the waiting time τ is calculated as follows:

$$K_I I_z \xrightarrow{\frac{\pi}{2} I_x - \tau - \pi(I_x + S_x) - \tau} K_I (I_y \cos(2\pi J_{IS}\tau) - 2I_x S_z \sin(2\pi J_{IS}\tau))$$

Consequently a 90_y° pulse is applied on spin I and a 90_x° pulse is applied on spin S :

$$K_I (I_y \cos(2\pi J_{IS}\tau) - 2I_x S_z \sin(2\pi J_{IS}\tau)) \xrightarrow{\frac{\pi}{2}(I_y + S_x)} K_I (I_y \cos(2\pi J_{IS}\tau) - 2I_z S_y \sin(2\pi J_{IS}\tau))$$

Using $\tau = \frac{1}{4J_{IS}}$ the magnetization from spin S detectable at the end of the *INEPT* sequence is $-\frac{\hbar\omega_I}{4k_B T} 2I_z S_y$. In contrast to this sequence, a direct excitation of spin S with a 90_x° pulse would have yield in the magnetization $-\frac{\hbar\omega_S}{4k_B T} S_y$, which is smaller by a factor $\frac{\omega_I}{\omega_S}$. The signal gain due to the *INEPT* sequence compared to a conventional one is therefore:

$$\frac{\text{INEPT}}{\text{conventional}} = \frac{\gamma_I}{\gamma_S} \quad (3.34)$$

The smaller the gyromagnetic ratio γ_S of spin S , the higher the signal gain.

3.5.3 Two Dimensional NMR Experiments

Signal overlap in the NMR spectrum will occur with increasing molecular size, and will reach the point that the separation of single resonances is impossible. In order to increase the resolution, selection criteria regarding the resonances of interest as well as the addition of a second dimension are applied.

Heteronuclear single quantum spectroscopy: HSQC

The $^1\text{H}, ^{15}\text{N}$ - HSQC is one of the essential two-dimensional NMR-experiments for proteins. It correlates the nitrogen atom of an NH_x -group with its directly attached protons. As the protein backbone consists of amid bonds, every amino acid, except for proline, gives rise to a cross peak signal in the spectrum. Side-chains containing ^{15}N -bound protons (i.e. Glutamine, Asparagine, Lysine, Arginine, Tryptophan and Histidine) can give rise to additional peaks. The HSQC spectrum is often referred to as the "fingerprint" spectrum, since it is unique for every protein. The acquisition of heteronuclear spectra presupposes the protein to be isotopically enriched with the required NMR active nuclei (^{13}C or ^{15}N).

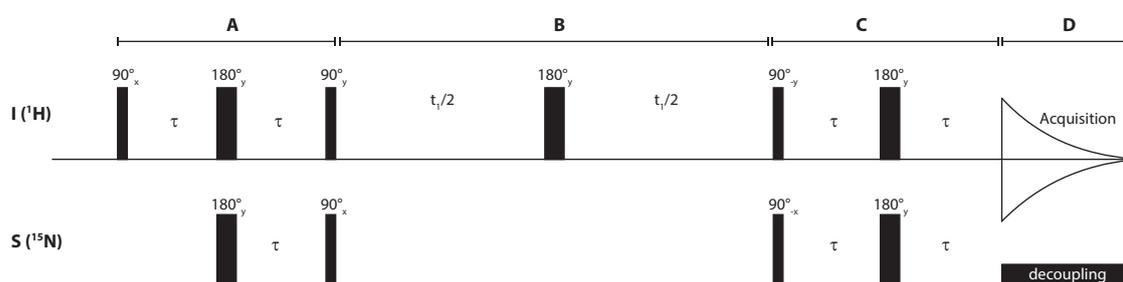


Figure 3.9: Basic HSQC pulse sequence

The pulse sequence illustrated, in Figure (3.9), can be divided in several sections and analyzed by using the product operator formalism. Section A is basically the above described spin-echo with refocussed chemical shift. In the following section B there is no chemical shift evolution for spin I due to the 180°_y pulse. Furthermore also the J-coupling is refocused. The chemical shift of spin S is evolving during the time t_1 . In the next step (C) the back transfer of the magnetization is performed, similar to section A. However, here the second 90° pulse is not applied to leave the magnetization precessing in the transversal plane, in order to induce a signal in the receiver coils of the probe head. During acquisition (D) a decoupling pulse on spin S is applied, therefore no J-coupling is active. Chemical shift remains active and will certainly influence the FID as well as relaxation processes. To sum up the observable magnetization at the

time of the acquisition start is:

$$I_y \cos \Omega_S t_1$$

The result indicates that I_y , which has been generated by the initial 90° pulse, has acquired the frequency/chemical shift of the spin S (Ω_S) during the evolution time t_1 . In other words, the frequency of the spin I is modulated with the frequency of spin S . The experiment is repeated with a stepwise increase of the delay t_1 , the so called evolution period. In order to obtain a 2D frequency spectrum from the recorded time-domain signal, a 2D-*Fourier* transform is performed.

Total correlation spectroscopy (TOCSY)

In a TOCSY experiment, magnetization is transferred over a complete spin-system of an amino acid by successive through-bond scalar coupling. Therefore the TOCSY experiment correlates all protons of a spin system and a characteristic signal pattern of each amino acid is observed. The presence of heteroatoms, such as oxygen, usually disrupts TOCSY transfer.

TOCSY transfers the coherence to other coupled spins through the 3J coupling (through molecular bonds) by isotropic mixing. During the mixing period, the coherence along an axis on xy plane is transferred throughout the spin system under the interaction of scalar coupling, at the same time the chemical shift is suppressed by the inhomogeneity of RF pulses in the isotropic mixing pulse train. Several methods for the isotropic mixing pulse sequences (also known as spin lock sequences) are used in different types of TOCSY experiments. DIPSI series mixing sequences are most commonly used for 1H - 1H TOCSY experiments. During the mixing time, the magnetization of amide protons is transferred to other protons within amino acid residues through three-bond scalar coupling. Since there is no through-bond scalar coupling between the interresidual amide proton H_N and a proton H_α , the magnetization cannot be transferred across the peptide bond. The transfer distance depends on several factors like the efficiency of the spin lock sequence, length of mixing time, coupling constants, and relaxation rate of the molecule. Protons further away from the H_N proton require a longer mixing time to reach. The objective is to optimize the magnetization transfer to reach every proton within the spin system. However, the magnitude of magnetization transfer to neighboring protons decreases quickly as the mixing time exceeds 30 ms. For remote protons, the magnetization transfer is achieved with mixing times as long as 100 ms. The correlations between H_N and H_α and H_β , respectively, can be observed in a TOCSY with a mixing time of 30 ms, whereas a 60 ms or longer mixing time is required to achieve magnetization transfer from H_N to other side chain protons. In this work for the acquisition of 2D Homo-nuclear TOCSY spectra of peptides a mixing time of 30 ms was applied, for the Pin1 WW domain it was set to 50 ms.

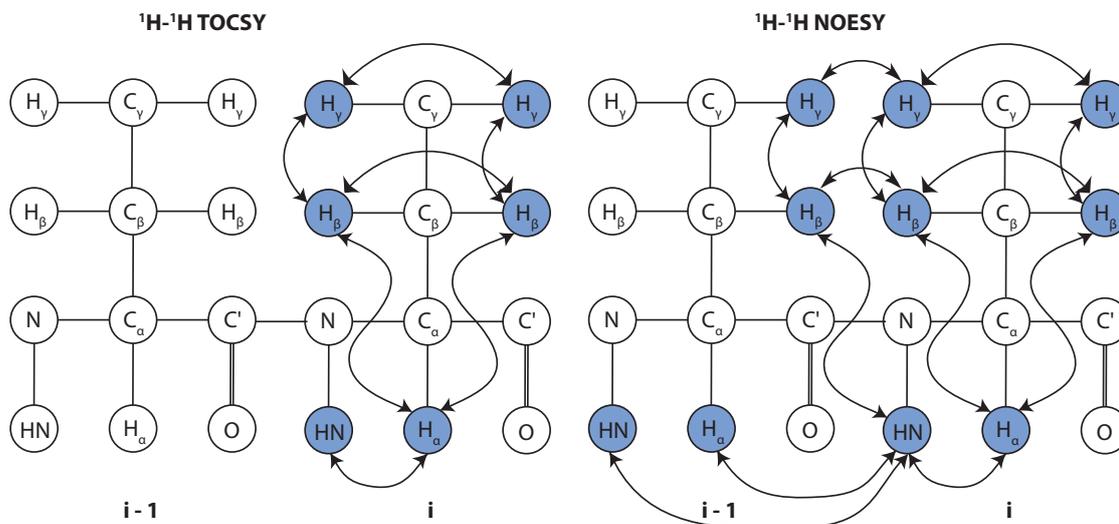


Figure 3.10: Basic scheme of the magnetization transfer for 2D Homonuclear TOCSY and NOESY experiments.

The Nuclear Overhauser Effect - Nuclear Overhauser Effect Spectroscopy (NOESY)

Given a spin system with spin I and spin S , which are coupled via dipolar interactions, a relaxation between those spins, which are not covalently bound but only in close spatial proximity, is observable. Transitions of spin S are saturated by a long RF-pulse with low power (e.g. equal population of spin S is achieved). The pulse is chosen such that spin I is not influenced. Nevertheless, the magnetization of spin I assumes a new value, which can be higher than at thermal equilibrium. Ergo, a saturation of spin S gives rise to a change in the magnetization of spin I . This phenomenon is called *Nuclear Overhauser Effect (NOE)*. The NOE-amplification-factor ϵ_{NOE} for steady state is described by the *Solomon-Equation*:

$$\epsilon_{NOE} = \frac{\gamma_I}{\gamma_S} \frac{W_2 - W_0}{W_0 + 2W_{1S} + W_2} \quad (3.35)$$

where W_0 , W_{1S} and W_2 are the possibilities for a zero quantum transition, a single quantum transition and a double quantum transition, respectively. For the homonuclear case, $\gamma_I = \gamma_S$, the NOE-amplification is up to 50 %. In practice, the NOE-effect allows to measure inter atomic distances by NMR, in general between hydrogen atoms. The dipolar coupling enables the transfer of spin polarization (magnetization along the z -axis), from one nuclei to another, while being modulated with its previous resonance frequency. This transfer allows to correlate a hydrogen with other hydrogens close in space ($\lesssim 5 \text{ \AA}$). In contrast to J-coupled systems, where the coherence transfer is via the molecular bonds, the NOE polarization is transferred through space. The distance

dependence is strong, and the transfer efficiency decreases with r^{-6} , where r is the atomic distance. In conclusion, the NOESY experiment correlates all protons which are close in space, this means it also correlates protons which are distant in the amino acid sequence but close in space due to tertiary structure. This is the most important information For the determination of protein structures this information is essential.

The 2D-Homonuclear-NOESY (nuclear Overhauser effect spectroscopy) pulse sequence can be written as follows:

$$90_x^\circ \longrightarrow t_1 \longrightarrow 90_x^\circ \longrightarrow t_m \longrightarrow 90_x^\circ \longrightarrow \text{acquisition}$$

The first 90_x° pulse generates magnetization in the transverse plane, where it is precessing during the time t_1 . The second 90_x° pulse is rotating the components of the magnetization along the -z-axis. Over the subsequent mixing period t_m , the z-magnetization components exchange under the influence of cross-relaxation due to dipolar coupling. The transverse components are not of interest and are therefore removed by phase-cycling. The third 90_x° returns the z magnetization in the transverse plane and renders it observable. Diagonal peaks are observed for magnetization which doesn't migrate during t_m , cross-peaks arise from magnetization transfer between two spins. The magnetization is exchanged between all hydrogens close in space. This spectrum was used to assign the ^1H - ^1H NOE restraints which are used for structure calculation. For the acquisition of 2D Homo-nuclear NOESY spectra of peptides a mixing time of 300 ms was applied, for the Pin1 WW domain it was set between 120-160 ms.

3.5.4 Three Dimensional NMR Experiments

Although the $^1\text{H},^{15}\text{N}$ - HSQC is one of the first spectra routinely acquired studying a new protein, the obtained signals cannot be correlated to the atom from which they arise without prior knowledge only by means of the HSQC spectrum. For the purpose of obtaining a correlation between the NMR signals and the atoms of the protein of interest, a set of multidimensional heteronuclear NMR experiments based on through bond J-couplings have been developed. Triple resonance experiments allow to correlate the protein backbone amide of each residue with the C_α and C_β chemical shifts of the current and the previous residue. A minimum isotopic labeling of ^{13}C and ^{15}N is required to apply these experiments. The experiments are named according to the atoms involved in the magnetization transfer.

For instance, in the CBCA(CO)NH / HN(CO)CACB experiment. Magnetization is transferred from $^1\text{H}_\alpha$ and $^1\text{H}_\beta$ to $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$, respectively, and then from $^{13}\text{C}_\beta$ to $^{13}\text{C}_\alpha$. From here it is transferred first to ^{13}CO , then to $^{15}\text{N}^H$ and then to $^1\text{H}^N$ for detection. The chemical shift is evolved simultaneously on $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$, so these appear in one dimension. The chemical shifts evolved in the other two dimensions are $^{15}\text{N}^H$ and $^1\text{H}^N$. The chemical shift is not evolved on ^{13}CO . Complementary to the

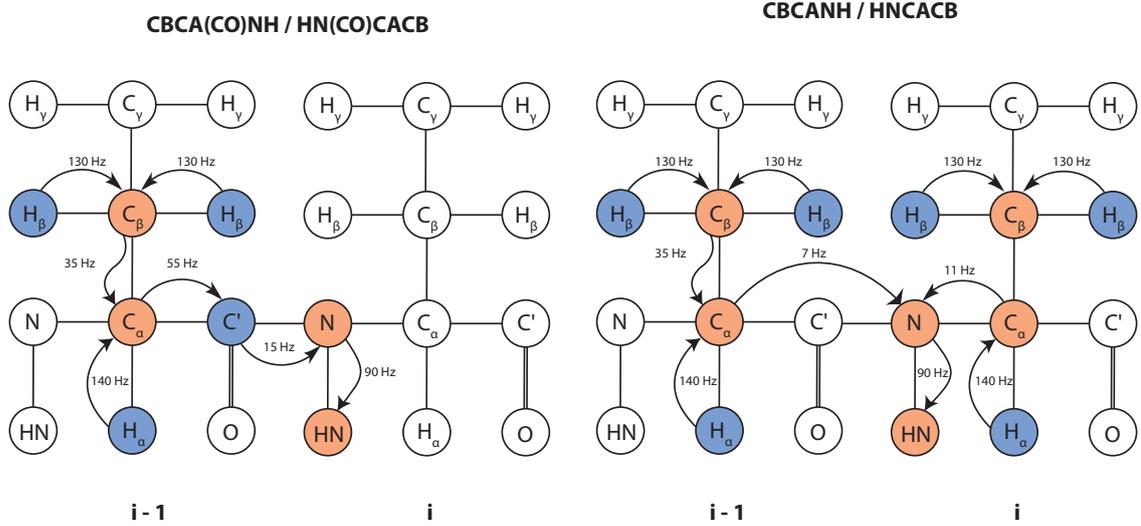


Figure 3.11: J-couplings utilized in backbone assignment experiments

CBCA(CO)NH / HN(CO)CACB experiment, the CBCANH / HNCACB is performed. Here, Magnetization is transferred from $^1\text{H}_\alpha$ and $^1\text{H}_\beta$ to $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$, respectively, and then from $^{13}\text{C}_\beta$ to $^{13}\text{C}_\alpha$. From here it is transferred first to $^{15}\text{N}^H$ and then to $^1\text{H}^N$ for detection. Transfer from $\text{C}_{\alpha_{i-1}}$ can occur both to $^{15}\text{N}_{i-1}$ and $^{15}\text{N}_i$, or from another point of view magnetization is transferred to $^{15}\text{N}_i$ from both $^{13}\text{C}_{\alpha_i}$ and $^{13}\text{C}_{\alpha_{i-1}}$. Thus for each amide group there are two C_α and C_β peaks visible. The objective of this sequential resonance assignment is to connect $\text{C}_{\alpha,\beta_i}$ resonances of residue $_n$ with matching $\text{C}_{\alpha,\beta_{i-1}}$ resonances from residue $_m$. These in this manner connected stretches of spin systems are then assigned to the known primary amino acids sequence of the protein construct. The chemical shift of the C_α and C_β is characteristic for each amino acid, it is, thus, possible to identify each amino acid based on these two carbon resonances.

Based on the assigned C_α and C_β chemical shifts, it is possible to derive information about secondary structure elements as there is a strong correlation between chemical shifts and local structure. The so called *secondary chemical shift* is defined as following:

$$\Delta\delta = \delta_{\text{observed}} - \delta_{\text{randomcoil}} \quad (3.36)$$

The secondary chemical shift is dependent on the secondary structure of a protein. The random coil conformation refers to conformation of a protein where it is a highly dynamic with no secondary or tertiary structural preference. In this state each amino acid type adopts very characteristic chemical shifts values called *random coil chemical shifts*. These shifts are essentially just averages over all possible conformations which the amino acid can adopt in the random coil. Comparing the observed chemical shifts with the random coil values allows to make prediction about the fold a specific amino acid in the protein is likely to adopt. Basically, C_α atoms in α helices will be likely

to have positive secondary chemical shifts, whereas C_α atoms in β -strands will tend so have negative values. For C_β atoms the behavior is exactly contrary: C_β atoms in β -strands will yield positive secondary chemical shift values, whilst being in α -helical conformation will result in negative values.

3.5.5 Ligand binding studies by NMR

As previously described, the chemical shift of an amino acid is very sensitive to its electronic environment. It is not only influenced by the covalent molecular structure, but can be also perturbed by non-covalent interactions with binding ligands. This characteristic makes chemical shifts a very sensitive parameter to monitor for the identification of interaction surfaces in the protein of interest. In general a series of ^1H , ^{15}N HSQC experiments are acquired, where an increasing amount of ligand is added to the protein and the chemical shifts are monitored. Upon addition of the binding partner, those residues, whose electronic environment is changed due to the protein ligand interaction, will experience chemical shift perturbations. Usually the affected residues are close to the interaction surface, but nevertheless it should be emphasized that the perturbation is no proof for a direct interaction with the binding partner. It merely indicates that there is a change in the electronic environment of a residue, which, for instance, could also occur because of structural rearrangement upon complex formation. The chemical shift perturbations observed upon titration depend on the exchange rate between free protein and complex. The exchange constant is $k_{ex} = k_{on}[L] + k_{off}$, where $[L]$ is the ligand concentration, k_{on} and k_{off} are the association and dissociation rates, respectively. On the NMR timescale three different exchange regime are distinguished:

$$\begin{aligned} k_{ex} &\gg \Delta\nu && \text{fast exchange} \\ k_{ex} &\approx \Delta\nu && \text{intermediate exchange} \\ k_{ex} &\ll \Delta\nu && \text{slow exchange} \end{aligned}$$

In fast exchange the chemical shifts change continuously upon titration. Basically the complex only exists for part of the time of acquisition. The observed chemical shift is ergo a weighted average of the chemical shifts for the free protein and the complex until the interaction is saturated. Intermediate exchange regime is observed when resonance peaks intensities become weaker and broader and finally disappear with increasing ligand concentration. At saturation the peaks may reappear at new positions as all protein molecules are forced into the complex form and the chemical shifts observed are those of the complex. High affinity binding interactions will be subject to slow exchange. Here, new resonance peaks will appear in the spectrum and increase in intensity while simultaneously some original resonances will gradually decrease in intensity. For the whole duration of the experiment any protein molecule is either free or in complex, therefore chemical shifts for both states are observed

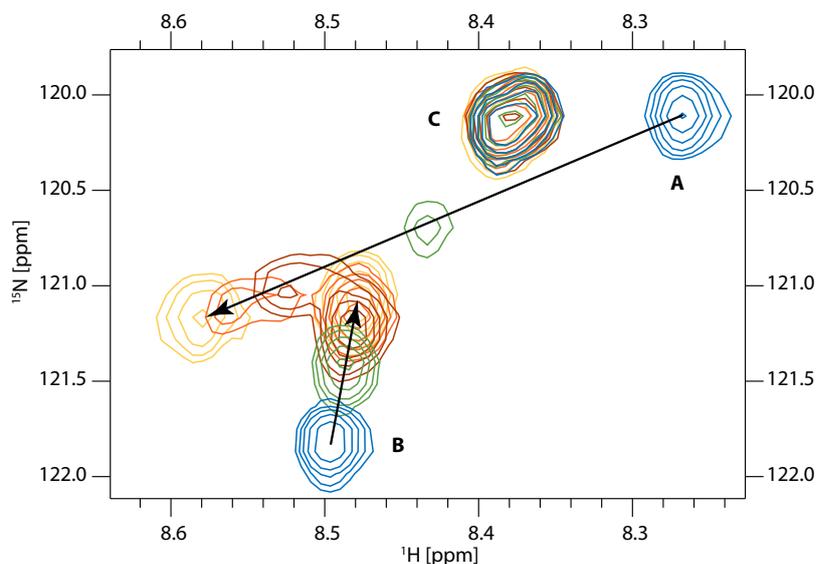


Figure 3.12: Ligand binding studies by ^1H , ^{15}N HSQC experiments. Residue A shows an interaction in the fast-intermediate exchange regime. Upon titration the signal is broadened and weaker. Towards saturation the initial signal intensity is reestablished. Residue B displays CSPs induced by binding in the fast exchange regime. C is an example for a residue that remains unaffected upon ligand binding.

simultaneously.

3.5.6 Relaxation Studies in NMR

The dynamic properties of a molecule are of great interest studying its function. Characterizing the NMR relaxation properties of a protein provides insight into the motions present in the molecule of interest. Applying a RF pulse is perturbing the thermal equilibrium of a sample, which subsequently seeks to return to its equilibrium state. This process, which is basically governed by two mechanisms, is referred to as relaxation.

Spin-lattice-relaxation

Due to interactions with the environment the spin system loses energy resulting in the relaxation of the longitudinal magnetization $M_z(t)$ towards M_0 :

$$M_z(t) = M_z(0)e^{-t/T_1} + M_0(1 - e^{-t/T_1})$$

where M_0 is the longitudinal magnetization at thermal equilibrium and $M_z(0)$ is the part of the magnetization parallel to \vec{B}_0 directly after the RF pulse. T_1 is a sample specific time constant which characterizes this process.

Spin-spin-relaxation

This relaxation mechanism describes the decay of the transverse magnetization, which is due to irreversible entropy effects. The energy exchange within the spin system leads to the loss of phase coherence of the spins. Dephasing as a function of time can be described as following:

$$M_{xy}(t) = M_{xy}(0)e^{-t/T_2}$$

Inhomogeneity of the magnetic field \vec{B}_0 give rise to additional dephasing. The resulting observable time constant T_2^* is therefore smaller than T_2 and consists of T_2 and T_2' (inhomogeneities).

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}$$

In summary the relaxation of the magnetization as a function of time can be described by the empirically determined *Bloch-equation*:

$$\frac{d\vec{M}}{dt} = \gamma\vec{M} \times \vec{B}_{ext} + \frac{1}{T_1}(M_0 - M_z)\hat{z} - \frac{1}{T_2}\vec{M}_{xy} \quad (3.37)$$

Generally relaxation is increasing with molecular size and therefore is a limiting factor for studies by NMR. On the other hand relaxation constants T_1 and T_2 provide valuable information about the relative flexibility of an amino acid within the protein structure. As a general rule T_1 of macromolecules increases with the molecular weight and the rigidity of a residue, whereas T_2 is decreasing. The flexibility of the protein backbone can be estimated by measuring T_1 and T_2 applying HSQC-based NMR experiments. A series of experiments with varying relaxation delay is recorded and the intensity of each signal is monitored. Higher relaxation delays result in a decrease of the signal intensity. The data are fitted to the exponential decay function and the time constants T_1 and T_2 are calculated.

3.5.7 Structure calculation and quality control

Structure determination by solution NMR is based on the assignment of inter and intra molecular NOEs. These restraints contain information about the distance between two protons. Since the intensity of the NOE is in first approximation proportional to r^{-6} , usually peaks are only observed when the distance between to protons is $\lesssim 5 \text{ \AA}$ (see

section 3.5.3). The distance restraints derived from NOEs can be supplemented by additional structural information such as dihedral angles, hydrogen bonds or residular dipole couplings (RDCs). The available NMR data is used in a molecular dynamics simulation applying the known restraints, which are converted into energy terms. Those restraints are applied to minimize the target function which consists of the potential energy and the energy of the restraints. This simulated annealing procedure starts with an initial high temperature state with a subsequent decrease of the system's energy. The objective is to minimize the penalties caused by the violated restraints during the cooling process. The problem of being trapped in a local minimum of energy is avoided by using a set of different starting structures (≈ 100), which are optimized by simulated annealing and the results are subsequently compared to each other. NMR structures are represented as an ensemble of the ~ 20 lowest energy structures of the simulation. The quality of a structure can be evaluated using the iCING web server providing a detailed residue based analysis and validation report [70].

3.6 Experimental Procedures

3.6.1 Cloning of CPEB and Pin1 constructs

Cloning of the CPEB4 RRM domains

Complementary DNA (cDNA) from human CPEB 4, Isoform 1 was used as a template to clone into the pETM-11 vector with kanamycin resistance, Tobacco Etch Virus (TEV) protease cleavage site and a N-terminal His₆-tag. The constructs cloned correspond to the following boundaries: RRM1 (463 - 573) and RRM1-RRM2 (463 - 665). Residue numbering is according to isoform 1 of CPEB4 (Uniprot: Q17RY0). *Escherichia coli* DH5 α strain was used for cloning. All clones were confirmed by DNA sequencing.

Cloning of the CPEB1 N-terminal fragments

Three fragments from the N-terminal region of CPEB1, Isoform 1 (Uniprot Q9BZB8) NF1 (48-183), NF1s (88-183) and NF2 (196-293) were cloned from cDNA into pETM-11 vectors

Cloning of Pin1 constructs

The Pin1 WW domain (Uniprot: Q13526, Residues: 1-41) was cloned via synthetic (template free) PCR (Eric Aragon, Maria J. Macias' group, IRB Barcelona) using partially overlapping, complementary oligonucleotides. This approach was used to remove some rare codons present in the template. The DNA insert was cloned into the pETM-30 expression vector containing an additional Gluthation-S-Transferase (GST) -tag between the His₆-tag and the TEV cleavage site.

The full-length Pin1 clone was obtained from (Addgene, Cambridge, USA) in ampicillin-resistant pMCSG7 vector containing a His₆-tag and the TEV cleavage site, provided by Dustin Maly's Lab.

3.6.2 Protein Expression and Purification

Transformation

Transformation is the process by which foreign DNA is introduced into a host cell. The transformation of the pure plasmid DNA into the competent *E.coli* (DH5 α or BL21(DE3)) cells was carried out by the heat shock method. Therefore 100 μ L of competent cells were thawed on ice and 1 μ L of pure plasmid DNA containing the protein of choice is added. The mix is incubated on ice for 20 min before a 45 second

heat shock at 42 °C without shaking. The tube is placed on ice immediately for at least 2 min and consequently 300 μ L of SOC medium (Super Optimal Broth with Catabolite Repression) is added. The culture is grown for 1h at 37°C in agitation 600 *rpm* and the centrifuged for 1 min at 6.000 *rpm*. The supernatant except 100 μ L is removed and the suspension is plated out on a LB agar plate containing the appropriate antibiotic. Plates are incubated overnight at 37°C to obtain bacterial colonies of an optimal size.

Recombinant Protein Expression and Protein Labeling

A starter culture was grown 3-4 hours at 37°C by inoculating 5 ml of LB medium with a single colony from a freshly streaked plate of the expression host containing the recombinant vector. The main culture was inoculated at 37°C and incubated until the optical density (OD₆₀₀) reaches a value of 0.6 (measured at $\lambda = 600$ nm). Depending on the expression temperature, cultures were cooled to the desired temperature of expression before induction with Isopropylthiogalactoside (IPTG) to a final concentration of 0.2 - 0.5 mM. Unlabeled and labeled samples were prepared using LB or TB and minimal media (M9) cultures, respectively. D₂O (99.89 %, CortecNet), ¹⁵NH₄ Cl and/or D-[¹³C] glucose were used as sole hydrogen, nitrogen and carbon sources respectively to prepare the labelled samples as described [58]. The overexpression was checked by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). After harvesting the cell pellet by centrifugation (20 min at 3500 g), pellets were purified immediately or stored at -20°C until purification.

Cell lysis and Protein Purification

The cell pellets were resuspended in the lysis buffer optimized for the construct (for buffer compositions, please refer to Section 3.7) and lysed using an EmulsiFlex-C5 (Avestin) in the presence of lysozyme and DNaseI. The Lysate was centrifuged for 20 min at 4°C and 23 000 g. The soluble supernatant was purified by nickel-affinity chromatography. Affinity chromatography separates proteins on the basis of a reversible interaction between a protein (or group of proteins) and a specific ligand coupled to a chromatography matrix. The Ni²⁺-column (HiTrap Chelating HP column, GE Healthcare Life Science, Uppsala Sweden) is equilibrated with the lysis buffer and the matrix is incubated with the supernatant. Unbound molecules are washed from the column using 1-2 column volumes of washing buffer. The target molecule was eluted using buffers either containing EDTA (50 mM) or Imidazole (250 mM). Eluted His₆ target proteins were digested with TEV protease at room temperature overnight, and further purified by size-exclusion chromatography on a HiLoad Superdex 30 or 75 16/60 pregrade columns (GE Healthcare) equilibrated in a suitable buffer. Cleavage and purity of the HPLC-fractions were monitored using SDS-PAGE.

Protein	MW [kDa]	Expression conditions	Samples
CPEB4 RRM1	12.8	LB, 37°C, o/n, 0.5 mM IPTG	unlabelled, ¹⁵ N, ¹³ C/ ¹⁵ N, ² H/ ¹³ C/ ¹⁵ N
CPEB4 RRM2	11.2	LB, 37°C, o/n, 0.2 mM IPTG	refolding not successful
CPEB4 RRM1-RRM2	22.9	LB, 37°C, o/n, 0.5 mM IPTG	unlabelled, ¹⁵ N, ² H/ ¹³ C/ ¹⁵ N
Pin1 WW	4.9	LB, 37°C, o/n, 0.5 mM IPTG	unlabelled, ¹⁵ N
Pin1 Fl	18.1	TB, 22°C, 4h, 0.4 mM IPTG	unlabelled, ¹⁵ N

Table 3.1: Overview of expression conditions and samples prepared

3.6.3 Manual SPPS of phosphorylated and unphosphorylated peptides

All peptides were manually synthesized in a 0.1 - 0.5 mmol scale with H-Rink Amide ChemMatrix® (substitution: 0.52 mmol/g) used as a solid support. ChemMatrix® is a 100% polyethylene glycol (PEG)-based polymer cross-linked through primary ether bonds, resulting in high chemical and thermal stability. Solid-phase synthesis was carried out in a polypropylene syringe fitted with a polyethylene porous filter. The solvents and soluble reagents were removed by suction. Prior to the first coupling reaction the resin was swollen for 30 min in DMF directly in the synthesis reactor. The Fmoc group was removed with piperidine (40 %, 20% Piperidine in DMF, v/v) for 5 min each. Washings, after deprotection and coupling steps, were carried out with DMF (5 x 0.5 min) and dichloromethane (DCM) (3 x 0.5 min) using 10 mL solvent/g resin each time.

Coupling was performed with a three-fold excess of Fmoc-protected amino acids. As activating agents 2.9 equivalents of HATU with or without the addition of 2.9 equivalents HOAt and 6 equivalents N,N-Diisopropylethylamine (DIPEA) or 2.9 equivalents of Oxyma Pure and 2.9 equivalents diisopropylcarbodiimide (DIC) were used. The Fmoc amino acid together with HATU/HOAt was completely dissolved in DMF, with

Peptide	Sequence	MW[Da]
CPEB1 S144	GKPSPLGFLT	1057.2
CPEB1 S248	AATVSPLG	1085.2
CPEB1 pS138	SMLN _p SPMGK	755.81
CPEB1 pS144	GKP _p SPLGFLT	1137.2
CPEB1 pS184	SRSS _p SSDSG	1086.93
CPEB1 pS210	LRI _p SPPLHFLPLGG	1637.9
CPEB1 pS210s	LRI _p SPPLPFG	1175.34
CPEB1 pS248	AATV _p SPLG	835.8
CPEB1 pS423	FVR _p SPSQRLDPG	1393.43

Table 3.2: Peptides from CPEB1 N-terminal regions synthesised by Fmoc-SPPS

subsequent addition of DIPEA. After vortexing the mixture it was immediately added onto the resin. In case of using Oxyma Pure as an activating agent, Oxyma and the Fmoc-amino acid were dissolved separately in the same final volume as for the HATU-coupling. The dissolved Oxyma, amino acid and DIC are mixed together, vortexed and directly added onto the resin. The peptidyl-resin mixture was mechanically stirred for 45 min. After each coupling, the ninhydrin [61] or chloranil test [62] was run to check the coupling. When the test showed a positive result, recoupling was performed, applying the same coupling conditions. Furthermore, if the colorimetric tests were inconclusive after the second coupling, the yield of the coupling was controlled by resin test cleavage, taking a representative sample of peptidyl-resin that was cleaved by TFA-H₂O (95 : 5) - mixture for 1 h. The intermediate was consequently analysed by liquid chromatography-mass spectrometry (LC-MS). After deprotection of the final amino acid, the unprotected N-terminus was acetylated with acetic anhydride solution (5 % acetic anhydride (Ac₂O), 8.5% DIPEA, 86.5% DMF for 30 min). Cleavage was performed adding 10 mL/g of TFA - H₂O - EDT (95 : 2.5 : 2.5) for 2-4 h. The mono-benzyl ester protecting group of the phospho-Ser/Thr is readily removed during the TFA-mediated acidolysis. TFA was evaporated until a final volume of 2-3 mL. The acidic solution was poured on cold diethyl ether, thereby precipitating the target peptide. The product was isolated by centrifugation (10 min at 4000 rpm and 4°)

of the mixture and decanting the solution. This procedure was repeated twice. The white solid was dissolved in H₂O-acetonitrile (AcN) (1 : 1) and lyophilized.

3.6.4 Peptide purification and analysis

All peptides were purified using reverse phase high performance liquid chromatography (RP-HPLC) on a ÄKTApurifier10 (GE Healthcare Life Science, Uppsala, Sweden) monitoring the 215 nm absorbance of the amide bond. Analytical HPLC was performed using a Vydac C18 column (5 μ m particle size, 4.6 x 250 mm); for preparative runs a SunFire C18 Sephasil preparative column (10 μ m particle size, 50 x 250 mm) was used. All runs were performed applying linear gradients of 5% ACN, 0.05% TFA (HPLC-grade) in deionized miliQ H₂O (solvent A) versus 70% ACN, 0.05% TFA (solvent B). Peptides eluted depending on their sequence and length around 20-40 % solvent B. Mass spectrometry analysis was routinely applied to all fractions of the purification. Pure fractions were lyophilised and resuspended in miliQ H₂O roughly estimating the final concentration to 10-15 mM. Subsequently the pH was neutralised with sodium hydroxide.

The exact concentration of each peptide sample was determined by amino acid analysis performed by the “Unidad de Técnicas Separativas de Análisis, Servicios Científico-Técnicos“, University of Barcelona. This technique basically implies a peptide hydrolysis in acidic conditions (6 M hydrochloric acid, for 24h at 110°) with consequent analysis of the sample by HPLC equipped with a fluorescence detector. During this process Met undergoes oxidation, Gln and Asn are deamidated, resulting in Glu and Asp, respectively. Chromatographic peak areas for each amino acid are identified and integrated. The value of each peak area is corrected according to the differing molar absorptivities of the various amino acids. The amount of each amino acid in the sample is calculated by first dividing the corrected peak area of each amino acid by the internal standard in the chromatogram and subsequently multiplying this value by the total amount of internal standard added to the original sample.

3.6.5 Ion Mobility - Mass Spectrometry

Our TW Ion Mobility mass spectrometry experiments were performed on a Synapt G1 HDMS mass spectrometer (Waters, Manchester, UK). Samples were placed on a 384-well plate refrigerated at 15 °C and sprayed using a Triversa NanoMate® (Advion BioSciences) automated Chip-Base nano-electrospray working in the positive ion mode. The instrument was calibrated over the 500-8000 Da m/z range using a caesium iodide solution. The software MassLynx 4.1 SCN 704 and Driftscope 2.1 were used for data processing. Protein samples (final concentrations of 30-50 μ M) were prepared in 20 - 50 mM NH₄OAc pH 7.2. Prior to analysis a 1D - ¹H NMR spectrum was acquired

on all samples in the IM-MS buffer conditions in order to check the stability of the samples and to compare it with spectra obtained in NMR buffer conditions. Spray voltage was set to 1.75 kV and delivery pressure to 0.5 psi. A reduction of the source pumping speed in the backing region (5.85 mbar) was set for optimal transmission of high mass non-covalent ions. Cone voltage, extraction cone and source temperature were set to 20 V, 6V and 20°C respectively. Ions passed through a quadrupole mass filter to the IM-MS section of the instrument.

3.6.6 ITC

ITC measurements were performed using a nano ITC calorimeter (TA Instruments) at 5 °C, 12 °C or 25°C. To be consistent with the condition used for the NMR studies protein samples were prepared in the NMR buffer. To avoid buffer mismatch, which would affect the quality of the data, buffer from concentrating the protein samples was used to resolubilise the ligand. All samples were centrifuged and degassed prior to the experiments. Depending on the expected affinity and stoichiometry, sigmoidal curves were optimised by injecting 5 to 15 fold concentrated ligand in 16 x 3 μ L steps in a cell containing 190 μ L of protein at 20 - 200 μ M adjusted concentration. The delay between injections was 3 minutes and data was collected while stirring at 200 revolutions per minute (rpm). The NanoAnalyze software (TA Instruments) was used to analyse the binding isotherms assuming a single binding site in each molecule. Baseline controls were acquired with buffer and pure RNA or peptide solutions. Measurements have been repeated at least twice.

3.6.7 NMR Chemical Shift Perturbation Experiments

^1H , ^{15}N HSQC experiments for CPEB samples were recorded at 303 K using a Bruker AVIII 600-MHz spectrometer equipped with a 5 mm TXI cryprobe, z-gradient. Protein samples of the CPEB4 RRM1 and the CPEB4 RRM1-RRM2 were equilibrated in a buffer containing 20 mM Tris- d_{11} , 130 mM NaCl and 5 % DMSO- d_6 . All samples were supplemented with 10% D_2O and pH adjusted to value 7. Spectra were acquired using 200 μ M ^{15}N -labeled protein samples equilibrated together with progressively increasing amounts of the unlabeled RNA fragments until saturation was achieved.

Spectra of Pin1 full-length and Pin1 WW were acquired at 300 K and 285 K, respectively, using a Bruker AVIII 600-MHz spectrometer, equipped with a room temperature triple resonance gradient probe. Samples were equilibrated in a 50 mM sodium phosphate, 50 mM sodium sulfate, 5 mM EDTA and 1 mM DTT buffer. The pH value was adjusted to 6.6 and the samples were supplemented with 10% D_2O . All spectra were recorded on 200 μ M ^{15}N -labeled protein samples equilibrated together with progressively increasing amounts of various phosphorylated CPEB peptides until saturation was achieved. Chemical shift perturbation of analyses were performed on

CcpNmr Analysis with a 0.15 weighting of ^{15}N with respect to ^1H .

One-dimensional ^{31}P -NMR-Titration experiments were acquired on 200 μM phospho peptide samples equilibrated in a buffer with 20 mM Tris- d_{11} and 130 mM NaCl. Progressively increasing amounts of the unlabeled Pin1 full-length (at 300 K) and Pin1 WW (at 285 K) were added until saturation was achieved. All samples were supplemented with 200 μM sodium phosphate buffer as a ^{31}P reference.

3.6.8 NMR Assignments

Spectra of the CPEB4 RRM1 and the CPEB4 RRM1-RRM2 were acquired at 303 K using a Bruker AVIII 600-MHz spectrometer, equipped with cryogenic or room temperature triple resonance gradient probes. Backbone ^1H , ^{13}C and ^{15}N resonance assignments were obtained by analyzing the 3D HNCACB and HN(CO)CACB experiment pair in the case of the CPEB4 RRM1 (fully protonated sample) or the CBCANH and CBCA(CO)NH pair in the case of the CPEB4 RRM1-RRM2 (^2H , ^{13}C , ^{15}N samples). Transverse Relaxation Optimized Spectroscopy (TROSY) versions [71] of these experiments and/or the Non-Uniform Sampling (NUS) acquisition strategy were used in selected cases to reduce experimental time and increase resolution. A limiting factor for NMR-Experiments in general is the relaxation rate. Typically relaxation for large proteins at high magnetic field strengths the transverse (T_2) is dominated by the dipole-dipole (DD) mechanism and the chemical shift anisotropy (CSA) mechanism. As these relaxation mechanisms are generally correlated but contribute to the overall relaxation rate of a given component with different signs, the multiplet components (due to J-coupling) relax with very different overall rates. The TROSY experiments are a class of experiments designed to select the component for which the correlation of the different relaxation mechanisms have almost cancelled, leading to a single, sharp peak in the spectrum. This significantly increases both spectral resolution and sensitivity, which during the assignment process can be a huge advantage.

Inter and intra-molecular proton distance restraints for Pin1 WW domain in complex with CPEB1 pS210 were assigned in the 2D homonuclear NOESY experiments. Furthermore peptides were assigned via NOE-based sequential resonance assignment using homonuclear 2D TOCSY and 2D NOESY spectra.

All buffer conditions were as mentioned. NMRPipe [72] was used for spectra processing. CARA [73] and CcpNmr Analysis [74] were used for spectra analysis and assignment.

3.6.9 Relaxation Measurements

Amide relaxation measurements were acquired on 500 μM ^{15}N -labeled RRM1-RRM2 and Pin1 full-length samples at 303 K and 300 K, respectively, on a Bruker AVIII 600-

MHz spectrometer, equipped with a room temperature triple resonance gradient probe. Twelve different relaxation delay time values (21.6, 54, 108, 162, 270, 432, 540, 702, 864, 1080, 1404 and 1728 ms) were measured to determine T_1 . To determine T_2 ten experiments were recorded with the following ^{15}N mixing times: 16.74, 33.48, 50.22, 66.96, 100.44, 117.18, 133.92, 167.40, 184.14, 200.88. All relaxation experiments were acquired as pseudo-3D experiments and converted to 2D data sets during processing. Peak integration values were fitted to a two-parameter function as following:

$$I(t) = I_0 e^{(-t/T_{1,2})} \quad (3.38)$$

where I_0 and $I(t)$ are the peak intensities at times 0 and t , respectively. The rotational correlation time of the RRM1-RRM2 pair was calculated with equation 3.39 derived from [75] equation 8, using the approximation of slow molecular motion τ_c larger than 0.5 ns and assuming that only $J(0)$ and $J(\omega_N)$ spectral density terms contribute to the overall value:

$$\tau_c \approx \frac{1}{4\pi\nu_N} \sqrt{6 \frac{T_1}{T_2} - 7} \quad (3.39)$$

where ν_N is the ^{15}N resonance frequency ($60,08 \cdot 10^6$ Hz).

For the $^1\text{H}, ^{15}\text{N}$ NOE experiment the reference and the presaturated HSQC spectra were acquired in an interleaved manner. The values of the steady-state $^1\text{H}, ^{15}\text{N}$ NOEs result from the ratios of the peak intensities measured in the reference and the presaturated spectra during the relaxation delay as described [76]. Background noise levels σ_S and σ_0 were measured and used to determine the NOE standard deviation through the following relationship:

$$\frac{\sigma_{NOE}}{NOE} = \sqrt{\left(\frac{\sigma_{I_S}}{I_S}\right)^2 + \left(\frac{\sigma_{I_0}}{I_0}\right)^2} \quad (3.40)$$

3.7 Buffers and solutions

PROTEIN EXPRESSION MEDIA			
LB medium (1L)		LB-Agar plates	
Bactotryptose	10g	LB medium	1.5%(w/v)
Yeast extract	5g	Autoclaved Agar	
NaCl	10g	appropriate antibiotic:	
appropriate antibiotic:		Kanamycin	25 $\mu\text{g}/\text{mL}$
Kanamycin (25mg/ml)	1mL	Ampicillin	25 $\mu\text{g}/\text{mL}$
Ampicillin (50mg/ml)	1mL		
(Autoclaved)			
Terrific Broth (TB) (1L)		SOC (0.5L)	
Tryptone	12g	Bactotryptose	10g
Yeast extract	24g	Yeast extract	2.5g
Glycerol (99.9%)	4mL	NaCl (5M)	1mL
dissolve in H ₂ O up to	900mL	KCl (1M)	1.25mL
		MgCl ₂ (1M)	5ml
KH ₂ PO ₄	0.17M	MgSO ₄	5mL
K ₂ HPO ₄	0.72M	Glucose (1M)	10mL
(Autoclaved)		(Autoclaved)	
add 100ml of Phosphate potassium solution for a final volume of 1L			
Medium A for ¹⁵ N, ¹³ C labelling (1L)		M9 medium (10x) (1L)	
¹⁵ M9 medium 10x	100mL	Na ₂ HPO ₄	60g
Trace Elements 100x	10mL	KH ₂ PO ₄	30g
Glucose/ ¹³ C ₆ -Glucose (20%)	20/10mL	NaCl	5g
MgSO ₄ (1M)	1mL	¹⁴ NH ₄ Cl/ ¹⁵ NH ₄ Cl)	5g
CaCl ₂ (1M)	300 μL		
Biotin (1 mg/ml)	1mL		
Thiamin (1 mg/ml)	1mL		
appropriate antibiotic			

Table 3.3: Cell culture media and solutions

PROTEIN BUFFERS			
Ni ²⁺ -Column Lysis/ washing buffer A		Ni ²⁺ -Column Elution buffer	
NaCl	150mM	NaCl	150mM
Tris-HCl pH 8.0	20mM	Tris-HCl pH 8.0	20mM
Imidazole	10mM	Imidazole	10mM
		EDTA	50mM
Lysis buffer (Pin1 FL)		Washing buffer (Pin1 FL)	
Tris-HCl pH 8.0	25mM	Tris-HCl pH 8.0	25mM
NaCl	500mM	NaCl	500mM
Imidazole	10mM	Imidazole	10mM
β -Mercaptoethanol	10mM	β -Mercaptoethanol	10mM
Tween 20	1% (v/v)	MgCl ₂ (1M)	5ml
Elution buffer (Pin1 FL)		Tris-HCl buffer	
Tris-HCl pH 8.0	25mM	Tris-HCl/Tris-d ₁₁	20 mM
NaCl	500mM	NaCl	130mM
Imidazole	250mM	NaN ₃	2mM
β -Mercaptoethanol	10mM	(adjust pH to 7.0)	
Sodium sulfate/phosphate buffer		HEPES buffer	
Na ₂ SO ₄	50mM	HESPES	10 mM
Na ₂ HPO ₄	50mM	NaCl	100mM
NaH ₂ PO ₄	50mM	NaN ₃	2mM
EDTA	5mM	(adjust pH to 7.7)	
DTT	1mM		
(adjust pH to 6.6)			
Ion-Mobility MS buffer			
Ammonium Acetate	50mM		
in LC-MS CHROMASOLV®			
water			

Table 3.4: Protein Purification, NMR, Crystallography and ITC buffers

4 Results

4.1 RNA Recognition and self-association of CPEB4 by its tandem RRM domains

4.1.1 Interactions of the RRM domains in the absence of RNA

To gain insight which role the two RRM domains play in RNA binding we prepared several constructs of the isolated RRM1, RRM2 and the RRM1-RRM2 tandem domains using His-tagged fusion expression vectors. The isolated RRM1 domain and the tandem construct were in the soluble fraction of the expression cultures and they were purified by Ni-affinity chromatography. However, overexpression of the isolated RRM2 using various experimental conditions resulted in the formation of inclusion bodies containing the RRM2 protein. Protein purified under denaturing conditions, in the presence of guanidium chloride, could not be refolded by dialysis under the various buffer conditions we have assayed.

We therefore focused our work on the construct containing the pair of RRM domains and also on the isolated RRM1, which was used for comparison. In order to assign the amide and backbone resonances and to facilitate the investigation into whether the first RRM domain adopts a similar fold when being independently expressed or in the construct of both consecutive RRM domains $^{13}\text{C}/^{15}\text{N}$ and $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled samples were prepared respectively. Due to its size and compact fold, the acquisition of the triple resonance experiments of the tandem required the deuteration of the sample in order to minimize the loss of the signal caused by transverse T2 relaxation. Applying TROSY versions of all NMR experiments was essential to improve the resolution and sensitivity of the back-bone experiments.

Analyzing the secondary ^{13}C chemical shifts obtained from the backbone assignments indicate that both RRM1 and RRM2 in the tandem construct adopt the canonical $\alpha\beta$ sandwich structure with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology with an additional β -strand β_4' at the C-terminus of RRM1 (see Figure 4.1, *upper panel*). Comparison of the ^{13}C chemical shifts of RRM1, when assigned independently or in the RRM1-RRM2 construct, indicate that the secondary structure is not altered due to the presence of RRM2 (cf. Figure 4.1, *lower panel*).

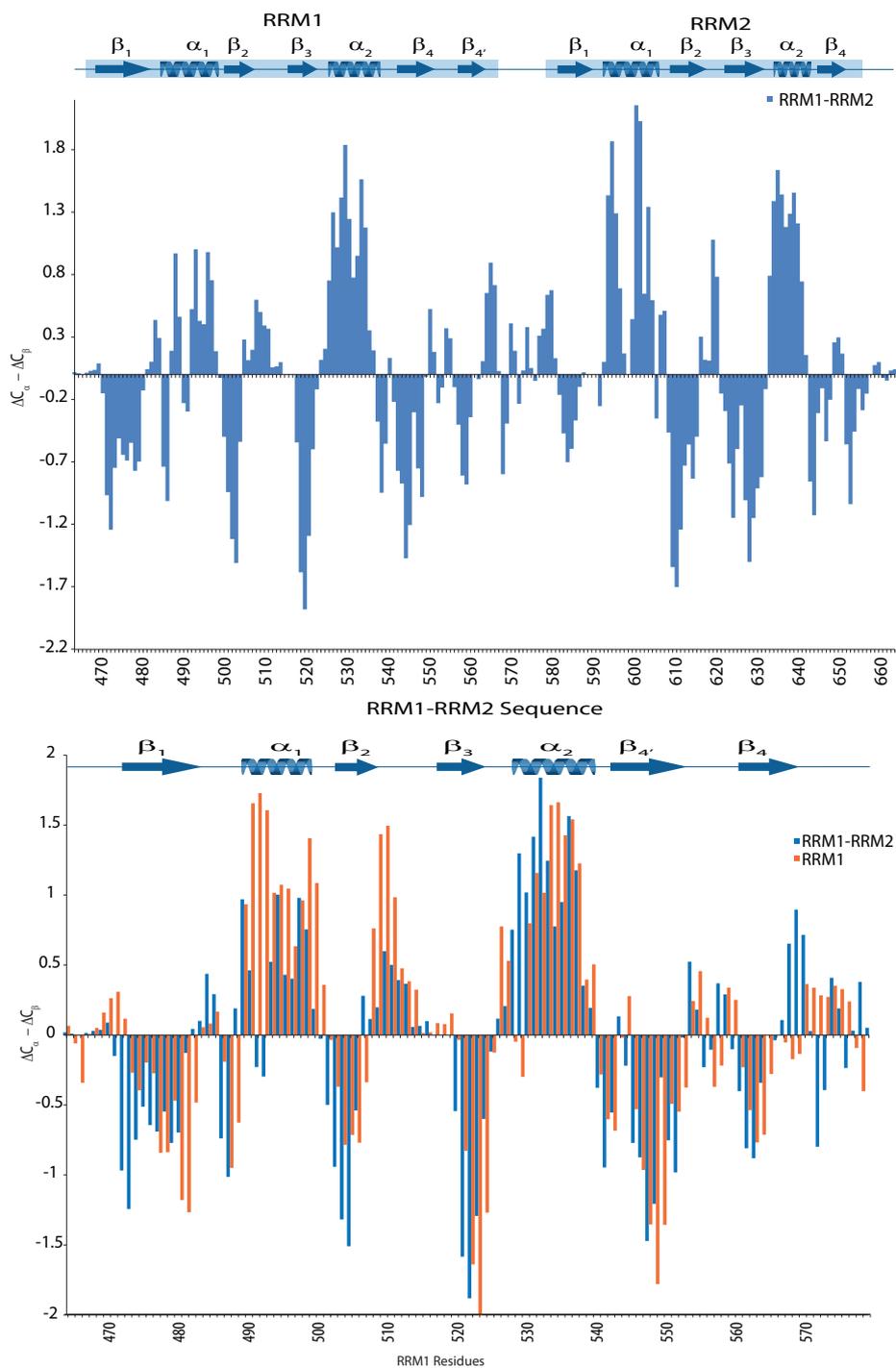


Figure 4.1: Secondary ^{13}C chemical shifts $\Delta C_{\alpha} - \Delta C_{\beta}$ of CPEB4 RRM1-RRM2. Residues with negative CSI indicate β -strand conformation whereas positive values indicate turn or α -helix (*upper panel*). Comparison of the CSI of RRM1 single (orange) and in the RRM1-RRM2 tandem (blue). It shows that the secondary structure of RRM1 is not significantly altered due to the presence of RRM2 (*lower panel*).

For RRM1 a homology model was built (cf. Figure 4.2) using SWISS-MODEL (Template: PDB entry 2DNL, RRM1 of CPEB3, sequence identity 97%, <http://swissmodel.expasy.org/>). The topology of the model obtained is consistent with the elements of secondary structure indicated by the analysis of the ^{13}C chemical shifts. The spatial arrangement of the canonical four-stranded antiparallel β -sheet is $\beta_4\beta_1\beta_3\beta_2$. The homology model we obtained indicates that the additional $\beta_{4'}$ strand is arranged antiparallel to β_4 resulting in the following order for the extended β -sheet: $\beta_{4'}\beta_4\beta_1\beta_3\beta_2$.

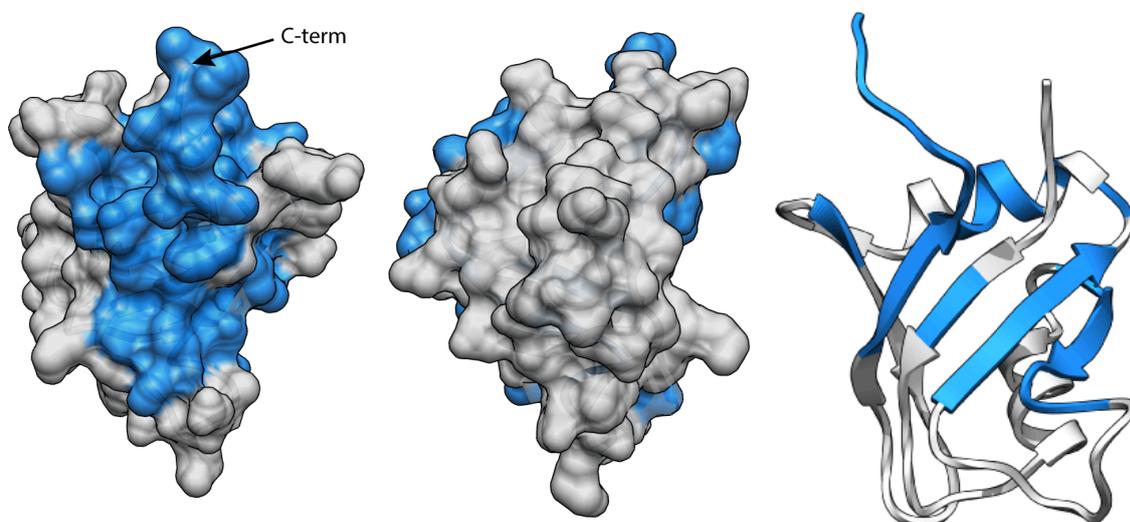


Figure 4.2: Significant differences in chemical shifts are mapped onto the homology model for RRM1. As those differences are clustering on one surface of the model, we assume that this surface lies in proximity of the RRM2

HSQC experiments of both constructs were measured under identical conditions to compare the dispersion of the amide resonances. The comparison of amide resonances can be used to assess the influence that the domains may have on the structure of each other. The superimposition of both spectra reveals that the signals only match reasonably well in certain areas. Significant differences in the chemical shift of some residues are observed between the RRM1 and the RRM1-RRM2 construct. The linker connecting both domains is very short in length (six residues), limiting the degrees of freedom of the domains in the tandem. The CSPs observed indicate that both RRM domains interact in their free state. Furthermore, when mapping the most affected residues onto the homology model of the RRM1 domain, they cluster on one surface including the β -sheet, indicating that this surface should be in proximity of the RRM2 domain (see Figure 4.2).

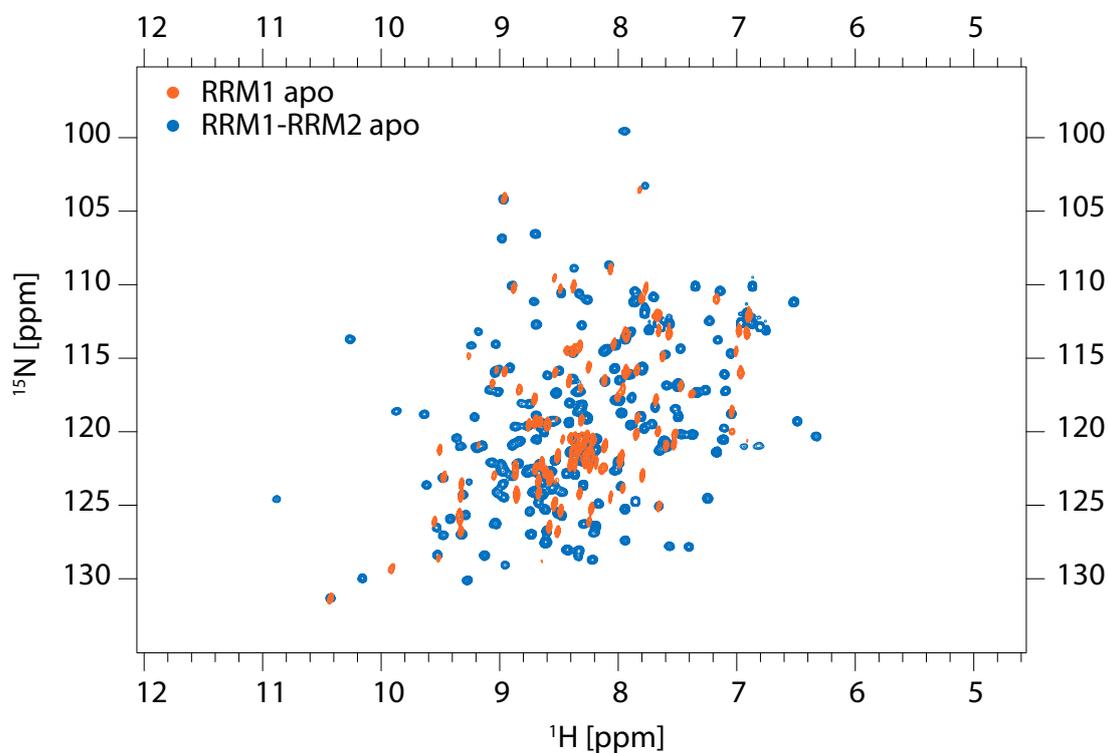


Figure 4.3: Comparison of ^1H , ^{15}N -HSQC spectra of the apo RRM1 (orange) and apo RRM1-RRM2

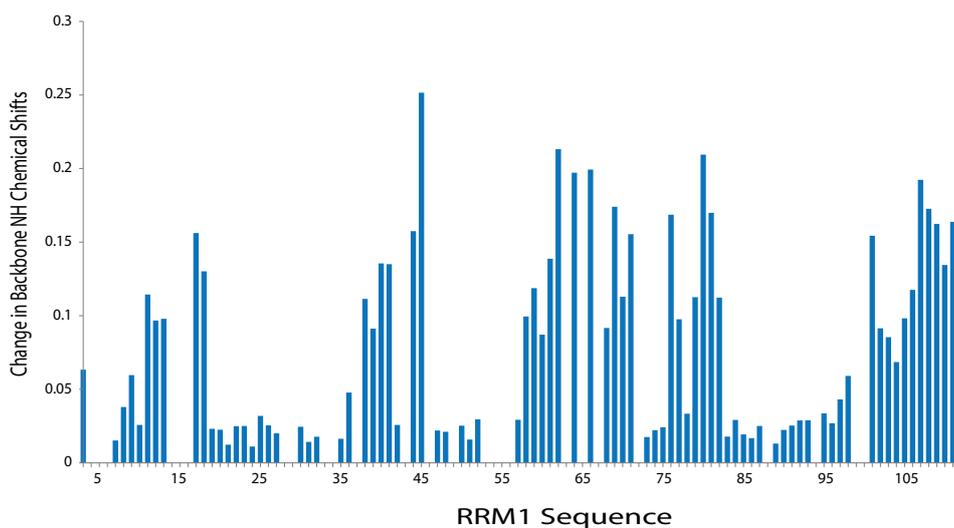


Figure 4.4: Comparison of NH CSP of the apo RRM1 and apo RRM1-RRM2 constructs. Residues mainly affected are clustering on the surface displayed in Figure 4.2.

4.1.2 RRM domains are tumbling as a single unit

To further investigate the backbone dynamics of the RRM1-RRM2 tandem domains,¹⁵N NMR relaxation data of CPEB4 RRM1-RRM2 in its free state was acquired. Already in the apo form, the values for ¹⁵N longitudinal T₁ and transverse T₂ are homogeneous along the whole sequence of the tandem (see Figure 4.6). Apart from the few N- and C-terminal residues, T₁ obtained are 1202 ± 113 ms and T₂ yielded 49.3 ± 5.1 ms. ¹H,¹⁵N - NOE data show that especially for the residues of the linker region between the two RRM domains no increased internal motion is observed.

Assuming isotropic rotational diffusion the rotational correlation time τ_c^{apo} was calculated to 15.6 ± 0.7 ns. The result is in agreement with the expected value for a globular protein of 23 kDa. This further confirms that the CPEB4 RRM1-RRM2 domains already in their free state tumble as a single unit in solution. Based on this evidence the chemical shift changes observed in the RRM1 domain are interpreted as an indication of a conformation where both RRM domains are close to one another.

To investigate whether upon RNA binding significant changes in the protein dynamics occur, ¹⁵N-NMR relaxation experiments of the RRM tandem in complex with U5A1U2 were recorded. Values obtained for T₁ (1134 ± 113 ms) and T₂ (50.8 ± 4.2 ms) show no significant differences compared to the T₁ and T₂ times of the free state. The rotational correlation time $\tau_c^{complex}$ was calculated to 14.8 ± 0.8 ns. Both values of $\tau_c^{complex}$ and τ_c^{apo} lie within the error limits of each other thus not allowing to significantly distinguish between the apo state and the complex. The results indicate that the RRM domains tumble as a unit already in their free state. Upon RNA binding no significant changes in the dynamic properties of the RRM tandem are observed.

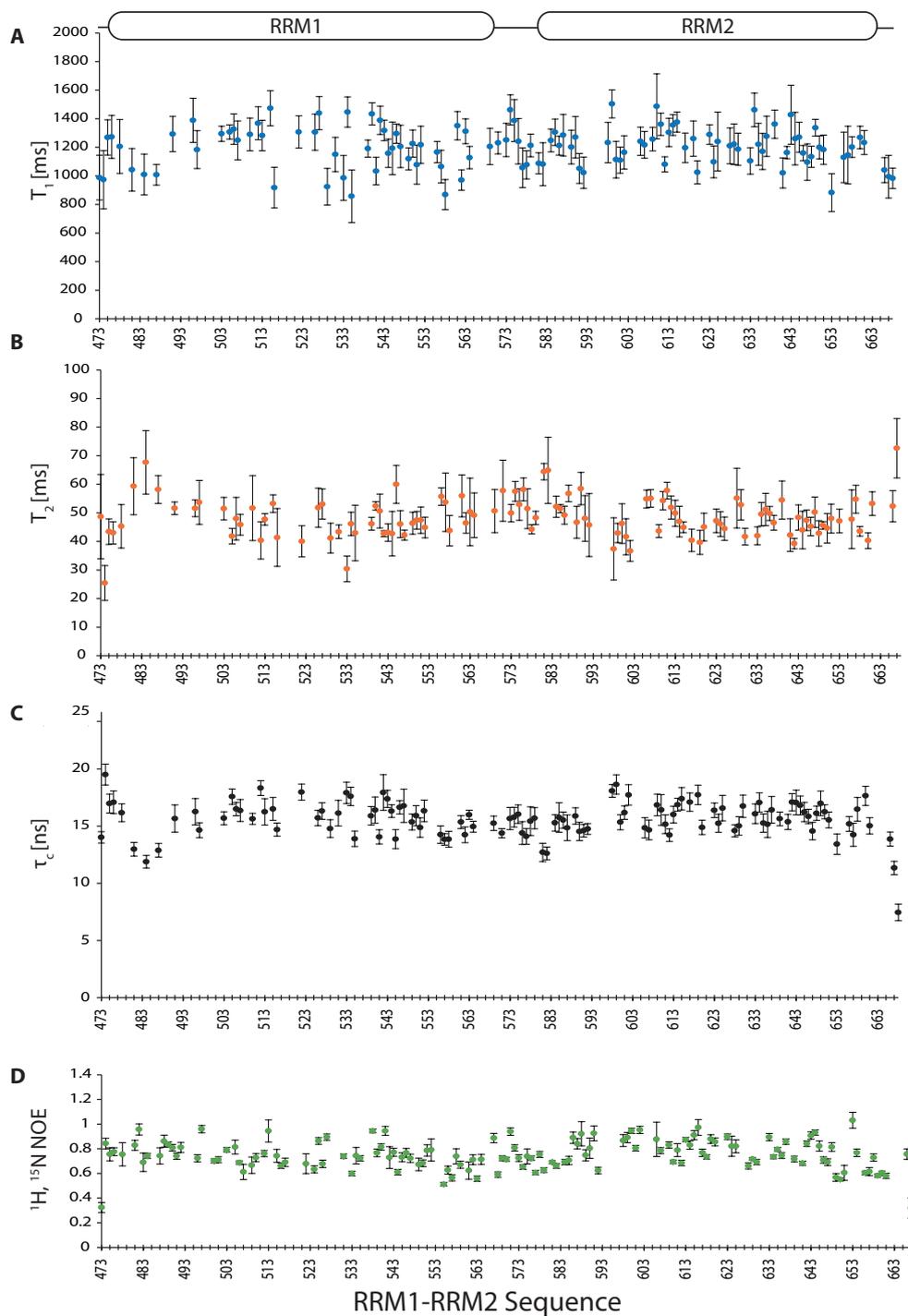


Figure 4.5: Amide ^{15}N relaxation properties of the CPEB4 RRM1-RRM2 apo form. Measurement of T_1 (longitudinal) (A) and T_2 (transverse) (B) relaxation constants and $^1\text{H}, ^{15}\text{N}$ NOE values (D). Local rotational correlation times τ_c (C) calculated from T_1/T_2 ratio. Overall τ_c is ~ 15.6 ns, consistent with a 23 kDa globular protein.

4.1 RNA Recognition and self-association of CPEB4 by its tandem RRM domains

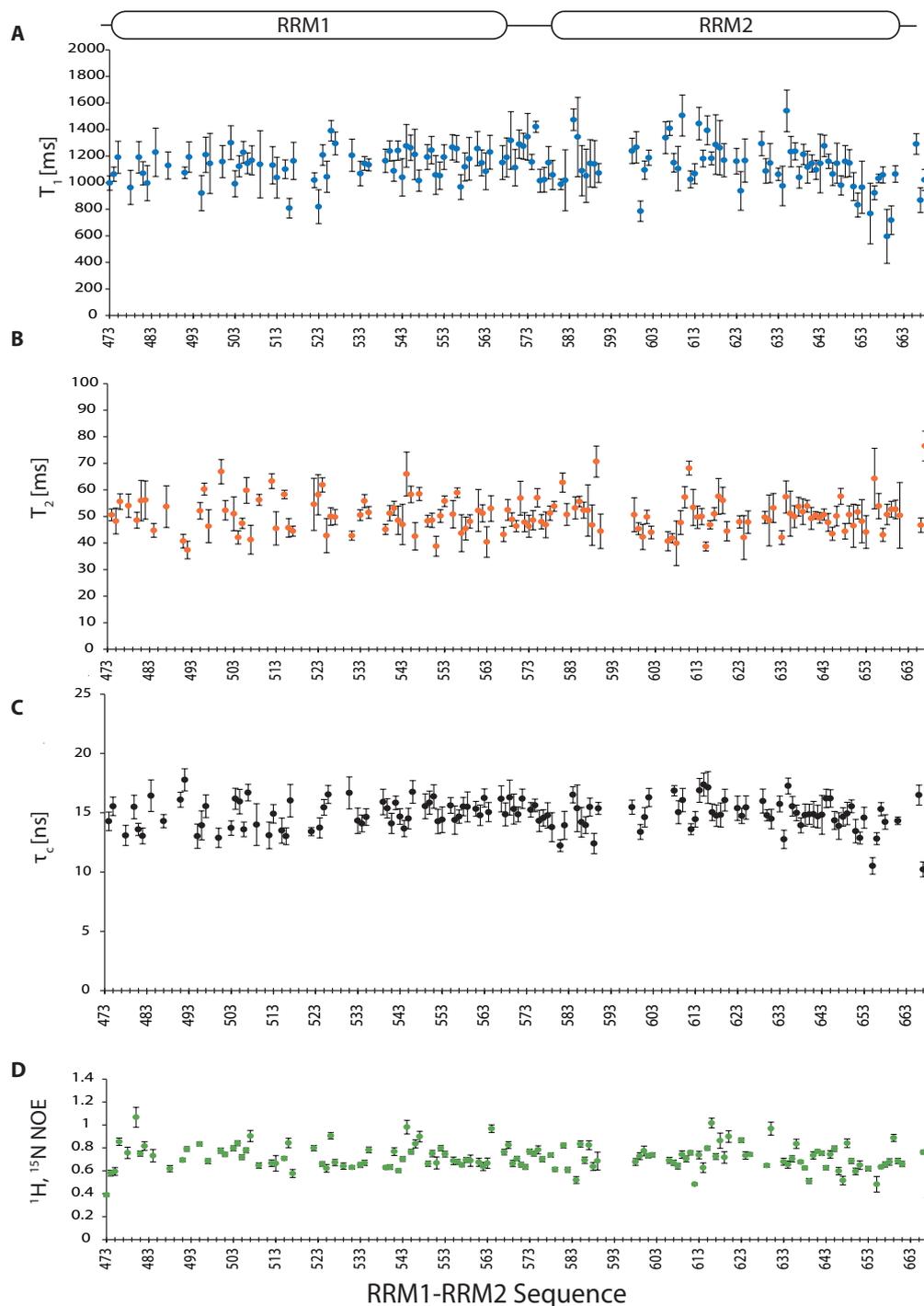


Figure 4.6: Amide ^{15}N relaxation properties of the CPEB4 RRM1-RRM2 in complex with U5A1U2. Measurement of T_1 (longitudinal) (A) and T_2 (transverse) (B) relaxation constants and $^1\text{H}, ^{15}\text{N}$ NOE values (D). Local rotational correlation times τ_c (C) calculated from T_1/T_2 ratio. Overall τ_c is ~ 14.8 ns, consistent with a 23 kDa globular protein.

4.1.3 CPEB4 uses both RRM domains to bind CPE motifs

In order to characterize the interaction of the RRM domains with short RNAs containing the CPE-motifs, Isothermal Titration Calorimetry (ITC) binding assays using both RRM1 and RRM1-RRM2 and different RNA fragments we performed. The binding ligands used for ITC experiments are listed in Table 4.1. Apart from two natural occurring octamer RNA fragments containing the two consensus CPE motifs (U5A2U1 and U5A1U2), a poly U stretch U8, a random control motif, C1U4A1 and the non-consensus motif C1U3A1 were investigated.

	RRM1	RRM1-RRM2
RNA fragment	K_D [μ M]	K_D [nM]
U5A2U1	3.1 ± 1.1	323 ± 34
U5A1U2	100	299 ± 28
U8	NB	248 ± 49
C1U3A1	NB	NB
C1U4A1	NB	5400 ± 734
Control (random motif)	NB	NB

Table 4.1: Binding affinities between CPEB4 RRM1, CPEB4 RRM1-RRM2 and RNA fragments (CPE-motif containing, poly U and random motifs) measured by ITC.

In general, the ITC experiments indicate that all ligands show higher binding affinities with RRM1-RRM2 compared to the single RRM1 domain. The RRM1 domain did not interact with U8, C1U3A1, C1U4A1 and the control motif, however dissociation constants in the μ M range were detected for both ligands U5A1U1 and U5A1U2.

The experiments with RRM1-RRM2 as titrate yielded dissociation constants of (323 ± 34) nM for U5A2U1 and (299 ± 28) nM for U5A1U2. Both ligands, which differ only in one nucleotide, show very similar K_D values (both lying within the error limit of each other), thus not allowing us to significantly distinguish between the two binding affinities. The affinity increase due to the presence of the RRM2 domain is about 100-fold with respect to the values obtained for the single RRM1 domain. RRM1-RRM2 binds the the poly U-stretch with a similar high affinity (248 ± 49 nM) as the consensus CPE-containing RNA fragments. For the non-consensus CPE-motif

C1U4A1, however, a considerable ten-fold decrease in binding affinity is observed ($5.4 \pm 0.7 \mu\text{M}$). Interestingly, the presence of the second RRM domain fails to recover the RNA binding activity towards the short C1U3A1 fragment, suggesting that the presence of four uracils is one of the necessary requirements for efficient RNA binding.

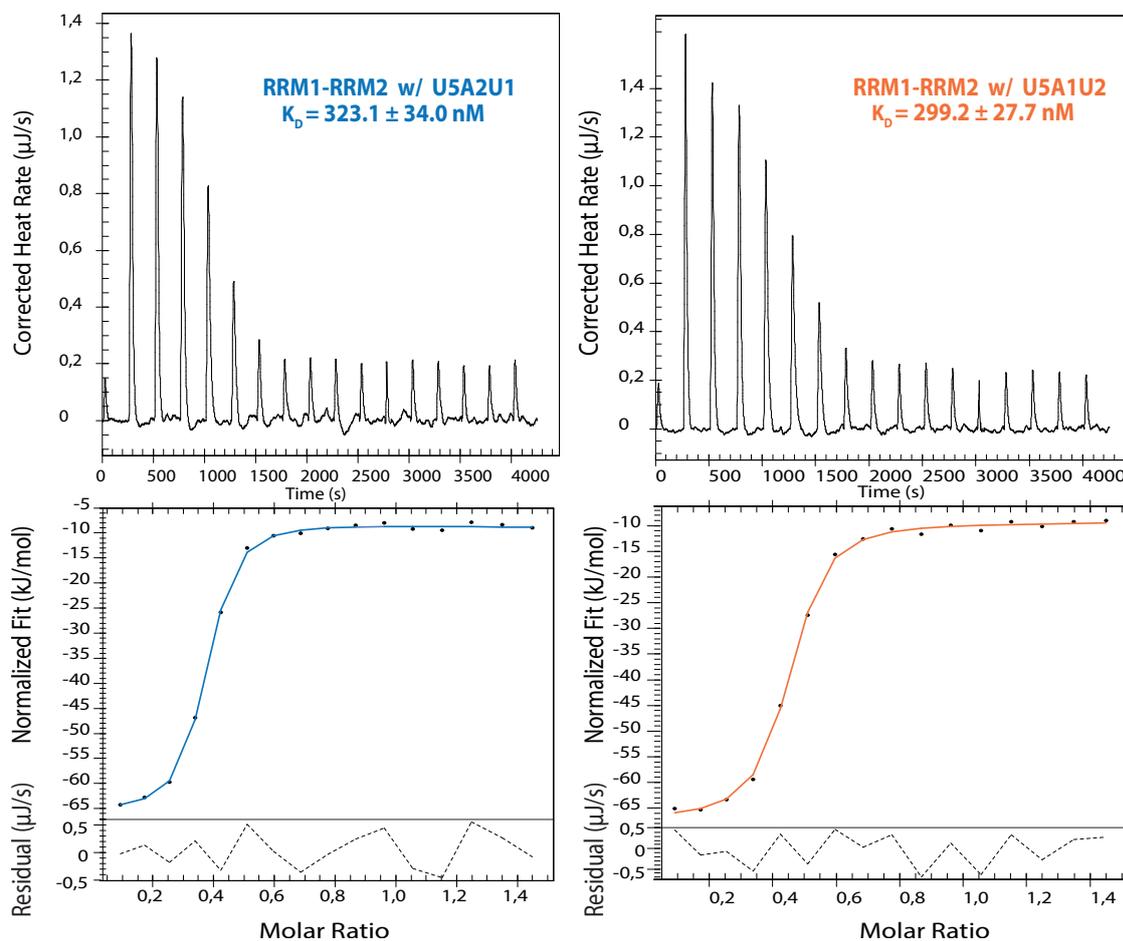


Figure 4.7: ITC curves and affinity values obtained for the titration of CPEB4 RRM1-RRM2 with U5A2U1 (blue) and U5A1U2 (orange) RNAs. Data were acquired at 5°C in Tris buffer pH 7.0. The data were fitted using the independent model assuming a single binding site

As the single RRM2 domain couldn't be obtained, the titration experiments could not be repeated for the second RRM domain. However, from our NMR titration data, which allows the identification of the regions affected by RNA-binding, it is clear that chemical shift perturbations are observed for residues in both domains. Therefore, our data indicate that both RRM domains are important to maintain optimal RNA binding activity and affinity. The improvement of the affinity by 100 times in using the RRM pair is quite remarkable since the RRM2 domain contains only a degenerate RNP motif and was assumed to be a poor RNA binder. Taking into account the increase of

the affinity and the length of the RNA recognized by the RRM1-RRM2 pair suggests that both RRM domains act cooperatively resulting in high affinity binding of the RNA. The data suggests that the role of the RRM2 in the RNA interaction is dual: it increases the global surface of the pair to recognize the RNA with respect to the single RRM1, and helps, adjusting the relative orientation of the two domains, to best accommodate the RNA ligand.

4.1.4 RRM1 and RRM1-2 have the ability to dimerize

It has been previously reported by Lin *et al.*, (2012), that CPEB1 dimerization via its RNA binding regions functions as a regulatory mechanism during cell cycle. We therefore consider the possibility that protein dimerization could also occur in CPEB4.

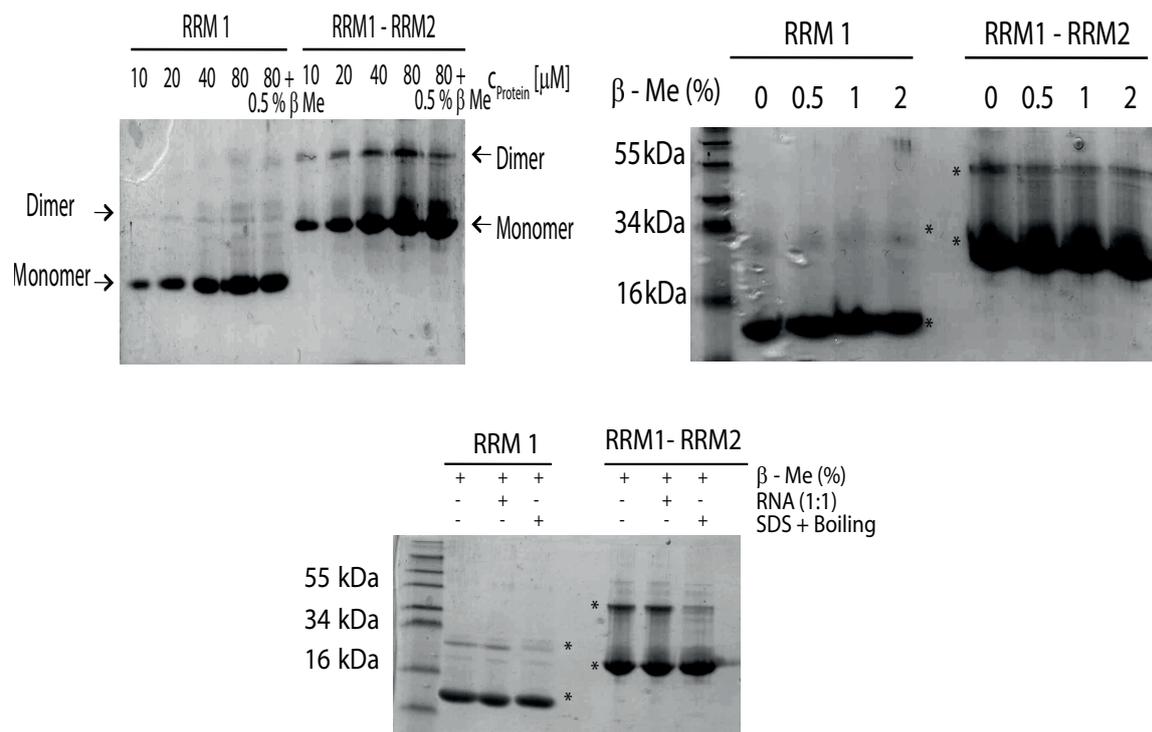


Figure 4.8: Protein samples were analyzed under semi native conditions using SDS 12% PAGE. SDS was not present in the loading buffer and the samples were not boiled unless indicated

The presence of highly populated dimers cannot be detected at the protein concentrations used during gel filtration chromatography and also our NMR data were acquired at low protein concentrations (0.2-0.5 mM), to minimize protein precipitation. At this concentration the main component of the protein solution corresponds to the monomeric protein, as revealed by the relaxation experiments and rotational correlation time measured for the tandem. However, when we performed the ITC experiments to determine the affinities of the different complexes we observed that the stoichiometries for both the interactions of RRM1 and RRM1-RRM2 proteins with the RNA-ligands were close to but smaller than 1. Repetition of the experiments at several temperature (5 °C, 12 °C and 25 °C) and using different buffer solutions (tris, phosphate and bis-tris) yielded similar values and stoichiometries. We attributed

these small discrepancies in the stoichiometries to the presence of a small population of oligomers that -together with the main complex formed by monomers- may also interact with the RNA.

In an attempt to understand the presence of monomers, dimers or other oligomers and to characterize their effect in the interaction with the RNA the proteins were analyzed by semi-native SDS PAGE. Figure 4.8 (*upper left*) shows that when SDS and reducing agent was absent from the loading buffer and the samples were not denatured by high temperature incubation; both RRM1 and RRM1-RRM2 migrate as two species on the gel representing both monomeric and dimeric forms of the proteins. The presence of the dimer is maintained through a series of protein concentrations (10-80 μ M).

Due to the presence of cysteins in the protein, we were interested in analysing whether the presence of di-sulfide interactions could play a role in the formation of the dimeric species. The addition of 0, 0.5, 1 and 2 % of β -mercaptoethanol as a reducing agent to the samples seems to have little effect on the propensity of the proteins to form a dimer, indicating that the dimeric species is not solely maintained through di-sulphide interactions (Figure 4.8, *upper right*). Furthermore, the addition of CPE RNA (1:1) to both the protein samples has no effect on the formation of the dimeric species (Figure 4.8, *lower center*). However, when SDS was added to both samples and the samples were denatured with temperature, a significant reduction in the presence of the dimeric species was observed. This result indicates that the oligomerisation of the proteins is not solely maintained through disulphide linkages and is dependent on the proteins maintaining a proper fold, therefore being mediated by surface electrostatic and/or hydrophobic interactions. Moreover, the proteins maintain the ability to form dimers in the presence of RNA.

4.1.5 Dimerization of RRM domains does not disrupt RNA binding

To further study the potential presence of oligomers at the protein concentration used in ITC experiments (20-50 μM) and to investigate the RNA binding activity of the dimeric species, we used Ion Mobility-Mass Spectrometry (IM-MS) on both constructs in their apo form as well as in complex with RNA.

Ion mobility coupled to mass spectrometry (IM-MS) is a technique that simultaneously separates gaseous ions on the basis of their mass, shape and size. The continuous advances in native mass spectrometry applied to IM-MS have prompted its application in the structural study of biomolecular complexes. Hence, challenging systems in terms of molecular size, complexity and heterogeneity can be transferred to the gas phase and their structural properties can be analyzed. IM-MS can provide valuable information about the occurrence and population of the species present in a sample. This technique has been efficiently applied in the characterization of protein-peptide interactions, in particular in the determination of the presence of several species and their role in the calculation of stoichiometries of complexes.

Ion	observed Mass [Da]	calculated Mass [Da]
RRM1 M6	2150.9	2149.3
RRM1 M7	1843.8	1842.4
RRM1 D9	2867.6	2865.3
RRM1 D10	2580.9	2578.9
RRM1 M1L7	2191.4	2189.9
RRM1 M1L8	2556.5	2554.8
RRM1-RRM2 M8	2294.5	2294.4
RRM1-RRM2 M9	2549.3	2549.3
RRM1-RRM2 M10	2867.8	2867.8
RRM1-RRM2 D12	3823.4	3823.4
RRM1-RRM2 D13	3529.4	3529.4
RRM1-RRM2 D14	3277.4	3277.4
RRM1-RRM2 ML8	3172.0	3172.0
RRM1-RRM2 ML9	2819.7	2819.7
RRM1-RRM2 ML10	2537.8	2537.8
RRM1-RRM2 D2L13	3903.8	3903.8
RRM1-RRM2 D2L14	3625.0	3625.0

Table 4.2: Table of ion-masses observed by IM-MS and theoretically calculated. All species detected were identified based on their m/z ratio.

Each ion was assigned to a given species based on its characteristic mobility and

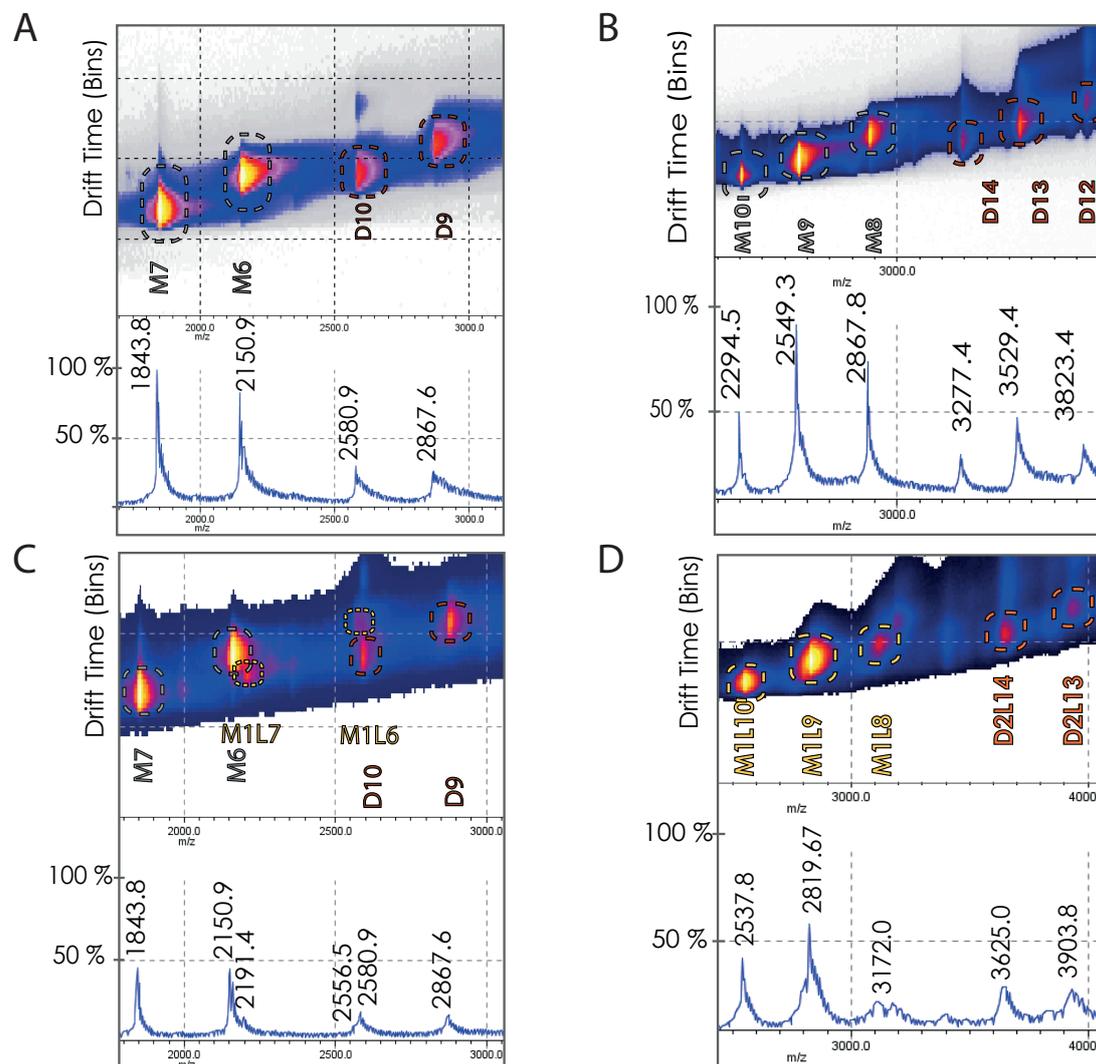


Figure 4.9: A region of the Ion Mobility-Mass Spectrometry data obtained for apo RRM1 (A), apo RRM1-RRM2 (B), RRM1-U5A2U1 complex (C) and RRM1-RRM2-U5A2U1 complex (D). We have unambiguously detected the presence of dimers for both RRM1 (A) and RRM1-RRM2 (B) in their apo state. Consistent with ITC results, for the RRM1-U5A2U1 complex (C) only a minor part of monomeric complex was detected (M1L7 and M1L6); the majority of the species represent the ones already detected for the apo state (A). (D) In the RRM1-RRM2-U5A2U1 sample monomers in complex with one ligand as well as dimers with two ligands were identified.

mass-to-charge ratio. Abbreviations used are M (Monomer, apo state), D (Dimer, apo state), M1L (Monomer with one ligand) and D2L (Dimer with two ligands). Numbers following the species' name indicate the ionization state. For exact ion-masses of both observed and theoretical values please refer to Table 4.2. Using this technique we could

identify in both the apo RRM1 (Figure 4.9 A) and in the apo RRM1-RRM2 (Figure 4.9 B) samples the presence of monomeric as well as dimeric species, consistent with the ITC results and the semi-native SDS PAGE. For the complex of RRM1 with RNA a weak binding affinity was expected. Accordingly, we detected only a minor part of the species in a monomeric complex, the majority remained in its free form (Figure 4.9 C). For the complex with the RRM-tandem domains, the presence of monomeric complexes as well as protein dimers bound to two ligands was detected (Figure 4.9 D). Moreover, no unbound species were detected. The different species (see schematic representation in Figure 4.10) were unambiguously identified based on their specific mass-to-charge ratio (MS) and/or their characteristic drift-times that correlate their size-to-charge ratio (IM).

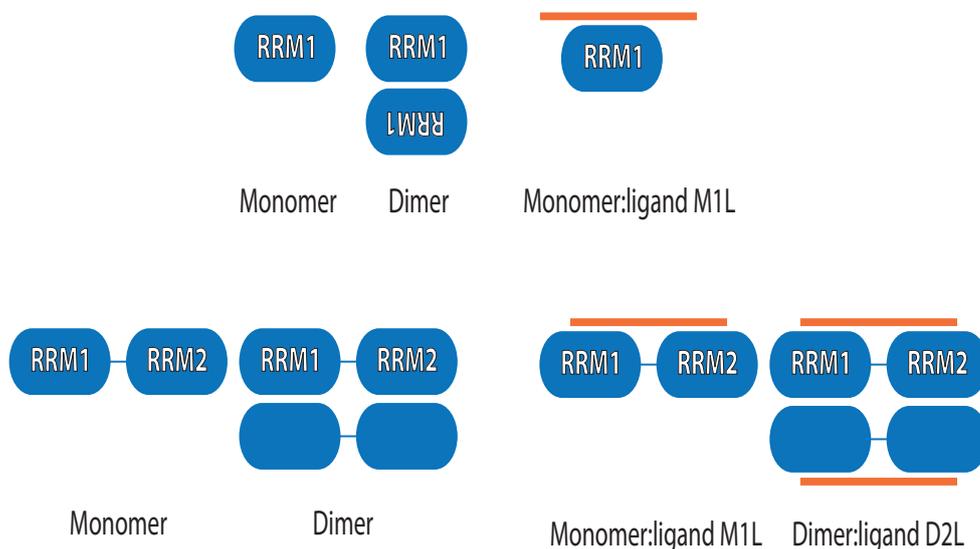


Figure 4.10: A schematic representation of the species identified by IM-MS. The RRM domains are represented in blue ellipses, the RNA octamer as an orange line.

Therefore, IM-MS confirms the hypothesis that indeed dimeric species are present. Perhaps the presence of monomeric and dimeric forms renders the determination of the protein concentration in native conditions inaccurate, explaining the stoichiometry values inferior to one. The results also clearly indicate that only 1:1 protein-RNA complexes are present in both the monomeric and dimeric forms (ML, D2L). The presence of the M1L and D2L species probably contributes to the global affinity measured by the ITC experiments, which is mainly governed by the contribution of the M1L form. The results of the IM-MS experiments with RRM1-RRM2 in complex with RNA reveal that the presence of the ligand does not prevent the formation of dimeric species and does not render the RNA binding site inaccessible. We therefore conclude that the dimerization surface of the RRM1-RRM2 is separate from its RNA-binding interface.

4.1.6 Identifying the RNA binding site of RRM1-RRM2 via NMR

To characterize the RNA binding properties of the RRM1-RRM2 construct to the two CPE containing ligands, U5A1U2 and U5A2U1, and to investigate whether there are different binding modes between the consensus CPE motifs, NMR chemical shifts of amides were monitored upon titration of the U5A1U2 and U5A2U1 to the ^{15}N -labeled RRM1-RRM2 domain. The chemical shift perturbations regarding their range and also the residues affected showed no significant difference between U5A1U2 and U5A2U1 (Figure 5B). We therefore assume that the number of adenosines within the CPE-motif does not alter the binding mode. Binding kinetics observed in the NMR titrations of RRM1-RRM2 are in fast to intermediate range on the NMR chemical shift time scale as some residues disappear and reappear with increasing U5A1U2 and U5A2U1 concentration (e.g. Arg⁵⁵⁹; Figure 4.11 (lower left)).

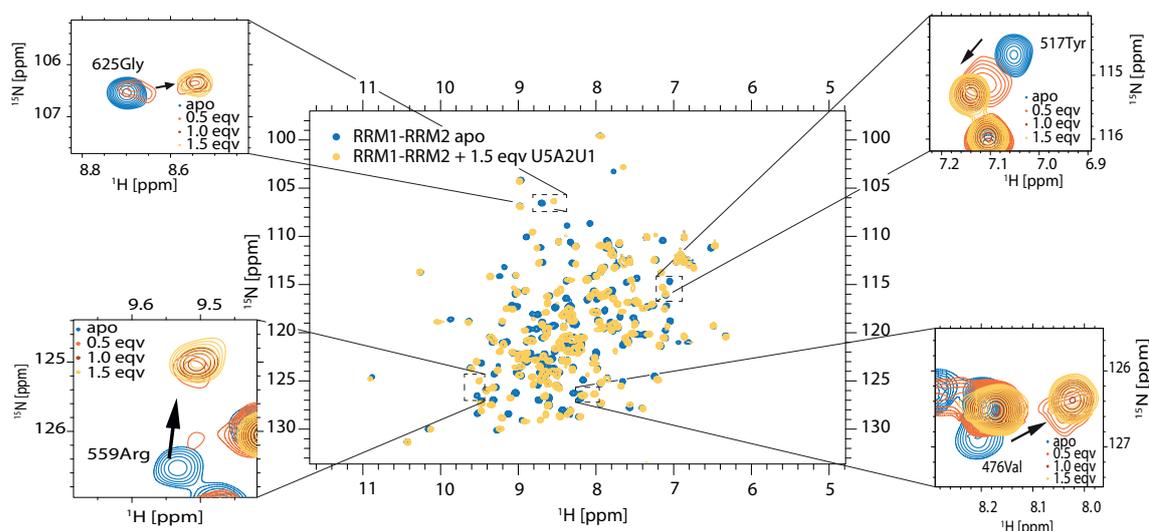


Figure 4.11: Superimposition of ^1H , ^{15}N - HSQC spectra of RRM1-RRM2 free (blue) and in complex with U5A2U1 (yellow) 1.5 molar equivalents. Zooms in various regions of the spectra are displaying all titration points. Spectra are colored as: blue (free), 0.5 (orange), 1.0 (dark-red) and 1.5 (yellow) molar equivalents.

For both RNA ligands we observe chemical shift perturbations in the canonical RNP motifs of RRM1. According to the secondary structure predictions, based on the ^{13}C secondary chemical shifts, and the homology modeling we have performed, the affected residues lie on the β_1 and β_3 strand. In addition large chemical shift perturbations are also observed for residues lying on β_2 (Trp⁵⁰², Lys⁵⁰⁵) and a C-terminal region of RRM1. These C-terminal residues are predicted to form the β_4 -strand (affected residues: Gln⁵⁵⁷, Ile⁵⁵⁸, Arg⁵⁵⁹) and an additional β -strand β_4' (affected residues:

Ser⁵⁴⁷, Thr⁵⁵⁰). When mapping the most affected residues of RRM1 on the homology model, they cluster on a region on the β -sheet (cf. Figure 4.12), which corresponds to the canonical RNA binding surface of RRM domains.

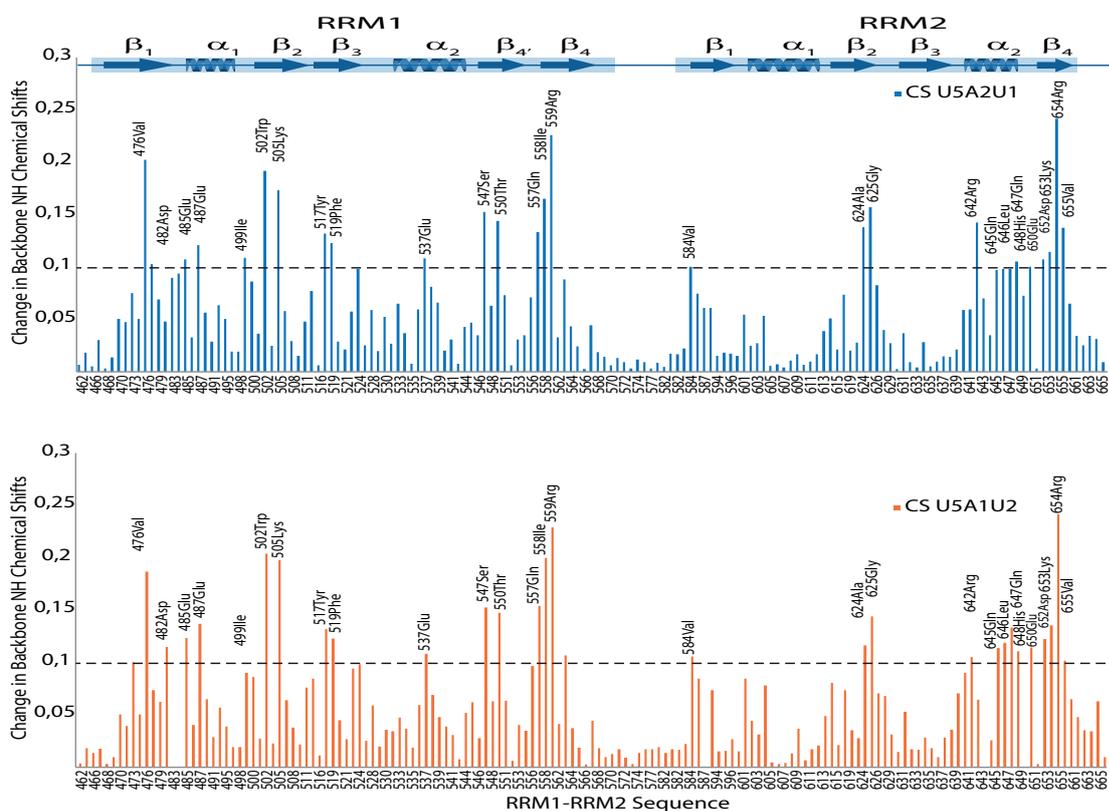


Figure 4.12: Chemical shift perturbations of RRM1-RRM2 domains upon binding to U5A2U1 (*upper panel*) and U5A1U2 (*lower panel*). The CSPs are plotted versus residue numbers. The dashed line in grey shows a cut-off $\Delta\delta$ 0.1 ppm (average chemical shift perturbation over all resonances + standard deviation). The CSPs for both RNA ligands do not show any significant differences. Within RRM1 the largest shift perturbations are found in residues that are supposed to lie on the β -sheet, for RRM2 however the clearly cluster in the C-terminal part. The RNP2 of RRM2 shows shifts just above the cut-off, which suggests that RNP2 isn't the main binding site for RRM2, instead the C-terminal positively charged residues seem to play an essential role.

For the RRM2 domain, which only features the RNP2 consensus sequence on β_1 , the strongest chemical shift perturbations were detected in its C-terminal region containing residues with positively charged side groups (e.g. His⁶⁴⁸, Lys⁶⁵³, Arg⁶⁵⁴). This indicates that these regions are affected by binding to the RNA fragments either by engaging in direct contacts to the RNA or resulting from conformational changes induced by the complex formation.

4.1.7 Crystallization Trials CPEB4 RRM1-RRM2

Screening for crystallization conditions were carried out at different protein concentrations, temperatures (4°C and 20°C) and various sample - screening condition ratios. Crystals of CPEB4 RRM1-RRM2 (free form and in complex with RNA) were grown in sitting drops at 20°C by mixing 200-300 nL of concentrated CPEB4 RRM1-RRM2 (10.8 mg/ml) with 100 nl of a reservoir solution consisting of 1.3 - 1.50 M sodium formate, 90-100 mM Cacodilate pH 6.4-6.6 and 30-50 mM Cobalt Chloride. Crystallization trials for the RRM1-RRM2 CPE complex were set up with a 5-Methyluridine 5'-Monophosphate- modified* RNA (U*UU UUA UU*) added in a two-fold excess. The modified RNA is less prone to degradation by RNAses and therefore enhancing the probability of the incorporation of the intact ligand during crystal growth. Crystals appeared within 4 days and grew in 3-4 weeks. Before freezing the crystals they were soaked in a cryo-protectant solution consisting of 15% PEG 500 MME, 85% reservoir solution.

Data were collected at ALBA Synchrotron Light Facility, BL13-XALOC beamline on a Pilatus 6M detector. The most promising data yield a resolution of 3.1 Å. Data processing with the XDS program package [77] indicates that the crystals belong to space group P4₁22 with unit cell dimensions of $a = b = 105.67$ Å, $c = 333.61$ Å, $\alpha = \beta = \gamma = 90^\circ$. Matthews coefficient indicates that there should be 8 molecules RRM1-RRM2 per asymmetric unit. The self-rotation function indicates that the arrangement of the molecules in the asymmetric unit leads to non-crystallographic symmetry in the asymmetric unit (see Figure 4.13). Molecular Replacement trials using the NMR structures of CPEB4 RRM1-RRM2 apo form (PDB code: 2MKJ) and in complex with RNA (PDB code: 2MKI), despite the 100% identity, have not been carried out successfully. The difficulties encountered here might be partly due to the large unit cell, which requires the search for eight copies of the molecule.

Since molecular replacement couldn't be carried out successfully, our objective was to obtain phase information by anomalous dispersion methods. To this end crystallization trials with a selenomethionine derivative sample were set up, but crystal growth could not be reproduced. Heavy Atom derivative screening [78] using native crystals was performed for the following compounds: Potassium tetrachloroplatinate(II), Gold(III) chloride and Thimerosal (Ethyl(2-mercaptobenzoato-(2-)-O,S) mercurate(1-) sodium). After back-soaking, the incorporation of the heavy-atom compounds was checked using the fluorescence detector at the beam line. The tunable beam line allows to adjust the wavelength to the known absorption edge of the specific compound so that the incident X-ray wavelength is optimized for the anomalous signal. The fluorescence detector shows that the heavy atom is present in the crystal, but does not indicate whether the compound is bound to defined sites in the protein. Derivatization was only successful when soaking the crystals in Thimerosal. The soaked crystals diffracted to lower resolution than the native crystals did (3.1 Å *vs.* 5.0 Å), the space group and the cell dimensions were the same as for the native crystal. Analysis of the data

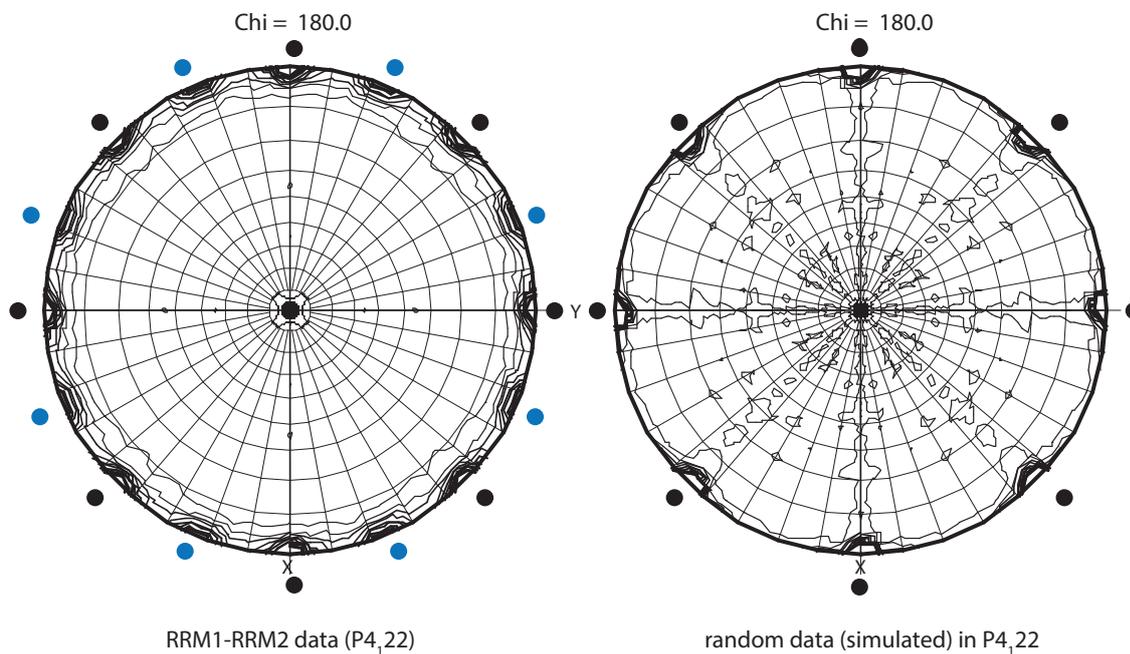


Figure 4.13: Self rotation function for the acquired data of a RRM1-RRM2 native crystal (*left*) and randomly generated data in the same space-group (*right*). Crystallographic symmetry axes are indicated by black dots. For the RRM1-RRM2 data we additionally observe non-crystallographic symmetry axes indicated by blue dots.

showed that the heavy-atom occupancy is very low and therefore phasing could not be performed successfully.

4.2 Pin1 a key factor in CPEB1 degradation

4.2.1 Pin1 WW domain interacts with CPEB prior to phosphorylation

Recently, CPEB1 has been identified as a target for Pin1. Since Pin1 target-recognition is mainly based on p(Ser/Thr)-Pro motif interaction, the association of Pin1 with CPEB1 prior to its phosphorylation has been an unexpected result [53]. The amino terminal portion of CPEB1 features no obvious structural motif. Within the N-terminal region of CPEB1 residues 48-183 have been identified to interact with Pin1 in a phosphorylation independent manner. Furthermore depletion of residues 211-290 abrogated the interaction, but due to the techniques applied size limitations precluded the direct detection of the interaction with the smaller fragment.

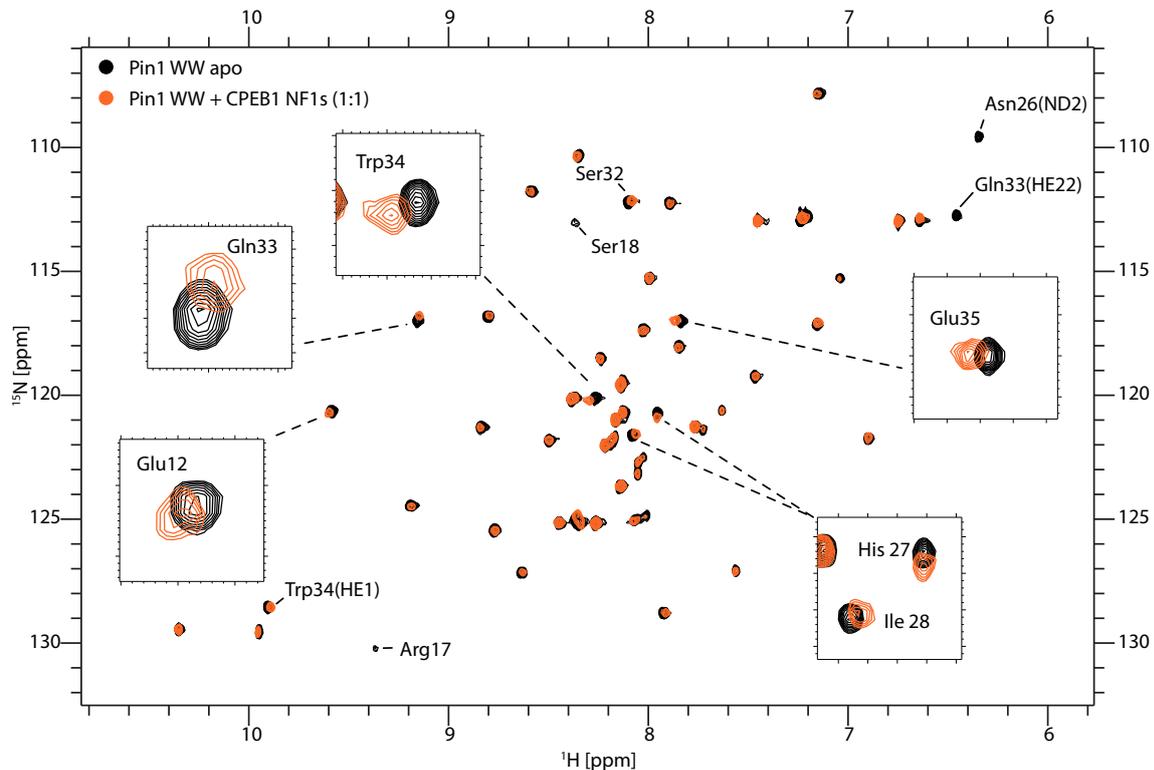


Figure 4.14: Superimposition of ^1H , ^{15}N - HSQC spectra of Pin1 WW domain (black) and in complex with CPEB1 NF1s (orange) ~ 1 molar equivalents.

To gain insight into which regions play an important role in the CPEB1-Pin1 interaction, three constructs of the following N-terminal fragments NF1 (48-183), NF1s (88-183) and NF2 (196-293) using His-tagged fusion expression vectors were prepared. The boundaries were chosen similar to the constructs used by Nechama *et al.*, how-

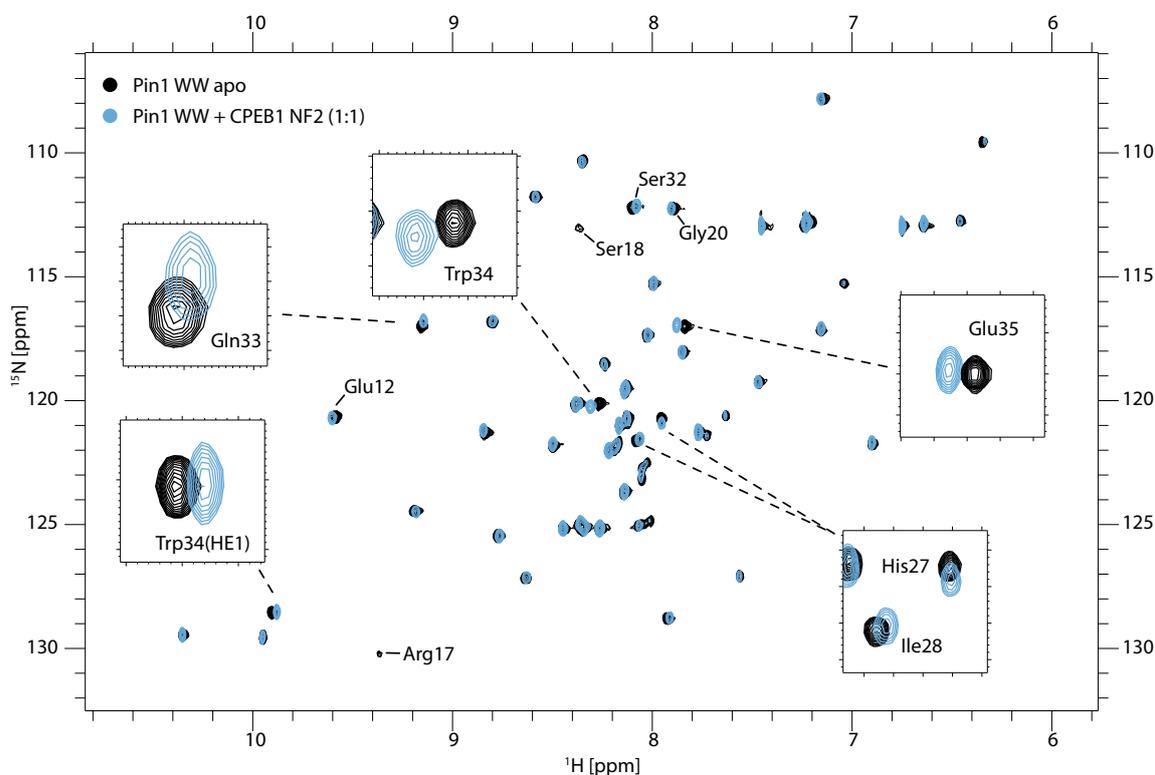


Figure 4.15: Superimposition of ^1H , ^{15}N - HSQC spectra of Pin1 WW domain (black) and in complex with CPEB1 NF2 (blue) ~ 1 molar equivalents.

ever one objective was to include the majority of the N-terminal Cdc2-phosphorylation sites: 138, 144, 210 and 248. The main part of the protein obtained in the expression cultures was in inclusion bodies, but the small amount obtained in the soluble fractions could be purified by Ni-affinity chromatography. Protein purified under denaturing conditions (6M guanidium chloride), except for 48-183, was dialyzed against an 100 mM Tris pH 7.0, 400 mM L-Arg, 20% glycerol 2 mM DTT buffer maintaining the protein at a maximum concentration of $\sim 300 \mu\text{M}$. Changing the buffer to NMR suitable conditions resulted in precipitation of the sample and only a maximum concentration of $\sim 100 \mu\text{M}$ could be obtained.

As previously reported, the interaction of the unphosphorylated CPEB1 is dependent on Pin1 WW domain. Therefore the work was focused on identifying the N-terminal regions of CPEB1 involved in the interaction with Pin1 WW domain. In order to directly detect the interaction of the N-terminal fragments with Pin1 WW domain, HSQC experiments were acquired. NMR chemical shifts of the ^{15}N -labeled Pin1 WW domain in its free state were compared to a sample containing Pin1 WW domain and NF1s or NF2 in a 1:1 ratio. Due to the limited stability of NF1s and NF2, it has not been possible to obtain stable samples containing an excess of NF1s or NF2 with

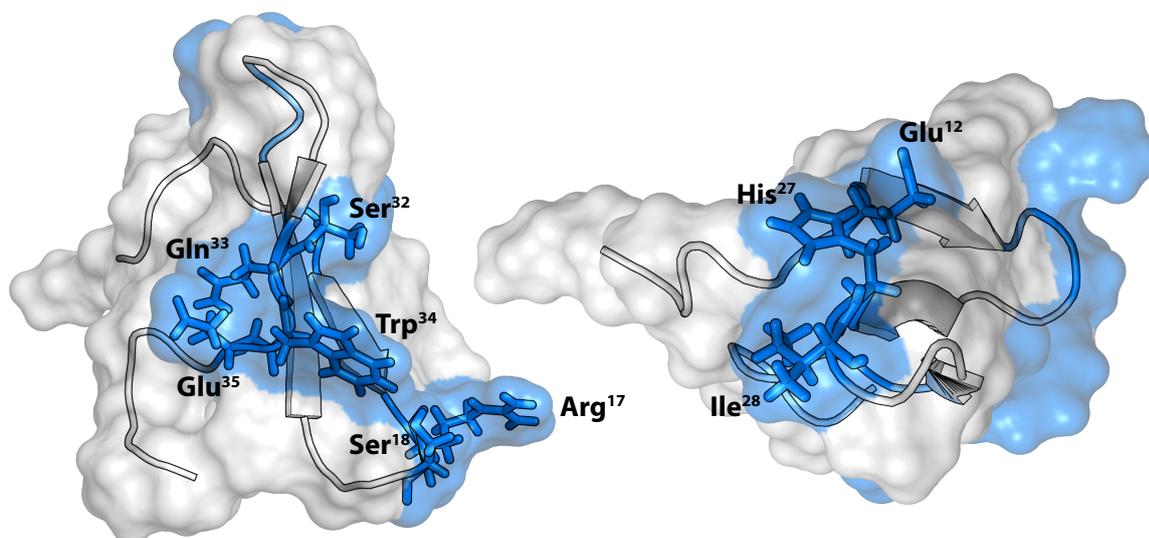


Figure 4.16: NH chemical shift perturbations are mapped on the solution structure of Pin1 WW domain (PDB entry:2M8I).

respect to Pin1 WW.

Despite the expected low affinity of the CPEB1-Pin1 association, we observe similar chemical shift perturbations introduced by both ligands NF1s and NF2 (see Figures 4.14 and 4.15 respectively). The only difference observable occurs in the spectrum of the complex with NF1s, in which two amide side-chain resonances of Asn²⁶ and Gln³³ disappear indicating that these side chains are affected in the presence of the ligand and the binding occurs in the intermediate exchange regime on the NMR time scale. Affected regions comprise residues in the beginning of the first β -strand (Glu¹²) and second loop (His²⁷, Ile²⁸), which are in spacial proximity to each other, and furthermore residues at the end of the third β -strand (Ser³², Gln³³, Trp³⁴ and Glu³⁵). It is noteworthy that the resonances from Arg¹⁷ and Ser¹⁸ disappear when CPEB1 associates with Pin1 WW. These two residues form part of the first loop between β_1 and β_2 , which is rather flexible in the free state (low relative intensity of the resonances) and becomes more rigid upon binding.

Upon titration with phosphorylated CPEB1 peptides (cf. section 4.2.2), the signals corresponding to Arg¹⁷ and Ser¹⁸ disappear (at a ligand:protein ratio 1:1). The resonances only reappear when the interaction is almost saturated. For the Pin1 WW : NF1s and NF2 sample, respectively, we assume that the interaction is not saturated. Interestingly, the same region at the end of the third β -strand and as well the residues from the first loop have been observed to be involved in the interaction of Pin1 WW domain and different phosphorylated (S/T)P motifs. In this case Arg¹⁷ forms a phosphate-binding pocket, and Trp³⁴ contacts the proline which is located directly downstream of the p(S/T)P motif. The second region affected by NF1s and

NF2, respectively, including residues in the beginning of the first β -strand and second loop, also experience minor chemical shift perturbations upon binding of p(S/T)P motifs containing ligands, however, those residues do not directly contact with the phosphorylated ligand. ITC experiments performed with Pin1 WW domain (titrant) and NF1s and NF2 (titrate), respectively, indicate binding affinities in the high μM range. Since the interaction could not be saturated due to the limitations of stability at high concentrations of the titrant, the exact affinity value couldn't be determined and therefore only a affinity range is provided.

4.2.2 Phosphorylation enhances binding affinity of Pin1 to CPEB1

In total there are six described Cdc2 phosphorylation sites within CPEB1. All of them contain a pSP-motif, which makes them potential Pin1 binding sites. It has been demonstrated that in *Xenopus* oocytes as well as in mammalian cells Pin1 interacts with the phosphorylated CPEB1. In order to characterize the interaction of phosphorylated CPEB1 with Pin1, several peptides containing the aforementioned phosphorylation sites were prepared by Fmoc SPPS and ITC binding assays were performed. Although Pin1 WW domain is known to be the substrate recognition module of the protein, inter domain communication between the WW and the catalytic domain are important for in vivo activity of Pin1. In order to detect any changes in binding affinity when using the isolated WW domain compared to the Pin1 full-length protein, all experiments were performed with both constructs (cf. Table 4.3).

	Pin1 WW	Pin1 FL
CPEB1 phospho peptide	K_D [μ M]	K_D [μ M]
CPEB1 pS138	71.3 ± 9.3	84.3 ± 10.6
CPEB1 pS144	75.4 ± 12.7	83.5 ± 11.5
CPEB1 pS184	28.6 ± 3.1	31.4 ± 2.9
CPEB1 pS210	30.3 ± 2.4	32.3 ± 5.9
CPEB1 pS248	47.6 ± 7.6	42.4 ± 8.7
CPEB1 pS423	157.7 ± 21.8	160.4 ± 25.6

Table 4.3: Binding affinities between Pin1 WW domain, Pin1 full-length and the six described Cdc2-phosphorylation sites of CPEB1 measured by ITC.

Binding affinities measured of the Pin1 WW domain, Pin1 full-length - CPEB1 pSP interaction lie in the low μ M range, which is typical for group IV WW domains. An exception to this is CPEB1 pS423, which is located in the linker region of the RRM domains. The affinity of this interaction is considerably lower than for the interactions with the N-terminal phosphorylation sites. No significant changes in binding affinity could be observed between the full-length protein and the isolated WW domain. Stoichiometric values for all complexes for both constructs determined clearly indicate a one to one interaction. An exception to this observation is the complex of Pin1

4 Results

full-length with CPEB1 pS210, where an n-values of 0.85 was obtained. This small discrepancy might indicate that for this complex, apart from the 1:1 protein:ligand population, a smaller population with a 1:2 interaction is present. Interestingly, CPEB1 pS210, which is known to be essential for CPEB degradation, together with pS184 have been identified to be the two preferred binding site for Pin1 *in vitro*.

Comparing the phosphorylation dependent interaction to the one with unphosphorylated CPEB N-terminal fragments, the results clearly indicate that the affinity is increasing about 10 fold. Furthermore ITC experiments as well as NMR titration studies show that for unphosphorylated versions of several CPEB1 peptides no interaction of the ligands is detectable, neither with the isolated WW domain nor with the full-length protein.

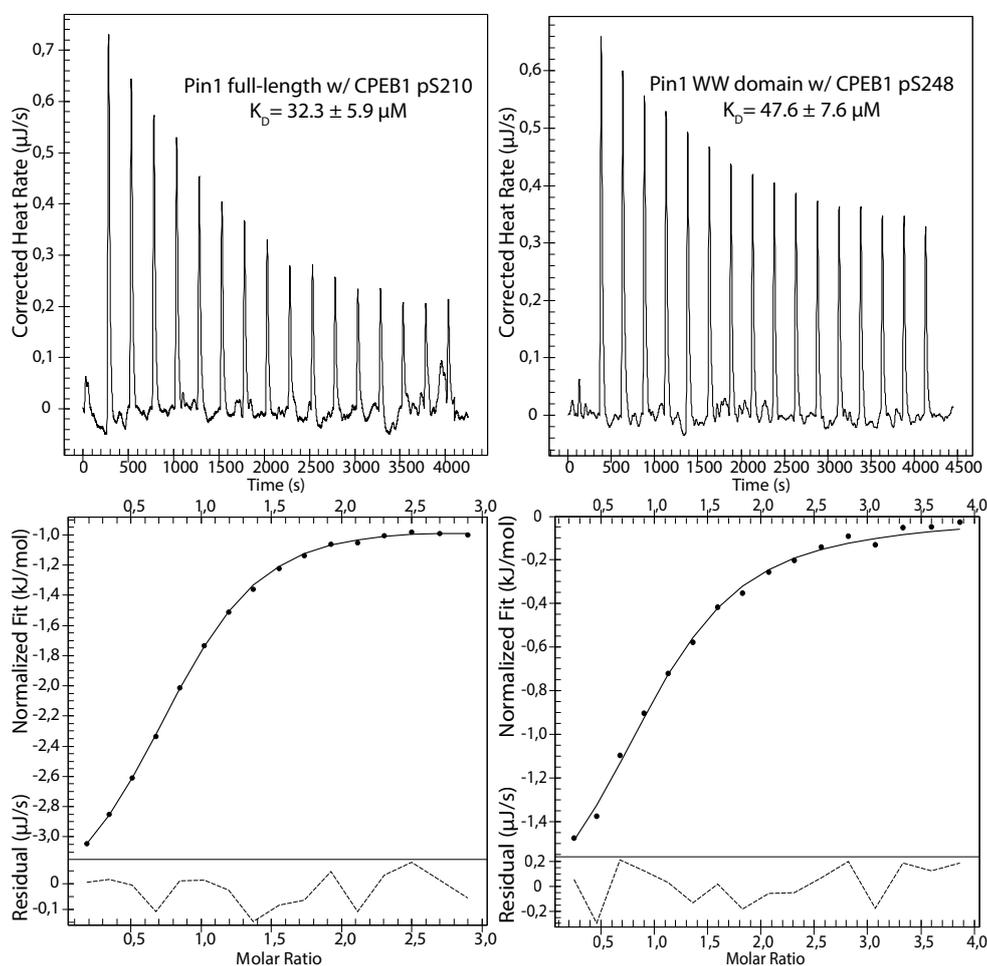


Figure 4.17: ITC curves and affinity values obtained for the titration of Pin1 full-length with CPEB1 pS210 (*left*) and Pin1 WW domain with CPEB1 pS248 (*right*).

In order to confirm the presence of Pin1 in complex with two CPEB1 pS210 ligands we applied Ion Mobility-Mass Spectrometry on the free Pin1 and in complex with CPEB1

pS210. Each ion was assigned to a given species based on its characteristic mobility and mass-to-charge ratio. Abbreviations used are M (Monomer, apo state), D (Dimer, apo state), M1L (Monomer with one ligand) and M2L (Monomer with two ligands). Numbers following the species' name indicate the ionization state. Exact ion-masses of both observed and theoretical values are provided in Table 4.4.

Ion	observed Mass [Da]	calculated Mass [Da]
Pin1 M7	2290.1	2290.2
Pin1 M8	2617.7	2617.2
Pin1 M1L7	2851.1	2851.2
Pin1 M1L8	2494.8	2494.8
Pin1 M2L8	2699.6	2699.6

Table 4.4: Table of ion-masses of Pin1 full-length apo and in complex with CPEB1 pS210 observed by IM-MS and theoretically calculated. All species detected were identified based on their m/z ratio.

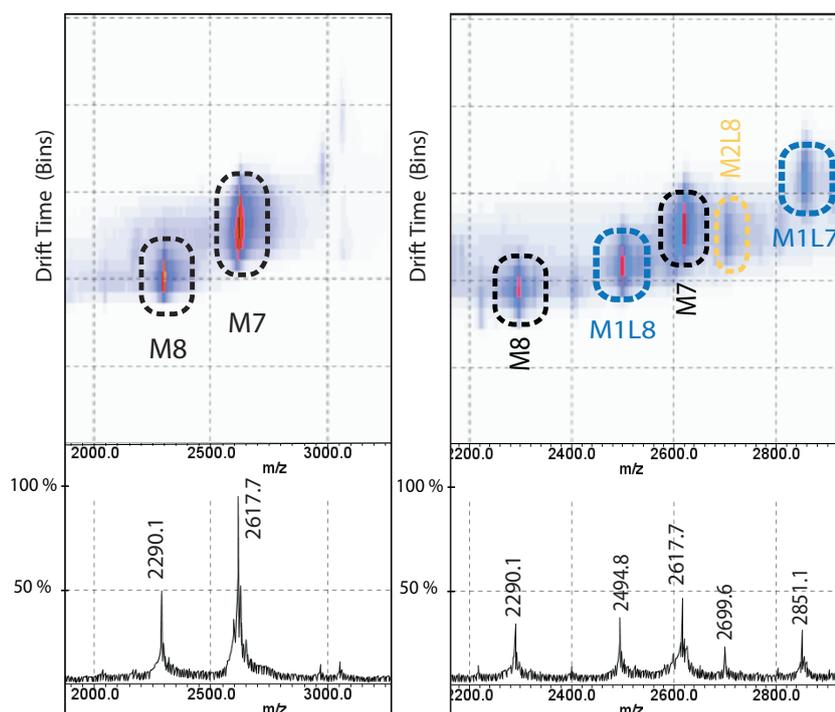


Figure 4.18: A region of the Ion Mobility-Mass Spectrometry data for Pin1 full-length in its apo form (*left*) and in complex with CPEB1 pS210 (*right*). In the free state solely monomers are detected. In complex two types of species are detectable: monomeric Pin1 bound to one ligand (ions M1L7 and M1L8) and in complex with two ligands (ion M2L8).

For the apo Pin1, ions corresponding to the monomeric species were detected (cf. Figure 4.18, *left*). For the complex, apart from the ions resulting from the apo state, we could identify monomer in complex with one ligand (M1L) as well as with two molecules of CPEB1 pS210 (M2L) (cf. Figure 4.18, *right*). The stoichiometry of the complex obtained by ITC suggests that the majority of the species is a 1:1 complex and only a minor population represents a complex with two ligands bound. Since for this peptide chemical shift perturbations are observed not only for the substrate binding site in the WW domain, but also in the catalytic domain, we assume that these perturbations represent a set of species. One species with the ligand bound to the the WW domain and another with the peptide bound to the catalytic side. Ion Mobility results and the stoichiometry obtained by ITC consistently indicate that a species where binding of two ligands simultaneously to the substrate binding site and the PPIase site of the catalytic domain occurs, represents a minority.

4.2.3 Identifying the binding site for CPEB1 phosphopeptides within Pin1

The results of our ITC titration study show that *in vitro* all of the CPEB1 peptides containing the six pSP-motifs are able to bind Pin1 with different affinity values between 30-85 μM for the N-terminal phosphorylation sites and 157 μM for the phosphorylation site located in the linker region between the two RRM domains. Since affinity values do not show a significant difference between the binding to the isolated WW domain and to the full-length protein, we have been particularly interested in whether chemical shift perturbations are also occurring within the catalytic domain. To investigate further the binding properties of these interactions and potential differences in binding modes, ^{15}N amide chemical shift perturbations within both Pin1 WW domain and Pin1 full-length were monitored upon titration of the phosphorylated CPEB1 peptides.

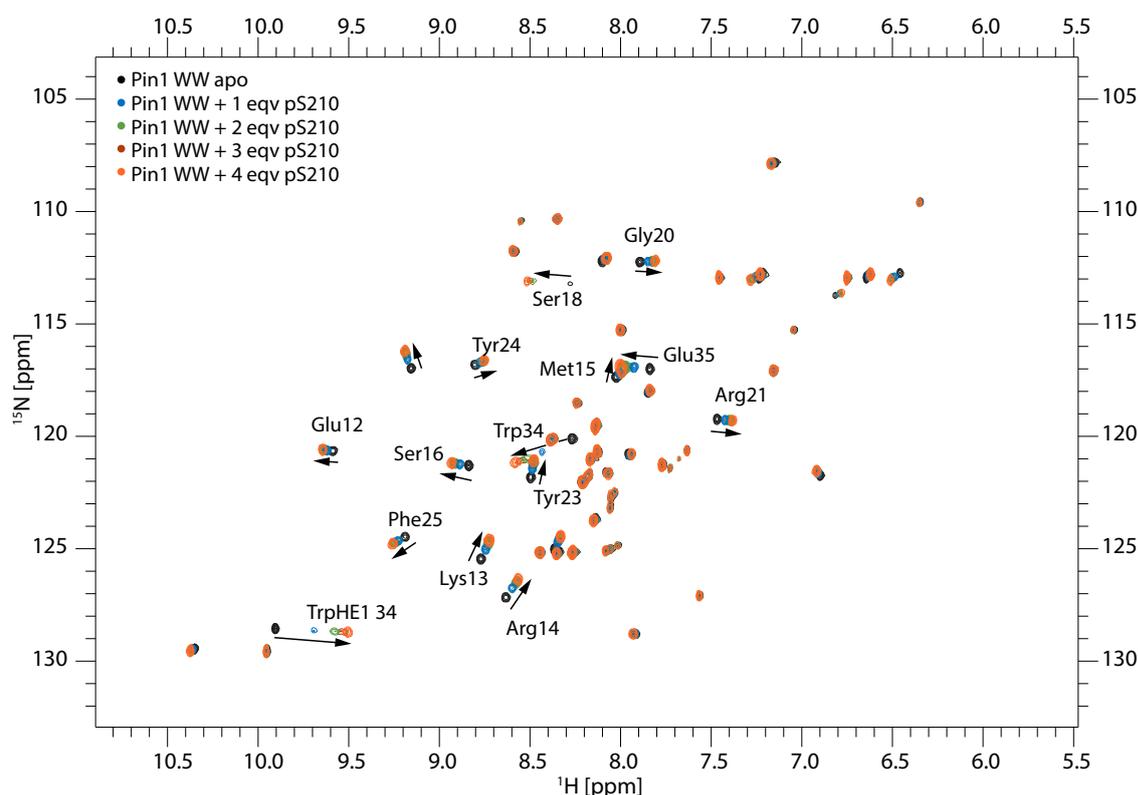


Figure 4.19: Superimposition of ^1H , ^{15}N - HSQC spectra of Pin1 WW domain (black) and in complex with CPEB1 pS210. Titration steps with 1-5 molar equivalents are shown in blue, green, red and orange, respectively. The most affected residues upon binding are labeled.

For the titrations experiments using the WW domain as a titrate, chemical shift perturbations detected display a similar pattern for all ligands tested (cf. Figure 4.20).

Affected residues are located mainly at the end of the β_1 -strand (Ser¹⁶, the loop region between β_1 and β_2 (Ser¹⁸, Gly²⁰), the β_2 -strand (Arg²¹, Tyr²³, Phe²⁵) and the residues at the end of the β_3 -strand (Trp³⁴, Glu³⁵). Residues of the first loop are quite flexible (cf. section 4.2.5) and therefore couldn't be detected in the free state. Upon binding the internal motion gets restricted, but still only Ser¹⁸, but not Arg¹⁷ and Ser¹⁹ could be detected. Since Ser¹⁸ displays a very distinct chemical shift perturbation and Arg¹⁷ has been shown to be important for binding: the side-chain and the backbone amide form the phosphate binding pocket in complex with different pSP-motifs. Furthermore Arg¹⁷Ala mutant decreases affinity by six-fold [51]), we assume that also for the CPEB1 complexes Arg¹⁷ plays an important role in binding. Apart from the complex with CPEB1 pS184 the most pronounced perturbation observed is for Trp³⁴ indicating its presence in the direct vicinity of the binding site. Upon titration with CPEB pS184 additional chemical shift perturbations are observed for Asn²⁶ and His²⁷.

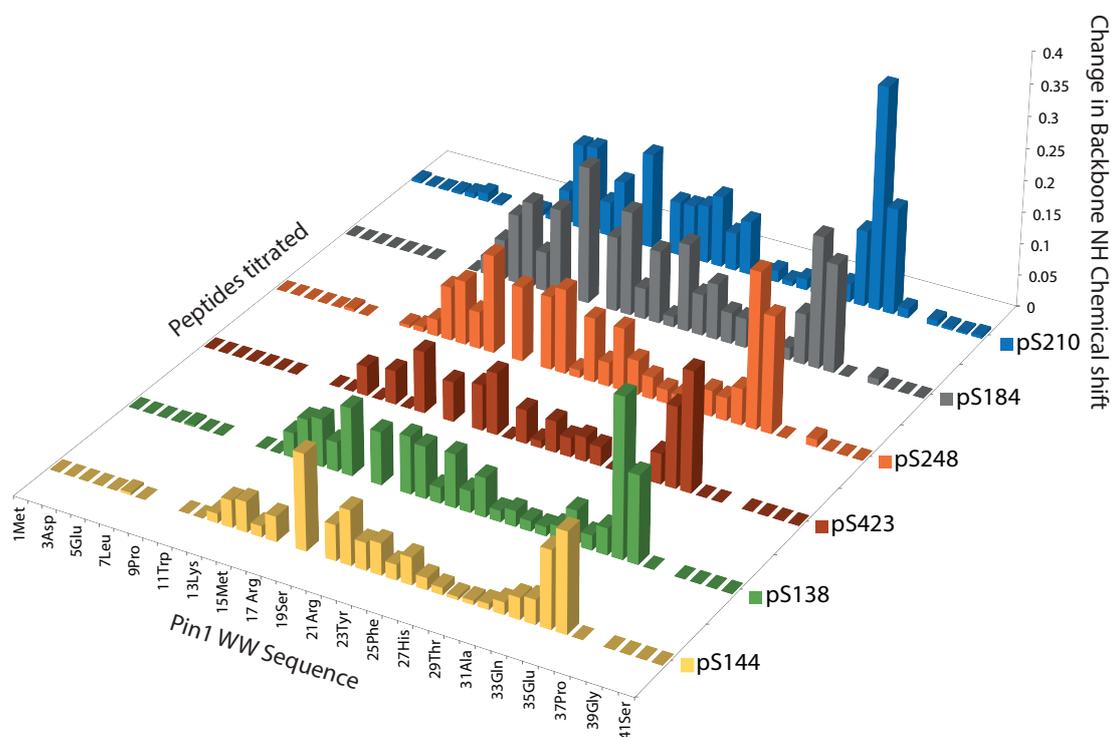


Figure 4.20: Chemical shift perturbations of Pin1 WW domain upon binding to CPEB1 pS210 (blue), pS184 (grey), pS248 (orange), pS423 (red), pS138 (green) and pS144 (yellow). The CSPs are plotted versus residue numbers. The CSPs for all phospho peptide ligands show a similar pattern.

Binding kinetics observed in the NMR titrations are in fast to intermediate range on the NMR chemical shift time scale since the majority of the chemical shifts change continuously upon titration. Only a few disappear and reappear upon saturation of the interaction (see Figure 4.19, for example Trp³⁴ H_ε and Ser¹⁸).

All experiments were repeated using ^{15}N -labeled full-length Pin1 as a titrate. Chemical shift perturbations observed from residues corresponding to the WW domain showed no significant changes in comparison to the results obtained when using the isolated WW domain (see Figure 4.20 and 4.22). This indicates that the binding mode observed upon binding of the substrate to the isolated WW domain does not significantly change compared to the full-length protein.

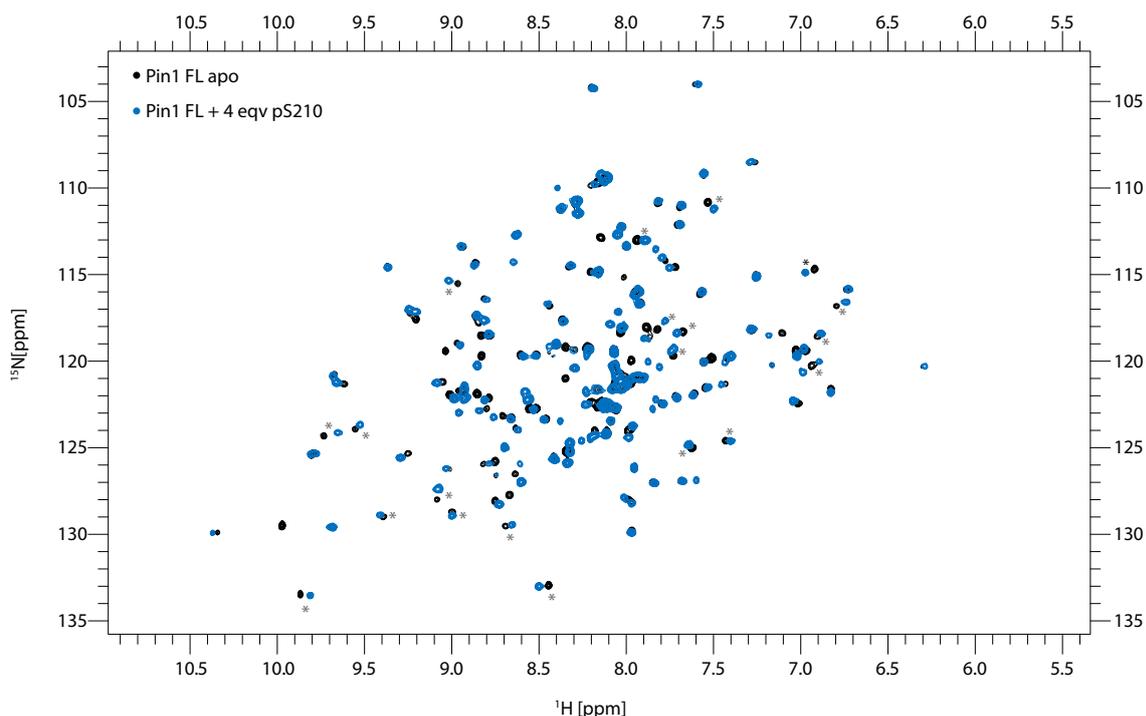


Figure 4.21: Superimposition of ^1H , ^{15}N - HSQC spectra of Pin1 full-length in its apo state (black) and in complex with CPEB1 pS210 (4 equivalents, blue). In contrast to the rest of the pSP motifs within CPEB1, CPEB1 pS210 is causing chemical shift perturbations not only in the WW domain, but also in the catalytic domain. Affected residues from the catalytic domain are marked with an asterisk.

Interestingly CPEB1 pS210 is the only ligand causing significant perturbations also within the catalytic domain. Mapping the observed chemical shift perturbations on the crystal structure, small perturbations at the domain interface, mainly on β_6 and α_4 , are detected. More distinct perturbations are observed for residues within the catalytic loop (Ser⁷², Ser⁷³, Trp⁷⁴), the PPIase active site (Leu¹²², Gly¹²³, Phe¹³⁴) and the proline substrate binding pocket (Gly¹²⁸) (see Figures 4.21 and 4.22).

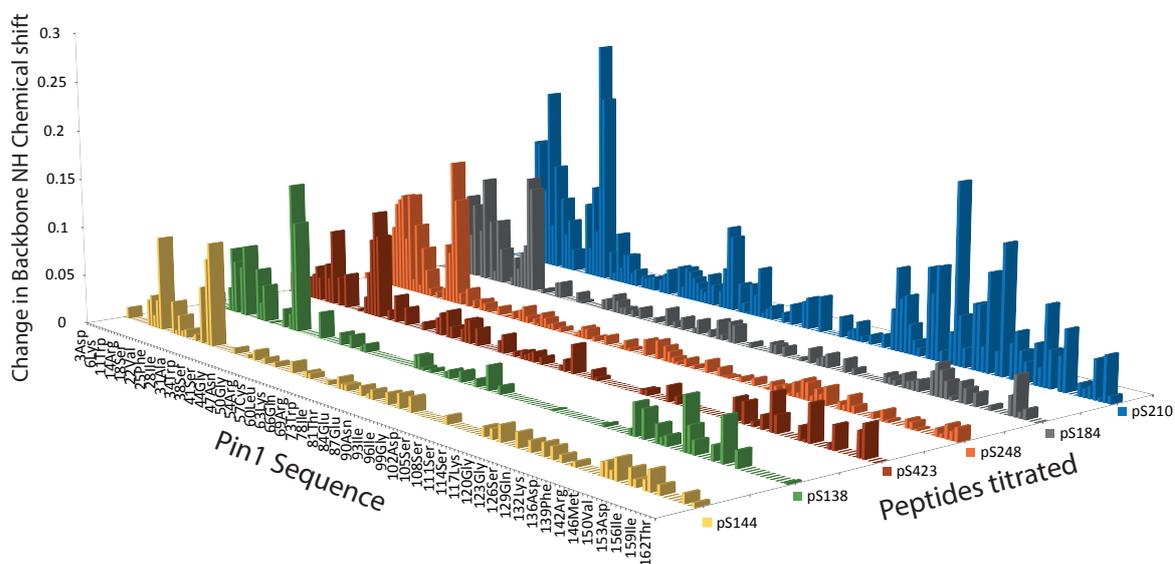


Figure 4.22: Chemical shift perturbations of Pin1 full-length upon binding to CPEB1 pS210 (blue), pS184 (grey), pS248 (orange), pS423 (red), pS138 (green) and pS144 (yellow). The CSPs are plotted versus residue numbers. Within the catalytic domain significant chemical shift perturbations are only observed for the complex with CPEB1 pS210 peptide.

4.2.4 Phosphate group directly involved in binding

In order to illustrate the effect of the phosphate group within the CPEB1 peptides on Pin1 binding, 1D- ^{31}P titrations were performed. The peptide samples were equilibrated in tris buffer and supplemented with 0.2 mM phosphate buffer as a ^{31}P -reference. Pin1 WW domain, or in case of CPEB1 pS210 also the full-length protein, was added to the peptide sample until saturation of the interaction was reached. As shown in Figure 4.23, the ^{31}P -reference peak doesn't change its chemical shift upon titration of the protein, whereas the CPEB1 pS peak shifts downfield during the titration experiments. The chemical shift perturbations of the phosphate-group in the different CPEB peptides are listed in Table 4.5.

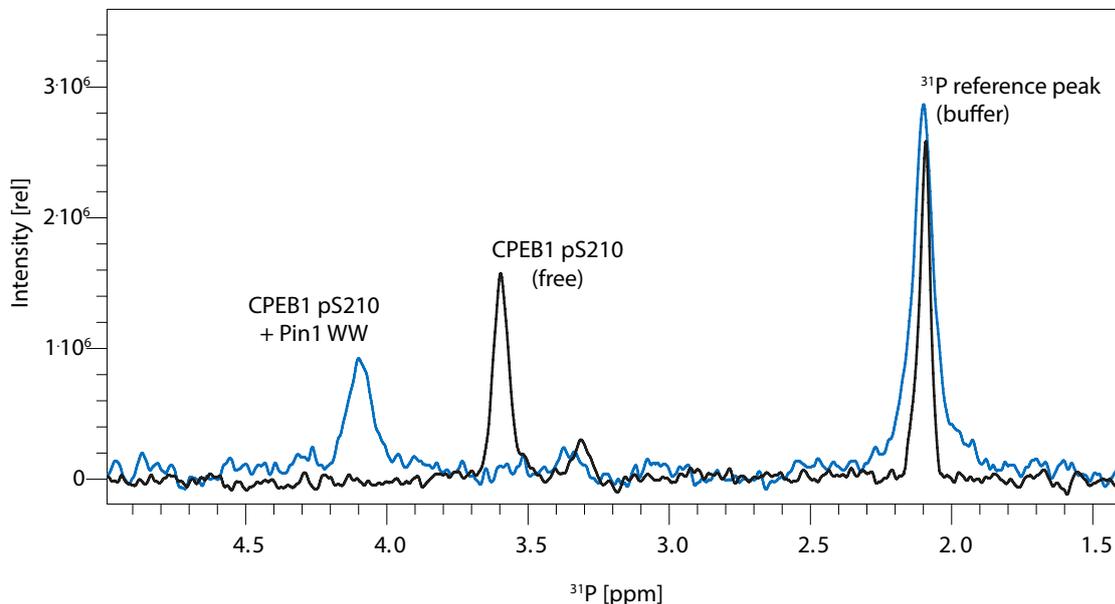


Figure 4.23: 1D- ^{31}P NMR titration experiment of CPEB1pS210 with Pin1 WW domain. The black spectrum shows the free peptide, in blue the saturated complex with Pin1 WW is displayed. The phosphate buffer reference peaks hardly shows any chemical shift perturbation.

Interestingly, CPEB1 pS210 and pS184, which according to the binding affinities obtained by ITC are the preferred Pin1 WW binding sites, also show the highest chemical shift perturbation of 0.5 ppm and 0.53 ppm, respectively. The difference in chemical shift perturbation comparing the titration of Pin1 WW and the full-length protein do not show a significant difference (0.5 ppm vs 0.45 ppm and 0.53 ppm vs. 0.5 ppm). The chemical shift perturbations of the rest of the phospho-peptides does not follow the ranking of affinities obtained by ITC. The ^{31}P NMR titration experiments confirm the importance of the phosphate group for Pin1 WW binding, as previously suggested by the ITC results.

CPEB1 peptide	$\Delta\delta$ [ppm]
CPEB1 pS138	0.25
CPEB1 pS144	0.19
CPEB1 pS184	0.53
CPEB1 pS210	0.50
CPEB1 pS248	0.19
CPEB1 pS423	0.11

Table 4.5: Chemical Shift Perturbations of the CPEB1 peptides' phosphate group upon titration with Pin1 WW domain.

4.2.5 Dynamics of Pin1 upon binding of CPEB1 pS210

CPEB1 pS210 is the only peptide of the six pSP-motifs within CPEB1, which induces significant chemical shift perturbations in the catalytic domain of Pin1. To investigate whether upon binding of CPEB1 pS210 any changes in the inter-domain mobility of the WW-domain and the catalytic domain are observable ^{15}N NMR Relaxation data of Pin1 full-length in its apo form as well as in complex with CPEB1 pS210 was acquired. ^1H , ^{15}N - NOE data show that for both the free state and the complex the catalytic domain and the WW domain are well structured. Within the WW domain in the free state an increased internal motion for residues of both loop 1 (Ser¹⁸, Gly²⁰, Ser²¹ and Ser²²) and loop 2 (His²⁷ and Ile²⁸) are observed. Upon binding of CPEB1 pS210, the flexibility of the residues within the first loop, which are forming part of the known substrate binding site, reduces notably. Within the catalytic domain the loop comprising residue Gln⁶⁶-Lys⁷⁷ is known to be involved in binding the phosphate moiety of the substrate. This area, in particular Ser⁶⁷-Arg⁶⁹, show an increased flexibility, which upon binding to CPEB1 pS210 slightly decreases. However, the dynamic nature of this catalytic loop region seems to remain upon substrate binding, which might be required for its function as for example observed in another *cis/trans* isomerase CypA [79]. The linker of Pin1 full-length shows high flexibility in its free state, in complex the obtained ^1H , ^{15}N - NOE values are notably higher, but the values are still indicating high flexibility compared to the restricted internal motions of residues within the two domains. This suggests that the linker region is neither involved in interactions with the domains nor with the peptide. Similar observations have been made for other Pin1 substrates [80], [50], [81], [82].

Futhermore longitudinal and transverse relaxation times T_1 and T_2 were determined. For both free and complexed form the values obtained are homogeneous along the single domains. The values for T_1 and T_2 in the apo state for the single domains obtained are: $T_1^{WW} = (612.7 \pm 53.6)\text{ms}$, $T_1^{Cat} = (792.7 \pm 81.8)\text{ms}$, $T_2^{WW} = (113.5 \pm 6.9)\text{ms}$ and $T_2^{Cat} = (85.4 \pm 13.1)\text{ms}$. In complex with CPEB1 pS210 the following values were obtained: $T_1^{WW} = (767.1 \pm 80.1)\text{ms}$, $T_1^{Cat} = (913.5 \pm 91.3)\text{ms}$, $T_2^{WW} = (94.1 \pm 5.8)\text{ms}$ and $T_2^{Cat} = (74.8 \pm 6.7)\text{ms}$.

The rotational correlation time τ_c was calculated for both domains in their free state and in complex assuming isotropic rotational diffusion. As described by Jacobs *et al.* [80] Pin1 is thought to consist of two rigid spheres of different sizes connected by a flexible linker. The model describes two extreme cases where (i) the two domains tumble independently of one another and (ii) the domains stick together and tumble as one intact moiety with a correlation time $\tau_c^{FL-rigid}$. Any deviation from these extreme cases is defined by an inter-domain interaction parameter x , which quantifies the dependence of τ_c of one domain on the other one: $\tau_c^{WW}(x) = \tau_c^{WW+L}(x) + x\tau_c^{Cat}$ and $\tau_c^{Cat}(x) = \tau_c^{Cat+L}(x) + x\tau_c^{WW}$, where $0 < x < 1$ and $\tau_c^{WW}, \tau_c^{Cat}$ proportional to the molecular weights of the respective domain. Based on this model using the same assumed $\tau_c^{FL-rigid}$ of 11.6ns, the inter-domain interaction parameter x has been

4 Results

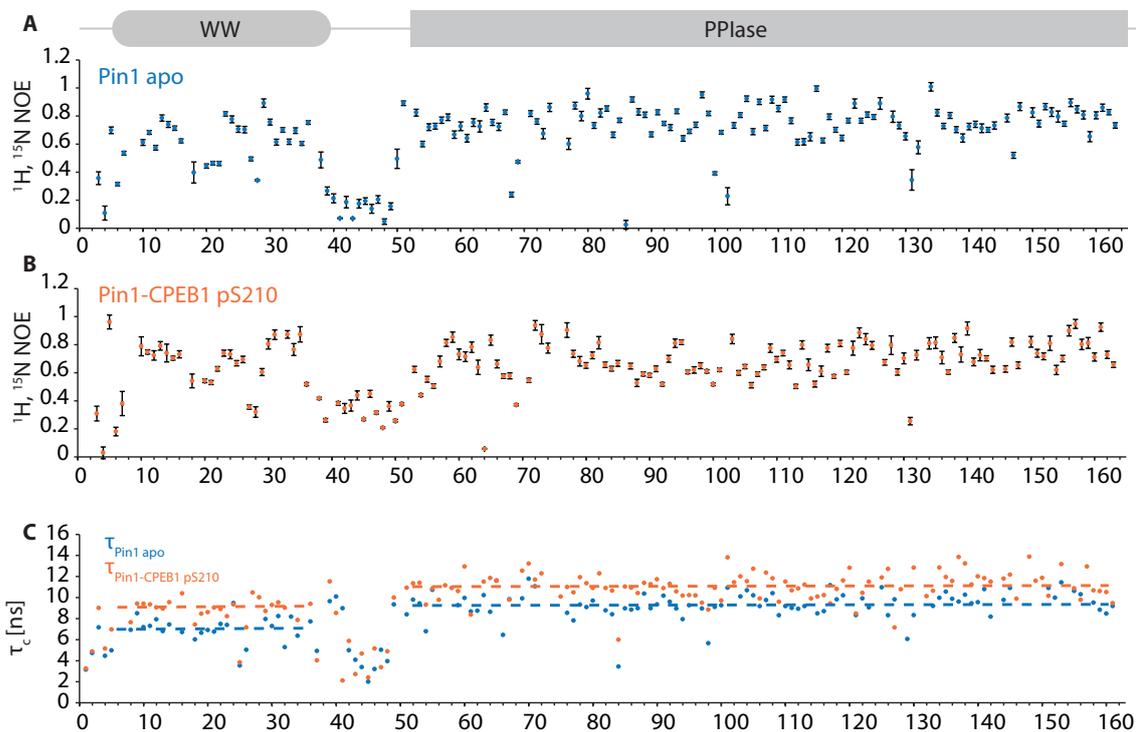


Figure 4.24: Amide ^{15}N relaxation properties of Pin1 apo form and in complex with CPEB1 pS210. Measurement of $^1\text{H}, ^{15}\text{N}$ NOE values are displayed for both the free (A) and the bound state (B). Local rotational correlation times τ_c (C) was calculated from T1/T2 ratio.

calculated. The correlation times of τ_c^{WW} (7.3 ns) and τ_c^{Cat} (9.4 ns) in the apo Pin1 full-length are larger than the expected values for two independently tumbling domains of their respective molecular weight. The inter-domain interaction parameter calculated is 0.46 (Jacobs *et. al* obtained a value of 0.47, measured at a slightly lower temperature, 298K vs. 300K). The notable difference of ~ 2 ns between the correlation times of the two domains indicates that the two domains tumble independently to a significant extent, although the linker restricts their flexibility indicated by the x value. Upon binding of CPEB1 pS210, the correlation time for the WW domain yields $\tau_c^{WW} = 9.2$ ns and $\tau_c^{Cat} = 10.9$ ns for the catalytic domain. The higher value of the inter-domain interaction parameter ($x_{complex} = 0.7$) indicates an increased restriction of the flexibility of the two domains upon binding of the ligand. Still the complex does not tumble as a single unit and a residual flexibility remains. Similar observations have been made using Pintide, a peptide designed to be an optimal substrate for Pin1 [80].

4.2.6 Structure of Pin1 WW CPEB1 pS210 complex

To gain insight into the binding mode of Pin1's WW domain and the CPEB1 pS210 substrate, sets of homonuclear TOCSY and NOESY experiments in both H₂O and D₂O were acquired of the WW domain bound to CPEB1 pS210.

The assignment of the apo Pin1 WW domain, previously obtained in the lab, was used as starting point to assign the TOCSY and NOESY spectra of the complex in order to obtain NOE restrictions for the structure determination of the WW domain in its bound state.

Applying the sequential resonance assignment strategy, the CPEB pS210 peptide was assigned in the free state using 2D homonuclear TOCSY and NOESY experiments. The bound state of the peptide was identified through sequential assignment of intramolecular NOEs and intermolecular NOEs with the protein's WW domain.

CPEB1 pS210 with Pin1 WW				CPEB1 pS210 with Pin1 WW							
AA	Proton	Intermolecular NOEs		AA	Proton	Intermolecular NOEs					
I209	HB	Y23	QE	P211	HD2	W34	HD1				
		F25	QE				HZ2				
			QD				Y23	QD			
	QD1	Y23	QE			HD3	W34	HZ2			
		F25	QE				Y23	QD			
			QD				HB3	QE			
	QG2	Y23	QE			HG2	W34	HD1			
		F25	QE					HZ2			
			QD				Y23	QD			
	S210	HA	W34		HZ2		HA		QE		
			HB2		R17	HD2/3		P212	HA	W34	HD1
						HB2/3					W34
Y23		QE		L213	HB2	W34	HE3				
W34		HZ2					HD1				
		HH2					HB2/3				
HB3		R17	HB2/3		QD1	W34	HD1				
			HG2				HH2				
		Y23	QE			QD2	W34		HD1		
W34		HZ2			HH2						
		HH2			HB2						

Table 4.6: Intermolecular NOEs identified by 2D NOESY assignment of the Pin1 WW CPEB1 pS210 complex.

Structure calculations were performed following the protocol described in section 3.5.7 using inter and intra molecular NOE structural restrictions as well as dihedral angle restraints (J and χ couplings). The quality of the calculated structure ensemble was analyzed with iCING web server and subjected to several rounds of refinement.

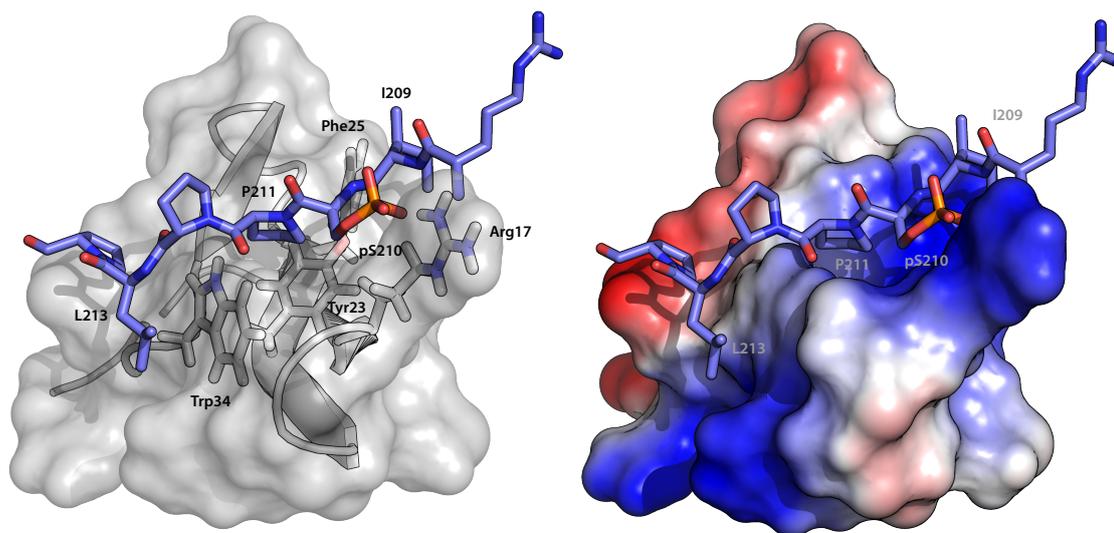


Figure 4.25: *left:* Lowest-energy structure of the Pin1 WW - CPEB1 pS210 complex in cartoon/surface representation. The domain is colored in grey and residues are labeled in the 3-letter code, the peptide is colored in blue with its residues labeled in the 1 letter code. *right:* Surface charge distribution of the Pin1 WW domain displayed in the same orientation as in the left panel.

The Pin1 WW domain adopts the typical three stranded anti-parallel β -sheet fold analogous to previously deposited structures [56]. The protein domain binds the CPEB1 pS210 peptide from the N- to the C-terminus. This orientation has been observed previously for several Pin1 WW - substrate complexes [51], [83], [36]. Pro²¹¹ is bound in *trans* configuration between the aromatic residues Tyr²³ and Trp³⁴ lying on the β_1 - and β_3 -strand, respectively. Further downstream intramolecular NOEs of Pro²¹² and Leu²¹³ with Trp³⁴ have been identified. The phosphate group of CPEB1 pS210 is coordinated by the side chain of Arg¹⁷ located in the β_1 -strand. Additional contacts with the upstream part of the peptide are observed between Ile²⁰⁹ and the aromatic rings of residues Tyr²³ and Phe²⁵.

Comparison of the Pin1 WW - CPEB1 pS210 complex to the binding mode observed in Pin1 WW complexes with other p(S/T)P-motifs shows that also in this complex

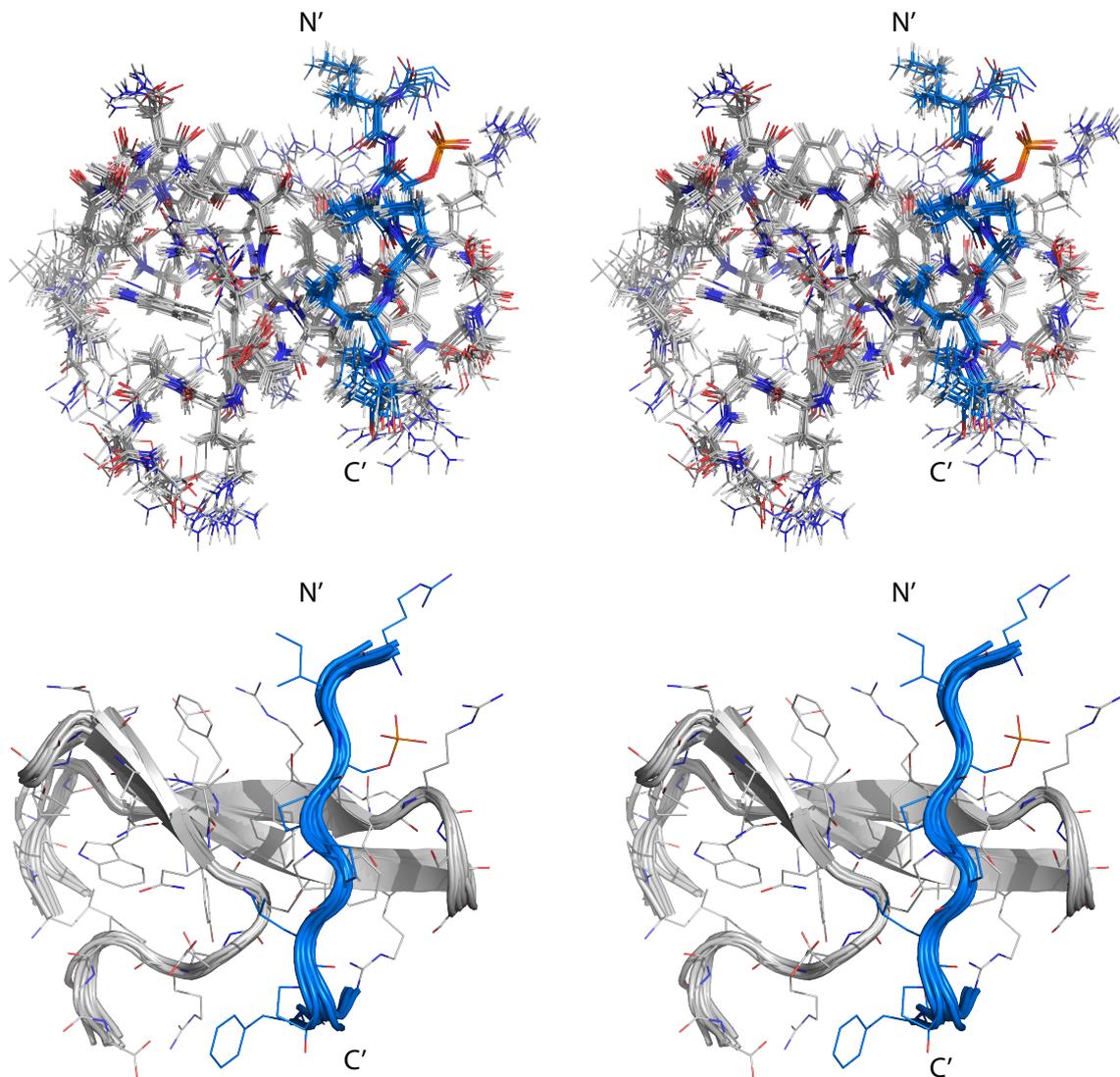


Figure 4.26: Stereo view of the best-fit backbone superposition (RMSD: 0.42) of the ten lowest energy structures after water refinement in sticks representation (*upper panel*) and cartoon representation (*lower panel*). Pin1 WW is displayed in grey, the ligand CPEB1 pS210 is colored in blue.

Arg¹⁷ plays an important role in phosphate binding [51], [52], [56]. It has been shown previously that an Arg¹⁷Ala mutation results in a six-fold affinity decrease [51]. The aromatic residues Tyr²³ and Phe²⁵, located in the second β -strand, contribute to the binding as observed also in the Pin1 WW Smad3 pT179 and pS208-pS213 complexes

[56]. However, in the aforementioned complexes the location of phosphate group is shifted towards the first β -strand, since Arg¹⁴ is either also involved in the coordinations of the phosphate group (Smad3 pS208-pS213) or in the interaction with another residue downstream the phosphorylated amino acid (Smad3 pT179).

Pin1 WW - CPEB1 pS210	
Restraints used for calculation	
Inter-domain	48
Sequential ($ i-j = 1$)	166
Medium-range ($1 < \text{dist} \leq 4$)	49
Long-range ($\text{dist} > 4$)	195
Dihedral	71
Hydrogen bonds	10
All restraints	539
Structure Statistics	
RMSD	
Bonds [\AA]	0.009 \pm 0.000
Angles [$^\circ$]	1.149 \pm 0.034
Average pairwise RMSD [\AA]¹	
Backbone	0.93 \pm 0.21
Heavy	1.38 \pm 0.20
CNS potential energy[kcal mol⁻¹]	
Total energy	-740.8 \pm 28.5
Electrostatic	-1428 \pm 46.9
Van der Waals	2.4 \pm 17.8
Bonds	68.18 \pm 3.9
Angles	258.4 \pm 13.56
Ramachandran	
Core	85.3 %
allowed	12.9 %
generous	1.7%
disallowed	0.0%

Table 4.7: NMR and refinement statistics for the Pin1 WW - CPEB1 pS210 ensembles. Statistics are calculated for the twenty lowest-energy structures of a total of 200 structures.

¹ Backbone atoms within Pin1 WW (11-43) - CPEB1 pS210(208-215).

In complex with CPEB1 pS210 however, no contacts between Arg¹⁴ and the peptide have been detected. Here the phosphate moiety is accommodated in the region of the first turn. This arrangement is similar to the structure of the Nedd4L WW3 in complex with the di-phosphorylated Smad3 pS204-pS208 peptide. In this complex the phosphate group of pS208 is coordinated by residues in and around the first loop of the domain. In the Pin1 WW - CPEB1 pS210 complex Arg¹⁴ is pointing away from the phosphate moiety and adopts a similar orientation as observed in the complex with Smad3 pS208-pS213.

5 Discussion

5.1 RNA recognition and self-association of CPEB4 is mediated by its tandem RRM domains

The first part of this work was devoted to characterize the RRM domains of CPEB4 and their binding properties using a combination of biochemical, biophysical and NMR techniques. Especially to address the question of the role of each single RRM domain upon RNA binding was of great interest.

Binding properties of CPEB4's RRM domains

Here, we provide evidence that CPEB4 uses both its RRM domains to maintain optimal RNA-binding. RRM1, whose sequence is the more conserved of the two tandem domains, relative to other RRM motifs, is only able to bind CPE-containing RNAs U5A2U1 and U5A1U2 in the μM range. Interaction could not be detected with neither the non-consensus motif C1U4A1 nor the poly U stretch. ITC experiments performed with the RRM1-RRM2 pair yielded affinity values around 300 nM for the RNA fragments containing the consensus CPE-motif, 10-100 times higher than the isolated RRM1 domain. Furthermore the poly U stretch displayed a high affinity in the same range as the two CPE-containing fragments. High affinity binding to poly U stretches have been observed for other RRM tandem domains such as the RRM domains of U2AF [84]. The tandem RRM1-RRM2 also interacts with the non-consensus CPE C1U4A1, although with a ten-fold decreased affinity compared to the consensus CPE-motifs. Interestingly, using a RNA fragment containing only a three nucleotide uracil stretch *in vitro* abrogates the interaction. We therefore assume, that a minimum of 4 consecutive uracils is required for recognition of the motif.

An overview of a repertoire of nucleic acid binding modes by proteins containing two RRM domains and the principles of multi-domain protein-RNA recognition has been recently reviewed by Mackereth and Sattler (2012) [32] and reported by Barraud and Allain (2013) [33]. Based on the structures available, they classified different domain arrangements in the free as well as in the bound state. In the free state the domains can either be tumbling independently, with pre-formed domain contacts or adopt a closed, autoinhibited form. Upon ligand binding RRMs can, for example, independently bind separate RNA motifs or form a continuous RNA-binding platform, stabilized by inter-

domain contacts or increased linker rigidity. Another mechanism to stabilize the RNA binding is mediated by protein-protein interactions with an auxiliary protein.

In the case of CPEB4 RRM domains, we observed that both domains contribute to the interaction with the consensus eight nucleotides CPE motif, presumably increasing not only the affinity but also the specificity of the interaction. Given the length of the RNA motif, we assume that the RRM domains form a continuous binding platform. Based on the ^{15}N amide relaxation properties the CPEB4 RRM1-RRM2 pair behaves as a compact molecule already in the unbound state. Dynamics of the tandem RRM domains in complex with a consensus CPE motif, does not show any significant changes comparing to its properties in the apo state. The results suggest that there are pre-formed domain contacts between RRM1 and RRM2 which establishes a certain orientation between the two domains. This arrangement might be tuned in the presence of RNA, but also resulting in an arrangement where both domains are tumbling as a single unit.

Our NMR titration data for CPEB4's tandem RRM shows that in RRM1 the most affected residues upon binding lie on its β -sheet, the canonical binding surface of RRMs, containing the conserved RNPs on the β_1 and β_3 strands. RRM2 is the less canonical domain of the pair containing 3 of 6 conserved residues of RNP2 motif (I-F-V) and completely lacking the RNP1 motif on the β_3 strand. Moreover, it seems to interact with the CPE in a non-canonical manner, since we detect chemical shift changes mainly at positively charged residues located in its C-terminus. As reported previously, the β -sheet surface can bind up to four nucleotides [85]. Therefore, the recognition of a longer RNA requires more than one RRM to create a sufficiently large binding platform. The two consensus CPE sequences contain each eight nucleotides (UUU UUA AU and UUU UUA UU), thus it seems plausible, that a motif of this length binds more than one RRM. Moreover, the use of regions other than the β -sheet binding platform has been demonstrated in other tandem RRM structures. For example, the RRMs in the protein Nucleolin use the linker and a loop between the β -sheets to interact with a stem loop RNA structure [86]. The seemingly different modes of RNA interaction of CPEB's RRMs, with RRM2 using C-terminal residues, suggest an asymmetrical orientation of the domains relative to each other.

A model for the spatial arrangement of the RRM domains

û In an attempt to visualize all the results obtained from this study, we built a homology model containing the two RRM domains and an octameric polyU stretch to illustrate the relative size of a single stranded octameric RNA compared to the RRM domains. For RRM1 the homology model mentioned above was used, for the RRM2 a homology model was built with SWISS-MODEL (Template: PDB entry 2FC8, RRM1 of NCL protein, sequence identity 28.95%). With a QMEAN score of -3.15 the RRM2 model cannot be considered well modeled, and must be treated with caution. Since

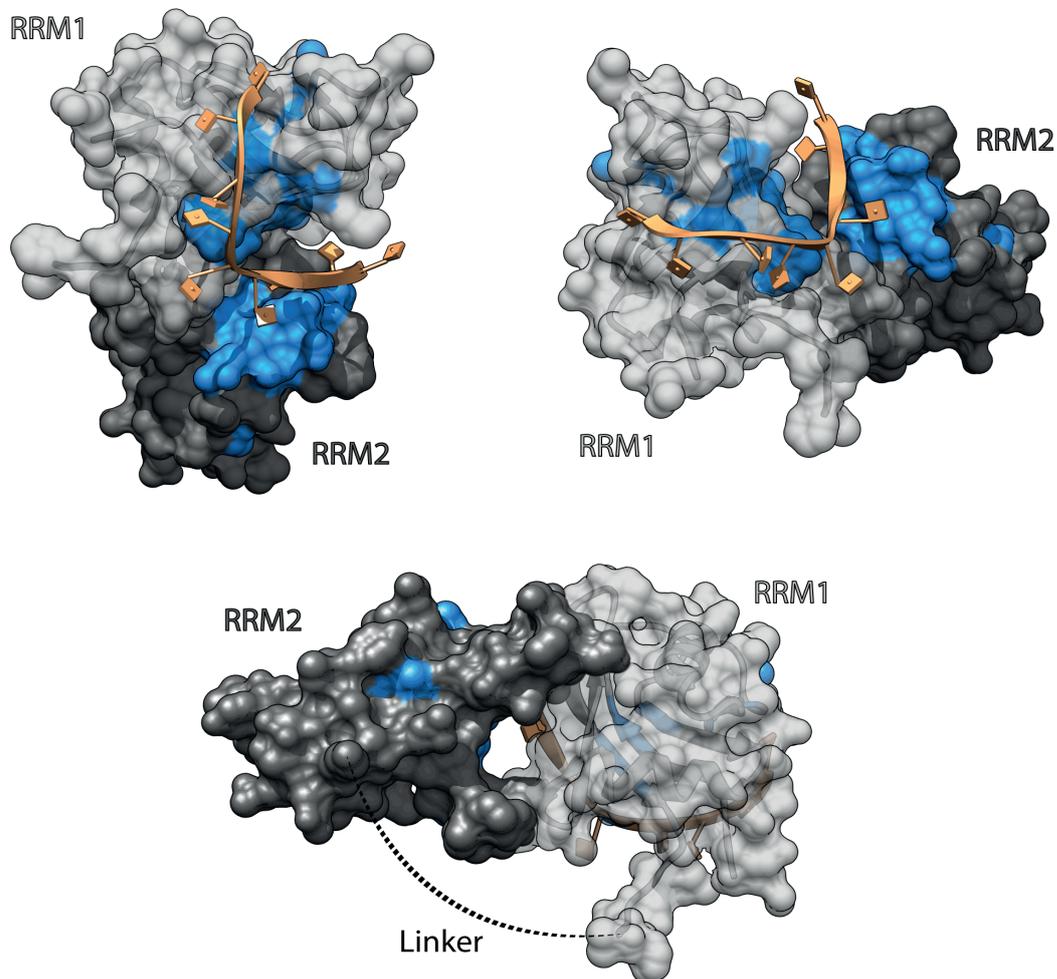


Figure 5.1: A model for the spatial arrangement of the RRM domains in complex with RNA. The arrangement of the two RRM domains as well as the RNA fragments was performed to be consistent with the results obtained in this work, especially from the NH CSPs

our interest only focused on the spatial arrangement of the two domains and not in any structural details of the RRM2, we consider the obtained model sufficiently good for that purpose. The linker region between RRM1 and RRM2 containing \sim ten amino acids has not been modeled. However the restrictions implied by the length of the linker were taken into account. The arrangement of the two RRM domains as well as the RNA fragments was performed to be consistent with the results obtained in this study, especially from the NH CSPs. As the model quality of RRM1 is much higher compared to RRM2, the focus regarding the restraints relies on the RRM1 results.

For each RRM domain a separate model excluding the linker region was built. The orientation of the domains was manually selected to satisfy the chemical shift differences observed for the RRM1 in the presence or absence of the RRM2 (isolated or in the pair). The orientation was selected assuming that the CS differences correlate with the proximity of both domains to one another and to satisfy as well all changes detected upon RNA binding (highlighted in blue in both RRM1 and RRM2). In order to consider all affected regions and to cover the maximum length of the RNA fragment, the RRM2 needs to be oriented with its C-terminus towards the RRM1 domain. We believe that this preliminary model based on the CS changes obtained for the amide resonances of the RRM1-RRM2 sample in the presence of RNA provides an idea about the relative orientation of the two domains.

Recently, both the apo state NMR structure of RRM1-RRM2 (PDB entry 2MKJ) as well as in complex with a C1U3A1 RNA fragment (PDB entry 2MKI) have been published [87]. The construct used in the study by Afroz *et. al* is similar to the one we have been working with, only featuring a slightly extended N-terminal part. According to the authors the extension corresponds to an unstructured area. Comparing the homology model we have obtained for RRM1 to the region of the lowest energy structure corresponding to RRM1 from PDB entry 2MKJ, a RMSD value of 1.8 Å is obtained. This value indicates, that the quality of the model prepared is very good. For the model for RRM2, despite its low sequence identity, the RMSD calculated yields a value of 3.1 Å, which indicates that our model can be considered sufficiently good for the purpose intended. Our assumption, based on the results obtained, that the β -sheet surface of RRM1 is oriented towards RRM2 is confirmed by the NMR structure of the apo CPEB4 RRM domains.

The structure of the RRM1-RRM2 in complex with RNA was solved using the fragment C1U3A1. For this complex the two RRM domain adopts a closed conformation with the β -sheet of RRM2 perpendicular to the one of RRM1. The cytosine as well as the three consecutive uracils are bound by residues of RRM1, only the adenine forms contacts with residues from RRM2. When comparing the chemical shift perturbations observed by Afroz *et al.* using the five-nucleotide C1U3A1 with the results from our NMR Titration experiments using either U5A1U2 or U5A2U1 as a ligand, the chemical shift of residues corresponding to the RRM1 domain are very similar. For the RRM2 domain, however, our results show several pronounced chemical shift perturbations in a C-terminal region containing residues with positively charged side groups. This region remains unaffected by the shorter C1U3A1 ligand. In general, the CSPs for RRM2 observed by Afroz *et al.* are less pronounced compared to the ones we observe with the octameric RNA fragments. Comparing the affinities reported by Afroz *et al.* to the results of our ITC study, the pentameric ligand yields an affinity value of 15 μ M, around 50-fold decrease compared to the values obtained for U5A1U2 or U5A2U1. We believe that the differences observed in the affinity range and in the occurrence of the chemical shift perturbations account most likely for the difference in length of the RNA fragment used. Unfortunately we have not been able to detect

interaction by ITC titration experiments for the C1U3A1 ligand under the conditions we used for the characterization of our complexes precluding a direct comparison of all affinity values reported.

When mapping the chemical shifts observed in our experiments on the complex structure (PDB entry 2MKI), the CSPs observed for RRM1 residues match well with the position of the RNA. However, the chemical shifts we observed for RRM2 are not located in spatial proximity to the ligand in the complex. Perhaps these differences can be attributed to the difference in length of the ligand and/or the relative orientation of RRM2 towards RRM1 adopted in the published complex. We therefore assume that upon binding a longer RNA sequence, the relative orientation of the two RRM domains towards each other might be slightly altered, resulting also in a higher affinity binding.

Both reported structures have been used as search models for molecular replacement trials with the native crystal data set. Unfortunately, phasing could not be carried out successfully. A different relative orientation of the two RRM domains, induced by either conformational changes upon binding of a longer RNA stretch and/or due to crystal packing, might be the reason why molecular replacement could not be performed successfully despite the 100 % sequence identity of the search model. Furthermore the resolution of 3.1 Å complete dataset should at least allow to trace the backbone of the protein. Additionally we need also to take into account that the large size of the unit cell imposes further difficulties, since the number of copies to search is quite high (8 copies per asymmetric unit). Nevertheless, the fact that both 2MKI and 2MKJ are not suitable as a search model for molecular replacement might indicate that the described conformation adopted by the two RRM domains in solution is different to the one its crystallized form.

Self-association of CPEB4's RRM domains

Our results demonstrate that CPEB4 RRM1 and RRM1-RRM2 can sample monomer and dimer conformations under the experimental conditions we have assayed. However, the inability to produce a soluble RRM2 fragment did not allow us to identify whether both domains are involved in the self-association. The dimeric species are present in low abundance, precluding its characterization using NMR spectroscopy. However, the potential to form dimers has been observed for other RRM domains. For example, the U1A protein forms a homodimer via its RRM domain. The dimer interface is separate from the polyadenylation inhibition element (PIE) RNA interaction interface [88]. Other RNA-binding domains like yeast She2p are only active in their oligomeric form [89], [90]. In AUF1 (AU-rich element RNA binding domain), for example, dimerization occurs via a N-terminal alanine rich region. Deletion of this region results in a 10-fold decrease in binding affinity. Therefore, in AUF1, dimerization is essential for optimal RNA binding [91]. In the cases of CPEB4's RRM domains, IM-MS assays

have unambiguously detected both monomeric as well as dimeric 1:1 protein-RNA complexes, indicating that both the monomeric and dimeric states of RRM1-RRM2 possess RNA-binding activity. This suggests that in the case of CPEB4 RRM domains dimerization is not required for RNA recognition.

Previously dimerization has been reported for another member of the CPEB-family CPEB1 [92]. Lin *et al.*, (2012) postulated that the dimerization functions to inactivate spare proteins, preventing them from inducing the polyadenylation of RNAs with low affinity binding sites and also serving as molecular hubs that release polyadenylation factors upon dimer destruction. The same study identified both the RRMs and the ZZ-domain to be essential for dimerization. Interestingly, and in contrast to our experiments, the results suggest that CPEB1 dimerization occurs at the expense of RNA binding. However, whereas in our study, isolated RRM domains were characterized, Lin *et al.*, (2012) assessed full length CPEB1 and two dimeric constructs: two full-length CPEB1 proteins separated by a linker sequence of 40 amino acids and full-length CPEB1 with an N-terminal coiled-coil dimerization domain.

The sequence identity of CPEB1 and CPEB4 RRM domains is 45% and the overall identity is 22%. Although it is possible that the sequence disparity between the two proteins RRM domains may account for differences in the mode of the RNA interaction and self-association, we suggest, it is more likely, that the reason for the loss of binding activity in the full-length CPEB1 dimer is due the spatial arrangement of the N-terminal unstructured regions and the C-terminal ZZ-domains, which renders the RNA binding interface inaccessible. Our IM-MS data clearly shows that the dimeric form of the RRM pair maintains their RNA-binding activity. We therefore conclude that the dimerization surface is separate from the identified RNA-binding interface.

Experiments with the full-length CPEB4, as investigated with CPEB1, would ascertain if the inhibition of RNA binding upon dimerization is exclusive of CPEB1 or is present in other family members. Further investigation into the function of the potential formation of CPEB oligomers is required since it may reveal the critical differences between the four functional CPEB paralogs of vertebrates. It would be interesting to see if the postulated molecular hubs are detectable using the available cell biology techniques or could indeed be purified for in-vitro analysis.

5.2 Pin1 mediated CPEB destruction - a two-step interaction

CPEB1 destruction has been shown to be vital for correct mitotic cell-cycle progression [16]. During cell-cycle progression CPEB1 exerts a dual function. Before stimulation with progesterone CPEB1 recruits several cofactors to maintain the mRNA in a translationally repressed state. Upon stimulation by progesterone CPEB1 becomes phosphorylated on Ser¹⁷⁴, resulting in the dissociation of the repression complex and the recruitment of the polyadenylation complex [25], [93]. At this early stage a class of mRNAs, such as those encoding Mos [15], are activated. However, the late translational events, such as the translation of cyclin B1, require on the one hand the previous synthesis of Mos and the activation of Cdc2 kinase resulting in hyperphosphorylation of CPEB1 and the subsequent partial destruction of CPEB1. The mechanism of degradation of CPEB1 during oocyte maturation is observed in many species [94], [95],[96] apart from *Xenopus* oocytes, the classic animal model for CPEB studies. Recently, the peptidyl-prolyl *cis/trans* isomerase Pin1 has been identified to be a crucial factor regulating CPEB1 degradation in *Xenopus* as well as in mammalian cells [53].

Step one: phosphorylation independent interaction of CPEB1 and Pin1

Our results demonstrate that prior to CPEB1 phosphorylation Pin1 is already interacting with CPEB via its WW domain. We have been able to ascertain two N-terminal regions of CPEB1 involved in this interaction, namely NF1s (88-183) and NF2 (196-293). Previously, it has been shown *in vivo*, that residues 48-183 are involved in Pin1 interaction. Furthermore a depletion of residues 211-290 resulted in the abrogation of interaction. It has been demonstrated *in vivo* that prior to oocyte maturation Pin1 is inactive due to its phosphorylation on Ser⁷¹. Interestingly Ser¹⁶, located on the first loop of the WW domain, is not phosphorylated as shown by Nechama *et al.* [53].

Our NMR based interaction studies allowed us to narrow down the region of interest to residues 88-183. Moreover, we could directly confirm the interaction of Pin1 with the second site NF2, comprising a slightly extended region of residues compared to the depletion region. Interestingly, the residues of Pin1 WW domain affected upon binding to NF1s and NF2, respectively, are similar to the ones affected upon binding to a phosphorylated pSP-motif. The binding affinity however lies only in the high μM range, as determined by ITC. Although Pin1 is described to specifically recognize and isomerize p(S/T)P-motifs, observations have been made that apart from the preferred phosphorylation dependent interaction, a phosphorylation independent interaction mode occurs, for instance in Cdc25 [97] and PKC isozyme [98]. For the latter example, a hydrophobic motif has been identified to interact with Pin1.

NF1s and NF2 contain 25.8% and 28.9% hydrophobic residues, respectively. Since no conserved motif is detectable and secondary structure prediction based on the sequence indicates only a few, short α helices as secondary structure elements, it is very likely that most of those hydrophobic amino acids are solvent exposed and therefore accessible as a binding platform for Pin1. Since the phosphorylation independent CPEB1-Pin1 interaction does not seem to regulate substrate degradation, its purpose in cell-cycle progression is yet to be determined. A possible function of the low affinity CPEB1-Pin1 association might be to ensure spatial proximity of Pin1 when CPEB1 becomes hyperphosphorylated and needs to be degraded rapidly for proper cell-cycle progression. Consistent with this notion, the N-terminal regions NF1s and NF2, which have been shown to be involved in Pin1 binding, comprise the six described Cdc2 phosphorylation sites, the proposed binding sites for Pin1.

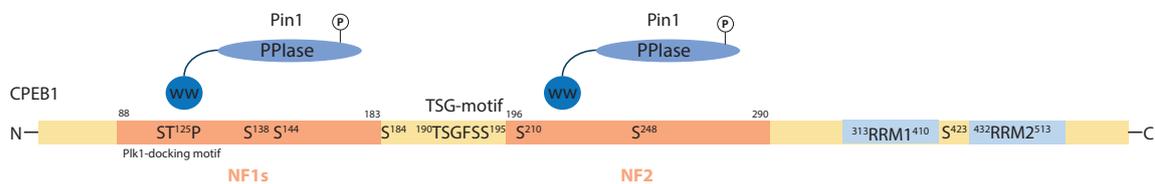


Figure 5.2: A model for the phosphorylation independent interaction of Pin1 with CPEB1 NF1s and NF2, respectively. This interaction is mediated via Pin1's WW domain.

Step two: phosphorylation dependent interaction of CPEB1 and Pin1

Upon progesterone stimulation CPEB becomes phosphorylated by Cdc2 on six pSP motifs. Simultaneously dephosphorylation of Pin1 and hence its activation occurs. NMR titration experiments and ITC assays show, that Pin1 WW recognizes all of the potential CPEB1 substrates *in vitro* with the following order of affinity $pS210 \approx pS184 > pS248 > pS138 \approx pS144 > pS423$. Except for pS423, which is located in the linker region between RRM1 and RRM2, all potential substrates are located in the N-terminal unstructured region of CPEB1. Interestingly, it seems to be a common feature of Pin1 substrates to be located in regions of a protein which are predicted to be unstructured [40].

The available biochemical information on Pin1 indicates that at least *in vitro* it is not very selective upon recognition p(S/T)P motifs. The WW domain recognizes a broad range of sequences whose common feature is the p(S/T)P motif. At low affinity binding ($\sim 100 \mu\text{M}$), WW's binding pocket mainly sequesters the phosphate moiety and the proline side-chains, whereas other amino acids around this motif are only marginally contributing to the binding [99]. Since pS423 *in vitro* is a poor Pin1 binder and located almost 200 amino acids upstream the ubiquitin ligase binding site,

we assume that Pin1 isomerization of this peptide bond is unlikely to play a role in Pin1 mediated CPEB1 degradation. It is noteworthy that the preferred Pin1 binding sites pS210 and pS184 are located very close to the TSG-motif (190-195), which when phosphorylated has been identified to be the SCF $^{\beta-TrCP}$ E3 ubiquitin ligase binding site[17]. Furthermore it has been demonstrated that apart from the phosphorylation in the TSG-motif, the Cdc2 phosphorylations are important for efficient SCF $^{\beta-TrCP}$ -binding. Consistent with this observation, pS210 has been identified to be essential for CPEB degradation [16]. The function of these multiphosphorylation sites is yet to be determined, especially for the site pS138 and pS144, which are located ~ 50 amino acids downstream the TSG-motif. However, a possible function might be to expose the nearby T¹²⁵ Plk1-docking motif or the TSG-motif (especially pS184) and therefore facilitating the SCF $^{\beta-TrCP}$ -binding. Similar suggestions have been made in the case of Cdc25A phosphatase, which is also targeted by SCF $^{\beta-TrCP}$ E3 ubiquitin ligase for degradation [100].

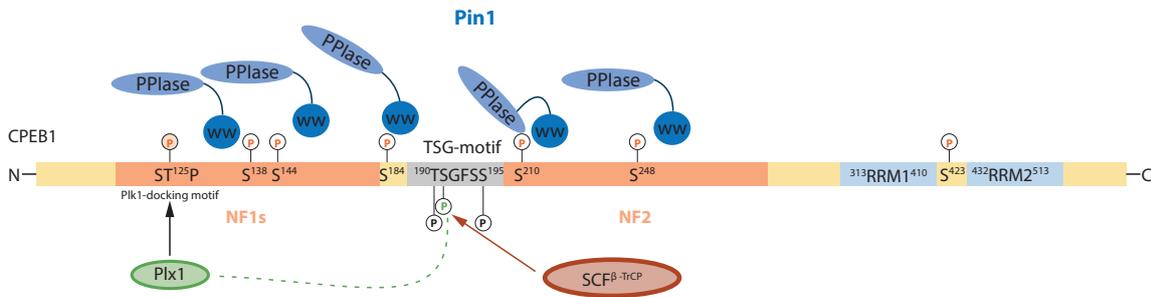


Figure 5.3: A model for the Pin1-phospho CPEB1 interaction including the molecular mechanisms of CPEB degradation.

Our results show that regarding the interaction with Pin1, CPEB1 pS210 features a unique characteristic compared to the the rest of the potential substrates from CPEB1. It is the only substrate tested, which apart from the binding to the WW domain, also interacts with the catalytic site of the PPIase domain of Pin1. Sequence comparison between CPEB1 homologs of different species reveals a preeminent conservation of the residues corresponding to pS210 and pS184 and adjacent ones (see Figure 5.4). Interestingly, the most conserved residues are also the preferred binding sites for Pin1.

Furthermore our ^{15}N NMR relaxation data demonstrate an increased inter domain interaction when in complex with pS210. The functional implication of this modulation regarding the inter domain flexibility are yet to be investigated. However studies by Namanja *et al.* [101] suggest that the interdomain interaction provides a intraprotein signaling mechanism, by which the WW domain might tune the binding affinity of the catalytic binding site. It has been suggested that conduit stiffening upon substrate binding by the WW domain may serve to regulate the conformations sampled by the catalytic site allowing a fine tuning of this remote functional site.

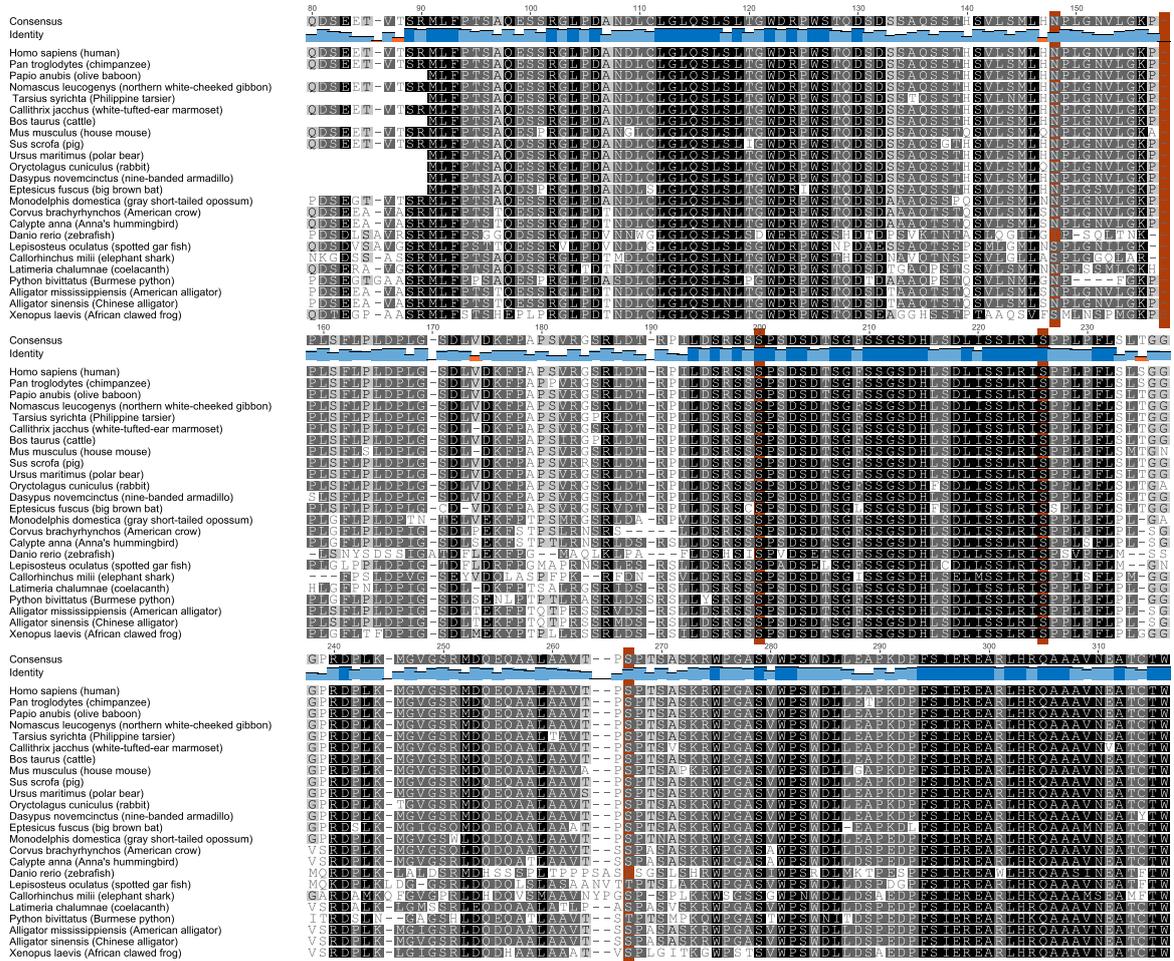


Figure 5.4: Sequence alignment of the N-terminal part of CPEB1. Cdc2 phosphorylation sites within the N-terminal part are indicated by a red box

The precise mechanism for Pin1 isomerization hasn't been elucidated yet, nevertheless the following two mechanistic models are proposed: A sequential model as proposed by Zhou *et al.* [102] and Wintjens *et al.* [52] describes the processes of target recognition via the WW domain and isomerization via the catalytic domain as two subsequently and independently occurring events at the same phosphorylation site. For substrates with repetitive target motifs an initial binding event may trigger a cascade of anchoring by the WW domain and subsequent isomerization by the catalytic domain. Consistent with this model, it has been shown that at high cellular concentrations the catalytic domain of Pin1 is sufficient to perform the essential Pin1 function. A second mechanism proposed by Lu *et al.* [103] describes a model where the WW domain anchors Pin1 to an already phosphorylated p(S/T)P motif on another protein, for example an activated kinase, phosphatase or another protein. This scenario would enable an efficient “tag and twist“ mechanism, in which the kinase would “tag“ the

substrate via phosphorylation and Pin1 would subsequently isomerize or “twist“ the p(S/T)P bond. Since alongside our *in vitro* studies also *in vivo* studies have revealed that the CPEB1-Pin1 interaction is mediated via Pin1’s WW domain, the sequential model of binding, with the WW domain directly anchoring on the substrate, seems to be more likely for the the case of CPEB1-Pin1 interaction.

The structure of the Pin1 WW - CPEB1 pS210 complex reveals that the WW domain binds the pSP-motif of its substrate in the *trans* configuration. The impact of the phosphate group on binding has been demonstrated via ^{31}P -NMR. The target recognition by the substrate binding domain involves mainly residues Arg¹⁷ and Trp³⁴, which have been also identified to play a key role in the interaction with other known substrates such as Smad3 [56] or RNA polymerase II subunit’s C-terminal domain [51]. However, in the complex with CPEB1 pS210, Arg¹⁴, which in the other aforementioned complexes participates either in the coordination of the phosphate moiety or contacts the ligand upstream the p(S/T)P motif, does not participate in the interaction.

The role of Pin1-induced conformational changes in CPEB during the process of ubiquitination and subsequent degradation is not described. However several lines of evidence indicate that Pin1 induces a conformational change in CPEB1, which favors the interaction of CPEB1 with SCF ^{β -TrCP}: for instance, SCF ^{β -TrCP}-binding is less efficient ($\sim 20\%$ of WT-binding) when the six Cdc2 phosphorylation sites are mutated to alanine and most kinases require *trans* configuration. Since the interactions sites for both Pin1 and the E3 ubiquitin ligase are located in unstructured regions, it is reasonable to assume that the predominant configuration at the pSP bonds will be *trans*. Nevertheless, it should be noted that the ratio is roughly 3:1 for X-Pro bonds compared to a 1000:1 ratio in most peptide bonds in unstructured regions. Moreover, phosphorylation of the adjacent serine or threonine further slows down the isomerization rate of prolines. Pin1’s catalytic activity only lowers the energetic barrier separating the two configurations, thus increasing the conversion ratio, but without changing the relative populations. Assuming that for the degradation of CPEB only the *trans* configuration can be targeted by the SCF ^{β -TrCP} ligase, the ratio between the *cis* and *trans* configuration will continuously change due to the degradation of the *trans* population. In this scenario, isomerization could accelerate changes in the *cis/trans* ratio and shift the equilibrium towards the *trans* configuration faster. In the case of CPEB1, where rapid degradation is crucial for correct cell-cycle progression, Pin1 might catalyze the degradation by ensuring a predominant configuration which is accessible for SCF ^{β -TrCP} E3 ubiquitin ligase mediated degradation.

6 Summary

Cytoplasmic polyadenylation is the process by which dormant, translationally inactive mRNAs become activated via the elongation of their poly(A) tails in the cytoplasm. This process is regulated by the conserved cytoplasmic polyadenylation element binding (CPEB) protein family.

The molecular mechanism of this process, which has been identified to play an important role in oocyte maturation, mitotic cell cycle progression, cellular senescence and synaptic plasticity, have been studied extensively and several models have been proposed. The majority of the characterization of cytoplasmic polyadenylation has been elucidated in *Xenopus* oocytes undergoing meiotic maturation stimulated by progesterone.

The CPEB family contains four members whose functional differences have not been elucidated in detail yet but their activities might also influence the pattern of translational regulation. The conservation between their RNA binding domains is high while the N-terminal unstructured regions are more divergent thus indicating that this region has a regulatory function. The relationship between CPEB1 and CPEB4 might be the best characterized so far. CPEB1 drives polyadenylation during oocyte maturation, after MI is heavily degraded to ensure correct cell cycle progression. It has been described that CPEB4 replaces CPEB1 in interkinesis through a translational loop involving polyadenylation and activation of CPEB4 mRNA by CPEB1 during MI. CPEB1 therefore establishes a positive feed-back loop essential for meiotic progression. Recently, the peptidyl-prolyl cis/trans isomerase Pin1 has been identified to be a crucial factor regulating CPEB1 degradation in *Xenopus* as well as in mammalian cells.

In the first part of this work we focused on studying the tandem RRM binding domains of CPEB4 with a special focus on their RNA Recognition properties. The aim was to characterize the role of each domain in RNA recognition and the differences in RNA binding activity for different CPE-motif containing RNAs. First, the characterization of the free RRM1 and the RRM1-RRM2 tandem construct via NMR revealed an interaction of the two RNA recognition domains in the absence of RNA. Subsequent NMR relaxation studies confirmed that the RRM domains of CPEB4 are tumbling as a unit already in their apo form.

For CPEB1, dimerization of a minor part of the population was observed *in vivo* and the RNA binding domains have been identified to be essential for dimerization. This

study furthermore revealed that RNA binding was abrogated upon self-association of CPEB1. Therefore it was postulated that the dimerization functions to inactivate spare proteins, preventing them from inducing the polyadenylation of RNAs with low affinity binding sites and also serving as molecular hubs that release polyadenylation factors upon dimer destruction. We have been interested whether dimerization would also occur for the RRM domains of CPEB4. To address this question semi-native SDS-PAGE gels and IM-MS were applied. Indeed, we observed minor dimer populations for both the isolated RRM1 domain and the RRM1-RRM2 pair. Interestingly and in contrast to CPEB1, in the case of CPEB4 dimerization does not preclude RNA binding activity. Further experiments with the full-length CPEB4, as investigated with CPEB1, would ascertain if the inhibition of RNA binding upon dimerization is exclusive of CPEB1 or is occurring also in CPEB4 when investigating the full-length protein.

The interaction of the RRM domains with several CPE- containing motifs (consensus and non-consensus) has been investigated by ITC and NMR titration studies. The results reveal that for high affinity RNA binding activity both RRM domains are important suggesting a cooperative interaction. Based on the data we assume that the role of the RRM2 in the RNA interaction is to increase the global surface of the tandem to recognize the RNA with respect to the single RRM1 and therefore helping to best accommodate the RNA ligand.

Since a possible application of our study is the design of inhibitor molecules, crystallization trials were set-up and preliminary results from diffraction data are presented.

In the second part of the project we focused on the N-terminal region of CPEB1 and its interaction with the peptidyl-prolyl isomerase Pin1. Our aim was to gain insight into the recognition of Pin1 and CPEB1 particularly with regard to the differences and similarities of the phosphorylation independent and phosphorylation dependent interaction. By NMR we have shown the interaction of Pin1 WW domain with two unphosphorylated N-terminal fragments of CPEB1 NF1s (88-183) and NF2 (196-293). A possible purpose of this low affinity association might be to establish spatial proximity of CPEB1 and Pin1 to guarantee rapid degradation of CPEB1 upon its phosphorylation by Cdc2.

Regarding the phosphorylation dependent interaction, we showed that Pin1's WW domain binds all of the six Cdc2 phosphorylation sites. These interactions were determined using recombinant proteins, phospho-peptides and applying biophysical techniques such as ITC and NMR. The influence of the phosphate group on binding the Pin1 WW domain was demonstrated by ^{31}P titrations. Interestingly, the Cdc2-phosphorylation sites CPEB1 pS184 and CPEB1 pS210, which are located in proximity to the SCF $^{\beta\text{-TrCP}}$ E3 ubiquitin ligase interaction site (TSG-motif:190-195), are the highest affinity binders. Among all the potential substrates studied CPEB1 pS210, which has also been identified to be essential for CPEB1 degradation, plays a special role, since it is the only ligand tested which showed binding activity not only to the

WW domain but also to the catalytic domain. Our NMR relaxation studies showed an increased inter domain interaction in Pin1 upon binding of CPEB1 pS210, whose function might be to regulate the conformations sampled by the catalytic site thus allowing a fine tuning of this remote functional site by substrate binding. The NMR structure of the Pin1 WW - CPEB1 pS210 complex revealed that the pSP motif is bound in *trans* configuration through contacts with amino acids located in the first turn of the WW domain and the conserved tryptophan in the β_3 -strand. For the role of Pin1 in CPEB1 degradation we suggest that Pin1 catalyzes the degradation by ensuring a predominant *trans* configuration which is accessible for SCF ^{β -TrCP} E3 ubiquitin ligase mediated degradation.

7 Appendix

7.1 CPEB4 RRM domains

CPEB4 RRM1 domain

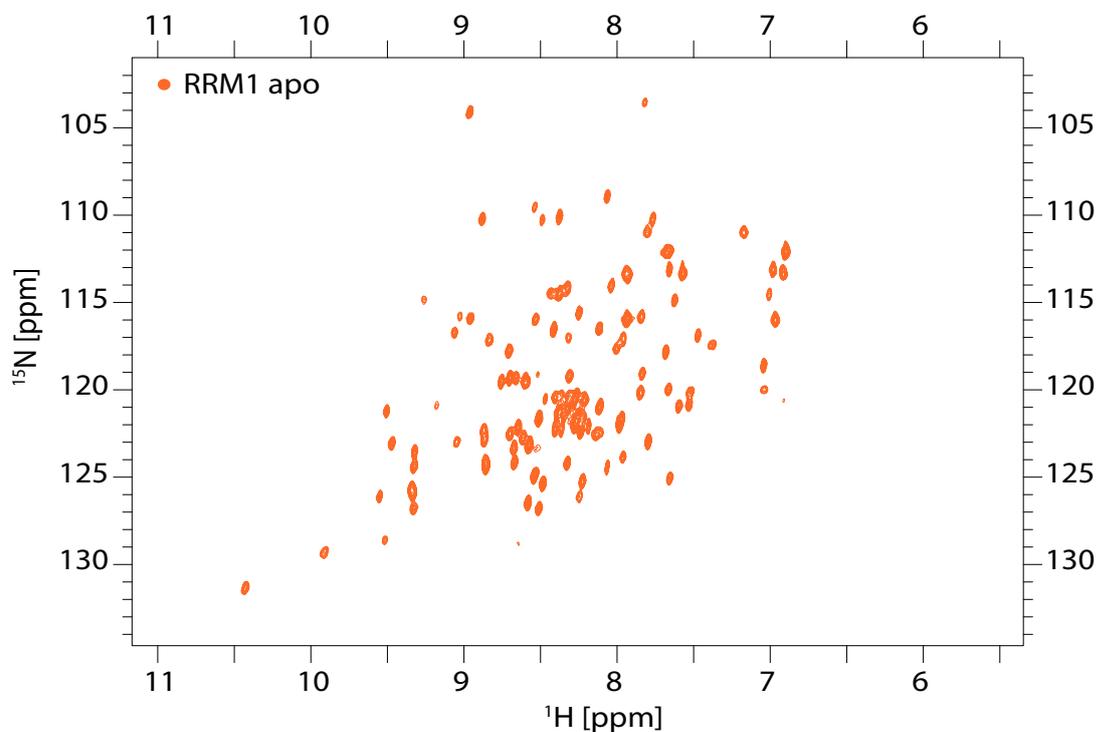


Figure 7.1: ^1H , ^{15}N SQC spectrum of RRM1 free with a good signal dispersion and little signal overlapping indicating a fully folded protein. The spectrum was acquired at 303 K on a 400 μM sample using a Bruker AVIII 600-MHz spectrometer equipped with a room temperature probe.

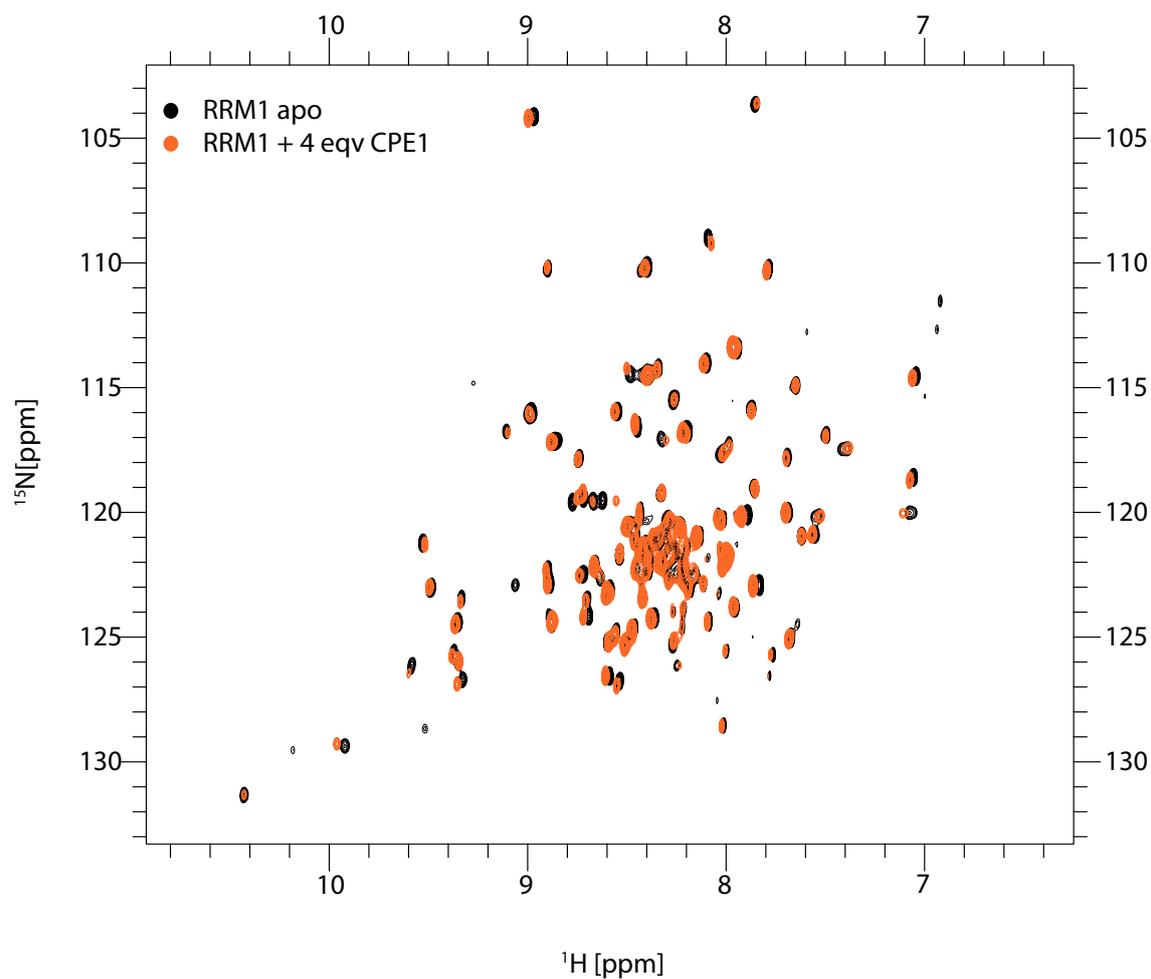


Figure 7.2: Superposition of ^1H , ^{15}N - HSQC spectra of RRM1 free (black) and after titration of 4 equivalents of U5A2U1 (orange). The Chemical Shift Perturbations observed indicate that the binding activity for the single RRM1 is only low affinity interaction. This result is consistent with ITC results obtained in this study.

CPEB4 RRM1-RRM2 tandem

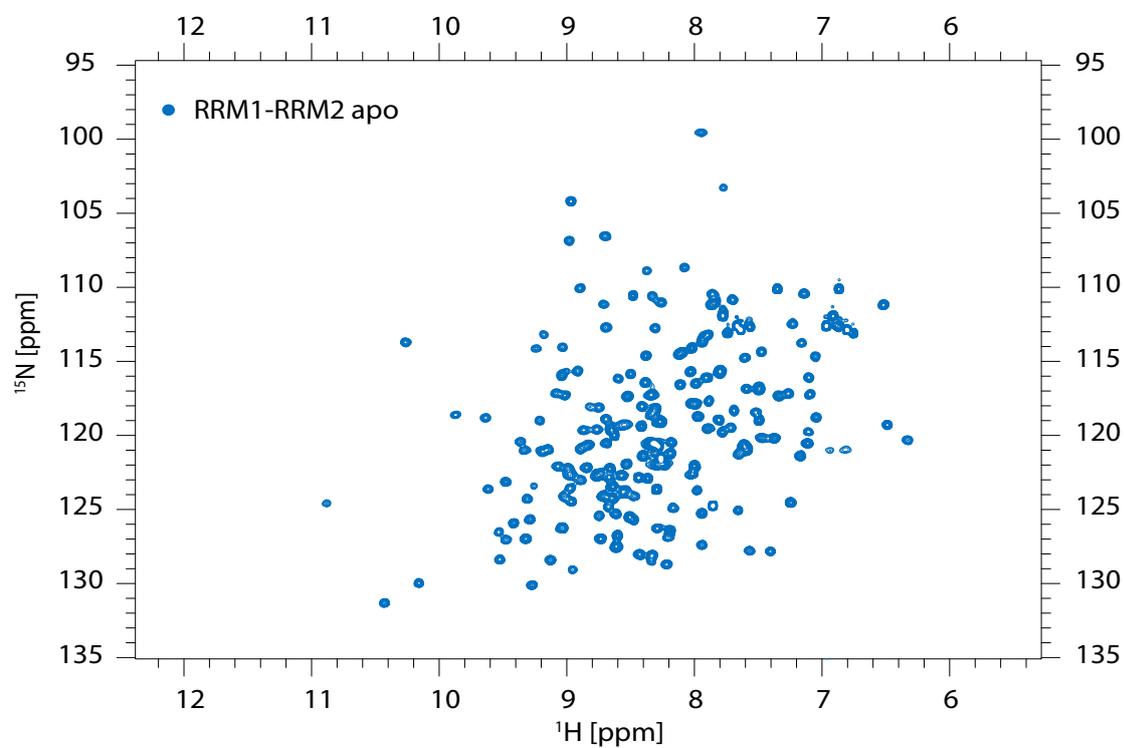


Figure 7.3: ^1H , ^{15}N - HSQC spectrum of RRM1-RRM2 free acquired at 303 K using a Bruker AVIII 600-MHz spectrometer equipped with a 5 mm TXI cryprobe, z-gradient.

Homology Models

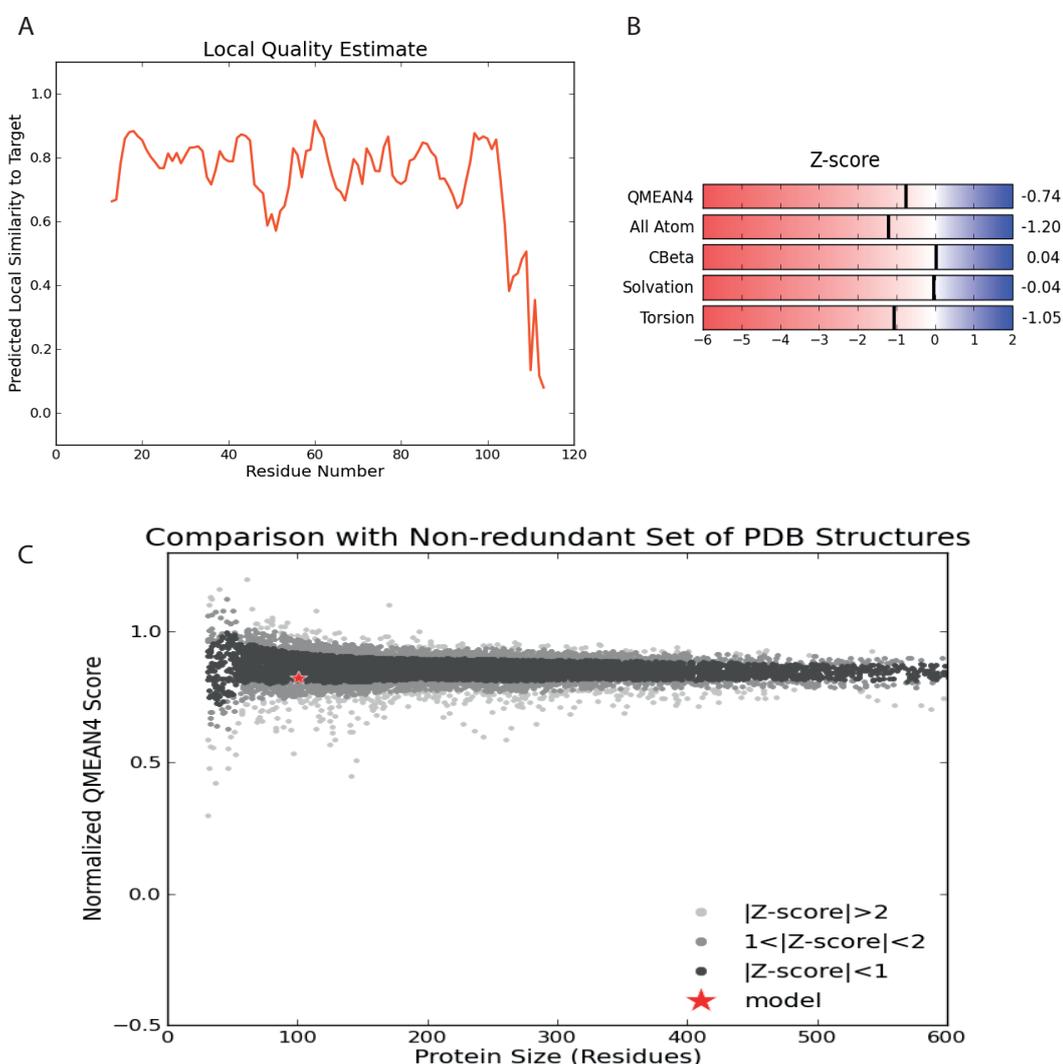


Figure 7.4: A model for CPEB4 RRM1 was built with SWISS-MODEL (<http://swissmodel.expasy.org/>), As a template served the NMR structure CPEB3 RRM1 (PDB entry 2DNL) which has a sequence identity of 97.06 % (alignment see Figure1). (A) Local estimates of the model quality based on the QMEAN scoring function are shown as per-residue plot. (B) Analysis of the Z-scores of the individual terms contributing to the QMEAN score. “Good structures“are supposed to have all sliders in the light red to blue region. With a QMEAN score of -0.74 we consider it a good model. (C) The model’s QMEAN score is compared to the scores obtain for experimental structures (high resolutions PDB-structures) of similar size (model size +/- 10 %) and a Z-score is calculated.

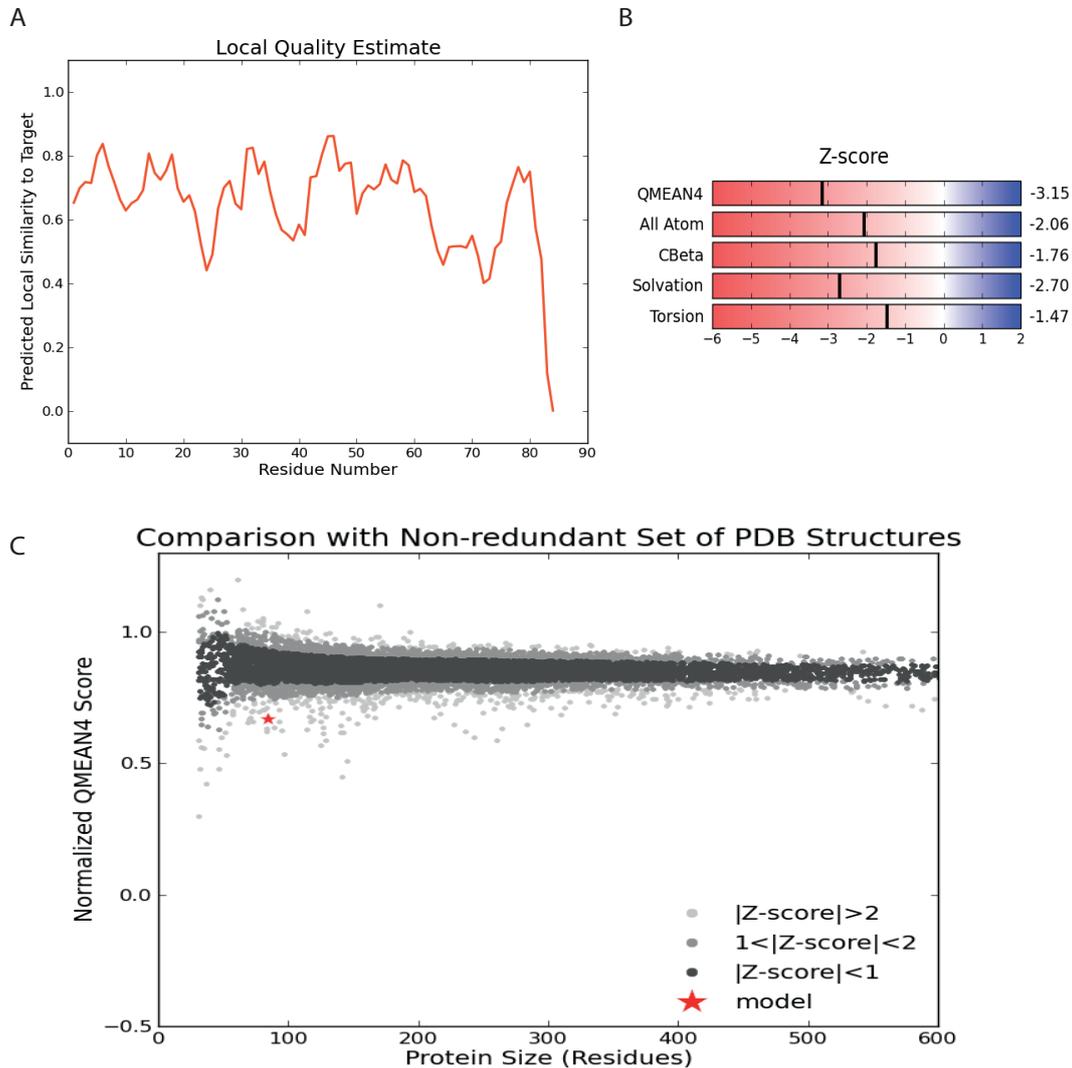


Figure 7.5: A model for CPEB4 RRM2 was built with SWISS-MODEL (<http://swissmodel.expasy.org/>), We used the structure of RRM1 of NCL protein (PDB entry 2FC8,) with a sequence identity of 28.95% as a template. (A) Local estimates of the model quality based on the QMEAN scoring function are shown as per-residue plot. (B) Analysis of the Z-scores of the individual terms contributing to the QMEAN score. With a QMEAN score of -3.15 the RRM2 model cannot be considered well modeled, and must be treated with caution. Since our interest only focused on the spatial arrangement of the two models and not in any structural details of the RRM2, we consider the obtained model sufficiently good for that purpose. (C) The model's QMEAN score is compared to the scores obtain for experimental structures (high resolutions PDB-structures) of similar size (model size +/- 10 %) and a Z-score is calculated.

7.2 Glossary

Å	Ångström
Ac ₂ O	Acetic anhydride
Bis-Tris	Bis(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane
Cdc2	Cell division cycle protein 2 homolog
CPE	Cytoplasmic Polyadenylation Element
CPEB	Cytoplasmic Polyadenylation Element Binding Protein
CSP	Chemical Shift Perturbation
cDNA	Complimentary DNA
Da	Dalton (g/mol)
DCM	Dichloromethane
DMF	N,N-Dimethylformamid
DMSO	Deoxyribonucleic acid
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
eIF4E	Eukaryotic translation initiation factor 4E
ExpPASy	Expert protein analysis system
FID	Free Induction decay
Fmoc	9H-Fluoren-9-ylmethoxycarbonyl
GST	Gluthation-S-Transferase
HATU	N,N-Methylmethan-aminium hexafluorophosphate
HOAt	1-hydroxy-7-aza-benzotriazole
HOBt	1-hydroxy-benzotriazole
HPLC	High Performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
IM	Ion Mobility
IPTG	Isopropyl-β-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K _D	Dissociation constant
LB	Lysogeny broth
mRNA	messenger RNA
MS	Mass Spectrometry
MW	Molecular weight
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
nt	Nucleotide
OD	Optical density
PCR	Polymerase chain reaction
PEG	Polyethylene glycol

pH	Pondus hydrogenium
Pin1	Protein interacting with NIMA (never in mitosis A) -1
poly(A)	polyadenosine
ppm	parts per million
RNA	ribonucleic acid
RNase	Ribonuclease
RRM	RNA recognition motif
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TOCSY	Total correlation spectroscopy
TROSY	Transverse relaxation optimized spectroscopy

RNA recognition and self-association of CPEB4 is mediated by its tandem RRM domains

Constanze Schelhorn¹, James M.B. Gordon¹, Lidia Ruiz¹, Javier Alguacil², Enrique Pedroso² and Maria J. Macias^{1,3,*}

¹Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10, Barcelona 08028, Spain, ²Departament de Química Orgànica and IBUB, Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona, Spain and ³Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, Barcelona 08010, Spain

Received June 3, 2014; Revised July 11, 2014; Accepted July 19, 2014

ABSTRACT

Cytoplasmic polyadenylation is regulated by the interaction of the cytoplasmic polyadenylation element binding proteins (CPEB) with cytoplasmic polyadenylation element (CPE) containing mRNAs. The CPEB family comprises four paralogs, CPEB1–4, each composed of a variable N-terminal region, two RNA recognition motif (RRM) and a C-terminal ZZ-domain. We have characterized the RRM domains of CPEB4 and their binding properties using a combination of biochemical, biophysical and NMR techniques. Isothermal titration calorimetry, NMR and electrophoretic mobility shift assay experiments demonstrate that both the RRM domains are required for an optimal CPE interaction and the presence of either one or two adenosines in the two most commonly used consensus CPE motifs has little effect on the affinity of the interaction. Both the single RRM1 and the tandem RRM1–RRM2 have the ability to dimerize, although representing a minor population. Self-association does not affect the proteins' ability to interact with RNA as demonstrated by ion mobility–mass spectrometry. Chemical shift effects measured by NMR of the apo forms of the RRM1–RRM2 samples indicate that the two domains are orientated toward each other. NMR titration experiments show that residues on the β -sheet surface on RRM1 and at the C-terminus of RRM2 are affected upon RNA binding. We propose a model of the CPEB4 RRM1–RRM2–CPE complex that illustrates the experimental data.

INTRODUCTION

The cytoplasmic polyadenylation element binding protein (CPEB) is involved in translational regulation of cytoplasmic polyadenylation element (CPE) containing mRNAs. The mechanism by which cytoplasmic polyadenylation controls the translation of many key mRNAs has been elucidated in meiotic maturation of *Xenopus* oocytes. Many maternal mRNAs are translationally repressed in oocytes arrested at the end of meiosis prophase I. This repression is achieved via the interaction of CPEB to the CPE U-rich sequence (U4-6A1-2U1-2) in the 3'UTR of the specific mRNAs (1). Moreover, CPEB recruits protein factors that facilitate translational regulation. The poly A-specific ribonuclease PARN has an antagonistic effect on the non-canonical poly A polymerase Gld2 resulting in mRNAs with a short poly A tail and therefore transitionally repressed (2). Other factors that interact with CPEB include members of the mRNA 3'-end processing machinery including symplekin and CPSF. An interaction with the protein Maskin and its interaction with the translational initiation factor eIF4E inhibits the assembly of the translation initiation complex (3,4). A progesterone initiated signaling cascade results in the activation of kinase Aurora A that in turn phosphorylates CPEB (5). The phosphorylation leads to the inhibition of the PARN–CPEB interaction resulting in the extension of the poly A tail of the mRNA by Gld2 allowing for the binding of poly A binding proteins and subsequent translational activation (6).

The CPEB family of proteins comprises of four paralogs (CPEB1–4). CPEBs 2–4 are closely related and CPEB1 is the most distant relative (7). The proteins are widely expressed in a variety of tissues, cell types and tumors with partially overlapping patterns (7). Additional functions to that of oocyte maturation have been identified including roles in synaptic plasticity and cellular senescence (8). Members of the family comprise of a largely conserved C-terminal region that comprises of two RNA recognition

*To whom correspondence should be addressed. Tel: +34 934037189; Fax: +34 934039976; Email: maria.macias@irbbarcelona.org

Present address:

Javier Alguacil, Institut des Biomolécules Max Mousseron (IBMM), Université Montpellier 2, Place E. Bataillon, Montpellier cedex 5, 34095, France.

© The Author(s) 2014. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

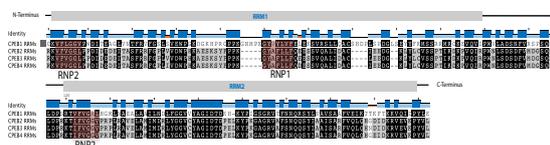


Figure 1. Alignment was prepared using the Geneious R6 software (biomatters). The position of the RRM domains is indicated and the RNP motifs highlighted. Black residues are conserved between all family members. Gray residues are partially conserved and blank residues are unconserved.

motifs (RRM1 and RRM2) that have been shown to be essential for the CPE interaction, and a ZZ-type domain that is postulated to be a protein–protein interaction module (9,10). The N-terminal half of the protein is more variable between the different members and contains a PEST-degradation motif, the site for Aurora A mediated phosphorylation and several uncharacterized phosphorylation sites. The two most commonly used CPE consensus motifs are UUUUAAU and UUUUAUU (11,12). Although there are other non-canonical motifs, studies have shown that CPEBs 2 and 4 recognize and regulate mRNA transcripts containing the same canonical CPE motifs that CPEB1 regulate (13,14–17).

The structure of the RRM1 domain of CPEB3, which is completely conserved when aligned to CPEB2 and 4, has been solved by Nuclear Magnetic Resonance (NMR) (PDB code: 2DNL) and shows the canonical RRM domain fold of a $\alpha\beta$ -sandwich with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology with the addition of two β strands immediately before the β_4 strand (β_4' and β_4''). Two conserved motifs, which are generally involved in RRM–RNA interactions, RNP1 and RNP2 lie in the central strands of the RRM β -sheet (18). The consensus sequence of RNP1 is (RK)-G-(FY)-(GA)-(FY)-(ILV)-X-(FY), while that of RNP2 is (ILV)-(FY)-(ILV)-X-N-L. The first RRM domain (RRM1) contains, for the most part, both RNP motifs. However, the second (RRM2), more C-terminal domain lacks the consensus sequence of an RNP1 motif. This characteristic is shared among all CPEB family members (Figure 1).

Previously, Lin *et al.* (19) showed that the full-length CPEB1 protein has the ability to dimerize. Deletions of the RRM motifs from the full-length protein showed that this oligomerization was mediated by the pair of RRM domains. Moreover, the dimers of CPEB1 seem to lose the ability to interact with RNA (19), suggesting that the formation of dimers prevents RNA recognition, maintaining the protein in an inactive state.

Prompted by these observations in CPEB1 and in order to characterize the binding properties of the RRM domains present in CPEB4, we have examined the RNA interactions of the CPEB4 RRM tandem and also of the individual RRM1 domain, using a combination of biochemical, biophysical and NMR-based techniques. The chemical shift effects seen by NMR of the apo form proteins, when comparing RRM1 to the tandem construct RRM1–RRM2, indicates that the domains are in close proximity to each other. Furthermore, isothermal titration calorimetry (ITC) and electrophoretic mobility shift assay (EMSA)

experiments have shown that both domains are necessary for optimal binding to the CPE RNA motifs and that the presence of one or two adenosines in the motif has little or no effect on the binding affinity. The binding site for this RNA interaction has been mapped using NMR titrations of the CPEs with the RRM1–RRM2 tandem construct and the observations made indicate that the binding site is localized largely on the β -sheet surface, including the conserved RNP motifs of RRM1, and the C-terminus of RRM2. We also report the ability of both the single RRM1 domain and the RRM1–RRM2 pair to form dimers, as previously reported for the full-length CPEB1 protein (19). In the case of the recombinant RRM1 and the RRM1–RRM2 pair, the dimer population represents a minority species in solution, below 10%. Remarkably, unlike CPEB1, the CPEB4–RRM dimers retain the ability to bind RNA as demonstrated by ion mobility–mass spectrometry (IM-MS). We have proposed a model of the RRM1–RRM2–CPE complex based on the data generated and reported in this manuscript. This model and the chemical shift assignments of the RRM1–RRM2 pair will provide a platform for further studies aimed at the design of protein–RNA inhibitors, target to the CPEB family of proteins.

MATERIALS AND METHODS

Cloning of human CPEB4 RRM domains

Complementary DNA (cDNA) from human CPEB4, isoform 1, was used as a template to clone into the pETM-11 vector (a gift from the EMBL-Heidelberg Protein Expression Facility). The constructs cloned correspond to the following boundaries: RRM1 (463–573) and RRM1–RRM2 (463–665). Residue numbering is according to isoform 1 of CPEB4 (Uniprot: Q17RY0). All clones were confirmed by DNA sequencing. *Escherichia coli* DH5 α strain was used for cloning.

Protein expression and purification

Both constructs were expressed at 37°C in *E. coli* BL21 (DE3) following induction at an optical density of 0.6 (600 nm) with 0.5 mM IPTG overnight. Unlabeled and labeled samples were prepared using LB and minimal media (M9) cultures, respectively. D₂O (99.89%, CortecNet), ¹⁵NH₄Cl and/or D-[¹³C] glucose were used as sole hydrogen, nitrogen and carbon sources respectively to prepare the labeled samples as described (20). All proteins were expressed fused to an N-terminal His₆-tag followed by a TEV-protease cleavage site. Cells were lysed using an EmulsiFlex-C5 (Avestin) in the presence of lysozyme and DNaseI. The soluble supernatant was purified by nickel-affinity chromatography (HiTrap Chelating HP column, GE Healthcare Life Science, Uppsala, Sweden) using 150 mM NaCl, 20 mM Tris-HCl, 10 mM imidazole pH 8.0 and 50 mM EDTA. Eluted His₆-RRM constructs were digested with TEV protease at room temperature overnight, and further purified by size-exclusion chromatography on a HiLoad™ Superdex 75 16/60 prepgrade columns (GE Healthcare) equilibrated in 20 mM Tris pH 7.0 and 130 mM NaCl.

EMSA

Binding reactions were carried out for 20 min at 4°C in 10 μ l of binding buffer [40 mM Tris pH 7.5, 150 mM NaCl, 0.1% (w/v) β ME]. Increasing amounts of protein were incubated with a fixed concentration of 32 P labeled RNA (~2 nM). Electrophoresis was performed in non-denaturing 8.0% (29:1) polyacrylamide gels. The gels were run for 4 h in 1 \times TBE buffer (90 mM Tris, 64.6 mM boric acid, 2.5 mM EDTA pH 8.4) at 100 V at 4°C. The gels were dried and exposed to a PhosphorImager screen overnight. The screen was scanned on a Molecular Dynamics Storm 840 PhosphorImager.

RNA synthesis and purification

The CPE-containing RNA fragments U5A2U1 and U5A1U2 were synthesized at the 4 \times 1 mmol scale using the standard phosphite triester approach in a 3400 Applied Biosystems synthesizer (21). After cleavage and deprotection, the oligonucleotides were purified by reverse-phase HPLC using a Jupiter C18 semipreparative column (10 μ m, 300 Å, 250 \times 10 mm) from Phenomenex with a Waters 600 HPLC System (buffer A: 0.1 M triethylammonium acetate, buffer B: acetonitrile). The RNA triethylammonium counter-ion was exchanged by sodium using a cation exchange resin (Dowex 50Wx4, 200–400 mesh, Sigma-Aldrich). The final products were analyzed by MALDI-TOF MS, lyophilized and stored at –20°C until further use.

Isothermal titration calorimetry

ITC measurements were performed using a nano ITC calorimeter (TA Instruments) at 5°C, 12°C or 25°C. To be consistent with the condition used for the NMR studies, protein samples were prepared in the NMR buffer (20 mM Tris-HCl pH 7.0 and 130 mM NaCl). Buffer from concentrating the protein samples was used to resuspend the RNA. All samples were centrifuged and degassed prior to the experiments. Protein concentration was measured in a NanoDrop™ 2000 measuring the UV absorption at 280 nm. RNA concentration was determined spectrophotometrically. Depending on the expected affinity, sigmoidal curves were optimized by injecting 5- to 10-fold concentrated RNA in 16 \times 3 μ l steps in a cell containing 190 μ l of protein at 20–50 μ M adjusted concentration. The delay between injections was 3 min and data was collected while stirring at 200 revolutions per minute (rpm). The NanoAnalyze software (TA Instruments) was used to analyze the binding isotherms assuming a single binding site in each molecule. Baseline controls were acquired with buffer and pure RNA solutions. Measurements have been repeated at least twice.

Ion mobility–mass spectrometry

An ion mobility mass spectrometer features an ion mobility cell before the mass analyzer. Gaseous ions are injected into the cell and accelerated by a weak electric field. Due to the presence of buffer gas in this cell, low-energy collisions with the buffer gas occur. The higher the collision cross sections (CCS) of the ion, the greater the number of collisions

with the buffer gas. As collisions increase, an energy loss of the ions occur and accordingly ions take a longer time to cross the IM cell (the ‘drift time’). Consequently, ions are injected into the mass analyzer, achieving a simultaneous separation on the basis of the CCS to charge ratio and the m/z ratio. The three-dimensional spectrum obtained consists of mass, drift time and relative intensity. Traveling wave IM-MS experiments were performed on a Synapt G1 HDMS mass spectrometer (Waters, Manchester, UK). Samples were placed on a 384-well plate refrigerated at 15°C and sprayed using a Triversa NanoMate® (Advion Bio-Sciences) automated Chip-Base nano-electrospray working in the positive ion mode. The instrument was calibrated over the 500–8000 Da m/z range using a cesium iodide solution. The software MassLynx 4.1 SCN 704 and Driftscope 2.1 were used for data processing. Samples containing the RNA complexes with either CPEB4-RRM1 or CPEB4 RRM1-RRM2 tandem or with the independent RRM domains (final concentrations of 30–50 μ M) were prepared in 20–50 mM NH₄OAc pH 7.2. Prior to analysis, a 1D-¹H NMR spectrum was acquired on all samples in order to check the stability of the samples and to compare it with spectra obtained in NMR buffer conditions. Spray voltage was set to 1.75 kV and delivery pressure to 0.5 psi. A reduction of the source pumping speed in the backing region (5.85 mbar) was set for optimal transmission of high mass non-covalent ions. Cone voltage, extraction cone and source temperature were set to 20 V, 6 V and 20°C respectively. Ions passed through a quadrupole mass filter to the IM-MS section of the instrument.

NMR chemical shift perturbation experiments

Experiments were recorded at 303 K using a Bruker AVIII 600 MHz spectrometer equipped with a 5 mm TXI cryoprobe, z-gradient. Protein samples of the CPEB4 RRM1 and the CPEB4 RRM1-RRM2 were equilibrated in a buffer containing 20 mM Tris-d₁₁, 130 mM NaCl and 5% DMSO-d₆. All samples were supplemented with 10% D₂O and pH adjusted to value 7. Spectra were acquired using 200 μ M ¹⁵N-labeled protein samples equilibrated together with progressively increasing amounts of the unlabeled RNA fragments until saturation was achieved. Chemical shift perturbation analyses were performed on CcpNmr Analysis (22) with a 0.15 weighting of ¹⁵N with respect to ¹H.

NMR backbone assignment

Spectra were acquired at 303 K using a Bruker AVIII 600 MHz spectrometer, equipped with cryogenic or room temperature triple resonance gradient probes. Backbone ¹H, ¹³C and ¹⁵N resonance assignments were obtained by analyzing the 3D HNCACB and HN(CO)CACB experiment pair in the case of the CPEB4 RRM1 (fully protonated sample) or the CBCANH and CBCA(CO)NH pair in the case of the CPEB4 RRM1-RRM2 (²H, ¹³C, ¹⁵N samples). Transverse relaxation-optimized spectroscopy (TROSY) versions of these experiments and/or the non-uniform sampling (NUS) acquisition strategy were used in selected cases to reduce experimental time and increase resolution. All buffer conditions were as mentioned. NMRPipe (23) was

used for spectra processing. CARA (24) and CcpNmr Analysis (22) were used for spectra analysis and assignment.

Relaxation measurements

Amide relaxation measurements were acquired on a 500 μM ^{15}N -labeled RRM1–RRM2 sample essentially as described (25). NMR experimental setup details were essentially as reported (26). Briefly, the T_1 and T_2 experiments were acquired using $135 (t_1) \times 2048 (t_2)$ total real points. Twelve different relaxation time values (22, 54, 108, 162, 270, 432, 540, 702, 864, 1080, 1404 and 1728 ms) were measured to determine T_1 . To determine T_2 ten experiments were recorded with the following ^{15}N mixing times: 16.74, 33.48, 50.22, 66.96, 100.44, 117.18, 133.92, 167.4, 184.14, 200.88 ms. All relaxation experiments were acquired as pseudo-3D experiments and converted to 2D data sets during processing. Peak integration values were fitted to a two-parameter function as described in Equation (1):

$$I(t) = I_0 e^{\left(\frac{-t}{T_{1,2}}\right)} \quad (1)$$

where I_0 and $I(t)$ are the peak intensities at times 0 and t , respectively. The optimum value of the I_0 and the $T_{1,2}$ parameters was determined using the Levenberg–Marquardt optimisation algorithm for minimization of χ^2 goodness of fit parameter as reported.

The rotational correlation time of the RRM1–RRM2 pair was calculated with Equation (2), using the approximation of slow molecular motion τ_c larger than 0.5 ns and assuming that only $J(0)$ and $J(\omega_N)$ spectral density terms contribute to the overall value. ν_N is the ^{15}N resonance frequency ($60,08 \times 10^6$ Hz)

$$\tau \approx \frac{1}{4\pi\nu_N} \sqrt{\left(6\frac{T_1}{T_2} - 7\right)} \quad (2)$$

Regarding the heteronuclear Nuclear Overhauser Effect (NOE) experiment, the reference and the presaturated Heteronuclear single quantum correlation spectroscopy (HSQC) spectra were acquired in an interleaved manner. The values of the steady-state ^1H - ^{15}N NOEs resulted from the ratios of the peak intensities measured in the reference (I_0) and the presaturated (I_S) spectra during the relaxation delay as described (27). Background noise levels σ_S and σ_0 were measured and used to determine the NOE standard deviation through the following relationship:

$$\frac{\sigma_{\text{NOE}}}{\text{NOE}} = \left(\left(\frac{\sigma_{I_S}}{I_S} \right)^2 + \left(\frac{\sigma_{I_0}}{I_0} \right)^2 \right)^{\frac{1}{2}}$$

RESULTS

Interactions of the RRM domains in the absence of RNA

To gain insight into which role the two RRM domains play in RNA binding we prepared several constructs of the isolated RRM1, RRM2 and the RRM1–RRM2 tandem domains using His-tagged fusion expression vectors. The isolated RRM1 domain and the tandem construct were in the

soluble fraction of the expression cultures and they were purified by Ni-affinity chromatography. However, overexpression of the isolated RRM2 using various experimental conditions resulted in the formation of inclusion bodies containing the RRM2 protein. Protein purified under denaturing conditions, in the presence of guanidium chloride, could not be refolded by dialysis under the various buffer conditions we have assayed.

We therefore focused our work on the construct containing the pair of RRM domains and also on the isolated RRM1, which was used for comparison. In order to assign the amide and backbone resonances and to facilitate the investigation into whether the first RRM domain adopts a similar fold when being independently expressed or in the construct of both consecutive RRM domains $^{13}\text{C}/^{15}\text{N}$ and $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ labeled samples were prepared respectively. Due to its size and compact fold, the acquisition of the triple resonance experiments of the tandem required the deuteration of the sample in order to minimize the loss of the signal caused by transverse T_2 relaxation. Applying TROSY versions of all NMR experiments was essential to improve the resolution and sensitivity of the backbone experiments.

A comparison of the secondary ^{13}C chemical shifts obtained for the samples to reference values indicate that both RRM1 and RRM2 in the tandem construct adopt the canonical $\alpha\beta$ sandwich structure with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology with an additional β -strand β_4 at the C-terminus of RRM1 (Figure 2A). Comparison of the ^{13}C chemical shifts of RRM1, when assigned independently or in the RRM1–RRM2 construct, indicates that the secondary structure is not altered due to the presence of RRM2 (Figure 2B). For RRM1, a homology model was built using SWISS-MODEL (Template: PDB entry 2DNL, RRM1 of CPEB3, sequence identity 97%, <http://swissmodel.expasy.org/>; a quality report is shown in Supplementary Figure S1). The topology of the model obtained is consistent with the elements of secondary structure indicated by the analysis of the ^{13}C chemical shifts. The spatial arrangement of the canonical four-stranded antiparallel β -sheet is $\beta_4\beta_1\beta_3\beta_2$. The homology model we obtained indicates that the additional β_4 strand is arranged antiparallel to β_4 resulting in the following order for the extended β -sheet: $\beta_4\beta_4\beta_1\beta_3\beta_2$.

HSQC-TROSY experiments for the single domain (Supplementary Figure S2) and for the pair (Supplementary Figure S3) respectively were compared to assess the influence that the domains may have on the chemical shifts of each other. The superimposition of both spectra (Figure 2C) acquired under identical conditions reveals that the comparison of the amide resonances only matches reasonably well in certain areas. Significant differences in the chemical shift of many residues are observed between the RRM1 and the RRM1–RRM2 construct. The linker connecting both domains is very short in length (six residues), limiting the freedom of the domains in the tandem. The comparison of the linewidths (Supplementary Figure S4) of the independent RRM1 domain to that of the construct containing both domains indicates a clear broadening of the amide signals characteristic of a 23 kDa sample and that the RRM tandem behaves as a unit. Moreover, this observation is supported by the homogeneous T_1 , T_2 and het-

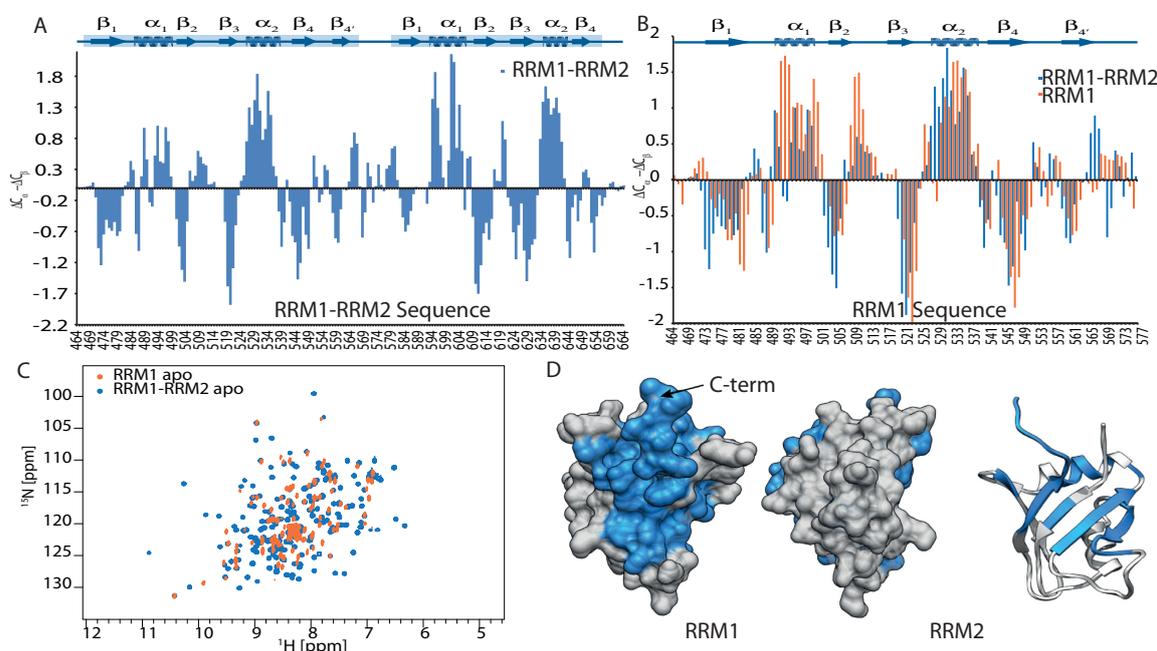


Figure 2. (A) Comparison of ¹H, ¹⁵N-HSQC spectra of the apo RRM1 (orange) and apo RRM1-RRM2 (blue). Backbone assignment of both constructs allows us to identify the chemical shift of the residues in both samples. (B) Significant differences in chemical shifts are mapped onto the homology model for RRM1. As those differences are clustering on one surface of the model, we assume that this surface lies in proximity of the RRM2. Secondary ¹³C chemical shifts $\Delta C_{\alpha} - \Delta C_{\beta}$ of CPEB4 RRM1-RRM2 (D) and comparison of the CSI of RRM1 single and in the RRM1-RRM2 tandem (C). Residues with negative CSI indicate β -strand conformation whereas positive values indicate turn or α -helix. (D) Secondary structure of RRM1 is not significantly altered due to the presence of RRM2.

eronuclear NOE values (Figure 3A, B and D respectively) measured for the tandem along its entire sequence. The calculated rotational correlation time ($\tau_c = 16$ ns) is in agreement with the average value obtained for a globular protein of this size (Figure 3C) (28). This further confirms that the CPEB4 RRM1-RRM2 domains do not tumble independently. Based on these evidence, we interpret that the chemical shift changes observed in the RRM1 domain indicate the presence of a conformation where both RRM domains are close to one another. Furthermore, when mapping the most affected residues onto the homology model of the RRM1 domain, they cluster on one surface (Figure 2D), indicating that this surface should be in proximity of the RRM2 domain.

CPEB4 uses both RRM domains to bind CPE motifs

In order to characterize the interaction of the RRM domains with short RNAs containing the CPE motifs, we have used several complementary techniques that allow us to detect and quantify the interactions. For this purpose, we performed EMSA as well as ITC binding assays using both RRM1 and RRM1-RRM2 and different RNA fragments. The EMSA experiments showed that RRM1-RRM2 tandem domain has a higher affinity for the CPE RNA when compared to the RRM1 alone (Figure 4A).

The binding ligands used for ITC experiments were two octamer RNA fragments containing the two consensus CPE motifs, U5A2U1 and U5A1U2. ITC experiments with RRM1 and both ligands showed low-affinity binding with dissociation constants in the high μ M range. The experiments with RRM1-RRM2 as titrate yielded dissociation constants of 323 ± 34 nM for U5A2U1 and 299 ± 28 nM for U5A1U2 (Figure 4B). Both ligands, which differ only in one nucleotide, show very similar K_D values (both lying within the error limit of each other), thus not allowing us to significantly distinguish between the two binding affinities. The affinity increase due to the presence of the RRM2 domain is about 100-fold with respect to the values obtained for the single RRM1 domain. As we have not been able to obtain pure RRM2, the experiment could not be repeated for the single RRM2. However, from our NMR titration data, which allows the identification of the regions affected by RNA binding (Figure 5), it is clear that chemical shift perturbations are observed for residues in both domains. The titration of the RRM1 with U5A2U1 is shown in Supplementary Figure S5. Therefore, our data indicate that both RRM domains are important to maintain optimal RNA-binding activity and affinity. The improvement of the affinity by 100 times in the construct containing the RRM2 domain is quite remarkable since this domain contains a degenerate RNP motif and it was assumed to be a poor

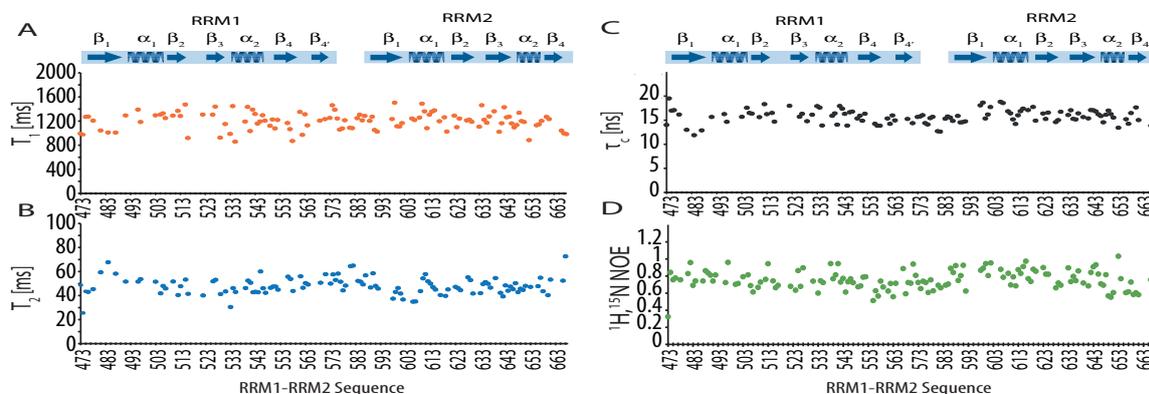


Figure 3. ^{15}N amide relaxation properties of CPEB4 RRM1-RRM2 in their free state. Measurements of longitudinal T_1 (A), transverse T_2 (B) relaxation were performed at 303 K and 600 MHz. Also shown is the rotational diffusion correlation time τ_c derived from the ratio of the ^{15}N T_1/T_2 relaxation times (C). The value $\tau_c \approx 16$ ns is consistent with a compact molecule of 23 kDa. Furthermore, ^1H - ^{15}N heteronuclear NOE (D) experiments were acquired and the results are consistent with the T_1/T_2 relaxation data. Proline and overlapped residues were not included in the analysis.

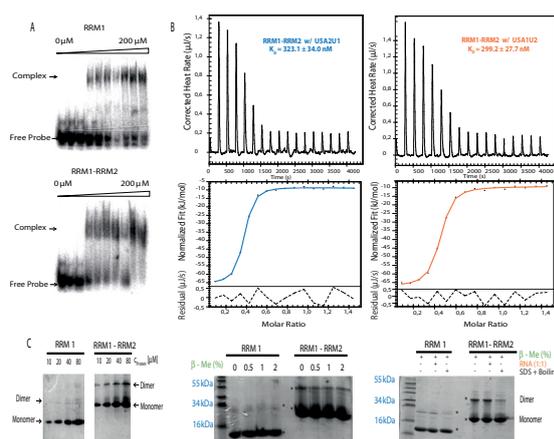


Figure 4. (A) EMSA conducted with increasing amounts of purified CPEB RRM1 or RRM1-RRM2 protein: 0, 1, 25, 50, 75, 100, 125, 150, 175 and 200 (μM) and radiolabelled U5A2U1 RNA (2 nM). Samples were fractionated in native gels and visualized by autoradiography. (B) ITC curves and affinity values obtained for the titration of CPEB4 RRM1-RRM2 with U5A2U1 (blue) and U5A1U2 (orange) RNAs. Data were acquired at 5°C in Tris buffer pH 7.0. The stoichiometry obtained was 0.6 and 0.7 respectively. The data were fitted using the independent model assuming a single binding site. (C) Protein samples were analyzed under semi-native conditions using SDS 12% PAGE. SDS was not present in the loading buffer and the samples were not boiled unless indicated.

RNA binder. This increment in the affinity together with the length of the RNA recognized by the RRM1-RRM2 pair suggests that both RRM domains act cooperatively resulting in high-affinity binding of the RNA. We suggest that the role of the RRM2 in the RNA interaction is dual: it increases the global surface of the pair to recognize the RNA with respect to the single RRM1 and helps adjusting the relative orientation of the two domains to best accommodate the RNA ligand.

Identifying the RNA binding site of RRM1-RRM2 via NMR

To characterize the RNA binding properties of the RRM1-RRM2 construct to the two CPE containing ligands, U5A1U2 and U5A2U1, and to investigate whether there are different binding modes between the consensus CPE motifs, amide chemical shift changes were monitored upon titration of the U5A1U2 and U5A2U1 to the ^{15}N -labeled RRM1-RRM2 domain (Figure 5A). The chemical shift perturbations regarding their range and also the residues affected showed no significant difference between U5A1U2 and U5A2U1 (Figure 5B). We therefore assume that the number of adenosines within the CPE motif does not alter the binding mode. Binding kinetics observed in the NMR titrations of RRM1-RRM2 are in fast to intermediate range on the NMR chemical shift time scale as some residues disappear and reappear with increasing U5A1U2 and U5A2U1 concentration (e.g. Arg⁵⁵⁹; Figure 5A (lower left)).

For both RNA ligands we observe chemical shift perturbations in the canonical RNP motifs of the RRM1 domain, which, according to the secondary structure predictions and the homology model, lie on the β_1 and β_3 strand. In addition, large chemical shift perturbations are also observed for residues lying on β_2 (Trp⁵⁰², Lys⁵⁰⁵) and a C-terminal region of RRM1. These C-terminal residues are predicted to form the β_4 -strand (affected residues: Gln⁵⁵⁷, Ile⁵⁵⁸, Arg⁵⁵⁹) and additional β -strand β_4 (affected residues: Ser⁵⁴⁷, Thr⁵⁵⁰). When mapping the most affected residues of RRM1 on the homology model, they cluster on a region on the β -sheet, which corresponds to the canonical RNA binding surface of RRM domains (Figure 5B).

For RRM2, which only features the RNP2 consensus sequence on β_1 , the strongest chemical shift perturbations were detected in its C-terminal region containing residues with positively charged side groups (e.g. His⁶⁴⁸, Lys⁶⁵³, Arg⁶⁵⁴), indicating that these regions are affected by binding to the RNA fragments either by engaging in direct con-

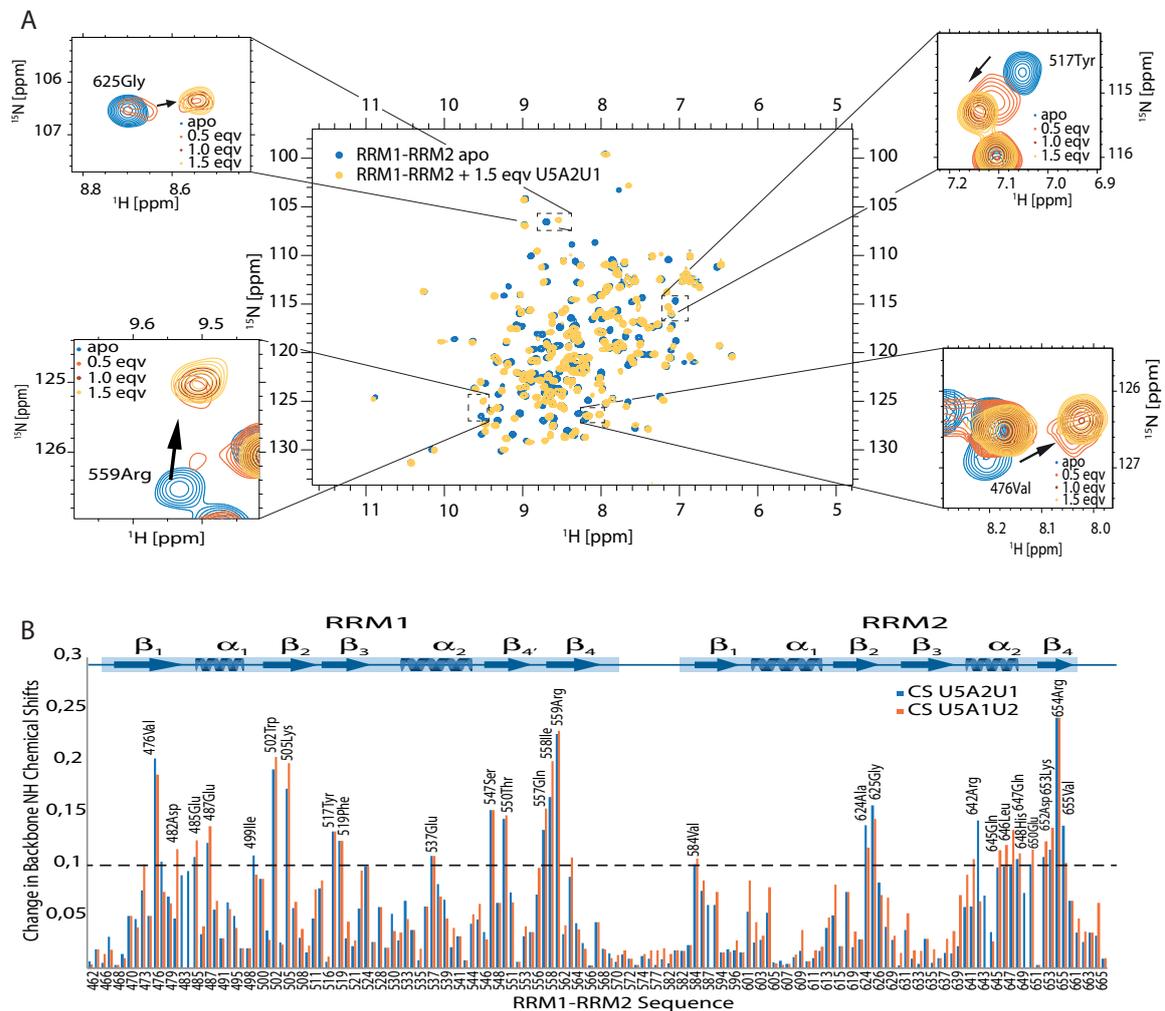


Figure 5. (A) Superimposition of ^1H , ^{15}N -HSQC spectra of RRM1-RRM2 free (blue) and in complex with U5A2U1 (yellow) 1.5 molar equivalents. Zooms in various regions of the spectra are displaying all titration points. Spectra are colored as blue (free); orange, 0.5; dark red, 1.0, and yellow, 1.5 molar equivalents. (B) Chemical shift perturbations of RRM1-RRM2 domains upon binding to U5A2U1 (blue) and U5A1U2 (orange). The Chemical Shift Perturbations (CSPs) are plotted versus residue numbers. The dashed line in gray shows a cut-off $\Delta\delta > 0.1$ ppm (average chemical shift perturbation over all resonances + standard deviation). The CSPs for both RNA ligands do not show any significant differences. Within RRM1, the largest shift perturbations are found in residues that are supposed to lie on the β -sheet; for RRM2, however, they clearly cluster in the C-terminal part. The RNP2 of RRM2 shows shifts just above the cut-off, which suggests that RNP2 is not the main binding site for RRM2, instead the C-terminal positively charged residues seem to play an essential role.

tacts to the RNA or by indirectly induced conformational changes.

RRM1 and RRM1-2 have the ability to dimerize

It has been previously reported by Lin *et al.* (19) that CPEB1 dimerization via its RNA binding regions functions as a regulatory mechanism during cell cycle. We therefore consider the possibility that protein dimerization could also occur in CPEB4. The presence of highly populated dimers

cannot be detected at the protein concentrations used during gel filtration chromatography and also our NMR data were acquired at low protein concentrations (0.2–0.5 mM) to minimize protein precipitation. At this concentration, the main component of the protein solution corresponds to the monomeric protein, as revealed by the relaxation experiments and rotational correlation time measured for the tandem. However, when we performed the ITC experiments to determine the affinities of the different complexes, we observed that the stoichiometries for both the interactions of

RRM1 and RRM1-RRM2 proteins with the RNA ligands were close to but smaller than 1. Repetition of the experiments at several temperatures (5°C, 12°C and 25°C) and using different buffer solutions (tris, phosphate and bis-tris) yielded similar values and stoichiometries. We attributed these small discrepancies in the stoichiometries to the presence of a small population of oligomers that—together with the main complex formed by monomers—may also interact with the RNA.

In an attempt to detect the presence of monomers or other oligomers, and to characterize their effect in the interaction with the RNA, the proteins were analyzed by semi-native sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE). Figure 4C shows that when SDS and reducing agent was absent from the loading buffer and the samples were not denatured by high-temperature incubation, both RRM1 and RRM1-RRM2 migrate as two species on the gel representing both monomeric and dimeric forms of the proteins. The presence of the dimer is maintained through a series of protein concentrations (10–80 μM).

Due to the presence of cysteins in the protein, we were interested in analyzing whether the presence of disulfide interactions could play a role in the formation of the dimeric species. Therefore, we have tested the effect of adding increasing amounts of a reducing agent (0, 0.5, 1 and 2% β -mercaptoethanol) to the loading buffer (Figure 4D). Again, little or no effect on the formation of the dimeric species was observed. Furthermore, the addition of CPE RNA (1:1) to both the protein samples has no effect on the formation of the dimeric species (Figure 4E). However, when SDS was added to both samples and the samples were denatured with temperature, a significant reduction in the dimeric species was observed (Figure 4E). This result indicates that the oligomerization of the proteins is not maintained through disulfide linkages and is dependent on the proteins preserving a proper fold, therefore being mediated by surface electrostatic and/or hydrophobic interactions.

The ratio of dimer to monomer as estimated from the semi-native gels is shifted toward the monomeric population with a maximum dimeric population of 10% of molecules in solution. Given this low percentage of dimer population, its large molecular weight (46 kDa), and the concentration used for the NMR experiments, its contribution to our NMR data is imperceptible.

Dimerization of RRM domains does not disrupt RNA binding

To further study the potential presence of oligomers at the protein concentration used in ITC experiments (20–50 μM) and to investigate the RNA-binding activity of the dimeric species, we used IM-MS on both constructs in their apo form as well as in complex with RNA.

Ion mobility coupled to mass spectrometry (IM-MS) is a technique that simultaneously separates gaseous ions on the basis of their mass, shape and size. The continuous advances in native mass spectrometry applied to IM-MS have prompted its application in the structural study of biomolecular complexes (29). Hence, challenging systems in terms of molecular size, complexity and heterogeneity can be transferred to the gas phase and their structural

properties can be analyzed. IM-MS can provide valuable information about the occurrence and population of the species present in a sample. This technique has been efficiently applied in the characterization of protein-peptide interactions, in particular in the determination of the presence of several species and their role in the calculation of stoichiometries of complexes (30).

Using this technique we could identify in the apo RRM1 sample (Figure 6A) and in the apo RRM1-RRM2 (Figure 6B) the presence of monomeric as well as dimeric species, consistent with the ITC results and the semi-native SDS PAGE. For the complex of RRM1 with RNA, a weak binding affinity was expected. Accordingly, we detected only a minor part of the species in a monomeric complex, the majority remained in its free form (Figure 6C). For the complex with the RRM-tandem domains, the presence of monomeric complexes as well as protein dimers bound to two ligands was detected (Figure 6D). Moreover, no unbound species were detected. The different species (see schematic representation in Figure 6E) were unambiguously identified based on their specific mass-to-charge ratio (MS) and/or their characteristic drift-times that correlate their size-to-charge ratio (IM). Therefore, IM-MS confirms the hypothesis that indeed dimeric species are present. Perhaps, the presence of monomeric and dimeric forms renders the determination of the protein concentration in native conditions inaccurate, explaining the stoichiometry values inferior to one. The results also clearly indicate that only 1:1 protein-RNA complexes are present in both the monomeric and dimeric forms (ML, D2L). The presence of the M1L and D2L species probably contributes to the global affinity measured by the ITC experiments, which is mainly governed by the contribution of the M1L form. The results of the IM-MS experiments with RRM1-RRM2 in complex with RNA reveal that the presence of the ligand does not prevent the formation of dimeric species and does not render the RNA binding site inaccessible. We therefore conclude that the dimerization surface of the RRM1-RRM2 is separate from its RNA-binding interface.

DISCUSSION

In this study we provide evidence that CPEB4 uses both its RRM domains to maintain optimal RNA binding. RRM1, whose sequence is the more conserved of the two tandem domains, relative to other RRM motifs, is only able to bind CPE-containing RNAs in the high μM range. ITC experiments performed with the RRM1-RRM2 pair yielded affinity values around 300 nM, 100 times higher than the isolated RRM1 domain.

An overview of a repertoire of nucleic acid binding modes by proteins containing two RRM domains and the principles of multi-domain protein-RNA recognition has been recently reviewed by Mackereth and Sattler (31) and reported by Barraud and Allain (32). Based on the structures available, they classified different domain arrangements in the free as well as in the bound state (31,32). In the free state, the domains can either be independent, with pre-formed domain contacts, or adopt a closed, autoinhibited form. Upon ligand binding, RRMs can, for example, independently bind separate RNA motifs or form a con-

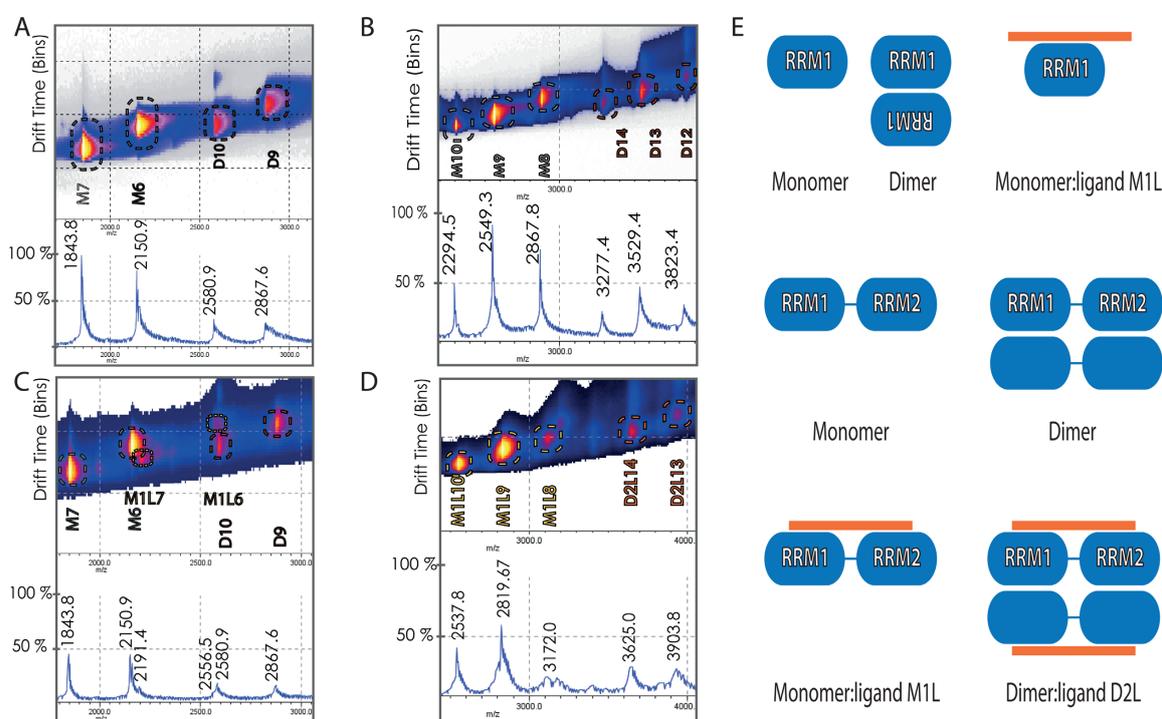


Figure 6. A region of the ion mobility–mass spectrometry data obtained for apo RRM1 (A), apo RRM1–RRM2 (B), RRM1–U5A2U1 complex (C) and RRM1–RRM2–U5A2U1 complex (D). Each ion was assigned to a given species based on its characteristic mobility and mass-to-charge ratio. Abbreviations used are M (monomer, apo state), D (dimer, apo state), M1L (monomer with one ligand) and D2L (dimer with two ligands). Numbers following the species' name indicate the ionization state. For exact ion masses, both observed and theoretical values, please refer to Supplementary Figure S6. We have unambiguously detected the presence of dimers for both RRM1 (A) and RRM1–RRM2 (B) in their apo state. Consistent with ITC results, for the RRM1–U5A2U1 complex (C) only a minor part of monomeric complex was detected (M1L7 and M1L6); the majority of the species represent the ones already detected for the apo state (A). (D) In the RRM1–RRM2–U5A2U1 sample, monomers in complex with one ligand as well as dimers with two ligands were identified. (E) A schematic representation of the species identified by IM-MS. The RRM domains are represented in blue ellipses and the RNA octamer as an orange line.

tinuous RNA-binding platform, stabilized by inter-domain contacts or increased linker rigidity. Another mechanism to stabilize the RNA binding is mediated by protein–protein interactions with an auxiliary protein.

In the case of CPEB4 RRM domains, we observed that both domains contribute to the interaction with the consensus eight nucleotide CPE motifs, presumably increasing not only the affinity but also the specificity of the interaction. Given the length of the RNA motif, we assume that the RRM domains form a continuous binding platform. Based on the ^{15}N amide relaxation properties and the residues affected upon RNA binding the CPEB4 RRM1–RRM2 pair behaves as a compact molecule already in the unbound state, suggesting that the pre-formed domain contacts provide an arrangement that might be finely tuned in the presence of RNA.

Our NMR titration data for CPEB4's tandem RRMs show that in RRM1 the most affected residues upon binding lie on its β -sheet, the canonical binding surface of RRMs, containing the conserved RNPs on the β_1 and β_3 strands. RRM2 is the less canonical domain of the pair containing three of six conserved residues of RNP2 mo-

tif (I-F-V) and completely lacking the RNP1 motif on the β_3 strand. Moreover, it seems to interact with the CPE in a non-canonical manner, since we detect chemical shift changes also at positively charged residues located in its C-terminus. As reported previously, the β -sheet surface can bind up to four nucleotides. Therefore, the recognition of a longer RNA requires more than one RRM to create a sufficiently large binding platform. The two consensus CPE sequences contain each seven nucleotides (UUU UAA U and UUU UAU U), thus it seems plausible that a motif of this length binds more than one RRM. Moreover, the use of regions other than the β -sheet binding platform has been demonstrated in other tandem RRM structures. For example, the RRMs in the protein nucleolin use the linker and a loop between the β -sheets to interact with a stem loop RNA structure (33). The seemingly different modes of RNA interaction of CPEB's RRMs, with RRM2 using C-terminal residues, suggest an asymmetrical orientation of the domains relative to each other.

In an attempt to visualize all the results obtained from this study, we built a homology model containing the two RRM domains and an octameric polyU stretch to illustrate

the relative size of a single-stranded octameric RNA compared to the RRM domains (Supplementary Figure S7). We believe that this preliminary model based on the CS changes obtained for the amide resonances of the RRM1–RRM2 sample in the presence of RNA provides an idea about the relative orientation of the two domains. Further investigation of the RRM's CPE interaction and their relative orientation to each other is required using structural techniques such as NMR or crystallography.

Our results demonstrate that CPEB4 RRM1 and RRM1–RRM2 can sample monomer and dimer conformations under the experimental conditions we have assayed. However, the inability to produce a soluble RRM2 fragment did not allow us to identify whether both domains are involved in the self-association. The dimeric species are present in low abundance, precluding its characterization using NMR spectroscopy. However, the potential to form dimers has been observed for other RRM domains. For example, the U1A protein forms a homodimer via its RRM domain. The dimer interface is separate from the polyadenylation inhibition element (PIE) RNA interaction interface (34). IM-MS assays have unambiguously detected both monomeric as well as dimeric 1:1 protein–RNA complexes, indicating that both the monomeric and dimeric states of RRM1–RRM2 possess RNA-binding activity. This suggests that dimerization is not required for RNA recognition. Previously, dimerization has been reported for another member of the CPEB-family CPEB1 (19). Lin *et al.* (2012) postulated that the dimerization functions to inactivate spare proteins, preventing them from inducing the polyadenylation of RNAs with low affinity binding sites and also serving as molecular hubs that release polyadenylation factors upon dimer destruction. The same study identified both the RRMs and the ZZ-domain to be essential for dimerization. Interestingly, and in contrast to our experiments, the results suggest that CPEB1 dimerization occurs at the expense of RNA binding. However, whereas in our study, isolated RRM domains were characterized, Lin *et al.* (2012) assessed full-length CPEB1 and two dimeric constructs, two full-length CPEB1 proteins separated by a linker sequence of 40 amino acids and full-length CPEB1 with an N-terminal coiled-coil dimerization domain. The sequence identity of CPEB1 and CPEB4 RRM domains is 45% and the overall identity is 22%, although it is possible that the sequence disparity between the two protein RRM domains may account for differences in the mode of the RNA interaction and self-association. We suggest, it is more likely, that the reason for the loss of binding activity in the full-length CPEB1 dimer is due to the spatial arrangement of the N-terminal unstructured regions and the C-terminal ZZ-domains, which renders the RNA-binding interface inaccessible. Our IM-MS data clearly show that the dimeric form of the RRM pair maintains their RNA-binding activity. We therefore conclude that the dimerization surface is separate from the identified RNA-binding interface.

Experiments with the full-length CPEB4, as investigated with CPEB1, would ascertain if the inhibition of RNA binding upon dimerization is exclusive of CPEB1 or is present in other family members. Further investigation into the function of the potential formation of CPEB oligomers is required since it may reveal the critical differences be-

tween the four functional CPEB paralogs of vertebrates. It would be interesting to see if the postulated molecular hubs are detectable using the available cell biology techniques or could indeed be purified for *in vitro* analysis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

We thank the Mass Spectrometry Core Facility at IRB Barcelona (Dr M. Vilaseca) for support with the IM-MS experiments as well as the NMR Facility of FMP Berlin (Dr H. Oschkinat and Dr P. Schmieder) for NMR measurement time and support during the acquisition of the experiments.

FUNDING

Ministry of Science and Innovation [Programme Consolidator RNAREG CSD2009-00080 to M.J.M. and E.P., CTQ2010-21567-C02-01 to E.P., SAF2011-25119 to M.J.M.]. C.S. is a recipient of a 'la Caixa'/IRB Barcelona International Ph.D. Programme fellowship. M.J.M. is an ICREA Programme Investigator. Source of open access funding: Publication charges are funded by the Programme Consolidator RNAREG CSD2009-00080 (to M.J.M.) grant. *Conflict of interest statement.* None declared.

REFERENCES

- Richter, J.D. (2007) CPEB: a life in translation. *Trends Biochem. Sci.*, **32**, 279–285.
- Kim, J.H. and Richter, J.D. (2006) Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Mol. Cell*, **24**, 173–183.
- Stebbins-Boaz, B., Cao, Q., De Moor, C.H., Mendez, R. and Richter, J.D. (1999) Maskin is a CPEB-associated factor that transiently interacts with eIF-4E. *Mol. Cell*, **4**, 1017–1027.
- Cao, Q., Kim, J.H. and Richter, J.D. (2006) CDK1 and calcineurin regulate Maskin association with eIF4E and translational control of cell cycle progression. *Nat. Struct. Mol. Biol.*, **13**, 1128–1134.
- Mendez, R., Hake, L.E., Andresson, T., Littlepage, L.E., Ruderman, J.V. and Richter, J.D. (2000) Phosphorylation of CPE binding factor by Eg2 regulates translation of c-mos mRNA. *Nature*, **404**, 302–307.
- Barnard, D.C., Ryan, K., Manley, J.L. and Richter, J.D. (2004) Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell*, **119**, 641–651.
- Wang, X.-P. and Cooper, N.G.F. (2010) Comparative in silico analyses of cpeb1–4 with functional predictions. *Bioinform. Biol. Insights*, **4**, 61–83.
- Villalba, A., Coll, O. and Gebauer, F. (2011) Cytoplasmic polyadenylation and translational control. *Curr. Opin. Genet. Dev.*, **21**, 452–457.
- Hake, L.E., Mendez, R. and Richter, J.D. (1998) Specificity of RNA binding by CPEB: requirement for RNA recognition motifs and a novel zinc finger. *Mol. Cell Biol.*, **18**, 685–693.
- Merkel, D.J., Wells, S.B., Hilburn, B.C., Elazzouzi, F., Pérez-Alvarado, G.C. and Lee, B.M. (2013) The C-terminal region of cytoplasmic polyadenylation element binding protein is a ZZ domain with potential for protein–protein interactions. *J. Mol. Biol.*, **425**, 2015–2026.
- Fox, C.A., Sheets, M.D. and Wickens, M.P. (1989) Poly(A) addition during maturation of frog oocytes: distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUU. *Genes Dev.*, **3**, 2151–2162.

12. McGrew, L.L., Dworkin-Rastl, E., Dworkin, M.B. and Richter, J.D. (1989) Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.*, **3**, 803–815.
13. Novoa, I., Gallego, J., Ferreira, P.G. and Mendez, R. (2010) Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nat. Cell Biol.*, **12**, 447–456.
14. Igea, A. and Méndez, R. (2010) Meiosis requires a translational positive loop where CPEB1 ensures its replacement by CPEB4. *EMBO J.*, **29**, 2182–2193.
15. Huang, Y.-S., Kan, M.-C., Lin, C.-L. and Richter, J.D. (2006) CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. *EMBO J.*, **25**, 4865–4876.
16. Ortiz-Zapater, E., Pineda, D., Martínez-Bosch, N., Fernández-Miranda, G., Iglesias, M., Alameda, F., Moreno, M., Elisovich, C., Eyraes, E., Real, F.X. *et al.* (2012) Key contribution of CPEB4-mediated translational control to cancer progression. *Nat. Med.*, **18**, 83–90.
17. Pavlopoulos, E., Trifilieff, P., Chevaleyre, V., Fioriti, L., Zairis, S., Pagano, A., Malleret, G. and Kandel, E.R. (2011) Neuralized1 activates CPEB3: a function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. *Cell*, **147**, 1369–1383.
18. Nagai, K., Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
19. Lin, C.-L., Huang, Y.-T. and Richter, J.D. (2012) Transient CPEB dimerization and translational control. *RNA*, **18**, 1050–1061.
20. Marley, J., Lu, M. and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR*, **20**, 71–75.
21. Bellon, L. (2001) Oligoribonucleotides with 2'-O-(tert-Butyldimethylsilyl) groups. In: *Current Protocols in Nucleic Acid Chemistry*. John Wiley & Sons, Inc., HOBOKEN, NJ.
22. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
23. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
24. Bartels, C., Xia, T.H., Billeter, M., Güntert, P. and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR*, **6**, 1–10.
25. Barbato, G., Ikura, M., Kay, L.E., Pastor, R.W. and Bax, A. (1992) Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry (Mosc.)*, **31**, 5269–5278.
26. Sánchez-Hernández, N., Ruiz, L., Sánchez-Álvarez, M., Montes, M., Macías, M.J., Hernández-Munain, C. and Suñé, C. (2012) The FF4 and FF5 domains of transcription elongation regulator 1 (TCERG1) target proteins to the periphery of speckles. *J. Biol. Chem.*, **287**, 17789–17800.
27. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry (Mosc.)*, **33**, 5984–6003.
28. Daragan, V.A. and Mayo, K.H. (1997) Motional model analyses of protein and peptide dynamics using 13C and 15N NMR relaxation. *Prog. Nucl. Magn. Reson. Spectrosc.*, **31**, 63–105.
29. Hoaglund, C.S., Valentine, S.J., Sporleder, C.R., Reilly, J.P. and Clemmer, D.E. (1998) Three-dimensional ion mobility/TOFMS analysis of electrosprayed biomolecules. *Anal. Chem.*, **70**, 2236–2242.
30. Aragón, E., Goerner, N., Xi, Q., Gomes, T., Gao, S., Massagué, J. and Macías, M.J. (2012) Structural basis for the versatile interactions of smad7 with regulator WW domains in TGF- β pathways. *Structure*, **20**, 1726–1736.
31. Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcárcel, J. and Sattler, M. (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature*, **475**, 408–411.
32. Barraud, P. and Allain, F.H.-T. (2013) Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRM influences the nucleic acid binding topology. *J. Biomol. NMR*, **55**, 119–138.
33. Allain, F.H., Bouvet, P., Dieckmann, T. and Feigon, J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.*, **19**, 6870–6881.
34. Varani, L., Gunderson, S.I., Mattaj, I.W., Kay, L.E., Neuhaus, D. and Varani, G. (2000) The NMR structure of the 38 kDa U1A protein-PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat. Struct. Biol.*, **7**, 329–335.

Curriculum Vitae

Constanze Schelhorn Passatge de Vilaret 28,2-2
08013 Barcelona, Spain
constanze.schelhorn@gmail.com

Personal data

Date of birth 10.12.1984
Place of birth Coburg
Nationality German

Education

09/2010 – present **IRB - Institute for Research in Biomedicine,
Barcelona, Spain
Structural & Computational Biology Programme,
Structural characterization of macromolecular assemblies
(Dr. Maria J. Macias)
PhD Student**

05/2010 – 07/2010 **University of Würzburg, Dept. of Experimental Physics 5,
Biophysics
MRB-Magnetic Resonance Bavaria:
Research Center for Life & Material Sciences
Scientific employee**

05/2009 – 03/2010 **University of Würzburg, Dept. of Experimental Physics 5,
Biophysics, Würzburg, Germany
MRB-Magnetic Resonance Bavaria:
Research Center for Life & Material Sciences
Diploma student
Diploma Thesis: Spectroscopic NMR - Flow Measurements using
ac – gradients; Pulse sequence development for both spectroscopic
and imaging methods**

10/2004 – 03/2010 **University of Würzburg, Faculty of Physics & Astronomy
Student in the *Diplom*- degree programm of physics**

- 10/2007 – 03/2008 **University of Padova, Faculty of Physics, Italy**
Exchange Semester in the course of the Bilateral Exchange Programme of the University of Padova and the University of Würzburg
- 09/1995 – 07/2004 **Grammar School, Gymnasium Alexandrinum, Coburg, Germany**
General Qualification for university entrance *Abitur* 2004

Scholarships

- 10/2007 – 03/2008 **University of Padova, Padova, Italy**
Received a merit based full ride scholarship for the period of studies
- 09/2010 – present **“La Caixa“/IRB Barcelona International PhD Programme Fellowship**

Languages

German:	native language
English:	fluent
Spanish:	fluent
Italian:	good
French:	basic knowledge

Conferences & Courses attended

- EMBO Practical Course: Structure, Dynamics and function of biomacromolecules by solution NMR at the Bavarian NMR Center at the TU Munich in Garching, Germany; Jul 29 - Aug 5, 2011; poster presentation
- CCPN Europa 2011, held at ITQB Oeiras, Portugal; Sep 28 - 30, 2011; poster presentation
- BioMed Conference Macromolecular Dynamics, Barcelona; Oct 24 - 26, 2011
- 2nd IRB Barcelona PhD Student Symposium: Life in Motion: Dynamics of Molecules&Systems, Barcelona, Spain; Nov 17 - 18, 2011; poster presentation
- The BCN Biosolids NMR and DNP Symposium, Barcelona; March 14, 2012
- III IMPPC Annual conference: RNA Biology in Cancer and other Diseases, Barcelona; May 3 - 4, 2012; poster presentation
- EUROMAR 2012 Magnetic Resonance Conference, University College Dublin, Ireland: July 1-5, 2012; poster presentation

- XXVth ICMRBS - International Conference on Magnetic Resonance in Biological Systems, Lyon, France; August 19 - 24, 2012; poster presentation
- EMBO Practical Course on Computational Structural Biology - from data to structure to function, EMBL-EBI, Hinxton, United Kingdom; April 17 - 11, 2014

Other

Heading the organization committee of the **3rd IRB Barcelona PhD Student Symposium: “The Clock of Life “**, held at Casa Milà–La Pedrera, Barcelona, Spain Nov 14 -15

Publications

Schelhorn C., Jakob P.M., Fidler F. : **Rapid spectroscopic velocity quantification using periodically oscillating gradients.** *J. Magn. Reson.* 214(1), 175-183, 2012

Schelhorn C., Gordon J.M.B., Ruiz L., Alguacil J., Pedroso E., Macias M.J. : **RNA recognition and self-association of CPEB4 is mediated by its tandem RRM domains.** *Nucl. Acids Res.* 42(5), 10185-10195, 2014

Schelhorn C., Martin P., Suñol D., Macias M.J. : **Characterisation of Pin1-CPEB1 interaction and its putative role in CPEB1 degradation.** *In preparation*

Bibliography

- [1] Fátima Gebauer, Thomas Preiss, and Matthias W. Hentze. From Cis-Regulatory Elements to Complex RNPs and Back. *Cold Spring Harb Perspect Biol*, 4(7):a012245, 2012. PMID: 22751153.
- [2] Raul Mendez and Joel D. Richter. Translational control by CPEB: a means to the end. *Nat Rev Mol Cell Biol*, 2(7):521–529, 2001.
- [3] Helois E. Radford, Hedda A. Meijer, and Cornelia H. de Moor. Translational control by cytoplasmic polyadenylation in *Xenopus* oocytes. *Biochim Biophys Acta*, 1779(4):217–229, 2008. PMID: 18316045 PMCID: PMC2323027.
- [4] Gonzalo Fernández-Miranda and Raúl Méndez. The CPEB-family of proteins, translational control in senescence and cancer. *Ageing Research Reviews*, 11(4):460–472, 2012.
- [5] Alan Sachs. 10 Physical and Functional Interactions between the mRNA Cap Structure and the Poly(A) Tail. *Cold Spring Harb. Monogr. Arch.*, 39(0), 2000.
- [6] Marina Chekulaeva, Hansruedi Mathys, Jakob T. Zipprich, Jan Attig, Marija Colic, Roy Parker, and Witold Filipowicz. miRNA repression involves GW182-mediated recruitment of CCR4–NOT through conserved W-containing motifs. *Nat Struct Mol Biol*, 18(11):1218–1226, 2011.
- [7] Barsanjit Mazumder, Vasudevan Seshadri, Hiroaki Imataka, Nahum Sonenberg, and Paul L. Fox. Translational Silencing of Ceruloplasmin Requires the Essential Elements of mRNA Circularization: Poly(A) Tail, Poly(A)-Binding Protein, and Eukaryotic Translation Initiation Factor 4G. *Mol. Cell. Biol.*, 21(19):6440–6449, 2001. PMID: 11533233.
- [8] Marc R. Fabian, Maja K. Cieplak, Filipp Frank, Masahiro Morita, Jonathan Green, Tharan Srikumar, Bhushan Nagar, Tadashi Yamamoto, Brian Raught, Thomas F. Duchaine, and Nahum Sonenberg. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. *Nat Struct Mol Biol*, 18(11):1211–1217, 2011.
- [9] Iren Wang, Janosch Hennig, Pravin Kumar Ankush Jagtap, Miriam Sonntag, Juan Valcárcel, and Michael Sattler. Structure, dynamics and RNA binding of

- the multi-domain splicing factor TIA-1. *Nucl. Acids Res.*, page gku193, 2014. PMID: 24682828.
- [10] Motoaki Wakiyama, Koji Takimoto, Osamu Ohara, and Shigeyuki Yokoyama. Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev.*, 21(15):1857–1862, 2007. PMID: 17671087.
- [11] Wang, Nigel Cooper, and Nigel Cooper. Comparative in Silico Analyses of Cpeb1–4 with Functional Predictions. *Bioinforma. Biol. Insights*, page 61, 2010.
- [12] Ana Villalba, Olga Coll, and Fátima Gebauer. Cytoplasmic polyadenylation and translational control. *Current Opinion in Genetics & Development*, 21(4):452–457, 2011.
- [13] Laura E. Hake, Raul Mendez, and Joel D. Richter. Specificity of RNA Binding by CPEB: Requirement for RNA Recognition Motifs and a Novel Zinc Finger. *Mol. Cell. Biol.*, 18(2):685–693, 1998. PMID: 9447964.
- [14] Daniel J. Merkel, Sarah B. Wells, Bryce C. Hilburn, Fatima Elazzouzi, Gabriela C. Pérez-Alvarado, and Brian M. Lee. The C-terminal region of cytoplasmic polyadenylation element binding protein is a ZZ domain with potential for protein-protein interactions. *J. Mol. Biol.*, 425(11):2015–2026, 2013. PMID: 23500490.
- [15] R. Mendez, L. E. Hake, T. Andresson, L. E. Littlepage, J. V. Ruderman, and J. D. Richter. Phosphorylation of CPE binding factor by Eg2 regulates translation of c-mos mRNA. *Nature*, 404(6775):302–307, 2000. PMID: 10749216.
- [16] Raul Mendez, Daron Barnard, and Joel D. Richter. Differential mRNA translation and meiotic progression require Cdc2-mediated CPEB destruction. *EMBO J*, 21(7):1833–1844, 2002.
- [17] Daiki Setoyama, Masakane Yamashita, and Noriyuki Sagata. Mechanism of degradation of CPEB during *Xenopus* oocyte maturation. *PNAS*, 104(46):18001–18006, 2007. PMID: 17986610.
- [18] M. Rechsteiner and S. W. Rogers. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.*, 21(7):267–271, 1996. PMID: 8755249.
- [19] Ana Igea and Raúl Méndez. Meiosis requires a translational positive loop where CPEB1 ensues its replacement by CPEB4. *EMBO J*, 29(13):2182–2193, 2010.
- [20] Isabel Novoa, Javier Gallego, Pedro G. Ferreira, and Raul Mendez. Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nat Cell Biol*, 12(5):447–456, 2010.

- [21] C. A. Fox, M. D. Sheets, and M. P. Wickens. Poly(A) addition during maturation of frog oocytes: distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. *Genes Dev.*, 3(12B):2151–2162, 1989. PMID: 2628165.
- [22] L. L. McGrew, E. Dworkin-Rastl, M. B. Dworkin, and J. D. Richter. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.*, 3(6):803–815, 1989. PMID: 2568313.
- [23] Jong Heon Kim and Joel D. Richter. Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Mol. Cell*, 24(2):173–183, 2006. PMID: 17052452.
- [24] Jong Heon Kim and Joel D. Richter. RINGO/cdk1 and CPEB mediate poly(A) tail stabilization and translational regulation by ePAB. *Genes Dev.*, 21(20):2571–2579, 2007. PMID: 17938241 PMCID: PMC2000322.
- [25] Daron C. Barnard, Kevin Ryan, James L. Manley, and Joel D. Richter. Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell*, 119(5):641–651, 2004. PMID: 15550246.
- [26] B. Stebbins-Boaz, Q. Cao, C. H. de Moor, R. Mendez, and J. D. Richter. Maskin is a CPEB-associated factor that transiently interacts with eIF-4E. *Mol. Cell*, 4(6):1017–1027, 1999. PMID: 10635326.
- [27] Quiping Cao, Jong Heon Kim, and Joel D. Richter. CDK1 and calcineurin regulate Maskin association with eIF4E and translational control of cell cycle progression. *Nat. Struct. Mol. Biol.*, 13(12):1128–1134, 2006. PMID: 17086181.
- [28] Joyce Tay, Rebecca Hodgman, Madathia Sarkissian, and Joel D. Richter. Regulated CPEB phosphorylation during meiotic progression suggests a mechanism for temporal control of maternal mRNA translation. *Genes Dev*, 17(12):1457–1462, 2003. PMID: 12815066 PMCID: PMC196075.
- [29] G. Dreyfuss, M. S. Swanson, and S. Piñol-Roma. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.*, 13(3):86–91, 1988. PMID: 3072706.
- [30] Antoine Cléry, Markus Blatter, and Frédéric H-T Allain. RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, 18(3):290–298, 2008.
- [31] K. Nagai, C. Oubridge, T. H. Jessen, J. Li, and P. R. Evans. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, 348(6301):515–520, 1990. PMID: 2147232.

- [32] Cameron D Mackereth and Michael Sattler. Dynamics in multi-domain protein recognition of RNA. *Current Opinion in Structural Biology*, 22(3):287–296, 2012.
- [33] Pierre Barraud and Frédéric H.-T. Allain. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRM s influences the nucleic acid binding topology. *J Biomol NMR*, 55(1):119–138, 2013.
- [34] Kun Ping Lu, Steven D. Hanes, and Tony Hunter. A human peptidyl–prolyl isomerase essential for regulation of mitosis. *Nature*, 380(6574):544–547, 1996.
- [35] Yusuke Kato, Mie Ito, Kunji Kawai, Koji Nagata, and Masaru Tanokura. Determinants of Ligand Specificity in Groups I and IV WW Domains as Studied by Surface Plasmon Resonance and Model Building. *J. Biol. Chem.*, 277(12):10173–10177, 2002. PMID: 11751914.
- [36] Marius Sudol and Tony Hunter. NeW Wrinkles for an Old Domain. *Cell*, 103(7):1001–1004, 2000.
- [37] R. Bruce Martin. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers*, 45(5):351–353, 1998.
- [38] R. L. Stein. Mechanism of enzymatic and nonenzymatic prolyl cis-trans isomerization. *Adv. Protein Chem.*, 44:1–24, 1993. PMID: 8317295.
- [39] Gunter Fischer. Peptidyl-Prolyl cis/trans Isomerases and Their Effectors. *Angew. Chem. Int. Ed. Engl.*, 33(14):1415–1436, 1994.
- [40] G. Lippens, I. Landrieu, and C. Smet. Molecular mechanisms of the phospho-dependent prolyl cis/trans isomerase Pin1. *FEBS J.*, 274(20):5211–5222, 2007.
- [41] Kun Ping Lu and Xiao Zhen Zhou. The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat Rev Mol Cell Biol*, 8(11):904–916, 2007.
- [42] Gerburg Wulf, Greg Finn, Futoshi Suizu, and Kun Ping Lu. Phosphorylation-specific prolyl isomerization: is there an underlying theme? *Nat. Cell Biol.*, 7(5):435–441, 2005. PMID: 15867923.
- [43] Tae Ho Lee, Lucia Pastorino, and Kun Ping Lu. Peptidyl-prolyl cis–trans isomerase Pin1 in ageing, cancer and Alzheimer disease. *Expert Rev. Mol. Med.*, 13:null–null, 2011.
- [44] Jane A. Driver and Kun Ping Lu. Pin1: A New Genetic Link between Alzheimers Disease, Cancer and Aging. *Curr. Aging Sci.*, 3(3):158–165, 2010.

- [45] Martin Theuerkorn, Gunter Fischer, and Cordelia Schiene-Fischer. Prolyl cis/trans isomerase signalling pathways in cancer. *Current Opinion in Pharmacology*, 11(4):281–287, 2011.
- [46] Yih-Cherng Liou, Xiao Zhen Zhou, and Kun Ping Lu. Prolyl isomerase Pin1 as a molecular switch to determine the fate of phosphoproteins. *Trends in Biochemical Sciences*, 36(10):501–514, 2011.
- [47] K. L. Rock, C. Gramm, L. Rothstein, K. Clark, R. Stein, L. Dick, D. Hwang, and A. L. Goldberg. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell*, 78(5):761–771, 1994. PMID: 8087844.
- [48] C. M. Pickart. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.*, 70:503–533, 2001. PMID: 11395416.
- [49] Rama Ranganathan, Kun Ping Lu, Tony Hunter, and Joseph P. Noel. Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition Is Phosphorylation Dependent. *Cell*, 89(6):875–886, 1997.
- [50] Elena Bayer, Sandra Goettsch, Jonathan W. Mueller, Bernhard Griewel, Elena Guiberman, Lorenz M. Mayr, and Peter Bayer. Structural Analysis of the Mitotic Regulator hPin1 in Solution INSIGHTS INTO DOMAIN ARCHITECTURE AND SUBSTRATE BINDING. *J. Biol. Chem.*, 278(28):26183–26193, 2003. PMID: 12721297.
- [51] Mark A. Verdecia, Marianne E. Bowman, Kun Ping Lu, Tony Hunter, and Joseph P. Noel. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat Struct Mol Biol*, 7(8):639–643, 2000.
- [52] R. Wintjens, J. M. Wieruszkeski, H. Drobecq, P. Rousselot-Pailley, L. Buée, G. Lippens, and I. Landrieu. 1H NMR study on the binding of Pin1 Trp-Trp domain with phosphothreonine peptides. *J. Biol. Chem.*, 276(27):25150–25156, 2001. PMID: 11313338.
- [53] Morris Nechama, Chien-Ling Lin, and Joel D. Richter. An Unusual Two-Step Control of CPEB Destruction by Pin1. *Mol. Cell. Biol.*, 33(1):48–58, 2013. PMID: 23090969.
- [54] Gustavo J. Gutierrez and Ze’ev Ronai. Ubiquitin and SUMO systems in the regulation of mitotic checkpoints. *Trends in Biochemical Sciences*, 31(6):324–332, 2006.
- [55] Stephen Orlicky, Xiaojing Tang, Andrew Willems, Mike Tyers, and Frank Sicheri. Structural Basis for Phosphodependent Substrate Selection and Orientation by the SCFCdc4 Ubiquitin Ligase. *Cell*, 112(2):243–256, 2003.

- [56] E. Aragon, N. Goerner, A.-I. Zaromytidou, Q. Xi, A. Escobedo, J. Massague, and M. J. Macias. A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes Dev.*, 25(12):1275–1288, 2011.
- [57] Aled M. Edwards, Cheryl H. Arrowsmith, Dinesh Christendat, Akil Dharamsi, James D. Friesen, Jack F. Greenblatt, and Masoud Vedadi. Protein production: feeding the crystallographers and NMR spectroscopists. *Nat Struct Mol Biol*, 7:970–972, 2000.
- [58] J. Marley, M. Lu, and C. Bracken. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR*, 20(1):71–75, 2001. PMID: 11430757.
- [59] R. B. Merrifield. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.*, 85(14):2149–2154, 1963.
- [60] Louis A. Carpino and Grace Y. Han. 9-Fluorenylmethoxycarbonyl amino-protecting group. *J. Org. Chem.*, 37(22):3404–3409, 1972.
- [61] V. K. Sarin, S. B. Kent, J. P. Tam, and R. B. Merrifield. Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. *Anal. Biochem.*, 117(1):147–157, 1981. PMID: 7316187.
- [62] T Vojkovsky. Detection of secondary amines on solid phase. *Pept. Res.*, 8(4):236–237, 1995. PMID: 8527877.
- [63] E. W. McDaniel, D. W. Martin, and W. S. Barnes. Drift Tube Mass Spectrometer for Studies of Low Energy Ion Molecule Reactions. *Rev. Sci. Instrum.*, 33(1):2–7, 1962.
- [64] Steven D. Pringle, Kevin Giles, Jason L. Wildgoose, Jonathan P. Williams, Susan E. Slade, Konstantinos Thalassinou, Robert H. Bateman, Michael T. Bowers, and James H. Scrivens. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry*, 261(1):1–12, 2007.
- [65] T. Wiseman, S. Williston, J. F. Brandts, and L. N. Lin. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.*, 179(1):131–137, 1989. PMID: 2757186.
- [66] M. M. Pierce, C. S. Raman, and B. T. Nall. Isothermal titration calorimetry of protein-protein interactions. *Methods*, 19(2):213–221, 1999. PMID: 10527727.
- [67] I. I. Rabi, J. R. Zacharias, S. Millman, and P. Kusch. A New Method of Measuring Nuclear Magnetic Moment. *Phys. Rev.*, 53(4):318–318, 1938.
- [68] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.*, 69(1-2):37–38, 1946.

- [69] F. Bloch, W. W. Hansen, and M. Packard. The Nuclear Induction Experiment. *Phys. Rev.*, 70(7-8):474–485, 1946.
- [70] Jurgen F. Doreleijers, Alan W. Sousa da Silva, Elmar Krieger, Sander B. Nabuurs, Christian A. E. M. Spronk, Tim J. Stevens, Wim F. Vranken, Gert Vriend, and Geerten W. Vuister. CING: an integrated residue-based structure validation program suite. *J Biomol NMR*, 54(3):267–283, 2012.
- [71] Konstantin Pervushin, Roland Riek, Gerhard Wider, and Kurt Wüthrich. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *PNAS*, 94(23):12366–12371, 1997. PMID: 9356455.
- [72] F Delaglio, S Grzesiek, G W Vuister, G Zhu, J Pfeifer, and A Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, 6(3):277–293, 1995. PMID: 8520220.
- [73] C. Bartels, T. H. Xia, M. Billeter, P. Güntert, and K. Wüthrich. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR*, 6(1):1–10, 1995. PMID: 22911575.
- [74] Wim F. Vranken, Wayne Boucher, Tim J. Stevens, Rasmus H. Fogh, Anne Pajon, Miguel Llinas, Eldon L. Ulrich, John L. Markley, John Ionides, and Ernest D. Laue. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, 59(4):687–696, 2005. PMID: 15815974.
- [75] D. A. Torchia L E Kay. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry (Mosc.)*, 28(23):8972–9, 1989.
- [76] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, 33(19):5984–6003, 1994. PMID: 7514039.
- [77] W. Kabsch. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.*, 26(6):795–800, 1993.
- [78] M. Gordon Joyce, Sergei Radaev, and Peter D. Sun. A rational approach to heavy-atom derivative screening. *Acta Crystallogr D Biol Crystallogr*, 66:358–365, 2010. PMID: 20382988 PMCID: PMC2852299.
- [79] Elan Z. Eisenmesser, Oscar Millet, Wladimir Labeikovskiy, Dmitry M. Korzhnev, Magnus Wolf-Watz, Daryl A. Bosco, Jack J. Skalicky, Lewis E. Kay, and

- Dorothee Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–121, 2005.
- [80] Doris M. Jacobs, Krishna Saxena, Martin Vogtherr, Pau Bernadó, Miquel Pons, and Klaus M. Fiebig. Peptide Binding Induces Large Scale Changes in Inter-domain Mobility in Human Pin1. *J. Biol. Chem.*, 278(28):26174–26182, 2003. PMID: 12686540.
- [81] Andrew T. Namanja, Tao Peng, John S. Zintsmaster, Andrew C. Elson, Maria G. Shakour, and Jeffrey W. Peng. Substrate Recognition Reduces Side-Chain Flexibility for Conserved Hydrophobic Residues in Human Pin1. *Structure*, 15(3):313–327, 2007.
- [82] Wladimir Labeikovsky, Elan Z. Eisenmesser, Daryl A. Bosco, and Dorothee Kern. Structure and Dynamics of Pin1 During Catalysis by NMR. *Journal of Molecular Biology*, 367(5):1370–1381, 2007.
- [83] Maria J. Macias, Silke Wiesner, and Marius Sudol. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Letters*, 513(1):30–37, 2002.
- [84] Cameron D. Mackereth, Tobias Madl, Sophie Bonnal, Bernd Simon, Katia Zanier, Alexander Gasch, Vladimir Rybin, Juan Valcárcel, and Michael Sattler. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature*, 475(7356):408–411, 2011.
- [85] Christophe Maris, Cyril Dominguez, and Frédéric H.-T. Allain. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, 272(9):2118–2131, 2005.
- [86] Frédéric H.-T. Allain, Philippe Bouvet, Thorsten Dieckmann, and Juli Feigon. Molecular basis of sequence specific recognition of preribosomal RNA by nucleolin. *EMBO J.*, 19(24):6870–6881, 2000.
- [87] Tariq Afroz, Lenka Skrisovska, Eulàlia Belloc, Jordina Guillén-Boixet, Raúl Méndez, and Frédéric H.-T. Allain. A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. *Genes Dev.*, 28(13):1498–1514, 2014. PMID: 24990967.
- [88] L. Varani, S. I. Gunderson, I. W. Mattaj, L. E. Kay, D. Neuhaus, and G. Varani. The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat. Struct. Biol.*, 7(4):329–335, 2000. PMID: 10742179.
- [89] Dierk Niessing, Stefan Hüttelmaier, Daniel Zenklusen, Robert H. Singer, and Stephen K. Burley. She2p is a novel RNA binding protein with a basic helical hairpin motif. *Cell*, 119(4):491–502, 2004. PMID: 15537539.

- [90] Marisa Muller, Klaus Richter, Alexander Heuck, Elisabeth Kremmer, Johannes Buchner, Ralf-Peter Jansen, and Dierk Niessing. Formation of She2p tetramers is required for mRNA binding, mRNP assembly, and localization. *RNA*, 15(11):2002–2012, 2009. PMID: 19710186 PMCID: PMC2764475.
- [91] Christine T. DeMaria, Yue Sun, Laura Long, Belinda J. Wagner, and Gary Brewer. Structural Determinants in AUF1 Required for High Affinity Binding to A + U-rich Elements. *J. Biol. Chem.*, 272(44):27635–27643, 1997.
- [92] Chien-Ling Lin, Yen-Tsung Huang, and Joel D. Richter. Transient CPEB dimerization and translational control. *RNA*, 18(5):1050–1061, 2012. PMID: 22456264 PMCID: PMC3334692.
- [93] Kirsten S. Dickson, Sunnie R. Thompson, Nicola K. Gray, and Marvin Wickens. Poly(A) Polymerase and the Regulation of Cytoplasmic Polyadenylation. *J. Biol. Chem.*, 276(45):41810–41816, 2001. PMID: 11551905.
- [94] Laure Lapasset, Bérengère Pradet-Balade, Jean-Claude Lozano, Gérard Peaucellier, and André Picard. Nuclear envelope breakdown may deliver an inhibitor of protein phosphatase 1 which triggers cyclin B translation in starfish oocytes. *Developmental Biology*, 285(1):200–210, 2005.
- [95] George Thom, Nicola Minshall, Anna Git, Joanna Argasinska, and Nancy Standard. Role of cdc2 kinase phosphorylation and conserved N-terminal proteolysis motifs in cytoplasmic polyadenylation-element-binding protein (CPEB) complex dissociation and degradation. *Biochem J*, 370:91–100, 2003. PMID: 12401129 PMCID: PMC1223136.
- [96] Svetlana Uzbekova, Yannick Arlot-Bonnemains, Joëlle Dupont, Rozenn Dalbiès-Tran, Pascal Papillier, Sophie Penner, Aurore Thélie, Christine Perreau, Pascal Mermillod, Claude Prigent, and Rustem Uzbekov. Spatio-temporal expression patterns of aurora kinases a, B, and C and cytoplasmic polyadenylation-element-binding protein in bovine oocytes during meiotic maturation. *Biol. Reprod.*, 78(2):218–233, 2008. PMID: 17687118.
- [97] Donna G. Crenshaw, Jing Yang, Anthony R. Means, and Sally Kornbluth. The mitotic peptidylprolyl isomerase, Pin1, interacts with Cdc25 and Plx1. *EMBO J.*, 17(5):1315–1327, 1998.
- [98] Hilde Abrahamsen, Audrey K. O’Neill, Natarajan Kannan, Nicole Kruse, Susan S. Taylor, Patricia A. Jennings, and Alexandra C. Newton. Peptidyl-prolyl Isomerase Pin1 Controls Down-regulation of Conventional Protein Kinase C Isozymes. *J. Biol. Chem.*, 287(16):13262–13278, 2012. PMID: 22318721.
- [99] Caroline Smet, Anne-Véronique Sambo, Jean-Michel Wieruszeski, Arnaud Leroy, Isabelle Landrieu, Luc Buée, and Guy Lippens. The Peptidyl Prolyl

- cis/trans-Isomerase Pin1 Recognizes the Phospho-Thr212-Pro213 Site on Tau†. *Biochemistry*, 43(7):2032–2040, 2004.
- [100] Yoshinori Kanemori, Katsuhiko Uto, and Noriyuki Sagata. -TrCP recognizes a previously undescribed nonphosphorylated destruction motif in Cdc25A and Cdc25B phosphatases. *PNAS*, 102(18):6279–6284, 2005. PMID: 15845771.
- [101] Andrew T. Namanja, Xiaodong J. Wang, Bailing Xu, Ana Y. Mercedes-Camacho, Kimberly A. Wilson, Felicia A. Etzkorn, and Jeffrey W. Peng. Stereospecific gating of functional motions in Pin1. *PNAS*, 108(30):12289–12294, 2011. PMID: 21746900.
- [102] X. Z. Zhou, P. J. Lu, G. Wulf, and K. P. Lu. Phosphorylation-dependent prolyl isomerization: a novel signaling regulatory mechanism. *Cell. Mol. Life Sci.*, 56(9-10):788–806, 1999. PMID: 11212339.
- [103] Kun Ping Lu, Yih Cherng Liou, and Xiao Zhen Zhou. Pinning down proline-directed phosphorylation signaling. *Trends Cell Biol.*, 12(4):164–172, 2002. PMID: 11978535.

Acknowledgements

It would not have been possible to write this thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

This thesis would not have been possible without the guidance and support of **Dr. Maria J. Macias**, my thesis director. **Maria**, since I started my PhD thesis you have always encouraged me to learn new things and I am extremely grateful for the countless opportunities you have provided me to do so. During the last four years, I have learned lots from you both on an academic and a personal level. Your positivity and courage inspired not only me.

I wish to thank **Prof. Dr. Michael Sattler** for giving me the opportunity to carry out this work as an external PhD student. I appreciate his support and advice as external university supervisor of my thesis.

I am grateful to all of my colleagues for the stimulating and fun environment in the lab. **Lidia**, I wish to thank you for your advice, support and patience especially in my early days in the lab. I am lucky to have had such an excellent teacher for molecular biology. Thanks a lot for all the fun times we spent together outside the lab. **Pau**, I wish to thank you for all your help with setting up NMR experiments and with all kinds of computational questions and problems. I would further like to acknowledge the post-docs of the lab: **Jimmy**, **Toni** and **Eric**. Your valuable advice and being open to discussion also made an essential contribution to the success of my experiments. **Albert**, I am grateful for your willingness to discuss my results and for tips and tricks with relaxation experiments. I wish you all the best for your future after finishing your PhD. Thanks to **Lluc** and **David** for giving me hand with peptide synthesis. **David** thanks for always helping out and infecting those around you with your good mood! I would further like to acknowledge **Tiago** and **Jordi**, who accompanied parts of my PhD time.

I am especially thankful to **Prof. Dr. Enrique Pedroso** for the fruitful collaboration and for giving me the opportunity to learn the basics of RNA synthesis in your lab.

I would further like to thank **Dr. Fernando Gil** from BL13-XALOC beamline at ALBA Synchrotron Light Facility for sharing his knowledge of data collection strategies and several other theoretical and practical aspects of X-ray diffraction.

Special thanks to **Dr. Joan Pous** for being a member of my thesis advisory committee

and his sensible advice and his openness to engaging in discussions about my countless questions and concerns regarding crystallography. Thanks to **Xandra** and **Sonia** for your help with setting up crystallization trials.

I am grateful for the advice and suggestions from **Prof. Dr. Miquel Pons** and **Prof. Dr. Antoni Riera**, who have been members of my thesis advisory committee.

I would like to thank **Dr. Marta Vilaseca** from the Mass Spectrometry Core Facility at IRB Barcelona for support with the IM-MS experiments as well as all the other IRB facilities.

Thanks to the NMR Facility of FMP Berlin, especially **Prof. Dr. Hartmut Oschkinat** and **Dr. Peter Schmieder**, for NMR measurement time and support during the acquisition of the experiments.

I would like to acknowledge the financial support of the Programme **Consolider RNAREG** and 'laCaixa' by rewarding me with the '**laCaixa**'/IRB Barcelona **International Ph.D. Programme fellowship**.

During my time at IRB Barcelona I met many people who contributed to making the time spent at the institute enjoyable. Unfortunately it is only possible to give a particular mention to only some of them here.

I would like to thank **Carles** for the talks, cups of teas and encouragement when experiments didn't succeed yet another time. **Giuseppe**, thank you for your friendship and support. I am happy to see we find ourselves approaching the finishing line together. **Elisa** and **Daria**, I am truly grateful for the encouragement, emotional support and caring you provided throughout my PhD time. I am glad that we have spent this special time together sharing our hopes and worries.

Though it might sound weird, there is a life outside the lab and science. I want to express my gratitude to the people that made me feel at home in this wonderful city: All the guys from the climbing gym, especially all the 'Climbats'- Crew. There's no better thing than climbing after a rough day in the lab. Thanks for the great times spent together after the training, the countless trips we made to all the amazing rock climbing destinations - it definitely helped me not to obsess about all the thesis and paper writing stress. **Laia**, you do not only deserve a special mention for your help with designing the cover of this thesis but also for your encouragement and support. **Saydi**, thanks for your friendship and all the fun times we spent together - I will definitely miss you and our hiking trips when you'll be in Canada. **Guillem**, a special thanks for all the pep talk, the support and the great moments we spent together.

Steffi 'Schatzi', I am truly thankful for your friendship, support, the endless skype-talks and kicks-up-the-backside, when needed.

Philipp, I would like to thank you for accompanying me during my PhD time, for your help, support and encouragement.

Last but not least I want to express my deepest gratitude to my whole family, especially to my parents **Gaby** and **Walter**. They have given me their unequivocal support and love throughout, as always, for which my mere expression of thanks likewise does not suffice.

To them I dedicate this thesis.