Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Nele Merret Hollmann

born in: Bochum (Germany)

Oral examination: 18.11.2020

Structure and dynamics of Upstream of N-Ras and their influence on RNA binding and translation regulation

Referees: Dr. Martin Beck.
Prof. Dr. Frauke Melchior

Abstract

The interplay of multiple RNA binding domains (RBD) in a single RNA binding protein (RBP) to achieve RNA target specificity is far from being understood. In this thesis, a multidisciplinary study of Upstream of N-Ras (Unr) is presented. Unr is an RBP that has been predicted to contain five single-stranded RNA binding cold shock-domains (CSD). The thesis aims to unravel how Unr binds RNA targets specifically in a cellular context.

Several NMR and crystal structures of multidomain constructs were determined. As a result, four non-canonical CSDs in addition to the five previously known canonical CSDs were discovered. These non-canonical CSDs play a scaffolding role between the canonical domains, but do not bind RNA independently. Using NMR relaxation and small-angle X-ray scattering, it could be shown that the linker between most of the canonical and non-canonical domains is rigid, leading to a restricted flexibility of the full-length protein. Different *in vitro* and cellular mutational studies, including a reporter gene assay and a RIP-seq experiment showed that a disruption of the fixed domain arrangement has influence on RNA recognition and the protein function.

Additionally, a crystal structure of a multidomain construct of Unr bound to poly-A RNA provides further information on the complexity of the multidomain RNA binding mechanism of Unr. Several non-canonical binding residues, some even in the non-canonical CSDs, contribute to cooperative RNA binding, suggesting that RNA binding of the full-length protein is likely to be of even higher complexity and plasticity.

Further insights into Unr-RNA binding are provided by a full-length protein model, that describes a restricted flexibility of the protein, which might play an essential role within target specificity. To expand the studies towards Unr-ribonucleotide complexes a quantitative mass spectrometry analysis was conducted, that defined several Unr interactors. The protein-protein interaction with the top hit of this result, namely pAbP, was further characterized, which paves the way towards future structure analysis of a larger translation repressor complex, as it assembles on the 3' UTR of the *msl2* mRNA.

Zusammenfassung

Das Zusammenspiel von mehreren RNA-bindenden Domänen (RBD) in einem einzigen RNA-bindenden Protein (RBP), welches die RNA-Sequenzspezifität erhöht, ist bei Weitem noch nicht verstanden. Im Rahmen dieser Arbeit werden Ergebnisse über das Multidomänen RBP ‚*Upstream of N-Ras*' (Unr) präsentiert, welches fünf RNA-Einzelstrang-bindende ‚*cold-shock*' Domänen (CSD) besitzt. Diese Arbeit soll zum Verständnis beitragen, wie Unr an seine RNA Interaktionspartner spezifisch im zellulären Kontext bindet.

Mehrere Multidomänen NMR- und Kristallstrukturen wurden gelöst, wodurch zusätzlich zu den fünf vorher bekannten CSDs vier weitere CSDs entdeckt wurden. Diese Domänen binden isoliert zwar keine RNA, haben allerdings eine stabilisierende Funktion auf die kanonischen CSDs. Durch NMR Relaxations- und Kleinwinkel-Röntgenstreuungsexperimente konnte gezeigt werden, dass der Linker zwischen den meisten kanonischen und nicht-kanonischen CSDs starr ist, welches zu einer eingeschränkten Beweglichkeit des gesamten Proteins führt. Verschiedene Mutationsstudien in *in vitro* und Zellexperimenten, wie beispielsweise Reportergenassays oder RIP-seq Studien, konnten zeigen, dass ein Aufbrechen dieser starren Anordnung Einfluss auf die RNA Erkennung und die Proteinfunktion hat.

Zusätzlich gibt eine Kristallstruktur eines Multidomänen-Unr-Konstruktes, welches eine poly-A RNA Sequenz bindet, Aufschluss über die Komplexität der Interaktion von Unr mit RNA. Verschiedenste, für die CSD-abhängige RNA Erkennung, nicht-kanonische Aminosäuren, auch in der nicht-kanonischen RNA bindenden Domäne, tragen zu einer kooperativen RNA Bindung bei. Der RNA-binde Mechanismus für diesen Proteinteil weist auf eine hohe Komplexität der RNA Bindung für das gesamte Protein Unr hin.

Ein Modell des gesamten Proteins, das eine beschränkte Beweglichkeit zeigt, die eventuell eine essentielle Rolle in der RNA Spezifität spielt, gibt zusätzliche Einblicke in eine potentielle Protein-RNA Bindung. Durch quantitative Massenspektrometrie wurden verschiedene Unr Proteininteraktionspartner identifiziert. Dies kann weiterhin verwendet werden um größere Unr-Ribonukleotidkomplexe zu charakterisieren. Die Protein-Protein Interaktion zwischen Unr und dem am meisten angereicherten Protein pAbP wurde weitergehend analysiert, was für zukünftige Strukturanalysen eines größeren Translationsrepressorkomplexes, wie er sich in der 3'UTR der *msl2* mRNA bildet, hilfreich sein.

Publications

[1] Nele Merret Hollmann, Pravin Kumar Ankush Jagtap, Pawel Masiewicz, Tanit Guitart, Bernd Simon, Jan Provaznik, Frank Stein, Per Haberkant, Lara Jayne Sweetapple, Laura Villacorte, Dylan Mooijman, Vladimir Benes, Mikhail M. Savitski, Fátima Gebauer and Janosch Hennig - "Pseudo-RNA-binding domains mediate RNA structure specificity in Upstream of N-Ras". Cell Reports, 2020 July 21, Volume: 32, Issue: 3, Article Number: 107930, doi: 10.1016/j.celrep.2020.107930.

[2] Ignacio L. Ibarra, Nele Merret Hollmann, Bernd Klaus, Sandra Augsten, Britta Velten, Janosch Hennig and Judith B. Zaugg - "Prediction of transcription factor cooperativity and its impact on protein-phenotype interactions". Nat Commun, 2020 Jan 8, Volume: 11, Issue: 1, Article Number: 124, doi: 10.1038/s41467-019-13888-7.

[3] Pravin Kumar Ankush Jagtap, Marisa Müller, Pawel Masiewicz, Sören von Bülow, Nele Merret Hollmann, Bernd Simon, Andreas W. Thomae, Peter B Becker and Janosch Hennig - "Structure, dynamics and roX2-lncRNA binding of tandem double-stranded RNA binding domains dsRBD1 and 2 of Drosophila Maleless". Nucleic Acids Res. 2019 May 7, Volume: 47, Issue: 8, Pages: 4319–4333, doi: 10.1093/nar/gkz125.

Contents

List of Figures

List of Tables

List of Abbreviations

Cryo-EM	cryo-electron microscopy
CSD	cold-shock domain
CSP	chemical shift perturbation
eIF	eukaryotic initiation factor
EMSA	electrophoretic mobility shift assay
HMQC	heteronuclear multiple quantum correlation
HSQC	heteronuclear single quantum coherence
ID	interdomain mutant
ILV	isoleucine, leucine and valine
IRES	internal ribosome entry site
K_D	dissociation constant
lncRNA	long non-coding RNA
MAD	multi-wavelength anomalous diffraction
Mle	maleless
Mof	male absent on first
MR	molecular replacement
mRNA	messenger RNA
Msl	male specific lethal
ncCSD	non-canonical cold-shock domain
NMR	Nuclear magnetic resonance spectroscopy
NOE	nuclear Overhauser effect
ORF	open reading frame
PABC	Poly(A)-binding protein C-terminal (domain)
pAbP	poly(A)-binding protein
RBD	RNA binding domain
RBP	RNA binding protein
RIP-seq	RNA immunoprecipitation coupled with next generation sequencing
RNA	ribonucleic acid
RNP complex	ribonucleoprotein complex
roX	RNA on X
RRM	RNA recognition motif
SAXS	small angle X-ray scattering
SL	stem-loop
ssRNA	single stranded RNA
Sxl	Sex-lethal
Unr	Upstream of N-Ras
UTR	untranslated region
WT	wild type

1. Introduction

1.1. RNA binding proteins

In 1868 Friedrich Miescher described nucleic acids for the first time naming them ‘nuclein’ based on their cellular localization[1]. Using chemical and biochemical studies differences between DNA and RNA were elucidated in the following decades. However, it took almost 70 more years until the role of RNAs was hypothesized to be linked to protein biogenesis[2] and again 20 more years until the central dogma of biology was formulated in 1958[3]. This together with the discovery of the ribosome in the mid 1950s[4] paved the way for new biological research areas. Additionally, with the discovery of tRNAs in 1965[5,6], the importance of non-coding RNAs for proper cell function showed that RNA is a highly versatile molecule, carrying out many different molecular functions[7].

Up to date it is hypothesized that almost 10% of the human genome codes for RNA binding proteins (RBPs)[8], making them interact with coding and non-coding RNAs as constitutive partners in ribonucleoprotein (RNP) complexes. However, structural and mechanistical knowledge about RNP assemblies is mainly restricted to large molecular machines, like the ribosome[9,10], the RNA polymerase[11–13] and the spliceosome[14–16]. These machines are often highly abundant in cells and the interaction with their targets is mostly constant and tight, which is advantageous for mechanistic studies. Nevertheless, most of the RBDs function as regulatory units within the cell, meaning their abundance is fluctuating and interaction with binding partners is weaker and versatile. Thereby these proteins are able to respond to environmental changes quickly. However, it also makes research very difficult and explains why there are still proteins, of which RNA binding properties remain unknown[8].

The dynamic nature of RBPs explains their involvement in many regulatory pathways of the cell, including gene expression at all levels ranging from transcription, splicing, polyadenylation, localization, stabilization and degradation to protein synthesis via their diverse roles in translation[7,8,17,18]. Consequently a dysregulation or mutations of RBPs can impact cell viability or function and has been linked to many diseases, such as cancer[19] and neurodegeneration[20].

1.2. Binding modes of multidomain RBPs

To ensure a specific regulation through RBPs in all the different cellular processes, a certain RNA target specificity of the protein is a prerequisite. RBPs employ RNA binding domains (RBD) to engage their target RNAs. RBDs are often small and very conserved domains, with specificities towards single stranded (ssRNA) or double stranded RNA (dsRNA).

One of the most abundant and best characterized RBDs in higher vertebrates is the RNA-recognition motif (RRM), which is binding to ssRNA[21]. Besides the canonical domains also several subclasses like the quasi-RRM (qRRM), the pseudo-RRM (ΨRRM) or U2AF Homology Motifs (UHM) were identified[22]. Further single stranded RBDs are for example zinc finger domains, hnRNP K homology (KH) domains or cold-shock domains (CSD). The most studied dsRNA binding domain is the dsRBD, which is recognizing minor and major groove elements on the target RNA. Besides these classical domains, also non-classical RBDs, like the NHL domain, exist. The increasing number of identified RBPs within the last years, is mostly due to the identification of these non-classical RBDs[8].

Although RBDs are the main drivers of protein-RNA interaction, the single domains are often not enough to discriminate target from non-target RNAs within the complex transcriptome of the cell. Most classical RBDs are around 10 kDa of size and can accommodate three to five contiguous RNA bases specifically, which is by far not enough to endow RNA target recognition. Thereby, a composition of multiple RBDs within one protein is likely to be required to endow the specificity[23–25]. The majority of RBPs is composed of multiple RBDs, either of the same or of different domain types. This results in a large combinatorial variety of different domain classes and the diversity of architectures influences the binding mode to the specific target RNA sequence. Often a flexible linker, which sometimes rigidifies upon binding, separates the different domains[17,26], leading to an increase of target specificity (Figure 1a).

An increased specificity is also mediated by RNP complexes, where two or multiple RBPs, each with its own selectivity, come together and bind synergistically or even cooperatively. This is exemplified by Sxl and Unr, both being involved in the translation repression of *msl2* mRNA in female flies. Unr itself binds to the 3' untranslated region (UTR) of the mRNA with a relatively low binding affinity of 10 μM. However, in the presence of Sxl the binding affinity increases to the low nanomolar range, indicating the strong cooperative effect of Sxl on the binding affinity and thereby also specificity of Unr to this target (Figure 1b)[27].

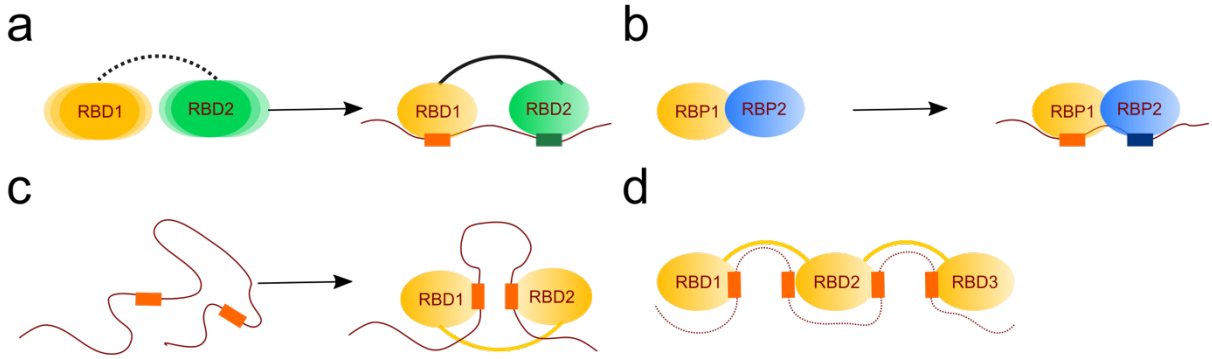


Figure 1: RNA binding proteins mediate target specificity through different types of interaction. **a**: cooperative binding of two different RBDs within one protein can increase target specificity. **b**: protein-protein interactions between two different RBPs can increase the target specificity. **c**: Two RBDs that are connected by a non-flexible linker can force loop formations or other secondary and tertiary structure elements on target RNAs. **d**: multiple RBDs recognize a target RNA through addition of single RBDs specificity.

Multiple RBDs within one protein can also mediate the formation of secondary loop elements within the target RNA. Here a non-flexible linker region between the binding motifs forces a bending of flexible RNA strands between each sequence recognition motif (Figure 1c)[28–34]. Further multiple RBDs may contribute to the RBP binding specificity by adding up to the domain specific engagement of the protein with clustered elements of the target RNAs. It is known that the recognition of target RNAs employs complex binding modes, that can also depend on the target itself, like secondary or tertiary structure elements, presence and distribution of bipartite motifs and the composition of flanking nucleotides (Figure 1d)[35,36]. Another, so far largely neglected option is that flexible linker regions are actually not flexible and force RBDs into fixed orientations and distances, which ultimately induce RNA tertiary structure specificity.

One prerequisite to understand mechanistic details of the different binding modes of RBPs that induce target specificity, is the determination of RNP structures at an atomistic level. Over the years, there have been a number of efforts to examine structural features that dictate RNP binding specificity[37,38]. The knowledge of how single RBDs engage their target sequences increased and in some cases these studies even offered insights into the role of multidomain arrangements in the recognition process[35,39]. *In vitro* and *in vivo* studies that mapped RBP binding landscapes have further highlighted that protein features beyond the boundaries of the RBDs may contribute to RNA binding specificity[40]. However, obstacles in the structural characterization of multidomain RBPs like the difficulties in expression, purification and generation of high-quality samples for crystallization or cryo-electron microscopy (cryo-EM) and the often modular heterogeneity and complicated low affinity interactions make it difficult to assess structures using single techniques. A usage of combinatorial structural techniques, like

the integrative approach that was used to determine the structure of the nuclear pore complex[41-46] will be necessary to gain more insights into the complexity of RBP-RNA interactions.

1.3. The role of RBPs during translation initiation

As a final step in gene expression, translation is very tightly regulated process. Often RBPs with a high target specificity are involved in translation reprogramming. Through direct binding to the target mRNAs, it is possible to rapidly react to changes in the cellular environment[47].

Translation can be divided into three main steps, namely initiation, elongation and termination. The initiation stage, which in eukaryotes mostly relies on the binding of the cap binding complex to the 5' 7-methylguanylate (^{7m}G) cap structure of the mRNA is the rate limiting step during translation regulation[48–52]. The cap binding complex, also called eukaryotic initiation factor 4F (eIF4F) complex, is a heterotrimeric protein complex, which is composed of the DEAD-box RNA helicase eIF4A, the cap-binding protein eIF4E, and the large scaffolding protein eIF4G[53,54]. First, eIF4E interacts with the cap, followed by binding of eIF4G and completed by the addition of eIF4A. Further eIF4G interacts with the poly(A)-binding protein 1 (PABP1), which forms a closed-loop structure between the 5' and 3' end of the mRNA, promoting the circularization of terminal ribosomes from the 3' end back to the 5' (Figure 2a, 2b and 2c)[47,55–60]. Concurrently, the bound eIF4F complex recruits the translation machinery, by interacting with the 43S ribosomal subunit and thereby promotes translation. After recruitment of further components, the recognition of the initiation codon and binding of the large ribosomal subunit, the end of the initiation phase is marked and the elongation begins (Figure 2d)[47].

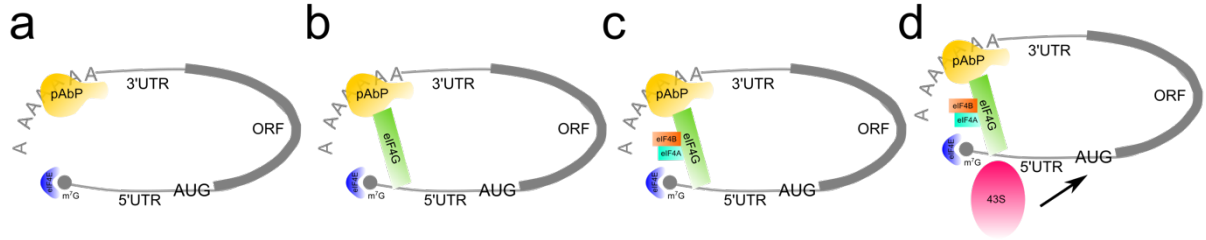


Figure 2: Schematic representation of the different steps of translation initiation.

Certain conditions, like viral infections, hypoxia or mitosis, inhibit the cap dependent translation initiation, by disruption of the eIF4F complex. However, in order to guarantee cell survival a cap independent mechanism evolved in eukaryotes[61,62]. Here, the 43S ribosomal subunit gets directly recruited to the mRNA via internal ribosome entry site (IRES) elements[61-63]. Although these elements do not exhibit a consensus sequence or universal

structure motifs, they share complex structural elements including stem loops and pseudoknots and are mostly located in the 5' UTR directly upstream of the initiation codon[64–66]. So called IRES trans-acting factors (ITAFs), which are RBPs, bind to these elements and are involved in the recruitment of the 43S ribosomal subunit leading to translation initiation. Besides eukaryotes, IRES elements evolved in viruses[67,68].

1.4. Upstream of N-Ras and its role in translation regulation

An example of an RBP with multiple RBDs that is involved in translation regulation is Upstream of N-Ras (Unr), a protein that is studied within this thesis to address the challenge of understanding sequence specificity of modular RBPs towards target RNAs. Unr is a highly conserved protein amongst metazoa and highly abundant in cells of different organisms, being ranked among the top 25% of all proteins of *Drosophila melanogaster* embryos and several human cell lines in the PaxDB[69]. In previous studies human Unr (CSDE1) could be linked to diverse cellular processes including cell migration, differentiation and apoptosis, where it predominantly acts as a cytoplasmic RBP, regulating translation and stability of its target RNAs[70–73]. Co-immunoprecipitation revealed a very large RNA interactome of Unr with more than hundreds of identified transcripts[74]. This reflects its widespread biological roles as a global regulator, which seems to act as a protein-RNA connector, that brings binding partners together or stabilizes their interaction (Figure 3a)[50]. The model further allows to explain the bidirectional role of Unr: it not only promotes, but also represses the translation of certain mRNAs and it increases, but also decreases the abundance of others (Figure 3b).

Unr positively regulates IRES-dependent translation of the transcripts of the cell-cycle kinase PITSLRE[75], the oncogene c-myc[76] and the key activator of caspase 9, APAF-1, explaining its influence on the apoptotic process[77]. Further Unr also promotes the protein synthesis of the human rhinovirus and poliovirus, by stimulating their cap independent translation initiation (Figure 3b)[78,79]. Besides the cap independent regulation, Unr plays also a role in the cap dependent translation reprogramming. A study could show, that a depletion of Unr led to reduction of cap-dependent translation initiation activity by 33% in HeLa and U2OS cells (Figure 3b)[80]. One explanation might be that Unr mediates and enhances interaction between eIF4G and PABP1, which further increases the interaction between the poly(A) tail and the 5' cap of the mRNA[80]. Another mechanism might be phase separation based. Recently a novel class of nucleoplasmic reticulum (NR), named CSDE1-NR, was characterized[81]. Except of Unr also poly(A)-RNAs, ribosomes and translation factors, like eIF4E, were concentrated in these

foci, leading to increased cap dependent translation initiation due to enhanced concentrations of the single components.

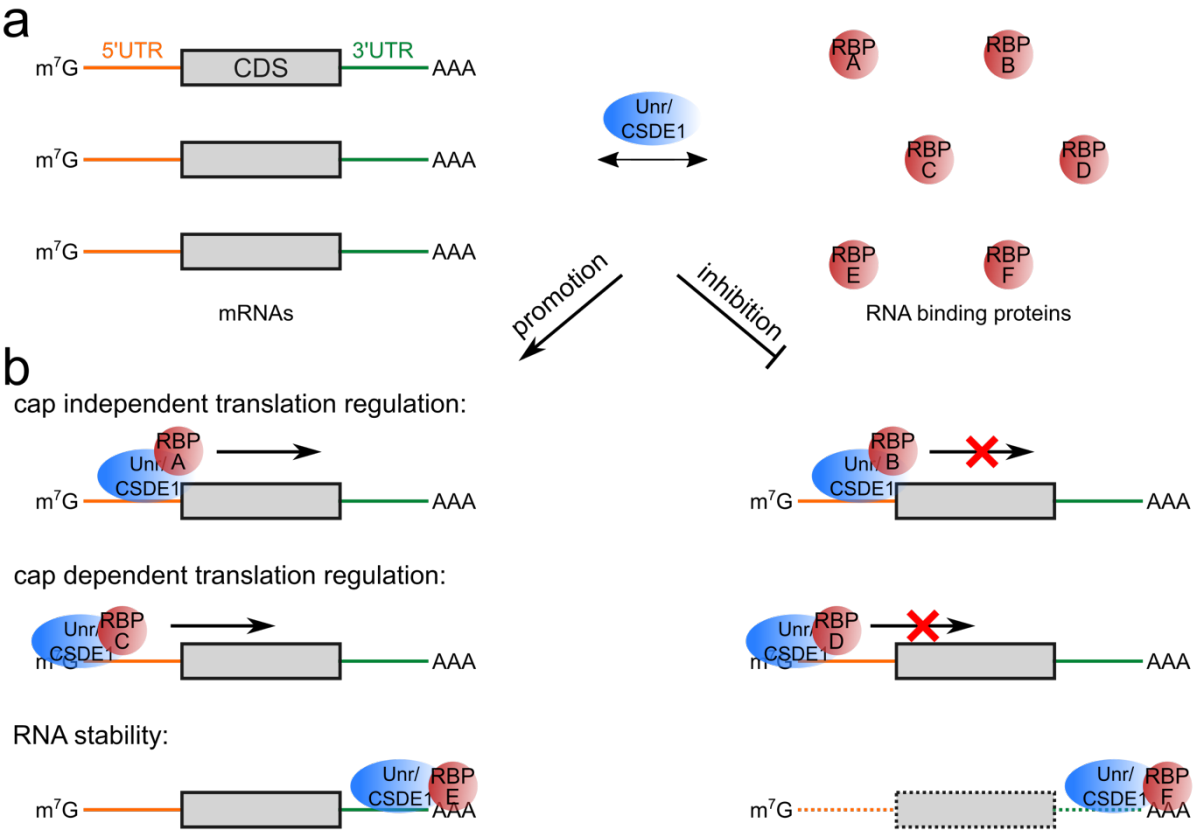


Figure 3: Unr/CSDE1 and its bidirectional role in translation regulation.
a: Unr/CSDE1 promotes the interaction of different RBPs (red; right) with their target RNAs (left), supposed to its function as a protein-RNA connector. **b**: Unr/CSDE1 is involved in cap-dependent and cap-independent translation regulation and also increases and decreases RNA abundance through interaction with both, the mRNA and another regulatory RBP. Adapted from Guo et al.[50].

Contrary to its positive regulation, Unr is also involved in translation repression. One example is the binding to its own mRNA, which results in an autoregulation of its own expression via repressing IRES activity[71]. Another one is its binding to adenine rich sequences in the 5' UTR of the *pAbP1* mRNA together with Imp, leading to cap dependent repression by stalling the movement of the 43S pre initiation complex (Figure 3b)[82,83]. Apart from the cap dependent or independent regulation, Unr is also involved in RNA stability or destabilization, contributing indirectly to the translation of the target mRNAs[84–89] (Figure 3b). Among several mRNAs also the non-coding RNA BC200 is shown to be affected[90].

Due to its diverse and as discussed above bidirectional roles in translation regulation within the cell, human Unr has been linked to several diseases, like Diamond-Blackfan anemia, autism and cancer progression[73,91–95], which is further reflected by high expression levels in sarcoma, melanoma and breast cancer cell lines[96].

1.5. The special role of *Drosophila* Unr in dosage compensation

A peculiarity of *Drosophila* Unr is its involvement during dosage compensation. In comparison to the X-chromosome, which encodes for genes that are necessary for survival, the Y chromosome became largely degenerated by coding just a comparatively low number of genes. In flies, just like in mammals, males have only one X chromosome (Figure 4a). Throughout the evolution different transcription regulation mechanisms evolved to overcome the lethal imbalance of the sex chromosomes. Whereas in mammals one female X-chromosome is packed into a Barr body, bypassing transcription[97], in flies an RNP complex called the male specific lethal (MSL) complex upregulates transcription of the single X chromosome via epigenetic modifications[98].

The *Drosophila* MSL complex consists of five multi-domain proteins, namely MLE (maleless), Msl1, Msl2, Msl3 and Mof (male absent on first) and two long non-coding RNAs (lncRNAs) roX1 and roX2 (RNA on X), that are redundant in their function, but differ in size (3.7 kb versus 0.6 kb) (Figure 4b, lower panel). Although, the entire complex works as a rather dynamic entity, a division into four separated mechanisms can be made: first targeting, second assembly, third spreading and fourth homeostasis. During the first two steps, which may take place simultaneously, MLE, an RNA helicase, remodels stem loop six and seven of the lncRNAs and thereby enables its incorporation into the remaining MSL complex, initiating chromatin binding[99,100]. Militti et al. could show that the interaction of MLE and roX is promoted by Unr, which may function as an RNA chaperone stabilizing the remodeled RNA structure until binding of the core proteins of the MSL complex[101]. After localization to high affinity binding sites on the X chromosome[102–106], the complex spreads to low affinity sites covering the whole chromosome (Figure 4b, upper panel). Two different mechanisms explain the twofold hyper transcription. One describes a higher accessibility of the DNA, through relaxation of the chromatin, which is achieved through methylation of lysine 16 of histone tail 4 (H4K16ac) by the histone acetyltransferase Mof[107–109]. The other one proposes the importance of Msl1 for efficient phosphorylation of Pol II Ser5 during the early transcriptional event, which promotes Pol II moving away from the initiation site[110,111].

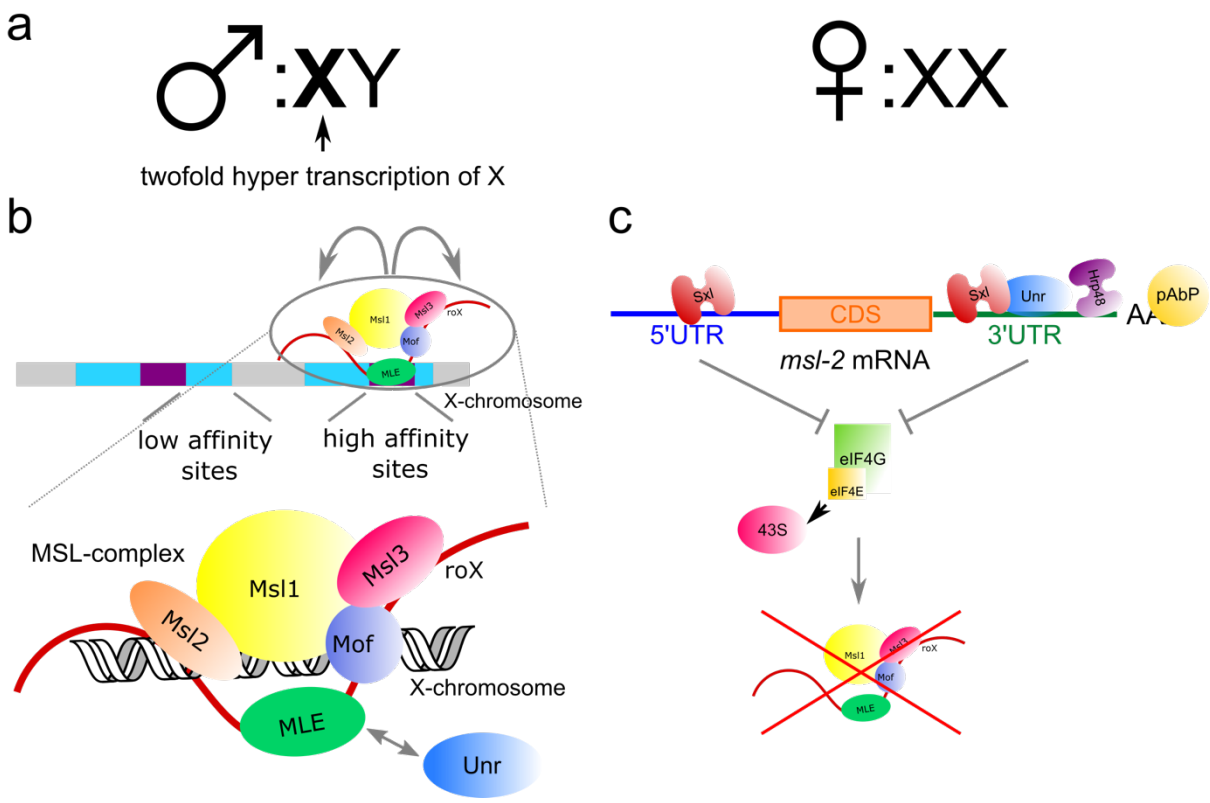


Figure 4: Dosage compensation in *Drosophila* indicating a dual sex specific role of Unr.
a: Like mammals *Drosophila* males lack one X-chromosome compared to the females. Equalization of X-linked genes is achieved by a twofold hyper transcription of the X-linked genes. **b:** The MSL complex mediates the hyper transcription of the whole X-chromosome, by spreading from high affinity sites to low affinity sites and from there throughout the whole chromosome. The complex consists of five proteins (MLE, Msl1, Msl2, Msl3 and Mof) and the long non-coding RNA roX. Its assembly relies on the presence of Unr. **c**: Translation of the *msl2* mRNA is inhibited in the 3' UTR by binding of Sxl and in the 5' UTR by an interaction of Sxl, Unr, Hrp48 and pAbP. The lack of the rate limiting factor Msl2 inhibits formation of the MSL-complex.

Contrary to its involvement in males, where it promotes the assembly of the MSL complex and thus is part of transcription regulation, *Drosophila* Unr is involved in the inhibition of the same complex formation in female flies via translation regulation. Together with other interaction partners it is responsible for the translation repression of the *msl2* mRNA. With Msl2 being one of the rate-limiting components[112–114], this leads to repression of the otherwise lethal assembly of the MSL complex (Figure 4c). One of the most studied translation inhibitory factors in this system is the protein Sex-lethal (Sxl). The primarily nuclear protein inhibits splicing of a facultative intron in the 5' UTR of the *msl2* pre-mRNA, promoting nuclear retention[115–117]. Additionally, Sxl is binding to uridine stretches in both untranslated regions (UTR) of the mRNA (Figure 4c)[118–120]. Thus, its binding to the 3' UTR inhibits the recruitments of the small ribosomal subunit[121,122], whereas the 5' UTR binding prevents the scanning of those subunits, that have escaped the 3' UTR mediated control by a mechanism that entails recognition of an

upstream AUG[112,113,123]. In contrast to the 5' UTR, where Sxl binds individually, a complex formation of multiple proteins is involved in the translation regulation at the 3' UTR. One of the critical components is Unr, which is interacting via its first CSD with Sxl and the mRNA (Figure 5e)[27,112–114]. Recently another protein Hrp48 (Hrb27C), which might be responsible for the interaction with the translation initiation factor eIF3d, was identified to interact with Sxl, Unr and the *msl2* mRNA (Figure 4c)[124]. Apart from the direct interaction with eIF3d, this inhibitory complex might repress translation via binding of Unr to the poly-A tail binding protein pAbP. It is hypothesized that this interaction might lead to an enhanced ability of the repressor complex to interfere with the 43S recruitment[125]. However, except for a crystal structure showing the trimeric interaction of Sxl RRM12-Unr CSD1-*msl2* mRNA[27], structural data are missing, leading to a lack of understanding of the molecular mechanisms of this RNP complex and potential additional interaction partners.

1.6. Structural characteristics of Unr

In previously published studies *Drosophila* and human Unr are delineated to contain five cold shock-domains (CSDs) distributed evenly throughout the protein sequence, that are engaging the target RNAs (Figure 5a)[126,127]. With *Drosophila* Unr being 1039 amino acids long, the CSDs account for only 31% of the entire protein, leading to the suggestion that the rest of the protein is unstructured. CSDs, that were first discovered in proteins that were induced after cold-shocks in *E. coli*, are about 70 amino acid long β-barrel protein domains, consisting of five antiparallel β-strands, similar to the oligonucleotide/oligosaccharide fold (OB-fold), that bind single stranded RNA (ssRNA) (Figure 5b and 5c)[128–133]. The binding with a preference towards purine rich motifs is predominantly achieved by four aromatic residues (**F**G**F** and (**F/Y**)F**H**), which are accessible to the RNA by pointing to the outside of the β-barrel, leading to a non-covalent π-π-stacking of the aromatic sidechains with the bases of the ssRNA (Figure 5b)[126,134,135]. Human and *Drosophila* Unr are highly conserved with a sequence identity of >42%, showing the biggest differences in the two additional Q-rich regions of *Drosophila* Unr, that are absent in the human version (Supplementary Figure 1).

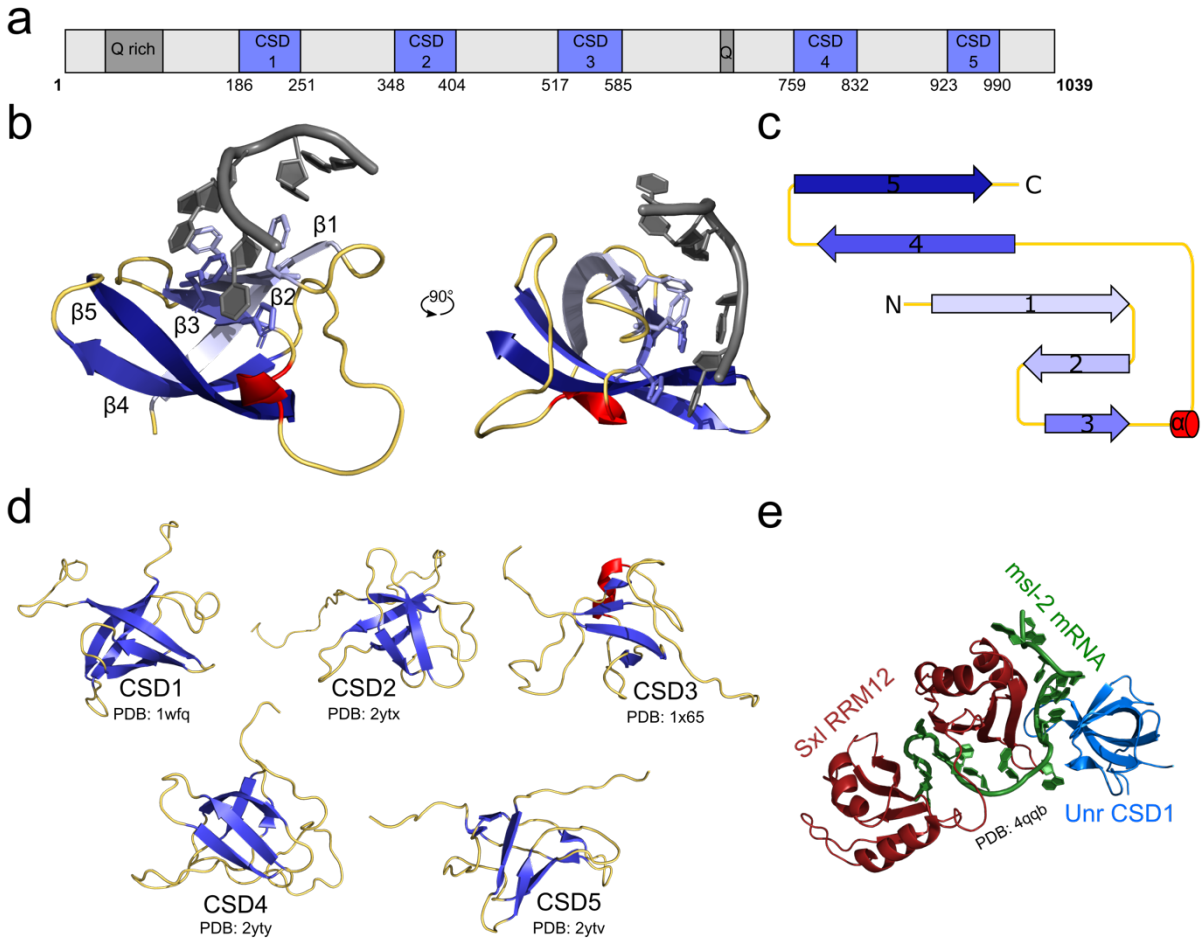


Figure 5: The cold shock-domain fold and structures of Unr.
a: Domain arrangement of the five CSDs and two q-rich domains in *Drosophila* Unr. **b**: Crystal structure of the CSD of Bs-CspB, showing the typical β-barrel fold composed out of five antiparallel β-sheets (blue). The domain binds a Uridine 5mer (grey) mainly via its aromatic RNA binding residues (sticks) (PDB: 3pf5)[129]. **c**: A 2D topology diagram of the CSD fold decomposed from the crystal structure shown in (a) using the Pro-origami webserver[136]. **d**: The five solution NMR structures of the single CSDs of human Unr/CSDE1, showing the typical CSD fold. For each domain one of the 20 lowest energy structures is shown (PBD: 1wfq; 2ytx; 1x65; 2yty; 2ytv)[127]. **e**: A crystal structure shows the interaction between the two RRMs of Sxl with the first CSD of Unr (PDB: 4qqb)[27].

Detailed structural knowledge of Unr is scarce, but important to understand its mechanistic involvement in the described biological functions. Moreover, structural data on Unr would help to understand its RNA specificity and its bidirectional function, represented by translation up- and down-regulation, within the cell. Up to date NMR structures of the five predicted single human CSDs have been determined (Figure 5d). Unfortunately, studies including their RNA binding properties are still missing[127]. For *Drosophila* Unr only a crystal structure of the first CSD is available with detailed information regarding its RNA recognition (Figure 5e)[27]. CSDs in general have shown to be rather promiscuous RNA binders with low sequence specificity when acting alone in *in vitro* contexts[129,137–141]. Also, the first CSD of Unr binds rather weakly to RNA and seem to lack base selectivity. Thus, the specificity of the single CSDs cannot

explain the target specificity of the full-length protein. An increased specificity of Unr might be achieved through an interplay of all CSDs or due to an interaction with other RBPs, like the Sxl-Unr-*msl2* RNP[27]. Unfortunately, larger multi-domain or even complex structures are not yet available.

1.7. Techniques used for integrated structural biology

Different methods for protein and RNA-protein structure determination were used to gain structural insights into the RNA recognition of Unr. Information from different methods were used in an integrative approach to bring together all collected restraints and obtain a full-length model of Unr. The main methods, that were used in this thesis are briefly explained.

For a long time, X-ray crystallography was used as the main technique to study macromolecular structures of proteins, nucleic acids or complexes of both. With the notice, that multi-domain proteins and complex ensembles are often of a dynamic nature, harboring flexible regions or linkers, the need to combine several structural biology methods has been revealed[142,143]. Such methods include X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), electron microscopy, small angle scattering, mass spectrometry, advanced light microscopy techniques and bioinformatics. An integrated approach ideally provides complementary structural information, that help to understand the function of biological macromolecules including their dynamics. In this thesis especially NMR, crystallography and small angle X-ray scattering (SAXS) were used and combined to study the multidomain RNA binding protein Unr.

1.7.1. Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique, that allows not only to study atomic structures of molecules, but also to study their dynamics or interactions with binding partners. Compared to other structural methods, a disadvantage in NMR is the size restriction of the observed molecule. However, progress in selective isotope labeling and deuteration of proteins, specific amino acids or nucleic acids made it possible to study significantly larger macromolecules[144–146].

The principles of NMR spectroscopy

NMR relies on the concept of a nuclear ‘spin’, which is a purely quantum mechanical quantity without a classical analogue. Only nuclei with an odd mass, like ^{1}H, ^{13}C and ^{15}N have this property and thus are candidates for NMR spectroscopy. The rotation of the charged nuclei around a certain axis is called the spin angular momentum I.

Nuclei as charged particles produce through their rotation a magnetic momentum μ. This is directly proportional to the spin angular momentum I with a proportionality constant, which is called the gyromagnetic ratio (γ) and is dependent on the nucleus (Equation 1).

$$\mu = \gamma I \qquad (Equation\ 1)$$

Without any external magnetic influence, the spin would be randomly oriented. As soon as an external magnetic field (B_0) is present, the magnetic moment μ processes in a cone like shape around it. This motion is called Larmor precession and the frequency is directly proportional to the strength of the external magnetic field, meaning the stronger the external field, the higher is the precession frequency (Equation 2)[147].

$$\upsilon_0 = \frac{|\gamma| B_0}{2\pi} \qquad (Equation\ 2)$$

If an additional weaker magnetic field B_1 exerts a torque on the magnetic moment of the nuclei, they change their precession angle θ around B_0. The resulting motion of μ can be described as caused by the resultant field B_0+B_1. If B_1 is perpendicular do B_0 and μ, the exerted torque on μ by B_1 would be away from B_0. In an NMR spectrometer B_0 is represented by the main magnetic field and B_1 by a radiofrequency field, that excites the spins. By using a vector formalism, the previous mentioned rotation of μ away from B_0 can be depicted. With the mere influence of B_0 the bulk magnetization M(t) experiences a torque towards the z-axis and parallel to B_0 (Equation 3).

$$\frac{dM(t)}{dt} = M(t) x\, \gamma B(t) \qquad (Equation\ 3)$$

A shorter radio pulse applied for example along the x-Axis will turn the magnetization vector towards the -y-axis. The direction of rotation can be determined by the right-hand rule, known from electromagnetism. The length of the applied pulse determines the angle of rotation θ. The rotating magnetization creates a voltage in a receiving coil, which is acquired for subsequent processing. By applying a Fourier analysis on the received signal, one can represent the magnetization in one dimension. After the radio pulse is switched off, the bulk magnetization will return to the ground state, due to relaxation effects. In addition to the radio pulse, also the specific environment of each spin influences its tumbling. Especially electrons, that orbit every nucleus, promote a local electronic environment, known as shielding. The shielding creates a specific signal for almost each nucleus in a molecule, which is helpful for later analysis.

Protein NMR

One of the easiest experiments in NMR is the ^{1}H-one-dimensional spectrum. It is unique for each molecule. Information gained by this spectrum is often sufficient for small molecules, however the spectrum of macromolecules like proteins becomes too complex with too many overlapping signals. Nevertheless, from recording this simple and quick spectrum, which also does not require any isotope labeling, one can see, whether a protein is folded (broad peak dispersion), unfolded (narrow line width and narrow dispersion) or aggregated (broad line width) (Figure 6).

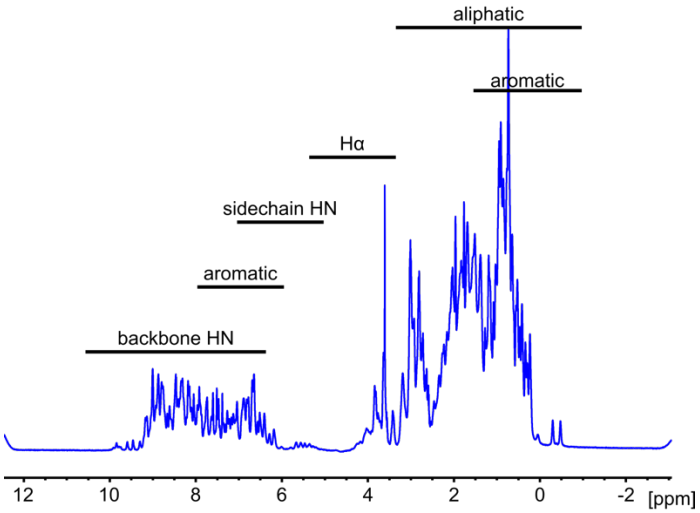


Figure 6: ^{1}H-Spectrum of one of the cold-shock-domains of *drosophila* Unr.
The spectrum is typically recorded from -2 - 12 ppm. The information content about special protein regions is highlighted. The broad peak dispersion and the extremely up-field shifted methyl group peaks, indicate that the protein is folded. The sharp peaks further indicate, that the protein is not aggregated.

In order to gain more information about a protein and to reduce signal overlap an additional dimension can be introduced. Two-dimensional spectra can be either homo- (e.g. ^{1}H-^{1}H) or heteronuclear. For proteins nitrogen is normally the first choice, but also ^{1}H-^{13}C spectra can be quite useful. The most common ^{1}H-^{15}N experiment is the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum[148], which shows a cross peak for every directly bonded proton-nitrogen pair, resulting in one peak for every backbone amide of a non-proline residue and additional peaks for the sidechains of asparagine, arginine, glutamine, tryptophan and lysine (often not visible due to fast chemical exchange). Due to its unique nature for each protein, the ^{1}H,^{15}N-HSQC is often called the fingerprint of a protein

The amino acid resolution and the high sensitivity towards changes in the local environment make the spectrum ideal to observe protein-ligand interactions. Upon binding of a ligand, the

chemical surrounding of an amino acid changes, leading to shifted positions in the spectrum. By mapping of these so-called chemical shift perturbations (CSP) one can get to know the binding site of a ligand on the protein and determine also the binding affinity. Unfortunately, for larger proteins many signals do overlap in the $^{1}H,^{15}N$-HSQC and the low T_2 relaxation (transverse relaxation) times making the analysis difficult. Nevertheless, certain methodological developments allow to increase this size limit, for example by deuteration of the protein[149] or by pulse programs, that reduce the relaxation effects and thereby result in better line shapes (Transverse relaxation optimized spectroscopy - TROSY)[150,151].

In order to assign the different resonances measured in the $^{1}H,^{15}N$-HSQC to the specific backbone NH group of the different amino acids in the protein, triple resonance spectra need to be recorded. Typical spectra include HNCA, HNCACB and CBCA(CO)NH triple resonance experiments[152], which correlate the backbone amide group with the Cα and eventually the Cβ carbon atoms of the same or the previous amino acid. Thereby, sequential connections between all amino acids, except prolines, can be made. Due to known averaged Cα and Cβ shifts for each amino acid type, which are also unique for some amino acids like alanine or threonine, these chains can be then linked to the protein sequence (Figure 7).

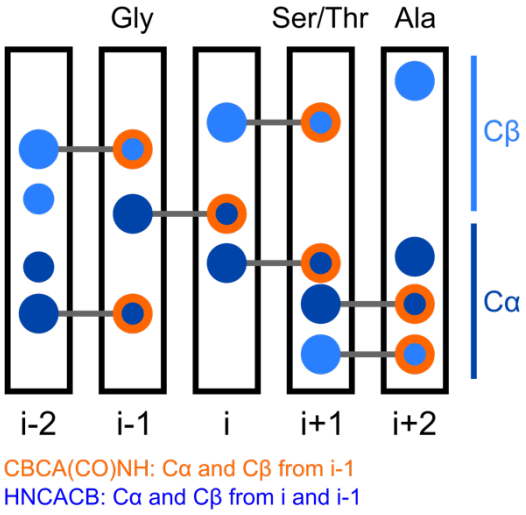


Figure 7: The theory of chain tracing in protein backbone assignment.
Overview of the backbone assignment of proteins using HNCACB and CBCA(CO)NH spectra. The i-1 signals from the CBCA(CO)NH spectrum (orange) needs to be linked to the HNCACB signals (blue) of another residue. This creates a chain that can be later assigned to the protein sequence, due to specific shifts of certain amino acids.

Structure calculations using NMR assignments

For structure calculation by NMR preferably all sidechain protons need to be assigned. In a next step so called nuclear Overhauser effect (NOE) measurements are carried out and analyzed[153]. NOEs are observed when magnetization is transferred between two spin-active

nuclei that are close in space (<5 Å). The volume of the peak reflects the relative distance between the two nuclei. By using a network of distance restraints, the structure is calculated by simulated annealing. Additionally, backbone dihedral angle restraints (Φ and Ψ), that are predicted from backbone chemical shifts by comparing them from known NMR structures can be incorporated as additional restraints during structure calculations[154]. For complex structures, that potentially lack a satisfying amount of NOE restraints, relative positions of structural elements can be obtained from residual dipolar couplings (RDCs)[155,156] or paramagnetic relaxation enhancement (PRE) measurements[157]. Further, chemical properties, like chirality or planarity of the aromatic ring or bond length and angles, are aimed to be fulfilled during the structure calculation. The program tries to find an ensemble of the lowest energy structures in several refinement rounds, that fulfills as many restraints as possible. The final amount of fulfilled restraints reflects the quality of the final ensemble.

Relaxation and dynamic measurements in NMR

NMR is a powerful technique to study protein dynamics. Nuclei that return to their thermodynamically stable states after being exited to higher energy levels is a phenomenon called relaxation in NMR. The energy that was absorbed during the excitation is released, which can be a complex process based on different timescales of the relaxation. The most common types of relaxation in NMR are spin-lattice relaxation (T_1 or longitudinal relaxation along z-magnetization) and spin-spin relaxation (T_2 or transverse relaxation along x-y magnetization).

T_1 is defined as the average time that it takes for the bulk magnetization to recover along the z-axis from the x-y plane after getting excited, meaning it reflects the average lifetime of nuclei in higher energy states. The sample in which the nuclei are kept can be referred as lattice and T_1 is induced by the interaction of the nuclei with it. Thereby not only the gyromagnetic ratio of the nucleus, but also the mobility of the lattice can influence the spin-lattice relaxation (e.g. viscosity, temperature and ionic content). The more efficient the relaxation is, the smaller is T_1.

T_2 is the time that it takes for spins to lose coherence with each other. Neighboring nuclei with identical precession, but different magnetic quantum states, can interact, meaning they can exchange their quantum states. Nuclei at lower energy states can get excited, while other nuclei relax to lower energy states, leading to the loss of coherence in x-y magnetization. Although the overall energy state within the population stays the same, the average lifetime of a single nucleus in an excited state decreases. In a spectrum, this is reflected in line broadening of the peaks. T_2 is always smaller than T_1.

For rigid protein molecules with an assumed isotropic tumbling, the ratio between the ^{15}N relaxation times T_1 and T_2 is represented in the rotational correlation time τ_c, where ν_N is the ^{15}N resonance frequency (in Hz) (Equation 4)[158].

$$\tau_C \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{T_1}{T_2} - 7} \qquad (Equation\ 4)$$

The rotational correlation time reflects the molecular tumbling of molecules in a solution and gives information about the protein size and flexibility of each amino acid. Flexible regions within a protein or smaller molecules show faster overall tumbling, whereas the tumbling rate decreases with increased protein size. As a general rule of thumb, τ_c of a monomeric protein in solution is approximately 0.6 times its molecular weight (kDa)[159].

1.7.2. X-Ray crystallography

Obtaining high resolution structures by X-ray crystallography is the primarily used tool, followed by NMR and EM. Although harboring several drawbacks including the lack of depicting dynamic systems or the need for crystal growth, the advantages of no size limitation and the ease of use still makes it a desired technique. However, in recent years and with the resolution revolution, cryo-EM develops into the method of choice for structural biologists, especially when studying larger complexes[160,161].

Protein crystallization

A prerequisite to obtain a high-resolution structure by X-ray crystallography is a well diffracting crystal. As for NMR highly pure and homogeneous samples are required. Growth of crystals is a multi-parametric process involving nucleation and growth. For nucleation the solution needs to be in the super-saturation phase, which is between precipitation and under-saturation phase and depends on the molecule's and precipitant's concentration. Usually in the beginning hundreds of different buffer compositions are screened over several weeks and at different temperatures to obtain the perfect crystallization condition. Often the screens are carried out in 96 well plates using vapor diffusion either by sitting or hanging drop methods. During the set up the macromolecular solution is mixed with the crystallization buffer. Afterwards the drop equilibrates over time with the reservoir solution, due to differences in the vapor pressure and the crystallization drop. Ideally the slow concentration increase of the precipitant results in reaching the super-saturation phase. In case no diffracting crystal was received in the initial screens, promising conditions, that show for example either phase

separation or needle growth, can be refined by screening various parameters of the initial crystallization buffer solution.

The principles of X-ray crystallography

The property of electrons to diffract X-rays is the key component of macromolecular crystallography. The lattice structure of a protein crystal, where the orientation of the macromolecule repeats within each unit cell is defined by three axes and the angles between them, which are named a, b, c and α, β, γ, respectively. To obtain a simplified crystal lattice each corner point of a unit cell could be represented by a point. Within the lattice infinite numbers of planes could dissect each site of the unit cell. So called Miller indices (h, k, l) describe how often a plane dissects a, b and c (Figure 8a)[162,163].

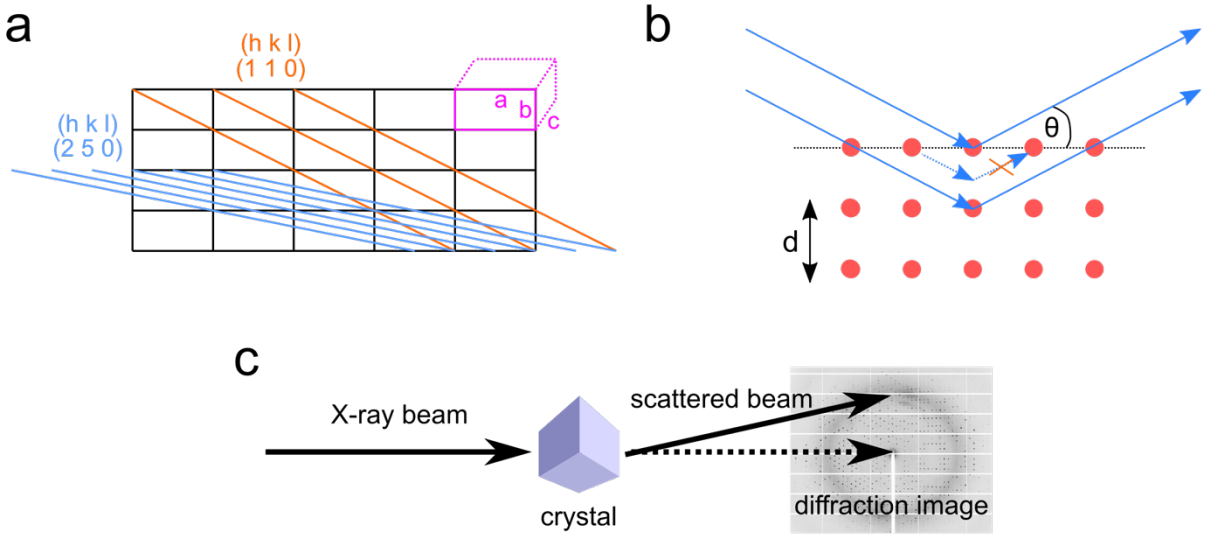


Figure 8: Principles of X-ray crystallography.
a: Schematic presentation of a two-dimensional crystal lattice, with exemplified planes with Miller indices, dividing the sides of the unit cell. **b**: Red dots represent the Miller plane formed by the lattice points. The X-ray beam (blue) is only diffracted by the Miller planes, which satisfy the Bragg's equation. **c**: Typical arrangement of an X-ray diffraction experiment. The crystal reflects the primary X-ray, which results in a typical diffraction pattern.

When X-rays hit the crystal, they get diffracted due to their interaction with the electrons of the macromolecule. According to Bragg's law, the different sets of planes defined by the Miller indices will produce a constructive interference at certain angles θ (Figure 8b, Equation 5). At all other angles, that do not fulfill the Bragg's equation, the waves extinct each other. Therefore, one observes the typical diffraction pattern of crystals (Figure 8c)[164].

$$n\lambda = 2dsin\theta \qquad (Equation\ 5)$$

In the equation n represents a positive integer, λ the radiation wavelength, θ the scattering angle and d the spacing between the Miller planes. Further, the distance d determines the resolution

of a certain reflection, meaning that the resolution of the reflection with maximal θ determines the resolution of the data set. Taken together, the lattice structure within a crystal enhances the scattering of X-rays into one direction, whereas it completely extinguishes it in others.

The final goal in crystallography is obtaining an electron density map, which can be used for the model building of the structure. To obtain a large number of scattering signals, the crystal is usually rotated 90-360° on a goniometer in the crystal beam. The resolution of the dataset reflects the minimum distance of two atoms, that can be resolved in the electron density map. The electron density $\sigma_{(x,y,z)}$ within one unit cell of the crystal is obtained via a Fourier transform using the following equation (Equation 6):

$$\sigma_{(x,y,z)} = \frac{1}{V}\sum_{hkl} F_{(hkl)} e^{-i\alpha_{hkl}} e^{[-2\pi i(hx+ky+lz)]} \qquad (Equation\ 6)$$

In order to calculate the electron density for each point (x, y, z) in the unit cell, the phase α_{hkl} and the amplitude needs to be known. The amplitude can be directly obtained from the reflection intensity, but the phase gets lost from the reduction of a three-dimensional event on a two-dimensional detector, depicting the term 'phase problem', one of the biggest hurdles in X-ray crystallography.

Examples to solve the phase problem

There are three different methods available, that can be used to solve the phase problem, namely molecular replacement (MR)[165], multiple isomorphous replacement (MIR)[166–168] and multi-wavelength anomalous diffraction (MAD)[169,170]. In the following MR and MAD will be described briefly, as they were used in this thesis.

For the usage of MR, the availability of a highly homologous protein structure is a prerequisite. The suitable structure is then used to obtain a model of the unknown structure, by taking a starting phase of the homologous structure and the amplitudes of the measured data set. In order to get a good fit between the model and the unknown structure, the orientation in the unit cell of both molecules needs to show the maximal overlap of the diffraction patterns. Afterwards the input phases are iteratively refined until a fitting final electron density map is calculated. One of the drawbacks of MR is the need of a high similarity between the known and unknown structure. Already small flexible regions may not superimpose with the novel structure, leading to difficulties in the calculations. Therefore, one can try to vary the input model, by cropping off flexible regions or sidechains. Despite this disadvantage, MR evolved to the method of

choice, not least because of the enormous number of available protein structures in the protein data bank, that can be used as input models.

In case there is no suitable atomistic model available, one can try to solve the phase problem using MAD. This technique uses the presence of anomalous scattering within a crystal, which is obtained by heavy atom derivatized crystals. Heavy atoms can be directly incorporated into the protein by the presence of selenomethionine during protein expression, or later by soaking the crystal into different heavy atom containing solutions. Diffraction data are then measured at a number of different X-ray wavelengths, that show significant differences in the scattering of the heavy atoms. The differences in resulting diffraction patterns allow to extract information about phase differences. Typically, datasets are acquired near the absorption edge, the absorption peak and at high energy wavelength remote from the absorption peak. To select the appropriate wavelength, recording of an X-ray fluorescence spectrum is required. Due to the use of wavelength tunable synchrotron radiation, which typically allow the selection of different wavelengths in a range of 0.5-3.0 Å, this became comparable easy within the last years.

1.7.1. Small-angle X-ray scattering

Compared to crystallography and NMR, small-angle X-ray scattering (SAXS) is neither dependent on crystal growth, nor restricted to any size of the measured macromolecule. Within the last years optimization of the measurement hardware, especially at the synchrotron bioSAXS beamlines, and analyzing software, which was mainly driven by the Svergun group[171], developed SAXS to a user-friendly technique. Another advantage is the low material cost of around 5-50 μg of the macromolecule, that is needed for one measurement.

One disadvantage of the technique is the low resolution of around 20 Å, which can give information about the overall shape of the protein or biomolecular complex. However, geometric parameters, like the molecular weight, the excluded particle volume, the radius of gyration (R_g) and the maximum dimension (D_{max}) can be determined from the scattering data. Therefore, the macromolecule in solution is exposed to monochromatic X-rays. The diffraction pattern of the non-orientated molecules of the scattered X-rays is finally recorded by a two-dimensional X-ray detector (Figure 8). The spatial averaged signal due to molecular tumbling in solution is the reason for the low resolution of the method.

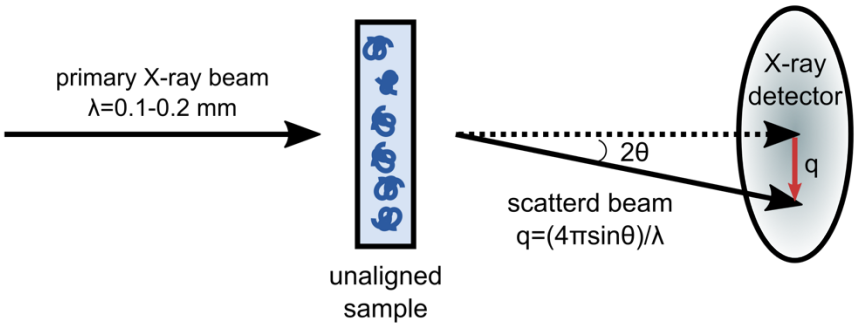


Figure 9: Schematic representation of a SAXS experiment.
The primary X-rays, that are scattered by the macromolecular sample are subtracted from the solvent's diffraction. To obtain the experimental 1D SAXS curve the scattering profile from the two-dimensional X-ray detector is circularly averaged.

A scattering vector q expresses the intensities of the diffracted X-rays. Q is inversely proportional to the wavelength λ and directly proportional to the scattering angle θ (Equation 7).

$$q = \frac{4\pi \sin(\theta)}{\lambda} \qquad (Equation\ 7)$$

Most information about the measured molecule is in the low q range of the scattering curve (0-5 mm^{-1}), which rapidly decays and describes the shape of the particle. In the medium resolution (5-10 mm^{-1}), the differences in the scattering vanishes between different macromolecules and is eradiated completely at higher q values. Beyond 20 nm^{-1} SAXS data are usually not recorded, since the water overlays the scattering of interest. To avoid that the final scattering curve of the macromolecule is influenced by the scattering of the buffer, the intensity has to be subtracted to obtain the scattering curve of the pure macromolecule. Further the scattering intensities correlate with the sample concentration, where higher concentrations result in a better signal-to-noise ratio of the buffer subtracted data. However, the aggregation potential of each molecule restricts the concentration limit, as well as potential radiation damage, which is more likely at higher concentrations.

The determination of the experimental R_g of a molecule, which describes the average root-mean-square distance to the center of density in the molecule, is based on the Guinier approximation. This is only true for regions without intramolecular interference. Oligomerization of the macromolecule can be observed in the guinier plot, due to the high influence of polydispersity and aggregation on the R_g. The Guinier plot (ln I(s) vs s^2) determines the R_g by the slope of the linear part, which should satisfy the condition $sx R_g < 1.3$. Additionally, SAXS data provide information on the presence of flexibility within the measured macromolecule. On the one hand the so-called Kratky plot (s^2I(s) vs q) visualizes particular

features of the scattering profile, with characteristic curves for globular, multidomain, unfolded or flexible proteins. Thereby, the identification of the folding state can be easily assessed. On the other hand, the comparison between the measured R_g and the D_{max} can help to assess the unstructured nature of a macromolecule, since globular molecules harbor a more tightly packed core than flexible ones.

SAXS data can be also used to determine low resolution *ab initio* shape models of the protein. Dummy beads fill a search volume, that is sufficient enough to represent the protein. A theoretical SAXS curve of the bead model is calculated and compared with the measured one. By using trial and error the beads are moved until the χ^2 of the superimposed theoretical and experimental curve is at its minimum. The possibility to theoretically calculate SAXS curves can be used further to validate high resolution structures in solution environment. The back calculated scattering curve of the structure can be compared to the experimentally measured one. This is especially helpful to validate crystal structures, where artifacts could occur due to crystal packing forces.

Overall SAXS data can be extremely useful to validate high resolution structures, assess the flexible nature of the macromolecule and observe polydispersity or oligomeric states of the macromolecule at different concentrations or changes of the shape in absence or present of potential ligands. In this thesis, besides using SAXS for structure validation, it is also used in combination with NMR and X-ray crystallography data to obtain a full-length structural model of Unr.

2. Aims of the thesis

This thesis aims to contribute to the understanding of the biological role of *Drosophila* Unr, by giving structural indices that help to understand how this protein conducts such bidirectional function within the cell. Thereby, structural data may also help to understand the basis of Unr-mediated translation repression, but also activation, via formation of different RNP complexes. Further, it aims to enlarge our knowledge about multidomain RBPs and how they orchestrate RNA target specificity, since structural information explaining different mechanisms is scarce.

To accomplish this, first the structural knowledge about Unr needs to get extended. Therefore, studies on the full-length protein, just as a divide-and-conquer approach will help to gain a large amount of structural data.

In order to connect these data to its biological function, the RNA recognition of Unr will be studied. This will not only help to understand binding mechanisms, that go beyond the canonical CSD-RNA recognition, but also increase the knowledge about cooperative binding, like it occurs in many multidomain RBPs.

The structural findings and the connected RNA binding properties will be validated in a functional and cellular context. Thereby, the biological relevance will be assessed using mutational studies.

Additionally, the RNP interactome of Unr will be analyzed in order to gain information about larger molecular assemblies. Together with the previously assessed structural data, this will help to pave the way towards the structural reconstitution of a translation regulation RNP complex.

Achieving these aims requires an integrative multipronged approach, that combines different methods, including structural techniques like nuclear magnetic resonance spectroscopy (NMR), X-ray crystallography and small angle X-ray scattering (SAXS), but also biochemical assays like electrophoretic mobility shift assays (EMSA) and fluorescence polarization assays as well as cell-based studies, including reporter gene assays, RNA immunoprecipitation coupled with next generation sequencing (RIP-seq) and proteomics approaches.

3. Material and Methods

3.1. Materials

3.1.1. Devices and consumables

Table 1: List of used devices

Device	Source
Nanodrop 2000	ThermoFisher
ÄktaExplorer	GE Healthcare
Synergy 4	BioTek
MicroCal PEAQ-ITC	Malvern Panalytical
nanoDSF	NanoTemperTech
J-815 CD spectrometer	Jasco
7500 fast Real-Time PCR System	Applied Biosystems
Trans Blot Turbo	BioRad
Bioruptor 300	Diagenode
2100 Bioanalzer	Agilent
Orbitrap Fusion Lumos mass spectrometer	ThermoFisher
Ultimate 3000 nano RSLC system	ThermoFisher
Mosquito LCP	TTPlabtech
UV Stratalinker 1800	Agilent Genomics
NextSeq 500 Sequencing System	Illumina
Typhoon Trio Imager 9000	GE Healthcare
semi dry Trans-Blot Turbo System	BioRad
Avance III 600 MHz	Bruker
Avance III 700 MHz	Bruker
Avance III 800 MHz	Bruker

All further devices correspond to the usual laboratory equipment.

Table 2: List of critical commercial assays.

Kit	Source
MinElute Gel Extraction Kit	Qiagen
QIAprep Spin Miniprep Kit	Qiagen
EndoFree Plasmid Maxi Kit	Qiagen
Plasmid Maxi Kit	Qiagen
Pierce BCA Protein Assay Kit	ThermoFisher
Dual Luciferase Assay	Promega
Turbo DNA free kit	ThermoFisher
MegaScript T7 Kit	ThermoFisher
Galacto Star (Tropix)	ThermoFisher
Renilla-Glo Luciferase Assay	Promega
NEBNext Ultra II Directional RNA library Prep Kit for Illumina	NEB
Pierce Silver Stain Kit	ThermoFisher

Table 3: List of used consumables.

Consumable	Source
RotiGarose-His/Ni NTA-HPBeads	Carl Roth
5 ml HiTrap Heparin Fast Flow	GE Healthcare
HiPrep 26/10 Desalting	GE Healthcare
HiLoad 16/600 Superdex 75 pg	GE Healthcare
HiLoad 16/600 Superdex 200 pg	GE Healthcare
Superdex 200 Increase 10/300 GL	GE Healthcare
Amicon Ultra Centrifugal Filters	Merck Millipore
Anti-FLAG M2 magnetic beads	Sigma-Aldrich
384 Well Low volume black round bottom polystyrene	Corning
G_{50} desalting columns	GE Healthcare
Drosophila overnight embryo extract	Provided by Fàtima Gebauer (CRG)
SybrGreen Master Mix	Applied Biosystems
Schneider's *Drosophila* medium	Gibco
FBS	Gibco
Effectene	Qiagen
Mini-Protean TGX Precast gradient gel 4-20%	BioRad
Pierce Protein A/G Magnetic beads	ThermoFisher
Sera-Mag Speedbeads	ThermoScientific
Precolumn C18 PepMap100	ThermoFisher
Acclaim PepMap 100	ThermoFisher

All further consumables correspond to the usual laboratory standards.

3.1.2. Chemicals

Table 4: List of used chemicals.

Chemical	Source
Ammonium-^{15}N chloride (>98% isotopic purity)	Sigma-Aldrich
D-glucose-$^{13}C_6$ (>99% isotopic purity)	Sigma-Aldrich
D-Glucose-1,2,3,4,5,6,6-d_7, 97-98%) (97% D)	Sigma-Aldrich
2-keto-3,3-d_2-1,2,3,4-^{13}C-butyrate	Cambridge Isotope Laboratories
2-keto-3-methyl-d_3-d_1-1,2,3,4-^{13}C-butyrate	Cambridge Isotope Laboratories
pCp-Cy5	Jena Biosciences
Trizol (RotiZol)	Carl Roth
Pierce ECL Western Blotting Substrate	ThermoFisher
TMT10plex	ThermoFisher

All further chemicals correspond to the usual laboratory standards.

3.1.3. Enzymes and antibodies

Table 5: List of used enzymes.

Enzyme	Source
His_6-TEV prescission	self-made
Restriction enzymes	NEB
DNaseI	Merck
T3 polymerase	ThermoScientific
SuperScript II Reverse Transcriptase	Invitrogen
Rnase OUT	ThermoFisher
Benzonase	Merck
Proteinase K	NEB
Sequencing Grade modified trypsin	Promega
Green Phusion PCR Mix	In house made by PEP Core facility (EMBL)
T4 RNA ligase	NEB
T7 RNA polymerase	In house made by PEP Core facility (EMBL)
Thermostable Inorganic Pyrophosphatase	NEB

Table 6: List of used antibodies.

Antibody	Source
Rabbit polyclonal antibody against *Drosophila* Unr	Provided by Fàtima Gebauer (CRG)
Mouse monoclonal antibody against V5 epitope	Invitrogen AB_2556564
Mouse monoclonal antibody against α-tubulin	Sigma T9026-100 µl; clone DM1A
Goat Anti-Rabbit IgG H&L (HRP)	Abcam ab6721
Goat Anti-Mouse IgG H&L (HRP)	Abcam ab6728

3.1.4. Composition of buffers

Table 7: List of composition of buffers.

Buffer	Chemical	Concentration
10x TBE	Tris	875 mM
	Boric Acid	890 mM
	EDTA pH 8.0	20 mM
10x SDS buffer	Tris-Cl	250 mM
	Glycin	2.5 M
	SDS	1% (w/v)
5x SDS loading dye	Tris-Cl (pH 6.8)	250 mM
	SDS	10% (w/v)
	-Mercaptoethanol	25% (v/v)
	Glycerole	50% (v/v)
	Bromphenolblue	1% (w/v)
Cell lysis buffer	Hepes/NaOH (pH 7.5)	20 mM
	NaCl	500 mM
	Urea	1 M
	Imidazole	30 mM
	β-mercaptoethanol	1.4 mM

NMR buffer	Sodium phosphate (pH 6.4)	20 mM
	NaCl	50 mM
	DTT	1 mM
	Sodium azide	0.2%
Crystallization buffer	Hepes/NaOH (pH 7.5)	20 mM
	NaCl	150 mM
	DTT	1 mM
CD buffer	Hepes/NaOH (pH 7.5)	20 mM
	NaCl	20 mM
	DTT	1 mM
ITC buffer	Sodium phosphate (pH 6.4)	20 mM
	NaCl	50 mM
	TCEP	0.5 mM
FP buffer	Hepes/NaOH (pH 7.5)	20 mM
	NaCl	50 mM
	DTT	2 mM
RIPA buffer	Tris-Cl pH 7.5	10 mM
	NaCl	150 mM
	SDS	0.1%
	DOC	1%
	EDTA	5 mM
	Triton X-100	1%
	Protease inhibitor	1x
	Leupeptin	
	Apoprotin	
	Pepstatin	
PBS-T	NaCl	137 mM
	KCl	2.7 mM
	Na_2HPO_4	10 mM
	KH_2PO_4	1.8 mM
	Tween 20	0.1% (w/v)
100x Trace-Elements	EDTA	5 g
	Add 800 ml ddH_2O	
	adjust pH 7.5	
	$FeCl_3$ x 6 H_2O	0.83 g
	$ZnCl_2$	84 mg
	$CuCl_2$ x 2 H_2O	13 mg
	$CoCl_2$ x 6 H_2O	10 mg
	H_3BO_3	10 mg
	$MnCl_2$ x 6 H_2O	1.6 mg
	ddH_2O	Add to 1 l
10x M9 salts	NaH_2PO_4	60 g
	K_2HPO_4	30 g
	NaCl	5 g
	ddH_2O	To 1 l
	adjust pH to 7.4	

3.1.5. Composition of media

Table 8: List of composition of media.

Medium	Chemical	Concentration
TB (Terrific Broth)	solution 1 (900ml):	
	trypton	12 g
	Yeast extract	24 g
	glycerin	4 ml
	solution 2 (100ml):	
	KH_2PO_4	2.31 g
	K_2HPO_4	12.54 g
	autoclave both solutions and mix afterwards.	
LB (Luria-Bertani)	Trypton	10 g
	Yeast extract	5 g
	NaCl	10 g
	H_2O	1 l
	Adjust pH to 7.4 and autoclave	
SOC	Trypton,	20 g
	Yeast extract	5 g
	NaCl,	0.58 g
	KCl	0.19 g
	$MgCl_2$ x 6 H_2O	2.03 g
	$MgSO_4$ x 7 H_2O	2.46 g
	ddH_2O	960 ml
	autoclave solution	
	add 40 ml of a 50% sterile glucose solution	
Minimalmedium	M9 medium (10x)	100 ml
	Trace elements solution (100x)	10 ml
	20% (w/v) glucose	20 ml
	$MgSO_4$ (1 M)	2 ml
	CaCl2 (1 M)	0.1 ml
	Thiamin (1 mg/ml)	6 ml
	NH4Cl	0.5 g
	ddH_2O	Add to 1 l
	filter solution sterile	

3.1.6. Bacterial and *Drosophila* cell lines

Table 9: List of cell lines.

Cell line	details
E. coli DH5 α	Genotype: *fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
E. coli BL21 (DE3)	*E. coli B dcm ompT hsdS*($r_B^-m_B^-$) *gal*
Schneiders *Drosophila* line 2	Male embryonic *Drosophila melanogaster* cell line (SL2); ATCC CRL-1963

3.1.7. RNA and DNA oligonucleotides

Table 10: List of RNA oligonucleotides.

RNA oligonucleotide sequence	Source
AAA AAA AUG (SL6 apical)	Biomers
AAA AAA AAA AAA AAA (A15mer)	IBA

Table 11: List of DNA oligonucleotides.

DNA oligonucleotide sequence	Name
TAATACGACTCACTATAGGGATTG CTGAAGAAGTTGTATAAGCAA	RNAi targeting Unr fwd
TAATACGACTCACTATAGGGACA CTTGGATTTCAATTTCGTTTTGC	RNAi targeting Unr rev
TAATACGACTCACTATAGGGCCTG AAGTTCATCTGCACCA	RNAi targeting GFP fwd
TAATACGACTCACTATAGGGTGA AATTTGTGATGCTATTGCTTT	RNAi targeting GFP rev
TTGTTTCCAAAAAGGGGTTG	fwd qPCR: firefly luciferase (*in vitro* translation assay)
CATCGACTGAAATCCCTGGT	rev qPCR: firefly luciferase (*in vitro* translation assay)
TATTGCTTTGATCTTATCTTGATGC	fwd qPCR: renilla luciferase (*in vitro* translation assay)
ACAAATATCTTACTGCATGGTTTG	rev qPCR: renilla luciferase (*in vitro* translation assay)
AACGTCGTGACTGGGAAAAC	fwd qPCR: β-galactosidase (cell assay)
GGCCTCTTCGCTATTACGC	rev qPCR: β-galactosidase (cell assay)
ACAAGTACCTCACCGCTTGG	fwd qPCR: renilla luciferase (cell assay)
GACACTCTCAGCATGGACGA	rev qPCR: renilla luciferase (cell assay)

3.1. Methods

3.2.1. Cloning

All plasmids used for large scale protein expression were derived from pETM11 (derived from pBR322; G. Stier) and comprise an N-terminal His_6-affinity tag connected via a tobacco etch virus protease (TEV)-cleavage site to the target protein sequence. However, in order to test the solubility of different construct boundaries, pETM22 (derived from pBR322; G. Stier) with a 3C-PreScission cleavable N-terminal His_6-affinity and a Thioredoxin solubility tag and pETM41 (derived from pBR322; G. Stier) with an N-terminal His_6-affinity and a maltose binding protein solubility tag (MBP) and a TEV-cleavage site were used as expression vectors as well. The insert sequences from *Drosophila* and human Unr were amplified from a vector that was kindly provided by Fátima Gebauer (CRG Barcelona). The pAbP constructs were directly cloned from SL2 cDNA that was generated during this study (Figure 10).

The protein constructs were cloned using the restriction free cloning approach. In a first step an insert amplification with target specific DNA oligonucleotides was conducted. A successful amplification was evaluated by an agarose gel, from which the DNA was also purified and eluted using a MinElute Kit according to the manufactures protocol. To this end, a chaotropic reagent dissolves the gel, before the DNA is immobilized into a spin column, cleaned with ethanol and finally eluted with H_2O or buffer. The eluted insert was used as a super primer for a second PCR amplification into the target vector[172]. Point mutations were inserted by one or multiple steps of site directed mutagenesis[173].

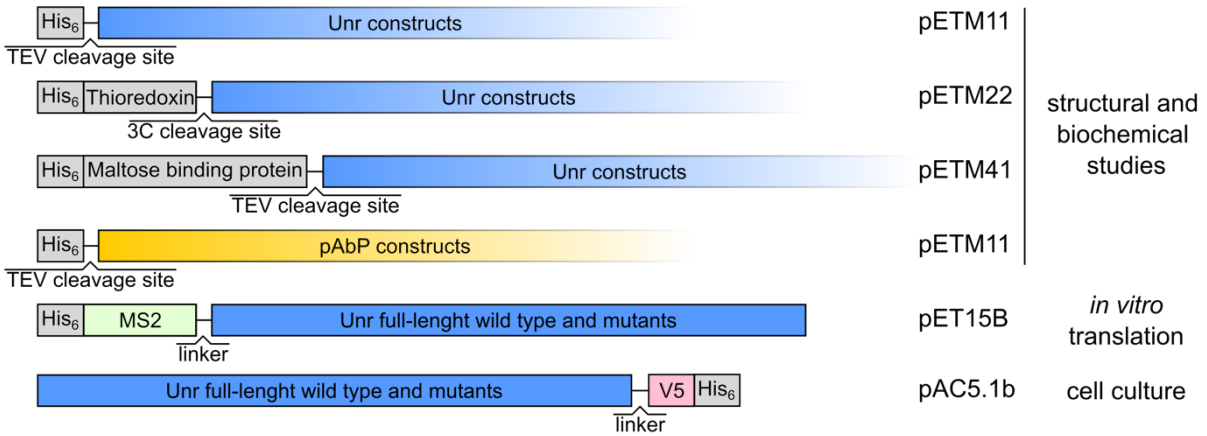


Figure 10: Overview of generated protein constructs, that were used during the study.
Different boundaries of Unr and pAbP were used for the structural and biochemical studies. Wild type and mutant versions of Unr full length were used for the *in vitro* and cell culture assays.

For the *in vitro* translation assays Unr full-length was cloned into a pET15B derived MS2 fusion vector, to express N-terminal His_6-MS2 fusion proteins. For SL2 cell culture experiments Unr

full-length was cloned into a pAc5.1B vector, which contains a C-terminal His_6 and V5 tag (Figure 10). The *msl2* promotor-constructs (msl2-FC-bGal[117] and BmutLMS2[114]), the pAc-Renilla and the pAc-Sxl plasmids were used as described before and also provided by Fátima Gebauer.

After successful cloning *E. coli* DH5 α cells were used to generate the plasmids in large scale. The cells were grown in LB medium at 37°C and harvested after 16 h. Plasmids for protein expression were extracted using alkaline lysis, followed by a column purification using the QIAprep spin Minirep Kit according to the manufacturer's protocol[174]. Plasmids for SL2 cell culture were purified using the EndoFree Plasmid Maxi Kit. The final plasmid concentration was assessed by a Nanodrop 2000 with the principle of the Beer-Lambert law ($E_\lambda = \varepsilon_\lambda$ c d)[175].

3.2.2. Protein expression and purification

E. coli BL21 (DE3) cells were used to express the different recombinant proteins. The cells were grown at 37°C. For NMR spectroscopy, expression was conducted in isotope labelled M9 minimal medium ($^{15}NH_4Cl$ and/or ^{13}C-glucose as sole nitrogen and/or carbon source) in H_2O or D_2O. Proteins that were used for other purposes than NMR were expressed in TB medium. After the cultures grew to an OD_{600} of 0.8 for minimal and 1.2 for TB media protein expression was induced using 0.2 mM IPTG, that represses T7 RNA polymerase inhibition and thereby promotes transcription of the protein of interest[176]. The cultures were left at 17°C shaking at 180 rpm overnight.

For the expression of the ^{13}C-isoleucine, leucine, valine-methyl group labeled samples, the cells were grown in minimal medium with fresh D_2O, containing deuterated Glucose and $^{15}NH_4Cl$ as sole carbon and nitrogen sources[177]. One hour prior to induction 120 mg of 2-keto-3,3-d_2-1,2,3,4-^{13}C-butyrate and 200 mg of 2-keto-3-methyl-d_3-d_1-1,2,3,4-^{13}C-butyrate per liter medium were added to the growth medium. After induction (0.5 mM IPTG) the growth was continued overnight at 17°C.

After expression of the proteins, the harvested cells (15 min at 4500 rpm) were resuspended in lysis buffer supplemented with 1 µg/ml DNaseI and 0.2 mg/ml lysozyme, incubated at room temperature for 30 min while continuously shaking and then lysed using a French press. The filtered cleared lysate (15 min at 18000 rpm and 4°C) was applied to a 5 ml Nickel-nitrilotriacetic acid (Ni-NTA) gravity flow column and after washing with 10 column volumes (CVs) lysis buffer, the protein was eluted by increasing the imidazole concentration to 500 mM. Except for MS2 tagged full-length Unr, all proteins were cleaved with

TEV-protease (1/100 w/w) and dialyzed overnight at 4°C against 10 mM imidazole and 150 mM NaCl using dialysis tubes with cut-offs between 3.5-10 kDa. After passing through a second Ni-NTA gravity flow column, all constructs, that included CSD1 were injected onto a 5ml FF Heparin column and eluted with a 2 M salt buffer to remove unspecifically bound bacterial RNAs. In a last step all proteins (except the His_6 tagged full-length Unr constructs, CSD1-6 and CSD4-9, which were only buffer exchanged on a HiPrep 26/10 Desalting column) were purified and buffer exchanged via size-exclusion chromatography on a S75 or S200 gel-filtration column and concentrated to desired concentrations using Amicon Ultra Centrifugal Filters with appropriate molecular weight cut-offs.

For NMR measurements the final sample buffer used already for size exclusion chromatography contained 20 mM NaP (pH 6.5), 50 mM NaCl, 1 mM DTT (10 mM for CSD12), 10% D_2O and 0.01% NaN_3. An exception was the ILV labeled samples. To increase protein stability of the longer protein constructs and measure all samples with the same salt concentrations, 150 mM NaCl were used. Proteins used for crystallization, SAXS, FP assays and EMSA were prepared using 20 mM Hepes/NaOH pH 7.5, 50 mM NaCl, 1 mM DTT and 0.01% NaN_3. The MS2 tagged full-length Unr constructs, that were used for the *in vitro* translation assays were after the first Ni-NTA gravity flow column further purified using Anti-FLAG M2 magnetic beads according to the manufactures protocol, eluted using a FLAG peptide and finally dialyzed against a buffer containing 20 mM Hepes/NaOH pH 7.4, 20% Glycerol, 1 mM DTT, 0.01% NP-40 and 0.2 mM EDTA.

The purity of the protein was assessed by SDS polyacrylamide gel electrophoresis, followed by coomassie staining[178]. The protein quantity was assessed by using a NanoDrop 2000 or a BCA Assay Kit, due to the lack of tryptophans, for CSD12, CSD123, CSD6, CSD8 and CSD9 according to the manufacturers' protocol based on the principle developed by Smith et al.[179].

3.2.3. Crystal structure determination

For initial crystallization trials each protein was tested using two different concentrations in sitting drops against five different screens each containing 96 different crystallization conditions at 7°C and room temperature. The drops, each composed of 0.1 µl protein sample and 0.1µl buffer solution, were observed for crystal growth regularly and potential conditions were refined later in a 24-well plate with hanging drops varying the concentration of precipitant and the pH (1 µl protein sample and 1 µl buffer solution). The initial screens were set up with a pipetting robot that was handled by Brice Murciano (Crystallization Facility; EMBL Heidelberg).

For CSD456 protein concentrations of 10 and 20 mg/ml were used for the initial screening. After refinement, crystals have grown in 0.1 M tri-sodium citrate at pH 5.5 and 20% PEG3000 at room temperature to a size of about 0.7x0.2x0.2 mm after two days without any visible macroscopic defects. Before freezing, crystals were soaked in mother liquor supplemented with 40% glycerol as a cryoprotectant. The crystals were soaked in the mother liquor containing 0.1 mM $(C_2H_5HgO)_2HPO_2$ over night at room temperature and multiwavelength anomalous diffraction (MAD) datasets were collected at the ID29 beamline of the European Synchrotron Radiation Facility (ESRF), Grenoble, France, that were used to solve the phase problem. The heavy atom-soaked crystals diffracted up to 2.2 Å and data was processed in XDS[180]. Phasing and initial automated model building was performed using AutoSol from the Phenix suite[181,182]. The resulting structure was further refined with several rounds of model building in COOT[183] and refinement in the Phenix suite. Structural statistics are listed in Supplementary Table 1.

CSD789 was crystallized in the presence of an A15-mer RNA. A previously measured NMR sample was dialyzed against 10 liter of crystallization buffer (20 mM Hepes/NaOH pH 7.5, 50 mM NaCl and 1 mM DTT) to reduce the amount of phosphate. The sample was then concentrated to 5 and 10 mg/ml. Initial crystallization screens and refinements were done as described before. Final crystals started to grow in 0.1 M Tris-Cl (pH 7.0), 0.2 M lithium sulfate and 2 M ammonium sulfate after three days. After three weeks and with a final size of 0.1x0.1x0.01 mm the platelet like crystals were frozen in mother liquor supplemented with 30 % glycerol. The diffraction was tested at the beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany) and a dataset up to 1.2 Å was collected. To solve the phase problem molecular replacement with three input structures of CSD1 with poly-alanine chains was used. Further refinement of the structure was done as described for CSD456.

Additionally, initial crystal screens were set up for CSD123, CSD12, CSD789, CSD1-9 and CSD1-9 in complex with an A50-mer RNA (protein and RNA were mixed in a 1:1 molar ratio prior to the initial screen set up). Due to the lack of crystal growth no refinements were conducted. Both crystal structures, the one of CSD456 and the one of CSD789 bound to the poly(A)-mer, were solved with the help of Pravin Kumar Ankush Jagtap (Postdoctoral Fellow; Hennig group).

3.2.4. NMR spectroscopy

All NMR measurements were performed on Bruker Avance III NMR spectrometers with magnetic field strengths corresponding to proton Larmor frequencies of 600 MHz, 700 MHz or

800 MHz. The spectrometers were equipped with a cryogenic triple resonance gradient probe head (600 and 800 MHz), a room temperature triple resonance probe head (700 MHz) or a room temperature quadrupole resonance probe head (600 MHz).

NMR sample concentrations for acquiring spectra for backbone and side chain assignments, as well as 3D-NOESY-type experiments were between 0.2-0.75 mM, depending on protein yield, solubility and stability. To increase sample stability and spectral quality CSD12 was measured in the presence of 1.2 molar excess of an RNA (AAA AAA AUG or AAU ACA). Experiments for backbone assignments have been performed on ^{13}C,^{15}N-labeled samples (using 70% D_2O in growth medium for CSD456 and CSD789) using conventional triple-resonance experiments (HNCO, HNCA, CBCA(CO)NH, HN(CO)CA and HNCACB)[152]. Side chain assignments were done using HBHA(CO)NH, HCCH-TOCSY, and CCH-TOCSY spectra. 3D ^{13}C-NOESY-HMQC and ^{15}N-NOESY-HSQC spectra with 100 ms mixing times and 3D (H)CCH HMQC-NOESY-HMQC, 3D HCH NOESY-HMQC and (H)CNH HMQC-NOESY-HSQC spectra with 70 ms mixing times were used for side chain and NOE assignments to derive distance restraints. 3D HMQC-based spectra were recorded in D_2O after sample lyophilization with a decoupling scheme as described by Schilling and colleagues[184]. All spectra were acquired using the apodization weighted sampling scheme[185] and processed using NMRPipe[186]. Resonance assignments for backbone, side chain and NOEs were done with the program Cara[187].

NOE-based structure calculation was done using CYANA 3.98[188] and dihedral angle restraints were derived from backbone chemical shifts, using TALOS[154]. A final water refinement was done using ARIA 1.2[189,190]. PROCHECK and WHATCHECK[191,192] were used for structure validation of the final ensemble of 20 structures with lowest energies. The structural statistics are listed in Supplementary Table S2.

For RNA titrations, a protein concentration of 0.1 mM was used for CSD6, CSD8, CSD9 and CSD78. CSD12, CSD456 and CSD789 were titrated at a concentration of 0.2 mM. The ^{15}N labeled proteins were titrated with various ratios of purchased RNA oligonucleotides. For some titrations of CSD789 with RNA a deuterated ^{15}N labeled sample was used to increase the spectral quality. ^{1}H,^{15}N-HSQCs were recorded for each titration point. An RNA stock solution concentration of 10 mM ensured a negligible dilution effect of the protein sample. The titration data was analyzed using Sparky[193] and chemical shift perturbations δ (ppm) were calculated according to: $\delta(\text{ppm})=\sqrt{(\Delta H)^2 + (0.2 * \Delta N)^2}$[86]. CCPNMR analysis software was used to determine the dissociation constants[194]. The chemical shift perturbations vs. the RNA

concentration of residues which shift at a protein:RNA ratio of 1:2 more than the average plus the standard deviation of all measured shifts were fitted using $A(B + x - \sqrt{((B + x)^2 + 4x)})$ as a function, where x is derived from the CSP value of both dimensions and A and B are estimated values, that are obtained by fitting the selected equation to the chemical shift data[194].

To measure the NMR relaxation parameters, R_1, R_2 and ^{1}H-^{15}N heteronuclear NOE experiments standard pulse sequences were used[158,195]. Relaxation delays for R_2 and R_1 were chosen dependent on the size of the protein (CSD78, R_1: relaxation delays of 1600, 20, 1300, 50, 800, 100, 500, 250, 650, 150, 1000, 400, 50 and 500 ms, R_2: 16, 128, 192, 48, 80, 160, 32, 112, 64, 96, 144, 16, 80 and 160 ms, CSD12, R_1: 2000, 50, 100, 700, 300, 400, 200, 1000, 150, 500, 1600 and 50 ms, R_2:16, 132, 64, 32, 50, 100, 116, 166, 200, 16, 132 and 64 ms, CSD789 with and without A15-mer RNA, R_1: relaxation delays of 1600, 20, 50, 800, 100, 500, 150, 650, 1000, 400, 150 and 20 ms, R_2: 25, 12.5, 50, 62.5, 100, 37.5, 75 and 25 ms, CSD678, R_1: relaxation delays of 1600, 20, 1300, 50, 800, 100, 500, 250, 650, 150, 1000, 400, 150 and 20 ms, R_2: 14.3, 114.4, 14.3, 42.9, 71.5, 28.6, 100.1, 57.2, 85.8 and 57.2 ms). The peak integration and anaylsis to derive spin relaxation parameters from which the rotational correlation time (τ_c) was calculated for each construct was done using PINT[196,197].

3.2.5. Small angle X-ray scattering data acquisition and analysis

Small-angle X-ray scattering (SAXS) data were collected at the BioSAXS beamline BM29[198] at the ESRF, Grenoble, using an X-ray wavelength of 0.992 Å and at the P12, operated by EMBL Hamburg at the PETRA III storage ring (DESY Hamburg, Germany)[199] using an X-ray wavelength of 1.24 Å. For the measurements 30 µl of protein sample or buffer were purged through a quartz capillary. Data acquisition details and statistics for each protein sample are listed in the Supplementary Table 3 according to community guidelines[200]. Each individual frame was checked for radiation damage and all frames without damage were merged. The buffer was measured before and after each sample and its contribution was subtracted from the merged data sets of the protein samples. A guinier analysis was carried out to assess data quality and to estimate the radius of gyration. The pair-wise distribution function was calculated by the indirect Fourier transform using GNOM[201]. The data were analyzed using the data analysis software package ATSAS 2.7.1[171]. EOM and CRYSOL calculations were done using the default settings in order to derive theoretical scattering curves from PDB files and fit them to experimental data[202,203].

3.2.6. Structure modeling of CSD1-9

All high-resolution structures from the single CSDs were taken from the calculated X-ray and NMR structures. CSD3 as the missing domain structure was modeled using swiss-model[204–208], where the sequence of *Drosophila* CSD3 and the NMR solution structure of human CSD3[127] were given as templates.

The script used to generate the modeled structures was generated by Bernd Simon (NMR facility; EMBL Heidelberg). Structures of CSD1-9 were calculated using CNS (1.2)[209,210] in an ARIA framework[189,190]. Structures were generated as described previously[211]. First the single domains were connected to a single molecule by the linker connecting residues. The parts for which we had interdomain contact information were kept rigid (NMR and crystal structures) and all other regions were randomized during structure calculations. 5000 structures were calculated and the corresponding regions within the structure were fitted against SAXS curves of CSD1-6 and CSD4-9. The two structures with the lowest χ for both curves and a chimera structure from the single top hits were selected for the final ensemble.

3.2.7. RNA synthesis and purification

SL67 (5'-ACAAUAUGCAAUACAAUACAAUACAAGACAAAAAAAUGUGUCUUGGA ACCAACAUUGUACAAGUCGCAAUGCAAACUGAAGUCUUAAAAGACGUGUAAAA UGUUGCAAAUUAAGCAAAUAUAUAUGCAUAUAUGGGUAACGUUUUACGCGCC UUAACCAGU-3'), SL678 (5'- ACAAUAUGCAAUACAAUACAAUACAAGACAAAAA AAUGUGUCUUGGAACGCAACAUUGUACAAGUCGCAAUGCAAACUGAAGUCUUA AAAGACGUGUAAAAUGUUGCAAAUUAAGCAAAUAUAUAUGCAUAUAUGGGUA ACGUUUUACGCGCCUUAACCAGUCAAAAUACAAAAUAAAUUGGUAAAUUUCAU AUAACUAGUGAAAUGUUAUACGAAACUUAACAAUUGCCAAAUAA-3') and the A50-mer (5'-ACAAACCCAAAACAAACCAACAAAAACAAAAAAACCAAAAAAAAC AAAAA-3') RNAs that were used for the EMSAs and crystallization were prepared by T7 *in vitro* transcription using unlabeled rNTPs and a template was cloned into pUC19 plasmid DNA and contained a hammerhead ribozyme (HH) cleavage site (in cis) at the 5′ end and a Varkud satellite (VS) ribozyme recognition sequence at the 3′ end (for cleavage in trans)[212–216]. For *in vitro* transcription the vectors were amplified and purified using a QIAGEN Maxi DNA Prep Kit according to manufacturer's protocol and linearized by PstI. 0.1 mg/ml vector, 40 mM Tris (pH 8.0), 50 mM $MgCl_2$, 10 mM spermidine, 15 mM DTT, 0.01% Triton X-100, 4 U/ml thermostable inorganic pyrophosphatase, 0.1 mg/ml T7 polymerase and 10 mM of each rNTP were incubated at 37°C for 5 h before 50 µg/ml VS ribozyme was added. The reaction was left

at 37°C overnight to ensure a successful cleavage. Afterwards the RNA was extracted using phenol-chloroform treatment[217]. The RNA was further purified by a preparative gel electrophoresis using a denaturing polyacrylamide gel (12% polyacrylamide and 6M UREA) and eluted by electroelution of the single isolated band. To remove oligomers or misfolded RNAs structured sequences were refolded prior to usage by heating the RNA to 95°C for five minutes followed by snap-cooling on ice. The RNA quality was measured using the Nanodrop 2000.

3.2.8. Fluorescence labeling of RNA

The SL67 RNA that was used for EMSAs was labeled with Cyanine 5 at the 3' end by ligating 100 pmol RNA with 100 pmol pCp-Cy5 and 20 U T4 RNA ligase in 20 µl volume (50 mM Tris-HCl (pH 7.5 at 25°C), 10 mM $MgCl_2$, 1 mM DTT and 0.5 mM ATP) and incubation over night at 4°C. At the next day the RNA was purified using Sephadex G_{50} columns to separate the labeled RNA from unincorporated dyes, enzymes and single nucleotides. The labeling efficiency, sample quantity was assessed using the Nanodrop 2000.

For fluorescence polarization assays short RNA oligonucleotides (AAA AAA AUG and an A15mer) were labeled at the 3' end with fluorescein-5-thiosemicarbazide by using 0.25 nmol of the short RNAs in 0.25 M sodium acetate (pH 5.6), which were oxidized with 2.5 nmol sodium periodate at 25°C in the dark for 90 minutes. After oxidization 5 nmol of sodium sulfite were added to the mixture followed by 15 min incubation at 25°C. For labeling 7.5 nmol of fluorescein-5-thiosemicarbazide were added and the reaction was performed for 3h at 37°C[218]. The labeled RNA was precipitated using one tenth of the reaction volume of 8 M LiCl and 2.5 times the reaction volume of 100% ethanol incubating for at least 3h at -80°C. Finally, the RNA was washed twice with 75% ethanol, before the labeling efficiency was assessed and concentration measured using the Nanodrop 2000.

3.2.9. Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSA) were used to assess semi-quantitatively protein-RNA affinities, based on the observation that protein-nucleic acid complexes typically migrate less than the free nucleic acid under native conditions in an electric field[219,220]. All RNA-binding reactions for the electrophoretic mobility shift assays were performed in a binding buffer containing 20 mM Hepes/NaOH pH 7.5, 50 mM NaCl, 10% [v/v] glycerol and 2mM DTT with a final volume of 10 µl. The reactions were equilibrated for 30 min at 20°C with an RNA concentration of 25 nM Cy5 labeled probe and varying protein concentrations.

Next the samples were resolved on a 6% native polyacrylamide gel in 0.5x TBE. After successful separation the gels were imaged with a Typhoon Trio imager 9000 adjusting the laser intensity for each gel.

3.2.10. Fluorescence polarization assay

Binding constants of protein-RNA interactions can be assessed by using fluorescence polarization due to the changes in the rotational correlation time between the free and the bound state. The unbound molecule labeled with a fluorophore tumbles quicker in solution than the bound complex. If the fluorophore gets excited with linearly polarized light, the changes in polarization in the emitted light can be measured and the differences between the two states recognized[221].

All RNA-binding reactions for the fluorescence polarization assays were performed in a binding buffer containing 20 mM Hepes/NaOH pH 7.5, 50 mM NaCl and 2 mM DTT in a final volume of 25 µl. 5 nM FAM labeled SL6a RNA (AAA AAA AUG) and 25 nM FAM labeled A15-mer RNA were incubated with different concentrations of protein for 30 min at 20°C. Each reaction was measured in a technical duplicate or triplicate. Finally, the fluorescence polarization was measured for each reaction using a black 384-well plate and a plate reader with a monochromator using the corresponding filters and automatic gain function.

3.2.11. Protein melting temperature

The protein melting temperatures were determined using the nano differential scanning fluorimetry (nanoDSF) technology, which measures the intrinsic tryptophan fluorescence of the sample. The direct surrounding of the amino acid impacts its fluorescence intensity and maximum, which makes changes for example due to protein unfolding detectable[222]. Proteins were soaked into a standard capillary and heated up 1°C/min. Depending on the protein concentration the excitation varied from 10-30%. The data analysis was done with the provided software. The temperature at which 50 % of the protein is unfolded, visible in changes of the fluorescence intensity ratio at 330 and 350 nm, was taken as melting temperature.

3.2.12. Circular dichroism

Circular dichroism (CD) was used to assess the secondary structure profile of proteins. Each secondary structure element has a representative CD spectrum, so that the amount of each element can be extracted from the spectrum of the measured protein[223]. The samples were dialyzed into a buffer containing 20mM Hepes/NaOH pH 7.5, 20 mM NaCl and 1 mM DTT.

The measurements were done at 10 μM concentration in a 0.2 mm cuvette at 20°C, using a Jasco J-815 CD spectrometer. The wavelength range was 240 to 190 nm, measured with 0.1 nm steps, and averaged over 5 points per wavelength. Analysis was done using SELCON3 to calculate the secondary structure content[224,225].

3.2.13. *In vitro* translation assay

The mRNAs used for the *in vitro* translation assay were *in vitro* transcribed from a linearized vector using a T3 polymerase[226]. All mRNAs contained a 5' M7GpppG cap and a poly(A) tail of 73 nucleotides. After *in vitro* transcription all RNAs were purified using G_{50} desalting columns and a phenol/chloroform extraction[217]. The homogeneity of the RNA samples was assessed by 1% agarose gels.

The *in vitro* translation reactions in overnight *Drosophila* embryo extracts were performed in a final volume of 12.5 μl, as described previously[226], with a final concentration of 60 μM amino acids, 0.6 mM DTT, 24 mM Hepes/KOH pH7.4, 0.26 mM $Mg(OAc)_2$, 48 mM KOAc, 16.8 mM creatin phosphate, 80 ng/μl creatin kinase, 0.4 ng/μl *Renilla* mRNA and 1.6 ng/μl *BmutL-MS2* mRNA. The overnight *Drosophila* extract was provided by Fátima Gebauer (CRG). Increasing amounts of wild type and mutant full-length *Drosophila*-MS2 tagged Unr were added prior to incubation.

The translation efficiency was measured using the Dual Luciferase Assay System according to manufacturer's protocol. The Renilla values were used as an internal control to correct the Firefly expression. To ensure the observed signal differences are not due to different mRNA levels within the reaction, a RT-qPCR was performed after each reaction.

3.2.14. *Real time quantitative PCR*

Real time quantitative PCR (RT-qPCR) was performed using SYBR Green on an Applied Biosystems 7000. The used primers are listed in Table 3.10. RNA was extracted using Trizol reagent and the DNA was digested using the Turbo DNA free kit according to manufacturer's protocol. Reverse transcription was done using the SuperScript II reverse transcriptase according to manufacturer's protocol. Per reaction 250-750 ng RNA were used as input material. For the final PCR reaction, the cDNA samples were diluted 1:6 and measured with 400 nM forward and reverse primer each in the presence of SYBR Green.

3.2.15. RNAi, transfection and reporter gene assay

SL2 cells were kept in culture at 25°C in Schneider's Medium with penicillin/streptomycin (1% v/v) and 10% FBS. RNAi was performed in 6 well dishes as described earlier[113]. The target sequence to neutralize the gene expression was designed in the 3' UTR of the endogenous Unr sequence to avoid a knock-down of the transfected gene. $2x10^6$ SL2 cells per well were pelleted and resuspended in 1 ml Schneider's medium without FBS. 15 µg/ml of dsRNA against the 3'UTR of the endogenous Unr or GFP as a control were added directly to each well. After 40 min with shaking the plates every 10 minutes, 1 ml of 20% supplemented FBS Schneider's medium was added. The Megascript T7 kit was used for *in vitro* transcription of the dsRNAs that were used for the knockdown. Amplified DNA strands functioned as a template. Oligonucleotides used to amplify the dsRNAs are listed in Table 3.10.

Three days after the knockdown the cells were transfected based on the principles of lipofection with Effectene according to the recommendations in the manufacturer's manual[227]. 2 ng of pAc-Sxl, 75 ng of the reporter gene construct, 100 ng of pAc-V5-Unr, 10 ng of pAc-Renilla and 163 ng of an empty pAc vector from endotoxin free isolated DNAs were transfected per well.

The cells were harvested two days after the transfection by scraping of and washing with PBS. β-galactosidase activity, which activates a chemiluminescent substrate based on 1.2-dioetane, was measured with Galacto-Star and Renilla luciferase activity, which is based on the catalyzation of coelenterazine oxidation to produce light, with Renilla substrate[228–230]. Both assays were done according to recommendations from two fifth of the total cell lysate. The luminescence activities were normalized against mRNA levels of β-gal and RLuc obtained by RT-qPCR as described above using one fifth of the input material (3.2.14). A western blot from two fifth of the cells was done to assess the quality of the knock-down and the transfection efficiency of Unr.

3.2.16. Western blot

The cells were lysed in RIPA buffer for 20 minutes on ice for the western blot. The concentration of the cleared lysate was measured by a BCA assay according to the manufacturer's protocol[179]. 15 µg total protein was loaded on a 4-20% Mini-Protean TGX precast protein gel, which was blotted on a nitrocellulose membrane using a semi dry Trans-Blot Turbo System. After blocking the membrane in 5% milk in PBS-T for 2h at room temperature, the primary antibodies were added and incubated overnight in the cold room under

agitation. For detection of Unr, a polyclonal antibody serum against amino acids 1-156 of Unr (1:2,000) and a monoclonal anti-V5 antibody (1:1,000) were used. Tubulin was used as a loading control and detected by a monoclonal anti-tubulin antibody (1:2,000). After washing of the membrane with PBS-T, the blots were incubated with anti-rabbit and anti-mouse secondary HRP-linked antibodies (1:10,000) and developed using an ECL substrate.

3.2.17. Unr immunoprecipitation

For Unr immunoprecipitation, endogenous Unr was knocked down to liberate rate-limiting targets and the SL2 cells were transfected with 2 µg pAc-Unr wild type and mutants (456 ID and 456-78 ID) as described before (3.2.16) in a 10 cm dish per reaction. To stabilize transient and weak RNA-protein interactions, UV crosslinking was done three days afterwards in a thin layer of ice-cold PBS at 300 mJ/cm^3. For cell lysis, the samples were sonicated in 20 mM Hepes/NaOH pH 7.5, 100mM NaCl, 1mM $MgCl_2$, 0.05% NP-40 and 40U/ml RNaseOUT for 3 cycles of 30 sec at a low energy level using a Bioruptor. Afterwards the cleared lysate of each sample was incubated with 1 µg of V5 antibody per 3 mg of total protein and incubated at 4°C for 2h on a rotating wheel. The total protein concentration was determined earlier by a BCA assay. After incubation, 4 µl of magnetic protein A/G beads per 1 µg of used antibody was added and incubated for another 10 minutes at 4°C on a rotating wheel. To get rid of unspecific binding, the samples were washed with 1 ml of 20 mM Hepes/NaOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.5 mM DTT, 40U/ml RNaseOUT and 1% Triton-X for three times, before the beads were resuspended in the final buffer and volume. In case of the RNase treated samples, 250 U of benzonase were added during washing steps. For the RIP-Seq experiments samples were resuspended in 125 µl of 20 mM Hepes/NaOH pH 7.5 and 150 mM NaCl. The samples for the IP-MS were resuspended in 30 µl 20 mM Hepes/NaOH pH 7.5, 150 mM NaCl and 10% SDS.

3.2.18. RNA sequencing and data analysis

To digest proteins and only extract bound RNAs, the immunoprecipitated samples (3.2.18) were incubated with 0.2 mg/ml Proteinase K for 30 min at 55°C. Afterwards the RNA was extracted from 400 µl Trizol per reaction according to recommendations. Finally, the RNA was resuspended in 10 µl water and the ribosomal RNA was depleted using an approach of fishing for ribosomal RNA with biotinylated oligonucleotides[231,232]. The successful RNA depletion was checked afterwards on a Bioanalyzer pico RNA chip. The following preparation steps of

the RNA library were done by Laura Villacorte and the data analysis by Jan Provaznik from the Genecore facility (Genomics Core Facility; EMBL Heidelberg).

In brief, a barcoded single stranded cDNA library was generated using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. Obtained libraries that passed the QC step were pooled in equimolar amounts. 1.9 pMol solution of this pool was loaded on the Illumina sequencer NextSeq 500 and sequenced uni-directionally, generating ~500 million reads, each 85 bases long.

The alignment of the sequencing reads was done using STAR aligner version 2.7.1a to a genome reference of *Drosophila melanogaster* BDGP6.22.97 from ENSEMBL[233]. The read counts were obtained using in-built implementation of HTSeq-count in STAR aligner with the '--quantMode GeneCounts' option. Finally the analysis of triplicate samples to generate PCA plots and assess differentially expressed genes was done in R v3.5.1 using DESeq2 v1.20.0[234,235]. The EnhancedVolcano package v1.3.5 was used to generate the volcano plots[236]. For resulting figure generation ggplot2 was used[237]. The VennDiagram package v1.6.20 was used to generate the Venn diagrams[238].

3.2.19. Sample preparation and LC-MS/MS analysis of IP-MS samples

The immunoprecipitated samples of control, wild type and mutant samples in the absence and presence of RNase were incubated for 5 min at 95°C and subjected to an in-solution tryptic digest using a modified version of the Single-Pot Solid-Phase-enhanced Sample Preparation (SP3) protocol[239,240]. In total three biological replicates of each sample were prepared (n=3). To check the pull-down efficiency a TGX 4-20% gradient polyacrylamide gel was silver stained using the Pierce Silver Stain kit according to manufacturer's protocol. The following sample preparation was done by Per Haberkant (Proteomics Core Facility; EMBL Heidelberg). Cell lysates were added to Sera-Mag Beads in 10 µl 15% formic acid and 30 µl of ethanol. Binding of proteins was achieved by shaking for 15 min at room temperature. SDS was removed by four subsequent washes with 200 µl of 70% ethanol. Proteins were digested overnight at room temperature with 0.4 µg of sequencing grade modified trypsin in 40 µl Hepes/NaOH, pH 8.4 in the presence of 1.25 mM TCEP and 5 mM chloroacetamide. Beads were separated, washed with 10 µl of an aqueous solution of 2% DMSO and the combined eluates were dried down. Peptides were reconstituted in 10 µl of H_2O and reacted for 1 h at room temperature with 80 µg of TMT10plex[241] label reagent dissolved in 4 µl of acetonitrile. Excess TMT reagent was quenched by the addition of 4 µl of an aqueous 5% hydroxylamine solution. The peptides were reconstituted in 0.1 % formic acid, mixed to achieve a 1:1 ratio across all TMT-channels and

purified by a reverse phase clean-up step (96-well µElution Plate). Peptides were subjected to an off-line fractionation under high pH conditions[239]. The resulting 12 fractions were afterwards analyzed by LC-MS/MS using a 2h gradient on an Orbitrap Fusion Lumos mass spectrometer as previously described[242]. To this end, peptides were separated using an Ultimate 3000 nano RSLC system equipped with a trapping cartridge (Precolumn C18 PepMap100, 5 mm, 300 µm i.d., 5 µm, 100 Å) and an analytical column (Acclaim PepMap 100. 75 × 50 cm C18, 3 mm, 100 Å) connected to a nanospray-Flex ion source. The peptides were loaded onto the trap column at 30 µl per min using solvent A (0.1% formic acid) and eluted using a gradient from 2 to 40% Solvent B (0.1% formic acid in acetonitrile) over 2 h at 0.3 µl per min (all solvents were of LC-MS grade). The Orbitrap Fusion Lumos was operated in positive ion mode with a spray voltage of 2.4 kV and capillary temperature of 275 °C. Full scan MS spectra with a mass range of 375–1500 m/z were acquired in profile mode using a resolution of 120,000 (maximum fill) time of 50 ms or a maximum of 4 x 10^5 ions and an RF lens setting of 30%. Fragmentation was triggered for 3 s cycle time for peptide like features with charge states of 2–7 on the MS scan (data-dependent acquisition). Precursors were isolated using the quadrupole with a window of 0.7 m/z and fragmented with a normalized collision energy of 38. Fragment mass spectra were acquired in profile mode and a resolution of 30,000 in profile mode. Maximum fill time was set to 64 ms or an AGC target of 4 x 10^5 ions. The dynamic exclusion was set to 45s.

Acquired data were analyzed using IsobarQuant[243] and Mascot V2.4 (Matrix Science) using a reverse UniProt FASTA *Drosophila melanogaster* database (UP000000803)[244] including common contaminants. The following modifications were considered: Carbamidomethyl (C, fixed), TMT10plex (K, fixed), Acetyl (N-term, variable), Oxidation (M, variable) and TMT10plex (N-term, variable). The mass error tolerance for full scan MS spectra was set to 10 ppm and for MS/MS spectra to 0.02 Da. A maximum of 2 missed cleavages were allowed. A minimum of 2 unique peptides with a peptide length of at least seven amino acids and a false discovery rate below 0.01 were required on the peptide and protein level[245].

3.2.20. Data analysis of mass spectrometry experiments

The following data analysis of the mass spectrometry experiments was done by Frank Stein (Proteomics Core Facility; EMBL Heidelberg). In brief, the protein.txt output files of IsobarQuant[243] were processed with the R v3.5.1[235]. Only proteins that have been identified in two out of three replicates were kept. The ‘signal_sum’ columns were cleaned for batch effects using the removeBatchEffect function of the limma package[246]. The data were normalized with

the vsn package[247]. A separate normalization was applied for control conditions, normal pull-down conditions and RNAse treated pull-down conditions. Potential missing values were imputed with the impute function of the Msnbase package[248]. Limma was used to test for differential abundance. Within this analysis, imputed values were given a weight of 5 %. When testing for differential abundance between conditions of different normalization groups (e.g. normal pull-down vs control condition), adjusted p-values from limma output were used as the false discovery rate (fdr). For all other tests, t-values from the limma output were used as an input to the fdrtool function of fdrtool[249] in order to calculate the fdr (qvalues were used). Proteins were classified as 'hit' with an fdr smaller 5 % and a fold-change of at least 100 % and as 'candidate' with an fdr smaller 20 % and a fold-change of at least 50 %. Hit and candidate proteins (tests: mutant vs WT, mutant_RNAse vs WT_RNAse, mutant vs WT / mutant_RNAse vs WT_RNAse) were clustered based on the Euclidean distance between normalized tmt reporter io signals (signal_sums) normalized by the WT or WT_RNAse condition using the kmeans algorithm[250,251].

3.2.21. Sequence alignment and HMMER prediction

The clustal omega tool was used to align the different sequences[252]. The alignments were graphically modified using ESPript[253]. A sequence alignment only from the non-canonical CSDs was used as an input for the hidden markov model-based search, which run on the HMMER webserver[254].

3.2.22. Data presentation

Graphs were plotted using either Gnuplot 4 or Prism 5. Structure representations were done using PyMOL 2.3.2. Structures were superimposed using either the align algorithm for molecules with sequence identity, or the super algorithm for proteins, that differ in their sequence. The surface potential of the different proteins was calculated with the APBS electrostatics plugin of PyMOL. The figures were generated using Inkscape version 0.92.3.

4. Results

4.1. Discovery of novel, non-canonical cold-shock domains in Unr and other CSD-containing proteins

To gain structural insights into *Drosophila* Unr, protein constructs with a high yield of soluble and stable protein were needed to fulfill the requirements of structural biology techniques like NMR and X-ray crystallography. Therefore, we employed a 'divide and conquer' approach and tested 117 different protein boundaries (Figure 11a; for a detailed list see Hollmann et al.[255]), which were designed based on the five predicted CSDs and their N- and C-terminal extensions, that were chosen based on secondary structure predictions (JPed4d[256]). Surprisingly most of the constructs needed to be extended beyond the boundaries of the predicted CSDs, to obtain soluble and stable protein samples.

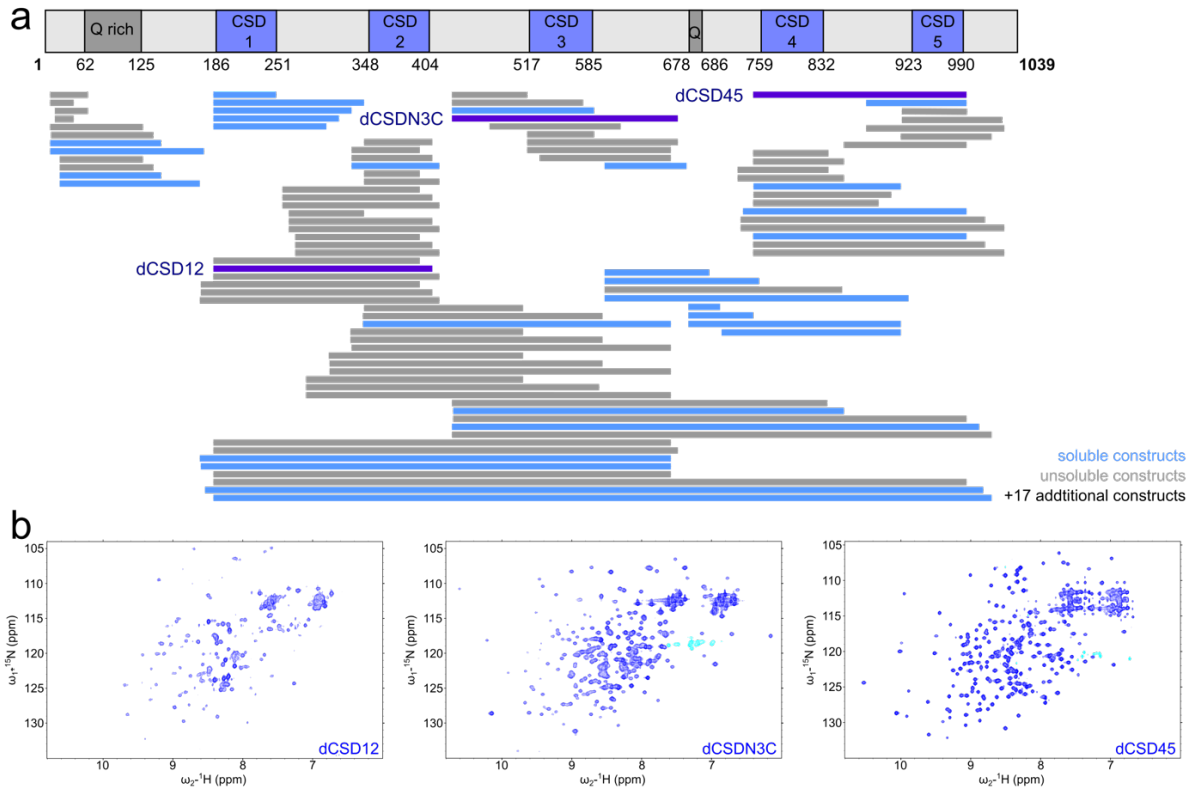


Figure 11: Construct screening of *Drosophila* Unr.
a: Hitherto annotated protein domain arrangement of *Drosophila* Unr showing the five cold-shock domains (CSD1-5) evenly distributed among the protein and two Q-rich domains, from which one is in the N-terminus and the other one in the middle of the protein. The bars underneath the scheme indicate different construct boundaries that were tested for solubility (blue: soluble; grey: unsoluble). **b:** ^{1}H,^{15}N-HSQC of three different protein constructs (dCSD12: 178-414, dCSDN3C: 424-677 and dCSD45: 756-990) (purple in a).

Three soluble constructs, which cover almost the whole region of full length Unr excluding the two Q-rich domains (Figure 11a, purple bars; CSD12: aa 178-414; CSDN3C: aa 424-677; CSD45: 756-990) were structurally assessed by $^{1}H,^{15}N$- HSQC spectra (Figure 11b). The broad dispersion of peaks in all three spectra indicates that these proteins are highly structured and do not possess larger disordered regions as has been suggested and could be assumed from the long unannotated regions in between predicted CSDs. The construct CSD12 showed aggregation during the purification and in the NMR measurements at high concentrations, visible in the low peak intensity. Therefore, this construct was exchanged with one that only covers CSD1 and its C-terminal extension (aa 179-344) to be used in assays and experiments described below.

Due to the apparent lack of unstructured regions within these constructs, structure determination efforts were pursued using X-ray crystallography or NMR. I was able to solve the crystal structure of CSDN3C using multiwavelength anomalous diffraction (MAD) scattering. Surprisingly, two additional domains are located in the N- and C-terminal extensions of the predicted CSD3 (Figure 12a; upper left box). The fold of these domains is highly similar to CSDs, harboring a β-barrel which is composed out of five antiparallel β-strands (Figure 12b). Similar observations were made for the other two constructs (CSD1+C-terminal extension: aa 179-344 and CSD4+C-terminal extension: aa 756-922), of which NMR structures were solved as no crystals could be obtained. The C-terminal regions of CSD1 and CSD4 show a similar additional domain (Figure 12a; Supplementary Figure 2a, 2b, 2c and 2d). An NMR structure of an N-terminally extended CSD5 construct was solved primarily by Lara Jayne Sweetapple, a master student under my supervision. This structure shows the expected CSD fold with an unstructured N-terminal extension. However, this extension would not be long enough to cover the entire additional domain discovered in the NMR structure of CSD4+C-terminal extension (Figure 12a).

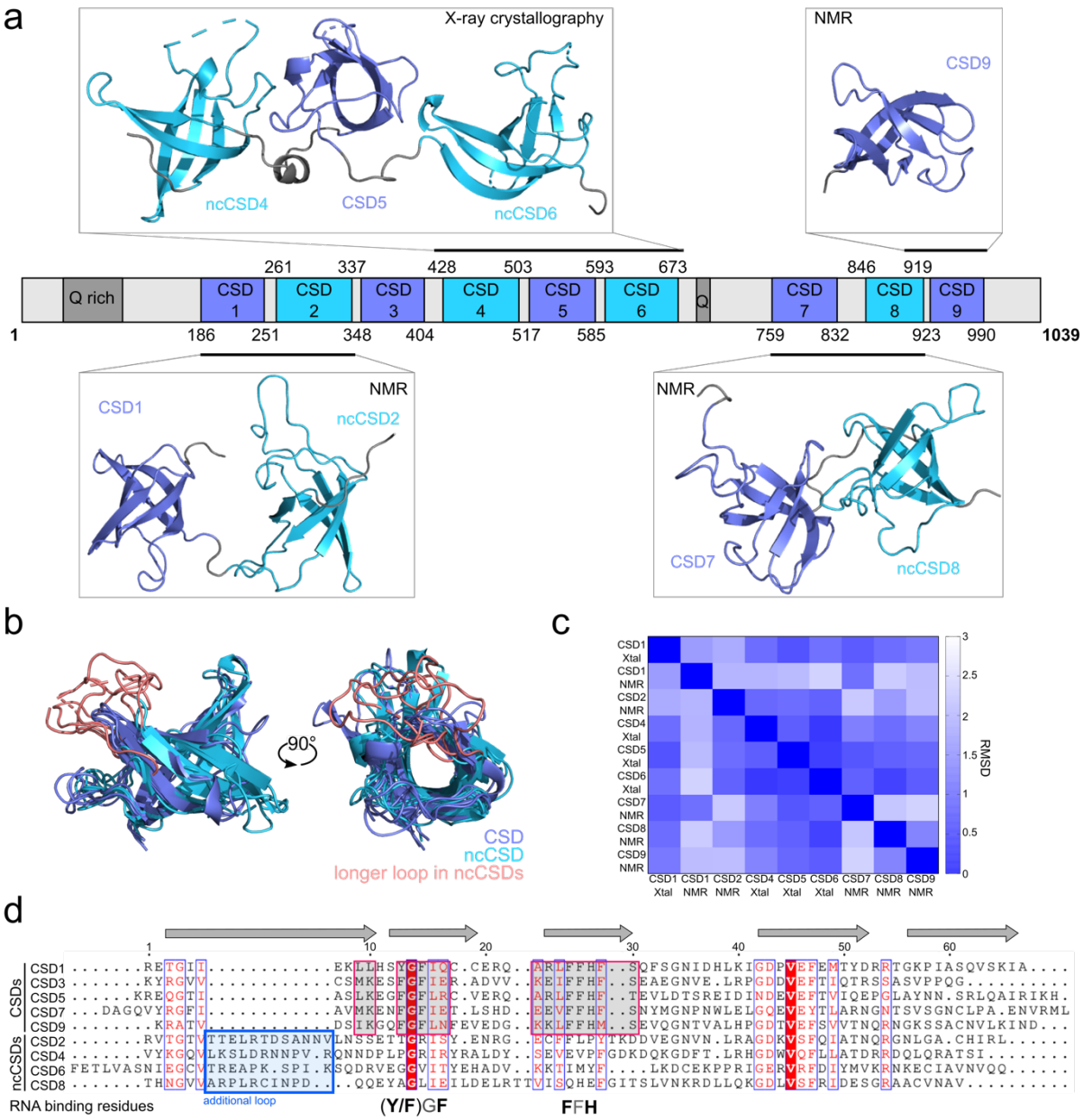


Figure 12: Novel non-canonical CSDs within *Drosophila* Unr.

a: An updated domain annotation scheme of *Drosophila* Unr showing the additional non-canonical CSDs discovered in this study (blue; 2, 4, 6 and 8) in spacer regions between the canonical ones (cyan; 1, 3, 5, 7 and 9) (middle). Above and below are solved structures indicating the similar fold of the two domain types using the same color code (X-ray crystallography structure of CSD456 (aa 424-677; PDB: 6Y6E) and NMR structures of CSD12 (aa 179-344; PDB: 6Y6M), CSD78 (aa: 756-922; PBD: 6Y4H) and CSD9 (aa 899-989; PDB: 6Y96)). The black bars indicate the protein boundaries of each structure. For clarity only a single representative NMR structure is shown. The full lowest-energy NMR ensembles are shown in Supplementary Figure 1b-e. **b**: Superimposition of canonical CSDs 1, 5, 7 and 9 (blue) with ncCSDs 2, 4, 6 and 8 (cyan). The additional loop in the ncCSDs is highlighted in pink. **c**: A matrix of RMSD values from superimpositions of each single CSD of *Drosophila* Unr. For the comparison with CSD1 the crystal structure from PDB entry 4QQB was used. **d:** Sequence alignment of *Drosophila* canonical and ncCSDs. The same or similar residues between all domains are colored, similar residue regions between the canonical CSDs are highlighted by red boxes, which align with the two RNA binding regions (**F**G**F** and (**F/Y**)F**H**). Additional loops only present in ncCSDs are highlighted by a blue box. The alignment has been done using Emboss Needle[257] and ESPript [258] has been used for illustration.

A sequence alignment between previously predicted CSDs and domains discovered in this study shows a low sequence similarity despite the high structural similarity (backbone RMSD between 0.7-2.5 Å; Figure 12c). The average sequence similarity between CSD1 to the other predicted CSDs is 31.5 ± 4.8%, but only 16.5 ± 2.1% to the domains discovered here. A striking difference is the low conservation of the RNA binding regions (**F**G**F** and (**F/Y**)F**H**) present in the predicted, canonical CSDs, but missing in the novel domains. Instead they feature an extended additional loop between β-strands β1 and β2 (Figure 12b and 12d). Due to the structural convergence and the sequence divergence these additional domains were called non-canonical cold-shock domains (ncCSDs). The position of the ncCSDs led to a renumbering of the domains, with the ncCSDs having the numbers 2, 4, 6 and 8, shifting the canonical ones to the odd numbers (1, 3, 5, 7 and 9) (Figure 12a).

The Hidden Markov model used to predict CSDs previously was trained based on the RNA binding motifs (Pfam: PF00313[259]), explaining why these novel domains passed unnoticed. To assess whether ncCSDs are a peculiarity of *Drosophila* Unr or do also exist in other proteins, a Hidden Markov model search using HMMer[254] based on a sequence alignment of the ncCSDs and their conserved hydrophobic core residues was performed against the sequence database of UniProtKB[244] (Figure 12d and 13b). Surprisingly, 1036 additional proteins besides *Drosophila* Unr seem to contain these ncCSDs (Figure 13a). Although many hits are related to Unr, other proteins, like the Cold-shock DNA binding domain protein of *Clostridium sp. CAP:1000* and the R3H domain-containing protein of *Phytophthora ramor* are not. Strikingly, almost all predicted proteins that encode at least for one CSD also feature an ncCSD, which is located close to the canonical or in case of multiple CSDs all distributed along the protein sequence in an alternating fashion.

Concomitantly, the presence of ncCSDs was further confirmed in human Unr/CSDE1 (hUnr) using $^{1}H,^{15}N$-HSQC spectra. Similar boundaries resulted in spectra with peak dispersions comparable to spectra of the *Drosophila* protein constructs, indicating a low amount of unstructured regions also within the human constructs (Figure 13c). Furthermore, a sequence alinment of the single canonical and ncCSDs of hUnr also shows a potential additional loop in the ncCSDs as observed for the *Drosophila* orthologue. Additional studies will show, whether this might be a conserved structural element amongst all ncCSDs (Figure 13d).

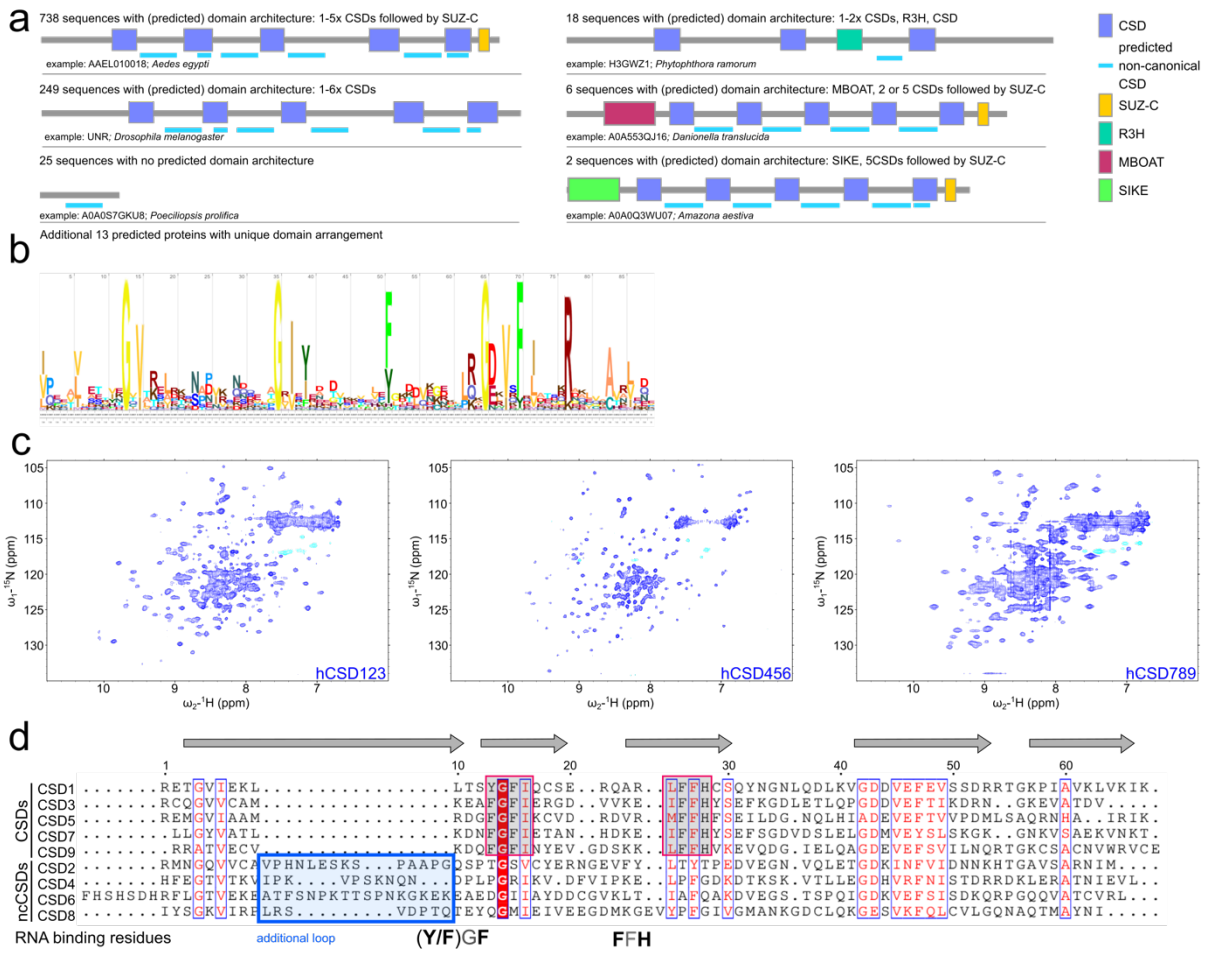


Figure 13: Non-canonical CSDs in different proteins.

a: Search results of UniProtKB, using a new Hidden Markov model based on an ncCSD sequence alignment (Figure 6b). Examples of the domain arrangement of different protein families containing the predicted ncCSDs are highlighted. **b:** Presentation of the sequence logo that was used for the Hidden Markov model search. **c:** $^{1}H,^{15}N$-HSQC spectra of CSD123, CSD456 and CSD789 of human Unr/CSDE1 exhibit peak dispersion similar to *Drosophila* Unr, indicative of highly structured proteins. **d:** Sequence alignment between CSDs and potential ncCSDs. The same or similar residues between all domains are colored, similar residue regions between the canonical CSDs are highlighted by red boxes, which align with the two RNA binding regions (**F**G**F** and (**F/Y**)F**H**). Additional loops only present in ncCSDs are highlighted by a blue box. The alignment has been done using Emboss Needle[257] and ESPript [258] has been used for illustration.

Taken together this unbiased and systematic approach to structurally characterize *Drosophila* Unr revealed the presence of ncCSDs. These domains lack RNA binding residues conserved in canonical CSDs. Further, genome sequence searches predicted ncCSDs in a subset of other proteins, which needs to be confirmed by future studies. Additionally, protein constructs were generated for high yield expression, stability and solubility, which can be used for in-depth mechanistic studies of how multi-domain RBPs that feature both canonical (CSDs) and non-canonical RBDs (ncCSDs) engage target RNAs.

4.2. RNA binding of multidomain and non-canonical CSD containing constructs

To assess RNA binding of *Drosophila* Unr as a multidomain RBP different methods were employed. Initially, RNA binding of Unr was investigated using electrophoretic mobility shift assays (EMSA). Here, Unr full-length and multiple shorter constructs (CSD123, CSD456 and CSD789) were tested in their ability to bind to an RNA, that derived from the well characterized target long non-coding RNA *roX2*[101]. To offer a large enough RNA surface to ensure binding of all CSDs, the binding affinity of the full-length protein was compared between two RNAs of different length, namely a stem-loop 6 and 7 (SL67) and a stem-loop 6, 7 and 8 (SL678) construct (Figure 14 and section 3.2.7.). The affinity to both RNAs is similar with a slight preference for the longer RNA. For SL67 there is still unbound RNA at a protein concentration of 1000 nM, whereas for SL678 at a protein concentration of 630 nM almost all RNA seems to be bound (Figure 14b). However, albeit speculative, this small difference indicates that the shorter RNA presumably provides enough single stranded RNA nucleotides for all CSDs to bind.

Binding of the shorter protein constructs were only tested against SL67. All three triple domain constructs (CSD123, CSD456 and CSD789) bind the RNA, whereas CSD123 shows the strongest binding affinity with a K_D at around 2 μM. The affinity for CSD456, featuring only a single canonical CSD, is with 16 μM comparable to the previously reported affinity of CSD1 in isolation[27]. Surprisingly, CSD789, which harbors two canonical CSDs showed a similar or even slightly weaker affinity (K_D of 32 μM; Figure 14c). However, the RNA binding affinities of both CSD78 and CSD9 constructs measured by NMR in isolation were significantly weaker (around 200-300 μM; Supplementary Figure 3a), indicating that synergistic binding within CSD789 may play a major role in enhancing its RNA binding affinity (see below). The higher affinity of CSD123 against the other two constructs might be explained by the higher theoretical isoelectric point of CSD1 (8.03, vs. average 6.06 for other CSDs). The higher affinity is further in agreement with the observation that in *Drosophila* the Unr N-terminal CSDs are sufficient for translational repression of *msl2* mRNA[114].

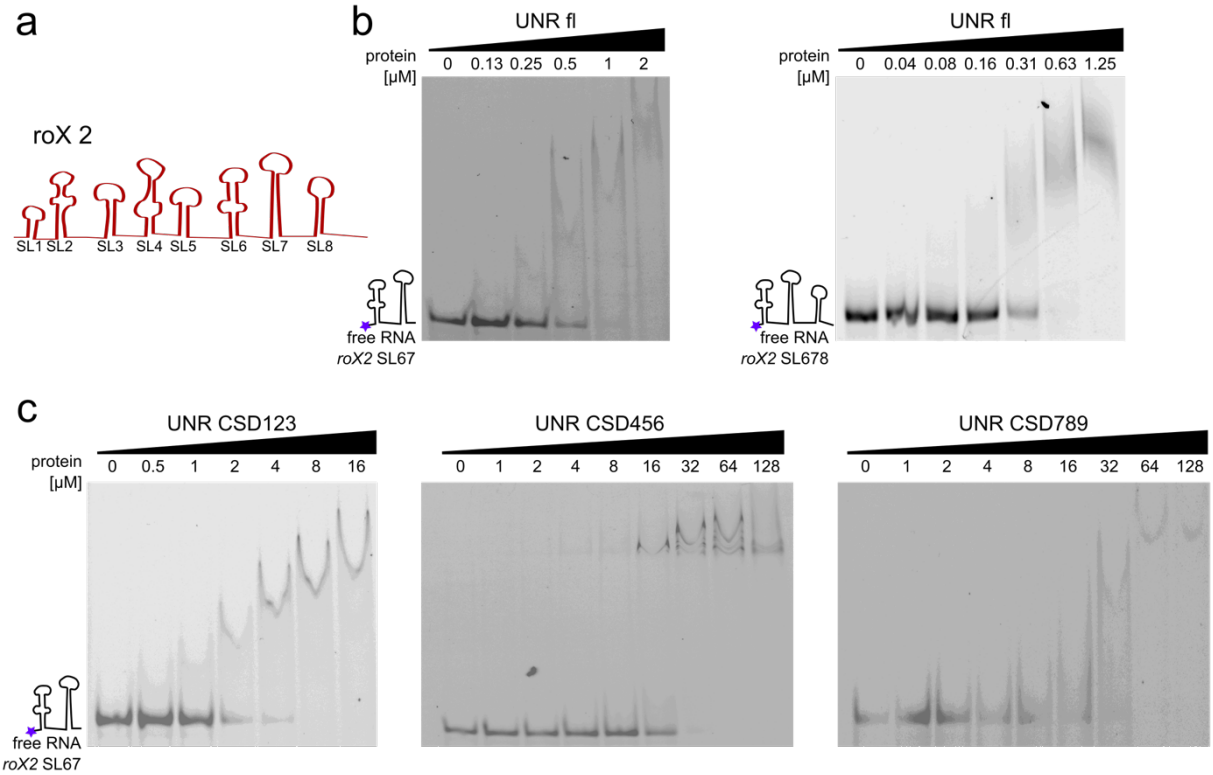


Figure 14: RNA binding of Unr multidomain constructs.
a: Schematic representation of the stem-loop arrangement within the lncRNA roX2 according to Militti et al.[101]. EMSAs show binding of Unr full-length (**b**) and different triple domain constructs (CSD123, CSD456 and CSD789) (**c**) to the stem loops 67 and 678 of the lncRNA *roX2* (see inset), which is 5' labeled with a Cy5 dye (blue star).

To evaluate whether ncCSDs, which lack the canonical RNA binding residues contribute to RNA binding within the multidomain constructs of Unr, we performed NMR titration experiments, which provide information on a single residue level. For $^{1}H,^{15}N$-HSQC NMR titration experiments different amounts of a polypurine-rich 9-mer RNA (AAA AAA AUG; adapted from the loop of lncRNA *roX2* SL6, later referred to as 'SL6a'), were titrated to different Unr protein constructs. Almost complete NMR backbone assignment allowed mapping of chemical shift perturbations (CSPs) onto the protein sequence and structure. For CSD12 (CSD123 shows aggregation at concentrations needed for this experiment), only CSD1, but not ncCSD2 interacts with RNA (Figure 15a). Similarly, only peaks corresponding to residues of CSD5 exhibit significant CSPs upon titration of RNA to CSD456. Both, ncCSD4 and ncCSD6 do not show major changes (Figure 15b). Additionally, no RNA-induced CSPs were observed during the titration of an isolated ncCSD6 construct (Supplementary Figure 3b). Contrary to this observation that ncCSDs do not participate in RNA binding, several ncCSD8 residues (e.g. R866, C867 and I868) exhibit significant chemical shift perturbations for RNA titration experiments against CSD789 (Figure 15c). This was confirmed by a titration against CSD78 (Figure 15d). However, titration of RNA to an isolated ncCSD8 did not induce chemical

shift perturbations (Supplementary Figure 3c), suggesting that the effects seen in the multi-domain constructs are due to proximity effects. Our NMR structure of CSD78 shows a positively charged area, located close to the RNA binding interface of CSD7 (Figure 15e). This region corresponds to the residues with significant CSPs on ncCSD8.

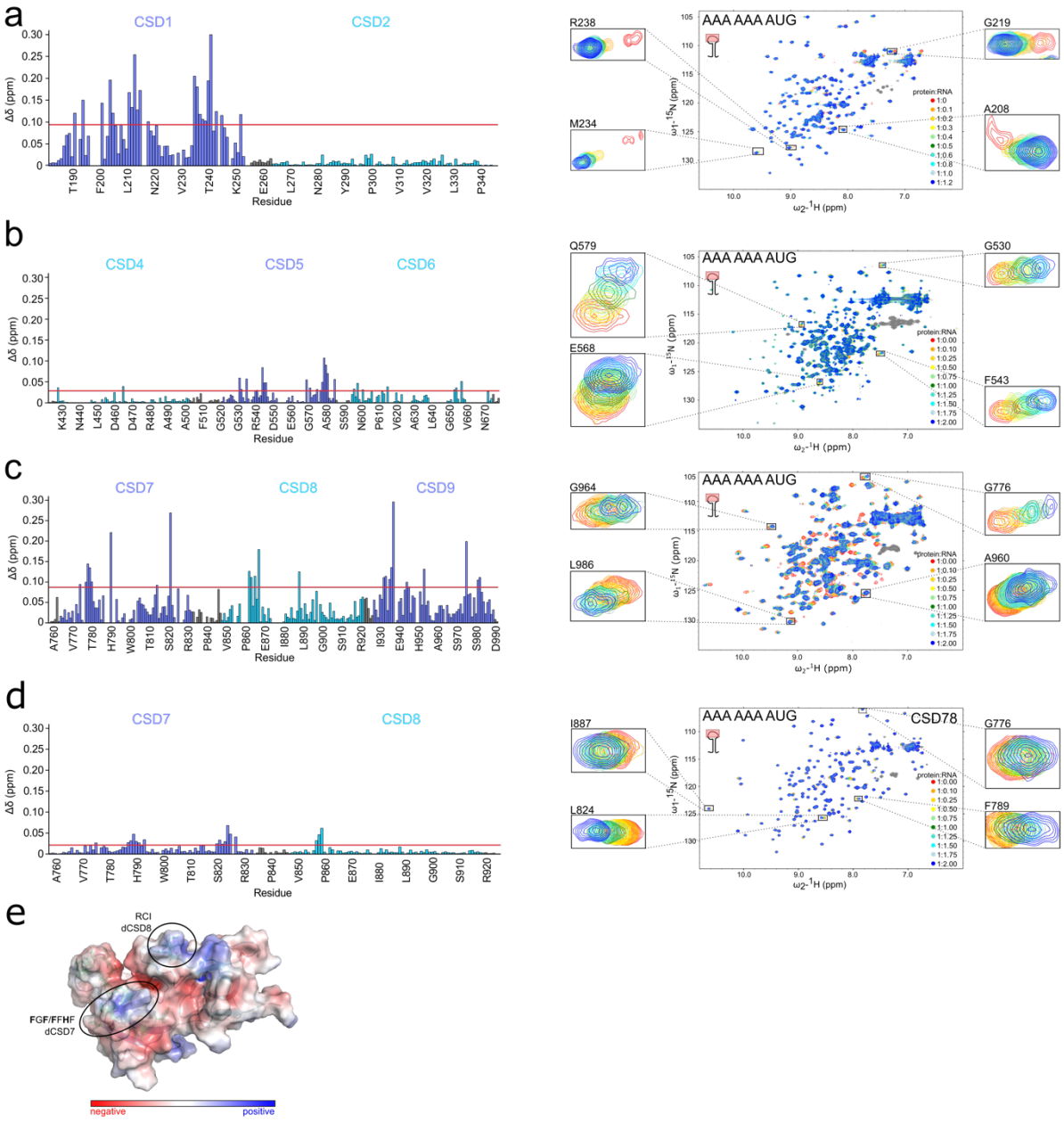


Figure 15: RNA binding of non-canonical CSDs measured by NMR titrations.
$^{1}H,^{15}N$-HSQC NMR titration experiments with a 9-mer RNA (AAA AAA AUG), derived from the single stranded loop region of stem loop 6 of roX2 (SL6a) and the calculated chemical shift perturbation (CSP) plots, for CSD12 (**a**), CSD456 (**b**), CSD789 (**c**) and CSD78 (**d**). The red line in the CSP plots indicates the average plus the standard deviation of all measured shifts, which was used to identify significant shifts[86]. **e**: Calculated surface potential for CSD78. Highlighted is the region of RNA binding residues of canonical CSD7 and the region of residues in ncCSD8, that show significant shifts after addition of RNA. Negative surface potential is highlighted in red and positive in blue.

Altogether, our RNA binding studies indicate that although ncCSDs do not interact with RNA in isolation, they may contribute to RNA binding, like ncCSD8, in a multidomain context (Figure 15). Additionally, we also observed synergistic effects when RNA affinity was measured in the context of multi-domain constructs that include both canonical and non-canonical CSDs (Figure 14).

4.3. Interdomain contacts mediate fixed orientation between canonical and non-canonical CSDs

To define the role of ncCSDs within Unr further, we analyzed the structures of our multi-domain CSD constructs to identify the residues that are involved in the formation of interdomain contacts. CSD5 and both ncCSD4 and ncCSD6 form several interdomain contacts in the crystal structure of CSD456 (F477 on ncCSD4, L505, T521, R533, Q538, E547, L549 and R582 on CSD5 and F593, N663 and R662 on ncCSD6) (Figure 16). To confirm the presence of these contacts in solution and exclude artefacts due to crystal packing, small-angle X-ray scattering (SAXS) was used. The experimentally observed scattering densities and the back-calculated scattering densities from the crystal structure show a high conformity (χ of 1.02; SASBDB: SASDHJ7; Supplementary Figure 4a), suggesting that the fixed domain arrangement seen in the crystal structure, which keeps the domains at a certain distance and orientation to each other, is maintained in solution (Figure 16).

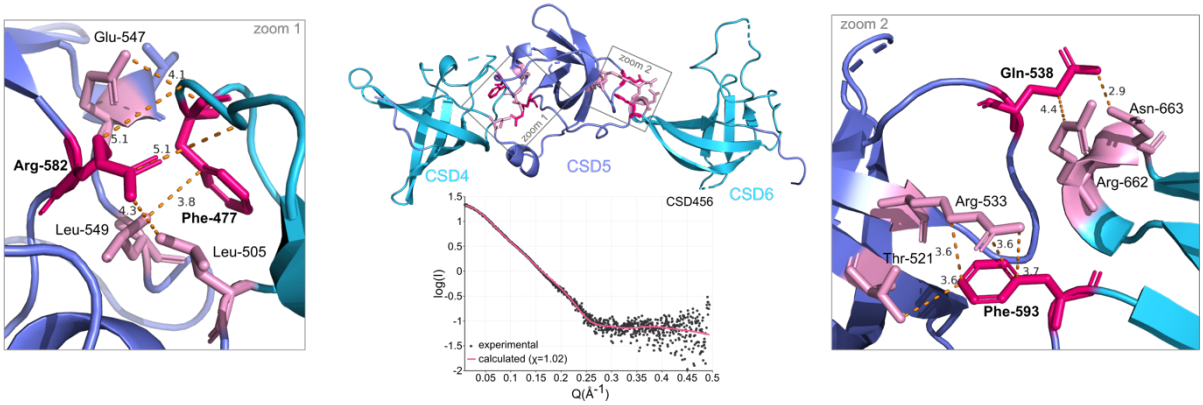


Figure 16: Interdomain contacts between canonical and ncCSDs within CSD456.
Potential interdomain contacts between ncCSD4 and CSD5 and between CSD5 and ncCSD6 from the crystal structure of CSD456 are highlighted in pink (bold-faced labeled residues colored in magenta are mutated for analyses in functional assays presented below). Middle: The SAXS scattering curve of CSD456 in solution (black dots) fits the back-calculated scattering density of the crystal structure (magenta) (χ=1.02; SASBDB: SASDHJ7).

Concomitant with these observations, a fixed domain arrangement was also observed between the two domains of the NMR structure of CSD78. For CSD78 49 interdomain and 48 domain-linker NOE-based distance restraints were identified on top of the 4.433 intradomain

NOEs. These molecular interactions form a defined domain-domain interaction network (Figure 17a and 17b).

The interface mostly consists of hydrophobic interactions formed by residues R765, F767, A769, L803 and E806 on CSD7, and I837, Y865, I887 and T888 on ncCSD8. Additionally, the overall fixed conformation was confirmed by SAXS (Figure 17c; χ=1.1; SASBDB: SASDHK7; Supplementary Figure 4a) and ^{15}N NMR relaxation data, which provide a measure of dynamics on a residue resolution level. The rotational correlation time τ_c (see 1.7.1) is similar over all residues within the molecule, with an exception between R813 and L824, a region that corresponds to a flexible intradomain loop between β-strand 4 and 5 of CSD7. Also, τ_c is too elevated if independent molecular tumbling of each domain is assumed (CSD78: τ_c = 12.7±1.2 ns) (Figure 17d). For single CSDs with an average size of 9 kDa a τ_c value of around 5.4 ns would be expected. Although, the time would increase for a dimeric construct with flexible linker, due to the influence of the second domain on the tumbling, the time would be much below 10 ns.

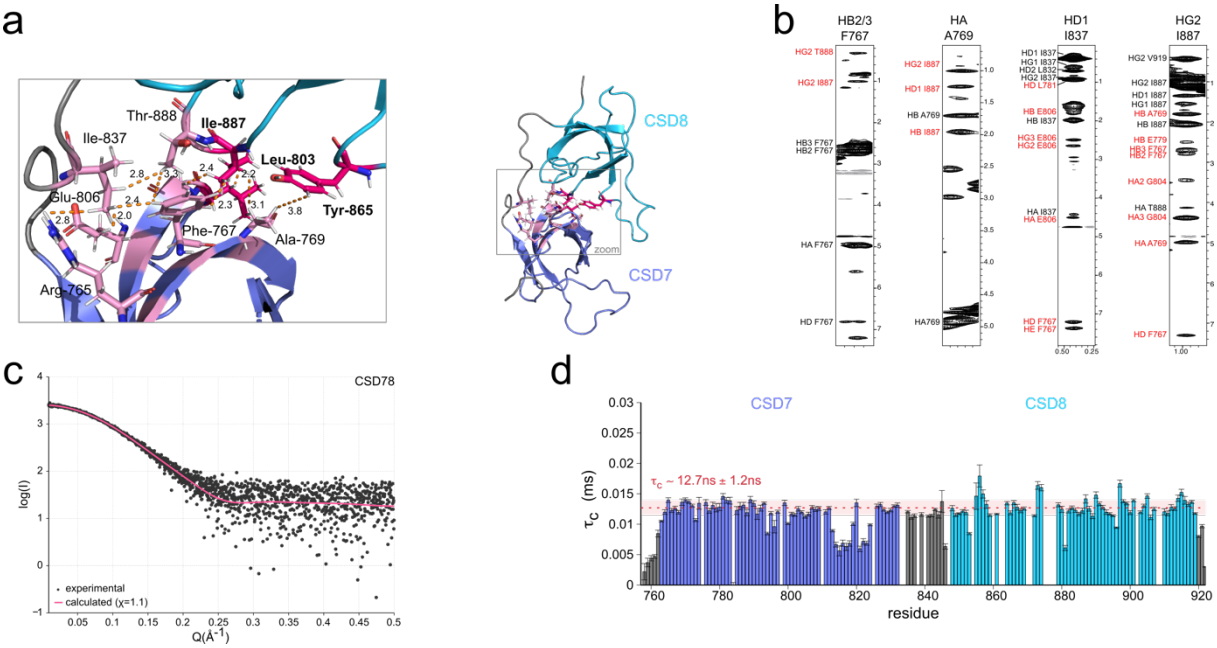


Figure 17: Interdomain contacts between CSD7 and ncCSD8 within CSD78.
a: A network of NOEs between mostly hydrophobic interactions of the two CSDs is highlighted in the NMR solution structure of CSD78 (highlighted in pink; residues labeled bold-faced and colored in magenta are mutated for functional assays). **b**: Exemplary NOE strips of the 3D ^{13}C,^{1}H,^{1}H-NOESY spectrum of CSD78 highlighting some of the assigned interdomain NOEs (red), that were detected and used for structure calculations. **c**: The SAXS scattering curve of CSD78 in solution (black dots) fits to the back-calculated scattering density of the NMR structure (magenta) (χ=1.1; SASBDB: SASDHK7). **d**: Relaxation data of CSD78 indicating joint tumbling of CSD7 and 8 in solution (τ_c = 12.7±1.2 ns), with flexible regions (residues between R813 and L824) within the domains (flexible loop between β-strands 4 and 5 of CSD7). The rotational correlation time (τ_c) is plotted per residue.

Similar to CSD78, ^{15}N relaxation experiments of CSD12 showed a high rotational correlation time (τ_c = 10.1±1.9 ns), indicating that these domains have a limited freedom of movement with respect to each other (Figure 18a). Although direct interdomain NOEs could not be identified in this case, linker-domain NOEs between CSD1 to linker residues up to A252 and NOEs between CSD2 to the linker starting after residue E254 were detected and may explain the joint tumbling. Thus, only three residues within the linker could move without restraint. An increased rigidity of the linker could further be explained by a proline (P253), which is in the center of this short linker (Figure 18b)[260,261].

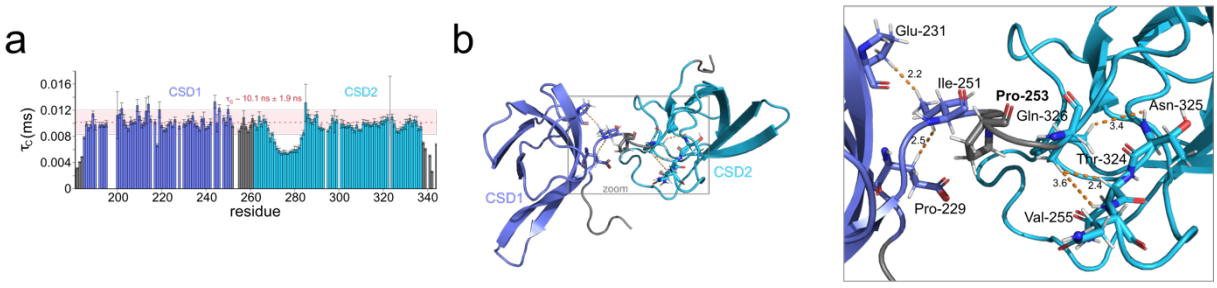


Figure 18: Joint tumbling of CSD1 and ncCSD2.
a: ^{15}N relaxation data of CSD12, indicating tumbling of the two single domains together in solution (τ_c = 10.1±1.9 ns). Flexible residues are located between L270 and T285, which corresponds to an interdomain loop between β-strand 1 and 2 in CSD2. **b**: CSD12 is lacking interdomain NOEs, but shows only a three-residue long linker, without NOEs, including one proline (labeled bold). The rotational correlation time (τ_c) is plotted per residue.

Interestingly, a flexible region within ncCSD2 seen in the ^{15}N relaxation data (Figure 18a), corresponds to the already mentioned extended loop between β-strands β1 and β2, that is a peculiarity of the ncCSDs. Concomitant with these observations, a weak or absent electron density in ncCSD4 and ncCSD6 of the X-ray diffraction data of CSD456 indicates flexibility of this loop. In contrast, the relaxation data of CSD78 do not indicate flexibility for the corresponding residues in ncCSD8, likely due to interdomain interactions with CSD7. Consequently, it cannot be concluded that this extended loop region is a general CSD-CSD interaction surface. Furthermore, the amino acid sequence in this region is not conserved and the described domain-domain interactions are of different compositions and locations (Supplementary Figure 4b).

In contrast to the described joint tumbling of some domains, ^{15}N relaxation measurements of CSD678 showed that ncCSD6 (τ_c = 13.0±3.3 ns) and CSD78 (τ_c = 16.4±4.4 ns) tumble independently in solution, presumably due to the presence of a long linker and a flexible Q-rich domain (Figure 19a). Relaxation data of CSD789 do confirm the joint tumbling between CSD7 and 8, but also indicates that CSD9 tumbles independently (CSD78: τ_c = 18.3±1.2 ns;

CSD9: τ_c = 14.5±1.8 ns) (Figure 19b). The linker between ncCSD8 and CSD9 is only 4 residues long and although the domains do not tumble together, SAXS data analyzed by the ensemble optimization method (EOM)[202] indicate spatial restrictions for CSD9 with respect to CSD78 (Figure 19c, 19d and Supplementary Figure 5). This is further confirmed by NMR data, which show CSPs in a region within CSD9 when compared to the longer construct of CSD789.

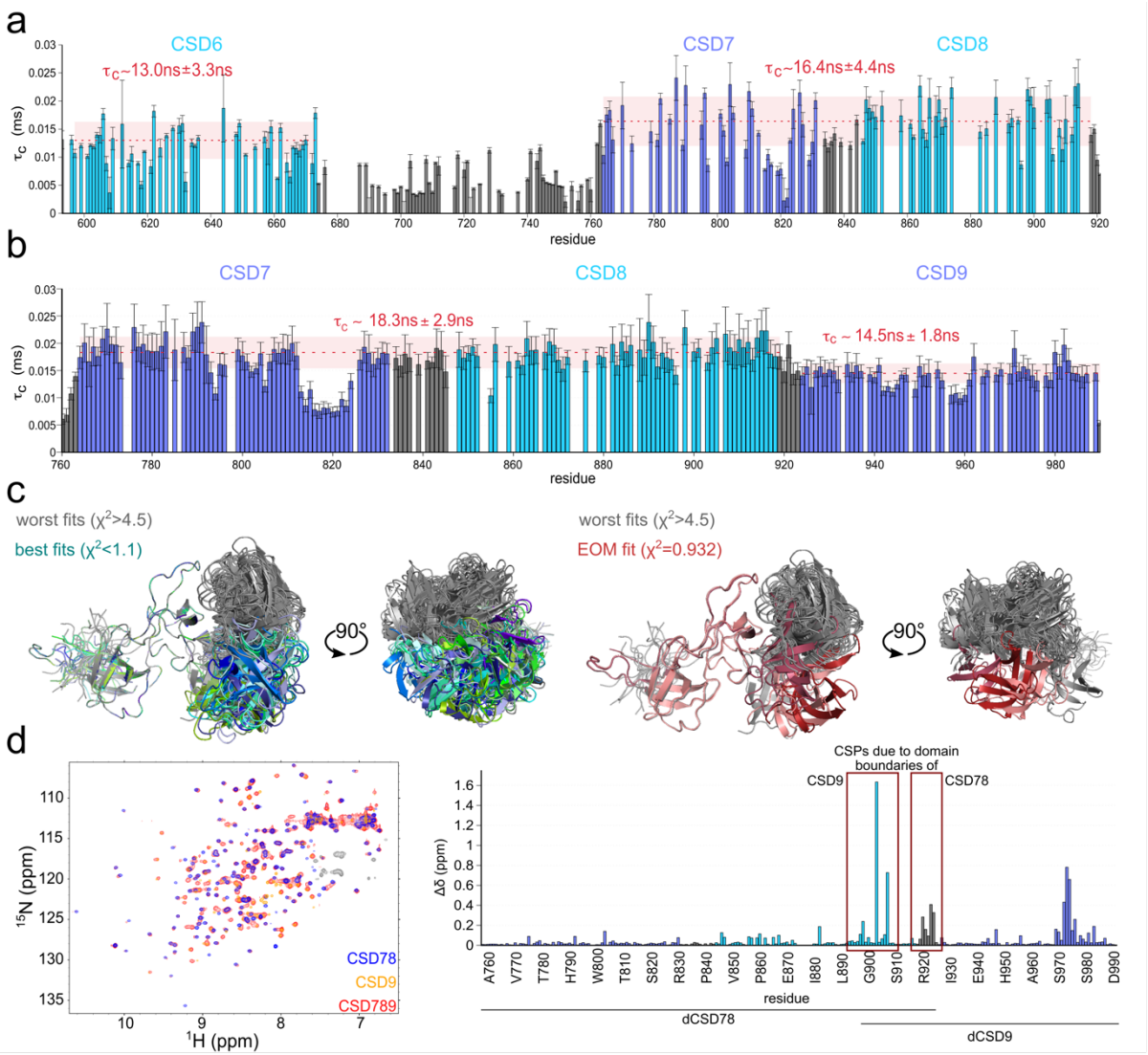


Figure 19: Flexible interdomain linkers between canonical and non-canonical CSDs within Unr. **a**: ^{15}N relaxation data of CSD678 indicating that the long linker between ncCSD6 and CSD7 is flexible (CSD6: τ_c = 13.0±3.3 ns vs. CSD78: τ_c = 16.4±4.4 ns). **b**: ^{15}N relaxation data of CSD789 indicating that the linker between ncCSD8 and CSD9 is flexible (CSD78: τ_c = 18.3±1.2 ns vs. CSD9: τ_c = 14.5±1.8 ns). **c**: EOM calculation and structure modeling using SAXS data suggests a restricted flexibility between ncCSD8 and CSD9. **d**: Overlaid ^{1}H,^{15}N-HSQC spectra of CSD78, CSD9 and CSD789 (left) showing CSPs especially in the termini of the single constructs (red boxes), but also within the single domains, especially CSD9. The rotational correlation time (τ_c) is plotted per residue.

Taken together, although some linkers between canonical CSDs and ncCSDs remain flexible, the identified domain interactions clearly impose spatial restrictions on the protein fold. The ncCSDs may act as scaffolding domains, that maintain distance and orientation between the

canonical CSDs. Thus, ncCSDs might position RNA binding motifs of the canonical CSDs (or potential protein-interaction surfaces) towards target RNAs and/or ribonucleoprotein complexes (RNPs) and as such contribute to RNA binding and induce RNA/RNP tertiary structure specificity.

4.4. Interdomain contacts in full-length Unr promote protein stability and RNA target specificity

Hereinafter, the importance of the interdomain contacts in stabilizing the overall structure of full-length Unr was assessed through mutational studies using several methods. To avoid misfolding of the protein, only residues identified to form interdomain contacts located in loop regions that lack secondary structure were mutated to alanine. Series of single mutants as shown in Figure 20a, as well as double and triple mutants (F477A/R582A referred to as 45 ID given that it disrupts interdomain interactions between ncCSD4 and CSD5; Q538A/F593A: 56 ID, disrupting CSD5 and ncCSD6 interactions; L803A/Y865A/I887A: 78 ID, disrupting the CSD7 and ncCSD8 interface) (Figures 16, 17a and 20a) were generated. The structural integrity of the CSD β-barrels in Unr full-length mutants was confirmed by circular dichroism spectroscopy (CD, Figure 20b). The derived secondary structure content in full-length WT protein and the mutants was similar, indicating that the mutations indeed only affect the interdomain contacts, but not the overall CSD domain fold.

Melting temperatures measured by differential scanning fluorimetry (DSF) showed that the stability of CSD456 and CSD78 mutants decreased compared to the wild type protein constructs (WT) (Figure 20c). Even the single mutants displayed a melting temperature of up to 10°C lower than the WT, suggesting a compromised overall structural integrity. This gets further supported by a decreased solubility observed during the purification process and in $^{1}H,^{15}N$-HSQC experiments and SAXS data, which both show a stronger aggregation propensity in the mutant constructs (Figure 20d, 20e and 20f). Aggregation leads to signal loss in $^{1}H,^{15}N$-HSQC spectra due to the increased size of the measured molecules and decrease of the monomeric species concentration. This signal loss for the different single mutants of CSD78 is not as drastic as for most of the CSD456 mutants, where barely any signal remains. An exception in both measurements (NMR and SAXS) is the single mutation of Q538A, suggesting that this may not disrupt the interdomain contacts efficiently (Figure 20d). Aggregation is also not visible for the L803A mutant disturbing the interaction between CSD7 and ncCSD8, whereas the I887A mutant is showing stronger aggregation than the Y865A mutant. Strong and numerous chemical shift perturbations occur in the non-aggregated L803A mutant, indicative

for a perturbed domain-domain interaction. The unvaried dispersion of peaks of all mutants shows that the overall CSD fold is retained (Figure 20f).

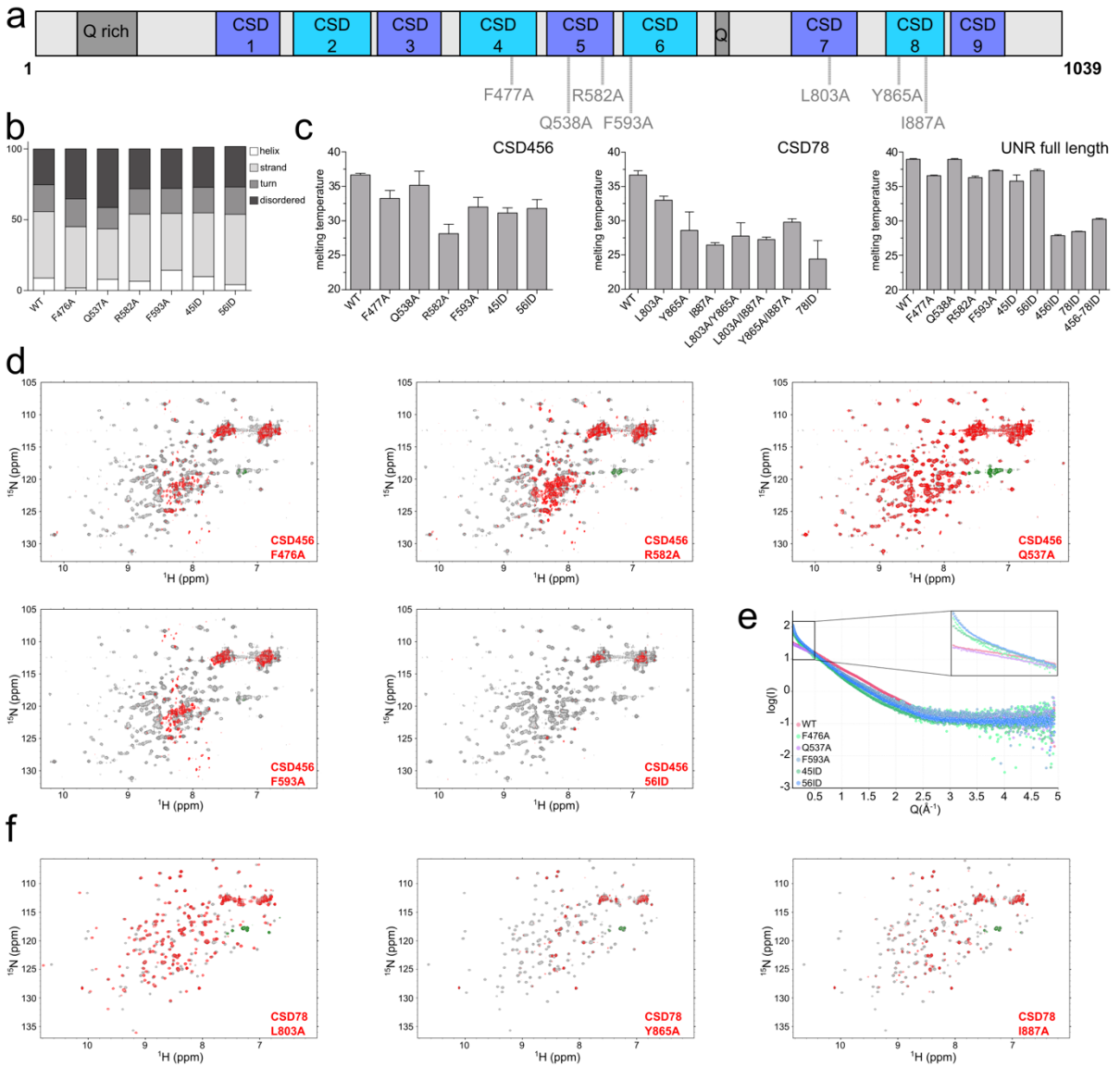


Figure 20: Interdomain contacts are involved in full-length Unr protein stability.
a: Schematic presentation of the inserted mutations within the full-length Unr protein. **b**: Prediction of different secondary structure elements (helix, strand, turn and disordered) from CD curves of Unr full-length wild type (WT) and different interdomain mutants, showing no impact of the mutations on the CSD fold. **c**: Melting temperatures for CSD456, CSD78 and full-length Unr WT and mutants as determined by nanoDSF. Measurements were done in duplicates. Shown is the mean and the error bars indicate the standard error. **d**: $^{1}H,^{15}N$-HSQC spectra of CSD456 wild type (grey) overlaid with the spectra of CSD456 mutants (red). All measured proteins had a concentration of 30 µM. **e:** SAXS curves of CSD456 wild type and the different mutants. All samples were measured at the same concentration (1 mg/ml). **f**: $^{1}H,^{15}N$-HSQCs of CSD78 wild type (grey) overlaid with the different spectra of CSD78 mutants (red). All measured proteins had a concentration of 80 µM.

The CSD45 and CSD56 mutations, which show a strong decrease in the melting temperature within the shorter constructs only have a minor influence on the full-length protein stability (Figure 20c). Further for all full-length mutants, melting temperatures are above 25°C, which

corresponds to the growth temperature of SL2 cells. Thus, *in vitro* and cellular functional studies of these mutants should yield interpretable results to obtain insights into the importance of the restricted flexibility mediated by the interdomain contacts within *Drosophila* Unr.

To test our hypothesis that interdomain contacts are important determinants of RNA tertiary structure specificity and are thereby involved in the selection of RNA targets, we used full-length Unr mutants (456 ID and 456-78 ID) in a RIP-seq experiment. To this end, endogenous Unr was knocked down is SL2 cells to liberate rate-limiting targets. These cells were then transfected with V5-tagged Unr WT, mutants or an empty V5-tag vector (background control). Three days later Unr was pulled down using an anti-V5 antibody and digested by proteinase K, so that the bound RNA targets could be extracted and sequenced.

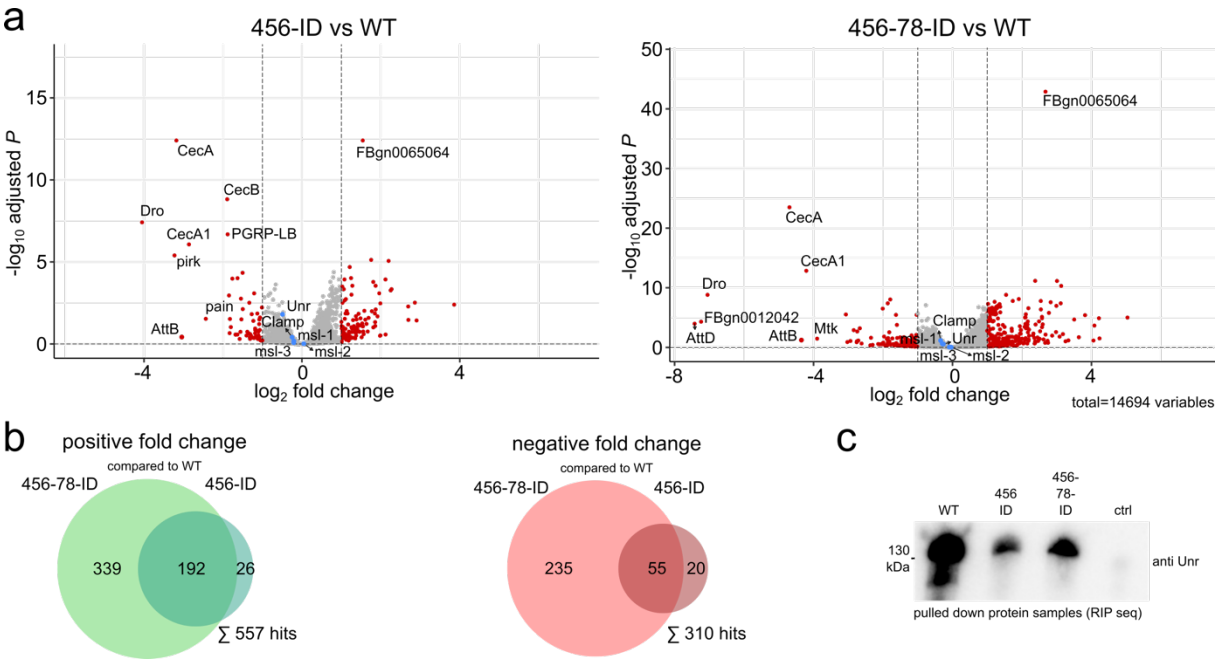


Figure 21: Unr interdomain contacts play a role in RNA target selectivity.
a: Volcano plots showing the difference of pulled down RNA targets of the RIP-seq experiment between the wild type (WT) and interdomain mutant samples (456 ID left graph, 456-78 ID right graph). The log_2 fold change is plotted against the log_{10} adjusted p value; N=3. **b**: Venn diagrams showing the numbers of positive (left graph; green) and negative (right graph; red) enriched targets of the different mutants compared to the WT. **c**: Western blot showing the efficiency of the Unr pull-down against the V5 tagged wild type and the two interdomain mutants of the samples (456ID, 456-78-ID) used for RIP-seq. No protein was detected in the control sample, which was transfected with an empty vector. A Unr antibody against the N-terminal part of the protein was used.

The WT and interdomain mutant protein samples showed differences in the numbers of pulled-down RNAs. The total number of significantly enriched genes is higher in the mutant pull-downs, with 218 genes being enriched in the 456 ID and 531 in the 456-78 ID mutant compared to the WT sample (Figure 21a and 21b). Contrary, the WT only shows 75 genes that are significantly enriched over the 456 ID and 290 over the 456-78 ID. 192 genes are

overlapping in the positively and 55 in the negatively enriched samples between the two different mutants. An increase in mutations led to an increase in the total number of differentially bound RNAs and most of the changes in 456 ID were also seen in the 456-78 ID mutants, indicative for a less discriminate RNA binding potentially due to the higher conformational heterogeneity of mutant Unr and thus higher adaptability to RNA structures (Figure 21a and 21b).

The higher protein level of Unr in the pulled down samples of the WT, which is visible in the western blot (Figure 21c) makes it problematic to say with certainty whether the enriched RNAs in the WT are due to different binding behavior to these RNAs or due to the different input amount, which may result in more hits above the significant enrichment. Nevertheless, these data indicate, that more RNAs are bound and pulled-down once the conformational heterogeneity of Unr is increased due to the mutations, showing that the scaffolding role of ncCSDs influences the RNA target specificity of *Drosophila* Unr.

4.5. Unr interdomain mutants impact translation regulation

To assess whether the changed RNA binding behavior of the Unr protein mutants also has influence on its translation regulation function, reporter gene assays in SL2 cells were performed using *msl2* mRNA as previously described[27,113,117]. To this end, endogenous Unr was depleted by an siRNA knock-down and three days later a β-galactosidase reporter gene construct containing the 5' and 3' UTRs of *msl2* mRNA, together with a Renilla control plasmid and plasmids encoding for Sxl and V5-tagged Unr were transfected (Figure 23a). SL2 cells are derived from male *Drosophila* flies, meaning they are free of an endogenous Sxl background level. Initial experiments were conducted to assess amounts of Sxl and Unr plasmids that result in a significant but not total translation repression in order to recognize eventual up- and downregulations of the mutants compared to the WT (Figure 22a and 22b).

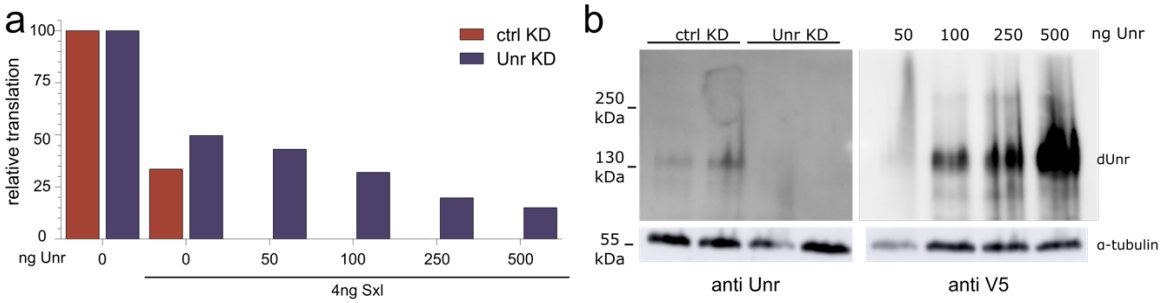


Figure 22: Translation inhibition of reporter gene RNA of different Unr plasmids amounts.
a: Relative translation of β-galactosidase after expression of different amounts of transfected Unr plasmid cotransfected with Sxl. Translation levels were further normalized to the ones from Renilla and the levels of the reporter RNA. **b**: The knock-down and transfection efficiency was assessed by western blot analysis using anti Unr and anti V5 antibodies. Tubulin was detected as a loading control.

Interestingly, translation of the target mRNA was significantly repressed by the 456 ID mutant compared to the WT background repression (Figure 23b). This higher activity is even more striking, considering the lower cellular protein levels of the mutant (Figure 23c). This either indicates strengthened translation repression or increased binding to the reporter gene. Mutation of only one of the interdomain interfaces does not show a significant difference with respect to the WT protein (Figure 23b) with the protein levels of the transfected protein being comparable to WT Unr.

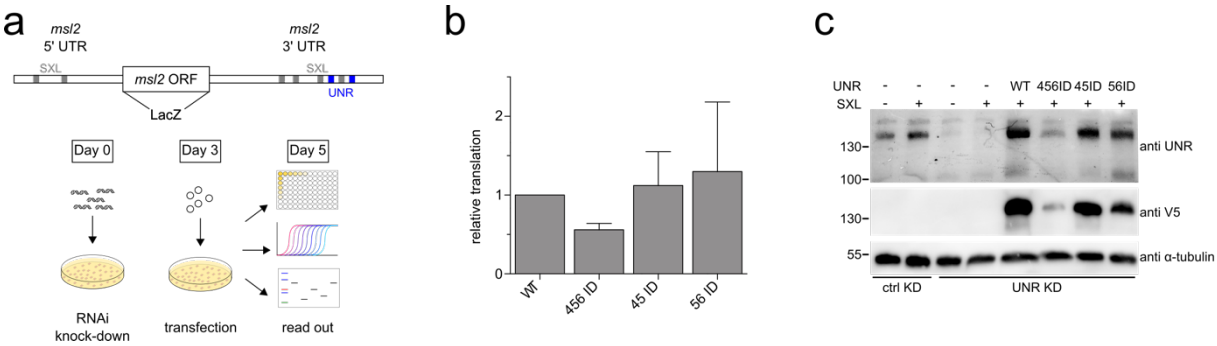


Figure 23: Proper controlled *msl2* translation repression of *Drosophila* Unr relies on intact interdomain contacts.

a: Schematic representation of the experimental set up of the cellular reporter gene assay and the used reporter gene construct. The 5' and 3' UTR are derived from the *msl2* mRNA exhibiting the binding sites for Sxl (grey) and Unr (blue). The open reading frame is coding for β-galactosidase. **b**: Relative translation of β-galactosidase after expression of different Unr mutants normalized to the level of the wild type protein. Translation levels were further normalized to the ones from Renilla and the levels of the reporter RNA; N=3. Shown is the mean and the error bars indicate the standard error. **c**: Knock-down and transfection efficiency was assessed by western blot analysis using anti-Unr and anti-V5 antibodies. Tubulin was detected as a loading control. The upper band in the anti-Unr blot is full-length Unr and the lower band is a C-terminal truncation.

In a next step, we investigated whether different binding of Unr to the target RNA or a direct effect on translation is responsible for the observed differences in translation repression. Therefore, an *in vitro* tethering translation assay using *Drosophila* embryo extract as described by Abaza et al.[114] was conducted. An mRNA construct containing the Firefly luciferase ORF, the 5' UTR of the *msl2* mRNA and a 3' UTR consisting of 9 MS2 binding sites was used for tethering (Figure 24a). Different amounts of recombinant MS2-tagged Unr were added to the RNAs and the embryo extract. The firefly luciferase levels were measured as a proxy of translation. As an internal translation control an mRNA encoding for Renilla luciferase under a Unr resistant promoter was added and the data corrected for the variations in its luciferase activity. To ensure specificity of the assay, untagged Unr and an unrelated MS2-tagged protein (MBP-MS2) were used as negative controls. Due to the instability of the 456ID and 456-78ID

recombinant proteins only 45ID and 56ID, which showed a higher melting temperature (Figure 20c), could be tested.

Compared to WT Unr, both mutants showed decreased activity in translation repression for low protein concentrations (2.5-7.5 µM; Figure 24b). At the highest tested protein concentration, the repression level equalizes between the early saturated WT and the mutants. The two measured controls do not show repression of translation indicating the reliability of the assay. Quantified mRNA levels after the translation by qPCR show that the observed differences are not due to variations in mRNA levels (Figure 24c), supporting the conclusion of differential translation regulation by the Unr constructs. Additionally, tested single interdomain mutants of *Drosophila* Unr showed a similar translation repression efficiency as the double mutants, with even less compensation at the highest measured protein concentration, indicating that already the single mutations have an influence on the protein function (Figure 24d and 24e).

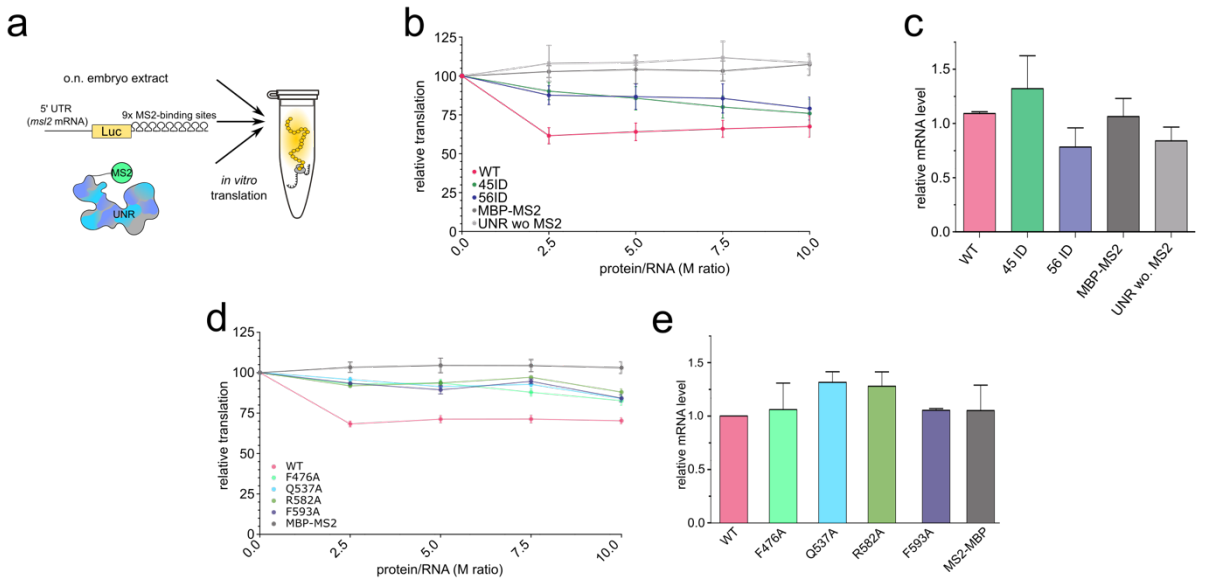


Figure 24: Influence of Unr interdomain contacts on cofactor binding.
a: Schematic representation of the *in vitro* translation assay. Recombinant MS2-tagged Unr gets tethered to the 3'-UTR, which contains nine MS2-binding loops. Binding leads to repression of translation within *Drosophila* embryo extracts. The same mRNA construct was used before[114] (BmutLMS2). **b/d**: Relative *in vitro* translation of the firefly reporter gene over the internal control (Renilla) after adding increasing amounts of Unr wildtype protein and different mutants; N=3. **c/e**: Relative reporter mRNA levels after *in vitro* translation (the point of 2.5 molar excess of Unr over RNA was taken), measured by RT-qPCR; N=3. The mean is shown for each data point and the error bars indicate the standard deviation.

As the tethering ensures the RNA binding of Unr to the 5' UTR of the mRNA, this assay clearly evaluates the role of Unr in translation regulation detached from its RNA binding behavior to the reporter gene construct. The decreased activity of the mutants in the *in vitro* tethering assay indicates that the fixed domain orientation between CSD456 promotes the regulation of

translation on a level that is independent of the initial RNA binding event. Compared to the cellular results shown in Figure 23, these data suggest an additional functional role of the interdomain contacts, besides mediating RNA tertiary structure specificity. This could well be role in RNP specificity, where ncCSDs and their interactions with CSDs provide protein binding platforms.

4.6. Interdomain contacts and non-canonical RNA binding residues are involved in RNA binding of CSD789

In order to investigate whether the restricted flexibility observed between ncCSD8 and CSD9 (section 4.3; Figure 19b, 19c and 19d) has a similar importance as the fixed domain arrangement observed between other domains, further RNA binding studies were performed. In a first step two poly adenine RNAs of different length (6-mer versus 15-mer) were titrated to the wild type protein. The 15-mer RNA was chosen to ensure that all domains of the CSD789 construct could bind the same RNA, whereas the 6-mer RNA would be too short and allows only the binding of a single CSD. Deuterated protein samples yielded a better signal-to-noise ratio and thus allow reliable interpretation.

Although the RNAs were not titrated to the same final endpoint, different binding modes are clearly discernable from the NMR titrations (Figure 25a and 25b). For the shorter A6-mer the CSPs induced by RNA binding is in the fast exchange regime, meaning that the chemical exchange rates between the bound and the unbound state are faster than the signal detection. This results in traceable CSPs, and for each titration point the average peak position between bound and unbound state depends on the ratio of the two states. Thus, a titration curve (peak position vs. RNA concentration) could be established from which a dissociation constant can be derived. In contrast, the binding to the A15-mer results in CSPs in the intermediate to slow exchange regime, indicative for a tighter interaction as there is a slower exchange between the bound and unbound state. Theoretically, both peaks of the bound and unbound state are visible simultaneously and the peak intensity ratio of both peaks depends on the bound and unbound fraction. Thus, these results show that the affinity of CSD789 towards the longer RNA is considerably stronger than for the shorter one, strengthening the previously observed synergistic binding of domains within CSD789 (section 4.2; Figure 14c and Supplementary Figure 3a).

To assess whether binding of the protein to one RNA strand may influences the independent tumbling behavior of CSD78 vs. CSD9 within CSD789 as described for other RBPs[262–266], ^{15}N relaxation data were measured for the RNA bound state. Indeed, upon RNA binding the

rotational correlation time of CSD9 increases to the overall tumbling time of CSD78 (14.5±1.8 ns unbound vs. 18.4±1.3 ns bound), meaning that the three domains tumble together in solution as one entity (Figure 25c).

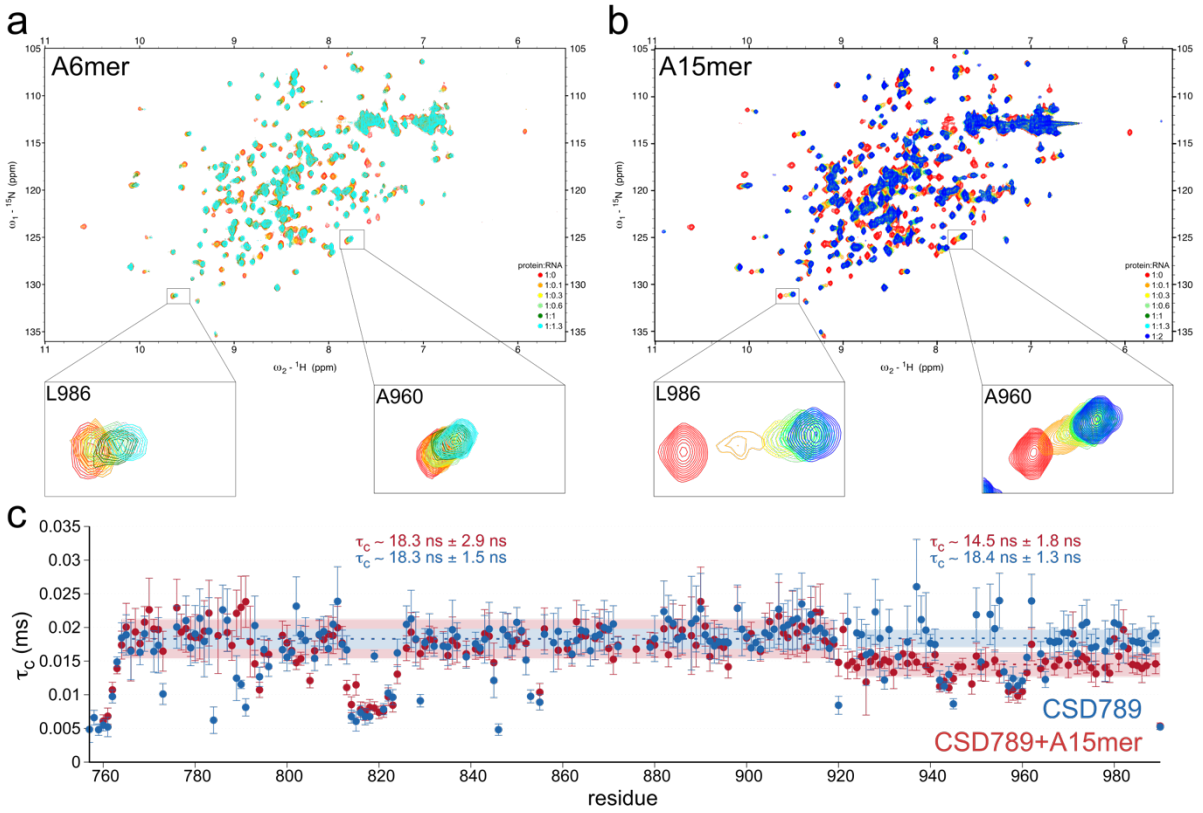


Figure 25: Joint tumbling of ncCSD8 and CSD9 within CSD789 upon RNA binding.
Overlaid $^{1}H,^{15}N$-HSQC spectra showing the titration of CSD789 with a 6-mer (**a**) and a 15-mer (**b**) RNA. **c**: ^{15}N relaxation data of CSD789 overlaid with the same protein bound to an A15-mer RNA indicating that CSD7-ncCSD8 and CSD9 tumble together when RNA is bound (τ_c of CSD9: 14.5±1.8 ns unbound vs. τ_c of CSD9: 18.4±1.3 ns bound). The rotational correlation time (τ_c) is plotted per residue.

The joint tumbling behavior of all three domains in the RNA bound state motivated us to try to crystallize this protein-RNA complex. Crystal growth could be observed in several conditions starting three days after setting up plates and resulted in a diffracting crystal of up to 1.2 Å (see 3.2.3.). By using molecular replacement with three poly-A-chain CSD1s derived from the ternary Unr-CSD1-Sxl-RRM12-*msl2* mRNA complex structure (PDB: 4qqb)[27] the structure of CSD789 bound to a poly-A chain could be solved.

From the 15 adenosines only five were visible in the structure, either due to flexibility of the remaining unbound nucleotides establishing no electron density or due to previous degradation within the sample drop. Furthermore, one RNA chain is bound by two molecules from the same unit cell. To be exact, CSD9 binds to RNA that is bound by CSD7 and ncCSD8 of its symmetry mate, resulting in one unit cell having two protein and two RNA molecules (Figure 26a).

From NMR titration and ^{15}N relaxation experiments it is known that despite high concentrations, the complex forms with a 1:1 (protein-RNA) stoichiometry, as the rotation correlation time fits to what can be expected of a 30 kDa complex (around 18 ns, for a 2:2 complex one would expect a τ_c of above 30 ns (Figure 25b and 25c). These observations were solidified by SAXS measurements of a pre-purified complex, that also indicates a molecular weight of around 30 kDa based on the normalized I(0), which is a direct proxy of the molecular weight (Supplementary Figure 6; Supplementary Table 3). These results suggest, that the peculiar assembly seen in the crystal structure is a result of crystal packing forces. To get a better impression of how the complex looks in solution, a structure was modeled that connects the termini of the two single RNA strands, by two additional nucleotides (Figure 26b; A6 and A7; dark red).

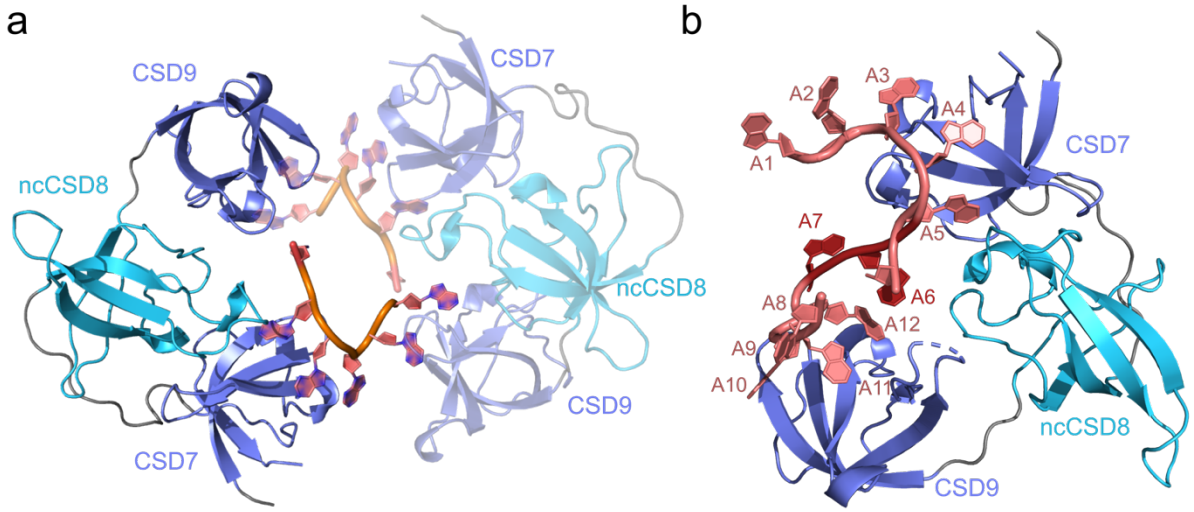


Figure 26: Crystal structure of CSD789 bound to RNA.
a: The crystal structure of CSD789 shows all three CSDs bound to 5 As of the RNA sequence. The symmetry mate in the same unit cell is highlighted by lower opacity. One RNA strand is bound to CSD7 of one and CSD9 of the other symmetry mate. **b**: Shown is a modeled structure in which the two RNA strands from both symmetry mates (A1-A5 and A8-A12; pale red) are linked together by two additional nucleotides (A6 and A7; dark red). Canonical CSDs are colored in blue and the ncCSD8 in cyan.

The validity of the resulting model is confirmed by already described NMR titrations. The two modeled nucleotides, linking both RNA strands establish contacts to residues in ncCSD8, of which corresponding NMR resonance shift significantly upon RNA titration (section 4.2.; Figure 15c and 15d).

Several details of this model are noteworthy. As described previously for many other CSD structures[27,128,129,134,140,141,267,268], the known RNA binding motifs **F**G**F** and **F**H**F** are also involved here in RNA binding by CSD7 and CSD9 (Figure 27a). Surface-exposed aromatic sidechains of F777, F788 and H790 of CSD7 and F934, F948 and H950 of CSD9 form

π-π-stacking to the bases of A3-A5 and A8-A9 of the RNA, resulting in a tightly packed interaction surface between the canonical CSDs and RNA. However, besides this canonical RNA binding residues previously unobserved non-canonical interactions are formed between N977 and A6 or K979 and A9 in CSD9, which also seem to be sequence specific towards adenine nucleotide (Figure 27b).

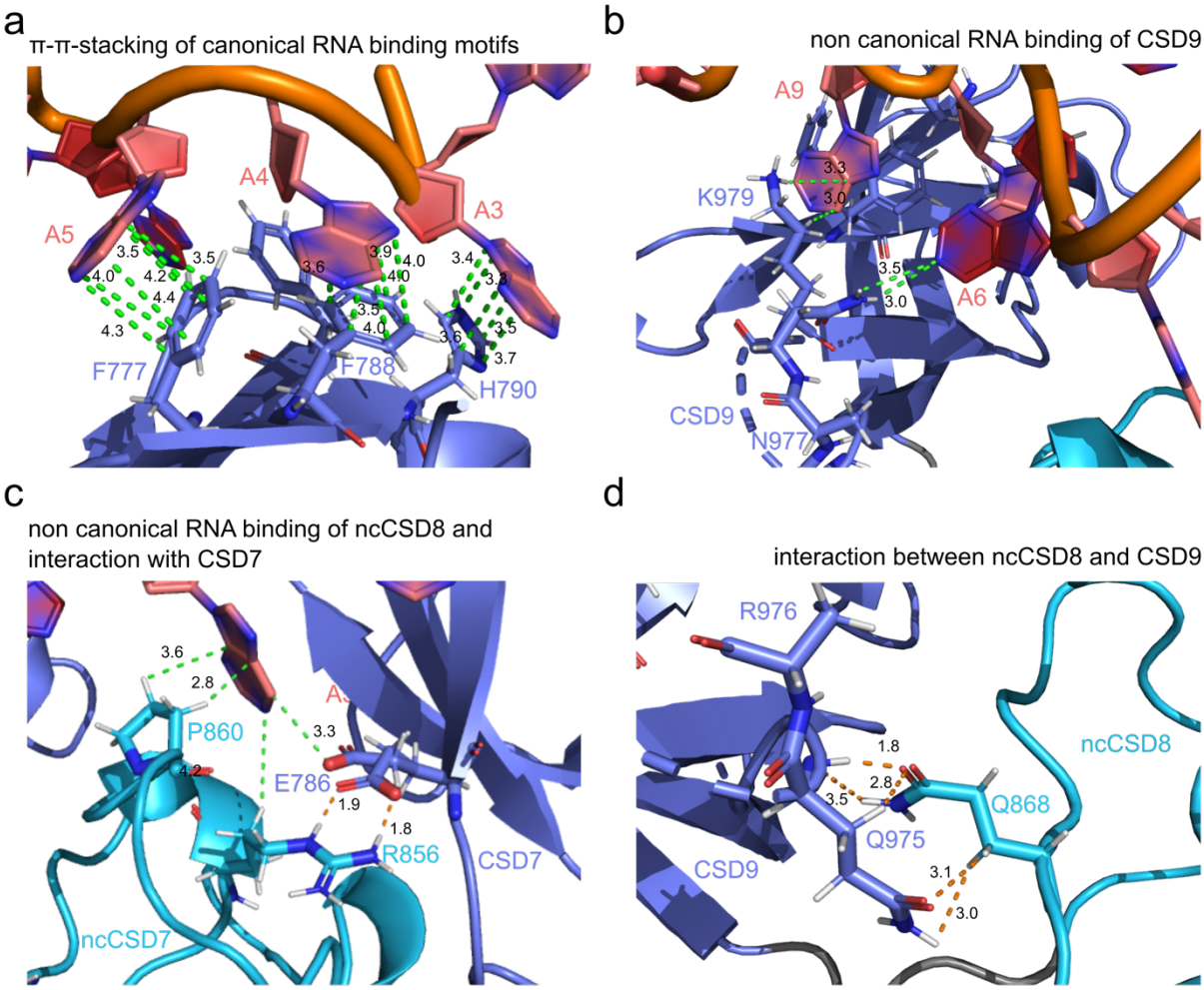


Figure 27: Canonical and non-canonical RNA binding and domain-domain interaction in the crystal structure of CSD789.

a: The RNA binding residues (F777, F788 and H790) within CSD7 form the CSD typical π-π-stacking of the RNA bases and the aromatic sidechains of the protein. **b**: Non-canonical binding residues of CSD9 (N977 and K979) interact with the RNA (A6 and A9). **c**: Residues of ncCSD8 (R856 and P860) interact with the RNA (A5) and CSD7 (E786). **d**: Two glutamines, one in ncCSD8 (Q868) and one in CSD9 (Q975), form a domain-domain interaction. Measurements show the distance between atoms of an amino acid and a nucleotide (green) or between two amino acids (orange). The RNA is colored in red, canonical CSDs in blue and the ncCSD8 in cyan.

In addition, residues located in ncCSD8, that were already identified as RNA binding residues in NMR titrations experiments (section 4.2.; Figure 15c and 15d) formed the following contacts: Nucleotide A6 points into the interaction surface of CSD7 and CSD8 and directly interacts with P860 and R856 of ncCSD8. The latter residue is also directly involved in an interaction with

two residues of CSD7 (E785 and E786), which in turn also interacts with the RNA in a sequence specific manner and results in sandwiching of the RNA between the two domains (Figure 27c).

Further tight domain-domain interactions are formed by glutamines in CSD8 (Q868) and one in CSD9 (Q975) that point to each other and form van der Waals contacts (Figure 27d). The fact that the joint tumbling is strengthened in presence of RNA may be due to conformational changes that happen upon binding, bringing the residues closer together and thereby resulting in a stable interface formation. This would also explain the restricted flexibility observed already in the absence of RNA (section 4.3.; Figure 19b, 19c and 19d). Potentially there is already a weak interaction between the two residues that is not strong enough to fix the two domains completely, but keeps them through ongoing on and off binding in close space.

To examine how much each non-canonical residue contributes to RNA binding and thereby also to validate the X-ray structure-based model, several mutants were tested for their binding affinity in fluorescence polarization assays. Mutations were generated to disturb the non-canonical RNA binding of CSD9 (N977A and K979A; Figure 27b), of ncCSD8 and its interface interactions to CSD7 (E786A, R856A and P860A; Figure 27c) and the interaction between ncCSD8 and CSD9 (Q868A, Q975A and R976A; Figure 27d). The mutated residues were located in loop regions without secondary structure elements to avoid misfolding of the individual domain fold. High yield and solubility during the purification process and the high peak dispersion in $^{1}H,^{15}N$-HSQC spectra (Figure 28) gives strong confidence that the overall domain fold of CSDs is not disrupted, making the binding affinity measurements meaningful.

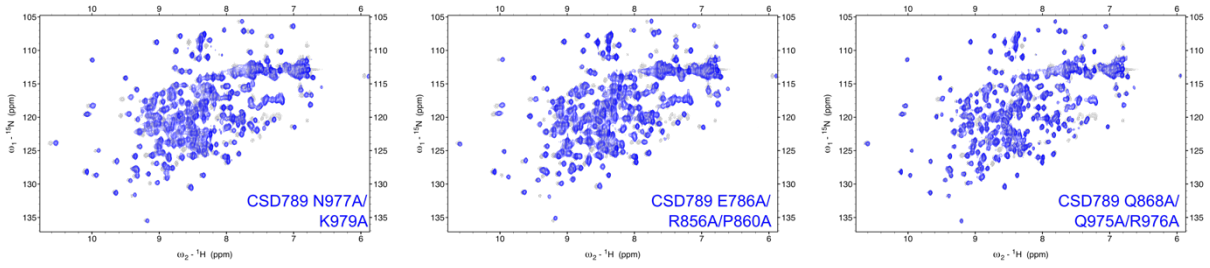


Figure 28: Inserted mutations do not disturb the overall protein shape of CSD789.
$^{1}H,^{15}N$-HSQC spectra of CSD789 wild type (grey) overlaid with the spectra of the different mutants (blue). (**a**: N977A/K979A, **b**: E786A/R856A/P860A and **c**: Q868A/Q975A/R976A). All measured proteins had a concentration of 100 µM.

The binding curve of the wild type protein results in a K_D of 22.0 ± 3.8 µM to the A15mer and of 18.1 ± 6.3 µM to the SL6a oligonucleotide (Figure 29a, 29b, 29c and 29d), giving comparable values to K_Ds derived from EMSAs (Figure 14b). All tested mutants show weaker binding to RNA. The mutant having a complete substitution of non-canonical RNA binding residues in ncCSD8 and its interface interactions to CSD7 (E786A/R856A/P860A; orange; Figure 29b and

29d) shows a more than 4-fold weaker binding affinity with a K_D of around 94.7 ± 32.8 µM to the A15-mer and of 114.0 ± 31.7 µM to the SL6a RNA compared to the wild type protein. Already the single mutations show a decreased binding affinity to the A15-mer RNA (E786A: 49.5 ± 9.3 µM, R856A: 72.0 ± 12.8 µM and P860A: 65.9 ± 12.2 µM; Figure 29b), meaning that each mutated amino acid contributes to the RNA binding of CSD789.

The binding affinity of the other two interface mutants decreases even stronger. Here the K_D drops more than 10-fold. Exact values cannot be calculated, since the binding does not even reach saturation with the measured protein concentrations (N977A/K979A: >1300 µM and Q868A/Q975A/R976A: >900 µM for both RNAs; Figure 29a, 29b and 29d). As observed previously, also the single mutants bind the RNA significantly weaker than the wild type protein (N977A: 71.4 ± 17.1 µM, K979A: 159.7 ± 65.3 µM, Q868A: 61.7 ± 15.4 µM, Q975A: 37.8 ± 9.7 µM and R976A: 158.5 ± 60.5 µM; Figure 29a and 29c), indicating again that all tested mutations influence the RNA binding.

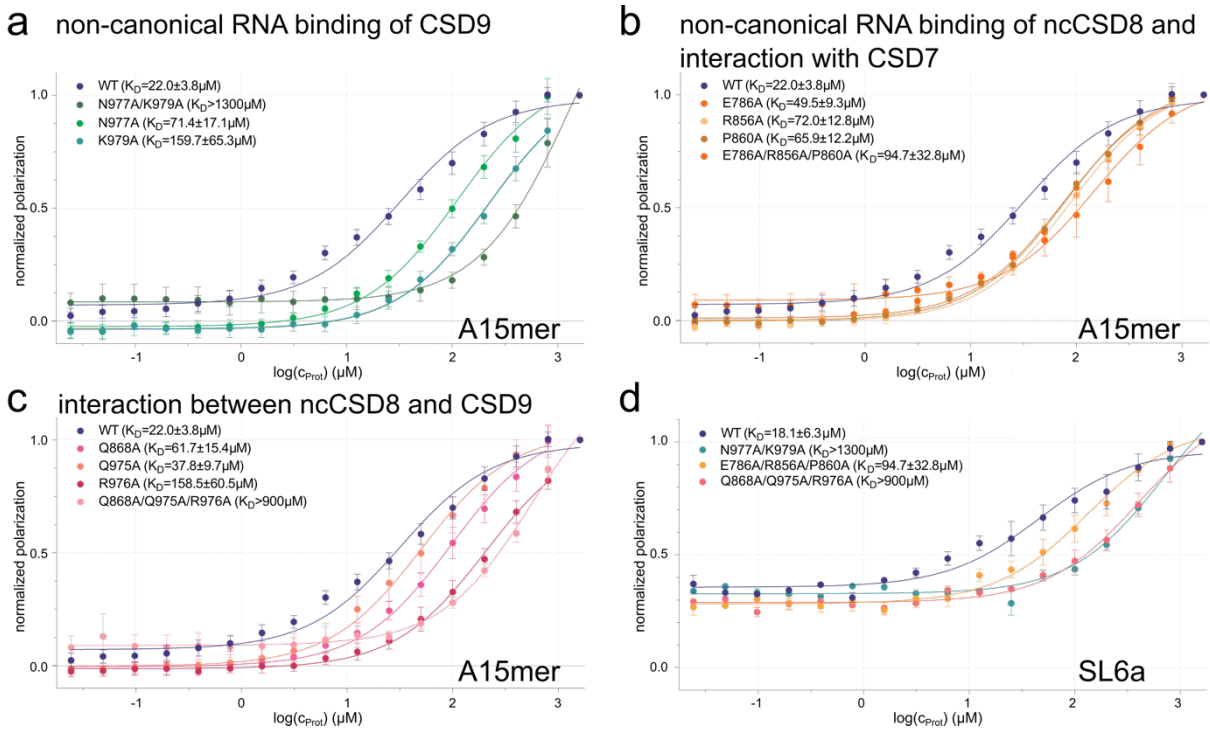


Figure 29: Effect of different mutations within CSD789 on RNA binding.
The fluorescence polarization binding curves of CSD789 wild type (blue) and different mutants to an A15mer RNA and SL6a (**d**) is shown. The mutations disturb the non-canonical RNA binding of CSD9 (green) (**a**), the non-canonical RNA binding of ncCSD8 and interaction with CSD7 (orange) (**b**) and the interaction between ncCSD8 and CSD9 (red) (**c**). The errors represent the standard deviation of three independent experiments.

Taken together the data show, that all the described structural peculiarities of CSD789, namely the non-canonical RNA binding within CSD9 and ncCSD8, just as the interface formation

between CSD7 and ncCSD8 and ncCSD8 and CSD9, contribute to the RNA binding affinity of CSD789. Thus, the previously observed synergistic binding between CSD7 and CSD9 in CSD789 (section 4.2. Figure 14 and Figure 25) could be explained by this structure and is further validated by the fluorescence polarization assays. Several non-canonical residues that are directly involved in RNA binding or indirectly via the formation of a domain-domain interface, contribute to strong RNA binding that is based on many tight residue interactions. This exemplifies also the complexity of RNA binding by a multidomain RBP to its target RNAs. To assess the importance of these non-canonical binding residues on the function of the full-length protein, cellular studies need to be conducted in the future.

4.7. Quantitative proteomics identifies the Unr interactome in a cellular context

To identify the *Drosophila* Unr interactome and to map its RNP composition, a protein immunoprecipitation assay followed by comparative quantitative mass spectrometry (MS) for V5-tagged Unr in SL2 cells after depletion of endogenous Unr was performed. To distinguish between direct and RNA-mediated protein-protein interactions, samples with and without RNase treatment were prepared.

Unfortunately, analysis of a WT pull-down sample against an empty vector control sample yielded in an enrichment of more than 1000 proteins (Figure 30). To overcome this limitation, we made use of the fact that the interdomain mutant (456-78ID) shows lower expression levels compared to the WT and thereby inevitably leads to lower Unr concentrations in the pull-downs (Figure 21c and 31a). Consequently, the number of significantly enriched protein hits in the single samples decreased (Supplementary Figure 7a and 7b). Based on their individual enrichment profile these hits were divided into five different clusters (Figure 31b and 31c). Enriched protein hits for the mutant sample against the wildtype, especially in cluster 5, were subunits of the proteasome, which may indicate that mutant Unr has a severely affected stability and is targeted by the protein quality control system, explaining the observed lower protein levels. The lower expression levels make it also impossible to claim that protein hits that are depleted in mutant Unr were pulled down due to lower affinity binding for the mutant with its increased conformational heterogeneity versus the wild type protein.

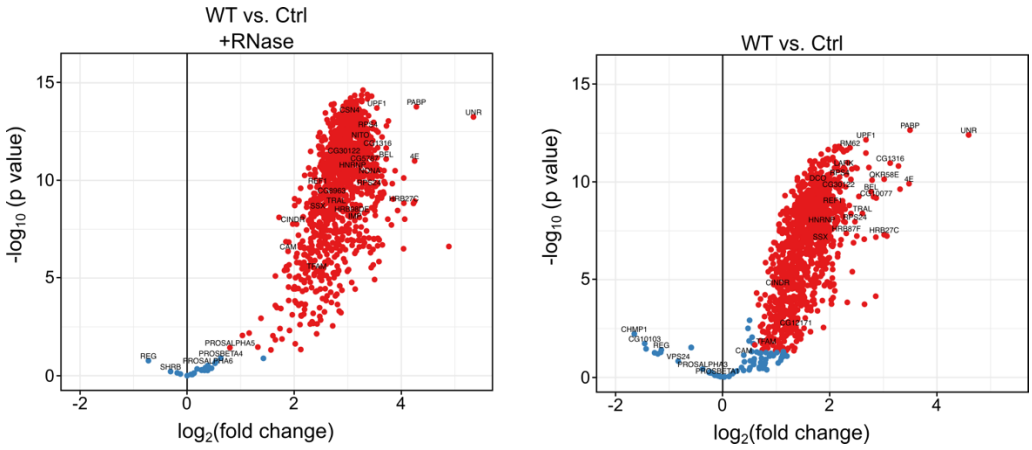


Figure 30: Protein enrichment of Unr pull-down samples against the mock control.
a: Volcano plots showing the difference between the WT Unr samples and the control with (upper graph) and without RNase treatment (lower graph). The log_2(fold change) is plotted against the negative log_{10}(p value). Differentially regulated proteins are colored in red. N=3.

Only 32 candidates are enriched in the WT, when focused on Unr interactors that are decreased in the 456-78ID background compared to WT (Figure 31b, 31c and 31d; cluster 1 and 2). Consequently, our strategy to use the mutation in order to bring down the Unr levels was successful and enabled us to define a Unr interactome and the composition of Unr RNPs. Interestingly, all of the 32 protein hits are RBPs and some of them are even well-known regulators of translation or mRNA degradation, which is also the best characterized function of Unr. Notably our analysis identified the poly-(A)-binding protein (pAbP), a previously reported binding partner of Unr[87,125]. The fact that this sample is enriched even in RNase treated samples indicates that this interaction is RNA-independent. On the contrary, most of the other targets show reduced binding after RNase treatment, indicative for a binding that is RNA mediated (Figure 31d). A well-established system of an RNA-dependent interaction is the binding of Sxl and Unr on the *msl2* mRNA[27,112,113,117]. One identified interaction partner in the pull-down analysis is Sister-of-sex-lethal (Ssx), which is a homolog of Sxl and known to exhibit comparable *msl2* mRNA binding activities, while incapable of engaging Unr directly[269,270]. The consensus of our results with previous studies demonstrates the reliability of our data. Additionally, there are further hits of special interest, like Hrb27c (Hrp48), another previously identified necessary component of the *msl2* mRNA translation repression complex[124].

Many of the other identified interaction partners could be subdivided based on their characterized functions. Some proteins play a role in RNA degradation (Upf1[271], and Edc3[272]), miRNA processing (Ago1[273]), RNA localization (Imp[274], tral[275], and Ref1[276]) and also splicing (Rm62[277,278], Nito[279], and Rump[280,281]). Strikingly, some of the identified targets (4E[282,283], tral[282], Hrb27C[284] and Imp[285]) are already characterized pAbP interaction partners, leading to

the suggestion, that these protein may be part of a larger RNP complex. The STRING analysis shows a potential network of known interactions between most of the hits (Figure 31d).

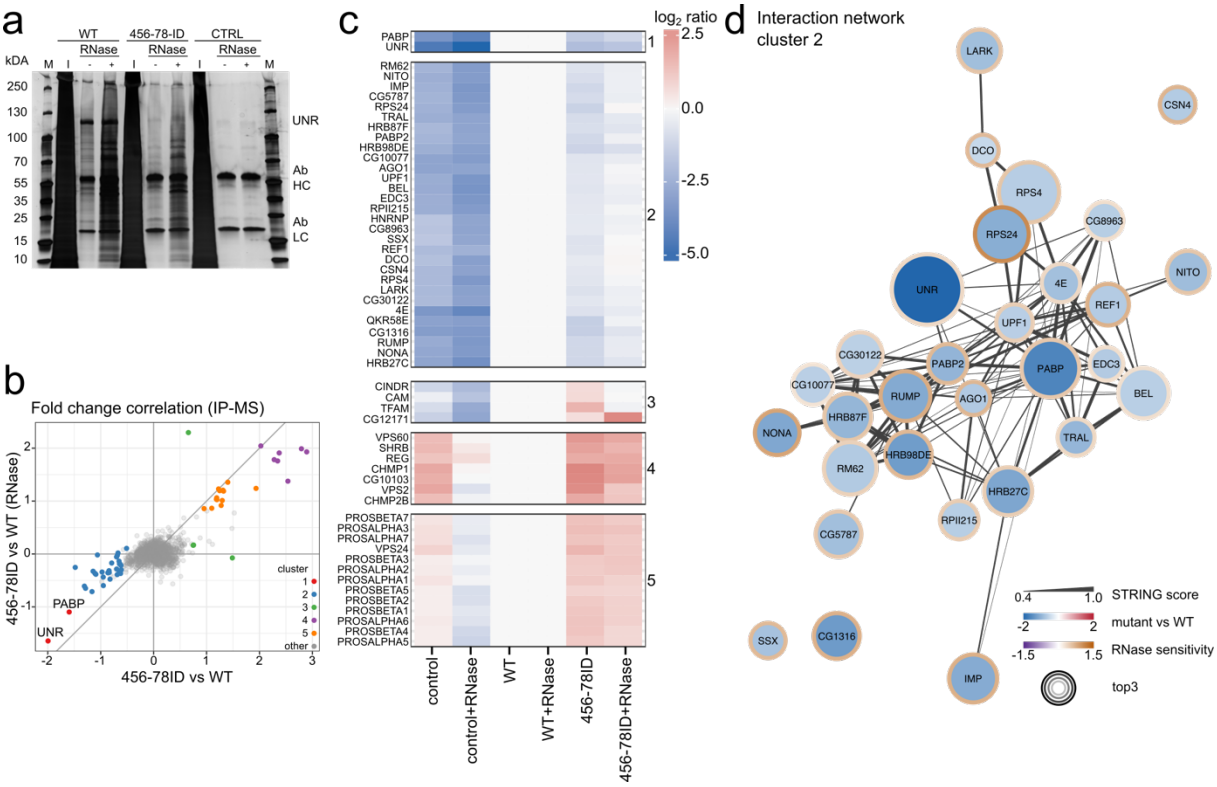


Figure 31: Proteomics data to characterized the *Drosophila* Unr interactome.
a: A silver stained polyacrylamide gel, showing the input (I) and elution samples with (+) and without (-) RNase treatment for wild type (WT), 456-78 ID and the empty vector control samples, that were used for the proteomics analysis. Unr, the antibody heavy chain (Ab HC) and the antibody light chain (ab LC) are labeled at the side of the gel. **b**: Scatter plot showing the correlation of the interdomain mutant vs WT (x-axis) and the RNase treated interdomain mutant vs RNase treated WT (y-axis). Colors indicate the cluster number from Figure 31c; N=3. **c**: Different protein targets of WT and mutant protein samples are represented in a heat map. Proteins were clustered in 5 clusters using kmeans algorithm. The RNase treated samples were normalized to the RNase treated WT sample and the non-treated samples were normalized to the non-treated WT sample. Red indicates upregulation and blue downregulation; N=3. **d**: A STRING interaction network, showing possible interactions between the targets that were enriched for the WT shown in cluster 1 and 2 of Figure 31c. The line thickness represents the STRING score, which represents the strength of data support for the individual interactions[286]. The inner color of each circle represents the fold-change between interdomain mutant and WT sample, the border line color indicates the difference of enrichments of interdomain mutant vs WT between RNase treated and non-treated samples (RNase sensitivity) and the size indicates the average abundance of a protein in the IP mass spec run (top3 value) which correlates with the pulldown efficiency.

Taken together, by using the mutant protein as a binding decreased interaction control of *Drosophila* Unr in a proteomics analysis, insights into the Unr interactome and its specific role in several cellular processes and complexes could be gained. Future studies will show the exact composition of different Unr containing RNP complexes, which will help to understand the broad function of Unr within the cell.

4.8. Structure modelling of Unr CSD1-9

The lack of flexible regions between several domains of *Drosophila* Unr raised the question whether full-length Unr adopts a specific shape. A full-length model would also provide a better understanding of how Unr might interact in a larger RNP complex. Thus, we set out to obtain a structural model of almost full-length Unr, excluding the flexible N-terminal region (residues M1-S175) (Supplementary Figure 8), which then represents Unr CSD1-9. As an input, we used all four high-resolution structures described above (CSD12, CSD456, CSD78 and CSD9) and a homology model of CSD3, which is based on the NMR structure of human CSD3[127]. The rigid domain distances and orientations determined above for the constructs CSD456 and CSD78 were kept fixed during structure calculations. The remaining linker regions were randomized and thus allowed to adopt different conformations in each structure calculation (Figure 32a).

The resulting structural ensemble of 5000 structures covered a large conformational space (Figure 32b and 32c, Supplementary Figure 9a). To further refine the model, we imposed additional restraints based on the experimental SAXS data for CSD1-6 and CSD4-9 (Fig. 32b, 32c, 32d and Supplementary Figure 9c). Full-length/CSD1-9 SAXS data was unfortunately of insufficient quality due to aggregation and radiation damage. Although several structures fitted different SAXS curves with a low χ (Supplementary Figure 9a), the whole ensemble showed only two structural arrangements that fitted all curves equally well, having χ values for both curves in the lowest fifth percentile (for CSD1-6: 2.973 & 2.172 and for CSD4-9: 2.172 & 1.751) (Supplementary Figure 9a). These two structures and a chimeric structure that was generated with the best fits of each SAXS curve share an overall shape, but differ in their domain orientation (Fig. 32c). Taken together, the currently available data suggests that Unr likely adopts multiple conformations restricted towards branched structures. Fully extended or overly elongated structures, as well as compact structures, do not fit the experimental data (Figure 32b and Supplementary Figure 9b). Therefore, the final ensemble provides insight into potential RNA binding domain orientations in the full-length protein context. Although some flexibility is maintained, the domains show substantial restriction with regards to domain-domain orientation in space. These results suggest that ncCSDs convey RNA/RNP target selectivity by scaffolding Unr into limited shape diversity. As the location and distribution of RNA binding surfaces suggests (Supplementary Figure 9b), full-length Unr forms an RNA binding platform poised to recognize only a subset of possible RNA tertiary structure arrangements.

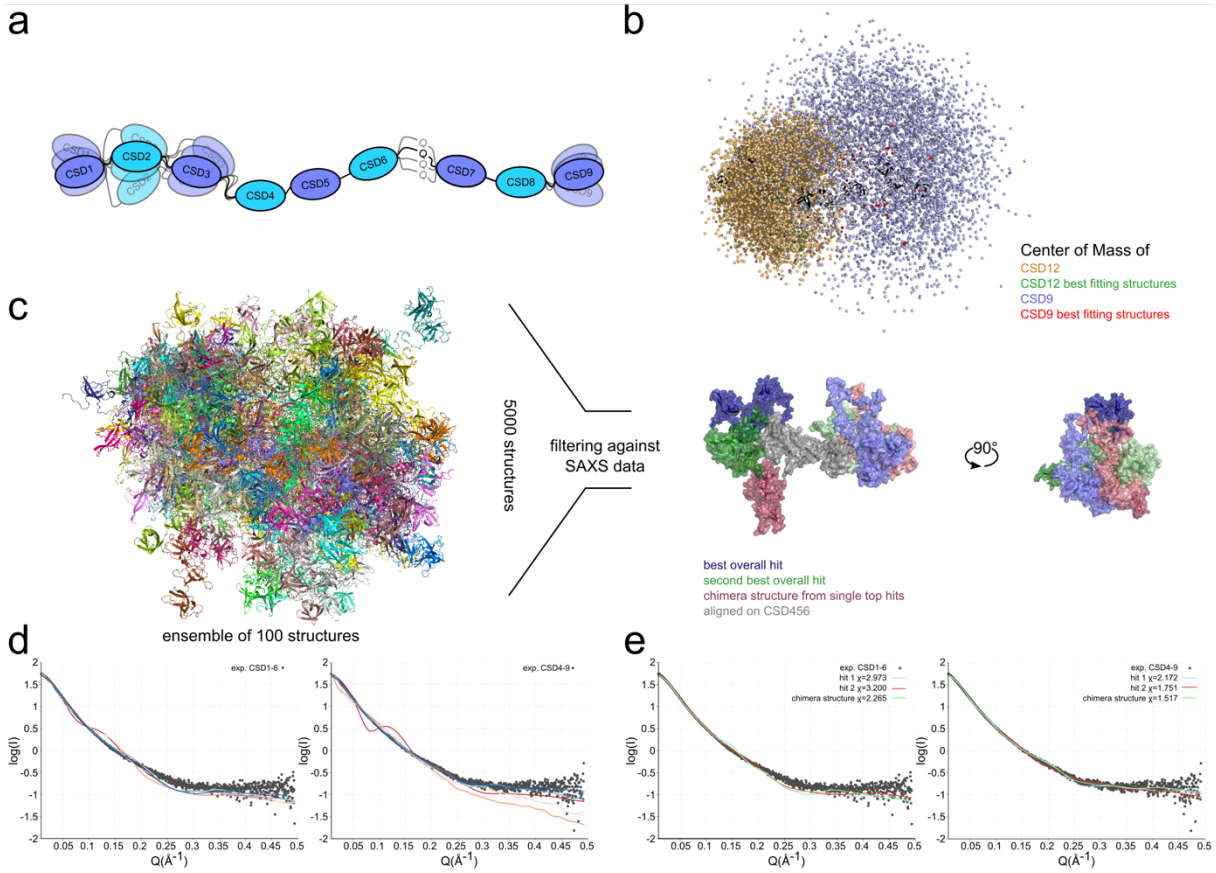


Figure 32: Ensemble modeling of Unr CSD1-9.

a: Schematic representation of single domains and linkers. Flexibility is indicated by multiple linkers with different shading, and flexible domain orientations are indicated by shaded, multiple domains in different orientations. **b**: The center of mass from CSD12 (orange) and CSD9 (blue) from each of the randomized structures are shown. In black is represented one exemplary structure from CSD1-9. The structures that fit the SAXS curves in the lowest 5th percentile are highlighted in green (CSD12) and red (CSD9), showing that the calculated structures cover a large conformational space and that the best fitting structures are restricted to certain areas. **c**: A representative ensemble of 100 structures out of 5000 covering a large conformation space. All structures are superimposed on CSD456 (left). The two best fitting structures and a chimera structure of the best hits of each SAXS curve are shown (right). **d**: The fit between back-calculated and experimental scattering densities for some structures against SAXS data of CSD1-6 (176-677) (left) and CSD4-9 (428-990) (right) are shown. For illustration, structures have been chosen that fit and do not fit the SAXS data well. **e**: The SAXS fits of the best fitting structures is shown (blue: overall best hit; green: second best overall hit and red: chimera structure). CSD789 is colored lighter. All models are superimposed on CSD456, which was kept rigid in the calculation and is colored grey in each structure.

To improve the modeled structure and to potentially obtain ambiguous long-range interaction restraints between domains for which information are so far lacking, $^{1}H,^{13}C$ methyl heteronuclear multiple quantum correlation transfer (HMQC) spectra, for detection of all $^{1}H,^{13}C$ moieties within the sample were recorded for CSD1-6 and CSD4-9. Due to the large size of the protein constructs (~60 kDa), only the methyl groups of isoleucine, leucine and valine (ILV) were ^{13}C labeled with specific precursors in an otherwise ^{15}N labeled-deuterated-^{12}C sample according to Tugarinov and Kay[177]. This labeling scheme removes most of the hydrogens and thereby reduces cross-relaxation effects, which in turn

increase line-broadening beyond detection. This effect gets worse the larger the object of study due to the slower molecular tumbling. Additionally, specific ILV methyl group labeling reduces peak overlap due to a reduction of signal number and allows to acquire decent spectra of even large-sized proteins not amenable to classical biomolecular NMR[287]. By tracking CSPs that occur between the larger protein constructs (CSD1-6 and CSD4-9) and the shorter ones (CSD12, CSD456 and CSD789), potential long-range interactions can be detected. For CSD12, CSD78 and CSD9 the assigned methyl groups of isoleucine, leucine and valine, which was done on a heterogenous labeled $^{1}H,^{15}N,^{13}C$-sample for the structure determination, were transferred to the ILV $^{1}H,^{13}C$-HMQC spectra. However, this transfer was not possible for all peaks as it is especially difficult in crowded spectral regions. Furthermore, the sample of CSD12 for structure determination was RNA bound. Thus, the transfer of several CSD12 peaks was not unambiguously possible. These were left out of the analysis.

The side chains for ncCSD6 were assigned using conventional HBHA(CO)NH, HCCH-TOCSY, and CCH-TOCSY spectra of an isolated ncCSD6 sample. However, in order to gain information about ncCSD4 and CSD5 a sidechain assignment of CSD456 would be necessary. Due to its comparably large size for NMR (>30kDa) the signal-to-noise is too low for 3D assignment experiments. Transverse relaxation rates (R_2) of molecules linearly enhance with increased molecular weight, leading to broad signals and poor spectral quality. But not only the large size, also a concentration limit of 300 µM will make an assignment of these sidechains difficult. The assignment transfer yielded an ILV methyl group assignment for CSD1-9 of 44 %.

Overlaid spectra of CSD12 and CSD456 onto CSD1-6 and CSD456 and CSD789 onto CSD4-9 respectively, showed clear CSPs between the longer and the shorter constructs (Figure 33a and 33j), indicative of long-range interactions between CSD12 to CSD456 and CSD456 to CSD789. Interestingly, a mapping of the CSPs from assigned residues on the sequence of the multidomain constructs CSD12 and CSD789, restricts these changes with only one exception within ncCSD2 to the canonical CSDs (Figure 33b and 33h). However, several significant CSPs are also located within ncCSD6. The changes between ncCSD6 and CSD1-6 are mainly located in the N-terminal part of the domain and more numerous compared to the changes to CSD4-9 (Figure 33c), where only two residues seem to be affected from long-range interactions (Figure 33g). Also, several residues from ncCSD4 or CSD5, that also show shifts (exemplified in the zoom of Figure 33a), seem to contribute to these interactions and future assignments will hopefully identify these.

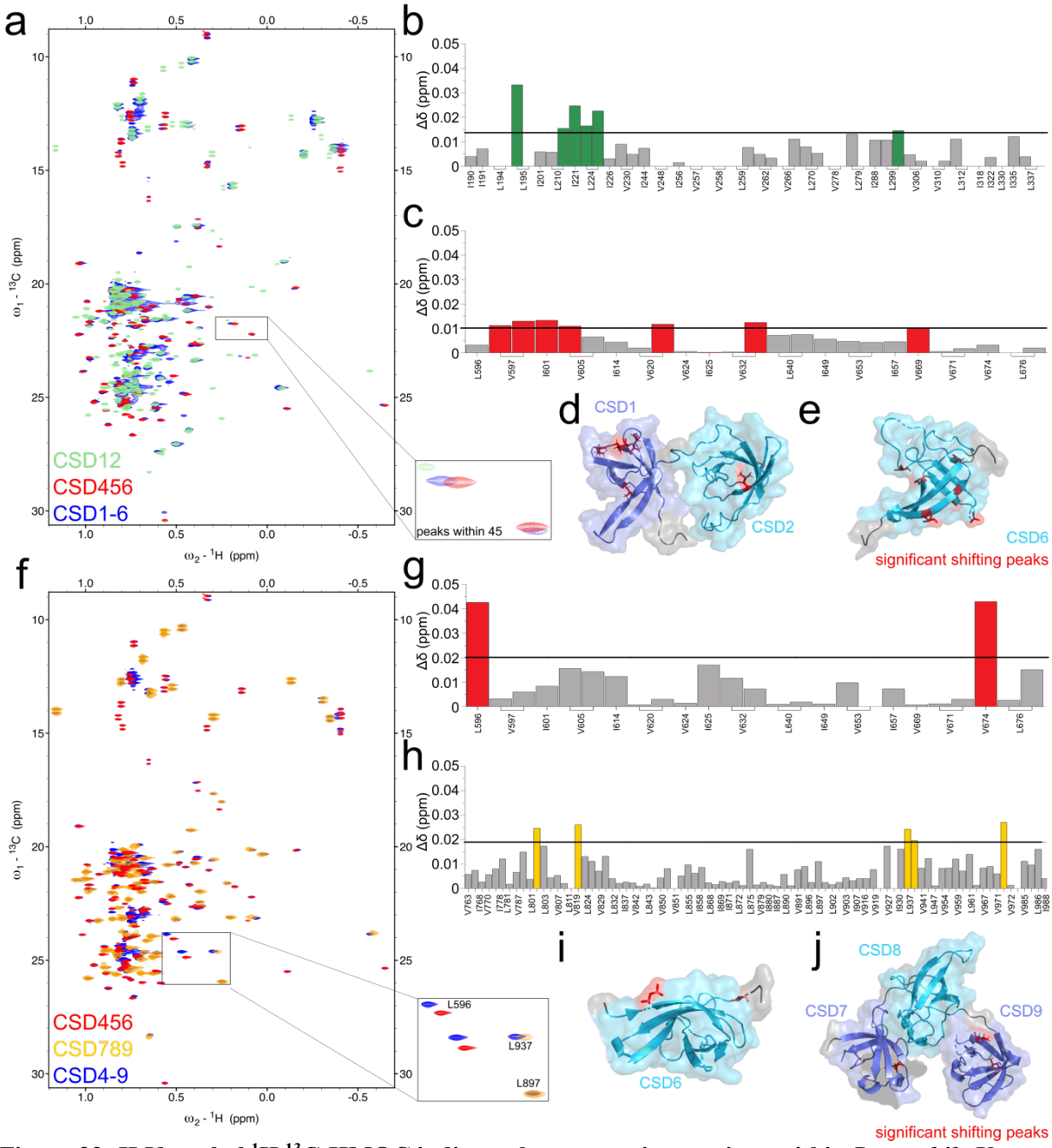


Figure 33: ILV methyl $^{1}H,^{13}C$-HMQC indicates long range interactions within *Drosophila* Unr.
a: Overlaid spectra showing the $^{1}H,^{13}C$-HMQCs of CSD12 (green), CSD456 (red) and CSD1-6 (blue). The zoomed in region shows shifts of non-assigned residues of ncCSD4 or CSD5. Additional weak peaks observed in CSD12, may result from a contamination of the sample with traces of CSD789. The chemical shift perturbations between CSD1-6 and CSD12 (**b**) and CSD456 (**c**) are plotted on the sequence. The significant shifts are highlighted on the PDB structure in red (CSD12: **d**; CSD456: **e**). **f:** Overlaid spectra showing the $^{1}H,^{13}C$-HMQCs of CSD456 (red), CSD789 (orange) and CSD4-9 (blue). The zoomed in region shows residues within ncCSD6 (L595), ncCSD8 (L897), CSD9 (L937) and one non-assigned of ncCSD4 or CSD5. The chemical shift perturbations between CSD4-9 and CSD456 (**g**) and CSD789 (**h**) are plotted on the sequence. The significant shifts are highlighted on the PDB structure in red (CSD456: **i**; CSD789: **j**). Residues with significant CSPs were defined to have the average shift plus the standard deviation of all measured shifts, except for CSD789 were the average plus twice the standard deviation was used[86]. The cut-off is highlighted with a black line in each CSP plot.

To assess the locations that may be affected from long range interdomain interactions, all residues harboring significant CSPs were highlighted on the protein structures (Figure 33d, 33e, 33i and 33j). Unfortunately, the changes are not clearly restricted to certain areas on the structures, making a meaningful analysis difficult. One could think of neglecting the residues that are involved in the core formation of the domains, since CSPs here are likely due to indirect allosteric effects, or comparing spectra between the shorter constructs and an even longer almost full-length construct (CSD1-9). However, despite all difficulties in the analysis of these data, at least transient interactions are present between the N- (CSD12) and C-terminal part (CSD789) of *Drosophila* Unr to the middle of the protein (CSD456).

Altogether the collected SAXS and NMR data on longer Unr constructs indicate that there remains flexibility in *Drosophila* Unr, which will make it difficult to determine a high-resolution structure of the isolated full-length protein using conventional methods like crystallography or even cryo-EM. Nevertheless, these data also show, that this flexibility is spatially restricted (Figure 32). Besides the identified direct and strong domain-domain interactions within the protein, also more transient long-range domain interactions (Figure 33), are involved in specifying the function of the protein.

4.9. Towards structure determination of an mRNP complex structure: Interaction of Unr with pAbP

The insights obtained in this study on the structure of full-length *Drosophila* Unr (section 4.8.), on the CSD789-RNA high-resolution structure (section 4.6.), on further RNA binding properties (section 4.2., 4.4. and 4.6.) and on Unr's potential interaction partners identified in the proteomics analysis (section 4.7.), convinced us that structure determination of an entire mRNP complex assembled on parts of the 3' UTR of *msl2* mRNA is soon within reach. This would give unprecedented structural information on translation regulation at high-resolution. The remaining flexibility of the protein may be further decreased when acting in a complex, where different domains are bound by both, protein or RNA interactors, leading to a rigidification of the whole complex and making it amenable to for cryo-EM. Further optimism is allowed as this complex would be a very tight binding complex, as even small amounts of Msl2 protein are lethal in female flies[288]. To reach this goal, further detailed insights into interaction interfaces of different partners could be helpful.

The top hit in our proteomics pull-down assay was pAbP, which is besides Unr, Sxl and Hrp48 known to be bound to the 3' UTR of *msl2*[87,113,119,124,125]. This interaction even seems to be RNA independent. A crystal structure showing the interaction of a trimeric Unr

CSD1-Sxl RRM12-*msl2* mRNA complex was already solved[27] and we decided to start characterizing the interaction between Unr and pAbP in more detail to identify further components to be included into the assembly of a translation repressor complex for structure determination.

PAbP is a phylogenetically conserved protein with crucial functions in translation and mRNA stability as demonstrated by P-element insertions in *Drosophila melanogaster* being embryonic lethal[289]. The roughly 630 amino acids long protein is composed of 4 RNA-recognition-motif domains (RRMs), that are separated by short linkers in the N-terminal part of the protein, followed by a long linker region and a C-terminal Poly(A)-binding protein C-terminal (PABC) domain (Figure 34a). It is known that the four RRMs bind the poly-A tail of mRNAs[290,291] and that the PABC domain interacts with proteins harboring a conserved motif of 12-15 amino acids, named pAbP- interacting motif (PAM)-2[292], which both leads to translation upregulation and protection against mRNA decay[293].
For initial interaction studies NMR was chosen, as it is sensitive to small structural changes and thereby recognizes even weak interactions. ^{1}H,^{15}N-HSQC spectra of different protein constructs of Unr (CSD12, CSD456 and CSD789) were acquired in the absence and presence of different pAbP constructs (linker PABC and RRM3; Figure 34 and 35).

For the titration of linker PABC with 1.5 molar excess of different Unr constructs, interaction was solely observed for the N-terminal Unr constructs (CSD12 and CSD123, Figure 34a, 34b, 34c and 34d). Addition of CSD456 or CSD789 to linker PABC resulted in no differences, neither CSPs nor intensity decrease of peaks ^{1}H-^{15}N-HSQC spectra (Supplementary Figure 10a and 10b). Already during sample preparation differences were observed for CSD12. Here, addition of any linker PABC construct, even at low concentrations, resulted in a turbidity of the solution and SDS-PAGE showed that this was due to precipitation of CSD12. Although the precipitation of one of the proteins indicates a certain reaction within the sample, it also makes it difficult to analyze a possible protein-protein interaction, as for most methods a certain sample quality of both proteins is required.

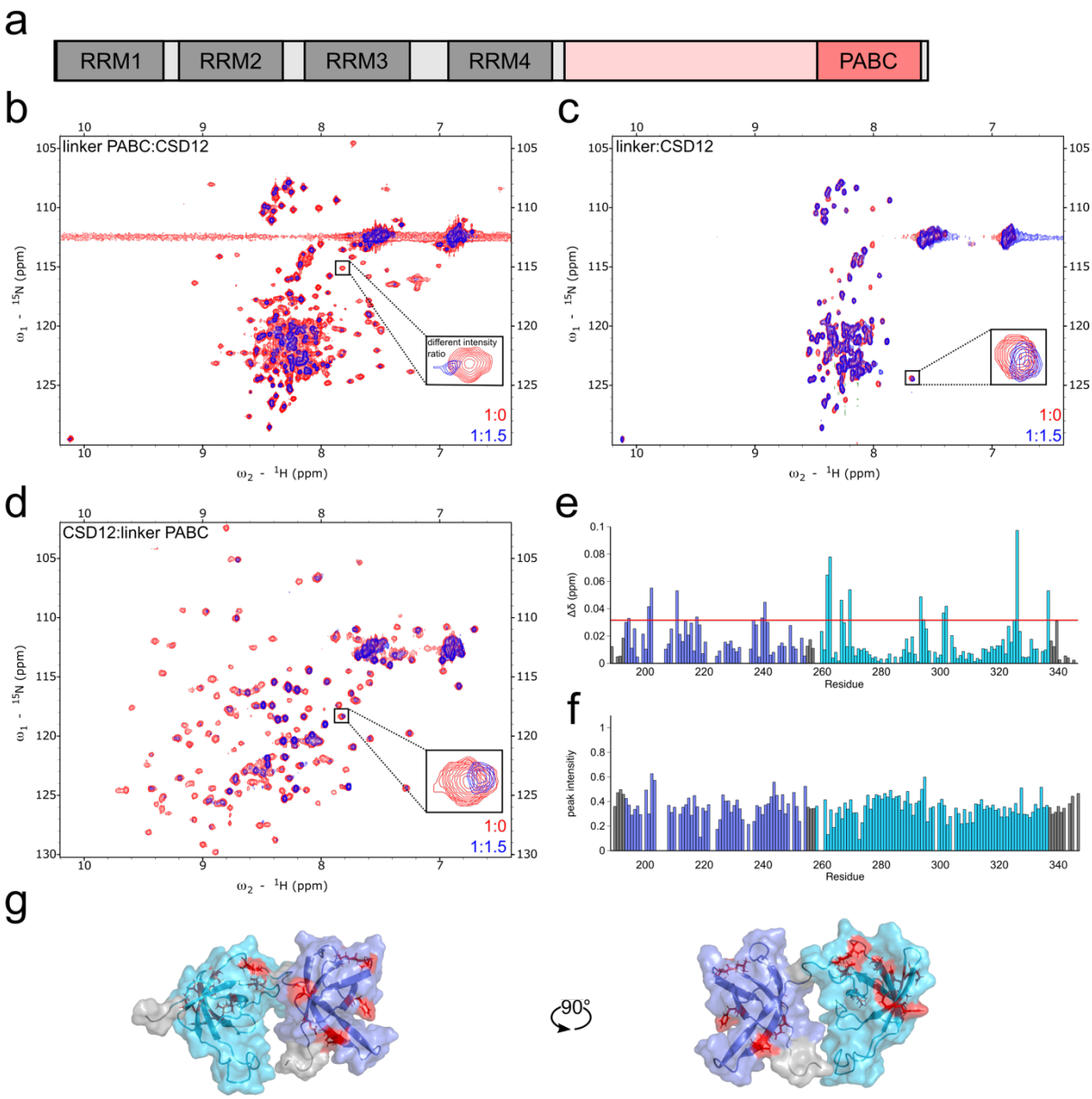


Figure 34: Protein-protein interaction mapping between *Drosophila* Unr and pAbP PABC domain.

a: Domain arrangement of *Drosophila* pAbP, showing the four N-terminal RRM domains and the C-terminal PABC domain. The boundaries are according to the Uniprot entry (ID: P21187). The part of the protein that was used for interaction studies is highlighted in red. **b/c/d**: Overlaid $^{1}H,^{15}N$-HSQC spectra and zoomed in regions of linker PABC (363-634) in absence (red) and presence (blue) of 1.5 excess unlabeled CSD12 (**b**), of the linker only (363-551) without and with 1.5 molar excess of unlabeled CSD12 (**c**) and CSD12 without and with 1.5 molar excess of unlabeled linker PABC (**d**). The intensities of the two spectra of the zoom in from b are not identical. **e**: Chemical shift perturbation (CSP) plot of CSD12 after addition of the linker PABC construct. The red line in the CSP plots indicates the average plus the standard deviation of all measured shifts, which was used to identify significant shifts[86]. **f**: relative peak intensities of the bound sample compared to the apo form. **g**: Significant CSPs of the interaction are highlighted on the protein structure of CSD12 in red.

Nevertheless, due to its sensitivity, NMR revealed differences between the apo and the CSD12 plus ^{15}N labeled linker PABC sample. Both, signal loss and CSPs could be observed

(Figure 34b). Due to the missing backbone assignment it is impossible to say, which parts of the linker or the PABC domain are involved. However, some of the shifts are observed for the well dispersed peaks, indicating binding of CSD12 to parts of the structured PABC domain. To assess whether the linker contributes to the binding, a ^{15}N labeled linker construct was measured in presence and absence of CSD12 (Figure 34c). Just as for the longer construct, CSPs and intensity loss of some peaks, suggests, that besides the domain also parts of the linker do interact with CSD12. Backbone assignments of resonances will help to understand which parts are interacting on a residue level. The reverse titration of linker-PABC to a ^{15}N-labeled CSD12 construct confirms the previous observations. Since a backbone assignment for this construct was done previously, CSPs and intensity changes could be mapped to the sequence. The previously described precipitation could be one explanation for the intensity loss, but the additionally observed CSPs, indicate that parts of the loss may also be due to an interaction dependent increase of the molecular size of the measured object (Figure 34e and 34f). Mapping of the CSPs on the sequence and on the structure shows, that although the major changes are located within ncCSD2, also residues in CSD1 seem to be affected from the pAbP interaction. Since the linker PABC construct is relatively big (~40 kDa), it is imaginable that both domains are interacting with different parts of the protein. A measurement with separate linker or PABC constructs will reveal whether one of the CSDs might bind to the linker and the other one to the domain.

In a next step binding of Unr to one of the RRMs of pAbP, namely RRM3, was assessed (Figure 35a). As described above the interaction of pAbP-RRM3 to CSD12, CSD456 and CSD789 was analyzed using ^{1}H,^{15}N-HSQC spectra, observing CSPs and intensity changes of the apo protein and in presence of 1.5 molar excess of the ligand.

For CSD456 basically no interaction was observed, as demonstrated by negligible CSPs and no intensity loss (Figure 35c, 35e and 35f). The CSPs for CSD12 were slightly stronger compared to CSD456 and an intensity decrease between 20-30% was detected (Figure 35b, 35e and 35f). However, the global decrease in intensity for all residues presumably indicates some kind of unspecific interaction, mediated by an increased τ_c or due to viscosity of the sample. On the contrary, interaction between CSD789 and RRM3 is doubtless. The measured CSPs are much stronger than for the other two constructs, forming even two patches within the sequence, one in the N-terminal part and another one in the middle (Figure 35d and 35e). Correspondingly, an overall but also region-dependent signal loss can be observed, which is not only stronger than the one for CSD12 (>50%), but also matches the CSP data. Residues showing larger CSPs also show a stronger decrease in the peak intensity (Figure 35d and 35f).

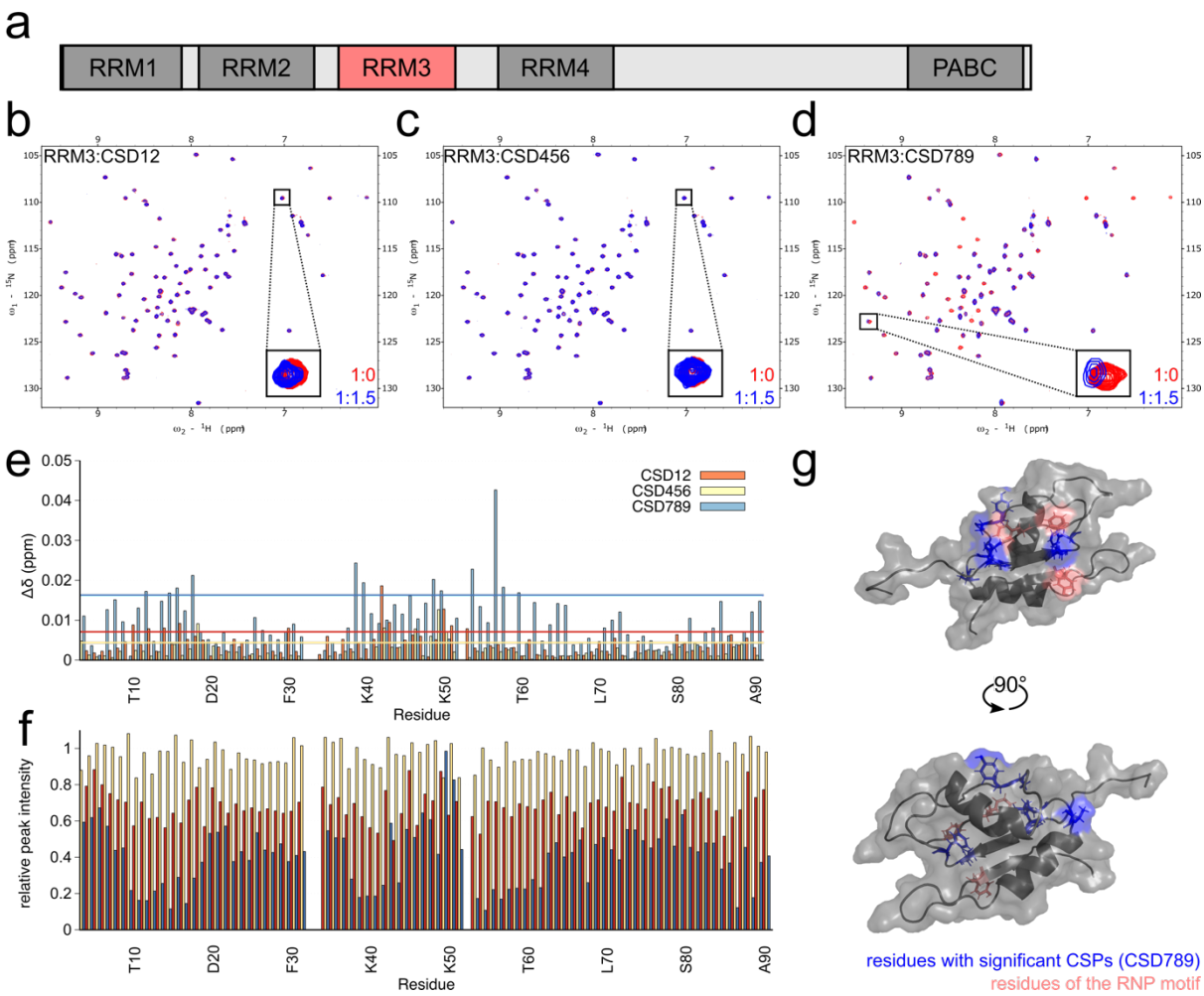


Figure 35: Protein-protein interaction between *Drosophila* Unr and RRM3 of pAbP.
a: Domain scheme of *Drosophila* pAbP. The part of the protein that was used for interaction studies is highlighted in red. **b/c/d**: Overlaid $^{1}H,^{15}N$-HSQC spectra of RRM3 (G176-A263) in absence (red) and presence (blue) of 1.5 excess unlabeled CSD12 (**b**), CSD456 (**c**) and CSD789 (**d**). **e**: Chemical shift perturbation (CSP) plot of RRM3 in addition of CSD12 (red), CSD456 (yellow) and CSD789 (blue). The colored lines indicate the average plus the standard deviation of all measured shifts for each interaction[86]. **f**: Relative peak intensities of the bound sample compared to the apo form. Red bars are in presence of CSD12, yellow ones of CSD456 and blue ones of CSD789. **g**: Significant CSPs of the interaction between CSD789 and RRM3 are highlighted in blue on the protein structure. Residues of the RNP motif are highlighted in pink.

The good sample and spectral quality of the protein enabled us to determine the solution structure of the RRM3 of pAbP using NMR. The resulting ensemble shows the typical RRM topology of β-α-β-β-α-β, with residues of the two RNP motifs pointing to the outside of the structure (Supplementary Figure 11a and 11b). Mapping of the significant CSPs on the structure reveals, that they are located exclusively on or close to the RNA binding surface (Figure 35g). Concomitant with this observation are the surface charges of the two protein constructs. The surface of the CSDs of *Drosophila* Unr are rather negatively charged (Result section 4.2), whereas the interaction surface of RRM3 shows positive potential (Supplementary Figure 11c),

promoting an interaction between both. The charge difference of the proteins might also explain the unspecific interaction observed between CSD12 and RRM3.

Interestingly, the residues of the RNP motif are not directly affected upon binding to CSD789, making two opposite scenarios imaginable (Figure 35g). Either RNA binding is somehow cooperative between CSD789 and RRM3, that are pre-formed in a complex, or binding of CSD789 to the RNA interaction surface of RRM3 inhibits its RNA binding. Further competitive binding studies with the two proteins or CSD789 as separate domains and RNA, just as a reverse titration of ^{15}N labeled CSD789 with unlabeled RRM3 will help to understand the mechanism in more detail. Additionally, binding studies with the remaining three RRMs of pAbP and the different Unr constructs will be necessary to understand how both proteins interact in a full-length manner and in order to optimize constructs towards crystallography or cryo-EM.

Taken together these results indicate that the C-terminal domain of pAbP interacts with the Unr N-terminus in an RNA independent manner (Figure 34). The same was observed for CSD789 and one of the RRMs (Figure 35). Future studies, that map these interactions on a residue level, and screening of an extended construct pool might show how crucial these are for translation regulation of the *msl2* mRNA and may elucidate how other proteins are linked to the interaction and form a large translation regulation mRNP complex.

5. Discussion

A prerequisite to obtain a complete understanding of the function of an RBP and a mechanistic description of how this influences RNA fate, is to gain information on its RNA recognition code. In order to increase the target specificity RBPs often engage multiple RBDs, since single domains only recognize short RNA sequences specifically[294]. Up to date a limited number of high-resolution structures of multidomain RBPs is available. Some of these structures show an increased sequence recognition platform through domain interaction, like in the case of the two RRMs of Hrp1[265] or the two KH domains of NusA[295]. In the case of two RRMs of PTB, the protein-RNA interaction, just as interdomain interactions bend the RNA in a certain manner[296].

However only recently a study on IMP3 could show how simultaneous contributions on RNA binding of multiple KH domains leads to a recognition of clustered RNA elements[35]. This described for the first time mechanistically the specificity of a protein not only towards RNA sequence specificity, but also structural assemblies within the RNA fold. Thus, this study reveals the need for systematic experimental approaches, that study combinatorial RNA recognition by multi-domain RBPs. Methods like CLIP, RIP or RNAcompete[297–299], that were used so far, aim to decipher short consensus sequences, that often do not describe the complexity of multi-domain RBPs satisfyingly.

During the integrative structural and functional studies of the *Drosophila* multidomain RBP Unr presented in this work, several NMR and crystallography structures were solved, dissecting its multi-domain organization. These data can contribute to our understanding of how multi-domains act together in RNA binding.

5.1. Evolution of non-canonical CSDs

Of particular note is the discovery of the previously unpredicted non-canonical cold shock domains. These domains were called ncCSDs, as they lack RNA binding motifs, which are otherwise conserved in CSDs. On the other hand, the overall fold is maintained in ncCSDs. The absence of RNA binding residues was the reason for being unrecognized until now. Our Hidden Markov model based iterative search though the UniProtKB[244,254] revealed that several other proteins throughout different phyla contain ncCSDs often located close to canonical CSDs in an alternating fashion.

The high structural similarity and the frequent coappearance of CSDs and ncCSDs within one protein makes it likely that they evolved from a common ancestor domain, possibly by domain duplication[300]. However, from the underlying data it is not clear whether the canonical CSDs gained the RNA binding function and lost the flexible loop that is present in ncCSDs during evolution or vice versa.

The age of a domain family can be assessed by finding the largest clade of organisms within which it is found, since this lineage's root is likely the age of the domain family[301]. Domains that are identified in all three kingdoms of life, or at least in two of them, are likely to be ancient[302–304]. For example canonical CSDs are present in eukaryotes (6.000 hits), in bacteria (138.000 hits) and archae (2.000 hits), classifying them as a very early evolved domain type[305]. Our iterative search detected ncCSDs to be present in two of the kingdoms, namely in bacteria and in eukaryotes. However, due to lateral gene transfer or annotation errors, domain families can be found in a few species of a kingdom without actually belonging to it. To avoid misallocations several studies determined thresholds[304,306,307], since ncCSDs were mainly identified in eukaryotes, with more than 1000 hits, and only to a small proportion in bacteria (13 hits), it is likely, that these domains originated not too long ago. Further bioinformatics studies will be necessary to deepen this analysis and to assess how frequent ncCSDs are in which kingdom or even their distribution within a kingdom, in order to estimate a time of origin more accurately.

5.2. ncCSDs promote RNA structure specificity

Due to the lack of RNA binding residues, ncCSDs do not bind to RNA independently. However, by forming tight interaction networks with their neighboring canonical CSDs, they impact the overall protein stability and have an essential scaffolding function in Unr. The necessity of ncCSDs for the protein stability is also reflected in the initial construct screening experiments, where stability and solubility of canonical CSD-constructs was much higher in presence of C- and N-terminal extensions. Many of the single canonical CSDs were not soluble, when lacking their neighboring ncCSD. The restriction of flexibility imposed through the ncCSDs orients the RNA binding surfaces within the full-length protein in a special conformation with respect to each other. Thus, ncCSDs might have an essential role in RNA tertiary structure recognition, translation activity, as well as Unr RNP assembly and composition (Figure 36a and 36b).

Unr mutational studies in *Drosophila* embryo extracts and SL2 cells showed that a disruption of the interdomain network influences translation regulation. That means that either RNA

binding, RNP composition and/or RNP interactions with other molecular machineries are disturbed. Concomitant with these observations and the proposed scaffolding role of ncCSDs, a differential bound RNA interactome of the interdomain mutants compared to the wildtype protein was identified using RIP-seq experiments. This leads to the interpretation that ncCSDs contribute to RNA target specificity.

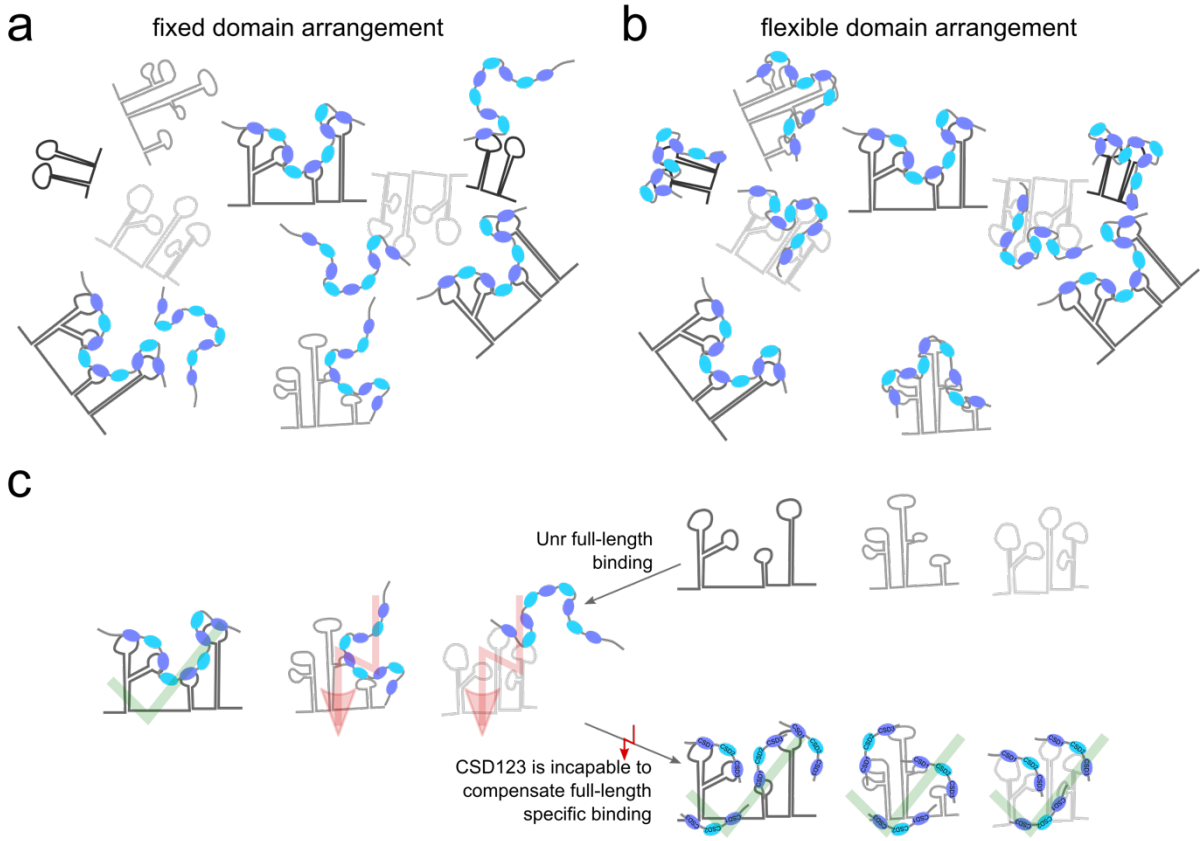


Figure 36: Restricted heterogeneity of *Drosophila* Unr contributes to RNA structure specificity. **a**: Restricted heterogeneity guarantees binding of all five CSDs only to specific RNA folds. **b**: A flexible domain arrangement would allow Unr to bind to different RNA structures, leading to less specific binding. **c**: Although harboring the same RNA binding domains, with similar sequence specificity CSD123 might be incapable to compensate for the target specificity of full-length Unr, due to high conformational heterogeneity.

Moreover, this hypothesis is strengthened by observations in previous studies. It has been shown that early fly development requires only the N-terminal half of Unr, however, later developmental stages are dependent on the full-length protein. Patalano and colleagues could show that insertion of a truncated version of Unr using a PiggyBac Transposon (CSD1-6 plus the first Q-rich domain) leads to death of the flies shortly after eclosion, most likely due to observed dramatic defects of dosage compensation[308]. This indicates that besides CSD123, additional CSDs and ncCSDs have an essential role in proper protein function. On the basis that all canonical CSDs share a similar fold, share the same RNA binding residues and represent

highly promiscuous RNA binders[27,309], CSD123 should be capable to compensate for the lack of the other domains (Figure 36c). Since this is not the case, the hypothesis, that ncCSDs mediate RNA structure specificity by reducing conformational heterogeneity is strengthened. However, the disability for compensation could also be due to a lower cellular abundance of CSDs in the shorter construct compared to the full-length protein.

Each RNA binding motif of the five canonical CSDs can be bound by single-stranded regions within RNAs that are otherwise structured. These three-dimensional structures could be very different and only some spatial arrangements could result in a CSD-cumulative high affinity binding of Unr. Contrary, RNAs with a different and less recognizable fold would be bound by fewer CSDs, resulting in weaker binding. In this model, conformational restrictions are imposed on the orientation of CSDs through interdomain contacts with the ncCSDs, defining target specificity (Figure 36a and 36b). The fact that Unr is known to bind to different IRES sequences[75–78,310], which do not exhibit a consensus sequence or universal structure motifs, but share complex structural elements[65], strengthens the hypothesis of a target specificity towards RNA structures instead of sequences.

Besides their role in RNA target specificity, ncCSDs could be involved in specific protein-protein interactions, maybe mediated through their additional loop that is missing in CSDs. In the case of other RBPs being the interactors, this would place further restrictions on the ultimate RNA sequence specificity. Despite high RNA binding promiscuity of CSDs, collectively and through spatial restraints that are imposed by ncCSDs and potentially by additional binding partners, Unr can select only in a limited scope of RNA features. To extend this hypothesis, ncCSDs may represent a new class of pseudo-RBDs, that structurally resemble classical RBDs, but lack residues critical for individual RNA binding. Future studies including other proteins with ncCSDs or even different domain classes will show whether this idea of pseudo-RBDs being a general phenomenon holds true.

5.3. Multidomain RNA recognition of *Drosophila* Unr

Although my hypothesis claims that the ncCSDs contribute to RNA binding without direct physical interaction, there is one exception. In presence of its neighboring CSDs, ncCSD8 clearly contributes directly to RNA binding. Already during NMR-based RNA titration experiments of CSD78 and CSD789, CSPs were observed within ncCSD8. The reason for CSPs in ncCSD8 were first speculated to be the result of unspecific interactions or allostery, as a single ncCSD8 construct did not harbor significant CSPs upon RNA titration and a positively charged surface patch was located close to the interaction surface of CSD7. However, the

crystal structure of CSD789 bound to a poly-A RNA clearly demonstrated that this interaction is specific. Indeed, one base of the RNA is pointing into the interaction surface between CSD7 and ncCSD8, leading to direct interactions between base A5 and amino acid E786 (CSD7), R856 and P860 (both ncCSD8). CSPs for the R856 were also observed in NMR titration experiments, which further validates the crystal structure. Signals of prolines cannot be observed in $^{1}H,^{15}N$-HSQC experiments, due to the lack of an amide proton. The importance of this interaction also on the binding affinity could be shown by mutational studies. Although the studies show, that the ncCSD8 contribution to RNA binding seems to be a peculiarity within Unr, further studies of other ncCSD containing proteins are inevitable to fully understand their exact biological functions.

Beside this ncCSD-RNA interactions, other non-canonical RNA interactions between N977 and K979 (both CSD9) were observed, which also contribute to the overall binding affinity and stabilize the interaction between the protein and the RNA. In difference to the canonical interactions, where the classical RNA binding residues (**F**G**F** and (**F/Y**)F**H**) of both CSDs show the typical π-π-stacking between the bases of the RNA and the aromatic side chains of the hydrophobic residues, the non-canonical residues are mostly charged or polar, and thereby not capable of forming the π-π-stackings. One exception is proline P860, which is also forming an interaction with the base via its π-aromatic face[311]. Altogether these residues are forming a tight RNA-protein interaction network, which is a good example of how complex multidomain RBP-RNA interactions can be. Considering that only one third of Unr is part of this structure, an impressive complexity can be imagined for full-length Unr-RNA interactions.

Interestingly, CSD789 showed high affinity binding to the poly-A RNA sequence and the crystal structure showed mainly contacts of RNA bases to the protein, indicating sequence specificity. A closer look at the contacts reveal adenine specificity for three of the non-canonical residues (E786, N977 and K979). Morozova and colleagues characterize discriminative binding to adenine from guanine and pyrimidine bases based on two types of interaction. One is hydrogen bonding of amino acids to the adenine-N1 or N6 and the other one is positioning of non-polar atoms to contact adenine-C2, which employs a steric hindrance of the exocyclic amino-N2 in guanine[312]. Both events occur in the crystal structure of CSD789 in complex with a poly-A sequence. E786 forms a hydrogen bond between its free oxygen to N6 of A5 (Figure 37a) and N977 and K979 form non-polar interactions to C2 of A6 and A9 respectively (Figure 37b and 37c).

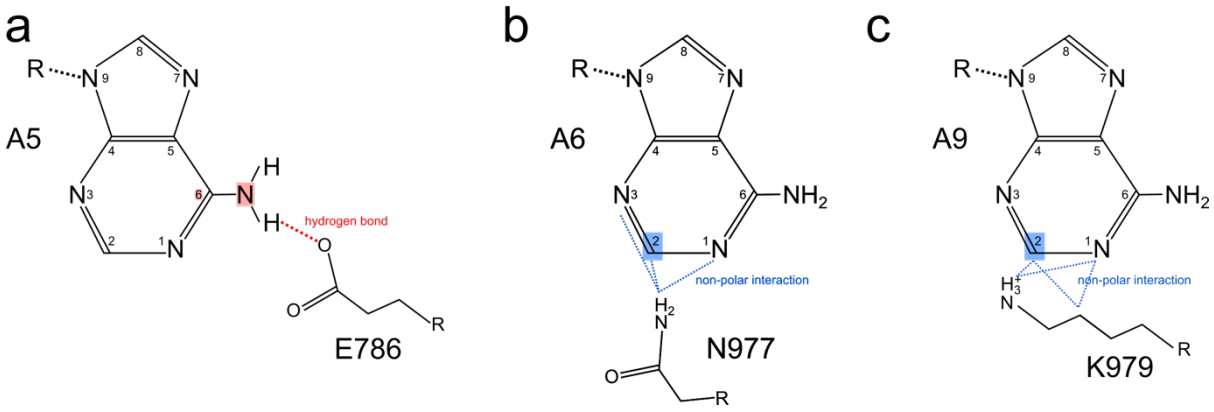


Figure 37: Selective adenine binding of residues of CSD789.
Adenine exclusive interactions of CSD789 are represented. **a**: A hydrogen bond is formed between a free oxygen of E786 and NH6 of A5. **b/c**: Non-polar interactions between C2 of A6 and A9 and the residues N977 (b) and K979 (c) respectively are highlighted.

In accordance with previous descriptions of CSDs being promiscuous RNA binders, the crystal structure shows that the non-canonical interaction residues may be the main determinants for a specificity towards A-rich sequences. Keeping in mind that the Unr binding sites within *msl2* mRNA are close to the poly-A tail (Unr binds preferentially in the 3'UTR of mRNAs[91]) and is known to interact with pAbP[82,125], it could be possible that the C-terminal part of the protein is binding to the poly-A-tail of mRNAs as well. Further foot printing or structural studies will show whether this hypothesis holds true.

5.4. Restricted structural heterogeneity of *Drosophila* Unr

To ultimately verify the hypothesis of RNA structure specificity that is mediated through the restricted conformational heterogeneity of the protein, structure determination of full-length Unr in complex with a target RNA would be required. This would help to understand the exact role of each domain in Unr biology and to characterize the complexity of multidomain RNA recognition. However, the challenges of preparing stable full-length Unr or CSD1-9 samples, might make structure determination using standard techniques impossible. Indeed, crystallization screens were tried for larger constructs of Unr (CSD1-6, CSD4-9 and CSD1-9) but crystal growth could not be observed. Flexibility that is retained in parts of the protein, like the one between ncCSD6 and CSD7 or between ncCSD8 and CSD9 in an RNA unbound form could very likely inhibit crystallization. Tools to characterize transient interactions between or within proteins, which are of weak affinity are for example NMR and SAXS. Both methods study molecules in solution and allow the generation of structure ensembles of even flexible and structurally heterogenous complexes[313]. Here, a SAXS-filtered modelling approach indicates a restricted structural heterogeneity within Unr. This is reflected by a better fit of structures that were neither extremely elongated nor compact. Additionally, collected NMR

data on ILV ^{13}C methyl group labeled samples, showed CSPs between shorter and longer protein constructs, indicating long-range interactions of the N- and C-terminal domains to the middle of the protein. Assignments by peak transfer only in crowded regions of the spectra were difficult. Additional NOE-based experiments will help to improve the number of assigned peaks. Nevertheless, even if all peaks can be assigned and the CSPs calculated it remains impossible to tell which domain interacts with each other. One possibility to gain more understanding towards a flexible model of a Unr full-length protein would be the usage of molecular dynamics simulations, that can finally be fitted against the available SAXS data[314,315]. This is ongoing work.

Another possibility to structurally characterize Unr in a full-length context is in its RNA-bound form. There are several examples of RBPs for which flexible interdomain linkers become ordered upon RNA binding alongside an increase in RNA binding affinity[262–266]. A scenario that was also observed between ncCSD8 and CSD9, where binding to the same RNA leads to a permanent interaction between the two domains. However, to enable structure determination of full-length Unr in complex with RNA, identification of the optimal RNA target is often a prerequisite. Consistent with previous studies, results presented here show binding of full-length Unr to one of its reported target RNAs (SL67 and SL678 of roX2 lncRNA) only with a slightly higher affinity compared to that of the N-terminal construct CSD123. This indicates, that roX2 lncRNA SL678 might not be the optimal binding partner for structural studies. Additional crystallization trials with an A50-mer RNA did not yield crystals. This could mean that indeed an optimal, probably folded, RNA partner is necessary or that flexibility remains even upon RNA binding, which prevents crystallization.

5.5. Unr RNP composition and its influence on translation regulation

Another approach to determine the structure of full-length Unr bound to RNA is the extension of the biomolecular complex to a Unr-dependent RNP, which may reduce flexibility by increasing the amount of interactions between single domains. By using proteomics analysis 31 proteins were identified as potential RNP interactors of *Drosophila* Unr, which will support future efforts in reconstituting an entire RNP for structural studies. Up to date the best studied Unr-RNP is the *msl2* mRNA repressor complex in dosage compensation, where single components like Hrp48 and pAbP are already characterized[124,125]. Also, structural knowledge about a subcomplex of Unr CSD1-Sxl RRM12 and parts of the mRNA is available[27]. However, in contrast to large and highly stable macromolecular machines, such as the ribosome and spliceosome[9,10,15,16,316,317], the analysis of transient interactions, as observed during signaling or

in regulatory networks is often difficult to conduct due to their dynamic nature. Nevertheless, since translation repression of the *msl2* mRNA in female flies is essential for the viability to be near constant, this system is a good candidate for promising structural characterization.

PAbP as the top candidate in the proteomics analysis seems to undergo an RNA independent interaction with Unr and is thus a promising candidate to start initial interaction studies. ^{1}H-^{15}N-HSQC experiments suggest binding of the C-terminal part of pAbP with the N-terminal part of Unr (CSD12). Although the PABC domain of pAbP is known to bind to proteins via a PAM2 motif (Pfam: PF07145)[318–320], this observation was unexpected, since the whole sequence of Unr does not contain this well characterized motif. Additional contribution of the long linker of pAbP to the binding with CSD12 suggests, that the interaction between the two proteins is of a so far uncharacterized manner. Additionally, the interaction between CSD789 and RRM3 of pAbP and polyA-RNA observed in this study makes me optimistic that structure determination of a large mRNP will be possible. Future studies have to show whether both proteins interact and still bind to the RNA or whether interaction competes with RNA of one or even both proteins.

The observed interaction is also in line with a previous study in humans that assigned the Unr-PABP interaction to CSD3 and CSD7[80], which results in the stimulation of translation. Although, CSD3 could not be tested due to the lack of a stable construct an interaction with pAbP could not be disproven. Moreover, it is possible that even further CSDs of Unr interact with parts of pAbP not tested in this study. Here, flexibility or loop formations within the RNA could bring the poly-A-tail closer to the characterized Unr binding sites, making interactions between the whole surface of both proteins possible.

As discussed previously, most studies connect binding of pAbP to mRNAs with translation upregulation, through recruitment of the small ribosomal subunit, mainly via interaction with eIF4G, mapped to RRM2 of pAbp. This results in the well-characterized closed-loop formation of the mRNA (Figure 38a)[58–60,321]. Nevertheless, there are few examples, that describe a contribution pf pAbP to translation repression, including its binding to msl2, but also to its own mRNA[82,83,125].

In the case of the neural RBP Musashi, eIF4G binding directly competes with pAbP binding, leading to an inhibition of translation enhancement of pAbP[322]. However, Duncan and colleagues could show, that in the case of *msl2* translation repression the binding between pAbP and eIF4G is not disturbed and they conclude, that the 3’ UTR repressor complex, including Unr, interferes with the recruitment of the small ribosomal subunit after formation of the stable closed-loop structure between the two mRNA ends[125]. PAbP might increase the repression

activity by reducing recruitment of the small ribosomal subunit[125]. It is possible that the structure remodeling of the mRNA mediated by the closed loop formation, presents the complex to a different set of specific targets. Further the repressor complex is sufficiently large to potentially repress scanning of the small ribosomal subunit for the initiation codon. This goes along with previous observations that the Sxl-Unr regulatory complex reduces ribosome recruitment (Figure 38a)[125].

Another potential repression mechanism would be similar to the one of Paip2, which suppresses translation by displacing pAbP and the poly-A tail[323]. The similar involvement of pAbP interaction domains between Paip2 and Unr, namely RRM3 and the PABC domain, would strengthen this repression. However, since pAbP-Paip2 interaction leads to a disruption of the closed loop formation[324], the mechanism in the *msl2* 3' UTR repressor complex must be different.

Considering that effects on the translation output depend on interaction partners of pAbP that either promote upregulation or repression of translation, and that Unr is known to interact in both cases with pAbP[80,82,125], RNA specific interaction partners remain to be identified. This specific regulatory interaction partner within the *msl2* 3' UTR repressor complex remains to be identified, but would most likely be a prerequisite to fully understand the mechanism. Potentially Imp is involved in the repression of *msl2* mRNA (Figure 38a). Imp is a known cofactor of pAbP and Unr in repression of the *pAbp* mRNA, leading to stalled movement of the small ribosomal subunit[82,83]. Interestingly, it was another hit in the proteomics analysis. Additionally, potential Imp binding sites, according to binding sequences from previous *Drosophila* iCLIP and parCLIP data, were identified downstream of the known repressor binding sites (Figure 38b)[285]. Of special interest is also the 32 nt spacing region between two different clusters of potential Imp binding sites, as Hansen and colleagues showed enriched binding sequences of binding sites with distances from 0 to 30 nt[285]. Since the CLIP studies were done in SL2 cells, which are derived from male flies without repressor complex formation, *msl2* could not be identified as a target of Imp. Further interaction and structural studies have to show whether Imp can be identified as an additional component in the 3' UTR repressor complex.

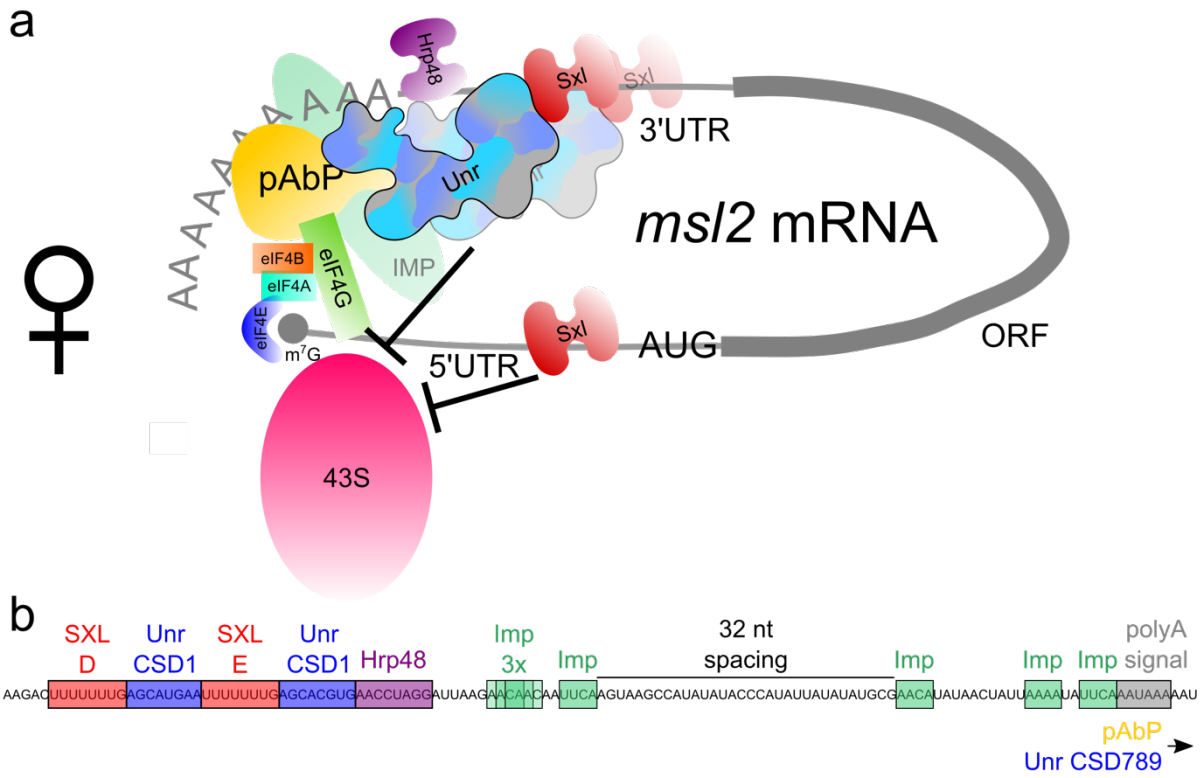


Figure 38: Schematic representation of a potential translation repression mechanism of the *msl2* mRNA in female flies.
a: In this model the 3' UTR repressor, consisting of dimeric Unr and Sxl, Hrp48, pAbP and potentially Imp, represses recruitment of the small ribosomal subunit. Due to the closed loop formation mediated by the pAbP bound eIF4G the sufficiently large complex prevents scanning of the ribosomal subunit for the initial codon at the 5' UTR. **b**: RNA binding sites of the different components of the 3' UTR repressor complex of the msl2 mRNA in a nucleotide resolution. Potential Imp binding sites, including a 32 nt spacing region are highlighted as well.

Taken together this thesis implicates a so far unknown role of a pAbP-Unr interaction within the translation repression of *msl2* mRNA. This indicates the need for more detailed structural and functional analysis.

Further, understanding this mechanism on a structural level may help to transfer knowledge to different systems, which involve a similar set of RBPs. However, although translation of some mRNAs might be similarly regulated, others could be regulated by an entirely different mechanism.

6. Conclusions and Outlook

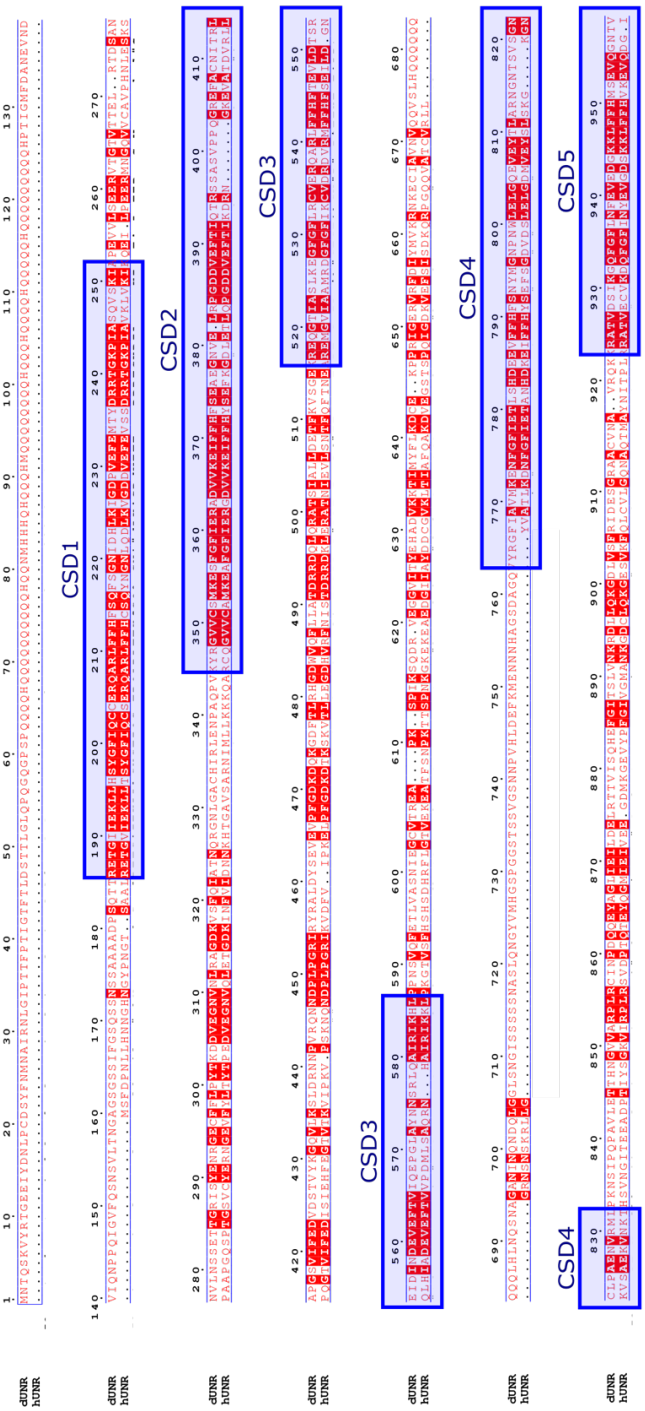
This work is the first study, that tries to understand RNA binding of Unr full-length with its five cold-shock domains (CSD) from a structural point of view. Surprisingly, four additional and previously unrecognized domains were identified, that turned out to share the CSD-fold despite low sequence similarity. Due to the lack of otherwise conserved RNA binding residues they were termed non-canonical CSDs. The database search showed that ncCSDs are present in more than 1000 proteins of different phyla. Compared to the classical CSDs, these domains lack the ability to bind RNA independently. Nevertheless, by forming interdomain contacts with the classical CSDs, they seem to contribute indirectly to the RNA target specificity of Unr. Although there remains some flexibility in the full-length protein, the restricted heterogeneity mediated by the domain-domain interaction, keeps the RNA binding surfaces in a special orientation, which makes them accessible to certain RNA folds. A disruption of the interaction surface impacts not only target specificity, but also translation regulation of Unr and thereby the protein function. This underlines the biological importance of the ncCSDs.

Together with a previous study on IMP3[35], these data show the need for the development of techniques, that not only focus on sequence or local RNA structure specificity of RBPs, but also consider and characterize the complexity of three dimensional folds of both RNA and protein.

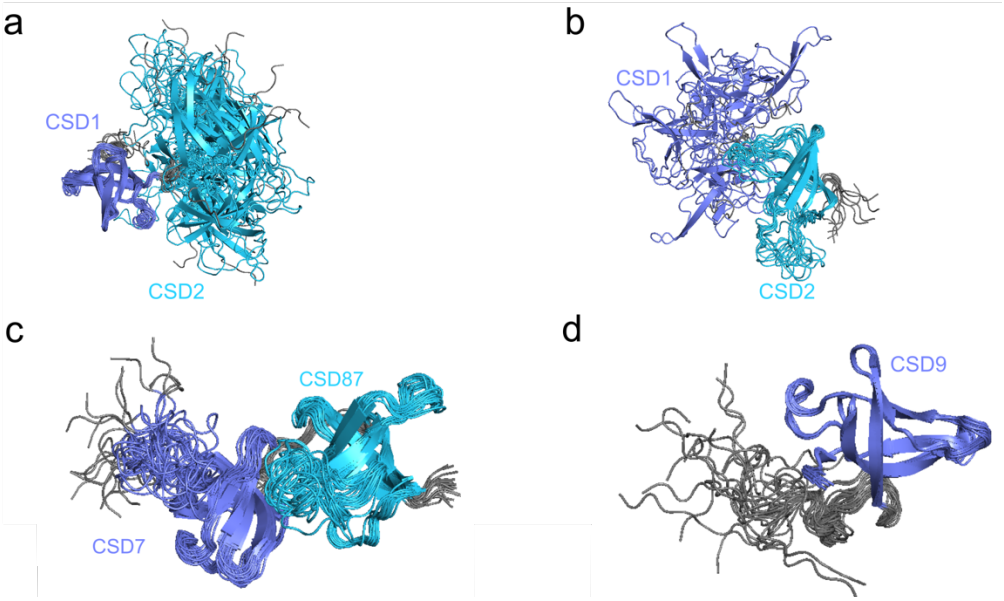
Additionally, structural data maps parts of Unr RNA binding to the poly-A tail and interaction studies, that add information to the Unr-pAbP interaction, could generate a more detailed model of the function of the 3' UTR *msl2* mRNA repressor complex. The interactions between Unr, pAbP, Sxl and Hrp48 may sterically repress the recruitment of the small ribosomal subunit. Together with the generation of sufficiently stable Unr constructs this work will support future work to dissect the complexity of translation regulation by structural biology. Moreover, the proteomics data, that mapped potential Unr-RNP compositions, may be helpful in analyzing different translation regulation complexes of Unr, that may be often of a transient character

Appendices

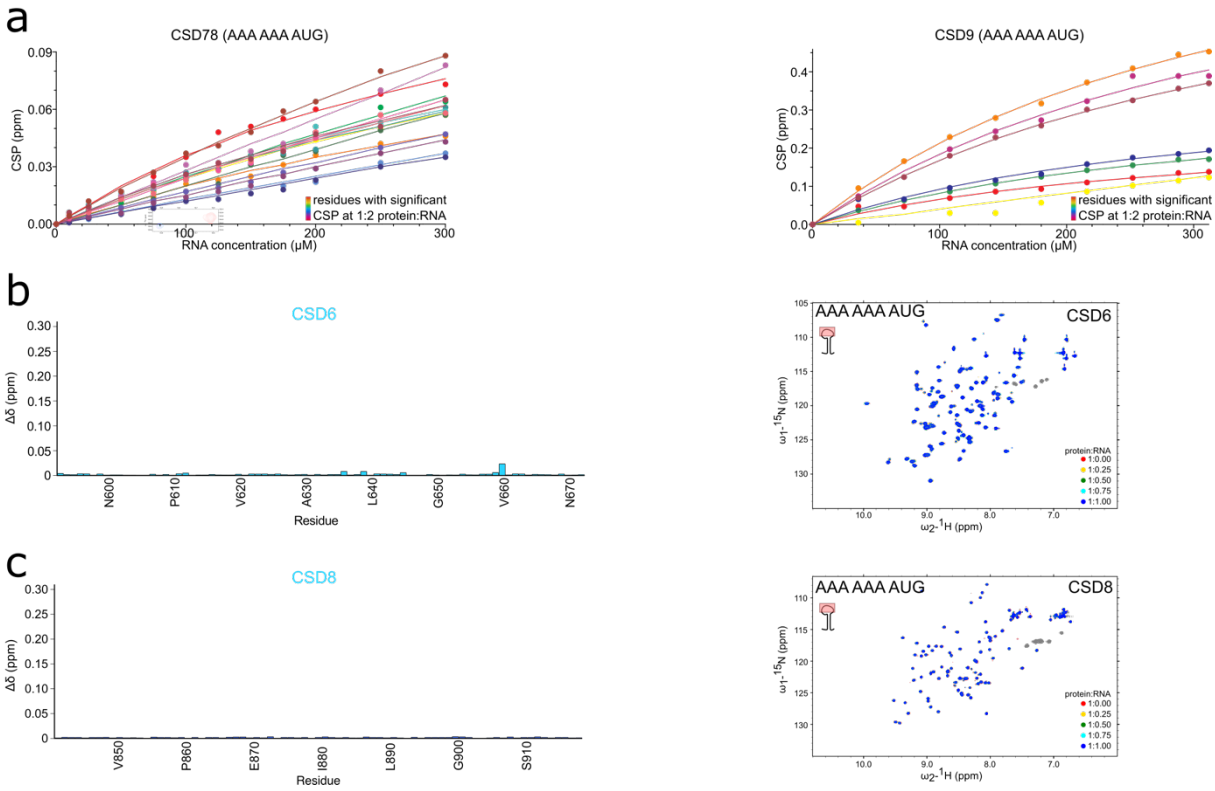
A. Supplementary Figures



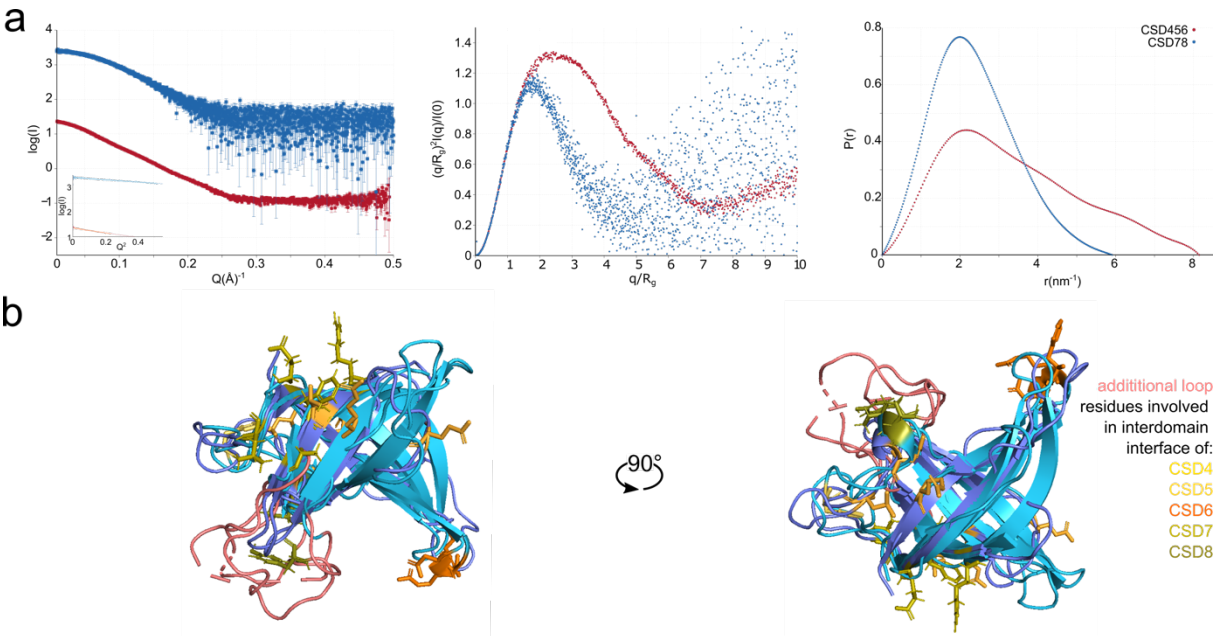
Supplementary Figure 1: Sequence alignment of *Drosophila* and human Unr.
The same residues between all domains are colored red. The CSD boundaries are highlighted in blue boxes. The alignment has been done using Emboss Needle[257] and ESPript [258] has been used for illustration.



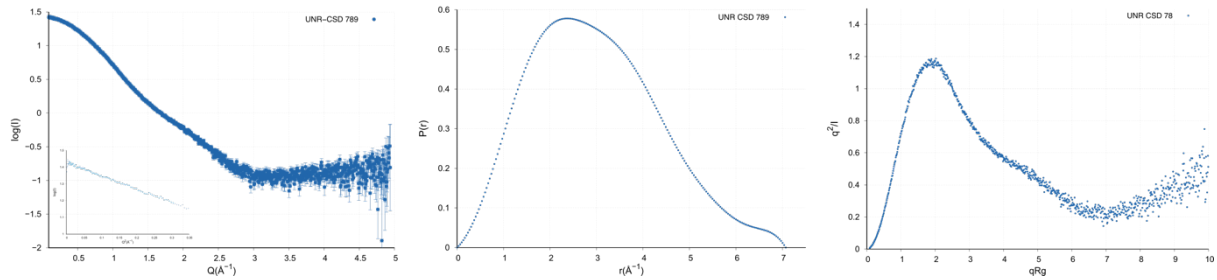
Supplementary Figure 2: Structural ensemble of NMR CSD structures
The ensemble of the 20 lowest energy NMR conformations of CSD12, superimposed on CSD1 (**a**) (residues: 197-237) and CSD2 (**b**) (residues: 269-274, 293-296 and 306-328), of CSD78 superimposed on secondary structure elements (**c**) (residues 764-813, 824-834, 843-854, 867-873 and 879-920) and of CSD9 superimposed on the core CSD domain (**d**) (residues 912-990).



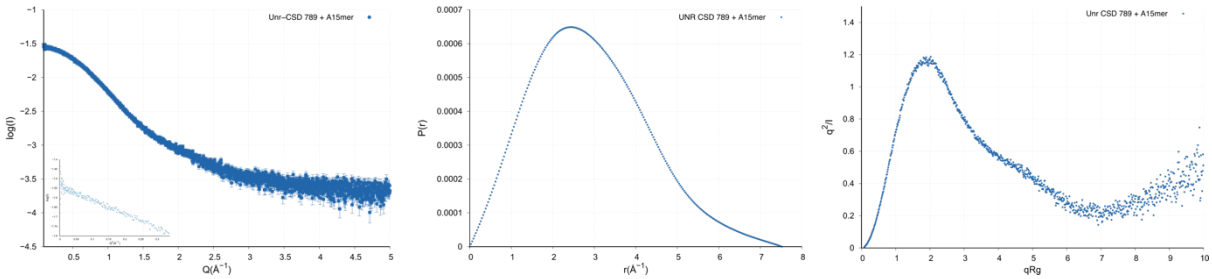
Supplementary Figure 3: RNA binding of single and multiple CSD constructs assessed by NMR.
a: For CSD78 (left) and CSD9 (right) chemical shift perturbations and the corresponding fit for different titration concentrations of the significantly shifting residues (CSPs larger than the average plus standard deviation of all shifts) are shown. **b**: ^{1}H, ^{15}N-HSQC NMR titration of CSD6 (593-677) with a 9-mer RNA (AAA AAA AUG) (right) devoid of significant chemical shift perturbations (left). **c:** ^{1}H, ^{15}N-HSQC NMR titration of CSD8 (842-922) with a 9-mer RNA (AAA AAA AUG) (right) devoid of significant chemical shift perturbations (left).



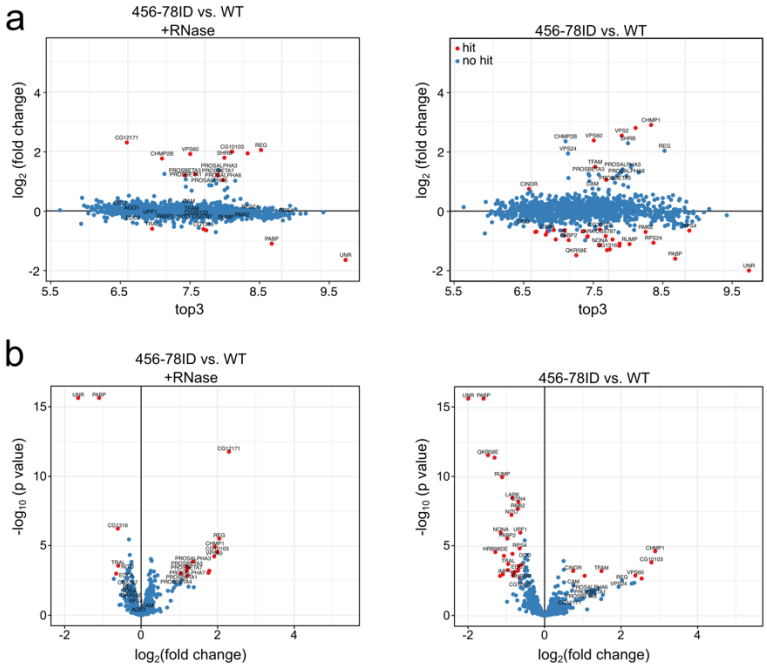
Supplementary Figure 4: Interdomain contacts of *Drosophila* Unr.
a: *I*(*q*) *versus q* as log-linear plots with the inset showing the Guinier fits for $qR_g < 1.3$ indicating good data quality and no aggregation for the curves of CSD456 (red) and CSD78 (blue) (left). Dimensionless Kratky plots indicate that proteins (CSD456: red; CSD78: blue) are mostly structured with low flexibility (middle). *P*(*r*) *versus r* profiles normalized to equal areas, showing the highest radius of gyration for CSD456 (red) and the lowest for CSD78 (blue) (right). **b**: Superimposition of CSD5 and 7 and ncCSD4, 6 and 8, highlighting the additional loop that is present in the ncCSDs (pink) and the residues that are involved in formation of the interdomain interaction surface (shades from yellow to green).



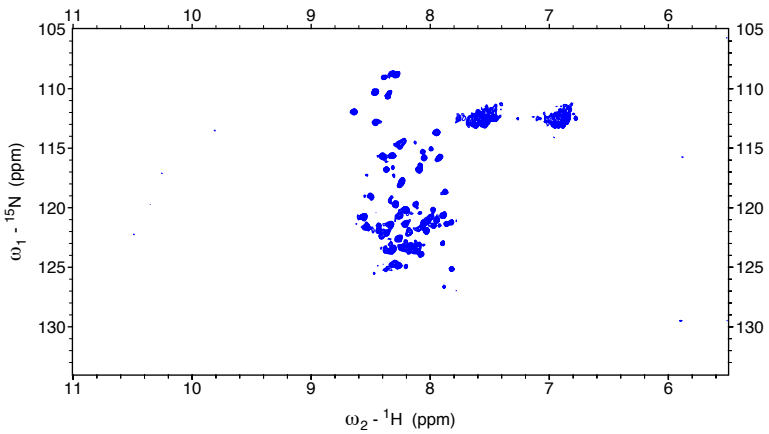
Supplementary Figure 5: SAXS plots of CSD789.
I(*q*) *versus q* as log-linear plots with the inset showing the Guinier fits for $qR_g < 1.3$ indicating good data quality and no aggregation for the curves of CSD789 (left). Dimensionless Kratky plots indicate that the protein is mostly structured with low flexibility (middle). *P*(*r*) *versus r* profiles showing the radius of gyration for CSD789 (right).



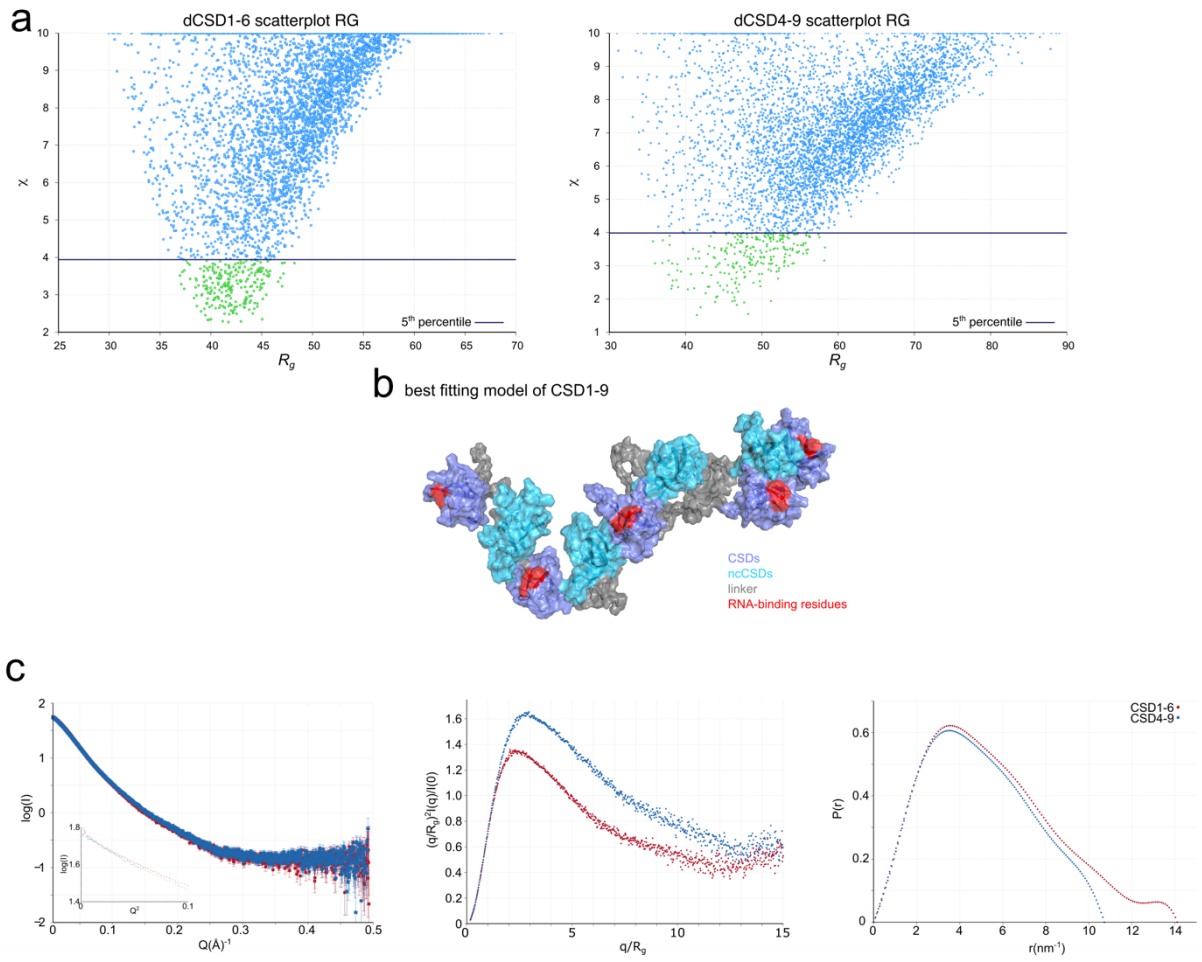
Supplementary Figure 6: SAXS plots of CSD789 in complex with an A15mer RNA.
I(*q*) *versus q* as log-linear plots with the inset showing the Guinier fits for $qR_g < 1.3$ indicating good data quality and no aggregation for the curves of the complex (left). Dimensionless Kratky plots indicate that the complex is mostly structured with low flexibility (middle). *P*(*r*) *versus r* profiles showing the radius of gyration for the complex (right).



Supplementary Figure 7: Proteomics data show an enrichment of different proteins in the Unr pull down samples.
a: Scatter plot showing the top3 value (average abundance of a protein in the mass spectrometry run) versus the log2(fold change) of interdomain mutant vs WT for RNase treated (top) and non-treated (bottom) conditions. N=3. **b**: Volcano plots showing the difference WT sample and the 456-78 interdomain mutant with (upper graph) and without RNase treatment (lower graph). The $\log_2$(fold change) is plotted against the negative $\log_{10}$(p value). Differentially regulated proteins are colored in red. N=3

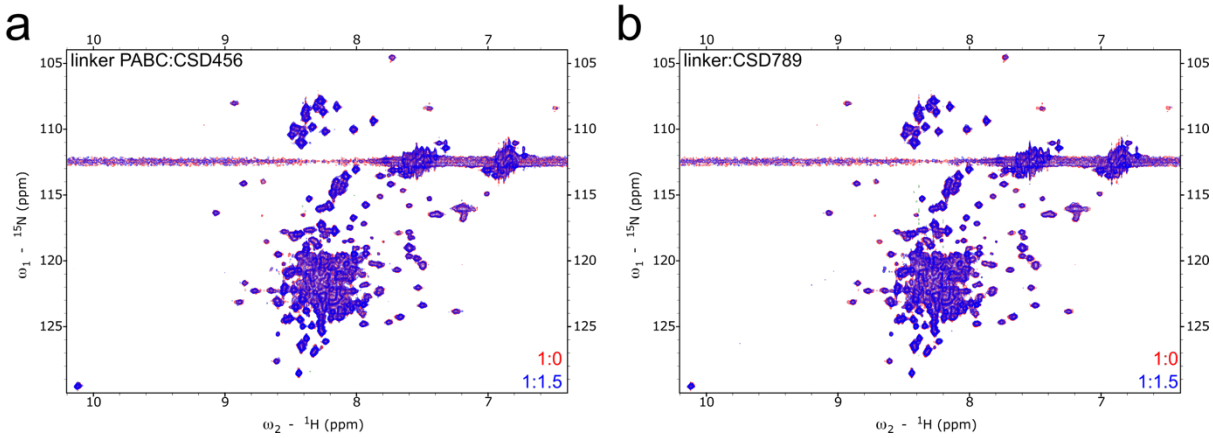


Supplementary Figure 8: ^{1}H,^{15}N-HSQC spectrum showing the N-terminal region of Unr.
The spectrum shows that M1-S175 within Drosophila Unr are unstructured.



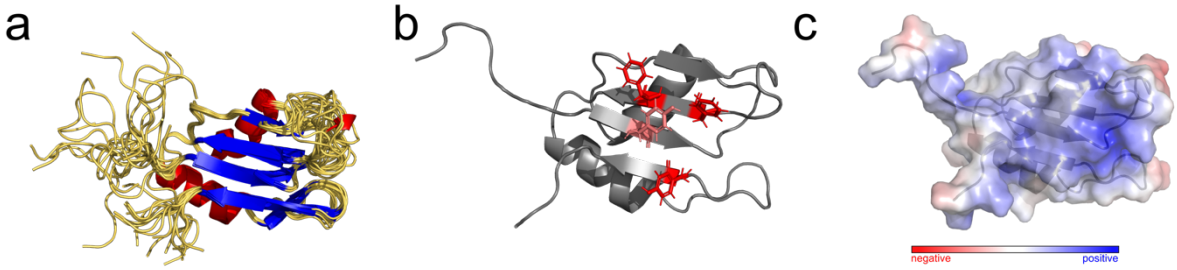
Supplementary Figure 9: SAXS data of CSD1-6 and CSD4-9 were used to analyzed modeled CSD1-9 structures.
a: Scatterplot of the radius of gyration versus the SAXS fit (χ) of CSD1-6 (left) and CSD4-9 (right) from each of the randomized structures. The structures that fit the individual SAXS curves best (lowest 5th percentile) converge towards an R_g-range (CSD1-6: 35-50; CSD4-9: 35-60). Data points of structures, for which the χ is out of range are assigned to an artificial χ of 10. **b**: RNA binding motifs (red), canonical (cyan) and non-canonical CSDs (blue) are highlighted on the best fitting overall structure. This indicates a certain orientation of the RNA binding residues in respect to each other.

c: *I*(*q*) *versus q* as log-linear plot with the inset showing the Guinier fits for $qR_g < 1.3$ indicating a good quality of the data (left). Dimensionless Kratky plot showing the scattering of a folded protein (middle). *P*(*r*) *versus r* profiles normalized to equal areas indicating a bigger radius of gyration for CSD1-6 compared to CSD4-9. The plots are shown for the SAXS curves of CSD1-6 (red) and CSD4-9 (blue) (right).



Supplementary Figure 10: Protein-protein interaction between the linker PABC construct and CSD456/789.

$^{1}H,^{15}N$-HSQC spectra showing the interaction between the linker PABC construct and CSD456 (**a**)/CSD789 (**b**). Red shows the apo form of the protein and blue shows it in the presence of 1.5 molar excess of the different unlabeled CSD constructs.



Supplementary Figure 11: NMR structure of pAbP RRM3.

a: The ensemble of the 20 lowest energy NMR conformations of pAbP RRM3, superimposed on the secondary structure elements. **b**: Structure of pAbP RRM3 showing the RNA binding residues of the two RNP motifs (red). One aromatic residue, pointing to the outside of the core located close to the RNP motif residues is colored in pink. **c**: The electrostatic surface of RRM3 shows positive charge on the RNA interaction surface.

B. Supplementary Tables

Supplementary Table 1: Data collection and refinement statistics of the crystal structures. Statistics for the highest resolution shell are shown in parentheses.

	CSD456	CSD789+A15mer
Wavelength	1.005 Å	0.976 Å
Resolution range	77.84 - 2.02 (2.092 - 2.02)	48.79 - 1.6 (1.657 - 1.6)
Space group	P 62	C 1 2 1
Unit cell	89.88 89.88 58.75 90 90 120	97.92 36.78 85.83 90 119.8 90
Total reflections	113928 (6419)	235687 (22904)
Unique reflections	34128 (2534)	35303 (3434)
Multiplicity	3.3 (2.5)	6.7 (6.7)
Completeness (%)	98.00 (87.3)	99.75 (98.76)
Mean I/sigma(I)	7.8 (1.0)	13.09 (1.31)
Wilson B-factor	39.76	21.23
R-merge	0.073 (0.78)	0.09 (1.22)
R-meas	0.087 (0.95)	0.099 (1.32)
R-pim	0.046 (0.54)	0.038 (0.50)
CC1/2	0.99 (0.57)	0.99 (0.52)
Reflections used in refinement	32428 (2409)	35297 (3434)
Reflections used for R-free	1698 (124)	1764 (171)
R-work	0.22 (0.32)	0.19 (0.34)
R-free	0.25 (0.32)	0.21 (0.39)
Number of non-hydrogen atoms	2062	2099
macromolecules	1842	1826
ligands	7	5
solvent	213	268
Protein residues	238	219
RMS(bonds)	0.002	0.008
RMS(angles)	0.38	1.00
Ramachandran favored (%)	96.93	98.60
Ramachandran allowed (%)	2.63	0.93
Ramachandran outliers (%)	0.44	0.47
Rotamer outliers (%)	0.00	0.00
Average B-factor	47.73	25.22
macromolecules	47.83	23.60
ligands	51.27	35.42
solvent	46.76	36.04

Supplementary Table 2: NMR restraints of the structure calculations.

	dCSD12+SL6 apical	dCSD78	dCSD9	pAbP RRM3
	Experimental restraints			
Total NOEs	3911	4530	3712	2462
distance restraints	1788	2323	1831	2321
Short range (|i − j| ≤ 1)	1005	1242	544	1413
Medium range (|i − j| < 5)	124	211	283	243
Long range (|i − j| > 5)	659	870	1004	665
Dihedral restraints (ϕ/ψ)	224	167	93	102
	Structural Quality			
Coordinate precision (Å)				
Backbone (N, Cα, C′)	0.40 (CSD1), 0.53 (CSD2)[a]	0.39 (CSD7), 0.39 (CSD8)[b]	0.38[c]	0.17[d]
Heavy atoms	0.99 (CSD1), 1.05 (CSD2)[a]	0.84 (CSD7), 0.86 (CSD8)[b]	0.80[c]	0.63[d]
Restraint RMSD				
Distance restraints, Å	0.02 ± 0.0019	0.0196 ± 0.002	0.0253 ± 0.003	0.0228805 ± 0.003
Dihedral restraints, °	0.666 ± 0.104	1.437 ± 0.101	0.565 ± 0.105	0.632473 ± 0.056
Deviation from idealized geometry				
Bond lengths, Å	0.0038 ± 0.00008	0.0036 ± 0.00009	0.00434 ± 0.00013	0.00351 ± 0.00014
Bond angles, °	0.504 ± 0.01	0.491 ± 0.011	0.542 ±- 0.022	0.458 ± 0.013
	Whatcheck analysis			
First generation packing	-3.271 ± 0.228	-2.751 ± 0.161	-1.912 ± 0.191	-1.430 ± 0.270
Second-generation packing	-2.491 ± 0.274	-2.402 ± 0.239	-1.560 ± 0.284	-1.587 ± 0.221
Ramachandran plot appearance	-3.806 ± 0.298	-3.843 ± 0.329	-4.442 ± 0.480	-3.783 ± 0.443
χ−1/χ−2 rotamer normality	-3.806 ± 0.532	-3.915 ± 0.473	-5.416 ± 0.431	-2.687 ± 0.691
Backbone confirmation	-1.261 ± 0.534	-2.329 ± 0.537	-1.326 ± 0.376	-1.202 ± 0.411
	Ramachandran analysis, %			
Favored regions	80.9	76.7	80.5	78.7
Allowed regions	16.7	20.9	18.9	21.3

Generously allowed	1.9	1.7	0.6	0.1
Disallowed	0.4	0.7	0.0	0.0

a For residues using 197-237 (CSD1) and 269-274, 293-296 and 306-328 (CSD2)

b For residues using 764-813 and 824-834 (CSD7) and 843-854, 867-873 and 879-920 for (CSD8)

c For residues using 912-990

d For residues using 11-41, 52-71 and 82-85

Supplementary Table 3: SAXS data analysis statistics, according to the common publishing guidelines.

	dCSD456	**dCSD78**	**dCSD789**	**dCSD789 A15mer**	**dCSD1-6**	**dCSD4-9**
(a) Sample Details						
Organism	*E. coli* BL2 (DE3)	*E. coli* BL2 (DE3)	*E. coli* BL2 (DE3)	*E. coli* BL2 (DE3)	*E. coli* BL2 (DE3)	*E. coli* BL2 (DE3)
Source	*this work*	*this work*	*this work*	*this work*	*this work*	*this work*
Uniprot sequence ID	Q9VSK3	Q9VSK3	Q9VSK3	Q9VSK3	Q9VSK3	Q9VSK3
Description	Unr E422-H677, with TEV-cleaved N-terminal His_6-tag	Unr A756-K922, with TEV-cleaved N-terminal His_6-tag	Unr A759-D990, with TEV-cleaved N-terminal His_6-tag	Unr A759-D990, with TEV-cleaved N-terminal His_6-tag in complex with A15mer RNA	Unr A176-H677, with TEV-cleaved N-terminal His_6-tag	Unr V428-D990, with TEV-cleaved N-terminal His_6-tag
Molecular mass M from chemical composition (Da)	29.532	18.597	25,750	30626	56.868	63.121
loading concentration (mg/ml)	1.16 and 4.65 mg/ml	0.8 mg/ml	1.6 and 6.2 mg/ml	1.5 and 5.8 m/ml	0.92 and 3.73 mg/ml	0.83 and 4.18 mg/ml
injection volume (ul)	30	30	30	30	30	30
concentration (uM)	40/160	45	58/225	50/190	15/65	13/66
Solvent composition and source	20 mM Hepes/NaOH pH 7.5, 150 mM NaCl and 1 mM DTT					
(b) SAS data collection parameter						

Source and instrument	Grenoble ESRF BM29 with Dectris Pilatus 1M	Hamburg PETRA-III P12 with Dectris Pilatus 6M (REF)	Grenoble ESRF BM29 with Dectris Pilatus 1M	Hamburg PETRA-III P12 with Dectris Pilatus 6M (REF)	Grenoble ESRF BM29 with Dectris Pilatus 1M	
Wavelength (Å)	0.9919	1.24	0.9919	1.24	0.9919	
Sample-detector distance (m)	2.867	3.0	2.867	3.0	2.867	
q-measurement range (nm^{-1})	0.0355-4.9391	0.0224-7.3176	0.0355-4.9391	0.0224-7.3176	0.0385-4.9335	
Radiation damage monitoring	frame-by-frame comparison					
Exposure time (s) & number	1.0x10	0.195x20	1.0x10	0.195x20	1.0x10	
Sample configuration	sample changer with flow through capillary measurement					
Sample temperature (°C)	20	25	20			
(c) Software employed for SAS data reduction, analysis and interpretation						
SAXS data processing	I(q) vs. q using Bsx cube, solvent subtraction and curve merging using PRIMUSqt from ATSAS[171]					
Basic analyses: Guinier, P(r), Vp	PRIMUSqt from ATSAS 2.7.1[171]					
Atomic structure modelling	CRYSOL 2.8.2 from PRIMUSqt in ATSAS 2.8[203]		EOM 2.1 from PRIMUSqt in ATSAS 2.7.1[202]	CRYSOL 2.8.2 from PRIMUSqt in ATSAS 2.8[203]		
Molecular graphics	--	--	PyMol v3.3.1	--	--	
(d) Structural parameters						
	Guinier analysis					
I(0) (raw)	22.8+/-0.06	25.4+/-0.05	26.74+/-0.04	29±0.05	56.04+/-0.17	58.27+/-0.19
R_g (Å)	26.7+/-0.1	18.1+/-0.01	23.4+/-0.07	22.3±0.02	43.6+/-0.31	49+/-0.34
qR_g max (q_{min} = 0.0066 $Å^{-1}$)	1.28	1.3	1.29	1.30	1.25	1.13
Coefficient of correlation, R^2	0.97	0.74	0.98	0.96	0.86	0.94
	P(r) Analysis from AUTOGNOM					
I(0) (cm-1)	22.8	25.43	26.51	28.9	55.19	48.99
R_g (Å)	22.61	18.2	23.1	22.6	43.3	38.1
d_{max} (Å)	81.5	59.1	70.7	75.2	141.1	107.2

q range (Å^{-1})	0.102-3.00	0.089-4.427	0.092-3.423	0.107-3.585	0.147-1.833	0.0714-1.631
χ^2 (total estimate from *GNOM*)	0.67	0.89	0.71	0.88	0.6	0.59
Porod volume (Å^{-3}) (ratio V_P/calculated *M)*	37490	27610	37470	36490	148820	132810
(f) Atomistic modelling						
Method	CRYSOL[203]		EOM[202]		CRYSOL[203]	
Crystal stucture	Unr CSD456	Unr CSD78	Unr CSD78 and 9			
Flexible linker definition			V919-R920			
CRYSOL	Constant subtraction allowed					
χ^2	1.026	1.218				
Predicted R_g (Å)	26.22	18.09				
Vol (Å), Ra (Å), Dro (e Å^{-3})	25504, 1.4, 0.07	17716, 1.8, 0.000				
EOM			default parameters, 10.000 models in initial ensemble, native-like models, constant subtraction allowed			
χ^2			0.932			
Constant subtraction			0.018			
No. of representative structures			3			

Acknowledgements

The last four years wouldn't have been the same without many people who were thereby directly or indirectly contributing to this thesis.

First of all, I'd like to thank Janosch for always having an open door to answer questions, discuss science, or just enjoy promising results together. Further I am grateful for his permanent support not only in science, but also in my future career development and for his trust in me four years ago, although I had no clue about NMR back then.

This brings me to my second big thank you. I want to thank Bernd for his patience in teaching me NMR, which includes answering many more or less intelligent questions. He is an expert in staying calm, even when I put my sample in the wrong spectrometer, that was of course not lifted and had another measurement running.

I'd like to thank Sandra for being a good friend, keeping the lab running and organizing the best retreats. Of course, thereby I'd like to thank also all the other former and current lab members, who became friends over the last years and without whom the retreats and every single day in the lab wouldn't have been as informative and as much fun. I also have to thank all my students, who gave their time, to conduct experiments for me and stayed motivated, regardless of my expectations.

Furthermore, I'd like to thank Fátima Gebauer and the members of her lab at the CRG in Barcelona, who hosted me with so much warmth and gave me the opportunity to escape one winter cycling to EMBL. I always felt welcomed and had the opportunity to learn so many new techniques.

Moreover, I'd like to thank different members of the core facilities, who helped realizing many experiments, which would never have come about without their help. Laura Villacorte, Dylan Mooijman and Vladimir Benes were a big support in the RIP-seq study and Frank Stein, Per Haberkant und Mikhail M. Savitski in the proteomics approach. Also, I want to thank Brice Murciano for pipetting all the different initial screens for crystallography. A special thanks belongs to Kathryn Perez, who helped me not only with many biophysics' measurements, but was also there for many relaxing evenings with one or the other bottle of wine.

I thank the members of my thesis advisory and defense committee Dr. Martin Beck, Prof. Dr. Frauke Melchior, Dr. Marco Marcia for the time and effort they spent over the past years for supporting me and my project. Further I want to thank Prof. Dr. Georg Stoecklin who put effort and time in reading my thesis and joining the defense.

A very big thanks goes to my parents, sisters and grandparents, whose pride has given me more strength and self-confidence than they were aware of and who have supported every vision of my future (although I guess we are all happy I didn't end up as a cashier).

Last but not least I want to thank Hennink, without whom the last years wouldn't have been as perfect. Above all, I want to thank him for supporting each of my wishes and to always take the load off, regardless of whether I wanted to spend several months in Marseille or Barcelona, or several years hundreds or (as in the future) thousands of kilometers away.

Bibliography

1. Dahm, R. Friedrich Miescher and the discovery of DNA. *Developmental Biology* **278**, 274–288 (2005).
2. Caspersson, T. & Schultz, J. Pentose nucleotides in the cytoplasm of growing tissues [12]. *Nature* **143**, 602–603 (1939).
3. Crick, F. On protein synthesis. *Biol. Replication Macromol. Symp. Soc. Exp. Biol. 12* 128–63 (1958).
4. PALADE, G. E. A small particulate component of the cytoplasm. *J. Biophys. Biochem. Cytol.* **1**, 59–68 (1955).
5. Holley, R. W. *et al.* Structure of a ribonucleic acid. *Science (80-.).* **147**, 1462–1465 (1965).
6. Bernfield, M. R. & Nirenberg, M. W. RNA codewords and protein synthesis. *Science (80-.).* **147**, 479–484 (1965).
7. Cech, T. R. & Steitz, J. A. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **157**, 77–94 (2014).
8. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology* **19**, 327–341 (2018).
9. Bieri, P., Greber, B. J. & Ban, N. High-resolution structures of mitochondrial ribosomes and their functional implications. *Current Opinion in Structural Biology* **49**, 44–53 (2018).
10. Voorhees, R. M. & Ramakrishnan, V. Structural Basis of the Translational Elongation Cycle. *Annu. Rev. Biochem.* **82**, 203–236 (2013).
11. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* **16**, 129–143 (2015).
12. Cramer, P. *et al.* Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **37**, 337–352 (2008).
13. Lee, J. & Borukhov, S. Bacterial RNA polymerase-DNA interaction-The driving force of gene expression and the target for drug action. *Frontiers in Molecular Biosciences* **3**, 73 (2016).
14. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).
15. Wan, R., Bai, R. & Shi, Y. Molecular choreography of pre-mRNA splicing by the spliceosome. *Current Opinion in Structural Biology* **59**, 124–133 (2019).

16. Wilkinson, M. E., Lin, P.-C., Plaschka, C. & Nagai, K. Cryo-EM Studies of Pre-mRNA Splicing: From Sample Preparation to Model Visualization. *Annu. Rev. Biophys.* **47**, 175–199 (2018).

17. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).

18. Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu. Rev. Biochem.* **84**, 325–354 (2015).

19. Pereira, B., Billaud, M. & Almeida, R. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends in Cancer* **3**, 506–528 (2017).

20. Conlon, E. G. & Manley, J. L. RNA-binding proteins in neurodegeneration: Mechanisms in aggregate. *Genes and Development* **31**, 1509–1528 (2017).

21. Craig Venter, J. *et al.* The sequence of the human genome. *Science (80-.).* **291**, 1304–1351 (2001).

22. Cléry, A., Blatter, M. & Allain, F. H. T. RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology* **18**, 290–298 (2008).

23. Auweter, S. D., Oberstrass, F. C. & Allain, F. H.-T. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**, 4943–59 (2006).

24. Hennig, J. & Sattler, M. Deciphering the protein-RNA recognition code: Combining large-scale quantitative methods with structural biology. *BioEssays* **37**, 899–908 (2015).

25. Hennig, J., Gebauer, F. & Sattler, M. Breaking the protein-RNA recognition code. *Cell Cycle* **13**, 3619–20 (2014).

26. Afroz, T., Cienikova, Z. & Cléry, A. One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence. *Methods Enzymol.* **558**, 235–278 (2015).

27. Hennig, J. *et al.* Structural basis for the assembly of the Sxl–Unr translation regulatory complex. *Nature* **515**, 287–290 (2014).

28. Afroz, T. *et al.* A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. *Genes Dev.* **28**, 1498–514 (2014).

29. Bae, E. *et al.* Structure and Interactions of the First Three RNA Recognition Motifs of Splicing Factor Prp24. *J. Mol. Biol.* **367**, 1447–1458 (2007).

30. Barraud, P. & Allain, F. H.-T. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRMs influences the nucleic acid binding topology. *J. Biomol. NMR* **55**, 119–138 (2013).

31. Cukier, C. D. *et al.* Molecular basis of FIR-mediated c-myc transcriptional control. *Nat. Struct. Mol. Biol.* **17**, 1058–1064 (2010).

32. Mackereth, C. D. *et al.* Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475**, 408–411 (2011).

33. Vitali, F. *et al.* Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J.* **25**, 150–162 (2006).

34. Zhang, W. *et al.* Crystal structures and RNA-binding properties of the RNA recognition motifs of heterogeneous nuclear ribonucleoprotein L: insights into its roles in alternative splicing regulation. *J. Biol. Chem.* **288**, 22636–49 (2013).

35. Schneider, T. *et al.* Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nat. Commun.* **10**, 2266 (2019).

36. Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell* **70**, 854-867.e9 (2018).

37. Cléry, A. & Allain, F. H.-T. *RNA Binding Proteins (form structure to function of RNA binding proteins) - Madame Curie Bioscience Database*. (2011).

38. Dimitrova-Paternoga, L., Jagtap, P. K. A., Chen, P. C. & Hennig, J. Integrative Structural Biology of Protein-RNA Complexes. *Structure* **28**, 6–28 (2020).

39. Gronland, G. R. & Ramos, A. The devil is in the domain: Understanding protein recognition of multiple RNA targets. *Biochemical Society Transactions* **45**, 1305–1311 (2017).

40. Sasse, A., Laverty, K. U., Hughes, T. R. & Morris, Q. D. Motif models for RNA-binding proteins. *Current Opinion in Structural Biology* **53**, 115–123 (2018).

41. Beck, M. & Hurt, E. The nuclear pore complex: Understanding its function through structural insight. *Nature Reviews Molecular Cell Biology* **18**, 73–89 (2017).

42. Hoelz, A., Glavy, J. S. & Beck, M. Toward the atomic structure of the nuclear pore complex: When top down meets bottom up. *Nature Structural and Molecular Biology* **23**, 624–630 (2016).

43. Kim, S. J. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018).

44. Kosinski, J. *et al.* Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science (80-.).* **352**, 363–365 (2016).

45. Lin, D. H. *et al.* Architecture of the symmetric core of the nuclear pore. *Science (80-.).* **352**, (2016).

46. Stuwe, T. *et al.* Architecture of the fungal nuclear pore inner ring complex. *Science (80-.).* **350**, 56–64 (2015).

47. Moore, K. S. & von Lindern, M. RNA binding proteins and regulation of mRNA translation in erythropoiesis. *Frontiers in Physiology* **9**, 910 (2018).

48. Kong, J. & Lasko, P. Translational control in cellular and developmental processes. *Nature Reviews Genetics* **13**, 383–394 (2012).

49. Sonenberg, N. & Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731–745 (2009).

50. Guo, A.-X., Cui, J.-J., Wang, L.-Y. & Yin, J.-Y. The role of CSDE1 in translational reprogramming and human diseases. *Cell Commun. Signal.* **18**, 14 (2020).

51. Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5′-

untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413–1416 (2016).

52. Hinnebusch, A. G. The Scanning Mechanism of Eukaryotic Translation Initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).

53. Merrick, W. C. eIF4F: A retrospective. *Journal of Biological Chemistry* **290**, 24091–24099 (2015).

54. Aitken, C. E. & Lorsch, J. R. A mechanistic overview of translation initiation in eukaryotes. *Nature Structural and Molecular Biology* **19**, 568–576 (2012).

55. Ivanov, A. *et al.* PABP enhances release factor recruitment and stop codon recognition during translation termination. 7766–7776 (2016).

56. Smith, R. W. P. *et al.* Viral and cellular mRNA-specific activators harness PABP and eIF4G to promote translation initiation downstream of cap binding. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 6310–6315 (2017).

57. Patel, G. P., Ma, S. & Bag, J. The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex. 7074–709 (2005).

58. Wei, C. C., Balasta, M. L., Ren, J. & Goss, D. J. Wheat germ poly(A) binding protein enhances the binding affinity of eukaryotic initiation factor 4F and (iso)4F for cap analogues. *Biochemistry* **37**, 1910–1916 (1998).

59. Tarun, S. Z. & Sachs, A. B. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J.* **15**, 7168–7177 (1996).

60. Wells, S. E., Hillner, P. E., Vale, R. D. & Sachs, A. B. Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* **2**, 135–140 (1998).

61. Shatsky, I. N., Terenin, I. M., Smirnova, V. V. & Andreev, D. E. Cap-Independent Translation: What's in a Name? *Trends in Biochemical Sciences* **43**, 882–895 (2018).

62. Walters, B. & Thompson, S. R. Cap-independent translational control of carcinogenesis. *Frontiers in Oncology* **6**, 128 (2016).

63. Yamamoto, H., Unbehaun, A. & Spahn, C. M. T. Ribosomal Chamber Music: Toward an Understanding of IRES Mechanisms. *Trends in Biochemical Sciences* **42**, 655–668 (2017).

64. Komar, A. A. & Hatzoglou, M. Internal ribosome entry sites in cellular mRNAs: Mystery of their existence. *Journal of Biological Chemistry* **280**, 23425–23428 (2005).

65. Baird, S. D., Lewis, S. M., Turcotte, M. & Holcik, M. A Search for Structurally Similar Cellular Internal Ribosome Entry Sites - PubMed. 4664–4677 (2007).

66. Grover, R., Candeias, M. M., Fhraeus, R. & Das, S. P53 and little brother p53/47: Linking IRES activities with protein functions. *Oncogene* **28**, 2766–2772 (2009).

67. Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62**, 2636–2643 (1988).

68. Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320–325 (1988).

69. Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).

70. Boussadia, O. *et al.* Unr is required in vivo for efficient initiation of translation from the internal ribosome entry sites of both rhinovirus and poliovirus. *J. Virol.* **77**, 3353–9 (2003).

71. Dormoy-Raclet, V., Markovits, J., Jacquemin-Sablon, A. & Jacquemin-Sablon, H. Regulation of Unr expression by 5′- and 3′-untranslated regions of its mRNA through modulation of stability and IRES mediated translation. *RNA Biol.* **2**, 112–120 (2005).

72. Elatmani, H. *et al.* The RNA-Binding Protein Unr Prevents Mouse Embryonic Stem Cells Differentiation Toward the Primitive Endoderm Lineage. *Stem Cells* **29**, 1504–1516 (2011).

73. Horos, R. & von Lindern, M. Molecular mechanisms of pathology and treatment in Diamond Blackfan Anaemia. *Br. J. Haematol.* **159**, n/a-n/a (2012).

74. Mihailovich, M. *et al.* Widespread generation of alternative UTRs contributes to sex-specific RNA binding by UNR. *RNA* **18**, 53–64 (2012).

75. Tinton, S. A., Schepens, B., Bruynooghe, Y., Beyaert, R. & Cornelis, S. Regulation of the cell-cycle-dependent internal ribosome entry site of the PITSLRE protein kinase: Roles of Unr (upstream of N-ras) protein and phosphorylated translation initiation factor eIF-2α. *Biochem. J.* **385**, 155–163 (2005).

76. Evans, J. R. *et al.* Members of the poly (rC) binding protein family stimulate the activity of the c-myc internal ribosome entry segment in vitro and in vivo. *Oncogene* **22**, 8012–8020 (2003).

77. Mitchell, S. A., Spriggs, K. A., Coldwell, M. J., Jackson, R. J. & Willis, A. E. The Apaf-1 Internal Ribosome Entry Segment Attains the Correct Structural Conformation for Function via Interactions with PTB and unr. *Mol. Cell* **11**, 757–771 (2003).

78. Boussadia, O. *et al.* Unr Is Required In Vivo for Efficient Initiation of Translation from the Internal Ribosome Entry Sites of both Rhinovirus and Poliovirus. *J. Virol.* **77**, 3353–3359 (2003).

79. Anderson, E. C., Hunt, S. L. & Jackson, R. J. Internal initiation of translation from the human rhinovirus-2 internal ribosome entry site requires the binding of Unr to two distinct sites on the ′ untranslated region. *J. Gen. Virol.* **88**, 3043–3052 (2007).

80. Ray, S. & Anderson, E. C. Stimulation of translation by human Unr requires cold shock domains 2 and 4, and correlates with poly(A) binding protein interaction. *Sci. Rep.* **6**, (2016).

81. Saltel, F. *et al.* Unr defines a novel class of nucleoplasmic reticulum involved in mRNA translation. *J. Cell Sci.* **130**, 1796–1808 (2017).

82. Patel, G. P., Ma, S. & Bag, J. The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex. *Nucleic Acids Res.* **33**, 7074–89 (2005).

83. Patel, G. P. & Bag, J. IMP1 interacts with poly(A)-binding protein (PABP) and the autoregulatory translational control element of PABP-mRNA through the KH III-IV domain. *FEBS J.* **273**, 5678–5690 (2006).

84. Moallem, E., Kilav, R., Silver, J. & Naveh-Many, T. RNA-protein binding and post-transcriptional regulation of parathyroid hormone gene expression by calcium and phosphate. *J. Biol. Chem.* **273**, 5253–5259 (1998).

85. Dinur, M., Kilav, R., Sela-Brown, A., Jacquemin-Sablon, H. & Naveh-Many, T. *In Vitro* Evidence that Upstream of N- *ras* Participates in the Regulation of Parathyroid Hormone Messenger Ribonucleic Acid Stability. *Mol. Endocrinol.* **20**, 1652–1660 (2006).

86. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* **73**, 1–16 (2013).

87. Chang, T. C. *et al.* UNR, a new partner of poly(A)-binding protein, plays a key role in translationally coupled mRNA turnover mediated by the c-fos major coding-region determinant. *Genes Dev.* **18**, 2010–2023 (2004).

88. Grosset, C. *et al.* A Mechanism for Translationally Coupled mRNA Turnover: Interaction between the Poly(A) Tail and a c-fos RNA Coding Determinant via a Protein Complex. *Cell* **103**, 29–40 (2000).

89. Ju Lee, H. *et al.* A post-transcriptional program coordinated by CSDE1 prevents intrinsic neural differentiation of human embryonic stem cells. *Nat. Commun.* **8**, (2017).

90. Booy, E. P. *et al.* Comprehensive analysis of the BC200 ribonucleoprotein reveals a reciprocal regulatory function with CSDE1/UNR. 11575–11591 (2018).

91. Wurth, L. *et al.* UNR/CSDE1 Drives a Post-transcriptional Program to Promote Melanoma Invasion and Metastasis. *Cancer Cell* **30**, 694–707 (2016).

92. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–41 (2012).

93. Xia, K. *et al.* Common genetic variants on 1p13.2 associate with risk of autism. *Mol. Psychiatry* **19**, 1212–1219 (2014).

94. Guo, H. *et al.* Disruptive variants of CSDE1 associate with autism and interfere with neuronal development and synaptic transmission. *Sci. Adv.* **5**, eaax2166 (2019).

95. Fishbein, L. *et al.* Comprehensive Molecular Characterization of Pheochromocytoma and Paraganglioma. *Cancer Cell* **31**, 181–193 (2017).

96. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science (80-.).* **347**, 1260419–1260419 (2015).

97. BARR, M. L. & BERTRAM, E. G. A Morphological Distinction between Neurones of the Male and Female and the Behaviour of the Nucleolar Satellite during Accelerated

Nucleoprotein Synthesis. *Nature* **163**, 676–677 (1949).

98. Muller, H. J. Further studies on the nature and cause of gene mutations. in *International Congress of Genetics Proc. 6th* 213–255 (1932).

99. Maenner, S., Müller, M., Fröhlich, J., Langer, D. & Becker, P. B. ATP-Dependent roX RNA Remodeling by the Helicase maleless Enables Specific Association of MSL Proteins. *Mol. Cell* **51**, 174–184 (2013).

100. Ilik, I. A. *et al.* Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. *Mol. Cell* **51**, 156–173 (2013).

101. Militti, C., Maenner, S., Becker, P. B. & Gebauer, F. UNR facilitates the interaction of MLE with the lncRNA roX2 during Drosophila dosage compensation. *Nat. Commun.* **5**, 4762 (2014).

102. Larschan, E. *et al.* MSL Complex Is Attracted to Genes Marked by H3K36 Trimethylation Using a Sequence-Independent Mechanism. *Mol. Cell* **28**, 121–133 (2007).

103. Bell, O. *et al.* Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster. *Mol. Cell. Biol.* **28**, 3401–9 (2008).

104. Straub, T., Grimaud, C., Gilfillan, G. D., Mitterweger, A. & Becker, P. B. The Chromosomal High-Affinity Binding Sites for the Drosophila Dosage Compensation Complex. *PLoS Genet.* **4**, e1000302 (2008).

105. Villa, R., Schauer, T., Smialowski, P., Straub, T. & Becker, P. B. PionX sites mark the X chromosome for dosage compensation. *Nature* **537**, 244–248 (2016).

106. Kelley, R. L. *et al.* Epigenetic Spreading of the Drosophila Dosage Compensation Complex from roX RNA Genes into Flanking Chromatin. *Cell* **98**, 513–522 (1999).

107. Smith, E. R. *et al.* A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol. Cell. Biol.* **25**, 9175–88 (2005).

108. Turner, B. M., Birley, A. J. & Lavender, J. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. *Cell* **69**, 375–384 (1992).

109. Taipale, M. *et al.* hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell. Biol.* **25**, 6798–810 (2005).

110. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes and Development* **20**, 2922–2936 (2006).

111. Chlamydas, S. *et al.* Functional interplay between MSL1 and CDK7 controls RNA polymerase II Ser5 phosphorylation. *Nat. Struct. Mol. Biol.* **23**, 580–589 (2016).

112. Abaza, I., Coll, O., Patalano, S. & Gebauer, F. Drosophila UNR is required for translational repression of male-specific lethal 2 mRNA during regulation of X-

chromosome dosage compensation. *Genes Dev.* **20**, 380–9 (2006).

113. Duncan, K. *et al.* Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3' UTR: translational repression for dosage compensation. *Genes Dev.* **20**, 368–79 (2006).

114. Abaza, I. & Gebauer, F. Functional domains of Drosophila UNR in translational control. *RNA* **14**, 482–90 (2008).

115. Förch, P., Merendino, L., Martínez, C. & Valcárcel, J. Modulation of msl-2 5′ splice site recognition by Sex-lethal. *RNA* **7**, 1185–1191 (2001).

116. Merendino, L., Guth, S., Bilbao, D., Martínez, C. & Valcárcel, J. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838–841 (1999).

117. Graindorge, A., Carré, C. & Gebauer, F. Sex-lethal promotes nuclear retention of msl2 mRNA via interactions with the STAR protein HOW. *Genes Dev.* **27**, 1421–33 (2013).

118. Bashaw, G. J. & Baker, B. S. The Regulation of the Drosophila msl-2 Gene Reveals a Function for Sex-lethal in Translational Control. *Cell* **89**, 789–798 (1997).

119. Kelley, R. L., Wang, J., Bell, L. & Kuroda, M. I. Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism. *Nature* **387**, 195–199 (1997).

120. Gebauer, F., Merendino, L., Hentze, M. W. & Valcárcel, J. The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. *RNA* **4**, 142–150 (1998).

121. Beckmann, K., Grskovic, M., Gebauer, F. & Hentze, M. W. z. *Cell* **122**, 529–540 (2005).

122. Medenbach, J., Seiler, M. & Hentze, M. W. Translational control via protein-regulated upstream open reading frames. *Cell* **145**, 902–913 (2011).

123. Grskovic, M., Hentze, M. W. & Gebauer, F. A co-repressor assembly nucleated by Sex-lethal in the 3'UTR mediates translational control of Drosophila msl-2 mRNA. *EMBO J.* **22**, 5571–5581 (2003).

124. Szostak, E. *et al.* Hrp48 and eIF3d contribute to msl-2 mRNA translational repression. *Nucleic Acids Res.* **46**, 4099–4113 (2018).

125. Duncan, K. E., Strein, C. & Hentze, M. W. The SXL-UNR Corepressor Complex Uses a PABP-Mediated Mechanism to Inhibit Ribosome Recruitment to msl-2 mRNA. *Mol. Cell* **36**, 571–582 (2009).

126. Jacquemin-Sablon, H. *et al.* Nucleic acid binding and intracellular localization of unr, a protein with five cold shock domains. *Nucleic Acids Res.* **22**, 2643–50 (1994).

127. Goroncy, A. K. *et al.* The NMR solution structures of the five constituent cold-shock domains (CSD) of the human UNR (upstream of N-ras) protein. *J. Struct. Funct. Genomics* **11**, 181–188 (2010).

128. Max, K. E. A., Zeeb, M., Bienert, R., Balbach, J. & Heinemann, U. Common mode of DNA binding to cold shock domains. *FEBS J.* **274**, 1265–1279 (2007).

129. Sachs, R., Max, K. E. A., Heinemann, U. & Balbach, J. RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *RNA* **18**, 65–76 (2012).

130. Schindelin, H., Marahiel, M. A. & Heinemann, U. Universal nucleic acid-binding domain revealed by crystal structure of the B. subtilis major cold-shock protein. *Nature* **364**, 164–168 (1993).

131. Schnuchel, A. *et al.* Structure in solution of the major cold-shock protein from Bacillus subtilis. *Nature* **364**, 169–171 (1993).

132. Doniger, J., Landsman, D., Gonda, M. A. & Wistow, G. The product of unr, the highly conserved gene upstream of N-ras, contains multiple repeats similar to the cold-shock domain (CSD), a putative DNA-binding motif. *New Biol.* **4**, 389–395 (1992).

133. Landsman, D. RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res.* **20**, 2861—2864 (1992).

134. Max, K. E. A., Zeeb, M., Bienert, R., Balbach, J. & Heinemann, U. T-rich DNA Single Strands Bind to a Preformed Site on the Bacterial Cold Shock Protein Bs-CspB. *J. Mol. Biol.* **360**, 702–714 (2006).

135. Triqueneaux, G., Velten, M., Franzon, P., Dautry, F. & Jacquemin-Sablon, H. RNA binding specificity of Unr, a protein with five cold shock domains. *Nucleic Acids Res.* **27**, 1926–34 (1999).

136. Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C. & Stuckey, P. J. Automatic generation of protein structure cartoons with Pro-origami. in *Bioinformatics (Oxford, England)* 3315–3316 (2011). doi:https://doi.org/10.1093/bioinformatics/btr575

137. Graumann, P., Wendrich, T. M., Weber, M. H. W., Schröder, K. & Marahiel, M. A. A family of cold shock proteins in *Bacillus subtilis* is essential for cellular growth and for efficient protein synthesis at optimal and low temperatures. *Mol. Microbiol.* **25**, 741–756 (1997).

138. Jiang, W., Hou, Y. & Inouye, M. CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *J. Biol. Chem.* **272**, 196–202 (1997).

139. Kljashtorny, V. *et al.* The Cold Shock Domain of YB-1 segregates RNA from DNA by non-bonded interactions. *PLoS One* **10**, (2015).

140. Yang, Y. *et al.* RNA 5-Methylcytosine Facilitates the Maternal-to-Zygotic Transition by Preventing Maternal mRNA Decay. *Mol. Cell* **75**, 1188-1202.e11 (2019).

141. Zou, F. *et al.* Drosophila YBX1 homolog YPS promotes ovarian germ line stem cell development by preferentially recognizing 5-methylcytosine RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3603–3609 (2020).

142. Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177**, 1384–1403 (2019).

143. Ward, A. B., Sali, A. & Wilson, I. A. Integrative structural biology. *Science* **339**, 913–915 (2013).

144. Foster, M. P., McElroy, C. A. & Amero, C. D. Solution NMR of large molecules and assemblies. *Biochemistry* **46**, 331–340 (2007).

145. Schütz, S. & Sprangers, R. Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Progress in Nuclear Magnetic Resonance Spectroscopy* **116**, 56–84 (2020).

146. Jiang, Y. & Kalodimos, C. G. NMR Studies of Large Proteins. *Journal of Molecular Biology* **429**, 2667–2676 (2017).

147. Larmor, J. Part 3, Relations with material media. in *On a Dynamical Theory of the Electric and Luminiferous Medium* 221–231 (1897).

148. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189 (1980).

149. Gardner, K. H. & Kay, L. E. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure* **27**, 357–406 (1998).

150. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. 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. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12366–12371 (1997).

151. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. Trosy in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13585–13590 (1998).

152. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**, 93–158 (1999).

153. Overhauser, A. W. Polarization of nuclei in metals. *Phys. Rev.* **92**, 411–415 (1953).

154. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).

155. Tjandra, N. & Bax, A. Measurement of Dipolar Contributions to 1 J CH Splittings from Magnetic-Field Dependence of J Modulation in Two-Dimensional NMR Spectra. *Journal of Magnetic Resonance* **124**, 512–515 (1997).

156. Tjandra, N., Szabo, A. & Bax, A. Protein backbone dynamics and 15N chemical shift anisotropy from quantitative measurement of relaxation interference effects. *J. Am. Chem. Soc.* **118**, 6986–6991 (1996).

157. Solomon, I. Relaxation processes in a system of two spins. *Phys. Rev.* **99**, 559–565 (1955).

158. Kay, L. E., Torchia, D. A. & Bax, A. Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal

Nuclease. *Biochemistry* **28**, 8972–8979 (1989).

159. Rossi, P. *et al.* A microscale protein NMR sample screening pipeline. *J. Biomol. NMR* **46**, 11–22 (2010).

160. Callaway, E. Revolutionary cryo-EM is taking over structural biology. *Nature* **578**, 201 (2020).

161. Callaway, E. The revolution will not be crystallized: A new method sweeps through structural biology. *Nature* **525**, 172–174 (2015).

162. Weiss, C. Ueber eine verbesserte Methode für die Bezeichnung der verschiedenen Flächen eines Krystallisationssystems, nebst Bemerkungen über den Zustand der Polarisierung der Seiten in den Linien der krystallinischen Structur. in *Abhandlungen der physikalischen Klasse der Königlich-Preussischen Akademie der Wissenschaften* 286–336 (1817).

163. Miller, W. *A treatise on crystallography*. (1839).

164. Bragg, W. & Bragg, W. The reflection of X-rays by crystals. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* **88**, 428–438 (1913).

165. Rossmann, M. & Blow, D. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **15**, 24.31 (1962).

166. The structure of haemoglobin - IV. Sign determination by the isomorphous replacement method. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **225**, 287–307 (1954).

167. Perutz, M. F. Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Crystallogr.* **9**, 867–873 (1956).

168. Blow, D. M. & Rossmann, M. G. The single isomorphous replacement method. *Acta Crystallogr.* **14**, 1195–1202 (1961).

169. Hendrickson, W. A. Analysis of Protein Structure from Diffraction Measurement at Multiple Wavelengths. *Trans. ACA* **21**, (1985).

170. Karle, J. Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int. J. Quantum Chem.* **18**, 357–367 (1980).

171. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).

172. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).

173. Braman, J., Papworth, C. & Greener, A. Site-Directed Mutagenesis Using Double-Stranded Plasmid DNA Templates. in *In Vitro Mutagenesis Protocols* 31–44 (Humana Press, 1996). doi:10.1385/0-89603-332-5:31

174. Birnboim, H. C. & Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. **7**, 1513–1523 (1979).

175. Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Ann. Phys.* **162**, 78–88 (1852).

176. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3**, 318–356 (1961).

177. Tugarinov, V. & Kay, L. E. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J. Am. Chem. Soc.* **125**, 13868–13878 (2003).

178. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–5 (1970).

179. Smith, P. K. *et al.* Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76–85 (1985).

180. Kabsch, W. *et al. XDS. Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

181. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861–877 (2019).

182. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D. Biol. Crystallogr.* **65**, 582–601 (2009).

183. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

184. Schilling, F. *et al.* Next-Generation Heteronuclear Decoupling for High-Field Biomolecular NMR Spectroscopy. *Angew. Chemie Int. Ed.* **53**, 4475–4479 (2014).

185. Simon, B. & Köstler, H. Improving the sensitivity of FT-NMR spectroscopy by apodization weighted sampling. *J. Biomol. NMR* **73**, 155–165 (2019).

186. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–93 (1995).

187. Keller, R. The Computer Aided Resonance Assignment tutorial. *Goldau: CANTINA Verlag* (2004).

188. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 453–471 (2015).

189. Rieping, W. *et al.* ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381–382 (2007).

190. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**, 315–316 (2003).

191. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).

192. Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. Errors in protein structures [3]. *Nature* **381**, 272 (1996).

193. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for

biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

194. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* **59**, 687–696 (2005).

195. Zhu, G., Xia, Y., Nicholson, L. K. & Sze, K. H. Protein Dynamics Measurements by TROSY-Based NMR Experiments. *Journal of Magnetic Resonance* **143**, 423–426 (2000).

196. Niklasson, M. *et al.* Comprehensive analysis of NMR data using advanced line shape fitting. *J. Biomol. NMR* **69**, 93–99 (2017).

197. Ahlner, A., Carlsson, M., Jonsson, B. H. & Lundström, P. PINT: A software for integration of peak volumes and extraction of relaxation rates. *J. Biomol. NMR* **56**, 191–202 (2013).

198. Pernot, P. *et al.* Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **20**, 660–664 (2013).

199. Blanchet, C. E. *et al.* Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**, 431–443 (2015).

200. Trewhella, J. *et al.* 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: An update. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 710–728 (2017).

201. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).

202. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).

203. Svergun, D., Barberato, C. & Koch, M. H. J. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).

204. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350 (2011).

205. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci. Rep.* **7**, 10480 (2017).

206. Bienert, S. *et al.* The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* **45**, D313–D319 (2017).

207. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30**, S162–S173 (2009).

208. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

209. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733 (2007).

210. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **54**, 905–21 (1998).

211. Simon, B., Madl, T., Mackereth, C. D., Nilges, M. & Sattler, M. An Efficient Protocol for NMR-Spectroscopy-Based Structure Determination of Protein Complexes in Solution. *Angew. Chemie Int. Ed.* **49**, 1967–1970 (2010).

212. Beckert, B. & Masquida, B. Synthesis of RNA by in vitro transcription. *Methods Mol. Biol.* **703**, 29–41 (2011).

213. Milligan, J. F. & Uhlenbeck, O. C. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51–62 (1989).

214. Saville, B. J. & Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in neurospora mitochondria. *Cell* **61**, 685–696 (1990).

215. Hutchins, C. J., Rathjen, P. D., Forster, A. C. & Symons, R. H. Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res.* **14**, 3627–3640 (1986).

216. Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. *Science (80-.).* **231**, 1577–1580 (1986).

217. Tannock, G. A. The nucleic acid of infectious bronchitis virus. *Arch. Gesamte Virusforsch.* **43**, 259–271 (1973).

218. Zearfoss, N. R. & Ryder, S. P. End-labeling oligonucleotides with chemical tags after synthesis. *Methods Mol. Biol.* **941**, 181–193 (2012).

219. Garner, M. M. & Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047–3060 (1981).

220. Fried, M. & Crothers, D. M. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505–6525 (1981).

221. Weber, G. Rotational Brownian Motion and Polarization of the Fluorescence of Solutions. *Adv. Protein Chem.* **8**, 415–459 (1953).

222. Senisterra, G., Chau, I. & Vedadi, M. Thermal denaturation assays in chemical biology. *Assay and Drug Development Technologies* **10**, 128–136 (2012).

223. Nakanishi, K., Berova, N. & Woody, R. *Circular dichroism : principles and applications*. (VCH, 1994).

224. Sreerama, N. & Woody, R. W. On the analysis of membrane protein circular dichroism spectra. *Protein Sci.* **13**, 100–12 (2004).

225. Sreerama, N. & Woody, R. W. A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism. *Anal. Biochem.* **209**, 32–44 (1993).

226. Gebauer, F., Corona, D. F. V., Preiss, T., Becker, P. B. & Hentze, M. W. Translational control of dosage compensation in Drosophila by Sex-lethal: cooperative silencing via the 5' and 3' UTRs of msl-2 mRNA is independent of the poly(A) tail. *EMBO J.* **18**, 6146–6154 (1999).

227. Felgner, P. L. *et al.* Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7413–7417 (1987).

228. Bronstein, I., Edwards, B. & Voyta, J. C. 1,2-dioxetanes: novel chemiluminescent enzyme substrates. Applications to immunoassays. *J. Biolumin. Chemilumin.* **4**, 99–111 (1989).

229. Cormier, M. J. *et al.* Evidence for similar biochemical requirements for bioluminescence among the coelenterates. *J. Cell. Physiol.* **81**, 291–297 (1973).

230. Matthews, J. C., Hori, K. & Cormier, M. J. Purification and Properties of Renilla reniformis Luciferase. *Biochemistry* **16**, 85–91 (1977).

231. Gaspar, I., Wippich, F. & Ephrussi, A. Enzymatic production of single-molecule FISH and RNA capture probes. *RNA* **23**, 1582–1591 (2017).

232. Gáspár, I., Wippich, F. & Ephrussi, A. Terminal Deoxynucleotidyl Transferase Mediated Production of Labeled Probes for Single-molecule FISH or RNA Capture. *BIO-PROTOCOL* **8**, (2018).

233. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

234. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

235. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing,. (2017). Available at: https://www.r-project.org/.

236. Blighe K, Rana S, L. M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.4.0. (2019).

237. Wickham, H. *ggplot2*. *ggplot2* (Springer New York, 2009). doi:10.1007/978-0-387-98141-3

238. Chen, H. Generate High-Resolution Venn and Euler Plots. (2018).

239. Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757 (2014).

240. Moggridge, S., Sorensen, P. H., Morin, G. B. & Hughes, C. S. Extending the Compatibility of the SP3 Paramagnetic Bead... Suppl. *J. Proteome Res.* **17**, 1730–1740 (2018).

241. Werner, T. *et al.* Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal. Chem.* **86**, 3594–601 (2014).

242. Sridharan, S. *et al.* Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP. *Nat. Commun.* **10**, 1155 (2019).

243. Franken, H. *et al.* Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat. Protoc.* **10**, 1567–1593 (2015).

244. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

245. Savitski, M. M., Wilhelm, M., Hahne, H., Kuster, B. & Bantscheff, M. A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets. *Mol. Cell. Proteomics* **14**, 2394–404 (2015).

246. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

247. Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. in *Bioinformatics* **18**, (Oxford University Press, 2002).

248. Gatto, L. & Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **28**, 288–9 (2012).

249. Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–2 (2008).

250. Steinhaus, H. Sur la division des corps matériels en parties. *Bull. Acad. Pol. Sci. (in French)* **4**, 801–804 (1957).

251. MacQueen, J. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability: Some Methods for classification and Analysis of Multivariate Observations*. (University of California Press, 1967).

252. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

253. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3 (2003).

254. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).

255. Hollmann, N. M. *et al.* Pseudo-RNA-Binding Domains Mediate RNA Structure Specificity in Upstream of N-Ras. *Cell Rep.* **32**, 107930 (2020).

256. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).

257. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

258. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

259. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

260. Poon, D. K. Y., Withers, S. G. & McIntosh, L. P. Direct demonstration of the flexibility of the glycosylated proline-threonine linker in the Cellulomonas fimi Xylanase Cex through NMR spectroscopic analysis. *J. Biol. Chem.* **282**, 2091–100 (2007).

261. Receveur, V., Czjzek, M., Schülein, M., Panine, P. & Henrissat, B. Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *J. Biol. Chem.* **277**, 40887–92 (2002).

262. Allain, F. H. T., Bouvet, P., Dieckmann, T. & Feigon, J. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* **19**, 6870–6881 (2000).

263. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845 (1999).

264. Handa, N. *et al.* Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–584 (1999).

265. Pérez-Cãadillas, J. M. Grabbing the message: Structural basis of mRNA 3′UTR recognition by Hrp1. *EMBO J.* **25**, 3167–3178 (2006).

266. Wang, X. & Hall, T. M. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **8**, 141–145 (2001).

267. Yang, X. J. *et al.* Crystal structure of a Y-box binding protein 1 (YB-1)–RNA complex reveals key features and residues interacting with RNA. *J. Biol. Chem.* **294**, 10998–11010 (2019).

268. Welte, H., Zhou, T., Mihajlenko, X., Mayans, O. & Kovermann, M. What does fluorine do to a protein? Thermodynamic, and highly-resolved structural insights into fluorine-labelled variants of the cold shock protein. *Sci. Rep.* **10**, 2640–2640 (2020).

269. Moschall, R., Strauss, D., García-Beyaert, M., Gebauer, F. & Medenbach, J. Drosophila sister-of-sex-lethal is a repressor of translation. *RNA* **24**, 149–158 (2018).

270. Moschall, R. *et al. Drosophila* Sister-of-Sex-lethal reinforces a male-specific gene expression pattern by controlling *Sex-lethal* alternative splicing. *Nucleic Acids Res.* **47**, 2276–2288 (2019).

271. Gatfield, D. Nonsense-mediated mRNA decay in Drosophila:at the intersection of the yeast and mammalian pathways. *EMBO J.* **22**, 3960–3970 (2003).

272. Fenger-Grøn, M., Fillman, C., Norrild, B. & Lykke-Andersen, J. Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. *Mol. Cell* **20**, 905–915 (2005).

273. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).

274. Nielsen, F. C., Nielsen, J. & Christiansen, J. A family of IGF-II mRNA binding proteins (IMP) involved in RNA trafficking. in *Scandinavian Journal of Clinical and Laboratory Investigation, Supplement* **61**, 93–99 (2001).

275. Wilhelm, J. E., Buszczak, M. & Sayles, S. Efficient protein trafficking requires trailer hitch, a component of a ribonucleoprotein complex localized to the ER in Drosophila. *Dev. Cell* **9**, 675–685 (2005).

276. Rodrigues, J. P. REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc. Natl. Acad. Sci.* **98**, 1030–1035 (2001).

277. Bates, G. J. *et al.* The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *EMBO J.* **24**, 543–553 (2005).

278. Lin, C., Yang, L., Yang, J. J., Huang, Y. & Liu, Z.-R. ATPase/Helicase Activities of p68 RNA Helicase Are Required for Pre-mRNA Splicing but Not for Assembly of the Spliceosome. *Mol. Cell. Biol.* **25**, 7484–7493 (2005).

279. Lence, T. *et al.* M6A modulates neuronal functions and sex determination in Drosophila. *Nature* **540**, 242–247 (2016).

280. King, M. R., Matzat, L. H., Dale, R. K., Lim, J. J. & Lei, E. P. The RNA-binding protein Rumpelstiltskin antagonizes gypsy chromatin insulator function in a tissue-specific manner. *J. Cell Sci.* **127**, 2956–2966 (2014).

281. Gattoni, R. *et al.* The human hnRNP-M proteins: Structure and relation with early heat shock-induced splicing arrest and chromosome mapping. *Nucleic Acids Res.* **24**, 2535–2542 (1996).

282. Wang, M. *et al.* ME31B globally represses maternal mRNAs by two distinct mechanisms during the Drosophila maternal-to-zygotic transition. *Elife* **6**, (2017).

283. Lim, C. *et al.* The novel gene twenty-four defines a critical translational step in the Drosophila clock. *Nature* **470**, 399–403 (2011).

284. Nelson, M. R. *et al.* A multiprotein complex that mediates translational enhancement in Drosophila. *J. Biol. Chem.* **282**, 34031–34038 (2007).

285. Hansen, H. T. *et al.* Drosophila Imp iCLIP identifies an RNA assemblage coordinating F-actin formation. *Genome Biol.* **16**, 123 (2015).

286. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

287. Ruschak, A. M. & Kay, L. E. Methyl groups as probes of supra-molecular structure, dynamics and function. *Journal of Biomolecular NMR* **46**, 75–87 (2010).

288. Kelley, R. L. *et al.* Expression of Msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila. *Cell* **81**, 867–877 (1995).

289. Sigrist, S. J. *et al.* Postsynaptic translation affects the efficacy and morphology of

neuromuscular junctions. *Nature* **405**, 1062–1065 (2000).

290. Deardorff, J. A. & Sachs, A. B. Differential effects of aromatic and charged residue substitutions in the RNA binding domains of the yeast poly(A)-binding protein. *J. Mol. Biol.* **269**, 67–81 (1997).

291. Baer, B. W. & Kornberg, R. D. The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein. *J. Cell Biol.* **96**, 717–721 (1983).

292. Kühn, U. & Pieler, T. Xenopus poly(A) binding protein: Functional domains in RNA binding and protein-protein interaction. *J. Mol. Biol.* **256**, 20–30 (1996).

293. Goss, D. J. & Kleiman, F. E. Poly(A) binding proteins: Are they all created equal? *Wiley Interdisciplinary Reviews: RNA* **4**, 167–179 (2013).

294. Maris, C., Dominguez, C. & Allain, F. H. T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal* **272**, 2118–2131 (2005).

295. Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R. & Taylor, I. A. Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J.* **24**, 3576–3587 (2005).

296. Oberstrass, F. C. *et al.* Structural biology - Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. *Science (80-.).* **309**, 2054–2057 (2005).

297. Ule, J. *et al.* CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science (80-.).* **302**, 1212–1215 (2003).

298. Zhao, J. *et al.* Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953 (2010).

299. Ray, D. *et al.* Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.* **27**, 667–670 (2009).

300. P. Bagowski, C., Bruins, W. & J.W. te Velthuis, A. The Nature of Protein Domain Evolution: Shaping the Interaction Network. *Curr. Genomics* **11**, 368–376 (2010).

301. Weiner, J., Moore, A. D. & Bornberg-Bauer, E. Just how versatile are domains? *BMC Evol. Biol.* **8**, (2008).

302. Apic, G., Gough, J. & Teichmann, S. A. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **310**, 311–325 (2001).

303. Kuznetsov, V. A., Pickalov, V. V., Senko, O. V. & Knott, G. D. Analysis of the evolving proteomes: Predictions of the number of protein domains in nature and the number of genes in eukaryotic organisms. *J. Biol. Syst.* **10**, 381–407 (2002).

304. Forslund, S. K., Kaduk, M. & Sonnhammer, E. L. L. Evolution of protein domain architectures. in *Methods in Molecular Biology* **1910**, 469–504 (Humana Press Inc., 2019).

305. Mitchell, A. L. *et al.* InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).

306. Ranea, J. A. G., Sillero, A., Thornton, J. M. & Orengo, C. A. Protein superfamily

evolution and the Last Universal Common Ancestor (LUCA). *J. Mol. Evol.* **63**, 513–525 (2006).

307. Chothia, C. & Gough, J. Genomic and structural aspects of protein evolution. *Biochemical Journal* **419**, 15–28 (2009).

308. Patalano, S., Mihailovich, M., Belacortu, Y., Paricio, N. & Gebauer, F. Dual sex-specific functions of Drosophila Upstream of N-ras in the control of X chromosome dosage compensation. *Development* **136**, 689–98 (2009).

309. Triqueneaux, G., Velten, M., Franzon, P., Dautry, F. & Jacquemin-Sablon, H. RNA binding specificity of Unr, a protein with five cold shock domains. *Nucleic Acids Res.* **27**, 1926–34 (1999).

310. Hunt, S. L., Hsuan, J. J., Totty, N. & Jackson, R. J. unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. *Genes Dev.* **13**, 437–48 (1999).

311. Zondlo, N. J. Aromatic-proline interactions: Electronically tunable CH/π interactions. *Acc. Chem. Res.* **46**, 1039–1049 (2013).

312. Morozova, N., Allers, J., Myers, J. & Shamoo, Y. Protein-RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics* **22**, 2746–52 (2006).

313. Acuner Ozbabacan, S. E., Engin, H. B., Gursoy, A. & Keskin, O. Transient protein–protein interactions. *Protein Eng. Des. Sel.* **24**, 635–648 (2011).

314. Hub, J. S. Interpreting solution X-ray scattering data using molecular simulations. *Current Opinion in Structural Biology* **49**, 18–26 (2018).

315. Chen, P. C. & Hub, J. S. Interpretation of solution x-ray scattering by explicit-solvent molecular dynamics. *Biophys. J.* **108**, 2573–2584 (2015).

316. Stein, A., Pache, R. A., Bernadó, P., Pons, M. & Aloy, P. Dynamic interactions of proteins in complex networks: A more structured view. *FEBS Journal* **276**, 5390–5405 (2009).

317. Bhattacharyya, R. P., Reményi, A., Yeh, B. J. & Lim, W. A. Domains, Motifs, and Scaffolds: The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits. *Annu. Rev. Biochem.* **75**, 655–680 (2006).

318. Siddiqui, N., Kozlov, G., D'Orso, I., Trempe, J.-F. & Gehring, K. Solution structure of the C-terminal domain from poly(A)-binding protein in Trypanosoma cruzi : A vegetal PABC domain . *Protein Sci.* **12**, 1925–1933 (2003).

319. Kozlov, G. *et al.* Structure and function of the C-terminal PABC domain of human poly(A)-binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4409–4413 (2001).

320. Albrecht, M. & Lengauer, T. Survey on the PABC recognition motif PAM2. *Biochem. Biophys. Res. Commun.* **316**, 129–138 (2004).

321. Safaee, N. *et al.* Interdomain Allostery Promotes Assembly of the Poly(A) mRNA

Complex with PABP and eIF4G. *Mol. Cell* **48**, 375–386 (2012).

322. Kawahara, H. *et al.* Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. *J. Cell Biol.* **181**, 639–653 (2008).

323. Khaleghpour, K. *et al.* Dual Interactions of the Translational Repressor Paip2 with Poly(A) Binding Protein. *Mol. Cell. Biol.* **21**, 5200–5213 (2001).

324. Khaleghpour, K. *et al.* Translational repression by a novel partner of human poly(A) binding protein, Paip2. *Mol. Cell* **7**, 205–216 (2001).

Erratum to my thesis

Nele Merret Hollmann

In my thesis the labeling of the Figure 25c is incorrect. The rotational correlation time of the unbound protein CSD789 is supposed to be highlighted in red and the RNA bound state in blue.