Interaction of the spliced Oskar localization element of Oskar mRNA with the protein PYM

Von der Naturwissenschaftlichen Fakultät der Gottffried Wilhelm

Leibniz Universität Hanover

zur Erlangung des Grades

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Veena Hegde, M.Sc. (India)

2023

Referentin: Prof. Dr. ric. Teresa Carlomagno

Korreferent: Prof. Dr. Russell J. Cox

Tag der Promotion: 23.03.2023

Dedication

To my Mahaganapati chaami,
Kendamahasati amnoru, Baba, Svarnavalli Krupe and my family.

ABSTRACT

mRNAs and the process of mRNA localization are the fundamental and pivotal parts of cellular functions. mRNA localization encompasses an important role in cellular differentiation and site-specific cellular functions, from the basic cellular biochemical mechanism to advanced abdomen formation. The study of mRNA, its localization mechanism along its binding partners have always been the main focus of study for several years. As they define life, in terms of cellular and sub-cellular mechanisms. Our study also involves one of the binding partners of the localization complex, which is Pym protein. Pym protein and exon junction complex are the common localization binding partners to many mRNA localization and *Oskar* mRNA is one of them. Pym being one of the recycling factors of the Exon Junction Complex shows binding interactions with many components, such as RNAs, Exon junction Complex, and Ribosomes.

Our results show interesting structural and binding features of the protein Pym. NMR studies reveal that Pym^{160}, the shorter construct of Pym is structurally unfolded, with the general characteristic of an intrinsically disordered protein. It has the long helical structural element in the middle part of the protein, while both N-terminal and C-terminal ends remain highly flexible with the structurally unfolded regions. The C-terminal part of the protein is not showing any direct involvement in the interaction with the SOLE RNA. However, it is structurally a very important part of the protein, as it stabilizes the ionic and hydrophobic interactions of the protein, so that protein could able to be a stable soluble protein. We have studied the binding motifs of the protein Pym^{160} with SOLE RNA and its isomers. Pym^{160} has binding motifs in the N-terminal region and in the middle helical region. Studies have confirmed that the N-terminal part of the protein binds to the Y14-Mago heterodimer, which is an essential part of the exon junction complex. In the absence of an Exon Junction Complex, the N-terminal part of the protein binds to the RNA. So, the study of the protein Pym^{160} is very much interesting and essential as it is a common protein for the wide range of mRNA localization mechanisms. Our studies explain the widespread binding nature of the Pym^{160}, which might be due to its functional significance of being a structurally unfolded protein.

Keywords: NMR, Protein, RNA, IDP.

Contents

Abbreviations and Symbols

1D, 2D, 3D	1-, 2-, 3- Dimensional
A	Adenine
ARIA	Ambiguous Restraints for Iterative Assignment
ATP	Adenine TRiphosphate
C	Cytosine
CNS	Crystallography and NMR System
CSA	Chemical Shift Anisotrophy
CSP	Chemical Shift Perturbation
CTP	Cytosine Triphosphate
Da	Dalton
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide triphosphate
dsRNA	Double stranded Ribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
G	Guanine
GTP	Guanosine triphosphate
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heternuclear Single Quantum Coherence
Hz	Hertz
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Kelvin
kDa	KiloDalton
MHz	MegaHertz
mRNA	Messanger Ribonucleic Acid
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
nt	Nucleotide
OD	Optical Density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Entry of the Protein DATA Bank
ppm	Parts per million
PRE	Paramagnetic relaxation enhancement
RDC	Residual dipolar coupling
RMSD	Root mean square deviation
RNA	Ribonucleic acid
rRNA	Ribosomal Ribonucleic acid
RNase	Ribonuclease
SDS	Sodium dodecyl sulphate
TEMED	*N,N,N,N*-Tetramethylethylenediamine
TROSY	Transverse relaxation optimized spectroscopy
tRNA	Transfer Ribonucleic acid
U	Uridine
UTP	Uridine triphosphate
UV	Ultraviolet
V	Volt

CHAPTER

1. Introduction

1.1. Ribonucleic acid

RNA, the Ribonucleic acid is one of the biological polymeric macro-molecules that exists in all forms of life, known to perform a wide range of functions. The central dogma of molecular biology apprises that genetic information passes onto the proteins via RNA through the transcription and translation process.

1.1.1 Types of RNA

mRNA: Messenger RNAs carry genomic information from DNA to the ribosome, where the protein translation takes place. mRNA encodes a genetic sequence of a gene, which can be read by the ribosomes during the translation process.
tRNA: tRNAs do not encode genomic information for translation. But they do involve different cytoplasmic functions, for example, Transfer RNA (tRNA) involve in the process of translation, they transfer amino acids to the ribosomes during translation.
rRNA: Ribosomal RNAs are the ribosomal units that synthesize proteins. Ribosomal RNAs bound to ribosomal proteins form small and large ribosomal units. The ribosomal units translate the mRNA.

1.1.2 Post-transcriptional RNA processing

RNA modifications to a newly transcribed primary RNA transcript preceding its translation into the protein are post-transcriptional or co-transcriptional processing[1]. This is prevalent in eukaryotes, however absent in prokaryotes. In prokaryotes, RNA synthesis during transcription is functionally organized for further downstream biochemical processes. RNA processes are biologically important as they facilitate efficient translation, binding, and producing functionally active proteins.

Modification of primary RNA transcript into mature RNA takes place mainly by three different processes. These modifications are 5′ processing, 3′ processing, and RNA splicing. 5′ processing: 5′ processing involves the capping of the 5′ end of primary RNA transcript that is the addiction of 7- methylguanosine to the 5′ end. 5′ cap is an altered nucleotide on the 5′ end of some RNAs like mRNAs and is a highly regulated and important creation of mature mRNAs which undergo translation during protein synthesis[2].
3′ processing: 3′ processing involves cleavage and polyadenylation processes at the 3′ end of the primary transcript RNA. Cleavage and adenylation reactions include the formation of poly(A) tail. In other words, the RNA stretch has a series of adenine bases. In eukaryotes, it is a vital part of the process of RNA processing. Poly(A) tail is vita for the nuclear export, stability, and translation processes[3].

RNA splicing: In other words also called intron splicing. It is the process where introns of pre-mRNA which do not code for proteins are removed to form exon ligated product[4,5]. The remaining exons after the specific splicing are connected to form a single RNA mature molecule, which is a functional motif of the translation process. Splicing also promotes the processes of localization, where the exon-exon junction complex binds to intron spliced exon-exon ligated product to facilitate the process of localization. Oskar mRNA localization process is one example[1,6–12].

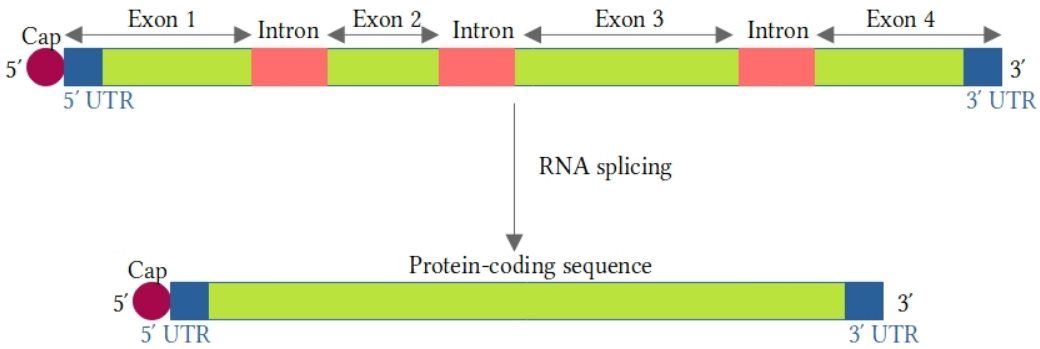


Figure: 1.1: Splicing representation, upon splicing introns will be removed from the pre-messenger RNA, to form mature RNA.

1.2 mRNA localisation

The mechanism of mRNA localization has exquisite site-specific control over the regulation of the gene expression. A large proportion of the mRNAs are transported to the specific subcellular cytoplasmic regions before gene expression. In Drosophila melanogaster, the embryogenesis shows out of the 3370 genes monitored, 71% of the

expressed are subcellular localized mRNAs[13–15]. Rather than proteins, transporting mRNAs have consequential benefits for cellular functions. Such as, it is very much effective in preventing the proteins from behaving ectopically, anywhere else out of their cytoplasmic site of action. Bio-genesis of multiple proteins can be possible with the single mRNA localization and also it facilitates the co-translation of other protein subunits, altogether construct into functional macromolecular complexes. mRNA targeting manages the functional activation of the RNA concerning time and space by transporting splice variants to different cellular regions, with much other wide range of functions mRNA localization holds a vital role in the biological systems[16–20].

The mechanism of mRNA localization has been studied widely and showed many different modes of localization patterns and localization elements that are involved during the process. The earliest results indicate that mRNAs localized within the cell are from in-situ hybridization, out of those distinct mRNAs have shown very specific patterns of localization, these results are from asymmetric cells like egg cells or fibroblasts. The studies have also indicated that cis-acting RNA elements are necessary for localization. The localization elements are most probably found in the 3' UTR and can be of varying in lengths from 5 or 6 to several hundred nucleotides, with repetitive nucleotide sequences for better binding. Cis-elements along with trans-acting factors altogether assist the process of the mechanism of localization. Trans-acting factors are the Ribonucleic Binding Proteins (RBPs) and can also be the small regulatory RNAs to form macromolecular complexes. mRNA to be localized embodies cis-elements that can accommodate zip code proteins that facilitate the localization processes[13,21–25]. Additionally, mRNA also recruits proteins that repress the translation, so the translation only begins when the mRNA is localized to the particular site. So, the ectopic translation can be prevented. mRNA localization is advantageous as the localized translation is an efficient way to localize a particular protein[26–29].

mRNA localization can take place through different mechanisms. One of the mechanisms is through cytoskeleton filaments. mRNA with the Exon Junction Complex (EJC) and other microtubule and actin components are translocated to the specific site of the cell. mRNA localization through active transport involves the specific recognition with the help of a core ribonucleoprotein complex and through cytoskeletal motor proteins. Altogether securing the mRNAs to the specific site of the cell. Pre-mRNA upon splicing becomes mature mRNA, that could incorporate the Exon Junction Complex to facilitate the process of mRNA localization. This is significant in many mRNA localization including *Oskar mRNA*.

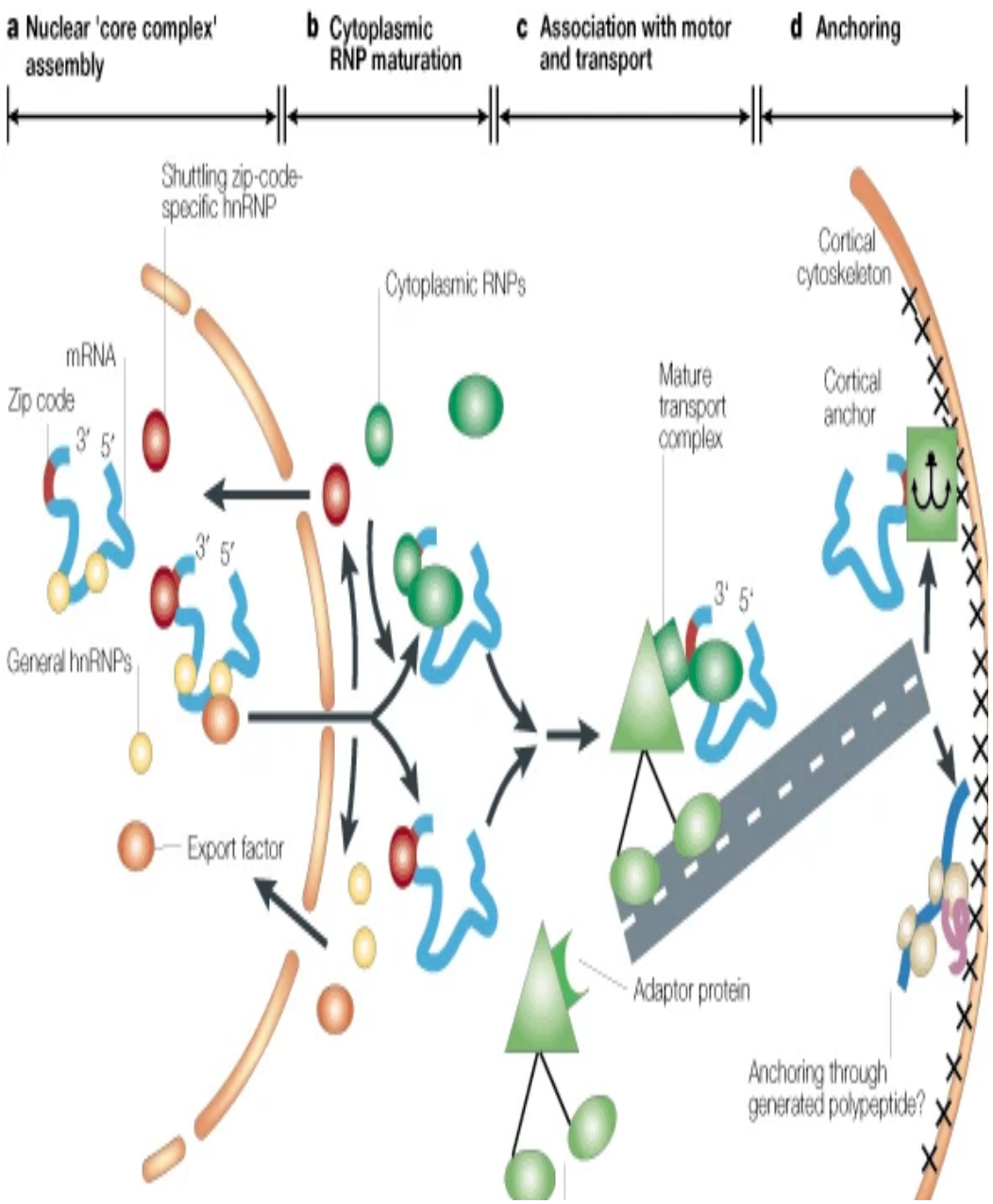


Figure:1.2-mRNA localisation mechanism [ansen, RP. mRNA localization: message on the move. *Nat Rev Mol Cell Biol* 2, 247–256 (2001)] a| Core complex assembly. Heterologous nuclear ribonucleoproteins (hnRNPs) bind to the transcript (blue). Apart from general hnRNPs (yellow circles), there are specific hnRNPs (red ovals) that recognize the mRNA's localization signal (zip code, red). In the following step, localized (and non-localized) mRNAs assemble with proteins that are involved in mRNA export (orange oval) and the mRNA–RNP complex is exported to the cytoplasm. b| Cytoplasmic maturation. General hnRNPs and export factors shuttle back to the nucleus, whereas specific hnRNPs stay associated with the mRNA (lower part). Alternatively, detach from the transcript and replaced by cytoplasmic zip-code-specific RNPs (green ovals, upper part). c| Transport. The mature RNP complex along with a motor protein (green triangle), probably with adaptor proteins, and is transported to the target site. d| Anchoring. Finally, the RNP is released from the motor.

1.3 *Oskar* mRNA

In *Drosophila melanogaster,* the localization of the *Oskar* mRNA towards the posterior pole of the oocyte marks the formation of the germ cells and abdomen of the future embryo during the process of embryogenesis. *Oskar* is that gene, which is essential for the differentiation event of the *Drosophila* embryo. Most of the posterior side of the oocyte is defined by *Oskar* mRNA localization. The O*skar* gene guides protoplasm assembly with which it also controls the germ cell precursors formed quantitatively at the posterior pole of the *Drosophila* embryo. Mislocalization of O*skar* RNA to the anterior pole ends up in the induction of germ cells at the anterior side. Eight genes are requisite for germ cell formation at the posterior pole, which are *oskar, vasa, tudor, nanos, cappucino, spire, staufen and valois.* Out of eight genes, only three genes *oskar*, *vasa,* and *tudor* are vital ectopically[30].

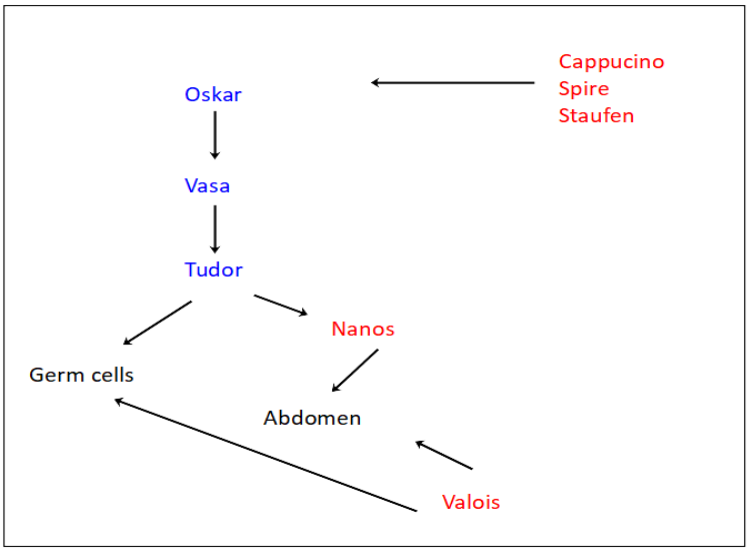


Figure: 1.3 - Pathway for germcell and abdomen format- -ion. Genes which are written in blue color are vital for the ectopic site, while *vasa* and *tudor genes* functions downwards to *oskar* gene. Genes written with in red color are genes responsible for the pole plasm formation. Image is modified from, *Nature* **358**, 387–392 (1992).

Oskar is answerable for assembling the germplasm, which is required for germ cell formation. The plasm involves polar granules, constituted by *Oskar* mRNA, mitochondrial coded ribosomal RNA, and polysomes and proteins like Oskar, Staufen, and Tudor (but they are void of mRNAs like Vasa, Staufen, or Tudor)[31]. Oskar protein function is especially restricted to the posterior pole of the embryo. During embryogenesis, embryos which lack *oskar* would fail form the abdomen and germline. Just in case of ectopic expression of the protein Oskar at the anterior region would induce the germ cells and abdominal structures within in the place of head[30–32].

Oskar mRNA is transcribed within the nurse cells during the first oogenesis, later during the mid oogenesis *Oskar* mRNA starts to localize towards the posterior pole through anterior margin[27,33]. From the mid oogenesis, the Oskar protein translation commences, by the end of oogenesis both *Oskar* mRNA and Oskar protein accumulated at the posterior pole of the oocyte until the first stage of embryogenesis[34–36]. This late phase accumulation is vital for the patterning of the embryo[37,38].

In *Drosophila,* mRNA localized through microtubule cytoskeleton and associated motor proteins[39]. The transport occurs along the polarized cytoskeleton. Trans-acting factors recognize the precise sequence within the *Oskar* mRNA transcript and form ribonucleoprotein particles that interact with the kinesin motor for transport along the microtubels[40]. *Oskar* mRNA localization towards the posterior pole requires non-coding sequence elements that regulate the localization mechanism are the 3′-untranslated region (UTR) and first intron splicing[41]. *Oskar* 3′UTR has an oocyte entry signal, which is significant for the mRNA transport into the oocyte. The secondary structural features and an AU rich nucleotide composition play a very important role in *Oskar* mRNA transport function[42]. EJC along with *Oskar* mRNA co-localizes to the posterior pole of the oocyte. For *Oskar* mRNA, EJC deposition event upon first intron splicing is the crucial occurrence for localized transportation. Upon splicing of the initial intron creates a brief stem-loop RNA structure, named as 'Spliced *Oskar* Localization Element' (SOLE)[41].

The SOLE RNA consists of 18 nucleotides from exon 1 and 10 nucleotides from exon 2, ligated together at the primary exon junction complex site upon intron splicing. *In vivo* mutational analysis established the relevance of the short proximal stem (PS, consists of 6 base pairs) for localization, suggesting that this structural element participates within the recognition of trans-acting factors (figure:1.3)[43]. Nucleotides 524-539 from *Oskar* mRNA were predicted to fold into a medial stem-loop element (MSL). Mutational analysis, designed on the already predicted structure of the MSL region appears to be non-essential for the function. However, this region can form an alternative secondary structure to the MSL. The SOLE RNA alone isn't sufficient for the localization event. Without the event of splicing mRNA tends to be mislocalized. On the other side mRNA bound to EJC, but without SOLE sequence is additionally mislocalized. These results strongly explain the role of SOLE sequence and EJC functions to facilitate the localization mechanism.

1.4. Exon Junction Complex

Exon-junction complex(EJC) is the protein complex deposited on a pre-messenger RNA at the junction of the 2 exons that have been ligated together during the RNA splicing[44]. In eukaryotic cells, the protein complex of EJC is made 20-24 nucleotides upstream at the 5' end of the spliced junction, during the second step of splicing[45,46]. The binding of EJC on the pre-messenger RNA is sequence-independent and form mature messenger ribonucleoprotein (mRNP)[47]. The EJC together with the bound mRNP then exported out of the nucleus to the cytoplasm, where it will remain until the initial stage of translation starts, and then it will be recycled back to the nucleus.
The EJC is deposited on mRNA 24 nucleotides upstream of spliced junctions after intron splicing. It is together with the bound mRNAs localized from nucleus to cytoplasm, where it will be within the bound form until the initial round of translation, and after recycled back to the nucleus.

The core of the EJC contains mainly four proteins. They are DEAD-box RNA helicase eIF4A3, MLN51, and Y14-Mago hetero-dimer. Structural studies revealed that DEAD-box RNA helicase eIF4A3 protein functions as a clamp that binds to the RNA in an exceedingly sequence-unspecific manner. Where because the Y14-Mago hetero-dimer locks the eIF4A3 onto the mRNA, while MLN15 provides the soundness for the complex[48–52]. The EJC has several regulating functions such as post-transcriptional processes. Including splicing, cellular localization, and Non-sense Mediated Decay (NMD). In *Drosophila,* mRNA localized through microtubule cytoskeleton and associated motor proteins[39]. EJC functions as s stable and sequence-independent complex that may bind to the mRNA, and can remain bound until the beginning of the translation process. EJC acting as a molecular shepherd holds and travels with mRNA throughout the localization process across the cell.

In Figure 1.4, the components, assembly, and functions of Exon Junction Complex deposited mRNA has been depicted. A pre-mRNA undergoes splicing, to get mature mRNA with the help of Spliceosome. Then, the exon junction complex is deposited upstream of 24 nucleotides nearby the exon junction point of the intron splicing. The proteins Y14, Magoh, eIF4AIII, and MLN51 are the core components f EJC. Further, EJC deposited mRNA involves itself in various functions such as localization, translation, etc.,

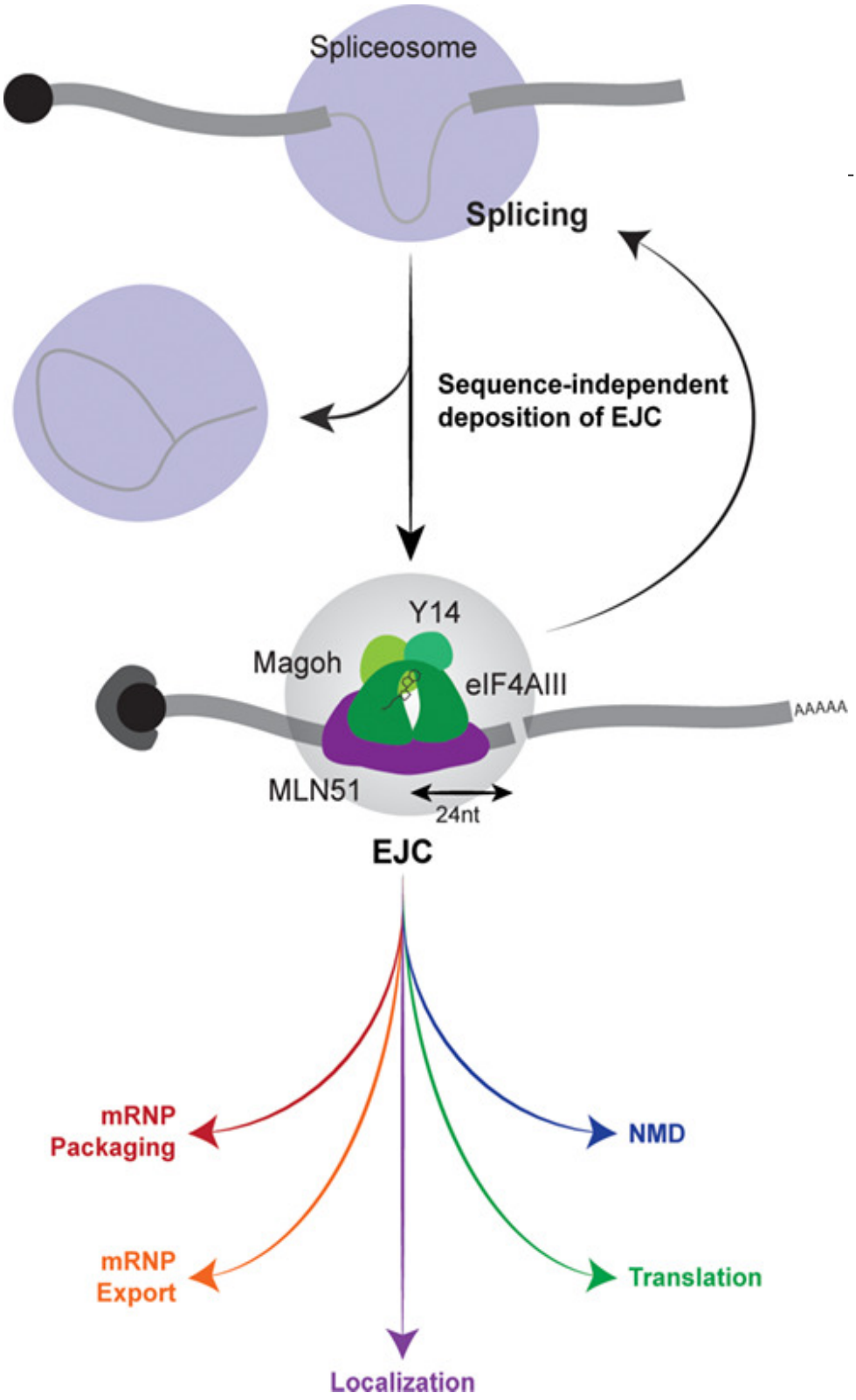


Figure 1.4: Assembly, structure, and function of the Exon Junction Complex (EJC) (Image is taken from the paper, Lauren et. al., 23 December 2016, Wires RNA)

1.5. Intrinsically Disordered Proteins (IDPs)

Intrinsically disordered proteins are also called unstructured proteins. They do not have any rigid 3-dimensional structure except for very few secondary structural elements. Because a large part of the protein is unfolded, it is very challenging to study these molecules' structural characteristics. Intrinsically disordered proteins or protein regions (IDPs/IDPRs) are characterized by exceptional conformational flexibility and structural plasticity, are very unusual when compared to well-folded proteins with functionality and folding. Hence, they are able to perform biological activities that are unlikely for the ordered proteins. The conformational flexibility of the proteins facilitates the possibility of interaction with the variants. Typically in the absence of the macro-molecular binding partner such as other proteins or RNAs, the IDPs remain to be unstructured[53–55].

Structural characterization for IDPs is very challenging, as they are structurally unfolded. Generally, structural characterization of structured proteins can be done through various biophysical analytical tools like the X-ray diffraction method, Crystallography, and NMR. But, for the intrinsically disordered proteins, only NMR is applicable as they fail to form crystals[56–58]. Nuclear Magnetic Resonance spectroscopy is well-suited to study the dynamics of IDPs, as well as basic structural characterization. NMR experiments like paramagnetic relaxation enhancements (PREs) or residual dipolar couplings (RDCs) can be adapted to analyze the intrinsically disordered regions, to get the distance restraints data[59–61].

Intrinsically disordered proteins exhibit the mechanism of ‘induced folding’. That means unfolded proteins will transform to folded structure, from fully folded to the partially folded state. But, this does not true for all the IDPs, in some, they remain structurally disordered in the bound state as well. This phenomenon is referred to as ‘fuzziness’. Thus, there is a range of possibilities from disorder-to-(complete)order to disorder-to-disorder transition. But, in every scenario, the binding will result in a change of structure and dynamics of the bound ground state[62].

Figure 1.5 reports two cases of bound IDPs. The complex between KIX – pKID[63,64] and the complex between the proteins GCN14 – Med15[65,66]. The protein pKID is an IDP, thorough its very specific interactions with the protein KIX undergoes the process of induced folding. Another example is GCN14-Med15, which is a large heterogeneous complex. The interaction is specific and through hydrophobic regions. Here, the

complex is fuzzy due to its dynamic state contributed by the structural disordered nature.

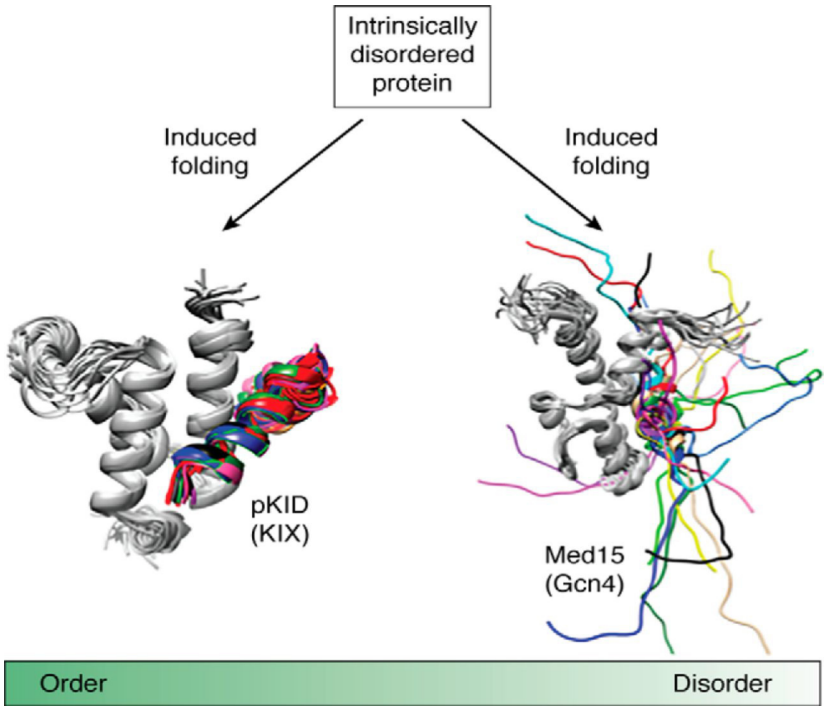


Figure 1.5: Representation of the pKID – KIX protein complex and GCN14 – Med15 protein complex, the structure is induced.(This image from the Journal JBC, volume 293issue 19, May2020 P6586-6593)

1.6. Protein Pym

Protein Pym is a cytosolic multi-functional protein that acts as a regulator protein of the exon junction complex[67], interacts with the translation machinery post mRNA localisation[64,] and is also involved in the Non-sense Mediated decay (NMD)[67]. Hence, the protein Pym is capable of directing the post – transcription processes[69,70].

Pym regulates the binding and dissociation activity of the exon junction complex. EJC binds to the mRNA during post-splicing events during the process of localization.

While Pym dissociates the EJC from the spliced mRNAs, on the other way EJC assembly intermediates are resistant to Pym. That means the protein Pym does not inhibit the deposition of exon junction complex onto the spliced mRNAs, but over-expression of the protein in the cells disrupts the bound EJC from spliced mature mRNAs and thus inhibits nonsense–mediated decay. Pym protein is the important molecule for the EJC dissociation. So, the recycling ceased in the cells with a reduced concentration of protein Pym, resulting in the accumulation of EJCs on spliced mRNAs. Even though this phenomenon is similar with respect to other EJC associated

mRNPs, such as *bicoid*, *gurken,* and *nanos* mRNAs, but the mechanism of localization may not always be interrupted by the dissociating EJC with ectopic action of Pym[69]. Because, in *oskar* and *nanos* mRNPs, the EJC plays a crucial role in the mechanism of localisation[41,43], while in *bicoid* and *gurken* mRNPs it is not[69]. Hence, protein Pym is indirectly involved in the process of localization, where the association of exon junction complex to the mRNAs plays an important role in the mechanism of localization.

1.6.1. Interaction of Pym with Y14-Mago heterodimer

The Mago-Y14 heterodimer is the core component of the exon junction complex. Pym is the cytoplasmic RNA binding protein. Pym interacts with Mago-Y14 dimer through its N-terminal region. In the crystal structure of the ternary complex, one can notice that Pym binds to Mago and Y14 simultaneously capping their hetero-dimerization interface at conserved surface residues. The interacting residues of Mago to the Pym have been implicated in the nonsense-mediated mRNA decay. Pym also binds directly to the RNA despite not showing any homology with the RNA binding proteins[67]. Gel shift assays showed that the Pym-RNA complex can be super-shifted on the addition of the Mago-Y14 hetero-dimer. These results confirm that Pym can simultaneously bind to RNA and the Mago-Y14 complex.

Full-length *Drosophila melanogaster* (Dm) Pym shows direct interaction with Mago-Y14 hetero-dimer[67]. Protein Mago is found to be a single structural unit, whereas the protein Y14 folds into three distinct structural domains, which are the N-terminal domain, the RNA binding-like domain (RBD), and a C-terminal low complexity region. In the image below, the interaction pattern of the N-terminal region of Pym can be seen. Pym binds at the edge of Y14 β-sheets and at the edge of Mago α-helices. Mago-Y14 heterodimer forms a rigid scaffold for Pym binding. The structure of the Mago-Y14 dimer is very much similar to the Pym-Mago-Y14 ternary complex. Mago has an anti-parallel β-sheet flanked on one side of two long and one short α-helices. The α-helical surface of the Mago interacts with the β-sheet surface of the Y14 RBD (RNA Binding like Domain). The amino acid residues of the Mago-Y14 complex superimpose with an overall RMSD of less than 1.2A°. The N-terminal region of Pym binds to the globular β-domain of both Mago and Y14, capping their heterodimerization interface. The structure of the complex Y14-Mago-Pym is very similar to the previously reported structure of Y14-Mago complex, which is devoid of Pym protein[71–73].

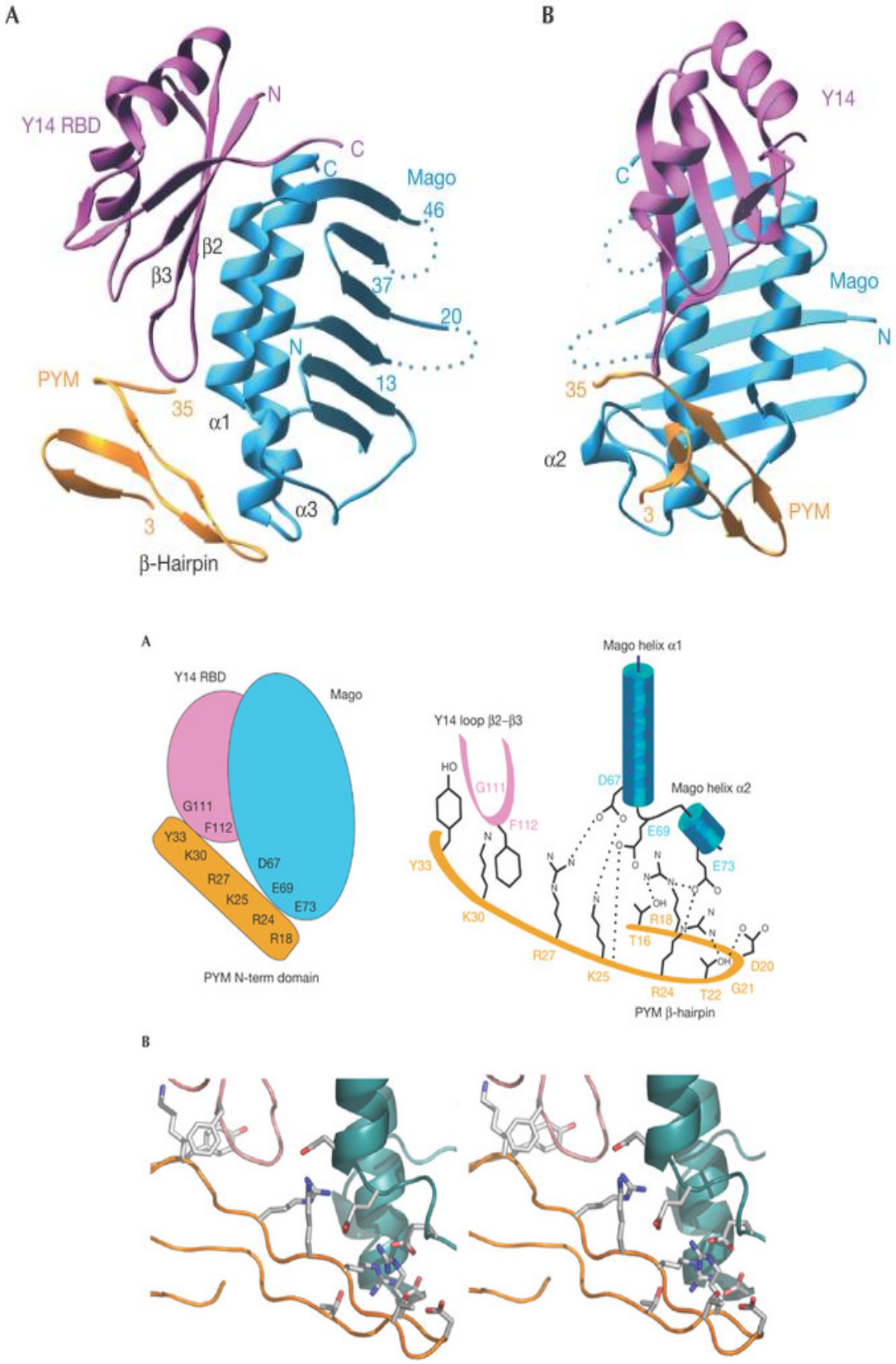


Figure 1.6: Pym-Mago-Y14 complex, **A** and **B** are different orientation of the complex. In the below image, A) is spatial representation, B) Stereo view.

The major difference is observed in the Mago protein at the α2- helix and at the 14-19 loop, where it is disordered in the crystal structure of the Y14-Mago-Pym complex. This data suggests the rigid scaffold of the Y14-Mago heterodimer for the interaction with the protein Pym.

Whereas from the protein Pym, the N-terminal domain (3-35) folds with a three-stranded β-sheet and a contiguous β-hairpin. However, the C-terminal residues are structurally disordered and do not contribute directly to the interaction with the Y14-Mago heterodimer. Pym binds with electrostatic interactions at the α-helices of Mago and with the hydrophobic interactions at the β2- β3 loop of Y14. Also, the solvent molecules enhance the interaction of Pym with the heterodimer, as at the interface at least 40 water molecules are found. Pym docks with its positively charged residues (R_{18}, R_{24}, K_{25} and R_{27}) to the acidic surface of the Mago α-helices (N_{67}, Q_{69}, N_{116}). Particularly, the interactions are contributed by the β-hairpin portion of Pym.

The extended stretch of the N-terminal domain of Pym wraps around β2- β3 loop of Y14. The rigidly conserved residues of Y14 are from the β2- β3 loop, which has a group of invariant residues for heterodimerization with the protein Mago and another group for Pym (K_{30}, Y_{33}) with the hydrophobic interactions.

1.6.2. Pym is ribosome -bound EJC disassembly factor

Pym is the dissociation factor of the exon-junction complex from the mature mRNA, post localization. However, the action of dissociation of EJC is also associated with interaction with the translation machinery. Pym interacts with exon-junction complex through N-terminal motif, while interacting with the ribosome units through C-terminal motif simultaneously. From all these data we can understand the functional significance of protein Pym. Pym embodies important binding motifs which can bind to RNA, Mago-Y14 complex and also interacts with ribosomal units.

In the image, Pym (represented as yellow) interlinks via its C-terminus (indicated by C) with the small ribosomal subunit and removes EJC during the process of ribosomal transit through interaction of its N-terminus (N) with Magoh-Y14 hetero-dimer. Cytosolic Pym minimizes by its concentration by its ribosomal association. Because, free Pym can dismantle EJC, independent of translation. Recycled EJC is transported back to the nucleus.

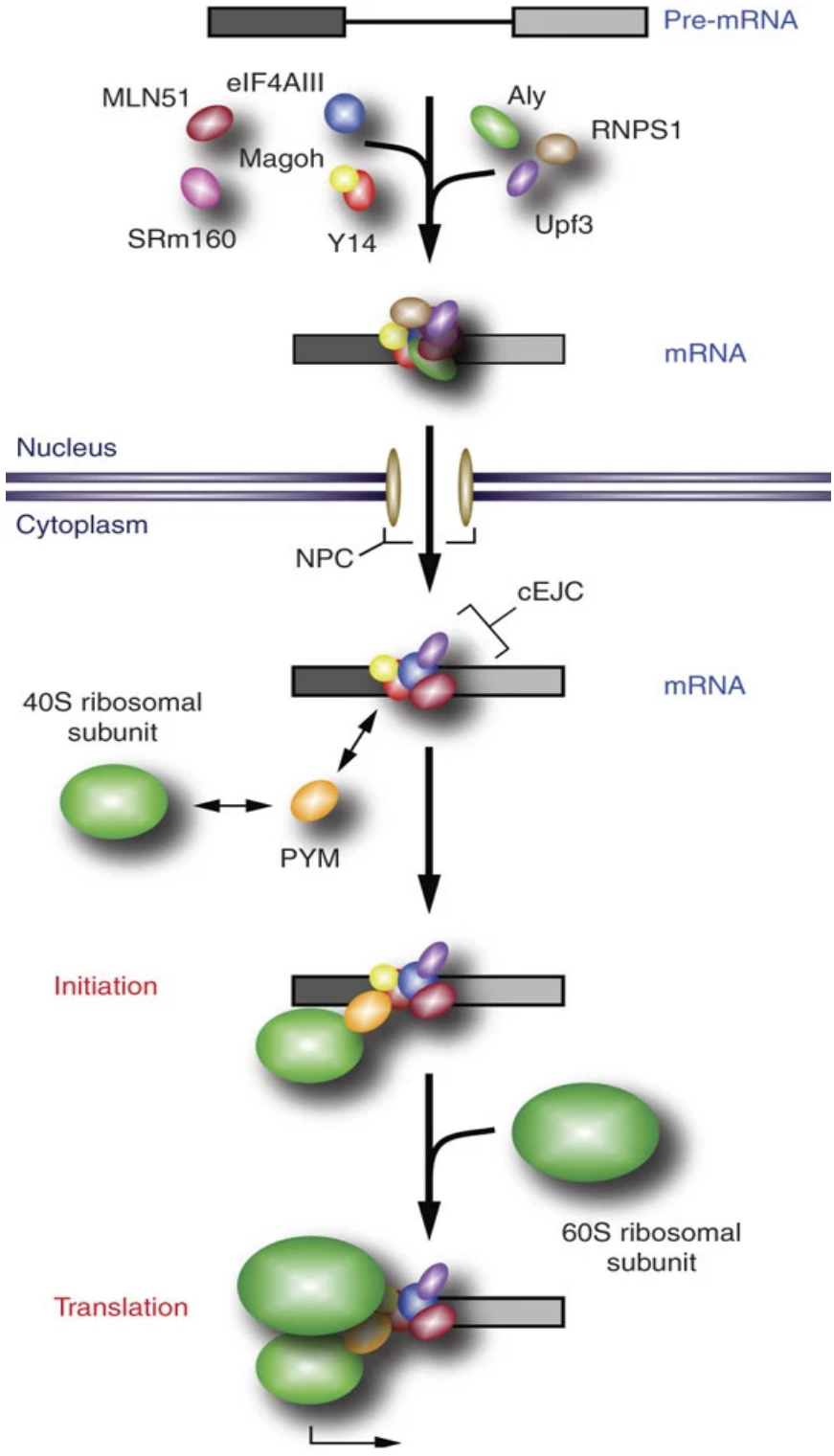


Figure 1.7: Representing the protein Pym, Exon junction complex, mRNA localisation and beginning of the translation process.

1.6.3. Pym over-expression disrupts *Oskar* localization

The mRNA processing such as splicing unlocks the events like mRNA export and localization translation enhancement[74,75].

In *Drosophila melanogaster*, over-expression of the Pym protein disrupts the *Oskar* mRNA localisation[69]. In this study, the protein Pym has been divided into N-terminal (N), middle (M) and C-terminal (C) domains. A set of eGFP-tagged Pym deletion transgenes were generated. Upon the expression on the female germline, the GFP signal in the Pym-GFP egg-chambers was distributed uniformly throughout the cytoplasm of nurse cells. However, in the N, M and C construct the signal was also observed in the nurse cells. That is because of the disrupted localization of the *Oskar*. This was monitored by observing the distribution of *oskar* mRNP component Staufen protein and Oskar protein by immunostaining. Because, Staufen and Oskar protein are only stable at the posterior pole of the oocyte. As, it can be seen from the image that Staufen was not able to enrich the posterior of ΔC- and N-Pym oocytes describe the failure of *Oskar* mRNA localization.

1.6.4. Spliced *Oskar* Localized Element (SOLE)

Oskar mRNA localization is evolutionary conserved, that results in the spatial and temporal restriction of protein synthesis to the specific parts within the cells[76]. Generally, the mRNA localization occurs through motor proteins and cytoskeleton. In *Drosophila melanogaster*, *Oskar* localization to the posterior pole of the oocyte takes place through the microtubules and kinesin heavy chain[77–79]. Along with this, the localization involves the complex but specific molecular arrangement of exon junction complex proteins like Y14, Mago, eIF4AIII and Barentsz proteins[70,80–83]. Consistent with this, the splicing at the position of the first intron is vital for the posterior localization of the *Oskar* RNA[41]. Splicing is followed by the formation of SOLE (the first intron spliced RNA) and the EJC deposition, both are vital for the localization.

The properly positioned EJC and *Oskar* 3′ UTR are the important components for the localization[43]. To prove this, the *oskar* coding region in *oskΔi(2,3)* of the mRNA is replaced by a *lacZ* coding sequence of identical length (*Lz* transgene) (Fig: 1.8 and 1.9), such that *oskar* intron 1 region was maintained, and both are observed through fluorescence. During the early stages of oogenesis, both the *Lz* mRNA and *oskΔi(2,3)* mRNA (fig) were enriched. However, from the mid-oogenesis onwards the distribution of both RNAs in the oocyte differed considerably. *OskΔi(2,3)* mRNA was transiently enriched at the center of the oocyte, then localized at the posterior pole (figure 1.8 and

1.9), whereas *Lz* mRNA remained diffusely distributed throughout the oocyte. Further in the later stage of oogenesis it is confirmed through qualitative analysis that there is no sharp increase of the *Lz* mRNA at the posterior pole, as it was for *oskΔi(2,3)* mRNA (fig). The failure of *Lz* mRNA to localize is may be because of the inability of Lz mRNA to assemble into a localization complex mRNP. In the figure the evident localization of bicoid and gurken towards anterior and anterodorsal positions is indicating the proper formation of polarized microtubule network[84,85]. These data indicate that the vital components of *Oskar* mRNA localization reside within the coding sequence, that is 3′ UTR of *oskar*.

Exon I and exon II are crucial for mRNA localization. In order to determine this, different regions were deleted and substituted in the *oskΔi(2,3)* transgene, and further evaluated the distribution of the transgenic mRNAs during oogenesis. To start with exon-1 was subdivided into regions I and II, deletion or replacement of region I with the *lacz* sequence (*oskΔI or oskΔ*^{Lz} transgenes), did not affect the posterior enrichment of the mRNA, whereas the mRNA produced from the transgene which was lacking region II, was mislocalized in the oogenesis, with little posterior enrichment (figure 1.9). The EJC disposition site in region II, 20-24 nucleotide upstream of the splice junction.

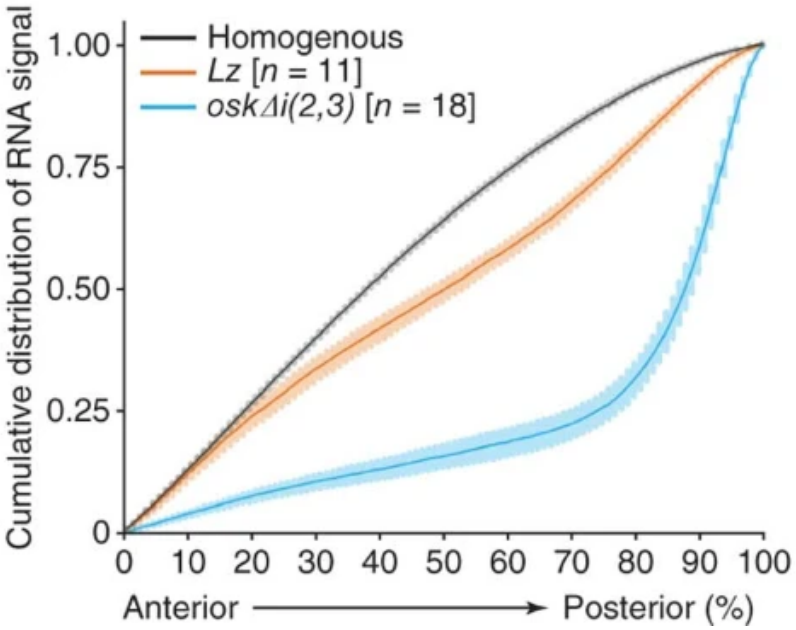


Figure1.8: Cumulative distribution of *oskΔi(2,3)* and *Lz* reporter mRNAs in stage 9 oocytes
This image is from *Nature Structural & Molecular Biology* 19,441–449 (2012))

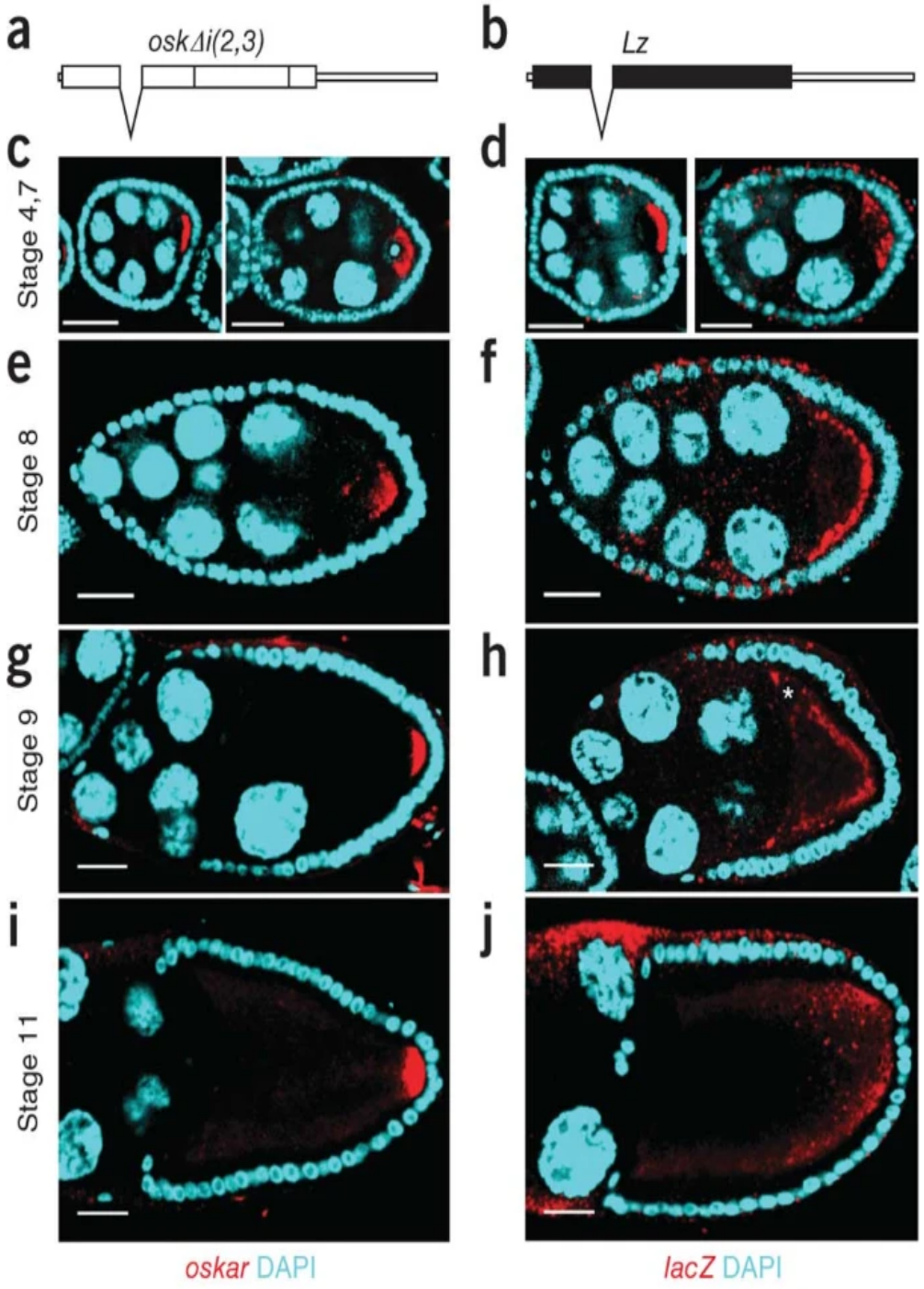


Figure1.9: Localization of Oskar mRNA, with mutated transgene. *LacZ* mutated transgene failed to localize

(This image is from 19,*Nature Structural & Molecular Biology*441–449 (2012))

So, Exon junction complex deposition and *Oskar mRNA* localization involve exon I and exon II together. This is the major and necessary component for the process of localization. The ligated element of exon I and exon II is hence called Spliced Oskar Localized Element (SOLE) RNA.

Our work includes the study of SOLE RNA, structural and binding studies, and the study of protein Pym both structural and functional aspects.

CHAPTER

2.Methodological Background

2.1 Nuclear Magnetic Resonance (NMR)

NMR is very unique in understanding the biological roles by studying the structure and dynamics of biological roles, the biological mechanism of the protein. One can also probe the biological aggregations, protein aggregations, protein-RNA interactions[86–91]. In general proteins have 30% of the genome their segments are unstructured which means the intrinsically disordered regions, point of flexibility. One can analyze the dynamic property of the protein, which is vital as the property is closely related to the biological process. NMR can cover a lot of ground with respect to biological mechanisms by solving the structure of individual molecules involved, to understand the dynamic property to the nuclear level, interactions between the molecules[92–97].
Nuclear Magnetic Resonance is one of the most important and fruitful analytical tools, especially in the field of Structural biology. Structural biology is concerned with the study of mechanistic systems as well as molecular structures of the biological macromolecules such as proteins, nucleic acids, membrane proteins[98–102]. In this regard, NMR can contribute structural particulars and can also monitor inter and intra bio-molecular interactions within the system[103–107].

The theoretical specifics of NMR briefed in this chapter be from J. Keeler. Understanding NMR Spectroscopy, 2nd edition. 2010, M. H. Levitt. Spin Dynamics: NMR Basics, 2nd Edition. 2008, I. Bertini. NMR of Biomolecules, 2012, T. D. W. Claridge. High-Resolution NMR Techniques in Organic Chemistry, vol 19. 1st edition.

2.2 Origin of NMR signal

NMR signal originates from the nuclei in a strong constant magnetic field is perturbed by a weak oscillating magnetic field. The essence of nuclear magnetism is the total angular momentum of the nucleus, defined by the *nuclear spin*. The property of the nuclear spin is fundamental for nuclear magnetic resonance spectroscopy. All the nuclei with non-zero spin quantum numbers (I) are NMR active. The spin quantum number is quantized, as integrals or half-integral multiples of (h/2p), h= Planck's

constant. Some of the useful atomic nuclei which have non-zero spin quantum numbers are ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P which have I=1/2. The naturally abundant nucleus of interest ^{12}C, has zero spin, hence NMR inactive. The total angular momentum (P) of the nuclei is associated with magnetic moment(μ).

$$\mu = \gamma P$$

Where,ɣ- Magnetogyric (gyromagnetic) ratio, constant for given nucleus.
For proton it is 2.67×108 rad·s−1·T−1.
Both angular momentum and magnetic momentum are vector quantities. In the presence of an external static magnetic field (B_0), the magnetic moments align themselves relative to the magnetic field[108–111]. The orientations are discrete, depending upon the energy states involved. Every spin with the spin quantum number I, has (2I+1) energy states or spin states prevail for the spin. Magnetic quantum number m, an integer, can be from -I to +I. Accordingly, for the nuclei with I=1/2, have 2 spin states +1/2 and −1/2 and have parallel (α, lower energy state) and anti-parallel (β, higher energy state) orientations in the applied magnetic field respectively. So, at equilibrium slightly excess of nuclei in the lower energy state. According to Boltzmann distribution[86,112,113]

$$\frac{N_\alpha}{N_\beta} = e^{\Delta E / k_B T}$$

Where Nα,β speak for the number of nuclei in the spin orientation, kB is Boltzmann constant and T is the temperature. A very slight excess of population difference of nuclei in the different spin states (1 part in 104) culpable for the NMR signal[114–117]. Hence, NMR is very insensitive relative to IR and UV spectroscopy.

This slight excess population of nuclear spins can be represented as a collection
of spins distributed randomly about the precessional cone. This slight excess of the nuclear spin population is called as Bulk magnetization. The whole NMR signal is situated on the bulk magnetization. So, various techniques are utilized in order to intensify the bulk magnetization. One of them is to increase the static magnetic field of the NMR spectrometers.

2.3 Excitation pulse and Free Induction Decay (FID)

In the static magnetic field (B_0), when the pulse of the oscillating electromagnetic field is applied transversely, which is the Larmor frequency(γB_0) of the corresponding nucleus, the bulk magnetization shifts to the traverse plane and start to precess for a duration of T_{pulse}. The excitation pulse, $T_{90°}=\pi/(2\gamma B1)$ shifts the bulk magnetization to the transverse plane, where, this induced current will be detected as NMR signals. But, after the time T_{pulse}, the magnetization starts to go back to the longitudinal axis, which is called free induction decay[118–121] (FID). The time required for the bulk magnetization to get back to longitudinal axis is termed as relaxation time. The more the relaxation time, the sharper NMR signal and shorter relaxation time result in a broad NMR signal (peak).

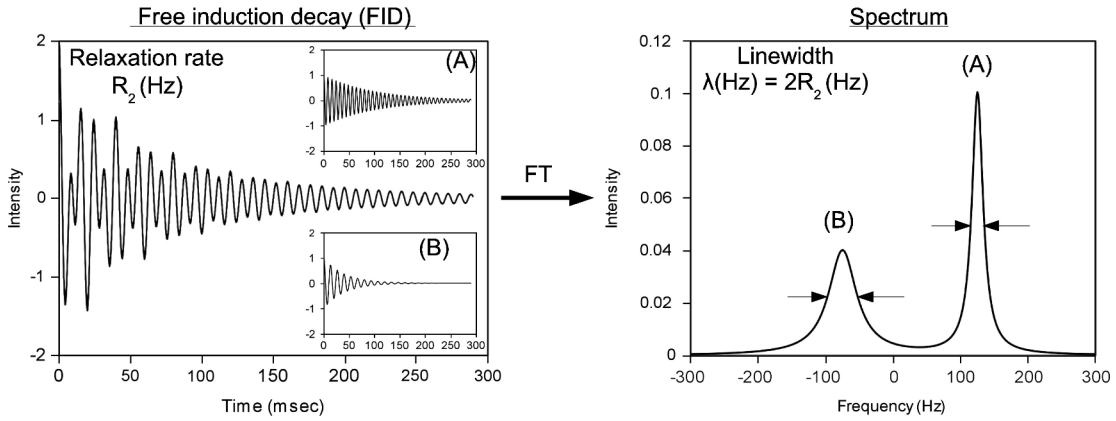


Figure 2.1: The free induction decay(FID)[122], which is fourier transformed to visualize the nmr spectrum. (A) is the FID for with competitively linger relaxation time than that of FID (B).

2.4 Relaxation

In the presence of the external magnetic field (B_0), the nuclear spins experience small torque which results in a net magnetization along B_0 direction[123,124]. This is how the bulk magnetization attains its equilibrium state. At equilibrium state, there is the presence of z-magnetization along the direction of the magnetic field, but there will be no transverse magnetization[125–128].

During the NMR experiments, when an RF pulse of a certain strength is applied to the magnetization at its equilibrium state the z-magnetization starts to rotate towards the transverse plane producing transverse magnetization at the end of the pulse, inducing a current which we detect as NMR signals in the form of FID. Once the RF pulse is switched off the transverse magnetization goes back to its original equilibrium state during the course of time. Thus, the recovery mechanism of the perturbed equilibrium distribution back to the Boltzmann distribution state is nothing but the relaxation[129–136].

2.5 Chemical Shift

Detection of the NMR signal, in terms of observance of the nuclear magnetic frequencies, is often termed as chemical shift. Chemical shift is dependent on the chemical and electronic environment of the nuclei[137–139]. Each and every nucleus in the molecule experiences the partial shielding of the external magnetic field (B_0) due to the electron density of the neighboring nuclei. Electrons revolving around the nucleus produce a small magnetic field that opposes B_0. So, the net magnetic field experienced by every nucleus is slightly different from each other depending upon the neighboring atoms. So, this is how the chemical shit of the nuclei differs from each other. The chemical shifts of strongly shielded nuclei will be up-field in the NMR spectrum, whereas the deshielded nuclei will have the chemical shift in the downfield regions.

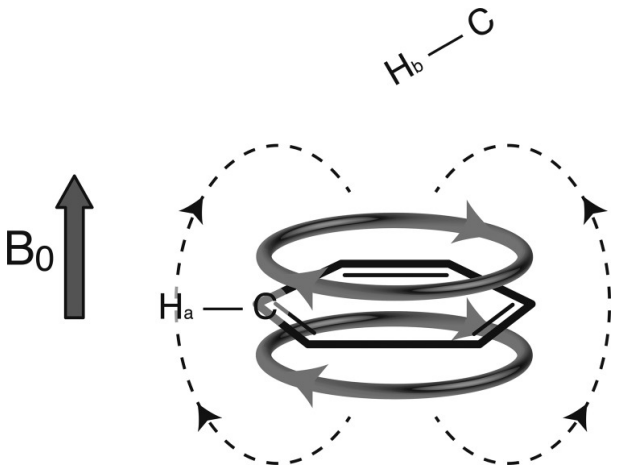


Figure 2.2: The magnetic field at the nucleus is generally different from the applied field B0: this additional contribution (or screening) arises from the interaction of the surrounding electrons with the applied field[103].

2.6 Spin – spin coupling

The magnetic interaction between the neighboring individual nuclear spins is transmitted through the bonding electrons. This is also called J-coupling. Coupling can be classified into two categories. Spin-spin coupling (through bond) and dipole-dipole coupling (indirect coupling). Spin coupling takes place between any magnetically active nuclei, which are close by either through a bond or through the space[140–143]. Coupling arises from the hyperfine interactions between the nuclei and local electrons present. J-coupling embodies the very important information about the relative bond distances and angles, also describes the connectivity of chemical bonds. The magnitude of the J-

coupling depends on the number of bonds present in between the nuclei. The closer the nuclei are, the stronger the chemical bond is. In the case of hetero-nuclear coupling, the coupling constant depends on the magnetic moment of the nuclei involved. The higher the magnetic moment, the higher is the coupling constant. J-coupling also gives information regarding the dihedral angles relating to coupling partners.

In general words, each nucleus observes neighboring magnetically active nuclei as tiny magnets, hence perturbed magnetic field around. So, the effect is observed as the splitting of NMR signals. So, an NMR spectrum, with all the coupling information may be overly crowded also. This may result in the severe overlapping of the NMR signals, which makes the spectrum difficult to analyze. Hence, the partial decoupling or complete decoupling NMR experiments are recorded, unless it is absolutely required. For, smaller molecules experiment with the coupling information may be beneficial, but in the case of larger bio-molecules or bio-polymers like proteins and nucleic acids, the NMR experiments are recorded by decoupling methods[144–148]. Decoupling is done by continuously irradiating the broad range frequency to the nuclei, either homonuclear or hetero-nuclear. This results in the fast flipping of the nuclei between the magnetic energy states. So, this results in the average magnetic unperturbed state, which is observed as a uniform field. Thus, the NMR signal will be without any splitting, which adds to the higher intensity of the signals[149–156].

2.7 Bio-molecular NMR

Biomolecular structure calculation is an essential requirement to address the biological problem and to understand the biological mechanism. For proteins, structural information can be derived from the different multi-dimensional NMR spectra recorded from isotopically (^{15}N, ^{13}C) labeled samples, information about sequential linking can be achieved by triple resonance backbone and side-chain assignment experiments, short distances between nuclei through space can be determined by 3D-NOESYs, orientation and longer-distance data can be obtained by the experiments like RDC (Residual Dipolar Coupling), PRE (Paramagnetic Relaxation Experiment) and relaxation experiments like diffusion spectroscopy.

2.8 Assignment theory

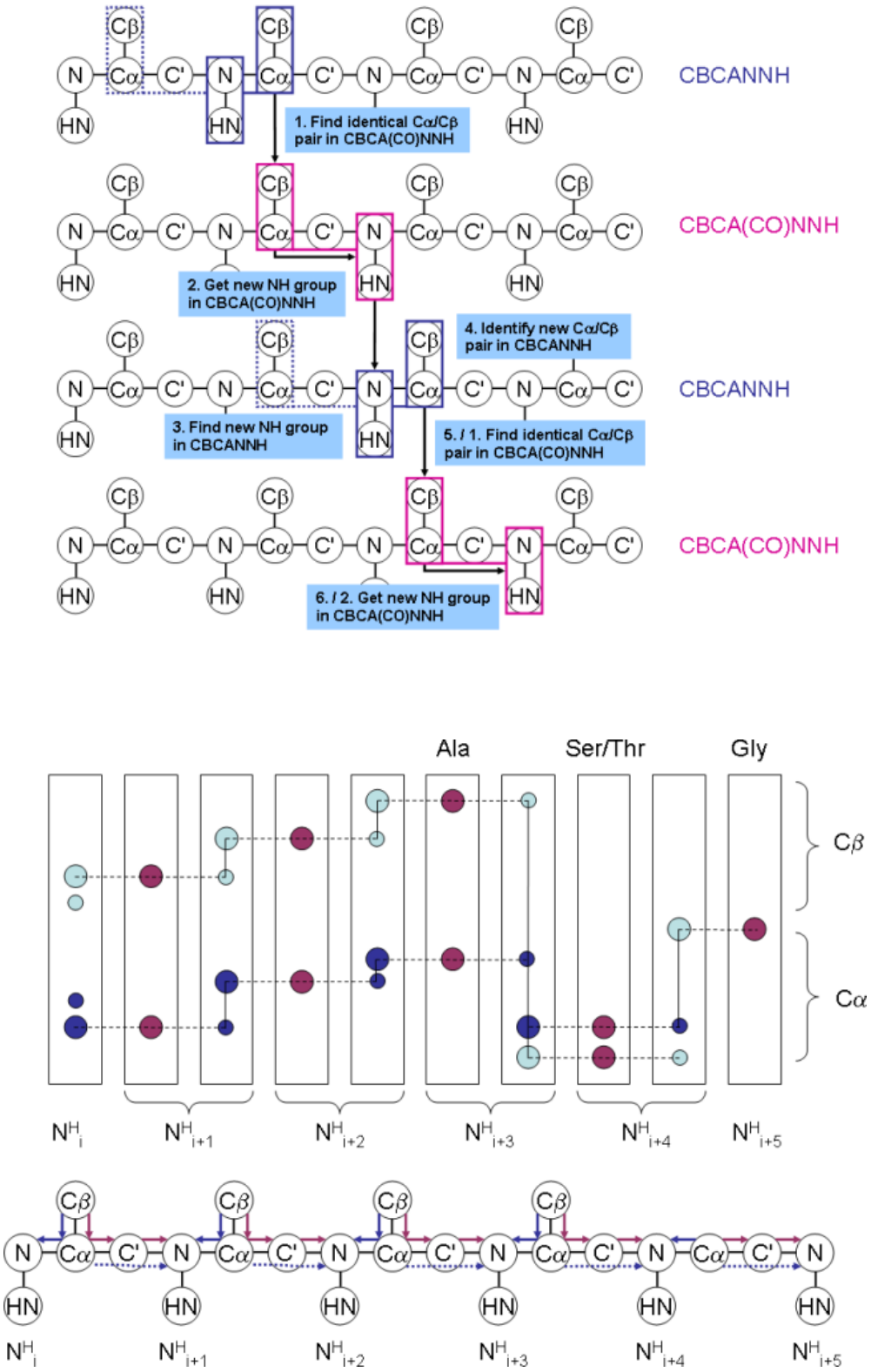


Figure 2.3: Schematic representation of backbone assignment experiments and the magnetization transfer mechanism (image is take from the page https://www.protein-nmr.org.uk/)

For the assignment experiments, both backbone assignments and side-chain assignments of the proteins, uniformly ^{15}N, ^{13}C labeled protein samples are used. For the specific interactions or for very large complexes, specific labeling will be employed according to the need. For the large molecular complexes, deuterated protein samples were prepared. As the molecular size increases the complexity of the NMR assignment would also increase. Higher magnetic fields are always better for the analysis of larger molecular weight biosamples. As the separation of the individual resonances is better in the higher magnetic field, the extent of overlapping can be considerably reduced. At the lower magnetic field, the quality of the NMR spectrum would be greatly reduced in terms of resolution. Hence, for the assignment purpose, it is always better to record the NMR experiments at the higher magnetic field[157–159].

2.8.1 Backbone assignment experiments

Standard triple-resonance backbone assignments experiments for proteins are NHCACB, NH(CO)CACB, HNCA, NHCO, HN(CA)CO.

HNCACB:

The 3D experiment HNCACB is well designed to analyze the correlation of ^{1}H-^{15}N amide resonances with the intra and inter-residual Cα and Cβ resonances, by means of $^{1}J(NH)$, $^{1,2}J(N, C\alpha)$ and $^{1}J(C\alpha, C\beta)$ coupling constants. This is very sensitive experiment can also be used for the proteins, with a shorter ^{13}C relaxation time and also for deuterated proteins. The HNCACB experiment is a 3D spectrum in which 1H,15N, and 13CA/13CB chemical shifts are displayed in three independent dimensions.

Here, the magnetization from the amide proton will be transferred to Cα and Cβ carbons through ^{15}N-^{13}C coupling [$^{1,2}J(N, C\alpha)$], and ^{13}C- ^{13}C coupling. With this experiment, one can get connectivity of particular amide resonances to its Cα and Cβ, of i^{th} and $(i-1)^{th}$ residues.

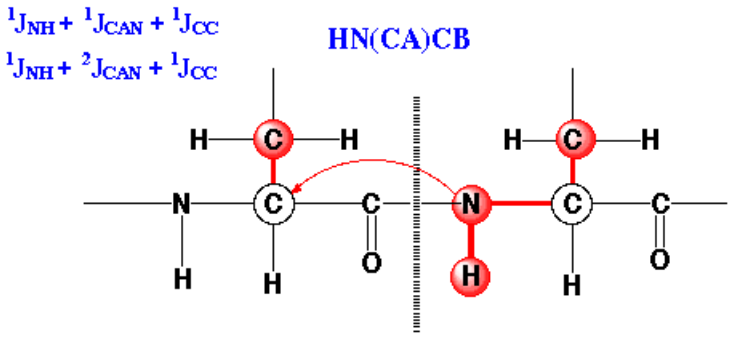


Figure 2.4: Schematic representation HNCACB experiment magnetization transfer interaction. (Image is from :http://triton.iqfr.csic.es/guide/eNMR/eNMR3Dhet/nohsqc3d.html)

NH(CO)CACB/CBCA(CO)NH[160]:

Here the magnetization is transferred from $^{1}H\alpha$ and $^{1}H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$, respectively. Then from $^{13}C\beta$ to $^{13}C\alpha$. Further, it is transferred to the ^{13}CO group and to ^{15}N to ^{1}H. Proton will be detected. The chemical shift is evolved on $^{13}C\alpha$ and $^{13}C\beta$ in one dimension and for the other two dimensions, ^{15}N and ^{1}H are evolved. This output of this experiment is similar to the above NHCACB experiment, but the difference is here only correlations of amide to its $(i-1)^{th}$ residues can be seen. Because, here selectively magnetization has to be transferred to the $(i-1)^{th}$ residues, via ^{13}CO group.

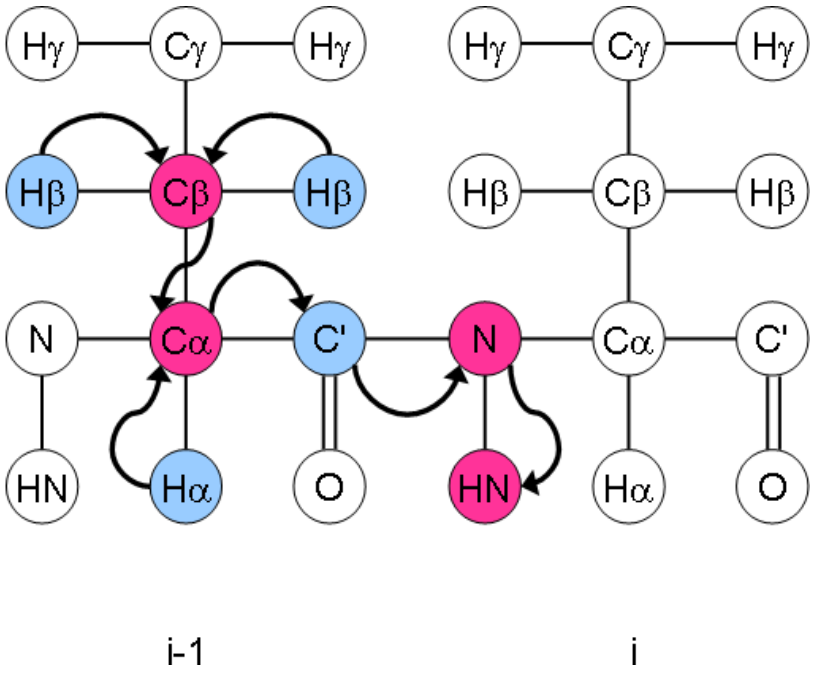


Figure 2.5: Schematic representation of HN(CO)CACB experiments and the magnetization transfer mechanism (image is taken from the page https://www.protein-nmr.org.uk/)

With the help of these two experiments, one can assign particular amide resonances to its the i^{th} and $(i-1)^{th}$ $^{13}C\beta$ to $^{13}C\alpha$. This is how sequential walking is done, for the NMR backbone assignments.

HNCO:

It is the 3D experiment, that outputs a 3D spectrum in which ^{1}H, ^{15}N, and ^{13}CO chemical shifts are depicted in three independent dimensions. It is designed to correlate the fundamental connectivity of the backbone of proteins. That is a ^{13}CO group – $^{15}N^{1}H$ correlation. Here, amide resonances are correlating with the ^{13}C chemical shift of the preceding carbonyl group.

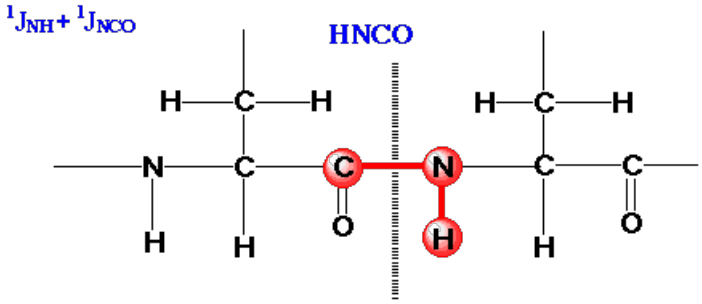


Figure 2.6: Schematic representation HNCO experiment magnetization transfer interaction. (Image is from :http://triton.iqfr.csic.es/guide/eNMR/eNMR3Dhet/nohsqc3d.html)

2.8.2 Side Chain assignment experiments

The basic set of side-chain assignment experiments is HBHA(CO)ONH, H(CCCO)NNH, and CC(CO)NNH. These spectra will provide the chemical shifts of hydrogen and carbon side-chain elements, for the preceding residues for each NH group[161–163].

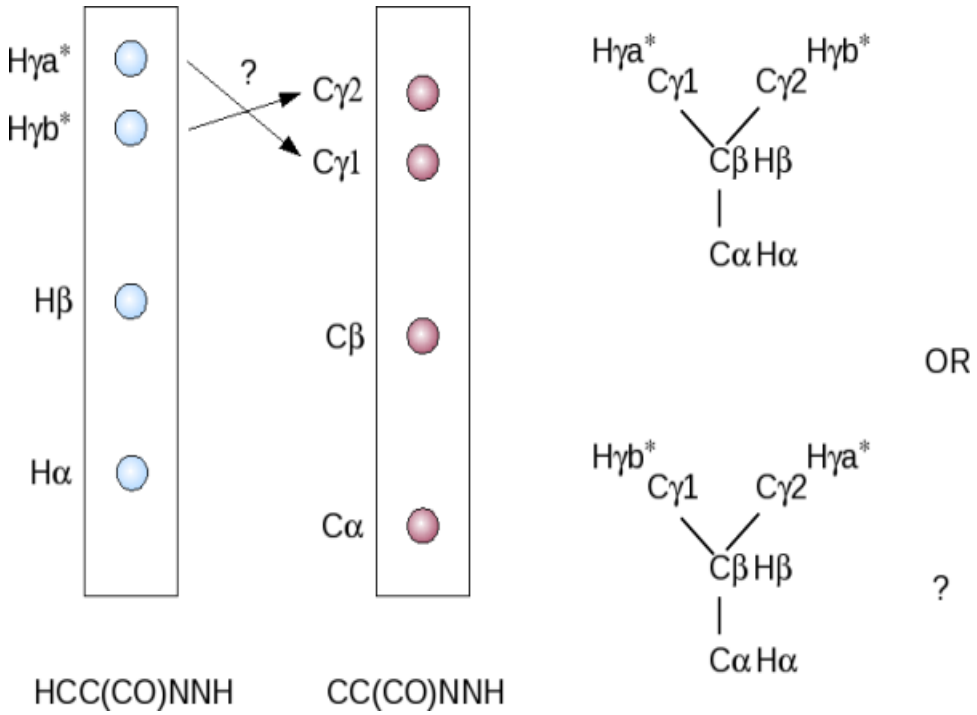


Figure 2.7: Schematic representation of side chain assignment experiments (image is take from the page https://www.protein-nmr.org.uk/)

The widely practiced experiment is the HCCH-TOCSY spectrum along with the HCCH-COSY spectrum, as they provide connectivity between the side-chain carbons and their respective bonded protons, as well as their connected side-chain protons. As COSY spectrum tells about the correlation between the directly bonded carbon-proton groups. Whereas TOCSY is the relayed COSY, where it shows the connectivity of long-range proton-carbon side-chains.

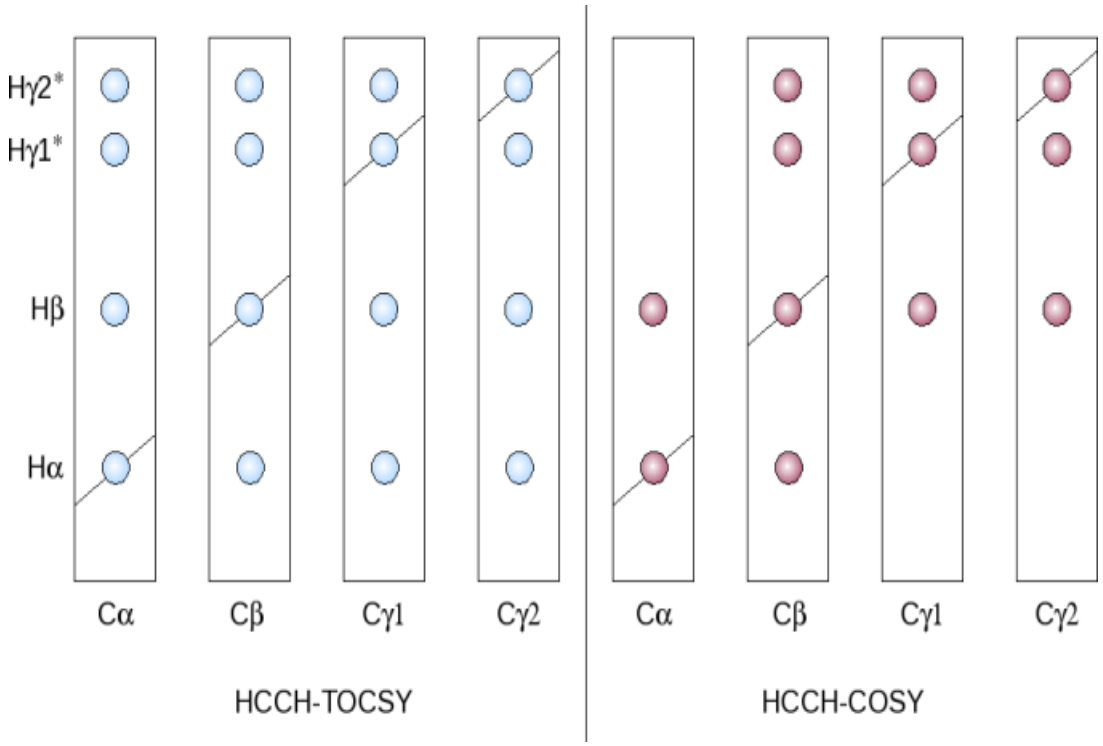


Figure 2.8: Schematic representation of side-chain assignment experiments, TOCSY and COSY magnetization transfer (image is taken from the page https://www.protein-nmr.org.uk/)

3D NOESY-HSQC experiment

Once, we have assigned all the nuclei of the protein, with the help of backbone assignment experiments and side-chain assignment experiments, we can get the distance restraint information by recording the 3D NOESY-HSQC experiment. As NOESY experiment provides information about spatially connected nuclei through hydrogen bonding. In the 3D NOESY-HSQC experiment, homonuclear ^{1}H-^{1}H NOEs can be assigned. So, here we will get to know the spatially connected nuclei information. So, the high number of NOEs data often exhibits the structurally well-defined bio molecule[164,165].

Figure 2.9: Schematic representation distant restraint experiments, 3D NOESY-HSQC magnetization transfer interaction.
(Image is from:http://triton.iqfr.csic.es/guide/eNMR/eNMR3Dhet/nohsqc3d.html)

Once we get all the information regarding sequential and distant restraints, one can go ahead with structural calculation.

CHAPTER

3. Experimental approach

3.1 Protein expression

The DNA sequence of the protein Pym, from *Drosophila melanogaster* (UniProt KB accession code – P82804) was amplified by PCR (Polymerase chain reaction) and cloned into the pETM-11 expression vector (EMBL collection) using NcoI and XhoI restriction sites. This T7 promoter plasmid allows the expression of the target protein fused to the N-terminal His6- affinity tag along with the cleavable TEV (Tobacco Etch Virus) Protease site. The shorter constructs of the protein Pym, which are Pym^{110}, Pym^{134} and Pym^{160} also cloned with the same protocol. Successful cloning was followed by the plasmid DNA -sequencing for the evaluation. All the shorter protein constructs were obtained by mutagenesis with the help of the Phusion High-Fidelity DNA Polymerase kit (NEB).

Mutagenesis protocol involved annealing at 63℃ for one minute, followed by 6 minutes of elongation step at 72℃ for 18cycles. An enzymatic digestion reaction was done after the mutagenesis to digest the wild-type template DNA, carried out at 37℃ for 2 hours with DpnI restriction enzyme (NEB). Subsequently, DpnI digested reaction mixture (10 µl) was used for the transformation reaction. Digested reaction mixture transformed to Top10 *Escherichia coli* cells. Finally, the mutated plasmid was amplified and extracted with the Miniprep protocol (QIAGEN) and subsequently sequenced.

The mutagenesis reaction mixture of 50 µl contains:

- Phusion HF buffer (5x): 10 µl
- Plasmid DNA (100ng/ µl): 1 µl
- dNTP (10mM NEB): 1 µl
- Forward primer (125ng/ µl): 1 µl
- Reverse primer (125ng/ µl): 1 µl
- Phusion HF polymerase: 0.5 µl
- DMSO: 1 µl
- H_2O: 34.5 µl

All the protein constructs were expressed in BL21(DE3) competent cells. This is a strain used for high-efficiency protein expression under the control of the T7 promoter and expressed by IPTG induction. For the production of unlabeled protein, nutrient-rich LB (Luria-Bertani) media was used for the cell culture. Single colony used for 25 ml of

overnight culture, 500 ml of LB media was inoculated by 5ml of overnight culture, and allowed to grow at 37℃ and with constant shaking at the rate of 200rpm in the shaking incubator. The cells were grown until the O.D. (Optical Density) of the cultures reached 0.6-0.8, then the cultures were cooled down for the induction. IPTG ((Isopropyl-D-1-thiogalactopyranoside) of 0.1mM final concentration was added to the culture for the protein expression and cells were allowed to express the protein at 16°C for 16hours. Later cells were harvested and stored at −20°C or −80°C until purification.

For the synthesis of isotopically labeled protein, minimal media(M9) was used. Isotopically labeled proteins are prerequisites for the NMR experiments. Isotopically labeled nutrients were used in accordance with the labeling requirement of protein. So, for the production of double-labeled (uniformly labeled ^{13}C,^{15}N) protein, ^{13}C- uniformly labeled Glucose(4g/l) and ^{15}N- ammonium chloride(1g/l) were used with the minimal media. For the nitrogen labeled protein, ^{15}N- ammonium chloride(1g/l) was used with the naturally labeled glucose for the cell culture. In the case of deuterium labeling, the entire M9 medium was prepared in the D_2O with isotopically labeled glucose and ammonium chloride. In every case, the 50ml of overnight culture was grown in the minimal media from the single colony of bacteria, for overnight. The overnight culture was centrifuged and the cell pellet was used for the inoculation of 500ml M9 media culture. Cells were grown upto 0.6-0.8 of O.D. at 37°C and 200rpm in the shaking incubator, followed by the induction with IPTG (0.1mM final concentration) for the expression of the proteins at 16℃ for 16hours. Cells were harvested and stored at -20℃ or −80℃ until purification.

3.2 Protein purification

Cell pellets were re-suspended with wash buffer (50mM Tris.Hcl, 50mM NaCl, 10mM Imidazole, 5mM β-mercaptoethanol, 5% Glycerol pH 7.5) with the addition of half a tablet of EDTA-free cOmplete™ protease inhibitor cocktails (Roche). Cells were lysed with the help of a sonicator (5 seconds of pulse with 10 seconds of pause, with the amplitude of 49% for 15-20 minutes). The cell lysate was centrifuged at 18 000G for 45 minutes and the supernatant was passed through the 0.45μm filter, before the purification.

The first step of purification is through the affinity chromatography with the help of His-trap FF column (GE Healthcare), where the protein would bind to the column, because of the His6-tag. The protein is washed with 0.5M of Lithium salt buffer(50mM Tris.Hcl, 50mM NaCl, 10mM Imidazole, 5mM β-mercaptoethanol, 5% Glycerol, 0.5M LiCl pH 7.5) in order to wash all the Rnases from the protein. As the protein Pym^{160} is

the RNA binding protein, it is necessary to remove all the cellular RNAs and RNases from the protein during the purification. Later, the protein was eluted with the help of a high concentration Imidazole (1M) buffer (50mM Tris.Hcl, 50mM NaCl, 400mM Imidazole, 5mM β-mercaptoethanol, 5% Glycerol pH 7.5). Before, the second step of purification the imidazole was removed by buffer exchange with the help of a desalting column.

The second step of affinity chromatography is by using the HiTrap Heparin FF column. This method is very much necessary for the purification of RNA or DNA binding proteins. This purification was done for the removal of cellular nucleic acids. Here, the protein would bind to the column (higher PI value of the protein than the pH of the buffer) and the column was washed with low salt buffer (50mM Tris.Hcl, 50mM NaCl, 5mM β-mercaptoethanol, 5% Glycerol pH 7.5) and the protein is eluted with high salt buffer(50mM Tris.Hcl, 1M NaCl, 10mM Imidazole, 5mM β-mercaptoethanol, 5% Glycerol pH 7.5).

After the second step of affinity chromatography, the is protein expected to be free from all cellular RNAs and RNases. So, the RNase Alert™ Lab test (ThermoFisher) was done to evaluate the same. If the protein was found to be bound with RNAs or RNases, the purification steps were repeated until the tests were negative. At this stage, TEV protease was used to cleave of His6-tag from the protein. The final step of purification is gel filtration chromatography.

The purified protein was buffer exchanged to the desired NMR buffer (50mM BisTRis, 50mM MES, 50mM NaCl, 0.05%NaN_3, 2.5mM Tcep, pH 6.0) later concentrated for the NMR experiments. The buffer used for the NMR experiments was optimized by Thermofluor assay[82,83].

3.3 Thermal shift assay

Thermofluor assay or Thermal shift assay (TSA) is very much helpful to measure the changes in the thermal denaturation temperature of the biomolecule. Hence, stability of the biomolecule such as protein under varying conditions such as buffer, pH, ionic strength can be monitored. The method for measuring the thermal stability is differential scanning fluorimetry (DSF) or thermofluor, which employs fluorogenic dyes[84]. SYPRO Orange dye has been used to measure the thermal stability of the proteins[85]. SYPRO Orange has an emission/excitation wavelength profile compatible with qPCR machines which are in recurrent everyday use. SYPRO Orange has the property of nonspecific binding to the hydrophobic surfaces, where water quenches the fluorescence of the dye. In the 64 qPCR plate protein is placed in different buffers of interest, during the experiment the temperature is gradually increased from 4℃ to 95℃

to monitor the thermal stability of the protein. When the temperature starts to increase from beyond its stability protein starts to unfold to expose the hydrophobic surfaces and hence the emission of fluorescence commences[86,87]. The melting point(Tm) is the temperature at which the emission of fluorescence is highest can also be measured. More the Tm more is the thermal stability of the protein in the respective buffer. With this, the optimum conditions for the experiments have been obtained[88–90].

3.4 RNA Synthesis

RNAs of the work interest were synthesized by in vitro transcription method using T7 RNA polymerase[91]. In vitro transcription is the template-directed synthesis process of RNA molecules, that includes a bacteriophage promoter sequence upstream of the sequence of work interest by transcription using the corresponding RNA polymerase[92]. This method of transcription allows transcribing RNA of any virtual length, in μg to mg quantities[93]. The desired sequence is cloned into any phage promoter vectors and the plasmid DNA is linearized with a restriction enzyme. Under the optimized reaction conditions, runoff transcription from these templates yields hundreds of moles of RNA per mole of DNA[91]. Thus, the method of in vitro transcription is very efficient for RNA synthesis. Requisite starting elements for the transcription are DNA template with the T7 promoter sequence, T7 RNA polymerase, nucleoside triophosphate (NTPs), and a buffer system containing DTT and magnesium ions[94,95].

3.4.1 DNA template preparation

The DNA sequence of interest with upstream T7 promoter and PstI restriction site at the 3' end of the sequence was cloned into the high copy number vector pUC19, by EcoRI (NEB) and HindIII (NEB) restriction digest and ligase (NEB). Successful cloning was transformed into the Top10 competent cells (Invitrogen). Transformed Top10 cells were grown in LB media at 37°C and 200rpm for overnight. Centrifuged cell pellets were used for the DNA extraction process. Transfection-grade plasmid DNA was extracted through QIAGEN Plasmid kits, which provide gravity-flow, anion-exchange tips for purification. Purified plasmid DNA with the target sequence was treated with PstI (NEB) restriction enzyme for linearization. As, this is crucial for the production of RNA transcripts of defined length, otherwise slightest impurity of circular plasmid would generate long heterogeneous RNA motifs in higher yield than the linear counterparts[96]. The linearized DNA template was used in the in-vitro transcription.

After the enzymatic digestion, the linearized DNA template was extracted by using the phenol-chloroform extraction. A mixture consisting of phenol—chloroform—isoamylalcohol (05:24:1, Roti) was added to the reaction mixture in an equimolar ratio. The nucleic acid and proteins are separated by phenol, separation of the aqueous and organic layer was facilitated by chloroform which also denatures the proteins. Foaming of the solution is controlled by isoamylalcohol[97]. The DNA was dissolved in the pH adjusted (4.5-5) phenol, while the RNA remained in the aqueous phase. The proteins were separated as a film between the organic and aqueous layers. The chloroform layer was collected and mixed with the chloroform—isoamylalcohol mix (24:1, v/v) to remove the phenol traces. The sample was mixed well and centrifuged. The chloroform layer was separated. To this 10% 5 M NaCl and 2.5 volume of pure ethanol was added, incubated at −20℃ for 3 hours the DNA precipitates. The sample was spun down at 14 000g for 1 hour. The DNA pellet was washed with 70% of ethanol. Centrifuged again, the pellet was dried and dissolved in pure water to the desired concentration.

3.4.2 Preparation of Nucleoside Triphosphates (NTPs)

The NTPs solid salts (from Sigma-Aldrich) were dissolved individually in nuclease-free water. The pH of the NTP solution was neutralized (pH=7-8) using 1M NaOH solution. Each NTP solution was then diluted to the desired concentration and mixed together in equimolar or non-equimolar ratios of different NTPs as of the transcription requirement for the corresponding RNA synthesis[98,99].

3.4.3 Transcription reaction optimization

In vitro transcription reaction is much sensitive to the concentration of NTPs, DNA template sequence, and other salts which contribute considerably to the yield of RNA transcripts. So, for every large-scale transcription reaction, it is very much required to optimize the concentrations of the transcription components, such as DNA template, NTPs, T7 RNA polymerase, and Mg^{+2} ions. Optimization would also be necessary for the new stock solutions of any one of the above mentioned components. 20 µl transcription reaction was set up for every reaction condition. Transcription reaction was carried out at 37℃ for 3 hours. Later all the reaction mixture was loaded to the analytical polyacrylamide denaturing gel. The optimum reaction condition is the one in which a higher amount of homogenized RNA was transcribed. This condition was used for the large-scale synthesis of RNA.

3.4.4 Large scale synthesis of RNA

SOLE RNA, 24mer, and ribozymes HDV(Hepatitis delta virus) and Hammerhead are synthesized in the lab by in vitro transcription method. Both the RNAs and ribozymes were optimized for large-scale transcription reactions. Transcription reaction was carried out at 37℃ for 5 hours in the presence of optimized quantities of Mg^{+2}, NTP, and T7 polymerase. After the transcription, the reaction was quenched with EDTA (Ethylenediaminetetraacetic acid) solution (final concentration is 50mM). Later, the reaction mixture was loaded to the polyacrylamide denaturing gel for RNA extraction. The polyacrylamide gel slice which encapsulated the RNA was transferred to the electroelution apparatus to extract RNA. Both SOLE and 24mer were sequentially homogenized by trans cleaving ribozymes.

SOLE RNA transcribed with few complementary sequences of Hammerhead ribozyme at the 3′end. Both SOLE and Hammerhead ribozyme are annealed together in the cleavage buffer at 95℃ and then cooled down to room temperature. To initiate the ribozyme cleavage reaction, $MgCl_2$ (final concentration is 30mM) is added to the reaction mixture. This trans-acting cleavage reaction was done at 37℃ for 12 hours. The reaction was quenched with EDTA solution (30mM final concentration). Later, the reaction mixture was loaded to polyacrylamide gel, followed by the electroelution method to elute the pure RNA.

Whereas in 24mer, both HDV and Hammerhead ribozymes were used for homogenizing 5′ and 3′ end respectively. Both the ribozymes were used in trans with RNA (to be cleaved), simultaneously for the cleavage reaction. The RNA was extracted from ethanol precipitation[100].

3.4.5 Poly Acrylamide Gel electrophoresis

This is the method of separation and analysis of macromolecules such as DNA, RNA, and proteins depending on their size and charge. By this method, the estimation of the size and population of macro-molecules can be done. Hence, the method is used to estimate the size of DNA and RNA or to separate the proteins by their charge[101]. Nucleic acids, the negatively charged molecules are separated by applying the electric field through the matrix of gel substances like agarose or polyacrylamide. The rate of movement of the molecules through the pores works on the phenomenon of sieving[102]. Smaller fragments move faster and farther. While proteins are separated through the charge. PAGE technique is very useful in the separation and analysis of the nucleic acids, as well as interaction studies[103]. Gel electrophoresis requires an electrical field; in

particular, the electric field applied makes one end of the gel positively charged while the other end is negatively charged. Nucleic acids are negatively charged bio-molecules, they will be pulled toward the anode. Proteins, however are not charged in particular; thus, the proteins are mixed with the detergent called sodium dodecyl sulfate. The treatment makes the protein degenerate and hence unfolds into a negative charge coated linear chain, and can be separated. Upon the separation, the bands representing molecules of different sizes can be detected.
Resolving gels are made in 6%, 8%, 10%, 12%, 15%. Smaller RNAs move faster, can be well separated using 12% or 15% gels. For the *Oskar* sole and 24mer RNAs, 12% gel was used to separate from their corresponding hammerhead and HDV ribozymes. Polyacrylamide gels include urea and acrylamide solution in TBE buffer. Depending upon the percentage of gels required the concentration of acrylamide solution is varied. After polymerization, the transcription reaction was loaded onto the gel and run for 6 to 10 hrs at 15-18V, as per the required resolution. RNAs were extracted from the gels in the denatured form. Extraction can also be done by preserving the high order structure state of the bio-molecules. However, this can little more complex than the extraction from the denatured state, as monitoring the size can be tricky. Because both charge and the molecular size of the RNAs contribute to the position of the band in the gel. RNA was purified through the Electroelution method.

3.4.6 RNA constructs used for NMR and EMSA analysis

SOLE RNA, ribozymes, and 24mer SOLE RNA are synthesized in the lab using the above RNA synthesis method. While the others listed here are obtained from Sigma for NMR and EMSA analysis.

SOLE RNA:
5'- Cy5-GACGAUAUCGAGCAUCAAGAGUGAAUAUCGUC-3'

24mer SOLE RNA:
5'- Cy5-AUAUCGAGCAUCAAGAGUGAAUAU-3'

20mer SOLE RNA:
5'- Cy5-AUCGAGCAUCAAGAGUGAAU-3'

16mer SOLE RNA:
5'- Cy5-CGAGCAUCAAGAGUGA-3'

31mer RNA:
5'- Cy5-GACGAUAUCGAGCUUCGGAGUGAAUAUCGUC-3'

DelA 23mer RNA:

5’- Cy5-AUAUCGAGCAUCAAGAGUGAUAU-3’

Ejpt-24mer RNA:
5’- Cy5-AUAUCGAGCAUCAAAGGUGAAUAU-3’

Alter-24mer RNA:
5’- Cy5-AUAUCGAGCAUCAAUUCGGAAUAU-3’

24mer MSL RNA:
5’- Cy5-GAUAUCGAGCUUCGGAGUGAAUAUC-3’

22mer MSL RNA:
5’- Cy5-AUAUCGAGCUUCGGAGUGAAUAU-3’

22mer MSL-II RNA:
5’- Cy5-AUAUCGCGCUUCGGCGUGAAUAU-3’

22mer MSL-III RNA:
5’- Cy5-AUAUCGCUCUUCGGCUUGAAUAU-3’

16mer MSL RNA:
5’- Cy5-AUAUCGAGAGUGAAUAU-3’

29mer RNA:
5’- Cy5-GGGAUAUCGAGCAUCAAGAGUGAUAUCCC-3’

3.5 Gel Electrophoretic Mobility Shift Assay (EMSA)

EMSA is the native gel electrophoresis, where the RNAs can be monitored in their native form without being structurally degenerated. Hence, it is used to analyze the RNA-protein interactions, where bound and unbound RNAs were easily distinguished. EMSA is based on the principle that nucleic acid-protein complexes are larger, so move slowly show lower electrophoretic mobility, when subjected to denaturing gel, when compared to respective unbound nucleic acid probes. Hence, the rate of shift of nucleic acids is retarded when it is bound to protein. Since the gel shift assay provides the information if the protein is able to directly interact with the specific sequence of nucleic acid, the technique was used to monitor the interaction between Pym[160] and isomers of *Oskar* SOLE RNA. Interaction study of SOLE RNA and all the other

isomeric constructs of SOLE were observed through EMSA, as holo and apo form with the protein Pym[160].

For the EMSA, both RNA and the protein Pym[160] are in the same buffer (50mM MES, 50mM BisTris, 50mM NaCl, 2.5mM Tcep) are mixed in the ratios of 1:0,1,2 and 5. The reaction complex was loaded into the Polyacrylamide gel. Electrophoresis for about 5 hrs at 5V at 4°C. Crisp bands can be obtained at the lower temperature, otherwise, bands will appear to be fuzzy because of thermal noise at room temperature. 0.1nMoles of the RNA was used for the EMSA.

3.6 NMR Spectroscopy

NMR experiments were recorded in the Bruker 600MHz and 850MHz spectrometers and are equipped 5mm HCN triple resonance probe with nitrogen and helium cryogenic probe cooling system respectively. NMR spectroscopy is the major analytical technique used for the interaction study, dynamics and structural characterization analysis of protein Pym[160].

3.6.1 Titration experiments

Interaction of the Pym[160] with SOLE RNA and with its other shorter constructs was monitored by NMR titration experiments. Protein was monitored by ^{15}N-labelled Pym[160] through NMR with unlabeled RNA constructs. The concentration of 50µM of the protein was used for all the NMR titration experiments, with RNA in the molar ratios of 1: 0, 1, 2, 5 for each protein-RNA titration.

3.6.2 Protein- RNA interactions: Calculation of Chemical shift Perturbation (CSP)

The nature of binding for SOLE RNA and its different structural and sequential isomers with the protein Pym[160] was monitored by a series of NMR titration experiments. Here, all the proton – nitrogen correlation shifts were observed through ^{15}N-^{1}H HSQC experiments. So, the chemical shift perturbations for all the amide resonances were calculated according to the below equation[166]

$$CSP = \sqrt{\frac{1}{2}[\Delta\delta_H^2 + (0.15 * \Delta\delta_N^2)]}$$

Where, $\Delta\delta^H$ and $\Delta\delta^N$ is the chemical shift difference, in the proton and nitrogen dimension respectively. Data analysis and graphical representation were done by using CCPNmr Analysis and Microsoft Excel.

3.6.3 Backbone assignments experiments

Backbone nuclei N, H^N, Cα, Cβ, and C' of the protein Pym and Pym[160] were sequentially assigned from the conventional triple resonances experiments like HNCA, HNCO, HNCACB, HN(CO)CACB, and HN(CA)CO and are explained below[167].

HSQC[168]: ^1H-^{15}N HSQC is a two-dimensional fingerprint experiment for proteins. Here, the magnetization is transferred from amide protons ($^1H^N$) to nitrogen ($^{15}N^H$) via J-coupling. This is the standard experiment with which all the N-H correlations can be observed. With all the amide groups, side-chain groups of Trp, Asn, Gln are also visible. Even though it is a basic vital experiment recorded for all the proteins, it has the potential to reveal the spectral and structural ambiguity of the protein sample.

HNCA[169–171]: This is a very sensitive 3D NMR experiment, where the magnetization is transferred from amide proton to nitrogen and then selectively transferred to carbonyl group via $^{15}N^H$-^{13}CO J-coupling and then magnetization is transferred back to $^1H^N$ via $^{15}N^H$ for the detection. Chemical shift evolves for all the three dimensions. The carbonyl chemical shift can be further used with TALOS[172] to predict the secondary structure during structure calculation.

HNCO[169,170,173]: This is the 3D NMR experiment, where the magnetization is transferred from $^1H^N$ to $^{15}N^H$and then to Cα and transferred back to ^{15}N and ^1H for detection via N-Cα J-coupling. The chemical shift evolved for $^1H^N$, $^{15}N^H$, and ^{13}Cα resulting in a 3D spectrum. Here, the both ^{13}Cα of one's own residue and of the preceding one can be observed, as amide proton is coupled to both of them. Since the coupling of the directly bonded Cα is stronger, two peaks are observed with the intensity difference.

HNCACB[170]: In the spectrum, each amide peak is observed with two Cα and two Cβ, which are from its own residue, and the other set is from its preceding residue. Here, the magnetization is transferred from Hα and Hβ. Then will proceed to Cα and Cβ respectively and finally from Cβ to Cα. From here it is transferred to $^{15}N^H$ and then to $^1H^N$ for detection. Because the magnetization is transferred to $^{15}N^H$ from both $C\alpha_i$ and

$C\alpha_{i-1}$. The chemical shift evolved simultaneously on Cα and Cβ in one dimension and the other two dimensions are for nitrogen and proton.

HN(CO)CACB[170]: Here, the magnetization is transferred from Hα and Hβ to Cα and Cβ respectively and then from Cβ to Cα. From here it is transferred first to ^{13}CO and then to $^{15}N^{H}_{i}$, then to $^{1}H^{N}_{i}$ for detection. The chemical shift evolved simultaneously on $C\alpha_{i-1}$ and $C\beta_{i-1}$ on one dimension but did not evolve on ^{13}CO. The other two dimensions are for $^{15}N^{H}_{i}$ and $^{1}H^{N}_{i}$. Together with the NHCACB spectrum, it is possible to link sequentially i^{th} and $(i-1)^{th}$ residue.

NH(CA)CO[174]: Here, the magnetization is transferred from $^{1}H^{N}$ to $^{15}N^{H}$ via the N-Cα J-coupling to the $^{13}C\alpha$. From here the magnetization is transferred to the ^{13}CO via $^{13}C\alpha$-^{13}CO J-coupling. Magnetization transferred back way from ^{13}CO, $^{13}C\alpha$,^{15}N and finally, ^{1}H to for the detection. The chemical shift evolved on ^{1}H, ^{15}N, and ^{13}CO and not on the $^{13}C\alpha$. Because the amide nitrogen is coupled to the Cα of its own residue and to its preceding residue. Magnetization transfer happens to both the ^{13}CO nuclei. Thus for each NH group, two carbonyl groups are observed in the spectrum with different intensities as coupling residue possesses stronger coupling.

3.7 Structure Calculation

Protein structure prediction is the essential part of structural biology, and essentially covers approaches enabling us to travel from the first sequence, via secondary and tertiary structure, to the Quaternary structure. This follows the central assumption that a protein's primary sequence and therefore the inherent properties of the bio-molecules like proteins or RNAs dictate the ultimate folded of three-dimensional structure. Besides the homology predictions, which are generally obtained through knowledge-based potentials or algorithms, or by comparing to already existing structures of systems with similar sequences, analysis of the standard of the resulting model is a necessary part of protein structure prediction.
For the structure calculation, the distance restraint data from the NMR has been used. All the data has been incorporated into ARIA for the structure calculation.
For the structural restraints, 3D 15N-edited ^{1}H-^{1}H NOESY (mixing time, 150ms) and 3d 13C edited ^{1}H-^{1}H NOESY[104], (mixing time, 150ms) were recorded. The chemical shift index was calculated from the CCPNmr 2.4. Unassigned peaks were discarded, so a total of 2163 unambiguously assigned distant restraints were used for the structure calculation. The total calculated structure was 100, of which the 20 lowest-energy structures were further refined in explicit water. NMR structural characterization (the

final tuning) for the protein and the molecular docking of the protein Pym[160] with SOLE RNA, was done by Dr. Deepshikha Verma.

3.8 Multi-angle light scattering (MALS)

MALS experiment was done by loading a pre-purified protein sample of Pym[160], to Superdex 75 10/300 (GE Healthcare) in line with Optilab, T-rEX and miniDAWN TREOS (Wyatt). The buffer used here is 50mM MES, 50mM BisTRis, 50mM NACl, 2.5mM Tcep. Serum Bovine Albumin (Sigma) was performed to calibrate the detectors. Data were processed with ASTRA7.1.4(Wyatt).

3.9 Materials, Kits and Chemicals

Table 3.1: NMR Spectrometers

NMR Spectrometer	
Avance III HD 600 MHz with N2-cooled inverse HCN triple-resonance cryogenic probehead	BRUKER, USA
Avance III HD 650 MHz with He2-cooled inverse HCN triple-resonance cryogenic probehead	BRUKER, USA

Table 3.2: Software used

	Name	Developer
NMR data processing ans analysis	Topspin 3.2	Bruker
	NMRPipe	Frank Delaglin
	CCPNmr Analysis 2.4	
	ARIA	
Molecular Visualization	PyMOL 2.3	Schrodinger LLC.
Biochemistry and Biophysics	UNICORN 7	GE Healthcare, USA
	ASTRA7.1.4	Wyatt Technology, USA
	NanoAnalysis	TA Instruments, USA
Data plotting	Microsoft Office Excel	Microsoft, USA
	Libre Impress	Libreoffice
	Origin Pro 8	OriginLab Corp, USA
	Illustrator CC 2018	Adobe, USA

Table 3.3: Biological Buffers

Buffer name	Composition
NMR buffer	50mM MES
	50mM BisTris
	50mM NaCl
	2.5 mM Tcep
	pH 6.0
Wash buffer	25mM Tris.Hcl
	25mM Nacl
	2.5mM 2-mercaptoethanol
	5% Glycerol
	pH 7.5
Elution Buffer	50mM Tris.Hcl
	50mM Nacl
	5% Glycerol
	2.5mM 2-mercaptoethanol
	pH 7.5
EMSA buffer	50mM MES
	50mM BisTris
	50mM NaCl
	2.5 mM Tcep
	100mM TBE
	pH 6.0
PBS Buffer	100mM Na-Phosphate
	250mM NaCl
	pH 7.5

Table 3.4: Kits

Product	Manufacturer
QuikChange Site-Directed Mutatgenesis Kit	Agilent, Germany
QIAprep spin miniprep kit, Qiagen PCR purification kit,QIAEX II gel extraction kit	Qiagen, Germany
cOmplete™EDTA-free Protease Inhibitor Tablets	Roche Diagnostics, Switzerland
HisTrap FF (5ml), HiLoad 16/600 Superdex 200 pg, HiLoad 26/600 Superdex 75 pg	GE Healthcare, USA

Table 3.5: Solutions for protein expression

Name	Amount	Chemical	Details
Kanamycin 1000x	50mg/ml	Kanamycin	Stored at −20°C
IPTG	1M	IPTG	Stored at −20°C
Amphicillin	100mg/ml	Amphicillin	Stored at −20°C
Trace Elements	50 ml	0.1M $FeCl_3$ in 0.12M HCl	For 100 ml, in H_2O
	2 ml	1M $CaCl_2$	
	1 ml	1M $MnCl_2 \cdot 4H_2O$	
	1 ml	1M $ZnSO_4 \cdot 7H_2O$	
	1 ml	0.2M $CoCl_2 \cdot 6H_2O$	
	2 ml	0.1M $CuCl_2 \cdot 2H_2O$	
	1 ml	0.2M $NiCl_2$	
	2 ml	0.1M $Na_2MoO_4 \cdot 2H_2O$	
	2 ml	0.1M Na_2SeO_4	
	2 ml	0.1M H_3BO_4	
	36 ml	H_2O	
M9 minimal medium	1000 ml	H_2O or D_2O	For medium
	100ul	1M $CaCl_2$	
	2ml	1M $MgSO_4$	
	6ml	5 mg/ml thiamine hydrochloride	
	4g	D-glucose	
	100ml	10x M9 salts	
	1g	NH_4Cl	
	1ml	Kanamycin 1000x	
	1ml	Trace elements 1000x	

CHAPTER

4. RESULTS

The exon junction complex is the functional binding partner for both Oskar mRNA and for the protein Pym. All together, they are involved in the process of localization of mRNA. Pym recycles the exon junction complex during early translation[68,69] by bridging between the mRNA-bound exon junction complex and with the translation machinery ribosomes[175]. To understand the mechanism of localization, it is necessary to understand the molecular interactions of all the components involved. As the protein Pym is an RNA binding protein and is also involved in the recyclization process of the exon junction complex. So, here we aim to study the structure and dynamics of protein Pym^{160} (the shorter construct of protein Pym) and also to look into the interaction between protein Pym^{160} and Oskar SOLE RNA. The approach to study the structural characterization and dynamics of protein Pym^{160} is through Nuclear Magnetic Resonance spectroscopy. Whereas for the interaction study both EMSA and NMR have been used.

4.1 Structure and Dynamics of protein Pym^{160}

4.1.1. Purification of Pym^{160}

Pym from *Drosophila melanogaster* is a 24 kDa protein comprised of 211 amino acid residues. This molecular weight is feasible for NMR characterization. But, the RNA-Protein interaction study, which is Pym and SOLE RNA complex, would have a total molecular weight of 34 kDa. In general, when the molecular weight of a target component exceeds 30 kDa, assignment and other characterization measurements would become difficult, as the challenges for the structure determination increase steeply with molecular size. This is because of the line broadening and increase in the degeneration of signals owing to fast transverse magnetization decay[97]. Hence, the shorter constructs, which are of lower molecular weight, are advantageous.

The first results of interaction showed that the initial hundred residues of the protein from the N-terminal region participated interactively with SOLE RNA. So, shorter constructs of protein Pym were made while retaining its biologically significant motifs. In this process, the shorter construct of Pym with 160 residues (Pym^{160}, 18kDa) was

found to be the best, as it embodied the functionally significant motif and could be purified as a soluble protein. Other shorter constructs, like Pym^{110} and Pym^{145}, were expressed as insoluble aggregates. So the protein Pym^{160} has been continued for structural characterization and binding study.

Pym^{160} expressed and purified using the optimized three-step purification method, which includes RNases and RNAs elimination steps. The purified protein (labeled /unlabeled) without any RNases and RNAs was used for further study. Finally, the purified protein was analyzed through the SDS-polyacrylamide gel electrophoresis to confirm the size and purity of the recombinant protein.

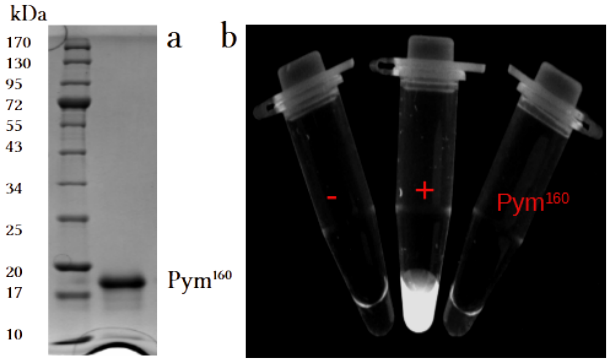


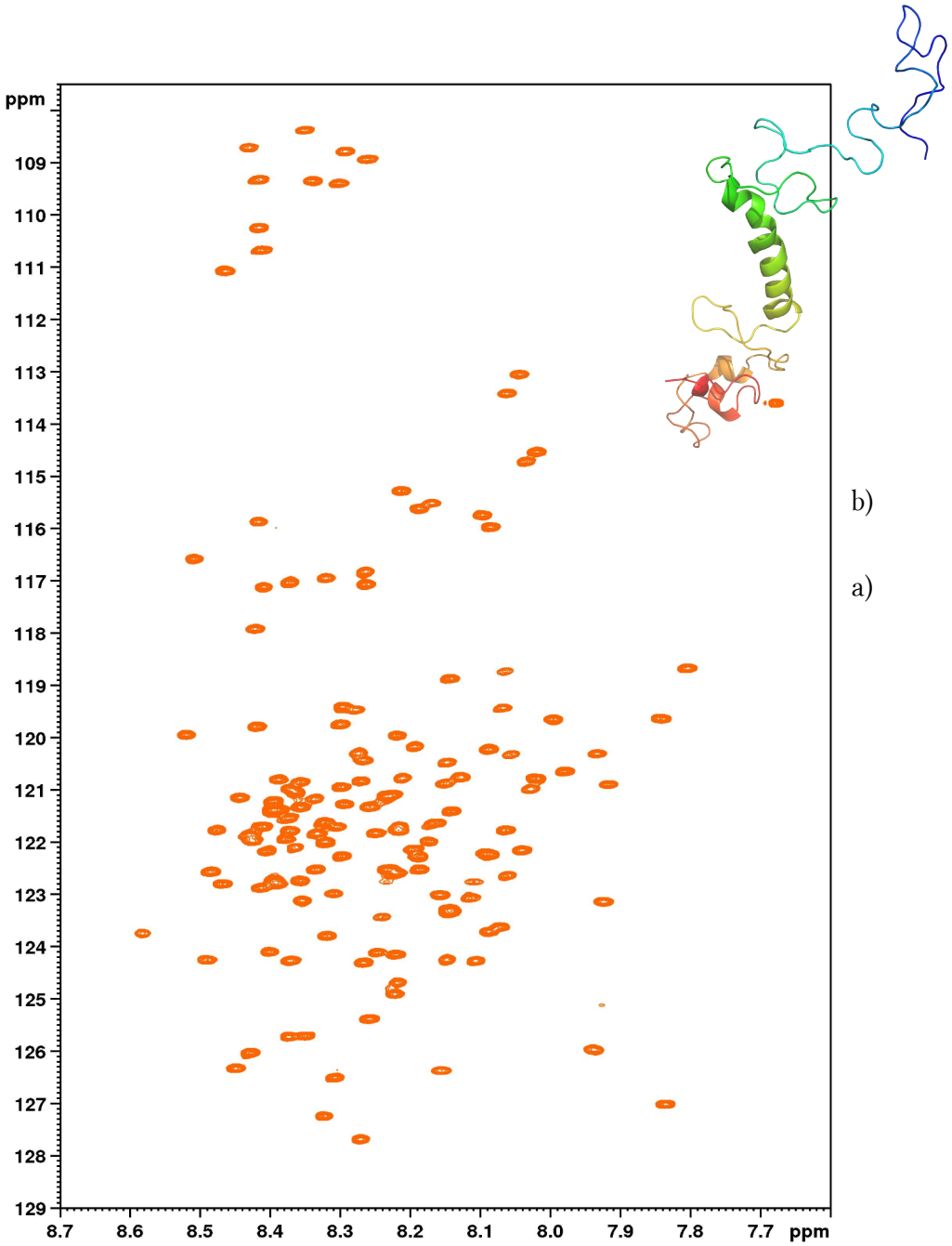
Figure 4.1 – **a**: SDS gel image for the purified Pym^{160} protein (18kDa), **b**: RNases kit Test-Image represents the absence of RNases or RNAs (Pym^{160}) with positive and negative control of RNases or RNAs.

4.1.2. Structural characterization of Pym^{160}

The solution-state NMR technique has been applied for the structural characterization of the protein, that is through collecting structural restraints data. Uniformly double-labeled Pym^{160} in the buffer (50mM MES, 50mM BisTris, 50mM NaCl, 2.5mM Tcep) has been used for the NMR experiments. All the resonances of the molecule have been assigned first followed by the measurement of distant restraints through three-dimensional NOESY experiments. Backbone assignment and side-chain assignment were done by using CCPNmr Analysis V2. Sequential assignments were successfully finished and then followed by the measurement of distant restraints. All this information was further used for the structural calculation by ARIA.

Backbone assignment experiments like HSQC, HNCA, HNCO, HNCACB, HN(CO)CACB, side-chain assignment experiments like H(CCCO)NH, HC(C)H - TOCSY, and structural restraint experiments like 3D NOESY – ^{13}C HSQC were recorded for the unbound protein Pym^{160} on AV600 MHz and AV850 MHz Bruker NMR spectrometers. The ^{1}H-^{15}N-HSQC (Heteronuclear Single-Quantum Correlation Spectroscopy) spectrum, is the fingerprint experiment for the proteins, from which the basic structural nature in terms of structural complexity can be predicted. All the amide resonances or the nitrogen-proton correlations peaks of amides were shown in

Figure 4.2: ^{15}N-^{1}H HSQC spectrum of protein Pym^{160}, depicting all the expected amide resonances(a). Pym^{160} three dimensional structure, calculated from the NMR structure calculation with the software ARIA (b).



the Figure 4.2, which is a ^{1}H-^{15}N-HSQC spectrum, recorded at 293K, for 200 µM Pym^{160} protein in the above mentioned buffer. In the spectrum, the amide resonances are distributed over a narrow range of chemical shifts in the X-axis, that is from the proton

chemical shift. This reflects the high mobility of the protein, which generally describes the structurally unfolded nature of the protein Pym[160]. The spectrum also has some severely overlapped regions, while the other cross-peaks are sharp without much line broadening.

The next step was to assign all the resonances of the spectrum for the protein Pym[160]. That means we have to find out which chemical shift corresponds to the particular nuclei of the protein. This was achieved by the sequential walking method, from the information derived from the various NMR experiments. The assignment was accomplished for all the resonances of the molecule, starting from the backbone assignment spectra to the side-chain assignment spectra for the protein Pym[160]. Once we have assigned all the nuclei of side-chain and backbone elements of the protein, we could then get the distance restraints by interpreting the three-dimensional NOESY spectrum. The intensity of the cross-peaks in the NOESY spectrum determines the spatial arrangement of the two nuclei involved. If the peak is highly intense, then it is termed as strong and the nuclei involved are spatially close to each other. While weaker cross-peaks describe the nuclei bound to each other are comparatively farther. This is how the distance restraints are obtained from the three-dimensional NOESY spectrum. Here, the distance could be a maximum of 5A° between two nuclei.

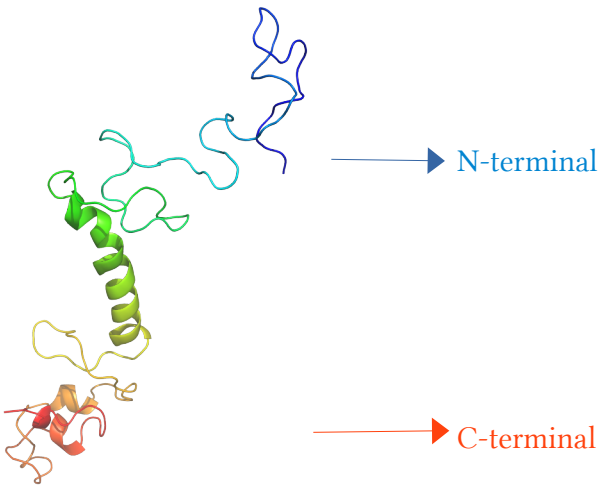


Figure 4.3: Three-dimensional structure of the protein, calculated from the NMR structural characterization method. This is from the first set of calculations, and one from 20 optimized structures. The structure shows the middle helical region, N-terminal unfolded part and C-terminal shows few helical elements. As one can see N-terminal is completely unfolded in the unbound protein Pym[160].

Secondary structure chart

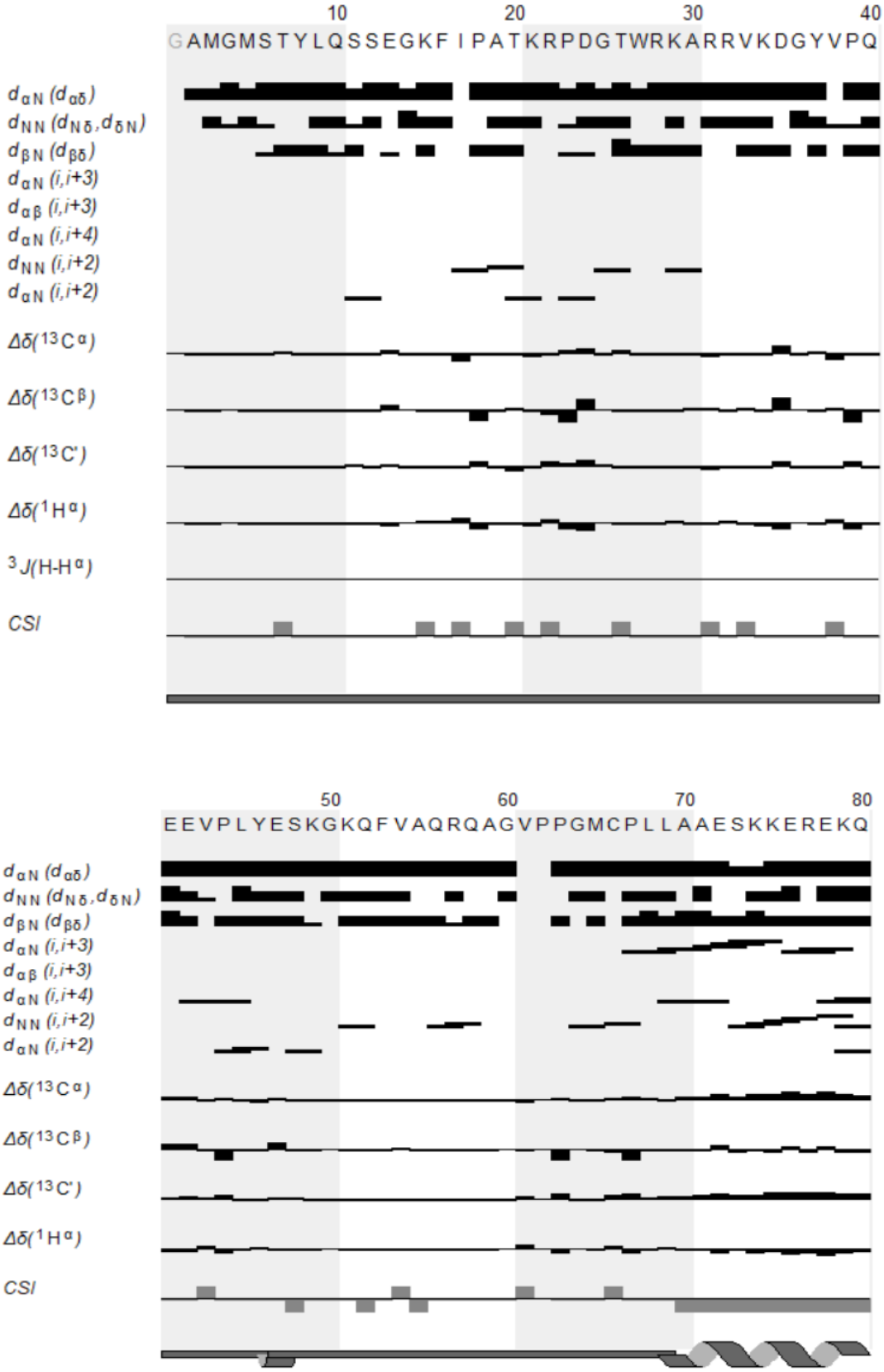


Figure 4.4: CCPNmr chemical shift, assignment table, depicting the assignment upto 80 residues. (Secondary structure chart)

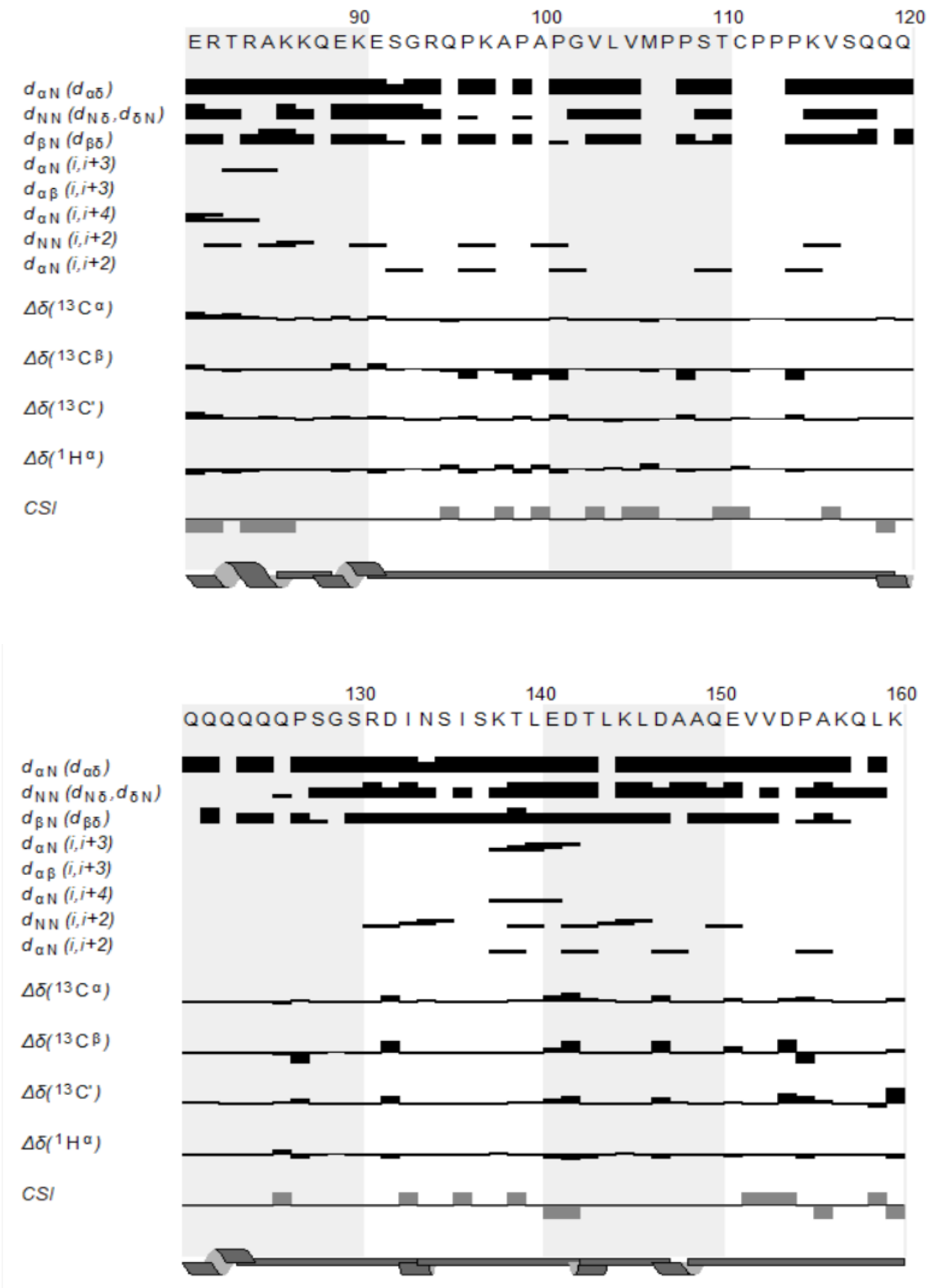


Figure 4.5: CCPNmr chemical shift, assignment table, depicting the assignment from 80 to 160 residues. (Secondary structure chart)

The Secondary structure chart represents NMR assignments, analysis of distant restraints, Chemical Shift Index (CSI), dihedral angles, and secondary structural elements. Figures 4.4 and 4.5 represent the secondary structural chart for the protein Pym[160], where all the residues from 1-160 have been represented. This chart has been generated using the program CcpNmr Analysis.

Chart table showing dαN, dNN, dβN a graphical representation of NOEs observed between protons HN and Hα, HNn and HNn+/−1, HN and Hβ, HNn and HNn+/−2, Hα and HNn+/−2, showing secondary chemical shifts calculated for Hα (Δδ(^{1}Hα)),^{13}Cα (Δδ(^{13}Cα)), ^{13}Cβ (Δδ(^{13}Cβ)) and ^{13}CO (Δδ(^{13}CO)), showing the Chemical Shift Index (CSI)[176,177]. The secondary chemical shifts are the difference in the chemical shift observed and the chemical shift of random coil value. So, the higher difference represents a stronger possibility of a secondary structural element. The chart also shows the DANGLE prediction (Dihedral Angles from Global Likelihood Estimates) of Pym[160] secondary structure[178–180].

The chart describes the presence of a strong α-helix, which forms between the residues 68-91 along with fewer possible small helical elements. Hence, the fewer distant restraints predict the random coil nature of the protein.

4.1.3 Dynamics of Pym[160]

To understand the internal dynamics of the protein Pym[160] hetNOE data has been recorded as the function of the primary sequence. It is the experiment to visualize the protein backbone dynamics from spin-relaxation data[181]. Here, the set of heteronuclear NOE NMR experiments was recorded for amide resonances with and without proton saturation. When RF pulse irradiated protons are allowed to saturate, there will be a transfer of nuclear spin polarization. That would take place through cross-relaxation. Thus, results in a change in the integrated intensity of the amide resonances. Nuclei that are closely situated through space are the ones directly affected by the RF perturbation. hetNOE data is measured as the difference in the peak intensities in two spectra before and after saturation. In other words, it is the integrated change in the intensity of resonances between saturated and unsaturated (reference) experiments. The backbone hetNOE experiment provides information motions of N-H bond vectors[181]. Molecules that have faster-tumbling motions would show a decrease in the NOE intensity when compared to the average observed majority of residues.
So, hetNOE calculated as,

$$NOE = \left(I_{saturated} / I_{reference} \right) - 1$$

In the figure 4.7,

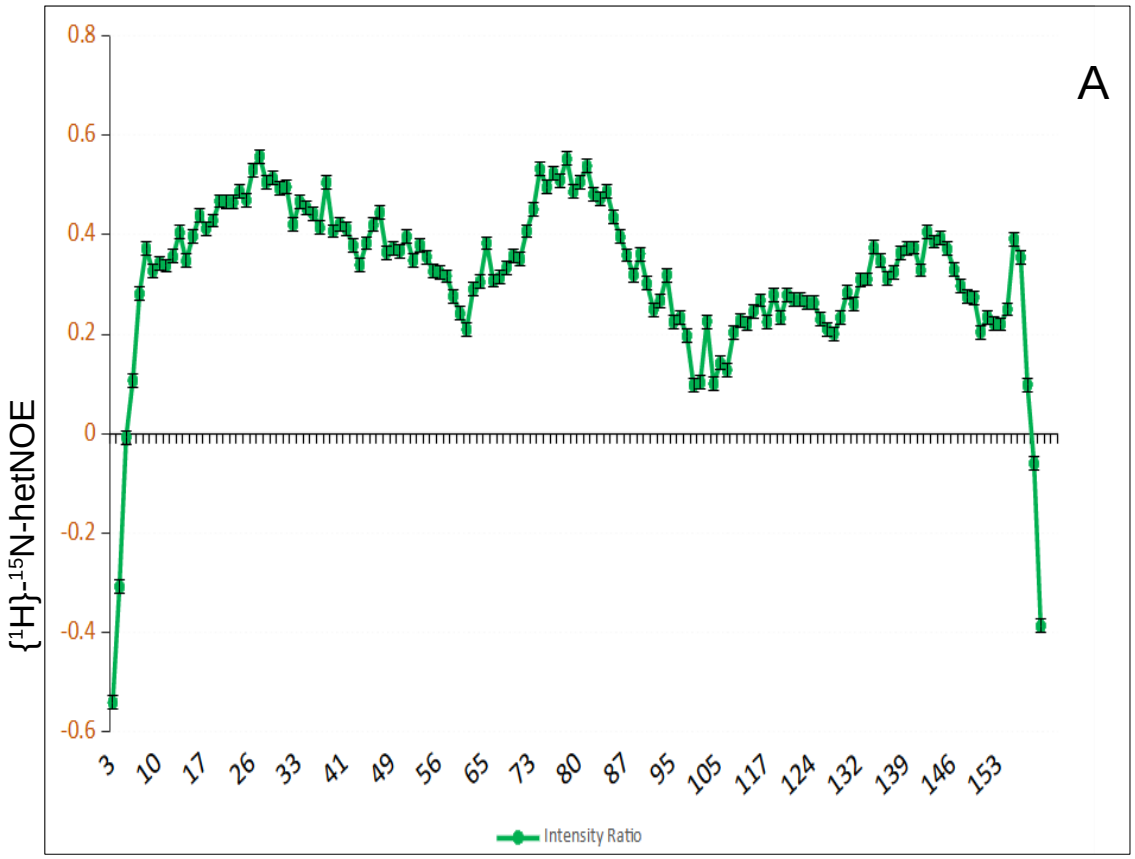


Figure- 4.7: (A)-Display of {^{1}H}-^{15}N heteronuclear NOE values are shown with errors.

Figure **(A)** displays the values of {^{1}H}-^{15}N heteronuclear NOE[182], which means values of magnetic relaxation of ^{15}N amide nuclei. The intramolecular motions of the protein can be estimated with the help of this experiment. {^{1}H}-^{15}N nuclear Overhauser effect is measured in the picosecond time scale of the protein Pym[160]. Figure 4.7 (A) does not show any definite groove, that could predict any secondary structural element, even though the protein has a strong α-helix in the middle of the protein. This could probably be because of the other stronger random coil elements of the protein. Thus, the chart depicts the stronger structural dynamic nature of the protein Pym[160].
Figure 4.7 **(B)** represents the carbon secondary chemical shift data. The chart identifies the secondary structural elements through the analysis of backbone ^{13}C chemical shifts. In the chart, the purple bars represent the presence of the α-helices and the blue line above the chart describes random coil elements.

All these analyses show that the protein Pym[160] is very dynamic and is identified as a 'structurally unfolded protein' with one long α-helix with other possible small helices.

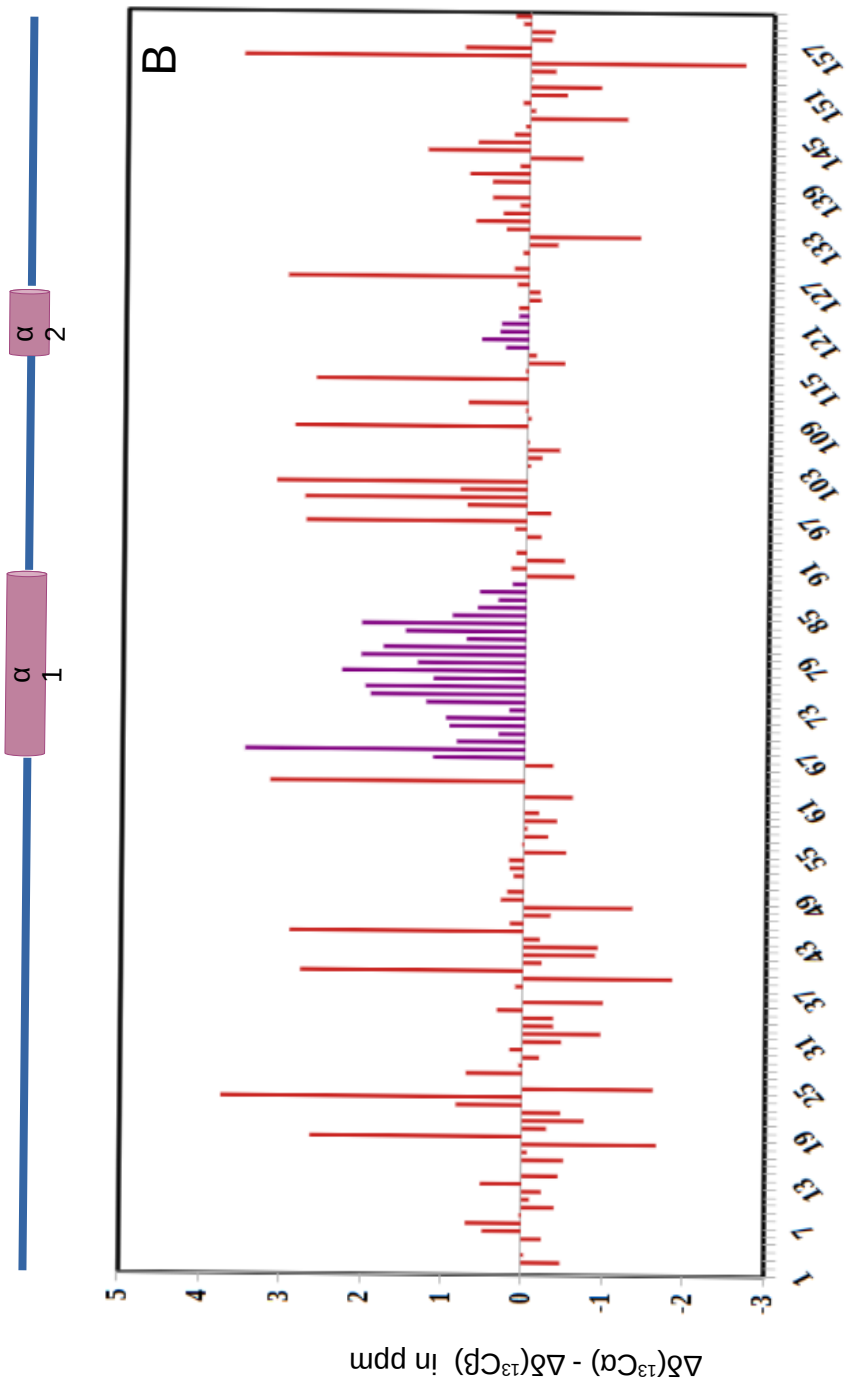


Figure- 4.7: (B)- combined Cα/Cβ carbon secondary chemical shift (SCS) values. SCS are interpreted with respect to their secondary structure elements as shown above the panel. Purple bars represent the helical elements and blue lines are unstructured part.

4.2. SOLE RNA

One of the necessary requirements for the Oskar mRNA localization is the SOLE (Spliced Oskar Localized Element) RNA. SOLE, the first intron spliced RNA, consists of exon-1 and exon-2 ligated together at the first exon junction point. Exon-1 has 18 nucleotides and exon-2 has 10 nucleotides. SOLE RNA embodies structurally short Proximal stem (PS, 6 base pair (bp)), Medial Stem Loop (MSL, 9nt), and Distal Loop (DL, 7nt), as shown in the below figure (). The residues (4) incorporated in the end are for the better RNA transcription.
The structural and dynamic characterization of SOLE RNA has been done through NMR biophysical technique[183].

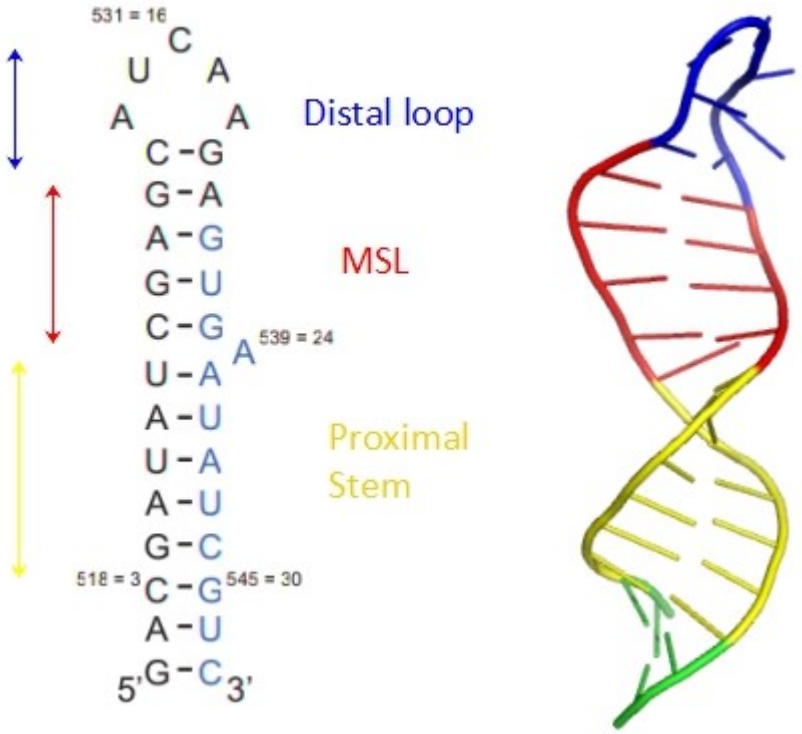


Figure 4.8: SOLE RNA (32mer) embodies Distal loop, Medial stem loop and Proximal stem. Sequential representation and 3D structure of SOLE RNA (5a18, pdb database).

SOLE RNA structurally has an upper bulged distal loop (blue), middle bulged-helical dynamic Medial stem loop(red) and strong helical proximal stem(yellow).

4.3. Dissecting SOLE RNA, for Pym binding motifs

In order to identify the protein binding motifs in the SOLE RNA, the dissection method was employed. The dissection of SOLE into various isomers that are both structural and sequential, were synthesized. Each isomer was observed as the complex with the protein Pym[160], through different analytical tools. The methods used for comprehensive

experimental identification of the RNA-protein interactions are ElectroMobility Shift Assay and NMR spectroscopy.
EMSA is being the first approach for the identification of the interaction of RNA-protein, by observing the RNA as an RNA-protein complex, through native polyacrylamide gel. The band for the binding is observed in case of any interaction present between the molecules. Later, NMR spectroscopy is used as the principal analytical tool. NMR spectroscopy provides an accurate interaction profile through chemical shift perturbation methods, where one molecule is titrated against the other with the concentration gradient of the latter.
The SOLE RNA isomers which exhibited interaction in the EMSA assay were proceeded through NMR titration experiments to study the interaction at the molecular level.
All the SOLE isomers have been screened by these approaches in order to have the complete interaction profile of SOLE RNA with the protein Pym[160]. The study has provided the binding map of SOLE and Pym[160] interaction.
Dissection of SOLE RNA was divided into different categories depending upon different structural and sequential elements that were being modified or deleted.

4.3.1. Sole structural elements

Structurally, SOLE RNA has been divided into three different parts. Proximal stem, Medial stem-loop, and Distal loop.
So, here five isomers were synthesized, 16mer, 20mer, 24mer, 22merDS, and 31mer, where each isomer is missing one of the structural elements.

Isomer 16mer consists of a distal loop and medial stem-loop and is devoid of proximal stem-loop, 20mer is a little longer construct of 16mer where it has initial nucleotides from the proximal stem, 22merDS (double-stranded) is the proximal stem part isomer, with missing nucleotides of the medial stem-loop region and distal loop, whereas the isomer 31mer is missing the distal loop having medial stem-loop and proximal stem intact. 24Mer is a shorter construct of Oskar SOLE, devoid of few nucleotides of the proximal stem. Sequences for the isomers and their corresponding secondary structures have shown in figure 4.8.

So, by analyzing each of these isomers one can find out the direct interaction of the structural elements with the protein Pym[160]. All these isomers with the protein Pym[160] were first analyzed through EMSA, in order to see the nature and strength of the binding and then followed by NMR titration experiments.

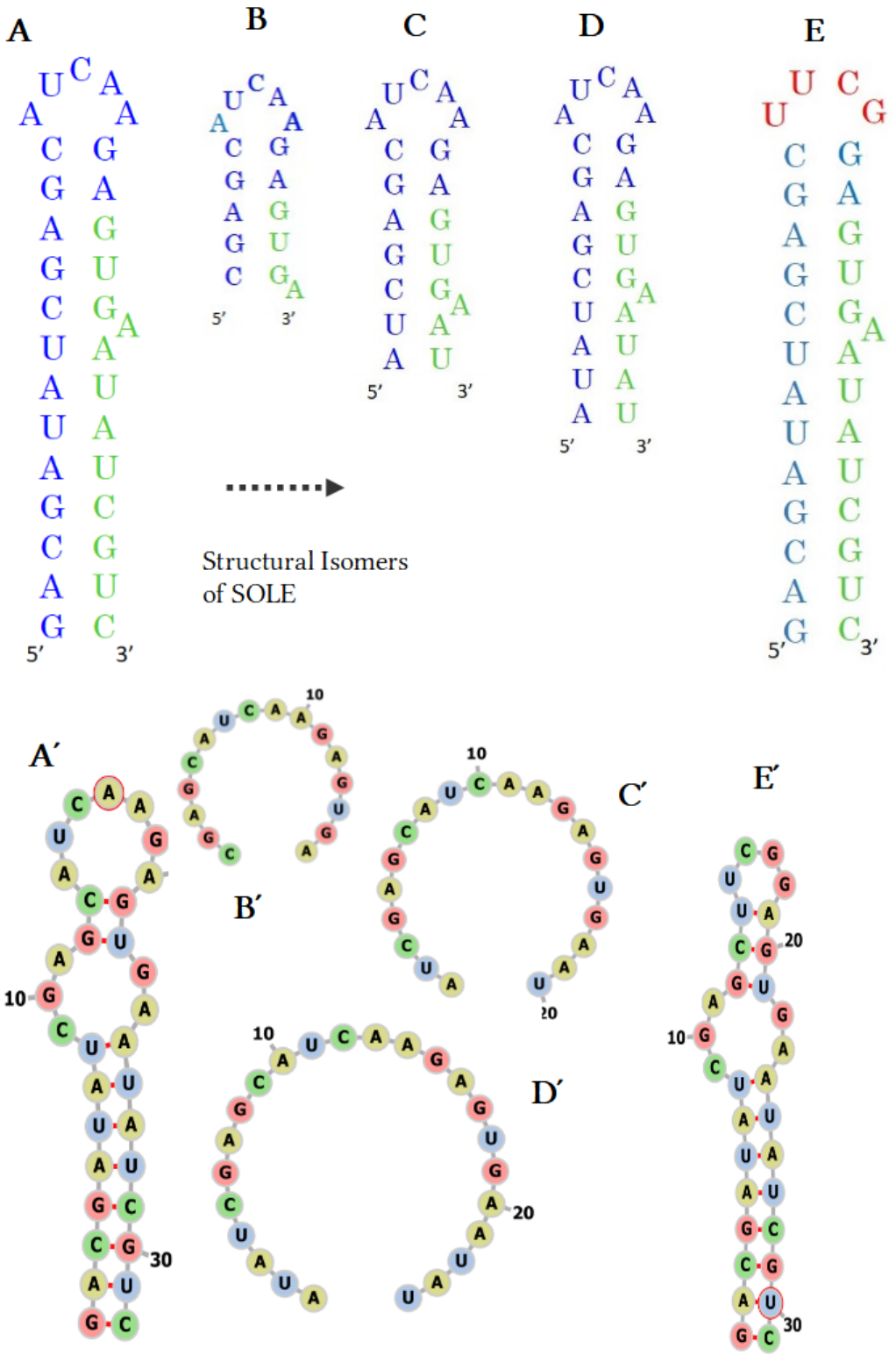


Figure: 4.8– A, B, C, D, and E are sequential representation of SOLE RNA, 16mer, 20mer, 24mer, and 31mer respectively. Where as, A′, B′, C′, D′, and E′ are their corresponding predicted secondary structures (ViennaRNA Web Services).

16mer: The shortest construct in this category, definitely shows the binding with the protein Pym[160], but the strength of the binding is very weak compared to Oskar SOLE RNA. Weak binding can be observed both in the EMSA, as well as in the NMR titration experiments. In figure 4.11, gel lanes from **a** to **d** are from the 16mer RNA, with the protein in the ratios of 0,1,2,5. As the binding is very weak, one cannot observe the band for the complex. But RNA band is weakening with a concentration of the protein that confirms the binding. The next step is to observe through the NMR titration experiments. ^{15}N-^{1}H HSQC experiments are recorded at 293K, for 50uM Pym[160], with the RNA gradient in the ratios of 1:0,1,2,5. In figures 4.9 and 4.10, an expanded region of the ^{15}N-^{1}H HSQC spectrum (1:2 -Pym[160]:16mer) shows weak binding through chemical shift perturbation of amide resonances (V38, A55, A98) for the 16mer-Pym[160] complex.

20mer: The construct with distal loop and medial stem-loop without a major part of the proximal stem, is having binding strength little stronger than that of 16mer, but when compared to Oskar SOLE the strength of binding is weak. EMSA shows the band for the complex 20mer-Pym[160] complex. ^{15}N-^{1}H HSQC spectrum (1:2, Pym[160]:20mer) shows the weak binding observed through the chemical shift perturbation (figure). The amide resonances (V38, A55, A98) are monitored for each RNA for the comparative observation.

24mer: The isomer which is a shorter construct of *Oskar* SOLE, shows an interaction pattern similar but stronger to 20mer, with respect to both chemical shift perturbation and strength of interaction. EMSA shows a definite band for the 24mer-Pym[160] complex. Adding four nucleotides from the proximal stem enhances the strength of the binding. This suggests that both the medial stem loop and proximal stem are actively participating in the interaction with the protein Pym[160]. In the image of ^{5}N-^{1}H HSQC spectrum, the resonances (V38, A55, A98) are depicted.

31mer: This construct, is devoid of the upper pentanucleotide distal loop. The interaction pattern is similar to *Oskar* SOLE. But the complex of 31mer-Pym[160] is more dynamic structurally, this may be due to the lack of a distal loop. The dynamic nature of the complex can be seen both in EMSA and as well as in the NMR spectrum. EMSA doesn't show a sharp band but instead, the broader fuzzy band can be seen. In the NMR spectrum, Figure 4.10, where few resonances are broader beyond observance. That may be due to either the dynamic nature of the complex or the structural heterogeneity. Here, we can conclude that the Distal loop does not involve directly in the interaction

process. Rather facilitate the binding indirectly by stabilizing or favoring the structure of the RNA-protein complex.

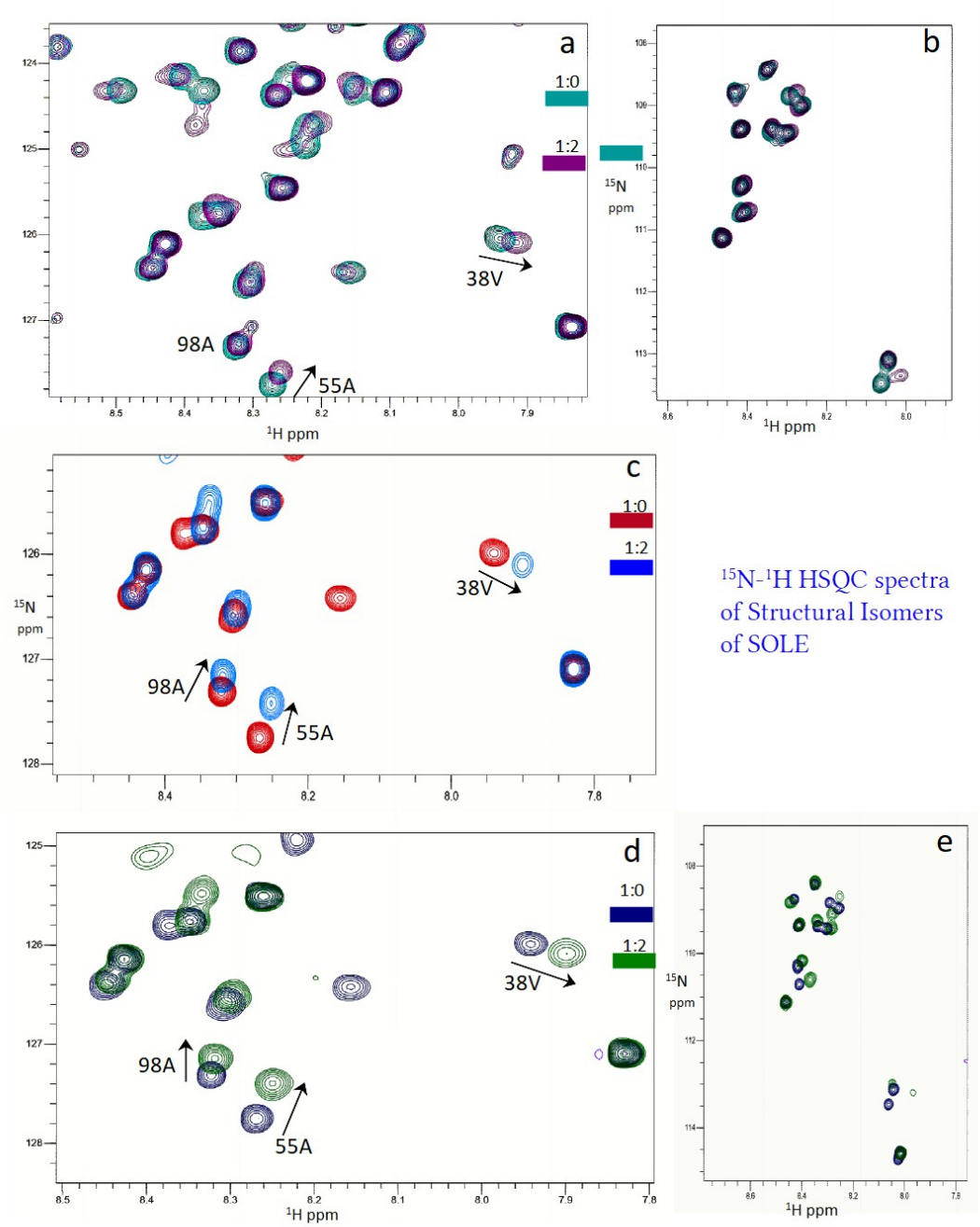


Figure: 4.9– ^{15}N-^{1}H HSQC spectrum of the protein, with the RNA bound. **a)** and **b)** are from the 16mer SOLE RNA, **c)** is from 20mer bound **d)** and **e)** are from the 24mer bound protein.

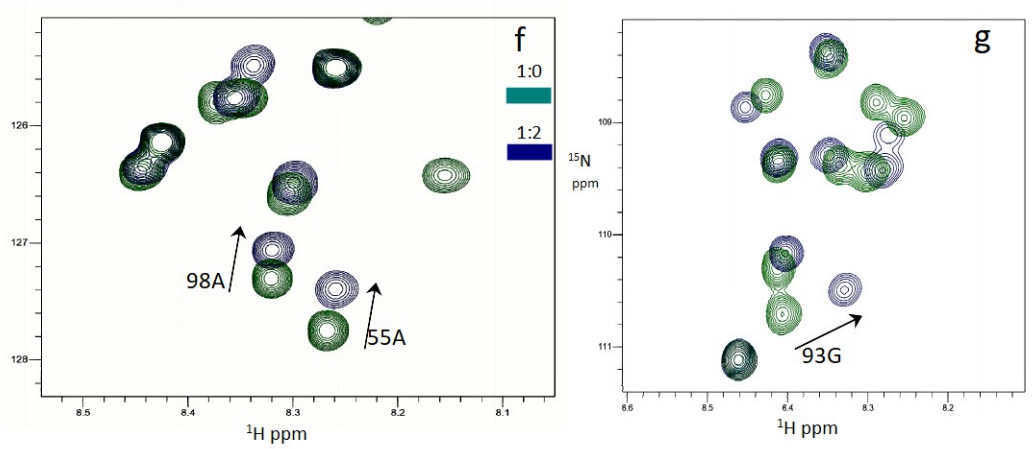


Figure: 4.10: ^{15}N-^{1}H HSQC spectrum of the protein, with the 31mer RNA bound. f) and g) are different parts of the spectrum showing the chemical shift perturbation.

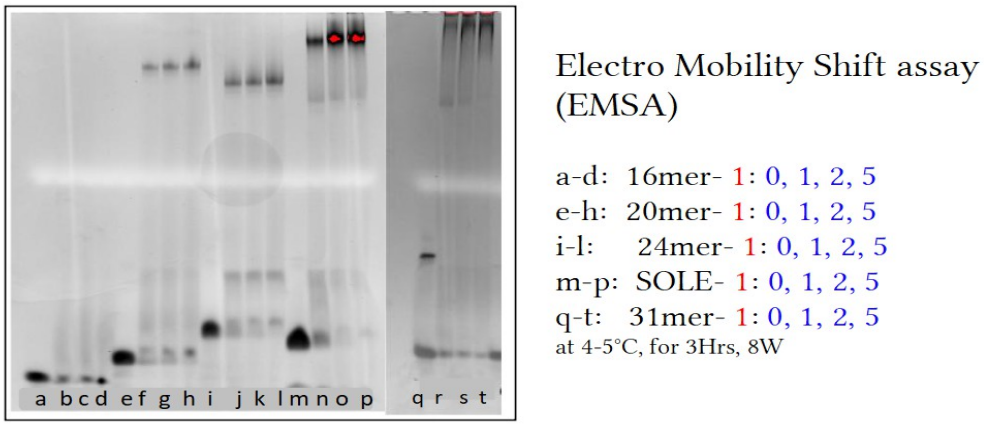


Figure: 4.11: EMSA gel images shows free RNA and RNA-Pym[160] complex in the different lanes.

After the analysis of the first set of structural based isomers, we can briefly conclude that that distal loop does not directly participate in the interaction, but definitely assists the binding as a secondary factor by structurally stabilizing the binding pockets to facilitate the complex formation. 16mer and 20mer show that the medial stem loop binds to the protein Pym[160], even though binding is weak. The involvement of proximal stem in the binding is very strong, as the incorporation of PS nucleotides shows strong binding and chemical shift perturbation results in the NMR and EMSA. However, at this point, we could not conclude anything about the binding pattern of the MSL and PS region. Hence further dissection was employed with respect to the MSL and PS

region of the SOLE. The detailed study of the interaction pattern of MSL and PS regions has been discussed in the next two sections.

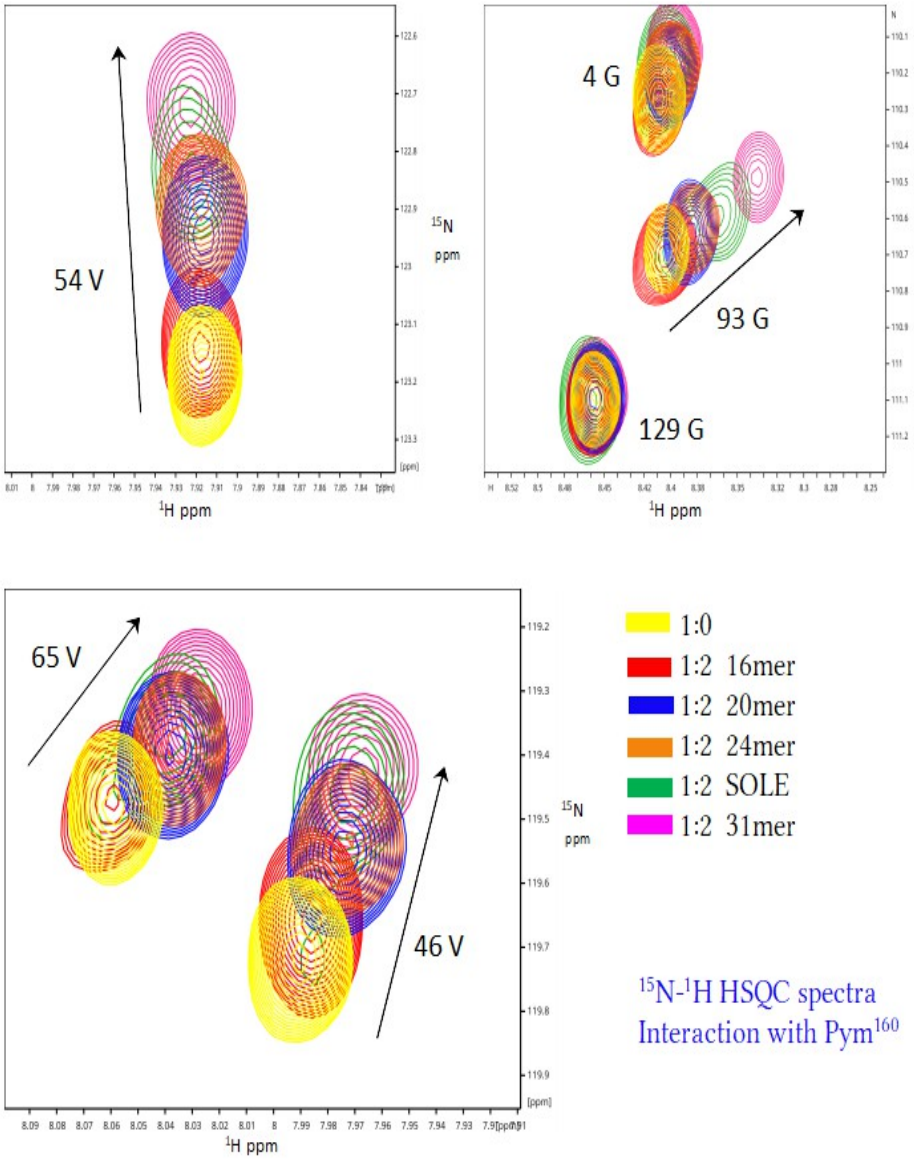


Figure: 4.12: Comparative observation of the different ^{15}N-^{1}H HSQC spectra for protein Pym[160,] bound to 16mer SOLE, 20mer SOLE, 24mer SOLE, 31mer RNA, and *Oskar* SOLE RNA.

We can observe the increase in the binding strength of the RNA, with the incorporation of proximal stem nucleotide. Hence, confirms the direct binding involvement of proximal stem nucleotides. Even though binding strength is weak, the MSL chunk also shows direct binding, which gets stronger with the incorporation of proximal stem nucleotides.

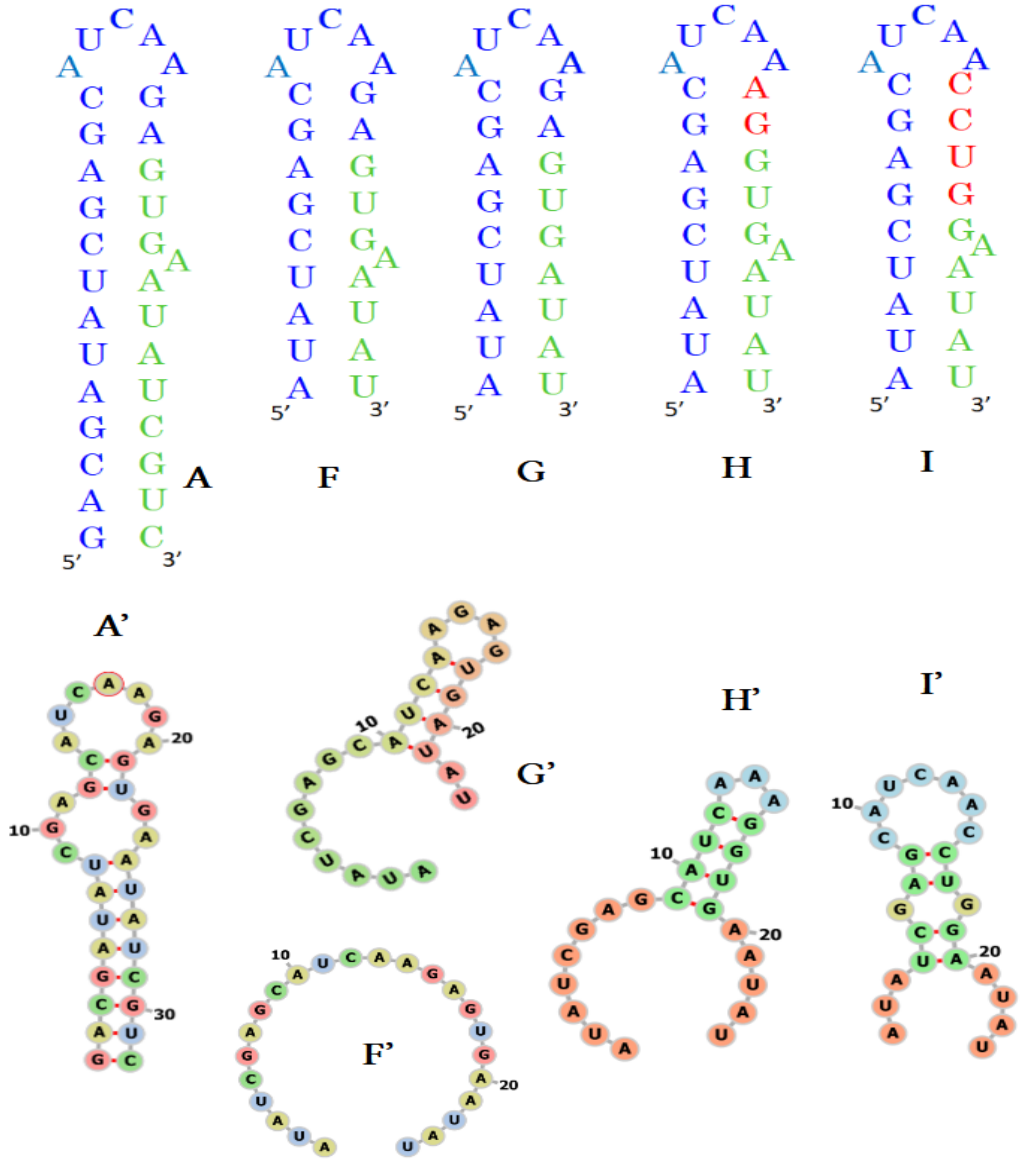
4.3.2 Medial Stem Loop isomers

The Medial stem loop region is very dynamic in nature. MSL region embodies the important motif of Oskar mRNA, that is the exon-exon junction point at the 18-19 nucleotides of SOLE, where the exon-1 and exon-2 are ligated together. Functionally, the exon junction point is very important, as the first intron splicing leads to the deposition of the exon junction complex on pre-mRNA. This is an important and prerequisite event prior to the localization as it facilitates the process of localization.
So, here the dissection involves two modes. The first is with the exon-exon junction point isomers (sequence modified) and the second one is medial stem loop isomers. Because, it is important to investigate, whether the exon – exon junction point has any sequence-specific interaction motifs that bind to the protein or not. Also, we already know that the MSL region involves in the interaction with the protein Pym[160]. So, the dissection process continued in order to identify the specific binding pattern, by making sequential isomers of the MSL region.

a) MSL exon junction point isomers:

Here, we have three isomers, Ejpt-24mer, Alter-24mer, and DelA-23mer. Individual isomers were analyzed through EMSA and NMR experiments. The interaction strength and pattern have been then compared with the 24mer and with SOLE RNA. Because all these isomers are of 24 and 23 nucleotides and technically isomers of 24mer. So, the binding strength should be of a similar value to the 24mer.

Ejpt-24mer: This exon junction point isomer. The 18^{th} and 19^{th} nucleotides of 24mer RNA have been replaced from G to A to A to G. As this is the exon – exon ligation point, at which exon junction complex is deposited on *Oskar* mRNA. At first Ejpt 24mer is passed through the EMSA gel, with the protein in the ratios of 0, 1, 2, 5. The first lane in the gel should have only RNA, the second lane would contain protein-RNA, in the ratios 1:1, and so on. EMSA result shows the interaction pattern same as that of 24mer. 24Mer and Ejpt-24mer have differences of only 2 nucleotides. So, EMSA results depict that this change of two nucleotides does not make any difference in the strength of the binding. Hence, exon junction point nucleotides do not themselves specifically contribute to the interaction, but instead, the whole MSL region involves.
NMR spectrum shows slight chemical shift perturbation in the specific nucleotides when compared to 24mer. This is due to the change in the nucleotides, which makes a slight difference in the secondary structure of the RNA (predicted structure). So, the RNA-protein complex also gets modified slightly.



SOLE: MSL exon junction point isomers

Figure: 4.13– A, F, G, H, and I are sequential representation of SOLE RNA, 24mer, delA, Ejpt-24mer, and Alter-24mer respectively. Where as, A′, F′, G′, H′, and I′ are their corresponding predicted secondary structures (ViennaRNA Web Service)

Alter-24mer: This is also an isomer of 24mer RNA, where 4 nucleotides from 18th nucleotides have been replaced from A, G, U, G to C, C, U, G. So, the secondary structure changes as the probability of these nucleotides to form a helical structure, when compared to the 24mer (figure). The EMSA results show the binding for the RNA -protein complex, but the intensity of the band is weaker compared to the 24mer. Further, continued NMR titration experiments show the chemical shift perturbation which is comparatively different from that of 24mer. But still, the isomer could be able to bind with the protein Pym[160]. From these two isomers, we can conclude that the MSL region binds to the protein in a partial sequence independent manner. The interaction

and the chemical shift perturbation is more concerned with the secondary structural elements.

DelA-23mer: This is 24mer RNA without 20th Adenine nucleotide. Hence, the name DelA-23mer. The EMSA and NMR experiments show similar chemical shift perturbation and interaction when compared to 24mer RNA.
With this, after analyzing MSL region isomers, we can conclude that 24mer and its isomers with respect to the MSL region show similar interaction strength, but with slightly different chemical shift perturbation in the NMR titration experiments. This may be due to the change in the secondary structural characters of each isomer. The predicted secondary structures of the RNAs also show structural differences, with a slight change in the nucleotide sequence. The Medial stem loop binds to the protein Pym[160] directly.

b) MSL sequential isomers:

In order to study the interaction behavior of the MSL region of SOLE without distal loop, we have made five sequential isomers of SOLE RNA, which are 24mer MSL, 22mer MSL, 22mer MSL alter I, 22merMSL alter II, and 16mer MSL. Out of which two isomers that are 22mer MSL and 24mer MSL have the same sequence of SOLE RNA. Whereas isomers 22mer MSL alter I and 22merMSL alter II have sequential modifications (Figure 4.14). 16mer MSL is missing two nucleotides from MSL. All these isomers are structurally helical, as predicted by the ViennaRNA Web Service.

So, in this section, we aim to investigate the direct involvement of the MSL region in the binding. From the EMSA (figure 4.15) we could see that the band for RNA-protein complex is weak for 22mer MSL alter I and 16mer MSL, in which few nucleotides are missing or changed from the MSL region.

Thus, in order to have the NMR clarification regarding the binding of the MSL region to the protein, we have compared 22mer MSL, 24mer MSL Ejpt and 24mer MSL alter. Out of three two isomers are missing the MSL nucleotides (figure 4.14). The amino acid G93 has been observed in the NMR titration spectra. Chemical shift perturbation of the NMR spectra for 24mer-Ejpt and 24mer Alter is less when compared to the 22mer MSL (similar to SOLE). Hence, we can broadly say that the sequence of the Medial stem loop (MSL) region is partially involved in the binding as they could probably change the structure of the RNA unbound.

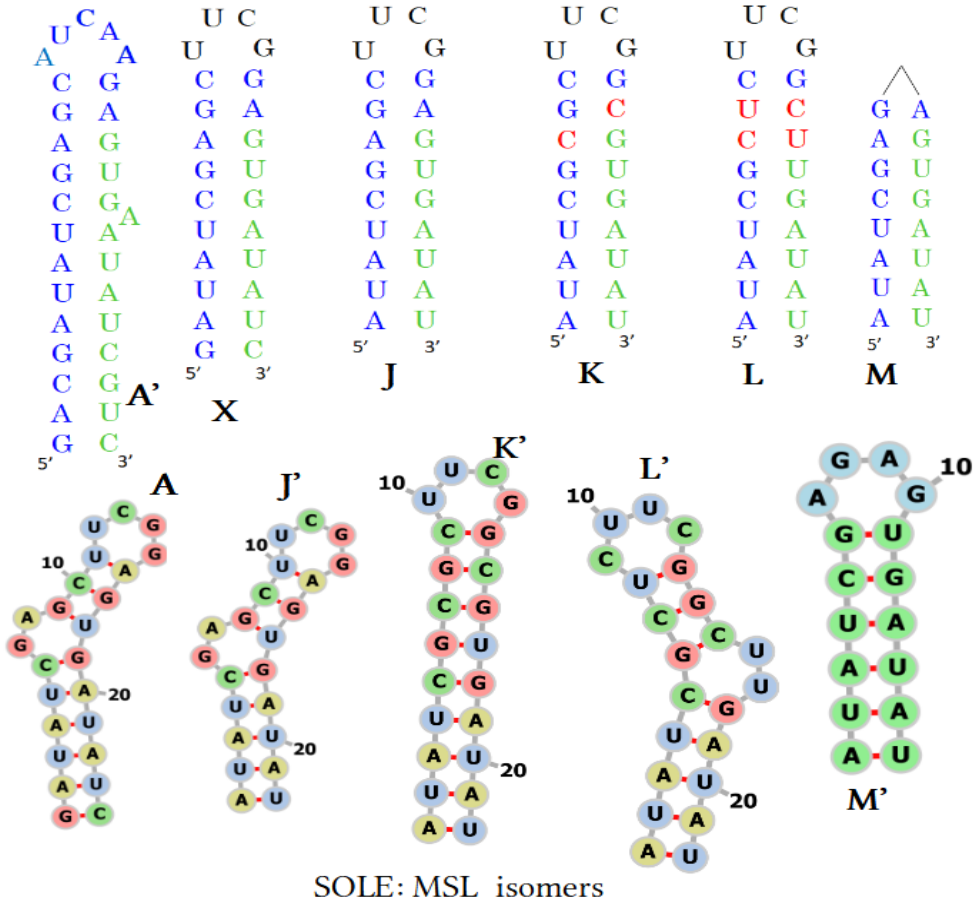


Figure: 4.14– A, X, J, L, and M are sequential representation of SOLE RNA, 24mer MSL, 22mer MSL, 22mer MSL alterI and 22mer MSL alterII, 16mer MSL respectively. Where as, A′, J′, K′, L′, and M′ are their corresponding predicted secondary structures (ViennaRNA Web Service)

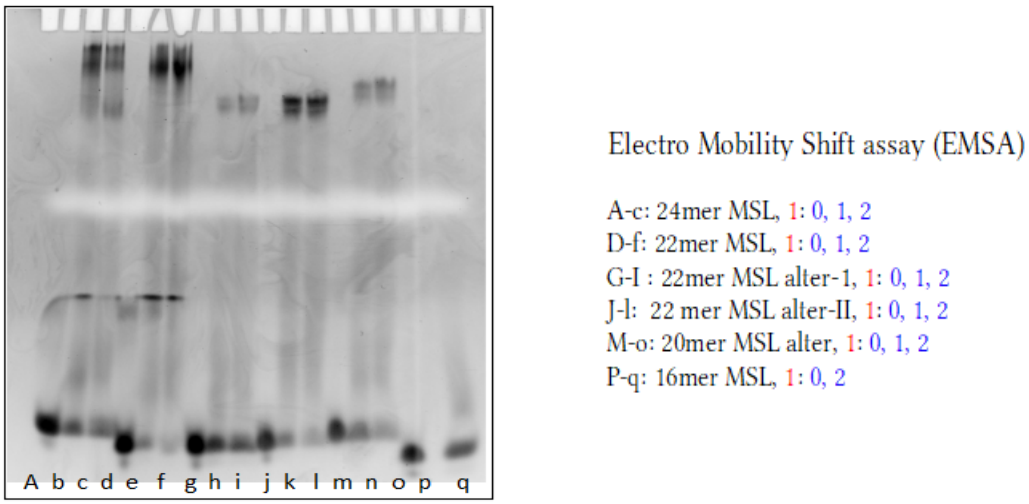


Figure: 4.15: EMSA gel images shows free RNA and RNA-Pym[160] complex in the different lanes.

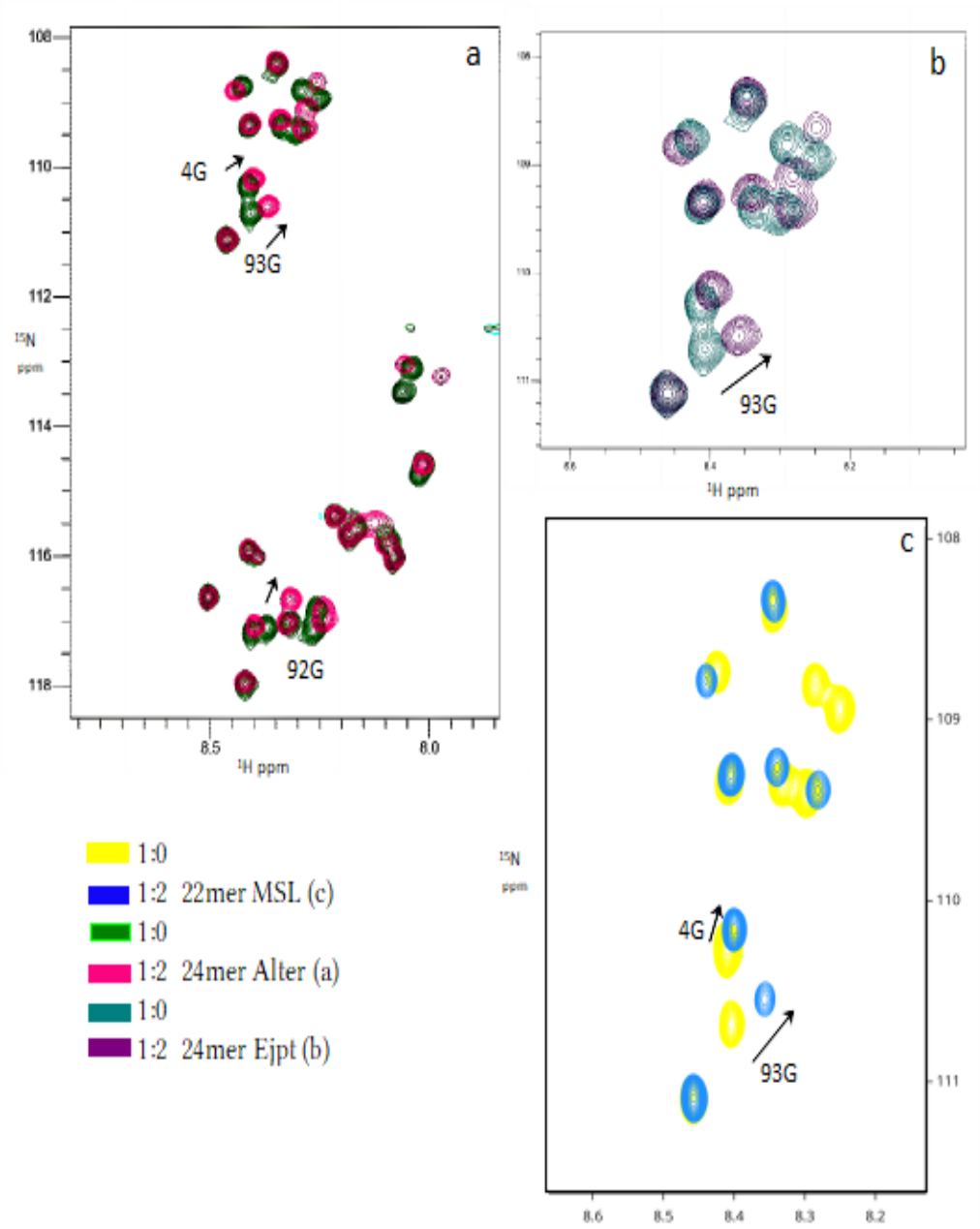


Figure: 4.16: Comparative observation of the different ^{15}N-^{1}H HSQC spectra for protein Pym[160,] bound to MSL isomers. We can observe the different binding strength of different isomers of MSL regions.

4.3.3 Proximal Loop isomers

From structural element isomers, it is confirmed that the proximal stem involves itself in the direct binding with the protein Pym[160]. Here, the dissection approach has been applied in order to investigate the sequential specificity for the binding.

Here, we are comparing 16mer, which is a proximal stem devoid isomer, containing only MSL and distal loop region. Whereas the 16mer MSL has MSL and proximal stem nucleotides. It is missing with a distal loop and a few MSL nucleotides. By comparing these two we can specifically identify the interaction of proximal stem with the protein. As we already have seen that 16mer shows very weak binding with the protein in the first part. But, 16mer MSL shows a stronger and specific binding pattern with protein Pym[160]. This is definitely due to the incorporation of the proximal stem nucleotides. Hence, it is again confirmed that the proximal stem is the primary and stronger binding motif, than the MSL region. Both together will make the protein-RNA complex much more stable. Further, the isomer 29mer has been made which has 6 nucleotides have replaced from the proximal stem, to have a stronger base-pairing helical element. 29Mer shows definitely interaction with the protein, but binding is comparatively weaker when compared to *Oskar* SOLE. Hence, the nucleotides of the proximal stem have specific interaction with the protein. This may be due to structural elements or sequential specificity or both.

From, this we can conclude that *Oskar* SOLE binds to the protein Pym[160], with its proximal stem region and medial stem loop region, whereas, the distal loop assists the binding by stable complex formation.

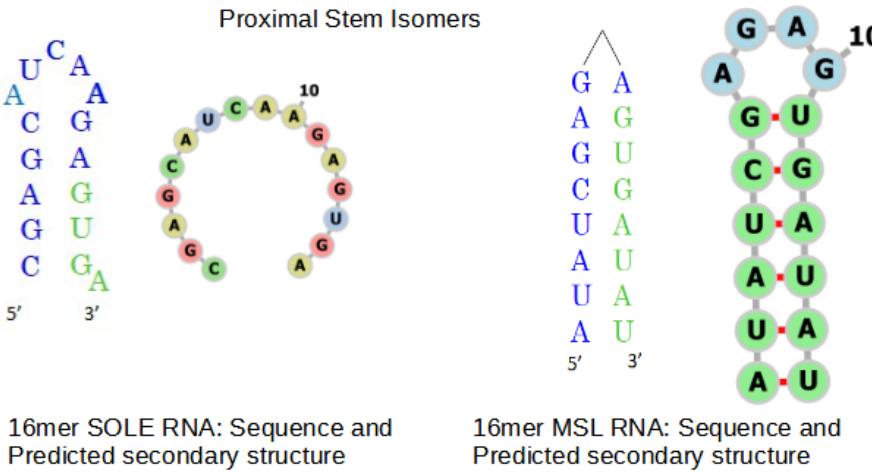
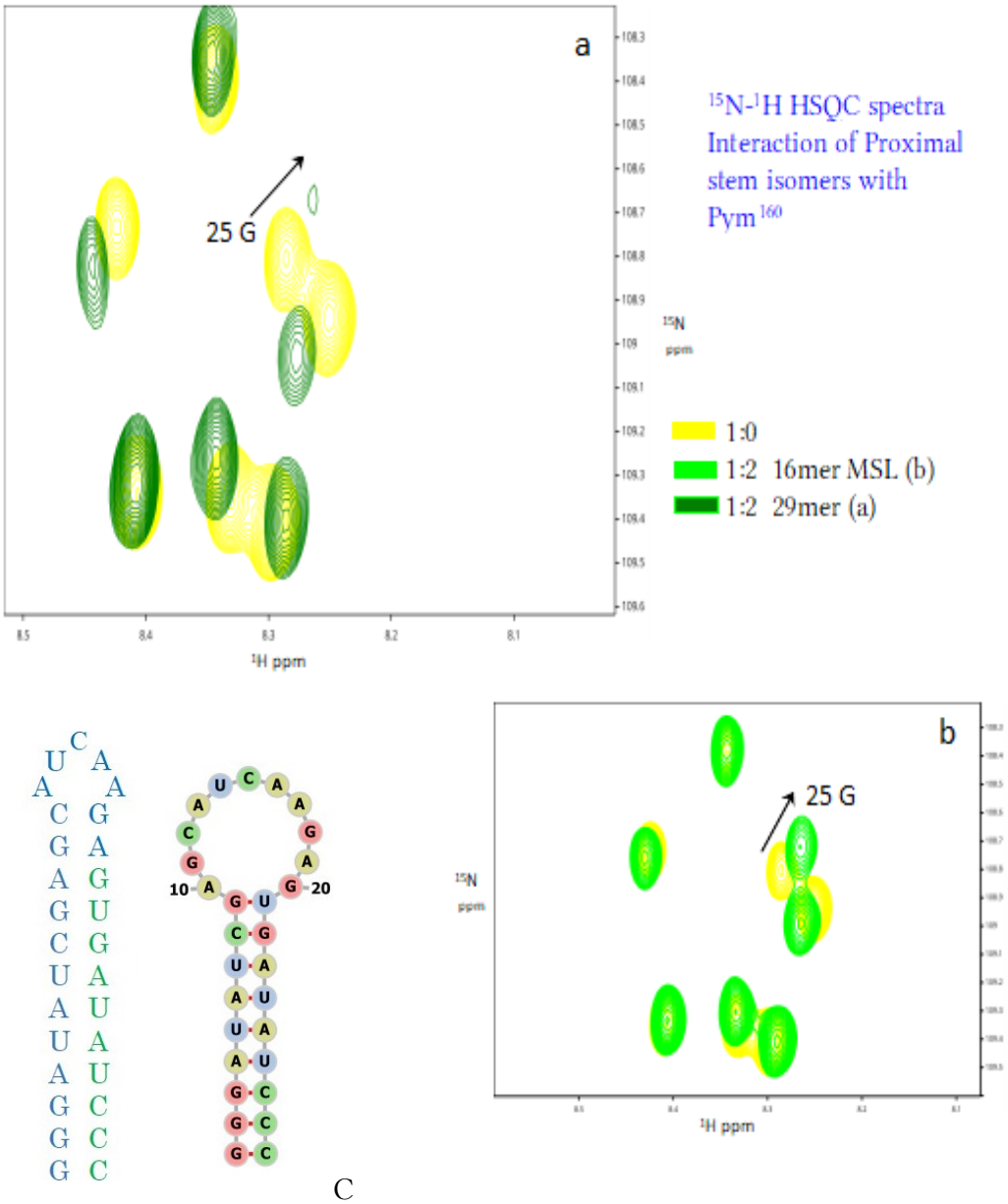


Figure: 4.17: Proximal Stem isomers

Figure: 4.18– Comparative observation of the different ^{15}N-^{1}H HSQC spectra for protein Pym[160,] bound to PS isomers. We can observe the 25G amide shows different interaction pattern with different RNAs (a) for 16mer, (b) for 29mer, (c) 29mer sequence and predicted structure from Vienna RNA webservice.



4.4 In the Pym^{160}-SOLE RNA complex

The binding of the SOLE RNA with the protein Pym^{160} has been studied with the various structural and sequential isomers of SOLE RNA. With this dissection method, we could able find out the binding map of RNA from a broad perspective. But still, we do not know the binding motifs of protein. So, in this regard, it is important to identify the interacting elements of the protein when it binds to RNA. We have made a number of isomers of RNA, and observed the protein in the NMR experiments, through the chemical shift perturbation of amide resonances. So, when comparing all the spectra of Pym^{160}, with every isomer together, we could able to see the specific binding pattern or interaction with respect to change in the SOLE isomers. We have compared every amide resonance, as to how it behaves with each RNA isomer. Some of the amides were very specific to the particular set of isomers. That means a different chunk of the protein interacting specifically with the different structural elements of the RNA. This is how we could able to identify the binding motifs of protein when it binds to *Oskar* SOLE.

4.4.1: Pym^{160} interaction with MSL isomers

^{15}N-^{1}H HSQC titration experiments of protein Pym^{160}, with respective RNAs are compared in order to investigate the mode of interaction of the protein with each amide resonances. Here, all the isomers have been compared for the study, but specific interactions were shown in the isomers of 24mer, Ejpt-24mer, Alter-24mer and 22merMSL. Here, 24mer is the normal construct of the SOLE RNA, whereas the other three are isomers of 24mer RNA, which have different nucleotides in the MSL region, they are sequentially modified constructs. So, by comparing the changes in the chemical shifts of amide resonances with the RNA-bound protein, one can identify the specific interaction pattern.

In the image, the overlay of ^{15}N-^{1}H HSQC titration experiments of protein Pym^{160}, with 24mer, Ejpt-24mer, Alter-24mer and 22merMSL. The amide resonance of Serine$_{92}$ (S92) and Glycine$_{93}$ (G93) have shifted differently with different RNAs. The shift is drastic if the RNA has a helical secondary structure in the MSL region. If the MSL region does not possess a helical structure then the shift is less. (Here all the predicted structures from the ViennaWeb Services have been compared). As, in the predicted structure 24mer, is not helical. So, the resonances of these two amides are comparatively less when compare to SOLE RNA or Alter 24mer or 22mer MSL. Similar behavior has also

been shown by the S92 amide. They both are very sensitive to the structural property of the MSL region. Hence, the region of the protein must be interacting with the protein.
Alanine$_{98}$ (A98) shows little chemical shift perturbation confirms the interacting behavior of S92 and G93 amide, as they might be the reason for the A98 shift.
Similarly, Valine$_{54}$, Valine$_{55}$, Alanine$_{59}$ and Valine$_{61}$ (V54, V55, A59, V61) show chemical shift perturbation in the NMR spectra. The shift is different with respect to the changes in the MSL region.
Similarly, T83 amide also shows chemical shift perturbation with the changes in the nucleotides of the MSL region.
By observing all this we can conclude that the middle part of the protein, mostly the helical part and its nearer regions are interacting with the MSL region of the SOLE RNA.

4.4.2: Pym^{160} interaction with Proximal Stem isomers

SOLE RNA has another binding motif in the proximal stem region while interacting with the protein Pym^{160}. Here, we have compared the 16mer RNA and 16mer MSL RNA, in order to identify the binding motif of the protein which may bind to the RNA. 16Mer does not have proximal stem nucleotides. Hence comparing the interaction patterns of these two RNAs, we have observed that few amide resonances interact differently. Amide resonances of Isoleucine17, Aspartic acid$_{24}$, Glycine$_{25}$, and Threonine$_{26}$ (I17, D24, G25, T26) behaved specifically when interacting with these two isomers. 16Mer RNA did not show any interaction or negligible shift in these amide resonances. But, 16mer MSL bound protein show a definite binding pattern of these amino acids. As the very distinct chemical shift perturbation can be seen in the NMR spectra. Glycine$_{25}$ amide broadens in the SOLE RNA, shows a clear shift in 16mer MSL, defines the binding of the amide with the proximal stem region. Similarly, Isoleucine$_{17}$ shows a clear shift when bound to the 16mer MSL, which is similar in the SOLE RNA as well. But failed to shift in the 16mer RNA, which does not possess any proximal stem nucleotides.
By comparing and analyzing all these results, we can conclude that the N-terminal region of the protein binds to the proximal stem nucleotides of the protein.
With these, we can surely say that the protein Pym^{160} and SOLE RNA have two binding motifs each when interacting together. So, because of the dynamics of the complex structure, the NMR spectrum broadens beyond observation while looking through the labeled RNA. This also may be due to the fact that protein itself is unfolded. So, the dynamic contribution of the complex on the protein side may not be affecting much for

the protein as it is already structurally very dynamic. Hence, the complex can be studied through the labeled protein NMR experiments.

The interaction pattern of Pym[160] with the different RNAs can be monitored through a chemical perturbation graph. So, one can clearly see that the N-terminal and middle helical part of the protein are mainly interacting with the RNA. Our results also show the same. The binding motifs of the protein lie in the N-terminal and middle helical part of the protein.

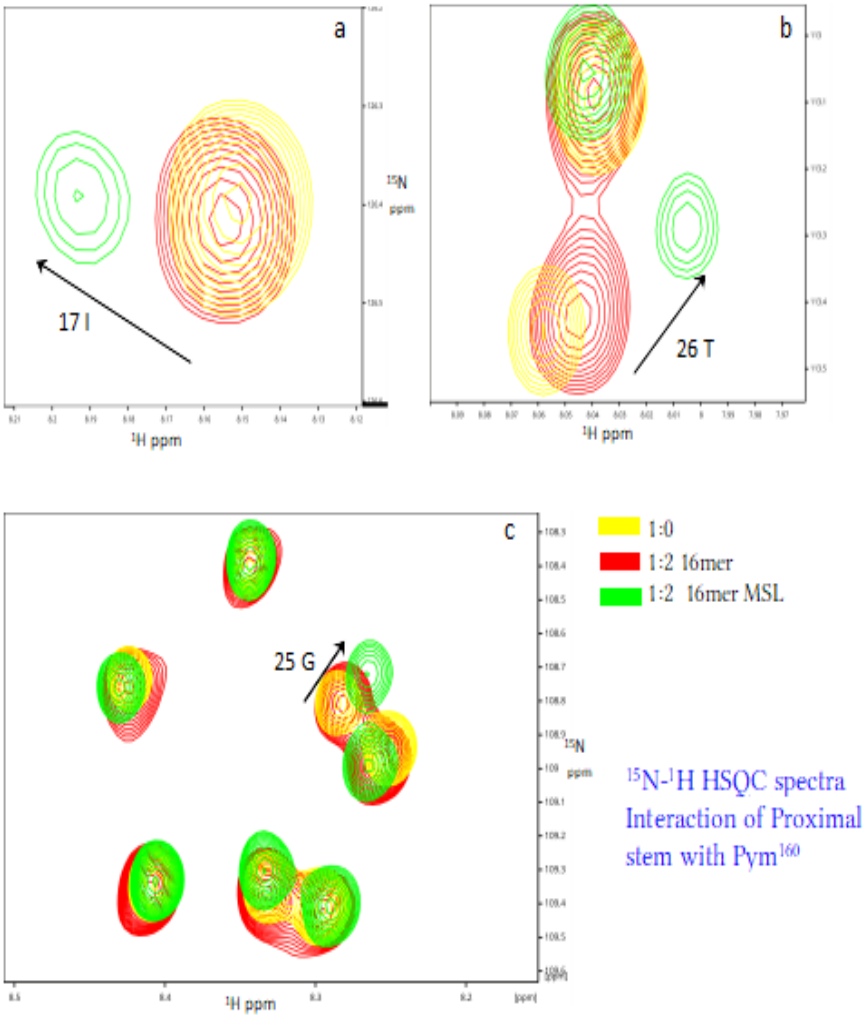


Figure: 4.19– Comparative observation of the different ^{15}N-^{1}H HSQC spectra for protein Pym[160,] bound to PS isomers. We can observe the 25G, 26T, 17I amides show different interaction patterns with different RNAs. This reveals the specific binding nature of these amino acids.

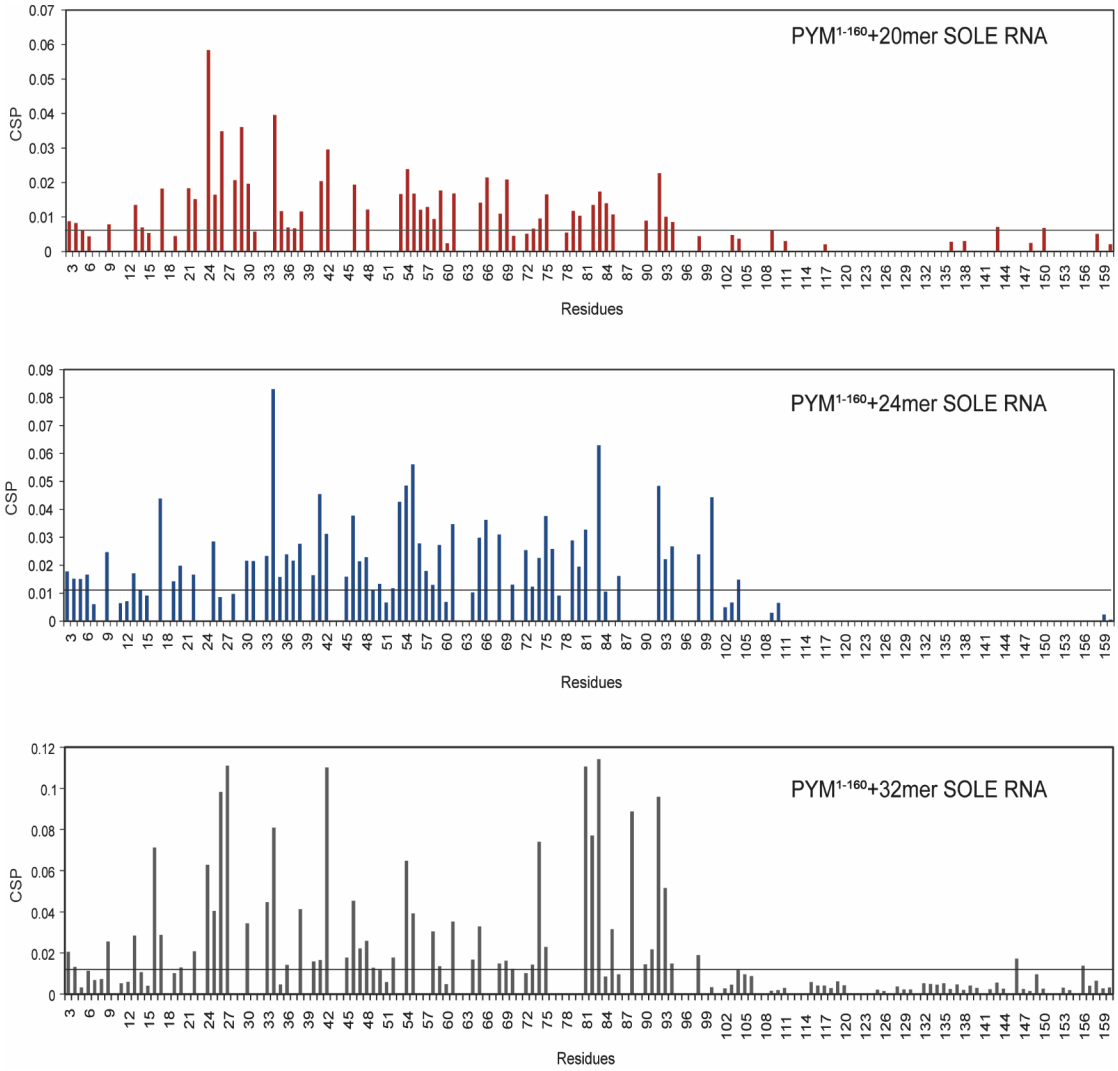


Figure: 4.20– Chemical Shift Perturbation graph for protein Pym^{160}, bound to 20mer SOLE, 24mer SOLE, and SOLE. From these, graph we get to know how differently each amide is interacting with the respective RNA. Also, we can notice that Protein does not show any binding pattern from its C-terminal end.(Image is from the submitted paper, Thanks to Dr. Deepshikha Verma, Dr. John Kirkpatrick, Prof. Dr. Teresa Carlomagno)

In figure 4.20, the chemical shift perturbation graph has been shown for the 20mer, 24mer and for the SOLE RNA, where the chemical shift deference if the bound and unbound protein Pym^{160} has been depicted. The graphs show that the protein binds to even the shorter construct of SOLE, which is 20mer RNA. Also, it is very much clear that the C-terminal of the protein does not bind to the RNA.

Further structure calculation, refining work and Pym^{160}-SOLE RNA Docking work has been done by Dr. Deepshikha Verma. I am very much thankful to Dr. Deepshikha Verma, Dr. John Kirkpatrick, Prof. Dr. Teresa Carlomagno for the data and the images.

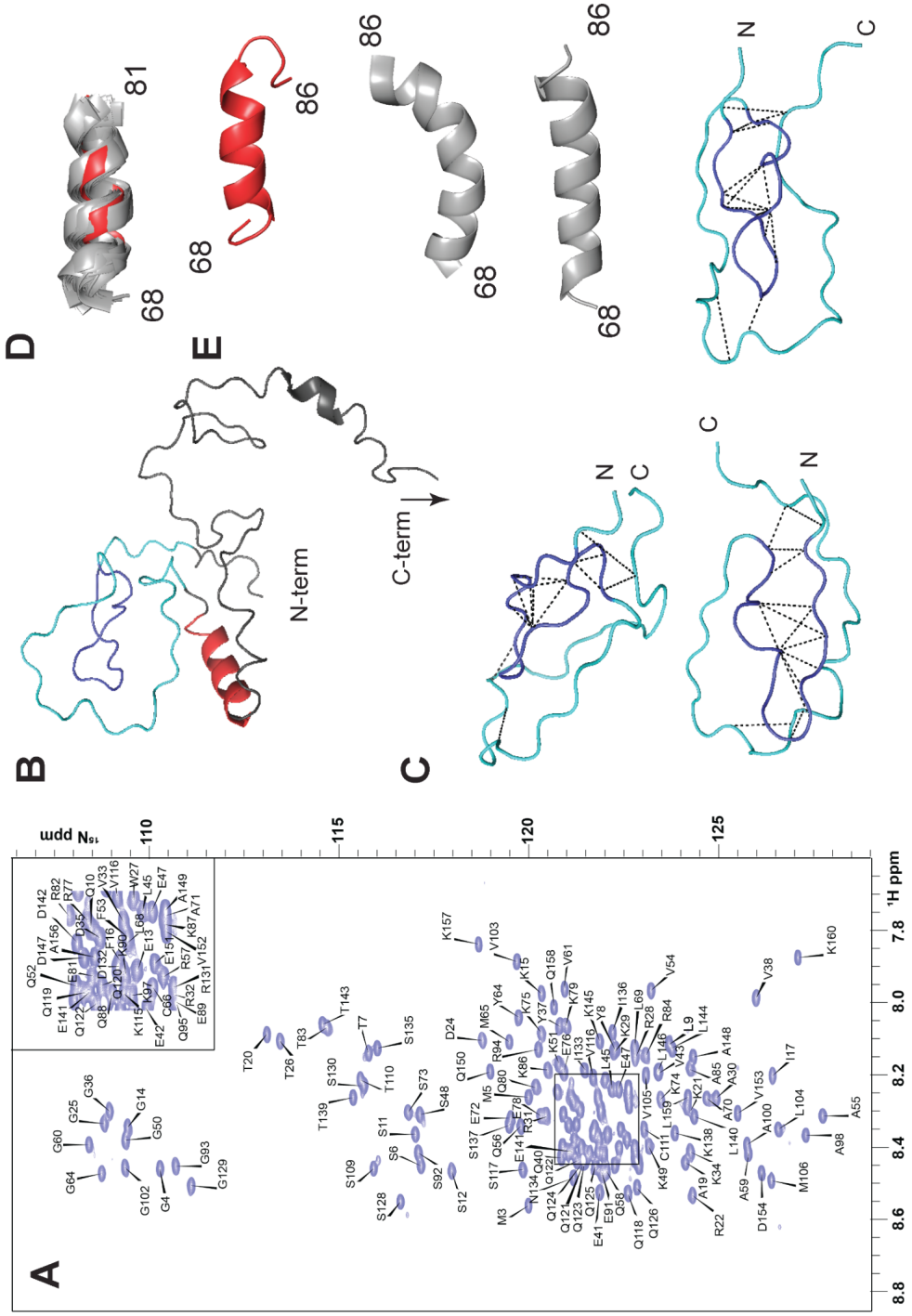


Figure 4.6: (A) ^{15}N-^{1}H HSQC spectrum of the protein depicting all the assigned amide resonances. B), C) D) E) are the individual elements of protein Pym[160]. (Image is from the submitted paper, Thanks to Dr. Deepshikha Verma, Dr. John Kirkpatrick, Prof. Dr. Teresa Carlomagno)

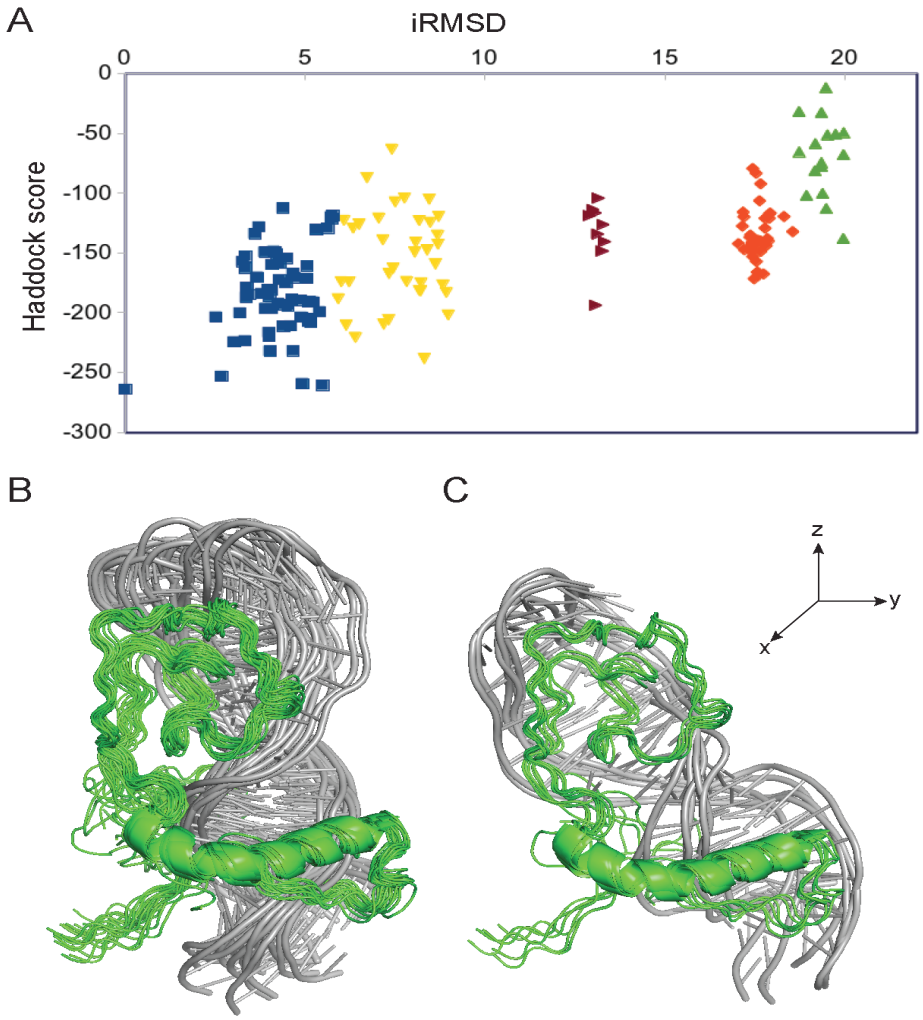


Figure 4.21: Models of the PYM$^{1-160}$–SOLE RNA complex obtained by HADDOCK. A. Plot of the HADDOCK score (y axis) versus the backbone iRMSD (protein–RNA interface RMSD) from the structure with the lowest score of 200 docked conformations of the PYM$^{1-160}$–SOLE RNA complex. The structures can be classified in five clusters, represented by five different symbols. Both clusters C1 and C2 contain complex structures with HADDOCK scores less than 200. B. Overlap of all complex structures of cluster C1 with score less than 200. The structures are aligned on the PYM protein aa 1–111. C. Overlap of all complex structures of cluster C2 with score less than 200. The structures are aligned on the PYM protein aa 1–111.
(Image is from the submitted paper, Thanks to Dr. Deepshikha Verma, Dr. John Kirkpatrick, Prof. Dr. Teresa Carlomagno)

The optimized structure of the protein shows a central helical part and a small helix at the C-terminal. Apart from these two structural elements, the protein remains to be structurally unfolded.

CHAPTER

5. Discussion and Outlook

The study of mRNA and mRNA localization has always been an important arena, in order to understand the different biological functions and cellular organization. The functional importance of mRNA localization has been evident from long ago. The study of mRNA and its localization mechanism, together with its binding partners are essential to understanding the biological problems. As they define life, in terms of cellular and sub-cellular mechanisms. Our study also involves one of the binding partners of the localization complex, that is Pym protein, that Pym^{160}, the shorter and the functionally important construct of the protein Pym. Exon Junction Complex and Pym are the common molecules involved in the many mRNA localization mechanisms and *Oskar* mRNA localization is one of them.

Pym^{160} is a structurally unfolded protein, with the general characteristic of an intrinsically disordered protein. It has a long helical structural element, i.e., α-helix in the middle part of the protein, along with both N-terminal and C-terminal ends are highly flexible and dynamic with the structurally unfolded random coil elements. Our structural studies show a few helical elements at the C-terminal end. The C-terminal part of the protein is not involved in the direct interaction with the SOLE RNA. However, it is a structurally very important part of the protein, as it stabilizes the ionic and hydrophobic interactions of the protein, so that protein could able to be a stable soluble protein. Without the C-terminal chunk of the protein, Pym was aggregated and cannot be purified as a soluble construct. So, the smaller constructs of the protein Pym, which are Pym^{110} and Pym^{135} are aggregated upon recombinant expression.

Pym^{160} is an RNA binding protein, that binds to the Exon Junction Complex as well[175,184,185]. Pym serves as the recyclization unit for Exon Junction Complex (EJC) during the process of localization. Once, the EJC bound mRNA reaches its target site, the protein Pym binds to the complex and releases Exon Junction Complex for the further localization of the remaining mRNAs. Pym would remain in the bound form with the mRNA, until the early stage of translation.

In this regard, our study unveils some important structural and binding features of Pym^{160}, which describes the broad range of binding property and structural flexibility of the protein.

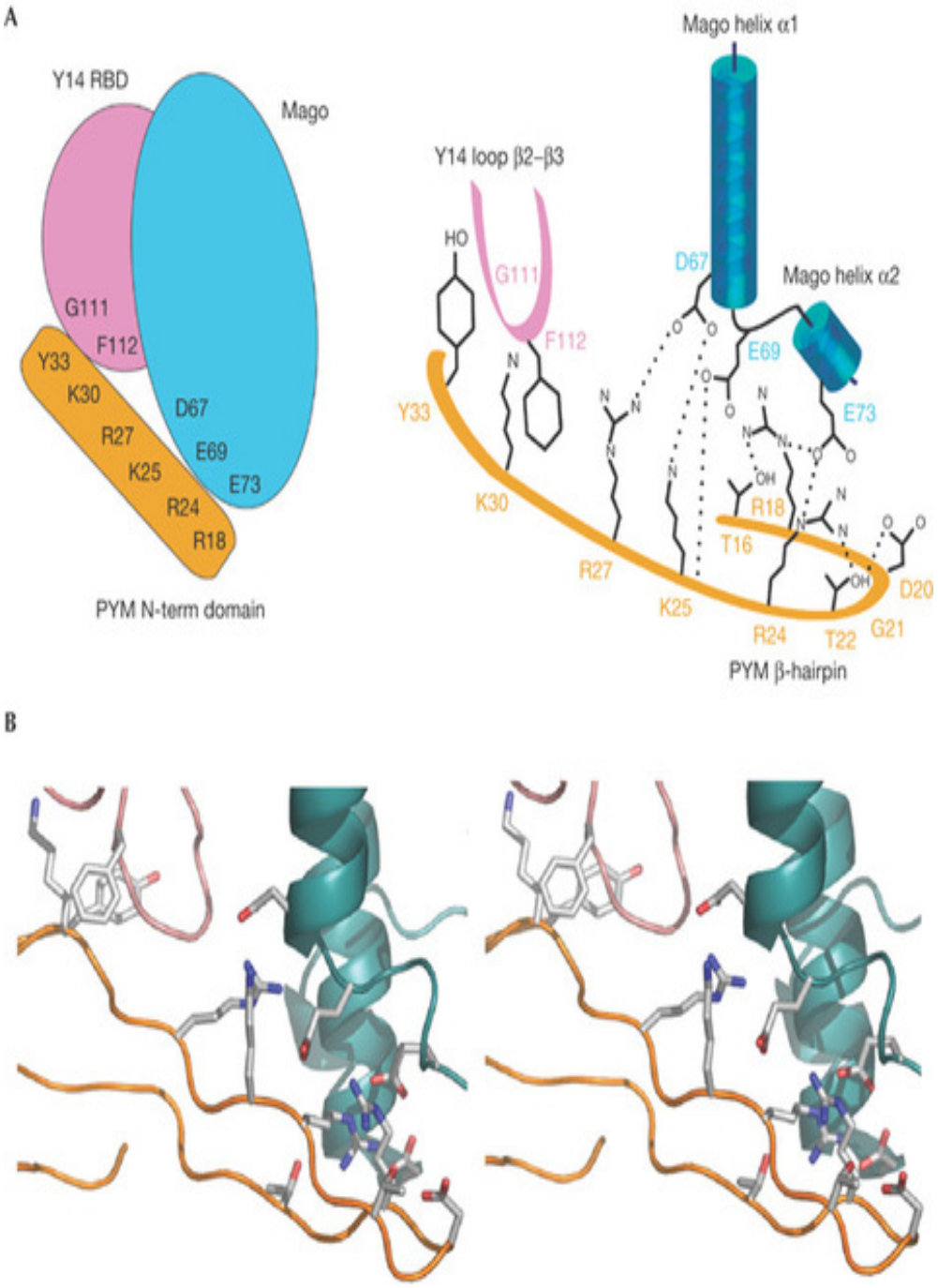


Figure 5.1: Pym shows interactions with Mago-Y14 hetero-dimer. The positively charged residues of the protein bind to the negatively charged residues of Mago at α1 and α2 helices, and through hydrophobic interactions with the β2-β3 loop of the Y14 protein. A) Schematic view B) Stereo view. (Image is from EMBO Reports (2004)5:304-310)

The crystallographic structural data of Pym-Y14-Mago[67], a multi-protein complex describes the binding motif of protein Pym which binds with the Exon Junction Complex core proteins, that are Y14 and Mago, a hetero-dimer. In figure 5.1, the binding map of all the proteins have been shown. Pym binds to the Y14 -Mago through hydrophobic and ionic interactions. The N-terminal of the protein from the residue R18 to Y33, the protein binds Mago via ionic interactions at α1 and α2 helices through its charged residues. While with Y14, Pym interacts through hydrophobic interactions via β2-β3 loop. In the image the N-terminal residues are clearly depicted which have shown interactions with Mago-Y14 hetero-dimer. So, the N-terminal part of the protein Pym[160], has binding motif with the exon junction complex proteins, that is Y14 and Mago.

Our results reveal that the N-terminal end of Pym[160] protein shows binding motif with the proximal stem of the SOLE RNA (Figure 5.2). The amino acid residues Isoleucine[17], Glycine[25] and Threonine[26] show significant change in the chemical shifts (CSP) in the bound form with the RNA, that embodies the proximal stem residues. Here in the image 16mer RNA does not contain any nucleotides from the proximal stem, show no perturbation in the chemical shifts, while the RNA 16mer MSL, which has proximal stem within shows clear and strong perturbation in the chemical shifts of the residues in the 16mer MSL bound-Pym[160].

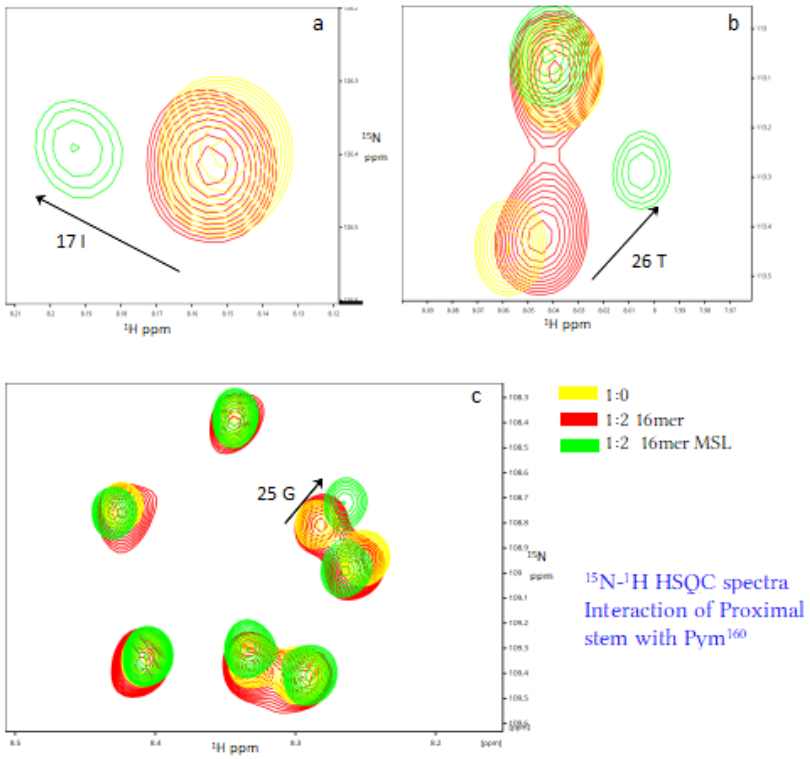


Figure 5.2: ^{15}N-^{1}H HSQC spectra overlay ed with bound and unbound Pym[160], Yellow spectrum is unbound Pym[160], Red spectrum is Pym[160] –16mer and Green spectrum is Pym[160] –16merMSL.

By comparing these two data we can clearly say that the N-terminal of the protein can bind to both Exon Junction Complex and to the SOLE RNA.
But the Exon Junction Complex binds to the protein Pym as well as to the mRNA, which means EJC deposits at the 20-24 nucleotides upstream of the exon-junction point (Intron spliced). This part is very much close to the proximal stem. So, with all this information, maybe during the localization process, Pym binds to the RNA with the helical motif and binds to the EJC with its N-terminal motif. After releasing the EJC, maybe the N-terminal end of the protein might interact with the mRNA as a competitive binding.

Also, Pym recycles the Exon junction complex at the early stage of translation. Moreover, some results have shown that Pym associates with ribosomes[69] and ribosome-bound Pym disassembles the EJC and remains bound to the RNA until the translation.
Adding to this, the overproduction of the Pym disrupts the localization process. This may be due to the fact that the N-terminal part of the protein competitively binds to the RNA, so overproduction of protein disrupts the EJC-mRNA association.

Well, with all these data, it is evident that Pym can bind to the Exon Junction Complex, it can also bind to the mRNAs and to the ribosomes as well. So, protein Pym has a broad range of functions and binding properties with many molecules. Thus, here one can conclude that Pym is an interacting partner for both RNAs and proteins.
The nature of this expansive binding capability of the protein Pym may be due to its structurally unfolded nature. As this is the important quality of an Intrinsically Disordered Protein, they do not possess any rigid three-dimensional structures, hence can bind easily to a wide range of molecules depending upon the functional requirement. In other words, the highly flexible non-structural feature is the reason behind this expansive binding nature. Our results explain the structural flexibility of the protein through NMR studies and through structural characterization. Our results show that Pym[160] is a structurally unfolded protein.

This binding nature of the protein justifies the fuzzy structural feature of the N-terminal region. Pym being very dynamic and multi-functional, could be able to bind SOLE RNA, EJC, and Ribosomes. So, the structural flexibility of the protein is a great advantage for the multi-functional activity.

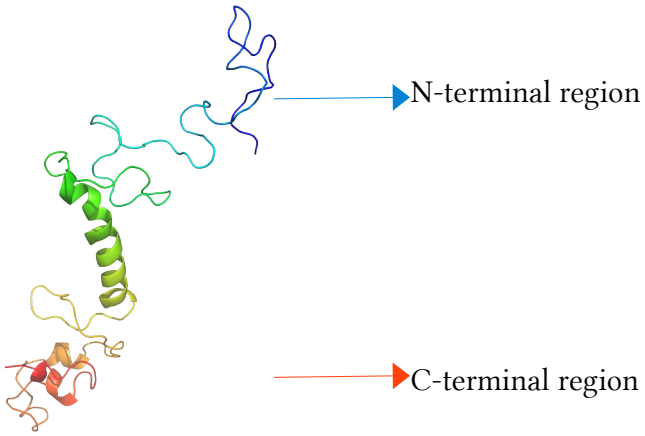


Figure 5.3: One of the optimized structures of Pym160, from the early stage of structure calculation

So, clearly, Pym has a characteristic feature of intrinsically disordered proteins (IDCs). Sometimes IDC in a bound form show induced structural elements due to complex formation. Pym is showing exactly similar behavior. Our structural studies of the protein in unbound form do not show any secondary structural elements in the N-terminal region. But X-ray crystallographic data shows beta-sheets in the bound N-terminal Pym protein.

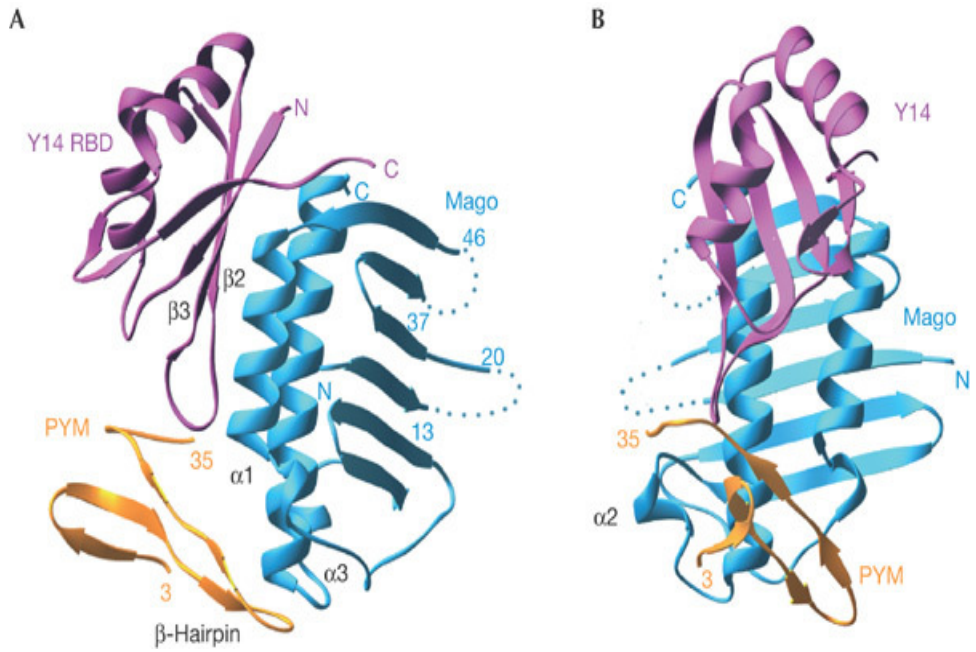


Figure 5.4: The structure of Drosophila melanogaster, Pym-Mago-Y14 structure. Here Pym shows the β-hairpin structure of the N-terminal region, in the bound form.

It is well known that structurally well-defined proteins address very specific functions in biological function. The same way being structurally unfolded nature of the protein may also be structurally so important that it could able to bind different RNA and protein complex. The flexibility of the protein may facilitate the broad range of binding properties to specific functional processes.
While studying Oskar RNA Pym interaction may not just shed light on this localization, but also may disclose the general strategy of the localization mechanisms.
Because, not just Oskar mRNA but many other mRNAs involved in the similar type of localization processes, where Exon Junction Complex and Pym protein are involved. mRNAs like *gurken, bicoid* also need exon junction complex for the localization process.
So, concluding the study of Oskar mRNA localization, the study of EJC and protein Pym can shed light on the arena of localization mechanism in general.

5.1 Future directions

1) Study of the *Oskar* SOLE-Pym unit, together with the exon junction complex, may disclose exact binding map of protein Pym's binding nature with respect to exon junction complex proteins, RNAs, and ribosomes.

2) The helical structure of the middle chunk of protein Pym, is an RNA binding motif. So, other mRNAs like *gurken, bicoid* might show similar interactions with the protein Pym. It would be interesting and vital to know the interactions of other mRNAs with protein Pym.

3) Pym is binding with the ribosomes and with the exon junction complex, from its N-terminal end. This suggests the competitive binding mechanism between the protein complex and the RNAs. This study would also shed light on the recycling of the exon junction complex and its interaction mechanism.

NMR Chemical Shift Table

for the protein Pym[160]

#	NmrResidue	Name	Shift
1	A.143.THR	H	8.06136983723719
2	A.143.THR	N	114.56274498323
3	A.26.THR	H	8.10798345916811
4	A.26.THR	N	113.446505312398
5	A.150.GLN	H	8.19158921301642
6	A.150.GLN	N	118.984461413797
7	A.6.SER	H	8.45203720091927
8	A.6.SER	N	117.166835521415
9	A.48.SER	H	8.30966469450592
10	A.48.SER	N	117.125313560722
11	A.35.ASP	H	8.29698196939937
12	A.35.ASP	N	121.349776575593
13	A.75.LYS	H	8.08736067240472
14	A.75.LYS	N	120.334937377093
15	A.20.THR	H	8.08830860947796
16	A.20.THR	N	113.098453751187
17	A.9.LEU	H	8.1149564675147
18	A.9.LEU	N	123.681993609067
19	A.24.ASP	H	8.10575199684222
20	A.24.ASP	N	118.784401385972
21	A.82.ARG	H	8.24842562989312
22	A.82.ARG	N	120.763615215132
23	A.92.SER	H	8.41677385063221
24	A.92.SER	N	117.089024254376
25	A.93.GLY	H	8.45403370610118
26	A.93.GLY	N	110.700272447697
27	A.55.ALA	H	8.31368635336513
28	A.55.ALA	N	127.740348011374
29	A.88.GLN	H	8.41738591572755
30	A.88.GLN	N	121.59225716038
31	A.11.SER	H	8.36510166418792
32	A.11.SER	N	117.011161737035
33	A.28.ARG	H	8.14942471158625
34	A.28.ARG	N	122.801431434044
35	A.94.ARG	H	8.13041461986654
36	A.94.ARG	N	120.25446879442
37	A.120.GLN	H	8.40030830017213
38	A.120.GLN	N	121.384116204253
39	A.142.ASP	H	8.31002599295438
40	A.142.ASP	N	120.91615854142
41	A.4.GLY	H	8.45992394436887
42	A.4.GLY	N	110.270780108789
43	A.147.ASP	H	8.34065990735462
44	A.147.ASP	N	121.053796322186

45	A.73.SER	H	8.30445984073042
46	A.73.SER	N	116.822439401678
47	A.98.ALA	H	8.36748426730127
48	A.98.ALA	N	127.300392309414
51	A.160.LYS	H	7.87610424323294
52	A.160.LYS	N	127.090687927084
53	A.116.VAL	H	8.20626241815883
54	A.116.VAL	N	121.684407112118
55	A.14.GLY	H	8.34624606385755
56	A.14.GLY	N	109.422533475448
57	A.110.THR	H	8.21386309851903
58	A.110.THR	N	115.56031787337
59	A.128.SER	H	8.55127090323666
60	A.128.SER	N	116.620732672899
61	A.30.ALA	H	8.26441543176603
62	A.30.ALA	N	124.931331762383
63	A.5.MET	H	8.26242850244265
64	A.5.MET	N	119.999402964481
65	A.104.LEU	H	8.35127442558476
66	A.104.LEU	N	126.57316855256
67	A.10.GLN	H	8.27656516338396
68	A.10.GLN	N	121.185738997871
69	A.102.GLY	H	8.45759108051256
70	A.102.GLY	N	109.356859183877
73	A.72.GLU	H	8.32201300895739
74	A.72.GLU	N	119.529273839345
75	A.135.SER	H	8.12733319433909
76	A.135.SER	N	116.007442289322
77	A.79.LYS	H	8.06737801419823
78	A.79.LYS	N	120.998773121081
79	A.3.MET	H	8.56221647376884
80	A.3.MET	N	119.993398770212
83	A.50.GLY	H	8.38154203560623
84	A.50.GLY	N	109.368217038813
85	A.42.GLU	H	8.44957734803405
86	A.42.GLU	N	122.253867408011
87	A.77.ARG	H	8.26053426048605
88	A.77.ARG	N	121.106125803491
89	A.109.SER	H	8.4593330548194
90	A.109.SER	N	115.913033701971
92	A.122.GLN	N	121.28030602039
93	A.37.TYR	H	8.0636964560707
94	A.37.TYR	N	120.819286816348
95	A.117.SER	H	8.46302678354155
96	A.117.SER	N	119.844296999502
97	A.106.MET	H	8.49312210086611
98	A.106.MET	N	126.388779375521
99	A.36.GLY	H	8.30124127931747

100	A.36.GLY	N	108.948346518549
101	A.158.GLN	H	8.01294589832808
102	A.158.GLN	N	120.667716573437
103	A.7.THR	H	8.1396726057061
104	A.7.THR	N	115.782952300593
105	A.56.GLN	H	8.34337932432591
106	A.56.GLN	N	119.792927357667
107	A.154.ASP	H	8.47142603384745
108	A.154.ASP	N	126.128072285993
109	A.87.LYS	H	8.26288564886512
110	A.87.LYS	N	122.616128333717
111	A.25.GLY	H	8.33598322313392
112	A.25.GLY	N	108.812133304549
113	A.133.ILE	H	8.18716675464157
114	A.133.ILE	N	121.475136941011
115	A.130.SER	H	8.23032649840091
116	A.130.SER	N	115.654751737563
117	A.46.TYR	H	8.04067672896458
118	A.46.TYR	N	119.73923327121
119	A.38.VAL	H	7.98638926163078
120	A.38.VAL	N	125.988379270147
121	A.40.GLN	H	8.42941602047254
122	A.40.GLN	N	120.856398708674
123	A.64.GLY	H	8.4737745840075
124	A.64.GLY	N	108.745899875119
125	A.139.THR	H	8.26313813312543
126	A.139.THR	N	115.370229636212
127	A.103.VAL	H	7.88748236183476
128	A.103.VAL	N	119.688418612784
129	A.29.LYS	H	8.13782804544974
130	A.29.LYS	N	122.239474947646
131	A.59.ALA	H	8.41862239493339
132	A.59.ALA	N	125.782228414756
133	A.60.GLY	H	8.39472694188904
134	A.60.GLY	N	108.410916123002
135	A.68.LEU	H	8.30999235244965
136	A.68.LEU	N	121.899411073542
137	A.83.THR	H	8.07023215037608
138	A.83.THR	N	114.68587118747
139	A.76.GLU	H	8.1936260853399
140	A.76.GLU	N	120.930489927773
141	A.47.GLU	H	8.24005051607308
142	A.47.GLU	N	122.358155216618
143	A.137.SER	H	8.33988234560849
144	A.137.SER	N	119.480802562103
145	A.100.ALA	H	8.39250817276898
146	A.100.ALA	N	125.752009081434
147	A.141.GLU	H	8.41252099485488
148	A.141.GLU	N	121.143659817089
149	A.89.GLU	H	8.43693978736577

150	A.89.GLU	N	122.809672123477
151	A.129.GLY	H	8.50727031390894
152	A.129.GLY	N	111.105392406489
153	A.124.GLN	H	8.43824528116722
154	A.124.GLN	N	121.429270241553
155	A.65.MET	H	8.10934716018911
156	A.65.MET	N	119.490875623357
157	A.152.VAL	H	8.28540240594171
158	A.152.VAL	N	122.630622455647
159	A.156.ALA	H	8.33897812440469
160	A.156.ALA	N	121.259578052718
161	A.153.VAL	H	8.30646429534772
162	A.153.VAL	N	125.502236174703
163	A.15.LYS	H	7.9755674693456
164	A.15.LYS	N	120.337006925447
165	A.58.GLN	H	8.46928939852844
166	A.58.GLN	N	122.020873700089
167	A.27.TRP	H	8.21394921003258
168	A.27.TRP	N	122.005945977307
169	A.86.LYS	H	8.18761009152359
170	A.86.LYS	N	120.492383494767
171	A.74.LYS	H	8.26755781802623
172	A.74.LYS	N	123.469135335437
177	A.70.ALA	H	8.26510390250802
178	A.70.ALA	N	124.684250212623
179	A.97.LYS	H	8.42063015955783
180	A.97.LYS	N	121.9623087738
181	A.146.LEU	H	8.1951868705851
182	A.146.LEU	N	123.403329989538
183	A.80.GLN	H	8.23479474947028
184	A.80.GLN	N	120.192475082722
185	A.57.ARG	H	8.38001151269596
186	A.57.ARG	N	122.580159758629
187	A.140.LEU	H	8.31604342537054
188	A.140.LEU	N	124.357530949333
189	A.91.GLU	H	8.47172522642605
190	A.91.GLU	N	121.923053449114
191	A.157.LYS	H	7.8399503927824
192	A.157.LYS	N	118.6826989977
193	A.54.VAL	H	7.96735947296294
194	A.54.VAL	N	123.210317470851
195	A.16.PHE	H	8.34584354560974
196	A.16.PHE	N	121.736420650001
198	A.123.GLN	N	121.413473190992
199	A.148.ALA	H	8.15038752790084
200	A.148.ALA	N	124.335689337832
201	A.53.PHE	H	8.29638048419517
202	A.53.PHE	N	121.887187974868

203	A.49.LYS	H	8.39753312136291
204	A.49.LYS	N	123.155500183985
205	A.61.VAL	H	7.96259139184437
206	A.61.VAL	N	120.93990536416
207	A.119.GLN	H	8.41450207224253
208	A.119.GLN	N	121.052452807665
209	A.22.ARG	H	8.53383848013807
210	A.22.ARG	N	124.309640750992
211	A.41.GLU	H	8.52580555382915
212	A.41.GLU	N	121.854122929562
213	A.95.GLN	H	8.4523053214923
214	A.95.GLN	N	122.893067971685
215	A.105.VAL	H	8.2022777438052
216	A.105.VAL	N	123.077848389432
217	A.12.SER	H	8.46729156252838
218	A.12.SER	N	117.96973009968
219	A.34.LYS	H	8.41527839250538
220	A.34.LYS	N	124.259017489555
221	A.151.GLU	H	8.35176821671364
222	A.151.GLU	N	122.409930465172
223	A.145.LYS	H	8.10776382997648
224	A.145.LYS	N	121.867714971003
225	A.84.ARG	H	8.14679888073703
226	A.84.ARG	N	123.078625116347
227	A.81.GLU	H	8.38025509607362
228	A.81.GLU	N	121.212187682192
229	A.121.GLN	H	8.44298791488788
230	A.121.GLN	N	121.434861496976
231	A.85.ALA	H	8.18406215146525
232	A.85.ALA	N	124.255757062419
233	A.90.LYS	H	8.37582851218139
234	A.90.LYS	N	121.84446211329
235	A.31.ARG	H	8.31320765285885
236	A.31.ARG	N	120.448850745192
237	A.111.CYS	H	8.351181203934
238	A.111.CYS	N	123.03750472086
239	A.159.LEU	H	8.26141270653009
240	A.159.LEU	N	124.190935845542
241	A.115.LYS	H	8.4152599377945
242	A.115.LYS	N	121.820189336002
243	A.51.LYS	H	8.16972750575749
244	A.51.LYS	N	120.78346285409
245	A.21.LYS	H	8.29081272883763
246	A.21.LYS	N	124.172051298072
247	A.33.VAL	H	8.26278997002589
248	A.33.VAL	N	121.817265260219
249	A.45.LEU	H	8.2392897224042
250	A.45.LEU	N	122.182878015085
251	A.66.CYS	H	8.40337985158974
252	A.66.CYS	N	122.376728272517

253	A.19.ALA	H	8.44331859575455
254	A.19.ALA	N	124.126099187667
255	A.125.GLN	H	8.45438068545083
256	A.125.GLN	N	121.754657129709
257	A.78.GLU	H	8.31398737658181
258	A.78.GLU	N	120.323772467818
259	A.131.ARG	H	8.40200181904703
260	A.131.ARG	N	122.786223879971
261	A.52.GLN	H	8.39834612849481
262	A.52.GLN	N	120.896875877982
263	A.138.LYS	H	8.36149264838829
264	A.138.LYS	N	123.850808922894
265	A.132.ASP	H	8.36266493369753
266	A.132.ASP	N	121.696343635667
267	A.69.LEU	H	8.12550230821998
268	A.69.LEU	N	122.789809003111
269	A.8.TYR	H	8.12559751281663
270	A.8.TYR	N	122.279355257927
271	A.144.LEU	H	8.13019327101823
272	A.144.LEU	N	123.774571533172
273	A.136.ILE	H	8.08340573420366
274	A.136.ILE	N	122.208271149118
275	A.118.GLN	H	8.52795579414328
276	A.118.GLN	N	122.607971527836
277	A.134.ASN	H	8.48593519211459
278	A.134.ASN	N	121.181750101798
279	A.149.ALA	H	8.23658020779958
280	A.149.ALA	N	122.626555585537
281	A.13.GLU	H	8.36671401620825
282	A.13.GLU	N	122.075995573751
295	A.150.GLN	CA	55.7069765894745
297	A.6.SER	CA	58.6276166528172
301	A.35.ASP	CA	54.7383621031309
303	A.75.LYS	CA	57.9550512545689
307	A.9.LEU	CA	55.0784846292944
311	A.27.TRP	CA	57.5184272338055
313	A.92.SER	CA	58.7427386014227
326	A.120.GLN	CA	56.3363066590262
331	A.147.ASP	CA	54.3795660413204
338	A.160.LYS	CB	33.8161882579038
341	A.116.VAL	CA	62.2064453743072
345	A.128.SER	CA	58.6555355181864
352	A.104.LEU	CA	55.045125790708
358	A.72.GLU	CB	30.0362611110177
372	A.109.SER	CA	58.3201582962761
380	A.106.MET	CA	52.8542087487518
382	A.36.GLY	CA	45.2970577730535
385	A.7.THR	CA	62.2390871334706

386	A.7.THR	CB	69.6336866076656
389	A.154.ASP	CA	51.5940949574596
390	A.154.ASP	CB	41.622018337595
396	A.130.SER	CA	58.4521872806027
400	A.38.VAL	CA	59.2325327146867
404	A.139.THR	CA	62.2445421033626
410	A.59.ALA	CA	52.6722624588938
423	A.100.ALA	CA	50.3721663430004
424	A.100.ALA	CB	17.9496843850564
433	A.65.MET	CB	33.0406276674158
442	A.58.GLN	CB	29.5553758912522
443	A.58.GLN	CA	55.6725565036655
444	A.27.TRP	CB	29.580190802571
449	A.74.LYS	CB	32.7824842608781
461	A.80.GLN	CA	57.2140855314386
467	A.91.GLU	CA	56.7666968129942
474	A.123.GLN	CA	56.1313578659017
476	A.148.ALA	CA	52.7562012238913
477	A.148.ALA	CB	19.1733020505696
478	A.53.PHE	CA	57.9893465059199
483	A.61.VAL	CA	59.7143887424048
484	A.119.GLN	CA	56.4856218711017
486	A.22.ARG	CB	29.9932289921877
494	A.12.SER	CA	58.8674393705341
496	A.34.LYS	CA	56.7522162379869
500	A.145.LYS	CB	32.7405709480691
507	A.121.GLN	CA	56.3177710929449
508	A.85.ALA	CA	53.0368503940606
510	A.90.LYS	CA	56.6875947434393
512	A.31.ARG	CB	31.3377013200903
513	A.31.ARG	CA	55.8644444145146
514	A.111.CYS	CA	56.4836908203491
515	A.111.CYS	CB	27.5506653271884
516	A.159.LEU	CA	55.2718412125765
528	A.66.CYS	CA	56.8926067253181
530	A.19.ALA	CA	52.6618873731187
538	A.52.GLN	CA	55.9240401409802
540	A.138.LYS	CB	32.9727841257802
549	A.144.LEU	CA	55.7453558611482
552	A.118.GLN	CB	29.3185859926449
553	A.118.GLN	CA	56.351914833244
559	A.13.GLU	CA	56.9658430448153
637	A.123.GLN	C	176.099082308967
652	A.43.VAL	H	8.18936436988068
653	A.43.VAL	N	123.412614004396
657	A.90.LYS	C	176.905638870602
709	A.142.ASP	CA	55.0249391298083
710	A.25.GLY	CA	45.5806133887293
713	A.5.MET	CB	33.0127722333865
714	A.5.MET	CA	55.5134062940874

715	A.47.GLU	CA	56.4361788123718
716	A.47.GLU	CB	30.5803179534086
717	A.34.LYS	CB	32.9310455744164
719	A.74.LYS	CA	58.2225583308603
721	A.19.ALA	CB	19.4135488070061
723	A.8.TYR	CA	58.084126060246
727	A.81.GLU	CB	29.8371435684749
728	A.81.GLU	CA	57.7639380401962
730	A.91.GLU	CB	30.1139822233949
732	A.92.SER	CB	63.9322813510633
735	A.87.LYS	CB	32.9072310975703
736	A.87.LYS	CA	56.9045717733053
737	A.10.GLN	CA	56.0234364061472
738	A.10.GLN	CB	29.4901376097788
740	A.93.GLY	CA	45.4428567826084
743	A.141.GLU	CA	57.44573162941
744	A.141.GLU	CB	30.0892471956903
746	A.3.MET	CA	55.6420584700253
748	A.45.LEU	CB	42.4391117316599
750	A.72.GLU	CA	57.4280347012399
754	A.125.GLN	CA	55.7937169700983
756	A.159.LEU	CB	42.2133924978897
758	A.115.LYS	CA	56.3995778612265
760	A.13.GLU	CB	30.085901048325
762	A.109.SER	CB	63.7574765983933
766	A.29.LYS	CA	56.4677546187386
768	A.129.GLY	CA	45.3522400282278
776	A.149.ALA	CB	19.0205874886682
778	A.134.ASN	CB	38.9113338683556
779	A.134.ASN	CA	53.6492226381679
780	A.78.GLU	CB	29.7103379868864
781	A.78.GLU	CA	58.1306386081544
782	A.149.ALA	CA	52.656256134282
784	A.16.PHE	CA	57.5799544452739
785	A.16.PHE	CB	39.7448929349449
786	A.49.LYS	CB	32.8060745448362
787	A.49.LYS	CA	56.6434679563353
788	A.41.GLU	CB	30.3529297476226
789	A.41.GLU	CA	56.6055011035944
790	A.76.GLU	CB	29.6794603554548
791	A.76.GLU	CA	57.9537770934171
797	A.116.VAL	CB	32.9479607478979
799	A.153.VAL	CB	32.8479525790331
801	A.35.ASP	CB	41.1480886730003
803	A.157.LYS	CB	32.9618266211915
805	A.157.LYS	CA	56.2369610763359
807	A.6.SER	CB	63.7026673745019
808	A.55.ALA	CB	19.0976279839047

809	A.55.ALA	CA	52.6129321111553
811	A.153.VAL	CA	62.1016182070185
813	A.86.LYS	CA	56.6789559640272
814	A.86.LYS	CB	32.8805991773681
815	A.24.ASP	CA	54.0086442998786
816	A.24.ASP	CB	40.9362148703042
821	A.45.LEU	CA	55.4209888656737
822	A.37.TYR	CA	58.1086417014226
823	A.37.TYR	CB	39.113140030062
829	A.138.LYS	CA	56.5755653893384
831	A.28.ARG	CA	56.1644662335979
835	A.59.ALA	CB	19.3010933371144
839	A.82.ARG	CB	30.4734880146801
841	A.75.LYS	CB	32.5686338670565
843	A.46.TYR	CA	57.7132946486184
844	A.46.TYR	CB	38.7144522617398
845	A.136.ILE	CB	38.6619182131423
846	A.136.ILE	CA	61.5433925723597
854	A.128.SER	CB	63.8299373150558
858	A.151.GLU	CA	56.5034133180451
859	A.151.GLU	CB	30.3960585156331
862	A.152.VAL	CA	62.3166227724393
864	A.4.GLY	CA	45.3175764956717
865	A.57.ARG	CB	30.7993335143813
866	A.57.ARG	CA	56.2845410257562
867	A.26.THR	CA	62.5383406231176
868	A.26.THR	CB	70.0179069772531
870	A.85.ALA	CB	19.0032083689008
871	A.73.SER	CA	59.3023846351148
872	A.73.SER	CB	63.5217675423412
878	A.69.LEU	CA	55.407535673744
882	A.145.LYS	CA	56.3443702062225
883	A.79.LYS	CA	58.1825192991813
884	A.79.LYS	CB	32.6426824881194
885	A.56.GLN	CB	29.505182354189
888	A.139.THR	CB	69.9072338173308
891	A.156.ALA	CA	53.0249174314247
894	A.53.PHE	CB	39.631537988467
896	A.15.LYS	CA	55.99833903398
899	A.122.GLN	CA	56.133684390757
900	A.147.ASP	CB	41.0608502674748
902	A.52.GLN	CB	29.3750404839212
904	A.48.SER	CA	58.5598693044381
905	A.48.SER	CB	63.736675199904
906	A.60.GLY	CA	45.1012652955102
910	A.21.LYS	CA	56.006865947308
911	A.40.GLN	CB	29.6263748431407
912	A.40.GLN	CA	56.0302216247189
917	A.11.SER	CA	58.2948608647703
918	A.11.SER	CB	63.9593388548365

919	A.33.VAL	CB	33.0354558611748
925	A.83.THR	CA	63.419488092615
926	A.83.THR	CB	69.4738969019303
927	A.80.GLN	CB	28.8470997396129
931	A.84.ARG	CB	30.6063996901653
933	A.89.GLU	CA	56.8982732057122
935	A.30.ALA	CB	19.4600303872156
936	A.30.ALA	CA	52.4727979315593
937	A.110.THR	CA	61.7188050065469
938	A.110.THR	CB	69.8105971491071
939	A.158.GLN	CB	29.4302124388994
940	A.158.GLN	CA	55.7076691985768
944	A.20.THR	CA	61.6189893780889
945	A.20.THR	CB	70.3931460648919
947	A.42.GLU	CA	56.3343094475697
951	A.65.MET	CA	55.2726837953708
956	A.77.ARG	CB	30.2238216836409
957	A.77.ARG	CA	58.0326605208026
958	A.130.SER	CB	63.8619014634046
959	A.51.LYS	CB	33.0201554551005
960	A.51.LYS	CA	56.5402147633076
961	A.137.SER	CA	58.5948595827339
962	A.137.SER	CB	63.6535715888842
965	A.68.LEU	CA	55.8731420127775
969	A.143.THR	CA	62.9928770808721
970	A.143.THR	CB	69.574440287881
971	A.135.SER	CA	58.7336857086761
973	A.117.SER	CA	58.3128837067115
974	A.117.SER	CB	63.8794683815011
975	A.133.ILE	CB	38.6432918455626
976	A.133.ILE	CA	61.8843434393149
977	A.70.ALA	CB	18.8990915633953
978	A.70.ALA	CA	53.2741556969475
980	A.12.SER	CB	63.6846540748902
990	A.8.TYR	CB	38.7171232284663
993	A.43.VAL	CA	60.1677521299441
997	A.143.THR	C	174.996250407438
998	A.26.THR	C	174.414657616042
999	A.150.GLN	C	175.907815234695
1000	A.6.SER	C	174.818571869628
1001	A.48.SER	C	174.891440445062
1002	A.35.ASP	C	176.504134639975
1003	A.75.LYS	C	178.243285940158
1004	A.20.THR	C	173.972529323423
1005	A.9.LEU	C	176.980107979702
1006	A.24.ASP	C	176.978730308006
1007	A.82.ARG	CA	57.5367007086889
1008	A.82.ARG	C	177.605247846526

1009	A.92.SER	C	175.308766154771
1010	A.93.GLY	C	174.123551965731
1011	A.55.ALA	C	177.727199173914
1012	A.88.GLN	CA	56.122662954687
1013	A.88.GLN	CB	29.3500804175918
1014	A.88.GLN	C	176.291408387541
1015	A.11.SER	C	174.964432367196
1016	A.28.ARG	C	175.969521319283
1018	A.94.ARG	CA	55.9065767983101
1019	A.94.ARG	C	176.231043829686
1020	A.120.GLN	C	176.327191453186
1021	A.142.ASP	C	177.069471861488
1022	A.4.GLY	C	174.186254906552
1023	A.147.ASP	C	176.247822360291
1024	A.73.SER	C	175.512936815172
1025	A.98.ALA	C	175.402419943982
1029	A.160.LYS	C	181.18665584263
1030	A.116.VAL	C	176.226508273708
1031	A.14.GLY	CA	45.3941821419572
1032	A.14.GLY	C	173.892166268366
1033	A.110.THR	C	174.179450087591
1034	A.128.SER	C	175.293659719927
1035	A.30.ALA	C	177.461689972189
1036	A.5.MET	C	176.586935320999
1037	A.104.LEU	C	176.783690059442
1038	A.10.GLN	C	176.12371440074
1039	A.102.GLY	CA	45.2231435026645
1040	A.102.GLY	C	173.978460025988
1042	A.72.GLU	C	177.483395261655
1043	A.135.SER	C	174.708038167447
1044	A.79.LYS	C	178.072089716513
1045	A.3.MET	C	176.809451957824
1048	A.50.GLY	CA	45.339846275029
1049	A.50.GLY	C	174.276927910125
1050	A.42.GLU	C	176.131501746191
1051	A.77.ARG	C	177.945433707558
1052	A.109.SER	C	174.96519263304
1053	A.122.GLN	C	176.249586361925
1054	A.37.TYR	C	175.121463921545
1055	A.117.SER	C	174.71564761001
1056	A.106.MET	C	173.654815876911
1057	A.36.GLY	C	173.588347295897
1058	A.158.GLN	C	175.712125336263
1059	A.7.THR	C	174.400285229859
1060	A.56.GLN	CA	55.8819285048
1061	A.56.GLN	C	176.054545057441
1062	A.154.ASP	C	175.334491679942
1063	A.71.ALA	H	8.25377932621954
1064	A.71.ALA	N	122.617437045257
1067	A.25.GLY	C	174.747804915607

1068	A.133.ILE	C	176.275669911928
1070	A.130.SER	C	174.630620100859
1071	A.46.TYR	C	175.761742998084
1072	A.38.VAL	C	173.554988166505
1073	A.40.GLN	C	176.017271730666
1074	A.64.GLY	C	174.202335293833
1075	A.139.THR	C	174.946901140174
1076	A.103.VAL	C	175.922032763148
1077	A.29.LYS	C	176.347230155584
1078	A.59.ALA	C	178.14535640444
1079	A.60.GLY	C	173.731180960414
1080	A.68.LEU	C	177.772229948198
1081	A.83.THR	C	175.130511267828
1082	A.76.GLU	C	177.894217803399
1083	A.47.GLU	C	176.25159131211
1084	A.137.SER	C	174.669865106067
1085	A.100.ALA	C	175.536150216567
1086	A.141.GLU	C	176.839757529869
1089	A.89.GLU	C	176.812609216049
1090	A.129.GLY	C	174.257269157502
1092	A.124.GLN	CA	55.9451364704595
1093	A.124.GLN	C	175.945527760583
1094	A.65.MET	C	175.906849985693
1095	A.152.VAL	C	176.037543401753
1096	A.156.ALA	C	178.447987361676
1097	A.153.VAL	C	175.404311662985
1098	A.15.LYS	C	175.886345003261
1099	A.58.GLN	C	175.570065165576
1100	A.27.TRP	C	176.412423725255
1101	A.86.LYS	C	177.065602989739
1102	A.74.LYS	C	177.534963936797
1103	A.70.ALA	C	178.370300475326
1104	A.97.LYS	C	176.001424054361
1105	A.146.LEU	C	177.400650586649
1106	A.80.GLN	C	177.441156513221
1109	A.57.ARG	C	176.270830663483
1110	A.140.LEU	C	177.795473418846
1111	A.91.GLU	C	176.84753730927
1112	A.157.LYS	C	176.461162564396
1113	A.54.VAL	CA	62.1857744997184
1114	A.54.VAL	CB	33.0331440606068
1115	A.54.VAL	C	175.40623568712
1117	A.148.ALA	C	177.726311019023
1118	A.53.PHE	C	175.479175048873
1119	A.49.LYS	C	177.172915963309
1120	A.61.VAL	C	174.237495757952
1121	A.119.GLN	C	176.34913446542
1122	A.22.ARG	C	175.107816613168

1123	A.41.GLU	C	176.257038449973
1124	A.95.GLN	CA	53.6534656604844
1125	A.95.GLN	C	173.988789645766
1126	A.105.VAL	CB	32.7606142134915
1127	A.105.VAL	CA	62.0765536666738
1128	A.105.VAL	C	175.778676358424
1129	A.12.SER	C	174.729199466944
1130	A.34.LYS	C	176.15104471111
1131	A.151.GLU	C	176.237972075554
1132	A.145.LYS	C	176.630127320374
1133	A.84.ARG	C	176.595630993195
1134	A.81.GLU	C	177.75540437839
1135	A.121.GLN	C	176.29399983525
1136	A.85.ALA	C	178.212352714851
1138	A.31.ARG	C	175.772587924801
1139	A.111.CYS	C	172.083719863832
1140	A.159.LEU	C	176.320629334251
1141	A.115.LYS	C	176.810071595741
1142	A.51.LYS	C	176.650469442554
1143	A.21.LYS	C	176.363666340899
1144	A.33.VAL	CA	62.3196482429755
1145	A.33.VAL	C	176.180180060219
1146	A.45.LEU	C	177.29023945018
1147	A.66.CYS	C	173.072650902742
1148	A.19.ALA	C	177.950242263922
1149	A.125.GLN	C	175.798715245998
1152	A.131.ARG	CA	56.0811988420802
1154	A.131.ARG	C	175.87184695068
1155	A.52.GLN	C	175.643148836455
1156	A.138.LYS	C	176.881063745699
1157	A.132.ASP	CB	41.1714654454989
1158	A.132.ASP	CA	54.1855056038716
1159	A.132.ASP	C	176.686787199387
1160	A.69.LEU	C	177.755891045485
1161	A.8.TYR	C	175.663246417209
1162	A.144.LEU	C	177.455003538304
1163	A.136.ILE	C	176.521670321145
1164	A.118.GLN	C	176.294348764017
1165	A.134.ASN	C	175.427900395901
1166	A.149.ALA	C	177.877772511594
1167	A.13.GLU	C	176.955536406927
1168	A.43.VAL	C	174.315931818673
1170	A.71.ALA	CA	53.4746687342458
1171	A.71.ALA	CB	18.9286650021711
1172	A.71.ALA	C	178.869306532514
1173	A.146.LEU	CA	55.3802115719315
1175	A.87.LYS	C	176.787196149085
1176	A.78.GLU	C	177.903627344909
1177	A.42.GLU	CB	30.4143475858989
1178	A.16.PHE	C	174.975391005211

1181	A.32.ARG	H	8.43719110176841
1182	A.32.ARG	N	122.840546122913
1183	A.32.ARG	C	176.322363366293
1184	A.32.ARG	CA	55.9253680129906
1185	A.32.ARG	CB	30.9709683837543
1187	A.18.PRO	CA	62.9111835136863
1188	A.38.VAL	CB	33.2492126959395
1189	A.64.GLY	CA	45.2847328598725
1190	A.84.ARG	CA	57.0313214823496
1191	A.97.LYS	CA	56.0074957681296
1193	A.103.VAL	CB	32.71140950463
1194	A.103.VAL	CA	62.3347835940383
1195	A.18.PRO	C	176.506894917299
1196	A.23.PRO	C	176.670043221592
1197	A.39.PRO	C	176.801022989451
1198	A.44.PRO	C	176.621772105454
1199	A.63.PRO	C	177.762560798086
1200	A.67.PRO	C	177.134765210292
1201	A.96.PRO	C	176.728720094729
1202	A.99.PRO	C	176.214203162691
1203	A.101.PRO	C	177.639837937504
1204	A.108.PRO	C	177.040823948106
1206	A.155.PRO	C	177.47739607483
1207	A.155.PRO	CB	32.1369810236726
1208	A.155.PRO	CA	63.9668843169411
1209	A.127.PRO	CA	63.2781894072325
1210	A.127.PRO	CB	32.1266135333795
1212	A.114.PRO	CA	62.8211176741013
1213	A.108.PRO	CA	63.082703960652
1216	A.101.PRO	CA	63.3882107663241
1218	A.99.PRO	CA	62.6967879280865
1220	A.67.PRO	CA	63.785999211263
1221	A.63.PRO	CA	63.4304768868387
1223	A.44.PRO	CB	32.0919496130638
1224	A.44.PRO	CA	63.0132660924989
1225	A.39.PRO	CA	63.0222390637452
1228	A.23.PRO	CA	63.8261244337032
1229	A.96.PRO	CA	63.0364565628197
1231	A.114.PRO	C	176.851866970656
1232	A.2.ALA	CA	52.5807875591505
1233	A.2.ALA	CB	19.4036037754273
1234	A.2.ALA	C	177.871416227831
1235	A.3.MET	CB	32.7824930568251
1241	A.159.LEU	HA	4.34593271029237
1243	A.159.LEU	HDx%	0.862779189728707
1244	A.159.LEU	CDx	23.4224196913386
1245	A.159.LEU	HDy%	0.92628001162214
1246	A.159.LEU	CDy	25.0279147310094

1247	A.159.LEU	HBx	1.59961087786167
1249	A.159.LEU	HBy	1.65119571395091
1250	A.159.LEU	HG	1.6307790591698
1251	A.159.LEU	CG	27.0006940684746
1252	A.158.GLN	HA	4.30012597271553
1253	A.158.GLN	HBx	1.97588225582574
1254	A.158.GLN	HBy	2.09300163820931
1255	A.158.GLN	CG	33.8212971703691
1256	A.158.GLN	HGx	2.34578444836778
1257	A.158.GLN	HGy	2.36472169037983
1258	A.157.LYS	HA	4.26084370712023
1261	A.157.LYS	HEy	2.99965502812004
1263	A.157.LYS	CE	42.1441993084821
1268	A.157.LYS	HEx	2.99754737536867
1269	A.157.LYS	HBy	1.90233387285441
1270	A.157.LYS	HBx	1.77061439674395
1271	A.157.LYS	HDy	1.69104080554575
1272	A.157.LYS	HDx	1.68456016156905
1273	A.157.LYS	HGy	1.44629274127376
1274	A.157.LYS	HGx	1.36417130341863
1276	A.157.LYS	CD	29.1115903515665
1277	A.157.LYS	CG	24.8606424955113
1280	A.156.ALA	HA	4.2498593381701
1281	A.156.ALA	HB%	1.39965047609678
1282	A.156.ALA	CB	18.7657511626117
1283	A.155.PRO	HA	4.34245869962761
1284	A.155.PRO	HDy	3.92258128820395
1285	A.155.PRO	HDx	3.89122300029296
1286	A.155.PRO	HBx	2.30844222441927
1287	A.155.PRO	HBy	2.32263407664154
1288	A.155.PRO	HGx	1.98812251495585
1289	A.155.PRO	HGy	2.02317764070114
1290	A.155.PRO	CD	51.0463203376373
1291	A.155.PRO	CG	27.3891221602869
1292	A.153.VAL	HA	4.0465597324817
1293	A.153.VAL	HB	1.98495362585278
1294	A.153.VAL	CGy	21.0668617838213
1295	A.153.VAL	HGy%	0.902143479382706
1297	A.153.VAL	CGx	20.8093570112858
1298	A.153.VAL	HGx%	0.863570274720582
1299	A.152.VAL	HA	4.08200427943161
1300	A.152.VAL	CGy	20.5912416808208
1301	A.152.VAL	HGy%	0.936750169088419
1302	A.152.VAL	CB	32.721458352783
1303	A.152.VAL	HB	2.02701320788928
1304	A.152.VAL	CGx	20.712883150427
1305	A.152.VAL	HGx%	0.891502210032151
1306	A.151.GLU	HA	4.26726946183747
1307	A.151.GLU	CG	36.2632416768304
1308	A.151.GLU	HGy	2.26231785106215

1309	A.151.GLU	HGx	2.19143839226636
1310	A.151.GLU	HBy	1.99139271972823
1311	A.151.GLU	HBx	1.92113634134158
1313	A.150.GLN	HA	4.28426291919863
1314	A.150.GLN	CG	33.8525275639315
1315	A.150.GLN	HGx	2.36093783744823
1316	A.150.GLN	CB	29.5476105111076
1317	A.150.GLN	HBx	1.97923429918856
1318	A.150.GLN	HBy	2.10704209336481
1319	A.150.GLN	HGy	2.37396993005915
1320	A.149.ALA	HA	4.2727775287601
1321	A.149.ALA	HB%	1.39407566012668
1322	A.148.ALA	HA	4.25985566851278
1323	A.148.ALA	HB%	1.38957782766467
1324	A.147.ASP	HA	4.54701804044039
1325	A.147.ASP	HBy	2.68495261363446
1326	A.147.ASP	HBx	2.61309153098765
1327	A.146.LEU	HA	4.30938983909402
1344	A.146.LEU	CB	42.3291701414111
1345	A.146.LEU	HBy	1.65489775462503
1346	A.146.LEU	CG	27.0754039383681
1347	A.146.LEU	HG	1.63014303924718
1348	A.146.LEU	HBx	1.56171381217065
1349	A.146.LEU	CDx	23.4175582685149
1350	A.146.LEU	HDx%	0.854723931639856
1351	A.146.LEU	CDy	24.8695900021027
1352	A.146.LEU	HDy%	0.91237390446095
1353	A.145.LYS	HA	4.28157291174897
1354	A.145.LYS	CE	42.130860122373
1355	A.145.LYS	HEy	2.99001666487751
1356	A.145.LYS	HEx	2.98545743732452
1359	A.145.LYS	HBx	1.76738106248331
1360	A.145.LYS	HBy	1.83749885999469
1361	A.145.LYS	CD	29.1297604797245
1362	A.145.LYS	HDy	1.67931811682275
1363	A.145.LYS	HDx	1.67297228405394
1365	A.145.LYS	HGx	1.38342151074254
1366	A.145.LYS	CG	24.7737974316183
1367	A.145.LYS	HGy	1.45392030830024
1368	A.144.LEU	HA	4.28784062155993
1372	A.144.LEU	HBy	1.69671316189585
1373	A.144.LEU	CG	27.2290363723878
1374	A.144.LEU	HG	1.63203893515715
1375	A.144.LEU	CB	42.0967997433029
1376	A.144.LEU	HBx	1.56449312127757
1377	A.144.LEU	CDx	23.5637058106199
1378	A.144.LEU	HDx%	0.857639410208882
1379	A.144.LEU	CDy	25.023553320571

1380	A.144.LEU	HDy%	0.914474045298984
1381	A.143.THR	HA	4.20582360877823
1382	A.143.THR	HB	4.25473472172001
1383	A.143.THR	HG2%	1.22176776725509
1384	A.143.THR	CG2	21.7364517897883
1387	A.142.ASP	HA	4.59005618239274
1388	A.142.ASP	CB	40.9935470624572
1389	A.142.ASP	HBy	2.69765547343716
1390	A.142.ASP	HBx	2.68783768966756
1391	A.141.GLU	HA	4.17613883010119
1392	A.141.GLU	CG	36.3476869833588
1393	A.141.GLU	HGx	2.25457948198911
1394	A.141.GLU	HGy	2.27330901971378
1395	A.141.GLU	HBx	1.95869974878193
1396	A.141.GLU	HBy	2.03275166022827
1397	A.140.LEU	HA	4.30863127443933
1398	A.140.LEU	CA	55.8880116812417
1399	A.140.LEU	CB	42.0986291430675
1400	A.140.LEU	HBy	1.66847910485909
1403	A.140.LEU	HBx	1.61905824514766
1404	A.140.LEU	CDx	23.6760330479552
1405	A.140.LEU	HDx%	0.875484336860635
1406	A.140.LEU	CDy	24.7659710390516
1407	A.140.LEU	HDy%	0.925332249043564
1408	A.140.LEU	HG	1.6379375984326
1410	A.140.LEU	CG	26.9258980213011
1412	A.139.THR	HB	4.24739257653625
1414	A.139.THR	HA	4.3068417783575
1416	A.139.THR	CG2	21.7519347349642
1417	A.139.THR	HG2%	1.21597527384229
1419	A.138.LYS	HA	4.38722885909223
1420	A.137.SER	HA	4.41894884534801
1421	A.137.SER	HBx	3.85413396015719
1422	A.137.SER	HBy	3.85944726939258
1423	A.136.ILE	HA	4.19024494391652
1424	A.136.ILE	HB	1.89997395077659
1425	A.136.ILE	CG1	27.335087299549
1426	A.136.ILE	HG1y	1.47931609012169
1427	A.136.ILE	HG1x	1.1880543725815
1428	A.136.ILE	CG2	17.5431866836096
1429	A.136.ILE	HG2%	0.914115387935074
1430	A.136.ILE	CD1	13.157877648825
1431	A.136.ILE	HD1%	0.8571378189312
1432	A.135.SER	HA	4.42289628513084
1433	A.135.SER	CB	63.7434123308843
1434	A.135.SER	HBy	3.89021618652063
1435	A.135.SER	HBx	3.86912052666171
1436	A.134.ASN	HA	4.72761721020033
1437	A.134.ASN	HBy	2.84540699930663
1438	A.134.ASN	HBx	2.77900092543862

1439	A.133.ILE	HA	4.11921687401832
1440	A.133.ILE	HB	1.90537498536869
1441	A.133.ILE	HG1y	1.42158219261946
1442	A.133.ILE	HG1x	1.20719756492315
1443	A.133.ILE	HG2%	0.905577261165483
1444	A.133.ILE	HD1%	0.861894838774748
1445	A.133.ILE	CG1	27.2680827292652
1446	A.133.ILE	CG2	17.5697147786534
1447	A.133.ILE	CD1	13.2314550120424
1448	A.132.ASP	HA	4.61415564312752
1449	A.132.ASP	HBy	2.72742875625872
1450	A.132.ASP	HBx	2.59862272387263
1451	A.131.ARG	HA	4.34553792298856
1453	A.131.ARG	CD	43.3548914014175
1454	A.131.ARG	HDy	3.18593971419932
1455	A.131.ARG	HDx	3.17306723353894
1456	A.131.ARG	CB	30.8206664692751
1457	A.131.ARG	HBy	1.8542331900828
1458	A.131.ARG	HBx	1.7462362425018
1460	A.131.ARG	HGy	1.63844447400698
1461	A.131.ARG	CG	27.1143365441161
1462	A.131.ARG	HGx	1.58572422301006
1463	A.130.SER	HA	4.43593679859534
1464	A.130.SER	HBx	3.84227946482086
1465	A.130.SER	HBy	3.85940731822489
1466	A.129.GLY	HAx	4.00052229203905
1467	A.129.GLY	HAy	4.01485998887632
1468	A.128.SER	HA	4.41938616256034
1469	A.128.SER	HBx	3.86682011493711
1470	A.128.SER	HBy	3.91332721335589
1471	A.127.PRO	HA	4.45657890069168
1472	A.127.PRO	HDy	3.79866313553834
1473	A.127.PRO	HDx	3.66804968956218
1478	A.127.PRO	CD	50.5568577367577
1480	A.127.PRO	CG	27.3787572422512
1481	A.127.PRO	HBy	2.3186620236163
1482	A.127.PRO	HGy	2.03875195265987
1483	A.127.PRO	HGx	2.01446790859257
1484	A.127.PRO	HBx	1.93753328611028
1485	A.126.GLN	H	8.5112160283652
1486	A.126.GLN	N	122.850713505174
1487	A.126.GLN	C	174.107023151638
1490	A.127.PRO	C	177.117396498901
1491	A.125.GLN	HA	4.30966274445921
1492	A.125.GLN	CG	33.8422070705296
1493	A.125.GLN	HGy	2.36550075487118
1494	A.125.GLN	HGx	2.35251993966271
1496	A.125.GLN	HBy	2.07573924550774

1497	A.125.GLN	CB	29.469010085867
1498	A.125.GLN	HBx	1.96433211652058
1499	A.124.GLN	HA	4.30112800513753
1500	A.124.GLN	CG	33.8440729415089
1501	A.124.GLN	HGy	2.37608251235239
1502	A.124.GLN	HGx	2.36069524712966
1504	A.124.GLN	HBy	2.09366647123898
1506	A.124.GLN	CB	29.4350418847383
1507	A.124.GLN	HBx	1.98487925437278
1508	A.123.GLN	HA	4.28079061452507
1509	A.123.GLN	CG	33.8466063900032
1510	A.123.GLN	HGy	2.38189907076223
1511	A.123.GLN	HGx	2.35985094772761
1513	A.123.GLN	HBy	2.09612485275272
1514	A.123.GLN	CB	29.3430761027391
1515	A.123.GLN	HBx	1.9992793691625
1516	A.122.GLN	HA	4.27969151760557
1517	A.122.GLN	H	8.43768592823593
1519	A.123.GLN	H	8.42896781421986
1520	A.122.GLN	CG	33.8447261036749
1521	A.122.GLN	HGy	2.37825876886851
1522	A.122.GLN	HGx	2.36732686319578
1524	A.122.GLN	HBy	2.09525904917655
1525	A.122.GLN	CB	29.3501490209449
1526	A.122.GLN	HBx	1.99679987604757
1527	A.121.GLN	HA	4.27227278131869
1528	A.121.GLN	CG	33.8420131592676
1529	A.121.GLN	HGy	2.376436215814
1530	A.121.GLN	HGx	2.36810593210305
1531	A.121.GLN	CB	29.318281339045
1532	A.121.GLN	HBy	2.09896881947831
1535	A.121.GLN	HBx	1.99720199863301
1536	A.120.GLN	HA	4.27717627958609
1537	A.120.GLN	CG	33.8523382753237
1538	A.120.GLN	HGy	2.37961212046511
1539	A.120.GLN	HGx	2.36056430559797
1541	A.120.GLN	HBy	2.09423234273857
1542	A.120.GLN	CB	29.3220816596852
1543	A.120.GLN	HBx	1.99978330793167
1544	A.119.GLN	HA	4.25889260305908
1545	A.119.GLN	CG	33.8384963908851
1546	A.119.GLN	HGy	2.37327595016506
1547	A.119.GLN	HGx	2.3605365363691
1549	A.119.GLN	HBy	2.07952973235842
1550	A.119.GLN	CB	29.2510747607776
1551	A.119.GLN	HBx	1.98526914002365
1552	A.118.GLN	HA	4.30461890172967
1553	A.118.GLN	HBx	1.98940802706058
1554	A.118.GLN	HBy	2.11673743064591
1555	A.118.GLN	CG	33.8262257010683

1556	A.118.GLN	HGy	2.37458254598136
1557	A.118.GLN	HGx	2.36330563679437
1558	A.117.SER	HA	4.43675144993761
1559	A.117.SER	HBx	3.85193578691457
1560	A.117.SER	HBy	3.911375401855
1561	A.116.VAL	HA	4.14381073279275
1562	A.116.VAL	HB	2.05632662960702
1563	A.116.VAL	CGy	20.5379837230624
1564	A.116.VAL	HGy%	0.930170945463184
1565	A.116.VAL	CGx	20.5276274260142
1566	A.116.VAL	HGx%	0.923605564684016
1567	A.115.LYS	HA	4.28762014307325
1568	A.115.LYS	HEx	2.98740076225972
1569	A.115.LYS	HEy	2.99185703282618
1570	A.115.LYS	CE	42.187693833458
1572	A.115.LYS	CB	33.0151421494223
1580	A.115.LYS	CG	24.8359243418864
1581	A.115.LYS	HGx	1.4017960915815
1593	A.110.THR	HA	4.36898838267622
1594	A.110.THR	HG2%	1.19837455501539
1595	A.110.THR	HB	4.24097621501401
1596	A.110.THR	CG2	21.7352919004586
1597	A.109.SER	HA	4.47186523207618
1598	A.109.SER	HBx	3.85377401996878
1599	A.109.SER	HBy	3.9152869814654
1600	A.108.PRO	HA	4.45756375178641
1601	A.108.PRO	HDx	3.65417017074291
1602	A.108.PRO	HDy	3.82346529328135
1603	A.108.PRO	CD	50.4962811997917
1606	A.108.PRO	HBy	2.31575166206849
1607	A.108.PRO	CB	32.0564001492355
1608	A.108.PRO	HBx	1.94415447717531
1609	A.108.PRO	CG	27.3713939612181
1610	A.108.PRO	HGx	2.03109536377629
1617	A.104.LEU	HA	4.3869343575922
1618	A.108.PRO	HGy	2.03771434081381
1619	A.104.LEU	CB	42.4435178439267
1620	A.104.LEU	HBy	1.60480132104794
1621	A.104.LEU	HBx	1.55922752798773
1622	A.104.LEU	CG	27.098458990091
1623	A.104.LEU	HG	1.57605559399345
1624	A.104.LEU	CDy	24.8897573656471
1625	A.104.LEU	HDy%	0.915081222204504
1626	A.104.LEU	CDx	23.7853075323337
1627	A.104.LEU	HDx%	0.856666212113661
1628	A.103.VAL	HA	4.07703224742336
1629	A.103.VAL	HB	2.05726645916189
1630	A.103.VAL	HGx%	0.902391099420806

1631	A.103.VAL	HGy%	0.911675522514003
1632	A.103.VAL	CGy	20.7159836326793
1633	A.103.VAL	CGx	20.7130404731713
1634	A.102.GLY	HAx	3.91937540121732
1635	A.102.GLY	HAy	3.93112493917727
1636	A.101.PRO	HA	4.39558488824955
1637	A.101.PRO	HDx	3.65555846401075
1638	A.101.PRO	HDy	3.79882573131361
1639	A.101.PRO	CD	50.7261989307862
1642	A.101.PRO	HBy	2.29425638250046
1643	A.101.PRO	CB	32.0520809513542
1644	A.101.PRO	HBx	1.9299811726398
1645	A.101.PRO	CG	27.4352103598358
1646	A.101.PRO	HGx	2.02660155340627
1647	A.101.PRO	HGy	2.04850731491983
1648	A.99.PRO	HA	4.38952371962519
1649	A.99.PRO	HDx	3.62179932778733
1650	A.99.PRO	HDy	3.78960727882887
1651	A.99.PRO	HBy	2.25972657518417
1652	A.99.PRO	CD	50.767892898256
1654	A.99.PRO	CB	32.0536143441479
1655	A.99.PRO	HBx	1.88892655194511
1656	A.99.PRO	CG	27.3750535058381
1657	A.99.PRO	HGy	2.00700675554919
1658	A.99.PRO	HGx	1.99614920467099
1659	A.97.LYS	HA	4.25898490029522
1660	A.97.LYS	CE	42.1418573596742
1661	A.97.LYS	HEy	3.00637522721424
1662	A.97.LYS	HEx	2.99983864556645
1664	A.97.LYS	HBy	1.79558236483224
1665	A.97.LYS	CB	33.1608755871241
1666	A.97.LYS	HBx	1.723295525594
1667	A.97.LYS	CD	29.1759980328758
1668	A.97.LYS	HDx	1.68005320729726
1669	A.97.LYS	HDy	1.69160831357868
1670	A.97.LYS	CG	24.7808982592502
1671	A.97.LYS	HGy	1.45607725089625
1672	A.97.LYS	HGx	1.44493632936858
1673	A.96.PRO	HDx	3.65617901329775
1674	A.96.PRO	HDy	3.78923249591701
1675	A.96.PRO	HA	4.40730559488931
1676	A.96.PRO	HBx	1.8857069279306
1677	A.96.PRO	CD	50.6630394169297
1679	A.96.PRO	CB	32.1705896429379
1680	A.96.PRO	HBy	2.28653438087224
1682	A.96.PRO	HGy	2.01791455712434
1683	A.96.PRO	CG	27.3256437848298
1684	A.96.PRO	HGx	2.00165268098789
1685	A.94.ARG	HA	4.33935605816637
1686	A.94.ARG	HDx	3.18139191200965

1687	A.94.ARG	HDy	3.18780767014796
1688	A.94.ARG	CD	43.3583993829887
1691	A.94.ARG	HBy	1.83871350655849
1692	A.94.ARG	CB	30.8132802941503
1693	A.94.ARG	HBx	1.72884232370324
1694	A.94.ARG	CG	27.0986350248964
1695	A.94.ARG	HGy	1.62199515247789
1696	A.94.ARG	HGx	1.57695566377993
1697	A.93.GLY	HAx	3.97282018388669
1698	A.93.GLY	HAy	3.98001204193982
1699	A.92.SER	HA	4.41728246006329
1700	A.92.SER	HBx	3.89110687512332
1701	A.92.SER	HBy	3.90803617018269
1702	A.91.GLU	HA	4.29216441921545
1703	A.91.GLU	HGx	2.2716721657447
1704	A.91.GLU	HGy	2.28760899830084
1705	A.91.GLU	HBx	1.97589127454738
1706	A.91.GLU	HBy	2.07567505703051
1707	A.91.GLU	CG	36.2852966809424
1709	A.90.LYS	HEx	2.99132505742728
1710	A.90.LYS	HEy	2.99568046472696
1711	A.90.LYS	HA	4.27782671612968
1712	A.90.LYS	CE	42.1439834765722
1714	A.90.LYS	HBy	1.849936342298
1715	A.90.LYS	CB	33.0196399861439
1716	A.90.LYS	HBx	1.79595334612906
1717	A.90.LYS	CD	29.1314342304525
1718	A.90.LYS	HDy	1.68606810679841
1719	A.90.LYS	HDx	1.68036658790727
1720	A.90.LYS	CG	24.8181409941688
1721	A.90.LYS	HGx	1.43360629431315
1722	A.89.GLU	HA	4.25527027340131
1724	A.89.GLU	HGy	2.31016044317368
1725	A.89.GLU	CG	36.1554650199385
1726	A.89.GLU	HGx	2.24505812208994
1727	A.89.GLU	HBy	2.05663669505195
1728	A.89.GLU	CB	30.3384393563978
1729	A.89.GLU	HBx	1.95915397271867
1730	A.88.GLN	HA	4.29296558573528
1731	A.88.GLN	HGx	2.37062343524414
1732	A.88.GLN	HGy	2.37491200043546
1733	A.88.GLN	HBx	1.99843993841625
1734	A.88.GLN	HBy	2.09565626581316
1735	A.88.GLN	CG	33.8431357236811
1737	A.85.ALA	HA	4.24859307665309
1738	A.85.ALA	HB%	1.41324573620675
1739	A.84.ARG	HA	4.2433564847682
1740	A.84.ARG	HDx	3.19960688310932

1741	A.84.ARG	HDy	3.20393730422087
1742	A.84.ARG	HBx	1.81651313048499
1743	A.84.ARG	HBy	1.87735664432275
1744	A.84.ARG	HGx	1.63160250574522
1745	A.84.ARG	HGy	1.69620911781389
1746	A.84.ARG	CD	43.3990237832927
1747	A.84.ARG	CG	27.2573085204514
1748	A.83.THR	HA	4.1948657667711
1749	A.83.THR	HB	4.25480578988466
1750	A.83.THR	HG2%	1.22904403541997
1751	A.83.THR	CG2	21.7243565132912
1752	A.82.ARG	HA	4.2881899983092
1753	A.82.ARG	HDx	3.21472763423399
1754	A.82.ARG	HDy	3.22075692425732
1755	A.82.ARG	HBx	1.89860129929343
1756	A.82.ARG	HBy	1.90666694046028
1757	A.82.ARG	HGx	1.6652328971316
1758	A.82.ARG	HGy	1.74384981875719
1759	A.82.ARG	CD	43.3958331585505
1760	A.82.ARG	CG	27.388275073907
1761	A.81.GLU	HA	4.17745130017821
1762	A.81.GLU	HGx	2.28341376050839
1763	A.81.GLU	HGy	2.35907531274447
1764	A.81.GLU	HBx	2.05022172871183
1765	A.81.GLU	HBy	2.057831971848
1766	A.81.GLU	CG	36.2546361359736
1768	A.80.GLN	HA	4.20857462532285
1769	A.80.GLN	HGx	2.41138472113317
1770	A.80.GLN	HGy	2.47050626996509
1771	A.80.GLN	HBx	2.11020847804396
1772	A.80.GLN	HBy	2.12135049033823
1773	A.80.GLN	CG	33.7834456860082
1774	A.78.GLU	HA	4.13130189528707
1775	A.78.GLU	HGx	2.25266909156567
1776	A.78.GLU	HGy	2.40861012085724
1777	A.78.GLU	HBx	2.04398234490854
1778	A.78.GLU	HBy	2.05021478038695
1779	A.78.GLU	CG	36.5331388612741
1781	A.77.ARG	HA	4.16108774061076
1782	A.77.ARG	HDx	3.20677491522365
1783	A.77.ARG	HDy	3.21632208120631
1784	A.77.ARG	HBx	1.87700831626818
1785	A.77.ARG	HBy	1.89282753274422
1786	A.77.ARG	HGx	1.6377287397626
1787	A.77.ARG	HGy	1.71403789288853
1788	A.77.ARG	CD	43.3895920382441
1789	A.77.ARG	CG	27.3424510654815
1790	A.76.GLU	HA	4.17464846964062
1791	A.76.GLU	HGx	2.30008961081031
1792	A.76.GLU	HGy	2.30839924911279

1793	A.76.GLU	HBx	2.04657358246942
1794	A.76.GLU	HBy	2.05576014746065
1795	A.76.GLU	CG	36.2765222110526
1796	A.73.SER	HA	4.38522337958823
1797	A.73.SER	HBx	3.92217317943385
1798	A.73.SER	HBy	3.98164846631712
1799	A.72.GLU	HA	4.22272772977353
1800	A.72.GLU	HGx	2.2961098163085
1801	A.72.GLU	HGy	2.31298122147593
1802	A.72.GLU	HBx	2.02669528762565
1803	A.72.GLU	HBy	2.07863700060126
1804	A.72.GLU	CG	36.016568655711
1805	A.71.ALA	HA	4.22082966878537
1806	A.71.ALA	HB%	1.4286508420472
1807	A.70.ALA	HA	4.23321448740663
1808	A.70.ALA	HB%	1.4071825197342
1809	A.69.LEU	HDx%	0.871525402817368
1810	A.69.LEU	HDy%	0.932595936562232
1811	A.69.LEU	HA	4.3145009425498
1813	A.69.LEU	HBy	1.66462906961075
1814	A.69.LEU	CB	42.1849876505442
1815	A.69.LEU	HBx	1.60607047198838
1816	A.69.LEU	CG	26.9152363876579
1817	A.69.LEU	HG	1.61777223237173
1818	A.69.LEU	CDx	23.6048842646074
1819	A.69.LEU	CDy	24.9107553556116
1820	A.68.LEU	HDx%	0.872989289128438
1821	A.68.LEU	HDy%	0.923197197889971
1822	A.68.LEU	HA	4.29038784249247
1823	A.68.LEU	CB	42.1242148288131
1824	A.68.LEU	HBy	1.65772637261637
1826	A.68.LEU	HBx	1.58003793464798
1827	A.68.LEU	CG	26.967541169112
1828	A.68.LEU	HG	1.64681538002338
1829	A.68.LEU	CDy	24.8658378835474
1831	A.68.LEU	CDx	23.6438266617348
1832	A.65.MET	HA	4.49678942089816
1833	A.65.MET	HGx	2.50276005515733
1834	A.65.MET	HGy	2.56636502027928
1835	A.65.MET	HBx	1.9872578312709
1836	A.65.MET	HBy	2.05987331531992
1837	A.65.MET	CG	32.047988603984
1838	A.64.GLY	HAx	3.91800604501216
1839	A.64.GLY	HAy	3.92504069357995
1840	A.63.PRO	HA	4.39953593420606
1841	A.63.PRO	HDx	3.66790421325903
1842	A.63.PRO	HDy	3.81328033054909
1849	A.63.PRO	CD	50.8013147013445

1851	A.63.PRO	CB	32.0451251826014
1852	A.63.PRO	HBy	2.29394640919929
1853	A.63.PRO	HBx	1.93586215648582
1854	A.63.PRO	CG	27.4712225792167
1855	A.63.PRO	HGy	2.04724247035742
1856	A.63.PRO	HGx	2.03412982479804
1857	A.60.GLY	HAx	3.92317697513452
1858	A.60.GLY	HAy	3.93532397405999
1859	A.59.ALA	HA	4.29919436387008
1860	A.59.ALA	HB%	1.39304648893588
1861	A.58.GLN	HA	4.30285517851641
1862	A.58.GLN	HGx	2.3543698518027
1863	A.58.GLN	HGy	2.36790457560354
1864	A.58.GLN	HBx	1.97004537524026
1865	A.58.GLN	HBy	2.08501345035993
1866	A.58.GLN	CG	33.7634486273693
1867	A.57.ARG	HGx	1.59806235226673
1868	A.57.ARG	HGy	1.61042322105089
1869	A.57.ARG	HBx	1.81251852881508
1870	A.57.ARG	HBy	1.82705668182495
1871	A.57.ARG	HDx	3.1735184844969
1872	A.57.ARG	HDy	3.17541867857546
1873	A.57.ARG	HA	4.28003555693741
1876	A.57.ARG	CD	43.3363517065314
1877	A.57.ARG	CG	27.1124975877871
1878	A.56.GLN	HA	4.26157014919718
1879	A.56.GLN	HGx	2.35779762103629
1880	A.56.GLN	HGy	2.36734362942344
1881	A.56.GLN	HBx	1.96637045584389
1882	A.56.GLN	HBy	2.07212658103009
1884	A.56.GLN	CG	33.8897930001245
1885	A.55.ALA	HB%	1.37970896209222
1886	A.55.ALA	HA	4.20340757882657
1887	A.54.VAL	HA	3.97590015081154
1888	A.54.VAL	HB	1.95132657420368
1889	A.54.VAL	HGx%	0.879171708135261
1890	A.54.VAL	HGy%	0.886892221855633
1891	A.54.VAL	CGy	21.1701487831078
1892	A.54.VAL	CGx	21.0198560498965
1893	A.53.PHE	HA	4.58450926390894
1894	A.53.PHE	HBx	3.00613196501271
1895	A.53.PHE	HBy	3.07266226562359
1896	A.52.GLN	HA	4.26353690298201
1897	A.52.GLN	HGx	2.18944622436513
1898	A.52.GLN	HGy	2.25985322597993
1899	A.52.GLN	HBx	1.90025488701196
1900	A.52.GLN	HBy	1.96279792174329
1901	A.52.GLN	CG	33.7224111931395
1902	A.50.GLY	HAx	3.94562889154758
1903	A.50.GLY	HAy	3.95634147927029

1905	A.48.SER	HBy	3.88195238801063
1906	A.48.SER	HBx	3.82717212779763
1909	A.48.SER	HA	4.35965653088688
1910	A.47.GLU	HA	4.27697686940049
1911	A.47.GLU	HGx	2.17219013535719
1912	A.47.GLU	HGy	2.18807479622337
1913	A.47.GLU	HBx	1.87391728745375
1914	A.47.GLU	HBy	1.99153357734106
1915	A.47.GLU	CG	36.2487283665202
1916	A.46.TYR	HA	4.54313305427123
1917	A.46.TYR	HBx	2.94598823808495
1918	A.46.TYR	HBy	3.02062339184027
1919	A.45.LEU	HDx%	0.825245714505351
1920	A.45.LEU	HDy%	0.883304954698109
1921	A.45.LEU	HA	4.21753128108985
1922	A.45.LEU	HBx	1.4221589228884
1923	A.45.LEU	HBy	1.54305575501187
1924	A.45.LEU	HG	1.53478969957695
1925	A.45.LEU	CG	26.8467969866062
1926	A.45.LEU	CDy	24.8200824781519
1927	A.45.LEU	CDx	23.6526291967901
1928	A.44.PRO	HA	4.33817121690423
1929	A.44.PRO	HDx	3.62506499651465
1930	A.44.PRO	HDy	3.82679418860429
1932	A.44.PRO	HGx	1.93849939727023
1934	A.44.PRO	HBy	2.17216698128737
1935	A.44.PRO	CD	51.0064714392915
1938	A.44.PRO	HGy	1.96450512319966
1939	A.44.PRO	HBx	1.65504402853089
1940	A.44.PRO	CG	27.4454221512263
1941	A.42.GLU	HA	4.28001560949753
1942	A.42.GLU	HGx	2.17998633933596
1943	A.42.GLU	HGy	2.24101123727044
1944	A.42.GLU	HBy	2.00391096013637
1945	A.42.GLU	HBx	1.88348904230539
1946	A.42.GLU	CG	36.2420550522492
1947	A.41.GLU	HA	4.25154476526003
1948	A.41.GLU	HGx	2.23316287835864
1949	A.41.GLU	HGy	2.24203391539141
1950	A.41.GLU	HBy	2.01818669153368
1951	A.41.GLU	HBx	1.90102704788915
1952	A.41.GLU	CG	36.2507204571156
1953	A.40.GLN	HA	4.24256265573491
1954	A.40.GLN	HGx	2.36598666800525
1955	A.40.GLN	HGy	2.3744855124977
1956	A.40.GLN	HBx	1.97972986048717
1957	A.40.GLN	HBy	2.06881624992939
1958	A.40.GLN	CG	33.7431352818706

1960	A.39.PRO	HA	4.26568204455371
1961	A.39.PRO	HDx	3.6081192245702
1962	A.39.PRO	HDy	3.61587865806606
1964	A.39.PRO	HBy	2.30307020475375
1967	A.39.PRO	CD	51.0508226845842
1968	A.39.PRO	CB	32.1574922696121
1969	A.39.PRO	HBx	1.89141341194842
1971	A.39.PRO	HGy	2.00865645615717
1972	A.39.PRO	CG	27.42062348002
1973	A.39.PRO	HGx	1.93745024441113
1974	A.37.TYR	HA	4.54288062973584
1975	A.37.TYR	HBx	2.88554529187074
1976	A.37.TYR	HBy	2.97731698248599
1977	A.36.GLY	HAy	3.94804612091165
1978	A.36.GLY	HAx	3.84504450054007
1979	A.35.ASP	HA	4.54752078401741
1980	A.35.ASP	HBx	2.65173720780944
1981	A.35.ASP	HBy	2.69197858655173
1982	A.33.VAL	HA	4.10567040204443
1983	A.33.VAL	HB	2.03873372190062
1984	A.33.VAL	HGx%	0.890896481830088
1985	A.33.VAL	HGy%	0.894463503500104
1986	A.33.VAL	CGy	21.1844987764611
1988	A.30.ALA	HA	4.27521359006069
1990	A.30.ALA	HB%	1.36884135604426
1991	A.27.TRP	HA	4.66675757165083
1992	A.27.TRP	HBx	3.21579519302663
1993	A.27.TRP	HBy	3.22236850837072
1995	A.26.THR	HA	4.26625087375674
1996	A.26.THR	HB	4.20393803892674
1997	A.26.THR	HG2%	1.13732363666313
1998	A.26.THR	CG2	21.6030732140224
1999	A.25.GLY	HAy	4.01810909997585
2000	A.25.GLY	HAx	3.82988349939127
2001	A.24.ASP	HA	4.53575603871603
2002	A.24.ASP	HBx	2.63695945329095
2003	A.24.ASP	HBy	2.80920510773473
2004	A.20.THR	HA	4.3518391520382
2005	A.20.THR	HB	4.14071675243478
2006	A.20.THR	HG2%	1.18533956410324
2007	A.20.THR	CG2	21.7020482219559
2008	A.19.ALA	HA	4.33337760000371
2009	A.19.ALA	HB%	1.40967080097919
2010	A.18.PRO	HA	4.28423696988349
2011	A.18.PRO	HDx	3.6218847436815
2012	A.18.PRO	HDy	3.71417779656822
2014	A.18.PRO	CD	50.9893731553356
2016	A.18.PRO	HBy	2.31500313757588
2017	A.18.PRO	CB	32.2611120768065
2018	A.18.PRO	HBx	1.91092507697091

2019	A.18.PRO	CG	27.3976898242952
2020	A.18.PRO	HGy	1.99986357016566
2021	A.18.PRO	HGx	1.9732209599398
2023	A.16.PHE	HBx	2.98288171185513
2027	A.17.ILE	H	8.20197090076536
2028	A.17.ILE	N	126.419914474837
2029	A.17.ILE	CB	38.9060677102055
2030	A.17.ILE	CA	58.0112800541766
2031	A.17.ILE	C	173.780673825981
2032	A.16.PHE	HBy	2.99513400839312
2033	A.14.GLY	HAy	3.93074723884924
2034	A.14.GLY	HAx	3.87922560107122
2035	A.13.GLU	HA	4.26541012522963
2036	A.13.GLU	HGx	2.25932237642381
2037	A.13.GLU	HGy	2.26731718129974
2038	A.13.GLU	HBx	1.95034165780484
2039	A.13.GLU	HBy	2.08104224706684
2040	A.13.GLU	CG	36.3378613291227
2041	A.11.SER	HA	4.47559683188173
2042	A.11.SER	HBx	3.84108760432861
2043	A.11.SER	HBy	3.95099063565062
2044	A.12.SER	HA	4.43927519729573
2045	A.10.GLN	HA	4.31041150568389
2046	A.10.GLN	HGx	2.32645050307526
2047	A.10.GLN	HGy	2.33633842112661
2048	A.10.GLN	HBx	1.97317116779498
2049	A.10.GLN	HBy	2.09180034062199
2050	A.10.GLN	CG	33.8705120448709
2051	A.9.LEU	HA	4.27276391185128
2052	A.9.LEU	HDx%	0.824092618483796
2053	A.9.LEU	HDy%	0.875682838078301
2054	A.9.LEU	CB	42.4729109046106
2055	A.9.LEU	HBy	1.56365075688043
2056	A.9.LEU	HBx	1.51660453845726
2057	A.9.LEU	CG	26.8422663143255
2058	A.9.LEU	HG	1.46298583500051
2059	A.9.LEU	CDy	24.9084209617418
2060	A.9.LEU	CDx	23.7076215063313
2061	A.8.TYR	HA	4.55258945763831
2062	A.8.TYR	HBx	2.94995711269602
2063	A.8.TYR	HBy	3.015519984748
2064	A.7.THR	HA	4.2697379900044
2065	A.7.THR	HB	4.1632708386058
2066	A.7.THR	HG2%	1.12765230364063
2067	A.7.THR	CG2	21.5801498524196
2068	A.6.SER	HA	4.47971945478136
2069	A.6.SER	HBx	3.82822687139169
2070	A.6.SER	HBy	3.87922242083039

2071	A.5.MET	HA	4.52124642735635
2072	A.5.MET	HGx	2.5222438618065
2073	A.5.MET	HGy	2.5979654369937
2074	A.5.MET	HBx	2.00377156826318
2075	A.5.MET	HBy	2.10089873421638
2076	A.5.MET	CG	32.0147002964033
2077	A.4.GLY	HAx	3.94400543812602
2078	A.4.GLY	HAy	3.96247840541024
2079	A.3.MET	HA	4.46291765250277
2080	A.3.MET	HGx	2.55307621043511
2081	A.3.MET	HGy	2.62164205647838
2082	A.3.MET	HBx	2.01941666679821
2083	A.3.MET	HBy	2.09391119230456
2084	A.3.MET	CG	32.0105577165419
2085	A.2.ALA	HA	4.35513941850525
2086	A.2.ALA	HB%	1.38780872290077
2087	A.15.LYS	HA	4.30764028755821
2088	A.15.LYS	HEx	2.9323355880537
2089	A.15.LYS	HEy	2.94056068489945
2090	A.15.LYS	HGx	1.26543124984384
2091	A.15.LYS	HGy	1.32391236476055
2092	A.15.LYS	CE	42.1368896804696
2093	A.15.LYS	CB	33.3026944030058
2094	A.15.LYS	HBx	1.65957093512442
2095	A.15.LYS	CD	29.0227908395332
2096	A.15.LYS	HDy	1.61911986136718
2097	A.15.LYS	HDx	1.60666027044179
2098	A.15.LYS	CG	24.7225295612547
2099	A.15.LYS	HBy	1.7038847645833
2100	A.29.LYS	HGx	1.39997438106971
2101	A.29.LYS	HGy	1.44023919089102
2102	A.29.LYS	HEx	2.96715964066583
2103	A.29.LYS	HEy	2.97405405066552
2104	A.29.LYS	HA	4.17159109942337
2105	A.29.LYS	CE	42.1397135299687
2107	A.29.LYS	HBy	1.80248762564049
2108	A.29.LYS	CB	33.0670318140782
2109	A.29.LYS	HBx	1.70551670589572
2110	A.29.LYS	CD	29.2651772641778
2111	A.29.LYS	HDy	1.68244676953018
2112	A.29.LYS	HDx	1.66849285711268
2113	A.29.LYS	CG	24.9750532481785
2114	A.21.LYS	HA	4.24115145659247
2115	A.21.LYS	HEx	2.8302592373504
2116	A.21.LYS	HEy	2.83708930542279
2118	A.21.LYS	HGx	1.12103891260454
2119	A.21.LYS	HGy	1.17674627673742
2120	A.21.LYS	CE	41.9982000752788
2121	A.21.LYS	CB	33.1427487209237
2122	A.21.LYS	HBy	1.65532170906307

2123	A.21.LYS	HBx	1.4313405132346
2124	A.21.LYS	CD	29.0664601986358
2125	A.21.LYS	HDy	1.52871158945231
2126	A.21.LYS	HDx	1.51893977608239
2127	A.21.LYS	CG	24.5542702425352
2128	A.134.ASN	HD2y	7.72503765376194
2129	A.134.ASN	ND2	113.146305137694
2130	A.134.ASN	HD2x	6.95198587587986
2131	A.134.ASN	CG	177.057699473995
2134	A.105.VAL	HA	4.06662161514726
2135	A.105.VAL	HB	1.99517601864276
2137	A.105.VAL	HGy%	0.904060544457363
2138	A.105.VAL	HGx%	0.889311722071523
2139	A.105.VAL	CGx	21.2340682992859
2140	A.105.VAL	CGy	21.2568670253385
2141	A.32.ARG	HA	4.3882912818562
2142	A.32.ARG	HDx	3.13202439380689
2143	A.32.ARG	HDy	3.140282786836
2144	A.32.ARG	HBx	1.70798212628495
2145	A.32.ARG	HBy	1.79617370131661
2146	A.32.ARG	HGx	1.54032339218314
2147	A.32.ARG	HGy	1.57633106601907
2148	A.32.ARG	CD	43.3516978993316
2149	A.32.ARG	CG	27.2816809600351
2152	A.16.PHE	HA	4.66344976222597
2153	A.31.ARG	HA	4.33453040003288
2154	A.31.ARG	HDx	3.12003728989685
2155	A.31.ARG	HDy	3.12374505983318
2156	A.31.ARG	HBx	1.70553520514752
2157	A.31.ARG	HBy	1.78161607626499
2158	A.31.ARG	HGx	1.56494763735708
2159	A.31.ARG	HGy	1.57255833860954
2160	A.49.LYS	HA	4.30111574559677
2161	A.49.LYS	HEx	2.93757550134662
2162	A.49.LYS	HEy	2.94379268397881
2163	A.49.LYS	HBx	1.75417221794146
2164	A.49.LYS	HBy	1.8501296578882
2165	A.49.LYS	HDx	1.6221255757545
2166	A.49.LYS	HDy	1.62986716627952
2167	A.49.LYS	HGx	1.37529373593668
2168	A.49.LYS	HGy	1.43767851049646
2169	A.49.LYS	CE	42.1352372087604
2170	A.49.LYS	CD	29.1202333980902
2171	A.49.LYS	CG	24.8309255944565
2172	A.51.LYS	HA	4.24637187036649
2173	A.51.LYS	HEx	2.95959045119871
2174	A.51.LYS	HEy	2.96608920833416
2175	A.51.LYS	HBx	1.71625958079938

2176	A.51.LYS	HBy	1.77670984658294
2177	A.51.LYS	HDx	1.64954052283911
2178	A.51.LYS	HDy	1.65738418570701
2179	A.51.LYS	HGx	1.34095084444654
2180	A.51.LYS	HGy	1.40516101403474
2181	A.51.LYS	CE	42.13692583323
2182	A.51.LYS	CD	29.1454588038029
2183	A.51.LYS	CG	24.79513308558
2184	A.74.LYS	HA	4.1527148327188
2185	A.74.LYS	HEx	2.97739700836304
2186	A.74.LYS	HEy	2.98966158046763
2187	A.74.LYS	HBx	1.83427206581047
2188	A.74.LYS	HBy	1.88748873335138
2189	A.74.LYS	HDx	1.66911590829895
2190	A.74.LYS	HDy	1.68784423617213
2191	A.74.LYS	HGx	1.38326663793932
2192	A.74.LYS	HGy	1.47439801250838
2193	A.74.LYS	CE	42.1724227814078
2194	A.74.LYS	CD	29.1461001780259
2195	A.74.LYS	CG	24.8015735076364
2196	A.75.LYS	HA	4.18242383618892
2197	A.75.LYS	HEx	2.97242950270746
2198	A.75.LYS	HEy	2.98632206598259
2199	A.75.LYS	HBx	1.83523265900631
2200	A.75.LYS	HBy	1.84846346962967
2201	A.75.LYS	HDx	1.68218844334972
2202	A.75.LYS	HDy	1.69327374056382
2203	A.75.LYS	HGx	1.43190105289413
2204	A.75.LYS	HGy	1.51633524371425
2205	A.75.LYS	CE	42.1418276146677
2206	A.75.LYS	CD	29.1102548432926
2207	A.75.LYS	CG	24.9482273363124
2208	A.79.LYS	HA	4.13074945949734
2209	A.79.LYS	HEx	2.96322323938393
2210	A.79.LYS	HEy	2.96729029776012
2211	A.79.LYS	HBx	1.86772052165112
2212	A.79.LYS	HBy	1.92151735266308
2213	A.79.LYS	HDx	1.68710286273561
2214	A.79.LYS	HDy	1.69486106482796
2215	A.79.LYS	HGx	1.38982162920295
2216	A.79.LYS	HGy	1.51998178958334
2217	A.79.LYS	CE	42.1365863992921
2218	A.79.LYS	CD	29.08960470504
2219	A.79.LYS	CG	24.8976639081556
2220	A.86.LYS	HA	4.24587661508435
2221	A.86.LYS	HEx	2.98591749393828
2222	A.86.LYS	HEy	2.99768363298057
2223	A.86.LYS	HBx	1.7880845815864
2224	A.86.LYS	HBy	1.8409905353345
2225	A.86.LYS	HDx	1.68352619347188

2226	A.86.LYS	HDy	1.68994275184332
2227	A.86.LYS	HGx	1.43468113745216
2228	A.86.LYS	HGy	1.49031311905671
2229	A.86.LYS	CE	42.0993290932291
2230	A.86.LYS	CD	29.1386875990413
2231	A.86.LYS	CG	24.7938981537239
2232	A.87.LYS	HA	4.24096834519579
2233	A.87.LYS	HEx	2.9738997549372
2234	A.87.LYS	HEy	2.98815008626038
2235	A.87.LYS	HBx	1.79168591754627
2236	A.87.LYS	HBy	1.83367984153824
2237	A.87.LYS	HDx	1.66792791637882
2238	A.87.LYS	HDy	1.68192550241524
2239	A.87.LYS	HGx	1.3956355061739
2240	A.87.LYS	HGy	1.46535154836129
2241	A.87.LYS	CE	42.1310917051936
2242	A.87.LYS	CD	29.1687009883863
2243	A.87.LYS	CG	24.803442614699
2244	A.115.LYS	HBx	1.73853178903069
2245	A.115.LYS	HBy	1.7954516118522
2246	A.115.LYS	HDx	1.67648090257516
2247	A.115.LYS	HDy	1.69461039086561
2249	A.115.LYS	HGy	1.46841590608316
2250	A.115.LYS	CD	29.1360233446742
2251	A.138.LYS	HEx	2.98590279387102
2252	A.138.LYS	HEy	2.99020032572774
2253	A.138.LYS	HBy	1.87209072403115
2254	A.138.LYS	HBx	1.77771947290138
2255	A.138.LYS	HDy	1.68126955237997
2256	A.138.LYS	HDx	1.66841453260597
2257	A.138.LYS	HGx	1.43201431156344
2258	A.138.LYS	HGy	1.46588271585053
2259	A.138.LYS	CE	42.1357548559197
2260	A.138.LYS	CD	29.0109473238479
2262	A.138.LYS	CG	24.7941498730542
2264	A.34.LYS	HA	4.24391245273453
2265	A.90.LYS	HGy	1.45396628211362
2266	A.34.LYS	HEx	2.95384856195857
2267	A.34.LYS	HEy	2.96293488786483
2268	A.34.LYS	HBy	1.80613285496752
2269	A.34.LYS	HBx	1.74057316058908
2270	A.34.LYS	HDx	1.64252255458614
2271	A.34.LYS	HDy	1.65401321455122
2272	A.34.LYS	HGx	1.39023090539793
2273	A.34.LYS	HGy	1.40057465185625
2274	A.34.LYS	CE	42.133464780928
2275	A.34.LYS	CD	29.1428160704768
2276	A.34.LYS	CG	24.7808781136666

2277	A.114.PRO	HA	4.39746169978169
2278	A.114.PRO	HDx	3.62644475016197
2279	A.114.PRO	HDy	3.81295149297541
2280	A.114.PRO	CD	50.4681271408809
2282	A.114.PRO	HBy	2.27957667889939
2283	A.114.PRO	CB	32.1146749169004
2284	A.114.PRO	HBx	1.85975148993459
2285	A.114.PRO	CG	27.3704113337053
2286	A.114.PRO	HGx	1.99759593826215
2287	A.114.PRO	HGy	2.01024890116111
2288	A.67.PRO	HA	4.38860210479927
2289	A.67.PRO	HDx	3.79128813373481
2290	A.67.PRO	HDy	3.8053419779187
2291	A.67.PRO	CD	50.5414570287207
2293	A.67.PRO	HBy	2.29903526962645
2294	A.67.PRO	CB	32.1145346342141
2295	A.67.PRO	HBx	1.90536954617886
2297	A.67.PRO	HGy	2.02392180648035
2298	A.67.PRO	CG	27.3710456150122
2299	A.67.PRO	HGx	2.0056786434634
2300	A.23.PRO	HA	4.35356740095581
2301	A.23.PRO	HDx	3.63597947277632
2302	A.23.PRO	HDy	3.79226041653669
2303	A.23.PRO	CD	50.8354564961771
2309	A.23.PRO	HGx	1.9634396272495
2310	A.23.PRO	CG	27.3744548067097
2311	A.23.PRO	HGy	2.03183792941796
2315	A.23.PRO	HBy	2.2788333508297
2316	A.23.PRO	CB	31.878137755025
2317	A.23.PRO	HBx	1.91442350264106
2318	A.28.ARG	HA	4.16432059391158
2319	A.28.ARG	HDx	3.05743652921175
2320	A.28.ARG	HDy	3.06833373686468
2321	A.28.ARG	CD	43.3456109583916
2322	A.28.ARG	CB	30.9204974096115
2323	A.28.ARG	HBy	1.70912250334489
2324	A.28.ARG	HBx	1.52341743419667
2325	A.28.ARG	CG	27.1525593360096
2326	A.28.ARG	HGy	1.34162529327919
2327	A.28.ARG	HGx	1.32711045418715
2328	A.31.ARG	CD	43.3716835789216
2329	A.31.ARG	CG	27.0887772379762
2330	A.154.ASP	HA	4.86703363123131
2331	A.154.ASP	HBx	2.56455691024574
2332	A.154.ASP	HBy	2.77126035721046
2333	A.111.CYS	HA	4.79198413815257
2334	A.111.CYS	HBx	2.82844303722973
2335	A.111.CYS	HBy	2.91265173094798
2336	A.38.VAL	HA	4.2959127764476
2337	A.38.VAL	HB	1.92158026796083

2338	A.38.VAL	CGx	20.4313252807347
2339	A.38.VAL	HGx%	0.84747910507244
2340	A.38.VAL	CGy	21.0652150600384
2341	A.38.VAL	HGy%	0.87138421443853
2342	A.43.VAL	HA	4.33192363861807
2343	A.43.VAL	CB	32.5302311111384
2344	A.43.VAL	HB	2.04872048625296
2345	A.43.VAL	CGx	20.6683566441739
2346	A.43.VAL	HGx%	0.91871833169369
2347	A.43.VAL	HGy%	0.944128428901217
2348	A.43.VAL	CGy	21.0622775236685
2349	A.61.VAL	HA	4.42693866858522
2350	A.61.VAL	CB	32.7300363526552
2351	A.61.VAL	HB	2.04768101390007
2352	A.61.VAL	CGx	20.5490115526309
2353	A.61.VAL	HGx%	0.892614718427411
2354	A.61.VAL	CGy	21.0994259259789
2355	A.61.VAL	HGy%	0.957262764134636
2356	A.100.ALA	HB%	1.36280077235359
2357	A.100.ALA	HA	4.56345523432659
2358	A.98.ALA	CB	18.0896162063268
2359	A.98.ALA	HB%	1.35119077750226
2360	A.98.ALA	CA	50.3926253447292
2361	A.98.ALA	HA	4.56440553929285
2362	A.106.MET	HA	4.81186887264659
2363	A.106.MET	CB	32.2642581522551
2364	A.106.MET	HBx	1.92770700100788
2365	A.106.MET	HBy	2.03397232003491
2368	A.106.MET	HGx	2.52852853911184
2369	A.106.MET	CG	32.0476722060328
2370	A.106.MET	HGy	2.63991040352246
2371	A.17.ILE	HA	4.39209600187596
2372	A.17.ILE	HB	1.75653619009029
2373	A.17.ILE	CD1	12.6329383677548
2374	A.17.ILE	HD1%	0.78941101685302
2375	A.17.ILE	CG2	17.0444952567612
2376	A.17.ILE	HG2%	0.868990266365021
2377	A.17.ILE	HG1x	1.08154356883261
2379	A.17.ILE	CG1	26.8307861374674
2380	A.17.ILE	HG1y	1.4316932494178
2381	A.22.ARG	CA	54.3156157286852
2382	A.22.ARG	HA	4.53842771116961
2383	A.22.ARG	HDx	3.10802251853016
2384	A.22.ARG	CD	43.3231541986783
2385	A.22.ARG	HDy	3.11416322069641
2386	A.22.ARG	HBx	1.74746156139222
2387	A.22.ARG	HBy	1.81924956446952
2389	A.22.ARG	HGy	1.63266783815629

2390	A.22.ARG	CG	27.1250852914433
2391	A.22.ARG	HGx	1.58503466019544
2392	A.95.GLN	HBx	1.92865180177496
2393	A.95.GLN	CB	28.8334970743756
2394	A.95.GLN	HBy	2.09402097929724
2395	A.95.GLN	CG	33.425249928749
2396	A.95.GLN	HGx	2.39417815117063
2397	A.95.GLN	HGy	2.4006370743586
2398	A.95.GLN	HA	4.60634061998682
2400	A.126.GLN	HBx	1.92376607590428
2401	A.126.GLN	CB	28.8010573231659
2402	A.126.GLN	HBy	2.09464803523008
2403	A.126.GLN	CG	33.4175233899395
2404	A.126.GLN	HGx	2.39309607176446
2405	A.126.GLN	HGy	2.40041339038111
2406	A.126.GLN	CA	53.6872131536835
2407	A.126.GLN	HA	4.61270625541253
2408	A.66.CYS	HA	4.74278427307144
2409	A.66.CYS	CB	27.5317190927354
2410	A.66.CYS	HBx	2.92537969949959
2411	A.66.CYS	HBy	2.92889733486219
2412	A.12.SER	HBx	3.9159719527844
2413	A.12.SER	HBy	3.9477069100664
2416	A.160.LYS	HA	4.15273396846479
2417	A.160.LYS	CA	57.5506570271811
2418	A.160.LYS	HE2	2.97920840754687
2419	A.160.LYS	HE3	2.97920840754687
2420	A.160.LYS	HG2	1.36936722229002
2421	A.160.LYS	HG3	1.36936722229002
2422	A.160.LYS	HBx	1.69834482767221
2423	A.160.LYS	HBy	1.81016225542144
2424	A.160.LYS	HD2	1.6560518049611
2425	A.160.LYS	HD3	1.6560518049611
2426	A.160.LYS	CE	42.2067916292002
2427	A.160.LYS	CD	29.1854520172919
2428	A.160.LYS	CG	24.6613491402907
2429	A.27.TRP	HD1	7.28943058002972
2430	A.53.PHE	HD%	7.20519156363846
2431	A.16.PHE	HD%	7.16021220405562
2432	A.37.TYR	HD%	7.0463652237999
2433	A.8.TYR	HD%	7.07768041894237
2434	A.46.TYR	HD%	7.07767402342037
2435	A.37.TYR	HE%	6.78091598843736
2436	A.8.TYR	HE%	6.79479330632421
2437	A.46.TYR	HE%	6.79480652335073
2438	A.37.TYR	CD%	133.1571172281
2439	A.37.TYR	CE%	118.088703212969
2440	A.53.PHE	CD%	131.889353448044
2441	A.16.PHE	CD%	131.81937498849
2442	A.8.TYR	CD%	133.222913150558

2443	A.46.TYR	CD%	133.222913150558
2444	A.8.TYR	CE%	118.208369509152
2445	A.46.TYR	CE%	118.208369509152
2446	A.27.TRP	CD1	127.262716757905
2447	A.27.TRP	HZ2	7.46719172281609
2448	A.27.TRP	CZ2	114.595659188224
2474	A.27.TRP	HE1	10.183645699476
2475	A.27.TRP	NE1	129.593357506976

Acknowledgment

The journey of my Ph.D. has started with the acceptance of my application by Prof. Dr. Teresa Carlomagno. I am grateful for this opportunity, to work in your group and be a part of A.K. Carlomagno member, and I cherish the experience throughout. I have been inspired by your work and the dedication you put into the scientific platform, which is commendable. I enjoy the academic discussion in the seminar sessions. I also would like to thank you for being so kind and understanding, that in lab outings Indian foods and celebrating Indian festivals were organized. Throughout this journey I have learned a lot on both academic and non-academic platforms, as being a foreigner in Germany was a great learning experience.

Dr. John Kirkpatrick, I can't thank you enough, for being so kind and humble, in every situation. Learning NMR with you was a great experience, that I always cherish. From setting up the experiment to analyzing the data, answering my stupid questions, and so on. Again, I can't thank you enough. You are admirable.

Dr. Alexander Marchanka, Thank you for being always helpful. My RNA research work has started with you. You being my RNA synthesis teacher was a fortunate opportunity. Thank you for being kind and helpful. Always grateful for you.

Dr. Luca Codutti, Thank you, Luca, Barbara, Martino, and Dario. Thank you for being so kind. I always cherish our Play sessions, with Natasha, Andrea, and with you all. I was welcomed whole with whole heart by you all. I can only have love and great respect for you all. From sharing lab benches to discussing work projects. It was a great experience.

Natasha… You were the one who spoke to me first when I joined the lab. You were the one who understood my anxiety being in a completely new place. You sat with me for my first protein purification in the lab, I can't thank you enough. From discussing work to playing games, it was a wonderful time. Andrea, thank you for being so kind and humble. I felt like I was one of the members in the house. This feeling was really special and I am very much grateful.

Deepshikha, thank you so much for being who you are. I always enjoyed our company. Thank you for everything. Megha, Thank you so much. You are the one with whom I could speak in my language. I am thankful for the lovely delicious dinner you cooked only for me when I was feeling a little low.

Simone, thank you for being my bio-hero. Olga, you are special to me. Thank you for being my friend. Thank you, Petra, Juliane, Arun, Michealanglo, Georg, Marco, Ying, Neha, Mumdooh, Phillip, Susanne, and Maria, for being great lab mates. Thank you for all the help and learning experience.

Frau Kandil and Dr. Katja, thank you so much for being so helpful and kind. I would like to thank all the members of BMWZ and Leibniz Universität Hanover.

Steffi, you are the window with which I use to see Germany and India in Germany. Thank you so much for being my friend, without you it could have been really difficult. Until this moment you are my constant support. Thank you so much dear. Prajna, thank you for submitting the documents and for all the help.
Parmida, you are an angel. I am really thankful for you for everything. I am fortunate that I met you. You inspired me a lot. I know I have a friend in the world, however far we may be.
I am very much thankful for M.M. Arts & Science college, for letting me be part of your member gain. I could able to find myself again and rebuild myself again. Always would be grateful. This moment wouldn't have been possible without your support.
Now, I would like to dedicate this to you, my heartbeats amma, appa puttu kanda, av, chaami, kendamahasati amnoru, baaba. This and all the goodness is for you. Vaniganeshanna thank you for all the support. I thank each and every one of you, because of you all, this moment could be possible. Trunamapi nachalathi tenavina.

Bibliography

1. RNA Processing, Post-Transcriptional - MeSH - NCBI. https://www.ncbi.nlm.nih.gov/mesh?Db=mesh&Cmd=DetailsSearch&Term=%22RNA+Processing,+Post-Transcriptional%22%5BMeSH+Terms%5D.
2. Potter, K., Cremona, N. & Wise, J. A. Messenger RNA Processing in Eukaryotes. in *Encyclopedia of Biological Chemistry (Second Edition)* (eds. Lennarz, W. J. & Lane, M. D.) 59–64 (Academic Press, 2013). doi:10.1016/B978-0-12-378630-2.00627-7.
3. Yalamanchili, H. K. *et al.* Chapter Nine - A computational pipeline to infer alternative poly-adenylation from 3′ sequencing data. in *Methods in Enzymology* (ed. Tian, B.) vol. 655 185–204 (Academic Press, 2021).
4. Berk, A. J. Discovery of RNA splicing and genes in pieces. *PNAS* **113**, 801–805 (2016).
5. Berg, J. M., Tymoczko, J. L. & Stryer, L. RNA Synthesis and Splicing. *Biochemistry. 5th edition* (2002).
6. Scherrer, K., Latham, H. & Darnell, J. E. DEMONSTRATION OF AN UNSTABLE RNA AND OF A PRECURSOR TO RIBOSOMAL RNA IN HELA CELLS*. *Proc Natl Acad Sci U S A* **49**, 240–248 (1963).
7. Harris, H., Watts, J. W. & Abraham, E. P. The relationship between nuclear and cytoplasmic ribonucleic acid. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **156**, 109–121 (1962).
8. Soeiro, E., Birnboim, H. C. & Darnell, J. E. Rapidly labeled HeLa cell nuclear RNA: II. Base composition and cellular localization of a heterogeneous RNA fraction. *Journal of Molecular Biology* **19**, 362–372 (1966).
9. Edmonds, M., Vaughan, M. H. & Nakazato, H. Polyadenylic Acid Sequences in the Heterogeneous Nuclear RNA and Rapidly-Labeled Polyribosomal RNA of HeLa

Cells: Possible Evidence for a Precursor Relationship. *PNAS* **68**, 1336–1340 (1971).

10. Rottman, F., Shatkin, A. J. & Perry, R. P. Sequences containing methylated nucleotides at the 5′ termini of messenger RNAs: Possible implications for processing. *Cell* **3**, 197–199 (1974).
11. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *PNAS* **74**, 3171–3175 (1977).
12. Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation - PubMed. https://pubmed.ncbi.nlm.nih.gov/9374866/.
13. Jansen, R.-P. mRNA localization: message on the move. *Nature Reviews Molecular Cell Biology* **2**, 247–256 (2001).
14. Martin, K. C. & Ephrussi, A. mRNA Localization: Gene Expression in the Spatial Dimension. *Cell* **136**, 719 (2009).
15. Medioni, C., Mowry, K. & Besse, F. Principles and roles of mRNA localization in animal development. *Development* **139**, 3263–3276 (2012).
16. Parton, R. M., Davidson, A., Davis, I. & Weil, T. T. Subcellular mRNA localisation at a glance. *Journal of Cell Science* **127**, 2127–2133 (2014).
17. Hughes, S. C. & Simmonds, A. J. Drosophila mRNA Localization During Later Development: Past, Present, and Future. *Frontiers in Genetics* **10**, 135 (2019).
18. Turner-Bridger, B., Caterino, C. & Cioni, J.-M. Molecular mechanisms behind mRNA localization in axons. *Open Biology* **10**, 200177.
19. Di Liegro, C. M., Schiera, G. & Di Liegro, I. Regulation of mRNA transport, localization and translation in the nervous system of mammals (Review). *International Journal of Molecular Medicine* **33**, 747–762 (2014).
20. Kloc, M., Zearfoss, N. R. & Etkin, L. D. Mechanisms of Subcellular mRNA Localization. *Cell* **108**, 533–544 (2002).
21. Kuhl, D. & Skehel, P. Dendritic localization of mRNAs. *Current Opinion in Neurobiology* **8**, 600–606 (1998).

22. Bashirullah, A., Cooperstock, R. L. & Lipshitz, H. D. Rna Localization in Development. *Annual Review of Biochemistry* **67**, 335–394 (1998).

23. Plant Paralog to Viral Movement Protein That Potentiates Transport of mRNA into the Phloem. https://www.science.org/doi/abs/10.1126/science.283.5398.94.

24. Mating Type Switching in Yeast Controlled by Asymmetric Localization of ASH1 mRNA. https://www.science.org/doi/abs/10.1126/science.277.5324.383.

25. Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast | Nature. https://www.nature.com/articles/38015.

26. Driever, W. & Nüsslein-Volhard, C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54**, 95–104 (1988).

27. Ephrussi, A., Dickinson, L. K. & Lehmann, R. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* **66**, 37–50 (1991).

28. Assembly of Tropomyosin Isoforms into the Cytoskeleton of Avian Muscle Cells | Pediatric Research. https://www.nature.com/articles/pr19982176.

29. Messenger RNA targeting of rice seed storage proteins to specific ER subdomains | Nature. https://www.nature.com/articles/35037633.

30. Ephrussi, A. & Lehmann, R. Induction of germ cell formation by oskar. *Nature* **358**, 387–392 (1992).

31. Lehmann, R. & Nüsslein-Volhard, C. Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila. *Cell* **47**, 141–152 (1986).

32. Overexpression of oskar directs ectopic activation of nanos and presumptive pole cell formation in Drosophila embryos - PubMed. https://pubmed.ncbi.nlm.nih.gov/1516136/.

33. Kim-Ha, J., Smith, J. L. & Macdonald, P. M. oskar mRNA is localized to the posterior pole of the Drosophila oocyte. *Cell* **66**, 23–35 (1991).

34. Kim-Ha, J., Kerr, K. & Macdonald, P. M. Translational regulation of oskar mRNA by bruno, an ovarian RNA-binding protein, is essential. *Cell* **81**, 403–412 (1995).

35. Jansen, M., de Moor, C. H., Sussenbach, J. S. & van den Brande, J. L. Translational control of gene expression. *Pediatr Res* **37**, 681–686 (1995).

36. Rongo, C., Gavis, E. R. & Lehmann, R. Localization of oskar RNA regulates oskar translation and requires Oskar protein. *Development* **121**, 2737–2746 (1995).

37. Snee, M. J., Harrison, D., Yan, N. & Macdonald, P. M. A late phase of Oskar accumulation is crucial for posterior patterning of the Drosophila embryo, and is blocked by ectopic expression of Bruno. *Differentiation* **75**, 246–255 (2007).

38. A late phase of germ plasm accumulation during Drosophila oogenesis requires Lost and Rumpelstiltskin | Development. https://dev.biologists.org/content/138/16/3431.

39. Theurkauf, W. E., Alberts, B. M., Jan, Y. N. & Jongens, T. A. A central role for microtubules in the differentiation of Drosophila oocytes. *Development* **118**, 1169–1180 (1993).

40. In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization - PubMed. https://pubmed.ncbi.nlm.nih.gov/18775316/.

41. Hachet, O. & Ephrussi, A. Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* **428**, 959–963 (2004).

42. Jambor, H., Mueller, S., Bullock, S. L. & Ephrussi, A. A stem-loop structure directs oskar mRNA to microtubule minus ends. *RNA* **20**, 429–439 (2014).

43. Ghosh, S., Marchand, V., Gáspár, I. & Ephrussi, A. Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA. *Nat. Struct. Mol. Biol.* **19**, 441–449 (2012).

44. Tange, T. Ø., Nott, A. & Moore, M. J. The ever-increasing complexities of the exon junction complex. *Current Opinion in Cell Biology* **16**, 279–284 (2004).

45. 5′ exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. http://genesdev.cshlp.org/content/16/21/2778.

46. Shibuya, T., Tange, T. Ø., Sonenberg, N. & Moore, M. J. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nature Structural & Molecular Biology* **11**, 346–351 (2004).

47. Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. *The EMBO Journal* **19**, 6860–6869 (2000).

48. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M. J. The exon–exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J* **20**, 4987–4997 (2001).

49. Hir, H. L., Saulière, J. & Wang, Z. The exon junction complex as a node of post-transcriptional networks. *Nature Reviews Molecular Cell Biology* **17**, 41–54 (2016).

50. Wang, Z., Ballut, L., Barbosa, I. & Le Hir, H. Exon Junction Complexes can have distinct functional flavours to regulate specific splicing events. *Sci Rep* **8**, 9509 (2018).

51. Boehm, V. & Gehring, N. H. Exon Junction Complexes: Supervising the Gene Expression Assembly Line. *Trends in Genetics* **32**, 724–735 (2016).

52. Schlautmann, L. P. & Gehring, N. H. A Day in the Life of the Exon Junction Complex. *Biomolecules* **10**, 866 (2020).

53. Intrinsically disordered proteins in cellular signalling and regulation | Nature Reviews Molecular Cell Biology. https://www.nature.com/articles/nrm3920.

54. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* **37**, 215–246 (2008).

55. Micelle-induced Folding of Spinach Thylakoid Soluble Phosphoprotein of 9 kDa and its Functional Implications ,. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2533273/.

56. Salmon, L. *et al.* NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **132**, 8407–8418 (2010).

57. Kosol, S., Contreras-Martos, S., Cedeño, C. & Tompa, P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **18**, 10802–10828 (2013).

58. Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Current Opinion in Structural Biology* **19**, 23–30 (2009).

59. Fu, B. & Vendruscolo, M. Structure and Dynamics of Intrinsically Disordered Proteins. *Adv Exp Med Biol* **870**, 35–48 (2015).

60. Jensen, M. R., Ruigrok, R. W. H. & Blackledge, M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol* **23**, 426–435 (2013).

61. Kurzbach, D., Kontaxis, G., Coudevylle, N. & Konrat, R. NMR Spectroscopic Studies of the Conformational Ensembles of Intrinsically Disordered Proteins. *Adv. Exp. Med. Biol.* **870**, 149–185 (2015).

62. Toto, A. *et al.* Templated folding of intrinsically disordered proteins. *Journal of Biological Chemistry* **295**, 6586–6593 (2020).

63. Bono, F. *et al.* Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. *EMBO Rep.* **5**, 304–310 (2004).

64. Gehring, N. H., Lamprinaki, S., Kulozik, A. E. & Hentze, M. W. Disassembly of exon junction complexes by PYM. *Cell* **137**, 536–548 (2009).

65. Ghosh, S., Obrdlik, A., Marchand, V. & Ephrussi, A. The EJC binding and dissociating activity of PYM is regulated in Drosophila. *PLoS Genet.* **10**, e1004455 (2014).

66. Palacios, I. M., Gatfield, D., St Johnston, D. & Izaurralde, E. An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature* **427**, 753–757 (2004).

67. Fribourg, S., Gatfield, D., Izaurralde, E. & Conti, E. A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat Struct Biol* **10**, 433–439 (2003).

68. Lau, C.-K., Diem, M. D., Dreyfuss, G. & Van Duyne, G. D. Structure of the Y14-Magoh core of the exon junction complex. *Curr Biol* **13**, 933–941 (2003).

69. Shi, H. & Xu, R.-M. Crystal structure of the Drosophila Mago nashi-Y14 complex. *Genes Dev.* **17**, 971–976 (2003).

70. Zhang, J., Sun, X., Qian, Y., LaDuca, J. P. & Maquat, L. E. At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. *Mol Cell Biol* **18**, 5272–5283 (1998).

71. Matsumoto, K., Wassarman, K. M. & Wolffe, A. P. Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO J* **17**, 2107–2121 (1998).

72. Besse, F. & Ephrussi, A. Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nat Rev Mol Cell Biol* **9**, 971–980 (2008).

73. Pokrywka, N. J. & Stephenson, E. C. Microtubules Are a General Component of mRNA Localization Systems in Drosophila Oocytes. *Developmental Biology* **167**, 363–370 (1995).

74. Levine, B. S., Long, R. & Chung, H. Subchronic oral toxicity of pyridostigmine bromide in rats. *Biomed Environ Sci* **4**, 283–289 (1991).

75. Hudson, M. A. *et al.* Choice of an optimal diluent for intravesical bacillus Calmette-Guerin administration. *J Urol* **142**, 1438–1441 (1989).

76. The case-control study: consensus and controversy. Final discussion by all participants. *J Chronic Dis* **32**, 117–137 (1979).

77. Hachet, O. & Ephrussi, A. Drosophila Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr. Biol.* **11**, 1666–1674 (2001).

78. Ohnishi, K., Kutsukake, K., Suzuki, H. & Iino, T. Gene fliA encodes an alternative sigma factor specific for flagellar operons in Salmonella typhimurium. *Mol Gen Genet* **221**, 139–147 (1990).

79. Shapira, S. K., Vercelli, D., Jabara, H. H., Fu, S. M. & Geha, R. S. Molecular analysis of the induction of immunoglobulin E synthesis in human B cells by interleukin 4 and engagement of CD40 antigen. *J Exp Med* **175**, 289–292 (1992).

80. Polarization of both major body axes in Drosophila by gurken-torpedo signalling | Nature. https://www.nature.com/articles/375654a0.

81. Roth, S., Shira Neuman-Silberberg, F., Barcelo, G. & Schüpbach, T. cornichon and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pattern formation in Drosophila. *Cell* **81**, 967–978 (1995).

82. Gerothanassis, I. P., Troganis, A., Exarchou, V. & Barbarossou, K. NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY: BASIC PRINCIPLES AND PHENOMENA, AND THEIR APPLICATIONS TO CHEMISTRY, BIOLOGY AND MEDICINE. *Chem. Educ. Res. Pract.* **3**, 229–252 (2002).

83. High Resolution NMR - 3rd Edition. https://www.elsevier.com/books/high-resolution-nmr/becker/978-0-12-084662-7.

84. Basic 1H- and 13C-NMR Spectroscopy - 1st Edition. https://www.elsevier.com/books/basic-1h-and-13c-nmr-spectroscopy/balci/978-0-444-51811-8.

85. Modern NMR Techniques for Chemistry Research, Volume 6 - 1st Edition. https://www.elsevier.com/books/modern-nmr-techniques-for-chemistry-research/derome/978-0-08-032513-2.

86. NMR Imaging in Biomedicine - 1st Edition. https://www.elsevier.com/books/nmr-imaging-in-biomedicine/mansfield/978-0-12-025562-7.

87. Mureddu, L. & Vuister, G. W. Simple high-resolution NMR spectroscopy as a tool in molecular biology. *FEBS J.* **286**, 2035–2042 (2019).

88. Nilges, M. Structure calculation from NMR data. *Current Opinion in Structural Biology* **6**, 617–623 (1996).

89. Bardiaux, B., Malliavin, T. & Nilges, M. ARIA for solution and solid-state NMR. *Methods Mol Biol* **831**, 453–483 (2012).

90. Allain, F., Mareuil, F., Ménager, H., Nilges, M. & Bardiaux, B. ARIAweb: a server for automated NMR structure calculation. *Nucleic Acids Research* **48**, W41–W47 (2020).

91. Automated NMR Structure Calculation With CYANA | SpringerLink. https://link.springer.com/protocol/10.1385/1-59259-809-9:353.

92. High-resolution structure determination of the CylR2 homodimer using paramagnetic relaxation enhancement and structure-based prediction of molecular alignment. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2758389/.

93. Sugiki, T., Kobayashi, N. & Fujiwara, T. Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists. *Computational and Structural Biotechnology Journal* **15**, 328–339 (2017).

94. Pomin, V. H. NMR Chemical Shifts in Structural Biology of Glycosaminoglycans. *Anal. Chem.* **86**, 65–94 (2014).

95. Kang, C. Applications of In-Cell NMR in Structural Biology and Drug Discovery. *Int J Mol Sci* **20**, 139 (2019).

96. Carlomagno, T. Present and future of NMR for RNA–protein complexes: A perspective of integrated structural biology. *Journal of Magnetic Resonance* **241**, 126–136 (2014).

97. RNA structure determination by solid-state NMR spectroscopy | Nature Communications. https://www.nature.com/articles/ncomms8024.

98. Williams, B., II *et al.* Structure modeling of RNA using sparse NMR constraints. *Nucleic Acids Research* **45**, 12638–12647 (2017).

99. Marion, D. An Introduction to Biological NMR Spectroscopy. *Mol Cell Proteomics* **12**, 3006–3025 (2013).

100. Lipchock, J. M. & Loria, J. P. Monitoring molecular interactions by NMR. *Methods Mol Biol* **490**, 115–134 (2009).

101. New techniques in structural NMR — anisotropic interactions | Nature Structural & Molecular Biology. https://www.nature.com/articles/nsb0798_517.

102. Markwick, P. R. L., Malliavin, T. & Nilges, M. Structural Biology by NMR: Structure, Dynamics, and Interactions. *PLoS Comput Biol* **4**, e1000168 (2008).

103. NMR mapping of protein interactions in living cells | Nature Methods. https://www.nature.com/articles/nmeth0206-80?proof=tNature.

104. Mikla, V. I. & Mikla, V. V. 3 - Physics of Magnetic Resonance Imaging. in *Medical Imaging Technology* (eds. Mikla, V. I. & Mikla, V. V.) 39–52 (Elsevier, 2014). doi:10.1016/B978-0-12-417021-6.00003-4.

105. Meyer-Baese, A. & Schmid, V. Chapter 1 - Introduction*This chapter contains material reprinted from chapter 1 of Biomedical Signal Analysis: Contemporary Methods and Applications, by Fabian Theis and Anke Meyer-Base, published by The MIT Press. Reprinted with permission from MIT Press.*. in *Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition)* (eds. Meyer-Baese, A. & Schmid, V.) 1–20 (Academic Press, 2014). doi:10.1016/B978-0-12-409545-8.00001-7.

106. Charles, S. W. MAGNETIC FLUIDS (Ferrofluids). in *Magnetic Properties of Fine Particles* (eds. Dormann, J. L. & Fiorani, D.) 267–276 (Elsevier, 1992). doi:10.1016/B978-0-444-89552-3.50035-9.

107. Afanas’ev, A. M. THEORETICAL POINT OF VIEW ON RELAXATION IN SMALL PARTICLES. in *Magnetic Properties of Fine Particles* (eds. Dormann, J. L. & Fiorani, D.) 13–20 (Elsevier, 1992). doi:10.1016/B978-0-444-89552-3.50007-4.

108. Camuffo, D. Chapter 4 - Consequences of the Maxwell–Boltzmann Distribution. in *Microclimate for Cultural Heritage (Third Edition)* (ed. Camuffo, D.) 61–71 (Elsevier, 2019). doi:10.1016/B978-0-444-64106-9.00004-3.

109. THE PHYSICS OF INFORMATION - ScienceDirect. https://www.sciencedirect.com/science/article/pii/B9780444517265500200.

110. Theory of NMR Spectroscopy. http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/theory.htm.

111. Hore, P. J. NMR Principles. in *Encyclopedia of Spectroscopy and Spectrometry* (ed. Lindon, J. C.) 1545–1553 (Elsevier, 1999). doi:10.1006/rwsp.2000.0206.

112. Dong, R. Y. NMR Relaxation Rates. in *Encyclopedia of Spectroscopy and Spectrometry* (ed. Lindon, J. C.) 1568–1575 (Elsevier, 1999). doi:10.1006/rwsp.2000.0264.

113. Autschbach, J. Chapter 4 - Relativistic Effects on NMR Parameters. in *Science and Technology of Atomic, Molecular, Condensed Matter & Biological Systems* (ed. Contreras, R. H.) vol. 3 69–117 (Elsevier, 2013).

114. Klein, W. Nuclear magnetic resonance: Free-induction decay and spin echoes in a 0.05-T magnetic field. *American Journal of Physics* **58**, 143–147 (1990).

115. Free induction decay. *Wikipedia* (2021).

116. Bainbridge, A. NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY | Overview. in *Encyclopedia of Analytical Science (Second Edition)* (eds. Worsfold, P., Townshend, A. & Poole, C.) 203–211 (Elsevier, 2005). doi:10.1016/B0-12-369397-7/00404-0.

117. NMR Data Processing - ScienceDirect. https://www.sciencedirect.com/science/article/pii/B0122266803003550.

118. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1814**, 942–968 (2011).

119. High Pressure Studies Using NMR Spectroscopy - ScienceDirect. https://www.sciencedirect.com/science/article/pii/B0122266803001174.

120. Wong, V. & Case, D. A. Chapter 8 - Comparing MD Simulations and NMR Relaxation Parameters. in *Annual Reports in Computational Chemistry* (eds. Wheeler, R. A. & Spellmeyer, D. C.) vol. 4 139–154 (Elsevier, 2008).

121. Bloembergen, N., Purcell, E. M. & Pound, R. V. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Phys. Rev.* **73**, 679–712 (1948).

122. Chavhan, G. B., Babyn, P. S., Thomas, B., Shroff, M. M. & Haacke, E. M. Principles, Techniques, and Applications of T2*-based MR Imaging and Its Special Applications. *Radiographics* **29**, 1433–1449 (2009).

123. Traficante, D. D. Relaxation. Can T2, be longer than T1? *Concepts in Magnetic Resonance* **3**, 171–177 (1991).

124. Gusev, D. G., Nietlispach, D., Vymenits, A. B., Bakhmutov, V. I. & Berke, H. Synthesis and NMR T1 relaxation study of rhenium and manganese hydride complexes. *Inorg. Chem.* **32**, 3270–3276 (1993).

125. Fullerton, G. D., Potter, J. L. & Dornbluth, N. C. NMR relaxation of protons in tissues and other macromolecular water solutions. *Magn Reson Imaging* **1**, 209–226 (1982).

126. Wokke, B. H. *et al.* T2 relaxation times are increased in Skeletal muscle of DMD but not BMD patients. *Muscle Nerve* **53**, 38–43 (2016).

127. Still, B. M., Kumar, P. G. A., Aldrich-Wright, J. R. & Price, W. S. 195Pt NMR--theory and application. *Chem Soc Rev* **36**, 665–686 (2007).

128. Dayie, K. T., Wagner, G. & Lefèvre, J. F. Theory and practice of nuclear spin relaxation in proteins. *Annu Rev Phys Chem* **47**, 243–282 (1996).

129. Caballero-Manrique, E., Bray, J. K., Deutschman, W. A., Dahlquist, F. W. & Guenza, M. G. A theory of protein dynamics to predict NMR relaxation. *Biophys J* **93**, 4128–4140 (2007).

130. Wood, M. L. & Hardy, P. A. Proton relaxation enhancement. *J Magn Reson Imaging* **3**, 149–156 (1993).

131. Korzhnev, D. M., Kloiber, K. & Kay, L. E. Multiple-quantum relaxation dispersion NMR spectroscopy probing millisecond time-scale dynamics in proteins: theory and application. *J Am Chem Soc* **126**, 7320–7329 (2004).

132. Naugler, null & Cushley, null. Spectral estimation of NMR relaxation. *J Magn Reson* **145**, 209–215 (2000).

133. Facelli, J. C. Chemical shift tensors: Theory and application to molecular structural problems. *Prog Nucl Magn Reson Spectrosc* **58**, 176–201 (2011).

134. Gardner, J. H. & Purcell, E. M. A Precise Determination of the Proton Magnetic Moment in Bohr Magnetons. *Phys. Rev.* **76**, 1262–1263 (1949).

135. Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **69**, 37–38 (1946).

136. Krivdin, L. B. Theoretical calculations of carbon-hydrogen spin-spin coupling constants. *Prog Nucl Magn Reson Spectrosc* **108**, 17–73 (2018).

137. Cherry, P. J., Rouf, S. A. & Vaara, J. Paramagnetic Enhancement of Nuclear Spin-Spin Coupling. *J Chem Theory Comput* **13**, 1275–1283 (2017).

138. Mugarza, A. *et al.* Spin coupling and relaxation inside molecule-metal contacts. *Nat Commun* **2**, 490 (2011).

139. Watson, M. A., Sałek, P., Macak, P., Jaszuński, M. & Helgaker, T. The calculation of indirect nuclear spin-spin coupling constants in large molecules. *Chemistry* **10**, 4627–4639 (2004).

140. Lin, L., Wei, Z., Lin, Y. & Chen, Z. Measuring JHH values with a selective constant-time 2D NMR protocol. *J Magn Reson* **272**, 20–24 (2016).

141. Parella, T. Current developments in homonuclear and heteronuclear J-resolved NMR experiments. *Magn Reson Chem* **56**, 230–250 (2018).

142. Yan, J., Kline, A. D., Mo, H., Shapiro, M. J. & Zartler, E. R. The absolute sign of J coupling constants determined using the order matrix calculation. *Magn Reson Chem* **42**, 962–967 (2004).

143. Wang, B., He, X. & Merz, K. M. Quantum Mechanical Study of Vicinal J Spin-Spin Coupling Constants for the Protein Backbone. *J Chem Theory Comput* **9**, 4653–4659 (2013).

144. Fiala, R. & Sklenár, V. 13C-detected NMR experiments for measuring chemical shifts and coupling constants in nucleic acid bases. *J Biomol NMR* **39**, 153–163 (2007).

145. Meyer, N. H. & Zangger, K. Boosting the resolution of 1H NMR spectra by homonuclear broadband decoupling. *Chemphyschem* **15**, 49–55 (2014).

146. Freeman, R. & Kupce, E. Decoupling: theory and practice. I. Current methods and recent concepts. *NMR Biomed* **10**, 372–380 (1997).

147. Luyten, P. R. *et al.* Broadband proton decoupling in human 31P NMR spectroscopy. *NMR Biomed* **1**, 177–183 (1989).

148. Freeman, D. M. & Hurd, R. Decoupling: theory and practice. II. State of the art: in vivo applications of decoupling. *NMR Biomed* **10**, 381–393 (1997).

149. Saner, M., McKinnon, G. & Boesiger, P. Glycogen detection by in vivo 13C NMR: a comparison of proton decoupling and polarization transfer. *Magn Reson Med* **28**, 65–73 (1992).

150. Holstein, P., Harris, R. K. & Say, B. J. Solid-state 19F NMR investigation of poly(vinylidene fluoride) with high-power proton decoupling. *Solid State Nucl Magn Reson* **8**, 201–206 (1997).

151. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* **131**, 373–378 (1998).

152. Jesmanowicz, A., Hyde, J. S., Froncisz, W. & Kneeland, J. B. Noise correlation. *Magn Reson Med* **20**, 36–47 (1991).

153. Leopold, M. F., Urbauer, J. L. & Wand, A. J. Resonance assignment strategies for the analysis of NMR spectra of proteins. *Mol Biotechnol* **2**, 61–93 (1994).

154. Gronwald, W. *et al.* Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). *J Biomol NMR* **23**, 271–287 (2002).

155. Jee, J. & Güntert, P. Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. *J Struct Funct Genomics* **4**, 179–189 (2003).

156. Grzesiek, S. & Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293 (1992).

157. Zheng, Y. & Yang, D. Measurement of dipolar cross-correlation in methylene groups in uniformly 13C-, 15N-labeled proteins. *J Biomol NMR* **28**, 103–116 (2004).

158. Malik, N. & Kumar, A. Resonance assignment of disordered protein with repetitive and overlapping sequence using combinatorial approach reveals initial structural propensities and local restrictions in the denatured state. *J Biomol NMR* **66**, 21–35 (2016).

159. Aitio, H. *et al.* NMR assignments, secondary structure, and global fold of calerythrin, an EF-hand calcium-binding protein from Saccharopolyspora erythraea. *Protein Sci* **8**, 2580–2588 (1999).

160. Pervushin, K. V., Wider, G., Riek, R. & Wüthrich, K. The 3D NOESY-[(1)H, (15)N,(1)H]-ZQ-TROSY NMR experiment with diagonal peak suppression. *Proc Natl Acad Sci U S A* **96**, 9607–9612 (1999).

161. Oezguen, N., Adamian, L., Xu, Y., Rajarathnam, K. & Braun, W. Automated assignment and 3D structure calculations using combinations of 2D homonuclear and 3D heteronuclear NOESY spectra. *J Biomol NMR* **22**, 249–263 (2002).

162. Protein NMR | A Practical Guide. https://www.protein-nmr.org.uk/.

163. Protein NMR Spectroscopy - 2nd Edition. https://www.elsevier.com/books/protein-nmr-spectroscopy/cavanagh/978-0-12-164491-8.

164. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *J. Magn. Reson.* **213**, 423–441 (2011).

165. Grzesiek, S. & Bax, A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969)* **96**, 432–440 (1992).

166. Muhandiram, D. R. & Kay, L. E. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B* **103**, 203–216 (1994).

167. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289–302 (1999).

168. Farmer, B. T., Venters, R. A., Spicer, L. D., Wittekind, M. G. & Müller, L. A refocused and optimized HNCA: Increased sensitivity and resolution in large macromolecules. *J Biomol NMR* **2**, 195–202 (1992).

169. Clubb, R. T., Thanabal, V. & Wagner, G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N 13C-labelled proteins. *Journal of Magnetic Resonance (1969)* **97**, 213–217 (1992). Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *J. Magn. Reson.* **213**, 423–441 (2011).

Veena Hegde, Ph.D.

hegde.v18@gmail.com

Summary

A passionate integrated molecular biologist with experience in NMR spectroscopy, biotechnology, biochemical and biophysical techniques. A dedicated and reliable person who can work independently and in fast-paced team environments with a proactive approach.

Employment History

2015 - 2023 **Ph.D. Student,** Leibniz Universität Hannover, Germany.

2010 – 2015 **Research Assistant.** NMR Research Center, Indian Institute of Science, India.

Education

2015 – 2023 **Ph.D., Leibniz Universität Hannover** Biomolekulares Wirkstoffzentrum.Germany

2005 – 2007 **M.Sc. Organic Chemistry,** Karnatak University, India.

2003 – 2006 **B.Sc. Chemisitry** Karnatak University, India..

Research Publications

1. "Antimicrobial Peptides with Potential for Biofilm Eradication: Synthesis and Structure Activity Relationship Studies of Battacin Peptides", Gayan Heruka De Zoysa, Veena V Hegde, Srinivasarao Raghothama, December 2014, Journal of Medicinal Chemistry.
2. "Synergetic effect of size and morphology of cobalt ferrite nanoparticles on proton relaxivity", N Venkatesha, Chandan Srivastava, Veena Hegde, December 2014, IET Nanobiotechnology.
3. "NMR Metabolomics Analysis of Stem Cell Culture Media: Human Umbilical Cord Stem Cell Fate, Differentiation and Metabolic Pathways", Veena V Hegde, Srinivasrao Raghothama, Puroshotham, (to be submitted).
4. "An inhibitor of nonhomologous end joining abrogates double-strand break repair and impedes cancer progression", Srivastava, M., Nambiar, M., Sharma, S., Karki, S. K., Goldsmith, G., Hegde, M., Kumar, S., Pandey, M., Singh, R. K., Ray, P., Natarajan, R., Kelkar, M., De, A., Choudhary, B. and Raghavan, S. C. (2012), Cell 151, 1474-1487, (Acknowledged).
5. "The EJC disassembly factor PYM is an intrinsically disordered protein and forms a fuzzy complex with RNA", Teresa Carlomagno, Deepshikha Verma, Veena Hegde, John Kirkpatrick, (2023), Frontiers in Molecular Biosciences, section Structural Biology.

List of Publications

1. “Antimicrobial Peptides with Potential for Biofilm Eradication: Synthesis and Structure Activity Relationship Studies of Battacin Peptides”, Gayan Heruka De Zoysa. Veena V Hegde. Srinivasarao Raghothama, December 2014, Journal of Medicinal Chemistry.
2. ”Synergetic effect of size and morphology of cobalt ferrite nanoparticles on proton relaxivity”, N Venkatesha. Chandan Srivastava. Veena Hegde, December 2014, IET Nanobiotechnology.
3. “The SOLE-RNA element of the spliced *oskar* mRNA binds the protein PYM, an EJC disassembly factor active in mRNA localization”, Deepshikha Verma, Veena Hegde, John Kirkpatrick, Teresa Carlomagno (submitted).
4. “NMR Metabolomics Analysis of Stem Cell Culture Media: Human Umbilical Cord Stem Cell Fate, Differentiation and Metabolic Pathways”, Veena V Hegde. Srinivasrao Raghothama. Puroshotham, (to be submitted).