7-2014

Leucine zipper interacts promiscuously? Analysis of leucine zipper specificity in the c-protein family.

Evelyn Rebecca Yambay-Tilman
Pittsburg State University

THE LEUCINE ZIPPER INTERACTS PROMISCUOUSLY? ANALYSIS OF LEUCINE ZIPPER SPECIFICITY IN THE C-PROTEIN FAMILY

A Thesis Submitted to the Graduate School
in Partial Fulfillment of the Requirements for the Degree of
Master of Science

Evelyn Rebecca Yambay-Tilman

Pittsburg State University

Pittsburg, Kansas

July, 2014

THE LEUCINE ZIPPER INTERACTS PROMISCUOUSLY? ANALYSIS OF LEUCINE ZIPPER SPECIFICITY IN THE C-PROTEIN FAMILY

Evelyn Rebecca Yambay-Tilman

APPROVED:

Thesis Advisor _____
Dr. Irene Zegar, Department of Chemistry

Committee Member _____
Dr. James McAfee, Department of Chemistry

Committee Member _____
Dr. William Shirley, Department of Chemistry

Committee Member _____
Dr. Peter Chung, Department of Biology

ACKNOWLEDGMENTS

I would like to express my gratitude to all the people that made it possible for me to finish my thesis. I am very thankful to Dr. Irene Zegar, and Dr. James McAfee for the encouragement and advice throughout all my college years. They have been patient in teaching me and guiding me during research and the writing of my thesis.

Furthermore, I appreciate the guidance that Dr. William Shirley provided by helping me every step of the way, advising me on the classes I needed to take and also patiently answering my questions and taking the time to help me whenever I needed.

I would also like to express my gratitude towards Dr. Peter Chung for agreeing to be a part of my thesis committee and for putting his time and effort in revising my thesis to make sure it is the best it can be.

I am also thankful to God for putting all of these people in my life to push me towards my goals whenever needed. None of this would have been possible without the support of my parents and friends. I would like to acknowledge Gary and Carrol Bell for opening their home to me and treating me like I was part of their family, for listening to me and supporting me through this process.

Most importantly, this would not have been possible without the support and encouragement of my husband, patiently waiting for me to achieve my personal goals and pushing me towards it when I did not feel like going on. All of these people have made it possible for me to attain this degree in one way or another, and I am very thankful for it.

THE LEUCINE ZIPPER INTERACTS PROMISCUOUSLY? ANALYSIS OF LEUCINE ZIPPER SPECIFICITY IN THE C PROTEIN FAMILY

An abstract of the thesis by

Evelyn Rebecca Yambay-Tilman

Heterogeneous nuclear ribonucleoprotein C is ubiquitous in vertebrates, and exists as two alternatively spliced isoforms, hnRNP C1 and hnRNP C2 (hnRNP C). hnRNP C has been associated with pre-mRNA packaging, pre-mRNA, splicing, mRNA stability, internal ribosome entry site medicated translation, and has even been reported to be an integral component of the telomerase enzyme. Two proteins, hRaly and hRalyl, exhibit a great deal of primary sequence similarity with the C proteins and also conserve structural and functional motifs that have been identified in hnRNP C. A leucine zipper motif has been shown to be the oligomerization domain of hnRNP C and this sequence is conserved in hRaly and hRalyl. To determine if the three proteins are truly separate or whether they form various combinations of homo and hetero-oligomers previous cloning experiments conducted using polycystronic vectors (Peetha, 2013) showed when cloning one gene (either hRaly or hRalyl), *Escherichia coli* remained virulent whereas when hRaly and hRalyl were cloned in the same polycystronic vector, *E. coli* cells died. This lethality was attributed to the hypothesis that the heterodimeric structure between hRaly and hRalyl is the physiologically relevant structure. The research presented here tests the efficacy of this hypothesis by using molecular docking studies. These studies were

conducted using FlexPepDock from Rosetta dock. It was shown that hetero-dimers consisting of hnRNP C/hRaly, and hnRNPC/hRalyl were equally and in some cases more stable than their homo-dimer counterparts. To investigate the positional relevance of heptads 1-4 in determining stability a “scrambled” leucine zipper was generated, this sequence contained random heptads from hRaly, hRalyl, and hnRNP C. The resulting structure was only slightly less stable than any of the other dimers. Analysis of all of the structures identified two salt bridges that were common to all of the dimers modeled but was lacking in the mutated sequence. To determine if the decreased stability of the scrambled sequence resulted from the loss of these two intermolecular salt bridges, these were incorporated. The residues forming these bridges were mutated into the sequence. To determine if this were indeed the case these residues were incorporated into the mutated sequence. The resulting structure’s binding energy was increased by 4 kcal/mol, and was not as stable as all of the other modeled structures. The lack of specificity between the different zippers suggests the possibility of compositionally diverse hnRNP C, hRaly, hRalyl proteins in the cell.

TABLE OF CONTENTS

LIST OF FIGURES

Figures **PAGE**

LIST OF TABLES

CHAPTER I

INTRODUCTION

The heterogeneous nuclear ribonucleoprotein C (hnRNP C) is one of the most abundant proteins in the nucleus of vertebrates. (Choi and Dreyfuss, 1984; Dreyfuss et al., 1984), hnRNP C binds to pre-mRNA cooperatively upon transcription and based upon its concentration appears to saturate the substrate. (Barnett et al., 1989) Its role in pre-mRNA biogenesis has remained obscure. However, different laboratories have suggested that it plays a key role in restraining the conformational diversity of the RNA as well as being directly involved with RNA splicing, and polyadenylation. (Jurica et al., 2002) Though it is localized to the nucleus, numerous laboratories have also suggested that it plays a role in regulating mRNA stability as well as being involved in Internal Ribosome Entry Site (IRES) mediated translation. (Huang et al., 1994; Kamma et al., 1999) There are two isoforms of hnRNP C that result from alternative splicing of the same transcript. (Merrill et al., 1989) hnRNP C2 differs from its alternatively spliced isoform hnRNP C1 by the addition of 13 amino acids in the primary transcript. (Merrill et al., 1989)

Several structural domains have been identified in hnRNP C. (Swanson et al., 1987) The NMR structure of the amino terminal RNA binding domain (RBD) has been determined and like all canonical RBDs it consist of a βαββαβ structure. (Ford et al.,

2002) Carboxy terminal to this region is a domain that has been called bZIP-like as a result of its similarity to the basic leucine zipper motifs found in transcription factors. (Burd et al., 1994; Görlach et al., 1992) Like the transcription factors, hnRNP C has a leucine zipper that is preceded by a basic region. This "B-ZIP like" motif has also been shown to bind RNA. The carboxy terminal region is characterized by a large number of acidic residues. (Burd et al., 1994; Görlach et al., 1992)

Over the past decade two proteins that exhibit a high degree of sequence similarity to the hnRNP C proteins have been identified. (Ji et al., 2003) One of these proteins, hRaly, was discovered independently by two different laboratories that were studying a genetic lesion at the mouse agouti locus. (Duhl et al., 1994; Michaud et al., 1994) Mice that were heterozygous for this lesion exhibit a wide range of physiological defects while homozygosity was embryonic lethal. (Duhl et al., 1994; Michaud et al., 1994) The latter phenotype was shown to result from the loss of a DNA sequence containing the hRaly gene. (Krylov and Vinson, 2001; Duhl et al., 1994; Michaud et al., 1994; Nakielny and Dreyfuss, 1997) Sequencing of this DNA revealed that its primary amino acid sequence is 54% identical to hnRNP C and more importantly preserved all of the structural motifs that had been characterized in the C proteins. A few years after the discovery of hRaly, another group identified hRalyl. (Ji et al., 2003) The name was inspired by its similarity to hRaly and like hRaly it too conserves all of the structural motifs associated with hnRNP C. (Ford et al., 2002)

The discovery of hnRNP C-like proteins raises a number of questions among which is the question, Are hRaly and hRalyl functionally redundant with hnRNP C? Embryonic lethality observed for hRaly deficient mice would argue against functional

redundancy. (Duhl et al., 1994) Moreover, studies have also shown that mice that are null for hnRNP C fail to develop beyond the blastocyst stage. (Duhl et al., 1994) However, these same cells can be rescued during development and propagated in tissue culture. This latter observation suggests that gene dosage is responsible for embryonic lethality, and propagation in tissue culture implies that another protein can functionally replace hnRNP C. Obviously, the replacement candidates would be hRaly or hRalyl. The conservation of all of the structural motifs in all three proteins also strongly supports functional overlap. (Ford et al., 2002)

Since the leucine zipper of hRaly, hRalyl, and hnRNP C have extensive homology it is also possible that these proteins form hetero-oligomers. A major objective of the research presented here is to test the feasibility of this hypothesis by assessing the stability of hetero-dimer interactions between the different leucine zippers.

Crosslinking studies using the native protein isolated from HeLa cells previously suggested that hnRNP C is a hetero-tetramer consisting of a $C1_3C2$ protomer composition. (Barnett et al., 1989) However, density gradient studies using recombinant hnRNP C1 or hnRNP C2 revealed that both proteins were homo-tetramers. (McAfee, Soltaninassab, Lindsay et al., 1996) Despite inconsistencies regarding the tetramer composition it has been clearly demonstrated that the leucine zipper is the oligomerization domain. This was illustrated by studies where mutation of a single leucine within the zipper disrupts oligomerization. (Wan et al., 2001)

Leucine zippers and their ability to function as oligomerization domains have been studied extensively. (Krylov and Vinson, 2001; Vaughan et al., 1995) The canonical leucine zipper is characterized by a 7 heptad repeat that typically begins with

the amino acid leucine or isoleucine in position 1 (also called position **a**) of the heptad. (Vaughan et al., 1995) Though not obligatory, it is also common for a hydrophobic residue to be found in position **d** of the heptad (the 7 residues of the heptad are labeled as **a**-**g**). The spacing of hydrophobic residues within zippers generates a hydrophobic face on one side of the helix and a hydrophilic one on the other side. The hydrophobic residues interact to stabilize the multimeric structures that in most cases are dimers, but can also exist as higher order structures. Depending on the composition of the heptad repeats they may be arranged in either a parallel or anti-parallel orientation. (Krylov and Vinson, 2001)

The NMR structure of synthetic hnRNP C leucine zippers has been solved and it has been shown that the zippers are oriented in an anti-parallel fashion. (Whitson et al., 2005) Crosslinking of these regions fused to an affinity tag identified the existence of monomers, dimers, trimers, and tetramers. The authors concluded that the native protein exists as a bundle of antiparallel dimers. (Whitson et al., 2005)

The similarity of the leucine zippers of hnRNP C, hRaly, and hRalyl prompted further investigation to study the efficacy of these forming hetero-oligomers between various protomers of the three proteins. An examination of Figure 1 shows that the leucine zipper of hnRnP C and hRalyl are 86% identical and hnRNP C and hRaly are 71% identical not considering conservative substitutions.

hnRNP C	LQAIKKE LTQIKQK VDSLLEN LEKIEKE
hRalyl	LQTIKKE LTQIKTK IDSLLGR LEKIEKQ
hRaly	LQAIKTE LTQIKSN IDALLSR LEQIAAE

Figure 1. Sequence homology between the leucine zippers of hnRNP C,hRalyl, and hRaly.

The hetero-oligomeric question was initially addressed by cloning hnRNP C, hRaly and hRalyl in a polycistronic vector fusing each protein to a different affinity tag. (Peetha, 2013) Though both proteins could be established independently in *E. coli*, the polycystronic vector containing two of the three genes was lethal. This led the researchers to believe that lethality resulted from the formation of a functionally distinct hetero-oligomer. Since we could not conduct the experiment in the cell, the work reported here discusses molecular modeling studies to investigate interactions between hnRNP C, hRaly, and hRalyl.

CHAPTER II

MATERIALS AND METHODS

Computer Programs

Chimera was developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco. (Pettersen et al., 2004) Chimera is used in this thesis for the visualization and analysis of structures of biomolecules. Modeller was developed and maintained by Andrej Sali, and Ben Webb, respectively, at the Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research at the University of California, San Francisco and is used as a plugin in Chimera to predict secondary structures of polypeptides by the comparative alignment of the sequence to be modeled with known structures with high homologies using spatial restraints. (Eswar et al., 2006)

Yasara (Yet Another Scientific Artificial Reality Application) View is the freeware part of the Yasara suite. This software was developed by Elmer Krieger, the founder of Yasara Biosciences in Vienna Austria**.** (Krieger et al., 2002) Yasara View was primarily used to interactively view and analyze protein structures and to create the structural images used in this thesis. (Krieger et al., 2002)

FoldX is an empirical force field developed in the Serrano lab at the Heidelberg Laboratory of the EMBL by Raphael Guerois, Jens Nielsen, Jesper Ferkinghoff-Borg, Joost Schymkowitz, Frederic Rousseau, Francois Stricher and Luis Serrano. (Van Durme et al., 2011) It was later implemented for use as a plugin in Yasara View by Joost Van Durme at the Switch Laboratory. (Van Durme et al., 2011) Due to its fast and fairly accurate estimation of difference in energies between folded and unfolded structures, it has become one of the most widely used empirical force field in predicting protein's stability. In FoldX, the following empirical equation is used to calculate the free energy (in kcal/mol) for the folded and unfolded proteins

$$\Delta G = a\Delta G_{vdw} + b\Delta G_{solvH} + c\Delta G_{solvP} + d\Delta G_{wb} + e\Delta G_{hbond}$$

$$+ f\Delta G_{el} + g\Delta G_{kon} + hT\Delta S_{mc} + iT\Delta S_{sc} + k\Delta G_{clash}$$

Equation 1. Free energy calculation using FoldX.

where a-k are relative weights of the different energy terms used to calculate the total free energy. (Van Durme et al., 2011) ΔG_{vdw} is the contribution of the Van der Waals interactions to the total free energy of the protein. The next three terms correspond to the contribution to the total free energy as a result of desolvation (removal of water) of hydrophobic residues being buried into a hydrophobic core (ΔG_{solvH}), the penalty to the total energy as a result of desolvation of polar residues (ΔG_{solvP}), and the free energy of interactions of persistent water molecules that are bound to the protein through more than two hydrogen bonds (ΔG_{wb}). ΔG_{hbond} is the contribution of hydrogen bonding between amino acids to the total free energy of the protein. The next two terms, ΔG_{el} and ΔG_{kon} represent the contributions of the intra polypeptide chain electrostatic interactions (ΔG_{el})

and the inter-polypeptide electrostatic interaction (ΔG_{kon}). The terms (ΔS_{mc} + $iT\Delta S_{mc}$) are the entropic penalty for fixing the backbone and side chains in given conformations as a result of protein folding, respectively. Finally ΔG_{clash} is a measure of steric overlap between atoms as a result of protein folding. All of these energy terms are derived empirically from experimental work on proteins, and amino acids. (Van Durme et al., 2011) The precision of the total free energy values (ΔG) calculated using FoldX was determined using experimentally determined mutational free energy changes obtained from 1000 proteins. And in its most current release, FoldX yields differences in calculated versus experimental mutational free energy values ($\Delta\Delta G$) with a standard deviation of 0.46 kcal mol^{-1}. To calculate the interaction energy ($\Delta G_{binding}$) of the docked proteins, the difference in total free energy of the folded and unfolded structures is determined.

The Rosetta software is a collection of algorithms used for computational docking of proteins using the Monte Carlo approach and is widely used for structure prediction of macromolecular complexes. Rosetta was initially developed by David Baker at the University of Washington for the prediction of molecular docking and was later taken over by the members of RosettaCommons and has since been adapted to solve a large array of computational macromolecular problems. (Raveh et al., 2010; London et al., 2011; Chaudhury and Gray, 2008)

FlexPepDock is a refinement high resolution docking protocol found in the Rosetta modeling framework which implements fully flexible and rigid body orientation for the backbone as well as full flexibility of the side chains for the peptides to be docked. (Raveh et al., 2010) The only requirement of this protocol is an input file which

represents the starting structure. This file must contain the two peptides to be modeled. The starting conformations of the two peptides can have varying degrees of conformational difference from the native structure. The first step in the docking protocol involves prepacking of the side chains of the input structure with the purpose of eliminating internal energy clashes in the peptides. Once the internal clashes are eliminated, the pre-packed structure becomes the starting structure for the next step. The second step involves the generation of 300 low- resolution structures by optimizing the rigid body orientation using Monte-Carlo search with energy minimization. In this routine, translational and rotational rigid body perturbations of 0.2 Å and 7^{o} are applied, respectively. The third step involves the generation of 300 high-resolution structures by optimizing the peptide backbone while allowing the backbone to be fully flexible in order to induce a better fit between the docked molecules.

MolProbity is an online software developed by David C. Richardson and Co-workers in the Department of Biochemistry at Duke University. It is primarily used to diagnose crystallographic, NMR and docked structures for errors that result from minor geometric distortions of protein side chains. MolProbity is primarily used in this thesis to determine the Ramachandran plots which represents contour diagrams of the ϕ and ψ dihedral angles of all the residues in the various modeled structures and compare them to that of the NMR solution structure of hnRNP C. (Chen et al., 2009)

Creating a Dimer out of the NMR hnRNP C Structure

The hnRNP C structure is used throughout this process as a reference structure. The NMR solution structure of hnRNP C was downloaded from the Protein Data Bank (PDB) and cut into its respective dimers. Yasara was used for viewing and editing this

homotetramer. Once the hnRNP C was loaded onto Yasara, the *Edit* menu was used to obtain the desired dimer. The *Delete* option was used and then *Molecules* was selected to delete the unwanted chains. The chains A and D were deleted from the tetramer and then the remaining structure was saved using the *File/Save/PDB* path.

Modeling of the Secondary Structures of the Amino Acid Sequence

The *Open Option* in the *File* pull down menu in the Windows version of Chimera was used to upload a PDB file of a known protein structure (NMR or crystal structure). Using the *Tools* pull down menu, the protein's sequence was displayed by choosing the *Sequence* option selection. If the protein is multimeric and consist of more than one polypeptide polymer, one of the monomers in the *Show Model Sequence Window* is selected and a new window labeled *"Chain X: Name of the uploaded protein"* appears. The *Add Sequence* option from the *Edit* pull down menu is then selected. A new window entitled : *"Add Sequence to Chain X: Name of the uploaded protein* "opens. To the "*Sequence name*" box, a name of the sequence to be modeled is added, and to the box labeled "*Sequenc*e" a FASTA format of the amino acid sequence of the protein to be modeled is added. The Modeller function from the *Structure* pull down menu of this window is then chosen followed by choosing the *OK* option. In the Alignment based on "*Chain X: Name of the uploaded protein"* window which now displays the sequence of the protein with the known structure and the sequence to be modeled, the *Structure* pull down menu is chosen and the *Modeller Homology* option is selected. In the window labeled *"Comparative Modeling with Modeller"*, the name of the sequence to be modeled is placed in the box titled *"Choose the target (sequence to be modeled)"* and in the *"Choose at least one template"* box the name of the protein with the known structure

which will be used as a template is selected. The *Run modeler via web server* button is selected and a *Modeller license key* obtained from the developer's website by filling out a license agreements is placed in the *"Modeller license Key"* box. The number of output models can be specified in the *Advanced options* selection. In this thesis, 5 output modeled were usually chosen. The option *OK* was selected. Once the models were generated, a *"Modeller Results"* window is opened and the *Numbers* of the 5 generated models are displayed in chronological order. The modeled structures are then displayed graphically in Chimera and analyzed for their degree of alignment with the known protein structure. The structure with the best RMSD value with respect to the known protein structure is chosen.

Structure Alignments Using Chimera

The structure alignments were made by selecting the pull down menu *Tools* and using the Matchmaker option under *Structure Comparison*. At that point, a window pops up; the reference structure and the structure to be matched have to be selected. This window lists all the modeled structures plus the structure listed as the reference to create the monomers. All other options were left as defaulted by the program and the *OK* button was pressed. The all-residue C_α-RMSD value is displayed at the bottom left hand on the main window, another option is to click on *Apply* instead of *OK*, the pop up window stays open but it is still possible to see the RMSD value in the main window, this way allows for faster processing of all the structures, by clicking on the different models and keeping the reference structure the same, always clicking *Apply* to display the all-residue C_α-RMSD value in the same determined position.

Once the lowest all-residue C_α-RMSD value is found, the monomer is selected and needs to be saved using *File/Save PDB*. The monomer is now ready to be docked.

Docking Dimers through FlexPepDock

FlexPepDock requires the monomers to be packed in the same pdb file and the monomers must be given two different letter labels, ex. Molecule A and molecule B. The structures are also required to have no more than 30 amino acids. A reference structure can be loaded by clicking on *Advanced options*. This toggles down a menu that allows one to upload the reference structure. The hnRNP C NMR dimer was used as a reference for this docking process. To ensure the highest possible resolution structures that can be obtained by FelxPepDock, the number of low and high resolution structures was set to 300 each. Everything else was left using the default session and an email address was necessary to retrieve the resulting structures. After the completion of the calculated 600 structures, the output files that are obtained include the ten structures with the lowest total energy and a score graph that plots the energies of the 600 generated structures with respect to their all-residue Cα-RMSD (rmsBB) relative to a provided native structure (the NMR structure of hnRNP C). The server sends an email with a link to the results, by opening that link, the results are displayed and the structures can be downloaded by scrolling down to the bottom of the page and clicking on "*Top 10 models(zip file)*" After extracting the file, ten individual files are created, they are named top 1 through top 10.

Alignment of Structures Using MUSTANG

Each one of the top 10 structures mentioned above is analyzed independently. The initial assessment of the goodness of the structure is to compare each one of these

structures to the hnRNP C NMR structure to determine backbone and side chain alignment. For purposesof this study, the **a** and **d** positions were compared by superimposing the NMR structure with each of the top docked structures. The MUSTANG plugin in Yasara View was used to superimpose the structures. *Analyze/Align/Object with MUSTANG* was the path used to access the MUSTANG tool. A pop up window titled "*Select source objects to align with another object*" will appear. The list of objects was now displayed and one was selected as a reference structure, after clicking *OK*, another window comes up "*Select target object for structural alignment with MUSTANG.*" *View/Show atoms/Residue sidechain* was the path utilized to show the desired side chains. A window titled "*Select residues (sidechain +bound backbone will be considered)*" pops up and the residues were selected by choosing them from the list. All structures were compared in the same manner. The structures that have the best alignment with the NMR structure of hnRNP C were chosen and subjected to the repair algorithm in FoldX. Repair is required to remove electronic clashes in side chains as a result of modeling. The repair was repeated a multiple of times until the total free energy of the structure reaches a minimum value.

The Determination of the Binding Free Energy of Docked Structures

The repaired structure discussed in the above section is used to determine the binding free energy of the two docked monomers. This is done by selecting the *Interaction Energy of Molecules* icon in the *FoldX* menu. The first interacting molecule (monomer) is then selected from a list of molecules that appear in *Sequence* box of "*Select First Molecule Range*" window. The second interaction molecule (monomer) is selected in a similar fashion. The total interaction energy along with the various

contributing energy values are then displayed in the bottom window of Yasara view which can be accessed by tapping the space bar twice.

Structure Validation using MolProbity

The PDB coordinate file of the structure to be analyzed can be uploaded in MolProbity by selecting *Choose File>* by selecting the *Upload>* option from the **Main Page** window. After the file has been uploaded, an **Uploaded file s**hort description page appears, and the *Continue>* option is selected. This is followed by an *Add hydrogens>* button, a recommended function to reduce electron cloud clashes which closely mimics conditions in real structures. The *Start Adding H>* button is then selected using all the default options originally selected in the **Add hydrgens** window. After all the hydrogen atoms have been added a **Review flips** window appears with a message indicating whether or not some of the residues have been flipped to fix incorrect orientations. *A continue>* button in this window is then selected. The *Analyze all-atom contacts and geometry* button is then selected from the top of the next **Main page** window. This is followed by selecting The *Run programs to perform these analyses>* option from the **Analyze all-atom contacts and geometry** window using all the default analyses. The outputs of all the performed calculations are then accessed in the **Analysis output: all-atom contacts and geometry for molprobity-C-CFH.pdb** window.

CHAPTER III

RESULTS

Positive Docking Controls

The high degree of sequence homology observed in the leucine zippers of hRaly, hRalyl and hnRNP C prompted us to investigate whether or not these proteins form hetero-dimers with one another. The homology data shown in Table 1 shows that amino acids of the three leucine zipper sequences of these proteins share from 60% to 75% identity, this indicates that hetero-dimers are highly likely. As mentioned earlier, the first approach is to investigate whether these proteins interact in the cell by cloning and co-expressing genes from two or three of these proteins in a polycystronic vector. However, due to the fact that the presence of at least two of these genes in the same vector proved to be lethal to *E. coli*, the feasibility of these interactions were investigated through molecular docking studies. Nevertheless, before proceeding with this approach the efficacy of the modeling and docking systems were first tested by conducting a positive control and a negative control. The positive control was to dock two hnRNP C monomers and determine if the resulting structure was comparable to that native NMR solution structure.

Table 1. Homology between different C proteins and Max.

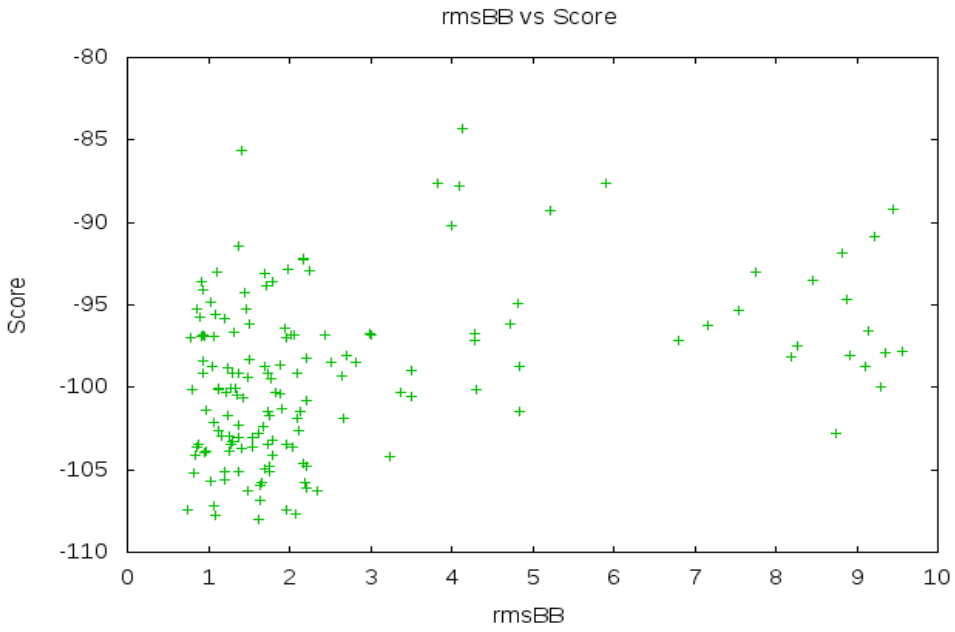
PROTEINS	PERCENT IDENTITY
hnRNP C + h Raly	60.71%
hnRNP C + hRalyl	75.00%
hnRaly + hRalyl	64.29%
HnRNP C+ Max	14.29%

The first step in the Docking of the hnRNP C homodimer involves the generation of an hnRNP C monomer in Modeller using the available NMR structure as the template. Rosetta FlexPePDock was used to generate the homo-dimer of hnRNP C. Shown in the top panel of **Figure 2** is the score plot generated for hnRNP C docked structures. Examination of the data shown indicates that more than 30% of the six hundred structures have low energies (< -85 kcal/mol) with rmsBB values with respect to the NMR hnRNP C structure between 1-2 Å. Of the six hundred modeled structures, the structure with the lowest energy represents the final structure. The final structure for hnRNP C superimposed over the NMR solution structure is shown in bottom panel of **Figure 2** superimposed over its NMR counterpart. Alignment analysis of both of these structures based upon an all-residue C_α-RMSD value using the multiple structural alignment algorithm (MUSTANG) in Yasara resulted in a C_α-RMSD value of 1.1 Å. The superposition of the modeled structures over the NMR dimer parallels the low RMSD value obtained.

Similarities between these structures were also assessed using Ramachandran plots that show two contour plot regions of the backbone ϕ and ψ dihedral angles of residues in right-handed α-helices. **Figure 3** shows these plots for the NMR and modeled hnRNP C structures. The regions shown in Panels A and B of this figure are

contour plots that represent ϕ and ψ angles combinations for all the residues in the NMR solution structure and the modeled structure of hnRNP C, respectively. The arrows in both figures point to the regions where the residues have the most favored α-helical conformations. In these most favored regions, the ϕ and ψ angles have average values of -57^{o} and -47^{o}, respectively. If 90% or more of the residues in a given structure show ϕ and ψ angles combinations in these most favored regions, the structure is then considered a valid structure with α-helical components. The rest of the regions in the Ramachandran plots represent allowed but less favored geometries for α -helical structures. Residues with geometries outside both of those regions are non-favored and are termed outliers and in α-helices those are typically occupied by terminal residues where there is more flexibility in the structure than observed for the internal residues. In fact for Isoleucine and Valine, ϕ and ψ combinations are typically shown separately in Ramachandran plots due to their bulky side chains which often result in large deviations in the ϕ and ψ angles relative to the ideal values observed in right handed α -helical structures. Examination of the data shown in Panel A of **Figure 3** reveals that 90.4% of the residues of the NMR structure are present in the favored region with only one isoleucine residue out of the five total valine/isoleucine residues present in hnRNP C found in an outlier region with ϕ and ψ angles of 65^{o} and 105^{o}, respectively. Furthermore, Panel B of **Figure 3** reveals that 96.2% of the residues in the Modeled hnRNP C structure with ϕ and ψ angles that lie in the favored region. And similar to the NMR solution structure with only one isoleucine residue (the same terminal residue found in the NMR structure) is present in the non-favored region of the Ramachandran plot. The strong

A



B

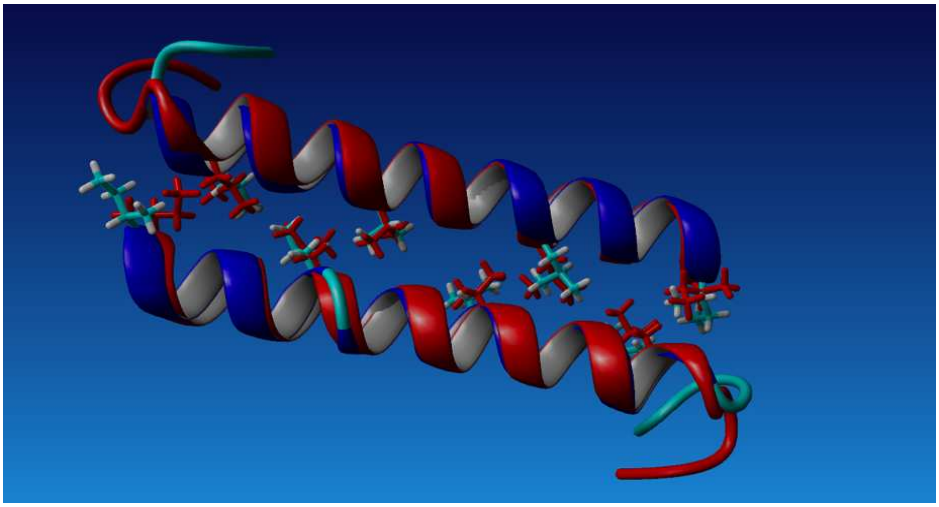


Figure 2. Panel A shows the docking score graph obtained from Flexpepdock which shows the total free energy of the generated 600 structures of hnRNPC vs. the all-residue C_α-RMSD values with respect to the hnRNP C NMR structure. Panel B shows the best fit modeled hnRNP C dimer superimposed over the hnRNP C NMR structure. The backbone of the modeled hnRNP C and the NMR structure of hnRNP C are colored blue and red respectively. The **a** residues are colored light blue and red, for the NMR and modeled structures, respectively.

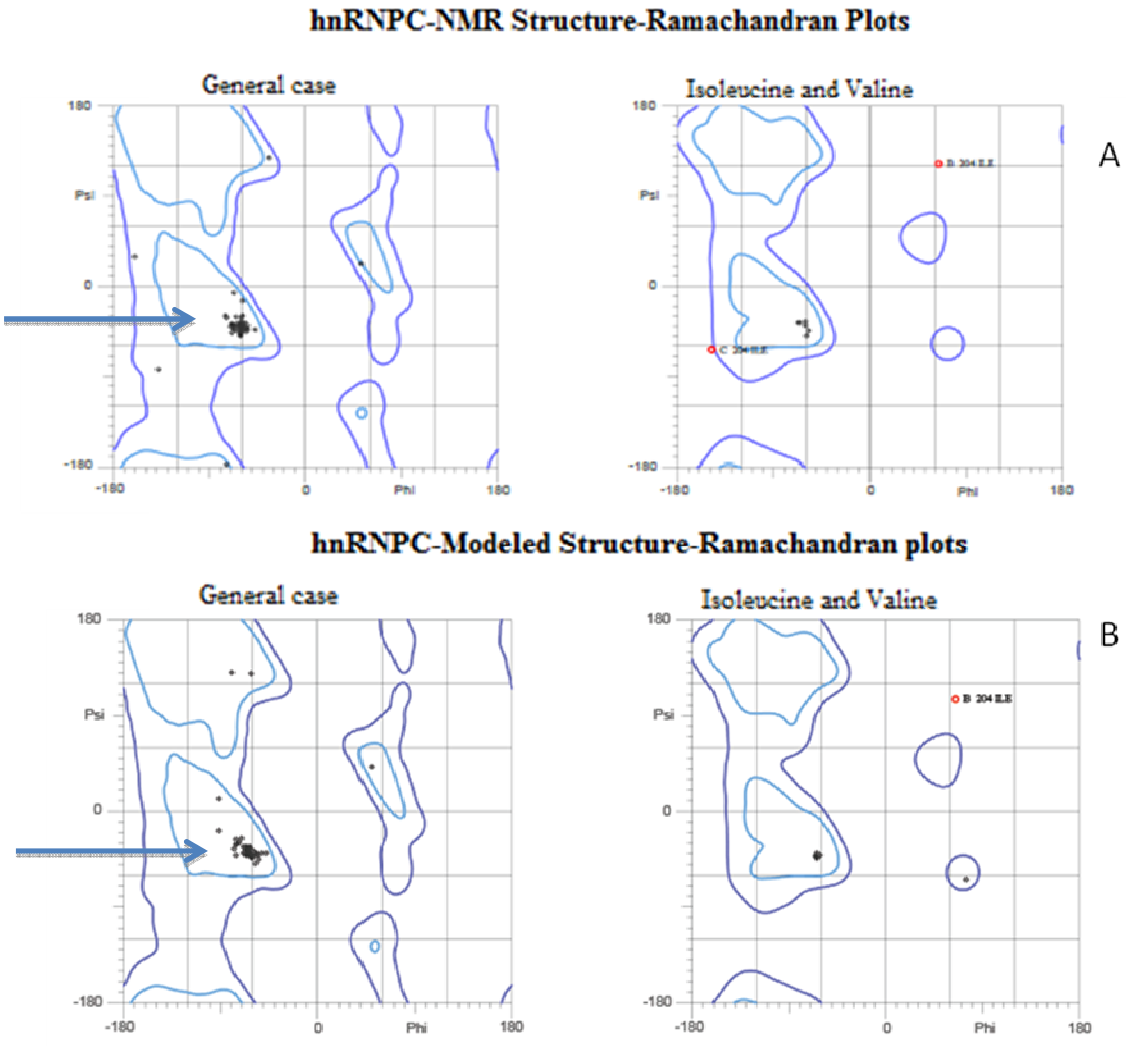


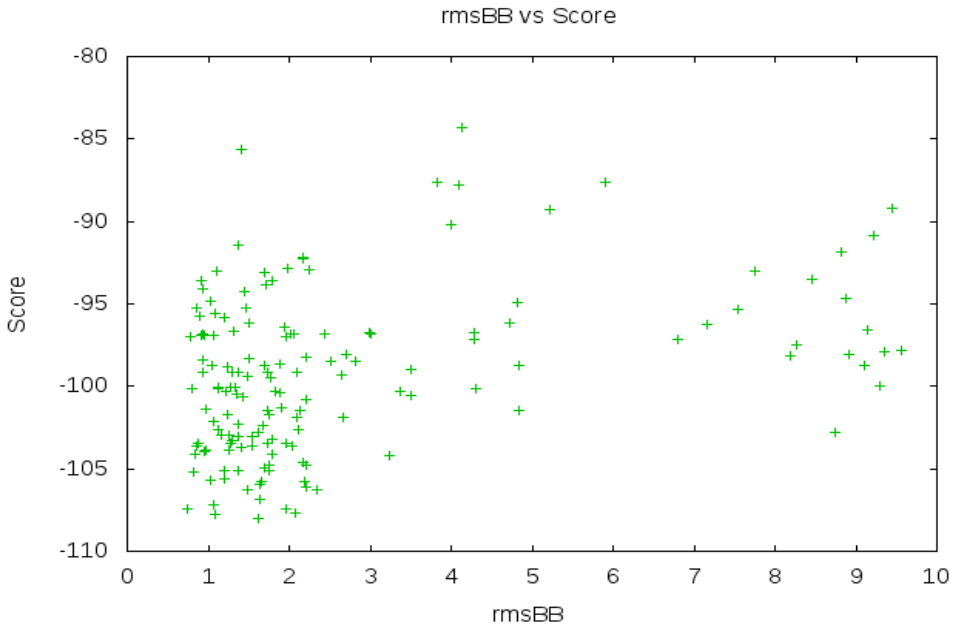
Figure 3. Ramachandran plots generated using MolProbity. The structures of hnRNP C NMR is in panel A and hnRNP C modeled in panel B. The contours indicated by the arrows represent favored regions containing ideal ϕ and ψ combinations for right-handed a-helical structures. The other contours represent less favored but allowed ϕ and ψ combinations for right-handed a-helical structures. The dots shown in black, represent amino acids with ϕ and ψ combinations that fall in the corresponding contours. Whereas the red dots which appear outside the contours (outliers) represent amino acids with ϕ and ψ combinations that are not favored for right-handed α-helical structures.

similarities in the Ramachandran values of ϕ and ψ angles between the NMR and the modeled hnRNP C structures further solidifies the accuracy of the modeled structure. In fact Ramachandran plots are considered the gold standards in assessing the validity of protein structures.

An additional positive control was obtained where by modeling the hetero-dimeric basic leucine zipper structure of the basic HLH-leucine zipper transcription factors Myc and Max. Shown in **Figure 4**, is the docking score graph for the Myc/Max dimer as well as the modeled Myc/Max dimer overlayed on the crystal structure of the same dimer. Alignment analysis of these two structures using MUSTANG resulted in an all-residue C_α-RMSD value of 0.89 Å. The superimposed images also shown in **Figure 4** validate this RMSD value.

As a negative control the hnRNP C monomer was docked with the Max monomer. This is used as a negative control due to that fact that the amino acid sequences for hnRNP C and for the Max protein share very low homology (14.29% as shown in Table 1). As a result of this low homology, it was not anticipated that the docking calculation will result in a valid leucine zipper dimeric structure. Shown in **Figure 5** is the score plot obtained in this docking. It is clear from the data that none of the calculated 600 structures have energies below -85 kcal/mol and none have all-residue C_α-RMSD values calculated with respect to either the NMR structure of hnRNP C or the crystal structure of the Myc/Max heterodimer below 5 Å. In fact the lowest energy structure calculated does not at all resemble a leucine zipper. Again this is an expected outcome considering the low sequence homology between these two proteins.

A



B

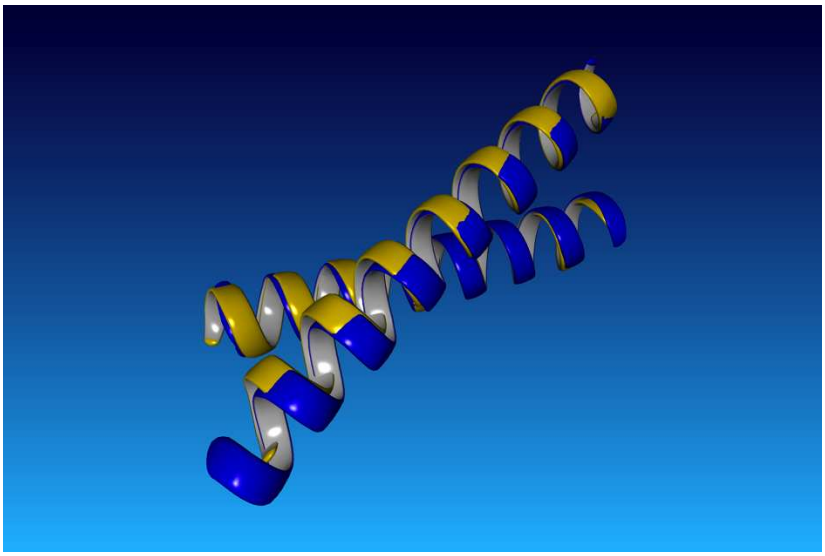
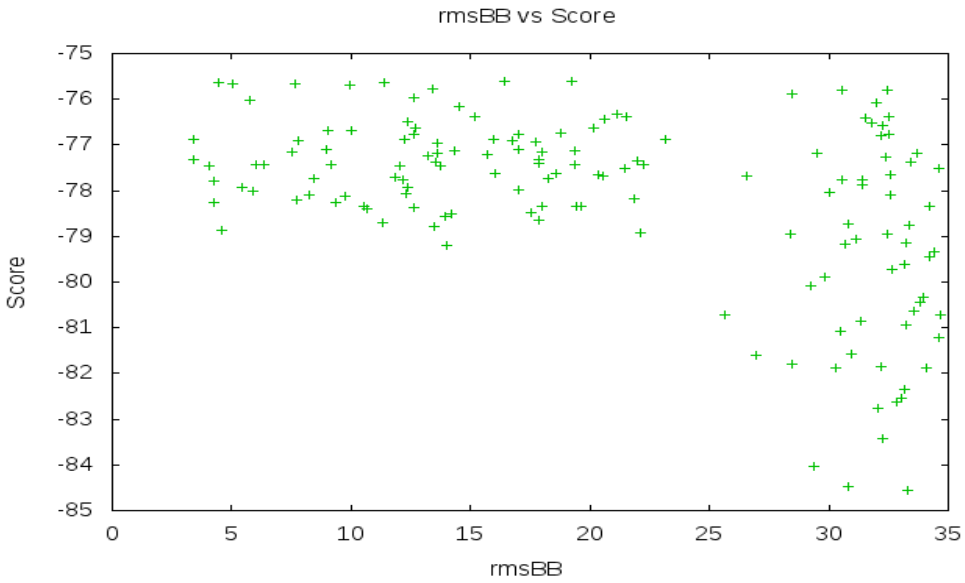


Figure 4. The docking score graph obtained from flexpepdock which shows the total free energy of the generated 600 structures of Myc/Max heterodimer vs. the all-residue C_{α}-RMSD (rmsBB) values calculated with respect to the crystal structure of the Myc/Max hetero-dimer. The superimposed structures in the bottom panel represent the best fit calculated Myc/Max heterodimer (yellow) overlaid on the Myc/Max crystal structure (blue).

A



B

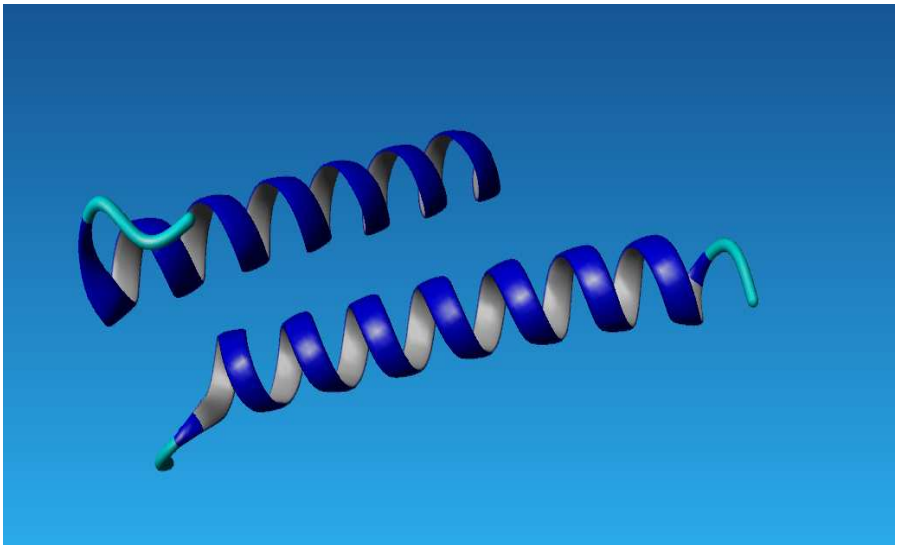


Figure 5. The docking score graph obtained from flexpepdock which shows the total free energy of the generated 600 structures of hNRNP C/Max sequences vs. the all-residue C_α-RMSD (rmsBB) values calculated with respect to the crystal structure of the Myc/Max hetero-dimer. The structure shown in the bottom pane represents the hnRNP C helix (top helix) positioned over the Max helix (bottom).

Generation of homo-and Heterodimers of hnRNP C, hRaly, and hRalyl

Now that the efficacy of the Rosetta Docking algorithm, Flexpepdock, was confirmed homo and heterodimers of hnRNP C, hRaly and hRalyl, were generated by docking various combinations of the monomers of the above mentioned proteins. In the sections shown below, only the calculated structures that have the lowest total energy values and the smallest all-residue C_α-RMSD calculated with respect to the NMR hnRNP C structure are presented.

The Docking of hRaly-hRaly and hRalyl Homodimers

Shown in **Figure 6** is the homodimer of hRaly compared to the modeled homodimer of hnRNP C (Panels A and B, respectively). Both structures are shown with the hydrophobic residue side chains that make up the **a** and **d** positions of each heptad. Upon visual inspection of the two structures it can be said that both structures exhibit a high degree of similarity. Both structures are antiparallel and pack according to the accepted "knobs in a hole" arrangements where the knob (the **a** residue) (knob) on one chain is packed into a hole which consists of a **d** residue and two **a** residues on the opposite chain. Moreover, an all atom alignment of both of these structures using MUSTANG results in 0.90 Å C_α-RMSD value. Similarly, **Figure 7** shows the best fit modeled structure of the hRalyl-hRalyl homodimer compared to that of the hnRNP C homodimer. Again, the hRalyl homodimer is virtually identical to the hnRNP C dimer with an all-residue C_α-RMSD of 0.91 Å and is an antiparalel leucine zipper with the **a** residues packing against the **d** residues according to the "knob in a hole" arrangement.

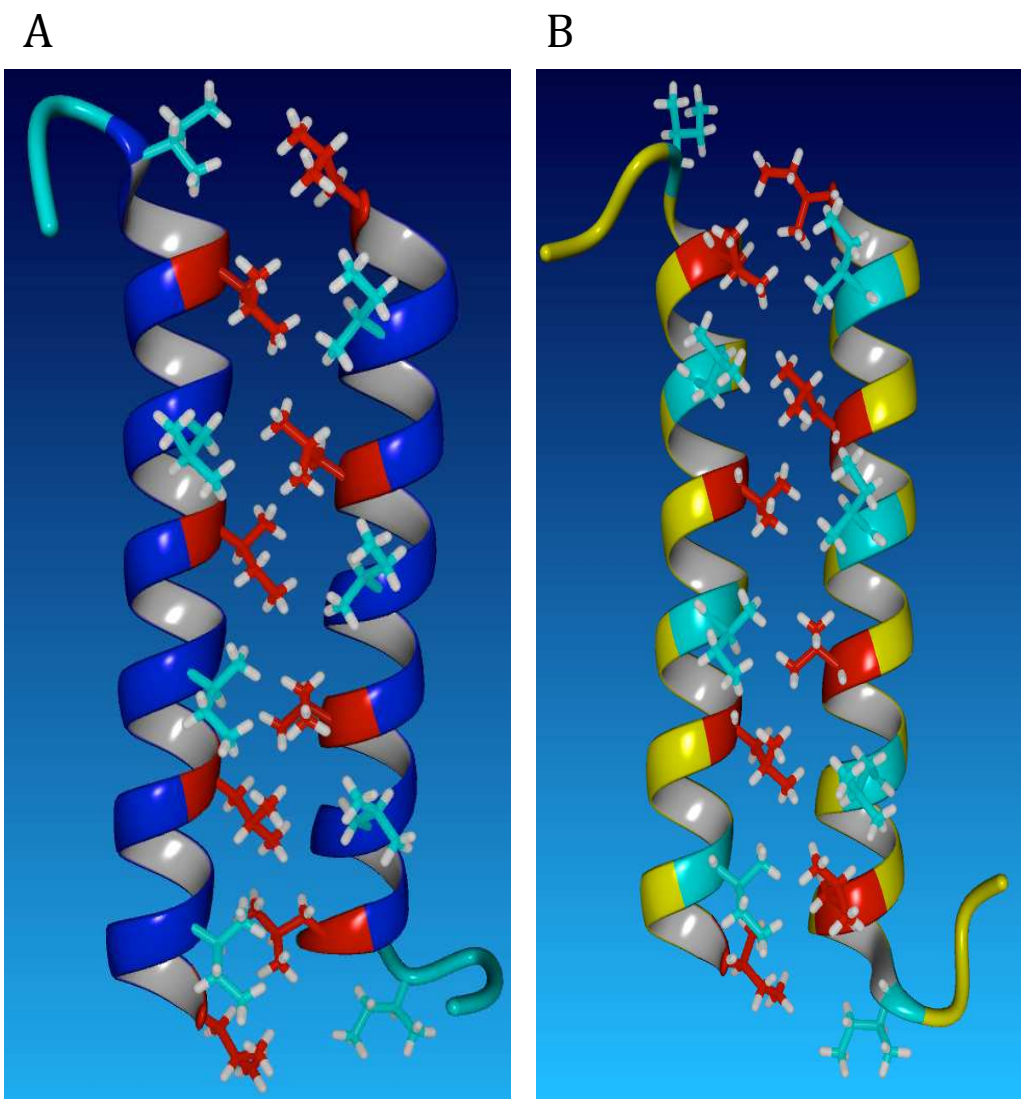


Figure 6. Panel A. The best fit modeled hRaly-hRAly homodimer. Panel B. The best fit modeled hnRNP C homodimer. In both structures, the **a** positions in the four heptads are shown in red and the **d** position in the four heptads are shown in light blue.

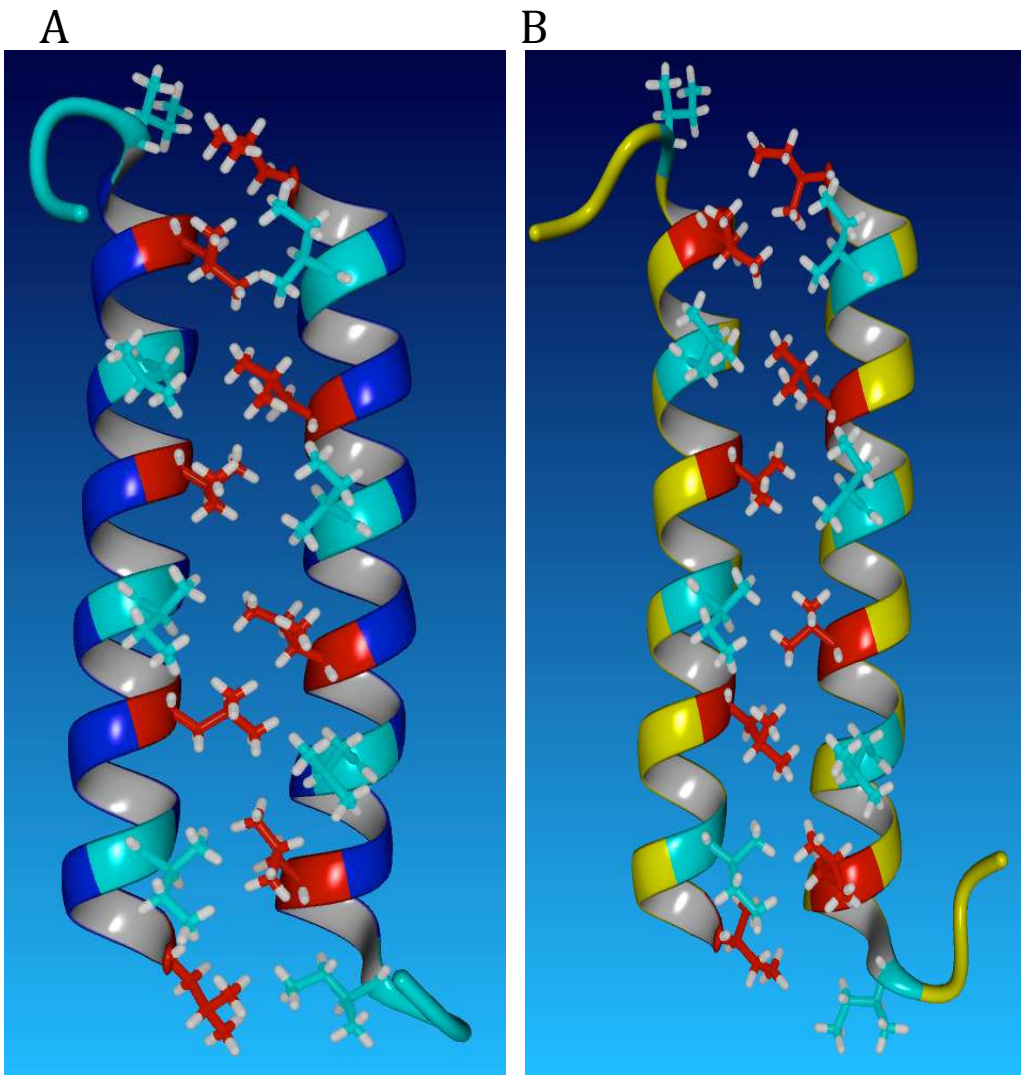


Figure 7. Panel A. The best fit modeled hRalyl-hRalyl homodimer. Panel B. The best fit modeled hnRNP C homodimer. In both structures, the **a** positions in the four heptads are shown in red and the **d** position in the four heptads are shown in light blue.

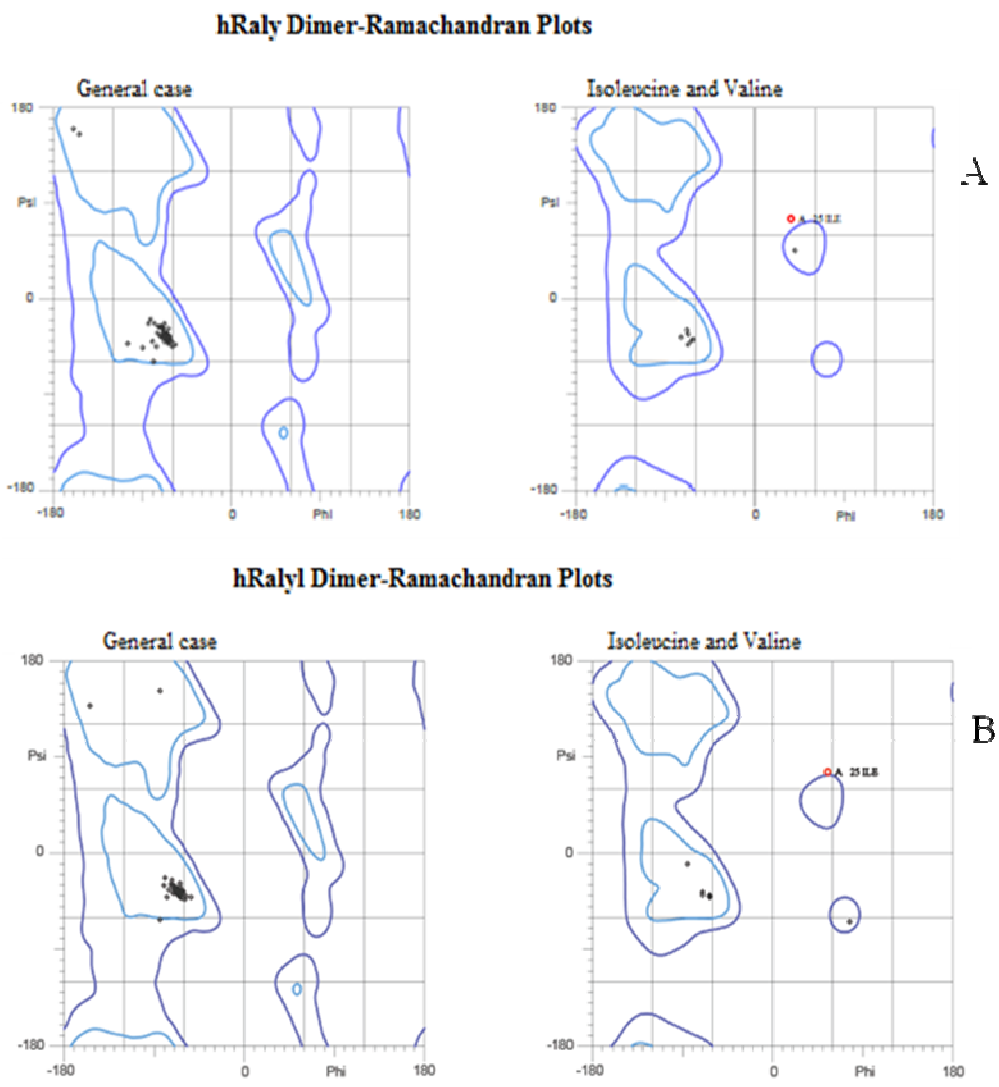


Figure 8. Homodimers corresponding to hRaly and hRalyl structures were represented to prove validity. These Ramachandran plots were generated using MolProbity. Both panel A and panel B show the only residue that was not encountered in a favorable region. The red dot represents that residue and the label on both panels read Ile 25 from chain A of the dimer.

To further validate the docked homodimers of hRaly and hRalyl, Ramachandran plots were also generated, as seen in **Figure 8**. Most of the residues (96.0% in the hRaly structure and 94.2% for the hRalyl structure) are located in the favored region of the Ramachandran plot with only one outliner residue in both cases that represents the same terminal isoleucine observed for the hnRNP C structures.

The Docking of the hNRNP C, hRaly and hRalyl Heterodimers

Structures were also calculated that resulted from the modeling of interactions between the different monomers of hnRNP C, hRaly, and hRalyl. The best fit structure from each docking was compared with the best hnRNP C docked structure. Considering the high sequence homologies between the three sequences, it is not surprising that not only the resulting heterodimers are all antiparallel leucine zippers with the **a** positions packed into two **a** and one **d** positions of the opposite chain in a "knob in the hole" pattern, but that all the resulting structures have high degree of alignment and extreme similarities in ϕ and ψ angles shown in the corresponding Ramachandran plots. Shown in **Figure 9** is the comparison between the best-fit structures of hnRNP C/Hraly heterodimer and hnRNP C homodimer. Considering the high sequence homology between hnRNP C and hRaly, the two structures have an all-residue-C_α RMSD of 0.82 Å. **Figure 10** shows the best fit structure of hnRNP C/hRalyl compared with the hnRNP C homodimer, and the two structures have an all residue backbone RMSD of 1.3 Å. These heterodimers were also analyzed with MolProbity and their respective ramachandran plots were generated. These plots can be seen in **Figure 11**. The plots still prove that the structures are accurate with 94.2% and 96.2% of the residues are in favored regions for for hnRNP C-hRaly, and hnRNP C-hRalyl, respectively. As shown before for the NMR

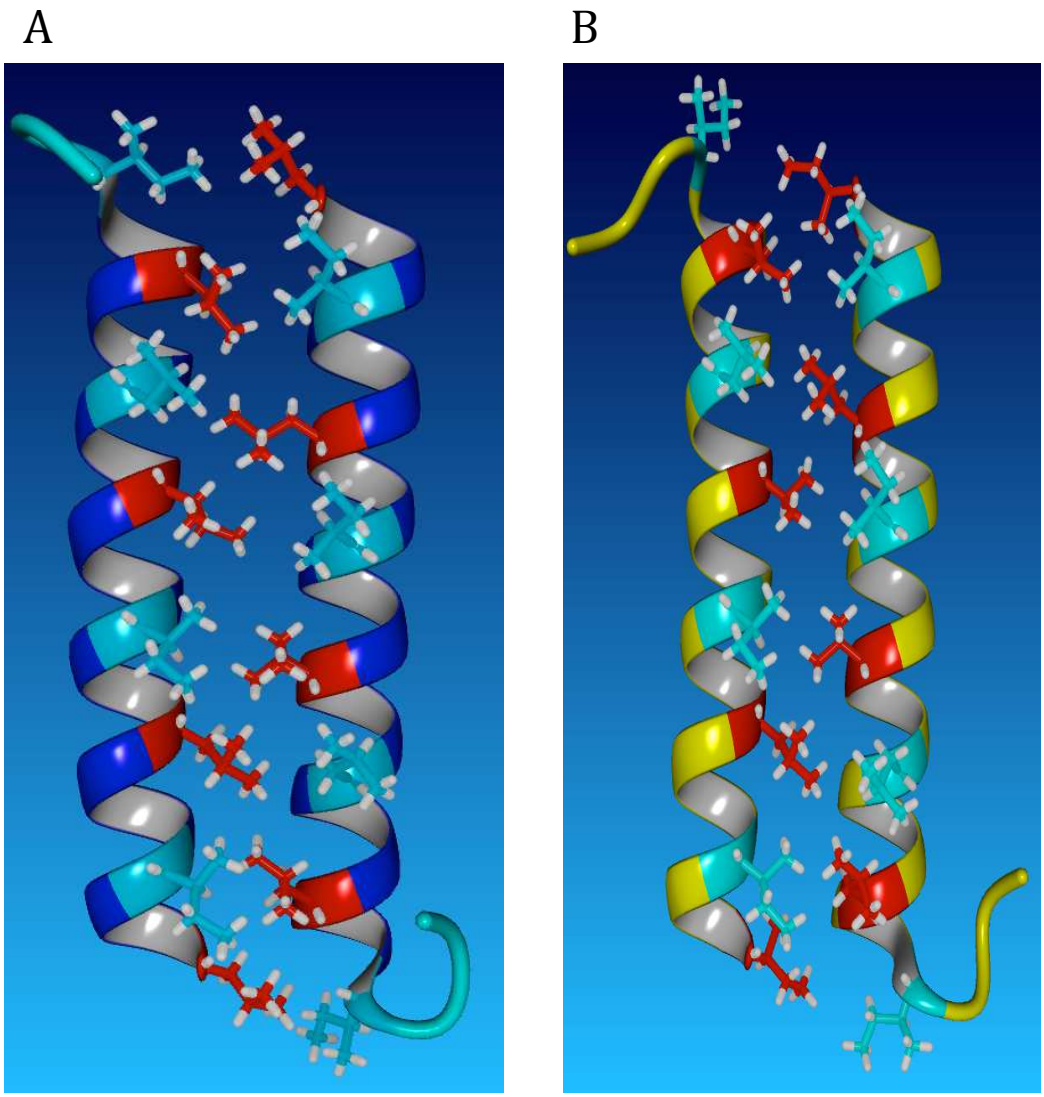


Figure 9. Panel A. The best fit modeled hnRNP C-hRaly hetro-dimer. Panel B. The best fit modeled hnRNP C homodimer. In both structures, the **a** positions in the four heptads are shown in red and the **d** position in the four heptads are shown in light blue.

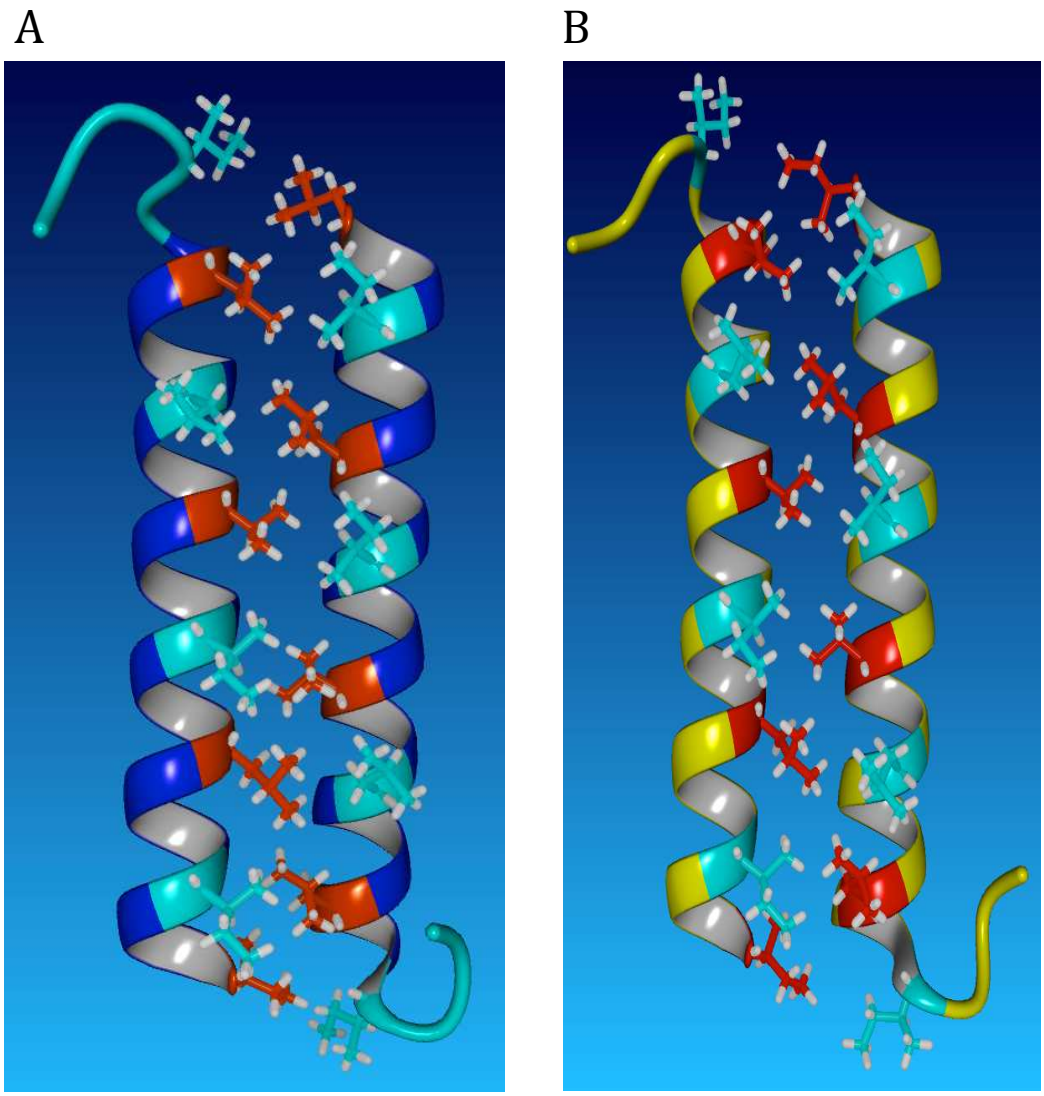


Figure 10. Panel A. The best fit modeled hnRNP C-hRalyl hetro-dimer. Panel B. The best fit modeled hnRNP C homodimer. In both structures, the **a** positions in the four heptads are shown in red and the **d** position in the four heptads are shown in light blue.

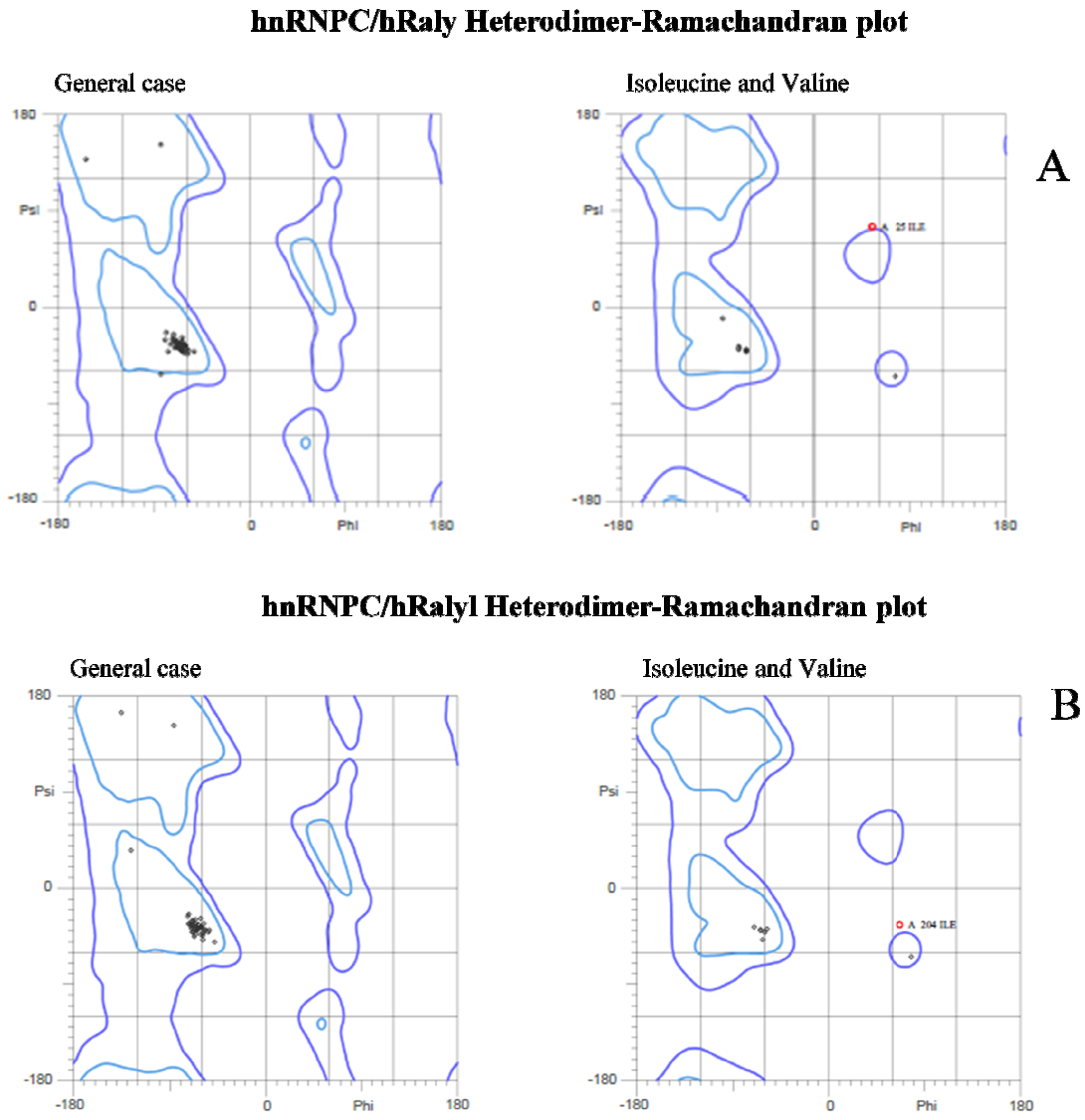


Figure 11. The Ramachandran plots representing the validity of the heterodimers hnRNP C-hRaly (Panel A) and hnRNP C and hRalyl (Panel B). The red dots at the right hand of the graph in panels A and B represent the residues outside of the favored regions.

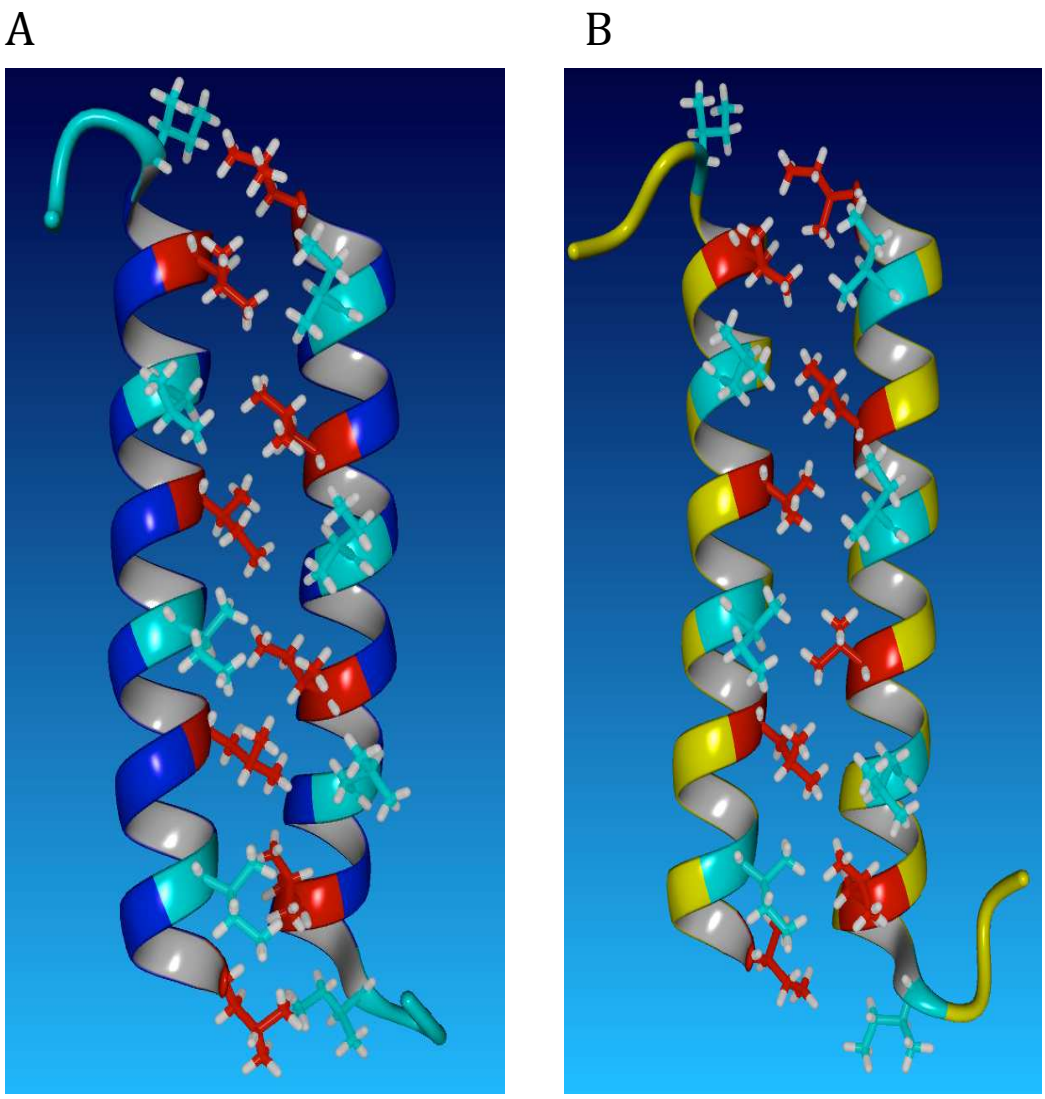


Figure 12. Panel A. The best fit modeled hRaly-hRalyl hetero-dimer. Panel B. The best fit modeled hnRNP C homodimer. In both structures, the **a** positions in the four heptads are shown in red and the **d** position in the four heptads are shown in light blue.

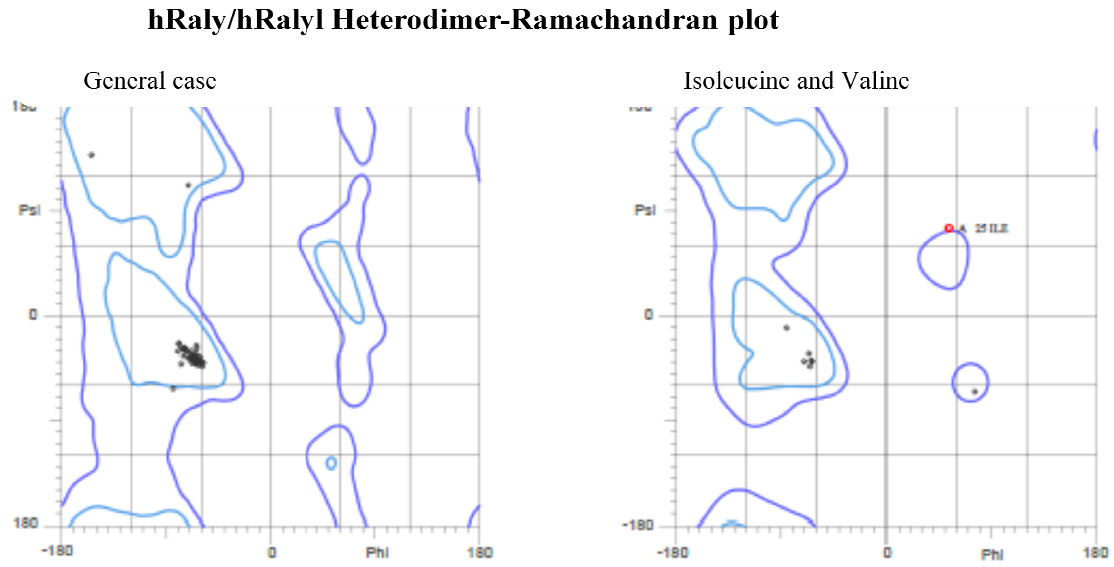


Figure 13. The Ramachandran plot of the heterodimer structure hRaly-hRalyl heterodimer. The red dot represents the Ile in chain A at position 25. This residue is not encountered in a favorable region.

structure of hnRNP C homodimer, the only residue out of the favored regions (represented as a red dot) is the c-terminus isoleucine. The final heterodimer generated from this group, hRaly/hRalyl, is shown in **Figure 12** which is compared with the hnRNP C homodimer where the all-residue C_α-RMSD of both structures was calculated to be also 1.3 Å. The Ramachandran plot corresponding to the hRaly-hRalyl heterodimer can be seen in **Figure 13.** The output encountered 94.2 % of the residues in the structure to be in the favored region of the plot. Again only the terminus isoleucine residue in the structure is found to be in an un-favored geometry of an α-helical structure.

The results discussed above that favors the formation of dimers between different protein sequences, prompted us to generate a mixed heptad monomer that consisted of heptad 2 of hRalyl, heptad 4 of hRaly, heptad 2 of hnRNP C and heptad 1 of hRaly in this

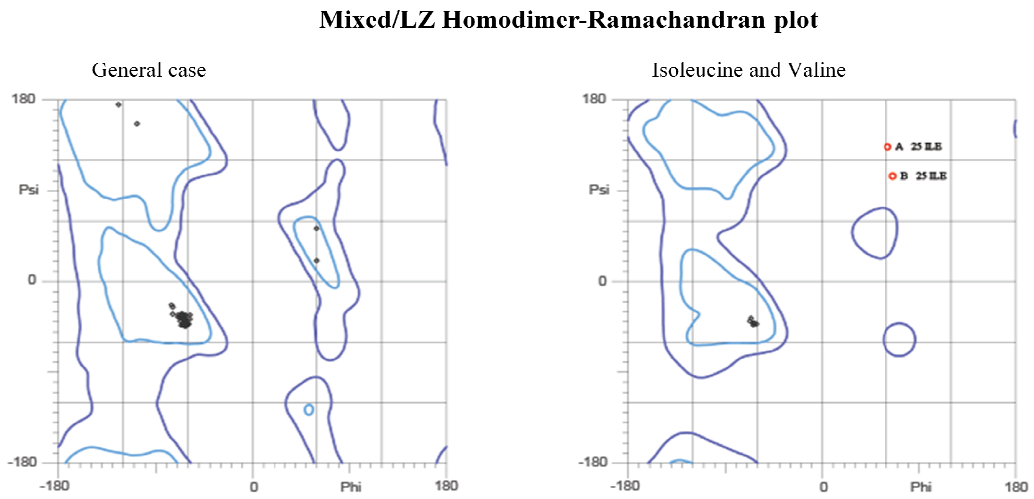


Figure 14. Ramachandran plot corresponding to the Mixed LZ homodimer. Two red dots (outliners) can be seen corresponding to each of the chains in the dimer.

order of N to C terminus. Molecular docking was then used to generate the resulting dimer (Mixed LZ). Shown in **Figure 15** is the Mixed LZ structure (Panel A) and the hnRNP C dimer. As shown in this figure, the Mixed LZ dimer is an anti-parallel leucine zipper with the accepted "knob in the hole" packing and has an all-residue C_{α}-RMSD of 1.1 Å which is as low and in some cases lower that hetero-dimers constructed from known protein sequences. Furthermore, analysis of the Ramachandran plot **(Figure 14)** for the Mixed LZ structure also indicate that 96.2 % of its residues lie in the favored region for α-helical structures with only the c-terminus isoleucine present as an outlier residue.

Although the modeled Mixed LZ structure overlaid quite well over the NMR structure of hnRNP C, there was concern about whether or not interactions beyond the hydrophobic ones were present in Mixed LZ or any of the other dimers. And as a result, all possible potential interactions needed to be considered. As expected for amphipathic helices, virtually all of the polar and charged residues were directed toward water.

However, none of these residues were involved in intra or inter-molecular interactions except for two that are conserved in the primary sequence of all three of the zippers. Specifically, the **e** and **b** positions of heptad 2 and 3 respectively contain a positively charged lysine and a negatively charged aspartate in the hnRNP C, hRaly, and hRalyl sequences. Shown in **Figure 16** are the intermolecular salt bridges that result from the

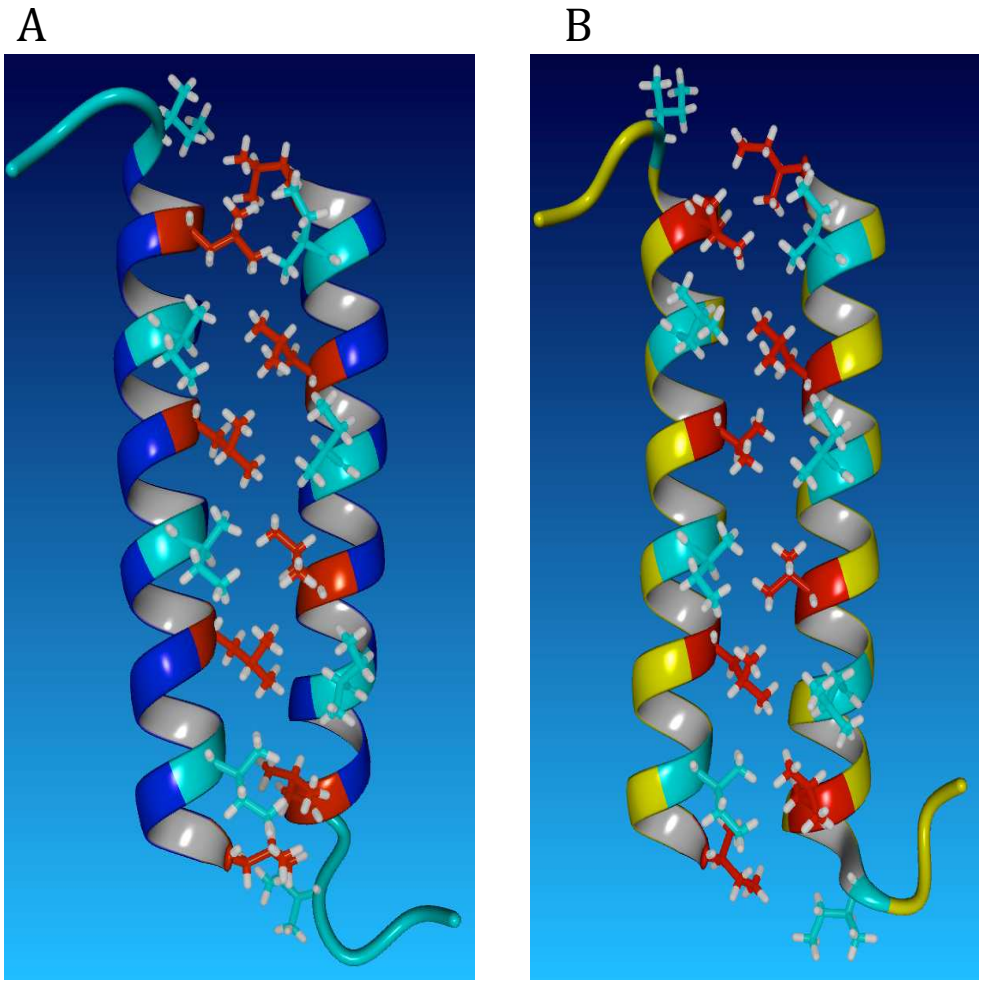


Figure 15. Panel A. The best fit modeled Mixed LZ homo-dimer. Panel B. The best fit modeled hnRNP C homodimer. In both structures, the **a** positions in the four heptads are shown in red and the **d** position in the four heptads are shown in light-blue.

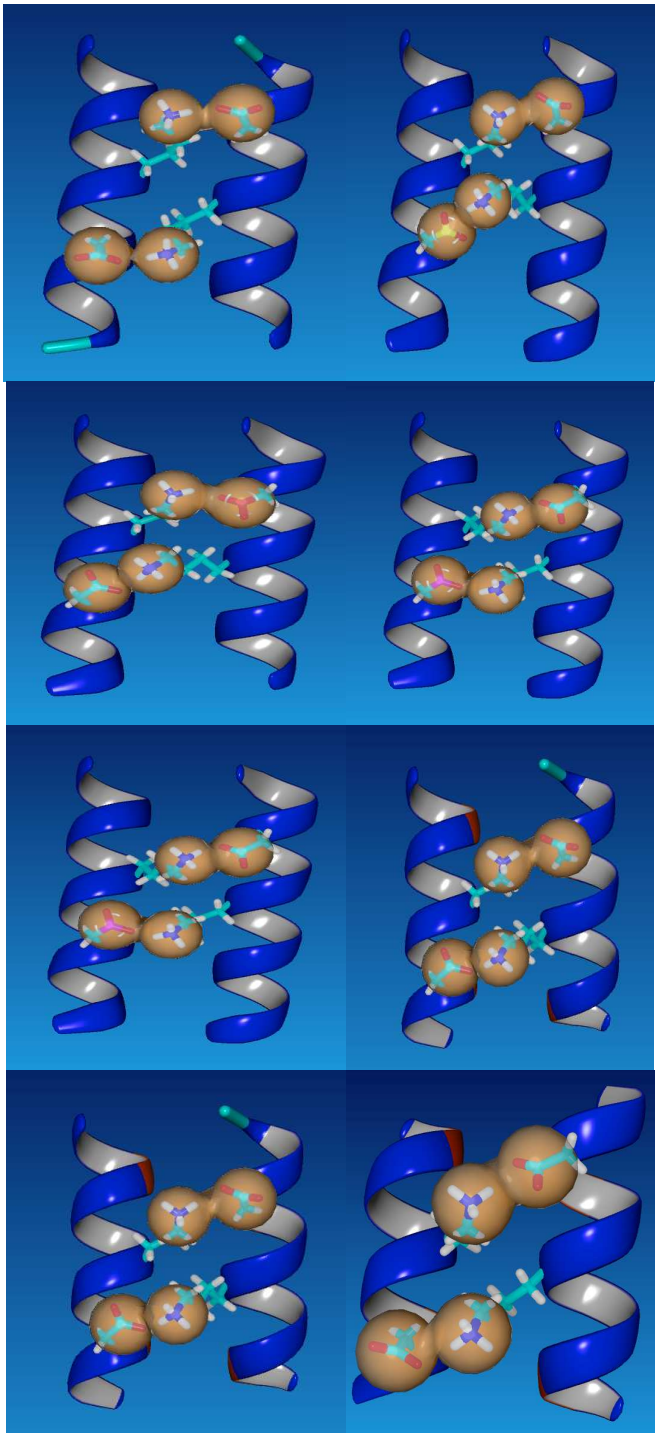


Figure 16. Zoomed images of the two salt bridges in hnRNP C, hRaly, hRalyl homo and hetero-dimers. Van der Waals surface contacts between the oxygen (atom shown in red) on the carbonyl group of aspartate #16 on one chain and the hydrogen on the amino group (atom shown in blue) on lycine #12 of the opposite chain. Panels A-H show the structures from hnRNP C homodimer (NMR), hnRNP C homodimer (modeled), hRaly (homodimer), hRalyl (homodimer), C-hRaly (heterodimer), C-hRalyl (homodimer).

interaction of these conserved residues. More specifically, the distances of the positively charged **e**-amino group of the lysine reside from the negatively charged carboxylate group on the aspartate ranged between a minimum of 2.3 Å to a maximum of 3.1 Å. These distances are well within the required distances for the formation of salt bridges.

To examine the effect of these two residues on the overall stability of the dimers, the **e** and **b** alanine and threonine residues of Mixed LZ were changed to the conserve lysine and aspartate to produce the Mixed LZ-SB sequence. This sequence was docked and the resulting dimer is shown in **Figure 17** along with the NMR structure of hnRNP C for comparison. The addition of these conserved residues did not alter the structure in that the C_α-RMSD value with respect to the NMR structure of hnRNP C is comparable to that observed for the other dimers (1.2 Å). Furthermore, the Ramachandran plot for this structure (**Figure 18**) indicates that all residues except for the c-terminus isoleucine are in regions with acceptable ϕ and ψ angles with 96.2% of those residues located in the favored region of the Ramachandran plot.

Binding Free Energies of the Docked Homo- and Hetero-Dimeric Structures Constructed from the hnRNP C, hRaly, hRalyl, and the Mixed Peptide Sequences

Shown in **Table 2** are the binding energies for all of the modeled structures. The ΔG_{solvH} which represents the difference between the hydrophobic residues being exposed to water versus their aggregation between the helices to exclude water molecules by far represents the major contribution to the stability of each dimer. Excluding the mixed leucine zipper (Mixed LZ) the average ΔG_{solvH} is similar for all of the structures with an average of -24.7 ± 0.7 kcal/mol. The value for the Mixed LZ was

A B

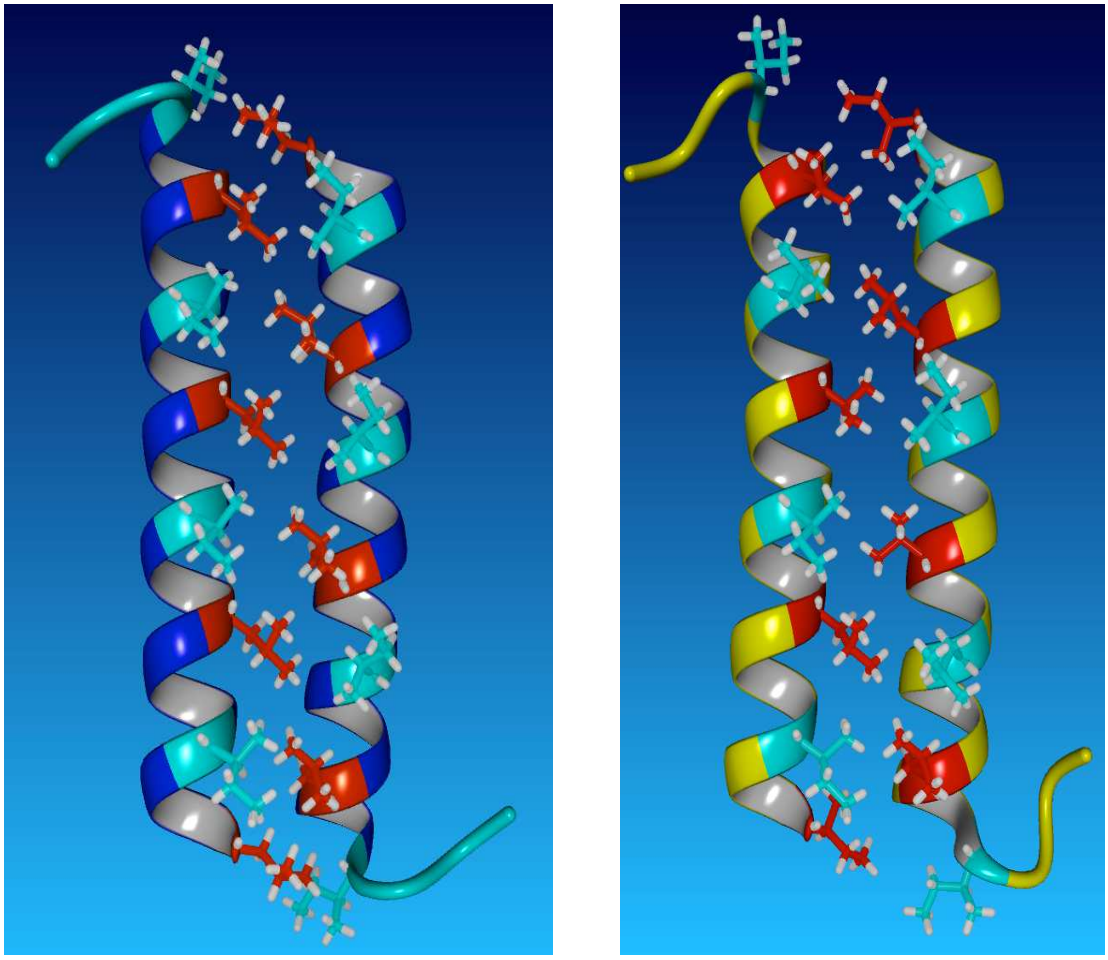


Figure 17. Panel A, the best fit modeled Mixed-Sequence homo-dimer mutated to include to salt bridges. Panel B. The best fit modeled hnRNP C homodimer. In both structures, the **a** positions in the four heptads are shown in red and the **d** position in the four heptads are shown in light blue.

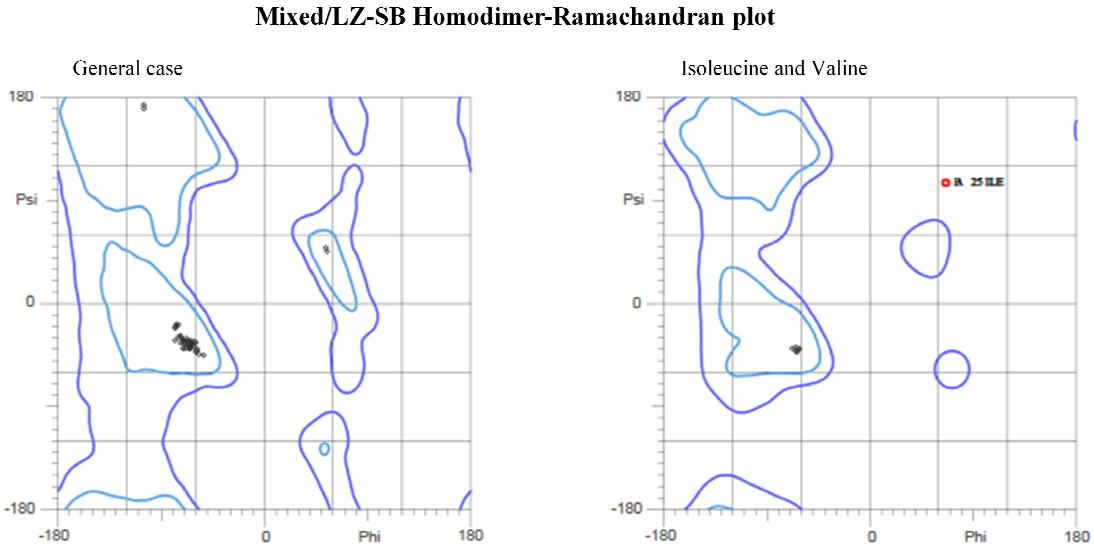


Figure 18. Ramachandran plots for the Mixed LZ-SB structure dimer. The black dots show the residues on the favored region and the red dots represent the residues outside of those regions. It is important to note that there are two red dots, one on top of the other and they correspond to each chain of the dimer.

-19.9 kcal/mol. The addition of a conserved salt bridge to Mixed LZ to produce Mixed LZ-SB brings the contribution of hydrophobic interactions back to a value consistent with the rest of the structures.

The change in the free energy associated with Van der Waals interactions was the second highest contributor to each dimer's stability. For all of the structures excluding the Mixed LZ, values were comparable ranging from a minimum of -12.10 kcal/mol for the hRaly/hRaly homodimer to a maximum value of -14.31 kcal/mol for the Mixed LZ-SB. The average free energy from Van der Waals interaction contributions for all of the dimers modeled was -12.87 ± 0.90 kcal/mol.

The free energy of electrostatic contributions (ΔG_{el}) to the stability of each structure was minor. All of the residues that would be positively charged at pH 7.0 (lysines) or those that would be negatively charged at this pH (glutamate and aspartate)

were not involved in any intramolecular salt bridges. However, a pair of residues that were conserved in the primary sequence of hnRNP C, hRaly, hRalyl, was noticed and inserted them into the sequence of Mixed-LZ-SB. This resulted in the formation of two intermolecular salt bridges that were observed in all the modeled structures listed above. Comparison of the binding energies of the Mixed LZ that lacked the salt bridges and the Mixed LZ-SB that has them, indicates that these salt bridges enhance the overall stability of the dimer by approximately 4 kcal/mol. Similar effects have been observed on GCN4, a basic leucine zipper that also functions as a transcription factor, where one salt bridge stabilizes the dimer by 1.7kcal/mol. (Spek et al., 1998)

The major destabilizing effect for all proteins is the decrease in the entropy associated with the packing of all of the amino acids into a globular structure where hydrophobic residues are excluded from water and hydrophilic ones are hydrated. This effect was minimized in the Mixed LZ dimer (8.4 kcal/mol) that lacked the salt bridges. The destabilizing effect was comparable in the other structures with an average value of 12.0 ± 1.4 kcal/mol.

The sum of these free energies led to values that were highly similar for all of the modeled dimers except the Mixed LZ dimer (-16.13 kcal/mol) that lacked the conserved salt bridge. The average value of the $\Delta G_{binding}$ for the remaining structures was determined to be -22.2 ± 1.5 kcal/mol.

Table 2. The Binding Free Energy for the Interaction of Monomer Leucine Zippers of Various Docked Dimers of the hnRNP C, hRaly, and hRalyl.

Molecule	ΔG_{SolvH} (kcal/mol)	ΔG_{vdw} (kcal/mol)	$\Delta G_{entropic}$ $(T\Delta S_{sc})$	ΔG_{el} (kcal/mol)	$\Delta G_{binding}$ (kcal/mol)	K_b
C-C-NMR	-25.12	-13.92	12.68	-0.85	-23.70	2.40438×10^{17}
C-C Modeled	-23.67	-12.40	13.59	-0.94	-20.81	1.82619×10^{15}
hRaly-hRaly	-25.29	-12.10	10.68	-0.72	-24.12	4.88675×10^{17}
hRalyl-hRalyl	-24.81	-12.22	11.34	-0.72	-21.07	2.83285×10^{15}
C-hRaly	-24.95	-12.27	11.32	-0.74	-21.79	9.55553×10^{15}
C-hRalyl	-25.04	-12.87	11.44	-0.75	-23.49	1.68653×10^{17}
Mixed-LZ	-19.85	-9.75	8.38	-0.00	-16.13	6.7507×10^{11}
Mixed-LZ-SB	-23.55	-14.31	14.58	-0.77	-20.56	1.01127×10^{15}

CHAPTER IV

DISCUSSION

hnRNP C has historically been viewed as a separate entity only linked to two polypeptide chains, C1 and C2. (Barnett et al., 1989) Virtually every function of the proteins has been linked to this perception. The discovery of hRaly in the 1990's and hRalyl in 2004 did little to dispel this perception despite the high degree of sequence homology between the proteins as well as the conservation of structural motifs. (Duhl et al., 1994; Michaud et al., 1994; Tomonaga and Levens, 1995). However, the research presented here questions the separate protein philosophy, in that this investigation clearly demonstrated that the oligomerization domains of each of the three proteins appear to be structurally equivalent. This conclusion is based upon the analysis of three findings. The first is obtained by overlaying the resulting modeled structures over the NMR structure of hnRNP C and visual inspection of how well the helical backbones and side chains align. The second is obtained through the analysis of the calculated all-residue C_α-RMSD value of the modeled dimers based in comparison the NMR structure of the hnRNP C dimer. And the third is obtained through the comparison of the ϕ and ψ angles of all the residues in the modeled structures relative to those observed in the NMR hnRNP C structure. From visual inspection, one could not discern the modeled structures from the NMR

structure of the hnRNP C dimer. This was reinforced quantitatively in that all-residue C_α-RMSD values ranged between a maximum of 1.3 Å for the hnRNP C/hRalyl to 0.8 Å Ralyl/Ralyl heterodimers. Two proteins are considered structurally identical with calculated Cα-RMSD value of less than 2 Å. (Raveh et al., 2010) Clearly, the resulted values are well below this threshold indicating structural identity.

In addition, it was possible to make a more detailed comparison of the structures to further confirm the validity of the modeled structures by analyzing their backbone geometries relative to the conformations of the residue side chains. This was done by generating contour plots of ϕ and ψ combinations of all residues in each modeled structure using Ramachandran plots. In all of the plots generated it was clear that more than 90% of all the residues for all of the modeled structures lie in the region representing favored a-helical conformations. These values are highly comparable to what was observed for the NMR structure of hnRNP C. This does not only confirm the validity of the modeled structures but clearly argues for the high probability that these structures represent physiologically accurate depiction of interactions between these three proteins in living cells.

The binding energies of all the structures presented here also confirm the structural identity and binding stability of all of the dimers modeled in this study. Equilibrium binding constants for monomer interactions ranged from a low of 7×10^{11} for the Mixed LZ structure to a high of 5×10^{17} for the hRaly homo-dimer. To gain a perspective of what these values mean, the lowest binding constant is characteristic of antigen antibody interactions which are among the strongest non-covalent interactions in living cells. The strongest non-covalent interaction known on earth is that between biotin

and streptovadin which approaches that of covalent interactions. The binding constant for this interaction has been determined to be in the range of 10^{15}. Based upon the very low cellular concentrations of proteins (10^{-9} M), reactions with these K_a values would represent irreversible binding. Though difference between the K_a of the weakest binding leucine zipper and the tightest binding leucine zipper obtained in this work represents several orders of magnitude, based on concentrations of these proteins in the cell virtually all of these reactions are irreversible (reaction quotient for these reactions is much larger than K_a).

One might argue that though the leucine zippers are compatible, the translation of each protein may be coupled to the formation of quaternary interactions. This is indeed a valid argument and one that was initially sought to be addressed by using recombinant DNA techniques. Specifically, the approach was to use a polycistronic vector to express different pairs of proteins at the same time as well as on different vectors. However, the expression of two copies of the proteins proved to be lethal in E. coli. Fortuitously, while preparing this research, recent findings have shown that hnRNP C and hRaly interact in an RNA independent fashion (Tenzer et al., 2013). This work reinforces the significance in the results reported in this work.

The studies reported here show that the binding energies for hnRNP C, hRaly, and hRalyl monomers to one another are comparable indicating the possibility of hetero-dimers or tetramers comprised of different combinations of each monomer. Since hnRNP C has been associated with so many cellular activities it sounds reasonable that one way to regulate its activity is through the generation of compositionally diverse proteins. For example, hnRNP C has been shown to bind RNA non-specifically

organizing it into a repetitive array of 40S monoparticles. The 40S monoparticle is comparable to the organization of chromatin into nucleosomes. Historically, nucleosomes were perceived to consist of the same protein composition and to be deposited in a sequence independent manner on DNA. Though histones are composed of two H2A-H2B dimers and one $(H3)_2(H4)_2$ tetramer, it is now known that there are histone variants for each of the four histones which allows functionally distinct nucleosomes. Based upon the findings mentioned in this paper, it can be suggested that the 40S monoparticle and other complexes that have been only linked to hnRNP C, probably are compositionally diverse and include hRaly and hRalyl monomers. As a result of these current findings immunoprecipitation studies will be conducted followed by western analysis to determine the viability of the hypothesis. It appears that one research group has confirmed aforementioned theories with regard to hnRNP C and hRaly. (Tenzer et al., 2013)

REFERENCES

Barnett, S. F., Friedman, D. L., and LeStourgeon, W. M. (1989). The C proteins of HeLa 40S nuclear ribonucleoprotein particles exist as anisotropic tetramers of (C1) 3 C2. *Molecular and Cellular Biology, 9*(2), 492-498.

Burd, C. G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science, 265*(5172), 615-621.

Chaudhury, S., and Gray, J. J. (2008). Conformer selection and induced fit in flexible backbone protein–protein docking using computational and NMR ensembles. *Journal of Molecular Biology, 381*(4), 1068-1087.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray L. W., Richardson J. S. and Richardson, D. C. (2009). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography, 66*(1), 12-21.

Choi, Y. D., and Dreyfuss, G. (1984). Monoclonal antibody characterization of the C proteins of heterogeneous nuclear ribonucleoprotein complexes in vertebrate cells. *The Journal of Cell Biology, 99*(6), 1997-1204.

Dreyfuss, G., Choi, Y. D., and Adam, S. A. (1984). Characterization of heterogeneous nuclear RNA-protein complexes in vivo with monoclonal antibodies. *Molecular and Cellular Biology, 4*(6), 1104-1114.

Duhl, D. M., Stevens, M. E., Vrieling, H., Saxon, P. J., Miller, M. W., Epstein, C. J., and Barsh, G. S. (1994). Pleiotropic effects of the mouse lethal yellow (Ay) mutation explained by deletion of a maternally expressed gene and the simultaneous production of agouti fusion RNAs. *Development, 120*(6), 1695-1708.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M., Eramian, D., Shen, M.-y., Pieper, U. and Sali, A. (2006). Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*. DOI:15:5.6:5.6.1–5.6.30.

Ford, L. P., Wright, W. E., and Shay, J. W. (2002). A model for heterogeneous nuclear ribonucleoproteins in telomere and telomerase regulation. *Oncogene, 21*(4),580-583.

Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L., and Dreyfuss, G. (1992). Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *The EMBO Journal*, *11*(9), 3289-3295

Huang, M., Rech, J. E., Northington, S. J., Flicker, P. F., Mayeda, A., Krainer, A. R., and LeStourgeon, W. M. (1994). The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. *Molecular and Cellular Biology*, *14*(1), 518-533.

Ji, C. N., Chen, J. Z., Xie, Y., Wang, S., Qian, J., Zhao, E., Jin W., Wu X. Z., Xu W. X., Ying K., and Mao, Y. M. (2003). A novel cDNA encodes a putative hRALY-like protein, hRALYL. *Molecular Biology Reports*, *30*(1), 61-67.

Jurica, M. S., Licklider, L. J., Gygi, S. P., Grigorieff, N., and Moore, M. J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, *8*(4), 426-439.

Kamma, H., Horiguchi, H., Wan, L., Matsui, M., Fujiwara, M., Fujimoto, M., Yazawa, T., and Dreyfuss, G. (1999). Molecular characterization of the hnRNP A2/B1 proteins: tissue-specific expression and novel isoforms. *Experimental Cell Research*, *246*(2), 399-411.

Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J., and Lesk, A. M. (2006). MUSTANG: a multiple structural alignment algorithm. *Proteins: Structure, Function, and Bioinformatics*, *64*(3), 559-574.

Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins: Structure, Function, and Bioinformatics*, *47*(3), 393-402.

Krylov, D., and Vinson, C. R. (2001). Leucine zipper. *eLS*. DOI:10.1038/npg.els.0003001

Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, *240*(4860), 1759-1764.

London, N., Raveh, B., Cohen, E., Fathi, G., and Schueler-Furman, O. (2011). Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions. *Nucleic Acids Research*, *39*(suppl 2), W249-W253.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*, *29*(1), 291-325.

McAfee, J. G., Shahied-Milam, L., Soltaninassab, S. R., and LeStourgeon, W. M. (1996). A major determinant of hnRNP C protein binding to RNA is a novel bZIP-like RNA binding domain. *RNA*, *2*(11), 1139.

McAfee, J. G., Soltaninassab, S. R., Lindsay, M. E., and LeStourgeon, W. M. (1996). Proteins C1 and C2 of heterogeneous nuclear ribonucleoprotein complexes bind RNA in a highly cooperative fashion: support for their contiguous deposition on pre-mRNA during transcription. *Biochemistry*, *35*(4), 1212-1222.

Merrill, B. M., Barnett, S. F., LeStourgeon, W. M., and Williams, K. R. (1989). Primary structure differences between proteins C1 and C2 of HeLa 40S nuclear ribonucleoprotein particles. *Nucleic Acids Research*, *17*(21), 8441-8449.

Michaud, E. J., Bultman, S. J., Klebig, M. L., Van Vugt, M. J., Stubbs, L. J., Russell, L. B., and Woychik, R. P. (1994). A molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow (Ay) mutation. *Proceedings of the National Academy of Sciences*, *91*(7), 2562-2566.

Nakielny, S., and Dreyfuss, G. (1997). Nuclear export of proteins and RNAs. *Current Opinion in Cell Biology*, *9*(3), 420-429.

Peetha, S., Analysis of heterotypic protein-protein interactions between hRaly and hRalyl protomers. M.S. thesis, Pittsburg State University, 2013

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605-1612.

Raveh, B., London, N., and Schueler-Furman, O. (2010). Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins: Structure, Function, and Bioinformatics*, *78*(9), 2029-2040.

Spek, E. J., Bui, A. H., Lu, M., and Kallenbach, N. R. (1998). Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Science*, *7*(11), 2431-2437.

Swanson, M. S., Nakagawa, T. Y., LeVan, K., and Dreyfuss, G. (1987). Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Molecular and Cellular Biology*, *7*(5), 1731-1739.

Tenzer, S., Moro, A, Kuharev, J, Francis A., Vidalino, L, Provenzani , A. , and Macchi , P. (2013). Proteome-Wide Characterization of the RNA-Binding Protein RALY-Interactome Using the in Vivo-Biotinylation-Pulldown-Quant (iBioPQ) Approach. *J. Proteome Research, 12 (6),* 2869–2884

Tomonaga, T., and Levens, D. (1995). Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *Journal of Biological Chemistry*, *270*(9), 4875-4881.

Van Durme, J., Delgado, J., Stricher, F., Serrano, L., Schymkowitz, J., and Rousseau, F. (2011). A graphical interface for the FoldX forcefield. *Bioinformatics*, *27*(12), 1711-1712.

Vaughan, J. H., Nguyen, M. D., Valbracht, J. R., Patrick, K., and Rhodes, G. H. (1995). Epstein-Barr virus-induced autoimmune responses. II. Immunoglobulin G autoantibodies to mimicking and nonmimicking epitopes. Presence in autoimmune disease. *Journal of Clinical Investigation*, *95*(3), 1316-1327

Vaughan, J. H., Valbracht, J. R., Nguyen, M. D., Handley, H. H., Smith, R. S., Patrick, K., and Rhodes, G. H. (1995). Epstein-Barr virus-induced autoimmune responses. I. Immunoglobulin M autoantibodies to proteins mimicking and not mimicking Epstein-Barr virus nuclear antigen-1. *Journal of Clinical Investigation*, *95*(3), 1306-1315

Wan, L., Kim, J. K., Pollard, V. W., and Dreyfuss, G. (2001). Mutational definition of RNA-binding and protein-protein interaction domains of heterogeneous nuclear RNP C1. *Journal of Biological Chemistry*, *276*(10), 7681-7688.

Whitson, S. R., LeStourgeon, W. M., and Krezel, A. M. (2005). Solution structure of the symmetric coiled coil tetramer formed by the oligomerization domain of hnRNP C: implications for biological function. *Journal of Molecular Biology*, *350*(2), 319-337.