Marquette University
e-Publications@Marquette

Dissertations (2009 -) | Dissertations, Theses, and Professional Projects

NMR-Based and Automated Docking Characterization of Protein Structure, Dynamics, and Ligand Binding

Andrew Lawrence Olson
Marquette University

Recommended Citation
Olson, Andrew Lawrence, "NMR-Based and Automated Docking Characterization of Protein Structure, Dynamics, and Ligand Binding" (2010). *Dissertations (2009 -)*. Paper 81.
http://epublications.marquette.edu/dissertations_mu/81

NMR-BASED AND AUTOMATED DOCKING CHARACTERIZATION OF PROTEIN STRUCTURE, DYNAMICS, AND LIGAND BINDING

by

Andrew L. Olson

A Dissertation Submitted to the Faculty
of the Graduate School, Marquette University,
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

Milwaukee, Wisconsin

December 2010

ABSTRACT

NMR-BASED AND AUTOMATED DOCKING CHARACTERIZATION OF PROTEIN STRUCTURE, DYNAMICS, AND LIGAND BINDING

Andrew L. Olson

Marquette University, 2010

NMR-based methods used in conjunction with a technique called docking are used to characterize ligand binding to proteins. Standard NMR methods were used to study the backbone dynamics of substrate binding to phosphomevalonate kinase (PMK) and it was observed that ligand binding caused PMK to undergo large conformational changes. These changes were reflected by the appearance of many chemical shift changes upon binding of the natural substrates of PMK (both the binary and ternary complexes) in ^{1}H-^{15}N HSQC NMR titration experiments. The same process was used to characterize the effect ligand binding has on the many arginines in the active site (and distal to the active site) to determine the effect of long-range coulombic interactions on ligand binding. While studying the backbone dynamics of PMK it was discovered that the N-terminal tail of PMK consisting of 10 residues was very disordered which is unlike every other monphosphate kinase. The function of this N-terminal tail was investigated by attempting to find other proteins in human liver cells that bind this peptide, monitored by ESI mass spectrometry.

The thioredoxin system of Mycobacterium Tuberculosis consists of a thioredoxin reductase and three thioredoxins. To help facilitate the understanding of this mechanism the solution structures of the oxidized and the reduced forms of thioredoxin C (TrxC) were solved by NMR and modeled with the crystal structure of the thioredoxin reductase complex. The two redox states of TrxC are very similar to each other with most of the differences coming from subtle changes in the active site of TrxC.

Automated docking is the process of computationally determining how a ligand binds to a protein and the correct orientation. A large scale docking study, termed virtual screening, was carried out by docking 10,590 compounds into three proteins to find inhibitors for each protein, and those predicted to bind best were tested experimentally. For each protein there were 3 compounds found to bind with reasonable affinity.

When ligands bind to a protein they can undergo dynamic changes. To explore this phenomenon, ^{15}N labeled NAD+ cofactor (and other derivatives) was synthesized and bound to oxidoreductases. Relevant binding motions were monitored using CPMG relaxation NMR experiments.

ACKNOWLEDGEMENTS

Andrew L. Olson

My time at Marquette University has been a memorable experience that has been made all the more special by the advisors, teachers, colleagues, and friends that I have come to know. In particular I would like to thank my research advisor, Dr. Daniel Sem, for the opportunity he has given me to grow as a scientist and as a person. The information, support, and guidance that he has given me are invaluable and I am extremely grateful for this.

I would also like to express my gratitude to Dr. Sheng Cai, for the knowledge he has given me and the help that he provided in using the NMR spectrometer, Dr. Craig Struble, for the assistance in setting up our virtual screening sessions, and to all of my group members and close friends; Xia Ge, Tyler Morin, Timothy Jonhera, Terrence Neumann, and Junghwa Kim.

I would especially like to thank my parents and brother, Kevin and Gayle Olson, and Steven Olson, for their continuous encouragement. Most of all, I would I like to thank my wife, Melissa Olson, for all the love and support she has given me. I am so grateful she was here to always encourage me and to help raise our beautiful son, Carter.

TABLE OF CONTENTS

CHAPTER I. Introduction to the Methods Used to Characterize Protein Structure, Function, and Ligand Binding.

1.1 Introduction

To study a protein and its interactions with substrates can be a difficult task since one needs knowledge of the protein structure as well as how the protein behaves upon binding its specific ligand. To help facilitate this task we employ standard NMR methods to study the protein structure and the dynamical properties associated with ligand binding. NMR is special in this way because we can observe ligand binding and residual dynamics information site specifically, since ^{15}N-labeled protein gives us information about each amino acid residue (amide bond) in the protein (except prolines) in the ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Correlation) experiment (Figure 1). After we have a structure and relevant dynamical information we can begin to assess how ligands bind to the protein, not only its natural substrates but also inhibitors, and other proteins. This can be done computationally, using a method called automated docking.

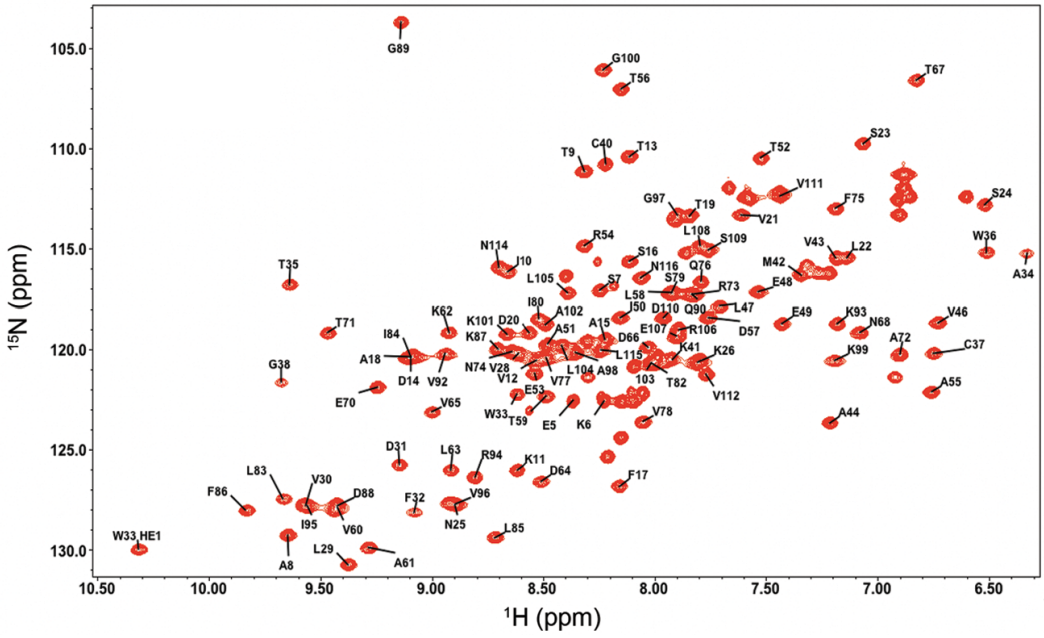


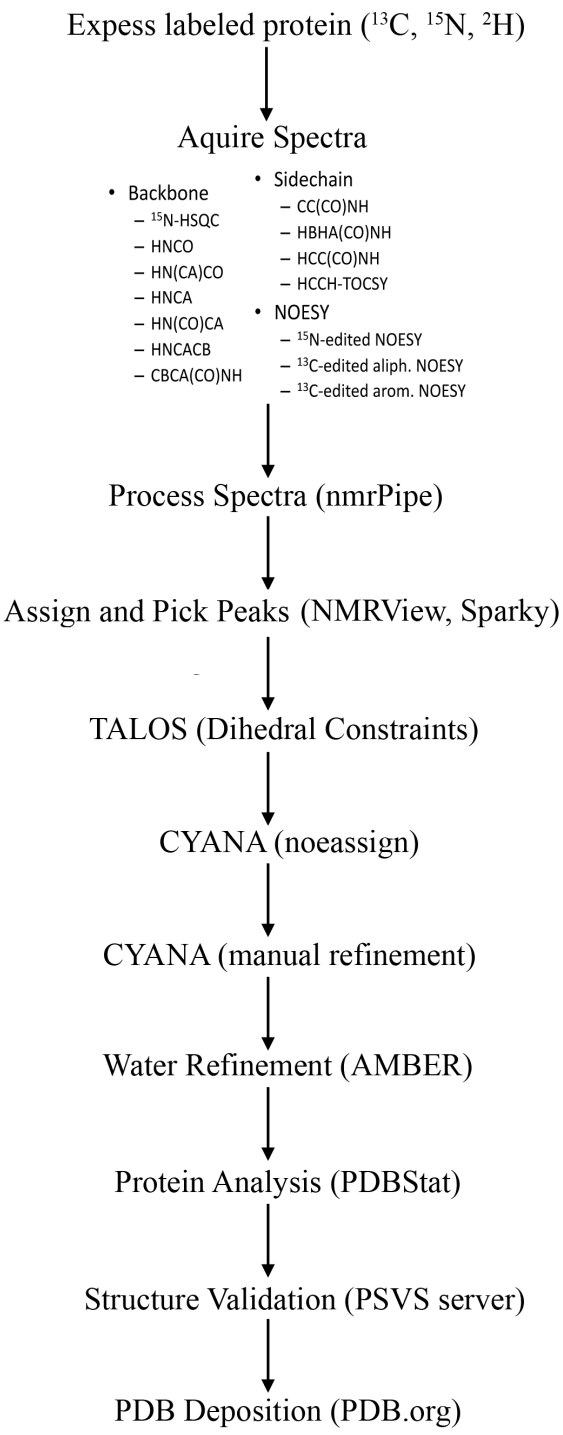
Figure 1.1. Heteronuclear single quantum correlation experiment (^{1}H-^{15}N HSQC) of Thioredoxin C protein with amide (N-H) crosspeaks labeled with amino acid residue numbers.

1.1.2 Protein Structure Determination using NMR

Protein structures, due to their size, can be difficult to solve. The two common methods for this are x-ray crystallography and NMR. The advantage of using NMR is that the protein is in its native state (being in solution) and not packed into a solid crystal. There are limitations however, mostly due to the size of the protein, in that for proteins > 25 kDa crosspeak overlap is an issue. To solve protein solution structures using NMR, an extensive set of experiments are acquired on ^{13}C-^{15}N-(^{2}D) labeled protein (Scheme 1.1). The backbone amides of the ^{1}H-^{15}N HSQC are generally assigned first using 6 3-D NMR experiments, which are 1) HNCO and HN(CA)CO, 2) CBCA(CO)NH and HNCACB, 3) HNCA and HN(CO)CA. Briefly, magnetization starts on the amide proton and transfers to the nitrogen, which then transfers to the following atoms in the sequence (for example, to Cα and Cβ carbons, in the HNCACB experiment), which are described in the title of the experiment. These experiments are run in the pairs stated above, in which one experiment's chemical shifts will align themselves in the exact same chemical shift as its partner experiment, allowing for that chemical shift to be assigned (Figure 1.2). This is done until every chemical shift in the ^{1}H-^{15}N HSQC is assigned.

The backbone assignments are then used to aid in the assignment of the sidechain protons and carbons of all the residues that appear in the ^{1}H-^{15}N HSQC (Figure 1.1). We do this with the 3-D experiments CBCA(CO)NH and CC(CO)NH which are used to assign the carbons of the side chains, HBHA(CO)NH and HCC(CO)NH which are used to assign all the protons in the side chains, and with HCCH-TOCSY which correlates all protons and carbons together. Getting as many assignments made as possible is very important as this will lead to more distance constraints for structure generation. These

constraints come from ^{13}C-NOESY (aromatic and aliphatic) and ^{15}N-NOESY spectra. These experiments are also 3-D experiments that have the heteroatom chemical shift as the third dimensions, and are directly used in structure calcuation. The NOESY experiments are based on the Nuclear Overhauser Effect in which cross peaks arise if there are protons in proximity of each other (< 5 Å). Since there are many residues in a protein, all whose protons "see" other protons through the NOE (cross peak), it is common to have up to 5000-6000 cross peaks for a (~12 kDa protein) amongst the three NOESY spectra. These cross peaks are used to determine the structure of the protein since they all provide distance restraints, which the program CYANA[1, 2] uses to calculate an ensemble of protein structures based on torsion angle dynamics, a molecular dynamics simulation that uses only torsion angles of the backbone as degrees of freedom instead of Cartesian coordinates. CYANA only calculates the proper folding of the protein based on these NOE-derived, TALOS (chemical shifts)[3], and covalent bonding (connectivity) constraints. To get the correct geometry of the side chains with full consideration of inter-atom energetic effects (ex. H-bonds and Van der Waals interactions), the calculated structure undergoes water refinement which is a simulated annealing protocol using the AMBER molecular force field where the protein is heated rapidly to overcome any local minima barriers, it is then allowed to explore higher energy states, and is then cooled slowly so that it finds its lowest energy minima. This is done repeatedly (ex. 100 times) to generate an "ensemble" of structures that are consistent with all constraints. Precision of the calculated structures is quantified using a RMSD (root mean squared deviation) for atoms in the protein backbone and again for all heavy atoms.



Scheme 1.1. Flow chart detailing the steps to solving protein structure using NMR.

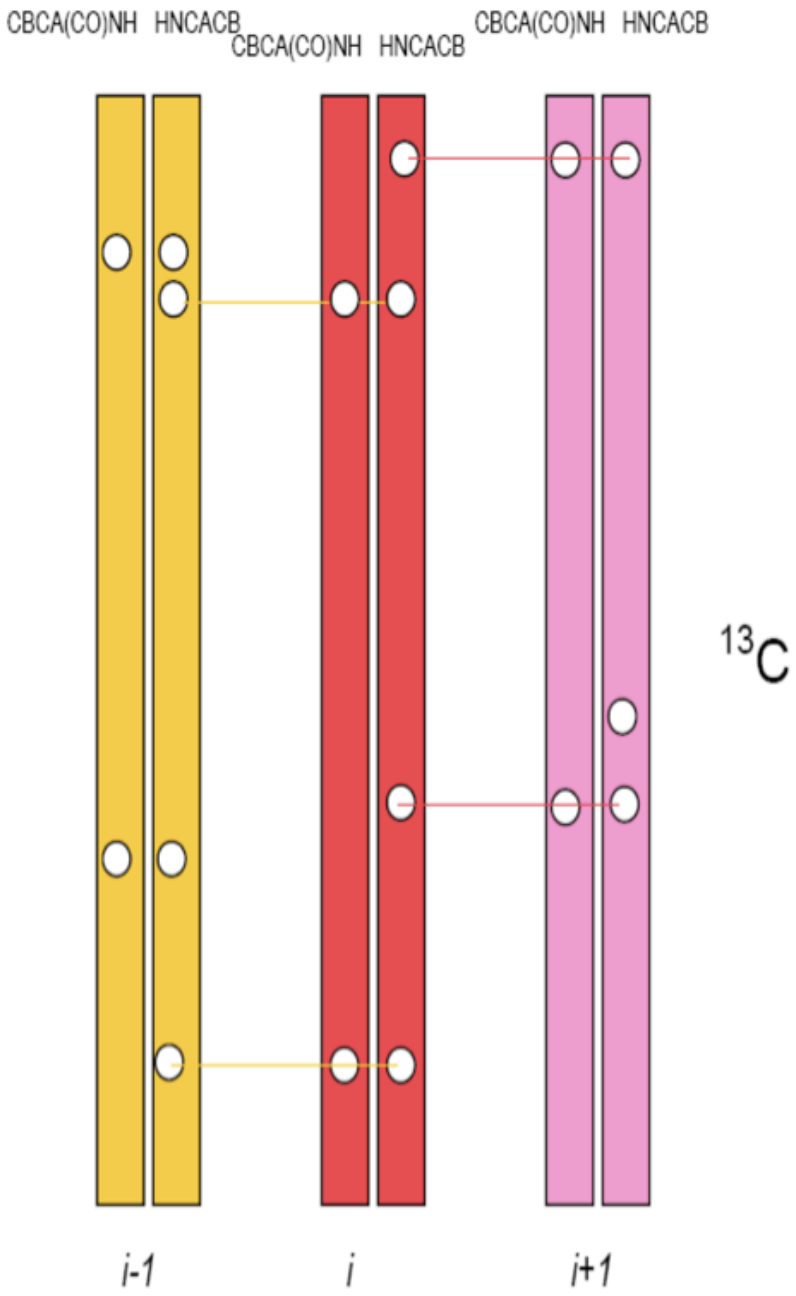


Figure 1.2. Illustration of sequential assignment process using 3D HNCACB and CBCA(CO)NH spectra. Residues are i (amino acid of interest), i-1 (amino acid proeceeding i), and i+1 (amino acid after i). Shown are pairs of strips from ^{1}H-^{13}C planes at a particular ^{15}N chemical shift. Typically, there would be one such pair of strips (^{15}N planes) per amino acid, in a protein (except for prolines and the N-terminus)

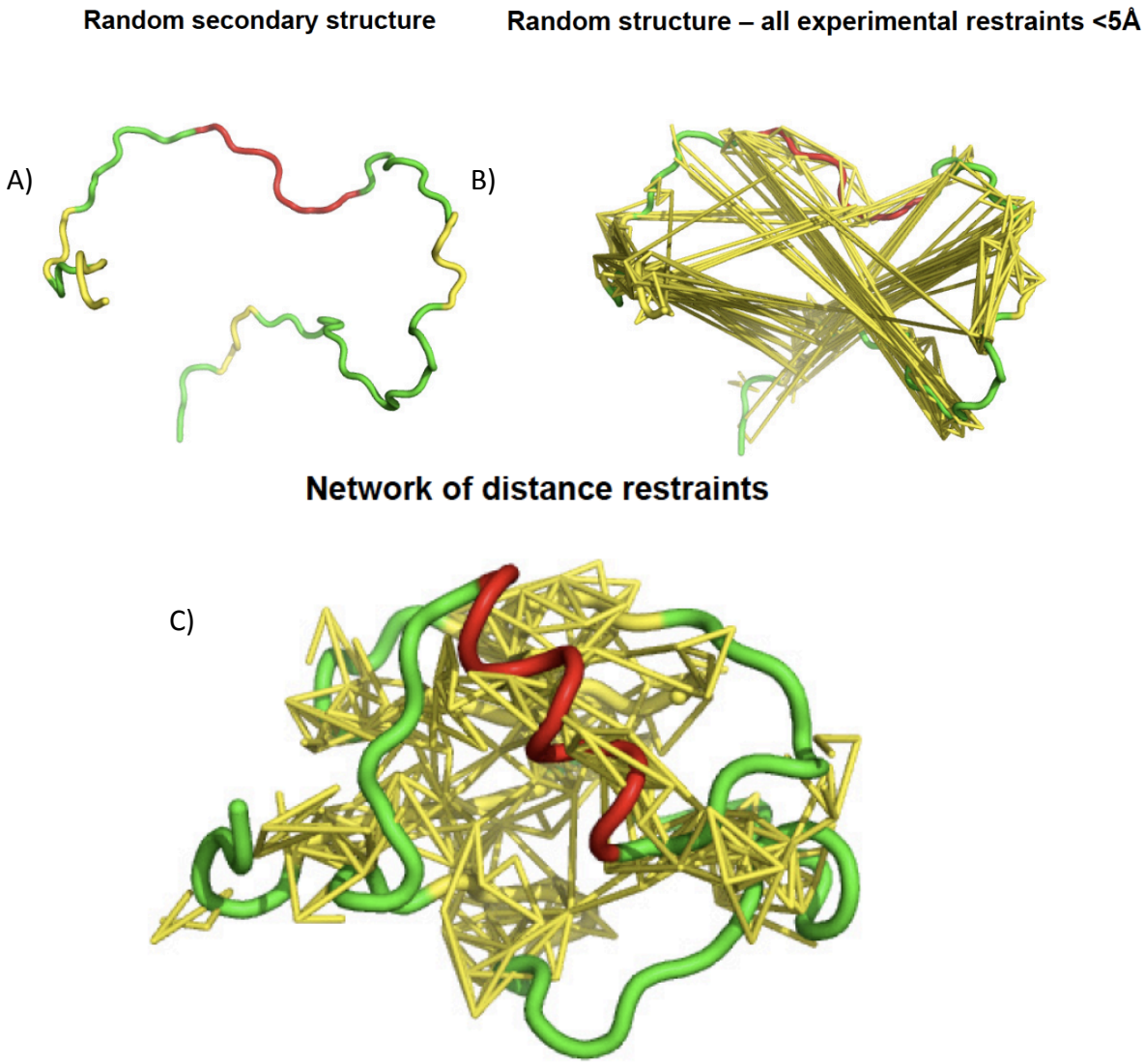


Figure 1.3. Representation of a A) random secondary structure, B) random secondary structure with all the restraints from the ^{13}C and ^{15}N NOESY experiments applied, and C) the experimental restraints turned into distance restraints in CYANA calculations, used to define the secondary and tertiary structure. Figures taken from NMRFAM Workshop.

1.1.3 Relaxation Dynamics Studies of Protein Backbone N-H's

Protein motions are critical for various biological functions and occur over a wide range of timescales. There is a strong correlation between structural dynamics and molecular function, and studying these dynamic processes can give site-specific information on motions that span timescales from psec to msec[4]. We can characterize these dynamic processes by use of NMR with ^{15}N-labelled protein and relaxation experiments (measure longitudinal relaxation rate R_1 ($1/T_1$), transverse relaxation rate R_2 ($1/T_2$), and the heteronuclear NOE for the amide N-H bond) in conjunction with a Modelfree analysis. The Modelfree analysis utilizes the R_1, R_2, and NOE values obtained from the relaxation dynamics experiments, and the spectral density equations:

$$R_1 = 1/4\ d^2[3J(\omega_N) + 7J(\omega_h) + c^2J(\omega_N) \qquad \textbf{Eq. 1.1}$$

$$R_2 = 1/8\ d^2[4J(0) + 3J(\omega_N) + 13J(\omega_h)] + 1/6c^2[4J(0) + 3J(\omega_N)] \qquad \textbf{Eq. 1.2}$$

$$NOE = 1 + \tfrac{1}{4}\ T_1 d^2(\gamma_H/\gamma_N)[5J(\omega_h)] \qquad \textbf{Eq. 1.3}$$

$$J(\omega) = 2/5[(S^2\tau_m/1+ (\omega\tau_m)^2) + ((S_f{}^2 - S^2)\tau/1+ (\omega\tau_m)^2 \qquad \textbf{Eq. 1.4}$$

in which the spectral density equations are fitted at each Larmor frequency ω (nitrogen and proton) and at given field strength using experimentally measured values of relaxation parameters (R_1, R_2, and NOE) to find S^2 and τ_m. In the above equations, c and d are both constants; γ is the gyromagnetic ratio for nitrogen or proton; S^2 is the generalized order parameter for movement of the N-H bond vector; S_f is the generalized order parameter on the sub-nsec timescale; τ_m is the isotropic rotational correlation time for the whole protein, and $\tau = \tau_m\ \tau_e / (\tau_e + \tau_m)$ where τ_e is the effective correlation time for internal motions within the protein (ex. Loop motions)[5-8]. This allows us to quantify any changes to the fast timescale motion (psec-nsec) of the protein backbone, especially for

catalytically relevant regions, by solving to find the generalized order parameter S^2 (S^2 =1 means completely restricted and immobile, S^2 =0 means no restriction and completely mobile). The Modelfree analysis can also provide insight into the state of the overall conformational motions of the protein, and how they change upon substrate addition, by measuring the correlation time (τ_m) for each complex (high τ_m means the protein tumbles slower and behaves like a larger protein; low τ_m means the protein tumbles faster and behaves like a smaller protein).[8]

1.1.4 Automated Docking and Virtual Screening

Docking is the process of positioning a ligand into the binding site of a protein, by computationally exploring different translations, orientations, and conformations until a lowest energy structure and "pose" orientation is found. There are many programs that are fully capable of virtual screening such as DOCK[9], FlexX[10], Glide[11], and AutoDock[12]. AutoDock is currently the docking program that is in use for virtual screening at Marquette University. AutoDock uses a genetic algorithm (GA) to sample multiple conformations of a ligand when binding to a protein. AutoDock will use those coordinates and make derivations to the ligand structure based on the "fitness" level of the previous generation of ligands, meaning if a certain ligand binds poorly the coordinates of that ligand will not be used for further "generations." However, if a ligand binds well, AutoDock will keep those coordinates and then optimize the binding by sampling other torsions of the ligand. "Fitness" is calculated using a target function that is similar to a forcefield equation that describes energy of interactions between the

protein and ligand. The GA optimizes the ligands pose to obtain the lowest energy function.

To quantitatively assess how well a ligand binds, AutoDock employs an "expanded" version of the free energy equation shown below,

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol} \quad \textbf{Eq. 1.5}$$

Where the first four terms are for the free energy terms for Van der Waals forces, hydrogen bonding, electrostatics, and deviations from covalent geometry and are typical for molecular mechanics free energy equations. The latter two terms are the expanded portion of the equation and are the free energy terms for rotation of internal torsions (internatl entropy) and desolvation upon binding (modeling hydrophobic effects and solvent entropic changes upon binding)[12].

The experimental information that AutoDock provides is written out at the end of the docking session in a docking log file (dlg file). After the docking is finished and the 50 poses are generated (Figure 1.4), AutoDock then puts them through a clustering analysis to define which poses are similar enough to each other to cluster them together (within an RMSD of 2 Å). Clustering offers an additional verification of the likelihood that a certain pose is the correct pose for the ligand binding to the protein. Figure 1.5 shows examples of this clustering analysis where A) had a very low energy of binding compound at (-10.11 kcal/mol) but there are many different clusters with few poses in each cluster. In practice, this indicates a low probability of binding. Figure 1.5B shows a cluster histogram for a compound with a slightly higher binding energy, but the

probability of this being an accurate pose is higher since there are fewer number of sparsely populated clusters, yet in the lowest energy cluster there are nearly half the number of poses clustered together. In general, an accurate pose (and prediction that a compound binds to a protein) is more likely if there is one populated cluster.

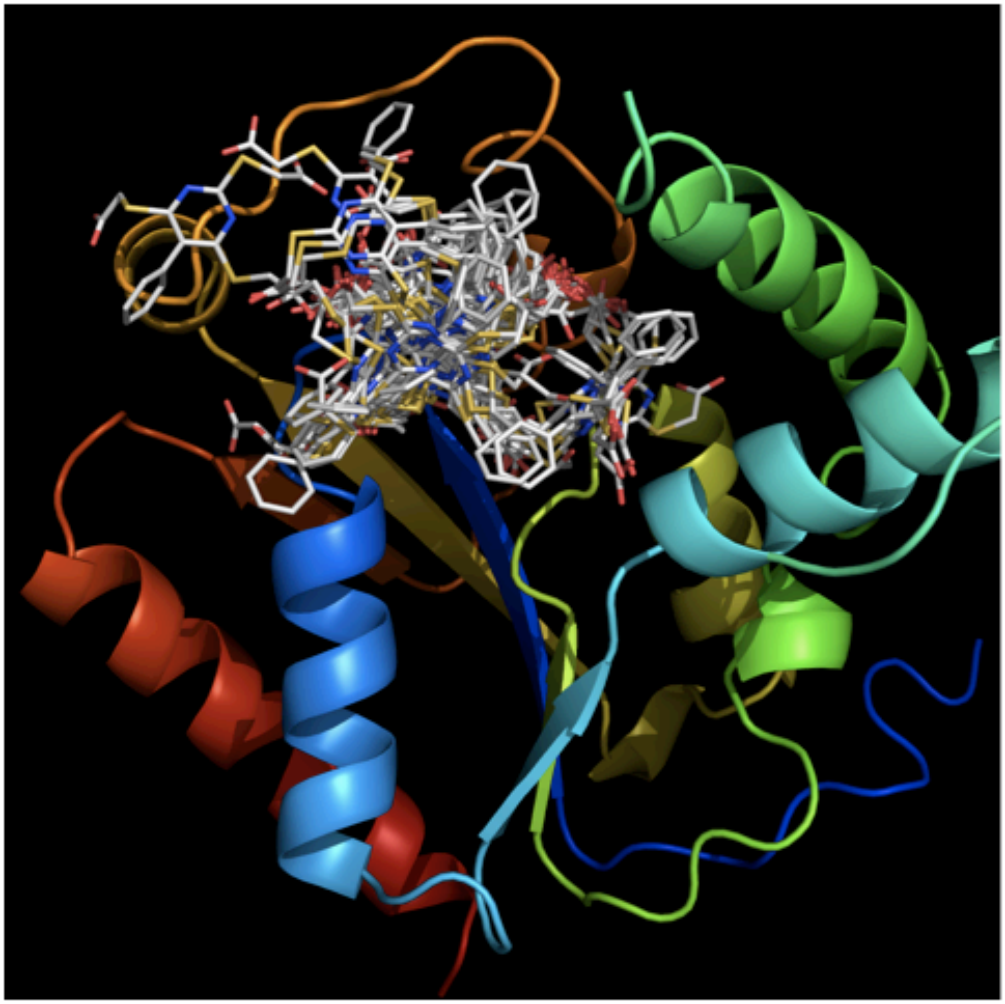


Figure 1.4. A docking experiment where AutoDock generated 50 poses (all 50 are shown) of a structure in the active site of phosphomevalonate kinase. An example structure clustering is shown in the circle.

A)

Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram 5 10 15 20 25 30 35
1	-10.11	44	-9.47	3	###
2	-7.58	10	-7.05	5	#####
3	-7.45	21	-6.93	5	#####
4	-7.32	48	-7.30	3	###
5	-7.27	35	-6.75	3	###
6	-7.12	13	-6.75	2	##
7	-7.10	3	-6.54	3	###
8	-6.96	24	-6.52	5	#####
9	-6.84	27	-6.66	2	##
10	-6.64	5	-6.64	1	#
11	-6.52	45	-6.44	2	##
12	-6.52	34	-6.27	2	##
13	-6.46	25	-6.46	1	#
14	-6.41	17	-6.41	1	#
15	-6.39	4	-6.39	1	#
16	-6.26	41	-6.07	3	###
17	-6.15	50	-5.95	2	##
18	-5.98	28	-5.98	1	#
19	-5.98	6	-5.98	1	#
20	-5.98	23	-5.98	1	#
21	-5.97	16	-5.97	1	#
22	-5.66	46	-5.66	1	#
23	-5.45	8	-5.45	1	#

B)

Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram 5 10 15 20 25 30 35
1	-9.73	26	-9.26	27	##########################
2	-9.44	9	-9.44	1	#
3	-9.06	19	-9.06	1	#
4	-8.94	22	-8.52	5	#####
5	-8.73	23	-8.73	1	#
6	-8.38	40	-7.91	5	#####
7	-8.36	6	-8.36	1	#
8	-8.29	11	-8.03	3	###
9	-7.67	14	-7.67	1	#
10	-7.56	48	-7.28	2	##
11	-7.46	39	-7.46	1	#
12	-7.27	30	-7.27	1	#
13	-7.24	50	-7.24	1	#

Figure 1.5. Clustering histograms from docking program AutoDock used to detail similarites between docked poses. Panel A is an example of poor clustering and B is an example of good clustering.

1.1.5 Protein Binding Analysis using Chemical Shift Perturbations

In NMR analysis, a change of chemical environment will cause the chemical shift (any nuclei used to detect NMR signal) of some compound or molecule to change or perturb, either upfield or downfield depending on the nature of the new environment. For example, when a protein binds a ligand, those residues that are involved in binding (or are causing a structural change in the protein such as domain movement) are perturbed as shown in Figure 1.6.

These changes in chemical shift can be monitored as a function of ligand concentration to give information about ligand binding, such as the affinity of the ligand (i.e. inhibitor or substrate) for the protein. This technique is also useful for screening of compounds that are thought to inhibit a certain protein. Multiple compounds can be screened at a time to determine if any of the compounds bind to the protein, by monitoring any chemical shift changes that arise in the ^{1}H-^{15}N HSQC. NMR-based titrations are usually complemented with a secondary assay for binding, to detect protein inhibition (or activation). For example, an enzyme kinetic assay can be executed utilizing fluorescence or UV spectroscopy, for detection of enzyme reaction at various inhibitor concentrations.

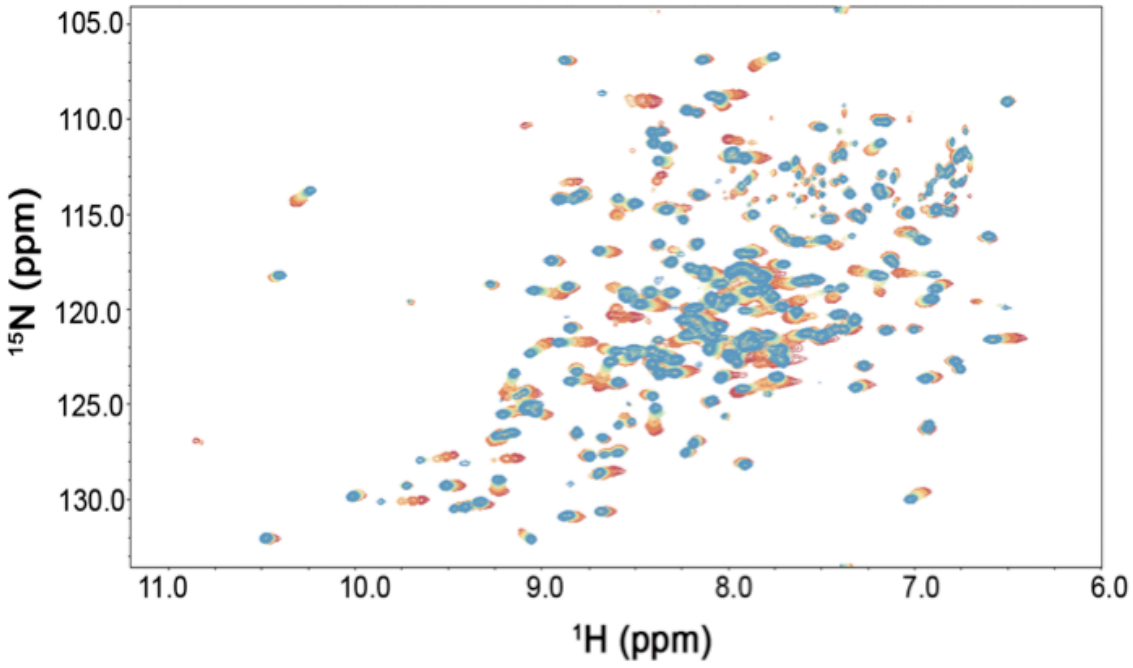


Figure 1.6. Overlay of ^{1}H-^{15}N HSQC titration experiments where Mg-ATP is being titrated onto phosphomevalonate kinase. Chemical shift changes are shown with the color change from red (free PMK) to blue (Mg-ATP saturated PMK). Note that the large number of chemical shift changes is an indication of domain reorientation to accept the new ligand, since the number of cross peaks (amino acid residues) that are perturbed is greater than those in the active site.

CHAPTER II. Substrate Induced Structural and Dynamics Changes in Human Phosphomevalonate Kinase and Implications for Mechanism

2.1 Introduction

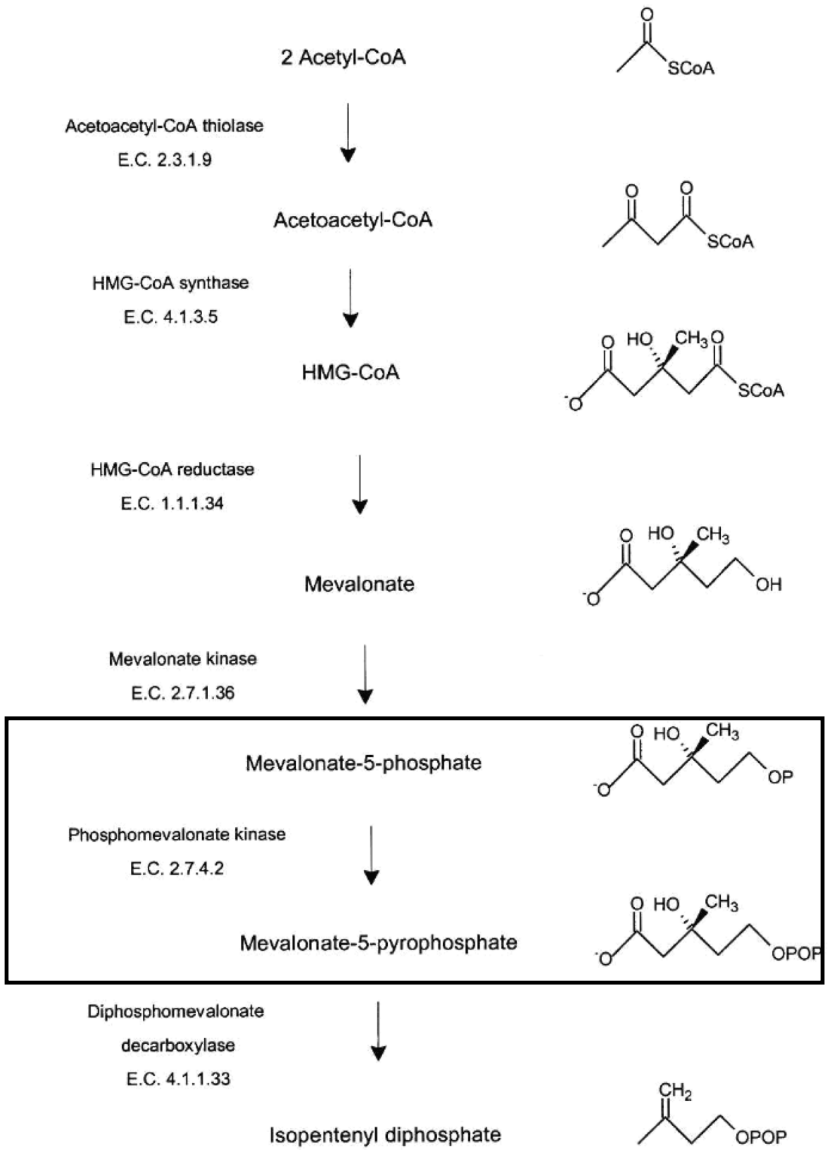
2.1.1 Phosphomevalonate Kinase and Rationale for Dynamics Study

Protein motions are critical for various biological functions and occur over a wide range of timescales. There is a strong correlation between structural dynamics and molecular function, and studying these dynamic processes can give site-specific information on motions that span timescales from psec to msec[4]. One class of enzymes that undergoes large structural and associated motional changes is the kinases. These changes occur due to the binding of their two ligands (substrates) and the need for reactive groups to move into close proximity for catalysis. A model kinase in this regard is adenylate kinase, which consists of two domains that adjust to bring their phosphate donor and acceptor ligands into proximity for reaction to occur[13].

To better understand the molecular mechanism of enzymatic catalysis, structural characterizations of protein-ligand complexes are required. Of particular interest are conformational and dynamic changes that are needed to bring the two highly charged substrates of PMK in proximity to react. To facilitate these studies (at the time of the study there was no crystal structure), a homology model for human phosphomevalonate kinase (PMK) had been generated using the PHYRE homology model method (≈24% overall sequence identity between PMK model and template), because there is no experimental structure yet available for this protein[14]. Our recent NMR chemical shift assignments and subsequent secondary structure analysis has largely validated this model, based on chemical shift index values that match helical and sheet regions in the model. But, the presence of disordered regions and loops has made it impossible to pursue a complete structure determination. Currently, 77% of the backbone atoms have

been assigned, excluding two unassigned gaps that may be disordered regions (residues 56–72 and 99–111). As such, studies herein refer to the homology model. This model consists of two domains. The larger ATP (adenosine triphosphate)-binding domain is comprised of a five-stranded parallel β-sheet interweaved with three α-helices. The smaller M5P (mevalonate 5-phosphate) domain is comprised of loop regions and two α-helices. In the model, two hinge regions, analogous to adenylate kinase, tether the ATP and M5P domains together. These may be involved in opening and closing motions to permit binding and release of substrates[15]. Also, recent site-directed mutagenesis studies have identified a ‘‘Walker A’’ ATP binding motif (K17, R18, K19, and K22)[16.] Mutagenesis of some of these residues can decrease catalysis up to 10,000-fold[14, 17], suggesting this as the active site location for catalysis of phosphoryl transfer. ‘‘Walker-A’’ loop residues that are most important for catalysis include K22 and R18. Outside the ‘‘Walker-A’’ catalytic loop, R110 is also important[14, 17]. The reaction catalyzed by PMK is shown in scheme 2.1.

To characterize the effects of substrate addition on protein dynamics, we have used NMR relaxation experiments in conjunction with a Modelfree analysis[8] to quantify any changes to the fast timescale motion (psec-nsec) of the protein backbone, especially for catalytically relevant regions. We also report on the conformational effects of substrate addition by comparing chemical shift changes and backbone ^{15}N relaxation data for various complexes of PMK: (a) Apo-PMK, (b) the binary complex with Mg-ADP (adenosine diphosphate), (c) the binary complex with M5P, and (d) the ternary complex with Mg-ADP and M5P.



Scheme 2.1: Mevalonate Pathway with PMK catalyzed reaction shown in black box. P represents phosphate ($-PO_3^{2-}$) and E.C. is "Enzyme Commission."

2.2 Materials and Methods

2.2.1 Docking of M5P and ATP into Human PMK Homology Model

The homology model used in our studies was that which was previously prepared and reported by Herdendorf and Miziorko[14]. The binding orientation of ligands in this PMK (human) homology model was determined using Autodock4[12]. Gasteiger charges and hydrogens were added using AutoDock Tools (ADT). The docking grids were also prepared using ADT, with a grid size of 60 x 60 x 60 Å and a spacing of 0.375 Å. These grids were centered on the ligands (AMP and GMP) of the homology model template, bacteriophage T4 deoxynucleotide kinase[18]. The homology model and template were superimposed on each other, in order to obtain coordinates for the grid box used to dock the ligands. The overlay was done using Sybyl 5.8 (Tripos Inc., St Louis, MO, USA). Default docking parameters were used, except that 50 genetic algorithm runs were used with 2,500,000 as the maximum number of evaluations. The docking of both ligands was done separately, due to inter-ligand repulsion from the negative charges on M5P and the ATP triphosphate, even after neutralization of two charges (as would be the case with bound Mg^{2+}). This repulsive interaction was recognized based on earlier docking attempts to form the ternary complex, which produced erroneously distorted structures with the phosphate groups on the two ligands as far apart as possible in the docking box. This led us to do independent docking of ATP and M5P to produce the separate binary complexes, and these pdb coordinate files were then merged to create the ternary complex. This approach positioned the phosphate groups on ATP and M5P in close proximity, as required for the phosphate transfer reaction, and therefore provides strong validation for the docking poses that were generated.

2.2.2 Protein Expression and Purification

PMK is normally a 192 residue 22.0 kDa protein, although it was expressed with an N-terminal histidine tag and additional linker residues, giving a total molecular weight of 24.2 kDa. Uniformly ^{15}N labeled PMK samples were prepared using $^{15}NH_4Cl$ as the sole source of nitrogen in minimal media and protein was expressed and purified as described previously[14, 17]. Briefly, *E. coli* BL21-(DE3) Rosetta cells were transformed with a pET15b(+) expression construct which encoded human PMK with an N-terminal His_6 affinity tag. The transformed cells were plated onto LB (Luria Bertani) agar containing ampicillin (amp) and chloramphenicol (chl) antibiotics. Plates were incubated overnight at 37 °C, and a single colony was picked to inoculate 2 mL of media, and grown to $A_{600} \sim 0.3$. This culture was then used to inoculate 20 LB-amp-chl plates. The plates were incubated overnight at 37 °C, and resulting bacterial lawns were used to inoculate 500 mL of LB-amp-chl to give $A_{600} \sim 1.0$. The liquid culture was then incubated at 30°C for 1 h prior to induction with 1 mM IPTG. The culture was harvested 4 h post induction at $A_{600} \sim 2.0$. Bacterial pellets were resuspended in 100 mL of a 50 mM KP_i (potassium phosphate) buffer containing 100 mM KCl, 5 mM imidazole, and 0.5 mM DTT at pH 7.8. Lysis was accomplished by passage through a microfluidizer at ~17 kpsi. The lysate was clarified by centrifugation at ~100,000 g and the supernatant was loaded onto ~0.5-1.0 mL of Ni-Sepharose Fast Flow resin. The column was washed with lysis buffer until $A_{280} < 0.005$, and the protein was eluted with lysis buffer supplemented with 300 mM imidazole. The fractions containing PMK were pooled and the concentration was determined spectrophotometrically using an extinction coefficient of $\epsilon_{280} = 32{,}290\ M^{-1}\ cm^{-1}$.

2.2.3 NMR Sample Preparation

All protein samples were buffer exchanged, using ultrafiltration with an Amicon (YM10) membrane, into 5 mM DTT, 20 mM KH_2PO_4, 10% D_2O, and 0.02% NaN_2, and concentrated to 400-600 μM. ADP was complexed with Mg^{2+} by adding a 1:1 ratio of $MgCl_2$. Mg-ADP and M5P were both concentrated to 50 mM prior to titration, with pH's of 5.0 for ADP and 7.0 for M5P. All experiments were performed at pH 6.5.

2.2.4 NMR Spectroscopy

All NMR experiments were performed on a 600 MHz Varian NMR System at 599.515 MHz using a triple resonance cryoprobe with z-axis gradients at 25 °C. Titrations were performed using 100 μM increments for both ligands, until saturation was achieved based on ^{1}H-^{15}N HSQC chemical shift changes (1 mM for M5P and 2 mM for MgADP). There were two sets of titrations performed, first starting with M5P then adding Mg-ADP. To assess binding order and synergy, the second titration started with Mg-ADP followed by M5P. ^{15}N relaxation experiments were performed on free PMK, PMK saturated with M5P (2 mM), PMK saturated with Mg-ADP (20 mM; higher concentration was needed due to slow hydrolysis of ADP), and then the ternary complex (using the same saturation concentrations). ^{15}N-T_1, ^{15}N-T_2, and $\{^{1}H\}$-^{15}N NOE experiments were all performed using the BioPack pulse sequences from Varian, Inc (Palo Alto, CA). Delay times for the T_1 experiment were 10.8, 108.3, 215.8, 379.2, 541.7, 758.4, 1083.4, 1515.8, and 2165.8 ms and 4.31, 8.62, 12.9, 17.2, 21.6, 30.2, 38.8, 47.4, 55.1 ms for T_2. T_1 experiments employed the standard inversion-recovery pulse sequence[19], while T_2 experiments employed the CPMG (Carr-Purcell-Meiboom-Gill) sequence, as implemented previously[20, 21]. NOE's were obtained by measuring HSQC

spectra with and without ^{1}H saturation for a time of 3 s (same for both), with an interscan delay of 1 sec. An interleaved approach was used for the T_1 and T_2 experiments, to average out any potential instability over time[22, 23].

2.2.5 NMR Spectroscopy Data Analysis

NMR data were processed using NMRPipe/NMRDraw[24], and analyzed using NMRView.[25] For all experiments, ^{1}H-^{15}N spectra were processed using a 90° shifted sine function in the ^{1}H and ^{15}N dimensions. For ^{15}N-T_1 and ^{15}N-T_2 experiments, the spectra with the shortest relaxation delay were peak picked using NMRView. For $\{^{1}H\}$-^{15}N NOE measurements only one spectrum was peak picked, and for each subsequent spectrum the peak ellipses were manually adjusted to fit each peak.

The R_1 and R_2 relaxation rates were determined by fitting the T_1 and T_2 curves to Equation 1:

$$I_t = I_o * e^{(-Rt)} \qquad \textbf{Eq.2.1}$$

where I_t is the peak intensity after time t, I_o is the intensity at time t=0, and R is either R_1 or R_2. Fitting was done using the Rate Analysis package included in NMRView. NOE values were obtained by taking the ratio of the intensity versus the control. This was done with two sets of experiments in order to obtain an error for the analysis. T_1 and T_2 values were measured for PMK at various concentrations and used to calculate τ_m values, to demonstrate that PMK remains monomeric under the conditions of our NMR studies (previous analytical gel filtration studies also established that PMK is monomeric).[14]

Dissociation constants were obtained by measuring the chemical shift changes in going from free to various bound states, monitoring peaks in fast exchange in both ^{1}H and ^{15}N dimensions. These changes were then combined using Equation 2:

$$\Delta shift_{obs} = [(^{1}H \text{ shift})^2 + (^{15}N \text{ shift}/5.51)^2]^{(1/2)} \qquad \textbf{Eq. 2.2}$$

The chemical shift change ($\Delta shift_{obs}$) as a function of the concentration of the ligand, at fixed concentration [P_o] were fitted to a quadratic equation, to determine the dissociation constant K_d:

$$\Delta shift_{obs} = (\Delta shift_{max}/(2[P_o])) * \{([L_o] + [P_o] + K_d) - \{([L_o] + [P_o] + K_d)^2 - 4[L_o][P_o]\}^{1/2}\} \qquad \textbf{Eq. 2.3}$$

using Sigmaplot8, where L_o is the total ligand concentration at a particular titration point, P_o is the total protein concentration, and $\Delta shift_{max}$ is the maximum chemical shift change observed for the particular peak in question. Fitting to the quadratic equation was required because [P_o] > K_d. Standard deviations resulted from the non-linear least squares fitting process (note: as K_d gets very small relative to [P_o], errors will become larger because K_d become less well-defined in the fitting process).

2.2.6 Modelfree Analysis

The parameters of internal motion were determined from the NMR relaxation data according to the model-free formalism established by Lipari and Szabo[5-7] using Modelfree4 software (version 4.20, A. G. Palmer, Columbia University). The residues

were optimized with an isotropic diffusion model using an initial estimate derived from the R2R1_diffusion program (A. G. Palmer, Columbia University). Backbone dynamics calculations were performed with 300 Monte Carlo simulations per run using an internuclear distance r_{NH} of 1.02 Å and a chemical shift anisotropy (CSA) for the ^{15}N nucleus of -172 ppm. Five models were used to fit our experimental data and were iteratively tested in order of increasing complexity (M1= S^2; M2= S^2, τ_e; M3= S^2, R_{ex}; M4= S^2, R_{ex}, τ_e; M5= S^2_f, S^2_s, τ_e) until an acceptable fit was achieved. S^2 is the generalized order parameter, τ_e is the internal correlation time, R_{ex} is the exchange contribution term, and S^2_f and S^2_s are for sub-nanosecond and nanosecond motions respectively. These models were tested until they could reproduce the experimental relaxation data within 90% confidence limits using appropriate χ^2 and F-tests[8], and this was done for each crosspeak (amino acid).

2.3 Results and Discussion

2.3.1 Docking of M5P and ATP to Human PMK

The docking of M5P and ATP (Fig. 2.1) seems reasonable based on the proximity of the reactive phosphate groups to each other, and also their location within 4.0 Å of the "Walker A" catalytic loop (which is a peptide motif found in proteins known to bind nucleotides, such as ATP), which includes Lys17, Arg18, Lys19, and Lys22 as well as other basic residues, especially Arg110. It is striking that the independent docking of M5P into PMK positioned its phosphate group within 2.7 Å of the ATP γ-phosphate and 2.8 Å from Arg110 and Arg138.

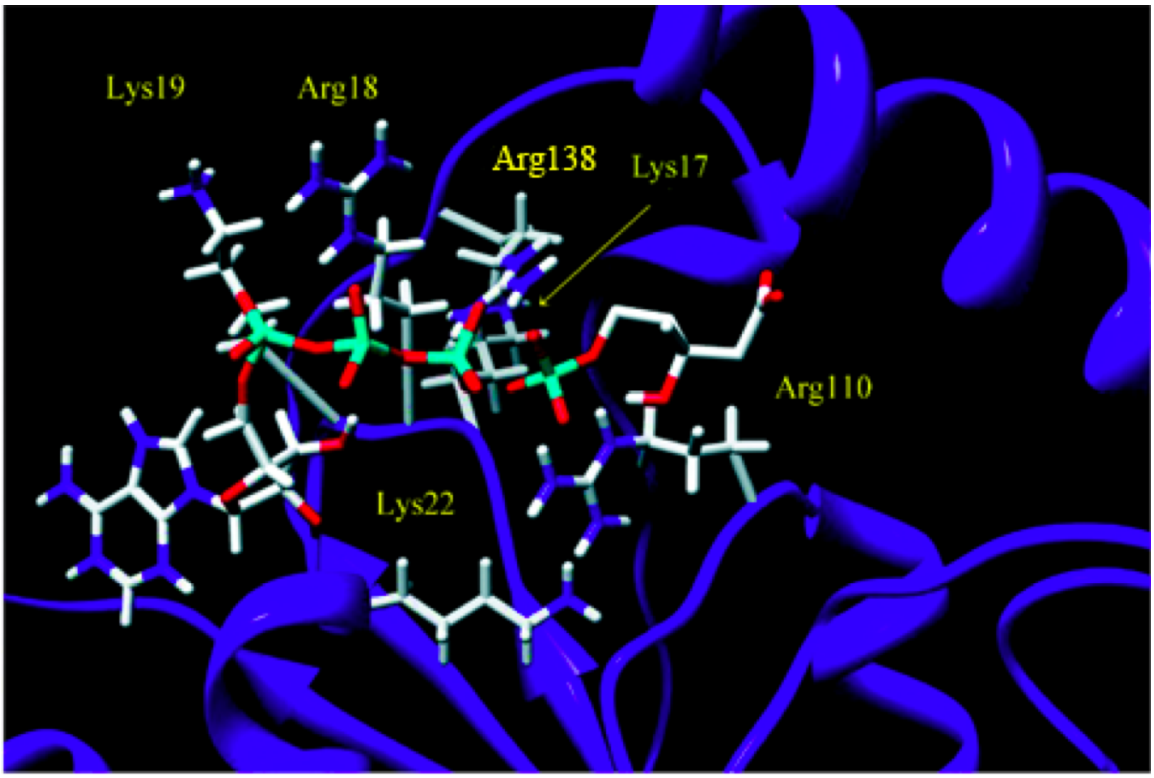


Figure 2.1. Ternary complex structure of docked ligands ADP and M5P in the active site, near the "Walker A" loop, of PMK. Catalytically important amino acids are labeled, including Lys's 17, 19, and 22, as well as Arg's 18, 138, and 110.

In our docked complex of human PMK with M5P and ATP (Fig. 2.1), the key catalytic residues identified by Herdendorf and Miziorko[14, 17] (Arg18, Lys22, and Arg110) are all reasonably well positioned for catalysis. The only possible exception is Lys22, which is in the active site but is pointed away from the phosphate; while this may reflect an inaccuracy of the homology model, it is noteworthy that a simple side chain rearrangement could easily position its ε-amino group near the ATP γ-phosphate. Arg18, Lys22, and Arg110, along with the other basic residues in Fig. 2.1, may be important for forming the ternary complex by neutralizing the negative charges on the two substrates,

so that their phosphate groups can move into proximity for phosphoryl transfer. Additional charge stabilization (from Lys22?) may also be needed to stabilize the transition state if it goes by an associative-type mechanism, which would place added charge density on the γ-phosphate. An important feature of the model in Fig. 2.1 is that the γ-phosphate of ATP is ideally positioned for nucleophilic attack by the M5P phosphate (separated by only 2.7 Å). This provides further validation that this structural model of the ternary complex is reasonably accurate.

2.3.2 Chemical Shift Perturbation Studies to Make the Binary and Ternary Complexes

To determine how different ligands affect the residues of the protein, titrations of the ligand to the protein can be performed by monitoring ^{1}H-^{15}N HSQC NMR spectra. The HSQC experiment monitors the magnetization transfer from the proton to the nitrogen on the amide group of the backbone of the protein. The spectra show crosspeaks for all the amide groups in the protein, one for each residue, except for proline. Addition of specific ligands that bind to protein can cause these amide groups to change their chemical shift due to environmental changes, giving rise to chemical shift perturbations. For chemical shift perturbation studies, a dead-end (i.e. non-reactive) complex with PMK was created, to ensure that there would be no catalytic turnover during the NMR measurements. AMP-PNP (Fig. 2.2 and 2.3b) is often used to create such dead-end complexes in kinases, but it binds only weakly to PMK and causes few structural changes in PMK compared to ATP (Fig. 2.3a) or ADP (Fig. 2.3c). As a result, ADP was used to make the dead end complex (Fig. 2.3c) in our studies.

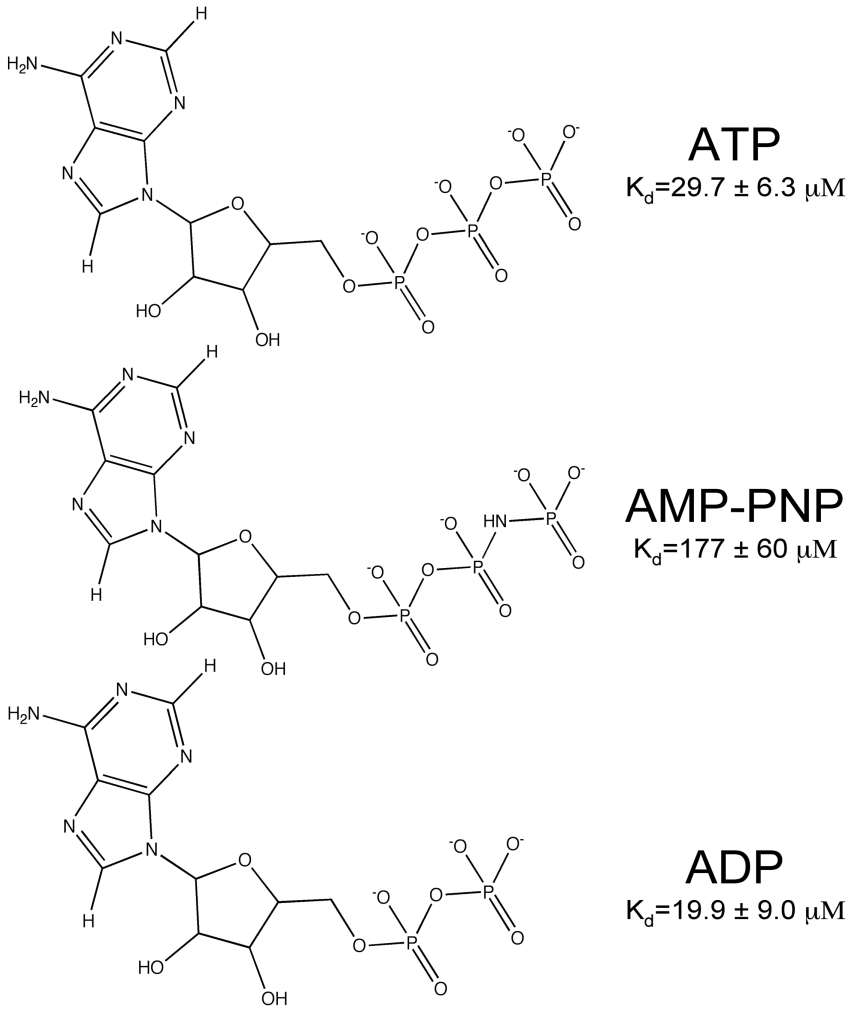


Figure 2.2. Structures of adenine nucleotides ATP, AMP-PNP, and ADP with corresponding K_d values.

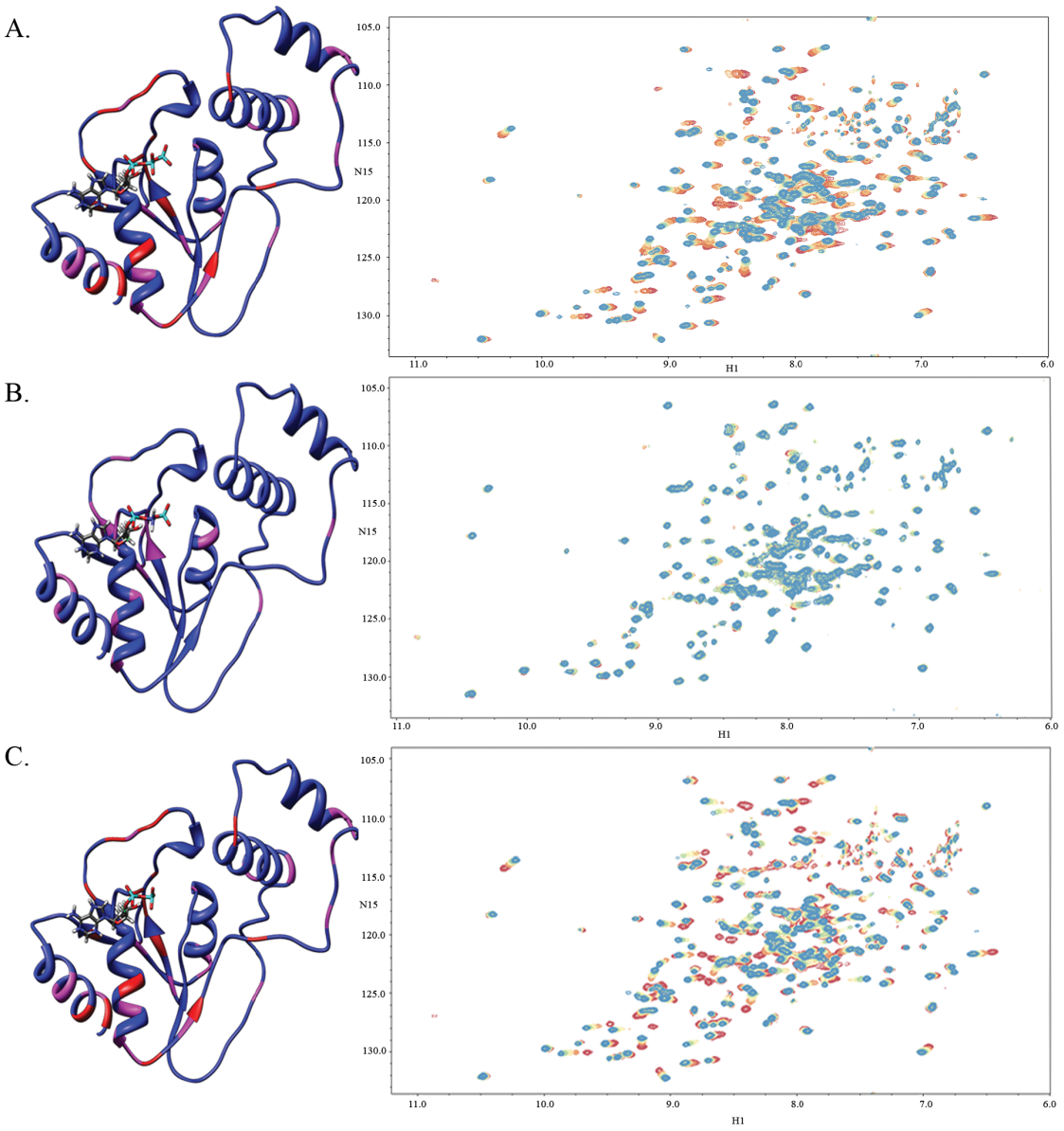


Figure 2.3. Chemical shift perturbations due to binding of nucleotides a) Mg-ATP titrated to saturation onto Apo-PMK, b) Mg-AMP-PNP titrated to saturation onto Apo-PMK, and c) Mg-ADP titrated to saturation onto Apo-PMK. All experiments begin at the red cross-peaks, with additions in 100 μM increments to saturation, which is indicated by the transition to the blue cross-peaks. The corresponding chemical shift changes are mapped on to the human PMK homology model, where the red indicates a large chemical shift change (>0.09 ppm) and pink indicates a medium chemical shift change (0.05-0.09 ppm), while blue indicates a small or no chemical shift change (or no data).

To justify the use of ADP as a suitable replacement for ATP, an ATP titration experiment was carried out and is shown in Fig. 2.3a. Comparing these spectra, it is clear that they show the same chemical shift changes for the same residues, which infers that the two protein-ligand complexes undergo very similar structural changes. From the HSQC spectra for the Mg-ADP titration (Fig. 2.3c), a pronounced conformational change is observed upon substrate addition, due to multiple residues that are affected when the complex is formed. Fig. 2.4a shows the chemical shift perturbations due to Mg-ADP binding mapped onto the PMK model, and is consistent with Mg-ADP causing the protein to undergo a gross conformational change; that is, the whole protein appears to adjust structurally in order to accommodate the addition of Mg-ADP. Notable changes are observed in the loops near the substrate phosphate groups, as well as the region where the two ATP-domain helices contact each other. The Mg-ADP addition was then followed by further titration with M5P to form the ternary complex (Fig. 2.4b). To determine if there is a preferred sequence for binding, due to synergy, M5P was then titrated first to Apo-PMK (Fig. 2.4c).

Based on the chemical shift perturbations that occurred during titration to form binary and ternary complexes, it can be seen that the largest changes occur in forming the binary complex, whether it is ADP or M5P that binds first (Fig. 2.4). The chemical shift changes include residues outside the binding site, such as hinge regions between domains, suggesting that a conformational change has occurred.

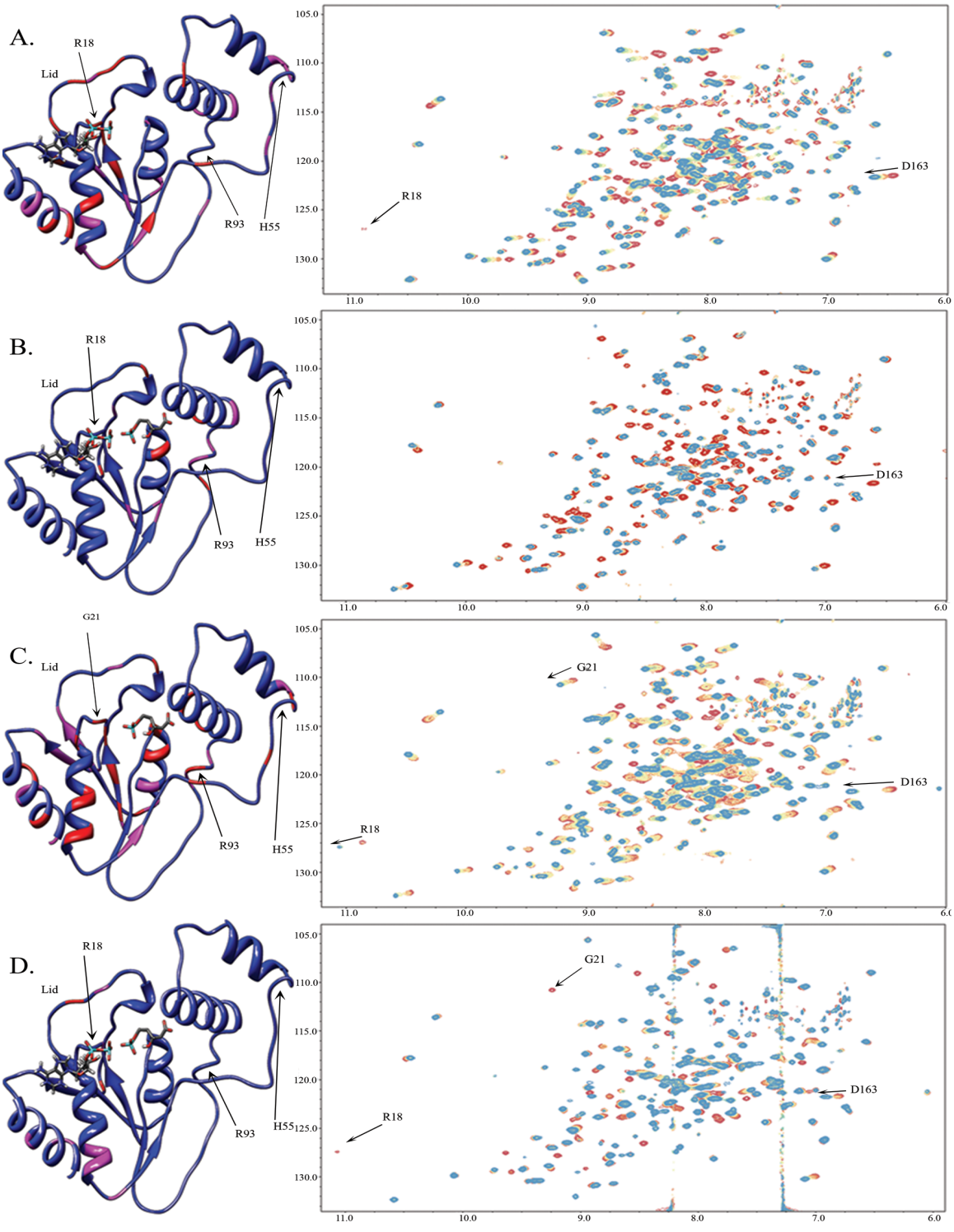


Figure 2.4. Chemical shift perturbations due to: a) Mg-ADP added to saturation to Apo-PMK b) M5P added to saturation to Mg-ADP/PMK c) M5P added to saturation to Apo-PMK and d) Mg-ADP added to saturation to PMK/M5P. All experiments begin with the red cross-peaks and titrations occur at 100 μM increments to saturation, which is indicated by transition to the blue cross-peaks. The corresponding chemical shift changes are mapped on to the human PMK homology model, where the red indicates a large chemical shift change (>0.09 ppm) and pink indicates a medium chemical shift change (0.05-0.09 ppm) while blue indicates small or no chemical shift change (or no data). Note: Gly21 is visible in panel (a) at a lower threshold.

In terms of adenine nucleotide binding, it is interesting that while the γ-phosphate of ATP seems to not be important for binding or inducing structural changes, based on comparisons of ATP and ADP (Figs. 2.2 and 2.3), the situation is actually more complicated - because simple O=>NH substitution of the bridging heteroatom between the β and γ phosphates has a dramatic effect on binding. This suggests there may be important interactions with the β-γ bridging heteroatom, which might be important for stabilizing this leaving group during phosphate transfer. It is also noteworthy that ADP (and ATP) binding causes chemical shift perturbations in the Walker A loop (ex. Arg18) and the lid that covers the Walker A loop, as well as the hinge residue for the lid (D163). Opening of this lid should only produce a modest change in size of the protein, and so would not be expected to produce a change in correlation time. In contrast, domain movement between ATP and M5P binding domains would be expected to produce a much larger change in size and correlation time, as had been observed in adenylate kinase, and appears to also occur in human PMK (*vide infra*).

2.3.3 Binding Affinity and Synergy

To quantitatively assess binding affinity and binding synergy, if any, cross-peak perturbations (Figs. 2.3 and 2.4) were fitted to obtain K_d values for each binding event. Monitoring chemical shift changes and fitting to the quadratic equation (Eq. 2.3) permitted determination of dissociation constants (K_d) for the various nucleotides (Fig. 2.2) and for all four complexes (Fig. 2.5). The K_d for the Mg-ADP titration to Apo-PMK is 19 ± 9 μM and <20 μM for subsequent M5P binding to form the ternary complex. For the titration of M5P to Apo-PMK the K_d is 6 ± 3 μM and 56 ± 16 μM for the subsequent

Mg-ADP addition to form the ternary complex. These values are consistent with the previously reported K_m values of 47 ± 5 μM for ADP and 34 ± 3 μM for M5P,[26] keeping in mind that K_m values can often deviate from K_d values due to kinetic effects.

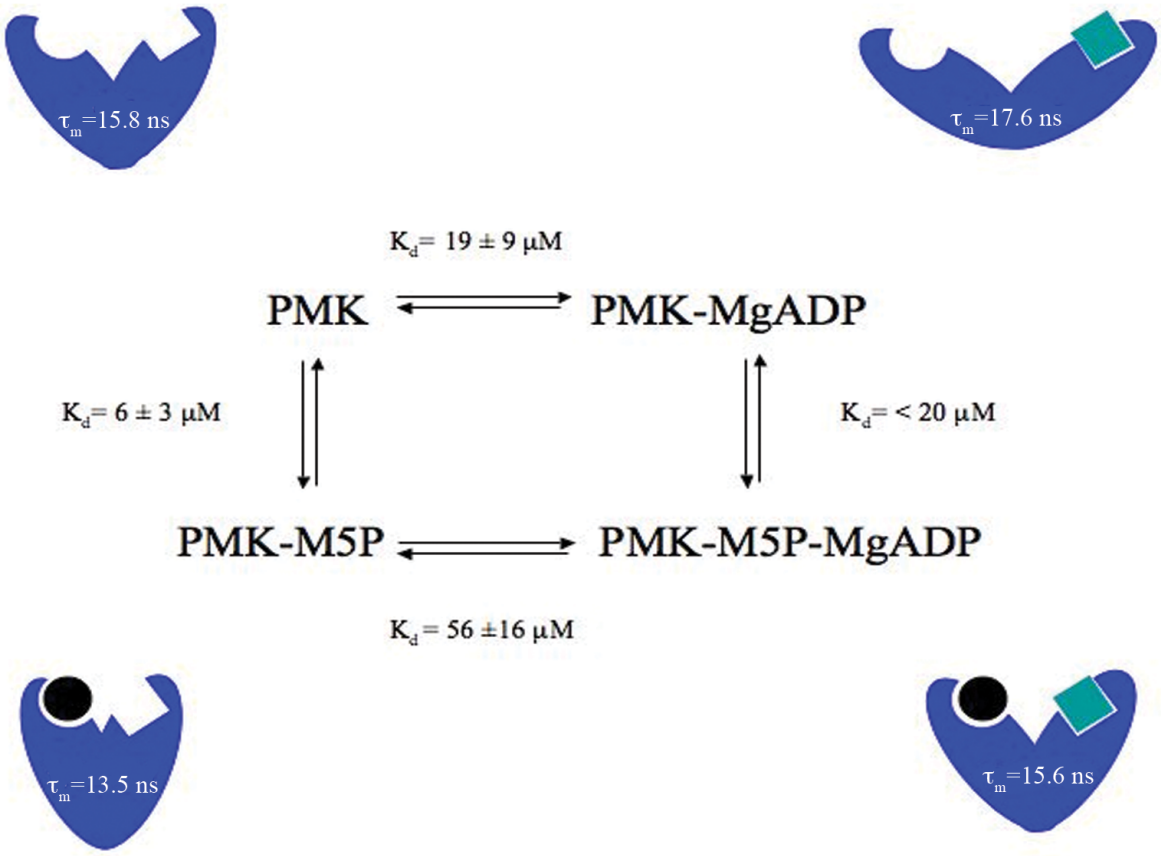


Figure 2.5. Thermodynamic box for the formation of the dead-end ternary complex of human PMK with corresponding K_d and τ_m values for each complex. Cartoon representations of each complex are shown, with the circles representing M5P and the squares representing ADP.

Based on chemical shift perturbations, it appears that binding of either M5P or ADP to apo PMK can induce a large conformational change. K_d values obtained from

fitting spectral changes are consistent with their being no preferred binding sequence or synergy for the ligands. This is in contrast to the report of an ordered sequential mechanism (M5P binding first and ADP released last) for the pig liver PMK enzyme,[26] but is consistent with recently reported steady state kinetics on human PMK, where competitive inhibition was observed using mevalonate 5-diphosphate as a product inhibitor versus M5P.[17] If anything, our data suggest there is a modest (~3-fold) anti-synergy, such that affinity for the second ligand is weakened due to presence of the first, which could be due to the charge/charge repulsion by the phosphate groups on both ligands. But, this anti-synergy could be unique to our dead-end complex (MgADP/M5P), and does not necessarily reflect the situation for the catalytically competent Michaelis complex (MgATP/M5P).[27]

2.3.4 Relaxation Dynamics

To obtain dynamic information on the psec-nsec time scale motions, longitudinal (R_1) and transverse (R_2) relaxation rates as well as $\{^1H\}$-^{15}N NOE values were obtained, and are summarized in Fig. 2.6 for Apo-PMK, Mg-ADP bound PMK, M5P bound PMK, and the ternary (Mg-ADP/M5P) complex. All of the complexes have relatively rigid structures (high S^2), indicative of a well-ordered protein backbone on the psec-nsec timescale (Fig. 2.6). Trends in R_1 and R_2 values in Table 2.1 can best be explained based on changes in the overall correlation time of the protein in the different complexes. It should again be noted that PMK remains in a monomeric state in our NMR studies (Fig. 2.7), so changes in correlation time are not associated with aggregation effects.

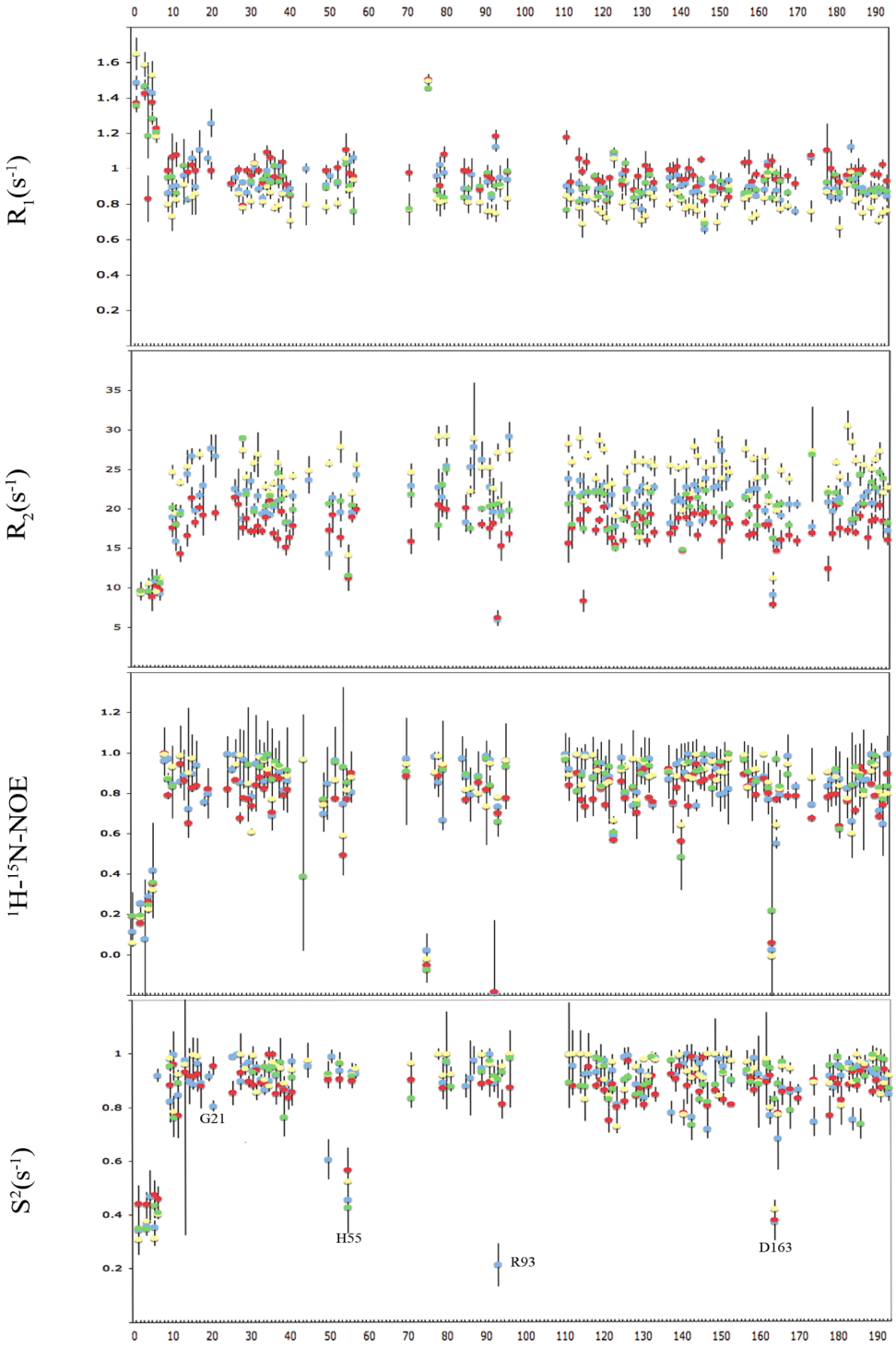


Figure 2.6. NMR dynamics data including R_1, R_2, NOE, and S^2 values, where blue is the Apo-PMK complex, red is the M5P saturated complex, yellow is the Mg-ADP saturated complex, and green is the ternary complex.

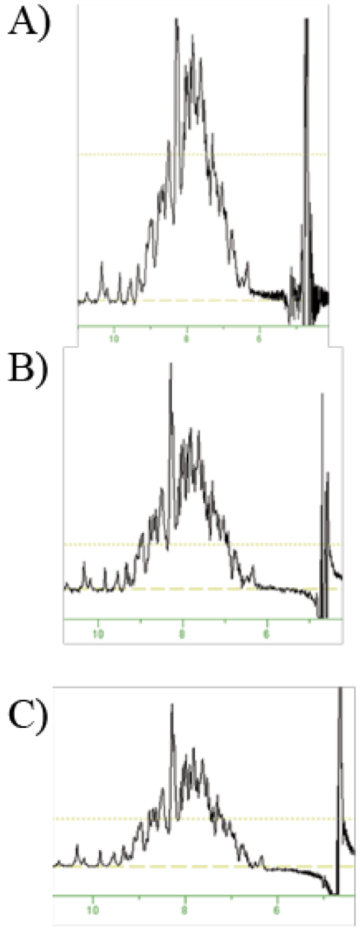


Figure 2.7: Representative 1-D spectra from T_2 relaxation measurement at A) 500 μM, B) 250 μM, and C) 125 μM PMK. The rotational correlation time τ_m for A) is 12.9 ± 1.2 ns, B) is 13.5 ± 0.9 ns, and C) is 12.0 ± 0.6 ns.

	Apo-PMK	PMK-M5P	PMK-ADP	PMK-ADP-M5P
R_1 (s^{-1})	0.945 ± 0.034	0.994 ± 0.046	0.856 ± 0.051	0.933 ± 0.036
R_2 (s^{-1})	21.9 ± 1.9	17.8 ± 1.9	25.4 ± 2.0	20.8 ± 1.7
NOE	0.820 ± 0.080	0.766 ± 0.028	0.834 ± 0.065	0.82 ± 0.11
S^2	0.881 ± 0.035	0.862 ± 0.040	0.898 ± 0.048	0.897 ± 0.036

Table 2.1. Summarization of relaxation dynamics parameters of apo and ligand induced states of PMK.

The changes in PMK correlation time were verified and quantified in the Modelfree analysis, where it was found that relaxation data could not be properly fitted for the different PMK complexes using the same correlation time. The correlation times obtained from the Modelfree analysis of each complex are summarized in Fig. 2.5. These correlation time changes are consistent with there being a ligand-induced change in overall shape of PMK in the different complexes, as had been previously observed in analogous NMR studies of adenylate kinase.

Based on Modelfree analysis of NMR dynamics parameters (Fig. 2.6), generalized order parameters were obtained. These clearly indicate that PMK adopts a fairly rigid structure (S^2 values >0.8) for all the binary and ternary complexes. Furthermore, the average S^2 values remain largely unchanged upon formation of the binary complex with either ADP or M5P, indicating that the local dynamics for all the complexes are virtually the same (Table 2.1) on the psec-nsec timescale. The large S^2 values, which are consistent with a rigid, highly ordered backbone, are not unexpected given the well-structured core of the protein, which includes a 5 stranded parallel β-sheet. One might have expected some ordering in the M5P domain upon ligand binding, given the high content of extended and loop regions, but this does not appear to be occurring.

2.3.5 Inter-domain and loop motion: ligand binding domains, lid and "Walker A" catalytic loop

Although the protein core may be largely rigid, it is possible that there is some motion of domains relative to each other, and of other defined regions. Indeed, chemical shift perturbation studies suggested there was a large substrate-induced conformational

change (Fig. 2.4). In fact, there were a small number of residues with small or negative NOE's that gave rise to small S^2 values (H55, R93, and D163), and these are located in loop and potential hinge regions in the homology model (*vide infra*). Interestingly, the "Walker A" ATP binding loop shows the same S^2 values as the rest of the protein (i.e. it is rigid), indicating that if there is any catalytically relevant change in dynamics, it might only occur during turnover with the actual Michaelis complex,[27] rather than with the "dead-end" complex used in our studies. One important exception is Gly21, which is adjacent to Lys22, the catalytically most important residue in PMK.[14] Gly21 undergoes an increase in S^2 in going from the apo (blue) to the M5P (red) complex, indicating it becomes less mobile in the binary complex (Fig. 2.6), perhaps as Lys22 moves into position for catalysis (note: Lys22 was not assigned, so changes to its dynamic state are not known). Interestingly, Gly21 exchange broadens in the ternary complex, suggesting it undergoes msec timescale motion. But, in the conversion of apo PMK to the PMK-M5P binary complex, one can see a slow exchange process for Arg18, and a fast exchange process for Gly21 (Fig. 2.4c, and expansions in Fig. 2.8).

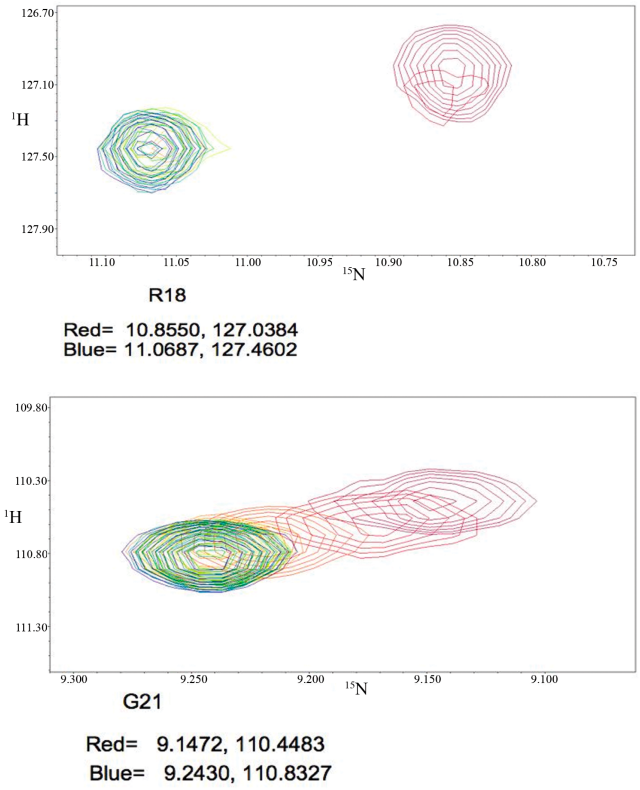


Figure 2.8: Expansions of HSQC spectra, showing crosspeak changes used to calculate exchange rates. R18 is in slow exchange between two states, while G21 is in fast exchange between two states. Crosspeak seperation, in sec^{-1}, therefore defines the exchange rate between these two states, for the loop containing R18 and G21.

Assuming the same motion occurs for these two "Walker A" loop residues, this timescale flanking permits an estimation of the rate constant for the "Walker A" catalytic loop exchanging between the two states associated with M5P binding as k_{ex} = 100-600 sec^{-1} (see Fig. 2.8). This is likely to be due to a lid opening/closing motion, which would be coupled to "Walker A" loop motion, based on analogy to adenylate kinase where a flexible lid caps the catalytic "Walker A" loop. [28-31]

It should be noted that while there doesn't appear to be significant motion of "Walker A" catalytic loop residues on the psec-nsec timescale (besides Gly21), there

does appear to be motion on slower (ex. msec) timescales. For example, in the ADP binary and ternary complexes, most crosspeaks for residues in the loop were no longer visible, due to exchange broadening (Fig. 2.4). For this reason, S^2 for ADP complexes could not be measured. Accordingly, it is simply not known in these complexes what motion might be occurring on the psec-nsec timescale.

Overall, the psec-nsec timescale dynamics data indicate that M5P binding drives the immobilization of catalytic residues (with Gly21 as our "Walker A" reporter), but there still appears to be msec timescale motion, especially when ADP is bound. This motion might be important for catalysis, but could also be unique to the particular dead-end inhibitory complex used in our study. There is certainly literature precedent for kinase active site loop dynamics being very sensitive to the nature of the inhibitory complex. For example, residual dipolar coupling and ^{15}N relaxation measurements of various ABL kinase complexes indicated that inhibitors differed significantly in their ability to immobilize the catalytic/activation loop.[32]

2.3.6 Ligand-induced changes to protein size: inter-domain conformational change?

The chemical shift perturbation studies indicated that binding of either ADP or M5P may cause a large conformational change (Fig. 2.4), but the above analysis of S^2 values indicated this did not correspond to any significant changes in dynamics within the two ligand binding domains, other than some changes to the "Walker A" catalytic loop, and the lid that covers it. What about inter-domain motion? The Modelfree analysis provided some insights in this regard. As mentioned in the Methods section, relaxation

data fitting was optimized for each complex to find the rotational correlation time τ_m for that complex, since it was not possible to fit all complexes using a single τ_m. Changes to fitted τ_m values for all the complexes (Fig. 2.5) reveal that the ADP binary complex has the largest τ_m of 17.6 nsec while the M5P binary complex has the smallest, with a τ_m of 13.5 nsec. The Apo-PMK and the ternary complex fall in between these two extremes, with similar τ_m values of ~15.7 nsec (Fig. 2.5). This suggests that both ADP and M5P binding induce conformational changes, but in opposite directions. That is, M5P causes a compression of the structure (so PMK tumbles faster), while ADP causes an opening up of the structure (so PMK tumbles slower). Such changes would be consistent with a movement of the two ligand binding domains relative to each other, as occurs in the nucleoside monophosphate kinases. [28-3133]

Why ADP binding causes the structure to open up is not clear. Since the product of the PMK reaction is ADP, it could be that the open conformation reflects the protein trying to release ADP. However, the chemical shift perturbation study with ATP (Fig. 2.3a) shows the same chemical shift perturbations as with ADP, arguing against there being a dramatically different conformation in the ATP vs. ADP binary complexes. The changes in τ_m for the various complexes indicates that protein closure occurs upon binding to M5P, with or without ADP present. This tightening/closure of the structure induced by M5P binding might be associated with domain movement around a His55/Arg93 hinge (Figs. 2.4 and 2.6). The chemical shift perturbation studies of PMK bound to M5P, with subsequent addition of Mg-ADP, demonstrate that the protein may open back up to bind ADP/ATP for catalysis.

2.3.7 Hinge regions for lid and domain movement

The previous discussion presented data in support of ligand-induced structural changes, which may be related to inter-domain (ATP/M5P) motion as well as to motion of the "Walker A" catalytic loop and the lid that covers it. Regarding movement of the lid, it is noteworthy that data on adenylate kinase indicates lid movement occurs on the same timescale as catalysis. This is consistent with the lid controlling access to the active site, as well as its packing up against the catalytic "Walker A" loop. In adenylate kinase, a key catalytic residue in the ATP binding site is Lys21, [33] which twists into an unfavorable double gauche rotamer in order to interact with the ATP phosphate, just as Lys22 would in the human PMK homology model (Fig. 2.1) if its sidechain were adjusted to permit interaction with the ATP phosphate. This lysine may cycle into a catalytically useful orientation in the transition state to stabilize the increasing negative charge on the γ-phosphate, which would be created in the associative mechanism that is thought to be operative for nucleoside monophosphate kinases.[33] Such a mechanism would permit selective stabilization of the transition state, and would require motion of the "Walker A" loop as the transition state is approached. So, lid motion is potentially relevant to catalysis. How does lid motion occur in PMK?

As noted earlier, a number of potential hinge residues had unusually small S^2 values (H55, R93, and D163), which indicates they are mobile on the psec-nsec timescale (note: often, fast timescale motion occurs along with msec timescale motion, which would be more relevant for a hinge region). Specifically, residues H55 and R93 are positioned (Fig. 2.4) such that they could operate as hinge residues that allow movement between the core ATP domain and the M5P domain, possibly for enabling an opening

and closing motion upon substrate binding[3]. Likewise, Asp163 is located in a potential hinge region for the "lid" loop region that is analogous to the "lid" that is present in the nucleoside monophosphate kinases. [28-31][33, 34] This lid caps the catalytic "Walker A" loop in the PMK model. The potential hinge residue, Asp163, also underwent dramatic chemical shift changes upon ligand binding (Fig. 2.4), as would be expected if there were motion around this hinge region induced by ligand binding. Motion around this hinge region would permit movement of the "Walker A" catalytic loop, which could be coupled to lid motion, as in adenylate kinase. The full extent to which catalytic loop and lid motions are coupled, and what role this motion might have in catalysis in human PMK, is not known, and will be the topic of future studies. But, extensive NMR studies on nucleoside monophosphate kinases,[33] and especially adenylate kinase as a prototype, [28-31][34] provide a rich literature that describes coupled motions between these loops, and their role in catalysis. These studies have also identified ligand-induced changes in protein correlation times (due to inter-domain movement; opening/closing), as we have now observed in human PMK. It will be interesting to see if human PMK follows the example of the adenylate kinase prototype, or whether it will provide unique and new insights into the role of such loop and domain motions in kinase-mediated phosphate transfer reactions. The studies reported herein provide a foundation for such future studies.

2.4 Conclusion

PMK is a very dynamic protein with large structural changes that occur upon ligand addition, especially the first ligand addition. However, as shown from the Modelfree analysis, we see that the protein conformation changes with different ligands

being added upon M5P binding, the protein goes to a closed conformation (τ_m of 13.5 nsec) and upon ATP/ADP binding, the protein opens up (τ_m of 17.6 nsec). The Modelfree analysis has also identified a few potential hinge regions, specifically R93 and H55, which may be involved in moving of the M5P binding domain to permit closure of the protein. This analysis has also indicated that D163 may be involved in the lid region movement. This lid caps the "Walker A" ATP binding site, like many other monophosphate kinases. Finally the K_d's determined from chemical shift perturbation studies were very similar to the K_m's previously reported[4], and we have determined that all ligands bind with similar dissociation constants: 6 +/- 3 μM for M5P binding to free PMK and 56 +/- 16 uM for subsequent ADP binding; 19 +/- 9 μM for ADP binding to free PMK and <20 μM for subsequent binding of M5P.

CHAPTER III. NMR Dynamics Investigation of Ligand-Induced Changes of Main and Side Chain Arginine N-H's in Human Phosphomevalonate Kinase

3.1 Introduction

3.1.2 Phosphomevalonate Kinase and Arginine Dynamics

Studies of the dynamical properties of proteins using NMR spectroscopy is an emerging field that has largely been limited to backbone N−H or side-chain methyl motions. It is often but not always true that changes to the dynamic state of the backbone reflect changes in the side chains. The NMR techniques for quantifying fast-time-scale motion involve measuring the longitudinal (R_1) and transverse (R_2) relaxation rates as well as the heteronuclear nuclear Overhauser effect (NOE) for each amino acid backbone N−H (or side-chain C−H) bond vector in the protein.[4, 7, 8, 35] These values can then be used to calculate the generalized order parameter (S^2), which is a measure of protein flexibility on the psec to nsec timescale.

We have recently reported on the ligand-induced structural and dynamical changes of human phosphomevalonate kinase (PMK).(4) PMK is the fifth enzyme in the mevalonate pathway in humans and is involved in steroid biosynthesis.[26, 36] PMK catalyzes phosphoryl transfer from adenosine triphosphate (ATP) to mevalonate 5-phosphate (M5P) to form adenosine diphosphate (ADP) and mevalonate 5-diphosphate. To permit phosphoryl transfer, the substrates are brought close together, resulting in a significant and repulsive buildup of negative charge. To facilitate this difficult task, PMK contains 17 arginines (Figure 3.1) and eight lysines, many of them in the active site to help neutralize the negative charge on the phosphates.

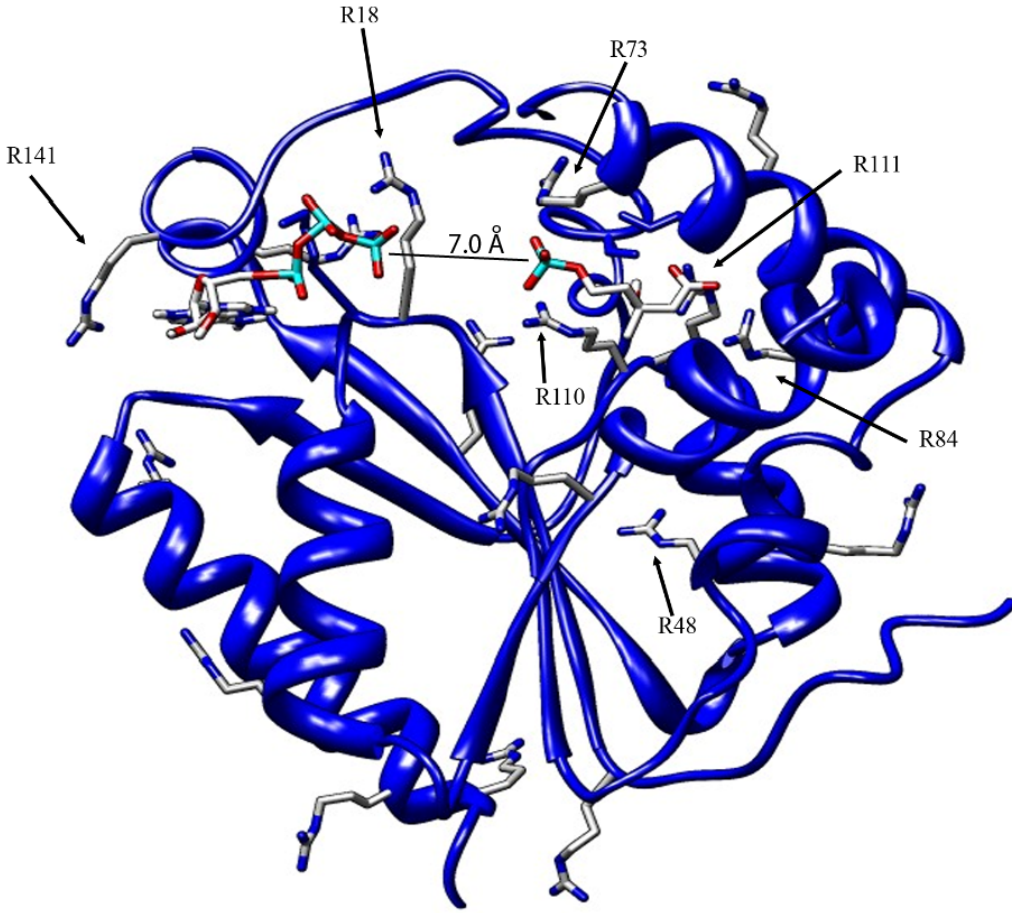


Figure 3.1. Ternary complex of human PMK, with M5P and ATP docked independently into the apo crystal structure (3CH4).[9] After docking, structure was optimized using molecular dynamics (50 ps at 300 K). All arginines are shown, and those known to be important for substrate binding or catalysis are labeled. Inter-phosphate distance is 7 Å, so phosphate transfer will require additional domain movement.

The most important arginines, on the basis of site-directed mutagenesis studies, are R18, R48, R73, R84, R110, R111, and R141.[14, 17] Here we describe the use of NMR dynamics methods to characterize changes in the mobility of arginine side chains upon ligand binding. NMR dynamics methods, which have recently been applied to arginine side chains,[35, 37, 38] allow study of the role that arginine side chains play in ligand binding and catalysis. Our studies provide surprising insights into ligand effects on arginine side

chains that are remote from the active site, perhaps due to long-range Coulombic attractions.

3.2 Methods and Materials

3.2.1 Docking of M5P and ATP into the Human PMK Crystal Structure

The binding orientation of ligands in the PMK (human) crystal structure (3CH4) was calculated using Autodock4.[12] Gasteiger charges and hydrogens were added using AutoDock Tools (ADT). The docking grids were also prepared using ADT, with a grid size of 60 x 60 x 60 Å and a spacing of 0.375 Å. These grids were centered on the ligands (AMP and GMP) of the homologous protein, bacteriophage T4 deoxynucleotide kinase.[18] The PMK crystal structure (apo) and template were superimposed on each other, to determine coordinates for the grid box used to dock ligands. Default docking parameters were used, except 50 genetic algorithm runs were used with 2,500,000 for maximum number of evaluations. The docking of both ligands was done separately, due to inter-ligand repulsion from the negative charges on M5P and the ATP triphosphate, even after neutralization of two ATP charges (as would be the case with bound Mg^{2+}). This repulsive interaction was recognized based on earlier docking attempts to form the ternary complex, which produced distorted structures with the phosphate groups on the two ligands as far apart as possible. This led us to do independent docking of ATP and M5P to produce the separate binary complexes, and these pdb coordinate files were then merged to create the ternary complex. The crystal structure of PMK has the R141 positioned in the proposed binding site of ATP, so that docking of ATP into that site caused the ATP to adopt a highly strained conformation with a steric clash involving

R141. So to achieve a proper docking, with minimal perturbation to the crystal structure, R141 was temporality changed to glycine. This allowed a proper docking of ATP into the active site. Then the R141 sidechain was added back and, as expected, there was still a steric clash with the R141 sidechain and ATP. This merged ATP/PMK pdb file was then minimized using Amber 99, as implemented by HyperChem 7.5 (Gainesville, FL). The minimized structure was then further refined by performing a brief molecular dynamics (MD) run (again using the Amber 99 force field, as implemented in HyperChem 7.5) to determine any lid motion that may be required to permit ATP binding. MD was performed using a 1 fsec step size, with a heating time of 10 ps and a production time at 300K of 50 ps. At the end of the dynamics run, the structure was again minimized to produce the final ATP/PMK complex. During the MD, motion was only permitted for residues 120 to 160, which included the lid that contains R141. This approach to docking, with subsequent MD-based refinement, positioned ATP and M5P in their respective binding sites, is positions that are consistent with previous binding studies. In the end, there were no dramatic changes to the lid. Rather, there was mostly just an adjustment in the orientation of R141 sidechain, with modest lid rearanegement.

3.2.2 Protein Expression and Purification

PMK is normally a 192 residue 22.0 kDa protein, but it was expressed herein with an N-terminal histidine tag and additional linker residues, giving a total molecular weight of 24.2 kDa. Mutated PMK constructs (R84M, R111M, R130M, and R141M) were described previously.[14, 17] Briefly, *E. coli* BL21-(DE3) Rosetta cells were transformed with a pET15b(+) expression construct of wild type and mutant PMK. The transformed

cells were plated onto LB (Luria Bertani) agar plates containing ampicillin (amp) and chloramphenicol (chl). Plates were incubated overnight at 37 °C, and a single colony was picked to inoculate 2 mL of media, which was then grown to A_{600} ~ 0.3. This culture was then used to inoculate 20 LB-amp-chl plates. The plates were incubated overnight at 37 °C, and resulting bacterial lawns were used to inoculate 500 mL of M9 minimal media-amp-chl with $^{14}NH_4Cl$ and ^{13}C-glucose, as the sole sources of nitrogen and carbon, except media was supplemented with .25g/L [$^{13}C^6$, $^{15}N^4$]-L-Arginine (Isotec Inc.), to give A_{600} ~ 1.0. The liquid culture was then incubated at 30°C for 1 h prior to induction with 1 mM IPTG. The culture was harvested 4 h post induction at A_{600} ~2.0. Bacterial pellets were resuspended in 100 mL of a 50 mM KP_i (potassium phosphate) buffer containing 100 mM KCl, 5 mM imidazole, and 0.5 mM DTT at pH 7.8. Lysis was accomplished by passage through a microfluidizer at ~17 kpsi. The lysate was clarified by centrifugation at ~100,000 g and the supernatant was loaded onto ~0.5-1.0 mL of Ni-Sepharose Fast Flow resin. The column was washed with lysis buffer until $A_{280} < 0.005$, and the protein was eluted with lysis buffer (stated above) supplemented with 300 mM imidazole. The fractions containing PMK were pooled and the concentration was determined spectrophotometrically using an extinction coefficient of ϵ_{280} = 32,290 M^{-1} cm^{-1}.

3.2.3 NMR Sample Preparation

All protein samples were buffer exchanged, using ultrafiltration with an Amicon (YM10) membrane, into 5 mM DTT, 20 mM KH_2PO_4, 10% D_2O, 10% d^6-glycerol, and 0.02% NaN_3, and concentrated to 400-600 μM. All experiments were performed at pH 6.5 and 298 K.

3.2.4 NMR Spectroscopy

All NMR experiments were performed on a 600 MHz Varian NMR System at 599.515 MHz using a triple resonance cryoprobe with z-axis gradients at 25 °C. Titrations were performed using 100 μM increments for M5P addition, until saturation was achieved based on ^{1}H-^{15}N HSQC chemical shift changes (1 mM for M5P). ^{15}N-T_1, ^{15}N-T_2, and $\{^{1}H\}$-^{15}N NOE experiments were all performed using the BioPack pulse sequences from Varian, Inc (Palo Alto, CA). Delay times for the T_1 experiments were 10.8, 108.3, 215.8, 379.2, 541.7, 758.4, 1083.4, 1515.8, and 2165.8 ms, and delay times for the T_2 experiments were 4.31, 8.62, 12.9, 17.2, 21.6, 30.2, 38.8, 47.4, 55.1 ms. T_1 experiments employed the standard inversion-recovery pulse sequence,[34] while T_2 experiments employed the CPMG sequence, as implemented previously.[32, 33] NOE's were obtained by measuring HSQC spectra with and without ^{1}H saturation for a time of 3 s (same for both), with an inter-scan delay of 1 sec. An interleaved approach was used for the T_1 and T_2 measurements, to decrease effects of field instability over long experiment times.[21]

The 3D HNCO spectrum was acquired on the same 600 MHz instrument, using 16 transients, with 24 ^{13}C and 16 ^{15}N increments, using ^{13}C and ^{15}N spectral widths of 2250 and 1200 Hz, respectively. In addition to HNCO, we attempted CC(CO)NH and related TOCSY experiments, in an effort to obtain sidechain assignments, but very little TOCSY transfer was observed due to relaxation. A 3D HNCA was acquired using 16 transients, and 13 ^{13}C and 16 ^{15}N increments, with ^{13}C and ^{15}N spectral widths of 2100 and 1200 Hz, respectively. Processing of the 2D planes was performed using 90° shifted sine bell window functions in ^{1}H and ^{15}N dimensions. In an another attempt to assign

sidechain NH's to backbone NH's, 3D ^{15}N-NOESY experiment were also acquired on apo-PMK with mixing times of 80 and 150ms using 16 transients, with 70 ^{13}C and 79 ^{15}N increments.

3.2.5 NMR Data Analysis

NMR data were processed using NMRPipe/NMRDraw,[24] and analyzed using NMRView.[25] For all experiments, ^{1}H-^{15}N spectra were processed using a 90° shifted sine function in the ^{1}H and ^{15}N dimensions. For ^{15}N-T_1 and ^{15}N-T_2 experiments, the spectra with the shortest relaxation delay were peak picked using NMRView. For {^{1}H}-^{15}N NOE measurements only one spectrum was peak picked, and for each subsequent spectrum the peak ellipses were manually adjusted to fit each peak.

The R_1 and R_2 relaxation rates were determined by fitting the T_1 ($1/R_1$) and T_2 ($1/R_2$) curves to Equation 3.1:

$$I_t = I_o * e^{(-Rt)} \qquad \textbf{Eq. 3.1}$$

where I_t is the peak intensity after time t, I_o is the peak intensity at time t=0, and R is either R_1 or R_2. Fitting was done using the Rate Analysis package included in NMRView. NOE values were obtained by taking the ratio of the intensity versus the control. This was done with two sets of experiments in order to obtain an error for the analysis.

3.2.6 Modelfree Analysis

The parameters of internal motion were determined from the NMR relaxation data according to the model-free formalism established by Lipari and Szab[5-7]o using Modelfree4 software (version 4.20, A. G. Palmer, Columbia University). R_2/R_1 values

were calculated for each complex for those assigned main chain residues and were used as the starting point for optimization of τ_c values. The optimized τ_c values (Table 3.1) were then fixed for each model afterward. Main chain and side chain dynamics calculations were performed with 300 Monte Carlo simulations per run using an internuclear N-H distance r_{NH} of 1.02 Å and a chemical shift anisotropy (CSA) for the ^{15}N nucleus of -172 ppm for backbone and -114 ppm for sidechain[38]. Five models were used to fit our experimental data and were iteratively tested in order of increasing complexity (M1= S^2; M2= S^2, τ_e; M3= S^2, R_{ex}; M4= S^2, R_{ex}, τ_e; M5= S^2_f, S^2_s, τ_e) until an acceptable fit was achieved. S^2 is the generalized order parameter, τ_e is the internal correlation time, R_{ex} is the exchange contribution term, and S^2_f and S^2_s are for sub-nanosecond and nanosecond motions respectively. These models were tested until they could reproduce the experimental relaxation data within 90% confidence limits using appropriate χ^2 and F-tests. Sidechain N-H's predominantly fit to model 5 for apo-PMK, while the three ligand bound complexes fit best to model 4 in most cases.

To address the concern of anisotropic rotation affecting the transverse relaxation of sidechain residues, we used our data from a previous dynamical study performed on the backbone residues of PMK. In this study, we assessed ligand binding and dynamic properties of PMK and at the time there was no structure (axial-anisotropic model needs a structure to optimize D_{ratio} for correlation time, which is the reason we could not use it in the current study of arginines as they are unassigned) so we could only use an isotropic model to assess correlation times of rotation (ns). Using the isotropic model and average R_2/R_1 as an initial estimate of τ_c, the correlation times were estimated for each complex (Table 3.1).

Table 3.1.

	Apo-PMK	M5P	MgADP	M5P/MgADP
τ_c (ns)	15.6	13.5	17.6	15.8

To test whether anisotropic rotation affects the correlation times of the previous study, we will rely on the crystal structure of human PMK and use the axial model of diffusion, which relies on the structure of the protein in which a global axially symmetric diffusion tensor is used for all spins. From the structure we have the calculated D_{ratio} for human PMK (0.88), giving it somewhat of a disc shape. This input as well as the data from the previous study gives us new estimated correlation times based on the anisotropic rotation of PMK (Table 3.2).

Table 3.2.

	Apo-PMK	M5P	MgADP	M5P/MgADP
τ_c (ns)	15.4	13.9	19.1	17.0

The S^2 values from these data are nearly the same as the reported values of the previous study, validating that the use of the isotropic model in calculating the dynamic properties of PMK.

3.3 Results and Discussion

3.3.1 Arginine Assignment and Chemical Shift Perturbation

Since PMK is a rather large protein (24 kDa), in terms of NMR studies, samples were prepared with only arginines having the ^{15}N label (Figure 3.2A). This allowed us to make sure that every arginine backbone residue was considered, without concern for signal overlap, along with the ε-N−H's of the side chain. Other arginine side-chain N−H's, though labeled, exchange too rapidly to be observed. Because PMK has

significant resonance overlap in the backbone N−H region and also because of significant relaxation that precluded side-chain assignments via TOCSY experiments, only half of the arginine residues could be assigned. The assigned arginines are labeled accordingly on the spectra, and the rest of the arginine backbone N−H's are labeled R-A through R-I. The ε-N−H's of the arginine side chains are shown at the top of Figure 3.2A and expanded in Figure 3.2B.

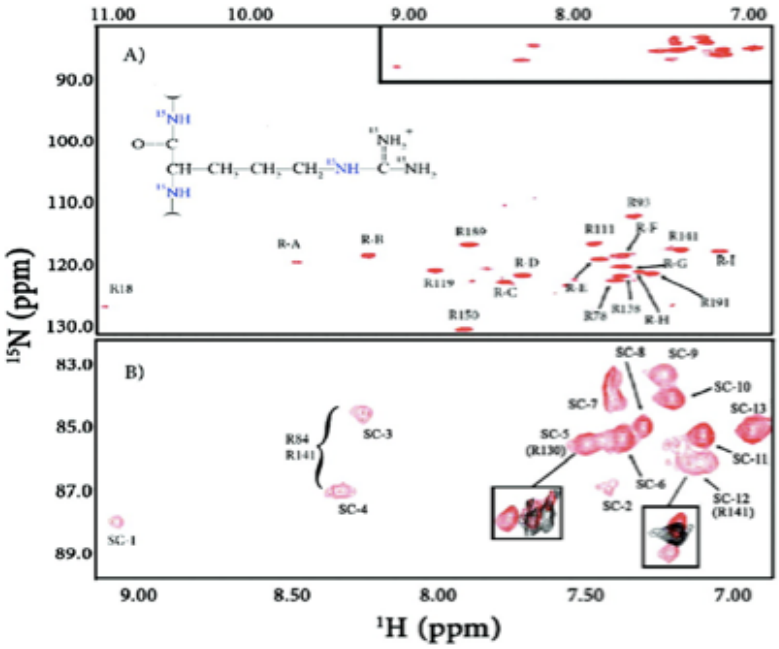


Figure 3.2. ^{1}H-^{15}N HSQC spectra of (A) ^{15}N-Arg-PMK with backbone residues labeled and (B) expansion of arginine sidechain region with cross-peaks labeled and spectra of mutational assignments (boxes). Spectra were of 500 μM PMK (20 mM potassium phosphate, 5 mM DTT, 10% D_2O, 10% d^6-glycerol) acquired at 298K on a 600 MHz Varian NMR System.

Many attempts were made to assign the arginine ε-N−H's to their respective backbone amides using various NMR spectroscopy experiments including HNCO, HNCA, HNCACB, CC(CO)NH, and ^{15}N-NOESY. Displayed in Figure 3.3 is the HNCA experiment of ^{15}N/^{13}C-Arg and U-^{13}C PMK to assign the CHd of the arginine side chain.

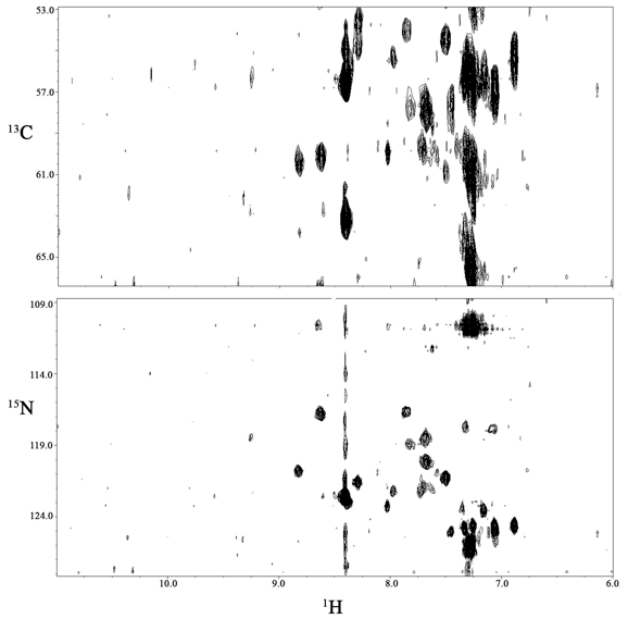


Figure 3.3. HNCA of ^{15}N/^{13}C-Arg and U-^{13}C PMK. Spectra are of PMK that is uniformly ^{13}C labeled, and has ^{15}N label present only in arginines. Spectra are 2D ^{1}H-^{13}C (top) and ^{1}H-^{15}N (bottom) planes from the 3D HNCA experiment.

These spectra provide confirmation for arginine specific labeling, versus metabolic scrambling of label into other amino acids. Sidechain crosspeaks show connection of the NHe proton to the CHd of the sidechain. Partnering experiments HNCACB and CC(CO)NH to assign the CHd of the side chain with the b and g protons as well as the amide N-H gave little signal; not enough to resolve assignments. Spectra in panels (A) and (B) of Figure 3.4 are standard ^{1}H-^{15}N HSQC spectra, of backbone and sidechain regions of ^{15}N-Arg labeled PMK, and are provided for comparison with HNCO 2D planes (panels (C) and (D)). The ^{1}H-^{15}N plane from the 3D HNCO spectrum (panel (C)) shows crosspeaks for arginine backbone ^{15}NH's, adjacent to ^{13}C=O carbonyls, but also for arginine sidechain ^{15}NHe's, adjacent to ^{13}C=NH_2 iminos, of the guanidinium groups. Panel (D) shows the corresponding ^{1}H-^{13}C HNCO plane, that correlates proton shift of

the arginine backbone or sidechain epsilon ^{15}NH with ^{13}C chemical shift of the adjacent carbonyl or imino carbon. Imino ^{13}Cz shifts are identified with red lines. In panel (C), the R150 backbone and all NHe sidechain crosspeaks are folded, as are the sidechain ^{13}Cz crosspeaks in panel (D).

The last attempt to assign the arginines using NMR was to use a ^{15}N-NOESY experiment in which magnetization starts on the backbone amide and transfers to any proton within 5 Å, including the ε-N-H of the arginine side chain allowing for assignment. This experiment was run with two isotropic mixing times of 80 and 150 ms with both experiments giving good signal to noise for the backbone amides however the signal on the ε-N-H of the arginines are very weak (Figure 3.5). This is likely due to the fact that magnetization, when on the ε-NH, was transferred rapidly to water and was subsequently destroyed with water suppression, which led to a decrease in signal intensity (red box in Figure 3.5) as compared to backbone NH intensity. Attempts to solve this issue using various forms of water suppression proved unsuccessful.

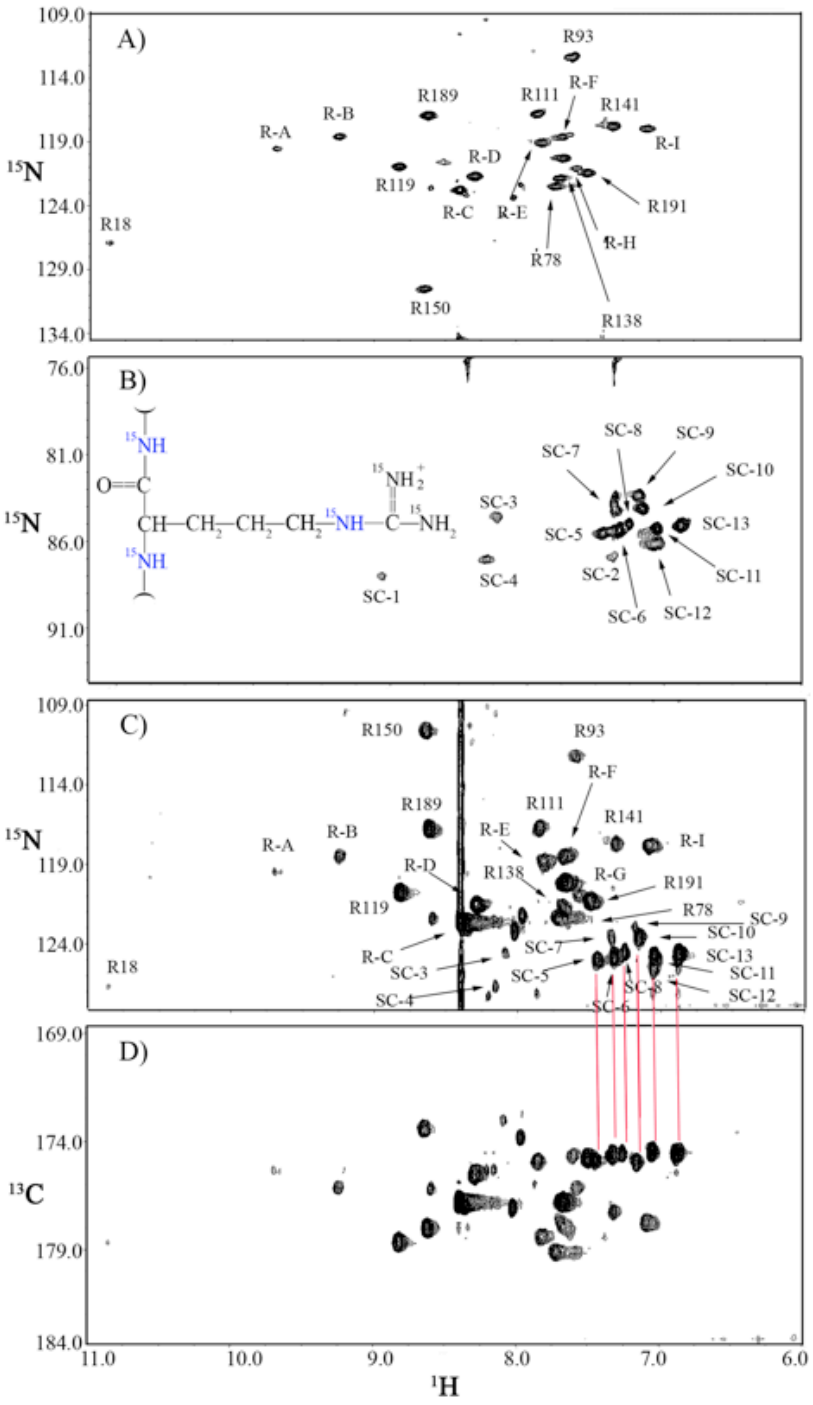


Figure 3.4. HNCO and HSQC spectra of $^{15}N/^{13}C$-Arg and U-^{13}C PMK. Spectra are of PMK that is uniformly ^{13}C labeled, and has ^{15}N label present only in arginines.

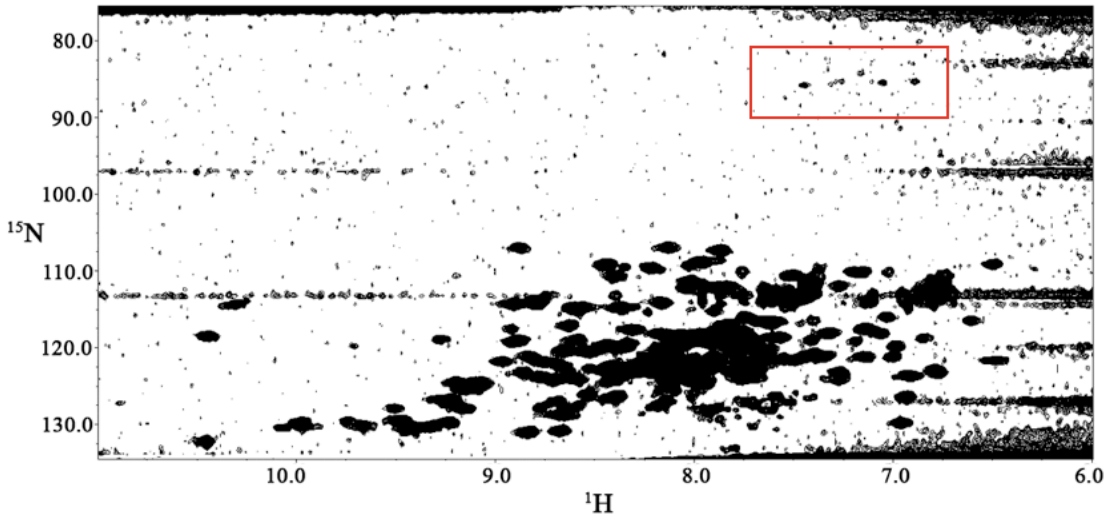


Figure 3.5. ^{15}N-^{1}H plane from a ^{15}N-NOESY experiment of 500 uM apo-PMK in 20 mM potassium phosphate at pH 5.5. ^{15}N-NOESY was run with a mixing time of 150 ms, ε-NH signal was weak (shown in red box) compared to backbone NH's, due to increased exchange with water. These data are shown, to demonstrate the infeasibility of arginine sidechain assignment using the NOESY experiment. Similar problems were encountered with TOCSY experiments as well.

We had to then resort to assigning these arginines using point mutations. The mutants R18M (arginine # 18 mutated to a methionine), R84M, R110M, R111M, R130M, and R141M proteins were all expressed in ^{15}N minimal media so that these mutated arginine side chains could be identified in ^{1}H-^{15}N HSQC based on perturbed chemical shifts. Figure 3.6 shows four of the mutant PMK ^{1}H-^{15}N HSQC spectra (R84M, R111M, R141M, and R130M) with the black spectra being native apo-PMK chemical shifts and red being the corresponding mutant PMK. In the R84M and R111M panels, notice that mutation causes the 3 crosspeaks on the left to be perturbed, broadening or shifting, suggesting that these arginine sidechains are likely to be in some proximity to each other. This is consistent with the titration of M5P onto apo-PMK (Figure 3.7), that showed these same 3 cross peaks broaden upon binding. The R141M mutant causes

some cross peaks to change slightly, however one crosspeak (SC-12) clearly disappears, suggesting it is most likely from the ε-NH of R141. As with the R141M mutant, the R130M mutant causes cross peaks to change slightly, however SC-5 completely disappears, implying this is probably the ε-NH of R130.

To monitor arginine involvement in the binding of PMK's substrates, chemical-shift titration experiments were performed (Figure 3.7). Increasing amounts of M5P and MgADP were added to PMK, and the arginine backbone and side chains were monitored using $^{1}H-^{15}N$ HSQC spectra. The majority of the arginine N−H's showed very small chemical shift changes, indicating little change in chemical environment for residues labeled SC-2 and SC-5−13. For residues labeled SC-1, -2, and -4, the chemical shifts were exchange-broadened, indicating some change in environment and motion due to binding of M5P (Figure 3.7C). This experiment was duplicated using MgADP (Figure 3.7D), monitoring chemical shift changes as before. Only SC-1, -2, and -4 showed large chemical shift changes, suggesting that they are involved in binding.

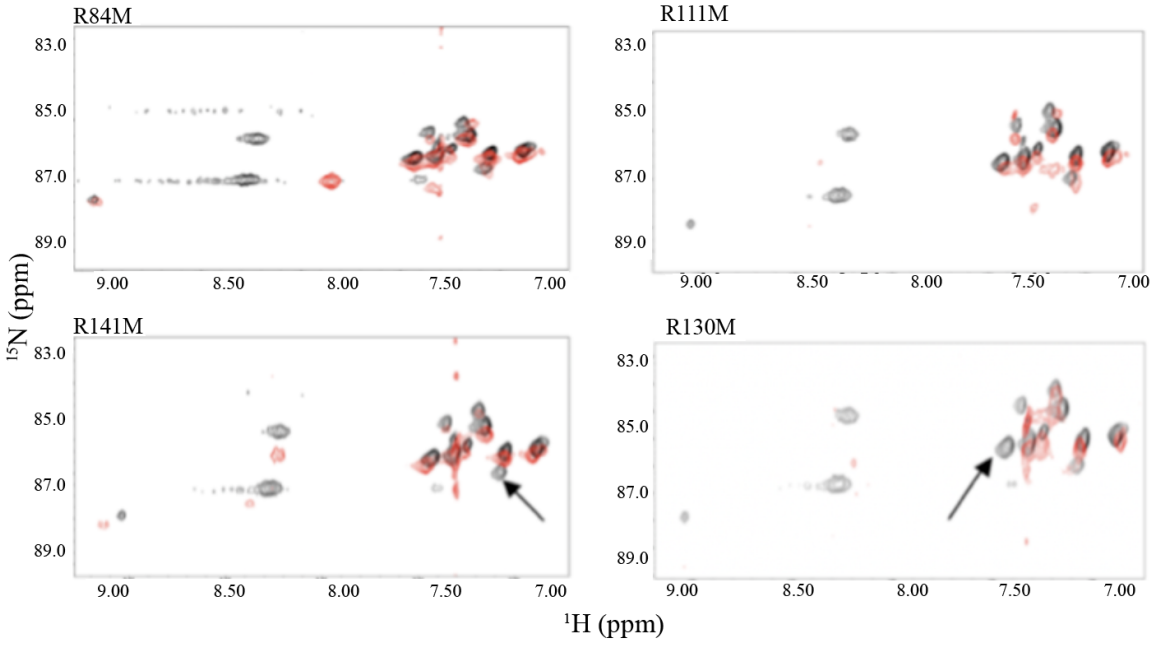


Figure 3.5. Assignment by site-directed mutagenesis. Overlaid ^{1}H-^{15}N HSQC spectra of the arginine ε-NH region where black indicates apo-PMK arginine chemical shifts, and red is for mutant (R84M, R141M, R111M, and R130M) PMK chemical shifts.

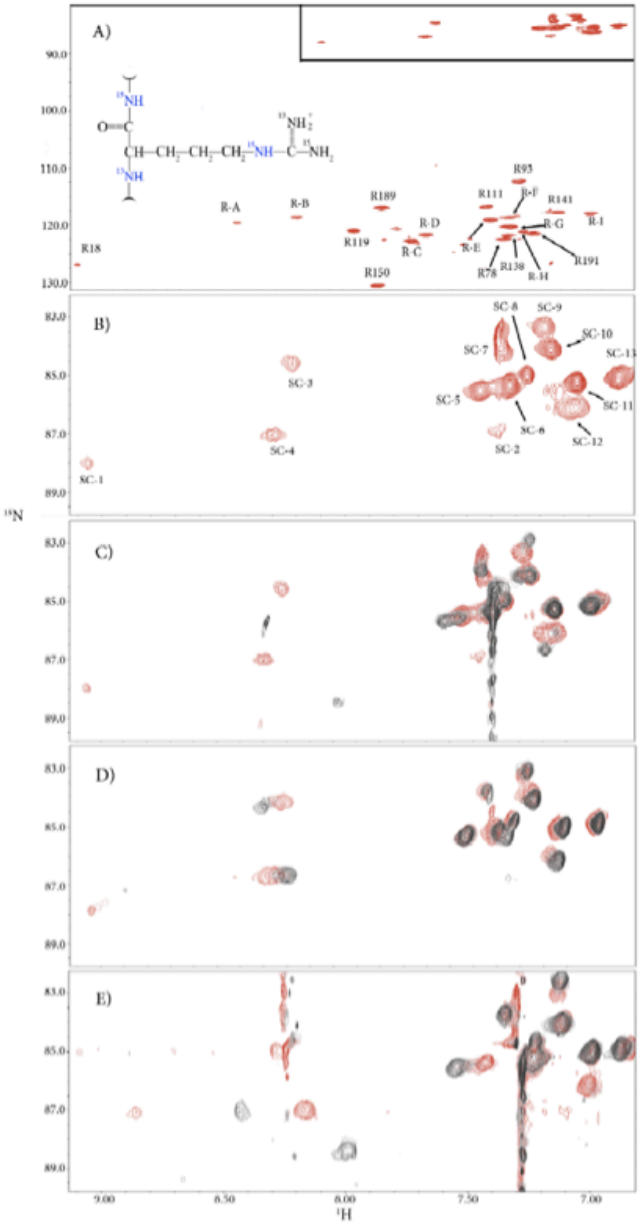


Figure 3.7. (A) ^{1}H-^{15}N HSQC spectrum showing Arg-PMK chemical shifts, including side chains (top) and backbone (bottom). (B) Expanded region of side chain chemical shifts. (C) Chemical shift titration showing effect of increasing amounts (100 uM increments) of M5P added to apo-PMK (red) to saturation (1 mM) shown in gray. (D) Chemical shift titration of increasing amounts of MgADP added to apo-PMK (red) to saturation (1 mM), shown in gray. (E) Chemical shift titration of increasing amounts of M5P added to the MgADP/PMK complex (red), forming the ternary complex (1 mM) shown in gray. Backbone and sidechain ^{15}N-H atoms, monitored in the spectra, are identified on the arginine structure (blue) shown in the inset in panel A, although the ε-N-H is the only observable sidechain NH, since it is the least acidic (least prone to exchange with solvent). Spectra were taken at 600 MHz at 298 K. Arginine backbone assignments are indicated, when available. SC-1 to SC-13 are unassigned arginine side chain N-H's (no side chains have been assigned). R-A to R-I are unassigned arginine backbone N-H's.

3.3.2 Arginine Dynamics

Dynamics experiments were performed as before, where we showed that substrate-induced domain movement occurs in PMK.[39] The longitudinal relaxation rates (R_1) were generally the same for the backbone and side chains for each particular complex and agreed with previous results[39] (Figure 3.8 and Table 3.3). The transverse relaxation rates (R_2) of the backbone N−H's also correlated with our previous studies.[39] The side chains, however, showed a decrease in transverse relaxation rates relative to the backbone. This difference in R_2 values for the apo-PMK arginine side chains was significant (half relative to backbone) while the difference in R_2 between backbone and sidechain was small for the M5P/PMK complex, with the other two complexes showing an intermediate decrease (Figure 3.8). The $^{1}H-^{15}N$ NOE values for the side chains varied from residue to residue, but this variability was consistent between complexes.

	Free	M5P	MgADP	MgADP/M5P
R_1 (s^{-1})	0.78 ± 0.06	1.1 ± 0.13	0.89 ± 0.07	0.82 ± 0.09
R_2 (s^{-1})	14.0 ± 1.3	15.6 ± 1.5	15.6 ± 2.1	15.2 ± 1.9
^{1}H-^{15}N NOE	0.30 ± 0.11	0.15 ± 0.08	0.28 ± 0.10	0.23 ± 0.11
S^2	0.47 ± 0.04	0.67 ± 0.09	0.64 ± 0.04	0.75 ± 0.06
τ_c (ns)	20.1	15.5	22.0	18.4

Table 3.3. Average dynamic parameters for the ε-NH's of arginine sidechains, with the correlation time (τ_c) of the backbone arginine residues used for calculating S^2 values of the arginine sidechain N-H's. Note: These are only averaged in those cases where all four complexes have data for a particular sidechain residue.

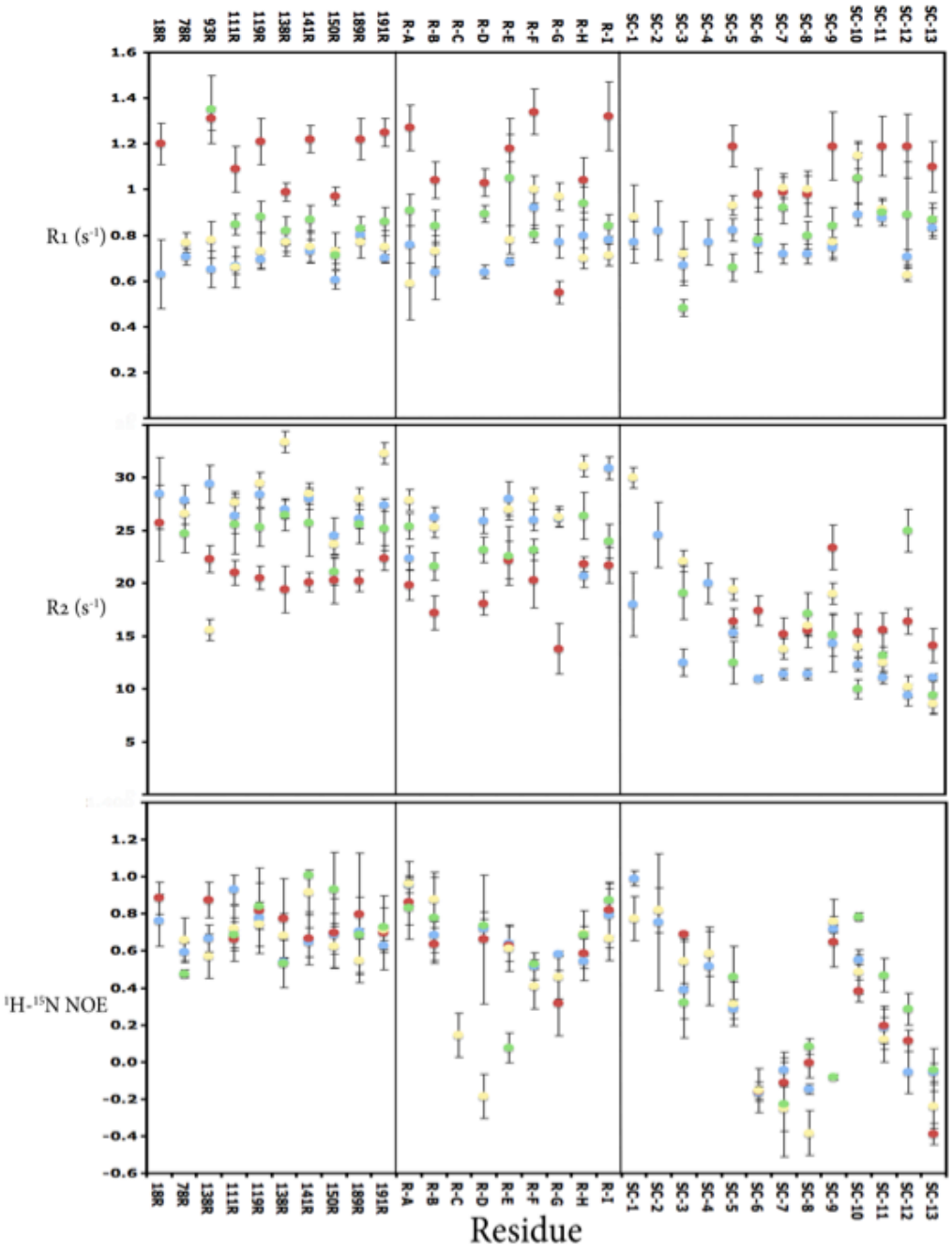


Figure 3.8. NMR dynamics data including R_1, R_2, and ^{1}H-^{15}N NOE values, where blue is apo-PMK, red is the M5P/PMK binary complex, yellow is the MgADP/PMK binary complex, and green is the M5P/MgADP/PMK ternary complex. The sections are divided into assigned backbone N-H's, unassigned backbone N-H's and side chain guanido N-H's. These are the data that were used to calculate order parameters in Figure 3.9.

As in previous studies,[39] a model-free analysis(2) was used to generate the generalized order parameter, S^2. The S^2 value is a measure of rigidity (where $S^2 = 1.0$ indicates complete rigidity and $S^2 = 0.0$ complete fluidity) on the nanosecond-to-picosecond time scale for the bond of interest,(1, 2) in our case the arginine backbone and side-chain N−H bond vectors. Model-free analysis can also differentiate between global and local motions. In our previous work on PMK, we showed that domain movement (global motion) was affected by ligand binding, with little attenuation of local motions. In the side-chain studies presented herein, it was observed that ligand binding causes the arginine ε-N−H's to increase S^2 on average from 0.47 in apo-PMK to 0.64−0.75 in the binary and ternary complexes (Figure 3.9 and Table 3.3). As a control to test whether it is coulombic attraction that is causing these arginines to rigidify upon ligand binding, the same dynamics experiments were repeated as before except in 100 mM KCl buffer (Figure 3.10). Addition of KCl into the buffer causes the R_2 values the MgADP bound PMK to decrease on average from 15 sec^{-1} to 10 sec^{-1} which subsequently seems to cause the S^2 values to decrease in every arginine NH (except SC-12/R141).

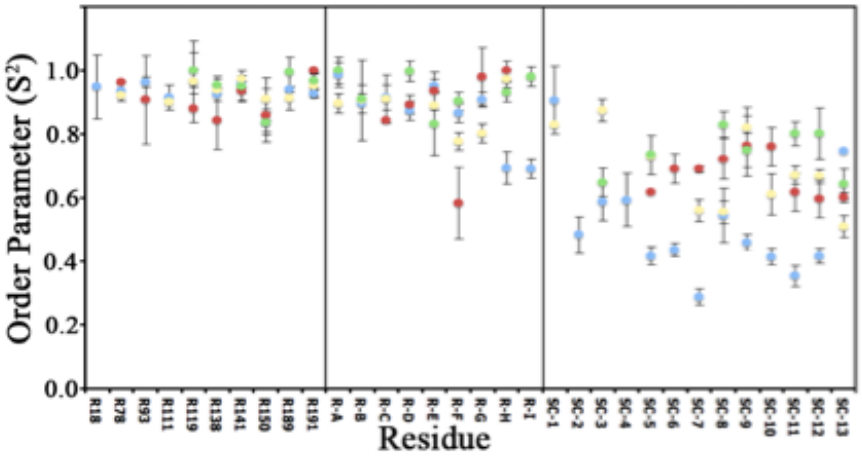


Figure 3.9. Generalized order parameters (S^2 values) for all assigned and unassigned arginine backbone and sidechain ε-N-H's, where blue is apo PMK, red is M5P/PMK complex, yellow is MgADP/PMK complex, and green is the M5P/MgADP/PMK ternary complex. Using mutant forms of PMK, we have identified SC-3 and 4 belonging to either R84 and R111, and SC-12 as probably belonging to R141. Nearly all arginine sidechain residues undergo a significant increase in S^2 (i.e. rigidification) upon binding either ligand.

Residue Number	Tentative Assignment	Distance to Nearest Neg. Charge (ε-NH)		S^2 Values of PMK Complexes			
		Distance	Atom	Apo	M5P	ADP	Ternary
SC-3	R84	3.0 Å	Carboxyl of M5P	0.588	-----	0.875	0.648
SC-4	R111	2.6 Å	Carboxyl of M5P	0.593	-----	-----	-----
SC-5	R130	14 Å	γ-phosphate of ATP	0.417	0.618	0.730	0.734
SC-12	R141	14 Å	α-phosphate of ATP	0.418	0.597	0.667	0.802

Table 3.4. Distance from arginine sidechain NH (for those that were tentatively assigned using site-directed mutagenesis) to the nearest negatively charged atom on either substrate. Also shown are changes in order parameter, when available. Note that while R141 is close to substrate, it is still distal from any charged atom on the substrate.

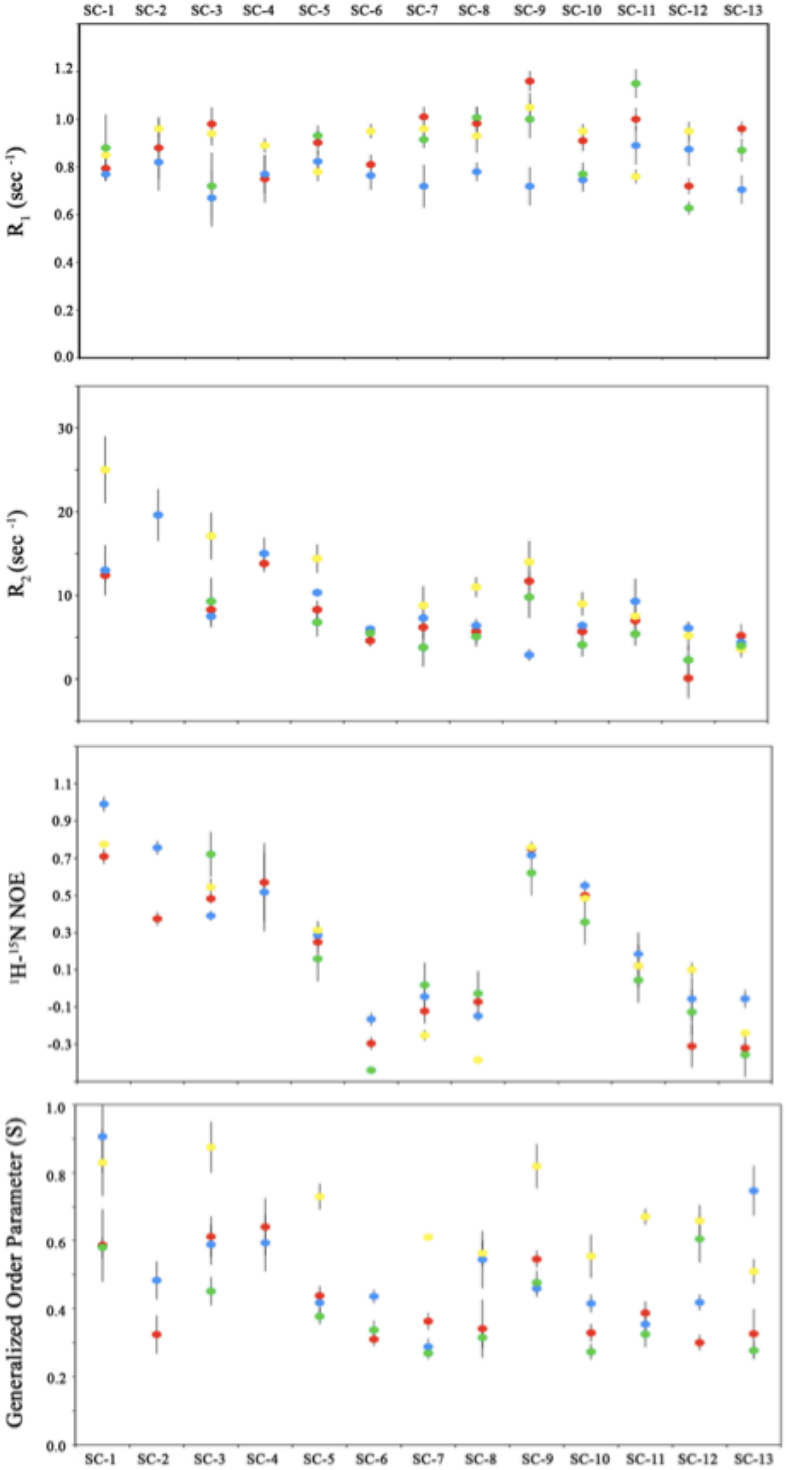


Figure 3.10. Effect of adding 100 mM KCl. NMR dynamics data of arginine side chain ε-NH's of PMK including the $R_1$and R_2 relaxation rates, ^{1}H-^{15}N NOE, and the generalized order parameter (S^2), where blue is Apo PMK, red is Apo PMK in NMR buffer including 100 mM KCl, yellow is MgADP bound PMK, and green is MgADP bound PMK in NMR buffer including 100 mM KCl.

It should be noted that MgADP bound PMK without salt was able to fit to model 4 while addition of salt seemed to cause the dynamical changes to fit only model 5. It should also be noted that the backbone NH's of MgADP bound PMK with in high salt behaved the same way as with regular NMR buffer (in terms of S^2 values), suggesting that only the ε-NH's are affected by the high salt. This leads to the conclusion that ligand binding causes arginine side chains to rigidify throughout the protein, even distal to the binding site, perhaps as a result of long-range Coulombic interactions with the substrates. The S^2 values for the arginine backbones corresponded to our previous results[39] on fully ^{15}N labeled PMK, where we reported changes in τ_c due to domain motion, analogous to those in adenylate kinase where the differences are attributed to different buffer conditions. Surprisingly, the backbone dynamics changes were minimal, in stark contrast to the side-chain effects.

3.4 Conclusion

It has been noted in studies of side-chain methyl groups that changes in the dynamic state of side chains may not be well-reflected in changes observed in the backbone.[40, 41] We also have observed quite large differences between the backbone and side-chain responses to substrate binding. Whether such dramatic differences between side-chain and backbone dynamics changes are common in arginines (vs aliphatic side chains) and/or in interactions with highly charged substrates (multiple phosphate groups) cannot be concluded until more arginine side-chain dynamics studies on other proteins are reported.

While it is possible that substrate binding simply causes a global rigidification of side chains in PMK, this seems unlikely. We propose that side chains rigidify because of Coulombic attraction from the highly charged substrate(s); in support of this hypothesis, the ADP-induced S^2 effect is abolished in high salt (Figure 3.10). This must be a long-range effect, since only four arginines are within 4 Å and only six arginines are within 8 Å, even though 10 arginine side chains rigidify (Figure 3.9).

In summary, we have shown that substrate binding to PMK causes arginine side chains to undergo a transition from a flexible to a rigid state. The magnitude and global nature of this effect is significant. Besides providing mechanistic insights, such studies will facilitate computational docking by identifying flexible side chains that can be allowed to sample conformational space during the docking calculations.

CHAPTER IV. Proteomic Investigation of the N- Teriminal Disordered Region of Unknown Function in Human Phosphomevalonate Kinase

4.1 Introduction

4.1.1 N-terminal Peptide of PMK and Aims to Discover Function

The most noticeable feature of the order parameter profile (Figure 4.1) of backbone amides in PMK is in the N-terminal region, which has no homology to adenylate kinase or to the T4 deoxynucleotide kinase template, upon which the homology model was based. But, there is a dramatic and sharp change in S^2 from a disordered state ($S^2 \approx 0.4$) to an ordered state ($S^2 > 0.8$), occurring immediately after Arg9 where the homology model starts. The fact that these nine residues are completely disordered in apo Phosphomevalonate Kinase (PMK), and all binary and ternary complexes, combined with the fact that organisms such as *C. elegans* lack this N-terminal segment entirely, suggests it is not an integral part of the PMK structure.[14] Indeed, inspection of the homology model reveals no potential structural role for these nine N-terminal residues. This led us to speculate that these residues may be part of a signal or regulatory peptide, which might be cleaved off by an as-yet unidentified protease (note: this region shows no homology to membrane-binding peptides, and there are no data suggesting PMK is membrane bound).

To determine the function, if any, of this N-terminal region, we mixed human liver proteins with ^{15}N-PMK to determine if any proteins in liver tissue bind to or cleave PMK. We also attempt to identify those proteins that do bind by purifying them using an affinity column with just the N-terminal disordered region (synthesized peptide) covalently attached to a column. Proteins that bind to the affinity column, and are eluted by free peptide, will be identified by mass spectrometry (MS). Finally, to ensure that those proteins that are identified do actually bind to the peptide, saturation transfer

difference STD NMR binding studies are performed with the liver proteins and the PMK-derived peptide.

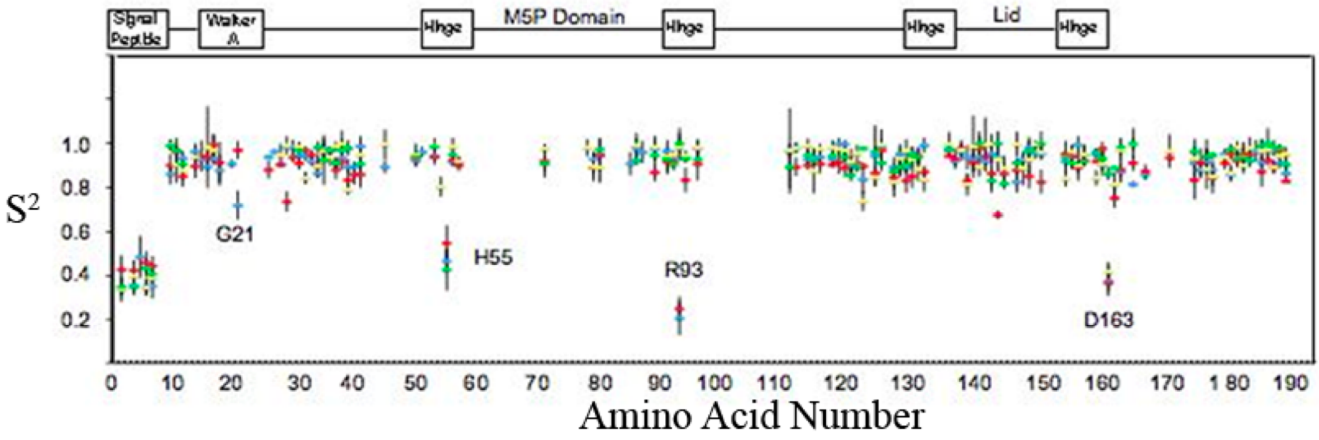


Figure 4.1: S^2 values of the backbone amide N-H's of human PMK with the hinge regions and domains labeled. Note the 9-residue signal peptide S^2 values at amino acid number 1-9.

4.2 Methods and Materials

4.2.1 Instruments and Materials

All NMR experiments were performed on a 600 MHz Varian NMR System instrument equipped with a triple resonance cryoprobe and *z*-axis gradients. Peptides were synthesized at the Medical College of Wisconsin where they were prepared with the N-terminus acetylated. All NuPage and Novex products used for SDS-PAGE experiments were purchased from Invitrogen. All other chemical reagents, including activated CH Sepharose 4B, as well as human liver proteins (cytoplasmic) were purchased from Sigma Aldrich.

4.2.2 ^{15}N-^{1}H HSQC Spectra of PMK with Human Liver Proteins Added

^{15}N-PMK was exchanged into a buffer containing 20 mM potassium phosphate and 5 mM DTT at pH of 5.5. Liver proteins and 500 uM PMK was combined into equal volumes in an NMR tube to attain 400 uL total volume. HSQC spectra were acquired after liver protein addition, at 6 hours, and at 24 hours to assess if any changes in chemical shifts occur with time. HSQC spectra were then processed using NMRPipe and NMRDraw using a 90° shifted sine bell function in ^{1}H and ^{15}N dimension.

4.2.3 Peptide Affinity Column Preparation and Purification of Liver Proteins

Both peptides were prepared for coupling to the column resin the same way. One gram of CH Sepharose was added to 200 mL of 1 mM HCl solution and the resin was allowed to swell for 15 minutes. This solution was then filtered into a column and washed with 500 mL of 1 mM HCl. The resin swelled to a volume of around 3 mL. To couple the ligand to the resin, 10 mg of ligand (1.7 mM) was dissolved into 3 mL of coupling buffer that consisted of 0.1 M $NaHCO_3$ and 0.5 M NaCl at pH 8.0. The 3 mL of peptide ligand solution was added to the resin in the column and shaken at room temperature for 2 hours, to permit chemical coupling by reaction of ligand amino groups with the N-hydrosuccinamide groups on the resin. After two hours, the residual ligand was washed with 30 mL of coupling buffer. The remaining active groups of the resin were blocked with 0.1 M Tris buffer at pH 8.0 by washing the resin with buffer and letting the resin incubate in the buffer for 1 hour. The resin was washed with alternating washes of 0.1 M acetic acid with 0.5 M NaCl at pH 4.0, then with 0.1 Tris-HCl with 0.5 M NaCl at pH 8.0. This was done to ensure that no free peptide ligand remained bound via ionic interactions. The column was then packed using a Pharmacia LKB peristaltic

pump with 50 mM phosphate buffer with the same pH as the human liver proteins, pH at 7.4. It was packed until the bed volume stabilized. 100 uL of human liver proteins were then added to 100 uL of phosphate buffer, and then added to the top of the resin. The resin was then washed with the phosphate buffer, and fractions were monitored by UV-Vis until $A_{280} < 0.005$, indicating that there was no more protein being eluted from the column. Competitive elution was used to elute any bound protein from the column: 20 mg of peptide (control or N-terminus) was dissolved into 5 mL of phosphate buffer with 0.5 M NaCl at pH 5.0 to give a 2 mM peptide solution, and all 5 mL of the peptide solution was added to the resin and phosphate/salt buffer was added to the column after the 5 mL went through, and all eluent was collected in a single fraction. The eluent was then filtered using a Millipore Amicon-4 Centrifugal Filter with a 10,000-dalton cutoff. This solution was concentrated down to approximately 100-200 uL and a SDS-PAGE gel was used to determine the molecular weights of the eluted proteins (Figure 4.4). To get a more precise MW value, the log value of all the MW marker proteins vs. the R_f values of protein bands was plotted and fitted. This gave a fitted equation of $y=-1.815x + 2.444$, where y is the log of MW and x is the R_f of the unknown protein in the eluent. The samples were then sent to the Medical College of Wisconsin Protein and Nucleic Acid Facility for tandem MS (ESI) based identification, based on sequencing of tryptic peptides from the eluted protiens. The trypsin digested peptides derived from the whole proteins were searched against a database of human protein mass spectra fingerprints to then identify the eluted proteins, using standard methods.

4.2.4 Saturation Transfer Difference (STD) of Hemoglobin with Peptide Fragments Present

STD NMR studies were recorded at 298 K with 1 mM N-terminal peptide and 100 μM hemoglobin, and using the Varian cyclenoe pulse sequence. 1D-^{1}H NMR spectra were taken in advance, for both protein and peptide, to select the proper frequency for irradiation. During the STD experiment, alternating on-resonance irradiation of the protein methyl region (around -2.5 ppm), and off-resonance irradiation at a region far from any protein or peptide signals (-10 ppm) was used. The *spacing* and *pattern* values were set to 20 and 5, while *cycle* was set to y and *tau* to 100 ms, with the saturation power at -10 dB. The total irradiation time was set to 4 s, with 512 transients collected for each experiment. All 1D-^{1}H NMR spectra and 1D STD spectra were processed using iNMR on a Mac workstation. Exponential multiplication was used with 2 Hz line broadening for all spectra.

4.3 Results and Discussion

4.3.1 Signal Sequence Database Search

To begin exploring the hypothesis that the 9 N-terminal residues of PMK have a functional/signaling role, and to identify which protease might cleave this N-terminal peptide, an exhaustive search of signal sequence databases was performed, searching only against the peptide sequence in question (i.e. Met1 to just past Arg9). While none of these searches yielded a strongly scoring motif hit, one did give a modest scoring prediction of cleavage immediately after Arg9:

PLGGAPR^{v}LV

This site was identified using the ProP 1.0 server, using a neural network model trained using literature sequence data.[42] While the score for the cleavage site was relatively low (0.112; scores in the 0.5-1.0 range are considered high probability), it is the best scoring motif hit obtained thus far, and should be interpreted in the context of NMR data that indicate cleavage, if it occurs, is expected at exactly this site (after Arg9). The enzyme predicted to cleave here is a *furin-like protease*, which belongs to a family of proteases broadly classified as "proprotein convertases."[43-46] Interestingly, there is growing evidence in recent years that proprotein convertases affect cholesterol levels and lipid metabolism, largely by undefined mechanisms.[47, 48] Furthermore, data have indicated that inhibiting the proprotein convertase PCSK9 may be an effective means to treat hypercholesterolemia and associated cardiovascular disease, as a complement to statin treatment,[45, 46] based on the observation of synergistic lipid lowering effects. Many of the newly discovered proprotein convertases, including PCSK9, do not have any known substrates, and therefore have poorly defined peptide recognition motifs. As such, we are now pursuing studies to identify which, if any, proprotein convertase might cleave the disordered PMK N-terminal peptide, and what affect this cleavage might have. Likewise, it is possible that a protein other then a proprotein convertase may recognize this 9 amino acid sequence, and our studies endeavor to identify this protein.

4.3.2 Human Liver Proteins Binding to PMK

A ^{15}N-^{1}H HSQC 2D NMR spectrum was acquired before human liver protein addition and at each time point after. If any protein binds to or cleaves the N-terminal region of PMK, then we should notice chemical shift changes in the N-terminal amino

acid residues, especially residues 1-9. As shown in Figure 4.2 there are no chemical shift changes in any residues when comparing purified PMK (black) to PMK after the liver proteins have been added (red), then after 4 hrs. (green), and after 24 hrs. (blue). So, HSQC data suggest that no liver proteins bind to or cleave PMK. This could be for any of several reasons, the first being that the concentration of the protein that may be doing the binding or cleaving is not in high enough concentration to produce any chemical shift changes. The approximate PMK concentration was 250 μM when mixed with liver proteins. To see any chemical shift changes, the protein(s) that bind would have to be close to that concentration (at least > 25 μM) to produce any detectable chemical shift changes. Another reason that we may not see binding is due to the additional 20 amino acid chain on the N-terminal region of PMK. This chain contains the histidine tag region that was engineered onto PMK to facilitate purification, and this tag may block binding to other proteins. Finally, it is of course possible that no human liver proteins bind to and/or cleave this N-terminal sequence.

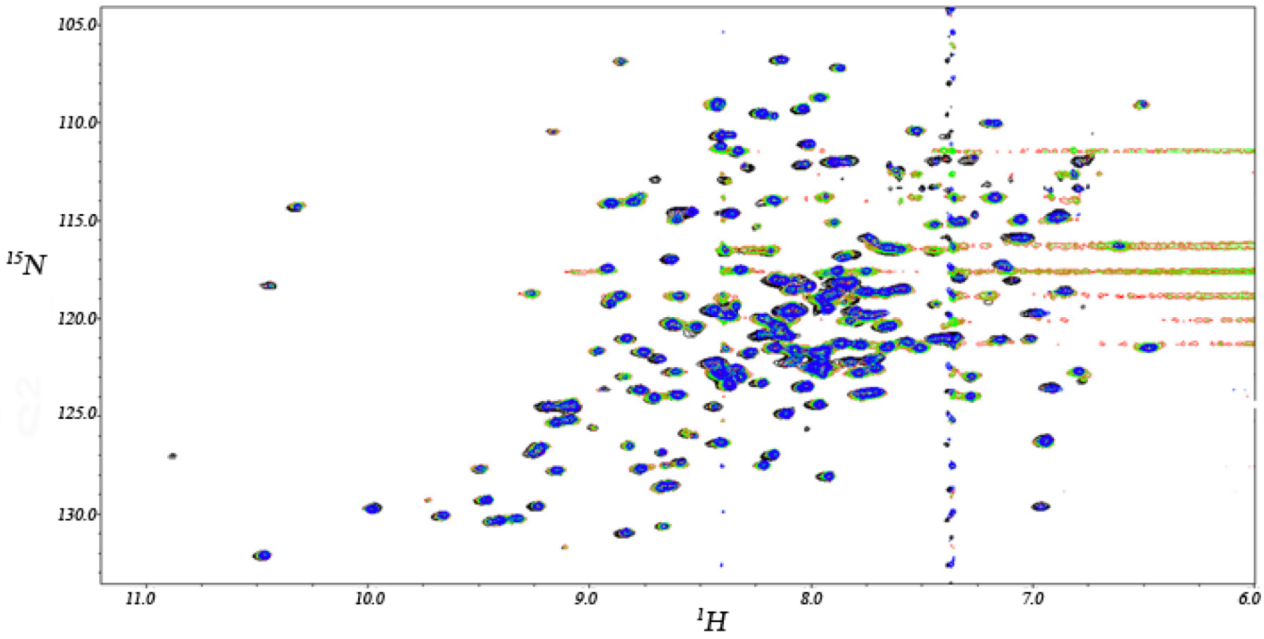


Figure 4.2: ^{1}H-^{15}N HSQC 2D spectra of human PMK (black) and PMK incubated with human liver proteins directly after addition (red), 4 hrs. after addition (green), and 24 hrs. after addition (blue).

4.3.3 Purification of Human Liver Proteins That Bind to the Synthesized N-terminal Peptide Sequence

Activated CH-Sepharose 4B resin is designed to be specific for reaction with primary amino groups. At lower pH, the NHS (N-hydroxysuccinimide) ester becomes available for attack (Figure 4.3) by ligands with active primary amino groups. As stated in the previous section, the disordered N-terminal region starts at Met1 and ends with Arg9, so we had this peptide sequence synthesized with those 9 residues with an additional 4 residues from the sequence and 7 more amino acids added as a flexible and hydrophilic linker, with the two-lysine residues at the C-terminal end to react with the NHS ester on the resin:

Acetyl-MAPLGGAPRLPLLGSGSKSK

These lysine residues (KSK) have the primary amino groups needed to react with the resin. To ensure that the amino group of Met1 does not react with the resin, we had the N-terminal residue acetylated.

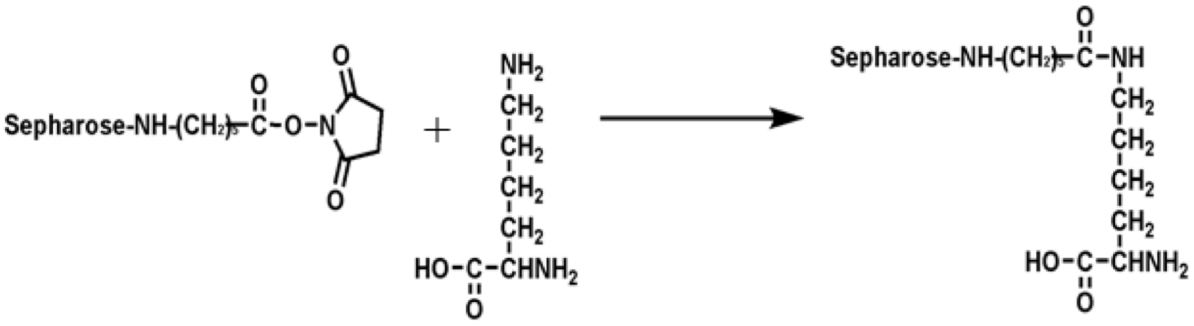


Figure 4.3: Chemical reaction of activated CH sepharose 4B with the primary amine of the amino acid lysine.

The column purification and subsequent concentration of the eluent yielded seven clear bands on the SDS-PAGE gel (Figure 4.4). The molecular weights, as determined by the measuring R_f values of the bands in the gel, are 13.7, 43.7, 52.4, 61.3, 65.1, 73.3, 142.8 kDa. Some of the proteins that correspond to the molecular weights, determined by mass spectrometry (MS), are represented in Table 4.1.

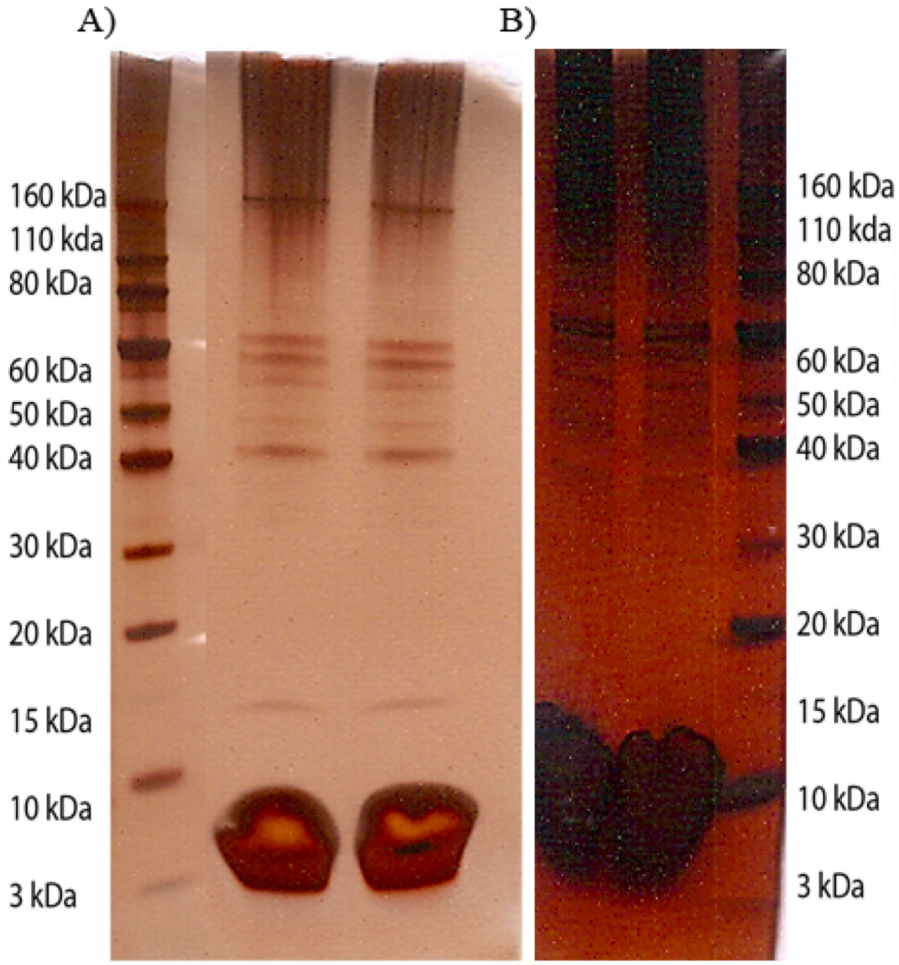


Figure 4.4: SDS-PAGE gel of eluted human liver proteins from the affinity column with covalently attached A) N-terminal PMK peptide and B) scrambled N-terminal PMK peptide. Note: Gel B is much darker due to longer staining time, to make sure that every band is resolved.

Protein	Protein Probability	Score	Peptide Count	Scan Count	Percent Coverage	MWT
Hemoglobin beta subunit	1	85.2	7	22	52.4	15868.26
L-xylulose reductase	1	45.6	4	17	25.5	25914.19
Succinate dehydrogenase	0.965	6.57	2	10	8.93	18611.31
Hemoglobin alpha subunit	1	9.67	4	7	37.1	15127.4
UDP-N-acetylglucosamine transporter	0.966	12.9	1	43	4.01	35985.56
Carbamoyl-phosphate synthase	.999	1.56	2	2	2.2	164940

Table 4.1. Human liver proteins with the highest probability of being bound to the N-terminal region of human PMK. These proteins were bound to the N-terminal peptide that was covalently attached to the affinity column.

In this MS anaylsis, protein probability is the likelihood that the peptides that were sequenced are for the "identified" protein. Peptide count is the number of sequenced peptides in the sample that are present in the "identified" protein; scan count represents the number of times the peptide was identified in the MS-scanning process (reflects S/N); percent coverage is the percent of the total protein sequence covered by the sequenced peptides, and MWT is the molecular weight of the identified protein. Of all the categories, probably the most important is percent coverage, along with score. According to these criteria, the highest rated proteins are hemoglobin subunits α (52.4% coverage) and β (37.1% coverage) as well as L-xylulose reductase (25.5% coverage). Hemoglobin is a tetramer with two α subunits and two β subunits that are used for oxygen transport from the lungs to the rest of the body, where the oxygen is released for cell uptake.[49] L-xylulose reductase is involved in the uronate cycle of glucose metabolism and catalyzes the reduction of xylulose into xylitol.[50] Currently, we have no

reason to suspect that either of these proteins would interact with PMK, however we are testing to determine whether hemoglobin binds the N-terminal peptide fragment (*vida infra*).

To make sure that the human liver proteins are not binding non-specifically to something else in the column, such as support material, we had another peptide synthesized using the same procedure as the first, except the amino acids that comprise the PMK signal peptide were scrambled. We did this by assigning a number to each amino acid then used a random number generator to scramble the order of the amino acids. To make sure there were no important amino acids still together in the sequence (Arg9 and those residues before and after it) we manually adjusted the sequence, as needed. This led to the following peptide, which was synthesized at MCW:

Acetyl-LARGPMLGLLAPVGSGSKSK

We preserved the final 7 amino acids so that they still bind the peptide to the sepharose support. Figure 4.4B shows the concentrated eluent from the scrambled N-terimanl region peptide of PMK. The bands look nearly the same between the two gels, however at 15 kDa, there appears to be no band present for the scrambled peptide indicating that there was no binding of any protein to affinity column at that molecular weight. This indicates that something does indeed bind to the natural N-terminal region of PMK (Figure 4.4A) with the molecular weight of 15 kDa. This strongly supports the MS data (Table 4.1) that hemoglobin (with a molecular weight of approx 15 kDa) could be the protein that binds to PMK's N-terminal disorderd region.

3.3.4 Saturation Transfer Difference STD NMR Binding Assay to Determine if Hemoglobin Binds to the N-terminal Disordered Region of PMK

Saturation transfer difference (STD) is a relatively new NMR technique in which a certain frequency region of the NMR spectra is irradiated; mainly the methyl region and difference spectra are measured. This irradiation of protein methyl protons leads to magnetization transfer via spin diffusion from the methyl protons to the rest of the protein.[51] If a ligand binds to the protein in fast exchange, magnetization is then transferred to the ligand, giving rise to a 1-D difference NMR spectrum for that ligand (the "NOE difference spectrum"). As mentioned in methods and materials, a 1-D ^{1}H NMR spectrum was taken of both the peptide and the hemoglobin, to determine the best spectral region for irradiating (Fig. 4.5). As shown, there is signal as low as 0.6 ppm for the ligand and -2.5 ppm for the protein. The STD experiment (Figure 4.6B) with hemoglobin and peptide, with irradiation of the methyl region at -2.5, gave only a modest increase in STD signal at 0.6 ppm, for peptide. To determine if we are observing artifactual STD of just the peptide, a control experiment was run in which only the peptide (no hemoglobin) was subjected to an irradiation at -2.5 ppm (Figure 4.6A).

When the spectra are overlaid on each other using the same noise level as reference what is observed is that the STD signal with hemoglobin present is roughly twice the signal for the control STD experiment with peptide alone. This doesn't necessarily mean that peptide binds to hemoglobin, as generally the STD signal is much larger. It is possible that STD signal is attenuated because binding is too tight (STD signal requires fast exchange binding). If a ligand binds to tightly, no STD signal will be observed, as the ligand won't be released before the magnetization that was transferred

from the protein decays. More experiments will have to be done to confirm whether the N-terminal peptide does in fact bind to hemoglobin.

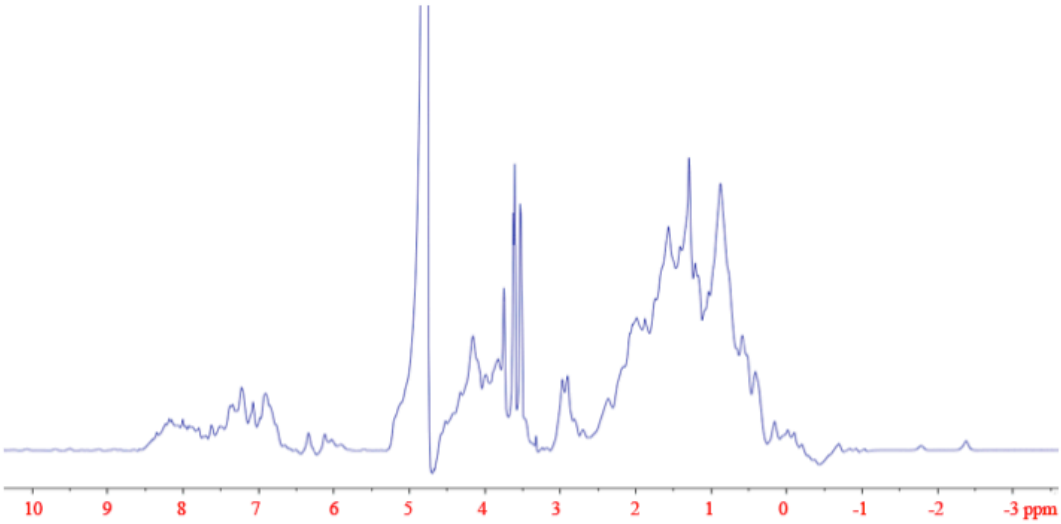


Figure 4.5. 1-D ^{1}H NMR spectra of hemoglobin to determine methyl irradiation for STD NMR spectra. Note: methyl regions extends to -2.4 ppm.

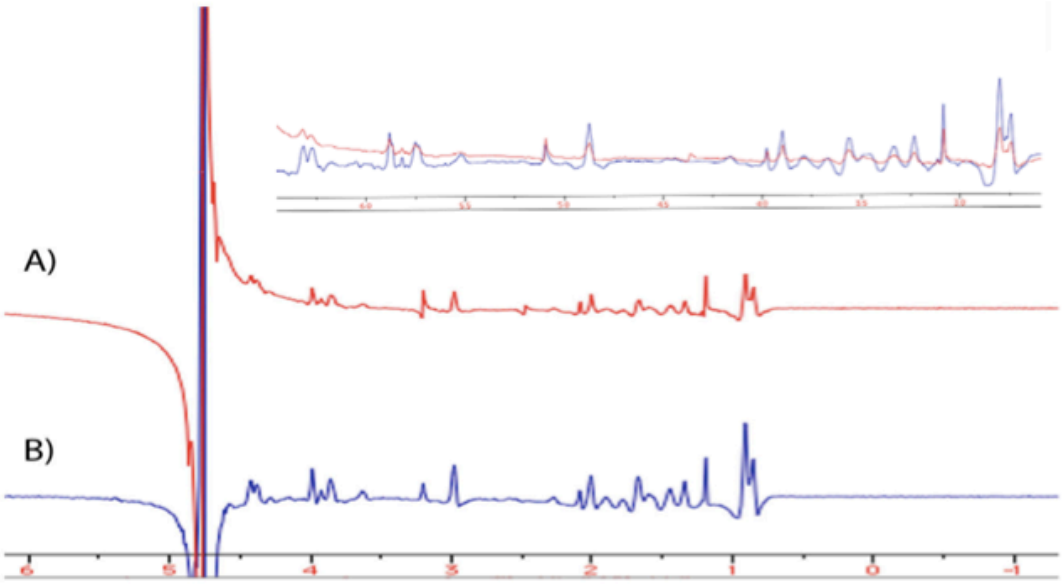


Figure 4.6: 1-D STD NMR spectra of A) the N-terminal region peptide and B) the N-terminal region peptide with 100 μM hemoglobin. An expanded overlay of the possible STD signal in the upper right corner.

4.3.4 Conclusion

The N-terminal region of PMK was synthesized and used in NMR and affinity column binding experiments to determine if there are proteins in the human liver proteome that bind to it and/or cleave it. We have identified proteins from human liver tissue that may bind to the N-terminal peptide of PMK, using tandem mass spectrometry (ESI), with hemoglobin being the most likely binding partner. Hemoglobin was then used in NMR STD binding experiments to determine if binding of hemoglobin to peptide does actually occur. However, it is still unclear if such binding does occur. An alternative strategy to verify binding is to have the peptide synthesized with one of the amino acids in the sequence labeled with ^{15}N, which will then give a signal in ^{1}H-^{15}N HSQC spectra, and which will permit monitoring of binding of the peptide to protein

based on chemical shift changes. The peptide could also be synthesized with a fluorescence tag, and we could monitor binding to hemoglobin based on fluorescence changes.

Chapter V. Solution Structures of Oxidized and Reduced forms of Thioredoxin C, and Models to Describe the Intact Thioredoxin System from Mycobacterium tuberculosis

5.1 Introduction

Mycobacterium tuberculosis (*M. tb*) is the causative agent of the infectious disease tuberculosis, and is a devastating pathogen that has affected (new and previously occurring infection) nearly 11 million people and a mortality rate of 1.3 million people in 2008.[52] It is also thought that a third of the population is infected, with a latent form of the disease, which is not infectious.[53] The pathogen resides in the alveolar phagocytes and can resist the oxidative killing of the cell however this mechanism is still unclear.[54] The method by which phagocytes kill the invading microbes is by the production of reactive oxygen and nitrogen species, including superoxides and hydrogen peroxide, which can create free radicals that become toxic to the microbe.[55] One method that *M. tb* uses to resist this oxidative killing is by the same process that eukaryotic cells use to combat oxidative stress. In other prokaryotes, this is done by the thioredoxin system, which keeps the cell in a reduced state, based on it's ability to control the redox state of cellular thiols/disulfides.[56, 57]

Thioredoxins are well-studied proteins, and found in almost every organism.[58] They share common features such as a similar three-dimensional structure, a conserved catalytic motif of WCXXC in the active site of the protein, and a molecular weight of about 12 kDa.[59] Thioredoxins catalyze thiol-disulfide exchange reactions using redox active cysteine thiols to reduce oxidized disulfide cysteines of other essential proteins, including metabolically essential enzymes.[60, 6162] Oxidized thioredoxins are then reduced by thioredoxin reductase in an NADPH dependant reaction, via redox active cysteine thiols.[63] Disulfide reductions occur via hydride transfer from NADPH to a bound FAD cofactor, followed by electron transfer from $FADH_2$ to bound thioredoxin. The

interaction and reaction between the thioredoxin reductase and the thioredoxin substrates are essential to keep the cell in a reduced state, and define what is called the thioredoxin system. While the thioredoxin system of E. coli is well characterized, that of *M. tuberculosis* is not.[64] If this system were disrupted, the *M. tuberculosis* pathogen would be more vulnerable to the oxidative attacks of our immune cells (neutrophils, macrophages).

The thioredoxin system in *M. tb* is comprised of three thioredoxins (TrxA, TrxB, and TrxC) and one thioredoxin reductase (TrxR).[54, 64, 65] Many organisms have multiple genes encoding thioredoxins, which is probably a function that enables coping with the amount and perhaps type of oxidative stress in a cell. It has been shown that TrxB and TrxC behave as general disulfide reductases with near equal reduction potential, while TrxA is observed to have a weak capacity to act as a disulfide reductase, and was found to not be a substrate of TrxR.[64] It was also observed that the ability of *M. tb* to survive redox stress isn't dependant on the expression of one thioredoxin over another, which suggests that having multiple thioredoxins may be a "redundant system" to ensure survival under oxidative conditions.[64]

It is for these reasons, and the severity of the tuberculosis disease, that make the thioredoxin system of *M. tb* a suitable target for drug inhibition, as long as a drug could be designed knock out the whole thioredoxin system (ex. inhibit TrxR binding).[66] There has been some work done on the inhibition of TrxC signaling.[67] The crystal structure of TrxC in the oxidized state has been solved to help facilitate those studies.[65] Previous studies of the different redox states of prokaryotic thioredoxins have shown that there are subtle changes in structure in going from oxidized to the reduced state.[68-70] Most of the

changes are occuring in the active site, around the redox active cysteines, with some conformational flexibility observed in the active site loop.[68, 71] The secondary structure of the whole protein remains the same, as the thioredoxin goes from oxidized to reduced state. Whether this is also true in *M. tb* is not known, since there is no structure of reduced TrxC (or of any TrxA or B).

Here we report the NMR solution structures of *Mycobacterium tuberculosis* thioredoxin C in both the oxidized and reduced states, with discussion of structural changes that occur in going between redox states. The NMR solution structure of the oxidized TrxC, when compared to the crystal structure, corresponds very closely. However, crystal packing may have caused an artifactual shift in the α4 helix in the reported crystal structure, compared to the solution structure. With the previous report of the crystal structure of *M. tb*'s thioredoxin reductase (not bound to a Trx), we have modeled the complete thioredoxin system in the mechanism (catalytic cycle) for disulfide exchange. This was based on existing structures of *M. tb* TrxR and both redox states of TrxC (our solution structures), and using the *E. coli* TrxR/Trx complex as a template for model building.[72]

5.2 Methods and Materials

5.2.1 Protein Expression and Purification

The plasmid containing the gene Rv3914 (TrxC) in the pET vector pET22b and gene Rv3913 (TrxR) in pET vector pET23a was obtained from Tuberculosis Vaccine Testing and Research Materials at Colorado State University and Shkhar C. Mande from the Center for DNA Fingerprinting and Diagnostics in Hyderabad, India.[73] The plasmid

containing the expression constructs (with a C-terminal histidine tag) was transformed into *E. coli* BL21-(DE3) Rosetta cells from Novagen and grown overnight at 37° C, which was then used to inoculate 2 L of LB (Luria Bertani) media. The media was closely monitored at O.D. 600 nm until the cell density reached an absorbance of 0.7, at in which point, the media containing the cells for expression of TrxR were induced with 1 mM IPTG and the cells containing TrxC, the cells were then spun down at 4000 rcf and washed with M9 minimal media salts at pH of 7.0. The cell pellets were transferred to 0.5 L of M9 minimal media containing 0.5 grams $^{15}NH_4Cl$ and 2.0 grams of ^{13}C-glucose as the only sources of nitrogen and carbon, to yield uniformly labeled protein.[74] The cells were allowed to acclimate for 1 hour and then induced with 1 mM IPTG for an additional 4 hours at 37° C, then cells were harvested. The bacterial pellet was resuspended in 50 mM Tris base, 300 mM NaCl, 5 mM imidazole, and 10% glycerol at pH of 7.8, and lysed using sonication. The lysate was clarified by centrifugation at 15,000 RPM for 30 minutes and the supernatant loaded onto 1 mL of nickel-Sepharose (Amersham Scienes) resin at 1 mL/minute. The bound proteins (TrxC/R) were then washed with lysate buffer, then with lysate buffer containing 25 mM imidazole, and then eluted using lysate buffer containing 300 mM imidazole. The eluted protein was than dialyzed with 40 mM potassium phosphate buffer at pH of 6.3 (NMR structures) and pH of 7.0 (TrxC/R binding experiments) and concentrated to 1 mM, with protein concentration determined by the absorbance at 280 nm using extinction coefficients of 11,000 and 14,440 L/mol-cm for TrxC and TrxR respectively by using the resource ProtParam (http://expasy.org/tools/protparam.html). To reduce the disulfide, 5 mM DTT was added to the protein samples.

5.2.2 NMR Spectroscopy and Structure Calculation

All NMR experiments were performed on a 600 MHz Varian NMR System at 599.515 MHz using a triple resonance cryoprobe with z-axis gradients at 25 °C. Backbone assignments were obtained using 2D ^{1}H-^{15}N HSQC, and 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCA(CO)NH, and HNCACB spectra (Figure 5.1). The side chains were assigned using CC(CO)NH, HBHA(CO)NH, HCC(CO)NH, and HCCH-TOCSY spectra. Distance restraints were obtained from 3-D ^{15}N-NOESY, ^{13}C aliphatic NOESY, and ^{13}C aromatic NOESY experiments which were obtained using 150 ms mixing times. Structures were calculated using these distance restraints with the program CYANA[1, 2], in conjunction with backbone phi and psi angle restraints as predicted from TALOS[3] based on chemical shifts of Cα, Cβ, Hα, Hβ, HN, and CO atoms. Initial structures were obtained using the noeassign macro of CYANA[2], followed by manual refinement using the calibration of peaklists by the CYANA function caliba,[75] with the disulfide bond added as a restraint for the oxidized structure. A final ensemble of 500 structures was calculated, and the 100 conformers with the lowest target function were then selected for water refinement using AMBER. Using implicit solvent (which treats the solvent as a high dielectric continuous medium surrounding the protein near its van der Waals surface so it can interact with the charges on the low dielectric protein), a 500 step energy minimization was first performed on the 100 structures (no restraints) followed by two cycles of 30 ps simulated annealing from 1000K to 0K, with distance and dihedral restraints applied. After the MD simulated annealing, a 2000 step minimization with restraints was applied in implicit solvent and analyzed for distance and dihedral penalties and rank ordered by lowest AMBER energy. The 20 lowest energy

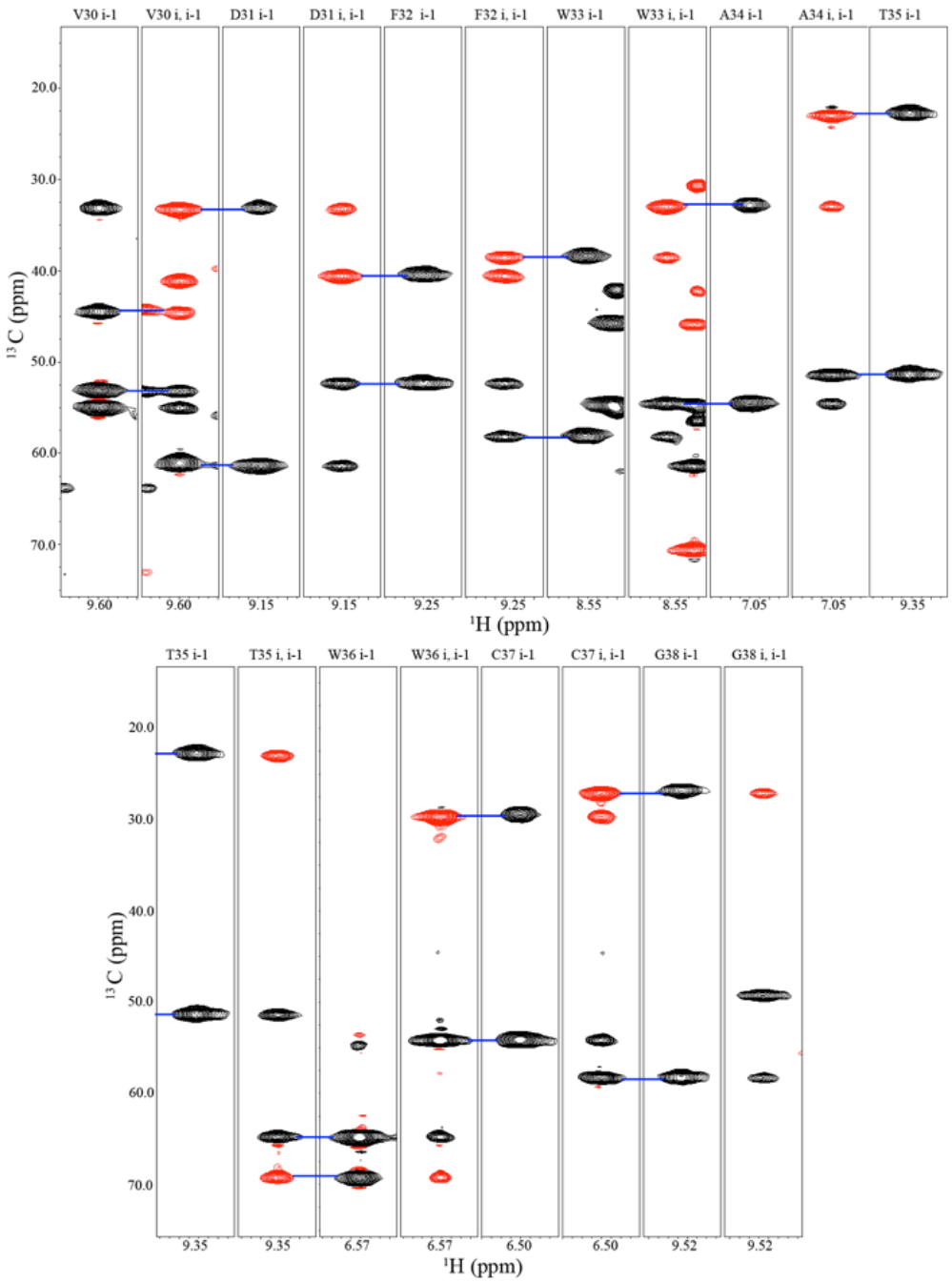


Figure 5.1. Selected strips from CBCA(CO)NH (left strip of residue) and HNCACB (right strip of residue where red peaks are Cβ and black peaks are Cα) spectra of residues 30-38. Blue lines indicate peaks used in sequential assignment of backbone residues.

structures were then selected for the final ensemble, and were subjected to analysis using the PSVS (http://psvs-1_4-dev.nesg.org/) server. NMR spectra were processed using NMRPipe[24] and analyzed using NMRView[25] and Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). In general, processing was performed with a cosine bell window function in the direct and indirect dimensions. All calculations were done using Pere and Condor submission process (NSF OCI-0923037).

5.2.3 Modeling of *M. tb* Thioredoxin Oxidoreductase Activity

The *M. tb* thioredoxin reductase structure (pdb code 2A87)[54] had been solved in the oxidized state. To create a model TrxR of in the reduced state, where the disulfide is converted to surface free thiols, available to react with and reduce thioredoxins, the coordinates of the *E. coli* homolog of TrxR bound to a thioredoxin (pdb code 1F6M)[72] were used as template. The *M. tb* TrxR (oxidized) structure was superimposed on the *E. coli* TrxR (reduced) structure; but, the reduced *E. coli* TrxR is known to have a 66° domain rotation relative to the oxidized TrxR,[54, 72] so this rotation was created in the *M. tb* structure. To do this, the pdb file of *M. tb* TrxR was split at amino acids 117 and 241 so that the FAD and NADPH domains could be separated. The separated NADPH and FAD domain coordinates were then superimposed onto the coordinates of the *E. coli* NADPH and FAD domain using Pymol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.), rejoined and then minimized using AMBER. To model the various oxidation states of the TrxR/TrxC complexes, including the mixed disulfide bond between TrxR and TrxC, structures were minimized with a restrained

disulfide bond between the cysteines (oxidized) or left with no restraint (reduced). All AMBER minimizations were performed with 2000 steps in implicit solvent.

5.2.4 Thioredoxin System NMR Binding Experiments

Concentrated TrxC and TrxR solutions were buffer exchanged in 50 mM potassium phosphate at pH 7.0, to 1 mM each. TrxC was diluted to 320 μM in the absence of DTT, so that it was in the oxidized state. TrxR was then added to the TrxC NMR sample, so that both proteins were at equal concentrations of 250 μM. NADPH was then added to the sample at 1 mM so that TrxR would reduce the disulfide of TrxC. As a control to test for binding in the presence of the different redox states of the cofactor, $NADP^+$ was added to the sample at 1 mM. ^1H-^{15}N HSQC experiments were taken after every step, and pH was measured to ensure that it remains at 7.0. As a control for the reduced state, ^1H-^{15}N HSQC spectra were acquired of TrxC in the oxidized state, and then reduced by DTT at pH 7.0.

5.3 Results and Discussion

5.3.1 NMR Structures of Oxidized and Reduced Thioredoxin C

The solution structures of *M. tb* Thioredoxin C were solved by NMR, with 2036 unique NOE derived restraints and 168 angle restraints in the oxidized state and 2114 NOE restraints in the reduced state with 169 angle restraints. Statistics describing the restraints used in calculation, precision to average coordinates, and quality of the structures are summarized in Table 5.1. The ensemble of structures for oxidized TrxC has a backbone RMSD of 0.54 Å, and 0.50 Å for the reduced conformer (Figure 5.2A-D). To ensure that the cysteines were reduced, 5 mM DTT was added to the protein sample

and chemical shifts were monitored before and after each experiment, using a ^{1}H-^{15}N HSQC (Figure 5.3A-B) spectrum, to ensure there was no disulfide reformation. Further confirmation of the reduction of the active site cysteines is provided by the decrease in the $^{13}C\beta$ chemical shift from 44.6 to 26.9 ppm and 35.1 to 27.6 ppm for Cys37 and Cys40 respectively. The HSQC spectrum of TrxC shows good dispersion of chemical shifts. After addition of DTT, those residues that are affected the most are located along that active site loop, including F32, A34, T35, C37, G38, and C40.

Table 1. NMR and Structure Refinement Statistics for $TrxC_{ox}$ and $TrxC_{red}$		
	$TrxC_{ox}$	$TrxC_{red}$
NMR distance and dihedral constraints		
Distance constraints		
Total NOEs	2036	2114
Intraresidue	207	257
Interresidue		
Sequential ([i-j] =1) Å	509	540
Medium range (1 < [i-j] < 5) Å	510	555
Long range ([i-i] ≥ 5) Å	810	762
Dihedral angle restraints		
ϕ	84	85
ψ	84	84
Water Refinement		
Mean AMBER energy (kcal/mol)	-3943	-3940
Mean violation	19.5	17.4
Distance (>0.3 Å)	0	0
Dihedral angle (>5°)	1	1
Structural Statistics		
Ramachandran statistics (%)		
Most favored regions	92.2	90.5
Additionally allowed	7.8	9.2
Generously allowed	0	0.2
Disallowed	0	0.1
Average pairwise RMSD (Å) 7-110		
Backbone	0.541	0.501
Heavy	0.942	0.875

Table 5.1. NMR and structure refinement statistics for the NMR solution structures of the oxidized and reduced states of *M. tb* TrxC, obtained using the program pdbstat.

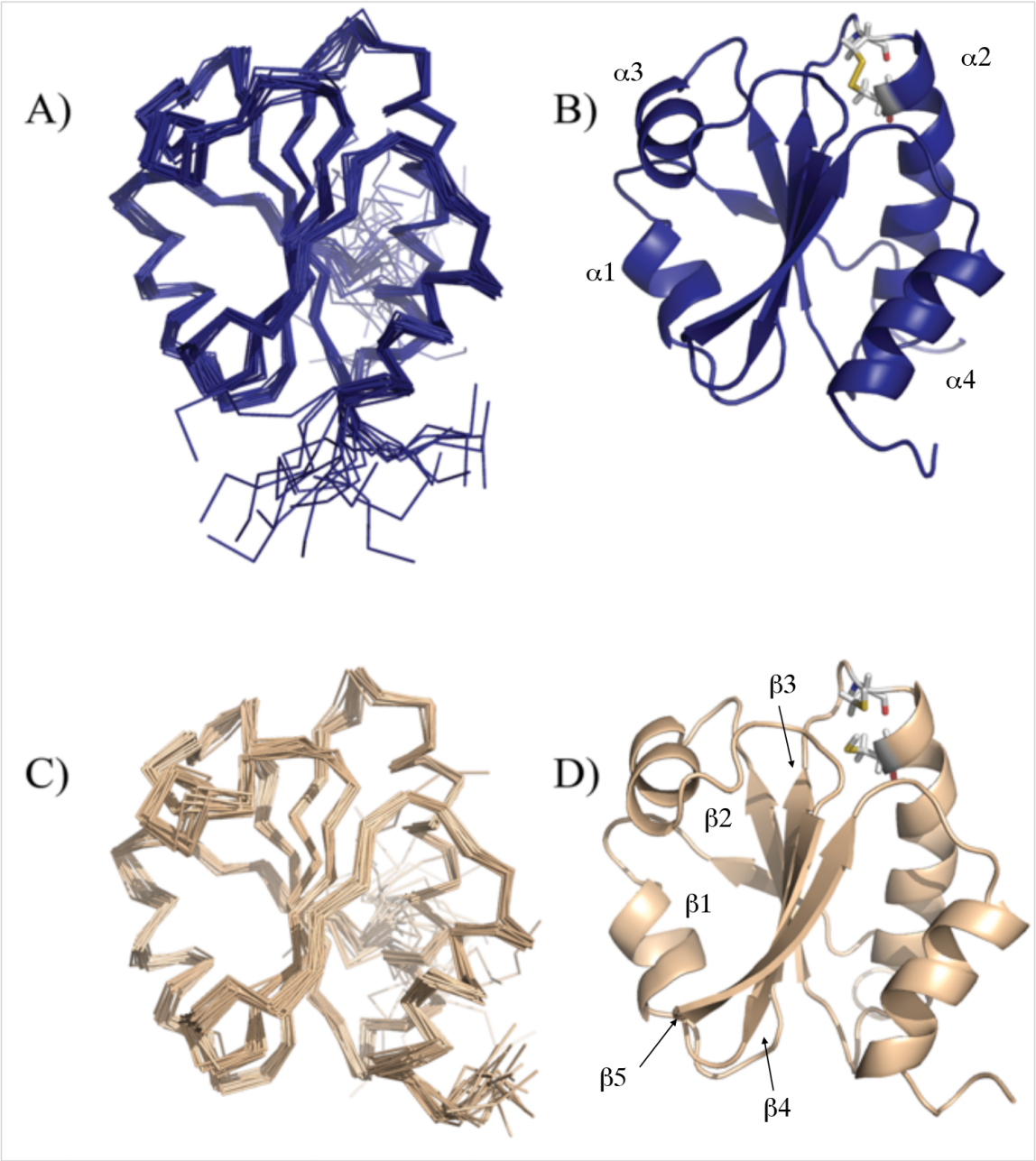


Figure 5.2. Solution NMR structures of A) the ensemble of oxidized TrxC with a backbone RMSD of 0.54 and B) is the median structure for this ensemble with the disulfide of C37 and C40 shown. Panel C shows the ensemble of the reduced state of TrxC with a backbone RMSD of 0.50 and D) is the median structure for this ensemble with the free thiols shown.

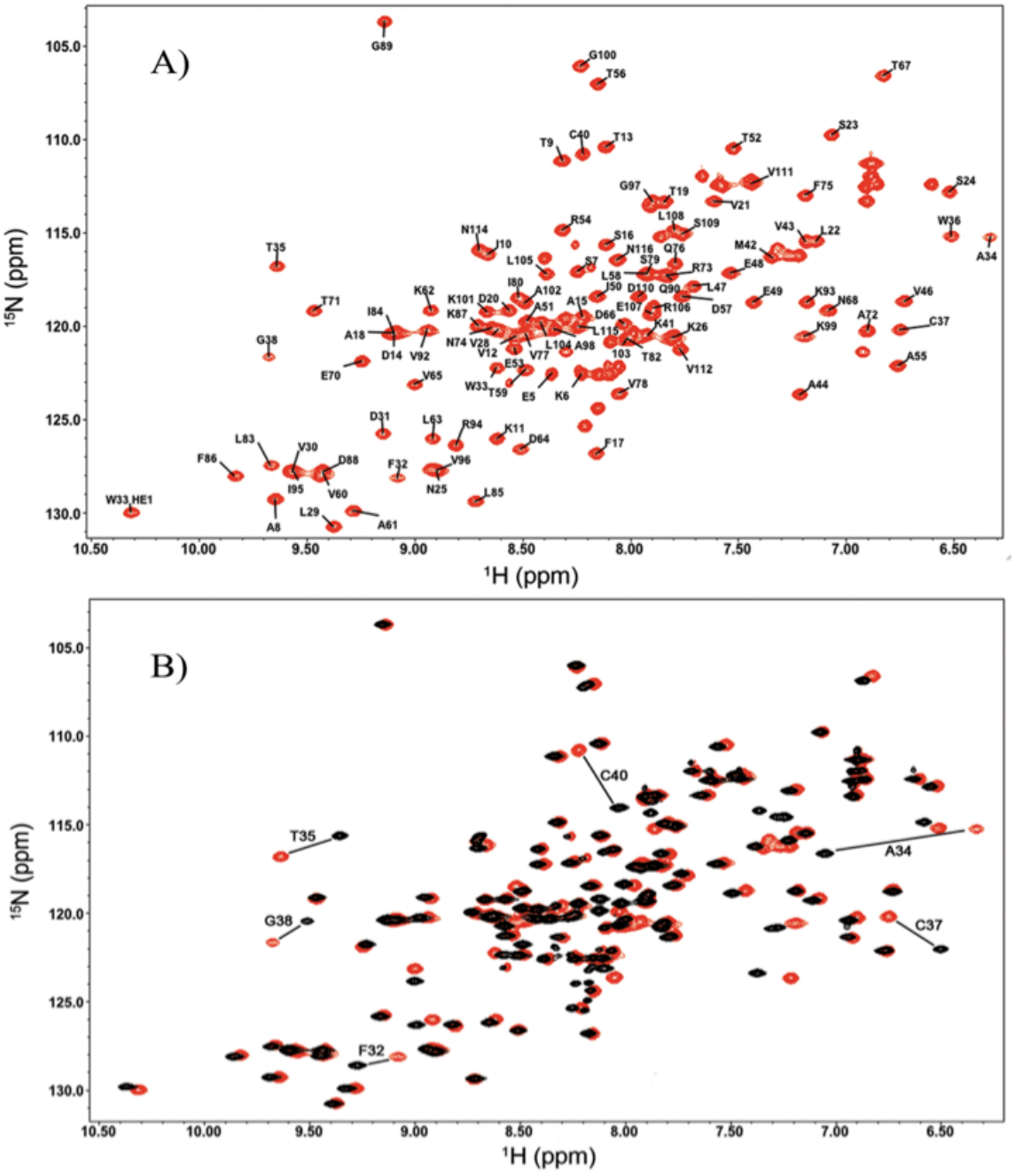


Figure 5.3 2D ^{1}H-^{15}N HSQC spectra of A) the oxidized state of TrxC (red) with all the backbone N-H chemical shifts assigned, and B) the reduced state of TrxC (black) with the large chemical shift changes associated with going from the disulfide to free thiols labeled.

Analogous to the situation with the solution structures of the *E. coli* thioredoxins,[68] the oxidized and reduced forms of TrxC are very similar to each other, showing only small differences between oxidized and reduced states, with a pairwise RMSD of 0.53 (calculated using Pymol) between ensembles (Figure 5.2A-D). Compared to other thioredoxins, TrxC in both forms has the characteristic folding pattern of 4 α-helices surrounding a 5 stranded beta sheet core, consisting of three parallel strands and two antiparallel. The redox active cysteine residues C37 and C40 are located in the conserved active site loop (C_{37}-G-P-C_{40}) that leads into the α2 helix, which in both states is a bent helix. The active site of TrxC includes a predominately hydrophobic core including two Ala (34 and 44), two Pro (39 and 81), an Ile (80), and a Val (43) residue. Both Trp (33 and 36) indole rings are all within 6 Å of the two Cys's, with W36 forming the interface with the thioredoxin reductase. The disulfide in the oxidized state has the typical 90° χ1 angle between the C-S-S-C linkage, and keeps roughly the same χ1 angle going to the free thiols. The most drastic change is the distance of the Cys residue sidechains, with the S-S distance in the oxidized state of the median structure in the ensemble having a distance of 2.0 Å, and 3.5 Å in the reduced state (Figure 5.4).

Although the α1 helices are nearly identical between the two states, it is somewhat of an irregular helix. It is also present in the crystal structure of *M. tb* $TrxC_{ox}$[65] as well both *E. coli* Trx solution structures. The α4 helix spans residues 101-111 however it begins to show disorder at L108, while still keeping its secondary structure. There is some degree of difference between the two states in the random coil and turn region linking the α1 helix to the β3 sheet (22-26), and the α3 helix connecting to the β4 sheet (75-81). The latter region closely contacts the active site region. These regions are

probably dynamic in nature, and are also evident in the *E. coli* solution structures of the different redox states, which resemble the *M. tb* solution structures (Figure 5.5).

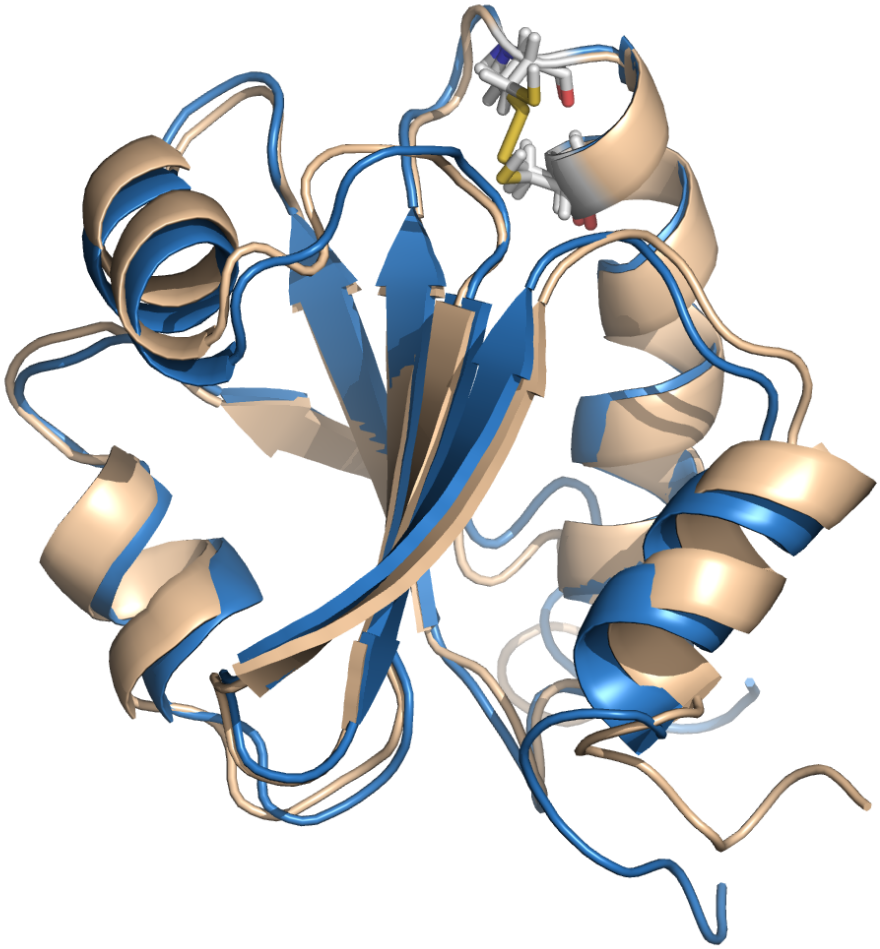


Figure 5.4. The overlay of backbones for the median structures from the solution structures of TrxC in both redox states (with active site cysteines shown), where blue is the oxidized form and wheat is the reduced form. Thiol and disulfide atoms from cysteines 37 and 40 are shown.

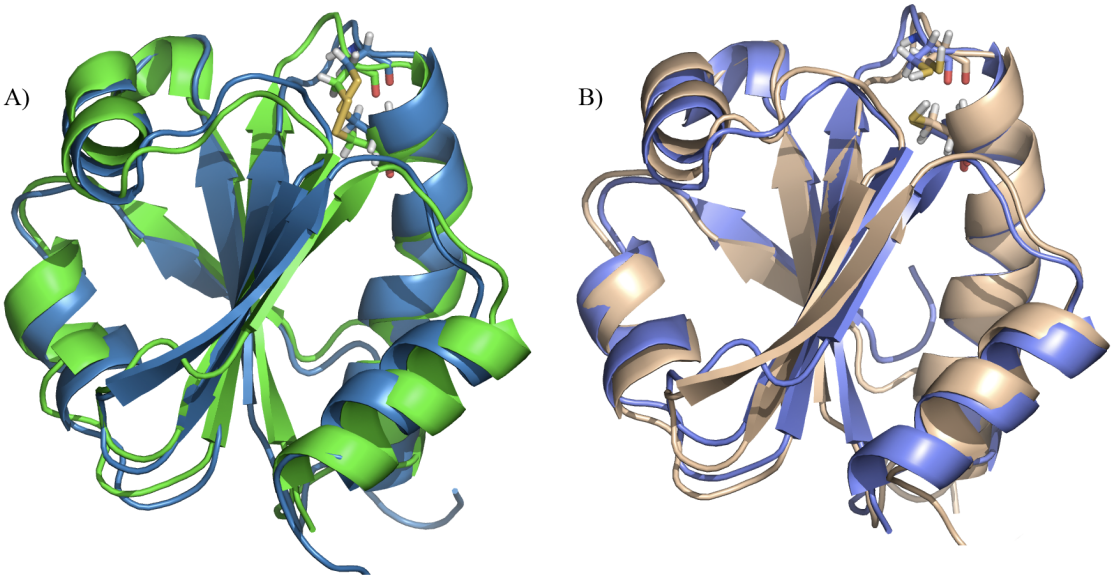


Figure 5.5. Overlays of both redox states of solution structures from *E. coli* Trx and *M. tb* TrxC, where A) is the overlay for the oxidized states, with blue being from *M. tb* and green from *E. coli* and B) is the overlay for the reduced states, with wheat being from *M. tb* and purple from *E. coli*.

5.3.2 Comparison to *M. tb* TrxC Crystal Structure

The *M. tb* TrxC crystal structure (pdb code 2i1u)[65] was solved in the oxidized state, and is nearly identical in secondary structure to our solution structure, with a pairwise rmsd of 0.83 Å between the crystal structure and the median structure from the ensemble (Figure 5.6). The major differences and the source of the high RMSD, is the orientation of the α4 helix. It was noted in the crystal structure study that the protein crystals packed such that the C-terminal residues (110-114) in the α4 helix rested in the active site groove of an adjacent protein in the crystal lattice. Based on our solution structure, this crystal packing seems to have caused the α4 helix to be pushed up 2.7 Å towards the β5 strand, and rotated in towards the β4 strand (Figure 5.6). Residues 108-115 in the crystal structure are disordered, perhaps so that those residues can occupy the

active site neighboring protein. However, residues 108-111 still retain helical shape (as mentioned above) before becoming disordered at residues 112-115 in our solution structure.

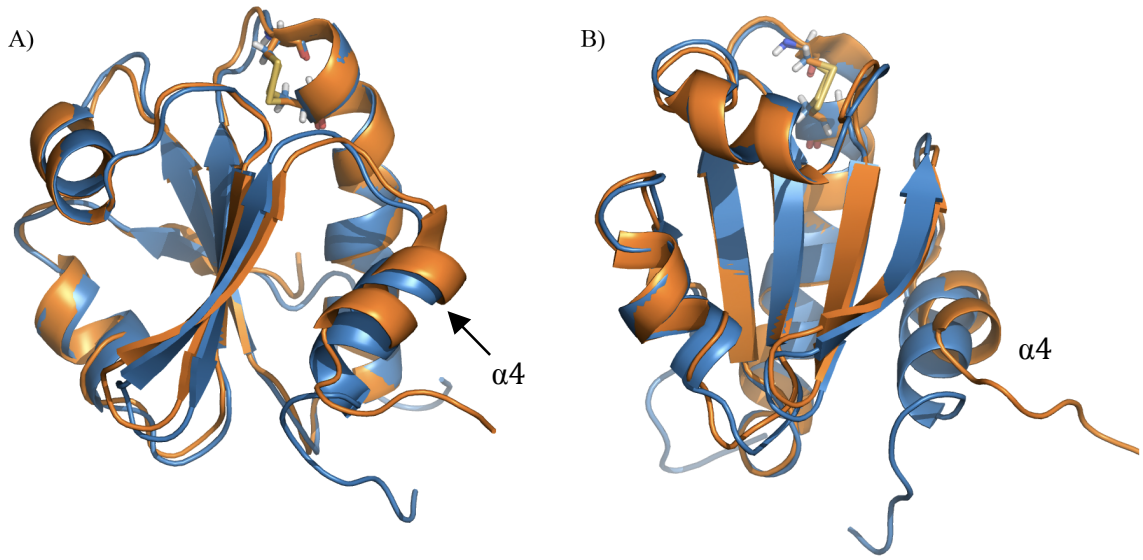


Figure 5.6. Overlay of the median structure of the NMR solution structure of oxidized TrxC (blue) and the crystal structure of oxidized TrxC orange with the active site disulfides shown.

5.3.3. Oxidoreductase Activity of *M. tb* Thioredoxin Reductase and Thioredoxin C

To complete the redox cycle, after having reduced other proteins in the cell, TrxC itself must be regenerated to the reduced state. As previously mentioned, it does this by serving as the substrate of thioredoxin reductase. Figure 5.7 shows our models of TrxR/TrxC in different redox states, depicting the complete redox cycle between TrxR and TrxC, using the *M. tb* TrxR crystal structure[54] and our TrxC structures (oxidized and reduced). It has long been known that bacterial TrxR's undergo a large conformational domain rearrangement, to place the reactive cysteines in a position so that they are able to react with the disulfide on its thioredoxin substrate. The crystal structure of the *M. tb*

TrxR is in the oxidized state (F_O), with it's disulfide buried in the protein interior, near the isoalloxazine ring of FAD so the reduced state of *M. tb* TrxR (F_R), with the cysteines at the surface and the NADPH close to the FAD, was modeled using the *E. coli* thioredoxin reductase structure, which was crystallized in complex with thioredoxin (pdb code 1f6m).[72] This *E. coli* structure was used as a template for proper positioning of the *M. tb* TrxR FAD and NADPH domains. All the steps in the reaction sequence have corresponding structures or models (Figure 5.7), and are based on previous kinetic and structural studies of the *E. coli* system.[76, 77]

As shown in Figure 5.7A the catalytic cycle for the *M. tb* thioredoxin system is shown starting out in a conformation described as the "F_R" state, so that the active site cysteines have just reduced TrxC and there are bound $NADP^+$ and $FADH_2$ cofactors in TrxR. It is called the F_R state because of the relative orientation of the NADPH and FAD domains such that the NADPH binding site is close to the FAD (and the thiol/disulfide is distal). Assuming *M. tb* TrxR behaves as the E. coli reductase, it will then undergo a 66° rotation around the NADPH binding domain to bring the disulfide bonded cysteines (C145 and C148 with a bond length of 2 Å) closer to the isoalloxazine ring of $FADH_2$, while the NADPH binding pocket moves away from the $FADH_2$. This positions the isoalloxazine ring within 3.7 Å of the S atom of Cys 148 (now TrxR is in the "F_O" state), while releasing $NADP^+$. The $FADH_2$ then reduces the disulfide by electron transfer (Figure 5.7B-C), now giving a sulfur inter-atom distance of 3.7 Å between the two cysteine thiols, although giving a slightly smaller $\chi 1$ angle. After reduction, a general base (thought to be D149 which is 5.3 Å from the thiol of C148) abstracts the proton from C148 to produce a thiolate ready for attack to the oxidized cysteines of the

thioredoxin.[77] NADPH then binds to TrxR, and the NADPH domain undergoes another 66° rotation to bring the reduced active site cysteines away from the FAD isoallxazine ring, and back to the surface (back to the F_R state) of the protein in the TrxC binding pocket (Figure 5.7D). At the same time, the NADPH nicotinamide ring is brought to within 3.4 Å of the isoallooxazine ring of FAD so that it can reduce the cofactor back to $FADH_2$. The $FADH_2$ can once again reduce the disulfide of TrxR, later in the catalytic cycle.

The TrxR is now in a conformation that can bind TrxC (Figure 5.7E), where the two proteins come together in a complex that is stabilized by hydrophobic interface produced by the F_R state of TrxR residues F142 and F143 in TrxR, interacting with TrxC residues A34, P39, V43, A44, A72, V78, I80, and P81. Figure 5.8 shows an expanded view of this binding interface, between the two proteins, with surface rendering showing the hydrophobic pocket. Notice how Trp36, which is conserved in all thioredoxins, has its indole ring situated in the middle of this hydrophobic pocket, with the indole ring nitrogen NH hydrogen bonding to the carboxyl group of Asp66, at a distance of 1.9 Å. Such interactions with TrxC's Trp36 are likely to be important for proper positioning of the TrxC C37 for reaction with the TrxR thiol/disulfide (C45,C148). The thiolate of C148 in TrxR is 4.9 Å from the disulfide linked cysteine (C37) of TrxC (Figure 5.7E). Attack of C37 on C148 (disulfide) produces the mixed disulfide between the two proteins (Figure 5.7F), which is analogous to the mixed disulfide structure of E. coli TrxR/Trx that was recently reported. During mixed disulfide formation, a proton is abstracted from C146, producing the thiolate that is capable of forming the disulfide with C148, and releasing the C37 leaving group of the thioredoxin. This ultimately leaves TrxC in the

thiol reduced state (Figure 5.7G). The newly reduced TrxC is released from TrxR, which is in the same state as panel A, completing the catalytic cycle.

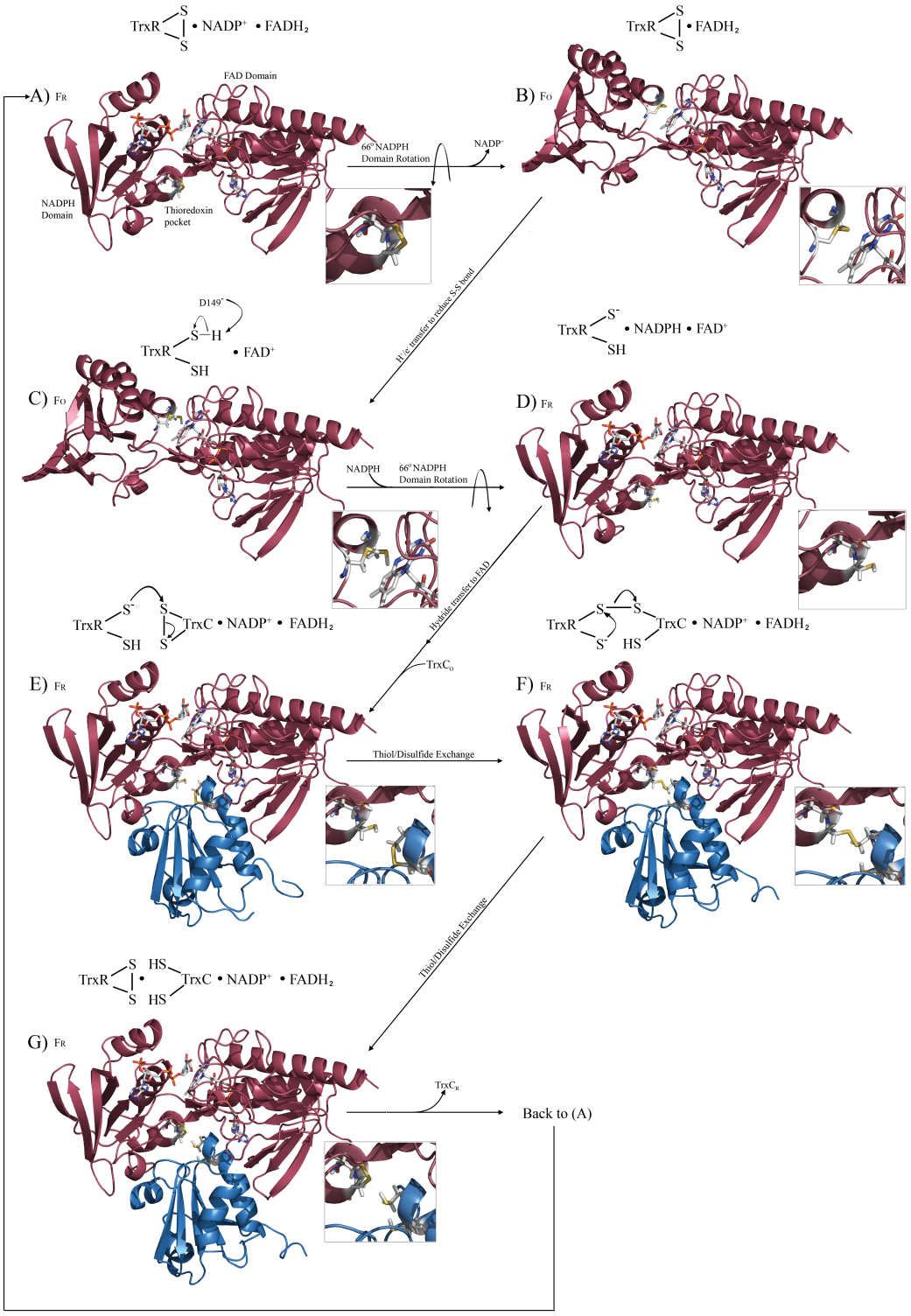


Figure 5.7. The complete redox and catalysis cycle of *M. tb* thioredoxin system of reduced TrxR reducing and oxidized TrxC. F_R and F_O refer to the reduced or oxidized conformations of TrxR, which are related by a 66° domain rotation. TrxC (ox and red) structures are those reported herein, TrxR (F_O) is the previously reported crystal structure, while models for the various complexes were constructed based on structural and functional data for the E. coli thioredoxin system. Each modeled complex was minimized with AMBER.

Figure 5.8. Expansion and surface rendering (white) of the hydrophobic interface formed when TrxR reduces an oxidized TrxC (from Figure 5.7F).

5.3.4 NMR Evidence for Binding Interactions between *M. tb* TrxC and TrxR

It has been proposed that the *E. coli* TrxR, in solution, exists in a dynamic equilibrium between the two conformations (F_O and F_R),[76] no matter what the state of oxidation is for the flavin or cysteines. The hypothesis is that changing redox state or binding complex merely shifts the F_O/F_R equilibrium. While are studies are not directed to testing this hypothesis, NMR binding and chemical shift perturbation measurements can be used for various catalytically competent and dead-end complexes, related to those shown in Figure 5.7. This can establish that binding occurs, and can define the interaction surface on TrxC. To determine what oxidation state the proteins (TrxC and TrxR) must be in order for binding to occur, a series of ^{1}H-^{15}N HSQC experiments were performed so that changes to the state of the proteins in different oxidative states could be monitored (Figure 5.9). Chemical shift changes could be due to binding or to conformational/structural changes induced by redox changes. The TrxC protein was partially reduced, so that we could track the changes that occur upon full reduction and binding, amongst the different redox states in the experiments. When oxidized TrxR in the absence of NADPH, was added to TrxC (Figure 5.9A), there were no chemical shift changes. Interestingly, the small amount of TrxC that was present in a reduced state was converted to a fully oxidized state, shown by the absence of red peaks over the reduced black peaks of TrxC. When the reducing agent (NADPH) for the thioredoxin system was added in excess (4-fold), all of the oxidized TrxC is converted to the reduced state (Figure 5.9B). But, there is no chemical shift changes from this reduced state of TrxC, and that observed in Figure 5.9C when of fully reduced TrxC is produced by adding the reducing agent DTT in 4-fold excess.

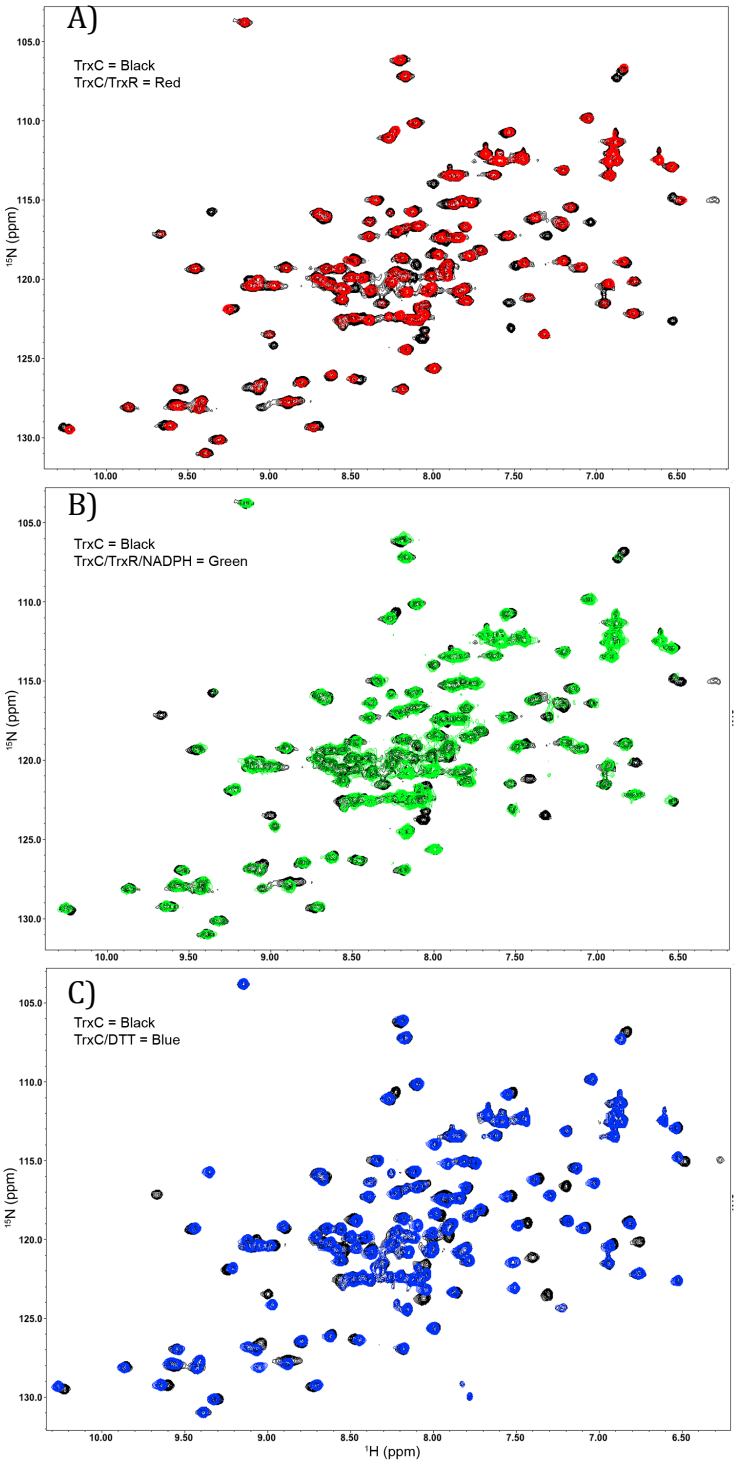


Figure 5.9. 2D ^{1}H-^{15}N HSQC spectra where the black cross peaks in all panels are for oxidized/reduced TrxC and A) TrxR has been added to the sample (red), B) TrxR has been added to the sample and the redox activity initiated by addition of with NADPH to reduce TrxC (green), and C) DTT is added to the sample to fully reduce TrxC as a control experiment (blue). Protein concentration is 250 μM for both TrxC and TrxR.

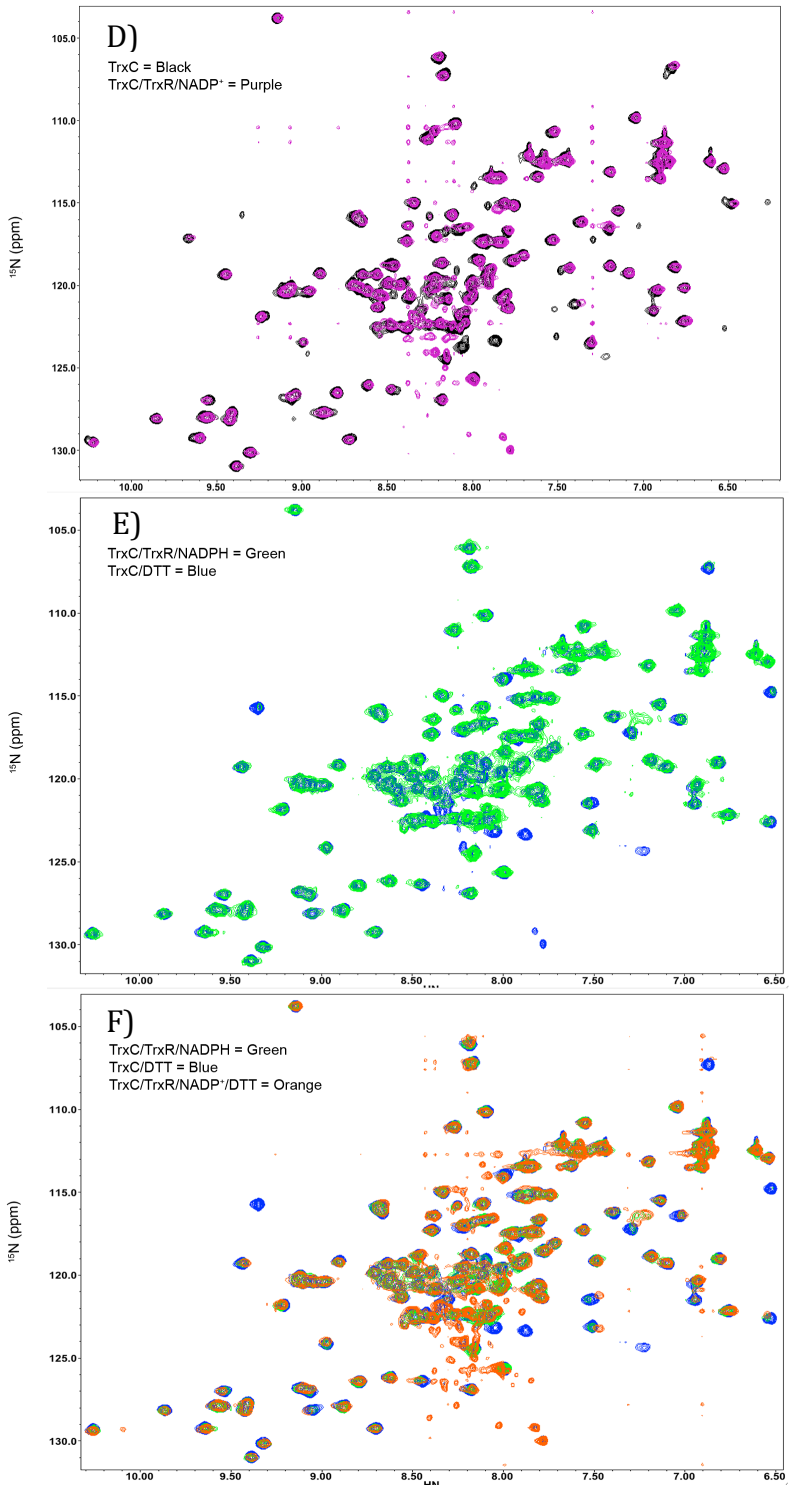


Figure 5.9 continued. 2D ^{1}H-^{15}N HSQC spectra where D) black cross peaks are for oxidized/reduced TrxC and TrxR is added to the sample along with the oxidized cofactor $NADP^{+}$. E) The overlay of the DTT reduced sample (blue) and the NADPH reduced sample (green), and F) is the same as panel E with the oxidized sample from panel D and DTT added to it (orange).

This indicates that while NADPH binds and reduces FAD, and subsequently reduces TrxC, that protein products of this reaction do not have high affinity for each other. In an effort to trap a dead-end complex, $NADP^+$ was added. But, when $NADP^+$ is added to the sample of TrxC/R there appears to be no chemical shift changes from the oxidized state of TrxR (Figure 5.9D).

When the spectra for TrxC that has been reduced using NADPH and DTT are overlaid onto each other (Figure 5.9E), there are no distinguishable chemical shift changes. However there are a few residues that broaden indicating an intermediate exchange process, perhaps due to binding interactions between TrxC and TrxR. To see whether binding was a function of the oxidation state of the flavin, the sample used in Figure 5.9D was reduced with DTT, so that both proteins were reduced, the flavin was still oxidized, and the cofactor $NADP^+$ was still bound. The results from this experiment are shown in Figure 5.9F, and are analogous to the results when both proteins were being reduced using NADPH, there were no distinguishable chemical shift changes, but some residues were exchange broadened. The residues that were in intermediate exchange in the complex with TrxR in Figures 5.9B and 5.9F, were mapped onto the reduced state solution structure of TrxC in Figure 5.10. The majority of the residues that were exchange broadened are in the active site region of the protein (F32, A34, T35, W36, C37, T67, V78), which is consistent with expectations from the modeling of the thioredoxin system in Figure 5.7E and 5.8, where the proteins interacting via a hydrophobic interface. These residues' are still present in Figure 5.9A (with no significant chemical shift change). The dead-end complex has both proteins in the oxidized state, with flavin being oxidized too. Based on these results, although the

protein may be sampling both F_O and F_R states, TrxC will only be bound to the reduced (free thiol) form of TrxR, no matter the state of oxidation of the flavin. And, the identity of the residues that exchange-broaden provide some confirmation of the protein-protein interface suggested in the model (Figure 5.7).

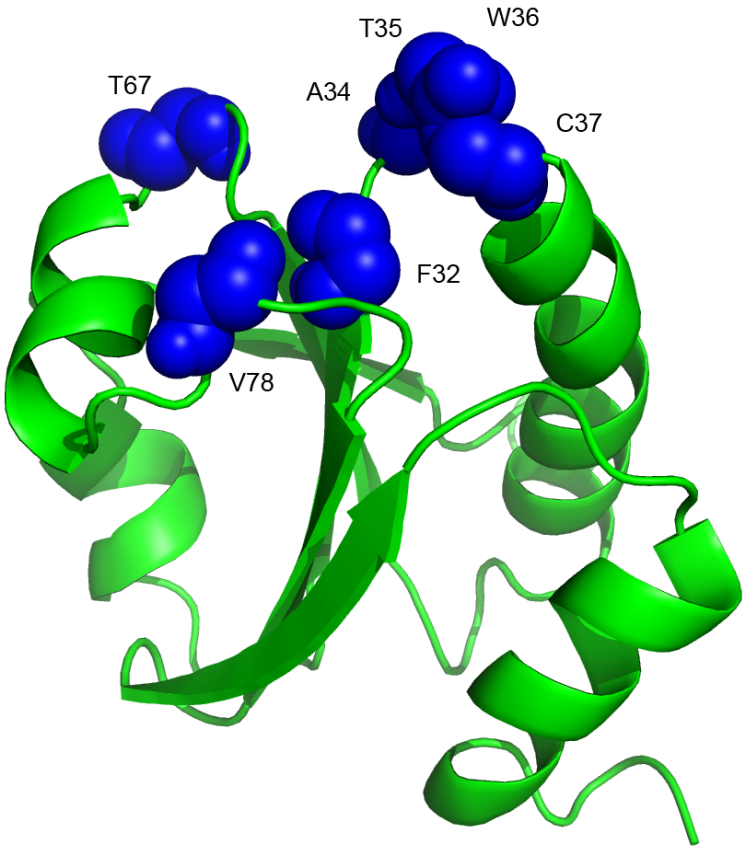


Figure 5.10. Solution structure of reduced TrxC with those chemical shifts that have exchanged broadened upon binding to TrxR mapped on it (blue spheres).

5.4 Conclusion

We have determined the solution structures of both redox states of thioredoxin C from *Mycobacterium tuberculosis* using NMR with high-resolution (backbone RMSD of

0.54 and 0.50 Å for $TrxC_{ox}$ and $TrxC_{red}$, respectively). Like the *E. coli* thioredoxins,[68] the secondary structure and folding of the different redox states of TrxC are nearly identical, with only subtle changes around the active site the protein, and with some dynamic loop regions. In the presence of the thioredoxin reductase from *M. tb*, it appears that TrxC binds to TrxR only when TrxR has a reduced disulfide. This was demonstrated by the disappearance of cross peaks for residues in the active site of TrxC. Using these data, and previous structures of the thioredoxin system from *E. coli* organism,[72] structural models depicting the steps of the oxidoreductase catalytic cycle of the thioredoxin system was developed to facilitate future mechanistic studies of this system, and perhaps aid in design of inhibitors as potential drug leads.

Further studies will include more experiments to define the binding interactions of TrxC and TrxR, with potential point mutations to trap mixed disulfide intermediates the proteins. This can help to determine whether TrxR really needs to be reduced for the proteins to bind to each other. The NMR solution structures of TrxA and B are also in the process of being solved, to provide a complete understanding of the *M. tb* thioredoxin system, and to aid in the drug design process. Inhibitors of the *M. tb* thioredoxin system may also help us to better understand this pathogen's ability to survive intracellular redox stress.

CHAPTER VI. Discovery of Inhibitors of Protein Drug Targets by Automated Docking: PMK, SDF1, and DUSP-5

6.1 Introduction

6.1.1 Docking and Virtual Chemical Libraries

Docking is the process of positioning a ligand into the binding site of a protein, by computationally exploring different translations, orientations, and conformations until a lowest energy structure and "pose" orientation is found. Virtual screening is the process of automated docking of a compound library into a known binding site of a protein, with the goal of discovering inhibitors with high affinity for the protein.[78] This method reduces time, cost, and material of high-throughput ligand screening processes such as random screening and structure based drug design.[79] Virtual screening is used in the early phase of drug discovery, since it can substantially reduce the number of compounds to be screened, so one can focus on the most relevant compounds which will go on to further testing.[80]

Chemical libraries are computational databases of chemical structures often broken down by specific needs (i.e. small fragments or larger chemical structures) and generally there are physical compounds accompanying the library for experimental testing. The first of two chemical libraries used in this study were a database of 2,000 Thai natural products coming from 7 plants from ChemieBase, which is a molecular database of 2D bioactive compounds developed by the Cheminformatic Research Unit at Kasetsart University in Bangkok, Thailand. This database houses approximately 30,000 compounds in 2D structure format and are grouped by the plants from which they come. There were no physical compounds associated with this database but if the virtual screening hits were strong enough, compounds similar in structure or the actual compound (if available) was purchased for testing. The second database is an in-house

physical collection of 10,590 compounds that contain numerous kinds of compound dyes, including sulfonates that had been selected or designed for binding dehydrogenases and kinases.

Automated docking, on average, takes approximately an hour per compound, which becomes very computationally expensive when trying to dock whole computational databases. To increase the speed of virtual screening the databases, we use a workload management system called Condor.[81, 82] Condor essentially provides a job queuing system that distributes jobs to computers, monitors progress, and compiles the data sent to those computers to a central location. Condor was set up on 400 computers throughout campus that are capable of running 400 jobs simultaneously which significantly decreases computational time for virtual screening of the compounds into our drug target proteins: phosphomevalonate kinase (PMK), dual specificity phosphotase-5 (DUSP-5), and stromal cell derived factor 1 (SDF1).

6.1.2 PMK as a Drug Target

Phosphomevalonate kinase (PMK) is a good drug target for such a docking study, as it relates to cardiovascular disease since it is in the mevalonate biosynthetic pathway. This pathway is a mammal's only route to biosynthesis of steroids and isoprenoids, including cholesterol.[26, 83] Much work has been done already on the inhibition of this pathway using statin drugs, which target the HMG-CoA Reductase enzyme.[84] However, PMK offers similar advantages to HMG-CoA Reductase inhibitors, in that the active site has been characterized and a structure is currently available for it (Figure 6.1).[85] In this study both chemical libraries were docked into the whole active site of PMK. Good

binding ligands, as predicted by AutoDock calculations,[12] were then tested experimentally for binding using both NMR and fluorescence-based titrations.

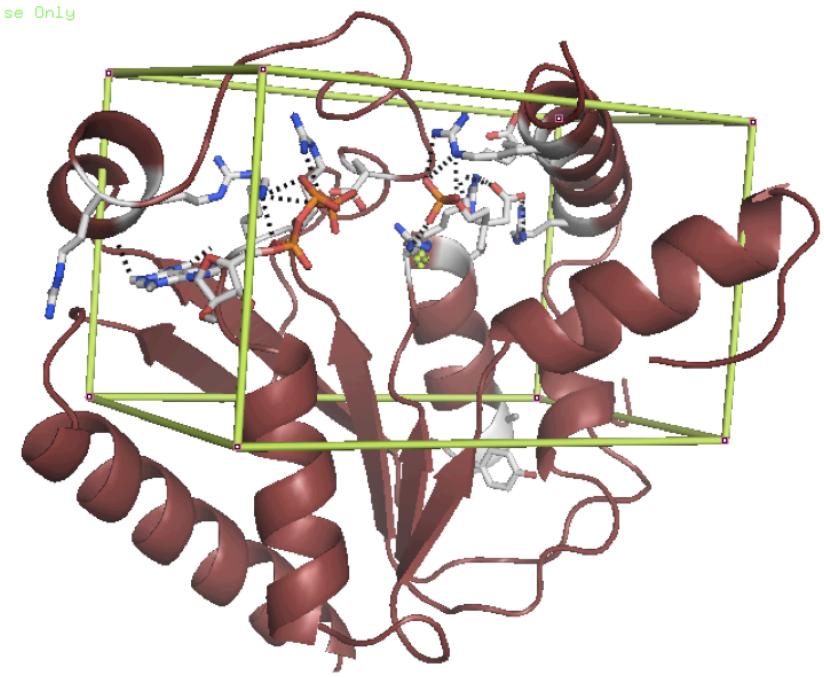


Figure 6.1. Crystal structure of PMK (pdb code 3ch4)[85] with natural substrates, ATP and M5P, docked into the active site. Nearby interactions from ligands to residues are shown in the black and yellow box, which indicates the site that was docked in to using AutoDock.

6.1.3 DUSP-5 as a Drug Target

Dual-specificity phosphatase 5 (DUSP-5) is part of a class of proteins that dephosphorylate phospho-serine/threonine and phospho-tyrosine substrates (DSPs) on proteins called mitogen-activated protein kinases (MAPKs).[86] MAPKs are involved in the mediation of cell signaling that is directly involved in diseases such as cancer, diabetes, and autoimmune disorders.[87-89] MAPKs are fully activated when the tyrosine and threonine in their active loops are both phosphorylated. As such DSPs counteract

such activity and if over expressed or the activity of DSPs were to increase, the cell signaling brought on by MAP would cease. In effect, the phosphorylation state of proteins such as MAPKs acts as an on/off switch for the proteins signaling activity.

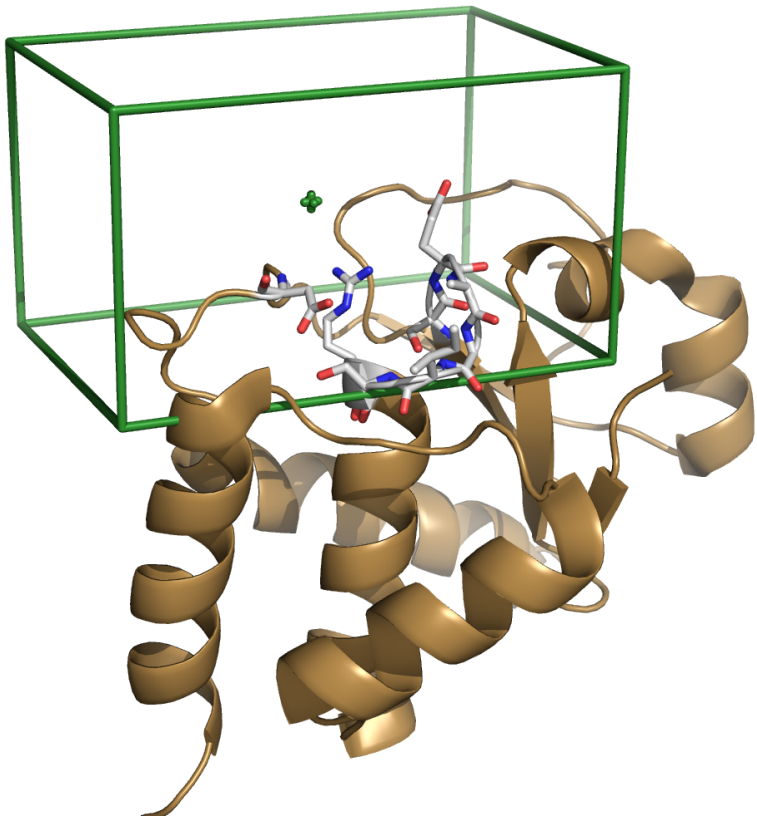


Figure 6.2. Crystal structure of the catalytic domain of DUSP-5 (pdb code 2g6z) is shown with the active p-loop residues displayed. The green box indicates the site that was docked into by the in house library of compounds.

DUSP -5 specifically dephosphrylates a MAPK called extracellular signal regulated kinase (ERK). It has also been shown that over expression of DUSP-5 suppresses the growth of several types of human cancer cells, by "shutting off" ERK (i.e. dephosphorylation).[86] There is also some interest in inhibition of DUSP-5, which would

lead to increased levels of phosphorylated ERK2, leading to angiogenesis (growth of blood vessels). Inhibition of DUSP-5 could, for example, be used to treat retinal diseases. Figure 6.2 shows the crystal structure of the catalytic domain of DUSP-5 as well as the binding site into which we docked.[90] The residues shown are those involved in binding the phosphorylated substrates, located in what is called the p-loop (residues 232, 263-269).[91] The in house library was docked into this site and the top binding compounds with good clustering information were used by our collaborators; Dr. Ramani Ramchandran (Medical College of Wisconsin) in cell-based assays to see whether these compounds inhibited or activated ERK. Of the 10 compounds that were tested, 1 of them showed promise as a potent in inhibitor.

6.1.4 SDF1 as a Drug Target

Stromal cell-derived factor 1 (SDF1) is a cytokine, a small protiein secreted by cells to carry signals to and from other neighboring cells, which in adulthood, plays an important role in angiogenesis (blood vessel growth).[92] Because of this, SDF1 is also important in the initial creation of tumor cells and formation of networks of blood vessels, which leads to tumor progression.[93] Figure 6.3 shows the dimer that SDF1 forms to permit signal transduction, with Arg 47 which is important for binding.[94, 95] The in-house chemical library was docked into the active site of SDF1, and then screened by NMR to verify the binding of compounds. Any compounds that were determined to bind by this screening assay were then sent to our collaborator on this project, Dr. Brian Volkman (Medical College of Wisconsin), to be further tested for binding affinity and inhibition.

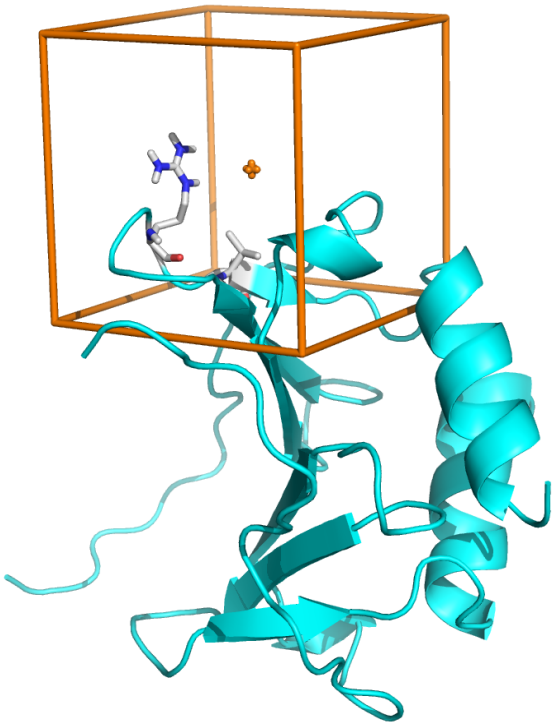


Figure 6.3. Shows one of the structures in the ensemble of solution structures of SDF1 (pdb code 2k05)[94]. Relevant side chains (Arg 47 and Val 49) are shown and the orange box shows the site into which the in-house chemical library was docked.

6.2 Methods and Materials

6.2.1 Ligand Preparation

Ligands were received as 2D MDL MOL files and were converted to 2D SDF MOL files using Babel, as CORINA was not equipped to handle 2D MDL MOL file extensions. The newly converted SDF MOL files were then converted to 3D PDB coordinate files by using CORINA (Molecular Networks, Erlangen, Germany), which generates low energy 3D structures. The files were then processed with the python script prepare_ligand4.py, which comes with AutoDock[12, 96] molecular modeling package. This

script generates a pdbqt file, and adds partial charges to the ligand, sets all torsions in the ligand to active (to permit rotation), as well as merging all non-polar hydrogen's.

6.2.2 Docking Protocol

The protein pdb files and grid maps were all prepared using the AutoDock Tools suite.[96] Grid maps are used in the energy calculations performed by AutoDock. The protein was treated just as the ligands were (done manually on AutoDock), adding partial charges and merging all non-polar hydrogens and saved as a pdbqt file. The 13 different grid maps, one needed for each of the different atoms in the chemical libraries (ex. C, H, F, Cl, etc.) were generated using AutoGrid4[15] with the grid box, the site used to dock the ligands, placed over the active sites of the protein (the colored boxes in Figures 6.1-6.3). Calculations are then limited to the region defined by the grid box.

The docking parameter file (dpf), which contains the parameters that AutoDock uses to dock the ligand into the protein, was prepared using the python script prepare_dpf4.py, and default docking parameters were used, except that 50 separate docking calculations were performed with each calculation consisting of 1,750,000 energy evaluations and an rmsd tolerance set to 2.0 angstroms (to define entry of structure into a given cluster). The dpf files were then automatically docked using MUGrid with Condor submission files and AutoDock4 using the Lamarckian genetic algorithm local search method to perform the optimization of docking pose.[12] The docking results were then clustered on the basis of the root mean square deviation (rmsd) between the coordinates of the atoms in a given ligand, and were ranked on the basis of calculated free energy of binding. The docking log files were then analyzed using the

python script summarize_results4.py contained in the shell script sumresults_4, which rank orders all the dockings, by binding energy. The results were then analyzed to find the best clustered compounds with highest free energy of binding as determined by AutoDock.

6.2.3 Sample Preparation and NMR Spectroscopy

The compounds that were predicted to bind best, based on AutoDock calculations of binding energy were located in our chemical library inventory and dissolved initially into 25 mM in deuterated DMSO and diluted to 5 mM as the working stock solution for experiments other than NMR. ^{15}N-labeled protein (PMK and SDF1) was concentrated to 300 μM for NMR screening, and titration experiments were performed by monitoring chemical shift changes in a 2-D ^{1}H-^{15}N HSQC experiment. For the titration experiments, aliquots of 25 mM of compound were added to the protein in increasing amounts until saturation was reached. For screening experiments, 5 compounds were pooled together to a final compound concentration of 1.25 mM for each compound and added to the protein. If there were any chemical shift changes in the screening experiment, each compound was then tested individually to identify which compound had bound to the protein. This process allows identification of best-binding compounds, with fewer NMR experiments and less protein consumption.

Dissociation constants were obtained by measuring the chemical shift changes in going from free to various bound states, monitoring NMR cross peaks in fast exchange in both ^{1}H and ^{15}N dimensions. These changes were then combined using Equation 2:

$$\Delta shift_{obs} = [(^{1}H\ shift)^2+(^{15}N\ shift/6.51)^2]^{(1/2)} \quad \textbf{Eq. 6.1}$$

The chemical shift change ($\Delta shift_{obs}$) is plotted as a function of the concentration of the ligand, then data were fitted to a quadratic equation, to determine the dissociation constant K_d:

$$\Delta shift_{obs} = (\Delta shift_{max}/(2[P_o]))*\{([L_o] + [P_o] + K_d) - \{([L_o] + [P_o] + K_d)^2 - 4[L_o][P_o]\}^{1/2}\} \quad \textbf{Eq. 6.2}$$

using Sigmaplot8, where L_o is the total ligand concentration, P_o is the total protein concentration, and $\Delta shift_{max}$ is the maximum chemical shift change observed for the particular peak in question. Fitting to the quadratic equation was required because $[P_o] > K_d$. Standard deviations are from the non-linear least squares fitting process. Note that as K_d gets very small relative to $[P_o]$, errors will become larger because K_d become less well-defined in the fitting process).

6.3 Results and Discussion

6.3.1 PMK

Twenty computationally selected compounds were chosen for testing using NMR titration experiments. Ten were from the in house library and ten from the Thai natural product library. Of those twenty, 4 compounds were found to bind to PMK and these were all from the in house library of compounds. Those compounds are labeled as mu_1633, mu_1750, mu_2260, and mu_2419 and shown in Figures 6.4-6.8 and summarized in Table 6.1.

Compound mu_1633 is a substituted anthraquinone flanked by two hydroxyls and two amino-sulfonate toluene groups and is shown in figure 6.4A. AutoDock calculations predicted mu_1633 to bind in the ATP – γ phosphate site (Figure 6.4B) with 25 poses in the lowest binding cluster, and mean calculated binding energy of -10.76 kcal/mol (the lowest in the cluster being -11.66 kcal/mol). Panel C in Figure 6.4 shows all the predicted interactions (less than 3.2 Å) between mu_1633 (the lowest energy docked pose) and the active site of PMK, including many ionic and hydrogen bonding interactions between the sulfonate groups and positively charged residues (R18, K19, R73, and R110). Following the chemical shifts of the residues in panel D of the titration experiment of mu_1633 and PMK and fitting the data as a function of concentration of ligand, gave a dissociation constant (K_d) of <150 μM (since the concentration of protein is so high, accurate K_d determination cannot be made). Following that up with a fluorescence titration (to get more accurate K_d measurement) and monitoring the emission of the tryptophan indole ring, gave the corresponding K_d of 127 ± 10 μM.

Compound mu_1750 is substituted naphthalene with a diazo linkage to a di-carboxylic acid substituted phenyl group (Figure 6.5A). AutoDock found l_cpfm1750 to bind in the ATP – γ phosphate site (Figure 6.5B) with the 8 poses in the lowest binding energy cluster having a mean binding energy of -9.8 kcal/mol (lowest in the cluster being -10.42 kcal/mol). Panel C in Figure 6.5 shows all the predicted interactions (less than 3.2 Å) between mu_1750 (the lowest docked pose) and the active site of PMK. Like the previous compound, the sulfonate groups (and one carboxyl group) show hydrogen bonding interactions with the basic residues R18, K22, and R73, and hydrogen bonds to the amide protons of S20, G21, and K22. Following the chemical shifts of the residues in

panel D in the titration experiment of mu_1750 gave a dissociation constant (K_d) of <40 μM and following the NMR titration with a fluorescence titration gave the corresponding K_d of 3.6 ± 1.0 μM.

Compound mu_2419 is naphthalene with a diazo linkage to a phenyl group with a single branched carboxylic acid figure 6.6A, which differs from mu_1750 by on carboxylic acid group on the phenyl ring. AutoDock found mu_2419 to bind in the ATP – γ phosphate site (Figure 6.6B) with the 2 poses in the lowest binding cluster with a mean binding energy of -10.99 kcal/mol and the lowest in the cluster being -11.02 kcal/mol. Panel C in Figure 6.7 shows the all the interactions (less than 3.2 Å) between mu_419 (the lowest docked pose) and the active site of PMK. The negatively charged sulfonate groups are involved in ionic interactions with R18, K22, and R73, with further hydrogen bonding to amide groups of K22 and G16. Following the chemical shifts of the residues in panel D in the titration experiment of mu_2419 gave a dissociation constant (K_d) of <800 μM and subsequent fluorescence titration gave a K_d of 48 ± 11 μM.

Compound mu_2260 is a tetra-substituted pyrimidine ring with three thio-ether linked acetyl groups and a phenyl ring (Figure 6.7A). AutoDock found mu_2260 to bind in the ATP – γ phosphate site (Figure 6.7B) with the 10 poses in the lowest binding cluster with a mean binding energy of -9.36 kcal/mol (lowest in the cluster being -10.36 kcal/mol). Panel C in Figure 6.7 shows all the predicted interactions (less than 3.2 Å) between mu_2260 (the lowest docked pose) and the active site of PMK. Following the chemical shifts of the residues in panel D of the titration experiment of mu_2260 gave a dissociation constant (K_d) of <20 μM, and subsequent fluorescence titration gave a K_d of 14.6 ± 4.6 μM.

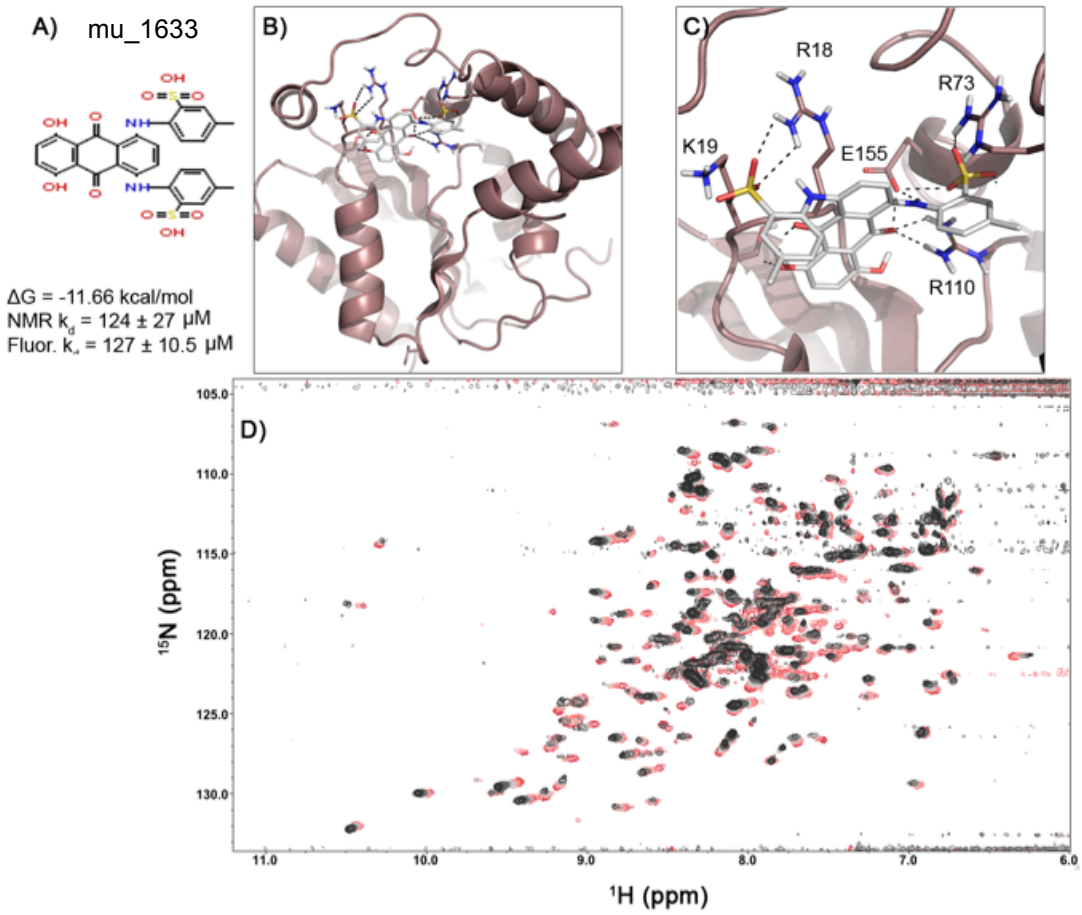


Figure 6.4. Docking and experimental data where A) summarizes the calculated binding energy of the compound and experimental binding affinity (dissociation constants) from NMR and fluorescence titrations. Panels B and C are active site with relevant ionic and hydrogen bonding interactions shown, and residues labeled. Panel D is the overlay of 2D ^{1}H-^{15}N HSQC spectra for the NMR titration, where black is free PMK and the cross peaks going to red are for increasing amounts of compound titrated onto PMK. The chemical shift changes indicate that the compound is binding to the protein, and fitting chemical shift values versus compound concentration provide the K_d for binding.

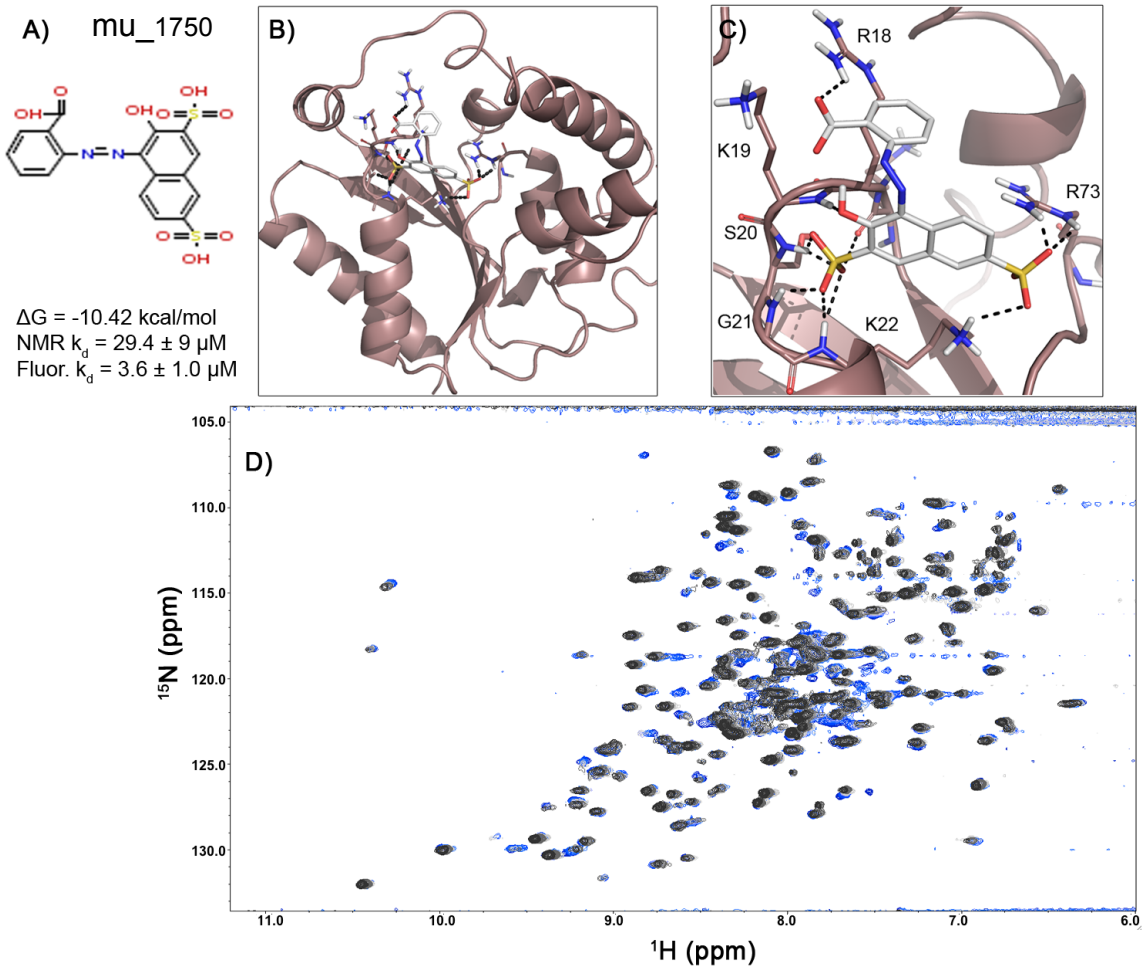


Figure 6.5. Docking and experimental data where A) summarizes the calculated binding energy of the compound and experimental binding affinity (dissociation constants) from NMR and fluorescence titrations. Panels B and C are active site with relevant ionic and hydrogen bonding interactions shown, and residues labeled. Panel D is the overlay of 2D ^{1}H-^{15}N HSQC spectra for the NMR titration, where black is free PMK and the cross peaks going to blue are for increasing amounts of compound titrated onto PMK. The chemical shift changes indicate that the compound is binding to the protein, and fitting chemical shift values versus compound concentration provide the K_d for binding.

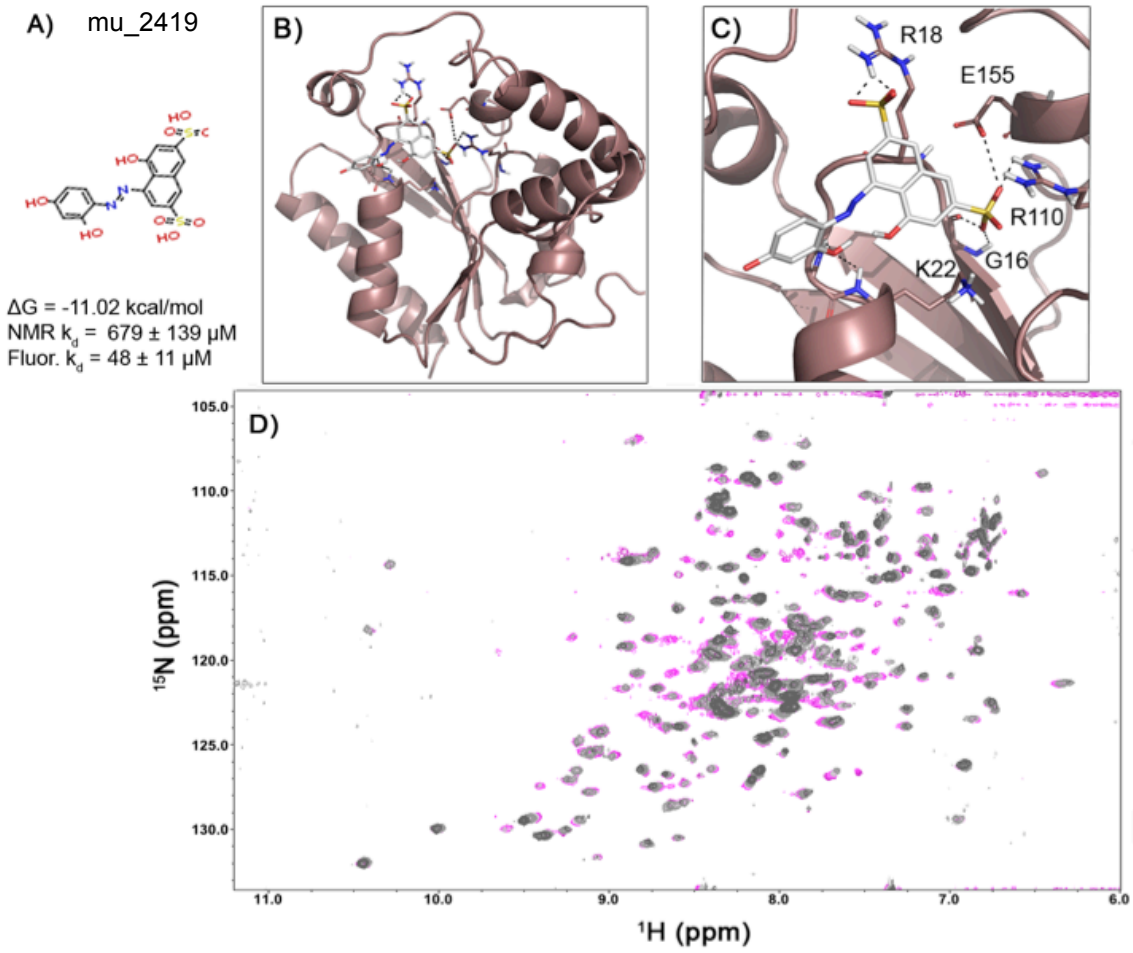


Figure 6.6. Docking and experimental data where A) summarizes the calculated binding energy of the compound and experimental binding affinity (dissociation constants) from NMR and fluorescence titrations. Panels B and C are active site with relevant ionic and hydrogen bonding interactions shown, and residues labeled. Panel D is the overlay of 2D ^{1}H-^{15}N HSQC spectra for the NMR titration, where black is free PMK and the cross peaks going to purple are for increasing amounts of compound titrated onto PMK. The chemical shift changes indicate that the compound is binding to the protein, and fitting chemical shift values versus compound concentration provide the K_d for binding.

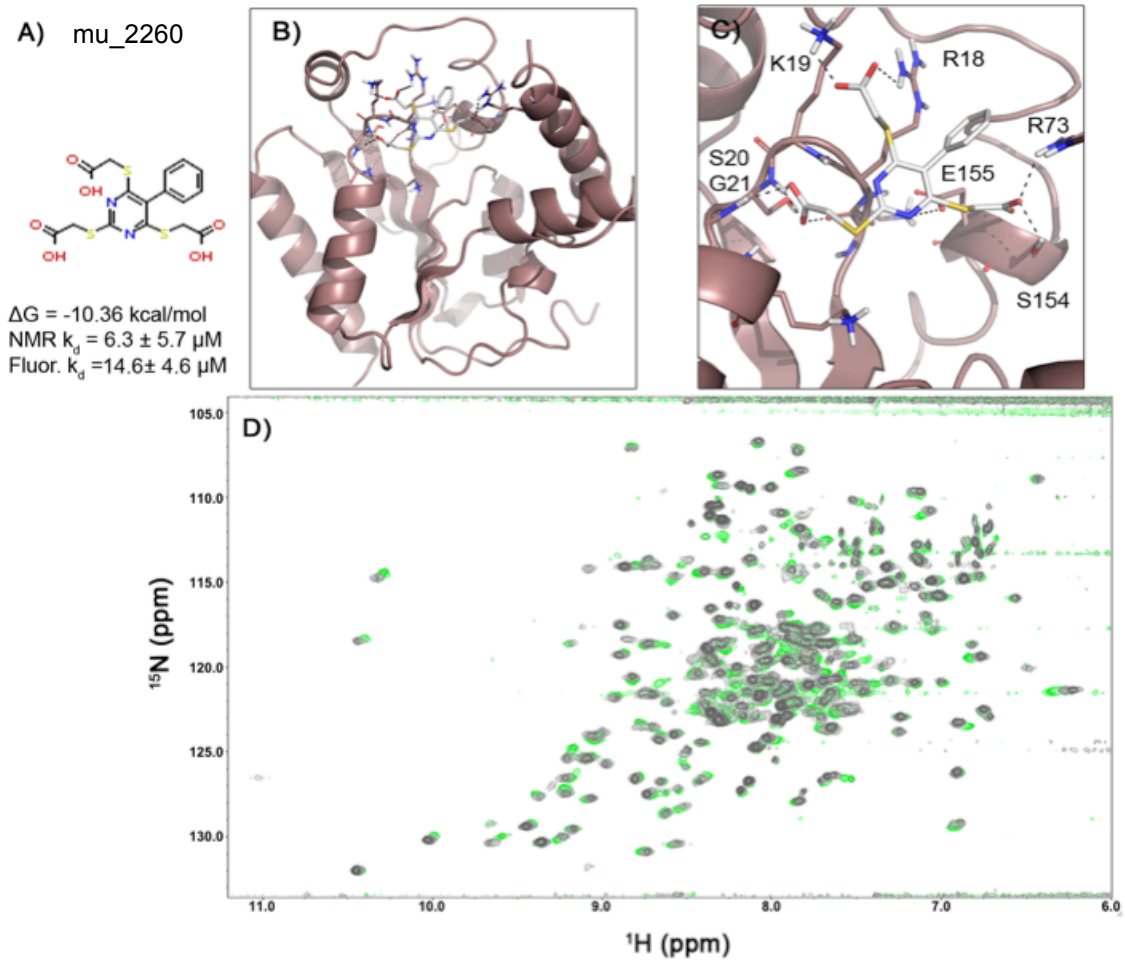


Figure 6.7. Docking and experimental data where A) summarizes the calculated binding energy of the compound and experimental binding affinity (dissociation constants) from NMR and fluorescence titrations. Panels B and C are active site with relevant ionic and hydrogen bonding interactions shown, and residues labeled. Panel D is the overlay of 2D ^{1}H-^{15}N HSQC spectra for the NMR titration, where black is free PMK and the cross peaks going to green are for increasing amounts of compound titrated onto PMK. The chemical shift changes indicate that the compound is binding to the protein, and fitting chemical shift values versus compound concentration provide the K_d for binding.

The four compounds that were found to bind to PMK bind with reasonable affinity all have similar structural qualities, having aromatic rings substituted with negatively charged functional groups. Table 6.1 summarizes the computational and experimental results of the virtual screening process. Compounds l_cpfm1750 and 2260 are the tightest binding of the four, with low micromolar affinity. It is curious that compound l_cpfm 1750 looks very similar to l_cpfm2419, but l_cfpm1750 binds nearly 10 times tighter. This could be attributed to having the negative charged carboxyl group on the phenyl ring of l_cpfm1750 instead of two hydroxyls on l_cfpm2419. These sort of compounds are model compounds for binding to PMK since this protein has many positively charged amino acids (17 arginines and 6 lysines), with the majority of them in the active site, that can neutralize the negative charge on the compounds. While these inhibitors are of value for in vitro studies of PMK, having too much negative charge on a molecule poses problems in vivo such as drug bioavailability, since charged molecules will have difficulty crossing the cell membrane. This will likely represent a significant challenge in any drug discovery effort that targets PMK.

mu_#	Lowest Energy (kcal/mol)	Average Energy in Cluster (kcal/mol)	# in Cluster	NMR K_d (μM)	Fluorescence K_d (μM)
1633	-11.66	-10.76	25	124 ± 27	127 ± 10
1750	-10.42	-9.84	8	29 ± 9	3.6 ± 1.0
2260	-10.36	-9.36	10	6.3 ± 5.7	15 ± 5
2419	-11.02	-10.99	2	679 ± 139	48 ± 11

Table 6.1. Summary of virtual and experimental screening of compounds that were found to bind to PMK, from our in house library of compounds.

6.2.2 SDF1

The in house library of compounds was docked into the Sy-21 site of SDF1 (specifically, in the vicinity of R47) where the signal peptide (sulfotyrosine 21) is known to bind, and mediate dimer formation (Figure 6.3). The top 50 scoring compounds were ranked by calculated binding energy of the lowest energy pose in the docking cluster and were then experimentally screened for binding to SDF1 protein in batches of 5 compounds. If any of the batches showed any chemical shift changes in NMR ^{1}H-^{15}N HSQC then each compound was added individually to SDF1 to identify which compound was indeed binding to SDF1. Since the compounds were dissolved in DMSO and DMSO has an effect on the spectra of SDF1, each batch of compounds was compared to a spectrum of SDF1 with only DMSO, and any chemical shift changes observed were in reference to this spectrum. Of the 50 compounds tested against SDF1, three were found to bind with reasonable affinity, and are shown in Figures 6.8-6.10.

The first compound coming from the batch screening process is mu_1452, which contains phenol sulfonate substituted napthlene rings, linked through diazo groups. Since this is a rather larger molecule (770 Da) there are a lot of torsional degrees of freedom that cause the docking to cluster rather poorly (32 different clusters out of 50 dockings), however the lowest energy cluster (-9.70 kcal/mol with lowest energy pose being -10.7 kcal/mol) is the most populated clustered with 5 poses within an RMSD of 2 Å. As the docking was confined to the Sy-21 binding site[95], the negatively charged sulfonate groups are interacting with the positively charged arginine group, which is important for binding (Figure 6.8A-B). There are also some hydrogen bonding interactions between the hydroxyl groups of the naphthalene and phenol rings and the carbonyl of E15. The

NMR spectra of mu_1452 shows chemical shift changes from green going to red and the binding affinity calculated from these chemical shift changes is < 50 μM (Figures 6.8C-D).

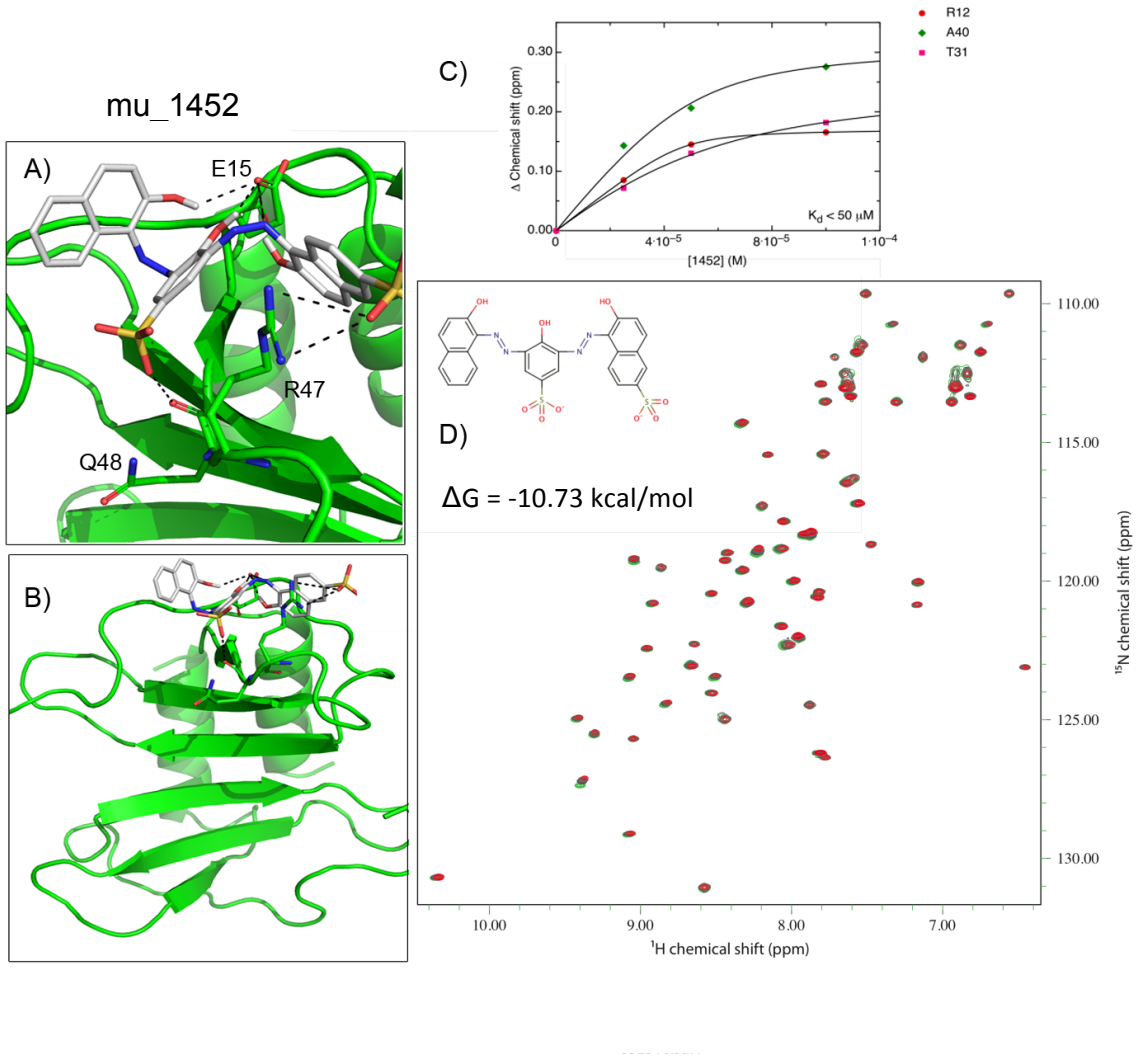


Figure 6.8. Docking and experimental data for the docking of compound l_cpfm1452 with SDF1 where panels A) and B) are the lowest energy pose of l_cpfm1452 and relevant interactions to SDF1. Panel C are the curves generated from the chemical shifts in panel D) to calculate the binding affinity for of the compound bound to SDF1 which was determined to be < 50 μM. Experimental data provided by Josh Ziarek of the Volkman lab at the Medical College of Wisconsin. SDF1 was at 50 μM in pH 6.8 predeutero MES buffer (25 mM).

The second compound identified in the batch screening process was mu_1784, which has an anthracene core with two amide groups stemming from it connected to a trimethyl sulfonate phenyl group. Although this compound is relatively large, there are less torsional degrees of freedom, so 8 clusters out of 50 poses were obtained, with 42 of the poses in the top two clusters (-8.35 kcal/mol and -8.95 kcal/mol for the lowest energy pose), where most of the variation is caused by the rotation of the phenyl rings. This compound's sulfonate groups are also located near R47, with interactions from the sulfonate groups to R47, N47, and D22, with some hydrogen bonding from the amide and hydroxyl groups of mu_1784 to the carbonyl of E15 (Figure 6.9A-B). The binding affinity, as determined from the chemical shifts (Figure 6.9C-D with chemical shifts going from green to red as compound is added) of multiple backbone residues is K_d < 100 μM. Based on which cross peaks perturbed, binding seems to be primarily confined to the Sy-21 site, as opposed to being at the dimer interface, as in the next case.

The last compound that was found to bind to SDF1 is mu_1661, which has sulfonate group attached to a pyrazole ring through a diazo linkage, and is connected to a dichloro phenyl sulfonate group. This compound's docking results had 16 clusters, with poor clustering due to significant rotation through the bonds coming from the pyrazole ring. The lowest energy cluster, with a binding energy of -8.62 kcal/mol (-8.92 kcal/mol for the lowest energy pose), had 18 of the 50 docking poses clustered into it, making it the most probable binding geometry. Many of the interactions between the lowest energy pose (Figure 6.10A-B) are between the sulfonate groups and the positively charged gunaido group of R47, as well as the backbone carbonyl and the amide of the sidechain of Q48. The binding affinity of this compound is slightly less than the previous two

compounds, with a K_d of ~ 200 μM (6.10C-D), however the chemical shift changes upon binding this compound are significantly greater than for the previous two compounds tested, with more of the changes being around the dimer interface. This indicates that this compound may cause dimerization of SDF1, which is known to occur upon binding ligand.

The docking and experimental data provided above suggests that the three compounds do indeed bind with reasonable affinity to SDF1. However, further testing will be needed to determine if these compounds do indeed inhibit this protein. As previously suggested that compound mu_1661 appears to induce dimerization based on chemical shift changes of ^{1}H-^{15}N HSQC spectra for residues in the dimer interface. This is of importance since dimerization of SDF1 causes the protein to cease cell signaling (proliferation of cancer). Since the docking box used for AutoDock calculations was located at this particular binding site, centered on R47, is probably the reason that the compounds identified were comprised of sulfonates and other negatively charged groups. Since the Sy21 binding site is another reason these compounds are so large, if we were to dock in a smaller binding site is quite large, it is not surprising that the compounds identified were also large. It should be noted that during titrations, protein appeared to precipitate. Subsequent SDS PAGE analysis by the Volkman lab indicated that the precipitate contained covalently linked SDF1 protein monomers. It is possible that the diazo groups, associated with many of the compounds, will form free radicals in the presence of light (when bound to protein active sites) and oligomerize the protein to the point in which it eventually precipitates out. Further work is being done to explore this phenomenon.

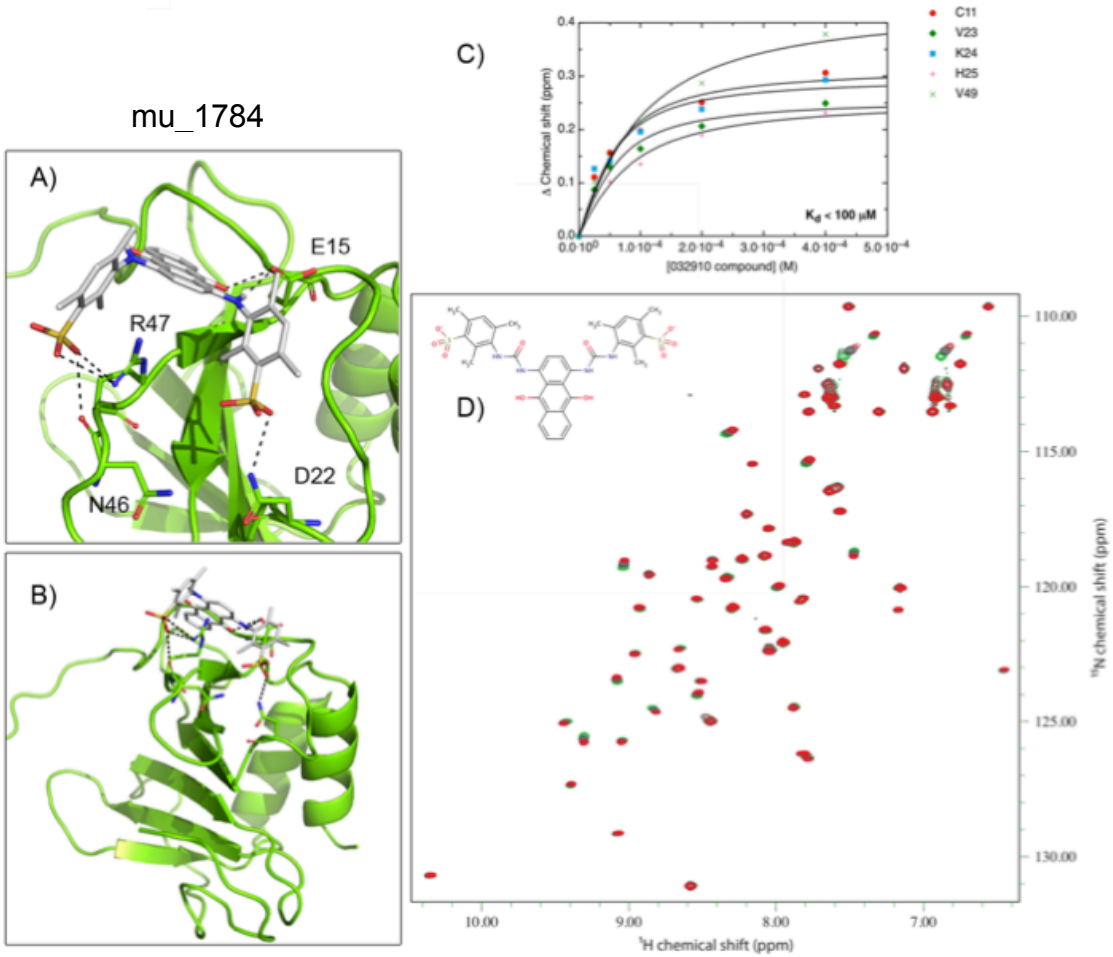


Figure 6.9. Docking and experimental data for the docking of compound l_cpfm1452 with SDF1 where panels A) and B) are the lowest energy pose of l_cpfm1452 and relevant interactions to SDF1. Panel C are the curves generated from the chemical shifts in panel D) to calculate the binding affinity for of the compound bound to SDF1 which was determined to be < 100 μM. Experimental data provided by Josh Ziarek of the Volkman lab at the Medical College of Wisconsin. SDF1 was at 50 μM in pH 6.8 predeutero MES buffer (25 mM).

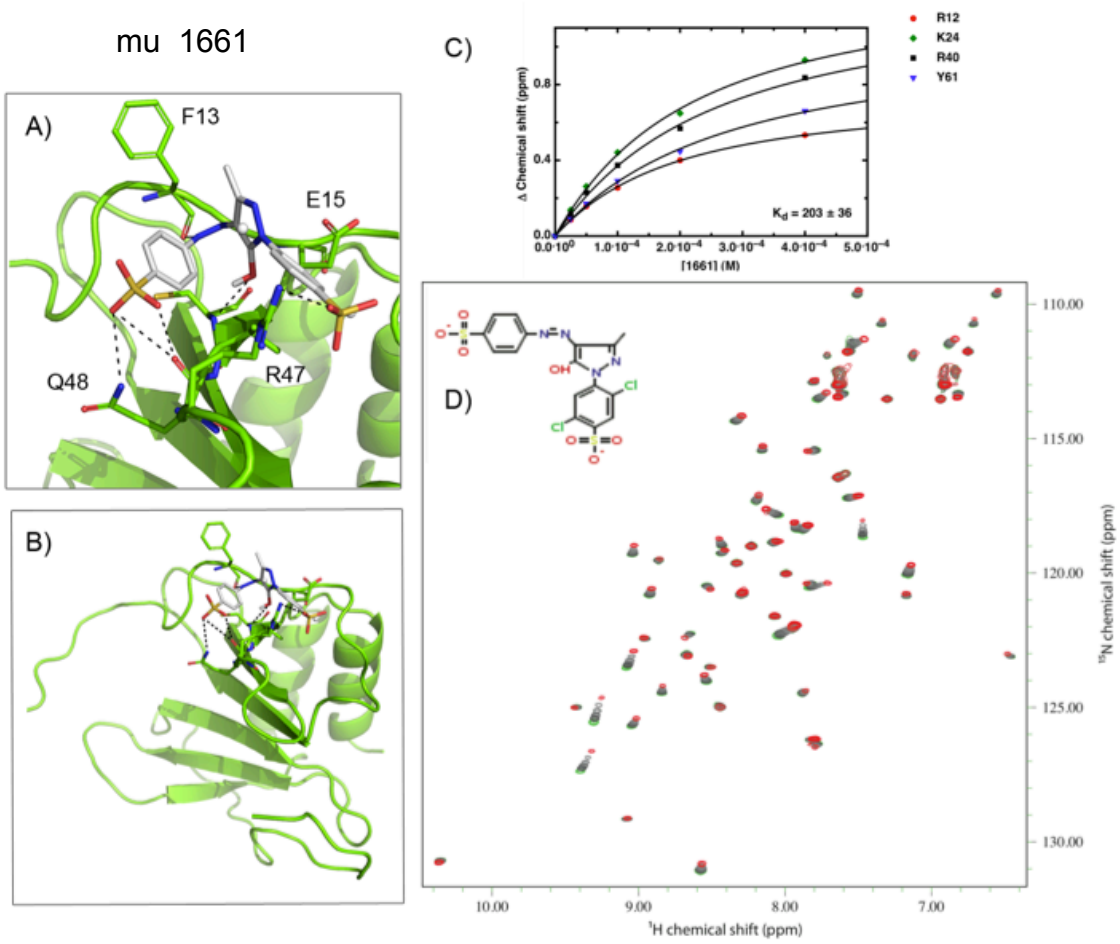


Figure 6.10. Docking and experimental data for the docking of compound l_cpfm1452 with SDF1 where panels A) and B) are the lowest energy pose of l_cpfm1452 and relevant interactions to SDF1. Panel C are the curves generated from the chemical shifts in panel D) to calculate the binding affinity for of the compound bound to SDF1 which was determined to be ~ 200 µM. Experimental data provided by Josh Ziarek of the Volkman lab at the Medical College of Wisconsin. SDF1 was at 50 µM in pH 6.8 predeutero MES buffer (25 mM).

6.2.3 DUSP-5

The in house library was docked into the p-loop active site region of DUSP-5, which is the binding site for dephosphorylation of the ERK protein. The presence of arginines 213,214, and 269 in the DUSP-5 active site make this an ideal binding site for the many sulfonated compounds that we have identified using docking analysis.[91] A select set of compounds from the docking (based on binding energy and clustering) was given to our collaborators (the Ramchandran group from Medical College of Wisconsin). Those compounds were tested directly in cell-based assays, using HUVECS (Human Umbilical Venous Endothelial Cells). Dose response experiments in these cells tests the effect of the compounds on a cell at differing levels of compound, by changes in phosphorylated ERK monitoring at a certain time after dosage. The HUVEC cells were stimulated by the addition of a compound that binds to the cell surface called VEGF receptor, which stimulates growth of blood vessels by stimulating the DUSP-5 catalytic reaction (i.e. dephosphorylating phospho-ERK). The amount of phospho-ERK (pERK) is monitored by a Western blot assay, in which antibodies specifically target pERK and thereby monitor pERK concentratio, when run on an SDS PAGE gel (concentration of pERK is related to intensity of band). If DUSP-5 has been inhibited in some way, after VEGF stimulation, the intensity of the Western blot band (for pERK) should be high. This intensity should decrease with time, and in the case of DUSP-5, the time it takes for the depletion of pERK is 45 minutes. But, if DUSP-5 is inhibited, the band will remain intense. In addition to monitoring pERK levels, ERK and DUSP-5 levels were also monitored. All compounds tested were monitored 45 minutes past stimulation by VEGF

and if the compounds are inhibitory, the relative concentration of pERK should be similar to that observed 2 minutes after VEGF stimulation in the control reaction.

The first of the compounds tested was mu_1472, which is comprised of two substituted naphthalene rings connected via a diazo group. Each naphthalene ring has a sulfonate and hydroxyl group, with one it with one also having an exocyclic amine. Compound mu_1472 has a median calculated binding energy of -8.60 kcal/mol in the lowest energy cluster as determined by AutoDock (-9.22 kcal/mol for the lowest docked compound). Of the 50 poses docked into the active site, 25 poses were clustered in the lowest energy cluster with the remaining 25 dispersed amongst 16 other clusters. As mentioned previously, the negative charged sulfonate groups of mu_1472 (Figure 6.11A-B) are involved in ionic interactions with the positively charged guanido groups of the three arginines in the DUSP-5 active site (R213,R214, and R269), with further hydrogen bonding interactions to H234 and the carbonyl of D232. The dose response curve (i.e. titrations) shows that compound mu_1472 inhibits DUSP-5 most effectively at 6 μM concentration (the amount of phospho-ERK after 45 minutes is nearly the same concentration as at two minutes after stimulation).

The second compound found to inhibit DUSP-5 is mu_1693 and was intended to be a control compound, since it was predicted to bind poorly according to AutoDock (-6.52 kcal/mol). In hindsight, this compound may not have been an appropriate negative control since it has many of the same structural features as mu_1472 and it had been selected in a previous docking study for another protein with a positively charged active site. This compound is rather large for a drug at 840 Da, consisting of a biphenyl bonded to two naphthalene groups through diazo groups. Each naphthalene ring is substituted

with two sulfonate groups as well as a hydroxyl and amine. Due to the large size of this compound and all the possible torsions it has, AutoDock had calculated 41 different clusters for this compound. The lowest energy pose seems to wrap around the outside of the active site as shown in Figure 6.12A-B with interactions between the sulfonates of the compound and R213 and R214, and hydrogen bonding from the hydroxyl to E231 and T235. According to the dose response curve, compound mu_1693 inhibits DUSP-5 most effectively at 3 μM.

The last compound that was found to inhibit DUSP-5 was mu_1842. This compound appears to be a nanomolar inhibitor, and an ideal starting point drug lead optimization project since it is smaller (404 Da), and is comprised of a reasonably drug-like core, a tri-sulfonated carbazole. Since this compound is rigid with fewer torsions, it clustered very well with the majority of the 50 poses clustering into 3 groups, with the lowest energy pose being -8.15 kcal/mol with average cluster binding energy of -7.93 kcal/mol. Note that calculated binding energies tend to scale with the size of the ligand, so given the small size of mu_1842 this is quite good. The lowest energy pose of the lowest energy cluster bound similarly to the previous two compounds with the sulfonate groups, interacting with the three arginines in the active site, with other hydrogen bonding interactions predicted with the sidechain carbonyl of E264 and backbone carbonyl of D232. The dose response curve in HUVEC cells indicates that at 500 nm, there is maximum inhibition of DUSP-5 by mu_1842 with inhibition all the way down to 10 nm. These studies need to be repeated at concentrations of mu_1842 < 10 nm, since pERK levels were still high at 10nm (Figure 6.13C). Interestingly, there appears to be two inflections in the dose response curve with mu_1842, especially based on panel D of

Figure 6.13. Furthermore, all compounds seem to show a loss of potency at higher concentrations, perhaps due to aggregations. These effects will need to be further characterized (ex. aggregation can be determined using nephelometry).

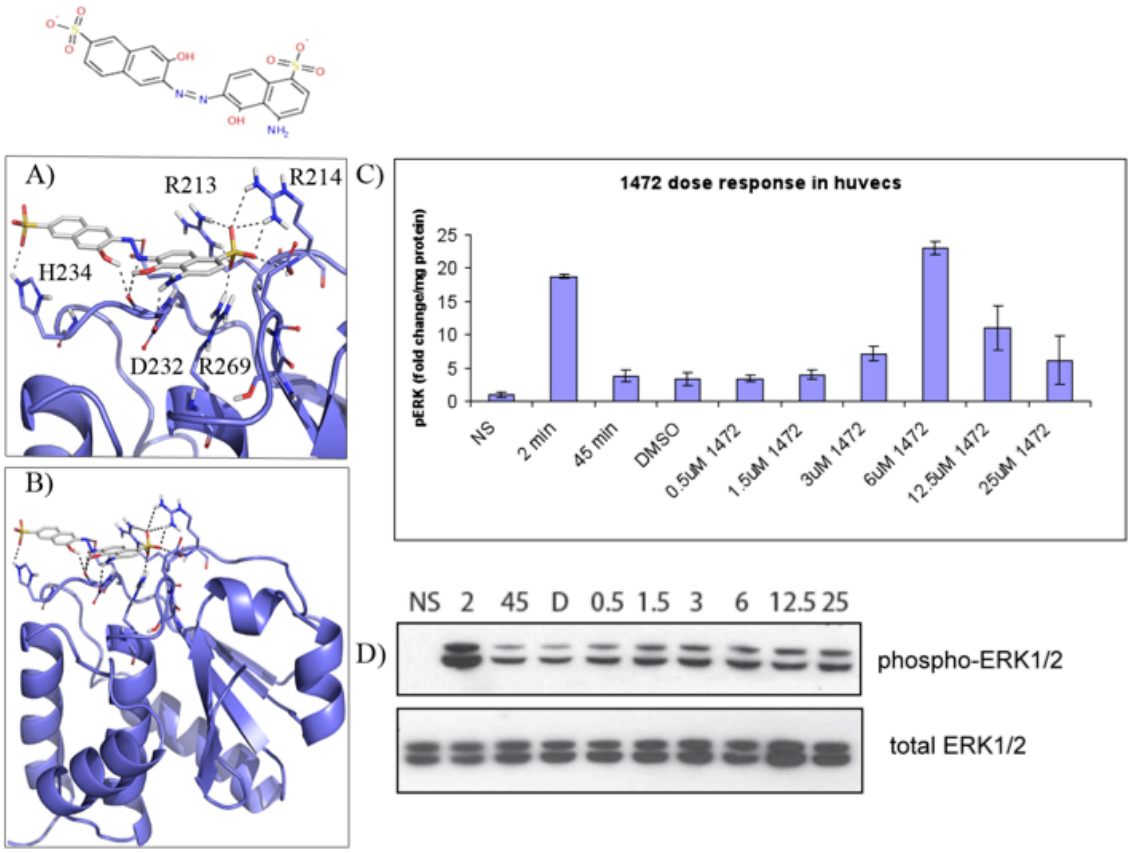


Figure 6.11. Computational and experimental results for the docking of compound mu_1472 into DUSP-5. The lowest energy pose for the docking of compound mu_1472 to DUSP-5 is shown in panels A and B, with relevant residues interactions labeled. Panel C is the dose response results for the compound incubated into HUVEC cells of the specified concentrations with relevant control experimental results shown. Specifically, NS is no VEGF stimulation, while 2 and 45 minutes after VEGF stimulation. Compound was incubated at 0, 0.5,…25 μ, and pERK or total ERK measured at 45 minutes. Panel D is the western blot of pERK and total ERK in all experiments, and corresponds to the data shown in Panel C. Experimental results provided by Indranil Sinha from the Ramchandran group at the Medical College of Wisconsin.

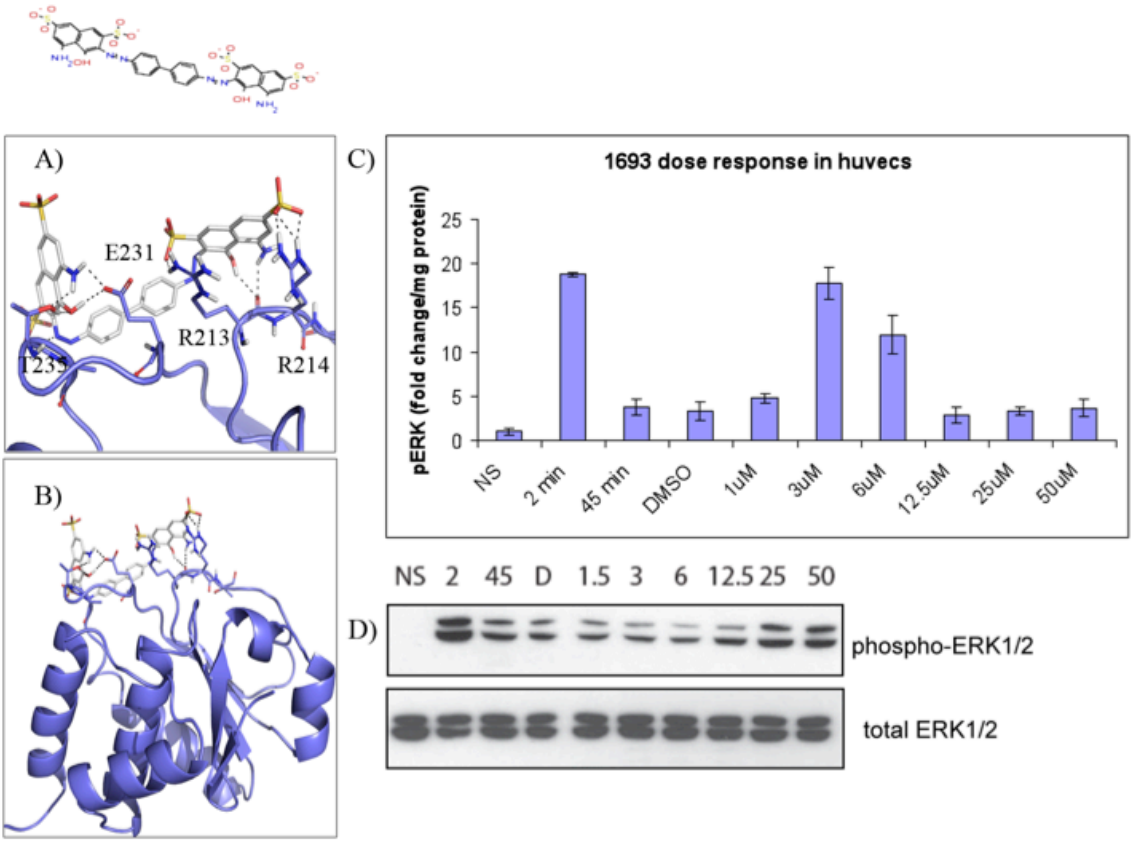


Figure 6.12. Computational and experimental results for the docking of compound mu_1693 into DUSP-5. The lowest energy pose for the docking of compound mu_1693 to DUSP-5 is shown in panels A and B, with relevant residues interactions labeled. Panel C is the dose response results for the compound incubated into HUVEC cells of the specified concentrations with relevant control experimental results shown. Specifically, NS is no VEGF stimulation, while 2 and 45 minutes after VEGF stimulation. Compound was incubated at 0, 1,3,…50 μM, and pERK or total ERK measured at 45 minutes. Panel D is the western blot of pERK and total ERK in all experiments, and corresponds to the data shown in Panel C. Experimental results provided by Indranil Sinha from the Ramchandran group at the Medical College of Wisconsin.

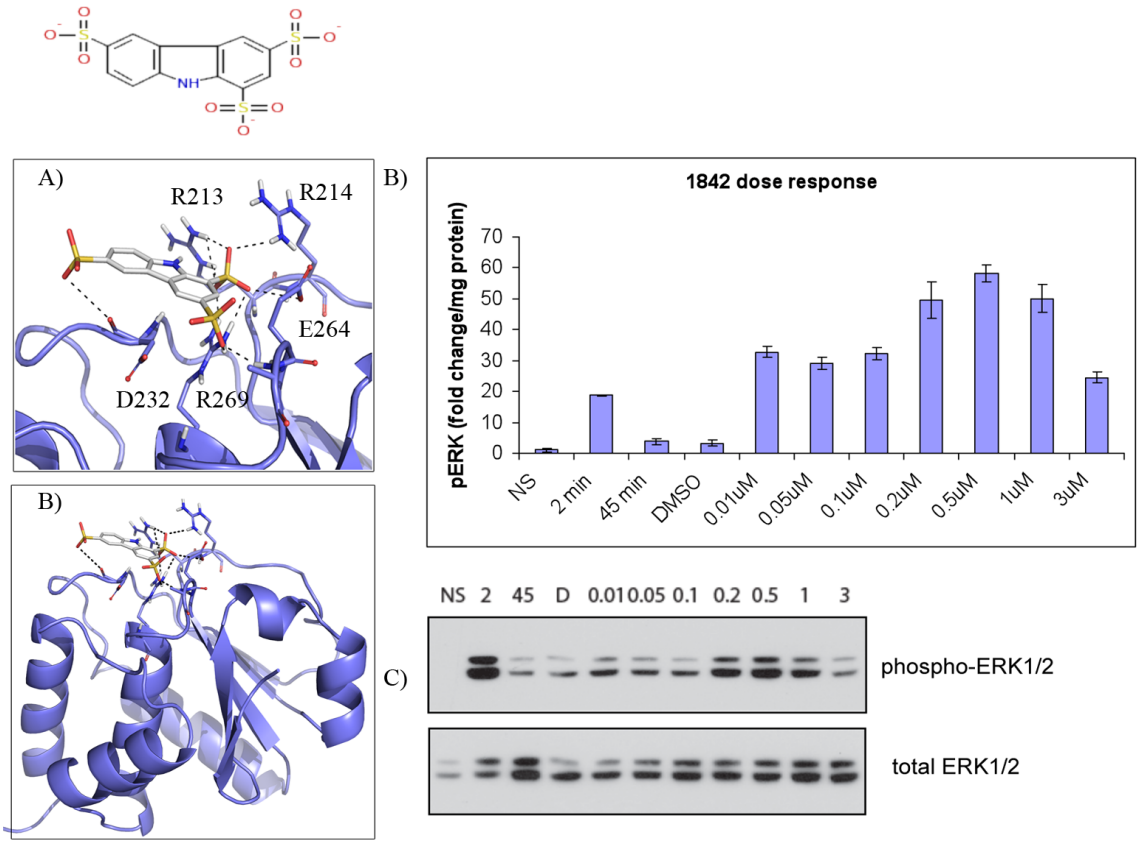


Figure 6.13. Computational and experimental results for the docking of compound mu_1842 into DUSP-5. The lowest energy pose for the docking of compound mu_1842 to DUSP-5 is shown in panels A and B, with relevant residues interactions labeled. Panel C is the dose response results for the compound incubated into HUVEC cells of the specified concentrations with relevant control experimental results shown. Specifically, NS is no VEGF stimulation, while 2 and 45 minutes after VEGF stimulation. Compound was incubated at 0, 0.01,0.05,…3 μ, and pERK or total ERK measured at 45 minutes. Panel D is the western blot of pERK and total ERK in all experiments, and corresponds to the data shown in Panel C. Experimental results provided by Indranil Sinha from the Ramchandran group at the Medical College of Wisconsin.

Unlike the other tests performed on compounds docked into SDF1 and PMK, these cell-based in-vitro assays cannot actually be used to unambiguously conclude that these compounds are binding and inhibiting DUSP-5. Rather these compounds will have to be tested with overexpressed and purified DUSP-5 protein, to study inhibition and binding. These cell-based assays do however provide evidence that the introduction of the compounds to the HUVECs have the desired cellular effect of preventing dephosphorylation of pERK. Compounds mu_1472 and mu_1963 seem to inhibit DUSP-5 with modest IC_{50} values of 6 μM and 3 μM, but are really no comparison to the results obtained by compound mu_1872 which has maximum inhibition at 500 nm with potency at as low as 10 nm. These results are very intriguing and will be investigated further not only with other various in-vitro assays but also with DUSP-5 protein.

6.4 Conclusion

Since the active sites of all the drug targets tested are rich in basic residues it is not surprising that most of the compounds identified using docking are rich in sulfonate groups. This poses other kind of problems in developing these molecules as drugs, since an excess of the negative charge makes it difficult to cross cell membranes. This problem can sometimes be addressed by masking charges as esters, which are then hydrolyzed in the cell. In the case of HUVEC studies, we know the compounds can make it into the cells. The next logical step in developing these compounds as drug leads is to try and optimize the binding of these ligands to the drug targets by changing and adding various functional groups to the compounds until binding potency is in the < 100nm range.[97, 98]

Of all the compounds tested the most promising is compound l_cpfm1842 and its inhibition of DUSP-5. The fact this compound seems to have potency as low as 10 nm is an exciting find and could potentially be very lucrative as a drug lead. From the results provided, virtual screening is an effective to way to find drug leads as we are getting a 10% hit ratio of compounds that bind to the drug targets that have been screened. The use of Condor and the resources of MUGrid have made it possible to virtual screen many compounds (ex. > 10,000) in a reasonable amount of time. The optimization of the virtual screening and experimental verification can lead this project down the road of offering this as a service to those groups (ex. collaborators) who wish find inhibitors and drug leads to their own drug targets.

CHAPTER VII. ^{15}N-Carboxamide NAD(H) Probes to Determine Conformational Binding Dynamics

7.1 Introduction

7.1.1 Rationale for Design of Conformational Dynamics Ligand Probe

Methods for charactering protein dynamics are becoming increasingly well developed, however dynamics from the ligand perspective has not been adequately explored. This is in part because such studies rely on the use ^{13}C or ^{15}N labels, and it has become routine to isotopically label proteins, but not ligands. Dynamic characterization of protein-ligand complexes using NMR has focused mostly on dynamics changes to the protein using a generalized order parameter which measures dynamics on the psec-nsec timescale (local backbone and sidechain dynamics) and relaxation dispersion which measures dynamics on the μsec-msec timescale (protein domain movement).[8, 99-101] In the unbound state, the ligand has a large amount of conformational entropy, which is lost upon binding to a protein, with enthalpic gains offsetting these entropic penalties. In certain proteins however, ligands can retain some flexibility in the bound state such as the matrix metalloproteinase 1 (MMP-1).[102, 103] It has been shown that tightly bound ligands of MMP-1 exist in multiple conformations that exchange slowly on the chemical shift timescale in which two sets of signals arise for the bound ligand.[97]

In an attempt to characterize dynamics changes from the ligand perspective, we have designed a ^{15}N-carboxamide ([^{15}N-CA])-labeled $NAD(P)^+$ and NAD(P)H cofactor (Figure 7.1) to be used as a probe in such studies. NAD(P)H are cofactors that are involved in redox reactions, which participate in hydride transfer to adjacently bound substrates.[104] Oxidoreductases (or dehydrogenases) are the class of enzymes that are responsible for this kind of reaction and are the class proteins of interest in this study. Previous bioinformatic studies from the Sem lab and others have indicated that genomes

are 2-5% oxidoreductases, so probes to study these enzymes could be broadly useful.[105, 106] Of particular interest to us for our initial studies is the protein ornithine monoxygenase (OMO) from *Aspergillus fumigates,*[107] which is an NADP(H) and FAD dependant enzyme that uses NADPH as an electron source and molecular oxygen to hydroxylate the sidechain amine of L-ornithine (Figure 7.2), an amino acid derivative of arginine that plays a role in the urea cycle to dispose of excess nitrogen.

Such dynamics studies are important for studying the mechanism of OMO in that cofactor immobilization/movement is a necessary part of catalysis, to permit electron transfer. Since a number of crystal structures show the nicotinamide displaced from that active site, sometimes showing no electron density for the nicotinamide ring, it is inferred that some domain movement and subsequent binding interactions are needed to bring the nicotinamide ring of the NAD(P)H cofactor into the right position for hydride transfer.[72, 108, 109] Since this process is generally occurring on the μsec-msec timescale, the most useful NMR experiment will be the relaxation dispersion experiment, using the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence.[32, 99, 101, 110] This pulse sequence relies on the measurement of the transverse relaxation rate by measuring the decay of the NMR resonance signal during the CPMG sequence. These NMR relaxation dispersion experiments have been used to study protein folding, protein-ligand interactions, and enzyme function, which makes them ideal experiments to study the conformational dynamics changes that the NAD(P)(H) cofactor appears to undergo upon binding to proteins, such as ornithine monooxygenase (OMO).

Besides studies of cofactor dynamics, the labeled NAD(P)(H) probes being developed can be used to address a structural questions related to oxidoreductase

function. The conformation of the carboxamide of the NAD(P)(H) cofactor when bound to proteins is sometimes ambiguous when determining protein structure with x-ray crystallography, since the nitrogen and oxygen of amides is not easily distinguished at typically achieved resolutions.[105] This was a problem that was recently recognized by the Richardson lab, with regard to carboxamides of asparagine and glutamine sidechains. Based on data compiled form crystal structures, the Sem lab has found that ~90% of the time the carboxamides are oriented so that the carbonyl is in the direction of hydride transfer (Figure 7.5B) with the remaining 10% of structures having the carboxamide in the opposite direction (Figure 7.5C). Since it is difficult to distinguish the orientation of the carboxamide, the question becomes, is the 90% number accurate? Could it be 50% or is it closer to 100%? If it is 100%, why is this geometry preferred? Using this NMR probe we can determine the orientation of the carboxamide when bound to proteins using simple ^{15}N-filtered NOE measurements. Besides addressing basic mechanistic questions about the role of the carboxamide in oxidoreductase function, such studies can also facilitate drug design efforts where inhibitors are rationally designed to bind cofactor binding sites. A detailed understanding of positioning and role of hydrogen bond donors and acceptors is needed to design inhibitors.

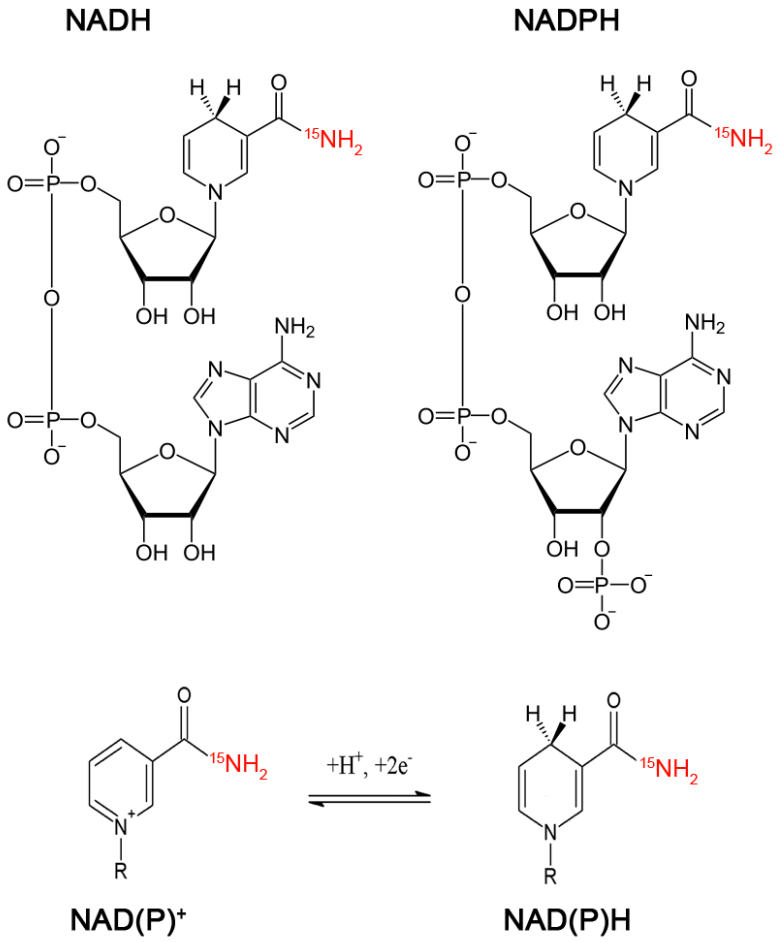


Figure 7.1. Cofactors NADH and NADPH, as well as their oxidized states. NAD(P)$^+$ and NAD(P)H shown with the ^{15}N-labeled carboxamide shown in red.

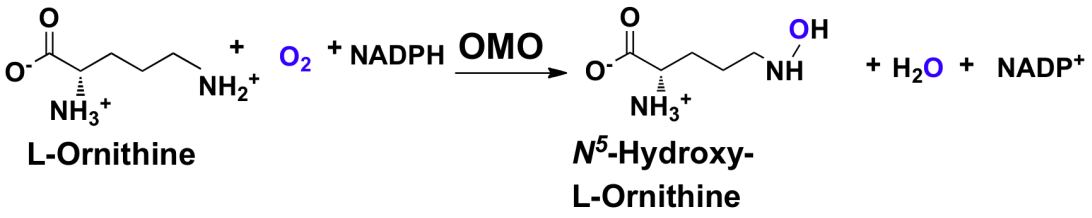


Figure 7.2. Catalytic reaction of ornithine monooxygenase (OMO) in which molecular oxygen is used to hydroxylate the sidechain amine of L-ornithine.

7.2 Materials and Methods

7.2.1 Synthesis of [^{15}N-CA]-nicotinamide

In an ice bath, 4.0 g NaOH and 60 mL deionized water was added to a 250 mL round bottom flask with a magnetic stirrer. When the NaOH dissolved, 1.0 g ^{15}N-NH_4Cl was added. To this solution, 3.9 g of nicotinic acid chloride was added over 20 minutes. The reaction was kept on an ice bath for another 20 minutes, and immediately frozen and water removed by lyophilizing. The product nicotinamide was extracted with hot benzene. Sample purity was tested by TLC using equal parts of DMF/Chloroform as the separating solvent. Product was 0.50 g, with a yield of 14.3%.

7.2.2 Synthesis of [^{15}N-CA]-NAD$^{+111\text{-}116}$

80 mg of [^{15}N-CA]-nicotinamide was dissolved in 10 mL of 40 mM potassium phosphate and brought to a pH of 7.0. 200 mg of NAD^+ was dissolved in 15 mL of 40 mM potassium phosphate buffer and brought to a pH of 7.0. In the same solution, 500-700 mg of the acetone dried powered form of porcine brain NADase (purchased from Sigma Aldrich) was added and pH brought to 7.0. For control experiments, 1 mL of NAD^+ solution and 1 mL of NAD^+/NADase solution were added to eppendorf tubes. Both solutions, [^{15}N-CA]-nicotinamide and NAD^+/NADase, were added together were incubated at 37°C, as were control reactions. The reaction mixture was tested periodically using the cyanide method, where 20 µL of reaction was added to 980 µL of a 1 M KCN solution and monitored spectrophotometrically at 326 nm with the same molar extinction coefficient as NADH (6220 L/mol-cm), to determine if the NADase has exchanged all the unlabeled for labeled nicotinamide rings of NAD^+, to produce [^{15}N-

CA]-NAD^+. In the assay, CN- attacks NAD+ to produce an adduct that absorbs at 326 nm. In the control reaction that has no nicotinamide, the NADase enzyme simply hydrolyzes the nicotinamide of NAD, rather than swapping the ring, (so absorbance at 326 nm decreases). After the reaction was complete (the control reaction lacking free dropped to base line levels, while the test reaction maintained similar absorbance levels as the NAD^+ control reaction), the reaction mixture was heated up to 70°C for 2 min to precipitate the protein, and spun down for 10 minutes at 4000 RPM.

The **[**^{15}N-CA]-NAD^+ was purified by using a BioGel P2 size exclusion column (100 ml bed volume) and washed with water to separate the **[**^{15}N-CA]-NAD^+ from all the free nicotinamide. Fractions were tested using TLC, and the same separation solvent as before and the fractions containing only [^{15}N-CA]-NAD^+ were pooled together. [^{15}N-CA]-NAD^+ concentration was determined in an enzymatic endpoint assay. An aliquot of this solution was mixed 20 mM of ethanol and 20 mM semicarbazide and 100 units of yeast alcohol dehyrogenase at pH 7.3. Absorbance increase was monitored at 340 nm, with the molar extinction coefficient of 6220 L/mol-cm. If **[**^{15}N-CA]-NAD^+ was the desired product, the solution was then flash frozen and lyophilized overnight. A 2D ^{1}H-^{15}N HSQC spectrum was acquired to confirm the ^{15}N labeling of the cofactor.

7.2.3 Synthesis of [^{15}N-CA]-NADH

[^{15}N-CA]-NADH was prepared by reduction of [^{15}N-CA]-NAD^+ using yeast alcohol dehydrogenase (ADH) and ethanol (Figure 7.3). The reaction was run at room temperature in the presence of excess alcohol (20 mM) in pH 7.3, 100 mM phosphate (KPi) buffer and monitored at 340 nm until reaction was complete. After the reaction was

done the ADH enzyme was denatured by heating at 70°C for two minutes, and then removed by filtration using an Amicon Ultra filtration device (Millipore, cutoff 10 kDa). The solution was then flash frozen in liquid nitrogen and lyophilized over night.

7.2.4 NMR Study of [^{15}N-CA]-NAD$^+$ in Complex with Ornithine Monooxygenase

OMO was obtained from the DuBois lab at Purdue University and concentrated to 600 μM in 200 mM KCl and 100 mM KPi buffer at pH 7.8. To this, 500 μM of [^{15}N-CA]-NAD$^+$ was added, and pH measured to be 7.5. To determine the orientation of the carboxamide on the nicotinamide ring, a 2D ^{15}N-filtered [^{1}H-^{1}H] NOESY experiment, was run with a mixing time of 150 ms. To determine possible conformational exchange, CPMG measurements were taken with 10 measurements, with a total delay of 30 μsec. The intensities of the NH protons were measured using the VNMRJ program, and fitted to Eq. 7.1, to obtain $R_2 = 1/T_2$, where I is intensity of a given measurement, I_o is the control intensity and τ_{cp} is the delay that surrounds the 180° pulse in the CPMG sequence. R_2 and τ_{cp} was fit using GraphPad Prism to the fast exchange equation (Eq. 7.2)[99],

$$I = I_o * e^{(-\tau cp/T2)} \quad \textbf{Eq. 7.1}$$

$$R_2 = R_2^0 + R_{ex} * (1-2\tanh(k_{ex}*\tau_{cp}/2)/(k_{ex}*\tau_{cp})),\ R_{ex} = p1p2\Delta\omega^2/k_{ex} \quad \textbf{Eq. 7.2}$$

$$R_{ex} = (p1p2*\Delta\omega)/k_{ex} \quad \textbf{Eq. 7.3}$$

where p1 and p2 are the relative site populations (concentration) of the two states that are in exchange, $\Delta\omega$ is the chemical shift difference between the two states (s^{-1}), and k_{ex} is the exchange rate constant for the two states. R_2^o represents the transfer relaxation rate, in the absence of exchange between states.

7.3 Results and Discussion

7.3.1 Preparation and Analysis of [^{15}N-CA]-NAD(H)

[^{15}N-CA]-NAD(H) was synthesized according to Figure 7.3 and characterized with UV-Vis spectroscopy to determine concentration and viability as a cofactor with the ADH enzymatic endpoint assay, and NMR to ensure ^{15}N labeling. Two crosspeaks are observed at 5°C in a 2D ^{1}H-^{15}N HSQC experiment, one for each proton on the carboxamide of NAD^+ (Figure 7.4). Analysis of a 1D slice ^{1}H-^{15}N HSQC experiment reveals that the crosspeak at 8.6 ppm (*anti* N-H) has 2/3 of the intensity of the crosspeak at 7.8 ppm (*syn* N-H), and previous studies showed that at higher temperature and pH, the crosspeak at 8.6 ppm significantly line broadens due to exchange with water because of the relatively higher acidity of this anti NH proton (however binding to a protein decreases the exchange, and permits the anti N-H to be observable). For this reason it was necessary to run all the experiments at 5°C and at neutral or slightly acidic pH, to decrease proton exchange with water, while keeping the necessary buffer requirements of the protein in mind.

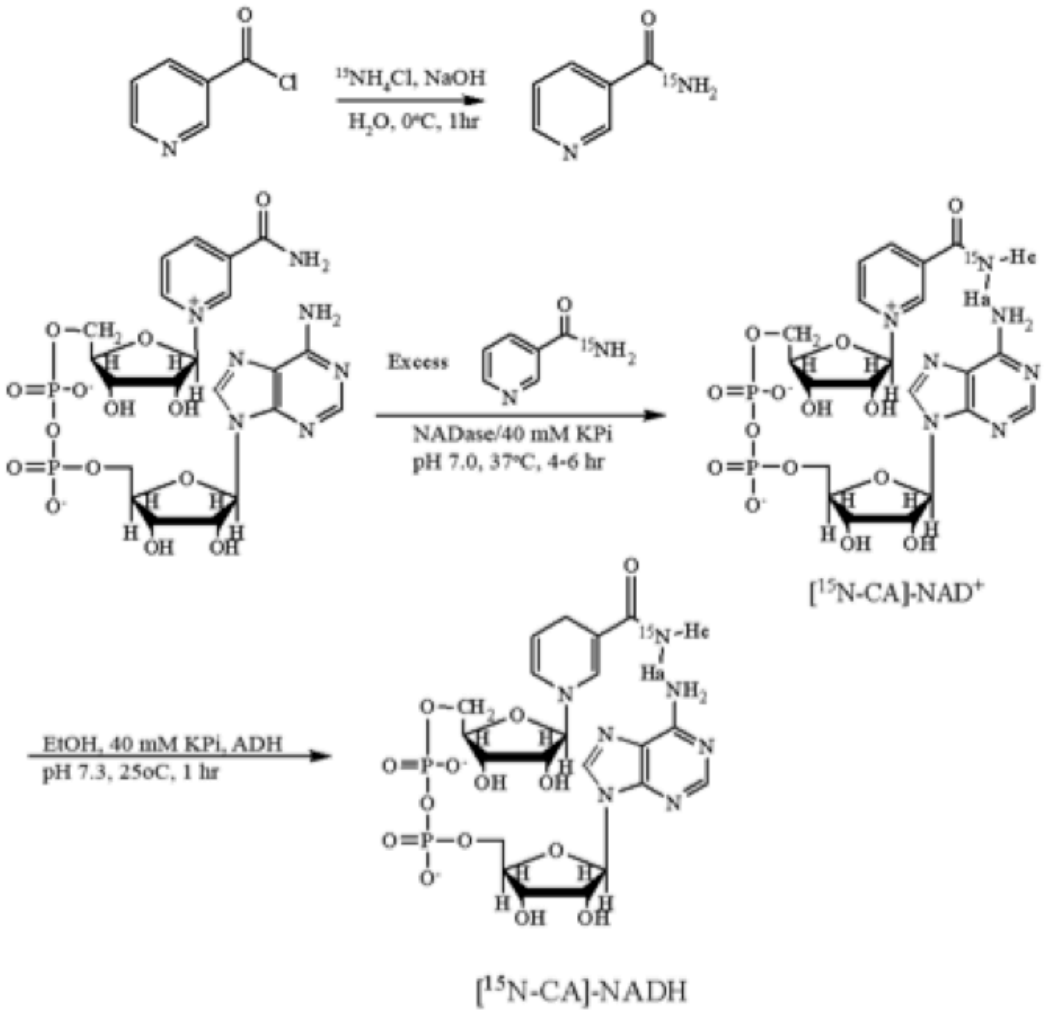


Figure 7.3. Synthetic scheme for preparation ^{15}N-nicotinamide and [^{15}N-CA]-NAD(H). ADH is alcohol dehydrogenase from yeast.

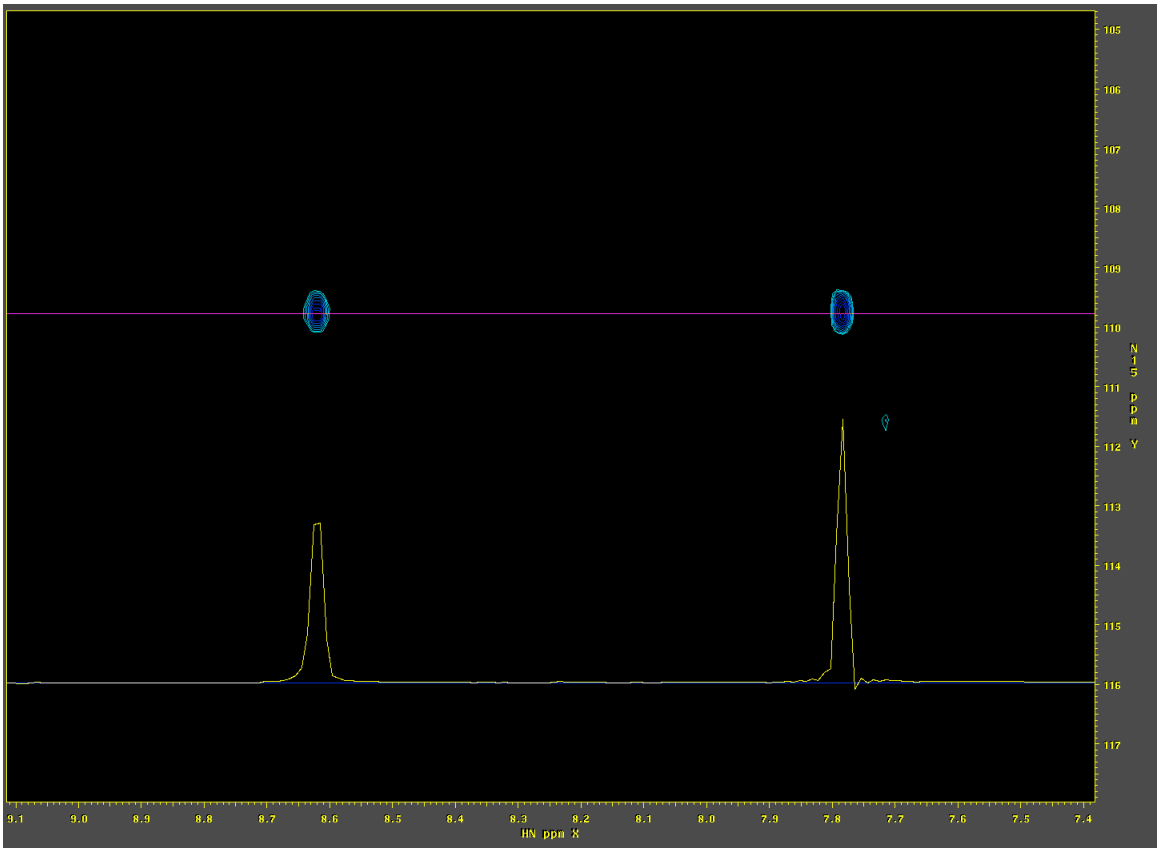


Figure 7.4. 2D ^{1}H-^{15}N HSQC spectrum of [^{15}N-CA]-NAD^{+} with 1D slice taken at the ^{15}N chemical shift indicated, showing the relative intensity of Ha (downfield) and He (upfield) protons. Spectrum was take at pH 7.0 and 5°C.

7.3.2 ^{15}N-edited NOESY studies of [^{15}N-CA]-NAD^{+}

A useful application of the [^{15}N-CA]-NAD^{+} probe is that it can be used to determine the orientation of the carboxamide group when binding to protein. The carboxamide, in the unbound state, should be able to rotate around the bond connecting to the pyridine ring of the nicotinamide group to sample two states (Figure 7.5). But, when bound to the protein, the amide protons are probably involved in hydrogen bonding to

amino acids in the active site, to stabilize the cofactor and permit hydride transfer. Since the ^{15}N isotope is on the amide group, the orientation of the amide can be determined with the ^{15}N-edited NOESY experiment, where NOEs are observed within 5 Å of the proton(s) stemming from the ^{15}N. NOE intensity is much greater, the closer two atoms are to each other, with intensity decreasing as $1/(\text{distance})^6$. So if the amide is oriented like Figure 7.5B then we should see an NOE around 9.3 ppm corresponding to the proton H_2, however if its oriented like Figure 7.5C the NOE should be around 9.0 ppm, corresponding to proton H_4.[117] The ^{15}N-[^1H-^1H] NOESY experiment of [^{15}N-CA]-NAD^+ bound to OMO shows the two characteristic diagonal cross peaks of the carboxamide protons of the cofactor at ~7.8 ppm (NH *syn*) and ~8.65 ppm (NH *anti*)with NOEs to each other. The strong peaks at approximately 4.9 ppm are due to magentization transfer with water, as this sample had to be in water (vs. D_2O) so the amide protons would be visible. At the bottom of the spectra there are NOE cross peaks correlating the amide protons (7.8, 8.6 ppm) with the H_2 protons at 9.26 ppm, indicating the carboxamide protons of [^{15}N-CA]-NAD^+ close to the H_2 proton of the nicotinamide ring. This information suggest the structure in Figure 7.5B is the relevant structure of [^{15}N-CA]-NAD^+ when bound to OMO. This chemical probe and NMR experiment provides a relatively simple test to verify the correct carboxamide geometry of any NAD(P)(H) cofactor when bound to an oxidoreductase. Further work will entail testing the cofactors with other oxidoreductases in literature and to confirm or disprove the reported geometry of the carboxamide in the crystal structure of cofactor/protein complexes.

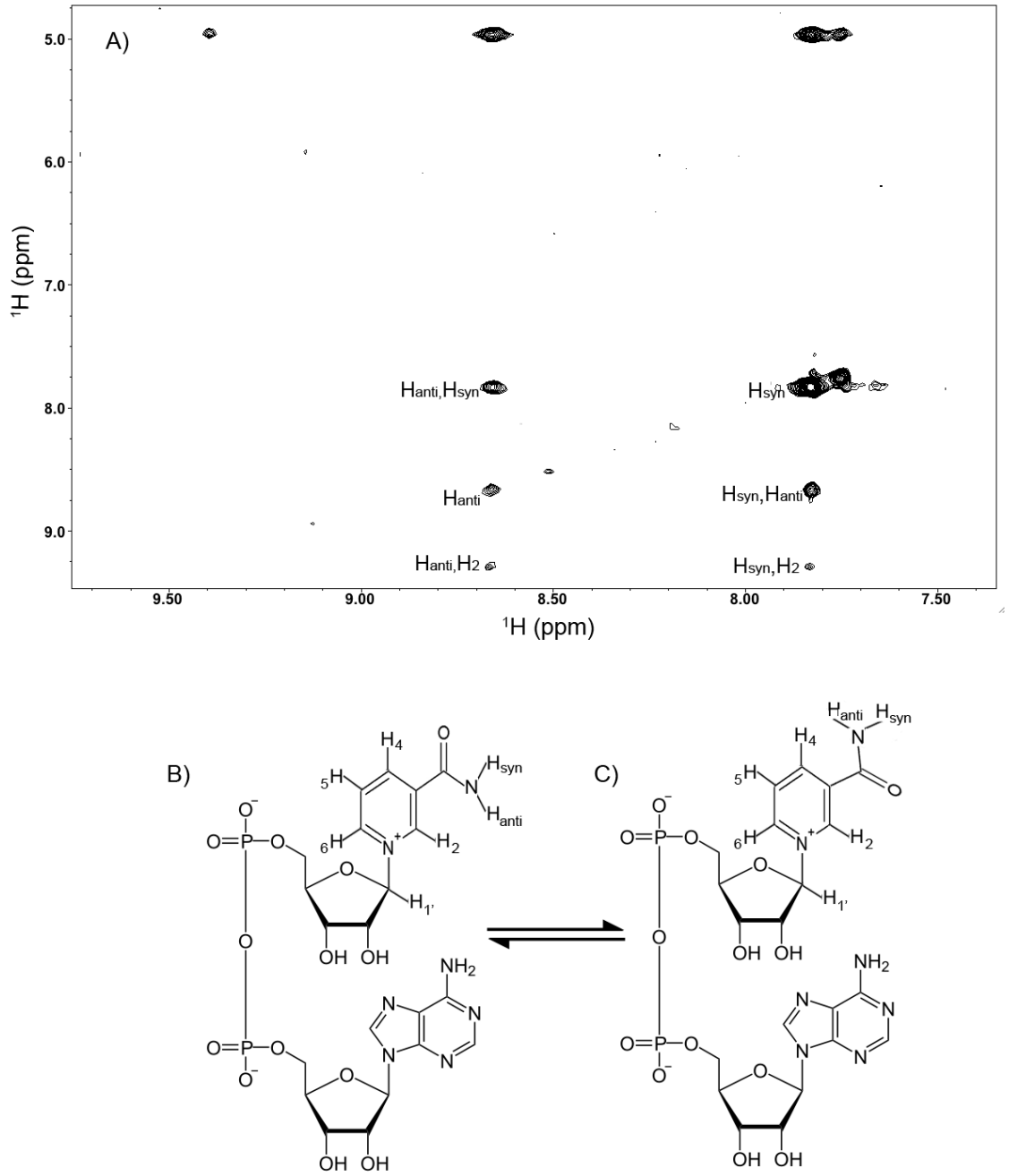


Figure 7.5. A) 2D ^{15}N-filtered [1H-1H}NOESY spectrum of [^{15}N-CA]-NAD$^+$ bound to OMO, and B) the two possible orientations of the carboxamide group of [^{15}N-CA]-NAD$^+$ The carboxamide could be oriented with the protons closer to the H2 proton (B) or to H4 proton (C). The NOE data indicates NAD+ binds OMO in the orientation shown in panel B.

7.3.3 Relaxation Dispersion Study to Determine Conformational Exchange of Bound [^{15}N-CA]-NAD$^+$ Cofactor

Conformational exchange is a phenomenon in which nuclei sample multiple conformations/environments, and is measured by the rate term R_{ex}, which is associated with transverse relaxation effects (T_2). Biological processes of proteins that undergo conformational exchange include protein folding, substrate binding, and catalysis, which often occur on the μsec-msec time-scale. One way to measure such motional processes is with NMR spectroscopy using relaxation dispersion experiments with the CPMG pulse sequence. These experiments measure the transverse relaxation rates while partially or fully suppressing the relaxation rate due to exchange R_{ex}, by a series of 180° refocusing pulses (or spin-echo's) known as CPMG.[20, 99, 118] The intensity of the peaks (i.e. the NH protons) are measured to get the transverse relaxation rates ($R_2 = 1/T_2$) using Eq.1, and are then plotted as a function of the delay (τ_{cp}) between successive spin-echo's and fit to Eq.2 to get R_{ex}. If there are R_{ex} contributions, the curve will demonstrate exponential decay in which the asymptote is the transverse relaxation rate (R_2). If there are no R_{ex} contributions, then only a flat line is observed.

Since OMO utilizes the cofactor FAD for electron transfer to its substrate, with NAD(P)H transferring a hydride ion to FAD, NAD$^+$ had to be used in place of NADH to avoid catalytic turnover. To study the conformational exchange effects of binding cofactor, relaxation dispersion profiles were acquired for unbound [^{15}N-CA]-NAD$^+$ and bound to OMO protein (in a 3:4 ratio so that all [^{15}N-CA]-NAD$^+$ was bound to OMO). The same relaxation dispersion experiments were used with both samples, except that the TROSY (Transverse Relaxation Optimized Spectroscopy)[119, 120] version of HSQC spectra

was acquired for OMO to sharpen the signals. Since the [^{15}N-CA]-NAD$^+$ is bound to a large protein OMO is (55 kDa), the slow rotation of the bound [^{15}N-CA]-NAD$^+$ would cause the signals to broaden, and TROSY compensates for this effect. For each experiment, a control experiment was run to determine the amide proton intensity (I_o) in the absence of delay (no CPMG) and was used in Eq.1 to determine R_2 and plotted against $1/\tau_{cp}$ and fit to Eq.2, with fitted curve shown in Figure 7.6.

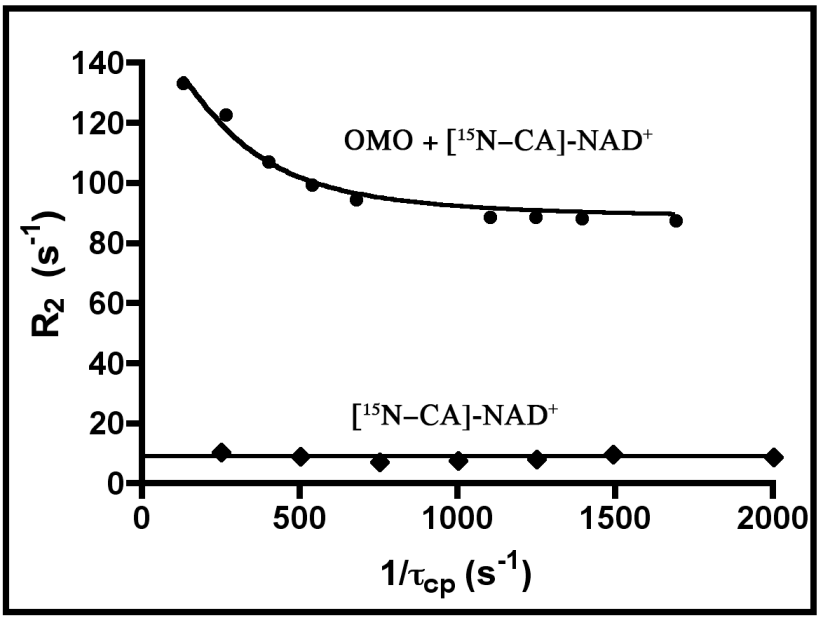


Figure 7.6. Relaxation dispersion plot for unbound [^{15}N-CA]-NAD$^+$ and [^{15}N-CA]-NAD$^+$ bound to OMO. Spectra were taken at pH 7.8 and 5°C with 500 μM [^{15}N-CA]-NAD$^+$. R_2 is the transverse relaxation rate for the amide protons, and τ_{cp} is the delay used in the CPFM pulse sequence.

The dispersion profile for unbound [^{15}N-CA]-NAD$^+$ in Figure 7.6 shows no conformational exchange occurs in this state, indicated by the flat line with an apparent transverse relaxation rate (R_2^o) of 9 s^{-1}. Upon binding to OMO the dispersion profile shows the exponential decay associated with conformational exchange due to binding

with a R_{ex} value of 67 ± 6 s^{-1}, an exchange rate constant (k_{ex}) of 922 ± 104 s^{-1}, and an R_2^o of 88 s^{-1}. The significant increase in R_2^o (9 to 88 s^{-1}) from unbound to bound states is most likely due to the slow tumbling of the bound labeled cofactor, which would lead to significant transverse relaxation. In effect, signal that is observed is like that of a large molecule (protein), even though we are still only observing the [^{15}N-CA]-NAD^+ amide protons. Since we also observe a transverse relaxation rate due to conformational exchange (R_{ex}) in the bound state of the cofactor, this is likely indicative of adopting multiple conformations in the bound state, perhaps due to domain motion for OMO to bind its other substrate (ornithine). If ornithine were bound to OMO in the presence of the [^{15}N-CA]-NAD^+, we would assume that the cofactor would then be locked down rigidly, for catalysis to occur, and would suppress most, if not all, of the R_{ex} contribution to R_2.

7.4 Conclusion

The ^{15}N-labeled cofactor NAD^+ probes, and their application described herein, are a useful tool for studies of NAD(P)(H) binding proteins, since they can be used to explore effects of complex formation and protein binding. We have shown that these probes can be used to determine carboxamide geometry (which crystallography often cannot) using the ^{15}N-filtered NOESY experiment. They can also be used to study cofactor binding dynamics, using CPMG-based relaxation dynamics methods. The synthesis of these probes is relatively straightforward and can be easily extended to the phosphorylated cofactor, since NAD^+ can be converted to $NADP^+$ using the protein NAD^+ kinase.[117] There are many ways that this probe can be used, for future experiments of binding [^{15}N-

CA]-$NADP^+$ to *M. tb* thioredoxin reductase. For example, it can be used to monitor the NADPH domain rotation (F_O and F_R states) in absence and presence of protein substrate thioredoxin C. This can be done with reduced and oxidized states of the cofactor (NADPH vs. $NADP^+$), of thioredoxin C, and of thioredoxin reductase.

BIBLIOGRAPHY

(1) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997***, 273*, 283-298.

(2) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002***, 319*, 209-227.

(3) Shen, ,Yang; Delaglio, ,Frank; Cornilescu, ,Gabriel; Bax, ,Ad *J. Biomol. NMR* **2009**, 213-223.

(4) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989***, 28*, 8972-8979.

(5) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982***, 104*, 4546-4559.

(6) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982***, 104*, 4559-4570.

(7) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990***, 112*, 4989-4991.

(8) Mandel, A. M.; Akke, M.; Palmer, A. G.,3rd *J. Mol. Biol.* **1995***, 246*, 144-163.

(9) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided Mol. Des. , 15*.

(10) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996***, 261*, 470-489.

(11) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004***, 47*, 1739-1749.

(12) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Computational Chemistry* **1998***, 19*, 1639-1662.

(13) Muller, C. W.; Schlauderer, G. J.; Reinstein, J.; Schulz, G. E. *Structure* **1996***, 4*, 147-156.

(14) Herdendorf, T. J.; Miziorko, H. M. *Biochemistry* **2006***, 45*, 3235-3242.

(15) Gerstein, M.; Lesk, A. M.; Chothia, C. *Biochemistry* **1994***, 33*, 6739-6749.

(16) Leipe, D. D.; Koonin, E. V.; Aravind, L. *J. Mol. Biol.* **2003***, 333*, 781-815.

(17) Herdendorf, T. J.; Miziorko, H. M. *Biochemistry* **2007***, 46*, 11780-11788.

(18) Teplyakov, A.; Sebastiao, P.; Obmolova, G.; Perrakis, A.; Brush, G. S.; Bessman, M. J.; Wilson, K. S. *EMBO J.* **1996***, 15*, 3487-3497.

(19) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994***, 33*, 5984-6003.

(20) Loria, J. P.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1999***, 121*, 2331-2332.

(21) Zhu, G.; Xia, Y.; Nicholson, L. K.; Sze, K. H. *J. Magn. Reson.* **2000***, 143*, 423-426.

(22) Tjandra, N.; Wingfield, P.; Stahl, S.; Bax, A. *J. Biomol. NMR* **1996***, 8*, 273-284.

(23) Gagne, S. M.; Tsuda, S.; Spyracopoulos, L.; Kay, L. E.; Sykes, B. D. *J. Mol. Biol.* **1998***, 278*, 667-686.

(24) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995***, 6*, 277-93.

(25) Johnson, B. A., and Blevins, R. A. *J. Biomol. NMR* **1994**, 603-614.

(26) Eyzaguirre, J.; Valdebenito, D.; Cardemil, E. *Arch Biochem Biophys.* **2006***, 454*, 189-196.

(27) Savard, P. -.; Gagne, S. M. *Biochemistry* **2006***, 45*, 11414-11424.

(28) Yan, H.; Tsai, M. D. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1999***, 73*, 103-34, x.

(29) Whitford, P. C.; Gosavi, S.; Onuchic, J. N. *J. Biol. Chem.* **2008***, 283*, 2042-2048.

(30) Burlacu-Miron, S.; Gilles, A. M.; Popescu, A.; Barzu, O.; Craescu, C. T. *Eur. J. Biochem.* **1999***, 264*, 765-774.

(31) Shapiro, Y. E.; Kahana, E.; Tugarinov, V.; Liang, Z.; Freed, J. H.; Meirovitch, E. *Biochemistry* **2002***, 41*, 6271-6281.

(32) Hanson, J. A.; Duderstadt, K.; Watkins, L. P.; Bhattacharyya, S.; Brokaw, J.; Chu, J. W.; Yang, H. *Proc. Natl. Acad. Sci. U. S. A.* **2007***, 104*, 18055-18060.

(33) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007***, 450*, 913-916.

(34) Temiz, N. A.; Meirovitch, E.; Bahar, I. *Proteins* **2004***, 57*, 468-480.

(35) Buck, M.; Boyd, J.; Redfield, C.; MacKenzie, D. A.; Jeenes, D. J.; Archer, D. B.; Dobson, C. M. *Biochemistry* **1995***, 34*, 4041-4055.

(36) Hellig, H., Popjak, G *J. Lipid Res.* **1961**, 235-243.

(37) Cai, M.; Gong, Y. -.; Wen, L.; Krishnamoorthi, R. *Biochemistry* **2002***, 41*, 9572-9579.

(38) Trbovic, ,Nikola; Cho, ,Jae-Hyun; Abel, ,Robert; Friesner, R.,A.; Rance, ,Mark; Palmer, A.,G. *J Am Chem Soc.* **2009***, 21*, 615-622.

(39) Olson, A. L.; Yao, H.; Herdendorf, T. J.; Miziorko, H. M.; Hannongbua, S.; Saparpakorn, P.; Cai, S.; Sem, D. S. *Proteins: Structure, Function, and Bioinformatics* **2008**.

(40) Hsu, S. T.; Cabrita, L. D.; Fucini, P.; Christodoulou, J.; Dobson, C. M. *J. Am. Chem. Soc.* **2009***, 131*, 8366-8367.

(41) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. *Chem. Rev.* **2006***, 106*, 1672-1699.

(42) Duckert, P.; Brunak, S.; Blom, N. *Protein Eng. Des. Sel.* **2004***, 17*, 107-112.

(43) Seidah, N. G.; Khatib, A. M.; Prat, A. *Biol. Chem.* **2006***, 387*, 871-877.

(44) Seidah, N. G.; Prat, A. *J. Mol. Med.* **2007***, 85*, 685-696.

(45) Dubuc, G.; Chamberland, A.; Wassef, H.; Davignon, J.; Seidah, N. G.; Bernier, L.; Prat, A. *Arterioscler. Thromb. Vasc. Biol.* **2004***, 24*, 1454-1459.

(46) Rashid, S.; Curtis, D. E.; Garuti, R.; Anderson, N. N.; Bashmakov, Y.; Ho, Y. K.; Hammer, R. E.; Moon, Y. A.; Horton, J. D. *Proc. Natl. Acad. Sci. U. S. A.* **2005***, 102*, 5374-5379.

(47) Scamuffa, N.; Calvo, F.; Chretien, M.; Seidah, N. G.; Khatib, A. M. *FASEB J.* **2006***, 20*, 1954-1963.

(48) Benjannet, S.; Rhainds, D.; Hamelin, J.; Nassoury, N.; Seidah, N. G. *J. Biol. Chem.* **2006***, 281*, 30561-30572.

(49) Silva, M.; Rogers, P.; Arnone, A. *J. Biol. Chem.* **1992***, 267*, 17248-17256.

(50) Verho, R.; Putkonen, M.; Londesborough, J.; Penttila, M.; Richard, P. *J. Biol. Chem.* **2004***, 279*, 14746-14751.

(51) Mayer, M.; Meyer, B. *J. Am. Chem. Soc.* **2001***, 123*, 6108-6117.

(52) Anonymous*WHO* **2009**.

(53) Jasmer, R. M.; Nahid, P.; Hopewell, P. C. *N. Engl. J. Med.* **2002***, 347*, 1860-1866.

(54) M. Akif, K. Suhre, C. Verma and S. C.,Mande .

(55) Shinnick, T. M.; King, C. H.; Quinn, F. D. *Am. J. Med. Sci.* **1995***, 309*.

(56) Holmgren, A. *Antioxid. Redox Signal.* **2000***, 2*, 811-820.

(57) Arnér, E. S. J.; Holmgren, A. *European Journal of Biochemistry* **2000***, 267*, 6102-6109.

(58) Holmgren, A. *Annu. Rev. Biochem.* **1985***, 54*, 237-271.

(59) Martin, J. L. *Structure* **1995***, 3*, 245-250.

(60) Gleason, F. K.; Holmgren, A. *FEMS Microbiol. Lett.* **1988***, 54*, 271-297.

(61) Ortenberg, R.; Gon, S.; Porat, A.; Beckwith, J. *Proceedings of the National Academy of Sciences of the United States of America* **2004***, 101*, 7439-7444.

(62) Powis, G.; Briehl, M.; Oblong, J. *Pharmacol. Ther.* **1995***, 68*, 149-173.

(63) Williams, C. H.; Arscott, L. D.; Müller, S.; Lennon, B. W.; Ludwig, M. L.; Wang, P.; Veine, D. M.; Becker, K.; Schirmer, R. H. *European Journal of Biochemistry* **2000***, 267*, 6110-6117.

(64) Akif, M.; Khare, G.; Tyagi, A. K.; Mande, S. C.; Sardesai, A. A. *J. Bacteriol.* **2008***, 190*, 7087-7095.

(65) Hall, G.; Shah, M.; McEwan, P. A.; Laughton, C.; Stevens, M.; Westwell, A.; Emsley, J. *Acta Crystallogr. D Biol. Crystallogr.* **2006***, 62*, 1453-1457.

(66) Zhang, Z.; Hillas, P. J.; Ortiz de Montellano, P. R. *Arch. Biochem. Biophys.* **1999***, 363*, 19-26.

(67) Shah, M. [.; Wells, G. [.; Bradshaw, T. D. [.; Laughton, C. A. [.; Stevens, M. F. G. [.; Westwell, A. D. [. *Letters in Drug Design & Discovery* **2006***, 3*, 419-423.

(68) Jeng, M.; Campbell, A. P.; Begley, T.; Holmgren, A.; Case, D. A.; Wright, P. E.; Dyson, H. J. *Structure* **1994***, 2*, 853-868.

(69) Weichsel, A.; Gasdaska, J. R.; Powis, G.; Montfort, W. R. *Structure* **1996***, 4*, 735-751.

(70) Qin, J.; Clore, G. M.; Gronenborn, A. M. *Structure* **1994***, 2*, 503-522.

(71) Peterson, F. C.; Lytle, B. L.; Sampath, S.; Vinarov, D.; Tyler, E.; Shahan, M.; Markley, J. L.; Volkman, B. F. *Protein Science* **2005***, 14*, 2195-2200.

(72) Lennon, B. W.; Williams, C. H., Jr.; Ludwig, M. L. *Science* **2000***, 289*, 1190-1194.

(73) M. Akif, R. Chauhan and S. C.,Mande .

(74) Marley, J.; Lu, M.; Bracken, C. *J. Biomol. NMR* **2001***, 20*, 71-75.

(75) Güntert, P.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1991***, 217*, 517-530.

(76) Lennon, B. W.; Williams, C. H. *Biochemistry (N. Y.)* **1997***, 36*, 9464-9477.

(77) Veine, D. M.; Mulrooney, S. B.; Wang, P. -.; Williams, C. H. *Protein Science* **1998***, 7*, 1441-1450.

(78) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004***, 3*, 935-949.

(79) Gohlke, H.; Klebe, G. *Angew. Chem. Int. Ed.* **2002***, 41*, 2644–2676.

(80) Shoichet, B. K. *Nature* **2004***, 432*, 862-865.

(81) Thain, D.; Tannenbaum, T.; Livny, M. *Concurrency and Computation: Practice and Experience* **2005***, 17*, 323-356.

(82) Frey, J.; Tannenbaum, T.; Livny, M.; Foster, I.; Tuecke, S. *Cluster Computing* **2002***, 5*, 237-246.

(83) Doun, S. S.; Burgner, J. W.,2nd; Briggs, S. D.; Rodwell, V. W. *Protein Sci.* **2005***, 14*, 1134-1139.

(84) Istvan, E. S.; Deisenhofer, J. *Science* **2001***, 292*, 1160-1164.

(85) Chang, Q.; Yan, X.; Gu, S.; Liu, J.; Liang, D. *Proteins: Structure, Function, and Bioinformatics* **2008***, 73*, 254-258.

(86) Ueda, K.; Arakawa, H.; Nakamura, Y. *Oncogene* **2003***, 22*, 5586-5591.

(87) Chang, L.; Karin, M. *Nature* **2001***, 410*, 37-40.

(88) Alonso, A.; Sasin, J.; Bottini, N.; Friedberg, I.; Friedberg, I.; Osterman, A.; Godzik, A.; Hunter, T.; Dixon, J.; Mustelin, T. *Cell* **2004***, 117*, 699-711.

(89) Camps, M.; Nichols, A.; Arkinstall, S. *FASEB J.* **2000***, 14*, 6-16.

(90) Jeong, D. G.; Cho, Y. H.; Yoon, T. -.; Kim, J. H.; Ryu, S. E.; Kim, S. J. *Proteins: Structure, Function, and Bioinformatics* **2007***, 66*, 253-258.

(91) Saraste, M.; Sibbald, P. R.; Wittinghofer, A. *Trends Biochem. Sci.* **1990***, 15*, 430-434.

(92) Zheng, H.; Fu, G.; Dai, T.; Huang, H. *J. Cardiovasc. Pharmacol.* **2007***, 50*, 274-280.

(93) Kryczek, I.; Wei, S.; Keller, E.; Liu, R.; Zou, W. *Am J Physiol Cell Physiol* **2007***, 292*, C987-995.

(94) Veldkamp, C. T.; Seibert, C.; Peterson, F. C.; De la Cruz, N. B.; Haugner, J. C.,III; Basnet, H.; Sakmar, T. P.; Volkman, B. F. *Sci. Signal.* **2008***, 1*, ra4.

(95) Veldkamp, C. T.; Seibert, C.; Peterson, F. C.; Sakmar, T. P.; Volkman, B. F. *J. Mol. Biol.* **2006***, 359*, 1400-1409.

(96) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *Journal of Computational Chemistry* **2009***, 30*, 2785-2791.

(97) Hajduk, P. J. *Nat Chem Biol* **2006***, 2*, 658-659.

(98) Hajduk, P. J.; Greer, J. *Nat Rev Drug Discov* **2007***, 6*, 211-219.

(99) Palmer III, A. G.; Kroenke, C. D.; Loria, J. P. *Methods in Enzymology* **2001**, *339*, 204-238.

(100) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006***, 106*, 1624-1671.

(101) Mittermaier, A.; Kay, L. E. *Science* **2006***, 312*, 224-228.

(102) Hajduk, P. J. *Nat Chem Biol* **2006***, 2*, 658-659.

(103) Hajduk, P. J.; Greer, J. *Nat. Rev. Drug Discov.* **2007***, 6*, 211-219.

(104) Pollak, N.; Dolle, C.; Ziegler, M. *Biochem. J.* **2007***, 402*, 205-218.

(105) Sem, D. S.; Bertolaet, B.; Baker, B.; Chang, E.; Costache, A. D.; Coutts, S.; Dong, Q.; Hansen, M.; Hong, V.; Huang, X.; Jack, R. M.; Kho, R.; Lang, H.; Ma, C.; Meininger, D.; Pellecchia, M.; Pierre, F.; Villar, H.; Yu, L. *Chem. Biol.* **2004***, 11*, 185-194.

(106) Kho, R.; Newman, J. V.; Jack, R. M.; Villar, H. O.; Hansen, M. R. *Journal of Proteome Research* **2003***, 2*, 626-632.

(107) Mayfield, J. A.; Frederick, R. E.; Streit, B. R.; Wencewicz, T. A.; Ballou, D. P.; DuBois, J. L. *Journal of Biological Chemistry* **2010**.

(108) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006***, 313*, 1638-1642.

(109) Sevrioukova, I. F.; Li, H.; Poulos, T. L. *J. Mol. Biol.* **2004***, 336*, 889-902.

(110) Wang, ,Chunyu; Grey, ,Michael; Palmer, ,Arthur *J. Biomol. NMR* **2001**, 361-366.

(111) Zatman, L. J.; Kaplan, N. O.; Colowick, S. P. *Journal of Biological Chemistry* **1953***, 200*, 197-212.

(112) Colowick, S. P.; Kaplan, N. O.; Ciotti, M. M. *Journal of Biological Chemistry* **1951***, 191*, 447-459.

(113) Lowry, O. H.; Passonneau, J. V.; Rock, M. K. *Journal of Biological Chemistry* **1961***, 236*, 2756-2759.

(114) Anderson, B. M.; Kaplan, N. O. *Journal of Biological Chemistry* **1959***, 234*, 1226-1232.

(115) Anderson, B. M.; Ciotti, C. J.; Kaplan, N. O. *Journal of Biological Chemistry* **1959***, 234*, 1219-1225.

(116) Burke, J.,R.; Frey, P.,A. **1993***, 32*, 13220-13230.

(117) Sem, D. S.; Kasper, C. B. *Biochemistry (N. Y.)* **1992***, 31*, 3391-3398.

(118) Boehr, D. D.; Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2006***, 106*, 3055-3079.

(119) Loria, J.,Patrick; Rance, ,Mark; Palmer, A.,G. *J. Biomol. NMR* **1999**, 151-155.

(120) Zhu, G.; Xia, Y.; Nicholson, L. K.; Sze, K. H. *Journal of Magnetic Resonance* **2000***, 143*, 423-426.

APPENDIX

Scripts and Procedure for NMR Structure Calculation

A) The files necessary to begin structure calculation coming from experimental data.

1) Protein sequence file (the exact sequence of the protein expressed, even if residues could not be assigned) called nameofprotein.seq with the format;

MET 1
THR 2
ASP 3
SER 4
GLU 5
LYS 6
SER 7
ALA 8
THR 9
ILE 10
LYS 11
VAL 12
THR 13
ASP 14
ALA 15
SER 16
PHE 17
ALA 18
THR 19
ASP 20
VAL 21
LEU 22
SER 23
SER 24
ASN 25
LYS 26
PRO 27
VAL 28
LEU 29
VAL 30
ASP 31
PHE 32
TRP 33
ALA 34
THR 35
TRP 36
CYSS 37
GLY 38
PRO 39
CYSS 40
LYS 41
MET 42
VAL 43
ALA 44
PRO 45
VAL 46
LEU 47
GLU 48
GLU 49
ILE 50
ALA 51
THR 52
GLU 53
ARG 54
ALA 55
THR 56
ASP 57
LEU 58
THR 59
VAL 60
ALA 61
LYS 62
LEU 63
ASP 64
VAL 65
ASP 66
THR 67
ASN 68
PRO 69
GLU 70
THR 71
ALA 72
ARG 73
ASN 74
PHE 75
GLN 76
VAL 77
VAL 78
SER 79
ILE 80
PRO 81
THR 82
LEU 83
ILE 84
LEU 85
PHE 86
LYS 87
ASP 88
GLY 89
GLN 90
PRO 91
VAL 92
LYS 93
ARG 94
ILE 95
VAL 96
GLY 97
ALA 98
LYS 99
GLY 100
LYS 101
ALA 102
ALA 103
LEU 104
LEU 105
ARG 106
GLU 107
LEU 108
SER 109
ASP 110

In which cysteines involved in disulfides have the name CYSS, and prolines in the cis conformation have the cPRO. This file is created manually.

2) The chemical shift list called nameofprotein.prot is every assignment that could be made from all backbone and sidechain spectra, and is written using the xe command in sparky. (note: every time a prot file is written, sparky should be closed and re-opened as sparky retains all chemical shift information and this can cause problems later on) This file will have the format;

1 55.063 0.000 CA 3	82 35.040 0.055 CB 12	163 22.621 0.050 CD2 22	244 4.587 0.006 HA 30	323 51.759 0.044 CD 39
2 41.232 0.000 CB 3	83 22.622 0.074 CG1 12	164 25.294 0.006 CG 22	245 2.337 0.008 HB 30	324 28.036 0.107 CG 39
3 4.698 0.000 HA 3	84 20.256 0.000 CG2 12	165 7.149 0.002 H 22	246 127.755 0.045 N 30	325 4.423 0.004 HA 39
4 2.627 0.000 HB2 3	85 8.576 0.002 H 12	166 3.738 0.010 HA 22	247 0.217 0.012 QG1 30	326 2.515 0.009 HB2 39
5 2.760 0.000 HB3 3	86 4.687 0.009 HA 12	167 2.312 0.012 HB2 22	248 0.914 0.010 QG2 30	327 1.779 0.008 HB3 39
6 58.682 0.000 CA 4	87 1.969 0.010 HB 12	168 1.608 0.008 HB3 22	249 52.327 0.012 CA 31	328 4.553 0.008 HD2 39
7 63.603 0.196 CB 4	88 120.675 0.056 N 12	169 1.205 0.005 HG 22	250 40.487 0.127 CB 31	329 3.770 0.007 HD3 39
8 8.414 0.001 H 4	89 0.820 0.009 QG1 12	170 115.492 0.037 N 22	251 9.169 0.007 H 31	330 2.056 0.007 HG2 39
9 4.393 0.007 HA 4	90 0.748 0.007 QG2 12	171 0.741 0.010 QD2 22	252 5.024 0.013 HA 31	331 2.276 0.009 HG3 39
10 3.847 0.008 HB2 4	91 58.711 0.004 CA 13	172 58.077 0.159 CA 23	253 2.226 0.004 HB2 31	332 64.025 0.037 CA 40
11 4.023 0.004 HB3 4	92 73.893 0.000 CB 13	173 63.339 0.051 CB 23	254 2.929 0.006 HB3 31	333 27.621 0.003 CB 40
12 116.388 0.014 N 4	93 8.135 0.002 H 13	174 7.068 0.002 H 23	255 125.859 0.036 N 31	334 7.875 0.003 H 40
13 56.626 0.011 CA 5	94 4.778 0.006 HA 13	175 4.551 0.007 HA 23	256 58.094 0.056 CA 32	335 4.121 0.011 HA 40
14 29.908 0.144 CB 5	95 4.777 0.004 HB 13	176 4.043 0.015 HB2 23	257 38.298 0.121 CB 32	336 3.410 0.002 HB2 40
15 36.218 0.031 CG 5	96 110.422 0.032 N 13	177 4.059 0.010 HB3 23	258 132.812 0.116 CD1 32	337 3.607 0.013 HB3 40
16 8.394 0.010 H 5	97 1.205 0.005 QG2 13	178 109.764 0.014 N 23	259 127.776 0.072 CZ 32	338 114.401 0.029 N 40
17 4.271 0.009 HA 5	98 58.096 0.059 CA 14	179 57.924 0.035 CA 24	260 9.266 0.007 H 32	339 58.954 0.093 CA 41
18 1.937 0.012 HB2 5	99 40.722 0.128 CB 14	180 63.904 0.010 CB 24	261 5.292 0.006 HA 32	340 32.060 0.000 CB 41
19 2.038 0.007 HB3 5	100 9.093 0.007 H 14	181 6.552 0.005 H 24	262 3.626 0.009 HB2 32	341 29.237 0.000 CD 41
20 2.225 0.010 HG2 5	101 4.328 0.004 HA 14	182 4.270 0.006 HA 24	263 2.647 0.011 HB3 32	342 42.167 0.124 CE 41
21 2.267 0.006 HG3 5	102 2.705 0.007 HB2 14	183 3.693 0.005 HB2 24	264 6.865 0.006 HZ 32	343 25.469 0.000 CG 41
22 122.604 0.079 N 5	103 2.814 0.015 HB3 14	184 3.787 0.003 HB3 24	265 128.566 0.030 N 32	344 8.116 0.005 H 41
23 56.050 0.000 CA 6	104 120.389 0.011 N 14	185 112.853 0.044 N 24	266 7.374 0.007 QD 32	345 4.121 0.010 HA 41
24 33.182 0.087 CB 6	105 54.086 0.014 CA 15	186 54.597 0.000 CA 25	267 7.001 0.002 QE 32	346 1.980 0.004 HB2 41
25 29.023 0.000 CD 6	106 18.715 0.079 CB 15	187 38.476 0.001 CB 25	268 54.545 0.046 CA 33	347 2.034 0.009 HB3 41
26 42.302 0.000 CE 6	107 8.225 0.005 H 15	188 8.943 0.002 H 25	269 32.813 0.000 CB 33	348 1.780 0.008 HD2 41
27 24.474 0.000 CG 6	108 4.310 0.009 HA 15	189 4.787 0.002 HA 25	270 126.151 0.092 CD1 33	349 1.741 0.000 HD3 41
28 8.241 0.011 H 6	109 119.446 0.028 N 15	190 2.893 0.013 HB2 25	271 124.582 0.074 CH2 33	350 1.512 0.002 HG2 41
29 4.341 0.002 HA 6	110 1.450 0.007 QB 15	191 2.920 0.005 HB3 25	272 114.248 0.051 CZ2 33	351 1.642 0.000 HG3 41
30 1.727 0.005 HB2 6	111 58.955 0.002 CA 16	192 127.681 0.019 N 25	273 121.359 0.117 CZ3 33	352 119.964 0.081 N 41
31 1.808 0.014 HB3 6	112 65.303 0.030 CB 16	193 53.783 0.000 CA 26	274 8.604 0.008 H 33	353 3.052 0.006 QE 41
32 122.518 0.029 N 6	113 8.132 0.003 H 16	194 33.680 0.000 CB 26	275 5.192 0.007 HA 33	354 57.324 0.001 CA 42
33 1.394 0.003 QD 6	114 4.701 0.009 HA 16	195 29.604 0.000 CD 26	276 3.061 0.003 HB2 33	355 32.204 0.000 CB 42
34 2.982 0.000 QE 6	115 4.360 0.006 HB2 16	196 41.790 0.056 CE 26	277 3.404 0.004 HB3 33	356 32.115 0.007 CG 42
35 1.672 0.000 QG 6	116 3.619 0.008 HB3 16	197 24.239 0.000 CG 26	278 7.232 0.008 HD1 33	357 7.397 0.004 H 42
36 57.132 0.180 CA 7	117 115.620 0.070 N 16	198 7.820 0.006 H 26	279 10.368 0.005 HE1 33	358 4.328 0.005 HA 42
37 64.338 0.000 CB 7	118 64.042 0.038 CA 17	199 4.790 0.003 HA 26	280 7.125 0.010 HH2 33	359 2.571 0.005 HG2 42
38 8.260 0.001 H 7	119 39.999 0.048 CB 17	200 1.652 0.009 HB2 26	281 7.484 0.003 HZ2 33	360 2.674 0.011 HG3 42
39 4.551 0.006 HA 7	120 131.972 0.114 CD1 17	201 1.544 0.012 HB3 26	282 6.863 0.005 HZ3 33	361 116.264 0.045 N 42
40 3.896 0.005 HB2 7	121 130.748 0.071 CE1 17	202 1.443 0.002 HD2 26	283 122.476 0.055 N 33	362 2.195 0.007 QB 42
41 3.652 0.002 HB3 7	122 128.638 0.082 CZ 17	203 1.559 0.009 HD3 26	284 129.830 0.012 NE1 33	363 62.943 0.007 CA 43
42 117.134 0.030 N 7	123 8.172 0.005 H 17	204 120.623 0.133 N 26	285 51.299 0.000 CA 34	364 33.395 0.167 CB 43
43 52.590 0.033 CA 8	124 3.603 0.004 HA 17	205 2.793 0.008 QE 26	286 22.663 0.003 CB 34	365 20.769 0.000 CG1 43
44 20.124 0.000 CB 8	125 3.541 0.014 HB2 17	206 1.416 0.013 QG 26	287 7.032 0.003 H 34	366 22.306 0.095 CG2 43
45 9.691 0.006 H 8	126 3.038 0.005 HB3 17	207 63.912 0.034 CA 27	288 3.728 0.013 HA 34	367 7.236 0.007 H 43
46 4.454 0.001 HA 8	127 6.444 0.006 HZ 17	208 32.082 0.054 CB 27	289 116.614 0.023 N 34	368 4.143 0.007 HA 43
47 129.351 0.089 N 8	128 126.834 0.083 N 17	209 50.983 0.016 CD 27	290 0.348 0.007 QB 34	369 2.098 0.011 HB 43
48 1.378 0.013 QB 8	129 6.926 0.003 QD 17	210 28.689 0.000 CG 27	291 21.858 0.046 CG2 35	370 115.791 0.029 N 43
49 61.020 0.047 CA 9	130 6.675 0.009 QE 17	211 4.867 0.006 HA 27	292 9.334 0.003 H 35	371 0.998 0.010 QG1 43
50 70.889 0.050 CB 9	131 55.676 0.034 CA 18	212 1.966 0.018 HB2 27	293 4.232 0.005 HA 35	372 1.158 0.013 QG2 43
51 21.194 0.072 CG2 9	132 17.528 0.016 CB 18	213 2.132 0.011 HB3 27	294 3.972 0.009 HB 35	373 57.560 0.031 CA 44
52 8.335 0.006 H 9	133 9.142 0.005 H 18	214 3.962 0.006 QD 27	295 115.455 0.052 N 35	374 15.381 0.031 CB 44
53 4.531 0.004 HA 9	134 3.937 0.010 HA 18	215 2.361 0.002 QG 27	296 1.314 0.009 QG2 35	375 7.369 0.008 H 44
54 3.843 0.010 HB 9	135 120.418 0.027 N 18	216 60.187 3.187 CA 28	297 54.193 0.098 CA 36	376 3.916 0.006 HA 44
55 111.119 0.009 N 9	136 1.558 0.007 QB 18	217 35.213 0.039 CB 28	298 29.479 0.000 CB 36	377 123.443 0.010 N 44
56 0.863 0.008 QG2 9	137 65.254 0.001 CA 19	218 22.495 0.193 CG1 28	299 129.058 0.066 CD1 36	378 1.368 0.007 QB 44
57 59.562 0.042 CA 10	138 69.205 0.032 CB 19	219 20.783 0.112 CG2 28	1015 121.133 0.022 CE3 36	379 65.892 0.062 CA 45
58 40.114 0.000 CB 10	139 21.777 0.027 CG2 19	220 8.638 0.010 H 28	300 115.188 0.094 CZ2 36	380 30.930 0.055 CB 45
59 13.430 0.064 CD1 10	140 7.864 0.005 H 19	221 5.016 0.016 HA 28	301 121.116 0.000 CZ3 36	381 50.571 0.052 CD 45
60 26.463 0.000 CG1 10	141 3.911 0.003 HA 19	222 1.910 0.010 HB 28	302 6.564 0.006 H 36	382 28.218 0.085 CG 45
61 17.374 0.000 CG2 10	142 4.076 0.004 HB 19	223 120.208 0.065 N 28	303 4.564 0.015 HA 36	383 4.323 0.005 HA 45
62 8.695 0.005 H 10	143 113.363 0.032 N 19	224 0.637 0.009 QG1 28	304 3.692 0.007 HB2 36	384 1.933 0.005 HB2 45
63 4.651 0.002 HA 10	144 1.192 0.007 QG2 19	225 1.064 0.009 QG2 28	305 3.167 0.009 HB3 36	385 2.315 0.008 HB3 45
64 1.800 0.010 HB 10	145 56.648 0.049 CA 20	226 53.023 0.003 CA 29	306 7.390 0.004 HD1 36	386 3.734 0.011 HD2 45
65 116.239 0.135 N 10	146 42.156 0.042 CB 20	227 44.474 0.001 CB 29	1016 7.385 0.002 HE3 36	387 3.625 0.007 HD3 45
66 0.642 0.010 QD1 10	147 8.576 0.006 H 20	228 23.741 0.064 CD1 29	307 7.165 0.000 HH2 36	388 2.053 0.010 HG2 45
67 1.311 0.007 QG1 10	148 4.274 0.007 HA 20	229 26.249 0.009 CD2 29	308 7.360 0.002 HZ2 36	389 2.028 0.010 HG3 45
68 0.843 0.007 QG2 10	149 119.229 0.025 N 20	230 26.967 0.000 CG 29	309 7.384 0.000 HZ3 36	390 65.533 0.058 CA 46
69 56.240 0.001 CA 11	150 2.271 0.009 QB 20	231 9.398 0.006 H 29	310 114.739 0.018 N 36	391 31.690 0.004 CB 46
70 33.245 0.000 CB 11	151 63.365 0.023 CA 21	232 4.932 0.006 HA 29	311 58.287 0.069 CA 37	392 21.243 0.070 CG1 46
71 29.625 0.000 CD 11	152 32.031 0.010 CB 21	233 2.126 0.011 HB2 29	312 26.869 0.003 CB 37	393 22.201 0.027 CG2 46
72 42.143 0.000 CE 11	153 20.602 0.040 CG1 21	234 1.167 0.008 HB3 29	313 6.521 0.009 H 37	394 6.742 0.002 H 46
73 25.259 0.020 CG 11	154 20.830 0.000 CG2 21	235 1.407 0.014 HG 29	314 4.629 0.011 HA 37	395 3.731 0.004 HA 46
74 8.643 0.004 H 11	155 7.640 0.004 H 21	236 130.791 0.011 N 29	315 1.550 0.010 HB2 37	396 2.363 0.011 HB 46
75 4.735 0.006 HA 11	156 3.843 0.008 HA 21	237 0.784 0.012 QD1 29	316 2.468 0.004 HB3 37	397 118.790 0.011 N 46
76 126.201 0.107 N 11	157 0.581 0.001 HB 21	238 0.811 0.012 QD2 29	317 122.120 0.018 N 37	398 0.854 0.011 QG1 46
77 1.790 0.005 QB 11	158 113.358 0.015 N 21	239 61.461 0.053 CA 30	318 49.232 0.026 CA 38	399 1.157 0.008 QG2 46
78 1.618 0.004 QD 11	159 -0.438 0.007 QG1 21	240 33.043 0.034 CB 30	319 4.012 0.006 HA2 38	400 57.447 0.044 CA 47
79 3.014 0.000 QE 11	160 0.001 0.007 QG2 21	241 22.307 0.029 CG1 30	320 4.307 0.010 HA3 38	401 40.806 0.059 CB 47
80 1.226 0.010 QG 11	161 58.022 0.031 CA 22	242 24.538 0.000 CG2 30	321 65.819 0.000 CA 39	402 22.211 0.000 CD1 47
81 59.077 0.082 CA 12	162 39.425 0.048 CB 22	243 9.595 0.004 H 30	322 32.618 0.000 CB 39	773 2.981 0.003 HB3 88

403	26.583	0.064	CD2 47	496	4.746	0.008	HA 57	589	7.113	0.004	H 68	680	64.785	0.013	CB 79
404	26.558	0.000	CG 47	497	2.459	0.004	HB2 57	590	2.227	0.006	HB2 68	681	7.879	0.008	H 79
405	7.722	0.005	H 47	498	2.573	0.005	HB3 57	591	119.223	0.106	N 68	682	4.706	0.009	HA 79
406	3.841	0.010	HA 47	499	118.403	0.103	N 57	592	64.195	0.000	CA 69	683	117.284	0.044	N 79
407	1.212	0.010	HB2 47	500	54.675	0.076	CA 58	593	32.587	0.018	CB 69	684	3.698	0.006	QB 79
408	1.886	0.008	HB3 47	501	45.758	0.001	CB 58	594	50.379	0.033	CD 69	685	56.994	0.129	CA 80
409	1.720	0.011	HG 47	502	26.361	0.000	CG 58	595	27.275	0.018	CG 69	686	40.133	0.049	CB 80
410	117.775	0.050	N 47	503	7.960	0.003	H 58	596	4.440	0.008	HA 69	687	14.042	0.000	CD1 80
411	0.634	0.009	QD1 47	504	4.764	0.013	HA 58	597	1.958	0.005	HB2 69	688	18.610	0.000	CG2 80
412	0.697	0.010	QD2 47	505	1.621	0.006	HB2 58	598	2.370	0.002	HB3 69	689	8.468	0.006	H 80
413	58.952	0.001	CA 48	506	1.269	0.005	HB3 58	599	3.764	0.006	HD2 69	690	4.702	0.003	HA 80
414	28.890	0.004	CB 48	507	1.417	0.007	HG 58	600	3.237	0.004	HD3 69	691	1.883	0.003	HB 80
415	34.875	0.000	CG 48	508	117.425	0.053	N 58	601	2.065	0.004	QG 69	692	120.767	0.036	N 80
416	7.573	0.002	H 48	509	0.654	0.011	QD1 58	602	59.716	0.127	CA 70	693	0.451	0.008	QD1 80
417	3.864	0.005	HA 48	510	61.446	0.215	CA 59	603	28.650	0.000	CB 70	694	0.796	0.007	QG2 80
418	2.135	0.006	HB2 48	511	70.758	0.000	CB 59	604	36.101	0.000	CG 70	695	63.334	0.003	CA 81
419	2.044	0.013	HB3 48	512	22.454	0.000	CG2 59	605	9.240	0.003	H 70	696	35.033	0.041	CB 81
420	117.254	0.049	N 48	513	8.510	0.002	H 59	606	4.019	0.007	HA 70	697	24.909	0.065	CG 81
421	2.315	0.007	QG 48	514	4.396	0.008	HA 59	607	2.187	0.002	HG2 70	698	5.113	0.007	HA 81
422	59.829	0.030	CA 49	515	3.906	0.007	HB 59	608	2.321	0.009	HG3 70	699	1.989	0.011	HB2 81
423	29.693	0.039	CB 49	516	122.451	0.031	N 59	609	121.728	0.023	N 70	700	2.791	0.003	HB3 81
424	36.677	0.004	CG 49	517	1.188	0.012	QG2 59	610	1.885	0.005	QB 70	701	1.513	0.014	HG2 81
425	7.498	0.003	H 49	518	61.309	0.004	CA 60	611	67.925	0.040	CA 71	702	1.783	0.010	HG3 81
426	4.059	0.004	HA 49	519	33.159	0.003	CB 60	612	67.953	0.030	CB 71	703	3.615	0.013	QD 81
427	2.029	0.009	HB2 49	520	22.000	0.000	CG1 60	613	20.512	0.024	CG2 71	704	63.911	0.017	CA 82
428	2.182	0.010	HB3 49	521	20.865	0.050	CG2 60	614	9.444	0.004	H 71	705	73.148	0.075	CB 82
429	2.533	0.006	HG2 49	522	9.452	0.008	H 60	615	3.774	0.009	HA 71	706	21.417	0.107	CG2 82
430	2.121	0.004	HG3 49	523	4.925	0.011	HA 60	616	3.717	0.004	HB 71	707	5.022	0.010	HA 82
431	118.949	0.035	N 49	524	1.827	0.004	HB 60	617	119.177	0.012	N 71	708	3.918	0.005	HB 82
432	66.070	0.039	CA 50	525	128.019	0.125	N 60	618	0.432	0.006	QG2 71	709	1.087	0.012	QG2 82
433	38.343	0.000	CB 50	526	0.824	0.008	QG1 60	619	55.707	0.021	CA 72	710	53.091	0.000	CA 83
434	29.625	0.093	CG1 50	527	0.797	0.009	QG2 60	620	17.895	0.006	CB 72	711	42.011	0.000	CB 83
435	16.530	0.040	CG2 50	528	48.742	0.038	CA 61	621	6.925	0.004	H 72	712	26.431	0.024	CG 83
436	8.169	0.009	H 50	529	24.366	0.022	CB 61	622	3.932	0.019	HA 72	713	9.680	0.005	H 83
437	3.491	0.007	HA 50	530	9.321	0.006	H 61	623	120.366	0.015	N 72	714	6.089	0.002	HA 83
438	1.686	0.008	HB 50	531	5.487	0.006	HA 61	624	1.511	0.008	QB 72	715	1.351	0.007	HB2 83
439	118.486	0.023	N 50	532	129.933	0.022	N 61	625	59.163	0.080	CA 73	716	1.823	0.006	HB3 83
440	0.816	0.021	QD1 50	533	0.974	0.007	QB 61	626	30.012	0.000	CB 73	717	0.881	0.004	HG 83
441	1.858	0.007	QG1 50	534	55.192	0.043	CA 62	627	43.370	0.078	CD 73	718	127.551	0.082	N 83
442	0.627	0.011	QG2 50	535	34.618	0.049	CB 62	628	27.412	0.003	CG 73	719	60.201	0.010	CA 84
443	54.679	0.052	CA 51	536	28.643	0.011	CD 62	629	7.849	0.010	H 73	720	41.735	0.000	CB 84
444	19.160	0.026	CB 51	537	24.481	0.160	CG 62	630	4.105	0.006	HA 73	721	17.709	0.041	CD1 84
445	8.497	0.003	H 51	538	8.960	0.007	H 62	631	1.728	0.006	HG2 73	722	28.237	0.000	CG1 84
446	3.832	0.010	HA 51	539	4.970	0.008	HA 62	632	1.555	0.005	HG3 73	723	13.350	0.056	CG2 84
447	119.695	0.025	N 51	540	119.181	0.029	N 62	633	117.280	0.016	N 73	724	9.136	0.005	H 84
448	1.332	0.011	QB 51	541	1.733	0.003	QB 62	634	1.917	0.004	QB 73	725	5.039	0.008	HA 84
449	65.423	0.074	CA 52	542	1.326	0.006	QD 62	635	3.210	0.004	QD 73	726	1.760	0.022	HB 84
450	69.350	0.014	CB 52	543	1.162	0.011	QG 62	636	55.601	0.001	CA 74	727	120.368	0.048	N 84
451	21.510	0.068	CG2 52	544	53.557	0.018	CA 63	637	38.168	0.133	CB 74	728	0.148	0.005	QD1 84
452	7.563	0.006	H 52	545	44.364	0.000	CB 63	638	8.618	0.004	H 74	729	0.800	0.013	QG1 84
453	4.060	0.003	HA 52	546	25.451	0.055	CD2 63	639	4.216	0.003	HA 74	730	0.637	0.012	QG2 84
454	4.265	0.010	HB 52	547	28.071	0.006	CG 63	640	2.208	0.010	HB2 74	731	52.736	0.001	CA 85
455	110.660	0.035	N 52	548	9.000	0.004	H 63	641	2.656	0.008	HB3 74	732	44.058	0.000	CB 85
456	1.279	0.007	QG2 52	549	4.716	0.010	HA 63	642	120.076	0.045	N 74	733	23.718	0.064	CD1 85
457	59.040	0.109	CA 53	550	1.251	0.013	HB2 63	643	58.431	0.000	CA 75	734	27.658	0.000	CG 85
458	30.696	0.003	CB 53	551	0.783	0.008	HB3 63	644	38.105	0.000	CB 75	735	8.719	0.008	H 85
459	36.682	0.029	CG 53	552	1.310	0.012	HG 63	645	131.856	0.050	CD1 75	736	5.183	0.008	HA 85
460	8.578	0.005	H 53	553	126.402	0.087	N 63	646	131.891	0.050	CE1 75	737	1.871	0.011	HB2 85
461	4.097	0.007	HA 53	554	0.718	0.010	QD1 63	647	129.886	0.022	CZ 75	738	1.052	0.004	HB3 85
462	2.100	0.009	HB2 53	555	0.850	0.013	QD2 63	648	7.218	0.003	H 75	739	1.423	0.007	HG 85
463	1.996	0.005	HB3 53	556	52.420	0.026	CA 64	649	4.650	0.001	HA 75	740	129.362	0.051	N 85
464	2.274	0.005	HG2 53	557	40.566	0.041	CB 64	650	2.273	0.007	HB2 75	741	0.786	0.011	QD1 85
465	2.492	0.008	HG3 53	558	8.532	0.006	H 64	651	3.319	0.002	HB3 75	742	57.087	0.042	CA 86
466	121.274	0.039	N 53	559	3.861	0.011	HA 64	652	7.628	0.007	HZ 75	743	41.684	0.063	CB 86
467	52.424	0.000	CA 54	560	2.357	0.012	HB2 64	653	113.044	0.031	N 75	744	132.073	0.098	CD1 86
468	27.725	0.043	CB 54	561	2.487	0.007	HB3 64	654	7.501	0.012	QD 75	745	130.815	0.155	CE1 86
469	42.008	0.011	CD 54	562	126.637	0.058	N 64	655	7.448	0.011	QE 75	746	128.598	0.110	CZ 86
470	26.283	0.049	CG 54	563	64.463	0.030	CA 65	656	56.756	0.000	CA 76	747	9.859	0.006	H 86
471	8.328	0.002	H 54	564	31.372	0.000	CB 65	657	25.873	0.001	CB 76	748	5.187	0.004	HA 86
472	4.943	0.007	HA 54	565	18.710	0.024	CG1 65	658	34.265	0.000	CG 76	749	2.792	0.013	HB2 86
473	1.804	0.008	HB2 54	566	23.428	0.000	CG2 65	659	7.819	0.002	H 76	750	2.734	0.011	HB3 86
474	2.185	0.007	HB3 54	567	9.014	0.004	H 65	660	3.854	0.004	HA 76	751	5.790	0.005	HZ 86
475	3.197	0.007	HD2 54	568	3.816	0.008	HA 65	661	2.076	0.006	HB2 76	752	128.109	0.081	N 86
476	3.127	0.010	HD3 54	569	2.227	0.009	HB 65	662	2.196	0.003	HB3 76	753	6.825	0.008	QD 86
477	1.478	0.007	HG2 54	570	123.712	0.066	N 65	663	2.195	0.004	HG2 76	754	6.945	0.014	QE 86
478	1.621	0.004	HG3 54	571	0.932	0.010	QG1 65	664	2.310	0.002	HG3 76	755	56.019	0.000	CA 87
479	114.925	0.019	N 54	572	1.204	0.008	QG2 65	665	116.553	0.066	N 76	756	36.456	0.000	CB 87
480	55.476	0.000	CA 55	573	57.094	0.150	CA 66	666	8.390	0.012	H 77	757	29.668	0.000	CD 87
481	19.174	0.003	CB 55	574	40.647	0.111	CB 66	667	3.830	0.000	HA 77	758	42.172	0.080	CE 87
482	6.758	0.003	H 55	575	8.045	0.002	H 66	668	1.703	0.013	HB 77	759	25.193	0.107	CG 87
483	4.366	0.006	HA 55	576	4.553	0.009	HA 66	669	120.482	0.064	N 77	760	8.736	0.003	H 87
484	122.065	0.036	N 55	577	2.702	0.005	HB2 66	670	-0.046	0.009	QG2 77	761	4.618	0.006	HA 87
485	1.610	0.011	QB 55	578	2.904	0.011	HB3 66	671	61.669	0.174	CA 78	762	1.460	0.006	HB2 87
486	69.088	0.000	CB 56	579	119.681	0.036	N 66	672	33.349	0.036	CB 78	763	1.784	0.010	HB3 87
487	22.027	0.018	CG2 56	580	61.223	0.208	CA 67	673	18.675	0.066	CG1 78	764	119.947	0.075	N 87
488	8.201	0.002	H 56	581	69.637	0.009	CB 67	674	8.033	0.011	H 78	765	1.675	0.006	QD 87
489	4.165	0.000	HA 56	582	22.766	0.029	CG2 67	675	4.391	0.009	HA 78	766	3.049	0.005	QE 87
490	4.014	0.002	HB 56	583	6.870	0.002	H 67	676	2.228	0.004	HB 78	767	1.377	0.005	QG 87
491	107.259	0.023	N 56	584	4.403	0.012	HA 67	677	121.955	0.042	N 78	768	55.688	0.143	CA 88
492	1.225	0.003	QG2 56	585	4.315	0.007	HB 67	678	0.815	0.011	QQG 78	769	39.709	0.024	CB 88
493	55.276	0.000	CA 57	586	106.834	0.058	N 67	679	56.994	0.126	CA 79	770	9.444	0.007	H 88
494	44.260	0.000	CB 57	587	1.224	0.005	QG2 67					771	4.534	0.001	HA 88
495	7.772	0.002	H 57	588	40.448	0.077	CB 68					772	2.750	0.003	HB2 88

774	127.788	0.081	N 88												
775	46.187	0.029	CA 89												
776	9.156	0.004	H 89												
777	4.367	0.008	HA2 89												
778	3.855	0.005	HA3 89												
779	103.729	0.023	N 89												
780	52.038	0.050	CA 90												
781	30.836	0.064	CB 90												
782	32.783	0.018	CG 90												
783	7.861	0.002	H 90												
784	5.197	0.006	HA 90												
785	1.929	0.008	HB2 90												
786	2.157	0.008	HB3 90												
787	117.288	0.013	N 90												
788	2.435	0.008	QG 90												
789	61.840	0.055	CA 91												
790	30.886	0.040	CB 91												
791	50.447	0.017	CD 91												
792	27.680	0.042	CG 91												
793	3.878	0.012	HA 91												
794	1.259	0.006	HB2 91												
795	1.697	0.012	HB3 91												
796	2.450	0.007	HG2 91												
797	2.148	0.008	HG3 91												
798	3.983	0.008	QD 91												
799	62.061	0.103	CA 92												
800	33.332	0.036	CB 92												
801	21.979	0.000	CG1 92												
802	20.481	0.035	CG2 92												
803	8.982	0.004	H 92												
804	4.394	0.004	HA 92												
805	2.224	0.004	HB 92												
806	119.911	0.054	N 92												
807	0.991	0.005	QG1 92												
808	0.836	0.010	QG2 92												
809	55.782	0.141	CA 93												
810	36.090	0.098	CB 93												
811	42.059	0.113	CE 93												
812	25.019	0.000	CG 93												
813	7.194	0.002	H 93												
814	4.538	0.006	HA 93												
815	1.558	0.004	HB2 93												
816	1.810	0.007	HB3 93												
817	118.773	0.011	N 93												
818	3.065	0.012	QE 93												
819	1.356	0.011	QG 93												
820	54.972	0.009	CA 94												
821	33.173	0.002	CB 94												
822	43.369	0.067	CD 94												
823	28.070	0.002	CG 94												
824	8.828	0.006	H 94												
825	5.076	0.005	HA 94												
826	1.653	0.006	HB2 94												
827	1.690	0.005	HB3 94												
828	1.303	0.009	HG2 94												
829	1.379	0.014	HG3 94												
830	126.303	0.048	N 94												
831	3.200	0.005	QD 94												
832	60.780	0.007	CA 95												
833	40.922	0.000	CB 95												
834	14.432	0.133	CD1 95												
835	27.475	0.000	CG1 95												
836	17.847	0.000	CG2 95												
837	9.593	0.001	H 95												
838	4.347	0.008	HA 95												
839	1.826	0.005	HB 95												
840	127.824	0.028	N 95												
841	0.681	0.012	QD1 95												
842	1.053	0.009	QG1 95												
843	0.839	0.010	QG2 95												
844	62.070	0.068	CA 96												
845	33.415	0.003	CB 96												
846	20.586	0.000	CG1 96												
847	21.321	0.000	CG2 96												
848	8.898	0.002	H 96												
849	4.533	0.004	HA 96												
850	2.013	0.005	HB 96												
851	127.788	0.013	N 96												
852	0.949	0.014	QG1 96												
853	0.932	0.007	QG2 96												
854	43.717	0.038	CA 97												
855	8.013	0.003	H 97												
856	3.589	0.004	HA2 97												
857	4.394	0.013	HA3 97												
858	113.977	0.022	N 97												
859	52.512	0.023	CA 98												
860	19.273	0.004	CB 98												
861	8.359	0.003	H 98												
862	4.065	0.006	HA 98												
863	120.317	0.032	N 98												
864	1.299	0.010	QB 98												
865	54.282	0.035	CA 99												
866	36.401	0.014	CB 99												

867 28.825 0.049 CD 99 868 42.328 0.000 CE 99 869 24.548 0.000 CG 99 870 7.285 0.003 H 99 871 4.546 0.008 HA 99 872 1.516 0.006 HB2 99 873 1.823 0.007 HB3 99 874 1.431 0.001 HD2 99 875 1.590 0.002 HD3 99 876 2.788 0.003 HE2 99 877 2.867 0.001 HE3 99 878 1.261 0.002 HG2 99 879 1.358 0.000 HG3 99 880 120.852 0.017 N 99 881 43.980 0.027 CA 100 882 8.231 0.003 H 100 883 3.863 0.004 HA2 100 884 4.171 0.010 HA3 100 885 105.881 0.021 N 100 886 61.261 0.060 CA 101 887 33.361 0.079 CB 101 888 30.102 0.035 CD 101 889 41.807 0.039 CE 101 890 25.219 0.119 CG 101 891 8.658 0.003 H 101 892 3.549 0.010 HA 101 893 1.820 0.007 HB2 101 894 1.906 0.005 HB3 101 895 1.678 0.005 HD2 101 896 1.761 0.007 HD3 101 897 2.685 0.005 HE2 101 898 2.893 0.008 HE3 101 899 1.343 0.012 HG2 101 900 1.148 0.009 HG3 101 901 119.193 0.069 N 101 902 55.381 0.010 CA 102 903 17.648 0.037 CB 102 904 8.484 0.002 H 102 905 4.012 0.005 HA 102 906 118.726 0.045 N 102 907 1.411 0.010 QB 102 908 54.750 0.047 CA 103 909 18.703 0.083 CB 103 910 8.022 0.007 H 103 911 4.050 0.012 HA 103 912 120.668 0.024 N 103 913 1.442 0.006 QB 103 914 58.304 0.004 CA 104 915 41.804 0.000 CB 104 916 24.901 0.000 CD1 104 917 8.426 0.007 H 104 918 3.868 0.008 HA 104 919 1.422 0.004 HB2 104 920 1.613 0.006 HB3 104 921 1.498 0.001 HG 104 922 119.833 0.023 N 104 923 0.692 0.014 QD1 104 924 0.700 0.014 QD2 104 925 58.225 0.000 CA 105 926 40.868 0.047 CB 105 927 22.510 0.000 CD1 105 928 8.421 0.001 H 105 929 3.836 0.010 HA 105 930 1.891 0.003 HB2 105 931 1.390 0.010 HB3 105 932 117.284 0.040 N 105 933 0.782 0.008 QD1 105 934 0.862 0.005 QD2 105 935 59.670 0.126 CA 106 936 29.892 0.000 CB 106 937 43.311 0.052 CD 106 938 27.418 0.000 CG 106 939 7.899 0.004 H 106 940 4.021 0.010 HA 106 941 1.724 0.021 HG2 106 942 1.600 0.010 HG3 106 943 119.038 0.106 N 106 944 1.955 0.004 QB 106 945 3.211 0.003 QD 106 946 58.984 0.081 CA 107 947 29.632 0.044 CB 107 948 36.165 0.001 CG 107 949 7.899 0.002 H 107 950 4.126 0.007 HA 107 951 2.131 0.007 HB2 107 952 2.044 0.004 HB3 107 953 2.357 0.008 HG2 107 954 2.278 0.007 HG3 107 955 119.217 0.109 N 107 956 55.545 0.008 CA 108				

The first value in the list is the atom number associated with the chemical shift which is the second value with the standard deviation of chemical shifts assigned to that particular atom of a residue which is the fourth and fifth lines. The chemical shift list that will be used for structure calculation can only consist of assignments made on the NOESY peaklists (more on this later). This is done so that there is less inconsistency in chemical shifts.

3) Peaklists, the ones which are needed for structure calculation come from the three NOESY spectra (c13no.peaks, c13ar.peaks, and n15no.peaks). The protocol for generating these peaklists goes as follows.

-Open all sidechain spectra in Sparky (CBCA(CO)NH, CC(CO)NH, HBHA(CO)NH, HCC(CO)NH) as well as the HCCH-TOCSY if you made any assignments using it.

-Write the chemical shift list (.prot) in xeasy format using the command xe.

-Need to convert this chemical shift list into peaks that correspond to the HCCH-TOCSY. This is done so that we can easily assign and pick peaks in the c13 aliphatic NOESY. This is done on the Mac workstation in room 530 using the program called GARANT. The terminal command is:

garant genPeaks HCCH24 sequence_file_name prot_file_name new_peak list

-This peak can then be read on the HCCH-TOCSY peaklist by using the command cg/cy in Sparky. One should then go through and fix any peaks that are not on the actual peak.

-This peaklist should be saved in sparky format so that it can be read onto the ^{13}C aliphatic NOESY.

-Once this peaklist is read onto the 13C aliphatic NOESY, one should then go through all the planes and pick each peak that isn't already picked. Its best to do this using strips in Sparky with the command sp and click all assigned peaks which generate stripts according to residue number. This should give strips in a format so that w1 is the x-dimension (^{1}H), w2 is the y-dimension (^{1}H), and w3 in the z-dimension (^{13}C). This should yield around 2000-4000 peaks depending on the size of the protein.

-The ^{15}N-NOESY should also be hand picked with the only assignments being from the ^{1}H-^{15}N HSQC. This should yield around 1500-3000 peaks.

-The 13C aromatic NOESY can be assigned using either the 2D experiments, which correlate Hδ or Hη to the Cβ (cbcgcdhd/cbcgcdcehe experiments) or by actually assigning them based on the Hδ protons or observed in the aliphatic ^{13}C NOESY and working through the aromatic ring. The rest of the peak should be manually assigned.

After all the peaks are assigned and picked in the NOESY experiments, the peaklists are written using the command xe and should have the following format.

```
# Number of dimensions 3
#CYANAFORMAT HhC
  1 4.508 3.615 59.401  1 U  1.35e+05      0 e 0   0   0   0
  2 0.627 3.323 13.364  1 U   1.8e+04      0 e 0   0   0   0
  3 0.787 5.201 20.846  1 U  1.05e+04      0 e 0   0   0   0
  4 0.711 1.823 23.030  1 U   1.5e+05      0 e 0   0   0   0
  5 1.772 3.057 42.102  1 U   1.6e+05      0 e 0   0   0   0
  6 4.102 1.282 59.019  1 U  4.73e+04      0 e 0   0   0   0
  7 1.384 4.484 40.852  1 U    1e+04      0 e 0   0   0   0
  8 0.710 0.462 24.901  1 U  1.54e+05      0 e 0   0   0   0
  9 0.828 0.449 22.000  1 U  3.59e+04      0 e 0   0   0   0
 10 4.626 3.747 58.214  1 U  9.56e+03      0 e 0   0   0   0
 11 6.312 3.208 42.066  1 U  1.66e+03      0 e 0   0   0   0
 12 0.941 4.782 22.385  1 U   2.4e+04      0 e 0   0   0   0
 13 1.511 0.688 17.899  1 U  9.06e+04      0 e 0   0   0   0
 14 1.508 1.360 17.958  1 U  2.62e+05      0 e 0   0   0   0
 15 0.937 3.727 22.385  1 U  3.66e+04      0 e 0   0   0   0
 16 0.937 3.603 22.385  1 U  2.51e+04      0 e 0   0   0   0
 17 1.511 1.676 17.890  1 U  1.79e+05      0 e 0   0   0   0
 18 0.933 2.392 22.385  1 U   2.5e+04      0 e 0   0   0   0
 19 4.497 3.734 64.083  1 U  1.32e+04      0 e 0   0   0   0
 20 4.499 3.573 64.083  1 U  2.01e+04      0 e 0   0   0   0
 21 1.938 3.983 30.931  1 U   2.7e+04      0 e 0   0   0   0
 22 4.499 1.718 64.083  1 U  1.13e+04      0 e 0   0   0   0
 23 2.330 3.715 32.285  1 U   4.3e+04      0 e 0   0   0   0
 24 4.130 2.801 64.123  1 U  1.07e+04      0 e 0   0   0   0
 25 4.139 3.871 64.011  1 U   5.3e+04      0 e 0   0   0   0
```

```
3811 0.923 2.081 22.112  1 U  4.76e+05      0 e 0  985  983  979
3812 0.923 0.923 22.112  1 U  1.03e+07      0 e 0  985  985  979
3813 0.930 0.930 22.385  1 U  8.72e+06      0 e 0  985  985  979
3814 0.925 0.917 21.740  1 U  1.04e+07      0 e 0  985  985  979
3815 0.925 0.930 21.740  1 U  1.04e+07      0 e 0  985  985  979
3816 0.930 0.917 22.385  1 U  8.72e+06      0 e 0  985  985  979
3817 0.923 0.893 22.112  1 U  1.03e+07      0 e 0  985  986  979
3818 0.904 1.807 20.824  1 U  1.23e+05      0 e 0  986  814  980
3819 0.899 3.062 20.868  1 U  6.08e+04      0 e 0  986  816  980
3820 0.889 4.169 20.765  1 U  8.14e+04      0 e 0  986  959  980
3821 0.893 4.398 20.765  1 U  1.31e+05      0 e 0  986  982  980
3822 0.893 2.081 20.765  1 U  2.73e+05      0 e 0  986  983  980
3823 0.893 0.923 20.765  1 U   7.8e+06      0 e 0  986  985  980
3824 0.893 0.893 20.765  1 U   7.8e+06      0 e 0  986  986  980
3825 0.868 0.868 21.120  1 U  6.48e+06      0 e 0  986  986  980
3826 4.502 0.765 59.401  1 U  2.55e+04      0 e 0  992  438  987
3827 4.504 0.672 59.401  1 U  1.07e+04      0 e 0  992  964  987
3828 4.503 4.524 59.401  1 U  5.18e+05      0 e 0  992  992  987
3829 4.503 2.090 59.401  1 U  4.92e+04      0 e 0  992  993  987
3830 4.503 0.917 59.401  1 U  2.34e+05      0 e 0  992  995  987
3831 4.503 0.930 59.401  1 U  2.34e+05      0 e 0  992  995  987
3832 4.503 0.917 59.401  1 U  2.34e+05      0 e 0  992  996  987
3833 4.503 0.930 59.401  1 U  2.34e+05      0 e 0  992  996  987
3834 4.504 1.959 59.401  1 U    4e+04      0 e 0  992 1003  987
3835 2.102 1.680 32.124  1 U   6.4e+04      0 e 0  993  436  988
3836 2.089 4.156 32.124  1 U  3.16e+04      0 e 0  993  959  988
3837 2.114 0.694 32.124  1 U  3.78e+04      0 e 0  993  964  988
```

The top two lines are needed so that CYANA can properly interpret the peaklists (#CYANAFORMAT HhN for the ^{15}N NOESY) and are added manually. The lines are the chemical shifts of the peaks with the inensity/volume of the peak. If the peaks are assigned (as in the right panel), the numbers correspond to the atoms from the chemical shift list. These peaklists should be saved with a name that tells the nature of the experiment (i.e. c13no.peaks, n15no.peaks, c13ar.peaks).

4) Dihedral restraints are generated using Talos+, which come from the chemical shifts of the Cα, Cβ, Hα, Hβ, CO, and HN atoms from the sidechain, HNCO, and ^{1}H-^{15}N HSQC experiments. To generate the dihedral constraints go to the Mac workstation and in the terimanl type;

-cyana
-read prot name.prot
-taloslist newname.tab

which creates the input needed for Talos called newname.tab. Dihedral constraints are generated by typing;

-talos+ -in newname.tab

You can look at the constraint values by typing;

-rama+ -in newname.tab

To generate the file that can be read by Cyana type, copy the file talos+2aco into the file and type;

-talos+2aco pred.tab > name.aco

This file has the format;

3 ASP PHI -76.7 -45.1	40 CYS PSI -48.6 -28.6	72 ALA PSI -52.6 -32.6	107 GLU PHI -74.6 -54.6
3 ASP PSI -63.5 0.5	41 LYS PHI -71.2 -51.2	73 ARG PHI -77.3 -57.3	107 GLU PSI -47.5 -27.5
4 SER PHI -96.1 -44.3	41 LYS PSI -49.9 -29.9	73 ARG PSI -49.1 -23.0	108 LEU PHI -97.8 -62.4
4 SER PSI -50.0 -16.4	42 MET PHI -75.1 -55.1	74 ASN PHI -76.2 -54.0	108 LEU PSI -45.1 -9.4
7 SER PHI -126.5 -60.4	42 MET PSI -50.1 -30.1	74 ASN PSI -44.2 -23.6	109 SER PHI -74.5 -54.5
7 SER PSI 107.7 161.9	43 VAL PHI -93.6 -68.5	75 PHE PHI -96.5 -76.3	109 SER PSI -46.6 -26.6
8 ALA PHI -105.7 -74.2	43 VAL PSI -39.5 -17.0	75 PHE PSI -14.2 9.7	110 ASP PHI -82.8 -60.2
8 ALA PSI 109.1 153.7	44 ALA PHI -61.3 -41.3	77 VAL PHI -139.0 -23.8	110 ASP PSI -44.9 -22.8
9 THR PHI -141.5 -89.7	44 ALA PSI -60.9 -40.9	77 VAL PSI 84.2 149.5	111 VAL PHI -123.8 -89.2
9 THR PSI 152.4 178.3	45 PRO PSI -42.1 -27.4	78 VAL PHI -147.3 -80.1	111 VAL PSI -16.4 11.5
10 ILE PHI -183.6 -59.7	46 VAL PHI -74.9 -54.9	# 78 VAL PSI 125.1 174.6	112 VAL PHI -138.3 -82.2
10 ILE PSI 125.1 154.6	46 VAL PSI -51.2 -31.2	# 79 SER PHI -128.6 -84.7	112 VAL PSI 67.3 122.7
11 LYS PHI -135.5 -83.7	47 LEU PHI -72.6 -52.6	79 SER PSI 136.3 166.8	114 ASN PHI -96.8 -50.1
11 LYS PSI 108.4 145.0	47 LEU PSI -49.3 -29.3	# 80 ILE PHI -144.8 -101.2	114 ASN PSI -50.0 -0.8
12 VAL PHI -139.4 -106.8	48 GLU PHI -71.6 -51.6	80 ILE PSI 76.2 146.5	
12 VAL PSI 151.0 171.0	48 GLU PSI -50.5 -30.5	81 PRO PSI 142.4 161.6	
13 THR PHI -123.0 -82.8	49 GLU PHI -75.1 -55.1	82 THR PHI -135.1 -109.7	
13 THR PSI 153.8 173.8	49 GLU PSI -55.2 -35.2	82 THR PSI 121.6 146.9	
14 ASP PHI -71.6 -51.6	50 ILE PHI -74.1 -54.1	83 LEU PHI -120.4 -96.8	
14 ASP PSI -50.7 -28.2	50 ILE PSI -53.4 -33.4	83 LEU PSI 119.7 139.8	
15 ALA PHI -84.2 -59.6	51 ALA PHI -70.2 -50.2	84 ILE PHI -136.4 -116.4	
15 ALA PSI -53.6 -13.9	51 ALA PSI -49.9 -29.9	84 ILE PSI 122.4 154.9	
17 PHE PHI -64.5 -44.5	52 THR PHI -75.4 -55.4	85 LEU PHI -123.2 -92.7	
17 PHE PSI -62.2 -42.2	52 THR PSI -48.8 -28.8	85 LEU PSI 106.8 132.8	
18 ALA PHI -69.8 -49.8	53 GLU PHI -77.6 -57.6	86 PHE PHI -122.7 -97.2	
18 ALA PSI -52.8 -32.8	53 GLU PSI -46.8 -21.9	86 PHE PSI 130.0 156.5	
19 THR PHI -79.0 -59.0	54 ARG PHI -128.3 -82.1	87 LYS PHI -154.0 -100.5	
# 19 THR PSI -51.9 -31.9	54 ARG PSI -5.9 35.5	87 LYS PSI 111.9 137.0	
# 20 ASP PHI -77.3 -55.6	55 ALA PHI -71.1 -51.1	89 GLY PHI 68.0 88.0	
20 ASP PSI -53.6 -33.1	55 ALA PSI -49.9 -28.5	89 GLY PSI -6.5 15.6	
21 VAL PHI -78.7 -58.7	56 THR PHI -84.0 -64.0	90 GLN PHI -125.9 -95.3	
21 VAL PSI -51.0 -24.0	56 THR PSI -46.2 -20.1	90 GLN PSI 108.6 151.8	
22 LEU PHI -73.3 -53.3	57 ASP PHI -116.9 -85.5	91 PRO PSI 140.4 156.7	
22 LEU PSI -47.5 -13.6	57 ASP PSI -39.1 -4.3	93 LYS PHI -163.1 -142.3	
23 SER PHI -104.4 -62.8	58 LEU PHI -163.5 -130.0	93 LYS PSI 132.6 152.6	
23 SER PSI -32.7 -11.3	58 LEU PSI 128.7 148.7	94 ARG PHI -128.4 -91.8	
24 SER PHI -139.4 -59.8	59 THR PHI -136.7 -109.2	94 ARG PSI 112.8 141.4	
24 SER PSI 84.7 150.4	59 THR PSI 114.9 138.1	95 ILE PHI -129.3 -108.0	
26 LYS PHI -141.9 -90.7	60 VAL PHI -128.5 -99.7	95 ILE PSI 122.9 145.1	
26 LYS PSI 102.5 166.1	60 VAL PSI 118.0 138.0	96 VAL PHI -133.2 -101.8	
27 PRO PSI 139.1 157.4	61 ALA PHI -143.2 -110.6	96 VAL PSI 115.5 143.8	
28 VAL PHI -135.6 -105.7	61 ALA PSI 137.1 165.4	97 GLY PHI -203.8 -73.5	
28 VAL PSI 110.9 140.6	62 LYS PHI -130.5 -104.6	97 GLY PSI 150.2 181.8	
29 LEU PHI -125.5 -93.7	62 LYS PSI 115.8 135.8	98 ALA PHI -119.5 -59.3	
29 LEU PSI 114.3 134.3	63 LEU PHI -134.4 -88.5	98 ALA PSI 102.6 158.0	
30 VAL PHI -119.4 -99.4	63 LEU PSI 117.3 143.5	99 LYS PHI -130.0 -109.0	
30 VAL PSI 115.7 144.4	64 ASP PHI -115.4 -81.9	99 LYS PSI 131.6 175.8	
31 ASP PHI -106.7 -84.4	64 ASP PSI 84.7 127.6	100 GLY PHI -146.2 -20.2	
31 ASP PSI 110.0 130.0	65 VAL PHI -70.6 -50.6	100 GLY PSI 157.0 177.0	
32 PHE PHI -100.3 -69.2	65 VAL PSI -38.0 -17.1	101 LYS PHI -65.2 -45.1	
32 PHE PSI 109.5 139.0	66 ASP PHI -78.5 -58.5	101 LYS PSI -54.3 -34.3	
33 TRP PHI -133.9 -110.4	66 ASP PSI -50.1 -19.9	102 ALA PHI -70.9 -50.9	
33 TRP PSI 148.2 172.5	67 THR PHI -118.5 -93.0	102 ALA PSI -53.1 -33.1	
34 ALA PHI -169.2 -107.3	67 THR PSI -17.9 15.6	103 ALA PHI -75.2 -55.2	
34 ALA PSI 135.3 168.5	68 ASN PHI -123.0 -63.7	103 ALA PSI -50.7 -30.7	
37 CYS PHI -136.5 -33.6	68 ASN PSI 78.0 141.4	104 LEU PHI -74.5 -54.5	
37 CYS PSI 83.9 196.8	70 GLU PHI -77.0 -57.0	104 LEU PSI -55.0 -35.0	
38 GLY PHI -109.4 61.9	70 GLU PSI -46.1 -26.1	105 LEU PHI -75.5 -53.9	
38 GLY PSI -210.7 -123.7	71 THR PHI -76.6 -56.6	105 LEU PSI -45.6 -25.6	
39 PRO PSI -42.4 -23.5	71 THR PSI -50.7 -30.7	106 ARG PHI -74.0 -54.0	
40 CYS PHI -73.4 -53.4	72 ALA PHI -72.7 -52.7	106 ARG PSI -53.7 -33.7	

The files needed for Cyana have now been generated (n15no.peaks, c13no.peaks, c13ar.peaks, nameofprot.seq, and nameofprot.prot).

B) To start structure calculation one has to perform the automatic assignment of the peaklists using the noeassign macro of Cyana.

1) Cyana is ran using the Condor submission system of Pere. So one has to log into Pere and transfer the files from above to a new directory in Pere.

2) Transfer files from the shared folder (NMR_STRUCTURE_INFO) in the Mac workstation to that directory.

-init.cya
-autorun.cya
-autorun.sub

3) The file init.cya is the initialization script for Cyana

```
# This file is read automatically when starting CYANA.
name:=trxc            # protein name, used for output file names
rmsdrange:=7..110       # residue range for RMSD calculations

cyanalib              # read standard library
read seq trxc         # read protein sequence
```

Input the name of the protein as well as the rmsd range used to determine rmsd as well as the name of the sequence file. This script is immediately called up when typing starting the Cyana program.

4) The file autorun.cya is the script that reads in the files that Cyana uses for automatic assignment (all those that were prepared previously).

```
nproc=8
peaks       := c13no,n15no,c13ar     # names of NOESY peak lists
prot        := trxc.prot                    # names of chemical shift lists
constraints := trxc.aco   # additional (non-NOE) constraints
tolerance   := 0.030,0.030,0.4          # chemical shift tolerances
calibration :=                          # NOE calibration parameters
structures  := 100,20                   # number of initial, final structures
steps       := 15000                    # number of torsion angle dynamics steps
rmsdrange   := 7..110                    # residue range for RMSD calculation
randomseed  := 450000                   # random number generator seed

noeassign peaks=$peaks prot=$prot autoaco
```

Input the name of your peaklists (peaks), chemical shift list (prot), and any additional constraints (dihedral, hbonds, or ssbonds). The rmsd range should be the same as the init file. The very last line of autorun.cya is the the noeassign macro and the flags needed to run automatic assignment (autoaco is optional to allow temporary angle restraints to favor Ramachandran plot).

5) To submit this to Condor on Pere to run noeassign of Cyana, one has to use autorun.sub.

```
Universe   = vanilla
Executable = /cluster/cyana/pgi/2.1/cyana
Arguments  = autorun.cya
Log        = cyana.log
Output     = autorun.out
Error      = autorun.error
Queue
```

Type condor_submit autorun.sub

6) The output of noeassign is 7 cycles of refinement with one final refinement. After every cycle a pdb file is written with an overview (.ovw) file summarizing the structural violations, target function, and rmsd. With cycle 7, assigned peaklists and new prot file are generated (name-cycle7.peaks and name-final.prot). The command cyanatable (shown below) summarizes the cycles.

Cycle	:	1	2	3	4	5	6	7	
Peaks:									
selected	:	6353	6353	6353	6353	6353	6353	6353	
assigned	:	5365	5292	5161	5163	5116	5072	5053	
unassigned	:	988	1061	1192	1190	1237	1281	1300	
with diagonal assignment	:	692	692	692	692	692	692	692	
Cross peaks:									
with off-diagonal assignment	:	4673	4600	4469	4471	4424	4380	4361	
with unique assignment	:	1182	2899	3208	3303	3549	3679	3687	
with short-range assignment |i-j|<=1	:	3080	3052	2985	2957	2912	2873	2871	
with medium-range assignment 1<|i-j|<5	:	803	722	645	659	655	649	653	
with long-range assignment |i-j|>=5	:	790	826	839	855	857	858	837	
Comparison with reference assignment:									
Cross peaks with reference assignment	:	4465	4465	4465	4465	4465	4465	4465	
with identical reference assignment	:	1136	2656	2950	3028	3193	3273	3271	
with compatible reference assignment	:	4152	4273	4145	4123	4016	3930	3902	
with incompatible reference assignment	:	166	139	198	228	301	356	368	
with additional reference assignment	:	145	51	120	112	146	177	193	
with additional assignment	:	355	188	126	120	107	94	91	
Upper distance limits:									
total	:	2902	2576	2403	2357	2283	2228	2282	2361
short-range, |i-j|<=1	:	1496	1336	1250	1194	1152	1116	1074	1125
medium-range, 1<|i-j|<5	:	1084	901	509	515	500	486	504	515
long-range, |i-j|>=5	:	322	339	644	648	631	626	704	721
Average assignments/constraint	:	4.95	2.12	1.30	1.26	1.18	1.14	1.00	1.00
Average target function value	:	288.82	100.24	123.18	17.53	9.73	6.51	7.33	3.36
RMSD (residues 7..110):									
Average backbone RMSD to mean	:	1.49	0.55	0.26	0.21	0.22	0.25	0.26	0.18
Average heavy atom RMSD to mean	:	1.87	0.93	0.74	0.68	0.71	0.75	0.70	0.61

As long as the initial target function is under 300 and rmsd is under 3, as well as having a cycle 7 target function under 10, the cycle 7 structure and assignments are suitable to go onto manual refinement. If any of these is violated, one can look at the ovw file from cycle 7 and look for any really violated distance or dihedral restraints and delete those and run again. Generally deleting the dihedral constraints is good enough to decrease the target function. Automatic peak picking of NOESY peaklists can lead to a lot of peaks being thrown out because they aren't real which leads very high target functions. This is why manual peak picking is preferred. The upper distance limits is the number of NOEs used in the automatic assignments.

C) Manual refinement involves assigning the peaks that noeassign missed.

1) In a new directory copy the files cycle7.pdb, all name-cycle7.peaks, name-final.prot, *.aco, *.seq, stereofound.cya, and init.cya.

2) (optional) Delete the ambiguous restraints the Cyana gave some peaks (if it wasn't sure it assigned more than one assignment to a peak) with the command;

- grep –v '0 e 0' name.peaks > newname.peaks

3) In sparky bring up the clean (no peaks) NOESY spectra and put the peaklists that noeassign from Cyana assigned onto them from step 2 by using the command cy/cg

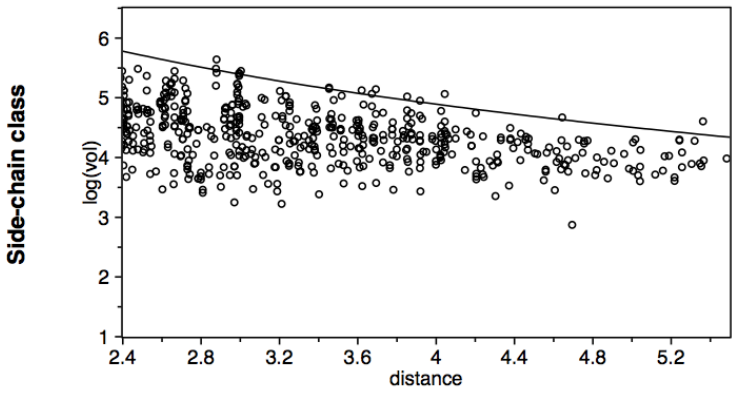
4) As in Step 1-3 generate strips of the particular NOESY spectra and go through each strip and make as many assignments as possible (using the cycle7.pdb as guide for close residues, since they have to at least be within 5 Å). Go through about 10 residues and and run through manual Cyana structure calculation.

5) Copy from the shared folder (NMR_STRUCTURE_INFO) in the Mac workstation the scripts to run manual refinement.
- man_calc.cya
- calc.cya
- calc.sub

Where man_calc.cya is the script to calibrate the peaklists and run the structure refinement in the initial process.

```
# Calibration

read prot trxc
read pdb cycle7

#the basic rule
# sc       = bb *0.17
# methyl   = sc *0.33

#peak n15no
read peaks n15no.peaks assigned integrated

#caliba avedis=3.7 plot=noen
caliba vmin=100 dmin=2.4 dmax=5.5 bb=5E+08 sc=3E+07 methyl=0.5E+07 plot=noen1

#peak c13no
read peaks c13no.peaks assigned integrated
#caliba avedis=3.7 plot=noec
caliba vmin=100 dmin=2.4 dmax=5.5 bb=3E+08 sc=2E+07 methyl=0.5E+07 plot=noec1

#peak c13ar
read peaks c13ar.peaks assigned integrated
#caliba avedis=3.7 plot=noec
caliba vmin=100 dmin=2.4 dmax=5.5 bb=5.00E+06 sc=7.00E+07 methyl=3.20E+07 plot=noea1


distance unique
write upl caliba_NC.upl

## 8DEMOS: GridSearch - Systematic local conformation analysis
##
## Local conformational data (NOE upper distance bounds and vicinal scalar
## coupling constants for phi/psi/chi1(/chi2) fragments are analyzed
## by a grid search using the FOUND algorithm in order to determine
## torsion angle restraints and stereospecific assignments.
##
## FOUND: Guntert et al. (1998) J. Biomol. NMR 12, 543-548.
## HABAS: Guntert et al. (1989) J. Amer. Chem. Soc. 111, 3997-4004.

read upl caliba_NC.upl                 # read noe upper distance limits
karplus                                # use standard Karplus curves (from karplus.cya)
read aco trxc.aco
expand=0
swap=0
habas range=2..108  angles="CHI1 CHI2*" tfcut=0.05   # alternative phi/psi/chi1/chi2 grid search
gridplot habas.ps
atom stereo list
write upl swapped.upl                  # write upper limits with swapped stereo pairs
write aco gridsearch.aco               # write upper limits with swapped stereo pairs
savestereo stereofound.cya             # write macro with stereospecific assignments
distance modify
write upl found-c.upl
write aco found-c.aco

# ---- structure calculation ----

read upl found-c.upl                          # read upper distance limits
read aco trxc.aco                           # read angle constraints
nproc=10
seed=34051
#cut_upl=0.05                                    # random number generator seed
expand=0
swap=0
dis stat
flip
calc_all structures=100  command=anneal steps=20000    # calculate 100 conformers
overview $name.ovw structures=20   pdb cor          # write overview file and coordinates

structures violate significant=20 delete=30
write upl filter.upl_
```

Where the macro caliba is used to calibrate the peaklists. At the end of caliba post script file is generated called noesyname.ps. The values must be altered until the post script file looks like this.



The line must be just barely above the dots for it to be calibrated properly. You get a ps file for each peaklist with backbone class, sidechain class, and methyl class, all of which must look like the figure above. If not calibrated properly the target function will go way up.

After the peaklists are calibrated properly, look at the overview file and delete any severely violated distance restraints and run again…repeat until all the bad restraints are gone and decent target functions remains. Copy these peaklists to the NOESY spectra using cy/cg command and repeat this process until all the spectra have gone through and many more assignments have been made.

In the case of the ^{15}N-NOESY look for NOEs that define secondary structure such as alpha helix peakse and beta sheets. For alpha helices the NH NOEs can be seen from i to i+4 residue.

6) Every now and again, its worth it to upload all the peaklists and prot files to another directory and run autoassignment again. This may make the structure look a little better. Treat the autoassigned peaklists like earlier and quickly go through the strips you have already gone through and make sure those peaks are still assigned.

7) When you are all done making assignments replace the man_calc.cya for calc.cya.

```
# ---- structure calculation ----

read upl found-c.upl                              # read upper distance limits
read aco trxc.aco                               # read angle constraints
nproc=10
seed=34051
cut_upl=0.05                                     # random number generator seed
expand=0
swap=0
dis stat
flip
calc_all structures=100  command=anneal steps=20000    # calculate 100 conformers
overview $name.ovw structures=20   pdb cor          # write overview file and coordinates

structures violate significant=5 delete=100
write upl filter.upl
```

8) Calc.cya is only the structure calculation script from man_calc.cya. This script is used to get your final structure and throw out NOEs to lower your target function to an acceptable value.

9) After you have gone through all your strips and added as many NOEs as possible you have to replace man_calc.cya with this script and change calc.sub so that it matches calc.cya.

10) Run this script but pay attention to the last two lines in this script;

Structures violate significant=15 delete=100
Write upl filter.upl

Structures violate will delete NOEs automatically that were violated upon structure calculation from the upl file generated. Upl file is just the list of upper distance limits (NOEs). This will not delete from the peaklist. Significant is the number of models in which an NOE is violated. So if an NOE is violated in 15 of the 20 structures it is considered significant. Delete is the percentage of NOEs that are significantly violated that are deleted. In this case any NOE that is violated in 15 or more structures it is deleted from the upl file and the new upl file is written to filter.upl. (significant should start at 20 and move to 15)

11) Change the name of filter.upl to found-c.upl.

mv filter.upl found-c.upl

12) Repeat until target function is between 0-3.

13) Once this is achieved, change calc.cya;

So that 500 structures are calculated and the 99 best structures are written out.

```
# ---- structure calculation ----

read upl found-c.upl                              # read upper distance limits
read aco trxc.aco                                # read angle constraints
ssbond 37-40
hbond O 37 H 41
write upl trxc_ox.upl
distance unique
nproc=10
seed=34051
cut_upl=0.05                                       # random number generator seed
expand=0
swap=0
dis stat
flip
calc_all structures=500  command=anneal steps=20000    # calculate 100 conformers
overview $name.ovw structures=99   pdb cor           # write overview file and coordinates

structures violate significant=20 delete=100
write upl filter.upl
```

D) This last structure of 99 conformers is then put through water refinement using simulated annealing and the AMBER molecular force field.

In the same directory as before, copy the files;

-mdin
-mdin_rst
-amber_anneal.sub
-amber_min1.sub
-amber_min2.sub
-anneal.in
-generate.inp
-refine.inp
-final_min.inp

1) generate.inp splits the 99 structures into individual files and performs and initial minimization on the structure in implicit solvent using the script mdin and submitted to Condor using amber_min1.sub.
(type ./generate.inp to run)

generate.inp

```
#! /bin/sh

grep -v 'ATOM      2  H   MET A   1' trxc.pdb > trxc_1.pdb

awk -f  /home/MARQNET/olsona/scripts/amber/splitpdb trxc_1.pdb


set -e
for f in model_0*
do
     d=${f%.*}
     if [ ! -d "$d" ]; then
        mkdir "$d"
     fi
     mv -f "$f" "$d"
done

for f in model_0*
do
cd $f
mv *pdb model.pdb
cp ../leap.in .
cp /home/MARQNET/olsona/scripts/amber/amber_min1.sub .
cp ../mdin .
tleap -s -f leap.in
condor_submit amber_min1.sub
cd ..
done
```

mdin

```
energy minimization for trx starting structures

 &cntrl

imin=1, maxcyc=1000, ncyc=500, ntpr=20,
        ntb=0, igb=1,cut=10.0
 /
```

amber_min1.sub

```
Universe   = vanilla
Executable = /cluster/amber/amber11/exe/sander
Arguments  = -O -i mdin -o model_min.out -c model.x -p prmtop -r model_min.x
Log        = amber.log
Output     = amber.out
Error      = amber.error
Queue
```

This initial minimization takes about 5-10 minutes to complete.

2) The output from the initial minimization (model_min.x) is fed into the simulated annealing refinement with the script refine.inp which calls the annealing script anneal.in and is submitted to Condor using amber_anneal.sub.
(type ./refine.inp to run)
refine.inp

```
#!/bin/sh


for f in model*
do
cd $f
cp ../found-c.upl .
cp ../trxc.aco .
sed -e "s/ H / HN /ig" found-c.upl > tempfile.tmp
mv tempfile.tmp found-c.upl
sed -e "s/ QE / HE /ig" found-c.upl > tempfile.tmp
mv tempfile.tmp found-c.upl
sed -e "s/ QD / HD /ig" found-c.upl > tempfile.tmp
mv tempfile.tmp found-c.upl
sed -e "s/ PRO  HD / PRO  QD /ig" found-c.upl > tempfile.tmp
mv tempfile.tmp found-c.upl
sed -e "s/ LYS  HD / LYS  QD /ig" found-c.upl > tempfile.tmp
mv tempfile.tmp found-c.upl
sed -e "s/ ARG  HD / ARG  QD /ig" found-c.upl > tempfile.tmp
mv tempfile.tmp found-c.upl
sed -e "s/ LYS  HE / LYS  QE /ig" found-c.upl > tempfile.tmp
mv tempfile.tmp found-c.upl
ambpdb -aatm -p prmtop < model.x > model.amb.pdb
cd ..
done

for f in model*
do
cd $f
makeDIST_RST -upb found-c.upl -pdb model.amb.pdb -rst RST.dist
cd ..
done

for f in model*
do
cd $f
makeANG_RST -pdb model.amb.pdb -con trxc.aco -lib $AMBERHOME/src/nmr_aux/prepare_input/tordef.lib >  RST.bb &
cd ..
done

for f in model*
do
cd $f
cat RST.bb RST.dist > RST
echo "cat: " RST
cd ..
done

for f in model*
do
cd $f
sed -e "s/rk2 =   2.0, rk3 =   2.0,/rk2 =   50.0, rk3 =   50.0,/ig" RST > tempfile.tmp
mv tempfile.tmp RST
sed -e "s/rk2=20.0, rk3=20.0,/rk2=25.0, rk3=25.0,/ig" RST > tempfile.tmp
mv tempfile.tmp RST
echo "Modified: " RST
cd ..
done



for f in model*
do

cd $f
cp ../amber_anneal.sub .
cp ../anneal.in .
condor_submit amber_anneal.sub
cd ..
done
```

The first part of refine.inp changes the atom names of some of the atoms in the upl file so that AMBER can convert this to restraints it can use. The second part converts the restraint files so that AMBER can use them and applies force constants to the restraints which can be altered. The last part is condor submission.

anneal.in

```
 simulated annealing protocol, 30 ps

  &cntrl
     nstlim=60000, pencut=-0.01, nmropt=1,
     ntpr=200, ntt=1, ntwx=200,
     cut=12.0, ntb=0, vlimit=12, rgbmax=12.0,
     igb=1, saltcon=0.2,
  /
  &ewald
  /
 #
 #Simple simulated annealing algorithm:
 #
 #from steps 0 to 5000: heat the system to 600K
 #from steps 5001-18000: re-cool to low temperatures with long tautp
 #from steps 18001-20000: final cooling with short tautp
 #
  &wt type='TEMP0', istep1=0,istep2=3000,value1=10.0,
             value2=1000.,    /
  &wt type='TEMP0', istep1=3001, istep2=8000, value1=1000.0,
             value2=1000.0,     /
  &wt type='TEMP0', istep1=8001, istep2=30000, value1=0.0,
             value2=0.0,     /
  &wt type='TEMP0', istep1=30000, istep2=33000,value1=10.0,
             value2=1000.,    /
  &wt type='TEMP0', istep1=33001, istep2=38000, value1=1000.0,
             value2=1000.0,     /
  &wt type='TEMP0', istep1=38001, istep2=60000, value1=0.0,
             value2=0.0,     /

  &wt type='TAUTP', istep1=0,istep2=8000,value1=0.2,
             value2=0.2,     /
  &wt type='TAUTP', istep1=8001,istep2=24000,value1=4.0,
             value2=2.0,     /
  &wt type='TAUTP', istep1=24001,istep2=27000,value1=1.0,
             value2=1.0,     /
  &wt type='TAUTP', istep1=27001,istep2=29000,value1=0.5,
             value2=0.5,    /
  &wt type='TAUTP', istep1=29001,istep2=30000,value1=0.05,
             value2=0.05,    /
  &wt type='TAUTP', istep1=30000,istep2=38000,value1=0.2,
             value2=0.2,     /
  &wt type='TAUTP', istep1=38001,istep2=54000,value1=4.0,
             value2=2.0,     /
  &wt type='TAUTP', istep1=54001,istep2=57000,value1=1.0,
             value2=1.0,     /
  &wt type='TAUTP', istep1=57001,istep2=59000,value1=0.5,
             value2=0.5,    /
  &wt type='TAUTP', istep1=59001,istep2=60000,value1=0.05,
             value2=0.05,    /

  &wt type='REST', istep1=0,istep2=8000,value1=0.1,
             value2=1.00,  /
  &wt type='REST', istep1=8001,istep2=30000,value1=1.00,
             value2=1.00,  /
  &wt type='REST', istep1=30000,istep2=38000,value1=0.1,
             value2=1.00,  /
  &wt type='REST', istep1=38001,istep2=60000,value1=1.00,
             value2=1.00,  /

  &wt type='END'  /
 LISTOUT=POUT
 DISANG=RST
```

This script defines the temperatures and step values used. This particular file heats up to 1000K and goes through 2 cycles of 30 ps simulated annealing sessions. Restraints are used with this.

amber_anneal.sub

```
Universe   = vanilla
Executable = /cluster/amber/amber11/exe/sander
Arguments  = -O -i anneal.in -o model_ann.out -c model_min.x -p prmtop -r model.ann.x -x model.ann.traj
Log        = amber.log
Output     = amber.out
Error      = amber.error
Queue
```

This portion of water refinement takes about 4 hours to complete.
If for some reason this fails, one can try and decrease the temperature and if that still fails look at the trajectory file and see where its failing.

3) The output from this file (model.ann.x) is fed into the last portion of water refinement, which is a final minimization using restraints in implicit solvent, which is uses the scripts final_min.inp which calls up mdin_rst and submits to condor using amber_min2.sub. (./final_min.inp is how to run this)
final_min.inp

```
#!/bin/sh


for f in model_0*
do
        cd $f
        cp ../amber_min2.sub .
        cp ../mdin_rst .
        condor_submit amber_min2.sub
        cd ..
done
[olsona@hpc-hn1 9]$ more final_min.inp
#!/bin/sh


for f in model_0*
do
        cd $f
        cp ../amber_min2.sub .
        cp ../mdin_rst .
        condor_submit amber_min2.sub
        cd ..
done
```

mdin_rst

```
#!/bin/sh


for f in model_0*
do
        cd $f
        cp ../amber_min2.sub .
        cp ../mdin_rst .
        condor_submit amber_min2.sub
        cd ..
done
[olsona@hpc-hn1 9]$ more mdin_rst
energy minimization for trx starting structures

 &cntrl

imin=1, maxcyc=2000, ncyc=1200, nmropt=1, ntpr=20,
        ntb=0, igb=1,cut=16.0
 /
&wt type='REST', istep1=0,istep2=500,value1=0.1,
            value2=1.00,  /
 &wt type='REST', istep1=501,istep2=2000,value1=1.00,
            value2=1.00,  /

 &wt type='END'  /
LISTOUT=POUT
DISANG=RST
```

amber_min2.sub

```
Universe   = vanilla
Executable = /cluster/amber/amber11/exe/sander
Arguments  = -O -i mdin_rst -o model_min2.out -c model.ann.x -p prmtop -r model_min2.x
Log        = amber.log
Output     = amber.out
Error      = amber.error
```

This portion takes about 15-20 minutes to complete.

3) Once all of AMBER water refinement is completed, the statistics are written out using the script make_collect.inp (./make_collect.inp on command line). This script turns the output from the last minimization into a pdb file, statistics using the command for each model are written using sviol (sviol2) and are merged together and sorted by lowest AMBER energy.

make_collect.inp

```
#!/bin/sh

for f in model_0*
do
        cd $f
        ambpdb -bres -p prmtop  < model_min2.x > model.min2.pdb
#        ambpdb -bres -p prmtop  < model.x > model.min.pdb
        cd ..
done

for f in model_0*
do
cd $f
sviol model_min2.out > temp.out
grep 'Averages:                                                    -' temp.out > $f.average.out

cd ..
done
cat model_0*/*average.out > averages.txt
/home/MARQNET/olsona/scripts/joinpdb -o ensemble.pdb model_0*/model.min2.pdb

sed = averages.txt | sed 'N;s/\n/\t/' > test.txt
mv test.txt averages.txt
sort -n +2 -3 averages.txt > averages.out
sed = averages.out | sed 'N;s/\n/\t/' > test.txt
mv test.txt averages.out
rm -rf averages.txt
#/home/MARQNET/olsona/scripts/joinpdb -o ensemble.pdb model_0*/model.min.pdb
```

All 99 pdb files are written into one called ensemble.pdb.

The sorted liste (averages.txt);

1	19	Averages:	-3961.71	15.80	0.0000	0.00	0.00	9.947	5.851	0.000
2	69	Averages:	-3949.57	21.01	0.0000	0.00	0.00	8.488	12.522	0.000
3	73	Averages:	-3948.27	24.56	0.0000	0.00	0.00	10.871	13.692	0.000
4	52	Averages:	-3945.62	13.77	0.0000	0.00	0.00	7.845	5.922	0.000
5	42	Averages:	-3944.73	18.80	0.0000	0.00	0.00	11.407	7.396	0.000
6	53	Averages:	-3943.33	16.75	0.0000	0.00	0.00	8.560	8.193	0.000
7	95	Averages:	-3942.93	22.94	0.0000	0.00	0.00	9.806	13.138	0.000
8	51	Averages:	-3940.64	17.43	0.0000	0.00	0.00	7.671	9.756	0.000
9	91	Averages:	-3940.33	15.95	0.0000	0.00	0.00	8.253	7.700	0.000
10	88	Averages:	-3939.96	19.08	0.0000	0.00	0.00	8.819	10.262	0.000
11	74	Averages:	-3939.04	12.94	0.0000	0.00	0.00	7.428	5.512	0.000
12	3	Averages:	-3938.17	15.20	0.0000	0.00	0.00	9.055	6.141	0.000
13	63	Averages:	-3937.19	20.15	0.0000	0.00	0.00	8.233	11.916	0.000
14	68	Averages:	-3936.73	15.17	0.0000	0.00	0.00	9.506	5.665	0.000
15	35	Averages:	-3936.43	16.35	0.0000	0.00	0.00	8.718	7.628	0.000
16	66	Averages:	-3934.85	19.80	0.0000	0.00	0.00	12.282	7.518	0.000
17	13	Averages:	-3934.17	16.19	0.0000	0.00	0.00	8.944	7.250	0.000

Where the first line is the rank, the second line is the model number, the fourth line is the amber energy, the fifth line is total penalty energy, and ninth and tenth lines are distance and torsion penalty.

4) In the second line of averages.txt is the model number, which corresponds to the files in the directory. The top 20 AMBER energy structures should be copied to a new directory along with make_collect.inp.

make_collect.inp should be ran again to get the top 20 structures into one file called ensemble.pdb.

5) To get statistics on this new file, run the program called pdbstat (type pdbstat on command line), and follow the directions on start up.

- rea coo pdb ensemble.pdb
- all
- rmsd
- coo
- backbone (or heavy)
- range of protein residues (ex 7-110)
- average
- new file name
- write
- new file name.pdb
- coo
- all
- all
- pdb
- exit

6) This can write the coordinates so that they are on top of each other as when concatenating the proteins right after AMBER causes them to be of center.

7) pdbstat can also give other statistics about number NOEs and dihedral restraints used. (consult this website for more details http://www.nmr2.buffalo.edu/nesg.wiki/Main_Page)

Other program such as molmol can do similar features.

E) The new structures must be validated to make sure they are acceptable for pdb submission. Since we don't have procheck or whatcheck installed on any computers we can use an online resource called Protein Structure Validation Software suite (PSVS) from http://psvs-1_4-dev.nesg.org/, which also uses pdbstat to qualify the structure. Just submit the ensemble written from pdbstat and selecy dyana or cyana and any other features you want and hit submit. This takes about 5 minutes to get feedback. In general one wants the protein to have Z scores > -3 for both Procheck G and Molprobity Clashscore.

F) The website http://www.nmr2.buffalo.edu/nesg.wiki/Main_Page offers good details on how to submit info to PDB and BMRB.