

1-1-2007

**Nmr Structure and Relaxation Studies of Dhfr from Haloferax
Volcanii at High Salt**

Bulent Binbuga

NMR STRUCTURE AND RELAXATION STUDIES OF
DHFR FROM *HALOFERAX VOLCANII*
AT HIGH SALT

By

Bulent Binbuga

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Chemistry
in the Department of Chemistry

Mississippi State, Mississippi

August 2007

NMR STRUCTURE AND RELAXATION STUDIES OF
DHFR FROM *HALOFERAX VOLCANII*
AT HIGH SALT

By

Bulent Binbuga

Approved:

John K. Young
Assistant Professor of Chemistry
Director of Dissertation and
Major Professor

David O. Wipf
Professor of Chemistry

William P. Henry
Associate Professor of Chemistry

Svein Saebø
Professor of Chemistry

Kang Xia
Assistant Professor of Chemistry
Graduate Coordinator of the Department of Chemistry

Stephen C. Foster
Associate Professor of Chemistry

Gary L. Myers
Interim Dean of the College of Arts & Sciences

Name: Bulent Binbuga

Date of Degree: August 11, 2007

Institution: Mississippi State University

Major Field: Chemistry

Major Professor: Dr. John K. Young

Title of Study: NMR STRUCTURE AND RELAXATION STUDIES OF DHFR FROM
HALOFERAX VOLCANII AT HIGH SALT

Pages in Study: 182

Candidate for Degree of Doctor of Philosophy

The studies of enzymes from extreme sources have gained significance due to their increasing potential applications. The proteins from halophiles (salt loving) have adapted to challenging environmental conditions and require salt for their structure and function. How halophilic proteins adapt to a hypersaline environment is still an intriguing question. It is important to mimic the environmental conditions of the sample under investigation with experimental techniques.

In this study, structure and dynamic features of a halophilic enzyme have been investigated under high salt conditions. The acquisition of NMR data on high salt samples has always been problematic. We have devised a simple and elegant approach for obtaining NMR data for a protein in a high salt buffer that allows for virtually complete ^1H , ^{13}C , and ^{15}N assignments. These data were then used to calculate the NMR derived structure of *Haloferax volcanii* dihydrofolate reductase in 3.5 M NaCl.

Structure calculations showed that this protein folds in a similar manner as investigated in the crystal structures of *Haloferax volcanii* dihydrofolate reductase and *Escherichia coli* dihydrofolate reductase. To understand the effect of salt on flexibility as well as activity, NMR relaxation studies at 3.5 M and 1.0 M salt concentration were carried out. NMR dynamics of this enzyme revealed that the loss of activity as the salt concentration is lowered is due to loss in the inherent flexibility across the backbone, particularly in the catalytic loops.

DEDICATION

I would like to dedicate this research to the memory of my father, Kadim Binbuga, and my lovely mother Melek Binbuga.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to many people who have helped me with this research. I would first like to sincerely thank my major professor Dr. John K. Young for his friendship, support, advice and, encouragement during my graduate study. Deep appreciation is due for giving me the opportunity to be a graduate student in the Department of Chemistry to the Graduate Study Committee: Dr. Stephen C. Foster, Dr. William P. Henry, and Dr. David O. Wipf. I would like to thank Dr. David O. Wipf, Dr. William P. Henry, Dr. Svein Saebø, Dr. Kevin L. Armbrust, and Dr. Kang Xia for serving as my committee members. My deepest affection and gratitude are due my wife, Nursen Binbuga, for her love, and invaluable support during my study. Finally, special thanks to my parent and my friends.

TABLE OF CONTENTS

	Page
DEDICATION	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS.....	xiv
CHAPTER	
I. INTRODUCTION	1
Archaea Microorganisms, Halophilic Archaea	1
Halophilic Archaea	2
Protein Adaptation to High Salt Conditions	4
The Dihydrofolate Reductase, DHFR.....	8
The Enzyme Dihydrofolate Reductase (DHFR) from <i>HaloferaxVolcanii</i>	21
Objectives of This Research	26
REFERENCES	28
II. PROTEIN ISOLATION AND PURIFICATION.....	35
Introduction.....	35
Purification of <i>Hv</i> DHFR1 by Column Chromatography	36
Isotope Labeling.....	41
REFERENCES	46
III. CHEMICAL SHIFT ASSIGNMENTS OF THE ENZYME <i>Hv</i> DHFR1	48
Introduction.....	48
NMR of a High Salt Protein Sample.....	49
Three Dimensional NMR Experiments	51
Backbone Chemical Shift Assignments.....	57

	Side Chain Chemical Shift Assignments	72
	Conclusion	80
	REFERENCES	89
IV.	NMR STRUCTURE CALCULATION OF H _v DHFR1 FROM EXPERIMENTAL RESTRAINTS.....	93
	Introduction.....	93
	Molecular Modeling:	94
	Secondary Structure Prediction Using Chemical Shift Values.....	97
	Distance Restraints Through 3D NOESY Experiments	100
	Structure Calculation Using CNS	108
	Comparisons of NMR derived H _v DHFR1 Structure to Crystal H _v DHFR1 and E _c DHFR Structures:	117
	Conclusion	123
	REFERENCES	126
V.	INVESTIGATION OF ENZYME FLEXIBILITY THROUGH NMR RELAXATION MEASUREMENTS	131
	Introduction.....	131
	Protein Dynamics.....	133
	Spin Relaxation and Motion of Molecules in Liquid:	135
	Nuclear Overhauser Effect (NOE):	137
	Microsecond-Millisecond Time Scale Motion:	140
	Relaxation Theory:	142
	Modelfree Theory:	144
	Extracting the Relaxation Rate Constants with the Software Felix:	147
	Results.....	152
	Overall Conclusion	161
	REFERENCES	166
VI.	EXPERIMENTAL.....	172
	Protein Production and Purification.....	172
	NMR Data Collection	174
	Protocols for Structure Calculation by CNS.....	178
	Extraction of Relaxation Data with the Software Felix and Protocols for the Modelfree Formalism.....	179
	REFERENCES	181

LIST OF TABLES

TABLE	Page
3.1. The assigned backbone and side chain chemical shift values of <i>Hv</i> DHFR1.	82
4.1. The statistics of the experimental distance restraints obtained from the ¹⁵ N-NOESY-HSQC and ¹³ C-NOESY-HSQC used in structure calculation by CNS.	106
4.2. Statistics of the 20 lowest-energy structures calculated by CNS.	115

LIST OF FIGURES

FIGURE	Page
1.1. The catalytic reaction performed by the enzyme DHFR is shown. The substrate DHF is converted to THF by the enzyme using NADPH.	9
1.2. The metabolic reactions catalyzed by DHFR, SHT and TMS are illustrated. The cycle highlight that the product of a catalyst is an input for the other enzymes. ⁶⁷	12
1.3. Amino acid sequence alignment of DHFRs from various sources. The sequence alignment was done by CLUSTAL W (1.83) which is a multiple sequence alignment program. ¹⁰²⁻¹⁰⁴ The conserved residues are indicated by * for the variants, conservatively replaced residues are indicated by :, and the residues replaced drastically are indicated by . representation.	14
1.4. The structural illustration of substrate and cofactor binding of <i>Ec</i> DHFR. The molecule in green is the DHF and the one in the red is NADPH. The reaction takes place at a hydrophobic cleft of the enzyme. The structure is redrawn from the PDB submission of 1bdr. ¹⁰⁵	15
1.5. The figure represents the superimposition of three X-ray structures of <i>Ec</i> DHFR. The open (yellow), closed (blue) and occluded (green) conformations of the M20 loop are shown. The PDB codes are 1rx6 ¹⁰⁹ , 1rx1 ¹⁰⁹ , and 1ra9 ¹⁰⁹ respectively.	17
1.6. The crystal structure of <i>Hv</i> DHFR1 is redrawn from the pdb file 1vdr ⁸³ by using the software MOLMOL. ¹²⁴	25
2.1. The structure of cross-linked Sepharose. ¹⁶	38

2.2.	The UV-Vis measurement at 280 nm of each fraction from Sepharose CL-6B column is shown on the left (A). The SDS-PAGE results from high absorbance fractions are shown on the right (the fraction numbers are shown at the bottom of the gel)(B). The first line in the gel represents the molecular weight markers which are in the range from 97 kDa (top marker) and 14 kDa (bottom marker).	39
2.3.	The structure of DEAE-cellulose is shown. The cellulose is bonded to quaternary amino group by ether linkage. ¹⁹	41
2.4.	The UV-Vis measurement at 280nm for each collected fraction from DEAE-Sephacel column is shown on the left (A) and the SDS-PAGE results from high absorbance fractions are shown on the right (B). The fraction numbers are shown at the bottom of the gel.	41
2.5.	¹ H- ¹⁵ N-HSQC spectrum of 2.0 mM <i>Hv</i> DHFR1 at 3.5 M NaCl and 10 mM tris buffer at pH 7 and 25 °C.	45
3.1.	The NMR tube used for high salt protein sample is shown. The tube geometry afforded acquisition of interpretable NMR data. The fill factor generated by volume reduction was compensated by increasing the protein concentration.	50
3.2.	Schematic pulse diagram illustrations of HNCA experiment. The timing of the pulses starts from the left where the thickness of vertical bars (90° and 180°) indicate pulses width (duration) on the related nuclei frequency. The subscripts on the pulse are used to denote the applied pulse axes and the pulse delay are denoted as tau, τ . ²⁴	52
3.3.	Fourier transform on the second dimension results in a FID form sinusoidal function in the t1 dimension.	53
3.4.	The results of two dimensional Fourier transformations of the first two data blocks. The result is a series of two dimensional spectrum where the peak intensity for the same spot vary sinusoidally. The blue and red indicates positive and negative signals, respectively.	54
3.5.	A 3D NMR data, a cross section from a 3D NMR experiment, and a strip from the cross section are illustrated. The NMR data are analyzed by cross sectioning the cube to planes at one of the largest frequency axis (typically ¹⁵ N) and planes are further analyzed by strips at each recorded ¹ H ^N chemical shift values.	55

3.6.	The one-bond (1J) and two-bond (2J) scalar coupling constants across the peptide backbone for a ^{15}N and ^{13}C labeled protein is shown. Correlations among homonuclear and heteronuclear spins are established through these couplings in triple resonance NMR experiments. ¹⁰	55
3.7.	The list of three dimensional triple resonance experiments for backbone chemical shift assignments is shown.	58
3.8.	Strips [f1(^1H)-f3(^{13}C)] were taken from ^{15}N plane (127.60 ppm) of HNCA (left) and HN(CO)CA (right) experiments. In the strip from HNCA, the more intense peak was assigned as residue's own $^{13}\text{C}^\alpha$ while the less intense one was assigned as $^{13}\text{C}^\alpha_{i-1}$ chemical shift values. The same $^1\text{H}/^{15}\text{N}$ chemical shift pair in HN(CO)CA experiment correlates to only $^{13}\text{C}^\alpha_{i-1}$ (right strip) chemical shift value.	60
3.9.	The figure illustrates the methodology followed to establish the $^{13}\text{C}^\alpha$ chemical shift correlations among the residues via the HNCA/HN(CO)CA experiments.	61
3.10.	Strips (f1-f3) from $^1\text{H}/^{15}\text{N}$ chemical shifts of the residues T145-F150 of 3D HNCA experiment is shown. The strips were obtained from different ^{15}N planes of HNCA experiments. The solid blue lines connect the intrer-residual $^{13}\text{C}^\alpha$ of residue i with the corresponding $^{13}\text{C}^\alpha$ of residue $i-1$	63
3.11.	Two strips (f1,f3) from the HNCACB and CBCA(CO)HN experiments are shown. The residue correlates both its own $\text{C}^\alpha/\text{C}^\beta$ chemical shifts at 53.78/42.49 ppm (intense positive and negative peaks) and previous residue's $\text{C}^\alpha/\text{C}^\beta$ at 54.95/19.67 ppm at the HNCACB. The residue only correlates to $\text{C}^\alpha/\text{C}^\beta$ of previous residues in the CBCA(CO)NH. The correlations of nuclei between the two experiments are shown by dashed lines.	65
3.12.	Strips (f1-f3) from the 3D HNCACB experiments are shown. The strips illustrate the C^α and C^β sequential correlations between L3 to A8. The correlations between C^β are shown by red line connectivity where a less intense negative peak becomes a strong negative peak in the previous strip.	67
3.13.	The labeled $^1\text{H}-^{15}\text{N}$ HSQC spectrum of <i>Hv</i> DHFR1 is shown. The $^1\text{H}/^{15}\text{N}$ chemical shifts were assigned to individual residue after completing the backbone chemical shift assignments.	68

3.14.	Two strips (f1-f3) from HNCO and HN(CA)CO experiments are illustrated. The chemical shift correlation between the two experiments is shown. The $^1\text{H}/^{15}\text{N}$ chemical shifts of Q153 correlate to $^{13}\text{C}'$ of L152 in HNCO. Both inter- and intraresidue $^{13}\text{C}'$ correlations were obtained in HN(CA)CO experiment.	71
3.15.	The list of 3D NMR experiments used for side chain proton and carbon chemical shift assignments.	73
3.16.	A strip (f1-f3) from (H)C(CO)NH-TOCSY demonstrates that $^1\text{H}^{\text{N}}/^{15}\text{N}$ chemical shift pair of the I100 correlating to the side chain carbon chemical shifts of V99 at 21.37, 19.31 and 32.39 ppm (identified by arrows).	74
3.17.	A strip (f1-f3) from H(CCO)NH-TOCSY experiment. The $^1\text{H}^{\text{N}}/^{15}\text{N}$ chemical shift pair of I100 correlates to three side chain proton chemical shifts of previous residue, V99, at 0.43 ppm, 0.87 ppm and 2.18 ppm. The H^{γ} chemical shifts were tentatively assigned to methyl groups. The more accurate $^1\text{H}^{\text{aliphatic}}-^{13}\text{C}^{\text{aliphatic}}$ correlations were accomplished via the 3D HCCH-TOCSY experiment.	76
3.18.	Strips [f1(^1H)-f3(^1H)] from HCCH-TOCSY experiment are shown. The strips were taken from $^{13}\text{C}^{\beta}$ and $^{13}\text{C}^{\gamma}$ s planes of V99. The diagonal peaks (indicated by arrows) represent the chemical shift value of the proton directly attached to carbon nucleus. The cross peaks correspond to proton chemical shifts of the neighboring carbon nuclei on the same residue.	79
3.19.	Strips [f1(^1H)-f3(^1H)] from the HCCH-TOCSY spectrum illustrating the assignment strategy used for R46. The strips from $^{13}\text{C}^{\alpha}$, $^{13}\text{C}^{\beta}$, $^{13}\text{C}^{\gamma}$, $^{13}\text{C}^{\delta}$ planes point up that there is no complete magnetization transfer within the spin system. By piecing the data from each plane, the chemical shift assignments were able to be accomplished.	80
4.1.	Deviation of H^{α} , C^{β} , C^{α} , and C' values from the random coil values are shown. The arrows indicate the presence of β -strands and tubes indicates helices.	98
4.2.	The screen of TALOS program illustrating predicted both the phi and psi angles for the residue L118 and secondary structure region. The X-ray structure of <i>Hv</i> DHFR1 was used as reference shown in blue dot, and the predicted region from the chemical shift values matching the database values are shown in green dot.	100

4.3.	Three strips from different ^{15}N planes of the ^{15}N -NOESY-HSQC experiment belonging to residues L115 (from a β -strand region), A85 (from a α -helical region), and L9 (from a random coil region) are shown. The diagonal peaks in each strip are indicated by arrows which correspond to $^1\text{H}^{\text{N}}$ of the residue. The peaks that are not in the plane or are not assigned are highlighted in green. The peaks that correlate different protons chemical shifts are labeled in red.	102
4.4.	A strip from the ^{13}C -NOESY-HSQC experiment belonging to $^1\text{H}^{\beta}$ of A89 at 1.40 ppm is shown. The diagonal peak belonging to methyl group protons (highlighted in blue) correlates to other side chain protons of different residues. The assignments in red indicate the peak overlaps. The ^{13}C -NOESY-HSQC experiment provides long-range proton restraints.	105
4.5.	The figure illustrates the identification of the secondary structure elements from the analysis of short range NOE data in T75-Y125 region. ^{53,75,76} The bar height specify the intensity of the observed NOE peak and length specify the range of correlations. Similarly, the strength of long range interactions are showed by colored lines where red indicates strongest and yellow indicates the weakest interactions.	107
4.6.	The β -sheet is characterized by long range $\text{H}^{\text{N}}\text{-H}^{\text{N}}$ (dashed lines) and $\text{H}^{\text{N}}\text{-H}^{\alpha}$ (solid lines) correlations. The $d_{\alpha\text{N}}(i, i+1)$ correlations along the β -strands are shown by the dashed red lines.....	109
4.7.	The structure calculation steps and TAD phase used in the CNS program are shown.	110
4.8.	The backbone superimposition of the 20 lowest-energy structures (left) and 180° rotated form (right) of <i>Hv</i> DHFR1 generated using the CNS is shown. The helical regions are highlighted in red and β -strands are in yellows.	114
4.9.	A representative NMR derived structure of <i>Hv</i> DHFR1 is shown in ribbon diagram. The helices and β -strands are numbered in the order of sequence. The structure is generated by the program MOLMOL. ⁸²	116
4.10A.	The results of PROCHECK is shown for one of the submitted structures.	118
4.10B.	The results of PROCHECK is shown for the 20 submitted structures.	119

4.11.	Superimpose of the X-ray crystallography structure (1vdr) of <i>Hv</i> DHFR1 (red ribbon) with the liquid state NMR structure (gray ribbon) resulting in a 1.35 Å rms deviation.	120
4.12.	One variation between the NMR (yellow) and X-ray crystallography structures (maroon) is in the β 3 region. Key NOEs were obtained for S59 and A60 that are not supported by the crystal structure.	121
4.13.	It is shown that H ^N of A28 is at a close enough proximity to carboxylic oxygen of E148.	122
4.14.	The superimposition of NMR derived structure of <i>Hv</i> DHFR1 (yellow) versus <i>Ec</i> DHFR (grey) (pdb code: 1RA9 ⁸⁶) is show. The loop GH of <i>Ec</i> DHFR is truncated by deletion of six residues in <i>Hv</i> DHFR1.	123
4.15.	The L21 and adenosine binding loop regions of <i>Hv</i> DHFR1 from NMR derived study are shown. The loop conformation is a time average conformation of the <i>open</i> , <i>occluded</i> , and <i>closed</i> conformation of <i>Ec</i> DHFR. ⁸⁶ The adenosine binding loop displayed no specific conformation.	124
5.1.	Illustration of ¹⁵ N- ¹ H bond vector motion relative to a reference frame. The protein molecule is illustrated by gray region. The amplitude of motion is characterized by S ² where S ² =1 indicates rigid, S ² =0 indicates isotropic motion. ⁴⁹	134
5.2.	The energy level diagram for two spin system is shown. The probabilities of transition across the spin states are shown. W _S and W _I are single quantum transition rates; W ₀ , and W ₂ are zero and double quantum transitions rates. ⁶⁴	138
5.3.	The transition probabilities after spin saturation of I is illustrated. The cross relaxation rate for zero quantum transition is proportional to W _I -W _S and for double quantum transition transition is proportional to W _I +W _S . ⁶⁴	139
5.4.	The figure illustrates the N-H bond vector experiencing two sites with an exchange rate of k _{ex}	141
5.5.	The ¹ H- ¹⁵ N HSQC spectrum of <i>Hv</i> DHFR1 at 3.5 M salt concentration recorded for R ₁ experiment at the relaxation delay times: 50 ms, 700 ms and 1500 ms.	149

5.6.	A spectrum region from the ^1H - ^{15}N HSQC of <i>Hv</i> DHFR1 at 3.5 M NaCl is shown. The spectra with proton saturation (noe_on) and without saturation (noe_off) are shown.	150
5.7.	The R_1 time course measurements for W134 of <i>Hv</i> DHFR1 at 3.5 M NaCl are shown.	151
5.8.	The R_2 time course measurements for L21 of <i>Hv</i> DHFR1 at 3.5 M NaCl are shown.	151
5.9.	The calculated T_1 , T_2 and heteronuclear ^1H - ^{15}N NOE values for the enzyme <i>Hv</i> DHFR1 at 3.5 M NaCl is shown.	153
5.10.	The calculated T_1 , T_2 and heteronuclear ^1H - ^{15}N NOE values for the enzyme <i>Hv</i> DHFR1 at 1.0 M NaCl is shown.	154
5.11.	The superimposed of order parameters of backbone ^{15}N resulting from 3.5 M (red) and 1.0 M NaCl salt (blue) concentrations. The order parameter varies between 1 and 0 where $S^2=1$ indicates completely restricted motion and the $S^2=0$ indicates large amplitude of motion. The flexible loop regions at 3.5 M NaCl are defined from their low S^2 values: L21 loop, $\alpha 2$ - $\beta 3$, adenosine binding loop (ABL), and $\beta 6$ - $\beta 7$ loop.	155
5.12.	The four flexible loop regions L21, adenosine binding, $\alpha 2$ - $\beta 2$, and $\beta 6$ - $\beta 7$ are shown on an NMR derived structure.	156
5.13.	The electrostatic surface of NMR derived structure is shown. The negatively charged regions are illustrated by red and positively charged regions are illustrated by blue.	160

LIST OF ABBREVIATIONS

<i>Hv</i> DHFR	<i>Haloferax volcanii</i> dihydrofolate reductase
<i>Ec</i> DHFR	<i>Escherichia coli</i> dihydrofolate reductase
DHFR	Dihydrofolate reductase
THF	Tetrahydrofolate
NADPH	Nicotinamide Adenine Dinucleotide Posphate
DNA	Deoxyribonucleic Acid
TMS	Thymidylate Syntase
dTMP	Deoxythymidilymono-phosphate
dUMP	Deoxyuridyl-monophosphayte
MTX	Methotrexaten
TMP	Trimethoprim
SHT	Serine Hyroxymethyl Transferease
Asp	Aspartic Acid
Glu	Glutamic Acid
Trp	Tryptophan
Lys	Lysine
Ala	Alanine
ddTHF	5-deazafolate, 5,10-dideazatetrahydrofolate

CD	Circular Dichroism
FL	Fluorescence Spectroscopy
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NOE	Nuclear Overhauser Effect
TOCSY	Total Correlation Spectroscopy
HSQC	Heteronuclear Single Quantum Coherence
FID	Free Inductive Decay
1D	One dimensional
2D	Two dimensional
3D	Three dimensional
CNS	Crystallography and NMR System
TAD	Torsion Angle Dynamics
MD	Molecular Dynamic
SA	Simulated Annealing
DG	Distance Geometry
TALOS	Torsion Angle Likelihood Obtained from Shift and sequence similarity
rmsd	root-mean-square deviation.
rf	radio frequency
ppm	part per million
Å	Angstrom
CH ₅ N ₂ HCl	Guandine

$(\text{NH}_4)_2\text{SO}_4$	Ammonium Sulphate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
DEAE	Diethylaminoethyl
η	Solvent Viscosity (cP)
τ_m	Correlation Time
T_1	<i>Spin-Lattice</i> Relaxation Time
T_2	<i>Spin-Spin</i> Relaxation Time
CSA	Chemical Shift Anisotropy
W_0	Zero Quantum Transition Rate
W_1	Single Quantum Transition Rate
W_2	Double Quantum Transition Rate
R_{ex}	Chemical Exchange Term
$J(\omega)$	Spectral Density Function
S^2	Squared Order Parameter

CHAPTER I
INTRODUCTION

Archaea Microorganisms, Halophilic Archaea

Living organisms are classified into three kingdoms: Eucarya, Eubacteria and Archaea.¹ Many microorganisms from Archaea Kingdom can survive in unusual environmental conditions and are called extremophiles.² Archeae refer to ancient microorganisms since this group is believed to have been one of the earliest forms of life on this planet.³ Extermophilic microorganisms are classified according to the physicochemical conditions at which they survive.⁴ Among them, piezophiles grow and function best under high pressure (deep sea), thermophiles at high temperature (hot springs and volcanic areas),⁵ psychrophiles at low temperature (Antarctica),⁶ acidophiles at low pH (volcanic vents),¹ alkalophiles at high pH (soda lakes),⁷ and halophiles (hals:salt, philes: loving) in hypersaline environments.⁸ This work centers on the explanation of halophilic adaptation through studying an enzyme from a halophilic organism.

Several enzymes from extremophilic organisms have industrial applications due to their ability to withstand and maintain function under harsh processing conditions. The extreme environments are considered a source of novel enzymes.^{9,10} Enzymes from extremophiles have been used in the pulp, detergent, textile, starch, cellulose, and

biocatalyst industries.¹¹⁻¹⁴ So far, halophilic enzymes and bacteria have found limited biotechnological applications but promise an increasing trend.¹⁵ Their ability to function under minimum low water content and their resistance to proteases makes the halophilic enzymes potential applicants in relevant industrial usage.¹⁶ Halophilic organisms have been used in waste water treatment in textile industry and degradation of organic pollutants.¹⁵ It has been shown that haloarcheas can degrade alkanes by utilizing sulfur anaerobically.¹⁵ Enzymes from haloacidophiles have been used in the starch industry to convert starch to more valuable products such as dextrin, glucose, and fructose.¹⁵ Due to outstanding stability and resistance to esterases, novel lipids produced by the halophilic organisms have been used in medicine, and cosmetic industry in the area of liposome application.¹⁷ The biosurfactant peptides produced from halophilic organisms have been used for oil recovery in petroleum reservoirs for environmental protection.^{18,19} The halophilic exopolysaccharides produced by halophilic microorganisms such as *Halofrex volcaii*, *Halobacterium salinarium* have been used as emulsifier in oil recovery.¹⁵ Carotenes obtained from halophilic bacteria are used as food coloring agents and as compatible solutes in pharmaceutical industry.²⁰

Halophilic Archaea

The total inorganic salt content of typical seawater is about 3.5-3.7% (w/v). A salt concentration exceeding this level is defined as hypersaline; saturation is reached at 35-37% (w/v) salt level.²¹ The salt composition and concentration varies from one location to another and can show variations within a location during a period of time; a gradient can also be present within a naturally occurring high salt source.²¹ Beside Na⁺

which being the major cation in most of the saline water bodies (except in the Dead Sea where divalent Mg^{+2} is the major cation), K^+ , Ca^{2+} , and Mg^{2+} can be found at different levels. Similarly Cl^- is the major counter anion in most of the high salt water bodies; Br^- , CO_3^{2-} , and SO_4^{2-} ions are also present at different compositions.²²

Halophiles can thrive at salt concentrations of near saturation. The Great Salt Lake^{23,24} (33% w/v, Utah, USA), Dead Sea²⁵ (34% w/v, Israel), Tuzgolu²⁶ (33% w/v, Turkey), and other similar salty lakes are examples of naturally developed hypersaline environments. Halophiles are subcategorized based on salt requirement for their growth. Extreme halophiles require 20-30% (corresponding 3.4-5.0 M) salt, moderate halophiles 5-20% (0.8-3.4 M), slight halophiles 2-5% (0.2-0.8 M) and halotolerant <2% salt for optimal growth condition.^{21,27}

Taxonomic classifications of halophilic organism show that halophiles exist in the three kingdoms.^{17,28,29} In the kingdom of Archaea they are called haloarchaea, in Eubacteria, they are called halobacteria, and in Eukaryat they are called haloeukaryotes.^{29,30} Most of the enzymes studied until now are mainly from halophilic Archaea and particularly from the *Halobacteriaceae* family.³¹ Some common genera and species (shown in parenthesis) from *Halobacteriaceae* group are: *Haloferax* (*Haloferax volcanii*, *Haloferax mediterranei*), *Halobacterium* (*Halobacterium salinarum*, *Halobacterium noricense*), and *Haloarcula* (*Haloarcula marismortui*, *Halobacterium marismortui*).^{30-32,33}

Halophilic microorganisms have developed unique strategies to adapt to hypersaline environments. Halobacteria and haloeukaryotes survive in hypersaline conditions by means of a 'compatible solute' strategy.³⁴ These microorganisms

accumulate only neutral and polar organic solutes and exclude inorganic salts in their cytoplasm.³⁴ The solutes are either synthesized in the cell or uptaken from the medium.²¹ Most common compatible osmolytes are polyols and their derivatives (sucrose, glucosylcerol, glycerol etc), glycine, betaine and some amino acids.³⁵ On the other hand, halophilic archaea utilize a ‘salt-in’ strategy to adapt to hypersaline environments.³⁵ This strategy includes the accumulation of inorganic salts in the cytoplasm at a level of the surrounding medium owing to a chloride ion gradient across the cell. Ion transports across the cell membrane are facilitated by specific ion channels.³⁶

Halobacteriaceae family microorganisms are the most abundant inhabitants in high salinity environments.^{33, 37-39} They can be identified by their typically red or orange colors, which originate from C₅₀ carotenoids present in the cytoplasm. The high osmotic stress is balanced by selectively accumulating K⁺ ions at a near saturation or exceeded level in the cytoplasm.³⁸ The cell mechanism favors K⁺ ions because they hydrate less water and because Na⁺ ions are eliminated by a process of pumping owing to the Cl⁻ gradient. This process protects the cells from dehydration.⁴⁰

Protein Adaptation to High Salt Conditions

Proteins attain functional three dimensional structures through weak inter- or intra-molecular interactions such as electrostatic, hydrophobic and hydrogen bonding. The changes in physical environment may alter these interactions through the weakening or strengthening of these interactions. The accumulated high salt in the cytoplasm makes the cell components and cell mechanism adapt to these challenging conditions.⁴¹ The proteins from halophiles require salt for their structural integrity and functions.⁴¹ How

proteins adapted to hypersaline environments is still a subject of debate. For a protein to be active, it should be hydrated in the medium. The limited availability of water molecules in hypersaline medium leads these groups of enzymes to develop adapted strategies.⁴¹

Salts have an interesting impact on protein stability and solubility.⁴²⁻⁴⁴ The effect of different salt ions on protein stability has been investigated.⁴⁵ The *Hofmeister series* ranks the effect of some anions and cations in the stabilization of proteins by ‘salting-in’ and ‘salting-out’ mechanisms.⁴⁵ The effect of anion on the protein ternary structure is in the following order: CO_3^{2-} (most stabilizing), SO_4^{2-} , $\text{S}_2\text{O}_3^{2-}$, H_2PO_4^- , F^- , Cl^- , Br^- , NO_3^- , I^- , ClO_4^- , SCN^- (most destabilizing).⁴⁶ The anions between ClO_3^{2-} and Cl^- are termed as kosmotropes while the anions from Cl^- to SCN^- are termed as chaotropes.⁴⁷ In this series, kosmotropes precipitate the macromolecule out of the solution by disrupting the structure; conversely, chaotropes destabilize the structure without precipitating.⁴⁷ These ions manipulate proteins solubility and stability either by disrupting the hydration shell or by interacting directly with proteins. The precipitation of protein is known as the “salting-out” effect.⁴³ In contrast to non-halophilic proteins, halophilic proteins are soluble in kosmotropes ions of the *Hofmeister series* owing to some unique structural features.⁴⁶ Generally halophilic proteins are sensitive toward the salt type as well the salt concentration.^{46,48,49}

Some common features of halophilic proteins have been identified. Primary amino acid composition analysis shows that halophilic proteins bear excess negatively charged amino acids (aspartic acid and glutamic acid) over positively charged ones.^{50, 51} According to the widely-accepted halophilic stabilization model, acidic residues at the

surface of halophilic proteins compete for water with the counter ions and coordinate water molecules that afford solubility and activity in hypersaline environments.⁵² It was shown that acidic residues can coordinate more salt and water molecules than the other amino acids.⁵³ It was proposed that this both eliminates unfavorable electrostatic repulsions between similar charged residues and provides a means for proteins to be solvated at high salt.⁵³ Studies have shown that lowering salt concentration has a drastic effect on halophilic protein stability through denaturation. According to the ‘competition model’ lowering salt concentration may reduce the shielding effect between similar charged residues and as a result increase unfavorable interactions.⁵³

Hydrophobic interactions also play a significant role in the stabilization of proteins at high salt concentration.⁵⁴ Halophilic and non-halophilic proteins bear different contents of hydrophobic residues.⁴¹ In general, the sequence analysis reveals that halophilic proteins contain fewer large hydrophobic residues such as tryptophan, isoleucine, and leucine but are rich in small size hydrophobic residues such as alanine and valine when compared to their non-halophilic counterparts.⁵⁵ High salt concentration can enhance hydrophobic interactions within a protein through minimizing the availability of water molecules that cause enhanced interactions among aliphatic side chains of a protein. The stabilizing effect of these hydrophobic interactions at high salt concentration was proposed to be disrupted through the increasing availability of water molecules around hydrophobic groups as the salt concentration is lowered.⁵²

In earlier studies, bio-physical techniques such as light scattering,⁵⁶ ultracentrifugation, neutron and small angle X-ray scattering⁵⁷⁻⁵⁹ were heavily employed to explain halophilic protein structural features and adaptation to high salt concentration.

The results of these studies show that unusually large amounts of salt and water interact with these proteins. Typically, non-halophilic proteins bind ~ 0.2 g H₂O /g proteins, while halophilic proteins have three to four times more water.⁶⁰ The role of acidic residues on a protein surface is believed to be the key issue in stabilization.⁵⁴ Typically, while non-acidic residues bind $\sim 2-4$ water molecules, acidic residues bind ~ 7 water molecules.⁶¹ The halophilic adaptation remained unclear until three dimensional halophilic protein structures became available. Three dimensional structures did reveal some interesting features of halophilic proteins and also lead to new questions since structural data did not completely support previous theories.⁶²⁻⁶⁵

The first halophilic protein structure, Malatedehydrogenase(hMDH) (PDB code: 1hlp) from *Haloarcula marismortui*, was resolved in 1995 by X-ray crystallography at 3.2 Å resolution.⁶² This protein is the most studied protein in the investigation of halophilic adaptation by different research groups using different techniques. The comparative structural analysis of hMDH and its homologous non-halophilic protein (dogfish malatedehydrogenase, dfDHM) displayed some interesting structural differences. The surface of hMDH was uniformly covered with negatively charged amino acids and more salt-bridges were involved across the surface, which was not found in the homologous non-halophilic dfDHM. In the latter structural studies at a higher resolution, 2.6 Å and 2.9 Å, the structure of hDHM and its mutant (PDB codes: 2hlp, 1d3a) were resolved for a further detail analysis.⁶³ Beside previous findings, an additionally large amounts of water molecules, and relatively fewer bound salt ions were identified on the surface of the protein. Ferredoxin (2Fe-2S) was the second halophilic protein which was resolved at a resolution of 1.9 Å from the same organism (PDB codes: p00216,

p00217).⁶⁴ Acidic residues are found over the surface of the protein. Three water layers were detected to be coordinated through hydrogen bonding at the surface of the protein. However, only a few cations bound to the protein were detected. The third structure investigated from haloarchaea was *Haloferax volcanii* dihydrofolate reductase (*HvDHFR*) in 1997 at a resolution of 2.6 Å.⁶⁵ The detailed structural discussion on this protein will be given in the next sections.

Today, around 10-15 protein structures from halophilic organism have been deposited in the Protein Data Bank. With this limited number of structural data, the explanation for their adaptation to hypersaline environment is not complete. By studying the structure of a halophilic protein in liquid state at the physiological condition, it is aimed to make a contribution to understand the relation between structural features of an halophilic protein and the the role of salt in protein stability. The concept of halophilic adaptation has been also investigated through studies of the enzyme DHFR.

The Dihydrofolate Reductase, DHFR

The dihydrofolate reductase (DHFR) enzyme, a crucial enzyme in the biochemical processes of folic acid, found in all living organisms. It catalyzes the reduction of 7,8-dihydrofolate (DHF) (**1**) to 5,6,7,8-tetrahydrofolate (THF) (**2**) by employing the cofactor, nicotinamide adenine dinucleotide phosphate (NADPH) (**3**), Figure 1.1.⁶⁶ To a lesser extent it also catalyses the reduction of folic acid (**4**) to dihydrofolate (**1**). The bacterial source of DHFRs only catalyses the reduction of dihydrofolate to tetrahydrofolate.⁶⁷ The reaction take places at the catalytic site of the enzyme where both substrate and cofactor are aligned to a favorable position by the

enzyme at which the conversion of DHF to THF takes place, Figure 1.1. One hydride transfer takes place from the C4 of NADPH to the C6 of DHF (2). The source of the second proton is unclear. It has been proposed that the second proton transfer to N5 of DHF takes place from ionizable side chain of an acidic residue at the active site of the protein, which is aspartic acid in bacterial DHFRs and glutamic acid in mammalian DHFRs. Alternatively, water molecules from bulk solvent is proposed to be the source.^{68, 69} In turn, NADP⁺(5) and THF are the products.⁷⁰ The rates of catalytic processes are different for different organisms's DHFRs and are pH dependent reactions. The product THF is a key precursor for many biosynthetic processes and is the functional form of folic acid used in metabolism. The pool of THF is utilized in the synthesis of purines, thymidine and some amino acids. The nucleotide thymine is one of the building blocks of DNA. This role has made DHFR an anti-cancer and anti-bacterial target. THF is used in the synthesis of one methyl group substitute, N⁵,N¹⁰-Methylene-THF.⁶⁶ In turns this intermediate is employed by Thymidylate Synthase (TMS) as a cofactor in the synthesis of deoxythymidilymono-phosphate (dTMP) from deoxyuridyl-monophosphate (dUMP) by one carbon atom transfer chemistry.⁷¹ Successively, dTMP is converted to

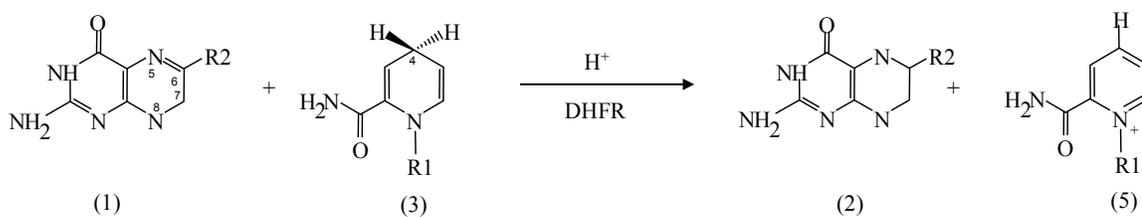
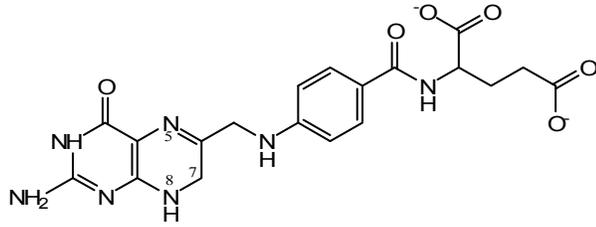
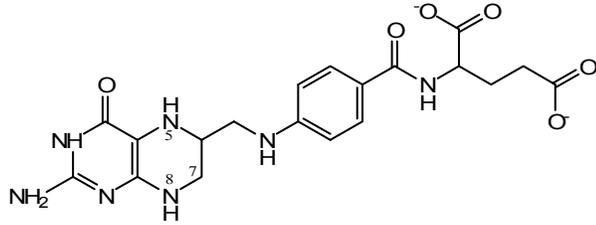


Figure 1.1. The catalytic reaction performed by the enzyme DHFR is shown. The substrate DHF is converted to THF by the enzyme using NADPH.



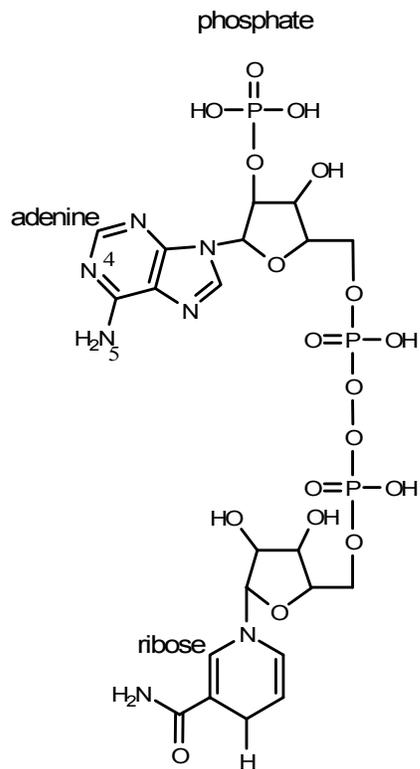
1

DHF (7,8-dihydrofolate)



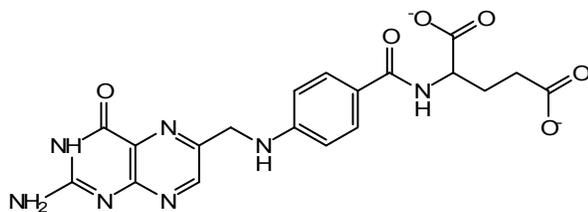
2

THF (5,6,7,8-tetrahydrofolate)



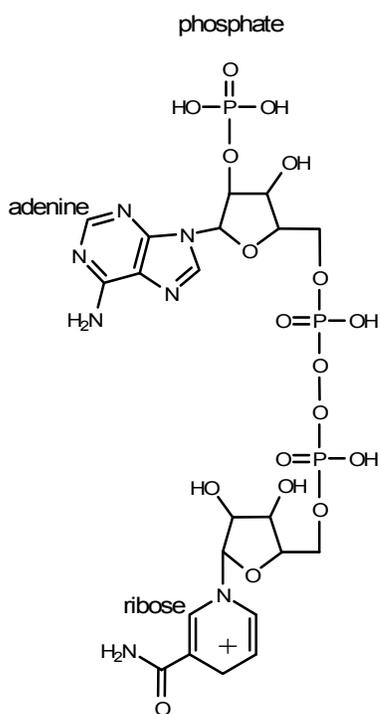
3

NADPH (nicotinamide adenine dinucleotide phosphate)



4

Folate



5

NADP⁺

deoxythymidine triphosphate (dTTP) to be utilized in the synthesis of DNA.⁷² The metabolic cycle is summarized in Figure 1.2. In these cascaded biosynthetic cycles, maintaining the pool of THF is essential for rapidly growing cells.⁷² Cancer cells grow rapidly, which requires a supply pool of THF. Inhibiting the enzyme DHFR from

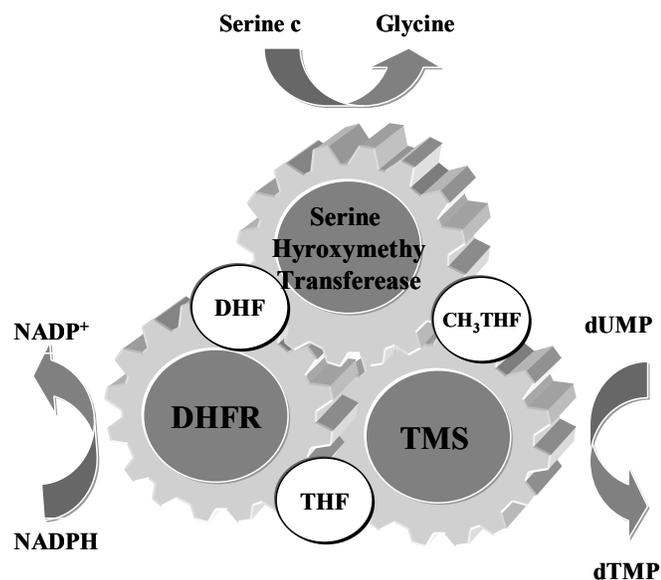
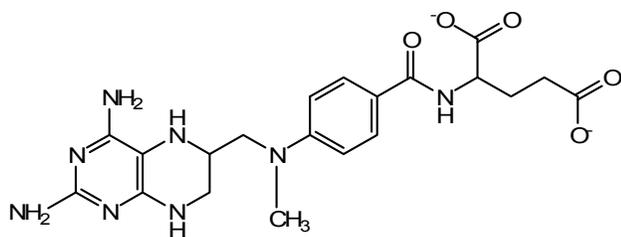
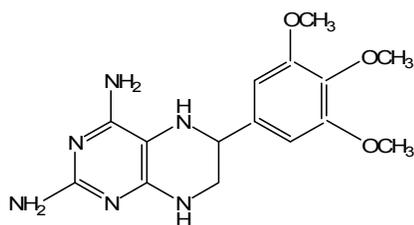


Figure 1.2. The metabolic reactions catalyzed by DHFR, SHT and TMS are illustrated. The cycle highlight that the product of a catalyst is an input for the other enzymes.⁶⁷



6

Methotrexate



7

Trimethoprim

functioning via inhibitors would result in retarding the DNA production for rapidly growing cells.⁷³ Several antifolate agents have been used to inhibit DHFR activity.⁷² The inhibitor molecules are typically similar to the substrate DHF.⁷² methotrexaten, (MTX) (6) is an anticancer drug which acts as an inhibitor for several DHFRs from different organisms, widely used in humans for the treatment of leukemias,^{74,75} lymphomas,^{76,77} and rheumatoid arthritis.^{78,79} The antibiotic trimethoprim (TMP) (7) is another inhibiting agent used for bacterial DHFRs to inhibit the enzymatic activity.^{80,81} These inhibitors bind to the active site of the enzyme at a higher affinity and halt the binding/releasing events. Under these circumstances the enzyme can not accomplish its functions at a regular rate.

DHFRs from various organisms share some similarities and dissimilarities in their biochemical and biophysical characteristics. For example, while some amino acids are conserved in all DHFRs, some are replaced conservatively or randomly. Figure 1.3 shows the amino acid sequence alignment of some selected DHFRs from various organisms. Even, in some cases sharing less than 30% sequence similarity, all known DHFRs structures exhibited similar folding which is referred to as Rossmann folding where four helices are packed across a central β -sheet.⁸²⁻⁸⁹ In addition, the fundamental function of all DHFRs are the same. The structure and activity aspects of DHFRs from several organisms have been extensively studied.^{68,70,84,90-98} Two consecutive glycine residue form a *cis* peptide bond geometry, which is also a unique feature of all known DHFRs.⁹⁹ The acidic residues Asp or Glu are catalytically important and are conserved in all DHFRs at the active loop site. In addition, catalytic mechanisms of DHFRs were determined to be similar.^{97, 98, 100} Based on *Ec*DHFR structural studies, residues Ala7,

```

DHFR_H.volcaii1 -----MELVSVAAALAE*NRVIGRDGELPWPSTLPADKKQY*RSRIAD----- 39
DHFR_E.coli -----MISLTAALAVDRVIGMENAMPWN-LPADLAWFKRNTLD----- 37 29%
DHFR_H.volcaii2 MSGEELSGSETEGDAAVRFVLVA*AVADNRVIGRDGTMFWH-LPEDLKQFKRTT*G----- 54 37%
DHFR_chicken -----VRSLSN*IVAVCONMGIGKDG*NLFWPPLRNE*YKYFORMTSTSHVEG 45 25%
DHFR_H.sapiens -----MFLLLNCTVAVSQNMGIGKNGDL*RPPLRNEFRYFORMTTTSVVEG 46 20%
L.casei -----TAF*LWAQDRDGLIGKDG*HLEWH-LPDDLHYFRAQ*TVG----- 36 20%
          : * : ** :. :* : : :.

DHFR_H.volcaii1 --DPVVLGR*TFESMRD----DLPGSAQIVMSRSERSFSVDTAHRAASVEEAVDIAASLD 93
DHFR_E.coli --KPVIMGRHTWESI*GR----PLPGRKNII*LSS-QPGTD-DRVTWVKS*VDEAIAACGDVP 89
DHFR_H.volcaii2 --HPVVMGRKTYESI*ARQLGGPLPERHSVLLTTRDL*DLP-EGAEAVESVESAVAAEDAA 111
DHFR_chicken KQNAVIMGKKTWFSI*PEK-NRPLKDRINIVLSRELKEAPKGAHYLSKSLDDALALLDSPE 104
DHFR_H.sapiens KQNLVIMGRKTWFSI*PEK-NRPLKDRINLVL*SRELKEPPQGAHFLARSLDDALKLTERPE 105
L.casei --KIMVVGRRTYE*SFPKR----PLPER*TNVVLT*H-QEDYQAQGA*VVVDVA*AVFAYAKQHP 90
          . :. :* : * : * : * : :. :.

DHFR_H.volcaii1 AET----AYVIGGAAIYALFQPHLD--RMVLSRVPGEGDTYPPEWDAAEWELDAETD- 146
DHFR_E.coli EIM-----VIGGGRVYEQFLPKAQ--KLYLTHIDAEVEGDT*HFPDYEPDDWESVFEF- 140
DHFR_H.volcaii2 DEMGVETVYVGGATVYEQFLGAA--GLVLT*ELDAAHEGDT*RFPEWDRGK*VVETDRDD- 168
DHFR_chicken LKSKVDMVIVGGTAVYKAAMEKPINHRLFVTRILHEFESDTFFPEIDYKDFKLLTEYPG 164
DHFR_H.sapiens LANKVDMIWIVGGSSVYKEAMNHLGHLKLFVTRIMQDFESDTFFSEIDLEKYKLLPEYPG 165
L.casei DQE----LVIAGGAQIF*TAFKDDVD--TLLVTRLAGSFE*GDTKMIPLNWDDE*TKVSSRT- 143
          : ** :. : : :. :. :. :.

DHFR_H.volcaii1 --HEGFTLQEWVRSASSR----- 162
DHFR_E.coli --HDADAENSHSYCFEILERR----- 159
DHFR_H.volcaii2 --RDGFSFVTYERKQPAADRGAESDE 194
DHFR_chicken VPADIQEEDGIQYKFEVYQKSVLAQ--- 189
DHFR_H.sapiens VLSDVQEGKH*IKYKFEVCEKDD----- 187
L.casei ---VEDTNPAL*THTYEVWQKA----- 161

```

Figure 1.3. Amino acid sequence alignment of DHFRs from various sources. The sequence alignment was done by CLUSTAL W (1.83) which is a multiple sequence alignment program.¹⁰²⁻¹⁰⁴ The conserved residues are indicated by * for the variants, conservatively replaced residues are indicated by :, and the residues replaced drastically are indicated by . representation.

Ser49, and Leu54 are involved in substrate and cofactor bindings which are conserved in all DHFRs.¹⁰¹ Figure 1.4 shows the cofactor and substrate bound ternary structure of *Ec*DHFR. This following section is a short literature review about structure, kinetic and dynamics aspects of *Ec*DHFR.

Benkovich and coworkers established the kinetic scheme of *Ec*DHFR¹⁰⁰ by studying numerous binary and ternary complexes of the enzyme with substrates and cofactors at varying concentrations and pH conditions. Binding studies showed that

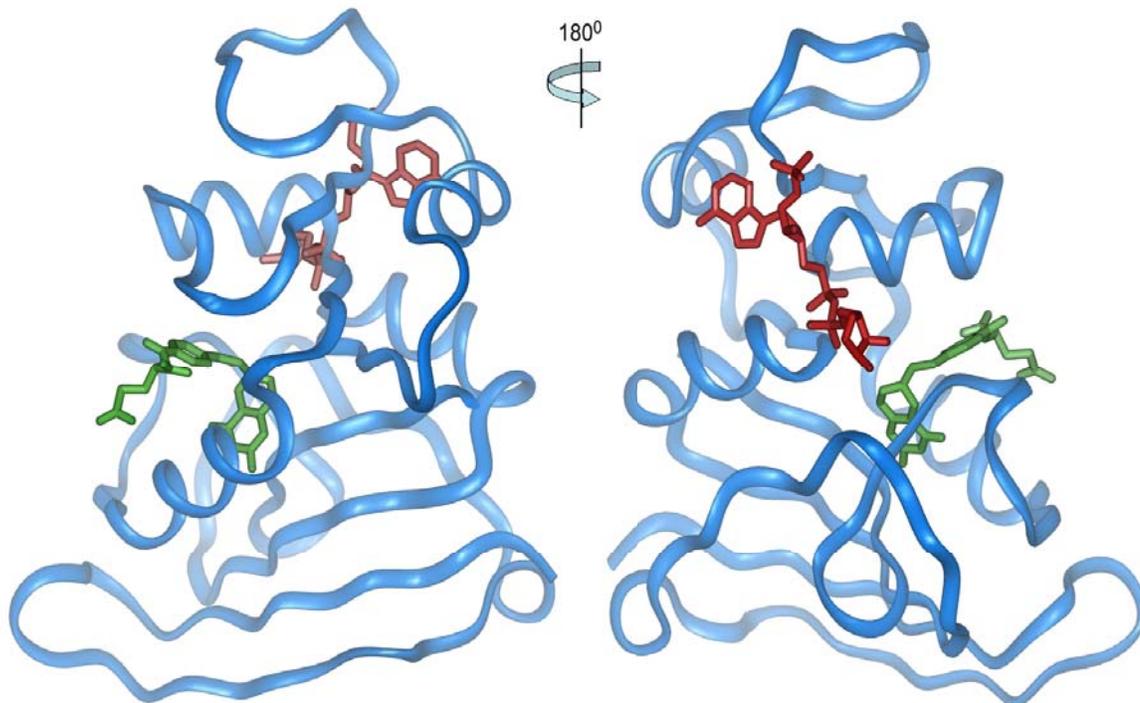


Figure 1.4. The structural illustration of substrate and cofactor binding of *EcDHFR*. The molecule in green is the DHF and the one in the red is NADPH. The reaction takes place at a hydrophobic cleft of the enzyme. The structure is redrawn from the PDB submission of 1bdr.¹⁰⁵

association and dissociation of the substrates occurs at different rates. The studies revealed that five intermediates are involved in the catalytic cycle. The hydride transfer takes place in the Michaelis complex, DHFR:H₂F:NADPH, and yields the DHFR:H₄F:NADP⁺. After the NADPH is released, the product is discharged only after replacing the oxidized coenzyme NADP⁺ with fresh NADPH. This is the rate determining step in the catalytic cycle. Overall, the catalytic cycle is sequentially occurs between the substrate, intermediates and coenzyme. The hydride transfer is a rate limiting process which is a pH dependent process with an optimal pK_a of 6.5. This pK_a was relatively high for the carboxylate group of Asp27 which is the only ionizable side group at the active site. This feature was attributed to the dielectric characteristic of

the hydrophobic pocket where the binding events take place. Shrimpton et.al. conducted molecular dynamic simulation studies and showed that a water molecule can access the active site to achieve one proton transfer to the N5 of DHF in the Michaelis complex.⁶⁹ Chen et.al. proposed that proton transfer may take place from a bound water molecule in the transition state.¹⁰⁶

The crystal structures of *Ec*DHFR have been studied with a variety of substrates, substrate derivatives, cofactors, and cofactor derivatives. These studies aimed to explain the mechanism involved in its catalytic cycle. Kraut and coworkers studied the structure of *Ec*DHFR in its apoenzyme state,¹⁰⁷ binary complex with MTX,¹⁰⁵ folic acid,¹⁰⁸ NADP⁺,¹⁰⁵ ternary complex with folate and NADP⁺,¹⁰⁵ folate and 5-deazafolate, 5,10-dideazatetrahydrofolate (ddTHF).¹⁰⁸ From the crystal structure data for different complexes representing different catalytic cycle stages, it was proposed that five intermediates were involved in catalytic turnover. These studies showed that the catalytic cycle is regulated by the successive movements of three loops; the M20 loop (residues 9-24), the F-G loop (residues 116-132), and the G-H loop (residues 142-150). The loop motions facilitate binding of substrate, cofactor and product release. Within these loops, the M20 loop plays a significant role in the catalytic cycle; as a result, special attention was paid to the conformations of this loop.⁹² Studies showed the M20 loop can be found in four identifiable conformations; *occluded*, *closed*, *open*, and *disordered*. Figure 1.5 shows the superimposition of three *Ec*DHFR structures in the *open*, *closed*, and *occluded* conformations. The M20 loop is in a *disordered* state in apoenzyme and in

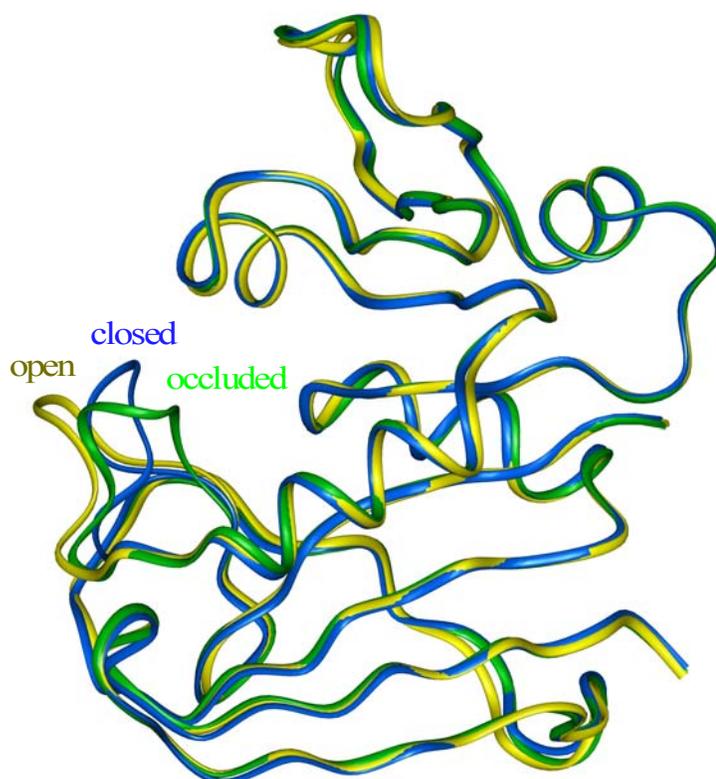


Figure 1.5. The figure represents the superimposition of three X-ray structures of *Ec*DHFR. The open (yellow), closed (blue) and occluded (green) conformations of the M20 loop are shown. The PDB codes are 1rx6¹⁰⁹, 1rx1¹⁰⁹, and 1ra9¹⁰⁹ respectively.

The MXT bound binary complex¹⁰⁹ which was considered to arise from conformational transitions taking between the *occluded* and *closed* states.¹⁰⁹ The *closed* conformation was observed for the ternary complexes (E:MTX:NADPH and E:FOL:NADP⁺) and binary complexes (E:NADPH and E:NADP⁺).¹⁰⁹ The *occluded* conformation was detected in the ternary complexes (E:THF:NADP⁺ and E:THF:NADPH) and binary complexes (E:THF and E:FOL).¹⁰⁹ The three conformations of the M20 loop are distinguished from each other according to specific conformations of the M20 loop residues and a hydrogen bonding pattern with the residues of the F-G and the G-H loops.¹⁰⁹ While the loop is in the *occluded* conformation, G15 (CO) and E17 (NH)

residues of the M20 loop form hydrogen bond with D122 (NH and O ϵ 2 atoms, respectively) of the F-G loop. The center residues of the M20 loop (17-20) hinder the coenzyme binding site.¹⁰⁹ In this conformation, another hydrogen bond forms between N23 of the M20 loop and S148 of G-H loop (N23 (CO) - S148 (NH) and N23 (NH) - S148 (O γ)). In the *closed* state, the inner residues of the M20 (16-19) loop form a short anti-parallel sheet and a type III turn. The hydrogen bonds form between G15-D122 and E17-D122 pairs from the M20 and F-G loops.¹⁰⁹ The substrate, cofactor binding, and product release result in the *open* conformation, which was not detected in any certain complex but was suggested to be a transition state between the *open* and *closed* conformations.¹⁰⁹ In the *closed* conformation, due to conformational change of residues 16-19, the loop interacts with the cofactor through hydrogen bonding and *van der Waals* interactions to facilitate binding of the cofactor.¹⁰⁹ In the *closed* conformation, the active site is protected from the solvent and it lets the substrate and cofactor align to the position of which the hydride transfers takes place.¹⁰⁹ Conversely, in the *occluded* conformation, the binding of cofactor is restricted. From the crystallographic data indication, it was proposed that while the enzyme complexes with substrate and cofactor, the M20 loop is in *closed* conformation. Once the hydride transfer takes places, the loop shifts to the *occluded* conformation and the cofactor is forced out of the active site.¹⁰⁹ During the catalytic cycle the M20 loop spends a substantial amount of time in *closed* and *occluded* conformations.¹⁰⁹ The binding and release events of cofactor are facilitated by the relative movement of the adenosine binding domain with respect to the major domain.¹⁰⁹ In the catalytic cycle, the *open* conformation of the M20 loop was suggested to be in a transition state between the *closed* and *occluded* conformations.¹⁰⁹

Write and co-workers studied the solution state structure and dynamic aspects of *Ec*DHFR by NMR Spectroscopy.¹¹⁰ The NMR studies showed that in the solution state, *Ec*DHFR exists in two slowly interchanging conformations in both apoenzyme and binary complexes with inhibitor MTX, each with different kinetic features.⁹¹ Similarly, the presence of this inter-conversion was detected in some other DHFRs such as *H.sapiens*⁹⁴ *L.casei*.¹¹¹ After completing the backbone and side chain chemical shift assignment of the folate bound enzyme it was demonstrated that the isomerization is not valid for the folate bound binary complex. In this study, a ¹³C labeled folate with unlabeled enzyme was studied. The *occluded* and *closed* conformations of the M20 loop were also detected in NMR structure studies.¹¹² Two dimensional NOESY experiments were used to detect the volume change of the proton cross peaks with respect to time belonging to side chain Trp22 residue (N^εH). As stated before, the M20 loop region, including the Trp22 was found in three different conformational states. The rate of conformational change of the M20 loop for apoenzyme state was calculated as 35 s⁻¹.¹¹³ Relaxation data were conducted to investigate the backbone and side chain motions for the complexes representing pre-Michaelis, Michaelis, and product release states.¹¹⁴ The chemical shift marker for binding of substrates and cofactor were identified from analysis of ¹⁵N, ¹³C^α, and ¹H^N chemical shifts of E:NADPH, folate, E:folate binary,^{112,113} E:folate:DHNADPH ternary¹¹² and E:folate:NADP⁺ ternary complexes.¹¹² The residue in the loop regions (M20 and FG loops) exhibited distinct chemical shift patterns in the different complexes at which the conformations of the loops are different. The timescale of motion of E:folate, E:folate:DHNADPH, E:folate:NADP⁺ complexes were investigated by NMR relaxation studies. The main flexibility variations were

detected in the loop regions due to different conformational states and interaction with each others. In the E:folate and E:folate:DHNADPH complexes, the M20 loop adopts the *occluded* conformation and the two complexes exhibited similar backbone dynamic features over the protein. The M20 loop, the adenosine binding loop, and F-G loops showed high flexibility from NMR dynamics calculations with a high internal correlation time, which indicated the motions are in the picosecond to nanosecond timescales in these regions. In both complexes, residues at the beginning of the M20 loop showed millisecond to microsecond timescale motion according to the R_{ex} term obtained from NMR dynamics calculations. The presence of picoseconds to nanosecond mobility of the M20 loop in the E:folate:NADP⁺ complex was not detected and the backbone mobility showed a drastic variation with respect to the two complexes. The M20 loop in these complexes adopts the *closed* conformation. The motion in the M20 and F-G loops were attenuated with decreasing internal correlation time. Examining the crystal structure gave an explanation for this observation. In the *occluded* state the M20 loop is solvent exposed and has a larger freedom of motion than in the *closed* conformation at which there is interaction between the M20 and F-G loops. Also in the *closed* state there is a steric hindrance due to NADPH packing against the residues in the M20 loops. The millisecond to microsecond time scale in this complex was observed for G121 which was explained from the formation of hydrogen bonds between Asn122 and Gly15. NMR relaxation studies showed that the M20 loop was fluctuating between the *closed* and the *occluded* conformations at about the rate of the catalytic cycle for the DHFR:folate:NADP⁺ ternary complexes. From chemical shift changes and relaxation

dispersion experiments, the exchange rate for the *closed* to *occluded* transition was calculated to be $31 \pm 1.7 \text{ s}^{-1}$ and $524 \pm 30 \text{ s}^{-1}$ for the reverse process.

The Enzyme Dihydrofolate Reductase (DHFR) from *Haloferax Volcanii*

Genetic and functional aspect of halophilic archaean and enzymes of those organisms from the Dead Sea have attracted much interest.^{39,115-117} The enzyme DHFR from *Haloferax volcanii* was identified and isolated from the Dead Sea. The enzyme has been subjected to numerous molecular biological studies. The enzyme is sensitive toward the inhibiting agents; methotrexate (MTX) and trimethoprim (TMP).¹¹⁸

Two genes, *hdrA* and *hdrB*, were encoded from *Haloferax volcanii*, which overexpresses two DHFR enzymes, *HvDHFR1* and *HvDHFR2*. The two enzymes share a 39% amino acid sequence similarity.¹¹⁹ *HvDHFR2* contains more acidic residues and has an acidic patch near the N-terminus like *Halobacterium ferredoxin*.¹¹⁹ No structural data are available for *HvDHFR2*. In spite of the high amino acid sequence similarities between *HvDHFR1* and *HvDHFR2*, the two enzymes exhibited entirely different biochemical features. The production levels of THF, salt concentration and pH dependent enzymatic activity of each enzyme are greatly different. While *HvDHFR1* has an increasing activity toward increasing salt concentrations (3.5 M KCl), *HvDHFR2* showed a decreasing activity as salt concentrations were increased.¹²⁰ While the maximum enzymatic activity for *HvDHFR1* was at a pH 4, it is 6.5 for *HvDHFR2*. These results showed that *HvDHFR1* exhibits more halophilic characteristics.¹²⁰

The enzyme DHFR1 from *Haloferax volcanii* is a monomeric enzyme; consisting of 162 amino acids and a molecular mass of 17.9 kDa. The enzyme *HvDHFR1* is a

moderate halophilic enzyme, and is an acidic enzyme. While it consists of 15 excess acidic residues (Glu and Asp) over basic residues (Lys and Arg), the homologous *Ec*DHFR consist of only 10 extra negatively charged residues. The ratio of acidic residues over basic residues is not significantly large compared to other halophilic-nonhalophilic homologous couples.⁵⁵ The analysis of primary amino acid sequences showed that the acidic residues are distributed across the protein. The halophilic adaptation of the enzyme to hypersaline environments has also been probed through the study of *Hv*DHFR1.¹²¹

The activity and stability of *Hv*DHFR1 and mesophilic *Ec*DHFR have been studied over a wide range of salt concentrations.¹²⁰ The enzyme exhibited relatively higher enzymatic activity in KCl than NaCl. This feature was attributed to the KCl being the naturally accumulated cosolute in the bacterial cell. The stability of the enzyme was studied by measuring the residual activity of the enzyme. The enzyme is stable as salt concentration is lowered to 1.5 M NaCl, lower than this resulted in diminished the enzymatic activity. Site directed mutagenic studies have been carried out to test the role of catalytic residues on the active site. The catalytic role of residue Asp27 has been well defined for *Ec*DHFR. The corresponding residue, Asp29 in *Hv*DHFR1 is followed by two positively charges residues, Lys30 and Lys31.¹²² The reaction rate and dissociation of THF were tested by mutation studies at varying salt concentrations. In one of these mutation studies, Lys30 was replaced by leucine, in another study Lys31 was replaced by alanine (Ala), and in another study both residues were replaced by both alanine and leucine. In each mutation, the k_{cat} value increased but did not change across varied salt concentrations. The highest rise in k_{cat} was obtained in the K30L mutation at which k_{cat}

value increased from 3.14 s^{-1} to 10.0 s^{-1} .¹²² The K_m value was higher in each mutation but decreased as salt concentration is increased. Decreasing the positive charge at the active site resulted in an increase in the activity of the enzyme with increasing substrate affinity.

It was shown that the enzymatic activity of *Hv*DHFR1 is a pH dependent process. The effect of varying pH on enzymatic activity at different salt concentrations were probed.¹²³ At pH 7, enzymatic activities exhibited a linear increase with increased salt concentration.¹²³ As the pH was reduced, the linearity vanished and at pH 4 the activity at low salt concentration exhibited the highest activity and remained constant as salt concentration was increased.¹²³

*Hv*DHFR1 has been used as model enzyme in the explanation of halophilic adaptation.¹²¹ The computational studies were performed by using the crystal structures of DHFR from *Escherichia coli* and *Lactobacillus casei* (since no crystal structure of *Haloferax volcanii* DHFR1 was available). From the generated protein topology, it was proposed that *Hv*DHFR1 has an unbalanced charge distribution over the protein surface.¹²¹ Positively charged residues are suggested to be located at the active site with a larger molecular dipole which was due to the unique insertation of Lys30 and Lys31.¹²¹ It was proposed that the large dipole is facilitating the substrate and cofactor binding.¹²¹ This close similar charge proximity resulted in unfavorable electrostatic interaction and was believed to cause the inactivation of the enzyme through unfolding as the salt concentration is decreased.¹²¹ The modeled structure was also used to explain the stability of *Hv*DHFR1 at low salt concentration and low pH.¹²¹ At low salt conditions, it was suggested that the unfavorable electrostatic interaction is increasing due to unshielding of the charged groups.¹²¹ The protonation of the charged groups at low pH

was proposed to screen the unfavorable electrostatic interaction between similarly charged ionizable groups.¹²¹

The crystal structure of *Hv*DHFR1 was only available in 1997 at 2.6 Å resolution by Herzberg and co-workers (PDB entry code:1vdr).⁸³ The structure of *Hv*DHFR1 was the third halophilic protein structure investigated and represents the only structural study for *Haloferax volcanii*. Two structures were modeled in the study with C α superimposition value of 0.6 Å.⁸³ Eighty bound water molecules were determined for one of the structures.⁸³ The X-ray crystallographically derived structure of *Hv*DHFR1 (in 2.4 M phosphate buffer) exhibited Rossmann folding where two pair of α -helices are packed against a β -sheet formed from eight β -strands (Figure 1.6).⁸³ A high crystallographic temperature factor obtained for the L21 loop (which corresponds to the M20 loop in *Ec*DHFR) indicates the presence of conformational fluctuations.⁸³ Nevertheless, the L21 loop was found to be similar to the closed conformation identified for *Ec*DHFR.⁸³ A common feature of all known DHFRs, the unusual *cis* conformation between two glycine residues, is also found in the *cis* conformation between G101 and G102.⁸³ Binding of methotrexate was not successful due to unfavorable directing of the carboxylate group of Asp55 toward the negatively charged side of methotrexate.⁸³ The residues Asp54 and Asp55 are in close proximity with the residues Lys30 and Lys31 where an electrostatic interaction is proposed to be present.⁸³

In previous studies, the mutation of Lys30 and Lys31 resulted in an increase of k_{cat} and K_{m} values. Based on the crystal structure, it was suggested that the binding events were enhanced by decreasing the positive charge at the vicinity of binding site. Electrostatic surface potential revealed that while positively charged residues are

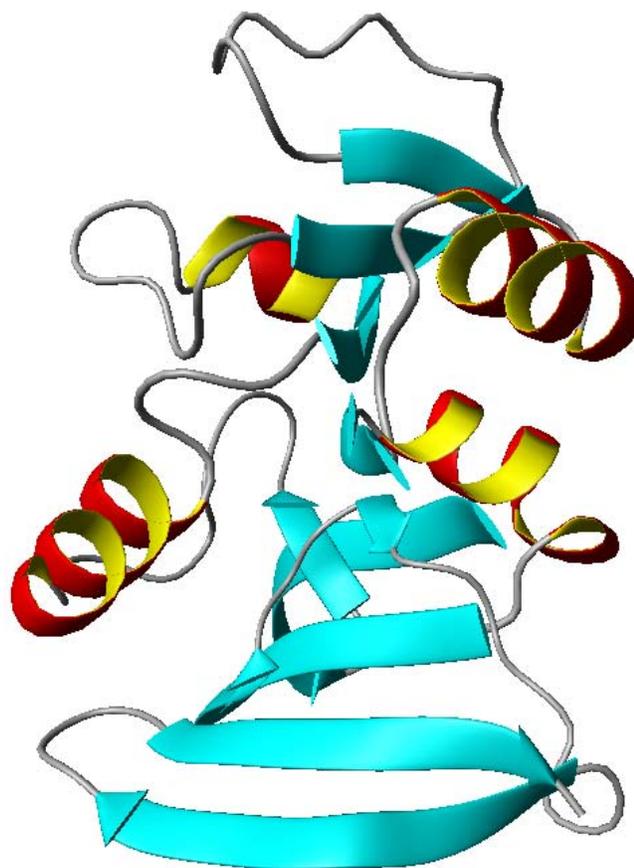


Figure 1.6. The crystal structure of *Hv*DHFR1 is redrawn from the pdb file 1vdr⁸³ by using the software MOLMOL.¹²⁴

centered at the vicinity of the active site, the negatively charged residues are spread over the surface. The stability of *Hv*DHFR1 has been studied comparatively with homologous *Ec*DHFR over a wide range of salt concentration by circular dichroism (CD) and fluorescence (FL) spectroscopy by Gloss and co-workers.¹²⁰ FL spectroscopy data revealed that lowering salt concentration has no major effect on the tertiary structure of *Hv*DHFR1. A similar effect was observed for *Ec*DHFR, for which over the salt concentration, *Ec*DHFR did not show a significant spectral change. The FL and CD spectroscopy data revealed that *Hv*DHFR1 did not unfold completely until the salt

concentration was lowered to 0.12 M KCl. The stability studies with different cosolutes showed that *Ec*DHFRs is more stable than *Hv*DHFR1 in KCl than CsCl. The comparative stability test with urea denaturation showed that the enzyme *Ec*DHFR is more stable than *Hv*DHFR1 by 6-8 kcal/mole over the range of salt concentrations. Binding studies of *Hv*DHFR1 showed an interesting affinity trend for substrate and cofactor as a function of salt concentration. There is a slight loss of affinity for DHF as salt concentration is lowered; however, the affinity for NADPH becomes stronger. On the other hand, *Ec*DHFR showed no significant affinity change for either DHF or NADPH as salt concentration was increased to 1.0 M, the concentration at which 80-90% of activity is lost. The loss of activity for *Ec*DHFR as salt concentration was increased was proposed to be associated with the loss of enzyme flexibility required in the catalytic steps rather than with structural changes. For *Hv*DHFR1 it has been proposed that the loss of activity as salt concentration was lowered was associated to the loss of enzyme flexibility.

Objectives of This Research

Very few proteins from different halophilic organisms have been investigated with X-ray crystallography at a range of salt concentrations between 0.3 M and 4.0 M^{62-64,125-129}. Only *Halobacterium salinarum* ferredoxin (HsFdx) has been studied by NMR at 0.45 M salt, which is far below the optimal salt concentration (~2.5-4.3 M NaCl) for this protein. Appropriate salt concentration for halophilic proteins is a vital requirement for these proteins to fold in their native forms, and also for structural stability and activity. Studies showed that maintaining less than the required salt resulted in destabilization and inactivation of halophilic proteins.^{34,38,46,50,53,64,83} However, only a

few of the reported halophilic proteins have been crystallized at the required salt concentration. *Haloferax volcanii* flourishes in an extremely high salt environment, the Dead Sea, which has 3-4 M monovalent salt concentration.²¹ Nevertheless, the structure of the enzyme was investigated by crystallizing in 2.4 M phosphate buffer.⁸³ There is a need for more three dimensional data of halophilic proteins for a better understanding of the structural features and how they relate to haloadaptation. In this study, three dimensional liquid state NMR derived structure of *Hv*DHFR1 at 3.5 M NaCl was investigated. This study represents the first protein studied by NMR at such an extreme salt concentration. Once the solution state structure is completed relaxation studies were conducted at 3.5 M and 1.0 M NaCl salt concentrations to probe flexibility. This research will lay the ground work for future studies that will investigate the structures and flexibilities of both *Hv*DHFR1 and *Ec*DHFR in their binary and ternary complexes.

REFERENCES

- (1) Chaban, B.; Ng, S. Y.; Jarrell, K. F. *Can. J. Microbiol.* **2006**, *52*, 73-116.
- (2) Pace, N. R. *Science* **1997**, *276*, 734-740.
- (3) Van Valen, L. M.; Maiorana, V. C. *Nature* **1980**, *287*, 248-250.
- (4) Rothschild, L. J.; Mancinelli, R. L. *Nature* **2001**, *409*, 1092-1101.
- (5) Segerer, A. H.; Burggraf, S.; Fiala, G.; Huber, G.; Huber, R.; Pley, U.; Stetter, K. *O. Origins Life Evol. Biosphere* **1993**, *23*, 77-90.
- (6) Kohshima, S. *Biol. Sci. Space* **2000**, *14*, 353-362.
- (7) Dubnovitsky, A. P.; Kapetaniou, E. G.; Papageorgiou, A. C. *Protein Sci.* **2005**, *14*, 97-110.
- (8) Dennis, P. P.; Shimmin, L. C. *Microbiol. Mol. Biol. Rev.* **1997**, *61*, 90-104.
- (9) Niehaus, F.; Bertoldo, C.; Kahler, M.; Antranikian, G. *Appl. Microbiol. Biotechnol.* **1999**, *51*, 711-729.
- (10) van den Burg, B. *Curr. Opin. Microbiol.* **2003**, *6*, 213-218.
- (11) Egorova, K.; Antranikian, G. *Curr. Opin. Microbiol.* **2005**, *8*, 649-655.
- (12) Antranikian, G.; Vorgias, C. E.; Bertoldo, C. *Adv. Biochem. Eng. Biotechnol.* **2005**, *96*, 219-262.
- (13) Marhuenda-Egea, F. C.; Piera-Velazquez, S.; Cadenas, C.; Cadenas, E. *Archaea* **2002**, *1*, 105-111.
- (14) Rodriguez-Valera, F. *Biochem. Soc. Symp.* **1992**, *58*, 135-147.
- (15) Margesin, R.; Schinner, F. *Extremophiles* **2001**, *5*, 73-83.
- (16) David C Demirjian, D. C.; Morís-Varas, F.; Cassidy, C. S. *Curr. Opin. Microbiol.* **2001**, *5*, 144-150.
- (17) Oren, A. *J. Ind. Microbiol. Biotechnol.* **2002**, *28*, 56-63.

- (18) Suriyamongkol, P.; Weselake, R.; Narine, S.; Moloney, M.; Shah, S. *Biotechnol. Adv.* **2007**, *25*, 148-175.
- (19) Pineda-Flores, G.; Mesta-Howard, A. M. *Rev Latinoam. Microbiol.* **2001**, *43*, 143-150.
- (20) Fujiwara, S. *J. Bios. Bioeng.* **2002**, *94*, 518-525.
- (21) Grant, W. D.; Germmell, R. T.; McGenity, T. J. Halophiles. In *Extremophiles*; Horikoshi, K., Grant, W. D., Eds.; Wiley-Liss: New York, 1998; pp 93-132.
- (22) Lanyi, J. K. *Bacteriol. Rev.* **1974**, *38*, 272-290.
- (23) Tsai, C. R.; Garcia, J. L.; Patel, B. K.; Cayol, J. L.; Baresi, L.; Mah, R. A. *Int. J. Syst. Bacteriol.* **1995**, *45*, 301-307.
- (24) Paterek, J. R.; Smith, P. H. *Appl. Environ. Microbiol.* **1985**, *50*, 877-881.
- (25) Jin, Y.; Weining, S.; Nevo, E. *Proc Natl. Acad. Sci. U. S. A* **2005**, *102*, 18992-18997.
- (26) Cemen, I. I.; Goncuoglu, M. C.; Dirik, K. *J. Geol.* **1999**, *107*, 693-706.
- (27) DasSarma, S.; Arora, P. *Encyclopedia Life Sci.* **2001**, 1-9.
- (28) Woese, C. R.; Kandler, O.; Wheelis, M. L. *Proc Natl. Acad. Sci. U. S. A* **1990**, *87*, 4576-4579.
- (29) Gutierrez, M. C.; Kamekura, M.; Holmes, M. L.; Dyall-Smith, M. L.; Ventosa, A. *Extremophiles* **2002**, *6*, 479-483.
- (30) Gupta, R. S. *Mol. Microbiol.* **1998**, *29*, 695-707.
- (31) Lake, J. A. *Prog. Nucleic. Acid. Res. Mol. Biol.* **1983**, *30*, 163-194.
- (32) <http://www.microbiol.unimelb.edu.au/people/dyallsmith/research/index.html#HalovirusRefs> (accessed Jan 2007).
- (33) <http://www.the-icsp.org/taxa/halobacterlist.htm> (accessed Dec 2005).
- (34) Rao, J. K.; Argos, P. *Biochemistry* **1981**, *20*, 6536-6543.
- (35) Scandurra, R.; Consalvi, V.; Chiaraluce, R.; Politi, L.; Engel, P. C. *Front. Biosci.* **2000**, *5*, D787-795.
- (36) Kloda, A.; Martinac, B. *Biophys. J.* **2001**, *80*, 229-240.

- (37) Grant, W. D.; Gemmell, R. T.; McGenity, T. J. *Extremophiles* **1998**, *2*, 279-287.
- (38) Madern, D.; Ebel, C.; Zaccai, G. *Extremophiles* **2000**, *4*, 91-98.
- (39) Soppa, J. *Microbiology* **2006**, *152*, 585-590.
- (40) Roberts, M. F. *Front. Biosci.* **2004**, *9*, 1999-2019.
- (41) Jaenicke, R. *Eur J. Biochem.* **1991**, *202*, 715-728.
- (42) Baldwin, R. L. *Biophys. J.* **1996**, *71*, 2056-2063.
- (43) Neagu, A.; Neagu, M.; Der, A. *Biophys. J.* **2001**, *81*, 1285-1294.
- (44) Gong, P.; Szleifer, I. *J. Colloid. Interface. Sci.* **2004**, *278*, 81-90.
- (45) Zhang, Y.; Cremer, P. S. *Curr. Opin. Chem. Biol.* **2006**, *10*, 1-6.
- (46) Ebel, C.; Faou, P.; Kernel, B.; Zaccai, G. *Biochemistry* **1999**, *38*, 9039-9047.
- (47) Zhou, H. X. *Proteins* **2005**, *61*, 69-78.
- (48) Ebel, C.; Costenaro, L.; Pascu, M.; Faou, P.; Kernel, B.; Proust-De Martin, F.; Zaccai, G. *Biochemistry* **2002**, *41*, 13234-13244.
- (49) Costenaro, L.; Zaccai, G.; Ebel, C. *Biochemistry* **2002**, *41*, 13245-13252.
- (50) Zaccai, G.; Eisenberg, H. *Trends Biochem. Sci.* **1990**, *15*, 333-337.
- (51) Eisenberg, H. *Arch. Biochem. Biophys.* **1995**, *318*, 1-5.
- (52) Mevarech, M.; Frolow, F.; Gloss, L. M. *Biophys. Chem.* **2000**, *86*, 155-164.
- (53) Elcock, A. H.; McCammon, J. A. *J. Mol. Biol.* **1998**, *280*, 731-748.
- (54) Christian, J. H.; Waltho, J. A. *Biochim. Biophys. Acta.* **1962**, *65*, 506-508.
- (55) Bohm, G.; Jaenicke, R. *Int. J. Pept. Protein Res.* **1994**, *43*, 97-106.
- (56) Zaccai, G.; Cendrin, F.; Haik, Y.; Borochoy, N.; Eisenberg, H. *J. Mol. Biol.* **1989**, *208*, 491-500.
- (57) Zaccai, G.; Bunick, G. J.; Eisenberg, H. *J. Mol. Biol.* **1986**, *192*, 155-157.
- (58) Zaccai, G.; Wachtel, E.; Eisenberg, H. *J. Mol. Biol.* **1986**, *190*, 97-106.
- (59) Calmettes, P.; Eisenberg, H.; Zaccai, G. *Biophys. Chem.* **1987**, *26*, 279-290.

- (60) Reich, M. H.; Kam, Z.; Eisenberg, H. *Biochemistry* **1982**, *21*, 5189-5195.
- (61) Kuntz, I. D. *J. Am. Chem. Soc.* **1971**, *93*, 516-518.
- (62) Dym, O.; Mevarech, M.; L., S. J. *Science* **1995**, *267*, 1344-1346.
- (63) Richard, S. B.; Madern, D.; Garcin, E.; Zaccari, G. *Biochemistry* **2000**, *39*, 992-1000.
- (64) Frolova, F.; Harel, M.; Sussman, J. L.; Mevarech, M.; Shoham, M. *Nat. Struct. Biol.* **1996**, *3*, 452-458.
- (65) Pieper, U.; Kapadia, G.; Mevarech, M.; Herzberg, O. *Structure* **1998**, *6*, 75-88.
- (66) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196-1202.
- (67) Levin, I.; Giladi, M.; Altman-Price, N.; Ortenberg, R.; Mevarech, M. *Mol. Microbiol.* **2004**, *54*, 1307-1318.
- (68) Basran, J.; Casarotto, M. G.; Barsukov, I. L.; Roberts, G. C.; Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochemistry* **1995**, *34*, 2872-2882.
- (69) Shrimpton, P.; Allemann, R. K. *Protein Sci.* **2002**, *11*, 1442-1451.
- (70) Casarotto, M. G.; Basran, J.; Badii, R.; Sze, K. H.; Roberts, G. C. *Biochemistry* **1999**, *38*, 8038-8044.
- (71) Herrmann, L.; Bockau, U.; Tiedtke, A.; Hartmann, M. W.; Weide, T. *BMC Biotechnol.* **2006**, *6*, 21.
- (72) Gangjee, A.; Kurup, S.; Namjoshi, O. *Curr. Pharm. Des.* **2007**, *13*, 609-639.
- (73) Pena, M. M.; Xing, Y. Y.; Koli, S.; Berger, F. G. *Biochem. J.* **2006**, *394*, 355-363.
- (74) Srimatkandada, S.; Dube, S. K.; Carmen, M.; Bertino, J. R. *Oncol. Res.* **2000**, *12*, 11-15.
- (75) Morganti, M.; Coronello, M.; Caciagli, B.; Biondi, C.; Quattrone, A.; Capaccioli, S.; Mazzei, T.; Mini, E. *Anticancer Drugs* **2000**, *11*, 285-294.
- (76) Budak-Alpdogan, T.; Alpdogan, O.; Banerjee, D.; Wang, E.; Moore, M. A.; Bertino, J. R. *Mol. Ther.* **2004**, *10*, 574-584.
- (77) Nano, R.; Invernizzi, R.; Gerzeli, G.; Michienzi, M. *Basic Appl. Histochem.* **1983**, *27*, 67-69.

- (78) Wessels, J. A.; de Vries-Bouwstra, J. K.; Heijmans, B. T.; Slagboom, P. E.; Goekoop-Ruiterman, Y. P.; Allaart, C. F.; Kerstens, P. J.; van Zeben, D.; Breedveld, F. C.; Dijkmans, B. A.; Huizinga, T. W.; Guchelaar, H. J. *Arthritis Rheum.* **2006**, *54*, 1087-1095.
- (79) Ramanan, A. V.; Whitworth, P.; Baidam, E. M. *Arch. Dis. Child.* **2003**, *88*, 197-200.
- (80) Hawser, S.; Lociuero, S.; Islam, K. *Biochem. Pharmacol.* **2006**, *71*, 941-948.
- (81) Navarro-Martinez, M. D.; Navarro-Peran, E.; Cabezas-Herrera, J.; Ruiz-Gomez, J.; Garcia-Canovas, F.; Rodriguez-Lopez, J. N. *Antimicrob. Agents Chemother.* **2005**, *49*, 2914-2920.
- (82) Schnell, J. R.; Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 119-140.
- (83) Pieper, U.; Kapadia, G.; Mevarech, M.; Herzberg, O. *Structure* **1998**, *6*, 75-88.
- (84) Martorell, G.; Gradwell, M. J.; Birdsall, B.; Bauer, C. J.; Frenkiel, T. A.; Cheung, H. T.; Polshakov, V. I.; Kuyper, L.; Feeney, J. *Biochemistry* **1994**, *33*, 12416-12426.
- (85) McTigue, M. A.; Davies, J. F., 2nd; Kaufman, B. T.; Kraut, J. *Biochemistry* **1992**, *31*, 7264-7273.
- (86) Davies, J. F., 2nd; Delcamp, T. J.; Prendergast, N. J.; Ashford, V. A.; Freisheim, J. H.; Kraut, J. *Biochemistry* **1990**, *29*, 9467-9479.
- (87) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650-13662.
- (88) Oefner, C.; D'Arcy, A.; Winkler, F. K. *Eur. J. Biochem.* **1988**, *174*, 377-385.
- (89) Volz, K. W.; Matthews, D. A.; Alden, R. A.; Freer, S. T.; Hansch, C.; Kaufman, B. T.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 2528-2536.
- (90) Zusman, T.; Rosenshine, I.; Boehm, G.; Jaenicke, R.; Leskiw, B.; Mevarech, M. *J. Biol. Chem.* **1989**, *264*, 18878-18883.
- (91) Falzone, C. J.; Cavanagh, J.; Cowart, M.; Palmer, A. G., 3rd; Matthews, C. R.; Benkovic, S. J.; Wright, P. E. *J. Biomol. NMR* **1994**, *4*, 349-366.
- (92) Li, L.; Falzone, C. J.; Wright, P. E.; Benkovic, S. J. *Biochemistry* **1992**, *31*, 7826-7833.
- (93) Falzone, C. J.; Wright, P. E.; Benkovic, S. J. *Biochemistry* **1994**, *33*, 439-442.

- (94) Johnson, J. M.; Meiering, E. M.; Wright, J. E.; Pardo, J.; Rosowsky, A.; Wagner, G. *Biochemistry* **1997**, *36*, 4399-4411.
- (95) Ivery, M. T.; Gready, J. E. *J. Med. Chem.* **1994**, *37*, 4211-4221.
- (96) Feeney, J. *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 290-312.
- (97) Andrews, J.; Fierke, C. A.; Birdsall, B.; Ostler, G.; Feeney, J.; Roberts, G. C.; Benkovic, S. J. *Biochemistry* **1989**, *28*, 5743-5750.
- (98) Thillet, J.; Adams, J. A.; Benkovic, S. J. *Biochemistry* **1990**, *29*, 5195-5202.
- (99) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. S. *J. Biol. Chem.* **1982**, *257*, 13650-13662.
- (100) Fierke, C. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 4085-4092.
- (101) Beverley, S. M.; Ellenberger, T. E.; Cordingley, J. S. *Proc. Natl. Acad. Sci. U.S. A* **1986**, *83*, 2584-2588.
- (102) Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. *Nucleic Acids Res.* **1997**, *25*, 4876-4882.
- (103) Higgins, D. G.; Thompson, J. D.; Gibson, T. J. *Methods Enzymol.* **1996**, *266*, 383-402.
- (104) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673-4680.
- (105) Bystroff, C.; Oatley, S. J.; Kraut, J. *Biochemistry* **1990**, *29*, 3263-3277.
- (106) Chen, Y. Q.; Kraut, J.; Blakley, R. L.; Callender, R. *Biochemistry* **1994**, *33*, 7021-7026.
- (107) Bystroff, C.; Kraut, J. *Biochemistry* **1991**, *30*, 2227-2239.
- (108) Lee, H.; Reyes, V. M.; Kraut, J. *Biochemistry* **1996**, *35*, 7012-7020.
- (109) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586-603.
- (110) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638-1642.
- (111) Birdsall, B.; Bevan, A. W.; Pascual, C.; Roberts, G. C.; Feeney, J.; Gronenborn, A.; Clore, G. M. *Biochemistry* **1984**, *23*, 4733-4742.
- (112) Osborne, M. J.; Schnell, J.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2001**, *40*, 9846-9859.

- (113) Epstein, D. M.; Benkovic, S. J.; Wright, P. E. *Biochemistry* **1995**, *34*, 11037-11048.
- (114) McElheny, D.; Schnell, J. R.; Lansing, J. C.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A* **2005**, *102*, 5032-5037.
- (115) Fine, A.; Irihimovitch, V.; Dahan, I.; Konrad, Z.; Eichler, J. J. *Bacteriol.* **2006**, *188*, 1911-1919.
- (116) Bidle, K. A.; Hanson, T. E.; Howell, K.; Nannen, J. *Extremophiles* **2006**.
- (117) Kapatai, G.; Large, A.; Benesch, J. L.; Robinson, C. V.; Carrascosa, J. L.; Valpuesta, J. M.; Gowrinathan, P.; Lund, P. A. *Mol. Microbiol.* **2006**, *61*, 1583-1597.
- (118) Rosenshine, I.; Zusman, T.; Werczberger, R.; Mevarech, M. *Eur. J. Biochem.* **1986**, *208*, 518-522.
- (119) Ortenberg, R.; Rozenblatt-Rosen, O.; Mevarech, M. *Mol. Microbiol.* **2000**, *35*, 1493-1505.
- (120) Wright, D. B.; Banks, D. D.; Lohman, J. R.; Hilsenbeck, J. L.; Gloss, L. M. *J. Mol. Biol.* **2002**, *323*, 327-344.
- (121) Bohm, G.; Jaenicke, R. *Protein Eng.* **1994**, *7*, 213-220.
- (122) Blecher, O.; Goldman, S.; Mevarech, M. *Eur. J. Biochem.* **1993**, *216*, 199-203.
- (123) Eisenberg, H.; Mevarech, M.; Zaccari, G. *Biochemical, Structural, and Molecular Genetic Aspects of Halophilism*. In *Advances in Protein Chemistry*; Academic Press: New York, 1992; pp 1-57.
- (124) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graph.* **1996**, *14*, 51-55, 29-32.
- (125) Bieger, B.; Essen, L. O.; Oesterhelt, D. *Structure* **2003**, *11*, 375-385.
- (126) Zeth, K.; Offermann, S.; Essen, L. O.; Oesterhelt, D. *Proc. Natl. Acad. Sci. U.S.A* **2004**, *101*, 13780-13785.
- (127) Kolbe, M.; Besir, H.; Essen, L. O.; Oesterhelt, D. *Science* **2000**, *288*, 1390-1396.
- (128) Besir, H.; Zeth, K.; Bracher, A.; Heider, U.; Ishibashi, M.; Tokunaga, M.; Oesterhelt, D. *FEBS Lett.* **2005**, *579*, 6595-6600.
- (129) Breiter, D. R.; Meyer, T. E.; Rayment, I.; Holden, H. M. *J. Biol. Chem.* **1991**, *266*, 18660-18667.

CHAPTER II

PROTEIN ISOLATION AND PURIFICATION

Introduction

In order to perform any structural studies by nuclear magnetic resonance (NMR) spectroscopy it is imperative to produce a highly pure sample.^{1,2} This is especially true for protein samples due to the overwhelming number of NMR peaks and peak shapes.³ There are variations in peak shape (height/volume) that are due to the tumbling of the molecule in solution (i.e. the freedom of individual nuclei's motion and the environments the nuclei are experiencing).⁴ Molecular tumbling which is characterized by the correlation time, is the key factor that affects the signal quality. Small molecules with short correlation times give rise to narrow, sharp peaks. As the molecular size increases correlation times get longer and the individual resonance lines are broadened. The local environment of nuclei can also broaden peaks. Broader peaks are not easily observed since they can be obstructed by the other peaks and result in lower resolution. Identification of each peak is important to both chemical shift and structural information. Therefore, it is essential to produce a sample that is > 95% pure to take advantage of filtering and dispersion gained in three dimensional NMR experiments.⁵

All the three dimensional (3D) NMR experiments rely on transferring magnetization through either ^{15}N or ^{13}C nucleus. The natural abundance of NMR active

nuclei ^{13}C (1.1%) and ^{15}N (0.37%) are extremely low.⁶ Therefore, data acquisition on a natural abundance sample takes a very long time, which is not feasible in terms of sample stability and NMR time. Isotope enrichment offers many advantages such as faster data acquisition, information on carbon and nitrogen, and effective means of one, two, and three step magnetization transfer (^1H - ^{15}N , ^1H - ^{13}C , ^{13}C - ^{13}C) among spin systems of amino acids.⁷

Typically, a 1-2 mM sample is required for NMR protein studies. For a medium-size protein, this corresponds to tens of milligrams of protein that is hard to obtain and isolate from natural systems. Protein stability becomes a problem in the course of NMR studies. Some proteins are intrinsically unstable and may only last for a few days. Most of the 3D NMR experiments require 3-7 days for data acquisition. The availability of protein should not be a bottleneck during the course of the NMR experiments. Recombinant protein techniques provide a means to overexpresses and label a protein in a bacterial system. These techniques also provide a means to obtain many milligrams of protein at a moderate cost. Once a protein has been expressed in a bacterial system, some type of chromatography must be employed to obtain a highly pure sample. While the choice of the purification technique and the optimum conditions vary for different proteins, the most commonly used techniques are size exclusion, ion exchange, hydrophobic, and affinity chromatography.

Purification of *Hv*DHFR1 by Column Chromatography

The purification protocols used for a protein should be optimized based on the proteins solubility and stability.⁸⁻¹³ The protein *Hv*DHFR1 which was grown in bacterial

system was found to be in insoluble the minimal media (M9). In order to solubilize the protein, it was denatured with guanidine buffer ($\text{CH}_5\text{N}_2\text{HCl}$) and rapidly re-folded in NaCl buffer. The folded proteins were soluble in 1 M NaCl and the unfolded proteins were precipitated out from the solution. The folded proteins were purified by chromatographic techniques. The choice of the technique depends on the characteristic features of the proteins such as hydrophobicity, charge, and affinity.

Gel filtration was chosen as the first chromatographic technique in the purification of *Hv*DHFR1. Gel filtration technique separates molecules based on variations in size.¹⁴ Molecules are retarded in the network of pores and elute according to their sizes.¹⁵ Molecules that are within the range of exclusion limits are excluded from the pore matrix and elute with the solvent without retention within the void volume. The rest of the molecules would diffuse into the gel pores where separation would be governed by molecular size. The relatively larger molecules in the pore would not be retained as much as the smaller molecules during the elution. The smaller molecules would access more pores and would be trapped more in the gel network which elute last from the column. The size related movement of molecules allows fractionating of different sizes of molecules. In practice, the mixture is eluted isocratically in the column where only one buffer system would be run as a mobile phase. In this study, Sepharose CL-6B[®] was employed as a gel filtration material.¹⁶ The column material, Sepharose, is in bead shaped produced from agarose. The repeating units of agarose, modified polysaccharides, are cross linked with 2,3-dibromopropanol to generate a continuous matrix and also to improve the durability of the material. The optimal separation range for the gel matrix is 10-400 kDa with 45-165 μm bead size.¹⁶ Figure 2.1 shows the

structure of the cross-linked form of Sepharose CL-6B. While separating the *HvDHFR1* protein sample, a decreasing concentration gradient of $(\text{NH}_4)_2\text{SO}_4$ was applied to elute the protein from the stationary phase. It has been known that the protein irreversibly binds this column material at salt concentrations $>2.0 \text{ M } (\text{NH}_4)_2\text{SO}_4$.¹⁷ This situation can be explained from the chemical features of the column material and the protein features. During the manufacturing processing of Sepharose, the content of ionizable groups is eliminated.^{16, 18} However, this removal process may not be complete and the gel matrix may still contain some residual charged groups such as sulfate and carboxylic groups.¹⁶ The charged groups on the polysaccharides would bind the ammonium ions by electrostatic interaction. In turn, the ammonium ions can bind the negatively charged proteins and retain the proteins in the column. The decreasing ionic strength gradient may facilitate a weakening of the attractive interaction and cause the protein to elute from the column. Figure 2.2-A shows UV-Vis spectrophotometric results for the individual

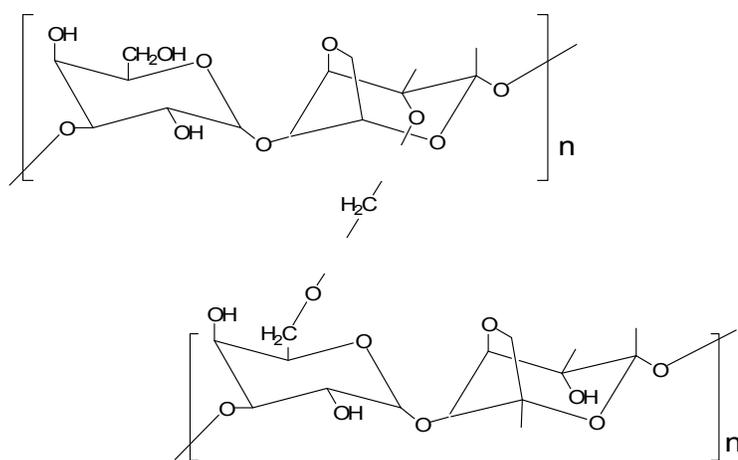


Figure 2.1. The structure of cross-linked Sepharose.¹⁶

fractions. The fractions with a high absorbance value indicate the presence of proteins. For each of the high absorbance fractions, the SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) was run to assure the presence of the protein in the fraction. Figure 2.2-B shows the result of the SDS- PAGE for the first column of fractions. The first band lane represents the marker proteins which have a range of 97-14 kD. The fraction lanes from 24 to 47 between 14 and 21 kDa markers represent the enzyme *HvDHFR1* (17.9 kDa). Due to very weak residual bands on the fraction lines in the gel, a further purification treatment is needed. In order to obtain a highly pure sample, the collected sample was submitted to ion exchange chromatography.

Ion exchange chromatography relies on separating molecules based on their charge features.¹⁹ The principle of separation is based on adsorbing the molecules on an oppositely charged stationary phase through electrostatic interactions and then desorbing

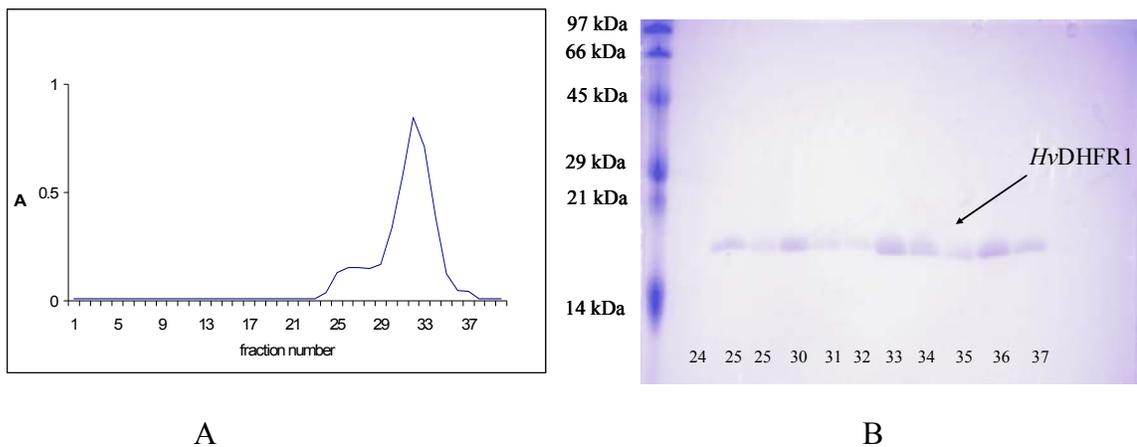


Figure 2.2. The UV-Vis measurement at 280 nm of each fraction from Sepharose CL-6B column is shown on the left (A). The SDS-PAGE results from high absorbance fractions are shown on the right (the fraction numbers are shown at the bottom of the gel)(B). The first line in the gel represents the molecular weight markers which are in the range from 97 kDa (top marker) and 14 kDa (bottom marker).

the solutes through weakening this interaction.²⁰ Initially, the stationary phase conditions would be equilibrated to the sample conditions (i.e. ionic strength, pH, and buffer) to facilitate the adsorption process. The oppositely charged molecules would be adsorbed by the stationary ion exchanger phase. The adsorbed solute molecules are eluted from the column by using a salt or pH gradient. This provides a means to isolate and purify the protein of interest. As the total charge of the protein *HvDHFR1* is negative, a strong anion exchanger, DEAE-Sephacel[®] was employed as the column material in the second purification method. The DEAE (Diethylaminoethyl)-cellulose is a bead shaped cellulose produced from microcrystalline celluloses (40-160 μm).¹⁹ The microcrystalline structure is modified to the form of a bead during processing and the functional charged groups are covalently attached. Figure 2.3 shows the chemical structure of the DEAE. Based on the electrostatic interaction between the positively charged quaternary amino groups and the negatively charged protein molecules, a reversible and specific binding event takes place. The desorption process takes place by running a higher ionic strength buffer that exchanges the adsorbed proteins molecules with counter ions.²¹ In this study, a high ionic strength of NaCl gradient was applied to the column. The interactions of counterions with a stationary phase matrix are stronger than the protein molecules. As the concentration of the counter ions increases, the proteins are desorbed and the counter ions are replaced. UV-Vis absorption measurement at 280 nm for each of the collected fractions was measured and are plotted in Figure 2.4-A. It was calculated that 90% of the first column sample was recovered at the end of the ion exchange chromatography. Each collected fraction with a high absorbance reading was again analyzed by SDS- PAGE. The results are shown the Figure 2.4-B. No other protein band besides the band

representing the *Hv*DHFR1 was observed in the gel. After optimizing the purification procedures for unlabelled protein samples, the procedure was applied to the triple labeled (^2H , ^{15}N , ^{13}C) *Hv*DHFR1 protein samples.

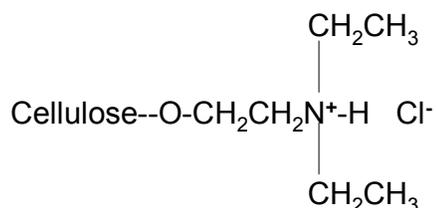


Figure 2.3. The structure of DEAE-cellulose is shown. The cellulose is bonded to quaternary amino group by ether linkage.¹⁹

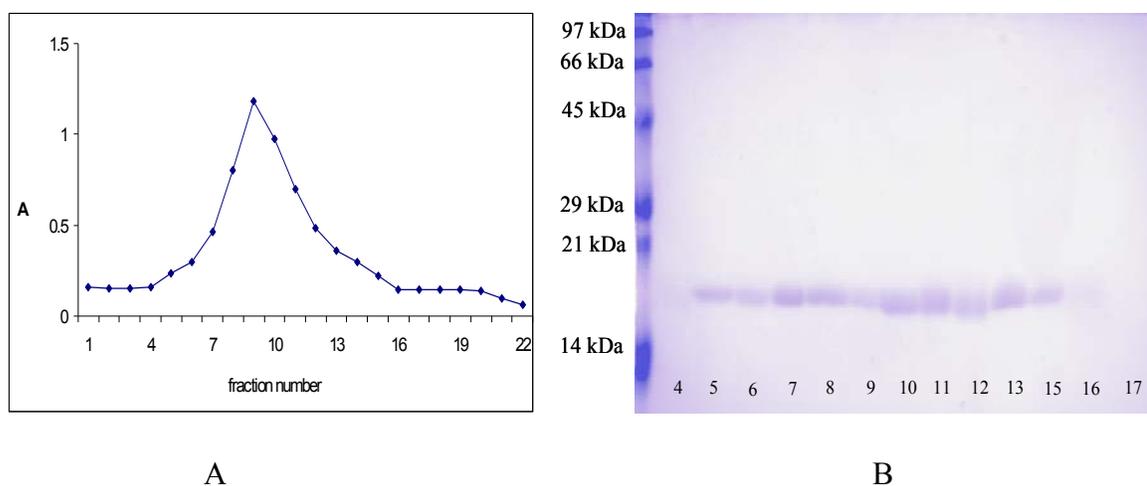


Figure 2.4. The UV-Vis measurement at 280nm for each collected fraction from DEAE-Sephacel column is shown on the left (A) and the SDS-PAGE results from high absorbance fractions are shown on the right (B). The fraction numbers are shown at the bottom of the gel.

Isotope Labeling

Proteins with molecular size up to 20 kDa require double isotope labeling (^{15}N , ^{13}C) since chemical shift information is obtained through a series of triple resonance

experiments where the magnetization is transferred through those nuclei.²² It is important to label proteins uniformly and establish spin ensembles over the entire protein so that heteronuclear spin coupling can effectively be used in triple resonance experiments. Molecules with molecular weight >20 kDa require triple labeling (¹⁵N, ¹³C, ²H) because the increased proton number can inhibit the effective magnetization transfer and behave like a magnetization “sink”.^{7,23,24} NMR spectra of larger proteins are difficult to study due to crowded spectrum with a large number of overlapping signals. Another problem with large proteins is the faster relaxation of transverse magnetization (T_2) as a result of accelerated spin-spin interactions that result in line broadenings and lower sensitivity.²² The transverse relaxation time of a protein is proportional to molecular weight and inversely proportional to the viscosity of the medium.²⁵ The viscosity of the solvent has a profound effect on the rotational motion of proteins and consequently on the relaxation properties. Correlation time (τ_m) is the time required for a molecule to rotate one radian (57°) and is given by the Stock-Einstein equation;

$$\tau_m = \frac{4\pi\eta r_H^3}{3k_B T} \quad [1]$$

where η is the solvent viscosity (cP), k_B is the Boltzmann constant (1.380×10^{-23} J/K), T is the temperature (K), and r_H is the hydrodynamic radius (Å) of the molecule. The viscosity at 3.5 M NaCl with respect to 0 M NaCl is 1.5 times larger (measured as 1.31 and 0.87 cP, respectively). The hydrodynamic radius of the enzyme was calculated by;

$$r_H = \left[\left(\frac{3VM}{4\pi N_A} \right) \right]^{1/3} + r_w \quad [2]$$

where V is the specific volume constant ($0.73 \text{ cm}^3/\text{g}$), M is the molecular weight of the protein (g/mol), N_A is Avogadro's number, and r_w is the hydrodynamic radius of water (1.6 \AA). From this equation the r_H for the *Hv*DHFR1 was approximated as 18.7 \AA . The calculated τ_m for the enzyme at 3.5 M NaCl was found to be 8.9 ns which correlate to a molecular weight of 25.9 kDa at 0 M NaCl . This suggests that the high viscosity (1.31 cP) makes the protein's correlation time more like a 26 kDa protein. This results in drastically attenuated data in the HNCACB and TOCSY type experiments. A general approach to minimize the spin-spin interactions and enhance the sensitivity of these experiments and avoid line broadening due to fast relaxation is the use of ^2H isotope in order to reduce the number of ^1H . In this study, $\sim 70\%$ ^2H isotope labeling was used to obtain interpretable NMR data.

Following the procedure outlined above enabled the production and purification of 34 mg *Hv*DHFR1 of triple labeled (^2H , ^{13}C , and ^{15}N) from 1 L minimal medium. The minimal medium contains $^{13}\text{C}_6$ -glucose as a sole source of ^{13}C , $^{15}\text{NH}_4\text{Cl}$ as a sole source of ^{15}N , and 70% D_2O as a sole source of ^2H in the protein production in bacterial system. After completing isolation and purification of *Hv*DHFR1, the initial globular structural and stability information were obtained from a two dimensional ^1H - ^{15}N HSQC experiment. Each residue (except prolines and the first residue in the sequence) is represented by a peak in the HSQC spectrum that corresponds to amide nitrogen (^{15}N) and the directly attached proton ($^1\text{H}^N$) on both spectral dimensions. The dispersion of peaks across an HSQC spectrum is an indication of a folded protein. Figure 2.5 shows a ^1H - ^{15}N -HSQC spectrum of 2.0 mM *Hv*DHFR1 recorded at 298 K at 500 MHz in 95% $\text{H}_2\text{O}/5\%$ D_2O , 3.5 M NaCl and 10 mM tris buffer at $\text{pH } 7$. The peaks are well spread

over the spectrum, which indicates that the protein is folded at 3.5 M NaCl. From the spectrum, 149 peaks were counted over 154 expected peaks (total 162 residue and 8 prolines). The shortage may be due to peak overlaps from chemical shift degeneracy.

The initial structural data obtained from 2D ^1H - ^{15}N HSQC indicated a folded protein of correct size was isolated and purified from optimized physical and chromatographic techniques. The protein sample prepared at 3.5 M NaCl, 10 mM tris buffer at pH 7, 95% H_2O /5% D_2O then was used to collect a series of 3D triple resonance experiment and 3D NOESY experiments. The 3D triple resonance experiment data will be used for backbone and side chain chemical shift assignments. The assigned backbone and side chain chemical shift values will then be used to NOE cross peaks in 3D NOESY experiments to obtain main geometrical information used in structure calculation.

^1H - ^{15}N HSQC of *Hv*DHFR1 at 3.5 M NaCl

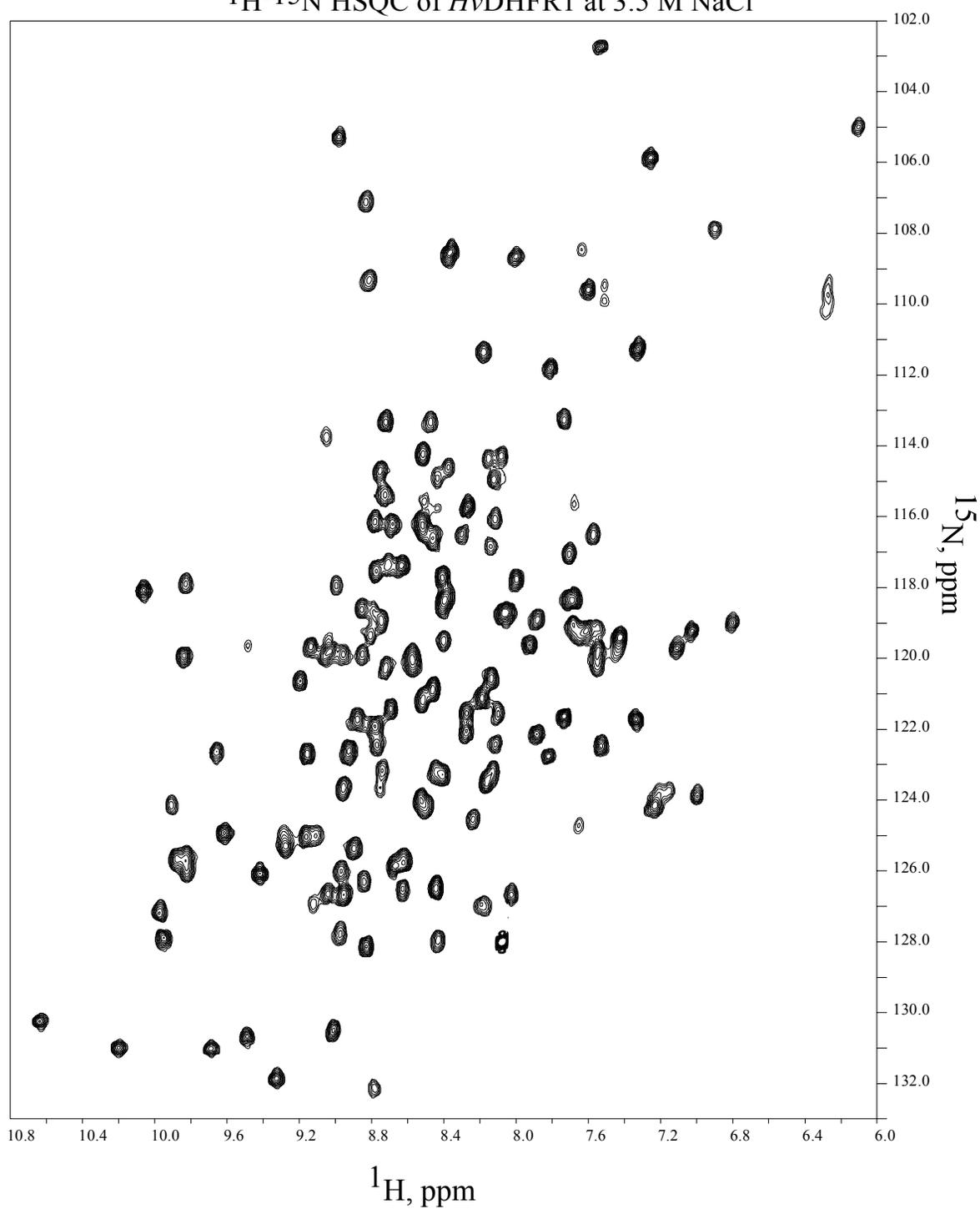


Figure 2.5. ^1H - ^{15}N -HSQC spectrum of 2.0 mM *Hv*DHFR1 at 3.5 M NaCl and 10 mM tris buffer at pH 7 and 25 °C.

REFERENCES

- (1) Wider, G. *Biotechniques* **2000**, *29*, 1278-1282.
- (2) Wagner, G.; Thanabal, V.; Stockman, B. J.; Peng, J. W.; Nirmala, N. R.; Hyberts, S. G.; Goldberg, M. S.; Detlefsen, D. J.; Clubb, R. T.; Adler, M. *Biopolymers* **1992**, *32*, 381-390.
- (3) Wuthrich, K. *Methods. Enzymol.* **1989**, *177*, 125-131.
- (4) Clore, G. M.; Gronenborn, A. M. *NMR of Proteins*; CRC Press: Boca Ratone, FL, 1993.
- (5) Roberts, G. C. K. *NMR of Macromolecules: A practical Approach*; Oxford Univeristy Press: New York, 1993.
- (6) Kainosho, M. *Nat. Struct. Biol.* **1997**, *4 Suppl*, 858-861.
- (7) Gardner, K. H.; Kay, L. E. *Ann. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357-406.
- (8) Gong, P.; Szleifer, I. *J. Colloid. Interface Sci.* **2004**, *278*, 81-90.
- (9) Melander, W. R.; el Rassi, Z.; Horvath, C. *J. Chromatogr.* **1989**, *469*, 3-27.
- (10) Stahlberg, J.; Jonsson, B.; Horvath, C. *Anal. Chem.* **1991**, *63*, 1867-1874.
- (11) Stahlberg, J.; Jonsson, B.; Horvath, C. *Anal. Chem.* **1992**, *64*, 3118-3124.
- (12) Heinitz, M. L.; Kennedy, L.; Kopaciewicz, W.; Regnier, F. E. *J. Chromatogr.* **1988**, *443*, 173-182.
- (13) Hallgren, E.; Kalman, F.; Farnan, D.; Horvath, C.; Stahlberg, J. *J. Chromatogr. A* **2000**, *877*, 13-24.
- (14) Popovici, S. T.; Schoenmakers, P. J. *J. Chromatogr. A* **2005**, *1099*, 92-102.
- (15) Polson, A.; Katz, W. *Biochem. J.* **1969**, *112*, 387-388.

- (16) [http://www1.gelifesciences.com/aptrix/upp00919.nsf/Content/6EEE47990D9F933EC1256F90000DD697/\\$file/18102218AI.pdf](http://www1.gelifesciences.com/aptrix/upp00919.nsf/Content/6EEE47990D9F933EC1256F90000DD697/$file/18102218AI.pdf) (accessed March 2007)
- (17) Eisenberg, H.; Mevarech, M.; Zaccari, G. *Biochemical, Structural, and Molecular Genetic Aspects of Halophilism*. In *Advances in Protein Chemistry*; Academic Press: New York, 1992; pp 1-57.
- (18) Uchiyama, H.; Fujimoto, N.; Sakurai, K.; Nagasawa, K. *J. Chromatogr.* **1984**, *287*, 55-66.
- (19) [http://www1.gelifesciences.com/aptrix/upp00919.nsf/Content/7185770671857706466D1AB8C1256EB40041805D/\\$file/11000421AA.pdf](http://www1.gelifesciences.com/aptrix/upp00919.nsf/Content/7185770671857706466D1AB8C1256EB40041805D/$file/11000421AA.pdf) (accessed March 2007).
- (20) Dorsey, J. G.; Cooper, W. T.; Wheeler, J. F.; Barth, H. G.; Foley, J. P. *Anal. Chem.* **1994**, *66*, 500R-546R.
- (21) Houwing, J.; van Hateren, S. H.; Billiet, H. A.; van der Wielen, L. A. *J. Chromatogr. A* **2002**, *952*, 85-98.
- (22) Foster, M. P.; McElroy, C. A.; Amero, C. D. *Biochemistry* **2007**, *46*, 331-340.
- (23) Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Guntert, P. *Nature* **2006**, *440*, 52-57.
- (24) Venters, R. A.; Huang, C. C.; Farmer, B. T., 2nd; Trolard, R.; Spicer, L. D.; Fierke, C. A. *J. Biomol. NMR* **1995**, *5*, 339-344.
- (25) Bax, A.; Grzesiek, S. *Acc. Chem. Res.* **1993**, *26*, 131-138.

CHAPTER III
CHEMICAL SHIFT ASSIGNMENTS OF THE ENZYME *Hv*DHFR1

Introduction

Chemical shift assignments play a key role in structural determinations since they are affected by the local chemical environment. Three dimensional NMR experiments simplify the spectrum of proteins with larger molecular weights where a large number of resonances can not easily be resolved by 1D or 2D experiments.^{1,2} ¹⁵N and ¹³C isotope labeling assist in the interpretation of NMR data for larger proteins by establishing correlations between homonuclear and heteronuclear spin systems.³⁻⁷ Labeling also provides a means of dispersing information since heteronuclei have a larger chemical shift range than proton.

A number of heteronuclear 3D NMR experiments have been developed for sequence specific backbone and side chain chemical shift assignments of amino acids.^{8,9} Most of these experiments utilize the three nuclei: ¹H, ¹⁵N, and ¹³C, therefore termed triple resonance experiments.^{10,11} These experiments typically employ one and two bond spin couplings to transfer magnetization via inter- and intraresidue correlations.¹² Complete backbone and side chain assignments require analyzing the experimental data from each experiment in a similar manner, very much like solving a puzzle. The

following section will illustrate the chemical shift assignments of *Hv*DHFR1 through a series of triple resonance experiments acquired in 3.5 M NaCl.

The acquisition of NMR data on high salt samples has always been problematic. Since the high salt changes the impedance of the coils, even relatively low concentrations of salt causes problems in tuning and matching of the probe. This in turn causes pulse widths to become very large, which makes collecting three-dimensional data virtually impossible.

NMR of a High Salt Protein Sample

Cosolute conditions can have a drastic effect on the quality of NMR data. Extreme salt concentrations have a detrimental effect on NMR spectrometer hardware, specifically on tuning and matching of probes, the relaxation properties of proteins and also magnetization transfer within a spin system.¹³ An NMR sample with a highly conductive buffer inductively couples to the coils of a probe which can dramatically affect the tuning and matching of the probe to the correct frequencies since the high salt buffer changes the impedance of the coils.¹⁴ The conductivity of the solvent due to high ion mobility causes the applied radio frequency (*rf*) pulse to develop an electric current and causes the sample to couple with the receiver coil.^{15,16} The enzyme *Hv*DHFR1 in this study requires 3.5 M NaCl for optimal activity. Using a regular 5 mm NMR tube with extremely high salt does not allow tuning and matching to the correct frequencies; therefore, *rf* pulses are >5 times longer which can create Joule heating.¹⁷ The electrical field coupling between the receiver coil and sample in a high ionic strength solvent was minimized by using an NMR reduction sample tube (New Era Enterprises Inc.) which has a 5.0 mm upper

section and a 3.0 mm lower section where the sample resides, Figure 3.1. This alternative tube geometry provided excellent tuning and matching of the probes as well as *rf* pulse lengths similar to low salt samples. This also eliminated any heating problems in the spin-lock experiments.¹³ The reduction in the cross-section of the NMR sample results in a low signal/noise (S/N). The fill factor generated by volume reduction is compensated by increasing the protein concentration to 2.0 mM.

Another difficulty which arises from a high salt sample is magnetization transfer, especially in experiments that utilize spin-lock pulses such as TOCSY¹⁸ and DIPSY.¹⁹ Experiments such as (H)C(C-CO)NH-TOCSY,²⁰ H(CC-CO)NH-TOCSY^{21,22} and HCCH-TOCSY²³ are heavily dependent on the magnetization transfer throughout a spin system which is used to obtain complete side chain assignments. For residues with large spin systems like I, R, and K, cross-peaks representing complete magnetization transfer

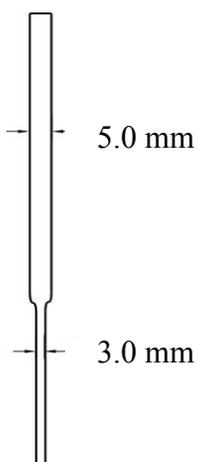


Figure 3.1. The NMR tube used for high salt protein sample is shown. The tube geometry afforded acquisition of interpretable NMR data. The fill factor generated by volume reduction was compensated by increasing the protein concentration.

were rarely observed in the spectra. However, one and two step transfers were still detected which allowed for complete side chain assignments, see Figure 3.19.

Three Dimensional NMR Experiments

The fundamental principle of multidimensional NMR experiments relies on manipulating homonuclear and/or heteronuclear spin interactions to move magnetization from one nucleus to another and encode multiple frequencies onto one FID. To acquire a 3D spectrum, a three dimensional data matrix is obtained through incremented time delays embedded into a pulse program. Figure 3.2 illustrates the HNCA pulse program. The specifics of HNCA will be discussed later. The concept of multidimensional NMR will be discussed using this pulse program.²⁴

It is the combination of pulses and delays that allows for the acquisition of a specific set of data. This combination changes from one 3D pulse program to another and will not be discussed here in detail. What will be illustrated is the concept of multidimensional NMR; in other words, how raw data are acquired and converted over to a 3D cube. In this example the ^1H frequency (f_3) is obtained by direct detection. The indirect frequencies, ^{15}N and ^{13}C , are labeled onto the proton frequency by the use of the incremented time delays, t_1 and t_2 respectively. Since there are two incremented time delays this requires the acquisition of many blocks of data. A single block of data represents the FID with these delays set to specific values for a certain number of scans. Since there is a t_1 and t_2 incremented delay for ^{15}N and ^{13}C , discrete blocks must be acquired for each dimension. For example, a typical 3D data set contains 64 blocks in the ^{15}N dimension and 128 blocks in the ^{13}C dimension. More blocks are needed in the ^{13}C

dimension due to its large chemical shift range. The size of the resulting matrix is 1024 x 64 x 128. If each scans takes 1 second this data will take a total of 18 hours to acquire. The incrementation of t1 and t2 must be performed in a specific manner. In the first block of data these delays are set to specific values determined by the sweep width of the indirect dimensions. In this matrix, t1 is incremented by a specific value in the first 64 blocks of data while t2 is kept constant. In the 65th block of data, t2 is incremented by one step and another 64 blocks of data are acquired by incrementing t1. This is continued until 8192 blocks of time domain data are acquired. These blocks of time domain data are converted over to frequency domain data through successive Fourier transforms. The first Fourier transform converts the direct detected dimension (t3) to the proton frequency

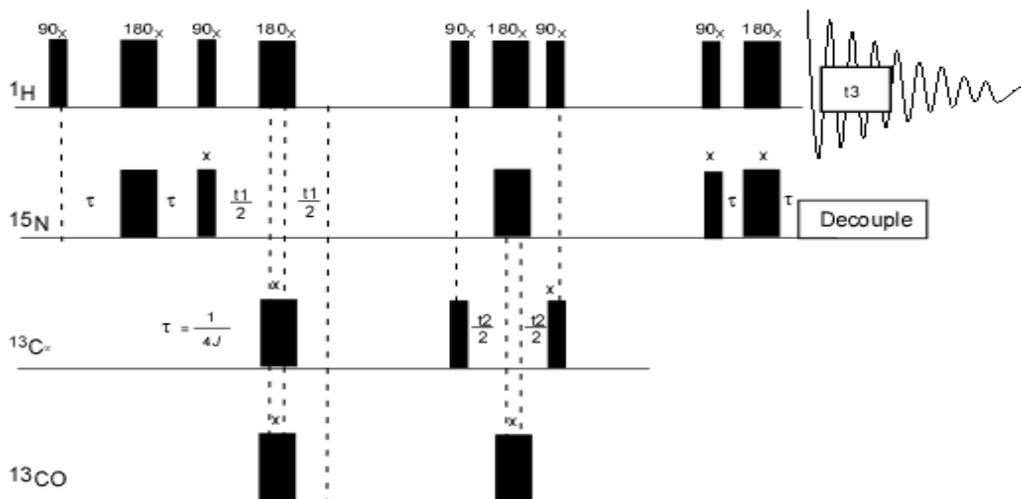


Figure 3.2. Schematic pulse diagram illustrations of HNCA experiment. The timing of the pulses starts from the left where the thickness of vertical bars (90° and 180°) indicate pulses width (duration) on the related nuclei frequency. The subscripts on the pulse are used to denote the applied pulse axes and the pulse delay are denoted as tau, τ .²⁴

(f1) for each block of data resulting in the interferograms shown in Figure 3.3. There should be 128 columns of data with each column containing 32 proton spectra. The process of a Fourier transform converts half the data to imaginary spectra which are used for phasing and then discarded. The incrementation of t1 causes the resonances in a single column (block) to vary sinusoidally due to the indirect frequencies. A new set of FIDs are created from each interferogram by transposing the data along discrete points in f1. These FIDs are then Fourier transformed to produce two dimensional planes (f1 and f2) where the signals again vary sinusoidally as shown in Figure 3.4. These data are transposed to a new set of FIDs by taking discrete points across the plane. A third Fourier transform gives the final cube of 3D data as shown in Figure 3.5 that contains three frequencies (f1, f2, f3) axes. The x-axis is ^1H while ^{15}N is usually placed along the z-axis and ^{13}C along the y-axis. It is impossible to interpret data directly from the cube therefore it is cut into discrete planes.

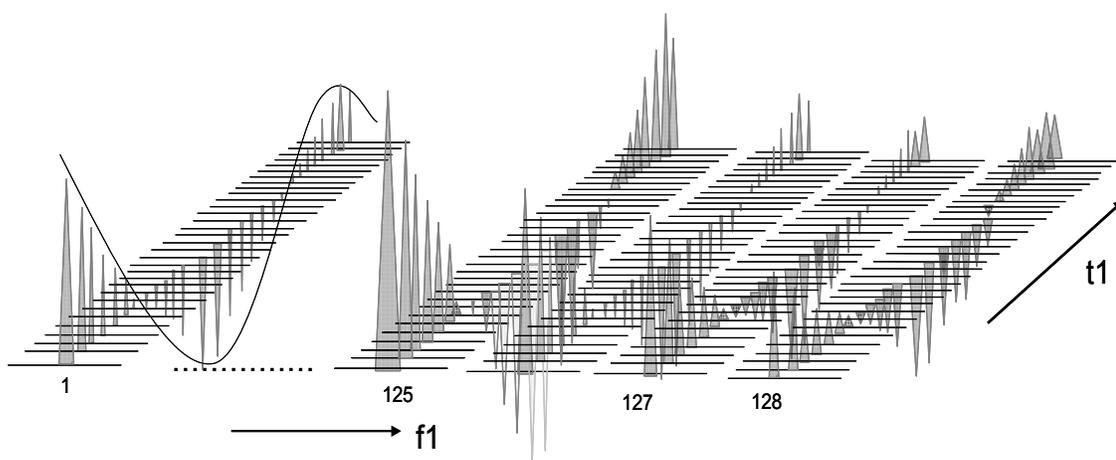


Figure 3.3. Fourier transform on the second dimension results in a FID form sinusoidal function in the t1 dimension.

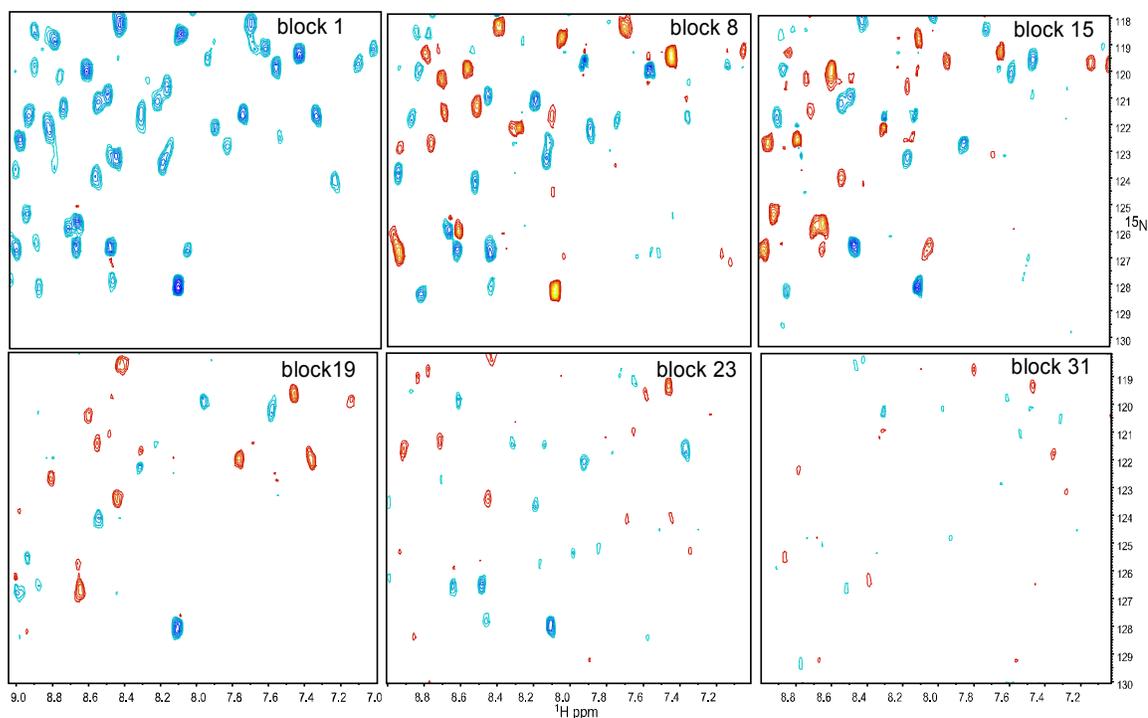


Figure 3.4. The results of two dimensional Fourier transformations of the first two data blocks. The result is a series of two dimensional spectrum where the peak intensity for the same spot vary sinusoidally. The blue and red indicates positive and negative signals, respectively.

Each plane reveals specific data that are determined by the particular 3D pulse program.

A single strip of data from the plane represents the correlations from a single amino acid.

The triple resonance experiments used for backbone and side chain assignments rely heavily on magnetization transfer among spin systems of inter- and intraresidues via through-bond coupling.⁶ The polarization of bonding electrons facilitates the spin-spin interaction through nuclei which is also known as scalar coupling (or *J*-coupling). A uniformly ¹⁵N and ¹³C isotope-enriched protein provides spin ensembles (same multiplicity) among the nuclei of a protein which enables the use of the heteronuclear spin coupling effectively across the peptide backbone and side chains.^{9,10} This also

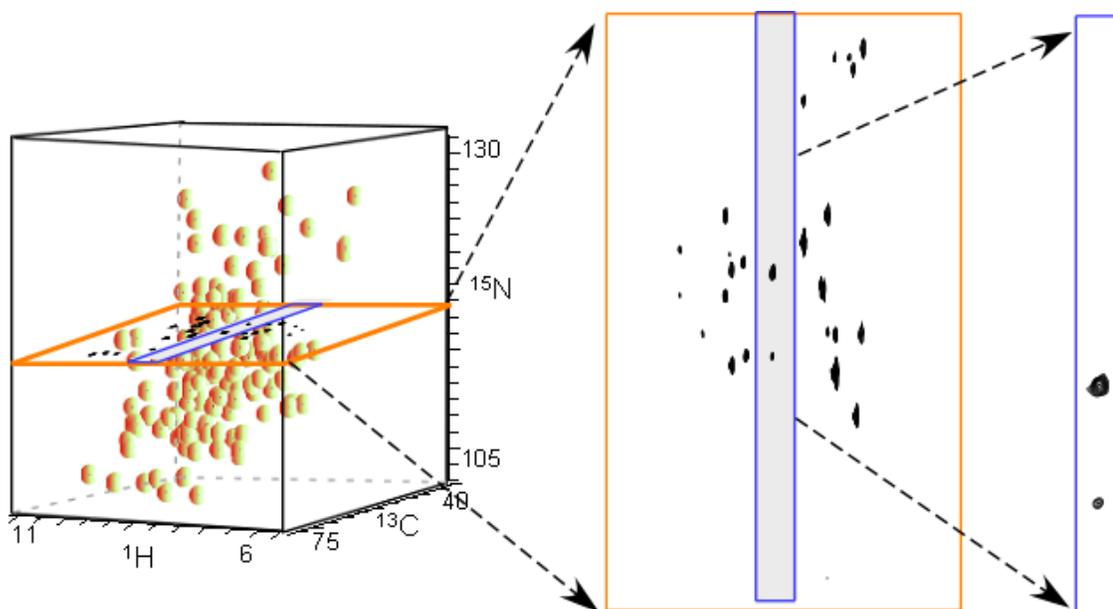


Figure 3.5. A 3D NMR data, a cross section from a 3D NMR experiment, and a strip from the cross section are illustrated. The NMR data are analyzed by cross sectioning the cube to planes at one of the largest frequency axis (typically ^{15}N) and planes are further analyzed by strips at each recorded ^1H chemical shift values.

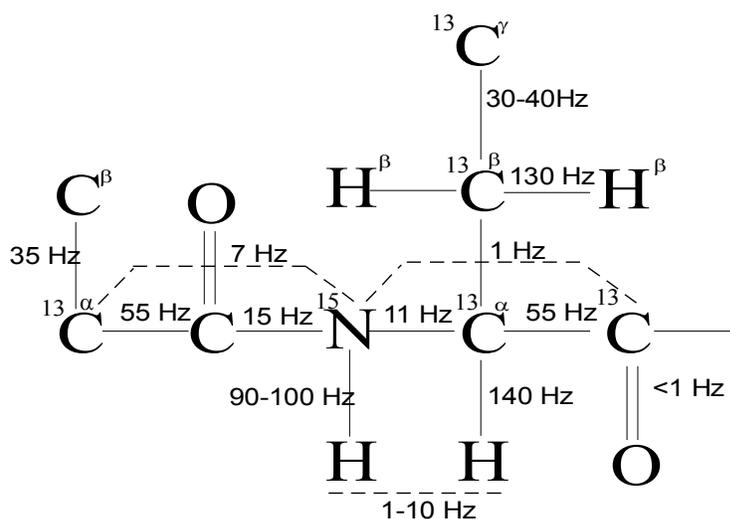


Figure 3.6. The one-bond (1J) and two-bond (2J) scalar coupling constants across the peptide backbone for a ^{15}N and ^{13}C labeled protein is shown. Correlations among homonuclear and heteronuclear spins are established through these couplings in triple resonance NMR experiments.¹⁰

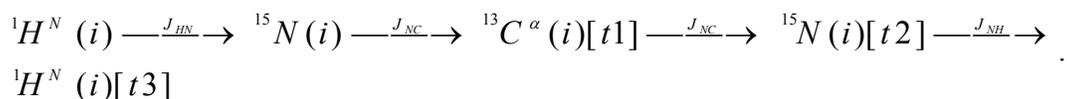
provides a means to selectively filter the magnetization transfer among certain nuclei in triple resonance NMR experiments. Figure 3.6 shows the typical size of 1J and 2J couplings across the peptide backbone and side chains.²⁵⁻²⁹ Magnetization transfer (i.e coherence transfer) among different nuclei is accomplished via applying rf pulses to specifically excite the frequency range of the spectrum. The magnetization of a set of nuclei can be transmitted to another set of nuclei through a delay (usually $1/2J$ or $1/4J$) - 180° pulse-delay, in other words magnetization is allowed to dephase during a delay and then is refocused onto another nuclei through the second delay.^{30,31}

A bank of heteronuclear triple resonance experiments has been developed for sequence specific backbone and side chain chemical shift assignments of proteins. These experiments use through-bond coupling to establish correlations among ^{15}N , ^{13}C , and ^1H nuclei for backbone and side chain assignments.³²⁻³⁵ The 3D HNCA,³⁶⁻³⁹ HN(CO)CA,³⁶⁻³⁹ HNCACB,⁴⁰⁻⁴² CBCANH,^{40,41} HNCO,^{36,39,41} HN(CA)CO,^{39,43-45} and HN(CA)HA^{46,47} experiments are used for backbone chemical shift assignments, each of which provides chemical shift information for a certain nucleus. The 3D H(CCO)NH-TOCSY,^{48,49} (H)C(CO)NH-TOCSY,⁴⁸ and HCCH-TOSCY²³ experiments are employed for side chain chemical shift assignments. Initially, during the course of assignment the backbone atoms (^1H , $^{15}\text{N}^{\text{H}}$, $^{13}\text{C}^\alpha$, C' (carbonyl carbon)) chemical shifts would be assigned and then these data would be used for side chain chemical assignments. The lists of 3D NMR experiments used for backbone chemical shift assignments in this study are given in Figure 3.7. The nomenclature of the 3D NMR experiment is informative in terms of magnetization transfer pathways and the nuclei involved in the experiment. The nucleus within a bracket indicates only used for magnetization transfer but frequencies are not

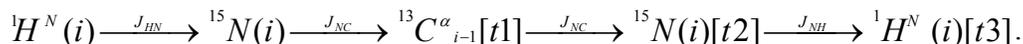
evolved and are not recorded. The pathways of magnetization among nuclei for inter- and intraresidues are illustrated by arrows in Figure 3.7. The nucleus that the frequency is recorded is shown either in blue (positive) or red (negative) circles. The nucleus that is used only for magnetization transfer is illustrated by square boxes.

Backbone Chemical Shift Assignments

A 3D HNCA experiment provides alpha carbon ($^{13}\text{C}^\alpha$) chemical shift information for inter- and intraresidues to be used in sequence specific backbone assignments. The mechanism of the experiment starts by exciting all protons and transferring magnetization through the backbone atoms in the order of:



In this experiment, the ${}^1\text{H}$ and ${}^{15}\text{N}$ chemical shifts of a residue correlate to ${}^{13}\text{C}^\alpha_i$ and ${}^{13}\text{C}^\alpha_{i-1}$ chemical shifts where the ${}^{13}\text{C}^\alpha_i$ chemical shift is observed as a more intense peak in the spectrum. However, since the one-bond coupling constant of ${}^{15}\text{N}$ - ${}^{13}\text{C}^\alpha_i$ (${}^1J_{\text{NC}} \sim 11$ Hz) is very close to the two-bond coupling constant of previous residue (${}^2J_{\text{NC}} \sim 7$ Hz), the magnetization transfer to the previous residue ${}^{13}\text{C}^\alpha_{i-1}$ will also take places in a similar fashion in the order of :



This correlation is observed as a less intense peak in the spectrum as the experiment conditions are optimized for the one-bond coupling of ${}^{13}\text{C}$ - ${}^{15}\text{N}$. Triple resonance experiments like this are referred to as ‘out and back’ because the magnetization is returned to ${}^1\text{H}$ for acquisition.

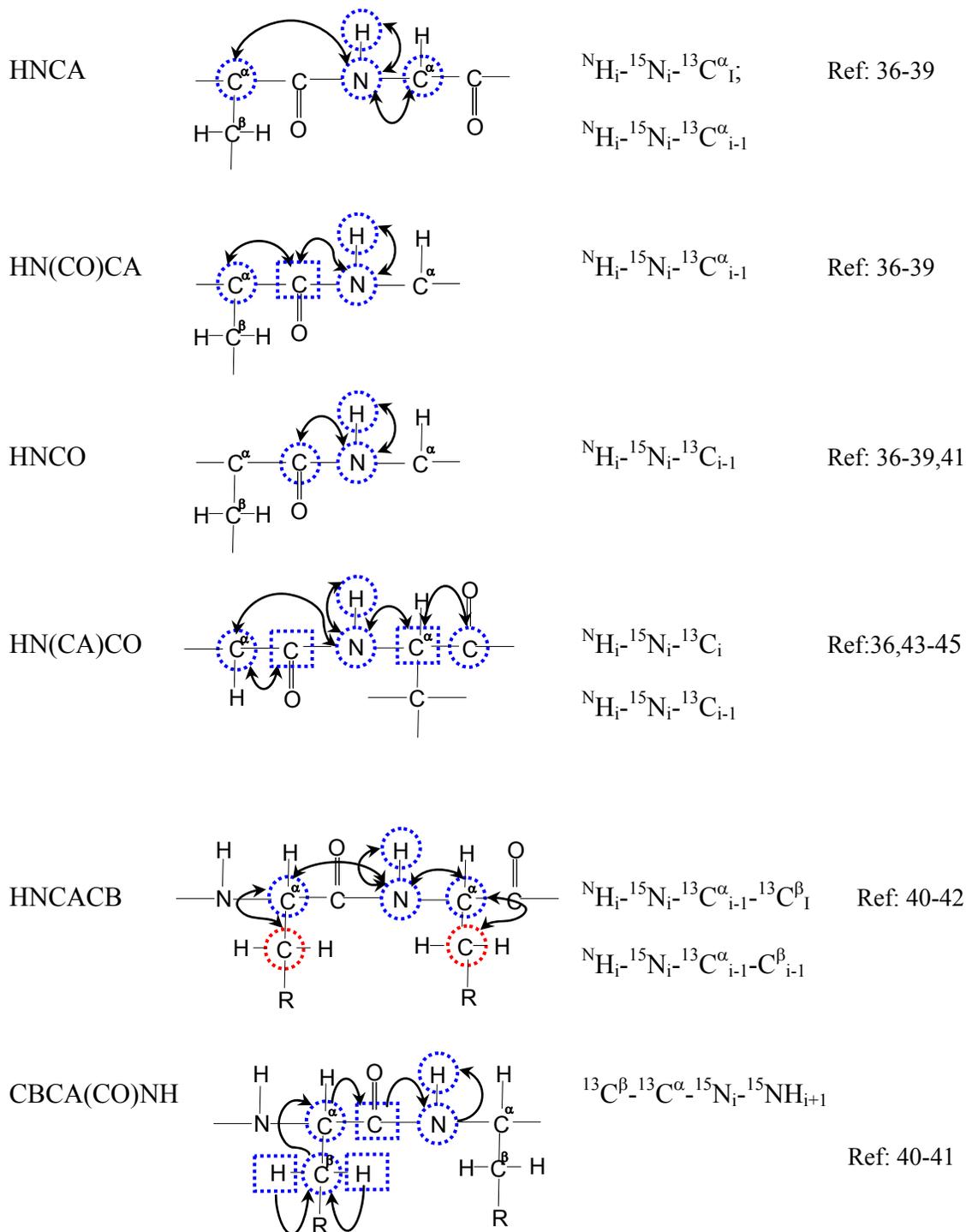
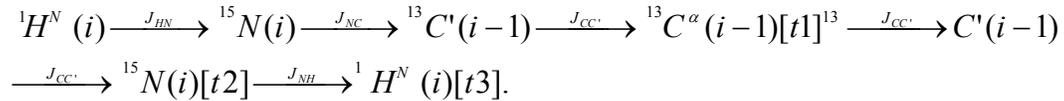


Figure 3.7. The list of three dimensional triple resonance experiments for backbone chemical shift assignments is shown.

The 3D HN(CO)CA experiment is a complementary experiment to the HNCA. Ambiguities in assignments due to the degeneracy of the $^{13}\text{C}^\alpha$ chemical shifts can be resolved by the 3D HN(CO)CA experiment. This experiment provides only the $^{13}\text{C}^\alpha$ chemical shift of the previous residue through correlating the $^1\text{H}^N$, $^{15}\text{N}^H$, and $^{13}\text{C}^\alpha_{i-1}$



The $^{13}\text{C}'$ nucleus in this experiment is utilized to transfer the magnetization from ^{15}N nucleus to $^{13}\text{C}^\alpha_{i-1}$ nucleus, but the $^{13}\text{C}'$ frequencies are not evolved and are not detected.

Figure 3.8 illustrates two strips, one from the HNCA and the complementary stripe from the HN(CO)CA. In the left strip (HNCA), the ^{15}N (127.86 ppm) and $^1\text{H}^N$ (9.0 ppm) of a residue correlate to two carbon chemical shifts at 55.3 ppm and 60.9 ppm. The strong $^{13}\text{C}^\alpha$ peak at 55.3 ppm was assigned to the $^{13}\text{C}^\alpha$ of the residue i while the less intense peak at 60.90 ppm was assigned as $^{13}\text{C}^\alpha$ of the residues $i-1$. The same $^1\text{H}^N$ and $^{15}\text{N}^H$ chemical shifts in the HN(CO)CA experiment correlate only to one peak which is the less intense peak observed in the HNCA experiment. The same type of interaction is obtained for every ^1H - ^{15}N pair in the protein.

Figure 3.9 illustrates the backward and forward $^{13}\text{C}^\alpha$ chemical shift tracking strategy in amino acid sequence format. The $^{13}\text{C}^\alpha$ chemical shift information from strips is shown on peptide backbone nuclei which demonstrates how the $^{13}\text{C}^\alpha$ correlation can be established across the peptide backbone via HNCA/HN(CO)CA experiments. It is illustrated that $^1\text{H}^N/^{15}\text{N}^H$ of a residue correlates to both its own as well as to the shifts of

$^{15}\text{N}^{\text{H}} = 127.86 \text{ ppm}$

$^1\text{H}^{\text{N}} = 9.00 \text{ ppm}$

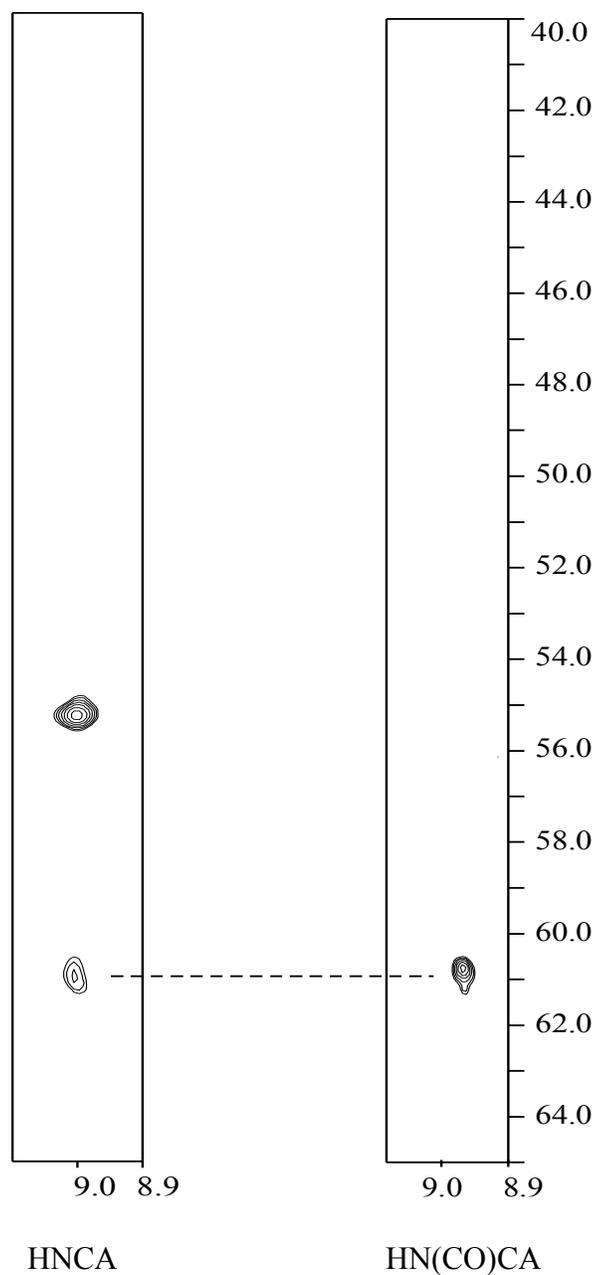


Figure 3.8. Strips [f1(^1H)-f3(^{13}C)] were taken from ^{15}N plane (127.60 ppm) of HNCA (left) and HN(CO)CA (right) experiments. In the strip from HNCA, the more intense peak was assigned as residue's own $^{13}\text{C}^{\alpha}$ while the less intense one was assigned as $^{13}\text{C}^{\alpha}_{i-1}$ chemical shift values. The same $^1\text{H}/^{15}\text{N}$ chemical shift pair in HN(CO)CA experiment correlates to only $^{13}\text{C}^{\alpha}_{i-1}$ (right strip) chemical shift value.

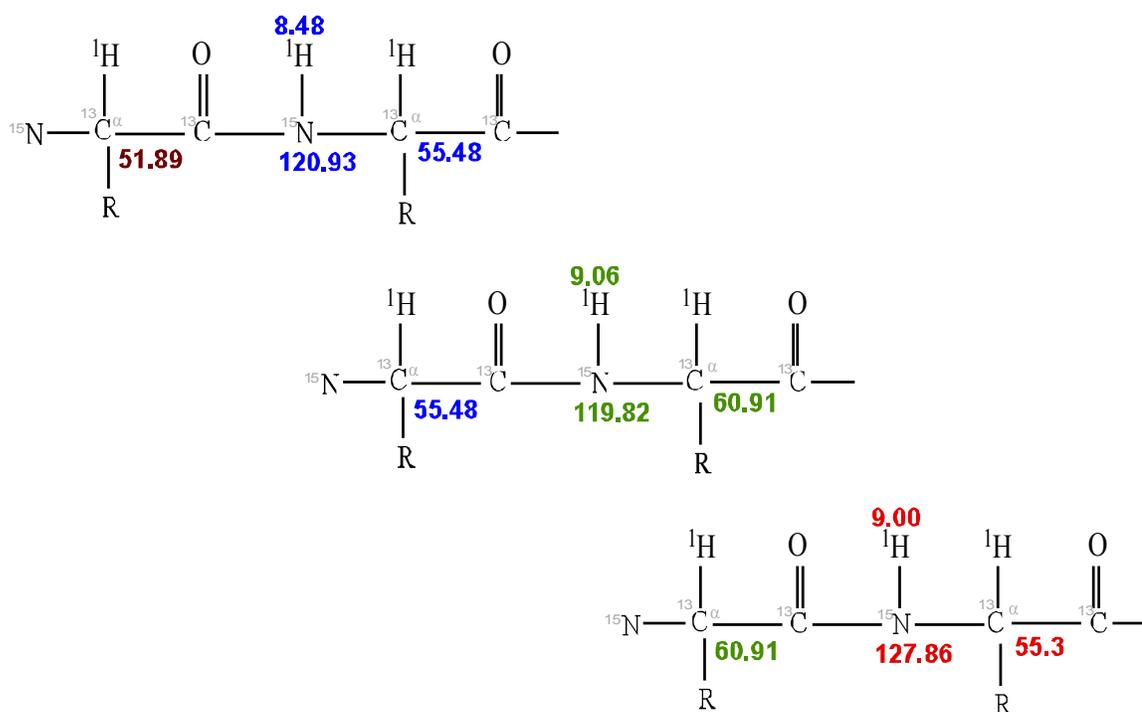


Figure 3.9. The figure illustrates the methodology followed to establish the $^{13}\text{C}^{\alpha}$ chemical shift correlations among the residues via the HNCA/HN(CO)CA experiments.

the previous residues $^{13}\text{C}^{\alpha}$ chemical shift. In turn, each residue would correlate to the previous residue such that the chemical shift correlations across the peptide backbone can be established.

Knowing the sequence of a protein aids in assigning the chemical shift values for an individual amino acid, which is termed sequence specific chemical shift assignments.^{34,47,50-56} The glycine residue is always a good starting point in sequence specific $^{13}\text{C}^{\alpha}$ chemical assignment since it has a unique $^{13}\text{C}^{\alpha}$ chemical shift value at ~ 45 ppm. The protein *Hv*DHFR1 contains 9 glycines where two consecutive glycines were chosen as the first starting point in sequential assignments (I100-G101-G102-A103).

Two correlating $^{13}\text{C}^{\alpha}$ glycines peaks were easily found on a ^{15}N plane. By backward and

forward correlating, a part of the amino acid sequence can be assigned to a specific $^1\text{H}^{\text{N}}$, ^{15}N , and $^{13}\text{C}^{\alpha}$ chemical shifts. This process is continued until there was either a break in the correlation or until too many $^{13}\text{C}^{\alpha}$ chemical shifts overlap. The magnetization transfer along the backbone atoms are interrupted at proline residues that do not have an amide proton. In this case, another starting point must be identified to continue sequence-specific assignments. One problem does arise from just using the HNCA/HN(CO)CA in which too many residues in a row is the degenerated chemical shift values. Therefore, only ~ 70% of the protein can be sequence-specifically assigned, additional chemical shift information must be obtained to clarify ambiguous regions. Figure 3.10 shows the sequential connectivity through T145-F150 residues of *HvDHFR1* obtained via the HNCA/HN(CO)CA experiments. The strong peak in each strip correlates to the residue's own $^{13}\text{C}^{\alpha}$ and the less intense peak correlates to the $^{13}\text{C}^{\alpha}_{i-1}$ chemical shift of the previous residues. The less intense peak in each strip becomes the more intense peak in the preceding strip. However, the complete sequential connectivity across the peptide backbone can not be easily established via these two experiments.

There is a need for more data to resolve any ambiguity or uncertainty in the chemical shift assignments. This can be achieved by extending the backbone assignments to the $^{13}\text{C}^{\beta}$ nuclei. The $^{13}\text{C}^{\beta}$ chemical shift values of amino acids have a broader chemical shift range (~10-80 ppm) and also allow for the identification of some residues from their specific patterns. Residues such as alanine, threonine, and serine have unique, distinguishable $^{13}\text{C}^{\beta}$ chemical shifts. Glycine can be easily distinguished from the rest of the amino acids because it does not possess a $^{13}\text{C}^{\beta}$ nucleus. The $^{13}\text{C}^{\beta}$ chemical

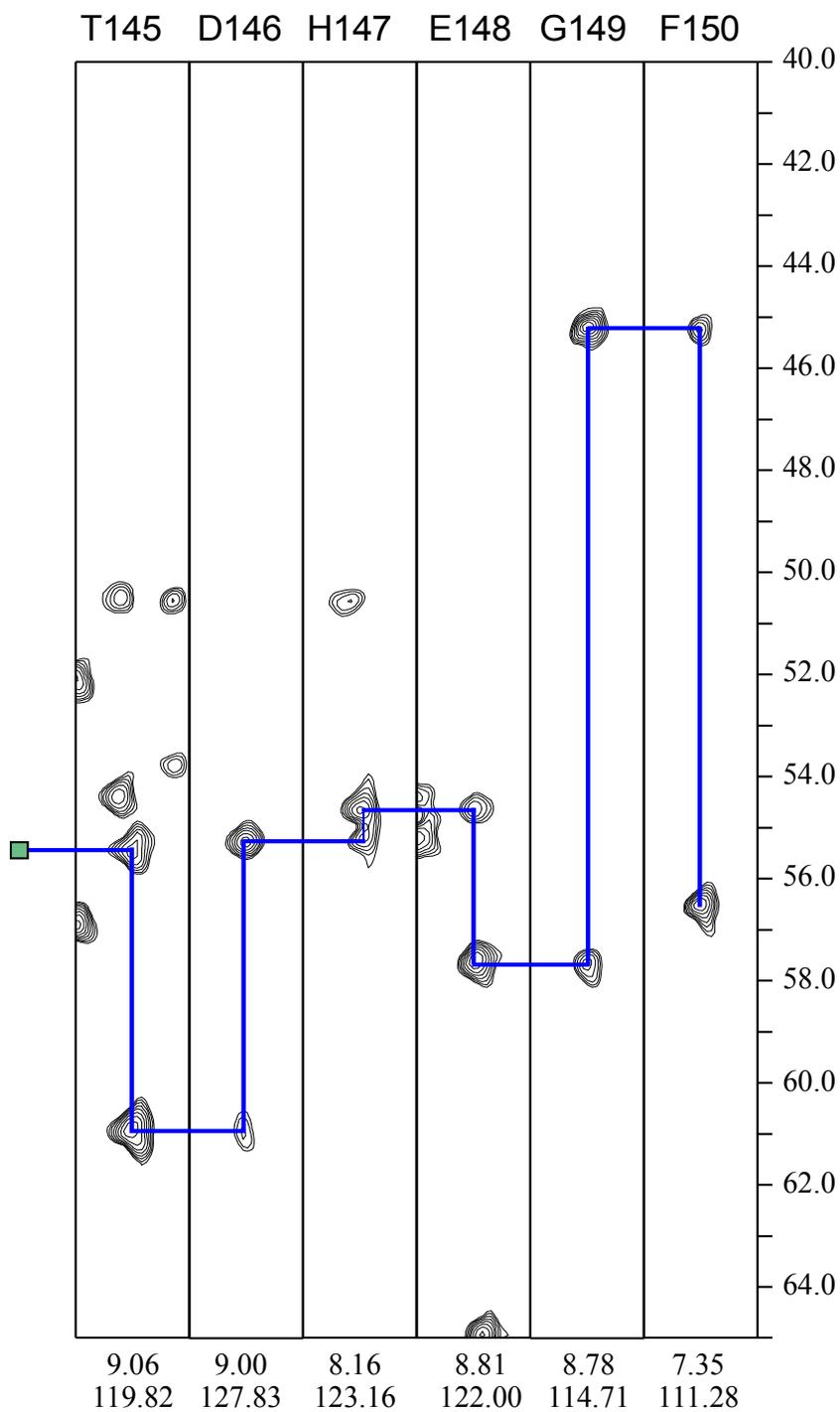
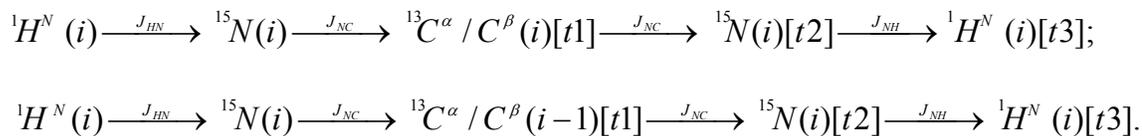


Figure 3.10. Strips (f1-f3) from $^1\text{H}/^{15}\text{N}$ chemical shifts of the residues T145-F150 of 3D HNCA experiment is shown. The strips were obtained from different ^{15}N planes of HNCA experiments. The solid blue lines connect the intrer-residual $^{13}\text{C}^\alpha$ of residue i with the corresponding $^{13}\text{C}^\alpha$ of residue $i-1$.

shift value of alanine is at ~15 ppm (21 alanine residue in *HvDHFR1*), threonine at ~63 ppm (7 threonine residue in *HvDHFR1*), and serine at ~75 ppm (15 serine residue in *HvDHFR1*) help to continue chemical shift assignments. The $^{13}\text{C}^\beta$ chemical shift values were obtained from HNCACB and CBCA(CO)NH experiments.

The 3D HNCACB experiment provides $^{13}\text{C}^\alpha/^{13}\text{C}^\beta$ chemical shift information for both residue *i* and residue *i-1* through correlating $^1\text{H}^N$, $^{15}\text{N}^H$, and $^{13}\text{C}^\alpha/^{13}\text{C}^\beta$ nuclei. The ^1H and ^{15}N of a residue in this experiment is correlated to both its own and previous residue's $^{13}\text{C}^\alpha/^{13}\text{C}^\beta$ chemical shifts. The magnetization transfer pathway in this experiment is as follows:



The relatively large coupling constant of $^1J_{\text{NC}i} \sim 11$ Hz with respect to the $^1J_{\text{NC}i-1} \sim 7$ Hz results in a stronger correlation for residue own $^{13}\text{C}^\alpha$. This would result in more intense peaks for the residue's own $^{13}\text{C}^\alpha/^{13}\text{C}^\beta$ peaks. The sign of $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ peaks are opposite where typically $^{13}\text{C}^\alpha$ peaks are phased positively and illustrated in black while C^β peaks are negative and illustrated in red.

The CBCA(CO)NH experiment is a complementary experiment to the HNCACB, homologous to the HNCA and the HN(CO)CA pair. The experiment assists to resolve ambiguities in the HNCACB experiment. The $^1\text{H}^N$ and $^{15}\text{N}^H$ chemical shift values of a residue correlate to only the $^{13}\text{C}^\alpha/^{13}\text{C}^\beta$ chemical shifts of the previous residue. The magnetization transfer mechanism is slightly different from the previous experiment. The magnetization of side chain H^β protons is filtered down to the directly bounded $^{13}\text{C}^\beta$

$^{15}\text{N}^{\text{H}} = 113.22 \text{ ppm}$
 $^1\text{H}^{\text{N}} = 8.74 \text{ ppm}$

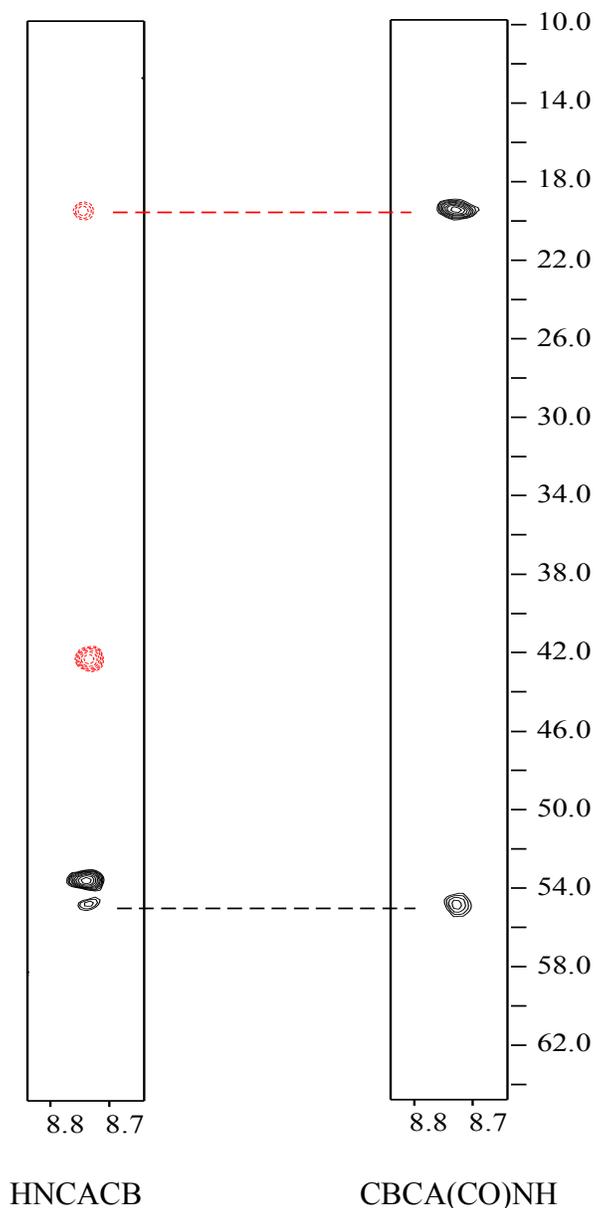


Figure 3.11. Two strips (f1,f3) from the HNCACB and CBCA(CO)HN experiments are shown. The residue correlates both its own $\text{C}^{\alpha}/\text{C}^{\beta}$ chemical shifts at 53.78/42.49 ppm (intense positive and negative peaks) and previous residue's $\text{C}^{\alpha}/\text{C}^{\beta}$ at 54.95/19.67 ppm at the HNCACB. The residue only correlates to $\text{C}^{\alpha}/\text{C}^{\beta}$ of previous residues in the CBCA(CO)NH. The correlations of nuclei between the two experiments are shown by dashed lines.

via $^1J_{\text{HC}}$ coupling (~ 130 Hz) and then transferred to $^{13}\text{C}^\alpha$ where carbon frequencies are recorded as the first dimension. The magnetization transfer pathway in this experiment is as follows:

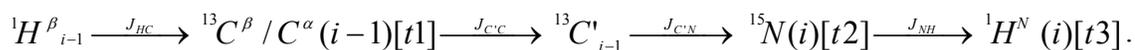


Figure 3.11 shows two strips from the HNCACB and CBCAHN experiments. The intense positive and negative peaks in the HNCACB experiment correlate to a residue's own $\text{C}^\alpha/\text{C}^\beta$ chemical shifts at 53.78/42.49 ppm; the less intense positive and negative peaks correlate to the previous residue's $^{13}\text{C}^\alpha/^{13}\text{C}^\beta$ chemical shifts at 54.95/19.67 ppm. The same $^1\text{H}^\text{N}$ and $^{15}\text{N}^\text{H}$ chemical shifts at CACB(CO)HN experiment correlate only to the $^{13}\text{C}^\alpha/^{13}\text{C}^\beta$ chemical shifts of the previous residue. The chemical shift patterns of the previous residue indicate that the residue is alanine since it is the only residue with $^{13}\text{C}^\beta$ at ~ 19 ppm. The sequential correlation obtained for $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift values across L3 through A8 residues via the HNCACB/CBCA(CO)NH experiments are shown in Figure 3.12. The near complete chemical shift assignments of $^1\text{H}^\text{N}$, ^{15}N , $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ were accomplished by means of four triple resonance experiments: HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH. By using this ^1H , ^{15}N chemical shift information, the individual peaks in the ^1H - ^{15}N HSQC spectrum can be identified and assigned to a specific residue, see Figure 3.13.

The 3D experiment HNHA is typically used for the backbone $^1\text{H}^\alpha$ chemical shift assignments.⁴⁶ For larger size proteins (>20 kDa), the sensitivity of this experiment is very low due to very small $^3J_{\text{HH}}$ coupling constants between $^1\text{H}^\text{N}$ and $^1\text{H}^\alpha$ (2-8 Hz) which results in a loss of magnetization due to the large delays needed for refocusing. As a

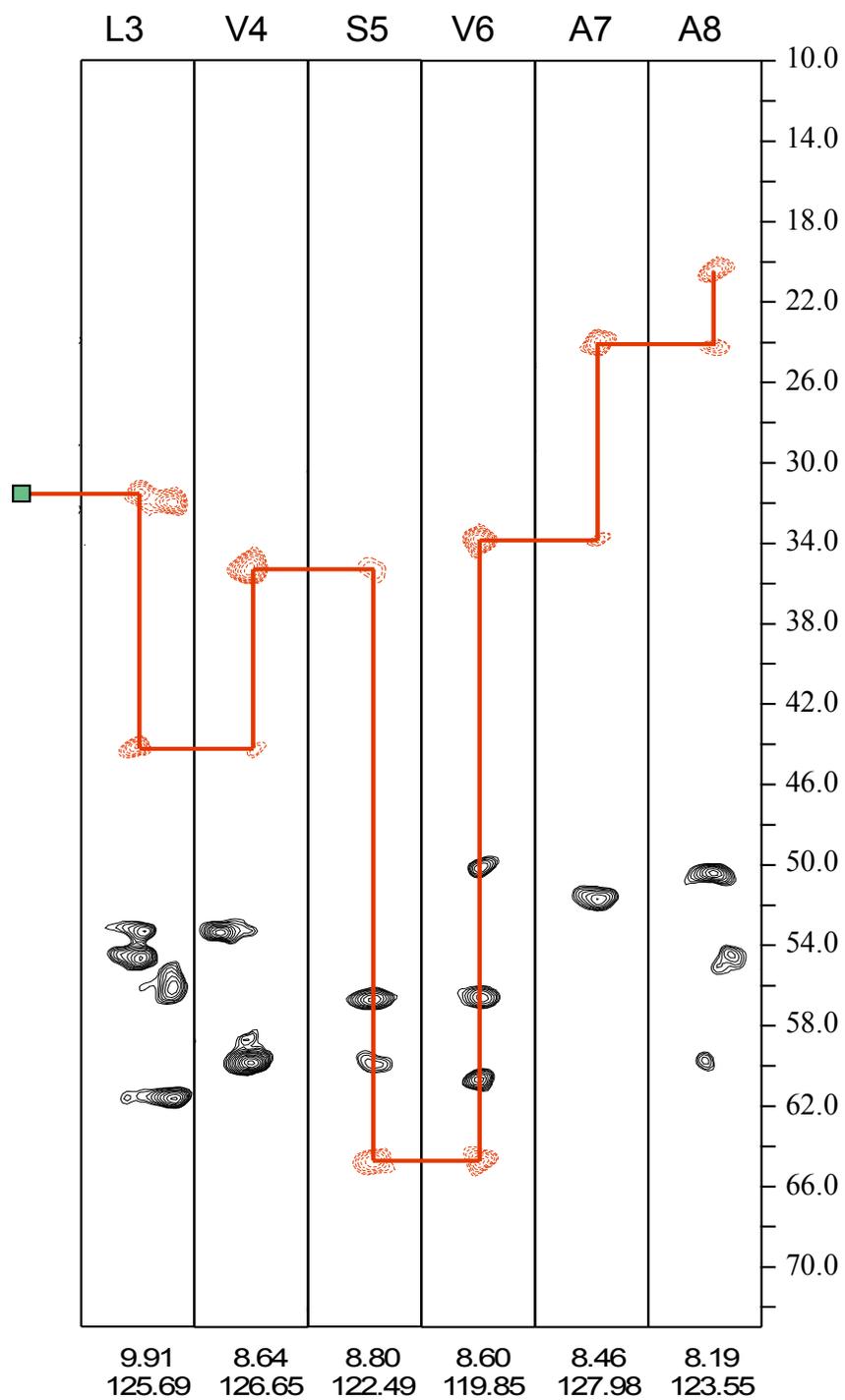


Figure 3.12. Strips (f1-f3) from the 3D HNCACB experiments are shown. The strips illustrate the C^α and C^β sequential correlations between L3 to A8. The correlations between C^β are shown by red line connectivity where a less intense negative peak becomes a strong negative peak in the previous strip.

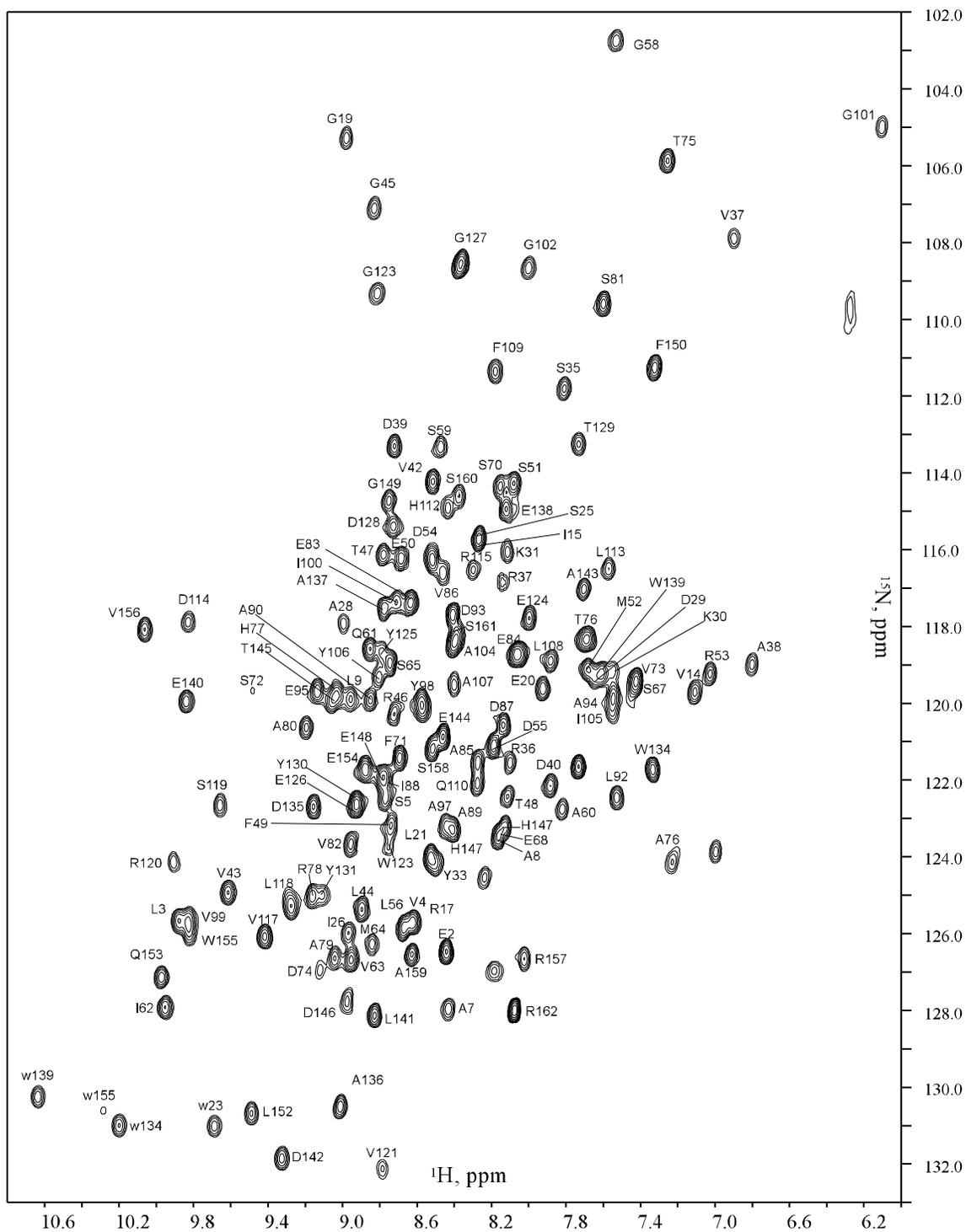
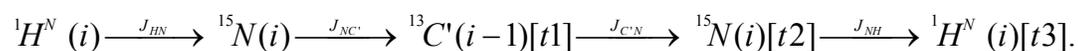


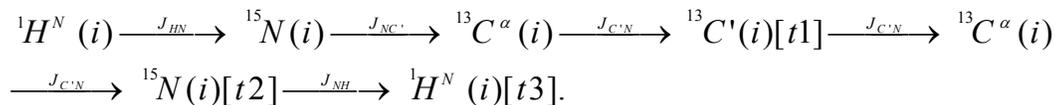
Figure 3.13. The labeled ^1H - ^{15}N HSQC spectrum of *HvdDHFR1* is shown. The $^1\text{H}/^{15}\text{N}$ chemical shifts were assigned to individual residue after completing the backbone chemical shift assignments.

result, the $^1\text{H}^\alpha$ chemical shift information was obtained from the 3D ^{15}N -NOESY-HSQC experiment. The detail of this experiment will be given in the next chapter. After predicting the secondary structure elements from the assigned chemical shift values, $^1\text{H}^\alpha$ resonances were able to be assigned. It has been known that there is a specific correlation between the amide protons ($^1\text{H}^{\text{N}}$) to other protons in the secondary structure elements. The $^1\text{H}^{\text{N}}$ of a residue always correlates with its own and previous $^1\text{H}^\alpha$, however the intensity of the peak varies depending on the secondary structure. In an α -helix, the correlation of $^1\text{H}^{\text{N}}$ to its own and previous residue's $^1\text{H}^\alpha$ result in a medium intensity NOE; whereas in a β -strand, $^1\text{H}^{\text{N}}$ correlates to its own $^1\text{H}^\alpha$ either weakly or very weakly and correlates to previous residue's $^1\text{H}^\alpha$ strongly. In a random coil, the intensity of the peak correlating to $^{15}\text{N}^{\text{H}}$ and its own $^1\text{H}^\alpha$ varies from medium to very weak. By using this information it was possible to assign most of the $^1\text{H}^\alpha$ chemical shift values.

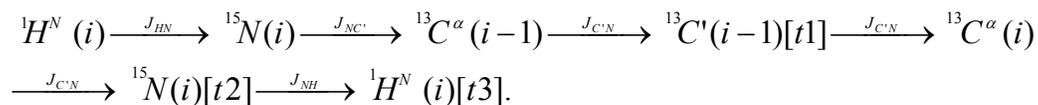
A 3D HNCO experiment provides carbon ($^{13}\text{C}'$) chemical shift values for the carbonyl group through correlating $^1\text{H}^{\text{N}}$, $^{15}\text{N}^{\text{H}}$, and $^{13}\text{C}'_{i-1}$ nuclei. In this experiment, the ^1H and ^{15}N chemical shifts of a residue correlate to the previous residue's $^{13}\text{C}'$ chemical shift. The pathway of the magnetization transfer is as follows;



The $^{13}\text{C}'$ chemical shift values of amino acids are confined in a narrow range at 165-190 ppm. In order to solve ambiguous chemical shift assignment in this region, a complementary experiment, 3D HN(CA)CO was used. In this experiment, ^1H and ^{15}N of a residue correlates to its own $^{13}\text{C}'_i$ chemical shift where the correlations among nuclei are established with the following magnetization transfer pathways:



In this experiment, the ${}^{13}C^\alpha$ nuclei are used to transfer the magnetization from ${}^{15}N$ to ${}^{13}C'$, of which the frequencies are not evolved and recorded. Due to the similar size of the ${}^1J_{NC\alpha i}$ (~ 11 Hz) and ${}^2J_{NC\alpha i-1}$ (~ 7 Hz) coupling constants, magnetization transfer to previous residue's C' would also take place in the order of:



However, while the HNC0 experiment is one of the most sensitive experiments, the HN(CO)CA is the least sensitive experiment where only limited information can be extracted. Since the experiment involves a large number of one step inter- and intraresidue magnetization transfers, the sensitivity is very low due to decay of transverse magnetization during the delay periods. In addition, the peak intensity does not provide a specific correlation for inter- and intraresidues in this experiment due to weak coupling. Figure 3.14 shows two strips from HNC0 and HN(CA)CO experiments, respectively. The chemical shifts of ${}^{15}N$ (127.14 ppm) and 1H (10.00 ppm) of Q153 correlate to the ${}^{13}C'$ chemical shift of L152 at 175.97 ppm in the HNC0 experiment. The same 1H and ${}^{15}N$ chemical shifts at the HN(CA)CO experiments correlate to ${}^{13}C'$ of both Q153 and L152 residues. The more intense peak at 175.97 ppm refers to ${}^{13}C'$ chemical shift of L152 and a less intense peak at 171.13 ppm correlated to the residue own (Q153) ${}^{13}C'$ chemical shift. Even though the coupling between the ${}^{15}N$ and its own ${}^{13}C^\alpha$ is larger, in this example it resulted in a less intense peak for its own ${}^{13}C'$. As a result, the connectivity along the peptide backbone shown as in Figure 3.10 and Figure 3.12 could

$^{15}\text{N}^{\text{H}} = 127.14 \text{ ppm}$
 $^1\text{H}^{\text{N}} = 10.00 \text{ ppm}$

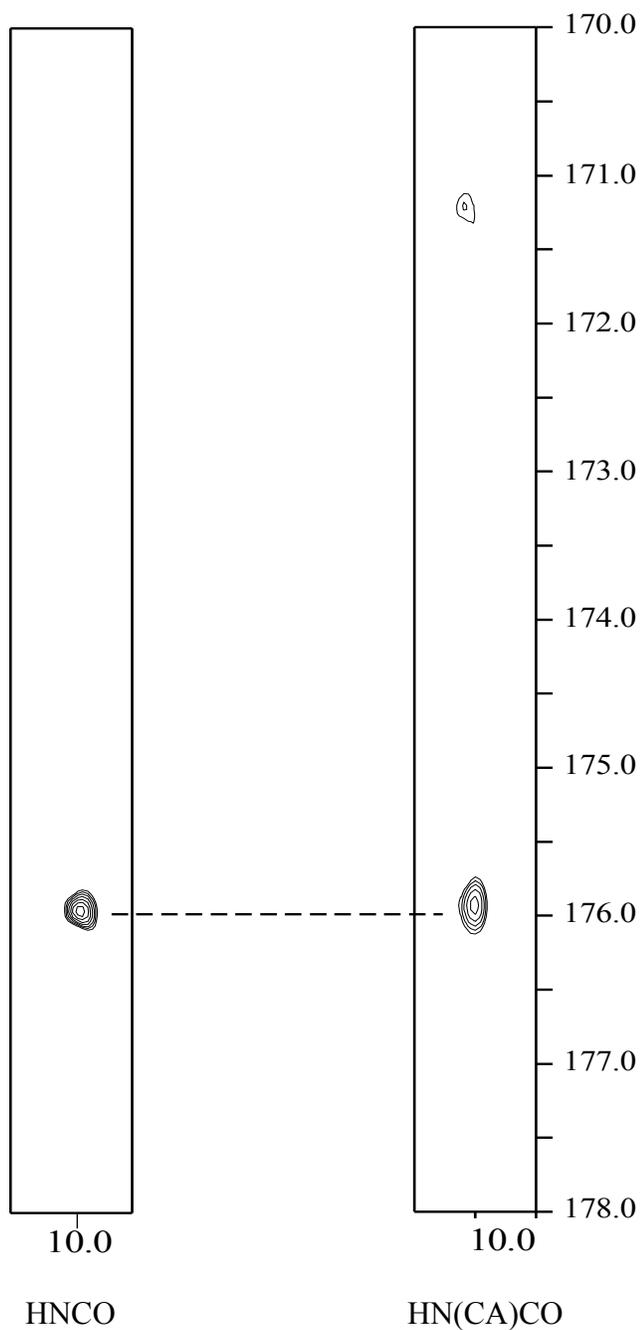


Figure 3.14. Two strips (f1-f3) from HNCO and HN(CA)CO experiments are illustrated. The chemical shift correlation between the two experiments is shown. The $^1\text{H}/^{15}\text{N}$ chemical shifts of Q153 correlate to $^{13}\text{C}'$ of L152 in HNCO. Both inter- and intrasidue $^{13}\text{C}'$ correlations were obtained in HN(CA)CO experiment.

not be obtained for $^{13}\text{C}'$ via the HNCO/HN(CA)CO experiments. As mentioned before, the reliability and the sensitivity of the experiment is very low. In some cases both correlations are obtained and the peak intensity may not be rationalized for its own or previous residue.

Side Chain Chemical Shift Assignments

Side chain chemical shift assignments were initiated once the backbone chemical shift assignments were completed. Assigning ^1H and ^{13}C nuclei of the aliphatic side chain chemical shift values of amino acids is crucial in order to obtain side chain distance restraints from the NOESY experiments. The aliphatic side chain carbon and proton chemical shift assignments were obtained via the 3D H(CCO)NH-TOCSY,^{48,49} (H)C(CO)NH-TOCSY,⁴⁸ and HCCH-TOCSY²³ experiments. Due to the large number of magnetization transfers in TOCSY step, the sensitivity of these experiments can be low for larger proteins. Figure 3.15 illustrates the mechanism and the nuclei involved in the magnetization transfer in these experiments.

A 3D (H)C(CO)NH-TOCSY experiment is utilized to assign the side chain ^{13}C chemical shift values of residues. In this experiment, $^1\text{H}^{\text{N}}$ and $^{15}\text{N}^{\text{H}}$ of a residue correlate to the C^{β} , C^{γ} , C^{δ} ...chemical shifts of the previous residue. The experiment is not designed to detect the $^{13}\text{C}^{\alpha}$ chemical shift value, but in some cases it may be observed. The magnetization on the side chain protons is transferred to the bound carbon nucleus and relayed to $^{13}\text{C}^{\alpha}$ via the TOCSY step. The flow of magnetization transfer during the experiment is as follows:

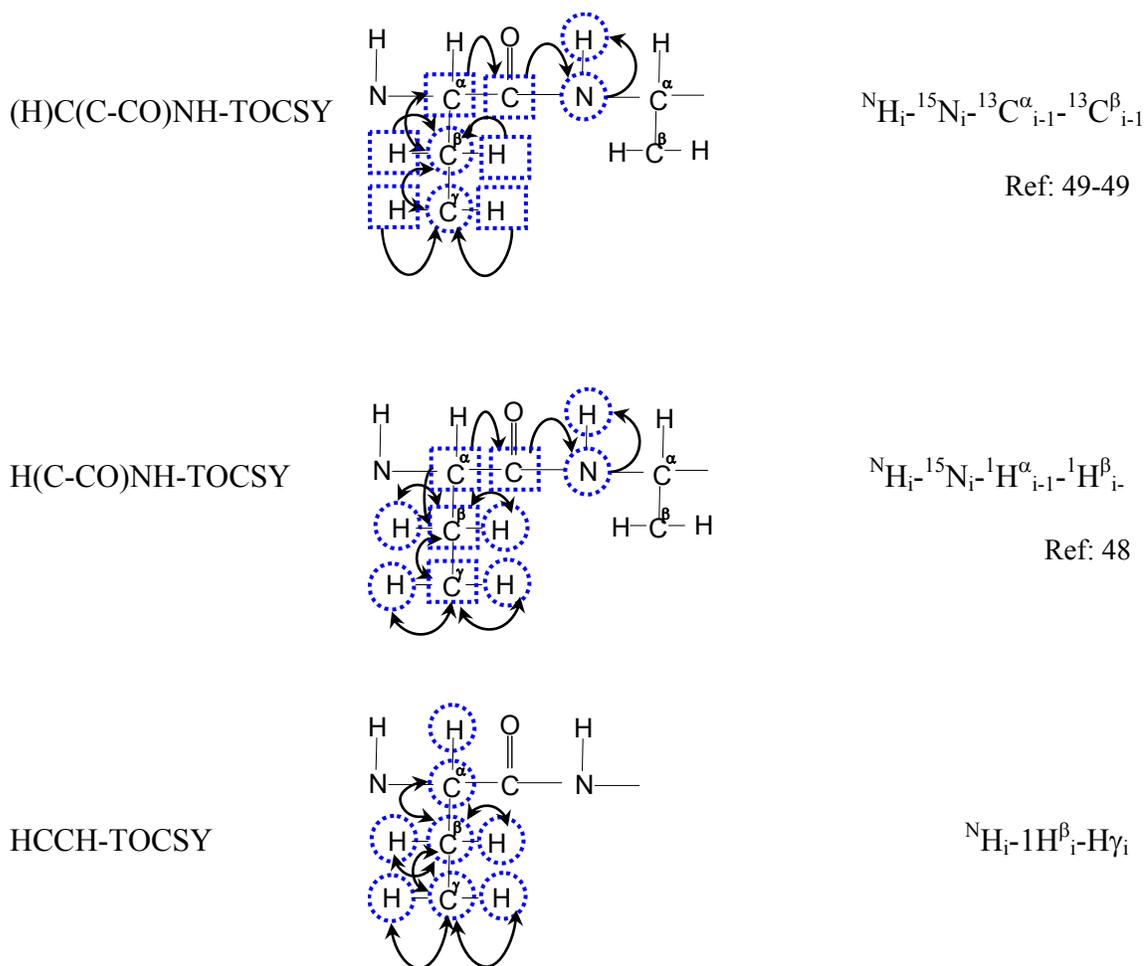


Figure 3.15. The list of 3D NMR experiments used for side chain proton and carbon chemical shift assignments.

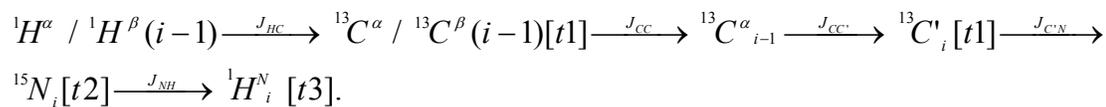


Figure 3.16 shows a strip from the (H)C(CO)NH-TOCSY experiment. The ${}^{15}\text{N}^{\text{H}}$ (117.36 ppm) and ${}^1\text{H}^{\text{N}}$ (8.73 ppm) chemical shifts of I100 are correlating to the ${}^{13}\text{C}^{\beta}$ and ${}^{13}\text{C}^{\gamma}$ chemical shifts of the previous residue, V99. The side chain carbon chemical shift values for the two methyl groups were assigned as 19.31 ppm and 21.37 ppm. The CH

$^{15}\text{N}^{\text{H}} = 117.36 \text{ ppm}$

$^1\text{H}^{\text{N}} = 8.73 \text{ ppm}$

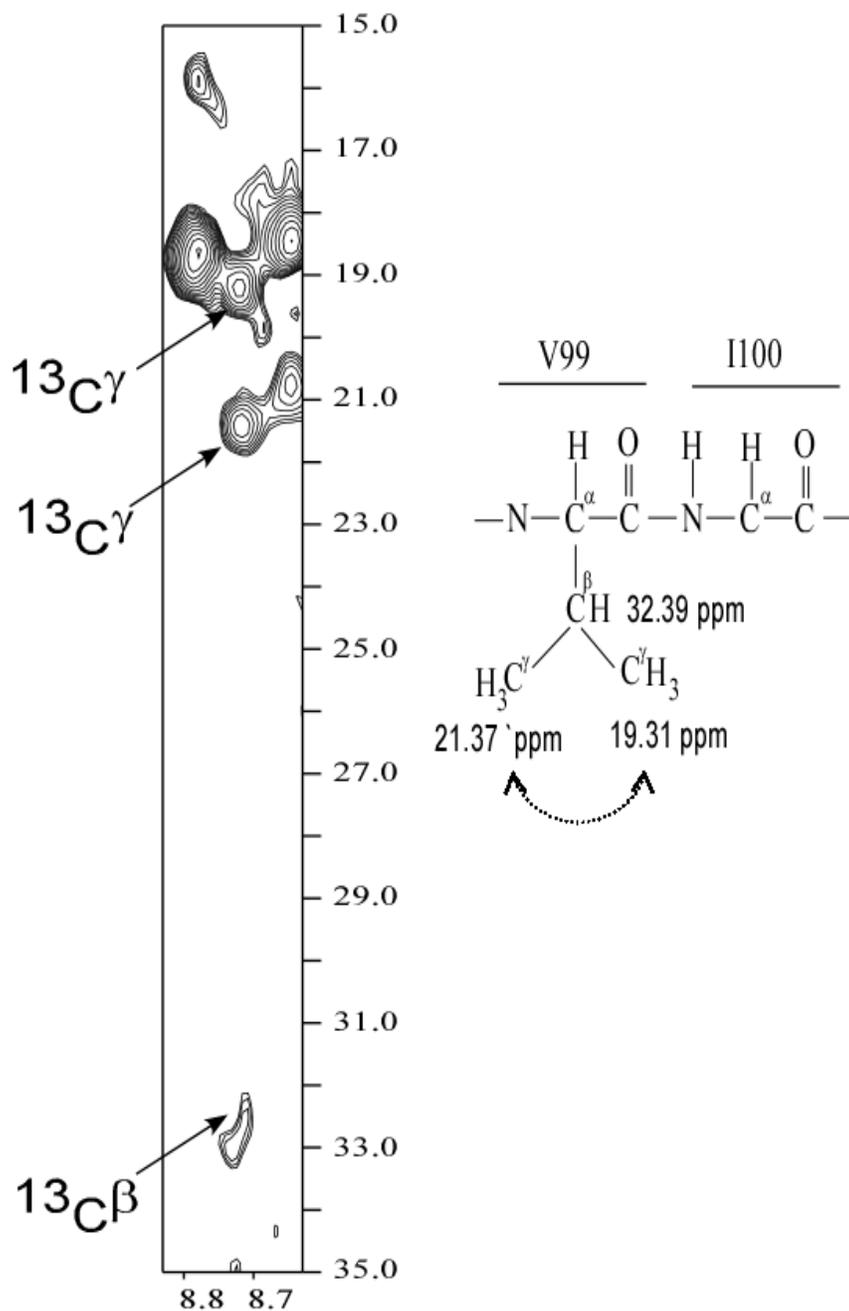


Figure 3.16. A strip (f1-f3) from (H)C(CO)NH-TOCSY demonstrates that $^1\text{H}^{\text{N}}/^{15}\text{N}$ chemical shift pair of the I100 correlating to the side chain carbon chemical shifts of V99 at 21.37, 19.31 and 32.39 ppm (identified by arrows).

group $^{13}\text{C}^\beta$ chemical shift is assigned as 32.39 ppm by using random chemical shift values as a guide.

A 3D H(C-CO)NH-TOCSY experiment is utilized to assign side chain proton chemical shift values. In this experiment, $^1\text{H}^{\text{N}}$ and $^{15}\text{N}^{\text{H}}$ of a residue correlate to the aliphatic side chain protons (H^β , H^γ , $\text{H}^\delta \dots$) chemical shift values of the previous residue. First, the protons' frequencies are recorded during the first evolutionary time and then magnetization is transferred to bonded ^{13}C nucleus where chemical shift information is intermixed via the TOCSY step. The magnetization pathway in this experiment is as follows:

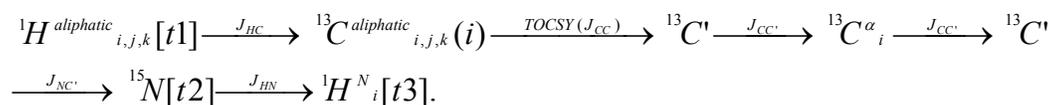


Figure 3.17 shows a strip from H(C-CO)NH-TOCSY experiment at which the ^{15}N and $^1\text{H}^{\text{N}}$ chemical shifts of I100 residue correlate to the side chain proton chemical shifts of V99 at 0.43 ppm, 0.87 ppm, and 2.18 ppm. The methyl groups' ($^1\text{H}^\gamma$) chemical shift values were tentatively assigned as 0.43 ppm and 0.87 ppm, and $^1\text{H}^\beta$ chemical shift value was assigned as 2.18 ppm. The more accurate carbon and proton chemical shifts assignment of side chain (CH_3 , CH_2 , CH) groups were obtained via the HCCH-TOCSY experiment.⁵⁷

The HCCH-TOCSY experiment correlates all of the aliphatic side chain ^1H and ^{13}C chemical shifts information of a residue. The big challenge of this experiment is to be able to assign the ^1H chemical shift to directly bound ^{13}C nuclei. The flow of magnetization transfer is slightly different from the previous experiments which are as follows:

$^{15}\text{N}^{\text{H}} = 117.36 \text{ ppm}$
 $\text{H}^{\text{N}} = 8.78 \text{ ppm}$

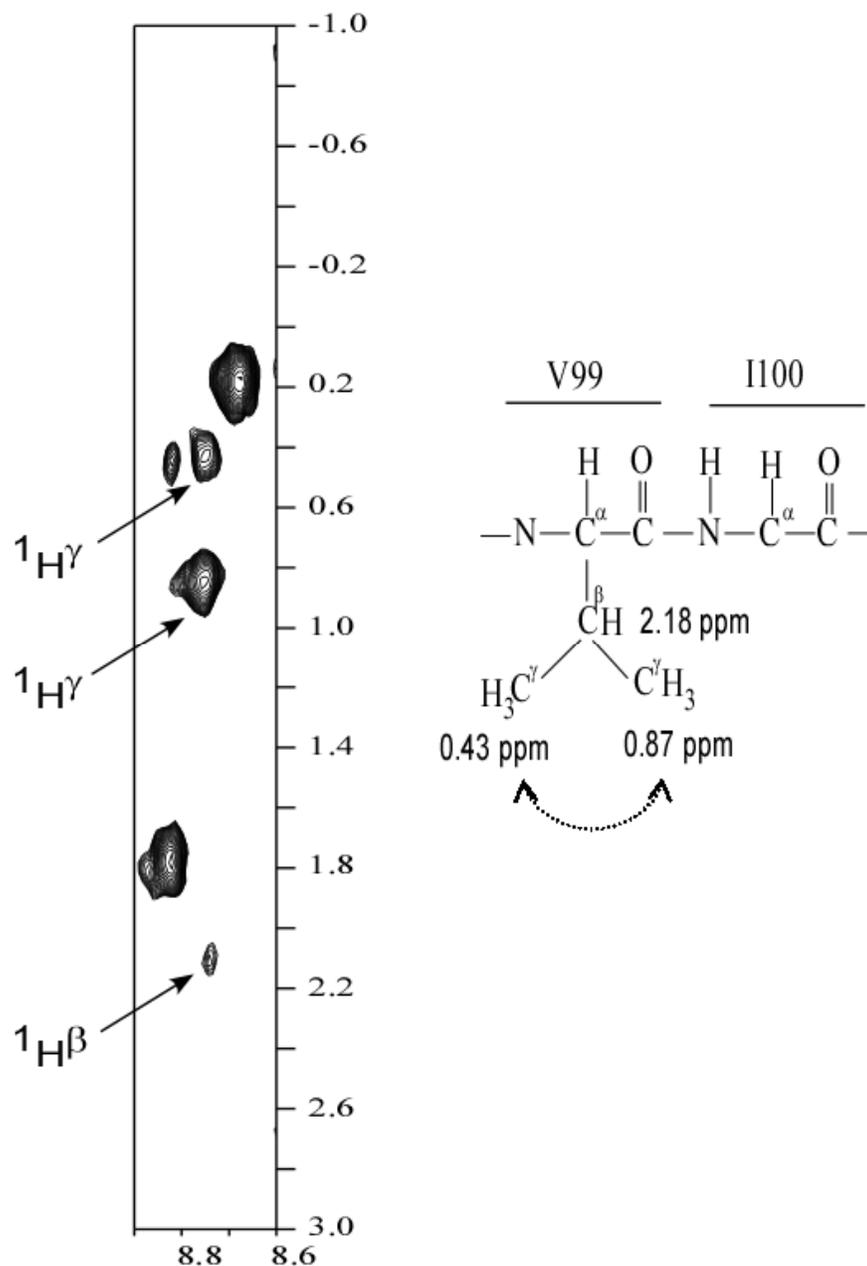
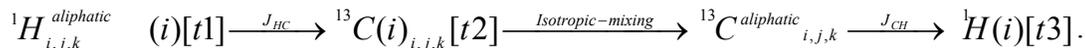


Figure 3.17. A strip (f1-f3) from H(CCO)NH-TOCSY experiment. The $^{1}\text{H}^{\text{N}}/^{15}\text{N}$ chemical shift pair of I100 correlates to three side chain proton chemical shifts of previous residue, V99, at 0.43 ppm, 0.87 ppm and 2.18 ppm. The H^{γ} chemical shifts were tentatively assigned to methyl groups. The more accurate $^{1}\text{H}^{\text{aliphatic}}\text{-}^{13}\text{C}^{\text{aliphatic}}$ correlations were accomplished via the 3D HCCH-TOCSY experiment.



The frequencies of the protons are recorded during the first evolutionary time (t1) and then encoded on the directly attached aliphatic ${}^{13}C$ nucleus via one bond coupling (${}^1J_{HC} \sim 130$ Hz). The frequencies of ${}^{13}C$ nuclei are allowed to evolve and are recorded as the second dimension before transferring magnetization along side chain ${}^{13}C$ nuclei. The magnetization information of ${}^{13}C$ nuclei along the side chain is intermixed within the spin system via the isotropic mixing period which employs one bond ${}^{13}C$ - ${}^{13}C$ ($J_{CC} \sim 30$ - 40 Hz) coupling interaction. The magnetization is transferred back to aliphatic 1H at which the frequencies are recorded as the third dimension. The accurate isotropic mixing time is important for maximum magnetization transfer within the spin system. An embedded isotropic mixing pulse scheme such as DIPSI (DIPSI-1, DISPSI-2, DIPSI-3) is a more common tool used in this experiment.⁵⁸⁻⁶² The DIPSI-3 isotropic mixing time was used in this study and was set to 18.7 ms.⁶³ The accuracy of the side chain 1H and ${}^{13}C$ assignments is validated with this experiment. The 3D spectra are analyzed by cross sectioning of ${}^{13}C$ nucleus axis. The diagonal peak in each 2D spectrum represents the proton chemical shift of the directly bonded carbon nuclei. The cross peaks correlate to other protons in a single spin system.

Figure 3.18 shows three strips of ${}^{13}C$ planes from the HCCH-TOCSY experiment. Previously assigned chemical shift values of ${}^1H^\beta$ and ${}^1H^\gamma$ from the H(CCO)NH-TOCSY were tentatively assigned to ${}^{13}C$ nuclei from which the chemical shifts were obtained from (H)C(CO)NH-TOCSY. The top strip shows the ${}^{13}C^\gamma$ chemical shift value of one of the methyl groups at 19.20 ppm. The diagonal peak at 0.43 ppm in this plane indicates

the methyl ^1H chemical shift value is 0.43 ppm and the cross peaks at 0.87 ppm and 2.17 ppm correlate to the other ^1H chemical shifts in the V99 spin system. Similarly, the middle strip is from the other methyl's $^{13}\text{C}^\gamma$ plane at 21.45 ppm of which the diagonal peak at 0.87 ppm is the chemical shift of the ^1H on this ^{13}C nucleus. The bottom strip is from the $^{13}\text{C}^\beta$ plane where the diagonal peak at 2.17 ppm is the chemical shift of the ^1H on this nucleus, which is correlated to the ^1H chemical shifts on methyl groups.

In practice, during the course of chemical shift assignments, the HCCH-TOCSY experiment would be conducted last. The sensitivity of the experiment can be dramatically increased by substituting the $^1\text{H}_2\text{O}$ with $^2\text{H}_2\text{O}$ by lyophilization process. Dissolving the powder form of a protein in $^2\text{H}_2\text{O}$ would provide the advantage of suppressing the residual water signal with lower power *rf* pulses instead of applying the larger power of *rf* pulses to $^1\text{H}_2\text{O}$. This allows saving the chemical shift information of the nuclei resonating close to water protons (mainly $^1\text{H}^\alpha$ region) from being dissipated during water suppression. Since the lyophilized *Hv*DHFR1 were not able to be re-dissolved in $^2\text{H}_2\text{O}$, the experiment had to be acquired in 95% $^1\text{H}_2\text{O}$ /5% $^2\text{H}_2\text{O}$.

For residues with longer side chain spin systems, such as such as I, R, P, and K, carbon-carbon magnetization transfer, which may not be complete along the side chain during the DISPI mixing period or transfer rate, may be attenuated due to high salt effect.^{64,65} However, one and two step transfers were still detected which allowed for complete side chain assignments. Figure 3.19 shows several $^{13}\text{C}^\alpha$ plane strips to illustrate the assignment strategy used for residue R46 in the HCCH-TOCSY experiment. The top strip corresponds to the $^{13}\text{C}^\alpha$ chemical shift of R46 at 60.27 ppm. At this plane, the diagonal peak at 4.35 ppm was assigned as the directly attached $^1\text{H}^\alpha$. This diagonal $^1\text{H}^\alpha$

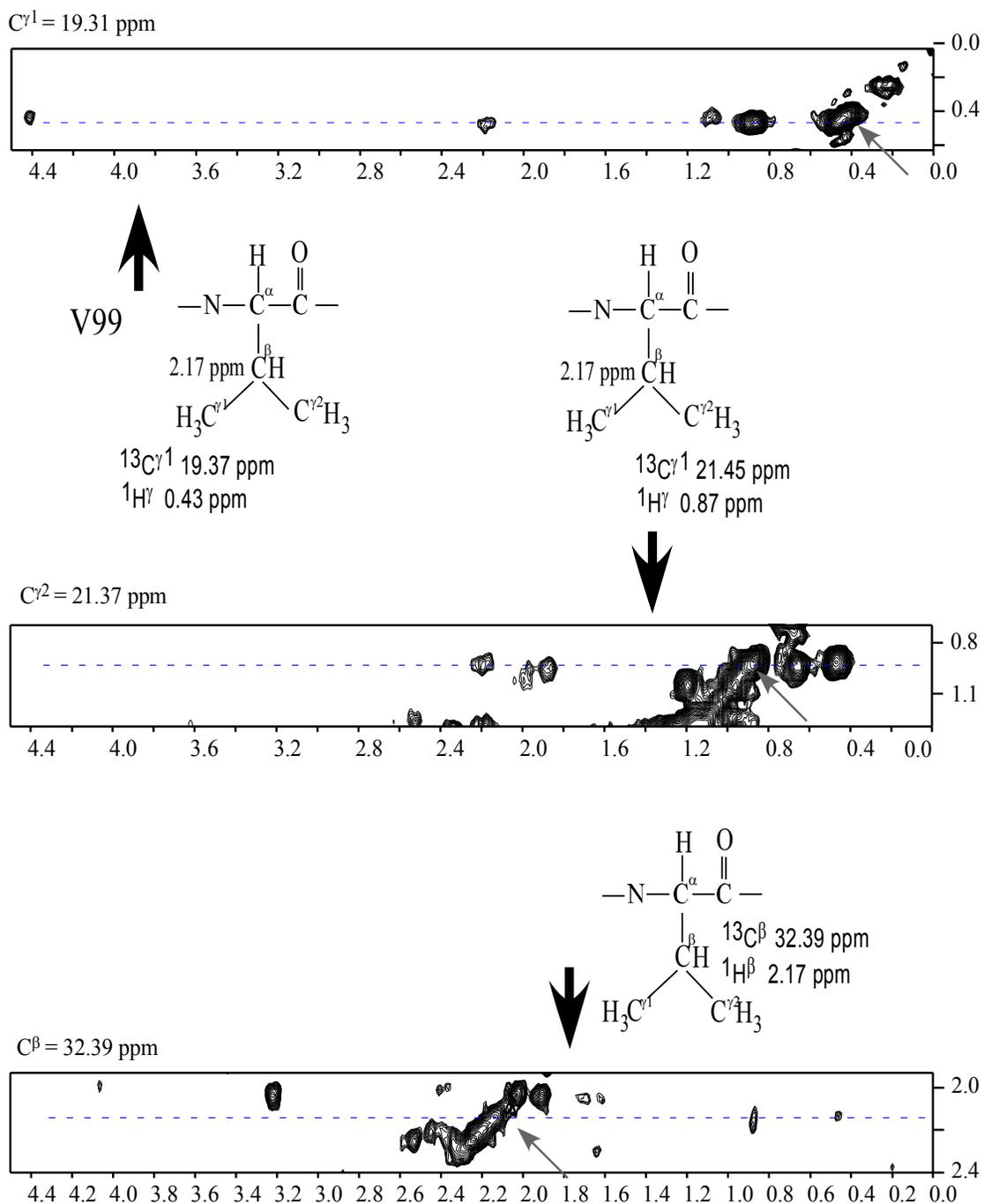


Figure 3.18. Strips $[f_1(^1\text{H})-f_3(^1\text{H})]$ from HCCH-TOCSY experiment are shown. The strips were taken from $^{13}\text{C}^{\beta}$ and $^{13}\text{C}^{\gamma}$ planes of V99. The diagonal peaks (indicated by arrows) represent the chemical shift value of the proton directly attached to carbon nucleus. The cross peaks correspond to proton chemical shifts of the neighboring carbon nuclei on the same residue.

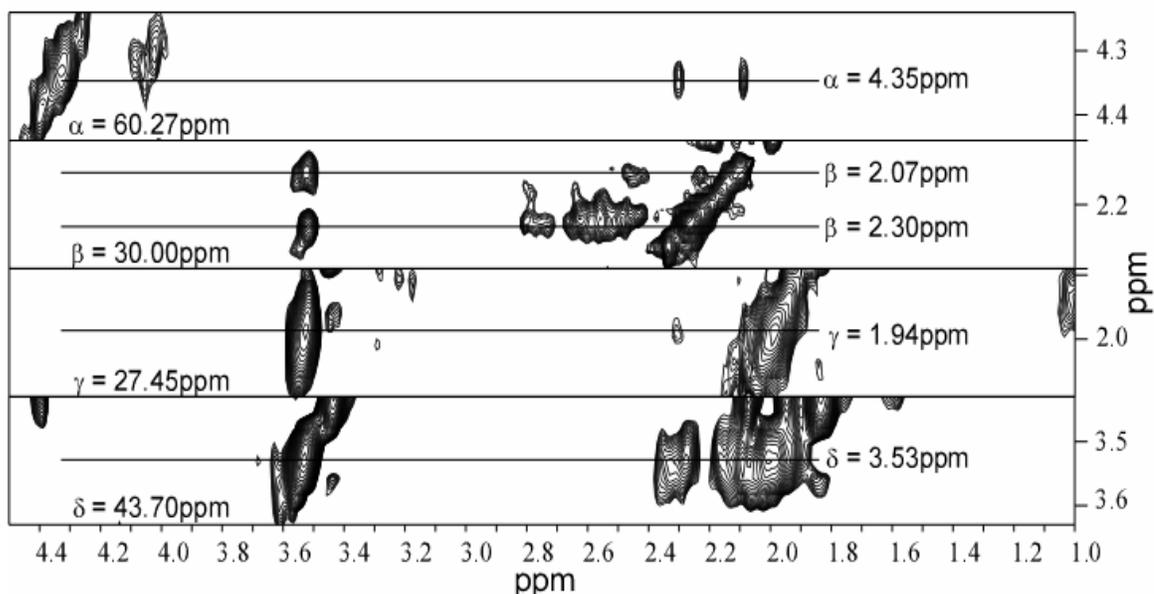


Figure 3.19. Strips [f1(^1H)-f3(^1H)] from the HCCH-TOCSY spectrum illustrating the assignment strategy used for R46. The strips from $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}^\gamma$, $^{13}\text{C}^\delta$ planes point up that there is no complete magnetization transfer within the spin system. By piecing the data from each plane, the chemical shift assignments were able to be accomplished.

peak correlates to two other peaks at 2.07 and 2.30 ppm. In the next strip, at the $^{13}\text{C}^\beta$ chemical shift of 30.00 ppm, the two peaks at 2.07 and 2.30 ppm were observed as diagonal peaks and were assigned as the directly attached protons' chemical shifts on the $^{13}\text{C}^\beta$. In this plane, these two diagonal peaks correlate to a proton resonance at 3.53 ppm. Similarly, by tracking the proton resonances on $^{13}\text{C}^\gamma$ and $^{13}\text{C}^\delta$ planes, proton resonances of the side chain were able to be assigned through piecing together the chemical shifts.

Conclusion

The near complete backbone and side chain chemical shift assignments for the enzyme *Hv*DHFR1 at 3.5 M NaCl were accomplished. By using a sample reduction tube the coupling resulted from the high ionic buffer was minimized. This enables the

acquisition of triple resonance experiments with standard experimental conditions. This study represents the first NMR structural study to have ever been done at such high salt. Out of a total of 1595, it was able to assign to 1454 of the chemical shifts values. This corresponds to 97% of the backbone amides; the residues A10, E11, R13, D18, and R69 were unassigned. In total, 98% of $^{13}\text{C}^\alpha$, 96% of $^{13}\text{C}^\beta$, 92% of ^{13}CO , 97% of $^1\text{H}^\alpha$ and 97% of $^1\text{H}^\beta$ were assigned. The backbone and side chain chemical shift values for individual residues are given in Table 3.1 and have been deposited in the BioMagResBank under accession number 6645. The chemical shift assignments of aromatic residues were not conducted because there are a few aromatic residues in protein sequence from which only limited information can be obtained.

The chemical shift information of residues is going to be used to predict the secondary structure of the protein. The chemical shifts are sensitive to the local environments nuclei are experiencing. Certain patterns of deviation of backbone chemical shift values of amino acids from their random chemical shift values relate to their secondary structure location. The chemical shifts will also be used to predict the phi and psi dihedral angles. Then the assigned chemical shifts values will be used in the NOE cross peaks assignments in ^{13}C -HSQC-NOESY and ^{15}N -HSQC-NOESY experiments. This will enable us obtain proton-proton distance restraints which will be used to elucidate the three dimensional protein structure.

Table 3.1. The assigned backbone and side chain chemical shift values of *Hv*DHFR1.

$^1\text{H}^{\text{N}}/^{15}\text{N}^{\text{H}}$	$^1\text{H}^{\alpha}/^{13}\text{C}^{\alpha}$	$^1\text{H}^{\beta}/^{13}\text{C}^{\beta}$	$^1\text{H}^{\gamma}/^{13}\text{C}^{\gamma}$	$^1\text{H}^{\delta}/^{13}\text{C}^{\delta}$	$^1\text{H}^{\epsilon}/^{13}\text{C}^{\epsilon}$	C(Carbonyl)
M1		4.25 54.77	2.13 - 2.30 25.91	2.51, 2.76 30.28		172.31
E2	8.47 126.51	5.08 54.77	2.24 31.76	2.48 - 2.78		174.35
L3	9.91 125.69	5.28 53.53	1.86 44.14	1.56 26.49	1.12 25.04	0.96 25.04 176.83
V4	8.64 125.65	5.53 59.95	1.99 35.37	1.17 20.91	0.93 20.84	176.76
S5	8.80 122.49	5.06 56.87	4.37 64.74			173.35
V6	8.60 119.85	5.50 60.80	2.26 33.85	0.10 22.20	1.12 22.09	175.79
A7	8.46 127.98	4.87 51.74	1.05 24.51			173.73
A8	8.19 123.55	4.67 50.62	1.59 20.66			175.23
L9	8.99 119.98	5.56 53.80	1.91 45.12	1.17 26.71	0.80 23.70	-0.08 19.03
A10	9.09 113.74					
E11		4.23 54.81	2.17 31.53	2.25 - 2.32 33.02		174.93
N12	8.40 118.50	4.50 54.92	3.01 - 3.51 41.42			
R13	8.38 118.19	4.02 58.75	1.22 - 1.56	1.49	2.81	174.49
V14	7.13 119.89	4.00 63.87	2.56 32.88	0.84 20.26	1.09 18.97	176.58
I15	8.31 115.72	4.26 59.29	2.50 - 2.21	1.57 - 1.98 25.73	1.07 17.06	1.07 11.17 176.37
G16	8.40 108.80	4.26 45.71				175.17
R17	8.65 125.96	5.31 53.45	1.82	1.82	2.89	
D18		4.33	2.68 - 3.15 40.04			176.04
G19	9.00 105.33	3.82 - 4.54 45.35				174.19
E20	7.95 119.66	4.39 53.69	2.01 - 2.16 31.69	2.37 35.50		175.74
L21	8.55 124.05	4.53 52.43	1.73 41.41	1.49 25.52	0.87 24.89	1.05 26.74
P22		4.55 65.57	2.56 - 2.36 25.44	2.02 26.07	3.93 - 4.09 46.83	177.73

Table 3.1. (continued)

	$^1\text{H}^{\text{N}}/\text{N}^{\text{H}}$	$^1\text{H}^{\alpha}/^{13}\text{C}^{\alpha}$	$^1\text{H}^{\beta}/^{13}\text{C}^{\beta}$	$^1\text{H}^{\gamma}/^{13}\text{C}^{\gamma}$	$^1\text{H}^{\delta}/^{13}\text{C}^{\delta}$	$^1\text{H}^{\epsilon}/^{13}\text{C}^{\epsilon}$	C(Carbonyl)
W23	8.78	4.55	3.00 - 3.48			9.69	
	123.22	57.71	32.26			131.01	
P24		4.58	2.56 - 2.02	2.31	3.59 - 3.76		
		62.59	31.40	25.32	51.50		176.59
S25	8.29	4.27	4.00				
	115.72	59.54	63.39				174.66
I26	8.99	4.55	2.03	1.27	0.90	0.86	
	126.15	58.70	39.10	26.56	16.84	14.24	
P27		4.47	2.23	1.77	3.95		
		64.89	28.20	26.05	51.10		179.18
A28	9.01	4.22	1.94				
	117.97	54.94	19.05				180.71
D29	7.68	4.45	2.50				
	119.25	56.21	36.47				177.23
K30	7.58	4.22	2.04 - 1.87	1.65	2.02	3.17	
	119.29	59.05	31.30	26.26			178.60
K31	8.15	3.55	2.44	2.01	2.16	3.11	
	116.05	59.15	32.09	22.85?	26.86?	41.02	179.68
Q32	7.72	3.66	1.73 - 1.96	2.77			
	118.35	58.12	31.58	29.89			177.70
Y33	8.52	4.14	3.28 - 3.23				
	124.14	60.35	37.74				176.23
R34	8.17	4.36	2.02	1.76	3.44		
	116.85	57.60	28.33	27.35	41.50		179.50
S35	7.83	4.11	3.71				
	111.85	60.78	63.07				176.75
R36	8.13	4.40	1.82 - 2.06	1.16	3.21		
	121.54	58.47	29.89	26.79	42.19		177.99
V37	6.93	4.15	1.38	0.15	0.55		
	107.75	60.66	30.15	14.77	14.33		175.89
A38	6.83	4.12	1.60				
	119.01	55.38	19.87				178.35
D39	8.74	5.08	2.64 - 3.12				
	113.31	53.83	42.60				174.92
D40	7.91	4.82	2.75				
	122.14	53.90	44.80				
P41		5.11	2.10	1.27 - 1.56	3.75		
		62.72	31.10	26.45	49.26		175.74
V42	8.54	5.70	2.23	1.09	0.88		
	114.27	57.86	34.22	22.48	20.48		174.19
V43	9.64	5.22	1.64	0.69	0.71		
	124.99	60.38	32.45	21.86	20.80		174.58
L44	8.93	5.66	1.96	1.31	0.96	0.96	
	125.37	53.45	43.82	27.10	18.64	18.95	175.59
G45	8.86	4.44 - 4.82					
	107.12	44.90					174.74
R46	8.74	4.35	2.07 - 2.30	1.94	3.53		
	120.29	60.27	30.00	27.45	43.70		176.37

Table 3.1. (continued)

	$^1\text{H}^{\text{N}}/\text{N}^{\text{H}}$	$^1\text{H}^{\alpha}/^{13}\text{C}^{\alpha}$	$^1\text{H}^{\beta}/^{13}\text{C}^{\beta}$	$^1\text{H}^{\gamma}/^{13}\text{C}^{\gamma}$	$^1\text{H}^{\delta}/^{13}\text{C}^{\delta}$	$^1\text{H}^{\epsilon}/^{13}\text{C}^{\epsilon}$	C(Carbonyl)
T47	8.80	4.29	4.05	1.48			
	116.06	66.88	67.65	21.82			177.34
T48	8.14	4.10	4.27	0.98			
	122.45	67.23	67.46	22.42			175.99
F49	8.77	4.05	3.00 - 3.51				
	123.12	62.34	38.93				176.32
E50	8.71	3.99	2.31	2.73			
	116.25	59.00	28.62	31.35			180.00
S51	8.11	4.51	4.25				
	114.33	60.62	63.15				175.60
M52	7.71	4.07	1.59	1.80		2.25	
	119.05	55.58	33.27			15.10	176.93
R53	7.01	4.31	1.81 - 1.89		1.63	3.29	
	119.26	58.97	29.82	26.80	43.33		176.57
D54	8.54	4.40	2.68				
	116.32	54.94	40.24				175.86
D55	8.22	4.79	3.43				
	121.15	53.57	40.03				174.96
L56	8.69	4.38	1.82	1.19	0.91	0.58	
	125.84	59.98	41.88	28.69	26.09	25.03	
P57		4.94	2.06	1.60	3.73		
		61.89	32.86	28.35	50.38		176.01
G58	7.56	4.17 - 4.57					
	102.73	45.10					174.95
S59	8.50	5.00	4.55				
	113.30	58.62	62.20				175.23
A60	7.84	4.80	1.55				
	122.99	51.21	22.61				175.54
Q61	8.87	5.30	2.06	2.27			
	118.64	55.71	28.28	33.98			173.14
I62	9.98	5.28	2.09	1.28 - 1.50		1.03	0.88
	127.92	60.99	37.09	26.14	18.03	14.11	173.46
V63	8.98	4.66	1.78	0.58	-0.05		
	126.70	60.48	31.37	20.51	20.51		176.64
M64	8.88		2.29	2.64			
	126.38	52.14	27.90	31.17			176.26
S65	8.77	4.62	4.02				
	119.01	57.40	64.33				174.73
R66							174.69
S67	7.45	4.93	4.10 - 4.72				
	119.53	58.00	63.91				174.13
E68	8.19	4.36	1.06	1.57			
	123.47	59.88					
R69			2.30				
		57.36	29.72				176.77
S70	8.18	4.59	3.98				
	114.37	57.42	63.79				173.76

Table 3.1. (continued)

	$^1\text{H}^{\text{N}}/\text{N}^{\text{H}}$	$^1\text{H}^{\alpha}/^{13}\text{C}^{\alpha}$	$^1\text{H}^{\beta}/^{13}\text{C}^{\beta}$	$^1\text{H}^{\gamma}/^{13}\text{C}^{\gamma}$	$^1\text{H}^{\delta}/^{13}\text{C}^{\delta}$	$^1\text{H}^{\epsilon}/^{13}\text{C}^{\epsilon}$	C(Carbonyl)
F71	8.72	4.64	2.99 - 3.43				
	121.47	57.33	40.65				176.55
S72	9.51	5.18	4.23				
	119.66	59.08	62.80				174.69
V73	7.45	4.35	2.31	1.26	1.31		
	119.53	61.33	33.35	21.29	20.97		177.62
D74	9.15	4.67	3.00				
	126.89	56.80	40.81				177.15
T75	7.28	4.61	4.17	1.48			
	105.80	61.49	68.32	21.25			172.82
A76	7.26	5.72	1.95				
	124.18	50.51	19.85				176.17
H77	9.09	4.97	2.89 - 3.34				
	119.37	54.39	29.65				173.87
R78	9.19	5.27	2.11 - 1.88	1.01	3.26		
	125.13	56.11	31.24	26.24	41.94		174.83
A79	9.08	4.95	1.55				
	126.71	50.55	22.87				175.79
A80	9.22	4.19	1.63				
	120.69	51.54	20.42				177.02
S81	7.61	4.45	4.07				
	109.60	55.80	66.47				174.21
V82	8.98	3.47	1.76	0.19	0.14		
	123.76	66.47	30.78	18.46	20.76		177.11
E83	8.65	4.00	2.09 - 2.14	2.46			
	117.38	59.78	28.40	36.77			179.09
E84	8.07	4.26	2.26 - 1.98	2.49 - 2.42			
	118.74	59.03	30.07	36.87			178.38
A85	8.30	4.06	1.76				
	121.61	55.26	18.31				179.19
V86	8.49	3.61	2.34	1.19	1.14		
	116.62	66.80	31.19	21.36	22.82		177.81
D87	8.16	3.92	2.74 - 2.99				
	120.63	57.34	40.31				180.37
I88	8.80	3.83	1.94	1.04 - 1.20	0.62	0.55	
	121.89	64.94	37.60	28.72	16.57	14.27	179.53
A89	8.47	3.90	1.45				
	123.16	55.95	16.14				179.17
A90	8.87	4.22	1.69				
	119.85	54.90	17.44				182.53
S91	8.43	4.58	4.28				
	118.73	61.01	62.60				175.11
L92	7.59	4.22	1.87	1.85	0.63	0.86	
	122.45	54.44	40.89	25.51	23.33	20.83	176.77
D93	8.44	4.58	2.78 - 2.31				
	117.73	55.11	38.65				175.13
A94	7.58	4.38	1.48				
	119.97	52.13	20.45				177.77

Table 3.1. (continued)

	$^1\text{H}^{\text{N}}/\text{N}^{\text{H}}$	$^1\text{H}^{\alpha}/^{13}\text{C}^{\alpha}$	$^1\text{H}^{\beta}/^{13}\text{C}^{\beta}$	$^1\text{H}^{\gamma}/^{13}\text{C}^{\gamma}$	$^1\text{H}^{\delta}/^{13}\text{C}^{\delta}$	$^1\text{H}^{\epsilon}/^{13}\text{C}^{\epsilon}$	C(Carbonyl)
A97	8.44	5.02	1.07				
	123.40	50.22	19.67				173.37
Y98	8.60	5.29	2.89				
	120.17	56.76	40.74				175.34
V99	9.85	6.11	2.18	0.43	0.87		
	125.63	61.69	32.39	19.31	21.37		177.51
I100	8.73	5.68	2.20	1.16	0.47	0.98	
	117.36	61.67	37.91	25.20	18.91	10.10	174.93
G101	6.13	4.34 - 4.36					
	102.73	42.04					174.84
G102	8.03	4.25					
	108.62	45.81					175.19
A103	9.31	4.23	1.70				
	125.02	57.23	19.43				182.49
A104	8.42	3.94	1.74				
	118.33	55.13	18.01				180.39
I105	7.58	4.14	2.26	1.39 - 1.19	0.93	1.00	
	120.11	61.70	34.79	22.11	18.17	9.85	179.23
Y106	8.83	4.42	2.77 - 3.02				
	119.36	60.36	37.72				179.59
A107	8.42	4.17	1.72				
	119.45	55.38	17.55				179.66
L108	7.91	4.23	2.04	1.52	0.96	1.06	
	118.97	57.31	42.88	25.53	25.53	23.13	178.45
F109	8.21	4.40	3.31 - 2.52				
	111.26	61.21	40.73				177.20
Q110	8.30	3.98	2.11	2.28			
	122.09	60.38	24.32	36.58			
P111		62.57	30.90	26.30 - 27.28	50.07		176.93
H112	8.47	4.60	3.26 - 3.48				
	114.94	55.33	31.76				175.74
L113	7.60	4.37	2.18 - 1.41	1.51	1.26	1.12	
	116.49	55.84	43.85	25.99	22.70	22.09	177.38
D114	9.85	4.85	2.81 - 2.72				
	117.87	56.33	44.95				176.58
R115	8.33	5.02	2.11 - 2.41	1.47 - 1.24	3.49		
	116.54	55.49	29.19	28.50	43.42		175.88
M116	7.75	5.23	2.25 - 2.56	2.65			
	121.65	55.53	29.01	31.52			174.78
V117	9.44	5.03	2.52	0.88	1.10		
	126.13	62.28	30.85	20.80	19.89		175.19
L118	9.30	5.54	0.12 - 1.14	0.41	0.20	0.11	
	125.36	52.11	43.97	26.38	20.84	19.47	176.56
S119	9.69	5.23	3.85				
	122.67	57.32	61.33				172.95
R120	9.93	5.48	1.51	1.11 - 1.03	3.28		
	124.19	52.35	26.05	25.06	41.94		175.85

Table 3.1. (continued)

	$^1\text{H}^{\text{N}}/\text{N}^{\text{H}}$	$^1\text{H}^{\alpha}/^{13}\text{C}^{\alpha}$	$^1\text{H}^{\beta}/^{13}\text{C}^{\beta}$	$^1\text{H}^{\gamma}/^{13}\text{C}^{\gamma}$	$^1\text{H}^{\delta}/^{13}\text{C}^{\delta}$	$^1\text{H}^{\epsilon}/^{13}\text{C}^{\epsilon}$	C (Carbonyl)
S119	9.69	5.23	3.85				
	122.67	57.32	61.33				172.95
R120	9.93	5.48	1.51	1.11 - 1.03	3.28		
	124.19	52.35	26.05	25.06	41.94		175.85
V121	8.80	6.09	2.12	-0.10	1.02		
	132.22	59.73	33.06	18.95	19.31		
P122		4.61	2.15	1.99 - 2.25	3.52 - 3.60		
		63.10	31.89	27.70	50.83		177.22
G123	8.85	3.89 - 4.48					
	109.35	44.05					172.23
E124	8.02	4.37	1.96	2.44			
	117.83	55.01	30.45	34.71			174.29
Y125	8.80	4.30	2.64 - 3.28				
	118.62	55.77	42.09				173.75
E126	8.95	4.23	2.24	2.50			
	122.67	56.07	29.92	36.35			176.37
G127	8.39	4.22					
	108.54	45.50					171.28
D128	8.76	4.50	2.96				
	115.48	53.35	42.32				175.99
T129	7.76	4.98	4.16	1.46			
	113.20	60.56	72.08	22.87			173.74
Y130	8.95	4.60	2.84 - 3.14				
	122.67	57.28	41.86				175.24
Y131	9.14	4.71	3.02				
	125.05	57.39	39.90				
P132		4.40	2.18 - 2.21	2.02	3.67 - 3.71		
		63.10	31.89	26.67	50.41		174.27
E133	8.09	3.94	1.86 - 1.99	2.15 - 2.12			
	118.76	56.84	29.94	36.38			175.79
W134	7.36	5.20	2.89 - 3.06			10.20	
	121.76	55.64	30.54			131.00	174.45
D135	9.18	4.54	2.87				
	122.70	53.48	41.73				177.80
A136	9.04	4.58	1.78				
	130.54	55.06	18.61				178.76
A137	8.80	4.63	1.75				
	117.58	53.83	18.74				179.73
E138	8.15	4.56	2.29	2.52			
	114.91	55.96	31.62	36.27			175.49
W139	7.64	5.33	3.89 - 3.59			10.64	
	119.25	56.68	34.17			130.25	174.67
E140	9.86	5.29	2.17	2.41			
	119.98	54.17	32.65	35.84			175.09
L141	8.85	4.84	1.57	1.56	0.60	-0.03	
	128.16	54.90	40.05	25.52	25.52	22.02	175.52
D142	9.34	4.83	2.67				
	131.87	55.61	44.52				176.05

Table 3.1. (continued)

	$^1\text{H}^{\text{N}}/\text{N}^{\text{H}}$	$^1\text{H}^{\alpha}/^{13}\text{C}^{\alpha}$	$^1\text{H}^{\beta}/^{13}\text{C}^{\beta}$	$^1\text{H}^{\gamma}/^{13}\text{C}^{\gamma}$	$^1\text{H}^{\delta}/^{13}\text{C}^{\delta}$	$^1\text{H}^{\epsilon}/^{13}\text{C}^{\epsilon}$	C(Carbonyl)
A143	7.73 117.07	4.72 51.89	1.49 22.21				174.98
E144	8.48 120.93	5.65 55.48	1.97 2.08 33.59	2.14 - 2.18 37.87			175.29
T145	9.06 119.82	5.53 60.91	4.11 71.33	1.15 21.50			173.22
D146	9.00 127.83	4.85 55.29	2.72 2.74 41.06				174.70
H147	8.16 123.16	4.92 54.66	2.88 - 3.00 32.73				174.64
E148	8.81 122.00	4.30 57.59	2.25 30.48	2.54 36.00			176.34
G149	8.78 114.71	4.04 - 4.55 45.23					173.14
F150	7.35 111.28	5.20 56.52	2.57 - 3.55 38.65				172.65
T151	8.55 116.08	4.67 61.73	4.36 71.02	1.30 21.70			173.06
L152	9.51 130.54	5.31 54.07	2.03 42.42	1.49 26.12	1.02 25.56		175.97
Q153	10.00 127.14	5.19 54.34	2.05 - 2.17 31.57	2.32 35.57			171.13
E154	8.91 121.72	5.43 55.21	2.14 30.48	2.33 2.22 36.86			174.89
W155	9.85 125.99	5.56 55.89	2.85 3.38 32.39			10.30 130.53	177.02
V156	10.09 118.14	5.50 60.00	2.44 35.35	1.18 19.76	1.24 21.23		174.79
R157	8.05 126.73	5.47 56.06	2.11 1.19 28.76	1.68 1.70 29.22	3.34 3.37 42.40		177.42
S158	8.54 121.23	4.51 58.43	3.90 63.72				174.82
A159	8.65 126.48	4.59 52.81	1.63 19.31				178.06
S160	8.40 114.62	4.71 58.49	4.14 63.81				175.19
S161	8.41 118.19	4.57 57.23	4.09 63.87				173.97
R162	8.11 128.14	4.39 57.74	1.97 2.08 31.01	1.80 27.24	3.43 43.75		

REFERENCES

- (1) Clore, G. M.; Gronenborn, A. M. *Crit. Rev. Biochem. Mol. Biol.* **1989**, *24*, 479-564.
- (2) Wuthrich, K. *Methods Enzymol.* **1989**, *177*, 125-131.
- (3) Kainosho, M. *Nat. Struct. Biol.* **1997**, *4*, 858-861.
- (4) Clore, G. M.; Gronenborn, A. M. *Protein Eng.* **1987**, *1*, 275-288.
- (5) Reid, D. G. *Protein NMR Techniques*; Human Press: Totawa, NJ, 1997.
- (6) Szyperski, T.; Wider, G.; Bushweller, J. H.; Wuthrich, K. *J. Biomol. NMR* **1993**, *3*, 127-132.
- (7) Pegan, S.; Kwiatkowski, W.; Choe, S.; Riek, R. *J. Magn. Reson.* **2003**, *165*, 315-319.
- (8) Clore, G. M.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1991**, *30*, 2315-2323.
- (9) Roberts, G. C. K. *NMR of Macromolecules: A practical Approach*; Oxford Univeristy Press: New York, 1993.
- (10) Clore, G. M.; Gronenborn, A. M. *NMR of Proteins*; CRC Press: Boca Ratone, FL, 1993.
- (11) Clore, G. M.; Gronenborn, A. M. *Curr. Opin. Chem. Biol.* **1998**, *2*, 564-570.
- (12) Wagner, G.; Thanabal, V.; Stockman, B. J.; Peng, J. W.; Nirmala, N. R.; Hyberts, S. G.; Goldberg, M. S.; Detlefsen, D. J.; Clubb, R. T.; Adler, M. *Biopolym.* **1992**, *32*, 381-390.
- (13) Voehler, M. W.; Collier, G.; Young, J. K.; Stone, M. P.; Germann, M. W. *J. Magn. Reson.* **2006**, *183*, 102-109.
- (14) Harpen, M. D. *Med. Phys.* **1989**, *16*, 781-784.
- (15) de Swiet, T. M. *J. Magn. Reson.* **2005**, *174*, 331-334.

- (16) Kelly, A. E.; Ou, H. D.; Withers, R.; Dotsch, V. *J. Am. Chem. Soc.* **2002**, *124*, 12013-12019.
- (17) McNair, D. S. *J. Magn. Reson.* **1981**, *45*, 490-502.
- (18) Young, J. K.; Moskau, D.; Hicks, R. P. *Concepts Magn. Reson.* **1994**, *6*, 115-130.
- (19) Cavanagh, J.; Rance, M. *J. Magn. Reson.* **1992**, *96*, 670-678.
- (20) Grzesiek, S.; Anglister, J.; Bax, A. *J. Magn. Reson., Ser. B.* **1993**, *101*, 114-119.
- (21) Logan, T. M.; Olejniczak, E. T.; Xu, R. X.; Fesik, S. W. *J. Biomol. NMR* **1993**, *3*, 225-231.
- (22) Lyons, B. A.; Montelione, G. T. *J. Magn. Reson., Ser. B* **1993**, *101*, 206-209.
- (23) Kay, L. E.; Xu, G. Y.; Singer, A. U.; Muhandiram, D. R.; Forman-Kay, J. D. *J. Magn. Reson., Ser. B* **1993**, *101*, 333-337.
- (24) John Cavanagh, W. J. F., III, Arthur G. Palmer, Nicholas J. Skelton *Protein NMR Spectroscopy: Principles and Practice* Academic Press: New York, 1996.
- (25) Giesen, A. W.; Bae, L. C.; Barrett, C. L.; Chyba, J. A.; Chaykovsky, M. M.; Cheng, M. C.; Murray, J. H.; Oliver, E. J.; Sullivan, S. M.; Brown, J. M.; Homans, S. W. *J. Biomol. NMR* **2002**, *22*, 21-26.
- (26) Otting, G.; Ruckert, M.; Levitt, M. H.; Moshref, A. *J. Biomol. NMR* **2000**, *16*, 343-346.
- (27) Carlomagno, T.; Schwalbe, H.; Rexroth, A.; Sorensen, O. W.; Griesinger, C. *J. Magn. Reson.* **1998**, *135*, 216-226.
- (28) Andersson, P.; Nordstrand, K.; Sunnerhagen, M.; Liepinsh, E.; Turovskis, I.; Otting, G. *J. Biomol. NMR* **1998**, *11*, 445-450.
- (29) Buffington, L. A.; Garsky, V.; Rivier, J.; Gibbons, W. A. *Biophys. J.* **1983**, *41*, 299-304.
- (30) Brutscher, B. *J. Magn. Reson.* **2001**, *151*, 332-338.
- (31) Parella, T.; Belloc, J.; Sanchez-Ferrando, F. *Magn. Reson. Chem.* **2004**, *42*, 852-862.
- (32) Palmer, A. G., 3rd; Fairbrother, W. J.; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Biomol. NMR* **1992**, *2*, 103-108.
- (33) Bax, A.; Ikura, M. *J. Biomol. NMR* **1991**, *1*, 99-104.

- (34) Weisemann, R.; Ruterjans, H.; Bermel, W. *J. Biomol. NMR* **1993**, *3*, 113-120.
- (35) Dotsch, V.; Oswald, R. E.; Wagner, G. *J. Magn. Reson. B* **1996**, *110*, 107-111.
- (36) Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1992b**, *96*, 432-440.
- (37) Ikura, M.; Kay, L. E.; Bax, A. *Biochemistry* **199a**, *29*, 4659-4667.
- (38) Ikura, M.; Marion, D.; Kay, L. E.; Shih, H.; Krinsis, M.; Klee, C. B.; Bax, A. *Biochem. Pharmacol.* **1990**, *40*, 153-160.
- (39) Yamazaki, T.; Lee, W.; Revingtom, M.; Mattiello, D. L.; Dahlquist, F. W. *J. Am. Chem. Soc* **1994**, *116*, 6464-6465.
- (40) Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1992**, *99*, 201-207.
- (41) Muhandiram, D. R.; Kay, L. E. *J. Magn. Reson.* **1994**, *B103*, 203-216.
- (42) Meissner, A.; Sorensen, O. W. *J. Magn. Reson.* **2001**, *151*, 328-331.
- (43) Clubb, R. T.; Thenbal, V.; Wagner, G. *J. Magn. Reson.* **1992**, *97*, 213-217.
- (44) Kay, L. E.; Xu, G. Y.; Yamazaki, T. *J. Magn. Reson.* **1994**, *109*, 129-133.
- (45) Matsuo, H.; Li, H.; Wagner, G. *J. Magn. Reson.* **1996**, *B110*, 112-115.
- (46) Clubb, R. T.; Thanabal, V.; Wagner, G. *J. Biomol. NMR* **1992**, *2*, 203-210.
- (47) Ikura, M.; Kay, L. E.; Bax, A. *Biochemistry* **1990**, *29*, 4659-4667.
- (48) Lin, Y.; Wagner, G. *J. Biomol. NMR* **1999**, *15*, 227-239.
- (49) Logan, T. M.; Olejniczak, E. T.; Xu, R. X.; Fesik, S. W. *J. Biomol. NMR* **1993**, *3*, 225-231.
- (50) Leopold, M. F.; Urbauer, J. L.; Wand, A. *J. Mol. Biotechnol.* **1994**, *2*, 61-93.
- (51) Vuister, G. W.; Boelens, R.; Padilla, A.; Kleywegt, G. J.; Kaptein, R. *Biochemistry* **1990**, *29*, 1829-1839.
- (52) Bax, A.; Ikura, M.; Kay, L. E.; Barbato, G.; Spera, S. *Ciba Found Symp* **1991**, *161*, 108-119.
- (53) Venters, R. A.; Farmer, B. T., 2nd; Fierke, C. A.; Spicer, L. D. *J. Mol. Biol.* **1996**, *264*, 1101-1116.
- (54) Driscoll, P. C.; Clore, G. M.; Marion, D.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 3542-3556.

- (55) Marion, D.; Driscoll, P. C.; Kay, L. E.; Wingfield, P. T.; Bax, A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1989**, *28*, 6150-6156.
- (56) Zuiderweg, E. R.; Fesik, S. W. *Biochemistry* **1989**, *28*, 2387-2391.
- (57) Wang, H.; Zuiderweg, E. R. *J. Biomol. NMR* **1995**, *5*, 207-211.
- (58) Peti, W.; Griesinger, C.; Bermel, W. *J. Biomol. NMR* **2000**, *18*, 199-205.
- (59) Kupce, E.; Schmidt, P.; Rance, M.; Wagner, G. *J. Magn. Reson.* **1998**, *135*, 361-367.
- (60) Kupce, E. *Methods Enzymol.* **2001**, *338*, 82-111.
- (61) Hadden, C. E. *Magn. Reson. Chem.* **2005**, *43*, 330-333.
- (62) Korzhnev, D. M.; Skrynnikov, N. R.; Millet, O.; Torchia, D. A.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 10743-10753.
- (63) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274-293.
- (64) Powers, R.; Garrett, D. S.; March, C. J.; Frieden, E. A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry* **1992**, *31*, 4334-4346.
- (65) Clore, G. M.; Bax, A.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 8172-8184.

CHAPTER IV
NMR STRUCTURE CALCULATION OF *Hv*DHFR1 FROM
EXPERIMENTAL RESTRAINTS

Introduction

NMR data provide experimental constraints that allow the calculation of the three dimensional structure of a protein.¹⁻⁹ Without supplying experimental restraints, the molecular modeling calculation program would result in an infinite number of different conformers. The major experimental constraints derived from NMR are proton-proton distances, backbone dihedral angles and H-bonding information. The distance restraints are obtained through the 3D NOESY experiments.^{3,10-25} These experiments rely on through space interactions from which proton-proton distances are derived. The intensity of the NOE (nuclear Overhauser effect) cross peaks provides a means to assign the distances. The backbone dihedral angles can be calculated from five backbone chemical shift values (^{15}N , $^1\text{H}^\alpha$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'$) using the program TALOS.¹⁰ Identification of H-bonding is important because most of the secondary structure elements are characterized by their specific hydrogen bonding patterns. Lyophilisation is the process during which the protein sample is dried to powder form and then re-dissolved in D_2O for the NMR study. The protons involved in hydrogen bonding will be replaced by deuterium atoms and this process can be observed from peaks disappearing in 2D ^1H - ^{15}N -

HSQC spectrum. It is crucial to obtain as much experimental data as possible, since the quality of the NMR structures are highly dependent on the number of restraints.

Molecular Modeling: There are a few molecular modeling programs that have been developed to calculate NMR derived protein structure such as CNS,¹¹ XPLOR,¹² CYANA,¹³ AMBER,¹⁴ CHARM,¹⁵ GROMOS,¹⁶ DIANA,¹⁷ DIG-II,¹⁸ and DYANA,¹⁹ using experimental restraints. These programs rely on Distance Geometry (DG), restrained Molecular Dynamic (rMD), and Simulated Annealing (SA) methods.²⁰⁻²⁴ Two algorithms have been developed in the DG method: Variable Target Function (which is implemented by the program DIANA) and Metrics Matrix (which is implemented by the program DIG-II). The DG method is based on generating distance matrices for every pair of atoms from experimental restraints. During the refinement, these matrices are reduced to Cartesian coordinate arrangements where the structure of the protein is to be calculated. Molecular Dynamics and Simulated Annealing are two closely related algorithms that minimize the total energy (energy resulted from deviation of ideal covalent geometry, $E_{\text{empirical}}$, and the energy resulted from experimental restraints, E_{exp}) of the structure using potential energy functions.²⁵ The algorithms use two techniques: Cartesian Space or Torsion Angle Space which are employed by the programs CNS, XPLOR, CYANA, AMBER,^{26,27} and GROMOS. DG and SA algorithms use covalent structure information such as torsion angles, dihedral angles, bond distances, and attractive/repulsive *van der Waals* interactions in the structure calculation. The atoms are exposed to a gradient of force to oblige the ideal covalent geometry. The ideal geometric information (protein parameters) is embedded in libraries such as CHARMM,²⁸ which is

used by the programs.²⁹ Besides the empirical energy contribution, the energy resulting from experimental distance and dihedral angle restraints (E_{exp}) are also added to the total energy. Similar to the DG algorithms, in the SA, the calculation takes place in Cartesian space. Unlike the DG, which involves many heating and cooling steps to minimize the penalty terms, SA involves a single high temperature dynamic phase followed by a slow cooling to a final temperature to arrive at a low energy structure.³⁰ The supplied heat enables the molecule to overcome the local energy barriers during molecular dynamic simulations.

In the rMD method, NMR derived structures are calculated with supplemental experimental restraints that are embedded in the energy function. The $E_{\text{empirical}}$ term includes E_{bond} , E_{dihedr} , E_{angle} , E_{improper} , and $E_{\text{non-bonded}}$ potentials.³¹⁻³⁷ The empirical energy function can be extended as follows and is the calculation of the cost of the deviations from ideal covalent geometries.³⁸⁻⁴⁰

$$E_{\text{emp}} = \sum_{\text{bonds}} k_b (r - r_0)^2 + \sum_{\text{dihedral}} k_\phi (1 + \cos(n\phi + \delta))^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{improper}} k_\phi (r\phi - \delta)^2 + \sum_{\text{non-bonded}} k_{\text{repel}} (r\phi - \delta)^2 \quad [1]$$

where k_b , k_ϕ , k_θ , k_{repel} stands for force constants, r , ϕ , θ , δ are the resulted distance, dihedral angles, bond angles, and offset value for dihedral and angles; r_0 , ϕ_0 , θ_0 are the ideal bond length, dihedral angles, and bond angles.⁴¹ Similarly, the experimental energy function (a pseudo-potential energy term⁴²⁻⁴⁵) can be extended as follows and is the calculation of violations resulted from the experimental restraints:³¹⁻³⁷

$$E_{\text{exp}} = \sum_{\text{distance-restraints}} k_d (\Delta d)^2 + \sum_{\text{angle-restraint}} k_a (\Delta a)^2 \quad [2]$$

The potential energy function (or penalty function) is a square-well function which calculates the cost of energy resulting from the supplied ^1H - ^1H - distance restraints and dihedral angle restraints. The potential function is given by³¹⁻³⁷

$$E_{\text{distance-restraints}} = \begin{cases} k_{\text{NOE}} [(r_{ij} - (r_{ij})^{\text{upper}})]^2 & r_{ij} > (r_{ij})^{\text{upper}} \\ 0 & (r_{ij})^{\text{lower}} < r_{ij} < (r_{ij})^{\text{upper}} \\ k_{\text{NOE}} [(r_{ij} - (r_{ij})^{\text{lower}})]^2 & r_{ij} < (r_{ij})^{\text{lower}} \end{cases} \quad [3]$$

where the k_{NOE} is the force constant of the NOE potential function (also sometimes referred to as the target function), r_{ij} is the distance between two atoms, $(r_{ij})^{\text{upper}}$ and $(r_{ij})^{\text{lower}}$ are the upper and lower limit distance restraints defined experimentally. The potential function for the dihedral angle restraints is given by³¹⁻³⁷

$$E_{\text{dihedral-angle}} = \begin{cases} k_{\phi} [(\phi - (\phi)^{\text{upper}})]^2 & (\phi)^{\text{upper}} > \phi \\ 0 & (\phi)^{\text{lower}} < \phi < (\phi)^{\text{upper}} \\ k_{\phi} [((\phi)^{\text{lower}} - \phi)]^2 & (\phi)^{\text{lower}} < \phi < (\phi)^{\text{upper}} \end{cases} \quad [4]$$

In this study, the molecular dynamic program CNS (Crystallography and NMR System) Version 1.1¹¹ was employed in the structure calculations of the enzyme *HvDHFR1*. It is a versatile program which uses various algorithms for structure calculations. The calculated NMR structures are reported as an ensemble of structure since the input restraints are defined in a range. Structure calculations continued until a family of structures is obtained which are both low in energy and best fit the experimental data. PROCHECK was used to check the structure quality based on how the dihedral angles fit to a Ramachandran plot.

Secondary Structure Prediction Using Chemical Shift Values

In three dimensional structure calculations, it is possible to estimate the secondary structure elements by using backbone chemical shift values.⁴⁶⁻⁵² It is known that there is a correlation between the backbone chemical shift deviations and the local protein conformation.^{38,50,53} The deviation from standard random coil chemical shift values of $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, C' and $^1\text{H}^\alpha$ is known to correlate with the presence of secondary structure elements.^{50, 54-57} Helices are predicted from the negative deviation of the $^1\text{H}^\alpha$ and $^{13}\text{C}^\beta$ and the positive deviation of $^{13}\text{C}^\alpha$ and $^{13}\text{C}'$ chemical shift values. The opposite trend is true for a β -strand. Figure 4.1 shows the plot of deviations of the assigned chemical shift values ($^1\text{H}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}^\alpha$, and $^{13}\text{C}'$) of *Hv*DHFR1 from random coil chemical shift values. Generally, four or more deviations in the same direction warrant presence of secondary structure, with the requirement that three of the four chemical indices be in agreement. Termination points were defined by junctions where the chemical shift deviation changed direction or became null. The resulting chemical shift deviations, which are plotted and shown in Figure 4.1, indicate the presence of four helices and eight β -strands in the protein structure.

Secondary structure prediction for an individual residue can also be accomplished using the program TALOS (Torsion Angle Likelihood Obtained from Shift and sequence similarity).¹⁰ The program is a database program that searches for similar chemical shift values and residue types in the database. This database contains ^{15}N , $^1\text{H}^\alpha$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'$ chemical shift values for 20 proteins as well as the high resolution X-ray crystallography structures. The ^{15}N , $^1\text{H}^\alpha$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'$ chemical shift values of *Hv*DHFR1 are used as

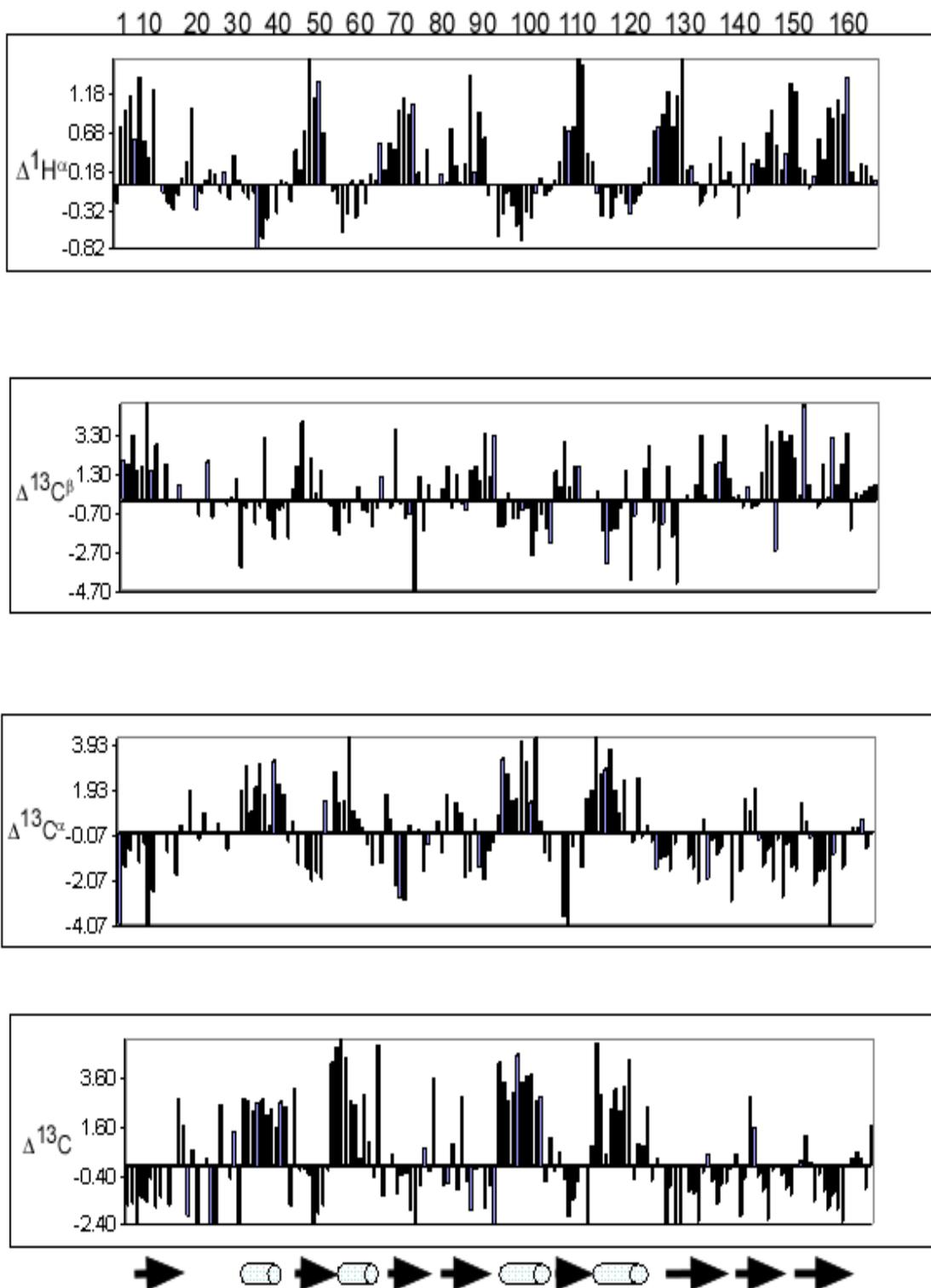


Figure 4.1. Deviation of H^α , C^β , C^α , and C' values from the random coil values are shown. The arrows indicate the presence of β -strands and tubes indicates helices.

input in the program TALOS. The program searches 10 times for the closest matches of three consecutive residues in the sequence and provides a prediction for the secondary structure location as well as the phi (Φ) and psi (Ψ) dihedral angles for individual residues. A dihedral angle can be defined as the bond rotation between two set of atoms along the peptide backbone. The Φ dihedral angle defines the bond rotation about $^{15}\text{N}^{\text{H}}$ and $^{13}\text{C}^{\alpha}$ atoms and is assigned from the four atoms $^{13}\text{C}_{i-1}$, $^{15}\text{N}_i$, $^{13}\text{C}^{\alpha}_i$, and $^{13}\text{C}'$. The Ψ dihedral angle defines the bond rotation about $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}'$ and is assigned from $^{15}\text{N}_i$, $^{13}\text{C}^{\alpha}_i$, $^{13}\text{C}'_i$, and $^{15}\text{N}_{i+1}$ atoms. The secondary structure prediction is plotted in a Ramachadran diagram for an individual residue where the allowed region for α -helices and β -strands are defined from characteristic dihedral angles. An average dihedral angle for Φ and Ψ is reported based on the distribution. Any residue in a β -strand falls in the upper left region ($\Phi = -120^\circ$, $\Psi = 120^\circ$), a residue in right handed helices ($\Phi = -57^\circ$, $\Psi = -47^\circ$) falls in the bottom region and any residue with left handed helices ($\Phi = 75^\circ$, $\Psi = 150^\circ$), falls in the right circled regions. The display shows the most energetically favored regions for the secondary structure elements with the dihedral angles. Figure 4.2 shows the display of TALOS for residue L118. The secondary structure for this residue was predicted to be in a β -strand with a Φ angle of $105 \pm 9^\circ$ and Ψ angle of $132 \pm 14^\circ$. Φ and Ψ dihedral angles are another important restraints in the structure calculations of proteins. In order to determine a high resolution NMR structure, the NOE restraints should also be supplemented by dihedral angle restraints. A total of 185 Φ (93) and Ψ (92) dihedral angles were obtained from TALOS and used in the structure calculations. The dihedral angles were only used for the residue that fit best on a value or displayed a good

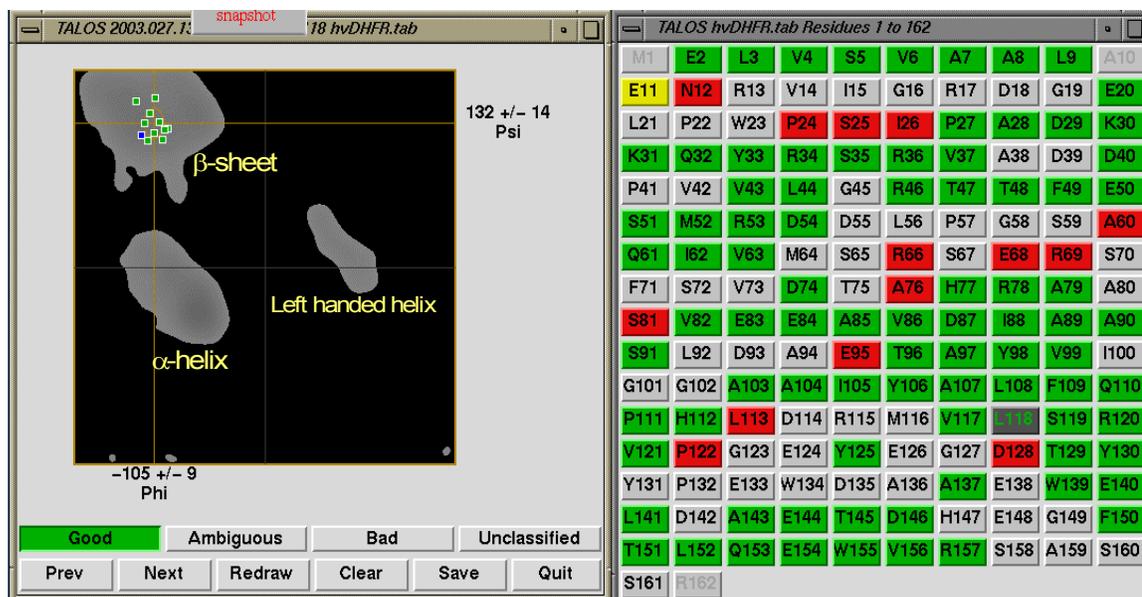


Figure 4.2. The screen of TALOS program illustrating predicted both the phi and psi angles for the residue L118 and secondary structure region. The X-ray structure of *HvDHFR1* was used as reference shown in blue dot, and the predicted region from the chemical shift values matching the database values are shown in green dot.

agreement among the predicted angles. The predicted dihedral angles were restricted within the intervals of $\pm 20^\circ$ in the structure calculation.

Distance Restraints Through 3D NOESY Experiments

The major experimental constraints used in NMR structure calculation are ^1H - ^1H distances which are derived from NOESY data. The distance restraints between proton pairs are obtained from NOE cross-peaks. NOEs provide important structural information for the secondary and tertiary structure of a protein.^{22,58} NOEs result from the through-space dipolar coupling of nuclei at a distance of $<5\text{-}6 \text{ \AA}$. Nuclei that are involved in dipolar coupling relax through cross relaxation which can be measured in the NOESY experiments.⁵⁹⁻⁷⁰ This process yields cross-peaks between the interacting nuclei

in the NOESY spectrum at which the intensity (I) or the cross-peak volume is a function of the distance between the nuclei and can be stated as; $I \propto f(\tau_m) \langle r^{-6} \rangle$ where r is the average distance between the interacting nuclei and the $f(\tau_m)$ is the correlation function of protein and depend on correlation time, τ_m .^{21,71,72}

NOE distance restraints are obtained from the 3D ^{15}N and ^{13}C - filtered NOESY-HSQC experiments. The first 3D NOESY experiment used in this study is ^{15}N -NOESY-HSQC. This experiment provides through-space correlations from the amide protons ($^1\text{H}^{\text{N}}$) of a residue to all inter- and intraresidue protons that are in close proximity. The first part of the experiment (NOESY) establishes through-space correlations between proton nuclei while the second part, HSQC, filters the proton magnetization between ^1H - ^{15}N pairs. The most important factor in NOESY experiments is setting the right mixing time. A short mixing time may result in the loss of NOE intensity while a long mixing time may result in spin-diffusion which generates NOEs for nuclei $<5\text{-}6 \text{ \AA}$. For small to medium size proteins, the optimum mixing time is typically set to 100-200ms. In this study, the mixing time in 3D NOESY experiments was set to 150 ms.

Chemical shift information is used to assign the NOE cross-peaks in the $^{15}\text{N}^{\text{H}}$ plane of the ^{15}N -NOESY-HSQC experiment. Figure 4.3 shows three strips selected from different $^{15}\text{N}^{\text{H}}$ planes of the ^{15}N -NOESY-HSQC experiment. The first strip corresponds to L118, the second to A85 and the last one to L9. The diagonal peak in each strip (indicated by blue arrows) refers to the amide proton ($^1\text{H}^{\text{N}}$) of the residue. The cross-peaks that correlated with the diagonal peaks are the NOEs resulted from through-space interactions of amide protons to other protons close in space. The assignments of the cross peaks are shown on the left. The first strip shows the NOEs for the $^1\text{H}^{\text{N}}$ of L118

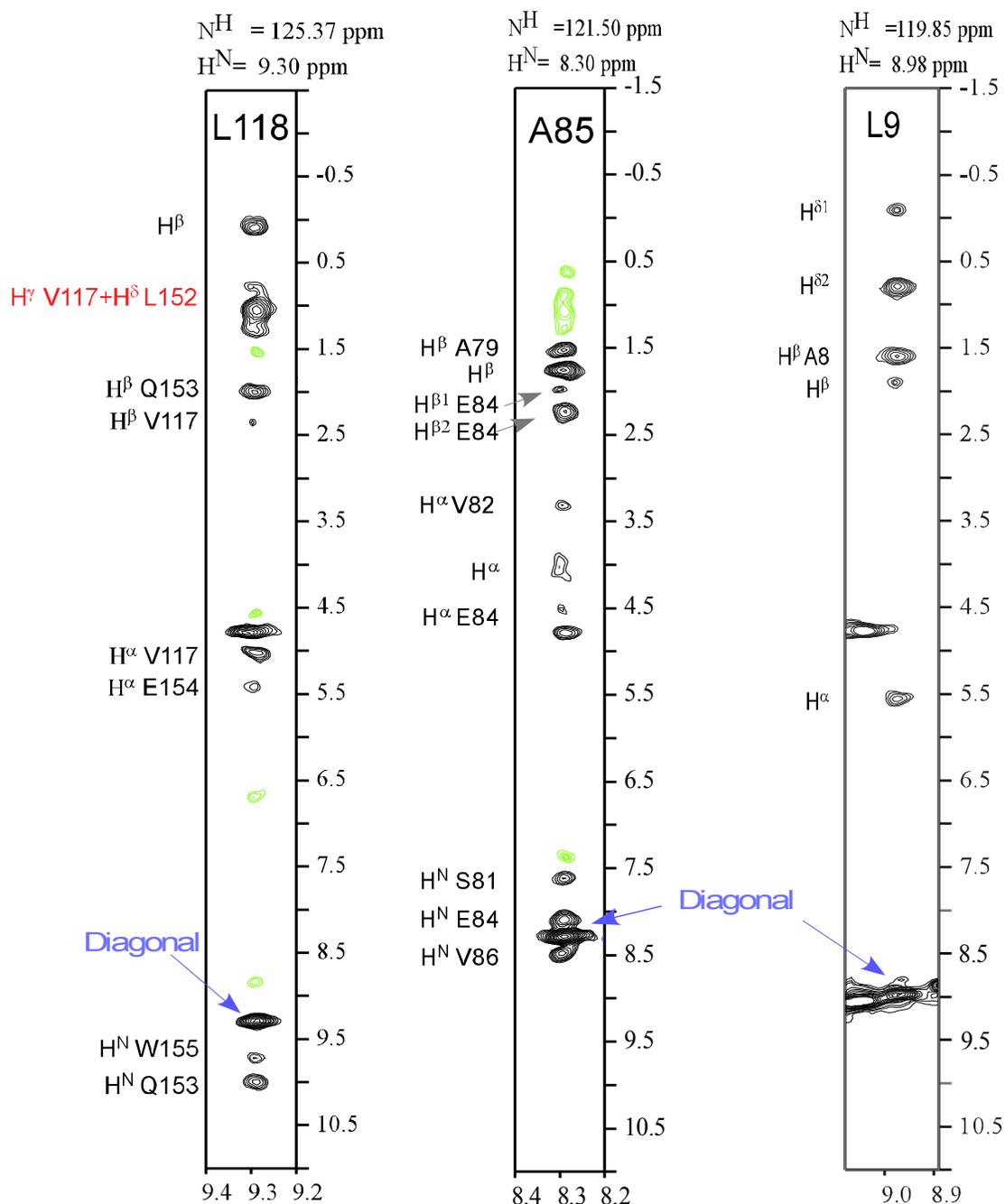


Figure 4.3. Three strips from different ^{15}N planes of the ^{15}N -NOESY-HSQC experiment belonging to residues L115 (from a β -strand region), A85 (from a α -helical region), and L9 (from a random coil region) are shown. The diagonal peaks in each strip are indicated by arrows which correspond to $^1\text{H}^{\text{N}}$ of the residue. The peaks that are not in the plane or are not assigned are highlighted in green. The peaks that correlate different protons chemical shifts are labeled in red.

which is in a β -strand. A residue in a β -strand is known to correlate strongly to the $^1\text{H}^\alpha$ of the previous residue. The $^1\text{H}^\text{N}$ (9.30 ppm) of L118 shows a strong correlation to the H^α (5.03ppm) of the V117. The $^1\text{H}^\text{N}$ of a residue in a β -strand can show correlations to $^1\text{H}^\text{N}$, $^1\text{H}^\alpha$, and side chain protons of residues in other β -strands across the β -sheet. The $^1\text{H}^\text{N}$ of the L118 shows correlations to $^{15}\text{N}^\text{H}$ of the Q153 and $^1\text{H}^\alpha$ of E154, two residues in another β -strand. Due to chemical shift degeneracy of nuclei belonging to different residues, some peaks may correlate to two or more nuclei such as the cross peak at 0.88 ppm correlating to both $^1\text{H}^\gamma$ of V117 and $^1\text{H}^\delta$ of L152 (highlighted in red). In such cases, the restraints were not employed in initial structure calculations due to ambiguities but as the calculations proceeded it was input if the calculated structure supported the correlation. The second strip, belonging to the $^1\text{H}^\text{N}$ of A85, shows an example from a helical region. The $^1\text{H}^\text{N}$ of a residue in a helical region is known to strongly correlate to the $^1\text{H}^\text{N}$ of residue $i-1$ and residue $i+1$. The $^1\text{H}^\text{N}$ of A85 (8.30 ppm) shows a strong correlation to $^1\text{H}^\text{N}$ of E84 and (8.07 ppm) as well as V86 (8.49 ppm). The $^1\text{H}^\text{N}$ of a residue in a helical region can correlate to $^1\text{H}^\alpha$ of the residues $i-1$, residue $i-2$, and residue $i-3$. The $^1\text{H}^\text{N}$ of A85 shows correlations to $^1\text{H}^\alpha$ of V82 and (8.07 ppm) and E84 (8.49 ppm). As seen from the strip, typically a residue in a helical region only shows medium-range (± 4 residue) correlations. The last strip belongs to $^1\text{H}^\text{N}$ of L9, is an example from a loop region where the minimal distance restraints are obtained. The $^1\text{H}^\text{N}$ of L9 (8.98 ppm) shows only correlations to its own and previous residue.

A total of 944 proton-proton distance restraints were defined via the 3D ^{15}N -NOESY-HSQC experiment. The global secondary and tertiary structure of a protein

can be calculated via the restraints from the ^{15}N -NOESY-HSQC experiment. However, these restraints may not be sufficient to fold the protein correctly due to lack of distance information defining the core of the protein. As a rule of thumb, more than 10 distance restraints per residue is ideal for NMR structure calculations.^{73,74} The number of distance restraints obtained from the ^{15}N -NOESY-HSQC experiment was 944 which result in 6.2 restraints per residue. In order to calculate a good converged structure more distance restraints are required.

Additional NOE restraints are obtained from the 3D ^{13}C -NOESY-HSQC experiment which provides side chain inter-proton distances. The experiment supplies essential distance restraints to fold the protein correctly. The majority of the side chain restraints would reflect the hydrophobic core. The hydrophobic groups tend to bury in the core of the protein from which the distance restraints assist to fold the protein correctly in the molecular dynamic calculations. The principle of the experiment is similar to 3D ^{15}N -NOESY-HSQC. The first part of the experiment, NOESY, establish through space correlations among the protons while the second part, HSQC, filters the correlations to ^1H - ^{13}C pairs. Figure 4.4 shows a strip from the ^{13}C -NOESY-HSQC experiment at a ^{13}C plane of 16.14 ppm which has been assigned to $^{13}\text{C}^\beta$ of A89. The diagonal peak (colored in blue) at 1.40 ppm corresponds to $^1\text{H}^\beta$ chemical shift value of methyl protons of A89. The methyl group protons of A89 show correlations to other side chain methyl and methylene protons belonging to the residues: I88, I62, A97, A90, P41, and L92. The assignments of the cross peaks are shown on the left. As seen from the strips, the experiment provides long range proton-proton distances. The methyl group region in this experiment has severe overlap due to degeneracy of chemical shift values.

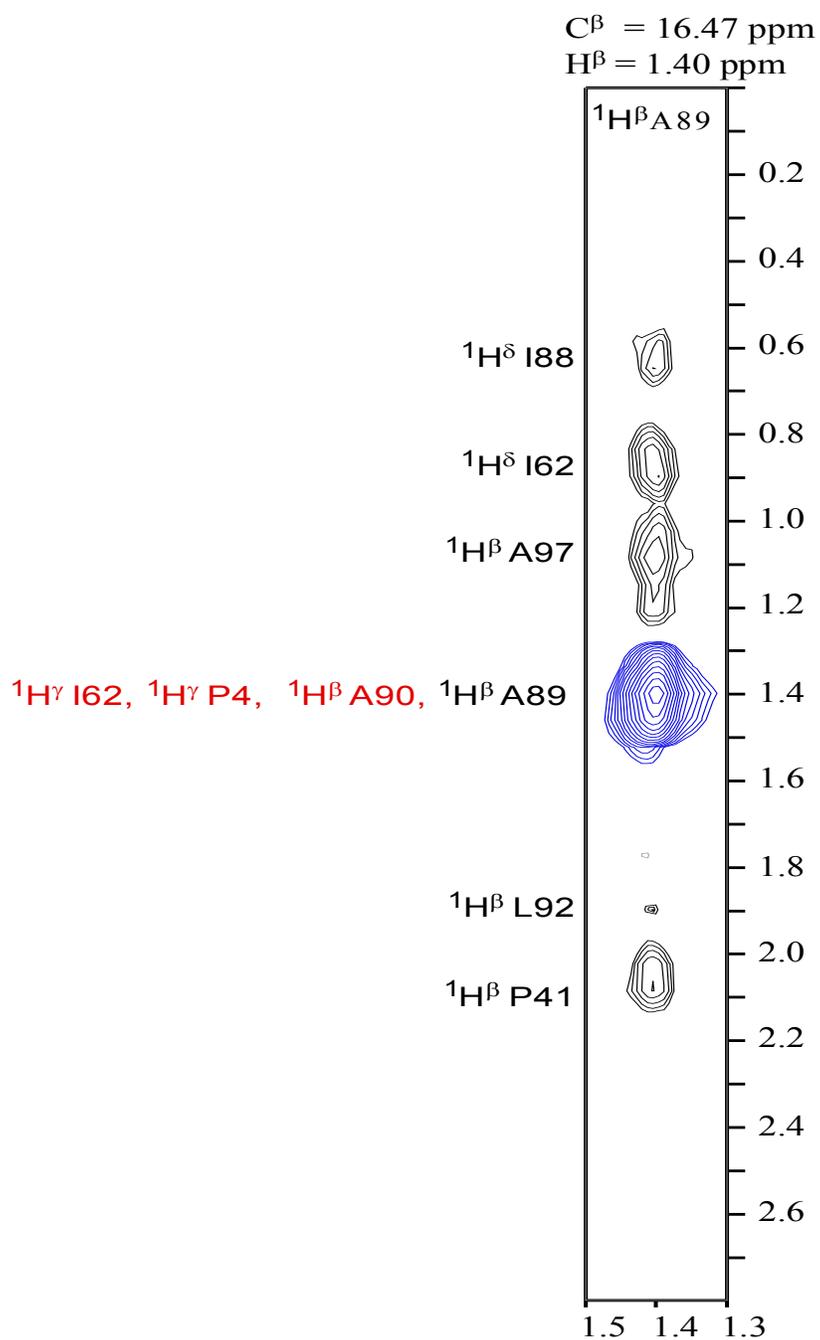


Figure 4.4. A strip from the ^{13}C -NOESY-HSQC experiment belonging to $^1\text{H}^\beta$ of A89 at 1.40 ppm is shown. The diagonal peak belonging to methyl group protons (highlighted in blue) correlates to other side chain protons of different residues. The assignments in red indicate the peak overlaps. The ^{13}C -NOESY-HSQC experiment provides long-range proton restraints.

The $^1\text{H}^\beta$ correlation of A90 and $^1\text{H}^\gamma$ of P41 overlap with the diagonal peak. These restraints were not used in structure calculation since the peak intensity can not be assigned specifically. A total of 427 side chain proton-proton distance restraints were defined via the 3D ^{13}C -NOESY-HSQC which raised the restraint to residue ratio to 8.5. The detail and category of distance restraints are given in Table 4.2. A total of 1556 derived restraints from the ^{15}N -NOESY-HSQC, ^{13}C -NOESY-HSQC, and TALOS were employed in the CNS structure calculation. The number of restraint per residue is 9.6. The NOE peak intensities (or peak volumes) are quantified to elucidate the proton-proton distance information. The intensities are quantified into distance limits by using the strong, medium, weak, and very weak method described by Clore et al.²¹ The peaks with intensity (volume) > 10 were defined as strong, 10-5 were defined as medium, <5 weak and low peak intensities were defined as very weak. The peak intensity categories were used to classify the upper and lower bounds of the proton-proton distances; strong (1.8 - 2.7 Å), medium (1.8 – 3.7 Å), weak (1.8 - 4.0 Å), and very weak (1.8 - 4.5 Å). The lowest distance range represents the sum of the *van der Waals* radius of two atoms.

Table 4.1. The statistics of the experimental distance restraints obtained from the ^{15}N -NOESY-HSQC and ^{13}C -NOESY-HSQC used in structure calculation by CNS

distance constraints	
no. of total restraints	1371
no. of intra residue (i=j)	335
no. of inter residue ((i-j)=1,(i+j)=1)	393
no. of medium (1<(i-j)≤5)	211
no. of long ((i-j)>5)	432
no. of hydrogen bond *	24

*Added through the last few calculations.

It has been known that there is a correlation between the NOE intensity and presence of the secondary structure elements. A motif in a protein sequence can be predicted from the NOE correlations.^{53, 75, 76} A α -helical region can be predicted when the following NOEs are observed: $d_{NN}(i, i+1)$: strong, $d_{NN}(i, i+2)$: very weak, $d_{NN}(i, i+3)$: medium, $d_{\alpha N}(i, i+1)$: weak, $d_{\alpha N}(i, i+3)$: medium, and $d_{\beta N}(i, i+3)$: medium.^{53, 75, 76} A β -strand can be predicted from $d_{NN}(i, i+1)$: weak, $d_{\alpha N}(i, i+1)$: strong, and $d_{N\alpha}(i, i)$ medium.^{53, 75, 76} The NOE correlations and the predicted secondary structures for the T75-E125 region of *Hv*DHFR1 are shown in Figure 4.5. The bar height or color specifies the

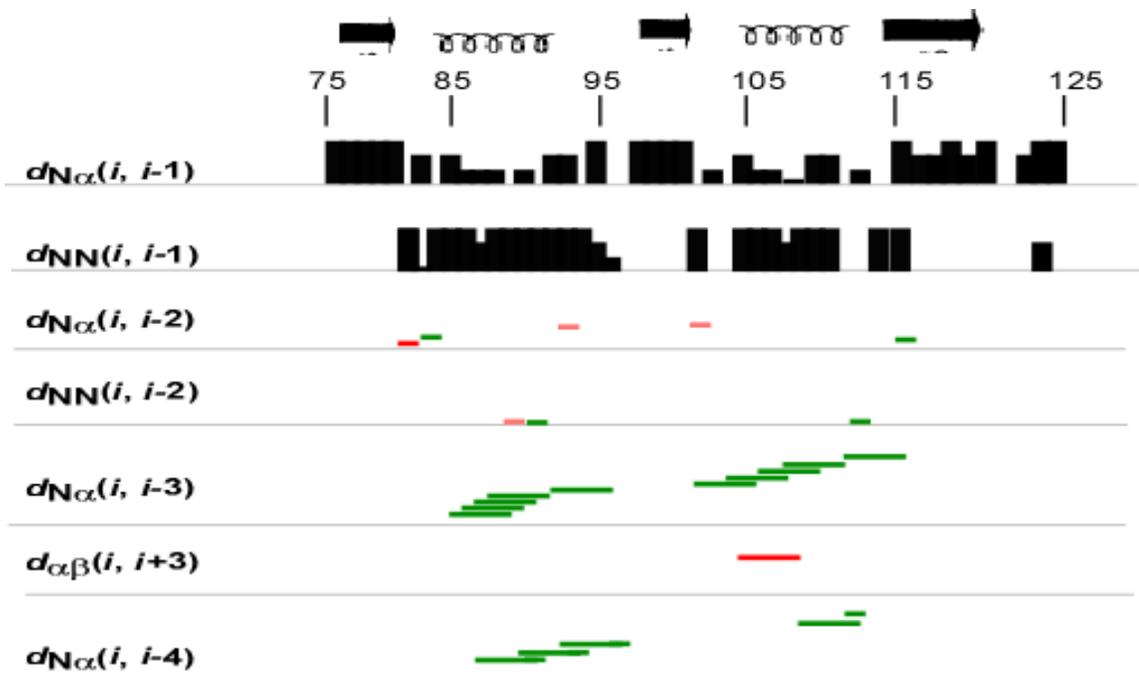


Figure 4.5. The figure illustrates the identification of the secondary structure elements from the analysis of short range NOE data in T75-Y125 region.^{53,75,76} The bar height specify the intensity of the observed NOE peak and length specify the range of correlations. Similarly, the strength of long range interactions are showed by colored lines where red indicates strongest and yellow indicates the weakest interactions.

strength of correlation and the bar length specifies the range of the correlation between residues. The analysis of short and medium range NOE peaks from ^{15}N -NOESY-HSQC revealed the presence of 8 β -strands and 4 helices. The resulting secondary structure elements from NOE data are shown on the top of the figure where the spring indicates the helical regions and an arrow indicates the β -strands. Further, by analyzing the β -strand regions, the long range ^1H - ^1H NOE cross peaks revealed that the 8 β -strands form a central β -sheet from specific NOE patterns, Figure 4.6. The long range $^1\text{H}^{\text{N}}$ - $^1\text{H}^{\text{N}}$ and $^1\text{H}^{\text{N}}$ - $^1\text{H}^{\alpha}$ NOE correlations across the β -strands indicate the presence of interactions between residues in the β -strands.

Structure Calculation Using CNS

In this study, the CNS (Crystallography and NMR System) program was used for the NMR structure calculations. Commonly, the algorithms implemented with this software are Distance Geometry (DG) combined with Simulated Annealing (SA), Simulated Annealing in Cartesian Space, and Simulated Annealing using Torsion Angle Dynamics (TAD).^{77,78} It has been reported that the last combination is more successful in the structure calculation of biological macromolecules.⁷⁸ In this combination, Φ and Ψ dihedral angles are restricted to those allowed for each amino acid (degree of freedom). Figure 4.7 shows the SA protocol used in CNS and steps followed during the structure calculation of *Hv*DHFR1. The proteins are folded toward the supplied experimental restraints during the annealing process (cooling) under the effect of gradient force constant.^{20,22,41} In the SA protocol used in this study, initially the generated structures were selected by a randomization process and were subjected to an increasing

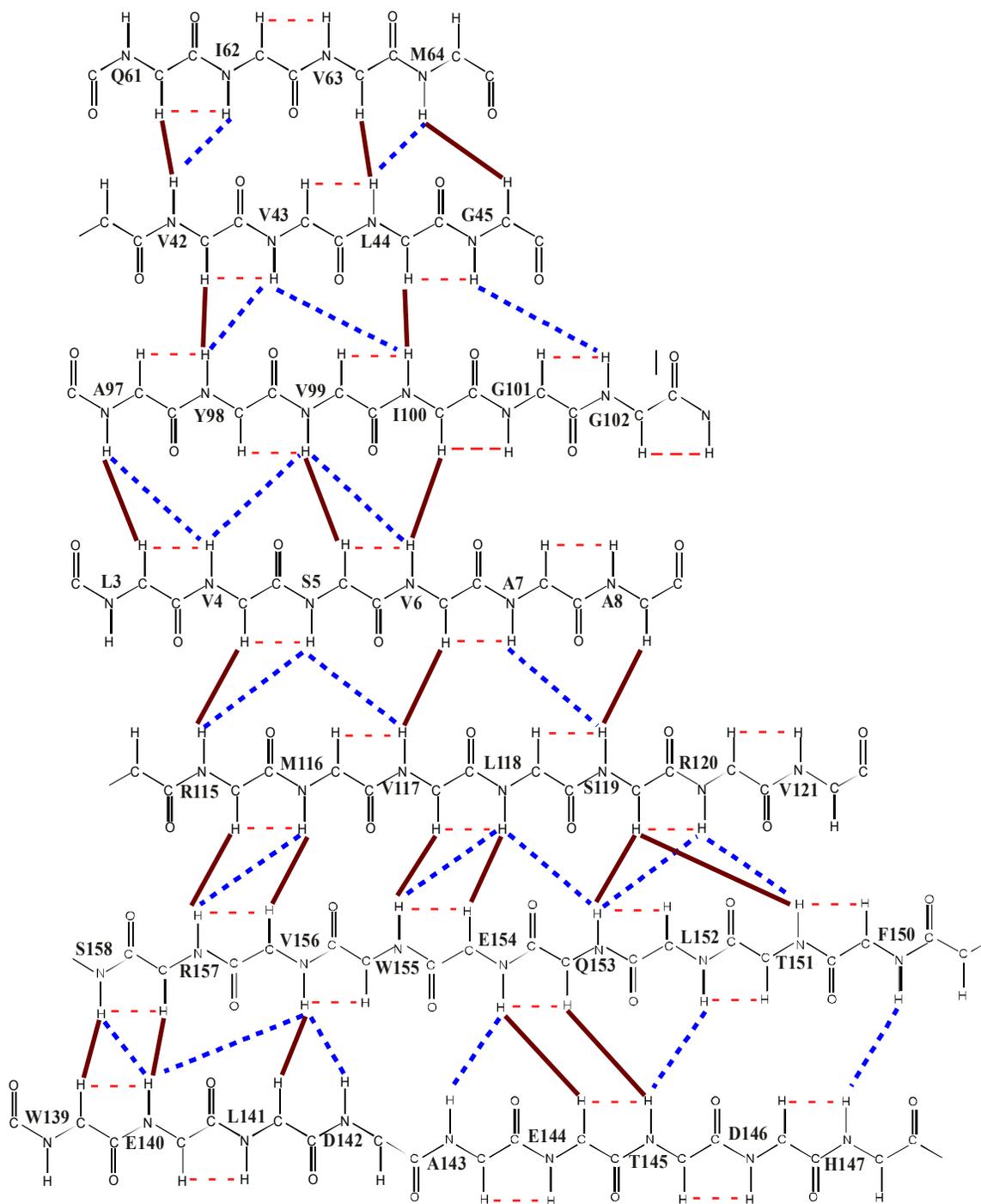


Figure 4.6. The β -sheet is characterized by long range H^N-H^N (dashed lines) and H^N-H^α (solid lines) correlations. The $d_{\alpha N}(i, i+1)$ correlations along the β -strands are shown by the dashed red lines.

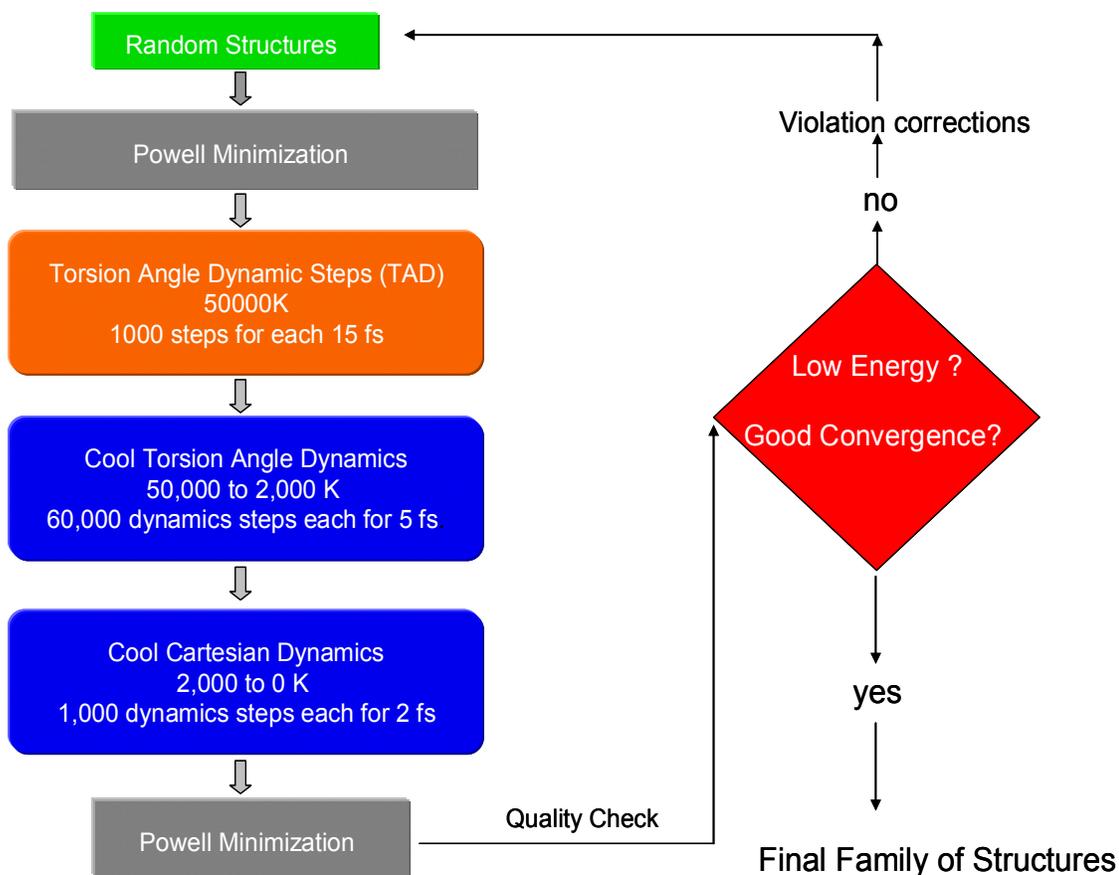


Figure 4.7. The structure calculation steps and TAD phase used in the CNS program are shown.

temperature with 1000 steps, each of which is 15 femtosecond to a final temperature of 50000 K. The high temperature provides the required velocities for atoms to overcome the local energy barriers and facilitates the molecule to reach the global minimum energy conformation during the cooling (annealing) steps. Typically during the initial phase of the SA, the force constants are set to low values (to successfully overcome the energy barriers and are proportional to kT) to maintain the experimental restraints and are gradually increased as the temperature decrease.³⁶ The force constants can be considered to be the scale factors to maintain the experimental restraints during the structure

simulations. The force constant for NOE restraints is in the range of 30-50 kcal mol⁻¹Å⁻² and for dihedral angles it is in the range of 5-10 kcal mol⁻¹rad⁻².³⁶ During the cooling process, the force constants were ramped to those values starting from unit values. The covalent structure parameters of amino acids in torsion angle space are set to standard values for bond length, bond angles, planarity, etc. The program uses a library that contains all the required covalent information of the amino acids.

As the structure calculations proceeded, resulted violations were fixed in the next rounds of calculations. A violation can be defined as the deviation of restraints between the defined range and that result in the structure. In the first round of structure calculations of *Hv*DHFR1, only unambiguous NOEs and dihedral restraints were employed. In subsequent calculations, the distance restraint range was moved to the next upper category for any NOE that was violated by more than 0.5 Å. Similarly, the variation in any dihedral angle was increased by 5°. As structural refinement proceeded, many of the ambiguous NOEs were clarified and used in the subsequent structure calculations unless it contradicted the generated structure. This iterative approach was repeated ~60 times. Hydrogen bonds were added (based on previous calculations) in secondary structure regions for the last few rounds of calculations. Lyophilized *Hv*DHFR1 cannot be redissolved in ²H₂O; therefore H/D exchange experiments were not obtained. The structure calculations were continued and finalized until no NOE distance was violated by >0.5 Å, no dihedral angle violated by >10°, and low energies were obtained. Obtaining a good convergence among the generated structures is another important criterion. The degree of the good convergence is a measure of the precision of experimental restraints used in structure calculations.^{74,79-81} This is achieved by

calculating the root mean square (rms) deviation of the backbone and heavy atom coordinates among the calculated NMR structures. This process also measures how similar structures are reproduced from the supplied restraints from the molecular modeling program. The rms deviations were achieved by superimposing atomic coordinates of the generated structures by the program MOLMOL.⁸² This program computes the pairwise rmsd values for each structure with respect to the rest of the structures. A low rmsd value for backbone and heavy atoms superimposition indicates the accomplishment of good convergence. Once all violations were eliminated, one hundred structures were generated for final analysis. The presence of a *cis*-peptide bond between two glycine residues has been identified in other DHFR enzymes by X-ray crystallography.⁸³ This *cis* peptide geometry was maintained by defining a restraint of 0° for the dihedral angle from C^α G101 to C^α G102, with an energy constant of 5 kcal mole⁻¹rad⁻². This constraint caused an improper violation in the structures output from CNS. The resultant structures were subjected to a further energy minimization due to high improper energy from imposing *cis* peptide geometry between two glycine residues. A fifty step conjugated minimization was performed on each structure using the program InsightII (Version 2005, Accelrys Inc, San Diego, CA) in the Discover Module (Version 2005, Accelrys). The InsightII software was used for graphical visualization.

The 20 lowest energy structures for *Hv*DHFR1 are shown in Figure 4.8 and the statistics are given in Table 4.2. The resulted penalties due to deviation from the ideal bond lengths, angles, and improper torsion angles are listed in Table 4.3. The final family of structures exhibits a good convergence with an rms deviation of 0.83±0.27 Å for all backbone atoms (C^α, N^H, C^γ) and 0.37±0.12 Å for secondary structure regions.

The heavy atoms (all atoms except hydrogen) superimposition in backbone and secondary structure regions resulted in an rms deviation of 1.52 ± 0.26 and 0.82 ± 11 , respectively. The total average energy for a structure is calculated as 1640 kcal/mol which includes the energy contributions from both empirical and experimental but not electrostatic contributions. The average number of distance restraints per residue in the secondary region is 11 and that for the loop regions is only 5. The low number of distance restraints in the loop regions resulted in relatively poor convergence (the fuzzy region in the structures) and these regions account for 25% of the number of the amino acid in the structure. The last few residues of C-terminus (A159-R162) resulted in random structure. The atomic coordinates of the 20 structures and the experimental restraints have been deposited in the RCSB Protein Data Bank under the code of 2ITH. A representative NMR structure is shown in Figure 4.9.

The calculated NMR structures reveal that four helices are packed across a central β -sheet with eight β -strands in the order of $\beta_1\alpha_1\beta_2\alpha_2\beta_3\beta_4\alpha_3\beta_5\alpha_4\beta_6\beta_7\beta_8$. The helical regions are α_1 : 27-37, α_2 : 46-52, α_3 : 83-92, α_4 : 105-112 and β -strands regions are β_1 : 3-9, β_2 : 42-44, β_3 : 61-64, β_4 : 76-79, β_5 : 97-99, β_6 : 115-120, β_7 : 139-147, β_8 : 151-157. The eight β -strands form a central β -sheet.

The structure quality of the generated structures were analyzed by the program PROCHECK.⁸⁴ Figure 4.10A and Figure 4.10B show the Ramachandran plot obtained from program PROCHECK. Figure 4.10A represents the Ramachandran plot for a representative NMR structure for clarity of view and Figure 4.10B represents the

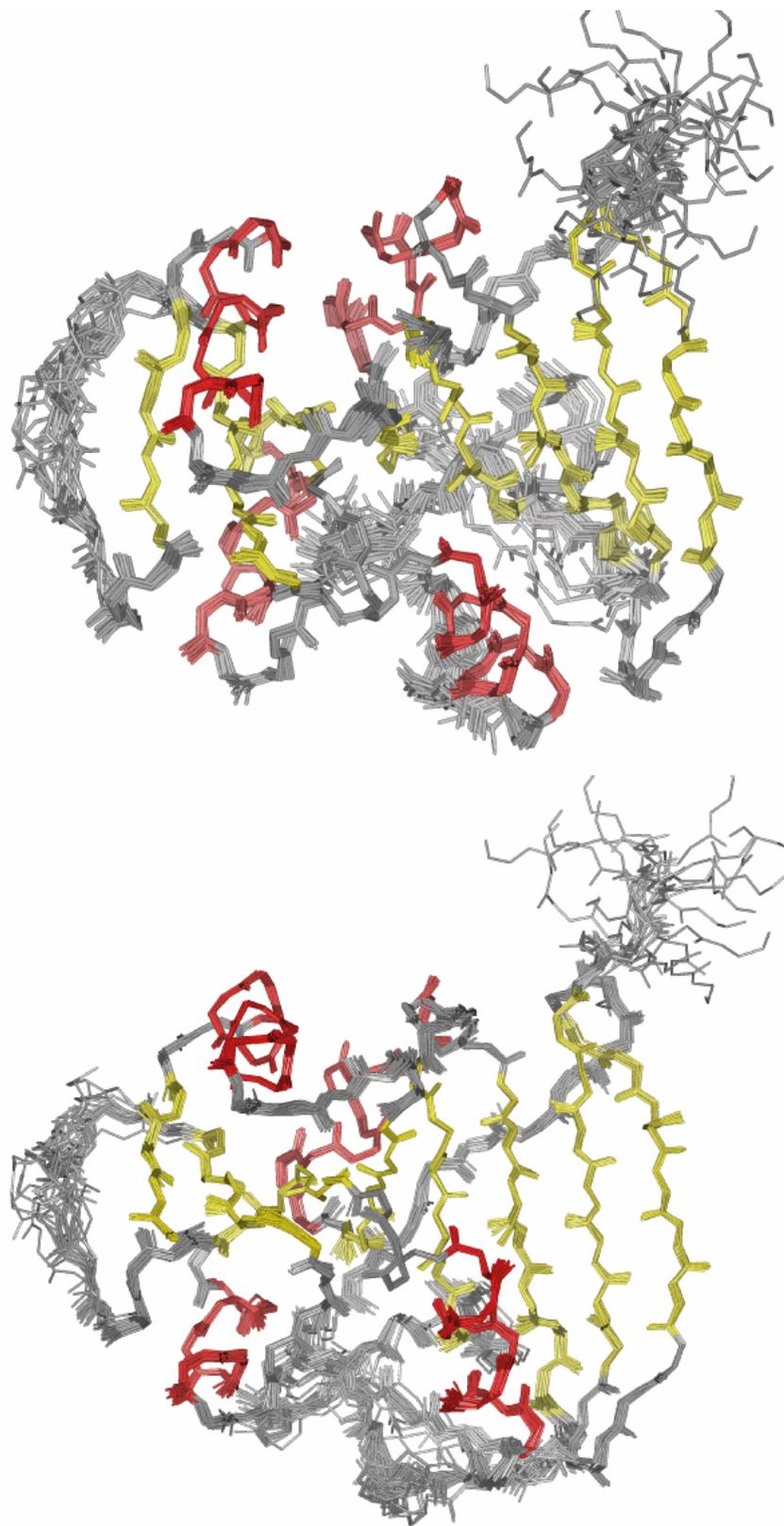


Figure 4.8. The backbone superimposition of the 20 lowest-energy structures (left) and 180° rotated form (right) of *HvdHFR1* generated using the CNS is shown. The helical regions are highlighted in red and β -strands are in yellows.

Table 4.2. Statistics of the 20 lowest-energy structures calculated by CNS

energy (kcal/mole)		
E _{total}		1689 ± 33
E _{bond}		134.6 ± 3.0
E _{angle}		399.2 ± 6.4
E _{improper}		67.4 ± 2.2
E _{NOE}		659.6 ± 12.6
E _{vdw}		379.9 ± 6.9
rmsds from experimental restraints		
bonds (Å)		0.074 ± 0.001
angles (°)		0.731 ± 0.150
impropers (°)		0.576 ± 0.013
NOEs (Å)		0.083 ± 0.0010
dihedrals (°)		1.053 ± 0.0310
PROCHECK		
% residues in most favorable region		73.7 ± 1.72
% residues in additionally allowed region		22.5 ± 2.70
% residues in generously allowed region		2.9 ± 1.48
% residues in disallowed region		0.8 ± 0.8
coordinate superimpose		
backbone (Å) (2-157)		0.83 ± 0.27
backbone heavy (Å)		1.52 ± 0.26
secondary backbone (Å)		0.37 ± 0.12
secondary backbone heavy (Å)		0.82 ± 0.11

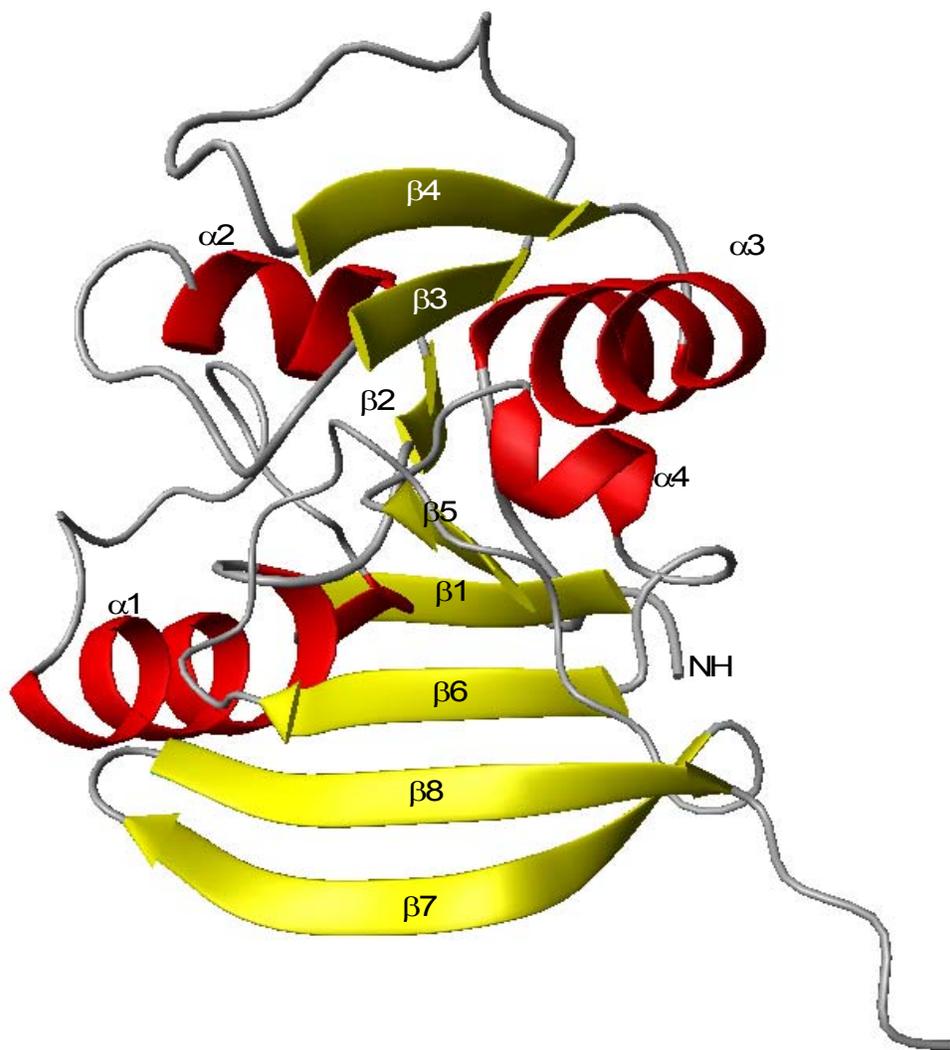


Figure 4.9. A representative NMR derived structure of *H_v*DHFR1 is shown in ribbon diagram. The helices and b-strands are numbered in the order of sequence. The structure is generated by the program MOLMOL.⁸²

result from 20 structures submitted to the Protein Data Bank. The helical and β -strand regions display specific Φ and Ψ dihedral angles. The dark red, yellow and pink regions represent the most favorable, allowed region, and the additionally allowed regions, respectively. Overall 96.2% of the residues of *HvDHFR1* are found in allowed and additionally allowed regions, see Table 4.2, which suggest well-calculated structures.

Comparisons of NMR derived HvDHFR1 Structure to Crystal HvDHFR1 and EcDHFR Structures: The overall fold of the NMR and crystal (1vdr⁸⁵) structures are similar with a rms deviation of 1.35 Å, Figure 4.11. Two different segments of the NMR structures did show a deviation from the crystal structure. First, there was a deviation in the secondary structure at the N-terminal residues of β 3 (S59-M64). The ¹H^N of A60 shows a strong NOE to both the ¹H^N of S59 and the ¹H^N of G58. Similarly, a strong NOE between the ¹H^N of S59 and the ¹H^N of G58 was observed in this region; however, these distances were not supported by the crystal structure. This divergence is considered to arise from the flexibility of the loop α 2- β 3. The difference in the secondary structure at the N-terminal residues of β 3 (59-64), causes a slight kink in the overall structure. The adenosine binding domain of the structure has a secondary backbone superimpose of 0.83 Å while the major is 0.66 Å. If either end of the molecule is superimposed it causes the whole other end of the molecule to deviate by as much as 2.5 Å. The deviation between the two structures in the N-terminus residues of β 3 is shown in Figure 4.12.

Secondly, there was a difference in the type of β -turn connecting β 7 and β 8 (H147-F150). The dihedral angles of the crystal structure define a Type-II β -turn⁸⁵

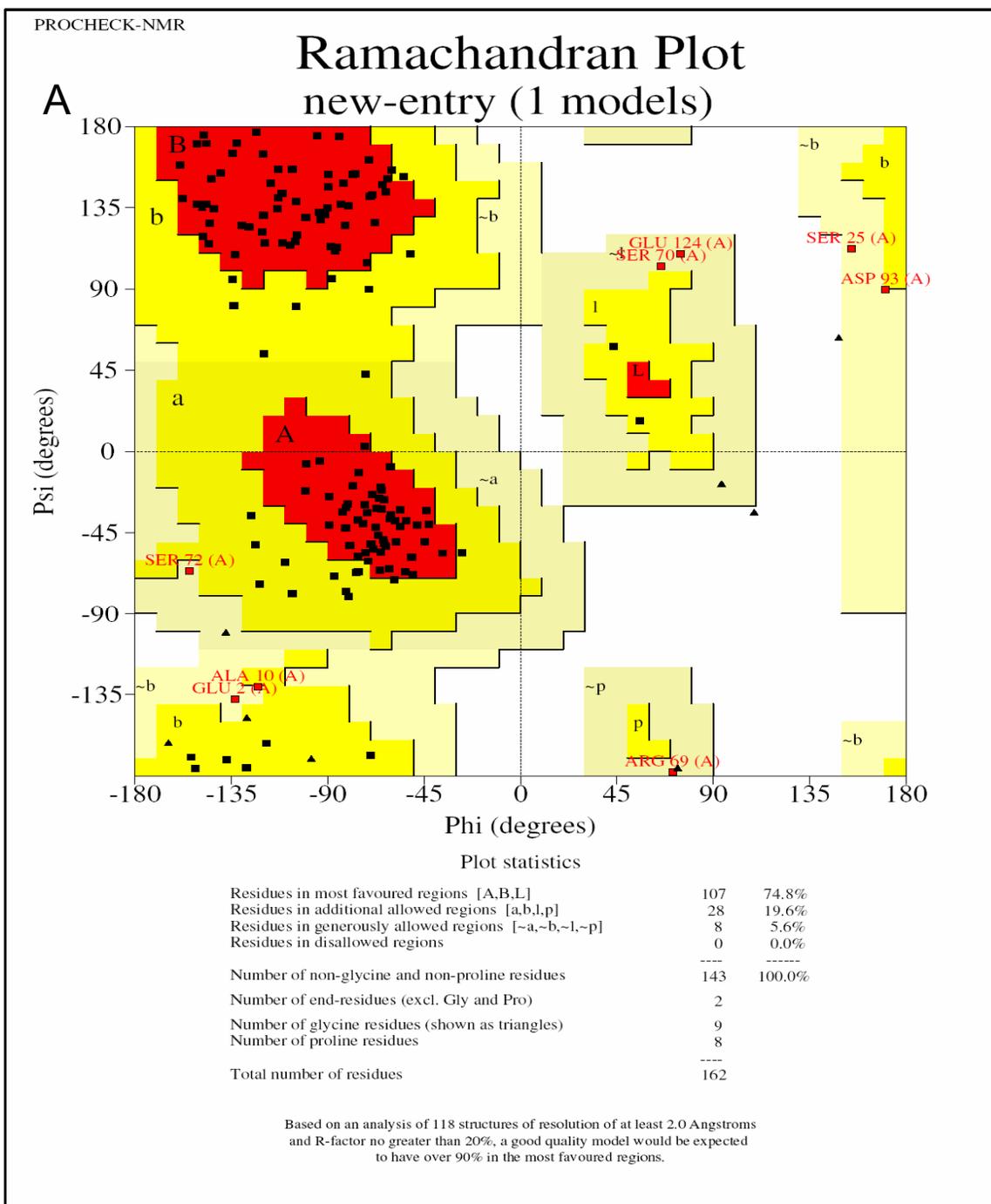


Figure 4.10A. The results of PROCHECK is shown for one of the submitted structures.

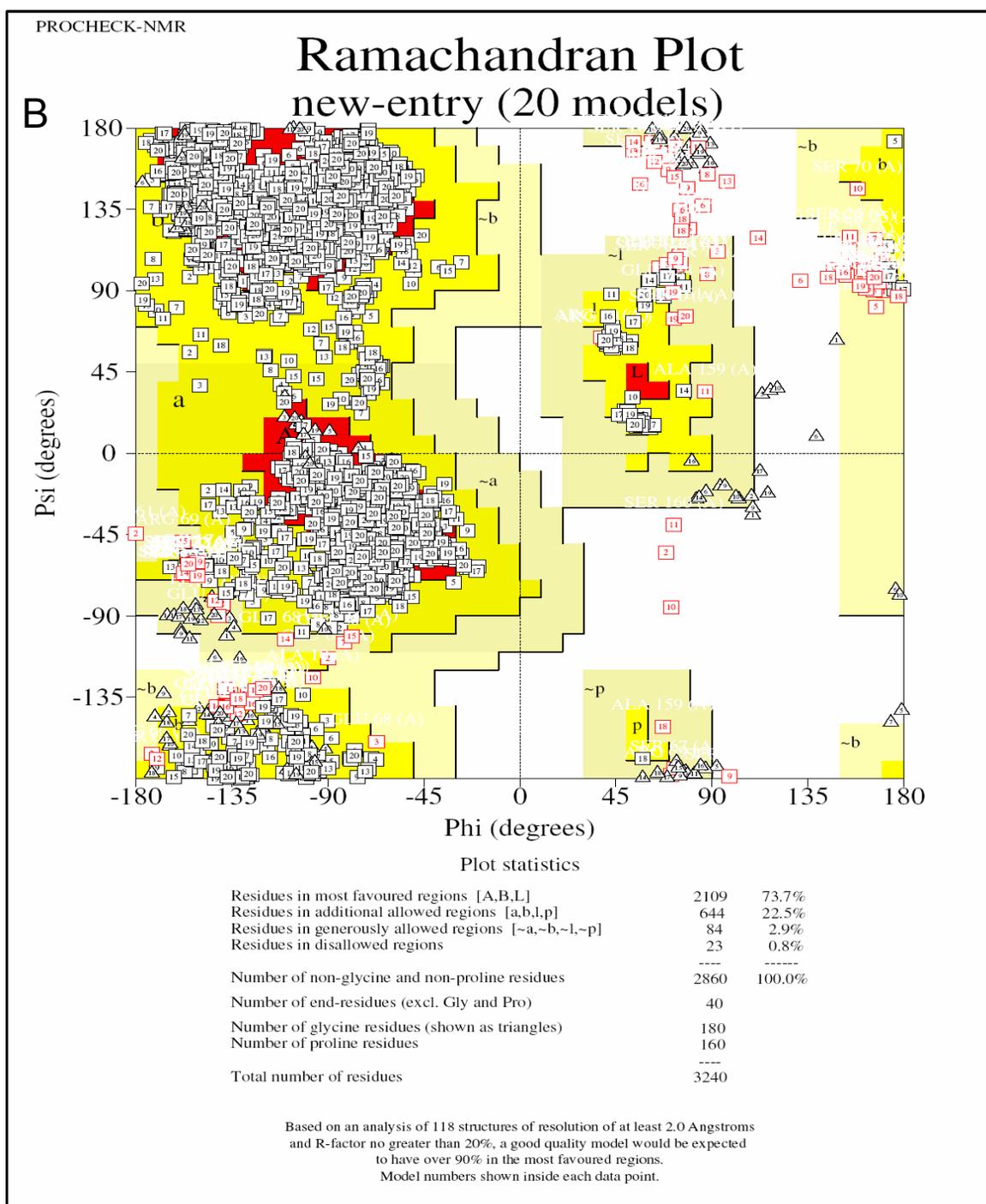


Figure 4.10B. The results of PROCHECK is shown for the 20 submitted structures.

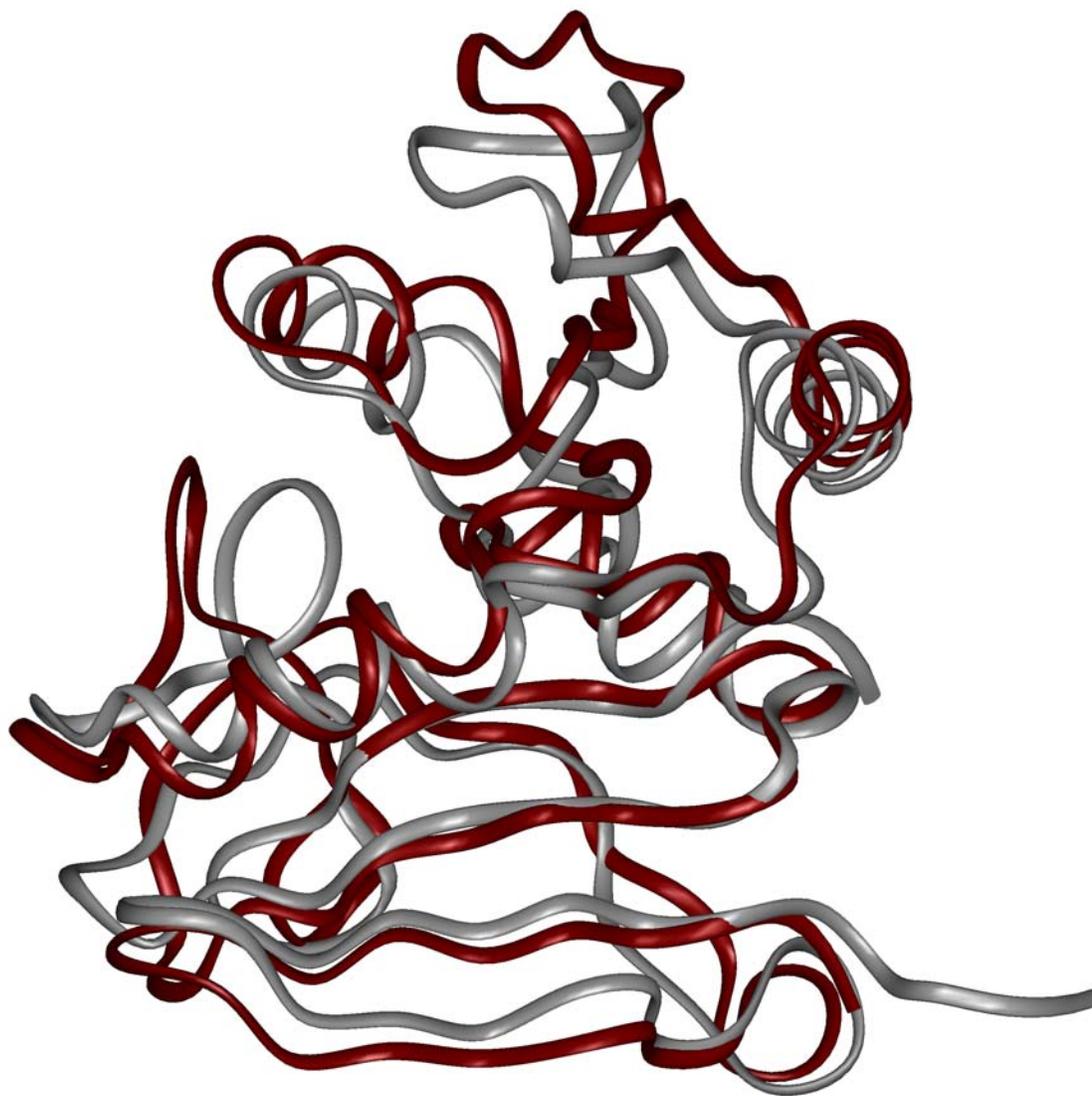


Figure 4.11. Superimpose of the X-ray crystallography structure (1vdr) of *HvDHFR1* (red ribbon) with the liquid state NMR structure (gray ribbon) resulting in a 1.35 Å rms deviation.

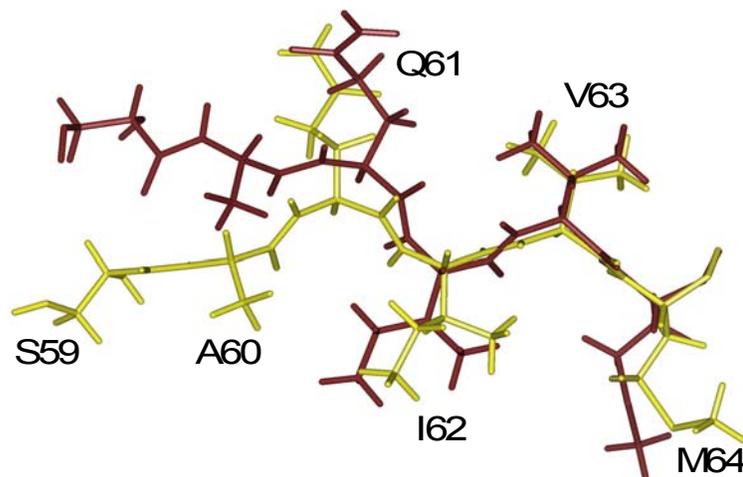


Figure 4.12. One variation between the NMR (yellow) and X-ray crystallography structures (maroon) is in the β_3 region. Key NOEs were obtained for S59 and A60 that are not supported by the crystal structure.

while the NMR structures define a Type-I β -turn. This Type-I β -turn is supported by the presence of an NOE from the $^1\text{H}^{\text{N}}$ of F150 to the $^1\text{H}^{\alpha}$ of H147 which should not be observable for a Type-II β -turn. This places this turn in closer proximity to the N-terminus of α_1 which is supported by an NOE from the $^1\text{H}^{\text{N}}$ of F150 to the δCH_3 of I26. It is possible that there may be a hydrogen bond between A28, a residue in α_1 helix and E148, a residue in this TypeI β -turn. The $^1\text{H}^{\text{N}}$ of A28 is at a distance of 2.7 Å to carboxylic group oxygen, Figure 4.13. The presence of a hydrogen bond is supported from the NOE between side chain groups of the two residues. An NOE between $^{\beta}\text{H}$ (methyl group) of A28 correlates to $^{\gamma}\text{H}$ (methenyl) group of E148. This TypeI β -turn is the main structural differences compared to other DHFR variants showed similar folding. There is a major sequence difference between *Ec*DHFR (159 aa) and *Hv*DHFR (162 aa) in this loop. This is a loop region in *Ec*DHFR which is termed the G-H loop (D142-

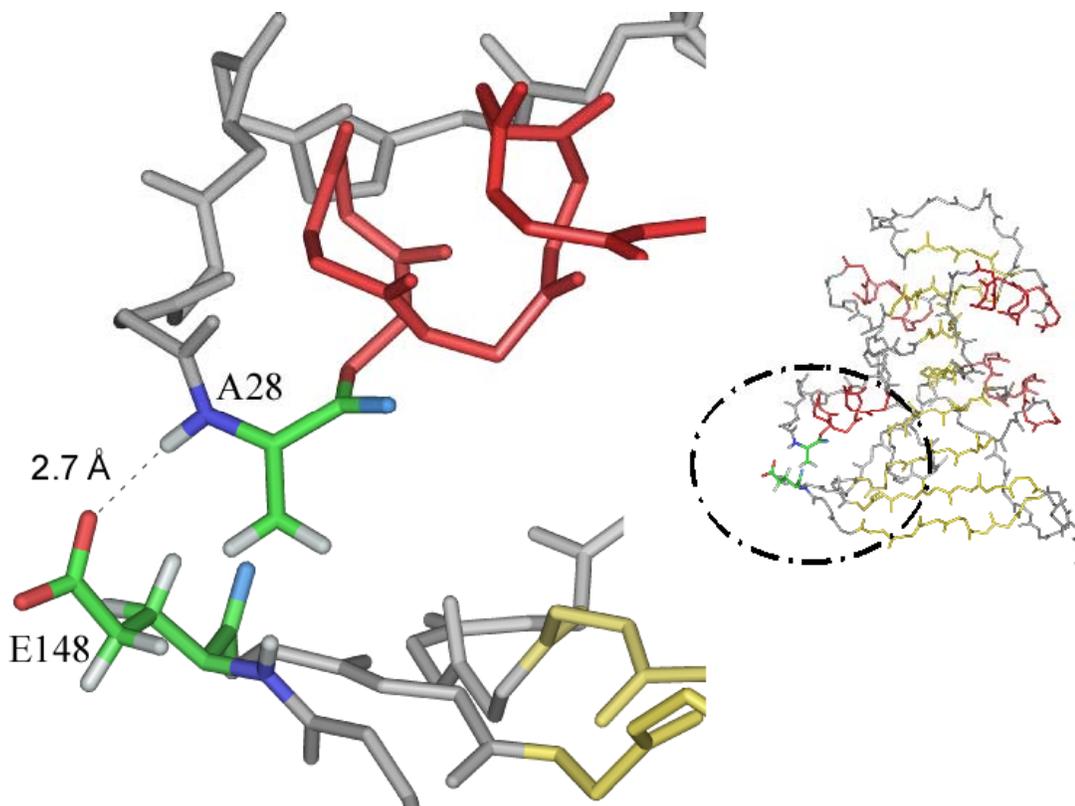


Figure 4.13. It is shown that H^N of A28 is at a close enough proximity to carboxylic oxygen of E148.

S150) and bears an important role in catalytic events.⁸⁶ In *Ec*DHFR, the residues in this loop form hydrogen bond, with the residues of M20 loop in the *occluded* conformation.⁸⁶ However, this loop is truncated in *Hv*DHFR1 by the deletion of six residues, (Figure 4.14) and there are no direct interactions with the L21 loop. This suggests (and as speculated by Pieper at al⁸⁵) that *Hv*DHFR1 may be using different loop motions than homologous mesophilic *Ec*DHFR. Based on the NMR derived structures, it can be speculated that the substrate binding may facilitate the motion of α 1 helix to form and deform a hydrogen bond. The NMR derived structures showed variations in the L21 and adenosine binding loop (ABL) regions. The loop regions are not defined well due to low NOE restraints derived from this region, Figure 4.15. The L21 loop which is the

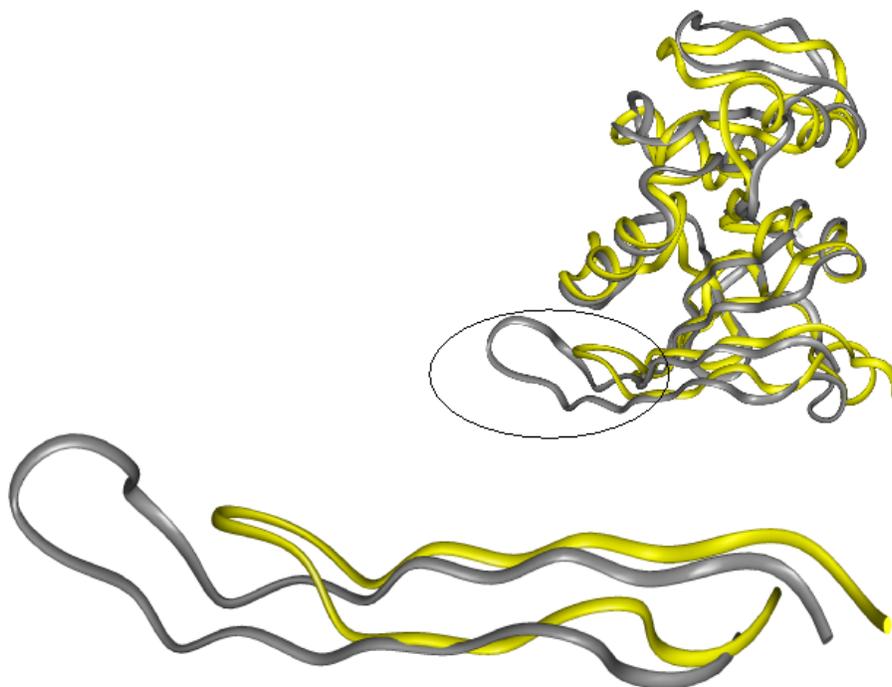


Figure 4.14. The superimposition of NMR derived structure of *Hv*DHFR1 (yellow) versus *Ec*DHFR (grey) (pdb code: 1RA9⁸⁶) is shown. The loop GH of *Ec*DHFR is truncated by deletion of six residues in *Hv*DHFR1.

equivalent of M20 in *Ec*DHFR was found as time averaging between the three conformational (*closed*, *open*, and *occluded*) states.⁸⁶

Conclusion

The liquid apo state NMR structure of *Hv*DHFR1 was calculated from the proton-proton distance restraints derived from ¹⁵N-HSQC-NOESY and ¹³C-HSQC-NOESY experiments in which the data have been collected at 3.5 M NaCl. The distance restraints are supplemented by the dihedral angle restraints which are derived from the program TALOS by using the backbone chemical shift values. The molecular dynamic program CNS has been used to calculate the NMR derived structures. The precision of the

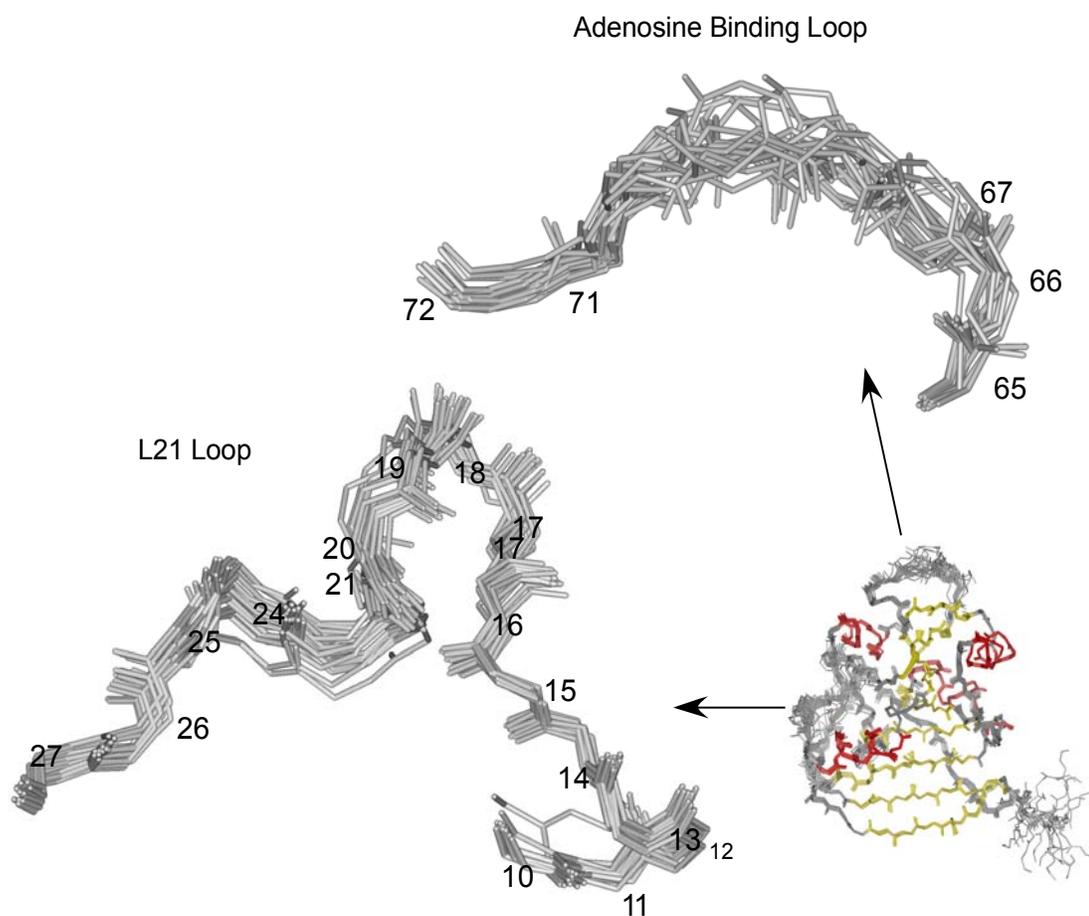


Figure 4.15. The L21 and adenosine binding loop regions of *HvDHFR1* from NMR derived study are shown. The loop conformation is a time average conformation of the *open*, *occluded*, and *closed* conformation of *EcDHFR*.⁸⁶ The adenosine binding loop displayed no specific conformation.

final family of structures are tested by measured rms deviations for backbone and side chain atoms. The low rmsd values of $0.83 \pm 0.27 \text{ \AA}$ and $1.52 \pm 0.26 \text{ \AA}$ for all backbone atoms and side chain atoms are also an indication of the accuracy of the experimental data. The overall protein folding, which is termed the Rossmann fold, was found to be similar to other DHFR variants. The NMR derived structure of *HvDHFR1* is comparable with that of the crystal structure with rms deviation of 1.35 \AA for backbone superimposition. The two structures are very similar in terms of the backbone fold and

secondary structure elements. The mainly differences are observed at loops region in the NMR structures which are considered to arise from the motion in this region. The noticeable difference in the N terminus of β_3 causes a slight kink in the overall structures as compared to the crystal structure. The G-H loop region of *Ec*DHFR is a β -turn in the NMR derived structure. In the crystal structure, it is defined as Type II β -turn, in the NMR derived structures, it is defined as Type I β -turn. The NMR derived structure is comparable to the crystal structure of the *Ec*DHFR:folate binary complex (1rx7) with an rms deviation of 1.03 Å for the secondary structure elements of the backbone atoms. In order to have a more complete understanding of the loop motions in *Hv*DHFR1, more structural studies must be carried out on the binary and ternary complexes.

The calculated NMR derived structures were used in relaxation studies. It has been speculated that the decrease of in activity of the enzyme as a response to lowered salt concentration may relate to changed inherent protein flexibility. This hypothesis was be tested by carrying out relaxation measurements at two salt concentrations.

REFERENCES

- (1) van Nuland, N. A.; Grotzinger, J.; Dijkstra, K.; Scheek, R. M.; Robillard, G. T. *Eur. J. Biochem.* **1992**, *210*, 881-891.
- (2) Nordstrand, K.; Ponstingl, H.; Holmgren, A.; Otting, G. *Eur. Biophys. J.* **1996**, *24*, 179-184.
- (3) van Nuland, N. A.; van Dijk, A. A.; Dijkstra, K.; van Hoesel, F. H.; Scheek, R. M.; Robillard, G. T. *Eur. J. Biochem.* **1992**, *203*, 483-491.
- (4) Uhrin, D.; Bramham, J.; Winder, S. J.; Barlow, P. N. *J. Biomol. NMR* **2000**, *18*, 253-259.
- (5) Fairbrother, W. J.; Gippert, G. P.; Reizer, J.; Saier, M. H., Jr.; Wright, P. E. *FEBS Lett.* **1992**, *296*, 148-152.
- (6) Ab, E.; Schuurman-Wolters, G.; Reizer, J.; Saier, M. H.; Dijkstra, K.; Scheek, R. M.; Robillard, G. T. *Protein Sci.* **1997**, *6*, 304-314.
- (7) Zangger, K.; Oberer, M.; Keller, W.; Sterk, H. *J. Magn. Reson.* **2003**, *160*, 97-106.
- (8) Moglich, A.; Koch, B.; Gronwald, W.; Hengstenberg, W.; Brunner, E.; Kalbitzer, H. R. *Eur. J. Biochem.* **2004**, *271*, 4815-4824.
- (9) Zieba, A.; Maslankiewicz, A.; Sitkowski, J. *Magn. Reson. Chem.* **2004**, *42*, 903-904.
- (10) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289-302.
- (11) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr.* **1998**, *54*, 905-921.
- (12) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65-73.
- (13) Guntert, P. *Methods Mol. Biol.* **2004**, *278*, 353-378.

- (14) [hppt://amber.scripps.edu](http://amber.scripps.edu) (accessed May 2007).
- (15) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *42*, 187-217.
- (16) Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719-1751.
- (17) Guntert, P.; Braun, W.; Wuthrich, K. *J. Mol. Biol.* **1991**, *217*.
- (18) Havel, T. F. *Prog. Biophys. Mol. Biol.* **1991**, *56*, 43-78.
- (19) Guntert, P.; Mumenthaler, C.; Wuthrich, K. *J. Mol. Biol.* **1997**, *273*, 283-298.
- (20) Nilges, M.; Gronenborn, A. M.; Brunger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27-38.
- (21) Clore, G. M.; Brunger, A. T.; Karplus, M.; Gronenborn, A. M. *J. Mol. Biol.* **1986**, *191*, 523-551.
- (22) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *239*, 129-136.
- (23) Clore, G. M.; Gronenborn, A. M.; Nilges, M.; Ryan, C. A. *Biochemistry* **1987**, *26*, 8012-8023.
- (24) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *229*, 317-324.
- (25) Karplus, M. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 255-266.
- (26) Wang, Z. X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. *J. Comput. Chem.* **2006**, *27*, 781-790.
- (27) Craft, J. W., Jr.; Legge, G. B. *J. Biomol. NMR* **2005**, *33*, 15-24.
- (28) Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L., 3rd *J. Comput. Chem.* **2004**, *25*, 1504-1514.
- (29) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 796-801.
- (30) Bruccoleri, R. E.; Karplus, M. *Biopolym.* **1990**, *29*, 1847-1862.
- (31) Cavanagh, J.; Fairbrothers, W.; Palmer III, A. G.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: New York, 1996.
- (32) Clore, G. M.; Gronenborn, A. M. *NMR of Proteins*; CRC Press: Boca Raton, FL, 1993.

- (33) John Cavanagh, W. J. F., III, Arthur G. Palmer, Nicholas J. Skelton *Protein NMR Spectroscopy: Principles and Practice* Academic Press: New York, 1996.
- (34) Reid, D. G. *Protein NMR Techniques*; Human Press: Totawa, NJ, 1997.
- (35) Roberts, G. C. K. *NMR of Macromolecules: A practical Approach*; Oxford Univeristy Press: New York, 1993.
- (36) Teng, Q. *Structural Biology: Practical NMR Application*; Springer Scinece + Business Media: New York, 2005.
- (37) van den Ven, F. J. M. *Multidimensional NMR in Liquids: Basic Principles and Experimental Methods*; Wiley-VCH, 1995.
- (38) Pardi, A.; Wagner, G.; Wuthrich, K. *Eur. J. Biochem.* **1983**, *137*, 445-454.
- (39) Wagner, G.; Thanabal, V.; Stockman, B. J.; Peng, J. W.; Nirmala, N. R.; Hyberts, S. G.; Goldberg, M. S.; Detlefsen, D. J.; Clubb, R. T.; Adler, M. *Biopolym.* **1992**, *32*, 381-390.
- (40) Wagner, G. *Nat. Struct. Biol.* **1997**, *4*, 841-844.
- (41) Guntert, P. *Q Rev. Biophys.* **1998**, *31*, 145-237.
- (42) Vendruscolo, M.; Dobson, C. M. *Philos. Transact. A Math. Phys. Eng. Sci.* **2005**, *363*, 433-450.
- (43) Abe, H.; Wako, H. *Phys. Rev. E Stat. Nonlin. Soft Matter. Phys.* **2006**, *74*, 11913.
- (44) Tavernelli, I.; Cotesta, S.; Di Iorio, E. E. *Biophys. J.* **2003**, *85*, 2641-2649.
- (45) Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128-132.
- (46) Morris, L. C.; Valafar, H.; Prestegard, J. H. *J. Biomol. NMR* **2004**, *29*, 1-9.
- (47) Wang, Y. *J. Biomol. NMR* **2004**, *30*, 233-244.
- (48) Wang, Y.; Jardetzky, O. *J. Biomol. NMR* **2004**, *28*, 327-340.
- (49) Beger, R. D.; Bolton, P. H. *J. Biomol. NMR* **1997**, *10*, 129-142.
- (50) Osapay, K.; Case, D. A. *J. Biomol. NMR* **1994**, *4*, 215-230.
- (51) Sharman, G. J.; Griffiths-Jones, S. R.; Jourdan, M.; Searle, M. S. *J. Am. Chem. Soc.* **2001**, *123*, 12318-12324.
- (52) Sibley, A. B.; Cosman, M.; Krishnan, V. V. *Biophys. J.* **2003**, *84*, 1223-1227.

- (53) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647-1651.
- (54) Wang, Y.; Jardetzky, O. *Protein Sci.* **2002**, *11*, 852-861.
- (55) Le, H.; Oldfield, E. *J. Biomol. NMR* **1994**, *4*, 341-348.
- (56) Mielke, S. P.; Krishnan, V. V. *J. Biomol. NMR* **2004**, *30*, 143-153.
- (57) Jee, J.; Guntert, P. *J. Struct. Funct. Genomics* **2003**, *4*, 179-189.
- (58) LeMaster, D. M.; Kay, L. E.; Brunger, A. T.; Prestegard, J. H. *FEBS Lett.* **1988**, *236*, 71-76.
- (59) Mumenthaler, C.; Braun, W. *J. Mol. Biol.* **1995**, *254*, 465-480.
- (60) Kraulis, P. J. *J. Mol. Biol.* **1994**, *243*, 696-718.
- (61) Xu, Y.; Jablonsky, M. J.; Jackson, P. L.; Braun, W.; Krishna, N. R. *J. Magn. Reson.* **2001**, *148*, 35-46.
- (62) Hare, B. J.; Wagner, G. *J. Biomol. NMR* **1999**, *15*, 103-113.
- (63) Huang, Y. J.; Tejero, R.; Powers, R.; Montelione, G. T. *Proteins* **2006**, *62*, 587-603.
- (64) Nilges, M. *J. Mol. Biol.* **1995**, *245*, 645-660.
- (65) Duggan, B. M.; Legge, G. B.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **2001**, *19*, 321-329.
- (66) Zhang, Q.; Chen, J.; Gozansky, E. K.; Zhu, F.; Jackson, P. L.; Gorenstein, D. G. *J. Magn. Reson. B* **1995**, *106*, 164-169.
- (67) Celda, B.; Biamonti, C.; Arnau, M. J.; Tejero, R.; Montelione, G. T. *J. Biomol. NMR* **1995**, *5*, 161-172.
- (68) Gorler, A.; Gronwald, W.; Neidig, K. P.; Kalbitzer, H. R. *J. Magn. Reson.* **1999**, *137*, 39-45.
- (69) Liu, H.; Spielmann, H. P.; Ulyanov, N. B.; Wemmer, D. E.; James, T. L. *J. Biomol. NMR* **1995**, *6*, 390-402.
- (70) Reid, D. G.; Saunders, M. R. *J. Biol. Chem.* **1989**, *264*, 2003-2012.
- (71) Oschkinat, H.; Griesinger, C.; Kraulis, P. J.; Sorensen, O. W.; Ernst, R. R.; Gronenborn, A. M.; Clore, G. M. *Nature* **1988**, *332*, 374-376.
- (72) Clore, G. M.; Gronenborn, A. M. *Protein Eng.* **1987**, *1*, 275-288.

- (73) Zagrovic, B.; van Gunsteren, W. F. *Proteins* **2006**, *63*, 210-218.
- (74) Clore, G. M.; Robien, M. A.; Gronenborn, A. M. *J. Mol. Biol.* **1993**, *231*, 82-102.
- (75) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363-392.
- (76) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171-180.
- (77) Brunger, A. T.; Adams, P. D.; Rice, L. M. *Structure* **1997**, *5*, 325-336.
- (78) Stein, E. G.; Rice, L. M.; Brunger, A. T. *J. Magn. Reson.* **1997**, *124*, 154-164.
- (79) Thomas, P. D.; Basus, V. J.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1237-1241.
- (80) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *Protein Sci.* **1996**, *5*, 1067-1080.
- (81) Shriver, J.; Edmondson, S. *Biochemistry* **1993**, *32*, 1610-1617.
- (82) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graph.* **1996**, *14*, 51-55, 29-32.
- (83) Zusman, T.; Rosenshine, I.; Boehm, G.; Jaenicke, R.; Leskiw, B.; Mevarech, M. *J. Biol. Chem.* **1989**, *264*, 18878-18883.
- (84) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S., and ; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283-291.
- (85) Pieper, U.; Kapadia, G.; Mevarech, M.; Herzberg, O. *Structure* **1998**, *6*, 75-88.
- (86) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586-603.

CHAPTER V
INVESTIGATION OF ENZYME FLEXIBILITY THROUGH NMR
RELAXATION MEASUREMENTS

Introduction

NMR Spectroscopy not only provides structural information it also provides a wealth of information about the internal motions of a protein.¹ These data are encoded onto the relaxation properties T_1 , T_2 and heteronuclear ^1H - ^{15}N NOE.² By analyzing the NMR relaxation properties of a protein, information about the timescale of motion on the picosecond to nanosecond range and microsecond to millisecond timescales can be extracted.^{2-5,6,7} A popular approach, Model-free formalism, can be used to characterize the amplitude of motion from the generalized order parameter (S^2) which is the measure of the spatial restriction to motion.^{2,5,6,8}

The activity of *Hv*DHFR1 is salt concentration dependent.⁹ By analyzing the relaxation data of *Hv*DHFR1 at different salt concentrations, it may be possible to obtain the relationship between the activity and flexibility of the enzyme. The enzyme *Hv*DHFR1 showed optimal enzymatic activity at a salt concentration of 3.5 M.⁹ As the salt concentration was decreased, there was a reduction in the activity. Spectroscopic studies (FL, CD) showed that the structure of the enzyme was stable at the range of 3.5-0.6 M salt concentration, below 0.6 M salt concentration the enzyme unfolds.⁹ Binding

studies (FL-based determination of K_D) of *Hv*DHFR1 exhibited an interesting affinity trend for substrate and cofactor as a function of salt concentration. There was a loss of affinity for DHF as salt concentration is lowered; however, the affinity for NADPH became stronger.⁹ However, these changes in affinity can not account for the dramatic decrease in activity. On the other hand, the homologous enzyme, *Ec*DHFR, revealed no significant affinity change for DHF or NADPH as salt concentration was increased to 1.0 M, the concentration at which 80-90% of activity was lost.^{9,10} The loss in the activity of *Ec*DHFR was speculated to be due to reduction in the internal flexibility over the salt concentration range. For *Ec*DHFR it has been shown that the loop motions are key factors that facilitate the catalytic activity.¹¹⁻¹³ It has been proposed that the activity lost for *Hv*DHFR1 may also be due to the changes in the inherent protein flexibility that facilitate the catalytic activity through either gaining excessive motion or losing motion.⁹ This may affect the binding events, halt the catalytic activity, and unfold the enzyme as the salt concentration is decreased.

In order to investigate the relation between activity loss and flexibility dependence as a function of salt concentration, ¹⁵N NMR spin relaxation studies of *Hv*DHFR1 at two different salt concentrations have been conducted. The 3.5 M salt concentration is the condition where the maximum activity was observed for this enzyme and 1.0 M salt concentration is the condition at which the enzyme activity was decreased by 80-90% but the ternary structure was retained. The enzyme *Hv*DHFR1 can be used as a model halophilic protein to give an explanation about the relationship between stability, flexibility, and activity as a function of salt concentration. This is the first NMR relaxation study performed for a halophilic protein at 3.5 M and 1.0 M salt conditions.

Protein Dynamics

Proteins are dynamic systems in solution.^{1,8,14-24} The internal flexibilities of proteins usually allow them to accomplish their biological activity. Commonly, the flexible regions of proteins are the catalytic regions. Characterizing the mobile regions of proteins is important in order to understand and identify their functions.²⁵⁻³⁰ There are a few experimental techniques that can be employed to study protein motions in the picoseconds to millisecond timescales. Time-resolved fluorescence anisotropy techniques provide information on mobility of a dye labeled sample.³¹⁻³⁷ A covalently attached dye reports the protein mobility and conformational changes. Solid state NMR also provides molecular motion data of a isotope labeled crystalline protein on the millisecond to nanosecond range.³⁸⁻⁴¹ X-ray crystallography can also yields information about the possible regions of a protein which undergoes motion. Regions with high temperature factors suggest the presence of motion. A recent crystallographic technique, time-resolved X-ray crystallography yields data for the regions undergoing motions in the ms-ns timescale.⁴²⁻⁴⁵ Molecular dynamic (MD) modeling is another technique that can be used to probe protein motions.⁴⁶⁻⁴⁸ The motion of protein modeling is simulated in the computer by the use of X-ray crystallographically or NMR-derived structure to elucidate the possible regions with motions. NMR spectroscopy, which is a powerful technique for investigation of protein flexibility, provides picoseconds-nanosecond and microsecond-millisecond timescale of the motions for individual residue in a protein. The dynamics of a protein can be analyzed either via backbone (¹⁵N or C') or side chain motions (¹³C), respectively.

Most molecular dynamic studies have focused on the analysis of backbone motions.^{6,7} The internal backbone motions of a protein can be investigated through the analysis of ¹⁵N relaxation parameters.^{7,12} The ¹⁵N-¹H bond vector is very sensitive to molecular motion under external magnetic fields and the local environmental conditions. The motion of each bond vector is hypothesized to be limited in a cone which is called “wobble in a cone” motion, Figure 5.1.⁴⁹ The magnitude of the bond motion is characterized from the order parameter, S^2 , and the internal effective correlation time, τ_e . The order parameter quantifies the restriction or freedom of the motion in the cone which is also a function of the angle θ and is given by the following equation:⁵⁰⁻⁵⁶

$$S^2 = \left[\cos \theta_0 \left(\frac{1 + \cos \theta_0}{2} \right) \right]^2. \quad [1]$$

The value of S^2 indicates the freedom of motion which ranges from 0 to 1. For a totally

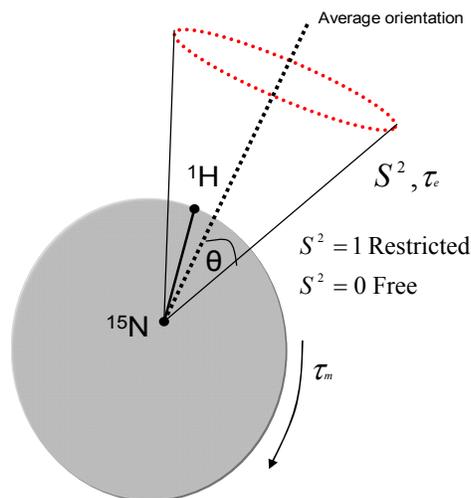


Figure 5.1 Illustration of ¹⁵N-¹H bond vector motion relative to a reference frame. The protein molecule is illustrated by gray region. The amplitude of motion is characterized by S^2 where $S^2=1$ indicates rigid, $S^2=0$ indicates isotropic motion.⁴⁹

restricted motion $S^2 = 1$ and for an unrestricted motion $S^2 = 0$. The backbone dynamics of proteins are analyzed through their relaxation parameters. The required relaxation data are obtained through manipulating the spin interactions between ^{15}N and the directly bonded ^1H . The relaxation mechanism is mainly governed by dipolar interactions between the two nuclei. Analysis of the ^{15}N longitudinal relaxation (T_1), transverse relaxation (T_2) and heteronuclear ^1H - ^{15}N NOE parameters from a series of modified ^1H - ^{15}N -HSQC spectrum provide the required relaxation parameters and enable the investigation of the timescale and the nature of the backbone motions. The analysis of the relaxation parameters are achieved by the Modelfree formalism which allow the characterization of protein motions.⁵⁷⁻⁵⁸

Spin Relaxation and Motion of Molecules in Liquid: Molecules in solution have three modes of motion: internal molecular motion, molecular translation, and molecular rotation. Internal molecular motion includes the vibrational motion of individual nuclei which is on the picosecond timescale.⁵⁹ Molecular transitions arise from the transport process of molecules in solution with two modes: *diffusion* and *flow* which is quantified by the *diffusion coefficient*, D .⁵⁹ The D value is size dependent: a large D value indicates relatively rapid motion and a low D value indicates slower motion.⁵⁹ Typically, the timescale of motion for small molecules in solution is in the picoseconds range and that for proteins is in the nanosecond range.^{5,7} Rotational motion is another form of molecular motion in solution which is defined as the rotation of a molecule in a random fashion around its center.⁸ Both diffusion and rotational motion are particularly

important in NMR studies of proteins since they are the key stimulators of intermolecular interactions and are the source of the relaxation phenomena.

Spin population between Zeeman states can be manipulated by the application of *rf* pulses.⁵⁴ The spin relaxation phenomena describe the re-establishment of the perturbed spin population back to thermal equilibrium. Two mechanisms govern spin relaxations: *spin-lattice* and *spin-spin*. The *spin-lattice* relaxation refers to exchange of energy of a nucleus with the surrounding fluctuating magnetic field that is generated by the other spins.⁵³ The *rf* pulse moves the net spin magnetization from the z-direction into the transverse plane (x-y plane). This is typically accomplished by a 90° pulse which perturbs the spin population between spin states.⁵³ Dipolar interaction and chemical shift anisotropy mechanisms act to re-establish the population distribution. *Spin-lattice* relaxation time (T_1) describes the rate of re-establishing the magnetization along the z-direction and relates to molecular tumbling. *Spin-spin* relaxation relates to dipolar spin interactions of nuclei. The time T_2 describes the rate of loss of magnetization in the transverse (x-y) plane.⁵² T_2 relates to molecular motion and is inversely proportional to molecular size. Larger molecules relax faster with shorter T_2 which is an undesirable feature in relaxation studies of large size proteins. Both *spin-spin* and *spin-lattice* relaxation processes are single exponential functions.⁵² Molecules with molecular weight <1 kDa have similar T_1 and T_2 values, typically at several seconds.⁵² For molecules with 5-30 kDa, typically T_1 is in 0.2 - 1.0 s and T_2 is in millisecond ranges. The time constant for *spin-lattice* relaxation rate (R_1) and *spin-spin* relaxation rate (R_2) are given by $1/T_1$ and $1/T_2$.⁵²

Nuclear Overhauser Effect (NOE): Dipole-dipole interactions, chemical shift anisotropy, spin rotation relaxation, and quadrupolar relaxations are the possible mechanisms that take place to re-establish the ground state equilibrium population when a spin state is perturbed.⁵¹ Nuclei with spin $\frac{1}{2}$ relax predominantly by dipolar interactions. Additionally, to a smaller extent, the chemical shift anisotropy (CSA) also contributes to the relaxation process which is the chemical shift differences of amide group being oriented parallel and perpendicular to the external magnetic field.⁶⁰⁻⁶⁵ One of the very important NMR phenomena arising from the dipolar interaction is the nuclear Overhauser effect (NOE). The NOE originates from dipole-dipole interactions of nuclear spins.⁵¹ The magnetically interacting spins affect the spin relaxation property of each other. The dipole interaction between the ^{15}N and its bonded ^1H are manipulated to study the spin relaxation properties of the ^{15}N nucleus.⁵¹ The magnitude of dipole-dipole coupling at the site of each ^{15}N nucleus is given by:⁵⁰⁻⁵⁶

$$B_N \propto -\gamma_N \gamma_H \left[\frac{3 \cos^3 \theta - 1}{r_{NH}^3} \right] \quad [2]$$

where θ is the angle between the N-H vector with respect to the direction of the static magnetic field vector, and γ_H and γ_N are gyromagnetic ratios of ^1H and ^{15}N . The fluctuating magnetic dipole-dipole interactions of the coupled spins induce a transition between the energy levels under the influence of external magnetic field. This is observed as signal enhancement in some spectrum.⁵⁰ This process also depends on the rotational correlation time of the protein. Once the spin population is perturbed for a set of nuclei, the relaxation processes result in building up of an NOE for the coupled nuclei.⁵⁰ Practically, *transient* and *steady-state* NOE are two NMR techniques used in

structure and relaxation studies of proteins. The *transient* NOE technique is used in NOESY experiments to obtain distance restraints for structure elucidation.⁶⁶ The NOE build up rate is distance dependent. NOE peak intensity is a means to quantify the distance between the interacting nuclei.⁶⁶ The *steady-state* NOEs are obtained by manipulating the spin population for one set of nuclei by selectively saturating and detecting the response on the coupled nuclei. Figure 5.2 represent the energy level for a coupled spin system for I and S nuclei.⁶⁷ W stands for transitions rate between the energy states, and W_0 , W_1 , W_2 are zero, single and double quantum transitions. W_0 and W_2 are also called cross-relaxation. The figure reveals that at equilibrium all possible transitions across the spin states can take place. The spin population of I nucleus between the α and β states is saturated (in case of H-N couples: I is ^1H and S is ^{15}N nucleus) via applying weak radio frequency pulses at frequency of spin I. Once the spin population is

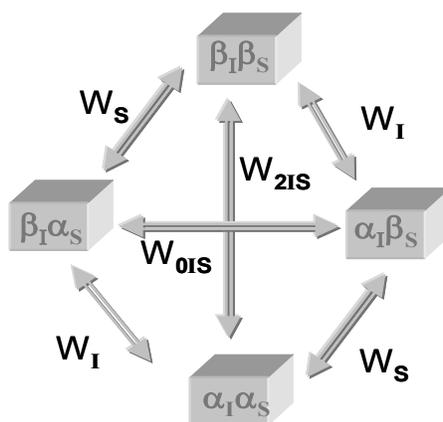


Figure 5.2. The energy level diagram for two spin system is shown. The probabilities of transition across the spin states are shown. W_S and W_I are single quantum transition rates; W_0 , and W_2 are zero and double quantum transitions rates.⁶⁴

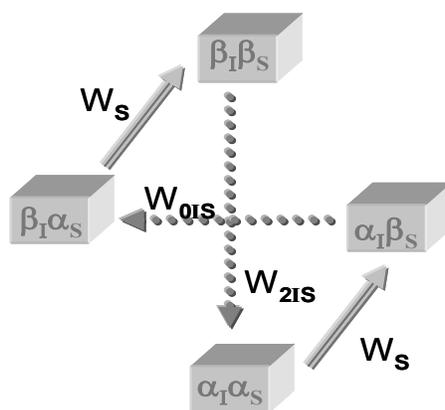


Figure 5.3. The transition probabilities after spin saturation of I is illustrated. The cross relaxation rate for zero quantum transition is proportional to $W_I - W_S$ and for double quantum transition transition is proportional to $W_I + W_S$.⁶⁴

saturated for I, the only mechanism that will act to reestablish the equilibrium is the zero- and double-quantum transitions (i.e. cross-relaxations mechanism will act), Figure 5.3.⁶⁸

As the equilibrium is establishing via these cross relaxations, NOEs are observed in the spectrum. This process is also molecular tumbling (τ_m) rate and γ dependent.

Molecular size controls the τ_m and dictates the type of cross-transition. For small organic molecules with shorter τ_m , the W_2 transition pathway dominates. On the other hand, for larger molecules such as proteins with longer τ_m , the W_0 transition pathway dominates.

The zero (W_0) and double (W_2) quantum transition rates are given by the relation:⁵⁰⁻⁵⁶

$$W_0 \propto \frac{12\tau_m}{r^6(1 + (\omega_I + \omega_S)^2 \tau_m^2)} \quad [3]$$

$$W_2 \propto \frac{12\tau_m}{r^6(1 + (\omega_I - \omega_S)^2 \tau_m^2)} \quad [4]$$

where τ_m is the overall correlation time, r is the distance between spin I and S and ω is

the Larmor frequency of I and S nuclei. The competition between zero and double transitions determine the NOE enhancement and also the sign of the NOE. If the double quantum transition is dominant during the relaxation process, NOE signal enhancements will result. If the zero quantum transition is dominant, there would be a decrease in signal intensities. The signal enhancement is given by;⁵⁰⁻⁵⁶

$$\eta_S = \frac{\gamma_I}{\gamma_S} \frac{(W_2 - W_0)}{(2W_1S + W_2 + W_0)} \quad [5]$$

where η_S is signal enhancement, γ_I , and γ_S are gyromagnetic ratio for spin I and S.

Microsecond-Millisecond Time Scale Motion: NMR spectroscopy is sensitive to ns-ps timescale backbone and side chain motions of proteins (which are at the rate of Larmor frequencies of nuclei). Some biochemical events such as conformational exchanges (sometimes referred as chemical exchange), substrate binding, and substrate release events may take place at slower timescales such as microsecond-millisecond (μ s-ms). It is also possible to detect μ s-ms timescale motions by NMR Spectroscopy. Figure 5.4 illustrates a nucleus experiencing two different chemical environments.⁵⁰⁻⁵⁶

The presence of any chemical exchange event affects the R_2 rate. A two sites conformational exchange between A and B states can be stated as⁵⁰⁻⁵⁶



where A and B denote the population of a nucleus (i.e ^{15}N nucleus) at A (pA) and B (pB) states and the constant k_1 and k_{-1} are the forward and backward rate constants. The rate constant for the exchange process is given by k_{ex} which is $k_1 + k_{-1}$. Two different

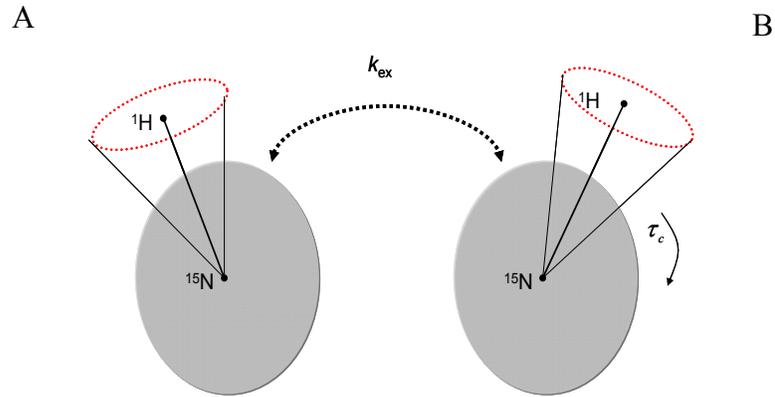


Figure 5.4. The figure illustrates the N-H bond vector experiencing two sites with an exchange rate of k_{ex} .

chemical shift values or single chemical shift value can be detected for a nucleus experiencing two different environments. The transverse relaxation time constant and chemical shift difference $\Delta\nu_{AB}$ ($\omega_A - \omega_B$) of the nucleus between the two conformations is a rate dependent process. The relationship between the $\Delta\nu_{AB}$ and chemical exchange can be simplified as:⁵⁰⁻⁵⁶

$$k_{ex} > \Delta\nu_{AB} \text{ fast exchange,}$$

$$k_{ex} \sim \Delta\nu_{AB} \text{ intermediate exchange,}$$

$$k_{ex} < \Delta\nu_{AB} \text{ slow exchange.}$$

The measured R_2 relaxation rate bears the chemical exchange information and is given by:⁵⁰⁻⁵⁶

$$R_2 = R_2^0 + R_{ex} \quad [7]$$

where R_2^0 is the inherent transverse relaxation rate which originates from dipolar interaction and to a lesser extend from CSA. The R_{ex} term is the contribution from the chemical exchange process. The R_2 relaxation rate is given by the equation:⁵⁰⁻⁵⁶

$$R_2 = p_A R_{2A}^0 + p_B R_{2B}^0 + p_A p_B \Delta\omega^2 k_{ex} \left[\frac{k_{ex} + \Delta\omega^2}{(k_{ex} + \Delta\omega^2) - p_A p_B (5k_{ex}^2 + \Delta\omega^2) \Delta\omega^2} \right] \quad [8]$$

where the R_{2A}^0 and R_{2B}^0 are inherent R_2 rate constants at the conformational sites A and B. A detailed analysis of chemical exchange analysis requires separate and complicated NMR experiments to be run. The advantage of the use of the Modelfree protocol would also provide the R_{ex} term from the analysis of T_2 data.

Relaxation Theory: Relaxation properties of proteins are strongly relate to their local and globular motions and can be quantified from the analysis of R_1 and R_2 , and steady-state NOE parameters. NMR provides dynamic information of a protein in two timescales: ps-ns and μ s-ms from the analysis of R_1 ($1/T_1$), R_2 ($1/T_2$), and heteronuclear ^1H - ^{15}N NOE parameters. In an isotropic liquid, molecules tumble and exhibit random Brownian type motions for which the time dependent fluctuation between two points is given by the correlation function, $C(\tau)$:⁵⁰⁻⁵⁶

$$C(\tau_m) = f(\tau_m) f(t + \tau_m) \quad [9]$$

where $f(\tau_m)$ and $f(t + \tau_m)$ refer to the position of a molecule between two points in time. The $f(\tau_m)$ is an exponential function of rotational correlation time (τ_m). For an isotropic molecular motion $f(\tau_m)$ is given by:⁵⁰⁻⁵⁶

$$f(\tau_m) = \frac{1}{5} e^{(-t/\tau_m)} \quad [10]$$

The correlation function is a time domain function and the Fourier transformation gives the frequency domain function which yields a Lorentzian shaped curve. The area under the curve is called the spectral density function, $J(\omega)$,⁵⁰⁻⁵⁶

$$J(\omega, \tau_m) = \frac{2\tau_m}{5(1 + \omega^2\tau_m^2)} \quad [11]$$

$J(\omega)$ describes the probability of a particular motion at a given frequency.⁶⁹ In this expression, the molecule is assumed to be completely rigid with an isotropic rotational diffusion model.⁷⁰⁻⁷³ The spectral density function includes the effect of the lattice on relaxation of nuclei as proteins tumble in solution and identifies the reorientation of individual N-H bond vectors during the tumbling process. The link between the R_1 and R_2 relaxation rates, heteronuclear NOE, and spectral density functions for ^{15}N spin is a linear combination of spectral density functions at five frequencies and given by⁶⁶

$$R_1 = \frac{1}{T_1} = \frac{d^2}{4} [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + C^2 J(\omega_N) \quad [12]$$

$$R_2 = \frac{1}{T_2} = \frac{d^2}{8} [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{C^2}{6} [4J(0) + 3J(\omega_N)] + R_{ex} \quad [13]$$

$$NOE = 1 + \left(\frac{d^2}{4R_1} \right) \left(\frac{\gamma_H}{\gamma_N} \right) [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]$$

[14]

and

$$d = \left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{8\pi^2} \right) \left\langle \frac{1}{r_{HN}^3} \right\rangle; C = \frac{\omega_N \Delta\sigma}{3^{1/2}} \quad [15]$$

where μ_0 is the permeability of free space ($4\pi \times 10^{-7} \text{ NA}^{-2}$), \hbar is Plank's constant, γ_H and γ_N are gyromagnetic ratios of ^1H and ^{15}N nuclei, r_{H-N} is internuclear distance (1.02 Å),

ω_H , and ω_N are Larmor frequencies of ^1H and ^{15}N nuclei, $\Delta\sigma$ is the chemical shift anisotropy (CSA) of ^{15}N nucleus (~ 172 ppm).

Modelfree Theory: Protein dynamic information from the relaxation studies can be accomplished by two methods: either treating the spectral density directly or indirectly. Direct spectral density mapping involves the direct determination of the spectral density from experimental relaxation data without prior assumptions.⁷⁴ The method necessitates the determination of five relaxation rates at $J(0)$, $J(\omega_H)$, $J(\omega_N)$, $J(\omega_H + \omega_N)$, and $J(\omega_H - \omega_N)$ frequencies, which is practically not feasible.^{4-6,75} An alternative and popular method, the Modelfree formalism, has been utilized in the analysis of NMR relaxation parameters to address the motion of the N-H bond vector by using models.^{57,58,76-78} In the Modelfree approach, the molecular motion of protein is assumed to be isotropic and characterized by the total correlation time, τ_m . The correlation function for internal motion is assumed to be a single exponentially decaying function and reaching a value of S^2 as $\tau \rightarrow \infty$ ($C(\infty) = S^2$). The correlation function can be re-stated as:⁷⁹

$$f(t + \tau) = S^2 + (1 - S^2)e^{-\frac{1}{\tau}}. \quad [16]$$

The Fourier transform of the correlation function ($C(\tau) = f(\tau)f(t + \tau)$) yields the spectral density function in the form of:⁵⁰⁻⁵⁶

$$J(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_m}{1 + \omega^2 \tau_m^2} + \frac{(1 - S^2) \tau}{1 + \omega^2 \tau^2} \right] \quad [17]$$

in which τ is given by

$$\frac{1}{\tau} = \frac{1}{\tau_m} + \frac{1}{\tau_e}. \quad [18]$$

The term τ_e describes the effective intermolecular correlation time for motion of an individual H-N bond vector. For the relation between the correlation times given above, τ_e is faster than τ_m but slower than 10 ps. The order parameter, S^2 , emphasizes the degree of N-H bond freedom as a result of motion and its value ranges from 0 and 1. For a completely restricted motion, S^2 is 1 and for fully non-restricted motion it is 0. The general order parameter has been extended by Clore et al.⁷⁸ In this extension, the general order parameter S^2 is split into a two-order parameters in which S_f^2 reveals the faster timescale motion (picoseconds range) and S_s^2 reveals the slower timescale internal motion (nanosecond range). The relation between the two parameters is $S^2 = S_f^2 S_s^2$. The extended spectral density function is given by the equation.⁸⁰

$$J(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_m}{1 + (\omega \tau_m)^2} + \frac{(1 - S_f^2) \tau_f'}{1 + (\omega \tau_f')^2} + \frac{(S_f^2 - S^2)}{1 + (\omega \tau_s')^2} \right] \quad [19]$$

in which

$$\frac{1}{\tau_f'} = \frac{1}{\tau_m} + \frac{1}{\tau_f} \quad \frac{1}{\tau_s'} = \frac{1}{\tau_m} + \frac{1}{\tau_s} \quad [20]$$

In the Lipari-Szabo Modelfree approach, five models are used to fit the relaxation data, and each of the model describes the motion of an H-N bond vector in the protein structure.⁸¹ In Model 1, only the order parameter, S^2 is optimized; in Model 2, S^2 and τ_e (the internal correlation time) are optimized. Model 3 is identical to Model 1 with the addition of R_{ex} term to R_2 ; similarly, Model 4 is identical to Model 2 with the addition of

R_{ex} term to R_2 . In Model 5, S_f^2 , S_s^2 , and τ_e are optimized. The Modelfree formalism treats the internal motion and overall motion separately in which the generalized order parameter S^2 , quantifies the spatial limitation of the N-H bond vector.⁸² The reported S^2 values in Modelfree calculations are motions that faster than τ_m .⁴⁹ Model 1 only includes very fast timescale motion < 20ps and ignores slower timescale motion. The assumptions made in Model 1 are $S_s^2=1$ (i.e. $S^2=S_f^2$), $\tau_e = \tau_f \rightarrow 0$, and $R_{ex}=0$. In Model3 and Model5, it is also assumed that $\tau_f \rightarrow 0$. The $S_s^2=1$ is the assumption made in Model 1, Model 2, Model 3 and Model 4. Model 2 ignores the slower timescale motion and considers only fast timescale motion. In this model, S_f^2 and the τ_e are optimized where the timescale of motion for τ_e is assumed in 20-500 ps range. Model 3 and Model 4 are

Models	Parameter(s)
1	S^2
2	S^2 , $\tau_e = \tau_f$
3	S^2 , R_{ex}
4	S^2 , τ_e , R_{ex}
5	S_f^2 , S_s^2 , τ_e

equivalent of Model 1 and Model 2, respectively. Additionally these models include the R_{ex} term. The R_{ex} term does not represent the spectral density value but originates from extremely large R_2 values.⁸³ It is not accepted as fast timescale motion since it results from chemical exchange process that takes place at the μ s-ms range. In the Modelfree calculations, any fitting to Model 3 and Mode 4 suggest the existence of

chemical exchange. Model 5 characterizes two different timescale motions: fast (S_f^2) and slow (S_s^2). The timescale for fast motion is faster than 20 ps and for slow motion it is slower than 500 ps.

The Modelfree 4.14 (<http://cpmcnet.columbia.edu/dept/gsas/biochem/lab/palmer>.) program extracts the order parameters of individual residues by fitting the experimental data to simulated data via minimization of the χ^2 (sum square error function) function for each ^{15}N nuclear spin⁵⁰⁻⁵⁶

$$\chi^2 = \sum \left(\frac{R_1^T - R_1^E}{\sigma_{R_1}^2} + \frac{R_2^T - R_2^E}{\sigma_{R_2}^2} + \frac{NOE^T - NOE^E}{\sigma_{NOE_1}^2} \right) \quad [21]$$

where in the relaxation parameters T stands for the theoretical value (simulated value) and E stands for the experimental value.⁸¹ The σ_R^2 is the experimental uncertainty for each parameter. F tests are used to choose the best model using the selection criteria described by Mandel et al.⁸¹

Extracting the Relaxation Rate Constants with the Software Felix: A series of spectrum were recorded at relaxation delays of 50 ms, 100 ms, 200 ms, 400 ms, 700 ms, 1000 ms, 1500 ms for R_1 experiment at 3.5 M and 1.0 M NaCl. Figure 5.5 shows the spectra collected at 50 ms, 700 ms, and 1500 ms time delays in 3.5 M NaCl. In the same way, seven spectrum were recorded with variable CPMG⁸⁴ (Carr, Purcell, Meiboom and Gill) delays at 50 ms, 70 ms, 90 ms, 110 ms, 150 ms, 150 ms, 190 ms, 250 ms for R_2 experiment at 3.5 M and 1.0 M NaCl. The ^1H - ^{15}N steady-state NOE experiments were carried out with and without proton saturation at 3.5 M and 1.0 M NaCl. Figure 5.6

shows the spectrum sections from with and without proton saturation spectrum at 3.5 M NaCl. The relaxation data were analyzed by using the software Felix.

The peak heights or volumes were calculated for each residue. The R_1 relaxation rates were calculated by fitting the peak heights for each residue to the following equation:

$$I = I_0 e^{R_1 t} \quad [22]$$

The relaxation rate R_1 for each of the residues is derived from the exponential term by fitting the height of each peak with respect to delay time to the least square function.

Figure 5.7 represents the time course measurements of peak intensity changes for residue W134 at 3.5 M NaCl. The R_2 rate was calculated by fitting the peak heights at the time course to the equation:

$$I = I_0 e^{R_2 t} \quad [23]$$

Figure 5.8 represents the time course measurements of peak intensity changes for residue L21 at 3.5 M NaCl. The ^1H - ^{15}N nuclear NOE for individual residue can be calculated from the peaks intensity with and without proton saturation by the equation:

$$NOE = \frac{I^s}{I^0} \quad [23]$$

where I^s is the peak intensity with proton saturation (noe_on) and I^0 is the peak intensity without proton saturation (noe_off). Figure 5.9 and Figure 5.10 show the derived T_1 , T_2 and heteronuclear ^1H - ^{15}N NOE values for the enzyme HvdDHF1 at 3.5 M and 1.0 M NaCl conditions. There is no specific relationship between T_1 and T_2 and flexibility. However, the ^1H - ^{15}N NOE values are indicative of the flexibility. The residues with low heteronuclear ^1H - ^{15}N NOE values ($\sim < 0.7$) are considered to be the flexible residues.

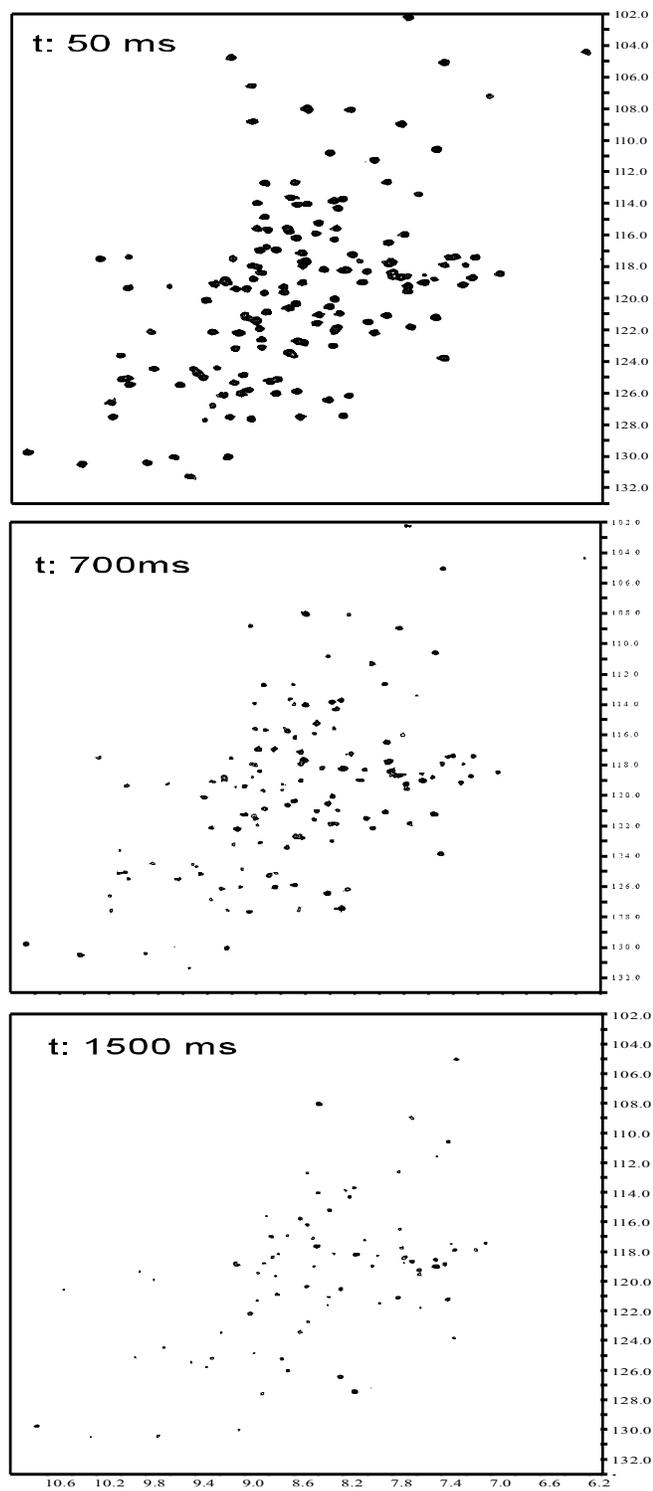


Figure 5.5. The ^1H - ^{15}N HSQC spectrum of *Hv*DHFR1 at 3.5 M salt concentration recorded for R_1 experiment at the relaxation delay times: 50 ms, 700 ms and 1500 ms.

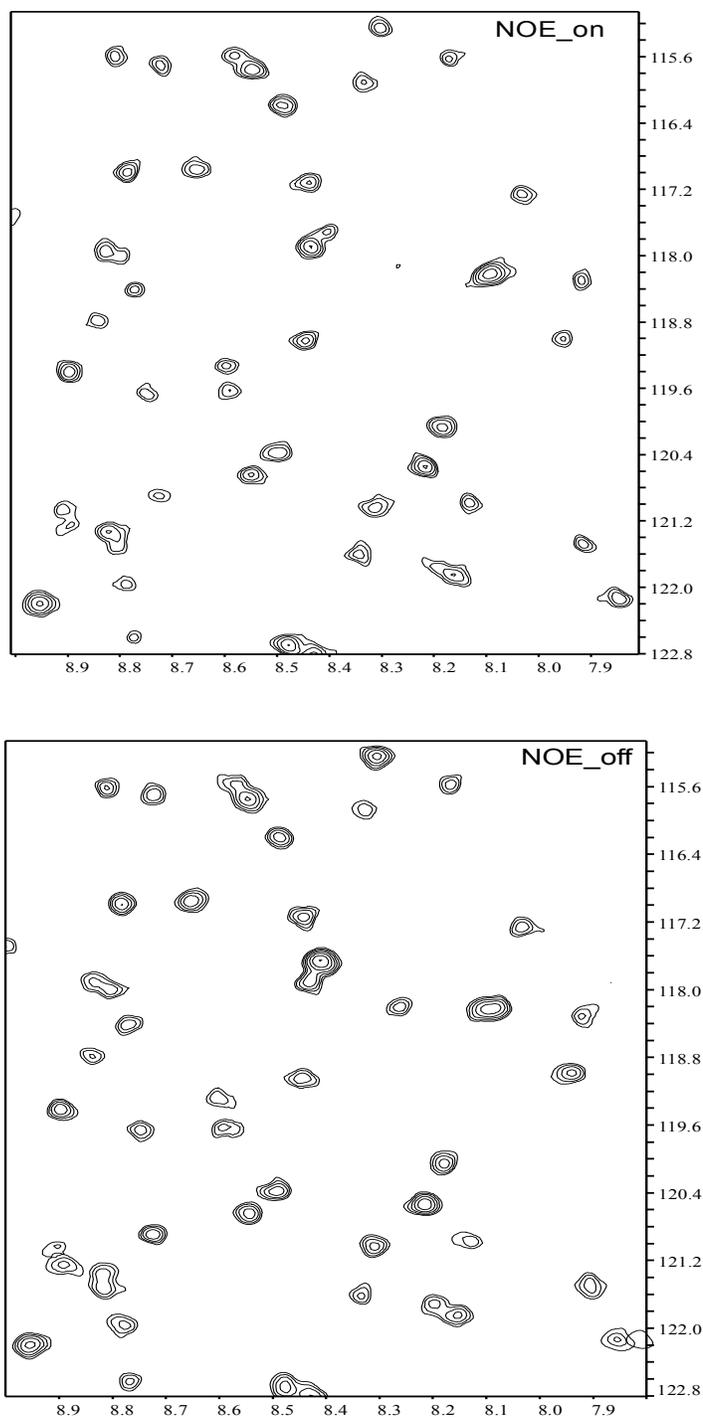


Figure 5.6. A spectrum region from the ^1H - ^{15}N HSQC of *Hv*DHFR1 at 3.5 M NaCl is shown. The spectra with proton saturation (noe_on) and without saturation (noe_off) are shown.

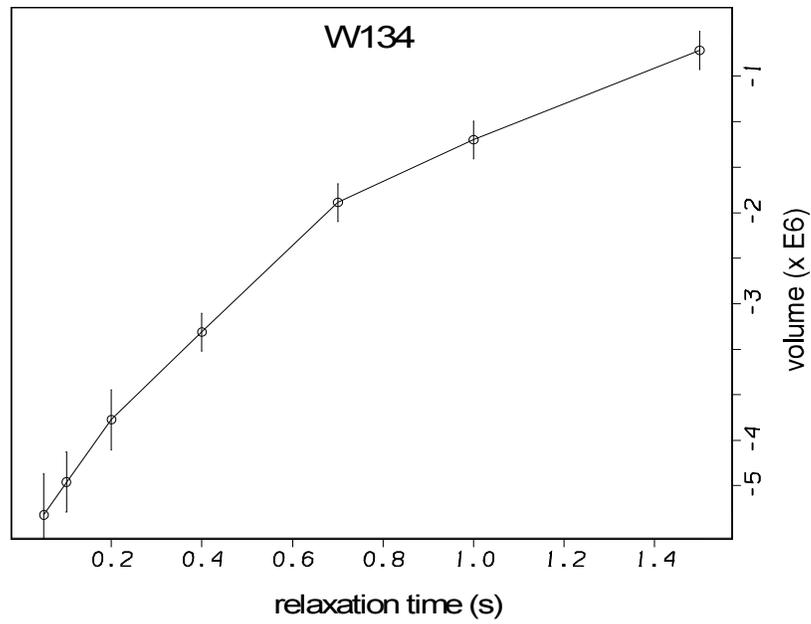


Figure 5.7. The R_1 time course measurements for W134 of *HvDHFR1* at 3.5 M NaCl are shown.

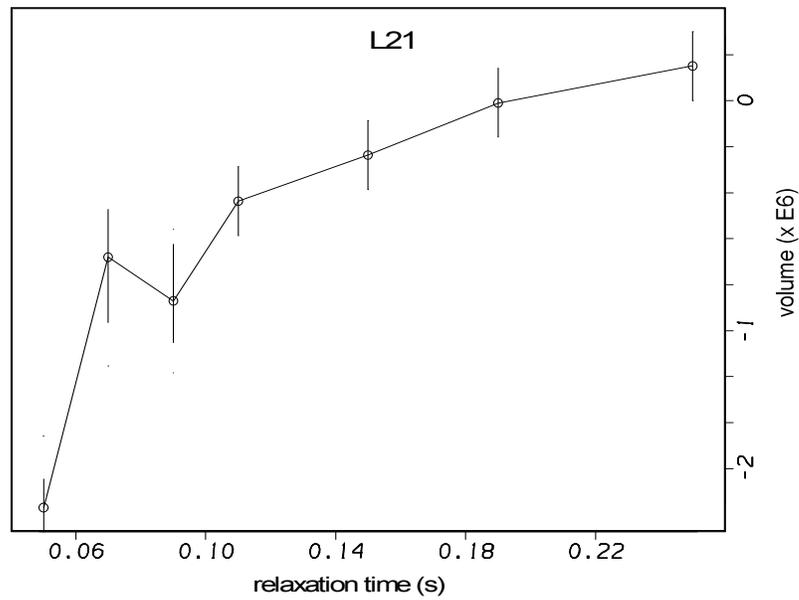


Figure 5.8. The R_2 time course measurements for L21 of *HvDHFR1* at 3.5 M NaCl are shown.

Initially, the τ_m was approximated from R_2/R_1 ratio for individual ^{15}N nuclei since this parameter is closely related to the overall tumbling part of spectral density function. Isotropic diffusion model was verified by using the diffusion protocol where $D_{\text{parallel}}/D_{\text{perpendicular}}$ was calculated as 1.19.

Results

The resulting order parameter (S^2) from the analysis of T_1 , T_2 , and heteronuclear ^1H - ^{15}N NOE data through the Modelfree approach at 3.5 M NaCl and 1.0 M NaCl are shown in Figure 5.11. The mobile regions are defined based on the S^2 values for which $S^2=1$ indicates totally restricted motion and $S^2=0$ indicates totally free motion. Typically for proteins the S^2 values average to 0.8 and regions below this value are considered as flexible. Based on the S^2 values, four mobile regions are defined as L21, α 2- β 3, adenosine binding (ABL), and β 6- β 7 loops. These flexible regions are labeled on an NMR derived structure and shown in Figure 5.12. The backbone-dynamic results of *Hv*DHFR1 at 3.5 M NaCl are comparable to the relaxation data of the *Ec*DHFR:folate binary complex.¹¹ Based on the crystal structure and NMR dynamic studies of *Ec*DHFR, four loop regions were defined, each with a specific role in the catalytic events.¹¹ Among the loops region of *Hv*DHFR1, the adenosine binding loop (S65 to T75) exhibited relatively highest mobility based on the lowest S^2 value obtained at 3.5 M NaCl. The loop β 6- β 7 (P122- E138) region displayed moderate mobility. The magnitude of the L21 loop motion of apoenzyme *Hv*DHFR1 is slightly larger than the equivalent M20 loop of the *Ec*DHFR:folate complex. This can be explained from the interaction of residues at

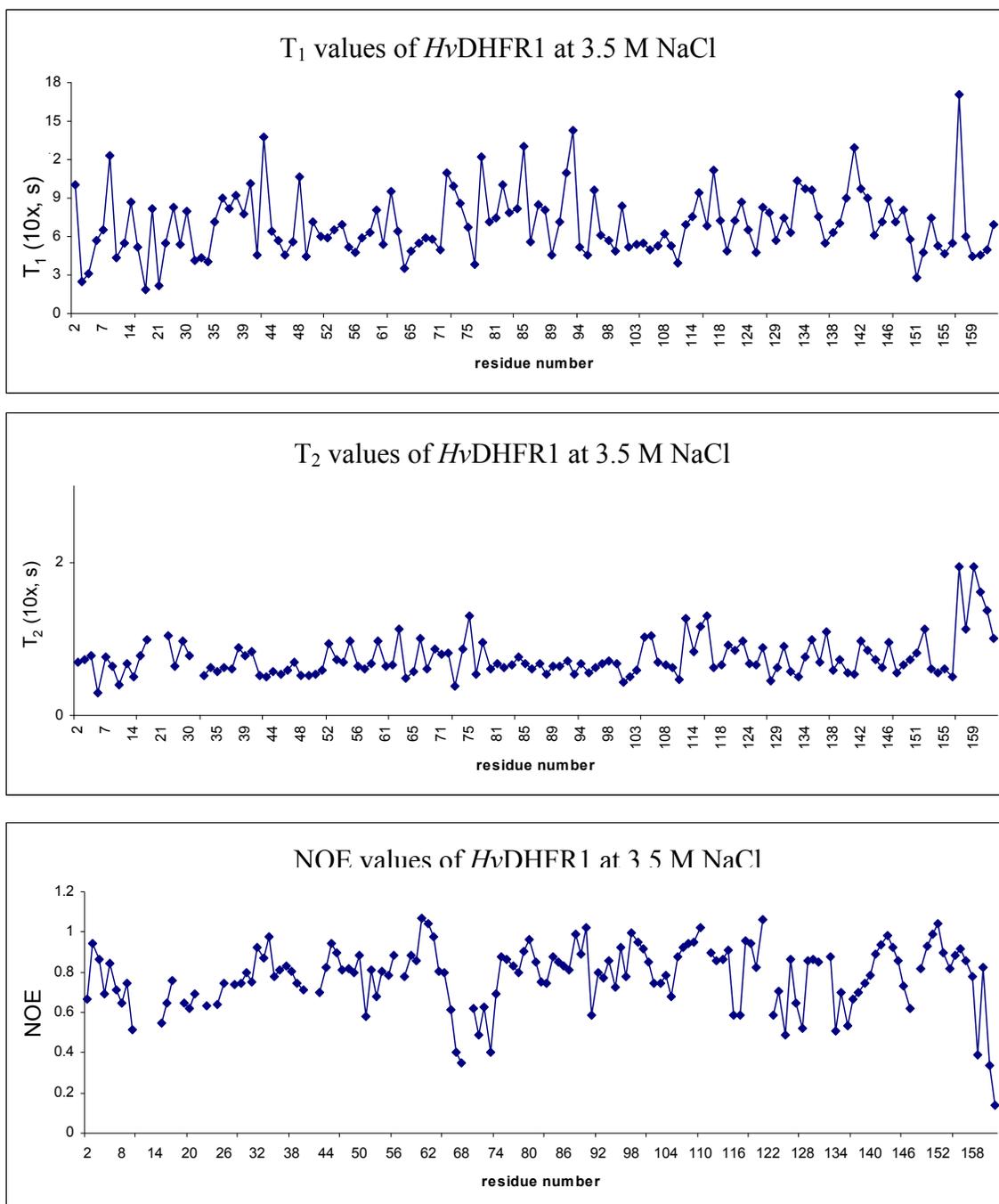


Figure 5.9. The calculated T_1 , T_2 and heteronuclear ^1H - ^{15}N NOE values for the enzyme *HvdHFR1* at 3.5 M NaCl is shown.

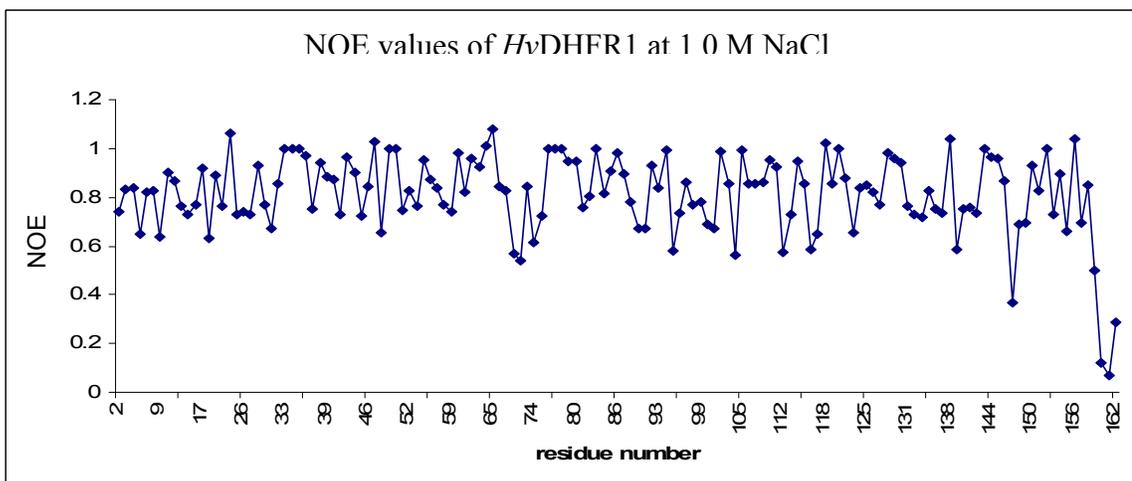
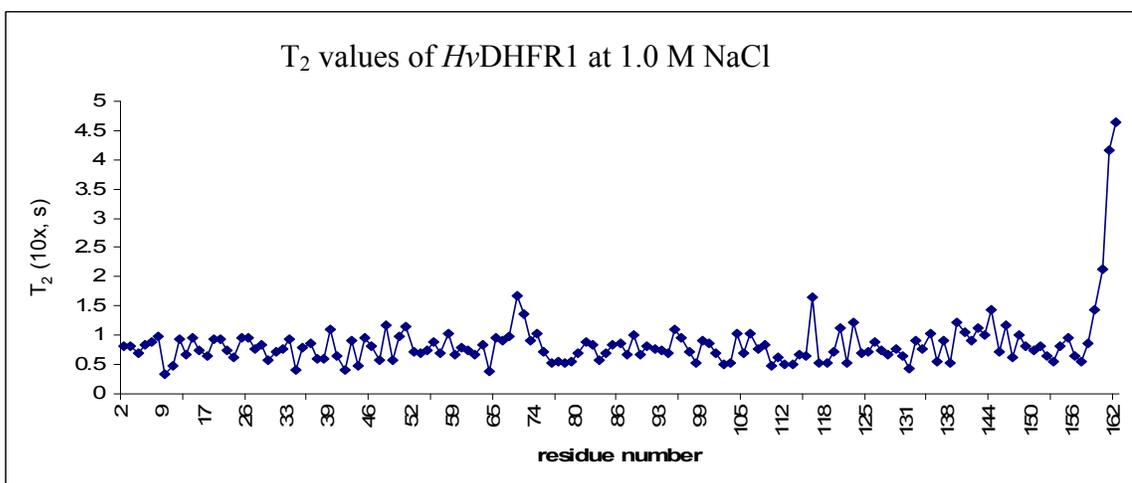
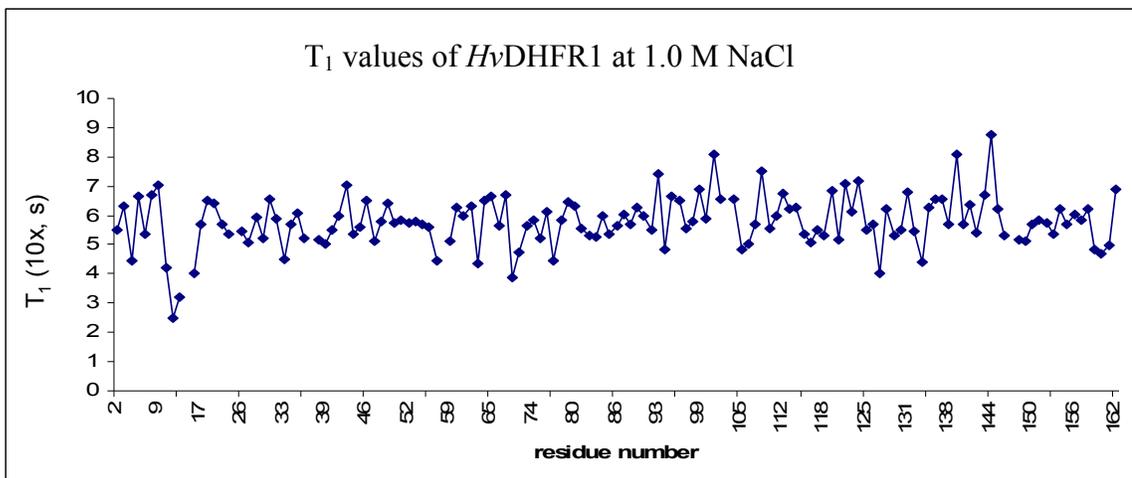


Figure 5.10. The calculated T_1 , T_2 and heteronuclear ^1H - ^{15}N NOE values for the enzyme *Hv*DHFR1 at 1.0 M NaCl is shown.

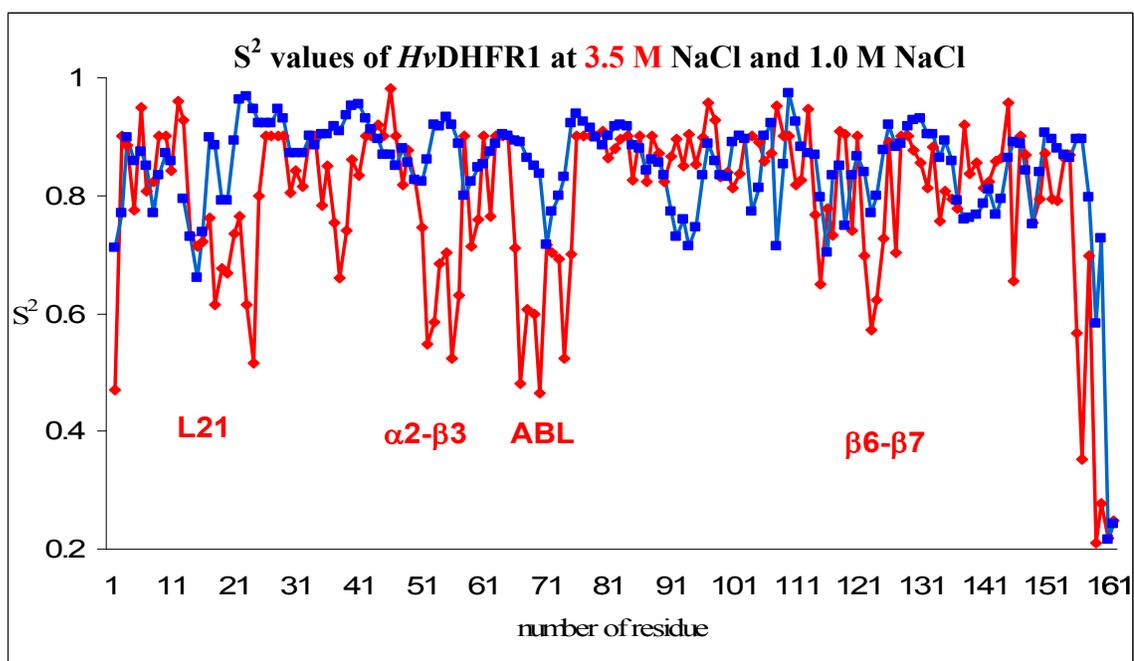


Figure 5.11. The superimposed of order parameters of backbone ^{15}N resulting from 3.5 M (red) and 1.0 M NaCl salt (blue) concentrations. The order parameter varies between 1 and 0 where $S^2=1$ indicates completely restricted motion and the $S^2=0$ indicates large amplitude of motion. The flexible loop regions at 3.5M NaCl are defined from their low S^2 values: L21 loop, $\alpha 2$ - $\beta 3$, adenosine binding loop (ABL), and $\beta 6$ - $\beta 7$ loop.

the M20 loop with folate, such as the E17 side chain being close to the benzoic acid moiety of folate. This proximity may be causing attenuation in flexibility of the M20 loop in *Ec*DHFR:folate complex. The flexibility of the L21 loop was also supported by the high crystallographic temperature factors obtained from the crystal structure study.⁸⁵ If data were available for apoenzyme *Ec*DHFR, this loop may exhibits a larger magnitude of flexibility. The residues of the $\beta 6$ - $\beta 7$ loop of *Hv*DHFR1 exhibited relatively low amplitude motion with respect to the equivalent F-G loop in *Ec*DHFR binary complex. The glycine residue is found to the most flexible residue in both enzymes (Gly123 in *Hv*DHFR1 and Gly121 *Ec*DHFR binary complex). This residue is known to play an

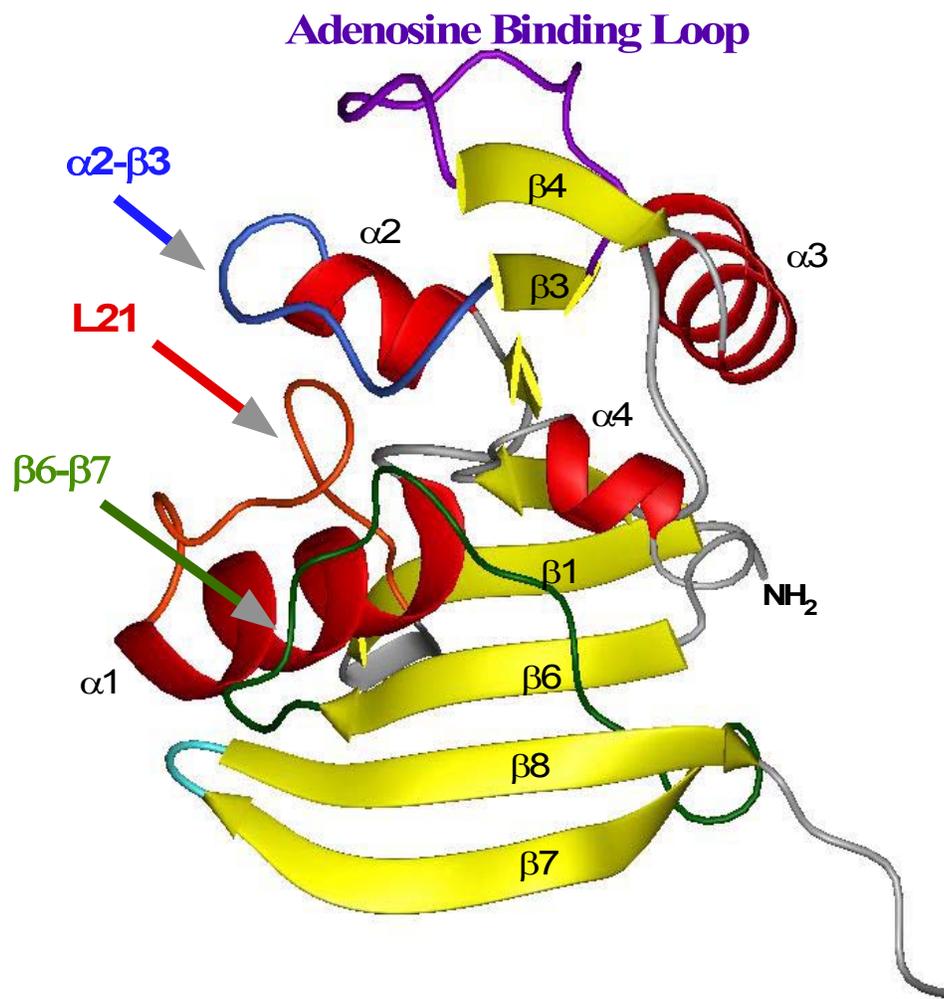


Figure 5.12. The four flexible loop regions L21, adenosine binding, $\alpha 2-\beta 2$, and $\beta 6-\beta 7$ are shown on an NMR derived structure.

important role in the catalytic activity for *Ec*DHFR. The mutation of G121 to valine in *Ec*DHFR resulted in 40 fold decrease in NADPH binding and 200 fold decrease in the hydride transfer rate.⁸⁶⁻⁸⁸ The backbone dynamic data of *Hv*DHFR1 at 3.5 M NaCl concentration suggest the existence of an additional flexible region, α 2- β 3 loop (R53-G58), whereas, in the *Ec*DHFR binary complex this flexibility is missing. The flexibility of S51 and R53 at this region may be the major reason causing the divergence between the NMR solution structures and the crystal structure. The last loop region (GH loop) defined for *Ec*DHFR comprising the residues 142-150, is a Type-I β -turn in *Hv*DHFR1 (D146-F150). In *Ec*DHFR, the residues of the GH loop form a hydrogen bond with residues of the M20 loop resulting in the *occluded* conformation of the M20 loop in substrate and cofactor complexes.^{89,90} It is not possible for there to be an interaction of this β -turn in *Hv*DHFR1 with the L21 loop due to deletion of six residues in this loop region. Instead, this turn was found to be close to the N-terminus of α 1 helix and the presence of a possible hydrogen bond between the residues in this turn (E148) and the α 1 helix (A28) was determined. The residue A9 in *Ec*DHFR:folate was determined with high R_{ex} term with $> 5.5\text{s}^{-1}$.¹¹ The R_{ex} terms indicate the presence of conformational exchange where a residue experiences two different chemical environments. The calculations for the 3.5 M salt concentration data resulted in R_{ex} term of 0.845 s^{-1} for residue A10 of *Hv*DHFR1. The residue Y128 of *Ec*DHFR:DHNADPH was found with R_{ex} term and the calculation for 3.5 M NaCl concentration resulted in R_{ex} term of 0.895 s^{-1} for *Hv*DHFR1 for the corresponding residue W134.

The enzyme *Hv*DHFR1 showed decreased enzymatic activity as salt concentration was decreased from 3.5 M to 1.0 M NaCl.⁹ It has been proposed that the

activity lost for *Hv*DHFR1 may be due to the changes in the inherent flexibility that facilitates the catalytic activity.⁹ This hypothesis was tested by carrying out the ¹⁵N relaxation studies of *Hv*DHFR1 at 1.0 M NaCl. Figure 5.14 shows the calculated order parameters for the enzyme in 1.0 M NaCl concentration (blue graph). The figure illustrates that there is a general decrease in the order parameters for 1.0 M NaCl concentration data. This indicates that the enzyme loses its flexibility across the peptide backbone particularly at loop regions as the salt concentration is lowered. This may connect to the activity loss as salt concentration is lowered.

The question to be answered now is how low salt concentration reduces the flexibility at the loop regions and which type of molecular interactions cause this loss in flexibility. The structural data for halophilic proteins illustrate that the surface of these proteins are mainly covered with negatively charged residues.⁹¹ The positively charged patches are centered at the substrate binding sites. The residues at the surface of the protein coordinate large amounts of water molecules which enable the proteins to be soluble and active in aqueous medium. It has been stated that the charged groups also coordinate a limited number of counterions.⁹²⁻⁹⁴ Yet, there has been no complete explanation about how halophilic enzymes retain their activities and flexibilities at such a high salt concentration. The reduction in enzymatic activity will be explained from changing the weight of two forces. One is the decreased hydrophobic interaction and the other is the increased electrostatic interactions. Figure 5.13 shows the electrostatic potential of the surface of an NMR derived structure where the negatively charged regions are shown in red and the positively charged regions are shown in blue. The loop regions are solvent exposed and generally rich in charged residues. One feature of the

L21 loop is the distribution of hydrophobic residues (A9, V14, I15, L21, W23, and I26) across the loop with separation by the charged residues. As the salt concentration is decreased, i.e. as the water activity is increased, the hydrophobic interaction may weaken in this region and cause the reduction in the flexibility. The increased number of water molecules around the hydrophobic residues can decrease the freedom of motion in the vicinity of these residues. The P22-W23-P23 combination in the L21 loop region may also intensify the rigidity due to hydrophobic effect on tryptophan and the additional constraints introduced from being between two prolines. The hydrophobic interaction between the aliphatic side chains are enhanced at high salt due to scarcing of water molecules. The hydrophobic interaction may be interrupted at low salt conditions by coordinating more water molecules around the residues. On the other hand, the electrostatic interaction between the charged residues in the M20 loop (E11, R13, R17, D18, and E20) and solvent molecules may be enhanced in the presence of increased water molecules. The carboxylic groups are known to coordinate a large amount of water molecules which facilitate the protein solubility at high salt media. The crystal structure of halophilic proteins showed that only small numbers of counterions are found at the surface.^{92,93,95-100} The decreased numbers of counter ions may cause an increase in repulsive interaction between similarly charged. At optimum salt concentration there may be a balance between repulsive interactions or/and both repulsive and attractive interaction that facilitate the flexibility of the enzyme.

The ABL facilitates the binding of cofactor (NADPH). The flexibility reduction in this region seems to be favoring the NADPH affinity. The phenylalanine (F71) is the largest hydrophobic residue residing at the center of this loop which may contribute to the

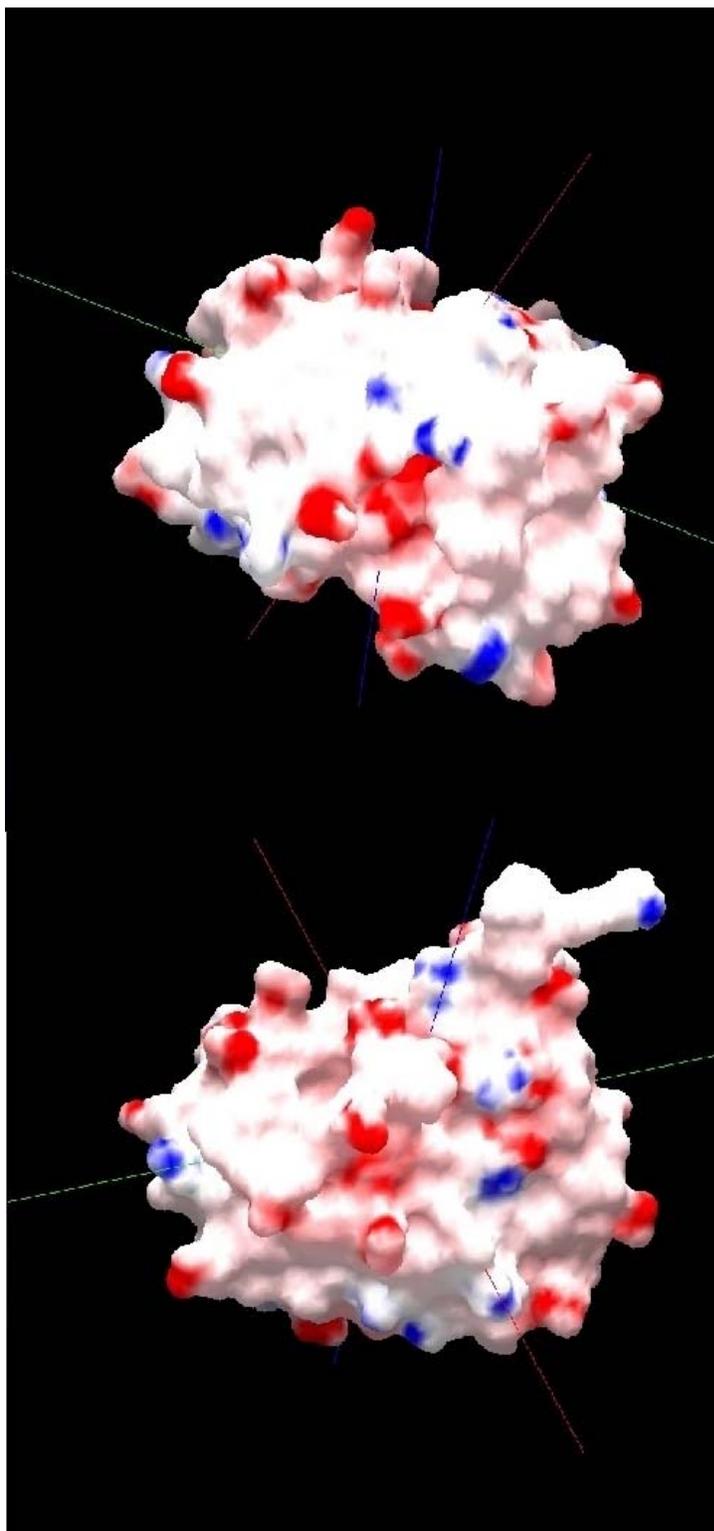


Figure 5.13. The electrostatic surface of NMR derived structure is shown. The negatively charged regions are illustrated by red and positively charged regions are illustrated by blue.

reduced flexibility. There are four serine residues in this loop; each of the hydroxyl side chain groups of the residues can form hydrogen bonds with water molecules. It has been shown that the enzyme *HvDHFR1* retains enzymatic activity at low pH even at low salt concentration.¹⁰¹ The protonation of the carboxylic and hydroxyl groups can screen electrostatic interaction between the ionizable charged groups.¹⁰¹

The flexibility at $\beta 6$ - $\beta 7$ loop is reduced overall. This loop encompasses hydrophobic residues, W134, A136, and A137 which may result in reduced flexibility. The loss of flexibility for the most flexible residue, G123, in the $\beta 6$ - $\beta 7$ loop can be explained from the rigidity resulting from P122 at low salt. The hydroxyl side chain of Y125, T129, Y130, Y131, may form extensive hydrogen bonding with water molecules which may play a role in the reduction of flexibility in the loop.

Overall Conclusion

The proteins from halophiles have adapted to challenging environmental conditions and require salt for their structure and function. How halophilic proteins adapted to a hypersaline environment is still an intriguing question. In this study, the enzyme *HvDHFR1* was used as a model enzyme to exploit the relationship between the activity and flexibility. The ¹⁵N relaxation studies at 3.5M and 1.0M NaCl concentrations address the effect of low salt concentration on enzyme flexibility, in turns enzymatic activity.

The enzyme *HvDHFR1* was produced in bacterial system by the recombinant protein techniques. It is important to label proteins uniformly and establish spin ensembles over the entire protein so that heteronuclear spin coupling can effectively be

used in triple resonance experiments. Double isotope labeling (^{15}N , ^{13}C) is required for proteins with molecular size <20 kDa. However, the viscosity of 3.5 M NaCl buffer has a significant effect on the correlation time of the proteins and makes the *Hv*DHFR1 to more like a 26 kDa proteins. The proteins with molecular weight > 20 kDa require triple labeling (^{15}N , ^{13}C , ^2H) due to short spin-spin relaxation times and peak broadening. Triple labeled enzyme *Hv*DHFR1 was successfully produced and purified by using size exclusion and anion exchange chromatographic techniques. Appropriate molar salt for halophilic proteins is a vital requirement for these proteins to fold in their native forms, and also for structural stability and activity. It is important to mimic the environmental conditions of the sample under investigation with experimental techniques. The acquisition of NMR data in high ionic strength buffers has been problematic. The electrical current generated by high ionic strength sample inductively couples with the radio frequency coils which make it impossible to tune and match the probe to the correct frequency. The inductive coupling between the receiver coil and sample in a high ionic strength solvent was minimized by using an NMR reduction sample tube. This tube geometry provided excellent tuning and matching of the probe. The fill factor generated by volume reduction is compensated by increasing the protein concentration to 2.0 mM.

Proton-proton distance restraints are the main geometrical information used in structure calculations and the derived via NOESY experiments. In order to obtain these restraints, backbone and side chain atoms chemical shifts must be assigned. For backbone atoms ($^1\text{H}^{\text{N}}$, $^{15}\text{N}^{\text{H}}$, C^{α} , C') chemical shift assignments were accomplished by using 3D HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO experiments. The assigned chemical shifts were deposited to Biological Magnetic

Resonance Bank (BMRB) under the accession code of 6645. This chemical shift values were used in 3D ^1H - ^{15}N NOESY-HSQC and ^1H - ^{13}C NOESY-HSQC experiments to obtain the proton-proton distance restraints. A total of 1371 distance restraints were obtained. Using the chemical shift information, 185 dihedral angle restraints were obtained via program TALOS. These restraints were used in the structure calculation program CNS. The calculated structure showed similar folding other DHFR variants where four helices are packed across a central β -sheet composed from eight β -strands. The atomic coordinates have been deposited in the RCSB Protein Data Bank under the code of 2ITH. The quality of the structures were checked by superimposing the atomic coordinates and the program PROCHECK. The final family of structures exhibits a good convergence with a root-mean-square-deviation of 0.83 ± 0.27 Å for overall backbone and 0.37 ± 0.12 Å for the secondary structure regions. The quality of the structures was evaluated using PROCHECK which showed that 96.2% of the residues fall in allowed and additionally allowed regions of the Ramachandran plot.¹⁰² Two different segments of the NMR structures did show a deviation from the crystal structure. First, there was a deviation in the secondary structure at the N-terminal residues of $\beta 3$ (S59-M64). This divergence was considered to arise from the flexibility of the $\alpha 2$ - $\beta 3$ loop. Secondly, there was a difference in the type of β -turn connecting $\beta 7$ and $\beta 8$ (H147-F150). The dihedral angles of the crystal structure defines a Type-II β -turn⁸⁵ while the NMR structures define a Type-I β -turn. This places this turn in closer proximity to the N-terminus of $\alpha 1$. The Type-I β -turn in the NMR structures causes the side chain of

E148 to come in close proximity to $\alpha 1$ while in the crystal structure this residue points away from $\alpha 1$ which could play a role in DHF binding.

It has been proposed that the activity loss for *Hv*DHFR1 as the salt concentration is decreased from 3.5 M to 1.0 M NaCl is due to the changes in the inherent enzyme flexibility that facilitates the catalytic activity.⁹ This hypothesis was tested by carrying out the ¹⁵N relaxation studies of *Hv*DHFR1 at both 3.5 M and 1.0 M NaCl. At both salt concentrations, the relaxation parameters T_1 and T_2 for individual residues were derived from a series of modified ¹H-¹⁵N HSQC spectrum acquired at different relaxation delay times. The ¹H-¹⁵N heteronuclear NOE parameters were obtained with the proton saturated and non-saturated HSQC experiments. Those parameters were input into the Modelfree program to calculate the squared order parameter for individual ¹⁵N nuclei. The magnitude of order parameter specifies the degree of motion. The calculation for 3.5 M NaCl salt condition showed that there are four mobile regions the L21, $\alpha 2$ - $\beta 3$, ABL, and $\beta 6$ - $\beta 7$ loops. From both the crystal and NMR dynamic studies of *Ec*DHFR, it is known that the loop region motion plays a key role in the catalytic activity. The calculation at 1.0 M NaCl concentration showed that there is a decrease in the magnitude of motion at the loop regions which can be account for the loss of activity. The weakened interaction of hydrophobic residues at the loop regions can be the one of the reason causing the decreased flexibility. The other possible interaction that plays a significant role in the decreased flexibility is electrostatic interaction. Halophilic enzyme studies stated the significance of electrostatic interaction and its effect on the halophilic enzyme stability and activity.⁹ However, these studies have not been able to quantify the role of electrostatic interaction on enzyme structure and activity at low salt concentration.

The effects of low salt concentration on halophilic enzyme structures have not been studied and explained at the molecular level. The ^{15}N relaxation studies at 1.0 M NaCl concentration address the lack of information regarding the loss of activity at low salt concentration. It was believed that as salt concentration is lowered, the enzyme structure destabilizes and this causes the deactivation of halophilic enzyme. The ^{15}N relaxation studies in this work showed that the enzyme can tolerate low salt concentration (retain ternary structure) and the repulsive electrostatic interaction between similarly charged groups results in the loss of flexibility.

Understanding the role of different salt concentration on halophilic enzyme structure and activity will assist in the design and modification of an enzyme to be utilized in different biotechnological and industrial applications in the area of fermentation, waste water treatment, environmental protection and novel medical applications. Since the enzyme *HvDHFR1* function at low water activity, it can be used in reduction processes in non-aqueous medium in relevant industrial applications.

These studies set the framework for further studies to elucidate the salt dependent activity of *HvDHFR1*. This is an apo-state structure and relaxation study. For a more detailed understanding, more relaxation studies at salt concentration between 3.5 M and 1.0 M salt concentration are required. The structure and relaxation studies with *HvDHFR1*:folate, *HvDHFR1*:NADPH and ternary *HvDHFR1*:folate:NADPH will give a complete understanding of protein activity at extreme salt condition and the relation between the activity and salt dependence. Relaxation measurements at different salt concentrations of these complexes will aid in understanding the salt dependent flexibility which could be linked to activity, electrostatics, and halophilic adaptation.

REFERENCES

- (1) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679-6685.
- (2) Palmer, A. G., 3rd *Curr. Opin. Struct. Biol.* **1997**, *7*, 732-737.
- (3) LeMaster, D. M.; Kay, L. E.; Brunger, A. T.; Prestegard, J. H. *FEBS Lett.* **1988**, *236*, 71-76.
- (4) Peng, J. W.; Wagner, G. *Biochemistry* **1995**, *34*, 16733-16752.
- (5) Wagner, G.; Thanabal, V.; Stockman, B. J.; Peng, J. W.; Nirmala, N. R.; Hyberts, S. G.; Goldberg, M. S.; Detlefsen, D. J.; Clubb, R. T.; Adler, M. *Biopolym.* **1992**, *32*, 381-390.
- (6) Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E. *J. Biomol. NMR* **1995**, *6*, 153-162.
- (7) Boehr, D. D.; Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2006**, *106*, 3055-3079.
- (8) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 796-801.
- (9) Wright, D. B.; Banks, D. D.; Lohman, J. R.; Hilsenbeck, J. L.; Gloss, L. M. *J. Mol. Biol.* **2002**, *323*, 327-344.
- (10) Blecher, O.; Goldman, S.; Mevarech, M. *Eur. J. Biochem.* **1993**, 199-203.
- (11) Osborne, M. J.; Schnell, J.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2001**, *40*, 9846-9859.
- (12) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638-1642.
- (13) Schnell, J. R.; Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 119-140.
- (14) Jones, D. T. *Science* **2003**, *302*, 1347-1348.

- (15) Yu, H.; Kohl, A.; Binz, H. K.; Pluckthun, A.; Grutter, M. G.; van Gunsteren, W. F. *Proteins* **2006**, *65*, 285-295.
- (16) Lattman, E. E. *Proteins* **2001**, *42*, 295.
- (17) Berendsen, H. J.; Hayward, S. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165-169.
- (18) Parak, F. G. *Curr. Opin. Struct. Biol.* **2003**, *13*, 552-557.
- (19) Walczak, A. M.; Antosiewicz, J. M. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2002**, *66*, 51911.
- (20) Gilquin, B.; Bourgoïn, M.; Menez, R.; Le Du, M. H.; Servent, D.; Zinn-Justin, S.; Menez, A. *Protein Sci.* **2003**, *12*, 266-277.
- (21) Wagner, G. *Nat. Struct. Biol.* **1997**, *4*, 841-844.
- (22) Kay, L. E. Torchina, D.A., Bax, A.; *Biochemistry* **1989**, *28*, 8972-8979.
- (23) Service, R. F. *Science* **2000**, *288*, 939-941.
- (24) Braxenthaler, M.; Unger, R.; Auerbach, D.; Given, J. A.; Moult, J. *Proteins* **1997**, *29*, 417-425.
- (25) Doniach, S.; Eastman, P. *Curr. Opin. Struct. Biol.* **1999**, *9*, 157-163.
- (26) Ladurner, A. G.; Itzhaki, L. S.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8473-8478.
- (27) Beck, D. A.; Daggett, V. *Methods* **2004**, *34*, 112-120.
- (28) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins* **2000**, *39*, 252-260.
- (29) Sullivan, D. C.; Kuntz, I. D. *Proteins* **2001**, *42*, 495-511.
- (30) Vendruscolo, M.; Dobson, C. M. *Philos. Transact. A Math. Phys. Eng. Sci.* **2005**, *363*, 433-450.
- (31) Wang, Y.; Iii, T. G. *J. Phys. Chem. B Condens. Matter Mater Surf. Interfaces Biophys.* **2007**, *111*, 327-330.
- (32) van der Weij-De Wit, C. D.; Doust, A. B.; van Stokkum, I. H.; Dekker, J. P.; Wilk, K. E.; Curmi, P. M.; Scholes, G. D.; van Grondelle, R. *J. Phys. Chem. B Condens. Matter Mater. Surf. Interfaces Biophys.* **2006**, *110*, 25066-25073.
- (33) Loura, L. M.; Ramalho, J. P. *Biochim. Biophys. Acta* **2006**.
- (34) Kim, T. Y.; Winkler, K.; Alexiev, U. *Photochem. Photobiol.* **2006**.

- (35) S., G. F.; Ulrike, A.; Helmut, G. *Biophys. J.* **2005**, *89*, 3757–3770.
- (36) Chattopadhyay, K.; Frieden, C. *Proteins* **2006**, *63*, 327-335.
- (37) Alexiev, U.; Rimke, I.; Pohlmann, T. *J. Mol. Biol.* **2003**, *328*, 705-719.
- (38) Traaseth, N. J.; Buffy, J. J.; Zamoon, J.; Veglia, G. *Biochemistry* **2006**, *45*, 13827-13834.
- (39) Lorieau, J. L.; McDermott, A. E. *J. Am. Chem. Soc.* **2006**, *128*, 11505-11512.
- (40) McDermott, A. E. *Curr. Opin. Struct. Biol.* **2004**, *14*, 554-561.
- (41) Mack, J. W.; Usha, M. G.; Long, J.; Griffin, R. G.; Wittebort, R. J. *Biopolym.* **2000**, *53*, 9-18.
- (42) Knapp, J. E.; Pahl, R.; Srajer, V.; Royer, W. E., Jr. *Proc Natl. Acad. Sci. U.S. A.* **2006**, *103*, 7649-7654.
- (43) Srajer, V.; Ren, Z.; Teng, T. Y.; Schmidt, M.; Ursby, T.; Bourgeois, D.; Pradervand, C.; Schildkamp, W.; Wulff, M.; Moffat, K. *Biochemistry* **2001**, *40*, 13802-13815.
- (44) Ren, Z.; Perman, B.; Srajer, V.; Teng, T. Y.; Pradervand, C.; Bourgeois, D.; Schotte, F.; Ursby, T.; Kort, R.; Wulff, M.; Moffat, K. *Biochemistry* **2001**, *40*, 13788-13801.
- (45) Schmidt, M.; Ihee, H.; Pahl, R.; Srajer, V. *Methods Mol. Biol.* **2005**, *305*, 115-154.
- (46) Rizzuti, B.; Sportelli, L.; Guzzi, R. *Biophys. Chem.* **2006**.
- (47) Smith, L. J.; Davies, R. J.; van Gunsteren, W. F. *Proteins* **2006**, *65*, 702-711.
- (48) Brown, F. K.; Kollman, P. A. *J. Mol. Biol.* **1987**, *198*, 533-546.
- (49) Palmer, A. G., 3rd *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129-155.
- (50) Cavanagh, J.; Fairbrothers, W.; Palmer III, A. G.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: New York, 1996.
- (51) Clore, G. M.; Gronenborn, A. M. *NMR of Proteins*; CRC Press: Boca Ratone, FL, 1993.
- (52) John Cavanagh, W. J. F., III, Arthur G. Palmer, Nicholas J. Skelton *Protein NMR Spectroscopy: Principles and Practice* Academic Press: New York, 1996.
- (53) Reid, D. G. *Protein NMR Techniques*; Human Press: Totowa, NJ, 1997.

- (54) Roberts, G. C. K. *NMR of Macromolecules: A practical Approach*; Oxford Univeristy Press: New York, 1993.
- (55) Teng, Q. *Structral Biology: Practical NMR Application*; Springer: New York, 2005.
- (56) van den Ven, F. J. M. *Multidimensional NMR in Liquids: Basic Principles and Experimental Methods*; Wiley-VCH: New York, 1995.
- (57) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559
- (58) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559-45570
- (59) Levitt, M. H. *Spin Dynamics: Basic of Nuclear Magnetic Resonance*; Wiley & Sons: New York, 2005.
- (60) Wang, T.; Weaver, D. S.; Cai, S.; Zuiderweg, E. R. *J. Biomol. NMR* **2006**, *36*, 79-102.
- (61) Vasos, P. R.; Hall, J. B.; Kummerle, R.; Fushman, D. *J. Biomol. NMR* **2006**, *36*, 27-36.
- (62) Loth, K.; Pelupessy, P.; Bodenhausen, G. *J. Am. Chem. Soc.* **2005**, *127*, 6062-6068.
- (63) Wider, G. *Methods Enzymol.* **2005**, *394*, 382-398.
- (64) Bouguet-Bonnet, S.; Mutzenhardt, P.; Canet, D. *J. Biomol. NMR* **2004**, *30*, 133-142.
- (65) Tugarinov, V.; Scheurer, C.; Bruschiweiler, R.; Kay, L. E. *J. Biomol. NMR* **2004**, *30*, 397-406.
- (66) Abragam, A. *Principle of Nuclear Magnetic*; Clarendon Press: Oxford, 1961.
- (67) Riek, R.; Wider, G.; Pervushin, K.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4918-4923.
- (68) Luy, B.; Glaser, S. J. *J. Magn. Reson.* **2000**, *142*, 280-287.
- (69) Cordier, F.; Caffrey, M.; Brutscher, B.; Cusanovich, M. A.; Marion, D.; Blackledge, M. *J. Mol. Biol.* **1998**, *281*, 341-361.
- (70) Blake-Hall, J.; Walker, O.; Fushman, D. *Methods Mol. Biol.* **2004**, *278*, 139-160.
- (71) Sahu, S. C.; Bhuyan, A. K.; Udgaonkar, J. B.; Hosur, R. V. *J. Biomol. NMR* **2000**, *18*, 107-118.

- (72) Lee, A. L.; Wand, A. J. *J. Biomol. NMR* **1999**, *13*, 101-112.
- (73) Luginbuhl, P.; Pervushin, K. V.; Iwai, H.; Wuthrich, K. *Biochemistry* **1997**, *36*, 7305-7312.
- (74) Chen, J.; Brooks, C. L., 3rd; Wright, P. E. *J. Biomol. NMR* **2004**, *29*, 243-257.
- (75) Peng, J. W.; Wagner, G. *Biochemistry* **1992**, *31*, 8571-8586.
- (76) King, R.; Jardetzky, O. *J. Chem. Phys. Lett.* **1978**, *55*, 15-18.
- (77) King, R.; Maas, R.; Gassner, M.; Nanda, R. K.; Conover, W. W.; Jardetzky, O. *Biophys.* **1978**, *6*, 103-117.
- (78) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C. *J. Am. Chem. Soc.* **1990**, *112*, 4989-4991.
- (79) Andrec, M.; Montelione, G. T.; Levy, R. M. *J. Magn. Reson.* **1999**, *139*, 408-421.
- (80) Andrec, M.; Montelione, G. T.; Levy, R. M. *J. Biomol. NMR* **2000**, *18*, 83-100.
- (81) Mandel, A. M.; Akke, M.; Palmer, A. G., 3rd *J. Mol. Biol.* **1995**, *246*, 144-163.
- (82) Palmer, A. G., 3rd *Chem. Rev.* **2004**, *104*, 3623-3640.
- (83) Palmer, A. G., 3rd; Kroenke, C. D.; Loria, J. P. *Methods Enzymol.* **2001**, *339*, 204-238.
- (84) Mulder, F. A.; van Tilborg, P. J.; Kaptein, R.; Boelens, R. *J. Biomol. NMR* **1999**, *13*, 275-288.
- (85) Pieper, U.; Kapadia, G.; Mevarech, M.; Herzberg, O. *Structure* **1998**, *6*, 75-88.
- (86) Rajagopalan, P. T.; Lutz, S.; Benkovic, S. J. *Biochemistry* **2002**, *41*, 12618-12628.
- (87) Cameron, C. E.; Benkovic, S. J. *Biochemistry* **1997**, *36*, 15792-15800.
- (88) Swanwick, R. S.; Shrimpton, P. J.; Allemann, R. K. *Biochemistry* **2004**, *43*, 4119-4127.
- (89) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586-603.
- (90) Osborne, M. J.; Venkitakrishnan, R. P.; Dyson, H. J.; Wright, P. E. *Protein Sci.* **2003**, *12*, 2230-2238.
- (91) Margesin, R.; Schinner, F. *Extremophiles* **2001**, *5*, 73-83.
- (92) Dym, O.; Mevarech, M.; L., S. J. *Science* **1995**, *267*, 1344-1346.

- (93) Richard, S. B.; Madern, D.; Garcin, E.; Zaccai, G. *Biochemistry* **2000**, *39*, 992-1000.
- (94) Frolow, F.; Harel, M.; Sussman, J. L.; Mevarech, M.; Shoham, M. *Nat. Struct. Biol.* **1996**, *3*, 452-458.
- (95) Ebel, C.; Costenaro, L.; Pascu, M.; Faou, P.; Kernel, B.; Proust-De Martin, F.; Zaccai, G. *Biochemistry* **2002**, *41*, 13234-13244.
- (96) Ebel, C.; Faou, P.; Kernel, B.; Zaccai, G. *Biochemistry* **1999**, *38*, 9039-9047.
- (97) Elcock, A. H.; McCammon, J. A. *J. Mol. Biol.* **1998**, *280*, 731-748.
- (98) Madern, D.; Ebel, C.; Zaccai, G. *Extremophiles* **2000**, *4*, 91-98.
- (99) Mevarech, M.; Frolow, F.; Gloss, L. M. *Biophys. Chem.* **2000**, *86*, 155-164.
- (100) Rao, J. K.; Argos, P. *Biochemistry* **1981**, *20*, 6536-6543.
- (101) Eisenberg, H.; Mevarech, M.; Zaccai, G. *Biochemical, Structural, and Molecular Genetic Aspects of Halophilism*. In *Advances in Protein Chemistry*; Academic Press: New York, 1992; pp 1-57.
- (102) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283-291.

CHAPTER VI

EXPERIMENTAL

Protein Production and Purification

The expression vector pET-11d containing DNA of *HvDHFR1* (obtained from Dr Lisa M. Gloss, Washington State University) was transformed to *Escherichia coli* strain, Rosetta(DE3)pLysS (Novagen) via CaCl₂ procedure. Antibiotics Ampicillin (Rpi) and Chloramphenicol (Rpi) were prepared at a concentration of 34 mg/ml and used at a ratio of 1/1000 for all Luria-Bertani (LB) and minimal media (M9) mediums. The cells were grown in 3L of LB medium, with appropriate antibiotics, at 37 °C until OD₆₀₀ reached 0.4. The growth was then spun down at 15000 rpm and resuspended in 1L of minimal media to produce ²H/¹³C/¹⁵N labeled protein. The minimal media contained 1L of D₂O (70%) as ²H source, 2.0 g/L of ¹³C₆-glucose (Cambridge Isotope Inc.) as the sole source of ¹³C and 0.5 g/L of ¹⁵NH₄Cl (Cambridge Isotope Inc.) as the sole source of ¹⁵N isotopes. This media was allowed to grow for 1 hour, induced with 0.4M IPTG (isopropyl β-D-1-thiogalactopyranoside) and grown overnight. The cells were precipitated by centrifugation. After the cell pellet was stirred in the presence of DeoxyribonucleaseI (Rpi) in 10 mM phosphate buffer for 20 min., it was lysed by sonication five times each of which was 1 min. The cell debris was removed from the lysate by centrifugation at 15000 rpm and the precipitate was dissolved in 6 M guanidine

buffer ($\text{CH}_3\text{N}_3\cdot\text{HCl}$, $\text{pH}=8.0$). The unfolded proteins were refolded by adding in 1 M NaCl buffer dropwise. This solution was dialyzed against 4 L of 2.5 M $(\text{NH}_4)_2\text{SO}_4$ and 10 mM tris-HCl buffer at $\text{pH}=7.0$ in dialysis tubes with MWCO of 10000 for overnight at 4 °C. The solution was centrifuged at 15000 rpm and the supernatant containing the proteins was further purified by column chromatography. Initially, size exclusion chromatography was accomplished using the column material Sepharose CL-6B as a gel matrix (Sigma-Aldrich). A 2.5 M $(\text{NH}_4)_2\text{SO}_4$ (100 ml)/1.2 M $(\text{NH}_4)_2\text{SO}_4$ (200 ml) gradient was run.¹ The fractions were collected at 50 ml solution and for each fraction the UV-Vis absorbance (Varian Cary1.0) at 280 nm was measured. The fractions with high absorbance were subjected to SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Electrophoresis) techniques. The fractions with the protein sample were collected and submitted to a second column. As a column material DEAE-cellulose (diethylamino ethyl)(Sigma-Aldrich) was used as an anion exchanger.¹ A 3.5 M NaCl buffer was run to elute the bound proteins from the column. For each fraction, UV-Vis absorbance at 280 nm was measured and the fractions with high absorbance were subjected to SDS-PAGE electrophoresis. The fractions with protein samples were pulled together and sample concentration was brought down to the desired levels using concentrators (Centricon Plus-80 and Centriprep from Amicon Inc with 10,000 MWCO). 10% of sample volume of D_2O (Sigma) was added to the protein sample for deuterium lock during the data acquisition. The final protein sample was concentrated to 2.0 mM and the solvent exchanged to 3.5 M NaCl in 10 mM tris at $\text{pH} 7.0$ in 95 % H_2O /5 % D_2O .

NMR Data Collection

All NMR data were collected either on a Varian Inova spectrophotometer operating at the base frequency of 499.86 MHz with a triple resonance (^1H , ^{13}C , ^{15}N) z-gradient probe or Bruker Avance spectrometer operating at the base frequency of 600.133 MHz with triple resonance (^1H , ^{13}C , ^{15}N) z-gradient probe at the Center for NMR Spectroscopy at Washington State University at 25 °C. All NMR data sets were processed and analyzed using the NMRPipe software² and viewed with PIPP³ on a Silicon Graphics Octane workstation. A 90° shifted sine-bell squared window function was applied in all dimension data. Prior to Fourier transformation, all data were zero-filled to 2048, 512, and 126 in each dimension except the HCCH-TOCSY experiment. The data were zero filled to 2048, 512, and 256 data points in the HCCH-TOCSY experiment.

- **HSQCs:**⁴ The 2D ^1H - ^{15}N -HSQC experiment was acquired with a total of 1024 complex data points in t2 (^1H) and 128 data blocks in t1 (^{15}N). For each data block 32 scans were recorded at spectral widths of 9614.385 (^1H) and 2149.96 Hz (^{15}N) in each dimension, respectively. The carrier frequencies were set to 4.773 ppm (^1H) and 117.87 ppm (^{15}N). The 2D ^1H - ^{13}C HSQC experiment was acquired with a total of 1024 complex data points in t2 (^1H) and 128 data blocks in t1 (^{13}C). For each data block 32 scans were recorded at spectral widths of 8000.00 (^1H) and 17597.095 Hz (^{13}C) in each dimension, respectively. The carrier frequencies were set to 4.773 ppm (^1H) and 37.058 ppm (^{13}C).
- **HNCA and HN(CO)CA:**^{5,6} The 3D HNCA experiment was acquired with a total of 1024 complex data points in t3 (^1H) for 64 data blocks in t2 (^{15}N) and 128 data blocks in t1 (^{13}C). The 3D HN(CO)CA experiment was acquired with a total of 1024 complex data

points in t3 (^1H) for 112 blocks of data in t2 (^{15}N) and 64 data blocks in t1 (^{13}C). For both experiments each block of data were recorded for 32 scans at spectral widths of 9615.400 (^1H), 2150.075 (^{15}N), and 6036.825 Hz (^{13}C) in each dimension, respectively. The carrier frequencies were set to 4.772 ppm (^1H), 117.086 ppm (^{15}N), and 56.738 ppm (^{13}C). The quadrature data were acquired in State-TPPI mode and MLEV pulse train used for decoupling during the data acquisition. The data acquisition was completed in 72.8 h (3 days, 50 min.) for HNCACB and 91.0 h (3 days, 50 min.) for HN(CO)CA.

- **HNCACB and CBCA(CO)NH:**^{7,8} Both the 3D HNCACB and 3D CBCA(CO)NH experiment were acquired with a total of 1024 complex data points in t3 (^1H) for 64 data blocks in t2 (^{15}N) and 128 data blocks in t1 (^{13}C). For each block of data 32 scans for HNCACB and 40 scans for CBCA(CO)NH were recorded at spectral widths 9615.385 (^1H), 2150.075 (^{15}N), and 11312.217 Hz (^{13}C) in each dimension, respectively. The carrier frequencies were set to 4.773 ppm (^1H), 41.740 ppm (^{13}C), and 117.086 ppm (^{15}N). The quadrature data were acquired in State-TPPI mode and MLEV pulse train used for decoupling during the data acquisition. The data acquisition was completed in 72.8 h (3 days, 50min.) for HNCACB and 91.0 h (3 days, 19.0h.) for CBCA(CO)NH.

- **HNCO and HN(CA)CO:**⁵ The 3D HNCO experiment was acquired with a total of 1024 complex data points in t3 for 64 data blocks in t2 (^{15}N) and 100 data blocks in t1 (^{13}C). The 3D HN(CA)CO experiment was acquired with a total of 1024 complex data points in t3 (^1H) for 64 data blocks in t2 (^{15}N) and 96 (^{13}C) data blocks in t1. For each block of data 20 scans for HNCO and 32 for HN(CA)CO were recorded at spectral widths of 9615.385 (^1H), 2150.075 (^{15}N), and 2399.952 Hz (^{13}C) in each dimension, respectively. The carrier frequencies were set to 4.773 ppm (^1H), 117.086 ppm (^{15}N), and

176.714 ppm (^{13}C). The quadrature data were acquired in State-TPPI mode and MLEV pulse train used for decoupling during the data acquisition. The data acquisition was completed in 35.5 h. (1 day, 11.5 h.) for HNCO and in 72.8 h (3 days, 50 min.) for HN(CA)CO.

- **(H)C(CO)NH-TOCSY:**⁹⁻¹² The 3D (H)C(CO)NH-TOCSY experiment was acquired with a total of 1024 complex data points in t3 (^1H) for 64 data blocks in t2 (^{15}N) and 128 data blocks in t1 (^{13}C). For each block of data 32 scans were recorded at spectral widths of 9615.400 (^1H), 2150.075 (^{15}N), and 11318.619 Hz (^{13}C) in each dimension, respectively. The carrier frequencies were set to 4.773 ppm (^1H), 150.911 ppm (^{13}C), and 117.086 ppm (^{15}N). The quadrature data were acquired in State-TPPI mode and MLEV pulse train used for decoupling during the data acquisition. The data acquisition was completed in 91.0 h (3 days, 50min.).

- **H(C-CO)NH-TOCSY:**⁹⁻¹² The 3D H(C-CO)NH-TOCSY experiment was acquired with a total of 1024 complex data points in t3 (^1H) for 64 data blocks in t2 (^{15}N) and 200 data blocks in t1 (^1H). For each block of data 64 scans were recorded at spectral widths of 8000.00 (^1H), 1800.018 (^{15}N), and 7800.00 Hz (^1H) in each dimension respectively. The carrier frequencies were set to 4.773 ppm (^1H), 119.800 ppm (^{15}N), and 4.77 ppm (^1H). The quadrature data were acquired in Rance-Kay mode and MLEV pulse train used for decoupling during the data acquisition. The data acquisition was completed in 5 days, 9 h.

- **HCCH TOCSY:**^{5, 13} 3D HCCH-TOCSY experiment was acquired with a total of 1024 complex data points in t3 for 96 data blocks in t2 (^{13}C) and 256 data blocks in t1 (^1H). For each block of data 64 scans were recorded at spectral widths of 8000.00 (^1H),

10055.30 (^{13}C) and 6600.00 Hz (^1H) in each dimension respectively. The carrier frequencies were set to 4.773 ppm (^1H), 37.078 ppm (^{13}C), and 4.77 ppm (^1H). The quadrature data were acquired in Complex mode and MLEV pulse train used for decoupling during the data acquisition. The data acquisition was completed in 126 h (5 days, 6h).

- **^{15}N NOESY-HSQC:**^{9,14} The 3D ^{15}N NOESY-HSQC experiment was acquired with a total of 1024 complex data points in t3 (^1H) for 64 data blocks in t2 (^{15}N) and 128 data blocks in t1 (^1H). For each block of data 32 scans were recorded at spectral widths of 9615.400 (^1H), 2150.075 (^{15}N), and 8000.00 Hz (^1H) in each dimension, respectively. Mixing time was set to 150 ms. The carrier frequencies were set to 4.773 ppm (^1H), 117.086 ppm (^{15}N), and 4.773 ppm (^1H). The quadrature data were acquired in State-TPPI mode and MLEV pulse train used for decoupling during the data acquisition. The data acquisition was completed in 91.0 h (3 days, 50 min.).

- **^{13}C NOESY-HSQC:**^{9,14} The 3D ^{13}C NOESY-HSQC experiment was acquired with a total of 1024 complex data points in t3 (^1H) for 96 data blocks in t2 (^{13}C) and 224 data blocks in t1 (^1H). For each block of data 16 scans were recorded at spectral widths of 8000.00 (^1H), 10055.304 (^{13}C), and 8000.00 Hz (^1H) in each dimension, respectively. The carrier frequencies were set to 4.773 ppm (^1H), 37.090 ppm (^{13}C), and 4.773 ppm (^1H). Mixing time was set to 150 ms. The quadrature data were acquired in Complex mode and MLEV pulse train used for decoupling during the data acquisition. The data acquisition was completed in 124 h (5 days, 4h.).

Protocols for Structure Calculation by CNS

The following steps were followed in order to calculate the protein structure using the CNS program.¹⁵ The one letter codes of amino acids sequence were edited in a file under the name of *hv.seq* to generate protein topology file for the structure calculation. This was done using the *generate-seq.inp* protocols. A protein topology file under the name of *hv.mtf* was generated and used in the next step. The generated *hv.mtf* file includes molecular connectivity information which was used as an input file in the next step. An extended protein structure data file was generated by using the *generate-extended.inp* protocol. The generated structure file, *hv.pdf*, is energy minimized starting structure which bears the covalent geometric information (bond lengths and bond angles) to be used during the Torsion Angle Dynamic phase in the CNS protocol. The CNS structure calculation was implemented by running the *anneal.inp* protocol. The protocol *anneal.inp* contains input file names of the NOE and the dihedral restrains, and the temperature program used in simulated annealing method. The calculation started from an extended conformation based on the amino acid sequence followed by an initial Powell minimization to normalize the geometry. Structures were then input into torsion angle dynamics (TAD). This TAD phase was performed at 50000 K for 1000 molecular dynamics steps, each at 15 fs, using the default force constant for the energy and the experimental functions. The structures were then cooled using TAD from 50000 K to 2000 K over 60000 steps, each at 5 fs while ramping the force constant with default values. Further cooling from 2000 K to 0 K was performed at 1000 steps, each at 2 fs in cartesian space. Again, default CNS values were used to ramp the force constants. The final stage of structure calculation employs Powell minimization.

Extraction of Relaxation Data with the Software Felix and Protocols for the Modelfree Formalism

NMR measurements were performed on a Varian Inova 500 MHz spectrophotometer at the Center for NMR Spectroscopy at Washington State University at 25 °C. A series of modified 2D ^1H - ^{15}N -HSQC spectrum were recorded. A total of 1024 complex data points in t_2 were acquired for 128 data blocks in t_1 . For each data block 32 scans were recorded at spectral widths of 8000.00 and 2000.00 Hz in each dimension, respectively. The carrier frequencies were set to 4.773 ppm (^1H) and 117.87 ppm (^{15}N). A Gaussian window function was applied to ^1H dimension and zero filled prior to Fourier transformation. The T_1 , T_2 and heteronuclear NOE relaxation data were processed and analyzed based on their volumes by the software Felix and converted to *.mat format. The R_1 and R_2 parameters were extracted by processing the seven spectra (for T_1 : 50*, 100, 200*, 400, 700, 1000, and 1500 ms; for T_2 : 50, 70*, 90, 110*, 150, 190, 250 ms). The steady-state heteronuclear NOEs were calculated from processing the spectra with and without proton saturation. The uncertainties in peak intensities were evaluated from duplicate measurements (indicated with *). The derived relaxation parameters, R_1 , R_2 , and ^1H - ^{15}N heteronuclear NOE, were used in Modelfree program. The PERL scripts provided by Dr Scott Scowalter (Florida State University) were used to prepare and evaluate the data for Modelfree analysis. All calculations were performed on a SGI workstation. Initially, overall molecular correlation time, τ_m , was approximated from R_2 and R_1 relaxation data. The τ_m was calculated from R_2/R_1 ratio of the individual ^{15}N nucleus since this parameter is closely relate to the overall tumbling part of spectral

density function. Any residue was excluded if $\text{NOE} < 0.65$, $\langle T_1 \rangle - T_{1,n} > 1 \text{ SD}$, and $\langle T_2 \rangle - T_{2,n} > 1 \text{ SD}$ where SD is the standard deviation, and $\langle T_1 \rangle$ and $\langle T_2 \rangle$ are the average values. For the subset residues the *r1r2_diffusion* protocol was used to estimate τ_m . The overall correlation time was estimated as 9.0 ns at 3.5 M NaCl and 8.8 ns at 1.0 M NaCl. The isotropic diffusion model was verified using the *diffusion* protocol. The $D_{\text{parallel}}/D_{\text{perpendicular}}$ ratio was calculated as 1.1 at 3.5 M NaCl and 1.3 at 1.0 M NaCl. The *mfdata.pl* and *mfpar.pl* scripts were used to generate data and parameter files for individual residue. The Modelfree program was run for five models. Model selection for individual residue was accomplished by running the *mod_sel.pl* scripts. The Modelfree program was run for the optimized selection to derive the order parameters.

REFERENCES

- (1) Eisenberg, H.; Mevarech, M.; Zaccai, G. *Biochemical, Structural, and Molecular Genetic Aspects of Halophilism*. In *Advances in Protein Chemistry*; Academic Press: New York, 1992; pp 1-57.
- (2) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.
- (3) Garrett, D. S.; Powers, R.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1991**, *95*, 214-220.
- (4) Palmer III, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1990**, *93*, 151-170.
- (5) Clore, G. M.; Bax, A.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 8172-8184.
- (6) Yamazaki, T.; Lee, W.; Revingtom, M.; Mattiello, D. L.; Dahlquist, F. W. *J. Am. Chem. Soc.* **1994**, *116*, 6464-6465.
- (7) Wittekind, M.; Mueller, L. *J. Magn. Reson.* **1993**, *B101*, 201-205.
- (8) Meissner, A.; Sorensen, O. W. *J. Magn. Reson.* **2001**, *151*, 328-331.
- (9) Muhandiram, D. R.; Kay, L. E. *J. Magn. Reson.* **1994**, *B103*, 203-216.
- (10) Grzesiek, S.; Dobeli, H.; Gentz, R.; Garotta, G.; Labhardt, A. M.; Bax, A. *Biochemistry* **1992**, *31*, 8180-8190.
- (11) Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1992**, *99*, 201-207.
- (12) Grzesiek, S.; Anglister, J.; Bax, A. *J. Magn. Reson., Ser. B.* **1993**, *101*, 114-119.
- (13) Kay, L. E.; Xu, G. Y.; Singer, A. U.; Muhandiram, D. R.; Forman-Kay, J. D. *J. Magn. Reson., Ser. B* **1993**, *101*, 333-337.
- (14) Muhandiram, D. R.; Farrow, N. A.; Xu, G. Y.; Smallcombe, S. H.; Kay, L. E. *J. Magn. Reson. Ser. B* **1993**, *102*, 317-321.

- (15) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr.* **1998**, *54*, 905-921.