To the Graduate Council:

I am submitting herewith a dissertation written by Brad C. Bennett entitled, "High Resolution X-ray and Neutron Crystallographic Studies of *Escherichia coli* Dihydrofolate Reductase." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biochemistry, Cellular and Molecular Biology.

Chris G. Dealwis
Major Professor

We have read this dissertation
and recommend its acceptance:

Elias J. Fernandez

Elizabeth E. Howell

Ronald B. Wetzel

Accepted for the Council:

Anne Mayhew
Vice Chancellor and Dean of
Graduate Studies

(Original signatures are on file with official student records.)

High Resolution X-ray and Neutron Crystallographic Studies of *Escherichia coli* Dihydrofolate Reductase

A Dissertation
Presented for the
Doctor of Philosophy Degree

The University of Tennessee- Knoxville

Brad C. Bennett
December 2005

Acknowledgements

There is not sufficient space to express all the gratitude I have for all who have helped me at UTK and in my life. But I will attempt to note the most important contributors. First, I want to thank Chris Dealwis for allowing me to join his lab and conduct research for *so* long. He continues to help me mature and think scientifically and critically, is always available for my endless questions and for advice, and is a strong motivator. Importantly, he is always interested in the "big picture", asking the tough questions, and brings enthusiasm to the research. This provides segue to the rest of my thesis committee: Elias Fernandez, Liz Howell, Richard Lee and Ron Wetzel. All of them have made invaluable suggestions and contributed constructive criticism (the helpful kind!) of my work and especially of my analyses; for that I am truly grateful. I believe it was conversations between Liz and Chris about neutron crystallography of DHFR which opened up this project for me. Not only did this save me from the Titanic-sinking iceberg that was BACE but it introduced me to X-ray and neutron crystallography. Neutrons are a fascinating tool for structural biologists and I hope that they become more widespread and applicable to all types of biological problems. I graciously thank Liz for all her insight and expertise on DHFR, from protein purification to analysis of the electron and nuclear density maps.

Members of the Dealwis, Fernandez, and Howell labs have been an important source of intellectual and social fellowship. Dr. Anna Gardberg is a great DHFR colleague in the lab, very intelligent, and is a crackshot crystallographer and scripter. Quite frankly, Anna solves problems. And she has been especially kind, helpful, and patient with me. Matt Wilkerson, Lezlee Dice, and Dr. Tomoaki Uchiki are former

members of the Dealwis lab who labored alongside me and provided scientific and humorous stimulation. I would definitely be remiss not to thank Drs. Vibha and Rakesh Gupta, Derike Smiley, Lori Stinnett, Michael Jackson, and Brad Strader and future medical doctor Jeremy Vincent for their suggestions and friendship (yes, Smiley actually has friends!). Since I've been around for far too long, all members of the Dealwis lab, past and present, have helped me at some point. But Israel Huff, James Goodin, Nathan Tanner, Kirril Borziak, Jim Fairman, Joseph Racca, and Richard Simmerman, for better or worse, have all directly participated in BACE or DHFR experiments; thank you for your help and patience. I am thankful to the Joint Institute of Neutron Sciences at UTK for a fellowship bestowed for parts of the 2003-2005 academic years.

Finally, there isn't a day that passes that I don't think how lucky I am for having such a fantastic and supportive (financial and emotional) family. Grandmother, I told you this day would eventually get here. Thanks for all your love! Mom and Dad, thanks for always being there and believing. You are phenomenal people and I am very fortunate to have you as my parents. My sister, Lee Ann, and brother, J.B., are perhaps the funniest people I know (obvious model bias there), actually don't mind talking science (!), and are great siblings and friends. My most special person in the world is Dr. Stephanie Hicks, who continually surprises me with her generosity, sense of humor, and special friendship. There are no words except: I love you, Stephanie, I love you. I thank God for this world, for scientific pursuit, and life in all its forms.

Abstract

Dihydrofolate Reductases (DHFRs) have been identified in nearly every proteome and are essential for most biosynthetic pathways involving one-carbon transfer reactions due to their recycling of tetrahydrofolate (THF). They catalyze the NADPH-dependent reduction of dihydrofolate (DHF), producing THF. Inhibition of DHFR ultimately depletes cellular pools of THF; causing a reduced supply of thymine nucleotides for DNA synthesis, resulting in genomic instability and cell death. Therefore, DHFRs remain important drug targets in antimicrobial and chemotherapeutic treatments. Despite exhaustive investigation of *E. coli* chromosomal DHFR, controversy persists over the dynamics of regulatory loops (the Met20, the βF-βG, and the βG-βH) and the nature of the interaction between methotrexate (MTX), a tight-binding anti-cancer drug, and Asp 27, the only ionizable residue in the active site. Also of importance is the ionization state of Asp 27 in the apoenzyme and other complexes. Hydrogen atoms (H) likely play a critical role in DHFR ligand binding and catalysis, yet are difficult to directly visualize. High resolution X-ray and neutron crystallography have been utilized in this dissertation to provide accurate positions of H within the DHFR active site and to probe dynamics of the enzyme. The ultrahigh resolution X-ray structures of DHFR/MTX (1.0Å; chapter 4), apo DHFR (1.05Å), and DHFR/MTX/NADPH (1.4Å; both chapter 5) have been solved. Novel features were observed in the electron density maps, including the ability to model the Met20 loop in the apoenzyme as closed (reported disordered previously) and alternate side chain conformations in all the structures. The high data-to-parameter ratio of the apoenzyme and the MTX data sets allowed anisotropic B-factor refinement and full-matrix refinement to calculate carboxylate bond lengths and estimates of their deviations.

The apoenzyme has highly different bond lengths for its Asp 27 carboxylate, thus, it is neutral at physiological pH. The carboxylate bond lengths of the Asp 27 in both the monomers of the asymmetric unit of the DHFR/MTX crystal are nearly equal, suggesting it is charged at physiological pH.

If H is substituted for deuterium (D), neutrons are especially powerful probes due to D's strong positive scattering length. To assign protonation states to the MTX and the Asp 27 by the *direct* identification of D, a neutron structure has been solved to 2.2Å resolution from nearly 80% complete data collected on a 0.3mm^3 crystal (chapter 4). Prerequisite to the neutron experiment was the growth and D_2O-soaking of large-volume crystals (chapter 3). The DHFR/MTX cocrystal possesses the largest primitive unit cell and is the smallest D_2O-soaked crystal used successfully in a neutron diffraction experiment. This is the 11th novel protein ever to be solved by neutron crystallography (the 16th total structure). Nearly 2/3 of the amide backbone has undergone H/D exchange, an indicator of protein dynamics. However, monomer B, where the Met20 loop is closed, is ~10% more exchanged than monomer A, where the Met20 loop is partially occluded. Based on results from D occupancy refinement and analysis of the neutron maps, it is concluded that the MTX N1 is protonated when bound to DHFR. Paired with the X-ray data, this is new strong evidence that the Asp 27•MTX interaction is ionic in nature.

To increase the signal-to-noise ratio in future neutron experiments, perdeuterated protein has been produced and its D enrichment measured by mass spectrometry. X-ray data (to 1.2Å) has now been collected on a perdeuterated DHFR/MTX cocrystal and it is isomorphous to the native cocrystals (chapter 3).

Table of Contents

Chapter 1.
Introduction……...1
1.1 Pyrimidine nucleotide & amino acid biosynthesis and tetrahydrofolate cofactors……..1
1.2 General aspects of DHFRs……...……2
1.3 *E. coli* DHFR: general structure……...…9
1.4 Met20 loop dynamics and catalysis……..…12
1.5 The catalytic mechanism and Asp 27……..15
1.6 *E. coli* DHFR mechanism: controversy and questions…….....17
1.7 The interaction of *E. coli* DHFR and MTX: more controversy and questions……...21
1.8 Objectives of the dissertation research……...22

Chapter 2.
Methods……25
2.1 Expression, purification, and crystallization of DHFR and DHFR complexes.....25
2.2 Expression and purification of *E. coli* DHFR from a Sumo fusion construct……....32
2.3 Initial X-ray diffraction analysis…….....33
2.4 Crystallography……...36
Data Collection: Oscillation method……36
Ultrahigh resolution data collection……..38
Laue data collection……...……39
Indexing, integration, and scaling of the X-ray intensities……....41
Polarization correction……..42
Data reduction and scaling……....42
Intensities, structure factors, and the electron density function……44
Molecular Replacement (MR)……45
Difference Fourier method……50
Refinement…….…52
Refinement: generalizations for reported structures……55
2.5 Growth of larger MTX binary crystals for neutron crystallography….....58
2.6 D_2O-soaking and harvesting of crystals for neutron diffraction experiments……..59
2.7 Preliminary neutron diffraction studies: Neutron data collection at the ILL (Bennett *et al*., 2005)…….…60
2.8 Neutron data collection and processing at the Protein Crystallography Station at LANSCE……...61
2.9 Neutron structure refinement……...…63
2.10 Preparation, crystallization and X-ray data collection and processing of perdeuterated DHFR/MTX Crystals……64

Chapter 3.
Preparation of samples for neutron crystallography and protein perdeuteration……68
3.1 Background……...…68
Neutron scattering and deuterium labeling…….…..69
Protein perdeuteration…….…..72
Biophysical effects of perdeuteration……73
3.2 Results and discussion……...……...76
Applying a "hybrid" crystallization method for larger volume DHFR/MTX crystals……...76
Perdeuteration of E. coli DHFR……77

Chapter 4.
X-ray and neutron crystallographic studies of the DHFR/MTX complex……87
4.1 Background……...…87
Characterization of Asp 27 and its role in ligand binding to E. coli DHFR…..…88
The powerful tool of X-ray crystallography in macromolecular structure determination…….……90
The limits of X-rays and re-emergence of neutrons in crystallography……92
Current NC capabilities: spallation vs. reactor sources……...95
4.2 Results and discussion……..97
X-ray crystallography: Synchrotron data collection and processing……97
Data reduction and scaling……98
Molecular Replacement (MR) solution of the DHFR/MTX complex……....99
Refinement……101
Comparison of the monomers within the AU……...102
Isotropic and anisotropic B-factor description and analysis of the MTX structure……...102
Comparison to the previously reported MTX X-ray structure……111
Difference electron density analysis to identify putative hydrogen atoms……..115
Full-matrix refinement to determine the protonation state of the active site carboxylate……...119
Neutron crystallography: Preliminary neutron diffraction studies of D_2O-soaked DHFR/MTX crystals at the ILL……122
Preparation for the latest NC experiments at the ILL and LANSCE……...128
Data collection and processing at LANSCE and refinement of the neutron structure……...131
Backbone H/D exchange assessed from occupancy refinement and nuclear density maps……132
Relevance of the H/D exchange results to previous reports on DHFR "dynamics"……..141
Correlation between solvent accessibility and H/D exchange……145

H/D exchange on side chains and identification of D_2O molecules within the maps……....145
Probing the Asp 27•MTX interaction: Occupancy refinement and density analysis……148

Chapter 5.
High resolution structures of Apo and MTX/NADPH-bound DHFR..153
5.1 Background…….....…153
E. coli DHFR loop dynamics: relationship to catalysis…….……153
Crystallographic evidence of other Met20 loop conformations……..155
Protonation state of Asp 27 in the apoenzyme……156
5.2 Results and discussion…….....…158
X-ray crystallography: Data collection and processing……..158
Difference Fourier calculation for apo DHFR……....160
MR solution of the DHFR/MTX/NADPH complex……..160
Structure refinement and analysis: Apo DHFR……...162
Full-matrix refinement results: the Asp 27 in Apo DHFR……..…166
A conformation for the Met20 loop in the apoenzyme crystal structure……..168
Differences between apoenzyme structures……175
Differences between the apo and substrate-bound forms……178
Structure refinement and analysis: DHFR MTX/NADPH ternary complex……182
Analysis of the MTX and NADPH binding sites……..182
Anisotropy analysis: the MTX/NADPH ternary complex……185

Chapter 6.
Conclusions and future objectives……...188
6.1 Conclusions from the present body of research results…….…....188
Preparation for NC experiments: crystal size and deuterium incorporation…..188
Effects of perdeuteration on proteins: structure and function (specifically DHFR) ……..189
Crystallographic characterization of the DHFR/MTX complex……..190
Protonation states of Asp 27 and MTX……....192
NC studies of DHFR/MTX……...193
DHFR dynamics as revealed by NC……195
Crystallographic characterization of apo DHFR……198
A conformation for the Met20 loop when the DHFR active site is vacant……..199
Protonation state of Asp 27 in apo DHFR……...201
Role of Thr 113 and implications of the apo results for the catalytic mechanism ……..204
6.2 Future directions……208
List of References……..…...212

Vita……221

List of Tables

Table **Page**

1.1 A sequence comparison of residues 5-32 of DHFRs from different species……...8

3.1 Potential costs for perdeuteration of a protein……...…74
3.2 Synchrotron X-ray diffraction data statistics: Perdeuterated DHFR/MTX…...…85

4.1 Neutron scattering lengths (or scattered wave amplitudes) and cross-sections for atoms in biological macromolecules……94
4.2 X-ray data processing statistics: DHFR/MTX cocrystal…….100
4.3 Refinement statistics for X-ray crystallography: DHFR/MTX cocrystal……103
4.4 Unit cell contents of the DHFR/MTX cocrystal (X-ray)……..…104
4.5 Residues which possess alternate side chain conformations in the two monomers of the DHFR/MTX AU……105
4.6 Refined carboxylate bond lengths, ESDs, isotropic B-factors, and charge estimates for Asp 11, Glu 17, and Asp 27 in monomers A and B of the 1.0Å DHFR/MTX structure….…….120
4.7 Neutron diffraction data statistics: preliminary studies on a D_2O-soaked DHFR/MTX crystal at the Institut Laue-Langevin (ILL) neutron source ….….124
4.8 Data reduction statistics for the ILL preliminary study……...125
4.9 Neutron data statistics from PCS for the DHFR/MTX cocrystal……133

5.1 X-ray data collection, indexing, integration, and scaling statistics: Apo DHFR.161
5.2 Refinement statistics for the apo and MTX/NADPH DHFR structures……..163
5.3 Contents of the unit cell for the apo and MTX/NADPH DHFR crystals……164
5.4 Refined carboxylate bond lengths, ESDs, isotropic B-factors, and charge estimation for Asp 11, Glu 17, and Asp 27 in monomers A and B of the 1.05Å apo DHFR structure…….167

List of Figures

Figure **Page**

1.1 DHF is the product of the Thymidylate Synthase (TS) reaction and must be recycled back to form N^5, N^{10}-methylene-THF……...3
1.2 The pathway to thymine nucleotides and the recycling of THF……..……4
1.3 Common antifolate drugs that bind tightly to DHFRs…….....6
1.4 A comparison of DHFR structures from representative species…….....8
1.5 Cartoon representation of the tertiary structure of *E. coli* DHFR bound to MTX……...10
1.6 *In vacuo* ($\varepsilon = 0$) electrostatic surface representation of *E. coli* DHFR bound to MTX……...11
1.7 One catalytic cycle involves conformational changes of the Met20 loop…...…..13
1.8 Observed conformations of the DHFR regulatory loops (PDB ID)…….....14
1.9 A proposed catalytic mechanism for proton and hydride transfer for *E. coli* DHFR.....19
1.10 Water molecules that form hydrogen-bonding contacts with the Asp 27 and the folate pteridine ring, as observed in the ultrahigh resolution structure of folate-bound DHFR……..20

2.1 Purification of *E. coli* DHFR from SK383 cells……27
2.2 Purification of *E. coli* DHFR using a 6xHis-Sumo fusion construct…….34
2.3 An example X-ray diffraction pattern from a DHFR crystal (MTX/NADPH) measured using the in-house diffractometer……..35
2.4 The relationship between real and reciprocal space relative to X-ray data collection……37
2.5 Profile fitting and integration of intensities on an X-ray detector…….43
2.6 Euler angles ($\alpha\beta\gamma$) and orthogonal coordinate system (xyz) used in the MR rotation function…….…48

3.1 The incorporation of deuterium (D) into proteins is a prerequisite for neutron experiments……71
3.2 Expression, purification, and crystallization of perdeuterated *E. coli* DHFR…...79
3.3 Intact mass analysis of purified perdeuterated *E. coli* DHFR using ESI-FTICR mass spectrometry and measurement of deuterium enrichment……81
3.4 Intact mass analysis of purified native *E. coli* DHFR using ESI-FTICR mass spectrometry……...82
3.5 X-ray diffraction recorded from a perdeuterated DHFR/MTX crystal at NE-CAT 8-BM at the Advanced Photon Source…….…..84

4.1 The conformation of MTX in the DHFR active site is altered significantly compared to FOL due to a differing rotamer of their pterin rings…….89
4.2 Comparison of the two monomers in the asymmetric unit of the 1.0Å DHFR/MTX model……..106

Figure **Page**

4.3 High resolution electron density maps allow the modeling of alternate side chain conformations for the DHFR/MTX structure…….107
4.4 Anisotropic displacement and B-factor analysis for the 1.0Å resolution DHFR/MTX structure (monomers A and B)……...110
4.5 Comparison of monomers A and B from the 1.0Å model reported here and the previously reported 1.9Å model (3DRC; (Warren *et al.* 1991)…….113
4.6 Positive difference density maps reveal putative hydrogen positions on the Trp 22 indole ring……116
4.7 High resolution electron density maps for the MTX pteridine ring and the DHFR active site…….117
4.8 A quasi-Laue neutron diffraction image from a 0.3mm^3 D_2O-soaked DHFR/MTX crystal…….123
4.9 Wavelength normalization curve determined by LSCALE (Arzt *et al.* 1999) for the neutron diffraction experimental data collected at the ILL……..127
4.10 Larger volume DHFR/MTX crystals could be grown using a microbatch-under-paraffin oil crystallization technique and diffract to high resolution at two different neutron facilities……....130
4.11 The DHFR/MTX nuclear density maps and assessment of H/D exchange at backbone amide nitrogens……134
4.12 Comparison of backbone amide H/D exchange, isotropic B-factors, and main chain anisotropy for DHFR/MTX……137
4.13 Backbone amide H/D exchange for DHFR/MTX monomer A deduced from NC ……...……138
4.14 Backbone amide H/D exchange for DHFR/MTX monomer B ……..139
4.15 Absolute solvent accessibility (ASA) values for all residues in DHFR/MTX monomers A and B……..146
4.16 Determination of H/D exchange on the side chains of His 114 and Lys 106 (both in monomer B) from neutron density maps of DHFR/MTX and comparison to atomic resolution electron density……...147
4.17 D_2O molecules identified in the neutron maps……149
4.18 Nuclear density maps at the MTX binding site for monomers A and B of the DHFR/MTX cocrystal…….151

5.1 B-factor and anisotropy distribution for Apo DHFR……..165
5.2 The positions of the residues in the apo DHFR structure used in the carboxylate bond length measurements in the full-matrix refinement……169
5.3 In the apo DHFR structure, solvent molecules form close contacts with the Asp 27 carboxylate that are conserved in substrate and inhibitor complexes….170
5.4 The ultrahigh resolution (1.05Å) of the electron density maps allowed modeling of the core residues of the Met20 loop, observed as disordered in the previously reported apo DHFR structure (5DFR; (Bystroff *et al.* 1991)…..171
5.5 The Met20 loop in apo DHFR is in a closed conformation……173

Figure **Page**

5.6 Crystal contacts at a metal coordination site near Asn 18 help to stabilize the Met20 loop in a closed conformation in the apo DHFR structure…….……..174
5.7 Superposition of the 5DFR structure (the previously reported apo DHFR model) onto the 1.05Å resolution apo DHFR structure reported here…….176
5.8 Superposition of a closed Met20 loop folate-bound DHFR structure (1.06Å resolution) onto the 1.05Å resolution apo DHFR structure…….179
5.9 Interactions at the active site of the DHFR MTX/NADPH ternary complex…..183
5.10 B-factor and anisotropy distribution for the ternary complex of DHFR bound to MTX and NADPH.……...…...186

6.1 A comparison of hydrogen bonding patterns in the *E. coli* DHFR active site when in the A) apo (unliganded), B) folate-bound, and C) MTX-bound forms**.** …….203
6.2 Deprotonation of Asp 27 must occur upon MTX (and folate) binding, as evidenced from crystallographic results reported here……....203
6.3 Hydrogen bonding patterns and possible proton relay events in the *E. coli* DHFR catalytic mechanism……..…...206

List of Abbreviations and/or Symbols

DHFR	Dihydrofolate Reductase
TS	Thymidylate Synthase
SHMT	Serine Hydroxymethyltransferase
DHF	7, 8-Dihydrofolate
THF	5, 6, 7, 8-Tetrahydrofolate
NADPH	Nicotinamide Adenine Dinucleotide Phosphate, Reduced Form
$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate, Oxidized Form
MTX	Methotrexate (Amethopterin)
TMP	Trimethoprim
PMT	Pyrimethamine
CC	Correlation Coefficient
A	Anisotropy ratio (E_{min} / E_{max})
APS	Advanced Photon Source
NC	Neutron Crystallography
ILL	Institut Laue-Langevin
LADI	Laue Diffractometer
LANSCE	Los Alamos Neutron Scattering Center
PCS	Protein Crystallography Station
TOF	Time-of-Flight
H	Hydrogen (^1H)
D	Deuterium (^2H)
Mg^{2+}	Magnesium cation
Ca^{2+}	Calcium cation
Cl^-	Chloride anion
$MgCl_2$	Magnesium Chloride
$CaCl_2$	Calcium Chloride
KCl	Potassium Chloride
EDTA	Ethyldiaminetriacetic Acid
DTE	Dithioerythritol
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
HCl	Hydrochloric acid
PEG	Polyethylene Glycol
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
NMR	Nuclear Magnetic Resonance
MALDI	Matrix-Assisted Laser Desorption Ionization
ESI	Electrospray Ionization
FTICR	Fourier Transform Ion Cyclotron Resonance
MS	Mass Spectrometry
r.m.s.d.	root mean square deviation
ESD	estimated standard deviation
MD	Molecular Dynamics
rpm	revolutions per minute

M	Molar; unit of concentration
mg	milligram (10^{-3} gram (g)); unit of mass
ml	milliliter (10^{-3} liter (l)); unit of volume
mm	millimeter (10^{-3} meter (m)); unit of distance
s	second; unit of time
°C	degree Celsius; unit of temperature
RT	room temperature
Å	Angstrom (10^{-10} m); unit of distance
kDa	kilodalton (10^{3} Da); unit of mass
AU	asymmetric unit
α	Alpha; Type of Helix and Crystallographic Unit Cell Constant (angle)
β	Beta; Type of Strand and Crystallographic Unit Cell Constant (angle)
ε	Epsilon; Dielectric Constant and Molar Absorbitivity (Extinction Coefficient)
χ	Chi; Side Chain Torsion Angle and Reliability Value in Squared form
λ	Lambda; Wavelength of Electromagnetic Radiation
μ	Mu; Micro, as in μm, or micrometer (10^{-4} m)
ϕ	Phi; Crystallographic Rotation Axis
σ	Sigma; Error and Density Map Contour Level Measurement
F_{obs}	Observed Structure Factors
F_{calc}	Calculated Structure Factors
R_{fac}	Crystallographic Reliability Factor Based on Comparison of Observed Structure Factors to Calculated Structure Factors (R_{cryst} or R_{work})
R_{free}	Crystallographic Reliability Factor Calculated as Above, Only for 5-10% of Data Sequestered from the Refinement (R_{test})
R_{merge}/R_{symm}	Error Calculated for the Merging of Symmetry-Related Reflections

Chapter 1. Introduction

1.1 Pyrimidine nucleotide & amino acid biosynthesis and tetrahydrofolate cofactors

A commonality shared among all known organisms is the maintenance of genetic material in the form of ribonucleic and/or deoxyribonucleic acid (RNA and DNA, respectively) and the requirement of its correct replication and transmission from parent to progeny. Additionally, proteins, which perform the bulk of the tasks within living cells, are universally composed of different linear sequences and percentages of the 20 common amino acids. Although the biosynthetic pathways which are used to create them are rigorously regulated at multiple levels (i.e. transcriptionally, post-translationally and others), the supply of the monomeric units which comprise the larger polymeric forms of RNA, DNA, and proteins must be kept in ample abundance. For the generation of one of the pyrimidine nucleotides in DNA, thymine, and the amino acid, methionine, tetrahydrofolate (THF) is a necessary cofactor for the enzymes catalyzing the direct chemical reactions. THF is widely utilized in these and other pathways because of its capability to transfer one-carbon (C1) units, even when it is in different oxidation states (Voet *et al.* 1995). A major entry point of THF into biosynthetic reactions is as the product of the serine hydroxymethyltransferase (SHMT) reaction, N^5, N^{10}-methylene-THF, where the methylene group branched between the N5 and N10 atoms can be used as the C1-unit for transfer. N^5, N^{10}-methylene-THF serves as cofactor for the thymidylate synthase (TS) reaction, converting deoxyuridylmonophosphate (dUMP) into deoxythymidylmonophosphate (dTMP) via methyl transfer to the dUMP C5 atom from the THF metabolite. Six individual enzymatic reactions are required in the *de novo* pyrimidine biosynthetic pathway in bacteria to generate UMP from ATP, glutamine, and

CO_2. From UMP, CTP can also be produced in addition to the dTMP from the TS reaction. The dTMP products from the TS reaction can then be phosphorylated by a nucleotide monophosphate kinase to dTTP which then can be incorporated into DNA at S phase or during a DNA damage response (Alberts *et al.* 1994). A by-product of the TS reaction is dihydrofolate (DHF) (**Figure 1.1A**). For most organisms, in order to replenish the stores of THF (**Figure 1.1B**) to serve as cofactor for future TS and other important reactions, DHF must be converted back into N^5, N^{10}-methylene-THF (**Figure 1.1C**) by sequential activities involving reduction by dihydrofolate reductase (DHFR) and methylation by SHMT (**Figure 1.2**).

1.2 General aspects of DHFRs

DHFRs are generally conserved across species from *Archaea* to the higher mammals, although recently it has been discovered that a large number of microbial species utilize an alternative flavin cofactor-dependent TS pathway to generate dTMP nucleotides (Myllykallio *et al.* 2002) (Myllykallio *et al.* 2003), thus, precluding the necessity for DHFR to help recycle THF. However, for most organisms, DHFRs are critical for multiple metabolic pathways (Kraut *et al.* 1987). They catalyze the NADPH-dependent reduction of 7,8-DHF to 5,6,7,8-THF. It is a two-step process where reduction occurs across the double bond between the N5 and C6 atoms of the DHF pteridine ring; most evidence indicates that proton transfer occurs first, possibly from solvent or from a titratable residue within or near the enzyme's active site, with concomitant or subsequent transfer of a hydride (H^-) from the C4 atom of the nicotinamide ring of the NADPH cofactor (a detailed mechanism and figure for *E. coli* DHFR is provided later in the

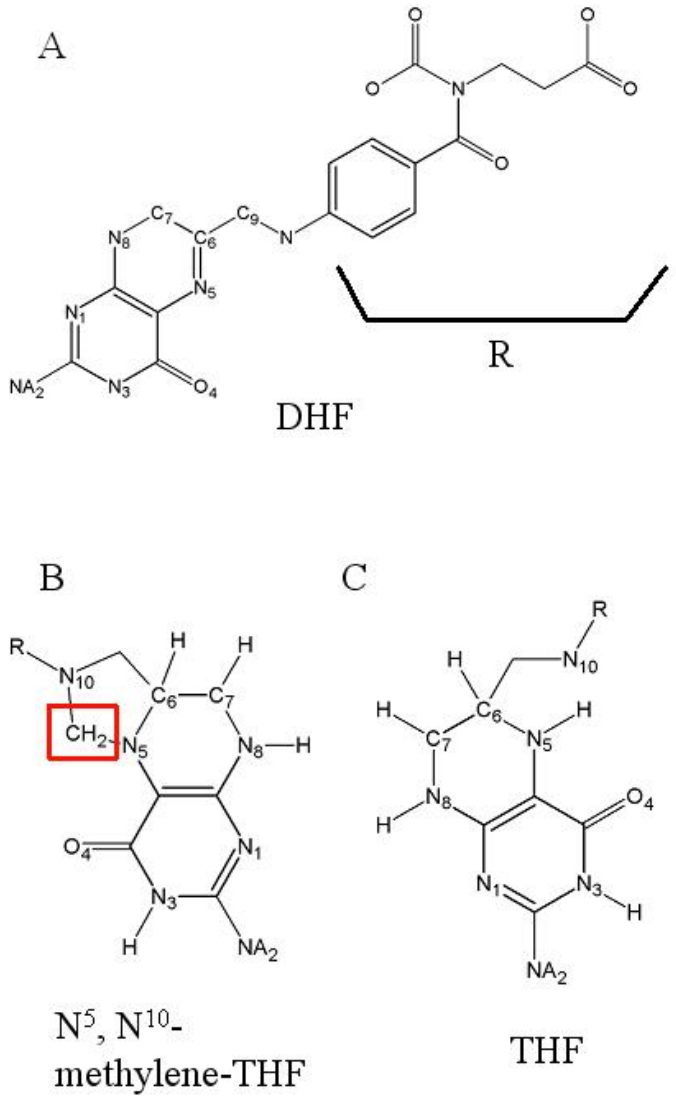


Figure 1.1: DHF is the product of the Thymidylate Synthase (TS) reaction and must be recycled back to form N^5, N^{10}-methylene-THF. A) The chemical structure of the TS-reaction product dihydrofolate (DHF) and **B)** the C1-unit (highlighted in the red box) transferred by the cofactor, N^5, N^{10}-methylene-tetrahydrofolate (THF), required for the TS reaction to make dTMP nucleotides. C) THF, the product of the DHFR reaction, is an intermediary metabolite in the recycling process to restore pools of N^5, N^{10}-methylene-THF.

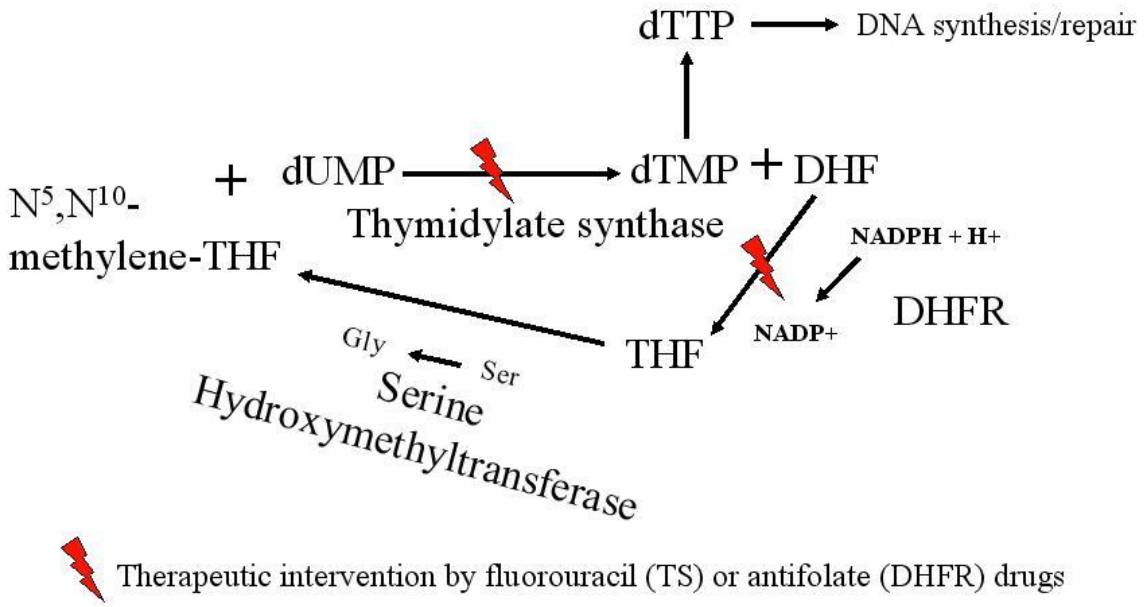


Figure 1.2: The pathway to thymine nucleotides and the recycling of THF. The synthesis of dTMP nucleotides can only occur in most organisms via a THF-dependent TS reaction. Restoration of the N^5, N^{10}-methylene-THF supply for continued production of dTMP occurs via two enzymatic reactions involving DHFR and SHMT. Direct or indirect (at the DHFR step) inhibition of TS (indicated by the red lightning bolts) by several well-known drugs such as fluorouracils (converted to F-dUMP in the cell) or antifolates such as amethopterin (or methotrexate (MTX)) and trimethoprim (TMP) have been extremely successful contributors in anti-cancer, anti-microbial, and anti-protozoan therapeutic regimes.

chapter). Mammals are unable to synthesize folate and its derivatives so they must rely on diet and symbiotic microorganisms living within their intestinal tract for its supply (Voet *et al.* 1995). DHFR can catalyze the reduction of folate into DHF, albeit with reduced efficacy as compared to reducing DHF directly to THF (Posner *et al.* 1996). For most microbial species and all multicellular species known, the proficient ability to block DHFR's catalytic activity leads to interruption of DNA synthesis, genomic instability, and, ultimately, cell death (Schnell *et al.* 2004). This is mainly due to the indirect inhibition of TS by depleting cellular pools of N^5, N^{10}-methylene-THF. Accordingly, several therapeutic agents have been discovered that can inhibit TS (the fluorouracil compounds) or DHFR, most notably, the anti-tumor drugs aminopterin and amethopterin (or methotrexate (MTX)), the anti-microbial trimethoprim (TMP), (**Figure 1.3A, B, C**) (Huennekens 1994; Huennekens 1996; Schnell *et al.* 2004). More recently, a vast number of analogs, mainly of TMP, have been synthesized and tested as successful lead compounds against a variety of DHFRs, including the anti-fungals pyrrolo- and pyrido-pyrimidines, and the anti-protozoan compounds, pyrimethamine (PYR) and WR9910 (Kuyper *et al.* 1996a; Kuyper *et al.* 1996b) (Hekmat-Nejad *et al.* 1997) (Li *et al.* 2000) (Zuccotto *et al.* 1998).

Due to its proficient binding to human DHFR, MTX is quite common today as part of a chemotherapeutic cocktail, especially for rapidly-growing tumors associated with acute leukemia, lymphomas, and choriocarcinomas (Voet *et al.* 1995), and is also used in anti-inflammatory treatments such as to combat rheumatoid arthritis, presumably DHFR being the primary target of its use in these patients. However, severe side effects

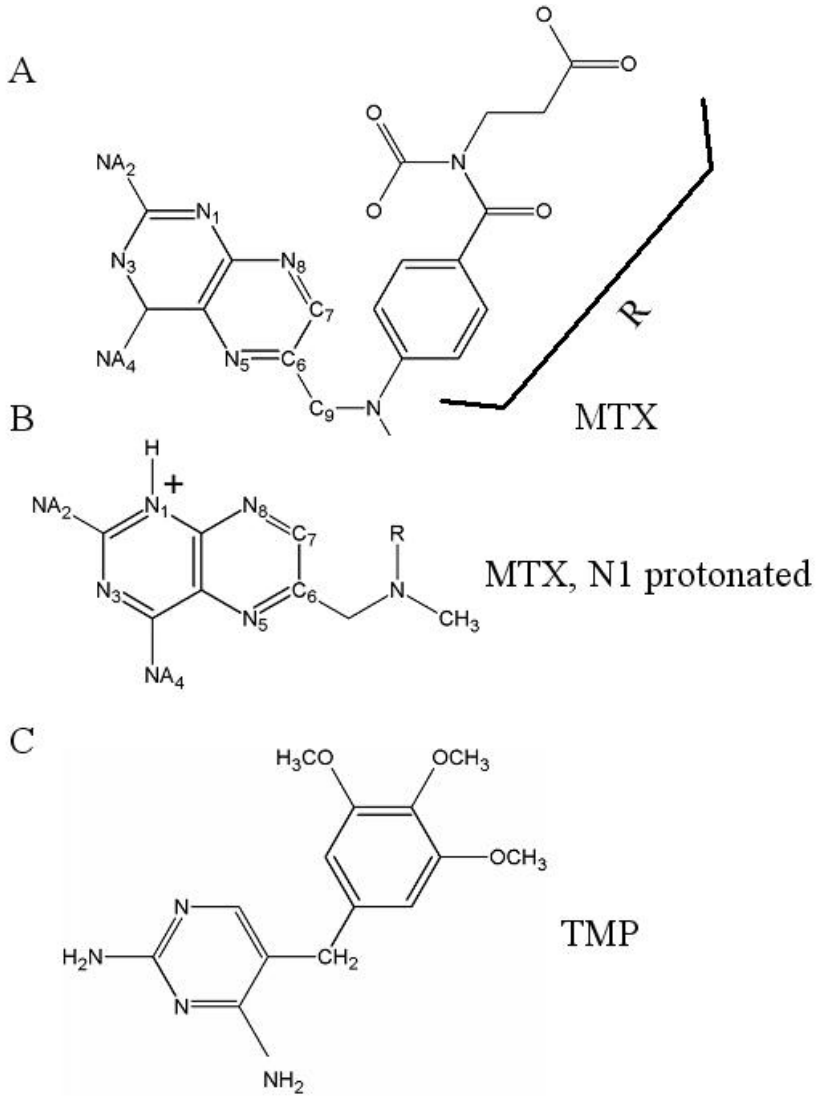


Figure 1.3: Common antifolate drugs that bind tightly to DHFRs. A) The anti-cancer and anti-inflammatory drug MTX, **B)** a positive charge is imparted on the N1 atom of MTX when it is protonated, and **C)** the anti-microbial and anti-protozoan drug TMP.

from its use persist; one primary example includes MTX, which inhibits DHFR in tumor and healthy cells. Interestingly, DHFR was one of the first enzymes to exhibit increasing expression levels as a means of resistance in tissues where MTX had been administered (Huennekens, 1996). Most of the above compounds are broad-spectrum inhibitors, binding tightly to several forms of the DHFR enzyme; efforts to narrow this spectrum are ongoing, with the aforementioned TMP antibiotic having a ~10^5 lower K_d for most microbial DHFRS compared to the human enzyme (Schnell *et al.* 2004). In fact, TMP binds *E. coli* chromosomal DHFR with a K_d of 20 pM (Howell 2005). Recently, from several pathogenic species, additional DHFR genes have been cloned (such as that from *Bacillus anthracis* (Barrow *et al.* 2004)) and a few of their three-dimensional structures have been solved by X-ray crystallography (for example, those from *Leishmania major* (Knighton *et al.* 1994) and *Mycobacterium tuberculosis* (Li *et al.* 2000)). This has led to suggestions on the design of novel and more *selective* inhibitors of DHFR and the folate synthesis pathways in these species. The sequence and, perhaps more importantly, the structural similarities of DHFRs from different species may make this effort quite difficult (**Figure 1.4**). However, slight differences in active site residues and their position may make this a feasible goal (**Table 1.1**), and hope is also provided from the success and efficacy from the use of TMP as an antimicrobial drug. For the remainder of this thesis and for the dissertation research described here, the focus is on high resolution structural characterization of the chromosomal DHFR enzyme from *Escherichia coli* (EC 1.5.1.3; 159 residues; 18 kilodaltons (kDa)).

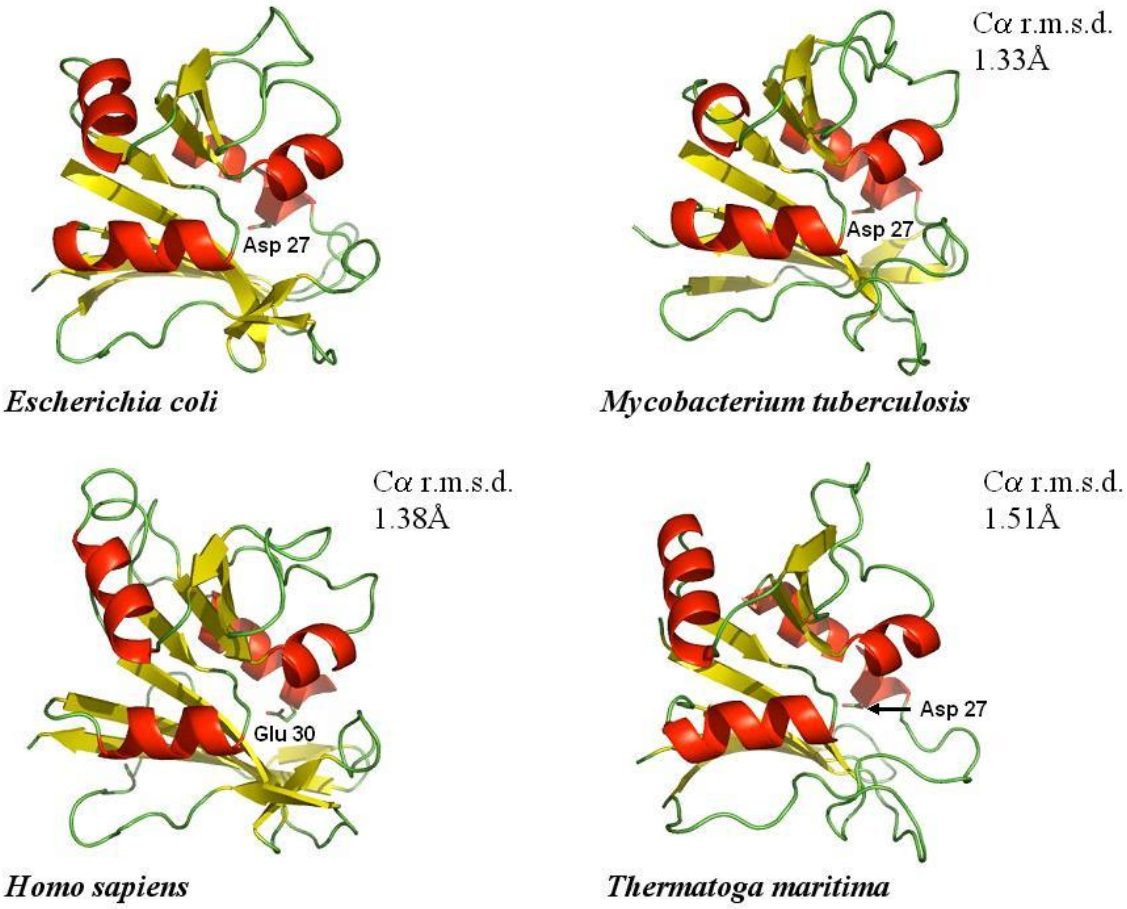


Figure 1.4: A comparison of DHFR structures from representative species. Cα atoms were used for least-squares superposition in the program O with the *E. coli* structure as the template.

Table 1.1: A sequence comparison of residues 5-32 of DHFRs from different species.

E. coli DHFR 17, 999 Da, 159 residues	AL**AVDRVIGMENAMPWNL**PA**D**LAWFKR
M. tuberculosis DHFR 18, 000 Da, 159 residues	AQ**ATSGVIGRGGDIPWRL**PE**D**QAHFRE
B. anthracis DHFR 19, 100 Da, 162 residues 39.4% identity to *E. coli* DHFR	VA**MDENRVIGKDNNLPWR**LPS**E**LQYVK
H. sapiens DHFR 21, 000 Da, 182 residues 36% identity to *E. coli* DHFR	CIVAVSQNMGIGKNGDLPWPPLRN**E**FR

Bold: Procaryotic active site regulatory loop sequences (e.g. Met20 for *E. coli*)
Red: Active site acidic residue **1.3 *E. coli* DHFR: general structure**

1.3 ***E. coli* DHFR: general structure**

There are more than 45 X-ray structures available for *E. coli* DHFR in 28 different liganded states as deposited in the Protein Data Bank (www.rcsb.org, (Berman *et al.* 2000). A recent listing of many of these structures with references, PDB codes, and basic crystallographic parameters (i.e. unit cell constants, resolution limits, and R-factors) is found in Table 1 of (Sawaya *et al.* 1997). DHFR's overall three-dimensional structure is dominated by an α/β pseudo-Rossman fold (Creighton 1993), composed of a central eight-stranded ß-sheet structure, with the upper four sheets (termed βB-βE) maintaining the adenosine nucleotide binding (minor) subdomain (residues 38-88) while the lower ones (termed βA and βF-βH) maintain the substrate binding (major) subdomain (Bolin *et al.* 1982; Bystroff *et al.* 1990; Matthews *et al.* 1977; Reyes *et al.* 1995) (**Figure 1.5**). The DHF substrate and the NADPH cofactor bind at a cleft formed at the interface of the two subdomains. The major subdomain (residues 1-37 and 89-159) is also known as the loop domain since nearly 50% of the residues in this region are within the Met20 (residues 9-24), βF-βG (residues 116-132), and the βG-βH (residues 142-150) loops. These loops can bend and contort on the ligand-binding face of the protein, surrounding the active site. In fact, the Met20 loop closes over the active site during catalysis and seems to enhance the turnover rate by occluding into the cofactor-binding site once hydride transfer has occurred (Falzone *et al.* 1994); Sawaya *et al.*, 1997). At physiological pH, *E. coli* DHFR has a net charge of -11, however, regions important for contacting the Glu tail of the substrates or of MTX are rich in basic residues, imparting a positive surface potential (**Figure 1.6**).

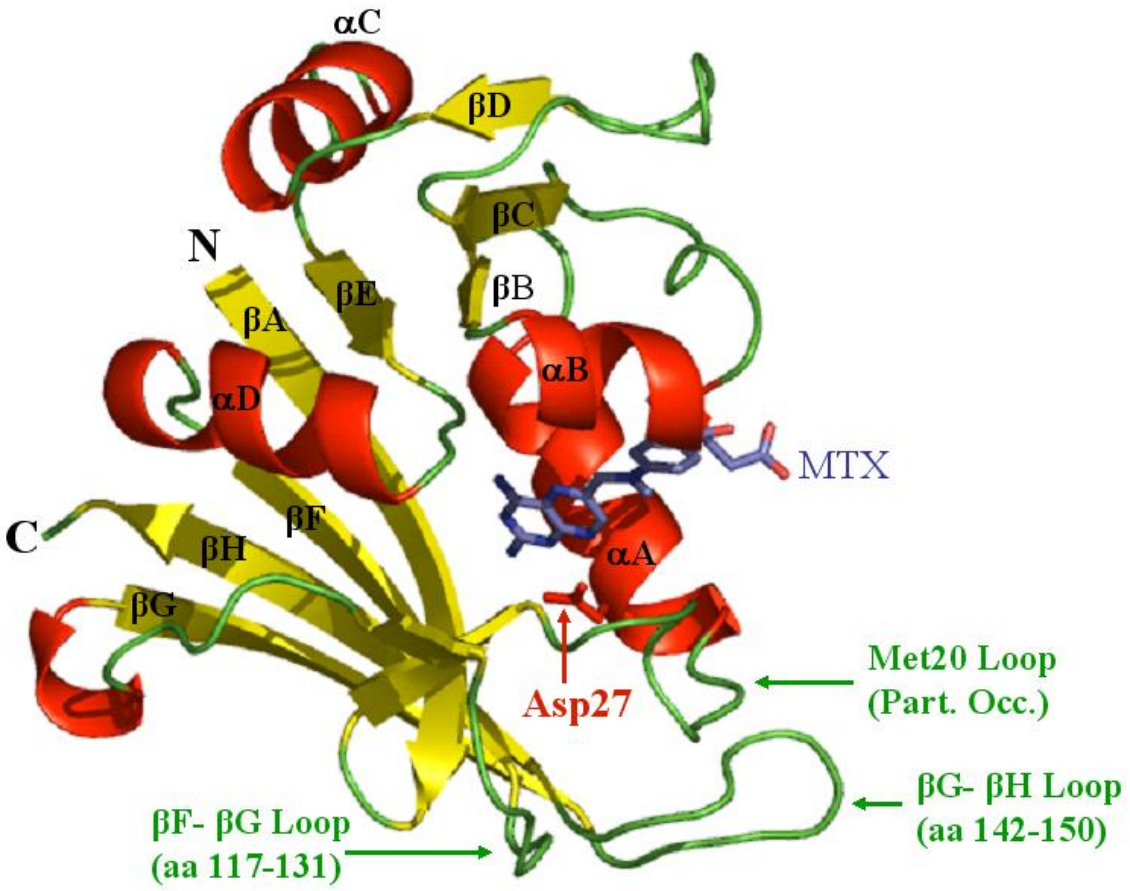


Figure 1.5: Cartoon representation of the tertiary structure of *E. coli* DHFR bound to MTX. Secondary structural elements (α-helices in red and β-strands in yellow), major loop regions (green), the MTX inhibitor, the N- and C-termini, and the active site Asp 27 (red) are labeled. The adenosine-binding loop (residues 62-70) is at the top of the structure, linking the βC and βD strands together.

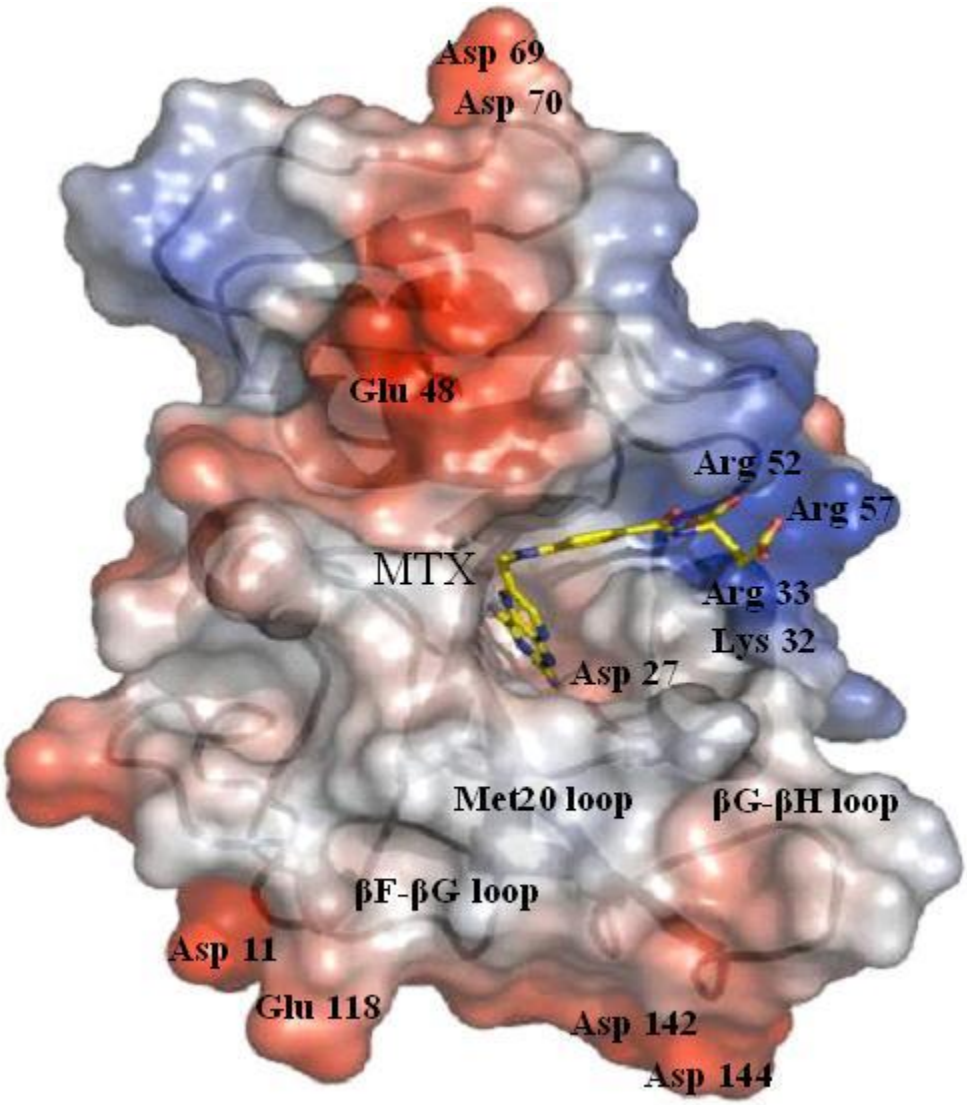


Figure 1.6: *In vacuo* (ε = 0) electrostatic surface representation of *E. coli* DHFR bound to MTX. Red represents regions of the structure with a net negative electrostatic potential whereas blue represents regions of the structure with a net positive electrostatic potential. Charge complementarity exists between the Glu tail of the MTX and a cluster of basic residues. The active site is predominantly without charge save a small patch at the Asp 27•MTX N1 contact. The surface is shown at approximately 20% transparency to reveal the cartoon representation of the structure, similar to what is shown in **Figure 1.5**. This orientation is rotated ~90° about an axis that is parallel to the plane of the paper as compared to the orientation in the previous figure.

1.4 Met20 loop dynamics and catalysis

From site-directed mutagenesis of the Met20 and βF-βG loop residues and from NMR and crystallographic studies, it is thought that the positioning of the Met20 loop is proposed to be indelibly linked to the positioning of the enzyme along the reaction coordinate (Miller *et al.* 1998; Miller *et al.* 2001; Sawaya *et al.* 1997); (Venkitakrishnan *et al.* 2004). The Met20 loop conformation in the context of the reaction scheme for *E. coli* DHFR is shown in **Figure 1.7** and different crystallographic conformations of the loop are shown in **Figure 1.8**. As determined in several different kinetic studies, the reaction catalyzed by *E. coli* DHFR is ordered, with NADPH binding first (K_m = 0.94μM) and aiding in release of the THF product (the rate-limiting step) from the previous reaction cycle (Stone *et al.* 1982); (Stone *et al.* 1988); (Fierke *et al.* 1987). The Met20 loop closes over the active site with the binding of NADPH and remains until the chemical step has occurred. After binding the DHF substrate (K_m = 1.2μM), the enzyme rapidly completes proton and hydride transfer (>200 s^{-1}) and the conformation of the Met20 loop dramatically changes from closed over the active site to occluding into the nicotinamide binding site. This is caused by a reorganization of the N-terminal and core part of the Met20 loop, the largest changes being a 180° rotation about ψ for Ile 14 and repositioning of Glu 17 and Asn 18, disrupting hydrogen bonding with residues within the βF-βG loop (backbone atoms in Gly 121 and Asp 122) that help to stabilize the closed conformation (Sawaya *et al.* 1997). Upon changing conformation to occluded, the N-terminal and core residues of the Met20 loop protrude into the cofactor pocket while the C-terminal Met20 loop residues (specifically Asn 23) form hydrogen bonds with (Schnell *et al.* 2004); (McElheny *et al.* 2005). A structural overlay of the closed, occluded, and

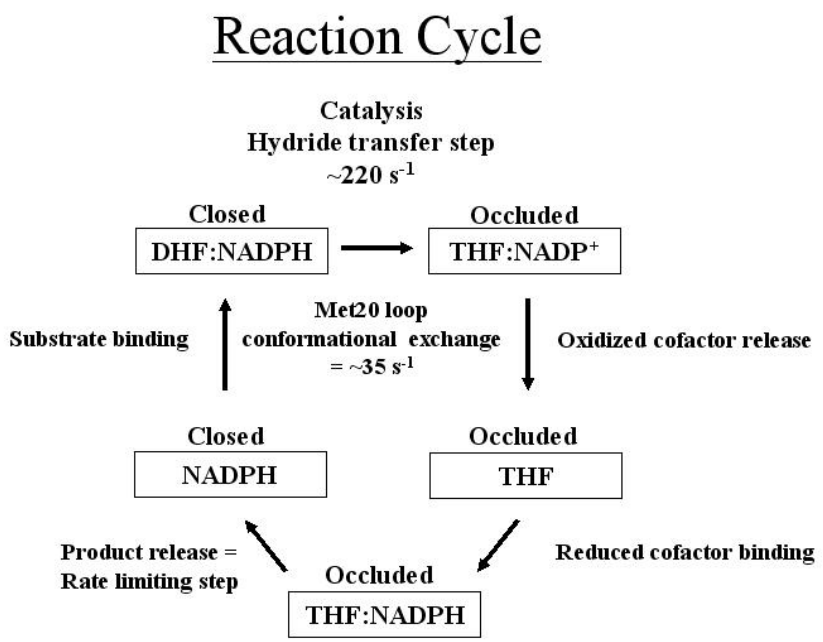


Figure 1.7: One catalytic cycle involves conformational changes of the Met20 loop.

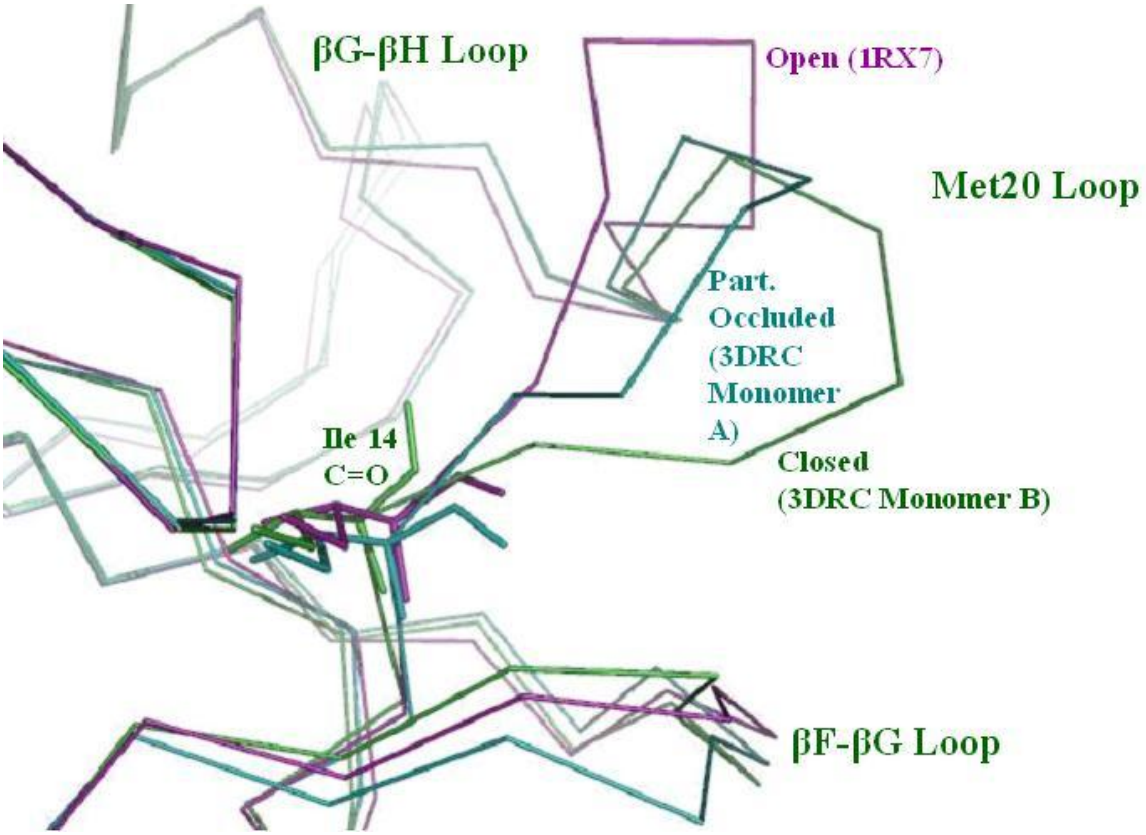


Figure 1.8: Observed conformations of the DHFR regulatory loops (PDB ID in parentheses).

open Met20 loop conformations are shown in **Figure 1.8**. The open conformation has not been identified in solution experiments such as NMR and stopped-flow fluorescence and backbone and side chain atoms of Ser 148 in the βG-βH loop. While the βF-βG loop seems to be the anchor for the Met20 loop when it is closed, it is the βG-βH loop that serves a similar purpose when the Met20 loop adopts an occluded conformation appears to be unique to certain crystallographic environments, where it is stabilized by lattice contacts. The open conformation extends further out from the active site than both the closed and occluded forms (Sawaya *et al.* 1997). The terminal residues of the open Met20 loop conformation also makes hydrogen-bonding contacts with the other regulatory loops but does so to both the βF-βG and the βG-βH loops while its central portion extends out and away from the active site. It may be that a population of open conformers is important in catalysis to allow entrance and exit of substrate and cofactor (Schnell *et al.* 2004). It is interesting to note that in the crystal structures of the apo and holoenzyme (where only NADPH is bound) electron density for the core of the Met20 loop (residues 16-21) is not observable, indicating disorder that may be due to time-averaged fluctuation between the closed and occluded conformers (Bystroff *et al.* 1991; Bystroff *et al.* 1990).

1.5 The catalytic mechanism and Asp 27

The active site of DHFR is quite hydrophobic with the only ionizable residue being aspartate 27 (Asp27), within α–helix B (αB), framing one side of the active site (Bolin *et al.* 1982; Matthews *et al.* 1977; Matthews *et al.* 1985), adjacent to the pteridine ring binding site. The protonation state of the Asp27 is the subject of great controversy. Early crystallographic studies of DHFR bound to DHF analogs (such as MTX) revealed

that the active site Asp27 would actually be >5Å away from the N5 atom on the pteridine ring if substrate was bound (Bolin *et al.* 1982; Matthews *et al.* 1977; Reyes *et al.* 1995). Even when the structures of DHFR bound to the weak substrate folate, to a dideoxy form of the product THF (ddTHF), or to other analogs such as folinic acid (also known as leucovorin) were elucidated, surprisingly, the Asp 27-to-N5 atom distance was about the same as observed in the earlier MTX structures, with the notable exception that MTX binds DHFR with a 180° rotation using the MTX C6-C9 bond as the rotation axis (Reyes *et al.*, 1995; (Sawaya *et al.* 1997); (Lee *et al.* 1996). Importantly, water molecules were observed within ~3.0Å of the Trp 22 and the Asp 27 residues, and the C4-O4, and the N5 atoms of the folate pteridine ring in the DHFR/folate structure, suggesting that solvent may participate in ligand binding and proton relay (Reyes *et al.* 1995). Crystal structures of ternary complexes, such as folate and $NADP^+$ bound to DHFR (Bystroff *et al.* 1990), also depict the rather long distance between the Asp 27 and the N5 proton acceptor of the substrate pteridine ring. However, mutagenesis of Asp 27 to Ser or Asn shows a severe decrease (several hundred-fold) in catalytic rate and efficiency (Howell *et al.* 1986; Villafranca *et al.* 1983).

The absolute role Asp 27 has in catalysis is therefore of great interest, and a number of groups have proposed different possibilities for its importance. While the initial mutagenesis studies suggested Asp 27 serves as a general acid, later studies propose that Asp 27 acts by elevating the N5 pK_a of bound DHF from 2.59 to 6.5, using long-range polarization effects and electrostatic forces to induce and maintain protonation of DHF once bound (Bajorath *et al.* 1991; Chen *et al.* 1994; Chen *et al.* 1997). Difference UV spectra suggested that an ionizable group from the enzyme has a pKa of ~6.3 but did

not speculate that this may be Asp 27; furthermore, they indicated that the ionization state of this group had minimal effects on DHF binding to the enzyme (Stone *et al.* 1983b). Since a pH profile of DHFR reveals a catalytic pK_a for the hydride transfer step of 6.5 (Fierke *et al.* 1987), this titration could possibly correlate with a severely perturbed pK_a for Asp 27. Recent ^{13}C NMR studies have shown that the homologous Asp 26 residue in *Lactobacillus casei* DHFR possesses a pK_a less than 4, so this residue appears to be negatively charged at physiological pH (Casarotto *et al.* 1999). An alternative possibility is that the pK_a could be ascribed to an ionizable group on the substrate, DHF. Resonance Raman spectra have established the pK_a of the N5 on DHF as 6.5 when bound to DHFR in a ternary complex with $NADP^+$ (Chen *et al.* 1994) whereas more recent Raman studies over a pH range of 5.6-9.0 revealed that the Asp27 most likely has a pK_a below 5 and, thus, is charged at least in the ground state (Chen *et al.* 1997). This suggests that the kinetic pK_a observed relates to the protonation state of bound ligand and not the Asp 27 residue.

1.6 *E. coli* DHFR mechanism: controversy and questions

Using computational approaches, different groups have proposed different pathways of proton donation and different ionization states for Asp 27 and bound ligands (Bajorath *et al.* 1991c); (Cannon *et al.* 1997a; Cannon *et al.* 1997b); (Greatbanks *et al.* 1999); (Cummins *et al.* 2001). A reorganization of electron distribution in folate and DHF has been suggested to occur once it has bound DHFR to counter the negative charge of Asp 27 (Bajorath *et al.* 1991a); more recent studies using varying theoretical quantum mechanical approaches have reiterated the substantial substrate polarization that occurs upon binding to DHFR, a shifting of ~0.5 electron from the N3-C4 bond towards the N5-

C6 bond of the DHF pteridine ring (Greatbanks *et al.* 1999). Recently, evidence has mounted for the catalytic pK_a possibly being ascribed to keto-enol tautomerization occurring at the N3-C4-O4 region of the substrate (Chen *et al.* 1994) (Lee *et al.* 1996) (Cannon *et al.* 1997b). It could be that this is a first step in a proton relay pathway involving the tautomeric group, water molecules, and ultimately the N5 atom on the substrate. Rajagopalan and Benkovic (2002) have suggested tautomerization could be induced by expulsion of bulk solvent upon substrate binding, lowering the dielectric constant at the active site, and triggering protonation of Asp 27 from the N3 atom of DHF. This would cause electron rearrangement within this region of the pteridine ring, facilitating protonation of the O4 atom from solvent and thus resulting in the enol tautomer. This sets the stage for proton donation from solvent to N5, possibly from a hydronium ion (H_3O^+) created from a water molecule that shares a proton with the O4 atom involved in tautomerization (Rajagopalan *et al.* 2002). A catalytic mechanism for *E. coli* DHFR is shown in **Figure 1.9**, with possible roles for Asp 27 and solvent molecules observed in recent substrate-bound crystal structures (noted above for the DHFR/folate structure; (Reyes *et al.* 1995)) that could help drive tautomerization at O4 and, ultimately, proton transfer to N5. **Figure 1.10** shows two water molecules participating in hydrogen-bonding contacts with the Asp 27 and folate in an ultrahigh resolution X-ray structure of DHFR/folate solved by a colleague in the Dealwis laboratory (Dr. Anna Gardberg). Cummins and Gready (2001) have used results from *ab initio* quantum chemical calculations on DHFR bound to DHF and folate to suggest that structurally conserved solvent bridges the Asp27 Oδ2 and the substrate O4 atom, the Asp 27 is protonated first or initially in catalysis, and that direct protonation of N5 (or N8 if folate is bound) occurs,

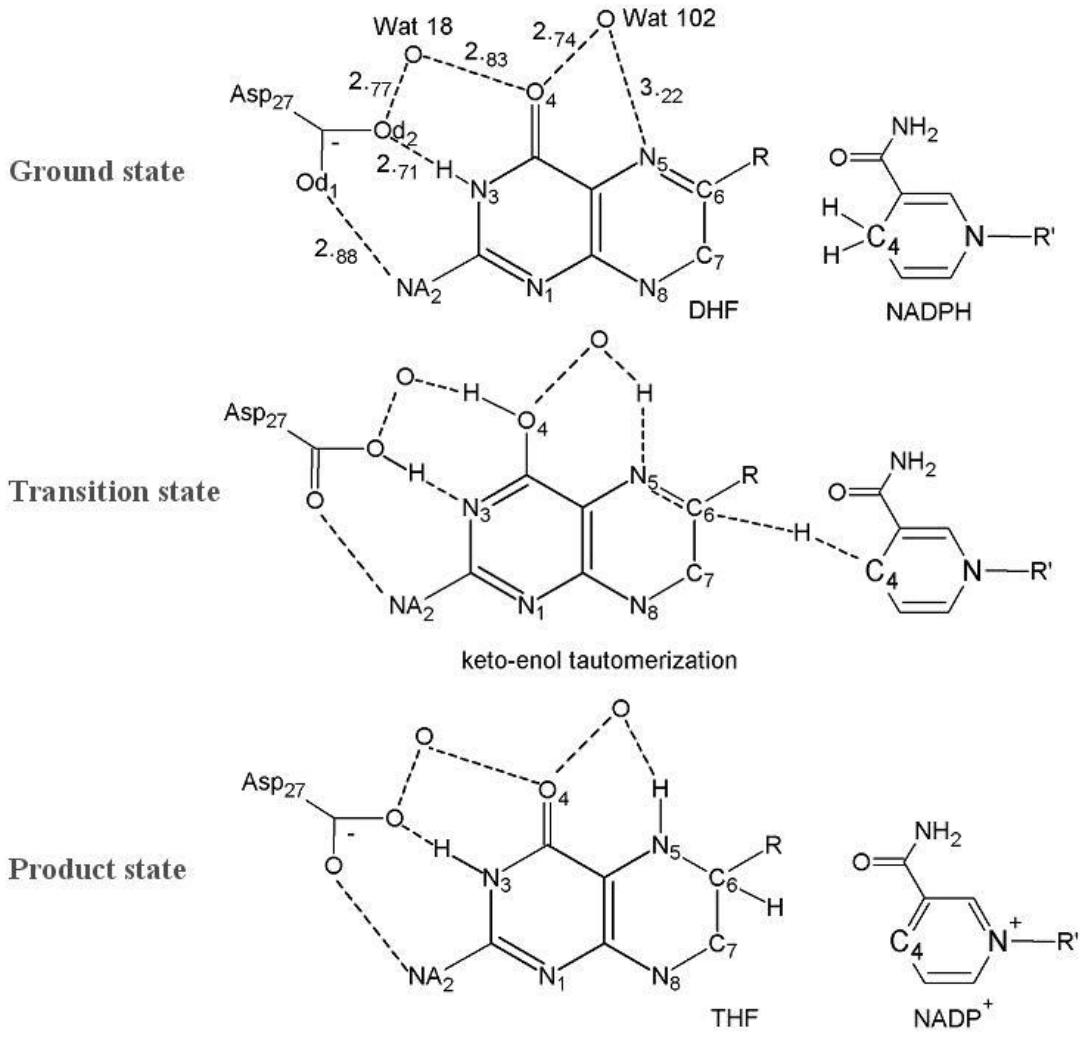


Figure 1.9: A proposed catalytic mechanism for proton and hydride transfer for *E. coli* DHFR. DHFR catalyzes the reduction across the N5-C6 double bond of DHF, using NADPH as a cofactor. Water molecules and distances to the substrate shown in the ground state have been observed in a 1.06Å resolution folate-bound DHFR structure (see **Figure 1.10**; coordinates kindly provided by Dr. Anna Gardberg in the Dealwis laboratory). It has been proposed that water molecules play a significant role in proton transfer, possibly first to O4, triggering a tautomerization event, and then to the N5 atom (Bystroff *et al.* 1990) (Chen *et al.* 1994) (Reyes *et al.* 1995) (Lee *et al.* 1996) (Rajagopalan *et al.* 2002). Most studies suggest that hydride transfer from NADPH to the C6 atom of DHF occurs concomitant with or immediately after N5 reduction.

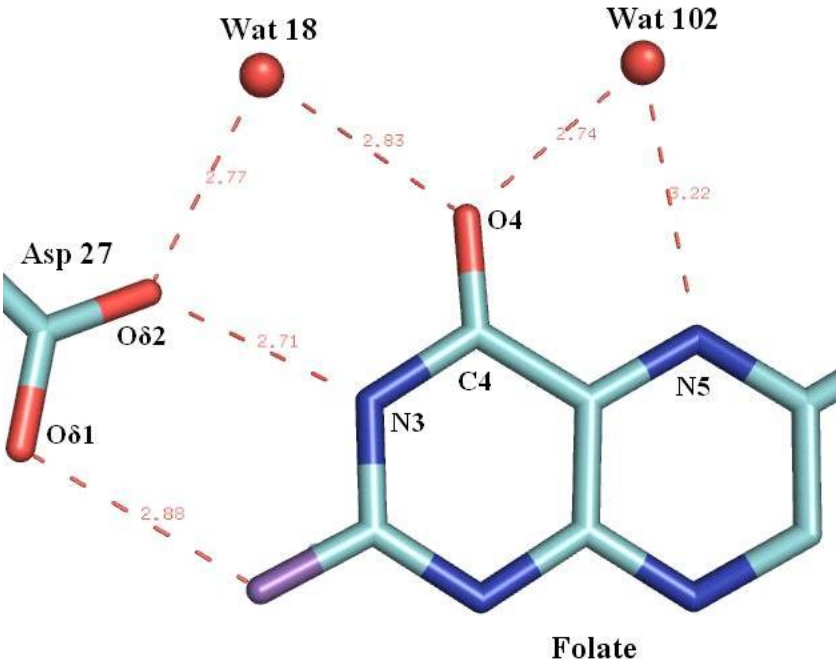


Figure 1.10: Water molecules that form hydrogen-bonding contacts with the Asp 27 and the folate pteridine ring, as observed in the ultrahigh resolution structure of folate-bound DHFR. Contacts are shown in red dashed lines and contact distances between atoms are shown in red. Solvent has been implicated to have mechanistic importance for proton transfer to DHF (see **Figure 1.9**). Even though direct reduction of N5 of folate is unlikely, it has been shown that DHF and folate bind in the same orientation in the DHFR active site (Reyes *et al.* 1995) (Sawaya *et al.* 1997), thus, it is expected that one or both of the water molecules shown would form similar contacts with DHF. (The coordinates for the DHFR/folate structure (1.06Å resolution; $P6_1$ space group) were kindly provided by Dr. Anna Gardberg.)

precluding a tautomerization step (Cummins *et al.* 2001) (Rajagopalan *et al.* 2002). The protonation state of the Asp 27 throughout the reaction, the origin and path of proton donation, and the extent to which solvent contributes to catalysis all remain controversial issues.

1.7 The interaction of *E. coli* DHFR and MTX: more controversy and questions

As stated earlier, MTX is a substrate analog that tightly binds *E. coli* chromosomal DHFR (pM K_d) and persists as an important chemotherapeutic agent. In contrast to the protonation states of substrate during the catalytic mechanism, there exists evidence from X-ray crystallography, NMR and Raman and UV difference spectra to suggest that the N1 atom of MTX is protonated and, thus, positively charged when bound to DHFR (Matthews *et al.* 1977);(Bolin *et al.* 1982); (Appleman *et al.* 1988; Coco *et al.* 1981; Coco *et al.* 1983); (Poe *et al.* 1972); (Ozaki *et al.* 1981); (Stone *et al.* 1983b). The binding of MTX in the DHFR active site is such that the N1 atom is within 2.7Å of the Oδ2 of Asp 27 and the NA2 amide nitrogen is within 2.9Å of the Oδ1 of Asp 27; as stated above, this stems from an orientation of the pteridine ring which is flipped as compared to folate and DHF binding (Appleman *et al.* 1988). The close proximity of N1 of MTX when bound to DHFR to the Asp 27 carboxylate group strongly suggests an electrostatic interaction, possibly ionic in nature. This interaction is seemingly so favorable that the N1 pK_a increases from 5.7 for free MTX to >10 for DHFR-bound MTX (Cocco *et al.* 1983) and is hypothesized to be the reason that the dissociation constants for folate and DHF for DHFR are so much higher than for MTX (Appleman *et al.* 1988; Stone *et al.* 1988). A recent computational study was performed measuring the affinities of protons on the Asp 27 carboxylate and the MTX N1 in low dielectric fields (ε

= 4 - 20) (Cannon *et al.* 1997a), attempting to mimic what has been evidenced to be the chemical environment at the core of globular proteins, especially at an enzyme active site (Creighton 1993). This has been suggested to be especially critical to consider when calculating the pK_a of an ionizable group which has been shown to be important in ligand binding and catalysis (Antosiewicz *et al.* 1996) (Demchuk *et al.* 1996). The results propose that the Asp 27 is protonated while MTX is not below an ε value of 10; this would mean that both groups are neutral when MTX is bound to ecDHFR, and, therefore, the interaction is suggested to be dipole-dipole rather than ionic (Cannon *et al.* 1997), refuting most of the available research on the subject.

1.8 Objectives of the dissertation research

Direct determination of hydrogen positions within the active site of DHFR has not been demonstrated and the lack of a consensus on DHFR's mechanism of proton donation arises from the question: what are the protonation states of Asp27 and bound ligands in the DHFR active site? The complementary use of ultrahigh resolution X-ray (beyond 1.2Å resolution) and high resolution neutron (beyond 2.5Å resolution) crystallography has been employed in the present dissertation research to probe this exact question, among others. X-ray crystallography is a well-established technique for the determination of macromolecular structure, allowing not only gross features of a molecule to be observed (such as secondary structure) but also, for resolutions beyond 3Å, distinction between different amino acid side chains of proteins. However, hydrogen atoms scatter X-rays weakly since scattering power of an atom is proportional to the number of electrons it possesses, of which hydrogen only has 1 (Z = 1). Even with atomic resolution data (beyond 1.2Å resolution), where distinct electron density peaks for

individual atoms can be resolved from one another, nearly all hydrogens in a protein structure cannot be observed. Those that can are often bound to a "heavier" atom which has a very low B-factor (Howard *et al.* 2004), and one must rely on single difference electron density maps ($\mathbf{F}_o$-$\mathbf{F}_c$) contoured at levels much higher than noise (perhaps, >+2.5σ) to be confident that a particular peak can be attributed to a hydrogen atom (Minichino *et al.* 2003) (Schmidt *et al.* 2003). Of course, even then, proper bonding geometry for the hydrogen to the atom to which it is bound must be maintained. In regions of a protein where there is high static (inherent) or dynamic (thermal fluctuations) disorder, it will most likely be impossible to observe hydrogen atoms, much less the heavier atoms to which they are bound (Drenth 1999). Possibly, with sub-atomic resolution crystallography at liquid helium temperatures (about -250 to -260°C), these features may be distinguished but the studies which have attained this level of detail from electron density maps is limited (Longhi *et al.* 1998) (Jelsch *et al.* 2000) (Ruiz *et al.* 2004) (Schmidt *et al.* 2003; Schmidt *et al.* 2005).

Neutron crystallography at moderate resolutions (beyond 2.5Å), on the other hand, can reveal hydrogen positions *directly* from $2\mathbf{F}_o$-$\mathbf{F}_c$ nuclear density maps (Schoenborn *et al.* 1996). The neutron structures of several important proteins have been elucidated, including trypsin (Kossiakoff *et al.* 1980; Kossiakoff *et al.* 1981), RNaseA (Wlodawer *et al.* 1983), insulin (Wlodawer*, et al.* 1989), concanavalin A (Habash *et al.* 2000), rubredoxin (Chatake *et al.* 2004; Kurihara *et al.* 2004), endothiapepsin (Coates *et al.* 2001; Cooper *et al.* 2000), xylose isomerase (Hanson *et al.* 2004), perdeuterated myoglobin (Shu *et al.* 2000) and perdeuterated aldose reductase (Hazemann *et al.* 2005). With the improvements in detector technology (Cipriani *et al.* 1996) (Niimura *et al.*

1997) and software development (Langan *et al.* 2004b), the re-emergence of spallation neutrons in protein crystallography (Langan *et al.* 2004a) (Schoenborn *et al.* 2004), and the application of deuterium labeling of macromolecules (Meilleur *et al.* 2005), neutron crystallography is becoming a more available and accessible tool in structural biology. It has been used in the present research to further probe the structure of DHFR bound to MTX, and a comparison is made to the ultrahigh resolution X-ray structure also reported here. In order to undertake a neutron crystallography experiment, special considerations and preparations must be made such as improving crystal diffraction quality and enhancing crystal volume and deuterium labeling either by direct soaking of samples or by perdeuteration. These are described in chapter 3, the first data chapter. The next chapter provides the results and analysis of the X-ray and neutron structures of DHFR/MTX. The final data chapter (chapter 5) describes results from high resolution X-ray crystallography experiments on apo DHFR and a ternary complex of DHFR bound to MTX and NADPH. Finally, chapter 6 provides a brief summary of the most important results from each of the data chapters (3-5), general conclusions about what these results may mean in the larger context of DHFR structure and function, and a few possible future directions for the research to continue in the next few years. The next chapter (2) is exclusively methods-based and provides much detail concerning the theory and practice implemented in the experiments described in chapters 3-5.

Chapter 2: Methods

2.1 Expression, purification, and crystallization of DHFR and DHFR complexes

DHFR was purified from the SK383 strain of *E. coli* (Zeig *et al.* 1978) that contains a pUC8 plasmid encoding the gene for expression of DHFR (Bystroff *et al.* 1990). The SK383 *E. coli* strain is grown in Terrific Broth in the presence of ampicillin (100 μg ml^{-1} final concentration). Expression of DHFR is constitutive, not induced, and the cells are grown by shaking at 250rpm at 37°C for ~72 hours. The cells are then harvested by centrifugation, flash frozen in liquid nitrogen, and stored at -80°C until ready for purification. Complete purification protocols can also be found in (Poe *et al.* 1972; Taira *et al.* 1987). Briefly, SK383 cells over-expressing DHFR are lysed with the addition of lysozyme and subsequent sonication (Fisherbrand Dismembranator) using a 2.5cm tip. After treatment with 0.05% (v/v final; from a freshly made 5% stock solution) Brij58 detergent solution to ensure complete lysis, the lysate is centrifuged for 45-60 minutes at 17,500 x g. The supernatant is subjected to streptomycin sulfate (0.01% v/v final; from a freshly made 0.1% stock solution) and ammonium sulfate precipitations (added gradually as a solid at 40 and 90% saturation). The final precipitate (the 90% ammonium sulfate pellet) contains DHFR, and this can be stored at -20°C indefinitely. After resuspension of the pellet and dialysis against MTX loading buffer (1 x P.E. (Phosphate and EDTA) buffer: 50 mM KPO_4 (pH 6), 1 mM dithioerythritol (DTE), and 1 mM EDTA), DHFR is purified using an MTX agarose inhibitor affinity column (Sigma Chemical Co., MO, USA). The column is washed with 4-5 bed volumes of loading buffer and then 4-5 bed volumes of loading buffer with 1 M KCl added to it. At this point, the

column is placed on-line with a Biologic HR HPLC chromatography system (BioRad, Hercules, CA), and at least 5 bed volumes of pre-elution buffer (0.2 M Sodium borate (pH = 9), 1 M KCl, 1 mM EDTA, and 1 mM DTE) are washed through the MTX affinity column at a flow rate of 1 ml min^{-1}. A major contaminant normally elutes in this step after 1 bed volume of pre-elution buffer has completely passed through the column. Elution of the DHFR from the MTX affinity column is performed by addition of 2 mM folate to the pre-elution buffer and a change of sodium to potassium for the borate buffer component (pH = 9). About 15 bed volumes of elution buffer are washed through the column; DHFR almost immediately begins to elute. Fractions of 1-5 ml (depending on the size of the preparation and the amount of MTX resin used) are collected; normally, DHFR completely elutes in the first 5 bed volumes. After the affinity column, DHFR is >90% pure as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). After dialysis against 1 x T.E. (Tris and EDTA) buffer (10 mM Tris-HCl (pH 7.2), 1 mM DTE, and 1 mM EDTA), the protein solution is loaded onto a DEAE-Sephacel weak anion exchange column (Pharmacia, NJ, USA). This is used to remove folate, the competitive ligand employed to elute DHFR from the MTX affinity column. DHFR and folate differentially elute from the DEAE column based on the ionic strength of the buffer. DHFR elutes at 0.2-0.3 M KCl while the folate remains bound until the resin is washed with 0.4-0.5 M KCl (**Figure 2.1 A**). After loading the dialysis solution onto the column, it is washed with 1-2 column volumes of 1 x T.E. The DEAE column is then placed on-line with the BioLogic HR system and a salt gradient elution step of 15 bed volumes is run. The 1 x T.E. buffer is mixed with the same buffer but with 1 M KCl added for the gradient run from 0 to 0.5 M KCl. Fractions of 1-5 ml are collected for the

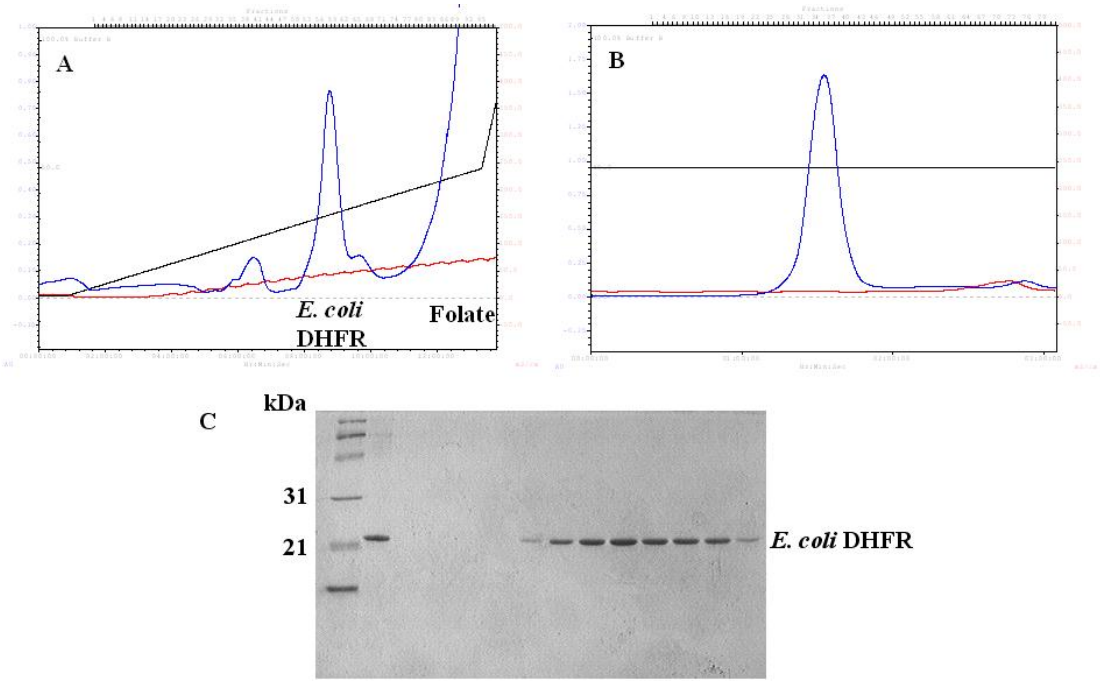


Figure 2.1: Purification of *E. coli* DHFR from SK383 cells. A) After MTX affinity chromatography, DHFR is loaded, washed and eluted via an ionic strength gradient from an anion exchange column. DHFR elutes at a weaker ionic strength than folate. **B)** A final polishing step over Superdex 75 gel filtration is used immediately prior to crystallization trials. **C)** Reducing SDS-PAGE (15% polyacrylamide) after gel filtration.

entirety of the salt gradient. An additional step of size exclusion chromatography (Superdex G-75; GE Healthcare, Piscataway, NJ) was incorporated into the purification to ensure sample homogeneity. The molecular size of species which elute from a gel filtration column can be determined if one has calibrated the column by injecting multiple pure proteins of differing but known monomeric size (e.g. standards like chymotrypsin and BSA). The apex of the DHFR elution peak is at about 35.5 ml after the end of the void volume. Using the calibration equation derived for the G-75 gel filtration column,

$K_{av} = (V_e - V_o)/ (V_t - V_o)$ (Equation 1)

where K_{av} is the partition coefficient of the eluted species, V_e is the elution volume (in mL), V_o is the void volume (in mL; the mobile phase; species which elute here are larger than the spacing between the gel beads; blue dextran (> 500,000 Daltons (Da)) is commonly used for this purpose), and V_t is the total volume of the column (in mL; packed gel and solution volume), one can closely estimate the mass of the eluted species because the K_{av} is related to the log of the molecular weight. Using a calibration equation and curve derived from the standards and from DHFR's elution volume, the molecular weight of the purified *E. coli* DHFR calculated to a range of 17,800 to 18,500 Da. DHFR's MW has been more accurately measured by FT-ICR and MALDI-TOF mass spectrometry (MS) (a description of the FT-ICR experiment is below in the perdeuteration section) and is 17,999.220 Da. For protein that would be used for crystallization, fractions from the extremes of the peak (i.e. peak "shoulders") were omitted. Normally, at this stage, the only elution peak could be attributed to DHFR (**Figure 2.1 B**).

Ultimately, purity was determined visually by SDS-PAGE (**Figure 2.1 C**) and by a UV spectrophotometer using an A_{280}/A_{260} ratio, a ratio of 1.7-1.9 being optimal for crystallization-quality ecDHFR. This allows us to quantify the ratio of protein to any contaminating nucleotides and also to folate. Protein concentration is measured by both the Beer-Lambert and the Bradford method (Bradford 1976). In the Beer-Lambert method, one must know with precision the molar extinction coefficient (ε) at a particular wavelength, for example near a λ_{max} like 280 nm. Most proteins have characteristic absorbances at 280 nm due to the presence of aromatic side chains on Tyr, Trp, and Phe. Then, using the equation,

$$A_\lambda = \varepsilon c l \qquad (2)$$

where A_λ is the absorbance of the protein at a given wavelength, c is the concentration of the absorbing species, and l is the path length of the cuvette used for the measurement. The ε for DHFR is 33.1 mM cm^{-1}. The Bradford protein estimation assay is based on the binding of Coomassie dye to proteins, specifically with aromatic and arginine side chains. After a brief incubation of the Coomassie solution with 0.25-50 μl of the protein sample, the absorbance of the mixture is measured at 595 nm. A standard curve where BSA was used as the control sample is consulted to estimate the amount of protein used in the assay. Solving for x in the standard line equation,

$$y = mx + b \qquad (3)$$

where y is the absorbance, m is the coefficient and b an integer known from the linear regression calculations from the BSA standardization data, provides the concentration of the protein sample. The yield of pure DHFR is normally 12-15 mg per liter of TB media.

The protein is concentrated to 0.75-1.5mg/ml with an Amicon or Centriprep YM10 membrane filtration device (Millipore, Bedford, MA). If a substrate or substrate analog (like MTX) was to be used, it was added (as a solid and at 5x molar ratio) while the protein is relatively dilute, due to these ligand's relative insolubility at high concentrations. After a 15-30 minute incubation with MTX in the dark at 4°C, DHFR/MTX is then concentrated with a Centricon YM10 (Millipore) device until the volume is 1/10 of the starting volume. It is then rediluted back to the starting volume with crystallization buffer (0.1 M Tris-HCl (pH 7.2), 2 mM dithioerythritol (DTE) and a trace amount of MTX) and reconcentrated in the same manner. This step is repeated twice. The protein and protein complexes were concentrated until they had reached $\geq$ 30 mg ml^{-1}. If $NADP^+$ or NADPH cofactor was to be incorporated into the complex for cocrystallization, it was added at this point (see below). All the ligands described here are light-sensitive and hygroscopic, so care was taken with the compounds to minimize exposure to light and to the atmosphere. For the apo crystallization, the protein was simply concentrated and rediluted as described above with no addition of ligands. To minimize oligomerization and nucleation events, immediately prior to crystallization, the protein or protein complex was spun at 40,000 rpm (155,000 x g) for 30-90 minutes in a Beckman TL-100 ultracentrifuge. Only the top 2/3 of the supernatant was decanted and used for crystallization. All crystallizations were performed and stored at 4°C. All initial crystal hits were grown from the Hampton Crystal Screens (Hampton Research, Aliso Viejo, CA), which employ the sparse matrix approach and test multiple crystallization parameters including pH, ionic strength, and precipitant concentration (Jancarik 1991). All hanging drop vapor diffusion crystallizations were performed in either Linbro 24-well

tissue culture plates or Nextal 24-well crystallization plates (Nextal Biotechnologies, Montreal, Quebec, Canada).

DHFR/MTX crystals were initially grown by hanging drop vapor diffusion by mixing equal volumes (1μl each) of the concentrated protein solution (30 mg ml^{-1}) in 0.1 M Tris-HCl (pH 7.3) and 1 mM DTE with 0.1 M Na-HEPES (pH 7.5), 0.3 M $MgCl_2$, and 30% PEG 400 (Crystal Screen #1, condition #23; Hampton Research). The size and diffraction quality of these crystals have since improved by changing $MgCl_2$ to $CaCl_2$ and lowering the molarity to 0.2 M and by lowering the precipitant concentration down to $\leq$ 18% (v/v). This is detailed in a later section of this chapter.

Apo DHFR crystals were grown by hanging drop vapor diffusion by mixing equal volumes (1μl each) of the concentrated protein solution (40 mg ml^{-1}) in 0.1 M Tris-HCl (pH 7), 2.5 mM Lee compound #568, and 2 mM DTE with 0.1 M Na-HEPES (pH 7.2), 50 mM $MgCl_2$, and 30% PEG MME 550 (an optimization of Index Screen condition 55; Hampton Research). Lee compound #568 is a proprietary sulfamethoxazole analog synthesized for and kindly provided to us as a gift from Dr. Richard Lee (University of Tennessee Health Sciences Center; Memphis, TN). It was used as an additive for our initial crystallization screens for apo DHFR.

To form the MTX/NADPH ternary complex, a 5-molar excess (v/v) of NADPH (Alexis Biochemicals, San Diego; from a 0.1 M stock solution) was added to a concentrated solution (30mg ml^{-1}) of DHFR/MTX and allowed to incubate in the dark for 30 minutes at 4°C. Crystals of the DHFR/MTX/NADPH ternary complex were grown by hanging drop vapor diffusion by mixing equal volumes (1μl each) of the complex in 0.1

M Tris-HCl (pH 7.4) and 2 mM DTE with the reservoir solution, 0.1 M Bis-Tris (pH 5.5), 0.2 M $(NH_4)_2SO_4$, and 25% PEG 3350 (Index Screen condition 66; Hampton Research).

2.2 Expression and purification of *E. coli* DHFR from a Sumo fusion construct

Recently, the gene for *E. coli* DHFR has been cloned into a pET-Sumo vector (Invitrogen) so as to create an alternative method for purification. Gradient PCR using *Taq* polymerase (NEB, Beverly, MA) and primers (synthesized by Operon) corresponding to the 5' and 3' ends of the *E. coli* DHFR gene were used for coding DNA amplification. The use of *Taq* is necessary to provide amplified gene products with A-overhangs. The TA cloning system, devised by Invitrogen, allows one to ligate genes of interest into expression vectors without the *a priori* use of restriction enzymes. Instead, the A-overhang is used to anneal the gene into the plasmid vector by the use of an inherent T-overhang on the vector. T4 DNA ligase (Fisher Scientific) is used to covalently link the gene to the vector. Mixed primer PCR (one T7-based from the vector and the other a gene-specific primer) and DNA sequencing (at the MBRF in Walters) were used to confirm the orientation and sequence of the clone. The pET-Sumo-DHFR plasmid was transformed into chemically competent BL21(DE3) *E. coli*, expression of the N-Sumo-DHFR-C fusion was induced by the addition of 0.5 mM isopropylthiogalactoside (IPTG) to a liquid culture at an OD_{600nm} of 0.5, and the culture was allowed to grow with shaking (225 rpm) at 37°C for an additional 4 hours. The fusion construct possesses a hexahistidine tag at the N-terminus of Sumo. So, after cell lysis, clarification of the supernatant by centrifugation and pH adjustment of the solution to 8.0, the supernatant was incubated with Ni-NTA resin (Qiagen) with slight shaking at

4°C for one hour. The fusion construct may be eluted by the addition of 0.5 M imidazole. To attain native DHFR, the purified fusion construct was incubated with GST-Ulp1 (the Sumo protease from *S. cerevisiae*) for 3 hours at room temperature with slight shaking (**Figure 2.2**). The proteolytic activity of Ulp1 is quite specific; DHFR prepared this way is assumed to have a native N-terminus and, unless occurring in the bacterial cell, DHFR is not bound to any pterin-ring containing ligand. GST-Ulp1 is used as bound to glutathione beads. Therefore, when the assay is completed, the reaction mixture is centrifuged, and the supernatant is native DHFR and 6xHis-Sumo while the GST-Ulp1 bound to glutathione beads remains in the pellet. (In the normal purification method of DHFR from SK383 *E. coli* cells described earlier in this chapter, folate is used to elute DHFR from the MTX affinity column in a middle step in the purification process. This alternative procedure, with the use of the His tag, exempts the MTX affinity step.)

2.3 Initial X-ray diffraction analysis

Crystals were tested for X-ray diffraction first using our in-house R-AXIS IV++ diffractometer (Rigaku/MSC; The Woodlands, TX) (**Figure 2.3**). Crystals were cryo-protected by soaking them in cold harvesting buffer (the crystallization buffer made with the incorporation of glycerol) for 15-30 seconds; 10% glycerol was used for the apo crystals, 20% glycerol for the MTX/NADPH crystals. Crystals were then either flash-cooled under LN_2 or in a stream of N_2 gas and diffraction images were collected at -160°C. Initial indexing revealed that the DHFR/MTX complex crystallized into a trigonal/hexagonal Bravais lattice with unit cell dimensions of a=b= 92.2 Å, c= 73.3 Å ($\alpha=\beta=90°$, $\gamma=120°$), the apo DHFR form crystallized in the *P*6 space group with unit

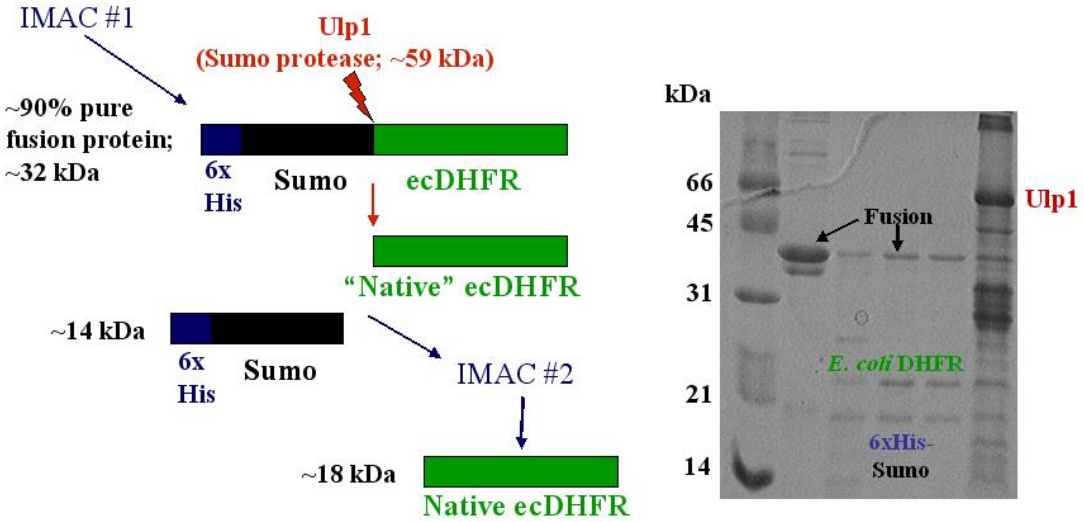


Figure 2.2: Purification of *E. coli* DHFR using a 6xHis-Sumo fusion construct. An N-terminal 6xHis tag is used to facilitate isolation of the DHFR fusion protein (IMAC #1). Ulp1 is a very specific cysteine protease that cleaves at the C-terminus of Sumo (Li *et al.* 2003; Li *et al.* 1999), thus, releasing native DHFR (no exogenous residues post-cleavage) and the His-Sumo section. A second metal affinity step (IMAC #2) is used to purify the DHFR from the His-Sumo. The Ulp1 is also a fusion construct with a GST tag at its N-terminus. The GST-Ulp1 is bound to glutathione beads during the proteolytic assay and, thus, does not contaminate the DHFR sample.

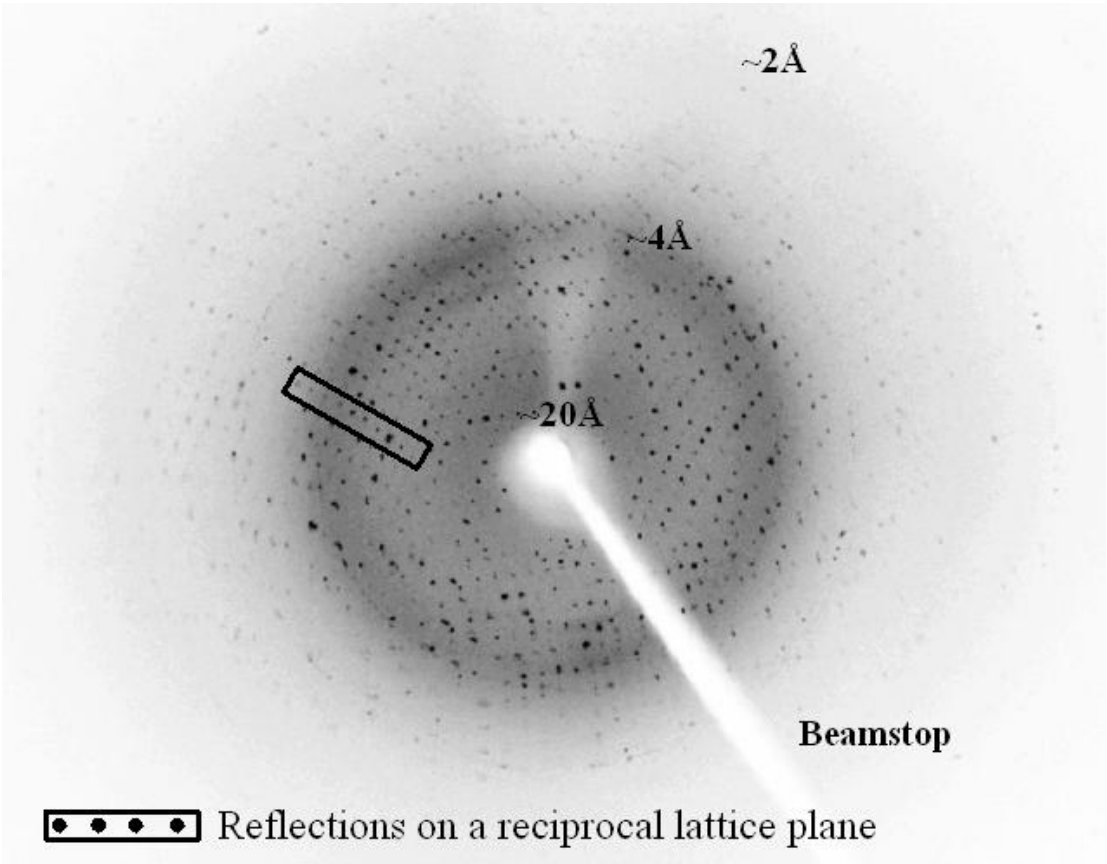


Figure 2.3: An example X-ray diffraction pattern from a DHFR crystal (MTX/NADPH) measured using the in-house diffractometer. Approximate resolution values (real space) for particular regions on the detector (reciprocal space) are shown, as is the position of the scattering shape and pattern of the beamstop used. A group of reflections comprising part of a circular line pattern has been highlighted with a black box and is most likely a group of points lying on the same plane of the reciprocal lattice.

cell dimensions of a=b= 95.8 Å, c= 34.3 Å (α=β= 90°, γ= 120°), and the MTX/NADPH ternary complex crystallized into the *P*3 space group with unit cell dimensions of a=b= 61.9 Å, c= 104.7 Å (α=β= 90°, γ= 120°). The best candidates for atomic resolution experiments at the synchrotron were immersed under LN_2 and stored in a dry shipping dewar until synchrotron data collection.

2.4 Crystallography

Data collection: Oscillation method

Strategies for the determination of some of the data collection parameters can be obtained by using processing programs like HKL2000 (Otwinowski *et al.* 1997); an example of this is the estimation of the oscillation range necessary to have good overall completeness ($\geq$ 90%). However, it is best to understand this in terms of the relationship between real and reciprocal space and by the sphere of reflection (also called the Ewald sphere; **Figure 2.4 A**). The spots on a diffraction pattern can be thought of as parallel planes comprising the reciprocal lattice which intersect the Ewald sphere. Reciprocal lattice points which do not intersect or lie within the Ewald sphere are not recorded as reflections. One must rotate the crystal about the spindle axis to make these spots intersect the sphere (Dauter 1997a). Two factors related to the full recording of reflections and data completeness are the crystal mosaicity (the smallest angle that the crystal can rotate and a reflection is fully recorded; it is a diagnostic value associated with lattice uniformity and/or deformation) and the magnitude of the individual oscillation angles per step rotation. If the rotation angle is too small and also happens to be smaller than the mosaicity, then many reflections are recorded as partials because the reflections at the extremes of the rotation range will just be starting to intersect the Ewald sphere

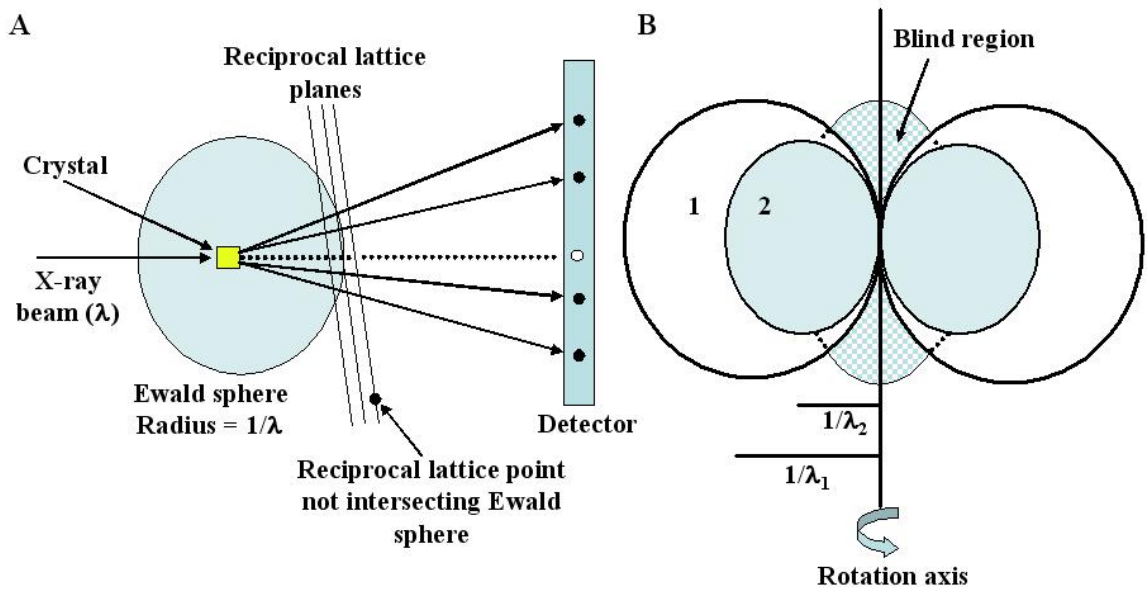


Figure 2.4: The relationship between real and reciprocal space relative to X-ray data collection. A) Conditions for X-ray diffraction are satisfied when points on the reciprocal lattice intersect the Ewald sphere. **B)** A region of the reciprocal lattice remains external to the Ewald sphere, even if full rotation of the crystal is accomplished. Two ways to ameliorate this is by using a shorter incident X-ray wavelength (i.e. λ_1 instead of λ_2), which provides a larger Ewald sphere (observe the magnitude of sphere 1 *vs.* sphere 2) and/or by rotating the crystal about another axis (or by simply off-setting the crystal).

(Gewirth 1999). However, programs such as HKL2000 can take care of partials quite well. An additional concern in data collection is the treatment of reflections which reside in the blind region (**Figure 2.4 B**), an area of the reciprocal lattice that never crosses the Ewald sphere during a ϕ-scan. The use of a shorter X-ray wavelength is helpful here because the Ewald sphere is much larger and the area of the cone encompassing the blind region is much smaller. (The X-ray wavelength can be changed at many synchrotron beam lines.) Additionally, the blind region can be recorded with a goniometer that has κ-geometry, which enable a 45 degree tilt from the crystal's orthogonal position with respect to the X-ray beam (offsetting the crystal).

Ultrahigh resolution data collection

Special considerations must be made when collecting diffraction data to atomic resolution (beyond 1.2Å). A synchrotron X-ray source allows one to do this, due to its strong flux and the ability to tune the wavelength. Use of shorter X-ray wavelengths (such as 0.9Å) expands the theoretical sphere of reflection (radius = $1/\lambda$) (Ladd *et al.* 1993) (Dauter *et al.* 1997b). Most X-ray detectors can be positioned quite close to the sample so as to record high θ-angle reflections, which would correspond to high resolution reflections. Additionally, one can set very long exposure times (30-60 sec) so that the high resolution reflections ultimately possess strong intensities for better data processing. To ensure good completeness ($\geq$ 90% overall, $\geq$ 70% in the highest resolution shell), one should collect highly redundant data about one spindle axis (dependent on the crystal symmetry but generally a 90-180° rotation) and, if possible, about an additional axis to provide coverage in the blind region. Unfortunately, the long exposure times and

the extensive total data collection can lead to radiation damage within the crystal, even when the experiment is performed at cryogenic temperatures (Burmeister 2000). During the process of collecting high resolution data, detector overloads are commonplace at areas of reflections which are medium to low resolution. It is now well-documented that one should collect multiple passes on the same crystal, with exclusive scans for the recording of high and lower resolution reflections (Howard *et al.* 2004) (Schmidt *et al.* 2005). The major differences between the scans are the positions of the detector and beam stop position, exposure time, and step width between images. In fact, the detector can even be off-set on an axis perpendicular to the incident X-ray beam to facilitate the collection of very high resolution reflections (beyond 0.9Å for the ADSC Quantum 4 detector).

Laue data collection

The neutron data sets described within this thesis were collected using pseudo-Laue (ILL) and wavelength-resolved Laue (LANSCE) methods with a wavelength range of 0.6-7.0Å for the PCS experiment and 2.8-3.7Å for the ILL experiments. The number of reflections which can be recorded directly depends on the wavelength range used. In contrast to the use of a monochromatic X-ray beam for the oscillation method, more coverage of reciprocal space from a stationary crystal is possible. Instead of fine-slice oscillation, the crystal can be moved into different settings which may be several degrees apart, and if the crystal is of higher symmetry (like the $P6_1$ DHFR/MTX crystals described in this thesis), then 15-25 settings may suffice to attain high data completeness ($\geq$ 75% for neutron data sets). However, there are three major obstacles to overcome in order to effectively process and analyze Laue diffraction data. The first is a problem of

multiple reflections arising from a set of planes that are harmonics of one another but contributing to one Laue spot (Moffat 1997). These reflections actually fall on a line of the reciprocal lattice that passes through the origin of the resolution sphere; the planes reflect in the same direction as *hkl* but with different *d*-spacings and a different wavelength. As it turns out, most of the recorded reflections (and nearly all those at the highest resolution) are singles (Drenth 1999). But deconvolution of the multiples is still necessary to acquire the low resolution data. A basis for unscrambling was proposed by Ren and Moffat (Ren *et al.* 1995) where a least-squares refinement procedure can produce accurate estimates for the structure factor amplitudes of separate reflections within each multiple. Secondly, spatial overlapping of neighboring spots in the Laue pattern makes data processing difficult. More painstaking profile fitting where non-overlapping spots are used as the basis profile and are applied to overlapped spots in similar detector regions (Yang *et al.* 1998). If the position of the detector can be changed (which is possible for most Laue set-ups at X-ray facilities but not feasible at the current neutron beamlines), the distance could be increased to increase the space between reflections. Lastly, wavelength normalization is undertaken so as to correct for factors, such as sample absorption, which are highly influenced by the wavelength range used and can have effects on the Laue intensities. A wavelength dependent factor is multiplied to every reflection intensity; often the factor is determined by using the intensity of symmetry-equivalent reflections measured at different wavelengths within the same image (Ren *et al.* 1995). This has been implemented recently for normalizing the wavelength range used in neutron experiments (including the DHFR work reported here); a normalization curve is generated and the extremes and the peak of the incident

spectrum can be assessed (Arzt *et al.* 1999) (Bennett *et al.* 2005). With the time-of-flight (TOF) Laue method for neutron data collection used at LANSCE, one can bin reflections based on the time the scattered neutrons reached the detector and, thus, the wavelengths of these neutrons can be derived (Langan *et al.* 2004b). This ability to resolve the wavelength of the neutrons and associate them directly with reflections that arise from their detection provide a way to deconvolute the diffraction data, reducing problems from spatial overlap of reflections and background accumulation (Schoenborn *et al.* 2004).

Indexing, integration, and scaling of the X-ray intensities

As one begins data collection and observes many diffraction spots (or peaks) which accumulate in characteristic patterns on the image, one can then deduce a crystal lattice system that may give rise to such a pattern. Reflections arising from a particular reciprocal lattice plane are indexed by the assignment of an *hkl* coordinate set; this occurs by a complete search of all possible indices from all selected diffraction peaks. When integer values have been identified for one index (such as *h*) for all reflections, this is essentially finding one real space direction (such as **a**) in the unit cell of the crystal, thus, this is real space indexing and the vector search is accomplished by a Fast Fourier Transform (FFT) (Gewirth 1999). When the search for reciprocal space vectors is completed, indexing of all reflections within the image is attempted. This provides an initial basis for description of the unit cell. Manual intervention is necessary to determine whether cell reduction is possible; the unit cell constants are determined for all the Bravais lattices and, based on how distorted the observed unit cell must become in order to fit the particular lattice type, one may decide to transform the initial basis cell to one of higher symmetry. Once the cell and symmetry have been agreed upon, one may further

refine values via the least squares method for crystal and detector parameters over several images in the indexing routine, and this adjustment normally improves error values (χ^x and χ^y) associated with the deviation of centers of reflections from predicted positions based on the indexing solution. An accurate prediction of spot positions is needed for proper profile fitting of diffraction peaks; background on the detector is subtracted from the intensity of the reflection (**Figure 2.5 A**). This is needed for a highly precise integration of the reflections (**Figure 2.5 B**).

Polarization correction

A polarization correction is applied to integrated intensities so as to correct for the differences in polarization of the incident and the diffracted X-ray beam and how this affects the intensities of reflections. Oscillations of electrons "induced" by the polarization component of the incoming beam perpendicular to the lattice plane do not necessarily contribute to the scattering; therefore, the expected scattered intensities are reduced. The X-ray beam generated at a synchrotron source is strongly polarized. So, the polarization factor as applied in common data processing programs like Denzo (Otwinowski *et al.* 1997) is similar to the formulation of Azaroff (Azaroff 1955),

$$P = (I_h - I_v) / (I_h + I_v), \qquad (4)$$

where I_h is the intensity of the horizontal component and I_v is the intensity of the vertical component of the polarized X-ray beam.

Data reduction and scaling

The merging of multiple observations of reflections (symmetry equivalent or redundant reflections) and the overall scaling of integrated intensities is commonly performed using the method of Fox and Holmes (Fox *et al.* 1966), the utility of which is

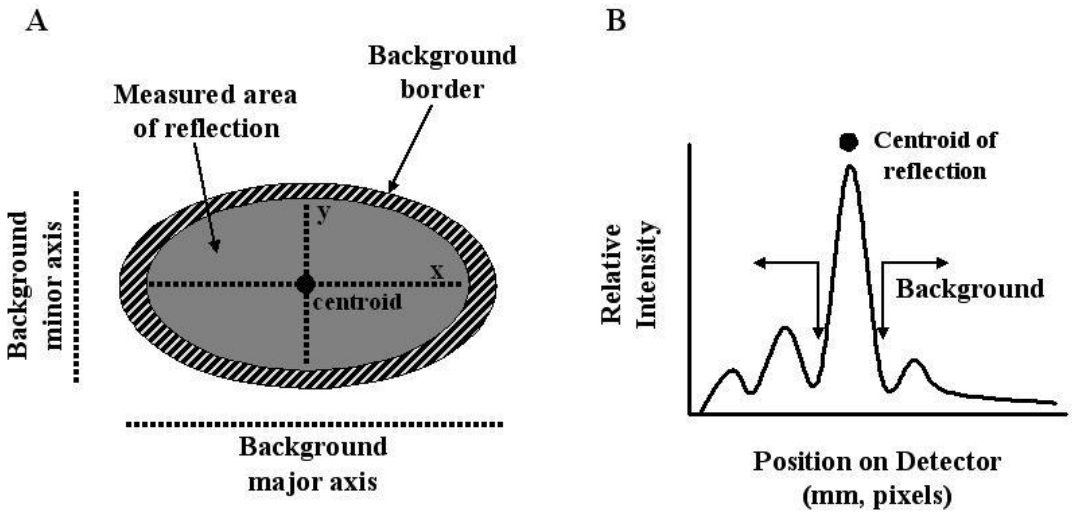


Figure 2.5: Profile fitting and integration of intensities on an X-ray detector. (A) One must decide on a spot size and shape (as well as the difference between spot and background) for the intensities that will be included in the **(B)** integration.

that one can scale data from multiple "layers" (or integrated intensity sets arising from individual images) and apply an estimated error σ to the overall scale factor as well as to input intensities. Poor agreement between multiple observations is cause for rejection. The calculation of an error factor based on the merging of symmetry-related data, or R_{merge},

$$R_{merge} = \Sigma_{hkl} \Sigma_i |I_{i(hkl)} - I_{hkl}|/\Sigma_{hkl} \Sigma_i I_{(hkl)}, \qquad (5)$$

where *hkl* is the reflection, i is the observation number of that reflection intensity, and $I_{i(hkl)}$ is the intensity of the reflection, and I_{hkl} is the average intensity of all the observations of that reflection, is performed by the scaling program Scalepack (Otwinowski *et al.* 1997). Additionally, Scalepack also allows one to globally refine the crystal mosaicity and the unit cell. Additionally, scaling affords one a statistical analysis of the quality and the practical extent of resolution of the diffraction data, by providing values for R_{merge}, χ^2, overall and individual I/σ. An R_{merge} of less than 10%, a χ^2 of less than 2, and an I/σ of >2 are generally considered satisfactory statistics (Gewirth 1999).

Intensities, structure factors, and the electron density function

Intensities are directly proportional to the square of the amplitudes,

$$I(hkl) = |F(hkl)|^2, \qquad (6)$$

and the overall structure factors (the **F**s) can be derived from the scattering factors,

$$\mathbf{F}(\mathbf{S}) = {}^{n}\Sigma_{j=1} f_j \exp[2\pi i r_j \mathbf{x} \mathbf{S}]. \qquad (7)$$

The total scattering from electrons within the unit cell is described in equation 7, where f_j is the atomic scattering factor for r_j positions of n atoms in the unit cell, **S** is a vector perpendicular to the plane of reflection (a vector between reciprocal lattice points, an *hkl*

reflections), and $2\pi i r_j$ **x S** is the modifier for phase angles if the scattering origin is shifted to the origin of the unit cell. The merged, averaged intensities from scaling are truncated to these mean structure factor amplitudes, and the method of French and Wilson is employed to produce sensible estimates of amplitudes, even ones derived from negative intensities (French *et al.* 1978). From this treatment, **F**s and their standard deviations (sigFs) are output.

Note that within the exponential of equation 7 there is an imaginary component. This applies to the phase component, α, of the structure factor. Essentially, one only knows the absolute value ($|\mathbf{F}(\mathbf{S})|$, or the magnitude) of the complex vector, $\mathbf{F}(\mathbf{S})$; the missing component is the phase information, or $\alpha(\mathbf{S})$. The real importance in terms of structure determination can be signified in the electron density equation,

$$\rho_{(xyz)} = 1/V\ \Sigma_h \Sigma_k \Sigma_l\ |F_{(hkl)}|\ \exp[-2\pi\ \mathbf{x}\ \ i(hx + ky + lz - \alpha_{hkl}], \qquad (8)$$

where $\rho_{(xyz)}$ is the electron density distribution in an xyz real space position. Equation 8 reveals that the relationship between electron density and the structure factors is a Fourier transformation. However, the equation cannot be completely "solved" without the phase angles, α_{hkl}, and, for macromolecular structure determination, must be found using indirect means, such as the molecular replacement method that is described next.

Molecular Replacement (MR)

The method of structure determination described in this thesis used for the solution of the initial DHFR/MTX complex was molecular replacement (MR). A homologous macromolecular structure (a target molecule) can be used to provide the phase information for the data collected from a crystal of a macromolecule whose

structure is not known (the search molecule). So, the intensities for the target can be calculated from its structure (both $|F_{(hkl)}|$ and α are known) while only the intensities are known for the search molecule. A six-dimensional search in two defined steps is undertaken in MR to detect the *orientation* (via a rotational search using the target molecule), and the *positioning* (via a translational search using the properly oriented target molecule) of the search molecule in the new unit cell (Taylor 2003). The method was first described by Rossman and Blow (Rossman *et al.* 1962).

The MR search process can be put in quantitative terms,

$$X_{new} = \mathbf{R}\{X\} + \underline{d}, \qquad (9)$$

where X_{new} is the position of the molecule to be identified, X is the position of the target molecule, **R** is rotation (or orientation) matrix, and $\underline{d}$ is the translation vector. The latter two factors are the computations necessary to bring the target into coincidence with the unknown molecule. A limited resolution range (e.g. between 8-3Å) is suggested for MR. To reveal the orientation of the search model, a rotation function describing the angular relationship between identical molecules in the asymmetric unit (a self-rotation function) or between similar molecules in different crystal forms (a cross-rotation function) must be implemented (Drenth 1999). A self-rotation function can also identify non-crystallographic symmetry. The best orientation is the one in which provides for the maximum overlap of the intramolecular vectors (the Patterson self-vectors) from the target molecule and the search model. The proposed Patterson overlap function (the cross-rotation function) by Rossman and Blow (1962) is given as an integral:

$$\mathbf{R} = \int U(x)P_T(x)P_S(\{X\}\underline{x})\,\delta V, \qquad (10a)$$

where $P_T(x)$ is the Patterson vector for the target molecule, $P_S(\{X\}\underline{x})$ is the Patterson vector for the search model, and U(x) is a shape function representing a sphere within which the rotational search occurs. Ultimately, the rotation function can be described as a summation:

$$\mathbf{R}(\alpha,\beta,\gamma) = \Sigma_h \Sigma_p \, |F_h|^2 \, |F_p|^2 \, \mathbf{x} \, G_{h,p}, \qquad (10b)$$

where F_h and F_p are respective structure factors for the target and the search molecules, $G_{h,p}$ is a Fourier transform of the shape and volume function, and $R(\alpha,\beta,\gamma)$ is the product rotation function and is dependent upon the angles of α,β,γ. These rotation angles and the magnitude of rotation are illustrated in **Figure 2.6** as angles in three-dimensional Euler space (Ladd *et al.* 1993). To simplify the rotation function calculation, the target molecule is arbitrarily placed within a P1 unit cell, and integration is performed over the volume, U, equal to the volume around the Patterson function origin where P_T (or P_{calc}) and P_S (or P_{obs}) are quite similar (Tollin *et al.* 1966). To avoid overlap of the self-rotation vectors, the search models must be placed in a sufficiently large unit cell, normally one and a half times the dimensions of the target molecule being used for the rotational search (Lifchitz 1983). Updated from (Rossman *et al.* 1962), a rotational search reciprocal space method for a fast rotation function was suggested by Crowther (Crowther 1972) and can be applied using the Eulerian angular and an orthogonal coordinate system (**Figure 2.6**).

Once a suitable rotation solution (the **R** matrix) has been found, a translation vector should be derived in order to overlap the target onto the search molecule in real space. Intermolecular vectors (cross-vectors) instead of self-vectors are used for several two-dimension translational searches, the translation function defined from (Crowther *et al.* 1967) as a sum of individual two-dimensional T-functions into one three-dimensional

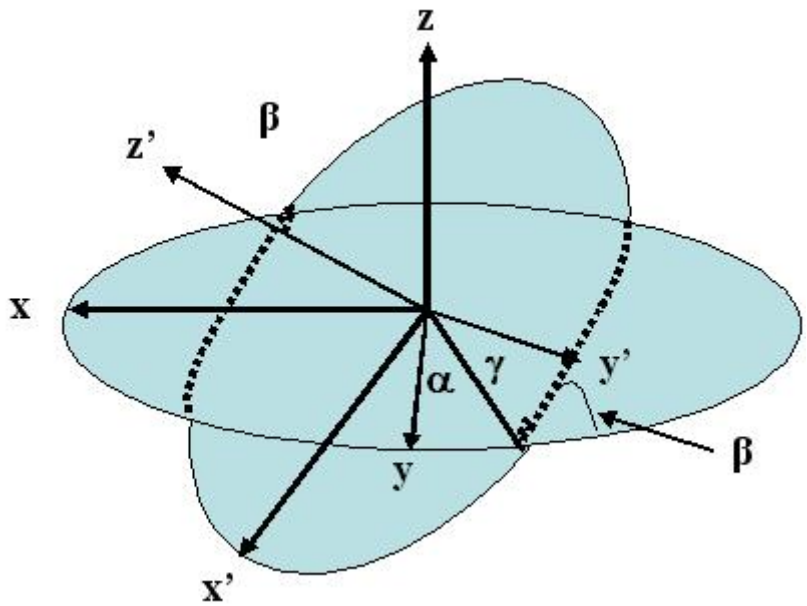


Figure 2.6: Euler angles (αβγ) and orthogonal coordinate system (xyz) used in the MR rotation function. From the origin, the rotational search is conducted in a counterclockwise manner, α along z, β along the "new" y axis (y'), and γ along the "new" z axis (z'). So, the xyz axes are rotated and the angles describe rotation of the molecule along αβγ while the reference frame (the new xyz axes after an incremental rotation) is fixed. The highest order symmetry axis is made to be along z in the Euler system.

T_2-function where many symmetry operators are utilized for the translational search. This form of the function is

$$T_2(t) = \int (P_{obs}(\underline{u}) - P_m(u)) \mathbf{x} P_{calc} (\underline{u},\underline{t})\, d\underline{u}, \qquad (11)$$

where $P_{obs}(\underline{u})$ is the Patterson vector for the target derived from the observed intensities, $P_m(u)$ is a sum of Patterson vectors for n asymmetric units for the model in their proper orientations, and P_{calc} $(\underline{u},\underline{t})$ is the calculated Patterson function (Drenth 1999). A non-crystallographic translation function can be implemented for crystals which possess more than one molecule in the AU where one molecule can be fixed and a translational search for the second molecule is performed (Tickle 1992). Additionally, standard R-factors (R) and correlation coefficients (CC) can be calculated as the oriented target is translated throughout the asymmetric unit (Navaza 2001); essentially, $|\mathbf{F}_{calc}|$ is calculated and is compared to the $|\mathbf{F}_{obs}|$ for agreement by:

$$R = \Sigma_{hkl} \,|\, |\mathbf{F}_{obs}| - |\mathbf{F}_{calc}| \,| \,/\, \Sigma_{hkl} |\mathbf{F}_{obs}| \qquad (12)$$

and

$$CC = \Sigma\, (|\mathbf{F}_{obs}| \mathbf{x} |\mathbf{F}_{calc}|) \,/\, [(\Sigma|\mathbf{F}_{obs}|^2) \mathbf{x} (\Sigma\, |\mathbf{F}_{calc}|^2)]^{1/2}. \qquad (13)$$

For the MR solutions described in this thesis, AMoRe (Navaza 2001) and PHASER (Storoni *et al.* 2004) (McCoy *et al.* 2005) were used. AMoRe is a fully automated MR package which uses fast rotation (ROTING) and translation functions (TRAING) to sample many different configurations and positions and calculates a CC for these positions. This allows one to establish a standard so as to contrast the CC values for the top solutions (Navaza 2001). AMoRe solutions are then subjected to rigid body refinement (FITING) of their orientation and position within the unit cell (Drenth 1999).

PHASER, an alternative to AMoRe, actually implements the AMoRe method of initial calculation of functions, then rescores this plausible list of solutions using a translation likelihood target (Read 2001). Now, both rotation and translation function solutions can be attained using approximations of maximum likelihood directly in PHASER (McCoy *et al.* 2005; Storoni *et al.* 2004).

Difference Fourier method

The position of a ligand bound to or sites of modification (such as mutagenesis or post-translational modification) on a "native" protein can be very accurately determined by calculating a single ($\boldsymbol{F}_o - \boldsymbol{F}_c$) or double difference ($2\boldsymbol{F}_o - \boldsymbol{F}_c$) Fourier map. The phase angles for the native protein are already known; the coordinates for this are first refined against the reflection data from an isomorphous "derivative" of the same protein ("native" meaning the originally solved structure, such as an apoenzyme, and the "derivative" here being a ligand-bound form of the same protein). If $\boldsymbol{F}_{obs}=I_{obs}^2$ for the Fourier series, ρ_{obs} are the coefficients designated and $\boldsymbol{F}_{calc}$ are the coeffici.ents designated for ρ_{calc} and one assumes that the phase term for α_{obs} is equivalent to α_{calc}, a difference Fourier synthesis providing the difference electron density can be calculated using

$$\Delta\rho_{(xyz)} = 1/V\ \Sigma_{hkl}\ |\ F_{obs} - F_{calc}\ |\ \exp\ [-2\pi i(hx + ky + lz)\ \exp\ [i\alpha_P(hkl)]. \qquad (14)$$

F_{obs} represents the structure factor amplitudes for the derivative whereas F_{calc} represents the calculated structure factor amplitudes for the native protein. The expression for F_{calc} is

$$F_{calc} = \Sigma f_j \exp[-B_j \sin^2\theta/\lambda^2]\exp 2\pi i[hx + ky + lz], \qquad (15)$$

where f_j are the scattering factors for the atoms in the unit cell and B_j is the isotropic temperature factor ($8\pi^2$Uj). The meaning and importance of atomic temperature factors (or displacement parameters) are discussed in detail in chapter 4. The resulting difference Fourier map (from the calculation in Eq. 14) would reveal peaks at half the height of a normal Fourier map, positive peaks should appear at regions for atoms of the derivative not present in the native. A single difference density map is useful to calculate because it can also reveal unknown or *unmodeled* density features at half height, both positive and negative,

$$\rho_{(xyz)} = 1/V\ \Sigma_{hkl}\,(|\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}|)\ \exp\,[-2\pi i(hx + ky + lz) + i\alpha_{calc}], \qquad (16)$$

even after extensive rebuilding of the model has occurred. A common final step in model building and verification is to display only a difference density map and ensure that all significant density has been identified and atoms modeled into them (Kleywegt *et al.* 1997). Peaks are displayed in these maps for unknown atoms at half the height and can be modeled fairly readily. Indeed, a recent study has shown that, using medium resolution data (~3Å), an $\boldsymbol{F}_o - \boldsymbol{F}_c$ synthesis can reveal unmodeled primary (bound) solvent structure as distinct spherical peaks above 3σ whereas a $2\boldsymbol{F}_o - \boldsymbol{F}_c$ synthesis cannot (Minichino *et al.* 2003).

Related to the single difference Fourier maps are the $2\boldsymbol{F}_{obs} - \boldsymbol{F}_{calc}$ (or $2\boldsymbol{F}_o - \boldsymbol{F}_c$) maps commonly known as the double difference Fourier electron density maps one utilizes in model building. They take the form

$$\rho_{(xyz)} = 1/V\ \Sigma_{hkl}\,(2|\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}|)\ \exp\,[-2\pi i(hx + ky + lz) + i\alpha_{calc}], \qquad (17)$$

and serve as a sum of the electron density of a model and of the difference density at double the peak height (Drenth 1999). This is mainly to reveal the electron density for the

model as well as the differences between the actual structure and the model at normal height. The $2\boldsymbol{F}_o - \boldsymbol{F}_c$ (or some use $3\boldsymbol{F}_o - 2\boldsymbol{F}_c$ calculations to see peak height differences) synthesis dampens peaks at regions where atoms should not be included while enhancing the peak height for correct atom positions to well above background (Main 1979). Most of the $2\boldsymbol{F}_o - \boldsymbol{F}_c$ maps reported here have been weighted as determined by a sigma-A (σ_A) calculation (Read 1986),

$$\sigma_A = D\,(\Sigma_P/\Sigma_N)^{1/2}, \qquad (18)$$

where D describes the probability distribution of electron density ($\Delta\rho\mathbf{r}$, and is a value describing agreement of $\boldsymbol{F}_o$ and $\boldsymbol{F}_c$), Σ_N is a sum of the total atoms in the structure and Σ_P is a sum of the total atoms in the partially modeled structure. The weighting is used simply to compensate for errors in the phases arising from *wrongly modeled* atoms or *unmodeled* atoms.

Refinement

The parameters of the structure model (which can include the *xyz*, B-factor, and occupancy for every atom) must be changed in order to minimize (*F*o-*F*c). This statement can be formalized as an agreement index (an R-factor) between $\mathbf{F}_{obs}$ and $\mathbf{F}_{calc}$ and takes on the recognizable form,

$$R = \Sigma_{hkl} \,|\, |\boldsymbol{F}_{obs}| - k \,|\, \boldsymbol{F}_{calc} \,|\,| \,/\, \Sigma_{hkl} \,|\boldsymbol{F}_{obs}|. \qquad (19)$$

One attempts to minimize a function (**Q**) describing the difference of a set of observed data, $\mathbf{F}_{obs}$, to a model set of data , $\mathbf{F}_{calc}$, by manipulation of the model used to describe the data (Tronrud 2004). In the simplest manner, this can be solved by taking a series of least-squares equations,

$$\mathbf{Q} = \Sigma_{hkl}\, w(hkl)\, (|\boldsymbol{F}_{obs}(hkl)| - |\boldsymbol{F}_{calc}(hkl)|)^2, \qquad (20)$$

and finding the minimum of the **Q** function by changing the atomic parameters, u_j (j = atom 1, 2, n…), that determine the $\boldsymbol{F}_{calc}$ for a given *hkl*. Essentially, **Q** for every u_j is set to zero,

$$\Sigma_{hkl}\, w(hkl)\, (|\boldsymbol{F}_{obs}(hkl)| - |\boldsymbol{F}_{calc}(hkl)|)\, \delta\, |\boldsymbol{F}_{calc}(hkl)| / \delta\, u_j = 0. \qquad (21)$$

$\boldsymbol{F}_{obs}$ values are constants and are set from the experimental intensities. However, $\boldsymbol{F}_{calc}$ values depend on the u_j values and will change by a given amount when the model structure is changed. So, equation 23 is set up as a vast Taylor expansion of normal equations where the solution is for new values of $\boldsymbol{F}_{calc}$ (*hkl;* u_j) and its' derivatives per refinement cycle; the common expression for this expansion is a matrix. This will be discussed below. Two major assumptions of least-squares mathematics as applied to the refinement of protein structures, which have led to a critical revision in how many phasing (de la Fortelle *et al.* 1997) and refinement problems (Pannu *et al.* 1996) are approached, are that errors associated with the data ($\boldsymbol{F}_{obs}$) follow a predicted distribution and the divergence of the observation to the model (the $\boldsymbol{F}_{calc}$) is known perfectly; thus, all parameters are known with certainty and one can provide a model which can describe the observations perfectly (Tronrud 2004). To alleviate the strict conditions for confidence in the solution of least-squares calculations, maximum likelihood (ML) has been implemented into many refinement packages because one calculates the probability (not the exactitude) that a set of observations correlates to a set of model parameters (Drenth 1999). The relationship between $\boldsymbol{F}_{obs}$ and $\boldsymbol{F}_{calc}$ is then based on this likelihood and takes a form similar to Bayes' Theorem:

$$\boldsymbol{L} = p[|\boldsymbol{F}_{calc}|; |\boldsymbol{F}_{obs}|]\ , \qquad (22)$$

where the likelihood (***L***) of the model (derived from the $\boldsymbol{F}_{calc}$; which can be modified) is measured given the known, fixed values of the data ($\boldsymbol{F}_{obs}$). ML is the optimization method used in the Refmac5 (Murshudov *et al.* 1997) refinements described in this thesis.

A common early step in macromolecular refinement is to perform a highly constrained rigid body refinement. This can be considered an extension of MR because, if one sets the rigid bodies to be whole molecules or domains of a protein, then only the parameters of the constrained bodies are refined, essentially only the global orientations and location of the molecule are modified (Tronrud 2004). Also early within the refinement process, one may need to correct for large errors in the model or to reduce the bias associated with using the phases from a homologous model in MR. A form of molecular dynamics refinement, simulated annealing (Brunger *et al.* 1999) as implemented in the program package CNS (Brunger *et al.* 1998), has been applied to all the X-ray structures described in this thesis; using Newtonian mechanics and applying (*in silico*) an increase in temperature, atoms move with increased velocity and possess increased kinetic energy (Tronrud 2004). This can allow the model to sample different energy minima, hopefully "pushing" it out of any local minima (including a homologous model from MR, thus reducing "model bias") and nearer to the global minimum, thereby expediting convergence of the model towards the experimental data (Brunger *et al.* 1999).

It was stated above that solution for all the $\boldsymbol{F}_{calc}$ values (dependent on the small changes applied to the atomic parameters) took the form of a matrix which possesses n **x** n terms, where n is the number of parameters in the model and this can be > 10,000 for

macromolecules at moderate resolutions. The diagonal terms of this matrix are directly linked to the effect that that parameter change has on minimizing **Q**. Off-diagonal terms describe the correlation between parameters that have changed. Many programs, including Refmac5 (Murshudov *et al.* 1997) and SHELXH (Sheldrick *et al.* 1997), use sparse-matrix approximation to reduce the number of terms that have to be considered in the refinement calculations and, thus, reduce the computational memory and time requirements. This means that many of the off-diagonal terms are not considered and are reduced to zero (Drenth 1999). Conjugate gradient least squares refinement (Tronrud 1992) was used in SHELXH (Sheldrick *et al.* 1997). Search vectors describing the gradient of descent for the **Q** function based on the shifting of parameters in the normal matrix are compared, and a later cycle in the process "learns" from a previous cycle by including a defined fraction of the shift vector used in the previous cycle. The shift vectors are conjugates to one another and help the model to reach the minimum (Tronrud 2004). If one has sufficiently high resolution (the data-to-parameter ratio is > 5), within SHELXH, very accurate bond lengths and angles (and their estimated standard deviations, ESDs) for all atoms in a protein structure can be calculated by using all the parameters in the normal matrix, on and off-diagonal. This is called full-matrix refinement and is *very* expensive computationally due to the fact that one is now considering a square or more parameter shifts than in the sparse-matrix and conjugate gradient methods (Cowtan *et al.* 2000).

Refinement: generalizations for reported structures

This section is a general refinement protocol for all the X-ray structures described in this thesis. Specific details of the refinement protocols for individual structures are

provided within chapters 4 (MTX) and 5 (Apo and MTX/NADPH). For all the data reported here, 5% of each data set was sequestered from the refinement for an R_{free} calculation for cross-validation. For all structures, at least three cycles of maximum-likelihood refinement in Refmac5 interspersed with manual rebuilding of the model in O (Jones *et al.* 1991) (Kleywegt *et al.* 1997) were performed. Alternating cycles of ARP/wARP (Lamzin and Wilson, 1997) and Refmac5 were invoked for solvent divination. Individual isotropic B-factors were also refined at this point. Alternate side chain conformations were built if atoms can be modeled into either $2\boldsymbol{F}_o - \boldsymbol{F}_c$ electron density above 1.5σ or $\boldsymbol{F}_o - \boldsymbol{F}_c$ electron density above 2.5σ and if they adopted geometrically favorable orientations (favorable side chain rotamers). For the MTX and apo structures, further cycles of refinement, this time using the conjugate gradient least squares (CGLS) method in SHELXH (Sheldrick *et al.* 1997), and rebuilding in O were performed. In both Refmac5 and SHELXH, all non-hydrogen atoms were ultimately refined with anisotropic displacement parameters (ADPs), causing a 3-6% decrease in the R_{free} values. This expands the description of the vibrational modes of an atom from isotropic (one displacement term) to a tensor matrix (multiple displacement terms), where the 6 ADP components describe a probability distribution for the electron density of an atom in three dimensions (Merritt 1999a). The resultant atomic model is more a sphere (for isotropic behavior) or an ellipsoid (for anisotropic behavior). In SHELXH, standard restraint and deviation values for SIMU, DELU (rigid body), and ISOR were employed. "Riding" hydrogens were generated on most residues using HFIX commands and refined with isotropic temperature factors, resulting in a 2-3% decrease in the R values. In SHELXH, a final weighting scheme of 0.2 was employed. A penultimate refinement in

SHELXH (with recommended weights) was performed until the R_{fac}, R_{free}, and Goodness-of-Fit values were stable for 20 cycles and the final maximum atom shift was < 0.02Å. A final refinement was performed within Refmac5, this time increasing the value for the sigma weighting matrix from 0.5 to 5-7 for all structures to allow molecular "flexibility" within the minimization procedure while maintaining ideal geometry. Essentially, reliance is placed more on the observed data for restraints rather than on the small molecule libraries (Engh *et al.* 1991) invoked in Refmac5. All final refined bond distance and angle deviations from the ideal are < 0.025Å and < 2.2°, respectively. Additionally, coordinates with hydrogen atoms were included in the final Refmac5 refinement. The identities of the metal ions were suggested by the composition of the crystallization buffer, their coordination geometries within the model, and their corresponding electron density.

For the MTX and the apo X-ray structures, when the refinement of the models had nearly converged (the agreement between the model and the data was seemingly as close as reasonably possible), the SHELXH program was used for refinement of the aspartate residues free of stereochemical restraints (by removing only the restraints associated with these residues) to allow unrestrained movement of the carboxylate oxygen positions. Coincident with this step, the structure was refined against all reflections (no R_{free} set). At this point, the measurements of the carboxylate Cγ-Oδ1 and Cγ-Oδ2 bond lengths were refined, quantified and very accurate. To determine the estimated standard deviations (ESDs) of *all* bond lengths, the MTX and apo models were then used in one cycle of least squares calculations for a full-matrix unrestrained refinement, where all the off-diagonal terms in the parameter matrix are included in the

calculations. All restraints are removed for this step. To lessen computational time and memory requirements, only the first 30 residues were included in the ESDs calculation. On a Linux machine with 4 processors, it took 10 minutes for this 30 residue calculation. $2\boldsymbol{F}_o - \boldsymbol{F}_c$ and $\boldsymbol{F}_o - \boldsymbol{F}_c$ difference electron density maps were calculated using FFT within CCP4 (CCP4 1994), and electron density was displayed for map inspection and model building in the program O (Jones *et al.* 1991). Sigma-A weighted $2\boldsymbol{F}_o - \boldsymbol{F}_c$ nuclear density maps were calculated using the mapmaking option ("M") within the SHELXpro interface. Maps were converted to CCP4-style maps using MAPMAN (Kleywegt *et al.* 1996). Alternatively, maps were also generated "on the fly" from the output .fcf file (which contains structure factors and phases) in the model building programs MI-Fit and Coot (Emsley *et al.* 2004). Model and map figures were made using PyMol and Coot.

2.5 Growth of larger MTX binary crystals for neutron crystallography

To generate large volume DHFR crystals for neutron diffraction, one needs to maximize crystal growth with minimal nucleation points within the drop while using highly concentrated protein. To achieve this, DHFR was ultimately concentrated to >50 mg ml^{-1} and large drops at >50 μl total volume were set-up in a modified sitting drop format using optimized precipitant and salt conditions. Moderate to large-sized crystals (a few grew to 1.4mm x 1.0mm x 0.3mm) were grown at 4°C by mixing equal volumes (25 μl) of the protein complex with the mother liquor, 0.1 M Na-HEPES (pH 7.5), 0.2 M $CaCl_2$, and 18% PEG 400 (v/v) (optimized from a condition in Hampton Crystal Screen #1 (Hampton Research) (Jancarik *et al.* 1991), on siliconized cover slips and placed on a Plexiglas support. Pyrex (Corning, NY, USA) custard dishes were used as the reservoirs (40 ml total volume of mother liquor) with the support sitting in the dish and the whole

apparatus sealed by a thick circle of Plexiglas, with vacuum grease (Dow Corning, MI, USA) used to finish the seal. Crystals appeared in 1 day and grew to full-size in about two weeks. The traditional crystal volume needed for neutron experiments is ~1mm^3 if the protein is not perdeuterated; theoretically, if the protein is perdeuterated then the volume can be reduced by 1/10th without a significant loss of signal. As reported in (Bennett *et al.*, 2005), a D_2O-soaked crystal of dimensions 1.4 x 1.0 x 0.2mm was used to collect a neutron data set at the ILL that is 56% complete to 2.2Å. The volumes of these crystals have been able to be significantly increased for at least 3 reasons: 1) further optimization of the mother liquor for DHFR/MTX crystallization (0.1M Na-HEPES (pH 7.5), 0.2M $CaCl_2$, and 16% PEG400); 2) a new microbatch-under-paraffin oil crystallization method using highly pure, concentrated protein complex (>50 mg ml^{-1}) has been employed; and 3), ultracentrifugation of the sample immediately prior to crystallization to ensure a minimal number of nucleation points. In the microbatch under oil method, 100 µl of DHFR/MTX complex at >50 mg ml^{-1} is mixed with the same volume of crystallization buffer in the bottom of a well of a Nextal 24-well crystallization plate (Nextal Biotechnologies). Immediately, a 4x drop volume (so, in this case, 800 µl) of paraffin oil was pipetted onto the top of the batch mixture. Screw-cap tops were used only after the crystals had grown to their largest possible size and the D_2O-soaking procedure began.

2.6 D_2O-soaking and harvesting of crystals for neutron diffraction experiments

Several DHFR/MTX crystals were subjected to H/D exchange prior to neutron data collection in order to reduce the large hydrogen incoherent scattering contribution to the background. To prevent "shocking" the crystals they were H/D exchanged

conservatively, against an increasing gradient of D_2O-based crystallization buffer (0.1 M Na-HEPES (pH 7.5) from a 1 M buffer stock made with D_2O, 0.2 M solid $CaCl_2$, and 18% PEG 400 (v/v), all components dissolved into D_2O) over the course of one week (i.e. from an initial ratio of 10% D_2O/90% H_2O, the D_2O% concentration was doubled every other day until 90-100% D_2O content was achieved). After one month, the crystals were mounted into custom-prepared quartz capillaries with a 2.9mm outer diameter (Vitrocom, Inc., Mountain Lakes, NJ) with a D_2O "plug" at one end, sealed at both ends with epoxy and paraffin wax, and stored securely in a sealed 50ml Falcon tube at 4°C for transport and for storage until data collection. The crystal mounting was performed in a D_2O-saturated environment: a tent was erected in a 4°C room with N_2-purged D_2O pumped into the tent to prevent back-exchange of crystal and buffer D atoms with H in H_2O vapor in the environment. The dimensions of the crystal used for preliminary neutron data collection at the ILL were 1.4 x 1.0 x 0.3 mm. Due to the hexagonal external morphology of these crystals (i.e. they are not a perfect cubic shape), the estimated volume of this crystal is ~0.3mm^3. For our most recent experiments at the ILL and at LANSCE, crystals grown from the microbatch under oil method have been used. Their dimensions are 1.7 x 1.2 x 0.4 mm and 1.8 x 1.0 x 0.4 mm, respectively. The volumes of these crystals, adjusted for the hexagonal morphology, are both ~0.25 mm^3.

2.7 Preliminary neutron diffraction studies: Neutron data collection at the ILL (Bennett *et al.*, 2005)

The D_2O-soaked DHFR/MTX crystal was tested for neutron diffraction on the quasi-Laue LADI instrument (λ= 3.5Å, $d\lambda/\lambda$ ~ 25%) at the ILL (Grenoble, France). This diffractometer uses a cylindrical neutron image-plate detector which completely

surrounds the sample (Cipriani *et al.* 1996; Myles *et al.* 1998). LADI has been utilized for a number of successful protein neutron diffraction experiments (Coates *et al.* 2001; Cooper *et al.* 2000; Habash *et al.* 2000; Habash *et al.* 1997; Niimura *et al.* 1997; Tuominen *et al.* 2004). The DHFR/MTX crystal was exposed for 34 hours per frame, and a total of 21 frames were collected at two different crystal orientations. The φ separation between frames was typically 8°. All data were collected at 20°C. The data were indexed and integrated using the program *LAUEGEN* (Campbell *et al.* 1998) which has been modified to account for the cylindrical geometry of the detector. The program *LSCALE* (Arzt *et al.* 1999) was used to derive the wavelength-normalization curve using the intensity of symmetry-equivalent reflections measured at different wavelengths. The data were then scaled using the *SCALA* program within the CCP4 suite (CCP4 1994).

2.8 Neutron data collection and processing at the Protein Crystallography Station at LANSCE

One crystal, mounted in one of the large-diameter quartz capillaries and with D_2O-based buffer plugs, was used for a time-resolved neutron Laue experiment on the Protein Crystallography Station (PCS) at LANSCE (Langan *et al.* 2004a) using previously described techniques (Hanson *et al.* 2004) (Li *et al.* 2004) (Sukumar *et al.* 2005). Fifteen total Laue images were collected. The capillary was mounted vertically and shifted 15° around the φ-axis at φ settings of 0, 15, 30, 45, and 60°, for a total of five frames about the φ-axis at κ=0. The crystal was then returned to its original position on φ and rotated to +30° on the κ-axis and 5 additional frames at the φ settings indicated above (0, 15, 30, 45, and 60°) were collected, again with a 15° rotation on the φ-axis between each image. This sequence was repeated using a κ-axis setting of 60°. The average

exposure time per setting was 36 hours. This is longer than the average exposure for a crystal of this volume and that has been D_2O-soaked but was necessary due to reduced beam flux (70-80 μAmps) for the initial part of the data collection. Normal flux (110-120 μAmps) was attained for the later exposures. Over 41,000 reflections were recorded (14,213 unique) for a resolution range of 38.2-2.20Å; the completeness of the data set is 79.7%.

Data over a wavelength range of 0.6Å to 7Å were processed using a version of *d*TREK* (Pflugrath 1997) modified for time-resolved neutron Laue protein crystallography (Langan *et al.* 2004b). Neutron intensity peaks were identified and a list of three coordinates was generated for each peak, two spatial (X, Y) and one based on TOF. Intensities were indexed and integrated within the modified *d*TREK* program, and integrated Laue reflections were normalized by wavelength and scaled using *LAUENORM* (Helliwell *et al.* 1989). In order to obtain reasonable values for R_{merge} (from *LAUENORM*, the R_{merge} was 0.216 for all measurements of a reflection, with 23,416 reflections; R_{merge} is 0.196 for measurements of a reflection of the same sign, with 17,788 reflections; R_{merge} is 0.174 for all measurements of a reflection of the same sign and within $\lambda = 0.1$ Å, with 4,533 reflections), the wavelength range was narrowed to 0.8Å-3.5 Å and only reflections with $I>2\sigma$ were used in determining the wavelength scaling normalization curve. Reflections in this wavelength range were binned into 15 wavelength intervals, and the normalization curve was determined from a Chebyshev polynomial of order 5 (Artz et al., 1999). The data were output in unmerged form so that *SCALA* (Evans 1997) and *TRUNCATE* (CCP4 1994) could be used for statistical analysis.

After merging with *SCALA* the overall value of R_{merge} was 0.136 and the completeness was 70%.

2.9 Neutron structure refinement

Solvent molecules and metal ions were removed from the coordinate file of the RT X-ray model of DHFR/MTX (again, 3DRC; solved to 1.9Å). This model was used as the input coordinates for refinement in SHELX against the neutron data set merged in SCALA. The instruction file for SHELX refinement was modified exhaustively to include the proper neutron scattering factors, restraints for the incorporation of deuteriums into the model and solvent and rigid body restraints within all residues throughout the model. An example of this is a restraint used to maintain all the atoms within one peptide bond as one rigid group. This effectively reduces the number of parameters which must be refined and thus increases the data-to-parameter ratio, an important factor when the number of reflections is limiting as in this neutron experiment (10,472 unique reflections used to refine about 6600 parameters). Many other restraints like this are included in the residue list so as to break up the structure into many rigid bodies. The resolution range chosen for the initial rounds of alternating restrained positional, occupancy, and B-factor refinement was 8.0 – 2.5Å. Occupancies of deuterium atoms (Ds) bound to backbone amide nitrogens were refined next. Occupancies of Ds bound to His imidazole rings, Arg guanidino groups, and Lys amine groups were then refined. The modeling of D on backbone amides or the side chains was based on the resulting values for their occupancies from the refinement, and the presence of $\geq 1.5\sigma$ nuclear density in the maps. An anisotropic scaling factor (the HOPE parameter in SHELX) was included in the refinement, more restraints were introduced for the D's

bound to the His, Arg, and Lys residues, and the lower limit of the refinement range was changed to 6.5Å. Sixty-six oxygens (as waters which are H_2O or have disordered D's) and 18 D_2Os were included in the refinement; the high resolution range was extended to 2.2Å. At this point, the R_{free} was at 36.7%, and progress in refinement halted, no matter the manipulation or the program that was used the R_{free} either improved only slightly or remained the same. A suggestion was made from a collaborator that maybe the crystal was twinned. Three facts would strengthen this argument: 1) the refinement had stalled; 2) larger crystals have an increased chance at malformations in the lattice and "domain" rotations of regions within the crystal that are not symmetrically-related; and, 3) crystals of the trigonal/hexagonal Bravais lattice system are known to suffer from merohedral twinning (Yeates 1997). So, a merohedral twinning correction (the twin law) was incorporated into the SHELX instruction file (TWIN 0 1 0 1 0 0 0 0 -1), allowing the contributions from the individual twinning components to be refined against the neutron data. Thus, SHELX can correct for data from twinned crystals in the refinement by fitting the sum of calculated intensities from the separate individual twin domain components to the observed intensities (Drenth 1999). Doing this led to a large decrease (> 10%) in the R_{free}.

2.10 Preparation, crystallization and X-ray data collection and processing of perdeuterated DHFR/MTX crystals

In order to obtain the highest level of deuterium exchange in the DHFR crystals, the growth of SK383 *E .coli* and the subsequent purification of DHFR from a completely deuterated media source were pursued. Based on low yields acquired from previous attempts to express DHFR from minimal media, a commercially available rich media

source (Celtone-D, D level >97%, Spectra Stable Isotopes, Columbia, MD) was chosen for perdeuterated cell growth (Tuominen *et al.* 2004) instead of existing alternative labeling methods (Leiting *et al.* 1998). The SK383 strain of *E. coli* requires an extensive period of adaptation before they will robustly grow in deuterated media. From work on the perdeuteration of proteins other than DHFR, it is known that BL21 (DE3) strains (Invitrogen, Carlsbad, CA) transformed with JM vectors and BL21 (DE3) pLysS strains (Novagen, San Diego) transformed with pGEX vectors (GE Healthcare, Piscataway, NJ) do not require extensive adaptation; hence, in only a few hours growing in 100% Celtone-D, these cells attain a relatively high optical density, as observed when measured at 600nm (OD_{600}nm). Conversely, it was necessary to adapt the SK383 strain to growth in deuterium over the course of 2-3 weeks by gradually increasing the ratio of D to H in the growth medium. Complete adaptation of the *E. coli* strain is assumed when cells growing in 90-100% D media grow at nearly the same rate as those growing in unlabeled rich media; this checked by OD_{600}nm measurements. To boost growth rates, 2-5% inoculum ratios were used and the media was supplemented with 0.4% D_8-glycerol (D level >98%, Spectra). An additional antibiotic, trimethoprim (TMP, Sigma-Aldrich) at 20 μg/ml, was added to the growth medium to elevate DHFR expression. D -glycerol stocks were made and stored at -80°C. The scale of cell growth could then be increase. Also, after centrifugation to harvest the cells for the first time, the media may be refiltered, inoculated, and used again for a second round of cell growth. So, from 1l of deuterated rich media, ~3 mg pure perdeuterated DHFR can be produced, totaling ~6 mg after recycling.

All details of cell lysis, protein purification, and concentration are the same as above except a pre-packed DEAE2 weak anion exchange column (Bio-Rad) was used for the final chromatography step instead of DEAE-Sephacel. Immediately after purification, the percent enrichment of D into the DHFR protein was measured by mass spectrometry. 1 mg ml^{-1} samples of native and perdeuterated DHFR were dialyzed separately overnight at 4°C against Millipore-Q H_2O, and the dialyzed solutions were desalted using a C_{18} ZipTip (Millipore). The pure, desalted DHFR samples were then subjected to intact mass analysis using Fourier Transform-Ion Cyclotron Resonance (FT-ICR) on a Voyager Workstation equipped on the front-end with an electrospray ionization (ESI) source to impart the samples with a charge. The sample was injected into the ESI source, and multiple scans were recorded for each sample to provide the associated deconvoluted m/z spectra. The m/z peaks were transformed to mass peaks, and the subsequent spectra was analyzed using Excalibur software and programs available on the PROWL website (www.prowl.rockefeller.edu).

After verification of significant incorporation of deuterium into the protein, a 1 mg ml^{-1} sample of perdeuterated DHFR was incubated with a 5x molar excess of MTX, was concentrated with a Centricon YM-10 (Amicon) device, and was then used for hanging drop vapor diffusion crystallization trials. All other aspects of the crystallization are the same conditions as for the native complex; however, reproducibility and consistency have been more difficult with the perdeuterated complex. For the largest crystals (~0.1 mm^3 volume), the protein concentration had to be >25 mg/ml and the precipitant concentration had to be decreased significantly, from 18 to 12-14% PEG 400 (v/v). Upon reaching their largest size, the crystals were soaked in the same buffer, only

D_2O-based. After one month of D_2O soaking, four perdeuterated DHFR/MTX crystals were mounted into quartz capillaries (Vitrocom, Inc.) as described above for the native DHFR/MTX crystals. Multiple perdeuterated crystals, too small for neutron diffraction, have been used for in-house X-ray diffraction. They are isomorphous with native DHFR/MTX crystals as they crystallize in space group $P6_1$ and have unit cell dimensions of a=b= 93.3Å and c= 74.4Å ($\alpha=\beta= 90°$, $\gamma= 120°$). The resolution of measured reflections extends to 1.5Å, the working limit of resolution on our R-AXIS IV++ detector (Rigaku/MSC).

Chapter 3. Preparation of samples for neutron crystallography and protein perdeuteration

3.1 Background

This chapter will summarize the requisite actions that have been taken in order to prepare samples for neutron experiments, including efforts to increase crystal volume and to effectively exchange H for D, both by D_2O-buffer soaking and by perdeuteration. The production of perdeuterated DHFR and its subsequent analysis by mass spectrometry and X-ray diffraction are also presented. Although this chapter is inherently technical, most of the methodological details are described in chapter 2 (Methods). However, the reiteration of some of these details is included here where necessary for clarification. Part of the results in this chapter have been published as a research article in *Acta Crystallographica* D in May 2005 (Bennett *et al.* 2005). The neutron crystallography results from the Institut Laue-Langevin (ILL) presented in this paper are provided in the next chapter.

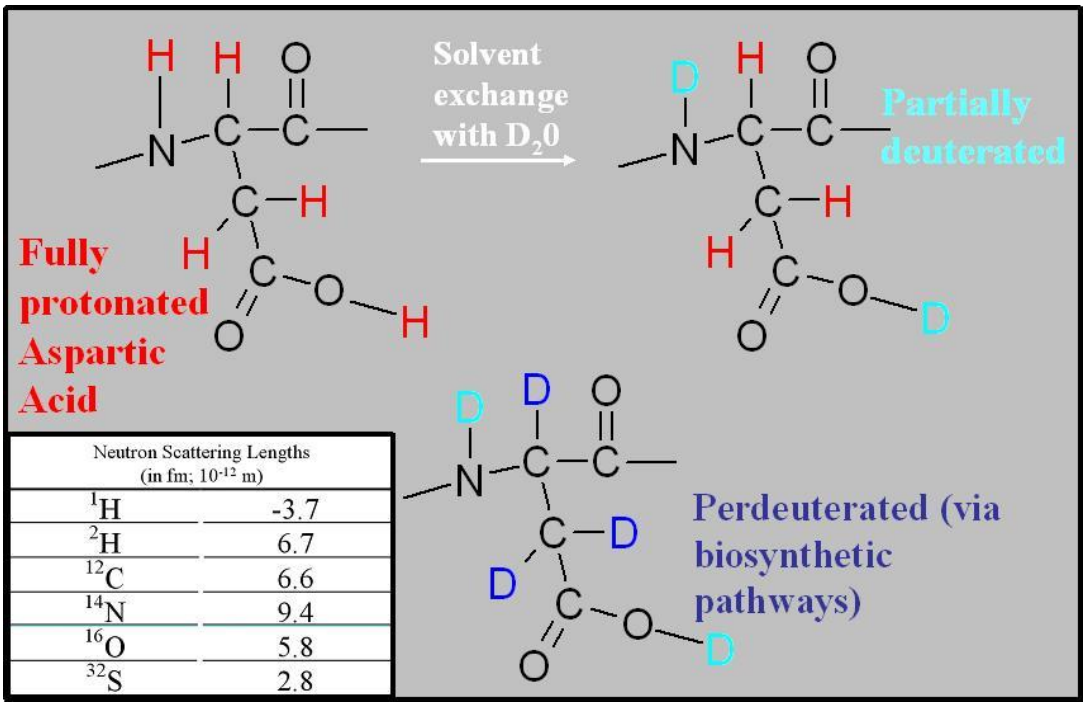
One of the major goals of the thesis project is to directly demonstrate protonation states in the active site of *E. coli* DHFR, and the proposal is to use X-ray and neutron crystallography to accomplish this goal. Many of the results presented in this thesis are on the binary complex of DHFR with MTX. Previous evidence from X-ray crystallography, NMR, UV/VIS difference, and fluorescence spectra suggest that the N1 atom of MTX is protonated and, thus, positively charged when bound to DHFR (Appleman *et al.* 1988; Bolin *et al.* 1982; Cocco *et al.* 1981; Cocco *et al.* 1983; Hood *et al.* 1978; London *et al.* 1986; Matthews *et al.* 1977; Ozaki *et al.* 1981; Poe *et al.* 1972). The binding of MTX in the DHFR active site is such that the N1 atom is within 2.6-2.7Å

of the Oδ2 of Asp27, stemming from an orientation of the pteridine ring which is flipped as compared to folate and DHF binding, a consequence of a 180° rotation about the C6-C9 bond. The close proximity of N1 of MTX when bound to DHFR to the Asp27 carboxylate group strongly suggests an electrostatic interaction, possibly a salt bridge. (Please see text and figures in chapter 4 for additional details on the DHFR/MTX complex and interaction.) In contrast, a recent computational study proposes that, when bound to MTX, the Asp27 is protonated while the N1 on MTX is not (and, thus, the MTX is neutral); therefore, the interaction is suggested to be neutral dipole-dipole rather than ionic in nature (Cannon *et al.* 1997). Essentially, the present controversy over DHFR's ligand binding and catalytic mechanism can be reduced to not knowing the active site and ligand protonation states. Ultra-high resolution X-ray crystallography provides highly precise macromolecular structures with atomic detail (Dauter *et al.* 1997b). However, having only one electron, hydrogen atoms scatter X-rays only weakly, causing ambiguities to arise upon inspection of electron density maps for hydrogen peaks.

Neutron scattering and deuterium labeling

Neutron crystallography can resolve this ambiguity by revealing the positions of hydrogen atoms –especially the deuterium isotope– at moderate resolution (at 2.5-2.0Å and beyond). Neutron scattering phenomena are the result of the strong-force nuclear interaction between the neutron particle beam and the atomic nucleus, the nature of the interaction and the scattering behavior being dependent on the nuclear composition (Shu *et al.* 2000). Due to the inherent weak fluxes at neutron sources, crystal volumes required for neutron data collection are prohibitively too large for most systems (Schoenborn *et al.* 1996). Another disadvantage to using neutron crystallography is that hydrogen nuclei

possess only a single proton and scatter neutrons incoherently and with negative scattering lengths (-3.75 femtometers (fm); (Tuominen *et al.* 2004)). Hydrogen atoms appear as negative peaks in neutron density maps and often "cancel" positive neutron scattering from the atoms to which they are covalently bound like carbon or nitrogen (Meilleur *et al.* 2004). The signal to noise ratio of the data can be improved dramatically by exchanging deuterium (D; M = 2.0126 Da) for hydrogen (H; M = 1.0063 Da) (called H/D exchange) in the sample, either by **(a)** growing or soaking the crystal in D_2O-based buffer (deuteration at chemically exchangeable positions) or **(b)** forcing the expression organism (i.e. *E. coli*) to incorporate deuterated amino acids into the target protein at the biosynthetic level (deuteration at chemically non-exchangeable positions or perdeuteration, **Figure 3.1**). H/D exchange in samples significantly reduces the background of neutron data collections, due to deuterium's positive and coherent scattering behavior (leading to its contribution to signal well above the noise level recorded on the detector) (Schoenborn *et al.* 1996; Tuominen *et al.* 2004). In fact, deuterium scatters neutrons with a similar strength as that of carbon and scatters more strongly than sulfur, phosphorus, oxygen, and nitrogen (Schoenborn *et al.* 1996). Upon soaking protein crystals in D_2O, H/D equilibrium exchange may occur for backbone amide, side chain carboxylate, side chain hydroxyl, and side chain amine hydrogens. Generally, the equilibrium exchange rates of the H/D atoms bound to the backbone amides are markedly slower than those rates of H/D atoms bound to side chain amine, carboxylate, and hydroxyl groups.



Neutron Scattering Lengths (in fm; 10^{-12} m)	
^{1}H	-3.7
^{2}H	6.7
^{12}C	6.6
^{14}N	9.4
^{16}O	5.8
^{32}S	2.8

Figure 3.1: The incorporation of deuterium (D) into proteins is a prerequisite for neutron experiments. Exchange of H for D at labile positions (such as backbone amide or side chain carboxylic acid groups, as shown here with **D**) can be accomplished either by incubating the sample in a D_2O-based buffer or D_2O itself. The exchange of H for D at chemically non-exchangeable positions (shown here as **D**) must occur at the level of amino acid biosynthesis in the expression organism, a much more complicated form of deuterium labeling called perdeuteration.
Inset table: The requirement for H/D exchange is based on H's negative and incoherent neutron scattering that can actually partially or completely cancel scattering arising from the atoms to which it is bound.

Protein perdeuteration

As stated above, proteins deuterated at chemically non-exchangeable positions (e.g. perdeuterated, where D is substituted for H on aliphatic and aromatic carbons) can only be generated at the level of biosynthesis by the organism used for expression of the target protein (Gamble *et al.* 1994); (Leiting *et al.* 1998); (Tuominen *et al.* 2004)). This increased enrichment of D can significantly lower the remaining background scattering and improve neutron diffraction quality even more. Perdeuteration has also been used to improve the signal from atoms other than D in NMR experiments where significant line broadening occurs for larger proteins or systems, and many of the overlapping chemical shift peaks can be attributed to H; exchanging for D allows one to "filter" the spectra to now only observe the chemical shifts arising from the contributions of "heavier" atoms (Gardner *et al.* 1998); (Mok *et al.* 1999). Conceptually similar to this, neutron scattering can also benefit from the use of perdeuterated molecules in contrast variation and matching experiments. The solvent D percentage can be changed until it matches the content within a native molecule or the perdeuterated component; once one has subtracted the matched components, the *difference* scattering arises from the other component in solution (Schoenborn *et al.* 1996; Trewhella *et al.* 1998). For the purpose of this thesis, the prevailing reasoning for the deuterium labeling of macromolecules is to aid in neutron crystallographic data acquisition and analysis. Indeed, nuclear density along the aliphatic side chain of Lys prior to the terminal ε-amino group and on hydrophobic side chains like Val and Leu are normally "invisible" in neutron maps due to H's negative scattering and its cancellation effects on carbon positive scattering. So, perdeuteration provides a layer of information about the H position (visualized as

deuterons in the neutron maps) on these side chins and, in the process, enhances the overall scattering of the crystal (Shu *et al.* 2000). However, perdeuteration can be quite expensive both in cost and in time requirements (**Table 3.1**) due to the inherent expense of deuterium itself, the possibility of lengthy deuterium adaptation periods for the expression organism, and the normally reduced expression levels of the target protein isolated from the deuterated organism relative to a native counterpart (Paliy *et al.* 2003); (Meilleur *et al.* 2005). Deuterium adaptation has been accomplished for several strains of *E. coli* for the heterologous expression of deuterated target proteins. It is simply the process of acclimating the strain to increasing percentages of deuterium over multiple platings (if done with deuterated agar medium) and/or over multiple inoculation cycles (if done in deuterated liquid medium), allowing each successive generation of cells to adapt and grow to high optical densities before the next plating onto or inoculation into a slightly higher deuterium percentage medium (Rokop *et al.* 1969); (Paliy *et al.* 2003).

Biophysical effects of perdeuteration

About ½ of the atoms in a protein are hydrogen, but exchange of hydrogen for deuterium at all or nearly all chemically non-exchangeable positions normally increases the mass of the protein by only 5-10%. This can be measured quite accurately with mass spectrometry, in a number of different modes (e.g. ESI-FT-ICR, MALDI-TOF). In addition to the change in mass, there exists at least mild concern that nearly uniform exchange of hydrogen for deuterium in proteins may significantly alter their overall three-dimensional and local secondary structure, their ability to perform their physiological function, and their inherent stability and flexibility. Pioneers to explore the latter two of these concerns were Crespi and Katz, who in the 1960's extensively studied

Table 3.1: Potential costs for perdeuteration of a protein

Media type	**Minimal (M9)**	**Rich (Celtone-D)**
D_2O	$350	N/A
D_7-glucose/D_8-glycerol[1]	$1100/$375 for 0.4% final	$550/$190 for 0.2% final[2]
D-algal hydrolysate powder/resuspension	N/A	$600/$1200
Total per liter media	$725-$1450	$600-$1750

[1]Many different deuterated carbon sources, more than can be listed here. However, examples of others commonly used are succinate and xylose. Glycerol is probably the best choice due to its relatively low cost, and the fact that it can be transformed into a glycolytic intermediate upon incorporation into the bacterial cell.

[2] The use of deuterated carbon sources in a rich media culture is strictly optional and is mainly used to boost growth rates during the adaptation process, not necessarily during the stage of large scale protein expression.

perdeuterated phycocyanin isolated from several different cyanobacterial and algal species. They found in a series of experiments that perdeuteration does lower both phycocyanin's thermal stability (as revealed by differential scanning calorimetry (DSC); (Hattori *et al.* 1965)) and its propensity to associate into oligomers (as revealed by sedimentation equilibrium; (Hattori *et al.* 1965)). Perdeuteration also seems to increase the catalytic rate of GST (Brockwell *et al.* 2001) compared to unlabeled enzyme while the expected isotope effect from replacing H with D (~ 1.4x decrease in rate; (Fersht 1998)) was reported earlier for perdeuterated alkaline phosphatase (Rokop *et al.* 1969). Little effect on K_m was reported for both of these studies. Cumulatively, the only marked difference between native and perdeuterated proteins seems to be in thermal stability and susceptibility to proteolytic degradation. Circular dichroism, DSC, and Fourier Transform-Infrared (FT-IR) spectroscopy analyses have revealed that the perdeuterated proteins used in these studies have slightly lower T_ms for unfolding and, thus, are less thermally stable than native counterparts (Brockwell *et al.* 2001); (Meilleur *et al.* 2004). Brockwell *et al.* also found that the rate of proteolytic cleavage was significantly enhanced for perdeuterated GST compared to the native form (Brockwell *et al.* 2001). Although the sampling size is quite small, X-ray structural studies of perdeuterated proteins, including SNase A (Gamble *et al.* 1994), myoglobin (Shu *et al.* 2000), inorganic pyrophosphatase (Tuominen *et al.* 2004), the elongation factor Tu (Cooper *et al.* 1998), and P450cam (Meilleur *et al.* 2005) reveal that perdeuteration seems to have little if any effect on global conformation and secondary structural composition when compared to the structure of an unlabeled reference. The r.m.s.d. values between perdeuterated and native Cα chains are always much less than 1Å, the traces being nearly identical (Cooper

et al. 1998). Stress should be placed on the fact that very little biophysical data has been accumulated on perdeuterated proteins, however, the evidence so far suggests that perdeuteration does not grossly alter the inherent structure and biochemical function of the protein under study. On the other hand, perdeuteration does seem to reduce the target protein's stability in solution, especially when effects from heat, solvent hydrophobicity, and proteolytic degradation are taken into account. Although the observations are limited in scope, perdeuterated DHFR seems to be very similar to the native form, with slight changes noted for solubility and exact conditions for cocrystallization with MTX.

3.2 Results and discussion

Applying a "hybrid" crystallization method for larger volume DHFR/MTX crystals

In order to identify hydrogen positions within the DHFR active site and on bound methotrexate, D_2O-soaked DHFR/MTX crystals suitable for neutron diffraction analysis were grown and a partially complete data set was collected on the Laue Diffractometer Instrument (LADI) at the ILL. The D_2O-soaked DHFR/MTX crystal diffracted neutrons to 2.2Å resolution, however, it is only 56% complete to this resolution. Therefore, to increase completeness, extend resolution, and maximize signal-to-noise, two strategies have been pursued: larger crystal growth with extended D_2O-soaking and perdeuteration, to meet these goals. A third is listed at the end of this chapter but has not been utilized.

First, larger volume DHFR/MTX crystals (~25-30% increase) have been able to be grown by an alternative crystallization method called microbatch under oil. This method is based on crystallization techniques described in (Chayen 1997) and (Rayment 2002) and relies on the ability to rapidly equilibrate the protein solution with the crystallization buffer while strictly limiting evaporation from the protein/buffer mixture

and also restricting the mixture's equilibration with the atmosphere. Large volumes (50-100 μl) of highly concentrated (40-60 mg ml^{-1}) DHFR/MTX complex were mixed with an equal amount of crystallization solution (0.1M Na-HEPES (pH 7.5), 0.2M $CaCl_2$, and 16-18% PEG 400) in the bottom of a well in a crystallization plate. An overlay of oil was applied to the batch protein/buffer mixture. For the least amount of crystal nuclei and the growth of the largest crystals of DHFR/MTX by this method, the precipitant (PEG 400) concentration can vary between 14-18% but paraffin oil is definitely the best diffusion control oil to use. After two weeks to allow for complete growth of the crystals, several of these were selected for D_2O-soaking and, once the buffer exchange began, a screw-seal cap was placed on top of the crystallization well to fully prevent diffusion of ^1H water vapor from the atmosphere into the well and the crystals. The crystal size for the preliminary ILL study in 2004 is 1.4 x 1.0 x 0.3 mm, or about 0.45mm^3, and this was produced in the modified sitting drop crystallization. Alternatively, the microbatch under oil technique produced at least two crystals used for the studies at ILL and LANSCE this year. Their sizes were 1.7 x 1.2 x 0.4 mm (0.6$mm^{3;}$ used at the ILL) and 1.8 x 1.0 x 0.4 mm (0.5$mm^{3;}$ used at LANSCE). When applying the correction in the volume calculations for the fact that these are hexagonal and not cubic crystals, the true volumes are about 30-40% less, or about 0.2-0.3mm^3. This is 1/3 to ¼ the volume that has routinely been required for neutron diffraction experiments.

Perdeuteration of E. coli DHFR

Second, to increase the signal-to-noise ratio in the neutron experiments, fully deuterated DHFR protein has been prepared and the growth of large volume perdeuterated DHFR/MTX crystals has been optimized for future neutron diffraction

studies. The improved signal to noise ratio of neutron data collected from perdeuterated protein crystals (Shu *et al.*, 2000) will allow us to collect more complete and higher resolution data, even from comparably sized crystals. The SK383 *E. coli* strain has been adapted to growth in >97% D background using both a commercially available rich media (Celtone-D) and a formulation devised and produced in-house at LANSCE (called "Altone") (**Figure 3.2**). Adaptation required about two weeks and consisted of incrementally increasing the D% in the growth medium if the cells had adjusted properly to the previous increase in D%. This was judged by comparing their rates of growth to cells which were growing in unlabeled rich media. To expedite the rate of adaptation, boost growth rates, and ensure that the expression plasmid was retained by the cells in the presence of an environmental stress, a two antibiotic system (ampicillin and trimethoprim) and D_8-glycerol (>99% D; Spectra Stable Isotopes) were exploited. Two similar protocols using deuterated succinate in minimal medium and carrying out adaptation over several cycles are described in (Leiting *et al.* 1998) and (Paliy *et al.* 2003). Very little study has been performed to understand exactly how bacteria adapt to a deuterated background and why adaptation even seems to work for most strains. This author is aware of several strains other than SK383 which have been utilized for recombinant protein expression in deuterated media, including BL21(DE3), MRE600D, and JM109 to list a few. A possible explanation for the success of adaptation is that *E. coli* can inherently grow in deuterated media, although slowly, and that, over the course of many generations, can accumulate non-lethal mutations which allow the cell to function and replicate in deuterated media more successfully (Gamble *et al.* 1994); (Paliy *et al.* 2003). The microbial and algal life cycle and their physiological responses to a

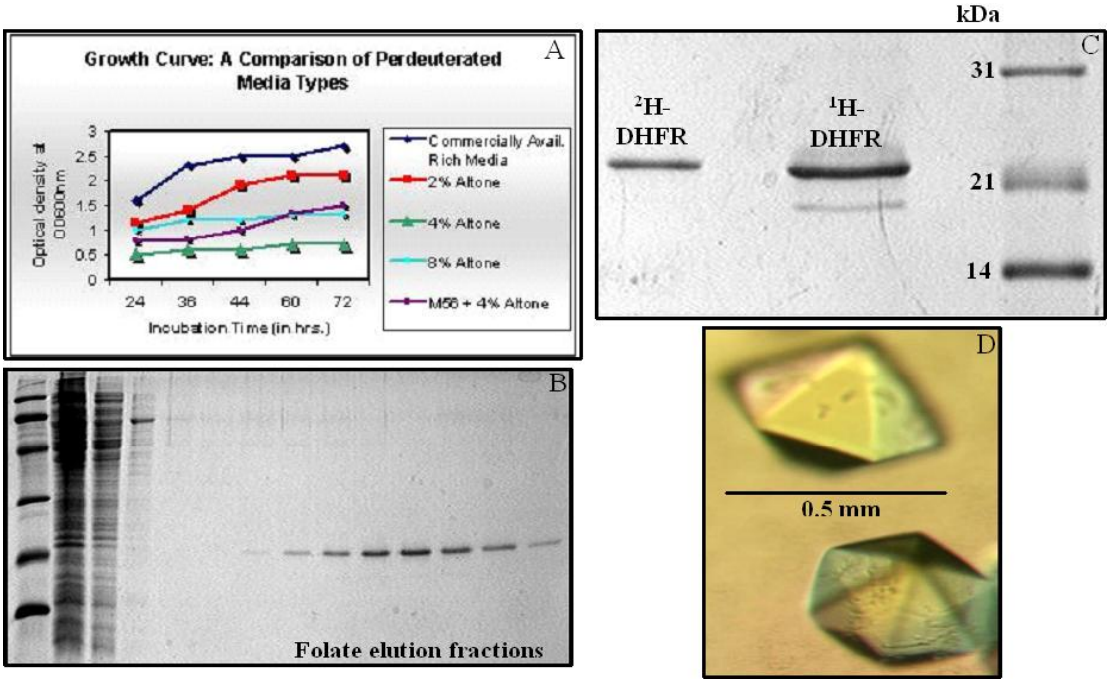


Figure 3.2: Expression, purification, and crystallization of perdeuterated *E. coli* DHFR. A: Mainly, a commercially available deuterated rich media (Celtone-D) has been used to culture SK383 *E. coli* cells for the production of perdeuterated DHFR. However, an alternative formulation from deuterated algal hydrolysate prepared at LANSCE ("Altone") has also been used successfully. **B** and **C**: The purification scheme is identical to the one used for the isolation of native *E. coli* DHFR. The yield is 6 mg l^{-1} culture, about 40-50% less than the native yield. **D**: Crystals of perdeuterated DHFR bound to MTX.

deuterated environment was actively explored in the 1960's with Crespi and Katz's work (Rokop *et al.* 1969) among others but, in my knowledge, is only now starting to draw more interest. This may be due to the expansion of NMR for studies on larger biological systems, and the reemergence of neutron scattering and crystallography for use in the life sciences.

Identical methods for the expression and purification of native DHFR were used for perdeuterated DHFR, save the use of trimethoprim and the addition of the deuterated glycerol to the growth media. A 1 mg ml^{-1} sample of pure perdeuterated DHFR was dialyzed against Millipore-Q H_2O and then desalted using a C18 Zip Tip (Millipore) for mass spectrometry. The percent enrichment of D into DHFR was checked by intact mass analysis using an Electrospray Ionization Fourier Transform- Ion Cyclotron Resonance (ESI-FTICR) mass spectrometer with Dr. Tomoaki Uchiki in the laboratory of Dr. Robert Hettich at ORNL (**Figure 3.3**). After deconvolution of the m/z data to derive direct mass values, a major peak was repeatedly observed in the deconvoluted spectra at 19,239.000 Da. A comparison to the average intact mass of native DHFR (17,999.220 Da; **Figure 3.4**) revealed that the perdeuterated sample was enriched with D at essentially every chemically non-exchangeable position; the mass difference was 1233.780 Da, possibly meaning that all 1233 non-exchangeable H positions were now enriched with D and an additional exchangeable D was still bound to the protein. Although not shown, the mass peak area associated with perdeuterated DHFR at full width half maximum spans about 40 Da and is only about 7 Da for native DHFR (**Figure 3.4**). This may reflect the isotopic distribution differences between the backgrounds in which the proteins were expressed as well as the tremendous mass accuracy of the ESI-FTICR mass spectrometer

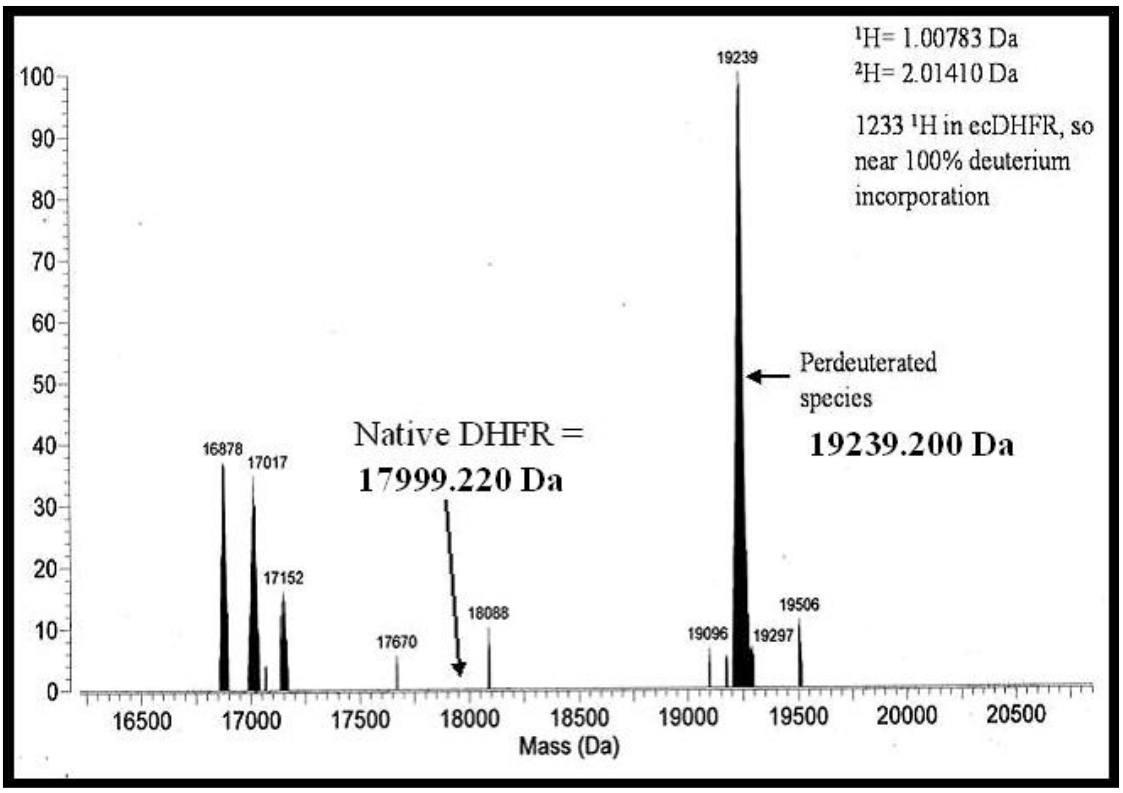


Figure 3.3: Intact mass analysis of purified perdeuterated *E. coli* DHFR using ESI-FTICR mass spectrometry and measurement of deuterium enrichment. The measured mass of the perdeuterated species is 19,239.200 Da, meaning that at essentially 100% of the chemically non-exchangeable positions D has been incorporated instead of H. This mass may also indicate that D persists at about 7 of the chemically labile positions as well. The peak distribution (full width at half maximum) is about 40 Da and, although it's not as small as the native mass distribution (**Figure 3.4**), may indicate the uniformity of D labeling in DHFR, resulting from the exhaustive *E. coli* adaptation process (Gamble *et al.* 1994) (Brockwell *et al.* 2001). The Y-axis for both figures 3 and 4 is percentage relative abundance, a comparative value for the peaks in a given spectra. The spectra in figures 3 and 4 were recorded with Dr. Tomoaki Uchiki in Dr. Robert Hettich's lab at the Organic and Biological Mass Spectrometry Division at Oak Ridge National Laboratory. Subsequent analysis of the mass spectra was performed in the Dealwis laboratory using IonSpec software.

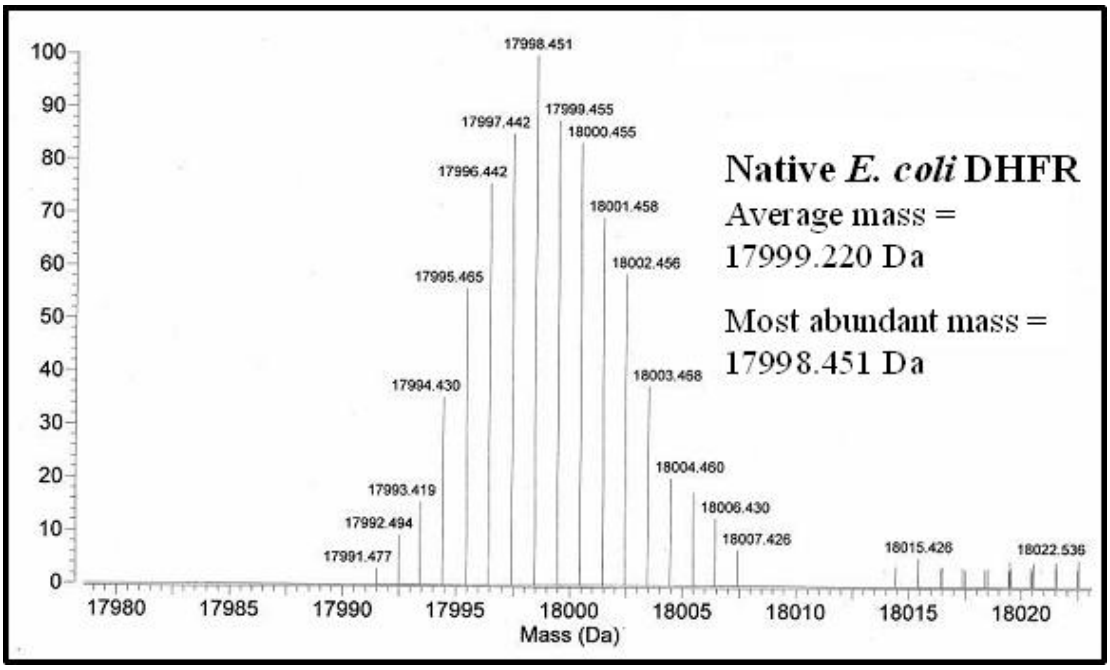


Figure 3.4: Intact mass analysis of purified native *E. coli* DHFR using ESI-FTICR mass spectrometry. The average calculated mass was 17999.220 Da and was used for comparison with the perdeuterated DHFR mass (**Figure 3.3**). Note the mass distribution is about 8 Da, most likely a distribution of natural isotopic species. Again, the Y-axis is percentage relative abundance.

(Brockwell *et al.* 2001); however, the possibility of ^{13}C contribution to the isotopic distribution cannot be overlooked. These analyses verified a high level of deuterium incorporation into DHFR, allowing us confidence to go forward with crystallization for X-ray and neutron diffraction experiments.

Crystallization of the perdeuterated DHFR/MTX complex was accomplished similar to using unlabeled DHFR except that the protein concentration used is much lower (15-25 mg ml^{-1}), and the precipitant concentration used in the crystallization buffer is slightly less (14% instead of 16-18% PEG 400). Multiple crystals were cryoprotected by incubating them in D_2O-based crystallization buffer with an addition of D_8-glycerol; this protects against ice formation in the buffer and helps to insulate the crystal against severe thermal fluctuations (Garman 2003). Some of these crystals were tested on our in-house X-ray source with a liquid nitrogen cold stream, and their diffraction extends to 1.5-1.6Å, the working limit of resolution on our detector. These were then dismounted and stored under liquid nitrogen for transport to the Advanced Photon Source (APS; Argonne, IL). Two synchrotron X-ray data sets were collected at the NE-CAT beamline on a perdeuterated DHFR/MTX crystal. The completeness for the combined data sets is 99% and the resolution extends to 1.25Å (**Figure 3.5**). The perdeuterated unit cell is isomorphous to the native DHFR/MTX crystals, and the space group ($P6_1$) is also conserved (**Table 3.2).** The resulting coordinates from refinement will be an excellent starting model for future neutron refinement of a perdeuterated DHFR/MTX complex. The perdeuterated crystals are first grown in an H_2O-based buffer. After they attain what is determined to be their largest size (about 1-2 weeks), they are then soaked in the same buffer, only D_2O-based. The largest size so far for a perdeuterated DHFR/MTX crystal is

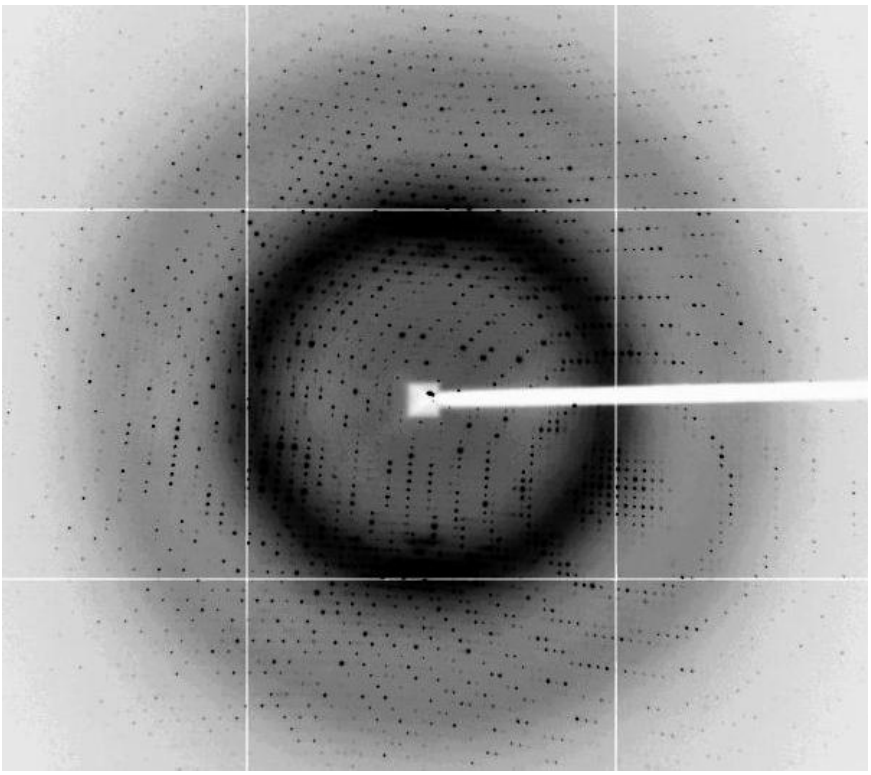


Figure 3.5: X-ray diffraction recorded from a perdeuterated DHFR/MTX crystal at NE-CAT 8-BM at the Advanced Photon Source. The reflections extend to 1.2Å, and the unit cell is isomorphous with the native DHFR/MTX crystals. The data revealed that these crystals are well-ordered and possess the same symmetry as the native crystals.

Table 3.2: Synchrotron X-ray diffraction data statistics: Perdeuterated DHFR/MTX

Synchrotron facility	Advanced Photon Source
Beamline	NE-CAT; 8-BM
Detector	ADSC Quantum-315
Temperature (K)	100
Wavelength used (Å)	0.980
Resolution range **Highest shell**	25.0-1.25Å 1.29-1.25Å
Space group	$P6_1$
Unit cell constants **a, b, c in Å** **α, β, γ in °**	 91.8, 91.8, 73.0 90, 90, 120
Total reflections [1]	222970 (15565)
Unique reflections	18497 (1852)
Completeness (%)	98.7 (86.6)
R_{symm} (%)[2]	6.0 (38.0)
I/σ	20.0 (2.0)
Multiplicity	2.7 (2.0)

[1]Numbers in parentheses correspond to values measured within the highest resolution shell

[2]$R_{symm} = \Sigma_{hkl} \Sigma_i |I_{i(hkl)} - I_{hkl}| / \Sigma_{hkl} \Sigma_i I_{(hkl)}$

0.5 x 0.5 x 0.2 mm, or about 0.02 mm^3, again taking into account the hexagonal morphology. Collaboration with scientists at LANSCE is ongoing to produce perdeuterated DHFR on a large scale to be able to grow larger volume crystals for future neutron experiments at LANSCE.

Finally, one possible strategy for modifying the data collection at the ILL that has not been pursued is to use a narrower wavelength band pass filter ($d\lambda/\lambda$=15%) and longer wavelength neutrons, centered at λ=3.85Å for data collection. This would effectively reduce the density of reflections and the number of spatial overlaps on the LADI cylindrical detector. Whilst this strategy doubles the number of data frames required, the narrower $d\lambda/\lambda$ bandpass would reduce the experimental background and the use of larger D_2O-soaked or perdeuterated crystals would deliver significant improvement in the signal to noise ratio of the data. However, the increase needed in total experimental time has not been feasible so far for the recent run cycles. In fact, as of the middle of 2005, the ILL will be shut down for experiments until 2006. Additionally, this has not been pursued because a larger volume crystal has now been used for another experiment at the ILL, and a complete data set has been collected on another large volume crystal at LANSCE. This latter data will be the focus of the neutron crystallography results discussed in the next chapter.

Chapter 4. X-ray and neutron crystallographic studies of the DHFR/MTX complex

4.1 Background

This chapter details results for crystallography experiments performed on the *E. coli* DHFR/MTX complex. The ultrahigh resolution X-ray structure reported here is compared to the 1.9Å resolution X-ray structure deposited into the PDB (3DRC) (Warren *et al.* 1991) and results are reported based on analysis which could not previously be undertaken due to the lack of atomic resolution. Outstanding questions concerning dynamics and hydrogen atom locations in the active site were left unanswered by this X-ray structure and, thus, a neutron structure of the same complex was solved. Additionally, preliminary neutron crystallography (NC) data were collected at the Institut Laue-Langevin (ILL), referred to here as "preliminary" because a more complete data set has since been collected on a larger volume D_2O-soaked crystal at the Protein Crystallography Station (PCS) at Los Alamos Neutron Scattering Center (LANSCE). This latter experiment has provided a more complete data set for the refinement of the neutron structure and analysis of active site protonation states and H/D exchange. The ILL data set is reported here because it serves as a benchmark for evaluating improvements in sample preparation, and if paired with a more recent data set collected at ILL on larger volume crystals, it then also allows at least a general comparison between reactor (ILL) and spallation neutron sources (LANSCE).

Characterization of Asp 27 and its role in ligand binding to E. coli DHFR

From Joseph Kraut's laboratory, the initial crystal structure of *E. coli* DHFR with substrate bound (a ternary complex with folate and $NADP^+$) revealed that the N5 atom of FOL was 5Å distant from the carboxylate group of the important catalytic residue, Asp 27 (Bystroff *et al.* 1990). It also revealed that the binding orientation of folate relative to MTX differs by the respective rotation of the pterin rings (**Figure 4.1**). The MTX pteridine ring is in a conformation which places its NA2 amide group and N1 atom within hydrogen bonding distance (~2.7-2.8Å) of the Asp27 carboxylate (Bolin *et al.* 1982; Matthews *et al.* 1977; Matthews *et al.* 1985). It has been speculated that this orientation is preferred so as to allow a salt bridge to form between the MTX N1 and the Asp27 Oδ2, a hypothesis seemingly defended by the fact that the binding affinity for MTX is much stronger than for folate or even the natural substrate, DHF (Stone *et al.* 1983a); (Stone *et al.* 1984); (Appleman *et al.* 1988). To form a true ionic interaction, the MTX N1 atom would have to be protonated and, thus, positively charged. It would follow that the Asp27 carboxylate would also be ionized, maintaining resonance and sharing a negative charge across the Oδ1-Cγ-Oδ2 atoms. A few experimental studies, none of which are crystallographic or absolutely direct, have broached this issue of charge states. Raymond Blakley's group used ^{13}C-NMR (Coco *et al.* 1981; Coco *et al.* 1983) to reveal that the pK_a of the MTX N1 atom increased tremendously when the MTX is bound to DHFR, from 5.7 in solution to >10 when enzyme-bound. Also using ^{13}C-NMR but observing the chemical shifts of these on labeled Asp residues, Gordon Robert's group (Casarotto *et al.* 1999) established that, in the homologous *L. casei* DHFR, no Asp carboxylate has a pK_a above 4.0, at least when folate, DHF or a ternary

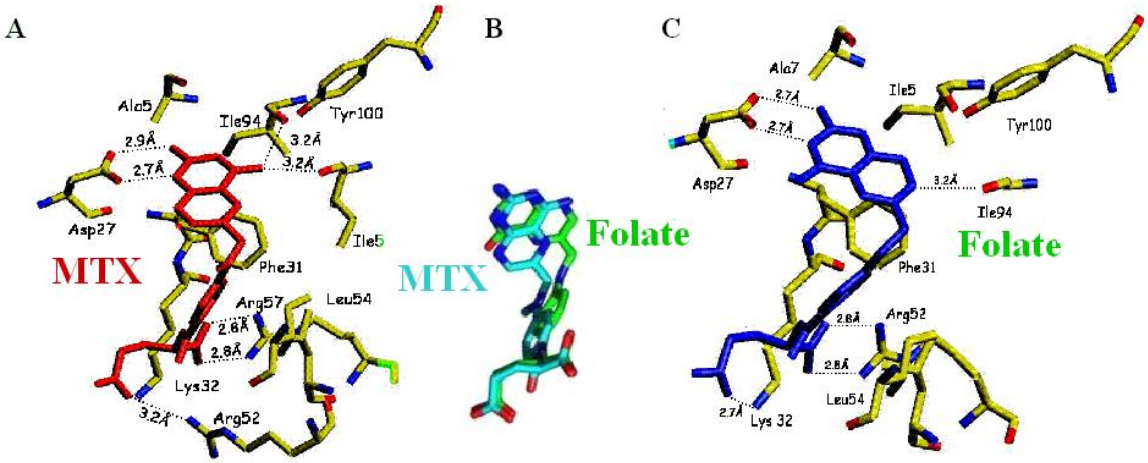


Figure 4.1: The conformation of MTX in the DHFR active site is altered significantly compared to FOL due to a differing rotamer of their pterin rings. A) Interaction distances between MTX and DHFR active site residues (this is the 1.0Å model reported here). **B)** MTX is oriented so as to have its N1 and NA2 atoms within contact distance of the Asp27 whereas FOL positions its NA2 atom and its N3-C4-O4 tautomeric group in an analogous fashion. **C)** Interaction distances between FOL and DHFR active site residues (this is also a 1.0Å model refined by Dr. Anna Gardberg in the Dealwis laboratory; it is shown here not as a report of my results but for comparison purposes only).

complex is bound. These results strongly suggested that the Asp27•MTX interaction could very well be ionic in nature. A computational study by Stephen Benkovic's group (Cannon *et al.* 1997) where theoretical pK_a calculations of the Asp27 and the MTX N1 were measured at dielectric constants nearer to what one would expect for the interior of a globular protein (ε = 4-10). They showed that the proton affinities for the two groups drastically change below an ε of 10; in fact, both the Asp27 and the N1 titrate, leaving them neutral. Their conclusion was that the interaction should be described as a neutral dipole-dipole and not ionic (Cannon *et al.* 1997a; Cannon *et al.* 1997b). It is one of the major aims of this dissertation research to reveal directly the nature of the interaction between the Asp27 carboxylate and the MTX N1. The best method to answer this question is protein crystallography by X-rays and neutrons.

The powerful tool of X-ray crystallography in macromolecular structure determination

X-ray crystallography has proven to be an eminently powerful technique for macromolecular structure determination. At moderate resolutions (2.5-3.0Å), obvious secondary structural features are revealed, the main chain can be traced (important in *de novo* structure determination), and residue side chains can normally be modeled into $2\boldsymbol{F}_o - \boldsymbol{F}_c$ electron density (Creighton 1993). Additionally, if one has collected data to high enough resolution (at or beyond 1.0Å) at cryogenic temperatures (-170 to -180°C (~100 K) can be achieved with liquid N_2, -250 to -260°C (~15 K) with liquid He), one can assign some (but not all) hydrogen atom positions, quantify and visualize anisotropic displacement (non-spherical atomic vibration), identify alternate side chain and even main chain conformations, and, at very high resolution (normally beyond 0.9Å),

deformation in the electron density distribution (Dauter *et al.* 1997b); (Longhi *et al.* 1998). Recently, Teeter and colleagues have refined the structure of two proteins, crambin (5kDa, to 0.54Å) and savinase (27kDa, to 0.9Å), and performed averaged density measurements for the repeating peptide bond units. Intriguingly, they discovered deformations in the electron density around the peptide bond atoms while clearly observing hydrogen peaks in their maps off the amide nitrogens and the alpha carbons (Lamzin *et al.* 1999); (Jelsch *et al.* 2000). Podjarny has reported similar behavior for the peptide bond for aldose reductase (to 0.66Å) and has also been able to model multiple hydrogen atoms (Howard *et al.* 2004). However, there are few studies so far which have revealed this level of detail. Crystals have inherent limitations on diffraction power, and a liquid He cryostat (similar to the one used for the crambin study at EMBL-Hamburg (Jelsch *et al.* 2000)) has only been implemented at a few synchrotron beamlines around the world. As far as the modeling of hydrogen atoms, the obvious importance for their visualization is for defining protonation states of "heavier" atoms (in the macromolecular case, these are C, N, O, S, and P); this would reveal the type of noncovalent interaction between two species. The covalent bond involving the Cα and the carbonyl carbon on the main chain is ~1.5Å in length, near the wavelength of X-rays used in most in-house diffractometers (Cu κ_A radiation = 1.514Å). From diffraction theory, the minimum Bragg spacing that can be resolved within a crystal (thus, the maximum resolution possible) is λ/2. So, theoretically, data to 0.78Å resolution (the d_{min}) could be obtained using an in-house diffractometer with a Cu rotating anode. (However, the X-ray detector would have to be very large or have the ability to be swung and/or offset.) To collect even higher resolution data, a synchrotron source can supply X-ray wavelengths nearer to the bond

distance between a Cα and its bound hydrogen, or ~1.0Å. Although it is dependent upon many factors inherent to the crystal itself (such as the lattice construction, packing, and any disorder of the molecule), theoretically one should be able to resolve hydrogen from other protein and solvent atoms that are sufficiently ordered using X-ray radiation.

The limits of X-rays and re-emergence of neutrons in crystallography

However, X-rays distinguish between different type atoms only by their electronic content, sulfur being the strongest scatterer common in proteins (Met and Cys) due to its 16 electrons and, thus, can normally be readily identified in electron density maps. Due to hydrogen's weak X-ray scattering (only one electron) and its inherent dynamic nature (normally, it maintains elevated thermal motion and its B-factor is linked to the B-factor of the atom to which it is bound (El-Kabbani *et al.* 2004), it is difficult at ultrahigh resolution (beyond 1.2Å) and essentially impossible at lower resolution (lower than 1.5Å) to confidently model hydrogen into electron density maps. In fact, it is a positive single difference density map ($\mathbf{F}_o - \mathbf{F}_c$) *above* a contour level of 2.5-3σ which one uses normally to attempt this task (Schmidt *et al.* 2003) (Minichino *et al.* 2003). Hydrogen constitutes ~50% of a macromolecule's atoms and its positioning can be extremely important in protein function (i.e. acid-base catalysis in enzyme mechanisms) yet its placement is normally only *implied* by inferences made using difference electron density, potential hydrogen bonding patterns, and standard geometries.

Macromolecular NC can reveal hydrogen atom positions at moderate resolution (2.0-2.5Å). Neutrons are scattered by the atomic nuclei instead of by the diffuse electron cloud. Since the initial description of a neutron structure of a protein (myoglobin) in 1969 (Schoenborn 1969) and into the 1980's, when the neutron structure of several important

enzymes were elucidated, including trypsin (Kossiakoff *et al.* 1980; Kossiakoff *et al.* 1981), RNase A (Wlodawer *et al.* 1983), and insulin (Wlodawer *et al.* 1989), there has been immense interest in NC. However, the practical use of neutrons has been arduous and time-consuming due to the weak fluxes (a neutron pulse is weaker in energy than an in-house X-ray source), the requirement of large volume crystals (>1mm^3), and the lack of dedicated protein crystallography instruments at neutron sources (only 1 in North America and 4 in the world). Even so, NC seems to be undergoing a resurgence in structural biology, as evidenced by the multiple publications that have been borne out over the last 8 years, from the bacterial cell-wall glycosidase, lysozyme (Niimura *et al.* 1997); to an aspartic protease, endothiapepsin (Coates *et al.* 2001); to a spallation neutron experiment on the glycolytic enzyme, xylose isomerase (XI; (Hanson *et al.* 2004); to a cryo experiment on a lectin, concanavalin A (Blakeley *et al.* 2004), just to select a few.

The ability to distinguish hydrogen in protein structures solved by neutrons is actually a misnomer. One can only visualize hydrogen directly by the use of negative difference nuclear density (e.g. an $\boldsymbol{F}_o - \boldsymbol{F}_c$ map contoured at -3σ). The reason is that the interaction of the incident neutron beam with the hydrogen nucleus (which contains only a single proton) is the major cause of incoherent scattering in NC and neutron scattering experiments; hydrogen's neutron scattering cross-section (a measure of the probability, magnitude, and area of an interaction event between the neutron beam and the hydrogen nucleus; measured in barns (10^{-24} cm^2)) is not only incoherent but greater in incoherence than any other atom normally found in proteins (Langan *et al.* 2004b) (**Table 4.1**). Incoherence is a discrepancy of scattered waveforms when compared to one another; there is no net contribution to a signal resulting from constructive or destructive

Table 4.1: Neutron scattering lengths (or scattered wave amplitudes) and cross-sections for atoms in biological macromolecules.

	^{1}H	^{2}H (D)	^{12}C	^{15}N	^{16}O
$B_{coherent}$ (fm)[1]	-3.74	+6.67	+6.65	+9.36	+5.81
(X-ray)	(2.80)	(2.80)	(1.69)	(1.97)	(2.25)
$\sigma_{coherent}$[2]	1.76	5.59	5.56	11.03	4.23
$\sigma_{incoherent}$[2]	80.27	2.05	0.0	0.49	0.0

[1] Scattering lengths are reported as measured in femtometers (10^{-12} m).
[2] Coherent and incoherent total cross-sections (σ) are reported as measured in barns, with 1 barn equaling 10^{-28} m^2 (about the size of uranium's nucleus) and quantifies a cross-sectional area of a nucleus or nuclear interaction. A cross section itself is a measure of probability of interaction between two interacting particles (Here, it would be the interaction between reactor or spallation neutrons and the nuclei of the atoms which comprise a macromolecular crystal.)

interference. It is similar to a randomness in scattering (not just waves that are out of phase but those arising from scattering events which have different frequencies) (Schoenborn *et al.* 1996). The result of hydrogen scattering is that the background on the detector will be increased, the signal-to-noise ratio will be reduced and it will take longer to accurately measure diffraction peaks, and, in the maps, nuclear density for atoms bound to hydrogen could be reduced or missing altogether (Meilleur *et al.* 2005). The remedy for this is hydrogen/deuterium (H/D) exchange, either by D_2O-soaking the crystal or by perdeuteration. (The details are given in chapter 3.) Briefly, the exchange of H for D increases overall coherent scattering and, thus, improves signal over background (**Table 4.1**); this is especially true if one perdeuterates the protein, where all the covalently-bound H are substituted for D at the biosynthetic level (Shu *et al.* 2000). Therefore, if the maps warrant its placement, one models deuterium directly into $2\boldsymbol{F}_o - \boldsymbol{F}_c$ nuclear density maps due to its substitution for hydrogen. Deuteration also reduces the sample volume required for NC (Hazemann *et al.* 2005); perdeuteration paired with spallation neutrons may make it possible for standard crystal volumes to be used (~0.05-0.1mm^3) in NC experiments (Schoenborn *et al.* 2004).

Current NC capabilities: spallation vs. reactor sources

Two source types, cold reactor and spallation, exist presently for the production of neutrons for structural biology. An example reactor source is the ILL and details concerning data collection and instrumentation at LADI for NC are presented later in this chapter. Reactors are at an inherent disadvantage to spallation sources due to flux limitations imposed by the fission process (limitations that may not be overcome in the near future) (Schoenborn *et al.* 2004). Spallation neutron sources provide a new tool for

protein crystallographers. Bombarding a metal target with μsec pulses of high-energy protons produces a beam of neutrons with a range of energies and a pulsed time-structure that allows data to be collected with high efficiency and low background noise using the time-of-flight (TOF) Laue method. The first macromolecular NC beam line to be built at a modern spallation source (PCS) has been operating at LANSCE since August 2002 (Langan *et al.* 2004a). LANSCE provides a pulse of neutrons 20 s^{-1} from the spallation impact of protons upon a tungsten target. PCS houses a gadolinium-doped, position-sensitive, large curved detector for faster data collections (Langan *et al.* 2004a). Since the neutrons are produced from a spallation source, the time that the protons strike the target is exactly known and is designated as T_0. Thus, the neutrons have time signatures and their TOF is based on their energy. Neutrons of different energies (therefore, of different wavelengths) arrive at the detector at different times and the TOF information (and the exact wavelength) is recorded for the detected neutron; so, not only does one detect the position of a reflection but also its time of arrival. Thus, the use of the wavelength-resolved Laue method provides data in three dimensions: two for the reflection position on the detector and one for the TOF (Langan *et al.* 2004b). The ability to time-stamp the reflections allows one to bin them based on the TOF, each bin or channel representing a small curvature of reciprocal space. One detector image is then several of these slices adding to a large wedge of reciprocal space (Hanson *et al.* 2004). The maximum wavelength range available at LANSCE is 0.6-7Å, with a normal experimental range of 1-5Å. Wavelength-resolved Laue NC is advantageous over quasi-Laue at reactor sources like ILL because of lower reflection overlap and lower background levels (Langan *et al.* 2004b). An overlay of all TOF channels displaying

each pixel position at the pixel value maximum resembles a conventional Laue pattern because of the integration of multiple time slices into one diffraction image. Although only 10-30 crystal settings are typically required for a complete data set, individual exposure times can be as high as 48 hours. Moderately large sample volumes ($\geq$ 0.2mm^3) that have been extensively D_2O-soaked are still required to enhance signal-to-noise in data collection.

High resolution X-ray and neutron crystallography has been used in this research to describe the positioning of H/D on residues and MTX in the active site of *E. coli* DHFR. Also of interest are the features in the 1.0Å electron density maps that were unavailable in previous crystallographic studies of this complex such as alternate side chain conformers, anisotropy, and highly accurate bond length measurements of the active site residue, Asp 27. By virtue of H/D exchange (and the fact that D scatters neutrons so strongly), an analysis has been made of structural flexibility based on the propensity of backbone amides and side chains to undergo H/D exchange.

4.2 Results and discussion

X-ray crystallography: Synchrotron data collection and processing

Ultrahigh resolution X-ray diffraction data were collected for a DHFR/MTX cocrystal at BioCARS beamline 14-BM-C at the Advanced Photon Source. All data were collected at -170°C using a Quantum 4 detector and an X-ray wavelength of 0.9Å. The program HKL2000 (Otwinowski *et al.* 1997) was used for indexing, integration, and initial scaling of all the data reported here.

An MTX cocrystal (0.4 x 0.4 x 0.2 mm^3) was chosen for high resolution data collection after observing reflections to 0.9Å. The detector was brought to 70 mm from

the crystal and the exposure times were set to 30 s for each image. The short crystal-to-detector distance and long exposure times were used to ensure recording of high resolution reflections and strong intensities for these reflections. Ninety degrees of data about the φ axis were collected at 0.5° per oscillation giving a total of 180 images; the resolution range of this set was 3.2 – 1.0Å. A foil attenuator was mounted directly in front of the low resolution area of the detector (4Å and lower) to protect it from overexposure. The next task was to collect a scan from the same crystal to provide data lower than 4Å that had been effectively blocked and/or weakened by the use of the attenuator. The detector was put back to 100 mm distance, the exposure times decreased to 5 s per oscillation, and the foil was removed. Unfortunately, the integration and scaling statistics for the crystal for this data collection progressively became worse as more frames were incorporated. It may have been advisable to collect the low resolution data first and then collect the high resolution reflections. This is a strategy that was adopted for the apo DHFR data collection described in the next chapter. So, a second MTX cocrystal of similar dimensions was ultimately used just for the purpose of collecting the lower resolution data. This scan encompassed 65° of φ space with 0.5° per oscillation, and the resolution range of this scan was 15.0 – 1.2Å. The mosaicity of both crystals was 0.4°.

Data reduction and scaling

After separate indexing and integration of the intensities from the two scans, the data sets were merged and scaled together. Owing to a swapped indexing convention common to *P*6, the data sets were rescaled using a reindexing matrix (*h k l* to *k h –l*) in the program Scalepack (Otwinowski *et al.* 1997). The overall R_{merge} was 8.9% and the

completeness was 95.3% for a resolution range of 15.0-1.04Å (for the highest resolution shell, 1.04-1.00Å, it is 89.6%). The combined scaling statistics are listed in **Table 4.2**.

Molecular Replacement (MR) solution of the DHFR/MTX complex

Early on in the thesis research (and well before the synchrotron data collection), an in-house X-ray data set was collected on a DHFR/MTX crystal at -170°C with resolution extending to 1.7Å. The unit cell is isomorphous to the one described in the above section, and the space group ($P6_1$) is identical. Based on previous results from *E. coli* DHFR/MTX crystals, it was suspected a screw axis existed on the 6-fold axis. A precession photograph was examined for one of the oscillation images using the program *HKLView* (CCP4, 1985). Along the 00*l* layer, systematic absences of reflections were observed at 6n + 1; this is due to a 6-fold symmetry element combined with a 1/6 cell edge translation. After integration and scaling of the intensities and conversion to structure factor amplitudes, a Mathews coefficient was calculated to determine the solvent content of the unit cell (34.5%) and that there were 2 molecules in the asymmetric unit (AU). The coordinate set from a room temperature model of DHFR/MTX (3DRC; $P6_1$ space group; 1.9Å resolution; only protein atoms were retained) was employed as the target molecule for MR using the program AMoRe (Navaza 2001) for a resolution range, 20.0-3.0Å. The search model was placed inside a model P1 unit cell (73.8 Å, 66.3 Å, and 64.5 Å) for the self-rotation search using a sphere of integration with a radius of 26.4 Å. The correlation coefficient (CC) for the top solution from a rotation and translation search of the unit cell was 0.32 (32%). This solution was then fixed at the 6_1 origin, and the target was translated from solution 1 along the z axis to search for the second molecule (the second solution) within the AU; the resultant CC was 34.6%. Two

Table 4.2: X-ray data processing statistics: DHFR/MTX cocrystal

DHFR crystal form	MTX
BioCARS Beamline;	14-BM-C
Detector	ADSC Q4
Unit cell (a,b,c, in Å)	92.2, 92.2, 73.3
(αβγ, in °)	90, 90, 120
Molecules/ A.U.	2
Solvent content (%)	34.5
Space group	$P6_1$
Resolution range (Å)	15.0-1.00
Highest shell	1.04-1.00
Unique reflections	172,284
Completeness (%)	95.3 (89.6)
R_{merge}[1] (%)	8.9 (67.5)
I/σ	13.7 (1.2)
Multiplicity	3.0 (1.6)
# Data sets used	2

$R_{merge}{}^{1} = \Sigma_{hkl} \Sigma_i |I_{i(hkl)} - I_{hkl}|/\Sigma_{hkl} \Sigma_i I_{(hkl)}$

solutions were then fit within the asymmetric unit (with one fixed at the origin), and the CC and the R-factor were calculated to be 75% and 35%, respectively.

Refinement

The output AMoRe MR solution models for the two molecules in the AU were combined into one coordinate file, and this was refined against the in-house 1.7Å reflection data using the rigid body method in CNS (Brunger *et al.* 1998). The refined model was then subjected to simulated annealing to reduce model bias from 3DRC (Brunger *et al.* 1997). The output model from annealing and energy minimization (R_{fac} and R_{free} = 36.0% and 36.7%) was used to refine against a reflection data set collected to 1.4Å at the Brookhaven synchrotron (the NSLS). (This was collected about 3 months prior to the APS synchrotron experiment. It is mentioned here because it was used in refinement in between the in-house and the APS data collection experiments.) Rigid body refinement and minimization in CNS along with initial rebuilding in O reduced the R_{free} to 34%. This model was ultimately used for minimization in CNS and refined against the reindexed reflection data collected to 1.0Å at the Argonne synchrotron (the APS). After switching over to Refmac5 for maximum likelihood (ML) refinement (Murshodov et al., 1998; invoked within the CCP4 suite; CCP4, 1985), the proper monomer library was invoked within CCP4 so as to include the MTX ligands in the refinement. The resultant R_{fac} and R_{free} values from the initial run were 29.3% and 30.3%, respectively. After the three cycles of refinement in Refmac5 and SHELX, solvent addition, refinement of anisotropy (see below) and modeling of several alternate side chain conformations, the R_{free} decreased to 23.4%. The addition of hydrogens in riding positions in SHELX further reduced the R_{free} to 18.4%. After a final refinement in Refmac5 to relax the sigma

weighting matrix, the R_{free} decreased to 16.5%. Refinement and analysis statistics are in **Table 4.3** and unit cell contents are listed in **Table 4.4**. A list of residues for which alternate side chain conformations have been modeled is in **Table 4.5**.

Comparison of the monomers within the AU

The two DHFR/MTX molecules in the asymmetric unit are shown in **Figure 4.2**. As expected for *E. coli* DHFR, these models possess a tertiary structure consisting of a pseudo-Rossman fold (Matthews *et al.* 1977); (Sawaya *et al.* 1997) of alternating β-strands (8, named βA- βH) and α-helices (4, named αB, αC, αE, and αF) with three major regulatory loop regions. The largest divergence between the two monomers is at these regions, termed the Met20 (aa 9-24), βF-βG (aa 117-131), and βG-βH (aa 142-149) loops, due to the fact that the Met20 loop differs in conformation from partially occluded (monomer A) to closed (monomer B). The overall Cα r.m.s.d. between monomers A and B is 0.52Å (calculated using a least squares superposition in the molecular graphics program O (Jones *et al.* 1991)). The largest Cα r.m.s.d. between monomers A and B is for Glu 17, in the core of the Met20 loop, at 5.2Å. Alternate side chain conformers were modeled if sufficient $2\mathbf{F}_o$-$\mathbf{F}_c$ ($\geq$1.5σ) and/or $\mathbf{F}_o$-$\mathbf{F}_c$ ($\geq$ 2.5σ) density existed at favorable rotamer positions. In the final refined model, there are 12 alternate conformers for monomer A and 10 for monomer B (see **Figure 4.3A, B** for examples). Over 600 solvent molecules could be modeled into $2\mathbf{F}_o$-$\mathbf{F}_c$ electron density contoured above 1.3σ.

Isotropic and anisotropic B-factor description and analysis of the MTX structure

Because X-rays are scattered from the electron cloud of an atom, any vibrational motion of the atom has an effect on its X-ray scattering power. Indeed, there is normally

Table 4.3: Refinement statistics for X-ray crystallography: DHFR/MTX cocrystal

DHFR crystal form	MTX
Resolution range (Å)	15.0-1.00
Highest shell	1.04-1.00
Final Isotropic R_{fac}/R_{free} (Refmac5) (%)	23.4/25.3
"Best" Anisotropic R_{fac}/R_{free} (Refmac5) (%)	18.7/21.0
Same coordinates; Isotropic in SHELXH (%)	21.2/24.0
"Best" Anisotropic R_{fac}/R_{free} (SHELXH) (%)	16.3/19.8
All riding hydrogens added; R_{fac}/R_{free} (SHELXH) (%)	15.1/18.4
Final R_{fac}/R_{free} (SHELXH) (%)	14.1/17.7
Final sigma weighting matrix (Refmac5)	7.0
Final R_{fac}^2/R_{free}^3 (Refmac5)	14.3/16.6
Final r.m.s.d. for bond distances	0.020Å
angles	2.036°

$R_{fac}^2 = \Sigma_{hkl} \mid |\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}| \mid / \Sigma_{hkl} |\boldsymbol{F}_{obs}|$
$R_{free}^3 = \Sigma_{hkl} \mid |\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}| \mid / \Sigma_{hkl} |\boldsymbol{F}_{obs}|$

Table 4.4: Unit cell contents of the DHFR/MTX cocrystal (X-ray)

Molecules/A.U.	2
Protein atoms (non-hydrogen)	2638
Water molecules	630
Metals	3 (2 Mg^{2+}; 1 Cl^{-})
Ligands	2 (MTX)

Table 4.5: Residues which possess alternate side chain conformations in the two monomers of the DHFR/MTX AU

Monomer/AU	A	B
# of alternate conformers	12	10
Residue/Secondary structural element	Leu 4/βA strand	Leu 8/βA strand
	Asn 18/Met20 loop	Ile 61/βC strand
	Leu 36/αB helix	Ile 82/αE helix
	Pro 39/βB strand	Val 88/βE strand
	Ser 49/αC helix	Thr 123/βF-βG loop
	Val 99/αF helix	Ser 138/βG strand
	Lys 109/βF strand	Asp 142/ βG-βH loop
	Ser 138/βG strand	Ser 150/ βG-βH loop
	Ser 150 (2[a])/ βG-βH loop	Cys 152/βH strand
	Glu 154/βH strand	Leu 156/βH strand
	Leu 156/βH strand	

[a] Ser 150 in monomer A possesses two alternate conformers for its side chain hydroxyl group

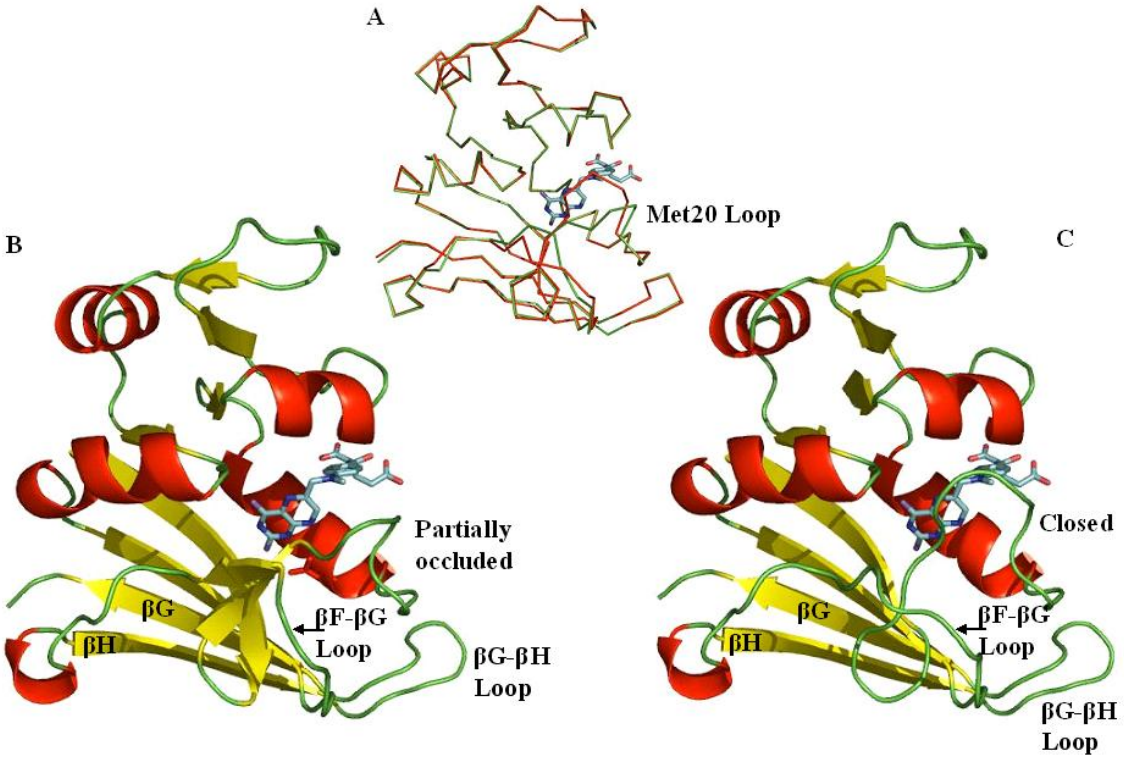


Figure 4.2: Comparison of the two monomers in the asymmetric unit of the 1.0Å DHFR/MTX model. A) Superposition of the two monomers, with the largest Cα deviations occurring at the Met20 loop. Major differences also exist for the βF-βG loops. The A monomer is in green and the B monomer is in red. **B)** The A monomer has a Met20 loop which adopts a partially occluded conformation and the βG-βH loop is positioned to stabilize it via hydrogen bonding. **C)** The B monomer has a Met20 loop which adopts a closed conformation. Part of the βF-βG loop (residues 120-122) makes contacts with the core of the closed Met20 loop, stabilizing it to block the substrate binding site and opening the nicotinamide-binding pocket.

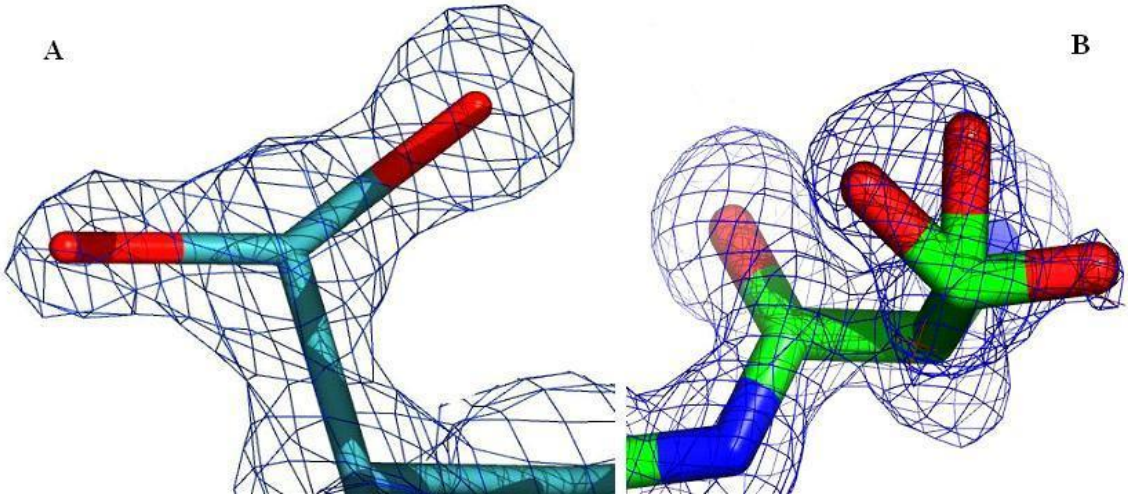


Figure 4.3: High resolution electron density maps allow the modeling of alternate side chain conformations for the DHFR/MTX structure. Examples include one alternate conformer for **(A)** Ser 49 (monomer A) and **(B)** two alternate conformers for Ser 150 (monomer B). Shown in blue is $2\boldsymbol{F}_o$-$\boldsymbol{F}_c$ electron density contoured at 2.0σ.

an associated "smearing" of electron density with atoms that possess large vibrational amplitudes. Therefore, a correction can be applied to each of the scattering factors, *f*, for the given vibrating atoms as

$f = f_0 \exp\,[-8\pi^2\{\mathbf{u}^2\}(\sin^2\theta/\lambda^2)]$, (Equation 1)

where $\mathbf{u}^2$ is the mean squared vibrational amplitude of the atom, θ is the scattering angle, and λ is the X-ray wavelength. The magnitude of the vibration is correlated to temperature, so $\mathbf{u}^2$ is normally referred to as a thermal parameter or temperature factor. However, atomic vibration and the smearing of density can also be related to inherent dynamics and disorder in the protein as well as to variation of the proper location of the atom's centroid among the cumulative unit cells (Merritt 1999a). Trueblood *et al.* (1996) have suggested that "displacement parameter" be used instead of temperature or thermal factor (Trueblood *et al.* 1996). It is normally referred to as the B-factor, with the term *B* being defined as

$B = 8\pi^2\{\mathbf{u}^2\}$. (2)

Here, $\mathbf{u}^2$ is one parameter (an area measurement) that describes isotropic distribution of atomic vibration, a uniform displacement that is best represented as a sphere. This is an inaccurate approximation for the vibration of most atoms. Thus, an anisotropic B-factor describing the non-uniform displacement of an atom is more appropriate and takes the form of a 6-parameter matrix. These 6 parameters define an ellipsoid within which the atom vibrates (Tronrud 2004). Although it is a more accurate description of an atom's position, refinement of anisotropic B-factors increases the total parameters to be determined and refined per atom from 5 (*xyz*, isotropic B-factor, and occupancy) to 10 (instead of one displacement parameter there are now 6). Only with a large ratio of

observations (unique reflections) to parameters (perhaps 5:1 or better) can anisotropic displacement parameter (ADP) refinement be justified (Sheldrick *et al.* 1997). Essentially, high resolution data is required. Inclusion of ADPs considerably reduces the R_{fac} and R_{free} by lowering the uncertainty in atomic positions.

For the MTX structure, ADPs were refined in two separate programs (Refmac5 and SHELX). Restrained refinement of the model in SHELX using the lone isotropic parameter (in $Å^2$) to describe the vibrational mode of each atom provided a data-to-parameter ratio of ~15:1 whereas restrained ADP refinement of the model in SHELX still provided a ratio of ~6.5:1 and decreased the R_{free} between 3-4% in both programs. Using the program PARVATI (Merritt 1999a), an analysis of the distribution of anisotropy and thermal motion from the input refined MTX coordinate file was performed and graphical displays are shown in **Figure 4.4**. Anisotropy is defined here as the ratio between the smallest and the largest eigenvectors of the ADP matrix, related to the length of the shortest and the longest principal axes (eigenvectors) of the ellipsoid (Trueblood *et al.* 1996),

$$A = \underline{E}_{min} / \underline{E}_{max.} \qquad (3)$$

So, a value for an atom near 1.0 would describe a sphere whereas values nearer to 0.5 describe an ellipsoid and values nearer to 0.1 can describe either a “cigar” or a “pancake” distribution (Merritt 1999a). Not surprisingly, the core residues for both MTX monomers in the AU can mainly be described as isotropic; that is they have a spherical, uniform distribution of motion (seen as blue spheres in **Figure 4.4A**). However, different regions between the two monomers show higher anisotropic behavior, and this seems to be strongly correlated to higher B-factors for the atoms in those regions (seen as orange-red

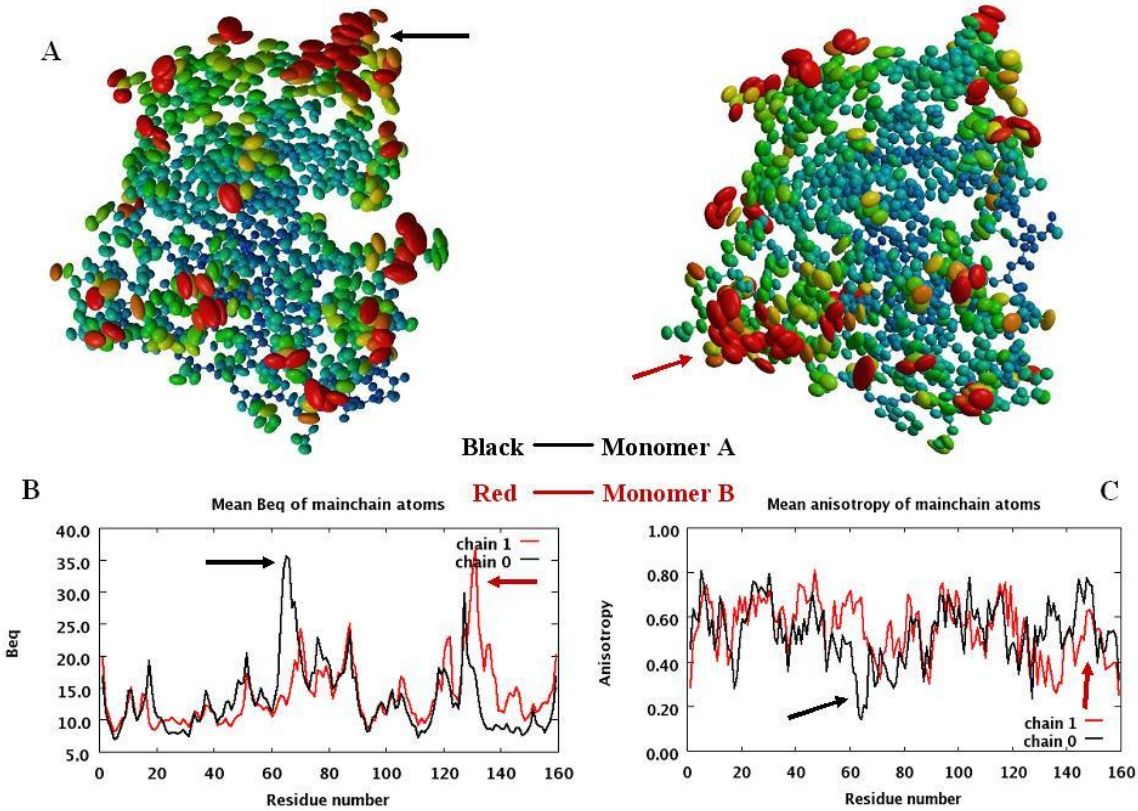


Figure 4.4: Anisotropic displacement and B-factor analysis for the 1.0Å resolution DHFR/MTX structure (monomers A and B). A) All non-hydrogen atoms for monomers A and B are shown as thermal ellipsoid probabilities, with blue-to-green spheres representing atoms displaying more isotropic motion (A = ~1.0) whereas yellow-to-red ellipses represent atoms displaying more anisotropic motion (A = ~0.5) **B)** B-factors (in $Å^2$) for peptide backbone atoms for monomers A (black) and B (red) **C)** Anisotropy ratio values (where $A = E_{min} / E_{max}$) for peptide backbone atoms; larger A values describe more isotropic behavior whereas smaller A values describe more anisotropic (non-spherical atomic vibrational distribution); higher B-factors and ADPs in the monomers are indicated by black and red arrows.

ellipsoids in **Figure 4.4B**, **C**). For example, the region between the βC and βD strands (aa 62-70) is fairly disordered in monomer A with high B-factors and ADP ratios around 0.3-0.55 (**Figure 4.4**, black arrows). In monomer A, Glu65 and Pro66 have little electron density for atoms other than the main chain. However, in monomer B, this region is quite ordered with lower overall B-factors and higher *A* values (more isotropic) and, in fact, much more than the previously deposited MTX structure (solved to 1.9Å; (Warren *et al.* 1991). A similar region of discrepancy, this time monomer B having a higher anisotropy, is in the βG-βH loop (aa 142-149) and into the C-terminus of the protein (**Figure 4.4**, red arrows). In monomer B, the Met20 loop adopts a closed conformation and the βG-βH loop is not needed to stabilize it, possibly contributing to this region's enhanced mobility (**Figure 4.4B**, **C**) as compared to monomer A (where the βG-βH loop forms hydrogen bonds with the core and C-terminal end of the partially occluded Met20 loop). Merritt (1999) established a consensus value for *A* (equation 3) for protein atoms from a sampling of reported high resolution structures (range = 1.4-0.8Å). The mean *A* is ~0.45, a significant deviation from an isotropic description of atomic motion in proteins (Merritt 1999b). The overall mean *A* for the MTX structure (including both monomers) is ~0.55. Therefore, the DHFR/MTX model at this resolution can be generally described as slightly more isotropic in nature, however, with large ADPs for most surface and loop regions.

Comparison to the previously reported MTX X-ray structure

A comparison between the two monomers of this 1.0Å MTX cryo model and the two monomers of the previously deposited 1.9Å MTX room temperature (RT) model (3DRC; Warren *et al.*, 1991) revealed many side chain rotamers that were different between the two structures, most conformational differences justified by the high

resolution data (in other words, the cryo model fit the 1.0Å electron density better than the RT model; this is normally expected but recall that with increased resolution one has increased precision for the structural model). Superposition of the monomer A models revealed an overall Cα r.m.s.d. of 0.31Å while superposition of the monomer B models revealed an r.m.s.d. of 0.35Å (**Figure 4.5A, C**). It should be reiterated that the unit cells are isomorphous and the space group is the same ($P6_1$) for the MTX structure reported here and for 3DRC. For the A molecules, there are three major areas of divergence between the structures. The side chain of His 45 adopts a different rotamer about the χ^1 torsion, causing the imidazole ring in the high resolution model to point towards a solvent channel and the Nε2 atom to form a potential hydrogen bond with Wat 274 (**Figure 4.5B**). This also orients the imidazole ring towards the backbone of Gly 15 and the core residues of the Met20 loop (~6.5Å away). The His 45 residue in the 3DRC model is pointed towards the surface and only part of the imidazole ring fits into 1σ $2\mathbf{F}_o$-$\mathbf{F}_c$ electron density. In the region between the βC and βD strands (aa 62-70), 3DRC is slightly more complete with Glu 65 in the coordinate file; for the model reported here, the Glu 65 had only sufficient electron density up to the Cβ atom. Two water molecules are currently modeled for the Glu carboxylate group. The average main chain difference for the models between aa 64-72 is ~0.5Å. The βF-βG loop is the region most different between the two structures; different side chain rotamers exist for Glu 118, Glu 120, and Asp 122 and are justified for the high resolution model based on the electron density. Due to ~1.0Å shift between the models for backbone atoms from Asp 127 to Asp 132, most of the side chains for these residues are also widely divergent. The βF-βG loop participates

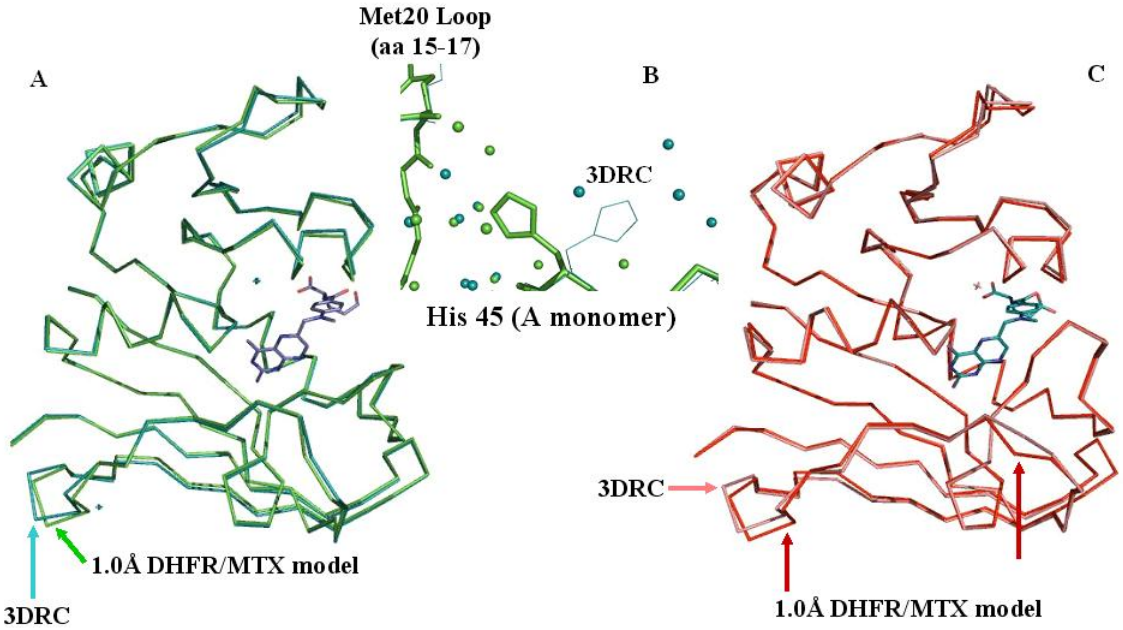


Figure 4.5: Comparison of monomers A and B from the 1.0Å model reported here and the previously reported 1.9Å model (3DRC; (Warren *et al.* 1991). Some of the regions of the largest Cα divergences are indicated by thick green and red arrows. **A)** Superposition of monomer A from the 1.0Å (green) and the 3DRC (cyan) model. The Cα r.m.s.d. between the two models is 0.32Å. **B)** A major difference in the χ^1 rotamer for His 45 leads to the imidazole ring pointed in opposite directions; the 1.0Å model (thick green line) directs the His 45 towards the Met20 loop whereas His 45 in the 3DRC model (thin cyan line) points toward a solvent channel and the surface. **C)** Superposition of monomer B from the 1.0Å (red) and the 3DRC (salmon) model. The Cα r.m.s.d. between the two models is 0.35Å.

less in Met20 loop stabilization in monomer A, due to the partially occluded conformation adopted by the Met20 loop. This may lend to its flexibility (though the B-factors compared to the B monomer with a closed Met20 loop conformation are nearly identical); the 3DRC diffraction data was collected at RT whereas the 1.0Å diffraction data reported here was collected at -180°C. The difference in temperature, with its effect on atomic motion, could explain some of the divergence between the models, especially in a region such as inherently dynamic as the βF-βG loop.

For the B monomers, the region between the βC and βD strands (aa 62-70) is much more complete in the 1.0Å model than in 3DRC. The backbone Cα atoms diverge by as much as 1.0Å here; 3DRC has only backbone atoms for Pro 66 and only an Ala (a Cβ) at residue 68 (the proper residue is Thr, which is in the high resolution model). There is also a major conformational change between the models for the Thr 68 carbonyl to Ala 69 amide nitrogen backbone, a shift of ~1.5Å between the Thr 68 carbonyl carbons. There is also a backbone shift of ~1.0Å at Asp 122 and a different side chain rotamer about the χ2 torsion angle, causing the Oδ2 atom in the 1.0Å model to make a 2.4Å contact with Wat246. Another large main chain divergence between the models occurs at the end of the βF-βG loop with ~1.2Å differences in Cα position for Pro 130 and Asp 131. Finally, the C-terminus is slightly different between models with varying rotamers adopted for Arg 158 and complete model available for Arg 159 in the 1.0Å model (there are no coordinates in 3DRC for this side chain).

Difference electron density analysis to identify putative hydrogen atoms

Single difference density maps ($\mathbf{F}_o$ - $\mathbf{F}_c$) were calculated from the high resolution data to identify positive density peaks at $\geq$ 2.5σ that could be attributed to hydrogen atoms in the structure. The resolution of the data (1.0Å) is equivalent to reported C-H and N-H bond lengths from spectroscopic data (Stout *et al.* 1989); so, with sufficient data completeness and fairly low B-factors for the "heavier" atoms to which they are bound, at least a few hydrogen atoms should be able to be identified in difference density at this contour level (Howard *et al.* 2004). However, the weak scattering power remains the main reason why hydrogen is difficult to visualize in electron density maps. On the Trp 22 indole ring in both monomers, difference peaks exist about 1.1Å distant from Cε3 (monomer A) or Nε1 (monomer B) (**Figure 4.6A, B**). The B-factors for the Cε3 and the Nε1 atoms are ~12Å^2. Other examples are peaks in the A monomer near the Cγ on Pro 39 and Pro 55, and in the B monomer near the Nε2 on His 45, the Oγ on Ser 138 and 150.

A major motive for undertaking this high resolution experiment was to examine the DHFR active site for any positive difference density peaks on the Asp 27 and on MTX. At an intermediate step in the refinement (when the R_{free} was ~20%, thus, the refinement had not yet converged), positive peaks adjacent to the Asp 27 Oδ2 and the MTX C7 and N8 atoms could be observed at a contour level of $\geq$ 2.5σ, well above background (**Figure 4.7A**). The peak ~1.0Å away from the MTX C7 atom is easily explained since it is supposed to have hydrogen covalently bound to it. The peak near the MTX N8 is not as easily understood since this N shares a double bond with C7 and a single bond with C8A and no pK_a has been ascribed to this atom. A possibility, albeit a

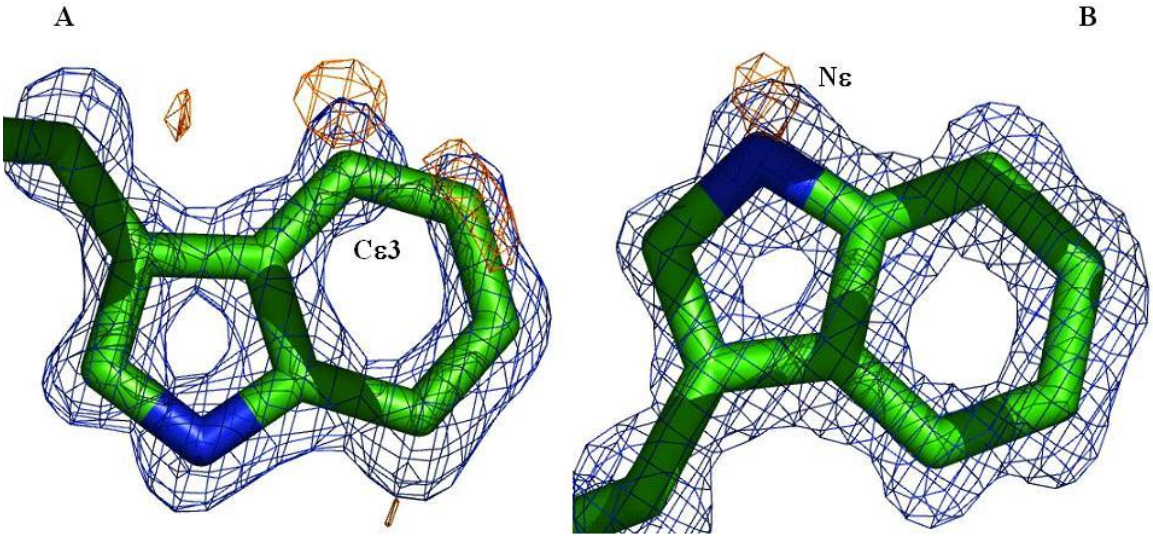


Figure 4.6: Positive difference density maps reveal putative hydrogen positions on the Trp 22 indole ring. Shown in blue is the $2\boldsymbol{F}_o$-$\boldsymbol{F}_c$ electron density map contoured at 2.0σ and in orange is the $\boldsymbol{F}_o$-$\boldsymbol{F}_c$ electron density map contoured at 3.0σ. **A)** Difference density peak adjacent to the Cε3 atom on Trp 22 in monomer A; **B)** Difference density peak adjacent to the Nε1 atom on Trp 22 in monomer B. Different orientations are shown because the difference peaks are offset from the plane of the paper (the peaks are sitting slightly "above" the Cε3 and Nε1 atoms, projecting out from the paper).

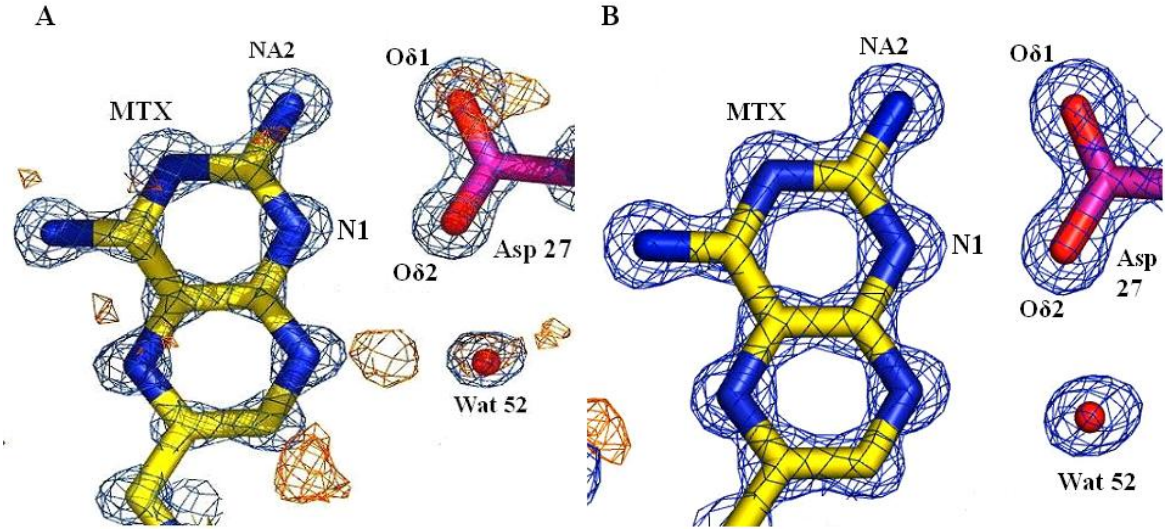


Figure 4.7: High resolution electron density maps for the MTX pteridine ring and the DHFR active site. Shown in blue is the $2\boldsymbol{F}_o$-$\boldsymbol{F}_c$ electron density map contoured at 2.0σ and in orange is the $\boldsymbol{F}_o$-$\boldsymbol{F}_c$ electron density map contoured at 2.5σ. **A:** Intermediate in the refinement process (R_{free} = ~20%), a +2.5σ difference density peak adjacent to the Asp 27 Oδ2 was observed and was initially speculated to be a possible hydrogen atom. **B:** After further refinement (R_{free} = ~16.5%), no difference density peaks (contoured $\geq$+2.5σ) were observed near the Asp 27.

weak one, is that this is observance of a hydrogen bond between the N8 and the Wat 52 about 2.8Å distant. Further complicating the analysis of the active site difference peaks is the position of the peak adjacent to the Asp 27 Oδ1 atom. The peak center contoured at 2.5σ is about 1.1Å distant from the Oδ1 atom in the model; however, it is the angle from the Oδ1 that is of most concern. The Asp 27 carboxylate group and the MTX pteridine ring lie in a near-perfect plane. So, the Asp 27 Oδ1 to MTX NA2 contact is a near-perfect linear orientation; the location of the difference peak near the Oδ1 is angled at about 120° off of this plane, facing away from the MTX NA2 and not within hydrogen-bonding contact with another residue or solvent. This seemed to be a strange and unfavorable orientation for hydrogen bound to the only ionizable group in the active site of a highly efficient enzyme. In collaboration with Dr. Hong Guo's laboratory (and in experiments *not* performed by the author), molecular dynamics simulations were performed to understand if a hydrogen in this position is favorable. When not constrained to remain at the difference peak adjacent to the Oδ1 atom, the hydrogen during the course of the dynamics simulation actually exchanges to the Oδ2 atom, seemingly to take advantage of a hydrogen bond to Wat 52. If strict constraints were applied to the hydrogen to remain bound to the Oδ1 atom in order to calculate the energetics of this position, the ΔG value derived was quite positive, thus indicating that this hydrogen was most likely in an unfavorable position and orientation. Fortunately, as the refinement converged and the R_{free} approached its final value of 16.7%, this difference peak had become much weaker, even when contouring at the 2.5σ level (**Figure 4.7B**).

Full-matrix refinement to determine the protonation state of the active site carboxylate

It was then sought to determine the ionization state of the Asp 27 residue not by direct observation of electron density but by assessing the highly precise and accurate measurements for bond lengths and angles one is afforded from the ultrahigh resolution of the diffraction data. In SHELX, refinement of the aspartate and glutamate residues free of stereochemical restraints to allow unrestrained adjustment of the carboxylate groups was performed. Coincident with this step, the structure was refined against all reflections (no R_{free} set). At this point, the measurements of the carboxylate Cγ-Oδ1 and Cγ-Oδ2 bond lengths are recorded and are very accurate. Differences in these bond lengths can be used to suggest the protonation states. To determine the estimated standard deviations (ESDs) of the bond lengths and angles, the MTX model was then used in one cycle of least squares calculations for a full-matrix refinement where all the restraints were removed. This allows one to extract positional ESDs from the inverted (full) matrix for the input coordinate model (Sheldrick *et al.* 1997). Only the first 30 residues were included in the ESDs calculation. This permitted the measurement of ESDs for the carboxylate groups of Asp11, Glu17, and, most importantly, Asp27 (**Table 4.6**). Asp11 is at the end of the β1 strand and the beginning of the Met20 loop, Glu17 is in the middle "hinge" region of the Met20 loop, and Asp27 is within the αA helix and is the lone ionizable residue in the DHFR active site. Neutral carboxylates have Cγ-Oδ1 and Cγ-Oδ2 bond lengths of ~1.210Å for the C=O and ~1.310Å for the C-OH while charged carboxylates have equivalent Cγ-Oδ1 and Cγ-Oδ2 bond lengths of ~1.250Å (Deacon *et al.* 1997). Asp and Glu residues whose bond

Table 4.6: Refined carboxylate bond lengths, ESDs, isotropic B-factors, and charge estimates for Asp 11, Glu 17, and Asp 27 in monomers A and B of the 1.0Å DHFR/MTX structure

Residue	Bond length (in Å)	ESD (in Å)	B-factor (in Å^2)
Asp 11			
Monomer A			Cγ 28.0
Cγ-Oδ1	1.242	0.057	Oδ1 36.0
Cγ-Oδ2	1.254	0.049	Oδ2 39.9
Monomer B			Cγ 24.0
Cγ-Oδ1	1.244	0.045	Oδ1 34.1
Cγ-Oδ2	1.230	0.040	Oδ2 34.4
Charge state	Negative		
Glu 17			
Monomer A			Cδ 41.6
Cδ-Oε1	N/A[1]	N/A[1]	Oε1 32.4
Cδ-Oε2	N/A[1]	N/A[1]	Oε2 50.0
Monomer B			Cδ 11.6
Cδ-Oε1	1.265	0.013	Oε1 13.8
Cδ-Oε2	1.254	0.013	Oε2 14.0
Charge state	Negative		
Asp 27			
Monomer A			Cγ 8.9
Cγ-Oδ1	1.244	0.011	Oδ1 10.0
Cγ-Oδ2	1.255	0.010	Oδ2 10.7
Monomer B			Cγ 10.0
Cγ-Oδ1	1.234	0.011	Oδ1 11.3
Cγ-Oδ2	1.245	0.012	Oδ2 12.1
Charge state	Negative		

N/A[1]: Not Applicable; the carboxylate bond lengths and ESDs are not shown for the Glu 17 residue in monomer A because they are likely to be highly inaccurate due to the large thermal parameters (B-factors) associated with these atoms (far right column). Glu 17 is in the core of the Met20 loop. Not surprising, the electron density for the side chain is very weak, indicating that it is disordered. The B-factors for the Glu 17 carboxylate in monomer B (where the Met20 loop is in a closed conformation) are quite low and, therefore, there is greater confidence in the accuracy of the bond length estimates and deviations.

length differences are greater than 3 times the calculated ESD (>3σ) can be designated as protonated while bond length differences that are less than 3 times the calculated ESD (<3σ) define the residue as negatively charged (Coates *et al.* 2002). For example, the calculated bond lengths for the Asp27 carboxylate (monomer A) are nearly equal: 1.244Å for the Cγ-Oδ1 and 1.255Å for the Cγ-Oδ2 with ESDs of 0.010Å and 0.011Å, respectively. The difference in the bond lengths is 0.11Å (or 1σ), identical to the ESDs, and, thus, the conclusion is that the Asp27 carboxylate is charged. Similar results were observed for monomer B: the Cγ-Oδ1 bond length was measured to be 1.234Å while the Cγ-Oδ2 bond length was 1.245Å, with ESDs of 0.011Å and 0.012Å, respectively. The bond length difference is 0.011Å, almost identical to the average ESD (so, 1σ). It should be noted that apparent bond lengths are influenced by the displacement parameters (the B-factors) of the atoms for which the lengths are being calculated and, possibly, the overall B-factor for the protein atoms (the overall B_{iso}) (Dauter *et al.* 1997b). In fact, for similar bond length calculations of Asp and Glu residues in different inhibitor-bound endothiapepsin crystals, Coates *et al.* (2002) omitted one of the structures from the analysis because the overall B_{iso} was >20Å^2 (Coates *et al.* 2002). The Asp27 (monomer A) Cγ-Oδ1-Oδ2 B-factors in the MTX structure are 8.9, 10.0, and 10.7Å^2, respectively, while the B-factors of the same atoms in monomer B are slightly higher, at 10.0, 11.3, and 12.1 Å^2, respectively. The overall B_{iso} is 12.9Å^2. From the inference by the full-matrix refinement of the MTX X-ray structure, it seems most likely that, for both monomers in the AU, Asp 27 is negatively charged when MTX is bound to DHFR, at least at the pH (7.5) of this experiment. The ultrahigh resolution model of DHFR/MTX

reported here provided new information concerning regions of the protein with higher mobility (alternate side chain conformations; B-factor and anisotropy refinement), possible positions of several hydrogen atoms, and protonation states of carboxylate groups (using full-matrix refinement). These can all be exploited for future research directives, such as molecular dynamics and possibly even directed mutagenesis and drug design. However, direct identification of hydrogen position is lacking for most of the structure, especially at the active site near Asp 27 and MTX. For this and additional information concerning protein dynamics in the crystalline state, NC was utilized.

Neutron crystallography: Preliminary neutron diffraction studies of D_2O-soaked DHFR/MTX crystals at the ILL

In order to identify hydrogen positions within the DHFR active site and on bound methotrexate, several moderately-large (0.1-0.3 mm^3) D_2O-soaked DHFR/MTX crystals suitable for neutron diffraction analysis were grown and a partial data set was collected at room temperature on the Laue Diffractometer (LADI) instrument at the Institut Laue-Langevin (ILL). The D_2O-soaked DHFR/MTX crystal (1.4 x 1.0 x 0.3 mm, or 0.3 mm^3 after correction for the crystal's hexagonal morphology, **Figure 4.8C**) diffracted neutrons to 2.2Å resolution (**Figure 4.8A, B**). The completeness in individual resolution shells dropped to below 50% between 2.75 and 2.97Å and the I/σ in individual shells dropped to below 2 at around 2.4Å. Reflections with I/σ greater than 2 were observed beyond these limits, and the cumulative completeness of the data set is 56% at 2.2Å (**Tables 4.7** and **4.8**). The DHFR/MTX crystal belongs to a high symmetry space group ($P6_1$); consequently, our strategy involved collecting 13 images in 8° steps about the spindle

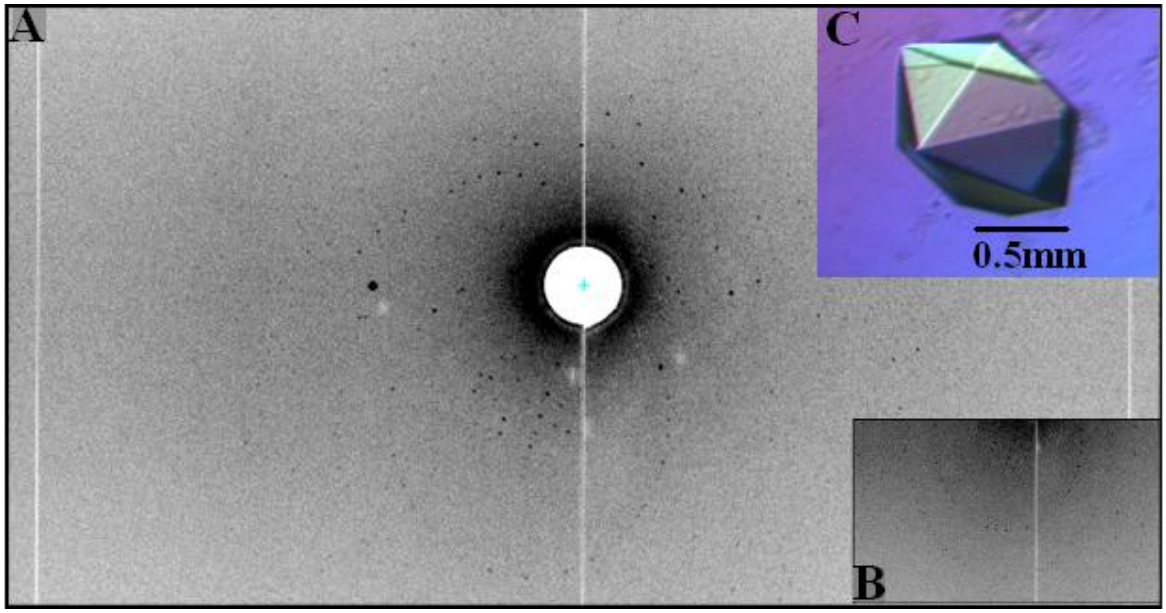


Figure 4.8: A quasi-Laue neutron diffraction image from a 0.3mm^3 D_2O-soaked DHFR/MTX crystal. A) The Laue diffraction pattern after a 34 hour exposure. The highest resolution reflections extend to 2.2Å. **B)** A magnified section of the diffraction pattern containing reflections at ~2.5 Å. **C)** The D_2O-soaked crystal's dimensions are 1.4 mm x1.0 mm x 0.3 mm, or 0.3 mm^3.

Table 4.7: Neutron diffraction data statistics: preliminary studies on a D_2O-soaked DHFR/MTX crystal at the Institut Laue-Langevin (ILL) neutron source

DHFR crystal form	MTX
Neutron facility	ILL
Instrument	LADI
Wavelength range used (Quasi-Laue)	2.80-3.70Å
Temperature (K)	293
Average exposure time	34 hours
Number of crystal settings	21
Total experiment time	30 days
Resolution range **Highest shell**	25.0-2.20Å 2.32-2.20Å
Space group	$P6_1$
Unit cell (a, b, c in Å) **(αβγ in °)**	90.93, 90.93, 72.36 90, 90, 120
Total reflections [A]	22970 (1665)
Unique reflections	8497 (852)
Completeness (%)	56.3 (56.3)
R_{symm} (%) [B]	20.0 (26.8)
I/σ	2.9 (1.5)
Multiplicity	2.7 (2.0)

[A] Numbers in parentheses correspond to values measured within the highest resolution shell

[B] $R_{symm} = \Sigma_{hkl} \Sigma_i |I_{i(hkl)} - I_{hkl}| / \Sigma_{hkl} \Sigma_i I_{(hkl)}$

Table 4.8: Data reduction statistics for the ILL preliminary study

D_{min}(Å)	R_{symm}	R_{cum}	I/σ	Mn(I)/sd	%Poss[A]	Cum%Poss[B]	Multipilcty
7.27	0.153	0.153	4.1	7.1	87.1	87.1	3.4
5.14	0.184	0.171	3.7	7.1	91.1	89.7	3.8
4.20	0.194	0.181	3.5	7.1	87.9	88.9	4.0
3.64	0.199	0.185	3.5	5.9	80.9	86.1	3.3
3.25	0.198	0.187	3.5	4.7	67.2	80.7	2.6
2.97	0.193	0.187	3.6	3.8	54.4	74.3	2.1
2.75	0.231	0.190	3.0	3.4	43.4	68.0	1.9
2.57	0.218	0.191	2.8	3.3	40.2	63.0	1.9
2.42	0.236	0.194	2.8	3.2	40.6	59.3	1.9
2.20	0.268	0.200	1.5	3.2	38.8	56.3	2.0
Overall		**0.200**	**2.9**	**4.8**	**56.3**	**56.3**	**2.7**

[A] **%Poss is the data completeness for the individual resolution shell** (To give an example, for the 3.25-2.97Å shell, the completeness is 54.4%, whereas the cumulative completeness from 25.0-2.97Å is 74.3%.)
[B] **Cum%Poss is the cumulative completeness for the data set.**

rotation axis at one crystal setting and 8 images at a second crystal orientation separated by a tilt of 20° in phi so as to record reflections in the "blind zone." The LADI instrument uses a limited quasi-Laue band pass ($\Delta\lambda/\lambda$~25%, λ=3.5 Å) to maximize the flux at the sample. As a result, reflections that are stimulated at the extremes of the wavelength range are significantly weaker than those recorded at the peak incident spectrum (**Figure 4.9**). Scale factors of up to 5 were required to normalize these reflections. The effective coverage of reciprocal space in each Laue diffraction pattern is dependent upon the wavelength range of neutrons in the incident spectrum that give rise to significantly recorded reflections. So, effective coverage is also sample dependent. Clearly, for small, weakly scattering crystals, the effective experimental wavelength range used can be narrow and the phi rotation angle between images should be reduced accordingly to ensure that the "recorded" regions of reciprocal space are truly contiguous.

To our knowledge, the primitive unit-cell volume of the DHFR crystal (5.2×10^5 Å^3; 2 molecules in the AU) is the largest so far investigated by high resolution NC. The ability to collect neutron diffraction data from a small H/D solvent exchanged crystal that is only 0.25 mm^3 in volume is due to the highly ordered lattice of the DHFR/MTX crystal. The solvent content of this crystal form is only 34.5%, allowing a more compact, ordered lattice and less diffuse scattering. Indeed, as described earlier, an ultra-high resolution X-ray dataset was collected from these crystals, indicative of tighter lattice packing and exclusion of excessive bulk solvent. However, the relatively large unit cell parameters of our crystal form (a = b = 90.93Å, c = 72.36Å) combined with the broad wavelength band pass utilized ($\Delta\lambda/\lambda$= 25%) led to some spatially overlapped neutron

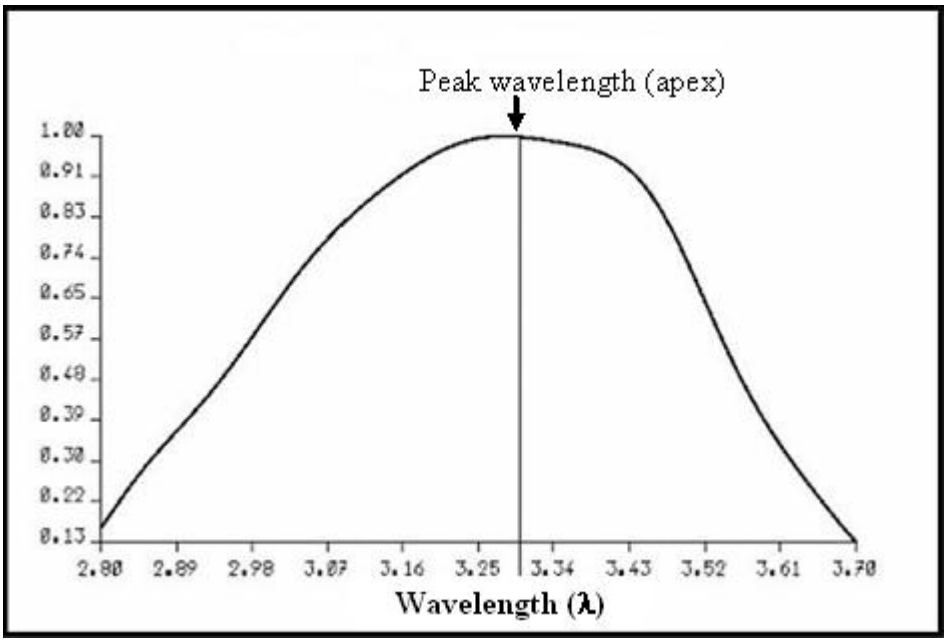


Figure 4.9: Wavelength normalization curve determined by LSCALE (Arzt *et al.* 1999) for the neutron diffraction experimental data collected at the ILL. The X-axis is the neutron wavelength used for diffraction, in Å. The Y-axis is an arbitrary scale for normalization and, thus, is dimensionless. The peak wavelength (the apex) occurs at ~3.3Å and the intensity falls off towards either extreme of the graph, at 2.8 and 3.7Å, respectively. Reflections arising from scattered neutrons at wavelengths off the apex are relatively weak, difficult to measure and integrate accurately due to the background, and led to low overall data completeness and I/σ at higher resolutions.

reflections at higher resolution from the partial data set collected at the ILL. The loss of these reflections contributes to the low data completeness, especially in the higher resolution shells.

Due to the reduced data completeness, it was questionable whether deuterium atoms can be modeled into the neutron density maps, even with the inclusion of the data that extend beyond the normal limit of around 3.0Å to 2.2Å. Previously, a neutron analysis of Con A at a similar resolution (2.75Å) revealed that some hydrogen/deuterium atoms could be modeled with confidence (Habash *et al.* 1997). More recently, the same group has collected neutron diffraction data on D_2O-soaked Con A that extends the resolution to 2.4Å, this improvement the authors attribute only to the H/D exchange method and a longer duration in which the crystal was soaked in D_2O prior to neutron data collection (Habash *et al.* 2000). The Con A data set to 2.75Å was 75.5% complete; the DHFR/MTX data set extends to higher resolution but possesses much lower overall completeness (56% to 2.2Å; 68% to 2.75Å). To reiterate from the previous chapter, the data completeness needed to be increased, the resolution extended, and the signal-to-noise ratio maximized. Two strategies were adopted to accomplish these goals: **(1)** larger crystal growth with extended D_2O-soaking and **(2)** perdeuteration. Details of **(1)** are provided in the previous chapter and immediately below while details for **(2)** are found entirely in the previous chapter.

Preparation for the latest NC experiments at the ILL and LANSCE

As stated in the chapter 3, a prerequisite for NC is the growth of large volume crystals. This makes the technique prohibitive for many systems. Using the sitting drop vapor diffusion method and a large apparatus for crystallization described in the previous

chapter, a few crystals on the order of 0.1-0.3 mm^3 volume were grown, D_2O-soaked, and one of these was used for a neutron experiment at the ILL. This required ~40-45 ml of D_2O per set-up and none of the crystals grown with this method and tested at LANSCE were very successful (one crystal diffracted neutrons to ~8Å resolution after a 24-hour exposure). An alternative crystallization method was then attempted using the microbatch-under-oil technique, employing oil as a diffusion/evaporation "regulator." This produced several crystals that were 0.2-0.4mm^3 in volume, and the requirement for the amount of D_2O that was used was reduced significantly (~400 fold; **Figure 4.10A**). To reduce the incoherent scattering from hydrogen in the neutron experiment, candidate crystals grown in the microbatch set-up were soaked extensively with D_2O-based crystallization buffer (0.1 M Na-HEPES (pH 7.5), 0.2 M $CaCl_2$, and 16% (v/v) PEG 400; both the Na-HEPES and $CaCl_2$ were added from 1 M stock solutions where the solutes had been dissolved in D_2O). Soaking the crystals should cause substitution of H for D at chemically exchangeable positions on DHFR and MTX, for example at backbone amide and side chain hydroxyl groups. To prevent shocking the crystals, the percentage of D_2O was gradually increased over several days. After about 10 days, the crystals were considered to be $\geq$ 90% exchanged and were left undisturbed in the D_2O solution for about 3 weeks. At that point, two crystals were mounted into 2.9 mm outer diameter custom quartz capillaries (**Figure 4.10B**) within a D_2O-rich environment and transported to LANSCE. A third crystal grown by microbatch-under-oil and D_2O-soaked for ~3 months was also mounted in the same way and transported to the ILL for further experiments. Crystals at both LANSCE and the ILL diffracted to high resolution (beyond 2Å) (**Figure 4.10C**). From this point, the NC results and discussion will be solely from

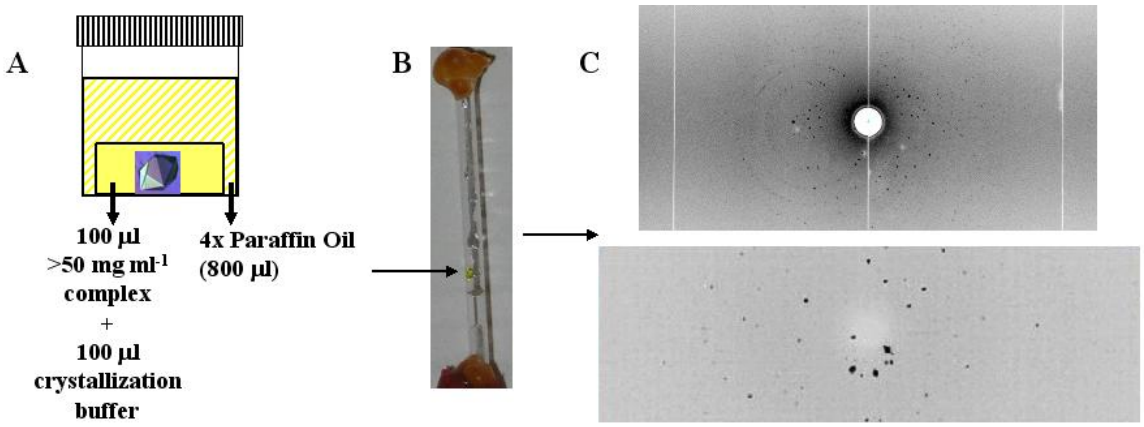


Figure 4.10: Larger volume DHFR/MTX crystals could be grown using a microbatch-under-paraffin oil crystallization technique and diffract to high resolution at two different neutron facilities. A) A schematic of the crystallization method and set-up. The protein complex solution was mixed with an equal volume of the crystallization buffer at the bottom of a crystallization well and, immediately, a 4x sample volume of paraffin oil was added onto the top of the mixture. The well was not sealed with the cap until the D_2O-soaking procedure began. **B)** D_2O-soaked crystals were mounted into large diameter quartz capillaries within a tent that had been purged with D_2O vapor by use of an LN2 tank. Two such mounts were transported to the **C)** ILL **(top)** and to PCS **(bottom)** for neutron diffraction experiments.

data collected at PCS. This is mainly due to the fact that, at the PCS, a data set could be collected from a single crystal attaining higher completeness in less than one month total experimental time.

Data collection and processing at LANSCE and refinement of the neutron structure

The crystal chosen for the room temperature neutron experiment at the PCS was (after correction for hexagonal morphology) ~0.24 mm^3 in volume, considerably smaller than most standard volumes previously used successfully in macromolecular neutron diffraction experiments. A neutron wavelength range of 0.6-7Å was employed for the pseudo-Laue experiment; due to the use of time-of-flight (TOF), reflections that are collected on the large, position-sensitive detector can be resolved in wavelength (**Figure 4.10C**, bottom image). Fifteen crystal settings were used with an average of 36 hours exposure per setting. To complete data collection, the experiment required 22 days.

Detailed aspects of processing wavelength-resolved Laue data collected at the PCS can be found in (Langan *et al.* 2004b) and also in chapter 2; only reflections recorded arising from a wavelength range of 1-5Å were processed. The previously deposited RT X-ray coordinate file (3DRC) (Warren *et al.* 1991) was used as the starting model in SHELX (Sheldrick *et al.* 1997) for restrained positional, isotropic B-factor and occupancy refinement against the merged neutron reflection data set; initial rounds of refinement reduced the R_{free} to 39.5%. The resolution range used for initial refinement was 8.0 – 2.5 Å; after several rounds, the high resolution limit was eventually set to 2.2Å. After repeated cycles of refinement and inclusion of several D atoms into exchangeable positions of the protein model and D_2Os into the solvent model, the R_{free} stalled at 36.7%. It was suggested that, due to the size and the space group of the DHFR/MTX crystals,

there was a possibility of physical twinning. Twinning is a crystal growth disorder where one crystal is composed of multiple physical domains whose orientations differ in a specific way (Yeates 1997). A merohedral twin has ≥ 2 real space lattices that coincide exactly in 3 dimensions, causing perfect overlap of reciprocal lattices and, thus, perfect overlap in the diffraction pattern. However, each observed intensity arises from contributions from reflections unrelated by crystallographic symmetry. In order to further successfully refine the neutron structure, a correction for merohedral twinning had to be applied. Incorporation of this twinning correction (a merohedral twin law; see chapter 2) in the SHELX instruction file significantly reduced the R_{free} to ~26%. The total scattering contribution from the twin domain approaches 40%. Data processing and refinement statistics are listed in **Table 4.9**.

Backbone H/D exchange assessed from occupancy refinement and nuclear density maps

From initial inspection of neutron maps (**top left and right**, **Figure 4.11**), there was significant nuclear density for the main chain and many of the side chain atoms, indicating a satisfactory fit between model and data. Additionally, it appeared that there was substantial nuclear density adjacent to and around many of the backbone amide nitrogen atoms (at this point, H were modeled as bound to these nitrogens) (**bottom left**, **Figure 4.11**). In these regions, H/D exchange at backbone amide nitrogens could explain the nuclear density due to the D nuclei's strong positive scattering of neutrons (recall that H nuclei are incoherent negative neutron scatterers). Compared to other groups which may undergo H/D exchange in proteins, amide nitrogens exchange at a fairly slow (and convenient) rate (~0.5 min^{-1} dependent on pH, temperature, and position in the protein), thus, the monitoring of their exchange is amenable to biophysical study and interpretation

Table 4.9: Neutron data statistics from PCS for the DHFR/MTX cocrystal

DHFR crystal form	MTX
Neutron facility	LANSCE
Beam line	PCS
Unit cell (a, b, c in Å)	90.93, 90.93, 72.20
(αβγ in °)	90, 90, 120
Molecules/ A.U.	2
Space group	$P6_1$
Resolution range (Å)	38.6-2.17
Highest shell	2.29-2.17
Unique reflections	14213 (1649)
Completeness (%)	79.7 (63.7)
R_{merge}[1] (%)	7.0 (32.5)
I/σ	3.7 (1.7)
Multiplicity	2.9 (1.9)
R_{fac}[2]/R_{free}[3] (%) (For refinement resolution range 6.5-2.2Å)	24.0/25.9

R_{merge}[1]$= \Sigma_{hkl} \Sigma_i |I_{i(hkl)} - I_{hkl}|/\Sigma_{hkl} \Sigma_i I_{(hkl)}$
R_{fac}[2]$= \Sigma_{hkl} | |\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}| | / \Sigma_{hkl} |\boldsymbol{F}_{obs}|$
R_{free}[3]$= \Sigma_{hkl} | |\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}| | / \Sigma_{hkl} |\boldsymbol{F}_{obs}|$
(R_{factor} calculated for 5% of the data sequestered from the refinement.)

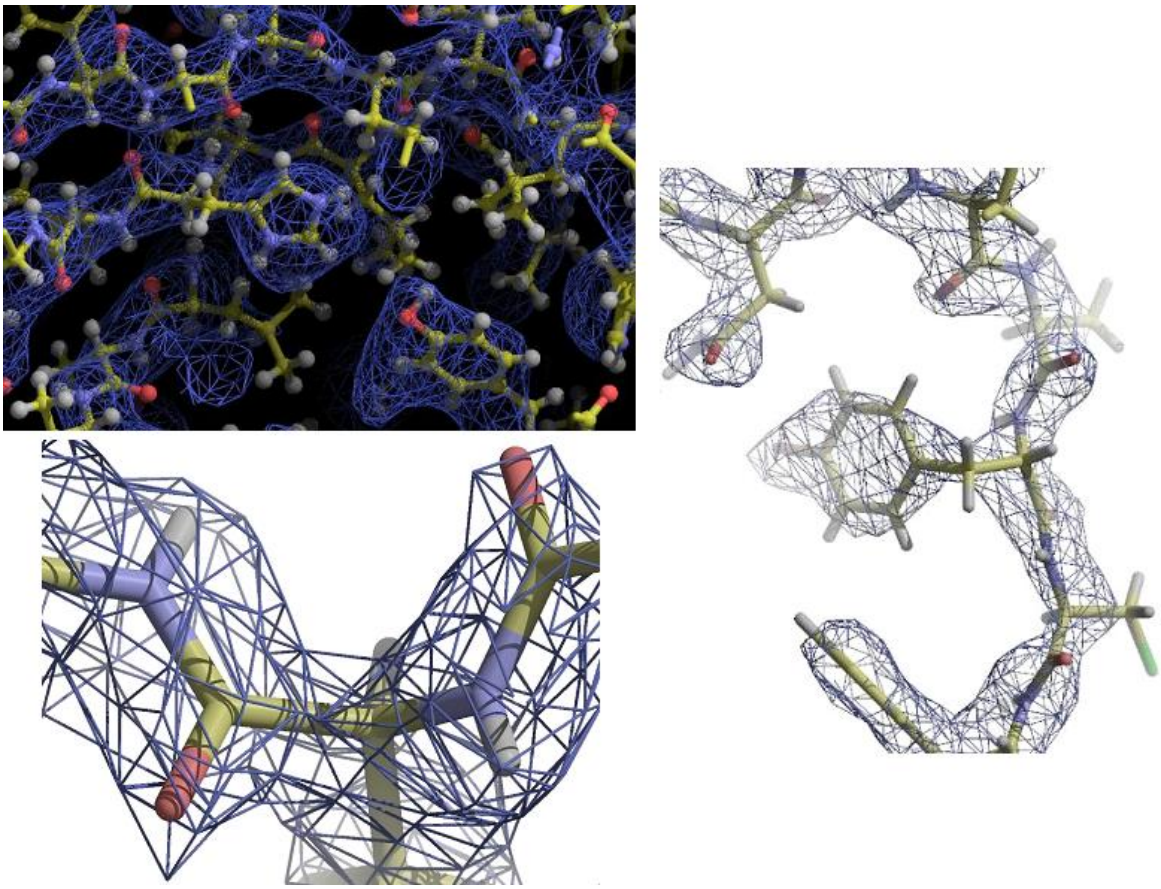


Figure 4.11: The DHFR/MTX nuclear density maps and assessment of H/D exchange at backbone amide nitrogens. 2***F***o-***F***c nuclear density map contoured at 1.5σ at the **(top left)** βF region of monomer A and the **(right)** βF region of monomer B, at Tyr 150-Cys 151-Phe 152. Note the continuous density for the main chain atoms and many of the side chains. Aliphatic carbons are especially difficult to visualize most likely due to H negative scattering cancellation of carbon neutron scattering. Observance of density in the maps and refinement of D occupancy values suggests the amide nitrogens at Ile 94-Gly 95 **(bottom left)** have exchanged H for D.

(Creighton 1993) (Korszun 1997). Also, it appears that the time scale of exchange (the exchange kinetics) is very similar for proteins in solution compared to in the crystalline state (Schoenborn *et al.* 1996). The occupancy refinement of D at these positions was performed in SHELX; essentially, the positional coordinates of the D atoms are constrained to be fixed and the probability that a D atom is located there is refined (measured as a fractional value from 0.0 to 1.0 but commonly communicated as a percentage, 0-100% occupied). Korszun (1997) defines three categories of H/D exchange occupancies: 0-15% is unexchanged, 15 to 60% is partially exchanged, and $\geq$60% is fully exchanged (Korszun 1997). Refined occupancy values were used as an initial indicator of amide H/D exchange. To strengthen evidence that an actual exchange event occurred, nuclear density maps were inspected $\geq 1.5\sigma$ along the backbone. Nearly 80% of refined occupancy values seemed to agree with what could be directly observed in the maps (see **Figure 4.11**); where little to no nuclear density existed (even at a 1.0σ contour level) at these positions adjacent to amide nitrogens often correlated to refined occupancy values that were quite low (<0.20 or <20% occupied)**.** The disagreement between the refined occupancy value for an amide D atom and what was observed in the map was ~18%. This was measured by taking the refined D occupancy values and assessing the nuclear density maps at each of the refined positions. If an amide D occupancy value was $\geq$0.2 but no density existed for the D atom, then it was identified as a disagreement. Due possibly to cancellation effects (Kurihara *et al.* 2004) of the H negative scattering, at a few positions, the amide nitrogen itself could not be accounted for in the neutron maps.

Including both monomers, there are 289 amide hydrogens possible to exchange. Of these, 187 did exchange for deuterium, or about 65%. However, monomer B (101 of

146, ~69%) possesses a higher number of amides which have exchanged over monomer A (86 of 143, ~60%). Both monomers have only a few amides which have exchanged in the βB to αC region (residues 36-50) and, surprisingly, in the loop segment between βC to βD (residues 64-72) on the surface of the protein. However, the major region of exchange differences between the two monomers occurs from near the end of the βF-βG loop (~residue 128) to the C-terminus (a segment of 31 residues), with 15 amides not exchanged in monomer A compared to only 3 amides that have not exchanged in monomer B. To reiterate, the major structural differences between the two monomers is at the Met20 (in A, this is partially occluded; in B, this is closed) and the βF-βG loops. A comparison of backbone amide D occupancies to main chain isotropic B-factor and anisotropy values (from the 1.0Å X-ray data) is shown in **Figure 4.12**. There is good agreement in the two monomers between regions of higher H/D exchange, higher B-factors, and higher anisotropy (lower *A* ratios indicates higher anisotropic motion). Models color-coded by amide H/D exchange (and based on both occupancy refinement and map observations) are shown in **Figures 4.13** and **4.14**. A major area of difference for H/D exchange between the A (Met20 occluded) and B (Met20 closed) monomers is in the C-terminal region, from the end of the βF-βG loop through the βG-βH loop and on to the C-terminus (residues 132-159), where the B monomer has significantly more amides that have exchanged. Even the majority of the amides *within* the final 2 β strands (G, H) have exchanged. This may be explained by the fact that in the B monomer the core of the Met20 loop rearranges so that Gly 15-Met 16-Glu 17 have inverted and moved away from the substrate and cofactor (nicotinamide) binding site and part of the core of the βF-βG loop (Gly 121-Asp 122) is, in turn, now in position to form additional hydrogen

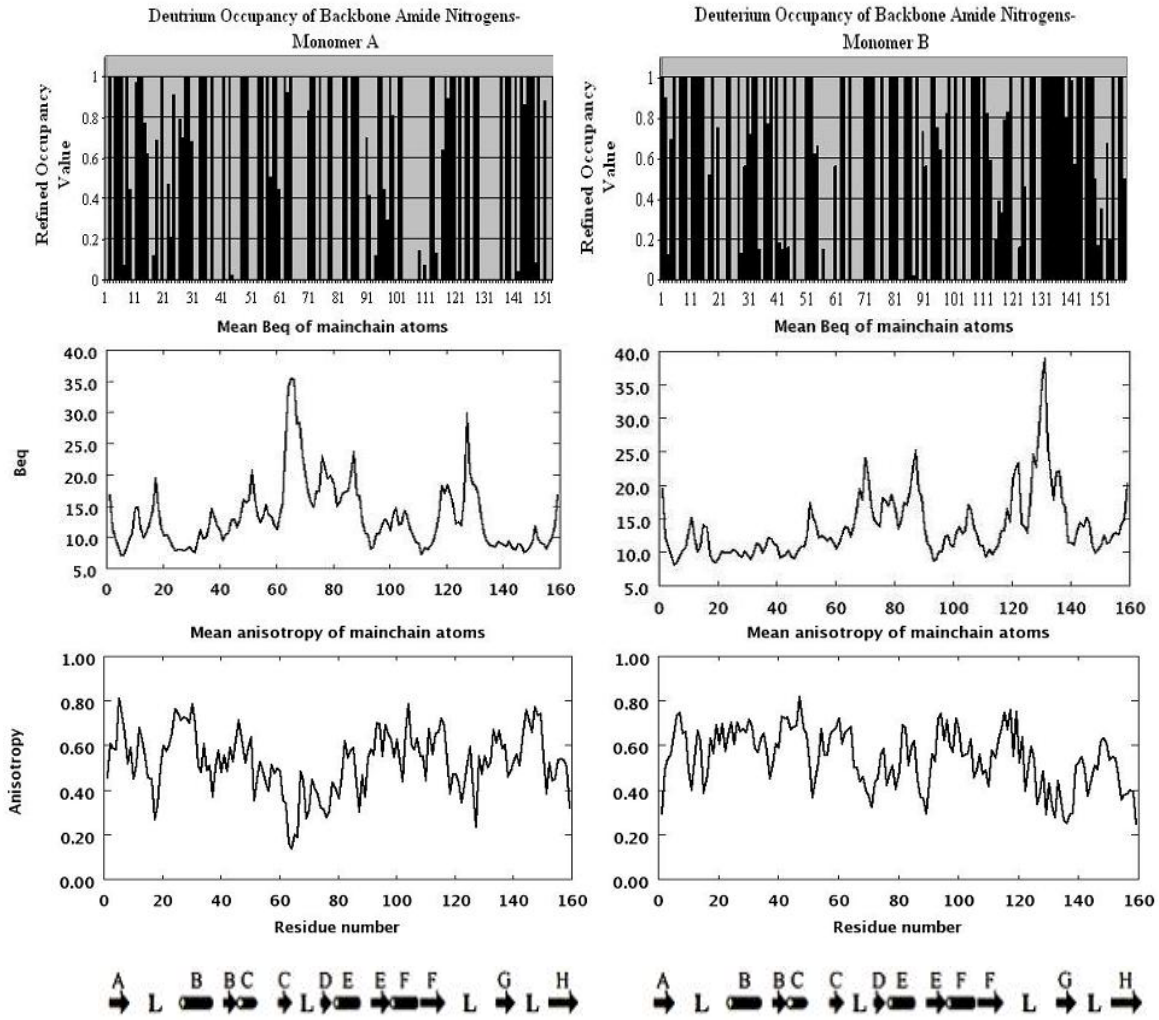


Figure 4.12: Comparison of backbone amide H/D exchange, isotropic B-factors, and main chain anisotropy for DHFR/MTX. The X-axis for all graphs is plotted as residue number. The left-hand panels are for monomer A while the right-hand panels are for monomer B. **Top panels)** Occupancy values for D atoms bound to backbone amide nitrogens after refinement in SHELX and inspection of nuclear density maps. Values on the Y-axis range from 0.0 to 1.0; essentially, occupancy is a probability index that a given atom (with associated scattering factor) is at the *xyz* position and possesses the particular B-factor it has been ascribed. **Middle panels)** Isotropic B-factors (in $Å^2$) for main chain atoms. **Bottom panels)** Anisotropy ratio values (*A*) for main chain atoms calculated with the PARVATI server (Merritt 1999a); the lower the value for *A*, the more non-spherical (or ellipsoidal) the vibrational behavior is for the corresponding atom. Below the anisotropy plots are secondary structure cartoons for DHFR. The arrows indicate β-strands (A-H), hollow tubes indicate α-helices (B, C, E, & F), and L indicates loops (Met20, adenosine-binding, F-G, and G-H). Generally, there is a correlation for residues which have amides that have undergone H/D exchange; for the most part, they also possess higher B-factors and lower *A* ratios.

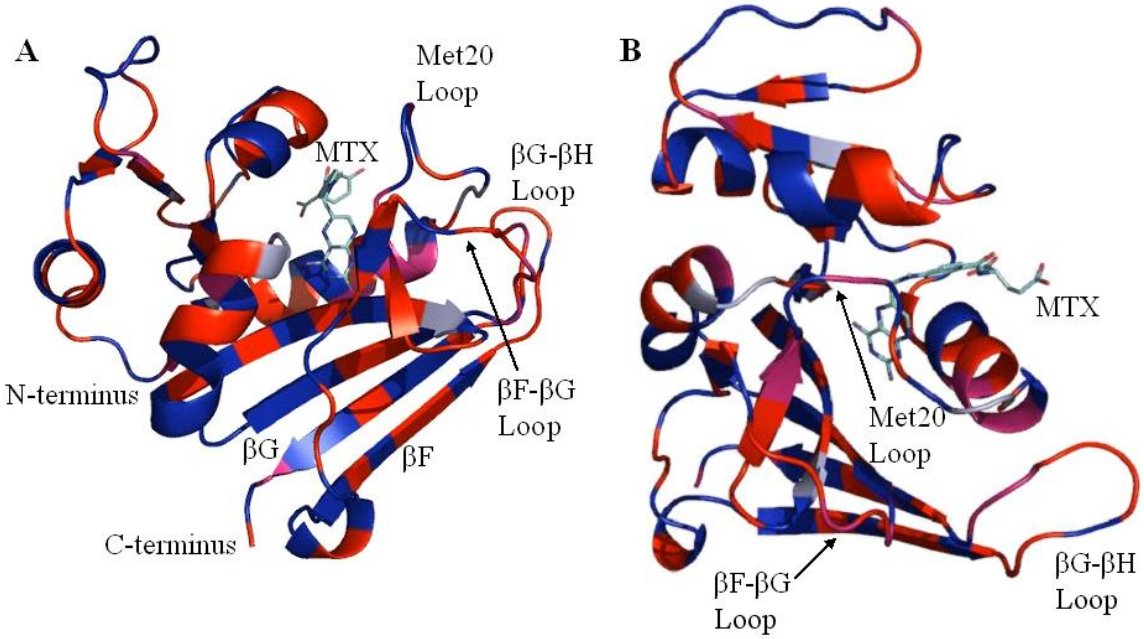


Figure 4.13: Backbone amide H/D exchange for DHFR/MTX monomer A deduced from NC. Exchange is determined by refined occupancy values for D bound to backbone amide nitrogens and by sufficient $2\mathbf{F}_o$-$\mathbf{F}_c$ nuclear density ($\geq$1.5σ) for these D atoms. Essentially, the higher the refined D occupancy value means the greater the probability that D has exchanged for H on a particular amide nitrogen. Different levels of D occupancy are indicated by color-coding; Blue = <20% occupancy, light Blue = 20-49% occupancy, Pink/Light Purple = 50-79% occupancy, and Red = >80% occupancy. Proline residues are colored in blue (There are 10 in each monomer). The A monomer has 6 residues which are disordered at the main chain and, thus, an exchange assignment isn't possible. The B monomer has 3 residues which are disordered to this extent. These residues are also shown in blue. **A:** Side view looking straight into the active site; **B:** Front view of the occluded Met20 loop partially obstructing the active site.

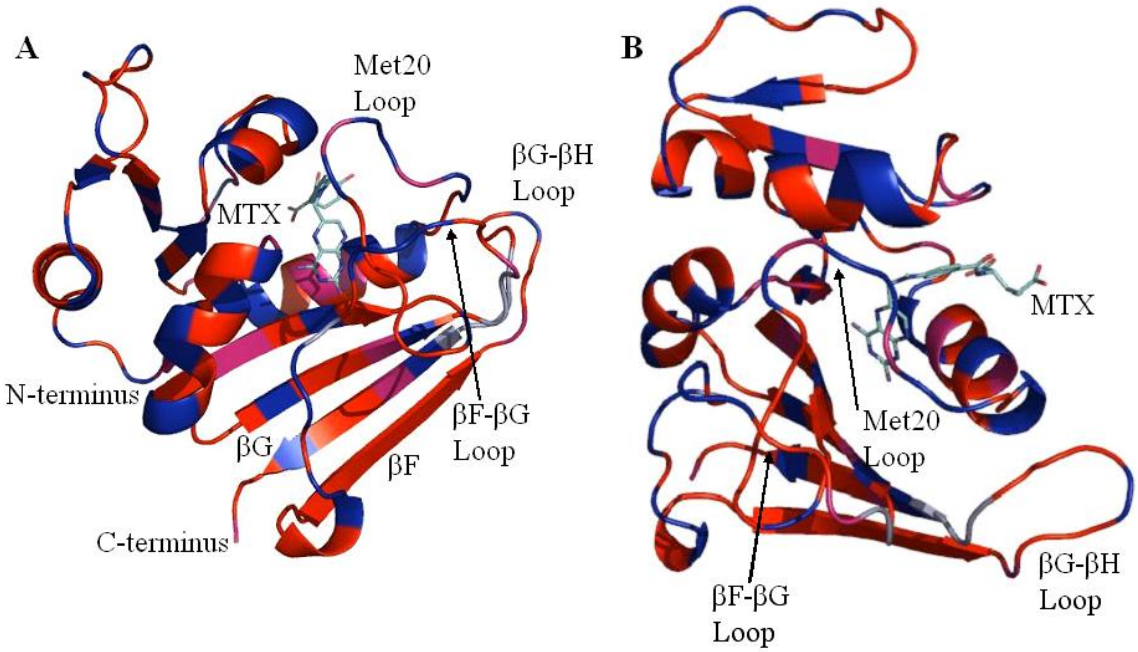


Figure 4.14: Backbone amide H/D exchange for DHFR/MTX monomer B. The color scheme and orientation of the model are as in Figure 4.13. Note the higher prevalence of deuterium exchange for β-strands G and H and at the C-terminus compared to monomer A. However, less exchange has occurred for most of the βF-βG loop (especially residues 121-127), possibly due to increased hydrogen-bonding with the closed Met20 loop.

bonding contacts with this part of the Met20 loop. For the amide nitrogens (121-123) in this section of the βF-βG loop, little to no exchange has occurred; additionally, 2 of the amide nitrogens (His 124, Phe 125) possess weak 1σ 2**F**o-**F**c density for D atoms which have correspondingly low refined occupancy values (<40%). This would be expected due to the enhanced interactions with and stabilization of the Met20 loop. In other words, H/D exchange at the backbone amide groups is hindered due to interloop packing and hydrogen bonding. Indeed, the refined occupancies and the nuclear density suggest that no exchange has taken place at Gly 121 or Asp 122 in the B monomer; backbone atoms of these two residues form hydrogen bonds with Gly 15 and Glu 17 in a closed loop conformation. Additionally, every B-factor for the C-terminal residues (from the end of the βF-βG loop to the terminus; residues 132-159) of the B monomer are higher than in the A monomer, indicating enhanced mobility.

In contrast, the A monomer has 2 more exchanged amides in the beginning and the core of the βF-βG loop (residues 116-128) than does the B monomer. For residues 116-119 and 125, both monomers' amides have exchanged H for D but the refined occupancy values for the A monomer are greater. However, from the end of this loop and throughout the C-terminus, H/D exchange is severely reduced compared to monomer B. This could be due to the necessary stabilization of the occluded Met20 loop by this region and especially the βG-βH loop (residues 142-150); two hydrogen bonds form between Asn 23 (backbone amide N and backbone carbonyl O) and Ser 148 (Oγ and amide N). In the A monomer, both of these residues have very low refined D occupancies for their backbone amides (Asn 23 = 0.2; Ser 148 = 0.1), indicating little to no H/D exchange. The βG-βH loop does not change position when the Met20 loop moves from occluded to

closed, but, when occluded, the βF-βG loop is not positioned close enough to form hydrogen bonds with the Met20 loop. Thus, the βG-βH loop is the *main stabilizer* of the Met20 loop when it adopts an occluded conformation. The Asn 23 and Ser 148 amides have exchanged in the B monomer, suggesting that the backbone hydrogen-bonding that helps to secure the Met20 loop when it is in an occluded conformation is not needed in the closed conformation. This agrees with earlier observations made by Sawaya and Kraut (Sawaya *et al.* 1997) and with the higher B-factors observed for this region of the B monomer reported here.

One other region of major difference for H/D exchange between the two monomers is at the adenosine-binding loop (residues 62-72); for monomer A, this region has higher B-factors, solvent accessibility, and anisotropy compared to monomer B yet it appears that both possess near-equivalent exchange patterns. However, for two of the residues in monomer A (Gly 67-Thr 68), there is not sufficient density to model all of the main chain atoms, much less H/D bound to amides. This region of monomer A is also somewhat disordered in the 1.0Å X-ray structure, with weak electron density for these residues; monomer B does not suffer from this. Since this region has been found disordered in DHFR/MTX occluded loop structures at different temperatures, it can be suggested that this region is inherently dynamic, possibly due to the fact that no cofactor is bound (Bolin *et al.* 1982) (Warren *et al.* 1991).

Relevance of the H/D exchange results to previous reports on DHFR "dynamics"

It is beneficial that the DHFR/MTX monomers are not equivalent in loop positioning; this way, a comparison can be made of the high resolution structures to the available solution data describing the these conformational differences. Several recent

NMR studies have indicated that the dynamics of many regions of DHFR can be correlated to the conformation of the Met20 loop and the presence of ligand bound at the active site. Examining the amplitude of internal motions (called S^2) (Schnell *et al.* 2004), changes in ^{1}H and ^{15}N chemical shifts from selectively-labeled ^{15}N-alanine (Venkitakrishnan *et al.* 2004), and difference ^{15}N chemical shifts from relaxation dispersion experiments (McElheny *et al.* 2005), the regions exhibiting the largest change in motion and chemical environment from an occluded to a closed Met20 loop conformation, of course other than residues in the Met20 loop itself, are residues in the βF-βG and βG-βH loops, as well as a few of the active site residues such as Ala 7 which shifts due to the presence of cofactor and co-requisite movement of the Met20 loop to become closed (Venkitakrishnan *et al.* 2004). Upon this conformational change (occluded to closed), the S^2 order parameter increases for the Met20 and slightly for the βF-βG loop (so, these regions are becoming more ordered and tumbling less) while the ^{15}N chemical shifts change most dramatically for the Met20, βF-βG, and βG-βH loops. As observed from the X-ray and NC data presented here, this seems to correlate to the "dynamics" of these same regions as revealed by the B-factor and anisotropy analysis paired with H/D exchange patterns for the two monomers. The enhanced motion of the loop regions and especially change in the Met20 loop conformation (the conformational exchange rate is 2-40 s^{-1}) is relevant to ligand binding and product release processes and is important for the catalytic mechanism (Osborne *et al.* 2001). The transition of the Met20 loop from occluded to closed seems to facilitate binding of the reduced cofactor nicotinamide ring prior to catalysis. Once the chemical step has occurred, the Met20 loop rearranges to occlude the oxidized nicotinamide ring from the active site, flipping it out towards the

surface (Sawaya *et al.* 1997). These movements and conformers must be facilitated and stabilized mainly by the regulatory adenosine, βF-βG, and βG-βH loop regions. Indeed, mutation of Gly 121 and Asp 122 disrupts a hydrogen bonding network that forms between the βF-βG and the Met20 loops and has severe effects on hydride transfer and, thus, the catalytic rate (Cameron *et al.* 1997) (Miller *et al.* 1998b) (Miller *et al.* 2001).

H/D exchange for backbone amides on surface-exposed loops and even on helices or sheets can be understood simply by positing that there is an enhanced accessibility of these residues to solvent, thus increasing the chances of exchange if D_2O is incorporated into the buffer solution that the crystal is stored. Possibly, the inability of certain residues to exchange could be due to crystal contacts, interactions and residue packing at interfaces between molecules in the AU and unit cell. What is difficult to explain is the mechanism by which H/D exchange can occur for some of the interior backbone amide nitrogens and even on side chains; generally there seems to be two hypotheses for the process, one invoking protein "breathing" or local unfolding of the polypeptide and the other invoking the diffusive capability of solvent to access core regions of the protein (Creighton 1993) (Korszun 1997). Conclusions made from early NC studies (Kossiakoff 1982) (Wlodawer *et al.* 1983) supported solvent permeation as the mechanism for H/D exchange at protein interiors, citing that unfolding should be highly impermissible due to the crystalline lattice. Even at the low solvent content of the DHFR/MTX crystals (~35%), perhaps the time course over which the D_2O-soaking and the diffraction experiment itself takes place (weeks) allows exhaustive permeation of the solvent throughout the available channels within the crystal. MALDI-TOF mass spectrometric (MS) analysis of proteolytic fragments of DHFR which had been subjected to H/D

exchange after proteolytic digestion revealed rapid exchange for the Met20 loop fragment and residues near to it. However, labile hydrogens within β-strands, even upon digestion into fragments, were very slow to exchange (Yamamoto *et al.* 2004a). Due to differences in the exchange kinetics and patterns they observed from intact protein compared to fragments, this study also indicated that local structural elements differentially influence the extent of exchange and are varied in their contributions to the overall fluctuation of the whole protein. The H/D exchange process is in solution for the MS studies whereas for the NC study here it is performed in the context of the crystalline lattice. Additionally, there are two monomers in the AU of the DHFR/MTX crystal and, as described above, differences in H/D exchange patterns are apparent and probably only due to loop conformational differences between the two monomers. In one of the MS experiments, it was observed that the incorporation of D into DHFR bound to MTX or folate along a time course was biphasic, with a rapid exchange step preceding a deceleration of the H/D exchange rate (Yamamoto *et al.* 2004b). This could correlate to the conformational isoforms observed for DHFR from crystallography, NMR (Schnell *et al.* 2004), and recently by single molecule fluorescence studies (Rajagopalan *et al.* 2002), further arguing that DHFR loops play important regulatory roles in ligand binding. Another important difference in the two studies is that for the MS experiments the protein sample was incubated in D_2O solution for ~10 minutes prior to preparing the MALDI plate for MS analysis. For the NC experiment at PCS, the DHFR/MTX crystal was soaked in $\geq$90% D_2O-containing buffer for ~3 weeks prior to mounting into the capillary for the diffraction experiment.

Correlation between solvent accessibility and H/D exchange

Solvent accessibility is measured by using a probe molecule with a radius of 1.4Å (equal to the radius of a water molecule) to roll upon the surface of the atoms in residues in a three-dimensional model to determine the accessibility of the probe to the different atoms. Absolute solvent accessibility plots (**Figure 4.15**) for the two DHFR/MTX monomers roughly correlate to amide exchange patterns. For example, the refined amide D occupancy values for Ser 64 are 1.0 (A) and 0.3 (B) whereas their solvent accessible volumes are 101Å^3 (A) and 77Å^3 (B). This is indicated for the red asterisks in **Figure 4.15**. Upon inspection of the neutron maps, the density suggested that a D atom is bound to the Ser 64 amide in monomer A but not for monomer B. Other accessibility/exchange correlative examples are provided in the figure. Sixty-one percent of the C-terminal region of the B monomer (residues 132-150; 17 of 28 residues) is more solvent accessible than the A monomer. Exactly 40% more of the backbone amides within this region have exchanged in monomer B (25 of 28 residues) than in monomer A (only 14 of 28 residues). Although there is a correlation between HD exchange and accessibility to solvent, this analysis excludes symmetry contacts as a determinant of accessibility. It is worthwhile to incorporate that aspect into any future analyses and its correlation to NC data.

H/D exchange on side chains and identification of D_2O molecules within the maps

Occupancy refinement of D at exchangeable positions on the functional groups of Arg, Lys, and His was also performed. For 13 of 17 Arg residues, 8 of 11 Lys residues, and 6 of 10 His residues, partial or full H/D exchange has occurred (see **Figure 4.16 A-D** for examples). The solvent molecules from 3DRC were displayed with the neutron maps.

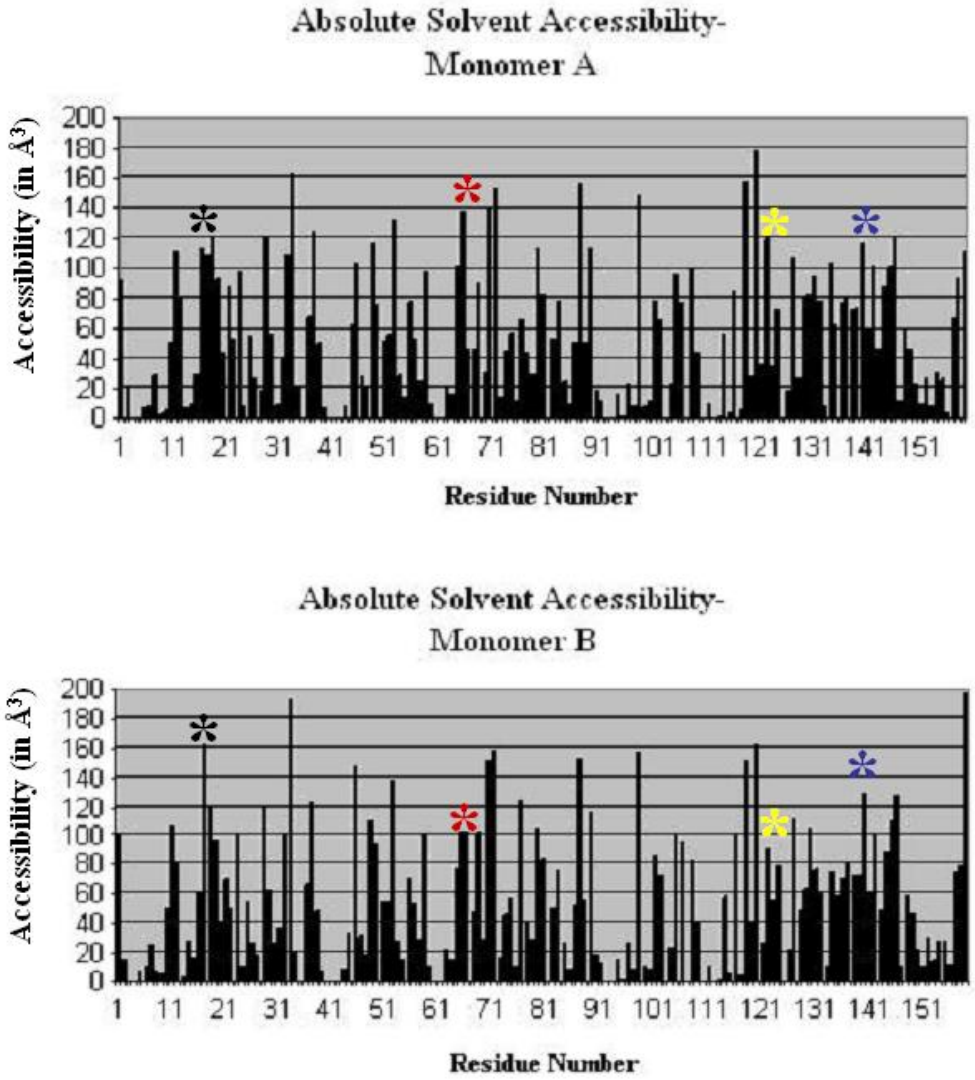


Figure 4.15: Absolute solvent accessibility (ASA) values for all residues in DHFR/MTX monomers A and B. Accessibility values (provided as surface area values) were calculated by the program NACCESS (Hubbard *et al.* 1996) for **(top)** Monomer A and **(bottom)** Monomer B. Indicated by color-coded asterisks are example residues where there are significant differences in accessibility volumes for the same residue between the two monomers and can be correlated to H/D exchange.
Black: Glu 17(A): ASA = 108Å^3, D occupancy = 12.1%, Map = H; Glu 17 (B): ASA = 162Å^3, D occupancy = 52.4%, Map = D.
Red: Ser 64 (A): ASA = 101Å^3, D occupancy = 1.0, Map = D; Ser 64 (B): ASA = 77 Å^3, D occupancy = 30.2%, Map = H.
Yellow: Asp 122 (A): ASA = 119Å^3, D occupancy = 1.0, Map = D; Asp 122 (B): 91Å^3, D occupancy = 53.3%, Map = H.
Blue: Phe 140 (A): ASA = 117Å^3, D occupancy = 0.0, Map = H; Phe 140 (B): ASA = 129Å^3, D occupancy = 1.0, Map = D.

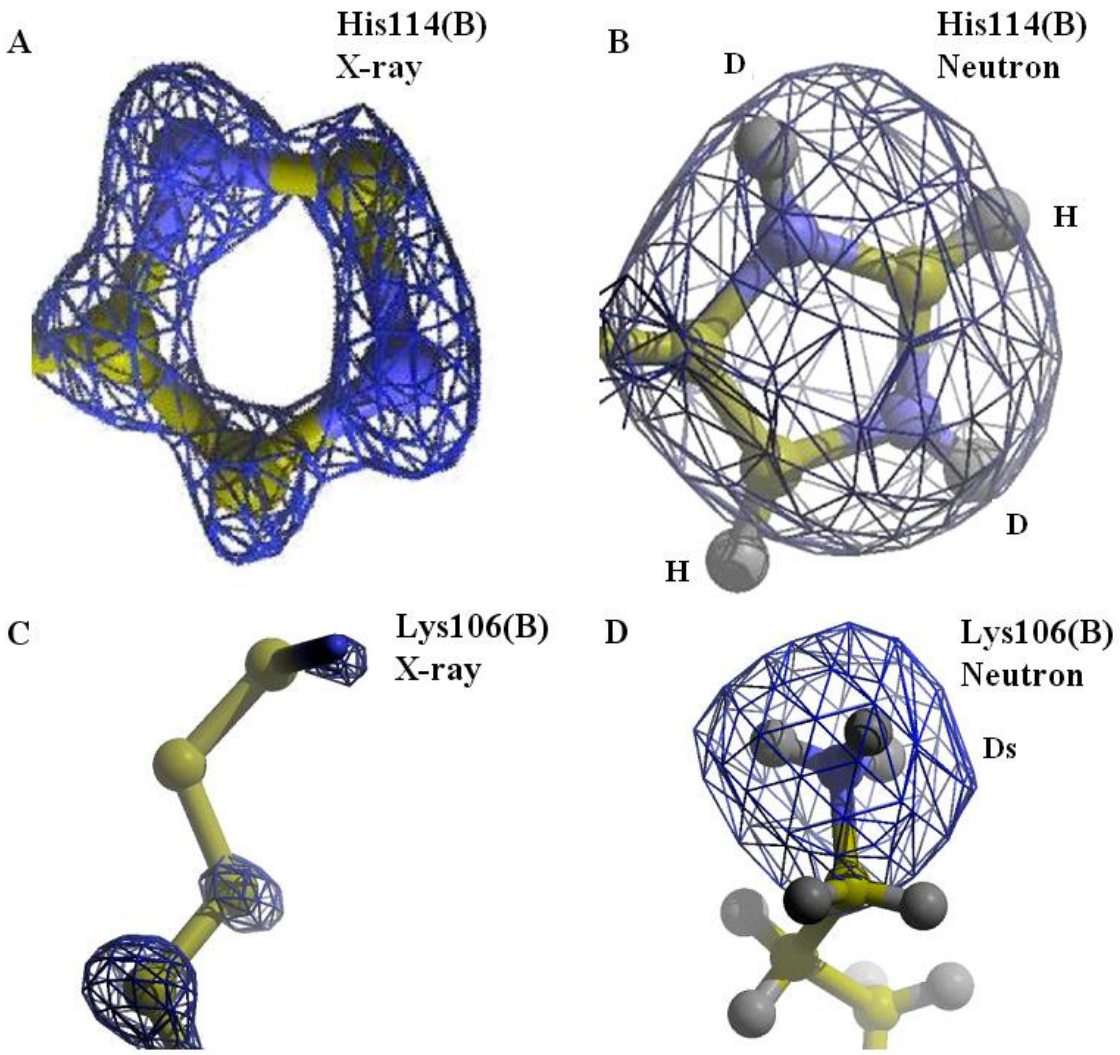


Figure 4.16: Determination of H/D exchange on the side chains of His 114 and Lys 106 (both in monomer B) from neutron density maps of DHFR/MTX and comparison to atomic resolution electron density. In blue are $2\boldsymbol{F}_o$- $\boldsymbol{F}_c$ nuclear or electron density maps contoured at 1.5σ. **A)** The imidazole ring of His 114 as modeled into 1Å electron density maps and into **B**) 2.2Å nuclear density maps. At this resolution, D is able to be modeled at chemically exchangeable positions with the neutron data whereas, with the X-ray data, it is not known whether hydrogens are bound to the Nε and Nδ atoms on the imidazole ring, thus, it is not known what the charge state of the His is. **C)** The terminal amine group of Lys 106 as modeled into 1Å electron density and into **D)** 2.2Å nuclear density maps. In the neutron maps, it is clear that D has exchanged for H at all three positions of the amine Nε atom.

There have been 18 D_2Os so far identified. These are mainly located at primary protein/solvent interfaces and more towards the interior of the protein; the stipulations for modeling D_2O into a region of unmodeled $2\mathbf{F}_o$-$\mathbf{F}_c$ or $\mathbf{F}_o$-$\mathbf{F}_c$ density is if there is sufficient nuclear density $\geq 1.5\sigma$ for all three atoms and also if at least 1 hydrogen bond donor/acceptor atom is within the proper distance (1.8-2.5Å depending on the atoms involved in the contact; **Figure 4.17 A, B**). Oxygen has been added to the model to represent 66 H_2O molecules. Again, H scatters neutrons incoherently, meaning that H_2O may very well be unobservable in neutron maps at this resolution. So, there is a possibility that these could represent oxygens in disordered D_2O molecules (Chatake *et al.* 2003). Many of the D_2Os (12) and oxygens (57) modeled as solvent molecules were initially oxygen atoms for H_2Os in 3DRC and needed only minor movement to place them into the nuclear density.

Probing the Asp 27•MTX interaction: Occupancy refinement and density analysis

Based on the observation of D atoms in many regions of the protein directly from the neutron maps, in both monomers a D atom was restrained to be 1.0Å from the N1 atom on MTX and its occupancy refined in SHELX. Restraints for a D at this position had to be manually added into the refinement instruction file. This part of the pteridine ring of MTX is very similar to part of a His imidazole ring at the Nε-Cδ-Nδ bonds. The bond angle and distance restraints for Ds bound to the Nε and Nδ atoms of a His imidazole ring were used as starting values for restraining the D bound to the N1 of MTX. The refined occupancy of this D atom was ~1.0 (or 100% occupancy) in monomer A and ~0.75 (75% occupancy) in monomer B. The distance between the N1 and the D

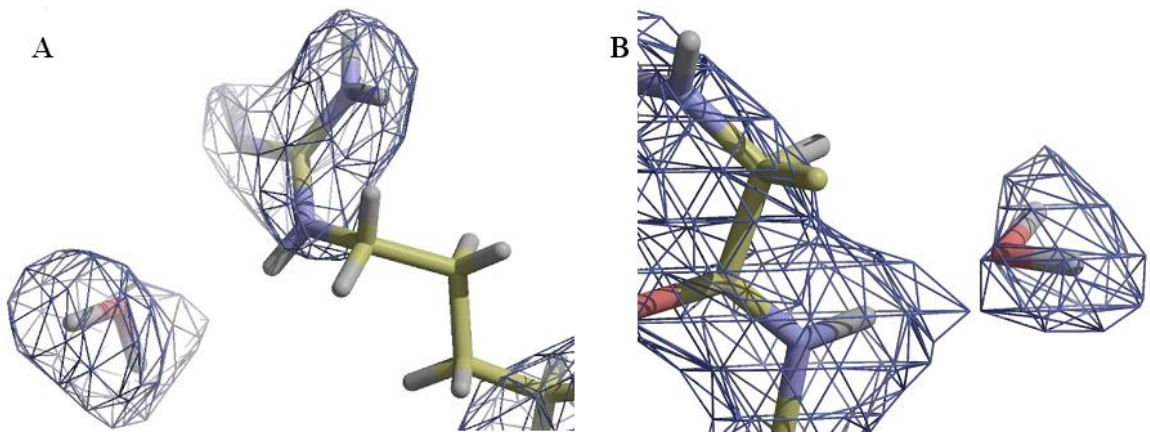


Figure 4.17: D_2O molecules identified in the neutron maps. $2\boldsymbol{F}_o$- $\boldsymbol{F}_c$ nuclear density maps contoured at 1.5σ are shown in blue. Eighteen D_2Os have been modeled into the neutron maps so far. **A)** The oxygen atom in D_2O 12 is ~3.2Å from the Arg 33 Nε atom (monomer B). **B)** The oxygen atom in D_2O 9 is ~1.9Å from the D on the backbone amide of Met 20 (monomer B).

atom refined to 1.00-1.02Å whereas the distance between the D atom and the Oδ2 atom of Asp 27 is 2.8-2.9Å. For monomer A, the N1, D, and Oδ2 atoms all are positioned on a near-perfect plane; the angle is 174°. (The same atoms in monomer B slightly deviate from this.) In addition to the occupancy values, the nuclear density around MTX and Asp 27 in both monomers (**Figure 4.18 A, B**) highly suggests that a D atom can be bound to the N1. At the same contour level (≥1.5σ), a D atom could not be constrained to be bound to the Oδ2 atom of Asp 27 and fit into the nuclear density. Put together, this would suggest that the N1 is protonated when MTX is bound to DHFR, imparting it with a positive charge. Based on the direct observation by NC of the D atom bound to the MTX N1 and the inference from the X-ray structure that the Asp 27 carboxylate is ionized, we propose that the Asp27•MTX interaction is ionic in nature. This is in agreement with nearly all of the available biochemical and complementary structural data (or suggestions made from analysis of the data) for at least *E. coli* and *L. casei* DHFRs (Coco *et al.* 1981) (Bolin *et al.* 1982; Coco *et al.* 1983; London *et al.* 1986; Stone *et al.* 1983b) (Appleman *et al.* 1988) (Gargaro *et al.* 1998). Evidence that this interaction was not ionic in nature but a neutral dipole-dipole was gleaned from a computational study measuring proton affinity for the Asp 27 carboxylate oxygen and the MTX N1 atoms in low dielectric environments (ε<10). This environment has been theorized to represent the interior of a protein, where there is normally an increase in hydrophobicity. It was found in this study that, at ε values below 10, the Asp 27 carboxylate would be protonated and the MTX N1 would not, making both these groups neutral (Cannon *et al.* 1997a). From the X-ray and the NC data reported here, it was observed that, at pH = 7.5 and regardless of the Met20

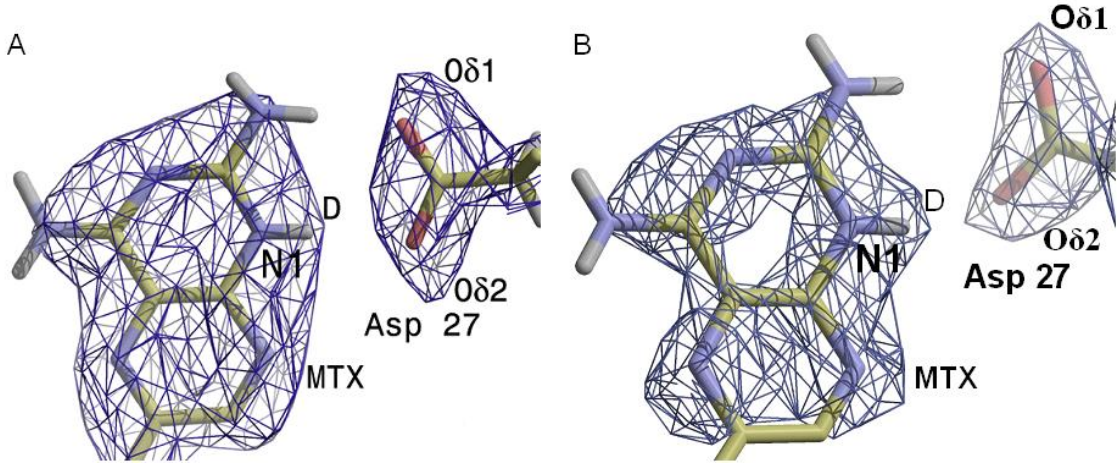


Figure 4.18: Nuclear density maps at the MTX binding site for monomers A and B of the DHFR/MTX cocrystal. $2\boldsymbol{F}_o$- $\boldsymbol{F}_c$ nuclear density maps contoured at 1.8σ are shown in blue. The occupancy of a D atom bound to the N1 atom of MTX for (A) monomer A and (B) monomer B was performed in SHELX. The refined occupancy values were 1.0 and 0.75 for monomers A and B, respectively. The D…Oδ2 distance is 1.9Å and 2.0Å for monomers A and B, respectively.

loop position, the Asp 27 carboxylate and the MTX N1 are both charged, resulting in a salt bridge which is most likely the driving force behind DHFR possessing an order-of-magnitude lower K_d for MTX (pM) than for substrates (nM) (Appleman *et al.* 1988).

Chapter 5. High resolution structures of Apo and MTX/NADPH-bound DHFR

5.1 Background

This chapter describes X-ray crystallographic studies of DHFR in apo and MTX/NADPH ternary complex forms. Both crystal forms are grown from newly identified crystallization conditions and their data sets extend to much higher resolution than had been previously reported; the resolution limits of the apo structure in the PDB (5DFR; space group $P3_121$) is 2.3Å and the one reported here is 1.05Å, while the resolution limits of the MTX/NADPH structure in the PDB (1RH3; space group $P3_221$) is 2.4Å and the one reported here is 1.4Å. Most of the results and discussion in this chapter focus on the apo DHFR structure because the ultrahigh resolution attained 1) allowed observation of the Met20 loop, whereas it is disordered in the lower resolution apo structure in the PDB, and 2) permitted anisotropic and full-matrix refinement due to the high data-to-parameter ratio. Unfortunately, the MTX/NADPH data did not extend to this level of resolution, and the reduced number of unique reflections did not justify the full-matrix refinement. Additionally, although the electron density maps are of high quality, analysis of single difference density for hydrogen atoms and protonation states was excluded for this structure due to the lower resolution limit.

E. coli DHFR loop dynamics: relationship to catalysis

As described in the previous chapters, loop regions are a determining factor of the catalytic functioning of *E. coli* DHFR. A striking example of this is how alteration of the loop sequence by mutagenesis can have significant effects on the catalytic cycle.

Replacement of the core residues 16-19 of the Met20 loop (which actually spans residues 9-24) with a single Gly residue has little effect on substrate (DHF) and cofactor (NADPH) binding but reduces the hydride transfer rate by several hundred fold. This effectively switches the rate-limiting step from product release to hydride transfer (which is normally ~220 s^{-1}) and indicates that the Met20 loop is indelibly tied to the enzymic mechanism (Li *et al.* 1992). From the >50 crystal structures so far determined for *E. coli* DHFR in several different liganded states (Berman *et al.* 2002), it is known that the Met20 loop is adjacent to the substrate and cofactor binding sites, and acts as a gate which can close over the active site (Sawaya *et al.* 1997). In fact, this is the conformation that is observed from crystallography, NMR and other methods when DHFR is in the ground state (only NADPH is bound) or during the hydride transfer chemical step (when both DHF and NADPH are bound) (Schnell *et al.* 2004). However, it has been demonstrated more recently that the substitution or deletion of residues distant from the active site have significant effects on cofactor binding and on hydride transfer rates. Specifically, substitution of Asp 122 for Asn or Ala, substitution of Gly 121 for Val or Gly 121 deletion, residues which are 10-15Å from the C6 atom on DHF (where hydride transfer takes place), increase the NADPH K_m up to 40-fold and decrease the k_{cat} of the reaction by a hundred-fold or more (Miller *et al.* 1998a) (Miller *et al.* 1998b). This has led to the idea that hydrogen-bonding between the core of the Met20 and the βF-βG loops is necessary to stabilize the Met20 loop in a conformation which maintains its closure over the active site during catalysis (Miller *et al.* 2001). This would effectively "lock in" DHF and NADPH while sequestering bulk solvent, ions, and other compounds from freely accessing the active site during chemistry. Additionally, mutagenesis of critical

loop residues also alter the closed-to-occluded Met20 loop conformation equilibration, possibly hindering the catalytic cycle by not allowing exchange of $NADP^+$ for NADPH for a new hydride transfer step (Rajagopalan *et al.* 2002). The closed and occluded conformations for the Met20 loop have been observed in solution experiments such as NMR; most recently described are the ^{15}N relaxation dispersion experiments by (McElheny *et al.* 2005) which further described that the Met20 mainly adopts a closed conformation with folate and $NADP^+$ bound but also reveals the presence of a small population of an "excited state" of occluded conformers in this complex. The movement of the Met20 loop from closed to occluded is essentially due to the flipping of the oxidized nicotinamide ring into and out of the active site, speculated to what actually happens after hydride transfer due to repositioning of Gly 15-Met 16-Glu 17 in the core of the Met20 loop. This causes them to be oriented into the cofactor binding site, thus, occluding $NADP^+$ from the active site after catalysis (Schnell *et al.* 2004) (Venkitakrishnan *et al.* 2004).

Crystallographic evidence of other Met20 loop conformations

Other conformations of the Met20 loop, called open and disordered, have been observed only in the context of the crystalline lattice. The open conformation seems to be mainly stabilized by crystal contacts, however, it could be important for cofactor binding and accessibility as well as product release because the core of the loop extends away from the active site, further outside and bent opposite of the closed or occluded conformations (Sawaya *et al.* 1997). The disordered loop conformation has been observed in the apo structure (Bystroff *et al.* 1991) and in a particular space group for the MTX complex ($P2_12_12_1$) (Sawaya *et al.* 1997). Essentially, there exists weak or no

electron density for residues 15-20 of the Met20 loop, however, it has been suggested that this loop is actually in a time-averaged configuration between closed and occluded (Schnell *et al.* 2004). In the report of the previous apo DHFR structure, it was noted that weak density existed at the cofactor nicotinamide and substrate pABG binding pockets, prompting the suggestion that parts of the Met20 loop could be modeled into these positions (Bystroff *et al.* 1991). This may indicate that the loop, at least in the apo DHFR crystal, is occluding into the cofactor-binding site or is possibly in a more closed position than had previously been observed.

Protonation state of Asp 27 in the apoenzyme

In addition to the Met20 loop, an area of continued interest in DHFR research is probing the importance and the ionization state of the Asp 27 residue along the reaction coordinate. Asp 27 is the lone ionizable residue in the hydrophobic active site and, as stated in previous chapters and inferred from crystallographic studies on DHFR bound to the weak substrate folate (Reyes *et al.* 1995), its carboxylate group is >5Å distant from the N5 atom on the DHF substrate. Mutagenesis has revealed that the Asp 27 is critically important to catalysis, markedly reducing the k_{cat} when it is replaced by Ser or Asn (Howell *et al.* 1986). Obviously, it is somehow required to facilitate protonation of the DHF N5, just not directly. Further confounding its possible role in ligand binding and catalysis is that the kinetic pK_a determined for the reaction is 6.5 (Fierke *et al.* 1987); therefore, a severe perturbation of the Asp carboxylate pK_a would exist if one ascribes the kinetic pK_a to titration of the Asp 27. Essentially, there are two proposed roles for Asp 27 in the proton donation pathway (Rajagopalan *et al.* 2002): **(1)** ionized Asp 27 polarizes bound DHF, triggering keto-enol tautomerization at the N3-C4-O4 of DHF with proton

donation to a water molecule and then to N5, or **(2)** Asp 27 accepts a proton from the DHF N3, causing direct protonation of N5 from a conserved water molecule. Combined results from quantum mechanical MD calculations (Bajorath *et al.* 1991a); (Bajorath *et al.* 1991c); (Greatbanks *et al.* 1999), multidimensional NMR (Blakley *et al.* 1993); (Casarotto *et al.* 1999), and difference Raman spectroscopy (Chen *et al.* 1994); (Chen *et al.* 1997) lend evidence to the former role: that Asp 27 is ionized at physiological pH, is responsible for polarizing bound substrate, and triggers tautomerization, thus initiating the proton relay to ultimately reduce the N5 atom. Cannon *et al.* (1997) additionally proposes that, when the enol tautomer is present, the Asp 27 is protonated, presumably the proton coming from the DHF N3 (Cannon *et al.* 1997b). Both of the above pathways have Asp 27 interacting with a conserved crystallographic water molecule (Reyes *et al.* 1995) but the latter role **(2)** involves direct protonation from another water molecule to the N5 atom of DHF or the N8 atom of folate (Cummins *et al.* 2001). Early difference spectroscopic studies indicated that the Asp 27 in the apoenzyme may be protonated at physiological pH (Stone *et al.* 1983b) (Stone *et al.* 1983a). This has since been refuted by the above NMR (Casarotto *et al.* 1999) and Raman (Chen *et al.* 1997) experiments as well as recent MD calculations of proton affinities of Asp 27 in low dielectric environments (ε = 4-20) more representative of protein interiors (Cannon *et al.* 1997a). These studies provide support that, in the apoenzyme, Asp 27 possesses a pK_a that is <5 and, thus, is negatively charged at physiological pH. A titration at some other group on the enzyme or substrate must be responsible for the kinetic pK_a of 6.5. It has been suggested that it this pK_a may be ascribed to either a titration occurring at the DHF N5

that is reduced in the reaction (Chen *et al.* 1994) (Rod *et al.* 2003) or the DHF O4 atom involved in keto-enol tautomerization (Cannon *et al.* 1997b).

Ultrahigh resolution crystallography has been used to probe unbeknownst features of DHFR in the unliganded form (the apoenzyme), including active site protonation states, the conformation of the Met20 loop, interloop contacts, anisotropy analysis, and alternate side chain conformations. A comparison to the previously determined apoenzyme structure and to an ultrahigh resolution substrate-bound structure is also provided. Additionally, some of the results here for the apo structure are cast as mechanistic implications for DHFR ground state ligand binding and catalysis. Finally, a report of a high resolution MTX/NADPH ternary complex and a brief analysis are provided. The structures presented in this and the previous chapter represents the highest resolution data attained for *E. coli* DHFR save a folate/$NADP^+$ ternary structure described by Dagmar Ringe at a recent conference (Ringe 2005). In fact, only two other DHFR structures from any species have been solved to ultrahigh resolution and deposited into the PDB; that of human DHFR, both structures bound to NADPH and two different proprietary compounds at 1.05 and 1.09Å resolution, respectively (Klon *et al.* 2002).

5.2 Results and discussion

X-ray crystallography: Data collection and processing

New crystallization conditions were identified for apo and MTX/NADPH-bound *E. coli* DHFR from a sparse matrix screen (Jancarik 1991) and were optimized by varying pH and precipitant concentration so as to obtain large volume (>0.05 mm^3), well-diffracting crystals. Similar to the strategy outlined in the previous chapter for synchrotron data collection for the DHFR/MTX crystal, an apo crystal (0.7 x 0.5 x 0.2

mm^3) was chosen for a two-step data collection in which 100° of φ space (at 0.5° per oscillation) was sampled in each scan. The difference here is that the low resolution data were collected first which may have been the reason that one apo crystal remained usable for both lower resolution and higher resolution scans. First, two low resolution scans were accomplished by keeping the detector back at a distance of 200 mm (the beam stop was set at 50 mm) from the crystal, limiting the exposure times per frame to 1 sec, and setting a pair of slits (JJ and front-end) to effectively weaken the incident X-ray beam. Merging the integrated intensities from these two scans together manually in Scalepack (Otwinowski *et al.* 1997) revealed a combined low resolution data set (20.0 – 1.4Å) that was 96% complete. More importantly, the scaling indicated that the crystal showed no signs of radiation damage. Thus, an ultrahigh resolution scan was then recorded by moving the detector up to 150 mm, the beam stop to very near the crystal (20 mm), increasing the exposure times per frame to 15 sec, and resetting the pair of slits to allow maximum beam. Again, a foil attenuator was used to protect the center of the detector from overexposure. Therefore, this region of data (20.0 – 2.0Å) was exempted from further analysis. Scaled separately, the ultrahigh resolution apo DHFR data set is 94% complete from 20 to 1.0Å. Due to much higher R_{sym} values for merged intensities beyond 2.0Å for the low resolution scans, these data sets were reintegrated from 20.0 - 2.0Å and then were scaled together with the ultrahigh resolution data using a resolution range of 20.0 – 1.05Å. Based on previous experience collecting *P*6 DHFR data on this detector for a separate crystal form, a *P*6 reindexing matrix (*hkl* to *kh-l*) was implemented into the Scalepack (Otwinowski *et al.* 1997) script. The merged apo data is 93% complete to 1.05Å with an R_{merge} of 8.9%.

The data collection strategy for the MTX/NADPH crystal (approx. 0.3 x 0.3 x 0.2 mm^3 in size) was identical to that for the apo form except that only one low resolution pass was necessary and that the detector was kept at a 200 mm distance from the crystal for both the low resolution and the high resolution scans. Only one crystal was used for data collection. The integrated data sets were scaled together in Scalepack (Otwinowski *et al.* 1997) in the $P3_221$ space group and the resulting merged data set is 98.6% complete to 1.35Å with an R_{merge} of 5.9%. The data collection and processing statistics for the apo and MTX/NADPH crystals are listed in **Table 5.1**.

Difference Fourier calculation for apo DHFR

The apo DHFR crystal possesses a unit cell which is isomorphous with only one other deposited DHFR structure, that of the $P6_5$ binary complex with folinic acid (1JOM; (Lee *et al.* 1996)). The previously reported apo DHFR crystal (5DFR; (Bystroff *et al.* 1991)) is in the $P3_121$ space group and has a non-isomorphous unit cell with the apo crystal reported here. Initially, rigid body refinement in Refmac5 (Murshudov *et al.* 1997) was conducted using a previously solved structure of DHFR bound to folinic acid (1JOM; (Lee *et al.* 1996)) as the starting model (F_{calc}) to refine against the 1.0Å apo DHFR reflection data (F_{obs}). After 10 cycles of rigid body refinement, R_{fac} and R_{free} were 45% and 44.3%, respectively.

MR solution of the DHFR/MTX/NADPH complex

The difference Fourier method was attempted for the MTX/NADPH 1.4Å reflection data using previously determined coordinates for DHFR bound to MTX and NADPH (1RH3; (Sawaya *et al.* 1997). The unit cell is isomorphous. Surprisingly, upon rigid body refinement in Refmac5 (Murshudov *et al.* 1997), the R factors remained

Table 5.1: X-ray data collection, indexing, integration, and scaling statistics: Apo DHFR

DHFR crystal form	Apo	MTX/NADPH
BioCARS Beamline;	14-BM-C	14-BM-C
Detector	ADSC Q315	ADSC Q315
Unit cell (a,b,c, in Å)	95.8, 95.8, 34.3	61.9, 61.9, 105.8
(αβγ, in °)	90, 90, 120	90, 90, 120
Molecules/ A.U.	1	1
Solvent content (%)	50.8	61.4
Space group	$P6_5$	$P3_221$
Resolution range (Å)	20.0-1.05	20.00-1.40
Highest shell	1.09-1.05	1.45-1.40
Unique reflections	83,656	43,422
Completeness (%)	92.9 (63.6)	98.6 (99.9)
R_{merge}[1] (%)	8.9 (27.6)	5.9 (67.3)
I/σ	12.4 (4.4)	29 (2.1)
Multiplicity	2.9 (1.6)	5.5 (2.9)
# Data sets used	3	2

R_{merge}[1]$= \Sigma_{hkl} \Sigma_i |I_{i(hkl)} - I_{hkl}| / \Sigma_{hkl} \Sigma_i I_{(hkl)}$

greater than 50%. Using this same coordinate set (1RH3) as a target molecule and a resolution range of 50.0-3.0Å, molecular replacement using PHASER (McCoy *et al.* 2005) as implemented within the CCP4 suite (CCP4 1985) was successful in finding a solution model and was then used as the starting coordinate set for refinement in Refmac5.

Structure refinement and analysis: Apo DHFR

The structure refinement statistics for apo DHFR are listed in **Table 5.2** and the apoenzyme unit cell contents are listed in **Table 5.3.** Further refinement of the apo structure in Refmac5 and, then, in SHELXH (Sheldrick *et al.* 1997), to 1.05Å included the refinement of anisotropic displacement parameters (ADPs) and the addition of hydrogens to all of the side chains (save side chain hydroxyls), ultimately reducing the R_{free} to 23.1%. The mean anisotropy (A, or E_{min} / $E_{max;}$ for details please see the last chapter) of the apo structure is 0.53, within range of what has been observed for many high resolution structures (0.45-0.55) (Merritt 1999b). However, anisotropy ratio values overall are lower in the C-terminal region of the protein (from the βF strand to the C-terminus), correlating to the regions with the highest B-factors in the structure (**Figure 5.1**). In contrast to either of the monomers in the MTX structure (chapter 4), the region around and containing the adenosine-binding loop (residues 60-80) is well-ordered in the apo structure, with very low B-factors ($\leq$12Å^2) and higher A values (more isotropic; average ~0.6). Using Refmac5 for further refinement of the model allowed loosening of the bond and angle restraints (by increasing the sigma weighting factor to 5; default is 0.3-0.5); this reduced the R_{free} to 22.6%. Essentially, this allows more reliance on the data

Table 5.2: Refinement statistics for the apo and MTX/NADPH DHFR structures

DHFR crystal form	Apo	MTX/NADPH
Resolution range (Å)	20.0-1.05	20.00-1.40
Highest shell (Å)	1.09-1.05	1.45-1.40
Final Isotropic R_{fac}[2]/ R_{free}[3] (Refmac5)	27.3%/28.9%	21.4%/23.6%
Anisotropic R_{fac}/R_{free} (Refmac5)	22.4/23.8	20.6/22.1
Same coordinates; Isotropic in SHELXH	32.0/34.1	N/A
Anisotropic R_{fac}/R_{free} (SHELXH)	23.9/25.9	N/A
Riding hydrogens added; R_{fac}/R_{free}	22.8/25.1	N/A
Final R_{fac}/R_{free} (SHELXH)	22.5/24.1	N/A
Final Weighting matrix (σ; Refmac5)	5.0	5.0
Final R_{fac}/R_{free} (Refmac5)	21.0/22.3	16.3/19.1
Final RMSDs for bond distances and angles	0.022Å 2.033°	0.019Å 1.994°

R_{fac}[2]$= \Sigma_{hkl} \mid |\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}| \mid / \Sigma_{hkl} |\boldsymbol{F}_{obs}|$; R_{free}[3]$= \Sigma_{hkl} \mid |\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}| \mid / \Sigma_{hkl} |\boldsymbol{F}_{obs}|$

Table 5.3: Contents of the unit cell for the apo and MTX/NADPH DHFR crystals

DHFR crystal form	**Apo**	**MTX/NADPH**
Unit cell contents		
Protein atoms	1264	1284
Water molecules	269	168
Metals	1 (Mg^{2+}) (Near Asp 70)	0
Ligands	N/A	2 (MTX; NADPH)

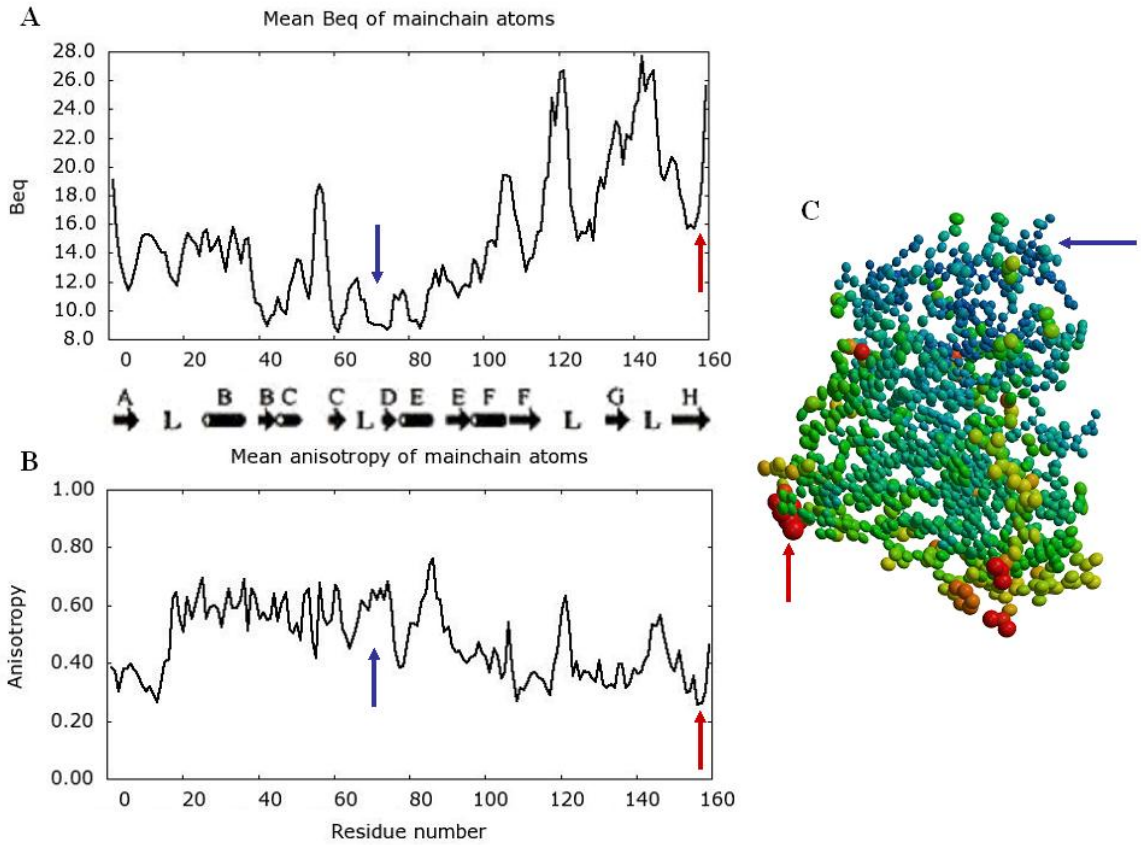


Figure 5.1: B-factor and anisotropy distribution for Apo DHFR. A) B-factor values (in Å^2) for the backbone atoms of the 1.0Å apo DHFR structure. **B)** Anisotropic ratio ($A = E_{min} / E_{max}$) for the backbone atoms; the lower the A value, the more anisotropic an atom is (its vibrational motion cannot be described as spherical) **C)** Apo DHFR model showing all protein atoms represented as thermal ellipsoids. Atoms exhibiting non-spherical vibrational motion are more elongated and ellipsoidal and are shown in yellow>orange>red (higher>lower A). Red arrows indicate regions of high B-factors and more anisotropic behavior. Atoms exhibiting displacement more representative of isotropic motion are shown as spheres (less ellipsoidal) and are colored in green>blue (lower>higher A). Blue arrows indicate regions of low B-factors and less anisotropic behavior.

than on the standard Engh and Huber (1991) small molecule libraries for geometric restraints that Refmac5 utilizes normally (Engh *et al.* 1991); therefore, matrix weighting should only be altered significantly when sufficiently high resolution data is available (beyond 1.2Å) and the data-to-parameter ratio is quite high (>10:1). An additional cycle of rebuilding and addition of more solvent molecules decreased the R_{free} to its current value of 22.3%. The final r.m.s.d. for bond lengths and angles were 0.021Å and 1.965°, respectively. The data-to-parameter ratio for isotropic B-factor refinement was ~12:1 whereas for refinement of ADPs (which essentially doubles the parameters that are refined) the ratio was ~5.5:1. Thus, unrestrained and full-matrix refinement of the first 30 residues of the apo structure was performed as described for the MTX structure detailed in the previous chapter and in chapter 2. This provides a precise and accurate measurement of bond lengths and angles. It was used here to assess carboxylate bond lengths and their calculated estimated standard deviations (ESDs) from ideality given the model so as to elucidate the charge state of the Asp and Glu residues.

Full-matrix refinement results: the Asp 27 in Apo DHFR

Surprisingly, the Cγ-Oδ1 and Cγ-Oδ2 bond lengths for Asp 27 are not close to equivalence. In fact, they are 0.092Å different, which is greater than 3 times the ESD (0.087Å) (**Table 5.4**). The B-factors of the Cγ, Oδ1, and Oδ2 atoms were 16.1, 17.4, and 21.3Å^2, respectively. The overall B_{iso} (main chain atoms) for the apo structure was 14Å^2. Overall, these thermal parameters seem reasonable and the ADPs are lower (*A* ratios are higher) and disorder is reduced within this part of the structure (all side chains can be modeled and there is no main chain disorder). Therefore, this analysis provides evidence that the Asp 27 is protonated (Coates *et al.* 2002; Deacon *et al.* 1997), at least in the

Table 5.4: Refined carboxylate bond lengths, ESDs, isotropic B-factors, and charge estimation for Asp 11, Glu 17, and Asp 27 in monomers A and B of the 1.05Å apo DHFR structure

Residue	Bond length (in Å)	ESD (in Å)	B-factor (in Å^2)
Asp 11			
Cγ-Oδ1	1.234	0.020	Cγ 13.5
Cγ-Oδ2	1.255	0.020	Oδ1 13.2
Charge state	Negative		Oδ2 17.2
Glu 17			
Cδ-Oε1	1.282	0.057	Cδ 14.0
Cδ-Oε2	1.321	0.049	Oε1 23.2
Charge state	Negative		Oε2 14.2
Asp 27			
Cγ-Oδ1	1.281	0.031	Cγ 16.1
Cγ-Oδ2	1.190	0.027	Oδ1 17.4
Charge state	Neutral		Oδ2 21.3

context of the crystalline lattice and under these conditions. The crystallization pH was 7.5; this suggests a dramatic perturbance of the Asp 27 pK_a in the apoenzyme. Additional results from the full-matrix refinement (**Table 5.4**) revealed that the Asp 11 and Glu 17 are both charged in the apoenzyme, with bond length differences much less than the mean calculated ESDs. The average B-factors for the atoms in these carboxylate groups were ~16-17$Å^2$. These residues, both within the Met20 loop, have side chains which are oriented towards the surface of the protein and away from the interior and interact with a number of solvent molecules (**Figure 5.2**). The Asp 27 carboxylate Oδ1 has three possible interaction partners, with the Thr 113 Oγ (2.7Å distance), the Wat 90 oxygen (2.9Å), and the Wat 131 oxygen (3.0Å). The Asp 27 Oδ2 atom maintains two possible contacts, with the Wat 83 oxygen (2.7Å) and the Wat 132 oxygen (2.5Å) (**Figure 5.3 A, B**). Superposition of the MTX closed Met20 loop structure (monomer B, chapter 4) onto the apo structure reveals that these solvent molecules are positioned very near (≤1.0Å distance) to where MTX pteridine ring atoms are when bound in the DHFR active site (**Figure 5.3 B**). In fact, two of these (Wats 90 and 132) are at nearly the same exact position where the NA2 and N1 atoms of MTX are; these atoms in MTX make direct hydrogen-bonding and ionic contacts with the Asp 27 carboxylate as detailed in chapter 4.

A conformation for the Met20 loop in the apoenzyme crystal structure

A striking feature in the maps early in refinement and rebuilding was the continuous stretch of $2\mathbf{F}_o$-$\mathbf{F}_c$ electron density for all of the Met20 loop (**Figure 5.4**) that was unobservable, at least for the core residues 16-20, in the previous apo structure

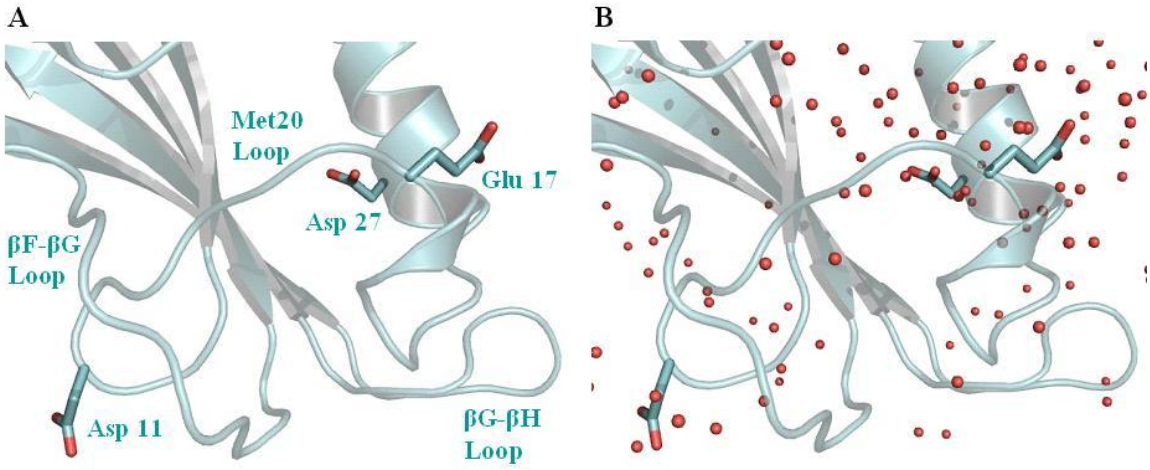


Figure 5.2: The positions of the residues in the apo DHFR structure used in the carboxylate bond length measurements in the full-matrix refinement. A) The Asp 11 and Glu 17 residues are both within the Met20 loop and their side chains are at the surface and directed away from the protein interior. Asp 27 is the only ionizable residue in the DHFR active site and is part of helix αB. **B)** The same image and orientation as in **(A)** only with solvent molecules shown as red spheres.

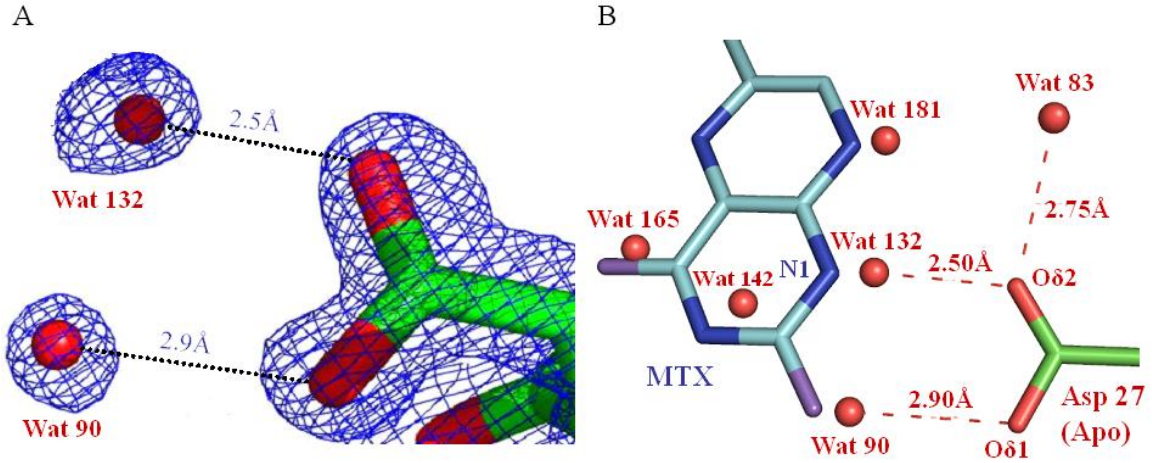


Figure 5.3: In the apo DHFR structure, solvent molecules form close contacts with the Asp 27 carboxylate that are conserved in substrate and inhibitor complexes. A) Shown in blue is the $2\mathbf{F}_o$-$\mathbf{F}_c$ electron density map contoured at 2σ. The Asp 27 carboxylate group forms possible hydrogen-bonding contacts with three solvent molecules (Wats 90 and 132 shown here, Wat 83 shown in B) and a nearby Thr residue (a 2.7Å contact with Thr 113 Oγ not shown). In the 5DFR structure, the only contact described to the Asp 27 is to the Thr 113. **B)** Solvent molecules occupy positions in the apo DHFR structure that are very near (most are <0.5Å in distance) to where substrate and inhibitor atoms (like those in MTX shown) would be. The closed Met20 loop MTX structure (monomer B) was used for superposition onto the apo structure.

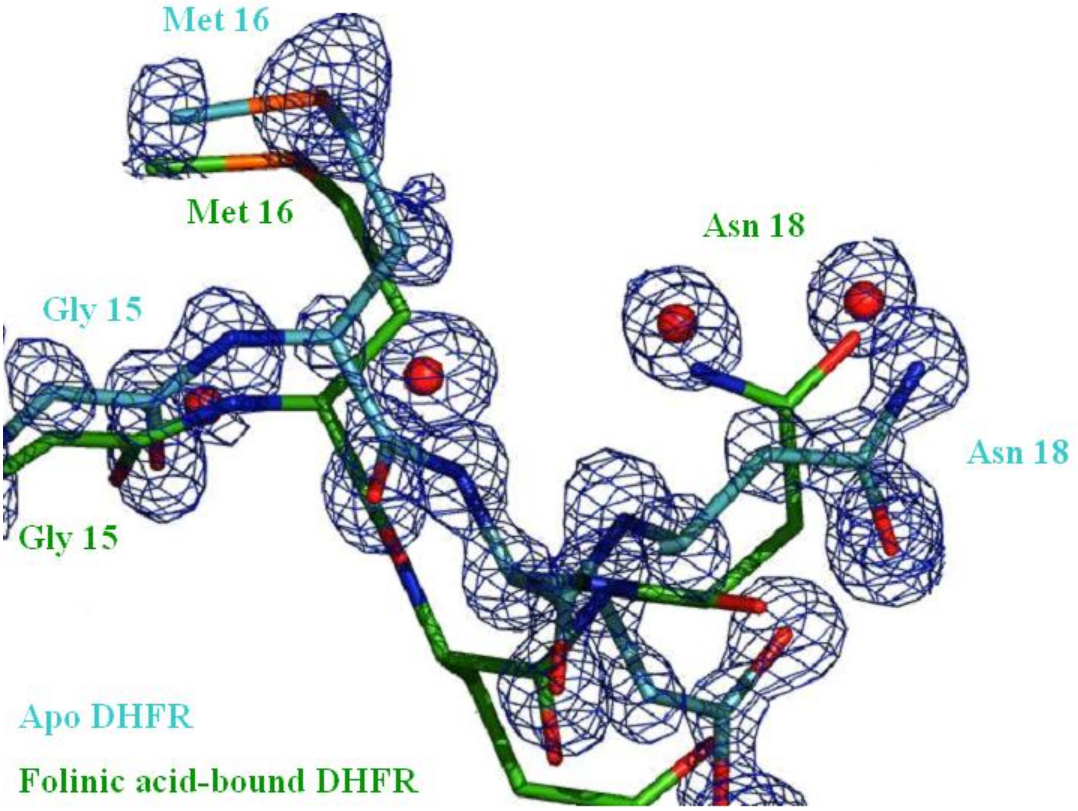


Figure 5.4: The ultrahigh resolution (1.05Å) of the electron density maps allowed modeling of the core residues of the Met20 loop, observed as disordered in the previously reported apo DHFR structure (5DFR; (Bystroff *et al.* 1991). Shown in blue is the $2\mathbf{F}_o$-$\mathbf{F}_c$ electron density contoured at 2σ for part of the Met20 loop. Although the folinic acid-bound DHFR structure (shown in green) could be used in difference Fourier calculations and as the initial model for apo DHFR, most of the core Met20 loop side chains and even regions of the main chain had to be rebuilt to better fit the electron density from the ultrahigh resolution apo DHFR data. The rebuilt and refined apo Met20 loop is shown in cyan.

(5DFR (Bystroff *et al.* 1991); no coordinates are reported from the amide nitrogen of Met 16 to the carbonyl carbon of Met 20). Using a loop modeling utility within the model building program, O (Jones *et al.* 1991) (Kleywegt *et al.* 1997), a preliminary loop based on the folinic acid-DHFR model (1JOM (Lee *et al.* 1996); the model used in the difference Fourier calculation) was placed into density and then atoms were manually refined into favorable positions. An alignment of loops from different structures reveals that this is a new type of closed conformation for the core residues of the apo Met20 loop (**Figure 5.5**). It was surprising that, in the absence of any ligand bound, the Met20 loop would adopt a closed conformation. However, the core region of the loop (residues 14-20) is very ordered, with clear electron density for all the side chains and average side chain B-factors of 14-15Å^2. Upon generating the atoms involved in symmetry-related contacts, it was observed that an Mg^{2+} ion was hexa-coordinated by four solvent molecules, Asp 70 Oδ2 from the symmetry molecule, and Asn 18 Oδ1 from the origin molecule (all contacts between 2.0-2.1Å) (**Figure 5.6**). Additionally, the Asn 18 side chain amide Nδ2 forms 3.0Å hydrogen-bonding contacts with Asp 70 Oδ2 and Water 51, both from the symmetry molecule. Therefore, this interface between crystallographic molecules seems to be stabilized by a solvent and ion network. Indeed, the average B-factor for the Asn 18 residue is 11.5Å^2. Most likely, it is this system of crystal contacts that induces the Met20 loop to adopt a closed conformation. It could be that, by virtue of the closed loop conformation, the active site is more sealed and prevents *en masse* the entrance of bulk solvent, providing a chemical environment more conducive to hydrophobicity. This could provide evidence to why the Asp 27 carboxylate is protonated and, thus, neutral. Perhaps there is decreased energetic favorability to burying a negative

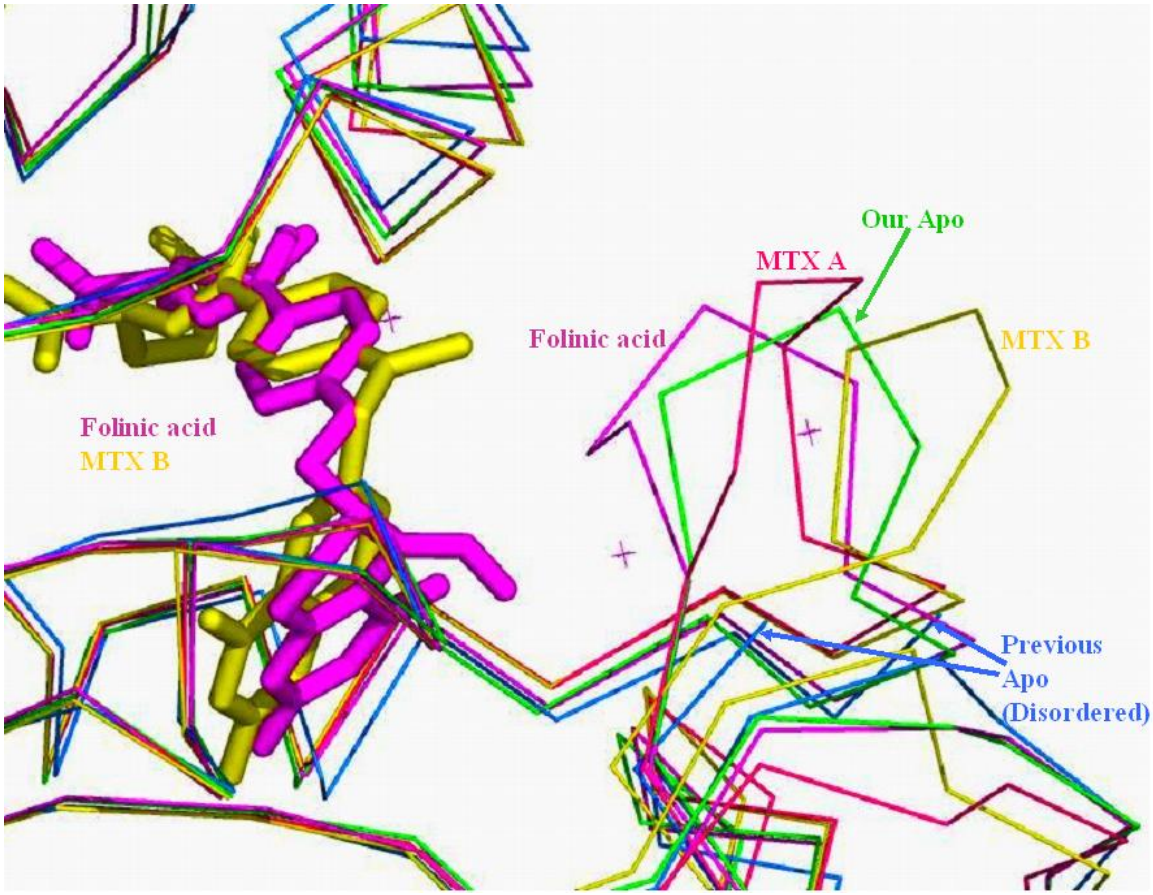


Figure 5.5: The Met20 loop in apo DHFR is in a closed conformation. Shown are color-coded Met20 loop Cα traces from representative *E. coli* DHFR structures. Nearest to the ligand binding site is the folinic acid-bound DHFR Met20 loop (1JOM; the folinic acid ligand and its Met20 loop are in purple), next in the sequence from left to right is the 1.05Å apo structure reported here (closed; light green; black arrow), MTX B molecule (closed; pink), MTX A molecule (partially occluded; loop and the MTX for the A molecule are in yellow), and, shown as disordered (no trace) is the previously reported apo loop (5DFR; light blue). There are no coordinates reported in 5DFR for the core of the Met20 loop, from the backbone carbonyl of Gly 15 to the backbone amide of Pro 21

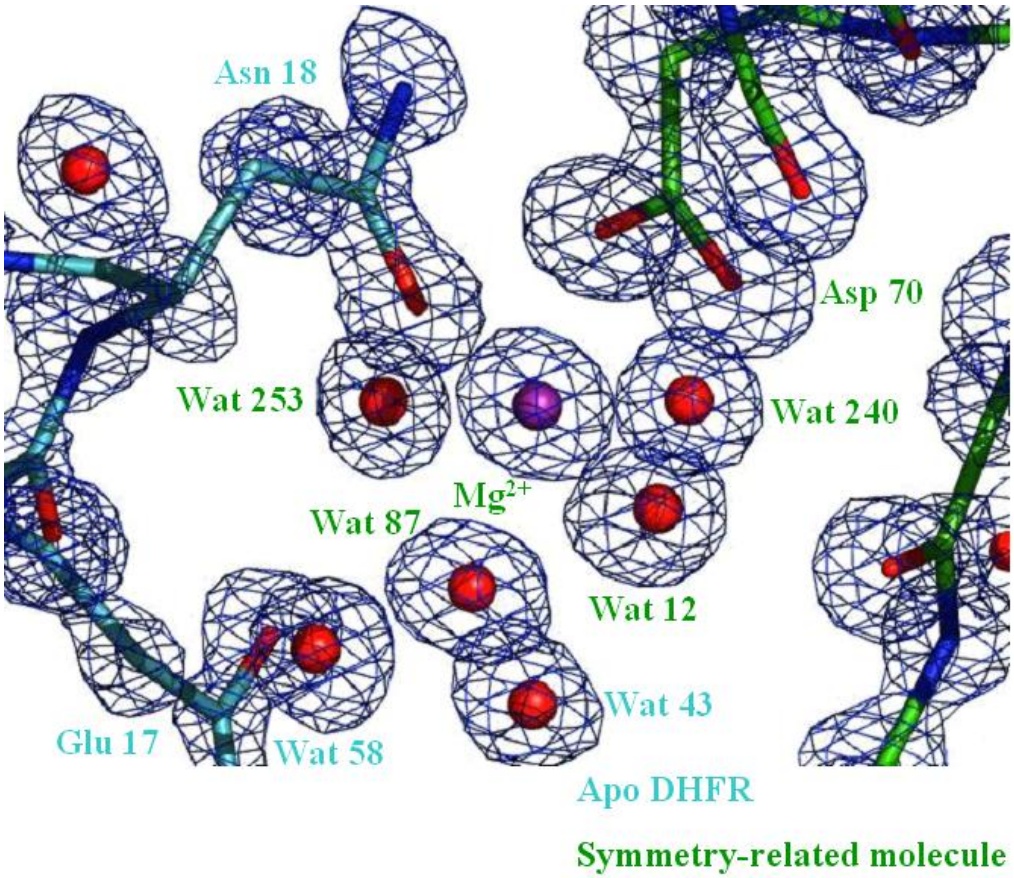


Figure 5.6: Crystal contacts at a metal coordination site near Asn 18 help to stabilize the Met20 loop in a closed conformation in the apo DHFR structure. Residues and solvent atoms for the origin apo DHFR molecule are shown in cyan whereas residues and solvent for the symmetry-related molecule are shown in green. $2\mathbf{F}_o$-$\mathbf{F}_c$ electron density contoured at 1.5σ is shown in blue. The Mg^{2+} ion possesses hexavalent coordination, with 4 contacts to oxygens of water molecules, the Oδ2 of Asp 70 (symmetry molecule), and the Oδ1 of Asn 18 (origin molecule). The average metal…oxygen coordination contact distance is ~2.1Å.

charge in an already-hydrophobic active site which is vacant of any ligand that could counter with a positive charge (such as MTX) for an ionic interaction or could form hydrogen-bonding contacts (such as folate).

Differences between apoenzyme structures

A comparison to the previous apo DHFR structure (5DFR; solved to 2.3Å resolution; (Bystroff *et al.* 1991)) was undertaken. 5DFR crystallized in the $P3_121$ space group whereas the 1.0Å apo model reported here crystallized in the $P6_5$ space group. Superposition of the two models by least-squares revealed that the r.m.s.d. between the Cα atoms was 0.6Å (**Figure 5.7 A**). The βA strands (residues 1-8) are very similar between the two structures; however, different side chain rotamers exist for Asp 11 and Arg 12. In the 1.0Å apo model, the positions of these side chains are stabilized by symmetry contacts, such as a 2.9Å hydrogen bonding distance between the Asp 11 Oδ1 and the Nς atom of Lys 76 of the symmetry molecule. The differences in the core of the Met20 loop are stated in the above section. Additionally, in the 5DFR structure, there are no coordinates reported for Leu 23-Pro 24; they could be readily modeled into the 1.0Å resolution $2\mathbf{F}_o$-$\mathbf{F}_c$ electron density maps. Also stated above, there are multiple interactions between the Asp 27 carboxylate group and surrounding solvent molecules. The only possible hydrogen-bonding partner to the Asp 27 in the 5DFR structure is between the Oδ1 and the Oγ atom of Thr 113, which is observed in the high resolution model. Not surprisingly, basic side chains which line the side of the active site opposite to where the Asp 27 resides and normally stabilize the diglutamic acid tail of folate ligands and analogs such as MTX (Arg 44, Arg 52, Lys 58) are disordered or in different orientations

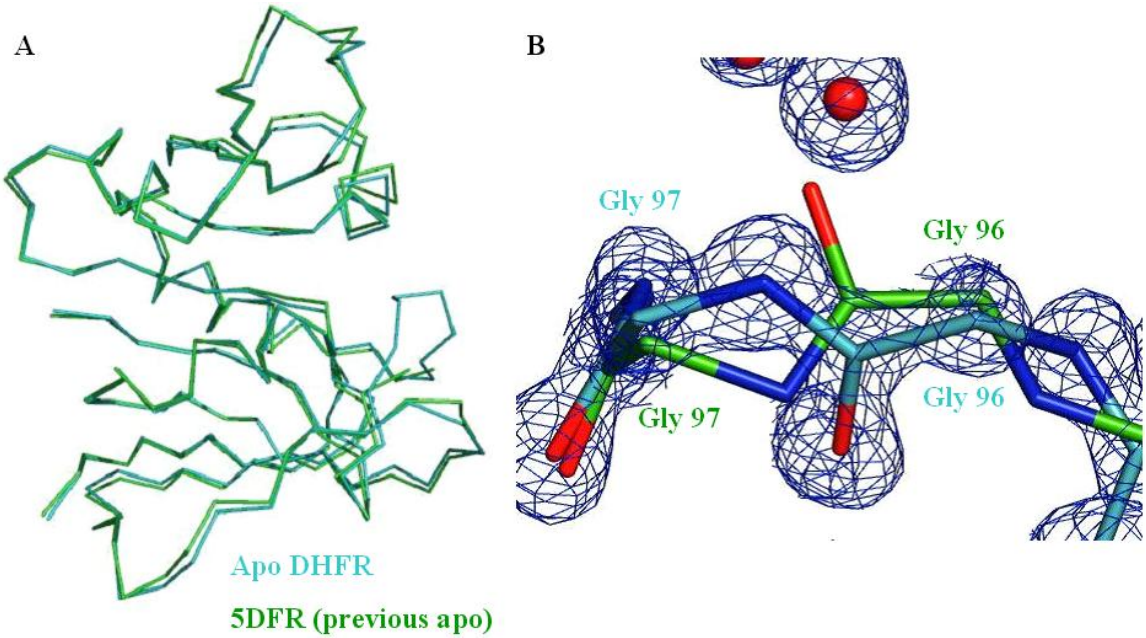


Figure 5.7: Superposition of the 5DFR structure (the previously reported apo DHFR model) onto the 1.05Å resolution apo DHFR structure reported here. A) Overlay of the two models was performed within the program O using least-squares superposition (Cα r.m.s.d = 0.6Å). Note that there is no Cα trace for residues 15-20 of the Met20 loop in the previous apo DHFR structure (green). **B)** A difference in the main chain conformation for Gly 96-Gly 97, which is modeled as trans in 5DFR yet modeled as cis in the ultrahigh resolution structure.

in the 5DFR structure as compared to the ultrahigh resolution model. The Arg 44 is rotated about χ^1 in a completely opposite orientation as compared to 5DFR; this allows a possible hydrogen-bonding contact with the Pro 66 carbonyl oxygen (2.9Å distance). This may help to stabilize the adenosine-binding loop (residues 62-70) in the high resolution structure, which has very low main chain B-factors. There exists ~1.1Å difference between main chains for the two models from Thr 46 to Pro 54 and from Thr 68 to Thr 73. In most of the *E. coli* DHFR structures deposited into the PDB (Berman *et al.* 2002) (and even structures from other organisms such as *Lactobacillus casei*) and in the ultrahigh resolution MTX structures reported in this thesis, the peptide bond for Gly 95-Gly 96 is found in the cis conformation (Bolin *et al.* 1982). In 5DFR, it was modeled as partially disordered between cis and trans, and this peptide bond was suggested to convert from trans to cis upon cofactor binding due to the fact that, in the trans conformation, the Gly 96 carbonyl oxygen could project into the pyrophosphate binding site of the cofactor and a switch to cis would alleviate this steric clash (Bystroff *et al.* 1991). In the 1.0Å model, it is clear that the Gly 95-Gly 96 peptide bond is in the cis conformation (**Figure 5.7 B**). Large χ^1 rotamer differences for Arg 98 and Lys 106 between the two models exist, both allowing the side chains in the 1.0Å model to make more hydrogen-bonding contacts and to fit the $2\mathbf{F}_o$-$\mathbf{F}_c$ electron density better. Another χ^1 rotamer difference between the models occurs at His 114. The Nε2 of the imidazole ring makes a 2.8Å contact with the Oε1 atom of Glu 154 in the 1.0Å apo model whereas in 5DFR the imidazole ring is pointed towards symmetry contacts and solvent molecules >3.0Å away. The remainder of the C-terminus is quite similar between the two models with no major main chain divergences or side chain differences.

Differences between the apo and substrate-bound forms

To determine changes that would occur in the apo DHFR structure upon substrate binding, an ultrahigh resolution (to 1.06Å) folate-bound DHFR structure was superposed using least-squares calculations onto the 1.0Å apo DHFR structure reported here. The folate-DHFR model was kindly provided by Dr. Anna Gardberg, a postdoctoral fellow in the Dealwis laboratory. The resolutions of the data are essentially the same; however, the space groups are different. The folate-DHFR crystals are very similar to the MTX crystals reported in the previous chapter: they crystallize into the $P6_1$ space group and have 2 molecules in the asymmetric unit (AU). The major difference between the molecules in the AU, like the MTX monomers A and B, is the conformation of the Met20 and the βF- βG loops. The closed Met20 loop monomer was used for the superposition and the analysis of differences with the 1.0Å $P6_5$ apo structure, and the r.m.s.d. between the Cα atoms of the two models is 0.7Å. As expected, the Cα traces for the two structures are very similar, with major divergences in the chain occurring at the Met20, the adenosine-binding, and the βF- βG loops (**Figure 5.8 A**). Beginning at the Ala 7 Cα and ending at the Val 10 Cα, there is ~0.5Å main chain difference between the two models. The position of the Arg 12 is also different, with the apo residue pointed towards the βF- βG loop, the Arg NH2 of the guanidino group within 3.3Å of the Tyr 128 hydroxyl group of the aromatic ring. The Arg 12 in the folate structure possesses a slightly different χ^1 rotamer, causing it to bend away from the βF- βG loop a bit more than the apo Arg 12; this causes the folate Arg 12 guanidino group to be about 4.5Å distant from the Asp 127 carboxylate. The positions of Asp 127 and Tyr 128 between the two models are the same. The largest divergence takes place at the core of the Met20

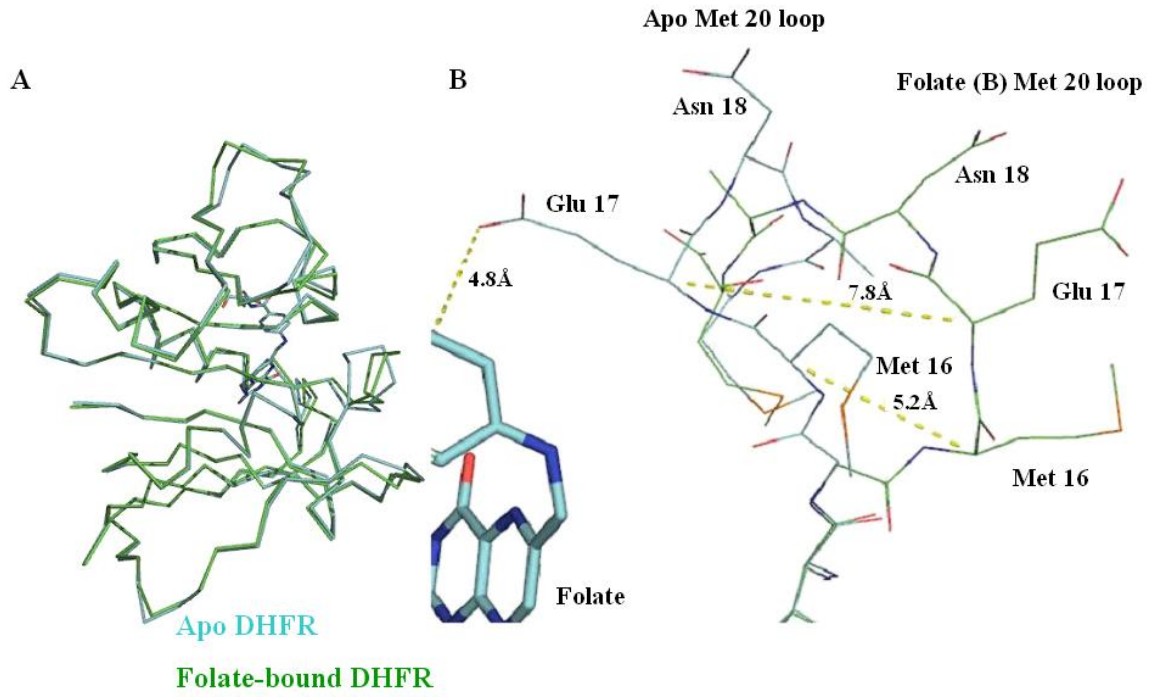


Figure 5.8: Superposition of a closed Met20 loop folate-bound DHFR structure (1.06Å resolution) onto the 1.05Å resolution apo DHFR structure. A) Overlay of the two models was performed within the program O using least-squares superposition (Cα r.m.s.d = 0.7Å). The Cα trace of the apo DHFR is shown in cyan while the Cα trace of folate-bound DHFR (monomer B) is shown in green. Folate is shown bound in the active site. **B)** The orientation of the core residues of the Met20 loop are significantly different between the folate-bound DHFR closed Met20 loop (green) and the apo DHFR closed Met20 loop (cyan) structures. The Glu 17 side chain of the apo loop protrudes into the active site very near (<5Å) to where the folate pABA-glu tail is positioned. The coordinates for the 1.0Å resolution model of folate-bound DHFR were kindly provided by Dr. Anna Gardberg.

loop, with the distances between Cα atoms for Met 16 (5.2Å), Glu 17 (7.7Å), Asn 18 (4.2Å), and Ala 19 (3.3Å) all being quite sizable (**Figure 5.8 B**). The reason for this difference is that the apo loop kinks inward to the protein interior just past the Met 16 backbone amide. The side chains for Met 16, Glu 17, and Asn 18 in the two models are oriented in the opposite directions, with the apo Glu 17 carboxylate slightly protruding into the active site and positioned less than 5Å from the benzoate moiety of the superposed folate ligand. The side chains of all these residues in the folate-bound structure are pointing away from the substrate-binding site. It should be noted, as observed in the MTX structure, that solvent molecules (Wats 90, 132, 142, 165, and 181) in the apo DHFR structure nearly overlay perfectly (≤0.5Å) at the positions of atoms in the folate pterin ring when the folate is bound to DHFR.

The main chain at Arg 44 and His 45 differs by about 1.5Å and the side chains extend in different directions; the Nδ1 atom on the imidazole ring of His 45 in the folate structure is ~4Å away from the Glu 17 Oε1 atom. This same atom pair in the apo structure is at a distance >15Å. The Arg 44 guanidino group in the apo structure is pointed towards the core of the adenosine-binding loop, the NH1 atom making a 3.1Å contact and the NH2 atom making a 2.9Å contact with the side chain carbonyl oxygen of Gln 65 and the main chain carbonyl oxygen of Pro 66, respectively. These interactions may help to stabilize the apo DHFR adenosine-binding loop and cause this region to be unexpectedly well-ordered. Indeed, from Gln 65 to Arg 71, there is an average main chain divergence of ~1.5Å between the two structures, the largest being a 3.0Å difference between the Pro 66 Cα atoms. Many of the longer and branched side chains have very

different rotamers and/or torsion angles about the χ^1 and χ^2 rotations, including Arg 52, Asp 79, Glu 80, and Arg 98.

In the apo DHFR structure, key residues within the βF- βG loop (Asp 120, Gly 121, and Asp 122) are actually 8-12Å distant from the core of the Met20 loop whereas these residues are packed closer to the Met20 loop in the folate-bound DHFR structure. An example of this is the side chain of Met 16; as stated above, it is projecting towards the active site in the apo structure but, in the folate-bound structure, it is projecting towards the Asp 122 of the βF- βG loop, with the distance between Met 16 and Asp 122 Cα atoms decreasing from 12.1Å (apo) to 7.9Å (folate-bound). With a decrease in packing against the βF- βG loop, normally a stabilizer of a closed Met20 loop, it further suggests that the crystal contacts are very important for helping to maintain the "ultra" closed conformation of the Met20 loop in the apo structure reported here. Indeed, in the $P6_5$ folinic acid-bound *E. coli* DHFR structure, it was suggested that components of the crystallization medium (Ca^+ and ethanol) played a role in stabilizing a "completely" closed Met20 loop, where part of the loop protruded into the cofactor-binding site (Lee *et al.* 1996). Although the apo Met20 loop does not protrude as far into the binding cleft as the Met 20 loop in the folinic acid-DHFR structure, recall it was the F_{calc} of this model (1JOM; Lee *et al.*, 1996) used successfully in a difference Fourier calculation in Refmac5 (Murshudov *et al.* 1997) with the apo DHFR reflection data (the F_{obs}). There also exists a difference at the end of the βF- βG loop between the two structures, with ~1.5Å between Cα atoms of Pro 130 through Asp 132. The remainder of the C-terminus for the two models is very similar.

Structure refinement and analysis: DHFR MTX/NADPH ternary complex

As stated above, a previous DHFR/MTX/NADPH structure (1RH3; $P3_221$ space group; (Sawaya *et al.* 1997)) solved to 2.3Å was used as the target molecule for MR. For the MTX/NADPH structure, 10 cycles of rigid body refinement in Refmac5 (Murshudov *et al.* 1997) using the MR output solution model against the 1.4Å reflection data resulted in an R_{fac} and R_{free} that were 42.6% and 43.1%, respectively. After multiple cycles of restrained and B-factor refinement, simulated annealing was performed using CNS (Brunger *et al.* 1998) to reduce model bias from 1RH3. Model building in O included the modeling of alternate side chain conformations and manual addition of several water molecules. Subsequent refinement in Refmac5 to 1.4Å with loosening of the geometric restraints reduced the R_{free} to 19.1%. The final r.m.s.d. for bond lengths and angles were 0.019Å and 1.994°, respectively. The data-to-parameter ratio for isotropic B-factor refinement was ~7:1 whereas, for refinement of ADPs, the ratio was ~3:1. Due to the smaller data-to-parameter ratio as compared to the apo and the MTX binary structures (stemming from the resolution limitations of the MTX/NADPH data), further refinement and analysis (i.e. a full-matrix calculation for carboxylate bond lengths and ESDs) in SHELX was not performed. The structure refinement statistics for the DHFR MTX/NADPH ternary complex are listed in **Table 5.2**.

Analysis of the MTX and NADPH binding sites

Electron density maps of the MTX and the cofactor binding sites are shown in **Figure 5.9**. It is expected that the Asp 27 is negatively charged in the MTX/NADPH ternary complex to conserve the ionic interaction identified from neutron crystallography for the MTX binary complex as reported in the previous chapter. Additionally, this is a

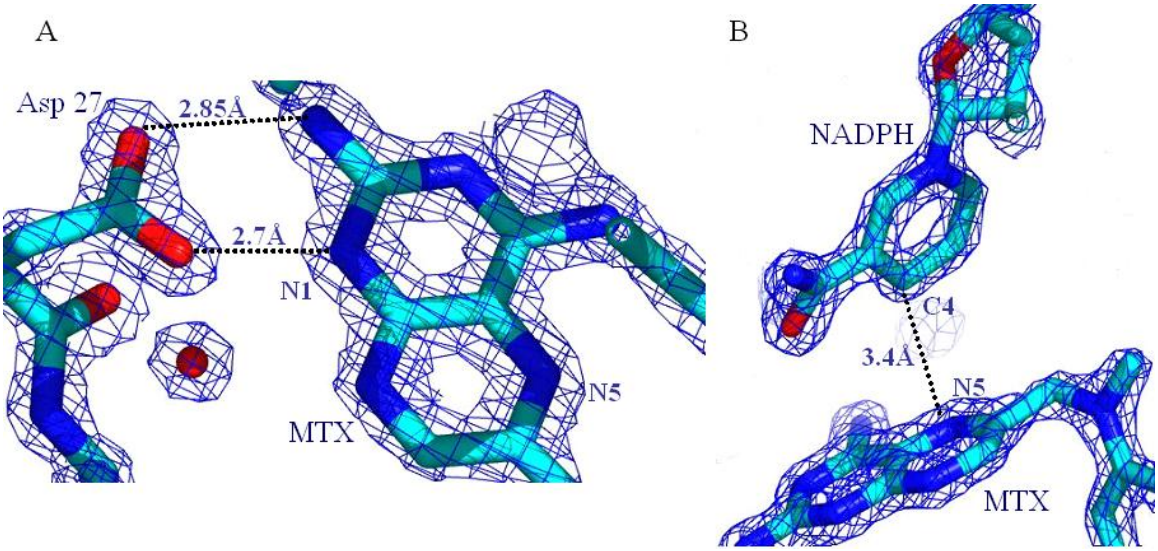


Figure 5.9: Interactions at the active site of the DHFR MTX/NADPH ternary complex. A) Contacts between the Asp 27 carboxylate and the pteridine ring of MTX. **B)** The flipped orientation of the MTX pteridine ring compared to folate results in a slightly larger distance (3.4Å) between the NADPH nicotinamide C4 atom and the MTX N5 atom (3.2Å for the folate N5 to the cofactor C4) and creates a vacancy in the position where the folate C6 would be. This, along with the position of the MTX pABG moiety, has the effect of pushing the αB helix (not shown) and providing more space for the nicotinamide ring of the cofactor. Essentially, less strain allows more optimal binding for both ligands. This arrangement has been suggested to be mimicked in the transition state once the DHF N5 atom has been protonated (Bystroff *et al.* 1990) (Bystroff *et al.* 1991), facilitating hydride transfer by reducing the distance between the cofactor and the DHF C6 atom (Wu 1987). Of course, the DHF would be flipped about its pteridine ring as compared to MTX.

dead-end complex due to the inhibitory nature of MTX; it is not expected that the NADPH would significantly alter or weaken MTX binding to DHFR. In fact, it has been reported that there is a tremendous positive cooperativity between MTX and NADPH for DHFR binding (Bystroff *et al.* 1990). The reason given for this phenomenon is that the MTX pteridine ring binds DHFR in a flipped orientation as compared to the weak substrate, folate (Reyes *et al.* 1995). This reduces the possible overlap between the nicotinamide ring of NADPH and the pteridine ring of MTX, affording less strain and steric clashing in the binding sites. This creates space where the folate C6 has been observed to be positioned in crystal structures of DHFR bound to it and to $NADP^+$ (Bystroff *et al.* 1990). It has been suggested that the conformation of the enzyme closely resembles a transition state structure when MTX is bound in the active site and the Met20 loop is closed (Bystroff *et al.* 1991). Indeed, if one superposes the DHFR/MTX/NADPH structure (1RH3) onto the DHFR/folate/$NADP^+$ structure (1RX2), the nicotinamide C4 from the former structure and the folate C6 of the latter structure are only ~2.9Å distant from one another (Sawaya *et al.* 1997). From theoretical calculations on the distance-dependence of the efficiency of hydride transfer reactions, it was revealed that 2.6-2.7Å is an ideal transfer distance (Wu 1987). These details seem to strengthen the argument that the binding of MTX and NADPH induce the protein to adopt a transition state-like structure and the position of the nicotinamide ring in the crystal structure of MTX/NADPH ternary complex is identical or very near to its position concurrent or just prior to hydride transfer. Even though the MTX/NADPH ternary complex crystallized in a pH 5.5 buffer (0.1 M Bis-Tris), it is not expected that this will titrate the Asp 27 carboxylate so it can maintain an ionic interaction with the MTX N1 atom. Recent NMR

results from the homologous *L. casei* DHFR in a binary complex with folate and a ternary complex with folate/$NADP^+$ revealed that the pKa for the active site Asp is not greater than pH 4.8 and is likely less than that (Casarotto *et al.* 1999). This seems to agree with results from NMR spectroscopy on the MTX complex with the *L. casei* enzyme that the Asp27•MTX interaction is a charge-charge contact (Gargaro *et al.* 1998), and Raman difference spectra on the *E. coli* enzyme in different liganded forms that provide evidence that the Asp 27 remains ionized at least between pH values of 5-9 (Chen *et al.* 1994; Chen *et al.* 1997).

Anisotropy analysis: the MTX/NADPH ternary complex

Analysis of the B-factor distribution and the anisotropy of the MTX/NADPH were performed using PARVATI (Merritt 1999a) as described previously (**Figure 5.10**). Immediately noticeable was that the overall B-factor for the protein atoms was higher (B_{iso} = 26.7Å^2) and nearly all surface residue atoms are highly anisotropic (lower *A* values) as compared to the apo and the MTX structures. However, the mean *A* value for protein atoms was 0.57, slightly higher than might be expected given the model shown in **Figure 5.10 C**. There is a general correlation between higher B_{iso} and lower *A* values (such as the adenosine-binding and the βF-βG loops), however, there existed regions such as the Met20 loop and the random coil region between αE and βE (residues 80-95) where this correlation breaks down (**Figure 5.10 A, B**). It could be that anisotropy was overestimated in the maximum likelihood refinement, and this could be due to the limitations of the data and the rather small data-to-parameter ratio for the anisotropic refinement (~3:1). It may be best to apply more stringent restraints on the anisotropic

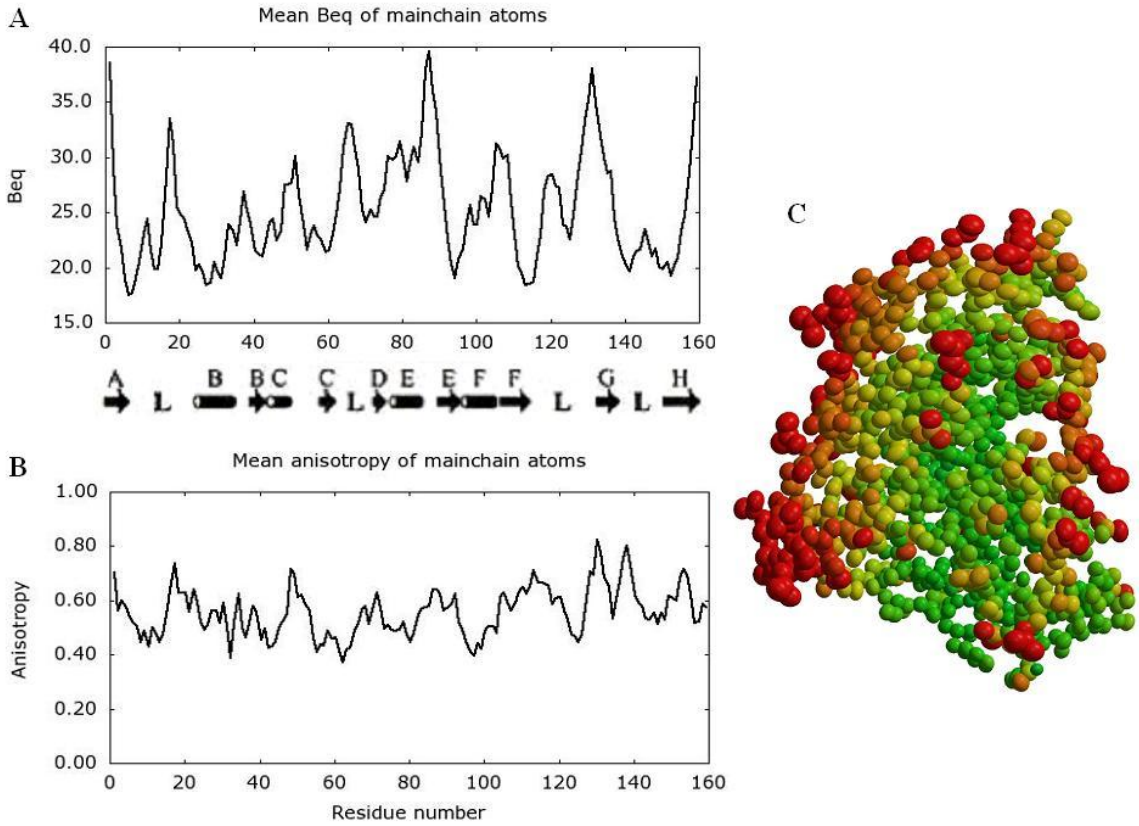


Figure 5.10: B-factor and anisotropy distribution for the ternary complex of DHFR bound to MTX and NADPH. A) B-factor values (in Å^2) for the backbone atoms of the 1.4Å DHFR/MTX/NADPH ternary complex. **B)** Anisotropic ratio ($A = E_{min} / E_{max}$) for the backbone atoms; the lower the A value, the more anisotropic an atom is (its vibrational motion cannot be described as spherical) **C)** MTX/NADPH model showing all protein atoms represented as thermal ellipsoids. Atoms exhibiting non-spherical vibrational motion are more elongated and ellipsoidal and are shown in yellow>orange>red (higher>lower A). Atoms exhibiting displacement more representative of isotropic motion are shown as spheres (less ellipsoidal) and are colored in green>blue (lower>higher A). In contrast to other structures presented in this thesis, most surface atoms in the MTX/NADPH structure have very high mobility and low A ratios.

refinement in Refmac5 or to attempt refinement in SHELX (Sheldrick *et al.* 1997), where further restraints can be implemented and modified as needed.

Chapter 6. Conclusions and future objectives

6.1 Conclusions from the present body of research results

This work provides new information about DHFR and its interactions with the anticancer drug, methotrexate (MTX), its overall dynamics, and the protonation state of Asp 27 by using a complementary X-ray and neutron crystallographic (NC) approach.

Preparation for NC experiments: crystal size and deuterium incorporation

Atypical as a prerequisite for most protein X-ray crystallography experiments is the growth of large volume crystals, those with one or more dimensions that exceed 1 mm and volumes that exceed 0.5-1.0mm^3. This, however, has traditionally been a necessity for NC. For DHFR/MTX crystals that were successfully utilized here, the volumes were ~0.2-0.3 mm^3. This is a significant decrease in the typical crystal volumes utilized in past macromolecular NC experiments. To improve the DHFR/MTX crystal size, a number of parameters were changed including using higher protein complex concentrations, decreasing the precipitant concentration, and switching to a microbatch-under-oil crystallization strategy (Rayment 2002) (Chayen 1997). This produced several crystals with volumes that approached and exceeded 0.2 mm^3; two of these diffracted neutrons beyond 2.2Å resolution. It should be restated that these crystal volumes are among the smallest to ever be reported to be successfully used in NC. They also possess the largest primitive unit cell (~6 x 10^5 Å) successfully tested. The D_2O-soaked crystal from the microbatch-under-oil set-up that was used in the NC data collection at Los Alamos provided moderately high resolution data (to 2.2Å) to sufficient completeness (~80%) in a decent amount of time (~22 days) from one 0.3 mm^3 crystal.

Prior to the NC experiments, the crystals were extensively soaked in D_2O-based crystallization buffer to substitute deuterium (D) for hydrogen (H) at chemically exchangeable positions (such as backbone amide nitrogens and side chain hydroxyl groups). This is done to reduce the incoherent neutron scattering from H in the crystal. Deuteration not only improves the signal-to-noise significantly in NC experiments but also aids in the *direct* observation and identification of labile H atoms (exchanged for D) in nuclear density maps. A more extensive form of deuterium labeling is perdeuteration and has been applied to *E. coli* DHFR. This is where H atoms at chemically non-exchangeable positions are substituted for D; this must occur within biosynthetic pathways of the cell and the growth medium is > 95% deuterated. This is especially important for labeling of all the aliphatic carbons in proteins. *E. coli* DHFR was expressed and purified from deuterium-adapted SK383 *E. coli* cells. Exhaustive adaptation of the bacterial strain to the deuterated background required about two weeks although adaptation is expedited to 2-3 days using more common strains (like BL21) transformed with pET and JM109 vectors. Deuterium enrichment of DHFR was nearly 100%, as measured by mass spectrometry. Perdeuterated DHFR was cocrystallized with MTX and ultrahigh resolution X-ray diffraction data (to 1.2Å) was collected at the APS.

Effects of perdeuteration on proteins: structure and function (specifically for DHFR)

Although very few examples exist in the literature, the three-dimensional structures of perdeuterated proteins do not appear to vary significantly from their native counterparts (Gamble *et al.* 1994) (Meilleur *et al.* 2005). Functional consequences of enzymes which have been perdeuterated are normally slight to moderate decreases in the catalytic rate (k_{cat}) and stability in solution, the latter demonstrated usually by calorimetry

(Hattori *et al.* 1965) (Rokop *et al.* 1969) (Brockwell *et al.* 2001) but also recently by NOE data from NMR (Mok *et al.* 1999) and infrared spectroscopy (Meilleur *et al.* 2004).

Deuterium solvent isotope effects have been measured for *E. coli* DHFR by Stone and Morrison (1988), and it was determined that turnover was limited by a combination of D_2O effects *and* by product release. However, the overall rate was only slightly decreased for the enzyme in D_2O; the kinetic pK_a in D_2O associated with a deuterium isotope effect using NADPD as the cofactor for hydride transfer was shifted significantly (~0.6 pH units) (Stone *et al.* 1988). The meaning of this shift is not well understood since now associating the kinetic pK_a with titration of Asp 27 is controversial. Although still a matter of dispute, it has been ascribed to protonation/deprotonation events on the DHF substrate itself (Chen *et al.* 1994) (Cummins *et al.* 2001) (Rajagopalan *et al.* 2002). For the perdeuterated DHFR/MTX data, the model and the electron density have not been examined, however, the space group is identical and the unit cell constants are isomorphous to the native DHFR/MTX cocrystals. Eventually, the model could be used as the starting coordinate set in refinement against future NC data collected on a perdeuterated DHFR/MTX crystal.

Crystallographic characterization of the DHFR/MTX complex

Prior to the NC experiments on DHFR/MTX, ultrahigh resolution X-ray data had been collected at the APS, and the structure was refined against data to 1.0Å resolution. Correlation was observed between anisotropic displacement and main chain B-factor values in the structure: the more anisotropic the behavior of atoms the higher their associated B-factor. An atom exhibiting anisotropic behavior means it has a tendency to vibrate with non-spherical (non-isotropic) displacements. There are 2 monomers in the

asymmetric unit (AU) of the DHFR/MTX crystals. Although their Cα r.m.s.d. is less than 0.5Å, striking differences for the main chain occur at the Met20 loop (residues 9-24) and at the core residues (118-123) of the supporting βF-βG loop (the whole loop comprises residues 116-132). In monomer A, where the Met20 loop is partially occluding the cofactor-binding site, residues in the adenosine-binding loop (residues 62-70) and at the beginning and core of the βF-βG loop exhibit both higher B-factors and anisotropic behavior. In monomer B, to contrast, the Met20 loop is closed over the active site. Its C-terminal region (residues 132-159) possesses higher B-factors and more anisotropic behavior than does the same region in monomer A. This could be due to the fact that the βF-βG loop must be positioned to hydrogen bond with the core of the Met20 loop when it is closed, possibly rigidifying that region of the protein. This stabilization is unnecessary in monomer A due to the fact that the Met20 loop adopts an occluded conformation and is directed more into the active site and away from the βF-βG loop, disrupting the hydrogen-bonding pattern between these two loops. So, it may be expected that the C-terminal region (residues 132-159) in DHFR would be more dynamic when the Met20 loop is closed (as in monomer B) because the βG-βH loop (residues 142-150) is not as important for stabilizing the Met20 loop in this conformation. When the Met20 loop occludes into the cofactor binding site (presumed to occur after catalysis to expedite exchange of $NADP^+$ for NADPH and release of the THF product), the βF-βG loop is no longer close enough to anchor the Met20 loop and, thus, the βG-βH loop is required to stabilize the C-terminal part of the Met20 loop (at Asn 23). These contacts are strengthened when the Met20 loop occludes because the βF-βG loop contacts are lost (Sawaya *et al.* 1997). As a consequence, this could reduce the flexibility of the C-

terminal region in monomer A, reflected by its lower B-factors and less anisotropic behavior compared to monomer B.

Protonation states of Asp 27 and MTX

In an attempt to clarify the possible protonation state of the Asp 27 and due to the high data-to-parameter ratio (>10:1) afforded by the ultrahigh resolution, full-matrix unrestrained refinement of carboxylate bond lengths was employed. This allows the carboxylate group to freely refine and all off-diagonal terms of the refinement parameter matrix are utilized in the calculation. It provides precise and accurate bond lengths for the Asp Cγ-Oδ1 and Cγ-Oδ2 bond lengths and their estimated standard deviations (ESDs) and allows comparison of the average bond length difference to the deviations (Deacon *et al.* 1997). For both monomers of the AU, the Cγ-Oδ1 and Cγ-Oδ2 bond lengths are nearly equal and the ESDs are essentially identical to the average bond length difference; this implies that the double bond is shared across the carboxylate. This means that the Asp 27 in both monomers is deprotonated and negatively charged, at least at the pH (7.5) of the crystallization. In the related DHFR/folate crystal structure solved to 1.06Å resolution (crystallization pH = 8.0), it has been shown that the Asp 27 carboxylate bond lengths are nearly equal and, thus, the Asp 27 is charged when folate is bound (Dr. Anna Gardberg). So, when the substrate binding site is occupied, the Asp 27 maintains a negative charge. It has been speculated that the maintenance of a negative Asp 27 is responsible for polarization of bound DHF, shifting its electrons from bonds of the pteridine ring facing the Asp 27 to the N5-C6 bond on the opposite part of the ring, inducing protonation and hydride transfer to that part of the ring (Bajorath *et al.* 1991a) (Bajorath *et al.* 1991c) (Greatbanks *et al.* 1999). Is a deprotonated (charged) Asp 27

conserved in productive and/or non-productive ternary complexes, when both the substrate and cofactor binding sites are occupied? This is yet to be demonstrated directly, an important future goal of X-ray and NC studies of DHFR.

NC studies of DHFR/MTX

NC of DHFR/MTX was performed because one can *directly*, using nuclear density maps, identify H (D) and, thus, directly determine protonation states. A spallation NC data set was collected on a 0.3 mm^3 crystal with reflections extending beyond 2.0Å. Ultimately, the high resolution limit for refinement was set to 2.2Å. After initial positional, B-factor, and D occupancy refinement, the nuclear density maps were examined at the active site around Asp 27 and MTX. In both monomers, sufficient nuclear density ($\geq$1.5σ) existed at the N1 atom of MTX that was not observable at the Asp 27 carboxylate. Indeed, the refined occupancy value of a D atom bound to N1 was 1.0 for monomer A and 0.75 for monomer B. The X-ray and neutron analysis provided the protonation states of Asp 27 and MTX, respectively: Asp 27 is deprotonated and, thus, negatively charged while the MTX N1 is protonated and, thus, positively charged, defining the Asp 27•MTX N1 contact as an ionic interaction. Early ^{13}C NMR experiments by Coco *et al.* (1981) and UV difference spectroscopy experiments by Stone and Morrison (1983) suggested this to be the case. Over a wide pH range, the N1 atom of DHFR-bound MTX, which has a pK_a of 5.7 in solution, did not titrate, suggesting it remains positively charged at the active site (Coco *et al.* 1981) (Stone *et al.* 1983b). An ionic interaction had also been suggested from NMR results on a complex of MTX with *L. casei* DHFR (Gargaro *et al.* 1998) but had recently been challenged by a computational study on the *E. coli* DHFR/MTX complex (Cannon *et al.* 1997a), where

the interaction between Asp 27 and the MTX N1 was speculated to be a neutral dipole-dipole. Clearly, from the combined X-ray and neutron results presented here, this is not the case; the computational predictions were made using the closed Met20 loop monomer from a lower resolution structure (3DRC). An obvious implication from this comparison of experimental and theoretical findings is that *in silico* models and predictions require intensive testing using other methods to help validate the computational results. A problematic assumption made in many MD simulations is about protonation states and hydrogen positions: neutrons can provide clear and direct evidence of these by identification of H (D) in the nuclear density maps and by their occupancy refinement. NC structures can then provide "updated" coordinate sets for MD, with residue and ligand protonation states understood because of H (D) identification.

Interestingly, drug compounds that have been shown to have marked selectivity for bacterial or fungal instead of vertebrate DHFRs, such as TMP, the pyrrolo-quinazolines, the pyrrolo-pyrimidines, and the antimalarial WR9910 family, share an orientation similar to MTX when bound at the active site (Matthews *et al.* 1985) (Kuyper *et al.* 1996a; Kuyper *et al.* 1996b) (Li *et al.* 2000). No DHFR complex structure is available, but based on its potent inhibition kinetics and chemical structure analogous to WR9910, it is likely that the anti-protozoan drug pyrimethamine (PYR) shares this feature as well (Hekmat-Nejad *et al.* 1997). This shared orientation conserves interactions to the Asp 27 (or homologous acidic residue, *i.e.* Asp 26 in *L. casei*) that utilize the N1 and the NA2 atoms of their pyrimidine rings. Applying the X-ray and neutron results here for the *E. coli* DHFR/MTX complex to these other bacterial/drug complexes, it seems that it is a common theme for tight-binding DHFR inhibitor compounds to use freedom to

rotate their pteridine/pyrimidine rings in order to exploit the possibility of a salt bridge interaction between N1 and Asp 27. This is an orientation that is not adopted by folate in binary or ternary complexes with *E. coli* or human DHFR (Reyes *et al.* 1995) (Davies *et al.* 1990) or by DHF in a binary complex with *E. coli* DHFR (Sawaya *et al.* 1997). Is an Asp 27 (or homologous residue) to N1 ionic contact the major determinant of inhibitor specificity and affinity to DHFRs, especially when taken in the context of competing with substrates for binding? Current drug design efforts with DHFRs now seem to be focusing on using the compounds TMP, PYR, and WR9910 as templates to generate leads against all manner of microbial, protozoan, and viral species, including *Mycobacterium tuberculosis*, *Plasmodium falciparum*, and even HIV (Li *et al.* 2000) (Hekmat-Nejad *et al.* 1997) (Wiktor *et al.* 1999). Why these compounds have higher selectivity for particular species whereas MTX binds tightly to nearly every DHFR studied is more than likely due to contacts made away from the Asp 27•MTX interface. Indeed, MTX possesses a glutamic acid tail that is not found in TMP or the other above compounds listed. TMP and these other compounds terminate with a substituted benzoic acid moiety and WR9910 (and its analogs) even possesses an extended linker sequence between its pyrimidine and chlorobenzene ring, imparting a more lipophilic character that could be important for cell membrane permeation and interaction in the hydrophobic active sites of DHFRs.

DHFR dynamics as revealed by NC

The D_2O-based soak of the crystal functioned as an H/D exchange experiment performed on the DHFR/MTX complex in the crystalline state. In terms of the backbone amide nitrogen atoms, monomer B has a higher H/D exchange percentage (~70%) than

does monomer A (~60%). At first this might seem counterintuitive since monomer A has an occluded Met20 loop and one would think it would be the less rigid of the two structures, more amenable to exchange and permeation of the D_2O. However, the B-factor and anisotropy analysis of the ultrahigh resolution X-ray data does not bear this out, at least in the C-terminal region of DHFR. Recall that the B monomer, especially from the middle of the βF-βG loop to the C-terminus, has higher B-factors and more atoms which exhibit anisotropy as compared to the A monomer. This may be due to the fact that the Met20 loop is closed in the B monomer and the N-terminal portion of the βF-βG loop is needed for hydrogen-bonding contact to anchor the Met20 loop in this conformation. The C-terminus of monomer B is not required to fully stabilize the Met20 loop, and the movement of the βF-βG loop is slightly towards the Met20 loop and the active site, as compared to monomer A. This repositioning may help to expose the βG and βH strands and make them more accessible to solvent. All the backbone amides save 2 have H/D exchanged from residue 132 to the C-terminus in monomer B whereas only 14 of 28 have exchanged for this same region in monomer A. In fact, accessibility measurements reveal that this region in monomer B is 60% more accessible to solvent than in monomer A. It should be reiterated here that lattice packing and crystal contacts have not been considered yet in the accessibility measurements. It was also observed that >75% of exchangeable protons on functional groups of basic residues (His, Lys, Arg) had indeed exchanged for D atoms, as assessed from occupancy refinement values ($\geq$ 0.2) and directly from the nuclear density maps.

From sequence conservation, mutational analysis, and molecular dynamics simulations, Agarwal *et al.* have suggested a correlated network of residues in DHFR

critical for the hydride transfer step of the reaction (Agarwal *et al.* 2002). Not surprisingly, many of the strictly conserved residues are responsible for maintaining hydrogen-bonding contacts with ligands, such as Thr 113 and an acidic residue adjacent to the pteridine ring (like Asp 27 in *E. coli*, Asp 26 in *L. casei*, and Glu 30 in *H. sapiens*). However, several of these key residues were quite distant from the active site, like Tyr 100 (its side chain hydroxyl oxygen is ~6Å from the N5 proton acceptor on DHF (or ~4Å from the N8 atom of folate)), Phe 31 (its aromatic ring is ~5Å from the DHF N5) and Met 42 (its Cβ ~10Å from the DHF N5). Double mutant studies on Met 42 and Gly 121 revealed adverse non-additive effects to the hydride transfer rate (Agarwal *et al.* 2002). During the course of the MD simulations and as progress is made towards the transition state, some of the residues identified move slightly closer to one another and/or to ligand atoms (like a 0.6Å decrease between the Ile 14 Cδ and Tyr 100 OH atoms), suggesting this motion is critical to catalysis, possibly by aiding in decreasing donor-to-acceptor distances (e.g. C4 of NADPH to C6 of DHF). Many of the conserved, distal residues that have effects on catalysis are located within the Met20 (such as Ile 14 and Gly 15), adenosine-binding (Ser 63), and the βF-βG (Gly 121-Asp 122) loops. It is less surprising that distant residues have dramatic functional effects if one frames this in the context of the importance of Met20 loop conformation and interloop stabilization to aid in regulating ligand binding and enhance catalysis (Schnell *et al.* 2004). A closed Met20 loop, stabilized by hydrogen-bonding to the βF-βG loop, is necessary for sealing the active site and is proposed to be the conformation during the hydride transfer step (Sawaya *et al.* 1997). Critical residues distal from the active site would have to be positioned in a particular way to maintain van der Waals and hydrogen-bonding contacts

with other key residues; this suggests that there is coupled "promoting motions" exploited by DHFR to enhance its catalytic rate (Agarwal *et al.* 2002) (Rajagopalan *et al.* 2002). Recent ^{15}N relaxation kinetics from NMR have also implicated the Met20 and the C-terminal loops (the βF-βG and βG-βH) to be major regions of change when the Met20 loop fluctuates from a closed to an occluded conformation (McElheny *et al.* 2005). If one allows application of increased H/D exchange propensity as a symptom of increased areas of mobility, the MD and NMR results can be corroborated with the neutron results here: nearly 75% of the amides in the Met20, the βF-βG and the βG-βH loop regions have exchanged in either monomer. However, in the closed Met20 loop monomer (monomer B), the core of the βF-βG loop has exchanged less than in monomer A, further suggesting its importance in closed Met20 loop stabilization. So, many of the loop residues implicated to have functionally important correlated motion identified from the work of (Agarwal *et al.* 2002), (Venkitakrishnan *et al.* 2004), and (McElheny *et al.* 2005) are supported by the NC and X-ray results here; however, a general correlation between the dynamics data here to the entire promoting motions network (involving non-loop residues like Met42 and Tyr 100) results from (Agarwal *et al.* 2002) is as of yet inconclusive.

Crystallographic characterization of apo DHFR

Ultrahigh (to 1.05Å) and high resolution (to 1.4Å) X-ray crystallographic studies have been performed on apo (unliganded) DHFR and a ternary complex of DHFR bound to MTX and NADPH, respectively. B-factor and anisotropy analysis revealed that the apo DHFR structure was surprisingly quite rigid and most of its atoms could be described isotropically. Comparison to the previously reported apo DHFR structure (5DFR; $P3_121$; Cα r.m.s.d. = 0.6Å; (Bystroff *et al.* 1991)) revealed significant (≥ 1.0Å) main chain

differences at the adenosine-binding loop (residues 62-70) and in the small randomly structured region between βE and αF (residues 94-97). In 5DFR, no solvent molecules were reported to be in contact distance with the Asp 27 carboxylate. In the present apo structure, 4 water molecules are of varying hydrogen-bonding distance to the Asp 27 carboxylate. Additionally, a possible hydrogen bond also exists between the Oγ atom of Thr 113 and the Oδ1 atom of Asp 27. This will be discussed later.

A conformation for the Met20 loop when the DHFR active site is vacant

From Met 16 through Met 20, there are no coordinates in 5DFR, but in the apo structure reported here these residues could readily be modeled into electron density.

The core of the Met20 loop in the present apo structure adopts a closed conformation, somewhat more closed over the active site than what was seen in monomer B of the DHFR/MTX structure. It seems that what stabilizes this conformation is a Mg^{2+} ion with hexavalent coordination from a symmetry-related molecule. One of these bonds is to a carbonyl oxygen of Asn 18 (distance = 2.04Å). The stabilization of the Met20 loop by a metal ion was also observed in a crystal structure of DHFR bound to folinic acid (1JOM; also $P6_5$) save the coordinating metal was Ca^{2+} in 1JOM and it contacted the Glu 17 carboxylate (Lee *et al.* 1996). Comparison to a closed Met20 loop structure of DHFR bound to folate ($P6_1$; Cα r.m.s.d. = 0.7Å) revealed that the largest differences between the unliganded and the substrate-bound structure was at the Met20, the adenosine-binding, and the βF-βG loops. The N-terminal and the core residues of the apo Met20 loop (from 14-19) are pointed more inwards and towards the active site than in the DHFR-folate Met20 loop; positioning of the residues as they are in the apo structure when ligand was bound would cause steric clashes. The largest divergence in this region

occurs between the Cα atoms of Met 16-Glu 17-Asn 18, with an average separation distance of ~5.5Å. The major difference in Met20 loop positions by default causes a major difference in interactions with the βF-βG loop. In the apo structure, the carbonyl oxygen of Gly 15 is >10Å distant from the amide nitrogen of Asp 122 in the βF-βG loop; in the DHFR/folate structure, this is a hydrogen-bonding distance of 2.8Å. Thus, the stabilization of the "further" closed Met20 loop conformation in the apo structure so far can be explained by crystal contacts with a symmetry-related molecule (and not by interloop contacts), which includes the Mg^{2+} coordination site and hydrogen-bonding contributions from the βG-βH loop. The Met20 loop has been characterized to have a rate of conformational exchange between closed and occluded conformers at 2-40 s^{-1} (Schnell *et al.* 2004), and recent NMR relaxation measurements have revealed that, even when DHFR is bound to folate and $NADP^+$ and the Met20 loop would be closed, a small population of the Met20 loop remains in another "excited" conformation, presumably occluded (McElheny *et al.* 2005). Visualizing the loop when no ligand is present is interesting but it must be noted in the context of the dynamic nature of the loop itself and the crystalline lattice. The lattice in the apo structure here seems to be helping to clamp the Met20 loop closed. However, it is worthwhile to discuss the mechanistic possibility that the loop serves as a gate to the binding site for both cofactor *and* substrate, shutting more when no ligand is present and opening somewhat (~4-7Å) at the core residues (16-18) while remaining in the "normal" closed conformation to allow catalysis to occur unperturbed by bulk solvent, exogenous compounds, and/or ions.

Protonation state of Asp 27 in apo DHFR

A full matrix refinement was performed on the apo DHFR structure. The calculated distances for the Asp 27 Cγ-Oδ1 and the Cγ-Oδ2 bonds were far from equivalent (1.281 and 1.190Å, respectively), resulting in a bond length distance difference of 0.092Å. This is >3 times the calculated average ESD for these bond lengths. No mixture of protonated and deprotonated species exist for the Asp 27, and there is no evidence of disorder of the carboxylate group. Taken together, this provides evidence that the Asp 27 carboxylate is protonated at the Oδ1 atom (the Cγ-Oδ1 is a single bond, the Cγ-Oδ2 is a double bond) and, thus, neutral at the pH (7.5) of the crystallization. This result agrees with early observations from spectroscopic studies but contradicts more recent data gleaned from Raman difference spectroscopy (Chen *et al.* 1997) and molecular dynamics (Cannon *et al.* 1997a; Cannon *et al.* 1997b) on the *E. coli* apoenzyme, and data from ^{13}C NMR with the homologous *L. casei* apoenzyme (Casarotto *et al.* 1999), all of which indicate that the pK_a of the Asp 27 (Asp 26 in *L. casei*) is less than pH 5. These combined results suggest Asp 27 is deprotonated and charged at physiological pH. All of these experiments were performed either in solution or *in silico*; perhaps the crystal environment, especially with the Met20 loop in a more closed position and its Glu 17 side chain projecting into the substrate binding site, alters the chemical environment within the active site such that the pK_a of Asp 27 shifts dramatically enough to be favorably protonated. Global and local conformational fluctuations and allowance of side chain "flexibility" seem to be important factors in MD calculations of pK_a values for ionizable groups in proteins (Yang *et al.* 1993) (You *et al.* 1995) (Antosiewicz *et al.* 1996); this provides evidence that conformation influences residue titration behavior. A

major perturbance of an Asp pK_a has been noted for the aspartic proteases, where one of the Asp pair is maintained protonated and is most likely stabilized by hydrogen bonds to the other Asp in the active site but also to a surrounding net of Ser/Thr residues (the so-called "fireman's grip") (Veerapandian *et al.* 1992) (Coates *et al.* 2001). Discussed below, a conserved Thr residue is positioned adjacent to the Asp 27 in the DHFR active site and makes hydrogen-bonding contact with, specifically, the Oδ1 atom of the Asp carboxylate. This contact is observed in multiple ligand-bound forms of DHFR as well, however, it is at its closest distance in the apo structure reported here and in the folate-bound structure (**Figures 6.1** and **6.2**). The aforementioned molecular dynamics calculations by Cannon and Benkovic measured proton affinities of the Asp 27 carboxylate as the dielectric constant (ε) was decreased over a range of 20 to 4. It was found that the carboxylate form was destabilized over the dielectric "titration", the final calculated pK_a being as low as 2.5 at ε =20 and increasing to 4.1 when the ε approached 4 (Cannon *et al.* 1997a). This suggests that, in the apoenzyme at least, proton affinity for the carboxylate strengthens when the dielectric is low, even if the pK_a at this dielectric does not reflect true protonation. Antosiewicz et al. (1996) have reported that pK_a values calculated using an ε of ~20 are more accurate and closer to what is observed from experimental data than those pKa values calculated using a lower dielectric (Antosiewicz *et al.* 1996). This phenomenon is not well understood but its finding, along with the protonation state data reported here, casts doubt on the accuracy of the computational predictions by Cannon and Benkovic (Cannon *et al.* 1997a). With a closed Met20 loop (and possibly even if a hydrophobic ligand is bound in the active site), it could be that

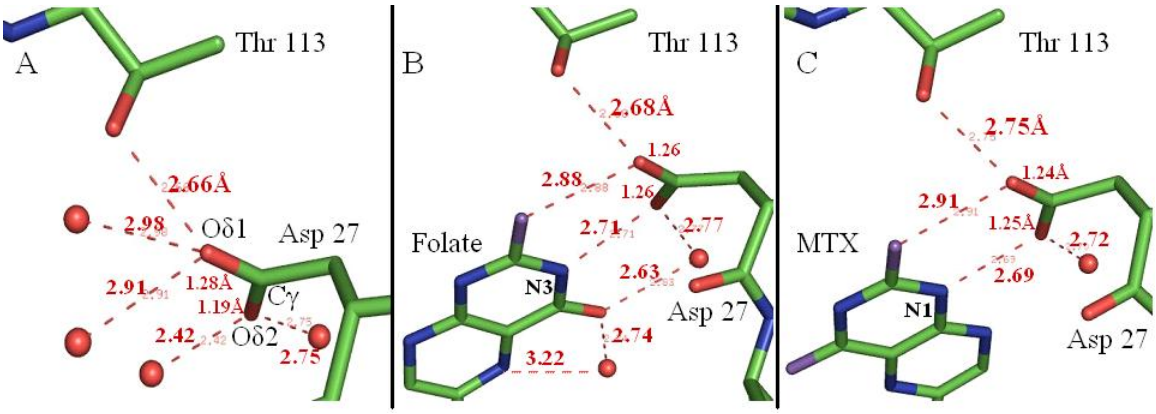


Figure 6.1: A comparison of hydrogen bonding patterns in the *E. coli* DHFR active site when in the A) apo (unliganded), B) folate-bound, and C) MTX-bound forms. The proton identified on the apo Asp 27 Oδ1 atom is possibly stabilized by three hydrogen-bonding contacts (two waters and Thr 113). The Asp 27-Thr 113 contact is slightly closer as compared to the folate-bound structure and significantly closer as compared to the MTX-bound structure.

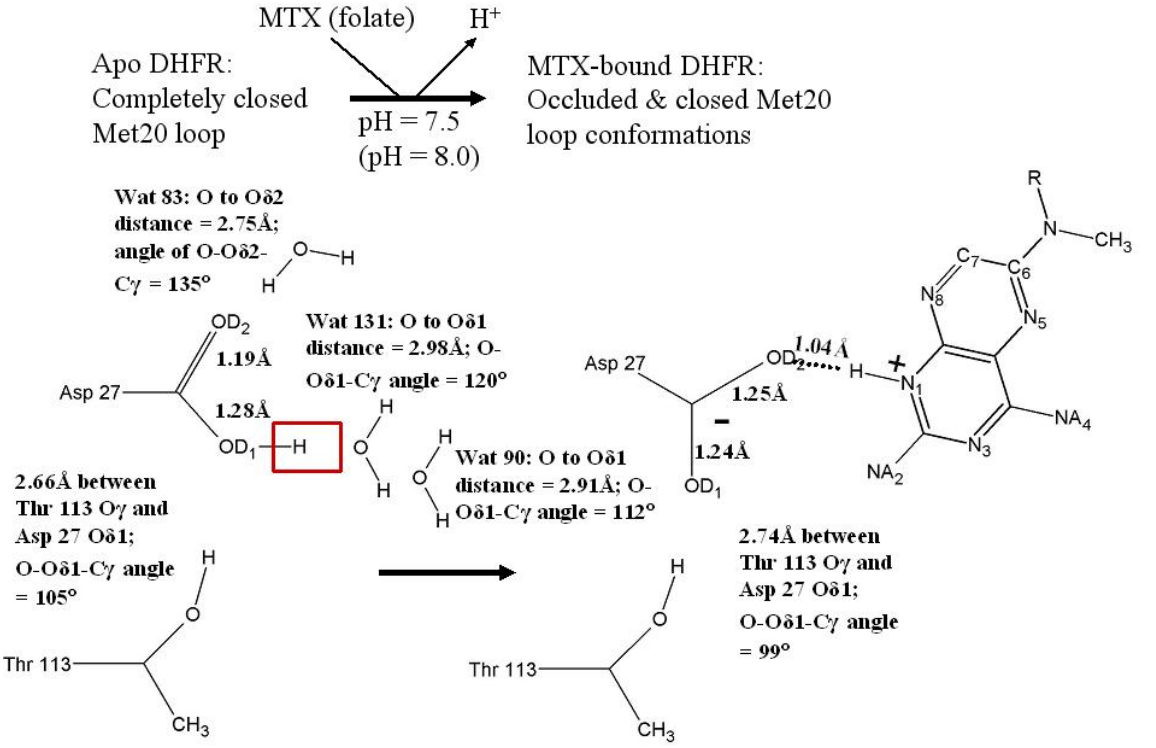


Figure 6.2: Deprotonation of Asp 27 must occur upon MTX (and folate) binding, as evidenced from crystallographic results reported here. The Asp 27 in the apo structure is protonated specifically at Oδ1, however, is deprotonated in binary ligand complexes. Therefore, transfer of the proton from Asp 27 (red box) to another group (possibly solvent) occurs concomitant with or soon after loop opening and/or ligand binding.

bulk solvent is excluded from the active site of apo DHFR, providing a more hydrophobic environment for the Asp 27, allowing it to have a perturbed pK_a and remain protonated. Solvent is definitely observed in the apo active site and seems to be in positions where ligand atoms would be if ligand was present. To assess the hydrophobicity of the active site, it would be pertinent to examine the active site of an apo DHFR structure with another Met20 loop conformation and/or from another space group. Unfortunately, the previously reported apo DHFR structure, although it crystallized in a different space group ($P3_221$ vs. $P6_5$), exhibits disorder for the core of the Met20 loop and has no solvent modeled near the Asp 27 residue (Bystroff *et al.* 1991). So, at this point, this comparison cannot be made.

Role of Thr 113 and implications of the apo results for the catalytic mechanism

However, we can attempt to understand this protonation/deprotonation phenomenon and cast our results in terms of the ligand binding and catalytic mechanism. In the apoenzyme, the "completely" closed conformation of the Met20 loop may be the main determinant on why the Asp 27 is protonated. What else could be constraining the protonation of the Asp 27 carboxylate? Possible hydrogen bonds can form between the Asp 27 and at least 3 water molecules. Most importantly, the Oγ atom from Thr 113 is within 2.66Å of the Asp 27 Oδ1 atom (**Figure 6.1 A**). The Cγ-Oδ1 bond distance as measured from full matrix refinement of the apo structure is 1.28Å, suggesting that it is a single bond and the Oδ1 is the oxygen specifically protonated in the carboxylate. Thus, the Thr 113 is most likely key in maintaining the protonation of the Asp 27 carboxylate *at Oδ1*, where the carboxylate hydrogen is utilized in the contact between the Oδ1 and the Oγ atoms. In other words, the Thr 113 may be *the* determinant for why the Oδ1 atom is

specifically protonated and not the Oδ2. Also, as stated earlier, there is no evidence of proton switching in the carboxylate as indicated from lack of a mixture of protonated/deprotonated species for the carboxylate and no evidence of disorder. Mutational studies where Thr 113 is replaced by a Val revealed its importance: binding affinity for substrates and the rate of hydride transfer was reduced (Fierke *et al.* 1989). When ligand is present, it must help to stabilize the Asp 27 by being a hydrogen bond donor (using its own hydroxyl hydrogen directly in the bonding interaction), and then plays a similar role when no ligand is bound and/or if the environment in the active site becomes favorable for the carboxylate to be neutral (e.g. hydrophobic); the apo structure refinement suggests that its Oγ atom may be playing the role of hydrogen bond acceptor. The Thr 113 Oγ and the Asp 27 Oδ1 atoms are slightly closer (~0.9Å) in the apo structure than in the MTX –bound structure; this distance is identical in the folate-bound structure (**Figures 6.1** and **6.2**). The X-ray and neutron results for the Asp 27 in the MTX and in the folate-bound complexes reveal that it is deprotonated in these forms. Without ligand bound and with the Met20 loop sufficiently closed, the Asp 27 has a perturbed pK_a so that it is protonated, even at physiological pH. Upon loop opening and binding of MTX or folate (and probably DHF), the proton may shift from Oδ1 to Oδ2 on Asp 27 (**Figure 6.3 panel A**) to facilitate its titration and its transfer of the proton to another group, possibly relaying through water molecules (**Figure 6.3 panel B**). Conserved solvent in the DHFR active site has been identified as important for ligand binding and possibly for catalysis by several studies (Bystroff *et al.* 1990) (Reyes *et al.* 1995) (Cummins *et al.* 2001) (Shrimpton *et al.* 2002). The proton shift on the Asp 27 may be

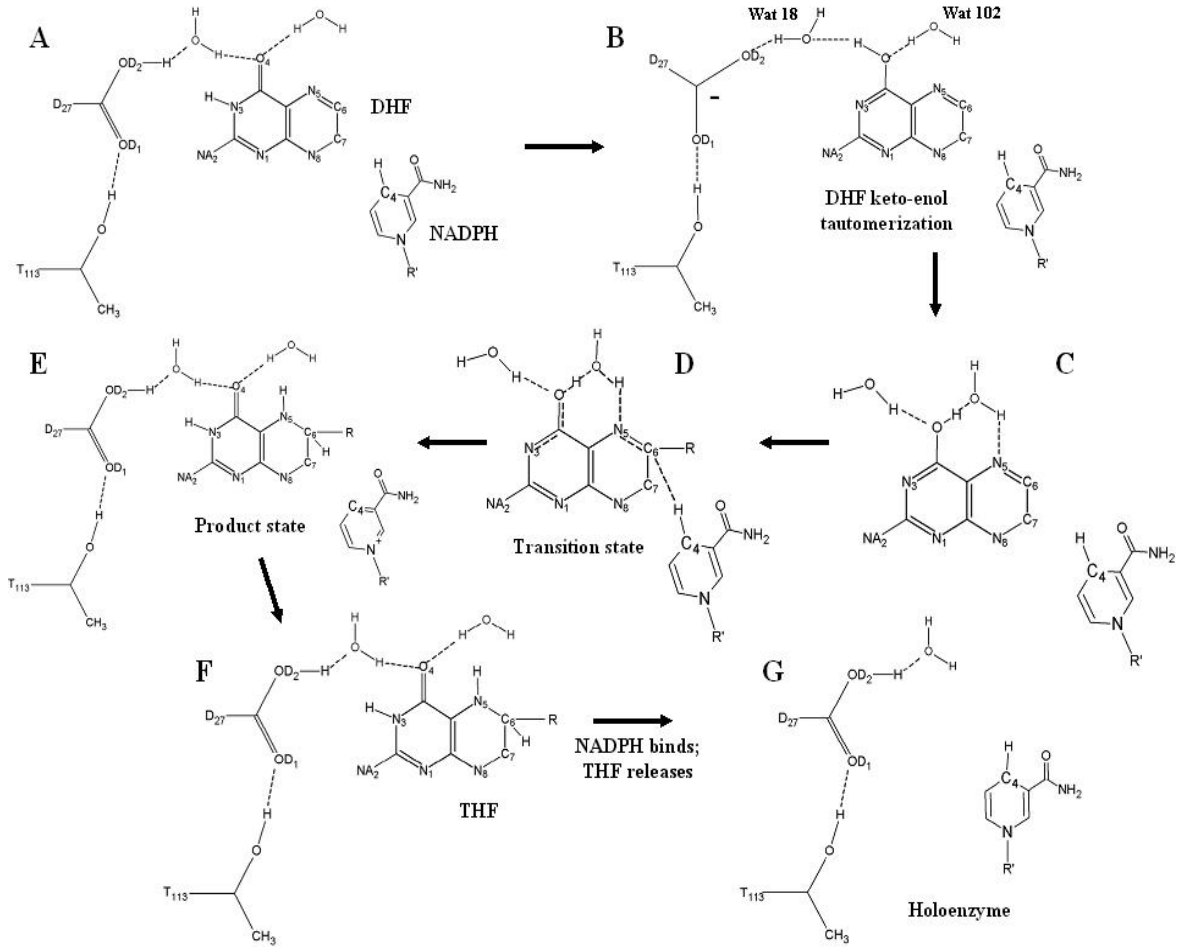


Figure 6.3: Hydrogen bonding patterns and possible proton relay events in the *E. coli* DHFR catalytic mechanism. The observance of a proton on Asp 27 in the apo structure with a closed Met20 loop may allow a mechanism in which direct proton transfer from Asp 27 to solvent plays a role in DHF reduction. **A)** Upon Met20 loop conformational change and substrate binding and to facilitate protonation of solvent adjacent to Asp 27 and also to the pteridine ring, the proton observed in the apo structure switches from Oδ1 to Oδ2 on the carboxylate. Protonation of Oδ2 has not been observed but has been implicated in previous mechanistic proposals (Bystroff *et al.* 1990) (Cummins *et al.* 2001). **B, C)** Transfer of a proton from Asp 27 to solvent may trigger DHF keto-enol tautomerization, producing an intermediate prior to attaining **D)** the transition state, the chemical step at which protonation of DHF N5 and hydride transfer to DHF C6 from NADPH occurs. **E)** THF and $NADP^+$ products bound in the active site and the Asp 27 is again protonated. **F)** The Met20 loop occludes, causing NADP+ to release. **G)** Formation of the holoenzyme occurs when a new NADPH cofactor binds, this helps to drive THF out of the active site; cofactor binding is concomitant with Met20 loop changing from occluded to closed conformations. Binding of DHF would begin a new cycle, as shown in **A**.

necessary so that it can be transferred to solvent which may participate in the proton relay pathway. However, there possibly would exist a type of ground state "intermediate" where Oδ2 is protonated but prior to solvent proton relay and DHF tautomerization (shown in **Figure 6.3, panel A**). The second step in the ground state would involve deprotonation of the carboxylate, resulting in a negative charge on the Asp 27; this charge on Asp 27 when MTX and folate is bound is inferred from bond lengths and deviations derived from full matrix refinement. Deprotonation of the Asp 27 would lead to protonation of solvent, possibly Wat 18 as shown in **Figure 6.3 panel B**, thus triggering proton relay (**panel C**) and ultimate reduction of DHF at N5 (transition state in **panel D**). Bystroff et al. (1990) proposed a catalytic mechanism which originated with a protonated Oδ2 atom on Asp 27, and this triggers keto-enol tautomerization on the DHF pteridine ring at the N3-C4-O4 group by initial proton transfer to a conserved water molecule (206) and then passage of the proton to the O4 atom (Bystroff *et al.* 1990). Ultimately, the proton is relayed to another water molecule before reduction of the DHF N5. Recent MD calculations by Cannon *et al.* (1997) suggest the kinetic pK_a of 6.5 should be ascribed to the tautomerization event and that Asp 27 becomes protonated simultaneous with tautomer formation (Cannon *et al.* 1997b). More recent *ab initio* quantum mechanical MD simulations revealed energetically favorable pathways for Asp 27 and substrate protonation states and the effects on catalysis (Cummins *et al.* 2001). Their results suggest that Asp 27 Oδ2 is protonated initially, with subsequent direct protonation of the DHF N5; therefore, this study argues against keto-enol tautomerization as a catalytic intermediate. The obvious important and contentious points with the mechanism proposed here is that: 1) in order to agree with previous proposals and

accommodate direct protonation of solvent near to the substrate, the proton on the Asp 27 must switch carboxylate oxygens (**Figure 6.3, panel A**), 2) the Asp 27 must donate a proton to some group, proposed here to be solvent, and 3) upon tautomerization, the N3 atom on DHF must become deprotonated but to which group this proton would transfer is not known. Again, it should be reiterated that the completely closed Met20 loop, stabilized in part by metal coordination, at the very least influences and may be the reason the Asp 27 is observed as protonated in the apo structure. The behavior of the Met20 loop in solution is fairly rapid conformational fluctuation between closed and occluded; does the Asp 27 remain protonated when no ligand is bound but the Met20 loop is more occluded or open?

6.2 Future directions

One of the long-term goals of this line of research is to be able to use X-ray crystallography to characterize an ensemble of *E. coli* DHFR structures at ultrahigh resolution in different liganded states. One critical question to address in each structure, if the data-to-parameter ratio is adequate, is the protonation state of the Asp 27. Additionally to full matrix calculations of bond lengths and in the absence of an NC structure, computational predictions of the Asp 27 pK_a could provide further evidence for the protonation states in the apo and the folate-bound structures. This would help to understand hydrogen-bonding patterns in the active site, especially between Asp 27, Thr 113, solvent in the apo structure. A major question is why specifically the Oδ1 is protonated in the Asp 27 carboxylate, as determined from the full matrix refinement of the apo structure. The in silico pK_a calculations would be performed using molecular dynamics methods as outlined in (Cannon *et al.* 1997a; Cannon *et al.* 1997b)

(Antosiewicz *et al.* 1996), and (You *et al.* 1995). A control is to perform this calculation on the Asp 27 in the MTX structure for which there already exists both X-ray and neutron evidence for the protonation state. An obvious missing piece in the present ensemble is a catalytic mimic structure such as a folate/$NADP^+$ ternary complex or other applicable substrate/cofactor combination. It would be clearly advantageous to complement each high resolution X-ray structure with a neutron structure. The experience gained by the laboratory with increasing crystal volume, D_2O-soaking, perdeuteration, LANSCE data collection, processing and refinement with the MTX complex and structure will be invaluable in the application to other forms, such as the folate binary complex and the apoenzyme. Once the crystal volumes could be increased slightly, a neutron experiment of perdeuterated DHFR bound to MTX could occur in the very near future. The steady partnership with LANSCE is now beginning to pay dividends, not only with the current DHFR/MTX neutron data, but now with their initiation of a deuterium labeling facility on-site. In fact, perdeuterated *E. coli* for the overexpression of DHFR have been produced at LANSCE and shipped to UTK for future purification and cocrystallization experiments.

Additionally, collaboration with Dr. Richard Lee at the University of Tennessee Health Sciences Center (Memphis) has provided derivatives of sulfamethoxazole drugs for kinetic study and cocrystallization. It has been demonstrated that at least one of these compounds has inhibitory activity against DHFR (data not shown); many of these compounds have been tested against a related bacterial enzyme, dihydropteroate synthase (DHPS). This could lead to several high resolution *E. coli* DHFR cocrystal structures and may provide a new template for drug design. A potential problem that has been observed

with cocrystallizing a ligand with DHFR other than MTX and folate is that MTX and folate are both used at some point in the DHFR purification procedure. Essentially, they "contaminate" the cocrystallization experiment by binding in the active site and preventing other ligands (such as new candidate inhibitor compounds) from being able to do so themselves. This problem has been addressed and solved utilizing an alternative purification procedure devised in the laboratory where an MTX affinity chromatography step is replaced by a metal affinity step. The DHFR is synthesized as a fusion protein in BL21(DE3) *E. coli* from a pET-Sumo vector (Invitrogen). A Sumo tag with an N-terminal hexahistidine tag is fused at the N-terminus of DHFR, and, after the metal affinity step, the Sumo tag may be cleaved specifically away right at the DHFR N-terminus to release native DHFR. After a second passage through the metal affinity column, the DHFR is >90% pure. The yield is ~6 mg l^{-1} and this is with little optimization of growth and expression conditions. Unless DHF or folate has bound to the enzyme within the cell, this preparation of DHFR should be free of "contaminating" ligands and satisfactory for cocrystallization experiments with new compounds.

Lastly, expression and purification of DHFR from *Bacillus anthracis* has been undertaken in the laboratory with the assistance of an undergraduate, Richard Simmerman. The gene had recently been cloned from *B. anthracis* at Oklahoma State University (Barrow *et al.* 2004). The short-term goals for this project are to crystallize the apoenzyme and also to cocrystallize the purified protein with MTX or folate and solve one or more of these structures; ultimately, it will be interesting to see how similar or divergent its structure is as compared to the *E. coli* and/or human forms of the enzyme. The above pET-Sumo system is being used for expression in BL21(DE3) *E. coli* cells.

Crystallization screens have already been attempted for the apoenzyme and the MTX complex. Functional characterization of the enzyme by kinetic assays, inhibitor screening, circular dichroism, and calorimetry are possible experimental avenues.

List of References

List of References

Agarwal, P. K., Billeter, S. R., Rajagopalan, P. T., Benkovic, S. J. & Hammes-Schiffer, S. (2002) *Proc Natl Acad Sci U S A* **99**, 2794-9.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. (1994) *Molecular Biology of The Cell* Garland Publishing, Inc. New York 1294

Antosiewicz, J., McCammon, J. & Gilson, M. (1996) *Biochemistry* **35**, 7819-7833.

Appleman, J., Howell, E., Kraut, J., Kuhl, M. & Blakley, R. (1988) *J Biol Chem* **263**, 9187-9198.

Appleman, J. R., Howell, E. E., Kraut, J., Kuhl, M. & Blakley, R. L. (1988) *J Biol Chem* **263**, 9187-98.

Arzt, S., Campbell, J. W., Harding, M. M., Hao, Q. & J.R.Helliwell. (1999) *J Appl Cryst* **32**, 554-562.

Azaroff, L. (1955) *Acta Crystallogr* **8**, 701-704.

Bajorath, J., Kitson, D. H., Fitzgerald, G., Andzelm, J., Kraut, J. & Hagler, A. T. (1991a) *Proteins* **9**, 217-24.

Bajorath, J., Kitson, D. H., Fitzgerald, G., Andzelm, J., Kraut, J. & Hagler, A. T. (1991) *Proteins* **9**, 217-24.

Bajorath, J., Kraut, J., Li, Z. Q., Kitson, D. H. & Hagler, A. T. (1991c) *Proc Natl Acad Sci U S A* **88**, 6423-6.

Barrow, E., Bourne, P. C. & Barrow, W. (2004) *Antimicrobial Agents and Chemotherapy* **48**, 4643-4649.

Bennett, B. C., Meilleur, F., Myles, D. A., Howell, E. E. & Dealwis, C. G. (2005) *Acta Crystallogr D Biol Crystallogr* **61**, 574-9.

Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 899-907.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) *Nucleic Acids Research* **28**, 235-242.

Blakeley, M. P., Kalb, A. J., Helliwell, J. R. & Myles, D. A. (2004) *Proc Natl Acad Sci U S A* **101**, 16405-10.

Blakley, R. L., Appleman, J. R., Freisheim, J. H. & Jablonsky, M. J. (1993) *Arch Biochem Biophys* **306**, 501-9.

Bolin, J., Filman, D., Matthews, D., Hamlin, R. & Kraut, J. (1982) *JBC* **257**, 13650.

Bradford, M. M. (1976) *Anal Biochem* **72**, 248-54.

Brockwell, D., Yu, L., Cooper, S., McCleland, S., Cooper, A., Attwood, D., Gaskell, S. J. & Barber, J. (2001) *Protein Sci* **10**, 572-80.

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. (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 905-21.

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. (1998) *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-21.

Brunger, A. T., Adams, P. D. & Rice, L. M. (1997) *Structure* **5**, 325-36.
Brunger, A. T., Adams, P. D. & Rice, L. M. (1999) *Prog Biophys Mol Biol* **72**, 135-55.
Burmeister, W. P. (2000) *Acta Crystallogr D Biol Crystallogr* **56 (Pt 3)**, 328-41.
Bystroff, C. & Kraut, J. (1991) *Biochemistry* **30**, 2227-39.
Bystroff, C., Oatley, S. J. & Kraut, J. (1990) *Biochemistry* **29**, 3263-77.
Cameron, C. & Benkovic, S. J. (1997) *Biochemistry* **36**, 15792-15800.
Campbell, J. W., Hao, Q., Harding, M. M., Nguti, N. D. & Wilkinson, C. (1998) *J. Appl. Cryst.* **31**, 496-502.
Cannon, W. R., Garrison, B. J. & Benkovic, S. J. (1997a) *J Mol Biol* **271**, 656-68.
Cannon, W. R., Garrison, B. J. & Benkovic, S. J. (1997) *J Mol Biol* **271**, 656-68.
Cannon, W. R., Garrison, B. J. & Benkovic, S. J. (1997b) *J Amer Chem Soc* **119**, 2386-2395.
Casarotto, M. G., Basran, J., Badii, R., Sze, K. H. & Roberts, G. C. (1999) *Biochemistry* **38**, 8038-44.
CCP4. (1994) in *Proceedings of the CCP4 Study Weekend*.
CCP4. (1985) *SERC, Daresbury Laboratory, Warrington, UK*.
CCP4, C. C. P. N. (1994) *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.
Chatake, T., Kurihara, K., Tanaka, I., Tsyba, I., Bau, R., Jenney, F. E., Jr., Adams, M. W. & Niimura, N. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 1364-73.
Chatake, T., Ostermann, A., Kurihara, K., Parak, F. G. & Niimura, N. (2003) *Proteins: Structure, Function, and Genetics* **50**, 516-523.
Chayen, N. E. (1997) *Structure* **5**, 1269-74.
Chen, Y. Q., Kraut, J., Blakley, R. L. & Callender, R. (1994) *Biochemistry* **33**, 7021-6.
Chen, Y. Q., Kraut, J. & Callender, R. (1997) *Biophys J* **72**, 936-41.
Cipriani, F., Castagna, J. C., Wilkinson, C., Oleinek, P. & Lehmann, M. S. (1996) *J. Neutron Res.* **4**, 79-85.
Coates, L., Erskine, P. T., Crump, M. P., Wood, S. P. & Cooper, J. B. (2002) *J Mol Biol* **318**, 1405-15.
Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. & Cooper, J. B. (2001) *Biochemistry* **40**, 13149-57.
Cocco, L., Groff, J., Temple, C., Jr, Montgomery, J., London, R., Matwiyoff, N. & Blakley, R. (1981) *Biochemistry* **20**, 3972-3978.
Cocco, L., Roth, B., Temple, C. J., Montgomery, J., London, R. & Blakley, R. (1983) *Arch Biochem Biophys* **226**, 567-577.
Coco, L., Groff, J., Temple, C., Jr., Montgomery, J., RE, L. & Blakley, R. (1981) *Biochemistry* **20**, 3972-3978.
Coco, L., Roth, B., Temple, C., Jr., Montgomery, J., RE, L. & Blakley, R. (1983) *Arch. Biochem. Biophys.* **226**, 567-577.
Cooper, J. B. & Myles, D. A. (2000) *Acta Crystallogr D Biol Crystallogr* **56**, 246-8.
Cooper, J. B. & Myles, D. A. (2000) *Acta Crystallogr D Biol Crystallogr* **56 (Pt 2)**, 246-8.
Cooper, S. J., Brockwell, D., Raftery, J., Attwood, D., Barber, J. & Helliwell, J. R. (1998) *J Chem Soc Chem Commun*, 1063-1064.
Cowtan, K. & Ten Eyck, L. (2000) *Acta Crystallogr D Biol Crystallogr* **56**, 842-856.

Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties* W.H. Freeman and Company New York 507
Crowther, R. (1972) in *The Molecular Replacement Method*. (M. Rossman, Ed.) pp 173-178, Gordon and Breach, Gordon and Breach
Crowther, R. & Blow, D. (1967) *Acta Crystallogr* **23**, 544-548.
Cummins, P. L. & Gready, J. E. (2001) *J Am Chem Soc* **123**, 3418-28.
Dauter, Z. (1997a) in *Macromolecular Crystallography part A*. (C. W. Carter and R. M. Sweet, Eds.). pp 326-344, Academic Press, Academic Press
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997b) *Curr Opin Struct Biol* **7**, 681-8.
Davies, J. F., 2nd, Delcamp, T. J., Prendergast, N. J., Ashford, V. A., Freisheim, J. H. & Kraut, J. (1990) *Biochemistry* **29**, 9467-79.
de la Fortelle, E. & Bricogne, G. (1997) in *Macromolecular Crystallography part A*. (C. W. Carter and R. M. Sweet, Eds.). pp 472-494, Academic Press, Academic Press
Deacon, A., Gleichmann, T., Kalb Gilboa, A. J., Price, H. J., Raftery, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997) *J Chem Soc, Faraday Trans* **93**, 4305-4312.
Demchuk, E. & Wade, R. (1996) *J Phys Chem* **100**, 17373-17387.
Drenth, J. (1999) *Principles of Protein X-Ray Crystallography* Springer New York City
El-Kabbani, O., Darmanin, C., Schneider, T. R., Hazemann, I., Ruiz, F., Oka, M., Joachimiak, A., Schulze-Briese, C., Tomizaki, T., Mitschler, A. & Podjarny, A. (2004) *Proteins* **55**, 805-13.
Emsley, P. & Cowtan, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32.
Engh, R. & Huber, R. (1991) *Acta Crystallogr A* **47**, 392-400.
Evans, P. (1997) in *Proceedings of the CCP4 Study Weekend*. (K. S. Wilson, G. Davies, A. W. Ashton and S. Bailey, Eds.). pp 97-102, Warrington, Warrington
Falzone, C., Wright, P. E. & Benkovic, S. J. (1994) *Biochemistry* **33**, 439-442.
Fersht, A. R. (1998) *Structure and Mechanism in Protein Science* W.H. Freeman and Company New York 631
Fierke, C. A. & Benkovic, S. J. (1989) *Biochemistry* **28**, 478-86.
Fierke, C. A., Johnson, K. A. & Benkovic, S. J. (1987) *Biochemistry* **26**, 4085-92.
Fox, G. & Holmes, K. (1966) *Acta Crystallogr* **20**, 886-889.
French, S. & Wilson, K. S. (1978) *Acta Crystallogr A* **34**, 517-525.
Gamble, T. R., Clauser, K. R. & Kossiakoff, A. A. (1994) *Biophys Chem* **53**, 15-25.
Gardner, K. H. & Kay, L. E. (1998) *Annu Rev Biophys Biomol Struct* **27**, 357-406.
Gargaro, A. R., Soteriou, A., Frenkiel, T. A., Bauer, C. J., Birdsall, B., Polshakov, V. I., Barsukov, I. L., Roberts, G. C. & Feeney, J. (1998) *J Mol Biol* **277**, 119-34.
Garman, E. (2003) *Curr Opin Struct Biol* **13**, 545-51.
Gewirth, D. (1999). pp 164,
Greatbanks, S. P., Gready, J. E., Limaye, A. C. & Rendell, A. P. (1999) *Proteins* **37**, 157-65.
Habash, J., Raftery, J., Nuttall, R., Price, H., Wilkinson, C., Kalb Gilboa, A. J. & Helliwell, J. (2000) *Acta Crystallogr D Biol Crystallogr* **56**, 541-550.
Habash, J., Raftery, J., Nuttall, R., Price, H. J., Wilkinson, C., Kalb, A. J. & Helliwell, J. R. (2000) *Acta Crystallogr D Biol Crystallogr* **56**, 541-50.

Habash, J., Raftery, J., Weisgerber, S., Cassetta, A., Lehmann, M. S., Hoghoj, P., Wilkinson, C., Campbell, J. & Helliwell, J. R. (1997) *J. Chem. Soc., Faraday Trans.* **93**, 4313-4317.
Hanson, B. L., Langan, P., Katz, A. K., Li, X., Harp, J. M., Glusker, J. P., Schoenborn, B. P. & Bunick, G. J. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 241-9.
Hattori, A., Crespi, H. L. & Katz, J. J. (1965) *Biochemistry* **4**, 1213-25.
Hattori, A., Crespi, H. L. & Katz, J. J. (1965) *Biochemistry* **4**, 1225-1238.
Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Dorsselaer, A. V., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005) *Acta Crystallogr D Biol Crystallogr* **D61**, 1413-1417.
Hekmat-Nejad, M. & Rathod, P. K. (1997) *Exp Parasitol* **87**, 222-8.
Helliwell, J. R., Habash, J., Cruickshank, D. W. J., Harding, M. M., Greenhough, T. J., Campell, J. W., Clifton, I. J., Elder, M., Machin, P. A., Papiz, M. Z. & Zurek, S. (1989) *J Appl Cryst* **22**, 483-497.
Hood, K. & Roberts, G. C. (1978) *Biochem. J.* **171**, 357-66.
Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004) *Proteins* **55**, 792-804.
Howell, E. E. (2005) *ChemBioChem* **6**, 590-600.
Howell, E. E., Villafranca, J. E., Warren, M. S., Oatley, S. J. & Kraut, J. (1986) *Science* **231**, 1123-8.
Hubbard, S. & Thornton, J. (1996).
Huennekens, F. M. (1994) *Adv Enzyme Regul* **34**, 397-419.
Huennekens, F. M. (1996) *Protein Sci* **5**, 1201-8.
Jancarik, J., and S.H. Kim. (1991) *J. Appl. Cryst.* **24**, 409-411.
Jancarik, J. & Kim, S. (1991) *J. Appl. Cryst.* **24**, 409-411.
Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000) *Proc Natl Acad Sci U S A* **97**, 3171-6.
Jones, T. A., Zou, J. Y., Cowman, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr A* **47**, 110-116.
Kleywegt, G. J. & Jones, T. A. (1997) in *Methods in Enzymology*. (C. W. Carter and R. M. Sweet, Eds.). Academic Press, Academic Press
Kleywegt, G. J. & Jones, T. A. (1996) *Acta Crystallogr D Biol Crystallogr* **52**, 826-8.
Klon, A., Heroux, A. & Borhani, D. (2002) *Journal of Molecular Biology*.
Knighton, D., Kan, C., Howland, E., Janson, C., Hostomska, Z., Welsh, K. & Matthews, D. A. (1994) *Nat Struct Biol* **1**, 186-194.
Korszun, Z. R. (1997) in *Methods in Enzymology*. (C. W. Carter and R. M. Sweet, Eds.). pp 218-232, Academic Press, Academic Press
Kossiakoff, A. A. (1982) *Nature* **296**, 713-721.
Kossiakoff, A. A. & Spencer, S. A. (1980) *Nature* **288**, 414-6.
Kossiakoff, A. A. & Spencer, S. A. (1981) *Biochemistry* **20**, 6462-74.
Kraut, J. & Matthews, D. A. (1987) *"DHFR" in Volume 3: Active Sites of Enzymes* John Wiley & Sons New York 1-71
Kurihara, K., Tanaka, I., Chatake, T., Adams, M. W., Jenney, F. E., Jr., Moiseeva, N., Bau, R. & Niimura, N. (2004) *Proc Natl Acad Sci U S A* **101**, 11215-20.

Kuyper, L. F., Baccanari, D. P., Jones, M. L., Hunter, R. N., Tansik, R. L., Joyner, S. S., Boytos, C. M., Rudolph, S. K., Knick, V., Wilson, H. R., Caddell, J. M., Friedman, H. S., Comley, J. C. & Stables, J. N. (1996a) *J Med Chem* **39**, 892-903.
Kuyper, L. F., Garvey, J. M., Baccanari, D. P., Champness, J. N., Stammers, D. K. & Beddell, C. R. (1996b) *Bioorg Med Chem* **4**, 593-602.
Ladd, M. F. C. & Palmer, R. A. (1993) *Structure Determination by X-ray Crystallography* Plenum Press London 586
Lamzin, V. S., Morris, R. J., Dauter, Z., Wilson, K. S. & Teeter, M. M. (1999) *J Biol Chem* **274**, 20753-5.
Langan, P. & Greene, G. (2004b) *Journal of Applied Crystallography* **37**, 253-257.
Langan, P., Greene, G. & Schoenborn, B. P. (2004a) *Journal of Applied Crystallography* **37**, 24-31.
Lee, H., Reyes, V. M. & Kraut, J. (1996) *Biochemistry* **35**, 7012-20.
Leiting, B., Marsilio, F. & O'Connell, J. F. (1998) *Anal Biochem* **265**, 351-5.
Li, L., Falzone, C., Wright, P. E. & Benkovic, S. J. (1992) *Biochemistry* **31**, 7826-7833.
Li, R., Sirawaraporn, R., Chitnumsub, P., Sirawaraporn, W., Wooden, J., Athappilly, F., Turley, S. & Hol, W. (2000) *J Mol Biol* **295**, 307-323.
Li, S. J. & Hochstrasser, M. (2003) *J Cell Biol* **160**, 1069-81.
Li, S. J. & Hochstrasser, M. (1999) *Nature* **398**, 246-51.
Li, X., Langan, P., Bau, R., Tsyba, I., Jenney, F. E., Jr., Adams, M. W. & Schoenborn, B. P. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 200-2.
Lifchitz, A. (1983) *Acta Crystallogr* **A39**, 130-139.
London, R. E., Howell, E. E., Warren, M. S., Kraut, J. & Blakley, R. L. (1986) *Biochemistry* **25**, 7229-35.
London, R. E., Howell, E. E., Warren, M. W., Kraut, J. & Blakley, R. L. (1986) *Biochemistry* **25**, 7229-7235.
Longhi, S., Czjzek, M. & Cambillau, C. (1998) *Curr Opin Struct Biol* **8**, 730-7.
Main, P. (1979) *Acta Crystallogr Part A* **35**, 779-785.
Matthews, D. A., Alden, R. A., Bolin, J. T., Freer, S. T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M. & Hoogsteen, K. (1977) *Science* **197**, 452-5.
Matthews, D. A., Bolin, J. T., Burridge, J. M., Filman, D. J., Volz, K. W., Kaufman, B. T., Beddell, C. R., Champness, J. N., Stammers, D. K. & Kraut, J. (1985) *J Biol Chem* **260**, 381-91.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005) *Acta Crystallogr D Biol Crystallogr* **61**, 458-64.
McElheny, D., Schnell, J. R., Lansing, J., Dyson, H. J. & Wright, P. E. (2005) *Proc Natl Acad Sci U S A* **102**, 5032-5037.
Meilleur, F., Contzen, J., Myles, D. A. & Jung, C. (2004) *Biochemistry* **43**, 8744-53.
Meilleur, F., Dauvergne, M. T., Schlichting, I. & Myles, D. A. (2005) *Acta Crystallogr D Biol Crystallogr* **61**, 539-44.
Merritt, E. A. (1999a) *Acta Crystallogr D Biol Crystallogr* **55 (Pt 6)**, 1109-17.
Merritt, E. A. (1999b) *Acta Crystallogr D Biol Crystallogr* **55 (Pt 12)**, 1997-2004.
Miller, G. & Benkovic, S. (1998) *Biochemistry* **37**, 6336-6342.
Miller, G., Wahnon, D. & Benkovic, S. (2001) *Biochemistry* **40**, 867-875.
Miller, G. P. & Benkovic, S. J. (1998b) *Biochemistry* **37**, 6336-42.

Miller, G. P. & Benkovic, S. J. (1998a) *Biochemistry* **37**, 6327-6335.
Minichino, A., Habash, J., Raftery, J. & Helliwell, J. R. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 843-9.
Moffat, K. (1997) *Methods Enzymol* **277**, 433-47.
Mok, Y. K., Kay, C. M., Kay, L. E. & Forman-Kay, J. (1999) *J Mol Biol* **289**, 619-38.
Murshudov, G., Vagin, A. & Dodson, E. (1997) *Acta Crystallogr D Biol Crystallogr* **D53**, 240-255.
Myles, D. A., Bon, C., Langan, P., Cipriani, F., Castagna, J. C., Lehmann, M. S. & Wilkinson, C. (1998) *Physica B* **241-243**, 1122-1130.
Myllykallio, H., Leduc, D., Filee, J. & Liebl, U. (2003) *Trends Microbiol* **11**, 220-3.
Myllykallio, H., Lipowski, G., Leduc, D., Filee, J., Forterre, P. & Leibl, U. (2002) *Science* **297**, 105-107.
Navaza, J. (2001) *Acta Crystallogr D Biol Crystallogr* **57**, 1367-72.
Niimura, N., Minezaki, Y., Nonaka, T., Castagna, J. C., Cipriani, F., Hoghoj, P., Lehmann, M. S. & Wilkinson, C. (1997) *Nat Struct Biol* **4**, 909-14.
Osborne, M., Schnell, J. R., Benkovic, S. J., Dyson, H. J. & Wright, P. E. (2001) *Biochemistry* **40**, 9846-9859.
Otwinowski, Z. & Minor, W. (1997) in *Macromolecular Crystallography part A*. (C. W. Carter and R. M. Sweet, Eds.). pp 307-326, Academic Press, Academic Press
Ozaki, Y., King, R. & Carey, P. (1981) *Biochemistry* **20**, 3219-25.
Paliy, O., Bloor, D., Brockwell, D., Gilbert, P. & Barber, J. (2003) *J Appl Microbiol* **94**, 580-6.
Pannu, N. S. & Read, R. J. (1996) *Acta Crystallogr A* **52**, 659-668.
Pflugrath, J. W. (1997) in *Macromolecular Crystallography Part A*. (R. M. Sweet, Ed.) pp 286-306, Academic Press, Academic Press
Poe, M., Greenfield, N., Hirshfield, J., Wiiliams, M. & Hoogsteen, K. (1972) *Biochemistry* **11**, 1023-1030.
Posner, B. A., Li, L., Bethell, R., Tsuji, T. & Benkovic, S. J. (1996) *Biochemistry* **35**, 1653-63.
Rajagopalan, P. T. & Benkovic, S. J. (2002) *Chem Rec* **2**, 24-36.
Rajagopalan, P. T., Zhang, Z., McCourt, L., Dwyer, M., Benkovic, S. J. & Hammes, G. G. (2002) *Proc Natl Acad Sci U S A* **99**, 13481-6.
Rayment, I. (2002) *Structure (Camb)* **10**, 147-51.
Read, R. J. (2001) *Acta Crystallogr D Biol Crystallogr* **57**, 1373-82.
Read, R. J. (1986) *Acta Crystallogr Part A* **42**, 140-149.
Ren, Z. & Moffat, K. (1995) *J Appl Cryst* **28**, 482-493.
Reyes, V. M., Sawaya, M. R., Brown, K. A. & Kraut, J. (1995) *Biochemistry* **34**, 2710-23.
Ringe, D. (2005) "Ultrahigh resolution X-ray structure of E. coli dihydrofolate reductase bound to folate and NADP+" in the Conference on Fronteirs in Macromolecular Neutron Crystallography, Oak Ridge, July 11, 2005.
Rod, T. H. & Brooks, C. L., 3rd. (2003) *J Am Chem Soc* **125**, 8718-9.
Rokop, S., Gajda, L., Parmerter, S., Crespi, H. L. & Katz, J. J. (1969) *Biochim Biophys Acta* **191**, 707-15.
Rossman, M. & Blow, D. (1962) *Acta Crystallogr* **15**, 24-31.

Ruiz, F., Hazemann, I., Mitschler, A., Joachimiak, A., Schneider, T. R., Karplus, M. & Podjarny, A. (2004) *Acta Crystallogr D Biol Crystallogr* **D60**, 1347-1354.
Sawaya, M. & Kraut, J. (1997) *Biochemistry* **36**, 586-603.
Sawaya, M. R. & Kraut, J. (1997) *Biochemistry* **36**, 586-603.
Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003) *J Biol Chem* **278**, 43357-62.
Schmidt, A. & Lamzin, V. S. (2005) *Acta Crystallogr D Biol Crystallogr* **61**, 1132-9.
Schnell, J., Dyson, H. & Wright, P. (2004) *Annu Rev Biophys Biomol Struct* **33**, 119-140.
Schnell, J. R., Dyson, H. J. & Wright, P. E. (2004) *Annu Rev Biophys Biomol Struct* **33**, 119-40.
Schoenborn, B. & Knott, R. (1996) in *Basic Life Sciences*. pp 452, Plenum Press, Plenum Press
Schoenborn, B. P. (1969) *Nature* **224**, 143-6.
Schoenborn, B. P. & Langan, P. (2004) *J Synchrotron Radiat* **11**, 80-2.
Sheldrick, G. & Schneider, T. (1997) in *Methods in Enzymology*. (C. W. Carter and R. M. Sweet, Eds.). Academic Press, Academic Press
Sheldrick, G. & Schneider, T. (1997) in *Macromolecular Crystallography Part B*. (R. M. Sweet, Ed.), Academic Press, Academic Press
Shrimpton, P. & Allemann, R. K. (2002) *Protein Sci* **11**, 1442-51.
Shu, F., Ramakrishnan, V. & Schoenborn, B. P. (2000) *Proc Natl Acad Sci U S A* **97**, 3872-7.
Stone, S. & Morrison, J. (1988) *Biochemistry* **27**, 5493-5499.
Stone, S. R., Montgomery, J. A. & Morrison, J. F. (1984) *Biochem Pharmacol* **33**, 175-9.
Stone, S. R. & Morrison, J. F. (1983a) *Biochim Biophys Acta* **745**, 237-46.
Stone, S. R. & Morrison, J. F. (1983b) *Biochim Biophys Acta* **745**, 247-58.
Stone, S. R. & Morrison, J. F. (1982) *Biochemistry* **21**, 3757-65.
Stone, S. R. & Morrison, J. F. (1988) *Biochemistry* **27**, 5499-5506.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 432-8.
Stout, G. H. & Jensen, L. H. (1989) *X-ray Structure Determination: A Practical Guide* John Wiley & Sons New York
Sukumar, N., Langan, P., Mathews, F., Jones, L., Thiyagarajan, P., Schoenborn, B. & Davidson, V. (2005) *Acta Crystallogr D Biol Crystallogr* **61**, 640-642.
Taira, K., Chen, J.-T., Mayer, R. & Benkovic, S. J. (1987) *Bulletin of the Chemical Society of Japan* **60**, 3017-3024.
Taylor, G. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 1881-90.
Tickle, I. J. (1992) in *Proceedings of the CCP4 Study Weekend*. (E. Dodson, S. Gover and W. Wolf, Eds.). pp 20-32,
Tollin, P. & Rossmann, M. G. (1966) *Acta Crystallogr* **21**, 872-6.
Trewhella, J., Gallagher, S. C., Krueger, J. K. & Zhao, J. (1998) *Sci Prog* **81 (Pt 2)**, 101-22.
Tronrud, D. E. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2156-68.
Tronrud, D. E. (1992) *Acta Crystallogr A* **48 (Pt 6)**, 912-6.
Trueblood, K., Burgi, H.-B., Burzlaff, H., Dunitz, J., Grammaccioli, C., Schulz, H., Shmueli, U. & Abrahams, S. (1996) *Acta Crystallogr A* **52**, 770-781.

Tuominen, V. U., Myles, D. A., Dauvergne, M. T., Lahti, R., Heikinheimo, P. & Goldman, A. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 606-9.
Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B. & Hoover, D. J. (1992) *Protein Sci* **1**, 322-8.
Venkitakrishnan, R., Zaborowski, E., McElheny, D., Benkovic, S. J., Dyson, H. J. & Wright, P. E. (2004) *Biochemistry* **43**, 16046-16055.
Villafranca, J. E., Howell, E. E., Voet, D. H., Strobel, M. S., Ogden, R. C., Abelson, J. N. & Kraut, J. (1983) *Science* **222**, 782-8.
Voet, D. H. & Voet, J. G. (1995) *Biochemistry* John Wiley and Sons New York City 1361
Warren, M. S., Brown, K. A., Farnum, M. F., Howell, E. E. & Kraut, J. (1991) *Biochemistry* **30**, 11092-103.
Wiktor, S. Z., Sassan-Morokro, M., Grant, A. D., Abouya, L., Karon, J. M., Maurice, C., Djomand, G., Ackah, A., Domoua, K., Kadio, A., Yapi, A., Combe, P., Tossou, O., Roels, T. H., Lackritz, E. M., Coulibaly, D., De Cock, K. M., Coulibaly, I. M. & Greenberg, A. E. (1999) *Lancet* **353**, 1469-75.
Wlodawer, A., Miller, M. & Sjolin, L. (1983) *Proc Natl Acad Sci U S A* **80**, 3628-31.
Wlodawer, A., Savage, H. & Dodson, G. (1989) *Acta Crystallogr B* **45**, 99-107.
Wu, Y. D. H., K. N. (1987) *Journal of the American Chemical Society* **109**, 906-8.
Yamamoto, T., Izumi, S. & Gekko, K. (2004a) *J Biochem (Tokyo)* **135**, 17-24.
Yamamoto, T., Izumi, S. & Gekko, K. (2004b) *J Biochem (Tokyo)* **135**, 663-671.
Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K. & Honig, B. (1993) *Proteins* **15**, 252-65.
Yang, X., Ren, Z. & Moffat, K. (1998) *Acta Crystallogr D Biol Crystallogr* **54 (Pt 3)**, 367-77.
Yeates, T. O. (1997) *Methods Enzymol* **276**, 344-58.
You, T. J. & Bashford, D. (1995) *Biophys J* **69**, 1721-33.
Zeig, J., Maples, V. & Kushner, S. (1978) *Journal of Bacteriology* **134**, 958-966.
Zuccotto, F., Martin, A. C., Laskowski, R. A., Thornton, J. M. & Gilbert, I. H. (1998) *J Comput Aided Mol Des* **12**, 241-57.

Vita

Brad Cooper Bennett was born in Chattanooga, TN on March 29, 1976 and was raised in the Harrison and Ooltewah communities just outside of Chattanooga. He graduated from Chattanooga Central High School in June 1994 and entered the University of Tennessee at Chattanooga (UTC) later that fall. He graduated with honors with a B.S in Biology from UTC in December 1998. There, he mainly studied plant taxonomy and physiology and performed independent research on the germination mechanism of *Phytolacca americana* (American Pokeweed) under the supervision of Dr. Maurice Edwards.

Brad began his graduate training in Biochemistry, Cellular and Molecular Biology at the University of Tennessee at Knoxville in August 1999 and joined the biochemistry and structural biology laboratory of Dr. Chris Dealwis in May 2000. He received his doctorate in December 2005. Brad is currently working in Dr. Dealwis' laboratory.