Towards Understanding Osmolyte Effects on Folate(s) and Dihydrofolate Reductase Proteins

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Purva Prashant Bhojane

December 2016

Dedication

This dissertation is dedicated to my father, Mr. Prashant Bhojane, for his never-ending love and support, my mother Mrs. Sandhya Bhojane for always being there for me, my younger brother, Mr. Abhishek Bhojane for making me laugh whenever I feel low and my fiancé, Mr. Shailander Rawat for loving me the way I am.

This dissertation is also dedicated to both my grandfathers, Mr. Sudhakar Bhojane and Mr. Sudhakar Joshi.

Acknowledgments

The first person I would like to acknowledge is my mentor, Dr. Liz Howell. I would like to express my sincere gratitude for giving me the opportunity to work in her lab and guiding me throughout my graduate studies. With her exceptional guidance and experience, I gained knowledge and valuable skills that helped me to achieve my goals. I not only learned to work in a research laboratory but also learned to be persistent in my efforts, which would help me develop into a scientific researcher. I appreciate her efforts to teach me the scientific way of thinking that asks for digging deeper. I will always look up to her for inspiration.

I am deeply grateful to my thesis committee members Dr. Engin Serpersu, Dr. Jerome Baudry, Dr. Christopher Stanley and Dr. Michael D. Best for their valuable suggestions and guidance. Next, I would like to extend my gratitude to our collaborator Dr. Pratul Agarwal for the support with his computational work on two of my research projects. I am also thankful to Dr. Susan Krueger from NIST for her help with the SANS data analysis using SASSIE.

Next, I would like to thank the members of the Howell lab. A special thanks to Mike, who helped me learn the techniques and understand the science. Our discussions facilitated my scientific thoughts. I would also like to thank my lab mates Timkhite, Deepika and Gabriel for being cordial. I am grateful to all the undergrads in the Howell lab and specially to Gabriella Rimmer for assisting me in my VPO project. I would also like to extend my gratitude to Khushboo Bafna for her help with the computational work.

One of the most important factors contributing to the successful completion of my graduate studies is my friends. I would like to thank my best friends back in India, Nicy,

Supriya, and Saheli who always stood by me in hard times and, all my friends in Knoxville, Vandna, Ishan and Mike for their constant support.

Last but not the least I would like to thank my beloved family for their love and support. I would not have been here without my parents Mr. Prashant Bhojane and Mrs. Sandhya Bhojane, my brother Abhishek and my fiancé Shailander. I can never thank them enough.

Abstract

Osmolytes are small molecules that alter water activity and probe role of water in biological processes. Osmotic stress approach explored the role of water in ligand binding to dihydrofolate reductase (DHFR). DHFR catalyzes NADPH dependent reduction of dihydrofolate (DHF) to tetrahydrofolate (THF), which is essential for the synthesis of DNA, amino acids and other metabolic intermediates. R67 DHFR is a plasmid-encoded DHFR that confers resistance against trimethoprim, which is a potent inhibitor of *E.coli* chromosomal DHFR.

Osmolytes addition decreases the affinity of the substrate towards both the DHFRs. Weak preferential interactions between the osmolytes and DHF impede substrate binding to the enzyme. Similar results were obtained for DHF binding to FolM, an *E. coli* enzyme which possesses weak DHFR activity. Binding of the cofactor to FolM was found to be tighter in presence of betaine but other osmolytes showed variable effects indicating interactions between FolM and osmolytes. Osmolytes (DMSO and ethylene glycol) showed decreased the stability of FolM further suggesting preferential interactions of osmolytes with the protein. Thus, ligand binding to FolM was hindered by interactions between osmolytes and the enzyme as well as the substrate.

Interaction potential (μ_{23}/RT value) of folate with betaine was quantified using a vapor pressure osmometry method. Folate interaction with betaine showed concentration dependence as folate dimerizes. A pH dependence owing to the deprotonation of folate's N3-O4 keto-enol group was also seen. The interaction of other heterocyclic aromatic compounds with betaine was monitored and deconvoluted into atomistic interaction potentials using an accessible surface area approach. Betaine preferentially interact with

aromatic surfaces, cationic and amide nitrogens whereas it is excluded from carboxylate oxygens and aromatic nitrogens. As folate contains a combination of surface types, the μ_{23}/RT value is predicted to be near zero, indicating folate interacts almost equally well with betaine and water.

Further, osmolyte effect on proteins was explored using SANS studies on R67 DHFR. The hydration studies yielded around 1200 water molecules excluding osmolytes from R67 DHFR surface. SANS also characterized the conformations sampled by the disordered tails of R67 DHFR under different conditions tested.

Table of Contents

PART 1. GENERAL INTRODUCTION TO OSMOLYTES AND THEIR EFFECTS ON PROTEIN AND PROTEIN PROCESSES..... 1

1.1 Water is Essential for Life 2

1.2 Mechanism of Osmolyte Action- Preferential Exclusion vs. Preferential Interaction..... 7

1.3 Osmotic Stress Studies Probe the Role of Water in Biological Processes..... 9

1.4 Introduction to Dihydrofolate Reductase (DHFR) 11

1.4.A. EcDHFR 12

1.4.B. R67 DHFR 12

1.4.C. A Novel DHFR: FolM 16

1.5 Osmotic Stress Studies 17

1.5.A. Osmotic Stress Studies on Ligand Binding to R67 DHFR 18

1.5.B. Osmotic Stress Studies on Ligand Binding to EcDHFR 22

1.6 Model for Preferential Interactional between DHF and Osmolytes 25

1.7 Do Osmolytes Weaken Substrate Binding to the Novel DHFR, FolM?..... 25

1.8 Folate – Osmolyte Interactions..... 27

1.8.A. Previous Studies using NMR and NOESY 29

1.8.B. Previous Osmotic Stress Studies with Folate 33

1.8.C. Vapor Pressure Osmometry (VPO) 35

1.8.D. Solubility Assays 42

1.9 How Strong are the Preferential Interactions between Folates and Osmolytes? 43

1.10 Further Osmotic Stress Studies on R67 DHFR 43

1.11 Small Angle Neutron Scattering (SANS) 44

1.12 Disordered Tails of R67 DHFR..... 48

1.13 Exploring Possible Role of the Disordered Tails 51

1.14 What is the Role of Disordered N-Termini of R67 DHFR? 53

1.15 References 56

PART 2. INVESTIGATION OF OSMOLYTE EFFECTS ON FOLM: COMPARISON WITH OTHER DIHYDROFOLATE REDUCTASES..... 61

2.1 Abstract..... 62

2.2 Introduction..... 63

2.3 Materials and Methods 68

2.3.A. Protein Expression *68*
2.3.B. Steady-State Kinetics *68*
2.3.C. Isothermal Titration Calorimetry (ITC) *69*
2.3.D. Ultracentrifugation *70*
2.3.E. Water Activity Measurements *70*
2.3.F. Differential Scanning Calorimetry *70*
2.3.G. Circular Dichroism *71*
2.3.H. Fluorescence Quenching *71*
2.3.I. Homology Modeling *72*
2.4 Results ***72***
2.4.A. Ultracentrifugation *72*
2.4.B. Homology modelling *74*
2.4.C. Stability *74*
2.4.D. Steady State Kinetics *78*
2.4.E. Osmolyte Effects *79*
2.5 Discussion ***97***
2.5.A. Betaine Effects on Ligand Binding *99*
2.5.B. Other Osmolytes Weaken Ligand Binding *103*
2.5.C. How do Osmolytes Alter Ligand Binding? *104*
2.5.D. Comparison of DHFR Enzymes *106*
2.6 Conclusion ***107***
2.7 References ***108***
PART 3. ASPECTS OF WEAK INTERACTIONS BETWEEN FOLATE AND GLYCINE BETAINE ***113***
3.1 Abstract ***114***
3.2 Introduction ***115***
3.3 Materials and Methods ***120***
3.3.A. Materials *120*
3.3.B. Vapor Pressure Osmometry (VPO) *121*
3.3.C. Folate Dimerization at pH 10 by NMR Spectroscopy *124*
3.3.D. Interaction of Betaine with Heterocyclic Test Compounds *124*
3.3.E. α Value Calculation by Analysis of μ_{23}/RT Values *129*
3.3.F. Solubility Assays *130*

3.3.G. Protein Purification *131*
3.3.H. Isothermal Titration Calorimetry (ITC) *132*
3.3.I. Simulation of Folate in Water *132*
3.3.J. Simulation of Folate in Betaine *133*
3.3.K. Comparison of $\Delta\mu_{23}$/RT Predictions of Ligand Binding with ITC Data *134*
3.4 Results ***135***
3.4.A. VPO Measurements of Folate at pH 7 ..... *135*
3.4.B. VPO Measurements of Folate at pH 10 ..... *138*
3.4.C. Folate Dimerization at pH 10 ..... *139*
3.4.D. VPO Measurements of Non-Heterocyclic Aromatic Compounds *139*
3.4.E. VPO Measurements of Compounds Containing Aromatic Nitrogen Atoms *142*
3.4.F. Pyridoxine-HCl *142*
3.4.G. Betaine-Nicotinic Acid Betaine-Imidazole and Interaction by VPO *150*
3.4.H. Analysis of μ_{23}/RT Values and Calculation of α Values *150*
3.4.I Solubility Assays *156*
3.4.J Binding of Folate and PG4 to R67 DHFR and EcDHFR *156*
3.4.K Prediction of Folate μ_{23}/RT Values from Simulation Data *160*
3.5 Discussion ***167***
3.5.A. Folate is an Interesting Molecule *167*
3.5.B. Deconvolution of μ_{23}/RT into α Values and K_p Values *169*
3.5.C. Solubility vs. VPO Assays *170*
3.5.D. Do these Results Provide any Insights into our Previous Studies where Betaine Weakens Binding of Folate to R67 DHFR and EcDHFR? *174*
3.6 Conclusion ***177***
3.7 References ***179***
PART 4. SMALL ANGLE NEUTRON SCATTERING (SANS) STUDIES ON R67 DHFR, A TETRAMERIC PROTEIN WITH INTRINSICALLY DISORDERED N-TERMINI ***183***
4.1 Abstract ***184***
4.2 Introduction ***185***
4.3 Methods ***189***
4.3.A. Protein Expression and Purification *189*
4.3.B. Small Angle Neutron Scattering (SANS) *189*

4.3.C. Analysis using MD and SASSIE *191*
4.3.D. Differential Scanning Calorimetry (DSC) *193*
4.4 Results ***194***
4.4.A. SANS Data Analysis for Apo R67 DHFR *194*
4.4.B. Predicting the Structures Associated with the R67 DHFR N-Termini using MD and SASSIE *194*
4.4.B Effect of Ligand Binding on the Disordered Termini in R67 DHFR *200*
4.4.C Effect of betaine on the disordered termini of R67 DHFR *211*
4.4.E. Osmolytes Probe Preferential Hydration of R67 DHFR *216*
4.4.E Effect of Osmolytes on the Thermal Stability of R67 DHFR *221*
4.5 Discussion ***225***
4.5.A Apo Protein Analysis *225*
4.5.B Analysis of Binary and Ternary Complexes *227*
4.5.C Effects of Osmolytes *227*
4.5.D Hydration Studies *229*
4.6 Conclusion ***232***
4.7 References ***235***
PART 5. CONCLUSIONS AND FUTURE DIRECTIONS ***238***
5.1 Role of Water in Biological Process ***239***
5.2 The Preferential Interaction Model ***240***
5.3 Characterization of Preferential Interactions ***241***
5.4 Can we use the Interaction Potentials to Predict the Osmotic Stress Effects In Vitro? ***243***
5.5 Osmolyte Effects on Proteins add a Layer of Complexity ***245***
5.6 SANS, a Structural Tool to Study Proteins and Osmolyte Effects ***247***
5.6. A Osmolyte Effects on Protein Hydration *247*
5.6.B SANS Studied the Disordered Tails of R67 DHFR *249*
5.7 Future Implications ***250***
5.8 How Relevant our In Vitro Studies are to Physiological In Vivo Systems? ***250***
5.9 References ***254***
VITA ***255***

List of Tables

Part 1

Table 1.1 - Comparison of structural and functional parameters for EcDHFR, R67 DHFR, PTR1 and FolM. Parameters obtained at pH 7.0 for EcDHFR and R67 DHFR and pH 6.0 for PTR1 and FolM. 15

Table 1.2. - Dimerization constants for folate obtained from NMR studies..... 32

Table 1.3 - Atomistic interaction potentials (α values) deconvoluted from the μ_{23}/RT values obtained for several test compounds using vapor pressure osmometry studies to quantify interactions with osmolytes..... 40

Table 1.4 - Comparison of kinetic parameters of full length and truncated R67 DHFR. Data is obtained from Reece., et al.[36] 50

Table 2.1 - Binding of NADPH to FolM in the presence of various osmolytes in MTA buffer, pH 6 at 25 °C. 82

Part 2

Table 2.2 - A comparison of the slopes for the ln (K_a) versus ln (a_o) plots describing cofactor binding to FolM, R67 DHFR and EcDHFR. 84

Table 2.3 - Steady state kinetic parameters for FolM obtained at pH 6, 30 °C. 89

Table 2.4 - A comparison of the slopes for the ln (K_a) versus ln (a_o) plots describing substrate binding to FolM, R67 DHFR and EcDHFR. 91

Table 2.5 - ITC measurements describing methotrexate binding to the FolM-NADPH binary complex in MTA buffer, pH 6.0 and 25 °C. 96

Table 2.6 - Global fitting of MTX and NADPH binding to FolM in MTA buffer, pH 6.0, at 25 ºC. 101

Part 3

Table 3.1 - A list of compounds tested for preferential interactions with betaine by the VPO method..... 125

Table 3.2 - Folate dimerization constants at pH 10. 141

Table 3.3 - A list of all test compounds with their experimental and predicted μ_{23}/RT values 146

Table 3.4 - A list of experimental and predicted μ23/RT values for test compounds in addition to compounds listed in Table 3.2. 152

Table 3.6 - Thermodynamic parameters for binding of folate and PG4 to the R67 DHFR-NADPH binary complex and EcDHFR by ITC..... 163

Table 3.7 - Calculation of K_p partition coefficient values for the different atom types. 173

Table 3.8 - Prediction of the $\Delta\mu_{23}$/RT values for betaine effects on the binding of ligands to two different dihydrofolate reductases 175

Part 4

Table 4.1 – Comparison of the radii of gyration of the full length R67 DHFR obtained upon analyzing the SANS data using GNOM and SASSIE modules..... 201

Table 4.2 - Comparison of the predicted and experimental numbers of water molecules (nw) hydrating R67 DHFR as obtained from the crystal structure and SANS data. 222

Table 4.3 - Comparison of melting temperatures of R67 DHFR with and without osmolytes 224

Part 5

Table 5.1 - Comparison of predicted and observed effects of betaine on ligand binding to R67 DHFR 244

List of Figures

Part 1

Figure 1.1 - Structures of osmolytes showing different functional groups..... 6

Figure 1.2 - Preferential exclusion and interaction mechanisms. 8

Figure 1.3 - DHFR reaction. 13

Figure 1.4 - Structures of EcDHFR, R67 DHFR and PTR1. 14

Figure 1.5 - Osmotic Stress Studies on Ligand Binding to R67 DHFR. 19

Figure 1.7 - Model depicting preferential interaction of osmolytes with free DHF. 26

Figure 1.8 - Structural features of folate. 28

Figure 1.10 - Osmolyte Dependent Changes in Folate NOEs. 34

Figure 1.11 - VPO Experiments Analyze Preferential Interaction and Exclusion between Osmolytes and Test Compounds. 37

Figure 1.12 - Small Angle Scattering Experimental Setup and Data Analysis. 46

Figure 1.13 - Sequence alignment of the type II DHFR variant sequences..... 52

Part 2

Figure 2.1 - A model showing the preferential interaction of osmolytes with free DHF. 65

Figure 2.2 - Structures of the various DHFRs. 67

Figure 2.3 - Sedimentation velocity data for FolM (15.7 μM) in MTA buffer, pH 6.0. .. 73

Figure 2.4 - A FolM homology model. 75

Figure 2.5 - Thermal denaturation of FolM by DSC. 76

Figure 2.6 - Effect of osmolytes on the secondary structure of FolM. 80

Figure 2.7- Representative ITC data. 81

Figure 2.8 - Effect of osmolytes on NADPH binding to FolM in MTA, pH 6.0..... 83

Figure 2.9 - Effect of DMSO and PEG400 on NADPH binding to FolM. 85

Figure 2.10 - The effect of osmolytes on the thermal stability of FolM in MTA, pH 6.0.87

Figure 2.11 - Effect of osmolytes on the DHF-reduction activity of FolM in MTA buffer, pH 6.0..... 90

Figure 2.12 - Effects of Viscosity and Dielectric Constant on the change in ln(kcat/Km) for the reduction of DHF by FolM. 92

Figure 2.13 - Fluorescence quenching data for MTX binding to FolM (2 μM) in MTA, pH 6.0..... 94

Figure 2.14 - Effect of osmolytes on MTX binding to the FolM-NADPH complex. 95

Figure 2.15 - Global fitting of ITC data for ligand binding to FolM using SEDPHAT. 100

Part 3

Figure 3.1 - A cartoon depicting preferential interaction of osmolytes with free DHF.. 117

Figure 3.2 - Structure of Folate. 119

Figure 3.4 - Folate dimerization at pH 10. 140

Figure 3.5 - Quantification of preferential interactions of betaine with test compounds. 143

Figure 3.6 - Vapor pressure osmometry studies of pyridoxine show pH effects..... 148

Figure 3.7 - Vapor pressure osmometry studies of imidazole 151

Figure 3.8 - pH dependence of folate solubility in 1 M betaine and water. 157

Figure 3.9 - The effect of osmolality on the binding affinities of folate (●) and PG4 (▲) to DHFRs 161

Figure 3.10 - Folate Simulation in Water. 164

Figure 3.11 - Folate Simulation in Betaine and Water. 166

Figure 3.12 - A comparison of predicted μ23/RT values vs. experimental μ23/RT values. 171

Figure 3.13 - A representation of Kp values for each atom type in folate is shown..... 172

Part 4

Figure 4.1- SANS profile and GNOM analysis for apo R67 DHFR. 195

Figure 4.2 - SASSIE analysis for apo R67 DHFR. 197

Figure 4.3 - Pairwise distribution of R_g for apo R67 DHFR, $NADP^+$ bound binary and $NADP^+$ and DHF bound ternary complexes. 202

Figure 4.4 – Conformational analysis of the disordered N-termini upon NADP+ binding to R67 DHFR by SASSIE 204

Figure 4.5 – Conformational analysis of the disordered N-termini upon DHF binding to the R67 DHFR-NADP+ complex by SASSIE 207

Figure 4.6 - Pairwise distribution of R_g for apo R67 DHFR with and without 20 % deuterated betaine 212

Figure 4.7 - Conformational analysis of the disordered N-termini of R67 DHFR in the presence of 20 % deuterated betaine 213

Figure 4.8 - SANS profiles of R67 DHFR in the presence and absence osmolytes and the analysis of the radii of gyration 217

Figure 4.9 - The ratio of I(0) in the presence and absence of osmolytes. 220

Figure 4.10 – Osmolyte effects on thermal denaturation of R67 DHFR. 223

Figure 4.11 - Crystal Structure of chymotrypsin truncated R67 DHFR (2GQV) at 1.1 Å resolution..... 233

Part 5

Figure 5.1 - Osmolyte effects on FolM..... 246

Abbreviations

5'AMP	Adenosine 5'-Monophosphate
ASA	Accessible Surface Area
AUC	Analytical Ultracentrifuge
BCA	Bicinchoninic Acid
BMRB	Biological Magnetic Resonance Bank
CB3717	10-Propargyl-5,8-Dideazafolate
CD	Circular Dichroism
2'CMP	Cytosine 2'-Monophosphate
DHF	Dihydrofolate
5'dTMP	Deoxythymidine 5'-Monophosphate
DMSO	Dimethyl Sulfoxide
DSC	Differential Scanning Calorimetry
EcDHFR	*E.coli* Chromosomal Dihydrofolate Reductase
f_v	Fraction Volume
5'GMP	Guanosine 5'-Monophosphate
GTFE	Group Transfer Free Energy
ITC	Isothermal Titration Calorimetry
MD	Molecular Dynamics
MTX	Methotrexate
MTA	100 mM MES, 50 mM Tris and 50 mM acetic acid
$NADP^+$/NADPH	Oxidized/Reduced Nicotinamide Adenine Dinucleotide Phosphate

NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NTA	Nitrilotriacetic Acid
n_w	Number of Water molecules
p-ABA	*p*-amino Benzoic Acid
p-ABA-Glu	*p*-amino Benzoyl Glutamic Acid
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PG4	Pteroyltetra-γ-L-Glutamate
PPC	Pressure Pertubation Calorimetry
ppm	Parts Per Million
PTR1	Pteridine Reductase 1
P6C	Pteridine-6-Carboxylate
R_g	Radius of Gyration
R67 DHFR	R67 dihydrofolate reductase
SANS	Small Angle Neutron Scattering
THF	Tetrahydrofolate
T_M	Melting Temperature
TMAO	Trimethylamine Oxide
TMP	Trimethoprim
5'UMP	Uridine 3'-Monophosphate
VPO	Vapor Pressure Osmometry

PART 1. GENERAL INTRODUCTION TO OSMOLYTES AND THEIR EFFECTS ON PROTEIN AND PROTEIN PROCESSES

1.1 Water is Essential for Life

Life originated in water and evolved subsequently. Water is the most abundant compound in all biological systems. Water is ubiquitous and makes up about 70 % of the weight of most living cells. Initial studies reviewing the physical chemistry of the prokaryotic and eukaryotic cytoplasm have emphasized a two-state model for cytoplasmic water including bound (hydration) water and free (bulk) water.[1, 2] Properties of water that make it a biological solvent include its hydrogen bonding ability to itself and to other molecules. Water is a polar solvent that dissolves all polar solutes and hydrates macromolecular surfaces.

An aqueous environment is essential to keep all intracellular biomolecules (proteins, DNA, RNA and small molecules) well hydrated thereby aiding biochemical processes. One of the major acting forces in protein folding is water driven hydrophobic collapse that minimizes the solvent exposure of hydrophobic groups forming the core of the protein structure. Water also plays a role in maintaining the stability of the folded protein by keeping the surface well hydrated. The role of hydration in protein conformation and conformational changes, in substrate binding and enzyme catalysis, and in molecular recognition has been studied and found to be important.[3]

Macromolecular interactions are largely governed by the structuring of water molecules (hydration shell) between the interacting surfaces. The hydration forces are stronger at close spacing than the van der Waals and electrostatic forces.[4]

Living cells face favorable and unfavorable environments. The extracellular environment affects the intracellular water concentration by osmosis. If the cells face a hypo-osmotic environment (low concentration of salts and solutes), water is taken up by

the cells causing swelling whereas under hyper-osmotic conditions (high concentration of salts\solutes), water is released from the cells causing cell shrinkage. The latter osmotic stress results in dehydration. Initially, the free cytoplasmic water is expelled from the cells rather than the bound water (hydration).[1, 5] Perturbations leading to dehydration cause adverse effects on the growth and survival of living cells. The concentration of free cytoplasmic water (unbound) that forms the bulk, sometimes referred to as the bathing solution, is the osmotically active water that is governed by the cytoplasmic osmotic coefficient which depends on the osmolality of the environment. On the other hand, bound water is osmotically inactive as it remains constant with the change in osmolality of growth media from 0.03 to 1 Osm.[1, 5, 6] When the amount of intracellular water continues to decrease, the bound water is lost, resulting in cell death. Maintaining normal cell function and growth under unfavorable conditions depends on stress response strategies and various counter mechanisms.

One mechanism that has been studied for over 3 decades now is the accumulation of intracellular osmolytes that function to combat osmotic stress in hyperosmotic conditions. The major players in the cell's response to osmotic stress include organic small molecules such as polyhydric alcohols (glycerol, ethylene glycols), sugars (sucrose, trehalose), free amino acids (proline, arginine, glycine), derivatives of amino acids (trimethylglycine, also known as glycine-betaine), methylamines (trimethylamine oxide) and urea.[7] These small molecules are either synthesized or transported into the hyperosmotically stressed cell, thereby increasing the intracellular osmolality (osmolyte concentration) and preventing water loss. Osmolytes take up space in solution and avoid cell shrinkage, thereby retaining the cellular volume, which supports cellular growth.[5] In

addition to this water retention mechanism, osmolytes also function as osmoprotectants that increase the cellular growth rate by stabilizing macromolecules in the cells.[5] (Structures of osmolytes-Figure 1.1) The choice of these osmolytes has been conserved throughout evolution because of their compatibility with macromolecular structure and function at high and variable osmolyte concentrations. Studies on *E.coli* cell growth and function have provided better understanding of the role of water in biological process and the part played by osmolytes under stressful conditions.

E.coli cells can modulate the intracellular concentration of osmolytes and water to adapt and grow over a wide range of external osmolalities.[1] Initial loss of intracellular water is followed by accumulation of potassium ions and glutamate. These two osmolytes were found to be effective when cells face low osmolality conditions. Secondary responses include synthesis of trehalose and putrescine, followed by uptake of betaine and proline, if available, to combat high osmolality surroundings.[1, 7, 8] Some osmolytes, also known as osmoprotectants (for example, betaine and proline), not only aid the cells to grow under stress but also make them more efficient by stabilizing the macromolecules and increasing the cell's growth rate. TMAO is another osmolyte that has been suggested to help unfolded proteins to fold to native-like structures.[9]

Betaine (N, N, N, -trimethyl glycine) is well-studied and is one of the most effective *E.coli* osmolytes.[5] Betaine was proposed to be an osmoprotectants as it is preferentially excluded from biomolecular surfaces. Preferential exclusion of osmolytes assures proper hydration leading to stabilization of macromolecules.[3, 10] This protective nature of excluded osmolytes is also supported by the measurements of volumes of "bound" water (presumably water of macromolecular hydration) and cytoplasmic osmotic coefficients for

cells grown in media of low (0.10 Osm) and moderate (0.28 Osm) osmolality. Osmotic coefficient (φ) characterizes deviation of the solvent from ideal behavior and is given by the ratio of observed to ideal osmotic pressures.

The volume of bound water is similar in the two osmotic conditions, suggesting no deleterious effects of osmolytes on the hydration. However, the free cytoplasmic water (bulk) decreases with increasing osmolality, and bound water becomes a larger fraction of the total volume as the osmolality of the growth medium increases. Growth appears to cease at the osmolality where the free water concentration is approximately equal to that of bound water.[1]

It was found that when *E.coli* cells are grown under hyperosmotic conditions (1 Osm), uptake of osmoprotectants (betaine) does not change the total concentration of intracellular osmolytes but increases the amount of cytoplasmic water significantly.[5, 6] Betaine was shown to replace other osmolytes (K^+, glutamate, trehalose) from the cytoplasm leading to no net change in osmolyte concentration. The overall increase in the cytoplasmic water can be attributed to a large change in the osmotic coefficient due to betaine uptake. This study indicated that betaine is the most preferred *E.coli* osmolyte as it alters the water activity to a greater extent than other osmolytes. The ratio of bound water in the biomolecular hydration layer to free water in the *E.coli* cytoplasm was noted to increase from 0.2 at (0.3 Osm-low osmolality) to 0.5 at higher osmolality (1 Osm), which resulted in bound water making a larger fraction of intracellular water. Therefore, the net increase in osmolality of the cytoplasm is large. Studies have shown that the amount of free cytoplasmic water, K^+ concentration and growth rate observed for *E.coli* under 1 Osm osmotic stress with 1 mM betaine supplied in the growth media are comparable to the

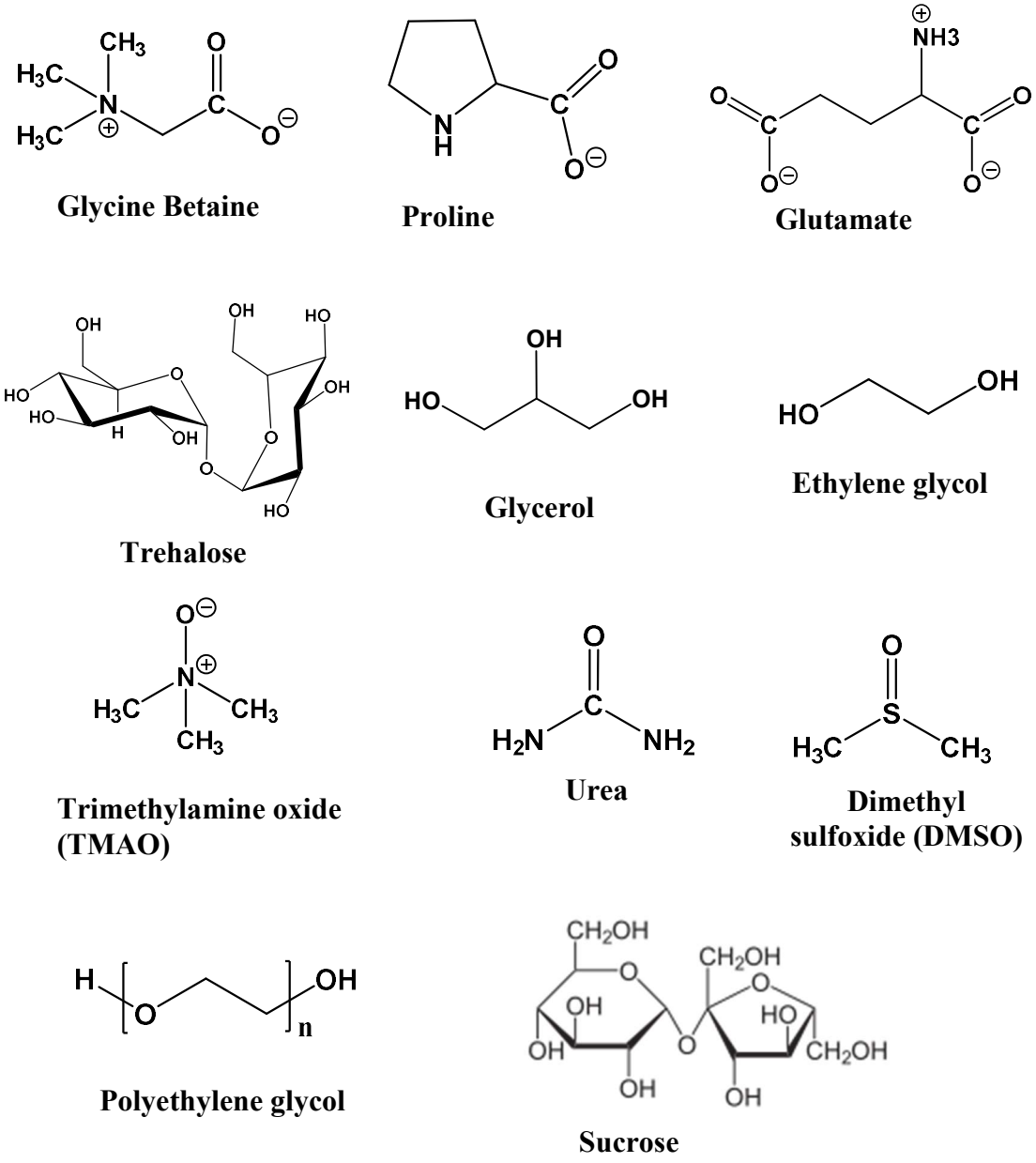


Figure 1.1 - Structures of osmolytes showing different functional groups. Most of the osmolytes except dimethyl sulfoxide (DMSO) and polyethylene glycol (PEG) are biologically relevant and are used to combat osmotic stress in living cells. Urea is another small molecule osmolyte that is known as a chemical denaturant for proteins.

values observed for *E.coli* growing at lower osmolality (0.4 Osm) with no betaine provided.[5]

1.2 Mechanism of Osmolyte Action- Preferential Exclusion vs. Preferential Interaction

Small molecule osmolytes acting as co-solutes exhibit their effects in multiple ways. Osmolytes can function by altering the water activity which is the measure of effective water concentration. All osmolytes take up volume in solution independent of their chemical nature, size and polarity. This changes the water concentration in the bulk media. Increasing osmolyte concentration leads to lowering of water activity, which alters the hydration of the macromolecules. In addition, osmolytes can also exhibit direct effects by forming favorable or unfavorable interactions with available functional groups in solution as seen in Figure 1.2. Osmolytes can be repelled from the macromolecular surfaces such that the concentration of osmolyte in the bulk solution is greater than in the hydration shell. This is the preferential exclusion mechanism of osmolyte action. The chemical nature as well as the size of the osmolyte can help determine its extent of exclusion from molecular surfaces. For example, betaine has been shown to be the most excluded osmolyte in *E.coli*.[8]

In a second mechanism, some osmolytes can associate with the molecular surfaces such that the number of osmolyte molecules in the proximity (hydration shell) increases; this decreases the number of osmolyte molecules present in the bulk solution. This is the preferential interaction mechanism. The thermodynamic quantitation of this phenomenon dictates the strength and favorability of osmolytes towards various functional groups in

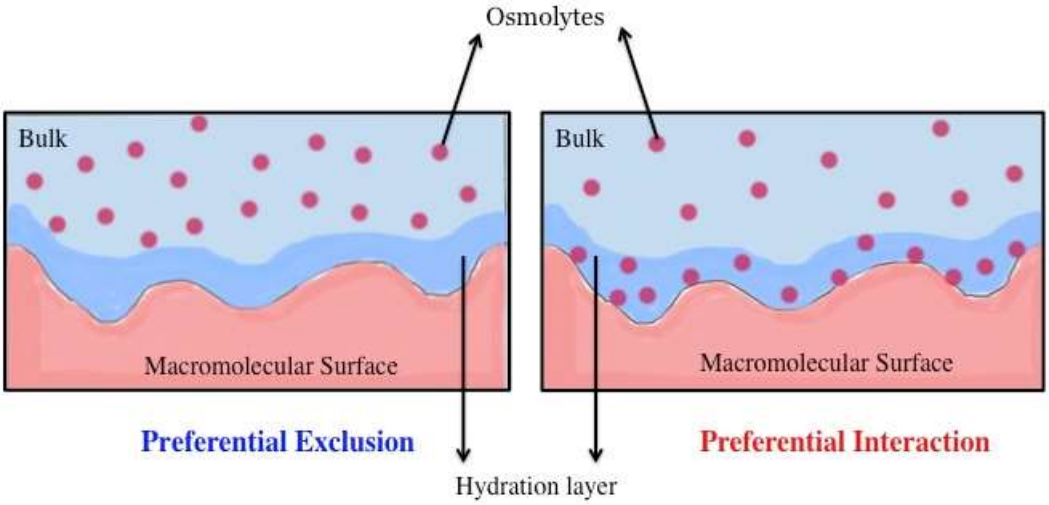


Figure 1.2 - Preferential exclusion and interaction mechanisms. Osmolytes (pink spheres) that are excluded do not enter the hydration shell (dark blue) and are retained in the bulk solution (light blue) whereas osmolytes that preferentially interact replace water in the hydration shell.

comparison to water. The differences in the interactions of *E.coli* cytoplasmic osmolytes with protein surfaces was shown to be correlated with their effectiveness as osmoprotectants. The order of preferential exclusion obtained for *E.coli* osmolytes was betaine>proline>TMAO>trehalose>K^+ glutamate> glycerol.[11]

1.3 Osmotic Stress Studies Probe the Role of Water in Biological Processes

Water plays a crucial role in all biological processes such as protein folding, stability and conformational change; protein-ligand, protein-protein and DNA-protein interactions; and enzymatic activity. This has led nature to develop an effective strategy of using osmolytes to maintain the structural and functional properties of macromolecules in cells exposed to denaturing environmental stresses. A lot of interest has developed in performing osmotic stress studies that measure the changes in hydration of macromolecules associated with biochemical processes.

Osmotic stress studies employ the addition of small molecule osmolytes that alter water activity in the experiment and the progress of the biochemical process is monitored. Upon comparison of the results obtained with and without externally added osmolytes, the effects of osmolytes can be directly correlated to the role of water in that particular process. Water release or uptake upon ligand binding have been demonstrated for proteins, for example, hexokinase activity results in a release of around 320 water molecules[12] and hemoglobin takes up about 60 water molecules upon changing from its oxygen deficient to its oxygen rich form.[13] The measurement of the change in the number of water molecules during a biomolecular reaction is a direct measure of forces between solute and solvent molecules that either support preferential exclusion or interaction between two molecules

in comparison to water. These studies are the key to understanding the strength and specificity of biological reactions involving water/hydration changes.

Studying the role of water and effects of osmotic stress on metabolic pathways allows a better understanding of mechanisms and strategies employed by the cell to overcome stress. Thus, changes in the number of macromolecule associated water molecules (in the hydration layer) and bulk solution can be studied using an osmotic stress approach. As betaine was shown to be the most excluded osmolyte from biological surfaces in an *E.coli* cell, [11] it is a preferred osmolyte used in osmotic stress studies. Thus, a highly excluded osmolyte like betaine is more likely to probe the changes in hydration in a biochemical process whereas an osmolyte that is not completely excluded will lead to an underestimation of the extent of changes in hydration. To understand the multifaceted effects due to preferential exclusion and/or interactions of osmolytes with molecular surfaces, a detailed analysis of these weak associations will help in better interpretation of the outcomes of osmotic stress experiments.

Our lab chose to study osmotic stress effects on enzymes of the folate metabolism pathway using a set of osmolytes including betaine, glycerol, ethylene glycol and sucrose to name a few. We also focus on interpreting our results based on both preferential exclusion and interaction mechanisms of osmolyte action. We measure the affinity of a ligand for the enzyme in buffer as well as in buffer containing osmolytes. The measurements are done in two concentrations of the osmolyte. The change in binding affinity with change in osmolality yields the change in the number of water molecules released or taken up upon ligand binding using the following equation, Eq (1.1):

$$\frac{\partial ln(K_a)}{\partial ln(osm)} = -\frac{\Delta n_w}{55.6} \qquad \text{Eq (1.1)}$$

where K_a is the binding association constant and osm is the osmolality of the buffer, Δn_w is the change in hydrating waters upon ligand binding. The following sections introduce our model system and previous osmotic stress studies performed.

1.4 Introduction to Dihydrofolate Reductase (DHFR)

Folic acid, vitamin B9, is an essential nutrient for cell growth and development. The folate metabolism pathway in *E. coli* consists of various enzymes that function to assimilate folate substrates to form metabolically important cofactors. Folate derived cofactors are utilized in one-carbon transfer reactions. Dihydrofolate reductase (DHFR) catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using NADPH as a cofactor. This reaction is important as THF and its derivatives serve as cofactors for reactions involved in the synthesis of nucleotides such as thymidine, amino acids such as methionine and glycine and various other metabolites. Effective blocking of DHFR activity leads to cell death and therefore, this enzyme is a potential target for developing anti-bacterial and anti-cancer drugs. EcDHFR is the enzyme encoded by a chromosomal gene in *E. coli* and is effectively inhibited by an antibacterial drug, trimethoprim (TMP) and an anti-cancer drug, methotrexate (MTX).

Occurrence of clinical resistance to TMP treatment was associated with emergence of a novel DHFR enzyme encoded by resistance plasmids (R-plasmids).[14] One of the interesting and most studied R-plasmid DHFRs is R67 DHFR. Both the chromosomally encoded (EcDHFR) and plasmid encoded R67 DHFR catalyze the same reaction (Figure

1.3), but possess unrelated structures and properties as seen in Figure 1.4 and Table 1.1.

1.4.A. EcDHFR

Chromosomally encoded dihydrofolate reductase *E.coli* (EcDHFR) is a well-evolved[15] and well-conserved protein in all kingdoms of life. It is an 18 kDa monomeric protein with discrete binding pockets for the substrate (DHF) and the co-factor (NADPH). The EcDHFR structure possesses an eight-stranded β sheet core with four surrounding α helices, which together form two rigid subdomains separated by a hinge region.[16, 17] Structural studies with the apo and ligand bound EcDHFR complexes suggest a significant role for the Met20 loop (residues 9–24) that switches between a closed and an occluded conformation during the catalytic cycle.[18, 19] Linked changes in the FG (residues 116–132) and GH (residues 142–149) loops also occur. Substrate and cofactor binding closes the Met20 loop over the active site and stable hydrogen bonding with the FG loop forms an optimal electrostatic environment to aid the hydride transfer.[19, 20] Upon formation of the products, the Met20 loop releases the oxidized cofactor and occludes the cofactor binding pocket. Product (THF) release is the rate-limiting step at pH 7 and is facilitated by binding of NADPH.[19]

1.4.B. R67 DHFR

The resistance plasmid encoded DHFR is not an efficient catalyst, but it confers resistance to the antibiotic drug, trimethoprim which is a potent competitive inhibitor of EcDHFR. R67 DHFR is unrelated in sequence and structure to EcDHFR and is thought to be a primitive enzyme.[21, 22] It is a 34 kDa homotetramer with all 4 subunits contributing to forming a single active site pore in the center of the structure. Each monomer of 78 amino acids contains five anti-parallel β strands. Dimerization occurs when three strands

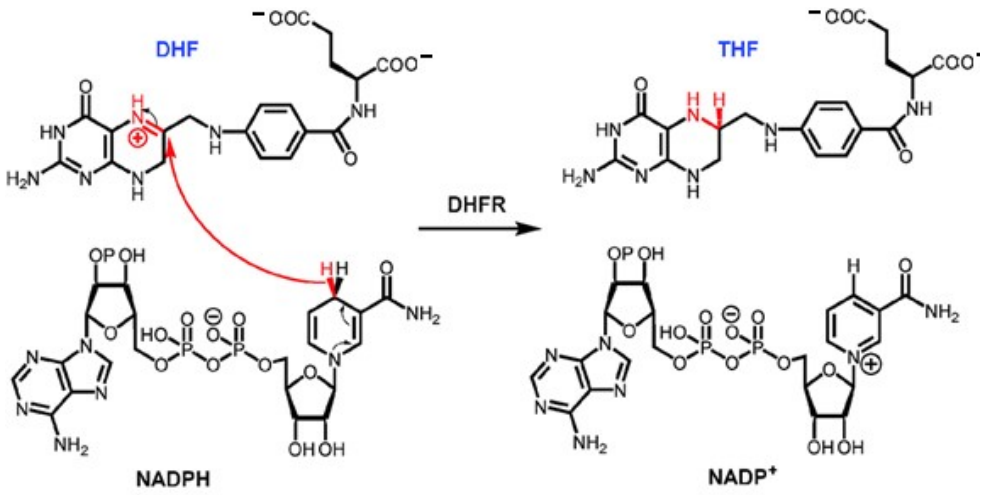


Figure 1.3 - DHFR reaction. Reduction of DHF to THF using NADPH as a cofactor. The substrate (DHF) is activated by pre-protonation (as shown in red) resulting in a positive charge intermediate that accepts a hydride from NADPH to form the product (THF). The cofactor is oxidized to $NADP^+$ as shown. Adapted from Reference. [23]

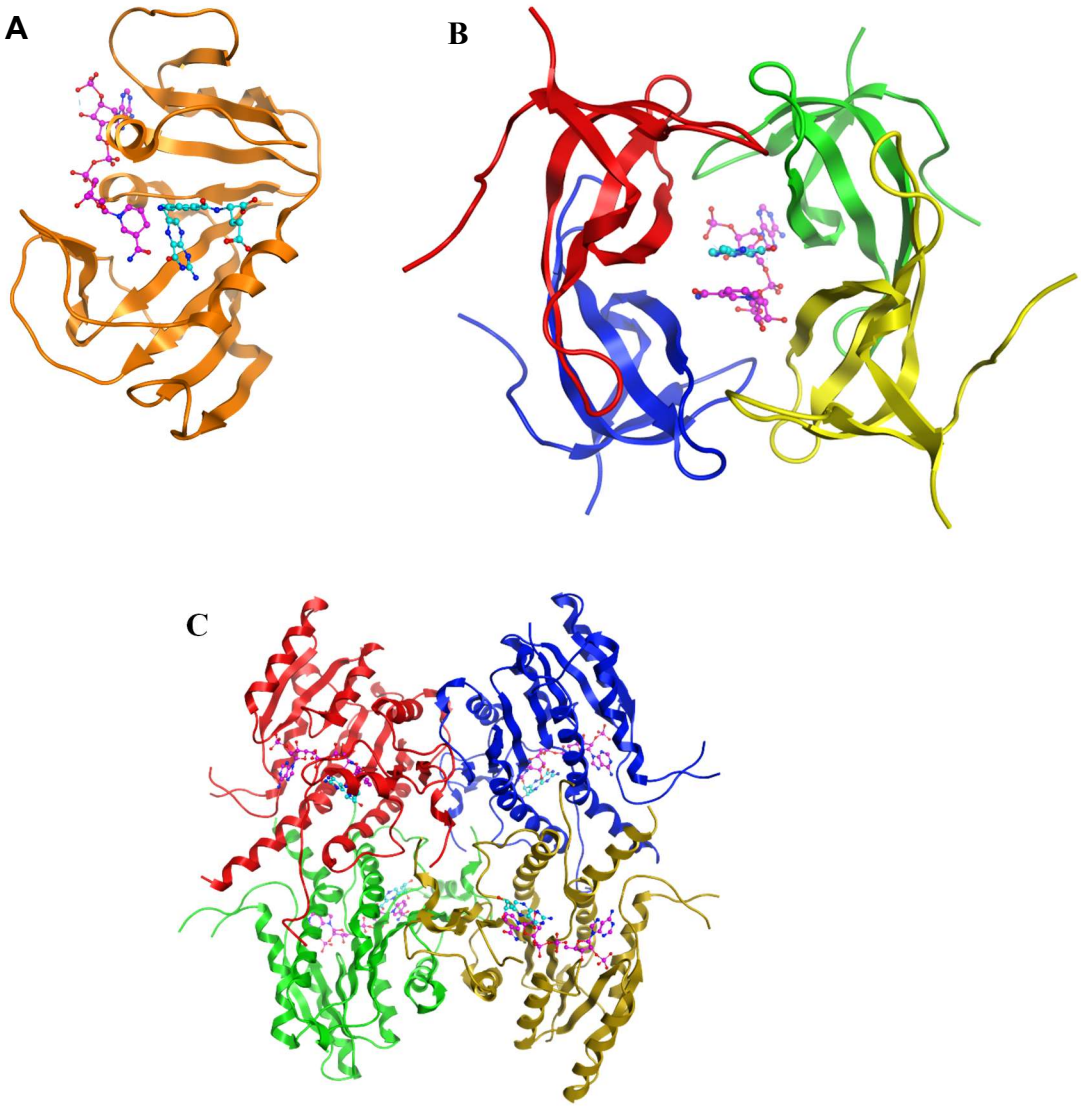


Figure 1.4 - Structures of EcDHFR, R67 DHFR and PTR1. EcDHFR (1RA2) is shown with the bound ligands ($NADP^+$ in magenta and folate in cyan) in panel A. R67 DHFR (2RK1) homotetramer shown in panel B with each monomer colored differently. The bound substrate (DHF in cyan) and cofactor ($NADP^+$ in magenta) is shown. Panel C shows the structure of a pteridine reductase, PTR1, from *L. major* (1E92) with bound $NADP^+$ (pink) and dihydrobiopterin (cyan).

Table 1.1 - Comparison of structural and functional parameters for EcDHFR, R67 DHFR, PTR1 and FolM. Parameters obtained at pH 7.0 for EcDHFR and R67 DHFR and pH 6.0 for PTR1 and FolM.

Parameter	EcDHFR	R67 DHFR	PTR1 (*Leishmania major*)	FolM
Monomer size[a]	159 aa; 17,999 daltons	78 aa; 8,430 daltons	288 aa; 30,457 daltons	240 aa; 27,496 daltons
Oligomeric state	Monomer[b]	Tetramer[c]	Tetramer[d]	Tetramer[e]
# of active sites	1	1	4	4
Structural features	8 stranded mixed β-sheet with 4 α-helical connecting strands[b]	4 β-barrels; single active site pore composed of residues from 4 subunits[c]	7-stranded parallel β-sheet sandwiched between 3 α-helices on either side[d]	ND
Volume of active site[f]	1677 $Å^3$ (1RA2)	3626 $Å^3$ (1VIE)	1392-1920 $Å^3$ (2BFA)	ND
Trimethoprim K_i	20 pM[g]	150μM[h]	ND	>1.4 mM[i]
Methotrexate K_i or K_d	0.07 nM[j]	>500μM[k]	30-255 nM[l]	5.9 μM[i]
NADPH K_m or K_d	0.94 μM[m]	3.0 μM[n]	14.2 μM[l]	1.9 μM (K_m)[i] 3.86 ± 0.29 (K_d)[e]
DHF K_m	1.2 μM[m]	5.8 μM[n]	3.4 μM[l]	9.0 μM[i] 4.3 ± 0.6 μM[e]
k_{cat}	28 s^{-1} (product release)[n] 238 s^{-1} (hydride transfer)[o]	1.3 s^{-1} [o]	0.38 μmol min^{-1} mg^{-1} (V_{max})[l]	0.083 μmol min^{-1} mg^{-1} (V_{max})[i] 0.240 ± 0.009 s^{-1} (k_{cat})[e]
Natural substrate	DHF	DHF	Biopterin, dihydro-biopterin[l]	Dihydro-monapterin[p]

[a] Molecular weight calculated from http://web.expasy.org/cgi-bin/protparam/protparam.
[b] From reference [24, 25]. [c] From reference [26, 27]. [d] From reference [28]. [e] From Part II. [f] Calculated from Castp[29] (http://cast.engr.uic.edu/cast/). [g] From reference [30]. [h] From reference [14]. [i] From reference[31]. [j] From reference [32]. [k] From reference [33]. [l] K_i value depends on substrate, from reference [34]. [m] From reference [35]. [n] From reference [36]. [o] From reference [37]. [p] From reference [38].

from each monomer form a six-stranded β barrel at the interface.[26] Two dimers associate

by loop–loop interactions to form the tetramer. The crystal structure of the dimeric species yielded weak electron density for the first 16-18 residues of the monomers suggesting disorder in the N-terminal tails.[26] The disordered tails were cleaved by chymotrypsin treatment and the resultant truncated tetramer was crystallized by Narayana et al.[39] The N-terminal truncation showed no significant effect on the enzyme activity.[36]

The presence of a single active site within a tetramer possessing a 222-symmetry requires both the cofactor (NADPH) and the substrate (DHF) to bind to symmetry related, promiscuous binding sites. Both ligands make contacts with the same residues from different monomeric surfaces. This enables R67 DHFR to bear 4 identical ligand binding sites, but only two ligands can be accommodated due to steric hindrance. R67 DHFR can form three types of ligand bound complexes-with two DHF molecules or with two NADPH molecules or with one DHF and one NADPH molecule.[40] The last combination forms a productive complex for the DHFR reaction.

1.4.C. A Novel DHFR: FolM

An interesting enzyme, FolM, encoded by a chromosomal gene, *ydgB,* catalyzes the reduction of dihydromonapterin to tetrahydromonapterin using NADPH as a cofactor.[38] It is a homologue of pteridine reductase (PTR1) from *Leishmania major,* which reduces biopterin and dihydrobiopterin. FolM additionally catalyzes the reduction of DHF, although not as efficiently as EcDHFR or even R67 DHFR.[31] FolM expression in the Δ *folA* (DHFR knockout) cells allowed the cells to grow on minimal media lacking folate pathway end products.[31] However, the double knockout *E. coli* mutant (Δ *folA* Δ *ydgB*) was a synthetic lethal.[41]

This novel enzyme - FolM - was found to be resistant to trimethoprim and less sensitive to methotrexate inhibition than EcDHFR.[31] FolM is genetically unrelated to either *E. coli* chromosomal or plasmid encoded R67 dihydrofolate reductases. FolM is a short chain dehydrogenase/reductase that utilizes NADPH as a hydride donor (cofactor). It is a tetramer with four active sites. FolM contains each of the amino acid residues in PTR1 that are important for binding of substrate and catalysis. It is thought to share the same catalytic triad Asp 181, Tyr 194, Lys 198 as PTR1.[28] It is noteworthy that FolM, which has a different scaffold and is unrelated in sequence and structure to the DHFRs, catalyzes the same reaction.

1.5 Osmotic Stress Studies

The major questions we ask are how metabolic pathways function under osmotic stress conditions and what are the effects of osmolytes on proteins, small molecule ligands or substrates and protein-ligand complexes? Osmotic stress experiments involve addition of osmolytes to the *in vitro* and *in vivo* experimental systems to probe the role of water in the addressed biological process. The effect of osmolytes on various macromolecular surface types has been studied as discussed earlier. The implications of the weak preferential exclusion and interaction effects on protein folding, stability and protein-DNA interactions have also been studied. The focus of our lab has been studying osmotic stress effects on folate metabolism in *E. coli*. [42-45] We tested the effects of betaine, glycerol, ethylene glycol, DMSO, sucrose and PEG to name a few.

Betaine was chosen as it is the most excluded osmolyte in *E.coli*. Other osmolytes were chosen to include compounds different chemical properties to ensure that any changes

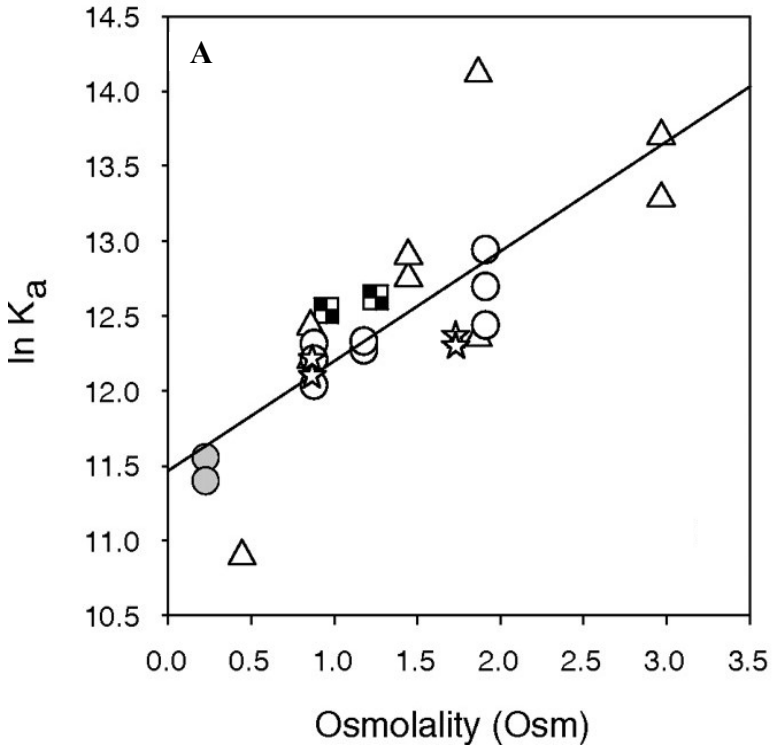
in binding affinity came from the change in water activity rather than changes in the solution dielectric or viscosity. For example, sucrose and betaine both alter water activity but they show opposite effects on the dielectric constant of the solution. To parse out multiple effects, osmolytes with differential characteristics were included in both the studies (R67 DHFR and EcDHFR). The following sections discuss previous studies addressing osmotic stress effects on ligand binding to DHFRs.

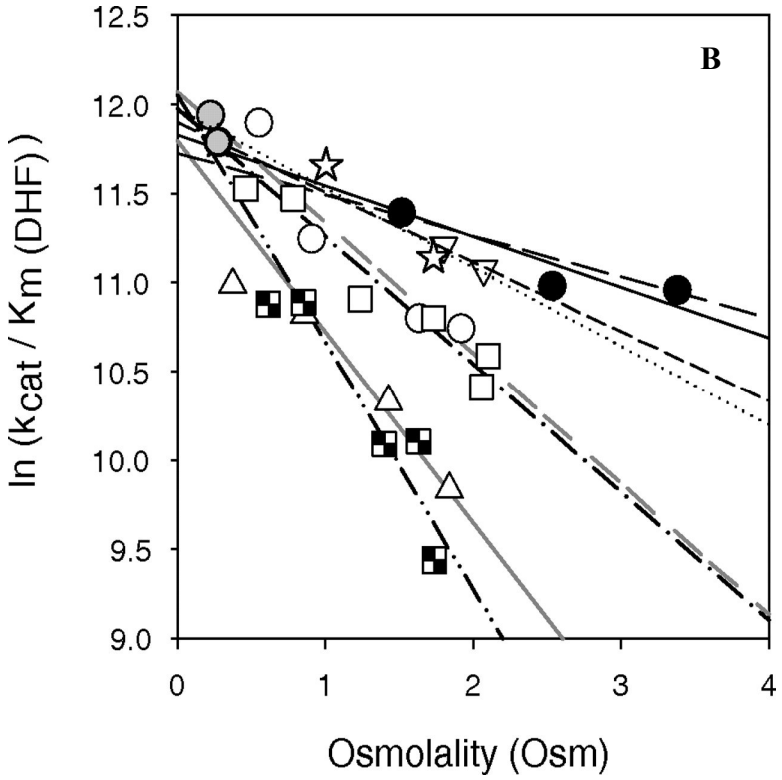
1.5.A. Osmotic Stress Studies on Ligand Binding to R67 DHFR

In vitro studies were performed to understand the role of water on ligand binding to R67 DHFR. Osmolytes were included in the binding assays and the binding affinities were compared to those with no osmolyte added. Several osmolytes such as betaine, glycerol, ethylene glycol, DMSO, sucrose and PEG400 were used. NADPH binding to R67 DHFR was monitored by isothermal titration calorimetry and osmolyte addition was found to tighten the binding affinity.[42] This observation is consistent with the dehydration effects of osmolytes which aid binding. The water molecules hydrating the protein and ligand molecules at the binding interfaces have to be shed to form contacts. Osmolytes lower the free water concentration in the system and therefore aid in dehydration of the interface. Tighter binding indicates preferential exclusion of osmolytes from the cofactor as well as protein surfaces. Another striking feature of these studies was that the extent of tightened binding of NADPH was the same for all osmolytes tested. Figure 1.5.A shows the data for all osmolytes could be fit to a single slope upon analysis using Eq (1.1). The positive slope yields a negative Δn_w indicating water release upon cofactor binding.[42]

DHF binding to R67 DHFR was studied using steady state kinetic assays that measure the K_m for the formation of enzyme-substrate complex as well as the catalytic rate

Figure 1.5 - Osmotic Stress Studies on Ligand Binding to R67 DHFR. In panel A, a plot of ln K_a values for $NADP^+$ binding to R67 DHFR vs. osmolality is shown. A single line is shown that fits the data obtained from multiple ITC experiments to examine effects of each osmolyte addition. Panel B shows the data obtained from steady state kinetics experiments. A plot of ln k_{cat}/K_m vs. osmolality yields multiple slopes for each osmolyte tested indicating an osmolyte specific effect on substrate binding. A positive slope in panel A suggest tightened binding of cofactor with increasing osmolality. The negative slopes in panel B suggest weakened binding of substrate to R67 DHFR upon osmolyte addition. Data for buffer (●), ethylene glycol (☆), DMSO (○) glycine betaine (△), PEG 400 (checkerboard), glycerol (●), TMAO (▽) and sucrose (□) are shown. Data are from Chopra et al.[42]


A
14.5
14.0
13.5
13.0
12.5
12.0
11.5
11.0
10.5
ln K_a
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Osmolality (Osm)


B
12.5
12.0
11.5
11.0
10.5
10.0
9.5
9.0
ln (kcat / Km (DHF))
0
1
2
3
4
Osmolality (Osm)

constant (k_{cat}). Osmolyte addition resulted in no significant effect on the k_{cat}, but the K_m for DHF increased in the presence of osmolytes. The ratio of k_{cat}/K_m gives the catalytic efficiency, which decreased with increasing osmolyte concentration (Figure 1.5.B).[42] This indicates weakened substrate-binding to R67 DHFR in the presence of osmolytes. As K_m values can contain kinetic information in addition to binding information, ITC experiments were also performed. A similar trend of weaker binding affinity with increasing osmolyte concentration was observed. The magnitude of the effect of each osmolyte was variable as can be seen from the differences in the slopes in Figure 1.5.B. This indicates that the extent of weakened binding depends on the individual osmolyte identity. Analysis of these data using Eq (1.1) resulted in positive Δn_w values suggesting water "uptake".

It was unusual to see weaker binding of DHF and tighter binding of NADPH to R67 DHFR as both the ligands interact with the symmetry-related residues in the active site pore. This behavior can be due to one of three possible effects of osmolytes – destabilization of the enzyme-DHF complex, stabilization of the free enzyme and/or stabilization of free DHF. Effects of osmolyte addition on free enzyme and enzyme-cofactor complex can be ruled out as many osmolytes show the same effect for NADPH binding (Figure 1.5A). Thus, NADPH binding results serve as an internal control as the substrate and cofactor binding sites are related. This is due to the 222 symmetry imposed on the R67 active site pore. This suggests that the osmolytes may not exert their effects on the enzyme. This leaves the possibility of osmolyte effects on free DHF which then affects its binding to the enzyme. The amplitude of effects was different for all the osmolytes tested indicating an osmolyte specific effect on DHF binding. The weakened binding

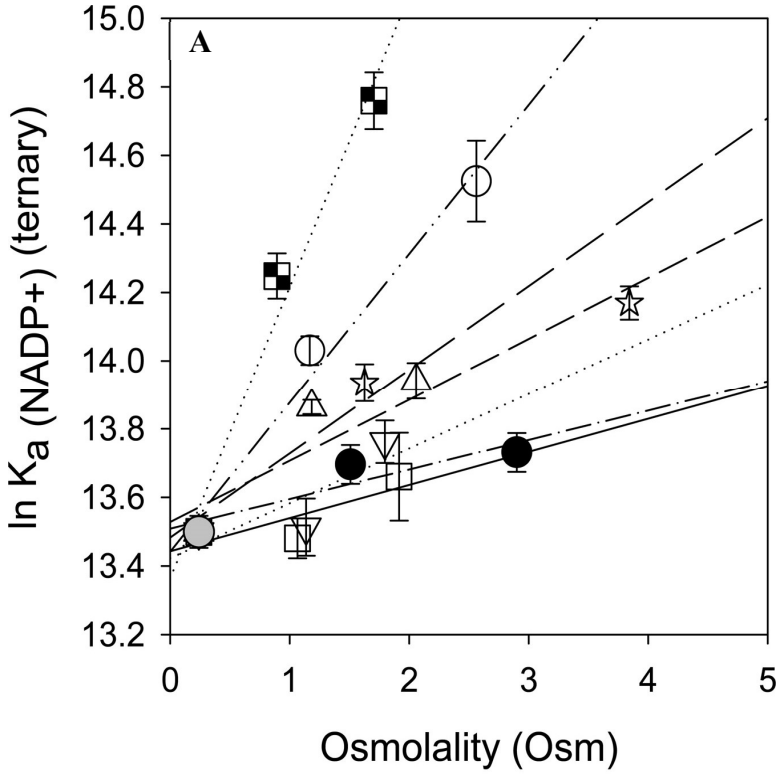
affinity with water activity indicates favorable interactions between osmolytes and DHF (relative to water) interfere with substrate binding to R67 DHFR.

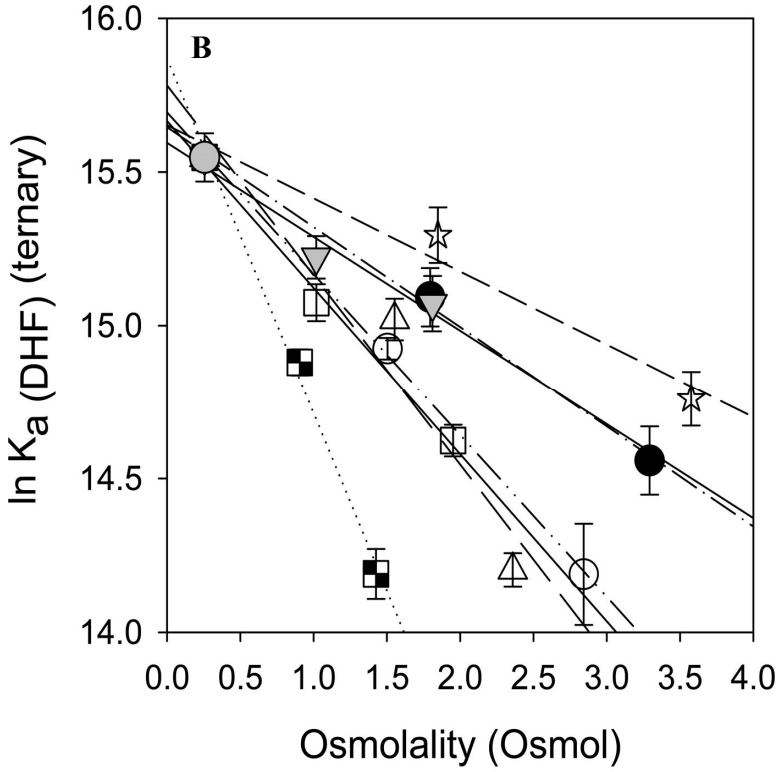
In vivo osmotic stress studies on R67 DHFR correlated with our *in vitro* findings as the osmotically stressed cells exhibited reduced enzyme activity and growth.[42] A variant of R67 DHFR with reduced catalytic efficiency was used to rescue *E. coli* cells grown in the presence of TMP. TMP addition inhibits chromosomal DHFR activity. The cells, when subjected to osmotic stress by growing in sorbitol containing media, showed no growth. Sorbitol induces osmotic stress leading to dehydration effects and production of intracellular osmolytes. The weak interactions between intracellular osmolytes and DHF can affect substrate binding to DHFR and the variant enzyme was unable to confer TMP resistance.

1.5.B. Osmotic Stress Studies on Ligand Binding to EcDHFR

To test the hypothesis that osmolytes weakly interact with folate/DHF and weaken binding, a second DHFR was used. EcDHFR catalyzes the same reaction as R67 but has a different structural scaffold. Similar studies were performed to monitor ligand binding to EcDHFR in the presence of osmolytes. $NADP^+$ binding to EcDHFR-DHF complex (ternary complex formation) was measured using ITC and the binding affinity (K_a) was noted to increase with addition of the osmolytes. These results, consistent with the dehydration effects of osmolytes, were similar for both DHFRs. With EcDHFR, variable slopes were obtained for each of the osmolytes tested (Figure 1.6.A), which indicated additional effects of osmolytes on $NADP^+$ and/or the protein or the complex. Also, the change in binding affinity of $NADP^+$ to EcDHFR (binary complex formation) showed an opposite trend upon addition of sucrose. This indicated additional effects of sucrose on the

Figure 1.6 - Osmotic Stress Studies on Ligand Binding to EcDHFR. Panel A shows a plot of ln K_a values for $NADP^+$ vs. osmolality. The data obtained from ITC experiments upon addition of each osmolyte gives a positive slope, indicating tightened binding of cofactor upon osmolyte addition. Variability in slopes indicates additional effects of osmolytes on the protein. Panel B plots ln K_a values for DHF binding to the EcDHFR·$NADP^+$ complex obtained using ITC vs. osmolality. Negative slopes indicate weakened substrate binding in the presence of osmolytes. Data for buffer (●), glycerol (●), ethylene glycol (☆), TMAO (▽), sucrose (□), DMSO (○), glycine betaine (△), and PEG 400 (checkerboard) are shown. Data are from Grubbs et al.[43]


A
15.0
14.8
14.6
14.4
14.2
14.0
13.8
13.6
13.4
13.2
ln Ka (NADP+) (ternary)
0
1
2
3
4
5
Osmolality (Osm)


B
16.0
15.5
15.0
14.5
14.0
ln Ka (DHF) (ternary)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Osmolality (Osmol)

binding event. In contrast, DHF binding to EcDHFR, as quantified using ITC, showed a trend of weakened binding upon osmolyte addition as can be seen with a negative slope in Figure 1.6.B.[43] Thus, consistent results with two structurally and genetically unrelated proteins supported the hypothesis that the small molecule osmolytes interact with DHF and need to be removed before DHF can bind to the enzyme.

1.6 Model for Preferential Interactional between DHF and Osmolytes

We proposed the model (shown in Figure 1.7) for preferential interaction between osmolytes and the substrate, dihydrofolate. This weak interaction shifts the binding equilibrium towards the unbound protein and free DHF as observed by a decrease in the affinity constant. The unusual weakening of DHF binding to both the DHFRs with different scaffolds supports this hypothesis. Our model posits that weak interactions between free DHF and osmolytes result in a competition between osmolyte and water for solvation of DHF. The model in Figure 1.7 represents weaker binding affinity of DHF, as the osmolyte molecules have to be released from DHF prior to its binding to DHFR.

1.7 Do Osmolytes Weaken Substrate Binding to the Novel DHFR, FolM?

Osmotic stress experiments for substrate binding to DHFRs resulted in weaker K_d values in presence of osmolytes. The results were interpreted using a preferential interaction of DHF with osmolytes, which affected DHF binding to the DHFRs. In order to strengthen our model, we decided to extend it to other DHFRs and other enzymes from the folate metabolism pathway. Similar osmotic stress studies were performed and the

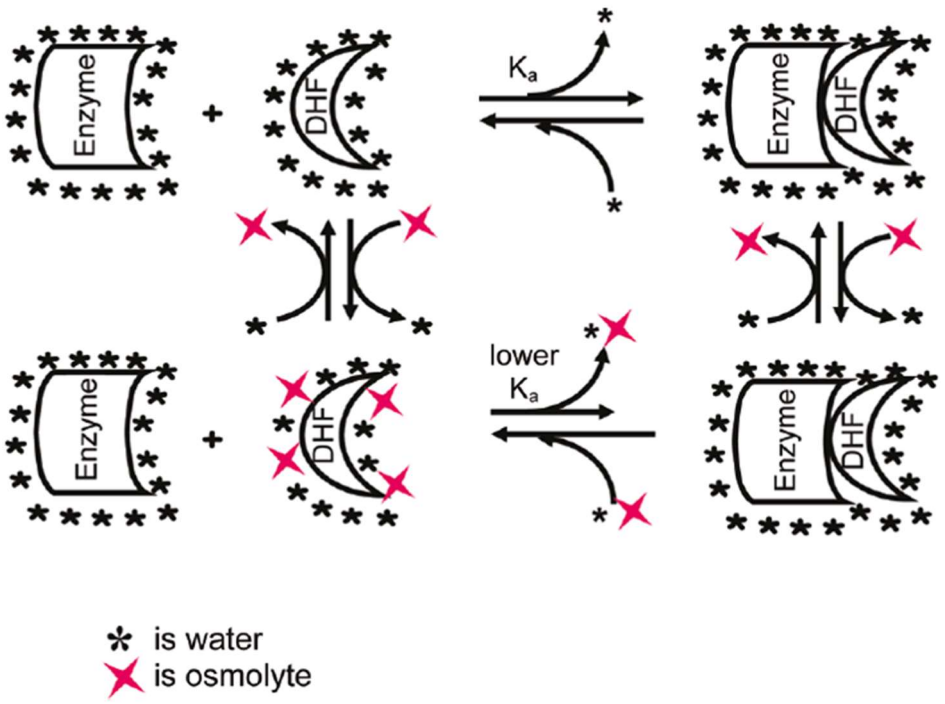


Figure 1.7 - Model depicting preferential interaction of osmolytes with free DHF. In the absence of osmolytes, DHF binds to either EcDHFR or R67 DHFR and water (*) is released as seen in the top panel. Osmolytes (pink star) interact with free DHF and replace the water molecules in the hydration layer. Both osmolytes and water must be released for DHF to bind to DHFRs, resulting in a lower binding affinity K_a. (This model does not exclude the possible binding of osmolytes to DHFRs, which could describe the differing effects of osmolytes on $NADP^+$ binding to EcDHFR, yielding the variation in slopes seen in Figure 1.6.A.)

preferential interaction model was tested for ligand binding to FolM as one of my projects and the results obtained are explained in Part 2. We predict all folate derivatives will be affected by osmolytes as per the model. Similarly, the model can be applied and tested for other enzymes of the folate metabolism pathway.

1.8 Folate – Osmolyte Interactions

Folate is the oxidized form of DHF with greater stability. It is known to bind and inhibit DHFRs.[33] Studies monitoring folate binding to R67 DHFR showed weaker interactions upon betaine addition suggesting similar effects to those seen with the substrate, DHF. Thus, folate can serve as a model for DHF to study its preferential interactions with osmolytes. The folate structure is comprised of a pterin ring connected to a *p*-amino benzoic acid (*p*-ABA) ring with a glutamate tail (Figure 1.8.A). A folate molecule can have more than one glutamate attached to its tail via the γ-carbon of glutamate.

Folate dimerizes at high concentrations. A head-to-tail dimer model has been proposed in which each pterin ring stacks with the *p*-ABA ring of the other monomer and the glutamate tails are free to rotate. [46] Thus, the charged glutamate tails are oriented in opposite directions. Poe found folate dimerization to be pH dependent.[46] Folate dimerizes easily if the N3−O4 amide of the pterin ring is fully protonated. The N3-O4 group undergoes a keto-enol tautomerization as shown in Figure 1.8.B. The pK_a of this amide group is reported to be 7.98 to 8.38.[46, 47] The dimerization constant (K_d) of neutral folate is 20 mM, which is less than the K_d of 340 mM for the basic form.[46] Deprotonation of the N3-O4 amide results in enol formation and a negatively charged

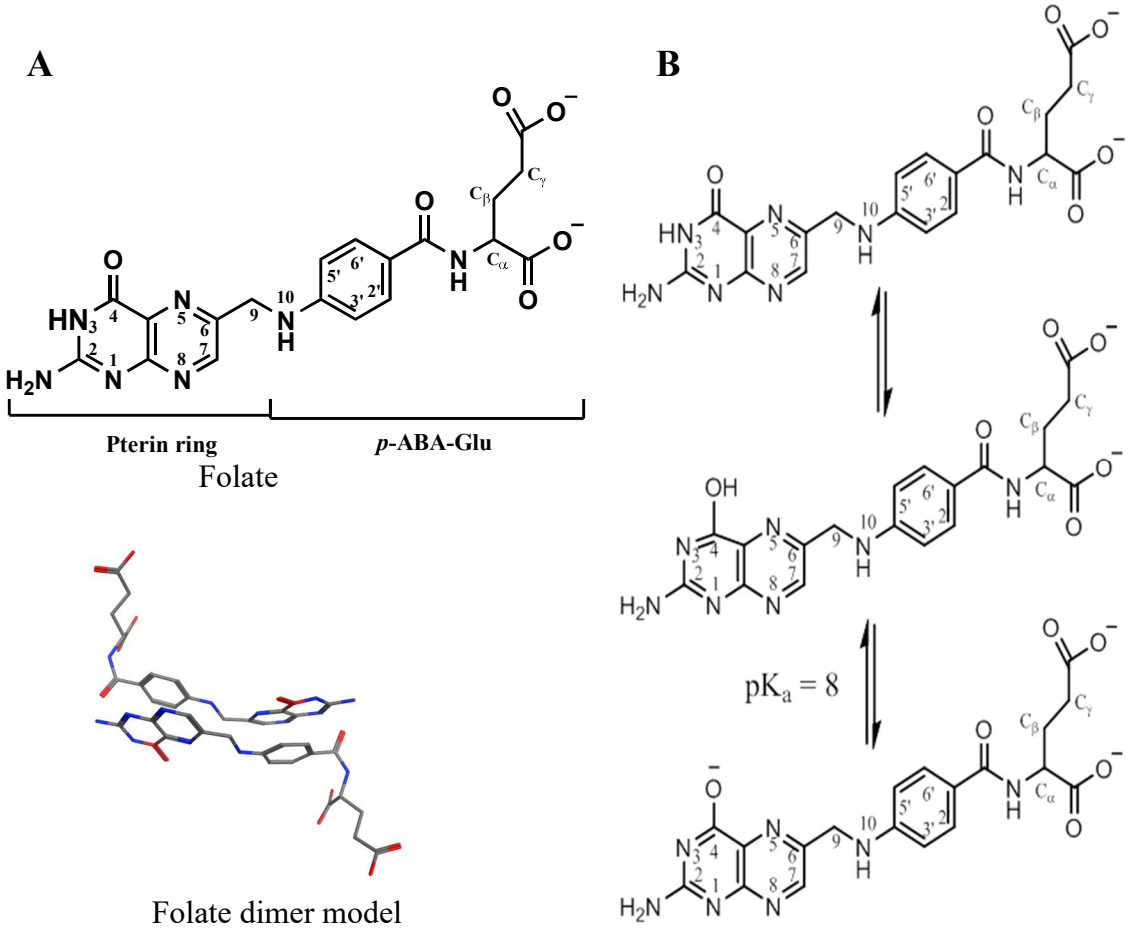


Figure 1.8 - Structural features of folate. Panel A shows the structures of folate with atom numbers and a model for the folate dimer. Atom colors are gray for carbon, blue for nitrogen and red for oxygen. Panel B shows the keto-enol tautomerization and deprotonation of the N3-O4 group of the pterin ring of folate.

oxygen, which repels the p-ABA ring of its partner folate, hinders the stacking of rings and weakens dimerization.

1.8.A. Previous Studies using NMR and NOESY

Previous studies in our lab tested the preferential interaction of osmolytes (betaine and DMSO) with folate using nuclear magnetic resonance (NMR) and nuclear Overhauser spectroscopy (NOESY).[44] NMR was performed to study folate dimerization in the presence of osmolytes. NMR experiments monitoring proton chemical shifts gave 2-fold and 2.5-fold increases in the folate dimerization K_d in the presence of 20 % (w/v) betaine and 20 % (v/v) DMSO, respectively (Figure 1.9 and Table 1.2).[44] These results indicated preferential interaction of betaine and DMSO with the folate monomer. In other words, these osmolytes interact with folate and need to be removed before two folate molecules can associate. The NMR data also supported the head to tail model of dimerization as the proton chemical shifts from the glutamate moiety showed no significant change with folate concentration and osmolyte addition as can be seen in the lower most panel of Figure 1.9. As folate dimerization was observed to be pH dependent, the pH titration of folate with and without osmolyte addition was studied using absorbance. The pK_a of the N3-O4 group was noted to be 7.94 with no significant effect upon betaine and DMSO addition.[44] Therefore, the weakening of dimerization was attributed to the weak interactions between the osmolytes and folate.

Homonuclear (^{1}H) Nuclear Overhauser Effect spectroscopy (NOESY) experiments were also done to identify the position(s) in the folate structure where betaine and DMSO may interact. In NOESY experiments, interactions between nuclei through space are detected. NOE spectra were collected at a low folate concentration with and without the

Figure 1.9 - Osmolyte Effects on Folate Dimerization. Effect of folate concentration on the chemical shifts in 10 mM Tris-*d*11 (□), 10 mM Tris-*d*11 with 20% deuterated betaine (△), and 10 mM Tris-*d*11 with deuterated DMSO (○). The pH was 7.1. Lines through the data show the fit to a dimerization equation (Eq 2 in Duff et. al.)[44]. Data from Duff et al.[44]

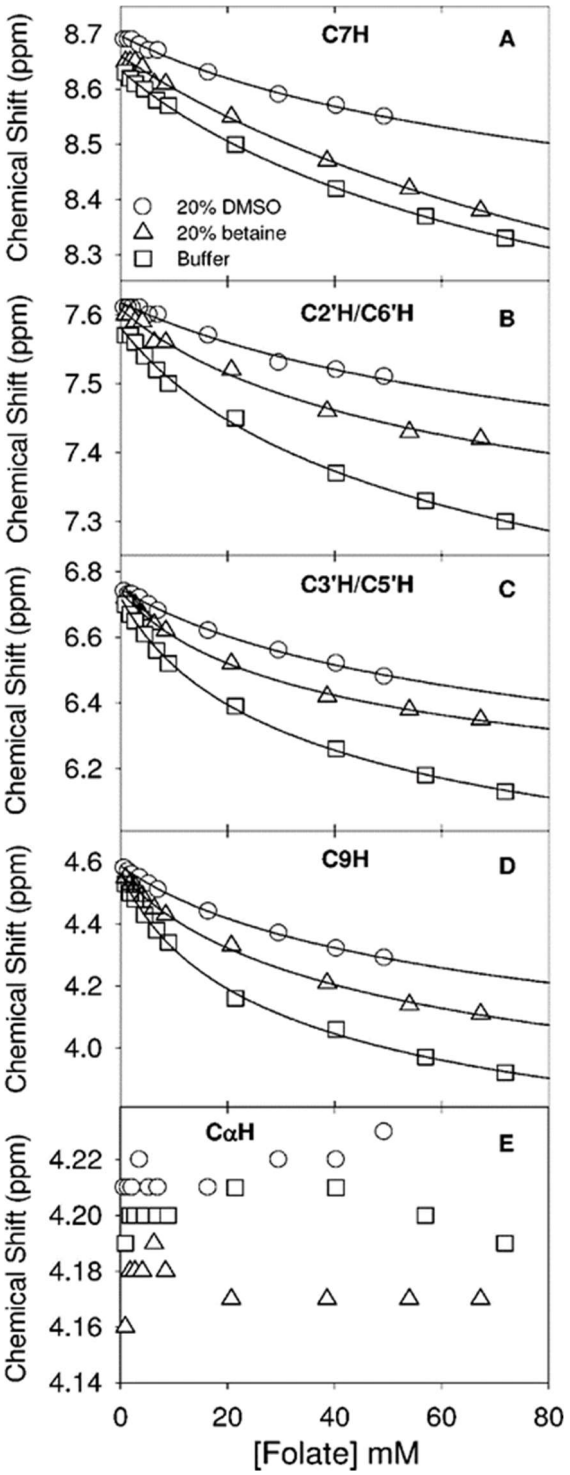

C7H
A
20% DMSO
20% betaine
Buffer
C2'H/C6'H
B
C3'H/C5'H
C
C9H
D
CαH
E
Chemical Shift (ppm)
[Folate] mM

Table 1.2. - Dimerization constants for folate obtained from NMR studies. The change proton chemical shifts in 10 mM Tris-DCl buffer at pH 7.1 with and without osmolyte addition was noted for protons at different positions and the dimerization constant was obtained upon fitting each data for proton to the dimerization equation (Eq 2 from Duff et.al.)[44] as listed in the Table. The sum of all these proton chemical shifts was also fit to the dimerization equation and the Kd obtained is shown in the last column. Data taken from Duff et.al.[44]

Osmolyte	**Osmolality (Osm)**	**K_d (mM)**				
		C7	**C9**	**C3'/C5'**	**C2'/C6'**	**Sum**
No osmolyte	0.20	250 ± 20	60 ± 10	80 ± 10	150 ± 30	80 ± 10
20 % betaine	2.34	430 ±130	110 ± 20	90 ± 10	160 ± 50	160 ± 40
20 % DMSO	3.23	260 ± 70	160 ± 10	200 ± 40	280 ± 170	200 ± 20

osmolytes. The NOEs noted in the spectra were between the protons on C7 with protons on the C9 and C3′/C5′ atoms, and between the C9 proton and the C3′/C5′ protons, which were expected for an extended monomeric folate. However, a change in sign for the NOE between the C9 and C3'/C5' protons from positive (without osmolytes) to negative (with betaine or DMSO) was observed as seen in Figure 1.10.[44] This change in the sign of the NOE suggested a slower rotational rate of the *p*-ABA ring protons, which is indicative of an interaction between betaine (or DMSO) with the *p*-ABA ring of folate.[44] Due to a lack of protons on the pterin ring, we were unable to monitor its interaction with betaine or DMSO.

1.8.B. Previous Osmotic Stress Studies with Folate

Osmotic stress experiments probe the changes in solvation of folate molecules upon addition of osmolytes. If osmolytes enter the hydration layer of folate, the bulk water activity increases, whereas if the osmolyte is retained in the bulk solution, the water activity decreases. The changes in bulk water activity (osmolality) are proportional to the number of water molecules that exclude osmolyte from the hydration layer. Experiments were performed to measure osmolalities of folate solutions with and without addition of osmolytes. The number of water molecules in the hydration layer of folate that exclude osmolytes were determined from the change in osmolality. The number of water molecules in the folate hydration shell were predicted to be higher than observed from the experimental data. The prediction was done from the accessible surface area of folate assuming a single layer of water with an area of 9 $Å^2$ per water molecule. The lesser number of water molecules in presence of osmolytes indicates interaction of osmolytes with folate. Thus, preferential solvation of folate by osmolytes betaine, DMSO, glycerol, ethylene

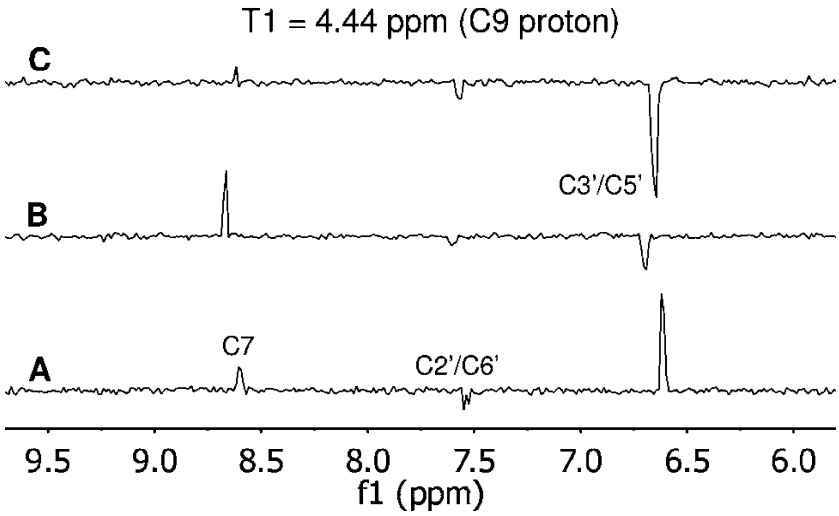


Figure 1.10 - Osmolyte Dependent Changes in Folate NOEs. Stacked NOESY spectra slices for 5 mM folate in (A) 10 mM Tris-DCl buffer (pH 7.0), (B) 10 mM Tris-DCl (pH 7.0) with 20% DMSO, and (C) 10 mM Tris-DCl (pH 7.0) with 20% betaine. Comparison of the spectra shows a change in sign of the NOEs for the protons from C9 with the C3′/C5′ protons. Data from Duff et al.[44]

glycol, sucrose and PEG400 was detected. Similar osmometry experiments with the model folate fragments *p*-aminobenzoyl glutamate (*p*-ABA-Glu) and pteridine-6-carboxylate (P6C) with osmolytes showed interactions with both the folate fragments.[44]

As betaine is the osmolyte that is most strongly excluded from protein surfaces, results obtained for betaine interactions with folate gave deeper insights on the preferential interactions between osmolytes and substrate resulting in weakened DHF binding to DHFRs (model in Figure 1.7). The pterin and *p*-ABA rings of folate are proposed to form cation-π interactions with the quaternary amine of betaine. Thus, our former studies indicate betaine interacts with the pterin and the *p*-ABA rings of the folate monomer and hinders dimerization of folate as noted by NMR. Betaine was also found to preferentially solvate folate with our osmometry studies.[44] In view of these background studies, we were interested in further characterization and quantification of this weak interaction between folate and betaine. As these interactions are comparable to interactions with water, they are challenging to be quantified. Two approaches to quantify these interactions use vapor pressure osmometry (VPO) and solubility assays.

1.8.C. Vapor Pressure Osmometry (VPO)

Extensive studies have been done by the Record lab to investigate the energetics of osmolyte interactions with functional groups as compared to water.[48-50] The Record lab has used a VPO method and a water-accessible surface area (ASA) analysis to quantify and analyze the thermodynamics of interaction of osmolytes with model compounds displaying biomolecular functional groups.[48] The VPO method measures the favorability of a small molecule interacting with another solute/osmolyte as compared to water in a three-component system (1-water, 2-test compound and 3-osmolyte). These studies

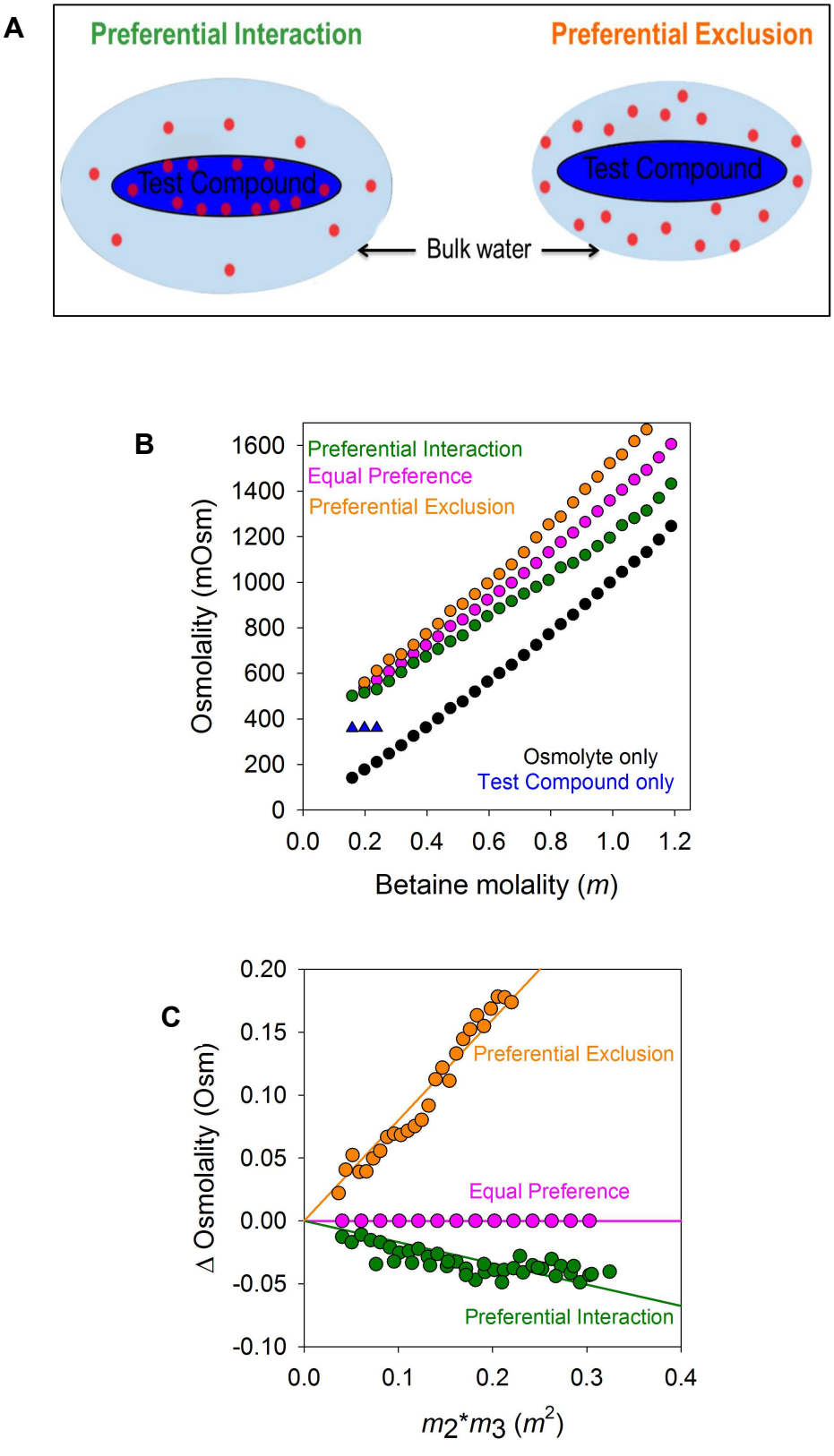
quantify the preferential interaction potentials (μ_{23}/RT values) for the osmolytes interacting with exposed functional groups on various compounds. The μ_{23}/RT value describes the change in chemical potential of the test compound with the change in molality of the osmolyte in solution. The change in osmolality going from the two component solutions (osmolyte in water and test compound in water) to the three component solution can be measured. Data analysis using Eq (1.2) gives the preferential interaction potential (μ_{23}/RT):

$$\Delta\text{Osm} \cong \left(\frac{\mu_{23}}{\text{RT}}\right) m_2\, m_3 \qquad \text{Eq (1.2)}$$

where ΔOsm is the difference in osmolalities of solution with the test compound and osmolyte and the corresponding two component solutions, m_2 and m_3 are the molal concentrations of test compound and osmolyte, respectively.

Figure 1.11.A displays two possible mechanisms of action of osmolytes. Preferential interaction mechanism allows the osmolyte molecules to enter the hydration shell and displace water molecules that in turn enter the bulk. The concentration of osmolytes in the bulk is lowered resulting in a lower osmolality. On the other hand, preferential exclusion of osmolyte molecules from the hydration shell, results in higher concentration of osmolytes in bulk and therefore higher osmolality. The raw data plots shown in Figure 1.11.B displays the measurements of osmolalities for the three solution conditions done in a VPO experiment. The osmolalities of the two component solutions (osmolyte in water and test compound in water) are measured initially, followed by measuring the osmolality of the three component solution (test compound and osmolyte in water). The changes in osmolalities with and without addition of the test compound

Figure 1.11 - VPO Experiments Analyze Preferential Interaction and Exclusion between Osmolytes and Test Compounds. The top panel represents the two possibilities in the three component samples of VPO experiments. If the osmolytes are preferred in the hydration layer, the osmolality in the bulk is lower whereas if the osmolytes are excluded from the hydration layer, they remain in the bulk and result in a higher osmolality. Panel B shows the osmolality versus osmolyte molality plots obtained for preferential interaction (●) and preferential exclusion mechanisms (●). Data points showing the osmolality for equal preference for water and osmolyte situation are given in magenta (●). The osmolalities with increasing concentrations of the osmolyte in water are shown as (●). The osmolality of the test compound at a fixed concentration is given by (▲) points. The corresponding slopes are shown in panel C.


A
Preferential Interaction
Preferential Exclusion
Test Compound
Test Compound
Bulk water
B
Preferential Interaction
Equal Preference
Preferential Exclusion
Osmolyte only
Test Compound only
Osmolality (mOsm)
Betaine molality (m)
C
Preferential Exclusion
Equal Preference
Preferential Interaction
Δ Osmolality (Osm)
m2*m3 (m2)

(ΔOsm) can be analyzed using Eq (1.2) and the slope of the plot of change in osmolality versus osmolyte concentration (in molality) gives the preferential interaction potential which is the μ_{23}/RT value. Negative slope (μ_{23}/RT) represents preferential interaction whereas positive slope (μ_{23}/RT) represents preferential exclusion of the osmolyte. A slope of zero is obtained when ΔOsm = 0 indicating equal preference of osmolytes and water in the hydration shell of the test compound. The sign and magnitude of interaction potentials determine the favorability and extent of preferential interaction, respectively. Negative interaction potentials indicate a preference of the test compound for osmolyte over water and positive interaction potentials indicate a preference for water. These interaction potentials were deconvoluted into atomistic interaction potentials (α values) that depend on the accessible surface area of each functional group.

Thus, interaction of osmolytes like betaine, proline, urea, glycerol and PEG with various functional groups were obtained. Table 1.3 compares the atomistic interaction potentials of betaine, proline and urea with different atom types. Using this information, the role of water in biomolecular processes like protein folding, assembly and protein-DNA interactions has been interpreted.[48-52]

Capp et al.[48] have studied the interaction of betaine with a set of model compounds containing carboxylate, phosphate, amide, hydroxyl, ammonium, guanidinium, aliphatic and aromatic hydrocarbon moieties. The μ_{23}/RT values obtained for these compounds were dissected into additive contributions from chemically distinct functional groups. The established set of values for each of the surface types coupled with the water-accessible surface areas (ASA) can be used to predict the μ_{23}/RT of any compound. Betaine was found to preferentially interact with aromatic groups, amide

Table 1.3 - Atomistic interaction potentials (α values) deconvoluted from the μ23/RT values obtained for several test compounds using vapor pressure osmometry studies to quantify interactions with osmolytes.[48-50] Negative interaction potentials mean preferential interaction (shown in red) and positive values mean preferential exclusion (shown in blue).

Atom type	Betaine	Proline	Urea
	α values ($10^4 \times \alpha$ (m^{-1} $Å^{-2}$))		
Aliphatic carbon	3 ± 3	5.3 ± 1.3	–1.1 ± 0.5
Aromatic C	–23 ± 4	–9.2 ± 0.9	–8.9 ± 0.5
Hydroxyl O	1 ± 2	–0.7 ± 1.3	–2.5 ± 0.6
Amide O	28 ± 10	14.5 ± 4.5	–8.5 ± 1.8
Amide N	–20 ± 7	–11.8 ± 3.2	–3.7 ± 1.6
Carboxylate O	29 ± 2	16.6 ± 4.3	–3.7 ± 1.6
Phosphate O	49 ± 4	18.0 ± 5.2	–5.8 ± 1.2
Cationic N	–12 ± 4	–12.6 ± 4.3	1.6 ± 1.7

nitrogens and cationic nitrogens with negative α values whereas betaine is strongly excluded from anionic and amide oxygens.[48]

As betaine is not a hydrogen bond donor, it was suggested to be a good hydrogen bond acceptor with amide and cationic nitrogens relative to water. The favorable interactions of betaine with aromatic surfaces was attributed to cation-π interactions between the cationic quaternary amine of betaine and the π electron cloud on aromatic surfaces. Formation of such interactions are also evident from crystal structures of betaine binding/transporting proteins. Betaine binding sites in these proteins are lined by aromatic residues forming a "box" that accommodates betaine via cation-π interactions.[53]

Betaine and proline are compatible solutes that have no to minimal deleterious effects on biomolecule stability and function. These two osmolytes were suggested to form energetically unfavorable interactions with aliphatic hydrocarbon, amide and/or oxygen surfaces that are exposed upon protein unfolding. Betaine and proline are also excluded from hydroxyl groups and negatively charged functional groups like carboxylate and phosphate oxygen. The basis of this exclusion is the inability of these osmolytes to be H-bond donors. The macromolecules prefer to be solvated by water and maintain their natively folded and functional states. Thus, the osmoprotectants work by the preferential exclusion mechanism and facilitate burial of these surface types. This finding indicates that betaine and proline are effective osmolytes as these molecules are preferentially excluded from the macromolecular hydration shell, and are retained in the bulk solvent thereby increasing the osmolality and countering osmotic stress. This mechanism of preferential exclusion of osmolytes from the macromolecular surfaces aids the cell to increase its volume thereby increasing the amount of intracellular water.[4]

Urea was found to interact favorably with most of the functional groups owing to its multiple H-bond donating and accepting capabilities.[50] Preferential interactions of urea with amide backbone, aliphatic and aromatic surfaces contribute to its ability to denature proteins. Urea is not used as an osmolyte by *E.coli* cells.[6]

In another study by the Record group, glycerol and tetra ethylene glycol were tested for preferential interactions with several small compounds displaying various macromolecular surface types.[52] This allowed modeling of interaction potentials of proteins for polyethylene glycol (PEG). Glycerol is known to stabilize proteins *in vitro* due to preferential exclusion from aliphatic and amide oxygen surfaces.

1.8.D. Solubility Assays

Solubility assays provide another approach to study thermodynamics of the weak preferential interactions to determine the transfer free energy of a test compound from one solvent system to another. The transfer free energy of a test compound from a pure aqueous system to a mixed aqueous osmolyte system is based upon the solubility of the solute in each of the systems. At the solubility limit of the test compound, i.e. at equilibrium, the chemical potential of the test compound in both the solvent systems must be equal.[54] Using this principle, transfer free energy measurements can be used to identify and understand weak preferential interactions. The magnitude and sign of the transfer free energy determines the preference of the test compound for one system over the other. Negative transfer free energy indicates favorable solvation of the test compound by osmolyte solution as compared to pure water.

1.9 How Strong are the Preferential Interactions between Folates and Osmolytes?

We aim to quantify interactions between folates and osmolytes. The quantitation of these weak short-range interactions between solutes relative to their interactions with water is challenging, however it is essential for better understanding of enzyme function *in vivo* where a crowded environment displays multiple functional groups available for participating in these preferential interactions. Thus, prior knowledge of the extent of these interactions will be beneficial in predicting the effects of osmolytes. VPO experiments would provide a scale for studying weak interactions between folate and betaine relative to water. Therefore, μ_{23}/RT values can be used to predict osmotic stress effects on ligand binding. In addition, solubility assays can quantify free energy of transfer of folate from water to 1 M betaine solution. The results from these studies are discussed in Part 3.

1.10 Further Osmotic Stress Studies on R67 DHFR

As discussed earlier, the role of water in biochemical process can be probed by addition of small molecule osmolytes that perturb water activity in osmotic stress studies. Alteration in water activity results in changes in the hydration of macromolecules like proteins. The protein associated water molecules exclude the osmolytes forcing them to remain in the bulk. Results from our previous osmotic stress experiments with R67 DHFR indicate preferential exclusion of osmolytes from the protein surface.[42] The tightened binding of cofactor was noted in presence of all osmolytes tested. Moreover, the data showed the similar extent of tightening effects, indicating preferential exclusion of all osmolytes tested. To further assess osmolyte effects on protein, we chose to perform further

studies using a coupled osmotic stress and small angle neutron scattering (SANS) approach.

1.11 Small Angle Neutron Scattering (SANS)

Previous applications have shown SANS to be an efficient technique that can be coupled with the osmotic stress experiments to study the solvation of proteins.[55-57] In SANS, a beam of neutrons is elastically scattered by a sample and the resulting scattering pattern is analyzed to provide information about the size, shape and orientation of some component of the sample. The sample is placed in a cuvette and the intensity of scattered neutrons is detected as a function of scattering angle. The SANS profile contains the scattering intensity as a function of the amplitude of the scattering vector or momentum transfer, q, given by Eq (1.3)

$$q = \frac{4\pi sin}{\lambda} \qquad \text{Eq (1.3)}$$

where λ is the wavelength of neutrons and θ is half the angle between the incident and scattered neutrons. At a constant λ, the scattering profile I(q) versus q represents the scattering intensity as a function of scattering angle.

The SANS profile gives information about the shape and composition of the sample molecules. The analysis gives the apparent radius of gyration (R_g) and zero angle scattering intensity (I(0)) of the protein molecules in sample. Guinier analysis uses a linear plot of ln I(q) versus q^2 with a slope that is equal to $-(R_g^2)/3$ and the intercept on the Y-axis gives the

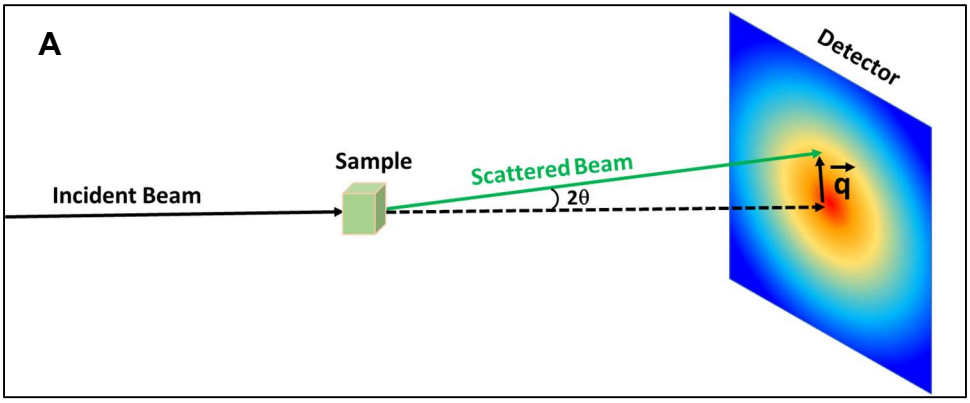
I(0) value. Guinier analysis gives an estimate of the radius of gyration of protein (R_g) and the zero angle scattering intensity I(0) using Eq (1.4) [58]

$$I(q) = I(0)e^{-q^2 {R_g}^2/3} \qquad \text{Eq (1.4)}$$

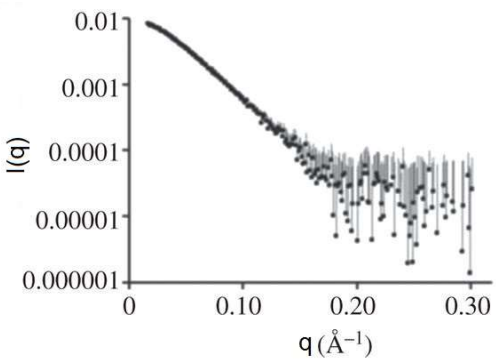
where I(q) and I(0) are the scattering intensities at small angles (q) and at zero angle respectively; and R_g is the protein radius of gyration. Guinier analysis uses the data points at low q value. Further analysis of the data can also be done using the GNOM program in the ATSAS package.[59] GNOM utilizes a Fourier transform of the scattering curve to provide the probabilities of distances between the scattering particles and the maximum dimension of the scattering species. It evaluates the particle distance distribution function, P(R), in a defined range of distribution and yields the apparent radius of gyration (R_g) and zero angle scattering intensity I(0). Figure 1.12 shows the experimental set up and preliminary data analysis. An advantage of neutron scattering is the difference in the scattering from hydrogen and deuterium allowing for the identification of scattering signal from individual components within a sample.

The scattering of neutrons from the target molecules (protein) and the bulk solvent can be differentiated by generating a contrast of a hydrogenated protein in D_2O buffer or deuterated protein in H_2O. Addition of hydrogenated osmolytes creates an additional contrast that differentiates scattering into three regions- the protein, the protein associated water and the bulk solution. Thus, the hydration layer also contributes to the scattering intensity determined for the protein. The scattering profiles obtained at varying osmolyte concentrations provide the changes in R_g and I(0), which gives information about the

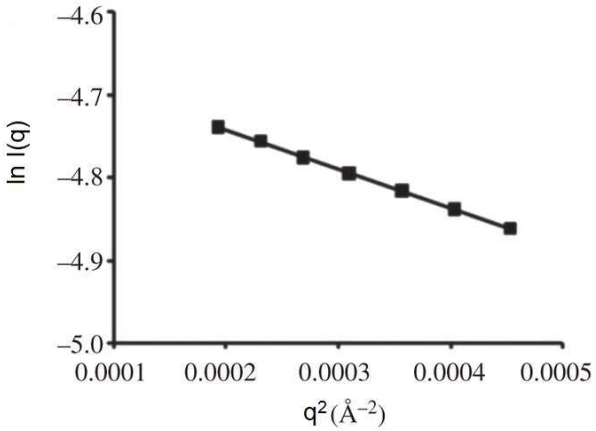
Figure 1.12 - Small Angle Scattering Experimental Setup and Data Analysis. Panel A shows the experimental set up of a neutron beam incident on the sample and the detection of scattered beam onto a detector. Panel B shows the representative scattering profile (I(q) versus q). Data fitting by Guinier analysis and the Fourier transform by GNOM that gives a pair-distance distribution are shown in panels C and D respectively. Panel B, C and D are adapted from Reference.[60]


A
Detector
Sample
Scattered Beam
Incident Beam
2θ
q

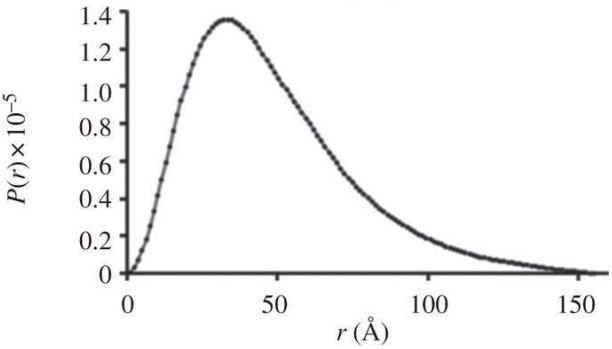
B


0.01
0.001
0.0001
0.00001
0.000001
I(q)
0
0.10
0.20
0.30
q (Å⁻¹)

C


–4.6
–4.7
–4.8
–4.9
–5.0
ln I(q)
0.0001
0.0002
0.0003
0.0004
0.0005
q² (Å⁻²)

D


1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
P(r) × 10⁻⁵
0
50
100
150
r (Å)

number of osmolyte excluding water molecules in the hydration layer. The variation in I(0) with osmolyte concentration gives the number of hydrating water molecules, whereas the variation in R_g gives both the number and location of the water molecules in the hydration layer. Thus, the waters responsible for osmolyte exclusion can be selectively studied in a multi-component solution using SANS. This method has been used by researchers to study the preferential hydration of small proteins.[55-57]

Recently, Stanley et al. probed the preferential hydration of two proteins- lysozyme and guanylate kinase- with addition of osmolytes such as betaine and polyethylene glycols.[56] The extent of osmolyte exclusion from these proteins as determined by SANS was supported by ITC studies monitoring osmolyte effects on ligand binding.[56] We utilize the SANS approach with an osmolyte mediated contrast variation by adding hydrogenated osmolytes to the protein solution in deuterated buffer, which enabled us to selectively obtain data for the hydration layer. The results from these experiments are discussed in Part 4 of this thesis.

As R67 DHFR possess an intrinsically disordered region at the N-terminus of each of its monomeric subunit, it offered us to explore additional aspects of R67 DHFR structure using SANS while concurrently understanding the osmolyte effects on R67 DHFR. The following section recaps the structural features of R67 DHFR and provides additional details about the disordered the N-termini.

1.12 Disordered Tails of R67 DHFR

R67 DHFR was described above (section 1.4.B). It is the plasmid-encoded DHFR that confers resistance against trimethoprim, a potent inhibitor of *E.coli* chromosomal

DHFR (EcDHFR). R67 DHFR is a homotetramer with a single active site pore.[39] Two monomers dimerize forming a six-stranded β barrel at the interface.[26] The tetramer is the dimer of dimers involving loop–loop interactions. The crystal structure of the inactive dimeric species yielded weak electron density for the first 16 -18 residues of the monomers suggesting disorder in the N-terminal tails.[26] Additionally, the electron density for the 21st residue was weak with a high thermal factor. Use of several disorder predictors indicate the N-terminal sequence is intrinsically disordered.[61] The N-terminal sequence can be cleaved after Phe16 by chymotrypsin treatment and the truncated tetrameric protein was crystallized by Narayana et. al. at a resolution of 1.7Å.[39] Further, the structure was refined to a 1.1 Å resolution.[62] In addition to the first 18 amino acids, residues 20 and 21 also exhibited diffused electron densities and high thermal factors in the refined structure obtained under cryo-cooling conditions at 100K. Thus, the stretch of residues 17–21 appears to be disordered independent of the temperature at which the data were collected. Electron densities for 21-23 were noted to be diffuse indicating high mobility.[62]

Truncated R67 DHFR lacking the first 16 residues at the 4 N-termini retains enzymatic activity with no significant changes in the k_{cat} and K_m parameters as determined by steady state kinetics (see Table 1.4).[36] Also, *in vitro* protein unfolding studies found the truncated R67 DHFR dimer to be less stable by 2.6 kcal/mol as compared to the full length protein.[36] In addition, when a gene encoding the truncated protein is constructed, no trimethoprim resistance is observed *in vivo*.[36] This suggests that either the N-termini play a role in the *in vivo* stability and folding of R67 DHFR or different mRNA stabilities result in protein expression differences.

Table 1.4 - Comparison of kinetic parameters of full length and truncated R67 DHFR. Data is obtained from Reece., et al.[36]

Enzyme	k_{cat} (s^{-1})	$K_{M(DHF)}$ (μM)	$K_{M(NADPH)}$ (μM)
R67 DHFR	1.3 ± 0.1	5.8 ± 0.1	3.0 ± 0.1
Truncated R67 DHFR	1.5 ± 0.1	7.2 ± 0.2	2.7 ± 0.1

Other type II variants of DHFR (e.g. R388, R751) show different N-terminal sequences but the same core sequence contributes to the β-barrel structure.[61] Figure 1.13 shows the sequence alignment of the three type II DHFR variants. The disordered N-terminal sequences of these variants show non-identity whereas the rest of the protein sequence is largely identical. A tandem array of four R67 DHFR gene copies encodes a protein where the C- and N-termini of 1st and 2nd monomers are fused. Similar fusions were generated between the 2nd and 3rd monomers and the 3rd and 4th monomers as well. The N-termini regions serve as linker sequences and the resulting Quad1 protein was a monomer with four times the molecular mass of the R67 DHFR subunit. Quad1 was found to be stable as well as functional.[63] Another mutational design strategy included alternate linker sequences using the disordered N-terminal sequences from other type II DHFRs. The resulting Quad4 protein was found to be quite functional, however, it is less stable than Quad1.[61] These various observations suggest that the N-terminal sequences of each of these variants are malleable as well as unique. They may also play a role in the stability of the overall protein structure.

1.13 Exploring Possible Role of the Disordered Tails

Disordered tails exhibit interconverting conformations that results in an ensemble structure in solution. This inherent structural plasticity can result in functional diversity of disordered tails. The regions of intrinsic disorder in proteins have been known to perform various functions ranging from forming display sites, effectors, assemblers to just being flexible entropic chains.[64, 65] The disordered sequences may function in the cellular signaling pathway by displaying readily accessible sites for post translational

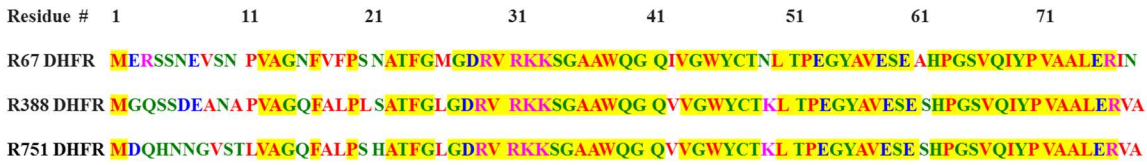


Figure 1.13 - Sequence alignment of the type II DHFR variant sequences. Sequence alignment is shown for comparison of three type II DHFR variants. Identical residues are highlighted in yellow. Residues are color coded as per the sidechain property. (red-non polar, blue-negatively charged, magenta- positively charged and green-neutral). The sequence of R67 DHFR N-terminus was modified to ^{1}MIRSSNEVSN^{10} to incorporate a restriction endonuclease site.

modifications. The effectors are the intrinsically disordered regions that are recognized by other proteins and/or ligands leading to interactions that may regulate the protein activity. These interactions often involve a disorder to order transition in the intrinsically disordered regions. The role played by intrinsically unstructured chains in recognition and interaction may also drive assembly of higher-order complexes. Intrinsically disordered segments in proteins can function as entropic chains that sample a wide array of conformations ranging from fully extended to compact forms. These entropic chains act as linkers or spacers between domains. Additionally, disordered entropic chains may act as bristles protruding from the ordered proteins. These entropic bristles stabilize the protein by preventing aggregation.[66] Studies have shown that the disordered tails of ordered proteins often undergo random movements and sample a large surface area. These unstructured extensions, entropic bristles, are proposed to aid in enhancing the soluble expression of proteins. The disordered tails can sample extended conformations exposing larger surface areas to the solvent. For example, studies support the role of the C-terminal entropic bristle in soluble expression of a human homologue of *E.coli* DNA glycosylase endonuclease VIII.[67] These extended tails are thought to provide larger surface areas for hydration thereby stabilizing the proteins. The stability of SUMO proteins is attributed to the disordered N terminus that samples a large volume and restricts intermolecular interactions which could lead to aggregation.[68]

1.14 What is the Role of Disordered N-Termini of R67 DHFR?

In vivo TMP resistance is not conferred on to the host cells by a gene encoding the truncated R67 DHFR, indicating a role of the disordered N-termini in protein expression

and stability. We used SANS to gain a deeper insight in the changes in the N-termini conformation in different solution conditions. The SANS data for the apo protein contains structural information for the core as well as the N-termini tails. As there is no structural information about the N-termini, SANS data could be utilized to model the ensemble of conformations sampled by the disordered tails in solution. The envelope of different possible conformations would give deeper insight on the functional role of the N-termini. One possibility is that the tails would form interactions with the ordered surface and/or with other tail and can render higher stability to the protein.

As the main focus of our SANS measurements was to study osmotic stress effects on R67 DHFR, we were curious to examine a possibility of structure attainment in the N-terminal tails upon osmolyte addition. Studies have shown that osmolytes exert protein stabilizing forces via preferential exclusion mechanism. The ability of TMAO to force folding of a modified RNase was attributed to its preferential exclusion from the peptide backbone also termed as the solvophobic effect.[9, 69] This effect of TMAO on the peptide backbone strongly destabilizes unfolded state thereby forcing the protein to fold. Thus, TMAO and other excluded osmolytes can drive folding of intrinsically disordered proteins. As betaine is one of the most excluded osmolyte, addition of betaine may drive a disorder to order transition in the N-termini of R67 DHFR that can be evident from the changes in the apparent radius of gyration (R_g) of the protein. SANS provides an excellent opportunity to unambiguously measure the protein's R_g with addition of deuterated betaine to create a contrast between the protein and its environment.

Also, coupled binding and folding often occurs in intrinsically disordered proteins as the Gibbs energy of the native state is lowered by using the binding energy of ligands to

drive folding. DNA binding was shown to induce proteins with disordered domains to fold.[70, 71] Ligand binding to R67 DHFR might result in structural changes in the flexible tails that enable them to attain order. Our previous pressure perturbation calorimetry (PPC) and densitometry measurements suggest R67 DHFR may become more compact when DHF binds to the protein-$NADP^+$ complex.[72] However, this prediction involves subtraction of large values from large values (e.g. molar expansivity values obtained by PPC experiments for the apo R67 DHFR, both the binary complexes and the ternary complex) leaving a small value with substantial error. Thus, the change in molar expansivity upon complex formation could not be robustly determined.[72] SANS is a better suited tool to investigate any changes in the R67 DHFR N-terminus that may occur upon ternary complex formation. The attainment of order may result in an overall compaction of the structure, which may affect the R_g of the protein. The data collected using this method are discussed in Part 4 of the thesis.

The four N-termini of R67 DHFR can potentially function as entropic bristles and aid in the soluble expression of the protein. Our data from SANS experiments could be systematically analyzed to get information about the conformations sampled by the N-termini in solution. This analysis employs generation of models for the sampling of the disordered tails using the experimental SANS data as constraints. The best models obtained provide a deeper insight on the realistic conformations of the N-termini in solution, which may suggest the entropic bristle like function. The details of this analysis and its outcomes are discussed in Part 4 of the thesis.

1.15 References

1. Cayley, S., et al., *Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo.* J. Mol. Biol., 1991. 222(2): p. 281-300.
2. Fulton, A.B., *How crowded is the cytoplasm?* Cell, 1982. 30(2): p. 345-347.
3. Timasheff, S.N., *The control of protein stability and association by weak interactions with water: how do solvents affect these processes?* Annu. Rev. Biophys. Biomol. Struct., 1993. 22: p. 67-97.
4. Parsegian, V.A., R.P. Rand, and D.C. Rau, *Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives.* Proc. Natl. Acad. Sci. U. S. A., 2000. 97(8): p. 3987-3992.
5. Cayley, S. and M.T. Record, Jr., *Roles of cytoplasmic osmolytes, water, and crowding in the response of Escherichia coli to osmotic stress: biophysical basis of osmoprotection by glycine betaine.* Biochemistry, 2003. 42(43): p. 12596-12609.
6. Cayley, D.S., H.J. Guttman, and M.T. Record, Jr., *Biophysical characterization of changes in amounts and activity of Escherichia coli cell and compartment water and turgor pressure in response to osmotic stress.* Biophys. J., 2000. 78(4): p. 1748-1764.
7. Record, M.T., Jr., et al., *Responses of E. coli to osmotic stress: large changes in amounts of cytoplasmic solutes and water.* Trends. Biochem. Sci., 1998. 23(4): p. 143-148.
8. Cayley, S., B.A. Lewis, and M.T. Record, Jr., *Origins of the osmoprotective properties of betaine and proline in Escherichia coli K-12.* J. Bacteriol., 1992. 174(5): p. 1586-1595.
9. Baskakov, I. and D.W. Bolen, *Forcing thermodynamically unfolded proteins to fold.* J. Biol. Chem., 1998. 273(9): p. 4831-4834.
10. Timasheff, S.N., *Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components.* Proc. Natl. Acad. Sci. U. S. A., 2002. 99(15): p. 9721-9726.
11. Courtenay, E.S., et al., *Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro.* Biochemistry, 2000. 39(15): p. 4455-4471.
12. Reid, C. and R.P. Rand, *Probing protein hydration and conformational states in solution.* Biophys. J., 1997. 72(3): p. 1022-1030.
13. Colombo, M.F., D.C. Rau, and V.A. Parsegian, *Protein solvation in allosteric regulation: a water effect on hemoglobin.* Science, 1992. 256(5057): p. 655-659.
14. Amyes, S.G. and J.T. Smith, *The purification and properties of the trimethoprim-resistant dihydrofolate reductase mediated by the R-factor, R388.* Eur. J. Biochem., 1976. 61(2): p. 597-603.
15. Fierke, C.A., et al., *Implications for enzymic catalysis from free-energy reaction coordinate profiles.* Cold Spring Harb. Symp. Quant. Biol., 1987. 52: p. 631-638.
16. Bolin, J.T., et al., *Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. I. General features and binding of methotrexate.* J Biol Chem, 1982. 257(22): p. 13650-13662.

17. Bystroff, C. and J. Kraut, *Crystal structure of unliganded Escherichia coli dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding.* Biochemistry, 1991. 30(8): p. 2227-2239.
18. Boehr, D.D., et al., *The dynamic energy landscape of dihydrofolate reductase catalysis.* Science, 2006. 313(5793): p. 1638-1642.
19. Sawaya, M.R. and J. Kraut, *Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence.* Biochemistry, 1997. 36(3): p. 586-603.
20. Loveridge, E.J., et al., *Evidence that a 'dynamic knockout' in Escherichia coli dihydrofolate reductase does not affect the chemical step of catalysis.* Nat. Chem., 2012. 4(4): p. 292-297.
21. Yahashiri, A., E.E. Howell, and A. Kohen, *Tuning of the H-transfer coordinate in primitive versus well-evolved enzymes.* Chemphyschem., 2008. 9(7): p. 980-982.
22. Howell, E.E., *Searching sequence space: two different approaches to dihydrofolate reductase catalysis.* Chembiochem., 2005. 6(4): p. 590-600.
23. Luk, L.Y., E.J. Loveridge, and R.K. Allemann, *Protein motions and dynamic effects in enzyme catalysis.* Phys Chem Chem Phys, 2015. 17(46): p. 30817-30827.
24. Bolin, J., et al., *Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. I. General features and binding of methotrexate.* J. Biol. Chem., 1982. 257: p. 13650.
25. Filman, D.J., et al., *Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. II. Environment of bound NADPH and implications for catalysis.* J. Biol. Chem., 1982. 257(22): p. 13663-13672.
26. Matthews, D.A., et al., *Crystal structure of a novel trimethoprim-resistant dihydrofolate reductase specified in Escherichia coli by R-plasmid R67.* Biochemistry, 1986. 25(15): p. 4194-4204.
27. Narayana, N., et al., *A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site.* Nat. Struct. Biol., 1995. 2(11): p. 1018-1025.
28. Gourley, D.G., et al., *Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites.* Nat. Struct. Biol., 2001. 8(6): p. 521-525.
29. Liang, J., H. Edelsbrunner, and C. Woodward, *Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design.* Protein Sci., 1998. 7(9): p. 1884-1897.
30. Stone, S.R. and J.F. Morrison, *Mechanism of inhibition of dihydrofolate reductases from bacterial and vertebrate sources by various classes of folate analogues.* Biochim. Biophys. Acta, 1986. 869(3): p. 275-285.
31. Giladi, M., et al., *FolM, a new chromosomally encoded dihydrofolate reductase in Escherichia coli.* J. Bacteriol., 2003. 185(23): p. 7015-7018.
32. Howell, E.E., et al., *Functional role of aspartic acid-27 in dihydrofolate reductase revealed by mutagenesis.* Science, 1986. 231(4742): p. 1123-1128.
33. Jackson, M., et al., *Calorimetric studies of ligand binding in R67 dihydrofolate reductase.* Biochemistry, 2005. 44(37): p. 12420-12433.

34. Nare, B., L.W. Hardy, and S.M. Beverley, *The roles of pteridine reductase 1 and dihydrofolate reductase-thymidylate synthase in pteridine metabolism in the protozoan parasite Leishmania major.* J. Biol. Chem., 1997. 272(21): p. 13883-13891.
35. Howell, E.E., et al., *A second-site mutation at phenylalanine-137 that increases catalytic efficiency in the mutant aspartate-27-serine Escherichia coli dihydrofolate reductase.* Biochemistry, 1990. 29(37): p. 8561-8569.
36. Reece, L.J., et al., *Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein.* Biochemistry, 1991. 30(45): p. 10895-10904.
37. Dion, A., et al., *How do mutations at phenylalanine-153 and isoleucine-155 partially suppress the effects of the aspartate-27-->serine mutation in Escherichia coli dihydrofolate reductase?* Biochemistry, 1993. 32(13): p. 3479-3487.
38. Pribat, A., et al., *FolX and FolM are essential for tetrahydromonapterin synthesis in Escherichia coli and Pseudomonas aeruginosa.* J. Bacteriol., 2010. 192(2): p. 475-482.
39. Narayana, N., et al., *A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site.* Nat Struct Biol, 1995. 2(11): p. 1018-1025.
40. Bradrick, T.D., J.M. Beechem, and E.E. Howell, *Unusual binding stoichiometries and cooperativity are observed during binary and ternary complex formation in the single active pore of R67 dihydrofolate reductase, a D_2 symmetric protein.* Biochemistry, 1996. 35(35): p. 11414-11424.
41. Levin, I., et al., *An alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea.* Mol. Microbiol., 2004. 54(5): p. 1307-1318.
42. Chopra, S., et al., *A balancing act between net uptake of water during dihydrofolate binding and net release of water upon NADPH binding in R67 dihydrofolate reductase.* J. Biol. Chem., 2008. 283(8): p. 4690-4698.
43. Grubbs, J., et al., *Thermodynamics and solvent effects on substrate and cofactor binding in Escherichia coli chromosomal dihydrofolate reductase.* Biochemistry, 2011. 50(18): p. 3673-3685.
44. Duff, M.R., Jr., et al., *Weak interactions between folate and osmolytes in solution.* Biochemistry, 2012. 51(11): p. 2309-2318.
45. Bhojane, P.P., et al., *Investigation of osmolyte effects on FolM: comparison with other dihydrofolate reductases.* Biochemistry, 2014. 53(8): p. 1330-1341.
46. Poe, M., *Proton magnetic resonance studies of folate, dihydrofolate, and methotrexate. Evidence from pH and concentration studies for dimerization.* J. Biol. Chem., 1973. 248(20): p. 7025-7032.
47. Szakacs, Z. and B. Noszal, *Determination of dissociation constants of folic acid, methotrexate, and other photolabile pteridines by pressure-assisted capillary electrophoresis.* Electrophoresis, 2006. 27(17): p. 3399-3409.
48. Capp, M.W., et al., *Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values.* Biochemistry, 2009. 48(43): p. 10372-10379.

49. Diehl, R.C., et al., *Quantifying additive interactions of the osmolyte proline with individual functional groups of proteins: comparisons with urea and glycine betaine, interpretation of m-values.* Biochemistry, 2013. 52(35): p. 5997-6010.
50. Guinn, E.J., et al., *Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer.* Proc. Natl. Acad. Sci. U. S. A., 2011. 108(41): p. 16932-16937.
51. Guinn, E.J., et al., *Quantifying Functional Group Interactions That Determine Urea Effects on Nucleic Acid Helix Formation.* J. Am. Chem. Soc., 2013. 135(15): p. 5828-5838.
52. Knowles, D.B., et al., *Chemical Interactions of Polyethylene Glycols (PEGs) and Glycerol with Protein Functional Groups: Applications to Effects of PEG and Glycerol on Protein Processes.* Biochemistry, 2015. 54(22): p. 3528-3542.
53. Schiefner, A., et al., *Cation-pi interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from Escherichia coli.* J. Biol. Chem., 2004. 279(7): p. 5588-5596.
54. Nozaki, Y. and C. Tanford, *The Solubility of Amino Acids and Related Compounds in Aqueous Urea Solutions.* J. Biol. Chem., 1963. 238: p. 4074-4081.
55. Lehmann, M.S. and G. Zaccai, *Neutron small-angle scattering studies of ribonuclease in mixed aqueous solutions and determination of the preferentially bound water.* Biochemistry, 1984. 23(9): p. 1939-1942.
56. Stanley, C., et al., *Protein structure and hydration probed by SANS and osmotic stress.* Biophys. J., 2008. 94(7): p. 2777-2789.
57. Stuhrmann, H.B., et al., *Low angle neutron scattering of ferritin studied by contrast variation.* J Mol Biol, 1976. 100(3): p. 399-413.
58. Guinier, A. and G.r. Fournet, *Small-angle scattering of X-rays*. Structure of matter series. 1955, New York,: Wiley. 268 p.
59. D.I., S., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria.* J. Appl. Crystallogr. , 1992. 25: p. 595-503.
60. Andrews, D.A., et al., *Design, purification and characterization of a soluble variant of the integral membrane protein MotB for structural studies.* J R Soc Interface, 2013. 10(79): p. 20120717.
61. Feng, J., et al., *Radical redesign of a tandem array of four R67 dihydrofolate reductase genes yields a functional, folded protein possessing 45 substitutions.* Biochemistry, 2010. 49(34): p. 7384-7392.
62. Narayana, N., *High-resolution structure of a plasmid-encoded dihydrofolate reductase: pentagonal network of water molecules in the D2-symmetric active site.* Acta Crystallogr D Biol Crystallogr, 2006. 62(Pt 7): p. 695-706.
63. Bradrick, T.D., et al., *Redesigning the quaternary structure of R67 dihydrofolate reductase. Creation of an active monomer from a tetrameric protein by quadruplication of the gene.* J. Biol. Chem., 1996. 271(45): p. 28031-28037.
64. Dunker, A.K., et al., *Intrinsic disorder and protein function.* Biochemistry, 2002. 41(21): p. 6573-6582.
65. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins.* Chem Rev, 2014. 114(13): p. 6589-6631.

66. Santner, A.A., et al., *Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression.* Biochemistry, 2012. 51(37): p. 7250-7262.
67. Bandaru, V., et al., *Overproduction, crystallization and preliminary crystallographic analysis of a novel human DNA-repair enzyme that recognizes oxidative DNA damage.* Acta Crystallogr D Biol Crystallogr, 2004. 60(Pt 6): p. 1142-1144.
68. Grana-Montes, R., et al., *N-terminal protein tails act as aggregation protective entropic bristles: the SUMO case.* Biomacromolecules, 2014. 15(4): p. 1194-1203.
69. Wang, A. and D.W. Bolen, *A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation.* Biochemistry, 1997. 36(30): p. 9101-9108.
70. Pan, B., et al., *Secondary and tertiary structural changes in gamma delta resolvase: comparison of the wild-type enzyme, the I110R mutant, and the C-terminal DNA binding domain in solution.* Protein Sci, 1997. 6(6): p. 1237-1247.
71. Weiss, M.A., et al., *Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA.* Nature, 1990. 347(6293): p. 575-578.
72. Timson, M.J., et al., *Further studies on the role of water in R67 dihydrofolate reductase.* Biochemistry, 2013. 52(12): p. 2118-2127.

PART 2. INVESTIGATION OF OSMOLYTE EFFECTS ON FOLM: COMPARISON WITH OTHER DIHYDROFOLATE REDUCTASES

This part is a slightly revised version of a manuscript by the same title published in the journal, *Biochemistry.*

Reprinted (adapted) with permission from Purva P. Bhojane, Michael R. Duff, Jr., Harini C. Patel, Melissa E. Vogt, Elizabeth E. Howell*., "Investigations of Osmolyte Effects on FolM: Comparison with Other Dihydrofolate Reductases", *Biochemistry*, 2014, 53 (8), pp 1330–1341. Copyright 2014 American Chemical Society.

(**DOI:** 10.1021/bi4014165)

*Corresponding Author: Elizabeth E. Howell, Department of Biochemistry, Cell and Molecular Biology, University of TN, Knoxville, TN 37996-0840; phone, 865-974-4507; fax, 865-974-6306; e-mail, lzh@utk.edu. †This work was supported by NSF grant MCB-0817827 and internal funds from UTK.

This author contributed the following to the manuscript: (1) protein expression and purification, (2) protein extinction co-efficient, (3) isothermal titration calorimetry data, (4) steady state kinetics data, (5) water activity measurements, (6) differential scanning calorimetry data, (7) fluorescence quenching data, (8) circular dichroism data and (9) most of the data analysis and organization.

2.1 Abstract

A weak association between osmolytes and dihydrofolate (DHF) decreases the affinity of the substrate towards the *E. coli* chromosomal and R67 plasmid dihydrofolate reductase (DHFR) enzymes. To test whether the osmolyte-DHF association also interferes with DHF binding to FolM, an *E. coli* enzyme which possesses weak DHFR activity, ligand binding was monitored in the presence of osmolytes. FolM affinity for DHF, measured by $k_{cat}/K_{m\,(DHF)}$, was decreased by addition of osmolyte. Additionally, binding of the antifolate drug, methotrexate, to FolM was weakened by osmolyte addition. The changes in ligand binding with water activity were unique for each osmolyte indicating preferential interaction between osmolyte and folate and its derivatives; however additional evidence found support for further interactions between FolM and osmolytes. Binding of the

NADPH cofactor to FolM was monitored by isothermal titration calorimetry as a control for protein-osmolyte association. In the presence of betaine (proposed to be the most excluded osmolyte from protein surfaces), the NADPH K_d decreased, consistent with dehydration effects. However other osmolytes did not tighten binding to cofactor. Rather DMSO had no effect on the NADPH K_d while ethylene glycol and PEG400 weakened cofactor binding. Differential scanning calorimetry of FolM in the presence of osmolytes found that both DMSO and ethylene glycol decreased the stability of FolM, while betaine increased the stability of the protein. These results suggest that some osmolytes can destabilize FolM by preferentially interacting with the protein. Further, these weak attractions can impede ligand binding. These various contributions have to be considered when interpreting osmotic pressure results.

2.2 Introduction

If water is involved in an interaction, perturbation of water activity will alter binding. For example, closer contact distances usually exclude water. Typically, increasing the concentration of small molecule osmolytes results in tighter binding, consistent with dehydration of the protein-ligand interface, which leads to stronger binding as water is released.[1] We previously probed the role of water in R67 dihydrofolate reductase (DHFR) by adding various osmolytes to steady state kinetic assays and ITC binding experiments. Tighter binding of the NADPH cofactor and weaker binding of the substrate, dihydrofolate, upon osmolyte addition were observed.[2] While different osmolytes had similar effects on NADPH binding, variable results were observed when DHF binding was probed.

Weaker binding of DHF in the presence of osmolytes can occur by either destabilization of the enzyme-ligand complex or by stabilization of the free enzyme or the free ligand. For R67 DHFR, as each symmetry related binding site accommodates either NADPH or DHF and different behavior is observed upon osmolyte addition (either weaker DHF binding or tighter cofactor binding, e.g. water release), we can use binding of NADPH to R67 as an internal control.[2-4] This analysis suggests effects on the free enzyme or the enzyme-cofactor complex are unlikely as numerous osmolytes have the same effect on cofactor binding, consistent with a preferential exclusion mechanism where osmolytes are excluded from the protein surface.[5-8] Elimination of these options for DHF binding leaves osmolyte effects on free DHF. A corollary of the hypothesis that DHF has differential interactions with osmolytes is that related osmolyte effects should then be observed in any enzyme that uses DHF, for example the non-homologous chromosomal DHFR from *E. coli* (EcDHFR).

Using the above logic, osmotic stress studies were performed using EcDHFR.[9] Tighter binding of $NADP^+$ and weaker binding of DHF were again observed. The slopes associated with plots of ln $K_{a\ (DHF)}$, the association binding constant, vs. ln water activity were similar for EcDHFR and R67 DHFR. Since positive slope values associated with ligand binding are unusual,[10-12] and as similar values are observed for DHF binding in two quite different DHFR scaffolds, this result supports the hypothesis that osmolytes associate weakly with free DHF. (If we consider the other side of the coin, as folate is hydrophobic with a logP value of -3.875 (logP is a partition coefficient reflecting solubility in water vs. octanol), water prefers to interact with the osmolytes rather than DHF.) This model, depicted in Figure 2.1, uses a variation of the preferential interaction model where

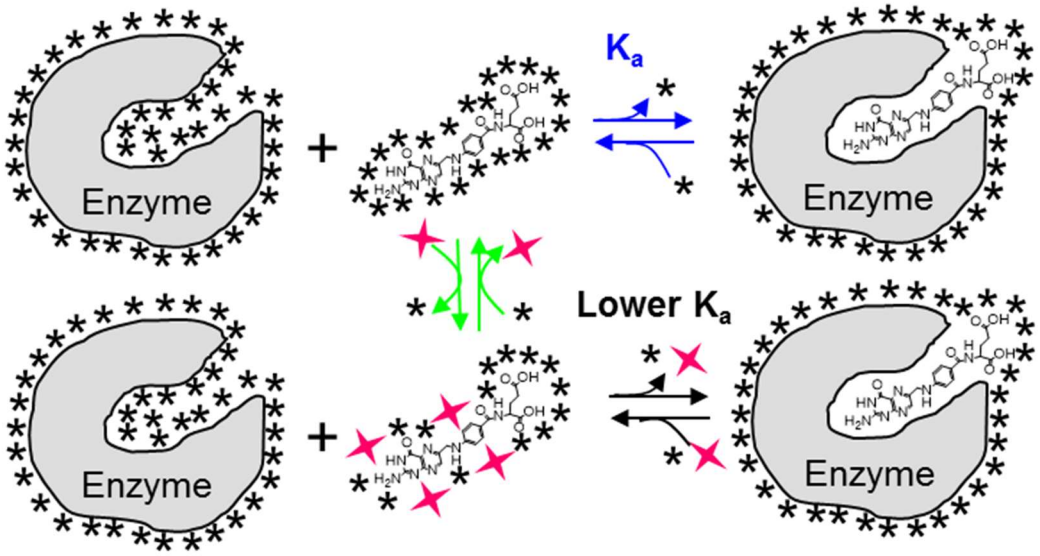


Figure 2.1 - A model showing the preferential interaction of osmolytes with free DHF. Removal of water (★) and/or osmolytes (✦) from the solvation shell of DHF is required for the ligand to bind to DHFR. If the DHF-osmolyte association is stronger than the DHF-water interaction, the binding equilibrium is shifted to the left, favoring the unbound state. This results in a decreased binding affinity for DHF to the DHFR. This model does not exclude interactions between osmolytes and the protein.

the osmolytes bind DHF, albeit weakly. If osmolytes are bound and more difficult to release than water, then weaker binding of substrate to DHFR results. In this scenario, the osmolytes shift the binding equilibrium towards the free DHF and enzyme species and inhibit complex formation.

In the next step, we question the prevalence of this phenomenon by investigating if other DHFRs continue to show the same behavior. Other enzymes capable of serving as DHFRs have been identified in organisms lacking chromosomal DHFR, also known as FolA.[13-15] For example, while pteridine reductase (PTR1) from *Leishmania major* normally reduces biopterin and dihydrobiopterin, it can also reduce DHF. The homologous gene in *E. coli* is *ydgB* (renamed *folM*).

The presence of FolM allows *E. coli* to grow even when the chromosomal DHFR gene has been deleted as a double *FolA* (encoding EcDHFR) plus *FolM* deletion in *E. coli* is synthetic lethal.[16] PTR1 and FolM are short chain dehydrogenases/reductases which utilize an entirely different structure and active site residues (catalytic triad of K198-Y194-D181 in PTR1).[17, 18] Most recently, FolM has been proposed to be a dihydromonapterin reductase where this substrate has a pteridine ring with a –CHOH-CHOH-CH_2OH tail.[19]

A summary of available information describing EcDHFR, R67 DHFR, PTR1 and FolM is given in Table 1.1 (Part 1).[20] Figure 2.2 compares the crystal structures for EcDHFR, R67 DHFR and PTR1. While a structure and more kinetic information are available for PTR1, we chose to work with the FolM protein from *E. coli* as PTR1 shows substrate inhibition,[21] which could complicate the analysis of osmolyte effects on DHF binding.

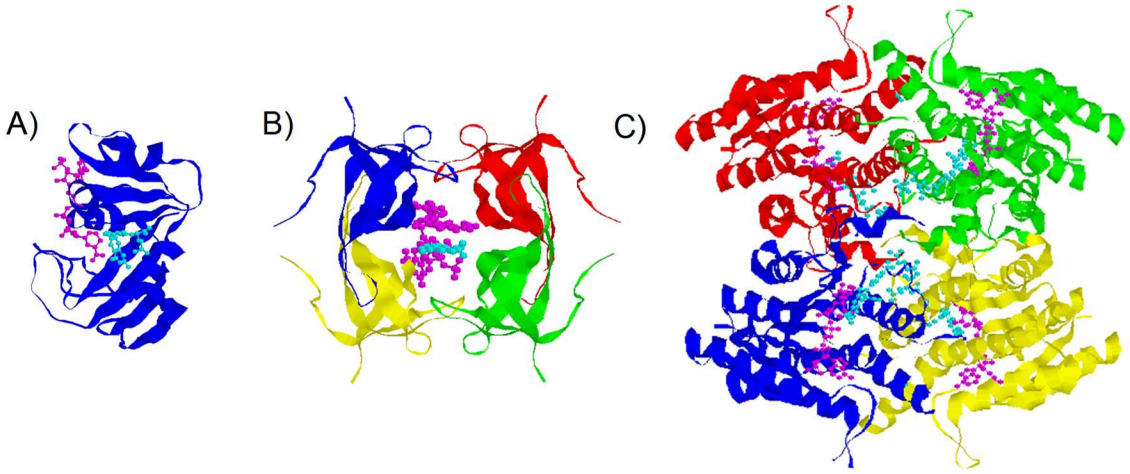


Figure 2.2 - Structures of the various DHFRs. Panel A shows the *E. coli* chromosomal DHFR structure (PDB code 1RA2).[22] Bound $NADP^+$ is shown in magenta and bound folate in cyan. Panel B gives the R67 DHFR structure (PDB code 1VIF). Each different color corresponds to a different monomer. The central doughnut hole is the active site. Bound $NADP^+$ and DHF are colored magenta and cyan, respectively.[4] Panel C provides the *L. major* PTR1 structure (PDB code 2BFA).[18] Each different monomer is colored differently. Bound NADPH and CB3717 are shown in magenta and cyan.

2.3 Materials and Methods

2.3.A. Protein Expression

The FolM gene from *E. coli* cloned into pET21B was a generous gift from Dr. Andrew Hanson, University of Florida, Gainesville.[19] This FolM construct carries an N-terminal His tag (MGHHHHHHH-), and expression is controlled by a lac promoter.[13] The plasmid was transformed into Rosetta 2 *E. coli* cells (EMD Millipore). For protein expression, cells were grown at 37 °C in TB media containing 100 μg ampicillin/ml and 30 μg chloramphenicol/ml. When the optical density reached 0.6 at 550 nm, IPTG was added to give a final concentration of 1 mM. Cells were grown for an additional 5 hours, centrifuged and frozen. For lysis, cells were resuspended in 50 mM sodium phosphate buffer (pH 8.0) containing 300 mM NaCl plus 20% (v/v) glycerol, and sonicated. Purification entailed loading and elution from a Ni^{2+} nitrilotriacetic acid agarose column (Qiagen). The protein was eluted in the same buffer using a gradient from 10-250 mM imidazole. SDS PAGE analysis showed a single band, and fractions were flash frozen in liquid N_2 and stored at -80°C. An Econo-Pac 10DG column (BioRad) was used to exchange buffers upon defrosting. Protein concentrations were determined using a bicinchoninic acid (BCA, Pierce) assay.

2.3.B. Steady-State Kinetics

Steady-state kinetic data were obtained at 30 ºC in MTA polybuffer at pH 6.0 using a Perkin-Elmer λ35 spectrophotometer as described previously.[23] MTA buffer consists of 50 mM MES plus 100 mM Tris plus 50 mM acetic acid; it maintains a constant ionic strength (μ=0.1 M) from pH 4.5-9.5.[24] Protein concentrations in the assay were 95-280

nM. To remove a lag, the enzyme was preincubated with NADPH and the reaction initiated by DHF addition. DHF K_m values were measured in the presence of saturating NADPH (32-76 μM). Initial rates were fit to the Michaelis-Menten equation in SigmaPlot. DHF was prepared by reduction of folate as per Blakley.[25] NADPH was purchased from Alexis Biochemicals. Concentrations of DHF and NADPH were measured using their respective extinction coefficients at 340 nm, 7.75×10^3 $M^{-1}cm^{-1}$ and 6.23×10^3 $M^{-1}cm^{-1}$.[26] The extinction coefficient for the DHFR reaction is 12.3×10^3 $M^{-1}cm^{-1}$.[27]

2.3.C. Isothermal Titration Calorimetry (ITC)

Affinities, stoichiometries and enthalpies of binding were determined as previously described.[3] At least two replicate titrations were performed using a VP-ITC microcalorimeter from MicroCal. 240 seconds separated each injection, allowing for baseline equilibration. FolM concentrations ranged from 8.5-15 μM in MTA buffer, pH 6. For titrations with osmolyte present, MTA buffer plus osmolyte was used in the reference cell. The "c value" (= $[P_{total}] / K_d$) ranged from 1-8, within the suggested values of 1-1000.[28]

Origin v7 software was initially used to analyze the ITC data. The data were then exported into SEDPHAT; this program allows global fitting of replicate data sets.[29] A single sites model ($A + B \leftrightarrow AB$) was used for the fitting process and errors were calculated using the Monte Carlo for non-linear regression option. For some experiments, baseline slopes and noise hindered integration of the ITC data by Origin. Therefore, the automated peak analysis program NITPIC[30] was used to integrate those data files to obtain the heat for each injection before further analysis with SEDPHAT.

2.3.D. Ultracentrifugation

Sedimentation velocity experiments were performed using absorbance optics in a Beckman Optima XL-I ultracentrifuge. The FolM sample was exchanged into MTA buffer, pH 6 and 15.7 μM FolM monomer was used in the experiment. Sedimentation velocity analysis was carried out at 50,000 rpm and 25 ºC using an An50 Ti eight-hole rotor. Sedimentation velocity analysis was performed by direct boundary modeling using the Lamm equation and the SEDFIT program (see www.analyticalultracentrifugation.com). [31] Partial specific volume (=0.7269) and buffer viscosity values were determined using the SEDNTERP program(see www.jphilo.mailway.com/download.htm). [32] The density of the buffer was determined using an Anton Paar DMA 35 vibrating tube density meter.

2.3.E. Water Activity Measurements

Solution osmolality was measured using a Wescor 5500 vapor pressure osmometer. This value was converted into water activity according to the equation:

$$a_{H2O} = e^{-0.018 \times \text{osmolality}} \qquad \text{Eq (2.1)}$$

where a_{H2O} is the water activity.[33]

2.3.F. Differential Scanning Calorimetry

Thermal unfolding of FolM was monitored between 25 and 95 ºC using a Microcal VP differential scanning microcalorimeter (DSC). The instrument was operated using the data acquisition and analysis program (Origin 7.0) supplied by the manufacturer. 9-11 μM FolM (monomer concentration) samples were prepared in MTA polybuffer, pH 6, plus or

minus osmolytes. Scan rates were 1 ºC per minute and 1.5 ºC per minute. Scans were repeated three times.

2.3.G. Circular Dichroism

CD was used to monitor the effect of cosolvents on the secondary structure of FolM using an AVIV Model 202 instrument. Briefly, at least 5 scans were accumulated on samples containing 9 μM protein in 50 mM phosphate buffer with 100 mM NaCl pH 6.0 plus or minus osmolyte using 0.5 nm steps and a two second integration. An average spectrum was calculated. The CD data were normalized as the mean residue ellipticity by using 110 g/mol as the mean residue molecular weight.[23]

2.3.H. Fluorescence Quenching

Binding of methotrexate (MTX, from Sigma-Aldrich) to 2 μM FolM was monitored in MTA buffer (pH 6) using tryptophan fluorescence as per Zhuang et al.[34] MTX concentrations were determined at pH 13.0 using an extinction coefficient of 22,000 M^{-1} cm^{-1} at 302 nm.[35] Spectra were collected on a Perkin Elmer LS55 fluorimeter. The sample was excited at 295 nm and emission spectra were recorded from 315 to 450 nm. Data were fit to Eq (2.2):

$$Fl = F_o - 0.5F_o\left[P_{tot} + K_d + L_{tot} - \sqrt{(P_{tot} + K_d + L_{tot})^2 - 4P_{tot}L_{tot}}\right] \quad \text{Eq (2.2)}$$

where Fl is the observed fluorescence, L_{tot} is the total ligand concentration, and P_{tot}, K_d and F_o are variables describing the number of enzyme binding sites, dissociation constant and fluorescence yield per unit concentration of enzyme, respectively.

2.3.I. Homology Modeling

A homology model was created for FolM in MOE v.2010 (Chemical Computing Group, Montreal, Canada) using the 2BFA PDB structure of *L. major* PTR1 as a template.[18] Primary sequences for FolM and PTR1 were aligned and FolM was modeled as a tetramer. Modeling was performed using the ligands bound in the 2BFA structure (NADPH and 10-propargyl-5,8-dideazafolate) as additional templates for the active site of the protein. The homology model with the fewest deviations in the φ and ψ angles from Ramachandran values is described below.

2.4 Results

2.4.A. Ultracentrifugation

A previous comparison of PTR1 and FolM primary sequences[13] using the ClustalW program yields a score of 22, where the score describes "the number of identities between the two sequences, divided by the length of the alignment, and represented as a percentage."[36] To determine if any homology extends past the primary sequence, the oligomerization state of FolM was assessed via sedimentation velocity analysis. The monomer mass of FolM with an N-terminal Histag is 27,496.5 daltons according to Expasy ProtParam[37] calculations. Figure 2.3 shows the ultracentrifugation data which indicate that FolM has an s value of 5.85 S, corresponding to a mass of 106 kilodaltons. This mass is consistent with FolM being a tetramer, suggesting a structural homology with PTR1 beyond that of the primary sequence as PTR1 is also a tetramer.

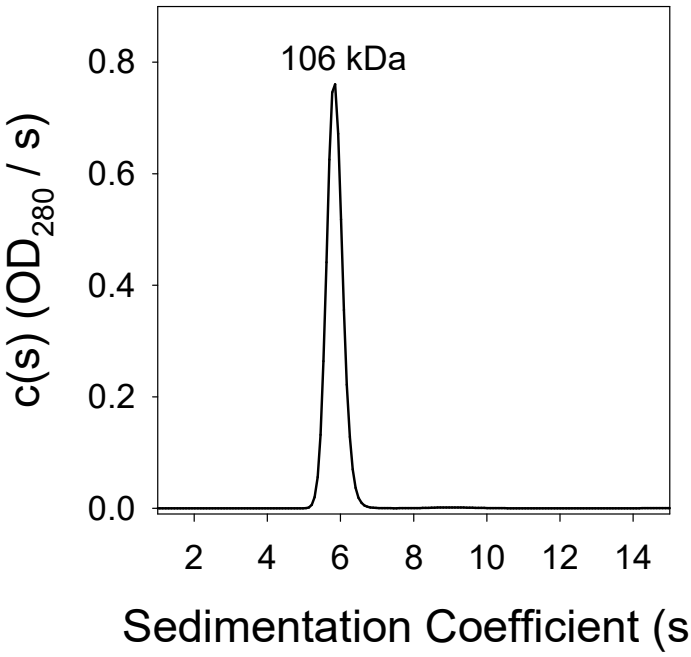


Figure 2.3 - Sedimentation velocity data for FolM (15.7 μM) in MTA buffer, pH 6.0. The ultracentrifugation of FolM was monitored by the change in absorbance at 280 nm. The data were fit to the sedimentation distribution constant, c(s), model in SEDFIT. A mass of 106 kDa was obtained, which indicates FolM forms a tetramer (monomer mass of 26.3 kDa). No evidence of other oligomerization states was noted.

2.4.B. Homology modelling

Using this information, a homology model for FolM was constructed from the 2BFA PDB structure of tetrameric PTR1 (with bound NADPH and 10-propargyl-5,8-dideazafolate, e.g. CB3717).[17, 18] Figure 2.4 shows this predicted structure. The predicted catalytic triad residues, D139, Y152 and K156, remain in the active site, suggesting a reasonable model. Also, R17, S87 and W89 occur in the active site of the FolM homology model; the comparable residues in PTR1 interact with the dihydrobiopterin substrate in the 1E92 structure.

2.4.C. Stability

As our initial forays with FolM found it tended to precipitate in low ionic strength phosphate buffer and at higher pH values, we investigated its stability using differential scanning calorimetry (DSC). Figure 2.5.A shows the resulting thermogram. Use of a two-state transition to fit the DSC thermogram did not accurately encompass the entire transition, so a three-state transition was used for fitting, yielding two T_ms of 60.5 ± 0.8°C and 63.0 ± 0.3 °C (1.5 °C/min scan rate). The calorimetric enthalpy (ΔH_d) for the first transition was 42.4 ± 3.4 kcal / mol, while the ΔH_d for the second transition was 32.4 ± 3.5 kcal / mol. Thermal denaturation of FolM results in protein precipitation; thus, the scans are not reversible. Irreversible unfolding prohibited obtaining further thermodynamic information, such as the van't Hoff enthalpy.[38, 39] We also considered whether FolM might display kinetic stability effects where a high free-energy barrier between the native and unfolded state keeps either the unfolded or an intermediate state from aggregating.[40, 41] This possibility can be tested by decreasing the scan rate.[40] As shown in Figure 2.5.B, a slower scan rate of 1°C/min results in lower T_ms for both transitions (58.7 ± 6.2 °C

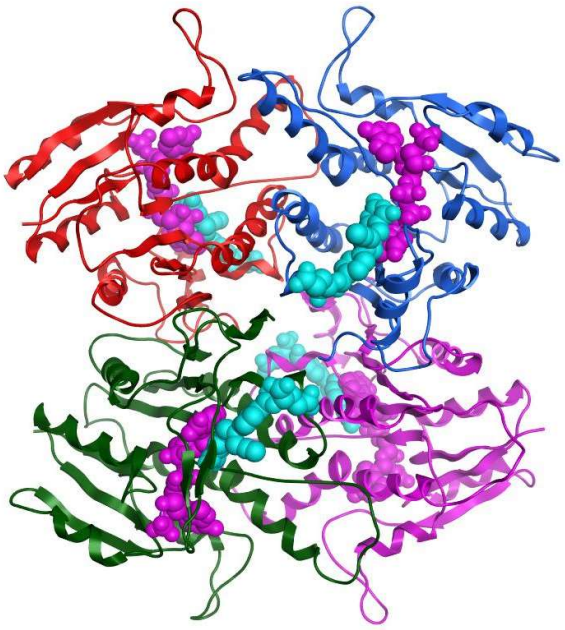
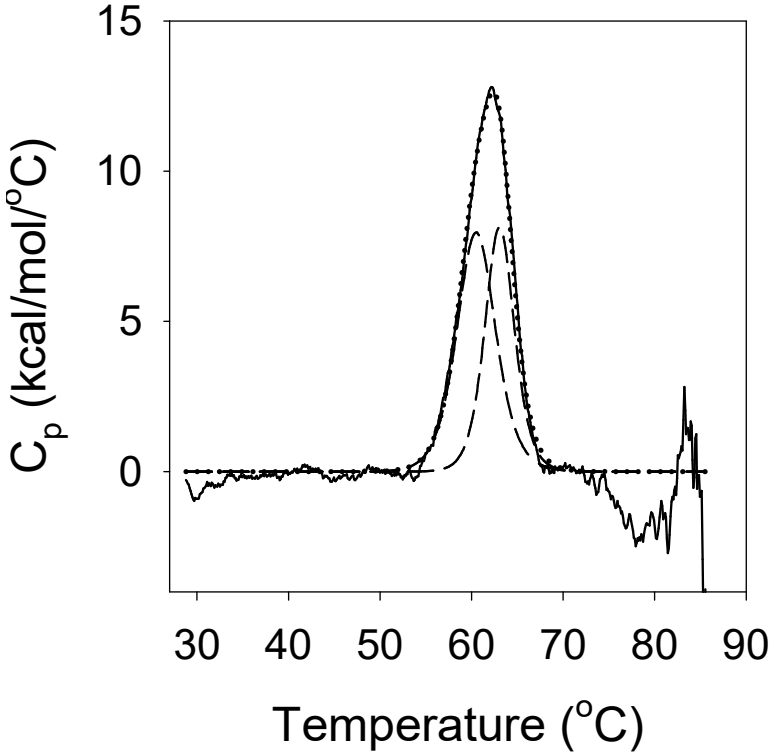
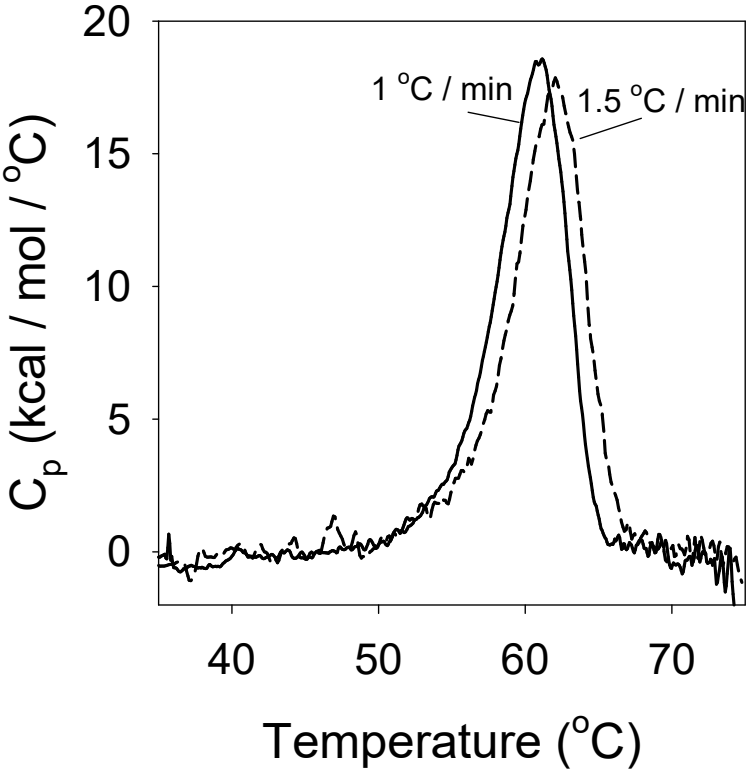


Figure 2.4 - A FolM homology model. The shown model constructed by MOE v.2010 (Chemical Computing Group, Montreal, Canada) based on the 2BFA crystal structure of *L. major* PTR1.[18] Each different monomer is shown in a different color. The putative binding sites for NADPH (CPK, magenta) and 10-propargyl-5,8-dideazafolate (CB3717, cyan) used in the homology modeling process are shown.

Figure 2.5 - Thermal denaturation of FolM by DSC. Panel A shows the fit of DSC data for FolM (11 μM) in MTA, pH 6.0 buffer at a scan rate of 1.5 °C / min. The solid black line is the DSC thermogram. The dashed black lines are fits of each of the transitions in the thermogram. The sum of the individual transition fits is shown as a dotted line which overlays the thermogram data. Fits of the data yield a T_{m1} of 60.5 ± 0.8°C and a T_{m2} of 63.0 ± 0.3 °C. Panel B shows DSC thermograms for 11 μM of the FolM protomer scanned at 0.5 °C / min (solid line) and 1.5 °C / min (dashed line). The data were fit with one-transition, cooperative unfolding models yielding T_ms of 58.7 ± 6.2 °C and 61.4 ± 1.0 °C for the 1 °C / min and 60.1 ± 0.6 °C and 62.7 ± 0.3 °C for 1.5 °C / min scan rates. Thermal denaturation of FolM was not reversible as no transition was noted for a second scan of the sample.


15
10
5
0
C_p (kcal/mol/^{o}C)
30
40
50
60
70
80
90
Temperature (^{o}C)


20
15
10
5
0
C_p (kcal / mol / ^{o}C)
1 ^{o}C / min
1.5 ^{o}C / min
40
50
60
70
Temperature (^{o}C)

and 61.4 ± 1.0 °C), suggesting FolM displays kinetic stability effects.[41] Additional DSC scans were performed for a range of FolM concentrations (4 to 11 μM) at a scan rate of 1.5 °C/min (data not shown). Both T_ms and the ΔH_d for the first transition were constant as the protein concentration changed. The ΔH_d for the second transition increased approximately two-fold as the FolM concentration was changed from 4 μM to 11 μM. The protein concentration dependence of the FolM thermogram suggests that a conformational change best describes the event accompanying T_{m1}, while T_{m2} is related to tetramer dissociation. Alternatively, increased aggregation of the denatured protein at higher concentrations of FolM could lead to an artificial change in T_{m2}.

2.4.D. Steady State Kinetics

Previous characterization of FolM found higher activity at lower pH.[13, 19] To balance increased FolM activity with DHF solubility and NADPH stability issues, we also performed our assays at pH 6. Our steady state kinetic values in MTA buffer are listed in Table 1.1. Both Giladi et al.[13] and Nare et al.[21] report V_{max} values of 0.083 μmol min^{-1} mg^{-1} for FolM and 0.38 μmol/min/mg for PTR1. While we report k_{cat} (per FolM monomer) in Table 1.1, our V_{max} value (0.52 μmol/min/mg) is slightly higher than for PTR1 (0.38 μmol/min/mg),[21] and both are higher than the value for FolM from Giladi et al. (0.083 μmol/min/mg).[13] Our DHF K_m value is also ~2 fold smaller than the value of Giladi et al.[13] These differences may arise due to variations in buffer and/or protein stability as Giladi et al.[13] used 0.1 M phosphate buffer and Pribat (personal communication)[19] found addition of 100 mM NaCl helped minimize protein precipitation in low ionic strength phosphate buffer. Additionally, the crystal structure of pteridine reductase (PTR1) from *Leishmania donovani* shows sulfate occupying the

phosphate binding site of the adenine-ribose phosphate for the NADPH cofactor.[42] The use of phosphate buffer by Giladi et al.[13] may provide some level of competition for cofactor binding.

2.4.E. Osmolyte Effects

To determine if osmolyte addition alters the secondary structure of FolM, CD scans were obtained in several osmolytes that do not absorb in the far UV range. Figure 2.6 shows only minor effects, suggesting osmolyte addition does not alter protein structure drastically.

Cofactor binding was also monitored by ITC. A representative thermogram is shown in Figure 2.7. Thermodynamic values for the binding of cofactor are given in Table 2.1. A linear trend between ln $K_{a\ (NADPH)}$ and ln water activity was noted for all the osmolytes used to examine cofactor binding (Figure 2.8). The slope of this plot indicates the preferential interaction or exclusion of the osmolytes involved in NADPH binding to the enzyme. The slopes associated with the individual osmolytes for this type of plot are given in Table 2.2. As betaine has been proposed to be the most excluded osmolyte from protein surfaces,[43] it is not surprising that addition of betaine decreased the cofactor K_d (or increased the cofactor K_a). Glycerol addition also resulted in tighter cofactor binding. In contrast, the change in the $K_{d\ (NADPH)}$ in the presence of DMSO (10%, 15% and 20%) was within error of the value for buffer alone (Figure 2.9.A). However, cofactor binding was weakened in the presence of ethylene glycol or PEG 400 (see Figure 2.9.B for a plot of the PEG400 data). This pattern of no effect or weakened binding of cofactor in the presence of osmolytes was not noted for R67 DHFR,[2] and only sucrose decreased the binding of $NADP^+$ to EcDHFR,[9] suggesting a more complex behavior associated with

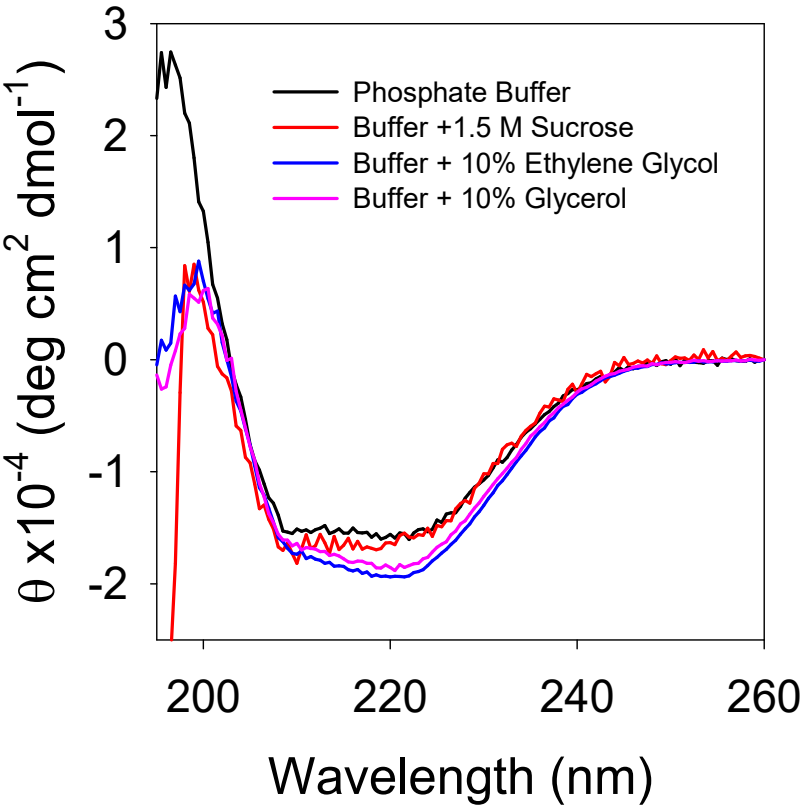


Figure 2.6 - Effect of osmolytes on the secondary structure of FolM. CD spectra were recorded for FolM (9 μM) in 50 mM Na_2HPO_4, 100 mM NaCl, pH 8.0 buffer (black line), or buffer plus 1.5 M sucrose (red line), buffer plus 10% ethylene glycol (blue line), or buffer plus 10% glycerol (magenta line).

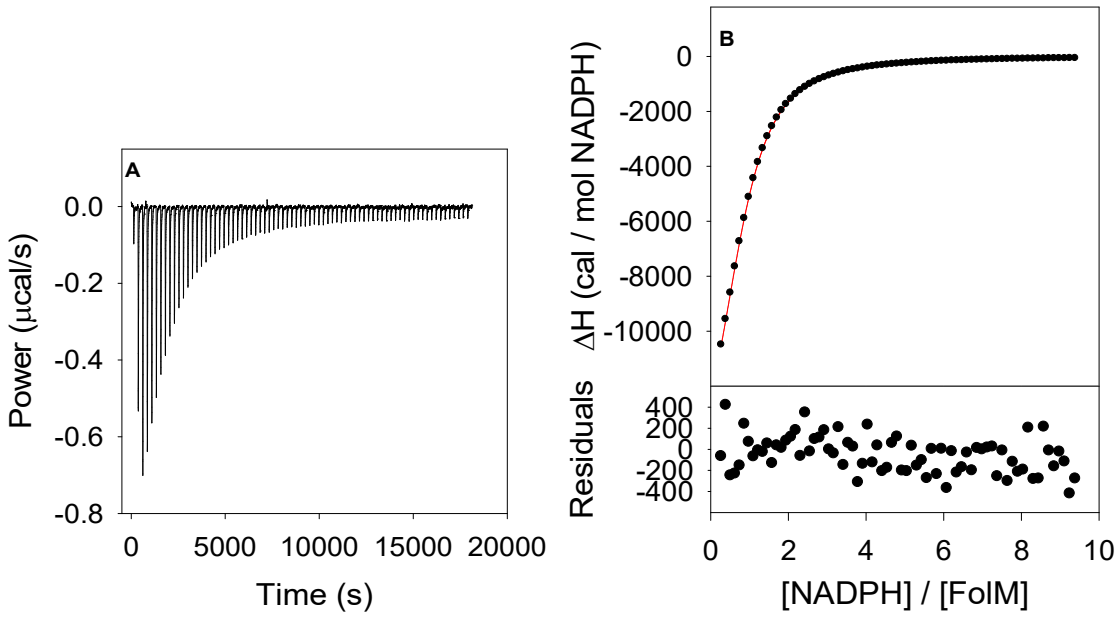


Figure 2.7- Representative ITC data. Panel A represents a raw ITC thermogram for the titration of NADPH (553 µM) into FolM (9.65 µM) in MTA, pH 6.0 at 25 °C. In panel B, the NADPH binary complex isotherm was fit in SEDPHAT using a single-site model. The fit yielded a binding affinity (K_d) of 3.25 µM with a stoichiometry (n) of 1.16 and a ΔH of -15.6 kcal/mol. Residuals for the fit are shown underneath the plot.

Table 2.1 - Binding of NADPH to FolM in the presence of various osmolytes in MTA buffer, pH 6 at 25 °C. The data are from global fits of at least two ITC data sets using SEDPHAT and a single sites model (A+B ↔ AB).[29] The ΔG values were calculated from the equation $\Delta G = -RT \ln K_d$ and TΔS values from $\Delta G = \Delta H - T\Delta S$. Data for DMSO were obtained with a different prep of FolM.

Buffer and/or osmolyte addition	K_d (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)	n	Osmolality (Osm)
MTA	3.86 ± 0.29	-7.38	-16.7 ± 0.68	-9.35	0.84 ± 0.01	0.22
MTA + 10 % betaine	1.94 ± 0.11	-7.79	-21.8 ± 0.3	-14.0	0.88 ± 0.01	1.31
MTA + 20 % betaine	1.61 ± 0.17	-7.90	-21.3 ± 0.7	-13.4	0.89 ± 0.01	2.17
MTA + 10 % glycerol	3.62 ± 0.46	-7.42	-18.2 ± 1.4	-10.8	0.69 ± 0.02	1.56
MTA + 20 % glycerol	3.11 ± 0.28	-7.51	-21.4 ± 1.1	-13.8	0.65 ± 0.02	2.69
MTA + 10 % ethylene glycol	5.13 ± 0.84	-7.22	-20.2 ± 1.8	-12.9	0.82 ± 0.05	1.90
MTA + 20 % ethylene glycol	4.79 ± 0. 80	-7.25	-19.2 ± 2.0	-12.0	0.73 ± 0.04	3.55
MTA buffer control for PEG400	2.69 ± 0.13	-7.60	-19.4 ± 1.3	-11.9	0.97 ± 0.01	0.22
MTA + 10 % PEG400	3.74 ± 0.28	-7.40	-19.8 ± 0.7	-12.4	0.94 ± 0.01	0.57
MTA + 20 % PEG400	5.05 ± 0.58	-7.22	-17.9 ± 1.5	-10.7	0.80 ± 0.03	1.05
MTA buffer control for DMSO	2.99 ± 0.54	-7.53	-15.1 ± 0.1	-7.57	0.98 ± 0.04	0.22
MTA + 10 % DMSO	2.91 ± 0.48	-7.54	-19.7 ± 1.3	-12.2	0.91 ± 0.02	1.56
MTA + 20 % DMSO	3.57 ± 0.64	-7.42	-21.1 ± 4.1	-13.6	0.88 ± 0.06	2.68

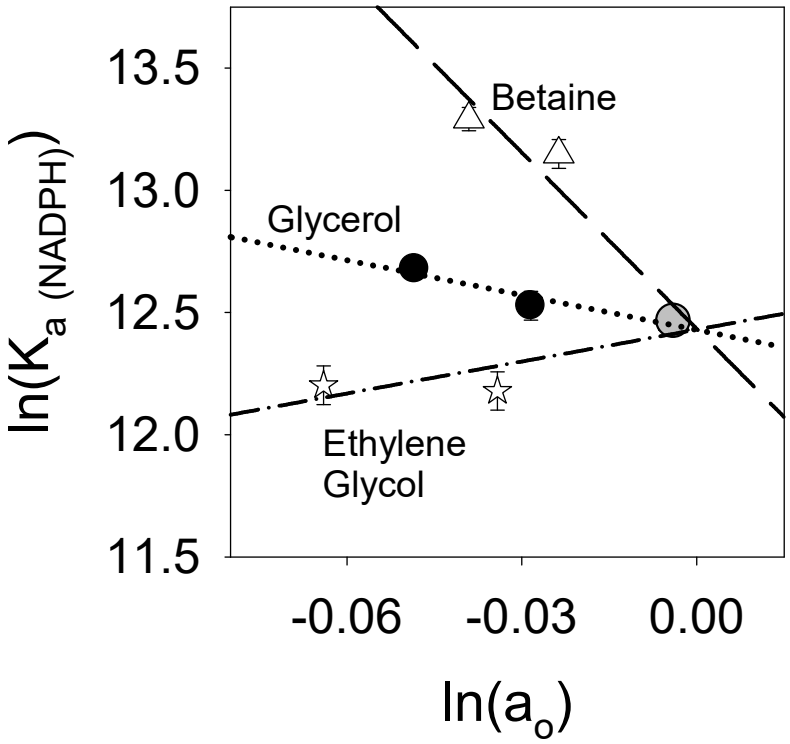


Figure 2.8 - Effect of osmolytes on NADPH binding to FolM in MTA, pH 6.0. Binding was measured using ITC. Data in buffer are given by ● points, buffer plus betaine (△ point, long dash line), glycerol (●, dotted line), or ethylene glycol (☆,dot-dash line). Error bars are shown and in some cases are smaller than the data symbols.

Table 2.2 - A comparison of the slopes for the ln (K_a) versus ln (a_o) plots describing cofactor binding to FolM, R67 DHFR and EcDHFR. A negative slope is consistent with release of water upon ligand binding, while a positive slope describes preferential binding effects that shift the equilibrium towards the unbound state.

Osmolyte	$\frac{\partial \ln(K_a)}{\partial \ln(a_o)}$ **for NADPH binding to FolM**	$\frac{\partial \ln(K_a)}{\partial \ln(a_o)}$ **for NADPH binding to R67 DHFR[a]**	$\frac{\partial \ln(K_a)}{\partial \ln(a_o)}$ **for $NADP^+$ binding to EcDHFR•DHF[b]**
Betaine	-24 ± 7	-38 ± 6	-14 ± 5
DMSO	5 ± 4	-38 ± 6	-24 ± 3
Sucrose	ND[c]	-38 ± 6	-5 ± 4
Ethylene Glycol	4 ± 3	-38 ± 6	-10 ± 3
Glycerol	-5 ± 1	ND	-5 ± 2
PEG400	38 ± 12	-38 ± 6	-47 ± 8

[a] Data for the non-homologous R67 DHFR were previously measured by steady state kinetics ($k_{cat}/K_{m\,(DHF)}$) (Reference [2]). ITC was also used to monitor ligand binding in R67 DHFR for a subset of the osmolytes (Reference [2]).
[b] Data for EcDHFR were previously measured by ITC (Reference [9]).
[c] Data not determined.

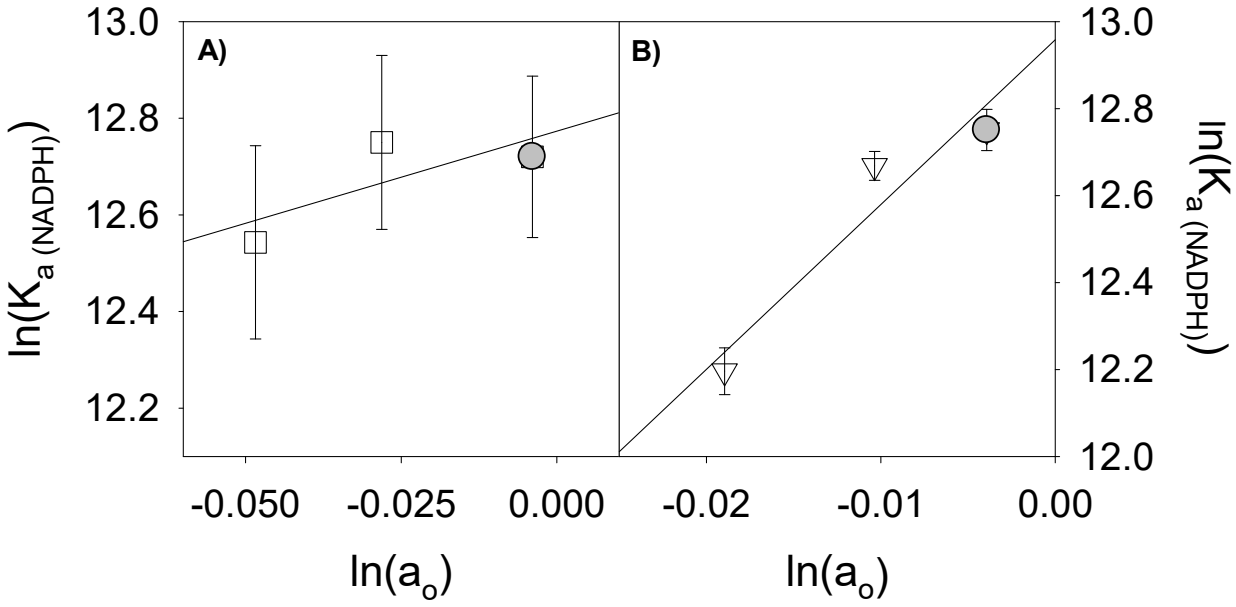


Figure 2.9 - Effect of DMSO and PEG400 on NADPH binding to FolM. Data from ITC experiments done with DMSO (panel A) and PEG (panel B) in MTA buffer, pH 6.0, 25 °C are shown separately from Figure 2.5 as they are from two different protein preparations and the value for the MTA buffer control (●) is 1.3 fold different for the DMSO data (□) and 1.4 fold different for the PEG400 data (▽).

FolM.

The thermal stability of FolM was studied to understand the weakening of cofactor binding by osmolytes. The effects of betaine, DMSO, and ethylene glycol on the thermal stability of FolM were examined by DSC (Figure 2.10.A). Betaine increased the temperature at which FolM melted and broadened the thermogram, while DMSO decreased the melting temperature and narrowed the thermogram. The denaturation peaks still fit to two transitions. Increases of 8.8 and 9.2 °C for T_{m1} and T_{m2} were noted in 20% betaine compared to FolM in buffer, see Figure 2.10.B. When DMSO was considered, addition of 20 % solute decreased T_{m1} by 8.5 °C and T_{m2} by 9.1 °C. Similarly, T_{m1} and T_{m2} decreased by 5.9 °C and 6.8 °C, respectively, in the presence of 20% ethylene glycol. The changes in T_ms were linear for betaine, ethylene glycol and DMSO osmolality, however opposite slopes were observed. For all three osmolytes, ΔH_d (calorimetric enthalpy of denaturation) was unchanged for the first and second thermal transitions. These results parallel other reports in the literature where osmolytes affect the hydration shell of the protein.[44-48] When osmolytes associate with the protein surface, they are destabilizing. When osmolytes are excluded from the protein surface, they are stabilizing.

The effects of osmolytes on $k_{cat}/K_{m\ (DHF)}$ were investigated next. Three different osmolytes (betaine, DMSO and sucrose) were initially chosen as they previously had significant effects on DHF binding to R67 DHFR and EcDHFR. These osmolytes also have different characteristics and can be used to parse out effects on viscosity and/or solution dielectric. For example, while sucrose and betaine both affect water activity, they provide opposite effects on the dielectric constant of the solution.[1, 49, 50] If both compounds show similar results in osmolality plots, then effects on the dielectric constant

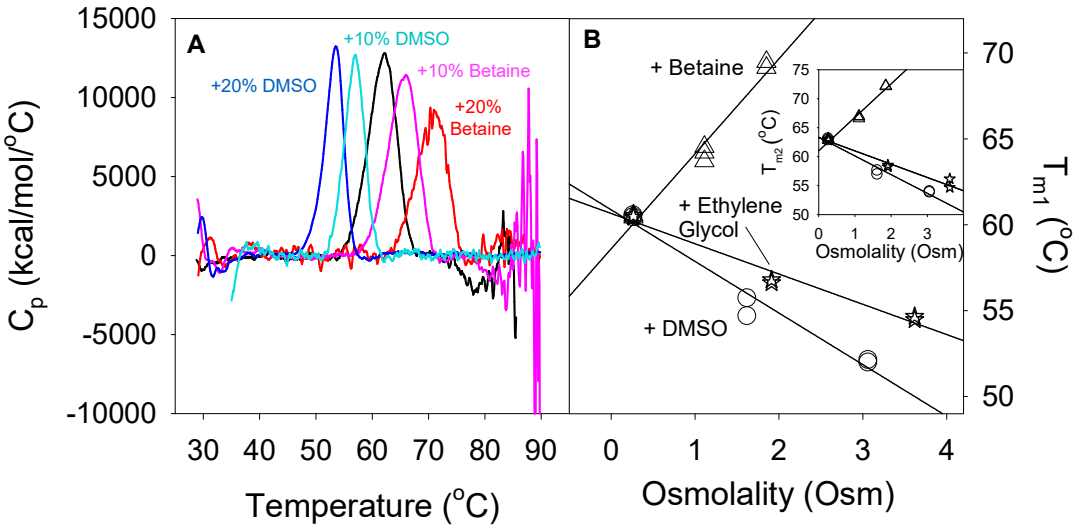


Figure 2.10 - The effect of osmolytes on the thermal stability of FolM in MTA, pH 6.0. Panel A shows the data obtained for FolM in buffer (black line), buffer + 10% betaine (magenta line), buffer + 20% betaine (red line), buffer + 10% DMSO (cyan line) and buffer plus 20% DMSO (blue line). All DSCs were performed with a scan rate of 1.5 °C / min. (B) Effect of betaine (△), DMSO (○) and ethylene glycol (☆) on the T_{m1} of FolM in MTA, pH 6.0. The effects of betaine, DMSO and ethylene glycol on T_{m2} are shown in the inset.

are not involved. Using steady state kinetics, we find osmolyte addition increases the $K_{m\ (DHF)}$ (Table 2.3). Figure 2.11 shows the linear relationships associated with plots of ln $k_{cat}/K_{m\ (DHF)}$ vs. ln water activity. As a precaution, we also plotted effects on solution viscosity or solution dielectric and overlapping data were not observed. These figures are presented as Figure 2.12, panels A and B. No effects on k_{cat} were noted except for ~1.5 fold increases in 20% glycerol, 20% ethylene glycol and DMSO. In general, the slopes for FolM compare with those previously determined for R67 DHFR and EcDHFR, providing further support for our model in Figure 2.1. The effects of larger molecular weight osmolytes on DHF binding to FolM were explored using polyethylene glycols (PEGs). Kinetic experiments using PEG400 and PEG3350 were performed with a separate FolM prep that yielded a 1.8x higher $k_{cat}/K_{m,(DHF)}$ value in MTA buffer. PEGs have a larger effect on $k_{cat}/K_{m\ (DHF)}$ compared to small molecule osmolytes.

As K_m can contain kinetic terms, K_d measurements would also be appropriate to test osmolyte effects on binding. ITC measurements of DHF binding to either apo FolM or the FolM-$NADP^+$ binary complex using ITC were unsuccessful due to a low signal and DHF degradation over the 2-3 hrs of the titration. Therefore, we turned to binding of the antifolate methotrexate (MTX). MTX is stable at pH 6 and provides a reasonable signal. It also provides a window into osmolyte effects on folate analogs. The titration of MTX into apo FolM gave no discernible heat signal. The lack of heat released or absorbed during the titration could be due to either no binding or no signal associated with binary complex formation. To differentiate between these scenarios, the quenching of tryptophan fluorescence upon MTX binding to FolM was measured. This titration was fit to Eq (2.2) as shown in Figure 2.13, yielding a $K_{d\ (MTX)}$ of 0.90 ± 0.24 μM with a stoichiometry of 0.84

Table 2.3 - Steady state kinetic parameters for FolM obtained at pH 6, 30 °C. The DHF concentration was varied in the presence of saturating NADPH concentrations and the data were fit to the Michaelis-Menten equation.

Buffer and/or osmolyte addition	k_{cat} (s^{-1})	$K_{m\,(DHF)}$ (μM)	Osmolality (Osm)
MTA buffer	0.24 ± 0.01	4.3 ± 0.6	0.256
MTA + 10% betaine	0.20 ± 0.01	7.0 ± 0.6	1.14
MTA + 20% betaine	0.23 ± 0.01	15.0 ± 0.9	2.10
MTA + 10% DMSO	0.30 ± 0.01	8.1 ± 0.6	1.62
MTA + 20% DMSO	0.32 ± 0.01	12.2 ± 1.4	3.10
MTA + 10% glycerol	0.23 ± 0.01	3.8 ± 0.3	1.56
MTA + 20% glycerol	0.28 ± 0.01	5.4 ± 0.4	2.92
MTA + 10% ethylene glycol	0.21 ± 0.01	5.5 ± 0.5	1.83
MTA + 20% ethylene glycol	0.32 ± 0.01	6.6 ± 0.3	3.46
MTA + 0.75M sucrose	0.21 ± 0.01	5.2 ± 0.8	1.31
MTA + 1.5M sucrose	0.21 ± 0.01	9.3 ± 1.4	2.40
MTA + 10% PEG400	0.31 ± 0.01	6.1 ± 0.7	0.575
MTA + 20% PEG400	0.22 ± 0.01	8.5 ± 0.8	1.05
MTA + 10% PEG3350	0.20 ± 0.01	3.8 ± 0.5	0.37
MTA + 20% PEG3350	0.21 ± 0.01	9.6 ± 1.0	0.54
MTA + 10% PEG8000	0.20 ± 0.01	12.7 ± 1.2	0.32
MTA buffer control used for PEG data	0.26 ± 0.01	2.6 ± 0.5	0.267

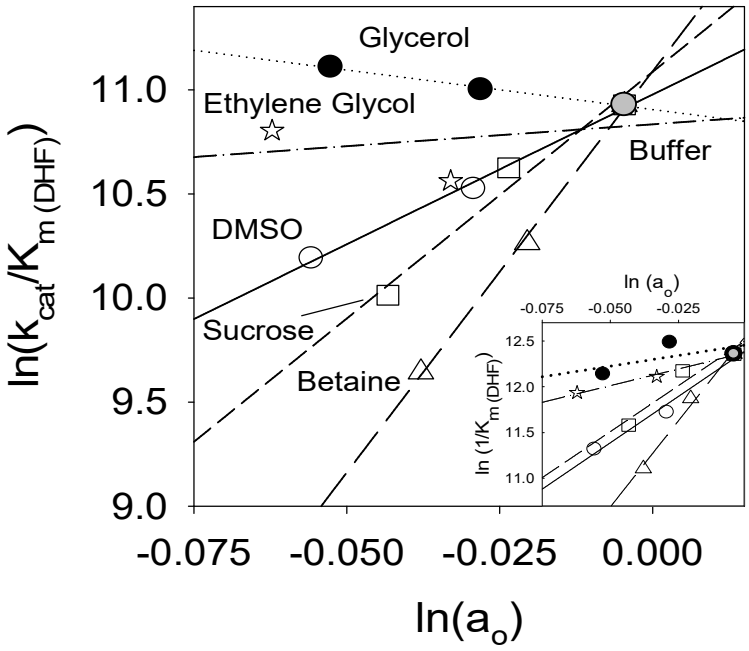


Figure 2.11 - Effect of osmolytes on the DHF-reduction activity of FolM in MTA buffer, pH 6.0. Activity studies were performed in the presence of buffer (●), betaine (△), DMSO (○), sucrose (□), glycerol (●) and ethylene glycol (☆). The change in $1/K_{m\ (DHF)}$ with water activity (a_o) is shown in the inset. Slopes of the plots are provided in Table 2.4.

Table 2.4 - A comparison of the slopes for the ln (K_a) versus ln (a_o) plots describing substrate binding to FolM, R67 DHFR and EcDHFR. A negative slope is consistent with release of water upon ligand binding, while a positive slope describes preferential binding effects that shift the equilibrium towards the unbound state. Slopes for ln (K_a) versus ln (a_o) plots describing methotrexate binding are also included in the last column.

Osmolyte	$\frac{\partial \ln(K_a)}{\partial \ln(a_o)}$ **for DHF binding to FolM•NADPH[a]**	$\frac{\partial \ln(K_a)}{\partial \ln(a_o)}$ **for DHF binding to R67 DHFR•NADP^{+}[b]**	$\frac{\partial \ln(K_a)}{\partial \ln(a_o)}$ **for DHF binding to EcDHFR•NADP^{+}[c]**	$\frac{\partial \ln(K_a)}{\partial \ln(a_o)}$ **for MTX binding to FolM•NADPH**
Betaine	39 ± 2	60 ± 13	34 ± 9	35 ± 10
DMSO	14 ± 1	41 ± 7	29 ± 1	ND
Sucrose	24 ± 4	40 ± 4	30 ± 2	ND
Ethylene Glycol	2 ± 6	25 ± 8	13 ± 1	11 ± 1
Glycerol	-4 ± 1	16 ± 3	18 ± 1	8 ± 1
PEG400	24 ± 4	78 ± 11	64 ± 5	ND[d]

[a] Data for FolM were measured using steady state kinetics ($k_{cat}/K_{m\ (DHF)}$).
[b] Data for the non-homologous R67 DHFR were previously measured by steady state kinetics ($k_{cat}/K_{m\ (DHF)}$) (ref [2]). ITC was also used to monitor ligand binding in R67 DHFR for a subset of the osmolytes (ref [2]).
[c] Data for EcDHFR were previously measured by ITC (ref [9]).
[d] Data not determined

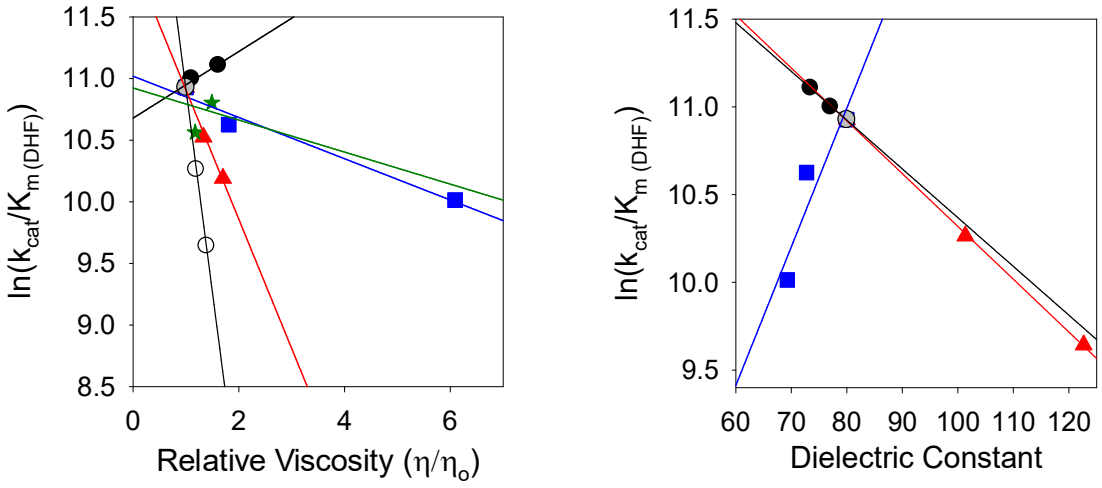


Figure 2.12 - Effects of Viscosity and Dielectric Constant on the change in ln(kcat/Km) for the reduction of DHF by FolM. Panel A shows plot of $\ln(k_{cat}/K_m)$ versus osmolyte solution viscosity. A lack of correlation between the FolM activity and the solution viscosity indicates the osmolyte effect on activity is not related to this parameter. Relative viscosities of the osmolyte solutions were taken from Chopra et al.[2] Plot of the change in $\ln(k_{cat}/K_{m\,(DHF)})$ for FolM versus the dielectric constant of the osmolyte solutions is shown in panel B. A lack of correlation between the FolM activity and the solution dielectric value indicates the osmolyte effect on activity is not related to the dielectric properties of the solution. Dielectric coefficients were calculated using a dielectric increment equation according to Edsall.[49] Activity studies were performed in the presence of buffer (●), betaine (▲), DMSO (○), sucrose (■), glycerol (●) and ethylene glycol (★).

± 0.20 MTX bound per FolM monomer. Previous fluorescence studies using apo PTR1 found folate binds with a K_d of 13 μM [51] and DHF binds with a K_d of 10 μM.[21] Product inhibition studies as well as the PTR1 crystal structure indicate an ordered mechanism with cofactor binding first, followed by substrate.[17, 21] For our ITC titration of MTX into FolM, the observation of no heat signal indicates entropy driven binding. As MTX is reasonably hydrophobic (logP = -2.1654), this may be due to desolvation effects.[52] MTX binding to the FolM-NADPH binary complex was also measured by ITC. A K_d of 3.68 ± 0.39 μM was obtained for MTX binding to the FolM-NADPH binary complex, which is close to the K_i of 5.9 μM determined by Giladi et al.[13] The effects of betaine, glycerol and ethylene glycol on MTX binding were also examined (Table 2.5). Compared to buffer alone, the $K_{d\,(MTX)}$ increased about 3-fold in 20% betaine. Likewise, the $K_{d\,(MTX)}$ increased 1.5-fold and 2-fold in 20% glycerol and 20% ethylene glycol, respectively. The change in ln $K_{a\,(MTX)}$ versus ln water activity of the osmolyte solutions was plotted (Figure 2.14). Slopes for the betaine, ethylene glycol and glycerol data in these plots were 35 ± 10, 11 ± 1 and 8 ± 1, respectively. To our knowledge, this is the first examination of osmolyte effects on MTX binding and the results parallel the FolM DHF binding effects as given in Table 2.4. We note similar positive slopes were previously obtained for DHF binding to the R67 DHFR-$NADP^+$ and EcDHFR-$NADP^+$ complexes in betaine and glycerol (Table 2.2).[2, 9]

To analyze the binding mechanism further,[53] a 1:1 mixture of NADPH and MTX was titrated into FolM. The data could be fit to a single binding site model. The enthalpy obtained was -22.1 ± 0.1 kcal/mol, which is close to the sum of the enthalpies for NADPH binding to apo FolM and MTX binding to NADPH-FolM binary complex (-25 kcal/mol).

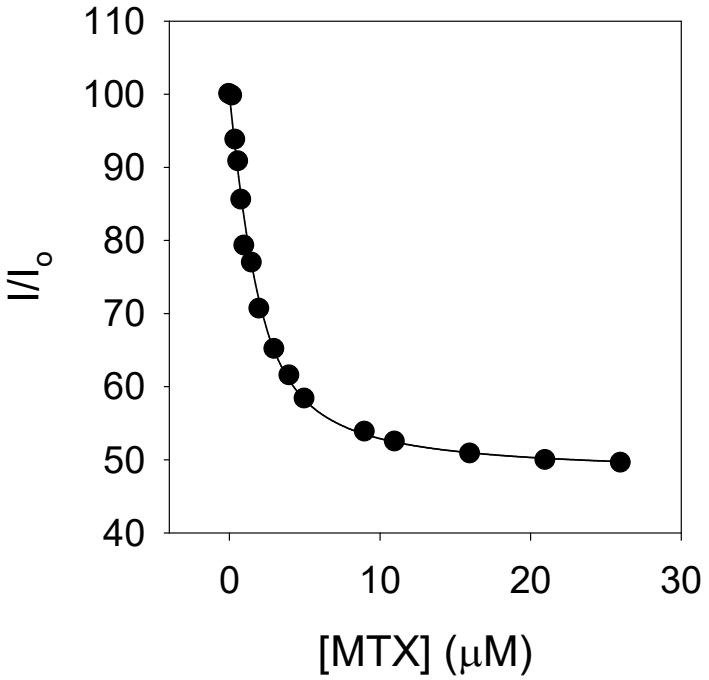


Figure 2.13 - Fluorescence quenching data for MTX binding to FolM (2 μM) in MTA, pH 6.0. Samples were excited at 295 nm and emission was measured at 350 nm. Data are plotted as I/I_o, where I is the corrected fluorescence intensity at a given concentration of MTX and I_o is the fluorescence intensity with no MTX present. A K_d of 0.90 ± 0.24 μM and a stoichiometry of 0.84 ± 0.20 MTX per FolM monomer were obtained from fitting the data.

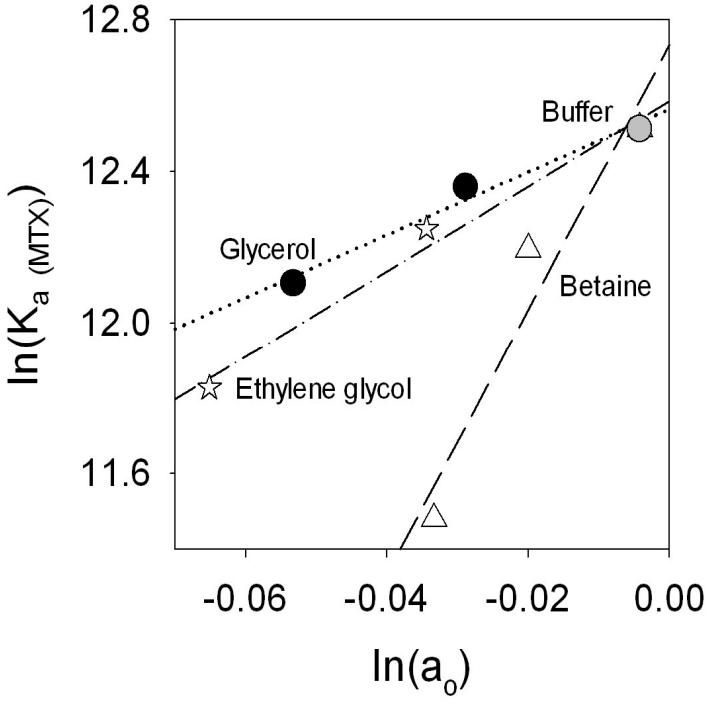


Figure 2.14 - Effect of osmolytes on MTX binding to the FolM-NADPH complex. ITC studies were performed in MTA, pH 6.0 at 25 ºC with no osmolyte (●), betaine (△), ethylene glycol (☆) and glycerol (●).

Table 2.5 - ITC measurements describing methotrexate binding to the FolM-NADPH binary complex in MTA buffer, pH 6.0 and 25 °C. At least two data sets were fit globally in SEDPHAT using a single-site binding model, A + B → AB.

Buffer and/or osmolyte addition	K_d (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)	n	Osmolality (Osm)
MTA	3.68 ± 0.39	-7.41	-8.73 ± 0.42	-1.32	0.85 ± 0.01	0.23
MTA + 10 % betaine	5.05 ± 0.70	-7.22	-6.86 ± 0.63	0.36	0.74 ± 0.04	1.11
MTA + 20 % betaine	10.3 ± 0.9	-6.80	-7.24 ± 0.49	-0.44	0.74 ± 0.04	1.85
MTA + 10 % glycerol	4.28 ± 0.29	-7.31	-7.16 ± 0.23	0.15	0.82 ± 0.01	1.60
MTA + 20 % glycerol	5.52 ± 0.70	-7.16	-6.46 ± 0.50	0.70	0.76 ± 0.03	2.96
MTA + 10 % ethylene glycol	4.78 ± 0.83	-7.26	-8.34 ± 0.82	-1.08	0.81 ± 0.04	1.91
MTA + 20 % ethylene glycol	7.31 ± 0.80	7.00	-9.35 ± 0.74	-2.35	0.85 ± 0.04	3.62

Similarly, the ΔG for the concurrent titration of both ligands was -15.5 ± 0.1 kcal/mol, while the sum of the ΔG values for the NADPH binary and MTX ternary titrations was -14.8 kcal/mol. The near equivalence of the ΔG and ΔH values for the titration of MTX+NADPH into FolM with the sum of the NADPH binary and MTX ternary experiments, coupled with no enthalpy associated with MTX binary binding, is consistent with an ordered binding mechanism where NADPH binds first, followed by MTX.[53] Two alternate possibilities are that MTX bind firsts and rearranges upon NADPH binding, or that the cofactor binding site is not totally occluded so that the cofactor can bind without release of MTX. Luba et al. find non-productive binding of DHF to PTR1, extending the similarities between FolM and PTR1.[54]

2.5 Discussion

One model of how enzymes work considers desolvation.[55-60] As cosolutes can compete with water to associate with molecular surfaces, we now expand the desolvation model of enzyme action to include removal of cosolutes such as osmolytes. If the DHF-osmolyte pairs are more difficult to break than the DHF-H_2O pairs (desolvation), then weaker binding of substrate to DHFR results. This is a solvent substitution scenario, and as shown in Figure 2.1, this situation shifts the binding equilibrium towards the free species.

Since our model posits the critical species for the osmolyte effects is the DHF substrate rather than the DHFR enzyme, we ask whether our observation can be extended to other DHFRs (and ultimately to other folate utilizing enzymes)? To test this model, we have previously used R67 DHFR and EcDHFR; we now add FolM to the list. We chose

FolM as a representative short chain dehydrogenase that can reduce DHF. While FolM has not been as well characterized as the canonical pteridine reductase, PTR1, it does not show substrate inhibition.

Previous characterization of the FolM protein has been minimal. In this study, we provide several additional details. Our ultracentrifugation studies find FolM is a tetramer. This observation is consistent with the structure of the canonical pteridine reductase, PTR1, which is also tetrameric.[17] A FolM homology model produced using the 2BFA PDB file for PTR1[18] leads to a FolM model with the proposed catalytic triad residues placed in the active site cavity. Nearby the FolM active site are R237 and R168 residues, which appear close enough to form ion pairs with the α- and γ-carboxylates of the Glu tail of DHF and provide tighter binding as compared to the proposed dihydromonapterin substrate.[19] (For comparison, the Nε atom of R287 is 3Å from the α-carboxylate of bound CB3717 (antifolate) in the active site of PTR1 in 2BFA.) Another pertinent observation is that DHF binding to FolM has a low enthalpic signal, making binding difficult to measure by ITC.

Binding of the antifolate methotrexate to FolM was also characterized. To garner additional information concerning the binding mechanism, SEDPHAT was used to globally fit both binary complex titrations as well as the ternary complex.[29] Figure 2.15 shows the global fit and Table 2.6 gives the fit values. In general, the global fit values are similar to those derived from the individual fits. Binding thermodynamics obtained from the global fit indicated that MTX binds to apo FolM, with a minimal ΔH of 0.81 kcal/mol. (Fit values are not appreciably altered when ΔH is set to 0 kcal/mol, though the errors are large.) The ΔG for binding (-8.12 kcal/mol, K_d of 1.09 μM) was surprisingly more negative

than for either the NADPH binary or MTX ternary experiments. This fit value concurs with the fluorescence quenching data for MTX binding to apo FolM, indicating tighter MTX binary than ternary binding. The difference between MTX binding to apo FolM and the FolM-NADPH binary complex suggests that MTX may bind somewhat differently depending upon whether NADPH is already bound, or not. This is not surprising since, in PTR1, NADPH forms part of the MTX binding site.[17, 18] While FolM has many other interesting features, we chose to pursue osmolyte effects.

2.5.A. Betaine Effects on Ligand Binding

As DHFR uses two substrates, an internal control monitors the effects of osmolytes on binding of the second ligand/cofactor. This allows us to determine if osmolytes associate with folate by observation of weaker folate binding in the presence of osmolytes coupled with tighter binding of the second ligand. In the presence of betaine, NADPH binds more tightly to FolM, while DHF binds more weakly. A decrease in $K_{d\ (NADPH)}$ with increasing betaine concentration is consistent with preferential exclusion of betaine from FolM and/or NADPH. Similar trends of betaine on cofactor and substrate binding were noted for both R67 DHFR and EcDHFR.[2, 9] These results continue to support our model in Figure 2.1 where betaine associates with free DHF. They also support more general models where betaine acts as a natural protective molecule in *E. coli* under times of osmotic stress and is the most excluded osmolyte from protein surfaces.[43, 61-64] A related observation is that betaine typically increases the stability of proteins by an increased hydration mechanism.[65, 66]

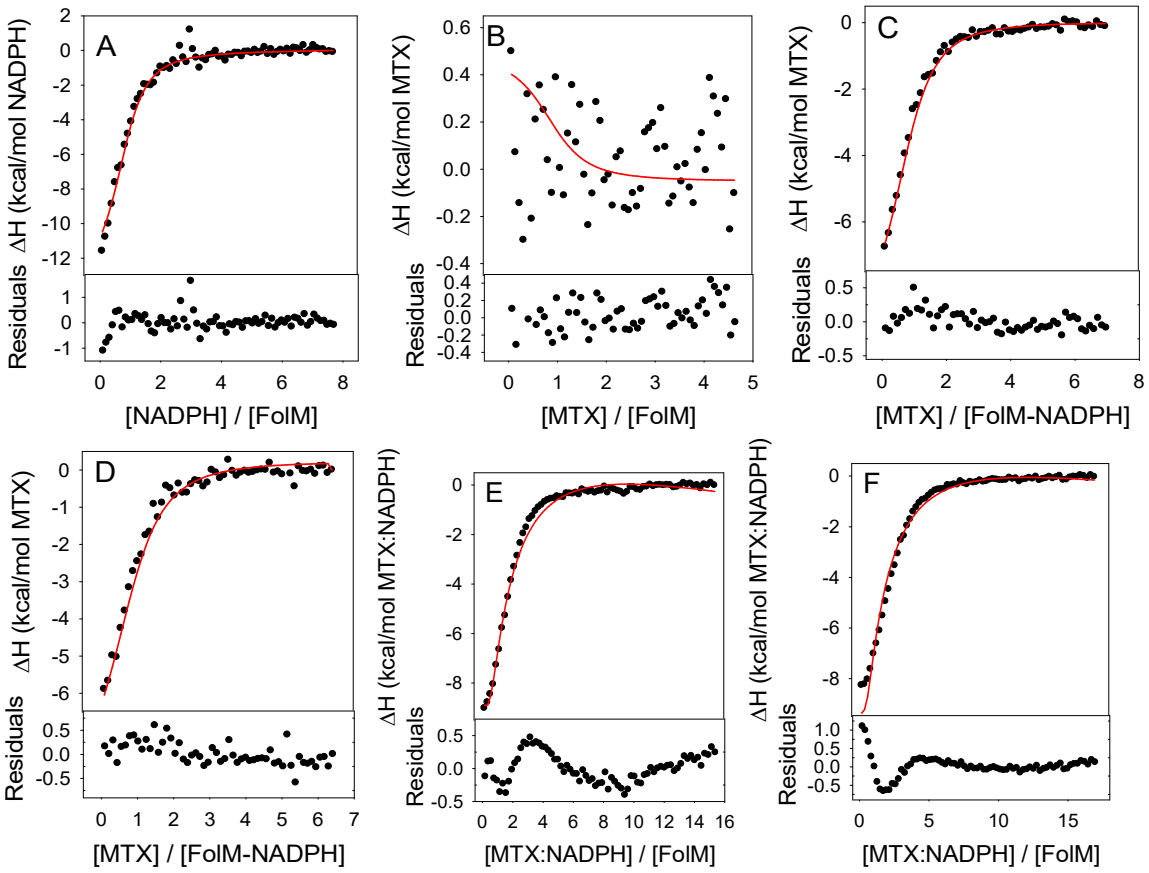


Figure 2.15 - Global fitting of ITC data for ligand binding to FolM using SEDPHAT. Plots are for binding of (A) NADPH to apo FolM, (B) MTX to apo FolM, (C) and (D) MTX to the FolM-NADPH binary complex and (E) and (F) a 1:1 mixture of MTX and NADPH binding to apo FolM. Best fit values are given in Table 2.6. If the ΔH for panel B is set equal to zero, e.g. MTX binary binding to FolM, the fits values are similar although the errors are much higher.

Table 2.6 - Global fitting of MTX and NADPH binding to FolM in MTA buffer, pH 6.0, at 25 ºC. Fits of the data were performed using the triple binding complex model, A+B+C → AB + C → AC + B → ABC model in SEDPHAT.[29] Forcing the MTX binary titration to an enthalpy of 0 kcal/mol did not affect the fit values, though the errors became appreciably higher.

Ligand	**Binding to**	**K_d (μM)**	**ΔG (kcal/mol)**	**ΔH (kcal/mol)**	**TΔS (kcal/mol)**	**n**
NADPH	FolM	2.34 ± 0.86	-7.67	-13.1 ± 2.2	-5.10	0.98 ± 0.08
MTX	FolM	1.09 ± 0.50	-8.12	0.81 ± 0.54	8.93	0.98 ± 0.08
MTX	FolM-NADPH	5.07 ± 0.66	-7.22	-12.9 ± 3.8	-5.65	0.98 ± 0.08

Betaine also weakens the binding of methotrexate. As MTX differs from folate by the substitution of an amine group for a carbonyl off C4 of the pterin ring as well as a methyl group off N10, these changes apparently do not greatly alter the weak attraction with betaine. Our previous NMR and osmometry results found betaine weakly associates with both the pterin and benzoyl rings of folate.[67] Also Capp et al.[63] has studied the interaction of betaine with small molecules and find that betaine associates with aromatic groups and amide nitrogens. Both these moieties are found in DHF as well as MTX, thus it is not surprising that betaine effects on MTX binding are also found. This analysis suggests that preferential interaction with osmolytes is likely a general property of folate derivatives. It also predicts that *in vivo* binding of antifolates will be weakened by osmotic stress conditions.

The interaction between betaine and other molecules can be predicted using μ_{23}/RT values established by the Record group.[63] μ_{23}/RT values, closely related to preferential interaction coefficients, measure the favorability of a molecule (in this case, betaine) to interact with another solute as compared to water. From Table 1 in Capp et al.,[63] the strongest association exists between betaine and sodium benzoate (μ_{23}/RT value of -0.091 m^{-1}). A negative (repulsive) interaction was found between betaine and tripotassium citrate with a μ_{23}/RT value of 1.2 m^{-1}. Intermediate μ_{23}/RT values span this range and allow prediction of interaction potentials, from strong to weak. NADPH has a predicted μ_{23}/RT value of 1.11 m^{-1} indicating that its interactions with betaine are unfavorable. From the data in Figure 2.8, a $\Delta\mu_{23}/RT$ of -0.43 ± 0.13 m^{-1} was calculated for betaine interaction with NADPH. This is slightly less (accounting for differences in sign) than the predicted value, but it still predicts minimal association between betaine and NADPH. Our results for

cofactor binding to FolM, as well as R67 DHFR[2] and EcDHFR[9] indicate negative interactions between betaine and NADPH, concurring with the predicted μ_{23}/RT value. Likewise, the predicted μ_{23}/RT values for DHF, folate and MTX (0.35, -0.01 and 0.01, respectively) indicate they are more likely to associate with betaine; this calculation agrees qualitatively with our experimental results. The $\Delta\mu_{23}$/RT values calculated from Figures 2.11 and 2.14 are 0.69 ± 0.03 m^{-1} and 0.63 ± 0.17 m^{-1} for DHF and MTX, respectively. While the experimental values are different from the predicted μ_{23}/RT values, the presence of other solutes in the experiment may potentially affect the actual $\Delta\mu_{23}$/RT value. Overall this analysis suggests exclusion of betaine from the solvation shell of NADPH and preferential interaction of betaine with folate/DHF/MTX.

2.5.B. Other Osmolytes Weaken Ligand Binding

In contrast to betaine, other osmolytes weaken the binding of NADPH to FolM. It is unlikely that these osmolytes are attracted to NADPH, as DMSO, ethylene glycol and PEG400 all increased the affinity of the cofactor for R67 DHFR and EcDHFR.[2, 9] The most likely alternative is that these osmolytes associate with FolM.[8, 68] This is an additional complicating factor.

For most osmolytes studied, the $k_{cat}/K_{m\ (DHF)}$ decreased with added osmolyte with most of the effects on $K_{m\ (DHF)}$. These results are similar to the decrease in $k_{cat}/K_{m\ (DHF)}$ for R67 DHFR with increasing osmolyte concentrations.[2] The most likely cause for the decrease in the FolM $k_{cat}/K_{m\ (DHF)}$ is osmolyte association with free DHF. The range in slopes for $k_{cat}/K_{m\ (DHF)}$ with water activity indicates that there are differences in the preferential interaction between glycerol or betaine (for example) with DHF, and/or that there are additional attractions between the osmolytes and FolM.[43, 68] Similar

preferential interactions were also noted for DHF binding to EcDHFR.[9] In addition, binding of the antifolate drug, methotrexate, was monitored and found to be weakened by the addition of osmolytes.

2.5.C. How do Osmolytes Alter Ligand Binding?

There are several options as to how osmolytes affect both cofactor and substrate binding to FolM. Osmolyte effects on NADPH affinity for R67 DHFR and EcDHFR were mostly due to changes in water activity by the osmolytes.[2, 9] However in the FolM case, some osmolytes increase the affinity of NADPH, while others decrease the affinity. One scenario that could account for these variable effects is that some osmolytes may bind to FolM in such a way as to prevent NADPH from binding. This could involve osmolyte binding/solvation of FolM in the active site. Alternatively, osmolyte association at another site(s) could alter the conformation of FolM or the population of apo FolM states such that the cofactor binding equilibrium is shifted towards the free state.[69, 70] Removal of these osmolytes would require input of energy to the system and lead to weaker binding of NADPH to FolM.

While osmolytes that are excluded from the protein surface stabilize proteins (e.g. trimethylamine oxide with chemically modified RNase T_1,[71] FolM with betaine), denaturants that interact with the protein (for example urea) destabilize proteins.[71-75] DMSO,[45, 46] as well as other osmolytes,[44] have also been found to destabilize proteins. Lin and Timasheff propose it is the difference between association of co-solvent to the native and denatured state that affects the stability of a protein compared to water.[76] From this point of view, while the DMSO or ethylene glycol interactions with FolM are not so apparent when comparing the CD spectra plus and minus osmolytes, these

weak interactions are clear when NADPH K_d values and T_m values are compared. While addition of DMSO and ethylene glycol did not weaken cofactor binding to R67 DHFR and EcDHFR, apparently the different sequence, surface, and/or structure of FolM result in an attractive effect. Thus each protein presents different contexts and effects. The case with betaine (most excluded) allows evaluation of the simplest case, while addition of other osmolytes can have variable effects that depend on the protein.

Once sufficient NADPH is added to overcome the negative effect of the osmolyte-FolM interaction, work has been done. It is not clear whether other osmolyte-FolM interactions are present and also exact a penalty on binding of DHF or MTX to FolM-NADPH. Again, the clearest case involves betaine, where the two effects are more clearly separated. Preferential exclusion of betaine from NADPH and FolM results in water release upon NADPH binding (tighter K_d values). However, association of betaine with DHF or MTX results in weaker substrate/inhibitor binding. Occam's razor suggests this pattern will continue with the other osmolytes with the added layer of osmolyte-FolM effects. When the slope values are compared for the ln $k_{cat}/K_{m\,(DHF)}$ vs. ln a_o plots, if osmolyte-FolM interactions affected DHF binding (in addition to osmolyte-DHF interactions), we might expect even larger positive numbers. However, the values are generally smaller than observed for R67 DHFR and EcDHFR. This observation may suggest osmolyte-FolM effects do not further weaken DHF ternary complex formation.

Another consideration that may come into play with respect to the smaller positive slope values (Table 2.4) associated with the FolM ln $k_{cat}/K_{m\,(DHF)}$ vs. ln a_o plot is the contact area between DHF and FolM-NADPH. Based on the crystal structure of PTR1 with the folate-based inhibitor CB3717,[18] portions of the pterin and *p*-ABA rings do not contact

the binding site, instead they are exposed to solution. Similarly, parts of the substrate may also be solvent exposed when bound to FolM. In this scenario, the number of osmolytes that have to be removed from DHF prior to its binding to FolM may be less than for R67 DHFR or EcDHFR because of this difference in solvent exposure. With fewer osmolytes removed, the increase in FolM $K_{m\ (DHF)}$ or $K_{d\ (MTX)}$ would not be as great as it is for R67 DHFR or EcDHFR.

2.5.D. Comparison of DHFR Enzymes

Osmolytes influence ligand binding to FolM differently compared to R67 DHFR and EcDHFR. For both R67 DHFR and EcDHFR, each osmolyte had a unique effect on DHF binding.[2, 9] These results are interpreted as preferential interaction of the osmolytes with the proteins.[43, 68] Different slopes were noted for DHF binding to FolM as well (Figure 2.11), which suggests preferential interaction of the osmolytes with FolM. However, unlike the other DHFRs, FolM binding of NADPH is also weakened in the presence of some osmolytes (Figure 2.8). The destabilizing interactions of osmolytes with FolM also perturb cofactor binding.

Though R67 DHFR, EcDHFR and FolM all have DHFR activity, they all have very different structures (Figure 2.2). All three have decreased substrate binding in the presence of osmolytes, indicating that osmolyte interactions with free DHF do shift the substrate binding equilibrium towards the free state. However, the three DHFR enzymes all interact to different extents with the osmolytes as well. The different sequence and structural characteristics of each enzyme make each enzyme more, or less, susceptible to associating with osmolytes. In the case of FolM, some of these osmolyte interactions can destabilize the enzyme, weakening ligand binding as well. Unfortunately, a comparison of the active

sites of R67 DHFR, EcDHFR, FolM and two pteridine reductases does not show large differences in character, precluding prediction of protein-osmolyte effects at this time.

2.6 Conclusion

Preferential interaction of osmolytes with DHF and the antifolate, methotrexate, decreases their affinity for FolM. In addition to the interaction of osmolytes with substrate/inhibitor, some osmolytes associate with, and destabilize FolM. Destabilization of FolM by DMSO, ethylene glycol and PEG400 weakens the binding affinity of NADPH. Exclusion of betaine whereas interaction of DMSO was observed with FolM. These osmolyte-FolM interactions may also contribute to the decrease in DHF affinity. Therefore, while interaction between osmolytes and DHF can be noted for FolM, additional interactions between some osmolytes with FolM complicate the analysis.

2.7 References

1. Fried, M.G., et al., *Role of hydration in the binding of lac repressor to DNA.* J. Biol. Chem., 2002. 277(52): p. 50676-50682.
2. Chopra, S., et al., *A balancing act between net uptake of water during dihydrofolate binding and net release of water upon NADPH binding in R67 dihydrofolate reductase.* J. Biol. Chem., 2008. 283(8): p. 4690-4698.
3. Bradrick, T.D., J.M. Beechem, and E.E. Howell, *Unusual binding stoichiometries and cooperativity are observed during binary and ternary complex formation in the single active pore of R67 dihydrofolate reductase, a D_2 symmetric protein.* Biochemistry, 1996. 35(35): p. 11414-11424.
4. Krahn, J., et al., *Structure of a type II dihydrofolate reductase ternary complex: Use of identical binding sites for unrelated ligands.* Biochemistry, 2007. 46: p. 14878-14888.
5. Arakawa, T. and S.N. Timasheff, *The stabilization of proteins by osmolytes.* Biophys. J., 1985. 47(3): p. 411-414.
6. Timasheff, S.N., *The control of protein stability and association by weak interactions with water: how do solvents affect these processes?* Annu. Rev. Biophys. Biomol. Struct., 1993. 22: p. 67-97.
7. Street, T.O., D.W. Bolen, and G.D. Rose, *A molecular mechanism for osmolyte-induced protein stability.* Proc. Natl. Acad. Sci. U.S.A., 2006. 103(38): p. 13997-14002.
8. Parsegian, V.A., R.P. Rand, and D.C. Rau, *Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives.* Proc. Natl. Acad. Sci. U. S. A., 2000. 97(8): p. 3987-3992.
9. Grubbs, J., et al., *Thermodynamics and solvent effects on substrate and cofactor binding in Escherichia coli chromosomal dihydrofolate reductase.* Biochemistry, 2011. 50: p. 3673-3685.
10. Dzingeleski, G.D. and R. Wolfenden, *Hypersensitivity of an enzyme reaction to solvent water.* Biochemistry, 1993. 32(35): p. 9143-9147.
11. Furukawa, Y. and I. Morishima, *The role of water molecules in the association of cytochrome P450cam with putidaredoxin. An osmotic pressure study.* J. Biol. Chem., 2001. 276(16): p. 12983-12990.
12. Xavier, K.A., et al., *Involvement of water molecules in the association of monoclonal antibody HyHEL-5 with bobwhite quail lysozyme.* Biophys. J., 1997. 73(4): p. 2116-2125.
13. Giladi, M., et al., *FolM, a new chromosomally encoded dihydrofolate reductase in Escherichia coli.* J. Bacteriol., 2003. 185(23): p. 7015-7018.
14. Levin, I., M. Mevarech, and B.A. Palfey, *Characterization of a novel bifunctional dihydropteroate synthase/dihydropteroate reductase enzyme from Helicobacter pylori.* J. Bacteriol., 2007. 189(11): p. 4062-4069.
15. Myllykallio, H., et al., *Life without dihydrofolate reductase FolA.* Trends Microbiol., 2003. 11(5): p. 220-223.
16. Levin, I., et al., *An alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea.* Mol Microbiol, 2004. 54(5): p. 1307-1318.

17. Gourley, D.G., et al., *Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites.* Nat. Struct. Biol., 2001. 8(6): p. 521-525.
18. Schuttelkopf, A.W., et al., *Structures of Leishmania major pteridine reductase complexes reveal the active site features important for ligand binding and to guide inhibitor design.* J. Mol. Biol., 2005. 352(1): p. 105-116.
19. Pribat, A., et al., *FolX and FolM are essential for tetrahydromonapterin synthesis in Escherichia coli and Pseudomonas aeruginosa.* J. Bacteriol., 2010. 192(2): p. 475-482.
20. Yahashiri, A., *Comparative investigations of H-transfer in dihydrofolate reductases from different families.* PhD. Thesis, 2010. University of Iowa, Chemistry Department.
21. Nare, B., L.W. Hardy, and S.M. Beverley, *The roles of pteridine reductase 1 and dihydrofolate reductase-thymidylate synthase in pteridine metabolism in the protozoan parasite Leishmania major.* J. Biol. Chem., 1997. 272(21): p. 13883-13891.
22. Sawaya, M.R. and J. Kraut, *Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence.* Biochemistry, 1997. 36(3): p. 586-603.
23. Strader, M.B., et al., *Role of S65, Q67, I68, and Y69 residues in homotetrameric R67 dihydrofolate reductase.* Biochemistry, 2001. 40(38): p. 11344-11352.
24. Ellis, K.J. and J.F. Morrison, *Buffers of constant ionic strength for studying pH-dependent processes.* Methods Enzymol., 1982. 87: p. 405-426.
25. Blakley, R.L., *Crystalline dihydropteroylglutamic acid.* Nature, 1960. 188: p. 231-232.
26. Horecker, B.L., and Kornberg, A., *The extinction coefficients of the reduced band of pyridine nucleotides.* J. Biol. Chem., 1948. 175: p. 385-390.
27. Baccanari, D., et al., *Purification and properties of Escherichia coli dihydrofolate reductase.* Biochemistry, 1975. 14(24): p. 5267-5273.
28. Wiseman, T., et al., *Rapid measurement of binding constants and heats of binding using a new titration calorimeter.* Anal. Biochem., 1989. 179(1): p. 131-137.
29. Houtman, J.C., et al., *Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling.* Protein Sci., 2007. 16(1): p. 30-42.
30. Keller, S., et al., *High-precision isothermal titration calorimetry with automated peak-shape analysis.* Anal. Chem., 2012. 84(11): p. 5066-5073.
31. Schuck, P., *Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling.* Biophys. J., 2000. 78(3): p. 1606-1619.
32. Laue, T.M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L., *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, S.E. Harding, Rowe, A. J., and Horton, J. C., Editor. 1992, Royal Society of Chemistry: Cambridge. p. 90-125.
33. Sweeney, T.E. and C.A. Beuchat, *Limitations of methods of osmometry: Measuring the osmolality of biological fluids.* Am. J. Physiol., 1993. 264(3 Pt 2): p. R469-480.

34. Zhuang, P., et al., *Artificial duplication of the R67 dihydrofolate reductase gene to create protein asymmetry: Effects on protein activity and folding.* J. Biol. Chem., 1993. 268(30): p. 22672-22679.
35. Seeger, D.R., et al., *Analogs of pteroylglutamic acid. III. 4-amino derivatives.* J. Am. Chem. Soc., 1949. 71(5): p. 1753-1758.
36. http://www.ebi.ac.uk/Tools/msa/clustalw2/help/faq.html#21, C.
37. http://web.expasy.org/protparam/, P.
38. Sturtevant, J.M., *Biochemical applications of differential scanning calorimetry.* Ann. Rev. Phys. Chem., 1987. 38: p. 463-488.
39. Johnson, C.M., *Differential scanning calorimetry as a tool for protein folding and stability.* Arch. Biochem. Biophys., 2013. 531: p. 100-109.
40. Sanchez-Ruiz, J.M., et al., *Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin.* Biochemistry, 1988. 27(5): p. 1648-1652.
41. Sanchez-Ruiz, J.M., *Protein kinetic stability.* Biophys. Chem., 2010. 148(1-3): p. 1-15.
42. Barrack, K.L., et al., *Structure of recombinant Leishmania donovani pteridine reductase reveals a disordered active site.* Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun., 2011. 67(Pt 1): p. 33-37.
43. Courtenay, E.S., et al., *Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro.* Biochemistry, 2000. 39(15): p. 4455-4471.
44. Singh, L.R., et al., *Protein and DNA destabilization by osmolytes: The other side of the coin.* Life Sci., 2011. 88(3-4): p. 117-125.
45. Torreggiani, A., et al., *Effect of sulfoxides on the thermal denaturation of hen lysozyme: A calorimetric and Raman study.* J. Molec. Struct., 2008. 891(1-3): p. 115-122.
46. Yang, Z.W., et al., *Dimethyl sulfoxide at 2.5% (v/v) alters the structural cooperativity and unfolding mechanism of dimeric bacterial NAD^+ synthetase.* Protein Sci., 2004. 13(3): p. 830-841.
47. Mitra, L., et al., *Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution--experiments and theoretical interpretation.* Phys. Chem. Chem. Phys., 2006. 8(11): p. 1249-1265.
48. Attri, P., P. Venkatesu, and M.J. Lee, *Influence of osmolytes and denaturants on the structure and enzyme activity of alpha-chymotrypsin.* J. Phys. Chem. B, 2010. 114(3): p. 1471-1478.
49. Edsall, J., *Dielectric constants and dipole moments of dipolar ions*, in *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, E. Cohn and J. Edsall, Editors. 1943, Reinhold: New York. p. 140-154.
50. Kiser, J.R., et al., *Hydration changes in the association of Hoechst 33258 with DNA.* Biochemistry, 2005. 44(51): p. 16988-16997.
51. Chang, C.-F., T. Bray, and J.M. Whiteley, *Mutant PTR1 proteins from Leishmania tarentolae: Comparative kinetic properties and active-site labeling.* Arch. Biochem. Biophys., 1999. 368(1): p. 161-171.
52. Kawasaki, Y. and E. Freire, *Finding a better path to drug selectivity.* Drug Discovery Today, 2011. 16(21-22): p. 985-990.

53. Perozzo, R., et al., *Compulsory order of substrate binding to herpes simplex virus type 1 thymidine kinase. A calorimetric study.* J. Biol. Chem., 2000. 275(21): p. 16139-16145.
54. Luba, J., et al., *Leishmania major pteridine reductase 1 belongs to the short chain dehydrogenase family: Stereochemical and kinetic evidence.* Biochemistry, 1998. 37: p. 4093-4104.
55. Lewis, C.A., Jr. and R. Wolfenden, *Orotic acid decarboxylation in water and nonpolar solvents: A potential role for desolvation in the action of OMP decarboxylase.* Biochemistry, 2009. 48(36): p. 8738-8745.
56. Wolfenden, R., *Waterlogged molecules.* Science, 1983. 222(4628): p. 1087-1093.
57. Cannon, W.R. and S.J. Benkovic, *Solvation, reorganization energy, and biological catalysis.* J. Biol. Chem., 1998. 273(41): p. 26257-26260.
58. Warshel, A., J. Aqvist, and S. Creighton, *Enzymes work by solvation substitution rather than by desolvation.* Proc. Nat. Acad. Sci. U.S.A., 1989. 86(15): p. 5820-5824.
59. Levy, Y. and J.N. Onuchic, *Water mediation in protein folding and molecular recognition.* Annu. Rev. Biophys. Biomol. Struct., 2006. 35: p. 389-415.
60. Dewar, M.J. and D.M. Storch, *Alternative view of enzyme reactions.* Proc. Natl. Acad. Sci. U. S. A., 1985. 82(8): p. 2225-2229.
61. Cayley, S., et al., *Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo.* J. Mol. Biol., 1991. 222(2): p. 281-300.
62. Cayley, S. and M.T. Record, Jr., *Roles of cytoplasmic osmolytes, water, and crowding in the response of Escherichia coli to osmotic stress: biophysical basis of osmoprotection by glycine betaine.* Biochemistry, 2003. 42(43): p. 12596-12609.
63. Capp, M.W., et al., *Interactions of the osmolyte glycine betaine with molecular surfaces in water: Thermodynamics, structural interpretation, and prediction of m-values.* Biochemistry, 2009. 48 p. 10372–10379.
64. Guinn, E.J., et al., *Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer.* Proc. Natl. Acad. Sci. U. S. A., 2011. 108(41): p. 16932-16937.
65. Santoro, M.M., et al., *Increased thermal stability of proteins in the presence of naturally occurring osmolytes.* Biochemistry, 1992. 31(23): p. 5278-5283.
66. Felitsky, D.J. and M.T. Record, Jr., *Application of the local-bulk partitioning and competitive binding models to interpret preferential interactions of glycine betaine and urea with protein surface.* Biochemistry, 2004. 43(28): p. 9276-9288.
67. Duff, M.R., Jr., et al., *Weak interactions between folate and osmolytes in solution.* Biochemistry, 2012. 51(11): p. 2309-2318.
68. Timasheff, S.N., *Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components.* Proc. Natl. Acad. Sci. U. S. A., 2002. 99(15): p. 9721-9726.
69. Boehr, D.D. and P.E. Wright, *How do proteins interact?* Science, 2008. 320(5882): p. 1429-1430.
70. Kumar, S., et al., *Folding and binding cascades: Dynamic landscapes and population shifts.* Protein Sci., 2000. 9(1): p. 10-19.

71. Baskakov, I.V., et al., *Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor.* J. Biol. Chem., 1999. 274(16): p. 10693-10696.
72. Diamant, S., et al., *Chemical chaperones regulate molecular chaperones in vitro and in cells under combined salt and heat stresses.* J. Biol. Chem., 2001. 276(43): p. 39586-39591.
73. Venkatesu, P., M.J. Lee, and H.M. Lin, *Osmolyte counteracts urea-induced denaturation of alpha-chymotrypsin.* J. Phys. Chem. B, 2009. 113(15): p. 5327-5338.
74. Wu, P. and D.W. Bolen, *Osmolyte-induced protein folding free energy changes.* Proteins, 2006. 63(2): p. 290-296.
75. Belluzo, S., et al., *Influence of stabilizers cosolutes on catalase conformation.* Int. J. Biol. Macromol., 2011. 49(5): p. 936-941.
76. Lin, T.Y. and S.N. Timasheff, *Why do some organisms use a urea-methylamine mixture as osmolyte - thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein.* Biochemistry, 1994. 33(42): p. 12695-12701.

PART 3. ASPECTS OF WEAK INTERACTIONS BETWEEN FOLATE AND GLYCINE BETAINE

This section is a slightly modified version of a manuscript by the same title submitted and accepted in the journal, *Biochemistry* in October, 2016

Reprinted (adapted) with permission from Purva P. Bhojane[‡], Michael R. Duff, Jr.[‡], Khushboo Bafna[¥], Gabriella P. Rimmer[‡], Pratul K. Agarwal[‡,¥,ǂ] and Elizabeth E. Howell[‡,¥] *., "Aspects of Weak Interaction between Folate and Glycine Betaine", *Biochemistry*. Copyright 2016 American Chemical Society.
(DOI: 10.1021/acs.biochem.6b00873)

[‡]Department of Biochemistry, Cellular, and Molecular Biology
University of Tennessee, Knoxville, TN 37996-0840, [¥]Genome Science and Technology Program, University of Tennessee, Knoxville, TN 37996-0840, [ǂ]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

Address correspondence to Elizabeth E. Howell, Department of Biochemistry, Cellular and Molecular Biology, University of TN, Knoxville, TN 37996-0840; phone, 865-974-4507; fax, 865-974-6306; e-mail, lzh@utk.edu.This work was supported by NIH grants GM 110669 (to EEH) and GM105978 (to PKA).

This author contributed the following to the manuscript: (1) vapor pressure osmometry, (2) solubility assays, (3) protein expression and purification, (4) protein extinction co-efficient, (5) isothermal titration calorimetry, (6) most of the data analysis and organization, (7) manuscript preparation.

3.1 Abstract

Folate, or vitamin B9, is an important compound in one carbon metabolism. Previous studies have found weaker binding of dihydrofolate to dihydrofolate reductase in the presence of osmolytes. In other words, osmolytes are more difficult to remove from the dihydrofolate solvation shell than water; this shifts the equilibrium towards the free ligand and protein species. This study uses vapor pressure osmometry to explore the interaction of folate with the model osmolyte, glycine betaine. This method yields a preferential interaction potential (μ_{23}/RT value). This value is concentration dependent as folate dimerizes. The μ_{23}/RT value also tracks the deprotonation of folate's N3-O4 keto-enol group, yielding a pK_a of 8.1. To determine which folate atoms, interact most strongly

with betaine, the interaction of heterocyclic aromatic compounds (as well as other small molecules) with betaine was monitored. Using an accessible surface area approach coupled with osmometry measurements, deconvolution of the μ_{23}/RT value into α values for atom types was achieved. This allows prediction of μ_{23}/RT values for larger molecules such as folate. Molecular dynamics simulations of folate show a variety of structures from extended to L-shaped. These conformers possess μ_{23}/RT values from -0.18 to 0.09 m^{-1}, where a negative value indicates a preference for solvation by betaine and a positive value indicates a preference for water. This range of values is consistent with values observed in osmometry and solubility experiments. As the average predicted folate μ_{23}/RT value is near zero, this indicates folate interacts almost equally well with betaine and water. Specifically, the glutamate tail prefers to interact with water while the aromatic rings prefer betaine. In general, the more protonated species in our small molecule survey interact better with betaine as they provide a source of hydrogens (betaine is not a hydrogen bond donor). Upon deprotonation of the small molecule, the preference swings towards water interaction due to its hydrogen bond donating capacities.

3.2 Introduction

How do two molecules come together and form a complex? Two steps are typically involved, desolvation and association. While forces that drive association are reasonably well understood, the role water plays is difficult to predict. For example, water can fill voids in structures and also provide a bridge between surfaces.[1-5] While high concentrations of water are present in test tube studies, the situation gets more complicated in the cell due to the presence of many other molecules. If other solutes, for example

osmolytes, interact with ligands and/or proteins, they need to be removed to form the protein-ligand complex. While these solute-ligand interactions are weak, the relative strength of the ligand-osmolyte interaction vs. that of the ligand-water interaction can affect binding to the protein partner. Binding will be either facilitated or made more difficult, resulting in altered K_d values between macromolecules and their ligands.

In most cases, the binding constant becomes tighter in the presence of osmolytes as the desolvation penalty is minimized.[6] An example of this is binding of the cofactor NADPH to R67 dihydrofolate reductase (DHFR).[7] However, if osmolytes prefer to interact with the ligand or protein, and if removing them is more difficult than shedding water, then the binding constant is weakened. This case is exemplified by binding of dihydrofolate to various DHFRs.[7-10] In this model, shown in Figure 3.1, the osmolytes shift the reaction equilibrium towards the free species of substrate and DHFR compared to the protein-ligand complex. One osmolyte that weakens DHF binding to R67 DHFR by 3.6 fold is glycine betaine (20% w/v). Note: DHFR catalyzes reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using NADPH as a cofactor. There are two types of DHFRs; type I is encoded by the chromosome and type II is carried by a resistance plasmid, an example is R67 DHFR. Neither the structures nor mechanisms are homologous in these DHFRs.[11]

Glycine betaine or N, N, N,-trimethyl glycine is proposed to be one of the most effective osmoprotectants in *E. coli* cells as it is efficient in maintaining growth under osmotic stress.[12] Betaine is unable to act as a hydrogen bond donor, thus it is highly excluded from the surface of proteins, facilitating macromolecular functions. How does betaine interact with folate? Homonuclear (^{1}H) Nuclear Overhauser Effect spectroscopy

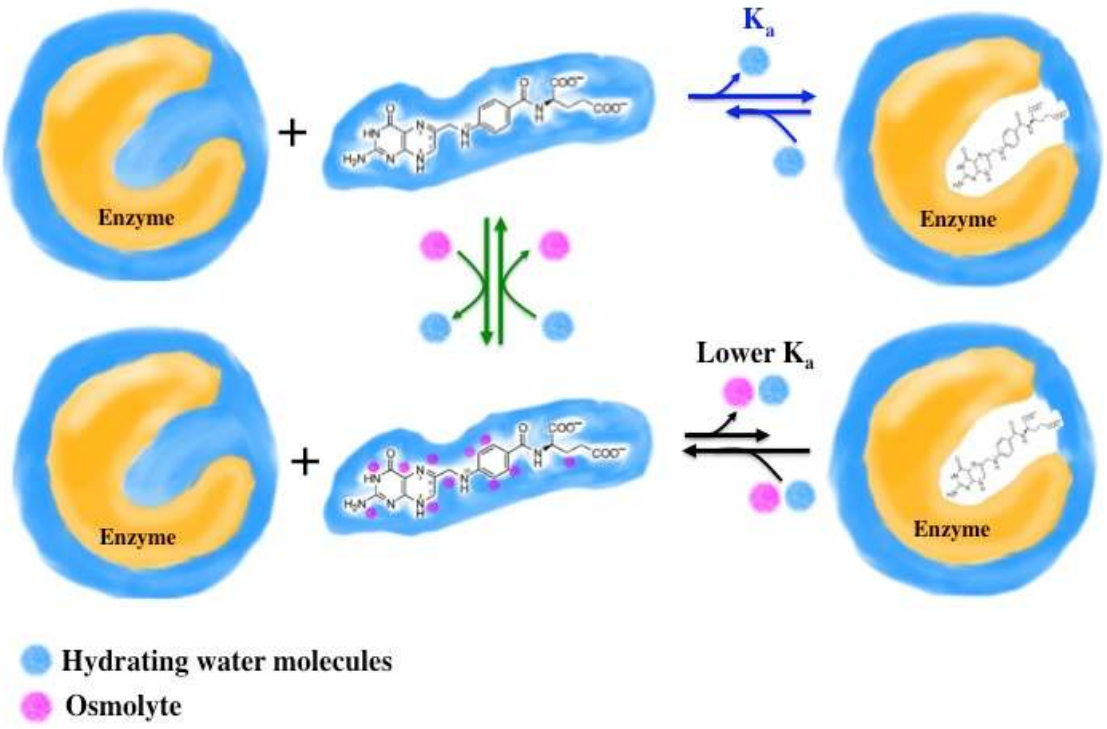


Figure 3.1 - A cartoon depicting preferential interaction of osmolytes with free DHF. In the absence of osmolytes, DHF binds tightly to its target enzyme and water (blue) is released. Added osmolytes (**magenta spheres**) interact weakly with DHF. For DHF to bind to the enzyme, both osmolytes and water must be released. Osmolytes that interact more strongly than water would have larger effects on the DHF K_a while the more weakly bound osmolytes would have smaller effects. (Note: this model does not exclude the possible binding of osmolytes to the enzyme.) We have used high hydrostatic pressure as an orthogonal technique to examine the top row of the model (**blue equilibrium arrows**).[13] We have also used NMR to observe interactions between folate and osmolytes (middle column, **green equilibrium arrows**).[9] Both sets of results are consistent with this model.

(NOESY) experiments found NOEs between the protons on the C7 atom of folate with protons on the C9 and C3′/C5′ atoms, and between the C9 proton and the C3′/C5′ protons.[9] (see Figure 3.2 for atom numbers folate structures). However, a change in sign for the NOE between the C9 and C3'/C5' protons from positive (without betaine) to negative (with betaine) was observed. The change in the sign for the NOE suggested a slower rotational rate for the *p*-amino-benzoyl ring protons, indicative of an interaction between betaine and this ring.[9] As folate has limited protons on its pterin ring, it is difficult to discern from our NMR results if osmolytes interact with this moiety. Thus we turned to alternate techniques.

Our previous osmometry studies support interactions between betaine with the folate fragments, *p*-amino benzoyl-glutamate (*p*-ABA-Glu) and pterin-6 carboxylate.[9] To extend this study, we use vapor pressure osmometry (VPO) to measure μ_{23}/RT values, which are closely related to preferential interaction coefficients. The μ_{23}/RT value measures the change in chemical potential of a small test compound with the change in molality of the osmolyte in solution. For our study, the μ_{23}/RT value measures the preference of small molecules to interact with betaine compared to water. The Record lab has pioneered this VPO approach along with a water-accessible surface area (ASA) analysis to quantify and analyze the thermodynamics of interaction of osmolytes (betaine, proline, PEG and urea) with model compounds displaying biomolecular functional groups.[14-17] The VPO method measures the favorability of a small molecule interacting with an osmolyte as compared to water in a three-component system (1- water, 2-test compound and 3- osmolyte). Capp et al. studied the interaction of betaine with a set of model compounds containing carboxylate, phosphate, amide, hydroxyl, ammonium,

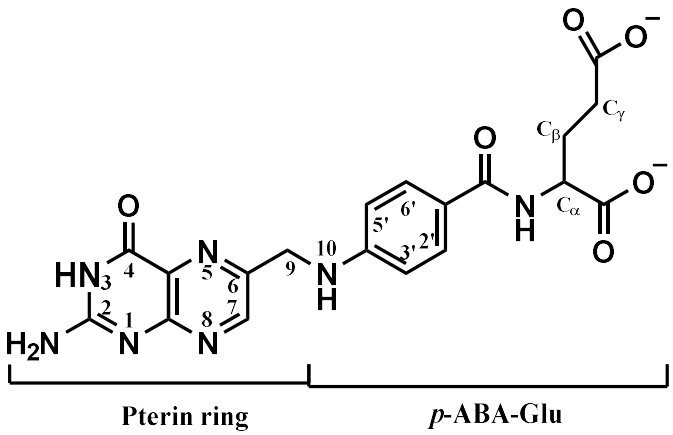


Figure 3.2 - Structure of Folate. The structure of folate with atom numbers is shown. Folate contains pterin and *p*-amino benzoate rings and a glutamate tail.

guanidinium, aliphatic and aromatic hydrocarbon moieties.[14] A positive μ_{23}/RT value for phosphate ($0.85 \pm 0.04\ m^{-1}$) and citrate ($1.2 \pm 0.1\ m^{-1}$) indicates a strong preference for water over betaine, whereas negative μ_{23}/RT values for benzoate ($-0.091 \pm 0.007\ m^{-1}$) and urea ($-0.093 \pm 0.005\ m^{-1}$) indicate a preference for betaine over water.

The preferential interaction potentials, or μ_{23}/RT values, obtained for those compounds were dissected into additive contributions from chemically distinct functional groups. The calculated set of atomistic preferential interaction potentials per unit water-accessible surface areas (ASA) of each surface type, also called α values, can be coupled with the ASA information to predict the μ_{23}/RT of any compound. We take this approach to understand how folate interacts with betaine as well as betaine effects on folate binding to DHFR.

3.3 Materials and Methods

3.3.A. Materials

Betaine, folic acid, indole acetate, *m*-aminobenzoate, *o*-aminobenzoate, *p*-toluic acid, *p*-aminobenzoate-glutamate, pyrrole-2-carboxylate, adenosine 5'-monophosphate, guanosine 5'-monophosphate, cytidine 2'-monophosphate and thymidine 5'-monophosphate were purchased from Sigma-Aldrich. Nicotinic acid, nicotinamide, pyrimidone, pyridoxine-HCl were from Acros Organics, *p*-aminobenzoate was from MP Biomedicals, and phenylalanine-HCl was from Fisher Scientific. Uridine 3'-monophosphate was from Chem-Impex International Inc. Pteroyltetra-γ-L-glutamate (PG4) was from Schircks Laboratories.

3.3.B. Vapor Pressure Osmometry (VPO)

We use folate in our studies as it is more stable than dihydrofolate. In VPO experiments, the change in osmolality of bulk water is measured in a multi-component system containing components 1, 2 and 3, which denote water, test compound and betaine (osmolyte) respectively. This technique monitors the change in osmolality (Δ Osm) of a solution, which is a quantitative measure of the favorable or unfavorable interaction of the two solutes (test compound and osmolyte), relative to their interactions with water.[14-17] As the solution osmolality increases due to increasing betaine concentrations, any change in measured osmolality arises due to the interaction of betaine with the test compound. If betaine is excluded from the surface of the test compound, the betaine concentration in the bulk media (relative to the betaine only control) is increased. This in turn decreases the bulk water concentration and increases the osmolality of the solution. If there is no preference for betaine or water to interact with the test compound, the osmolalities of betaine and small molecule are additive. If betaine prefers to interact with the test compound, the betaine concentration in the bulk media is decreased, which increases the bulk water concentration and decreases the solution osmolality. The difference in osmolality between the solution of the test compound with and without betaine , Δ Osm, when plotted versus the product of betaine and the test compound molality, $m_2 m_3$, yields a linear plot, the slope of which is the μ_{23}/RT value,

$$\Delta\,\mathrm{Osm} = \mathrm{Osm}(m_2, m_3) - \mathrm{Osm}(m_2, 0) - \mathrm{Osm}(0, m_3) \cong \left(\frac{\mu_{23}}{\mathrm{RT}}\right) m_2\, m_3 \qquad \text{Eq (3.1)}$$

where m_2 and m_3 are molal concentrations of test compound and betaine, respectively and μ_{23}/RT is the relative chemical potential of the test compound in betaine. If μ_{23} is independent of m_2 and m_3, it approximates the preferential interaction potential.

Experiments were performed on a Wescor Vapro 5520 osmometer. The instrument was calibrated using standard solutions of 0.100, 0.290, and 1.000 osmol. An additional linear calibration curve was made by measuring 1.000, 1.500 and 2.000 osmol standards to correct for osmolality readings above 1.000 osmol. A betaine stock solution (2 *m*) was prepared daily using a gravimetric method. Betaine (2 gm) was weighed and dissolved to make a 10 ml stock solution in a pre-weighed tube. The weight of water was determined by subtracting the weight of betaine from the weight of the solution, which was then used to calculate the molal concentration of the stock. Typically, 30-500 mg of the test compound (for example, folate) was added to a pre-weighed microfuge tube and stock solutions were prepared fresh daily in water. The molality of the stock solutions was determined using the weight of the solution. A series of betaine solutions were prepared and the osmolality of each was measured in triplicate. Then, solutions containing a desired concentration of the test compound with equivalent betaine concentrations as for the betaine only line were prepared and incubated at room temperature for 10 minutes. The osmolalities of the solutions were then measured in triplicate. The concentration of the test compound was constant in each experiment. Solutions were prepared such that the osmolality ranged between 0.1 and 2 Osm, which typically spanned the range of betaine concentrations from 0.1 to 1.25 *m*, and test compound concentrations from 0.04 to 0.5 *m*. The data were fit to Eq (3.1).

We used this method to determine μ_{23}/RT values for folate at pHs 7 and 10. Capp

et al. suggested not adjusting the pH of the stock solutions to avoid additional components in the system.[14] However as folic acid has a low solubility, we adjusted the pH using sodium hydroxide (1 N) to form folate. To ensure that we could compensate for NaOH addition, we precipitated the sodium folate salt at pH 10 in acetone and isopropanol, lyophilized it and re-did the VPO experiments. The μ_{23}/RT values were the same.

As folate dimerizes at high (non-physiological) concentrations,[18] we additionally monitored the μ_{23}/RT value as a function of folate concentration. The data were fit to a dimerization function, Eq (3.2), adapted from Duff et al.[9]

$$\frac{\mu_{23}}{\mathrm{RT}\,obs} = \frac{\mu_{23}}{\mathrm{RT}\,(\mathrm{D})} + \left(\frac{\mu_{23}}{\mathrm{RT}\,(\mathrm{M})} - \frac{\mu_{23}}{\mathrm{RT}\,(\mathrm{D})}\right)\left[-\mathrm{K_d} + (\mathrm{K_d}^2 + 8\mathrm{K_d}[\mathrm{F}]_{\mathrm{tot}})^{\frac{1}{2}}\right]/(4[\mathrm{F}]_{\mathrm{tot}}) \quad \text{Eq (3.2)}$$

where $\mu_{23}/\mathrm{RT}_{\mathrm{obs}}$ is the observed μ_{23}/RT, $\mu_{23}/\mathrm{RT}_{(\mathrm{M})}$ and $\mu_{23}/\mathrm{RT}_{(\mathrm{D})}$ are the μ_{23}/RT values for monomer and dimer respectively, K_d is the dimerization constant and $[F]_{tot}$ is the total folate concentration.

Folate also undergoes a keto-enol tautomerization at the N3-O4 atoms, and can deprotonate at the O4 position at high pH.[18] Thus we studied the effect of pH on the folate μ_{23}/RT value. The data were fit to a pK_a titration, Eq (3.3), adapted from Duff et al. [9]

$$\frac{\mu_{23}}{\mathrm{RT}\,\mathrm{obs}} = \frac{\frac{\mu_{23}}{\mathrm{RT}\,(\mathrm{fh})} - \frac{\mu_{23}}{\mathrm{RT}\,(\mathrm{f})}}{1+\frac{10^{-pKa}}{10^{-p}}} + \frac{\mu_{23}}{\mathrm{RT}\,(\mathrm{f})} \qquad \text{Eq (3.3)}$$

where μ_{23}/RT_{obs} is the observed value while $\mu_{23}/RT_{(fh)}$ and $\mu_{23}/RT_{(f)}$ are the μ_{23}/RT values of protonated and deprotonated folate respectively.

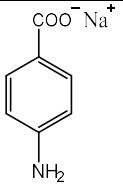
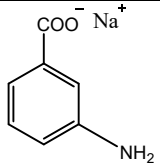
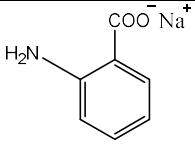
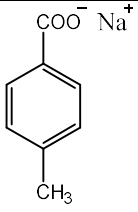
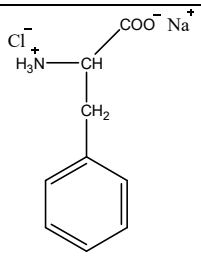
3.3.C. Folate Dimerization at pH 10 by NMR Spectroscopy

A 1D proton NMR experiment was performed as described by Duff et al.[9] to study dimerization of deprotonated folate. Stock solutions of folate were prepared in 10 mM deuterated Tris (pD 10) with and without 20 % deuterated betaine. An NMR sample with 300 mM folate was prepared at pH 10 and the spectrum was recorded. The sample was diluted with buffer and the same procedure was repeated until a folate concentration of 0.5 mM was reached. Spectra were recorded on a Varian 500 MHz NMR spectrometer with a pulse length of 3.7 μs using 16 scans from 14 to −0.5 parts per million (ppm) per spectrum. Data analysis used MestreNova version 10.0 (Mestrelab Research, Compostela, Spain).[19] The spectra were phase and baseline corrected, and the peaks were referenced to the water peak (chemical shift for water, 4.80 ppm). The proton chemical shifts were fit to a dimerization equation as described previously.[9] Similar NMR experiments and analysis were done in the presence of 20 % deuterated betaine.

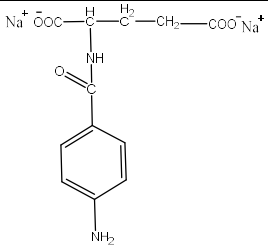
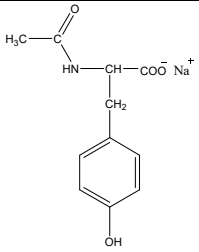
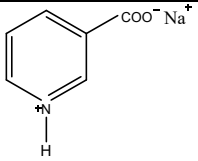
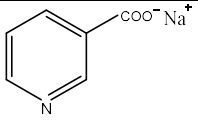
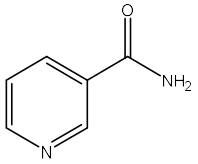
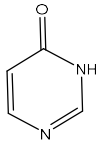
3.3.D. Interaction of Betaine with Heterocyclic Test Compounds

To examine how other small molecules containing aromatic carbons and/or nitrogens interact with betaine, we performed additional VPO studies, mostly at pH 7.0. The test compounds and their structures are shown in Table 3.1.

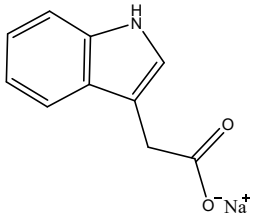
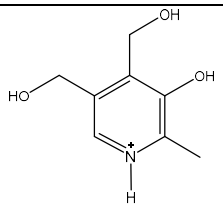
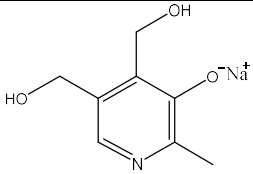
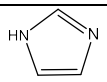
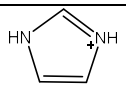
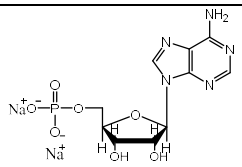
Table 3.1 - A list of compounds tested for preferential interactions with betaine by the VPO method. Structures with correct protonation states are shown along with the source from which each of the structures were obtained.

Compound	Structure	Source
p-Amino-benzoate		MOE[a]
m-Amino-benzoate		MOE[a]
o-Amino-benzoate		MOE[a]
p-Toluic acid		MOE[a]
Phenylalanine-HCl		BMRB[b]

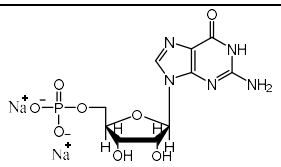
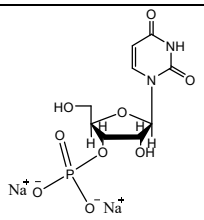
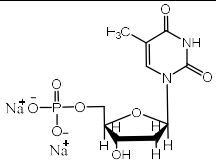
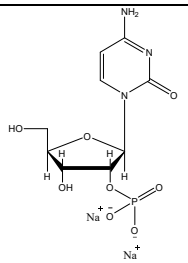
Continued Table 3.1

Compound	Structure	Source
p-Amino-benzoyl-Glutamate		PDB 3DL6[c]
N-acetyl-tyrosine		MOE[a]
Nicotinic acid (Protonated)		BMRB[b]
Nicotinic acid (Deprotonated)		BMRB[b]
Nicotinamide		BMRB[b]
Pyrimidone		MOE[a]

Continued Table 3.1

Compound	Structure	Source
Indole Acetate monosodium salt		BMRB[b]
Pyridoxine- HCl (protonated)		BMRB[b]
Pyridoxine (deprotonated)		BMRB[b]
Imidazole (deprotonated)		MOE[a]
Imidazole (protonated)		MOE[a]
5'AMP monosodium salt		PDB 12AS[c]

Continued Table 3.1

Compound	Structure	Source
5'GMP disodium salt		PDB 5C46[c]
3'UMP disodium salt		PDB 4J7L[c]
5'dTMP disodium salt		BMRB[b]
2'CMP disodium salt		PDB 1ROB[c]

[a] built using MOE (versions 2012.10 and 2015.1001, Chemical Computing Group, Montreal, QC). [b] obtained from the Biological Magnetic Resonance Bank, (http://www.bmrb.wisc.edu/). [c] obtained from the Protein Data Bank (PDB, http://www.rcsb.org)

3.3.E. α Value Calculation by Analysis of μ_{23}/RT Values

The μ_{23}/RT values of test compounds are proposed to be additive contributions from the interaction of betaine with individual functional surface types on the test compounds. Specifically, the contribution of each type of surface to the molecule's μ_{23}/RT value is the product of a chemical interaction potential $(\mu_{23}/RTASA)_i$ and the accessible surface area (ASA) of that surface type *i*. Capp et al. deconvoluted molecular μ_{23}/RT values into surface type μ_{23}/RT values (also called α values) using Eq (3.4).[14]

$$\frac{\mu_{23}}{RT} = \sum_i \left(\frac{\mu_{23}}{RTASA}\right)_i (ASA)_i + v_j\left(\frac{\mu_{23}}{RT}\right)_j \quad \text{Eq (3.4)}$$

where the μ_{23}/RTASA value is the α value, which is the measure of interaction of betaine with 1 Å^2 of surface type *i* on any compound, $(ASA)_i$ is the water accessible surface area in Å^2 of the surface type *i* and $v_j(\mu_{23}/RT)_j$ is the product of the number of salt ions per salt test compound and the assigned contribution $(\mu_{23}/RT)_j$ or the β value, of that type of ion to μ_{23}/RT. The Record lab has calculated α values for many atom types using a β value for the sodium ion of zero.[14]

The structure files of the test compounds were either obtained from the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/) or the Protein Data Bank (PDB, http://www.rcsb.org) or were built in MOE (versions 2012.10 and 2015.1001, Chemical Computing Group, Montreal, QC). Table 3.1 lists the structures and sources of each test compound. Some of the small molecule structures were obtained from the ligand bound protein complex structures in the PDB after deleting the protein and minimizing the ligand in MOE. The water-accessible surface area (ASA) for each atom in the molecule was then

calculated using the SurfaceRacer program.[20] The van der Waal radii from Richards[21] were used as well as a 1.4 Å probe radius for water. Conformational sampling of nucleotides was done using MOE to account for areas from all conformers. There was no significant difference in the average areas for all the conformations when compared to the areas from the minimized structures of each nucleotide. In a multi-linear fit, all experimental μ_{23}/RT values, along with the $(ASA)_i$ information were fit to Eq (3.4) and the α values $(\mu_{23}/RTASA)_i$ were calculated for each surface type. MATLAB (version R2016A) was used for fitting. Errors were calculated using Eq (S17) from Knowles et al.[17]

3.3.F. Solubility Assays

The solubilities of folate in water or in 1 M betaine were determined at pHs 7 and 10 using the method of Liu and Bolen.[22] Folate was weighed in increasing amounts in 10 pre-weighed plastic vials. The range of concentrations was selected so that approximately half of the solutions were unsaturated while the remaining suspensions were saturated. The concentrations ranged from 20 – 500 mM. Solutions were adjusted to the desired pH using 1 N NaOH and the vials were weighed again. The vials were then capped and incubated in the dark in a shaker at 25 °C. After 24 hours, the vials were centrifuged at 4000 RPM for 5 minutes and the supernatant collected. The density of each supernatant was measured using an Anton Paar DMA 35 density meter and plotted against the molality of folate. Solubilities of folate were determined from the density vs. molality plots as described in Auton and Bolen.[23] The apparent free energy of transfer of folate from water to betaine was determined using Eq (3.5).[23]

$$\Delta G^{\circ}_{app,m} = RT\ ln\left(\frac{n_{i,w}}{n_{i,bet}}\right) + RT\ ln\left(\frac{wt_{bet}}{wt_{w}}\right) \quad \text{Eq (3.5)}$$

where ΔG° is the apparent transfer free energy for folate measured on a molal (*m*) scale; $n_{i,w}$ and $n_{i,bet}$ are the number of moles of folate soluble in 1000 g water and in 1000 g of 1M betaine solution respectively; and wt_w and wt_{bet} are the total masses of water and of 1 M betaine solution, respectively.

An acidic pH was also used; however, folate is sparingly soluble at pH 5. Thus only 2 - 100 μM folate (in water) and 2 - 150 μM (in 1M betaine) were used in a similar fashion as described above. After incubation, centrifugation and filtration, the concentration of folate in the supernatant was measured by absorbance at 282 nm. Absorbance was measured upon diluting the samples in MTA buffer, pH 7.0 (ε of folate at 282 nm, pH 7.0, 27000 M^{-1} cm^{-1}).[24] This concentration was plotted against the composition (weight/100 g) of folate for each sample. The solubility of folate and its apparent free energy of transfer were determined as described above.

3.3.G. Protein Purification

R67 DHFR was expressed and purified as described previously.*[25]* Briefly, ammonium sulfate precipitation and ion-exchange column chromatography were used to purify the protein to homogeneity. EcDHFR was expressed and purified as published earlier.[8] His-tagged protein was purified using two affinity chromatography columns - a nickel-NTA column followed by a methotrexate (MTX) affinity column. Elution of EcDHFR from the MTX affinity column required addition of folate, which was subsequently removed with a DEAE column. Purified samples were dialyzed against

distilled, deionized water and then lyophilized. Protein concentrations were determined using a bicinchoninic acid (BCA) (Pierce) assay.

3.3.H. Isothermal Titration Calorimetry (ITC)

Binding affinities, stoichiometries and enthalpies were determined using either a Nano-ITC (TA Instruments) or a VP-ITC (Microcal). For studies with R67 DHFR, binding of folate or pteroyltetra-γ-L-glutamate (PG4 from Schircks Laboratories) to a 1:1 R67 DHFR- NADPH complex was monitored. Titrations were performed in duplicate at 13 °C and pH 8.0 to minimize catalysis. The R67 DHFR concentration was 100- 150 μM and the buffer was MTA pH 8.0. The ligand concentration ranged from 1.2 – 1.4 mM for the experiments with no betaine and from 1.8 mM-1.95 mM for experiments with 5 % and 10 % betaine in the MTA buffer. The time between injections was 240-300 seconds, allowing for baseline equilibration. The software supplied by the manufacturer was initially used for analysis. The data were then exported into SEDPHAT; this program allows global fitting of replicate data sets.[26] A single sites model ($A + B \leftrightarrow AB$) was used for the fitting process. Similar experiments were performed with binding of folate and PG4 to EcDHFR in MTA buffer pH 7.0 at 25 °C. EcDHFR concentrations ranged from 10-15 μM. Folate and PG4 concentrations ranged from 350 – 550 μM for titrations in the absence of betaine. The folate concentration for binding to EcDHFR in presence of 10 % and 20 % betaine was in the range of 600-850 μM. The "c value" ($= [P_{total}] / K_d$) ranged from 1-10, within the suggested values of 1-1000.[27]

3.3.I. Simulation of Folate in Water

Computer simulations of folate in water were performed using the AMBER simulations package.[28] For system preparation, a single folate molecule was placed in

the center of a periodic box surrounded by water (SPC/E water model)[29] such that the boundary of the box was at least 10 Å away from the edges of folate molecule. AMBER's *parm 14SB* force-field was used, and the folate molecule was parameterized using the procedure outlined in the AMBER manual. The charges for folate atoms were calculated using electronic structure calculations at the Restricted Hartree-Fock 6-31G** level of theory. The prepared system was slowly equilibrated as previously described.[30] The production run of 200 nanoseconds was performed at 300 K in an NVE ensemble using 2 femtosecond time-steps. A total of 200 conformations (every 1 ns) were used for analysis.

3.3.J. Simulation of Folate in Betaine

Computer simulations of folate surrounded by betaine and water were performed using the AMBER simulations package.[28] A folate molecule was placed in the center of a periodic box surrounded by betaine and water (SPC/E water model). Betaine, or tri-methyl-glycine, was modeled using AMBER's *parm 14SB*; the aliphatic carbons, hydrogens and nitrogen were parameterized using the lipid related parameters while the remaining atoms were parameterized based on glycine.[31] The charges for betaine were calculated using a procedure similar to that used for the folate molecule. The ratio of folate to betaine molecules was 1:76, corresponding to a 1.35 M concentration of betaine in a periodic box of 46.74 Å x 49.64 Å x 50.12 Å. The initial placement of betaine around folate was performed using PackMol software [4], followed by filling the remaining space with SPC/E water using AMBER's xleap module. The prepared systems were slowly equilibrated using a procedure developed in our group, and described previously.[30] The production run of 200 nanoseconds was performed for each system at 300 K in an NVE ensemble using a 2 femtosecond time-step. A total of 200 conformations (every 1 ns) were

used for analysis.

3.3.K. Comparison of $\Delta\mu_{23}$/RT Predictions of Ligand Binding with ITC Data

To determine how accurately the predicted μ_{23}/RT values reflect experimental data, $\Delta\mu_{23}$/RT values were calculated for ligands binding to the two DHFR enzyme types. Using data previously obtained by ITC, the $\Delta\mu_{23}$/RT for binding can be calculated from the slopes of ln(K_a) versus molality using Eq (3.1) from Guinn *et al*.[15]

$$-\frac{lnK_a}{m_3} = \frac{\Delta\mu_{23}}{RT} \qquad \text{Eq (3.6)}$$

where K_a is the association constant and m_3 is molality of betaine. Predicted μ_{23}/RT values were calculated for the apo-proteins or protein-ligand complexes using Eq (3.4) with a Python script. Waters were removed from the PDB file and the surface areas of each of the atom types were calculated using SurfaceRacer.[20] The product of the atomic surface areas and the corresponding atom-type α value were summed to obtain the predicted μ_{23}/RT. The μ_{23}/RT values for the ligands in their bound conformation were calculated in a similar manner. To calculate the $\Delta\mu_{23}$/RT for the binary protein-ligand complexes, the sum of the μ_{23}/RTs for the apo-protein and the unbound ligand was subtracted from the μ_{23}/RT of the complex. To obtain the $\Delta\mu_{23}$/RT for ternary complexes, the μ_{23}/RT of the of the binary protein-ligand complex plus the unbound ligand was subtracted from the μ_{23}/RT value of the ternary complex.

3.4 Results

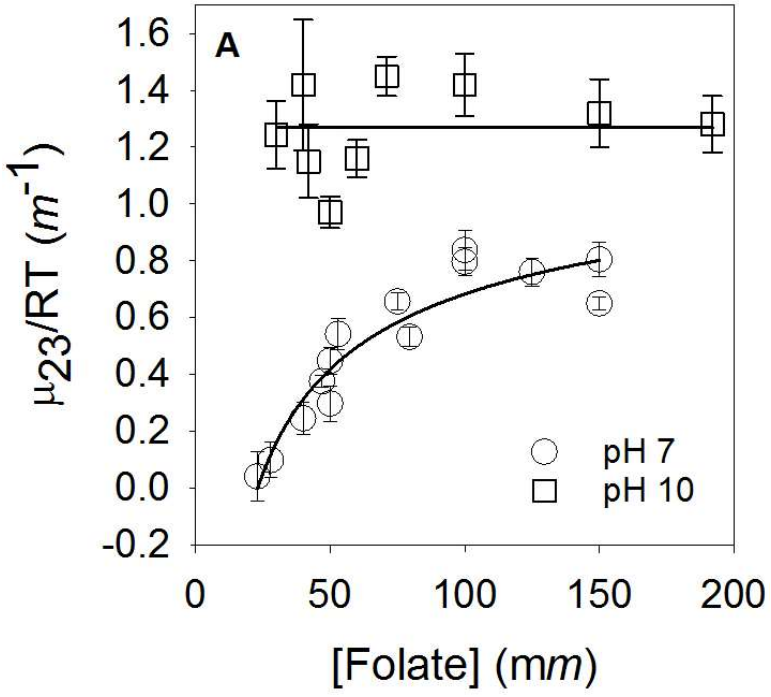
3.4.A. VPO Measurements of Folate at pH 7

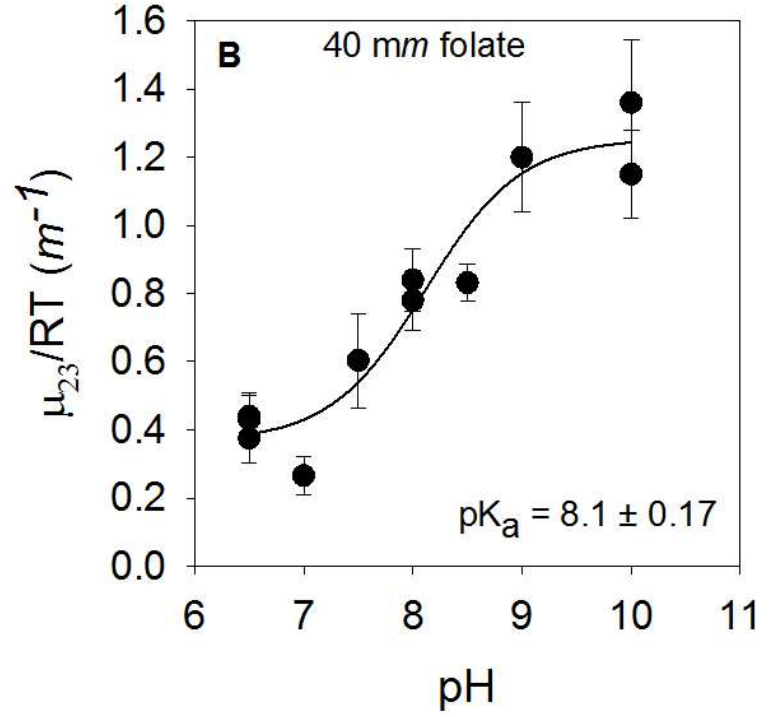
Figure 3.3.A shows the concentration dependence of folate interaction potentials (μ_{23}/RT) with betaine measured by VPO experiments at pH 7. A value near zero indicates similar interaction preferences of folate for water and betaine. A positive value predicts an interaction preference for water while a negative value indicates a preference for betaine. We observed an increase in the folate μ_{23}/RT values from 0.04 ± 0.09 m^{-1} (at 23 m*m*) to 0.80 ± 0.06 m^{-1} (150 m*m*). These observed μ_{23}/RT values indicate that at low concentrations, folate interacts with both water and betaine while at higher concentrations, folate favors water.

The concentration dependence in Figure 3.3.A is consistent with previous observations of folate dimerization, which occurs in a head-to-tail fashion such that each pterin ring stacks with the *p*-ABA ring of the other monomer and the glutamate tails are free to rotate.[18] Previously Capp et al. have found betaine interacts with aromatic carbons, amide nitrogens and cationic nitrogens and is excluded from aliphatic carbons, hydroxyl oxygens, amide oxygens, carboxylate oxygens and phosphate oxygens.[14] Thus the observed increase in μ_{23}/RT values at high folate concentrations is consistent with decreased accessible surface area for the aromatic ring surfaces due to ring stacking.

Unfortunately, we obtained a poor fit when the concentration dependent data were fit to Eq (3.2) describing dimerization. The poor fit may be due to not having a good lower limit for the μ_{23}/RT of monomeric folate (due to poor signal to noise levels at low folate concentration) as well as the variable effects of different folate and betaine concentrations

Figure 3.3 - Preferential interactions between folate and betaine show folate concentration and pH effects. Panel A shows the folate concentration dependence of μ_{23}/RT at pH 7 (○) and at pH 10 (□). A fit to Eq (3.2) describing folate dimerization was poor for data at pH 7 and no concentration dependence was noted at pH 10, thus the lines provided are to aid the eye. Panel B shows the pH dependence of μ_{23}/ RT values for 40 m*m* folate (●). The data were fit to Eq (3.3) to yield a pK_a describing deprotonation of the N3-O4 enol tautomer.


A
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
μ_{23}/RT (m^{-1})
pH 7
pH 10
0
50
100
150
200
[Folate] (m*m*)


B
40 m*m* folate
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
μ_{23}/RT (m^{-1})
pK_a = 8.1 ± 0.17
6
7
8
9
10
11
pH

associated with each point on the plot. As betaine addition alters the K_d describing folate dimerization,[9] there may be additional effects contributing to the titration observed in Figure 3.3.A. Another contributor may be the possible formation of higher oligomerization states.[32]

3.4.B. VPO Measurements of Folate at pH 10

Using an NMR approach, Poe found folate dimerization is pH dependent.[18] The N3-O4 atoms in the pterin ring undergo a keto-enol tautomerization. Deprotonation of the enol (pK_a ~8) results in a negatively charged O4 atom. The dimerization constant for neutral folate is 20 mM while the value for basic folate is 340 mM.[18] To potentially determine a μ_{23}/RT value for monomeric folate, we repeated the VPO study at pH 10. We measured the μ_{23}/RT values of folate at concentrations ranging from 30 m*m* to 190 m*m*. The average μ_{23}/RT at this pH was 1.27 ± 0.36 m^{-1}, which indicates strong exclusion of betaine from the anionic folate surface. No concentration dependence of μ_{23}/RT values was observed (see Figure 3.3.A), consistent with folate being monomeric at pH 10. This observation is also consistent with our 1D H-NMR experiments performed at pH 10 (see below). A high μ_{23}/RT value for anionic, monomeric folate is surprising, however quantum mechanical calculations by Soniat et al. on anionic pterin report delocalization of the negative charge on the ring.[33] This view supports the studies of Felitsky et al.[34] who found betaine was strongly excluded from anionic surfaces.

Due to the large difference in μ_{23}/RT values for neutral and anionic folate, we monitored preferential interaction coefficients for 40 m*m* folate from pH 6.5 to 10. While dimers are likely present at neutral pH at this concentration, the data display higher signal to noise levels and possess lower errors. Figure 3.3.B shows a plot of μ_{23}/RT values vs pH.

The data were fit to Eq (3.3), yielding a pK_a of 8.1 ± 0.17. This value is similar to values of 7.94 and 8.38 obtained by NMR[9, 35] and 7.98 by capillary electrophoresis studies.[36] The fit also yields μ_{23}/RT values of 0.36 ± 0.06 m^{-1} and 1.25 ± 0.07 m^{-1} for the neutral (protonated) and basic (deprotonated) forms respectively. Our data indicate that VPO experiments can be used to monitor pK_a values if the protonated and deprotonated species possess different μ_{23}/RT values.

3.4.C. Folate Dimerization at pH 10

NMR experiments at pH 10 noted the change in chemical shifts for the pteridine (C7H), C9H, benzoyl ring protons (C2'H/C6'H, C3'H/C5'H) with increasing folate concentration while the glutamate proton shifts were unchanged (See Figure 3.2 for numbering of atoms). The data and fits for each of the proton chemical shifts are shown in Figure 3.4. Fitting the sum of the C7, C9, C3′/C5′, and C2′/C6 proton chemical shifts with no betaine to a dimerization equation yielded a K_d of 960 ± 140 mM for folate at pH 10 (Table 3.2). As the K_d was higher than the highest folate concentration used for the experiment, it suggests that folate dimerization at pH 10 is very weak and its K_d cannot be accurately determined. Higher concentrations of folate cannot be achieved because of limited solubility. Although the K_ds obtained were much higher than the folate concentrations used for the experiments, we can qualitatively see a trend for a lower K_d in the presence of betaine.

3.4.D. VPO Measurements of Non-Heterocyclic Aromatic Compounds

To extend the list of aromatic compounds used to predict α values for aromatic carbons, the μ_{23}/RT values for *p*-amino-benzoate, *m*-amino-benzoate, *o*-amino-benzoate, *p*-amino-

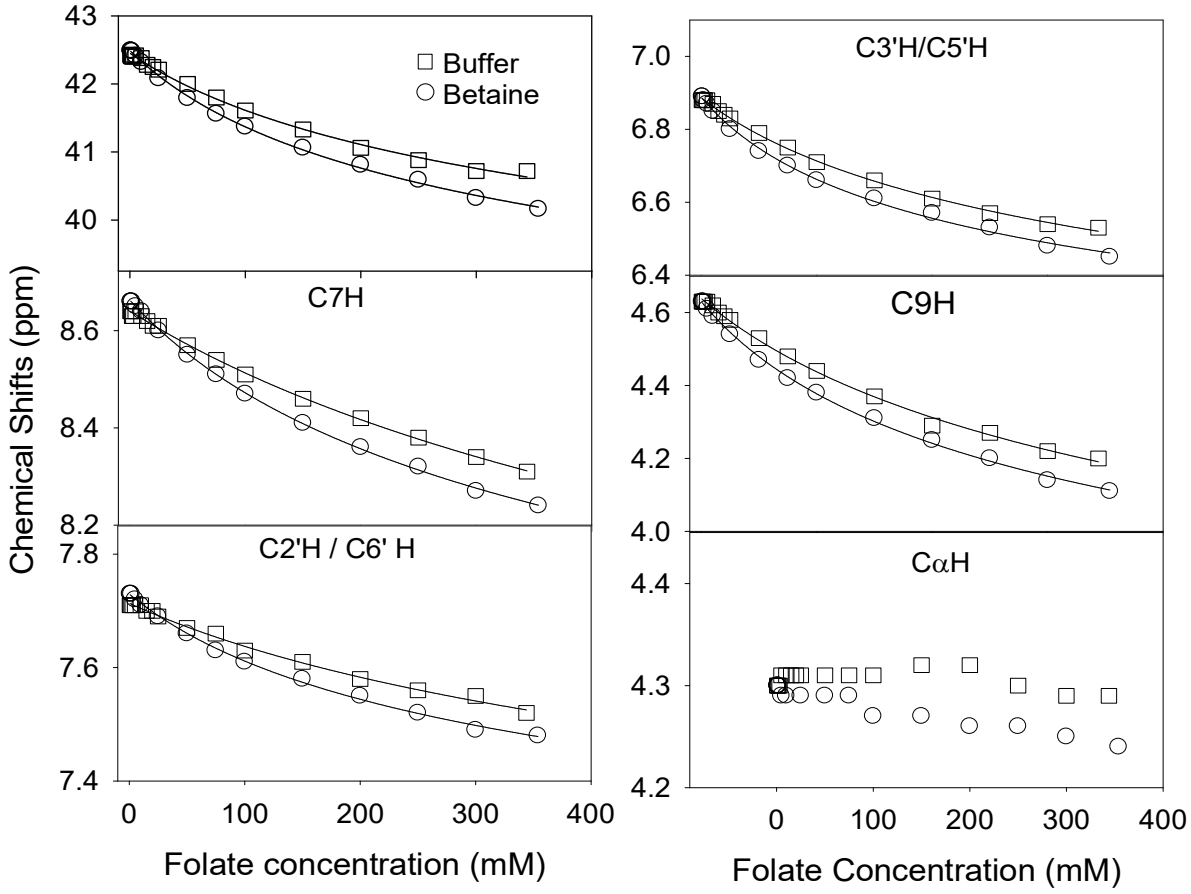


Figure 3.4 - Folate dimerization at pH 10. The panels show chemical shifts noted for folate protons (numbered as in Figure 3.2) in 10 mM deuterated Tris (□), 10 mM deuterated Tris with 20 % deuterated betaine (○). The lines are the fits to the dimerization equation in Duff et al.[9]

Table 3.2 - Folate dimerization constants at pH 10. Values obtained from fitting the concentration dependence of chemical shifts obtained by NMR as described in Duff et al. [9] Data and fits shown in Figure 3.4.

Chemical shifts	**No betaine**	**20 % betaine**
	K_d (mM)	**K_d (mM)**
Sum of all proton chemical shifts	960 ± 140	710 ± 60
C7H	2200 ± 260	980 ± 60
C2'H	1800 ± 370	830 ± 90
C6'H	1500 ± 440	900 ± 90
C3'H	860 ± 80	510 ± 60
C5'H	880 ± 90	590 ± 80
C9H	1100 ± 140	700 ± 60

benzoyl-glutamate (*p*-ABA-Glu), *p*-toluic acid, quinolinic acid, phenylalanine-HCl and N-acetyl-tyrosine were measured. The experimental μ_{23}/RT values for the amino-benzoates and phenylalanine are listed in Table 3.3. These values for the rest of the aromatic compounds are listed in Table 3.4. All amino benzoates and phenylalanine slopes are negative (Figure 3.5.A), consistent with aromatic carbon atoms preferring to interact with betaine compared to water. For the *o-, m- and p*-amino-benzoate series, the μ_{23}/RT values were within error of each other, suggesting the relative ring position of the substituents does not have a large effect.

3.4.E. VPO Measurements of Compounds Containing Aromatic Nitrogen Atoms

As folate contains aromatic nitrogen atoms and its μ_{23}/RT value showed pH effects, we were interested in studying interactions of betaine with compounds containing titratable aromatic nitrogens. Compounds for this study were chosen based on their solubility, lack of dimerization and pK_a values. Table 3.1 gives the structures of the compounds while Figures 3.5.B and 3.5.C show the experimental VPO data. The measured μ_{23}/RT values are listed in Tables 3.3 and 3.4. The below sections provide more detail on the pH effects observed in a few of our studies.

3.4.F. Pyridoxine-HCl

As pyridoxine possesses titrations in the physiological pH range,[37] we measured μ_{23}/RT values for pyridoxine from pH 2 to 12. Figure 3.6.A shows the slopes (μ_{23}/RT) for protonated and deprotonated pyridoxine. The μ_{23}/RT values at lower pHs are slightly negative while at higher pH, the values are positive. The pH dependence of the μ_{23}/RT values is shown in Figure 3.6.B. The data were fit to Eq (3.3) and a pK_a of 5.98 ± 0.25 was obtained. The upper and lower limits for the μ_{23}/RT values were 0.26 ± 0.02 m^{-1} and 0.017

Figure 3.5 - Quantification of preferential interactions of betaine with test compounds. The panels show the raw data plots of Δ Osm vs. the product of molal concentration of test compound and betaine obtained from VPO experiments. Panel A shows the data for the non-heterocyclic aromatic compounds. Panels B and C show data for heterocyclic (nitrogen containing) aromatic compounds. Panel B data include nicotinamide, pyrrole-2-carboxylate, guanosine 5'-phosphate (5'GMP), cytosine 2'-phosphate (2'CMP) and uridine 3'-phosphate (3'UMP). These are compounds with lower solubilities and therefore span shorter concentration ranges. Panel C shows plots for adenosine 5'-phosphate (5'AMP), deoxythymidine 5'-phosphate (5'dTMP), pyrimidone and indole acetate, which have higher solubilities.

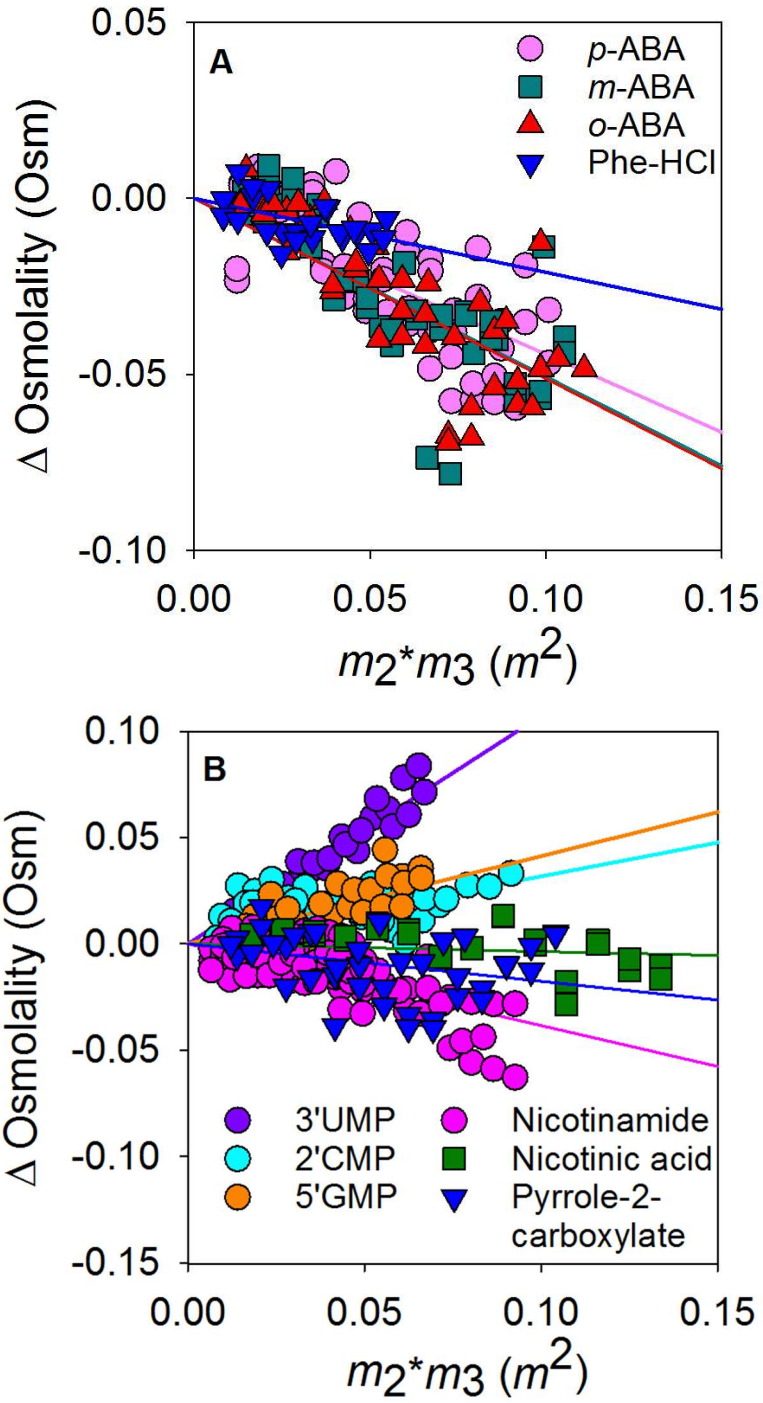

A
p-ABA
m-ABA
o-ABA
Phe-HCl
Δ Osmolality (Osm)
0.05
0.00
-0.05
-0.10
0.00
0.05
0.10
0.15
m2*m3 (m2)
B
3'UMP
2'CMP
5'GMP
Nicotinamide
Nicotinic acid
Pyrrole-2-carboxylate
Δ Osmolality (Osm)
0.10
0.05
0.00
-0.05
-0.10
-0.15
0.00
0.05
0.10
0.15
m2*m3 (m2)

Figure 3.5 continued

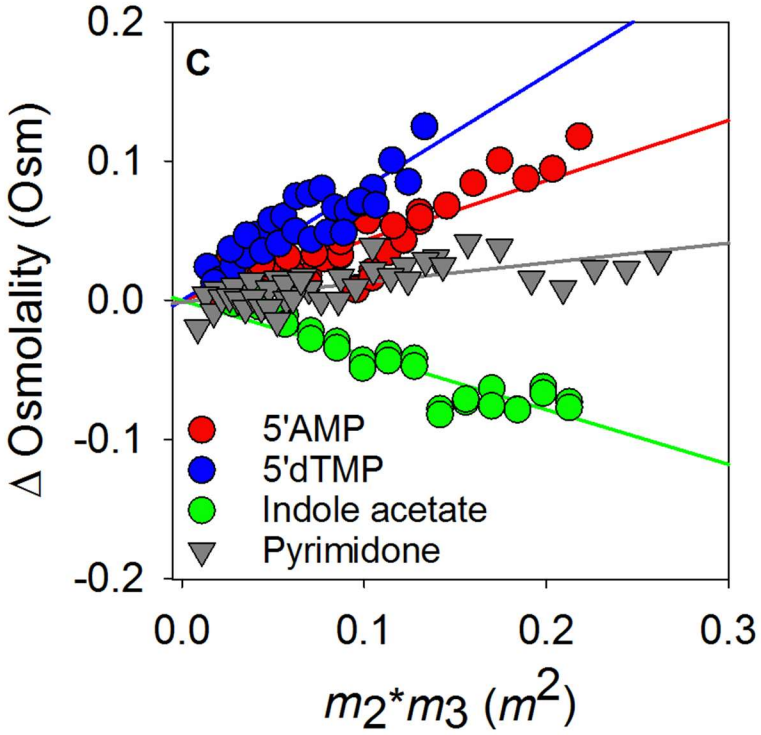


Table 3.3 - A list of all test compounds with their experimental and predicted μ_{23}/RT values. The predicted μ_{23}/RT values were obtained using α values in Table 3.5 and Eq (3.4). The pH at which each compound was tested is also supplied.

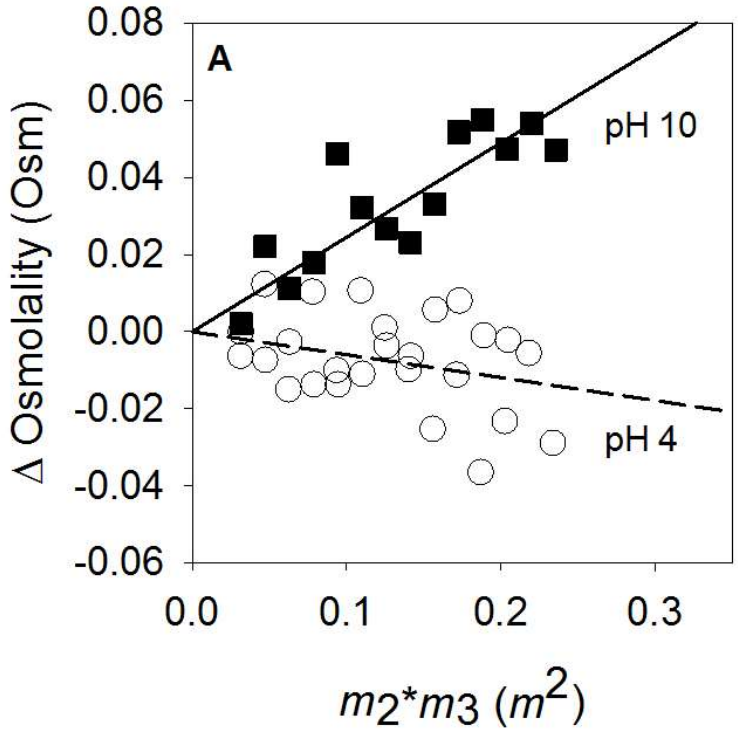
Compound	**Experimental μ_{23}/RT (m^{-1})**	**Predicted μ_{23}/RT (m^{-1})**	**pH**
Non-Heterocyclic Aromatic			
p-Amino-benzoate	-0.44 ± 0.03	-0.46 ± 0.02	7
m-Amino-benzoate	-0.50 ± 0.03	-0.46 ± 0.02	7
o-Amino-benzoate	-0.51 ± 0.03	-0.46 ± 0.02	7
Phenylalanine-HCl	-0.21 ± 0.03	-0.24 ± 0.02	5
Heterocyclic Aromatic			
Nicotinamide	-0.38 ± 0.03	-0.27 ± 0.02	Unadjusted (6.3)
Pyrimidone	0.13 ± 0.01	0.08 ± 0.03	Unadjusted (5)
Indole acetate monosodium	-0.39 ± 0.02	-0.20 ± 0.02	7
Pyrrole-2-carboxylate	-0.18 ± 0.04	-0.12 ± 0.02	7
5'AMP disodium	0.33 ± 0.03	0.18 ± 0.04	7
5' GMP disodium	0. 41 ± 0.04	0.46 ± 0.04	8.1
3' UMP disodium	1. 07 ± 0.03	0.9 ± 0.04	7.6
5' dTMP disodium	0.81 ± 0.03	1.00 ± 0.04	8
2' CMP disodium	0.32 ± 0.04	0.34 ± 0.04	7
pH dependent Heterocyclic Aromatic			
Pyridoxine-HCl	0.02 ± 0.02	0.02 ± 0.02[a]	2
Pyridoxine- HCl	0.01 ± 0.02	-	Unadjusted (2.6)

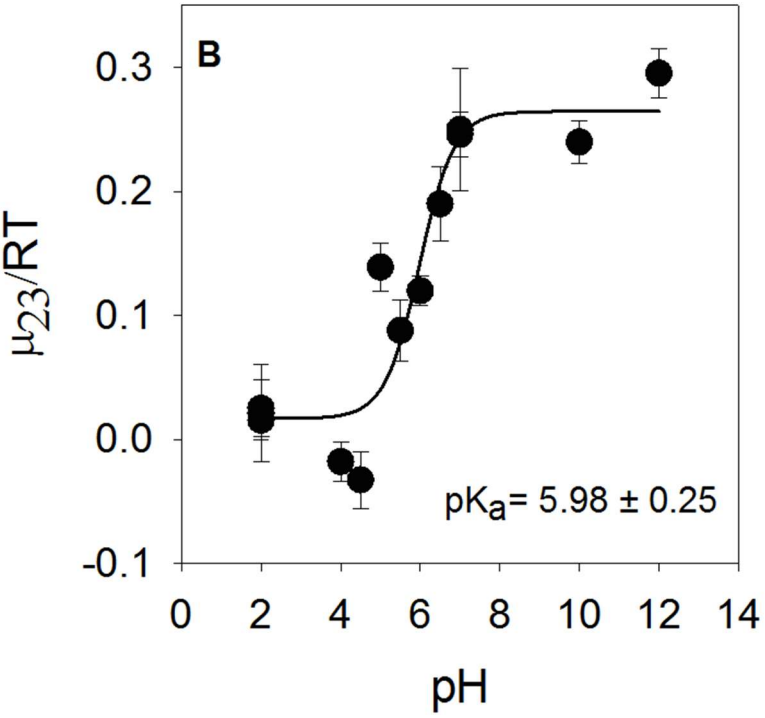
Continued Table 3.3

Compound	**Experimental** μ_{23}**/RT (**m^{-1}**)**	**Predicted** μ_{23}**/RT (**m^{-1}**)**	**pH**
Pyridoxine	0.25 ± 0.02	0.06 ± 0.02 0.26 ± 0.02[a]	10
Nicotinic acid	-0.27 ± 0.03	-	Unadjusted (3.5)
Nicotinic acid	-0.04 ± 0.02	-0.04 ± 0.02	7

[a]Values obtained from the fit limits of the pH titration data for pyridoxine-HCl using Eq (3.3)

Figure 3.6 - Vapor pressure osmometry studies of pyridoxine show pH effects. Panel A, data for pyridoxine at pHs 4 (○) and 10 (■). The dashed and solid lines represent the slopes of the plots for pH 4 and 10 data respectively. Panel B, pH titration of μ_{23}/RT for pyridoxine. Data were fit to Eq (3.3) and best fit values are 0.017 ± 0.018 m^{-1} for the protonated form and 0.26 ± 0.02 m^{-1} for the deprotonated form.


A
pH 10
pH 4
Δ Osmolality (Osm)
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
0.0
0.1
0.2
0.3
m2*m3 (m2)


B
μ23/RT
0.3
0.2
0.1
0.0
-0.1
0
2
4
6
8
10
12
14
pH
pKa= 5.98 ± 0.25

± 0.018 m^{-1} respectively. These results indicate that the protonated form of pyridoxine interacts more strongly with betaine than the deprotonated form.

3.4.G. Betaine-Nicotinic Acid Betaine-Imidazole and Interaction by VPO

Nicotinic acid (vitamin B3) is an aromatic heterocyclic compound with nitrogen in a six-membered ring. VPO experiments found this compound possessed a slightly negative preferential interaction potential as seen in Figure 3.5.B. The μ_{23}/RT value for nicotinic acid was observed to change with pH, consistent with titration of the aromatic nitrogen, which has previously been observed to have a pK_a of 4.9.[38] The acidic form of nicotinic acid at pH 3 yielded a more negative μ_{23}/RT value, indicating a stronger preference for interaction with betaine than the deprotonated form at pH 7.

Imidazole, a small molecule, has a pK_a of 6.5;[39] it also dimerizes with a K_d of 1 mM for the protonated form and a K_d of 33 μM for the deprotonated form.[40] VPO studies of imidazole at pH 4 showed a slightly negative μ_{23}/RT value, whereas at pH 10, its μ_{23}/RT was near zero and had large errors. Figure 3.7 shows the data for 250-270 m*m* of imidazole. As the change in μ_{23}/RT values for protonated and deprotonated imidazole was not large, scatter was observed in the data, and dimerization was a concern, we did not analyze these data.

3.4.H. Analysis of μ_{23}/RT Values and Calculation of α Values

To deconvolute which atoms of folate are involved in the interactions with betaine, we use the α value analysis developed by the Record lab.[14-17] This approach uses multiple linear regressions (based on the number of compounds used), which describe all the surface types present in the molecules. We added our 15 compounds to the list of 27

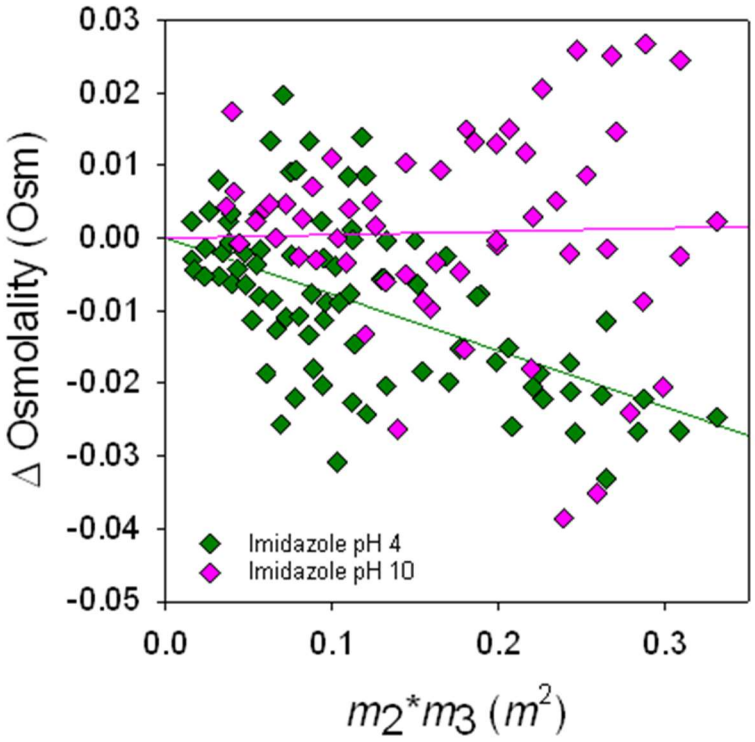


Figure 3.7 - Vapor pressure osmometry studies of imidazole. Data for imidazole at pH 4 (green diamonds) and pH 10 (magenta diamonds). The concentration of imidazole used in the experiment was 250 m*m*.

Table 3.4 - A list of experimental and predicted μ23/RT values for test compounds in addition to compounds listed in Table 3.3. These compounds were not used in our fits to obtain α values.

Compound	**Experimental μ_{23}/RT (m^{-1})**	**Predicted μ_{23}/RT (m^{-1})**	**pH**	**Reason for not including in α value fit**
p-Toluic acid	-0.46 ± 0.04	-0.21 ± 0.03	7	Outlier on plot of experimental vs. predicted μ_{23}/RT values
p-ABA-Glu	0.50 ± 0.02	-0.12 ± 0.02	7	Changed R^2 of fit from 0.93 to 0.86
N-acetyl-tyrosine	-0.66 ± 0.03	-0.05 ± 0.02	4	Low solubility → low concentration
Quinolinic acid	0.86 ± 0.04	0.20 ± 0.01	7	Low solubility → low concentration
Imidazole	-0.07 ± 0.01	-	4	Dimerization
Imidazole	0.01 ± 0.01	-	10	Dimerization

molecules published by Capp et al.[14] As our model compounds were mostly nitrogen containing aromatic heterocycles, we aimed at calculating α values for aromatic N surface types in addition to the surface types analyzed by the Record lab.[14-17] Since several atom types that appeared in our molecules were not included in the Record lab study, we added an amine N off aromatic rings to our atom types.

Aromatic ring systems are complicated. In our fittings, we considered other atom types in the AMBER (ff14SB) force field that describe different aromatic carbons and nitrogens.[41] Some considerations on whether to include atom types were whether its ASA was significant and whether the amplitudes for the related atom type in our fits overlapped and/or whether the error was low. While we tried many combinations, ultimately, we just added an aromatic nitrogen and an amine nitrogen off an aromatic ring to the list of atom type as too many variables can affect error analysis.

All compounds were included in our fit except *p*-ABA-Glu, N-acetyl-tyrosine, imidazole, quinolinic acid, and the acidic forms of pyridoxine and nicotinic acid. We did not include imidazole as it dimerizes at the concentrations needed to obtain a VPO signal. As only 2 compounds with protonated aromatic nitrogens were available (acidic pyridoxine and acidic nicotinic acid), we were concerned with the ability of only 2 atoms to provide good statistics for this atom type. Addition of *p*-ABA-Glu, N-acetyl-tyrosine, *p*-toluic acid and quinolinic acid significantly caused the R^2 of our fit to drop, from 0.93 to 0.8 (with all compounds added). For N-acetyl-tyrosine and quinolinic acid, this is likely due to their low solubilities which necessitated using low concentrations, a potential source of error. It is not clear why *p*-ABA-Glu and *p*-toluic acid were outliers in our fit. Perhaps mixed effects from the different electron donating and withdrawing groups off the aromatic rings

play a role.

Our α values are listed in Table 3.5 along with those from the Record lab. While our α values are different in magnitude from those from the Record lab, the overall trend is the same. Both this study and Guinn et al.[15] obtain positive α values for oxygens in hydroxyl, amide, carboxylate, and phosphate groups. We add the information that aromatic nitrogens display positive α values. The Record group and our present study find that amide nitrogens show a negative α value. We add that amine nitrogens off aromatic rings do as well. Finally, while the Record group had a positive α value for aliphatic carbon ($3 \pm 3\times10^{-4}\ m^{-1}Å^{-2}$), the addition of our compounds tip the balance towards a small negative value. On the other hand, we obtain a positive β value for Cl^- ($8 \pm 1\times10^{-2}\ m^{-1}$), which is outside the range ($-4 \pm 4\times10^{-2}\ m^{-1}$) of Guinn et al.[15]

Finally, we note Diehl et al.[16] compared their proline VPO results with those from solubility or group transfer free energy (GTFE) assays. While the preferences of many amino acids to interact with betaine vs water were similar for the two techniques, they also found significant differences. For example, solubility assays noted a weak preference of valine and leucine for betaine compared to water while the VPO results indicated a weak preference of valine for water. In another difference, GTFE experiments found sodium salts of glutamate and aspartate strongly prefer to interact with betaine while VPO results indicate a strong preference for water. These differences suggest that while our α values are somewhat different than those from Capp et al.[14] and Guinn et al.[15], this variability is not surprising, given that the compounds used in the analysis are different and that GTFE assays can show somewhat different patterns. On the other side of the coin, the α values reflect the compounds used in the calculations. Accordingly, our fits likely converge to

Table 3.5 - A comparison of α and β values from this study vs. those from Guinn et al.[15] Calculations used Eq (3.4). Data for 15 compounds from this study were used in addition to 27 compounds from Capp et al.[14] α values for an amine N off an aromatic ring and an aromatic N atom types were calculated in addition to the atoms types in Guinn et. al.[15]

α values from this study		**α values from Guinn et al.[15]**	
Surface type, *i*	**$10^4\alpha_i$, $m^{-1}Å^{-2}$**	**Surface type, *i***	**$10^4\alpha_i$, $m^{-1}Å^{-2}$**
Aliphatic C	-3 ± 1	Aliphatic C	3 ± 3
Hydroxyl O	7 ± 1	Hydroxyl O	1 ± 2
Amide O	49 ± 3	Amide O	28 ± 10
Amide N	-33 ± 2	Amide N	-20 ± 7
Carboxylate O	28 ± 1	Carboxylate O	29 ± 2
Cationic N	-14 ± 1	Cationic N	-12 ± 4
Aromatic C	-31 ± 1	Aromatic C	-23 ± 4
Phosphate O	48 ± 2	Phosphate O	49 ± 4
Amine N off aromatic rings	-53 ± 3	Amine N off aromatic rings	-
Aromatic N	27 ± 3	Aromatic N	-
Inorganic ion	$10^2\beta_{ion}$, m^{-1}	Inorganic ion	$10^2\beta_{ion}$, m^{-1}
K^+	8 ± 2	K^+	5 ± 2
Cl^-	7 ± 1	Cl^-	-4 ± 4

somewhat different values due to the extra information provided by the additional compounds used in our experiments.

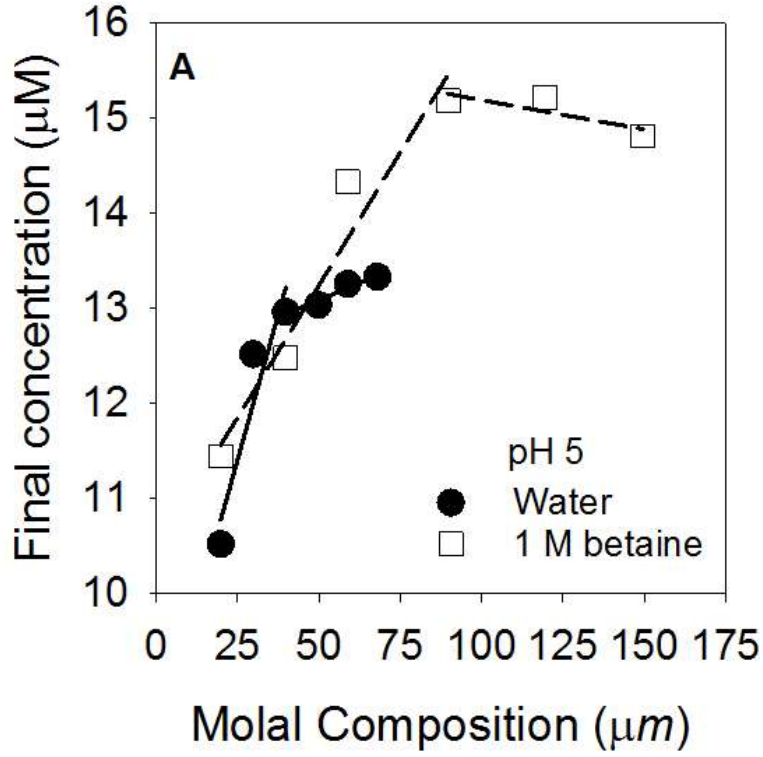
3.4.I Solubility Assays

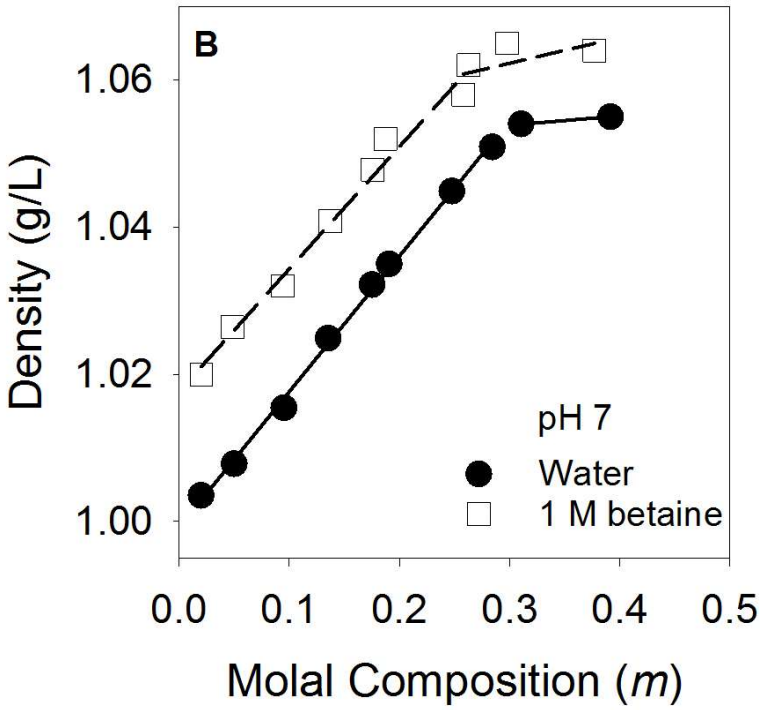
A second approach to investigate how betaine interacts with folate uses a solubility assay. Thus we measured the solubility of folate in water vs. 1M betaine at various pH values. Figure 3.8 shows the data, which were analyzed as per the Bolen lab.[22, 23] The composition vs. density plots for pHs 7 and 10 were each fit to two lines as shown. The intersection of the lines provided the concentration at which the solution was saturated with folate in either water or 1 M betaine. At pH 5, the solubility of folate in 1 M betaine was higher than in water (Figure 3.8.A). The transfer free energy of folate was calculated as -297 ± 22 cal/mol where a negative free energy indicates a preference for the betaine solution over water. At pH 7, folate is almost equally soluble in water and 1 M betaine (Figure 3.8.B). The transfer free energy from water to betaine was found to be 89 ± 30 cal/mol. The data at pH 10 (Figure 3.8.C) indicate that folate is more soluble in water than in betaine with a transfer free energy of 500 ± 150 cal/mol. These solubility assays indicate folate prefers to interact with betaine compared to water in the lower pH range. In contrast, folate prefers to interact with water over betaine as the pH increases and the deprotonated enol tautomer of folate predominates. The general trend observed in the solubility and VPO experiments is the same. We also note that depending on the pH, dimer K_d and the folate concentration, monomer and/or dimer species may be present.

3.4.J Binding of Folate and PG4 to R67 DHFR and EcDHFR

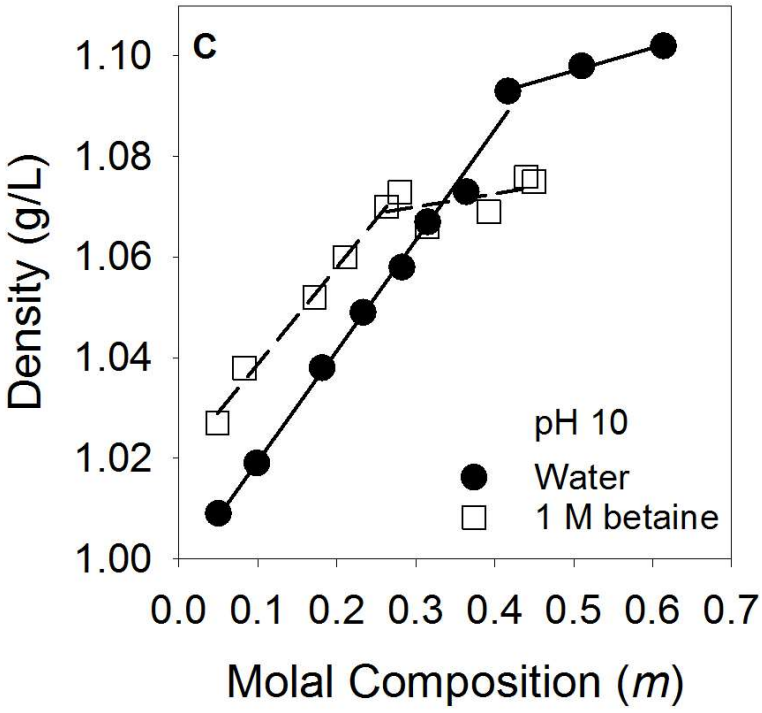
α values can be used to predict ligand-osmolyte interactions. Can this information be used to predict effects on ligand binding to proteins? The caveat is whether all the

Figure 3.8 - pH dependence of folate solubility in 1 M betaine and water. The data for solubility in 1M betaine and water are shown as (□) and (●) respectively. Panel A plots the folate concentration measured by absorbance vs. the folate composition at pH 5. Panels B and C plot the solution density vs. the molal composition at pHs 7 and 10 respectively. The data were fit to two solid lines for water and two dashed lines for betaine. The intersection of the lines for each solution condition gave the saturation concentration of folate. The transfer free energies at pHs 5, 7, and 10 are -297 ± 22 cal/mol, 89 ± 30 cal/mol and 500 ± 150 cal/mol, respectively.


A
Final concentration (μM)
16
15
14
13
12
11
10
0
25
50
75
100
125
150
175
Molal Composition (μm)
pH 5
Water
1 M betaine


B
Density (g/L)
1.06
1.04
1.02
1.00
0.0
0.1
0.2
0.3
0.4
0.5
Molal Composition (m)
pH 7
Water
1 M betaine

Continued Figure 3.8



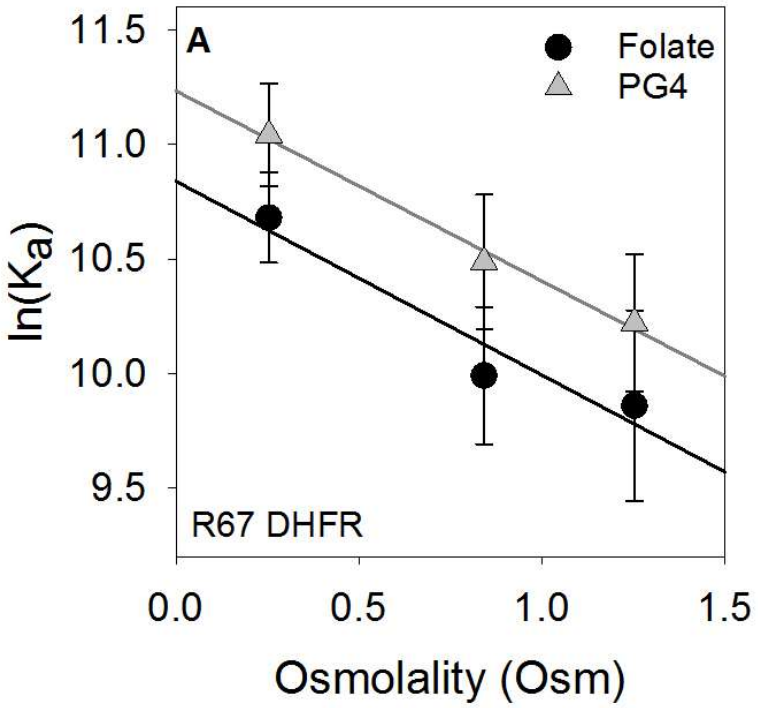
surfaces of the ligand are used in the binding interaction. For example, we consider the case of folate polyglutamylation. As glutamate excludes betaine, addition of extra glutamates to folate (extended conformer) increases the predicted μ_{23}/RT value from -0.09 ± 0.04 m^{-1} to 1.22 ± 0.04 m^{-1} for pteroyltetra-γ-L-glutamate (PG4). If the polyglutamate tail is involved in binding to DHFR, this increase in μ_{23}/RT predicts lesser osmotic stress effects. However, our ITC experiments found betaine addition weakens binding of folate or PG4 to R67 DHFR (see Figure 3.9.A and Table 3.6). Thus use of a calculated μ_{23}/RT value for a ligand is not sufficient to predict effects of betaine on binding.

While many folate pathway enzymes show tighter binding to polyglutamylated folate redox states, we did not find any information addressing this issue in EcDHFR. Thus we measured the affinity for folate to EcDHFR and found it also decreased linearly with increasing betaine concentration. The affinity for PG4 binding to EcDHFR was similar to that of folate (see Figure 3.9.B). These results predict that the additional glutamates will not contribute to binding to EcDHFR. Thus an important parameter in predicting betaine effects on binding is whether all the ligand atoms are used in the interaction.

3.4.K Prediction of Folate μ_{23}/RT Values from Simulation Data

In the MD simulation of folate in water, folate adopted a range of conformations. For each of these conformations, a μ_{23}/RT was calculated from α values. Similar calculations were also performed for the simulations of folate in water and 1.35 M betaine. A relatively large variation in the predicted μ_{23}/RT values for folate was noted over the course of both simulations (see Figures 3.10.A and 3.11). The average μ_{23}/RT value for folate in water was -0.03 ± 0.05 m^{-1} while the value for folate in 1.35 M betaine was -0.05

Figure 3.9 - The effect of osmolality on the binding affinities of folate (●) and PG4 (▲) to DHFRs. Panel A plots ln K_a vs. osmolality for folate and PG4 binding to R67 DHFR-NADPH. The negative slopes indicate weaker binding of both folate and PG4 in the presence of betaine. Panel B shows similar results obtained for EcDHFR. No significant difference can be noted in the binding affinities of folate and PG4 to EcDHFR, predicting no effect of polyglutamylation on binding.


A
Folate
PG4
11.5
11.0
10.5
10.0
9.5
ln(Ka)
R67 DHFR
0.0
0.5
1.0
1.5
Osmolality (Osm)

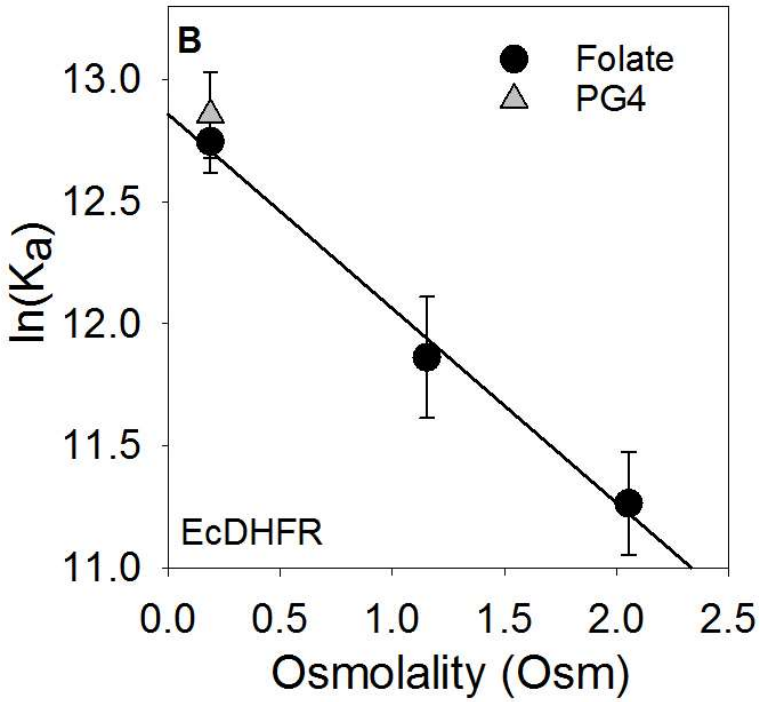
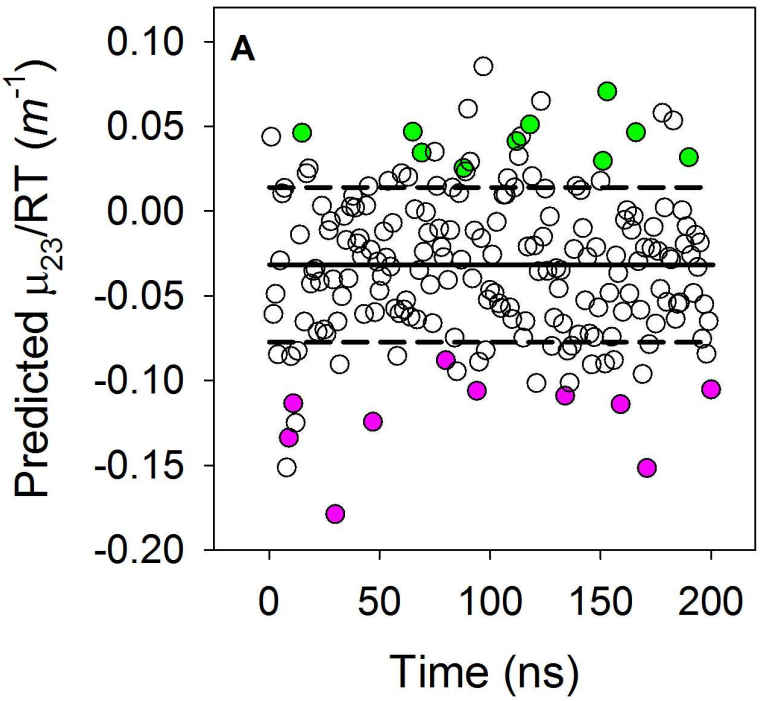
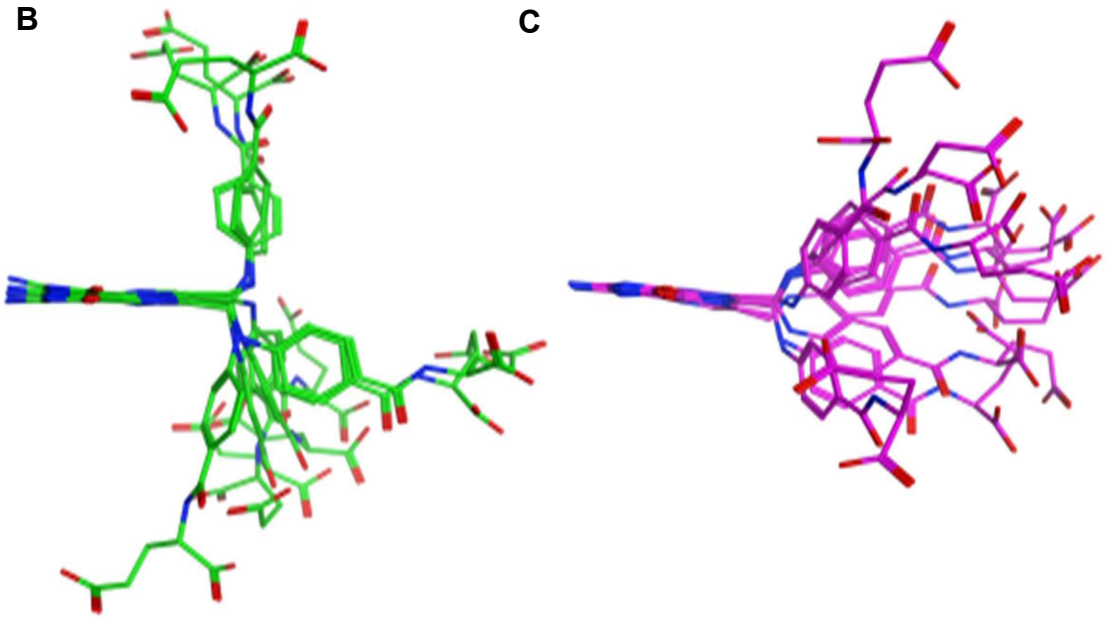

B
Folate
PG4
13.0
12.5
12.0
11.5
11.0
ln(Ka)
EcDHFR
0.0
0.5
1.0
1.5
2.0
2.5
Osmolality (Osm)

Table 3.6 - Thermodynamic parameters for binding of folate and PG4 to the R67 DHFR-NADPH binary complex and EcDHFR by ITC. The osmolalities of the buffer with and without betaine are also listed.

Complex	Buffer and/or osmolyte addition	K_d (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)	n	Osmolality (Osm)
Folate binding to R67 DHFR-NADPH	MTA pH 8	23 ± 5	-6.1 ± 0.1	-8.1 ± 0.8	-1.9	0.60 ± 0.1	0.25
	MTA pH 8 + 5 % betaine	46 ± 16	-5.7 ± 0.2	-5.6 ± 1.2	0.12	0.58 ± 0.1	0.84
	MTA + 10 % betaine	52 ± 27	-5.6 ± 0.4	-2.0 ± 0.6	3.6	0.68 ± 0.1	1.25
PG4 binding to R67 DHFR-NADPH	MTA pH 8	16 ± 6	-6.3 ± 0.3	-6.8 ± 1	-0.52	0.61 ± 0.1	0.25
	MTA pH 8 + 5 % betaine	29 ± 10	-5.9 ± 0.2	-4.4 ± 0.8	1.6	0.58 ± 0.1	0.84
	MTA pH 8 + 10 % betaine	37 ± 13	-5.8 ± 0.3	-1.3 ± 0.2	4.5	0.55 ± 0.1	1.25
Folate binding to EcDHFR	MTA pH 7	2.9 ± 1.7	-7.6 ± 0.3	-9.0 ± 2.3	-1.5	0.80 ± 0.1	0.19
	MTA pH 7 + 10 % betaine	7.1 ± 1.8	-7.1 ± 0.2	-9.6 ± 1.3	-2.5	0.93 ± 0.1	1.15
	MTA pH 7 + 20% betaine	13 ± 4	-6.7 ± 0.2	-9.7 ± 2.5	-2.9	0.83 ± 0.1	2.05
PG4 binding to EcDHFR	MTA pH 7	2.6 ± 0.7	-7.6 ± 0.2	-8.9 ± 0.9	-1.4	0.70 ± 0.1	0.19

Figure 3.10 - Folate Simulation in Water. Predicted μ_{23}/RT values for folate associated with its MD simulation in water (○) are shown in panel A. The average of the μ_{23}/RT values is shown by a solid line. The dashed lines show 1 standard deviation from the average value. Ten representative folate conformers are superimposed on their pterin rings and are shown in panel B for higher (green) and panel C for lower (magenta) than one standard deviation corresponding to the filled circles in panel A. Oxygen and nitrogen atoms are shown in red and blue, respectively.


A
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
Predicted μ_{23}/RT (m^{-1})
0
50
100
150
200
Time (ns)


B
C

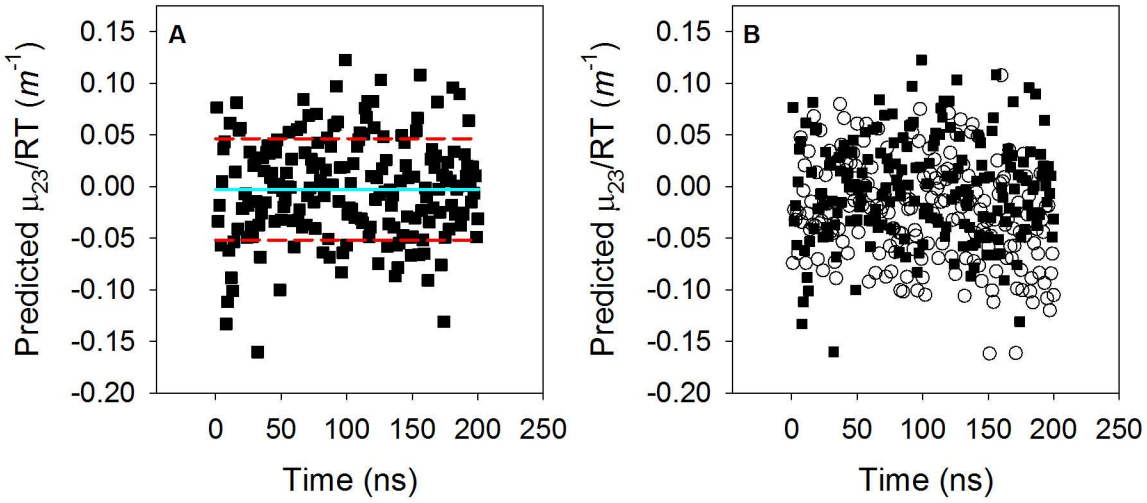


Figure 3.11 - Folate Simulation in Betaine and Water. Predicted μ_{23}/RT values for folate associated with the MD simulations of folate in water with betaine (■). The average of the μ_{23}/RT values is shown by a cyan line. The red dashed lines show one standard deviation from the average value. Panel B shows predicted μ_{23}/RT values from the frames of simulations of folate in just water (○) and in water with 1.35 M betaine (■).

± 0.05 m^{-1}. Slightly more than 30 % of the structures fall above, or below, one standard deviation of the average. This suggests that folate can adopt a range of conformations that can have significantly different interactions with betaine. The average μ_{23}/RT values for folate in water and 1 M betaine are within error, suggesting that betaine has no effect on folate conformations.

3.5 Discussion

3.5.A. Folate is an Interesting Molecule

Folate display many differences from the small molecule compounds and proteins previously studied by the osmometry approach. First, it contains aromatic nitrogen atoms. Our deconvolution of μ_{23}/RT values down to α values indicates aromatic nitrogens prefer to interact with water rather than betaine. This is consistent with betaine not being a H-bond donor, leaving water to interact with the aromatic nitrogens. Second, folate dimerizes, allowing the pterin and *p*-ABA rings to stack. This results in a concentration dependent μ_{23}/RT value. Using our α values, we can predict μ_{23}/RT values. For the dimeric folate model proposed by Poe,[18] this value is 0.81 ± 0.03 m^{-1}. We note the predicted values are based on specific structures of folate while the experimental value describes the solution conformation(s). Differences between the predicted and experimental values can describe variances in the solution conformation(s) vs. our minimized structures. We find that our predicted μ_{23}/RT value is sensitive to the monomeric folate conformation. For example, an extended folate structure from R67 DHFR[42] yields a μ_{23}/RT of -0.11 m^{-1}, while L-shaped folates from EcDHFR (PDB ID 1RX7) and FolT, a folate transporter (PDB ID 4Z7F), provide μ_{23}/RT values of -0.02 m^{-1} and -0.01 m^{-1}, respectively. To assess the

possible folate conformations present in solution, we performed a population analysis. We analyzed 200 folate conformations from a MD trajectory of folate in water, calculated their ASAs with SurfaceRacer and used MATLAB to calculate μ_{23}/RT values. Figure 3.10.A plots the range of μ_{23}/RT values predicted, which is -0.18 to 0.09 m^{-1}. This range of μ_{23}/RT values easily corresponds to the lower limit of the titration seen in Figure 3.3.A. As shown in Figure 3.10.B, the folates with negative μ_{23}/RT values show extended structures while folates with positive values show more bent structures. Analysis of the ASA contributions to the change in μ_{23}/RT value indicates alterations in the N10 and aromatic ring areas are most important. We note the biological relevance of the *p*-ABA-Glu tail flexibility was explored previously by covalent tethering of folate to R67 DHFR, which results in lower enzyme activity.[43] In addition, MD simulations found that flexibility in the *p*-ABA-Glu tail orients the pterin ring for the hydride transfer event in the active sites of both R67 DHFR[42] and EcDHFR.[44]

A third interesting characteristic associated with folate is deprotonation of the N3-O4 enol tautomer, which affects folate's μ_{23}/RT value. The pK_a measured by VPO (8.1 ± 0.17) is similar to those previously monitored by NMR (7.94, 8.38)[9, 35] and capillary electrophoresis (7.98).[36] As O4 titrates from an enol to an enolate and N3 concomitantly loses its proton, a high μ_{23}/RT value results (1.25 ± 0.07 m^{-1}). As the N3 can no longer serve as a H-bond donor, this part of the folate molecule prefers to interact with water. Another consideration arises from quantum mechanical calculations by Soniat et al. on anionic pterin which report delocalization of the negative charge on the ring.[33] Exclusion of betaine from a delocalized negative charge on the pterin ring is consistent with Felitsky et al.[34] who found betaine was strongly excluded from anionic surfaces.

Other compounds with aromatic nitrogens such as pyridoxine and nicotinic acid also showed pH effects on their μ_{23}/RT values. Our measured pyridoxine pK_a was 5.98 ± 0.25. This compares to pK_a values of 5.1 ± 0.02 and 9.0 ± 0.03 for the aromatic nitrogen and phenol hydroxyl measured by potentiometry.[45] However other studies indicate pyridoxine in aqueous solution at neutral pH exists as a mixture of neutral and zwitterionic species.[46-48] While the identity of the titrating species is not clear, the pH dependence of μ_{23}/RT is evident. The general trend is for protonated species to be more interactive with betaine than the deprotonated species. This is true for folate (pK_a ~8), pyridoxine (pK_a 5-6) and nicotinic acid (pK_a ~5). Again, this is consistent with neither betaine nor the small molecule (at the position of interest) being a good H-bond donor. In contrast, water competes well under these conditions.

3.5.B. Deconvolution of μ_{23}/RT into α Values and K_p Values

Our α values are listed in Table 3.5. As mentioned above, our α values mostly show the same sign as those from the Record group, however the amplitudes are different. This may be due to different ASAs calculated for the small molecules. Other differences may be due to whether dimerization occurs as we add aromatic compounds to the list of small molecules. Dimerization was observed in our folate studies as well as imidazole.[49] Another possible difference is the influence of ionization state on μ_{23}/RT values. We (mostly) maintained pH 7 conditions and also considered relevant pK_a values. The Record lab also considered ionization states in their study of PEG interactions as they included 2 different oxygen atom types, -COOH and $-COO^-$. The α values for interaction of these atom types with glycerol are 0.0446 m^{-1} and 0.467 m^{-1}, respectively.[17] An additional issue is whether uracil is aromatic. While a recent publication suggested it is not, we treated

the ring atoms as aromatic.[50] Even with all these caveats, the R^2 for our MATLAB fit of 42 compounds was 0.93. We found removing each of the 15 compounds and refitting to Eq (3.4) yielded similar R^2 values. Also the α values did not change significantly in these various fits.

To test our α value calculations, we predicted μ_{23}/RT values for our test compounds using Eq (3.4) and compared them to the experimental values. A plot of predicted vs. experimental μ_{23}/RT values is linear as can be seen in Figure 3.12. The good correlation between predicted and experimental values supports this type of analysis for betaine interactions with small molecules.

K_p values represent the microscopic local bulk partition coefficients that can be calculated from the α values and Table 3.7 shows K_p values obtained for each surface type. A value less than one indicates water accumulates around the atom more than betaine. K_p values above 1 indicate the opposite, where betaine accrues more readily around the atom surface. Carbon and nitrogen atoms, except for aromatic nitrogens, have K_p values above 1. On the other hand, all types of oxygens, as well as aromatic nitrogens, have K_p values below 1. Therefore, these atom types prefer to be hydrated by water over betaine. A representation of K_p values for the atom types in folate is shown in Figure 3.13.

3.5.C. Solubility vs. VPO Assays

We studied the interaction of folate with betaine using solubility assays and VPO experiments. Both approaches yielded similar results. At pH 7, solubility assays find that folate interacts with both water and betaine with a transfer free energy of 89 ± 30 cal/mol. In our VPO studies in Figure 3.3.A, the μ_{23}/RT value approaches zero at low folate

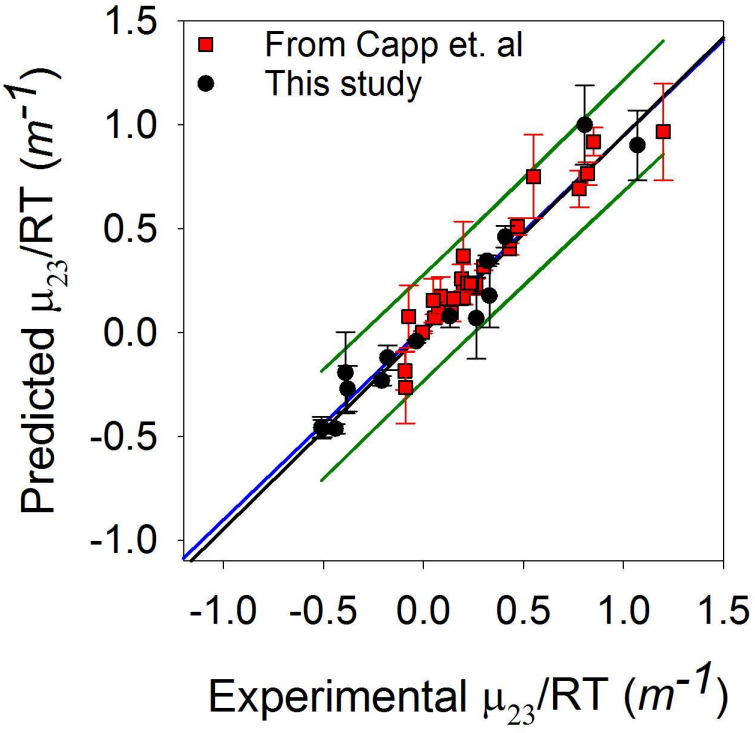


Figure 3.12 - A comparison of predicted μ23/RT values vs. experimental μ23/RT values. The red squares are from Capp et al.[14] and the black circles are from our additional compounds. Many of our compounds have negative μ_{23}/RT values. The black line shows a slope of 1 for a fit through 0,0. The blue line shows the best linear fit of the data with an R^2 of 0.93. The 90% confidence intervals for the fit are shown in green lines.

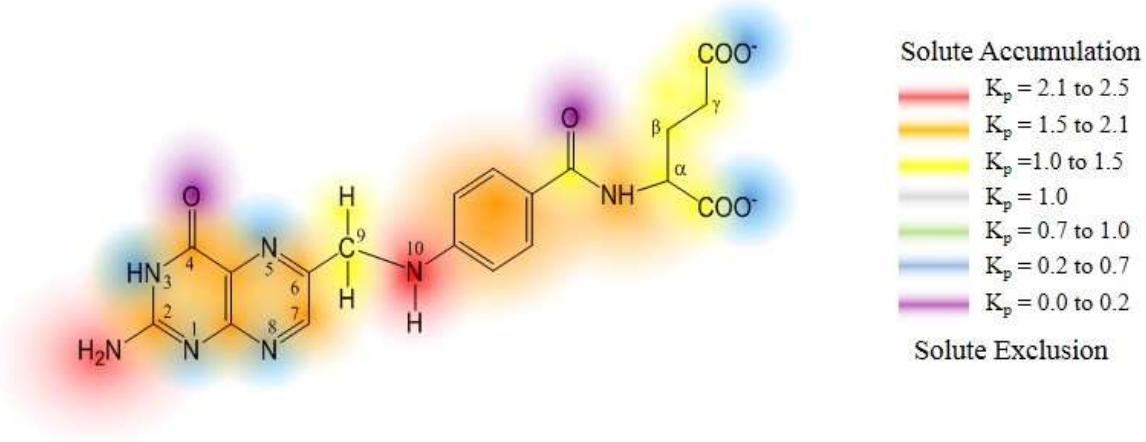


Figure 3.13 - A representation of Kp values for each atom type in folate is shown. Aromatic carbons, amide nitrogen, amine nitrogens off aromatic rings and aliphatic carbons accumulate betaine whereas aromatic nitrogens, carboxylate oxygens and amide oxygen exclude betaine.

Table 3.7 - Calculation of K_p partition coefficient values for the different atom types.

Atom Type	K_p
Aliphatic C	1.08 ± 0.03[a]
Hydroxyl O	0.81 ± 0.03[a]
Amide O	0.11 ± 0.05[b]
Amide N	1.89 ± 0.05[a]
Carboxylate O	0.24 ± 0.03[a]
Cationic N	1.37 ± 0.03[a]
Aromatic C	1.83 ± 0.03[a]
Phosphate O	0.13 ± 0.04[c]
Amine N off Aromatic rings	2.43 ± 0.08[a]
Aromatic N	0.27 ± 0.08[a]

[a] b_i is set to 0.18, which is approximately 2 layers of water.[15] [b] b_i is set to 0.27, which gives a K_p value greater than 0. $b_i = 0.27$ is equivalent to three layers of water surrounding the amide oxygen atoms [c] b_i is set to 0.27, or three layers of water surrounding the phosphate oxygens, in accordance with Capp et. al.[14]

concentrations. Within error, the solubility and VPO techniques converge to similar conclusions. They also qualitatively agree with the prediction of μ_{23}/RT values from our α values for the various folate conformers as shown in Figure 3.12. At pH 10, the solubility assays indicate folate prefers water over betaine interaction with a transfer free energy of 500 ± 150 cal/mol. Our VPO studies agree, yielding a μ_{23}/RT of 1.27 ± 0.36 m^{-1}.

To conclude, at neutral pH, betaine interacts strongly with aromatic carbon surfaces of folate. This interaction is likely due to formation of cation-π pairs.[51-53] Betaine also strongly interacts with the folate amine groups, indicating betaine is a better H-bond partner for this group than water. In contrast, betaine is excluded from aromatic nitrogens, carboxylates, and amide oxygens. This scenario occurs as water can provide H-bonds to these groups while betaine cannot.

3.5.D. Do these Results Provide any Insights into our Previous Studies where Betaine Weakens Binding of Folate to R67 DHFR and EcDHFR?

A means of checking the adequacy of predicting μ_{23}/RT values is to look at the effects of betaine on folate, or DHF, binding to enzymes. Previous ITC studies have looked at the effects of betaine on DHF binding to *E. coli* chromosomal DHFR (EcDHFR) and to R67 DHFR.[7, 8] To determine how accurately the current α values predict betaine's effects, $\Delta\mu_{23}$/RT values for DHF and folate binding to EcDHFR and to the R67 DHFR•$NADP^+$(NADPH) complex were calculated using available protein structures (Table 3.8). Similar calculations were done using the α values from Guinn et. al.[15] The signs of the predicted and experimental $\Delta\mu_{23}$/RT values match, although the amplitudes vary. Also sometimes the Guinn et al.[15] α values provide a better match to experiment

Table 3.8 - Prediction of the $\Delta\mu_{23}$/RT values for betaine effects on the binding of ligands to two different dihydrofolate reductases. The $\Delta\mu_{23}$/RT values were calculated by subtracting the sum of the μ_{23}/RTs of the ligand and the apo-enzyme (or binary complex) from the μ_{23}/RT values for the binary complex (or the ternary complex). The predicted values for complex formation were compared with the $\Delta\mu_{23}$/RT values calculated from ITC data.

Protein-Ligand Complex Formed	**Ligand**	**$\Delta\mu_{23}$/RT (m^{-1})**		
		α values from Guinn et al.[k]	**Using α values from Table 3.5**	**Using Eq (3.6) with ITC data**
EcDHFR•NADP$^+$ [a,b]	$NADP^+$	-0.51	-0.88	-0.23 [l]
EcDHFR•NADPH [b,c]	NADPH	-0.43	-0.77	-0.28 [l]
EcDHFR•DHF [b,d]	DHF	0.52	0.59	0.57 [l]
EcDHFR•Folate [b,e]	Folate	0.57	0.37	0.90 [m]
EcDHFR•NADP$^+$•DHF [f]	DHF	0.30	0.54	0.68 [l]
R67 DHFR•NADP$^+$ [g,h]	$NADP^+$	-0.77	-0.52	-0.84 [n]
R67 DHFR•NADP$^+$•DHF [i]	DHF	0.23	0.46	0.61 [n]
R67 DHFR•NADPH•Folate [j]	Folate	0.19	0.34	0.86 [m]

[a]PDB ID 1RX9 was used in the calculations.[54] [b]The apo-enzyme in PDB ID 5DFR[55] was also used in the calculations. [c]PDB ID 1RX1.[54] [d]PDB ID 1RF7[54] was used with the missing terminal carboxylate group of glutamate tail added to the bound DHF. [e]PDB ID 1RX7.[54] [f]PDB ID 4PDJ.[56] [g]PDB ID 2RK2.[57] [h]The first two residues of the apo protein (PDB ID IVIE)[58] were removed to be consistent with the other structures. [i]The DHF structure has the pABA-glu tail added.[42] [j]The structure from Kamath et. al. with the pterin ring converted to folate was used. [42] [k]α values from Table 1 in Guinn et. al.[15] [l]Data from Grubbs et. al.[8] [m]Data from Table 6. [n]Data from Chopra et.al.[7]

and sometimes the values from Table 3.5 provide a better match. Variations between predicted and experimental values may be due to the Met20 loop, which is disordered in the apo-enzyme and occluded in the $NADP^+$ binary, folate binary and DHF binary complexes.[59] Another factor concerning apo EcDHFR is that it exists in 2 conformations (E1 and E2) prior to binding ligand.[60] Thus conformational heterogeneity could play a role in the ability of computational predictions to match experimental values.

Comparison of the predicted and experimental effects of betaine on ligands binding to R67 DHFR is quite different than for EcDHFR. Again, the sign of the prediction matches that of the experiment, with variations in the amplitude. A possible issue affecting the ability of the calculation to match experiment is the disordered *p*-ABA-Glu tail of bound substrate.[42, 43, 57] Different poses can yield different protein surfaces involved in binding and different substrate conformers, which would both affect the calculated μ_{23}/RT value. Finally, water bridges between R67 DHFR and DHF occur and SurfaceRacer does not take these bridging atoms into account.

A general issue that may affect both experimental data sets is uptake or loss of protons upon binding. Our ITC results have previously found uptake of a proton by R67 DHFR upon binding folate.[61] Additionally, resonance Raman studies find protonation of DHF by the active site of EcDHFR in the ternary complex.[62, 63] This event is not necessarily identified by ITC, which only measures the net number of protons taken up or released.[64] However, binding of NADPH and $NADP^+$, as measured by ITC, does involve release of a proton.[8] Discrepancies between our experimental and predicted $\Delta\mu_{23}$/RT values may arise due to these protonation effects not being accounted for in our

predictions of $\Delta\mu_{23}$/RT values. Another potential issue in our prediction of μ_{23}/RT values using Eq (3.4) may arise due to deviations from the additivity principle for macromolecules.[65, 66] While chemical additivity of small molecules is common, additivity does not always occur in large biochemical molecules. The predicted μ_{23}/RT values of the DHFRs may be overestimated if the interaction potentials of individual groups with betaine are non-additive.

We conclude that this approach to analyze binding has limitations. As with folate (Figure 3.10), proteins are likely to have conformational changes associated with their structures. Indeed loop movement and other dynamics have long been associated with ligand binding to EcDHFR.[54, 67, 68] This suggests that it will likely be difficult to predict μ_{23}/RT values for proteins that release/uptake protons upon binding, undergo dynamic motion, that use "wet interfaces" for binding, for intrinsically disordered sequences and for protein folding, although the Record lab has had some success with the latter case.[16, 69]

3.6 Conclusion

While betaine is an excellent osmolyte for protein stability and folding, it is less helpful for folate to function as a substrate and/or cofactor as the aromatic pterin and *p*-ABA rings prefer to interact with betaine compared to water. This preferential interaction results in weaker binding affinities of folate(s) to DHFRs. As the aromatic pterin ring is lost in dihydrofolate (DHF) and tetrahydrofolate (THF), the predicted μ_{23}/RT values for these more reduced states increases to -0.06 ± 0.03 (for the DHF conformation in the

EcDHFR•$NADP^+$•DHF complex). As DHF and THF contain the same atom types, the predicted μ_{23}/RT values do not change. These values predict osmotic stress effects on other folate pathway enzymes. These effects could be mitigated if the enzymes involved prefer polyglutamylated substrates.

3.7 References

1. Levy, Y. and J.N. Onuchic, *Water mediation in protein folding and molecular recognition.* Annu. Rev. Biophys. Biomol. Struct., 2006. 35: p. 389-415.
2. Biela, A., et al., *Dissecting the hydrophobic effect on the molecular level: the role of water, enthalpy, and entropy in ligand binding to thermolysin.* Angew. Chem. Int. Ed. Engl., 2013. 52(6): p. 1822-1828.
3. Krimmer, S.G., et al., *Methyl, ethyl, propyl, butyl: futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors.* ChemMedChem., 2014. 9(4): p. 833-846.
4. Breiten, B., et al., *Water networks contribute to enthalpy/entropy compensation in protein-ligand binding.* J. Am. Chem. Soc., 2013. 135(41): p. 15579-15584.
5. Homans, S.W., *Water, water everywhere--except where it matters?* Drug Discov. Today, 2007. 12(13-14): p. 534-539.
6. Duff, M.R., Jr. and E.E. Howell, *Thermodynamics and solvent linkage of macromolecule-ligand interactions.* Methods, 2015. 76: p. 51-60.
7. Chopra, S., et al., *A balancing act between net uptake of water during dihydrofolate binding and net release of water upon NADPH binding in R67 dihydrofolate reductase.* J. Biol. Chem., 2008. 283(8): p. 4690-4698.
8. Grubbs, J., et al., *Thermodynamics and solvent effects on substrate and cofactor binding in Escherichia coli chromosomal dihydrofolate reductase.* Biochemistry, 2011. 50(18): p. 3673-3685.
9. Duff, M.R., Jr., et al., *Weak interactions between folate and osmolytes in solution.* Biochemistry, 2012. 51(11): p. 2309-2318.
10. Bhojane, P.P., et al., *Investigation of osmolyte effects on FolM: comparison with other dihydrofolate reductases.* Biochemistry, 2014. 53(8): p. 1330-1341.
11. Howell, E.E., *Searching sequence space: two different approaches to dihydrofolate reductase catalysis.* Chembiochem., 2005. 6(4): p. 590-600.
12. Cayley, S. and M.T. Record, Jr., *Roles of cytoplasmic osmolytes, water, and crowding in the response of Escherichia coli to osmotic stress: biophysical basis of osmoprotection by glycine betaine.* Biochemistry, 2003. 42(43): p. 12596-12609.
13. Timson, M.J., et al., *Further studies on the role of water in R67 dihydrofolate reductase.* Biochemistry, 2013. 52(12): p. 2118-2127.
14. Capp, M.W., et al., *Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values.* Biochemistry, 2009. 48(43): p. 10372-10379.
15. Guinn, E.J., et al., *Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer.* Proc. Natl. Acad. Sci. U. S. A., 2011. 108(41): p. 16932-16937.
16. Diehl, R.C., et al., *Quantifying additive interactions of the osmolyte proline with individual functional groups of proteins: comparisons with urea and glycine betaine, interpretation of m-values.* Biochemistry, 2013. 52(35): p. 5997-6010.
17. Knowles, D.B., et al., *Chemical Interactions of Polyethylene Glycols (PEGs) and Glycerol with Protein Functional Groups: Applications to Effects of PEG and Glycerol on Protein Processes.* Biochemistry, 2015. 54(22): p. 3528-3542.

18. Poe, M., *Proton magnetic resonance studies of folate, dihydrofolate, and methotrexate. Evidence from pH and concentration studies for dimerization.* J. Biol. Chem., 1973. 248(20): p. 7025-7032.
19. Willcott, M.R., *MestRe Nova.* J. Am. Chem. Soc., 2009. 131(36): p. 13180-13180.
20. Tsodikov, O.V., M.T. Record, Jr., and Y.V. Sergeev, *Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature.* J. Comput. Chem., 2002. 23(6): p. 600-609.
21. Richards, F.M., *Areas, volumes, packing and protein structure.* Ann. Rev. Biophys. Bioeng., 1977. 6: p. 151-176.
22. Liu, Y. and D.W. Bolen, *The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes.* Biochemistry, 1995. 34(39): p. 12884-12891.
23. Auton, M. and D.W. Bolen, *Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale.* Biochemistry, 2004. 43(5): p. 1329-1342.
24. Williams, E.A. and J.F. Morrison, *Human dihydrofolate reductase: reduction of alternative substrates, pH effects, and inhibition by deazafolates.* Biochemistry, 1992. 31(29): p. 6801-6811.
25. Reece, L.J., et al., *Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein.* Biochemistry, 1991. 30(45): p. 10895-10904.
26. Houtman, J.C., et al., *Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling.* Protein Sci., 2007. 16(1): p. 30-42.
27. Wiseman, T., et al., *Rapid measurement of binding constants and heats of binding using a new titration calorimeter.* Anal. Biochem., 1989. 179(1): p. 131-137.
28. Case, D.A., et al., *AMBER14.* 2014: p. University of California, San Francisco.
29. H. J. C. Berendsen, J.R.G., T. P. Straatsma, *The missing term in effective pair potentials.* J. Phys. Chem. , 1987(91): p. 6269-6271.
30. Ramanathan, A., et al., *Discovering conformational sub-states relevant to protein function.* PLoS One, 2011. 6(1): p. e15827.
31. Ma, L., et al., *Preferential interactions between small solutes and the protein backbone: a computational analysis.* Biochemistry, 2010. 49(9): p. 1954-1962.
32. Lam, Y.-F. and G. Kotowycz, *Self Association of Folic Acid in Aqueous Solution by Proton Magnetic Resonance.* Can. J. Chem., 1972. 50: p. 2357-2363.
33. Soniat, M. and C.B. Martin, *Theoretical Study on the Relative Energies of Anionic Pterin Tautomers.* Pteridines, 2009. 20(4): p. 124-128.
34. Felitsky, D.J., et al., *The exclusion of glycine betaine from anionic biopolymer surface: why glycine betaine is an effective osmoprotectant but also a compatible solute.* Biochemistry, 2004. 43(46): p. 14732-14743.
35. Poe, M., *Acidic dissociation constants of folic acid, dihydrofolic acid, and methotrexate.* J. Biol. Chem., 1977. 252(11): p. 3724-3728.
36. Szakacs, Z. and B. Noszal, *Determination of dissociation constants of folic acid, methotrexate, and other photolabile pteridines by pressure-assisted capillary electrophoresis.* Electrophoresis, 2006. 27(17): p. 3399-3409.

37. Carlin, H.S. and A.J. Perkins, *Predicting pharmaceutical incompatibilities of parenteral medications.* Am. J. Hosp. Pharm., 1968. 25(6): p. 270-279.
38. Appleby, C.A., B.A. Wittenberg, and J.B. Wittenberg, *Nicotinic Acid as a Ligand Affecting Leghemoglobin Structure and Oxygen Reactivity.* Proc. Natl. Acad. Sci. U. S. A., 1973 70(2): p. 564–568.
39. Pace, C.N., G.R. Grimsley, and J.M. Scholtz, *Protein ionizable groups: pK values and their contribution to protein stability and solubility.* J .Biol. Chem., 2009. 284(20): p. 13285-13289.
40. Peral, F. and E. Gallego, *Self-association of imidazole and its methyl derivatives in aqueous solution. A study by ultraviolet spectroscopy.* J. Mol. Structure, 1997. 415(1-2): p. 187-196.
41. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J. Chem. Theory Comput., 2015. 11(8): p. 3696-3713.
42. Kamath, G., E.E. Howell, and P.K. Agarwal, *The tail wagging the dog: insights into catalysis in R67 dihydrofolate reductase.* Biochemistry, 2010. 49(42): p. 9078-9088.
43. Duff, M.R., Jr., et al., *Tales of Dihydrofolate Binding to R67 Dihydrofolate Reductase.* Biochemistry, 2016. 55(1): p. 133-145.
44. Agarwal, P.K., et al., *Network of coupled promoting motions in enzyme catalysis.* Proc. Natl. Acad. Sci. U. S. A., 2002. 99(5): p. 2794-2799.
45. dos Santos, T.A.D., et al., *Potentiometric and conductimetric studies of chemical equilibria for pyridoxine hydrochloride in aqueous solutions: Simple experimental determination of pKa values and analytical applications to pharmaceutical analysis.* Ecletica Química, 2010. 35(4): p. 81-86.
46. Kiruba, G.S.M. and M.W. Wong, *Tautomeric Equilibria of Pyridoxal-5'-phosphate (Vitamin B6) and 3-Hydroxypyridine Derivatives: A Theoretical Study of Solvation Effects.* J. Org. Chem. , 2003. 68: p. 2874-2881.
47. Ristila1, M., et al., *pH-Dependent Electronic and Spectroscopic Properties of Pyridoxine (Vitamin B6).* J. Phys. Chem. B 2006. 110: p. 16774-16780.
48. Llor, J. and L. Mun, *Tautomeric Equilibrium of Pyridoxine in Water. Thermodynamic Characterization by 13C and 15N Nuclear Magnetic Resonance.* J. Org. Chem. , 2000. 65(2716-2722).
49. Peral, F. and E. Gallego, *Self-association of imidazole and its methyl derivatives in aqueous solution. A study by ultraviolet spectroscopy.* J. Mol. Struct., 1997. 415: p. 187-196.
50. Galvão, T.L.P., et al., *Is Uracil Aromatic? The Enthalpies of Hydrogenation in the Gaseous and Crystalline Phases, and in Aqueous Solution, as Tools to Obtain an Answer.* J. Phys. Chem. A 2013. 117: p. 5826−5836.
51. Cheng, J., et al., *The cation-pi box is a specific phosphatidylcholine membrane targeting motif.* J. Biol. Chem., 2013. 288(21): p. 14863-14873.
52. Diaz-Sanchez, A.G., et al., *Amino acid residues critical for the specificity for betaine aldehyde of the plant ALDH10 isoenzyme involved in the synthesis of glycine betaine.* Plant Physiol., 2012. 158(4): p. 1570-1582.
53. Schiefner, A., et al., *Cation-pi interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-*

binding protein ProX from Escherichia coli. J. Biol. Chem., 2004. 279(7): p. 5588-5596.
54. Sawaya, M.R. and J. Kraut, *Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence.* Biochemistry, 1997. 36(3): p. 586-603.
55. Bystroff, C. and J. Kraut, *Crystal structure of unliganded Escherichia coli dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding.* Biochemistry, 1991. 30(8): p. 2227-2239.
56. Wan, Q., et al., *Toward resolving the catalytic mechanism of dihydrofolate reductase using neutron and ultrahigh-resolution X-ray crystallography.* Proc. Natl. Acad. Sci. U. S. A., 2014. 111(51): p. 18225-18230.
57. Krahn, J.M., et al., *Crystal structure of a type II dihydrofolate reductase catalytic ternary complex.* Biochemistry, 2007. 46(51): p. 14878-14888.
58. Narayana, N., et al., *A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site.* Nat. Struct. Biol., 1995. 2(11): p. 1018-1025.
59. Venkitakrishnan, R.P., et al., *Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle.* Biochemistry, 2004. 43(51): p. 16046-16055.
60. Cayley, P.J., S.M. Dunn, and R.W. King, *Kinetics of substrate, coenzyme, and inhibitor binding to Escherichia coli dihydrofolate reductase.* Biochemistry, 1981. 20(4): p. 874-879.
61. Jackson, M., et al., *Calorimetric studies of ligand binding in R67 dihydrofolate reductase.* Biochemistry, 2005. 44(37): p. 12420-12433.
62. Chen, Y.Q., et al., *Determination by Raman spectroscopy of the pKa of N5 of dihydrofolate bound to dihydrofolate reductase: mechanistic implications.* Biochemistry, 1994. 33(23): p. 7021-7026.
63. Deng, H. and R. Callender, *Structure of dihydrofolate when bound to dihydrofolate reductase.* J. Am. Chem. Soc., 1998. 120(31): p. 7730-7737.
64. Baker, B.M. and K.P. Murphy, *Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry.* Biophys. J., 1996. 71(4): p. 2049-2055.
65. Dill, K.A., *Additivity principles in biochemistry.* J. Biol. Chem., 1997. 272(2): p. 701-704.
66. Szwajkajzer, D. and J. Carey, *Molecular and biological constraints on ligand-binding affinity and specificity.* Biopolymers, 1997. 44(2): p. 181-198.
67. Carroll, M.J., et al., *Evidence for dynamics in proteins as a mechanism for ligand dissociation.* Nat. Chem. Biol., 2012. 8(3): p. 246-252.
68. Boehr, D.D., et al., *The dynamic energy landscape of dihydrofolate reductase catalysis.* Science, 2006. 313(5793): p. 1638-1642.
69. Guinn, E.J., et al., *Quantifying Functional Group Interactions That Determine Urea Effects on Nucleic Acid Helix Formation.* J. Am. Chem. Soc., 2013. 135(15): p. 5828-5838.

PART 4. SMALL ANGLE NEUTRON SCATTERING (SANS) STUDIES ON R67 DHFR, A TETRAMERIC PROTEIN WITH INTRINSICALLY DISORDERED N-TERMINI

This part is a version of a manuscript in preparation.

Small Angle Neutron Scattering (SANS) Studies on R67 DHFR, a Tetrameric Protein with Intrinsically Disordered N-termini

Purva P. Bhojane,[†] Michael R. Duff, Jr.,[†] Khushboo Bafna,[‡] Pratul Agarwal,[§] Arvind Ramanathan,[¶] Christopher Stanley,[‖] and Elizabeth E. Howell[†],*

[†]Department of Biochemistry, Cellular and Molecular Biology, University of Tennessee, Knoxville, TN, 37996-0840
[‡]Genome Science and Technology Program, University of Tennessee, Knoxville, TN, 37996-0840
[§]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37831
[¶]Health Data Sciences Institute, Computational Science and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830
[‖]Biology and Soft Matter Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37831

Address correspondence to Elizabeth E. Howell, Department of Biochemistry, Cellular & Molecular Biology, University of TN, Knoxville, TN 37996-0840; phone, 865-974-4507; fax, 865-974-6306; e-mail, lzh@utk.edu.

Keywords: Disorder, hydration, osmotic stress, SASSIE

4.1 Abstract

Dihydrofolate reductase (DHFR) catalyzes the NADPH dependent reduction of dihydrofolate to tetrahydrofolate, which serves as a source for one-carbon donation reactions in cellular metabolism. R67 DHFR is a plasmid-encoded DHFR that confers resistance against trimethoprim, which is a potent inhibitor of *E.coli* chromosomal DHFR. R67 DHFR is a homo-tetramer with a single active site pore. The dimer crystal structure indicates 16-18 amino acids at the N-terminus of each monomer are intrinsically

disordered. Truncation of 16 N-terminal amino acids results in almost full activity but a lowered stability.

We investigated the effect of ligand binding on the disordered N-termini that might induce a coupled binding and folding of the unstructured tails using small angle neutron scattering (SANS). The binary complex with the oxidized cofactor ($NADP^+$) and the ternary complex with the substrate (dihydrofolate) resulted in radii of gyration comparable to that of the apo protein, suggesting minimal, if any changes in the overall shape of the protein.

We did not observe compaction of the overall structure in the presence of betaine as the radius of gyration (R_g) of the protein indicated slightly higher values. A combined analysis using molecular dynamics and a program called SASSIE gives better insight into the ensemble of states sampled by the disordered tails of the apo R67 DHFR in the presence and absence of betaine. A similar analysis was done for the binary and ternary protein complexes. The disordered N-termini seem to sample collapsed as well as partially extended conformations and remain mostly disordered in all the conditions tested.

We also studied the hydration of R67 DHFR in presence of osmolytes (glycine betaine and DMSO) and our results indicated around 1200 water molecules hydrating the full-length protein in the presence of betaine as well as DMSO.

4.2 Introduction

Dihydrofolate reductase (DHFR) catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using NADPH as a cofactor. THF and its derivatives serve as cellular cofactors for one-carbon transfer reactions involved in the synthesis of nucleotides

such as thymidine, amino acids such as methionine and glycine and various other metabolites. Effective blocking of DHFR activity leads to cell death. Therefore, this enzyme is a target for anti-bacterial and anti-cancer drugs. Trimethoprim is a potent inhibitor of *E. coli* chromosomal DHFR (EcDHFR) that has been used widely as an antibacterial drug.

The gene encoding R67 DHFR, carried by an R-plasmid, confers resistance against trimethoprim. This type II DHFR is genetically and structurally unrelated to EcDHFR. R67 DHFR is a homotetramer and each monomer has five antiparallel β-strands that assemble into a dimer with a six-stranded β-barrel at the subunit interface. Using loop-loop interactions, two dimers assemble into a tetrameric "doughnut" with a single active site pore.[1]

Numerous experiments indicate the first 21 residues at the N-terminal of R67 DHFR are disordered and can tolerate various sequences. For example, several disorder predictors indicate the N-terminal sequence to be intrinsically disordered.[2] Also the first 18 amino acids for each monomer do not appear in the dimer crystal structure.[3] The N-termini can be cleaved after the 16^{th} residue (Phe) by chymotrypsin treatment and the truncated protein is almost fully active, although somewhat less stable.[4] When a gene encoding the truncated protein is constructed, no trimethoprim resistance was observed *in vivo*. The truncated tetrameric protein was crystallized by Narayana et al. and the structure was first solved at a resolution of 1.7Å[1] and a later resolution of 1.1 Å. [5] High thermal factors in the refined structure obtained under cryo-cooling conditions at 100K suggested the stretch of residues from 17–21 to be disordered independent of the temperature at which

the data were collected. In addition, electron densities for residues 21-23 were noted to be diffuse, indicating high mobility.[5]

Other type II DHFR variants (e.g. R388, R751) show different N-terminal sequences but the same core sequence contributes to the β-barrel structure.[2, 6, 7] This can also be seen from the sequence alignment of the type II DHFR variants suggesting non-identity in the first 21 residues at the N-termini. His tags can be added to the N-termini.[8, 9] Further a tandem array of four R67 DHFR gene copies encodes a protein where the C and N-termini of the first and second monomers are fused as well as the second and third monomers and the third and fourth monomers. The resulting Quad1 protein possessing four times the molecular mass of the R67 DHFR monomer was stable as well as functional.[10] Similarly, the N-terminal sequences from R388 and R751 can be used as the linker domains to give a functional monomeric Quad4 protein.[2] These various experiments and constructs indicate the N-termini can be modified without loss of function.

In vivo TMP resistance is not conferred onto the host cells by a gene encoding the truncated R67 DHFR, indicating a role of the disordered N-termini in protein expression and stability. To gain information on the conformational space occupied by the disordered N-terminal sequences in R67 DHFR, we performed small angle neutron scattering (SANS) experiments in different solution conditions. As there is no structural information about the N-termini, SANS data could be utilized to model the ensemble of conformations sampled by the disordered tails in solution.

Disordered sequences often undergo coupled binding and folding as the Gibbs energy of the native state is lowered by using the binding energy of ligands or other protein partners to drive folding. We monitored if there was any change in the conformational

sampling of the disordered tails upon binary complex (R67 DHFR-$NADP^+$) or ternary complex (R67 DHFR-$NADP^+$-DHF) formation. Also, osmolytes have been shown to exert protein stabilizing forces via a preferential exclusion mechanism. To determine whether osmolyte addition leads to folding of the termini, we added deuterated betaine to examine any changes in the R67 DHFR shape with SANS.

The water associated with the protein surface comprises the hydration layer which can be differentiated from the bulk solvent. The first hydration shell can contain tightly bound water as well as water that can freely exchange. These differences are due to the varied environments associated with the protein surface which can display different clefts and bumps as well as different atom types.[11] Computational studies have shown the water molecules hydrating the disordered chains exhibit different properties than those surrounding the globular domains, both in terms of number of waters and structural order of the water molecules in the hydration layer.[12, 13]

Various methods can be employed to study protein hydration. A typical approach calculates the accessible surface area (ASA) and divides the value by 9 $Å^2$ to predict the number of solvent waters. Several experimental techniques including NMR probe hydration water.

We aimed to monitor the preferential hydration of full length R67 DHFR using SANS experiments, upon addition of hydrogenated osmolytes in D_2O buffer solution; this is analogous to a H_2O/D_2O contrast variation approach, i.e. the contrast created by hydrogenated osmolyte addition allows measurement of the hydration shell associated with R67 DHFR. The contrast created by osmolytes differentiates between the hydration layer and the bulk solvent. The information obtained from the scattering contrast can be used to

obtain the number of water molecules in the hydration layer that are responsible for exclusion of the added osmolyte from the protein surface.

4.3 Methods

4.3.A. Protein Expression and Purification

R67 DHFR was expressed and purified as per Reece et al.[4] Briefly, cell lysates were subjected to ammonium sulfate precipitation and ion-exchange column chromatography to purify the protein to homogeneity. Purified samples were dialyzed against distilled, deionized H_2O and lyophilized. Protein concentrations were determined using a bicinchoninic acid (BCA) (Pierce) assay.

4.3.B. Small Angle Neutron Scattering (SANS)

Experiments were performed on the EQ-SANS instrument at the Spallation Neutron Source at the Oak Ridge National Laboratory. In 60 Hz operation mode, a 4 m sample-to-detector distance with 2.5-6.1 Å wavelength band was used.

Samples of R67 DHFR were prepared in 20 mM deuterated Tris buffer in D_2O (pD 7.0) with no osmolyte and with the osmolytes betaine and DMSO. The osmolytes were hydrogenated to create a contrast with the deuterated buffer conditions, allowing measurement of the changes in preferential hydration of apo R67 DHFR.[14] The concentrations of osmolytes ranged from 2.5 % to 20 % (w/v) for betaine and 2.5 % to 17.5 % for DMSO (v/v). The protein concentration ranged from 4.5 - 7.5 mg/ml.

All samples were prepared, centrifuged and loaded into banjo-shaped quartz cuvettes (Hellma USA, Plainville, NY) of 2 mm path length. Neutron exposure times were approximately 1 h and the scattered neutrons were detected on a 1 × 1 m two-dimensional

detector at 25 °C. The data collected for all experiments were reduced using MANTID Plot[15] and the total two-dimensional scattering was corrected by the scattering from the empty quartz cell. Then, the scattering was normalized by the incident beam flux and radially averaged to obtain the absolute scale intensity, I(q) versus scattering angle, q. The background scattering for the respective buffers was subtracted from the total scattering. Guinier analysis with a linear plot of ln I(q) versus q^2 for low q data gave a slope of $-(R_g^2)/3$ and the intercept on the Y-axis gave the I(0) value. An estimate of the radius of gyration of the protein (R_g) and the zero angle scattering intensity I(0) was obtained using Eq (4.1): [16]

$$I(q) = I(0)e^{-q^2 {R_g}^2/3} \qquad \text{Eq (4.1)}$$

where I(q) and I(0) are the scattering intensities at small angles (q) and at zero angle respectively; and R_g is the radius of gyration.

The data were also analyzed using the GNOM program in the ATSAS package.[14] GNOM reads the scattering profile and evaluates the particle distance distribution function, P(R), in a defined range of distances and yields the apparent radius of gyration (R_g) and zero angle scattering intensity I(0). Data for each sample were fit using Guinier analysis and the GNOM program.

The R_gs and the zero angle scattering intensities, I(0), of R67 DHFR in the presence of varying concentrations of osmolytes (betaine and DMSO) were determined from the GNOM fitting. The data were normalized for the protein concentration of each sample. To obtain information on the preferential hydration of R67 DHFR and effect of osmolytes on

the hydration, the change in I(0) with increasing concentration of osmolytes was fit to Eq (4.2) from Stanley et al.[15]

$$I_s(0)/I(0) = (1 + f_v\,(\frac{\rho_w-\rho_s}{\rho_p-\rho_w})(\frac{V_p+V_w}{V_p}))^2 \quad \text{Eq (4.2)}$$

where $I_s(0)$ and I(0) are the zero angle scattering intensities in the presence and absence of osmolyte respectively, f_v, or fractional volume, is the concentration of osmolyte added (w/v for betaine and v/v for DMSO), ρ_w, ρ_s, ρ_p, are the scattering length densities of water, solute (=osmolyte), and protein, respectively, and V_p and V_w are the volumes of protein and protein-associated water, respectively. The scattering length densities of protein, betaine, DMSO and the protein volume were calculated using the online tool MULCh.[16] The volume of protein associated water gives the number of water molecules in the hydration layer of R67 DHFR upon osmolyte addition.

To study the effect of betaine on the disordered N-termini of R67 DHFR, the change in overall shape and compaction of apo protein in the presence of 20 % deuterated betaine was explored. Experiments were also done to study any changes in the protein ordering upon binding of $NADP^+$ to apo R67 DHFR (binary complex formation) and of DHF to R67 DHFR-$NADP^+$ (ternary complex formation). Buffer controls were run for detecting the background scattering. Data were analyzed using Guinier analysis and GNOM to determine the R_g and I(0) values. GNOM also gives the pairwise distance distribution.

4.3.C. Analysis using MD and SASSIE

Our next step was to analyze the data using SASSIE (http://www.smallangles.net/sassie/SASSIE/SASSIE_HOME.html).[17] This program

suite creates atomistic models of the protein using Monte-Carlo simulations, calculates theoretical scattering data for these models using a SasCalc tool and compares it to the experimental data. The experimental SANS data were interpolated into SASSIE in a defined q range using the data interpolation module. SASSIE required full length protein as the starting structures to generate a large number of models for fitting.

The N-termini were added to the crystal structure of truncated R67 DHFR (2RH2)[18] using Modeler (version 11) with energy minimization generating 10 models. To get a broad range of starting structures for SASSIE, molecular dynamics simulations on 10 models of full length R67 DHFR were run for 100 ns using the AMBER force field. Further 1 μs MD runs were performed for 4 full length protein models to provide additional sampling time. The modelling of the full length protein and molecular dynamic simulation was done in collaboration with Khushboo Bafna and Dr. Pratul Agarwal.

Representative frames were extracted from the course of the simulations and the SasCalc module in SASSIE was used to generate theoretical SANS profiles, which were compared to the experimental SANS data (apo R67 DHFR in buffer with no osmolytes) using the χ^2 analysis module. Those structures with a low χ^2 value (<10) were chosen as good fits to the experimental SANS data and five such frames were chosen for further analysis in SASSIE. Additionally, two more models were built – one with the two N-termini interacting with each other on either side of the tetramer and a second one with all the four termini blocking the active site pore. This approach allowed us to obtain a set of 27 starting structures where the N-termini sample a large conformational space.

All the selected input frames were used to run a Complex Monte-Carlo simulation generating 10,000 frames from each starting structure of which the accepted frames

avoiding atom overlap were used for further analysis. The core of the protein remained constant, only alternate conformations of the N-terminal 21 amino acids were generated. Based on the average R_g obtained (~ 21.5 Å), directed Monte-Carlo sampling was additionally performed to generate 20,000 structures with R_g values limited to a range from 20.5 Å – 22.5 Å.

The resultant frames from Monte-Carlo sampling were subjected to a 500 step minimization using NAMD and the theoretical SANS profiles were calculated using the SasCalc module in SASSIE. The next step in SASSIE was a χ^2 analysis, which compares the theoretical profile for each frame to the experimental data. Similar comparisons were performed with the experimental SANS profiles obtained for the ligand bound complexes (binary and ternary) as well as the apo R67 DHFR in 20 % deuterated betaine.

4.3.D. Differential Scanning Calorimetry (DSC)

Thermal unfolding of R67 DHFR was monitored between 25 and 95 °C using a Microcal VP differential scanning microcalorimeter. The concentration of R67 DHFR was 150-160 μM in MTA buffer (100 mM MES, 50 mM Tris and 50mM acetic acid), pH 8. Samples were also prepared in MTA buffer with 20 % betaine or 15 % DMSO. Scan rates were 1 °C/min. Scans were repeated two times. The data obtained were analyzed using the Origin program (version 7.0) supplied by the manufacturer and the melting temperatures were obtained.

4.4 Results

4.4.A. SANS Data Analysis for Apo R67 DHFR

The SANS profile shown in Figure 4.1.A examines the overall shape and radius of gyration of R67 DHFR. The data have been corrected by subtracting the background scattering from the buffer and the intensity was normalized by I(0). The primary analysis using GNOM fits shown in Figure 4.1.B yields a R_g value of 21.89 ± 0.12 Å. The crystal structures of truncated apo R67 DHFR gave R_g values of 17.2 Å (2RH2)[18] and 17.6 Å (2GQV)[19] as determined using MOE. (version 2015.10). The differences in the R_g values for the two truncated proteins may be due to Ser 20 be present in one of the structures (2GQV)[19] whereas the other structure (2RH2)[18] starts at the Asn 21.

4.4.B. Predicting the Structures Associated with the R67 DHFR N-Termini using MD and SASSIE

To gain information about the space sampled by the N-termini, we used a MD approach for analyzing our SANS data. As described in the methods section, 25 frames of the full length protein were selected from MD simulations and used as inputs for a Monte-Carlo atomistic simulation module in SASSIE. Additionally, two more models were hand built where the termini either blocked access to the active site pore or the termini were interacting and collapsed on both sides of the protein. The complex Monte-Carlo module in SASSIE sampled various conformations of the initial 21 residues at the four N-termini of R67 DHFR holding the rest of the structure constant. The output frames were filtered for accepted structures based on steric hindrance. Thus, a total of ~150,000 frames were generated with the Monte-Carlo simulations over a wide range of R_g, from 19.5 Å to 29.5 Å.

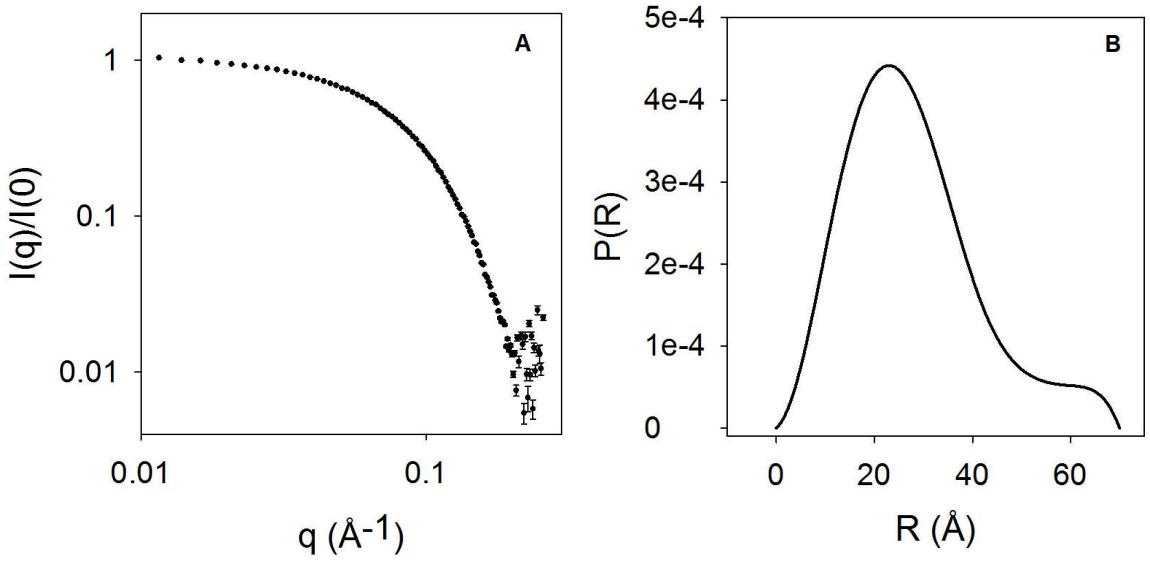
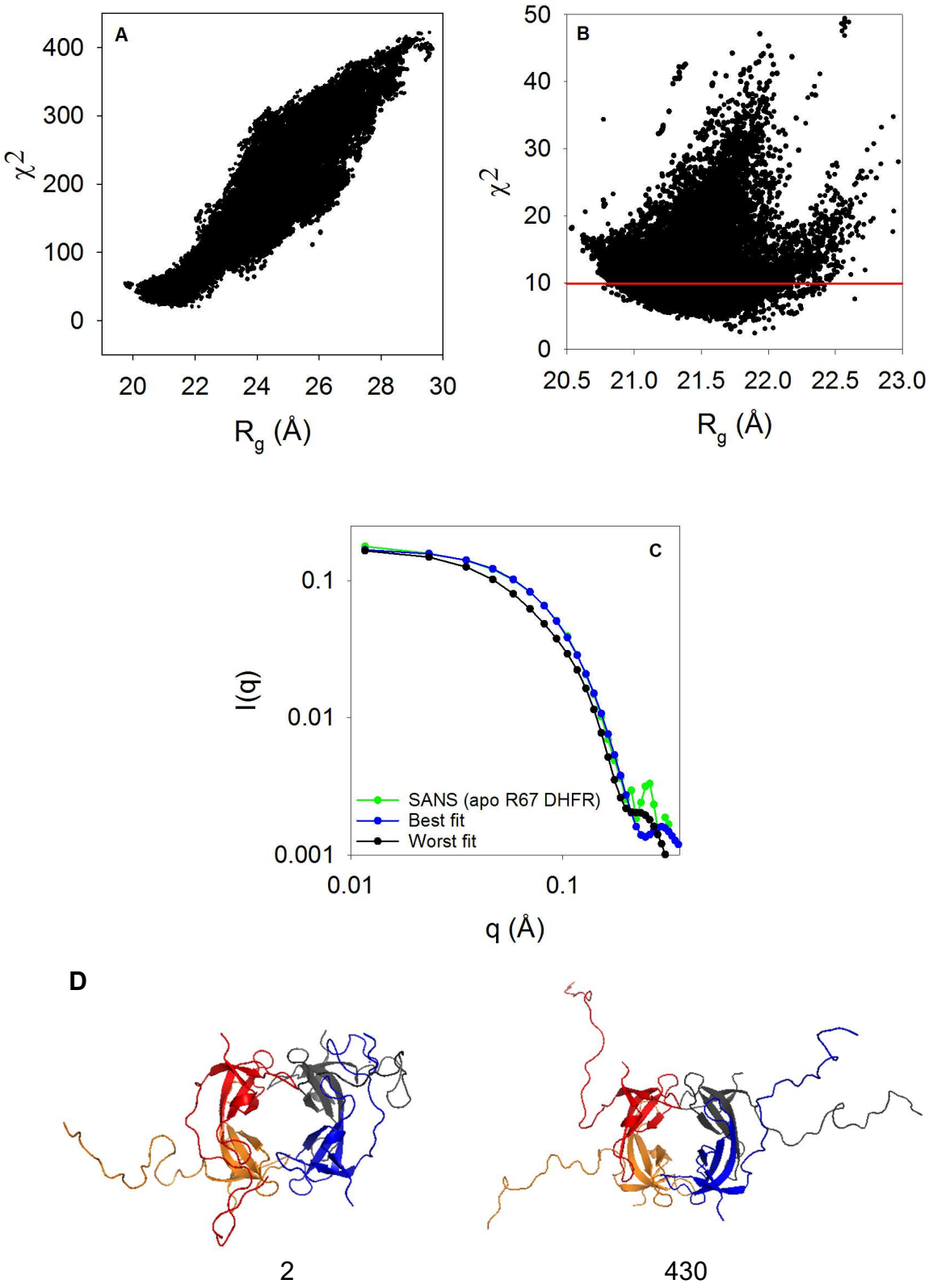


Figure 4.1- SANS profile and GNOM analysis for apo R67 DHFR. Panel A shows the normalized scattering intensity of the protein, I(q)/I(0), with increasing q. The SANS profile was obtained upon subtracting the scattering contribution from the buffer and normalizing the scattering intensity by I(0). A GNOM fit of the profile, shown in panel B, gave the pairwise distance distribution and a R_g value of 21.89 ± 0.12 Å.

Our next step was to use SasCalc in SASSIE to obtain theoretical SANS profiles for the accepted structures. These were compared to the experimental SANS data obtained for apo R67 DHFR in the absence of osmolytes. A goodness of fit analysis gave an output of χ^2 for each fit against the R_g. All the 150,000 structures had high χ^2 values (12-430). This suggested that our initial sample set of 150,000 frames did not contain conformations that could provide the best fits to the experimental SANS data. Figure 4.2.A plots the χ^2 versus R_g distribution for all the frames, which indicates that the χ^2 value increases with increasing R_g value. Thus, to obtain frames to model our SANS data with lower χ^2 values, we performed directed Monte-Carlo simulations (see Methods 4.3.C for details) that generated frames yielding better fits. Figure 4.2.B shows the χ^2 versus R_g plot for ~ 25,000 frames sampled by directed Monte-Carlo simulations. Thus, we obtained a set of ~175,000 frames from our SASSIE analysis. The space sampled by all those structures is represented by the gray mesh in Figure 4.2.F

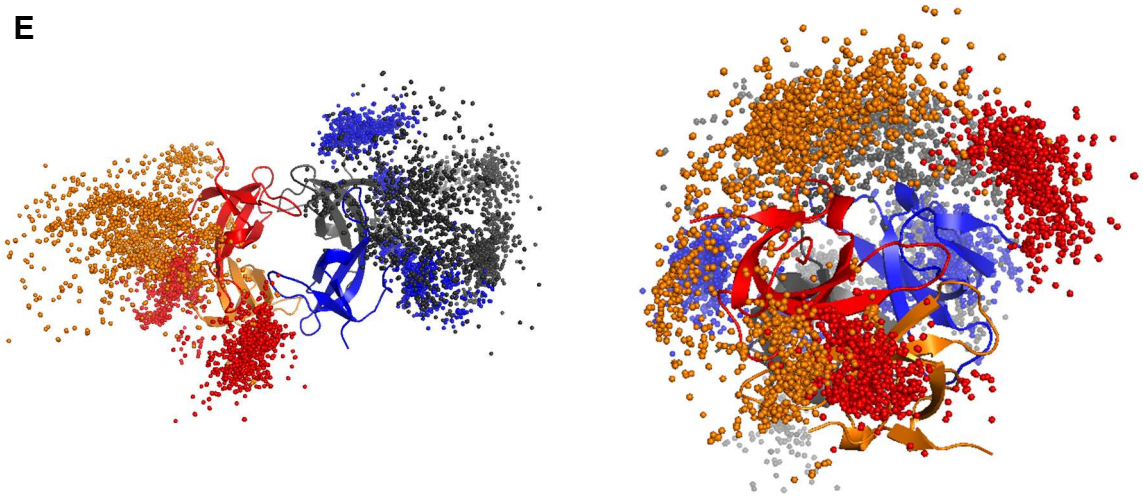
Figure 4.2.C shows the overlay of the SANS profiles for the best fit (χ^2=2) and the worst fit (χ^2=430) to the experimental SANS data. The corresponding structures for the best and worst fits are shown in Figure 4.2.D. The analysis provided 6791 frames with χ^2 values lower than10 representing better fits to the experimental data. The positions of the N-termini with respect to the core of the protein are shown in a figure of the center of mass of the first residue (methionine) for each chain (see Figure 4.2.E). The asymmetrical sampling of the disordered tails was observed with two N-termini sampling collapsed conformations, positioned near the core of the protein (red and blue spheres) as compared to the other two termini sampling both the collapsed conformations placing them near the core as well as relatively extended conformations away from the core of the protein (orange

Figure 4.2 - SASSIE analysis for apo R67 DHFR. Panel A shows the χ^2 vs. R_g plot comparing the SANS profiles of representative frames generated by Monte-Carlo sampling to the experimental SANS data for apo R67 DHFR in absence of osmolytes. The red line indicates a cut off for χ^2 =10. Panel C overlays the theoretical SANS profiles for the best (χ^2 = 2) and worst fits (χ^2 = 430) compared to the experimental SANS data while Panel D provides the corresponding best and worst structures, respectively. Panel E shows the center of mass of the N-terminal methionine for each of the 4 chains for the frames with χ^2 <10 (good fits). The four monomers and the center of mass points for each N-termini are colored coded. Also, the center of mass points for two termini are shown with 90° rotation to visualize the interface between two monomers. Panel F shows the overlay of the density plots for the structures sampled by MD and Monte-Carlo (gray mesh) and the structures providing good fits to the experimental data (light blue mesh). The gray mesh indicates the MD simulations and Monte-Carlo approaches sample most of the available "structural space". The overlay plots are also shown with a 90° rotation (side view of the pore). The density plot representing the sampling of each termini is shown in panel G. The mesh and the monomers are color coded.

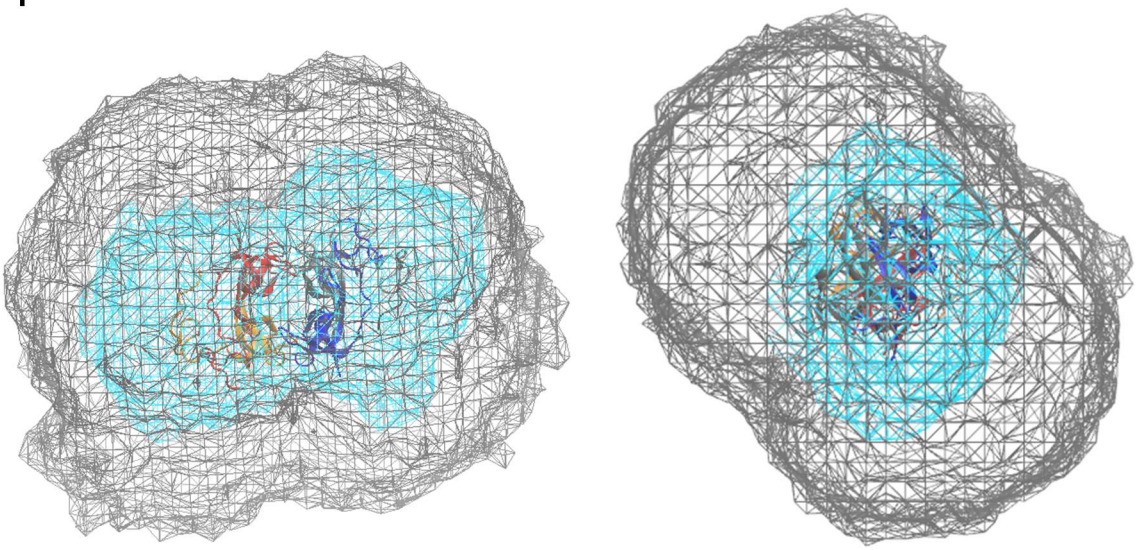

A
400
300
200
100
0
χ2
20
22
24
26
28
30
Rg (Å)
B
50
40
30
20
10
0
χ2
20.5
21.0
21.5
22.0
22.5
23.0
Rg (Å)
C
0.1
0.01
0.001
I(q)
0.01
0.1
q (Å)
SANS (apo R67 DHFR)
Best fit
Worst fit
D
2
430

Continued Figure 4.2

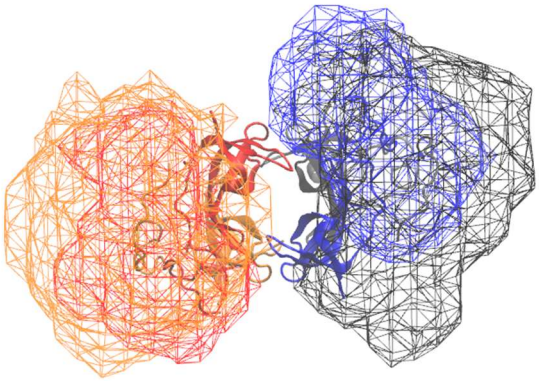
E



F



G



and gray spheres). Interactions between two disordered tails and also between the disordered and the ordered protein surface also seem possible as seen from the center of mass positions in Figure 4.2.E.

Another way of representing the outcome of SASSIE analysis is shown in Figure 4.2.F. A large sampling of conformations for the N-termini observed using the Monte-Carlo approach can be seen from the gray mesh in Figure 4.2.F representing the density plot for all the frames generated by Monte-Carlo sampling. Best fits by SASSIE for the apo protein sample regions near the structured protein as shown by the blue mesh (see Figure 4.2.F). The overlay of density indicates the tendency of the N-termini to compact and sample space mostly near the sides of the structured protein. An asymmetric density plot for the good fits can be attributed to the conformations disordered tails extending out from the ordered protein core. A range of R_g values for the good fits is given in Table 4.1. Thus, we obtain the conformational sampling of the disordered tails of R67 DHFR with an average R_g value of 21.27 ± 0.21 Å, which is within error of the R_g value obtained from the GNOM fitting of the SANS data for apo R67 DHFR as can be seen from Table 4.1.

4.4.B Effect of Ligand Binding on the Disordered Termini in R67 DHFR

SANS data were collected for R67 DHFR to monitor any changes in the disordered N-termini upon ligand binding. Data collected for binary (R67 DHFR-$NADP^+$) and ternary (R67 DHFR-$NADP^+$-DHF) complexes were analyzed using GNOM. A comparison of the pairwise distribution plots for the apo, binary and ternary complexes is shown in Figure 4.3. The R_g values for the apo protein, $NADP^+$ binary and $NADP^+$-DHF ternary complexes are 21.89 ± 0.12 Å, 21.45 ± 0.14 Å, and 21.45 ± 0.18 Å respectively. As these values are close, they suggest minimal changes in the shape of the protein occur upon ligand binding.

Table 4.1 – Comparison of the radii of gyration of the full length R67 DHFR obtained upon analyzing the SANS data using GNOM and SASSIE modules.

Protein Samples	**GNOM R_g (Å)**	**SASSIE Good fits**		
		Number	**R_g Range (Å)**	**Mean R_g (Å)**
Apo R67 DHFR	21.89 ± 0.12	6791[a]	20.77 – 22.64	21.27 ± 0.21
R67 DHFR - $NADP^+$	21.45 ± 0.14	569 [a]	20.65 – 21.89	21.26 ± 0.22
R67 DHFR - $NADP^+$- DHF	21.45 ± 0.18	86 [a]	20.68 – 22.01	21.35 ± 0.28
Apo R67 DHFR in 20 % deuterated betaine	22.84 ± 0.30	15305 [b]	21.89 – 24.7	22.72 ± 0.48

[a]Frames with χ^2 values lower than 10
[a]Frames with χ^2 values lower than 5

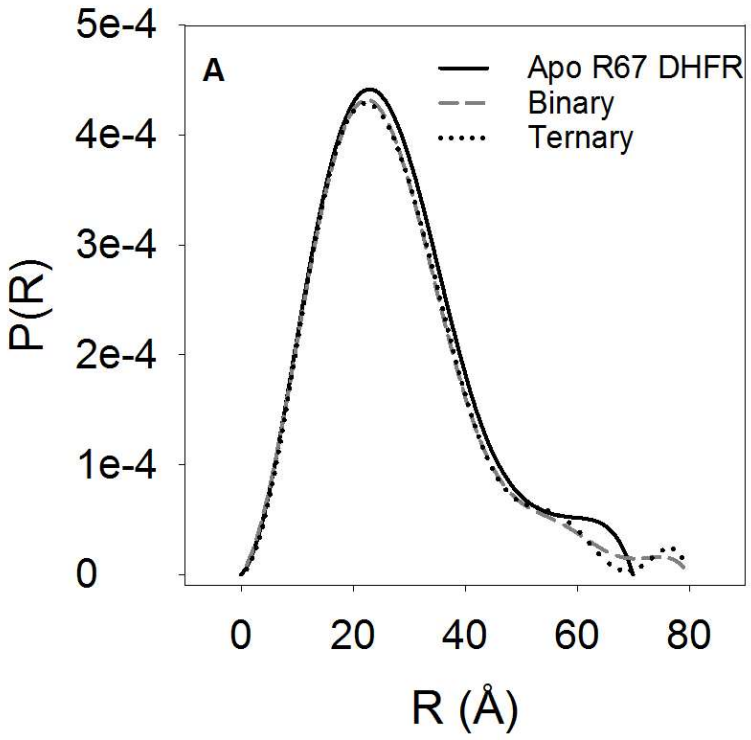
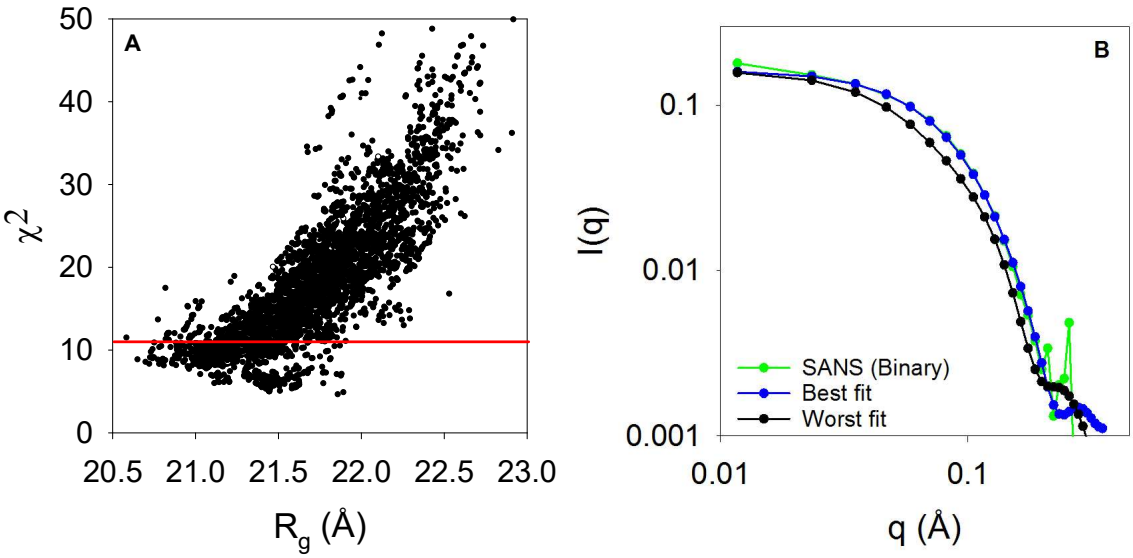


Figure 4.3 - Pairwise distribution of R_g for apo R67 DHFR, $NADP^+$ bound binary and $NADP^+$ and DHF bound ternary complexes. SANS data were collected using 6.5 mg/mL DHFR in 20 mM deuterated Tris buffer in D_2O (pD 7.0) with no osmolyte. Binary and ternary complexes were formed by adding 3 mM $NADP^+$ or $NADP^+$ plus 2 mM DHF respectively. The pairwise distributions of R_g for R67 DHFR by GNOM (black line), R67 DHFR- $NADP^+$ binary (dashed line) and R67 DHFR-$NADP^+$-DHF ternary (dotted line) complexes are shown. The R_g values for the apo protein, binary and ternary complexes are 21.89 ± 0.12 Å, 21.45 ± 0.14 Å and 21.45 ± 0.18 Å respectively, indicating minimal differences in the overall protein compaction upon ligand binding.

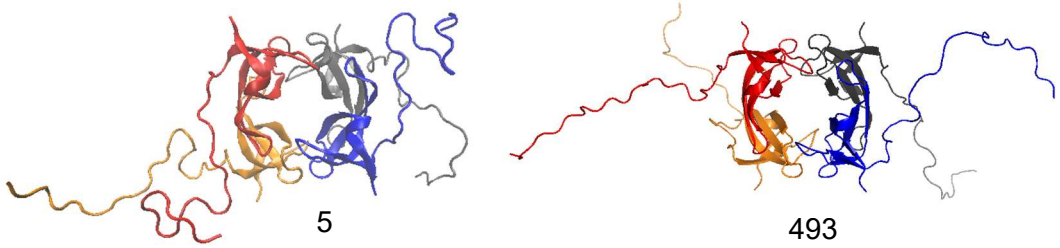
To get deeper insight on the disordered tails, SANS data for the R67 DHFR-$NADP^+$ binary and R67 DHFR- $NADP^+$-DHF ternary complexes were further analyzed using SASSIE. SASSIE analysis using the same set of ~175,000 frames is shown in Figure 4.4 and 4.5 for the binary and ternary complexes, respectively. Preliminary analysis yielded high χ^2 similar to that for the apo protein as in Figure 4.2.A. The frames obtained from directed Monte-Carlo simulation generated structures that provided better fits to the binary and ternary SANS data (see Figures 4.4.A and 4.5.A). The structures with a χ^2 of 10 or below were considered to be the good fits of the experimental data for both binary and ternary complexes. Out of all the frames, our analysis gave 569 frames with an acceptable χ^2 value of less than 10 for the binary data analysis. Further, only 86 frames yielded good fits to the ternary data. The best χ^2 values for the binary and ternary complexes were 5 and 8 respectively. The structures obtained as good fits for ternary data were the same as that for the binary data. The conformations sampled exhibit a similar range for both the binary and ternary fits. The mean R_g value obtained for 569 structures with a $\chi^2 < 10$ for the binary complex data was found to be 21.26 ± 0.22 Å. This value is comparable to the R_g obtained by GNOM analysis. Also, the mean R_g value for 86 structures with a $\chi^2 < 10$ for the ternary complex was 21.35 ± 0.28 Å, again within error of the R_g value by GNOM analysis. Thus the disordered tails do not undergo drastic changes in their conformational sampling behavior upon binding of substrate to the cofactor bound binary complex as can be seen from the density plots in Figure 4.5.G.

The center of mass representation in panel D of both Figures 4.4 and 4.5 depicts a similar behavior of the N-terminal tails as seen with the SASSIE analysis for the apo protein. Also, the good fits for the binary and ternary data mostly overlapped with the good

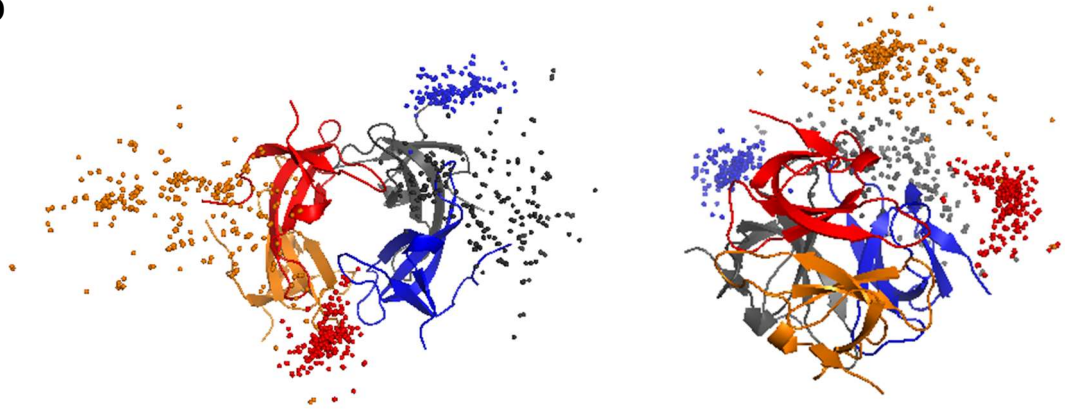
Figure 4.4 – Conformational analysis of the disordered N-termini upon NADP+ binding to R67 DHFR by SASSIE. Panel A shows the resulting χ^2 vs. R_g plot generated by Monte-Carlo sampling. The red line indicates a cut off for χ^2 =10. Panel B shows an overlay of the theoretical and experimental SANS profile for the best and worst fits (lowest and highest χ^2 values respectively). Panel C shows the corresponding structures of the best and worst fits. Panel D represents the center of mass points of the N-terminal methionine residue of each of the four N-termini of the good fits from SASSIE with a χ^2 <10. The four monomers are colored differently and the center of mass points are color coded for each monomer. In panel E the pink mesh depicting the density plot for all the good fits (569) to the binary complex data overlaid on the density plot of all sampled structures (gray mesh) is shown. Panel F shows the density plot represented for each of the N-termini color coded with the monomers shown in the structure.


A
50
40
30
20
10
0
χ2
20.5
21.0
21.5
22.0
22.5
23.0
Rg (Å)
B
0.1
0.01
0.001
I(q)
SANS (Binary)
Best fit
Worst fit
0.01
0.1
q (Å)

C

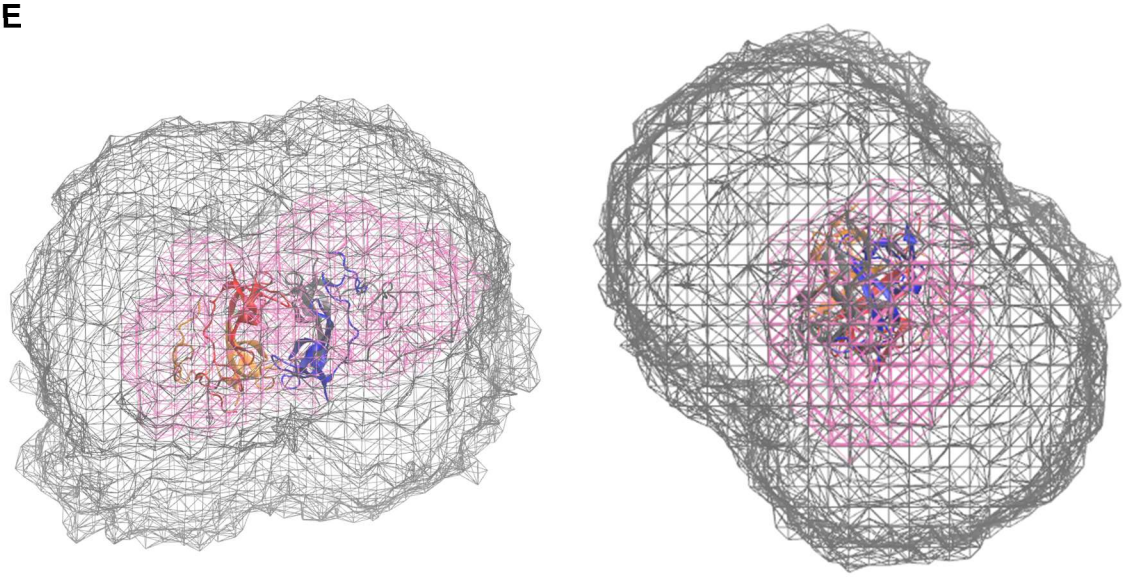

5
493

D



Continued Figure 4.4

E



F

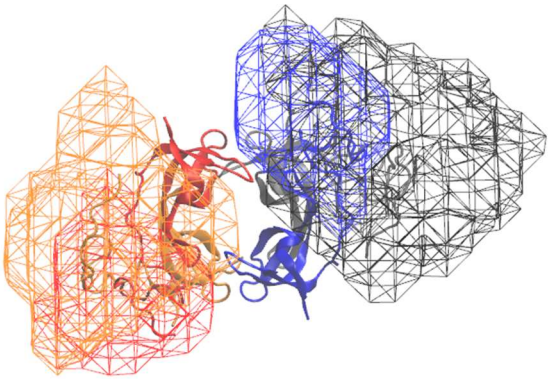
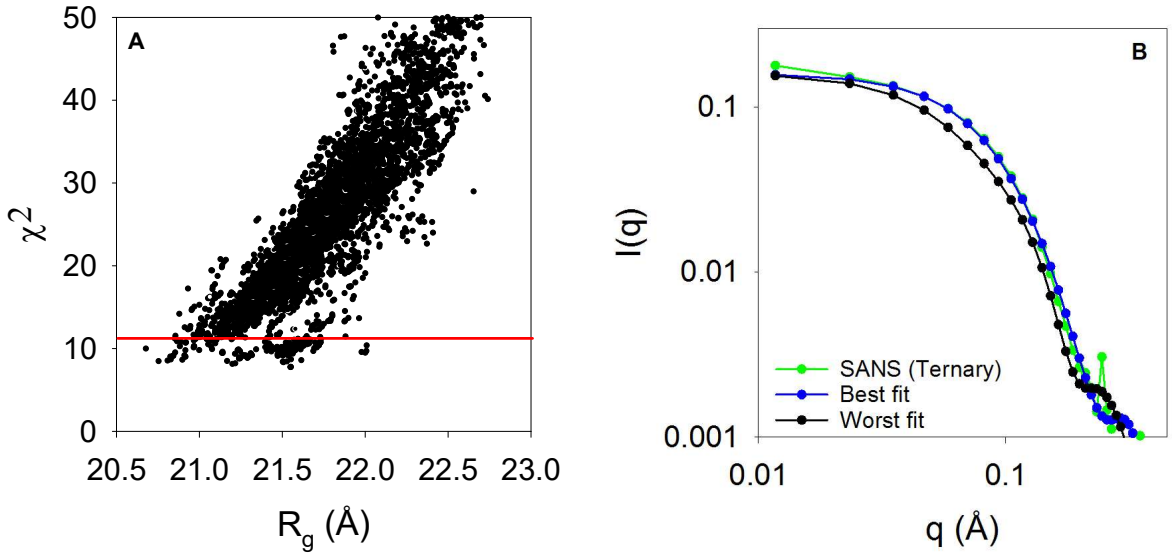
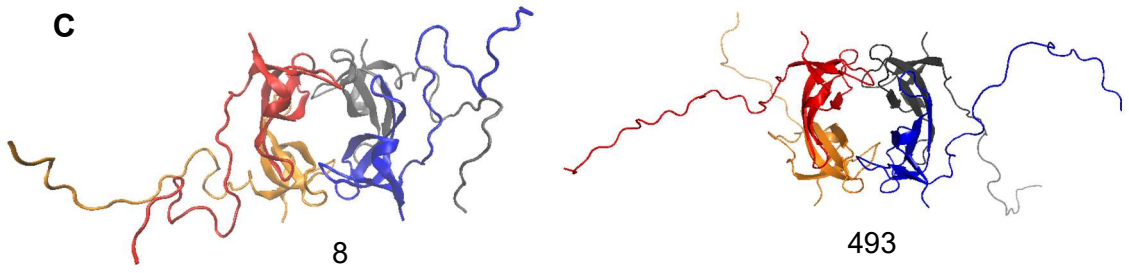
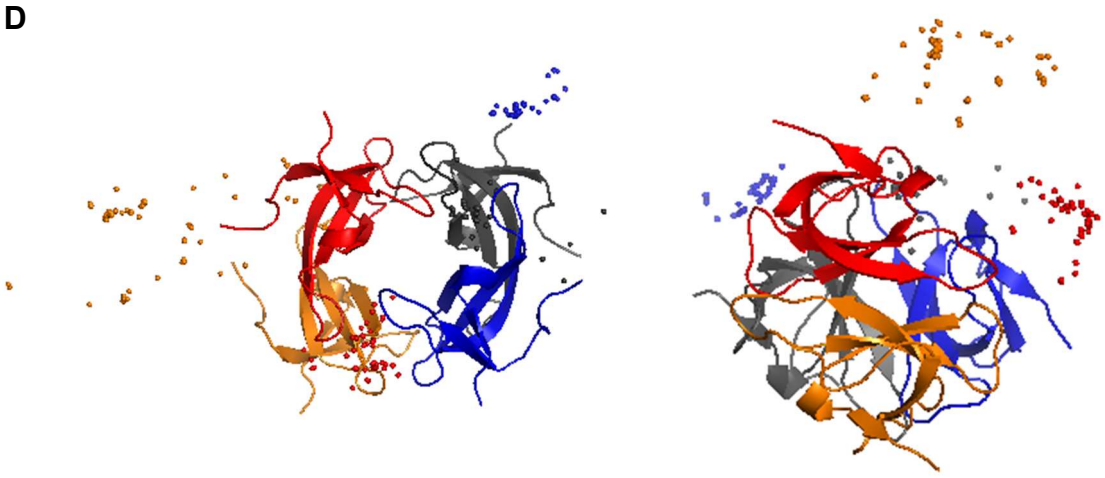


Figure 4.5 – Conformational analysis of the disordered N-termini upon DHF binding to the R67 DHFR-NADP+ complex by SASSIE. The χ^2 vs. R_g plot shown in panel A compares the representative frames generated by Monte-Carlo sampling to the experimental SANS data for R67 DHFR-$NADP^+$-DHF ternary complex. The red line indicates a cut off for χ^2 =10. An overlay of the theoretical and experimental SANS profiles for the best (χ^2 = 8) and worst (χ^2 = 493) fits, respectively) is shown in panel B. The corresponding best and worst frames are shown in panel C and panel D represents the center of mass points of the N-terminal methionine residue of each of the four N-termini of the best frames from SASSIE with a χ^2 <10. The four monomers are colored differently and the center of mass points are color coded for each monomer. Panels E shows the overly of the density plots obtained for the 86 frames best fits to the SANS data for the ternary complex (green mesh) and all the structures generated in SASSIE (gray mesh). In panel F, the density plot describing the sampling of each N-termini is shown in different colors. Panel G shows overlaid density plots for the best fits for the binary and ternary data. The density plots obtained from the SASSIE analysis for the apo (light blue) and ligand bound states (pink-binary and green-ternary) are shown in panel H.

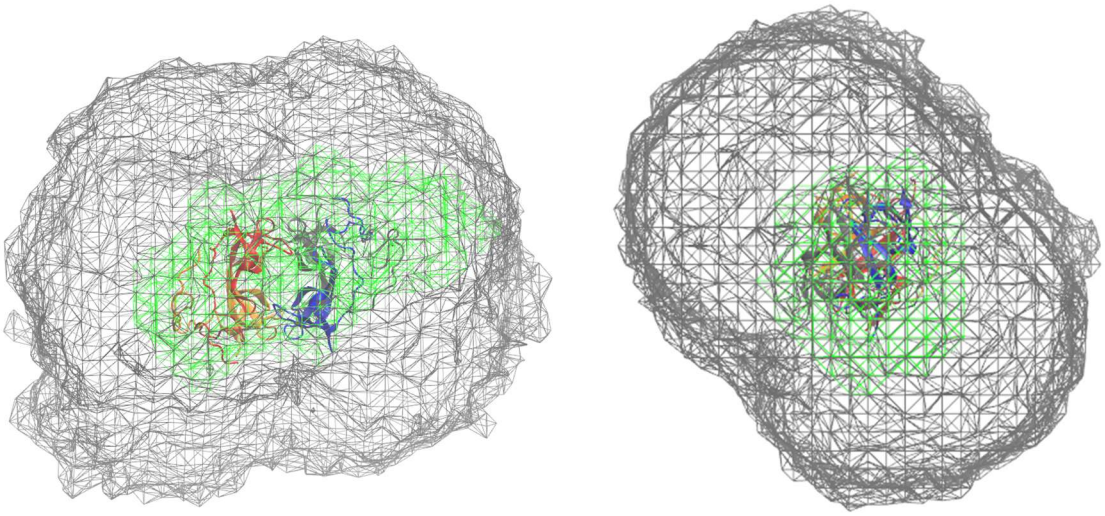

A
50
40
30
20
10
0
χ^2
20.5
21.0
21.5
22.0
22.5
23.0
R_g (Å)
B
0.1
0.01
0.001
I(q)
SANS (Ternary)
Best fit
Worst fit
0.01
0.1
q (Å)


C
8
493

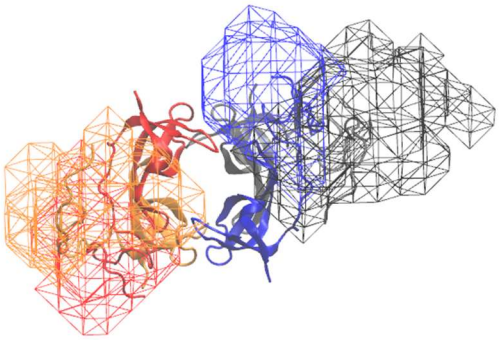

D

Continued Figure 4.5

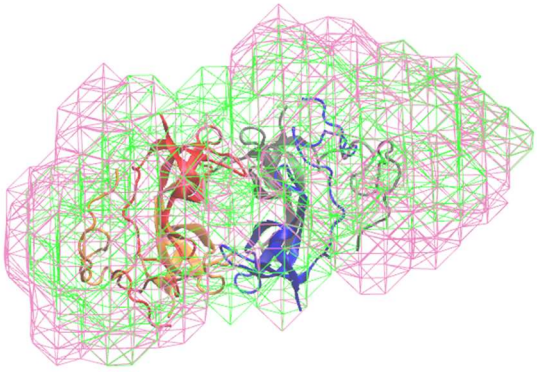
E



F

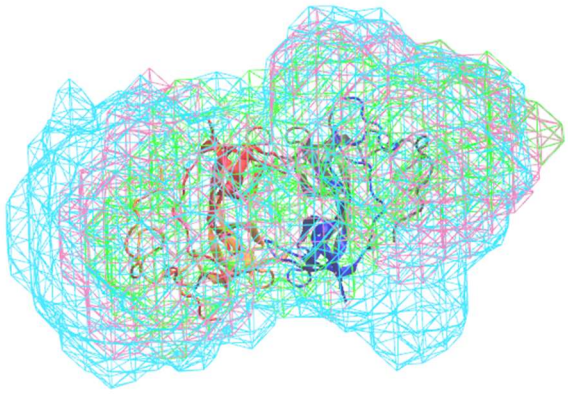


G



Continued Figure 4.5

H



fits for apo data, suggesting that the conformations sampled by the N-termini in the ligand bound form were similar to that of the apo form (see Figure 4.5.H)

4.4.C Effect of betaine on the disordered termini of R67 DHFR

To monitor the effects of betaine on the disordered N-termini, SANS data were collected for R67 DHFR in 20 % deuterated betaine. Data were analyzed using GNOM to determine the pairwise distribution, which is shown in Figure 4.6. The R_g was 22.84 ± 0.31 Å, which is slightly larger than 21.89 ± 0.12 Å, the value for R67 DHFR in the absence of betaine. Both the plot and the resulting R_g indicate a more swollen state in presence of betaine.

SASSIE analysis of these SANS data for apo R67 DHFR in 20 % deuterated betaine was performed using the same set of 175,000 frames of the full length protein. The χ^2 versus R_g plot shows a distribution of states that were assigned low χ^2 values as can be seen from Figure 4.7.A. For example, some compacted structures (lower R_g) as well as extended structures (higher R_g) fit the data. This suggests that the N-termini sample both types of conformations in the presence of betaine. Out of all the structures generated using MD and Monte-Carlo sampling, ~38,000 fit to the experimental SANS data with acceptable χ^2 values that are lower than 10. The lowest χ^2 was 3 and ~15,000 frames gave χ^2 values < 5. The plot in Figure 4.7.A shows the $\chi^2 = 5$ cut off for the good fits. The overlay of the theoretical SANS profiles for the best and worst fits and the frames associated with it are shown in Figures 4.7.B and 4.7.C, respectively. The center of mass of the four N-termini shown in Figure 4.7.D represents all frames that fit the data well, indicating the termini can sample many positions. The points represent the N-terminal methionine residue of each monomer sampling areas near the core of the protein as well as distant from the ordered

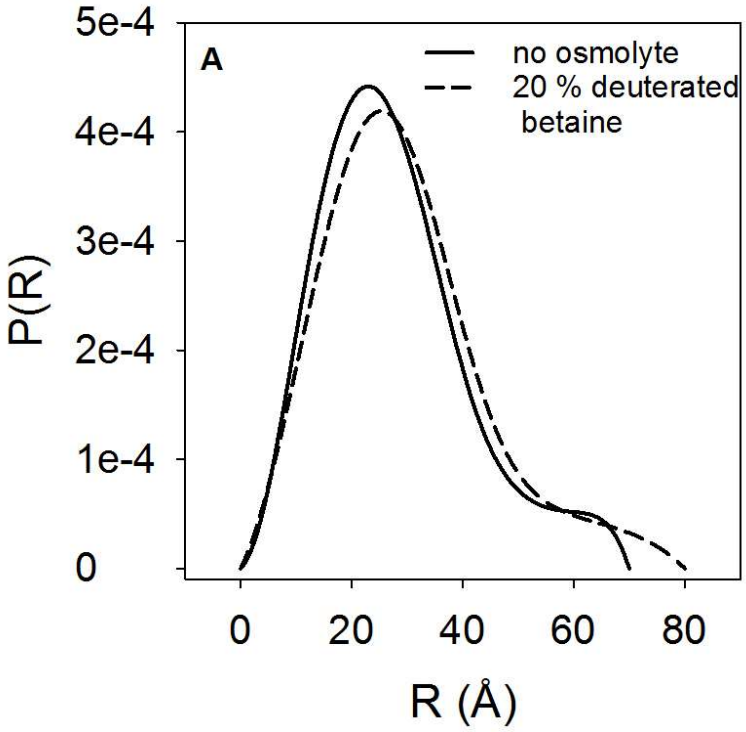
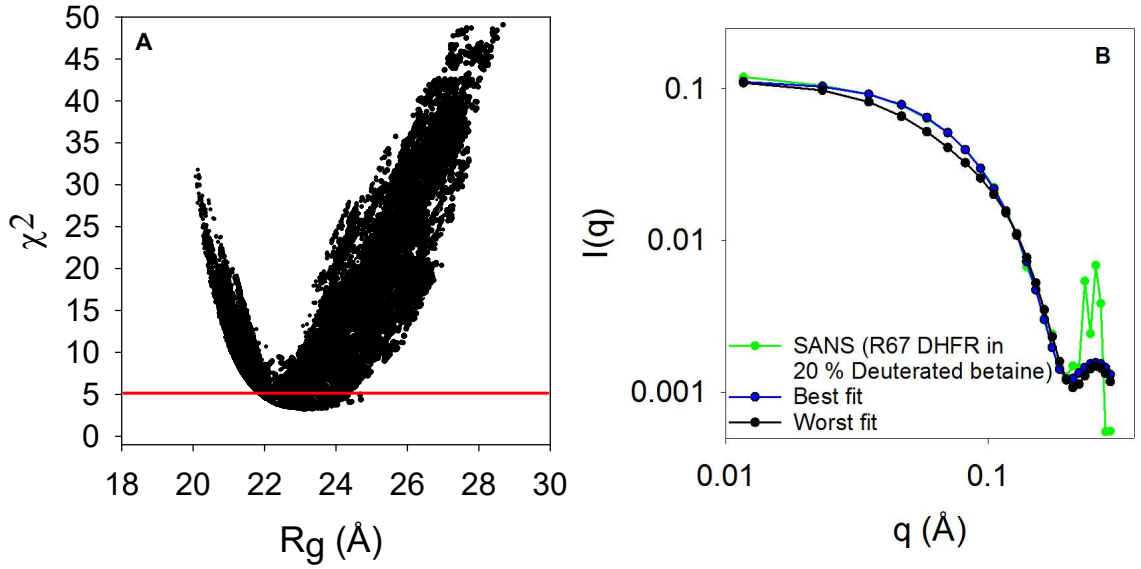
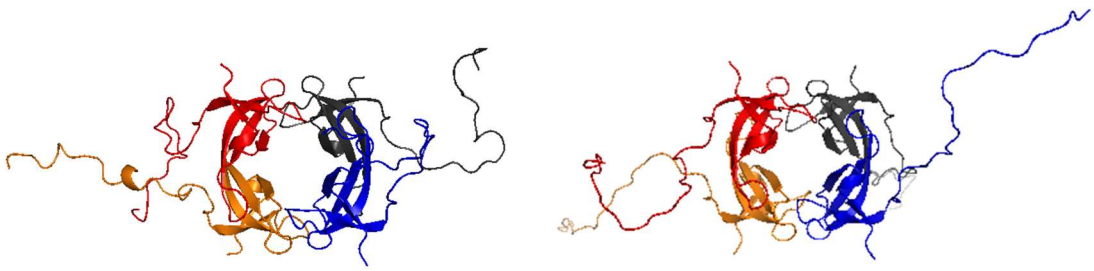


Figure 4.6 - Pairwise distribution of R_g for apo R67 DHFR with and without 20 % deuterated betaine. SANS data were collected on R67 DHFR at 6.5 mg/mL in 20 mM deuterated Tris buffer in D_2O (pD 7.0) with 20 % deuterated betaine (dashed line). A wider distribution in the presence of betaine and an R_g value of 22.80 ± 0.31 Å indicates an increased number of slightly more extended conformations for the termini of R67 DHFR.

Figure 4.7 - Conformational analysis of the disordered N-termini of R67 DHFR in the presence of 20 % deuterated betaine. Panel A shows the χ^2 vs. R_g plot that compares the experimental SANS data for apo R67 DHFR in the presence of 20 % deuterated betaine with the theoretical SANS profiles. The red line indicates a cut off for χ^2 =5. Panel B shows an overlay of the theoretical and experimental SANS profiles for the best and worst fits. The corresponding structures are shown in panel C. Panel D represents the center of mass of the first residue of each of the N-termini associated with good fits to the experimental data. The 4 chains and center of mass points are color coded. In panel E, the density plot for those frames with $\chi^2 < 5$ (magenta mesh) are overlaid on the entire structural space (gray mesh) sampled and shown in both the end on and sideways orientation of the pore of R67 DHFR. Panel F shows the density plot for each of the N-termini.


A
50
45
40
35
30
25
20
15
10
5
0
χ2
18
20
22
24
26
28
30
Rg (Å)
B
0.1
0.01
0.001
I(q)
SANS (R67 DHFR in
20 % Deuterated betaine)
Best fit
Worst fit
0.01
0.1
q (Å)

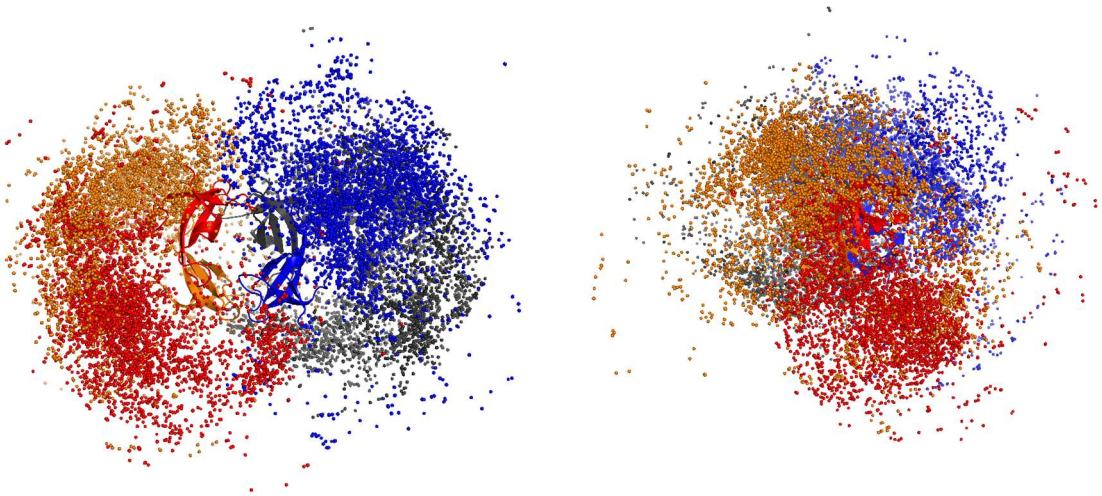
C



3

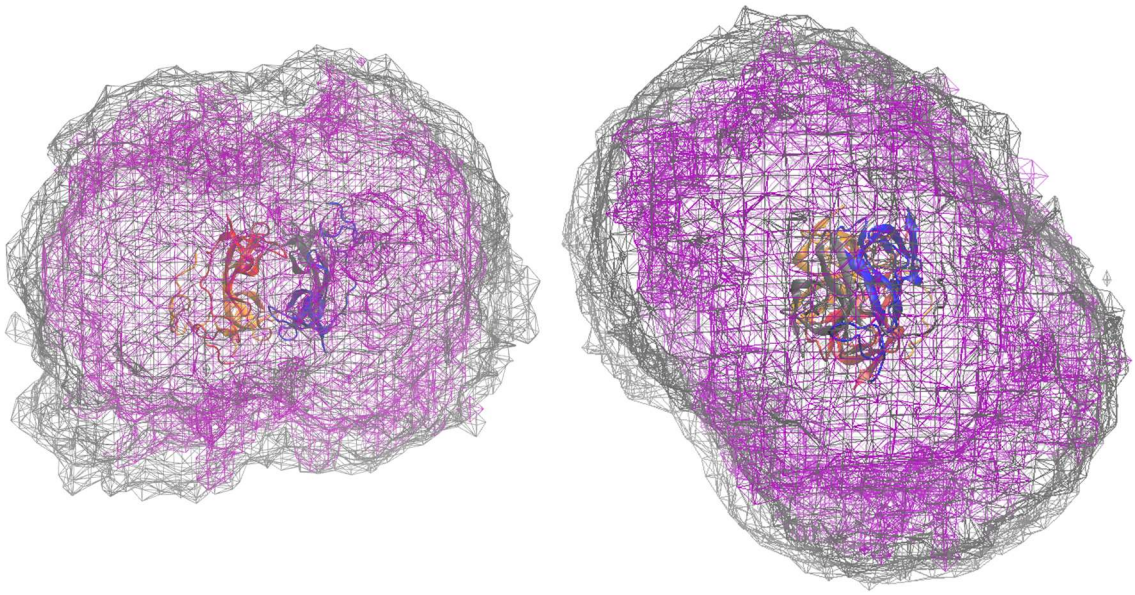
58

D

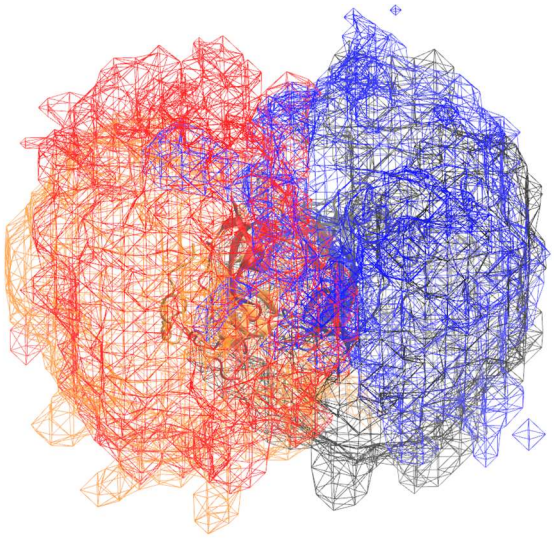


Continued Figure 4.7

E



F



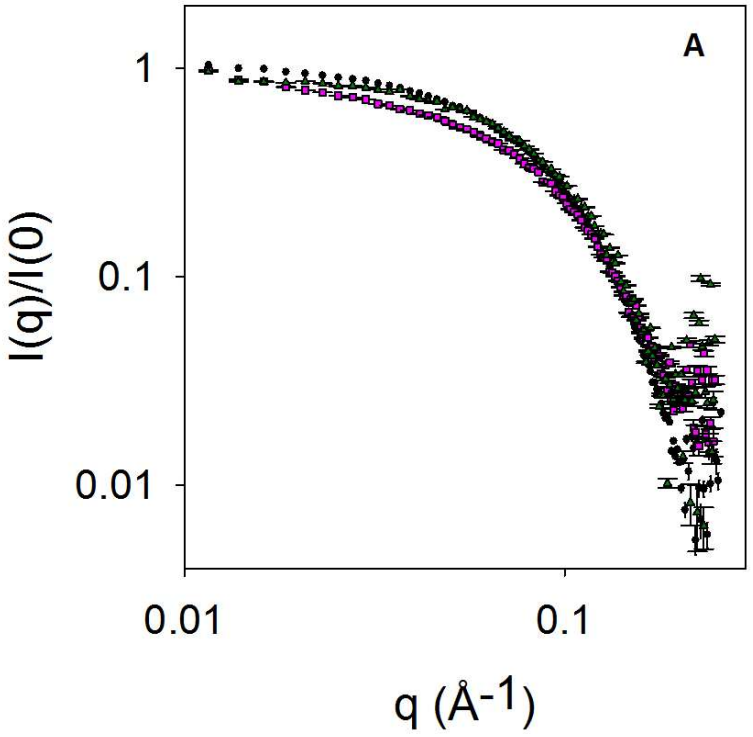
regions. This can also be clearly seen from the overlay of density plots depicting the good fits taking up most of the sampled space generated by MD and Monte-Carlo sampling. The R_g values for the good fits obtained using SASSIE ranged from 21.89 Å to 24.7 Å with an average R_g value of 22.72 ± 0.48 Å. The wider sampling range and higher average R_g both corroborated the outcome from the GNOM analysis, indicating that the N-termini sample extensive conformations in the presence of betaine. Although the density plot for the good fits (magenta mesh) seems to occupy most of the space sampled by our set of ~175,000 frames (gray mesh), the range of the R_g values obtained indicate no sampling of fully extended conformations for all the four N-termini, which would have resulted in higher R_g values. The highest Rg sampled by Monte-Carlo simulation is 29.49 Å while our model of R67 DHFR with four fully extended N-termini has an Rg of 36.25 Å.

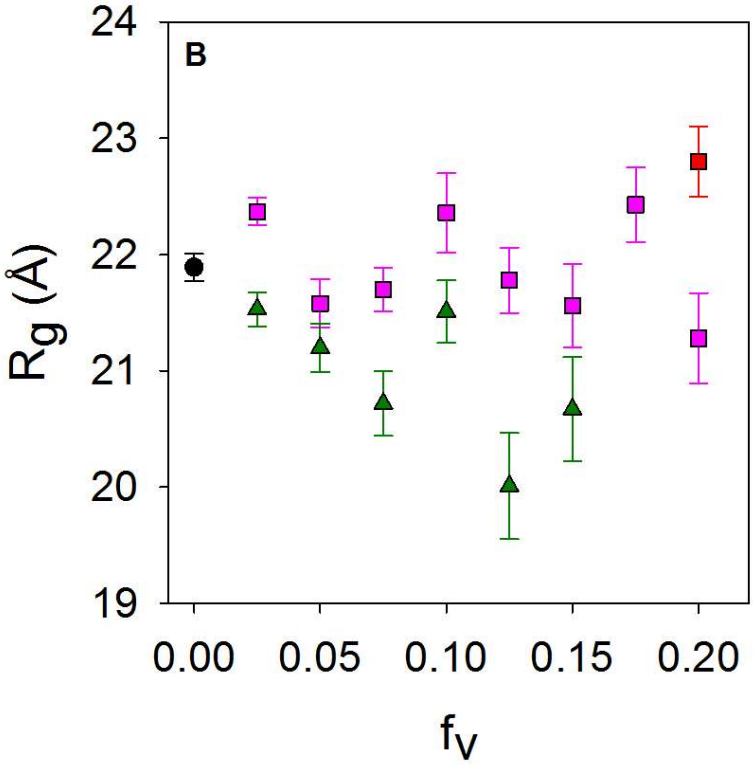
Thus, the wide sampling behavior of the best fits (see Figure 4.7.D and E) suggests that the N-termini can sample large volumes around the core of the protein. The disorder in the N-termini was found to be retained upon betaine addition.

4.4.E. Osmolytes Probe Preferential Hydration of R67 DHFR

We tested the hydration of R67 DHFR using two osmolytes- betaine and DMSO. The SANS profiles shown in Figure 4.8.A examined the effects of osmolytes (betaine or DMSO) on the overall shape and radius of gyration of R67 DHFR. The primary analysis of the reduced data using GNOM fits yields an R_g value for apo R67 DHFR of 21.89 ± 0.12 Å with no osmolyte. The R_g values showed slight differences upon osmolyte addition with values ranging from 20.01 ± 0.2 Å to 22.36 ± 0.34 Å as seen in Figure 4.8.B. The data obtained in the presence of betaine suggested no trend as all the values were within error. Our data for R67 DHFR in 20 % deuterated betaine (preceding section) indicate an increase

Figure 4.8 - SANS profiles of R67 DHFR in the presence and absence osmolytes and the analysis of the radii of gyration. Panel A shows the overlay of the SANS profiles at 6.5 mg/mL R67 DHFR in 20 mM deuterated Tris buffer in D_2O (pD 7.0) with no osmolyte (●) and in the presence of 15% betaine (■) or 15% DMSO (▲). The SANS intensity was normalized by the scattering at zero angle, I(q)/I(0). The scattering profiles did not show any large changes upon osmolyte addition. GNOM fits to these curves yield radii of gyration of 21.89 ± 0.12 Å for no osmolyte, 21.28 ± 0.38 Å with 20 % betaine and 20.67 ± 0.45 Å with 15 % DMSO, respectively. Panel B shows the variation in R_g for R67 DHFR with increasing fractional volume, f_v, of betaine (■) and DMSO (▲). The R_g in the absence of osmolytes is shown (●). Additionally, panel B shows the R_g value obtained for R67 DHFR in the presence of 20 % deuterated betaine (■).


A
1
0.1
0.01
I(q)/I(0)
0.01
0.1
q (Å-1)


B
24
23
22
21
20
19
Rg (Å)
0.00
0.05
0.10
0.15
0.20
fV

in R_g. The data obtained in the presence of DMSO indicate slightly lower R_g values, suggesting a slight compaction of the protein. However, there is also an outlier point at 10 % DMSO. Also, the errors on R_g increase with increasing concentration of osmolytes because osmolyte addition reduces the contrast between the protein and the osmolytes. Hydrogenated osmolytes also increase the incoherent background scattering that arises due to the protons in osmolytes.[15] Since we do not have data for R67 DHFR in more than 15 % DMSO, we are unable to discern whether R_g truly varies as a function of DMSO concentration or whether these R_g values are within error of each other.

Figure 4.8.B also shows the R_g value obtained from the SANS data collected for apo R67 DHFR in 20 % deuterated betaine (Rg = 22.8 ± 0.3 Å). As deuterated osmolyte (betaine) was added to the deuterated buffer, the contrast between the hydration layer and bulk was masked and the R_g value represents of the overall shape of the protein without any contributions from the hydration layer.

The zero angle scattering intensity, I(0) is sensitive to changes in hydration. A decreasing I(0) for apo R67 DHFR was observed with increasing concentrations of both osmolytes tested and the data were fit to Eq (4.2) as shown in Figure 4.9. The fits obtained yield the volume of the hydration layer for R67 DHFR in the presence of betaine or DMSO. The number of water molecules in the hydration layer is determined by dividing the observed water volume by the volume of a single water molecule (30 $Å^3$).[20] The number of osmolyte excluding water molecules associated with the protein was found to be 1202 ± 113 or 1237 ± 138 using betaine or DMSO, respectively. This result indicates a similar number of water molecules exclude betaine and DMSO from the hydration layer of R67 DHFR.

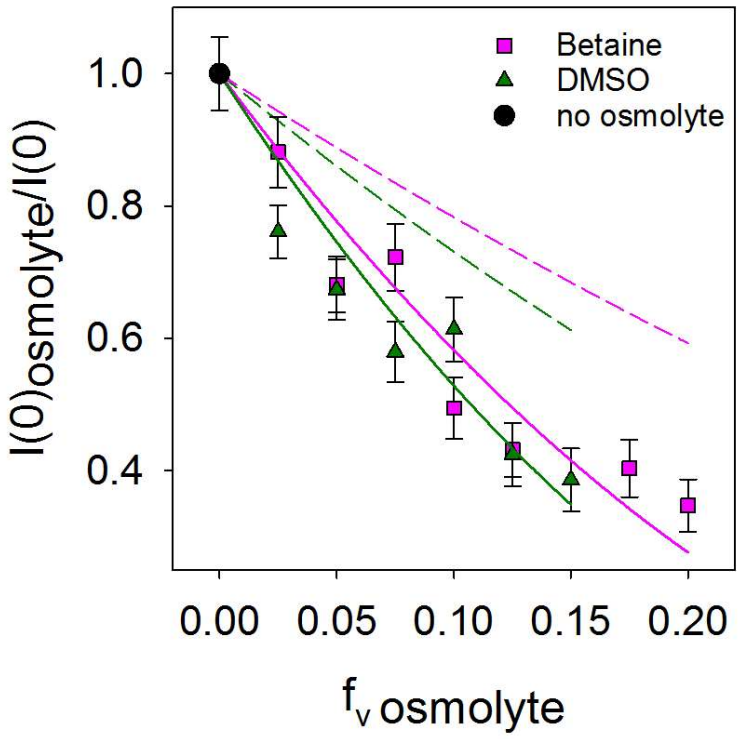


Figure 4.9 - The ratio of I(0) in the presence and absence of osmolytes. Small angle neutron scattering intensity ratio with and without osmolyte $I(0)_S/I(0)$ as a function of osmolyte concentration, f_v, for betaine (w/v) (■) and DMSO (v/v) (▲) are shown, respectively. Solid lines are fits to Eq. 4.2 to calculate the number of protein-associated waters, n_w. The dashed lines represent the theoretical fit for no water present ($v_w = 0$) to show the expected intensity ratio dependence on f_v due to the contrast generated by betaine (dashed magenta) and DMSO (dashed green).

To compare the experimental value with a theoretical value, the solvent accessible surface area (ASA) of tetrameric, truncated apo R67 DHFR (2RH2) was calculated to be 11,072 $Å^2$ using Molecular Operating Environment (MOE 2015 version). If we assume the area of a water molecule to be 9 $Å^2$ [21], this yields approximately 1230 water molecules potentially hydrating the truncated protein. For the refined tetrameric crystal structure of the truncated R67 DHFR (2GQV)[19], the solvent accessible surface area was 11,673 $Å^2$. This structure yields around 1297 water molecules. Comparing this range of values for waters in the hydration shell, to our SANS data fit, there is a reasonable match. However, our SANS data describe full length R67 DHFR while the crystal structure describes protein lacking N-terminal amino acids. When we use the good fits obtained from our SASSIE analysis, the number of water molecules in the hydration layer rises to 1700-2100. Table 4.2 compares our experimental results with the predicted values from the truncated crystal structures as well as two of the full length models of R67 DHFR.

4.4.E Effect of Osmolytes on the Thermal Stability of R67 DHFR

DSC scans were performed to monitor the effects of betaine and DMSO on the thermal stability of R67 DHFR . This is another way to determine if osmolytes are excluded from the protein surface. Previous studies on thermal denaturation of R67 DHFR at pH 8 have shown reversible folding with a melting temperature of 70.95 °C and evidence of an intermediate state.[22] An overlay of DSC scans is shown in Figure 4.10. Fitting the data to a three state model gives two melting temperatures that correspond to two events in the thermal unfolding of R67 DHFR. The R67 DHFR tetramer is known to unfold via formation of a dimeric intermediate prior to unfolding of four monomers.[22] The T_M1 and T_M2 values for R67 DHFR in the absence of osmolyte are 66.8 °C and 68.7 °C. The melting

Table 4.2 - Comparison of the predicted and experimental numbers of water molecules (nw) hydrating R67 DHFR as obtained from the crystal structure and SANS data. The n_w determined by SANS gives the number of water molecules that exclude the added osmolytes from the protein hydration layer.

Protein	**Source**	**Predicted/Experimental**	**Number of water molecules in the hydration layer (n_w)**
Truncated R67 DHFR	Crystal Structure (2RH2)	Predicted	1230[a]
Truncated R67 DHFR	Crystal Structure (2GQV)	Predicted	1297[b]
Full length R67 DHFR	SASSIE Analysis ($\chi^2 = 2$, $R_g = 21.89$ Å)	Predicted	2000
Full length R67 DHFR	SASSIE Analysis ($\chi^2 = 7$, $R_g = 21.19$ Å)	Predicted	1840
Full length R67 DHFR in presence of Betaine	SANS	Experimental	1202 ± 113
Full length R67 DHFR in presence of DMSO	SANS	Experimental	1237 ± 138

[a]2RH2 structure lacks 20 residues at the N-termini.
[b]2GQV structure lacks 19 residues at the N-termini. Serine 20 was removed for calculations for comparison with 2RH2.

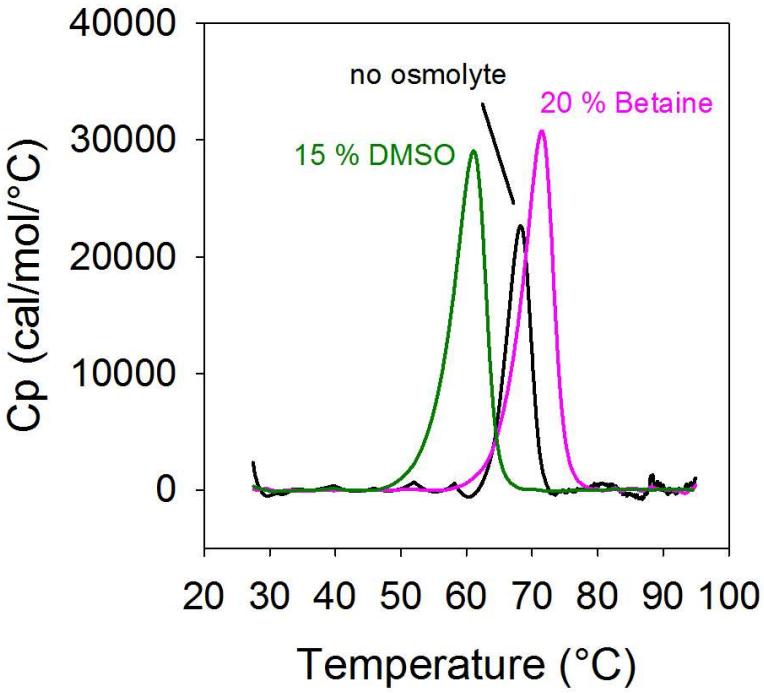


Figure 4.10 – Osmolyte effects on thermal denaturation of R67 DHFR. DSC scans were performed with 150-160 μM R67 DHFR in MTA buffer with and without 20 % betaine (magenta) and 15 % DMSO (green). Betaine increases the melting temperature of the protein by 2-3 °C whereas DMSO decreases it by 7-9 °C.

Table 4.3 - Comparison of melting temperatures of R67 DHFR with and without osmolytes. The thermal denaturation studied by DSC fits to a three state model that yields two melting temperatures.

R67 DHFR	**TM_1 (° C)**	**TM_2 (° C)**
Buffer	66.8 ± 0.2	68.7 ± 0.1
20 % Betaine	68.9 ± 0.1	71.7 ± 0.1
15 % DMSO	58.2 ± 0.2	61.3 ± 0.1

temperatures are slightly lower than the reported value. The variation may come from fitting the thermogram to the three state model in our analysis. The addition of 20 % betaine increased the melting temperature of R67 DHFR by 2-3 °C, while 15 % DMSO decreased it by 7-9 °C as can be seen from Table 4.3. Stabilization of R67 DHFR in the presence of betaine is consistent with preferential exclusion of betaine from the protein surface. DMSO slightly destabilizes R67 DHFR, which indicates the likely interaction of DMSO with R67 DHFR, either in the native or unfolded state. A similar destabilizing effect of DMSO was noted for FolM, a pteridine reductase1 (PTR1) homologue in *E.coli* that shows DHFR activity.[23]

4.5 Discussion

R67 DHFR is a homotetramer with an intrinsic disorder in the first 16-18 residues of each monomer. The disordered N-termini are not seen in the crystal structure, but are important for protein expression and stability. The truncated gene product is not expressed *in vivo* and therefore, does not confer TMP resistance upon the host cells.[4] This implies that the N-terminal sequence plays a role in stable expression of the protein inside the cells. Chymotrypsin treatment of the full-length protein results in a truncated product, which has previously been shown to be active. Thus, the N-termini are essential for protein expression and/or stability but not for catalysis. To understand the role of N-termini of R67 DHFR, we characterized its structural properties by SANS using full length R67 DHFR.

4.5.A Apo Protein Analysis

Our SANS experiments suggest that the disordered N-termini retain their flexibility and sample interconvertible conformations in solution. The conformational diversity of the

N-termini was modeled using frames from molecular dynamics simulations. Monte-Carlo sampling was also performed using SASSIE. This program also generated SANS profiles, which upon comparison to the experimental SANS data, provided structures that fit to the SANS data. The best fits for apo R67 DHFR indicate a compaction of two N-termini near the ordered tetramer core whereas the other two N-termini preferred to remain partially extended (see Figure 4.2.D). In many of these poses, the N-terminal residues interacted with the β-strands in the protein core. The termini also frequently interacted with each other. These interactions resulted in compaction of the overall shape. These interactions seem likely to be why the N-termini provide stability to R67 DHFR. Future data mining analysis of the best fits from the simulations will be required to obtain insights into these interactions. As the N-terminal sequence contains hydrophobic residues (Met, Ile, Val, 2 Phe, Ala), it could potentially form hydrophobic interactions with similar exposed side chains on the ordered protein surface. In addition, the N-termini also contain polar residues (2 Ser and 1 Arg) that could form H-bonds. In addition, electrostatic interaction of Glu is also possible. Thus, further information would be obtained from analyzing the best frames obtained from SASSIE.

The crystal structure of truncated R67 DHFR shows 222-symmetry. This symmetry could also apply to each of the disordered N-termini in that they could sample similar conformations. The best fits obtained from SASSIE analysis of the apo R67 DHFR ($\chi^2 <$ 10) show one N-termini on each side of the pore exploring collapsed conformations whereas the other termini on the same side explore slightly extended conformations. Any asymmetry in the structures probed likely arises from incomplete convergence of the MD trajectories as only microsecond time scales were explored. Likewise, symmetry related

conformations are likely present in the best frames from SASSIE analysis. A future step will explore symmetry in the SASSIE frames.

4.5.B Analysis of Binary and Ternary Complexes

No substantial effect of ligand binding was observed on the conformations sampled by the N-termini. The conformations that provided the best fits to the SANS data for both the binary and ternary complexes were similar to those sampled by the apo protein. Similar outcomes were obtained from GNOM as well as SASSIE analysis. These results are consistent with the truncated protein retaining its enzyme activity with comparable kinetic parameters.[4] However, our present SASSIE analysis did not account for any contributions from the ligands in the theoretical SANS profile determination. This may be why fewer frames were identified as good fits for the SANS data (binary and ternary complexes). The pore, when occupied by the ligand atoms, would likely yield a different SANS profile, which may provide better fitting. Thus, our next step will position the bound ligands in all the frames and these structures will be re-analyzed by SasCalc.

4.5.C Effects of Osmolytes

Osmolytes that are excluded from protein surfaces are known to stabilize the protein via the preferential exclusion mechanism. The ability of TMAO to force folding of a modified RNase was attributed to its preferential exclusion from the peptide backbone (also termed as the solvophobic effect).[24, 25] While R67 DHFR was found to be stabilized upon betaine addition by our DSC studies, no disorder to order transition was observed for the disordered tails from our analysis of the SANS data. This was indicative of the inability of betaine to force the N-termini in R67 DHFR to fold. On the contrary, according to SASSIE analysis, betaine addition resulted in larger conformational sampling

of the disordered tails, from being collapsed near the core of the protein to being partially extended. Thus, betaine seems to aid in maintenance of wider sampling of the disordered N-termini in R67 DHFR. This may be due to preferential interactions of betaine with aromatic surfaces, cationic and amide nitrogen atoms (see Part 3 of this thesis). Thus, betaine can compete with water to form stable interactions that may possibly hinder the collapsed conformations from being sampled. Comparison of the interactions involving the N-terminal residues in the frames obtained as best fits in the presence and absence of betaine would provide better insights into the differences in the intramolecular interactions. Also, one concern is that most of the starting structures in our SASSIE analysis were obtained from MD trajectories that did not include betaine. Thus, any potential intermolecular interactions between betaine and the protein were not accounted for in our MD and Monte-Carlo simulations.

Another possible explanation for the extensive sampling of the disordered tails upon betaine addition may be attributed to changes in the solvent structure. Studies have characterized effects of solutes on the structure of bulk as well as hydrating water molecules around proteins.[26, 27] The nature and extent of these alterations depend on the chemical properties of the solutes. Polar and hydrophilic surfaces were found to be water structure breakers whereas hydrophobic surfaces were described as water structure makers.[26] Sucrose at 1.5 M, stabilized RNase A, however, accompanying pressure perturbation calorimetry studies showed nonlinear effects on α, the apparent coefficient of thermal expansion, Specifically, RNase is less compact at 0.5M sucrose as indicated by an increased α than in the presence of no sucrose, while the protein becomes more compact at

1.5M sucrose yielding a decreased α. [27] The differences in α were attributed to changes in protein hydration.

Betaine can potentially affect the structural order of the bulk as well as hydration water. These changes in solvent structure (environment) may have effects on the sampling behavior of the disordered tails of R67 DHFR, allowing partially extended and collapsed conformations. Also, intrinsically disordered regions are known to exhibit high hydration capacities owing to their high solvent accessibility and polar/charged nature.[13] The dynamics and/or stability of the water molecules in the hydration network around disordered regions has been reported to be different than the rest of the folded surface.[12, 13] Betaine induced changes in the order of the hydration water molecules around the disordered tails may also contribute to the broader sampling of the N-termini. Thus, effects of betaine on R67 DHFR could be attributed to changes in the bulk solvent structure as well as interactions between the solute and the protein.

4.5.D Hydration Studies

We were interested in getting insights into the hydration of full length R67 DHFR. Experiments to study protein hydration have used varying techniques. A typical approach calculates the accessible surface area (ASA) and divides the value by 9 $Å^2$ to predict the number of solvent waters. This yields a high value. In contrast, experimental approaches often yield lower numbers of hydration waters. For lysozyme, ASA calculations predict ~900 waters of hydration. [15] Experimental techniques to study lysozyme hydration include NMR,[28] excess heat capacity,[29] dielectric relaxation[30] and x-ray diffraction.[31] The experimental approaches yield from 121-900 hydration waters,

indicating the value is sensitive to the technique used as well as the experimental conditions employed.

A previous SANS study of hydration in lysozyme used different osmolytes.[15] With added betaine, triethylene glycol, PEG400 or PEG1000, 84 ± 5, 114 ± 24, 156 ± 8 or 347 ± 11 hydration waters were observed, respectively. The increase in the number of waters (n_w) may be due to osmotic stress effects combined with volume exclusion as the osmolyte size gets larger.[32, 33] Alternately, fewer waters may be observed if the osmolyte interacts with the protein surface. Both factors may play a role in observation of a lower n_w value than the predicted, upper limit.

We used SANS coupled with an osmotic stress approach to assess the osmolyte excluding water molecules that preferentially hydrate the R67 DHFR surface. Betaine is a compatible solute that is usually excluded from the protein surface, [34] which results in protein stabilization. Our DSC results support this notion as upon addition of 20 % betaine, R67 DHFR melts at a higher temperature, implying preferential exclusion of betaine from the protein. Our SANS experiments yielded the number of water molecules (n_w) responsible for betaine exclusion to be 1200 ± 110. This value matches the number obtained from ASA calculations using the chymotrypsin truncated structure, however this species lacks the first 16 amino acids and strong electron density only appears at residues 20 or 21 (see Table 4.2). The full length R67 DHFR model built with the 4 N-termini added to the structure yielded a range of n_w from 1800 – 2000 (see Table 4.3), depending upon the conformations sampled by the N-termini.

Our experimental value is lower than the predicted upper limit, but as described above, this is a common result and could be due to some level of osmolyte interaction with

the protein surface. For example, as mentioned earlier, betaine can interact with aromatic and cationic surfaces. The disordered sequence has two phenylalanine and one arginine residues that may be accessible to betaine. Similar results were also reported for sucrose effects on RNase A stability and hydration as discussed in the previous section.[27]

The other osmolyte tested in our SANS experiments, DMSO, provided a comparable n_w of 1240 ± 140 hydrating R67 DHFR. However, our DSC results suggest that addition of 15 % DMSO lowers the melting temperature of R67 DHFR. The thermal destabilization can be attributed to preferential interaction of DMSO with the protein, resulting in exclusion of water molecules. The number of waters in the hydration layer can be altered by solute penetration into the hydration layer.

The observation of a similar number of water molecules hydrating R67 DHFR in the presence of a stabilizing osmolyte, betaine, and a destabilizing osmolyte, DMSO, was interesting. Though the number of waters were similar, their location may vary. DMSO can form hydrophobic interactions whereas betaine interacts with aromatic, amide and cationic nitrogens exposed on the protein. Thus, both the osmolytes may lead to water exclusion from different protein surfaces. This can also result in the variable effects on protein stability. The melting temperature of R67 DHFR decreased by 7-9 °C with DMSO addition, whereas betaine only stabilized the protein by 2-3 °C. Thus, the effects of betaine interactions with exposed surfaces as mentioned earlier are evident from the lower number of water molecules hydrating the protein. However, the effects of betaine interaction on the overall stability of the protein may be mild and may have been counter balanced by the preferential exclusion of betaine from amide oxygen, hydroxyl and carboxylate oxygen surfaces of the protein. This may result in a net stabilizing effect of betaine on R67 DHFR.

At this point we consider another possibility and ask what is known about water in the R67 DHFR crystal structure? A 1.1 Å resolution structure (PDB ID:2GQV) finds 148 water molecules with full occupancy (per monomer) and 43 waters with half occupancy. This yields a total of 764 waters per tetramer. Included in this number are 168 waters found in well-ordered pentagonal arrays in the active site pore. The hydrated structure is shown in Figure 4.11. Also, due to the high resolution and low temperature factors of this structure, 85 waters per monomer were identified in the first hydration shell and 106 in higher level shells. This yields 340 waters in the first hydration shell of the tetramer. When we consider our SANS data in light of this information, 1200 waters (measured by SANS for the full length protein) minus 340 waters (per chymotrypsin truncated R67 DHFR) would leave 215 waters to hydrate each N-terminal sequence. This assumes SANS measures the first hydration shell.

4.6 Conclusion

R67 DHFR, a tetrameric protein, contains a 21 residue disordered region at the N-termini of each of the monomer. A truncation at the 16th residue yields a functional protein with hampered protein stability, indicating a crucial role of the 4 N-termini. To understand the contribution of the N-termini towards protein stability, we determined the structural features of the full length protein using SANS. The conformational ensemble obtained for the apo protein suggested interactions of the disordered tails either with the ordered protein structure elements or with another disordered tail. These weak interactions result in compaction of the disordered tails near the sides of the protein. The *in vivo* stability of the full length protein may be due to the contributions of these interactions. The pattern of the

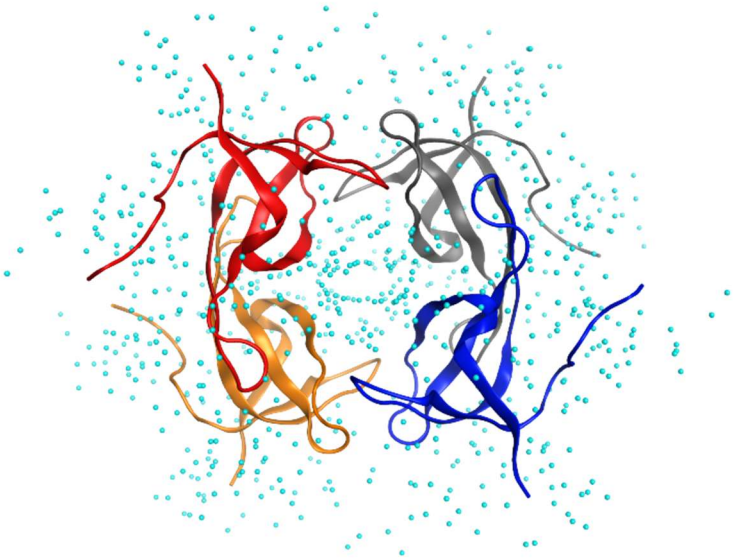


Figure 4.11 - Crystal Structure of chymotrypsin truncated R67 DHFR (2GQV) at 1.1 Å resolution.[19] The monomers are depicted in different colors and the water molecules seen in the crystal structure are shown by cyan spheres.

conformational sampling was similar for the protein in both the ligand bound complexes (binary and ternary). No influence of the disordered tails was observed on ligand binding as none of the conformations sampled show the N-termini blocking the active site pore of the protein. Addition of osmolytes such as betaine did not force the disordered tails to attain any order. On the contrary, the conformational ensembles indicate wider sampling of both extended as well as collapsed/compact forms of the N-termini. Preferential hydration of the full length protein studied by SANS yields an equal number of waters hydrating the protein upon addition of the stabilizing osmolyte betaine as well as the destabilizing osmolyte DMSO.

Acknowledgements - We acknowledge Susan Krueger from NIST for her help with SASSIE analysis.

4.7 References

1. Narayana, N., et al., *A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site.* Nat. Struct. Biol., 1995. 2(11): p. 1018-1025.
2. Feng, J., et al., *Radical redesign of a tandem array of four R67 dihydrofolate reductase genes yields a functional, folded protein possessing 45 substitutions.* Biochemistry, 2010. 49(34): p. 7384-7392.
3. Matthews, D.A., et al., *Crystal structure of a novel trimethoprim-resistant dihydrofolate reductase specified in Escherichia coli by R-plasmid R67.* Biochemistry, 1986. 25(15): p. 4194-4204.
4. Reece, L.J., et al., *Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein.* Biochemistry, 1991. 30(45): p. 10895-10904.
5. Narayana, N., *High-resolution structure of a plasmid-encoded dihydrofolate reductase: pentagonal network of water molecules in the D2-symmetric active site.* Acta Crystallogr D Biol Crystallogr, 2006. 62(Pt 7): p. 695-706.
6. Amyes, S.G. and J.T. Smith, *The purification and properties of the trimethoprim-resistant dihydrofolate reductase mediated by the R-factor, R388.* Eur. J. Biochem., 1976. 61(2): p. 597-603.
7. Broad, D.F. and J.T. Smith, *Classification of trimethoprim-resistant dihydrofolate reductases mediated by R-plasmids using isoelectric focussing.* Eur J Biochem, 1982. 125(3): p. 617-622.
8. Strader, M.B., *Identifying the Catalytic and Ligand Binding Roles of Active Site Residues in Homotetrameric R67 Dihydrofolate Reductase.* University of Tennessee, 2003.
9. Duff, M.R., Jr., et al., *Tales of Dihydrofolate Binding to R67 Dihydrofolate Reductase.* Biochemistry, 2016. 55(1): p. 133-145.
10. Bradrick, T.D., et al., *Redesigning the quaternary structure of R67 dihydrofolate reductase. Creation of an active monomer from a tetrameric protein by quadruplication of the gene.* J. Biol. Chem., 1996. 271(45): p. 28031-28037.
11. Patel, A.J., et al., *Sitting at the Edge: How Biomolecules use Hydrophobicity to Tune Their Interactions and Function.* J. Phys, Chem. B, 2012. 116: p. 2498-2503.
12. Rani, P. and P. Biswas, *Diffusion of Hydration Water around Intrinsically Disordered Proteins.* J Phys Chem B, 2015. 119(42): p. 13262-13270.
13. Rani, P. and P. Biswas, *Local Structure and Dynamics of Hydration Water in Intrinsically Disordered Proteins.* J Phys Chem B, 2015. 119(34): p. 10858-10867.
14. D.I., S., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria.* J. Appl. Crystallogr. , 1992. 25: p. 595-503.
15. Stanley, C., et al., *Protein structure and hydration probed by SANS and osmotic stress.* Biophys. J., 2008. 94(7): p. 2777-2789.
16. A. E. Whitten, S.C.a.J.T., *MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies.* J. Appl. Cryst., 2008. 41: p. 222-226.

17. Curtis, J.E., et al., *SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints.* Computer Physics Communications, 2012. 183(2): p. 382-389.
18. Krahn, J.M., et al., *Crystal structure of a type II dihydrofolate reductase catalytic ternary complex.* Biochemistry, 2007. 46(51): p. 14878-14888.
19. Narayana, N., *High-resolution structure of a plasmid-encoded dihydrofolate reductase: pentagonal network of water molecules in the D2-symmetric active site.* Acta. Crystallogr. D. Biol. Crystallogr., 2006. 62(Pt 7): p. 695-706.
20. Whitten, A.E., et al., *Cardiac myosin-binding protein C decorates F-actin: implications for cardiac function.* Proc. Natl. Acad. Sci. U. S. A., 2008. 105(47): p. 18360-18365.
21. Fried, M.G., et al., *Role of hydration in the binding of lac repressor to DNA.* J Biol Chem, 2002. 277(52): p. 50676-50682.
22. Zhuang, P., E. Eisenstein, and E.E. Howell, *Equilibrium folding studies of tetrameric R67 dihydrofolate reductase.* Biochemistry, 1994. 33(14): p. 4237-4244.
23. Bhojane, P.P., et al., *Investigation of osmolyte effects on FolM: comparison with other dihydrofolate reductases.* Biochemistry, 2014. 53(8): p. 1330-1341.
24. Baskakov, I. and D.W. Bolen, *Forcing thermodynamically unfolded proteins to fold.* J. Biol. Chem., 1998. 273(9): p. 4831-4834.
25. Wang, A. and D.W. Bolen, *A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation.* Biochemistry, 1997. 36(30): p. 9101-9108.
26. Lin, L.N., et al., *Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry.* Anal Biochem, 2002. 302(1): p. 144-160.
27. Mitra, L., et al., *Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution--experiments and theoretical interpretation.* Phys Chem Chem Phys, 2006. 8(11): p. 1249-1265.
28. Torres, A.M., et al., *Strong and weak binding of water to proteins studied by NMR triple-quantum filtered relaxation spectroscopy of (17)O-water.* Biophys Chem, 1997. 67(1-3): p. 187-198.
29. Yang, P.H. and J.A. Rupley, *Protein--water interactions. Heat capacity of the lysozyme--water system.* Biochemistry, 1979. 18(12): p. 2654-2661.
30. Pethig, R., *Protein-water interactions determined by dielectric methods.* Annu Rev Phys Chem, 1992. 43: p. 177-205.
31. Kodandapani, R., C.G. Suresh, and M. Vijayan, *Crystal structure of low humidity tetragonal lysozyme at 2.1-A resolution. Variability in hydration shell and its structural consequences.* J Biol Chem, 1990. 265(27): p. 16126-16131.
32. Fried, M.G., et al., *Role of hydration in the binding of lac repressor to DNA.* J. Biol. Chem., 2002. 277(52): p. 50676-50682.
33. Chopra, S., et al., *A balancing act: Net uptake of water during dihydrofolate binding and net release of water upon NADPH binding in R67 dihydrofolate reductase.* J. Biol. Chem., 2008. 283: p. 4690-4698.
34. Courtenay, E.S., et al., *Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the*

interpretation of "osmotic stress" experiments in vitro. Biochemistry, 2000. 39(15): p. 4455-4471.

PART 5. CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Role of Water in Biological Process

Osmotic stress studies examined the role of water in biological processes by addition of small molecule osmolytes that alter water activity. This strategy to probe the biological importance of water was developed from the observation that living cells facing dehydrating conditions accumulate these small organic molecules and ions (osmolytes). These intracellular small molecules increase the osmolality and render the cells to survive under osmotic stress but can also exhibit additional effects owing to their different chemical properties.

The two potential mechanisms of osmolyte action- preferential exclusion and preferential interaction with biomolecular surfaces relative to water were considered while interpreting the outcomes of the osmotic stress experiments. Understanding effects of osmolytes on biochemical processes shed light on the role of intracellular water. Water in biological systems can be divided into two layers, a hydration layer that surrounds all the biomolecules and the bulk water that forms the solvent of cytoplasmic milieu. If the osmolytes are excluded from macromolecular surfaces and retained mostly in the bulk solution, the outcomes of osmotic stress experiments probe the changes in the bulk water occurring during the course of the biochemical process. However, if the osmolytes potentially interact with the macromolecules and other small molecules and enter the hydration shell, the contributions from these additional effects are also accounted for in the osmotic stress experiments.

We have previously noted effects of osmolytes on ligand binding to dihydrofolate reductase (DHFR) enzymes- (R67 DHFR and EcDHFR).[1, 2] In this study we observe a

range of effects including both preferential exclusion and interaction mechanisms of osmolytes with ligand binding to a novel DHFR, FolM.

Osmolytes such as betaine are known to be excluded from protein surfaces.[3] Preferential exclusion of betaine from the surface of all two DHFRs was manifested by the tightened cofactor binding upon betaine addition, indicating dehydrating effects on the enzyme-ligand interface.[2, 4] A similar trend was noted in Part 2 of this thesis for NADPH binding to FolM becoming tighter with betaine addition. However, interesting results were obtained for the substrate binding in presence of betaine and other osmolytes that could be explained by the following preferential interaction model.

5.2 The Preferential Interaction Model

Substrate binding to DHFRs was found to be weakened by osmolytes.[1, 2] Similar results obtained for two structurally unrelated proteins led us to propose effects of osmolytes on the free DHF species rather than the protein. We proposed the preferential interaction model for weak interactions between folate(s) and osmolytes that interfere with substrate binding to DHFRs. The removal of the osmolyte molecules from DHF is essential prior to its binding to DHFR. This model was found to be applicable to a third DHFR, FolM, as osmolyte addition weakened substrate binding.[5] It is interesting that all three DHFRs bearing unique protein scaffolds showed similar trends of weakened substrate binding in presence of osmolytes.

The results in Part 2 emphasize the weak interactions between the substrate and osmolytes, while it also tests the preferential interaction model for binding of an antifolate drug, methotrexate. Weakened binding of methotrexate suggests its preferential

interactions with osmolytes. Thus, the model can also be applied to the antifolate drugs that are structural analogues of folate(s). The interactions between drugs and intracellular small molecules may hamper the efficiency of the drug to bind its target thereby highlighting the necessity of further characterization of these weak interactions.

In addition, the preferential interaction model between osmolytes and folate(s) can be applied to all the substrates in the folate metabolism pathway suggesting large effects on the functioning of the pathway under osmotic stress.

5.3 Characterization of Preferential Interactions

Although the interactions between folate(s) and osmolytes are proposed to be weak, they are suggested to be favorable than the interactions between these molecules (folates and osmolytes) and water, hence termed as preferential interactions. These favorable interactions can potentially exhibit various amplitudes as osmolytes present different functional groups that can form multiple types of interactions with folate(s) such as H-bonding, hydrophobic interactions, cation-π interactions, and ionic interactions.

Understanding the nature of these interactions would be necessary to gain further insights in the process occurring in a living cell under normal as well as osmotic stress conditions. As mentioned earlier, numerous functional groups would be exposed in the intracellular milieu and it would be challenging to imagine the range of effects imposed by these omnipresent weak interactions. One of the way to understand this complexity is to quantify the strengths of interactions between each osmolyte and a set of compounds accounting various functional groups. As each osmolyte differs in its composition and

chemical properties, a large sample of multiple possible interactions can be quantified, which would collectively help to predict the degree of these interactions inside a cell.

We, in Part 3, quantified the preferential interaction potential (μ_{23}/RT value) between one of the osmolytes in *E.coli* – betaine with different surface types displayed by folate(s) using the vapor pressure osmometry method. Betaine is a poor H-bond donor and therefore is mostly excluded from macromolecular surfaces. However, it is known to favorably interact with aromatic surfaces forming cation-π interactions.[6, 7] Folate contains two aromatic rings and hence was predicted to favorably interact with betaine. A positive μ_{23}/RT value indicates preferential exclusion of betaine whereas a negative μ_{23}/RT value indicates preferential interaction with betaine relative to water. The interaction potential between betaine and folate was found to be near zero suggesting equal preference of water and betaine. Differences in the interaction potentials were noted upon dimerization, protonation and conformational sampling of folate in solution. Quantification of interactions with betaine and compounds in addition to folate enabled us to parse out atomistic interaction potentials per unit surface area of each atom type studied. Our results suggest interaction of betaine with aromatic carbon and amide nitogen surfaces but exclusion from the aromatic nitrogens and carboxylate oxygens.

The extensive experimental data suggests the range of these weak interactions between numerous surfaces types depending upon the physical and chemical properties of the interacting partners. In Part 3, we note a trend of betaine interacting with protonated species favorably than the deprotonated species owing to betaine not being a good H-bond donor. The deprotonated groups favor water over betaine as water can be the H-bond donor.

This would suggest that betaine can form H-bond by being an acceptor with protonated surface types.

5.4 Can we use the Interaction Potentials to Predict the Osmotic Stress Effects In Vitro?

A wide range of interaction potentials can be expected for different osmolytes displaying unique functional groups with folate as well as other biomolecules (ligands). We can apply this information to predict the effects of osmolyte addition on biological process thereby also predicting the role played by water. Our ITC experiments in Part 3 explored this idea. The μ_{23}/RT of DHFR ligands interacting with betaine can be used to predict the outcome of the ligand binding experiments performed upon betaine addition. The cofactor NADPH excludes betaine (high μ_{23}/RT value) and its binding to all three DHFRs was found to be tightened by the osmolyte. The low μ_{23}/RT values for folate and DHF interacting with betaine resulted in weakened binding owing to favorable interactions between folate(s) and betaine. On the contrary, PG4 (folate with four glutamates) with a high μ_{23}/RT value showed a similar trend of weakened binding to R67 DHFR upon betaine addition. As not all atoms of PG4 participate in forming interactions with the protein, their interaction potentials should not be considered in the predictions. (See Table 5.1). Thus, changes in preferential potentials depend on the changes in surface areas (burial or exposure) associated with the biochemical event. As discussed in Part 3, the prediction of osmotic stress effects would be reasonable if the details of ligand binding are known. Consideration of the preferential interaction potential of the protein is also required for prediction of the effects of osmolytes on the ligand binding event. Part 3 computes the

Table 5.1 - Comparison of predicted and observed effects of betaine on ligand binding to R67 DHFR

Ligand	**μ_{23}/RT (m^{-1})**	**Betaine-Ligand**	**Effect of betaine on Enzyme-Ligand Binding**	
			Expected/Predicted	**Observed**
$NADP^+$	0.85	Excluded	Tightened	Tightened[a]
Folate	0	Equal	No effect	Weakened
PG4	0.98	Excluded	Tightened	Weakened

[a]Data from Chopra et.al.[1]

$\Delta \mu_{23}/RT$ values that yielded a better prediction of the ligand binding experiments.

5.5 Osmolyte Effects on Proteins add a Layer of Complexity

Osmolytes such as betaine and proline are known as the osmoprotectants in *E.coli* as they function not only to maintain cell growth under osmotic stress but also increase the growth rate. The increase in growth rate was correlated to the stabilizing effects these osmoprotectants impose on macromolecules. This is attributed to their preferential exclusion from macromolecular surfaces that aid in preferential hydration.[8] The effectiveness as osmoprotectants depends on the level of preferential exclusion obtained for *E.coli* osmolytes which was betaine>proline>TMAO>trehalose>K^+ glutamate> glycerol.[3]

Betaine, one of the osmolyte tested in our osmotic stress experiments was found to have stabilizing effects on EcDHFR, R67 DHFR and FolM.[1, 2, 5] Differential scanning calorimetry results describe an increase in melting temperature for all the proteins tested. Our previous results along with the results from Part 2 suggest tightened cofactor binding to DHFRs in presence of betaine, consistent with its preferential exclusion from the protein and NADPH surfaces. However, other osmolytes like DMSO, ethylene glycol and PEG400 weaken the binding of NADPH to FolM, indicating additional effects of these osmolytes on the protein that results in an altered binding affinity. This is further validated by our DSC experiments showing a lower melting temperature of the protein in presence of DMSO indicating preferential interaction between DMSO and FolM (see Figure 5.1). The adverse effects of osmolytes on proteins complicate the prediction of osmotic stress outcomes *in vitro* systems that may further amplify *in vivo*. As osmolytes are known to

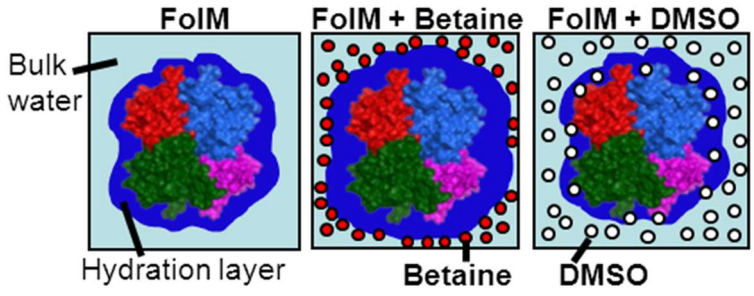


Figure 5.1 - Osmolyte effects on FolM. FolM-osmolyte interactions. FolM monomer surfaces are shown in different colors and the hydration layer represented in blue layer surrounding the surface of the tetramer. Betaine represented as red circles are preferentially excluded from the protein surface whereas DMSO (white circles) interact with FolM. Preferential exclusion of betaine results in protein stabilization by preferential hydration and preferential interaction between FolM and DMSO results in protein destabilization.

function *in vivo* to combat osmotic stress, their deleterious effects on protein stability are undesirable. Similary, the weakened binding of substrate to FolM in presence of osmolytes may be due to a combination of the deleterious effects of osmolytes on FolM and the preferential interactions between osmolytes and substrate. The effects of osmolytes on proteins complicate the prediction of osmotic stress outcomes *in vitro* systems that may further amplify *in vivo*.

5.6 SANS, a Structural Tool to Study Proteins and Osmolyte Effects

We chose to study R67 DHFR using small angle neutron scattering (SANS). R67 DHFR is a homoterameric protein with disordered N-termini that are not seen in the crystal structure, but are important for protein expression and stability.[9] SANS, in addition of providing insights on the structural characterization of the disordered N-termini, was employed to study R67 DHFR hydration. Part 4 of this thesis focusses on studying the full length R67 DHFR protein for its hydration properties as well as charachterizing the conformations sampled by disordered tails under different conditions.

5.6. A Osmolyte Effects on Protein Hydration

SANS has proven to be a powerful tool to study the preferential hydration of proteins. Part 4 of this thesis focuses on applying this approach to study preferential hydration of R67 DHFR in presence of betaine and DMSO. The scattering of neutrons from hydration layer can be differentiated from that of the protein and the bulk solvent using a contrast variation method. Hydrogenated osmolytes when added to the deuterated bulk solvent, creates a contrast as the scattering length densities of hydrogen and deuterium are

distinct. This method selectively determines the scattering data of the hydration layer and give the volume of water layer that hydrate the protein thereby excluding osmolytes.

Also, R67 DHFR possess four disordered tails (one per monomer) at the N-termini. The disordered sequences tend to sample larger volumes, they are highly exposed to solvent and are known to be highly hydrated.[10] The crystal structure of tetrameric R67 DHFR lacking the first 20 or 21 residues of each monomer yield a calculation of a total number of 1200-1300 water molecules in the hydration layer from the ASA information. The models for the full length protein predict around 1800-2000 waters. Thus, the total numbers of hydration waters around the full length protein were expected to be high.

Around 1200 waters were found to be hydrating R67 DHFR upon addition of betaine as well as DMSO. The volume of water hydrating the full length protein in presence of osmolytes coincidently resembles the number of waters hydrating the truncated protein. The experimental value was found to be inconsistent with the predicted upper limit from the ASA calculation. This is not an unusual result as previous studies exploring protein hydration has reported variations in the number of hydration waters depending upon the technique used for its determination. Also, lower number of hydration waters could be due to some level of osmolyte interaction with the protein surface. R67 DHFR hydration layer was found to contain similar number of water molecules in presence of betaine as well as DMSO. Betaine, though stabilized R67 DHFR, indicating its preferential exclusion from the protein surface, it may form weak interactions with the solvent exposed aromatic, cationic as well as amide nitrogen surfaces on N-termini and protein core. This may result in a lower number of waters than expected. Similarly, DMSO may preferentially interact

with the protein as also suggested by the thermal destabilization noted by our DSC results, thereby lowering the number of waters in hydration layer of the full length protein.

5.6.B SANS Studied the Disordered Tails of R67 DHFR

SANS also enabled us to gain information about the conformational envelope sampled by the disordered tails. A computational modeling and simulation approach was employed to obtain relevant structural information from the experimental data using modules in SASSIE. The results from this combined method of data analysis indicated that the disordered tails of R67 DHFR sample collapsed and partially extended conformations on either sides of the ordered core of the protein. The N-terminal residues were also found to interact with each other and with the β-strands in the protein core. These interactions may provide stability to the full length R67 DHFR.

Ligand binding to R67 DHFR did not seem to significantly change the conformations sampled by the N-termini. This is consistent with no significant role known for the N-termini in catalysis as the truncated protein retains activity.

While betaine stabilized R67 DHFR, the disorder in the N-termini was retained in presence of betaine. Further betaine led to a wider conformational sampling in both the collapsed and extended regimes. This may be attributed to the preferential interactions of betaine with aromatic surfaces, cationic and amide nitrogen atoms as studied in Part 3 of this thesis. Thus, our SANS studies in Part 4 provided further insights on osmolyte effects on protein hydration as well as disordered regions of the proteins.

5.7 Future Implications

Taking all the possible effects of osmolytes on proteins as well as their ligands in to account, we would like to extend our studies to understand the *in vivo* effects of osmolytes. The preferential interactions between folate and osmolytes were shown to weaken its binding affinity towards DHFRs. Also, the deleterious effects of osmolytes on proteins can potentially alter the ligand binding to the proteins. These *in vitro* observations are expected apply to *in vivo* osmotic stress situation.

5.8 How Relevant our In Vitro Studies are to Physiological In Vivo Systems?

Folate is predicted to be a sticky molecule, it may interact weakly and transiently with several functional groups it encounters. The cell is a crowded milieu in which such interactions become more relevant with multiple functional groups available to non-specifically interact with folates. This would result in larger effects on substrate binding to enzymes of folate metabolism pathway *in vivo*.

One of our current focus in the Howell lab is to quantify the weak interaction potentials for another *E.coli* osmolyte, trehalose interacting with compounds with different functional groups. The outcomes of these studies would help us to model the different possibilities of osmolytes interacting with ligands and other small molecules as compared to water. We can further apply these models to understand the collective effect of the osmolytes accumulated in *E.coli* under osmotic stress conditions.

Studies in our lab are testing the preferential interaction model by growing E. coli cells in an exogenous osmotic stressor, sorbitol, which induces synthesis and accumulation of intracellular osmolytes. Under osmotic stress *E. coli* accumulates osmolytes (for

example betaine, trehalose and glutamate). The weak preferential interactions between osmolytes and substrate affect substrate binding and thus functionality of the enzyme *in vivo*. Previous studies in our lab have shown titration of enzyme activity *in vivo* with osmotic stress. R67 DHFR function to rescue the trimethoprim treatment of cells lacking EcDHFR was studied as a function of increasing osmotic stress.[1] The enzyme activity could be titrated from higher in low to moderate osmotic stress to low under high osmotic stress finally resulting in complete loss of enzyme function. This can be attributed to lower substrate binding in presence of the intracellular osmolytes. Similar experiments are on-going to test osmolyte effects on other enzymes in folate pathway and the results obtained so far suggest similar effects as seen before with R67 DHFR. Inefficiency of enzyme function *in vivo* is attributed to weak interactions between osmolytes and folates but osmotic stress might impose additional effects on the cell such as changes in intracellular substrate and/or enzyme concentration, which could also affect cell growth. Also "domino effects" can occur where blockage of one reaction leads to substrate buildup. The increased substrate concentration can then inhibit other enzymes in the folate pathway. An example of this is inhibition of EcDHFR by trimethoprim, which results in buildup of the DHF concentration. Higher [DHF] in turn inhibit folylpolyglutamate synthetase, which adds glutamate groups to the tail of tetrahydrofolate. These possible effects can be delineated by studying an overall effect of osmotic stress on the levels of folate metabolites in the cytosol.

Quantitation of the *in vivo* folate pool will further test our hypothesis and provide additional observations from *in vivo* osmotic stress studies, which are limited to phenotypic changes. We expect to detect accumulation of the substrate and depletion of the product for a particular enzyme under osmotic stress conditions. As folate-metabolizing enzymes

relay folates from one enzyme to other in the pathway, it would be interesting to get an overall picture of folate metabolomics in osmotic stress. Our initial aim is to examine the mechanism by which osmotic pressure affects DHFR function in *E.coli*, by monitoring the concentrations of DHF and THF for various *in vivo* conditions. Further, these studies can be extended to other enzymes in the folate metabolic pathway. We also aim to monitor the *in vivo* kinetic flux of various folates in the pathway.

Absolute quantification of intracellular folates will be performed by a LC MS/MS approach in collaboration with Shawn Campagna in the Chemistry Department at UTK. The Rabinowitz group has developed a method for absolute quantification of folates from *E.coli*.[11] We will follow the protocols of the Rabinowitz group, which use chromatographic retention time, parent ion mass and fragmentation patterns to identify DHF and all the folate species. Osmotic stress effects on folate metabolomics will be studied by quantifying levels of folates in *E. coli* cells grown in high sorbitol concentration. The Rabinowitz lab has also quantitated NADPH, $NADP^+$, osmoprotectants (betaine glutamate, proline) concentrations in *E. coli*.[12] We will monitor the intracellular levels of these metabolites to get a more complete picture of the *E.coli* metabolome under osmotic stress that would help us interpret our results for both *in vitro* and *in vivo* studies.

In addition to osmotic stress conditions, our studies can be extended to model *in vivo* conditions. In the cells not exposed to osmotic stress, the typical high intracellular concentration of biomolecules result in crowding conditions. Numerous functional groups would be exposed in the intracellular milieu that can form preferential interactions with folate substrates as well as other small molecules. Our present studies using betaine along with future studies using different osmolytes (with different functional groups) can

potentially predict the effects imposed by these omnipresent weak interactions on the cellular and biochemical pathways. Thus, the outcomes of our studies provide better understanding in unveiling the complex effects of weak interactions in a crowded intracellular environment.

5.9 References

1. Chopra, S., et al., *A balancing act between net uptake of water during dihydrofolate binding and net release of water upon NADPH binding in R67 dihydrofolate reductase.* J. Biol. Chem., 2008. 283(8): p. 4690-4698.
2. Grubbs, J., et al., *Thermodynamics and solvent effects on substrate and cofactor binding in Escherichia coli chromosomal dihydrofolate reductase.* Biochemistry, 2011. 50(18): p. 3673-3685.
3. Courtenay, E.S., et al., *Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro.* Biochemistry, 2000. 39(15): p. 4455-4471.
4. Chopra, S., et al., *Effects of temperature and viscosity on R67 dihydrofolate reductase catalysis.* Biochemistry, 2006. 45(21): p. 6596-6605.
5. Bhojane, P.P., et al., *Investigation of osmolyte effects on FolM: comparison with other dihydrofolate reductases.* Biochemistry, 2014. 53(8): p. 1330-1341.
6. Capp, M.W., et al., *Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values.* Biochemistry, 2009. 48(43): p. 10372-10379.
7. Schiefner, A., et al., *Cation-pi interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from Escherichia coli.* J. Biol. Chem., 2004. 279(7): p. 5588-5596.
8. Timasheff, S.N., *The control of protein stability and association by weak interactions with water: how do solvents affect these processes?* Annu. Rev. Biophys. Biomol. Struct., 1993. 22: p. 67-97.
9. Reece, L.J., et al., *Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein.* Biochemistry, 1991. 30(45): p. 10895-10904.
10. Rani, P. and P. Biswas, *Local Structure and Dynamics of Hydration Water in Intrinsically Disordered Proteins.* J Phys Chem B, 2015. 119(34): p. 10858-10867.
11. Lu, W., Y.K. Kwon, and J.D. Rabinowitz, *Isotope ratio-based profiling of microbial folates.* J Am Soc Mass Spectrom, 2007. 18(5): p. 898-909.
12. Bennett, B.D., et al., *Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli.* Nat Chem Biol, 2009. 5(8): p. 593-599.

VITA

Purva Prashant Bhojane was born in 1988 in Dhule, India. She attended G.E.S HAL High School and Junior College in Ozar Township, Nashik. She then obtained her Bachelor's degree in Biotechnology, Biochemistry and Genetics from Garden City College, affiliated to the Bangalore University, Bangalore. After completing her Bachelor's program, she went on to successfully pursue a Master's degree in Biochemistry from Fergusson College, affiliated to the University of Pune, Pune. Upon completion of her Master's program, Purva gained research experience when she worked as a Project Assistant II at National Chemical Laboratory, Pune for 9 months. Purva then joined the University of Tennessee, Knoxville to obtain her doctoral degree in Biochemistry, Cellular and Molecular Biology. Upon completion of her doctoral degree, she plans to pursue research in an industrial setup.