Ligand binding studies of a plasmid encoded dihydrofolate reductase by fluorine NMR: towards quantifying ligand binding inside the cell

Gabriel Jose Fuente Gomez

University of Tennessee, Knoxville, gfuente@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Recommended Citation


https://trace.tennessee.edu/utk_graddiss/9910

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.
To the Graduate Council:

I am submitting herewith a dissertation written by Gabriel Jose Fuente Gomez entitled "Ligand binding studies of a plasmid encoded dihydrofolate reductase by fluorine NMR: towards quantifying ligand binding inside the cell." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biochemistry and Cellular and Molecular Biology.

Gladys Alexandre, Major Professor

We have read this dissertation and recommend its acceptance:

Dan Roberts, Francisco Barrera, Engin Serpersu

Accepted for the Council:

Dixie L. Thompson
Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)
Ligand binding studies of a plasmid encoded dihydrofolate reductase by fluorine NMR: towards quantifying ligand binding inside the cell

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Gabriel José Fuente Gómez
August 2020
Dedication

This dissertation is dedicated to my parents, Manuel Fuente and Maria Teresa Gómez, my sisters Mayra and Jessica and brother Manuel, my nice Jessica and to the memory of my PhD advisor Dr. Liz Howell.
Acknowledgments

I would like to express my gratitude to Liz Howell for giving the opportunity to learn from her and guidance through my PhD. Unfortunately, you are not with us anymore to see my graduation but your memory and kindness is deep in my heart. Thank you Liz for all your patience with me and willingness to help me in all times. Thank you for showing me to work with pottery and holding lab parties for the lab. I will always keep you in my mind for such a fighter, hardworking and exceptional woman.

Next, I would like to thank to my friend and mentor Mike Duff because also without his advice and guidance I would not being able to finish all this work. Thank you for your patience and willingness to teach us over and over again.

I would like to thank to Gladys Alexandre for all her support to Mike and me to be able to complete our studies and her guidance. Special thank you to my PhD committee Dr. Dan Roberts, Dr. Engin Serpersu and Dr. Barrera for all their valuable suggestions and guidance.

Thank you to my undergrads Lexi Miranda and Creighton Kellum for all their hard work.

Thank you to the University of Tennessee and the BCMB department for giving the opportunity to teach Biochemistry 401 and allowing to work with Dr. Wright, it was a pleasure to work for him. Furthermore, I would like to thank Dr. Jain and Dr. Serpersu for all your advice on the use of NMR.

Thank you all my friends inside and outside of the BCMB department: Justin, Jesy, Brian, Rosela, Mike, Katheryn,

Thank you to the American people and all the tax payers for their contributions that allow students like me to obtain scholarships to come to the U.S.

Thank you to all my professors in the BCMB department for your lessons and guidance.

Thank you to the BCMB office members, LaShel, Lori, Cheryl and Chrissy for always been ready to help us.
Abstract

Antibiotic resistance is a worldwide problem. The excessive use and misuse of antibiotics have resulted in the development of drug resistance. One example involves the antibiotic trimethoprim (TMP). TMP selectively inhibits the bacterial enzyme dihydrofolate reductase (DHFR) over its human counterpart. Inhibition of DHFR results in inhibition of DNA synthesis and ultimately in cell death. Unfortunately, extensive use and dissemination of TMP over the last forty years has resulted in the development of drug resistance in the form of highly transmissible plasmids. R67 dihydrofolate reductase is a plasmid-encoded enzyme that confer resistance to TMP. Our lab has studied how R67 DHFR works in vitro using diluted conditions. However, most proteins function inside cells under crowded and heterogeneous environment. As $^{19}$F labeled proteins in complex mixtures can be characterized by NMR, we report our initial expression and characterization of R67 DHFR. To obtain such information, we have labeled R67 DHFR with fluorine at different positions in the indole ring and showed that fluorine incorporation into R67 DHFR does not affect the structure and function of the protein. We have characterized the $^{19}$F NMR spectra of apo R67 DHFR and optimized fluorine incorporation. Furthermore, fluorinated R67 DHFR shows different signals for binding of cofactor, NADP$,^+$, and substrate, dihydrofolate (DHF). We have obtained similar dissociation constants for the cofactor and substrate of R67 DHFR by both NMR and ITC. In our next step, we will measure dissociation constants in more complex solutions. We envision using fluorinated protein to quantify the effect of osmolytes and crowders on the binding affinity of substrate and cofactor towards R67 DHFR.
Table of contents

Chapter 1. Introduction to fluorine NMR as a tool to study ligand binding in crowded, cell-like environments ................................................................. 1
  Macromolecular crowding ........................................................................ 2
  Volume-excluded effect ............................................................................ 2
  Weak transient-interactions ..................................................................... 5
  Methods to study weak transient interactions ....................................... 6
  Effects of crowding on enzyme activity .................................................. 9
  Weak interactions between osmolytes and folate .................................. 13
  Dihydrofolate reductases ....................................................................... 16
    Structure of DHFRs ............................................................................. 16
      Ligand binding studies with R67 DHFR ........................................... 18
      Enzyme mechanism of R67 DHFR .................................................... 20
      Role of the disordered N-terminus of R67 DHFR ............................. 22
  Nuclear magnetic resonance ................................................................... 23
    NMR labeling strategies ..................................................................... 24
    Advantages of using fluorine NMR ..................................................... 25
    In-cell NMR ..................................................................................... 25
  References ............................................................................................. 27

Chapter 2. Effect of fluorotryptophans on the stability and function of R67 dihydrofolate reductase ......................................................... 34
  Abstract ............................................................................................... 36
  Introduction ........................................................................................... 36
  Materials and Methods ......................................................................... 37
    Expression and purification of $^{19}$F-labeled R67 DHFR .................. 37
    Determination of fluorine incorporation levels .................................. 37
    pH dependence of the tetramer to dimer equilibrium ....................... 37
    Circular dichroism ............................................................................. 39
    Differential scanning calorimetry (DSC) ............................................ 39
    Isothermal titration calorimetry ......................................................... 39
    Steady-state kinetics .......................................................................... 39
  Results .................................................................................................... 39
    Percent fluorine incorporation ........................................................... 39
    Effects of fluorine incorporation on the structure and function of R67 DHFR ................................................................. 41
    Fluorine incorporation effects on the activity of R67 DHFR ............ 41
    Impact of fluorine labeling on ligand binding ................................... 48
  Discussion ............................................................................................ 48
    Effects of fluorine incorporation on R67 DHFR stability and structure ................................................................. 48
  Conclusions .......................................................................................... 57
  References ............................................................................................. 58

Chapter 3. Ligand binding studies of a plasmid encoded dihydrofolate reductase by fluorine nuclear magnetic resonance .......... 61
  Abstract ............................................................................................... 63
  Introduction ........................................................................................... 63
  Materials and Methods ......................................................................... 64
Expression and purification of $^{19}$F-labeled R67 DHFR ........................................... 64
NMR experiments .................................................................................................................. 65
Determination of fluorine incorporation levels ..................................................................... 65
Ligand binding studies by $^{19}$F NMR .................................................................................. 65
Results ................................................................................................................................... 66
$^{19}$F NMR assignment of R67 DHFR ................................................................................... 66
Titration of NADP$^+$ and DHF to fluorinated R67 DHFRs .................................................... 71
Discussion ............................................................................................................................... 82
NMR assignment ..................................................................................................................... 82
Quantifying ligand binding affinity ......................................................................................... 84
Effects of ligand binding on changes in the $^{19}$F NMR spectra of R67 DHFR ....................... 84
Choosing an optimal fluorine probe to differentiate ligand binding in R67 DHFR .............. 86
Conclusion ............................................................................................................................... 86
References .............................................................................................................................. 87
Chapter 4. Conclusion and future directions ........................................................................ 91
Macromolecular crowding .................................................................................................... 92
Effects of fluorine incorporation on protein structure and function ..................................... 92
Effects on stability .................................................................................................................. 93
Ligand binding studies with $^{19}$F labeled R67 DHFR .......................................................... 93
Towards quantifying ligand binding inside the cell by $^{19}$F NMR ......................................... 94
Future directions ..................................................................................................................... 96
Using $^{19}$F R67 DHFR for drug discovery ............................................................................. 101
References .............................................................................................................................. 104
Vita ......................................................................................................................................... 109
List of figures

Figure 1.1. Entropic effects of crowding ................................................................. 3
Figure 1.2. Enthalpic interactions can occur between crowders and other molecules in solution. 4
Figure 1.3. Structures of typical macromolecules used in crowder studies. Including examples of polyethyleneglycols, ficoll, bovine serum albumin (BSA), dextran, sugars and hemoglobin................................................................. 7
Figure 1.4. Osmolytes and crowders have similar functional groups that may form weak interactions with other molecules. ................................................................. 12
Figure 1.5. Model for how preferential interaction of osmolytes and crowders with the free ligands, like DHF, affects their binding affinity ................................................................. 14
Figure 1.6. Structures of the substrate and cofactor of dihydrofolate reductases ................................................................. 15
Figure 1.7. Structures of dihydrofolate reductases .................................................... 17
Figure 1.8. Stacked arrangement of NADP⁺ and DHF, Y69, Q67, and W38 in the binding site of R67 DHFR ................................................................. 19
Figure 1.9. The glutamate or aspartate residue in enzymes that bind folate that interacts with N3 and N2 of the pterin ring. ................................................................. 21
Figure 2.1. Structure of R67 DHFR with the location of tryptophan residues .................... 38
Figure 2.2. Integrated NMR spectra for R67 DHFR labeled with fluoroindoles .................. 40
Figure 2.3. Fluorine incorporation minimally affected the secondary structure of R67 DHFR... 42
Figure 2.4. pH titration of R67 DHFR monitored by intrinsic fluorescence for unlabeled R67 DHFR ................................................................. 44
Figure 2.5. DSC thermogram for unlabeled R67 DHFR ............................................ 45
Figure 2.6. Fluorination of R67 DHFR altered the thermal stability of the protein as measured by differential scanning calorimetry ................................................................. 46
Figure 2.7. Example titration of NADP⁺ and DHF binding to 5F R67 DHFR at 25 °C .......... 49
Figure 2.8. The proximity of the fluorine atoms at position 4 of the indole ring of W38 to the carbonyl of the amide backbone of A37 suggests a potential for steric clash. ............ 53
Figure 2.9. The proximity in R67 DHFR (PDB ID: 2RK1) of the fluorine atom at position 4 of W38 (black stick model) could disrupt the packing of H62 (magenta stick model) from the adjacent protomer at the dimer interface when A) a proton is attached to C4 and B) a fluorine (cyan) is bonded to C4 ................................................................. 54
Figure 2.10. Fluorine replacement of a hydrogen atom increases the surface area of van der Waals contact which could explain the increase in thermostability for 5F and 6F R67 DHFR (PDB ID: 2RK1)18. ................................................................. 55
Figure 2.11. The proximity of the fluorine atom at position 7 could sterically clash with the nearby carbonyl of S65 in R67 DHFR (PDB ID: 2RK1)18. ................................................................. 56
Figure 3.1. Fluorine incorporation in R67 DHFR exhibited distinct NMR spectra. ............... 67
Figure 3.2. 19F-NMR spectra of the W45F mutant of R67 DHFR was used to assign resonances observed in the spectrum of the WT enzyme ................................................................. 69
Figure 3.3. Effects of incomplete fluorine incorporation in R67 DHFR ................................ 70
Figure 3.4. Temperature effects on fluorinated R67 DHFRs exhibit increased rates of exchange of W38 and W45. ................................................................. 72
Figure 3.5. Sequence of pRSETb R67 DHFR. ............................................................ 73
Figure 3.6. W45 peaks sharpened, and split into two peaks, when the disordered N-termini were truncated from 4F R67 DHFR. ................................................................. 74
Figure 3.7. W45 peaks sharpened when the disordered N-termini were truncated from 5F R67 DHFR ................................................................. 75
Figure 3.8. Representative $^{19}$F NMR spectra for the titration of 4F and 5F R67 DHFR with ligands. ........................................................................................................ 76
Figure 3.9. Representative $^{19}$F NMR spectra for the titration of 6F and 7F R67 DHFR with ligands. ........................................................................................................ 77
Figure 3.10. $^{19}$F NMR spectra of 4F, 5F and 6F W45F R67 DHFR in the absence of ligands and saturated with NADP$^+$ and DHF. ................................................................. 79
Figure 3.11. A representative $^{19}$F NMR spectra of 4F R67 DHFR at different temperatures. ..... 80
Figure 3.12. The two conformations of W45 in the crystal structure of R67 DHFR. ............... 83
Figure 4.1. HSQC NMR spectrum for pRSETb R67 DHFR was complicated by the disordered, intact N-termini ........................................................................................................ 95
Figure 4.2 A representative in-cell NMR spectra of pRSETb 5F R67 DHFR. A) 5F R67 DHFR in buffer, B) cell lysates, C) supernatant and D) inside of E. coli cells. ......................... 97
Figure 4.3 A representative in-cell NMR spectra of normal length 5F R67 DHFR. ............... 98
Figure 4.4. Inhibitors of R67 DHFR. Ki$s$ are from references $^{64}$ and $^{71}$. ................................. 103
List of tables

Table 2.1. Best-fit values for the pH-dependent dissociation of the tetramer + 2nH⁺ ↔ 2 dimers-Hₙ ......................................................................................................................... 43

Table 2.2. Comparison of steady-state kinetics parameters of fluorinated R67 DHFR in MTA buffer, pH 7 at 30 °C. ......................................................................................................................... 47

Table 2.3: ITC thermodynamic parameters from the global fits of NADP⁺ binding to either unlabeled or labeled apo-R67 DHFR at 25 °C in MTA, pH 8.0 buffer. ................................. 50

Table 2.4: Thermodynamic parameters obtained from ITC for DHF binding to the binary complex of R67 DHFR-NADP⁺ in MTA, pH 8.0 buffer at 25 °C. ......................................................... 51

Table 3.1: ¹⁹F chemical shifts (ppm) for R67 DHFR in the indicated ligand complexes........... 68

Table 3.2: Dissociation constants of R67 DHFR obtained by NMR........................................ 81
Chapter 1. Introduction to fluorine NMR as a tool to study ligand binding in crowded, cell-like environments
Macromolecular crowding

The interior of the cell is a very crowded environment, where the concentration of macromolecules can reach more than 300 mg/mL, but the impact such high concentrations of macromolecules have on biomolecular interactions, such as macromolecular complex formation and metabolite or drug binding, are not well appreciated. All this crowded environment takes up space inside the cell and it is generally described as volume occupied or volume excluded for the simple fact that two molecules cannot occupy the same space at the same time. Despite these crowded conditions biochemists perform most of their in vitro experiments under dilute conditions where the concentration of macromolecules is less than 10 mg/mL. The disparity between the crowded environment of the cell and dilute biochemical experiments has led biochemists to wonder about effects that large concentrations of macromolecules may have on a system.

Macromolecular crowding is expected to affect several chemical equilibria and kinetics of associations such as protein folding, protein oligomerization and ligand binding due to the presence of different types of interactions between macromolecules. The most obvious and often underappreciated interaction arises from the impenetrable nature of molecules and is referred to as volume excluded effect, or sometimes called hard-core repulsions (Figure 1.1). The nature of these interactions are purely entropic as reduces the space available. These entropic effects favor the formation of the most compact state. If unfolded protein takes up more volume than the folded protein, then volume excluded effects will favor the more compact folded state, and push the equilibrium towards folded protein. Macromolecules inside the cell can form transient higher-order structures, a so called fifth-order of protein structure, or quinary structure. Formation of quinary structures can lead to either protein complexes that are more efficient at producing metabolites, or to cause phase separation of macromolecules that must be regulated by the cell and open up regions of the cell that are mostly water that can be used as pools of metabolites.

The second type of interactions are chemical interactions and they could be electrostatic, hydrogen bonds or hydrophobic (Figure 1.2). These interactions are mainly enthalpic. These enthalpic interactions can also perturb biochemical equilibria by forming more interactions with one state over another; whether it is the folded versus the unfolded state of macromolecules or between a molecule binding versus the free molecule. Although macromolecular crowding is currently referred to as the mixture of entropic and enthalpic effects seen by concentrated solutions, the original term was conceived to refer only to the volume excluded effects.

Volume-excluded effect

The effects of macromolecular crowding on equilibria were investigated for a two-state protein folding reaction and the formation of protein heterodimers. Theory of protein folding predicts that folded proteins occupy less space as they are more compact while unfolded proteins sample conformations that are more expanded. Addition of crowders results in more of the solution’s volume occupied, reduction in the number of water molecules in the system, and less volume available for the protein. This reduced volume limits the space available for the unfolded protein and as a consequence, favors the more compact folded state. Furthermore, if the activity or function of a protein depends on the protein being in the folded state, the according to the volume excluded
Figure 1.1. Entropic effects of crowding. Macromolecules that crowd a solution (grey spheres) will favor the formation of the most compact state of a protein of interest (black line at center). If the unfolded state (left panel) has a larger volume than the folded state (right panel), then crowders will shift the equilibrium from the unfolded state to the folded state. Thus, crowder macromolecules will stabilize the more compact folded state.
Figure 1.2. Enthalpic interactions can occur between crowders and other molecules in solution. Crowders (grey spheres) contain functional groups such as amines and carboxylates. The functional groups on crowder macromolecules can form attractive or repulsive interactions with functional groups on other molecules in solution, such as the protein being studied (black line in the center). Positively charged amine groups will form electrostatic interactions with negatively charged residues on the protein being studied. Carboxyl groups on the crowder macromolecules will repel negative charges on the protein being studied.
model, the addition of crowders will increase the function of the protein by favoring the shifting the equilibrium from unfolded proteins to the more compact folded state.

A second example considers the association of a monomers of a protein that is only active as dimer. Under diluted conditions the chances of forming an active dimer will depend more on Brownian motion, stochastically forming dimers as the two monomer diffuse in vast volume available in dilute solution. Crowding the system, however, would enhance the tendency to form dimers as the volume available to the protein monomers is reduced. This increases the effective concentration of the protein, which through Brownian motion will interact with another protein monomer more often in the decreased space available. This in turn pushes the equilibrium towards the favored more compact state (dimer) and results increase the activity of the enzyme increasing.

Experimental evidence of these theoretical examples was pioneered by the Minton group at NIH. The enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPD) exists in equilibrium of monomers and tetramers, of which the monomers are 30-fold more active than the tetramers. The crowding hypothesis predicts that increasing concentrations of crowders will promote association of GAPD into tetramers by decreasing the total volume available. The formation of tetramers of GAPD under crowded conditions should reduce the specific activity relative to the same concentration of the enzyme in dilute conditions. To test this idea, increasing concentration of ribonuclease A, β-lactoglobulin and bovine serum albumin were added to fixed concentrations of GADP. The addition of crowders did not affect the enzyme activity of the monomer or tetramer but facilitated the formation of the tetramer by increasing the effective concentration of GADP through reduction of the available volume in solution.

In a second example, the effects of adding the synthetic crowder, dextran, to a solution of cytochrome c at pH 2, which is primary unfolded were investigated. The addition of the synthetic crowder resulted in the increase of protein stability and the formation of a more compact, molten globule-like structure relative to the unfolded form in diluted solution. These results suggested that crowding promotes protein stability. As crowders reduce the volume available to unfolded cytochrome c, the protein must adopt more compact conformations, which will favor the formation of the molten globule-like structure. Subsequent efforts by other groups helped provide further experimental evidence of the effects of crowding on protein stability.

**Weak transient-interactions**

The crowding hypothesis requires the assumption that crowders are considered as “inert” macromolecules so that only the effects of volume excluded are being studied. However, this model lacks the possibility of realistic chemical interactions; that the functional groups on the crowders will form hydrophobic, van der Waals, hydrogen bond, polar, or electrostatic interactions with protein being studied. Understanding the types of interactions that are formed by macromolecules under crowding conditions is important for understanding how proteins behave inside the cell. In order to address this problem biochemical experiments should be studied under crowded conditions. Adding macromolecules at concentrations similar to those in the cell to the solution and determining how proteins behave will help bridge this gap in knowledge. The synthetic crowder most typically used is the water-soluble polymer polyethylene glycol (PEG,
Figure 1.3), which are commercially available in a range of sizes from 200 Da to >20,000 Da. High concentrations of PEG tend to have repulsive interactions with proteins and avoid associating with the surface protein surface. This favors macromolecular association and compaction in agreement with the crowding hypothesis. However, it became evident from several studies that volume excluded effects were not the only interactions present when PEG was used as a crowder. Several groups found attractive interactions between PEGs and the hydrophobic side chains on the surface of proteins. 

Attractive interactions between PEGs and hydrophobic residues should decrease the stability of a protein by stabilizing the denatured state as most of the core of the folded protein consists of hydrophobic residues. PEGs will stabilize the unfolded state by interacting with exposed hydrophobic residues as the protein begins to partially unfold. These interactions between PEGs and the partially unfolded and fully unfolded states stabilizes them, shifting the equilibrium towards the unfolded protein. Therefore, increased stability of the protein of interest in the presence of PEGs indicates that the entropic volume excluded effects contribute more to the overall stability of the protein than does the attractive interaction between the protein and the crowder. These enthalpic interactions will become destabilizing for proteins of interest that have a larger quantity of hydrophobic groups, particularly those that are buried in the core of the protein, relative to the overall size of the protein. Proteins that are made up of a larger fraction of hydrophobic residues will form more interactions with PEGs, thus favoring the denatured state. More hydrophilic proteins, on the other hand, will have fewer hydrophobic residues that can interact with PEGs, and thus the denatured state will not be destabilized relative to the native state. As a result of these attractive interactions between protein and crowder, the use of PEGs was discouraged and instead the use of other water-soluble polymers (such as dextran and Ficoll) and proteins (hemoglobin and BSA) were encouraged as they “lack” such attractive interactions with other proteins (Figure 1.3). However, again contrary to the crowding hypothesis, several papers suggested that some globular proteins are not stabilized by protein crowders. The destabilization was attributed to the formation of numerous weak, non-specific attractive interactions between the the protein being studied and the crowders and sometimes weak interactions can overcome the stabilizing effects of volume excluded effects. In summary, these results suggested that macromolecular crowding arises from two phenomena, volume excluded effect and weak transient interactions.

Methods to study weak transient interactions

Until recently, most efforts to understand macromolecular crowding have focused on the entropic effects, however weak interactions also play an important role in the cell. For example, weak transient interactions slow the rates on protein diffusion. There are two components of diffusion, rotational and translational, both of which can be altered by the viscosity of the solution. Increasing viscosity affected both the rotation and translational diffusion of chymotrypsin inhibition. However, different results were observed depending upon whether synthetic or protein crowders were used. Synthetic crowders increased translational diffusion of chymotrypsin inhibitor, while protein crowders only slowed translational diffusion slightly. Rotational diffusion dramatically decreased when protein crowders were used, but increased in synthetic crowders. The reason for the difference between protein crowders and synthetic crowders was attributed to the presence of weak, non-specific, transient interactions between the protein crowders and chymotrypsin
Figure 1.3. Structures of typical macromolecules used in crowder studies. Including examples of polyethyleneglycols, ficoll, bovine serum albumin (BSA), dextran, sugars and hemoglobin.
inhibitor, while no interactions with the synthetic crowders noted.\textsuperscript{27} The decrease protein rotational diffusion, particularly, broadens NMR signals, which make proteins hard to detect in viscous solutions, such as inside the cell.

Another illustrative example of how weak interactions affect the rotational diffusion of proteins inside the cells was described for the three model proteins, the B1 domain of protein G (GB1, 6.2 kDa), the N-terminal metal-binding domain of mercuric reductase (NmerA, 6.9 kDa) and ubiquitin (8.7 kDa).\textsuperscript{28} These globular proteins have similar sizes (6-9 kDa) and structures, but different surface properties (i.e., pI, volume, surface area). The ability to detected in-cell signals of these three proteins in \textit{E. coli} was tested using two-dimensional NMR spectroscopy. In \textit{E. coli}, the majority of the most abundant proteins have acidic pIs, and their surfaces are negatively charged. GB1 has a pI of 4.8 and a net charge of -4 at pH 7.\textsuperscript{28} Consequently, steric repulsion should be favored and diffusion would not be affected. This is in agreement with the detection of clear in-cell NMR signals for GB1. However, no signal was detected for NmerA or ubiquitin, which have a similar net charge near zero. An equal number of positively and negatively charged residues on the surface of these near-neutral proteins indicates that they may form electrostatic interactions with negatively charged macromolecules in the cell. Additionally, they possess nearly the same number of hydrophobic residues on their surfaces, but the distribution of hydrophobic residues differs. In NmerA most of the hydrophobic residues are dispersed across the surface of the protein. The hydrophobic residues in ubiquitin are mainly located in one hydrophobic patch at the C-terminus. This difference in surface distribution of hydrophobic groups results in different propensities of the two proteins to form weak interactions which ultimately hinders detection of the proteins by in-cell NMR. These results not only highlighted the importance of repulsive interactions but also showed how hydrophobic interactions could play an important role in the rotational diffusion of proteins.

Similarly, positive charges in the surface of proteins could slow the diffusion of proteins inside the cell. This was tested by detecting in-cell NMR signals of GB1 and the positively charged protein, cytochrome c (11.5 kDa).\textsuperscript{29} As predicted GB1 had a well-resolved in-cell NMR spectrum; however, cytochrome c did not. Cytochrome c is a basic protein with a net charge of +7, which suggested the possibility of interactions with the negatively charged cytosolic proteins in the interior of \textit{E. coli}. Surprisingly, when the cells were lysed and subjected to size-exclusion chromatography, cytochrome c eluted with an apparent molecular weight 10 times higher than its mass in dilute solutions. Addition of salt (200 mM NaCl or K-glutamate) or the use of charge-inverted mutants (replacing 1-3 lysines/arginines with glutamate) of cytochrome c disrupted the formation of these larger complexes.\textsuperscript{29} This indicates that cytochrome c forms weak, electrostatic interactions with other proteins, or the lipid membrane, in cells. The decreased diffusion due to these interactions leads to the inability to detect cytochrome c by in-cell NMR.

How many of these weak interactions are needed to drive interactions between proteins and macromolecules in the \textit{E. coli} cytoplasm? For the small and biologically “inert” protein GB1 (i.e., it does not significantly interact with other proteins in the cytosol), only the addition of a few residues was needed to yield poor quality in-cell NMR spectra.\textsuperscript{30} As mentioned above, GB1 gives a well resolve in-cell NMR spectrum, therefore it was used as a model system to test whether
increasing the net charge of a protein would diminish the in-cell NMR signal due to increasing the interactions of the protein with macromolecules in the cytosol. Small motifs of arginine (from one to five residues) were added to its C-terminus. The in-cell NMR spectra became undetectable when four or five arginines were fused to the C-terminus.\textsuperscript{30} The lack on an in-cell NMR spectrum with the additional positively charged residues suggests that the C-terminal arginines in GB1 electrostatically interact with other macromolecules in the cytosol, which decreased the mobility of the protein and dampened the NMR signals. After cell lysis all the arginine-containing constructs became detectable suggesting that these interactions between GB1 and the other macromolecules no longer occurred upon cell lysis and thus are weak. In summary, these studies also highlight the impact of charge-charge interactions inside the cell, and how they can affect detection of proteins by in-cell NMR.

Weak transient interactions also play an important role in the stability of proteins inside the cell. The presence of weak transient interactions inside the cells could also overcome the stabilizing effects of volume excluded effects.\textsuperscript{31} Crowding effects were studied by in-cell NMR of the small globular protein L (ProtL) and a variant (K\textsubscript{x7E}) in which seven lysine residues have been replaced with glutamates. The glutamates lower the stability of the mutant, more than 80\% of which is unfolded at room temperature. Addition of salts (>0.3 M) folded the variant which prompted the question whether the crowded interior of the cell would have the same effect. If volume exclusion was the only effect on protein folding, then proteins in the interior of the cell would be folded. However, the K\textsubscript{x7E} remains unfolded inside the cell, even when the cells were grown in hyperosmotic media (1 osmolal). These results suggest that weak interactions could overcome the volume excluded effects of the cell and folding of K\textsubscript{x7E}. This indicates that weak interactions occur in the cell and highlights the importance of studying weak interactions under cell and cell-like conditions. Similar non-stabilizing effects have been reported for other proteins when placed in cells.\textsuperscript{32} These results suggest that weak interactions may be a more general effect that is not protein specific.

**Effects of crowding on enzyme activity**

In addition to the importance that crowding and weak interaction plays in protein stability, a lot of interest has been paid to its role on enzymatic reactions. In general, similar to when studying the effects of crowding on protein stability, the effects on enzyme activity are studied by adding synthetic or protein crowders of different sizes and chemical compositions. Enzyme activity varies quite a bit with the enzyme being tested with activity increasing,\textsuperscript{33-38} decreasing,\textsuperscript{3, 35, 39-42} or unaffected,\textsuperscript{43, 44} by the presence of crowders, depending upon the enzyme and crowders being used. These changes in enzyme activity have been attributed to changes in enzyme oligomeric state,\textsuperscript{3, 35, 45-47} conformation changes,\textsuperscript{37, 48, 49} protein diffusion,\textsuperscript{39, 50, 51} and water activity.\textsuperscript{35, 52} Such varied interpretations of the role of crowders on enzyme activity indicates that a simple volume excluded effect may not explain the data adequately.

Few studies have focused on the effects of crowding on enzyme activity from the point of view of small-molecules substrates. Macromolecular crowding is predicted to shift the binding equilibrium of the substrate to the smallest volume state, typically the bound state.\textsuperscript{20, 53, 54} For example, small-molecules substrates are made of recurrent functional groups that have
hydrophobic properties or charged atoms that could interact with the protein pockets.\textsuperscript{55} Some authors have reported additional effects to those that can be interpreted through the volume excluded hypothesis. Substrate diffusion, reversible reactions, substrates nature, and enzyme mechanism also play important roles when considering the effects of crowders on enzyme activity. For example, for the reversible reaction of malate dehydrogenase, crowders affected the $K_m$ of the oxidation reaction more drastically than the forward reaction.\textsuperscript{56} The authors attributed these results to chemical properties of the crowders and not to volume excluded effect. Further analysis also suggested that the maximum effects on $V_{\text{max}}$ were lower when neutral crowders were used instead of charged crowders. The different effects of synthetic versus protein crowders on malate dehydrogenase highlights the need to use crowders that better mimic the interior of the cell to gain true insight into how proteins behave \textit{in vivo}.

Another example of how weak interactions affect enzyme activity was reported for two crowders, PEG (8 kDa) and dextran (10 kDa), modulated the reaction of the horseradish peroxidase (HRP) with two substrates, 3,3′,5,5′-tetramethylbenzidine (TMB) and o-phenylenediamine (OPD).\textsuperscript{57} The two substrates differ in hydrophobicity as TMB is more hydrophobic than OPD. Measurements of the maximal velocity ($V_{\text{max}}$) decreased more drastically for TMB in comparison to OPD when either crowding agent was added. Additionally, PEG decreased TMB oxidation 2-times more than dextran did. This result is most likely due to the more hydrophobic nature of PEG compared to the hydrophilic dextran. These results suggested that substrate may interact with the crowders, particularly with PEG. As mentioned earlier PEGs are known to form hydrophobic interactions with the surface sidechain of proteins\textsuperscript{4} and hydrophobic molecules.\textsuperscript{58,59} To focus only on the effects of weak transient interactions without the volume excluded effects, the monomeric counterparts of the synthetic crowders were used, PEG 400 and glucose. The monomeric units of the crowders have the same chemical properties, and form similar weak interactions with ligands and proteins, but take up less volume per molecule and thus the volume effects are not as significant. The $V_{\text{max}}$ and $K_m$ did not significantly change for OPD in either cosolute. However, the $K_m$ for TMB was 24-fold higher in PEG400. This result was in agreement with the results obtained with the crowder PEG 8 kDa. Furthermore, measurements of the substrate diffusion by diffusion order NMR spectroscopy in buffer and crowders exhibited that the diffusion for TMB decreased greatly compared to OPD. These results highlight how weak interactions between crowders and substrates affect enzyme activity and substrate diffusion.

In addition to synthetic crowders, protein crowders will also bind to small-molecules substrates.\textsuperscript{60} Crowding effects on the activity and ligand binding to two different dihydrofolate reductase (DHFR) enzymes, R67 plasmid DHFR and \textit{E. coli} chromosomal DHFR (EcDHFR), which have different protein scaffolds and active site structures but that catalyze the same reaction, were studied to differentiate the effects of crowders on the ligands from the interactions of the crowders with the DHFRs.\textsuperscript{60} Binding of the cofactor, NADPH, to the two DHFRs exhibited a mixture of effects with some crowders increasing the binding affinity of the ligand, while others weakening it. However, binding of DHF to both DHFRs decreased for all crowders suggesting the presence of weak interactions between the crowders and substrate. Several correlations between protein properties and both substrates were observed. Binding of either NADPH or the cofactor product, NADP$^+$, correlated with the net charge of protein crowder, which concurs with computer
simulations of a simplified bacterial cytosol that indicated that small, negatively charged molecules diffuse slower in the crowded cytosol due to forming weak charge-charge interactions with macromolecules. The effects of crowders on DHF binding correlated with the size of binding pockets on the crowder proteins. These results were not surprising as others have reported interactions between folate and the central cavity of hemoglobin and to serum albumin. Additionally, protein binding sites tend to be promiscuous, able to weakly bind ligands with an array of structures. The recurrent presence of these weak interactions between different small-molecules substrates and proteins suggest that the in vivo behavior of enzymes may be much different than expected from in vitro.

As many of the atoms present in the surface of proteins are also present on many small molecules called osmolytes in the cell (Figure 1.4), insight from the interactions with between osmolytes and DHF can be used to glean information about crowders interactions with ligands. In vitro studies on R67 DHFR catalysis in the presence of several osmolytes (such as glycerol, ethylene glycol, sucrose and betaine) resulted in tighter binding for the cofactor and weaker binding for DHF. However, while for NADPH the effects are independent of the osmolyte used, the effects on DHF varied with the osmolyte. After ruling out effects of viscosity and dielectric constant of the solvent and volume excluded effects on the conformation of the protein the next plausible explanation was that the osmolytes interacted with free DHF. Our group has tested this idea with two additional, structurally unrelated DHFRs, EcDHFR and FolM. Again, similar results to R67 DHFR were observed with tighter binding for NADP⁺ and weaker binding for DHF. It is unlikely the osmolytes were having an effect on the enzymes as all three are structurally different enzymes. This suggests that the osmolytes are having an effect on DHF binding by interacting with the free ligand, and not with the DHFRs. Additionally, the effect of the osmolytes will depend upon which part of the DHF they are interacting. Binding of the pterin ring analog, pterine pyrophosphate, to dihydropteroate synthase was unaffected by the presence of trehalose, but became tighter in vitro.

To address whether weak interactions seen in vitro correlate with in vivo effects, osmotic stress studies were performed. Bacteria cells, like many organisms, adapt to changes in their surroundings. During times of salt, heat, dehydration, and hydrostatic pressure stresses, bacteria release water to the media and accumulate or produce osmolytes. WT and some more efficient mutant clones of R67 DHFR allowed E. coli in the presence of increasing concentration of sorbitol, up to 1.9 osmolar, in media containing the anti-folate, trimethoprim, which selectively targets chromosomal DHFR, but does not inhibit R67 DHFR. Trimethoprim makes the cells dependent upon R67 DHFR to survive. The Y69L mutant with poor catalytic efficiency nearly 250-fold lower than the WT enzyme, however, could not rescue the cells at sorbitol concentrations above 0.8 osmolal. These results were interpreted osmolytes produced upon osmotic stress in the cell interacted with DHF, which weakened binding to the Y69L R67 DHFR mutant, which could not produce enough THF for the cells to grow at high osmolalities. These results were corroborated by another study where R67 DHFR was expressed at the minimum levels needed for cells to survive, and cells were grown under osmotic stress. The cells with limited R67 DHFR expression stopped growing at lower osmolalities than cells supplemented with the metabolic end products of the folate cycle. Likewise, when other folate cycle enzymes, serine hydroxymethyl transferase and methylene tetrahydrofolate reductase, were also expressed at low levels in knockout cell lines,
Figure 1.4. Osmolytes and crowders have similar functional groups that may form weak interactions with other molecules.
the cells were not able to grow at high osmotic stress levels compared to the strains supplemented with the end product metabolites of the enzyme reaction. These data indicate that osmolytes in the cell can interact with folates and may stop cell growth by decreasing the efficiency of the enzymes being expressed at low levels.

These results led to the proposal of a model of preferential interactions (Figure 1.5). In this model osmolytes interact with the free DHF ligand and require more energy to remove than water. The additional energy required to remove the osmolytes then shifts the equilibrium to the unbound state. This model does not exclude potential effects of the osmolytes on the proteins as well. For both FolM and dihydropteroate synthase some osmolytes caused weaker binding of ligands through interacting with the proteins as well, destabilizing the protein structure. Since osmolytes have similar functional groups to crowder macromolecules, this model can also explain how crowders may weaken ligand binding as well. The ligands may form weak interactions with the surfaces of the crowders, either through charge-charge interactions or hydrophobic interactions, and need to dissociate from the surface of the crowder prior to binding to their protein target. Overall, our results support the idea of recurrence of weak transient interactions at the interior of the cell.

**Weak interactions between osmolytes and folate**

To further study the interactions between DHF and the osmolytes our group utilized osmometry and NMR approaches. First, interactions of osmolytes with folate were elucidated through the effects of the osmolytes on folate dimerization. Folate is a more stable analogue of DHF and is composed of three moieties, a pteridine ring, a p-amino-benzoyl (pABA) and glutamate tail. Folate dimerizes with a dissociation constant of 80 mM at pH 7. Addition of osmolytes weakened dimerization 2 and 3-fold for betaine and DMSO, respectively, suggesting interaction of the osmolytes with folate. Similar to how osmolytes affect ligand binding to proteins, osmolytes associating with the monomeric form of folate stabilize the formation of the monomer over the dimer. To further understand which part of the folate structure the osmolytes interact with, nuclear Overhauser effect (NOE) was used. NOE is a phenomenon that describes how the local magnetic field of one nucleus is affected by the presence of another nucleus. This effect is sampled in the NMR spectra with changes in intensity from one peak, resulting in positive or negative NOEs. Furthermore, since the NOE is distance dependence, it can be used to measure interatomic distances between two protons. From the spectra NOEs between the C7 with C9 and C3’/C5’ were noted, and also between the C9 proton and the C3’/C5’ protons (Figure 1.6). These results suggested that folate is an extended molecule in solution. However, the NOE switched signs from positive to negative upon addition of osmolytes. Negative NOE suggested a slower rotational rate for the bonds between the pterin and pABA rings. These slower rotations of the bonds between the rings could be rationalized by an interaction between osmolytes and these rings. The folate-osmolyte complex has a higher effective mass than just the solvated folate molecule, thus spins slower because of this higher mass. Osmolytes were found not to interact with the glutamate tail. As the pterin ring has only one non-exchangeable proton, it was difficult to extract information of an interaction between osmolytes with this moiety of folate. Therefore, a different approach was used.
Figure 1.5. Model for how preferential interaction of osmolytes and crowders with the free ligands, like DHF, affects their binding affinity. In dilute solutions, without osmolytes, water needs to be removed from the protein binding site and the ligand solvation shell prior to binding. When osmolytes are added to the solution, and the ligand is preferentially solvated by the osmolytes compared to water, then the osmolytes also need to be removed from the solvation shell of the ligand prior to binding (lower panel). More energy is required to remove the osmolytes from the solvation shell of the ligand compared to water alone, which manifests as a decrease in the binding affinity of the ligand in the presence of the osmolytes.
Figure 1.6. Structures of the substrate and cofactor of dihydrofolate reductases. Left, the structure of folate labeled with the atom numbers and the pterin ring and pABA-glu tail indicated in the structure. Right, NADPH structure with the nicotinamide ring pointed out.
Vapor-pressure osmometry (VPO) measures the change in osmolality when a small molecule is added to a three-component system containing water (1), test compound (2), and an osmolyte (3). The chemical potential for the test compound in the presence of the osmolyte, $\mu_{23}$, measures the test compound-osmolyte interaction and is calculated when the product of test compound and an osmolyte molality as a function of changes in osmolality yields a linear plot. The slope is used to calculate the $\frac{\mu_{23}}{RT}$, which is defined as the preferential interaction coefficient. A negative value suggests that the test compound interact with the osmolyte, while a positive value suggests interaction with water and zero value indicates equal preference with water or osmolyte. Betaine interacted with folate with a slightly negative $\frac{\mu_{23}}{RT}$ of $-0.03 \text{ m}^{-1}$ at dilute folate concentrations. The experimental $\frac{\mu_{23}}{RT}$ values suggest a preferential interaction between betaine and pteridine and pABA rings, while the glutamate tail prefers to interact with water. Betaine was found to prefer to interact with the aromatic carbons that make up the pterin and pABA rings, while it is highly excluded from the carboxylate groups of the glutamate tail. These results were in agreement with simulation data, which indicated that the interaction between betaine and folate was dependent upon the latter’s conformation in solution. The osmolyte interacted with the extend conformation more than a bent, L-shaped conformation which had less exposed aromatic surface area than the extend conformation. In summary, weak interactions between osmolytes and DHF are likely to play an important role in vivo as the cell is crowded with proteins and small molecules decorated with groups that could compete with interactions with DHF and have an effect in the enzyme activity of DHFRs. Furthermore, while we have studied the effects of homogeneous proteins and osmolytes on DHFRs, we have not done with a heterogeneous protein mixture as those seen in the interior of the cell. As these in vivo experiments are complicated, we propose to use a bottom up approach using as a model enzyme, R67 DHFR, and fluorine NMR to test how a more complex crowded milieu will affect ligand binding.

**Dihydrofolate reductases**

Dihydrofolate reductases are enzymes that catalyze the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using NADPH as a cofactor. DHFR is the gateway enzyme to the folate cycle, where THF, an important cellular metabolite, is used as a one-carbon donor in the synthesis of purines, pyrimidines, methionine, glycine and other essential metabolic intermediates. Therefore, DHFRs is an important target for several anticancer and antibacterial drugs such as methotrexate and trimethoprim (TMP). However, in some cases excessive use and misuse of antibiotics, such as TMP, have resulted in the development of drug resistance with the production of new forms of DHFR. One example is the plasmid encoded R67 DHFR which shares no sequence or structural homology with the chromosomal DHFR despite of catalyzing the same reaction.

**Structure of DHFRs**

Chromosomal DHFR and R67 DHFR have strikingly different structural features (Figure 1.7). For example, chromosomal DHFR is an 18 kDa monomeric protein with 8 beta-strands core and four alpha-helices conforming a two subdomain structures separated by a hinge region. Chromosomal DHFR possess specific binding pockets for the cofactor and the substrate is oriented for relatively easy transfer of the hydride from NADPH to DHF. Crystallographic studies suggest an important
Figure 1.7. Structures of dihydrofolate reductases. A) *E. coli* chromosomal DHFR (PDB ID: 1RA2\textsuperscript{80}) with bound NADP\(^+\) (green space fill) and DHF (magenta space fill). B) R67 DHFR (PDB ID: 2RK1\textsuperscript{81}) with the monomers individually colored with NADP\(^+\) (green) and DHF (magenta) bound in the central active site pore.
role for a stretch of residues (known as the M20 loop) in minimizing access of solvent into the active site. This loop adopts different positions depending upon which ligand is bound. In contrast to the chromosomal enzyme, R67 DHFR is a 34 kDa homotetramer with all four subunits forming a single active site. Each monomer is 78-residues long where the first 16-residues are disordered and the remaining residues form a five anti-parallel beta strand resembling an SH3 domain. Two monomers associate to form a dimer with three beta strands of each monomer intercalating to form a six beta strand dimer. Subsequently, dimers associate to form a tetramer which is stabilized by symmetry-related tryptophans and histidines residues at the dimer interface. Due to the 222-symmetry of the molecule the active site is of R67 DHFR is unusual as is exposed to the solvent and transverses the length of the protein. This unusual geometry of the active of R67 DHFR is quite different from chromosomal DHFR which led to several questions about ligand recognition and specificity as well as the binding modes of the cofactor and substrate.

Ligand binding studies with R67 DHFR

One of most striking features of R67 DHFR is it unusual pore. With a length of 25 Å, the pore transverses the entire length of the molecule from top to bottom. The pore has an hourglass shape. The pteridin ring binds to the center of the pore. In the crystal structure of R67 DHFR with folate bound, while density for the pABA-glutamate tail was not observed. Docking of the nicotinamide ring of the cofactor to the R67 DHFR-folate complex was used to model the ternary complex from which the nicotinamide ring was proposed to bind near the center of the pore to reduce DHF. This model suggested that R67 DHFR uses symmetry-related residues to accommodate, in each half of the pore, NADPH or DHF in different orientations.

Subsequent fluorescence anisotropy and ITC studies indicated that a maximum of two ligands bind in the pore of R67 DHFR. Either two molecules of NADPH, two molecules of DHF or a combination of DHF and NADPH and only the last complex results in hydride transfer. Furthermore, it was clear that when various pairings of ligands bound they interacted with each other in a way that the first molecule to bind affected the second molecules affinity. For example, binding of one molecule of NADPH decreases the affinity for a second NADPH molecule, while the first DHF to bind increases the affinity for a second DHF. These binding effects are referred as negative and positive cooperativity and help to suggest a sequential productive mechanism of catalysis where one molecule of NADPH bind first to the free enzyme, follow by DHF binding to facilitate hydride transfer.

Successful crystallization of the R67 DHFR-NADP⁺-DHF ternary complex aided to understand the non-specificity of the active site for substrate and cofactor. As initially reported for the binary complex, no electron density was observed for the pABA-glutamate tail. The ternary complex is characterized as being a part of a multiple-layered stacking arrangement that includes four W38s residues and a hydrogen network formed by Q67 and Y69 residues (Figure 1.8). The pteridine and nicotinamide rings are sandwiched between these layers. Such arrangement has been dubbed as the club sandwich motif and helps explain why in the absence of binding site in R67 DHFR that are specific to DHF or NADPH the enzyme is able to facilitate hydride transfer. Such prowess is achieved by the ability of active site residues to play dual roles. For example, the backbone of I68 provide recognition for the caboxamide on the nicotinamide ring and the N3-O4 amide group of
Figure 1.8. Stacked arrangement of NADP\(^+\) and DHF, Y69, Q67, and W38 in the binding site of R67 DHFR. Monomers of R67 DHFR (PDB ID: 2RK1\(^8\)) are individually colored, DHF (yellow), NADP\(^+\) (magenta), and the residues that make up the club sandwich motif (light grey) are shown.
the pteridine ring. As a result of these and other interactions, the two rings are brought into close proximity to facilitate hydride transfer.

To further understand the role of pABA-glu tail in catalysis several approaches have been used. Crystallography and NMR studies suggested that the glutamate tail is disordered when bound.\(^{82, 83, 85}\) Modeling of the glutamate tail suggested that the alpha and gamma carboxyl groups in the glutamate could interact with symmetry-related K32 residues at each side of the mouth of the pore. However, because the mouth of the pore is too wide for the two carboxylates to interact with the two lysines at the same time a poor fit was observed. This poor fit indicates that the lysines could potentially switch between interacting with the two lysines, which is consistent with the glutamate tail being disordered.\(^{86, 87}\) Furthermore, ITC binding studies with the DHF analogue, dihydropteroate (DHP), which lacks the glutamate tail exhibited weaker binding affinity and resulted in 1600-fold decreased in \(k_{\text{cat}}\). These results suggested that the tail plays an important role for catalysis. Further computational studies of the hydride transfer reaction revealed that glutamate tail sample\(^{88}\) two conformations and are moving back and forth forming ionic interactions with K32 at the mouth of the pore.\(^{88}\) This movement appears to be linked to the puckering of the pteridine ring to facilitate the hydride transfer. To test the hypothesis that movement of the glutamate tail improves the catalytic rate, several experimental strategies were used to constrain the glutamate tail. For example, reduction of the active pore volume by 35% by introducing a Y69W mutation did not alter the \(k_{\text{cat}}\). Tethering of the glutamate tail via chemical crosslinking to the nearby K32 residue did not prevent reduction, but slow the reaction rate. Finally, chemically modifying the glutamate carboxylates with positively charged amines weaken binding by ~9-fold. Overall, these results suggest that R67 DHFR evolved to allow the disorder of the pABA-glu tail to facilitate enzyme function.\(^{89}\)

**Enzyme mechanism of R67 DHFR**

The crystal structure of ternary complex also helped to provide a better understanding of R67 DHFR catalytic mechanism.\(^{81}\) The reduction of dihydrofolate to tetrahydrofolate takes place in two steps, protonation first, then hydride transfer.\(^{80}\) Many enzymes that utilize folates contain an aspartate or glutamate group that acts as a “pterin hook” to increase the affinity of the folate.\(^{91}\) For bacterial and human chromosomal DHFRs, protonation of the N5 nitrogen of the pteridine ring is facilitated by the presence of this glutamate or aspartate residue in the active site (Figure 1.9). This residue is not in direct contact with N5, instead is positioned directly with N3 and the C2-exo amino group (N2 in Figure 1.9). This close proximity has been suggested to be involved in elevating the pKa of N5 from 2.6 to 6.5 and promoting a tautomeric equilibrium which facilitates direct protonation from water follow up by the hydride transfer by the nicotinamide ring of NADPH. Mutation of this carboxylate, like Asp27 in *E. coli* chromosomal DHFR, decreases the hydride transfer rate by 100-fold.\(^{92}\) In contrast to chromosomal DHFR, protonation and hydride transfer are suggested to be concerted in R67 DHFR.\(^{81}\) Several structural differences support this hypothesis. First, it is unlikely that the pK\(_a\) of N5 would be affected in the bound DHF substrate as there are not acidic residues in its active that would stabilize the protonated form. Second, invocation of tautomer equilibrium between the enol and keto conformations of the N3-O4 moiety to facilitate protonation of N5 seems not to be supported as the formation of hydrogen bonds...
Figure 1.9. The glutamate or aspartate residue in enzymes that bind folate that interacts with N3 and N2 of the pterin ring. In *E. coli* chromosomal DHFR, Asp27 interacts with these residues and increases the pKa of N5.82,90 There is no similar residue in the active site of R67 DHFR.
between the backbone of I68 with N3-O4 of the pteridine ring should restrict formation of the enol form. Furthermore, ITC supports the notion that in the enol form of pteridine ring, the N3-O4 is deprotonated with no possibility to H-bond with the amide carbonyl of I68. Rather, a simpler catalytic model has been proposed which invokes water as the proton donor to N5 and exploit a subtle symmetry between the nicotinamide amide group and the N3-O4 amide of the pteridine to hydrogen bond with the backbones of I68 residues. As a result, substrate and cofactor are in close proximity. In this model, protonation and hydride transfer may be concerted. Finally, measurement of kinetic isotope effects suggests that for R67 DHFR, the rate limiting step is the hydride transfer as oppose to product release of chromosomal DHFR.

Role of the disordered N-terminus of R67 DHFR

The first 16-20 residues of the N-terminus of R67 DHFR are disordered. The first crystal structure of dimeric R67 DHFR exhibited poor electron density for the first 16 residues implying disorder of the N-terminus. Removal of these disordered residues and subsequent crystallization revealed still a high mobility for residues 17-21 judged by its diffuse density and high thermal factors. Furthermore, when the sequence of R67 DHFR was analyze by several disordered predictors, all algorithms converged to predict that the residues of the N-terminus are disordered. Finally, small angle neutron scattering studies indicated that the disordered tail sampled large regions of space near the monomer-monomer interface of each dimer while they still remained disordered. This suggested that the N-termini interacted with themselves as well as with residues at the monomer interface. Collectively, these experiments suggested that the N-termini of R67 DHFR are intrinsically disordered.

Intrinsically disordered proteins or regions play multiple role in molecular recognition, protein modification and much more. However, it is unclear what exact role the R67 DHFR terminus play on its function. Several experiments suggest that the disordered N-terminus of R67 DHFR does not play a role in enzyme catalysis. For example, the truncation of the first 16 residues by α-chymotrypsin treatment revealed that the protein remains fully active. However, attempts to genetically express a truncated version of the protein did not produce protein in vivo, suggesting a possible role of the first 16 residues in stabilizing the protein. GuanidineHCl stability studies indicated that full length R67 DHFR is 2.6 kcal/mol more stable than truncated protein. The increased stability of R67 DHFR with the N-termini intact may be due to their collapsed interactions between themselves and the monomer-monomer interface. This suggests that while the N-terminus seems unnecessary for catalytic activity, it is required for in vivo expression. Subsequent sequence analysis of five R67 DHFR homologues revealed that the N-terminus is the most variable region of the protein and experimental data suggested that incorporation of histidines tags into the N-termini did not affect the protein function. Indeed, when four gene copies of R67 DHFR are fused in tandem from C-terminus of one gene into the N-terminus of the next, the gene product resulted in four times the initial molecular mass and exhibited no changes in protein function. These results indicated that the N-termini can be manipulated without affecting enzyme function.

The symmetric binding sites that can bind either cofactor or DHF make R67 DHFR an excellent protein to study ligand binding under osmotic stress or macromolecular crowding. As both ligands
bind to symmetry-related residues in the active site, any effects of osmolytes or macromolecules that differ between cofactor and substrate binding can be attributed to effects on the ligands themselves. This is illustrated by the effect of the osmolyte betaine on NADPH versus DHF binding to R67 DHFR. NADPH binding became tighter, while DHF binding was weakened. Any effects on the protein would alter both NADPH and DHF binding in the same way, because DHF and NADPH bind to the same residues, but on opposite sides of the active site pore. Therefore, we wondered if in vitro experiments correlate with what will happen in vivo? To address this question, we turned to fluorine NMR as the optimal tool for studying ligand binding to R67 DHFR in crowded solutions.

**Nuclear magnetic resonance**

Nuclear magnetic resonance is a spectroscopy technique in which a nucleus under a magnetic field, and subject to a radio frequency (rf) pulse, absorbs energy to produce a transition between two states. Only certain atoms possess the quantum nuclear spin property to absorb rf and give rise to an NMR signal. Among the most common nuclear spins are $^1\text{H}$, $^2\text{H}$, $^{13}\text{C}$, $^{11}\text{B}$, $^{15}\text{N}$, $^{17}\text{O}$, $^{19}\text{F}$, $^{31}\text{P}$ among others. These nuclei have the property of “spin”, and when a charge particle spins, it produces a magnetic field with positive and negative poles. Thus, it is said that these nuclei behave like tiny magnet bars. When these magnets are subject to an external magnetic field they either align with the magnetic field (low energy state) or against it (high energy state). The difference in energy to change the spin from the low energy state to the high energy state. When there is no magnetic field the difference between energies is zero, but as you increase the magnetic field the difference between states increases. Atoms follow the Boltzmann distribution and therefore the majority will align with the magnetic field and a small excess will align against it. However, if spins are subjected to rf energy exactly equal to the difference in energy between spins, this energy will be absorbed by some of the nuclei and the orientation of the spin will flip. This phenomenon of energy absorption is known as resonance.

A nuclear spin within a molecule with multiple nuclear spins will not sample or “feel” the same magnetic field as each of the other nuclei in the molecule as they are shielded by electrons and other nuclear spins systems. Therefore, each spin will have a slightly different chemical environment and absorb energy with different frequencies. These frequencies are called chemical shifts and are represented in the NMR spectrum as peaks. In the NMR spectrum, chemical shifts of shielded atoms, or with an increased electron density, appear upfield while deshielded atoms appear upfield. Since chemical shifts correlate to the chemical environment of the nucleus, several efforts have been made to relate chemical shifts with structural features. For the fluorine atom an empirical correlation between chemical shifts and their close intermolecular contacts with proteins was found. The study found that shielded fluorine atoms are observed primarily in close contact with hydrogen bond donors, suggesting the possibility to form hydrogen bonds. In contrast, deshielded fluorine atoms are mainly found in close contact with hydrophobic side chain and carbonyl groups of the backbone of proteins. In summary, chemical shifts give information about the chemical environment of nuclear spin probe.

Chemical shifts and peak intensities from NMR spectra can be used to quantify the binding affinity of a ligand. Two primary features of the NMR spectra are chemical shifts and peak
Important thermodynamic, kinetic and structural information can be obtained by a careful analysis of these features. One example is the quantification of dissociation constants by following changes in chemical shifts of a protein upon titration of ligand (or a protein). An important factor in the appearance of chemical shifts during a titration relates to the exchange rate of the ligand between the free and bound state \( k_{ex} \) and the chemical shift difference between these states \( \Delta \omega \). When the \( k_{ex} \) is greater than \( \Delta \omega \), it is considered as fast exchange on the chemical shift time scale. This is characterized by a smooth movement of a single chemical shift from the peak for the free state to the bound state peak upon titration, where the peaks have the same shape throughout the titration. When the \( k_{ex} \) is smaller than \( \Delta \omega \), it is considered as slow exchange. This is characterized with the gradual disappearance of the free state peak as a new peak signal appears for the bound state. The decrease and increase in intensities reflect the concentrations of the free and bound protein. An intermediate exchange process is considered when \( k_{ex} \) is on the same order as \( \Delta \omega \). A titration in the intermediate NMR time scale results in a combination of chemical shift transitions along with changes in peak intensities as a function of the ligand concentration. Dissociation constants from fast exchange processes are calculated by monitoring changes in chemical shift or peak intensity as function of the ligand concentration. However, because an intermediate exchange processes represents a more complex behavior with a combination of chemical shifts and peak intensities a different method of quantitative analysis is used called, NMR lineshape analysis. This analysis involves fitting the titration of one or two dimensional spectra to the equations that govern evolution of magnetization (or relaxation) in an exchange system. Overall, primary features of the NMR spectra such as chemical shift and peak intensities are used to extract thermodynamics parameters such as dissociation constants.

**NMR labeling strategies**

There are two strategies to incorporate NMR-active nuclei into proteins, through chemical modification and via biosynthetic incorporation of fluorinated amino acids. The most common method is the latter. In this method fluorinated amino acids are added to the media and incorporated biosynthetically by cellular machinery upon activation of protein expression of the target protein. To facilitate incorporation, auxotrophic cell lines of the amino acid of interest are used. When aromatic residues are to be labeled, inhibitors of their synthesis pathway are added, such as glyphosate. An updated method for the incorporation of fluorinated tryptophan reported that glyphosate is not needed and the precursor of the tryptophan amino acid, indole, can be used. Several advantages of biosynthetic incorporation include a wide variety of fluorinated aliphatic and aromatic amino acids, all which are commercially available. Furthermore, fluorinated amino acids can be monofluorinated or highly fluorinated. One drawback of biosynthetic labeling is that several of the fluorinated amino acids are relatively toxic to the cell which could hamper bacterial cell growth and consequently results in poor levels of incorporation.

Fluorine incorporation via chemical modification is achieved by conjugating a fluorinated moiety to a reactive functional group in the protein. The most common reactive functional group is the thiol cysteines, but other amino acids can also be labeled such as lysine, serine, threonine, histidine and tyrosine. In this method, the protein of interest is mixed with an excess of the fluorinated
reagent, the reaction is quenched by the addition of an excess of a competitive nucleophilic reagent, and removal of excess reagent and reaction byproducts is achieved by using size exclusion chromatography, dialysis or centrifugation.\textsuperscript{108} One of the advantage of this technique is the ability to label proteins in which biosynthetic labeling is cost-prohibited, such as protein expression in mammalian cells.\textsuperscript{112} Furthermore, the low abundance of free cysteines in a protein makes this approach amenable to a small number of labeling site or facilitates the opportunity to selectively engineer labeling site via site directed mutagenesis.\textsuperscript{109} One of the drawback of chemical modification is that removal of fluorinated reagents, particularly those with phenyl groups, requires the use of conditions that could result in denaturation of the protein target.\textsuperscript{108}

**Advantages of using fluorine NMR**

Fluorine NMR possess several advantages in monitoring ligand binding.\textsuperscript{108, 109} First, the fluorine atom is almost absent in all biological systems which eliminates background signals and simplifies the NMR spectra as we only look for one protein signal. Second, fluorine is the second most sensitive nuclear spin after hydrogen. This sensitivity of the fluorine probe facilitates detection of changes upon ligand binding, and can be used in other biological applications such as local and distal conformational changes, domain swapping and protein folding, among others. Third, the variety of fluorinated amino acids that can be incorporated into proteins is myriad, which increases the application. One advantage of using aromatic amino acids, specifically tryptophan, is its lower abundance relative to the other amino acids, which minimizes structural and functional protein perturbations. Finally, fluorine labeling has become simpler and less expensive. Recent reports showed that fluorine incorporation into tryptophan can easily be performed using the tryptophan precursor indole and standard \textit{E. coli} expression strains.\textsuperscript{111} The method does not require the addition of tryptophan or tryptophan synthesis inhibitors which greatly decreases the cost. Overall, fluorine NMR offer several advantages such as sensitivity, simplicity in the NMR spectra which favor its use over other NMR probes. These advantages make it also attractive for the use of in-cell NMR approaches.

**In-cell NMR**

The usage of NMR spectroscopy to study biological macromolecules inside intact cells is a referred as in-cell NMR.\textsuperscript{113} This technique offers the advantage to study macromolecules in the native, crowded, and heterogeneous environment of the cell.\textsuperscript{114} Similar to conventional NMR methods, proteins need to be labeled with NMR-active nuclei. The most common active nuclei for in-cell NMR include $^{13}$C, $^{15}$N and $^{19}$F.\textsuperscript{114} To collect in-cell NMR spectra, the protein of interest could either be recombinantly overexpressed or delivered into the cell via different approaches. One advantage of the former approach is that no purification is needed to perform the experiment. Protein delivery approaches are varied, including the attachment of cell penetrating peptides to the target protein, the use of pore forming toxins that permeabilize the cell membrane and facilitates protein translocation, direct microinjection of the protein into the cell (usually oocytes) and recently, via electroporation.\textsuperscript{114, 115} One disadvantage of protein delivery approaches is the direct manipulation of the protein and that the protein solution needs to be highly concentrated. In-cell NMR offer the advantage to study proteins in several cell types, such as \textit{E. coli}, oocytes, insect
and human cells. This technique has aided in understanding protein folding, determining structure \textit{in vivo}, and studying post-translational modifications in the cell. While in-cell NMR is a relatively new technique, it still has several challenges to overcome. For example, severe line broadening due to the presence of weak, transient interactions; the same type of weak interactions that occur between osmolytes or macromolecules and the ligands of DHFR. One possible way to overcome this is by changing protein surface residues. Another disadvantage is cell viability. $^{15}$N experiments require a long time (hours) to acquire data, which limits the survival of the cell. Fluorine NMR approaches have been developed to decrease the acquisition data. One dimensional NMR requires less time. Other disadvantage is the insensitivity of the technique, large amount of protein are needed in order to detect signals in the cell, which are often many times larger concentrations than the endogenous protein level. Ultimately, this is counter to the purpose of studying the protein in its native environment of the cell. Finally, the limited size of the proteins that can be studied is another important factor. Most of the in-cell NMR have been determined for relative small proteins (<20 kDa). This could be overcome by the use of trifluoromethylated labels, but the best labeling strategy needs to be optimized for each protein. In summary, in-cell NMR is a technique that offers the advantage to study proteins in their native environment, however, multiple challenges need to still to be overcome in order to broadening its usage.
References


Chapter 2. Effect of fluorotryptophans on the stability and function of R67 dihydrofolate reductase
This part is a version of a manuscript in preparation entitled: Effect of fluorotryptophans on the stability and function of R67 dihydrofolate reductase.

Gabriel J. Fuente-Gómez, Creighton L. Kellum, Alexis C. Miranda, Michael R. Duff, Jr.*, and Elizabeth E. Howell

Department of Biochemistry & Cellular and Molecular Biology Department, University of Tennessee-Knoxville, Knoxville, Tennessee 37996, United States

Corresponding Author: Michael R. Duff, Jr., mduff5@utk.edu

Funding: This work was supported by NIH grant GM 110669 (to EEH).
Abstract

Labeling proteins with fluorine atoms can be used to increase protein stability as well as in $^{19}$F NMR. It is important to study the effects of incorporating fluorine on protein stability and function to ensure optimal labeling. We have explored how fluorotryptophans impact the activity of the plasmid-encoded R67 dihydrofolate reductase (R67 DHFR). Circular dichroism indicated that the fluorotryptophans do not impact the secondary structure. The pH-dependent dissociation of R67 DHFR into dimers was stabilized by 5- and 6-fluorotryptophan, while the 4 and 7-fluorinated analogs decreased tetramer stability up to 2-fold. Similarly, the $T_m$ of R67 DHFR increase by 3 °C through incorporation of 5- and 6-fluorotryptophans and decreased 1-3 °C with the 4 and 7 analogs. Activity and ligand binding affinity decreased for 7-fluorotryptophan labeled R67 DHFR, but was unaffected by the other fluorine labels. Overall, our data indicate that labeling the tryptophans of R67 DHFR at the 5 and 6 positions with fluorine slightly increased tetramer stability and did not impact enzyme active, while the 4 position did decrease the stability of R67 DHFR without altering its function. Universally, 7-fluorotryptophan negatively impacted R67 DHFR, decreasing stability and activity.

Introduction

Fluorine has a slightly larger van der Waals radius to hydrogen, which suggests incorporation of fluorine-modified amino acids should have only small effects on the protein structure.\textsuperscript{1} One use of fluorine modification of proteins is to increase protein stability through protein engineering by increasing the hydrophobicity of the modified residues.\textsuperscript{2, 3} Additionally, $^{19}$F NMR has been used to study protein stability, dynamics, and ligand binding.\textsuperscript{1, 4} Fluorine is a particularly useful nucleus for NMR studies because of its absence from natural proteins; it must be incorporated biosynthetically.\textsuperscript{2} One such residue that can be labeled with fluorine is tryptophan.

One challenge of labeling proteins with fluorotryptophans is determining the optimal probe, including the position of the fluorine in the residue functional group. Several groups have studied the impact of the position of fluorine in the indole ring of tryptophan on protein stability and function.\textsuperscript{5-12} The effects of fluorine incorporation is protein specific with increased or decreased stability and function having been shown. Additionally, there can be a mixture of effects, depending upon the position of fluorine in the indole ring, where a fluorine at one position will alter activity, while fluorine at another position has no effect.\textsuperscript{7, 11, 12} A better understanding of how and why fluorine impacts the stability of proteins is needed.\textsuperscript{13} A larger data set with a range of proteins with different structures would help improve our ability to predict the effects of fluorine incorporation. Therefore, to progress towards understanding how the position of fluorine in the indole ring affects protein stability and function, we have explored the effects of 4-, 5-, 6-, and 7-fluorotryptophan on the R-plasmid form of dihydrofolate reductase (R67 DHFR).

R67 DHFR catalyzes the reduction of dihydrofolate (DHF) with NADPH as a cofactor, and is not sensitive to anti-folate inhibitors, such as trimethoprim or methotrexate, that target the chromosomal forms of DHFR.\textsuperscript{14-16} R67 DHFR is a homotetramer, that is a dimer of dimers, and possesses a single active site pore containing a single active site (Figure 2.1).\textsuperscript{17, 18} There are two tryptophan residues per R67 DHFR monomer, W45 at the monomer-monomer interface and W38, which is an integral part of the dimer-dimer interface.\textsuperscript{19} Mutation of W38 and other dimer interface
residues can destabilize the tetramer. Only the tetrameric form of R67 DHFR is active, the dimer is inactive. Therefore, determining which fluorotryptophan probes alter R67 DHFR stability is important in discerning the effect of the probes on the function of the enzyme.

**Materials and Methods**

**Expression and purification of 19F-labeled R67 DHFR**

Using the method of Crowley et al., fluorotryptophans were incorporated into R67 DHFR using the 4-, 5-, 6-, and 7-fluoroindoles. A his-tagged construct of R67 DHFR cloned into a pRSETb vector was expressed in pLysS (DE3) cells. Cells were grown to an O.D. of 0.6 at 600 nm in rich media, were harvested by centrifugation, and the pellet was resuspended in an equal volume of pre-warmed minimal medium (7.5 mM (NH4)2SO4, 50 mM Na2HPO4 and 50 mM KH2PO4 pH 7.0). Fluoroindoles were added to the media (final concentration of 60 mg/L), which was allowed to acclimatize for 30 minutes prior to protein induction by 1 mM IPTG. Protein was expressed for 10 hours and purified on a Ni(II)-NTA resin (Qiagen). Purified protein was dialyzed against water and lyophilized. The labeled R67 DHFR are referred to as 4F, 5F, 6F, and 7F R67 DHFR.

**Determination of fluorine incorporation levels**

The percent incorporation of fluorine was measured using 19F NMR. Samples were prepared in D2O containing 10 mM Tris-d11, pH 8.0 (uncorrected) with 5 mM trifluoroacetic acid. Protein concentrations in the sample (~300 μM) were measured using a bicinchoninic acid (BCA) assay from Pierce. Spectra were measured for 512 scans at 470 MHz on a Varian 500 MHz NMR with a OneNMR probe. A pulse angle of 87 degrees and delay of 1 s were used. Spectra were analyzed using MestReNova (Mestrelab).

**pH dependence of the tetramer to dimer equilibrium**

The tetramer of R67 DHFR is pH sensitive due to the protonation of H62 residues at the dimer-dimer interface. As W38 is also located at the dimer-dimer interface, it could impact tetramer stability and be utilized to monitor changes in the local environment. Fluorine incorporation effects on the pH-dependent equilibrium between tetramer and two dimers of 4F, 5F, 6F, and 7F R67 DHFR were assessed by fluorescence at room temperature as a function of pH on a Perkin-Elmer LS-5B spectrometer. Protein (2 μM) in MTA buffer (50 mM MES, 100 mM Tris, and 50 mM acetic acid), was excited at 295 nm and emission was monitored from 315 to 500 nm in a 1 cm cuvette. The samples were titrated with small aliquots of 2 M HCl over a decreasing pH range from 8.5 to 4.5. The intensity averaged emission wavelength, <λ>, was monitored and plotted versus pH and globally analyzed in SAS using Eq. 4 in Nichols et al.
Figure 2.1. Structure of R67 DHFR with the location of tryptophan residues. A) Tetrameric R67 DHFR where each monomer is individually colored (PDB ID: 2RK1)\(^8\). NADP\(^+\) (magenta) and DHF (light green) are shown bound in the central active site pore. B) A 90° rotated structure showing the W38 residues that make up part of the dimer-dimer interface (yellow sticks). C) Location of W45 (cyan stick structure) where two monomers of R67 DHFR interact to form the dimer. D) Numbering of the carbon atoms in the indole ring of tryptophan.
Circular dichroism

The effect of fluorine incorporation on the secondary structure of unlabeled and fluorinated R67 DHFR was measured by circular dichroism (CD) on an AVIV model 202 CD spectrometer at 25 °C in a 1 mm cuvette with 2–3.5 μM DHFR in 45 mM Na₂HPO₄, pH 7.0 buffer. The respective buffer spectra were subtracted from all samples.

Differential scanning calorimetry (DSC)

The thermal unfolding of fluorinated R67 DHFRs (63-100 μM) were measured using a MicroCal VP-DSC in 45 mM Na₂HPO₄, pH 7.0. Samples were scanned from 30 to 95 °C at a scan rate of 60 °C/hour. Data were fitted to a non-two state model in the Origin software supplied by the MicroCal.

Isothermal titration calorimetry

Binding affinities, stoichiometries, and enthalpies of binding were determined using isothermal titration calorimetry (ITC) as described by Bradrick et al. Isotherms were obtained at 25 °C on a MicroCal VP-ITC. NADP⁺ (14-25 mM) was titrated into 150-250 μM R67 DHFR in MTA buffer, pH 8. DHF (3-5 mM) was titrated into the binary complex of 150-250 μM R67 DHFR saturated with 2 mM NADP⁺. Experiments were performed at least in duplicate. The peaks were integrated in NITPIC, and the data sets were globally analyzed in SEDPHAT. A single-site model (A + B ↔ AB) was used for fitting.

Steady-state kinetics

Steady-state kinetic data were measured at 30 °C in MTA buffer, pH 7.0 using a PerkinElmer λ35 spectrophotometer. Ligand concentrations of DHF and NADPH were determined at 360 nm using their respective extinction coefficients, 2630 M⁻¹ cm⁻¹ and 4020 M⁻¹ cm⁻¹, and an extinction coefficient for the reaction of 5020 M⁻¹ cm⁻¹. Catalysis was measured in either 2 mm or 10 mm cuvettes. NADPH (0-1500 μM) oxidation was achieved with R67 DHFR (13-47 nM) saturated with 350 μM DHF. Likewise, DHF (0-600 μM) reduction by NADPH (1200 μM) saturated R67 DHFR (13-47 nM) was measured to determine the k_cat and K_m. The data were fit to the nonlinear Michaelis-Menten equation using Sigma Plot 12.

Results

Percent fluorine incorporation

The level of fluorine incorporation into R67 DHFR was measured by ¹⁹F NMR. There are two tryptophans per monomer of R67 DHFR, which accounted for the two peaks noted in the ¹⁹F NMR spectrum (Figure 2.2). As both peaks were broad and not well resolved from each other, the total area underneath both peaks was used to measure the percent incorporation for both tryptophans. All four fluoroindoles were incorporated at levels greater than 90%.
Figure 2.2. Integrated NMR spectra for R67 DHFR labeled with fluorooindoles in 10 mM Tris-d11, 5 mM trifluoroacetic acid, pH 8.0. Data are shown for A) 0.72 mM 4-, B) 0.63 mM 5-, C) 0.95 mM 6-, and D) 0.25 mM 7F R67 DHFR and contained 2 mM trifluoroacetic acid.
Effects of fluorine incorporation on the structure and function of R67 DHFR

The CD spectra for the labeled proteins were similar to unlabeled R67 DHFR which indicated a folded structure (Figure 2.3). The presence of a minimum at 205 nm for all samples is consistent with the presence of a disordered N-termini. These results suggest that fluorotryptophans did not significantly perturb the secondary structure of R67 DHFR.

The effects of fluorine incorporation on the tetrameric structure of R67 DHFR were explored using the pH-sensitivity of the tetramer and the change in fluorescence that occurs upon dimer formation (Figure 2.4). The tetramer is destabilized by mutations of residues, including W38, that are part of the interface. Fluorination of W38, more than W45, which is at the monomer-monomer interface, could potentially destabilize the R67 DHFR tetramer as well. For unlabeled protein a $K_{\text{overall}}$ of $4.2 \pm 0.5 \times 10^{-14}$ M$^2$ was obtained (Table 2.1), which was slightly higher than for the normal length R67 DHFR that does not possess the his-tag extension of its N-terminus, and may be due to the longer N-terminus in the pRSETb construct. Labeling of tryptophans with fluorines in the indole rings did impact the tetramer stability slightly (Figure 2.4 and Table 2.1). The tetramer was 2-fold less stable for 4F R67 DHFR, while 7F was only slightly less stable than the unlabeled enzymes. Incorporation of fluorine at positions 5 and 6 of the indole ring increased the stability of the R67 DHFR tetramer to pH dissociation 0.5- and 4-fold, respectively.

There were two transitions in the thermal unfolding of unlabeled and labeled R67 DHFRs in agreement with a two state unfolding process where a tetramer dissociates to two dimers (Figure 2.5). R67 DHFR unfolding was not reversible, therefore any thermodynamic parameters obtained from DSC are only approximate. This is in contrast to the 92% reversibility of the normal length R67 DHFR at pH 8. It was unclear why the longer N-termini would make the unfolding of R67 DHFR irreversible. The $T_{\text{m}}$s for unlabeled R67 DHFR, however, were similar to previously reported for the normal length R67 DHFR (Table 2.2). Fluorine incorporation at positions 5 and 6 of the indole ring increased the $T_{\text{m}}$s of R67 DHFR 2-3 °C compared to the unlabeled protein (Figure 2.6 and Table 2.2). 4F and 7F R67 DHFR decreased the $T_{\text{m}}$s by 3 °C and ~1 °C, respectively. Differences in protein thermal stability due to fluorotryptophan substitutions have been associated with either the structural perturbations to the protein caused by fluorines at particular positions in the residue or due to the chemical nature of fluorine that favors interaction with other fluorines over hydrocarbons and hydrophilic groups, i.e. the fluorous effect.

Fluorine incorporation effects on the activity of R67 DHFR

Incorporation of fluorotryptophans may also alter the function of R67 DHFR, therefore enzyme kinetics were monitored. The unlabeled R67 DHFR had a $k_{\text{cat}}$ of 1.0 s$^{-1}$, which was similar to the normal length construct, while the $K_{\text{m}}$s of 10 μM for DHF and 9 μM for NADPH were approximately two-fold higher (Table 2.3). This is consistent with the N-terminus not having a role in the activity of R67 DHFR. The $k_{\text{cat}}$s for 4F, 5F and 6F were 2-fold faster than the unlabeled protein, with similar DHF $K_{\text{m}}$s as well (Table 2.3). Likewise, fluorine at the 6 position in the indole ring did not affect the $K_{\text{m}}$ for NADPH binding, but 2-fold higher $K_{\text{m}}$s were observed for 4F and 5F. 7F R67 DHFR showed the greatest effects on R67 DHFR function. While the $k_{\text{cat}}$ for 7F was similar to WT, the $K_{\text{m}}$s were 3- and 7-fold higher for DHF and NADPH, respectively. The position
Figure 2.3. Fluorine incorporation minimally affected the secondary structure of R67 DHFR. CD spectra for unlabeled (green line), 4F (blue line), 5F (red line), 6F (dark yellow line), and 7F (orange line) R67 DHFRs.
Table 2.1. Best-fit values for the pH-dependent dissociation of the tetramer + 2nH⁺ ↔ 2 dimers-Hₙ

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$K_{overall}$ (M²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlabeled R67 DHFR</td>
<td>4.2 ± 0.5 x 10⁻¹⁴</td>
</tr>
<tr>
<td>4F R67 DHFR</td>
<td>1.9 ± 0.3 x 10⁻¹⁴</td>
</tr>
<tr>
<td>5F R67 DHFR</td>
<td>7.0 ± 1.4 x 10⁻¹⁴</td>
</tr>
<tr>
<td>6F R67 DHFR</td>
<td>1.8 ± 0.4 x 10⁻¹³</td>
</tr>
<tr>
<td>7F R67 DHFR</td>
<td>2.8 ± 0.5 x 10⁻¹⁴</td>
</tr>
</tbody>
</table>
Figure 2.4. pH titration of R67 DHFR monitored by intrinsic fluorescence for unlabeled R67 DHFR, (●), 4F R67 DHFR (△), 5F R67 DHFR (■), 6F R67 DHFR (▲), and 7F R67 DHFR (★).
Figure 2.5. DSC thermogram for unlabeled R67 DHFR. R67 DHFR (solid line) was fit to a two-state unfolding model (dotted lines) with two thermal transitions (dashed lines) which was consistent with tetramer dissociation into two dimers and unfolding of the dimers.
Figure 2.6. Fluorination of R67 DHFR altered the thermal stability of the protein as measured by differential scanning calorimetry. DSC thermograms were obtained for unlabeled (green line), 4F (blue line), 5F (red line), 6F (yellow line), and 7F (orange line) R67 DHFRs (63-100 µM) in 45 mM Na₂HPO₄, pH 7.0 buffer.
Table 2.2. Comparison of steady-state kinetics parameters of fluorinated R67 DHFR in MTA buffer, pH 7 at 30 °C.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_{m}$ DHF (µM)</th>
<th>$K_{m}$ NADPH (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>unlabeled R67 DHFR</td>
<td>1.0 ± 0.1</td>
<td>10 ± 1</td>
<td>9 ± 1</td>
</tr>
<tr>
<td>4F R67 DHFR</td>
<td>1.9 ± 0.1</td>
<td>10 ± 1</td>
<td>22 ± 4</td>
</tr>
<tr>
<td>5F R67 DHFR</td>
<td>2.0 ± 0.1</td>
<td>12 ± 2</td>
<td>21 ± 2</td>
</tr>
<tr>
<td>6F R67 DHFR</td>
<td>1.9 ± 0.1</td>
<td>11 ± 1</td>
<td>13 ± 2</td>
</tr>
<tr>
<td>7F R67 DHFR</td>
<td>0.9 ± 0.1</td>
<td>27 ± 6</td>
<td>67 ± 9</td>
</tr>
</tbody>
</table>
of fluorines in the benzyl ring of phenylalanine in the endonuclease, \textit{Pvu} II, similarly impacted the effect of fluorine incorporation on enzyme activity.\textsuperscript{36}

**Impact of fluorine labeling on ligand binding**

Since the $K_m$S for NADPH, particularly, were impacted by incorporation of fluoroindoles, ligand binding was measured by ITC (Figure 2.7). Binding of NADP$^+$ to unlabeled R67 DHFR yielded a $K_d$ of 210 ± 70 µM (Table 2.3), which is two-fold higher than the normal length R67 DHFR construct.\textsuperscript{17} A similar two-fold higher $K_m$ for NADPH was also noted for the labeled histagged proteins compared to the normal length enzyme.\textsuperscript{24} NADP$^+$ binding to the fluorine-labeled R67 DHFRs resulted in $K_d$ values within error of the unlabeled enzyme for 4F and 6F R67 DHFR, while 5F and 7F R67 DHFR were 1.5 and 4-fold higher (Table 2.3).

Binding of DHF to the unlabeled R67 DHFR-NADP$^+$ elicited a $K_d$ of 7.3 ± 0.5 µM (Table 2.5), which was only slightly (1.5-fold) higher than the normal length R67 DHFR.\textsuperscript{17} DHF binding for 4F R67 DHFR yielded $K_d$S 1.2-fold lower, within error for 6F R67 DHFR, and 1.3 to 2-fold higher for 5F and 7F R67 DHFR.

**Discussion**

**Effects of fluorine incorporation on R67 DHFR stability and structure**

Fluorine incorporation is thought to be a conservative replacement as the van der Waals radius is 0.2 Å larger than hydrogen.\textsuperscript{35} This would be exemplified by incorporation of 5F-tryptophan in the hydrophobic core of GB, which did not affect the global and local structure and had no influence in the thermodynamic stability.\textsuperscript{9} Similarly, fluorine incorporation in phenylalanine, or in multiple positions of tryptophan elicited no significant differences in stability for Cold shock protein B from \textit{Bacillus subtilis}.\textsuperscript{12} However, in other cases, fluorine incorporation did alter protein stability. Incorporation of 5-fluoroaleucine residues in the core ubiquitin decreased the $T_m$ by 8 °C.\textsuperscript{37} In contrast, stability of 4-helix bundle proteins or leucine zipper domains increased when highly fluorinated residues were placed in the hydrophobicity cores of the proteins.\textsuperscript{35, 38-42} We found similar effects of fluorine incorporation on thermal unfolding of R67 DHFR, and this perturbation depended upon the position of the fluorine atoms in the indole ring. Although fluorine incorporation did not perturb the secondary structure of fluorinated R67 DHFR, there were a mixture of effects on the thermal unfolding of the protein. Incorporation of 4-fluoroindole, and to a lesser extent 7-fluoroindole had mildly negative effects on the thermal unfolding of R67 DHFR and also increased the pH at which the tetramer dissociates. In contrast, both 5F and 6F R67 DHFR had higher denaturation temperatures, and increased the stability of the tetramer to dissociation at low pHs. These data indicated that the effects of fluorine on the melting and tetramer dissociation of R67 DHFR were dependent upon the position of the fluorine in the indole ring. Similar positional effects of fluorine incorporation on protein stability were noted for human transthyretin.\textsuperscript{11} 6-fluorotryptophan-substituted human transthyretin unfolded at lower urea concentrations, while 5-fluorotryptophan increased the urea stability relative to the unlabeled protein.
Figure 2.7. Example titration of NADP$^+$ and DHF binding to 5F R67 DHFR at 25 °C. A) Raw thermogram for the titration of 10.2 mM NADP$^+$ into 232 μM apo 5F R67 DHFR in MTA, pH 7.0. B) Fit (red line) of the integrate isotherm to a single site model. C) Raw thermogram for the binding of 4.3 mM DHF to 157 μM 5F R67 DHFR saturated with 2 mM NADP$. D) Integrated isotherm fit (red line) to a single site model. Lower panels in B) and D) show the residuals for the fits.
<table>
<thead>
<tr>
<th>Protein</th>
<th>n</th>
<th>$K_d$ (μM)</th>
<th>$\Delta G$ (kcal/mol)</th>
<th>$\Delta H$ (kcal/mol)</th>
<th>$T\Delta S$ (kcal/mol/K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlabeled R67 DHFR</td>
<td>1.0 ± 0.1</td>
<td>210 ± 70</td>
<td>-5.0 ± 0.2</td>
<td>-8.1 ± 2.1</td>
<td>-3.2</td>
</tr>
<tr>
<td>4F R67 DHFR</td>
<td>1.1 ± 0.1</td>
<td>230 ± 40</td>
<td>-5.0 ± 0.1</td>
<td>-4.2 ± 0.3</td>
<td>0.8</td>
</tr>
<tr>
<td>5F R67 DHFR</td>
<td>1.3 ± 0.1</td>
<td>390 ± 80</td>
<td>-4.7 ± 0.1</td>
<td>-5.0 ± 1.0</td>
<td>-0.4</td>
</tr>
<tr>
<td>6F R67 DHFR</td>
<td>1.2 ± 0.1</td>
<td>190 ± 20</td>
<td>-5.1 ± 0.1</td>
<td>-7.0 ± 0.2</td>
<td>-1.9</td>
</tr>
<tr>
<td>7F R67 DHFR</td>
<td>1.5 ± 0.2</td>
<td>840 ± 50</td>
<td>-4.2 ± 0.1</td>
<td>-6.2 ± 1.6</td>
<td>-2.0</td>
</tr>
</tbody>
</table>

Table 2.3: ITC thermodynamic parameters from the global fits of NADP$^+$ binding to either unlabeled or labeled apo-R67 DHFR at 25 °C in MTA, pH 8.0 buffer.
Table 2.4: Thermodynamic parameters obtained from ITC for DHF binding to the binary complex of R67 DHFR-NADP⁺ in MTA, pH 8.0 buffer at 25 °C.

<table>
<thead>
<tr>
<th>Protein</th>
<th>n</th>
<th>$K_d$ (μM)</th>
<th>$\Delta G$ (kcal/mol)</th>
<th>$\Delta H$ (kcal/mol)</th>
<th>$T\Delta S$ (kcal/mol/K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlabeled R67 DHFR</td>
<td>1.1 ± 0.1</td>
<td>7.3 ± 0.5</td>
<td>-7.0 ± 0.1</td>
<td>-14.0 ± 0.3</td>
<td>-7.0</td>
</tr>
<tr>
<td>4F R67 DHFR</td>
<td>1.1 ± 0.1</td>
<td>5.7 ± 0.4</td>
<td>-7.2 ± 0.1</td>
<td>-10.9 ± 0.2</td>
<td>-3.8</td>
</tr>
<tr>
<td>5F R67 DHFR</td>
<td>1.0 ± 0.1</td>
<td>9.8 ± 0.4</td>
<td>-6.8 ± 0.1</td>
<td>-13.0 ± 0.2</td>
<td>-6.2</td>
</tr>
<tr>
<td>6F R67 DHFR</td>
<td>1.2 ± 0.1</td>
<td>7.0 ± 0.3</td>
<td>-7.0 ± 0.1</td>
<td>-15.3 ± 0.1</td>
<td>-8.3</td>
</tr>
<tr>
<td>7F R67 DHFR</td>
<td>1.2 ± 0.1</td>
<td>14 ± 3</td>
<td>-6.6 ± 0.1</td>
<td>-16.4 ± 1.0</td>
<td>-9.8</td>
</tr>
</tbody>
</table>
The position of fluorine in the indole ring seems to have caused the effects on R67 DHFR stability as 5F and 6F R67 DHFR enhanced stability, while 4F and 7F decreased stability. It is unlikely that W45 contributed to the changes in the Tm of R67 DHFR as the W45F mutant does not significantly affect guanidinium HCl unfolding of dimeric R67 DHFR. Therefore, the alteration of R67 DHFR stability most likely arise from interactions between the fluorines present in W38 indole rings at the dimer interface. We examined the crystal structure of R67 DHFR to determine if there are any potential interactions between W38 and other residues in the enzyme that may explain the effects of fluorine incorporation. Inspection of the crystal structure suggests that 4-fluoroindole in W38 could affect the packing of R67 DHFR in two ways. First, the proximity of the fluorine atoms at position 4 of the indole ring to the carbonyl of the amide backbone of A37 suggests a potential steric clash (Figure 2.8). Replacing a hydrogen atom for a fluorine decreases the interatomic distance 0.4 Å in Pymol. This shorter distance suggests that A37 and W38 will be pushed further apart due to the proximity of the electronegative fluorine and carbonyl oxygen atoms. Electron-rich atoms as fluorine are repelled by the partial negative charge of the carbonyl oxygens. It is unclear how the repulsive interaction between the fluorine and the A37 carbonyl oxygen would destabilize the protein other than through general effects on hydrogen bonding in the β-strands or generally affecting the protein structure. A second reason why 4F R67 DHFR might be slightly destabilized is that the fluorine at position four could disrupt the packing of H62 from the adjacent protomer at the dimer interface (Figure 2.9). W38 is stacked against H62, which is also involved in pH-dependent tetramer to dimer stability. Though, incorporation of 4-fluoroindole did not significantly affect the pH-dependent dissociation of R67 DHFR, perhaps the fluorine at the fourth position of the indole ring has particularly destabilizing effects on the stacking of W38 and H62, relative to the fluorines in the other positions in the indole ring.

The increased thermal stability for 5F or 6F R67 DHFR likely arises from the close contact between W38 residues in each protomer at the dimer interface. Replacement of a hydrogen atoms by fluorine increases the van der Waals radius and the hydrophobicity of the molecule while preserving the shape of the residue. It is likely that due to the stacking arrangement of W38, where the edge of the indole ring that contains the 5 and 6 carbons forms van der Waals interactions, the larger surface areas of the fluorines contribute to stability by forming increased contacts (Figure 2.10). The increased stability of the 6F R67 DHFR is consistent with the tighter tetramer formation of this fluorine probe (Figure 2.4), which was somewhat surprising as mutation of W38 to phenylalanine, as well as other mutations at the dimer-dimer interface, destabilized the tetramer forming only dimeric R67 DHFR.

The similar Kd's and Km's obtained for both ligands with 4F, 5F, and 6F R67 DHFR indicated that fluorine incorporation at these positions did not significantly affect ligand binding. In general, for R67 DHFR the rate limiting step is hydride transfer, which is much slower than substrate binding, and suggests that the Km's mostly report on substrate or cofactor binding, and are less affected by other steps in the kinetic mechanism. This suggests that the effects of these fluorine probes on the thermal unfolding of R67 DHFR did not notably alter ligand binding to the protein.
Figure 2.8. The proximity of the fluorine atoms at position 4 of the indole ring of W38 to the carbonyl of the amide backbone of A37 suggests a potential for steric clash. A) Space fill model for A37 and W38 with the carbon atoms in the indole ring numbered showing the distance between the proton on the C4 in the indole ring and the carbonyl oxygen. B) Replacement of the proton on C4 with fluorine (cyan) in Pymol indicates that there could be a potential steric clash between the fluorine and carbonyl oxygen (red) of A37 in the structure (PDB ID: 2RK1).
Figure 2.9. The proximity in R67 DHFR (PDB ID: 2RK1) of the fluorine atom at position 4 of W38 (black stick model) could disrupt the packing of H62 (magenta stick model) from the adjacent protomer at the dimer interface when A) a proton is attached to C4 and B) a fluorine (cyan) is bonded to C4.
Figure 2.10. Fluorine replacement of a hydrogen atom increases the surface area of van der Waals contact which could explain the increase in thermostability for 5F and 6F R67 DHFR (PDB ID: 2RK1)\textsuperscript{18}. A) Space filled structure of W38 (black) where the indole carbons have been numbered and possess a proton on C5. B) 5F W38 (space fill cyan) indicates a larger van der Waals contact surface area.
Figure 2.11. The proximity of the fluorine atom at position 7 could sterically clash with the nearby carbonyl of S65 in R67 DHFR (PDB ID: 2RK1)\textsuperscript{18}. A) W38 (black) with a proton attached to C7 and S65 (orange) in space fill models. B) Replacement of the C7 proton with a fluorine (cyan) in Pymol\textsuperscript{43} suggests that there may be a potential for steric clash between the fluorine and the carbonyl oxygen of S65 (red).
In contrast to the other three fluorine probes, 7F R67 DHFR displayed weaker binding for NADP⁺, and a 7-fold higher Kₘ for NADPH than for the unlabeled protein. Even though 7-fluoroindole did not affect the structure or unfolding temperature of R67 DHFR, it perturbed ligand binding. Likewise, 7F R67 DHFR had a 3-fold higher Kₛ and Kₘ than the other fluorine-labeling positions, consistent with a fluorine in position 7 decreasing NADP⁺ binding. This indicates that the effects of 7F-indole are on the active site of R67 DHFR, since both ligands are similarly affected, and these effects most likely are manifested through the structure of the enzyme. Analysis of the crystal structure suggests that fluorine at position seven of W38 could sterically clash with the nearby carbonyl of S65 (Figure 2.11). The latter residue forms part of the binding surface of R67 DHFR as well as part of a β-strand that is located next to hydrogen bonding network involving Q67 and Y69.¹⁸ Although S65 does not interact directly with NADP⁺/NADPH, it does samples large shifts in the HSQC spectrum upon NADP⁺ binding. Most likely any effect of 7F W38 on S65 produces structural perturbations in the active site by altering the positioning of the β-strand containing the active site residues. The altered position of the active site residues could result in a less optimal orientation of the ligands in the active site, and therefore weaken binding.³² These interactions may not significantly affect the stability of R67 DHFR since S65, Q67, and Y69 are at the surface of the active site pore in the middle of the R67 DHFR tetramer. Though, mutations in the active site that do not significantly impact the Kₘs, they can decrease the stability of R67 DHFR by ~10 °C.³⁰

Conclusions

R67 DHFR was labeled with fluorotryptophans containing fluorine at one of four positions in the indole ring. None of the probes altered the secondary structure of the protein. We found that the location of fluorine in the indole ring did have a slight effect on the unfolding temperatures of the protein as well as the tetramer dissociation with pH. Both 4F and 7F decreased the Tₘs slightly and made the tetramer more sensitive to pH, while 5F and 6F increased both the Tₘs and tetramer stability. Neither the Kₛ nor Kₘs were affected by fluorine probes at positions 4, 5 and 6, and even enhanced the kₗₛ 2-fold. Conversely, 7F R67 DHFR decreased ligand affinity in both ITC and enzyme kinetics, indicating that fluorine in this position of the indole ring may affect the structure such that it weakens ligand binding. Our results indicate that the effects of the fluorotryptophans on R67 DHFR ligand binding, enzyme activity, and stability of the tetramer are dependent upon position of the fluorine. These results will help advance the understanding of how fluorines impact protein stability. Additionally, incorporation of fluorine at positions 4, 5, and 6 into R67 DHFR could be used in further studies of the enzyme by NMR.
References


Chapter 3. Ligand binding studies of a plasmid encoded dihydrofolate reductase by fluorine nuclear magnetic resonance
This part is a version of a manuscript in preparation entitled: Ligand binding studies of a plasmid encoded dihydrofolate reductase by fluorine nuclear magnetic resonance.

Gabriel J. Fuente-Gómez, Creighton L. Kellum, Alexis C. Miranda, Michael R. Duff, Jr.*, and Elizabeth E. Howell

Department of Biochemistry & Cellular and Molecular Biology Department, University of Tennessee-Knoxville, Knoxville, Tennessee 37996, United States

Corresponding Author: Michael R. Duff, Jr., mduff5@utk.edu

Funding: This work was supported by NIH grant GM 110669 (to EEH).
Abstract

R67 dihydrofolate reductase (R67 DHFR) is a plasmid-encoded enzyme that confers resistance to the antibacterial drug trimethoprim. The R67 DHFR active site is unusual as both cofactor and substrate are recognized by symmetry-related residues. Such promiscuity has limited our previous efforts to differentiate binding by NMR. To address this problem, we incorporated fluorine at positions 4, 5, 6, or 7 of the indole ring of tryptophans 38 and 45 and characterized the spectra to determine which probe is optimal for studying ligand binding. Two resonances were observed in the NMR spectra for all apo proteins. Unexpectedly, the W45 resonance appears broad, suggesting a dynamic process and interaction with the disordered N-terminus of R67 DHFR. Truncation of first 16 disordered residues resulted in the appearance of two new sharp resonances for W45 consistent with two conformations of the indole ring in the crystal structure. Addition of ligands caused changes to the W38 peak despite not being in direct contact with the ligand, while W45 remained unchanged. Binding of the cofactor broadened W38 for all fluorine probes, while dihydrofolate (DHF) binding resulted in the appearance of three new resonances for 4- and 5-labeled protein and severe line broadening for 6- and 7-labeled R67 DHFR. The split of W38 upon DHF binding is associated with breaking the symmetry of the tetramer and may reflect the dynamic nature of the glutamate tail of DHF. Our NMR analysis suggests that 4- and 5-labeled R67 DHFR are the best probes to differentiate ligand-protein interactions, as these positions offer the best peak resolution upon ligand binding.

Introduction

Drug resistance enzymes are a major medical problem that lead to the ineffectiveness of antibiotics. One such enzyme targeted by antibiotics is dihydrofolate reductase (DHFR). DHFRs catalyze the reduction of dihydrofolate (DFH) to tetrahydrofolate (THF) using nicotinamide adenine dinucleotide phosphate (NADPH) as a cofactor. THF is a carrier of carbon groups that are required by cells for the synthesis of thymidine, methionine and other important metabolic intermediates. Inhibition of this enzyme can lead to cell death, thus DHFR has been a target for numerous drugs, such as the antibacterial drug trimethoprim (TMP). Excessive use and misuse of TMP has resulted in the development of drug resistance in the form of a highly transmissible plasmid encoding the protein R67 DHFR. R67 DHFR shares no sequence or structural homology with chromosomal DHFR despite both enzymes catalyzing the same reaction. Due to the structural differences between R67 DHFR and chromosomal DHFR, including the active site, none of the canonical drugs that target chromosomal DHFRs (i.e., TMP, pyrimethamine, or methotrexate) significantly inhibit R67 DHFR. Only recently have two studies designed drugs that will specifically target R67 DHFR with micromolar inhibition constants.

R67 DHFR is a homotetramer with a single active site in a pore at the center of the structure (Figure 2.1). Each R67 monomer is 78 amino acids long, composed of five antiparallel β-strands. Association of three β-strands from one monomer with three β-strands from the second monomer form a β-barrel at the dimer interface. Two dimers then associate to form an active tetramer with 222-symmetry. Each dimer-dimer interface is stabilized by interactions between symmetry-related tryptophan 38 and histidine 62 residues. Two W38 residues on each side of the pore contribute to the packing by forming hydrophobic interactions while protonation of H62 destabilizes the tetramer into two dimers. The active site of R67 DHFR is unusual as the cavity transverses the length of the protein and is mostly water accessible. Furthermore, the 222-symmetry of R67 DHFR
means that the V66, Q67, I68, and Y69 residues on each of the four monomers make up the active site surface, which results in binding sites on both sides of the pore that can bind either the substrate or the cofactor. Therefore, it is possible to form nonproductive complexes between two substrates or two cofactors as well as the productive complex with one substrate and one cofactor.

The pterin ring of DHF forms the primary contacts with residues in the active site of R67 DHFR, while the tail interacts with lysine residues at the top of the pore. Crystal structure and molecular dynamics simulations indicated that the tail of the substrate is dynamic, and crosslinking folate to lysine residues at the top of the R67 DHFR pore reduces the catalytic rate <30-fold indicating that the dynamic movement of the glutamate tail of folates is integral to the catalysis. NADP$^+$ binding monitored by HSQC NMR found shifts in several active site, as well as other nearby, residues. However, the slow exchange kinetics of the tighter binding DHF substrate broke up the symmetry in the protein, which increased the number of signals making the R67 DHFR-NADP$^+$-DHF spectra difficult to analyze. In this regard, $^{19}$F NMR could provide an alternative, less complicated, approach to study ligand binding to R67 DHFR.

$^{19}$F NMR offers several advantages to study ligand binding. First, fluorine chemical shifts are extremely sensitive to changes in local chemical environments. Second, the fluorine nucleus is almost totally absent from biological systems, which eliminates background signals and facilitates interpretation of the NMR spectra. Third, aromatic residues, which are low in abundance, can be labeled specifically, helping to minimize the complexity of the spectra and the possibility of structural perturbations. Fourth, fluorine incorporation has become more accessible without the necessity to add fluoro-tryptophan, tryptophan synthesis inhibitors, and the use of knock-out strains. Finally, $^{19}$F NMR has been successfully applied to characterize protein-ligand interactions, and for drug discovery. Such studies highlight the sensitivity of fluorine NMR to monitor ligand binding changes by observing changes in chemical shift and line shapes. Therefore, we sought to use $^{19}$F labeling of the tryptophan residues to obtain complementary information regarding the ligand binding process of R67 DHFR.

Materials and Methods

Expression and purification of $^{19}$F-labeled R67 DHFR

Fluorine-labeled tryptophans were incorporated as described by Crowley et al. To facilitate protein purification, a his-tagged construct of R67 DHFR (UniProt ID P00383) cloned into a pRSETb vector, which has an additional 30 residues added to the disordered N-terminus compared to the normal length R67 DHFR, was expressed using pLysS (DE3) cells. A W45F mutant was made using Agilent’s QuickChange mutagenesis kit as described by West et al. Cells were grown in Terrific Broth media at 37 °C to an O.D. 600 of 0.6. The cells were harvested by centrifugation and the pellet was resuspended in an equal volume of pre-warmed minimal medium (7.5 mM (NH$_4$)$_2$SO$_4$, 50 mM Na$_2$HPO$_4$ and 50 mM KH$_2$PO$_4$ pH 7.0), 4-, 5-, 6- or 7-fluoroindoles were dissolved in dimethyl sulfoxide and added to the media (final concentration of 60 mg/L). After 30 minutes of acclimatization in the presence of fluoroindole, protein expression was induced with the addition of 1 mM IPTG for ten hours before harvesting and lysing the cells. Protein was purified on a nickel-NTA column (Qiagen). Purified protein was dialyzed against deionized H$_2$O, lyophilized, and stored at 4 °C. From now on we will refer to R67 DHFR labeled with 4-, 5-, 6-, or 7-fluoroindole as 4F, 5F, 6F, and 7F, respectively.
α-chymotrypsin-truncated R67 DHFR was prepared as described.\textsuperscript{27, 28} α-chymotrypsin cleaves after F16 in the normal length R67 DHFR sequence (without the his-tag), or after F47 in the his-tagged sequence, resulting in a 62-amino acid monomer that assembles into an active tetramer, but without the disordered N-termini. Protein concentrations were determined by measuring the absorbance at 280 nm of the solution using an extinction coefficient determined with a bicinchoninic acid assay (Pierce).

NMR experiments

Lyophilized R67 DHFR was dissolved in 10 mM Tris-d\textsubscript{11} buffer, pH 8.0 in D\textsubscript{2}O. The total tetramer concentration ranged between 0.1 to 0.75 mM. \textsuperscript{19}F NMR measurements were acquired on a Varian 500 MHz spectrophotometer equipped with a OneNMR probe operating at 470 mHz. \textsuperscript{19}F spectra were obtained with 512 scans, a spectral width of 230 ppm, relaxation delay of 1 second, and 87 degrees pulse angle. The \textsuperscript{19}F chemical shifts were referenced to 10 mM trifluoroacetic acid (TFA). All NMR data were processed using MestReNova (Mestrelab). A 20 Hz line-broadening was applied to each spectrum.

Determination of fluorine incorporation levels

The spectrum of apo R67 DHFR was first integrated and the area compared to an internal 10 mM trifluoroacetic acid standard. Next, the total protein concentration in the NMR sample was calculated using the extinction coefficient.

To determine whether the presence of the W38 shoulder corresponds to incomplete incorporation of 5- and 6-fluoroindole into R67 DHFR, we took advantage of the tetramer to dimer dissociation of R67 DHFR. The NMR spectra for 5F and 6F R67 DHFR were obtained at pH 8. Next, the fluorine labeled R67 DHFR was mixed with a molar equivalent of unlabeled R67 DHFR, equilibrated for 5 min, and the spectra were recorded. Subsequently, the pH of the mixture of labeled and unlabeled R67 DHFR was decreased to pH 5 in order to produce dimeric R67 DHFR and the spectra recorded. Finally, the pH of the solution was adjusted back to pH 8 to reform the tetramer, which should be a mixture of labeled and unlabeled R67 DHFR dimers, and the NMR spectra were recorded.

Ligand binding studies by \textsuperscript{19}F NMR

NMR titrations were performed at 25 °C by titrating 0.6-1 mM \textsuperscript{19}F-R67 DHFR with increasing concentrations (0-20 mM) of the cofactor NADP\textsuperscript{+}. Binding data for the ternary complex were obtained by titrating DHF (0-3.5 mM) into a binary complex of R67 DHFR (0.17-1 mM) saturated with of 6 mM NADP\textsuperscript{+}. Ligand concentrations were measured using their respective extinction coefficients, 28,000 M\textsuperscript{-1} cm\textsuperscript{-1} at 282 nm for DHF and 18,000 M\textsuperscript{-1} cm\textsuperscript{-1} at 259 nm for NADP\textsuperscript{+}. Dissociation constants were calculated by plotting either changes in chemical shifts (for 5F R67 DHFR) or peak intensities (for 4F, 6F, and 7F R67 DHFR) as a function of ligand concentration and fitting the curve to Eq. 1:\textsuperscript{29}

\[
\frac{L_P}{P} = \frac{-K_d + (L+P)}{2P} - \frac{(K_d + L + P)^2 - 4B}{2P}
\]

(Eq. 1)
where B is the concentration of ligand bound to R67 DHFR, K_d is the dissociation constant of the ligand, L is the total ligand concentration, and P is the total protein concentration. The data were fit using Sigma Plot 12.

Results

19F NMR assignment of R67 DHFR

Incorporation of the 19F probe at four different positions of the indole ring (Figure 2.1D) of the tryptophans of R67 DHFR resulted in distinct NMR spectra in terms of chemical shifts and line width (Figure 3.1). Two peaks were observed for all fluorinated proteins, a downfield sharp peak around -44 and an upfield broad one at -47.5 ppm with the exception of 7F R67 DHFR, which had peaks that were located 12 ppm further upfield (Table 3.1). Similar upfield shift for 7F indole compared to the other indoles was noted for the apelin receptor, likely due to the close proximity of the fluorine probe to the indole nitrogen. The presence of only two peaks in the R67 DHFR spectra is consistent with the 222 symmetry of the protein, where each symmetry related W38 (or W45) residue in the tetramer samples a similar environment, and therefore the chemical shifts for all four symmetry-related tryptophan residues were either the same or significantly overlapping. Fluorine incorporation for R67 DHFR were above 90% for all the four different fluoroindoles.

The NMR assignment was achieved using the W45F mutant of R67 DHFR. Spectra for 4F, 5F, and 6F W45F R67 DHFR only contained the sharp peak, or two sharp peaks for 6F, while the broad peak was missing (Figure 3.2). Therefore, the sharp peaks that remained correspond to W38, while the broad peaks correspond to W45. Though W45F R67 DHFR was not labeled with 7-fluoroindole, it was assumed based on the assignment of the peaks for the other three fluoroindoles that the sharp peak at -56 ppm was W38 and the broad shoulder at -57 ppm was due to W45. Incomplete incorporation of 5F and 6F led to inhomogeneous line broadening. A small shoulder corresponding to approximately 10% of the area of the sharp peak was present in the 5F R67 DHFR spectrum, while there were two W38 peaks for 6F R67 DHFR (Figure 3.3). It has been reported that incomplete fluorine incorporation can lead to inhomogeneous line broadening. These shoulders in the 5F and 6F R67 DHFR spectra may correspond to partially labeled protein where the W38 on one face of the dimer is labeled, while the symmetry partner with which it interacts on the opposite side of the dimer interface is unlabeled. To determine whether these extra peaks for 5F and 6F correspond to incomplete incorporation of the fluoroindole into R67 DHFR, we took advantage of the reversible dissociation of R67 DHFR from tetramer to dimers as a function of pH. Thus, we first added in a 1:1 ratio unlabeled R67 DHFR to labeled R67 DHFR. In the mixture of labeled and unlabeled proteins the shoulder adjacent to the sharp peak increased in intensity (Figure 3.3B and F). As the pH was decreased to 5, protonation of H62 resulted in dissociation of the tetramer into dimers as a function of pH. Thus, we first added in a 1:1 ratio unlabeled R67 DHFR to labeled R67 DHFR. In the mixture of labeled and unlabeled proteins the shoulder adjacent to the sharp peak increased in intensity (Figure 3.3B and F). As the pH was decreased to 5, protonation of H62 resulted in dissociation of the tetramer into dimers. In the NMR spectra, dimerization exposes W38 to the solvent, which shifted the W38 peak upfield to -49.5 ppm for 5F and downfield to -45.0 for 6F (Figure 3.3C and G). Next, increasing the pH back to 8 returned the mixture of labeled and unlabeled dimers of R67 DHFR to tetramers. The population of R67 DHFR should be a statistical distribution of fully labeled tetramers with four W38s with fluorine incorporated in the indole ring, fully unlabeled R67 DHFR which does not contain any fluorine, and tetramers that are a mixture of one dimer that is 19F-labeled and the other that is unlabeled. This mixture of labeled and unlabeled R67 DHFR will have a chemical shift that corresponds to partially labeled R67 DHFR. Upon reformation of the tetramer, the shoulder peak increased in intensity, while the intensity of the
Figure 3.1. Fluorine incorporation in R67 DHFR exhibited distinct NMR spectra. Spectra are shown for A) 4F R67 DHFR, B) 5F R67 DHFR, C) 6F R67 DHFR (W45 is underneath the W38 peak) D) 7F R67 DHFR.
Table 3.1: $^{19}$F chemical shifts (ppm) for R67 DHFR in the indicated ligand complexes

<table>
<thead>
<tr>
<th>Condition</th>
<th>Tryptophan resonance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W38</td>
</tr>
<tr>
<td>4F R67 DHFR</td>
<td>-44.3</td>
</tr>
<tr>
<td>4F R67 DHFR + NADP*</td>
<td>-44.1</td>
</tr>
<tr>
<td>4F R67 DHFR:NADP* + DHF</td>
<td>-43.7, -44.8, -45.7</td>
</tr>
<tr>
<td>5F R67 DHFR</td>
<td>-45.4</td>
</tr>
<tr>
<td>5F R67 DHFR + NADP*</td>
<td>-44.8</td>
</tr>
<tr>
<td>5F R67 DHFR:NADP* + DHF</td>
<td>-43.2, -44.2, -44.5</td>
</tr>
<tr>
<td>6F R67 DHFR</td>
<td>-45, -46</td>
</tr>
<tr>
<td>6F R67 DHFR + NADP*</td>
<td>-45.6</td>
</tr>
<tr>
<td>6F R67 DHFR:NADP* + DHF</td>
<td>-43.3, -43.9, -44.8, -45.2, -45.6, -46.3, -46.5</td>
</tr>
<tr>
<td>7F R67 DHFR</td>
<td>-56.4</td>
</tr>
<tr>
<td>7F R67 DHFR + NADP*</td>
<td>-55.9</td>
</tr>
<tr>
<td>7F R67 DHFR:NADP* + DHF</td>
<td>-54.9, -55.1, -55.2, -55.9</td>
</tr>
</tbody>
</table>

*Not determined because of the overlap of the W38 peaks with the W45 peak.*
Figure 3.2. $^{19}$F-NMR spectra of the W45F mutant of R67 DHFR was used to assign resonances observed in the spectrum of the WT enzyme. Spectra are shown for A) WT 4F R67 DHFR, B) W45F 4F R67 DHFR, C) WT 5F R67 DHFR, D) W45F 5F R67 DHFR, E) WT 6F R67 DHFR, and F) W45F 6F R67 DHFR. The broad, upfield peak is missing from all of the W45F spectra (B, D, and F) indicating that the sharp peak is due to W38 and the broad band arises from W45.
Figure 3.3. Effects of incomplete fluorine incorporation in R67 DHFR. A) Apo 5F R67 DHFR exhibited a shoulder in W38 at -45.06 ppm. B) When unlabeled R67 DHFR is added, the shoulder increases in intensity. C) Separation of the tetramers in the mixture of labeled 5F R67 DHFR and unlabeled R67 DHFR into dimers at pH 5 and D) readjusting back to pH 8 to reform tetramers that are a mixture of labeled and fluorine-labeled proteins led to further increase in the shoulder at -45.06 ppm. A similar method was used for E) 6F R67 DHFR at pH 8, F) addition of unlabeled R67 DHFR to 6F R67 DHFR, G) the labeled and unlabeled R67 DHFR mixture at pH 5.0, and H) the mixture readjusted to pH 8, which had an increased downfield W38 peak indicating it comes from the partial labeling of 6F R67 DHFR.
upfield peak decreased. This indicates that the shoulder corresponded to the partially labeled protein, while the upfield peak arises from fully labeled R67 DHFR.

Another aspect of the NMR spectra was W45 exhibiting severe line broadening relative to W38. The broadness of W45 compared to W38 was surprising, as W38 is buried at the interface between the two dimers, while W45 is exposed to the solvent at the monomer-monomer interface where it exchanges between two orientations. The less restricted environment at the protein surface would suggest that W45 would be more dynamic, and form a sharp peak. Line broadening in the absence of ligands is often ascribed to a slow exchange mechanism between two different states. To determine whether the line broadening was caused by conformational exchange, the NMR spectra for 19F-labeled R67 DHFRs were measured at 25, 35 and 45 °C. As the temperature increased the broad W45 peaks sharpened and shifted upfield (Figure 3.4). This suggests that W45 samples two conformations in the slow exchange regime at 25 °C; in contrast, at 45 °C, the exchange rate becomes faster as evidenced by the sharpening of the peak. However, even after increasing the temperature, the W45 peak was still broader than the W38 peak which suggested broadening may arise for other reasons as well.

Transient interactions between the disordered N-terminus of R67 DHFR and W45 were investigated. The first 16-20 residues of R67 DHFR are disordered and need to be proteolytically cleaved to form the tetramer in crystal structures. In solution, the disordered N-termini collapse, interacting with themselves and the monomer-monomer interface of the core. To test whether weak, dynamic interactions between residues in the N-termini and W45 were responsible for the broader W45 signal, we truncated the first 47 residues of 4F and 5F R67 DHFR (Figure 3.5). Both the W38 and W45 peaks were sharper for the truncated protein (Figures 3.6 and 3.7), which has a molecular weight 40% lower than the full length pRSETb construct. In addition, W45 of 4F R67 DHFR splits into two new peaks at -48.2 and -48.7 ppm, while W45 sharpens into one peak for 5F R67 DHFR. Two peaks in the truncated 4F R67 DHFR spectrum indicate W45 is undergoing a slow exchange process.

Titration of NADP+ and DHF to fluorinated R67 DHFRs

Whether 19F NMR could differentiate the binding of the cofactor NADP+ from the substrate DHF, was examined by NMR titrations. Addition of NADP+ to all fluorinated proteins affected the W38 resonance while the broad W45 peak remained unchanged. Binding of NADP+ to 4F and 6F R67 DHFR caused line broadening of the W38 peak with no change in the chemical shift, while 5F and 7F R67 DHFR sampled a downfield chemical shift in addition to line broadening (Figures 3.8 and 3.9). Severe line broadening for the W38 peaks in 6F R67 DHFR caused an overlap of the W45 peak, forming a broad envelop.

Similar to the NADP+ titration, the W45 peaks remained unchanged upon DHF titration into the fluorinated R67 DHFRs, while W38 split into smaller peaks. Binding of DHF to 4F and 5F R67 DHFR-NADP+ complexes resulted in the appearance of three new peaks. Addition of DHF to the 7F binary complex caused W38 to form at least two peaks, with further line broadening. In contrast, DHF titration into the 6F R67 DHFR-NADP+ binary complex resulted in the appearance of seven new peaks with severe broadening. Overall, our NMR titrations showed that W38 sensed the presence of both ligands while W45 remains unaffected during the titrations of both NADP+ and DHF.
Figure 3.4. Temperature effects on fluorinated R67 DHFRs exhibit increased rates of exchange of W38 and W45. A) 4F R67 DHFR, B) 5F R67 DHFR, C) 6F R67 DHFR, D) 7F R67 DHFR. Data are shown for 25 °C (lower panel), 35 °C (middle panel), and 45 °C (top panel) for each R67 DHFR.
Full length pRSETb R67 DHFR
MRGSHHHHHHGMASMTGGQQMGRDLYDDDKDPSSNEVSNPVAGNFVFPSNATFGMGRV
KKSGAAWQGQIVGWYCTNLTPEGYAVESEAHPGSVQIYPVAALEIR

Truncated R67 DHFR
VFPSNATFGMGRVRKKSGAAWQGQIVGWYCTNLTPEGYAVESEAHPGSVQIYPVAALEIR

Figure 3.5. Sequence of pRSETb R67 DHFR. The α-chymotrypsin cleavage site indicated by the red arrow (upper panel) and the sequence of the truncated R67 DHFR.
Figure 3.6. W45 peaks sharpened, and split into two peaks, when the disordered N-termini were truncated from 4F R67 DHFR. A) Full length apo 4F R67 DHFR and B) 4F R67 DHFR after treatment with α-chymotrypsin.
Figure 3.7. W45 peaks sharpened when the disordered N-termini were truncated from 5F R67 DHFR. A) Full length apo 5F R67 DHFR and B) 5F R67 DHFR after treatment with α-chymotrypsin.
Figure 3.8. Representative $^{19}$F NMR spectra for the titration of 4F and 5F R67 DHFR with ligands. Titration of A) 0, 0.4, 1.2, 2.4 and 6 mM NADP$^+$ into 4F R67 DHFR, B) 0, 0.2, 0.3, 0.8 and 5 mM DHF into 4F R67 DHFR saturated with 6 mM NADP$^+$, C) 0, 0.8, 1.5, 2.4 and 9 mM NADP$^+$ to 5F R67 DHFR, and D) 0, 0.1, 0.4, 1.8 and 4.5 mM DHF into 5F R67 DHFR saturated with 6 mM NADP$^+$. 
Figure 3.9. Representative $^{19}$F NMR spectra for the titration of 6F and 7F R67 DHFR with ligands. A) Titration of 0, 0.5, 1.8, 4 and 21 mM NADP$^+$ into 6F R67 DHFR, B) titration of 0, 0.3, 0.6, 1.5 and 3.2 mM DHF into 6F R67 DHFR saturated with 6 mM NADP$^+$, C) addition of 0, 0.6, 1.9, 4 and 13 mM NADP$^+$ to 7F R67 DHFR, and D) titration of 0, 0.3, 0.6, 1.5 and 5 mM DHF into 7F R67 DHFR saturated with 6 mM NADP$^+$. 
To further explore whether it is only W38 that is and not W45, we repeated our NMR titrations with the W45F mutant (Figure 3.10). Previous fluorescence quenching suggest that W38 is sensitive to ligand binding while emission from W45 is mostly unaffected. Saturation of NADP$^+$ for 4, 5 and 6F resulted in broadening of W38 in agreement with WT results. Similarly, saturation of fluorinated proteins with DHF resulted in the splitting of W38 as WT.

To probe indole ring dynamics upon ligand binding and improve our spectra resolution, we recorded the $^{19}$F NMR spectra for 4F R67 DHFR at different temperatures. Side-chain dynamics of aromatic residues via temperature-dependent line width analysis have been used to study ligand binding of the intestinal fatty acid binding protein$^{33}$ and the molten globule state of $\alpha$-lactalbumin.$^{36}$ Increasing the temperature for the apo and binary complexes resulted in the sharpening of W38 and W45 (Figure 3.11). However, for the ternary complex, the downfield most W38 peak either resolved or split into two peaks, while the other peaks sharpened above 25 °C. For the W45 peak, sharpening was also observed at higher temperatures. The splitting of the W38 peak and the temperature dependence of line width upon ligand binding suggest a dynamic process for W38 upon DHF binding. It is likely that this dynamic behavior is caused by the motion of the pterin ring and particularly the glutamate tail of DHF switching its interaction between lysines at the top of the active site pore.$^{13,16}$

We calculated the dissociation constants of NADP$^+$ binding to R67 DHFR, or DHF binding to the R67 DHFR-NADP$^+$ binary complexes, using either peak intensities or chemical shift changes as a function of increased ligand concentrations (Tables 3.1 and 3.2). The $K_d$s obtained for NADP$^+$ binding were all similar, within error, irrespective of the fluorine probe. These results were slightly higher to those reported previously by NMR and ITC, 132 μM and 99 μM, respectively.$^{12,17}$ Most likely this difference arises from the difference in the length of the disordered N-termini. Truncated R67 DHFR was used for previous NMR experiments, while the normal length R67 DHFR was used in ITC experiments. The longer N-terminus of the pRSETb his-tag construct, with a molecular weight of 47 kDa, also increases the $K_m$s of NADPH and DHF compared to the 33 kDa normal length R67 DHFR.$^{26}$ Another possible reason for the higher $K_d$s may due to the high protein concentrations used in NMR (discussed below).

DHF binding to the 4F, 5F, and 7F R67 DHFRs yielded similar $K_d$s; however, when 6-fluoroindole was used as the probe, the $K_d$ was 3-fold higher, with an error of 100%. These values are also 6 to 8 times higher to previous ITC reported value 4.8±1.0 μM. Like the NADP$^+$ titration, the high concentrations of protein used in the titrations most likely are a factor.
Figure 3.10. $^{19}$F NMR spectra of 4F, 5F and 6F W45F R67 DHFR in the absence of ligands and saturated with NADP$^+$ and DHF. A) 4F W45F R67 DHFR, B) 5F W45F R67 DHFR, and C) 6F W45F R67 DHFR apo, binary and ternary complexes.
Figure 3.11. A representative $^{19}$F NMR spectra of 4F R67 DHFR at different temperatures. A) Apo R67 DHFR at 25 °C, 35 °C and 45 °C. B) 4F R67 DHFR bound to NADP$^+$ at 25 °C, 35 °C and 45 °C. C) 4F R67 DHFR bound to NADP$^+$ and DHF at 25 °C, 35 °C and 45 °C.
Table 3.2: Dissociation constants of R67 DHFR obtained by NMR

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>( K_d ) (µM)(^a)</th>
<th>NADP(^+)</th>
<th>DHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truncated WT</td>
<td>132 ± 51(^b)</td>
<td>ND(^c)</td>
<td></td>
</tr>
<tr>
<td>4F R67 DHFR</td>
<td>360 ± 66</td>
<td>42 ± 13</td>
<td></td>
</tr>
<tr>
<td>5F R67 DHFR</td>
<td>220 ± 20(^d)</td>
<td>33 ± 20</td>
<td></td>
</tr>
<tr>
<td>6F R67 DHFR</td>
<td>320 ± 70</td>
<td>110 ± 120</td>
<td></td>
</tr>
<tr>
<td>7F R67 DHFR</td>
<td>380 ± 60</td>
<td>30 ± 28</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)K\(_d\)s were obtained using peak intensity. \(^b\)data were obtained from reference.\(^{17}\) \(^c\)Not determined. \(^d\)K\(_d\) obtained by measuring chemical shift.
Discussion

NMR assignment

One of the many advantages of using $^{19}$F NMR is its sensitivity to changes in the chemical environment surrounding the fluorine probe. As a result, several attempts have been made to relate chemical shifts to structural features. This includes an empirical derived computational and experimental correlation between the electron distribution around the fluorine probe and its interactions with proteins. Fluorine atoms with increased electron distribution are more “shielded” from the magnetic field applied. These shielded fluorines are more likely to be in close contact with hydrogen bond donors of a protein, suggesting the possibility of hydrogen bond formation. In contrast, “deshielded” fluorine atoms are predominantly found in close contact with hydrophobic side chains and carbonyl groups of the backbone of the protein. In terms of chemical shifts, fluorine atoms that are more shielded appear at more negative ppm (upfield) values while deshielded atoms appear at more positive values (downfield). Hence, in view of these interpretations we discuss our results.

R67 DHFR is an homotetramer with two tryptophans per monomer, for a total of eight tryptophans in the protein, though only two resonances were observed due to the symmetry of structure. All four symmetry-related W38 tryptophans had the same chemical shift. Likewise, the four W45 residues had the same chemical shifts, though the W45 peaks are broader for the reasons discussed below. Single peaks for W38 and W45 are consistent with the HSQC NMR data that yield single peaks for each residue due to the symmetry of R67 DHFR. Each set of four tryptophans are in the same environment in the tetramer. The sharpness of the W38 peak is indicative of rigid structure and consistent with the presence of single conformations in the crystal structure for W38. On the other hand, broadening of W45 indicates there is an exchange process between two alternative conformations.

The chemical shifts observed for W38 are downfield compared to W45 irrespective of where the fluorine is positioned in the indole ring, and likely arise from the chemical environments surrounding the fluorinated residues. W38 and its symmetry-related partners are buried in a hydrophobic environment at the dimer-dimer interface, while the W45 residues are located at the monomer-monomer interface, exposed to the solvent. Our data were consistent with W38 being in a more hydrophobic environment slightly downfield in the NMR spectra in contrast to the upfield W45 peak, which is exposed to the solvent, possibly forming hydrogen bonds.

One of the interesting features of the NMR spectra is the broadness of the W45 peak relative to W38. The broad peak for W45 indicates a chemical exchange process is occurring on a slow timescale. Line broadening is usually associated with an exchange process between conformers. One possible explanation for the broader W45 peak is the presence of two conformations of the residue in the crystal structure (Figure 3.12). If the line broadening is the result of an exchange process between the two conformations, increasing the temperature should result in an increase of the exchange rate and narrow the W45 peak. However, while the W45 peak narrowed with an increase in temperature, the W38 peak sharpened to a similar degree. Therefore, slow exchange between the two W45 conformations did not completely explain the broadness of W45 peak, and suggested that there were additional factors that caused line broadening.
Figure 3.12. The two conformations of W45 in the crystal structure of R67 DHFR. One conformer is shown in pink, the other in cyan (PDB ID: 2RH2)\textsuperscript{15}.
Transient interaction of multiple residues in the disordered N-termini with W45 may also broaden the peak. The first 16-20 residues of R67 DHFR are disordered and collapse into compact structures, most likely interacting with themselves and the monomer-monomer interface of the core of the protein. The disordered N-termini possess many charged and hydrophobic residues that can potentially interact with the hydrophobic patch at the monomer-monomer interface where W45 is located. The broad W45 bands were consistent with our previous SANS data that suggested that the N-termini collapse and interact with the monomer-monomer interface. Indeed, truncation of the disordered N-termini resulted in the appearance of two new, sharp resonances for 4F R67 DHFR and a single sharp resonance for 5F R67 DHFR (Figures 3.6 and 3.7). The two sharp resonances in the truncated 4F R67 DHFR indicated that W45 populated two distinct states, consistent with the two conformations of W45 present in the crystal structure. Two well resolved resonances suggested a slow exchange between the two conformations of the indole ring of W45, as a fast exchange processes on the NMR timescale would have led to the appearance of a single average resonance. Integration of the areas of chemical shifts yielded populations of W45 in the two conformations at 45 and 55%, which agrees with the 45 and 55% occupancies of the two W45 conformations observed in the crystal structure. The single peak for truncated 5F R67 DHFR does not exclude an explanation of two conformations of the W45 indole ring if the environment surrounding the fluorine probe in the two indole conformations is the same for 5-fluoroindole, but not the same for 4-fluoroindole.

Quantifying ligand binding affinity.

Our NMR titrations with the four different fluorinated probes obtained binding affinities for NADP+ that were within error, indicating that all four probes are adequate to study cofactor binding. For DHF binding, 4F, 5F, and 7F R67 DHFR all had K_d within error, they were much higher than previously reported for R67 DHFR. Additionally, the K_d obtained for 6F R67 DHFR had large (100%) errors. This indicated that 19F NMR may not be able to measure substrate binding. The protein concentration used in our 19F NMR experiments may account for the discrepancy between the K_d obtained here and those from ITC. From Eq. 1, an estimate of K_d can be obtained if the protein concentration is closer to the K_d. Optimal protein concentrations should be less than half of the K_d. However, in our substrate binding experiments, the protein concentrations was about 40-fold higher than the K_d, and thus are not able to give accurate estimates of the binding affinity. Several groups have also noted differences in dissociation constants obtained from NMR when compared to surface plasmon resonance results. Lowering our R67 DHFR concentrations below 0.2 mM could potentially have improved our analysis of the K_d. However, R67 DHFR from pRSETb has a molecular weight of 47 kDa, and somewhat broad 19F peaks that hinder using lower protein concentrations in NMR. Increasing the number of scans would have improved the signal-to-noise, however DHF is a relatively unstable molecule that decomposes on ice after several hours, which precluded longer scan times.

Effects of ligand binding on changes in the 19F NMR spectra of R67 DHFR

The broadening and chemical shifts exhibited by R67 DHFR upon ligand binding may reflect changes in the local chemical environment of the fluorine probe. Binding of the cofactor, NADP+, to all fluorinated proteins resulted in the broadening of W38 resonance while W45 remained unchanged (Figures 3.8 and 3.9). This differs from the HSQC data, where both W38 and W45 shifted upon addition of NADP+, which might have been due to changes in hydrogen bonding.
between the β-sheets upon ligand binding. Perhaps $^{19}$F-NMR is less sensitive to the factors that caused the change in the chemical shift of W45 in the HSQC experiments. W45 being unresponsive to binding of NADP$^+$ is consistent with the distance of the residue form the active site and the minimal quenching of the fluorescence of W45 upon NADPH binding relative the quenching of W38 fluorescence. Broadening of W38 upon NADP$^+$ binding, on the other hand, likely arises from its vicinity to the active site. While W38 is not involved in any direct interactions with the bound ligands, the active site of R67 DHFR is composed of a multi-layered stacking arrangement that includes the indole rings of W38 residues and the hydrogen network formed between Q67 and Y69 residues. The latter two residues directly interact with the bound ligands. From the apo R67 DHFR structure, the Q67 side chain bends into the pore to form a hydrogen bond with the amide proton of I68 in the absence of ligand. In this conformation, the W38 indole rings are in a loose state. However, upon NADP$^+$ binding, the side chain of Q67 is displaced from the active site to form the hydrogen bond network between Q67 and Y69. The movement of the side chain of Q67 in turn packs against the W38 residues. Therefore, it is likely that although W38 is not in direct contact with the cofactor, the change in Q67 conformation upon NADP$^+$ binding alters the environment around the indole ring. The packing of Q67 against the indole ring may shift the W38 peak when ligand is bound. The tighter packing of W38 in the ligand-bound complex may lead to the broadening of the chemical shift relative to the looser-packed W38 in the apo structure.

Upon DHF binding, W38 split into three smaller peaks (Figures 3.8 and 3.9). The upfield-most W38 peak in 4F R67 DHFR, or the downfield most peak in 5F R67 DHFR, had twice the area of the other two smaller peaks, while the W38 peaks were less well defined in 6F and 7F R67 DHFR due to line broadening made it difficult to determine if one of the peaks had twice the area of the others. The larger W38 peaks in the 4F and 5F R67 DHFR spectra most likely describes the W38 in the half of the pore nearest the nicotinamide ring of the cofactor. The other two peaks arise from the environment sampled by bound DHF. Two symmetry related lysine 32s at the top of one side of the pore constrain the position of NADP$^+$ by forming ionic interactions with the phosphate groups. However, on the other half of the pore, the glutamate tail of DHF is not well resolved in the crystal structure and switches between forming direct ion pairs with the K32s on each side of the pore. This could result in two different environments for the W38 residues on the side of the pore with bound DHF, one where the glutamate tail is nearby, and the other where the glutamate tail is further away. Our results are consistent with NMR and computational simulations of the glutamate-tail interacting with symmetry related lysine 32 residues at the edge of the pore. Furthermore, increasing the temperature of the ternary complex caused the three peaks to become sharper and shift slightly (Figure 3.11). These temperature effects indicate an exchange process occurring that becomes faster with temperature, which is consistent with the switching of the glutamate tail interactions between the two K32s at the top of the active site pore.

One complicating factor with 6F R67 DHFR was the incomplete incorporation of fluorine probe which led to two W38 peaks of similar intensity (Figure 3.3). The downfield peak corresponded to the partially labeled protein, with only one labeled tryptophan next to an unlabeled tryptophan in at least one of the two dimer interfaces. The upfield peak arising from the fully labeled 6F R67 DHFR, where both tryptophans at the dimer interface have incorporated the fluorooindole. While the analysis of NADP$^+$ binding was not hindered by the two W38 peaks, the splitting of W38 due to DHF binding made the spectra much more complicated. A total of seven peaks were observed due to presence of the W45 peak in addition to the splitting of the both of the W38 peaks into three new peaks each (Figure 3.9B). Due to severe line broadening of the two W38
peaks in the 6F R67 DHFR-NADP$^+$ binary complex, the subsequent splitting of both peaks into three additional peaks upon DHF binding, and the broadening of the peaks it was difficult measure binding of DHF as indicated by a $K_d$ with 100% error (Table 3.2). This suggests that the 6-fluoroindole is not a good probe to use for R67 DHFR due to the incomplete labeling and the splitting of the peaks upon DHF binding. 5F R67 DHFR also had a shoulder for W38 that was due to incomplete incorporation. However, either because of its low intensity, or fortuitous chemical shifts, only three W38 peaks were noted upon DHF binding. This indicates that the incomplete labeling of 5F R67 DHFR did not affect analysis of DHF binding. Similar modest effects of incomplete labeling were noted for the bromodomain Brd4.47

Choosing an optimal fluorine probe to differentiate ligand binding in R67 DHFR

R67 DHFR active site is unusual in the way that both cofactor and substrate are recognized by symmetry-related residues. Previous studies in our lab with $^{15}$N-labeled R67 DHFR probed for ligand binding of cofactor and substrate were able to quantify cofactor binding by following chemical shift perturbations.17 However, NMR data for the ternary complex resulted in a more complex NMR spectra which prohibited analysis of DHF binding by HSQC. Labeling of tryptophans with fluorine and using $^{19}$F-NMR on the other hand was able to distinguish NADP$^+$ and DHF binding, and determine $K_d$s for both ligands. Fluorotryptophans offer great chemical shift dispersion and are highly sensitive to changes in its local environment.43, 48, 49 While each fluorine probe exhibited line broadening upon NADP binding, 4F and 5F offered the best resolution as their peaks had better chemical shift dispersion and were better defined. Others have reported that 5F yields greater chemical shift dispersion48 or better resolution49 over 4F and 6F probes. The resonances for 6F and 7F were less optimal due to overlap from broadening. It is interesting that 6F was not a good probe for R67 DHFR as it was used to monitor binding in E. coli chromosomal DHFR.43 Thus, the different structural folds, and tryptophan environments in R67 DHFR versus chromosomal DHFR yield alternate optimal fluoroindole probes for studying ligand binding.5, 34, 50 Thus seems that fluorine probe selection is protein dependent as for R67 DHFR, 6F resulted in incomplete labeling of R67 DHFR was significant, which lead to two distinct W38 peaks. The two peaks for the labeled and partly label proteins complicated the NMR spectra, particularly upon DHF binding, which caused W38 to split into seven peaks. In summary, for R67 DHFR the 4F- and 5F-labels would be the better probes for potential exploration of inhibitors of this antibiotic-resistant enzyme.

Conclusion

Fluorine NMR provided a simple and useful way to differentiate binding of the cofactor from the substrate in R67 DHFR overcoming the challenges arising from more complex $^{15}$N HSQC spectra upon ternary complex formation. Probing different fluorine positions labels allowed us to rationally select optimal probe for ligand binding detection. Incomplete incorporation of 6F-indole resulted in the appearance of two W38 peaks in the NMR spectra, which complicated the analysis of DHF binding, suggesting that 6F R67 DHFR is also not a good probe for studying ligand binding. We conclude that the best probes to monitor ligand binding for R67 DHFR are position 4F and 5F. As there are no commercially available drugs that target R67 DHFR, our results will help in the ongoing effort to develop inhibitors of this plasmid encoded DHFR.
References


Chapter 4. Conclusion and future directions
Macromolecular crowding

The interior of the cell is a crowded environment where the concentration of macromolecules can reach more than 300 mg/mL. A typical cell is composed of thousands of different proteins, DNA, RNA, lipids, sugars and small molecules like metabolites, ions, and salts, which create a vastly different environment from the one used in in vitro studies. In a typical in vitro experiment, the highest concentration of macromolecule is no more than 10 mg/mL. For a long time, biochemists have wondered about the implications of neglecting the environment of the cell interior on biological equilibriums. This work set out to determine how does the crowding environment of the cell affect the function of R67 DHFR.

In order to study the effects of macromolecular crowding, synthetic and protein crowders are added to the test tube with the idea of mimicking the cell interior. While these studies have yielded insight to the effects of crowding, we are still far from measuring biochemical equilibriums in the native cellular environment. In our own effort to study crowding, our lab showed that upon the addition of small molecules, called osmolytes, weaker binding for the substrate and tighter binding for the cofactors were observed. These results suggested possible interactions between osmolytes and the substrates and subsequent vapor pressure osmometry and NMR studies concurred with this hypothesis.

To further explore the role of protein and synthetic crowders, enzyme activity of two dihydrofolate reductases was assayed in the presence of these crowders and similar weak interactions were found. We have proposed a model of preferential interactions between osmolytes and crowders with DHF and found these interactions were consistent for ligand binding to three different protein scaffolds (Figure 1.5). While we have worked with homogeneous samples of crowders and osmolytes, we have not studied the effects heterogeneous mixtures of proteins on ligand association. This work aimed to address this problem. We used a stepwise approach employing fluorine NMR as a primary tool to monitor ligand binding as this technique offers the advantage of selectively labeling R67 DHFR, and second the fluorine atom is almost absent in biological samples. The absence of fluorine from natural amino acids simplifies our NMR spectra as the labeled protein of interest can be easily distinguished from the crowded environment of the cell. First, we sought to determine whether fluorine incorporation has an effect on the structure and function of R67 DHFR. Next, we determined in which position in the indole ring was fluorine substitution the best candidate to differentiate binding of the cofactor from the substrate. Finally, we probed the feasibility of detecting in-cell NMR signals from R67 DHFR.

Effects of fluorine incorporation on protein structure and function

Analysis of the literature suggests caution about generalizing the effects of fluorine incorporation and little attention has been given to the effects on protein function. Likely, the reason behind this effect being neglected arises from the several early studies that showed proteins were functionally unperturbed by fluorine incorporation. However, more recently others have noticed mixed effects of fluorine incorporation on protein function. In order to monitor ligand binding by fluorine NMR, we sought to determine first whether fluorine incorporation had an effect on the structure and function of R67 DHFR. Several properties of the fluorine atom, such as
its similar van der Wall radius to that of hydrogen, makes this substitution likely to be inconsequential to R67 DHFR function. On the other hand, some features, such as its electronegativity, are very different from that of hydrogen and can affect residue polarity and the types of interactions the residue can form. Fluorine incorporation does not affect the secondary structure. Three of the fluoroindole probes (4, 5, and 6) did not alter the function of R67 DHFR. However, steady state kinetics, along with ITC, suggested that fluorine at position 7 weakens the binding of both the cofactor and substrate, while not altering the $k_{cat}$. These results are important to advance our goal of understanding the potential secondary effects from fluorine incorporation and fluorine position in the indole ring as highlighted by the differing effects on R67 function by fluoroindoles.

**Effects on stability**

Except for 7F, fluorine incorporation did not have significant effects of the structure and function of R67 DHFR. Therefore, we determined if there were any effects on stability, especially as both tryptophans are present at protein-protein interfaces in the R67 DHFR tetramer. Using fluorine incorporation to monitor ligand binding possesses several advantages such as simple NMR assignment, easy interpretation of the NMR spectra, and minimization of possible effects on the structure and function of the protein target. Our studies revealed for R67 DHFR that fluorine incorporation does have an effect on the thermal unfolding of the protein. We found that while 4F decreased the $T_m$, 5F and 6F increased them, and 7F only had a minimal effect. Our results for R67 DHFR suggest that there is a not clear trend yet established on what fluorine incorporation can do to protein function and stability. Our results concur with other work that fluorine incorporation could increase, decrease or not alter protein stability in different proteins or even in the same protein. For example, incorporation at position 4F into staphylococcal nuclease is slightly stabilizing, while positions 5F and 6F do not alter the thermal stability. Likewise, for the protein annexin, positions 4F and 6F decreases the stability while 5F increased it. Opposite effects have also been reported for proteins with similar molecular sizes. For example, incorporation of 5-fluorotryptophan at the center of the hydrophobic core of GB1 did not affect the thermodynamic stability while incorporation of 5-fluoreleucine in the interior of ubiquitin exhibited a decrease in thermal stability of $8 ^{\circ}C$. Thus, our results suggest that the effects of fluorine incorporation on the stability of a protein seems to be position and protein dependent and needs to be considered for each protein to be studied by $^{19}$F NMR. A larger data set with proteins of different structures and properties are needed to be able to potentially predict fluoroindole effects on protein stability a priori. Our thermal unfolding studies get us closer to this goal.

**Ligand binding studies with $^{19}$F labeled R67 DHFR**

In addition to studying the effects of fluorine incorporation of R67 DHFR, we sought to determine which position of the fluorine probe would allow us to differentiate binding of the cofactor from the substrate. This problem as arises due to the active site of R67 DHFR being composed of four monomers where both cofactor and substrate bind to symmetry-related residues unlike other proteins which contain specific binding pockets for each ligand.
We hypothesized that cofactor and substrate binding might elicit similar changes in the NMR spectra as they are interacting with the same residues in the different protomers. The ligands binding on opposite sides of the active site pore to symmetry-related residues has the potential to hamper NMR detection. Furthermore, NMR spectra became too complex upon substrate binding for analysis due to breaking the symmetry in previous $^{15}$N NMR studies with truncated R67 DHFR. In addition, our initial attempts to monitor ligand binding with full length R67 DHFR by $^{15}$N labeling exhibited severe line broadening due to the presence of the disordered N-terminus (Figure 4.1). Thus, we took advantage of the simplicity of the fluorine NMR spectra.

Fluorine labeling simplified the NMR spectra by selectively targeting residues in comparison to labeling the entire protein backbone with $^{15}$N. For R67 DHFR, we sought to label its two tryptophan residues, W38 and W45. Previous studies reported that W38 is sensitive to monitor binding of the cofactor while W45 is not, which further suggested that the NMR spectra were likely to be simplified. Our results exhibited the appearance of two different $^{19}$F peaks, a sharp and a broad peak assigned to W38 and W45, respectively (Figure 3.1). Incorporation of the fluorine probe at four different positions of the indole ring resulted in different NMR spectra for each fluorine labeled protein. Cofactor binding broadened the W38 peak, while W45 remained unchanged for all protein. Upon binding of DHF, W38 split into new signals, while W45 remain again unchanged. Thus, these results are in agreement with previous fluorescence data that showed W38 is sensitive to ligand binding. The fact that W45 is not affected upon binding can be explained by its distance from the active site. Foremost, our results indicate that 4F and 5F are the best probes to differentiate ligand binding in R67 DHFR as they offer the best peak resolution as opposed to 6F and 7F in which resonances overlap due to line broadening. Thus, our results suggest that fluorine incorporation into R67 DHFR does simplify the NMR spectra, and monitoring ligand binding can be measured by the change in a single resonance. These results not only overcome the problems faced with the complexity of the HSQC spectra, but provide the basis to study R67 DHFR in context of the myriad of proteins, and other macromolecules, inside the cell.

Towards quantifying ligand binding inside the cell by $^{19}$F NMR

The development of new NMR approaches to monitor molecules in their native environment offers a venue to bridge the gap between in vitro and in vivo experiments. Our lab has studied enzyme kinetics of R67 DHFR in the test tube for decades. Interestingly, while probing the role of water, we found that addition of osmolytes weaken binding of the substrate, DHF. We proposed a model of preferential interaction, which suggests that small molecules such as osmolytes, amino acids and sugars could interact with the surfaces the substrate of R67 DHFR (Figure 1.3). We have tested this model and found similar results with three different structural scaffolds of DHFR. Furthermore, as several of the functional group present on osmolytes are also present on the surface of proteins (Figure 1.4), we have tested the effects of adding different protein crowders and found, again, weaker binding of DHF. Additionally, cells expressing R67 DHFR at low levels stopped growing at lower osmotic stress levels (lower medium osmolality) compared to the parent strain indicating that there is a potential for weak interactions between osmolytes and ligands in vivo to impact enzyme activity, and thus cell viability. These results made us wonder...
Figure 4.1. HSQC NMR spectrum for pRSETb R67 DHFR was complicated by the disordered, intact N-termini. Some residues that have previously been assigned\textsuperscript{23} are shown in the spectrum.
how enzymes work inside the crowded environment of the cell. Studying ligand binding in vivo is not a trivial problem, today few studies have reported ligand binding inside the cell. Typical methods to measure ligand binding in vivo introduce a fluorescently- or isotope-labeled proteins to distinguish the protein of interest among all the other macromolecules.

Thus, we have turned to fluorine NMR to take advantage of the simplicity that NMR spectra offer. We started our experiments by selectively labeling the tryptophan of R67 DHFR and checking for in-cell NMR signals. Unfortunately, we were not able to detect signals inside the cell due to severe line broadening (Figure 4.2). These results are not surprising as many other groups have reported severe line broadening from in-cell NMR studies, mainly attributed to the presence of weak interactions between the protein and intracellular components which hampers rotation of the protein. Next, we look for signal from the supernatant but no peak were observed. It is reported that protein overexpression in E. coli exceeding 20% of the total cellular protein could result in protein leakage. This leakage may cause the apparent signal expected from inside the cell be the product of an artifact and instead corresponds to signal from protein in the liquid surrounding the cells. The more dominant signal from the leaked protein is attributed to the faster tumbling of the protein in diluted conditions relative to the in-cell. Therefore, as a control, the supernatant was subject to NMR analysis, however no signals were detected. Subsequently, measurements of signals in the cell lysate exhibited a weak, yet still broad, signal was observed. Several groups have reported that the presence of positively charged residues on the surface of protein hampers in-cell NMR detection due to the interaction of the test protein with cytosolic components affecting the mobility of the proteins. Therefore, we rationalized that removal of the his tag from the N-terminus to decrease the number of positive charges in R67 DHFR would improve the quality of the in-cell NMR signal. Again in-cell NMR signals were broad and labeled protein inside the cell could not be easily identified (Figure 4.3). Our results represent an attempt to observed in-cell NMR from a protein with a molecular weight of 35 kDa as opposed to the average 10 kDa molecular weight of proteins normally studied by 19F NMR, such as GB1 or ubiquitin. While the in-cell NMR field has advanced to detect signals from a few additional proteins, more research needs to be done to overcome the problems faced with R67 DHFR to enable successful analysis of more proteins by in-cell NMR.

Future directions

Although the ultimate goal of this project was to monitor ligand binding in the crowded environment of the cell, our attempts to detect in-cell NMR signals of R67 DHFR failed, likely due to the presence of weak transient interactions between our test protein and the surface of cytosolic proteins which resulted in severe line broadening. While our results represent a first attempt toward in-cell detection of our protein, we envision several strategies that could be employed help to achieve this goal for R67 DHFR and other proteins. A common strategy would involve the reversal of positively charge residues on the surface of R67 DHFR. Several groups have reported the use of this strategy to improve in-cell NMR signals. An average of 80% of the most abundant proteins in E. coli are negatively charged. Therefore, positively charged residues on the surface of R67 DHFR could be interacting with the these negatively charged cytosolic proteins. This would decrease the rotation of the 19F-labeled protein, which ultimately
Figure 4.2 A representative in-cell NMR spectra of pRSETb 5F R67 DHFR. A) 5F R67 DHFR in buffer, B) cell lysates, C) supernatant and D) inside of *E. coli* cells.
Figure 4.3 A representative in-cell NMR spectra of normal length 5F R67 DHFR. 5F R67 DHFR in A) cell lysate, B) inside of cells and C) without addition of IPTG.
hampers in-cell NMR detection. Thus, we envision that R29A and R31A or R29E and R31E mutants might improve the sharpness of the peaks in the in-cell NMR of R67 DHFR. These mutants offer the advantage that the mutations do not affect enzyme activity.\textsuperscript{40-42} A successful example of this reversal of charge strategy used in human cell lines was reported on the human protein profilin 1, where in-cell NMR signals were recover after a careful analysis of surface residues and the introduction of mutations.\textsuperscript{38}

Another relatively easy way to improve NMR detection is to increase the sample size. Because NMR detection depends the number of nuclei sampled in the detection region, increasing the path length of the NMR tube could be used as simple strategy.\textsuperscript{43} Currently, we employ a 5 mm diameter NMR tube, but there are also 10 mm tubes available. Using larger tubes not only allow us to use a larger sample volume (4 mL vs 0.5 mL) but also facilitates the addition of the viscous cell slurry samples to the NMR tube for our in-cell NMR experiments. Unfortunately, one of the biggest concerns of this approach is that the sample preparation remains a long process that includes several stressful conditions for the cells such as centrifugation, changes in buffer conditions, exposure to different temperatures and packing of the cells, and reduction of oxygen availability (hypoxia and medium acidification) during the experiment which could results in diminishing cell viability.\textsuperscript{44} To overcome these problem several groups have incorporated the use of bioreactors.\textsuperscript{31, 45, 46}

The development of new bioreactors in the field of in-cell NMR could be incorporated in our research. A recurrent challenge of in-cell NMR is cell viability.\textsuperscript{31, 47} Currently, in-cell NMR experiments are limited by the short lifetime of the cells and length of experiments, which can last several hours. The new development of bioreactors allows for maintaining human cell lines, keeping them alive and active for up to 72 h, and could be incorporated into our \textsuperscript{19}F NMR analysis to keep the \textit{E. coli} strains used for R67 DHFR viable for longer times.\textsuperscript{31} If it is possible to overcome all these challenges of detecting proteins by in-cell NMR, then the technique can be used for cell-based assays to develop inhibitors for R67 DHFR.

Another way to increase in-cell NMR detection employs the use of trifluoromethyl labels. The trifluoromethyl group has several advantages for the use of \textsuperscript{19}F in-cell NMR for several reasons. First, trifluoromethylation increases the sensitivity of proteins given that three fluorine atoms are present. Second, the trifluoromethyl groups, like methyl groups, exhibit narrow lines because of the free rotation of the bond between the carbon and the methylated atom. One of the most common trifluoromethyl labeled amino acids is phenylalanine, and has been used to monitor ligand binding in several small,\textsuperscript{48, 49} and larger,\textsuperscript{50} proteins. This approach not only offers the advantage of narrower signals but could help to minimize NMR acquisition time which could also help to avoid protein leakage problems. One drawback of adding this probe into a target protein it is its bulkiness which could be disruptive if place in hydrophobic interfaces. Fortunately, several aliphatic amino acid analogs such as leucine, valine, alanine methionine are also available with trifluoromethyl probes which expand the application of this strategy.\textsuperscript{9, 10} Structural and functional analysis needs to be done to avoid potential destabilizing effects upon incorporation.

One additional problem with in-cell NMR experiments involves the overexpression of the protein target for NMR biolabeling. In this approach, the end goal is to only label your target protein,
however, because the cell protein production machinery are being used to biolabel the protein target it is also possible to label off-target components of the cell which will contribute to the NMR signal. Several ways to get around this problem are addressed by different advances of protein delivery. For example, microinjection of purified protein into oocytes not only overcomes the problem of background signals, but also provides a way to work with eukaryote cells. Other techniques like the attachment of cell penetrating peptides, pore-forming toxins, and electroporation, not only expand the use of in-cell applications, but offer the biochemist different tools to optimize specific systems.

If in-cell detection continues to be a problem, other still biological, cell-like, media such as cell lysates could be used. The recent development of a reconstituted cytosol could be use an alternative approach. This reconstituted cytosol is prepared from E. coli extracts where cytosol has been extensively dialyzed to remove low-molecular-weight components and then lyophilized to obtain a powered cytosol. This reconstituted cytosol has been used to measure the effects of macromolecular crowding on the stability of the chymotrypsin inhibitor 2 with varying concentrations of reconstituted cytosol. These cell extracts are attractive because they are easier to handle and can be spiked with different labeled proteins. While the concentrations of macromolecules in E. coli has been calculated to be around 400 mg/mL, working with E. coli extracts provides a decent approach to work with more cell-like conditions where macromolecular concentrations can reach close to ~90 mg/mL. Another advantage of adding E. coli lysates is the lack of metabolites which further simplify our NMR analysis. If line broadening continues to be a problem with could still work with mixture of protein crowders and osmolytes. We have already performed in vitro binding and kinetic assays with homogenous samples of protein solutions of lysozyme, a rhamnose synthesis protein (RmlC), ovalbumin, hemoglobin or casein. One advantage of using these proteins is that can easily obtain in large quantities and protein concentrations could be modulate to tease apart potential effects.

Advances in the field of NMR such as the development of solid state NMR (ssNMR) would also expand our tools to study cell like conditions. Detection of in-cell NMR signals in solution depends largely on the molecular tumbling of the protein. However, due to the presence of weak interactions with other cytosolic component, crowding, viscosity and lack of diffusion of the cells themselves, tumbling is largely minimized resulting in poor detection. In contrast, ssNMR is not limited by molecular tumbling and has been largely used to study systems devoid of mobility, such as bacterial cell walls, membrane proteins, and proteins in large complexes. A notable example is the case of the human FK506-binding protein in which detection by solution in-cell NMR spectra, like our studies, failed due to the weak interaction with cellular components. However, using ssNMR in combination with flash freezing of cells allowed the detection of backbone chemical shifts. Similarly, ssNMR has been used to study folding pathways of intrinsically disordered regions of the yeast prion protein Sup35 using cell lysates. ssNMR is a powerful technique to study protein in their native environment and we envision its adaptability to the study of quinary interactions.

If our protein-based NMR approach fails, we could turn to ligand-based approaches. Studying small molecule binding to target protein by NMR has been mainly applied to the detection of
proton signal and transfer of NOEs. We could probe the applicability of fluorine-labeled DHF or folate to monitor in-cell NMR signals. These compounds could be directly added to the cells and, if imported, used to look for in-cell NMR signals. Several variants of these experiments could also be performed, such as using cell lysates or a mixture of protein crowders with labeled DHF/folate. If incorporation of labeled DHF/folate faces problems, we could engineer E. coli cells to facilitate import of the ligand. An example of engineering cell to facilitate transport of molecules has been reported for the production of 5′-fluorodeoxyadenosine from S-adenosyl-L-methionine (SAM) and fluoride. The authors required the introduction of a transporter for SAM and the deletion of endogenous fluoride efflux to facilitate retention of metabolites in E. coli. We could explore similar strategies such as the usage of transporter of folate and the use of E. coli strains that prevent folate efflux.

Using $^{19}$F R67 DHFR for drug discovery

Use of the broad-spectrum antibacterial drug trimethoprim (TMP) has resulted in the development of drug resistance in the form of highly transmissible plasmids such as R67 DHFR. Today, only a few non-specific inhibitors for R67 DHFR have been reported (Figure 4.4), and there is no drug currently on the market that targets R67 DHFR. However, some efforts are under way towards solving this issue. For example, the Pelletier group used a fragment-based design with the structures of congo red and novobiocin as base templates to identify several low molecular weight fragments that inhibited R67 DHFR while also displaying good selectivity against the human DHFR. After optimization of fragments for their inhibitory effects, symmetric versions of the molecule were made and elongation of the linker between the symmetric inhibitors were used to helped improve potency (Figure 4.4). These results highlighted two important features of these inhibitors. First, the presence of carboxylic acid functional groups greatly improved binding likely due to interacting with the K33 residues located at the surface of the active site pore. Secondly, the length of the inhibitors helps to create more contacts with the hydrophobic residues at the interior of the active site. Subsequent optimization of these leads resulted in the design a new inhibitor with improved potency. These results highlight the need for different venues to combat the emergence of TMP resistant.

Our fluorine NMR studies with R67 DHFR could be helpful in this regard for the following reasons. We have shown that R67 DHFR $^{19}$F NMR is sensitive enough to differentiate cofactor and the substrate and monitor binding. This could be used as an advantage for the screening of new small molecules fragments using protein-observed $^{19}$F NMR (PROF). PROF uses the advantage of the simplicity of the NMR spectra and the high sensitivity of the fluorine probe to changes in its chemical environment. PROF has been successfully applied for the discovery of small-fragments leads for the antimalarial drug target AMA1, GPCRs and bromodomains proteins. Furthermore, with the increasing generation of fluorinated fragments libraries, and the development of fluorine NMR experiments, PROF could be useful to extract structural information for the design of new drugs for R67 DHFR. An application of this method was applied in Abbott Laboratories to obtain ligand-protein NOEs for the anti-apoptotic protein BCL-xL, a drug target for anticancer therapy. In addition, probably one of the best advantages of PROF is the absence of fluorine atom inside the cell which could be used to monitor inhibitors inside the cell.
This approach has been used to monitor the binding of inhibitors to human carbonic anhydrase in cells.\textsuperscript{70} Overall, the use of fluorine NMR for the drug discovery offer several advantages such as simplicity over other NMR labeling strategies. We envision that our studies could be used to augment traditional structure determination of protein-ligand complexes for the development of new inhibitor for R67 DHFR.
Figure 4.4. Inhibitors of R67 DHFR. $K_{i}$s are from references 64 and 71.
References


bisbenzimidazole inhibitors to an emergent trimethoprim-resistant type II dihydrofolate reductase guides the design of monomeric analogues, *ACS Omega* 4, 10056-10069.


Vita

Gabriel José Fuente Gómez was born in the sunny city of La Paz, Baja California Sur, México. After high school, he attended the Instituto Tecnológico de La Paz and obtained a degree in Biochemical Engineering. Immediately after obtaining his degree, he pursued a Master’s in Science at the Centro de Investigación en Alimentación y Desarrollo A.C. in Hermosillo, Sonora, México. Upon completion of his Master’s program, he worked as a food engineer at the food processing company Sigma Alimentos in Guadalajara, Jalisco, México. Gabriel then joined the University of Tennessee to obtain a degree in Biochemistry & Cellular and Molecular Biology.