

# **Computer Modelling and Simulations of Enzymes and their Mechanisms**

**Hernán Alonso**



A thesis submitted for the degree of Doctor of Philosophy of  
the Australian National University

**September 2006**



---

## Publications from this work

---

*Alonso H, Bliznyuk AA and Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev 26:531-568 (2006).*

*Alonso H and Gready JE. Integron-sequestered dihydrofolate reductase, a recently redeployed enzyme. Trends Microbiol 14:236-242 (2006).*

*Alonso H, Gillies MB, Cummins PL, Bliznyuk AA and Gready JE. Multiple ligand-binding modes in bacterial R67 dihydrofolate reductase. J Comput Aided Mol Des 19:165-187 (2005).*

A pdf file of each publication can be found in a supplementary CD at the back of the thesis.



---

## Statement of Originality

---

The contents of this thesis are the result of my original research, which has been conducted under the principal supervision of Prof. Jill E. Gready (The John Curtin School of Medical Research, ANU).

Prof. Elizabeth E. Howell (University of Tennessee) provided the coordinates of some published docked complexes of folate and NADPH within R67 DHFR, which were used to generate a complete R67•DHFH<sup>+</sup>•NADPH structure, which I referred to as *howH* in the analysis of Chapter 3.

Dr Malcolm Gillies produced the topology files for the ligands dihydrofolate and NADPH used for the GROMACS MD simulations presented in Chapter 3.

The X-ray crystal structure of the enzyme MeTr complexed with methyl-tetrahydrofolate, used for the analysis in Chapter 4, was provided to Prof. Jill Gready in confidence by Prof. Stephen W. Ragsdale (University of Nebraska).

I declare that the work presented in this thesis is, to my belief, original, except as acknowledged in the text above, and that the material has not been submitted, either in whole or in part, for a degree at this or any other university.

Hernán Alonso

September 2006



---

## Acknowledgements

---

This PhD thesis would have not been possible without the invaluable supervision of Prof Jill Gready. I am grateful for her honest and thorough guidance, constructive advice and constant support. Her enthusiasm for science and inquisitive nature have set a clear example for me to follow. I hope there will be many opportunities in the future to further work and learn from her.

My special thanks to Dr Peter Cummins and Dr Andrey Bliznyuk who have guided me during these three and a half years, and have provided me with helpful and essential comments on my work, while helping me to become a computational biologist. I am particularly thankful to Dr Malcolm Gillis, who acted as my nanny during the first months of my thesis, showing me how to deal with scientific problems and tackle them with computers.

I would like to express my gratitude to all those people who facilitated my work by sharing their results with me: to Prof Elizabeth Howell (University of Tennessee), for providing the coordinates of R67 DHFR complexes and disclosing unpublished experimental results; and to Prof Stephen Ragsdale and Dr Tzanko Doukov (University of Nebraska), for their unpublished X-ray structure of MeTr complexed with CH<sub>3</sub>THF.

I would also like to acknowledge the benefit of the valuable discussions with Prof Ruth Hall (University of Sydney) regarding integrons and gene cassettes, as well as Prof Rowena Matthews (University of Michigan) for her constructive advice on experimental results for methionine synthase, and to Dr Ivan Rostov (ANU Supercomputer Facility), for his help and advice with calculations in general and the ONIOM method in particular.

From a personal point of view, all these years in Australia would not have been the same without the invaluable friendship of Alex Zelensky, and those hundreds of lunches shared with the coolest Indians ever, Babu Kannappan and Nagesh Chakka.

My gratitude extends to all the members of the Computational Proteomics and Therapy Design Group, for creating a great work environment where I have been able to grow both as a scientist and as a person over the past several years.

As a proud member of the Fenner Hall community, I am sincerely grateful to all those superb people that form this little world. Fenner Hall is a place where students from all over the world live together in a friendly and supportive environment, and a place that I have grown up to call my home.

Finally, I would like to thank my family, especially my Mum, my sister Lucía, my grandma Chola and my grandpa Coco, for their unconditional support and love.

This thesis is dedicated to the courageous reader, who dares to go through hundreds of pages in search of some (hopefully) valuable information...

Hernán Alonso,

September 2006.



---

## Abstract

---

Although the tremendous catalytic power of enzymes is widely recognized, their exact mechanisms of action are still a source of debate. In order to elucidate the origin of their power, it is necessary to look at individual residues and atoms, and establish their contribution to ligand binding, activation, and reaction. Given the present limitations of experimental techniques, only computational tools allow for such detailed analysis. During my PhD studies I have applied a variety of computational methods, reviewed in Chapter 2, to the study of two enzymes: DfrB dihydrofolate reductase (DHFR) and methyltetrahydrofolate: corrinoid/iron-sulfur protein methyltransferase (MeTr).

The DfrB enzyme has intrigued microbiologists since it was discovered thirty years ago, because of its simple structure, enzymatic inefficiency, and its insensitivity to trimethoprim. This bacterial enzyme shows neither structural nor sequence similarity with its chromosomal counterpart, despite both catalysing the reduction of dihydrofolate (DHF) using NADPH as a cofactor. As numerous attempts to obtain experimental structures of an enzyme ternary complex have been unsuccessful, I combined docking studies and molecular dynamics simulations to produce a reliable model of the reactive DfrB•DHF•NADPH complex. These results, combined with published empirical data, showed that multiple binding modes of the ligands are possible within DfrB.

Comprehensive sequence and structural analysis provided further insight into the DfrB family. The presence of the *dfrB* genes within integrons and their level of sequence conservation suggest that they are old structures that had been diverging well before the introduction of trimethoprim. Each monomer of the tetrameric active enzyme presents an SH3-fold domain; this is a eukaryotic auxiliary domain never found before as the sole domain of a protein, let alone as the catalytic one. Overall, DfrB DHFR seems to be a poorly adapted catalyst, a ‘minimalistic’ enzyme that promotes the reaction by facilitating the approach of the ligands rather than by using specific catalytic residues.

MeTr initiates the Wood-Ljungdahl pathway of anaerobic CO<sub>2</sub> fixation. It catalyses the transfer of the N5-methyl group from N5-methyltetrahydrofolate (CH<sub>3</sub>THF) to the cobalt centre of a corrinoid/iron-sulfur protein. For the reaction to occur, the N5 position of CH<sub>3</sub>THF is expected to be activated by protonation. As experimental studies have led to conflicting suggestions, computational approaches were used to address the activation mechanism.

Initially, I tested the accuracy of quantum mechanical (QM) methods to predict protonation positions and p*K<sub>a</sub>s* of pterin, folate, and their analogues. Then, different protonation states of CH<sub>3</sub>THF and active-site aspartic residues were analysed. Fragment QM calculations suggested that the p*K<sub>a</sub>* of N5 in CH<sub>3</sub>THF is likely to increase upon protein binding. Further, ONIOM calculations which accounted for the complete protein structure indicated that active-site aspartic residues are likely to be protonated before the ligand. Finally, solvation and binding free energies of several protonated forms of CH<sub>3</sub>THF were compared using the thermodynamic integration approach. Taken together, these preliminary results suggest that further work with particular emphasis on the protonation state of active-site aspartic residues is needed in order to elucidate the protonation and activation mechanism of CH<sub>3</sub>THF within MeTr.

---

# Table of Contents

---

<b>Publications from this work .....</b>	<b>i</b>
<b>Statement of Originality.....</b>	<b>iii</b>
<b>Acknowledgements.....</b>	<b>v</b>
<b>Abstract.....</b>	<b>vii</b>
<b>Table of Contents.....</b>	<b>ix</b>
<b>List of Figures.....</b>	<b>xiii</b>
<b>List of Tables.....</b>	<b>xvi</b>
<b>List of Abbreviations.....</b>	<b>xvii</b>
<b>CHAPTER 1. INTRODUCTION.....</b>	<b>1</b>
1.1 ENZYMES.....	1
1.2 ENZYMATIC CATALYSIS .....	2
1.2.1 Substrate binding.....	3
1.2.2 Enzyme Kinetics .....	6
1.2.3 Catalytic efficiency .....	7
1.3 ORIGIN OF THE CATALYTIC POWER OF ENZYMES.....	8
1.3.1 Desolvation Hypothesis .....	9
1.3.2 Electrostatic Effects .....	10
1.3.3 Entropic Effects.....	11
1.3.4 Strain Mechanism.....	12
1.3.5 Reactive Near Attack Conformers (NACs).....	13
1.3.6 Orbital Steering Mechanism.....	13
1.3.7 Low-Barrier Hydrogen Bonds.....	14
1.3.8 Dynamic Effects.....	15
1.3.8.2 Quantum Effects.....	18
1.3.8.3 Covalent Hypothesis .....	19
1.4 GOALS OF THE THESIS .....	20
<b>CHAPTER 2. COMPUTATIONAL APPROACHES FOR THE STUDY OF ENZYMES AND THEIR REACTIONS.....</b>	<b>23</b>
2.1 DOCKING.....	23
2.1.1 Programs and Algorithms.....	25
2.1.2 Ligand Flexibility.....	26
2.1.2.1 Fragment-Based Methods or Incremental Construction Algorithms.....	27
2.1.2.2 Monte Carlo Methods .....	28
2.1.2.3 Evolutionary Algorithms .....	28
2.1.2.4 Pregenerated Conformational Libraries .....	29

2.1.2.5	Implicit Ligand Flexibility .....	29
2.1.3	Protein Receptor Representation and Flexibility .....	30
2.1.3.1	Soft Docking .....	31
2.1.3.2	Sidechain Flexibility .....	32
2.1.3.3	Combined Protein Grid .....	32
2.1.3.4	United Description of the Receptor .....	33
2.1.3.5	Docking into Several Individual Protein Conformations .....	34
2.1.4	Scoring Functions .....	35
2.1.4.1	First Principles Methods .....	36
2.1.4.2	Semiempirical Methods .....	36
2.1.4.3	Empirical Methods .....	37
2.1.4.4	Knowledge-Based Potentials .....	37
2.1.5	Refinement of Docked Complexes .....	38
2.2	MD SIMULATIONS .....	39
2.2.1	Programs and algorithms .....	40
2.2.1.1	Treatment of solvation .....	41
2.2.1.2	Boundary conditions .....	42
2.2.1.3	Long-Range Interactions .....	43
2.2.1.4	Temperature and Pressure Control .....	44
2.2.2	Limitations and sampling improvement .....	45
2.2.2.1	Modified Potentials .....	45
2.2.2.2	Modified Sampling .....	46
2.2.2.3	Modified Dynamics .....	46
2.2.3	MD Simulations and Enzyme Flexibility .....	47
2.3	QUANTUM MECHANICAL METHODS .....	48
2.3.1	<i>Ab Initio</i> Calculations .....	49
2.3.2	Semi-Empirical Calculations .....	50
2.3.3	Density Functional Theory Methods .....	51
2.3.4	Linear-scaling QM methods .....	52
2.4	HYBRID POTENTIAL METHODS .....	53
2.5	FREE ENERGY CALCULATIONS .....	55
2.5.1.1	Free Energy Perturbation and Thermodynamic Integration .....	56
2.5.1.2	Linear Interaction Energy Method .....	57
2.5.1.3	Molecular Mechanics/Poisson-Boltzmann Surface Area Method .....	59
<b>CHAPTER 3. DFRB DIHYDROFOLATE REDUCTASE .....</b>		<b>61</b>
3.1	RESISTANCE TO TRIMETHOPRIM .....	62
3.2	R67 VS CHROMOSOMAL DHFR .....	64
3.3	MOTIVATIONS AND GOALS .....	67
3.4	GENETIC FRAMEWORK AND SEQUENCE ANALYSIS .....	68
3.4.1	Methods .....	68
3.4.2	Genetic Framework: Integrons and Gene Cassettes .....	69
3.4.3	Sequence analysis: the DfrB Cassettes .....	72

3.5	STRUCTURAL ANALYSIS OF THE DFRB PROTEINS.....	77
3.5.1	Methods.....	77
3.5.2	SH3 Domain.....	77
3.5.3	Structural Homologues.....	80
3.5.4	DfrB and HIV-1 integrase.....	81
3.6	COMPILATION OF MUTATIONAL ANALYSES .....	83
3.7	R67 DHFR AND LIGAND BINDING .....	85
3.7.1	Methods.....	88
3.7.1.1	Docking .....	88
3.7.1.2	MD simulations .....	94
3.7.2	Results.....	95
3.7.2.1	Docking .....	95
3.7.2.2	MD simulations .....	102
3.7.3	Discussion .....	110
3.7.3.1	Interligand interactions .....	110
3.7.3.2	Protein – ligands interactions.....	112
3.7.3.3	Multiple conformations.....	114
3.7.3.4	Role of R67 DHFR in the catalysis of the reaction.....	116
3.7.4	Conclusions.....	116
3.8	FINAL CONSIDERATIONS .....	117

#### **CHAPTER 4. METHYLTETRAHYDROFOLATE:CORRINOID/IRON-SULFUR**

<b>PROTEIN METHYLTRANSFERASE (METR) .....</b>	<b>121</b>	
4.1.1	Structure.....	123
4.1.2	Reaction .....	125
4.2	COBALAMIN-DEPENDENT METHIONINE SYNTHASE (METH).....	128
4.2.1	Protonation and activation of CH <sub>3</sub> THF within Meth.....	129
4.2.2	Role of aspartic residues within the Meth active site .....	130
4.3	COMPARISONS OF METR AND METH.....	130
4.4	MOTIVATIONS AND GOALS.....	133
4.5	METHODS .....	134
4.5.1	Theoretical calculations of protonation energies.....	134
4.5.2	QM fragment calculations.....	136
4.5.3	ONIOM calculations .....	137
4.5.3.1	Preparation of initial structures.....	138
4.5.3.2	ONIOM Energy Calculations.....	139
4.5.4	Free energy calculations.....	140
4.5.4.1	Preparation of initial structures.....	143
4.5.4.2	Transformations <i>in vacuo</i> .....	143
4.5.4.3	Transformations in water .....	144
4.5.4.4	Transformations inside the protein.....	144
4.6	RESULTS AND DISCUSSION .....	145

4.6.1	Protonation Patterns and $pK_a$ of folates and derivatives.....	145
4.6.2	Influence of the protein environment on the protonation pattern and energy of CH <sub>3</sub> THF .....	148
4.6.2.1	Fragment QM calculations.....	148
4.6.2.2	ONIOM calculations.....	152
4.6.2.3	Free Energy Calculations Using Thermodynamic Integration .....	155
4.7	CONCLUSIONS.....	162
	<b>SUMMARY AND FINAL REMARKS .....</b>	<b>165</b>
	<b>REFERENCES.....</b>	<b>169</b>

---

## List of Figures

---

Figure 1-1 Protein flexibility.....	4
Figure 1-2 Dynamical recrossing of the transition state.....	7
Figure 1-3 Schematic representation of a catalysed reaction ( $\Delta G^{\ddagger}_{\text{cat}}$ ) and its uncatalysed counterpart ( $\Delta G^{\ddagger}_{\text{non}}$ ). ....	8
Figure 1-4 Desolvation hypothesis.....	9
Figure 1-5 Electrostatic effects.....	10
Figure 1-6 Entropic effects.....	11
Figure 1-7 Strain mechanism.....	12
Figure 1-8 Reactive Near Attack Conformers.....	13
Figure 1-9 Dynamic effects.....	15
Figure 1-10 Quantum effects.....	18
Figure 2-1 Drug design process.....	24
Figure 2-2 Docking with an incremental construction algorithm. ....	27
Figure 2-3 Protein flexibility.....	31
Figure 2-4 Soft docking.....	31
Figure 2-5 United description of the protein receptor. ....	34
Figure 2-6 Differential ligand binding. ....	48
Figure 2-7 Partitioning of the protein system for the application of hybrid potential methods.....	53
Figure 3-1 Crystal structure of <i>E. coli</i> DHFR (PDB code 1RA2) (A) and R67 DHFR (PDB code 1VIF) (B).....	65
Figure 3-2 Stereochemistry of the reaction. ....	66
Figure 3-3 Schematic representation of the structure of an integron and mobile gene cassettes. ....	70

Figure 3-4 Sequence alignment of the six known DfrB proteins.....	73
Figure 3-5 Phylogenetic tree of the <i>dfrB</i> gene family.....	74
Figure 3-6 <i>attC</i> recombination site. ....	75
Figure 3-7 Nucleotide variability along the <i>dfrB</i> cassette. ....	76
Figure 3-8 Structure of R67 DHFR (1VIE). ....	78
Figure 3-9 Monomers association in R67 DHFR. ....	79
Figure 3-10 Superposition of SH3 domain structural homologues. ....	81
Figure 3-11 R67 DHFR and HIV-1 integrase.....	82
Figure 3-12 Compilation of mutations observed for active DfrB enzymes. ....	83
Figure 3-13 Schematic representation of the reactants DHF and NADPH and the interactions they establish with the enzyme R67 DHFR as observed experimentally.....	86
Figure 3-14 Docking sequence performed with AutoDock [166] (A) and FlexX [587] (B).....	90
Figure 3-15 Relative positioning of ligands in the final docked structures. ....	97
Figure 3-16 Scoring analysis of docked structures. ....	101
Figure 3-17 Ligand mobility within the active site pore.....	103
Figure 3-18 MD simulations of selected complexes.....	105
Figure 3-19 Evolution of H-bond interactions along an MD trajectory. ....	108
Figure 3-20 Water distribution within the active-site pore.....	111
Figure 4-1 Crystal structure of MeTr.....	122
Figure 4-2 Disposition of the ligand methyltetrahydrofolate (CH <sub>3</sub> THF) within the active site of MeTr. ....	124
Figure 4-3 Structural comparison of MeTr and the CH <sub>3</sub> THF-binding domain of MetH. ....	131
Figure 4-4 Schematic representation of the MeTr active site.....	132
Figure 4-5 Representation of the active site and ligand used during fragment QM calculations. ....	136



Figure 4-6 QM region of the MeTr•CH <sub>3</sub> THF system used during ONIOM calculations. ....	140
Figure 4-7 N5-protonated form of CH <sub>3</sub> THF used in alchemic transformations during free energy calculations. ....	142
Figure 4-8 Mutational sequence followed during the calculation of relative free energies.....	142
Figure 4-9 Protonation patterns of pterin and derivatives.....	146
Figure 4-10 Correlation of theoretical protonation energies (HF/6-31+G** PCM) and experimental p <i>K</i> <sub>a</sub> values. ....	147
Figure 4-11 Relative energies of single- (A) and double-protonated (B) forms of a simplified representation of a MeTr•CH <sub>3</sub> THF complex.....	150
Figure 4-12 Estimated p <i>K</i> <sub>a</sub> values within the protein environment. ....	151
Figure 4-13 ONIOM energies of different protonated forms of the MeTr•CH <sub>3</sub> THF complex. ....	153
Figure 4-14 ONIOM energies of MeTr•CH <sub>3</sub> THF complexes (larger QM region). ....	155
Figure 4-15 Calculation of relative energies of solvation. ....	156
Figure 4-16 Relative solvation free energies.....	157
Figure 4-17 Free energy of solvation and interaction energies. ....	158
Figure 4-18. Thermodynamic cycles used to determine relative protein-solvation energies ( $\Delta\Delta G_{\text{prot-solv}}$ ) (A) and relative binding energies ( $\Delta\Delta G_{\text{binding}}$ ) (B). ....	159
Figure 4-19 Free energy of solvation in water and within the protein environment.....	160
Figure 4-20 Relative binding free energies. ....	161

---

## List of Tables

---

Table 3-1 Collection of <i>dfrB</i> genes reported to the GenBank. ....	63
Table 3-2 General properties of <i>E. coli</i> DHFR and DfrB DHFR. ....	64
Table 3-3 Experimental observations pertinent to the conformations adopted by the ligands DHF/folate and NADPH/NADP <sup>+</sup> within the active site of R67 DHFR. ....	87
Table 4-1. Experimental observations regarding the protonation and activation of CH <sub>3</sub> THF within the active site of MeTr and MetH. ....	133
Table 4-2. Simulation conditions used during the thermodynamic integration calculations <i>in vacuo</i> , water, and the protein environment. ....	144
Table 4-3. Compilation of pterin-derived compounds used for the QM prediction of p <i>K<sub>a</sub></i> . ....	145

---

## List of Abbreviations

---

B3LYP	Becke 3-Parameter (Exchange), Lee, Yang and Parr (DFT)
C/Fe-SP	Corrinoid/iron-sulfur protein
CH <sub>3</sub> THF	N5-methyl-5,6,7,8-tetrahydrofolate
DFT	Density functional theory
DHF	7,8-dihydrofolate
DHFR	Dihydrofolate reductase
DHPS	Dihydropteroate synthase
FEP	Free energy perturbation
GA	Genetic algorithm
HF	Hartree Fock theory
LBHB	Low-barrier hydrogen bond
LGA	Lamarckian genetic algorithm
LIE	Linear integration energy
MC	Monte Carlo
MD	Molecular dynamics
MetH	Cobalamin-dependent methionine synthase
MetH <sup>(2-649)</sup>	Homocystein- and CH <sub>3</sub> THF-binding domains of MetH
MeTr	Methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase
MM	Molecular mechanics
MM-PBSA	Molecular Mechanics/Poisson-Boltzmann Surface Area
MP2	Second-order Moller-Plesset perturbation theory
NAC	Near attack conformers
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form
NADP <sup>+</sup>	Nicotinamide adenine dinucleotide phosphate, oxidized form
NMR	Nuclear magnetic resonance
ONIOM	Our own N-Layered Integrated molecular Orbital and molecular Mechanics
ORF	Open reading frame
PBC	Periodic boundary conditions

PCM	Polarisable continuum model
PDB	Protein data bank
PME	Particle mesh Ewald
PMF	Potential of mean force
QM	Quantum mechanics
REMD	Replica exchange molecular dynamics
RESP	Restrained electrostatic potential
RMSD	Root mean square deviation
THF	5,6,7,8-tetrahydrofolate
TI	Thermodynamic integration
TMP	Trimethoprim
TS	Transition state
vdW	van der Waals