

Structural Studies of Baeyer-Villiger Monooxygenases

Brahm Jonathan Yachnin

Doctor of Philosophy

Department of Biochemistry

McGill University

Montreal, Quebec, Canada

Dec. 09, 2013

A thesis submitted to McGill University in partial fulfillment of the requirements of the
degree of Doctor of Philosophy

©Copyright 2013 All rights reserved.

ACKNOWLEDGMENTS

The work presented in this thesis would not have been possible without the contributions of many people. First and foremost, I would like to thank my supervisor, Albert Berghuis, for giving me the opportunity to work in his lab. His unconditional support has made my time here easy, both during the good times and the more difficult times, both personally and professionally. His guidance, and the working environment that he fosters in the lab, permeates every page of this thesis. I would like to also thank all of the past and present members of the Berghuis lab, for their help, friendship, and countless Coffee Times. In particular, I would like to thank Ahmad Mirza, whose guidance when I first started in the lab helped me find my way both with respect to my own project and lab work in general. Without Magdalena Korczynska, it is entirely possible that I would still be trying to get my first crystal. I would also like to thank Desiree Fong, Dave Burk, Wayne Miller, Oliver Baettig, Jonathan Blanchet, Jae Park, and Nozhat Safaee for helping me with countless favours, big and small, including reading parts of this thesis, as well as well as being incredible friends over the years. I would also be remiss not to make special mention of Dmitry Rodionov, who seems to single-handedly keep the entire fourth floor running, is one of the best sounding boards anyone could ask for, and an incredibly generous friend (not to mention the only one I know of powered by a butane-potato-ethanol mixture). It often seems impossible for all of us together to have been as much help as he is to any one of us. I also had the privilege of working with five excellent project students who helped move this project forward. Vicky Cheng, Jason Ivall, Roderick MacCuish, Michelle McEvoy, and Jude Cheng helped me immensely with protein purification, crystallization, SAXS, and enzyme kinetics. Their hard work and enthusiasm made my life much easier than it would have been otherwise.

Outside of the lab, I would like to thank the members of my RAC, Dr. Robert MacKenzie, Dr. Nicolas Moitessier, and Dr. John Silvius, for their help and advice over the years. I would also like to recognize my collaborators in the Lau and Pelletier labs. I'd like to thank Dr. Peter Lau, who has been a valuable

resource on all things BVMO. His enthusiasm and support for this project is always appreciated. From his lab, I would also like to thank in particular Stephane Grosse and Krista Morley for their help. I have also had the chance to collaborate with Dr. Joelle Pelletier's group. This has led to the two publications that are the appendices of this thesis, and hopefully some more to come. I would like to thank Jordan Volpato, Jonathan Blanchet (again), Lucie Poulin, Damien Collin, Maximilian Ebert, and Jacynthe Toulouse for their help on these projects. I would also like to thank Dr. Pelletier for her generous help whenever I needed advice on an enzyme kinetics problem.

The NMR data would never have been collected without the help of Tara Sprules and Sameer Al-Abdul-Wahid at QANUC, and I would also like to thank Dr. Anthony Mittermaier for his valuable assistance in designing and interpreting the CPMG NMR experiments in Chapter 4. I would like to also recognize Shaun Labiuk at CLS for his help with crystallography data collection, and Kevin Dyer at ALS for his help with SAXS data collection.

Of course, I would also like to thank the agencies that funded both my stipend and this work: NSERC, CIHR (chemical biology), and McGill University.

Finally, I would like to thank my parents, Michael and Lina, my brothers, Daniel, Aaron, and David, my grandparents, and the army of aunts, uncles, and cousins for their unending and unconditional love and support throughout my life. While their efforts to make sure that I don't starve to death are always appreciated, far more valuable is the knowledge that whatever happens, I have them to lean on. I could not have done this without it.

I remember my uncle once saying that going through life is like walking a tightrope. It is easy to fall off, but fortunately there is not a long way to fall, and there are lots of people to help you climb back on. I have been extremely privileged to have so many people to help me climb back on that rope.

AUTHOR CONTRIBUTIONS

The text and images in this thesis have been adapted from the following manuscripts:

Chapter 2:

Yachnin, BJ, Sprules, T, McEvoy, MB, Lau, PCK, Berghuis, AM. (2012) The Substrate-Bound Crystal Structure of a Baeyer-Villiger Monooxygenase Exhibits a Criegee-like Conformation. *J. Am. Chem. Soc.* **134**: 7788-7795.

Author contributions: BJY: Subcloning of the *chnB1* gene, mutagenesis, protein expression and purification, crystallization, data collection, structure solution, refinement, and analysis; BJY and TS: NMR data collection and analysis; BJY and MBM: protein expression and purification for kinetic studies, and enzyme kinetics data collection and analysis; BJY, PCKL, AMB: experimental design and manuscript preparation.

Copyright © 2012 American Chemical Society.

Chapter 3:

Yachnin, BJ, McEvoy, MB, MacCuish, R, Morley, K, Lau, PCK, Berghuis, AM. Lactone-bound structures of cyclohexanone monooxygenase reveal the basis for substrate selection. **Manuscript in preparation.**

Author contributions: BJY: Crystal structure data collection, structure solution, refinement, and analysis, substrate profiling bioconversions; BJY, MBM, RM: protein expression, purification, and crystallization; BJY, KM: gas chromatography data collection and analysis; BJY, PCKL, AMB: experimental design and manuscript preparation.

Chapter 4:

Yachnin, BJ, Lau, PCK, Mittermaier, AK, Berghuis, AM. Baeyer-Villiger monooxygenase loop dynamics examined by SAXS and NMR. **Manuscript in preparation.**

Author contributions: BJY: Subcloning of the *cpdB* gene, protein expression and purification, SAXS data collection and analysis, NMR data collection and analysis; BJY, PCKL, AKM, AMB: experimental design and manuscript preparation.

Appendix A:

Volpato, JP*, Yachnin, BJ*, Blanchet, J, Guerrero, V, Poulin, L, Fossati, E, Berghuis, AM, Pelletier, JN. (2009) Multiple Conformers in Active Site of Human Dihydrofolate Reductase F31R/Q35E Double Mutant Suggest Structural Basis for Methotrexate Resistance. *J. Biol. Chem.* **284**: 20079-20089. ***Co-first authors.**

Author contributions: JPV, VG, EF: molecular biology and enzyme kinetics; JPV, JB, LP: protein expression and purification; JPV, BJY, JB, LP: protein crystallization and data collection; BJY: structure solution and refinement; JPV, BJY: structural analysis; JPV, JNP: docking studies; JPV, BJY, EF, AMB, JNP: experimental design and manuscript preparation.

Copyright © 2009 The American Society for Biochemistry and Molecular Biology, Inc.

Appendix B:

Yachnin, BJ, Colin, DY, Volpato, JP, Ebert, M, Pelletier, JN, Berghuis, AM. (2011) Novel crystallization conditions for mutant R67 DHFR yields wild-type crystal structure. *Acta Crystallogr. F.* **F67**: 1316-1322.

Author contributions: BJY: Crystal structure data collection, structure solution, refinement, and analysis; JPV, ME: protein expression and purification, crystallization; DYK: SDS-PAGE and MS analysis; BJY, DYK, JPV, JNP, AMB: experimental design; BJY, DYK, JNP, AMB: manuscript preparation.

Copyright © 2011 International Union of Crystallography.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	ii
AUTHOR CONTRIBUTIONS	iv
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES.....	x
LIST OF ABBREVIATIONS	xii
ABSTRACT	xvi
ABRÉGÉ.....	xviii
Chapter 1: Introduction	1
1.1 Biocatalysis and Green Chemistry.....	1
1.2 Baeyer-Villiger Oxidation	2
1.2.1 Discovery, Mechanism, and the Criegee Intermediate.....	2
1.2.2 Stereochemistry of Baeyer-Villiger Oxidation	3
1.2.3 Applications of the Chemical Baeyer-Villiger Oxidation	5
1.3 Discovery of Baeyer-Villiger Monooxygenases.....	6
1.3.1 Cyclohexanone Monooxygenase and Cyclopentanone Monooxygenase.....	6
1.3.2 Genome Mining	9
1.3.3 Directed Evolution and Rational Design of Baeyer-Villiger Monooxygenases.....	10
1.3.4 Baeyer-Villiger Monooxygenases in Organic Synthesis.....	13
1.4 Structural Studies of Baeyer-Villiger Monooxygenases.....	14
1.4.1 Phenylactone Monooxygenase.....	14
1.4.2 Cyclohexanone Monooxygenase	17
1.4.3 Other Baeyer-Villiger Monooxygenases.....	21
1.5 Research Objectives	23
Chapter 2: The Substrate-Bound Crystal Structure of a Baeyer–Villiger Monooxygenase Exhibits a Criegee-like Conformation	25
2.1 Preface.....	25
2.2 Introduction	25
2.3 Materials and Methods	28
2.3.1 Subcloning the <i>chnB1</i> gene into the pJW234 expression vector.....	28
2.3.2 Creation of the L145N, L145D, and F507Y Mutants	29
2.3.3 Expression and Purification of CHMO	29
2.3.4 Crystal Structure Determination	31
2.3.5 Saturation Transfer Difference NMR Spectroscopy	32
2.3.6 Enzyme Kinetics	33
2.4 Results	34
2.4.1 Crystal Structure of Substrate-Bound CHMO	34
2.4.2 Rotation of the NADP ⁺ Cofactor	35
2.4.3 The Binding of the Substrate, Cyclohexanone.....	36
2.4.4 Structure–Function Studies Probing the Rotated Conformation.....	37
2.5 Discussion	42
2.5.1 A Third NADP ⁺ Conformation.....	42
2.5.2 The Catalytic Position of Cyclohexanone.....	45
2.5.3 A Structural Mechanism for BVMO Catalysis	48
2.6 Conclusion.....	50
Chapter 3: Lactone-bound structures of cyclohexanone monooxygenase reveal the basis for substrate selection	51
3.1 Preface.....	51
3.2 Introduction	51
3.3 Materials and Methods	54
3.3.1 Purification and Crystallization	54
3.3.2 Data Collection and Structure Solution.....	55
3.3.3 Creation of the W492A Mutant	56
3.3.4 Substrate Profiling of Wild-type and W492A CHMO	56

3.4 Results	57
3.4.1 Crystal structures of CHMO in complex with its product.....	57
3.4.2 Presence of ϵ -caprolactone in the substrate binding site	59
3.4.2.1 CHMO _{Tight} Crystal Structure	59
3.4.2.2 CHMO _{Loose} Crystal Structure	60
3.4.2.3 Comparison of the lactone positions	61
3.4.3 Binding Site Properties of the Lactone-Bound Structures	61
3.4.4 Effects of the W492A on Substrate Profile	63
3.5 Discussion	64
3.5.1 CHMO is a dynamic enzyme	64
3.5.2 CHMO _{Tight} is likely the conformation where stereospecificity is determined	66
3.5.3 The stereospecificity of CHMO can be predicted by the CHMO _{Tight} structure	68
3.5.4 Transition from CHMO _{Tight} to CHMO _{Rotated}	73
3.5.5 The role of the loop in the catalytic cycle	76
3.5.6 CHMO _{Loose} and CHMO _{Tight} are consistent with the BVMO structural mechanism	77
3.6 Conclusion.....	80
Chapter 4: Baeyer-Villiger Monooxygenase Loop Dynamics Examined by SAXS and NMR.....	81
4.1 Preface.....	81
4.2 Introduction	81
4.3 Materials and Methods	83
4.3.1 Subcloning the <i>cpdB</i> gene into the pJW234 expression vector.....	83
4.3.2 Small-Angle X-ray Scattering.....	84
4.3.2.1 Protein Expression and Purification	84
4.3.2.2 Sample Preparation and Data Collection.....	84
4.3.2.3 BVMO Ligand Comparison Primary Data Analysis	85
4.3.2.4 NADP ⁺ Titration Data Analysis	87
4.3.2.5 Ensemble Modelling of the Loop.....	87
4.3.3 Nuclear Magnetic Resonance.....	88
4.3.3.1 Protein Expression and Purification	88
4.3.3.2 Sample Preparation	88
4.3.3.3 Data Collection	88
4.4 Results	90
4.4.1 NADP ⁺ Decreases the R _g of BVMOs	90
4.4.2 Loop Positioning Can Account for the R _g Change.....	94
4.4.3 The Loop Behaves as a Peptide in Solution.....	95
4.4.4 Loop Dynamics.....	96
4.5 Discussion	97
4.6 Conclusion.....	104
Chapter 5: Conclusion.....	106
Appendix A: Multiple Conformers in Active Site of Human Dihydrofolate Reductase F31R/Q35E	
Double Mutant Suggest Structural Basis for Methotrexate Resistance	114
A.1 Introduction	114
A.2 Experimental Procedures	117
A.2.1 Construction of Vectors hDHFR F31R/Q35E-pET24, hDHFR F31R-pQE32, and	
hDHFR Q35E-pQE32.....	117
A.2.2 Expression and Purification of hDHFR F31R/Q35E	118
A.2.3 Expression and purification of His ₆ -hDHFR F31R/Q35E	119
A.2.4 Crystallization and X-ray Data Collection of hDHFR F31R/Q35E	120
A.2.5 Crystallization and X-ray data collection of His ₆ -hDHFR F31R/Q35E.....	122
A.2.6 Structure Determination and Refinement of hDHFR F31R/Q35E.....	123
A.2.7 Structure Determination and Refinement of His ₆ -hDHFR F31R/Q35E.....	123
A.2.8 <i>In Silico</i> Automated Docking of Folate.....	123
A.3 Results	124
A.3.1 Kinetic and Inhibition Double Mutant Cycle of hDHFR Variant F31R/Q35E	124
A.3.2 Overall Structure and General Characteristics of hDHFR F31R/Q35E Complexed to	
MTX	126

A.3.3 Interactions between F31R/Q35E and Bound MTX at the Active Site.....	128
A.3.4 Differences in MTX Binding Resulting from the F31R and Q35E Substitutions	133
A.3.5 Docking of Folate upon the Crystallized F31R/Q35E Structure	138
A.4 Discussion.....	139
Appendix B: Novel crystallization conditions for tandem variant R67 DHFR yield a wild-type crystal structure	143
B.1 Introduction.....	143
B.1.1 Structural features of R67 DHFR.....	143
B.1.2 Mutational analysis of key active-site residues	144
B.2 Materials and Methods.....	145
B.2.1 Creation of the R67 DHFR INSF tandem variant construct and protein purification	145
B.2.2 Crystallization and data collection	147
B.2.3 Structure solution and refinement	148
B.2.4 Analysis of the mutation sites	150
B.2.5 Analysis of protein samples before and after digestion.....	150
B.3 Results and discussion	151
B.3.1 Protein purification.....	151
B.3.2 Crystallization and structure determination.....	151
B.3.3 Examination of the mutation sites in the crystal structure.....	153
B.3.4 Rationale for the variant protomers being excluded from the crystal.....	153
B.3.5 Implications for structural and functional studies of wild-type and heterodimeric R67 DHFR variants	158
B.3.6 Implications for the use of <i>in situ</i> proteolysis as a crystallization method	159
LIST OF REFERENCES	161

LIST OF TABLES

Table 2.1: PCR Primers for subcloning and site-directed mutagenesis.....	29
Table 2.2: Data Collection and Refinement Statistics for the CHMO _{Rotated} Crystal Structure (PDB ID 3UCL).....	31
Table 2.3: Enzyme kinetics data.....	40
Table 3.1: Data Collection and Refinement Statistics for the CHMO _{Tight} and CHMO _{Loose} Crystal Structures.....	55
Table 3.2: Substrate profiling of wild-type and W492A <i>Rm</i> CHMO.....	64
Table 3.3: Regiospecificity of CHMO and CPMO with respect to four terpenones.....	71
Table 4.1: Ligand concentrations for SAXS data collection.....	85
Table 4.2: SAXS data collection and analysis statistics for CHMO.....	86
Table 4.3: SAXS data collection and analysis statistics for CPDMO.....	86
Table 4.4: Chemical exchange parameters obtained from the NADP ⁺ -free CPMG relaxation dispersion experiments.....	96
Table A.1: Crystallographic data.....	121
Table A.2: Kinetic and inhibitory parameters of WT hDHFR and hDHFR variant F31R/Q35E.....	125
Table A.3: Polar interactions in F31R/Q35E and WT hDHFR (1U72) complexed with MTX ...	134
Table B.1: Sample information.....	147
Table B.2: Data-collection and structure-solution statistics.....	148
Table B.3: Structure refinement and model validation.....	149

LIST OF FIGURES

Fig. 1.1: The Baeyer-Villiger oxidation reaction.....	3
Fig. 1.2: The three proposed mechanisms for BVO.....	4
Fig. 1.3: Proposed pathway to convert cyclohexanol to adipate, and then to acetyl-CoA and succinyl-CoA.....	7
Fig. 1.4: BVMO enzyme mechanism.....	8
Fig. 1.5: A selection of substrates that can be converted by CHMO.....	11
Fig. 1.6: The crystal structure of PAMO (PDB ID 1W4X).....	15
Fig. 1.7: Overall structure of the BVMOs.....	17
Fig. 1.8: Rotation of the NADPH-binding domain in the CHMO _{Open} (orange, PDB ID 3GWF) and CHMO _{Closed} (blue, PDB ID 3GWD) structures.....	18
Fig. 1.9: "Sliding" of NADP ⁺ when comparing the CHMO _{Open} (A, PDB ID 3GWF) and CHMO _{Closed} (B, PDB ID 3GWD) structures.....	19
Fig. 1.10: Putative substrate binding pocket in the CHMO _{Closed} (A, PDB ID 3GWD) and CHMO _{Open} (B, PDB ID 3GWF) structure.....	20
Fig. 1.11: Conformations of R337 (PAMO)/R329 (<i>Rm</i> CHMO).....	21
Fig. 1.12: The BVMO signature motif (non-transparent blue) in NADP ⁺ -free PAMO (A, PDB ID 1W4X) and NADP ⁺ -bound CHMO _{Open} (B, PDB ID 3GWF).....	22
Fig. 2.1: View of the active site in the CHMO _{Rotated} (green) structure.....	34
Fig. 2.2: NADP ⁺ blocks cyclohexanone from diffusing into the bulk solvent in the Rotated conformation.....	35
Fig. 2.3: The CHMO _{Rotated} crystal structure (green) is shown superimposed on a model of the Criegee intermediate (cyan).....	37
Fig. 2.4: The percent STD effect for the nicotinamide (left) and adenine (right) protons, using A2 as a reference, for the wild-type and the three mutant enzymes.....	38
Fig. 2.5: The K _M and k _{cat} of NADPH for the wild-type and the three mutants are shown.....	39
Fig. 2.6: The uncoupling ratio is compared to k _{cat} /K _M (NADPH) and k _{cat} /K _M (cyclohexanone).....	41
Fig. 2.7: The NADP ⁺ conformations in the CHMO _{Closed} (blue) and CHMO _{Rotated} (green) structures are compared to the rotated NADP ⁺ conformation of <i>m</i> FMO (purple, PDB ID 2XLR).....	46
Fig. 2.8: Comparison of the "back" of CHMO (panel A, green, CHMO _{Rotated} conformation) and PAMO (panel B, wheat, PDB ID 2YLT).....	47
Fig. 2.9: Schematic of the proposed mechanism for CHMO.....	49
Fig. 3.1: Rotation of the NADPH-binding domain.....	58
Fig. 3.2: Electron density map for ϵ -caprolactone in the (A) CHMO _{Tight} and (B) CHMO _{Loose} crystal structures.....	60
Fig. 3.3: Comparison of the CHMO _{Tight} and CHMO _{Loose} active sites.....	62
Fig. 3.4: Comparison of the ϵ -caprolactone binding site in the (A) CHMO _{Loose} and (B) CHMO _{Tight} crystal structures.....	63
Fig. 3.5: Position of ϵ -caprolactone relative to the residues important for substrate specificity in the CHMO _{Tight} structure.....	67
Fig. 3.6: Enantiospecificity of CHMO for 4-hydroxycyclohexanone.....	70
Fig. 3.7: Regiospecificity of CHMO and CPMO for various terpenones.....	72
Fig. 3.8: R329 plays a role in guiding the substrate and product from the Tight position to the Rotated position and back.....	74
Fig. 3.9: The Criegee intermediate of large substrates like decalone cause clashes in the Tight conformation.....	75
Fig. 3.10: Schematic of the proposed mechanism for CHMO.....	79
Fig. 4.1: Raw SAXS curves and probability distance distribution plots for (A, B) CHMO and (C, D) CPDMO with and without ligands.....	90
Fig. 4.2: Radius of gyration and flexibility analysis for (A) CHMO and (B) CPDMO.....	91
Fig. 4.3: SAXS similarity maps for (A) CHMO and (B) CPDMO.....	92
Fig. 4.4: Change in R _g of CHMO with the titration of NADP ⁺	93
Fig. 4.5: Ensemble modelling of the loop.....	94
Fig. 4.6: CPMG relaxation dispersion data.....	98
Fig. 4.7: Summary of dynamic properties of the BVMOs.....	102

Fig. 5.1: Schematic of the proposed mechanism for CHMO.....	108
Fig. 5.2: Summary of dynamic properties of the BVMOs.	109
Fig. A.1: Chemical structures of hDHFR ligands.....	115
Fig. A.2: Double mutant cycle of F31R/Q35E for DHF and MTX affinity.	126
Fig. A.3: Electron density around MTX and NADPH from low resolution structure of His ₆ - hDHFR F31R/Q35E.	127
Fig. A.4: Electron density around MTX molecule in 3EIG.....	128
Fig. A.5: Stereo view of structure of hDHFR variant F31R/Q35E (3EIG).	129
Fig. A.6: Active site residues observed as two or more conformers in the crystal structure of the F31R/Q35E mutant.	130
Fig. A.7: Bound MTX in hDHFR variant F31R/Q35E.	131
Fig. A.8: Alternate side chain conformers at active site residues Tyr33 (A) and Arg31 (B) in hDHFR variant F31R/Q35E (3EIG).	132
Fig. A.9: Multiple conformers of Arg31 in the lower resolution His ₆ -hDHFR F31R/Q35E (in lines) relative to 1.7 Å resolution hDHFR F31R/Q35E (in sticks).	133
Fig. A.10: Shift of loop 17–27 in hDHFR variant F31R/Q35E.....	135
Fig. A.11: Comparison of WT hDHFR and variant F31R/Q35E by modeling.	136
Fig. B.1: Organization of the INSF tandem dimer gene.....	146
Fig. B.2: Change in magnitude of map σ and R factor with increasing wild-type character.....	154
Fig. B.3: Electron-density map of the mutation sites.	155
Fig. B.4: Comparison of the untreated and chymotrypsin-treated INSF tandem dimers on SDS– PAGE.	156
Fig. B.5: Proposed model for the degradation of the INSF tandem dimer by chymotrypsin.	158

LIST OF ABBREVIATIONS

*Ac*CHMO: CHMO obtained from *Acinetobacter* sp. NCIMB 9871

ACMO: Acetone monooxygenase

BV: Baeyer-Villiger

BVMO: Baeyer-Villiger monooxygenase

BVO: Baeyer-Villiger oxidation

CAST: Complete active site saturation test

CDMO: Cyclodecanone monooxygenase

CHMO: Cyclohexanone monooxygenase

CHMO_{Closed}: Crystal structure of CHMO in the "Closed" conformation (PDB ID: 3GWD)

CHMO_{Loose}: Crystal structure of CHMO in the "Loose" conformation

CHMO_{Open}: Crystal structure of CHMO in the "Open" conformation (PDB ID: 3GWF)

CHMO_{Rotated}: Crystal structure of CHMO in the "Rotated" conformation (PDB ID: 3UCL)

CHMO_{Tight}: Crystal structure of CHMO in the "Tight" conformation

CPDMO: Cyclopentadecanone monooxygenase

CPMG: Carr-Purcell-Meiboom-Gill

CPMO: Cyclopentanone monooxygenase

CV: Column volume

DHF: 7,8-dihydrofolate

DHFR: Dihydrofolate reductase

D_{max}: Maximum interatomic distance

DTT: Dithiothreitol

EDTA: ethylenediaminetetraacetate

EE: enantiomeric excess, defined as $|[R] - [S]| / ([R] + [S])$

EtaA: ethionamide monooxygenase

FAD: Flavin adenine dinucleotide, oxidized form

FMO: Flavin-dependent monooxygenase

GR₂: Glucocorticoid receptor subfamily

HAPMO: 4-hydroxyacetophenone monooxygenase

H-bond: Hydrogen bond

hDHFR: Human dihydrofolate reductase

HSQC: Heteronuclear single quantum coherence

IC₅₀: Half maximal inhibitory concentration

IPTG: isopropyl-1-thio-β-D-galactopyranoside

k_{cat}: Michaelis-Menten enzyme rate constant

K_d: Dissociation equilibrium constant

K_i: Inhibitory equilibrium constant

K_{iso}: Isomerization equilibrium constant

K_M: Michaelis constant

k_{off}: Dissociation rate constant

MEKMO: methyl ethyl ketone monooxygenase

MES: 2-(N-morpholino)ethanesulfonic acid

*m*FMO: FMO from *Methylophaga* sp. strain SK1

MPD: 2-methyl-2,4-pentanediol

MS: Mass spectroscopy

MtmOIV: premithramycin B monooxygenase

MTX: Methotrexate

MTXO: *N*-[4-[(2,4-diaminofuro[2,3-*d*]pyrimidin-5-yl)methyl]methylamino]benzoyl]-L-glutamate

MWCO: Molecular weight cut-off

NADP⁺: β -Nicotinamide adenine dinucleotide phosphate, oxidized form

NADPH: β -Nicotinamide adenine dinucleotide phosphate, reduced form

Ni-NTA: Nickel-nitrilotriacetic acid

NMR: Nuclear magnetic resonance

NOE: Nuclear Overhauser effect

OD: Optical density

OTEMO: 2-Oxo- Δ^3 -4,5,5-Trimethylcyclopentenylacetyl-Coenzyme A Monooxygenase

p-ABA: *p*-aminobenzoic acid

PAMO: Phenylacetone monooxygenase

PCR: Polymerase chain reaction

PEG: polyethylene glycol

PtIE: 1-deoxy-11-oxopentalenic acid monooxygenase

R_g: Radius of gyration

*Rm*CHMO: CHMO obtained from *Rhodococcus* sp. HI-31

RMSD: Root mean squared deviation

SAXS: Small angle X-ray scattering

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SMO: Steroid monooxygenase

STD: Saturation transfer difference

TCEP: Tris(2-carboxy)phosphine

TEV: Tobacco etch virus

THF: 5,6,7,8-tetrahydrofolate

TLS: Translation libration screw

TMOS: Tetramethyl orthosilicate

TMP: Trimethoprim

TOCSY: Total correlation spectroscopy

Tr-NOE: Transferred nuclear Overhauser effect

V_{\max} : Maximum reaction velocity

WT: Wild-type

ABSTRACT

Baeyer-Villiger oxidation (BVO) is a synthetically useful reaction that involves the conversion of ketones to esters, or cyclic ketones to lactones. For industrial and lab-scale applications, peracids are typically used as the oxidizing agent. The Baeyer-Villiger monooxygenases (BVMOs) are a group of bacterial flavoproteins that have garnered a great deal of interest in recent years as potential industrial biocatalysts, as they would replace the hazardous peracid reagents with NADPH and molecular oxygen. Extensive enzymology studies have been performed on these enzymes, revealing a highly-ordered, complex reaction mechanism that includes a ketone-peroxyflavin adduct known as the Criegee intermediate, which mirrors the Criegee intermediate observed in the chemical BVO. In addition, many studies have been aimed at solving the practical problems involved in using BVMOs as industrial-scale biocatalysts.

In spite of the interest in these enzymes, structural data for the BVMOs has only recently been obtained, and remains relatively limited. In particular, there is a lack of crystal structures of BVMOs with a ketone substrate or lactone product. In addition, while the available crystal structures have provided evidence for domain and loop movements, our description of these dynamic properties remain incomplete, and their role in the enzyme mechanism is not fully understood.

Here, we describe three substrate- and product-bound crystal structures of cyclohexanone monooxygenase (CHMO), known as the CHMO_{Rotated}, CHMO_{Tight}, and CHMO_{Loose} crystal structures. Together, these structures provide much insight into the substrate/product-associated states of the enzyme. The CHMO_{Rotated} structure illustrates a Criegee intermediate-like state that can explain the broad substrate specificity of the enzyme. The CHMO_{Tight} structure places the lactone in a position that allows the substrate specificity and stereospecificity to be determined, and could be used to guide attempts to re-engineer the BVMOs for different substrate specificity. The CHMO_{Loose} crystal structure describes a loose binding mode for the substrate and product that likely reflects the enzyme state upon initial binding of the ketone and just prior to the release of the lactone.

In addition, we describe solution studies that provide greater insight into the dynamic properties of CHMO and a second BVMO, cyclopentadecanone monooxygenase (CPDMO). SAXS studies demonstrate that addition of the NADP⁺ cofactor causes both BVMOs to become more compact, likely reflecting changes in the dynamic properties of a large loop that, based on our crystal structures, can be either folded in or disordered. NMR studies allowed us to further describe and quantify these changes in the presence and absence of NADP⁺, the substrate, and the product.

These studies have allowed us to propose a comprehensive structural mechanism to complement the chemical mechanism of the BVMOs. The crystal structures allow us to propose a series of structural states that are consistent with the broad substrate specificity and high stereospecificity of the BVMOs. The solution studies allow us to describe the dynamic properties of the enzymes during and between the catalytic steps. Together, this has greatly increased our understanding of the BVMOs from a structural perspective, and will hopefully pave the way for rational BVMO engineering for specific biocatalytic applications.

ABRÉGÉ

L'oxydation de Baeyer-Villiger (OBV) est une réaction chimique utile qui permet la conversion des cétones en esters ou des cétones cycliques en lactones. Pour les applications industrielles et en laboratoire, les peracides sont généralement utilisés comme agent oxydant. Les mono-oxygénases de Baeyer-Villiger (MOBV) sont un groupe de flavoprotéines bactériennes qui ont suscité un grand intérêt ces dernières années comme biocatalyseurs industriels potentiels, car ils peuvent remplacer les réactifs peracides dangereux par de l'oxygène moléculaire et du NADPH. Des études d'enzymologie approfondies ont été réalisées sur ces enzymes, révélant un mécanisme de réaction complexe et hautement ordonnée qui comprend un intermédiaire de la cétone-peroxyflavine appelé l'intermédiaire Criegee, qui est analogue à l'intermédiaire Criegee observé dans l'OBV chimique. De plus, de nombreuses études ont visé à résoudre les problèmes pratiques liés à l'utilisation des MOBV comme biocatalyseurs à l'échelle industrielle.

Malgré l'intérêt de ces enzymes, les données structurales pour les MOBV n'ont été obtenues que récemment et restent relativement limitées. Il y a, en particulier, un manque de structures cristallines de MOBV avec un substrat de cétone ou d'un produit de lactone. De plus, alors que les structures cristallines disponibles démontrent l'existence de mouvements des domaines et d'une boucle, notre description de ces propriétés dynamiques reste incomplète et leur rôle dans le mécanisme de l'enzyme n'est pas entièrement élucidé.

Ici, nous décrivons trois structures cristallines de l'enzyme mono-oxygénase de la cyclohexanone (MOCH) lié au substrat ou au produit, connu sous les noms MOCH_{Tourné}, MOCH_{Serré}, et MOCH_{Relâché}. Ensemble, ces structures permettent une meilleure compréhension des états de l'enzyme, associée soit avec le substrat ou avec le produit. La structure MOCH_{Tourné} illustre un état de l'enzyme liée à l'intermédiaire Criegee qui peut expliquer sa spécificité pour une large gamme de substrats. La structure MOCH_{Serré} place la lactone dans une position qui permet de déterminer la spécificité du substrat et la stéréospecificité de l'enzyme et pourrait être utilisée pour guider des travaux de réingénierie des

MOBV pour changer la spécificité du substrat. La structure MOCH_{Relâché} décrit un mode de liaison relâché, pour le substrat et le produit, qui reflète probablement l'état initial de la liaison de la cétone à l'enzyme et l'état final juste avant la sortie de la lactone.

De plus, nous décrivons des études en solution qui permettent de mieux comprendre les propriétés dynamiques de MOCH et d'un deuxième MOBV, l'enzyme mono-oxygénase de la cyclopentadécanone (MOCPD). Des études de diffusion des rayons X aux petits angles ("SAXS") démontrent que l'ajout du cofacteur NADP⁺ rend les deux MOBV plus compacts, ce qui reflète probablement les changements des propriétés dynamiques d'une grande boucle qui, selon nos structures cristallines, peut être soit repliée ou désordonnée. Des études de RMN ont permis de décrire et quantifier ces changements en présence et en absence de NADP⁺, du substrat et du produit.

Ces études nous ont permis de proposer un mécanisme structurel global pour compléter le mécanisme chimique des MOBV. Les structures cristallines nous permettent de proposer une série d'états structurels qui sont compatibles avec la spécificité des MOBV envers une large gamme de substrats et leur haute stéréospecificité. Les études en solution nous permettent de décrire les propriétés dynamiques des enzymes pendant et entre les étapes catalytiques. Ensemble, ces études ont grandement amélioré notre compréhension des MOBV d'un point de vue structurel et pourrait ouvrir la voie à l'ingénierie rationnelle des MOBV pour des applications industrielles spécifiques.

Chapter 1: Introduction

1.1 Biocatalysis and Green Chemistry

Since the earliest days of civilization, human beings have adapted the materials around them for their own purposes. Just as we used chemistry before we understood atoms and molecules, we used biotechnology before we understood biology. For example, we have used fermentation and other "biotechnology" methods of preparing food even beyond the very beginnings of recorded history (Cavalieri *et al.*, 2003).

Initially, "biotechnology" methods were used because they were the only way of catalyzing a reaction. In more modern times, biocatalysis has been seen as an attractive way of developing more efficient catalytic methods. In this respect, biocatalysis fulfills many of the so-called Twelve Principles of Green Chemistry (Anastas & Warner, 1998). In particular, it avoids the use of dangerous chemicals, minimizes the use of toxic solvents, and is inherently both renewable and biodegradable. In addition, the high specificity of many biocatalysts eliminates the need for costly and complex protecting groups, and frequently improves the efficiency of the reaction (Bornscheuer *et al.*, 2012). As is often the case, increasing the efficiency of a process can result not only in a more environmentally friendly and sustainable process, but also in a more economical one, creating an economic incentive for the development of biocatalytic syntheses.

While great strides have been made in the last decade in the *de novo* design of enzymes (Kries *et al.*, 2013), we have yet to produce a completely novel enzyme with sufficient efficiency to be useful as an industrial biocatalyst. This means that our greatest resource in the short-term for identification of novel biocatalysts is in the microbial world. Bacteria have evolved mechanisms to digest a wide variety of "foreign" compounds, a process which can be referred to as "xenobiotic biodegradation" (Galvão *et al.*, 2005). Improvements in molecular biology technologies, combined with structural data, have created the capability to use directed evolution and rationale enzyme engineering approaches to optimize

these enzymes for industrial applications and environmental bioremediation (Galvão *et al.*, 2005).

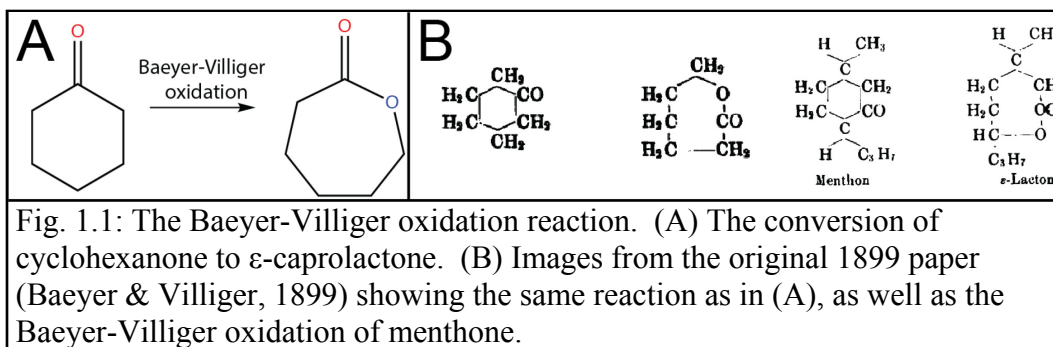
1.2 Baeyer-Villiger Oxidation

1.2.1 Discovery, Mechanism, and the Criegee Intermediate

In 1899, Adolf von Baeyer and Victor Villiger (Baeyer & Villiger, 1899; Baeyer & Villiger, 1900) discovered that a ketone could be converted to ester, or a cyclic ketone to a lactone, using Caro's acid, which had been identified by Heinrich Caro the year before (Caro, 1898). This reaction, which can be performed using various other peracids, became known as the Baeyer-Villiger oxidation (BVO) reaction (Fig. 1.1).

In the decades following the discovery of the Baeyer-Villiger rearrangement, there was significant debate as to the nature of the reaction's mechanism. By the middle of the twentieth century, three possible mechanisms had been proposed for the BVO reaction. Upon discovery of the reaction, Baeyer and Villiger had proposed a dioxirane-based intermediate, which was analogous to the mechanism proposed for the Beckmann rearrangement at that time. This intermediate was termed the "von Baeyer" intermediate (Fig. 1.2A) (Baeyer & Villiger, 1899). A second possibility was proposed by Wittig and Pieper in 1940. They proposed that an OH^+ group from the peracid would be transferred to the carbonyl oxygen, thereby producing the "Wittig" intermediate (Fig. 1.2B) (Wittig & Pieper, 1940). The last proposal was made by Criegee in 1948. He proposed that the peracid would undergo a nucleophilic addition at the carbonyl carbon, thereby forming the "Criegee" intermediate (Fig. 1.2C) (Criegee, 1948).

The question of which of the three intermediates was produced was definitively resolved in 1953 by radiolabelling the carbonyl oxygen with ^{18}O (Fig. 1.2) (Doering & Dorfman, 1953). If the reaction went through a Wittig intermediate, it would be expected that the labelled oxygen would be found entirely in the "ether" position. By contrast, if the reaction occurred using a Criegee intermediate, the labelled oxygen would be located in the "carbonyl" position. Finally, the dioxirane structure proposed in the von Baeyer intermediate would render the two oxygens chemically indistinguishable, thereby leading to

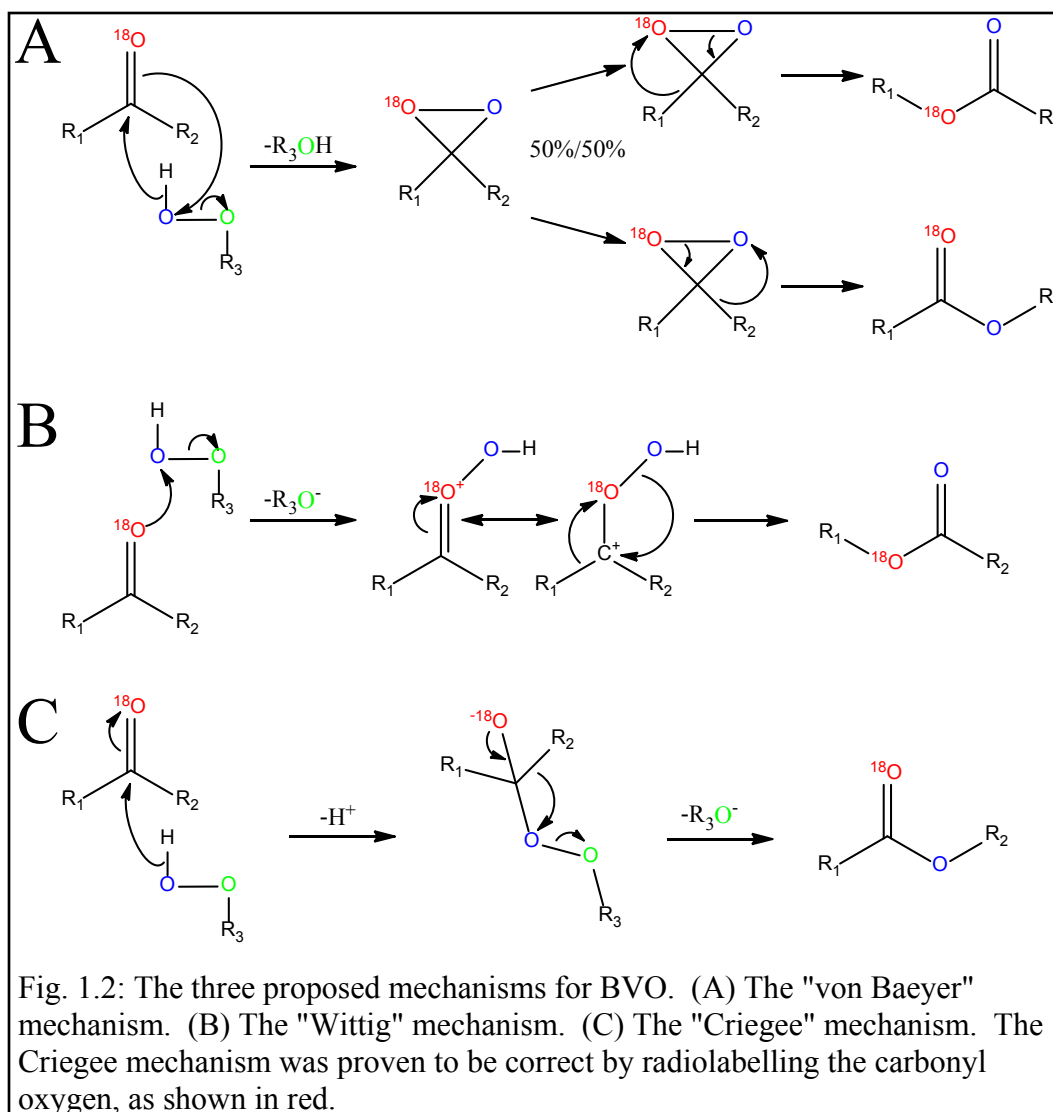


the oxygen being found at both positions in a 50/50 mixture. Doering and Dorfman showed that the labelled oxygen was found only in the carbonyl position, thereby excluding the von Baeyer and Wittig intermediates and suggesting that the Criegee intermediate was in fact formed during BVO (Doering & Dorfman, 1953). It should be noted that in a later study (Camporeale *et al.*, 1990), it was shown that under specific conditions and with specific oxidants, the Criegee intermediate can be forced to re-arrange to form the dioxirane proposed by Baeyer and Villiger, thereby "scrambling" the labelled oxygen.

1.2.2 Stereochemistry of Baeyer-Villiger Oxidation

As the question of the nature of the intermediate in BVO was being resolved, others were working on establishing the stereochemical properties of BVO. In the early 1950s, a number of groups established that the BVO reaction occurred with retention of stereochemistry of the migrating group (Gallagher & Kritchevsky, 1950; Turner, 1950; Mislow & Brenner, 1953). In 1983, this was confirmed to be true even with achiral migrating groups using deuterium-labelled cyclohexanone (Rozzell & Benner, 1983). This property provides some of the synthetic utility of this reaction, as the chirality of the starting material is retained in the resulting products.

Another important question involves the regiochemistry of BVO. A great deal of work has gone into determining which of the two possible groups will migrate from the carbonyl carbon to the peroxide oxygen to form the ester. A number of factors come into play, but it has been determined that as a general rule, the more substituted carbon will be the one to migrate (Doering & Speers, 1950), and that electron donating groups will also favour migration (Doering & Speers, 1950). This generally preferred product is known as the "normal"



product, so as to differentiate from the alternate regioisomer, known as the "abnormal" product. A number of other factors appear to play a role in determining the regioselectivity, however, including the nature of the peracid and steric considerations (Hawthorne *et al.*, 1958).

An additional mechanistic question involves geometric requirements for successful migration of the Criegee intermediate. A recent computational study has demonstrated that in order for migration to occur, the two peroxide oxygens, the carbonyl carbon, and the migrating carbon must have antiperiplanar geometry in the transition state (Snowden *et al.*, 2004). They found that a transition state was unable to form using a gauche configuration. This stereoelectronic requirement appears to supersede the regioselective preferences (Kayser, 2009).

In other words, constraining the Criegee intermediate so that only the less substituted carbon can form the antiperiplanar transition state will result in the formation of the abnormal product.

1.2.3 Applications of the Chemical Baeyer-Villiger Oxidation

The BVO reaction is unquestionably versatile when the body of literature over the past 100 years is considered. This is clearly illustrated in a review by Krow (1993), which cites 1092 references and includes a 420 page table of applications of the BVO reaction. A vast range of carbonyl-containing compounds can be converted, including straight-chain, cyclic, and polycyclic ketones, as well as dicarbonyl compounds. In certain cases, aldehydes can be converted to the corresponding carboxylic acids.

In general, esters are common starting materials for a variety of polymers, including in the synthesis of nylon and a variety of polyesters. They are also known for having "fruity" aromas, and have been utilized in perfumes as well as in the food industry.

Krow (1993)'s review highlights many of the applications of the Baeyer-Villiger oxidation reaction. Antibiotics and their precursors, including erythronolide B, thienamycin, brefeldin A, rifamycin S, nonactic acid, and the Prelog-Djerassi lactone can all be synthesized using Baeyer-Villiger oxidation. It is also applicable to the synthesis of antitumor agents, including pseudoguaianolides, helenanolide, daunosamine, quassinoids, eriolangin, eriolanin, terrecyclic acid A, spatane diterpenes, and mitomycin. Plant and animal hormones and pheromones, as well as prostaglandins and prostacyclins, are also accessible using Baeyer-Villiger oxidation, including estrone, lineatin, norsteroids, gibberellic acid, paniculide A, and a variety of other sterols. Other bioactive compounds produced using Baeyer-Villiger oxidation include reserpine (an antipsychotic and antihypertensive agent), Inhoffen-Lythgoe diol (precursor to vitamin D), L-dopa (a dopamine analogue), zoapatanol (a contraceptive), and insecticides like xylomollin and ryanodol. As expected for a reaction that produces ester products, fragrances like methyl dihydrojasmonate and epi-beta-santalene can also be produced. Finally, some compounds with unusual carbon

skeletons that are difficult to synthesize are produced using Baeyer-Villiger oxidation, including pentaprismane and peristylane. It is clear that the synthetic utility of the Baeyer-Villiger oxidation reaction, both in industry and in the laboratory, cannot be understated (Krow, 1993).

1.3 Discovery of Baeyer-Villiger Monooxygenases

At around the same time that the questions regarding the nature of the intermediate in the chemical BVO reaction were being resolved, the first evidence of the existence of a microbial BVO reaction was uncovered. A microbe that degrades steroids was found to oxidize the carbonyl group on a steroid that had been previously oxidized from a hydroxyl group (Turfitt, 1948). While the lactone product was not identified, it was likely that it was an intermediate in the steroid degradation pathway. Subsequently, a pathway was identified in which both the ketone and lactone forms of the steroid were identified (Fried *et al.*, 1953), demonstrating conclusively for the first time that microbes were capable of converting ketones to esters, the chemical equivalent of Baeyer-Villiger oxidation.

The era of Baeyer-Villiger monooxygenases truly began in 1975. Donoghue and Trudgill (1975) reported that the bacterial strain *Acinetobacter* NCIMB sp. 9871 was able to grow on cyclohexanol and cyclohexanone as its carbon source. They were able to identify a metabolic pathway in which cyclohexanol was converted to adipic acid (Fig. 1.3), which could then be further catabolized by the cell. In one step, they proposed that cyclohexanone would be converted to ϵ -caprolactone. This was followed up in 1976 (Donoghue *et al.*, 1976) when they reported the purification and characterization of a flavoprotein from *Acinetobacter* NCIMB sp. 9871, and a homologue from *Nocardia globerula* sp. CL1, that could catalyze this reaction. These enzymes and their homologues have since been researched intensively as potential biocatalysts to replace the environmentally dangerous peracid catalysts that are routinely used.

1.3.1 Cyclohexanone Monooxygenase and Cyclopentanone Monooxygenase

In 1976, reports of the purification and characterization of two BVMOs were published: cyclohexanone monooxygenase (CHMO) (Donoghue *et al.*,

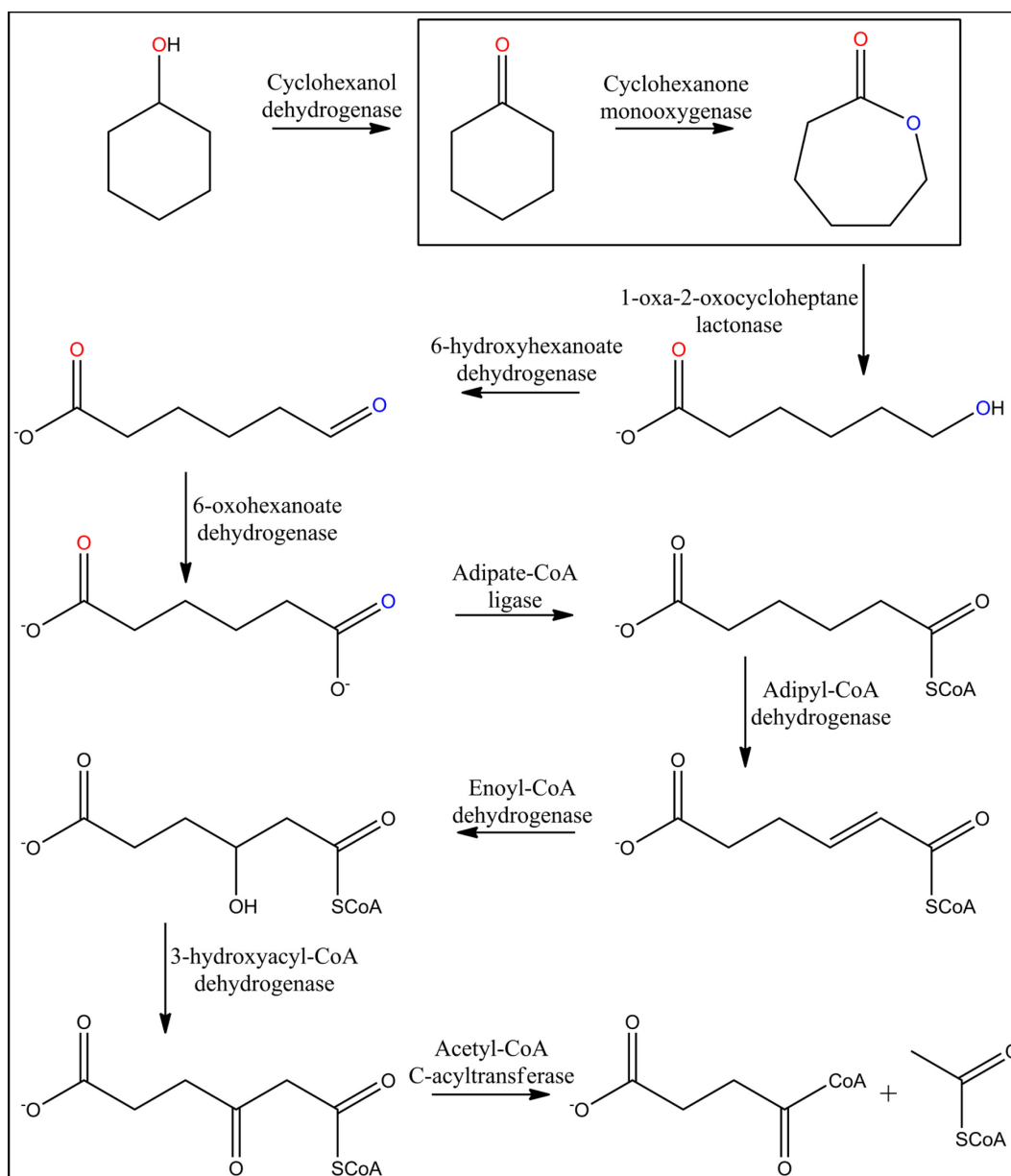


Fig. 1.3: Proposed pathway to convert cyclohexanol to adipate, and then to acetyl-CoA and succinyl-CoA. These products can then feed into the Krebs cycle. The box indicates the reaction catalyzed by CHMO.

1976) and cyclopentanone monooxygenase (CPMO) (Griffin & Trudgill, 1976).

These two enzymes, which have gone on to become the two best-characterized BVMOs, were found to generally exhibit similar behaviour. Both enzymes were characterized as having a single tightly-bound FAD per polypeptide, and were found to be strictly NADPH-dependent. They were both found to act on a range of different cyclic ketones, with CPMO's substrate spectrum being directed towards smaller substrates than that of CHMO. An interesting difference between

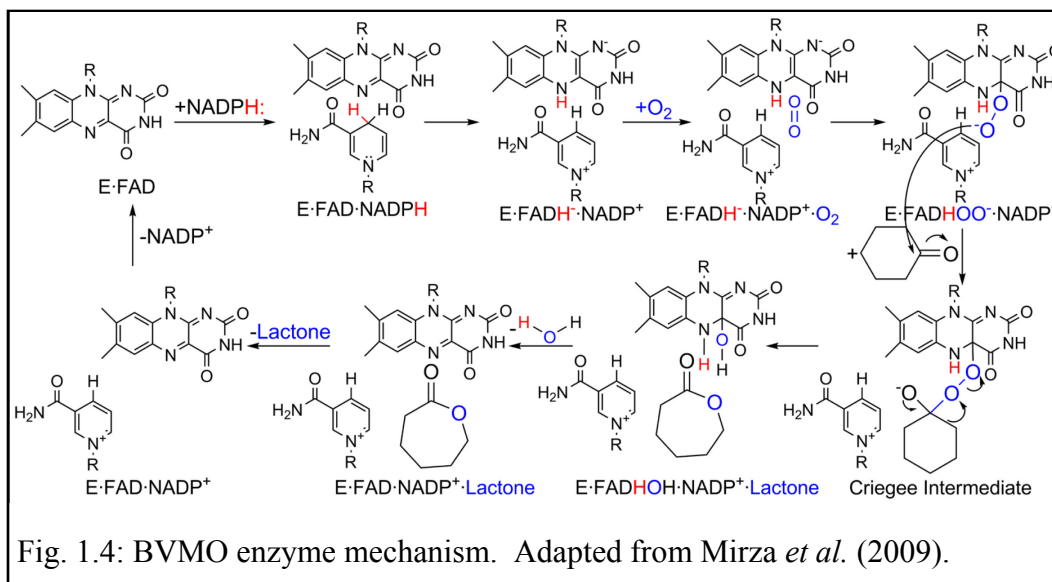


Fig. 1.4: BVMO enzyme mechanism. Adapted from Mirza *et al.* (2009).

the two enzymes is that while CHMO was found to be monomeric, CPMO was determined to be a homotetramer.

The group led by Dr. Christopher T. Walsh produced a number of studies in the 1980s that led to a more detailed understanding of the mechanism and range of substrates available to CHMO. Perhaps the most critical is the 1982 study in which the complete mechanism of CHMO was elucidated (Fig. 1.4) (Ryerson *et al.*, 1982). The spectral properties of FAD permit the tracking of its chemical state based on changes of its absorbance spectrum. This made it possible, using a combination of pre-steady state and steady state kinetics experiments under both anaerobic and aerobic conditions, to propose a detailed description of the highly ordered reaction mechanism. In particular, it was shown that binding of NADPH and reduction of FAD occurred first, followed by binding of molecular oxygen and the formation of an oxygen-flavin adduct, the peroxyanion intermediate. Only at that stage can the substrate bind, react with the peroxyanion, and ultimately form the lactone product. After dehydration of the hydroxylated FAD, the lactone is released prior to release of NADP⁺. Of note, the ketone-peroxyflavin adduct (Ryerson *et al.*, 1982) was proposed to be analogous to the Criegee intermediate (Criegee, 1948) that is formed in the chemical BVO. In addition, NMR studies using deuterated cyclohexanone demonstrated that like the chemical BVO, BVMOs are able to transform ketones with retention of configuration (Schwab *et al.*, 1983).

Walsh's group also discovered that, in addition to catalyzing the BVO reaction, CHMO was also able to oxidize sulfides to sulfoxides (Ryerson *et al.*, 1982). A later study demonstrated that in addition to ketones and sulfides, CHMO could oxidize aldehydes, selenides, boronic acids, phosphite esters, and sodium iodide (Branchaud & Walsh, 1985). Highlighting the versatility of CHMO was the fact that while the oxidation of ketones, aldehydes, and boronic acids requires the formal nucleophilic addition of oxygen to the substrate, the oxidation of sulfides, selenides, phosphite esters, and iodide requires the formal electrophilic addition of oxygen (Branchaud & Walsh, 1985). This implies that CHMO can catalyze two chemically distinct reactions.

A remaining ambiguity in enzyme mechanism was the question of whether the reactive species in CHMO was an anionic peroxyflavin intermediate or the protonated hydroperoxyflavin intermediate. In 2001 (Sheng *et al.*, 2001), a study similar to that performed by Ryerson *et al.* (1982) was able to demonstrate a pH dependence on the rate of formation of the Criegee intermediate. This strongly suggested that it was the anionic form of the peroxyflavin adduct that reacted with the substrate (Sheng *et al.*, 2001). This allowed the proposal of the chemical enzyme mechanism for BVMOs that is accepted today (Fig. 1.4) (Sheng *et al.*, 2001).

1.3.2 Genome Mining

The development of rapid genome sequencing technologies has allowed the identification of a large number of BVMOs in recent years. By 2002, a small number of new BVMOs had been sequenced, including steroid monooxygenase (SMO) (Morii *et al.*, 1999), two new CHMO enzymes (Brzostowicz *et al.*, 2002), 4-hydroxyacetophenone monooxygenase (HAPMO) (Kamerbeek *et al.*, 2001), and cyclododecanone monooxygenase (CDMO) (Kostichka *et al.*, 2001). Using these sequences, it was possible to determine a so-called BVMO signature motif: FxGxxxHxxxW(P/D) (Fraaije *et al.*, 2002). This signature motif was located prior to the second of two GxGxxG motifs, characteristic of Rossmann folds. The conserved residues in this motif were found to be important for the expression of soluble protein and for the tight binding of FAD.

Once this critical sequence was identified, it became possible to quickly identify a variety of BVMOs through genome mining. A large number of "CHMOs" (Brzostowicz *et al.*, 2003; Brzostowicz *et al.*, 2005; Choi *et al.*, 2006; Kim *et al.*, 2008; Rial *et al.*, 2008) and another "HAPMO" (Rehdorf *et al.*, 2009) have been identified, as have BVMOs that target larger cyclic ketones, including CPDMO (Iwaki *et al.*, 2006). In addition, a few BVMOs that can act on linear ketones, like ACMO (Kotani *et al.*, 2007), BmoF1 (Kirschner *et al.*, 2007), and MEKMO (Onaca *et al.*, 2007), have been identified. In other cases, a large number of BVMOs have been identified in the same or closely related species (Brzostowicz *et al.*, 2003; Bonsor *et al.*, 2006; Park *et al.*, 2007; Szolkowy *et al.*, 2009), especially in genera like *Rhodococcus*, *Pseudomonas*, *Mycobacterium*, and *Streptomyces*. While most BVMOs are thought to be used to catabolize molecules for use as a carbon source, some recently discovered BVMOs, like PtlE (Jiang *et al.*, 2009), are involved in biosynthesis, in this case in the production of an antibiotic. In another case, the BVMO EtaA was found to be involved in activating a prodrug used to treat tuberculosis (Fraaije *et al.*, 2004). Two noteworthy examples of BVMOs identified with the help of their BVMO signature sequence are phenylacetone monooxygenase (PAMO) (Fraaije *et al.*, 2005) and a CHMO isolated from *Rhodococcus* sp. HI-31 (*RmCHMO*) (Mirza *et al.*, 2009), which were the first two BVMOs to have their crystal structures solved (Malito *et al.*, 2004; Mirza *et al.*, 2009). With the large number of BVMOs that have now been sequenced, differences in their substrate profile, as well as their regio- and enantiospecificity, create a great deal of versatility in selecting biocatalysts for industrial biocatalytic applications. For example, it was shown that using a panel of different BVMOs, it was possible to access both the "normal" and "abnormal" regioisomers from a series of terpenones (Cernuchova & Mihovilovic, 2007).

1.3.3 Directed Evolution and Rational Design of Baeyer-Villiger Monooxygenases

The increasing number of cloned and characterized BVMOs has expanded the number of substrates that can be oxidized, as well as the regiochemistry and

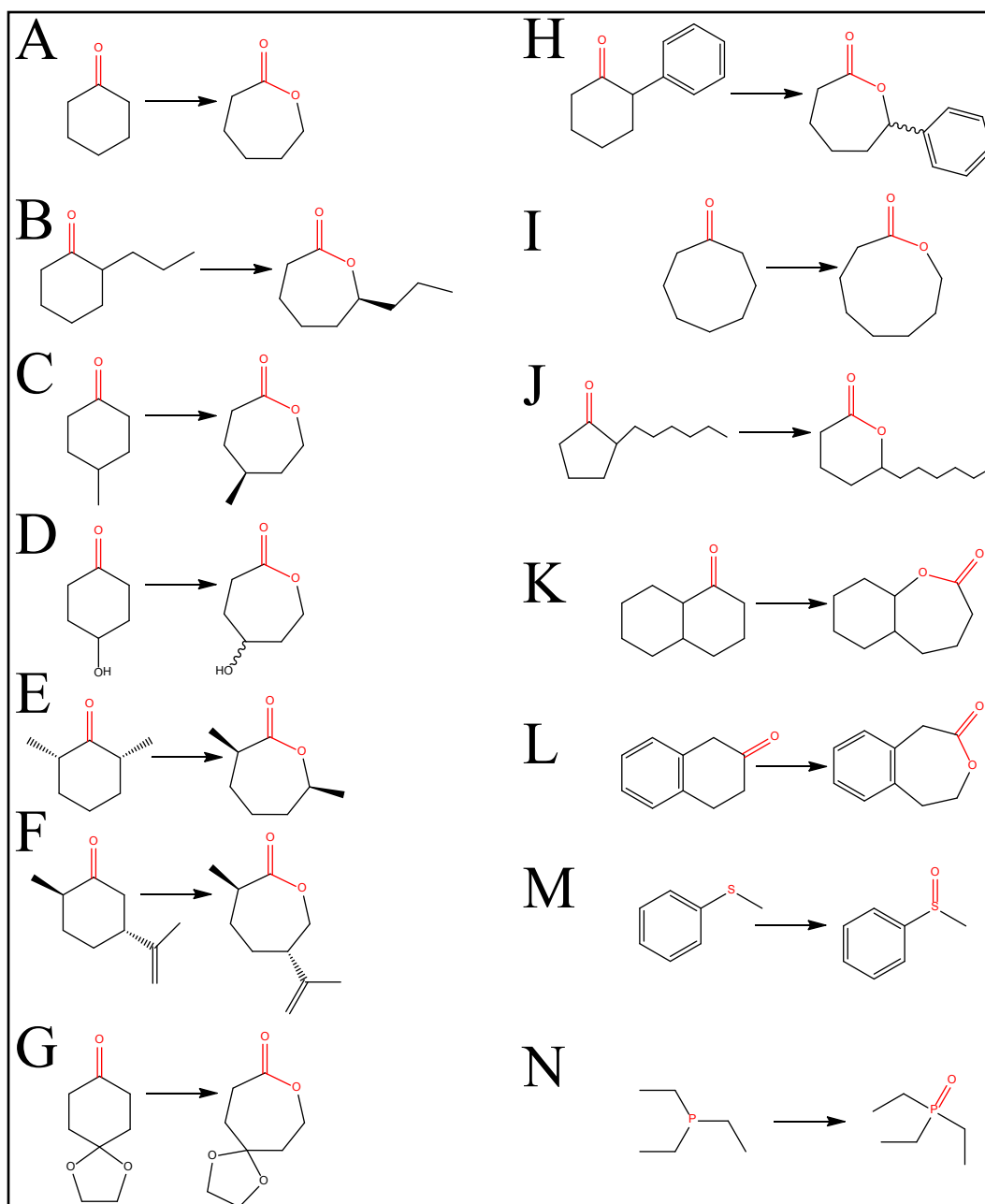


Fig. 1.5: A selection of substrates that can be converted by CHMO. B, C, and E show examples of high enantiospecificity. F, H, J, K, and L show examples of high regiospecificity. M and N are examples of non-BVO reactions. A more complete list for *RmCHMO* can be found in Mirza *et al.*, 2009.

enantiochemistry of the products. In spite of this, the majority of BVMOs prefer aliphatic cyclic ketones with around six carbons as their substrates, similar to *AcCHMO* and *RmCHMO* (Fig. 1.5), with relatively few BVMOs able to target linear and aromatic ketones. With advances in molecular biology techniques, directed evolution has become an increasingly attractive avenue for further

increasing the diversity of BVMOs. The first example of the directed evolution of BVMOs involved back-to-back papers describing the evolution of CHMO to increase its enantiospecificity towards the prochiral substrates 4-hydroxycyclohexanone (Reetz *et al.*, 2004a) and methyl-p-methylbenzyl thioether (Reetz *et al.*, 2004b). Later, BmoF1, which acts on linear ketones, was evolved to improve its enantioselectivity towards 4-hydroxy-2-decanone (Kirschner & Bornscheuer, 2008).

With the publication of the crystal structure of PAMO, the first BVMO to have its structure reported (Malito *et al.*, 2004, and see section 1.4.1), semi-rational approaches were used to complement "pure" directed evolution studies. An attempt to remove a loop present in PAMO, but not in CHMO, was able to slightly broaden PAMO's substrate spectrum (Bocola *et al.*, 2005). A similar approach, in which those residues in PAMO that were both close to the flavin ring system and different in CPMO were mutated, allowed the identification of another mutant that had altered substrate specificity (Torres Pazmiño *et al.*, 2007). In another case, Complete Active Site Saturation Test (CAST) was used to evolve CPMO (Clouthier *et al.*, 2006). Based on the PAMO structure, CPMO's enantioselectivity was modified using this approach. A similar approach was later attempted with PAMO itself (Reetz & Wu, 2008; Dudek *et al.*, 2011). Residues close to the active site, but not directly contacting the binding pocket (Reetz & Wu, 2009), as well as distant, "allosteric" residues (Wu *et al.*, 2010), have also been targeted for saturation mutagenesis. Recently, a simple but effective approach, in which the C-terminal portion of PAMO was replaced with the equivalent amino acids from other BVMOs, generated thermostable BVMOs with differing substrate specificity (van Beek *et al.*, 2012).

In addition to attempts to alter the substrate specificity of BVMOs, attempts have also been made to generate enzymes that are better suited for industrial purposes. The Cys and Met residues in CHMO were mutated in order to increase its oxidative stability towards H₂O₂ (Opperman & Reetz, 2010). Another example involves the generation of a BVMO fusion protein with an NADPH-recycling enzyme. While not necessary for biocatalysis involving whole

cells, in cases where purified enzyme is used, an NADPH regeneration system is necessary to avoid using stoichiometric amounts of the expensive NADPH cofactor. Several BVMO-phosphite dehydrogenase chimeras were produced for this purpose (Torres Pazmiño *et al.*, 2008b). They were later altered to include a thermostable variant of phosphite dehydrogenase (Torres Pazmiño *et al.*, 2009).

These extensive protein engineering studies have highlighted that the BVMOs can be modified, both in terms of their substrate profile and in terms of practical considerations in industrial biocatalysis. At the same time, the difficulty in designing efficient biocatalysts for a specific application without a complete structural description of substrate binding is clear.

1.3.4 Baeyer-Villiger Monooxygenases in Organic Synthesis

With the efforts to extend the number of BVMOs, both by genome mining and by protein engineering, it should not be surprising that attempts have been made to use BVMOs in organic synthesis. Recent reviews (Alphand & Wohlgemuth, 2010; Leisch *et al.*, 2011) have described a number of syntheses that have used BVMOs. These include the synthesis of a polyether antibiotic, ionomycin (Taschner & Chen, 1991), Corey's lactone, an intermediate in prostaglandin synthesis (Fantin *et al.*, 2006; Szolkowy *et al.*, 2009), the Geisman-Waiss lactone, an intermediate in alkaloid synthesis (Luna *et al.*, 2005), lauryl lactone, which is used commercially in detergents and cosmetic products (Yang *et al.*, 2009), and others.

The publication of these syntheses speaks to their potential industrial utility. Perhaps more importantly, though, is the filing of at least three patent applications for syntheses involving BVMOs. The first of these describes the synthesis of lipoic acid, a nutritional supplement (Adger *et al.*, 2000). More recently, the preparation of (–) modafinil using a BVMO has been patented (Riva *et al.*, 2008). A third recent patent, which takes advantage of engineered BVMOs, employs the sulfide oxidation activity of BVMOs to synthesize prazole compounds (Bong *et al.*, 2013). The recent and ongoing developments in the BVMO field are likely to lead to the patenting of more BVMO-catalyzed syntheses, and ultimately to large scale industrial syntheses.

Progress has also been made in overcoming some of the practical problems in the use of BVMOs as biocatalysts. The development of a scaled-up reaction in a 200 L bioreactor has been reported (Baldwin *et al.*, 2008). The stability of the biocatalyst has been improved through immobilization, both when using purified enzyme (Zambianchi *et al.*, 2002) and when using whole cells (Hucík *et al.*, 2010). Yields could also be improved by continuously feeding the substrate instead of adding it all at once (Zambianchi *et al.*, 2002). Extensive work has gone into the development of adsorbent resins to simultaneously feed substrate and remove product, thereby avoiding substrate and product inhibition (Simpson *et al.*, 2001; Hilker *et al.*, 2004a; Hilker *et al.*, 2004b; Hilker *et al.*, 2005; Rudroff *et al.*, 2006; Hilker *et al.*, 2008; Geitner *et al.*, 2010). For similar reasons, performing the reactions using biphasic media (aqueous and organic phases) has been found to improve yields (Schulz *et al.*, 2005; Yang *et al.*, 2009). Finally, the use of non-growing cells has also been found to improve biocatalytic efficiency (Walton & Stewart, 2002).

1.4 Structural Studies of Baeyer-Villiger Monooxygenases

The efforts to expand the chemical space covered by the BVMOs by genome mining and directed evolution have been quite successful. These methods, however, are limited in that they do not allow for a true rational approach to biocatalyst development. It is clear that a complete understanding of the structural state of BVMOs throughout the complex enzyme mechanism, and in particular the binding modes of the substrate and product, are necessary to take full advantage of this family of enzymes.

1.4.1 Phenylactone Monooxygenase

The first crystal structure of a BVMO was published in 2004 (Malito *et al.*, 2004). This structure was of the thermostable phenylacetone monooxygenase (PAMO). PAMO's substrate spectrum, in contrast to CHMO and CPMO, is primarily directed towards linear ketones with aromatic substituents, though a few linear and cyclic non-aromatic substrates can also be converted (Fraaije *et al.*, 2005). Like CHMO, PAMO is also able to oxidize sulfides (Fraaije *et al.*, 2005).

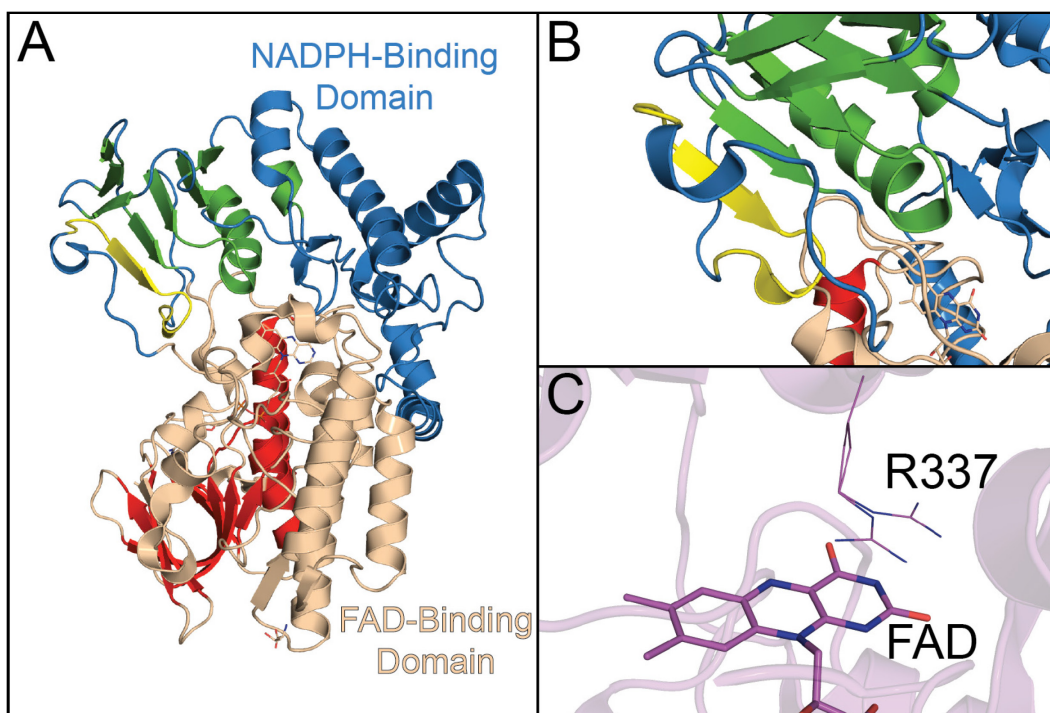


Fig. 1.6: The crystal structure of PAMO (PDB ID 1W4X). (A) The FAD-binding domain is shown in wheat, and the NADPH-binding domain is shown in blue. The glutathione reductase family Rossmann folds for the NADPH-binding and FAD-binding domains are indicated in green and red, respectively. The BVMO signature motif, which is part of the NADPH-binding domain Rossmann fold, is shown in yellow. (B) Expanded view of (A) highlighting the BVMO signature motif. (C) The two rotameric conformations of R337 in PAMO are shown.

This first BVMO crystal structure (Malito *et al.*, 2004) had only the tightly-bound FAD co-factor bound. This allowed the identification of two domains in the PAMO structure: the FAD-binding domain and the NADPH-binding domain (Fig. 1.6A). Both of these domains include a characteristic Rossmann dinucleotide binding fold. The FAD-domain was found to be discontinuous, as it includes the N-terminal and C-terminal 150 residues and is interrupted by the central NADPH-binding domain.

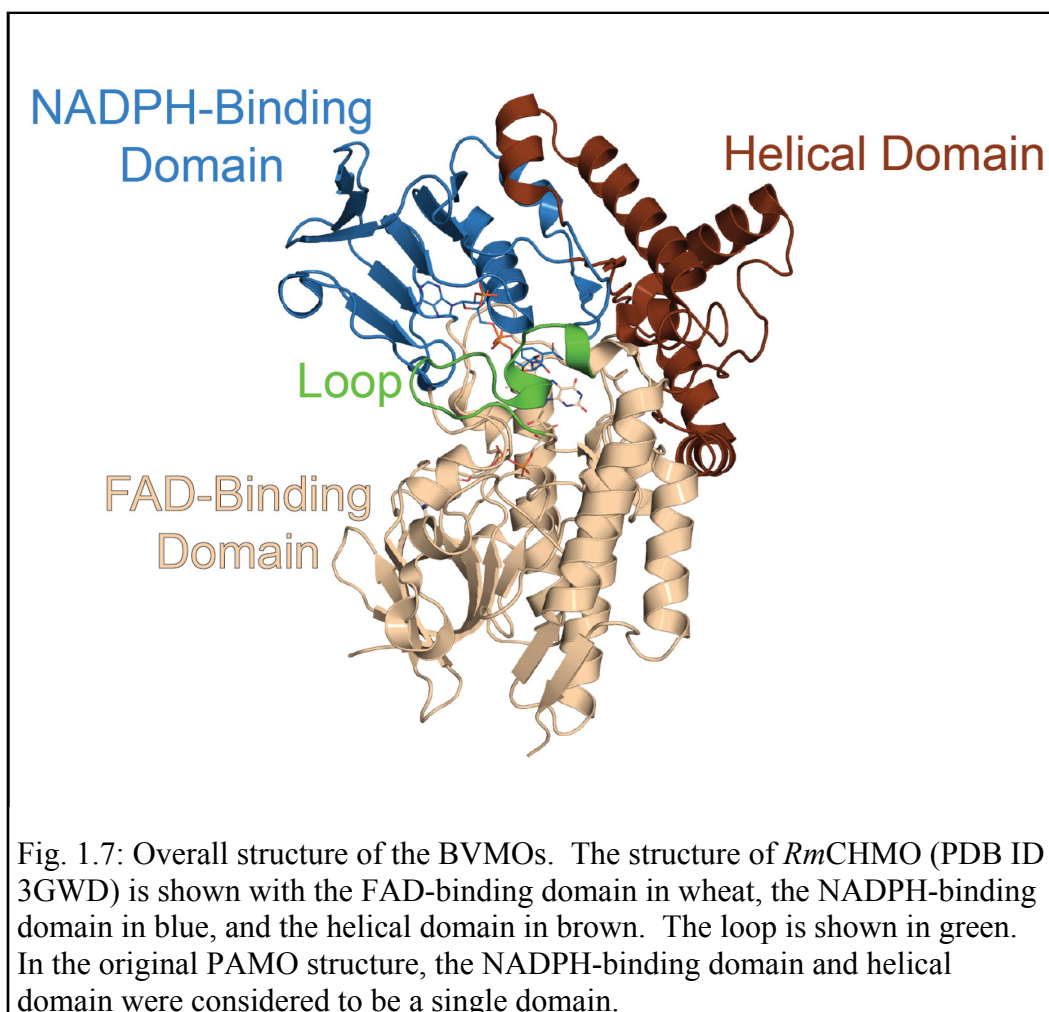
The FAD molecule adopts an extended conformation in the PAMO structure, with the flavin ring system sitting in a cleft in the center of the protein between the FAD- and NADPH-binding domains. R337 was observed in two conformations in close proximity to the flavin rings (Fig. 1.6C). The homologous residue in HAPMO had been previously found to be required for oxygenation of

substrates, but not necessary for reduction of FAD by NADPH (Kamerbeek *et al.*, 2004). Due to its position close to the flavin C4a atom, where molecular oxygen forms the adduct with FAD, it was speculated that while one conformation, the "OUT" conformation, provided the space required for reduction of FAD by NADPH, the "IN" conformation was required for the stabilization of the peroxyanion intermediate. Given the domain architecture of the protein, it was postulated that this change in conformation would be driven by domain movements that would occur throughout the catalytic cycle.

The position of the "BVMO signature motif" was also notable (Fig. 1.6B). Somewhat surprisingly, this motif was not located near the active site, but rather near the N-terminal end of the NADPH-binding domain. Given that this sequence is conserved (Fraaije *et al.*, 2002), and also has been shown to be required through mutagenesis studies (Fraaije *et al.*, 2002; Cheesman *et al.*, 2003), this implied that this sequence must be critical in orchestrating the domain movements that occur throughout the catalytic cycle.

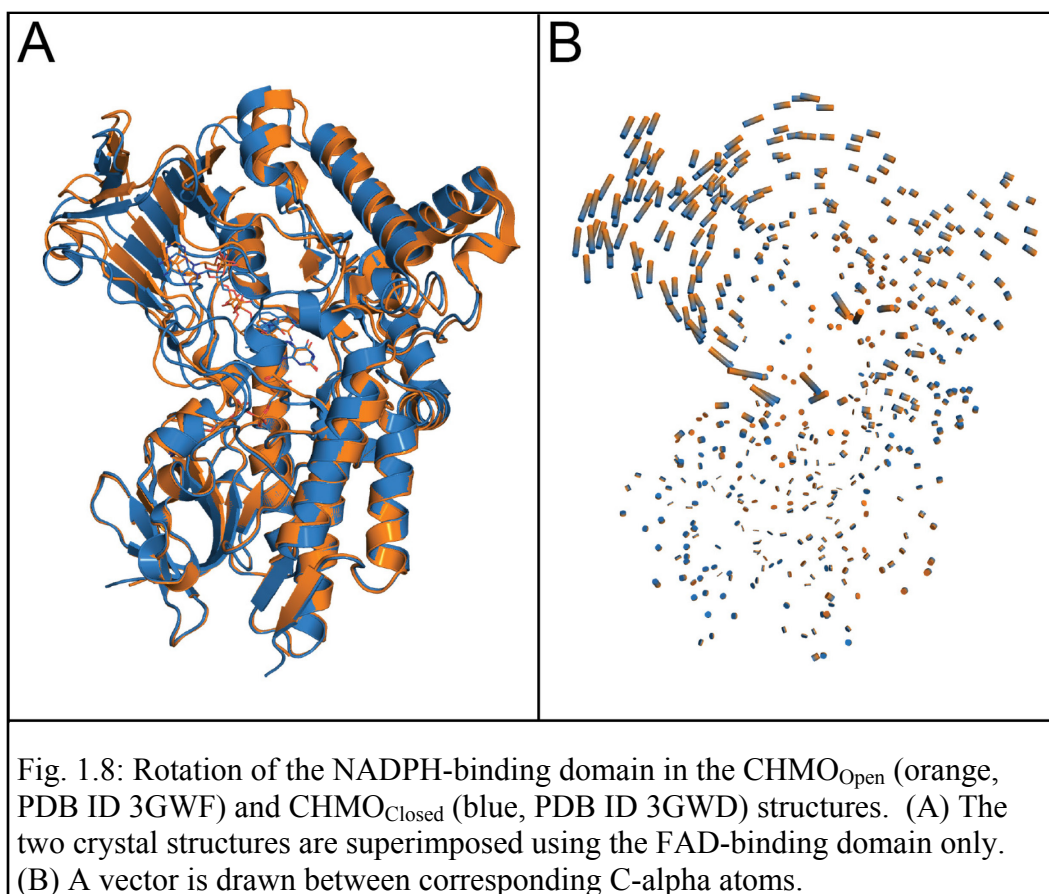
While it was not possible at that time to obtain a structure of PAMO with NADP⁺ bound, the authors were able to identify the likely binding site. This was done based on the Rossmann fold of the NADPH-binding domain (Fig. 1.6A), and also from previous mutagenesis studies in which the binding of NADPH was affected. There was a sulfate ion from the ammonium sulfate precipitant present in the position where the NADP⁺ phosphates were predicted to bind, and in addition, a large loop spanning residues 506-513 from a symmetry-related protein molecule occluded part of the NADPH binding site. As PAMO is expected to be a monomer in solution, this interaction was proposed to be a crystal packing artifact, and no catalytic role for this loop was proposed.

This first crystal structure of PAMO revealed the general architecture of the classical BVMOs, suggested the importance of domain movements in the BVMO catalytic cycle, and allowed an initial proposal for the catalytic mechanism from a structural perspective. A more complete picture, however, would require crystal structures with NADP⁺ and the substrate or product bound.

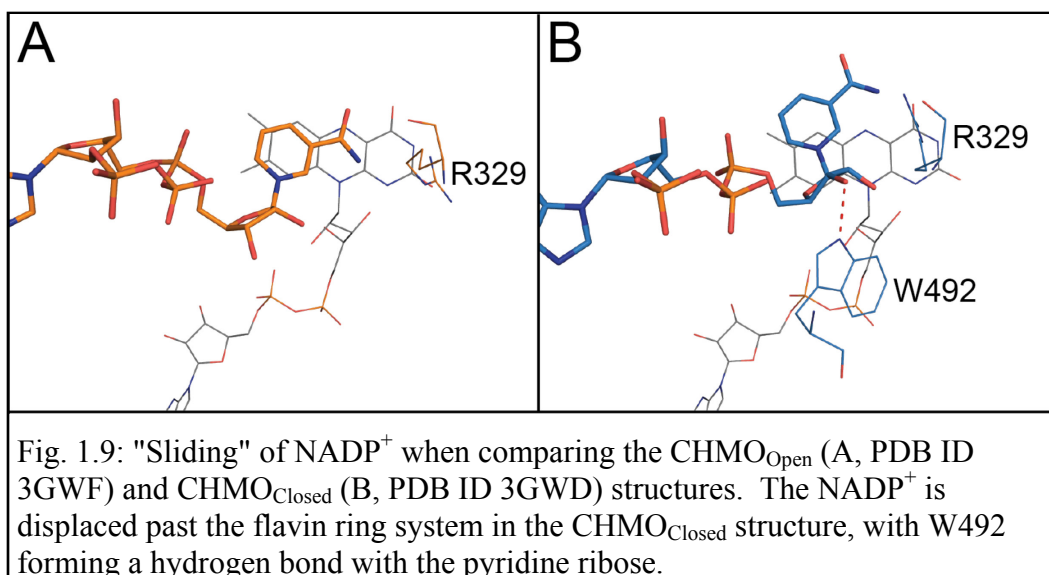


1.4.2 Cyclohexanone Monooxygenase

The next crystal structures to be reported were those of a CHMO homologue from *Rhodococcus* sp. HI-31 (*RmCHMO*) (Mirza *et al.*, 2009). This enzyme has 55% sequence identity to the classical *Acinetobacter* enzyme (*AcCHMO*) and 43% identity to PAMO. Accordingly, the substrate spectrum was found to be quite similar, though the newly identified enzyme was found to have a slightly broader substrate spectrum than the classical CHMO (Fig. 1.5) (Mirza *et al.*, 2009). Two distinct crystal structures were reported of *RmCHMO*, both with FAD and NADP⁺ bound (Mirza *et al.*, 2009). This not only allowed a detailed understanding of the nicotinamide cofactor binding mode, but also provided more information regarding the domain movements proposed based on the PAMO structure (Malito *et al.*, 2004).



While the overall architecture of CHMO closely resembled that of PAMO, a subtlety in the domain organization was apparent that could not have been identified using the PAMO structure alone. The NADPH-binding domain originally identified in the PAMO structure (Fig. 1.6A) could be subdivided into two based on a comparison of the two *Rm*CHMO structures (Fig. 1.7). In the central portion of the NADPH-binding domain was an entirely helical segment which was reclassified as a distinct domain, referred to as the "helical domain." When the two structures of CHMO were aligned based on the FAD-binding domain, it was noted that the remaining NADPH-binding domain underwent a pronounced rotation (Fig. 1.8). In comparison, the helical domain exhibited much less movement than the NADPH-binding domain. This effectively causes the NADPH-binding domain to close over the active site in one structure, and leaves it open in the other. For this reason, these two structures were referred to as the CHMO_{Closed} and CHMO_{Open} structures, respectively.



In addition to the distinct domain positions, these two structures also revealed distinct NADP⁺ binding modes. In the CHMO_{Open} structure, the pyridine moiety of NADP⁺ sits above and parallel to the flavin ring system (Fig. 1.9A). The pyridine ring sits above the xylene ring of FAD, which is in contrast to the expected position for flavin reduction above the central flavin ring. This discrepancy may be due to the fact that the charged, oxidized pyridine group was present instead of the neutral reduced pyridine. In contrast, the CHMO_{Closed} structure has the pyridine ring displaced laterally beyond the flavin ring system, eliminating any pi stacking interactions between the two cofactors (Fig. 1.9B). This suggested a "sliding" movement of the NADP⁺ cofactor during catalysis.

The loop, spanning residues 487-504 in *RmCHMO* (Fig. 1.7), which was observed in the NADP⁺-binding cleft in the PAMO structure, was found in two very different states in the two *RmCHMO* structures. In CHMO_{Open}, this loop was completely disordered and not visible in the electron density map. In contrast, the CHMO_{Closed} structure has the loop folded in and in contact with the pyridine ribose of NADP⁺ through the strictly conserved residue W492. This interaction may be linked to the two conformations of NADP⁺ observed in the two crystal structures.

A combination of the domain movements, the shift of NADP⁺, and the order-disorder transition of the loop creates two very distinct environments in a putative substrate binding pocket. In the CHMO_{Closed} structure, a small pocket

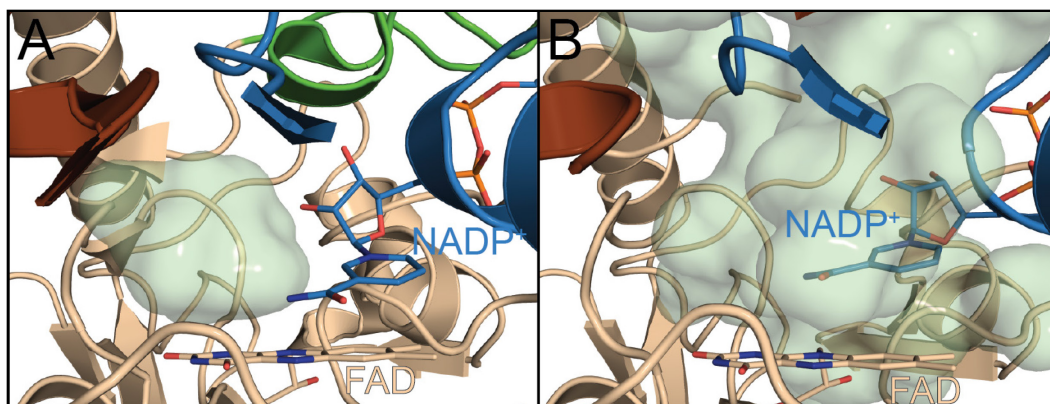
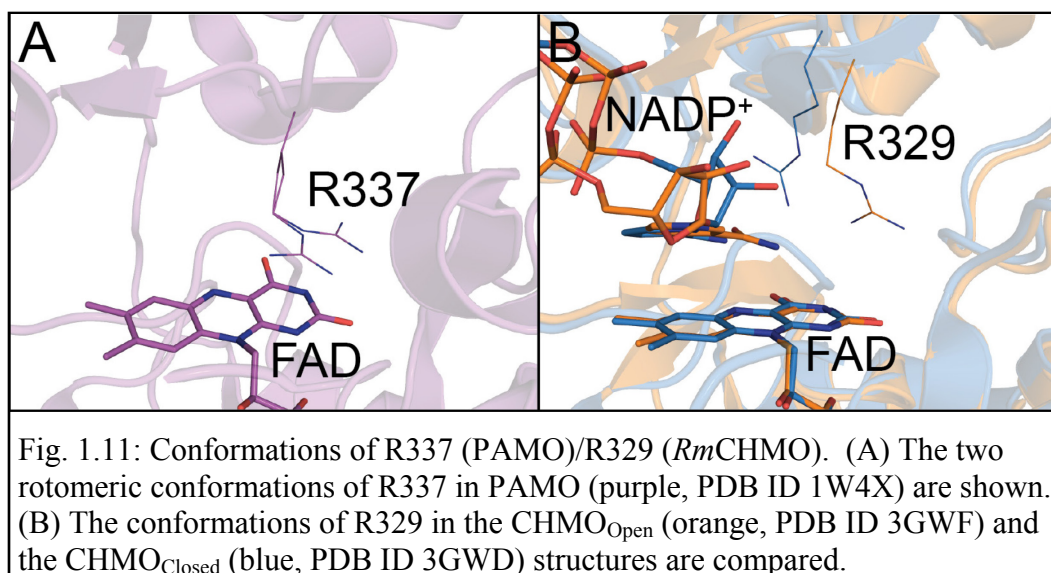


Fig. 1.10: Putative substrate binding pocket in the CHMO_{Closed} (A, PDB ID 3GWD) and CHMO_{Open} (B, PDB ID 3GWF) structure. The space available for a substrate is shown as a transparent green surface.

that is sealed off from the solvent is observed (Fig. 1.10A). This pocket is just big enough for cyclohexanone, and would perhaps accommodate some smaller substituents on cyclohexanone. The CHMO_{Open} structure has a much larger, solvent accessible pocket (Fig. 1.10B).

The strictly conserved Arg residue, numbered R329 in *Rm*CHMO (R337 in PAMO, Fig. 1.11A), was observed in both structures in a similar rotameric conformation as the "OUT" conformation from the PAMO structure. This is consistent with the "OUT" conformation making sufficient room for pyridine moiety of NADP⁺. In CHMO_{Closed}, however, a shift in the small loop spanning residues 327-330 results in a lateral displacement of R329 without significant changes in its rotameric conformation. This new configuration, referred to as the "PUSH" configuration, is speculated to be coordinated with the overall domain rotation, the loop order-disorder transition, and the sliding of the NADP⁺ cofactor during the catalytic cycle (Fig. 1.11B).

Unlike in the PAMO structure, where the central His of the BVMO signature sequence was solvent exposed, H166 was involved in hydrogen bonding interactions with G381 in both of the *Rm*CHMO structures. This residue is in one of the interdomain linkers between the NADPH-binding domain and the FAD-binding domain, and this linker forms contacts with the NADP⁺ cofactor. It appears that this interaction is likely important in coordinating the complex molecular movements that occur once NADP⁺ is bound (Fig. 1.12).



1.4.3 Other Baeyer-Villiger Monooxygenases

Following the publication of the first three crystal structure of BVMOs, a few additional papers describing BVMO structures were released. A second publication, describing eight new PAMO structures, was released in 2011 (Orru *et al.*, 2011). By changing the crystallization conditions used to obtain the first PAMO structure to rely on PEG instead of ammonium sulfate, it was possible to obtain structures of PAMO in the presence of NADP⁺. These structures revealed domain rotations that were similar to what was observed in the CHMO crystal structures, and also that the large loop could fold in and contact NADP⁺ in the ligand-bound structure. This last point was notable, as the adoption of a similar loop conformation in two homologous enzymes indicates that this conformation is unlikely to be a crystallization artifact.

It was possible to chemically reduce the FAD of the PAMO crystals by soaking them in sodium dithionite. They also found that X-ray irradiation was capable of reducing FAD, though warming of the crystal to room temperature is required for the conformational changes to be visible in the crystal structure. While the reduced enzyme structure did not exhibit large scale conformational changes, R337 moves to form an interaction with the negatively charged reduced flavin, as does the carboxamide moiety of NADP⁺. In addition, structures of the enzyme bound to MES, a weak inhibitor of PAMO, were also obtained. These structures place this molecule close to the location of the small pocket observed in

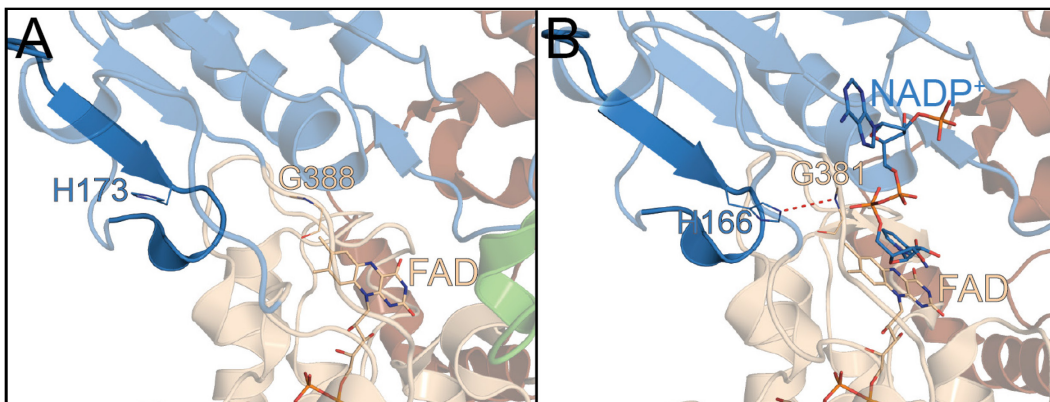


Fig. 1.12: The BVMO signature motif (non-transparent blue) in NADP⁺-free PAMO (A, PDB ID 1W4X) and NADP⁺-bound CHMO_{Open} (B, PDB ID 3GWF). The FAD-binding domain is shown in wheat, the NADPH-binding domain in blue, and the helical domain in brown. In the NADP⁺-bound structure (panel B), the central His links to the FAD-binding domain, while in the NADP⁺-free structure (panel A), it is solvent exposed. This trend is maintained in all available BVMO structures, including CHMO, PAMO, and OTEMO.

the CHMO_{Closed} structure. The authors speculated that this inhibitor is taking the place of the substrate. A "back door" channel to the exterior of the enzyme that is distinct from the large cleft in which NADP⁺ binds was identified as a potential route of entry for the inhibitor (and therefore the substrate) into the active site while NADP⁺ remains bound to the enzyme. They also examined the structures of three mutants: R337K, D66A, and M446G. The first removes the strictly conserved R337. This structure revealed that while the Lys residue can adopt similar conformations as the Arg, the presence of a primary amino group instead of a guanidinium group precludes the possibility of participating in multiple simultaneous interactions as required for catalysis. The D66A mutant demonstrated that this residue plays a non-essential role in correctly orienting NADPH for flavin reduction. Finally, the M446G mutant, which alters the substrate specificity of PAMO, creates more room around the MES inhibitor.

A series of six crystal structures of 2-Oxo- Δ^3 -4,5,5-Trimethylcyclopentenylacetyl-Coenzyme A Monooxygenase (OTEMO) was also published (Leisch *et al.*, 2012). This included two FAD-bound structures and four FAD- and NADP⁺-bound structures in four distinct crystal forms. OTEMO is distinct from CHMO and PAMO in that it is a dimer, and also targets a substrate linked to the very large coenzyme A. The most striking difference

between this structure and CHMO and PAMO is the behaviour of the large loop. In five of the six structures, the loop is completely ordered in an open β -hairpin conformation that sticks out and away from the enzyme. This structure is likely at least partly influenced by crystal contacts. In the last structure, part of the loop folds in and contacts the pyridine ribose of NADP^+ through W501, as seen in CHMO and PAMO. In this structure, however, the loop is partially disordered. This unusual behaviour of the loop, coupled with an apparent increase in the flexibility of the enzyme in the vicinity of the active site, could be due to the large coenzyme A moiety which needs to protrude from the active site. Also notable was the absence of a "back door" channel observed in PAMO, especially given the large size of coenzyme A.

Another BVMO had its structure solved in 2009 (Beam *et al.*, 2009). This enzyme, MtmOIV, is quite different from the family of "classical" BVMOs whose prototype is *AcCHMO*. Unlike most BVMOs, which are thought to be part of a catabolic pathway, MtmOIV is involved in the biosynthesis of the anticancer antibiotic mithramycin (Beam *et al.*, 2009). This enzyme, which is considered to be part of the GR₂ subfamily of FAD-dependent enzymes, is more closely related, both in structure and sequence, to hydroxylases PgaE and CabE. Correspondingly, it is structurally and mechanistically distinct from the classical BVMOs (Beam *et al.*, 2009). It was notable, however, that it also employs a catalytic Arg residue, and is expected to use a somewhat similar chemical mechanism (Beam *et al.*, 2009). The structural details from this enzyme are unlikely to be directly relevant to understanding the mechanism or substrate profile of the classical BVMOs like CHMO and PAMO.

1.5 Research Objectives

The importance of BVMOs as potentially useful biocatalysts is clear. Unfortunately, the lack of a crystal structure of any BVMO that shows the enzyme in complex with its substrate or product hampers attempts to do rational protein engineering. Furthermore, the complex domain motions that are apparent from the available crystal structures create additional hurdles in understanding the enzyme mechanism. Without a better understanding of the broad substrate

spectrum coupled with the high regio- and enantiospecificity, as well as the complex protein dynamics required for the multi-step enzyme mechanism, protein engineering efforts will remain limited. The present work aims to address these issues, and can be divided into three sections:

- A substrate-bound crystal structure of CHMO explains the broad substrate spectrum of the BVMOs. Specifically, the enzyme is captured with a unique NADP⁺ conformation that allows catalysis to occur in a pocket large enough to accommodate the wide range of substrates that can be oxygenated by CHMO. This work is described in **Chapter 2**.
- Two product-bound crystal structures of CHMO highlight a three-step binding mechanism for BVMOs. They also show the product in a position that would determine the regio- and enantiospecificity of the enzyme. This work is described in **Chapter 3**.
- Solution studies, using small-angle X-ray scattering and nuclear magnetic resonance, allow probing of the important role of protein dynamics in the BVMO catalytic mechanism. This work is described in **Chapter 4**.

Chapter 2: The Substrate-Bound Crystal Structure of a Baeyer–Villiger Monooxygenase Exhibits a Criegee-like Conformation

2.1 Preface

The interest in the BVMOs is linked to their potential use as biocatalysts. While the enzyme mechanism is well understood, structural biology can provide unique information that can pave the way for rational protein engineering. The crystal structures reported up to now have been able to describe the overall structure of the protein, provided some insight into the flexibility of this family of proteins, and allowed some initial speculation as to possible substrate binding modes. In spite of this, questions of how the substrate could bind after the binding of the NADP(H) cofactor, as well as the high degree of stereospecificity and broad substrate specificity remain unanswered.

This chapter describes the first high resolution crystal structure of a BVMO with a true substrate or product bound. This structure can be used to rationalize the broad substrate specificity of the enzyme, and can also explain how the enzyme can bind the substrate after binding of NADPH. This structure will increase our understanding of the structural mechanism of the BVMOs to complement the well-understood chemical mechanism.

2.2 Introduction

The century-old Baeyer–Villiger (BV) reaction is the oxidative cleavage of a carbon–carbon bond adjacent to a carbonyl; this process transforms ketones to esters and cyclic ketones to lactones using peracids (Baeyer & Villiger, 1899; Baeyer & Villiger, 1900). While this reaction has proven to be very useful to synthetic chemists, it has a number of shortcomings that include the use of costly and hazardous reagents or chlorinated solvents while producing more waste than product for lack of functional group selectivity and enantioselectivity. Although metal-catalyzed enantioselective BV reactions have also been developed, both cost and contamination of the products with metals may present a serious problem (ten Brink *et al.*, 2004). For a variety of reasons, Baeyer–Villiger monooxygenases (BVMOs), a family of bacterial flavoproteins that catalyze the

BV reaction (Kayser, 2009; Alphand & Wohlgemuth, 2010; de Gonzalo *et al.*, 2010; Lau *et al.*, 2010; Torres Pazmiño *et al.*, 2010; Leisch *et al.*, 2011), have garnered much attention. Most BVMOs are involved in the second step of biodegradation of C5 to C15 alicyclic alcohols such as cyclopentanol, cyclohexanol, and cyclopentadecanol, providing a carbon source for the organisms via the formation of Krebs cycle intermediates (Donoghue & Trudgill, 1975; Mirza *et al.*, 2009; Leisch *et al.*, 2011). These enzymes can function in water, and require NADPH as a cofactor and molecular oxygen as the oxidative reactant. While one oxygen atom is introduced adjacent to the carbonyl group of a ketone substrate, the other atom is reduced to water as a by-product (Donoghue *et al.*, 1976; Alphand & Wohlgemuth, 2010; de Gonzalo *et al.*, 2010). These green characteristics have made them intriguing targets for use as biocatalysts since the prototypical BVMO, cyclohexanone monooxygenase (CHMO) from *Acinetobacter* sp. NCIMB 9871 (*Ac*CHMO), was first characterized nearly 40 years ago (Donoghue & Trudgill, 1975; Donoghue *et al.*, 1976). Their frequently exquisite regio- and enantiospecificity, combined with their broad substrate profiles that include substituted and bicyclic ketones, make them of particular interest to the pharmaceutical industry (Kayser, 2009; Alphand & Wohlgemuth, 2010; de Gonzalo *et al.*, 2010; Lau *et al.*, 2010; Torres Pazmiño *et al.*, 2010; Leisch *et al.*, 2011). In recent years, a multitude of homologues of these enzymes have been cloned and characterized (Kayser, 2009; Alphand & Wohlgemuth, 2010; de Gonzalo *et al.*, 2010; Lau *et al.*, 2010; Torres Pazmiño *et al.*, 2010; Leisch *et al.*, 2011), thereby expanding the range of potential substrates catalyzed by BVMOs.

While the BVMOs have been studied extensively, only recently has any structural data been obtained. The first BVMO to have its structure solved was phenylacetone monooxygenase (PAMO) from the thermophile *Thermobifida fissa* (Malito *et al.*, 2004). This structure revealed the overall fold of the enzyme and the location and conformation of the FAD prosthetic group, which remains bound throughout the catalytic cycle. Subsequently, we reported two structures of CHMO from *Rhodococcus* sp. HI-31 (*Rm*CHMO) in complex with both NADP⁺

and FAD (Mirza *et al.*, 2009). This enzyme is notable for being a close homologue and having a similar substrate profile to the canonical *Ac*CHMO while being substantially more stable. The two crystal structures, designated as CHMO_{Open} and CHMO_{Closed}, revealed that NADP⁺ can bind to the enzyme in two distinct conformations. The sliding of the NADP⁺ cofactor deeper into the protein in the CHMO_{Closed} structure was revealed to be coupled with a rotation of the NADPH-binding domain so as to create a well-defined substrate binding pocket. More recently, a series of structures of PAMO with NADP⁺ and a weak inhibitor, 2-(*N*-morpholino)ethanesulfonic acid (MES), have corroborated these findings and provided insight into the conformation of the reduced flavoenzyme (Orri *et al.*, 2011). Intriguingly, the PAMO structures identified a funnel-shaped cavity that may provide an alternative entrance for the substrate to gain access to the active site. Another recent paper describing NADP⁺-bound and NADP⁺-free crystal structures of 2-oxo- Δ^3 -4,5,5-trimethylcyclopentenyl acetyl-Coenzyme A 1,2-monooxygenase (OTEMO), which is notable for being a dimeric BVMO with a large substrate, also confirmed the dynamic nature of the BVMO family of proteins (Leisch *et al.*, 2012). It should be noted that in addition, a crystal structure of an atypical BVMO, mithramycin monooxygenase (MtmOIV) has been obtained (Beam *et al.*, 2009). This enzyme has high sequence and structural similarity to a FAD-dependent hydroxylase of the glucocorticoid receptor (GR₂) subfamily. This makes it a distant relative of the prototypical BVMOs and precludes its detailed comparison with CHMO or PAMO.

Structural studies of BVMOs have provided much insight into these enzymes' catalytic mechanism; however, they have also raised questions. Notably, the CHMO_{Closed} structure reveals that for the majority of CHMO's substrates, which are larger than cyclohexanone, there is actually insufficient room in the active site for the formation of the critical Criegee intermediate, the adduct of the flavin peroxide and the ketone substrate that is required for catalysis. Models of the Criegee intermediate for the cyclohexanone substrate, one of CHMO's smallest substrates, suggests that this intermediate can be achieved in the Closed conformation (Mirza *et al.*, 2009; Polyak *et al.*, 2012). In

contrast, the Criegee intermediates of larger substrates, such as substituted or bicyclic ketones on which CHMO readily acts, would result in severe steric clashes with one or more of NADP⁺, L146, F279, and F434 in the Closed conformation. It is clear that in order for a Criegee intermediate to form, a significant reorganization of the active site is required.

Here, we present the first crystal structure of a prototypical BVMO, represented by *Rm*CHMO, bound to its substrate, cyclohexanone, as well as NADP⁺. This structure reveals a major rotation in the NADP⁺ cofactor, allowing CHMO to adopt a catalytically relevant conformation. The rotation permits the substrate to be positioned above the FAD, thus revealing a Criegee intermediate-like conformation, a snapshot of the most important stage of the catalytic mechanism. It also highlights the role of NADP⁺ in organizing the active site and provides a structural explanation for the broad substrate specificity of the enzyme, which is necessary for a complete understanding of the catalytic mechanism.

2.3 Materials and Methods

2.3.1 Subcloning the *chnB1* gene into the pJW234 expression vector

The *chnB1* gene was amplified using *Pwo* DNA polymerase (Roche) from the pSDRmchnB1 plasmid (Mirza *et al.*, 2009). PCR primers with *Nsi*I and *Eco*RI restriction sites were used to amplify the gene (restriction sites are underlined), as shown in Table 2.1. The desired band was purified using the QIAquick Gel Extraction Kit (Qiagen) and subcloned into the pPCR-Script Cam vector (Stratagene) using the PCR-Script Cam Cloning Kit (Stratagene). Plasmid DNA from positive clones was prepared using the QIAprep Spin MiniPrep kit (Qiagen), and then digested using *Nsi*I and *Eco*RI (New England BioLabs). The pJW234 expression vector (obtained from Dr. Mirek Cygler, University Saskatchewan, Saskatoon, Canada) was digested with the same enzymes. The desired bands from both digests were purified using gel extraction (MO BIO Laboratories UltraClean 15 DNA Purification Kit). The *chnB1* insert was ligated into the pJW234 vector, and this plasmid was used to transform *E. coli* DH5 α cells. Plasmid DNA from positive clones was prepared using the QIAprep Spin

MiniPrep kit (Qiagen), and DNA sequencing was used to confirm the sequence (McGill University and Genome Quebec Innovation Centre Sequencing Service, Montreal, Canada). The resultant plasmid is called His₈-TEV-ChnB1.

Table 2.1: PCR Primers for subcloning and site-directed mutagenesis. Restriction sites are in bold and mutated bases are underlined.

Construct	Primer	Sequence
His ₈ -TEV-ChnB1	Sense	5'-CAGGAAACAAT GCATATGACCGCACA G-3'
His ₈ -TEV-ChnB1	Antisense	5'-CAGTACAGA ATTCTCAGACCGTGACC ATCTC-3'
His ₈ -TEV-ChnB1-L145N	Sense	5'-GTCGTCAACGCCGTGGGCA ATCTCTCC GCGATCAACTTC-3'
His ₈ -TEV-ChnB1-L145N	Antisense	5'-GAAGTTGATCGCGGAGAG ATTGCCCA CGGCGTTGACGAC-3'
His ₈ -TEV-ChnB1-L145D	Sense	5'-GTCGTCAACGCCGTGGG AGATCTCTCC GCGATCAACTTC-3'
His ₈ -TEV-ChnB1-L145D	Antisense	5'-GAAGTTGATCGCGGAGAG ATCTCCCAC GGCGTTGACGAC-3'
His ₈ -TEV-ChnB1-F507Y	Sense	5'-GCCGAGCGTACTGT ATTACCTGGGCGG CCTG-3'
His ₈ -TEV-ChnB1-F507Y	Antisense	5'-CAGGCCGCCAGGTA ATACAGTACGCT CGGC-3'

2.3.2 Creation of the L145N, L145D, and F507Y Mutants

The His₈-TEV-ChnB1 plasmid was amplified using the protocol from the Agilent QuikChange II Site-Directed Mutagenesis Kit. The primers used are shown in Table 2.1, with the altered bases underlined. Plasmid DNA from positive clones was prepared and sequenced as described above (section 2.3.1).

2.3.3 Expression and Purification of CHMO

The expression and purification protocols for wild-type CHMO and its mutants are identical. *E. coli* Rosetta(DE3) (Novagen) cells were transformed with His₈-TEV-ChnB1 plasmid and the strain was grown in 50 mL of LB supplemented with 100 µg/mL of ampicillin and 30 µg/mL of chloramphenicol and grown for 16-18 hours at 37°C. This was then subcultured into a fresh 50 mL culture and grown for an additional 6 hours. After another subculturing in fresh 1

L culture and growth at 37°C to an OD600 of 0.5-0.6, the cells were induced with 1 mM IPTG and allowed to grow at 15°C for 18-20 hours.

All subsequent steps were performed at 4°C. The cells were harvested using centrifugation (6000×g for 10 minutes). The pellets were resuspended in 50 mM sodium phosphate, 0.3 M sodium chloride, 10 mM imidazole, pH 8.0, and a Complete Mini (EDTA-free) protease inhibitor tablet (Roche) was added. The cells were lysed by sonication, and the crude extract was clarified using centrifugation (21,000×g for 30-60 minutes). The lysate was then filtered. Purification procedures were performed on an Akta Explorer (GE Healthcare). The lysate was loaded onto a 5 mL Ni-NTA Superflow Cartridge (Qiagen) equilibrated with 50 mM sodium phosphate, 0.3 M sodium chloride, 10 mM imidazole, pH 8.0. The protein was eluted using a linear gradient to 50 mM sodium phosphate, 0.3 M sodium chloride, 300 mM imidazole, pH 8.0. Fractions containing CHMO were pooled. For samples to be used for enzyme kinetics or NMR, the protein was dialysed against 50 mM sodium phosphate pH 8.0 and then concentrated to approximately 30 mg/mL. For samples to be used for crystallization, the pooled sample was digested using His-tagged tobacco etch virus (TEV) protease S219V (purified in house, plasmid obtained from Dr. Mirek Cygler) while dialysing the sample against 50 mM sodium phosphate, 150 mM sodium chloride, 5 mM sodium citrate, 1 mM DTT, pH 8.0 overnight. The sample was then re-loaded onto the Ni-NTA column to remove undigested CHMO and the TEV protease. The unbound protein was pooled and concentrated (Millipore Centrprep concentrators) to about 5 mL. The sample was supplemented with FAD and allowed to stand for about 4 hours. The sample was then loaded onto a HiLoad Superdex 75 26/60 prep grade column (GE Healthcare) equilibrated with 50 mM sodium phosphate, 150 mM sodium chloride, pH 8.0. Fractions containing CHMO were pooled and concentrated. Purity was assessed using SDS-PAGE throughout the purification. The buffer was exchanged to 50 mM Tris pH 8.0 by running the sample on a PD-10 desalting column (GE Healthcare). The sample was concentrated to 5 mg/mL and supplemented with a 5-fold molar excess of FAD and NADP⁺.

2.3.4 Crystal Structure Determination

Crystals were obtained using the hanging drop vapor diffusion method. Wild-type CHMO was prepared at 5 mg/mL in 50 mM Tris pH 8.0 and supplemented with a 5× molar excess of FAD and NADP⁺. In brief, 2 µL of protein solution was mixed with 2 µL of reservoir solution on an 18 mm siliconized coverslip (Hampton Research). The reservoir solution consisted of 0.1 M imidazole, pH 8.0, 0.2% TMOS, 20% PEG 3350, and 0.1 M cyclohexanone. This drop was suspended over a 1 mL reservoir in a 24-well ComboPlate (Greiner Bio-One), and the plate was incubated at 4°C. Crystals suitable for diffraction studies grew in about one week.

Data were collected under standard cryogenic conditions on a Rigaku MicroMax-007HF generator equipped with VariMax HF optics and a Saturn 944+ CCD detector. The data were processed using the HKL2000 suite of programs (Table 2.2) (Otwinowski & Minor, 1997). The structure was subsequently solved using Phaser (McCoy *et al.*, 2007), employing CHMO_{Open} (PDB ID 3GWF) as the search model (Mirza *et al.*, 2009). It should be noted that this crystal form is distinct from both the CHMO_{Open} and CHMO_{Closed} crystal forms. The model was subjected to multiple rounds of positional and B-factor refinement using Refmac (Murshudov *et al.*, 1997). Manual model building was performed regularly during refinement using Coot (Emsley & Cowtan, 2004). The coordinates and structure factors have been deposited in the Protein Data Bank as PDB ID 3UCL. Structural figures were prepared using PyMOL (Schrödinger LLC).

Table 2.2: Data Collection and Refinement Statistics for the CHMO_{Rotated} Crystal Structure (PDB ID 3UCL).

Data collection statistics	
Space Group	P2 ₁ 2 ₁ 2 ₁
a, b, c (Å)	55.5, 67.1, 131.4
Resolution Range (Å)	30.6 - 2.4 (2.5-2.4)
Completeness (%)	97.9 (92.7)
Redundancy	11.2 (8.3)
Reflections with I/σ < 2 (%)	8.7 (47.2)
Refinement statistics	
Total number of reflections (reflections in R _{free} set)	20,214 (2061)
R _{factor} (%) (Work + Free/Free)	19.8 / 26.4

Number of atoms	4215
Protein	3986
Water	121
Cofactors and Substrate	108
RMSD	
Bond length (Å)	0.017
Bond angle (°)	1.669
Ramachandran plot	505 (100%)
Residues in favoured positions	492 (97.4%)
Residues in allowed positions	10 (2.0%)
Residues in disallowed positions	3 (0.6%)

2.3.5 Saturation Transfer Difference NMR Spectroscopy

The buffer used for dialysis of the protein was used to dissolve NADP⁺ at 5 mM. Samples of the buffer and NADP⁺ solution were placed under vacuum until dry, and then reconstituted in an equal volume of D₂O. A 600 μ L sample of 75 μ M protein and 1.5 mM NADP⁺ was prepared. Samples of the protein without its ligand were also prepared as a control to confirm that all of the STD signals were derived from protein–ligand interactions.

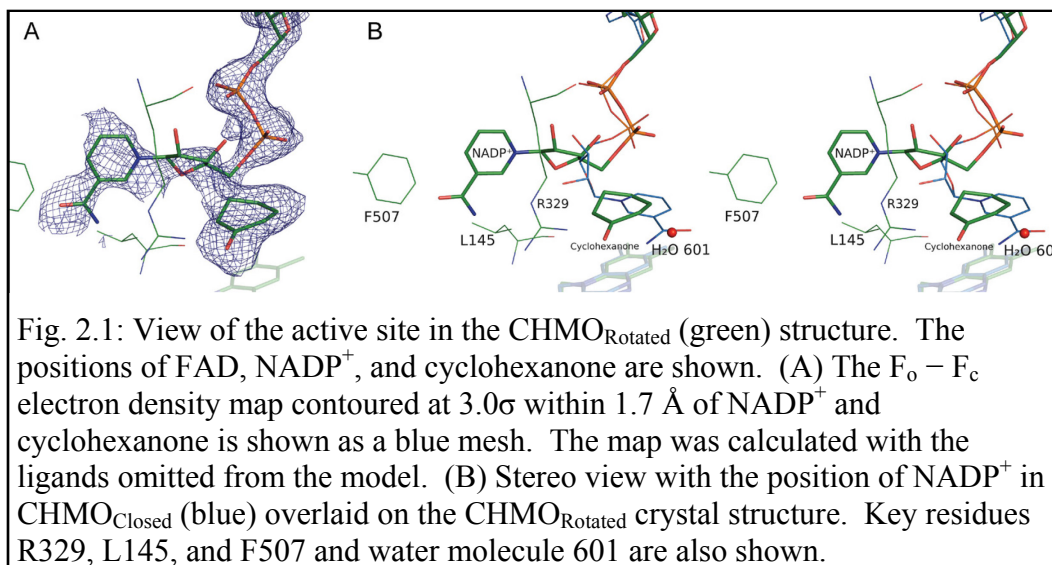
All NMR experiments were performed at 15°C on a Varian INOVA 500 MHz spectrometer equipped with a triple resonance HCN cold probe. A 1D saturation transfer difference pulse sequence with internal subtraction via phase cycling was employed to record difference spectra (Mayer & Meyer, 2001). Residual HDO signal was removed using a double pulsed-field gradient spin echo. On-resonance irradiation of the protein was performed at -0.5 ppm, with off-resonance irradiation at 36.5 ppm. Reference spectra were recorded using the same pulse sequence with saturation pulses applied at 36.5 ppm, and no internal subtraction. The irradiation power was $(\gamma/2\pi)B_1 = 80$ Hz applied through a train of 50 ms Gaussian pulses with a 100 μ s delay between the pulses. Experiments were recorded with saturation times of 2 s. The relaxation delay was set to 3 s. A 30 ms spin-lock pulse with a strength of $(\gamma/2\pi)B_1 = 3000$ Hz was used to eliminate background protein signals. Spectra were recorded with 32k points, a sweep width of 16 ppm, and 128 transients. The experiments were repeated the next day in order to ensure that the samples were stable. The intensity of each

peak in the STD spectrum was divided by the intensity in the reference spectrum. In order to determine the percent STD effect, this ratio was divided by the same ratio for the largest peak in the STD spectra, the A2 proton of the adenine ring. In those cases where multiplets were observed for a single proton, the percent STD effect was averaged for each peak in the multiplet. The data from the two replicates were averaged to obtain a final value for each proton.

2.3.6 Enzyme Kinetics

All enzyme assays were performed at 22°C using a Cary 50 Bio UV–Visible Spectrophotometer equipped with a Peltier-thermostatted cell. The enzyme, NADPH, and cyclohexanone were dissolved at the appropriate concentration in 50 mM sodium phosphate, pH 8.0. A 500 μ L reaction volume was used. To determine the K_M and k_{cat} for NADPH, a series of assays with 100 μ M cyclohexanone ($>40\times K_{M(cyclohexanone)}$) and between 3 μ M and 200 μ M of NADPH were used. These experiments were repeated with varying concentrations of $NADP^+$ to determine the K_i of $NADP^+$. To determine the K_M and k_{cat} for cyclohexanone, a series of assays with 50 μ M NADPH ($>8\times K_{M(NADPH)}$) and between 0.4 μ M and 40 μ M cyclohexanone were used. A higher concentration of NADPH could not be used due to substrate inhibition. In all cases, the NADPH, 10 μ L of an appropriate concentration of enzyme, and the substrate were mixed together to start the reaction. The decrease in absorbance at 340 nm was monitored to determine the initial rate of the oxidation reaction. The initial rates were plotted against the substrate (NADPH or cyclohexanone) concentration for each reaction series, and nonlinear regression was performed using SigmaPlot (Systat Software, Inc.) to determine the K_M and k_{cat} .

A second set of experiments were performed to determine the uncoupling ratio, defined as the rate of NADPH oxidation in the presence of cyclohexanone (apparent BV activity) divided by the rate of NADPH oxidation in the absence of cyclohexanone (NADPH oxidase activity). Using saturating concentrations of both NADPH ($>6\times K_{M(NADPH)}$) and cyclohexanone ($>40\times K_{M(cyclohexanone)}$), the decrease in absorbance at 340 nm was assessed both before and after the addition of cyclohexanone. Three different NADPH concentrations were used (40–80



μM), all of which saturated the enzyme, and the experiments were conducted in quadruplets.

2.4 Results

2.4.1 Crystal Structure of Substrate-Bound CHMO

The crystal structure of *Rm*CHMO in complex with FAD, NADP⁺, and cyclohexanone was obtained to a resolution of 2.4 Å (Table 2.2). The polypeptide backbone could be modeled completely from residues 5–534, with the exception of two unstructured loop regions from residues 146–149 and 489–503. The first loop region corresponds to a long linker between the FAD- and NADPH-binding domains, while the second corresponds to the large loop that is unstructured in the CHMO_{Open} structure, but folds in to contact NADP⁺ in the CHMO_{Closed} structure. Density consistent with FAD, NADP⁺, and cyclohexanone was also visible (Fig. 2.1A). To differentiate it from the CHMO_{Open} and CHMO_{Closed} structures that were previously reported, we will refer to this structure as CHMO_{Rotated}.

In general, the CHMO_{Rotated} structure most closely resembles the CHMO_{Open} structure (RMSD of 0.56 Å between backbone atoms present in both CHMO_{Open} and CHMO_{Rotated} structures when the FAD-binding domains are aligned, vs. 1.31 Å for CHMO_{Closed}). When aligning the two structures based on their FAD-binding domains, the FAD-binding and helical domains show very little structural divergence. The NADPH-binding domain shows a slight rotation, though this is significantly less pronounced than the substantial domain

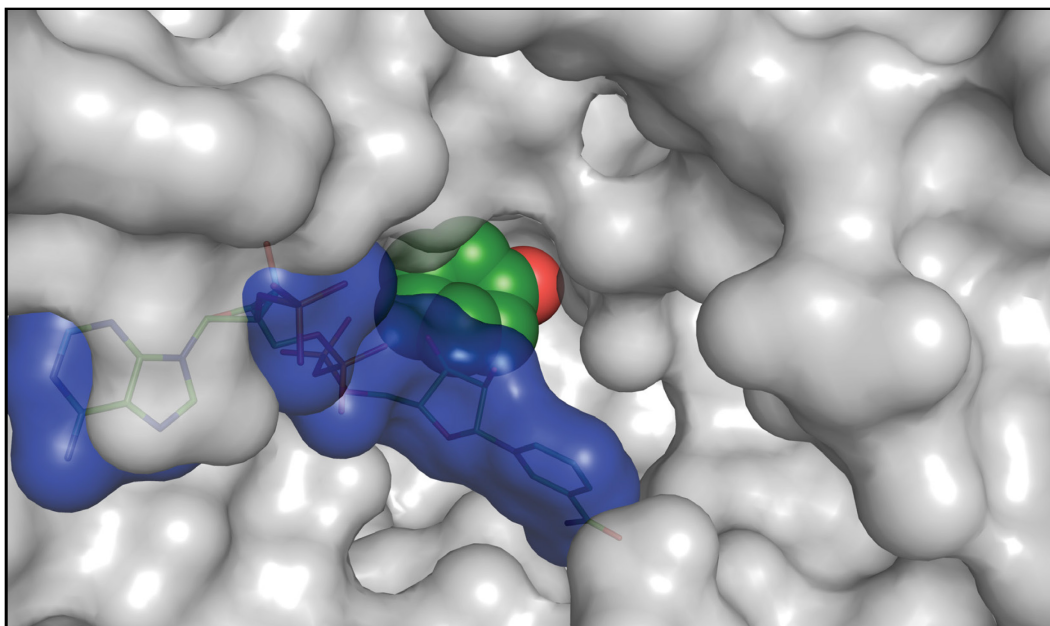


Fig. 2.2: NADP^+ blocks cyclohexanone from diffusing into the bulk solvent in the Rotated conformation. The protein surface is shown in gray. NADP^+ is shown in stick representation and its surface is outlined in blue. Cyclohexanone is shown using green space-filling representation.

movements seen when comparing the $\text{CHMO}_{\text{Open}}$ and $\text{CHMO}_{\text{Closed}}$ structures.

This rotation allows for a subtle shift in the position of the adenine portion of NADP^+ , which may in turn trigger the rotation of its nicotinamide moiety.

2.4.2 Rotation of the NADP^+ Cofactor

One of the most significant features in the crystal structure is the large rotation of the nicotinamide phosphoribose of NADP^+ when compared to $\text{CHMO}_{\text{Open}}$ and $\text{CHMO}_{\text{Closed}}$ (Fig. 2.1B). It is for this reason that we refer to it as the $\text{CHMO}_{\text{Rotated}}$ structure. This rotation serves to displace the nicotinamide head away from its position above the isoalloxazine ring system of FAD and into the large, primarily hydrophobic cavity that is seen in the $\text{CHMO}_{\text{Open}}$ structure. This movement has two notable effects. By rotating away from the FAD ring system, space for the substrate to enter into the catalytic position is created. Second, the nicotinamide head serves to partially block the active site region from the bulk solvent, which prevents the substrate from diffusing back into solution while the protein is in the Rotated conformation (Fig. 2.2). It is notable that the nicotinamide moiety forms relatively few strong interactions in this conformation.

The new conformation of NADP^+ allows the formation of intramolecular hydrogen bonds between the 3' hydroxyl of the nicotinamide ribose and both the nicotinamide and the adenine phosphates. This intramolecular interaction stabilizes this conformation of the NADP^+ and may be a driving force in the adoption of this conformation.

2.4.3 The Binding of the Substrate, Cyclohexanone

Electron density corresponding in size and shape to what is expected for cyclohexanone was observed above the isoalloxazine ring system of FAD; therefore, we proceeded to model the substrate at this position (Fig. 2.1A). The carbonyl oxygen of cyclohexanone appears to be oriented through interactions with D59 and R329. Both of these residues have been previously implicated in controlling the position of NADP^+ (Malito *et al.*, 2004; Torres Pazmiño *et al.*, 2008a; Mirza *et al.*, 2009), as well as in the positioning of a weak inhibitor of PAMO (Orri *et al.*, 2011). A nearby water molecule (H_2O 601) also forms a weak hydrogen bond with the carbonyl oxygen, linking it to Q192, the backbone atoms of residues L57 and D59, and a larger solvent network.

It is known that the formation of the Criegee intermediate, the key catalytic step in the BVMO reaction mechanism, involves a nucleophilic attack on the carbonyl carbon of cyclohexanone by the peroxyanion intermediate, in which the peroxyanion is covalently linked to the C4X carbon of FAD. The distance between C4X and the carbonyl carbon is 3.9 Å in this crystal structure. With the FAD-linked oxygen of the peroxide group expected to occupy an axial position relative to the FAD ring system, if the anionic oxygen were to point toward cyclohexanone, the position occupied by cyclohexanone in this structure would correspond approximately to the expected position for the nucleophilic attack and formation of the Criegee intermediate. A model of the Criegee intermediate that imposes antiperiplanar geometry required for migration of the carbon to form the lactone (Snowden *et al.*, 2004) shows that cyclohexanone nearly coincides with its position in the model of the intermediate (Fig. 2.3). A shift of ~ 0.5 Å is required to superimpose the carbonyl carbon of the cyclohexanone molecule and

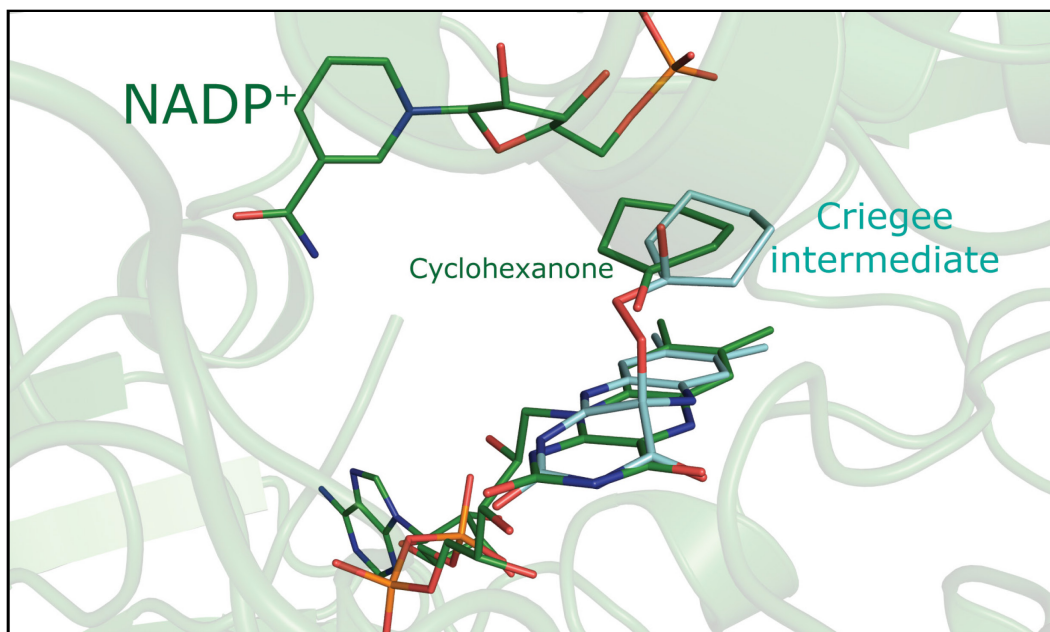
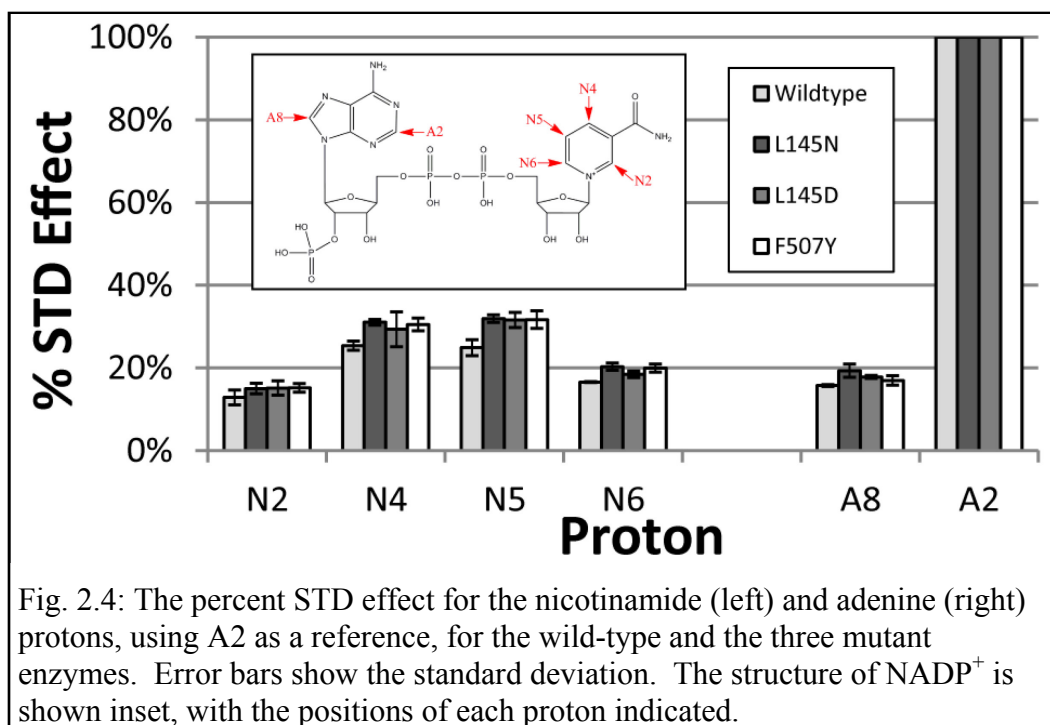


Fig. 2.3: The CHMO_{Rotated} crystal structure (green) is shown superimposed on a model of the Criegee intermediate (cyan). The model was produced using molecular dynamics followed by energy minimization in Chem3D, while enforcing the antiperiplanar geometry required for migration of the carbon.

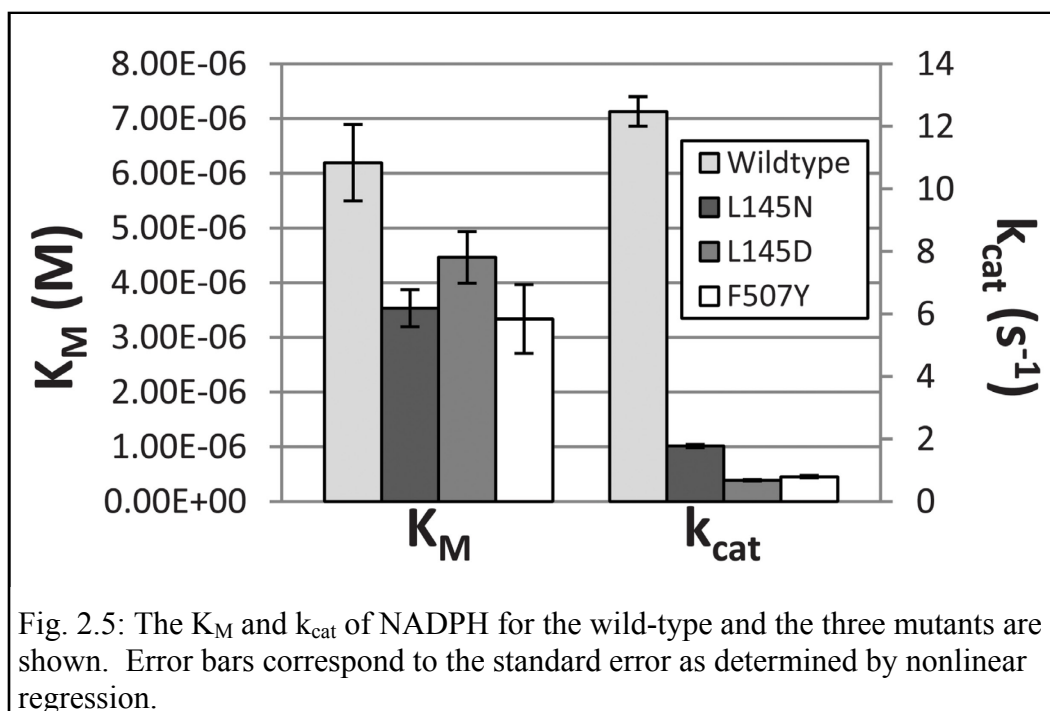
the intermediate, after which a $\sim 30^\circ$ rotation is sufficient to make the molecules overlap.

2.4.4 Structure–Function Studies Probing the Rotated Conformation.

To probe the relevance of the observed Rotated conformation for catalysis, we designed three mutants that were predicted to specifically perturb nicotinamide cofactor binding in the Rotated conformation, while having no impact on the cofactor in either the Open or Closed conformations. The three mutants, L145N, L145D, and F507Y, are anticipated to stabilize the catalytic conformation by forming a polar contact with the carboxamide group of NADP⁺ in the Rotated conformation. In contrast, in the CHMO_{Open} and CHMO_{Closed} structures, these residues should not interact directly with the cofactor, with the nearest distance in either structure being more than 5 Å. Given that many of the residues implicated in NADP⁺-binding in the CHMO_{Rotated} structure are also involved in substrate specificity, the rational design of destabilizing mutations specific for the Rotated conformation was not feasible.



To assess the degree of perturbation in the NADP⁺ binding mode caused by these mutations, transferred nuclear Overhauser effect (Tr-NOE) and saturation transfer difference (STD) NMR experiments were performed. No major differences were observed between the Tr-NOESY spectra of NADP⁺ in the presence of wild-type CHMO or any of the three mutants, as only some slight changes of NOE intensities and appearance and disappearance of the weakest NOEs were observed. This suggests that the overall set of conformations available to NADP⁺ in the mutants is not significantly perturbed. In the STD experiments, the aim was to assess if there was a difference in the relative enhancement of the STD effect of the pyridine protons as compared to the adenine protons when the mutants were compared to the wild-type protein. The differences observed were subtle, and preclude a discussion of the effects of the mutations on any individual proton; however, when the STD effect of the four protons on the pyridine ring (N2, N4, N5, and N6) was compared to the STD effect for the two adenine ring protons (A2 and A8, where A2 is used as the reference for the STD effect), a trend could be observed (Fig. 2.4). The interactions between the protein and the pyridine protons appeared closer than the protein–adenine proton interactions in the mutants as compared to the wild-type



CHMO. This is consistent with a minor selective stabilization of the nicotinamide moiety binding mode in the Rotated conformation.

While the structural impact of the mutants was subtle, as assessed by NMR methods, the impact on the kinetic properties of CHMO was substantial (Fig. 2.5, see Table 2.3 for complete kinetic data). In comparing the K_M values for NADPH, the wild-type had the highest K_M , suggesting that all three mutants bind the cofactor more strongly during catalysis. It should be noted that all the K_M values were within an order of magnitude of each other. Furthermore, in terms of k_{cat} , all three mutants had substantially lower rates than the wild-type, with the k_{cat} dropping by at least a factor of 6 and by as much as a factor of 100. This effect was seen for the k_{cat} of both NADPH and cyclohexanone. This shows that, although the binding affinity of NADPH had been slightly improved, the mutations appear to have significantly slowed the reaction rate.

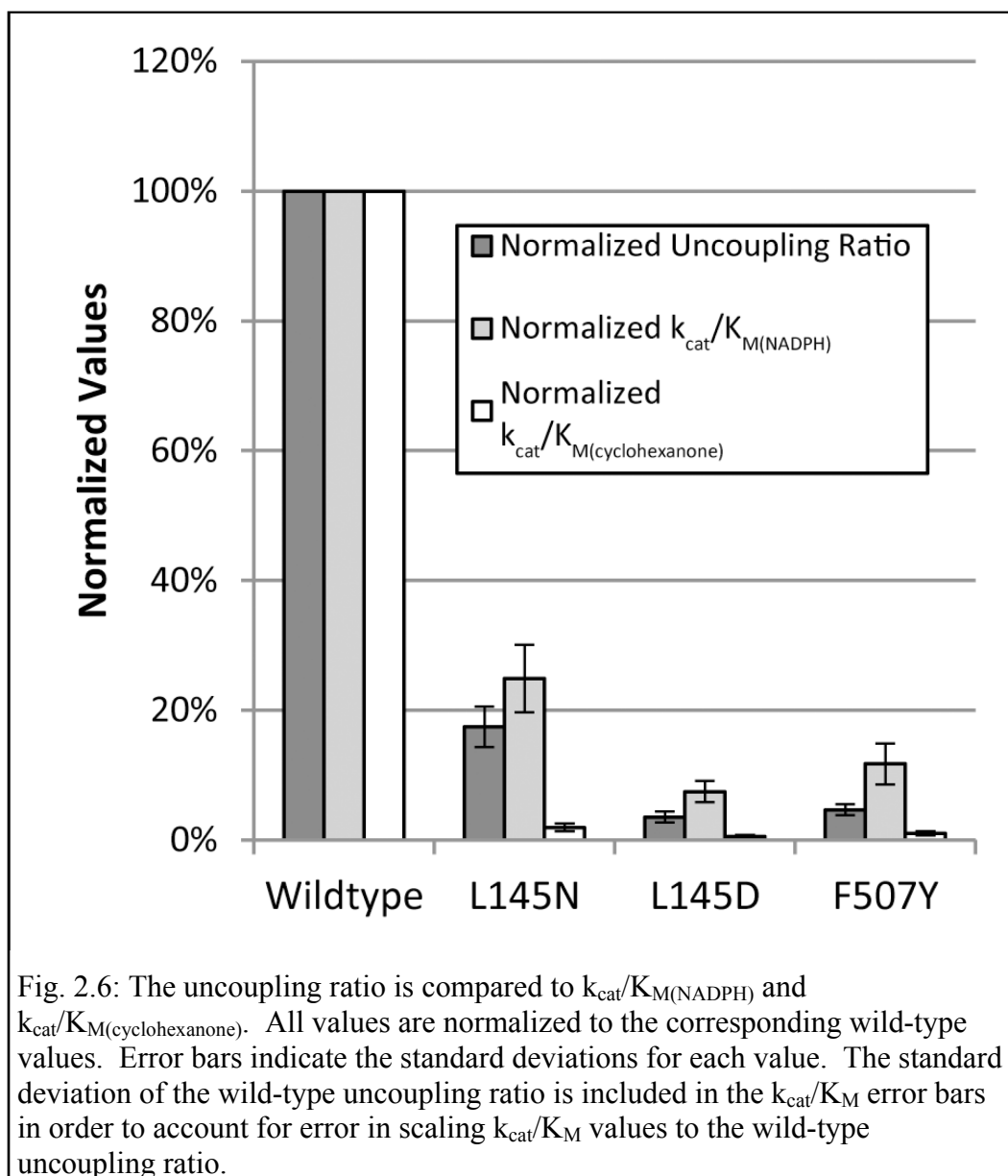
In order to estimate the affinity of $NADP^+$, the K_i values of the wild-type and the mutants were determined using $NADP^+$ as the inhibitor and varying NADPH as the substrate (Table 2.3). As expected, $NADP^+$ appeared to behave as a competitive inhibitor for CHMO. $NADP^+$ proved to be only a weak inhibitor of *Rm*CHMO, in contrast to *Ac*CHMO (Sheng *et al.*, 2001) and PAMO (Torres

Pazmiño *et al.*, 2008a), which have K_i values in the low micromolar range, but similar to 4-hydroxyacetophenone monooxygenase (HAPMO) (van den Heuvel *et al.*, 2005). In comparison to the *RmCHMO* wild-type enzyme, the L145N mutant is more sensitive to inhibition by NADP^+ . The L145D mutant also shows this same trend, but an accurate K_i could not be determined due to the low activity of this mutant. The F507Y mutant reveals that it is similar to the wild-type enzyme, in this respect.

Table 2.3: Enzyme kinetics data. Errors are the standard errors determined by regression, or standard deviations for the Uncoupling Ratio.

		Wild-type	L145N	L145D	F507Y
K_M (M)	NADPH	$(6.19 \pm 0.70) \times 10^{-6}$	$(3.53 \pm 0.34) \times 10^{-6}$	$(4.46 \pm 0.47) \times 10^{-6}$	$(3.34 \pm 0.63) \times 10^{-6}$
	Cyclo-hexanone	$(2.95 \pm 0.72) \times 10^{-7}$	$(2.46 \pm 0.26) \times 10^{-6}$	$(4.93 \pm 1.95) \times 10^{-7}$	$(1.14 \pm 0.11) \times 10^{-6}$
k_{cat} (s^{-1})	NADPH	12.47 ± 0.47	1.77 ± 0.05	0.67 ± 0.02	0.79 ± 0.04
	Cyclo-hexanone	8.86 ± 0.47	1.45 ± 0.05	0.080 ± 0.004	0.36 ± 0.01
k_{cat}/K_M ($\text{M}^{-1}\text{s}^{-1}$)	NADPH	$(2.01 \pm 0.24) \times 10^6$	$(5.02 \pm 0.51) \times 10^5$	$(1.51 \pm 0.17) \times 10^5$	$(2.37 \pm 0.46) \times 10^5$
	Cyclo-hexanone	$(3.00 \pm 0.75) \times 10^7$	$(5.89 \pm 0.66) \times 10^5$	$(1.63 \pm 0.65) \times 10^5$	$(3.16 \pm 0.31) \times 10^5$
Uncoupling Ratio		114 ± 16	20 ± 2	4 ± 1	5 ± 1
K_i (M)	NADP^+	$(1.65 \pm 0.19) \times 10^{-4}$	$(4.20 \pm 0.75) \times 10^{-5}$	N.D. ($< 10^{-4}$)	$(1.92 \pm 0.38) \times 10^{-4}$

The NADP^+ inhibition studies imply that the affinity of NADP^+ is much weaker than would have been suggested by the values for $K_{M(\text{NADPH})}$; however, as NADP^+ is known to stay bound to the enzyme throughout the catalytic mechanism (see Fig. 2.9), and readily co-crystallizes with *RmCHMO*, it must be concluded that this apparent weak affinity is at least partially caused by complex binding kinetics. Despite this puzzling aspect, the mutants display K_i values that are less or statistically equal to that of the wild-type enzyme, mirroring the trend in affinity seen by monitoring $K_{M(\text{NADPH})}$.



The uncoupling ratio is an indicator of the efficiency of the complete Baeyer–Villiger reaction (BV activity) relative to the efficiency at which NADPH would be used in the absence of cyclohexanone (NADPH oxidase activity). The wild-type enzyme exhibits the largest uncoupling ratio of 114. For the three mutants, this ratio is significantly reduced, ranging from 4 to 20. To assess how much of this change is caused by the decrease in BV activity and how much is due to changes in the amount of unproductive NADPH oxidase activity, the normalized uncoupling ratio was compared to the normalized $k_{cat}/K_M(NADPH)$ and $k_{cat}/K_M(cyclohexanone)$ values, using the wild-type values as 100% (Fig. 2.6). Note

that both k_{cat}/K_M values were determined using the apparent BV activity (which includes unproductive NADPH oxidase activity in the presence of cyclohexanone), not the uncoupled reaction (NADPH oxidase activity). For all three mutants, the uncoupling ratio was significantly greater than would be predicted by the normalized $k_{\text{cat}}/K_{M(\text{cyclohexanone})}$, and moderately smaller than predicted by the normalized $k_{\text{cat}}/K_{M(\text{NADPH})}$.

2.5 Discussion

2.5.1 A Third NADP⁺ Conformation

Previous structural studies of BVMOs have raised a critical question: how is the enzyme capable of effecting catalysis with a structure that has its active site blocked by the NADP⁺ cofactor? The NADP⁺ has been shown to be bound throughout the catalytic cycle (Ryerson *et al.*, 1982; Sheng *et al.*, 2001; Torres Pazmiño *et al.*, 2008a), presumably to stabilize the peroxyanion intermediate, precluding the possibility of dissociation of NADP⁺ prior to substrate binding. At the same time, much of the space required for the substrate to form the expected Criegee intermediate is blocked by NADP⁺ and other key residues. The rotation of the nicotinamide away from the FAD provides an explanation for how the substrate can enter the space close to the peroxyanion intermediate while NADP⁺ remains bound. In this conformation, the substrate sits in a putative catalytic position that is large enough to accommodate CHMO's larger substrates. While the structure resembles the Open conformation, the rotation of the NADP⁺ nicotinamide blocks off much of the space that is available in the CHMO_{Open} structure (Fig. 2.2), and a shifting of the key residue, R329, narrows the channel further to prevent the substrate from diffusing away from the peroxyanion intermediate. In this way, the substrate is "locked" into the active site pocket in the Rotated conformation. We propose that in order to avoid the rapid collapse of the peroxyanion intermediate, the substrate is bound in the previously reported Open and Closed conformations. The relatively weak interactions between the nicotinamide and the protein in the Rotated conformation would disfavor this conformation unless the substrate was able to replace some of the lost interactions. In other words, the enzyme would be locked in a conformation that

stabilizes the peroxyanion intermediate until the substrate is present in the substrate binding site observed in the CHMO_{Closed} structure. At that point, the enzyme can switch into the Rotated conformation, allowing catalysis to occur.

The NMR and kinetic data for wild-type and CHMO mutants are in agreement with this proposal. The three mutants were designed such that the altered residues would form favorable interactions specifically with the observed Rotated conformation of the nicotinamide cofactor while having no interactions with this cofactor in either the Open or Closed conformations of the enzyme. The NMR data supports that the design of the mutants was successful as the STD data for the mutants show that the pyridine moiety is bound with higher affinity relative to the adenine moiety. This is consistent with the stabilization of a minor conformational state for the pyridine moiety akin to that observed in the CHMO_{Rotated} structure. Kinetic studies reveal that the mutants display a moderate increase in the affinity of NADP⁺, as seen by the decrease in K_M . This observation strongly suggests that the Rotated conformation is an enzyme state that must occur sometime during the reaction cycle of CHMO. Moreover, the reaction rate is significantly decreased (6- to 100-fold). This change in k_{cat} for the mutants implies that the enzyme is spending longer than usual in one or perhaps more catalytically critical steps during the reaction cycle.

Finally, the uncoupling ratio data provide insight into which specific step in the reaction mechanism might be delayed. The uncoupling ratio represents the ratio between the enzyme catalyzing the full BV reaction (BV activity), and a short-circuited reaction in which NADPH is merely oxidized (NADPH oxidase activity). In the wild-type enzyme, the full BV reaction is dramatically favored, as indicated by a ratio of 114. The three mutants are all less efficient in favoring the BV reaction, as shown with their uncoupling ratios of 20 or less. This implies that the mutants are either defective in catalyzing the full BV reaction, enhanced in the short-circuited NADPH oxidation reaction, or both. The steady-state kinetics data indicate that the mutants do have compromised BV activity. To access whether the mutations also enhanced NADPH oxidase activity, we compared both the normalized $k_{cat}/K_{M(NADPH)}$ and the normalized

$k_{\text{cat}}/K_{\text{M(cyclohexanone)}}$ to the normalized uncoupling ratio (with wild-type values set to 100%). This analysis provides insights into how efficiently each substrate is used by the enzymes (see Fig. 2.6). Of course, the mutants are less efficient catalysts both with respect to NADPH and cyclohexanone. Intriguingly, using the drop in uncoupling ratios as a base for the extent of anticipated efficiency loss, the mutants are actually using NADPH more effectively than predicted. Conversely, the mutants use cyclohexanone as a substrate much less efficiently than would be anticipated. This observation implies that the mutations have differential impact on the BV and NADPH oxidase reactions. Given that the only unique step in the NADPH oxidase pathway, as compared to the BV activity pathway, is the unproductive collapse of the peroxyanion intermediate without formation of the Criegee intermediate, we conclude that the mutants destabilize this peroxyanion intermediate. In so doing, the peroxyanion intermediate is more likely to collapse unproductively prior to the arrival of the ketone substrate. The CHMO_{Rotated} crystal structure readily rationalizes this property of the mutants. In contrast to the Closed conformation, NADP⁺ cannot play a role in stabilizing the peroxyanion intermediate in the Rotated conformation. The decreased ability of the mutants to stabilize the peroxyanion intermediate is in agreement with the adoption of the Rotated conformation prior to the binding of the ketone substrate in the mutants, leading to the uncoupling of NADPH consumption from BV activity. Given that this is the case, it would be expected that the Rotated conformation most likely occurs when the ketone substrate is present, as we propose based on our structural data.

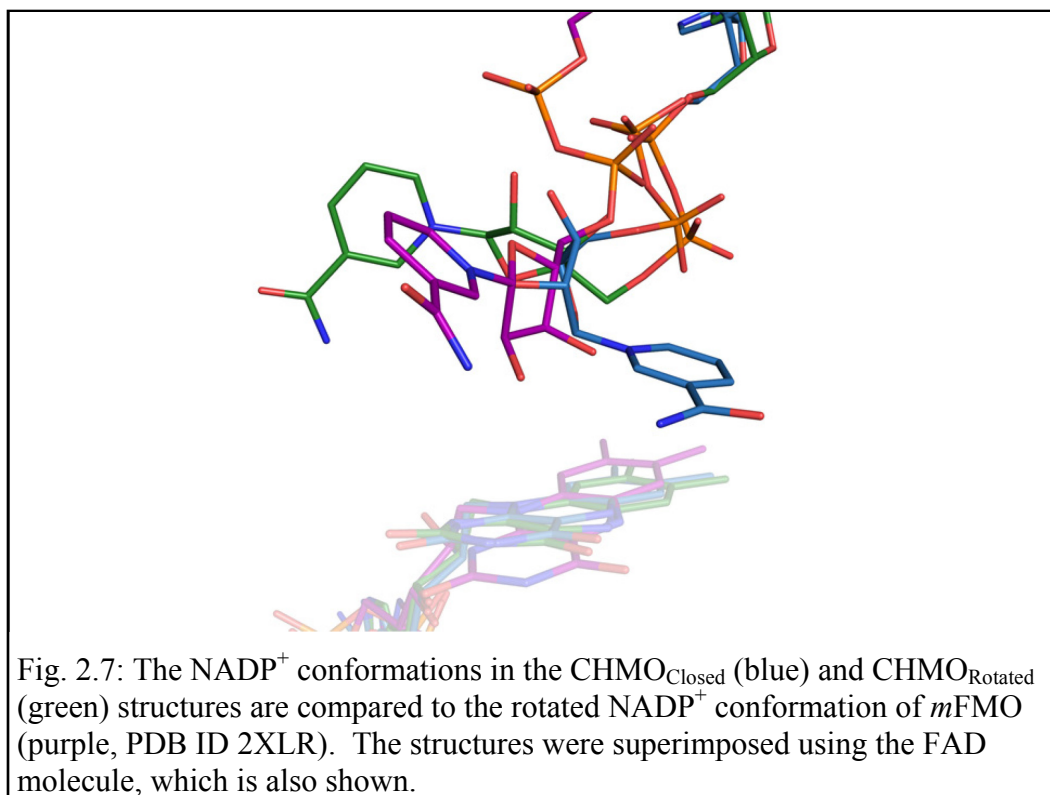
The Rotated structure highlights the need of this enzyme to adopt several arrangements of the substrates and cofactors in order to catalyze the elaborate chemical mechanism. This is a phenomenon seen in other flavoenzymes, including the well-studied *para*-hydroxybenzoate hydroxylase (Schreuder *et al.*, 1988; Schreuder *et al.*, 1989; Gatti *et al.*, 1994; Schreuder *et al.*, 1994; Wang *et al.*, 2002) and phenol hydroxylase (Enroth *et al.*, 1998; Xu *et al.*, 2001), the former being one of the first flavoenzymes to have its structure solved. Like the BVMOs, these hydroxylases face the problem of needing to bring the substrate in

close proximity to a flavin peroxide intermediate that is inherently unstable in protic solvents (Entsch *et al.*, 2005). The details and order of substrate and cofactor binding and release are different, as are the solutions employed to overcome this problem. In spite of this, both the hydroxylases and the BVMOs employ changes in the relative arrangements of the substrates and cofactors at different stages of their catalytic mechanisms to protect the peroxide while permitting the reaction of the substrate to occur (Entsch *et al.*, 2005). The hydroxylases employ three conformations, involving a large rotation of the FAD cofactor, to move from a substrate binding position (Wang *et al.*, 2002) to a flavin reduction position (Gatti *et al.*, 1994; Schreuder *et al.*, 1994) and finally to a substrate hydroxylation position (Schreuder *et al.*, 1988; Schreuder *et al.*, 1989; Gatti *et al.*, 1994). With CHMO, it is the NADP^+ that moves from a flavin reduction position to a substrate binding position and finally to the Criegee position. In both cases, the peroxyanion intermediate is maintained in a protected environment until the moment that the reaction with the substrate can occur (Entsch *et al.*, 2005).

This does not mark the first time a flavin-dependent monooxygenase (FMO) has been observed to have NADP^+ present in a rotated conformation. In a recent study, two crystal structures of an FMO from *Methylophaga* sp. strain SK1 (*m*FMO) were obtained with a similar conformation of NADP^+ (Fig. 2.7) (Orri *et al.*, 2010). These structures were obtained using a mutant enzyme or a NADP^+ analogue. As *m*FMO has substantial structural differences as compared to CHMO, the minor differences in the conformation are to be expected. As *m*FMO has a much more open structure than CHMO, there is no need for the enzyme to adopt an alternate conformation to allow catalysis to occur. As such, this conformation is considered to be a product of the high promiscuity of *m*FMO (Orri *et al.*, 2010). With CHMO, we can see this as being a critical conformation during the catalytic cycle.

2.5.2 The Catalytic Position of Cyclohexanone

The $\text{CHMO}_{\text{Rotated}}$ structure unveils the position of CHMO immediately prior to the reaction with the peroxyanion intermediate to form the Criegee



intermediate. Indeed, we see the substrate close to the ideal geometry for catalysis to occur. Assuming that the peroxyanion intermediate were to form prior to adoption of this conformation, a slight shift in the position of the substrate would result in the necessary geometry for nucleophilic attack on the carbonyl carbon of the ketone. In addition, a fairly minor rotation will result in the antiperiplanar geometry required for the migrating group to form the necessary bond with the peroxide oxygen.

Intriguingly, many of the residues previously implicated in determining the substrate specificity of CHMO and other BVMOs (Reetz *et al.*, 2004a; Bocola *et al.*, 2005; Kayser & Clouthier, 2006; Mihovilovic *et al.*, 2006; Torres Pazmiño *et al.*, 2007; Reetz & Wu, 2008; Reetz & Wu, 2009; Dudek *et al.*, 2011), which generally line the pocket observed in the CHMO_{Closed} structure, are not in the vicinity of the substrate in the CHMO_{Rotated} structure. This suggests that the Rotated conformation is not the most important conformation for determining substrate specificity, regiospecificity, and enantiospecificity of the enzyme. Rather, the putative substrate binding pocket observed in the CHMO_{Closed} structure is critical in determining the substrate profile of the enzyme. We can

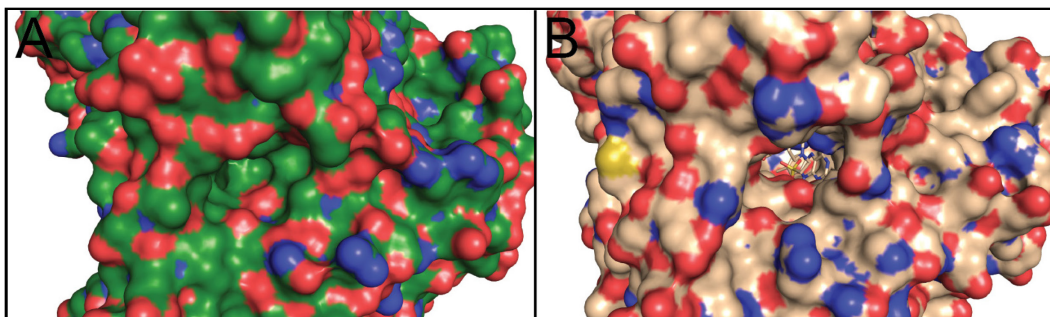


Fig. 2.8: Comparison of the "back" of CHMO (panel A, green, CHMO_{Rotated} conformation) and PAMO (panel B, wheat, PDB ID 2YLT). The protein is shown as a surface, and the ligands (FAD, NADP⁺, cyclohexanone, and MES) are shown as sticks. The large tunnel seen in PAMO is obstructed in CHMO, obscuring the view of the FAD, NADP⁺, and cyclohexanone in the CHMO structure.

speculate that the key residue R329 may play a role in guiding the substrate from the Closed conformation pocket to the Rotated conformation position while maintaining the chemo-, regio-, and enantiospecificity dictated by the CHMO_{Closed} structure.

Given the apparent necessity for a substrate-bound form in the Closed conformation, it appears likely that CHMO employs a series of conformational changes to gradually move the substrate from the solvent into the position observed in the CHMO_{Rotated} structure. In contrast to this, a recent study has reported multiple structures of PAMO with bound NADP⁺ (Orru *et al.*, 2011), all of which were obtained in a conformation similar to the CHMO_{Closed} structure, and some of which also had bound MES, a weak inhibitor of PAMO. The MES inhibitor was observed in a funnel-shaped cavity leading to the catalytic site, suggesting that this is a potential route for the substrate to enter the active site. This arrangement has been previously observed in other flavoproteins, like ornithine hydroxylase (Olucha *et al.*, 2011). In CHMO, this funnel is blocked by a dipeptide insert that is missing in PAMO (residues 278–279, CHMO numbering) (Fig. 2.8), precluding its use for substrate binding in *Rm*CHMO. On the basis of sequence alignments, this dipeptide insert is conserved in almost all CHMOs, including *Ac*CHMO, as well as a number of other closely related BVMOs. The funnel is also blocked in the structures of OTEMO (Leisch *et al.*, 2012). It might be that BVMOs like CHMO and OTEMO use different

mechanisms for substrate binding as compared to PAMO: the former by adopting a series of conformations that progressively move the substrate toward the catalytic position, and the latter via a tunnel that allows for a “back door” access. The possibility that CHMO and PAMO employ slightly different mechanisms is not new; when the detailed kinetic mechanism for PAMO was elucidated (Torres Pazmiño *et al.*, 2008a), it was seen that a spectral change associated with a conformational change during NADP⁺ release in CHMO (Sheng *et al.*, 2001) was not observed. It is also possible that the more limited substrate profile of PAMO as compared to other BVMOs (Bocola *et al.*, 2005; Fraaije *et al.*, 2005; Torres Pazmiño *et al.*, 2007; Reetz & Wu, 2009) is related to this differing substrate binding mechanisms.

2.5.3 A Structural Mechanism for BVMO Catalysis

The CHMO_{Rotated} structure allows us to propose a mechanism for BVMO catalysis from a structural perspective that is consistent with the previously elucidated kinetic mechanism of BVMOs (Ryerson *et al.*, 1982; Sheng *et al.*, 2001; Torres Pazmiño *et al.*, 2008a). The structural states referred to here will correspond to the states shown in Fig. 2.9. The catalytic mechanism begins in the NADP(H)- and substrate-free form (*state A*). Following binding of NADPH (*state B*) and reduction of FAD in an Open-like conformation, the NADP⁺, R329, and D59 will occupy positions that stabilize the reduced flavin. Molecular oxygen will react with the reduced flavin, forming the peroxyanion intermediate (*state C*), which is also stabilized by the same residues. This would also likely take place in a state resembling the Open conformation. The substrate will then bind weakly in the diffuse binding pocket observed in the CHMO_{Open} structure (*state D*). This will trigger the reorganization of the large, unstructured loop, permitting the enzyme to adopt a tight-binding, CHMO_{Closed}-like structure (*state E*). It is this structure that will determine whether a substrate will be accepted by the enzyme, as well as what the regio- and enantiospecificity will be. From there, the enzyme will switch into the peroxyanion intermediate version of the Rotated conformation. This involves the rotation of NADP⁺, the migration of the substrate into the catalytic position, and the shifting of R329 (*state F*). We

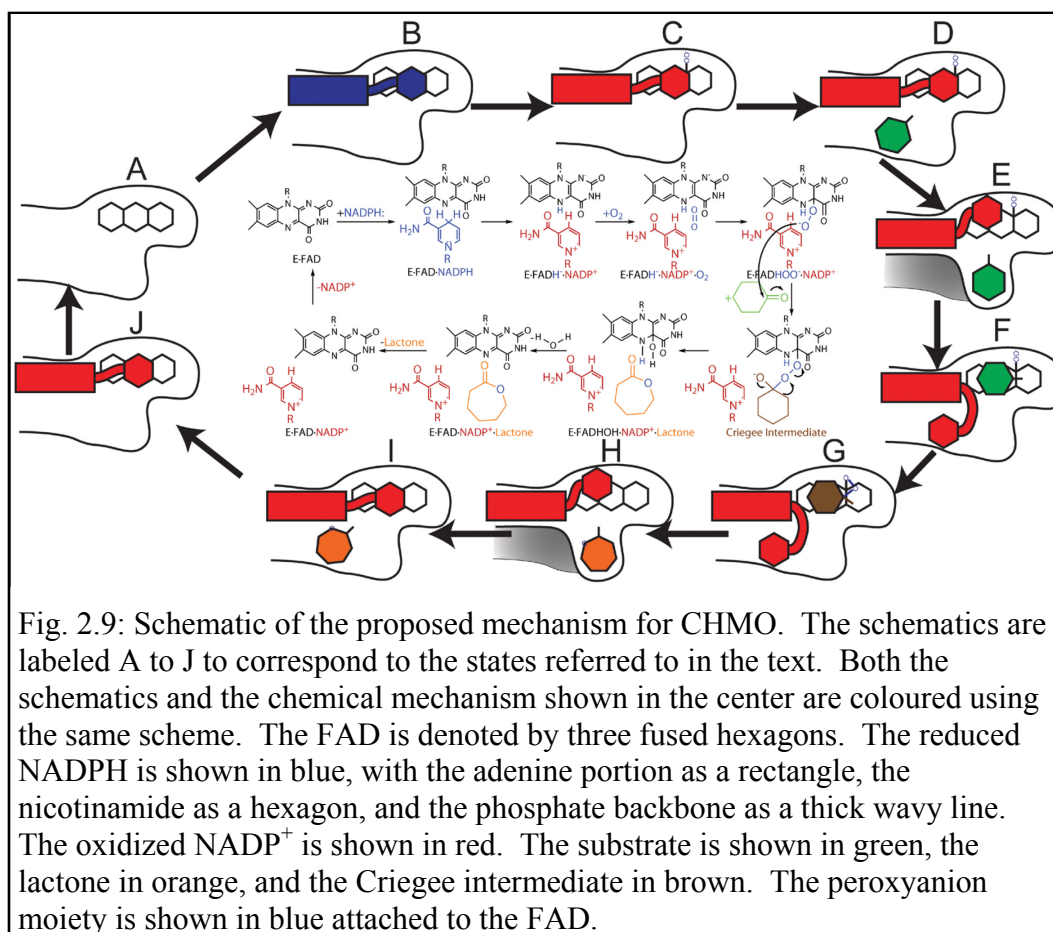


Fig. 2.9: Schematic of the proposed mechanism for CHMO. The schematics are labeled A to J to correspond to the states referred to in the text. Both the schematics and the chemical mechanism shown in the center are coloured using the same scheme. The FAD is denoted by three fused hexagons. The reduced NADPH is shown in blue, with the adenine portion as a rectangle, the nicotinamide as a hexagon, and the phosphate backbone as a thick wavy line. The oxidized NADP⁺ is shown in red. The substrate is shown in green, the lactone in orange, and the Criegee intermediate in brown. The peroxyanion moiety is shown in blue attached to the FAD.

speculate that R329 plays the role of a chaperone guiding the substrate into the catalytic position without allowing it to reorient. This would allow the preservation of the stereochemical requirements imposed in the Closed conformation. At the same time, NADP⁺ obstructs the exit pathway, preventing the substrate from diffusing away from the reaction site. Formation of the Criegee intermediate will occur (*state G*), followed by the formation of the lactone product. Once the product is formed, the enzyme will reverse its steps. The R329 will reposition the product, allowing the NADP⁺ to return to its position above the flavin rings. This may occur in two stages, with the lactone returning to the tight-binding pocket in the CHMO_{Closed} structure (*state H*) before switching to the Open conformation (*state I*), or it may switch to the Open conformation directly (*state I*). Either way, the Open conformation causes the product to be bound again in the loose-binding mode. The product can then be released to the

solvent (*state J*), followed by the release of the oxidized NADP⁺ cofactor, and the return of the enzyme to the initial state (*state A*).

The schematic shown in Fig. 2.9 emphasizes the fact that the protein accommodates a variety of ligand positions, of which three have been observed directly with crystal structures. In spite of this, it appears to do so using only two major global protein conformations: Open-like conformations and Closed-like conformations.

2.6 Conclusion

Up to now, there has been a considerable gap in our understanding of the structural mechanism of the BVMOs. The lack of a crystal structure with the substrate bound to the enzyme has prevented a structural description of the enzyme state allowing for the formation of the Criegee intermediate. The structure presented here shows for the first time a catalytically competent structure of a BVMO. Indeed, it places the ketone substrate in an ideal position for the formation of the Criegee intermediate. At the same time, it provides the necessary space and flexibility for the accommodation of a wide range of substrates of various sizes. Finally, it provides a solution to the problem of how the substrate can reach the active site while the NADP⁺ cofactor remains bound that is consistent with the established kinetic mechanism. This fills a critical gap in understanding of the catalytic mechanism of the BVMOs, and has allowed us to propose a structural description of the mechanism of this family of enzymes.

Chapter 3: Lactone-bound structures of cyclohexanone monooxygenase reveal the basis for substrate selection

3.1 Preface

The crystal structure presented in Chapter 2 provides some important insight into the conformation of the enzyme at a critical stage in the reaction mechanism: the formation and collapse of the Criegee intermediate. In so doing, the broad substrate specificity of the enzyme could be explained. This structure, however, also raises several questions regarding the mechanism of the enzyme. In particular, the binding pocket observed in the CHMO_{Rotated} structure seems unlikely to explain the stereospecificity of the enzyme.

This chapter describes two crystal structures of CHMO with the product, ϵ -caprolactone, bound. These are the first crystal structures of a BVMO to be reported with a product bound. In contrast to the CHMO_{Rotated} structure, these structures place the product in the pocket predicted from the CHMO_{Closed} structure. This can be used to explain the stereospecificity of the enzyme, and also extends the structural mechanism proposed in Chapter 2. A three-state binding mechanism is proposed, and the dynamic properties of the enzyme are once again highlighted.

3.2 Introduction

For over a hundred years, the Baeyer-Villiger oxidation (BVO) reaction has been used for the synthesis of esters from ketones (Baeyer & Villiger, 1899; Baeyer & Villiger, 1900). A wide range of synthetic applications, including in the production of polymers, perfumes, drugs, and other bioactive molecules, have been developed (Krow, 1993). The "classical" oxidative reagents in BVO, however, are peracids, with the original BVO reaction being performed with Caro's acid. These reagents are quite dangerous, especially on an industrial scale, and require the extensive use of organic solvents. The reaction can also lack the stereospecificity required for synthesis of pharmaceutical compounds, necessitating the use of expensive cleanup steps. This has led to attempts to develop alternative reagents, such as transition metal catalysts (Strukul, 1998; ten

Brink *et al.*, 2004) for the BVO reaction. These catalysts unfortunately tend to be expensive, and the reaction is still generally performed in organic solvents.

An alternative to conventional chemical catalysts are the Baeyer-Villiger monooxygenases (BVMOs), a group of bacterial enzymes that can catalyze the BVO reaction. These enzymes have been extensively studied since the 1970s (Donoghue & Trudgill, 1975; Donoghue *et al.*, 1976). The BVMOs are FAD-dependent enzymes that are notable for exhibiting high stereospecificity in many cases, while still having a broad substrate spectrum (Fig. 1.5), making them quite versatile (Kayser, 2009; Alphand & Wohlgemuth, 2010; de Gonzalo *et al.*, 2010; Lau *et al.*, 2010; Torres Pazmiño *et al.*, 2010; Leisch *et al.*, 2011). They only require NADPH and oxygen as co-substrates for the reaction, and produce only NADP⁺ and water as by-products. Genome mining has led to the identification of a number of BVMOs in recent years, increasing the spectrum of substrates that can be targeted by BVMOs (Leisch *et al.*, 2011).

As with any potential biocatalyst, it is important to be able to re-design the biocatalyst for specific industrial applications. While a number of directed evolution studies have led to improved or re-designed BVMOs (Reetz *et al.*, 2004a; Bocola *et al.*, 2005; Kayser & Clouthier, 2006; Mihovilovic *et al.*, 2006; Torres Pazmiño *et al.*, 2007; Reetz & Wu, 2008; Reetz & Wu, 2009; Dudek *et al.*, 2011), a complete understanding of the structural mechanism of BVMOs is necessary to perform rational protein design. The first step towards this goal came a decade ago, with the publication of the first crystal structure of a BVMO, phenylacetone monooxygenase (PAMO), in complex with FAD (Malito *et al.*, 2004). This was followed by two distinct crystal structures of cyclohexanone monooxygenase (CHMO) from *Rhodococcus* sp. HI-31 (*Rm*CHMO) in complex with FAD and NADP⁺, known as the CHMO_{Open} and CHMO_{Closed} crystal structures (Mirza *et al.*, 2009). These two structures highlighted the dynamic properties of the enzyme, characterized by domain rotations and "open" and "closed" forms of the active site (Mirza *et al.*, 2009). In addition, a twenty residue loop was observed to be disordered in the CHMO_{Open} structure, while adopting a folded in, ordered conformation in the CHMO_{Closed} structure. This

folding in of the loop creates an interaction with the NADP⁺ cofactor, which caused it to be displaced laterally past the FAD ring system (Mirza *et al.*, 2009).

Following the publication of these three structures, reports of several more PAMO structures, in complex with NADP⁺ as well as a very weak inhibitor, 2-(*N*-morpholino)ethanesulfonic acid (MES) (Orru *et al.*, 2011), and a series of structures of 2-Oxo- Δ^3 -4,5,5-Trimethylcyclopentenylacetyl-Coenzyme A Monooxygenase (OTEMO) (Leisch *et al.*, 2012) were solved. The PAMO structures confirmed the importance of the large loop, and also provided some insight into the structure of the enzyme with various redox states of the FAD molecule. The OTEMO structures were the first BVMO crystal structures of an enzyme that acts on a very large substrate.

The first crystal structure of a BVMO with its true substrate or product bound was that of CHMO bound to FAD, NADP⁺, and cyclohexanone (Chapter 2). In this crystal structure, the pyridine moiety of the NADP⁺ cofactor was rotated away from the flavin ring system, allowing the substrate to take its place. For this reason, the structure was called the CHMO_{Rotated} structure (Chapter 2). This places the substrate in an ideal position for the formation of the Criegee intermediate without imposing steric constraints on the size of the substrate. In this way, the broad substrate specificity (Fig. 1.5) of the enzyme could be rationalized.

While the broad substrate specificity of CHMO could be rationalized based on the CHMO_{Rotated} structure, the stereospecificity and substrate profile could not. The size of the cavity where the substrate sits in the CHMO_{Rotated} structure is very large, making it difficult to explain substrate selection or stereospecificity. In addition, a number of residues that have been implicated, through mutagenesis and directed evolution studies (Reetz *et al.*, 2004a; Bocla *et al.*, 2005; Kayser & Clouthier, 2006; Mihovilovic *et al.*, 2006; Torres Pazmiño *et al.*, 2007; Reetz & Wu, 2008; Reetz & Wu, 2009; Dudek *et al.*, 2011), in the determination of the substrate profile and stereospecificity of the enzyme are distant from the substrate in the CHMO_{Rotated} structure. It was suggested that the substrate should bind in the pocket previously predicted based on the CHMO_{Closed}

structure for the determination of substrate acceptance and stereospecificity. The enzyme would then switch to the Rotated conformation to allow the formation of the Criegee intermediate (Chapter 2).

Here, we present the first two crystal structures of any BVMO with its product, ϵ -caprolactone, bound. In both crystal structures, the lactone sits in roughly the position that would have been predicted based on the CHMO_{Closed} structure (Mirza *et al.*, 2009), making them the first crystal structures with a natural ligand at this position. This places the product in close proximity to those residues implicated in substrate specificity and stereospecificity (Reetz *et al.*, 2004a; Bocola *et al.*, 2005; Kayser & Clouthier, 2006; Mihovilovic *et al.*, 2006; Torres Pazmiño *et al.*, 2007; Reetz & Wu, 2008; Reetz & Wu, 2009; Dudek *et al.*, 2011), and allows us to propose a model for the basis of the stereospecificity of the enzyme. Furthermore, the differences between the two structures allow us to present a three state binding mechanism that is consistent with and extends the previously proposed structural mechanism (Chapter 2).

3.3 Materials and Methods

3.3.1 Purification and Crystallization

*Rm*CHMO was purified as previously described (section 2.3.3), excepting that the HiLoad Superdex 75 26/60 prep grade column was equilibrated and the sample eluted with 50 mM Tris buffer, pH 8.0. The protein was then concentrated to 5-8 mg/mL and supplemented with a 5-fold molar excess of FAD and NADP⁺. Crystals were obtained using the sitting drop vapour diffusion method in MRC2 crystallization plates. In brief, 2 μ L of protein sample were mixed with 1 μ L of reservoir solution over an 85 μ L reservoir of 100 mM bicine or imidazole buffer, pH 8.0, 45-60% PEG 3350, and 0.1 - 0.25 M ϵ -caprolactone. The plate was incubated at 4°C, and crystals were obtained within one week.

For the cross-linking studies, the method of Lusty was used (Lusty, 1999). A 4 μ L drop consisting of 100 mM imidazole pH 8.0, 50% PEG 3350, and 0.2 M ϵ -caprolactone was placed on an 18 mm siliconized coverslip (Hampton Research) and suspended over a 1 mL reservoir containing the same solution in a 24-well ComboPlate (Greiner Bio-One). Crystals were transferred from the

mother liquor to this new drop for cross-linking. This is to avoid reactivity with the primary amine of Tris. A microliter of concentrated hydrochloric acid was added to 1 mL of a 25% aqueous solution of glutaraldehyde obtained from Sigma-Aldrich (G5882). A 2 μ L drop of this solution was added to the cover slip without allowing it to mix with the drop containing crystals. The plate was incubated at room temperature for 30 minutes, and then at 4°C for 30 minutes, after which the glutaraldehyde drop was removed. The plate was allowed to stand overnight at 4°C, after which 1 μ L of 1 M ϵ -caprolactone was added to the crystal drop. The plate was incubated at 4°C overnight.

3.3.2 Data Collection and Structure Solution

Data were collected under standard cryogenic conditions. For the CHMO_{Tight} crystal, data were collected at beamline 08ID-1 at the Canadian Light Source. For the CHMO_{Loose} crystal, data were collected on a Rigaku MicroMax-007HF generator equipped with VariMax HF optics and a Saturn 944+ CCD detector. The data were processed using the HKL2000 suite of programs (Table 3.1) (Otwinowski & Minor, 1997). The CHMO_{Rotated} structure and the CHMO_{Tight} structure, with ligands removed, were used as the starting models for the refinement of the CHMO_{Tight} and CHMO_{Loose} structures, respectively. The models were subjected to multiple rounds of positional and B-factor refinement using Refmac (Murshudov *et al.*, 1997). Manual model building was performed regularly during refinement using Coot (Emsley *et al.*, 2010). Structural figures were prepared using PyMOL (Schrödinger LLC).

Table 3.1: Data Collection and Refinement Statistics for the CHMO_{Tight} and CHMO_{Loose} Crystal Structures.

	CHMO _{Tight}	CHMO _{Loose}
Data collection statistics		
Space Group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
a, b, c (Å)	55.7, 67.1, 131.6	55.1, 67.0, 133.6
Resolution Range (Å)	24.03 - 1.94	42.62 - 2.50
Completeness (%)	98.3 (91.4)	94.4 (77.5)
Redundancy	13.4 (12.7)	3.7 (1.7)
Reflections with I/ σ <2 (%)	19.6 (53.2)	27.6 (43.8)
Refinement statistics		
Total number of reflections (reflections in	36,235 (3670)	16,591 (1678)

	CHMO _{Tight}	CHMO _{Loose}
R _{free} set)		
R _{factor} (%) (Work + Free/Free)	17.1/21.9	19.4/26.0
Number of atoms	4324	4044
Protein	4066	3872
Water	138	53
Cofactors and Product	109	109
Other	11	10
RMSD		
Bond length (Å)	0.01	0.02
Bond angle (°)	1.56	1.87
Ramachandran plot	526 (100%)	505 (100%)
Residues in favoured positions	504 (95.82%)	482 (95.45%)
Residues in allowed positions	21 (3.99%)	19 (3.76%)
Residues in disallowed positions	1 (0.19%)	4 (0.79%)

3.3.3 Creation of the W492A Mutant

The His₈-TEV-ChnB1 plasmid was amplified using the protocol from the Agilent QuikChange II Site-Directed Mutagenesis Kit. The following primers were used (altered bases are underlined): Sense—5'-ACCAAGGGCGATTCGGCGATCTTCGGCGCGAACATCCCAGGGAAGAAGCCGAGC-3'; Antisense—5'-GCTCGGCTTCTTCCCTIGGGATGTTCGCGCCGAAGATCGCCCGAATCGCCTTGGT-3'. Plasmid DNA from positive clones was prepared and sequenced as described in section 2.3.2.

3.3.4 Substrate Profiling of Wild-type and W492A CHMO

Expression and purification protocols for the wild-type and W492A mutant are identical to that used in section 2.3.3 for enzyme kinetics and NMR, except that after Ni-NTA purification, the protein was dialysed against 50 mM Tris pH 8.0. After dialysis, the protein was concentrated to ~10 mg/mL.

Substrates were prepared in 100% isopropanol at 1 M concentration. Substrate profiling was performed in 5 mL reaction volumes of 50 mM Tris pH 8.0, with 1.5 mM NADPH and 3 mM of substrate, in triplicate. The reaction was

started by adding either wild-type or W492A CHMO, and allowed to proceed at 4°C until the NADPH was depleted based on the absorbance at 340 nm. In addition, controls with isopropanol and no substrate, as well as with substrate but no enzyme, were prepared. The following substrates were used: 2-methylcyclohexanone (Aldrich M38400), 2-propylcyclohexanone (TCI P1114), (+)-dihydrocarvone (Aldrich 37275), and 2,6-dimethylcyclohexanone (Aldrich 102261).

The compounds were extracted by mixing the aqueous reaction solution with an equal volume of ethyl acetate. The organic layer was removed and dried over sodium sulfate. The products were analysed using chiral gas chromatography on a Hewlett Packard 6890 Series gas chromatograph equipped with a Varian Chirasil-Dex CB column (25 m x 0.25 mm x 0.25 μ m), a flame ionization detector, and a Hewlett Packard 6890 Series auto-sampler. A 1 μ L volume from the extracted sample was injected under pulse split (1:5) condition with 8 mL/min helium flow. For most substrates, the following gas chromatography conditions were used: 75°C for 5 minutes, 2.5°C/minute until 175°C, 175°C for 10 minutes, 10°C/minute until 200°C, and 200°C for 2.5 minutes. In order to achieve better separation, the increase from 75°C to 175°C was performed at 1.5°C/minute in certain cases. As well, for 2-methylcyclohexanone, the following conditions were used: 80°C for 5 minutes, 2°C/minute until 130°C, 130°C for 1 minute, 15°C/minute until 200°C, and 200°C for 2 minutes. For dihydrocarvone and 2,6-dimethylcyclohexanone, the mutant samples were analysed again with a 4 μ L injection volume to obtain more reliable peak sizes for the products.

3.4 Results

3.4.1 Crystal structures of CHMO in complex with its product

The CHMO_{Tight} and CHMO_{Loose} structures were solved to 1.9 and 2.5 Å, respectively. The CHMO_{Tight} structure could be modelled from residues 6-534, whereas the CHMO_{Loose} structure could be modelled from residues 6-533. In addition, a large loop was not visible in the CHMO_{Loose} electron density map ranging from residues 489-505, and was therefore omitted from the model. This

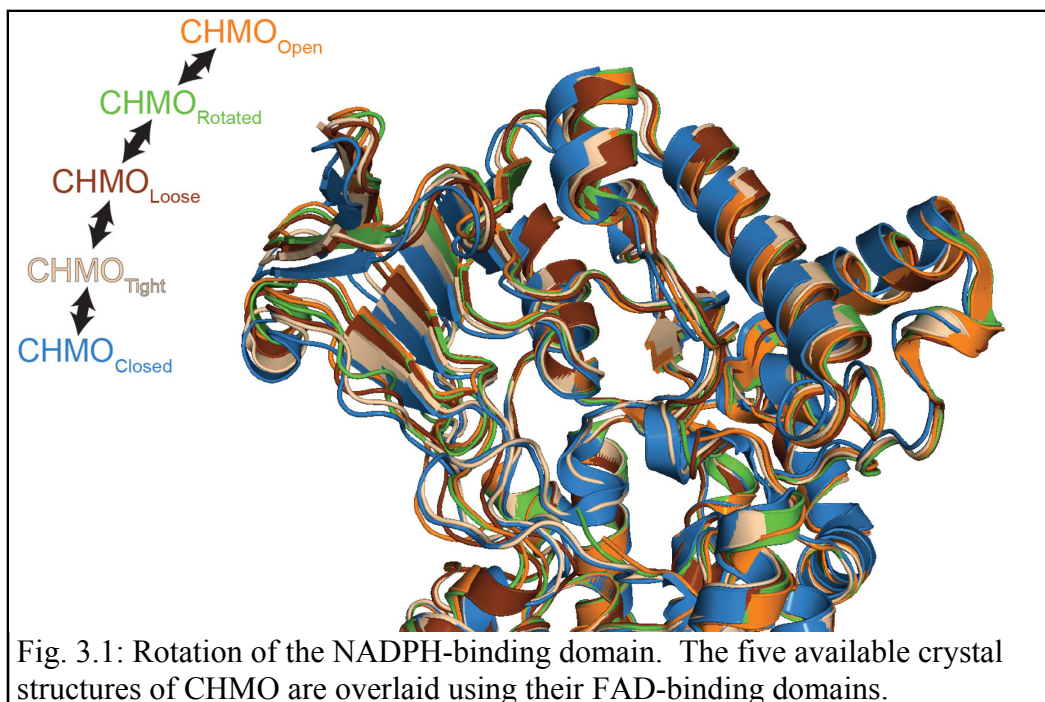


Fig. 3.1: Rotation of the NADPH-binding domain. The five available crystal structures of CHMO are overlaid using their FAD-binding domains.

has been previously observed in the CHMO_{Open} (Mirza *et al.*, 2009) and CHMO_{Rotated} (see Chapter 2, especially section 2.4.1) structures. Both structures had clearly defined density for both the FAD and NADP⁺ cofactors. Of note, in spite of the disordering of the loop in the CHMO_{Loose} structure, both structures had the NADP⁺ molecule in a similar conformation, as was previously observed in the CHMO_{Closed} structure (Mirza *et al.*, 2009). The CHMO_{Loose} crystal structure, which was cross-linked using glutaraldehyde, had density visible beyond the end of Lys40 and also connecting Lys174 to Arg179. The former was modelled as a glutaraldehyde modification to Lys40, whereas the latter was modelled as a glutaraldehyde bridge between Lys174 and Arg179 as described by Salem *et al.* (2010).

In spite of the fact that both of these structures have the same set of ligands bound, these structures are not identical. Indeed, the CHMO_{Tight} structure more closely resembles the CHMO_{Closed} structure, with an RMSD of 0.42 Å for backbone atoms present in both structures after overlaying the FAD-binding domains. This compares to an RMSD of 0.61 Å when comparing the CHMO_{Loose} and CHMO_{Tight} structures. The CHMO_{Loose} structure resembles the CHMO_{Rotated} structure marginally more than it resembles CHMO_{Tight}, with an RMSD of 0.59 Å.

If the two new CHMO structures are overlaid onto the three available CHMO structures, using the FAD-binding domains to superimpose them, the rotation of the NADPH-binding domain can be compared between all of these structures. The two extremes of domain rotation remain the CHMO_{Open} and CHMO_{Closed} structures. For the three structures that have a ketone substrate or lactone product bound, the CHMO_{Rotated} structure is closest to the CHMO_{Open} structure, while the CHMO_{Tight} structure is closest to the CHMO_{Closed} structure. This means that starting with the CHMO_{Open} structure, a gradual closing of the NADPH-binding domain can occur by transitioning from structure to structure in the following order: CHMO_{Open} → CHMO_{Rotated} → CHMO_{Loose} → CHMO_{Tight} → CHMO_{Closed} (Fig. 3.1).

3.4.2 Presence of ϵ -caprolactone in the substrate binding site

3.4.2.1 CHMO_{Tight} Crystal Structure

Upon initial solution of the CHMO_{Tight} crystal structure, a number of characteristics of the structure suggested that it was distinct from others obtained under similar conditions, but in the absence of ϵ -caprolactone. The most obvious of these was the ordering of the loop from residues 487-505, which was consistently disordered in the absence of either ketone or lactone, and also in the presence of cyclohexanone. In addition, the conformation of NADP⁺, which resembled either the CHMO_{Open} or CHMO_{Rotated} structure in the absence of the lactone, adopted the conformation observed in the CHMO_{Closed} structure. Finally, crystals obtained in the absence of either cyclohexanone or ϵ -caprolactone would tend to break the orthorhombic symmetry, and yield a monoclinic crystal with unit cell angles all equal to 90° and with two protomers in the asymmetric unit.

Close examination of the electron density map in the area identified in the CHMO_{Closed} structure as a putative substrate binding pocket (Mirza *et al.*, 2009) revealed a ring-shaped region of electron density corresponding to the size and shape of ϵ -caprolactone. Upon modelling the lactone in this site, this ring of density remained visible in the weighted 2F_o-F_c map, but only when the map was contoured at 0.6 sigma (Fig. 3.2A). At 1.0 sigma, density corresponding to the carbonyl group and the ring oxygen were visible, while density for the remaining

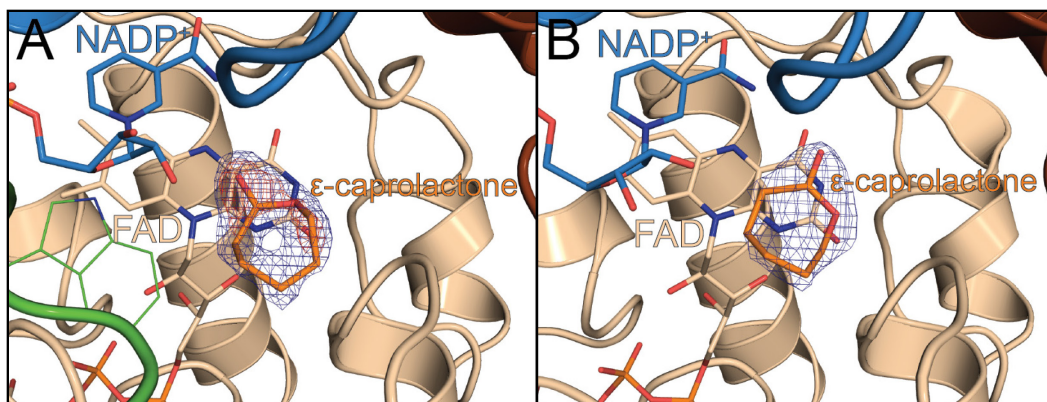


Fig. 3.2: Electron density map for ϵ -caprolactone in the (A) CHMO_{Tight} and (B) CHMO_{Loose} crystal structures. The $2F_o-F_c$ map is shown contoured at 1.0 sigma (blue) and 0.6 sigma (red, panel A only) within 1.7 Å and 1.3 Å of ϵ -caprolactone, respectively.

carbons was very weak or not visible at all. This suggested while the product may be present in this crystal structure, it was unlikely to be present at full occupancy. Based on the contour level of the weighted $2F_o-F_c$ electron density map at which the full lactone ring was visible, the occupancy for the lactone was set to 0.6.

3.4.2.2 CHMO_{Loose} Crystal Structure

In order to corroborate the presence of the lactone in the CHMO_{Tight} structure, it was decided to cross-link the crystals with glutaraldehyde and soak in more ligand than could normally be tolerated by the crystals. It was expected that this would lead to a very similar structure with better defined density for the lactone. Surprisingly, there were a number of alterations in the crystal structure that suggested a more dynamic conformation than the original, non-cross-linked crystals. Note that since these changes require alterations within an already formed crystal, it is likely that the lower resolution and slightly elevated R-factors of the CHMO_{Loose} structure are due to perturbations in the crystal lattice caused by the cross-linking reaction. The details of the changes introduced by cross-linking will be discussed in more detail in section 3.5.1.

One characteristic of this structure was the disordering of the large loop spanning residues 489-505, resulting in a much more open active site (see section 3.4.3). It is likely for this reason that the soaking experiments appear to have been so successful. There was clearer density in the same location in which the ligand was observed in the CHMO_{Tight} structure. When ϵ -caprolactone was

modelled in at this position, the weighted $2F_o - F_c$ map contoured at 1.0 sigma clearly shows a fairly strong region of electron density corresponding to the lactone (Fig. 3.2B). This is in spite of the lower resolution of this crystal structure.

3.4.2.3 Comparison of the lactone positions

As the two ϵ -caprolactone-bound crystal structures are not isomorphic, it became necessary to compare the binding mode of the lactone in the two structures. When the two structures are overlaid, the lactone is found to sit in almost exactly in the same place. There is, however, a rotation of roughly 30° that causes a displacement of the position of the carbonyl oxygen. This shift could be related to the disordering of the large loop in the CHMO_{Loose} structure. The lack of strong features in the electron density map of the CHMO_{Loose} structure, however, means that unlike in the CHMO_{Tight} structure, there is some ambiguity in the rotational position of the lactone in the CHMO_{Loose} structure. This could reflect greater freedom of movement for the ligand in the Loose conformation.

When comparing all of the residues that are within 6.0 Å of either lactone in either structure, most of the residues sit in the same position in both structures. The RMSD when comparing all atoms in all of these residues is 1.1 Å. When residues 145 and 146, which form part of the linker loop between the FAD-binding and NADPH-binding domains, are excluded from this analysis, this number drops to 0.7 Å. This suggests that the residues critical for recognizing the lactone remain in the same position in spite of the conformational changes observed in the global protein conformation (Fig. 3.3).

3.4.3 Binding Site Properties of the Lactone-Bound Structures

While the residues in close proximity to the lactone remain in a similar position in both lactone-bound crystal structures, there remains a significant difference in the ligand binding site in these structures. The rotation of the NADPH-binding domain and concurrent order-disorder transition of the large loop spanning residues 489-505 results in a change in the solvent accessibility of the binding site. In the CHMO_{Loose} crystal structure, the position of the NADPH-binding domain and the disorder of the large loop results in a fairly open binding

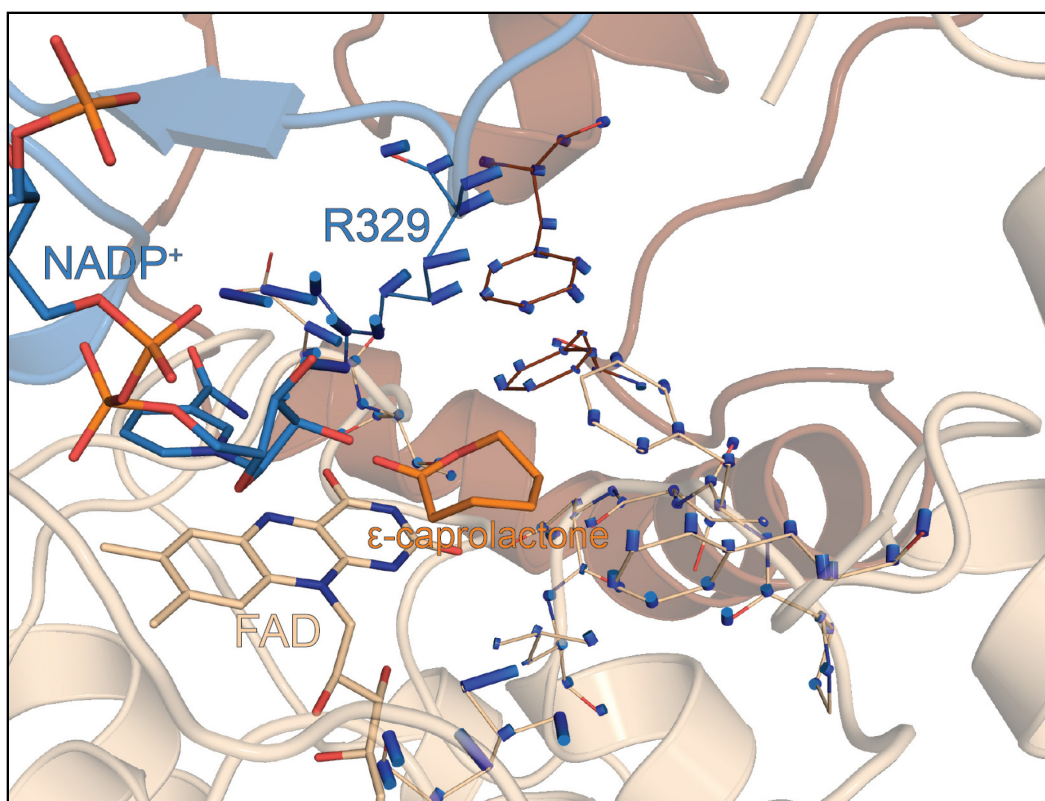
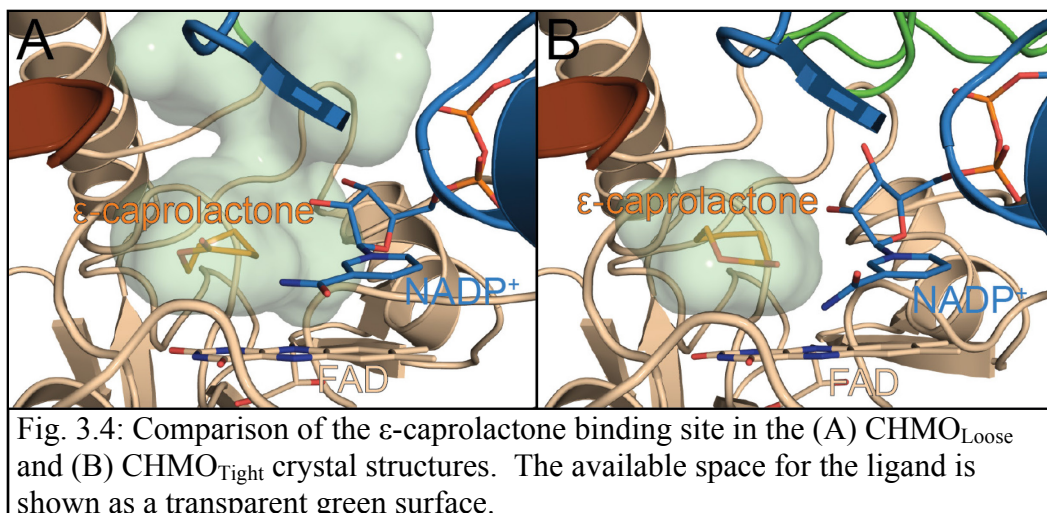


Fig. 3.3: Comparison of the CHMO_{Tight} and CHMO_{Loose} active sites. Vectors are drawn between matching residues within 6.0 Å of ε-caprolactone in either structure. The CHMO_{Tight} crystal structure is shown as a reference.

site (Fig. 3.4A), though not quite as large as what is observed in the CHMO_{Open} crystal structure (Fig. 1.10) (Mirza *et al.*, 2009). The CHMO_{Tight} structure, in contrast, has a binding site that is roughly the same size as the small binding pocket observed in the CHMO_{Closed} crystal structure (Mirza *et al.*, 2009) (Fig. 3.4B). This appears to suggest that the lactone is being bound in two alternative binding modes, which provides the rationale for the naming of these two structures. The CHMO_{Tight} structure reflects a tighter binding mode that is likely important for substrate selection and orientation. The CHMO_{Loose} structure appears to reflect a looser binding mode which may reflect an initial association between the ligand and the enzyme. This structural state is likely more "fluid" than the Tight conformation, and may be a representative of a "family" of Loose conformations that vary slightly depending on the particular substrate and protein dynamics. The similarity to the CHMO_{Open} structure, however, does suggest that the Loose conformation is a distinct structural state, and not simply an artifact of the glutaraldehyde cross-linking.



3.4.4 Effects of the W492A on Substrate Profile

In order to assess the impact of the loop ordering on the substrate profile of the enzyme, the substrate profile of the wild-type enzyme was compared to that of the W492A mutant. W492 is a critical residue for the interaction of the loop with NADP⁺, and the W492A mutant has previously been shown to have substantially decreased activity (Mirza *et al.*, 2009), implying that the efficiency of loop ordering is diminished or abrogated completely. The isopropanol controls demonstrated that the mutant is prone to increased uncoupling of NADPH consumption from substrate conversion, as the mutant completely consumed the NADPH within one day in the absence of substrate, whereas the NADPH was stable in the presence of the wild-type enzyme for several days.

Of the substrates tested, the mutant was found to be more selective in all cases, and in some cases became almost completely unable to transform certain substrates (see Table 3.2). Both 2-methyl and 2-propylcyclohexanone show gains in their E-values, becoming increasingly S-selective. For 2,6-dimethylcyclohexanone, of the three possible isomers (*cis*, R/R-*trans*, and S/S-*trans*), it appears that the *cis*-isomer and one of the *trans*-isomers are both converted efficiently by the wild-type enzyme. In the case of the mutant, the enzyme loses the ability to convert the *cis*-isomer, converting only the same *trans*-isomer in significant quantities. For (+)-dihydrocarvone, the wild-type is already highly selective for the *trans*-isomer, and produces essentially no lactone

from the *cis*-isomer. In the mutant, we see almost complete loss of activity, with the conversion rate dropping from 51% to 2%.

Table 3.2: Substrate profiling of wild-type and W492A *Rm*CHMO.

Substrate	Enzyme	% Conversion			E-value [*]
		Isomer 1	Isomer 2	All Isomers	
2-methyl-cyclohexanone	Wild-type	71±5 (S)	33±5 (R)	51±5	3.5
	W492A	64±4 (S)	10±2 (R)	36±3	15.5
2-propyl-cyclohexanone	Wild-type	85±3 (S)	23±10 (R)	52±7	9.7
	W492A	28±7 (S)	-1±9 (R)	13±8	N/A [†]
(+) - dihydrocarvone	Wild-type	-6±4 (<i>cis</i>)	62±6 (<i>trans</i>)	51±6	N/A [†]
	W492A	-1±1 (<i>cis</i>)	2±1 (<i>trans</i>)	2±1	N/A [†]
2,6-dimethyl-cyclohexanone	Wild-type	46±7 (<i>cis</i>)	46±4 (<i>trans</i>)	46±7	N/A [†]
	W492A	0±2 (<i>cis</i>)	17±5 (<i>trans</i>)	3±3	N/A [†]

[†]N/A: Not applicable.

^{*}E-value is an estimation of the ratio of an enzyme's k_{cat}/K_M values towards the R and S enantiomers of a substrate (Chen *et al.*, 1982). It is calculated using the following equation if two substrate enantiomers are converted exclusively into two product enantiomers by CHMO with non-zero percent conversion:

$$E = \frac{\ln[(1-c)(1-EEs)]}{\ln[(1-c)(1+EEs)]} \quad (\text{equation 3.1})$$

where

$$c = 1 - \frac{\sum [Isomers]_{final}}{\sum [Isomers]_{initial}} \quad \text{and} \quad EEs = \frac{|[R \text{ Isomer}]_{final} - [S \text{ Isomer}]_{final}|}{[R \text{ Isomer}]_{final} + [S \text{ Isomer}]_{final}}$$

[Isomer] refers to the concentration of an isomer. "Initial" and "final" refer to the substrate concentrations before and after BVO bioconversion.

3.5 Discussion

3.5.1 CHMO is a dynamic enzyme

The CHMO_{Open} and CHMO_{Closed} crystal structures first suggested that the BVMOs are highly dynamic proteins, and are able to change conformations even without a trigger from a ligand being bound or released. This was shown when both of these original CHMO structures had the same ligands bound, but with markedly different conformations (Mirza *et al.*, 2009). The CHMO_{Loose} and CHMO_{Tight} structures once again highlight this feature. While both of these structures have FAD, NADP⁺, and ϵ -caprolactone bound, they are characterized by a domain rotation that causes the closing up of the substrate binding site. In addition, the large loop is ordered in CHMO_{Tight} and disordered in CHMO_{Loose}. This emphasizes the dynamic properties of this loop, and is consistent with the

large loop coming on and off the enzyme at least twice per catalytic cycle (see Fig. 3.10). It also provides more evidence that this loop is catalytically important, as previously proposed (Mirza *et al.*, 2009).

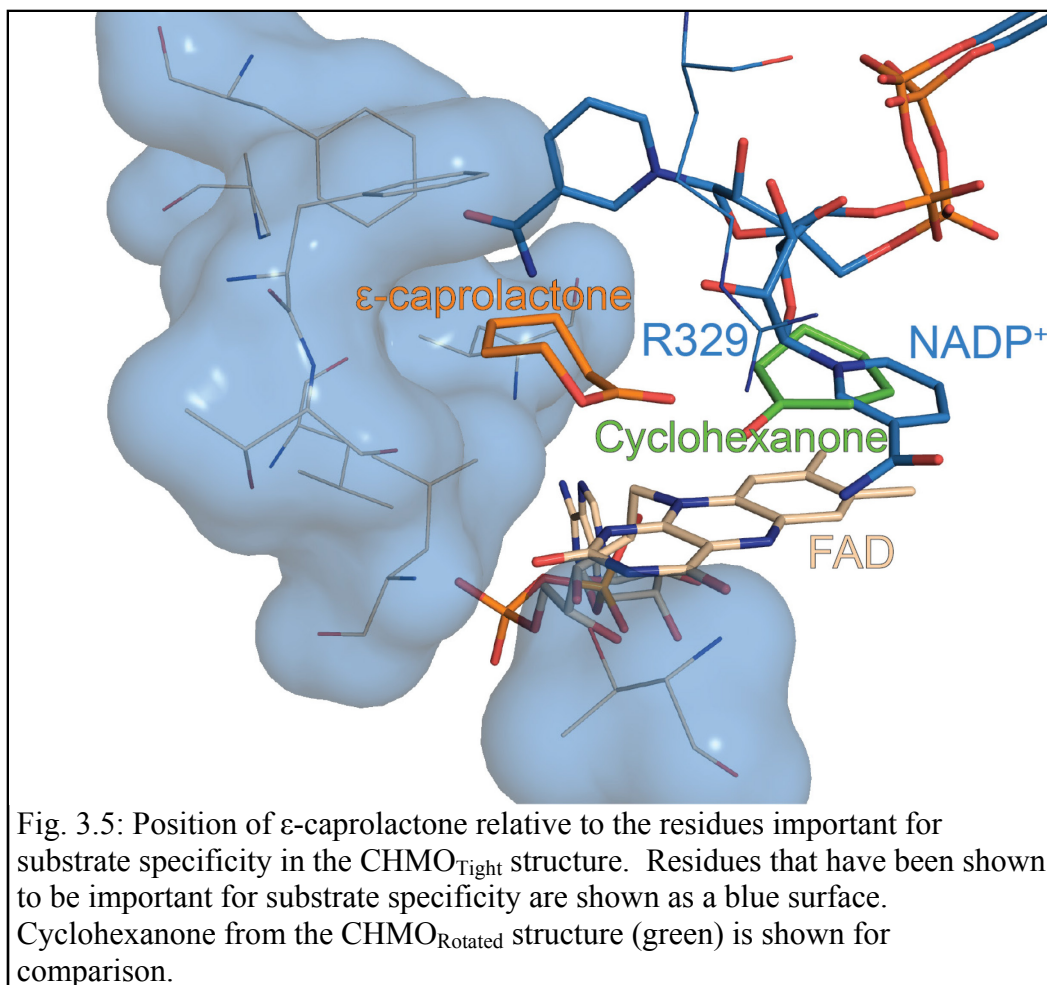
The crystal cross-linking, which was performed with glutaraldehyde, was expected to stabilize the CHMO_{Tight} structure, allowing the enzyme to maintain the same conformation, while at the same time being able to tolerate high concentrations of ϵ -caprolactone in a soaking experiment. It was therefore quite surprising that the CHMO_{Loose} structure was in fact in a different conformation, characterized by a rotation of the NADPH-binding domain and disordering of the large loop. This implies that even in the CHMO_{Tight} crystal, there must be some mobility in the NADPH-binding domain and some fraction of the loop must be present in the disordered conformation at room temperature. The freezing of the crystal prior to data collection may be responsible for locking the rotational position of the NADPH-binding domain in the crystal, as well as the ordered conformation of the loop.

Glutaraldehyde is expected to react with the primary amines of lysine residues, though it is also able to react with arginine. CHMO contains eighteen lysine residues, of which almost all can be described as surface residues, with K328 being the notable exception. In addition, three of these residues are present in the large loop. Two lysines had visible electron density for a glutaraldehyde modification. K40, which is in the FAD-binding domain, does not have its position significantly perturbed by the modification as compared to the other crystal structures, and does not form a covalent link with another residue that is visible in the electron density map. K174, which cross-links with R179 of the same protein molecule, is located in the NADPH-binding domain. These residues are in similar positions as in the CHMO_{Open} structure, which is the structure that most closely resembles the CHMO_{Loose} structure. As the crystals were found to be more durable, it is expected that other lysine residues are modified, but in a way that is not consistent throughout the crystal and therefore not observable in the electron density map.

The surprising observation that the loop goes from being ordered in the CHMO_{Tight} structure to disordered in the CHMO_{Loose} structure is indicative of a variety of loop conformations being present even in the CHMO_{Tight} crystals. In the CHMO_{Tight} structure, the loop should be almost entirely in the ordered position, at least after freezing the crystal. In order for glutaraldehyde to change the conformation of the loop, however, it is necessary for the loop to occasionally transition to a disordered, likely short-lived conformation, at least at room temperature. The fact that the loop is not visible in the CHMO_{Loose} structure in either the ordered conformation seen in the CHMO_{Tight} structure or in a second, "out" conformation also indicates that there must be a population of disordered loop conformations that are being stabilized by glutaraldehyde cross-linking, thereby preventing any of them from being observed in the crystal structure. This can be explained by the presence of three lysine residues in the loop. These lysines create an opportunity for loop to be readily modified by glutaraldehyde in any "out" conformation, and would also create a variety of different cross-linking configurations that would prevent a single conformation from being observable in the crystal structure. Further study of the mechanistic importance of this loop will be described in Chapter 4.

3.5.2 CHMO_{Tight} is likely the conformation where stereospecificity is determined

A number of directed evolution and mutagenesis studies have identified residues that, when altered, trigger changes in the substrates that can be accommodated by the BVMOs, as well as perturbations in the stereospecificity of the enzymes. Close to the flavin ring system, these include T60 (Dudek *et al.*, 2011), L145 (Reetz *et al.*, 2004a; Kayser & Clouthier, 2006; Dudek *et al.*, 2011), L428 (Reetz *et al.*, 2004a; Mihovilovic *et al.*, 2006; Dudek *et al.*, 2011), P430 (Reetz *et al.*, 2004a), F434 (Reetz *et al.*, 2004a; Bocola *et al.*, 2005; Kayser & Clouthier, 2006; Mihovilovic *et al.*, 2006; Reetz & Wu, 2008; Reetz & Wu, 2009; Dudek *et al.*, 2011), T435 (Reetz *et al.*, 2004a; Bocola *et al.*, 2005; Reetz & Wu, 2008; Reetz & Wu, 2009; Dudek *et al.*, 2011), L437 (Torres Pazmiño *et al.*, 2007; Dudek *et al.*, 2011), and F507 (Reetz *et al.*, 2004a; Kayser & Clouthier, 2006),



with all residues indicated with *RmCHMO* residue numbering. It has been previously noted that the CHMO_{Rotated} structure, which is in an ideal position for formation of the Criegee intermediate, does not place cyclohexanone close to these residues (see section 2.5.2). In contrast, both CHMO_{Tight} and CHMO_{Loose} place ϵ -caprolactone in close proximity to these key residues (Fig. 3.5). Assuming that the substrate, in the presence of the peroxyanion intermediate, would also bind in the same manner as the product in our structures, this strongly suggests that the acceptance of substrates, as well as the regio- and enantiospecificity of the enzyme, is determined in the Loose and/or Tight conformations.

The existence of two distinct lactone-bound structures is indicative of a multi-step binding mechanism. The more open binding pocket in the CHMO_{Loose} structure is likely to be the first, loose-binding mode. In this state, the enzyme

will undergo the initial interaction with a potential substrate, and likewise will be the last state prior to product release. It would be expected that this step would be fairly non-specific. It may also vary slightly depending on the substrate and dynamic protein motions, thereby acting as a representative of a "family" of structural states. Once the substrate is associated with the enzyme, a conformational change will occur causing the enzyme to adopt a second binding mode in the Tight conformation. The small, closed off binding pocket in this conformation would not be able to form in the presence of a molecule that is not an appropriate substrate for the enzyme. Likewise, if the substrate were to be in the "wrong" orientation for the proper stereospecificity of the enzyme, the conformational change would either not occur, or force the substrate into the correct orientation. Once properly in the Tight conformation, the enzyme could then transition into the Rotated conformation to form the third binding mode and allow the formation of the Criegee intermediate. This three-state binding mechanism is consistent with what was initially proposed in section 2.5.3.

3.5.3 The stereospecificity of CHMO can be predicted by the CHMO_{Tight} structure

One of the most important questions in understanding the structural mechanism of the BVMOs is the structural basis for the stereospecificity of the enzyme. Unfortunately, the resolution of the CHMO_{Tight} and CHMO_{Loose} structures are insufficient to differentiate between the oxygen and carbon atoms within the lactone ring. In order to attempt to determine on which side of the carbonyl the oxygen is located, cases where the stereospecificity differs, either in a BVMO mutant or in a homologue, were examined to justify the placement of the oxygen atom in the lactone ring. We will distinguish this characteristic of the enzyme, the "decision" as to which side of the carbonyl the oxygen is incorporated from the frame of reference of the protein, by referring to it as the "directionality" of oxygen incorporation. The directionality can be seen as a single property that covers all ketone substrates. In contrast, chemical terms like "regiospecificity" and "enantiospecificity" will be determined by the directionality

of the enzyme, but will always depend on the stereochemical properties of the substrate and product in question.

Two cases where the directionality of oxygen incorporation will be assessed are considered. The first case is the mutation of F434 to a serine, thereby improving the enantiospecificity of CHMO with respect to the prochiral substrate, 4-hydroxycyclohexanone (Reetz *et al.*, 2004a). The second case examines the differences in regiospecificity of the terpenones in CHMO and CPMO (Cernuchova & Mihovilovic, 2007).

In order to assess the directionality of the enzyme, it is assumed that all of the cyclohexanone-based substrates will bind in the CHMO_{Tight} conformation with the plane formed by the carbonyl group roughly parallel to the flavin ring system, allowing the formation of pi-pi interactions. Furthermore, as these are cyclohexane-based rings, it is expected that they will adopt chair-like conformations, causing carbons 3, 4, and 5 to sit either above or below the carbonyl plane, producing two possible configurations of the substrate: "up" or "down." This position should also be predictive of the eventual position of the substrate in the CHMO_{Rotated} conformation, which is required to form the Criegee intermediate. Finally, it is assumed that in all BVMOs and for all substrates, the directionality of oxygen incorporation will be the same.

The substrate 4-hydroxycyclohexanone is prochiral, with the possibility of being converted into the R- or S-enantiomer, depending on which of the two carbons migrates. It has been shown that CHMO shows almost no preference for either isomer, with an almost negligible EE of 9% in favour of the R-enantiomer. In contrast, a single point mutation to serine causes a large improvement in EE to 79% in favour of the S-enantiomer (Reetz *et al.*, 2004a). If ϵ -caprolactone is replaced with 4-hydroxycyclohexanone in the CHMO_{Tight} structure, it can be seen that there is very little reason for the substrate to prefer either the "up" or the "down" configuration (Fig. 3.6A). In contrast, if F434 is mutated to serine, a hydrogen bonding interaction can be formed in the "up" configuration between the serine hydroxyl group and the 4-hydroxyl group of the substrate (Fig. 3.6B). In order to form the preferred S-isomer, the directionality of the enzyme must

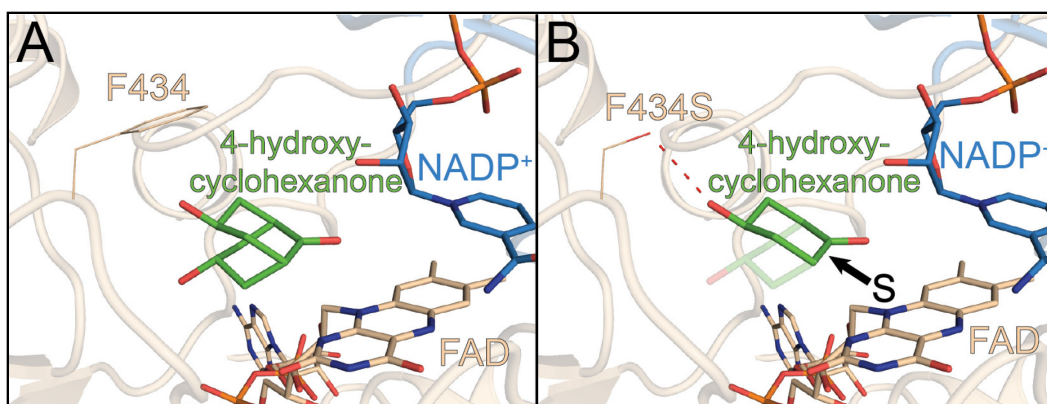
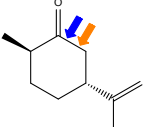
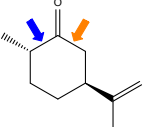
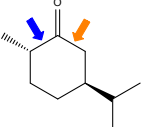
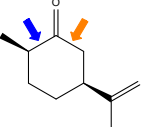


Fig. 3.6: Enantiospecificity of CHMO for 4-hydroxycyclohexanone. (A) In the wild-type structure, neither the "up" nor the "down" configuration of the substrate is preferred. (B) In the F434S mutant, the "up" configuration is preferred due to the formation of a hydrogen bond. The positions of the substrate are modelled based on the position of ϵ -caprolactone in the CHMO_{Tight} crystal structure. The directionality of the enzyme is indicated with the black arrow.

place the oxygen on the close side (as oriented in Fig. 3.6) of the carbonyl, as modelled in the crystal structure.

In order to confirm this preference for which side of the substrate will migrate, the differing regiospecificity of CHMO and CPMO towards various terpenones was considered (Cernuchova & Mihovilovic, 2007). The regiospecificity of various CHMO homologues, as well as CPMO, has previously been examined for a panel of terpenones. Some of the terpenones either were not converted, or the regiospecificity was rather poor. These will not be considered. Remaining are (+)- and (-)-*trans*-dihydrocarvone, (-)-*cis*-dihydrocarvone, and (-)-carvomenthone (Table 3.3). In all cases, CPMO yielded the abnormal lactone. By contrast, CHMO yielded the normal lactone for (-)-*trans*-dihydrocarvone, (-)-*cis*-dihydrocarvone, and (-)-carvomenthone, and yielded the abnormal lactone for (+)-*trans*-dihydrocarvone (Cernuchova & Mihovilovic, 2007). It should be noted that the *trans*-isomers can accommodate both substituents in the preferred equatorial position, whereas the *cis*-isomers must place one of the two substituents, likely the smaller methyl group, in the disfavoured axial position.

Table 3.3: Regiospecificity of CHMO and CPMO with respect to four terpenones. The formation of the normal or abnormal lactone by each enzyme is indicated in the table, and also denoted by a blue (CHMO) or orange (CPMO) arrow.

Substrate				
	(+)- <i>trans</i> -dihydrocarvone	(-)- <i>trans</i> -dihydrocarvone	(-)-carvomenthone	(-)- <i>cis</i> -dihydrocarvone
CHMO	Abnormal	Normal	Normal	Normal
CPMO	Abnormal	Abnormal	Abnormal	Abnormal

In examining altered residues between CHMO and CPMO lining the CHMO_{Tight} binding pocket, four residues undergo notable changes (Fig. 3.7AB). Most importantly, L437, which sits below the substrate at the end of a small loop that passes behind the substrate, becomes a glycine. This creates more room below the substrate, and will likely make the loop more flexible. F507, which sits to the back-left of the substrate, is converted to an asparagine, which is expected to take up less space and should be more flexible than the rigid phenylalanine. L145, which sits to the bottom-left of the substrate, becomes a phenylalanine, which should crowd the substrate slightly on that side. Finally, L428, which is somewhat distant from the substrate but also occupies some space to the bottom-left, becomes a tyrosine. This larger, more rigid residue may not directly affect the substrate, but the crowding in that region would impact neighbouring residues, thereby reducing the space available below and to the left of the substrate. It should be noted that "second sphere" residues, that is residues that are close to the active site but not in direct contact with the substrate, have been implicated in BMVO substrate specificity in the past (Reetz & Wu, 2009; van Beek *et al.*, 2012; Zhang *et al.*, 2012). In summary, CPMO has more crowding to the bottom-left of the substrate, but more space to the back-left and below the substrate, than CHMO.

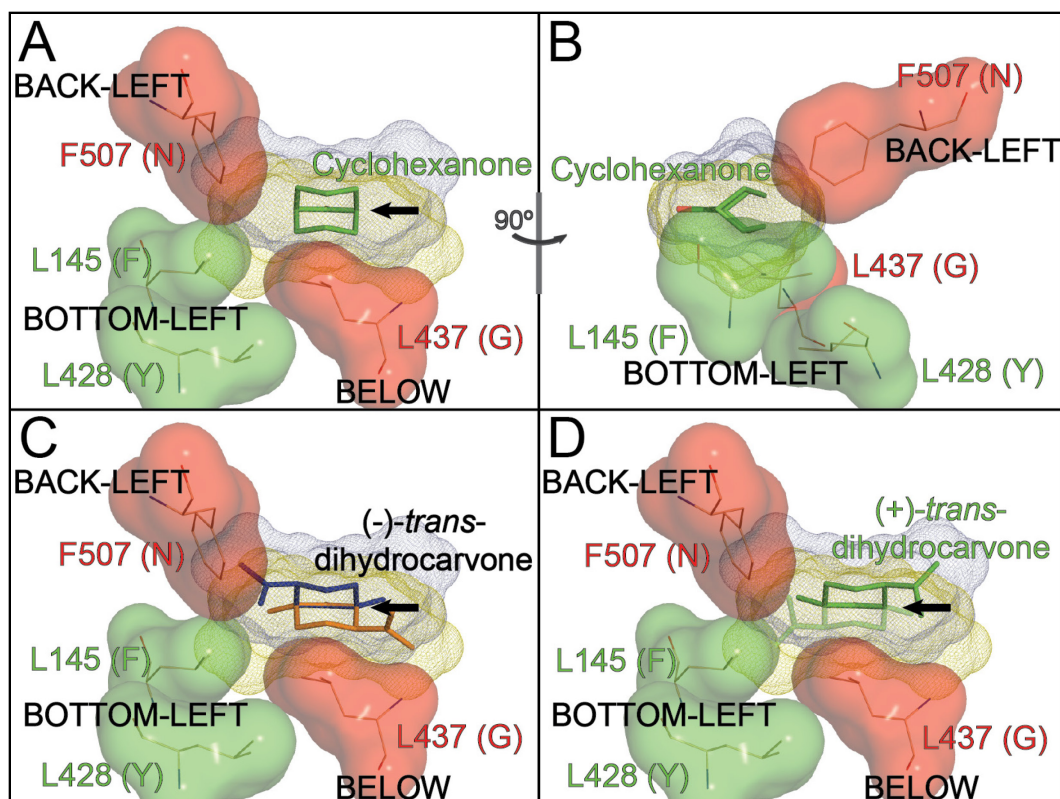


Fig. 3.7: Regiospecificity of CHMO and CPMO for various terpenes. The residues close to the binding site that are altered in CPMO compared to CHMO are shown as surfaces. Those residues that get larger or more rigid in CPMO are shown in green, whereas those that get smaller or more flexible are shown in red. The positions of those residues, as described in the text, are indicated in black. The space taken up by all of the terpenes that are considered, with the isopropyl or isopropenyl group in the equatorial position, is indicated as a blue (up configuration) or yellow (down configuration) mesh. The directionality of the enzyme is indicated with the black arrow. (A) The up and down configurations of cyclohexanone are shown (green sticks). (B) The same image as in (A), but rotated by 90°. (C) The preference of CHMO (blue sticks) and CPMO (orange sticks) for (-)-*trans*-dihydrocarvone is shown. (D) The preference of both CHMO and CPMO for the "up" configuration of (+)-*trans*-dihydrocarvone (green sticks) vs. the "down" configuration (transparent green sticks) is shown. The substrate positions are modelled based on the position of ϵ -caprolactone in the CHMO_{Tight} structure.

Because of the restricted space below the substrate in CHMO, it is expected that for terpenes with the large isopropyl and isopropenyl groups bonded to carbon 5, CHMO will favour the "up" configuration. By contrast, the crowding to the bottom-left of the substrate in CPMO will tend to force the isopropyl/isopropenyl group to the right, while the space below the substrate will allow more access to the "down" configuration. This means that for (-)-*trans*-

dihydrocarvone and (-)-carvomenthone, CHMO will prefer the all equatorial "up" configuration, placing the large group up and to the left, while CPMO will prefer the all equatorial "down" configuration, with the large group down and to the right (Fig. 3.7C). For (-)-*cis*-dihydrocarvone, the same reasoning would apply, as the only difference is the placement of the methyl group in the axial position. Conversely, CHMO can only maintain the isopropenyl group of (+)-*trans*-dihydrocarvone in the all equatorial "up" configuration if it is on the right, and CPMO will also prefer to have it on the right to avoid the crowding on the left (Fig. 3.7D). In summary, all four terpenones should have the large isopropyl or isopropenyl group on the right for CPMO. For CHMO, however, since the "up" configuration is favoured, the isopropyl/isopropenyl group will sit on the right for (+)-*trans*-dihydrocarvone and to the left for the other three. Correspondingly, the abnormal lactone was found to be produced by CPMO in all cases, and in CHMO for (+)-*trans*-dihydrocarvone. The other three terpenones produce the normal lactone for CHMO. In order to obtain these results, the directionality of the enzyme would require the right-hand carbon (as oriented in Fig. 3.7ACD) to migrate in all cases, as predicted for the 4-hydroxycyclohexanone example and as modelled in the crystal structure. We have used a single property, the "directionality" of oxygen incorporation, to explain two distinct stereochemical phenomena: the enantiospecificity towards 4-hydroxycyclohexanone and the regiospecificity towards terpenones. While independent experimental data will ultimately be required to validate this trend, we feel that this provides sufficient justification for modelling the lactone as we have in the CHMO_{Tight} and CHMO_{Loose} crystal structures. As such, it provides valuable insight into the structural basis for the substrate specificity and stereospecificity of the BVMOs.

3.5.4 Transition from CHMO_{Tight} to CHMO_{Rotated}

As previously described (Chapter 2), the Rotated conformation is expected to resemble the enzyme conformation in which the Criegee intermediate would form. The data presented here, however, suggest that substrate acceptance and stereospecificity are determined in the Tight conformation. The question then arises as to how the substrate can be guided from the Tight position to the Rotated

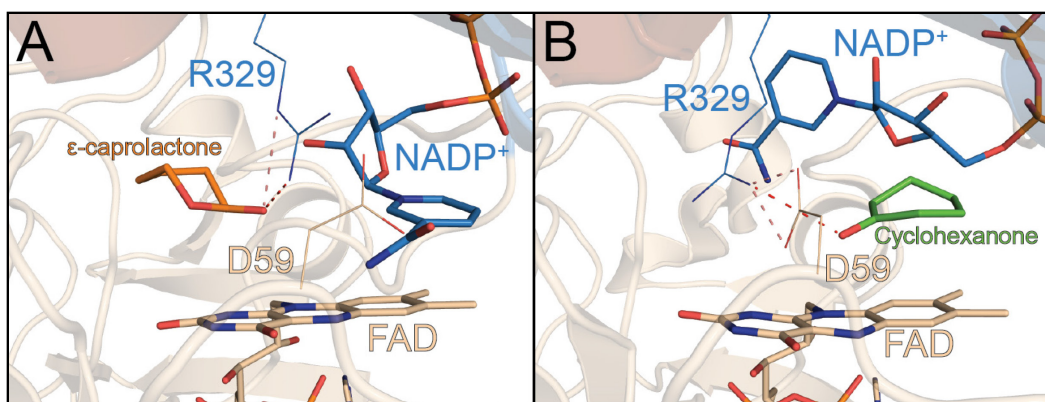


Fig. 3.8: R329 plays a role in guiding the substrate and product from the Tight position to the Rotated position and back. (A) In the CHMO_{Tight} structure, R329 forms a strong (dark red) and weak (light red) hydrogen bonding interaction with ϵ -caprolactone. (B) In the CHMO_{Rotated} structure, the interaction between R329 and cyclohexanone (dark red) and D59 (light red) is shown.

position without reorienting itself. This can be explained by the position of the key residue R329 (Fig. 3.8). In the Tight structure, R329 forms a hydrogen bonding interaction with the carbonyl oxygen of ϵ -caprolactone. A strong interaction is formed with one of the terminal nitrogens, and a weaker interaction is formed with the ϵ -nitrogen. In the Rotated structure, R329 remains at a hydrogen bonding distance from the carbonyl oxygen of cyclohexanone. This suggests that R329 could play a role in guiding the substrate from the Tight position to the Rotated position without giving it the opportunity to reorient itself. It further highlights the important role of the strictly conserved R329 residue in the BVMO enzyme mechanism.

One issue with this interpretation is that while the nitrogen of R329 is within a typical hydrogen bonding distance, it is not oriented in such a way as to form a strong hydrogen bond in the CHMO_{Rotated} structure. Such an interaction would require a slight rotation of the arginine sidechain. This can be rationalized, however, due to the absence of a negative charge that would be produced by the peroxyanion intermediate, and subsequently the Criegee intermediate. In the absence of this negative charge, R329 shifts to form an interaction with the nearby negatively charged D59.

With the enzyme adopting the Tight conformation, the question arises as to whether the Rotated conformation is still necessary. Indeed, the structure of

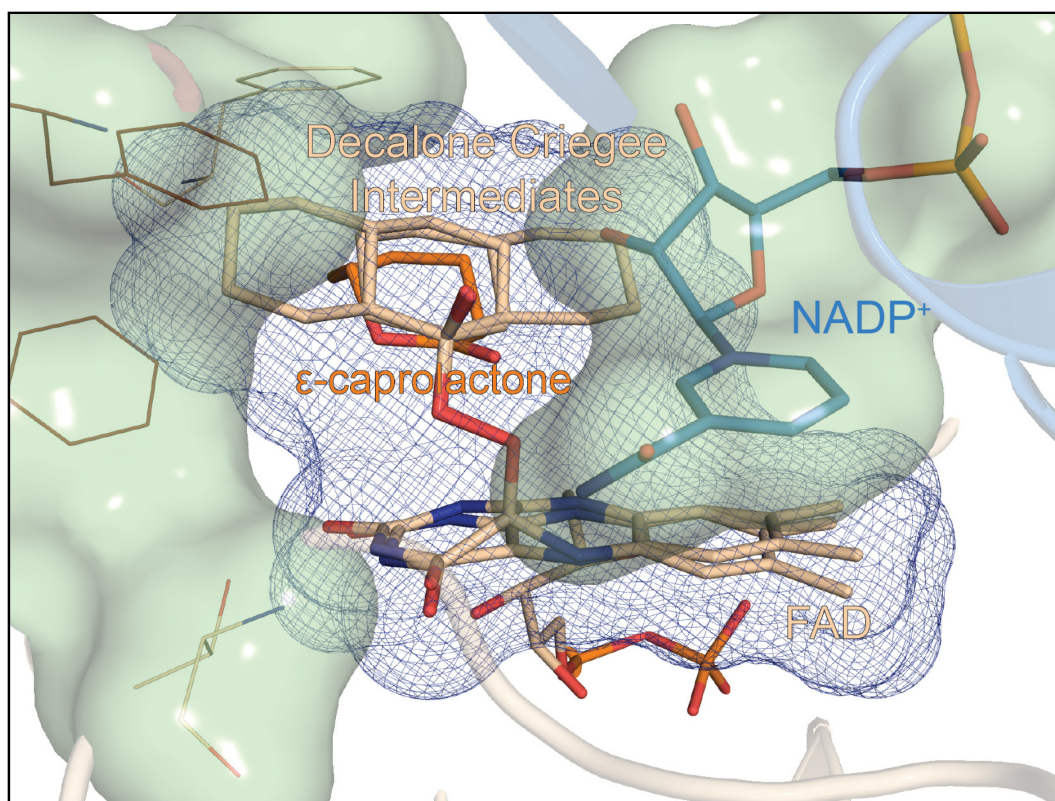


Fig. 3.9: The Criegee intermediate of large substrates like decalone cause clashes in the Tight conformation. Models of the Criegee intermediate of decalone with the second ring in both possible orientations are shown overlaid on the CHMO_{Tight} structure. The space taken up by the intermediates is shown as a blue mesh, and the space taken up by NADP⁺ and other nearby residues is shown as a green surface. The large intermediate clashes with the residues present in the substrate binding pocket.

PAMO bound to the inhibitor MES has previously been used to suggest that the Criegee intermediate would form with the substrate in a Tight-like pocket (Orru *et al.*, 2011), and computational studies have been used to model the reaction with cyclohexanone and its derivatives with small substituents forming the Criegee intermediate at that position (Polyak *et al.*, 2012; Polyak *et al.*, 2013). The drawback of this interpretation is the inability to explain the larger substrates that can be accommodated (see Fig. 1.5). Even if active site plasticity is invoked, which could be used to explain how a large, bicyclic substrate could bind in a Tight-like conformation, if the substrate were to get close enough to actually form the Criegee intermediate, it would clash with either NADP⁺ or the residues on the opposite side of the lactone (Fig. 3.9). For the former possibility, such a mechanism would require NADP⁺ to slide deeper into the enzyme, for which

there is no structural evidence and appears unlikely based on the proximity of the protein surface to NADP^+ on that side. For the latter case, the residues on the opposite side of the lactone are in a relatively immobile region of the protein, and therefore are unlikely to move. In contrast, if the Criegee intermediate is not required to form in a Tight-like conformation, the substrate could move out towards the opening seen in the $\text{CHMO}_{\text{Open}}$ and $\text{CHMO}_{\text{Loose}}$ structures without the constraint of being in close enough proximity to the FAD molecule to form the Criegee intermediate. The Tight-Rotated mechanism appears to be the only available explanation that is consistent with both the high stereospecificity and the ability to accommodate large substrates that is characteristic of CHMO.

3.5.5 The role of the loop in the catalytic cycle

As the large loop is disordered in the $\text{CHMO}_{\text{Loose}}$ structure and ordered in the $\text{CHMO}_{\text{Tight}}$ structure, it is easy to suggest that the ordering of the loop assists the enzyme in adopting the more selective, Tight conformation, thereby restricting the amount of space available to the substrate. It was therefore quite surprising to find that the W492A mutant, which is believed to favour the disordered state of the loop, actually resulted in an enzyme that was more selective, if less active. This suggests that the role of the loop may be to act as a "gatekeeper" to prevent the substrate from diffusing away from the active site. The W492A mutant may be decreasing the frequency of the "gate closing," allowing less preferred substrates or substrate conformations (ie. substrates/conformations with a rapid k_{off} rate) to diffuse away from the active site before the loop closes it off. This may point to the loop playing a more complex role than simply defining the binding site. It also seems to suggest that the substrate specificity and stereospecificity are at least in part determined prior to loop ordering.

It is also readily apparent that in all cases, the W492A mutant is less efficient than the wild-type. There is no isomer of any of the substrates tested that was consumed to a greater extent by the mutant than the wild-type in absolute terms. The proposed lost interaction between NADP(H) and the loop tryptophan should slow down the loop ordering steps, which is projected to play a role in multiple steps in the reaction mechanism (sections 2.5.3 and 3.5.6). This is likely

the cause of the 14% drop in specific activity that was previously reported (Mirza *et al.*, 2009). On top of this, the observed uncoupling of NADPH consumption from product production, combined with the previously reported 3.4-fold increase in the K_i of NADP^+ for the mutant (Mirza *et al.*, 2009), indicates that the mutant is less able to keep NADP^+ bound throughout the catalytic cycle, leading to the unproductive collapse of the peroxyanion intermediate. In spite of the higher specificity, the role of the loop in maintaining an efficient catalyst is clear.

These data could nonetheless suggest that modification of other loop residues could be helpful for protein engineering applications. The W492A mutant appears to alter the balance between high specificity of the enzyme and high efficiency. While the lack of efficiency is clearly of concern, as it would result in a slower biocatalyst that consumes far more NADPH than the wild-type, it is conceivable that altering other residues in the loop could produce an enzyme that is more specific than the wild-type while retaining some of its efficiency. In particular, increasing the strength of interaction between the loop and the adenine portion of NADP^+ while weakening the interactions with the loop near the substrate binding site may result in an enzyme that is able to tightly bind NADP^+ while still favouring a more Loose-like conformation for enough time to allow less preferred substrates to diffuse away from the enzyme. This might reduce uncoupling of NADPH consumption as compared to the W492A mutant while improving selectivity relative to the wild-type. It is clear that further study of this critical loop region, both from a mechanistic perspective and as a potential target site for rational protein modification, is needed.

3.5.6 CHMO_{Loose} and CHMO_{Tight} are consistent with the BVMO structural mechanism

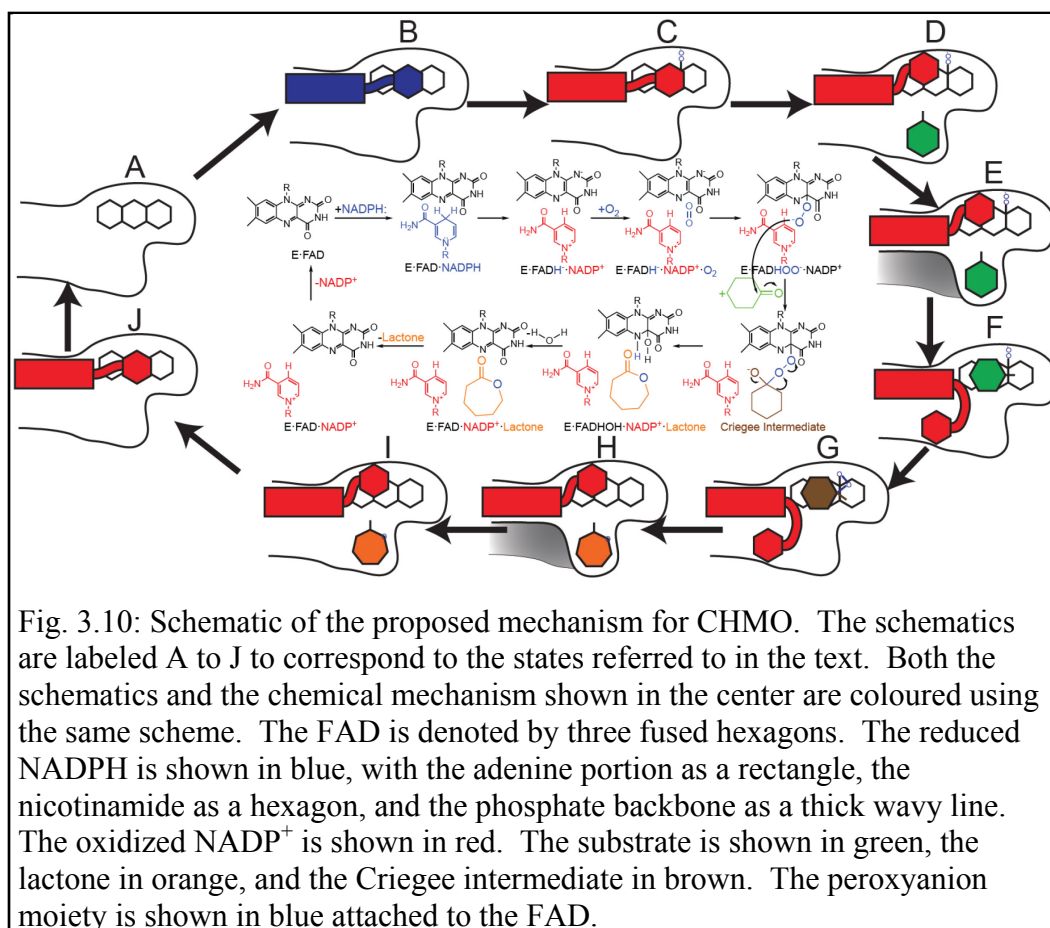
With the report of the CHMO_{Rotated} crystal structure (Chapter 2), it was possible to propose a complete structural mechanism for the BVMOs (see section 2.5.3). This proposed mechanism was able to rationalize both the broad substrate specificity and high stereospecificity of CHMO (see Fig. 1.5). As part of this mechanism, it was proposed that the substrate would bind first in a loose-binding mode characterized by a more open structure with the large loop disordered (*state*

D). This state represented the initial association of the substrate with the enzyme. The enzyme would then transition to a tight-binding mode (*state E*), in which the loop would become ordered and the binding pocket would become more compact. This proposed state was suggested to be the position in which substrate acceptance and stereospecificity would be determined. These two states were expected to resemble the CHMO_{Open} and CHMO_{Closed} structures, but no structures had yet been obtained with ligands at those positions. The enzyme was then proposed to assume the reported Rotated conformation, allowing catalysis to occur. Once the product was formed, the enzyme would revert back through the tight-binding and loose-binding modes (*states H and I, respectively*) before releasing the product.

With the new CHMO_{Tight} and CHMO_{Loose} crystal structures, we now have direct evidence for the Loose and Tight conformations first proposed based on the CHMO_{Rotated} crystal structure (Chapter 2) and the previously determined CHMO_{Open} and CHMO_{Closed} structures (Mirza *et al.*, 2009). The Tight conformation corresponds to the tight-binding conformation of CHMO with the product bound (*state H*), which should be related to the substrate-bound Tight conformation (*state E*). Likewise, the Loose conformation corresponds to the product-bound loose binding state (*state I*), which should be related to the substrate-bound loose binding state (*state D*).

There is one minor alteration to the mechanism proposed in section 2.5.3. In that mechanism, it was believed that the disordered loop states with substrate or product bound (*states D and I*) would have NADP⁺ in the Open-like conformation, sitting above the flavin ring system. This assumption was based on the absence of a crystal structure of a BVMO with the disordered loop and NADP⁺ in the Closed-like conformation. Based on the CHMO_{Loose} structure, we can now revise the mechanism to have NADP⁺ in a Closed-like conformation in *states D and I*.

With the report of these two new crystal structures, we now have a crystal structure with either substrate or product bound for each of the substrate- and product-bound states proposed in the BVMO structural mechanism (Fig. 3.10).



The CHMO_{Rotated} crystal structure corresponds to the Criegee-like states, *states F and G*, which are characterized by a closed off binding pocket that is large enough to accommodate a wide range of larger substrates. The CHMO_{Tight} structure corresponds to the states just prior to and just following the Criegee-like states, *states E and H*, which are characterized by a more specific tight-binding pocket that is not solvent exposed. Finally, the CHMO_{Loose} structure corresponds to the states just prior to and just following the tight-binding states, *states D and I*, which allow a loose association between the enzyme and the substrate or product by virtue of the fact that the loop is disordered, the NADPH-binding domain is rotated away from the active site, and the binding pocket is solvent exposed. This third structure is the only one of the three that has the substrate or product completely solvent exposed and able to associate or dissociate easily with the enzyme. Evidently, this assumes that the product-associated CHMO_{Tight} and CHMO_{Loose} structures are structurally analogous to the substrate-bound states with

the peroxyanion intermediate present, and likewise that the CHMO_{Rotated} structure is not substantially perturbed by the absence of the peroxyanion intermediate and/or the ketone-peroxyflavin adduct.

In obtaining direct evidence for the six critical solvent- and product-associated enzyme states, these new structures are consistent with our proposed structural mechanism. They allow us to simultaneously rationalize the broad substrate specificity, the high stereospecificity, and the ability of the enzyme to bind and release the substrate and product with NADP(H) bound throughout the catalytic cycle. Most importantly, this can be done using crystal structures of the same enzyme that have either a substrate or a product bound at every step.

3.6 Conclusion

The CHMO_{Loose} and CHMO_{Tight} crystal structures provide the first examples of crystal structure of BVMOs with a true substrate or product bound in the position initially predicted based on the CHMO_{Closed} structure. They reinforce the structural mechanism for catalysis that we proposed in Chapter 2, and also allow a model for the stereospecificity of the enzyme to be proposed. While structures involving large or asymmetric substrates or products are still needed to fully understand the substrate profile and stereospecificity of the BVMOs, these structures provide an important foundation for future protein design studies.

In addition, the role of protein dynamics in the complex catalytic cycle is re-emphasized. The domain rotations that have been observed when comparing previous crystal structures are highlighted again when these two new structures are added to the mix. Furthermore, the role of the large loop can be expanded. There is increased evidence that the loop undergoes multiple order-disorder transitions per catalytic cycle. In addition, evidence for the loop playing a role in determining the substrate profile and stereospecificity of the enzyme has been provided, which highlights the need for future studies on the role of the loop in catalysis.

Chapter 4: Baeyer-Villiger Monooxygenase Loop Dynamics

Examined by SAXS and NMR

4.1 Preface

That the BVMOs are structurally dynamic enzymes was first proposed based on the initial PAMO structure, and was corroborated by the differences between the CHMO_{Open} and CHMO_{Closed} structures. The data presented in Chapters 2 and 3 continue to emphasize the importance of the dynamic properties of CHMO to its structural mechanism. The changing state of the large, twenty residue loop is clearly important, and is likely linked to the rotation of the NADPH-binding domain.

While X-ray crystallography can highlight protein dynamics by providing "snapshots" of different states of the enzyme, a more complete understanding of protein dynamics requires the use of solution techniques. In this chapter, SAXS and NMR studies are employed to study the dynamic properties of CHMO and CPDMO, with particular emphasis on the twenty residue loop. The changes in structure, in particular triggered by the addition of NADP⁺, can further explain the dynamic properties of the enzyme that were apparent based on the crystal structures.

4.2 Introduction

The Baeyer-Villiger monooxygenases (BVMOs) are a group of bacterial proteins that are able to catalyze the synthetically useful Baeyer-Villiger oxidation (BVO) reaction, which involves the conversion of ketones into esters (Baeyer & Villiger, 1899; Baeyer & Villiger, 1900). Since their discovery in the 1970s (Donoghue & Trudgill, 1975; Donoghue *et al.*, 1976), these enzymes have been increasingly looked upon as potential biocatalysts. This has led to the extensive characterization of these proteins, especially the prototypical BVMO, cyclohexanone monooxygenase (CHMO) from *Acinetobacter* sp. NCIMB 9871 (*AcCHMO*). The mechanism of this enzyme was extensively studied as early as the 1980s (Ryerson *et al.*, 1982). From this study, as well as a later mechanistic study of *AcCHMO* (Sheng *et al.*, 2001), the BVMOs were found to have a complex, highly ordered enzyme mechanism. The first step in this reaction cycle

involves the binding of NADPH, which reduces the tightly bound FAD molecule in the core of the protein. Molecular oxygen then binds to the enzyme and reacts with the reduced flavin to form the peroxyanion intermediate. The ketone substrate binds last, reacts with the peroxyanion intermediate, and forms the Criegee intermediate, which is analogous to the Criegee intermediate that is formed in the chemical BVO reaction (Criegee, 1948). The Criegee intermediate then rearranges to form the lactone product. Following dehydration of the flavin, the lactone is released back to the solvent, and finally NADP^+ is released. This ordered mechanism was also found in a second BVMO, phenylacetone monooxygenase (PAMO) (Torres Pazmiño *et al.*, 2008a), suggesting that it is likely to be a general mechanism for all BVMOs.

The first crystal structure of a BVMO was reported a decade ago, with the publication of the FAD-bound structure of PAMO (Malito *et al.*, 2004). Even at this early stage, the importance of protein flexibility was apparent, as a strictly conserved residue, R337, was found to be in multiple conformations that were believed to be associated with domain rotations of the enzyme. In addition, a large flexible loop was observed to be sticking out to contact a symmetry-related protein molecule in the crystal structure. This interaction was believed to be a crystallization artifact, implying that the loop was unstructured in solution.

This structure was followed by two NADP^+ -bound crystal structures of CHMO from *Rhodococcus* sp. HI-31 (*RmCHMO*) (Mirza *et al.*, 2009). These structures confirmed the domain rotations and introduced additional conformations in the conserved arginine residue. In addition, the large loop was found to be disordered in the first structure, known as $\text{CHMO}_{\text{Open}}$, and folded in and contacting the NADP^+ cofactor in the second structure, known as $\text{CHMO}_{\text{Closed}}$. This ordered conformation of the loop was unlikely to be a crystallization artifact, as mutation of W492, which interacts with the NADP^+ molecule, resulted in an enzyme with drastically reduced activity.

More recently, crystal structures of PAMO with NADP^+ bound have corroborated the loop order-disorder transition observed in *RmCHMO*, as well as the observed domain rotations (Orru *et al.*, 2011). Crystal structures of 2-Oxo- Δ^3 -

4,5,5-Trimethylcyclopentenylacetyl-Coenzyme A Monooxygenase (OTEMO) have demonstrated alternate loop configurations that are required to accommodate the very large, coenzyme A linked substrates of this enzyme (Leisch *et al.*, 2012). Finally, substrate-bound (Chapter 2) and product-bound (Chapter 3) crystal structures of *Rm*CHMO, known as CHMO_{Rotated}, CHMO_{Loose}, and CHMO_{Tight}, highlighted the multiple conformations of NADP⁺, the domain rotations, and the requirement for the large loop to undergo multiple order-disorder transitions per catalytic cycle. Intriguingly, the W492A mutant was also found to increase the stereospecificity of the enzyme, albeit at the cost of enzyme efficiency, highlighting the importance of this loop (section 3.5.5).

It is clear that the increasing number of crystal structures of BVMOs provides us with a wealth of knowledge regarding the dynamic properties of this enzyme. With X-ray crystallography, however, we are limited to looking at "snapshots" of the enzyme without direct experimental evidence as to how these enzymes states are related to one another at various points in the reaction cycle.

In order to address this limitation, we have pursued small-angle X-ray scattering (SAXS) experiments and nuclear magnetic resonance (NMR) experiments to study the dynamic properties of *Rm*CHMO and a second BVMO, cyclopentadecanone monooxygenase (CPDMO) (Iwaki *et al.*, 2006). These studies allow us to examine the dynamic properties of these enzymes, in particular considering the large loop. The changes in flexibility of this loop, in particular with the addition of NADP⁺, can be linked to the enzyme mechanism and the available crystal structures to give us a more complete picture of the properties of the BVMO family of proteins.

4.3 Materials and Methods

4.3.1 Subcloning the *cpdB* gene into the pJW234 expression vector

The *cpdB* gene was amplified using *Pwo* DNA polymerase (Roche) from the pCD201 vector (Iwaki *et al.*, 2006). The following PCR primers with *Nsi*I and *Eco*RI restriction sites were used to amplify the gene (restriction sites are underlined): Sense—5'-CAGGAAACAATGCATATGAGTCAGCTAATTCAAG-3'; Antisense—5'-CACAACGGAATTCTCAACGCTTGCG-3'. The desired

band was purified using the QIAquick Gel Extraction Kit (Qiagen) and subcloned into the pPCR-Script Cam vector (Stratagene) using the PCR-Script Cam Cloning Kit (Stratagene). Plasmid DNA from positive clones was prepared using the QIAprep Spin MiniPrep kit (Qiagen), and then digested using *NsiI* and *EcoRI* (New England BioLabs). The pJW234 expression vector (obtained from Dr. Mirek Cygler, University Saskatchewan, Saskatoon, Canada) was digested with the same enzymes. The desired bands from both digests were purified using gel extraction (MO BIO Laboratories UltraClean 15 DNA Purification Kit). The *cpdB* insert was ligated into the pJW234 vector, and this plasmid was used to transform *E. coli* DH5 α cells. Plasmid DNA from positive clones was prepared using the QIAprep Spin MiniPrep kit (Qiagen), and DNA sequencing was used to confirm the sequence (McGill University and Genome Quebec Innovation Centre Sequencing Service, Montreal, Canada). The resultant plasmid is called His₈-TEV-CpdB.

4.3.2 Small-Angle X-ray Scattering

4.3.2.1 Protein Expression and Purification

CHMO was expressed and purified as previously described (section 3.3.1). For CPDMO, an identical expression and purification protocol was used, excepting that the His₈-TEV-CpdB plasmid was used instead of His₈-TEV-ChnB1. In addition, the lysis buffer and purification buffers were all supplemented with 1 mM dithiothreitol for the CPDMO purification. Following gel filtration, the samples were concentrated to approximately 15 mg/mL or 3 mg/mL, depending on the application.

4.3.2.2 Sample Preparation and Data Collection

Samples of CHMO and CPDMO were prepared for SAXS data collection by adding appropriate amounts of ligands to the protein sample (~15 mg/mL protein stock solutions), and then diluting the sample to 10 mg/mL. These samples were microdialysed overnight against buffer with identical ligands, but without any protein. For CPDMO samples, 2 mM Tris(2-carboxy)phosphine (TCEP) was also added to both the samples and the buffer. Ligand concentrations for each sample are given in Table 4.1.

Table 4.1: Ligand concentrations for SAXS data collection.

Sample	[NADP ⁺] (mM)	[Ligand] (mM)	TCEP (mM)
CHMO (No ligands)	0	None	0
CHMO+NADP ⁺	5	None	0
CHMO+NADP ⁺ +Substrate	5	5 (Cyclohexanone)	0
CHMO+NADP ⁺ +Product	5	5 (ϵ -caprolactone)	0
CPDMO (No ligands)	0	None	2
CPDMO+NADP ⁺	5	None	2
CPDMO+NADP ⁺ +Substrate	5	1 (Cyclopentadecanone)	2
CPDMO+NADP ⁺ +Product	5	1 (ω -pentadecalactone)	2

Following dialysis, the samples were quantified using UV absorbance. Using the dialysis buffer, each sample was diluted 1:2 and 1:4. Data were collected at the SIBYLS beamline (beamline 12.3.1) at the Advanced Light Source using 0.5, 1.0, 2.0, and 4.0 second exposures. Data were collected from dialysis buffer with the same exposure times, and these buffer scattering curves were used for buffer subtraction.

In addition, a titration of CHMO with NADP⁺ was prepared using a series of CHMO samples with increasing NADP⁺ concentrations. Microdialysis was performed as previously described, but by diluting a ~3 mg/mL stock solution to 2 mg/mL, with NADP⁺ concentrations ranging from 0 mM to 10 mM. Following dialysis, the protein samples were quantified, and the absorbance of the buffer at 260 nm was used to quantify the exact concentration of NADP⁺ in the samples. Data were collected on a SAXSess mc² nanostructure analysis system equipped with an Anton Paar GmbH detector for both the dialysed sample and the dialysis buffer. All data collection was performed at 4°C. For each sample and buffer, a series of 360 10 second exposures were measured and then averaged. In addition, dark current images of the same length were measured.

4.3.2.3 BVMO Ligand Comparison Primary Data Analysis

The data from each of the four sets of scattering curves obtained from the Advanced Light Source were scaled to the protein concentration. Curves in which the detector was saturated or the data exhibited pronounced concentration

dependent effects were omitted from the analysis. The remaining curves were merged in PRIMUS (Konarev *et al.*, 2003) such that noisy data at higher q-ranges and interparticle repulsion at low q-ranges were eliminated from the final, merged scattering curves. The curves were then normalized to $I(0)$. The radius of gyration (R_g) was determined using the Guinier plot in PRIMUS (Konarev *et al.*, 2003). The particle distance distribution function was then evaluated for each merged, normalized scattering curve using GNOM (Svergun, 1992). The Porod volume and Porod exponent were determined using ScÅtter (Rodriguez & Rambo, www.bioisis.net). The four scattering curves for each protein were compared using the volatility ratio mode with the SAXS Similarity Map tool (Hura *et al.*, 2013). SAXS data collection and analysis statistics are presented in Table 4.2 and Table 4.3.

Table 4.2: SAXS data collection and analysis statistics for CHMO

	CHMO	CHMO + NADP ⁺	CHMO + NADP ⁺ + cyclohexanone	CHMO + NADP ⁺ + ϵ -caprolactone
Guinier R_g (Å)	26.94±0.06	25.95±0.14	25.39±0.05	25.33±0.04
Real space R_g (Å) [*]	27.43±0.03	26.08±0.02	25.70±0.02	25.59±0.01
D_{max} (Å) [*]	93	80	78	75
Porod volume (Å ³) [†]	1.1×10^5	9.9×10^4	1.0×10^5	9.9×10^4
Porod exponent [†]	3.7	3.9	3.8	3.8
Total [*]	0.696	0.735	0.702	0.672
EOM χ^2 [‡]	4.3	1.3	N/D	N/D

^{*}Calculated by GNOM (Svergun, 1992).

[†]Calculated by ScÅtter (Rodriguez & Rambo, www.bioisis.net).

[‡]Calculated by EOM using the CHMO_{Closed} model in random mode without any ligands (Bernadó *et al.*, 2007).

Table 4.3: SAXS data collection and analysis statistics for CPDMO

	CPDMO	CPDMO + NADP ⁺	CPDMO + NADP ⁺ + cyclopentadecanone	CHMO + NADP ⁺ + ω -pentadecanone
Guinier R_g (Å)	27.95±0.16	26.58±0.05	26.34±0.05	26.22±0.05
Real space R_g (Å) [*]	28.44±0.07	26.22±0.02	26.12±0.02	26.00±0.02
D_{max} (Å) [*]	104	87	90	90

	CPDMO	CPDMO + NADP ⁺	CPDMO + NADP ⁺ + cyclopentadecanone	CHMO + NADP ⁺ + ω -pentadecalactone
Porod volume (\AA^3) [†]	1.1×10^5	1.2×10^5	1.1×10^5	1.0×10^5
Porod exponent [†]	4	3.5	3.7	4
Total [*]	0.614	0.704	0.683	0.681

^{*}Calculated by GNOM (Svergun, 1992).

[†]Calculated by ScÅtter (Rodic & Rambo, www.bioisis.net).

4.3.2.4 NADP⁺ Titration Data Analysis

Data were analysed using SAXSquant (Anton Paar). The dark current was subtracted from all sample and buffer scattering curves, which were then normalized to the intensity of the attenuated direct beam. The appropriate buffer curve was then subtracted from the sample scattering curves. A Porod plot was constructed in order to evaluate any constant background scattering, and this constant was subtracted from the scattering curve as appropriate. The data were then truncated to a q -range of 0.17 nm^{-1} to 2.00 nm^{-1} and desmeared. The R_g was determined using the Guinier plot in PRIMUS (Konarev *et al.*, 2003).

4.3.2.5 Ensemble Modelling of the Loop

The program EOM (Bernadó *et al.*, 2007) was used to generate an ensemble model of possible loop configurations based on the SAXS scattering curves obtained for CHMO in the presence and absence of NADP⁺. Starting with the CHMO_{Open} and CHMO_{Closed} crystal structures (Mirza *et al.*, 2009), two "domains" were defined. The N-terminal domain consisted of residues 6-486, and ensemble models were built with no cofactors, with FAD only, and with FAD and NADP⁺ included in the N-terminal domain. The C-terminal domain consisted of residues 512-534 for CHMO_{Closed} and 505-534 for CHMO_{Open}. All non-protein, non-cofactor atoms were removed. EOM was run with the N- and C-terminal domains fixed for both the CHMO_{Open} and CHMO_{Closed} models. A pool of 10,000 random models was generated based on the CHMO sequence and the two fixed domains using both random and native mode, and a genetic algorithm was run fifty times to select the ensembles of structures from that pool that best represent

the SAXS scattering curves obtained in the presence and absence of NADP⁺. This process was repeated ten times for each model. The R_g distribution data were averaged across the fifty genetic algorithms, and then across the ten EOM runs. The EOM data analysed using the CHMO_{Closed} model without any ligands in random mode are shown, as they had the best fit to the experimental data.

4.3.3 Nuclear Magnetic Resonance

4.3.3.1 Protein Expression and Purification

E. coli Rosetta (DE3) (Novagen) cells were transformed with His₈-TEV-ChnB1 plasmid and the strain was grown in 50 mL of LB supplemented with 100 µg/mL of ampicillin and 30 µg/mL of chloramphenicol and grown for 16-18 hours at 37°C. From this culture, 10 mL were used to subculture into fresh N5052 autoinduction media (Studier, 2005), using 25 mM ¹⁵NH₄Cl. Instead of using purified vitamins to supplement the media, half of the soluble portion of a Centrum tablet was added to the media. In addition, the media was supplemented with 100 µg/mL of ampicillin, 30 µg/mL of chloramphenicol, and 5 µM riboflavin, which is required for FAD synthesis. The culture was grown at 37°C for about 7 hours, at which point the OD₆₀₀ had reached ~0.7. The culture was then transferred to 22°C and allowed to grow for another 16-18 hours.

The harvesting, sonication, and purification procedures were performed as previously described (section 3.3.1). Following gel filtration, the protein was concentrated to about 35 mg/mL.

4.3.3.2 Sample Preparation

Four samples were prepared for NMR. All four samples contained 30 mg/mL (0.5 mM) CHMO in 50 mM Tris pH 8.0 and 10% D₂O. One sample contained no additional ligands. A second sample contained 10 mM NADP⁺, while the third and fourth samples contained 10 mM NADP⁺ in addition to 5 mM of either cyclohexanone or ε-caprolactone.

4.3.3.3 Data Collection

All NMR data were collected on Varian INOVA 800 MHz and 500 MHz spectrometers equipped with a high-sensitivity cold probe at the Quebec/Eastern Canada High Field NMR Facility at 14°C. Prior to each CPMG experiment, a ¹H-

^{15}N HSQC experiment was performed. Constant relaxation time CPMG experiments for NH with proton decoupling were performed on all samples using both the 800 MHz and 500 MHz spectrometer. A constant inter-scan relaxation delay of 40 ms was used, with ν_{CPMG} values of 0, 25, 50, 75, 100, 125, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, and 1000 Hz, with the 100 and 800 Hz experiments repeated. A cold probe compensation pulse was applied to account for ν_{CPMG} -dependent differences in cold probe sensitivity. The intensities of those peaks with narrow linewidths and large intensities in each set of spectra were analysed using the program NESSY (Bieri & Gooley, 2011) in order to calculate transverse relaxation rates (R_2). Using NESSY (Bieri & Gooley, 2011), the dispersion curves were then fitted to three exchange models:

1. The no exchange model

$$R_2^{\text{eff}} = R_2^0 \quad (\text{equation 4.1})$$

2. The two-state, fast exchange model (Meiboom equation) (Luz & Meiboom, 1963)

$$R_2^{\text{eff}} = R_2^0 + \frac{\Phi}{k_{\text{ex}}} \left[1 - \frac{4\nu_{\text{CPMG}}}{k_{\text{ex}}} \tanh\left(\frac{k_{\text{ex}}}{4\nu_{\text{CPMG}}}\right) \right]$$

where

$$\Phi = p_a p_b \partial \omega^2 \quad (\text{equation 4.2})$$

3. The two-state, slow exchange model (Carver-Richards equations) (Carver & Richards, 1972)

$$R_2^{\text{eff}} = R_2^0 + \frac{k_{\text{ex}}}{2} - \nu_{\text{CPMG}} \cosh^{-1} [D_+ \cosh(\eta_+) - D_- \cosh(\eta_-)]$$

where

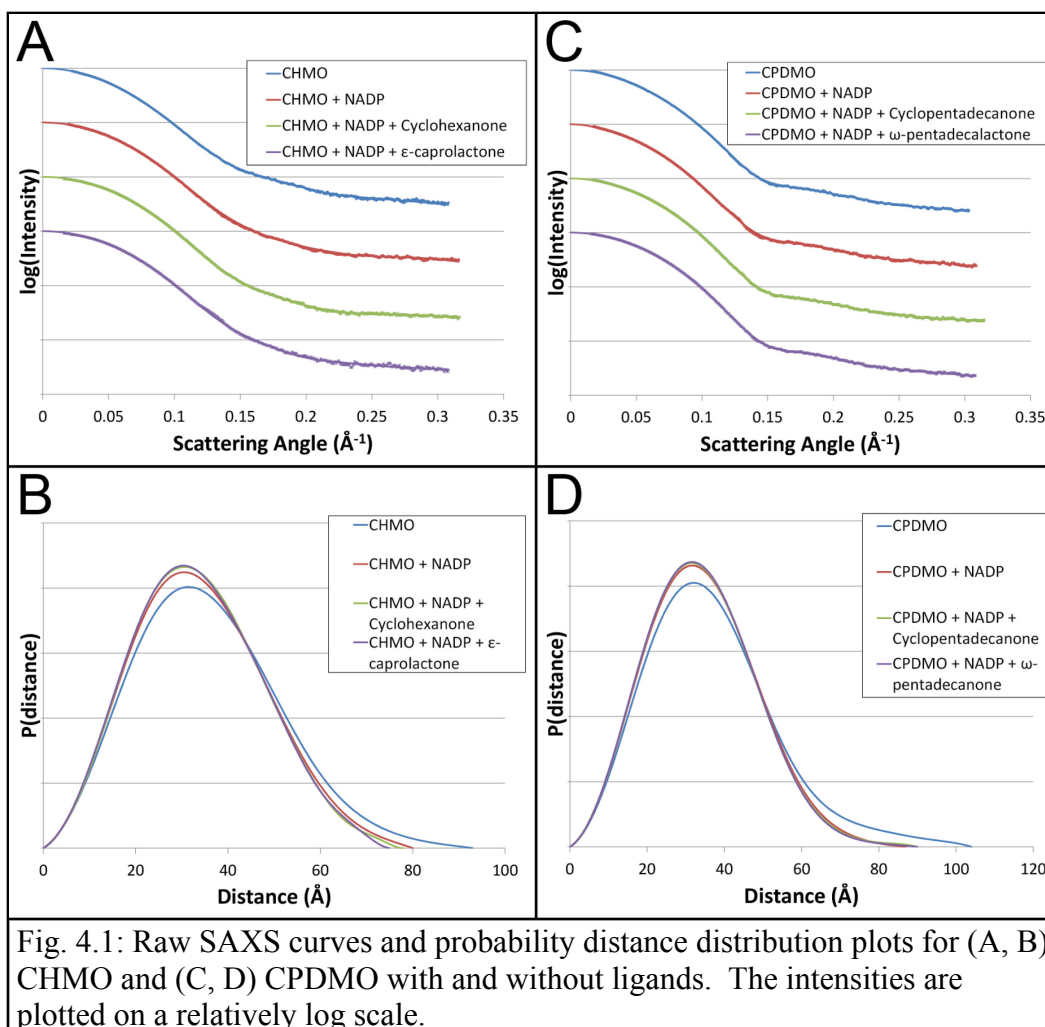
$$D_{\pm} = \frac{1}{2} \left[\pm 1 + \frac{\Psi + 2\partial \omega^2}{(\Psi^2 + \xi^2)^{1/2}} \right]$$

$$\eta_{\pm} = \frac{[\pm \Psi + (\Psi^2 + \xi^2)^{1/2}]^{1/2}}{2\sqrt{2}\nu_{\text{CPMG}}}$$

$$\Psi = k_{\text{ex}}^2 - \partial \omega^2$$

$$\xi = -2\partial \omega (p_a k_{\text{ex}} - p_b k_{\text{ex}})$$

(equation 4.3)

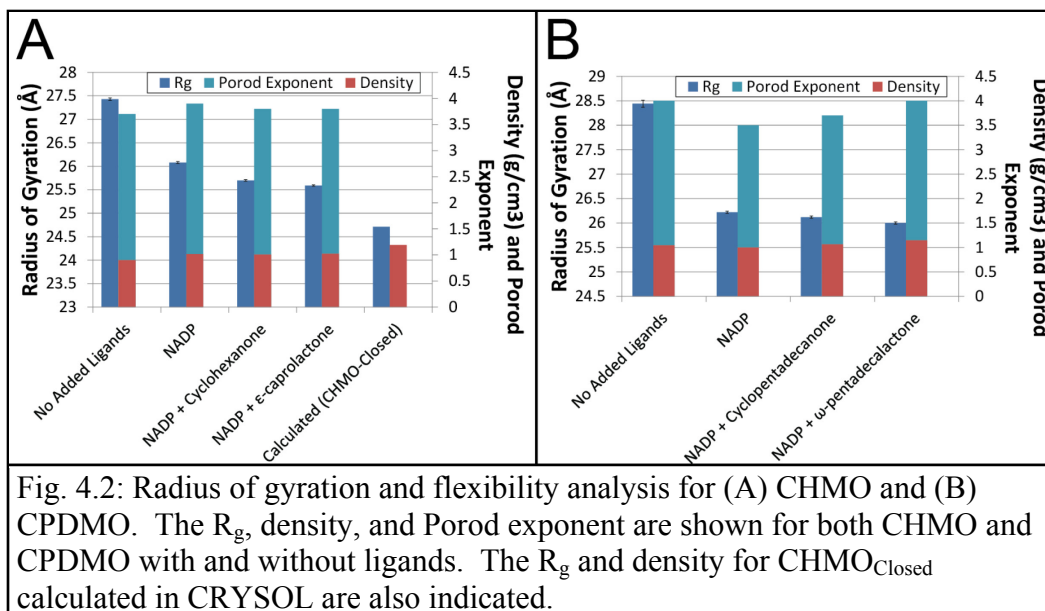


For each signal, the model that best described the data was selected.

4.4 Results

4.4.1 NADP^+ Decreases the R_g of BVMOs

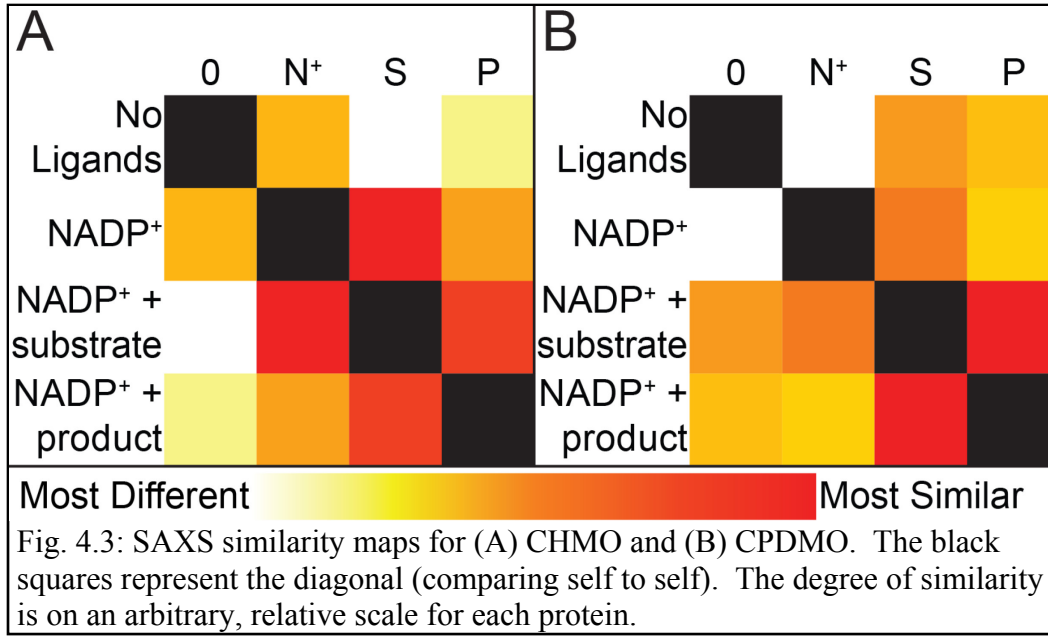
Examination of the four scattering curves for both CHMO and CPDMO revealed that there were no large-scale conformational changes that occurred upon addition of NADP^+ , substrate, or product (Fig. 4.1AC). Based on the particle distance distribution functions calculated from those curves (Fig. 4.1BD), however, it was possible to detect changes in the overall particle size based on the R_g and D_{max} . For both CHMO and CPDMO, addition of NADP^+ triggered a significant drop in both R_g and D_{max} . Further addition of either the ketone substrate or lactone product did result in a further decrease in R_g , but to such a small extent as to not be significant (Fig. 4.2). Of note, the smallest R_g for



CHMO is still significantly higher than that calculated from the CHMO_{Closed} structure using CRY SOL (Svergun *et al.*, 1995).

The SAXS similarity maps in volatility ratio mode (Fig. 4.3) allow the scattering curves to be compared on a relative scale without bias towards the high intensity, low q -range region (Hura *et al.*, 2013). The similarity maps for both CHMO and CPDMO show that in general, the ligand free scattering curve is the most distinct from the rest, though this effect is much more pronounced with CHMO than with CPDMO. Among the curves with NADP⁺ present, for both CHMO and CPDMO, the greatest difference lies between the sample with NADP⁺ only and the sample with NADP⁺ and product.

To show that this change is related to the degree of binding of NADP⁺, a series of SAXS curves were measured with NADP⁺ titrated in. A gradual decrease in the R_g allowed the generation of a titration curve, from which the dissociation constant (K_d) could be determined by fitting the data to the general ligand binding equation (equation 4.4, see Fig. 4.4).



$$\Delta R_g = \frac{\Delta R_{g,\max}}{[M]_T} \times \frac{1}{2} \left[([M]_T + [L]_T + K_d) - \sqrt{([M]_T + [L]_T + K_d)^2 - 4[M]_T[L]_T} \right]$$

where

$-\Delta R_g = R_g - R_g$ in the absence of NADP⁺

$[M]_T$ = The total concentration of CHMO

$[L]_T$ = The total concentration of NADP⁺

K_d = The dissociation constant

$\Delta R_{g,\max} = \Delta R_g$ at infinite NADP⁺ concentration

(equation 4.4)

The K_d of NADP⁺ obtained from this equation, $330 \pm 140 \mu\text{M}$, is roughly equivalent to that obtained previously based on the K_i of NADP⁺ (Mirza *et al.*, 2009, and section 2.4.4), suggesting that the changes in R_g are related to the degree of binding of NADP⁺. Of note, this is the first time that the K_d of NADP⁺ has been measured directly for *Rm*CHMO.

The observed differences between the samples upon addition of NADP⁺, and in particular the reduction in R_g and D_{\max} , imply that the protein either becomes more compact or exhibits a higher degree of flexibility. In order to assess whether the changes in R_g are caused by an increase in the flexibility of the protein, the density of the particles for each scattering curve was assessed using the Porod-Debye law (Fig. 4.2) (Rambo & Tainer, 2011). This analysis permits the estimation of the volume, and therefore the density, of the particle from the

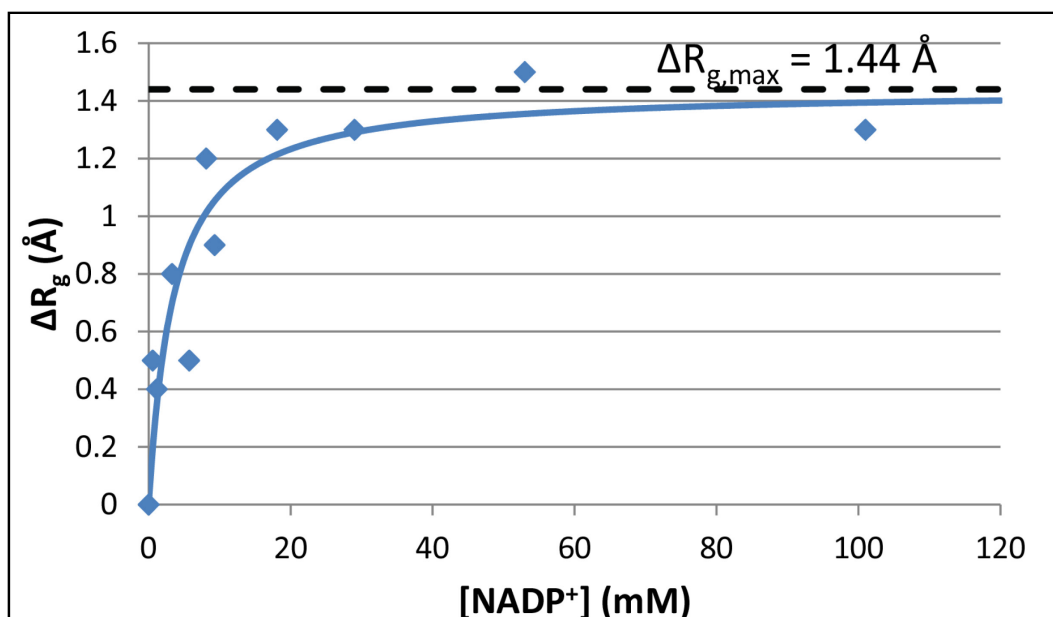


Fig. 4.4: Change in R_g of CHMO with the titration of NADP^+ . The data are fit to the general ligand binding equation (equation 4.4). The $\Delta R_{g,\text{max}}$ asymptote, determined by fitting the data to equation 4.4, is indicated by the dashed line.

SAXS scattering curve. A protein that is more flexible should appear to have a smaller density than a rigid protein. While some variability in the density was observed, this variability is relatively small and does not correlate with the change in the R_g of the particle, indicating that it is likely due to the imprecision in determining the particle volume using SAXS techniques. In addition, the particle density calculated based on the $\text{CHMO}_{\text{Closed}}$ crystal structure is higher than those obtained from the SAXS scattering curves, and approaches the experimentally-determined density of proteins (Charlwood, 1957; Rambo & Tainer, 2011). Flexibility can also be assessed using the Porod exponent (Rambo & Tainer, 2011), with an exponent of 4.0 indicating a rigid structure and a Porod exponent of 2.0 indicating a completely disordered structure. The Porod exponents for all four scattering curves were between 3.5 and 4.0, indicating a consistently fairly rigid particle, for both CHMO and CPDMO. These observations suggest that while there is some degree of flexibility for both BVMOs in the presence and absence of ligands, the flexibility does not appear to vary greatly upon addition of NADP^+ , substrate, or product. This implies that the change in R_g and D_{max} observed upon addition of NADP^+ cannot be explained by changes in protein flexibility alone, and must therefore involve a conformational change.

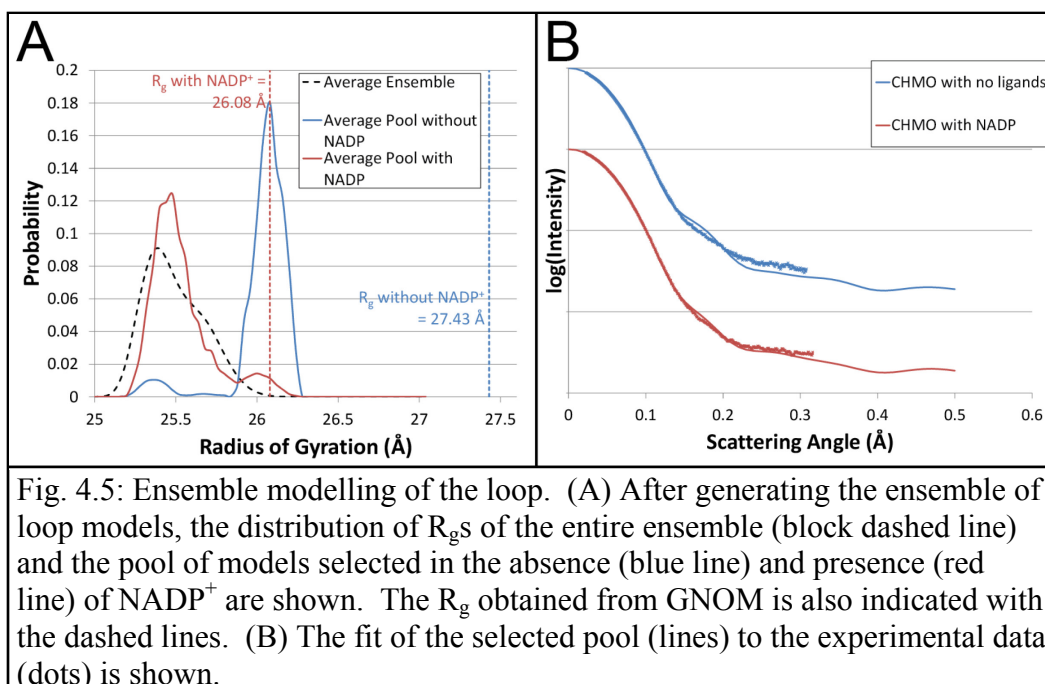


Fig. 4.5: Ensemble modelling of the loop. (A) After generating the ensemble of loop models, the distribution of R_g s of the entire ensemble (block dashed line) and the pool of models selected in the absence (blue line) and presence (red line) of NADP^+ are shown. The R_g obtained from GNOM is also indicated with the dashed lines. (B) The fit of the selected pool (lines) to the experimental data (dots) is shown.

4.4.2 Loop Positioning Can Account for the R_g Change

A variety of conformational changes could account for the increased compactness observed upon addition of NADP^+ . By using ensemble modelling of the loop with EOM, it is possible to determine if changes in the loop structure are able to account for this change in structure. In all of the EOM runs, a similar pattern emerged. When fit to the CHMO with NADP^+ scattering curve, the R_g distribution has a major peak at around 25.5 Å, slightly below the value of 26.1 Å obtained from GNOM (Fig. 4.5). In many cases, there was a second, smaller peak closer to 26.0 Å, suggesting a bimodal distribution of conformations. In all cases, the χ -values were in the 1.0 - 1.5 range, and χ^2 -free values (Rambo & Tainer, 2013) were consistent with the χ -values, indicating a good fit with the experimental data. This was also observed based on visual comparison of the calculated scattering curve based on the ensemble and the experimental scattering curve.

When the ensemble was selected based on the scattering curve of CHMO in the absence of any ligands, the inverse pattern was observed. In all cases, a sharp R_g peak was observed around 26.0 Å, roughly coinciding with the minor peak described above (Fig. 4.5). This lies near the upper extreme of possible R_g values generated in the pool. In some cases, a second minor peak was observed at

around 25.5 Å, suggesting an inversion of the bimodal distribution observed in the sample with NADP⁺. In addition, the χ -values were significantly higher. In most cases, they were between 4.0 - 5.0, but when NADP⁺ was included in the model, this rose to 6.0 - 7.0. Once again, these values were consistent with the χ^2 -free values (Rambo & Tainer, 2013). It should be noted, however, that the agreement between the calculated and experimental scattering curves was reasonably good in the low q-range, with significant deviations occurring at $q > 0.15 \text{ Å}^{-1}$.

4.4.3 The Loop Behaves as a Peptide in Solution

In general, it is difficult to obtain NMR spectra of large proteins due to the large number of signals and the broad linewidths of those signals, leading to a spectrum with heavily overlapping weak signals. It was somewhat surprising that for CHMO, a number of sharp, strong signals were clearly visible above the noise of the majority of the protein signals. While some of these residues are caused by the flexible termini, there was evidence that a number of these signals were derived from the loop residues behaving as a small peptide in solution. A TOCSY spectrum revealed a number of spin systems that were consistent with the residues that are present in the loop. HSQC spectra revealed signals corresponding to Asn and Trp sidechains, both of which are present in the loop.

In order to probe the behaviour of these residues, a series of HSQC spectra were recorded. The first, which was recorded for the protein in the absence of any ligands, revealed approximately twelve strong signals. When NADP⁺ was added to the sample, the number of signals roughly doubled to approximately 20, with a number of signals shifting and/or splitting into two or more signals. The addition of cyclohexanone or ϵ -caprolactone to the sample did not change the number of signals substantially, but did lead to changes in both the chemical shifts of those signals as well as the splitting of those signals. As the addition of NADP⁺ is unlikely to alter the chemical environment of terminal residues, it is reasonable to conclude that many of these strong signals arise from the loop behaving as a small peptide in solution.

4.4.4 Loop Dynamics

Given that the loop is behaving as a peptide and should be fairly flexible, it should be possible to obtain dynamics information using CPMG experiments. With each of the four samples (no ligands, NADP⁺ only, NADP⁺ and cyclohexanone, and NADP⁺ and ϵ -caprolactone), CPMG data were recorded at two magnetic field strengths. In the sample without any ligands, three of the visible signals showed very clear CPMG dispersion curves (Fig. 4.6ABC). All three of these curves could be fitted to the two-state Carver-Richards equations (equation 4.3). The parameters obtained (Table 4.4) indicated that these residues are exchanging between two states at a frequency of 500-1000 Hz, with the minor state populated at 1-2%. The remaining residues showed either very little or no CPMG relaxation dispersion, indicating that the exchange is either too slow or too fast, or that the chemical shifts in the two states are too similar to be distinguishable.

Table 4.4: Chemical exchange parameters obtained from the NADP⁺-free CPMG relaxation dispersion experiments.

	Signal A (Fig. 4.6A, D)	Signal B (Fig. 4.6B, E)	Signal C (Fig. 4.6C, F)
ω_1 - ¹⁵ N (ppm)	126.8	124.6	121.7
ω_2 - ¹ H (ppm)	8.39	8.31	8.45
R_2^0 800 MHz (s ⁻¹)	8.6±0.5	5.1±0.5	4.5±0.9
R_2^0 500 MHz (s ⁻¹)	7.7±0.4	6.0±0.6	6.2±1.2
k_{ex} (s ⁻¹)	900±120	740±50	570±20
Population of b (%)	1.0±0.1	1.8±0.1	1.86±0.03
d ω (ppm)	1.9±0.2	2.36±0.08	3.53±0.03
χ^2	103	28	34

In the CPMG experiments with NADP⁺, either in the presence or absence of cyclohexanone or ϵ -caprolactone, none of the signals were observed to have significant relaxation dispersion, indicating once again that either the two exchanging states were too similar, or the exchange regime was too slow or fast. For two of the three signals, the R_2^0 values, the transverse relaxation rates in the absence of exchange, were lower in the samples with NADP⁺ than in the absence of ligands, suggesting that they have relaxation properties consistent with a

smaller molecule (Fig. 4.6BC). For the third signal, the R_2^0 values were larger in the samples with NADP⁺ at a field strength of 800 MHz (Fig. 4.6A).

Given that the three residues in which CPMG dispersion could be observed are likely to be loop residues, these three peaks were examined in the four HSQC spectra obtained in the presence and absence of the ligands (Fig. 4.6DEF). In the spectrum obtained in the absence of ligands, there was no indication that any of those signals were splitting due to multiple conformations in slow exchange with each other. In the sample with NADP⁺ only, two of the three appear to split into two signals, while the third shifts slightly downfield in the ¹H dimension. With the addition of cyclohexanone, one signal disappears entirely, while the other two also undergo some slight shifts and splitting. With ϵ -caprolactone, one remains in a similar position as in the sample without ligands, whereas the other two once again split into multiple peaks and undergo some slight shifting. This demonstrates that the addition of NADP⁺, cyclohexanone, and ϵ -caprolactone are able to perturb the chemical environment of the residues observed to be exchanging based on the CPMG dispersion experiments obtained in the absence of ligands.

4.5 Discussion

The SAXS and NMR data provide the first insight into the behaviour of the loop in solution. The loop is at minimum an indicator of the state of the enzyme throughout the reaction mechanism, and may be important in coordinating the transition from state to state. As such, these data provides entirely novel data regarding the BVMO catalytic mechanism. Examination of the five crystal structure of *Rm*CHMO (Mirza *et al.*, 2009 and Chapters 2 and 3) reveals that the loop is ordered in some states, including the CHMO_{Closed} and CHMO_{Tight} states, and disordered in others, including the CHMO_{Open}, CHMO_{Rotated}, and CHMO_{Loose} states. Similar observations have been made in other BVMOs (Malito *et al.*, 2004; Orru *et al.*, 2011; Leisch *et al.*, 2012). The order-disorder transition is not exclusively dependent on the ligands present, as the loop can be observed in both the ordered and disordered state with the same ligands bound. It is therefore reasonable to expect that the loop is always in an

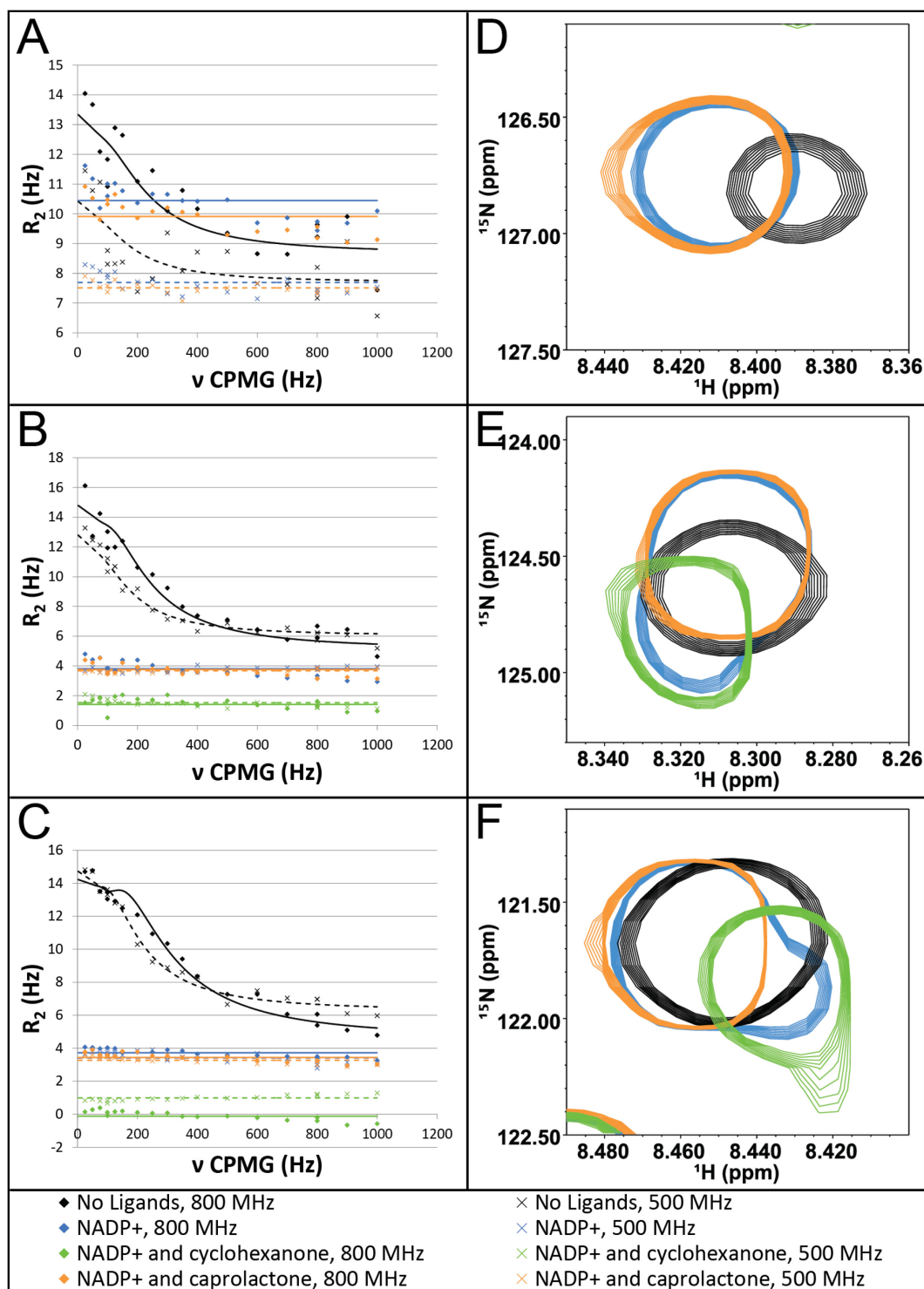


Fig. 4.6: CPMG relaxation dispersion data. (A, B, C) The CPMG relaxation curves for the signals that show relaxation dispersion without added ligands. (D, E, F) The corresponding signals in the HSQC spectra are overlaid using the same colour scheme as in panels A, B, and C.

equilibrium between at least two states: the ordered state observed in the crystal structures, and one or more disordered states that cannot be observed using

crystallography. Furthermore, it would be expected that the addition of ligands, and most importantly NADP^+ , would perturb this equilibrium.

The SAXS data demonstrate that the addition of NADP^+ does cause a change in the structure of both CHMO and CPDMO. The proteins appear to become more compact, based both on the R_g and the D_{max} , and the SAXS similarity maps demonstrate that addition of NADP^+ causes the greatest change to the SAXS scattering curve when the low intensity, high q -range region is considered as well. This effect appears to be a direct result of the binding of NADP^+ , given that the decrease in R_g can be used to produce a saturating binding curve that yields a K_d corresponding to the expected affinity of NADP^+ . This phenomenon could be explained based on domain movements or changes in global protein flexibility. The Porod-Debye analysis of protein density appears to preclude the latter interpretation. The rotation of the NADPH-binding domain, as observed when comparing $\text{CHMO}_{\text{Open}}$ and $\text{CHMO}_{\text{Closed}}$, could be responsible for this difference, but calculated R_g values vary no more than 0.5 Å between these two structures, as compared to the observed differences of 1.4 and 2.2 Å for CHMO and CPDMO, respectively. In addition, changes in R_g of a particle of fixed mass require an elongation of the particle, whereas the domain rotations observed based on the crystal structures are primarily rotations roughly around the protein's center of mass, which will have a limited effect on R_g . It therefore appears likely that a significant component of this reduction of R_g in the presence of NADP^+ is caused by changes in the loop structure. As NADP^+ binds to the enzyme, the equilibrium between the ordered and disordered states is expected to shift towards the more compact ordered state. In turn, this will lead to a decrease in the R_g of the protein, as observed in the SAXS data. The loop accounts for approximately 3% of the total protein mass, so even a very large movement away from the protein's center of mass will result in a relatively modest change in R_g . For this reason, the magnitude of the observed change in R_g is sufficient to account for the loop order-disorder transition, even concurrent with some domain movements that would have a minor effect on R_g .

The EOM analysis also appears to corroborate this interpretation. Using the scattering curve obtained in the absence of NADP^+ , a major peak at the high end of the R_g ensemble was consistently obtained, with a small peak sometimes present in the middle of the R_g ensemble, depending on the starting model and software parameters used. This would be consistent with the loop being primarily in the disordered conformation, which is also consistent with the structural models selected in the EOM pool. The high χ -value may be caused by minor differences in the domain organization in the absence of NADP^+ , and these χ -values are in line with what is obtained when comparing the NADP^+ -free scattering curve with the $\text{CHMO}_{\text{Open}}$ and $\text{CHMO}_{\text{Closed}}$ crystal structures using CRY SOL. In contrast, when the scattering curve obtained in the presence of NADP^+ was used, the position of the major and minor peaks were inverted, with the minor peak once again being present in only some EOM analyses, depending on the starting model and software parameters. The shift in the major peak is consistent with the ordering of the loop. It is tempting to suggest that the minor peaks in either scattering curve correspond to the alternative conformation of the loop. The fact that this minor peak was not obtained consistently, however, means that conclusions drawn from its presence should be treated with caution.

In summary, the SAXS data appear to suggest a disordered loop in the absence of NADP^+ that, upon binding NADP^+ , adopts an ordered conformation that increases the compactness of the protein. There is some evidence, when the EOM data and crystal structures are considered, that these two states always exist in an equilibrium that is perturbed by addition of NADP^+ . The NMR studies expand on these data with a more detailed analysis of the loop's configuration.

Given that CHMO is a 60 kDa protein, it is reasonable to expect that any residues that form part of the core, folded structure would have NMR properties consistent with a very large protein by NMR standards. The small intensities and broad linewidths cause the signals to merge into one another, making it very difficult to resolve them. Conversely, those signals that do not form part of the core structure and behave more like small peptides can be expected to have strong intensities and narrow linewidths, allowing them to be observed by NMR. It was

expected that a subset of those residues would correspond to the residues that make up the loop. It should be noted that not all of the roughly twenty loop residues need to be visible, as part of the loop, especially the loop termini, may exhibit properties that are more consistent with the folded protein.

The loop is expected to be mostly disordered in the absence of NADP^+ , and then become more ordered in the presence of NADP^+ ; therefore, it would be reasonable to expect that the largest number of signals would be visible in an HSQC spectrum in the absence of NADP^+ , and that addition of NADP^+ would cause some of those signals to disappear, as they spend more time in close contact with the protein and take on its NMR properties. It was therefore quite surprising to see the opposite effect occur, as the number of signals roughly doubles in the presence of NADP^+ .

Examination of the CPMG data provides a possible explanation for this observation. In the absence of NADP^+ , three of the observable residues exhibited CPMG relaxation dispersion. When fitted to the Carver-Richards equations for two-state exchange (equation 4.3), these residues were found to exchange at a rate of 500-1000 Hz, while occupying the minor state 1-2% of the time. This is consistent with the loop spending 98-99% of the time in the peptide-like state while rapidly sampling the ordered, large protein-like state, perhaps to "detect" if NADP^+ has bound to the protein.

When NADP^+ is added to the sample, irrespective of the presence of cyclohexanone or ϵ -caprolactone, the CPMG relaxation dispersion is lost, a number of additional signals appear, and some signals show signs of shifting and/or splitting into two. This may suggest that instead of rapidly sampling the ordered state, the loop is now exchanging much more slowly. Some of the residues that were spending a large portion of the time behaving like the large, ordered protein in the absence of NADP^+ were suddenly observable. This can be explained by the slower exchange between the observable, peptide-like state and the unobservable, protein-like state; in rapid exchange, if the residue spent too much time in the protein-like state, the signal would become too broadened to be observed. This would be compounded by the R_2 relaxation caused by chemical

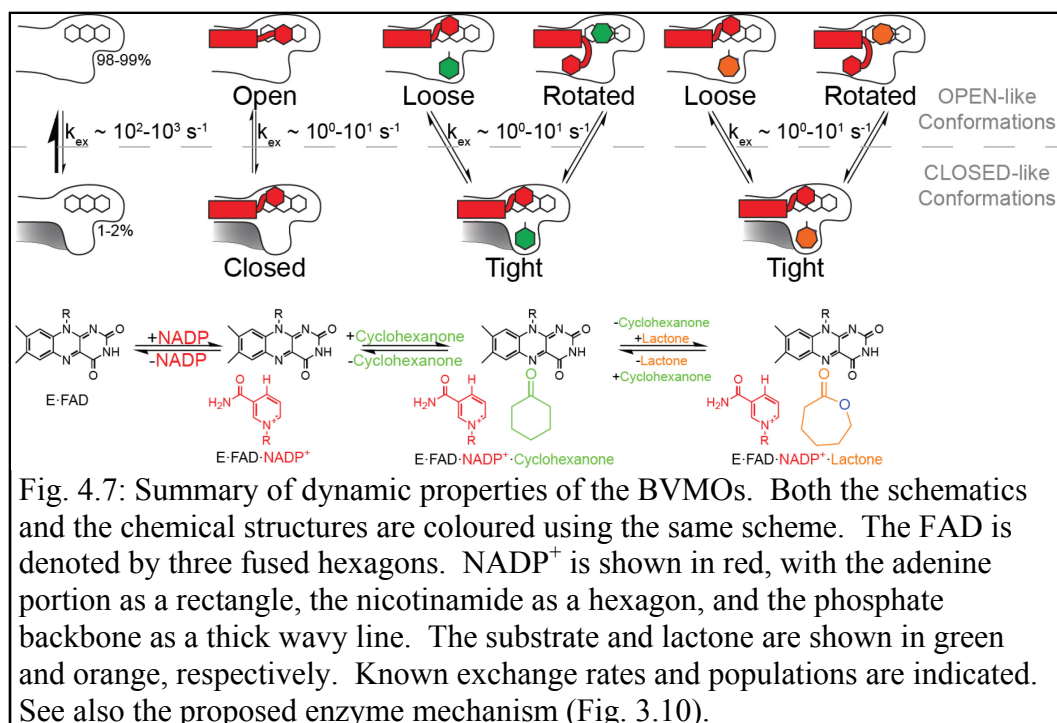


Fig. 4.7: Summary of dynamic properties of the BVMOs. Both the schematics and the chemical structures are coloured using the same scheme. The FAD is denoted by three fused hexagons. NADP⁺ is shown in red, with the adenine portion as a rectangle, the nicotinamide as a hexagon, and the phosphate backbone as a thick wavy line. The substrate and lactone are shown in green and orange, respectively. Known exchange rates and populations are indicated. See also the proposed enzyme mechanism (Fig. 3.10).

exchange under that exchange regime. Under slower exchange, the peptide-like state would have its own signal, which would be expected to be sharp. This is also consistent with the decrease in R_2^0 observed in two of the three signals in the presence of NADP⁺, as the slowly exchanging loop will not have the transverse relaxation properties caused by the rapid sampling of the large protein-like, ordered state. In addition, the signals that appeared split in two indicate two states that are significantly populated and chemically distinct. It should be noted that neither of these states can be the ordered state, as this state would be expected to exhibit NMR properties consistent with the large protein. These states likely represent an ensemble of two or more peptide-like states. It should be emphasized that all three of the residues that showed CPMG relaxation dispersion in the NADP⁺-free state shifted position, split, or disappeared in the three spectra with NADP⁺ and ligands present.

Taken together with the NADP⁺-free sample, the following picture emerges (Fig. 4.7). In the absence of NADP⁺, the loop residues are rapidly exchanging, with a minor conformation perhaps being involved in sampling for the presence of NADP⁺. In some regions of the loop, the residues may spend too much time in the protein-like state to be visible. In other cases, the two states

may be too chemically similar to each other to exhibit CPMG relaxation dispersion. Upon addition of NADP^+ , the loop changes to a slow exchange regime, with some of the residues that were previously unobservable in the rapid exchange regime becoming observable in the slow exchange regime. The addition of cyclohexanone or ϵ -caprolactone leads to local changes in the chemical environments of individual residues, but no large scale changes as seen when NADP^+ is added. Regardless of the presence of the substrate or product, the loop exchange in the NADP^+ -free state of the enzyme may be related to the enzyme determining whether or not to enter the catalytic cycle, whereas the loop exchange in the presence of NADP^+ is mimicking what would be observed during the catalytic cycle. In both cases, this illustrates that the dynamic properties of the BVMOs are occurring on an ongoing basis throughout the catalytic cycle, rather than occurring as a linear sequence of structural changes that occur as the reaction proceeds from one step to the next.

In the SAXS data, the rapidly exchanging, but predominantly peptide-like conformation is visible in the absence of NADP^+ as the less compact structure. Upon addition of NADP^+ , the two slowly exchanging states lead to a particle that is more compact on average. In some cases, when ensemble modelling of the loop is performed, the two states can in fact be identified in the SAXS data both in the presence and absence of NADP^+ . Unfortunately, none of the available data allow us to determine a quantitative value for the populations of the ordered and disordered state in the presence of NADP^+ , to compare the 98-99% value in the absence of NADP^+ . Based on the relative differences between the R_g values in the presence and absence of NADP^+ , as well as the calculated R_g based on the $\text{CHMO}_{\text{Closed}}$ structure, it is reasonable to assume that both the disordered, peptide-like conformation and folded-in, ordered conformation are present in significant quantities. The population of each would be expected to be roughly equal, given the presence of clear NMR signals and the notable change in R_g .

When the rates of exchange obtained from the NADP^+ -free state are compared to the rates of the various steps in the reaction mechanism that have been determined for *AcCHMO* and *PAMO* (Ryerson *et al.*, 1982; Sheng *et al.*,

2001; Torres Pazmiño *et al.*, 2008a), as well as the overall reaction rate determined for *RmCHMO* (Chapter 2), it is clear that the rate of loop exchange is far too fast to be driving the reaction. The reaction rates generally fall in the range of 1-100 s⁻¹. On the other hand, these reaction rates are very typical of exchange rates that give rise to two distinct NMR signals rather than a single signal that represents the average. As such, it appears likely that the NADP⁺-free loop state represents the pre-catalytic behaviour of the loop. Upon binding of NADP(H), the loop will "detect" the presence of the cofactor and switch to the slower exchange regime that is likely consistent with the behaviour of the loop during catalysis. This change in loop behaviour may trigger the domain movements observed through the various crystal structure of *RmCHMO*. This changes the view of the catalytic mechanism of the BVMOs as presented in Chapters 2 and 3. That mechanism describes the BVMO catalysis as a series of concerted steps, which is typically how we describe reaction mechanisms. These data, however, suggest that the conformational changes are likely occurring multiple times during and between catalytic steps, and only a subset of those multiple cycles are "productive," leading to the next stage in the mechanism. This is likely a more realistic view of enzyme behaviour in general, especially in those cases where conformational changes occur without any chemical modification of the enzyme or its ligands.

4.6 Conclusion

The complex domain movements of the BVMOs have been suggested through kinetic studies as well as crystallographic studies. There has as of yet been no attempt to structurally characterize these changes in solution, however. We have shown that the large loop, which we have previously demonstrated to be important for catalysis (Mirza *et al.*, 2009) and have proposed to play a role in the domain and active site organization of the enzyme (Chapters 2 and 3), can be studied in solution both by NMR and SAXS. We have demonstrated that the rate of exchange of the loop decreases in the presence of NADP⁺, leading to a structure that is globally more compact. Both solution techniques seem to point to an equilibrium between the two states, with NMR allowing a quantification of the

exchange parameters and populations in the NADP^+ -free state. Given the crystallographic evidence that the loop positioning is related to global domain rotations, we are able to extend the mechanistic implications of the loop dynamics to global enzyme dynamics. As such, these studies provide valuable insight into the dynamics of the BVMOs, and the behaviour of this loop in particular. It also allows us to link the protein dynamics to the chemical and structural enzyme mechanisms, while emphasizing that the structural changes likely represent a constantly shifting equilibrium that only sometimes leads to a "productive" advancement in the catalytic mechanism. In addition, given that these are the first BVMO solution studies to be published, they can pave the way for future studies of BVMO protein dynamics using solution techniques.

Chapter 5: Conclusion

Well over a century has passed since Anton von Baeyer and Victor Villiger discovered the reaction that bears their names (Baeyer & Villiger, 1899; Baeyer & Villiger, 1900). In the decades that followed, the Baeyer-Villiger oxidation (BVO) reaction has been studied extensively, providing us with detailed knowledge of its mechanism, including the nature of the Criegee intermediate (Criegee, 1948), as well as the basis of the stereochemistry of BVO (Doering & Speers, 1950; Hawthorne *et al.*, 1958; Rozzell & Benner, 1983). While this reaction has been applied to a vast range of important synthetic reactions (Krow, 1993), the need for alternative catalysts and reagents remains, as current methods can be inefficient, costly, and hazardous both for workers and for the environment (ten Brink *et al.*, 2004). In addition, as with any process with pharmaceutical applications, improvements in stereospecificity are always desirable.

One potential alternative BVO catalyst is the group of enzymes known as the Baeyer-Villiger monooxygenases (BVMOs), which have been studied extensively for decades. These enzymes are attractive catalysts owing to their high stereospecificity, broad substrate specificity, and "green" biocatalytic properties. Even in the 1976 paper first describing cyclohexanone monooxygenase (CHMO) (Donoghue *et al.*, 1976), the potential use of these enzymes and the bacteria that harbour them in the degradation of compounds present in fossil fuels was implied. Shortly after that, the idea that the BVMOs could be used as tools in synthetic chemistry, rather than just to clean up toxic substances, was proposed. As stated in a 1983 paper, "we have also set out to investigate [CHMO's] potential as a general tool for the production of chiral, bifunctional synthons of high enantiomeric purity. (Many enzymes are known which discriminate quite efficiently between enantiomeric substrates)" (Schwab *et al.*, 1983). Since then, reports detailing the chemical mechanism of these enzymes (Ryerson *et al.*, 1982; Sheng *et al.*, 2001; Torres Pazmiño *et al.*, 2008a) have led to a very complete view of the chemistry behind this family of enzymes. At the same time, advances in molecular biology techniques have allowed the

identification of many new BVMOs and the development of new bioprocess strategies for the industrial use of BVMOs (Leisch *et al.*, 2011).

While these developments are encouraging, the complexity of the BVMO mechanism and the lack of structural data have hampered our ability to adapt them to specific biocatalytic applications. Some recent crystal structures have shed light on the overall architecture and flexibility of these proteins (Malito *et al.*, 2004; Mirza *et al.*, 2009; Orru *et al.*, 2011; Leisch *et al.*, 2012), but they have left us with many important questions in our understanding of the BVMOs. First, on the basis of the crystal structures that were already available, it was not possible to satisfactorily explain how these enzymes can have such broad substrate specificity while retaining a high degree of regio- and enantiospecificity. The necessity for the substrate to bind and the product to be released while NADP^+ is bound to the enzyme further complicated this question. Second, the mechanism by which the stereospecificity of the BVMOs is determined, a critical piece of information for rational protein design, could not be explained. Finally, from the available PAMO and CHMO structures, it was also clear that a better understanding of protein dynamics was needed. To this end, we have pursued structural studies of CHMO and a homologue, cyclopentadecanone monooxygenase (CPDMO). The $\text{CHMO}_{\text{Rotated}}$, $\text{CHMO}_{\text{Tight}}$, and $\text{CHMO}_{\text{Loose}}$ crystal structures, along with the dynamics and mutagenesis data presented here, have allowed us to make progress in answering all three of these questions. We are now able to propose a mechanism that is consistent with an enzyme with broad substrate specificity and high stereospecificity (Fig. 5.1). These structures, combined with solution studies, also give us a much improved picture of the dynamic properties of the protein (Fig. 5.2).

The $\text{CHMO}_{\text{Rotated}}$ crystal structure is critical in explaining the broad substrate specificity of CHMO. The structures of BVMOs that were previously available, especially the $\text{CHMO}_{\text{Open}}$ and $\text{CHMO}_{\text{Closed}}$ structures, suggested a dynamic substrate binding site that could transition between a large solvent exposed pocket and a small closed off pocket. It was tempting to suggest that the substrate bound first in the $\text{CHMO}_{\text{Open}}$ cavity before switching to the $\text{CHMO}_{\text{Closed}}$

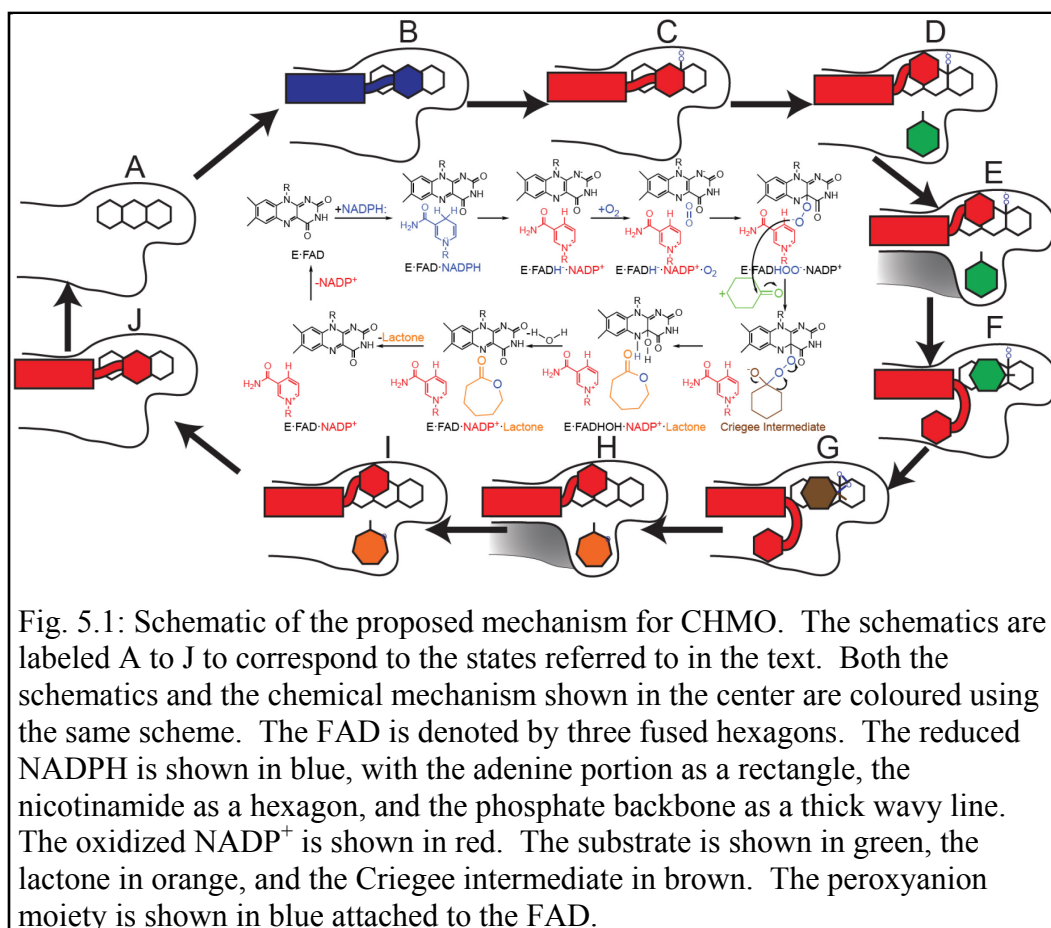


Fig. 5.1: Schematic of the proposed mechanism for CHMO. The schematics are labeled A to J to correspond to the states referred to in the text. Both the schematics and the chemical mechanism shown in the center are coloured using the same scheme. The FAD is denoted by three fused hexagons. The reduced NADPH is shown in blue, with the adenine portion as a rectangle, the nicotinamide as a hexagon, and the phosphate backbone as a thick wavy line. The oxidized NADP⁺ is shown in red. The substrate is shown in green, the lactone in orange, and the Criegee intermediate in brown. The peroxyanion moiety is shown in blue attached to the FAD.

conformation to undergo catalysis. The difficulty with this mechanism was that for larger substrates of CHMO, especially bicyclic ketone substrates, to form a Criegee intermediate with the enzyme in either of those conformations would introduce steric clashes with either the enzyme or NADP⁺. Furthermore, the various conformations of CHMO available at that time would suggest that flexibility in the active site to accommodate these larger substrates would likely move the enzyme out towards the exterior of the protein, rather than towards the core of the flavin ring system where the Criegee intermediate would be formed.

The CHMO_{Rotated} structure introduces a new conformation of NADP⁺ that involves a drastic rotation of the nicotinamide moiety, for which the structure is named. This rotation is necessary to create enough space to allow the substrate to move above the flavin ring system and react with the peroxyanion intermediate (Fig. 5.1, *state G*). Relatively few interactions are formed with the pyridine moiety in the Rotated conformation. This is likely important in preventing the

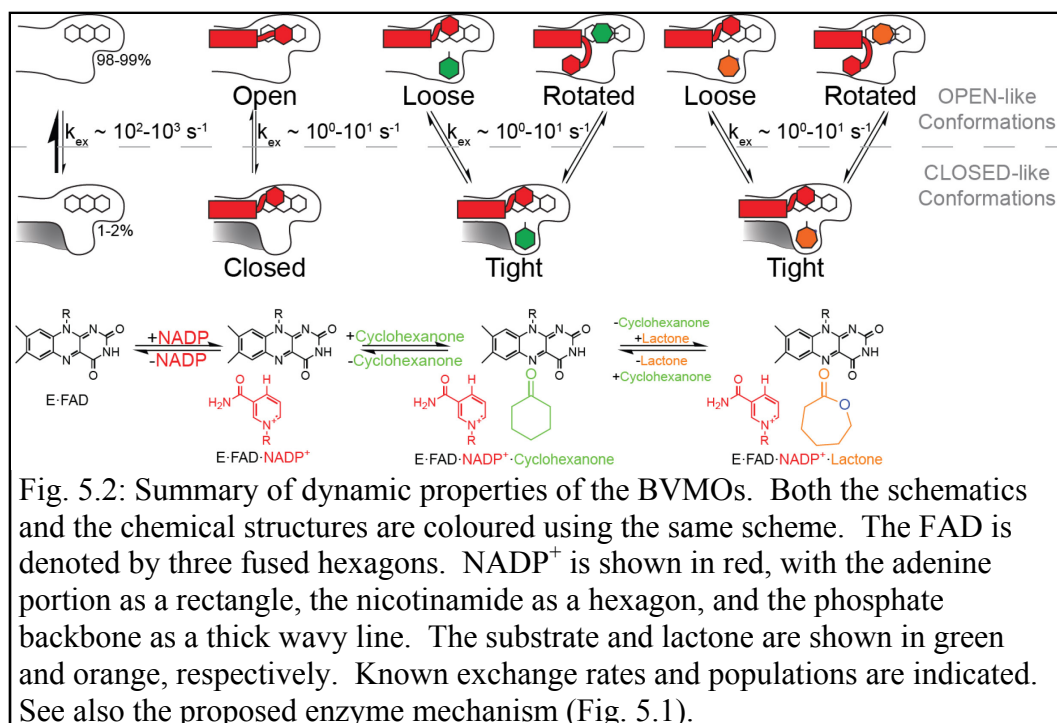


Fig. 5.2: Summary of dynamic properties of the BVMOs. Both the schematics and the chemical structures are coloured using the same scheme. The FAD is denoted by three fused hexagons. NADP⁺ is shown in red, with the adenine portion as a rectangle, the nicotinamide as a hexagon, and the phosphate backbone as a thick wavy line. The substrate and lactone are shown in green and orange, respectively. Known exchange rates and populations are indicated. See also the proposed enzyme mechanism (Fig. 5.1).

adoption of this conformation in the absence of substrate, which would lead to a destabilization of the peroxyanion intermediate and, by extension, uncoupling of NADPH consumption from substrate oxidation. We demonstrated this effect through the mutagenesis experiments that we performed, which showed that mutants in which this conformation was stabilized have properties that are consistent with a destabilized peroxyanion intermediate. The rotation of the nicotinamide moiety also serves to seal the substrate away from the bulk solvent. This is an important characteristic of the enzyme, as if the NADP⁺ were to move away from the peroxyanion intermediate while the substrate diffuses back into the bulk solvent, it would also lead to the destabilization of the peroxyanion intermediate and uncoupling of NADPH consumption.

While the CHMO_{Rotated} structure is able to explain the broad substrate specificity of the enzyme, examination of the Rotated pocket suggests that the high degree of stereospecificity seemed highly unlikely to be determined in that conformation. In addition, residues that appear to have an effect on the substrate profile of the BVMOs are too distant from the binding site of cyclohexanone in the CHMO_{Rotated} structure. The CHMO_{Loose} and CHMO_{Tight} structures, on the other hand, place the product in close proximity to those residues, and in the case

of the CHMO_{Tight} structure, in a pocket that is small enough to explain the stereospecificity of the enzyme. While in principle the Tight conformation could allow the formation of the Criegee intermediate for smaller substrates like cyclohexanone, it appears too small to accommodate larger substrates without substantial conformational changes that have not yet been observed, suggesting that the Rotated conformation remains important for the formation of the Criegee intermediate. The CHMO_{Loose} structure is the only structure of CHMO that has a substrate or product bound that is solvent accessible. In CHMO_{Rotated}, the NADP⁺ rotation blocks cyclohexanone from the bulk solvent, whereas in CHMO_{Tight}, the domain rotations and loop ordering do the same.

The CHMO_{Tight} structure can also be used to explain the stereospecificity of the enzyme with respect to cyclohexanone-based substrates. Based on past mutagenesis studies, as well as examination of the differences between BVMOs, it is possible to use the positioning of the lactone in the CHMO_{Tight} structure to project how an asymmetric substrate would bind in the Tight conformation.

Given that formation of the Criegee intermediate is expected to occur in the Rotated conformation, it is necessary for the substrate to move from the Tight position to the Rotated position without having the opportunity to reorient itself. The positioning of the strictly conserved active site residue, R329, may be playing a role in directing this movement. R329 forms a hydrogen bond with the carbonyl oxygen of the lactone in the Tight conformation. In the Rotated conformation, R329 remains within hydrogen bonding distance of the carbonyl oxygen of cyclohexanone, though the angle is perturbed due to the lack of the negatively charged peroxyanion intermediate. This would allow a substrate to move from the Tight position to the Rotated position while maintaining the stereospecificity preferences of the Tight position. It is also possible that the Criegee intermediate is formed in concert with this rotation.

In examining the five crystal structures of CHMO, and the three substrate and product bound structures in particular, it is clear that the rotation in the NADPH-binding domain and the ordering of the large loop do not have a simple relationship with the ligand-bound state of the enzyme (see Fig. 5.1). The degree

of rotation of the NADPH-binding domain can be quite different with the same ligands bound, and also does not seem to correlate well with a logical progression through the enzyme mechanism. This implies that the degree of domain rotation and the order/disorder transition of the large loop are dynamic enzyme properties that may be occurring continuously through the catalytic cycle. The SAXS and NMR data probe this behaviour. From the SAXS data, it is clear that the enzyme becomes more compact upon the addition of NADP^+ . The main cause of this is likely the order/disorder transition of the loop, which is likely linked to domain rotations. A more detailed analysis using ensemble modelling suggests that in absence of NADP^+ , the loop is primarily in an extended conformation and folds in in the presence of NADP^+ , as would be predicted from the crystal structures. There is also evidence, however, of the loop, and therefore the domain rotations, being present in the "opposite" state both in the presence and absence of NADP^+ , suggesting that the loop is always partly ordered and partly disordered.

The NMR data corroborates these observations, and provides more detail into the nature of this transition. In the absence of NADP^+ , the CPMG data seem to suggest that the loop residues are rapidly sampling a minor chemical state. The sampling rate of this minor conformation is much too fast to be related to the majority of the steps in the enzyme mechanism, but may represent the loop "sensing" if NADP^+ has been bound by the enzyme, perhaps to determine whether to enter the catalytic cycle. In contrast, this rapid exchange is not observed in the presence of NADP^+ . Instead, a number of additional signals become visible in the HSQC spectrum. This is consistent with the exchange of the large loop slowing down, allowing direct observation of some residues whose signals were undetectable in the absence of NADP^+ . The slow exchange NMR regime is roughly consistent with the timescale of the individual steps in the BVMO reaction mechanism. This suggests that the loop exchange in the presence of NADP^+ could be mimicking that which would be seen in the true catalytic cycle. This seems to point to a dynamic equilibrium of domain rotations and loop order-disorder transitions occurring throughout the catalytic cycle, with "productive"

changes in this equilibrium causing the enzyme to move from one major mechanistic state to the next.

Together, these data give us a much better appreciation for the complexities of the BVMO catalytic cycle from a structural perspective (see Fig. 5.1 and Fig. 5.2). The binding of NADPH, which will cause the enzyme to adopt an Open-like conformation, will trigger a change in loop dynamics from a rapid exchange regime to a slower exchange regime. This will allow the NADPH to reduce FAD. Subsequently, molecular oxygen will bind and react with the reduced flavin, with the enzyme still in an Open-like conformation. The substrate will then be bound, first in the Open-like Loose conformation, then in the Closed-like Tight conformation. The latter conformation will allow the determination of substrate acceptance as well as the stereospecificity of the reaction. The binding of the substrate is expected to cause the minor alterations in the chemical environment of the loop, as detected by NMR. The enzyme will then switch to the Rotated conformation, allowing the formation of the Criegee intermediate and production of the lactone. The flavin will then undergo dehydration before switching back to the Tight conformation and then the Loose conformation. The product will then be released to the solvent, and finally NADP^+ will be released to the solvent. Once NADP^+ is released, the loop will resume the rapid sampling of the minor conformation in preparation for another catalytic cycle.

While our structural knowledge of the BVMO enzyme mechanism has clearly been advanced as a result of this work, unanswered questions remain. Most importantly, it is necessary to confirm the basis for stereospecificity proposed in Chapter 3. This piece of information is very important for rational protein design, and could best be confirmed by obtaining a crystal structure of CHMO in the presence of an asymmetric or large ketone or lactone. Examination of additional BVMOs, in particular those with unusual substrate profiles, would also be very helpful in this regard. In addition, the non-BVO reactions that can be catalyzed by the BVMOs, including the oxidation of sulfides and other electrophilic centers, should be studied to fully understand the substrate profile of the BVMOs. Finally, continued study of the dynamic properties of the enzyme,

and the large loop in particular, could provide additional insight on how to approach the rational engineering of modified BVMOs.

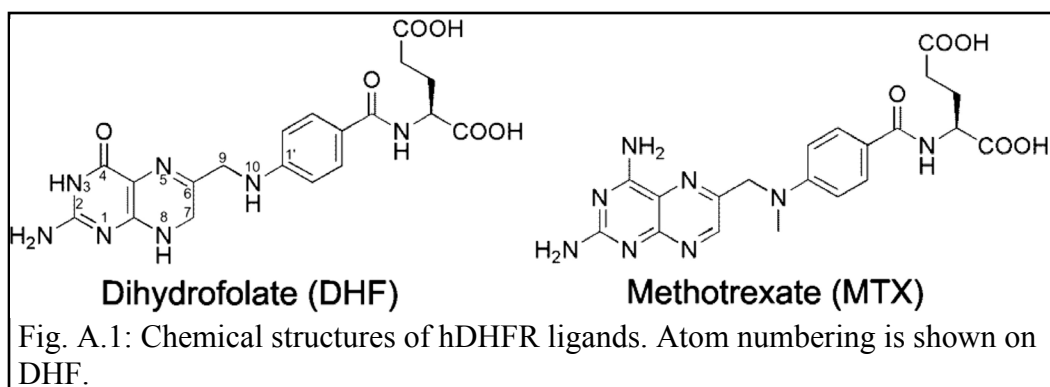
That said, the three crystal structures, enzymology and mutagenesis data, and solution structural data presented here have vastly increased our understanding of the BVMO enzyme mechanism. Through our crystal structures, we can rationalize the broad substrate profile, high degree of stereospecificity, and the mechanistic requirement to have the substrate binding and product release occur with NADP⁺ bound. With our solution structural data, we can better describe the nature and timescale of the complex conformational changes that are observed in our crystal structures. The enzymology and mutagenesis data have been critical in confirming the observations made based on the structural data. Together, these data have allowed us to propose a detailed structural mechanism that not only describes the "linear" sequence of structural events required for a complete catalytic cycle, but also provides some insight into the structural fluctuations occurring during and between each step. When combined with our detailed knowledge of the BVMO chemical mechanism, this information provides the most comprehensive description of molecular motions involved in BVMO catalysis. This will further our ultimate goal of being able to easily engineer a BVMO suited for specific biocatalytic applications. With recent efforts to overcome many of the practical problems of industrial-scale biocatalysis using BVMOs, this will hopefully lead to the adoption of BVMOs as "green" industrial biocatalysts to replace the environmentally unfriendly and hazardous peracid reagents that are most commonly used today.

Appendix A: Multiple Conformers in Active Site of Human Dihydrofolate Reductase F31R/Q35E Double Mutant Suggest Structural Basis for Methotrexate Resistance

A.1 Introduction

Human dihydrofolate reductase (hDHFR) catalyzes the reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate in a NADPH-dependent manner. 5,6,7,8-Tetrahydrofolate is a cofactor in purine and thymidylate biosynthesis, which are essential metabolites in cell division and proliferation. As a consequence of its essential role in nucleoside biosynthesis, hDHFR has been extensively exploited as a drug target. Inhibition with folate antagonists, or antifolates, arrests cell proliferation. The most effective clinical antifolate to date is methotrexate (MTX) (Fig. A.1), a slow, tight-binding competitive inhibitor that displays high affinity for hDHFR ($K_i^{\text{MTX}} = 3.4 \text{ pM}$). MTX is currently used to treat a variety of diseases, including cancer (Strojan *et al.*, 2005; Daw *et al.*, 2006; Slamon *et al.*, 2006), and autoimmune diseases such as juvenile idiopathic arthritis (Ramanan *et al.*, 2003). A number of resistance mechanisms to MTX have been observed in cancer patients, including impaired transport of MTX to the cytoplasm (Flintoff *et al.*, 2004) and decreased retention of MTX in the cell (Takemura *et al.*, 1999). Numerous *ex vivo* studies have reported mutations in the hDHFR gene resulting in an enzyme variant with decreased affinity for MTX (Blakley & Sorrentino, 1998; Volpato & Pelletier, 2009). These have contributed to increase our understanding of the molecular basis for active-site discrimination between the substrate, DHF, and its competitive inhibitor, MTX. Understanding the molecular interactions that affect tight binding of MTX to the active site of DHFR will contribute to our understanding of antifolate binding to DHFR, which can in turn contribute to the design of more efficient inhibitors.

A considerable number of DHFR active-site variants have been identified in MTX-resistant cancer cell lines (although never in patients) (Spencer *et al.*, 1996) or engineered *in vitro* to elucidate the role of active site residues in the binding of MTX. Amino acid substitutions at residues Ile-7 (Patel *et al.*, 1997), Leu-22 (Lewis *et al.*, 1995; Ercikan-Abali *et al.*, 1996b), Phe-31 (Chunduru *et al.*,



1994), Phe-34 (Nakano *et al.*, 1994), Arg-70 (Thompson & Freisheim, 1991), and Val-115 (Fossati *et al.*, 2008) have yielded MTX-resistant variants. These residues are all present in the folate-binding pocket (Cody *et al.*, 2005). Because MTX and DHF bind to the active site of hDHFR in a similar manner, all known substitutions causing a decrease in MTX affinity also decrease DHF affinity and overall catalytic efficiency (Blakley & Sorrentino, 1998; Volpato *et al.*, 2007; Fossati *et al.*, 2008). However, the loss of DHF affinity and catalytic efficiency is generally smaller than the loss of MTX affinity. This is often attributed to formation of different contacts with either ligand due to the 180° inversion of the pterin ring of bound DHF relative to MTX (Oefner *et al.*, 1988; Cody *et al.*, 2005).

Crystal structures of MTX-resistant point mutants have offered insight into the causes of decreased binding of MTX or other antifolates (Cody *et al.*, 1997; Gangjee *et al.*, 1998; Cody *et al.*, 2003a; Cody *et al.*, 2003b; Cody *et al.*, 2004; Cody *et al.*, 2005). To this day, crystal structures of MTX-resistant hDHFR variants L22F, L22R, and L22Y (Lewis *et al.*, 1995), as well as F31G and F31S (Cody *et al.*, 1998), complexed to various antifolates, have been reported. Only the L22Y variant has been co-crystallized with MTX. Despite its decreased affinity for MTX (L22Y $K_i^{\text{MTX}} = 11$ nM versus WT $K_i^{\text{MTX}} = 31$ pM (Volpato *et al.*, 2007)), the inhibitor in the variant structure was bound in the same way as in the native enzyme, making interpretation of decreased affinity difficult to assess. Nonetheless, the low probability conformation of residue Tyr-22 suggested that the presence of a bulky aromatic residue in this area of the folate-binding pocket generated unfavorable hydrophobic interactions with the 2,4-diaminopter

moiety of the inhibitor (Lewis *et al.*, 1995). This is also expected to reduce DHF substrate binding. Structures of MTX-resistant variants F31G and F31S were obtained complexed to N-[4-[(2,4-diaminofuro[2,3-d]pyrimidin-5-yl)methyl]methylamino]benzoyl]-L-glutamate (MTXO) (Cody *et al.*, 1998), a MTX analog in which the 2,4-2,4-diaminopterine moiety is replaced by a 2,4-diaminofuropyrimidine moiety. Superposition of MTXO-bound variants with MTX-bound WT hDHFR revealed that the ligands bind to the active site in an analogous manner. It was suggested that decreased MTX binding in the substituted variants resulted from the loss of van der Waals and hydrophobic contacts established between the native Phe-31 and the *p*-ABA and 2,4-diaminopterine moieties of MTX. F31G and F31S display a 10-fold decrease in affinity for MTX relative to WT hDHFR ($K_i^{\text{MTX}} = 31 \text{ pM}$ (Volpato *et al.*, 2007)). Further Phe-31 variants (i.e. F31R; $K_i^{\text{MTX}} = 7 \text{ nM}$, 200-fold decrease in MTX affinity) (Patel *et al.*, 1997) display larger decreases in affinity relative to F31G and F31S. This cannot be rationalized by reduction of side-chain contacts with the inhibitor due to the presence of a smaller side chain.

These results illustrate the difficulty of gaining insight into the molecular causes for altered MTX binding. This may be partly attributed to the very tight binding of MTX to the native enzyme, such that binding to resistant variants often remains in the sub-nanomolar or low nanomolar range, where the general mode of ligand binding has not changed appreciably relative to the native enzyme. Combining active-site mutations in hDHFR by protein engineering has been shown to generate variants with greatly decreased affinity to MTX (Ercikan-Abali *et al.*, 1996a; Volpato *et al.*, 2007). Studying the molecular interactions in highly MTX-resistant hDHFR variants offers the possibility of capturing more important changes in enzyme-ligand interactions.

Here, we report detailed observations for the mode of MTX resistance in the combinatorial variant F31R/Q35E. Variant F31R/Q35E is a relevant candidate for better understanding the specific interactions that govern ligand recognition in the folate binding site, because it displays a >650-fold decrease in MTX affinity ($K_i^{\text{MTX}} = 21 \text{ nM}$) accompanied by a modest, 9-fold decrease of

affinity for the substrate DHF relative to WT hDHFR (Volpato *et al.*, 2007). In addition, we have recently shown that this variant is an efficient selectable marker for various mammalian cell types, including murine hematopoietic stem cells (Volpato *et al.*, 2007; Volpato *et al.*, 2011). Because mutations giving rise to MTX resistance are not observed in mammals, and because MTX is approved for human treatment, engineered resistant DHFRs offer great potential as human selective markers *ex vivo* or *in vivo* (Flasshove *et al.*, 1995; Patel *et al.*, 1997; Allay *et al.*, 1998). To better understand the effect of either amino acid substitution on each ligand, a kinetic double mutant cycle was constructed with the simple variants F31R and Q35E. The crystal structure of the F31R/Q35E variant was obtained with bound MTX at 1.7-Å resolution, to elucidate the structural basis of MTX resistance in this variant. In addition, molecular docking was performed with the F31R/Q35E structure to evaluate the role of the two substitutions toward folate binding. Overall, the results reveal synergistic effects of the combined substitutions toward loss of MTX binding, characterized by increased disorder of specific residues throughout the active site of the highly MTX-resistant F31R/Q35E variant.

A.2 Experimental Procedures

A.2.1 Construction of Vectors hDHFR F31R/Q35E-pET24, hDHFR F31R-pQE32, and hDHFR Q35E-pQE32

The hDHFR F31R/Q35E gene was amplified by PCR using the following primer set: forward (5'-CACACACCATATGGTTGGTTCGCTAAACTG-3', NdeI restriction site in italics) and reverse (5-GTTCTGAGGTCATTACTGG-3', external primer) from the hDHFR F31R/Q35E-pQE32 template (Volpato *et al.*, 2007). The amplified gene was subcloned in the modified pET24 vector (Doucet *et al.*, 2007) between the NdeI and HindIII restriction sites using T4 DNA ligase, and the ligation mixture was transformed into electrocompetent BL21(DE3) cells. The expected sequence was confirmed by DNA sequencing.

The F31R and Q35E substitutions were created by mega-primer PCR using external primer set 2 described in a previous study (Volpato *et al.*, 2007) and the mutagenic primers 5'-TCTCTGGAAATATCTACGTTTCGTTCTTAAG

G (F31R, reverse) and 5'-GTTGTGGTCATTTCTTTTCGAAATATCTAAATTCTG T (Q35E, reverse), respectively. The amplified gene was subcloned in the pQE32 vector between the BamHI and HindIII restriction sites using T4 DNA ligase, and the ligation mixture was transformed into electrocompetent SK037 cells (Volpato *et al.*, 2007). Expression, purification, and determination of kinetic and inhibitory constants were performed as previously described (Volpato *et al.*, 2007). Briefly, kinetic and inhibition assays were conducted in MATS buffer (25 mM MES, 25 mM acetate, 50 mM Tris, 100 mM sodium acetate, and 0.02% (w/v) sodium azide) (pH 7.6) at 23°C, by monitoring the NADPH and DHF depletion ($\Delta\epsilon_{340\text{ nm}} = 12\,800\text{ M}^{-1}\text{ cm}^{-1}$). All assays were performed in at least four independent experiments, and the average values are reported. The initial rates during the first 15% of substrate conversion were recorded for all assays. Kinetic and inhibition parameters were obtained from a non-linear regression fit to the Henri Michaelis-Menten equation using Prism (GraphPad Software, San Diego, CA). The k_{cat} values were determined in the presence of saturating substrate concentrations (100 μM each of DHF and NADPH) in 1 cm cells according to $k_{\text{cat}} = V_{\text{max}}/[E]$. K_{M} DHF values were obtained using 10 cm cells containing 1 nM enzyme, 10 μM NADPH and a range of DHF concentrations (0.05 μM to 10 μM). $\text{IC}_{50}^{\text{MTX}}$ was determined in the presence of saturating substrate concentrations and variable MTX concentrations (0.025 μM to 100 μM). Inhibition constants for MTX (K_i^{MTX}) were calculated from $\text{IC}_{50}^{\text{MTX}}$ according to the equation for competitive inhibitor binding (Segel, 1993).

A.2.2 Expression and Purification of hDHFR F31R/Q35E

An overnight culture of BL21(DE3)/hDHFR F31R/Q35E-pET24 was used to inoculate 1 liter of LB medium. The culture was grown at 37°C until the $\text{OD}_{600\text{ nm}}$ reached ~ 0.7 . Protein expression was induced with the addition of 1 mM of isopropyl 1-thio- β -D-galactopyranoside (IPTG), after which the cells were grown for 16 h at 22°C. Induced cells were harvested by centrifugation ($4000 \times g$ for 30 min at 4°C). The cell pellet was resuspended in 10 mM Tris-HCl, pH 8.3, at 4°C. The cells were lysed on ice using a Branson sonicator (four pulses at 200 watts for 30 s with a tapered micro-tip). The cellular debris was pelleted by centrifugation

(4000 × g for 30 min at 4°C), and the supernatant was filtered through a 0.2 µm filter before purification.

Purification was performed following a two-step purification protocol on an AKTA fast-protein liquid chromatography system (Amersham Biosciences) at 5°C. First, the supernatant was applied to an anion-exchange DEAE-Sepharose column (1.6 × 30 cm) followed by a 3 column volume wash with 10 mM Tris-HCl, pH 8.3, at 2 mL/min. A linear gradient of 5 column volumes with NaCl (0–200 mM) in 10 mM Tris-HCl, pH 8.3, was used to elute the F31R/Q35E variant. hDHFR activity was monitored in MATS buffer, pH 7.6, in the presence of 100 µM each of NADPH and DHF. Activity was measured in flat-bottom plates (Costar #3595) by monitoring concurrent depletion of NADPH and DHF ($\Delta\epsilon_{340\text{ nm}} = 12\,800\text{ M}^{-1}\text{ cm}^{-1}$) on a FLUOstar OPTIMA UV-visible plate reader (BMG Laboratories, Offenburg, Germany). Active fractions were pooled and dialyzed overnight at 4°C against 50 mM phosphate buffer, pH 7.5. Following dialysis, the sample (45 mL) was concentrated to 1.5 mL using an Amicon concentrator (molecular weight cut-off 10000, Millipore), for injection on a Superose 12 column (1.6 × 55 cm). The sample was eluted with 50 mM phosphate buffer, pH 7.5, at a flow rate of 1.5 mL/min. hDHFR activity was monitored as described above. Enzyme purity was evaluated using separation by SDS-PAGE (15% (w/v) polyacrylamide gel) stained by the zinc-imidazole method (Fernandez-Patron *et al.*, 1992) and quantified using the public domain image analysis software Scion Image (NIH, rsb.info.nih.gov/nih-image). Protein concentration was quantified using the Bradford assay (Bio-Rad).

A.2.3 Expression and purification of His₆-hDHFR F31R/Q35E

An overnight pre-culture of SK037/His₆-hDHFR F31R/Q35E-pQE32 was used to inoculate 1 L of LB medium. The culture was grown at 37°C until the OD_{600 nm} ~0.7. Protein expression was induced with the addition of 1 mM of isopropyl 1-thio-β-D-galactopyranoside (IPTG), after which the cells were grown for 3 h at 37°C. Induced cells were harvested by centrifugation (4000 × g for 30 min at 4°C). The cell pellet was resuspended in 0.1 M phosphate buffer, pH 8.0, at 4°C. The cells were lysed on ice using a Branson sonicator (four pulses at 200

W for 30 s with a tapered micro-tip). The cellular debris was pelleted by centrifugation (4000g for 30 min at 4°C) and the supernatant was filtered through a 0.2µm filter before purification.

Purification was performed following a 2-step purification protocol on an AKTA FPLC (GE Healthcare, Piscataway, NJ) at 5°C. First, the supernatant was applied to a HisTrap FF 1 mL pre-packed cartridge (GE Healthcare) followed by a 10 column volumes (CV) wash with 0.1 M phosphate buffer, pH 8.0, at 1 mL/min. A step-wise gradient of imidazole (10, 20, 50 and 200 mM; 5 CV for each step) in 0.1 M phosphate buffer, pH 8.0, was used to elute the F31R/Q35E mutant. hDHFR activity was monitored in MATS buffer, pH 7.6, in the presence of 100 µM each NADPH and DHF. Activity was measured in flat-bottom plates (Costar #3595) by monitoring concurrent depletion of NADPH and DHF ($\Delta\epsilon_{340\text{nm}} = 12\,800\text{ M}^{-1}\text{ cm}^{-1}$) on a FLUOstar OPTIMA UV-vis plate reader (BMG Laboratories, Offenburg, Germany). Active fractions were pooled and dialysed overnight at 4°C against 50 mM phosphate buffer, pH 7.5. Following dialysis, the sample (15 mL) was concentrated to 1.5 mL using an Amicon concentrator (MCWO 10000, Millipore), for injection on a Superose 12 column (1.6 × 55 cm). The sample was eluted with 50 mM phosphate buffer, pH 7.5, at a flow rate of 1.5 mL/min. hDHFR activity was monitored as described above. Enzyme purity was evaluated following separation by SDS-PAGE (15% (w/v) polyacrylamide gel) stained by the zinc-imidazole method (Fernandez-Patron *et al.*, 1992) and quantified using the public domain image analysis software Scion Image (NIH, rsb.info.nih.gov/nih-image). Protein concentration was quantified using the Bradford assay (Biorad, Hercules, CA).

A.2.4 Crystallization and X-ray Data Collection of hDHFR F31R/Q35E

Purified hDHFR F31R/Q35E enzyme was concentrated to 10 mg/mL using an Amicon concentrator (molecular weight cut-off 10000). MTX and NADPH were prepared as described previously (Volpato *et al.*, 2007) and were added at a final concentration of 2 mM each (5-fold molar excess) to the protein sample. Crystallization experiments were set up using hanging drop, vapor-diffusion experiments, with a reservoir volume of 1 mL and a drop size of 4 µL of

equal volumes of protein and reservoir solutions. A reservoir solution containing 0.2 M cadmium phosphate and 2.2 M ammonium sulfate yielded crystal-like formations that diffracted poorly. These crystals were crushed using a Hampton Seed Bead Kit and used as seeds (1/10 dilution). Rod-shaped crystals were obtained from crystallization experiments with 0.2 M sodium phosphate and 2.2 M ammonium sulfate as the reservoir solution and with drops containing 1.5 μL of protein, 2 μL of reservoir solution, and 0.5 μL of seeding solution. The crystals were soaked in the mother liquor supplemented with 15% glycerol as a cryoprotectant and frozen in a nitrogen cryostream (model X-stream 2000). Data were collected using a Rigaku RU-H3R generator, equipped with Osmic focusing mirrors, and an R-axis IV++ image plate detector, and processed using HKL-2000 (Otwinowski & Minor, 1997) (Table A.1).

Table A.1: Crystallographic data.

	hDHFR F31R/Q35E	His ₆ -hDHFR F31R/Q35E
Data collection statistics		
Space group	P2 ₁ 2 ₁ 2 ₁	P1
Number of molecules per asymmetric unit	1	6
a (Å)	42.348	62.605
b (Å)	47.868	83.334
c (Å)	90.715	83.272
α (°)	90	62.92
β (°)	90	78.68
γ (°)	90	71.81
Wavelength (Å)	1.5418	1.100
Resolution range (Å) [†]	1.70-10.52 (1.70-1.76)	2.21-40.89 (2.21-2.29)
Completeness (%) [†]	97.2 (93.9)	92.1 (94.2)
Redundancy [†]	11.7 (9.5)	3.6 (3.1)
R _{merge} (%) [†]	6.4 (47.3)	6.2 (15.6)
Refinement statistics		
Total number of reflections (reflections in R-free set)	20,304 (2066)	65,777 (3284)
R _{factor} (%)	17.93	31.420
R _{free} (%)	22.30 (10% of reflections)	40.028 (5% of reflections)
Number of atoms	1719	10,446
Protein	1507	9720
Water	138	240

	hDHFR F31R/Q35E	His ₆ -hDHFR F31R/Q35E
Co-factor	0	288
Ions	41	0
Inhibitor	33	198
RMSD		
Bond length (Å)	0.012	0.010
Bond angle (°)	1.417	1.847
Average atomic B-factor (Å ²)	16.372	32.789
Protein (Å ²)	14.99	32.85
Water (Å ²)	30.57	22.74
Co-factor (Å ²)	N/A	41.01
Ions (Å ²)	35.20	N/A
Inhibitor (Å ²)	15.78	30.06
Wilson B-factor (Å ²)	21.585	32.253
Luzzati sigma A coordinate error (observed) (Å)	0.12	N/D
Luzzati sigma A coordinate error (R-free set) (Å)	0.11	N/D
Ramachandran plot (non-Gly, non-Pro residues)	159 (100%)	1026 (100%)
Residues in favored positions	146 (91.8%)	822 (80.1%)
Residues in allowed positions	13 (8.2%)	137 (13.4%)
Residues in generously allowed positions	0 (0%)	34 (3.3%)
Residues in disallowed positions	0 (0%)	33 (3.2%)

[†]Items in parentheses refer to the highest resolution shell.

A.2.5 Crystallization and X-ray data collection of His₆-hDHFR F31R/Q35E

Purified His₆-hDHFR F31R/Q35E enzyme was buffer-exchanged into 50 mM Imidazole pH 7.5 and concentrated to 20 mg/mL using an Amicon concentrator (MWCO 10000). MTX and NADPH were prepared as described previously (Volpato *et al.*, 2007) and were added at a final concentration of 4 mM each (5-fold molar excess) to the protein sample. Crystallization experiments were set up using hanging drop vapour diffusion experiments, with a reservoir volume of 1 mL and a drop size of 4 µL of equal volumes of protein and reservoir solutions. A reservoir solution containing 0.1 M Tris pH 8.5, 0.01 M NiCl₂ and 21% PEG 2000 MME yielded crystals that diffracted at 2.2 Å. The crystals were soaked in the mother liquor supplemented with 30% PEG 400 as a cryoprotectant, frozen in a nitrogen cryostream (model X-stream 2000), and stored in liquid

nitrogen. Data collection was performed with a X8C beam at the National Synchrotron Light Source, Brookhaven National Laboratory, and processed using HKL2000 (Otwinowski & Minor, 1997) (Table A.1).

A.2.6 Structure Determination and Refinement of hDHFR F31R/Q35E

The structure was determined by molecular replacement using Phaser (Storoni *et al.*, 2004), which found a single protein molecule in the asymmetric unit (Resolution Range Used: 1.70 – 25.99 Å; Log Likelihood Gain (refined): 726.123). A lower quality His-tagged F31R/Q35E-MTX-NADPH structure (see A.2.7) was used as a molecular replacement model. Reciprocal-space refinement was performed using Refmac (Murshudov *et al.*, 1997) and included individual isotropic B-factor refinement as well as TLS refinement in the final stages of refinement. Manual model building was performed periodically using Coot (Table A.1) (Emsley & Cowtan, 2004).

A.2.7 Structure Determination and Refinement of His₆-hDHFR F31R/Q35E

The structure was determined by molecular replacement using Phaser (Storoni *et al.*, 2004), which found six protein molecules in the asymmetric unit (Resolution Range Used: 2.21-40.89; Log Likelihood Gain (Refined): 1932.291). PDB ID 1U72 was used as a molecular replacement model. Reciprocal-space refinement was performed using Refmac (Murshudov *et al.*, 1997), and included individual isotropic B-factor refinement. Manual model building was performed periodically using Coot (Table A.1) (Emsley & Cowtan, 2004). The full length of the hDHFR backbone could be modelled from the electron density map in all six molecules in the asymmetric unit, as could some portion of the N-terminal His₆-tag. In all six protein molecules in the asymmetric unit, large regions of electron density were observed in the DHF- and NADPH-binding sites. MTX and NADPH were modelled into these sites, respectively.

A.2.8 *In Silico* Automated Docking of Folate

The ligands were prepared as PDB files using ChemDraw 8.0 and Chem3D 8.0 (CambridgeSoft, Cambridge, MA). Energy minimization of ligands was performed with the integrated MM2 energy minimization script in Chem3D. Automated docking experiments were performed using the Autodock 4 software

package (Scripps Research Institute, La Jolla, CA). Macromolecules PDB ID 1U72 and 3EIG were stripped of all ligands and heteroatoms, with the exception of the highly conserved active site water molecule (H₂O #216 in 1U72, H₂O #244 in 3EIG), and were prepared using default settings. A box covering the entire folate binding site and more than half the NADPH binding site was generated as a docking grid. 50 runs of a Lamarckian genetic algorithm using default settings were performed. Following docking, clusters were evaluated according to total binding energies calculated by Autodock 4, and the minimal energy conformation within the lowest energy cluster was retained for comparison with crystal structures.

A.3 Results

A.3.1 Kinetic and Inhibition Double Mutant Cycle of hDHFR Variant F31R/Q35E

To determine the effect of each constituent amino acid substitution of the F31R/Q35E variant on DHF and MTX binding, the singly-substituted F31R and Q35E were created and their kinetic and inhibition parameters were determined (Table A.2). The reactivity (k_{cat}) of F31R and Q35E are 5-fold lower ($1.9 \pm 0.3 \text{ s}^{-1}$) and 2-fold lower ($4.6 \pm 0.2 \text{ s}^{-1}$) than WT hDHFR, respectively. The k_{cat} of F31R/Q35E is similar to that of the F31R variant, indicating that loss of reactivity in F31R/Q35E is primarily due to the F31R substitution. The Michaelis constants ($K_{\text{M}}^{\text{DHF}}$) of variants F31R and Q35E are 110 ± 60 and $250 \pm 90 \text{ nM}$, respectively, illustrating a slight decrease in DHF affinity (1.5- and 3-fold, respectively) relative to the WT. The high % error on $K_{\text{M}}^{\text{DHF}}$ results from the low values of $K_{\text{M}}^{\text{DHF}}$ and the correspondingly low spectrophotometric signal, which was enhanced by the use of 10 cm path length cuvettes. Kinetic data for the F31R variant had previously been reported (Patel *et al.*, 1997), and compares well with our data, although we determined a $K_{\text{M}}^{\text{DHF}}$ value that is 6-fold lower than previously reported. Because we expected a low value for $K_{\text{M}}^{\text{DHF}}$ and a correspondingly low spectrophotometric signal, we used 10 cm cuvettes in $K_{\text{M}}^{\text{DHF}}$ determination, rather than 1 cm cuvettes (Patel *et al.*, 1997). This enabled a more precise measurement in the target range. Inhibition constants for MTX ($K_{\text{i}}^{\text{MTX}}$)

revealed that the F31R substitution ($K_i^{\text{MTX}} = 1.1 \text{ nM}$) confers a 35-fold loss in MTX affinity, whereas the Q35E substitution ($K_i^{\text{MTX}} = 0.048 \text{ nM}$) displays a modest decrease of MTX affinity relative to WT (1.5-fold decrease). Like the F31R variant, the F31R/Q35E variant displayed larger decreases in MTX affinity than DHF affinity relative to the WT. However, the Q35E substitution modestly decreased DHF affinity, while having a negligible effect on MTX affinity. The F31R/Q35E variant presented both of these features, as DHF affinity decreased 9-fold relative to WT, whereas MTX affinity decreased >650-fold relative to the WT. Thus, addition of the Q35E substitution to variant F31R increased the K_M^{DHF} of F31R 6-fold while increasing the K_i^{MTX} of F31R nearly 20-fold.

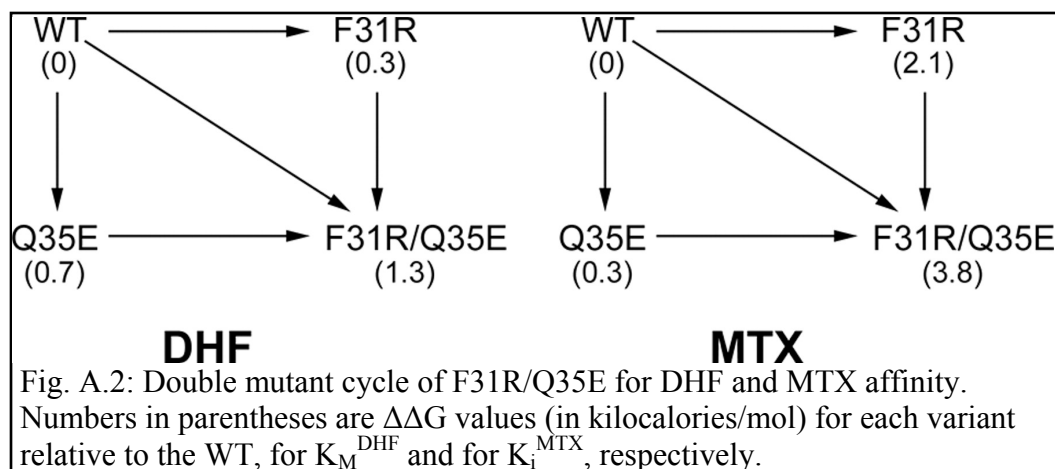
To better quantify the effect of the combined substitutions on MTX and DHF binding, we calculated the loss of binding free energy ($\Delta\Delta G$) for each variant relative to WT hDHFR. For K_M^{DHF} , the sum of the $\Delta\Delta G^{\text{F31R}}$ and $\Delta\Delta G^{\text{Q35E}}$ values was comparable to the value of $\Delta\Delta G^{\text{F31R/Q35E}}$ (Table A.2 and Fig. A.2), indicating additive effects of the two substitutions toward loss of DHF affinity. However, for K_i^{MTX} , $\Delta\Delta G^{\text{F31R/Q35E}}$ was considerably greater than the sum of $\Delta\Delta G^{\text{F31R}}$ and $\Delta\Delta G^{\text{Q35E}}$, indicating a synergistic effect of the two substitutions on loss of MTX affinity (Table A.2 and Fig. A.2). These data demonstrate that the F31R substitution is the most important contributor to the binding properties of the F31R/Q35E variant, while the addition of the Q35E substitution synergistically decreases MTX affinity with only a modest, additive reduction of DHF affinity.

Table A.2: Kinetic and inhibitory parameters of WT hDHFR and hDHFR variant F31R/Q35E

Variant	k_{cat} s^{-1}	K_M^{DHF} nM	$k_{\text{cat}}/K_M^{\text{DHF}}$ $\text{s}^{-1} \mu\text{M}^{-1}$	$\Delta\Delta G^{\text{DHF}\dagger}$ kcal/mol	$\text{IC}_{50}^{\text{MTX}}$ nM	K_i^{MTX} nM	$\Delta\Delta G^{\text{MTX}\ddagger}$ kcal/mol
WT*	10±2	<75	>130	0	41±14	<0.031	0
F31R	1.9±0.3	110±60	17±12	0.3	1100±600	1.1±0.6	2.1
Q35E	4.6±0.2	250±90	18±7	0.7	19±4	0.048±0.009	0.3
F31R/Q35E*	1.3±0.2	690±13	1.9±0.2	1.3	3100±1600	21±11	3.8

$\dagger \Delta\Delta G = -RT \times \ln(K_M^{\text{DHF}} \text{ WT} / K_M^{\text{DHF}} \text{ variant})$; T = 293 K.

$\ddagger \Delta\Delta G = -RT \times \ln(K_i^{\text{MTX}} \text{ WT} / K_i^{\text{MTX}} \text{ variant})$; T = 293 K.

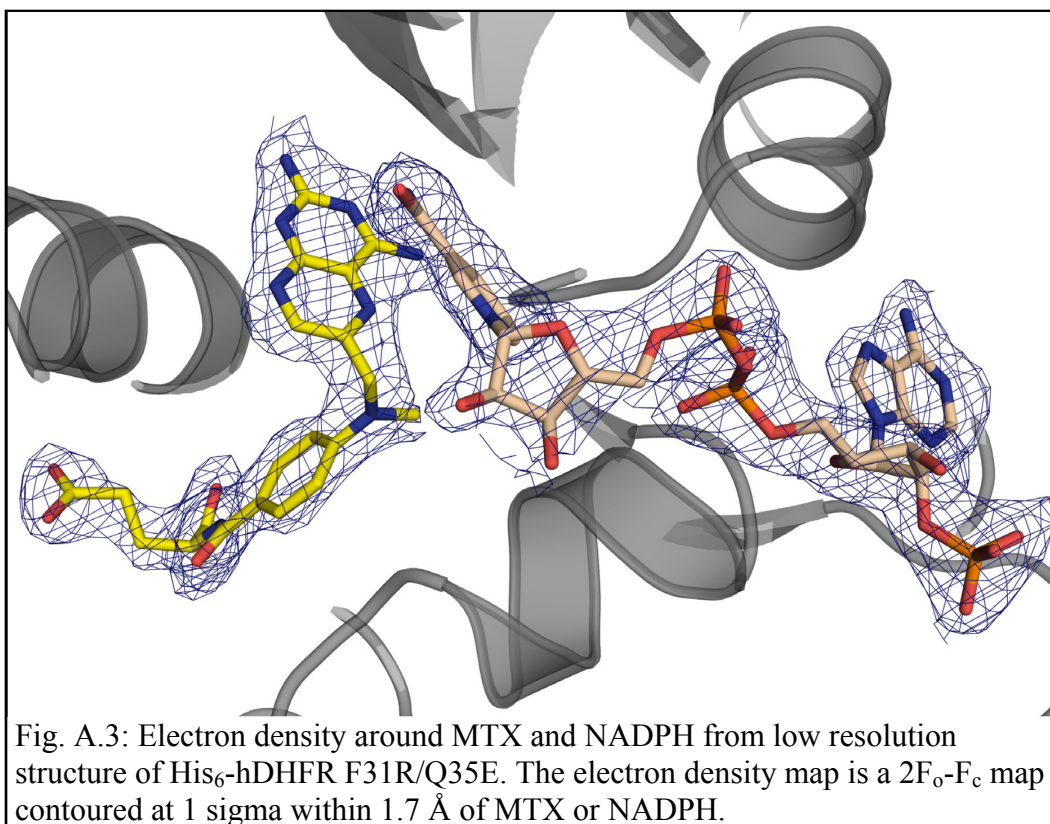


*Values were taken from Fossati *et al.* (2008) and Volpato *et al.* (2007).

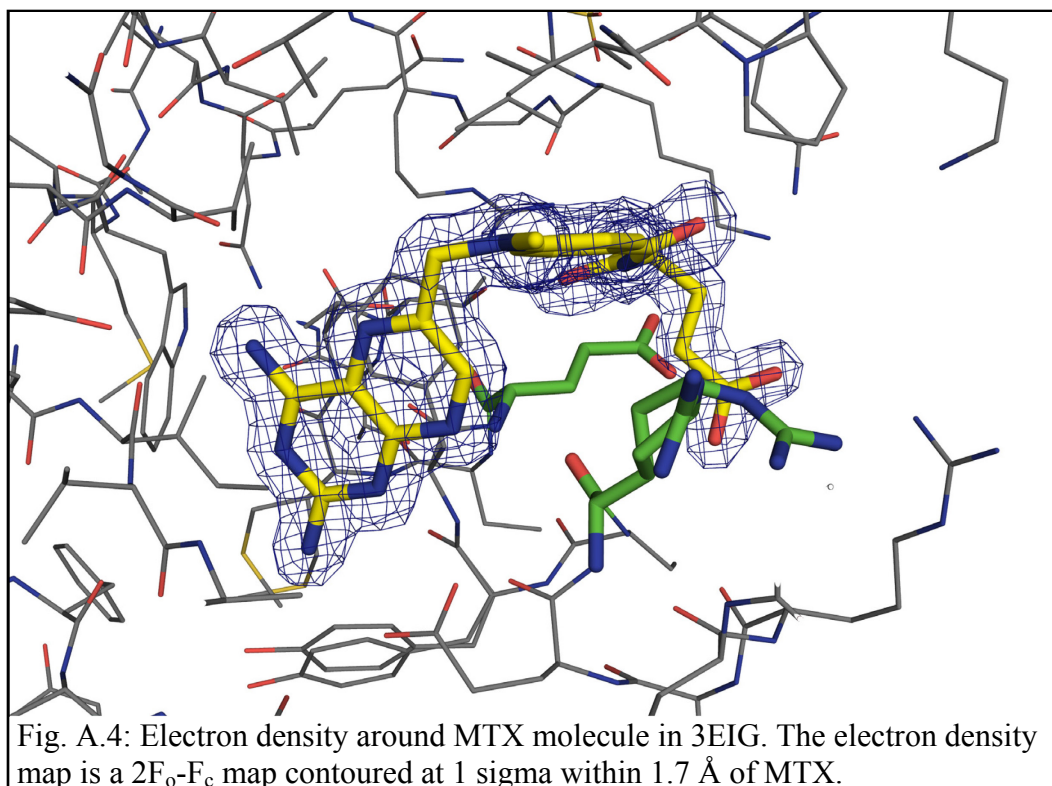
A.3.2 Overall Structure and General Characteristics of hDHFR F31R/Q35E Complexed to MTX

To further understand the structural basis of the important loss in MTX affinity, we obtained two crystal structures of variant F31R/Q35E. The first structure had bound NADPH and MTX and contained six hDHFR molecules within the asymmetric unit (Fig. A.3). However, considerable twinning of the crystal lattice resulted in poor data quality, leading to an inability to refine the crystal structure to acceptable R_{factor} and R_{free} values (Table A.1). This precluded detailed analysis of molecular contacts; however, we used this structure to corroborate observations in the second structure. The full-length of the hDHFR backbone was modeled from electron density map in the second structure, with several side chains exhibiting multiple conformations. Some side chains did not display well defined electron density, and thus were excluded from the model. A large region of very well defined electron density was modeled as MTX in the DHF-binding site (Fig. A.4). There was no electron density in the NADPH-binding site with which to model in NADPH, although two peaks modeled in as sulfate ions were visible where the phosphates of NADPH have been observed in other structures. Six other sulfates, as well as a cadmium ion, which originates from the crystal seed stabilization solution, were also built into the model.

The overall fold and tertiary structure of variant F31R/Q35E are very similar to those reported for other hDHFRs (Oefner *et al.*, 1988; Davies *et al.*,



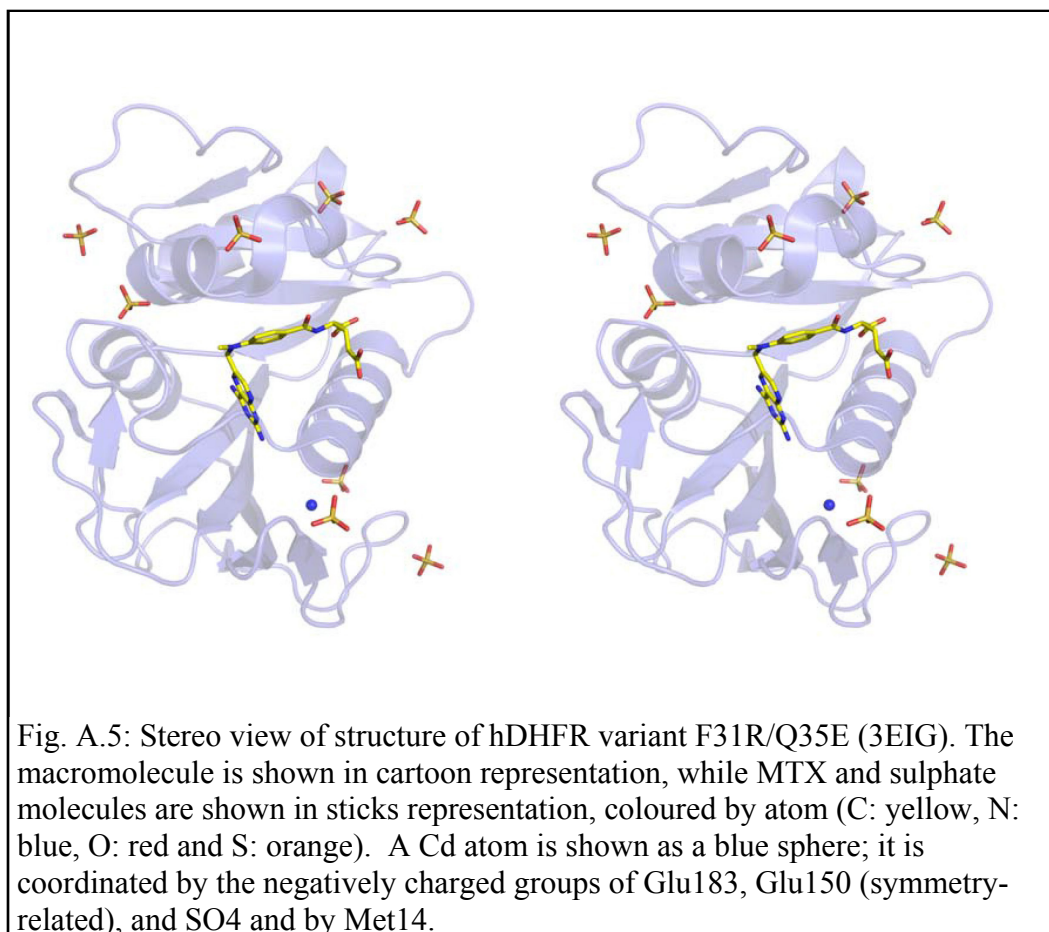
1990; Lewis *et al.*, 1995; Cody *et al.*, 1997; Cody *et al.*, 1998; Gangjee *et al.*, 1998; Klon *et al.*, 2002; Cody *et al.*, 2003a; Cody *et al.*, 2003b; Cody *et al.*, 2004; Cody *et al.*, 2005). hDHFR F31R/Q35E is composed of a central β -sheet containing seven parallel and one anti-parallel strands and four α -helices interconnected by a series of loops (Fig. A.5). Density surrounding the side chains of residues Arg-32, Arg-36, Glu-78, Glu-81, His-87, Arg-91, Lys-98, Glu-101, and Glu-161 was either absent or poorly defined. Those side chains, which are all at the surface of the protein, were not included in the model. Nonetheless, a number of residues were revealed to have two conformers that were clearly visible. Interestingly, these residues are clustered either in the folate (F31R, Tyr-33, Met-37, and Ser-41) or NADPH (Ser-59, Ser-118, Asp-145, and Thr-146) binding sites (Fig. A.6). Other high resolution structures have reported two conformers at Ser-41, Ser-118, and Thr-146 when NADPH was not present in the active site (Klon *et al.*, 2002), but the presence of two conformers at the other residues (F31R, Tyr-33, Met-37, Ser-59, and Asp-145) has never been previously



reported and suggests that this is a consequence of the substitutions at residues 31 and/or 35.

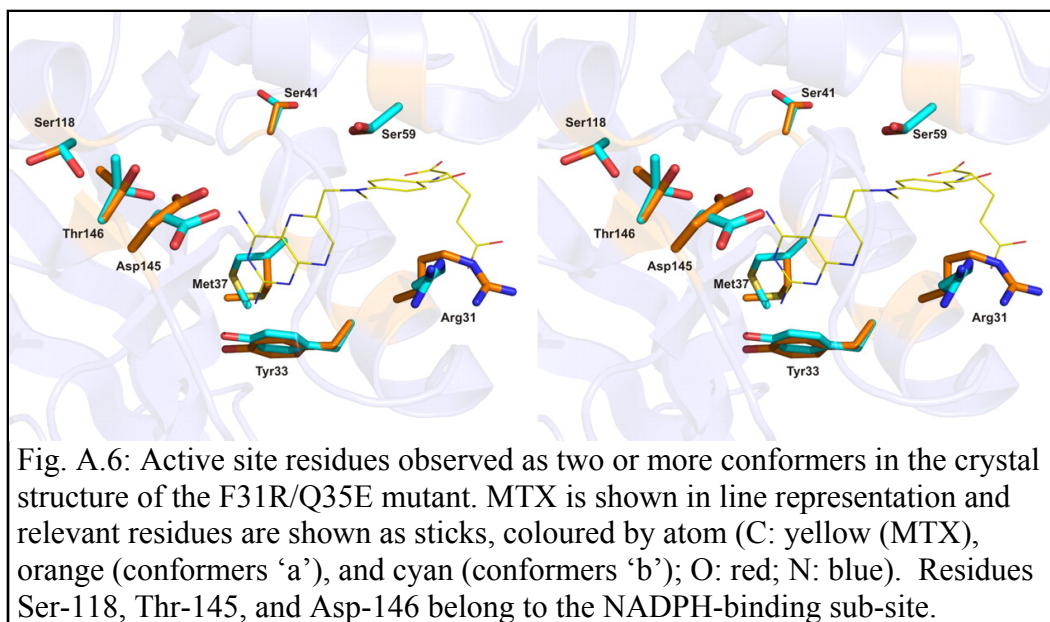
A.3.3 Interactions between F31R/Q35E and Bound MTX at the Active Site

To identify structural changes in variant F31R/Q35E complexed with MTX, the 1.7 Å resolution structure was superimposed onto WT human DHFR complexed with NADPH and MTX (1U72, 1.9 Å resolution) (Cody *et al.*, 2005) and WT human DHFR complexed with folate (1DRF, 2.0 Å resolution) (Oefner *et al.*, 1988) (RMSD 0.66 and 0.69 Å, respectively). Those structures were selected because they contain structurally similar ligands in the folate binding site, and because 1DRF lacked NADPH, which enabled identification of structural differences originating from NADPH binding. In contrast to the two WT structures, variant F31R/Q35E displays a considerable shift of loop 17–27 toward the active site (discussed below), in addition to the residues present as two conformers in the NADPH-binding cleft. The low RMSD value (0.304 Å) between 1U72 and 1DRF was also an incentive for comparison with variant F31R/Q35E. MTX binds in a similar fashion in the F31R/Q35E active site (Fig. A.7) as in WT hDHFR (1U72). There was little difference in the orientation of



the side chains of residues involved in binding, with the exception of two active site residues, Arg-31 and Tyr-33, which were each present in two distinct conformations. Therefore, as in the WT enzyme, residues 31 and 35 in variant F31R/Q35E do not form specific contacts with the pterin moiety of MTX.

Binding of the pterin ring involves characteristic H-bonding with specific amino acids and with conserved water molecules (Fig. A.7A). H₂O #210 is also H-bonded to the hydroxyl group of one of the two distinct conformers of Tyr-33 observed in the F31R/Q35E structure (Fig. A.7A). To our knowledge, this is the first hDHFR structure that reports two distinct Tyr-33 conformers in the same macromolecule, identified as Tyr-33A and Tyr-33B (Fig. A.8). The Tyr-33A conformer is homologous to that observed in all but one structure of hDHFR. It is positioned to form the same close ring stacking interaction with Phe-179 that has been previously described for murine DHFR (Cody *et al.*, 2006) and that exists in WT hDHFR. A $\sim 30^\circ$ rotation around the C $_{\alpha}$ –C $_{\beta}$ bond toward the active site slides



the hydroxyl group 2 Å away from its initial conformation, slightly increasing the ring-stacking distance with Phe-179 while bringing the hydroxyl within H-bonding distance of H₂O #210 (Tyr-33B (Fig. A.7A)); this may result in the Tyr-33B conformer being slightly less stabilized than the Tyr-33A. A similar H-bond is observed in WT hDHFR complexed with NADPH and PT523 (PDB ID 1OHK), a MTX-like inhibitor (Cody *et al.*, 1997). Hydrophobic and van der Waals interactions are also formed between the pterin moiety of MTX and variant F31R/Q35E (Fig. A.7B). The side chains of Ile-7, Ala-9, Leu-22, Phe-34, and Val-115 are all within van der Waals distance of the pterin ring.

Hydrophobic contacts are predominant in binding of the *p*-aminobenzoic acid (*p*-ABA) moiety of MTX, via residues Phe-34, Ile-60, Pro-61, and Leu-67 (Fig. A.6B). The two side-chain conformers of the mutated Arg-31 (Arg-31A and Arg-31B; Fig. A.6C and Fig. A.8B) are also within van der Waals distance of the *p*-ABA phenyl ring. However, Arg-31 cannot establish the hydrophobic and possibly edge-to-face aromatic contacts occurring between with WT residue Phe-31 and the *p*-ABA phenyl ring, the consequences of which will be discussed below. Importantly, the side-chain conformation adopted by four of the six Arg-31 residues from the lower resolution structure we obtained clustered about conformation Arg-31B (Fig. A.9A). The other two Arg-31 residues from the lower resolution structure clustered together in a new conformation, whereas none

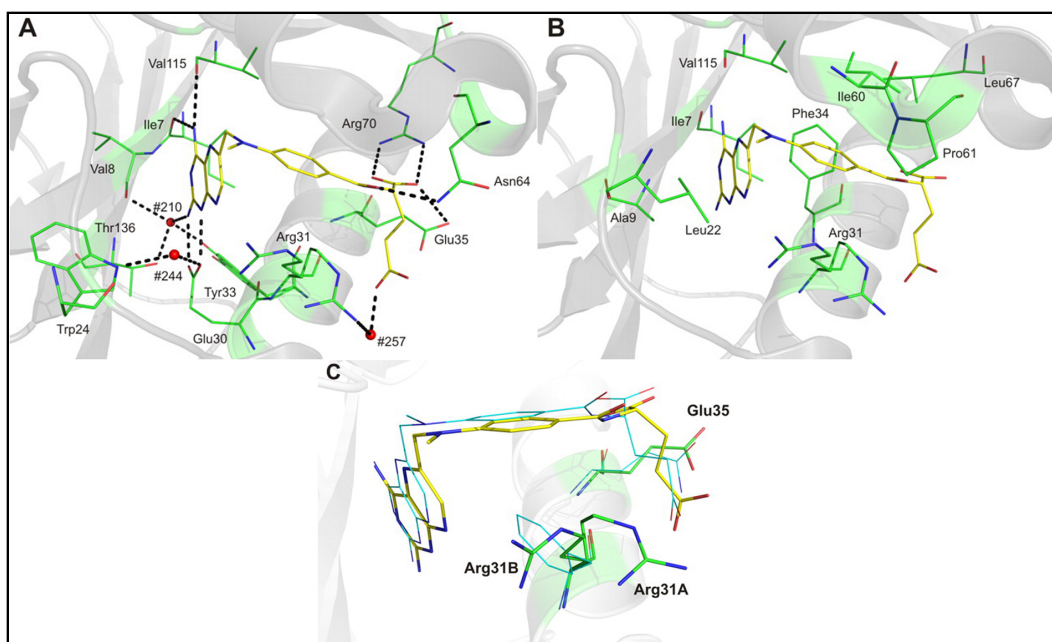
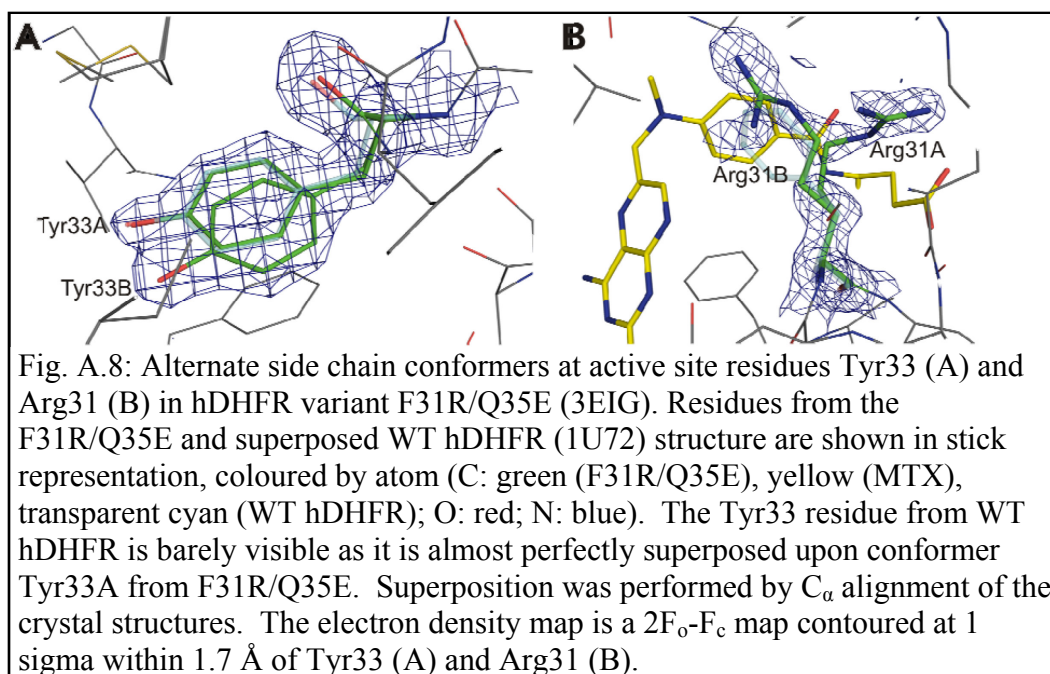
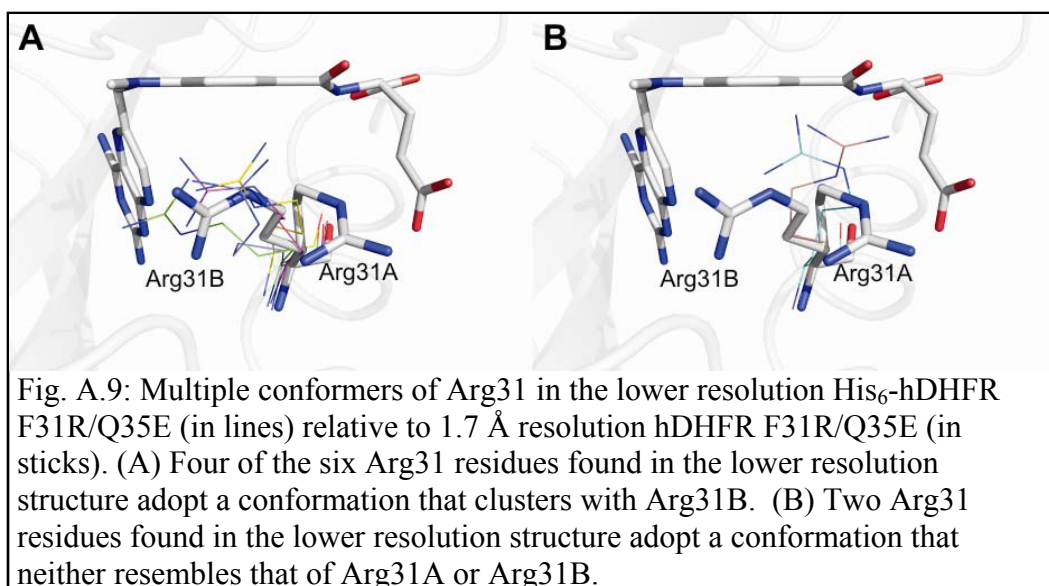


Fig. A.7: Bound MTX in hDHFR variant F31R/Q35E. Polar (A) and non-polar interactions (B). MTX is shown in stick representation, and relevant residues are shown as lines, coloured by atom (C: yellow (MTX) and green (active-site residues); O: red; N: blue). In A, H-bonds and salt bridges are shown as dashed black lines, while active-site water molecules #210, #244, and #257 are shown as red spheres. The backbone carbonyls of Ile-7 and Val-115 are within H-bonding distance of the pterin 4-amino group, as is the hydroxyl group of Tyr-121. The carboxylate of the catalytic Glu-30 residue forms a salt bridge with the pterin N₁ and 2-amino group. A conserved active-site water molecule (H₂O #257), coordinated via H-bonding interactions with the indole ring of Trp-24 and the Glu-30 carboxylate group, is within H-bonding distance of the pterin N₈. Another highly conserved water molecule (H₂O #210) present in the active site can H-bond with the pterin 2-amino group, the backbone carbonyl of Val-8, and the hydroxyl group of Thr-136. The *p*-ABA moiety of MTX interacts mainly with residues Phe-34, Ile-60, Pro-61, and Leu-67 through hydrophobic contacts. A sole H-bond is formed between the carbonyl group of the *p*-ABA moiety and the γ -amino group of residue Asn-64. C, position of MTX, Arg-31, and Glu-35 in F31R/Q35E relative to the position observed in WT hDHFR (1U72). Residues and MTX from the F31R/Q35E structure are shown in stick representation, whereas residues and MTX from superposed WT hDHFR (1U72) are shown as lines, coloured by atom (C: green (F31R/Q35E), cyan (WT hDHFR), and yellow (MTX from 3EIG); O: red; N: blue). Superposition was performed by C α alignment of the crystal structures.

adopted the Arg-31A conformation (Fig. A.9B). The high conformational variation of this residue provides evidence that the conformers result from the amino acid substitutions, rather than from crystallization artifacts.



The glutamate moiety of MTX is mostly solvent-exposed and interacts mainly via polar contacts with residues of the active site and water molecules at the surface of the protein. The most characteristic contact is the salt bridge formed between the guanidinium group of Arg-70 in variant F31R/Q35E and the α -carboxylate of the glutamate moiety. Arg-70 is strictly conserved in DHFRs from all species, and this interaction is present in all crystal structures complexed with ligands containing a glutamate moiety. In the F31R/Q35E variant, the α -carboxylate of MTX is also within H-bonding distance of a network involving three water molecules (H₂O #232, 256, and 257) and the carbonyl group of the *p*-ABA moiety. This H-bond network has not been observed in other hDHFR structures. Molecule H₂O #257 may also H-bond with the γ -carboxylate of MTX and the backbone carbonyl of Arg-28. The backbone nitrogen of Arg-32 is within H-bonding distance of the γ -carboxylate group of MTX as is the ϵ NH group of mutated Arg-31A. The Arg-31A conformer places the side chain in proximity to the glutamate portion of MTX and enables H-bonding with H₂O #257 and the MTX γ -carboxylate.



A.3.4 Differences in MTX Binding Resulting from the F31R and Q35E

Substitutions

Despite similar interactions formed with MTX in variant F31R/Q35E and WT hDHFR (1U72), differences were observed that are consistent with weakened MTX binding in the variant (Fig. A.7C). Table A.3 lists the distances for apparent polar contacts between the enzyme and the inhibitor, in F31R/Q35E and 1U72. With the exception of unique contacts related to each structure, which mainly involve H-bonds mediated with H₂O #210 and 257, the most important differences were observed in the Trp-24-H₂O #244 H-bond (0.8 Å shorter) and the Glu-30 εO1-N₈ of MTX H-bond (0.6 Å shorter). The change in distance of the Trp-24-H₂O #244 H-bond can be attributed to the shift of residues 17–27 (Fig. A.10). This shift enables closer contacts between the Trp-24 indole ring and the conserved water molecule in variant F31R/Q35E. Another difference results from a specific rotation around the C6–C9 of the pterin ring of MTX in F31R/Q35E relative to 1U72. The slight rotation (~7°) around the C6–C9 bond brings the pterin 2-amino group slightly closer (0.6 Å) to the catalytic Glu-30 residue. These structural changes in residues that interact with the pterin-moiety appear to be induced by the substitutions at positions 31 and 35, which do not form direct contacts with the pterin moiety.

Table A.3: Polar interactions in F31R/Q35E and WT hDHFR (1U72) complexed with MTX

Polar contact	Distance F31R/Q35E (Å)	Distance 1U72 (Å) (Cody <i>et al.</i> , 2005)
Ile-7 O-4-NH ₂ MTX	2.9	2.7
Val-8 N-HOH#210	4.1	
HOH#210-2-NH ₂ MTX	3.1	
Trp-24 N-HOH#244	3.1	3.9
HOH#244-N8 MTX	3.2	3.2
Arg-28 O-HOH#257	2.8	
HOH#257-Oε1 Glu MTX	3.2	
Glu-30 Oε1-2-NH ₂ MTX	2.7	3.3
Glu-30 Oε2-N1 MTX	2.8	2.9
Glu-30 Oε2-HOH#244	2.8	2.6
Arg-31A NH1-HOH#257	3.3	
Arg-32 N-Oε2 Glu MTX	3.2	
Tyr-33B OH-HOH#210	2.7	
Gln-35 Oε-O1 Glu MTX		3.3
Asn-64 Nδ-O <i>p</i> -ABA MTX	2.8	2.8
Arg-70 NH1-O2 Glu MTX	3.0	2.4
Arg-70 NH ₂ -O1 Glu MTX	2.8	2.7
Val-115 O-4-NH ₂ MTX	2.9	3.2
Tyr-121 O-4-NH ₂ MTX	3.3	3.3
Thr-136 OH-HOH#210	2.7	

The *p*-ABA portion of MTX is shifted by ~ 0.6 Å in F31R/Q35E relative to WT hDHFR, bringing this moiety closer to the mutated Arg-31, for which two conformers were resolved. Arg-31B (Fig. A.8B) points in roughly the same direction as the WT Phe-31 residue. Its χ_1 angle (-81°) is similar to the WT residue (-84°), such that the guanidinium side chain occupies the area of the active site normally occupied by Phe-31. However, a rotation of $\sim 90^\circ$ about the C $_{\beta}$ -C $_{\gamma}$ bond of Arg-31B takes the δCH_2 and ϵNH groups out of the plane of the Phe-31 phenyl group (Fig. A.8B). This results in a loss of hydrophobic and van der Waals contacts, consistent with decreased MTX affinity in variant F31R/Q35E. Nonetheless, the bulky Arg side chain conserved some van der Waals contacts with the *p*-ABA moiety of the inhibitor, and so this may not be sufficient to rationalize the large decrease in MTX affinity. The shift of the *p*-ABA moiety also brings the MTX N₁₀-methyl group closer to residue 22 in F31R/Q35E as a

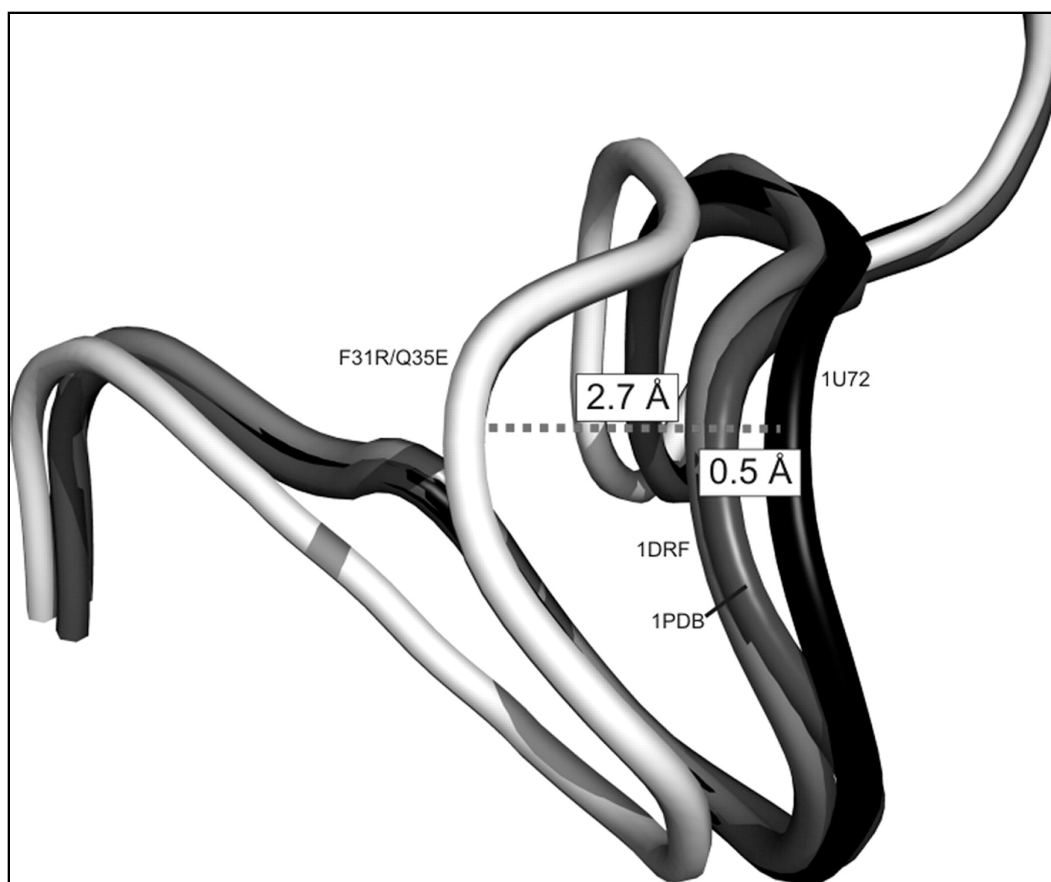


Fig. A.10: Shift of loop 17–27 in hDHFR variant F31R/Q35E. Loops are shown in schematic representation. Residues 17–27 are shown for variant F31R/Q35E (white), and WT hDHFR from three structures: 1U72 (complexed with NADPH and MTX; black), 1DRF (complexed with folate; light gray), and 1PDB (apoenzyme; dark gray). Distances were calculated between the 1U72 C α of Gly-20 of 1U72 and the respective C α of Gly-20 for the three structures. Superposition was performed by C α alignment of the crystal structures.

result of a 7° rotation about the N₁₀–C_{1'} bond. Leu-22 belongs to the 17–27 loop, which displays an important shift toward the active site in the variant structure. This loop shift may be partly attributable to the change of position of residue Leu-22 (0.8 Å C α –C α distance, relative to 1U72) to maximize van der Waals interactions with MTX, and also with the side chain of Arg-31B. In variant F31R/Q35E, one of the η NH groups of Arg-31B is in close proximity to the side chain of Leu-22 (Fig. A.11B), mimicking an interaction observed with Phe-31 in WT hDHFR (Fig. A.11A). Considering that the van der Waals radius for nitrogen is 0.2 Å smaller than for carbon, Leu-22 must therefore be closer to the Arg-31 side chain to optimize van der Waals interactions.

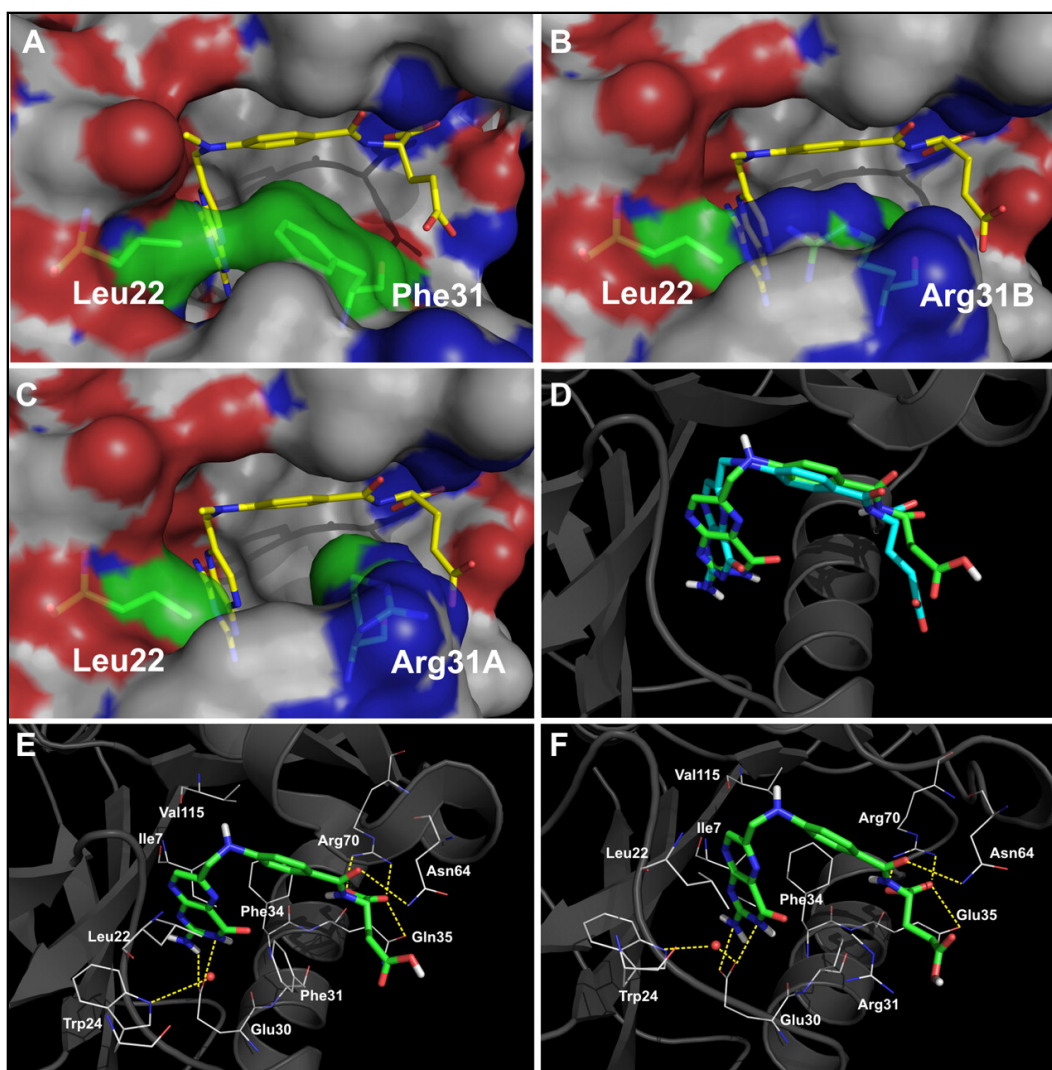


Fig. A.11: Comparison of WT hDHFR and variant F31R/Q35E by modeling. (A-C) Surface representation of the contacts established between residues 22 and 31 in WT hDHFR (A, 1U72), variant F31R/Q35E with Arg-31B (B) or with Arg-31A (C) bound to MTX. MTX and residues 22 and 31 are in stick representation, coloured by atom (C: yellow (MTX) and green (residues 22 and 31), O: red, N: blue). Surface is coloured by atoms (C: white, O: red, N: blue). (D-F) Docking of folate onto WT hDHFR and variant F31R/Q35E. (D) Superposition of the original crystal structure 1DRF (WT DHFR bound to folate in blue) and the docking model of 1U72 (WT DHFR) docked with folate (in green). Results for the minimum energy binding conformers are shown for folate (E) docked onto WT hDHFR (1U72), as well as for folate (F) docked onto F31R/Q35E with Arg-31A conformer. The ligands are shown in stick representation while the residues are shown as lines, coloured by atom (C: green (ligands) and white (residues), O: red, N: blue). Superposition was performed by C_{α} alignment of the crystal structures.

The second Arg-31 conformer, Arg-31A, is rotated about C_{α} – C_{β} by $>70^{\circ}$ relative to Arg-31B and the native Phe-31. This rotation brings the guanidinium

side chain closer to the γ -carboxylate group of MTX, where it H-bonds with the γ -carboxylate via a water molecule. As shown in Fig. A.11C, considerable hydrophobic and van der Waals contacts with MTX are lost when Arg-31 adopts this second conformation, again consistent with the important decrease of MTX affinity in the variant.

The Gln-35 \rightarrow Glu substitution also appears to contribute, however slightly, to the decreased MTX affinity. Residue 35 is located at the surface of the protein. In the WT, the side chain of Gln35 H-bonds with the MTX α -glutamate, but in F31R/Q35E, the Glu-35 side chain is slightly shifted away from the glutamate moiety of MTX relative to the WT 1U72 structure. As a result, the Glu-35 side chain points away from the MTX α -glutamate group such that H-bonding would be weakened or absent. An important rotation around the N-C $_{\alpha}$ bond of the MTX glutamate moiety (34° relative to 1U72), which tilts the α -carboxylate group closer to helix $\alpha 1$, could be caused by unfavorable electrostatic interactions due to two close negative charges (the Glu-35 side chain and the glutamate moiety of MTX). This tilt of the MTX glutamate portion results in the lengthening of the salt bridge between Arg-70 ϵ NH1 and α O $_2$ of MTX (0.6 Å difference) in variant F31R/Q35E.

A further potential role of the Q35E substitution in reducing MTX affinity may be a reorientation of Glu-35 in the absence of MTX or other ligands, to form a salt bridge with the guanidinium group of Arg-70. This speculation is based on the presence of a H-bond between Gln-35 and Arg-70 in the WT hDHFR apoenzyme structure following repositioning of the Gln-35 side chain relative to WT hDHFR bound to MTX (Cody *et al.*, 2003b). An intramolecular salt bridge between Glu-35 and Arg-70 would hinder the binding of the α -carboxylate group of MTX or DHF, thereby decreasing affinity for either compound. However, the kinetic data obtained for the Q35E variant suggests that, if these effects are present, they are not the predominant cause for important loss of binding of either ligand.

A.3.5 Docking of Folate upon the Crystallized F31R/Q35E Structure

One of the key features of variant F31R/Q35E is the more important decrease in MTX affinity relative to DHF affinity. This may be attributed to the flip of the pterin ring of folate relative to MTX, resulting in establishment of different, specific contacts of either ligand with the active-site residues. As mentioned above, the variant residues 31 and 35 do not interact directly with the pterin moiety of MTX. It is thus unlikely that the selective decrease in MTX affinity relative to DHF could result from different contacts with the variant residues. The crystal structure presented herein shows that the substitutions mainly perturb interactions with the *p*-ABA and glutamate portions of the bound inhibitor. WT hDHFR structures complexed with folate, the more oxidized form of the substrate, have shown that the *p*-ABA and glutamate portions bind similarly to those of MTX. Therefore, loss of contacts with these portions due to substitutions would be expected to reduce the affinity by a similar factor for either ligand. As our attempts to obtain high quality crystals of F31R/Q35E with folate have been unsuccessful thus far, molecular docking studies were performed with variant F31R/Q35E to provide insight into the potential binding mode of folate in the variant enzyme, using the WT hDHFR (PDB ID 1U72) as a reference. Because NADPH was not present in the structure of variant F31R/Q35E, it was removed from the 1U72 structure. For F31R/Q35E, two independent structures, containing either of the observed conformers at Arg-31 (Fig. A.8B), were created. Because the minimal energy conformers for all tested ligands were indistinguishable for the two conformers of Arg-31, we report only the results obtained for F31R/Q35E Arg-31A.

As a control for the docking protocol, MTX was docked onto 1U72 and F31R/Q35E structures. The docking results confirmed that the protocol enabled good prediction of interactions between MTX and the macromolecules. The minimal energy conformers of docked MTX for WT hDHFR or F31R/Q35E closely resembled the binding observed in the respective crystal structures (RMSD MTX to MTX = 0.9 Å, results not shown). Furthermore, the minimal energy conformer of folate docked onto the 1U72 structure was superimposable

with the crystallized folate molecule contained in the 1DRF crystal structure (RMSD folate to folate = 1.2 Å, results not shown). The orientation of the pterin ring of folate and MTX was correctly predicted in all the minimal energy conformers obtained. Additionally, all known contacts (except when involving non-active site water molecules) were present, including those of the *p*-ABA and glutamate moieties (Fig. A.11D–F). The docking studies with folate suggest that the *p*-ABA and glutamate moieties of folate and MTX bind similarly to the active site of variant F31R/Q35E. If the loss of MTX affinity in this variant were uniquely attributable to loss of interactions with residues 31 and 35, we would expect the MTX and DHF affinities to decrease by a similar factor. It thus appears that the main effect of substitutions F31R and Q35E is to decrease local order within the active-site area, as evidenced by the number of residues observed as more than one conformer. The resulting active site composition is more detrimental to MTX affinity than to DHF affinity, resulting in reduced discrimination between these two ligands.

A.4 Discussion

We report the structure of the highly MTX-resistant, doubly substituted hDHFR variant F31R/Q35E bound to MTX. A major obstacle to effective gene therapy for treatment of hematologic disorders is the low transduction efficiency that reduces effectiveness of bone marrow repopulation with cells carrying the desired gene. We have shown that this variant of hDHFR allows rapid and efficient selection of hematopoietic cells, offering the potential to address this issue (Volpato *et al.*, 2011). The distinguishing feature of this variant, relative to others that have previously been tested toward this goal (Flasshove *et al.*, 1995; Ercikan-Abali *et al.*, 1996a; Ercikan-Abali *et al.*, 1996b), is the much larger decrease in MTX affinity (>650-fold decrease of MTX affinity relative to WT hDHFR) than DHF affinity (9-fold relative to WT hDHFR). When performing directed evolution of this enzyme, we had selected for a decrease in MTX affinity and concomitant retention of catalytic activity (hence, substrate binding) (Volpato *et al.*, 2007). This is a particularly difficult task, considering the structural similarity between the substrate and the inhibitor, and the observation (from a

number of DHFR mutants) that a decrease in affinity for the one has generally been accompanied by a similar decrease in affinity for the other (Blakley & Sorrentino, 1998). Nonetheless, it has been recently suggested that the native function of enzymes tend to be more resistant to mutations than drug binding, which is a promiscuous (secondary) property of enzymes (Nobeli *et al.*, 2009).

In this variant, the mutations resulting in the desired phenotype provided weaker discrimination between the ligands. The MTX resistance appears to be attributed to loss of van der Waals contacts between Arg-31 and the MTX *p*-ABA moiety resulting in part from multiple conformations adopted by this side chain, as well as unfavorable electrostatic contacts between Glu-35 and the glutamate portion of MTX. These differences could also account for the decrease of DHF affinity, which also contains *p*-ABA and glutamate moieties. Kinetic characterization of the simple F31R and Q35E variants showed that the decrease of MTX affinity in variant F31R/Q35E was attributable to a synergistic effect of the combined substitutions, while the decrease of DHF affinity was attributable to an additive effect. Docking studies predicted that folate would bind to variant F31R/Q35E in a similar manner to what is observed in WT hDHFR crystal structures bound to folate or MTX (Oefner *et al.*, 1988; Davies *et al.*, 1990), and that loss of contacts with the *p*-ABA and glutamate moieties of MTX should also prevail with DHF or folate. This suggests that the larger decrease in MTX affinity is not solely caused by loss of contacts between the enzyme and inhibitor.

In addition to the broad distribution of conformers observed for our two structures for the mutated F31R residue, seven non-mutated residues in proximity to the folate-binding site were observed in two distinct conformers (Fig. A.6). Such a far-reaching disruption of active-site order has never been observed in any other vertebrate DHFR structure reported to date, where few or no secondary conformers are observed even at high resolution. This observation is striking because two neighboring mutations on a single α -helix (F31R/Q35E) have led to apparent disorder throughout the active-site region. This suggests that variant F31R/Q35E has a more dynamic character than the WT hDHFR and suggests the mode by which the increased ligand discrimination operates. There has been a

great amount of work reported over the past 10 years relating protein dynamics and function (reviewed by Doucet and Pelletier (2009)), including human immunodeficiency virus protease and antibodies and, more specifically, with drug binding (reviewed by Teague (2003)). Further work has shown that variations of ligand binding and catalytic activity can be related to amino acid substitutions that change the dynamics of an enzyme (Eisenmesser *et al.*, 2005; Doucet *et al.*, 2007; Watt *et al.*, 2007). Here, putative dynamic effects throughout the active-site area as a result of two substitutions appear to provide a selective decrease of inhibitor binding. The structural resolution of hDHFR variant F31R/Q35E provides insight into an unsuspected path to drug resistance, representing a significant advancement into our understanding of active-site mutations on drug binding. Although our structural data provide information on the MTX bound enzyme, we may assume that yet more active-site disorder would be observed in the free enzyme. Ligand binding would occur at the expense of greater entropy, which would be reflected in the binding constants; nonetheless, the data currently available does not allow us to provide an entropic basis for ligand discrimination.

As indicated by the inhibition constants, the F31R substitution is the main contributor to decreased MTX affinity. Structural data has suggested that the decrease in MTX affinity for variants F31A, F31S, and F31G could be due to the loss of van der Waals interactions (Chunduru *et al.*, 1994). Although some loss of contacts between MTX and residue 31 is apparent in the F31R/Q35E structure, it is hard to reconcile with the fact that variants F31R or F31R/Q35E display larger decreases in MTX affinity than the F31G variant (Chunduru *et al.*, 1994), where no contacts can be formed between the ligand and residue 31. This suggests that the basis for the greater decrease of MTX affinity relative to DHF affinity is attributable to a feature that specifically promotes MTX affinity.

It has been shown experimentally that an isomerization step following initial cofactor and MTX binding increases MTX affinity >60-fold in the native enzyme, leading to a non-dissociating hDHFR·NADPH·MTX complex (Appleman *et al.*, 1988). It has been suggested that Phe-31 is a key residue in this isomerization (Chunduru *et al.*, 1994). This assumption was based on two

observations: the presence of a second conformer at the homologous Tyr-31 residue in chicken DHFR bound to NADP⁺ and biopterin (McTigue *et al.*, 1992) and a non-native-like Phe-31 conformer present in one of the two macromolecules observed in WT hDHFR complexed with folate (Davies *et al.*, 1990). This suggested a possible readjustment of this residue following ligand binding, which would be required for isomerization.

The multiple conformers at residue 31 when MTX is bound to variant F31R/Q35E suggest two possibilities. The substitution of Phe-31 may decrease the isomerization constant (K_{iso}) (Appleman *et al.*, 1988), altering the dynamic process required for tight MTX binding. This isomerization effect has never been reported in the binding of DHF or folate; amino acid substitutions causing perturbation of K_{iso} would not affect the binding of these ligands as much as it would MTX, consistent with the greater decreases in MTX affinity observed for variant F31R/Q35E. The F31R substitution, which results in multiple distinct conformers, may perturb K_{iso} , thus reducing MTX binding. Biophysical data will be required to confirm the disruption of a putative isomerization step in MTX binding in the F31R and F31R/Q35E variants.

Appendix B: Novel crystallization conditions for tandem variant R67 DHFR yield a wild-type crystal structure

B.1 Introduction

Dihydrofolate reductase (DHFR) is a highly conserved enzyme that catalyses the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using NADPH as the hydride-donating cofactor. This process is critical in the synthesis of key metabolites, including purines and thymidylate. The central importance of DHFR in these biosynthetic processes has led to its use as a drug target for the treatment of a variety of conditions, including cancer and bacterial and protozoan infections (Volpato & Pelletier, 2009). Trimethoprim (TMP) has been shown to be an effective competitive inhibitor of the chromosomally encoded bacterial DHFRs. Despite the extensive similarity between bacterial and mammalian DHFRs (Margosiak *et al.*, 1993), TMP displays a much lower affinity for mammalian DHFRs, making it an effective antibiotic. In response to the broad use of this compound as an antibiotic in humans and livestock, some bacteria have acquired a plasmid-encoded DHFR known as R67 DHFR (*dfrB* gene family), which is structurally unrelated to the chromosomally encoded DHFRs (*dfrA* gene family) (Pattishall *et al.*, 1977). This lack of similarity, combined with the loose structural resemblance between TMP and the substrate DHF, allows R67 DHFR to evade inhibition by TMP, thus providing resistance to this antibiotic.

B.1.1 Structural features of R67 DHFR

The structure of wild-type R67 DHFR has been solved numerous times (Matthews *et al.*, 1986; Narayana *et al.*, 1995; Narayana, 2006; Divya *et al.*, 2007; Krahn *et al.*, 2007). The enzyme is a homodimer of homodimers, with four identical protomers forming a toroidal structure possessing a central pore. The fold of each protomer resembles an SH3-like domain, the first 16 N-terminal residues of which are presumably unstructured (Narayana *et al.*, 1995). For structural studies, these residues are removed by limited proteolysis to enable crystallization (Narayana *et al.*, 1995; Narayana, 2006; Divya *et al.*, 2007; Krahn *et al.*, 2007). Importantly, the removal of these residues does not alter the activity or stability of the purified enzyme (Reece *et al.*, 1991). With respect to the

quaternary structure, the first dimerization interface is mediated by residues 26 and 45–47 and effectively expands a β -sheet to span the two protomers. The "dimer-of-dimers" interface is mediated by loop regions that link the β -strands, primarily through residues 60–65 (Narayana *et al.*, 1995). It should also be noted that only minimal contact exists between protomers that are "diagonal" to each other in the complete tetramer, which includes van der Waals interactions between residues 36 and 50.

The central pore has been identified as the active site, with the four protomers together being able to bind two ligands: two DHFs, two NADPHs or a combination of one DHF and one NADPH, with this last configuration representing the productive ternary complex (Bradrick *et al.*, 1996). The active-site residues are for the most part distinct from those residues that are important for stabilization of the tetramer. The residues involved in ligand binding and catalysis include residues 32, 35–36, 50, 64–70 and 72–73. Of these, only residues 64–65 and 67 play any role in stabilizing the quaternary structure (Narayana *et al.*, 1995).

B.1.2 Mutational analysis of key active-site residues

The key ligand-binding portion of the pore is defined by residues 66–69 (sequence VQIY), which are responsible for binding both the pteroyl moiety of DHF and the nicotinamide ribose of NADPH (Strader *et al.*, 2001; Schmitzer *et al.*, 2004; Krahn *et al.*, 2007). As the binding of a single molecule of the substrate and the cofactor breaks the symmetry of the pore, this implies that each of these four residues can in fact play up to four different roles, depending on where they sit relative to the substrate and cofactor at any given time (Schmitzer *et al.*, 2004). This also implies that the mutation of any of these residues would cause more convoluted perturbations in the binding and catalytic properties of the enzyme compared with typical structure–function studies.

In order to better examine the roles of individual active-site residues, we prepared a construct of R67 DHFR that reduced the symmetry in the active site. This variant of the enzyme consisted of two protomers connected by a short linker and allowed mutational manipulation of only two, instead of all four, of the

protomers within the homotetramer. It is conceptually similar to the tandem dimer of Zhuang *et al.* (1993), but carries one mutated monomer. Specifically, for this study, residues 66–69 in the N-terminal protomer were mutated from VQIY to INSF (V66I/Q67N/I68S/Y69F). This quadruply mutated construct is of interest since the two first substitutions (V66I/Q67N) are included in a previously characterized monomeric triple variant V66I/Q67N/I68R, which had a catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}^{\text{DHF}}$) threefold greater than the wild type (Schmitzer *et al.*, 2004), while the point substituent Y69F had a catalytic efficiency fivefold lower than the wild type (Stinnett *et al.*, 2004). In the following, we will refer to this construct as the INSF tandem variant; as it resulted from a functional selection, it is necessarily catalytically active (Schmitzer *et al.*, 2004). We therefore characterized its kinetic properties and, to complement enzymatic studies, we pursued structure determination of this tandem variant. For this, we modified the most common crystallization method for R67 DHFR (Narayana *et al.*, 1995; Narayana, 2006; Divya *et al.*, 2007; Krahn *et al.*, 2007) by incorporating *in situ* proteolysis to simplify the crystallization process. Surprisingly, the resulting crystallized tetramer does not include the mutated protomers and instead appears to have reassembled in an all-wild-type form in the crystal structure.

B.2 Materials and Methods

B.2.1 Creation of the R67 DHFR INSF tandem variant construct and protein purification

The INSF tandem dimer was functionally selected from a library of tandem R67 DHFR variants. The detailed procedure for library creation will be described elsewhere. Briefly, the DNA sequence encoding the first protomer was combinatorially mutated at active site residues 66–69 (Schmitzer *et al.*, 2004) while keeping the second protomer native. The two protomers were linked by a Glu-Leu dipeptide, similar to a previous report (Zhuang *et al.*, 1993). The tandem dimer was preceded by an N-terminal hexahistidine tag in a pQE-32-derived vector (Qiagen). DNA sequencing was used to confirm the successful generation of the INSF tandem dimer construct (DNA-sequencing platform at IRIC, Montreal). Clones that were active and TMP-resistant were selected from the

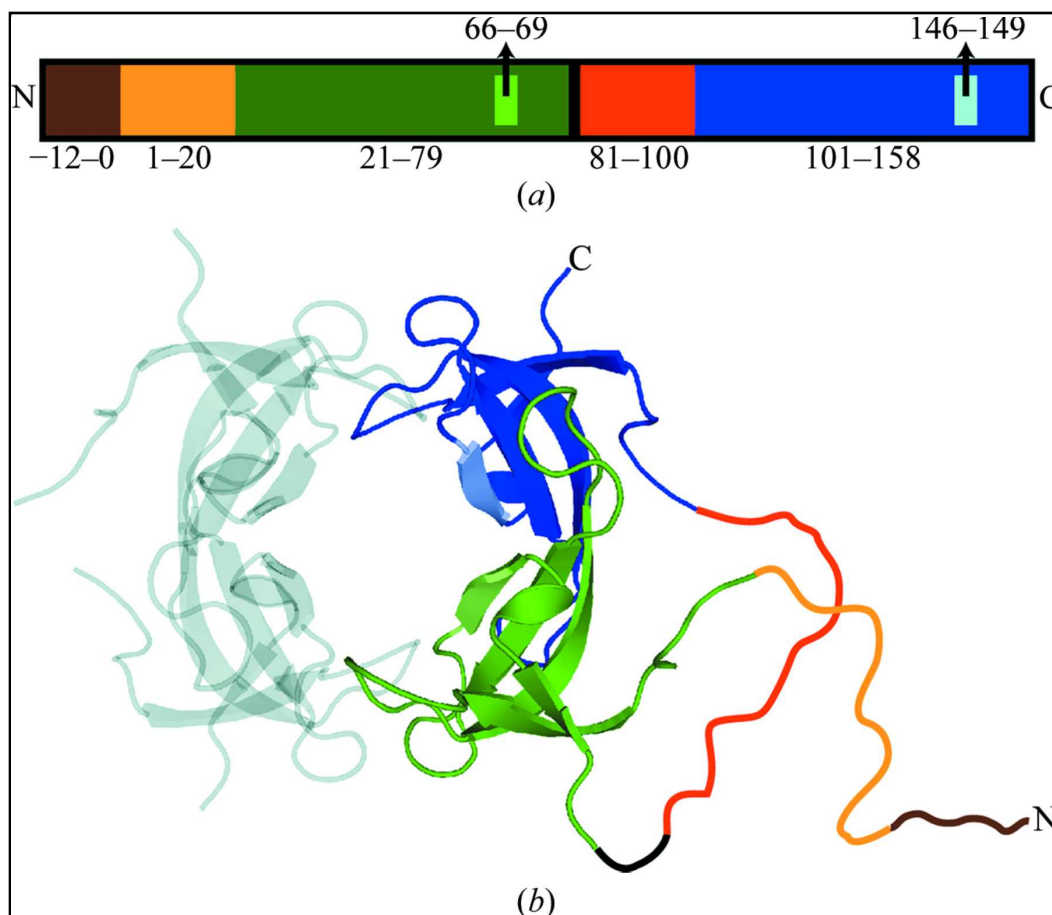


Fig. B.1: Organization of the INSF tandem dimer gene. (A) The different segments of the primary structure of the INSF tandem dimer are shown to scale. The colours represent the various segments of the gene. Brown (residues -12 to 0) represents the His tag, light orange (residues 1–20) and orange (residues 81–100) represent the N-terminal unstructured regions in the variant and wild-type protomers, respectively, green (residues 21–78) and blue (residues 101–158) represent the structured regions in the variant and wild-type protomers, respectively, and black (residues 79–80) represents the interprotomer linker. Lighter green (residues 66–69) and blue (residues 146–149) represent the mutation sites in the variant (INSF) and wild-type (VQIY) protomers, respectively. (B) The same colour scheme is used to show the proposed structure of the tandem dimer in the context of the assembled tetramer. The positions of the unstructured regions and linker (drawn in) are speculated.

resulting library, thus identifying the INSF tandem variant. The plasmid containing the INSF tandem variant (Fig. B.1) was transformed into *Escherichia coli* BL21 (pREP4) and subsequently used for protein expression using standard procedures. Purification of the protein employed a nickel-affinity column followed by size-exclusion chromatography in 100 mM Tris buffer pH 8.0. SDS-PAGE was used to assess the purity of the protein samples. The protein was

concentrated to 20 mg mL⁻¹ for crystallization. The molecular weight of the purified protein was confirmed to be 18 639 Da by mass spectroscopy, which agreed well with the predicted molecular weight of 18 638.6 Da.

B.2.2 Crystallization and data collection

Immediately prior to setting up crystal trays, chymotrypsin was added to the protein sample to a final ratio of 1:100 chymotrypsin:R67 DHFR by mass. In addition, the protein sample was diluted to 3/4 using MPD, thus reducing the protein concentration to 15 mg mL⁻¹ and resulting in a final MPD concentration of 25%. Crystals were obtained using the hanging-drop vapour-diffusion method in a Greiner 24-well hanging-drop crystallization plate. Reservoirs were prepared using 1 mL 100 mM sodium phosphate pH 7.2 and 55% MPD. On a siliconized glass cover slip (Hampton Research), 2.5 µL protein sample was mixed with 1.5 µL reservoir solution. The plate was incubated at 277 K and crystals were obtained within a week (Table B.1). Data were collected (Table B.2) under standard cryogenic conditions using a Rigaku RU-H3R generator equipped with an R-Axis IV⁺⁺ detector and Confocal Blue optics. The data were processed using the HKL-2000 suite of programs (Otwinowski & Minor, 1997).

Table B.1: Sample information.

Macromolecule details	
Database code(s)	PDB code 3SFM; UNP code dyr21 ecolx
Macromolecular assembly	R67 DHFR is a homotetramer. The other three parts of the biological assembly are generated by the following symmetry operations: (-y, -x, -z), (-x, -y, z) and (y, x, -z).
Mass of macromolecular assembly (Da)	26912 (method: mass spectrometry)
Source organism	<i>E. coli</i>
Crystallization and crystal data	
Crystallization method	Vapor diffusion, hanging drop
Temperature (K)	277
Crystallization solutions	
Macromolecule	2.5 µl 15 mg mL ⁻¹ R67 DHFR tandem dimer, 75 mM Tris pH 8, 25% (v/v) 2-methyl-2,4-pentanediol, 0.15 mg mL ⁻¹ chymotrypsin

Precipitant	1.5 μ l 100 mM sodium phosphate pH 7.2, 55% (v/v) 2-methyl-2,4- pentanediol
Reservoir	1 mL 100 mM sodium phosphate pH 7.2, 55% (v/v) 2-methyl-2,4- pentanediol
Crystal data	
Matthews coefficient V_M ($\text{\AA}^3 \text{ Da}^{-1}$)	2.35
Solvent content (%)	47.70
Unit-cell data	
Crystal system, space group	Tetragonal, $I4_122$
Unit-cell parameters (\AA , $^\circ$)	$a = b = 67.69$, $c = 51.76$, $\alpha = \beta = \gamma = 90$
No. of molecules in unit cell Z	16

Table B.2: Data-collection and structure-solution statistics. Values in parentheses are for the outer shell.

Diffraction source	Rigaku RU-H3R rotating anode
Diffraction protocol	Single wavelength
Monochromator	Rotating copper anode
Wavelength (\AA)	1.5418
Detector	Rigaku R-Axis IV ⁺⁺ image plate
Temperature (K)	93
Resolution range (\AA)	50-1.40 (1.43-1.40)
No. of unique reflections	12169 (802)
No. of observed reflections	12169
Completeness (%)	100 (100)
Multiplicity	23.8 (13.9)
$\langle I/\sigma(I) \rangle$	18.5
R_{merge}	0.061 (0.371)
Data-processing software	HKL-2000
Phasing method	Molecular replacement
Starting model data set	PDB entry 2RH2
Solution software	REFMAC

B.2.3 Structure solution and refinement

The integrated and scaled data were imported into the CCP4 suite of programs (Winn *et al.*, 2011). The isomorphous crystal structure of wild-type R67 DHFR (PDB entry 2RH2, Krahn *et al.*, 2007), which has one protomer in the asymmetric unit, was used to obtain phases for the new crystal structure. As the INSF tandem variant R67 DHFR crystallized in the same space group and with

the same unit-cell parameters as the wild-type model, care was taken to ensure that the same R_{free} test set was selected as was used in the 2RH2 structure. Refinement of the structure was performed using the program REFMAC (Table B.3) (Murshudov *et al.*, 1997). TLS refinement was introduced part way through the refinement and was ultimately switched to anisotropic B-factor refinement in the late stages of refinement. Manual model building was employed periodically using Coot (Emsley & Cowtan, 2004). Residues and solvent molecules that were revealed to have multiple positions in the electron-density maps were modelled accordingly. The finalized coordinates and structure factors have been deposited in the PDB (PDB entry 3SFM). All figures of protein structures were prepared using PyMOL (Schrödinger LLC).

Table B.3: Structure refinement and model validation. Values in parentheses are for the outer shell.

Refinement software	REFMAC v.5.5.0109
Refinement on	F
Resolution range (Å)	23.97–1.40 (1.435–1.40)
No. of reflections used in refinement	11533 (826)
No. of reflections above σ cutoff in final cycle	11533 (826)
Final overall R factor	0.146
Atomic displacement model	Anisotropic
Overall average B factor (Å ²)	17.8
No. of protein atoms	443
No. of ligand atoms	24
No. of solvent atoms	59
Total No. of atoms	526
Bulk-solvent model	Mask
Final R_{work}	0.144 (0.525)
No. of reflections for R_{free}	626 (42)
Final R_{free}	0.164 (0.565)
Ramachandran plot analysis	
Most favoured regions (%)	100
Additionally allowed regions (%)	0
Generously allowed regions (%)	0
Disallowed regions (%)	0

B.2.4 Analysis of the mutation sites

A priori, given the INSF tandem variant construct and the presence of a single protomer in the asymmetric unit, the density for residues 66–69 should reflect 50% occupancy for the wild-type sequence VQIY and 50% occupancy for the INSF sequence. In order to assess whether this was the case, a series of 11 ‘hybrid’ models were generated. These models included both the wild-type (VQIY) and variant (INSF) residues modelled as alternate conformers for residues 66–69. In the first model, the wild-type occupancies were set to 0 and those of the variant residues were set to 1. In each subsequent model the wild-type occupancies were increased by 0.1, while the variant occupancies were decreased by the same amount. For the variant residue Ser68, two alternate conformations for serine at half the occupancy of the other variant residues were used. An additional ten cycles of refinement in REFMAC were performed and the R-factors and the sizes of key peaks in $F_o - F_c$ maps were noted.

B.2.5 Analysis of protein samples before and after digestion

In order to better understand the results from the crystal structure, three analyses were performed: SDS–PAGE, an activity assay and mass-spectral analysis. A protein sample was prepared in an identical manner as for crystallization and the above three analyses were performed prior to the addition of chymotrypsin and then repeated on the same sample the day after addition of chymotrypsin. One week later, the SDS–PAGE and activity-assay analyses were repeated.

Enzyme activity was monitored in 50 mM phosphate buffer pH 7 in the presence of 100 mM each of NADPH and DHF. Activity was measured with a Cary100 Bio UV–Vis spectrophotometer (Varian Canada, Montréal, Quebec, Canada) by monitoring NADPH and DHF depletion ($\Delta\epsilon_{340} = 12\,800\text{ M}^{-1}\text{ cm}^{-1}$). In those cases where the protein had been digested with chymotrypsin prior to the activity assay, the protein was diluted in buffer and then re-concentrated to remove small peptide fragments.

Mass-spectral analysis was performed using an HPLC-MS system composed of an Agilent 1100 HPLC coupled with a TOF instrument equipped

with an electrospray source in positive mode. The chromatographic column was a Poroshell 300SB-C8 (2.1 × 75 mm, 5 µm particle size; Agilent Technologies). The eluents used were 0.1% formic acid in H₂O and 0.1% formic acid in acetonitrile.

B.3 Results and discussion

B.3.1 Protein purification

The INSF tandem dimer was purified as described and determined to be pure by SDS–PAGE. The oligomerization state of the protein under native conditions was determined to be the expected dimer of tandem dimers (i.e. a tetramer of the 8.4 kDa natural protomer) by gel filtration. The activity of the tandem dimer was determined to be 20% of the wild-type activity under conditions that are saturating for the wild type ($0.078 \pm 0.008 \text{ U mg}^{-1}$ for mutant INSF versus $0.45 \pm 0.05 \text{ U mg}^{-1}$ for the wild type). This strongly suggests that the INSF tandem dimer is correctly folded and assembles as a tetramer. We cannot preclude, however, that the quadruple mutation introduces some minor perturbations to the native fold of the protein.

B.3.2 Crystallization and structure determination

The conditions used to crystallize R67 DHFR represent a modification (and simplification) of the crystallization conditions previously used to crystallize R67 DHFR (Narayana *et al.*, 1995; Narayana, 2006; Divya *et al.*, 2007; Krahn *et al.*, 2007). Previously, purification of this enzyme for use in crystallization had incorporated a limited proteolysis step using chymotrypsin-linked beads. After removal of the beads, the enzyme was further purified to remove the proteolysed fragments. In this case, chymotrypsin was simply added to the crystallization experiment and allowed to function *in situ*, a technique that is increasingly being adopted by structural genomics groups (Dong *et al.*, 2007; Wernimont & Edwards, 2009). Crystals were produced easily using this simplified technique and the crystal quality did not appear to suffer, as demonstrated by the high resolution of the crystals obtained by this method. It should be noted that the composition of both the protein crystallization buffer and the reservoir conditions

were modified to make the protocol more amenable to commonly used automated crystallization protocols.

The crystals obtained diffracted to a resolution of 1.4 Å in space group I4₁22 with one protomer per asymmetric unit (crystallization, data-collection and refinement statistics are reported in Table B.1, Table B.2, and Table B.3). Upon examination of symmetry-related molecules, the characteristic toroidal structure of R67 DHFR could be observed. Residues 21–78 (numbering using equivalent wild-type residue numbers) could be modelled into the density, while residues 1–20 were not visible. Based on MS results, a portion of these residues were removed by proteolysis by chymotrypsin, while the rest are likely to be disordered in the crystal structure. The N-terminal His tag and linker which connected the two protomers in the modified gene were also not visible. As would be expected in a crystal structure at this resolution, six residues were observed in multiple conformations, as were some water molecules. Several larger peaks in the solvent were modelled as MPD.

As the crystals that were obtained were isomorphous to those used to obtain previously published structures (Matthews *et al.*, 1986; Narayana *et al.*, 1995; Narayana, 2006; Divya *et al.*, 2007; Krahn *et al.*, 2007), it is expected that the structure reported here would be very similar to these. The highest resolution structure of R67 DHFR that has been solved to date is an apo structure with a resolution of 0.96 Å (PDB entry 2RH2, Krahn *et al.*, 2007). As such, this structure was used as the basis for all further comparisons. Of the six residues that are present in dual conformations (Arg31, Trp45, Cys47, Asn49, Leu50 and Ser65), all six were also present in multiple conformations in 2RH2. However, there were seven residues (Lys32, Trp38, Gln41, Pro52, Pro63, Gln67 and Glu75) that were only modelled as having a single conformation even though two conformations were modelled in 2RH2. This variation is likely to be a consequence of the higher resolution of 2RH2 and does not represent a significant difference. A comparison of both proteins in their entirety revealed that the main chain essentially did not vary between the two structures and that most side-chain atoms occupied effectively identical positions in both structures. The only

residues to vary significantly from one structure to the other were the first and last residues in the chain, as well as some of the residues that were modelled in dual conformations. The RMSD for all main-chain atoms, excluding the first and last residues, is 0.08 Å and that for all atoms is 0.38 Å; however, if multiple conformations are omitted from the comparison the all-atom value drops to 0.08 Å. From this analysis, it is clear that the two structures are essentially identical to each other, with the side chains of those residues known to be somewhat flexible being the only sites to show some differences between the two structures.

B.3.3 Examination of the mutation sites in the crystal structure

The mutated residues are located in a well ordered part of the structure that is critical for binding both of the ligands. The crystallization of the protein in the presence of chymotrypsin led us to expect that the linked dimers would be cleaved apart. This would be likely to abolish the ordered architecture of the homotetramer over the timescale of crystallization. Irrespective of whether the architecture is affected, since the crystallization condition contained an exact 1:1 stoichiometric ratio of wild-type to variant protomers, there should be an equal probability of each type being incorporated into the growing crystal. The two sequences are sufficiently different to be easily distinguishable in the crystal structure at this resolution. It was therefore quite surprising that the series of models with varying occupancies of the wild-type and variant forms revealed that the best agreement with the diffraction data was achieved when the wild-type sequence was assigned an occupancy of 0.9–1.0 (Fig. B.2). This was confirmed both in reciprocal space using R-factors and in real space by assessing the σ level in F_o-F_c maps for distinguishing atoms in each of the four mutated residues (Fig. B.3). This implies that the variant protomer is selectively excluded from the crystal and that it is likely that not even one variant protomer is incorporated into the majority of functional R67 DHFR tetramers in the crystal structure.

B.3.4 Rationale for the variant protomers being excluded from the crystal

The observation of the variant protomers being excluded from the growing crystal may be the result of preferential degradation of the variant protomer and/or destabilization of the inter-protomer interactions in the variant relative to the

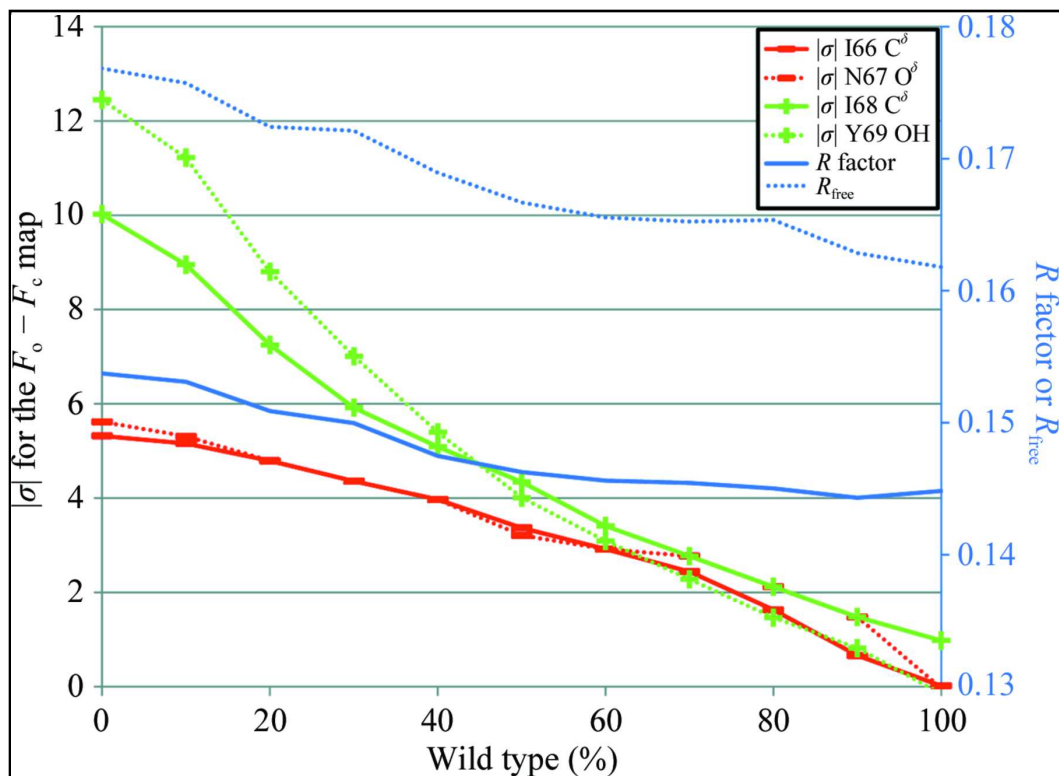
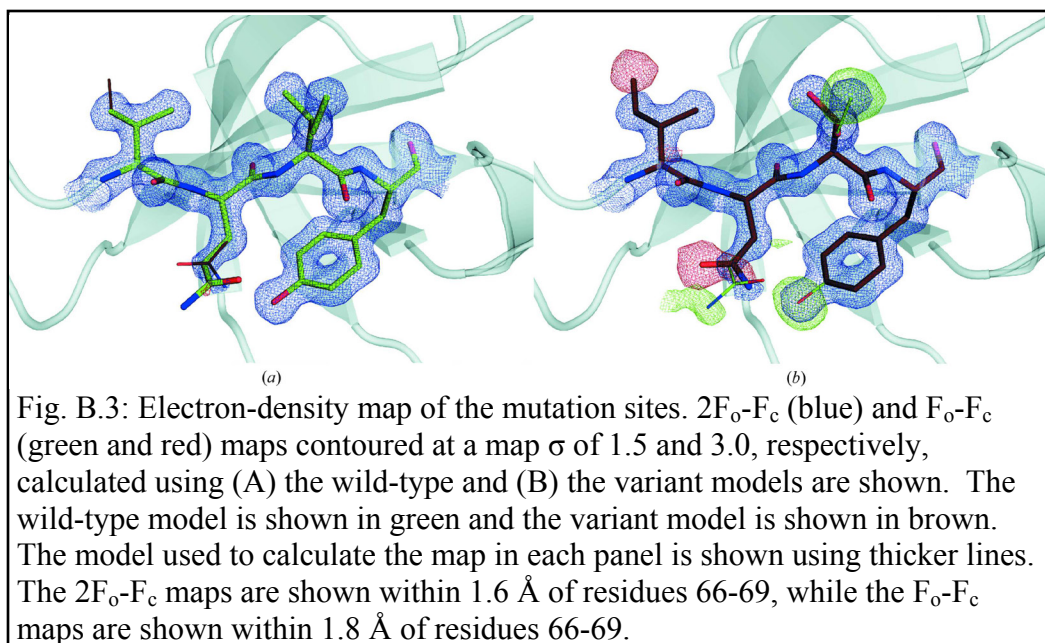


Fig. B.2: Change in magnitude of map σ and R factor with increasing wild-type character. The map σ value is evaluated at the position of one atom per residue that is unique to either the wild-type or the variant protomers. The map σ s (left axis) that are positive are shown in green and denoted with a + sign, while the map σ s that are negative are shown in red with a - sign. R factors (right axis) are shown in blue.

wild-type protomer. We have demonstrated that the variant protomer is degraded by chymotrypsin prior to its possible incorporation into the crystal (see below), despite the fact that no canonical chymotrypsin cleavage site is introduced by mutating VQIY to INSF. We carefully considered the possibility that destabilization of the inter-protomer interactions is what caused the variant protomers to be excluded. This would most likely have to be mediated by loss of the interaction of Asn67 with Asn67 in the neighbouring protomer, as this is the only one of the four mutated residues that forms an inter-protomer interaction. However, the results shown below clearly indicate that a single species the size of the monomer is present after chymotrypsin digestion, and molecular dynamics simulations failed to demonstrate any significant difference in interaction energy between the wild-type and variant forms.



When the INSF tandem dimer is exposed to chymotrypsin, as is required to remove the unstructured N-terminus for crystallization, the inter-protomer linker is efficiently cleaved under crystallization conditions. This is clearly demonstrated by the elimination of at least 90% of the variant protomers from the crystal. In order to further investigate this, the protein sample was analysed using SDS-PAGE, activity assays and mass spectrometry before and after chymotrypsin digestion. Following 24 h of exposure to chymotrypsin at 277 K, resolution on tricine-SDS-PAGE showed no trace of the dimer (18.6 kDa) and the appearance of a 6 kDa band (Fig. B.4). This corresponds roughly to the molecular weight expected for a single protomer having no additional residues (His tag, linker or termini). The 6 kDa band was estimated to contain approximately 25% of the protein initially digested, with some faint bands of lower molecular weight. This indicates full degradation of a fraction of the initial tandem dimer; a further fraction of at least one of the protomers remained intact, suggesting a fold that is sufficiently stable to prevent proteolysis. When these same samples were assayed for enzyme activity at 24 h, the chymotrypsin-digested tandem dimer revealed an unexpected increase in specific activity (in units per milligram of protein) from 20% of the wild-type specific activity for the undigested INSF tandem dimer to wild-type-like levels following chymotrypsin digestion. These data are consistent with the variant protomer being degraded more quickly than the wild-type

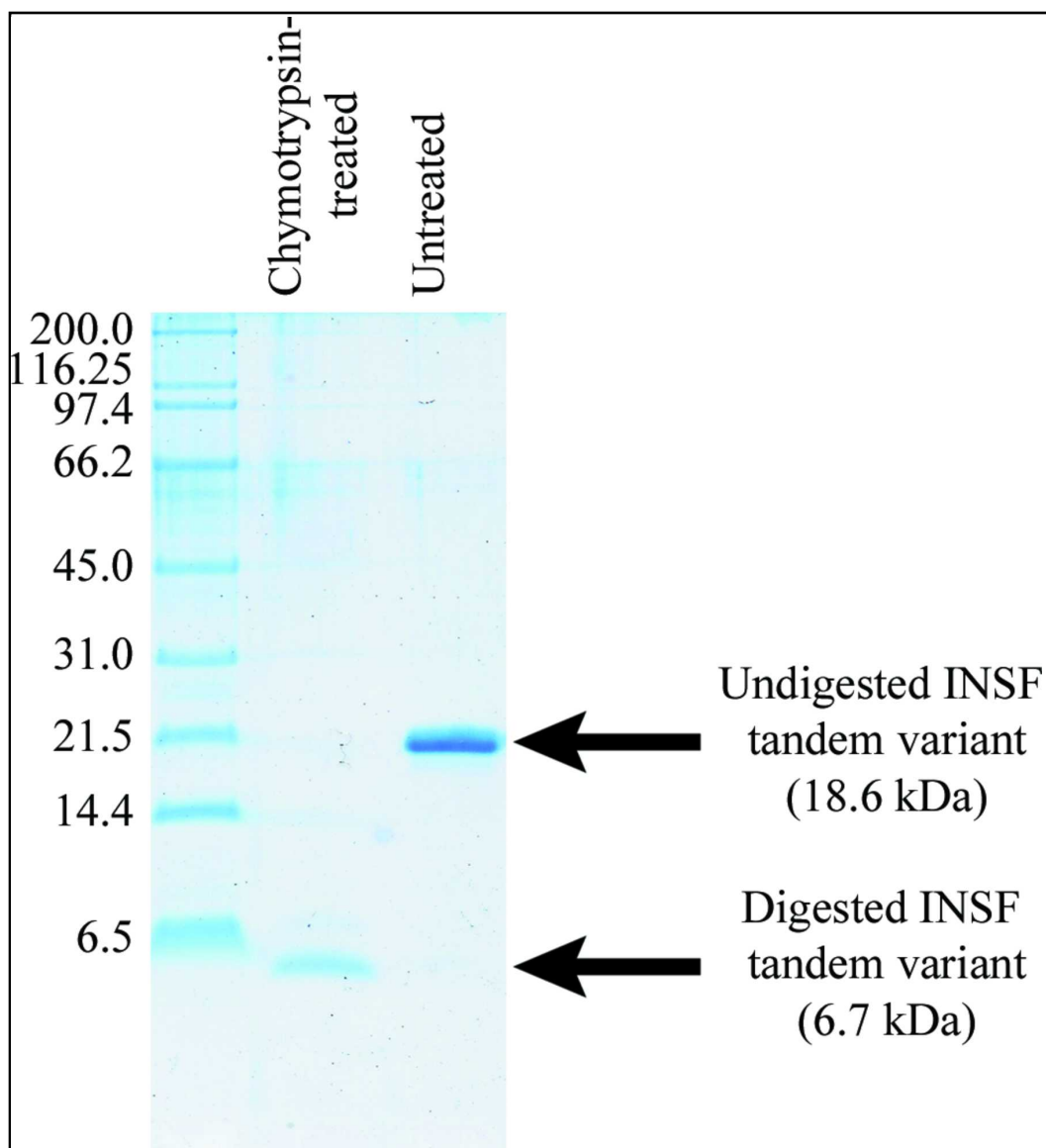


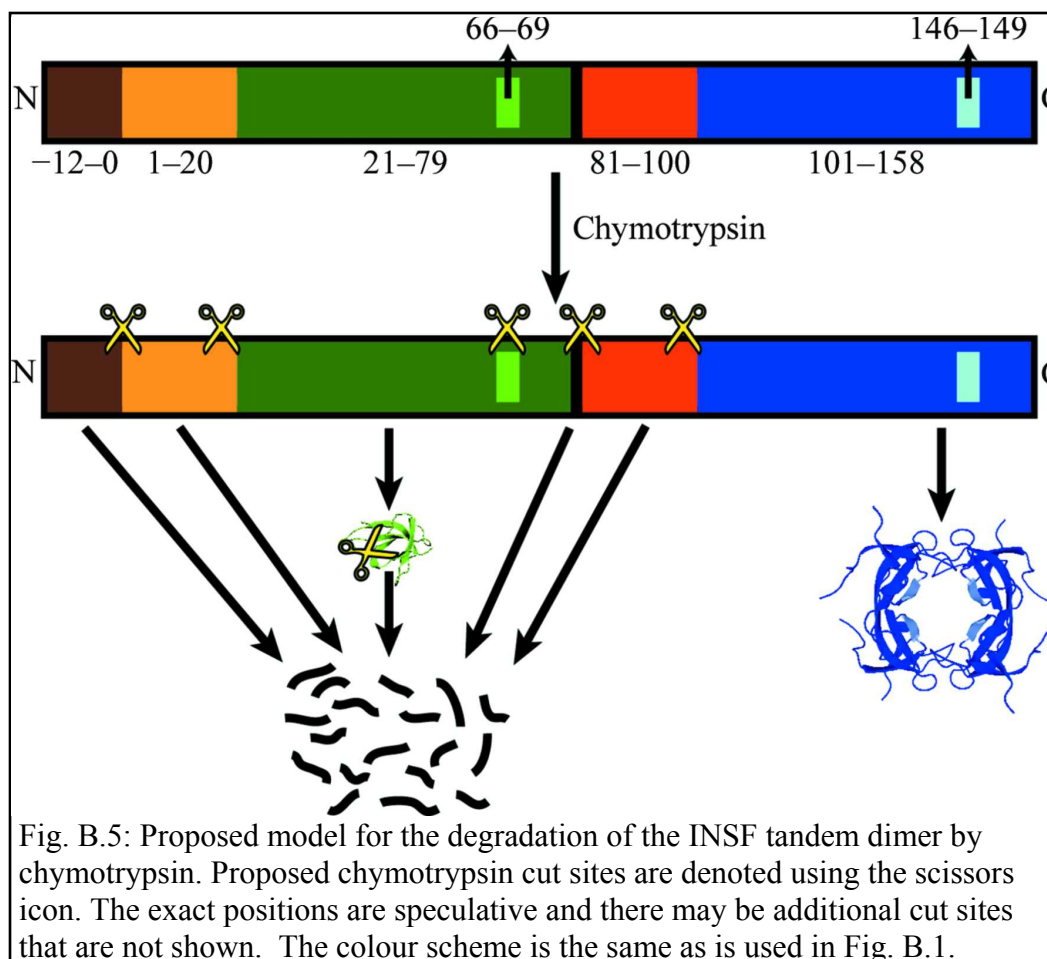
Fig. B.4: Comparison of the untreated and chymotrypsin-treated INSF tandem dimers on SDS-PAGE. The positions of the uncleaved (tandem dimer) and digested monomeric proteins are shown and the expected molecular weight as determined by mass spectroscopy is given.

protomer: as the variant protomers are degraded, the wild-type protomers are allowed to re-associate into a stable and fully wild-type tetramer. These analyses were repeated after one week and identical results were obtained, indicating that the sample was stable after 24 h.

Mass-spectral analysis of the same sample was performed 24 h following chymotrypsin treatment. For the digested INSF tandem dimer one predominant species at 6728 Da was observed (relative to 18 639 Da for the starting material),

suggesting that one of the two protomers is the dominant species. Furthermore, this corresponds well to the theoretical molecular weight of 6727.5 Da for residues 17–78 of the wild-type protomer. As residue 16 is a phenylalanine, this would indicate cleavage C-terminal to a phenylalanine, which is typical of chymotrypsin activity. This is consistent with the digestion of almost the entire N-terminal region, as only the last four residues (17–20) of this unstructured region remain. In contrast, the same span of residues for the mutant protomer would have a theoretical molecular weight of 6685.4 Da. However, it is interesting to note that neither the wild-type VQIY sequence nor the variant INSF sequence contain a "canonical" chymotrypsin cleavage site. Chymotrypsin typically cleaves C-terminal to Trp, Tyr or Phe, but not if the residue C-terminal to the aromatic residue is a Pro. The first residue after the Tyr69 or Phe69 in these sequences is Pro70, eliminating this site as a good chymotrypsin cleavage site. We are conducting more detailed analysis of the degradation in the context of this and other tandem dimers of R67 DHFR by modifying the cleavage conditions (time and the ratio of target to chymotrypsin) and analyzing the resulting fragments.

In addition to the apparent degradation of the variant protomer by chymotrypsin, the mutations may also destabilize the structure of the variant protomers themselves or impair the inter-protomer interactions. The latter could only be caused directly through the loss of the Gln67 interaction, although this seems unlikely given the lack of evidence from previous studies or from molecular-dynamics simulations. It is also possible that the mutations introduce disorder throughout the protein, as has recently been reported for a human DHFR variant (Volpato *et al.*, 2009). This disordering could be sufficient to allow chymotrypsin to more easily degrade the variant protomers. We can also speculate that if the linker and N-terminus is cleaved prior to the degradation of the structured portion of the protein, then the mixture of wild-type and variant subunits in each complete tetramer would be allowed to reach an equilibrium. If the variant was excluded from the tetramer, either because of decreased stability or the loss of a direct interaction, it would follow that it would be less structured



and more prone to proteolytic degradation. This could then lead to the rapid degradation of the monomeric INSF variant protomers prior to formation of the crystal (Fig. B.5). This could explain the degradation of the variant even without a canonical chymotrypsin cleavage site.

B.3.5 Implications for structural and functional studies of wild-type and heterodimeric R67 DHFR variants

The crystallization protocol presented here provides a simplified route to rapidly and readily obtain high-quality crystals of R67 DHFR which will be useful for further studies of this enzyme aimed at addressing the issue of resistance to trimethoprim. However, the method does highlight the difficulties in obtaining the structure of a tandem variant of R67 DHFR. To begin with, the fact that *in situ* proteolysis appears to cause the inter-protomer linker and the INSF variant protomer to be digested on a comparable timescale will be likely to make it impossible to use *in situ* proteolysis to crystallize the tandem variant. In

addition, the requirement to have the N-terminal unstructured region within R67 DHFR removed in order to obtain crystals jeopardizes the structure elucidation of an ordered tetramer from a tandem variant without the use of proteolysis. It may be necessary to prepare a tandem variant that contains specific protease cleavage sites engineered in locations at the N- and C-termini of the SH3-like domain, eliminating the use of a nonspecific protease such as chymotrypsin. Alternatively, a tandem variant in which the first unstructured N-terminal sequence is made to be cleavable and the second has been eliminated in favour of a linker of appropriate length between the two SH3-like domains might prove effective. This would eliminate the need for nonspecific proteases and maintain both the N-terminal region that allows the stable expression of the protein in *E. coli* (Reece *et al.*, 1991) and the linker that enforces the ordered tandem dimer in the crystal structure.

B.3.6 Implications for the use of *in situ* proteolysis as a crystallization method

With the rise of structural genomics consortia, *in situ* proteolysis is gaining popularity as a means to crystallize proteins that initially yield few or no crystals. The crystallization of R67 DHFR can be seen as one of the earliest examples of the use of this technique to obtain high-quality crystals. The present work demonstrates that while powerful, this technique can have many unforeseen consequences. While it was not surprising that the linkers were cleaved, it would have been very difficult to predict the destruction of the variant protomers of a protein that had previously been shown to be resistant to proteolytic degradation. This is especially true given that there is no canonical chymotrypsin cleavage site that is being introduced into the variant protomers, making a simple explanation of the degradation of the variant protomers difficult to make. While it is known that proteases can cleave proteins at non-canonical sequences, it is difficult to predict these additional cut sites prior to performing an *in situ* proteolysis experiment, especially when the susceptibility of the protein to proteolysis based on structural accessibility is taken into account. This demonstrates that our understanding of the effects of mutations on the global stability of protein structure is still lacking. Overall, we can see here that adapting a previously

obtained *in situ* proteolysis protocol to a slightly different system, be it a mutant, a homologue or a ligand-bound structure, can be more difficult than one might initially expect.

LIST OF REFERENCES

- Adger BM, McCague R, Roberts SM. (2000) Process for the Production of Lipoic Acid. *US6140512 A*. Asta Medica Aktiengesellschaft.
- Allay JA, Persons DA, Galipeau J, Riberdy JM, Ashmun RA, Blakley RL, Sorrentino BP (1998) *In vivo* selection of retrovirally transduced hematopoietic stem cells. *Nat. Med.* **4**: 1136-1143
- Alphand V, Wohlgemuth R (2010) Applications of Baeyer-Villiger Monooxygenases in Organic Synthesis. *Curr. Org. Chem.* **14**: 1928-1965
- Anastas PT, Warner JC (1998) *Green Chemistry: Theory and Practice*, New York: Oxford University Press.
- Appleman JR, Prendergast N, Delcamp TJ, Freisheim JH, Blakley RL (1988) Kinetics of the formation and isomerization of methotrexate complexes of recombinant human dihydrofolate reductase. *J. Biol. Chem.* **263**: 10304-10313
- Baeyer A, Villiger V (1899) Einwirkung des Caro'schen Reagens auf Ketone. *Ber. Dtsch. Chem. Ges.* **32**: 3625-3633
- Baeyer A, Villiger V (1900) Ueber die Einwirkung des Caro'schen Reagens auf Ketone. *Ber. Dtsch. Chem. Ges.* **33**: 858-864
- Baldwin CVF, Wohlgemuth R, Woodley JM (2008) The First 200-L Scale Asymmetric Baeyer-Villiger Oxidation Using a Whole-Cell Biocatalyst. *Org. Process Res. Dev.* **12**: 660-665
- Beam MP, Bosserman MA, Noinaj N, Wehenkel M, Rohr J (2009) Crystal Structure of Baeyer-Villiger Monooxygenase MtmOIV, the Key Enzyme of the Mithramycin Biosynthetic Pathway. *Biochemistry* **48**: 4476-4487
- Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**: 5656-5664
- Bieri M, Gooley PR (2011) Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics* **12**: 421
- Blakley RL, Sorrentino BP (1998) In vitro mutations in dihydrofolate reductase that confer resistance to methotrexate: Potential for clinical application. *Hum. Mutat.* **11**: 259-263

- Bocola M, Schulz F, Leca F, Vogel A, Fraaije MW, Reetz MT (2005) Converting Phenylacetone Monooxygenase into Phenylcyclohexanone Monooxygenase by Rational Design: Towards Practical Baeyer-Villiger Monooxygenases. *Adv. Synth. Catal.* **347**: 979-986
- Bong Y, Clay M, Collier S, Mijts B, Vogel M, Zhang X, Zhu J, Nazor J, Smith D, Song S. (2013) Synthesis of Prazole Compounds. *US20130017580 A1*. Codexis, Inc.
- Bonsor D, Butz SF, Solomons J, Grant S, Fairlamb IJS, Fogg MJ, Grogan G (2006) Ligation independent cloning (LIC) as a rapid route to families of recombinant biocatalysts from sequenced prokaryotic genomes. *Org. Biomol. Chem.* **4**: 1252-1260
- Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K (2012) Engineering the third wave of biocatalysis. *Nature* **485**: 185-194
- Bradrick TD, Beechem JM, Howell EE (1996) Unusual Binding Stoichiometries and Cooperativity Are Observed during Binary and Ternary Complex Formation in the Single Active Pore of R67 Dihydrofolate Reductase, a D2 Symmetric Protein. *Biochemistry* **35**: 11414-11424
- Branchaud BP, Walsh CT (1985) Functional Group Diversity in Enzymatic Oxygenation Reactions Catalyzed by Bacterial Flavin-Containing Cyclohexanone Oxygenase. *J. Am. Chem. Soc.* **107**: 2153-2161
- Brzostowicz PC, Blasko MS, Rouvière PE (2002) Identification of two gene clusters involved in cyclohexanone oxidation in *Brevibacterium epidermidis* strain HCU. *Appl. Microbiol. Biotechnol.* **58**: 781-789
- Brzostowicz PC, Walters DM, Jackson RE, Halsey KH, Ni H, Rouvière PE (2005) Proposed involvement of a soluble methane monooxygenase homologue in the cyclohexane-dependent growth of a new *Brachymonas* species. *Environ. Microbiol.* **7**: 179-190
- Brzostowicz PC, Walters DM, Thomas SM, Nagarajan V, Rouviere PE (2003) mRNA Differential Display in a Microbial Enrichment Culture: Simultaneous Identification of Three Cyclohexanone Monooxygenases from Three Species. *Appl. Environ. Microbiol.* **69**: 334-342
- Camporeale M, Fiorani T, Troisi L, Adam W, Curci R, Edwards JO (1990) On the mechanism of the Baeyer-Villiger oxidation of ketones by bis(trimethylsilyl) peroxomonosulfate. Intermediacy of dioxiranes. *J. Org. Chem.* **55**: 93-98
- Caro H (1898) Zur Kenntniss der Oxydation aromatischer Amine. *Angew. Chem.* **11**: 845-846

- Carver J, Richards R (1972) General 2-site solution for chemical exchange produced dependence of T2 upon Carr-Purcell pulse separation. *J. Magn. Reson.* **6**: 89-96
- Cavalieri D, McGovern PE, Hartl DL, Mortimer R, Polsinelli M (2003) Evidence for *S. cerevisiae* fermentation in ancient wine. *J. Mol. Evol.* **57**: S226-232
- Cernuchova P, Mihovilovic MD (2007) Microbial Baeyer-Villiger oxidation of terpenones by recombinant whole-cell biocatalysts-formation of enantiocomplementary regioisomeric lactones. *Org. Biomol. Chem.* **5**: 1715-1719
- Charlwood P (1957) Partial Specific Volumes of Proteins in Relation to Composition and Environment. *J. Am. Chem. Soc.* **79**: 776-781
- Cheesman MJ, Kneller MB, Rettie AE (2003) Critical role of histidine residues in cyclohexanone monooxygenase expression, cofactor binding and catalysis. *Chem.-Biol. Interact.* **146**: 157-164
- Chen C-S, Fujimoto Y, Girdaukas G, Sih CJ (1982) Quantitative analyses of biochemical kinetic resolutions of enantiomers. *J. Am. Chem. Soc.* **104**: 7294-7299
- Choi J-H, Kim T-K, Kim Y-M, Kim W-C, Park K, Rhee I-K (2006) Cloning and characterization of a gene cluster for cyclohexanone oxidation in *Rhodococcus* sp. TK6. *J. Microbiol. Biotechnol.* **16**: 511-518
- Chunduru SK, Cody V, Luft JR, Pangborn W, Appleman JR, Blakley RL (1994) Methotrexate-resistant variants of human dihydrofolate reductase. Effects of Phe31 substitutions. *J. Biol. Chem.* **269**: 9547-9555
- Clouthier CM, Kayser MM, Reetz MT (2006) Designing New Baeyer-Villiger Monooxygenases Using Restricted CASTing. *J. Org. Chem.* **71**: 8431-8437
- Cody V, Galitsky N, Luft JR, Pangborn W, Blakley RL, Gangjee A (1998) Comparison of ternary crystal complexes of F31 variants of human dihydrofolate reductase with NADPH and a classical antitumor fuopyrimidine. *Anti-Cancer Drug Des.* **13**: 307-315
- Cody V, Galitsky N, Luft JR, Pangborn W, Gangjee A (2003a) Analysis of two polymorphic forms of a pyrido[2,3-*d*]pyrimidine N9-C10 reversed-bridge antifolate binary complex with human dihydrofolate reductase. *Acta Crystallogr. D* **D59**: 654-661
- Cody V, Galitsky N, Luft JR, Pangborn W, Rosowsky a, Blakley RL (1997) Comparison of two independent crystal structures of human dihydrofolate reductase ternary complexes reduced with nicotinamide adenine dinucleotide phosphate and the very tight-binding inhibitor PT523. *Biochemistry* **36**: 13897-13903

- Cody V, Luft JR, Pangborn W (2005) Understanding the role of Leu22 variants in methotrexate resistance: comparison of wild-type and Leu22Arg variant mouse and human dihydrofolate reductase ternary crystal complexes with methotrexate and NADPH. *Acta Crystallogr. D* **D61**: 147-155
- Cody V, Luft JR, Pangborn W, Gangjee A (2003b) Analysis of three crystal structure determinations of a 5-methyl-6-*N*-methylanilino pyridopyrimidine antifolate complex with human dihydrofolate reductase. *Acta Crystallogr. D* **D59**: 1603-1609
- Cody V, Luft JR, Pangborn W, Gangjee A, Queener SF (2004) Structure determination of tetrahydroquinazoline antifolates in complex with human and *Pneumocystis carinii* dihydrofolate reductase: correlations between enzyme selectivity and stereochemistry. *Acta Crystallogr. D* **D60**: 646-655
- Cody V, Pace J, Chisum K, Rosowsky A (2006) New insights into DHFR interactions: Analysis of *Pneumocystis carinii* and mouse DHFR complexes with NADPH and two highly potent 5-(omega-carboxy(alkyloxy) trimethoprim derivatives reveals conformational correlations with activity and novel parallel ring stacking interactions. *Proteins Struct. Funct. Bioinf.* **65**: 959-969
- Criegee R (1948) Die Umlagerung der Dekalin-peroxydester als Folge von kationischem Sauerstoff. *Liebigs Ann. Chem.* **560**: 127-135
- Davies JF, Delcamp TJ, Prendergast NJ, Ashford Va, Freisheim JH, Kraut J (1990) Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafoate. *Biochemistry* **29**: 9467-9479
- Daw NC, Billups CA, Rodriguez-Galindo C, McCarville MB, Rao BN, Cain AM, Jenkins JJ, Neel MD, Meyer WH (2006) Metastatic osteosarcoma - Results of two consecutive therapeutic trials at St. Jude Children's Research Hospital. *Cancer* **106**: 403-412
- de Gonzalo G, Mihovilovic MD, Fraaije MW (2010) Recent Developments in the Application of Baeyer-Villiger Monooxygenases as Biocatalysts. *ChemBioChem* **11**: 2208-2231
- Divya N, Griffith E, Narayana N (2007) Structure of the Q67H mutant of R67 dihydrofolate reductase-NADP⁺ complex reveals a novel cofactor binding mode. *Protein Sci.* **16**: 1063-1068
- Doering WvE, Dorfman E (1953) Mechanism of the Peracid Ketone—Ester Conversion. Analysis of Organic Compounds for Oxygen-18. *J. Am. Chem. Soc.* **75**: 5595-5598
- Doering WvE, Speers L (1950) The Peracetic Acid Cleavage of Unsymmetrical Ketones. *J. Am. Chem. Soc.* **72**: 5515-5518

- Dong A, Xu X, Edwards AM (2007) *In situ* proteolysis for protein crystallization and structure determination. *Nat. Methods.* **4**: 1019-1021
- Donoghue NA, Norris DB, Trudgill PW (1976) The Purification and Properties of Cyclohexanone Oxygenase from *Nocardia globerula* CL1 and *Acinetobacter* NCIB 9871. *Eur. J. Biochem.* **63**: 175-192
- Donoghue NA, Trudgill PW (1975) The Metabolism of Cyclohexanol by *Acinetobacter* NCIB 9871. *Eur. J. Biochem.* **60**: 1-7
- Doucet N, Pelletier JN (2009) Gaining Insight into Enzyme Function through Correlation with Protein Motions. In *Protein Engineering Handbook*, Lutz S, Bornscheuer UT (eds), Vol. 1, 7, pp 187-211. Weinheim, Germany: Wiley-VCH
- Doucet N, Savard P-Y, Pelletier JN, Gagne SM (2007) NMR investigation of Tyr(105) mutants in TEM-1 beta-lactamase - Dynamics are correlated with function. *J. Biol. Chem.* **282**: 21448-21459
- Dudek HM, de Gonzalo G, Torres Pazmiño DE, Stepniak P, Wyrwicz LS, Rychlewski L, Fraaije MW (2011) Mapping the Substrate Binding Site of Phenylacetone Monooxygenase from *Thermobifida fusca* by Mutational Analysis. *Appl. Environ. Microbiol.* **77**: 5730-5738
- Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skaliky JJ, Kay LE, Kern D (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**: 117-121
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **D60**: 2126-2132
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr. D* **D66**: 486-501
- Enroth C, Neujahr H, Schneider G, Lindqvist Y (1998) The crystal structure of phenol hydroxylase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis. *Structure* **6**: 605-617
- Entsch B, Cole LJ, Ballou DP (2005) Protein dynamics and electrostatics in the function of *p*-hydroxybenzoate hydroxylase. *Arch. Biochem. Biophys.* **433**: 297-311
- Ercikan-Abali EA, Mineishi S, Tong YZ, Nakahara S, Waltham MC, Banerjee D, Chen W, Sadelain M, Bertino JR (1996a) Active site-directed double mutants of dihydrofolate reductase. *Cancer Res.* **56**: 4142-4145

- Ercikan-Abali EA, Waltham MC, Dicker AP, Schweitzer BI, Gritsman H, Banerjee D, Bertino JR (1996b) Variants of human dihydrofolate reductase with substitutions at Leucine-22: Effect on catalytic and inhibitor binding properties. *Mol. Pharmacol.* **49**: 430-437
- Fantin G, Giovannini PP, Guerrini A, Maietti S, Medici A, Pedrini P (2006) Enantioselective Baeyer-Villiger oxidation of bicyclo[3.2.0]hept-2-en-6-One with fungi: optimization of biotransformation and use of TiO₂ as support of cell growth. *Biotechnol. Lett.* **28**: 805-810
- Fernandez-Patron C, Castellanos-Serra L, Rodriguez P (1992) Reverse Staining of Sodium Dodecyl-Sulfate Polyacrylamide Gels by Imidazole-Zinc Salts - Sensitive Detection of Unmodified Proteins. *Biotechniques* **12**: 564-&
- Flasshove M, Banerjee D, Mineishi S, Li MX, Bertino JR, Moore MAS (1995) *Ex vivo* Expansion and Selection of Human CD34(+) Peripheral Blood Progenitor Cells After Introduction of a Mutated Dihydrofolate Reductase cDNA via Retroviral Gene Transfer. *Blood* **85**: 566-574
- Flintoff WF, Sadlish H, Gorlick R, Yang R, Williams FMR (2004) Functional analysis of altered reduced folate carrier sequence changes identified in osteosarcomas. *BBA - Mol. Basis Dis.* **1690**: 110-117
- Fossati E, Volpato JP, Poulin L, Guerrero V, Dugas D-A, Pelletier JN (2008) 2-tier bacterial and *in vitro* selection of active and methotrexate-resistant variants of human dihydrofolate reductase. *J. Biomol. Screen.* **13**: 504-514
- Fraaije MW, Kamerbeek NM, Heidekamp AJ, Fortin R, Janssen DB (2004) The prodrug activator EtaA from *Mycobacterium tuberculosis* is a Baeyer-Villiger monooxygenase. *J. Biol. Chem.* **279**: 3354-3360
- Fraaije MW, Kamerbeek NM, van Berkel WJH, Janssen DB (2002) Identification of a Baeyer-Villiger monooxygenase sequence motif. *FEBS Lett.* **518**: 43-47
- Fraaije MW, Wu J, Heuts DPHM, van Hellemond EW, Spelberg JHL, Janssen DB (2005) Discovery of a thermostable Baeyer-Villiger monooxygenase by genome mining. *Appl. Microbiol. Biotechnol.* **66**: 393-400
- Fried J, Thoma RW, Klingsberg A (1953) Oxidation of Steroids by Micro Örganisms. III. Side Chain Degradation, Ring D-Cleavage and Dehydrogenation in Ring A. *J. Am. Chem. Soc.* **75**: 5764-5765
- Gallagher TF, Kritchevsky TH (1950) Perbenzoic Acid Oxidation of 20-Ketosteroids and the Stereochemistry of C-17. *J. Am. Chem. Soc.* **72**: 882-885

- Galvão TC, Mohn WW, de Lorenzo V (2005) Exploring the microbial biodegradation and biotransformation gene pool. *Trends Biotechnol.* **23**: 497-506
- Gangjee A, Vidwans AP, Vasudevan A, Queener SF, Kisliuk RL, Cody V, Li RM, Galitsky N, Luft JR, Pangborn S (1998) Structure-based design and synthesis of lipophilic 2,4-diamino-6-substituted quinazolines and their evaluation as inhibitors of dihydrofolate reductases and potential antitumor agents. *J. Med. Chem.* **41**: 3426-3434
- Gatti DL, Palfey BA, Lah MS, Entsch B, Massey V, Ballou DP, Ludwig ML (1994) The Mobile Flavin of 4-OH Benzoate Hydroxylase. *Science* **266**: 110-114
- Geitner K, Rehdorf J, Snajdrova R, Bornscheuer UT (2010) Scale-up of Baeyer-Villiger monooxygenase-catalyzed synthesis of enantiopure compounds. *Appl. Microbiol. Biotechnol.* **88**: 1087-1093
- Griffin M, Trudgill PW (1976) Purification and properties of cyclopentanone oxygenase of *Pseudomonas* NCIB 9872. *Eur. J. Biochem.* **63**: 199-209
- Hawthorne MF, Emmons WD, McCallum KS (1958) A re-examination of the peroxyacid cleavage of ketones. I. Relative migratory aptitudes. *J. Am. Chem. Soc.* **80**: 6393-6398
- Hilker I, Alphand V, Wohlgemuth R, Furstoss R (2004a) Microbial Transformations, 56. Preparative Scale Asymmetric Baeyer–Villiger Oxidation using a Highly Productive “Two-in-One” Resin-Based *in situ* SFPR Concept. *Adv. Synth. Catal.* **346**: 203-214
- Hilker I, Gutiérrez MC, Alphand V, Wohlgemuth R, Furstoss R (2004b) Microbiological transformations 57. Facile and efficient resin-based *in situ* SFPR preparative-scale synthesis of an enantiopure "unexpected" lactone regioisomer via a Baeyer-Villiger oxidation process. *Org. Lett.* **6**: 1955-1958
- Hilker I, Gutierrez MC, Furstoss R, Ward J, Wohlgemuth R, Alphand V (2008) Preparative scale Baeyer-Villiger biooxidation at high concentration using recombinant *Escherichia coli* and *in situ* substrate feeding and product removal process. *Nat. Protoc.* **3**: 546-554
- Hilker I, Wohlgemuth R, Alphand V, Furstoss R (2005) Microbial transformations 59: first kilogram scale asymmetric microbial Baeyer-Villiger oxidation with optimized productivity using a resin-based *in situ* SFPR strategy. *Biotechnol. Bioeng.* **92**: 702-710

- Hucík M, Bucko M, Gemeiner P, Stefuca V, Vikartovská A, Mihovilovic MD, Rudroff F, Iqbal N, Chorvát D, Lacík I (2010) Encapsulation of recombinant *E. coli* expressing cyclopentanone monooxygenase in polyelectrolyte complex capsules for Baeyer-Villiger biooxidation of 8-oxabicyclo[3.2.1]oct-6-en-3-one. *Biotechnol. Lett.* **32**: 675-680
- Hura GL, Budworth H, Dyer KN, Rambo RP, Hammel M, McMurray CT, Tainer JA (2013) Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods.* **10**: 453-454
- Iwaki H, Wang S, Grosse S, Bergeron H, Nagahashi A, Lertvorachon J, Yang J, Konishi Y, Hasegawa Y, Lau PCK (2006) Pseudomonad Cyclopentadecanone Monooxygenase Displaying an Uncommon Spectrum of Baeyer-Villiger Oxidations of Cyclic Ketones. *Appl. Environ. Microbiol.* **72**: 2707-2720
- Jiang J, Tetzlaff CN, Takamatsu S, Iwatsuki M, Komatsu M, Ikeda H, Cane DE (2009) Genome mining in *Streptomyces avermitilis*: A biochemical Baeyer-Villiger reaction and discovery of a new branch of the pentalenolactone family tree. *Biochemistry* **48**: 6431-6440
- Kamerbeek NM, Fraaije MW, Janssen DB (2004) Identifying determinants of NADPH specificity in Baeyer-Villiger monooxygenases. *Eur. J. Biochem.* **271**: 2107-2116
- Kamerbeek NM, Moonen MJH, van der Ven JGM, van Berkel WJH, Fraaije MW, Janssen DB (2001) 4-Hydroxyacetophenone monooxygenase from *Pseudomonas fluorescens* ACB. *Eur. J. Biochem.* **268**: 2547-2557
- Kayser MM (2009) 'Designer reagents' recombinant microorganisms: new and powerful tools for organic synthesis. *Tetrahedron* **65**: 947-974
- Kayser MM, Clouthier CM (2006) New Bioorganic Reagents: Evolved Cyclohexanone Monooxygenase—Why Is It More Selective? *J. Org. Chem.* **71**: 8424-8430
- Kim Y-M, Jung S-H, Chung Y-H, Yu C-B, Rhee I-K (2008) Cloning and characterization of a cyclohexanone monooxygenase gene from *Arthrobacter* sp. L661. *Biotechnol. Bioproc. E.* **13**: 40-47
- Kirschner A, Altenbuchner J, Bornscheuer UT (2007) Design of a secondary alcohol degradation pathway from *Pseudomonas fluorescens* DSM 50106 in an engineered *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **75**: 1095-1101
- Kirschner A, Bornscheuer UT (2008) Directed evolution of a Baeyer-Villiger monooxygenase to enhance enantioselectivity. *Appl. Microbiol. Biotechnol.* **81**: 465-472

- Klon AE, Héroux A, Ross LJ, Pathak V, Johnson Ca, Piper JR, Borhani DW (2002) Atomic Structures of Human Dihydrofolate Reductase Complexed with NADPH and Two Lipophilic Antifolates at 1.09Å and 1.05Å Resolution. *J. Mol. Biol.* **320**: 677-693
- Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**: 1277-1282
- Kostichka K, Thomas SM, Gibson KJ, Nagarajan V, Cheng Q (2001) Cloning and characterization of a gene cluster for cyclododecanone oxidation in *Rhodococcus ruber* SC1. *J. Bacteriol.* **183**: 6478-6486
- Kotani T, Yurimoto H, Kato N, Sakai Y (2007) Novel Acetone Metabolism in a Propane-Utilizing Bacterium, *Gordonia* sp. Strain TY-5. *J. Bacteriol.* **189**: 886-893
- Krahn JM, Jackson MR, DeRose EF, Howell EE, London RE (2007) Crystal Structure of a Type II Dihydrofolate Reductase Catalytic Ternary Complex. *Biochemistry* **46**: 14878-14888
- Kries H, Blomberg R, Hilvert D (2013) *De novo* enzymes by computational design. *Curr. Opin. Chem. Biol.* **17**: 221-228
- Krow GR (1993) The Baeyer–Villiger Oxidation of Ketones and Aldehydes. In *Organic Reactions*, Paquette LA (ed), Vol. 43, 3, pp 251-798. New York, NY: John Wiley & Sons, Inc.
- Lau PCK, Leisch H, Yachnin BJ, Mirza IA, Berghuis AM, Iwaki H, Hasegawa Y (2010) Sustained Development in Baeyer-Villiger Biooxidation Technology. In *Green Polymer Chemistry: Biocatalysis and Biomaterials*, Cheng HN, Gross RA (eds), Vol. 1043, 24, pp 343-372. Washington, DC: American Chemical Society
- Leisch H, Morley K, Lau PCK (2011) Baeyer–Villiger Monooxygenases: More Than Just Green Chemistry. *Chem. Rev.* **111**: 4165-4222
- Leisch H, Shi R, Grosse S, Morley K, Bergeron H, Cygler M, Iwaki H, Hasegawa Y, Lau PCK (2012) Cloning, Baeyer-Villiger Biooxidations, and Structures of the Camphor Pathway 2-Oxo- Δ^3 -4,5,5-Trimethylcyclopentenylacetyl-Coenzyme A Monooxygenase of *Pseudomonas putida* ATCC 17453. *Appl. Environ. Microbiol.* **78**: 2200-2212
- Lewis WS, Cody V, Galitsky N, Luft JR, Pangborn W, Chunduru SK, Spencer HT, Appelman JR, Blakley RL (1995) Methotrexate-Resistant Variants of Human Dihydrofolate Reductase with Substitutions of Leucine-22 - Kinetics, Crystallography, and Potential as Selectable Markers. *J. Biol. Chem.* **270**: 5057-5064

- Luna A, Gutiérrez M-C, Furstoss R, Alphand V (2005) Microbial Baeyer–Villiger oxidation applied to the synthesis of the N-protected (1R,5R)-Geisman–Waiss lactone. *Tetrahedron - Asymmetr.* **16**: 2521-2524
- Lusty CJ (1999) A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. *J. Appl. Crystallogr.* **32**: 106-112
- Luz Z, Meiboom S (1963) Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution—Order of the Reaction with Respect to Solvent. *J. Chem. Phys.* **39**: 366-370
- Malito E, Alfieri A, Fraaije MW, Mattevi A (2004) Crystal structure of a Baeyer-Villiger monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 13157-13162
- Margosiak SA, Appleman JR, Santi DV, Blakley RL (1993) Dihydrofolate Reductase from the Pathogenic Fungus *Pneumocystis carinii*: Catalytic Properties and Interaction with Antifolates. *Arch. Biochem. Biophys.* **305**: 499-508
- Matthews DA, Smith SL, Baccanari DP, Burchall JJ, Oatley SJ, Kraut J (1986) Crystal structure of a novel trimethoprim-resistant dihydrofolate reductase specified in *Escherichia coli* by R-plasmid R67. *Biochemistry* **25**: 4194-4204
- Mayer M, Meyer B (2001) Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **123**: 6108-6117
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**: 658-674
- McTigue MA, Davies JF, Kaufman BT, Kraut J (1992) Crystal Structure of Chicken Liver Dihydrofolate Reductase Complexed with NADP⁺ and Biopterin. *Biochemistry* **31**: 7264-7273
- Mihovilovic MD, Rudroff F, Winninger A, Schneider T, Schulz F, Reetz MT (2006) Microbial Baeyer-Villiger Oxidation: Stereopreference and Substrate Acceptance of Cyclohexanone Monooxygenase Mutants Prepared by Directed Evolution. *Org. Lett.* **8**: 1221-1224
- Mirza IA, Yachnin BJ, Wang S, Grosse S, Bergeron H, Imura A, Iwaki H, Hasegawa Y, Lau PCK, Berghuis AM (2009) Crystal Structures of Cyclohexanone Monooxygenase Reveal Complex Domain Movements and a Sliding Cofactor. *J. Am. Chem. Soc.* **131**: 8848-8854
- Mislow K, Brenner J (1953) Stereospecificity of the Baeyer-Villiger Rearrangement. *J. Am. Chem. Soc.* **75**: 2318-2322

- Morii S, Sawamoto S, Yamauchi Y, Miyamoto M, Iwami M, Itagaki E (1999) Steroid monooxygenase of *Rhodococcus rhodochrous*: sequencing of the genomic DNA, and hyperexpression, purification, and characterization of the recombinant enzyme. *J. Biochem.* **126**: 624-631
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D* **D53**: 240-255
- Nakano T, Spencer HT, Appleman JR, Blakley RL (1994) Critical Role of Phenylalanine 34 of Human Dihydrofolate Reductase in Substrate and Inhibitor Binding and in Catalysis. *Biochemistry* **33**: 9945-9952
- Narayana N (2006) High-resolution structure of a plasmid-encoded dihydrofolate reductase: pentagonal network of water molecules in the D₂-symmetric active site. *Acta Crystallogr. D* **D62**: 695-706
- Narayana N, Matthews DA, Howell EE, Xuong N-h (1995) A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D₂-symmetric active site. *Nat. Struct. Biol.* **2**: 1018-1025
- Nobeli I, Favia AD, Thornton JM (2009) Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* **27**: 157-167
- Oefner C, Darcy A, Winkler FK (1988) Crystal Structure of Human Dihydrofolate Reductase Complexed with Folate. *Eur. J. Biochem.* **174**: 377-385
- Olucha J, Meneely KM, Chilton AS, Lamb AL (2011) Two Structures of an N-Hydroxylating Flavoprotein Monooxygenase. *J. Biol. Chem.* **286**: 31789-31798
- Onaca C, Kieninger M, Engesser K-H, Altenbuchner J (2007) Degradation of Alkyl Methyl Ketones by *Pseudomonas veronii* MEK700. *J. Bacteriol.* **189**: 3759-3767
- Opperman DJ, Reetz MT (2010) Towards practical Baeyer-Villiger-monooxygenases: design of cyclohexanone monooxygenase mutants with enhanced oxidative stability. *ChemBioChem* **11**: 2589-2596
- Orru R, Dudek HM, Martinoli C, Torres Pazmiño DE, Royant A, Weik M, Fraaije MW, Mattevi A (2011) Snapshots of Enzymatic Baeyer-Villiger Catalysis. *J. Biol. Chem.* **286**: 29284-29291
- Orru R, Pazmiño DET, Fraaije MW, Mattevi A (2010) Joint Functions of Protein Residues and NADP(H) in Oxygen Activation by Flavin-containing Monooxygenase. *J. Biol. Chem.* **285**: 35021-35028

- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* Vol. 276, pp 307-326. Academic Press
- Park J, Kim D, Kim S, Kim J, Bae K, Lee C (2007) The analysis and application of a recombinant monooxygenase library as a biocatalyst for the Baeyer-Villiger reaction. *J. Microbiol. Biotechn.* **17**: 1083-1089
- Patel M, Sleep SEH, Lewis WS, Spencer HT, Mareya SM, Sorrentino BP, Blakley RL (1997) Comparison of the protection of cells from antifolates by transduced human dihydrofolate reductase mutants. *Hum. Gene Ther.* **8**: 2069-2077
- Pattishall KH, Acar J, Burchall JJ, Goldstein FW, Harvey RJ (1977) Two distinct types of trimethoprim-resistant dihydrofolate reductase specified by R-plasmids of different compatibility groups. *J. Biol. Chem.* **252**: 2319-2323
- Polyak I, Reetz MT, Thiel W (2012) Quantum Mechanical/Molecular Mechanical Study on the Mechanism of the Enzymatic Baeyer–Villiger Reaction. *J. Am. Chem. Soc.* **134**: 2732-2741
- Polyak I, Reetz MT, Thiel W (2013) Quantum mechanical/molecular mechanical study on the enantioselectivity of the enzymatic baeyer-villiger reaction of 4-hydroxycyclohexanone. *J. Phys. Chem. B* **117**: 4993-5001
- Ramanan AV, Whitworth P, Baildam EM (2003) Use of methotrexate in juvenile idiopathic arthritis. *Arch. Dis. Child.* **88**: 197-200
- Rambo RP, Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**: 559-571
- Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**: 477-481
- Reece LJ, Nichols R, Ogden RC, Howell EE (1991) Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein. *Biochemistry* **30**: 10895-10904
- Reetz MT, Brunner B, Schneider T, Schulz F, Clouthier CM, Kayser MM (2004a) Directed Evolution as a Method To Create Enantioselective Cyclohexanone Monooxygenases for Catalysis in Baeyer–Villiger Reactions. *Angew. Chem. Int. Ed.* **43**: 4075-4078
- Reetz MT, Daligault F, Brunner B, Hinrichs H, Deege A (2004b) Directed Evolution of Cyclohexanone Monooxygenases: Enantioselective Biocatalysts for the Oxidation of Prochiral Thioethers. *Angew. Chem. Int. Ed.* **43**: 4078-4081

- Reetz MT, Wu S (2008) Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions. *Chem. Commun. (Camb.)*: 5499-5501
- Reetz MT, Wu S (2009) Laboratory Evolution of Robust and Enantioselective Baeyer–Villiger Monooxygenases for Asymmetric Catalysis. *J. Am. Chem. Soc.* **131**: 15424-15432
- Rehendorf J, Zimmer CL, Bornscheuer UT (2009) Cloning, expression, characterization, and biocatalytic investigation of the 4-hydroxyacetophenone monooxygenase from *Pseudomonas putida* JD1. *Appl. Environ. Microbiol.* **75**: 3106-3114
- Rial DV, Cernuchova P, van Beilen JB, Mihovilovic MD (2008) Biocatalyst assessment of recombinant whole-cells expressing the Baeyer-Villiger monooxygenase from *Xanthobacter* sp. ZL5. *J. Mol. Catal. B: Enzym.* **50**: 61-68
- Riva S, Fassi P, Allegrini P, Razzetti G. (2008) Process for the preparation of (–) modafinil. *US7316918 B2*. Dipharma S.P.A.
- Rozzell JD, Benner SA (1983) Combining enzymatic and chemical steps in the synthesis of biochemically valuable compounds: isotopically chiral methyl acetate. *J. Org. Chem.* **48**: 1190-1193
- Rudroff F, Alphand V, Furstoss R, Mihovilovic MD (2006) Optimizing fermentation conditions of recombinant *Escherichia coli* expressing cyclopentanone monooxygenase. *Org. Process Res. Dev.* **10**: 599-604
- Ryerson CC, Ballou DP, Walsh CT (1982) Mechanistic Studies on Cyclohexanone Oxygenase. *Biochemistry* **21**: 2644-2655
- Salem M, Mauguén Y, Prangé T (2010) Revisiting glutaraldehyde cross-linking: the case of the Arg-Lys intermolecular doublet. *Acta. Crystallogr. F* **F66**: 225-228
- Schmitzer AR, Lepine F, Pelletier JN (2004) Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations. *Protein Eng. Des. Sel.* **17**: 809-819
- Schreuder HA, Mattevi A, Obmolova G, Kalk KH, Hol WGJ, van der Bolt FJT, van Berkel WJH (1994) Crystal Structures of Wild-Type p-Hydroxybenzoate Hydroxylase Complexed with 4-Aminobenzoate, 2,4-Dihydroxybenzoate, and 2-Hydroxy-4-aminobenzoate and of the Tyr222Ala Mutant Complexed with 2-Hydroxy-4-aminobenzoate. Evidence for a Proton Channel and a New Binding Mode of the Flavin Ring. *Biochemistry* **33**: 10161-10170

- Schreuder HA, Prick PAJ, Wierenga RK, Vriend G, Wilson KS, Hol WGJ, Drenth J (1989) Crystal structure of the p-hydroxybenzoate hydroxylase-substrate complex refined at 1.9 Å resolution: Analysis of the enzyme-substrate and enzyme-product complexes. *J. Mol. Biol.* **208**: 679-696
- Schreuder HA, van der Laan JM, Hol WGJ, Drenth J (1988) Crystal structure of p-hydroxybenzoate hydroxylase complexed with its reaction product 3,4-dihydroxybenzoate. *J. Mol. Biol.* **199**: 637-648
- Schulz F, Leca F, Hollmann F, Reetz MT (2005) Towards practical biocatalytic Baeyer-Villiger reactions: applying a thermostable enzyme in the gram-scale synthesis of optically-active lactones in a two-liquid-phase system. *Beilstein J. Org. Chem.* **1**
- Schwab JM, Li W, Thomas LP (1983) Cyclohexanone oxygenase: stereochemistry, enantioselectivity, and regioselectivity of an enzyme-catalyzed Baeyer-Villiger reaction. *J. Am. Chem. Soc.* **105**: 4800-4808
- Segel IH (1993) In *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*, pp 100-120. New York: John Wiley and Sons
- Sheng D, Ballou DP, Massey V (2001) Mechanistic Studies of Cyclohexanone Monooxygenase: Chemical Properties of Intermediates Involved in Catalysis. *Biochemistry* **40**: 11156-11167
- Simpson H, Alphan V, Furstoss R (2001) Microbiological transformations: 49. Asymmetric biocatalysed Baeyer-Villiger oxidation: improvement using a recombinant *Escherichia coli* whole cell biocatalyst in the presence of an adsorbent resin. *J. Mol. Catal. B: Enzym.* **16**: 101-108
- Slamon DJ, Romond EH, Perez EA, Inc. CC (2006) Advances in adjuvant therapy for breast cancer. *Clin. Adv. Hematol. Oncol.* **4**: Suppl. 1, 4-9
- Snowden M, Bermudez A, Kelly DR, Radkiewicz-Poutsma JL (2004) The Preference for Anti over Gauche Migration in the Baeyer-Villiger Reaction. *J. Org. Chem.* **69**: 7148-7156
- Spencer HT, Sorrentino BP, Pui CH, Chunduru SK, Sleep SEH, Blakley RL (1996) Mutations in the gene for human dihydrofolate reductase: An unlikely cause of clinical relapse in pediatric leukemia after therapy with methotrexate. *Leukemia* **10**: 439-446
- Stinnett LG, Smiley RD, Hicks SN, Howell EE (2004) "Catch 22," the Effects of Symmetry on Ligand Binding and Catalysis in R67 Dihydrofolate Reductase as Determined by Mutations at Tyr-69. *J. Biol. Chem.* **279**: 47003-47009

- Storoni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **D60**: 432-438
- Strader MB, Smiley RD, Stinnett LG, VerBerkmoes NC, Howell EE (2001) Role of S65, Q67, I68, and Y69 Residues in Homotetrameric R67 Dihydrofolate Reductase. *Biochemistry* **40**: 11344-11352
- Strojan P, Soba E, Budihna M, Auersperg M (2005) Radiochemotherapy with Vinblastine, Methotrexate, and Bleomycin in the treatment of verrucous carcinoma of the head and neck. *J. Surg. Oncol.* **92**: 278-283
- Strukul G (1998) Transition Metal Catalysis in the Baeyer-Villiger Oxidation of Ketones. *Angew. Chem. Int. Ed.* **37**: 1198-1209
- Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expres. Purif.* **41**: 207-234
- Svergun D, Barberato C, Koch MHJ (1995) CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**: 768-773
- Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**: 495-503
- Szolkowy C, Eltis LD, Bruce NC, Grogan G (2009) Insights into Sequence–Activity Relationships amongst Baeyer–Villiger Monooxygenases as Revealed by the Intragenomic Complement of Enzymes from *Rhodococcus jostii* RHA1. *ChemBioChem* **10**: 1208-1217
- Takemura Y, Kobayashi H, Miyachi H (1999) Variable expression of the folylpolyglutamate synthetase gene at the level of mRNA transcription in human leukemia cell lines sensitive, or made resistant, to various antifolate drugs. *Anti-Cancer Drug.* **10**: 677-683
- Taschner MJ, Chen Q-Z (1991) The Enzymatic Baeyer-Villiger Oxidation: Synthesis of the C11-C16 Subunit of Ionomycin. *Bioorg. Med. Chem. Lett.* **1**: 535-538
- Teague SJ (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* **2**: 527-541
- ten Brink GJ, Arends IWCE, Sheldon RA (2004) The Baeyer–Villiger Reaction: New Developments toward Greener Procedures. *Chem. Rev.* **104**: 4105-4124
- Thompson PD, Freisheim JH (1991) Conversion of arginine to lysine at position 70 of human dihydrofolate reductase: generation of a methotrexate-insensitive mutant enzyme. *Biochemistry* **30**: 8124-8130

- Torres Pazmiño DE, Baas B-J, Janssen DB, Fraaije MW (2008a) Kinetic Mechanism of Phenylacetone Monooxygenase from *Thermobifida fusca*. *Biochemistry* **47**: 4082-4093
- Torres Pazmiño DE, Dudek HM, Fraaije MW (2010) Baeyer-Villiger monooxygenases: recent advances and future challenges. *Curr. Opin. Chem. Biol.* **14**: 138-144
- Torres Pazmiño DE, Riebel A, de Lange J, Rudroff F, Mihovilovic MD, Fraaije MW (2009) Efficient Biooxidations Catalyzed by a New Generation of Self-Sufficient Baeyer-Villiger Monooxygenases. *ChemBioChem* **10**: 2595-2598
- Torres Pazmiño DE, Snajdrova R, Baas B-J, Ghobrial M, Mihovilovic MD, Fraaije MW (2008b) Self-Sufficient Baeyer-Villiger Monooxygenases: Effective Coenzyme Regeneration for Biooxygenation by Fusion Engineering. *Angew. Chem. Int. Ed.* **47**: 2275-2278
- Torres Pazmiño DE, Snajdrova R, Rial Daniela V, Mihovilovic MD, Fraaije MW (2007) Altering the Substrate Specificity and Enantioselectivity of Phenylacetone Monooxygenase by Structure-Inspired Enzyme Redesign. *Adv. Synth. Catal.* **349**: 1361-1368
- Turfitt GE (1948) The Microbiological Degradation of Steroids. *Biochem. J.* **42**: 376-383
- Turner RB (1950) Stereochemistry of the Peracid Oxidation of Ketones. *J. Am. Chem. Soc.* **72**: 878-882
- van Beek HL, de Gonzalo G, Fraaije MW (2012) Blending Baeyer-Villiger monooxygenases: using a robust BVMO as a scaffold for creating chimeric enzymes with novel catalytic properties. *Chem. Commun. (Camb.)* **48**: 3288-3290
- van den Heuvel RHH, Tahallah N, Kamerbeek NM, Fraaije MW, van Berkel WJH, Janssen DB, Heck AJR (2005) Coenzyme Binding during Catalysis Is Beneficial for the Stability of 4-Hydroxyacetophenone Monooxygenase. *J. Biol. Chem.* **280**: 32115-32121
- Volpato JP, Fossati E, Pelletier JN (2007) Increasing Methotrexate Resistance by Combination of Active-site Mutations in Human Dihydrofolate Reductase. *J. Mol. Biol.* **373**: 599-611
- Volpato JP, Mayotte N, Fossati E, Guerrero V, Sauvageau G, Pelletier JN (2011) Selectively weakened binding of methotrexate by human dihydrofolate reductase allows rapid *ex vivo* selection of mammalian cells. *J. Mol. Recognit.* **24**: 188-198

- Volpato JP, Pelletier JN (2009) Mutational 'hot-spots' in mammalian, bacterial and protozoal dihydrofolate reductases associated with antifolate resistance: Sequence and structural comparison. *Drug Resist. Update*. **12**: 28-41
- Volpato JP, Yachnin BJ, Blanchet J, Guerrero V, Poulin L, Fossati E, Berghuis AM, Pelletier JN (2009) Multiple Conformers in Active Site of Human Dihydrofolate Reductase F31R/Q35E Double Mutant Suggest Structural Basis for Methotrexate Resistance. *J. Biol. Chem.* **284**: 20079-20089
- Walton AZ, Stewart JD (2002) An Efficient Enzymatic Baeyer–Villiger Oxidation by Engineered *Escherichia coli* Cells under Non-Growing Conditions. *Biotechnol. Progr.* **18**: 262-268
- Wang J, Ortiz-Maldonado M, Entsch B, Massey V, Ballou DP, Gatti DL (2002) Protein and ligand dynamics in 4-hydroxybenzoate hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 608-613
- Watt ED, Shimada H, Kovrigin EL, Loria JP (2007) The mechanism of rate-limiting motions in enzyme function. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 11981-11986
- Wernimont A, Edwards AM (2009) *In Situ* Proteolysis to Generate Crystals for Structure Determination: An Update. *PLoS ONE* **4**: e5094
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **D67**: 235-242
- Wittig G, Pieper G (1940) Über das monomere Fluorenon-peroxyd. *Ber. Dtsch. Chem. Ges.* **73**: 295-297
- Wu S, Acevedo JP, Reetz MT (2010) Induced allostery in the directed evolution of an enantioselective Baeyer–Villiger monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 2775-2780
- Xu D, Ballou DP, Massey V (2001) Studies of the Mechanism of Phenol Hydroxylase: Mutants Tyr289Phe, Asp54Asn, and Arg281Met. *Biochemistry* **40**: 12369-12378
- Yang J, Wang S, Lorrain M-J, Rho D, Abokitse K, Lau PCK (2009) Bioproduction of lauryl lactone and 4-vinyl guaiacol as value-added chemicals in two-phase biotransformation systems. *Appl. Microbiol. Biotechnol.* **84**: 867-876
- Zambianchi F, Pasta P, Carrea G, Colonna S, Gaggero N, Woodley JM (2002) Use of isolated cyclohexanone monooxygenase from recombinant *Escherichia coli* as a biocatalyst for Baeyer–Villiger and sulfide oxidations. *Biotechnol. Bioeng.* **78**: 489-496

Zhang Z-G, Parra LP, Reetz MT (2012) Protein engineering of stereoselective Baeyer-Villiger monooxygenases. *Chem. Eur. J.* **18**: 10160-10172

Zhuang P, Yin M, Holland JC, Peterson CB, Howell EE (1993) Artificial duplication of the R67 dihydrofolate reductase gene to create protein asymmetry. *J. Biol. Chem.* **268**: 22672-22679