STRUCTURAL CHARACTERIZATION OF UDP-GALACTOPYRANOSE MUTASE FROM EUKARYOTIC PATHOGENS

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
Of the Requirements for the Degree

Doctor of Philosophy

By

RICHA DHATWALIA

Prof. John J. Tanner, Dissertation Supervisor

DECEMBER 2013

The undersigned, appointed by the Dean of the Graduate School, have examined

the dissertation entitled

STRUCTURAL CHARACTERIZATION OF UDP-GALACTOPYRANOSE MUTASE FROM EUKARYOTIC PATHOGENS

presented by Richa Dhatwalia,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor John J. Tanner

Professor Lesa J. Beamer

Professor Kent S. Gates

Dr. Thomas J. Reilly

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Prof. John J. Tanner for his patience in teaching me during the first years of my PhD. He had spent a considerable amount of time in teaching me crystallography and the very minute details of solving three-dimensional crystal structures. I would never have been so much detail oriented in solving structures, if Dr. Tanner had not corrected me every time I made silly mistakes while solving my first LpHAP structure. These corrections made me think about the basic chemistry behind all these structural problems. I would also like to thank his wife, Mrs. Evelyn Tanner for all her help and the baby shower.

I would like to thank my committee members- Dr. Thomas J. Reilly, Prof. Lesa J. Beamer and Prof. Kent S. Gates for their time and guidance. Also thanks to Dr. Reilly for always offering his help at all times.

Thanks to our collaborators at Virginia Tech, Prof. Pablo Sobrado and his student Dr. Michelle Oppenheimer for sending us everything we needed before time. Also, thanks to Dr. Jay Nix at ALS beamline 4.2.2 for collecting data on UGM and to Dr. Jonathan Schuermann for his help in data collection and processing at APS beamline 24-ID-C.

My heartiest thanks to Dr. Harkewal Singh, my colleague and life partner. I would not have reached this far in my PhD, if Harkewal was not there to support me in both professional and personal lives. Protein biochemistry was something I never imagined I could do, but Harkewal not only taught me in the lab, in fact, made sure that I do not escape without learning every concept behind cloning and protein purification. I remember the days when I would not remember how to set up the superloop for protein

purification. Though annoyed he would still teach me until I would not stop making any more mistakes in performing experiments. Whether it was LpHAP purification or sub-cloning of PPK or solving UGM structures, I would always cherish those days of hardwork and late nights with IHOP pancakes. I would never have been able to accomplish what I have during my PhD without your help and constant source of motivation. Thank you for helping me grow at both personal and professional levels. I wish we could again get a chance to work together with the same passion as 2 years ago.

I would also like to thank my parents, especially, my mother Mrs. Pushpa Dhatwalia and mother-in-law Mrs. Jaswinder Kaur, who helped me during my last months in graduate school. Without your help at home, I would not have been able to finish off my graduate school with a newborn baby.

Thanks to Tanner lab members, Dale Karr for introducing me to my PhD project, Travis Pemberton and Min Luo for all their help and a good time in the lab.

At last, I would like to thank the most important part of my life, my lovely daughter, Mehar for bringing all the happiness in our lives. Your presence makes me feel a strong and confidant woman. I would try my best in life to achieve things that would make you feel proud of your mother.

Sincerely,

Richa Dhatwalia

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABBREVIATIONS

AfUGM	*Aspergillus fumigatus* UDP-galactopyranose mutase
ADP	Adenosine dinucleotide phosphate
ALS	Advanced light source
AMP	Adenosine monophosphate
APS	Advanced photon source
Arap	Arabinopyranose
Araf	Arabinofuranose
COOT	Crystallographic object-oriented toolkit
DEAE	Diethyl aminoethyl cellulose
DrUGM	*Deinococcus radiodurans* UDP-galactopyranose mutase
FAD	Flavin adenine dinucleotide
FADH	Flavin adenine dinucleotide reduced
FMN	Flavin mononucleotide
FtHAP	*Francisella tularensis* Histidine Acid Phosphatase
Galp	Galactopyranose
Galf	Galactofuranose
GIPL	Glycoinositolphospholipid
GPI	Glycophosphatidylinositol
HAP	Histidine Acid Phosphatase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hPAP	Human Prostatic Acid Phosphatase

KIE	Kinetic isotope effect
Km	Michaelis Constant
KpUGM	*Klebsiella pneumoniae* UDP-galactopyranose mutase
LPG	lipophosphoglycans
LpHAP	*Legionella pneumophila* Histidine Acid Phosphatase
LmUGM	*Leishmania major* UDP-galactopyranose mutase
MAP	Major Acid Phosphatase
MOLREP	Molecular replacement
MtUGM	*Mycobacterium tuberculosis* UDP-galactopyranose mutase
NADH	Nicotinamide adenine dinucleotide reduced
NADPH	Nicotinamide adenine dinucleotide phosphate reduced
NCS	Non-crystallographic symmetry
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PHENIX	Python based hierarchical environment for integrated crystallography
PISA	Protein Interactions, Surfaces and Assemblies
Pi	Inorganic phosphate
pNP	*p*- nitrophenol
pNPP	*p*- nitrophenol phosphate
polyP	Inorganic polyphosphate
PPOX	Protoporphyrinogen oxidase from *Myxococcus xanthus*
RMSD	Root mean square deviation

SAD	Single-wavelength Anomalous Diffraction
SAXS	Small Angle X-ray Scattering
SeMet	Selenomethionine
TcUGM	*Trypanosoma cruzi* UDP-galactopyranose mutase
TLS	Translational Libration Screw-motion
TPS	Translational Pseudosymmetry
UDP	Uridine diphosphate
UMP	Uridine monophosphate
UGM	UDP-galactopyranose mutase
V*M*	Matthew's volume
V*max*	Maximal Velocity
XDS	X-ray detector software

STRUCTURAL CHARACTERIZATION OF UDP-GALACTOPYRANOSE MUTASE FROM EUKARYOTIC PATHOGENS

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
Of the Requirements for the Degree

Doctor of Philosophy

By

RICHA DHATWALIA

Prof. John J. Tanner, Dissertation Supervisor

DECEMBER 2013

The undersigned, appointed by the Dean of the Graduate School, have examined

the dissertation entitled

STRUCTURAL CHARACTERIZATION OF UDP-GALACTOPYRANOSE MUTASE FROM EUKARYOTIC PATHOGENS

presented by Richa Dhatwalia,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor John J. Tanner

Professor Lesa J. Beamer

Professor Kent S. Gates

Dr. Thomas J. Reilly

1.1 Introduction

The emergence of drug-resistant strains of pathogenic bacteria, fungi and protozoans requires a consistent effort from scientists worldwide in developing new drugs. The cell wall core of these pathogens is composed of several different glycoconjugate structures, lipids and polysaccharides; essential for the cell wall rigidity, adhesion to mammalian host and pathogen-host interactions. Among these components is the monosaccharide galactofuranose (Galf) present in abundance in opportunistic fungi, *Aspergillus fumigatus* and protozoan parasites such as *Trypanosoma cruzi* and *Leishmania major*[1].

UDP-galactopyranose mutase (UGM) is a central enzyme involved in the biosynthesis of galactofuranose in these pathogens[2]. UGM is an atypical flavoenzyme that catalyzes the interconversion between UDP-galactopyranose (UDP-Gal_p) and UDP-galactofuranose (UDP-Gal_f) (Fig. 1.1). Flavoenzymes are typically known to catalyze redox reactions, but UGM (although active in its reduced state) catalyzes the reaction without any net transfer of electrons.

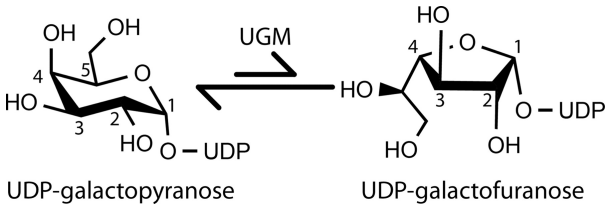

Fig. 1.1 Reaction catalyzed by UGM.

Bacterial UGMs from *Escherichia coli*[3], *Mycobacterium tuberculosis*[4] and *Klebsiella pneumoniae*[4b] have been somewhat characterized in detail using structural and

biochemical methods, but major questions about the catalytic and structural properties of eukaryotic UGMs remain unanswered. Furthermore, eukaryotic UGMs share only ~15 % sequence homology with prokaryotic UGMs. Thus, the unique chemistry of UGMs and its implication in the virulence of pathogenic bacteria, fungi and protozoans make it a potential drug target[5].

1.1 Biosynthesis of Galactofuranose

Galactofuranose (Galf) is the five-membered ring form of galactose present in most microbial pathogens and lower eukaryotes. All mammals have this sugar, but only in the six-membered pyranose ring form. Several enzymes are involved in the biosynthesis of Galf and its transport to different cell wall and extracellular structures of pathogenic bacteria, fungi and protozoans (Fig. 1.2). The major component involved in the biosynthesis of galactofuranose is UDP-Galf, the only known donor of Galf in these pathogens.

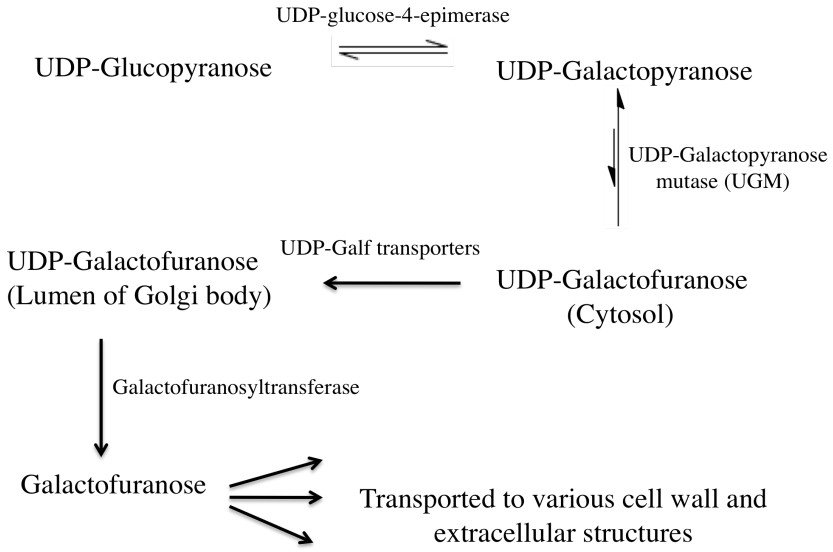

Fig. 1.2 Schematic of enzymes involved in Galactofuranose biosynthetic pathway.

1.2.1 UDP-Glucose-4'-epimerase (GalE)

The first enzyme required for Galf biosynthesis is UDP-glucose-4'-epimerase (GalE) that catalyzes the reversible reaction between UDP-glucopyranose (UDP-Glc_p) and UDP-galactopyranose (UDP-Gal_p). In *T. cruzi,* GalE metabolism is important for the survival of the parasite[6]. Several inhibitors have been identified against this enzyme but due to the presence of GalE homolog in humans, it poses toxic side effects. On the contrary, in *L. major*, galactose can be obtained from the environment and studies have shown that enzymes involved in the synthesis of GalE are not essential for the survival of *Leishmania* spp.[7]

1.2.2 UDP-Galactopyranose mutase (UGM)

UGM catalyzes the interconversion between UDP-Gal_p and UDP-Gal_f. It is the sole biological source of UDP-Gal_f and is absent in humans thus, an ideal drug target. Moreover, deletion of the gene-encoding UGM (*glf*) in *L. major* shows that it is essential for its survival and pathogenesis[5a]. Previous studies have also shown that Galf is a virulence factor in the pathogenic fungus *Aspergillus fumigatus*[5b].

1.2.3 UDP-Galf transporters

The product of UGM catalysis, UDP-α-Galf is transported from cytosol to the lumen of Golgi body by the enzyme, UDP-Galf transporter for galactofuranosylation of glycoproteins and glycolipids in the cell wall structures. These enzymes are not very well characterized until recently, the first UDP-Galf transporter specific for UDP-Gal_f was identified in *Aspergillus fumigatus*[8].

1.2.4 UDP- Galactofuranose transferases (GalfT)

Finally, another enzyme directly involved in Galf metabolism is Galactofuranosyltransferase that transfers Galf to various glycoconjugate structures present at the cell surface or extracellular matrix. Among these enzymes is *LPG*-1, that has been extensively studied in *L. major* and is responsible for the addition of Galf to lipophosphoglycans (LPG), but not glycoinositolphospholipids (GIPLs)[9]. Sequence alignment results have shown that more than 30 GalfTs are present in *T. cruzi* that share significance sequence homology with *L. major LPG*1[10]. Due to the large number of putative GalfTs reported in *T. cruzi*, targeting this enzyme might not be an effective strategy.

Therefore, UGM is a central enzyme involved in Galf biosynthesis and thus, a potential drug target.

1.2 Role of galactofuranose in eukaryotic UGMs

1.2.1 Role of Galf in the pathogenesis of Aspergillus fumigatus

Aspergillus fumigatus is the cause of invasive aspergillosis (IA) and bronchopulmonary aspergillosis (BPA) in immunocompromised individuals. The overall mortality rates associated with IA are > 50 % and as high as 95 % in certain cases[11]. Over 10 % of the patients receiving bone marrow transplantation develop invasive aspergillosis during clinical treatment[12]. Despite the use of antifungal agents in treating aspergillosis, overall mortality rate associated with IA remains high[13]. This suggests that new antifungal agents and co-therapies are needed for treating aspergillosis.

The cell wall of *A. fumigatus* is composed of galactomannan residues consisting

of a linear mannan core branched with galactofuranose side chains (Fig. 1.3). The deletion of the gene-encoding UGM (*glf*) results in a decreased cell wall thickness of *A. fumigatus*, thus making it more susceptible to antifungal drugs[2a, 5b]. Also, tests for pathogenecity in a low-dose mouse model of IA have shown that Δ*glf* mutant has an attenuated virulence as compared to those inflicted with the wt[5b]. Thus, the decreased virulence and increased susceptibility to antifungal drugs make UGM a potential drug target for treating aspergillosis.

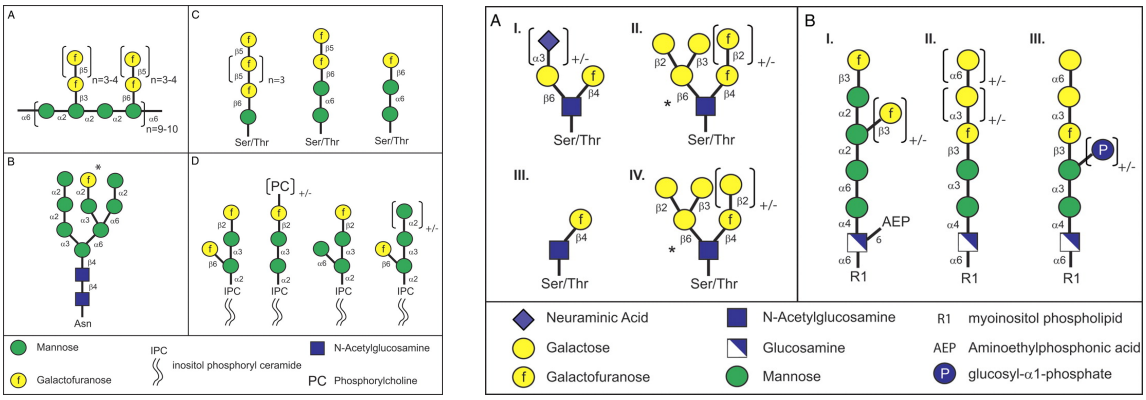

Fig. 1.3 Different Galf containing glycoconjugates found in (A) *Aspergillus* spp. and (B) protozoans[2b].

1.2.1 Role of Galf in the pathogenesis of Leishmania major

Leishmania major is the cause of cutaneous and lethal visceral leishmaniasis; not only in tropical and mediterranean regions of Africa, India, Middle East and Asia, but also in parts of Southern Europe[14]. *Leishmania* parasite is transmitted to humans by female sandfly bite and causes the third most important vector-borne disease after malaria and African trypanosomiasis[15]. Approximately 58,000 cases of visceral leishmaniasis and 220,000 cases of cutaneous leishmaniasis are reported officially each year[16]. Current treatments use pentavalent antimony compounds that are highly toxic or amphotericin B

treatment that is highly expensive[17]. Thus, new drug development strategies are required for combating this lethal disease.

The cell wall and extracellular matrix of *Leishmania major* is composed of glycoconjugates such as glycophosphatidylinositol (GPI), lipophosphoglycan (LPG) and glycoinositolphospholipids (GIPLs) that consist of galactofuranosyl residues (Fig. 1.2b). Gene deletion studies in *L. major* have shown that mice infected with Δ*glf* mutant delayed the appearance of lesions at the site of infection by ~3-4 weeks (Fig. 1.4) thus, attenuating the virulence[5a].

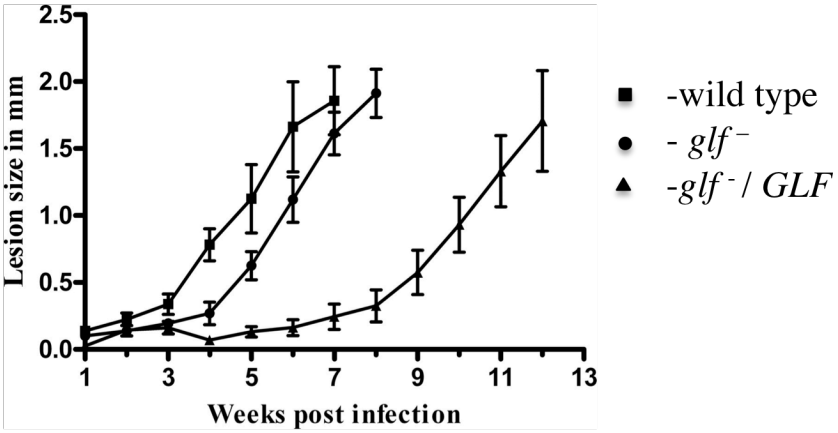

Fig. 1.4 Virulence of the *glf*- mutant is attenuated in a murine infection model[5a].

1.2.3 Role of galactofuranose in the pathogenesis of Trypanosoma cruzi

Chagas disease or American trypanosomiasis is a vector-borne disease, transmitted to humans by triatomine bug. The disease can be transmitted by blood transfusion, organ transplantation, congenital transmission and by ingesting contaminated food and drinks. Chagas disease was initially thought to be affecting only people living in

endemic areas of Latin America, but recent reports have suggested that more than 300,000 people living in United States are infected with this disease[18].

In *T. cruzi*, Galf is present in glycoinositolphospholipids (GIPLs), N- and O-linked glycans and GPI-linked glycoproteins of the cell surface structures[1a]. As UGM is a central enzyme involved in Galf metabolism, targeting this enzyme would provide a novel strategy for developing inhibitors against this disease.

1.3 Structural characterization of prokaryotic UGMs

The first structure of UGM was determined from *Escherichia coli* in 2001[3]. Since then, several structures of UGMs from *Mycobacterium tuberculosis*[4a], *Klebsiella pneumoniae* and *Dienococcus radiodurans*[19] have been determined with and without substrate; but no eukaryotic UGM structure was determined until this work. Although the residues in flavin-binding domain and substrate-binding site are mostly conserved, bacterial UGMs share only ~15% global sequence identity with eukaryotic UGMs.

Even after the determination of structures of prokaryotic UGMs, several questions about UGMs are still unanswered. UGMs are active in their reduced state but so far, not much is known about the biological reducing partners of the enzyme[3, 20]. Moreover, structures of bacterial UGMs show no variation in the hydrogen-bonding pattern in the oxidized and reduced states. Both the states have a hydrogen bond formed between N5 of FAD isoalloxazine and backbone carbonyl of Gly residue[4a]. However, this pattern is characteristic of the reduced state of enzyme. Thus, structural characterization of eukaryotic UGMs might help us in determining the structural changes associated with different redox states of FAD and understand the enzymatic mechanism.

1.4 Proposed enzymatic mechanism for UGM

UGM is an atypical flavoenzyme that catalyzes the reversible reaction between UDP-Gal_p and UDP-Gal_f without any net redox change[3]. Positional isotope exchange experiments have shown that the reaction proceeds with the cleavage of the bond between the anomeric carbon C1 and UDP moiety[21].

Earlier evidences suggested that the reaction occurs by one-electron transfer forming an oxocarbenium intermediate[22] . Recently, a novel role for flavin was proposed where N5 of the FAD isoalloxazine acts as a nucleophile and attacks the anomeric carbon C1 of the galactose moiety to form a covalent flavin-iminium adduct (Fig. 1.5)[23]. The UDP moiety is then displaced from UDP-Gal_p or UDP-Gal_f via SN_2- type mechanism[24].

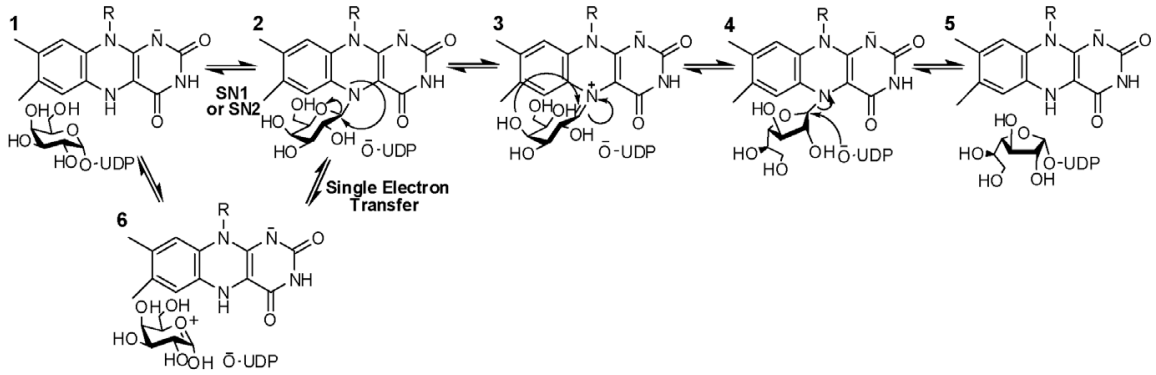

Fig. 1.5 Mechanism involving flavin-iminium adduct as the intermediate in the reaction catalyzed by UGM[25].

1.6 Conclusion

UGM plays a central role in the virulence of pathogenic fungi and protozoa such as *Aspergillus fumigatus*, *Trypanosoma cruzi* and *Leishmania major*. These pathogenic

protozoa are implicated in neglected tropical diseases that have gained tremendous attention worldwide because of high mortality rates[16, 18]. Due to emigration and globalization, these diseases that were once endemic to tropical and mediterranean regions, have now also emerged in parts of United States and Europe[14, 18].

In addition to its medical relevance, UGM has a unique chemistry among flavoenzymes that has never been observed before. Thus, the determination of three-dimensional structures of eukaryotic UGMs might help us in elucidating the enzymatic mechanism of this class of enzymes and potential inhibitor design.

1.7 References

1. (a) de Lederkremer, R. M.; Colli, W., Galactofuranose-containing glycoconjugates in trypanosomatids. *Glycobiology* **1995,** *5* (6), 547-52; (b) Tefsen, B.; Ram, A. F.; van Die, I.; Routier, F. H., Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact. *Glycobiology* **2012,** *22* (4), 456-69; (c) Richards, M. R.; Lowary, T. L., Chemistry and biology of galactofuranose-containing polysaccharides. *Chembiochem : a European journal of chemical biology* **2009,** *10* (12), 1920-38.

2. (a) Beverley, S. M.; Owens, K. L.; Showalter, M.; Griffith, C. L.; Doering, T. L.; Jones, V. C.; McNeil, M. R., Eukaryotic UDP-galactopyranose mutase (GLF gene) in microbial and metazoal pathogens. *Eukaryot Cell* **2005,** *4* (6), 1147-54; (b) Pedersen, L. L.; Turco, S. J., Galactofuranose metabolism: a potential target for antimicrobial chemotherapy. *Cell Mol Life Sci* **2003,** *60* (2), 259-66.

3. Sanders, D. A.; Staines, A. G.; McMahon, S. A.; McNeil, M. R.; Whitfield, C.; Naismith, J. H., UDP-galactopyranose mutase has a novel structure and mechanism. *Nat Struct Biol* **2001,** *8* (10), 858-63.

4. (a) Partha, S. K.; van Straaten, K. E.; Sanders, D. A., Structural basis of substrate binding to UDP-galactopyranose mutase: crystal structures in the reduced and oxidized state complexed with UDP-galactopyranose and UDP. *J Mol Biol* **2009,** *394* (5), 864-77; (b) Beis, K.; Srikannathasan, V.; Liu, H.; Fullerton, S. W.; Bamford, V. A.; Sanders, D. A.; Whitfield, C.; McNeil, M. R.; Naismith, J. H., Crystal structures of Mycobacteria tuberculosis and Klebsiella pneumoniae UDP-galactopyranose mutase in the oxidised state and Klebsiella pneumoniae UDP-galactopyranose mutase in the (active) reduced state. *J Mol Biol* **2005,** *348* (4), 971-82.

5. (a) Kleczka, B.; Lamerz, A. C.; van Zandbergen, G.; Wenzel, A.; Gerardy-Schahn, R.; Wiese, M.; Routier, F. H., Targeted gene deletion of Leishmania major UDP-galactopyranose mutase leads to attenuated virulence. *J Biol Chem* **2007,** *282* (14), 10498-505; (b) Schmalhorst, P. S.; Krappmann, S.; Vervecken, W.; Rohde, M.; Muller, M.; Braus, G. H.; Contreras, R.; Braun, A.; Bakker, H.; Routier, F. H., Contribution of galactofuranose to the virulence of the opportunistic pathogen Aspergillus fumigatus. *Eukaryot Cell* **2008,** *7* (8), 1268-77.

6. MacRae, J. I.; Obado, S. O.; Turnock, D. C.; Roper, J. R.; Kierans, M.; Kelly, J. M.; Ferguson, M. A., The suppression of galactose metabolism in Trypanosoma cruzi epimastigotes causes changes in cell surface molecular architecture and cell morphology. *Mol Biochem Parasitol* **2006,** *147* (1), 126-36.

7. Lamerz, A. C.; Damerow, S.; Kleczka, B.; Wiese, M.; van Zandbergen, G.; Lamerz, J.; Wenzel, A.; Hsu, F. F.; Turk, J.; Beverley, S. M.; Routier, F. H., Deletion of UDP-glucose pyrophosphorylase reveals a UDP-glucose independent UDP-galactose salvage pathway in Leishmania major. *Glycobiology* **2010,** *20* (7), 872-82.

8. Engel, J.; Schmalhorst, P. S.; Dork-Bousset, T.; Ferrieres, V.; Routier, F. H., A single UDP-galactofuranose transporter is required for galactofuranosylation in Aspergillus fumigatus. *J Biol Chem* **2009,** *284* (49), 33859-68.

9. Spath, G. F.; Epstein, L.; Leader, B.; Singer, S. M.; Avila, H. A.; Turco, S. J.; Beverley, S. M., Lipophosphoglycan is a virulence factor distinct from related glycoconjugates in the protozoan parasite Leishmania major. *Proc Natl Acad Sci U S A* **2000,** *97* (16), 9258-63.

10. Oppenheimer, M.; Valenciano, A. L.; Sobrado, P., Biosynthesis of galactofuranose in kinetoplastids: novel therapeutic targets for treating leishmaniasis and chagas' disease. *Enzyme research* **2011,** *2011*, 415976.

11. Abad, A.; Fernandez-Molina, J. V.; Bikandi, J.; Ramirez, A.; Margareto, J.; Sendino, J.; Hernando, F. L.; Ponton, J.; Garaizar, J.; Rementeria, A., What makes Aspergillus fumigatus a successful pathogen? Genes and molecules involved in invasive aspergillosis. *Revista iberoamericana de micologia* **2010,** *27* (4), 155-82.

12. Pappas, P. G.; Alexander, B. D.; Andes, D. R.; Hadley, S.; Kauffman, C. A.; Freifeld, A.; Anaissie, E. J.; Brumble, L. M.; Herwaldt, L.; Ito, J.; Kontoyiannis, D. P.; Lyon, G. M.; Marr, K. A.; Morrison, V. A.; Park, B. J.; Patterson, T. F.; Perl, T. M.; Oster, R. A.; Schuster, M. G.; Walker, R.; Walsh, T. J.; Wannemuehler, K. A.; Chiller, T. M., Invasive fungal infections among organ transplant recipients: results of the Transplant-Associated Infection Surveillance Network (TRANSNET). *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2010,** *50* (8), 1101-11.

13. Mansour, M. K.; Tam, J. M.; Vyas, J. M., The cell biology of the innate immune response to Aspergillus fumigatus. *Annals of the New York Academy of Sciences* **2012,** *1273*, 78-84.

14. Christodoulou, V.; Antoniou, M.; Ntais, P.; Messaritakis, I.; Ivovic, V.; Dedet, J. P.; Pratlong, F.; Dvorak, V.; Tselentis, Y., Re-emergence of visceral and cutaneous

leishmaniasis in the Greek Island of Crete. *Vector Borne Zoonotic Dis* **2012,** *12* (3), 214-22.

15. Hotez, P. J.; Remme, J. H.; Buss, P.; Alleyne, G.; Morel, C.; Breman, J. G., Combating tropical infectious diseases: report of the Disease Control Priorities in Developing Countries Project. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2004,** *38* (6), 871-8.

16. Alvar, J.; Velez, I. D.; Bern, C.; Herrero, M.; Desjeux, P.; Cano, J.; Jannin, J.; den Boer, M., Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* **2012,** *7* (5), e35671.

17. Gradoni, L.; Soteriadou, K.; Louzir, H.; Dakkak, A.; Toz, S. O.; Jaffe, C.; Dedet, J. P.; Campino, L.; Canavate, C.; Dujardin, J. C., Drug regimens for visceral leishmaniasis in Mediterranean countries. *Tropical medicine & international health : TM & IH* **2008,** *13* (10), 1272-6.

18. Bern, C.; Montgomery, S. P., An estimate of the burden of Chagas disease in the United States. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2009,** *49* (5), e52-4.

19. Karunan Partha, S.; Bonderoff, S. A.; van Straaten, K. E.; Sanders, D. A., Expression, purification and preliminary X-ray crystallographic analysis of UDP-galactopyranose mutase from Deinococcus radiodurans. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2009,** *65* (Pt 8), 843-5.

20. Koplin, R.; Brisson, J. R.; Whitfield, C., UDP-galactofuranose precursor required for formation of the lipopolysaccharide O antigen of Klebsiella pneumoniae serotype O1 is synthesized by the product of the rfbDKPO1 gene. *J Biol Chem* **1997,** *272* (7), 4121-8.

21. John N. Barlow, M. E. G., John S. Blanchard, Positional isotope exchange catalyzed by UDP-galactopyranose mutase. *J Am Chem Soc* **1999,** *121*, 2.

22. Huang, Z.; Zhang, Q.; Liu, H. W., Reconstitution of UDP-galactopyranose mutase with 1-deaza-FAD and 5-deaza-FAD: analysis and mechanistic implications. *Bioorg Chem* **2003,** *31* (6), 494-502.

23. (a) Soltero-Higgin, M.; Carlson, E. E.; Gruber, T. D.; Kiessling, L. L., A unique catalytic mechanism for UDP-galactopyranose mutase. *Nat Struct Mol Biol* **2004,** *11* (6), 539-43; (b) Miller, S. M., A new role for an old cofactor. *Nat Struct Mol Biol* **2004,** *11* (6), 497-8.

24. Sun, H. G.; Ruszczycky, M. W.; Chang, W. C.; Thibodeaux, C. J.; Liu, H. W., Nucleophilic participation of reduced flavin coenzyme in mechanism of UDP-galactopyranose mutase. *J Biol Chem* **2012,** *287* (7), 4602-8.

25. Oppenheimer, M.; Valenciano, A. L.; Kizjakina, K.; Qi, J.; Sobrado, P., Chemical mechanism of UDP-galactopyranose mutase from Trypanosoma cruzi: a potential drug target against Chagas' disease. *PLoS One* **2012,** *7* (3), e32918.

Chapter 2

Crystal structures and small-angle X-ray scattering analysis of UDP-galactopyranose mutase from the pathogenic fungus *Aspergillus fumigatus*

Richa Dhatwalia, Harkewal Singh, Michelle Oppenheimer, Dale B. Karr, Jay C. Nix, Pablo Sobrado, and John J. Tanner

Abstract

UDP-galactopyranose mutase (UGM) is a flavoenzyme that catalyzes the conversion of UDP-galactopyranose to UDP-galactofuranose, which is a central reaction in galactofuranose biosynthesis. Galactofuranose has never been found in humans but is an essential building block of the cell wall and extracellular matrix of many bacteria, fungi, and protozoa. The importance of UGM for the viability of many pathogens and its absence in humans make UGM a potential drug target. Here we report the first crystal structures and small-angle x-ray scattering data for UGM from the fungus *Aspergillus fumigatus*, the causative agent of aspergillosis. The structures reveal that *Aspergillus* UGM has several extra secondary and tertiary structural elements that are not found in bacterial UGMs yet are important for substrate recognition and oligomerization. Small angle x-ray scattering data show that *Aspergillus* UGM forms a tetramer in solution, which is unprecedented for UGMs. The binding of UDP or the substrate induces profound conformational changes in the enzyme. Two loops on opposite sides of the active site move toward each other by over 10 Å to cover the substrate and create a closed active site. The degree of substrate-induced conformational change exceeds that of bacterial UGMs and is a direct consequence of the unique quaternary structure of *Aspergillus* UGM. Galactopyranose binds at the *re* face of the FAD isoalloxazine with the anomeric carbon atom poised for nucleophilic attack by the FAD N5 atom. The structural data provide new insight into substrate recognition and the catalytic mechanism and thus will aid inhibitor design.

2.1 Introduction

UDP-galactopyranose mutase (UGM) is a unique flavoenzyme that catalyzes the reversible conversion of UDP-galactopyranose (UDP-Galp) to UDP-galactofuranose (UDP-Galf) (Fig. 2.1). Unlike flavin-dependent oxidoreductases, the redox state of the flavin in UGM is unchanged upon product formation[1]. Although it is known that the FAD must be reduced for maximal catalytic activity, the precise role that the flavin plays in catalysis remains controversial.

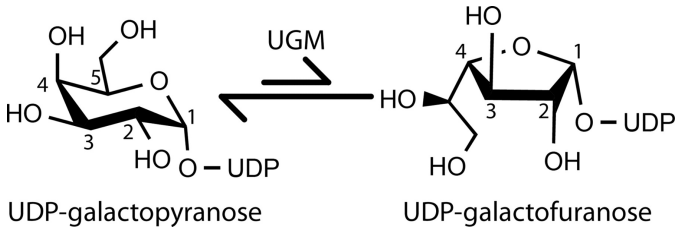

Fig. 2.1 Reaction catalyzed by UGM.

UGM is also an attractive target for drug design. The carbohydrate moiety of the product of the reaction, galactofuranose (Galf), is the five-membered ring form of the common monosaccharide galactose. Galf has never been found in mammals and higher plants but is an essential building block of the cell wall and extracellular matrix of many bacteria, fungi, and protozoa[2]. The UGM reaction is central to Galf biosynthesis. Gene deletion studies have shown that UGM is essential for mycobacterial growth[3] and contributes to the virulence of the pathogenic fungus *Aspergillus fumigatus*[4] and the protozoan parasite *Leishmania major*[5]. In summary, the importance of UGM for the viability of many pathogens and its absence in humans make UGM a potential drug target.

This study focuses on the UGM from *A. fumigatus*. *Aspergillus* spp. are ubiquitous fungi that cause diseases ranging from allergic reactions and lung infections to sepsis and death[6]. Acquired by inhalation of airborne spores, *Aspergillus* infections (mainly *A. fumigatus*) can lead to invasive pulmonary aspergillosis, particularly in immunocompromised persons[6,7,8]. Invasive pulmonary aspergillosis is also an emerging serious infection in patients with chronic obstructive pulmonary disease[9]. The overall case fatality rate associated with invasive aspergillosis is 58%[10]. Also, bronchial colonization by *A. fumigatus* in people with asthma or cystic fibrosis causes allergic bronchopulmonary aspergillosis, a hypersensitivity lung disease that can lead to airway destruction and widening and scarring of the lung, resulting in significant morbidity and mortality[11]. The ubiquity of the fungus, the increasing occurrence of *Aspergillus* infections in humans, and the high mortality rate associated with invasive pulmonary aspergillosis demonstrate the need for new anti-*Aspergillus* drugs.

Inhibition of *A. fumigatus* UGM (AfUGM) is a potential strategy for the development of new drugs to treat patients with aspergillosis. In *Aspergillus*, Galf is present in the galactomannan fraction of the cell wall, N - and O-glycans of secreted glycoproteins, glycosylphosphatidylinositol-anchored lipophosphogalactomannan, and sphingolipids[12]. It is estimated that Galf constitutes 5% of the dry weight of *A. fumigatus*[12]. Deletion of the UGM gene in *A. fumigatus* results in a thinner cell wall and increased susceptibility to drugs[4]. Furthermore, the mutated strain exhibits attenuated virulence in a low-dose mouse model of invasive aspergillosis[4]. In another study, deletion of the UGM gene resulted in a reduced growth phenotype, but virulence as measured in a higher dose mouse model was unimpaired[13]. The different outcomes of the two deletion

studies may reflect the different doses used, with the low-dose model generally considered to mimic more closely the in vivo situation[2]. Because of the importance of UGM for the fitness of *A. fumigatus*, UGM is a promising target for the development of new drugs for adjunctive treatment for aspergillosis [2,4].

Here we report four crystal structures of AfUGM along with an analysis of the oligomeric state and quaternary structure in solution using small-angle x-ray scattering (SAXS). The data reveal interesting differences from bacterial UGMs, including additional secondary structure elements, unique mode of oligomerization, and profound conformational changes induced by substrate binding. The structures should aid inhibitor design efforts.

2.2 Materials and Methods

2.2.1 Crystallization

Crystallization screens of AfUGM and Se-Met AfUGM were performed at 21°C and 4°C using the sitting-drop and hanging drop methods of vapor diffusion, as well as microbatch. These screens were performed using ligand-free protein and protein that had been incubated with UDP-glucose, UMP, or UDP. Initial structure determination efforts were hampered by translational pseudosymmetry (TPS). Crystal screening and optimization yielded two crystal forms that exhibit TPS. One of these forms appeared as hexagonal blocks in ammonium sulfate and diffracted to 3.4 Å resolution. The optimal reservoir contained 1.6 M ammonium sulfate, 5 % PEG 400, and HEPES buffer at pH 7.5 The apparent Laue symmetry is 6/mmm with unit cell lengths of a = 202 Å, c = 355 Å. The native Patterson map calculated using phenix.xtriage [15] exhibited a peak at (u = 0, v =

0, w = 0.49) with a height of 48 % relative to the origin peak, which suggests strong TPS. This crystal form was also obtained under low ionic strength conditions, such as 13% polyethylene glycol (PEG) 8000 and HEPES buffer at pH 8.0. Curiously, the external morphology of the PEG form was thin blades instead of hexagonal blocks. Nevertheless, the lattice and TPS were identical to that of the ammonium sulfate form. Various PEG solutions also yielded an orthorhombic crystal form, which diffracted to 3.2 Å. The apparent Laue symmetry is mmm with a = 111 Å, b = 207 Å, and c = 234 Å. The native Patterson map exhibited a strong (22 - 45 %) off-origin peak near (0, 0.25, 0.5), indicating TPS. Se-Met SAD phasing was attempted with these crystal forms, but the maps were not interpretable. Additive screening was used to identify new crystal forms. The most promising result was obtained using ammonium sulfate as the precipitant and isopropanol as the additive. The inclusion of 5 % isopropanol changed the morphology from hexagonal block to hexagonal bipyramid. The apparent Laue symmetry is 6/mmm with cell dimensions of a = 217 Å, c = 325 Å. Note that these dimensions are about 8 % different from those obtained in the absence of isopropanol. Moreover, the native Patterson map did not indicate TPS. However, these crystals diffracted to only about 5.5Å resolution. Structure determination via Se-Met SAD phasing failed, due presumably to poor data quality. Several site-directed mutants of AfUGM were screened to identify crystal forms that are free of pathologies and diffract to higher resolution. The mutations were designed from a homology model of AfUGM calculated using Phyre [16]. The mutant enzymes that were subjected to crystal screening included the predicted active site mutants W167A, R211A, R327K, and W204A, as well the predicted surface double mutants K231A/K233A, K239A/K241A, K344A/K345A, and K363A/E364A. Most of

the aforementioned undesirable crystal forms were also observed with the mutant enzymes. This difficulty was circumvented by using the double mutant enzyme K344A/K345A for structure determination. The kinetic constants of K344A/K345A are virtually identical to those of the native enzyme (Table 2.1), and the structures show that residues 344 and 345 are on the surface of the enzyme, far from the active site and oligomerization interfaces.

Table 2.1 Steady state kinetics of AfUGM and K344A/K345A.

	$\mathbf{k_{cat}}$	$\mathbf{K_m}$	$\mathbf{k_{cat}/K_m}$
Native	60 ± 3	70 ± 10	0.84 ± 0.13
K344A/K345A	55 ± 2	70 ± 9	0.75 ± 0.07

AfUGM mutant K344A/K345A (referred to hereafter as AfUGM) was expressed and purified as described for the native enzyme[14]. Se-Met AfUGM was produced using the metabolic inhibition method[15]. Crystals were grown in sitting drops at room temperature using drops formed by mixing 1.5 μl of the protein stock solution with an equal volume of the reservoir. The protein stock solution consisted of 8 mg/ml protein in 25 mM HEPES, 125 mM NaCl, 5 mM L-cysteine, and 0.5 mM Tris(hydroxypropyl)phosphine at pH 7.5. The optimal reservoir contained 1.5 M ammonium sulfate and 0.1 M sodium acetate, pH 4.5. The crystals appeared as large, yellow hexagonal bipyramids. The crystals were cryoprotected by exchanging the mother liquor with 1.8 M ammonium sulfate, 0.1 M sodium acetate, pH 4.5, and 25% glycerol. The space group is $P6_522$ with unit cell dimensions of a = 218 Å and c = 320 Å. Although

a conventional estimate of the solvent content using the method of Matthews[16] yields the expectation of eight molecules in the asymmetric unit, there are, in fact, only four molecules present. The solvent content and VM are thus 75% and 4.9 $Å^3$/Da.

2.2.2 Crystal Soaking

Crystals of reduced AfUGM with and without bound active site ligands were prepared by soaking the aforementioned crystals. Crystals of ligand-free reduced AfUGM were prepared by soaking crystals in the cryobuffer supplemented with 80 mM dithionite. At this concentration, the enzyme is fully reduced in solution[14]. Indeed, the yellow color of the crystals was bleached upon soaking, indicating that the FAD is reduced. When the transformation was complete (2–3 min), the crystals were plunged into liquid nitrogen to trap the reduced state. Crystals of the reduced enzyme complexed with UDP were prepared by soaking with the cryobuffer supplemented with 80 mM dithionite and 40 mM UDP. The soaking time was ~30 min. Crystals of the reduced enzyme complexed with UDP-Galp were prepared similarly using 80 mM dithionite and 100 mM UDP-Galp. As with the ligand-free enzyme, the crystals of the ligand complexes were colorless when frozen.

2.2.3 X-ray Diffraction Data Collection, Phasing, and Refinement

Crystals were analyzed at beamlines 19ID and 24-ID-C of the Advanced Photon Source and beamline 8.2.2 of the Advanced Light Source. The 19ID data sets were processed with HKL3000 [17]. The other data sets were integrated with XDS and scaled with SCALA[18] via CCP4i[19].

The first structure of AfUGM was determined using single-wavelength anomalous diffraction phasing based on data collected at 19ID from Se-Met AfUGM crystals (Table 2.2). Several single-wavelength anomalous diffraction data sets were collected at the wavelength corresponding to the experimentally measured peak of f″. The phasing potential of each data set was assessed with the HKL2MAP interface[20] to the SHELXC/D/E programs [21,22]. A 2.65 Å resolution data set with good anomalous signal was identified, from which a 56-atom selenium constellation was deduced. This constellation was input to PHENIX AutoSol[23] for single-wavelength anomalous diffraction phasing, density modification, and automated building. After density modification, the figure of merit improved from 0.42 to 0.77 for all reflections to 2.65 Å resolution. The model from automated building consisted of 1956 residues, 1778 of which were assigned to sequence. This model was adjusted and extended manually in COOT [24], aided by the density-modified single-wavelength anomalous diffraction map, and subsequently refined against a 2.35 Å resolution native data set using PHENIX[25]. Several additional rounds of building and refinement were performed. The resulting structure provided the starting coordinates for the refinements of the other structures reported here. Data collection and refinement statistics and Protein Data Bank (PDB) deposition codes are listed in Table 2.2.

2.2.4 Small angle X-ray Scattering

SAXS experiments were performed at the SIBYLS beamline (12.3.1) of the Advanced Light Source[26]. Prior to analysis, a sample of AfUGM at 9.4 mg/ml was dialyzed into a buffer of 20 mM HEPES, 45 mM NaCl, 0.5 mM Tris(hydroxypropyl)phosphine at pH 7.5. Scattering intensities (I) were measured at three

nominal protein concentrations to ensure concentration-independent scattering. Exposures of 0.5, 1.0, and 5.0 s were used to check for radiation damage. The scattering curves collected from the protein sample were corrected for background scattering using intensity data collected from the dialysis buffer. A composite scattering curve was generated with PRIMUS[27] by scaling and merging the background corrected high q region data from the 5.0-s exposure with the low q region data from the 0.5-s exposure. The pair distribution function was calculated with GNOM [28]. GASBOR [29] was used to calculate shape reconstructions, and DAMAVER [30] was used to average and filter the resulting dummy atom models. The Situs module pdb2vol was used to convert the averaged, filtered models into volumetric maps [31]. SUPCOMB was used to superimpose the crystallographic tetramer onto the dummy atom model [32].

2.2.4 Kinetics

The mutase activity was tested with UDP-Galf as the substrate following procedures previously described [14]. The enzyme concentration was determined using the flavin extinction coefficient at 450 nm of 10.6 $mM^{-1}cm^{-1}$ [14].

2.2.5 Structure Analysis

SSM [33] was used for structure superposition. The PDBePISA server was used to analyze protein interfaces in crystal lattices [34]. COOT and PyMOL [35] were used to analyze noncovalent interactions.

Table 2.2 X-ray Diffraction Data Collection and Refinement statistics.

	Se-Met	**Sulfate complex**	**AfUGM$_r$**	**AfUGM$_r$-UDP**	**AfUGM$_r$-UDP-Gal*p***
Beamline	APS 19ID	APS 19ID	ALS 8.2.2	ALS 8.2.2	APS 24-ID-C
Active site ligand	Sulfate	sulfate	None	UDP	UDP-Gal*p*
Space group	$P6_522$	$P6_522$	$P6_522$	$P6_522$	$P6_522$
Unit cell parameters (Å)	a= 218.1, c= 319.1	a=217.8, c=319.7	a= 217.9, c= 322.8	a= 218.2, c= 320.8	a= 219.1, c= 322.4
Wavelength	0.97915	0.97915	1.00000	1.00000	0.97949
Resolution (Å)	55.00-2.65 (2.74-2.65)	50.0-2.35 (2.43-2.35)	48.6-2.20 (2.37-2.25)	48.6-2.25 (2.37-2.25)	48.8-2.25 (2.37-2.25)
Observations	1,833,668	900,659	1,560,709	1,036,765	1,670,151
Unique reflections	128,441	167,455	211,192	205,755	207,614
$R_{merge}(I)$	0.087 (0.351)	0.086 (0.568)	0.097 (0.599)	0.086 (0.676)	0.097 (0.599)
$R_{meas}(I)$			0.104 (0.643)	0.096 (0.753)	0.104 (0.645)
$R_{pim}(I)$			0.038 (0.231)	0.041 (0.323)	0.035 (0.232)
Mean I/σ	32.3 (5.5)	20.1 (3.0)	16.2 (3.5)	17.6 (2.8)	13.9 (3.4)
Completeness (%)	99.0 (99.6)	90.5 (91.1)	99.8 (100.0)	97.9 (99.7)	97.3 (97.6)
Multiplicity	14.3 (12.7)	5.4 (5.4)	7.4 (7.4)	5.0 (5.0)	8.0 (7.0)
No. of protein chains		4	4	4	4
No. of protein residues		2018	2020	2019	2019
No. of protein atoms		15597	15580	15551	15562
No. of FAD atoms		212	212	212	212
No. of ligand atoms		0	0	50	72
No. of water molecules		516	813	657	659
R_{cryst}		0.188 (0.249)	0.190 (0.244)	0.192 (0.260)	0.185 (0.249)
R_{free}[b]		0.218 (0.297)	0.215 (0.281)	0.218 (0.297)	0.210 (0.294)
rmsd bond lengths, Å[c]		0.007	0.007	0.007	0.007
rmsd bond angles, °		1.06	1.04	1.07	1.08
Ramachandran plot[d]					
Favored (no. residues)		1985	1983	1979	1977
Allowed (no. residues)		25	29	32	34
Outliers (no. residues)		0	0	0	0
Average B-factor (Å^2)					
Protein		35	30	33	35
FAD		32	25	30	31
Active site ligand		56		26	39
Water		34	34	34	37
Coordinate error (Å)[e]		0.33	0.32	0.32	0.32
PDB code		3UTE	3UTF	3UTG	3UTH

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]A common set of test reflections (5 %) was used for refinement of all structures.
[c]Compared to the parameters of Engh and Huber[46].
[e]The Ramachandran plot was generated with RAMPAGE[47].
[f]Maximum likelihood-based coordinate error estimate.

2.3 Results

2.3.1 Overall Fold and Flavin Binding Site

The AfUGM protomer has a mixed α/β-fold (Fig. 2.2) and exhibits the same three domain architecture that was first identified in the structure of Escherichia coli UGM [36]. Domain 1 is the largest of the three domains. It is a tripartite unit consisting of residues 3–90, 205–291, and 421–507 and includes a Rossmann fold core that binds the FAD. Domain 2 is a bundle of α-helices (residues 105– 204). This domain mediates one of the major interfaces of the tetramer. Domain 3 (residues 91–104, 292–420) is situated between the other two domains and features a twisted, seven-stranded β-sheet. This domain participates in substrate binding. The FAD isoalloxazine binds in a crevice between domains 1 and 3.

AfUGM is about 100 residues longer than bacterial UGMs, and the extra residues form several structural elements that appear to be important for catalysis and oligomerization (Fig. 2.2A, red sections). The first of these extra elements is a fourstranded β-sheet in domain 1 formed by residues 6–8 and 243–262. This β-sheet is located near the adenine of the FAD. Domain 2 has an extra helix (residues 188–197) not found in bacterial UGMs. This helix is situated between two mobile flaps that close and open in response to substrate binding and product release, respectively (Fig. 2.2A). Also, the second helix of the domain (residues 115–134) is about seven residues longer in AfUGM and rotated by about 90° from the corresponding helix in bacterial UGMs. The

115–134 and 188–197 helices form a four-helix bundle in one of the dimer interfaces of the UGM tetramer (see below).

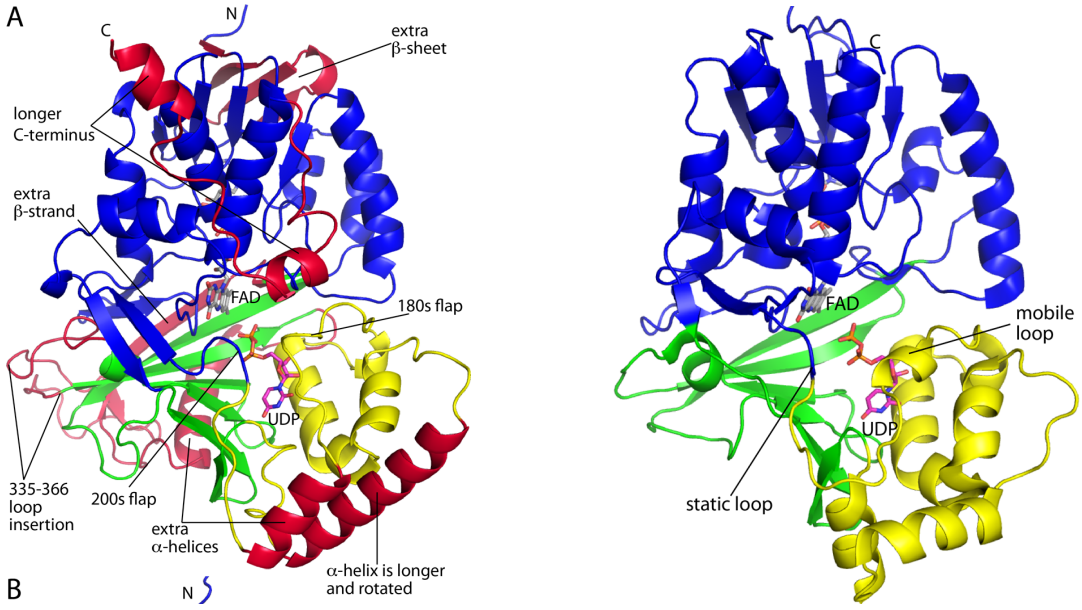

Fig. 2.2 Overall fold of AfUGM. (A) Structure of the AfUGM protomer. Domains 1, 2, and 3 are colored blue, yellow, and green, respectively. Features that distinguish AfUGM from bacterial UGMs are colored red. (B) Protomer structure of a bacterial UGM (*Deinococcus radiodurans* UGM, PDB code 3HE3). This figure and others were prepared with PyMol.

Domain 3 of AfUGM has two large inserts when compared with bacterial UGMs. The first is a long loop formed by residues 335–366. The second consists of residues 378–418, which fold into an α-helix followed by a β-strand. Because of the latter secondary structure element, the sheet of domain 3 has seven strands rather than six as in bacterial UGMs. Finally, AfUGM has ~30 extra residues at the C terminus. These residues form a U-shaped substructure that packs against domain 1 and features two α-helices. This extra structural element is involved in oligomerization.

2.3.2 Quaternary Structure

SAXS was used to obtain solution structural information for AfUGM. A composite scattering curve is shown in Fig. 2.3A. The associated Guinier plot exhibits good linearity (R^2 = 0.9996) and suggests a radius of gyration of 47.3 Å (Fig. 2.3A, inset). Calculations of the pair distribution function yield a R_g of 47.6 Å and maximum particle dimension of 140 Å (Fig. 2.3B). For reference, the R_g of the crystallographic protomer is only 24 Å. These results suggest that AfUGM self-associates in solution.

The SAXS data were further analyzed to determine the oligomeric state and quaternary structure. The experimental scattering profile was compared with theoretical profiles calculated from models of AfUGM oligomers. These models were obtained by analyzing the protein interfaces in the crystal lattice to identify potentially stable assemblies. The calculations revealed an octamer with Rg of 52 Å, a tetramer exhibiting point group 222 symmetry and having an R_g of 47 Å, and two symmetric dimers with Rg values of 34 Å (OP dimer) and 43 Å (OQ dimer). The R_g data suggest that AfUGM forms a tetramer. Indeed, only the profile calculated from the tetramer shows good agreement with the experimental curve (Fig. 2.3C).

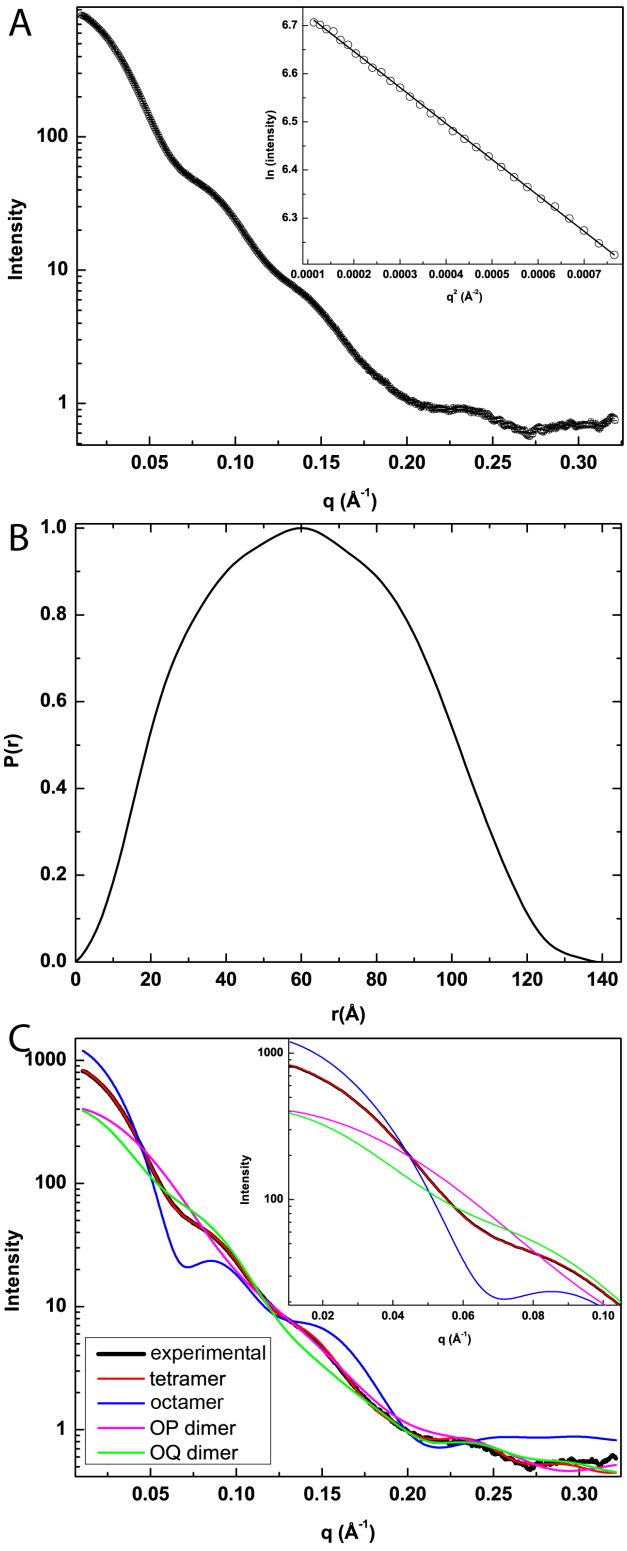

Fig. 2.3 SAXS analysis. (A) SAXS curve and Guinier plot (inset). The range of the Guinier plot corresponds to qR_g = 0.503 - 1.31. (B) Pair distribution function. (C) Comparison of the experimental SAXS curve (black) with theoretical curves calculated from various oligomeric models of AfUGM. The inset highlights the fits in the low q region (q < 0.1 Å^{-1}).

Note in particular that the agreement is exceptional for the low q region ($q < 0.1$, Fig. 2.3C, inset). These results are consistent with size exclusion chromatography data, which suggested that AfUGM forms a tetramer in solution[14]. Finally, shape reconstruction calculations performed with GASBOR assuming a 222 tetramer yield an envelope that is in good agreement with the crystallographic tetramer (Fig. 2.4). It is concluded that the tetramer identified in the crystal lattice represents the solution structure of AfUGM. The AfUGM tetramer is a dimer of dimers with point group 222 symmetry (Fig. 2.4). The symmetry of the tetramer can be described with three mutually orthogonal 2- fold axes, denoted P, Q, and R that intersect at the center of the tetramer. The P-, Q-, and R-axes relate protomer O to protomers P, Q, and R, respectively. The axes define three potential dimer interfaces corresponding to protomer O contacting the other three protomers. Only the OP and OQ interfaces are significant.

The OP interface buries 5000 $Å^2$ of area and is primarily hydrophilic, featuring 10 intersubunit hydrogen bonds. The interface is noncontiguous. The major part of the interface consists of the C-terminal residues 468–502 of two P-related chains interacting across the P-axis at the distal ends of the tetramer. Two such interfaces are located at opposite ends of the tetramer (Fig. 2.4). The minor part of the OP interface is located at the intersection of the two-fold axes and involves just the guanidinium of Arg-133 forming a hydrogen bond with the carbonyl of Val-134 of the P-related chain (Fig. 2.4A, boxed inset). Thus, there are four of these intersubunit hydrogen bonds located in the center of the tetramer.

The OQ interface is located in the middle of the tetramer and buries 4600 $Å^2$ of contiguous surface area. Two helices (residues 115–134 and 188–197) from domain 2 of

Q-related chains interact to form a four-helix bundle oriented parallel to the Q-axis (Fig. 2.4). Note that these two helices are unique to AfUGM (Fig. 2.4 A). The tetramer has two of these bundles. The four-helix bundle has a hydrophobic interior featuring Val, Leu, and Ile side chains.

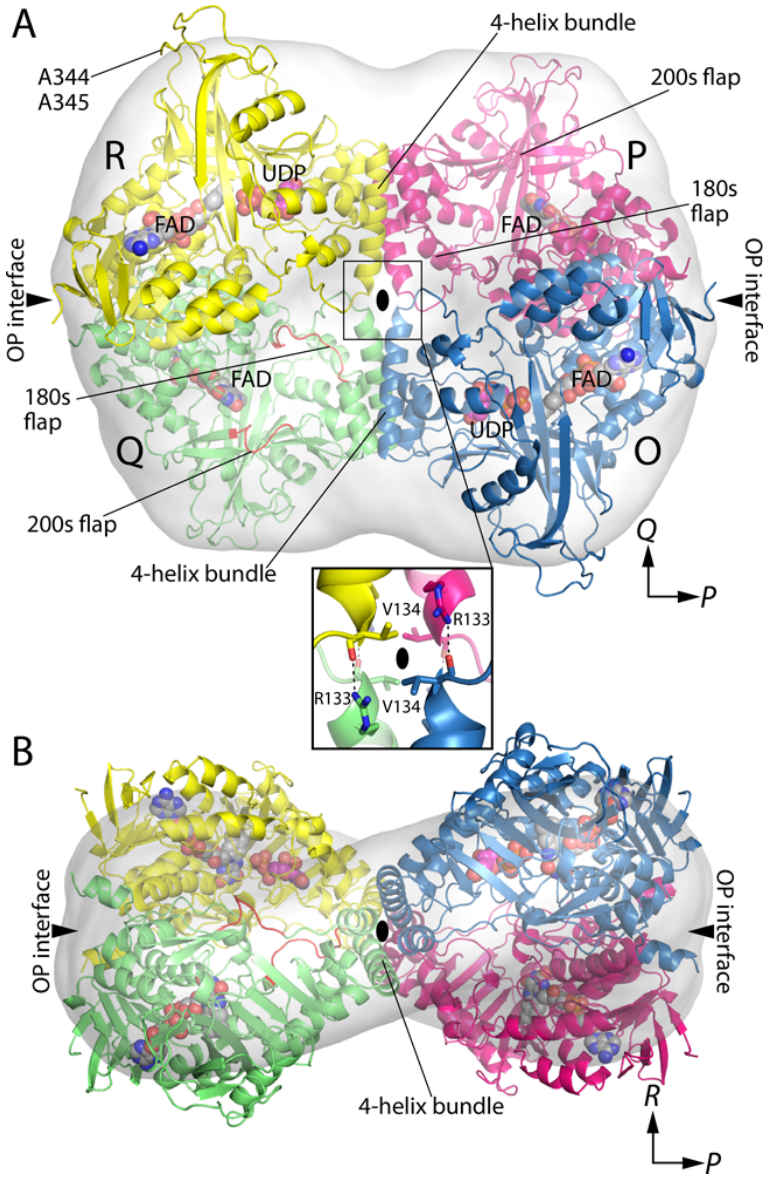

Fig. 2.4 Quaternary structure of AfUGM. (A) The tetramer is viewed down the *R* molecular 2-fold axis. Each chain has a different color. The active site flaps of chain Q are colored red. The surface represents the SAXS shape reconstruction. Inset: Intersubunit hydrogen bonds at the intersection of molecular 2-fold axes. Chains related by the *P*-axis engage in hydrogen bonding. The oval represents the *R* two-fold molecular axis. (B) The tetramer is viewed down the *Q*-axis.

2.3.3 Active Site of Sulfate Complex

The crystals used for structure determination were grown in 1.5 M ammonium sulfate, and as crystallized, two sulfate ions are bound in the active site (Fig. 2.5). The ions bind on the re side of the isoalloxazine. One of the ions binds next to the center ring of the isoalloxazine, which is the binding site for the Galp moiety of the substrate. His-63, an absolutely conserved residue in UGMs, is likewise located on the re face and interacts with both sulfate ions. The location of the conserved histidine at the re face is unprecedented and likely a crystallization artifact.

The isoalloxazine is planar, which is consistent with the FAD being oxidized. It is noted that the crystals are yellow, which is also indicative of the oxidized state.

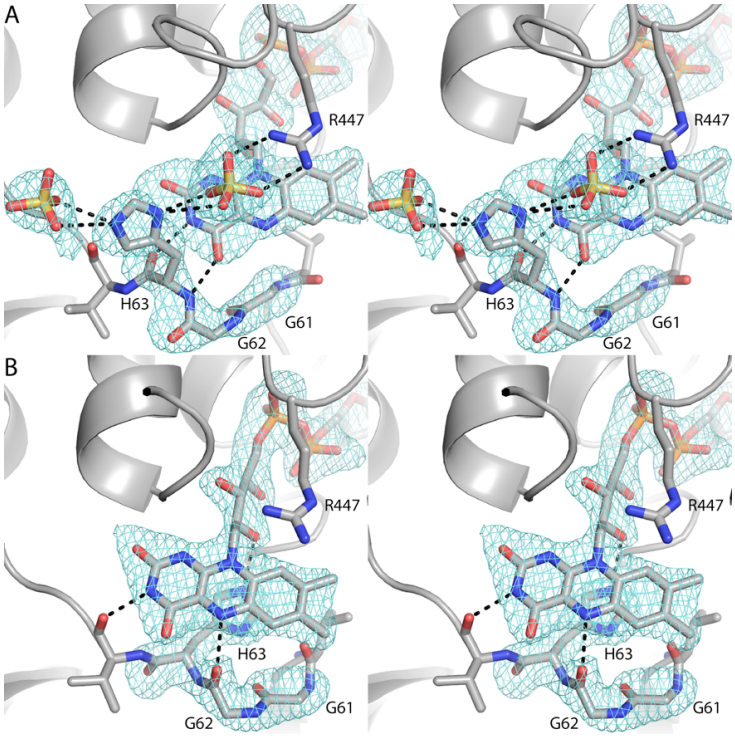

Fig. 2.5 Stereographic views of histidine loop in the (A) sulfate complex and (B) AfUGMr. In both panels, the cage represents a simulated annealing **σ**A-weighted **F**o - **F**c omit map contoured at 3.0 **σ**.

2.3.4 Histidine Loop of Reduced AfUGM

The structure of the reduced enzyme ($AfUGM_r$) was determined from a crystal that was soaked in 80 mM sodium dithionite. The crystal changed from yellow to colorless, which indicated that the FAD was reduced.

Soaking with dithionite causes significant changes in the active site, which occur in all four chains. Electron density features for the sulfate ions are absent, and the conformation of the histidine loop is dramatically different (Fig. 2.5B). In particular, His-63 has moved to the *si* face of the FAD, where it stacks in parallel with the middle ring of the isoalloxazine and forms a hydrogen bond with the FAD 2′-OH. Furthermore, the carbonyl of Gly-62 forms a hydrogen bond with the FAD N5. This conformational change in the protein is accompanied by a 3 Å shift in the FMN part of the FAD. This shift brings the N1 atom of the FAD within hydrogen-bonding distance of the backbone N-H group of Gln-458, which is consistent with the FAD adopting the anionic hydroquinone state.

The conformation of the histidine loop in $AfUGM_r$ is very similar to those of bacterial UGMs. In particular, in all other structures, the conserved histidine is located at the *si* face as in $AfUGM_r$. Furthermore, in all other UGMs, the carbonyl of the residue preceding the conserved His accepts a hydrogen bond from the N5 of reduced FAD. The carbonyl-N5 interaction is important for stabilizing the reduced state of FAD, a necessity for activity. As mentioned, Gly-62 forms this critical hydrogen bond in $AfUGM_r$. It is concluded that the dithionite-reduced AfUGM structure represents the active, substrate-free form of the enzyme.

2.3.5 FAD Conformation and Binding Site

The FAD of $AfUGM_r$ binds primarily to domain 1 (Fig. 2.2A). The conformation of the FAD is nearly identical to those of bacterial UGMs, and the flavin-protein interactions are highly conserved (Fig. 2.6). In particular, there are several interactions with the protein backbone (Fig. 2.6B); these are conserved, as one would expect. Asp-38 forms two hydrogen bonds with the ribose. This interaction appears to be universal in UGM. Some of the side chain interactions with the FAD in AfUGM are found in other bacterial UGMs. For example, Ser-461 and Thr-18, which interact with the ribityl 3′-OH and pyrophosphate, respectively, are analogous to Thr-355 and Ser-14 of *Klebsiella pneumoniae* UGM (KpUGM).

The FAD of $AfUGM_r$ exhibits a butterfly-like distortion of the isoalloxazine in which the pyrimidine ring bends 7° out of the plane such that the *si* face is concave (Fig. 2.6C). The direction of the bending is the same as that of *Deinococcus radiodurans* UGM (DrUGM) [37]and opposite to that of KpUGM[38]. The angle is close to that of DrUGM and about half that of KpUGM (~13°).

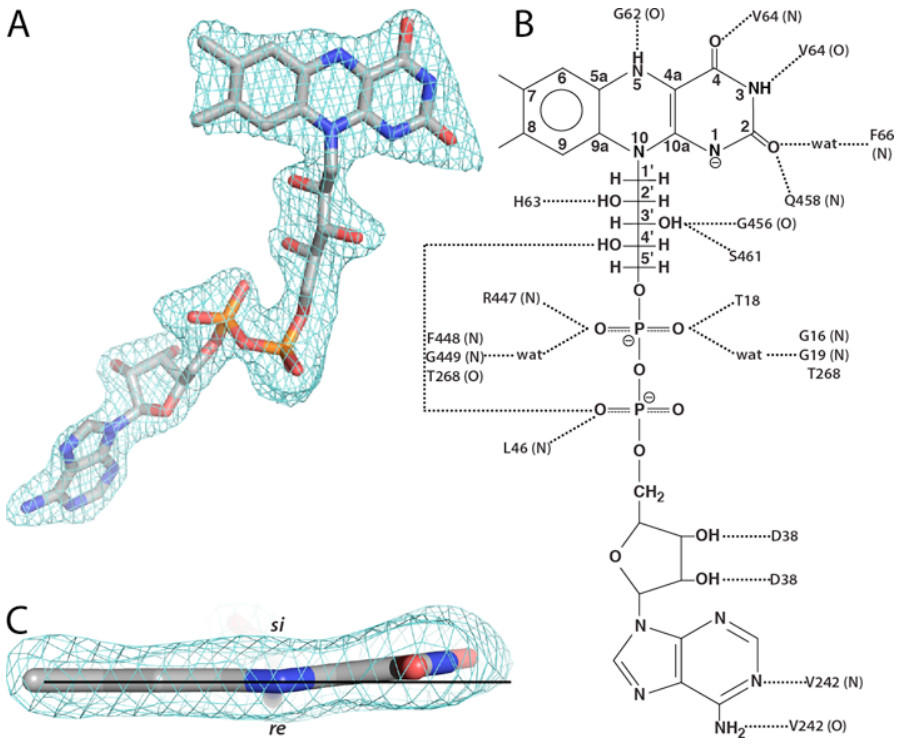

Fig. 2.6 Electron density and interactions for the FAD in AfUGMr. (A) View of the FAD oriented with the *re* side of the isoalloxazine system facing the viewer. The cage represents a simulated annealing **σ**A weighted Fo - Fc omit map contoured at 3.0 **σ**. (B) Schematic diagram of protein-FAD interactions in AfUGMr. Backbone interactions are indicated by N or O in parentheses. (C) Edge-on view of the isoalloxazine. The horizontal line assists in seeing the 7° butterfly-like bend angle.

2.3.6 Structure of AfUGM$_r$ Complexed with UDP

The structure of AfUGM$_r$ complexed with UDP was determined from a crystal that had been soaked in sodium dithionite and UDP. Electron density maps clearly indicated that UDP is bound in two of the four active sites of the tetramer (Fig. 2.7). The occupancy of UDP is estimated to be 0.9, which indicates essentially full occupancy and thus tight binding.

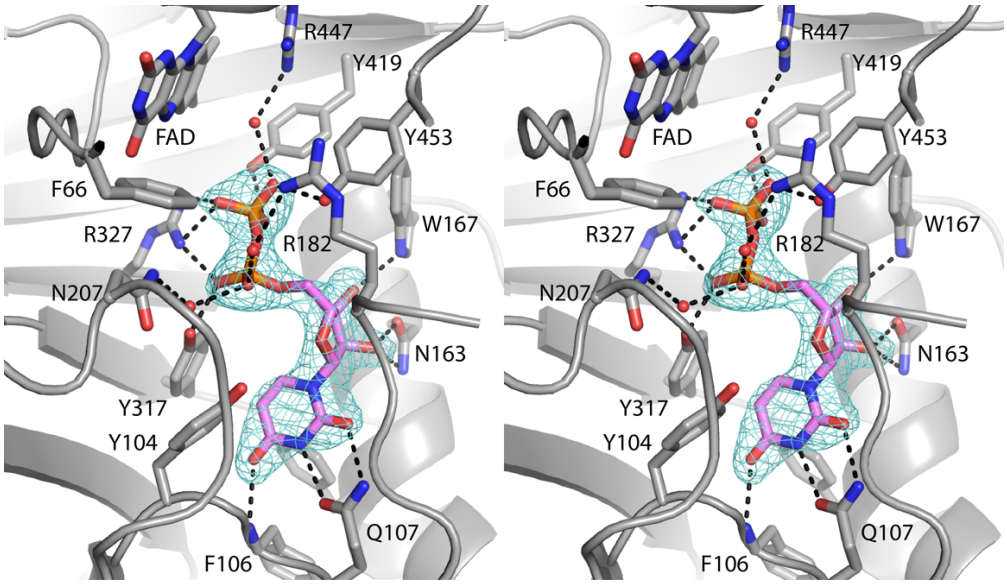

Fig. 2.7 Electron density maps UDP bound to AfUGM$_r$ (stereographic view). The cage represent a simulated annealing σ_A-weighted F_o - F_c omit maps contoured at 3.0 σ.

The binding of UDP induces profound conformational changes in the enzyme (Fig. 2.8). The largest of these changes involves residues 179–187 and 203–209. In the absence of UDP, these loops reside on the periphery of the enzyme, and thus, the active site is open. Upon binding, the loops move toward each other like the flaps of a box top to create a closed active site. The conformational changes are substantial. For example, the middle residues of the two loops, Val-183 and Pro- 206, move by 11 and 13 Å, respectively, upon ligand binding. One consequence of the flaps folding inward is that the side chain of Phe-66 swings into the active site to avoid a clash with Pro-206 (Fig. 2.8). As shown below, Phe-66 contacts the Galp moiety of the substrate.

Closing of the flaps is accompanied by smaller conformational changes in residues 104–107 and 146–161 that fine-tune the uridine binding pocket. Tyr- 104 and Phe-158 move apart from each other by 3 Å to create space for the incoming uracil. Phe-106 and Gln-107 shift by 1.0 and 2.0 Å, respectively, to form three hydrogen bonds with

the base (interactions are shown in Fig. 2.7). Finally, movements of ~1.0 Å bring Asn-163, Trp-167, Tyr-317, Arg-327, and Tyr-453 into contact distance of the ribose and pyrophosphate (Fig. 2.7).

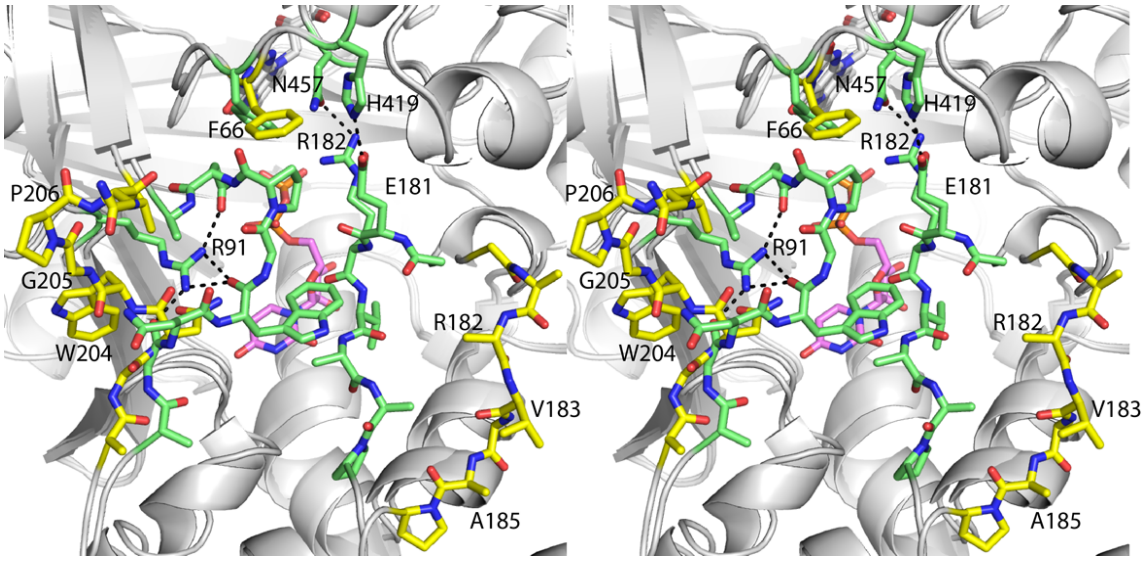

Fig. 2.8 Comparison of the open and closed forms of AfUGM$_r$ (stereographic view). AfUGM$_r$ is shown in yellow. The AfUGM$_r$-UDP complex structure is shown in green. UDP is colored pink. Selected interactions in the closed state are indicated by dashed lines.

The dramatic closing of the flaps also results in the formation of new protein-protein interactions that stabilize the closed active site conformation (Fig. 2.8). In particular, Glu-181 and Arg-182 of the 180s flap interact with His-68 and Asn- 457, respectively, in the closed state, whereas in the open state, the interacting partners are separated by 17–19 Å. Another interaction that stabilizes the closed state involves Arg-91, whose side chain moves to make four hydrogen bonds with the 200s flap.

2.3.7 Structure of AfUGM$_r$ Complexed with UDP-Galp

The structure of AfUGM$_r$ complexed with the substrate UDP-Galp was

determined from a crystal that was soaked in sodium dithionite and UDP-Galp. As with the UDP complex, electron density maps indicated binding in two of the four active sites (Fig. 2.8A). The occupancy of UDP-Galp is estimated to be 0.8. Electron density is weak for the O5 atom of the Galp moiety, suggesting the possibility of conformational disorder in this part of the ring. We note that conformational disorder of the Galp moiety has been observed previously[38] (PDB code 3INT). As seen with UDP, the substrate induces closure of both flaps.

The conformation and interactions of the UDP moiety of the substrate are essentially identical to those of the AfUGM$_r$-UDP complex, so we will focus on the Galp moiety. Galp binds at the re face of the FAD isoalloxazine with the anomeric carbon atom (C1) poised 3.4 Å from the FAD N5 atom (Fig. 2.9A, red dashes). The hydroxyl groups of Galp make direct hydrogen bonds with Arg-182, Asn-207, and Asn-457 (Fig. 2.9B). Water-mediated interactions are formed with Thr-329, Tyr-334, and Arg-447. Also, the O4 hydroxyl forms a hydrogen bond with the FAD O4 carbonyl. In addition to these electrostatic interactions, nonpolar contacts are formed with Phe-66 and Trp-315. These steric interactions presumably help enforce specificity for Galp and orient the carbohydrate for catalysis.

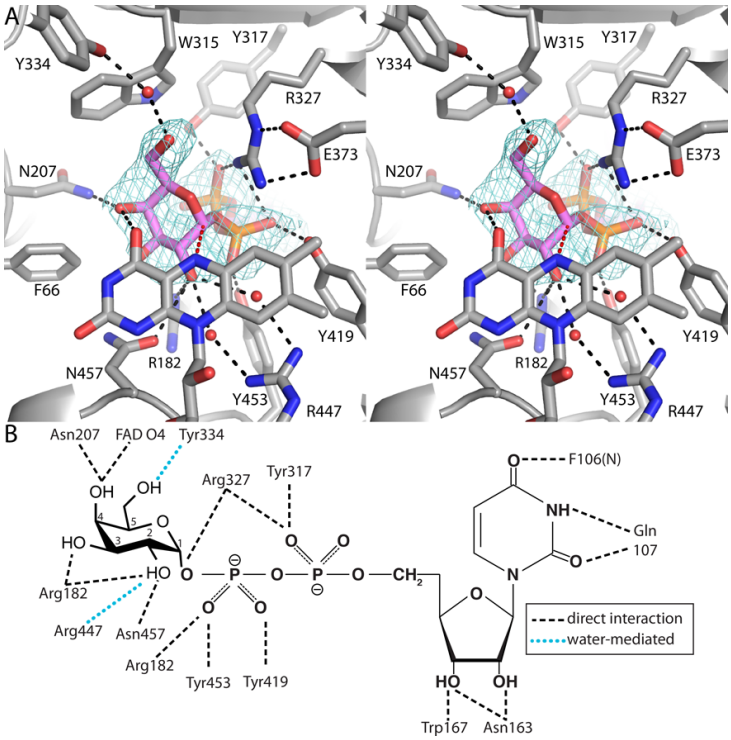

Fig. 2.9 Electron density map (stereographic view) and interaction diagram for UDP-Gal*p* bound to AfUGM$_r$. (A) The cage represent a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 2.0 σ. The red dashes in panel indicate the 3.4 Å separation between the anomeric C atom of Gal*p* and the FAD N5. (B) Schematic diagram of noncovalent electrostatic interactions.

2.3.8 Comparison of UDP-Galp Interactions in AfUGM and Bacterial UGMs

The structures of the UDP-Galp complexes of AfUGM, KpUGM, and DrUGM were compared to identify similarities and differences in substrate recognition (Fig. 2.10). In all three structures, the substrate adopts an extended conformation, the O4 hydroxyl of Galp forms a hydrogen bond to the FAD O4, and the anomeric carbon atom is near the FAD N5. Several conserved residues participate in substrate recognition, including Arg-182, Arg-327, Trp-167, Tyr- 419, and Tyr-453 (AfUGM numbering).

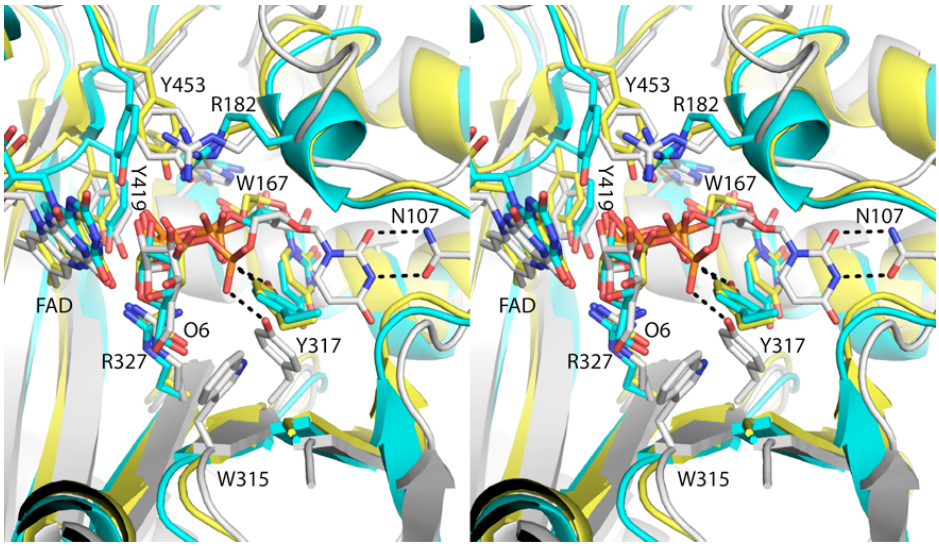

Fig. 2.10 UDP-Galp recognition by eukaryotic and bacterial UGMs (stereographic view). The structuresof AfUGM (white), KpUGM (cyan, 3INT), and DrUGM (yellow, 3HDY) complexed with UDP-Galp are shown. Selected residues of AfUGM are noted.

Differences between the two classes of UGM are also evident (Fig. 2.10). In particular, the UMP moiety of AfUGM is displaced by 3–5 Å when compared with the bacterial enzymes. The uracil ring is shifted by ~4 Å and rotated by almost 90°. This difference allows the uracil to form two hydrogen bonds with Gln-107, a residue without a counterpart in bacterial UGMs. Displacement of the UMP also reflects a different arrangement of Tyr residues that interact with the pyrophosphate (Tyr-317 in AfUGM, Tyr-185 in KpUGM, Tyr-209 in DrUGM). Finally, the O6 hydroxyl of Galp in AfUGM is rotated by 110° from that in the bacterial enzymes. This difference is due to the potential for steric clash with Trp-315.

2.4 Discussion

The core structure of UGM appears to be conserved across the bacterial and eukaryotic domains of life despite low sequence identity (~16%). The root mean square

deviation between AfUGM and bacterial UGMs is 2.4–2.7 Å over 279–322 residues. Also, the majority of the FAD-protein interactions are conserved.

Curiously, protoporphyrinogen oxidase from *Myxococcus xanthus* (PPOX) [39], rather than another UGM, is the closest structural homolog of AfUGM in the PDB (Fig. 2.11). PPOX has FAD binding, membrane binding, and substrate binding domains, which correspond to UGM domains 1, 2, and 3, respectively. PPOX (PDB code 2IVD) and AfUGM superimpose with a root mean square deviation of 2.5 Å covering 370 residues, despite sharing only 15% sequence identity. Furthermore, PPOX has some of the extra structural features of AfUGM that are not found in bacterial UGMs, including the inserted β-sheet in domain 1 and the extra helix and strand of domain 3. The significance of the structural homology to PPOX is not obvious.

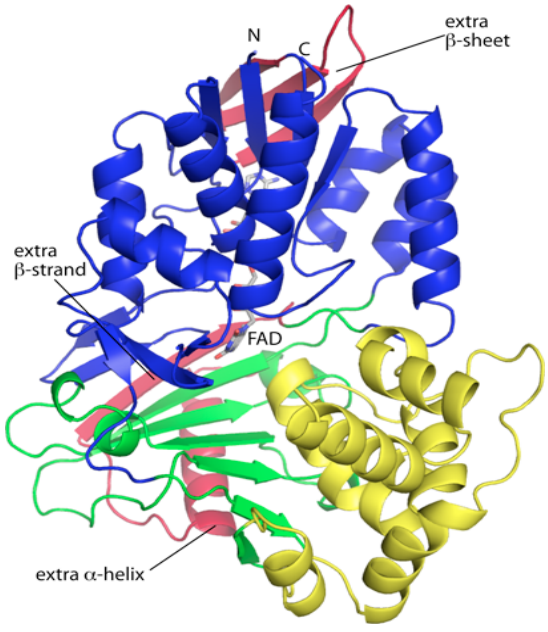

Fig. 2.11 Protoporphyrinogen oxidase from Myxococcus xanthus (PDB code 2IVD). The FAD-binding (blue), membrane-binding (yellow), and substrate-binding (green) domains correspond to UGM domains 1, 2, and 3, respectively. Red indicates elements that are common to AfUGM and PPOX but not found in bacterial UGMs.

Neither oligomeric state nor quaternary structure is conserved in UGM. Dimers are common for bacterial UGMs, whereas AfUGM forms a tetramer in solution. UGM from E. coli was shown to be dimeric in solution based on dynamic light scattering[40], and the crystal structure revealed a semicircular dimer that likely represents the dimer in solution[36]. This dimer is also found in the crystal lattices of KpUGM and *Mycobacterium tuberculosis* UGM[41]. DrUGM appears to have a different quaternary structure, although this has not been confirmed with solution studies. Analysis of the interfaces in the primitive orthorhombic lattice of DrUGM suggests that the enzyme forms a decamer in solution [37]. The decamer is a pentamer of dimers in which the propagated dimer is unlike the semicircular dimer of the other bacterial UGMs. The DrUGM dimer is formed by interactions involving domains 1 and 3. The constituent dimers of the AfUGM tetramer are unlike any of the dimers of bacterial UGMs.

Unique structural features of AfUGM preclude the assembly of the dimers formed by bacterial UGMs. For example, if the classic UGM semicircular dimer is constructed from AfUGM protomers, several clashes across the dimer interface are evident. These clashes involve the 200s flap and the extra helix of domain 2, both of which are unique to AfUGM. Likewise, a hypothetical dimer built like DrUGM exhibits clashes in the four-stranded sheet in domain 1 and the long inserted loop in domain 3. Thus, the protomer structure of AfUGM is incompatible with the quaternary structures of bacterial UGMs.

The observation of a tetramer raises the possibility of cooperativity between the subunits. Interestingly, previous studies have shown that 50% of the bound flavin remains reduced under normal laboratory conditions (absence of reducing agent), which suggests nonequivalence of the subunits [14]. One possibility is that electrostatic repulsion between

reduced flavins in the tetramer could contribute to this phenomenon. FADs related by the P-axis are 53 Å apart, whereas those related by the Q- and R-axes are separated by 75 and 79 Å, respectively. If electrostatic repulsion is important over the ~50 Å scale, these distances suggest that protomers related by the R-axis, i.e. those along the diagonal of the tetramer, are equivalent. Interestingly, crystal soaking populated only the diagonal protomers. Future studies of cooperativity in AfUGM seem warranted.

The quaternary structure of AfUGM enables conformational changes that distinguish AfUGM from bacterial UGMs. AfUGM has two flaps that move substantially upon substrate binding and product release. Bacterial UGMs have just one mobile loop that enters the active site upon substrate binding (Fig. 2.2B, mobile loop) [37, 38]. This loop is analogous to the 180s flap of AfUGM. The region of bacterial UGMs corresponding to the 200s flap also participates in substrate binding (Fig. 2.2B, static loop). In particular, Tyr-185 and Phe-186 of KpUGM (Tyr-209 and Phe-210 of DrUGM) contact the pyrophosphate and Galp moiety of UDP-Galp, respectively. However, these residues are static because the loop is in a protein-protein interface, and a neighboring protein molecule of the oligomer restricts the motion of the loop, effectively locking it in the closed conformation. In contrast, the 200s flap of AfUGM resides on the surface of the tetramer (Fig. 2.4A) and is able to adopt open and closed conformations. Thus, tetramerization of AfUGM allows an additional degree of conformation freedom.

It is notable that very few of the flap residues directly interact with the substrate. (Exceptions are Arg-182 and Asn-207.) This observation suggests that the purpose of flap closure is not only to assemble the constellation of residues needed for substrate recognition but also to create a protected environment for catalysis. The latter role is

consistent with mechanisms in which UDP is displaced during the catalytic cycle[1] because closure of the flaps presumably prevents the severed UDP from migrating out of the active site.

Finally, the AfUGM structures provide additional insight into the catalytic mechanism of UGM. Two prevailing mechanisms have been proposed based on studies of the bacterial enzymes[1]. One mechanism begins with the flavin functioning as a nucleophile that attacks the anomeric carbon of galactose to form a covalent intermediate and displace UDP[42]. The other mechanism involves single electron transfer from the reduced flavin to an oxocarbenium sugar intermediate followed by the formation of a flavin-sugar adduct [43, 44]. Which of these mechanisms, if either, applies to eukaryotic UGMs is unknown. However, the Galp moiety in the $AfUGM_r$-UDP-Galp complex is poised for nucleophilic attack by the FAD N5 at the anomeric carbon atom (Fig. 2.9A). Also, the direction of the bend of the FAD isoalloxazine in AfUGM is consistent with formation of a covalent intermediate [41]. For bacterial UGMs, crystal structures, reductive trapping experiments, and studies with modified flavins support the covalent mechanism [37, 38, 44]. Thus, it seems likely that bacterial and eukaryotic UGMs share a common catalytic mechanism.

2.5 References

1. Tefsen, B., Ram, A. F., van Die, I., and Routier, F. H. (2011) *Glycobiology*, in press

2. Pan, F., Jackson, M., Ma, Y., and McNeil, M. (2001) *J Bacteriol* **183**, 3991-3998

3. Kleczka, B., Lamerz, A. C., van Zandbergen, G., Wenzel, A., Gerardy-Schahn, R., Wiese, M., and Routier, F. H. (2007) *J. Biol. Chem.* **282**, 10498-10505

4. Kradin, R. L., and Mark, E. J. (2008) *Arch Pathol Lab Med* **132**, 606-614

5. Virnig, C., and Bush, R. K. (2007) *Curr Opin Pulm Med* **13**, 67-71

6. Chong, S., Lee, K. S., Yi, C. A., Chung, M. J., Kim, T. S., and Han, J. (2006) *Eur J Radiol* **59**, 371-383

7. Samarakoon, P., and Soubani, A. O. (2008) *Chron Respir Dis* **5**, 19-27

8. Lin, S. J., Schranz, J., and Teutsch, S. M. (2001) *Clin Infect Dis* **32**, 358-366

9. Knutsen, A. P., and Slavin, R. G. (2011) *Clin Dev Immunol*, in press

10. Latge, J. P. (2009) *Med Mycol* **47 Suppl 1**, S104-109

11. Schmalhorst, P. S., Krappmann, S., Vervecken, W., Rohde, M., Muller, M., Braus, G. H., Contreras, R., Braun, A., Bakker, H., and Routier, F. H. (2008) *Eukaryot Cell* **7**, 1268-1277

12. Lamarre, C., Beau, R., Balloy, V., Fontaine, T., Wong Sak Hoi, J., Guadagnini, S., Berkova, N., Chignard, M., Beauvais, A., and Latge, J. P. (2009) *Cell Microbiol* **11**, 1612-1623

13. Oppenheimer, M., Poulin, M. B., Lowary, T. L., Helm, R. F., and Sobrado, P. (2010) *Arch Biochem Biophys* **502**, 31-38

14. Doublie, S. (1997) *Methods Enzymol.* **276**, 523-530

15. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr., Sect. D* **66**, 213-221

16. Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J., and Kelley, L. A. (2008) *Proteins* **70**, 611-625

17. Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491-497

18. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326

19. Evans, P. (2006) *Acta Cryst.* **D62**, 72-82
20. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) *Acta Cryst.* **D59**, 1131-1137

21. Pape, T., and Schneider, T. R. (2004) *J. Appl. Crystallogr.* **37**, 843-844

22. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772-1779

23. Sheldrick, G. M. (2002) *Z. Kristallogr.* **217**, 644-650

24. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Cryst.* **D58**, 1948-1954

25. Emsley, P., and Cowtan, K. (2004) *Acta Cryst.* **D60**, 2126-2132

26. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W., and Tainer, J. A. (2009) *Nat. Methods* **6**, 606-612

27. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* **36**, 1277-1282

28. Svergun, D. I. (1992) *J. Appl. Crystallogr.* **25**, 495-503

29. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) *Biophys. J.* **80**, 2946-2953

30. Volkov, V. V., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* **36**, 860-864

31. Wriggers, W. (2010) *Biophys. Rev.* **2**, 21-27

32. Kozin, M. B., and Svergun, D. I. (2001) *J. Appl. Crystallogr.* **34**, 33-41

33. Krissinel, E., and Henrick, K. (2004) *Acta Cryst.* **D60**, 2256-2268

34. Krissinel, E., and Henrick, K. (2007) *J. Mol. Biol.* **372**, 774-797

35. DeLano, W. L. (2002) *The PyMOL User's Manual*, DeLano Scientific, Palo Alto, CA, USA

36. Sanders, D. A., Staines, A. G., McMahon, S. A., McNeil, M. R., Whitfield, C., and Naismith, J. H. (2001) *Nat Struct Biol* **8**, 858-863

37. Partha, S. K., van Straaten, K. E., and Sanders, D. A. (2009) *J Mol Biol* **394**, 864-877

38. Gruber, T. D., Westler, W. M., Kiessling, L. L., and Forest, K. T. (2009) *Biochemistry* **48**, 9171-9173

39. Corradi, H. R., Corrigall, A. V., Boix, E., Mohan, C. G., Sturrock, E. D., Meissner, P. N., and Acharya, K. R. (2006) *J Biol Chem* **281**, 38625-38633

40. McMahon, S. A., Leonard, G. A., Buchanan, L. V., Giraud, M. F., and Naismith, J. H. (1999) *Acta Crystallogr D Biol Crystallogr* **55**, 399-402

41. Beis, K., Srikannathasan, V., Liu, H., Fullerton, S. W., Bamford, V. A., Sanders, D. A., Whitfield, C., McNeil, M. R., and Naismith, J. H. (2005) *J Mol Biol* **348**, 971-982

42. Richards, M. R., and Lowary, T. L. (2009) *Chembiochem* **10**, 1920-1938

43. Soltero-Higgin, M., Carlson, E. E., Gruber, T. D., and Kiessling, L. L. (2004) *Nat Struct Mol Biol* **11**, 539-543

44. Fullerton, S. W., Daff, S., Sanders, D. A., Ingledew, W. J., Whitfield, C., Chapman, S. K., and Naismith, J. H. (2003) *Biochemistry* **42**, 2104-2109

45. Huang, Z., Zhang, Q., and Liu, H. W. (2003) *Bioorg Chem* **31**, 494-502

46. Engh, R. A., and Huber, R. (1991) *Acta Cryst*. **A47**, 392-400

47. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50**, 437-45

Chapter 3

Crystal Structures of *Trypanosoma cruzi* UDP-Galactopyranose Mutase Implicate Flexibility of the Histidine Loop in Enzyme Activation

Richa Dhatwalia, Harkewal Singh, Michelle Oppenheimer, Pablo Sobrado, John J. Tanner

Abstract

Chagas disease is a neglected tropical disease caused by the protozoan parasite *Trypanosoma cruzi*. Here we report crystal structures of the galactofuranose biosynthetic enzyme UDP-galactopyranose mutase (UGM) from *T. cruzi*, which are the first structures of this enzyme from a protozoan parasite. UGM is an attractive target for drug design because galactofuranose is absent in humans but is an essential component of key glycoproteins and glycolipids in trypanosomatids. Analysis of the enzyme–UDP noncovalent interactions and sequence alignments suggests that substrate recognition is exquisitely conserved among eukaryotic UGMs and distinct from that of bacterial UGMs. This observation has implications for inhibitor design. Activation of the enzyme via reduction of the FAD induces profound conformational changes, including a 2.3 Å movement of the histidine loop (Gly60-Gly61-His62), rotation and protonation of the imidazole of His62, and cooperative movement of residues located on the *si* face of the FAD. Interestingly, these changes are substantially different from those described for *Aspergillus fumigatus* UGM, which is 45% identical to *T. cruzi* UGM. The importance of Gly61 and His62 for enzymatic activity was studied with the site-directed mutant enzymes G61A, G61P, and H62A. These mutations lower the catalytic efficiency by factors of 10–50, primarily by decreasing k_{cat}. Considered together, the structural, kinetic, and sequence data suggest that the middle Gly of the histidine loop imparts flexibility that is essential for activation of eukaryotic UGMs. Our results provide new information about UGM biochemistry and suggest a unified strategy for designing inhibitors of UGMs from the eukaryotic pathogens.

3.1 Introduction

Chagas disease (aka American trypanosomiasis) is a major global health concern[1]. The disease is caused by the protozoan parasite *Trypanosoma cruzi* and spread by triatomine insects, which carry the parasite in their gut. Chagas disease is one of several neglected tropical diseases, i.e., chronic infectious diseases that are prevalent in poor countries and underemphasized by society relative to the number of people affected. The disease is endemic in Latin America, and at least 10 million people worldwide are estimated to be infected with *T. cruzi*[1]. Infection mainly occurs through contact with the feces of triatomine bugs, but transmission also occurs via blood transfusions, organ transplants, ingestion of contaminated food, and mother-to-child[1]. It is estimated that 30% of Chagas patients will develop heart damage in the late chronic stage of the disease, leading to death caused by arrhythmia in early adulthood[1]. More than 10 000 deaths occur annually from Chagas disease, and the burden of disease is the highest for any parasitic disease in the Western hemisphere [(1,2)]. Chagas is unusual among neglected diseases in that it is spreading to nonendemic areas, including the United States, Canada, and Europe [(2,3)]. The two drugs used for treatment, benznidazole and nifurtimox, have significant side effects and uncertain efficacy [(4,6)], and there is no vaccine available. The prevalence of the disease, the spread to new regions of the globe, and the lack of adequate medications emphasize the need for new drugs to treat Chagas disease.

The flavoenzyme UDP-galactopyranose mutase (UGM) has received attention recently as a drug design target for neglected tropical disease (7,9). UGM plays a central role in the biosynthesis of galactofuranose (Gal*f*) by catalyzing the conversion of UDP-galactopyranose (UDP-Gal*p*) to UDP-galactofuranose (UDP-Gal*f*) (Fig. 3.1). Galf has

never been found in humans but is an essential component of the cell wall and extracellular matrix of many pathogenic bacteria, fungi, and protozoa [8,9].

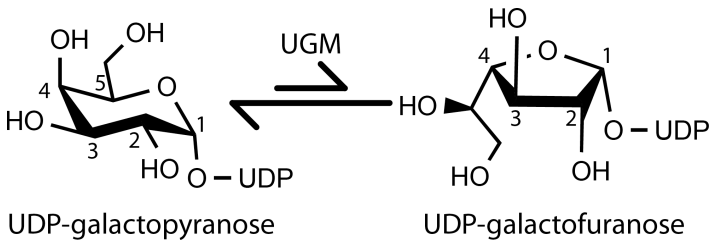

Fig. 3.1 Reaction catalyzed by UGM.

UGM and Gal*f* are widely distributed in pathogenic protozoa [7,8]. In particular, Gal*f* is present in glycoinositolphospholipids and glycosylphosphatidylinositol anchor proteins of *T. cruzi* [10,11]. In the related parasite, *Leishmania major*, which causes leishmaniasis, Gal*f* is present in the membrane anchor of the lipophosphoglycan and in glycoinositolphospholipids[12]. These glycoconjugates are highly expressed throughout the life cycle of these parasites and are important for their survival and proliferation [12,14]. Gal*f*-containing glycoconjugates are thought to be involved in the mechanism of myocardiac invasion by *T. cruzi* [15]. In *Leishmania*, lipophosphoglycans are essential for the binding and detachment of the parasite to the midgut of the vector insect and thus for transmission of the parasite to the human host [16,17]. Furthermore, studies with lipophosphoglycan deletion mutants in *L. major* demonstrated that these glycosylated structures are involved in resistance to oxidative stress and evasion of the human immune system [16,17]. Moreover, a UGM deletion mutant of *L. major* exhibits attenuated virulence [7]. In summary, Gal*f*-containing molecules of protozoan parasites function in host-specific cell recognition, growth, and pathogenesis. Since UGM is essential for the biosynthesis

of Gal*f*, inhibition of the enzyme is an attractive approach for finding new drugs for Chagas disease and leishmaniasis.

The potential for inhibitor design and the uniqueness of the chemical mechanism of UGM have motivated structural studies of the enzyme. Several crystal structures of bacterial UGMs have been determined [18-23]. These structures revealed the essential UGM fold and provided insight into several aspects of UGM biochemistry, including the structural basis of substrate recognition and the catalytic mechanism. Eukaryotic UGMs have received less attention. We recently reported crystal structures and small-angle X-ray scattering analysis of UGM from the pathogenic fungus *Aspergillus fumigatus* (AfUGM), which was the first structural data for any eukaryotic UGM[24]. Shortly thereafter, Sanders' group published structures of AfUGM based on a different (space group *P*1) crystal form[25]. Our analysis of the data showed that AfUGM has several extra secondary and tertiary structural elements that are not found in bacterial UGMs yet are important for substrate recognition and tetramerization[24]. The AfUGM structures also revealed large conformational changes that accompany substrate binding, which is highly relevant for inhibitor design[24].

As part of our ongoing studies of UGM from eukaryotic pathogens, we report crystal structures of oxidized and reduced *T. cruzi* UGM (TcUGM) complexed with the inhibitor UDP. Analysis of the UDP binding site suggests a common strategy for designing inhibitors of UGMs from eukaryotic pathogens, including *T. cruzi*, *L. major*, and *A. fumigatus*. Comparison of the structures of oxidized and reduced TcUGM reveals profound conformational changes induced by reduction of the FAD cofactor, which provides new information about the molecular mechanism of enzyme activation by

reducing agents. These results provide a foundation for inhibitor design and insight into UGM biochemistry.

3.2 Materials and Methods

3.2.1. Crystallization

TcUGM was expressed and purified as described previously[26]. Crystallization studies used TcUGM at 7–8 mg/mL in 150 mM NaCl buffered at pH 7.5 with either 50 mM phosphate or 50 mM HEPES. Prior to crystallization, the enzyme was incubated with 10 mM UDP for 30 min. Crystallization experiments were performed at 20 °C using sitting drop vapor diffusion with the drops formed by mixing 1.5 μL each of the protein and reservoir solutions. Several commercially available crystallization screens were used to identify initial crystallization conditions. Promising conditions were obtained with reservoirs containing ammonium sulfate and HEPES buffer. Small yellow crystals appeared after 2 weeks. Larger crystals were obtained within a week via microseeding. The optimized crystallization reservoir contains 1.0 M ammonium sulfate, 0.5% poly(ethylene glycol) 8000, and 0.1 M HEPES at pH 7.5. The crystals were cryoprotected in 1.2 M ammonium sulfate, 0.1 M HEPES pH 7.5, and 25% glycerol before plunging into liquid N_2.

The space group is $P6_522$ with unit cell dimensions of a = 143 Å and c = 354 Å. Based on the method of Matthews[27] and assuming ~50% solvent content, the asymmetric unit is predicted to contain four protein molecules, which implies V_M of 2.4 Å^3/Da (48% solvent). However, molecular replacement calculations show that the

asymmetric unit contains just two protein molecules (*vide infra*), which corresponds to 74% solvent and V_M of 4.8 Å^3/Da.

Crystals of the TcUGM–UDP complex with the FAD in the reduced state were prepared by soaking the aforementioned crystals in 1.2 M ammonium sulfate, 0.1 M HEPES pH 7.5, 60 mM sodium dithionite, and 25% ethylene glycol. Once the crystals turned from yellow to colorless, they were flash-cooled by plunging into liquid N_2.

3.2.2 X-ray Diffraction Data Collection, Phasing, and Refinement

Diffraction data were collected at the Advanced Photon Source. The data set for oxidized TcUGM–UDP was collected at beamline 19-ID and processed using HKL3000[28]. The data set for reduced TcUGM–UDP was collected at beamline 24-ID-C, and the data were integrated using XDS[29] and scaled with SCALA[30] via CCP4i[31]. Data processing statistics are listed in Table 3.1.

The phase problem for oxidized TcUGM–UDP was solved using molecular replacement as implemented in MOLREP[32]. The search model was derived from the structure of AfUGM (PDB code 3UTE[24]). Chainsaw was used to create a model in which all the side chains were pruned to the β-carbon atom. The calculations produced a solution having two molecules in the asymmetric unit with *R*-factor of 0.6 and score of 0.3. The model was manually edited and built using COOT[33] and refined using PHENIX[34]. An advanced model of oxidized TcUGM-UDP was used as the starting point for refinement of reduced TcUGM–UDP. Refinement statistics are listed in Table 3.1. We note that, for both the oxidized and reduced TcUGM–UDP structures, the two molecules in the crystallographic asymmetric unit are identical within experimental error.

Table 3.1 X-ray Diffraction Data Collection and Refinement[a]

	Oxidized	Reduced
Space group	$P6_522$	$P6_522$
Unit cell parameters (Å)	a = 143.4, c = 354.2	a = 143.8, c = 354.4
Wavelength (Å)	0.9791	0.9795
Resolution (Å)	19.88 - 2.25 (2.33 - 2.25)	47.08 - 2.25 (2.37 - 2.25)
Observations	523354	841035
Unique reflections	101725	102978
$R_{merge}(I)$[b]	0.105 (0.542)	0.089 (0.526)
$R_{meas}(I)$[b]	0.105 (0.647)	0.095 (0.566)
$R_{pim}(I)$[b]	0.045 (0.278)	0.032 (0.204)
Mean I/σ	14.1 (2.2)	17.6 (3.7)
Completeness (%)	99.2 (100.0)	99.9 (99.8)
Multiplicity	5.1 (5.2)	8.2 (7.3)
No. of protein residues	939	938
No. of protein atoms	7495	7482
No. of FAD atoms	106	106
No. of ligand atoms	50	50
No. of water molecules	381	246
R_{cryst}	0.183 (0.246)	0.180 (0.238)
R_{free}[c]	0.212 (0.298)	0.209 (0.283)
rmsd bond lengths (Å)[d]	0.007	0.007
rmsd bond angles (°)[d]	1.12	1.11
Ramachandran plot[e]		
Favored (no. residues)	912	914
Allowed (no. residues)	21	18
Outliers (no. residues)	0	0
Average B-factor (Å^2)		
Protein	30	32
FAD	23	24
UDP	21	25
Water	32	33
Coordinate error (Å)[f]	0.32	0.32
PDB code	4DSG	4DSH

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Definitions of R_{merge}, R_{meas}, and R_{pim} can be found in Weiss {Weiss, 2001 #3422}[1].
[c]A common set of test reflections (5 %) was used for refinement of both structures.
[d]Compared to the parameters of Engh and Huber {Engh, 1991 #980}.
[e]The Ramachandran plot was generated with RAMPAGE {Lovell, 2003 #2719}.
[f]Maximum likelihood-based coordinate error estimate from PHENIX.

3.3 Results

3.3.1 Overall Fold and Oligomeric State

The structures of oxidized and reduced TcUGM complexed with the inhibitor UDP were determined at 2.25 Å resolution (Table 3.1). These are the first structures of UGM from a parasitic pathogen and the second structure of a eukaryotic UGM.

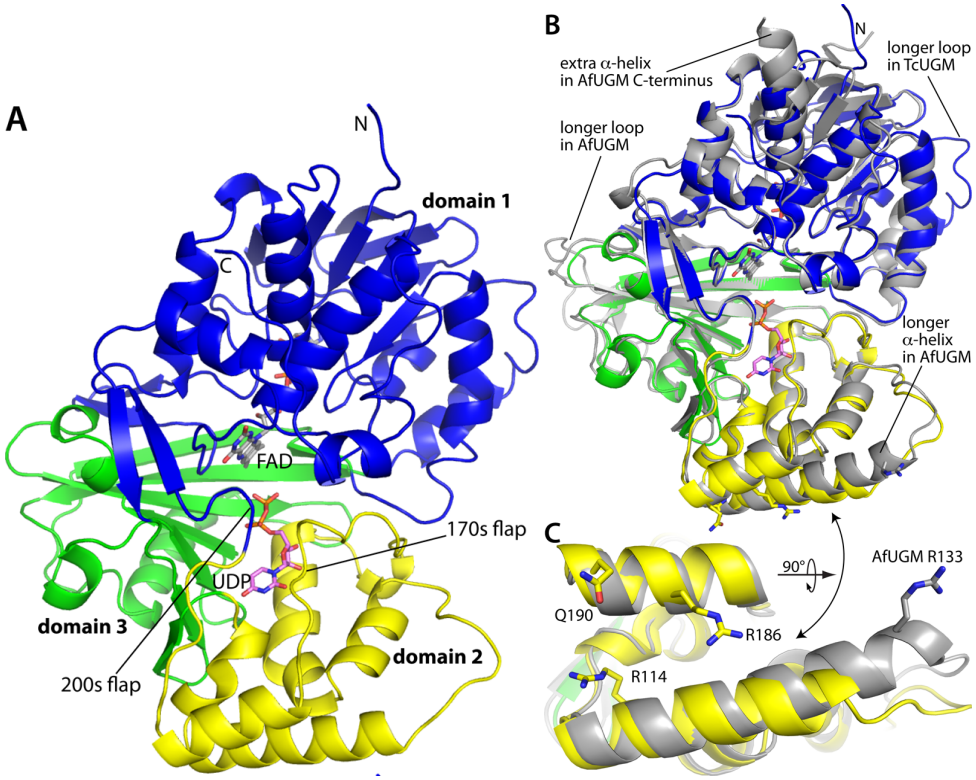

Fig. 3.2 Structure of TcUGM. (A) Structure of the TcUGM monomer. Domains 1, 2, and 3 are colored blue, yellow, and green, respectively. FAD and UDP are colored gray and pink, respectively. (B) Superposition of TcUGM (blue, yellow, green) and AfUGM (gray). (C) Close-up view of two helices on the periphery of domain 2. The view is rotated from panel B by about 90° around the horizontal axis. TcUGM and AfUGM are colored as in panel B. This figure and others were created with PyMOL {DeLano, 2002 #2088}.

AfUGM is the closest structural neighbor of TcUGM in the Protein Data Bank (PDB). The two enzymes are 45% identical in sequence, and the two structures

superimpose with a root-mean-square deviation of 1.1 Å (Fig. 3.2B). Although they share a common overall fold, several local variations between the two structures are evident (Fig. 3.2B). For example, a loop of domain 1 is longer in TcUGM, whereas a loop in domain 3 is longer in AfUGM. The significance of these differences is not obvious. In contrast, other differences are substantive in that they appear to dictate the oligomeric state formed in solution. These include the longer α-helix in domain 2 of AfUGM and the additional α-helix at the C-terminus of AfUGM (Fig. 3.2B).

The crystal structure is consistent with TcUGM being monomeric in solution. Analysis of protein–protein interfaces in the crystal lattice using the PDBePISA[37] revealed no oligomers predicted to be stable in solution. Furthermore, neither the AfUGM tetramer nor its constituent dimers are present in the TcUGM crystal lattice. Moreover, none of the dimeric assemblies of bacterial UGMs are observed in the lattice. It is concluded that TcUGM forms a monomer in solution, which is consistent with recent size exclusion chromatography data[26].

3.3.2 Binding of UDP

The structures of TcUGM complexed with UDP were determined from crystals that had been grown in the presence of the inhibitor. Electron density maps clearly indicated that UDP is bound in the active site with full occupancy (Fig. 3.3A). The protein–ligand interactions are identical in the reduced and oxidized enzymes, so we will focus on the reduced enzyme complex.

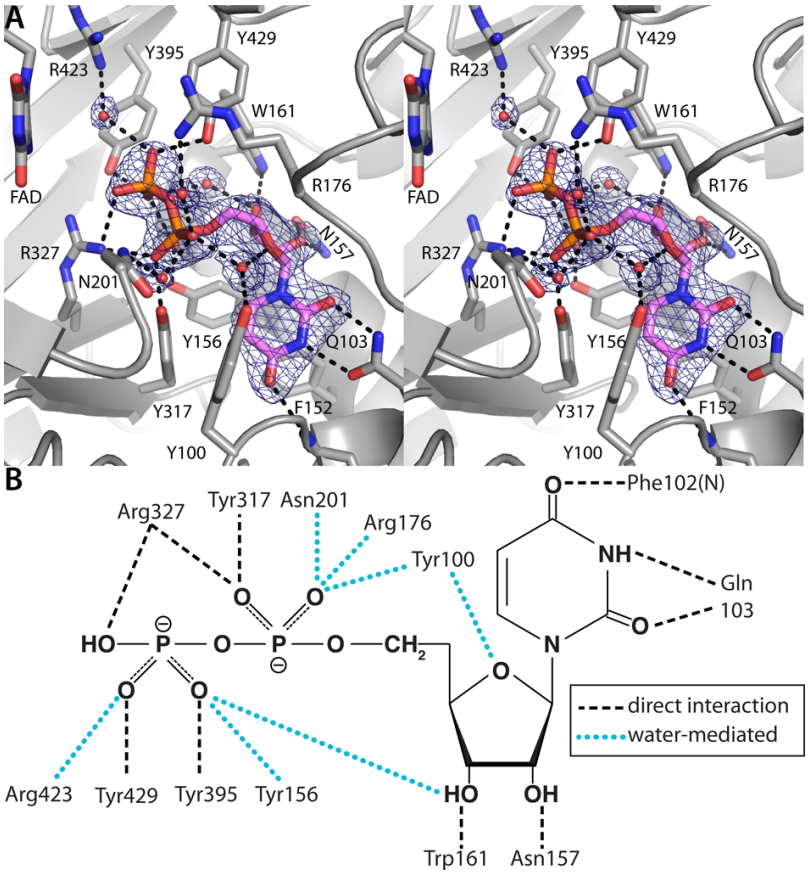

Fig. 3.3 Electron density and interactions for the UDP bound to reduced TcUGM. (A) Stereographic view of the TcUGM active site. The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3.0 σ. (B) Schematic diagram of protein-UDP interactions in TcUGM. Backbone interactions are indicated by N in parentheses.

UDP binds at the *re* face of the isoalloxazine and is covered by the 170s and 200s flaps (Fig. 3.3A). Structures of AfUGM showed that the analogous loops move by 10 Å from the open to closed conformation upon UDP binding. The flaps of TcUGM–UDP are in the closed conformation, as expected. Although it was not possible to crystallize TcUGM in the ligand-free state, presumably the 170s and 200s flaps open and close as in AfUGM.

UDP forms several interactions with TcUGM (Fig. 3.3B). The uracil ring is wedged between the aromatic rings of Tyr100 and Phe152, which positions the base to

form three hydrogen bonds with the side chain of Gln103 and backbone of Phe102. Thus, the hydrogen-bonding potential of uracil is fully satisfied. The ribose hydroxyl groups of UDP interact with Asn157 and Trp161. The pyrophosphate directly interacts with Arg327 and three tyrosine residues (317, 395, and 429). Several water-mediated hydrogen bonds are also observed. In summary, all three chemical groups of UDP—base, ribose, and pyrophosphate—are in direct electrostatic contact with the enzyme, which results in a plethora of stabilizing interactions.

3.3.2 Conformational Changes Induced by FAD Reduction

The FAD of UGM functions as a nucleophile that attacks the anomeric carbon atom of the galactose moiety (C1 in Fig. 3.1) in an S_N2-like reaction; thus, enzymatic activity requires that the FAD be in the reduced state ($FADH^-$) [26,38,39]. Structures of oxidized and reduced TcUGM were determined to understand the mechanism by which the enzyme is activated by FAD reduction.

The structure of the oxidized enzyme was determined from crystals that were grown and cryoprotected without reducing agent. These crystals were yellow, which is characteristic of the oxidized state of FAD. The structure of the reduced enzyme was determined from crystals of the oxidized enzyme that were soaked in reducing agent (dithionite) prior to flash-cooling in liquid nitrogen. The crystals turned colorless during soaking, which indicated that the FAD was reduced.

Reduction of the FAD causes substantial conformational changes in the active site (Fig. 3.4). In the oxidized enzyme, the conserved histidine loop (Gly60-Gly61-His62) adopts an unprecedented conformation for UGMs in which it is retracted from the FAD isoalloxazine (Fig. 3.4A). This conformation is stabilized by hydrogen bonds between

Asp58 and the backbone of Gly61 and His62. Upon reduction, the histidine loop shifts 2.3 Å toward the isoalloxazine (Fig. 3.4B). In this conformation, the carbonyl of Gly61 accepts a hydrogen bond from the flavin N5, which is a hydrogen bond donor in the reduced state. This hydrogen bond is observed in all reduced UGM structures and is thus thought to be essential for stabilizing the reduced flavin. Movement of the histidine loop toward the reduced flavin triggers other conformational changes (Fig. 3.4C). Asp58 and Thr212 rotate to engage each other in a hydrogen bond. Note that the rotation of Asp58 also allows a hydrogen bond with Ser 48. Finally, Met347 moves into the space vacated by Gly61 and Asp58, while Gln434 rotates away from His62.

The imidazole ring of His62 becomes protonated and flips by 180° upon flavin reduction. Although the protonation state and value of $\chi 2$ of histidine cannot be determined solely from electron density at 2.25 Å resolution, these attributes can be inferred from hydrogen-bonding considerations and knowledge of the preferred tautomer of neutral histidine. Analysis of hydrogen bonding with MolProbity[40] suggests that in the oxidized enzyme, His62 adopts the neutral τ tautomer with the imidazole accepting a hydrogen bond from Gln434 and donating a hydrogen bond to the carbonyl of Leu59 (Fig. 3.4A). We note that τ is the preferred tautomer of neutral His[41]. In the reduced enzyme, on the other hand, hydrogen bonding is maximized with the imidazole protonated and flipped by 180° to allow hydrogen bonds with the carbonyl of Gly60 and the FAD 2′-hydroxyl (Fig. 3.4B). Note that the assigned protonation states of His62 are consistent with the fact that the oxidized FAD is uncharged and the reduced FAD is anionic ($FADH^-$).

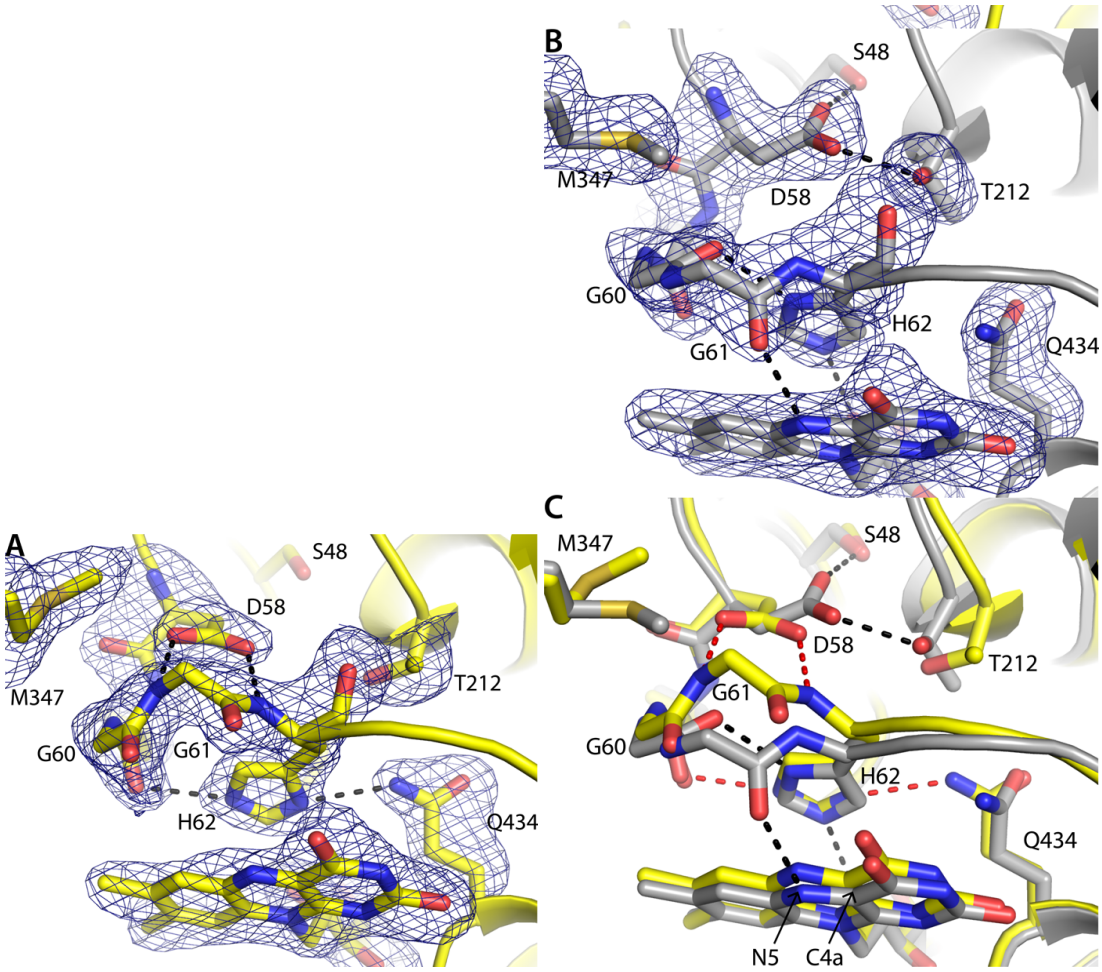

Fig. 3.4 Structural changes induced by FAD reduction. (A) Electron density for the histidine loop region of oxidized TcUGM. The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3 σ. (B) Electron density for the histidine loop region of reduced TcUGM. The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3 σ. (C) Superposition of oxidized (yellow) and reduced (gray) TcUGM. Red and black dashes represent hydrogen bonds for oxidized and reduced TcUGM, respectively.

Reduction also affects the conformation of the isoalloxazine (Fig. 3.5). In the oxidized state, the electron density maps are consistent with a planar isoalloxazine (Fig. 3.5A). In the reduced enzyme, the FAD exhibits a butterfly like conformation in which the pyrimidine ring bends 7° out of the plane such that the *si* face is concave (Fig. 3.5A). This conformation is identical to that of reduced AfUGM.

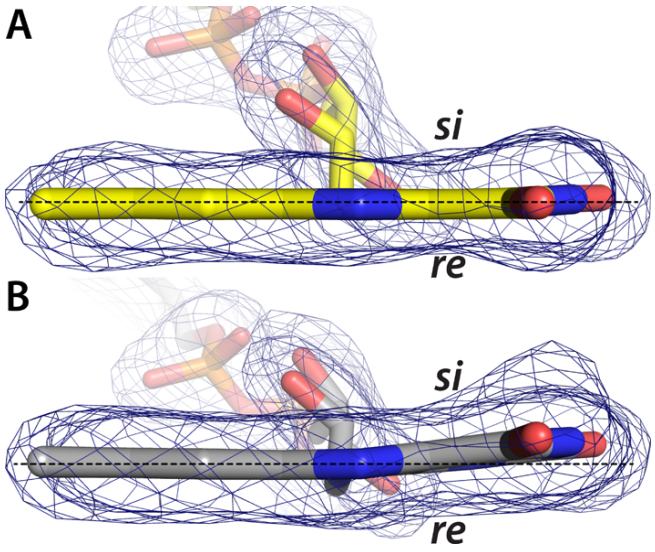

Fig. 3.5 Electron density for the isoalloxazine rings of (A) oxidized and (B) reduced TcUGM. The cages represent simulated annealing σ_A-weighted F_o - F_c omit maps contoured at 3 σ. The horizontal line assists in seeing the 7° butterfly-like bend angle of the reduced cofactor.

3.3.3 Site-Directed Mutagenesis of the Histidine Loop

The importance of the histidine loop for catalytic activity was investigated using site-directed mutagenesis. Gly61 and His62 were targeted for mutagenesis because they form hydrogen bonds to $FADH^-$ and exhibit large conformational changes upon flavin reduction. Gly61 is conserved among eukaryotic UGMs but appears as Ala or Pro in bacterial UGMs. Therefore, the mutant enzymes G61A and G61P were created. His62 is universally conserved among UGMs, and the H62A mutant enzyme was created. (For reference, Gly60 is also conserved in all UGMs.)

Mutation of the histidine loop of TcUGM is highly detrimental to activity (Fig. 3.6 and Table 3.2). Mutation of Gly61 to Ala reduces k_{cat} by a factor of 70 but reduces K_m by only a factor of 3. As a result, the catalytic efficiency of G61A is only 4% of that of TcUGM. Similarly, mutation of Gly61 to Pro substantially decreases k_{cat} (factor of 16) but has less effect on K_m. The catalytic efficiency of this mutant enzyme is 10% of

that of TcUGM. Mutation of His62 to Ala has a profound effect on k_{cat}; this mutation decreases k_{cat} by over 300. The catalytic efficiency of H62A is just 2% compared to that of TcUGM. These results suggest that Gly and His at positions 61 and 62 are important for efficient catalysis by TcUGM. Furthermore, the histidine loop sequences found in bacterial UGMs (GAH, GPH) are poorly tolerated by TcUGM.

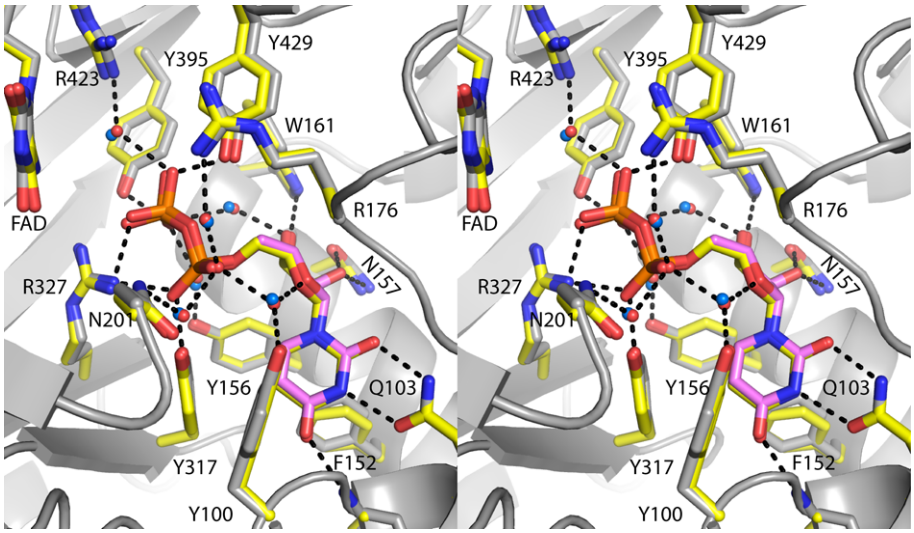

Fig. 3.6 Superposition of the active sites of TcUGM and AfUGM.

Table 3.2 Steady-state kinetic constants for TcUGM and TcUGM mutant enzymes.

	k_{cat} (s^{-1})	K_m (mM)	k_{cat} (s^{-1})/K_m ($s^{-1} M^{-1}$)	k_{cat} (s^{-1})/K_m (%)
TcUGM	13.4 ± 0.3	140 ± 10	93 ± 6	100 ± 6
G61A	0.198 ± 0.011	50 ± 10	4 ± 1	4 ± 1
G61P	0.83 ± 0.06	90 ± 20	9 ± 2	10 ± 2
H62A	0.041 ± 0.001	24 ± 4	1.8 ± 0.2	2 ± 0.2

3.4 Discussion

Inhibitor design is aided by knowledge of substrate recognition, and substrate recognition appears to be exquisitely conserved among eukaryotic UGMs. Comparison of the structures of TcUGM and AfUGM complexed with UDP show that the UDP binding sites are identical (Fig. 3.7). All residues that contact UDP, either directly or via water molecules, are present in both enzymes. Furthermore, the conformations of these residues, as well as their interactions with UDP, are identical in the two structures. The structural similarity extends even to the water molecules that mediate protein–inhibitor interactions (Fig. 3.7). Moreover, all residues that contact UDP or UDP-Gal*p* in the TcUGM and AfUGM crystal structures are also present in many other eukaryotic UGMs, including *L. major* UGM (Fig. 3.7, triangles). Thus, it is likely that the AfUGM and TcUGM structures are representative of eukaryotic UGMs with regard to substrate binding.

The TcUGM structures provide new information about conformational changes associated with activation of the enzyme via reduction of the FAD (Fig. 3.4 and 3.5). Reduction of the FAD causes several concerted changes in the protein. The histidine loop moves by 2.3 Å (Fig. 3.4C). His62, a universally conserved residue in UGMs, flips by 180° and becomes protonated. The protonation of His62 may help stabilize the negative charge of the reduced FAD. The side chains of Asp58 and Thr212 rotate by 180°. Met 347 and Gln434 move by 1.6 Å. It is notable that all of these residues are located on the side of the FAD that is opposite to the substrate-binding site. Thus, two critical aspects of function are delegated to distinct regions of the protein: maintaining the redox state is the

responsibility of residues on the *si* face FAD, while substrate binding is performed by residues on the *re* side.

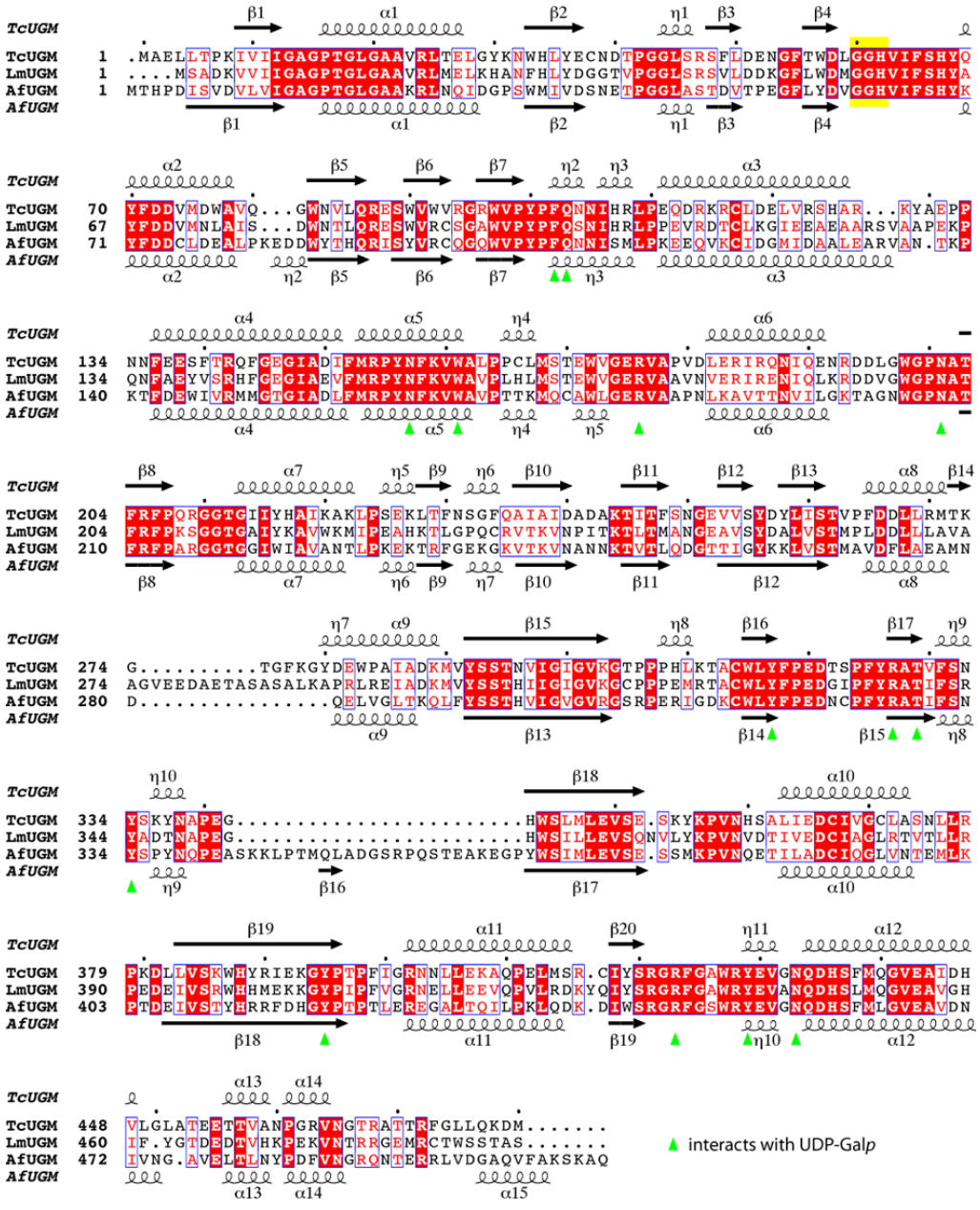

Fig. 3.7 Amino acid sequence alignment of UGMs from *T. cruzi*, *L. major*, and *A. fumigatus*. The secondary structure elements above and below the alignment are from the TcUGM and AfUGM structures, respectively. The yellow box denotes the histidine loop, which moves in response to changes in the FAD redox state. The green triangles denote residues that contact UDP-Gal*p*; note that they are 100 % identical in the three enzymes.

These conformational changes are consistent with the generally accepted chemical mechanism of UGM. The prevailing mechanism is an S_N2-type displacement in which the N5 atom of the reduced FAD functions as the nucleophile that attacks the anomeric carbon of galactose to form a covalent intermediate and displace UDP [38,39]. This mechanism was recently validated for TcUGM[26]. Activity thus requires that the FAD be reduced. The coordinated movements of the histidine loop, Asp58, Thr212, and the FAD isoalloxazine in TcUGM have two salient effects. First, a hydrogen bond is created between the N5 atom of the reduced flavin and the carbonyl oxygen atom of the residue preceding the conserved histidine (Gly61). This interaction is seen in all reduced UGMs and is therefore thought to be essential for stabilizing the reduced flavin. Second, curvature is induced in the flavin isoalloxazine such that the *si* face is concave. Bending of the isoalloxazine in this direction is consistent with the FAD functioning as a nucleophile [19].

Curiously, the conformational changes observed for TcUGM are different from those described for the close homologue AfUGM (Fig. 3.8). We previously reported structures of oxidized and reduced AfUGM based on a $P6_522$ crystal form[24]. Sanders' group subsequently reported $P1$ structures of AfUGM[25]. The structures of the reduced active sites are nearly identical in TcUGM and AfUGM (Fig. 3.8C). However, in oxidized AfUGM, the His residue of the histidine loop is substantially displaced toward the pyrimidine portion of the isoalloxazine (Fig. 3.8B), whereas in oxidized TcUGM the His residue is near the middle ring of the isoalloxazine (Fig. 3.8A). Furthermore, in oxidized AfUGM, the carbonyl bond vector of the middle Gly residue is directed away from the isoalloxazine (Fig. 3.8B, Gly62), whereas the corresponding carbonyl of

oxidized TcUGM points toward the isoalloxazine (Fig. 3.8A, Gly61). Reduction of the FAD in AfUGM causes a 6–8 Å movement of the imidazole ring of the conserved histidine (Fig. 3.8). As described above, the movement of the histidine loop in TcUGM is more subtle. It should be noted that elucidation of redox-linked conformational changes for AfUGM is complicated by the adventitious binding of sulfate ions to the oxidized active site of the $P6_522$ crystal form (PDB code 3UTE) and weak electron density for the histidine loop and bound UDP in the oxidized $P1$ form (3UKH). Nevertheless, taken together, the AfUGM and TcUGM structures implicate flexibility of the histidine loop in the mechanism of enzyme activation.

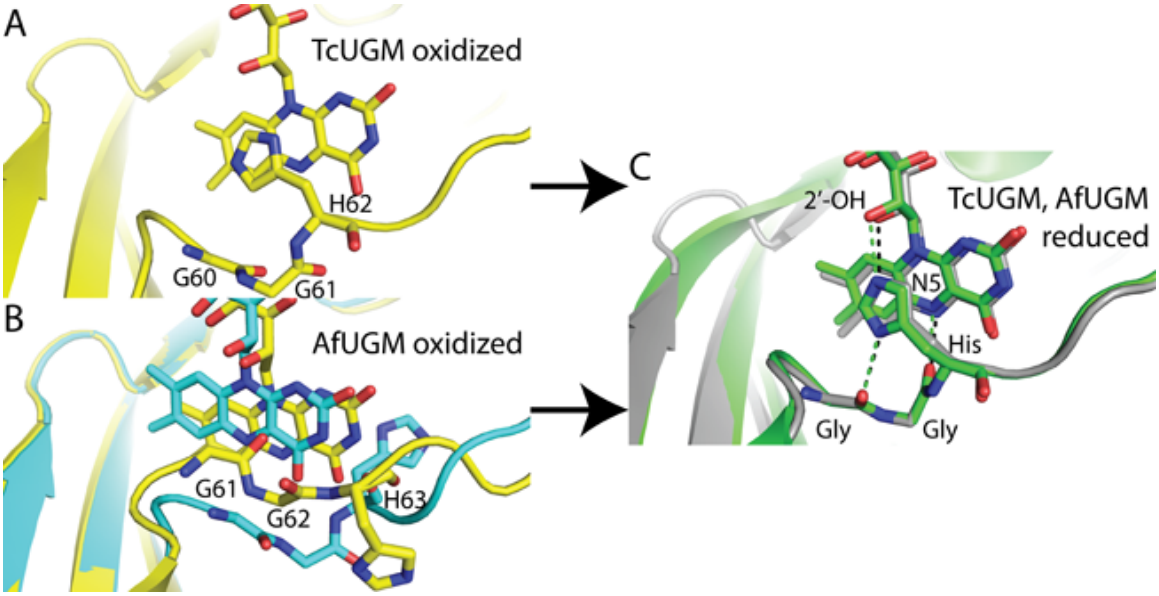

Fig. 3.8 Summary of conformational changes induced by flavin reduction in TcUGM and AfUGM. (A) Oxidized TcUGM (PDB code 4DSG). (B) Oxidized AfUGM crystallized in space groups $P6_522$ (cyan, PDB code 3UTE) and $P1$ (yellow, PDB code 3UKH). (C) Superposition of reduced TcUGM (gray, 4DSH) and reduced AfUGM (green, 3UTF).

The large conformational changes in the histidine loop observed in TcUGM and AfUGM appear to be unique to eukaryotic UGMs. Comparison of oxidized and reduced

bacterial UGM structures shows that flavin reduction induces bending of the isoalloxazine but no substantial changes in the protein conformation [19,21,22]. nalysis of amino acid sequence conservation provides a rationale for why the histidine loop is static in bacterial UGMs and dynamic in eukaryotic UGMs. The histidine loop sequence of GGH is conserved among eukaryotic UGMs (Fig. 3.8). In contrast, bacterial UGMs have either Ala or Pro in place of the second Gly. The extra Gly in the histidine loop most likely accounts for the increased flexibility of the loop in AfUGM and TcUGM.

This additional flexibility of the histidine loop appears to be important for function in TcUGM. Our mutagenesis data show that transplanting the bacterial sequences of GAH and GPH into TcUGM substantially decreases catalytic efficiency (Table 3.2). It is possible that the additional flexibility afforded by the extra Gly residue in TcUGM is needed to establish optimal hydrogen bonding between the histidine loop and $FADH^-$. Although structures of the TcUGM histidine loop mutant enzymes are not available, it is possible that the loop-flavin hydrogen bonding is suboptimal in these impaired enzymes, which likely decreases the nucleophilic character of the reduced flavin. This idea is consistent with the markedly decreased k_{cat} values of the histidine loop mutants.

Why eukaryotic UGMs undergo such large conformational changes upon activation is an open question. One possibility is that these conformational changes control the access of O_2 and reductant to the flavin. In the reduced enzyme, positioning the histidine loop close to the isoalloxazine N5–C4a edge helps protect the C4a atom of the reduced FAD from attack by O_2 (Fig. 3.4B). Retraction of the loop upon oxidation perhaps allows open space for reducing agents to gain access to the oxidized flavin (Fig.

3.4A). Regulation of function is another possibility. Conformational changes that are linked to the flavin redox state often serve a regulatory purpose[42]. Whether eukaryotic UGMs are subject to some sort of redox-linked regulation remains to be determined.

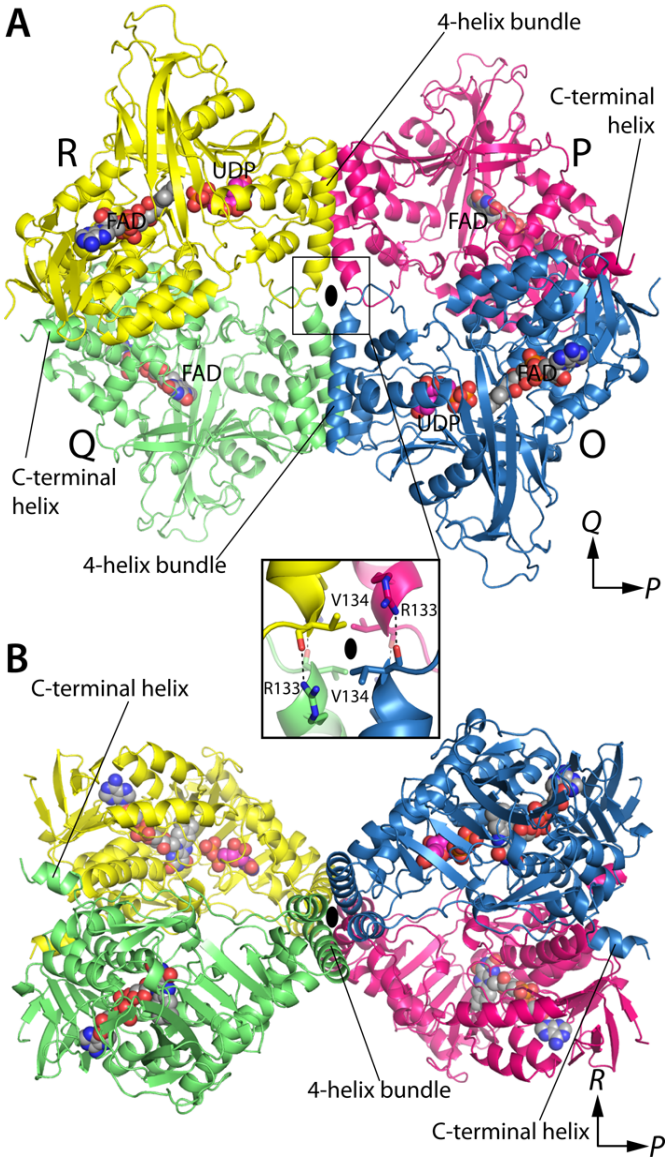

Fig. 3.9 Structure of the AfUGM tetramer. (A) The tetramer is viewed down the *R* molecular 2-fold axis. Each chain has a different color. Inset: Intersubunit hydrogen bonds formed by Arg133 at the intersection of molecular 2-fold axes. (B) The tetramer is viewed down the Q-axis.

3.5 References

1. Savioli, L., and Daumerie, D. (2010) First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases, (Crompton, D. W. T., Ed.), World Health Organization, Geneva.

2. Bern, C., and Montgomery, S. P. (2009) An estimate of the burden of Chagas disease in the United States, *Clin. infect. Dis. 49*, e52-54.

3. Tanowitz, H. B., Weiss, L. M., and Montgomery, S. P. (2011) Chagas disease has now gone global, *PLoS Negl. Trop. Dis. 5*, e1136.

4. Bern, C., Montgomery, S. P., Herwaldt, B. L., Rassi, A., Jr., Marin-Neto, J. A., Dantas, R. O., Maguire, J. H., Acquatella, H., Morillo, C., Kirchhoff, L. V., Gilman, R. H., Reyes, P. A., Salvatella, R., and Moore, A. C. (2007) Evaluation and treatment of chagas disease in the United States: a systematic review, *JAMA 298*, 2171-2181.

5. Teixeira, A. R., Cordoba, J. C., Souto Maior, I., and Solorzano, E. (1990) Chagas' disease: lymphoma growth in rabbits treated with Benznidazole, *Am. J. Trop. Med. Hyg. 43*, 146-158.

6. Issa, V. S., and Bocchi, E. A. (2010) Antitrypanosomal agents: treatment or threat?, *Lancet 376*, 768; author reply 768-769.

7. Beverley, S. M., Owens, K. L., Showalter, M., Griffith, C. L., Doering, T. L., Jones, V. C., and McNeil, M. R. (2005) Eukaryotic UDP-galactopyranose mutase (GLF gene) in microbial and metazoal pathogens, *Eukaryot. Cell. 4*, 1147-1154.

8. Tefsen, B., Ram, A. F., van Die, I., and Routier, F. H. (2011) Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact, *Glycobiology*, in press.

9. Oppenheimer, M., Valenciano, A. L., and Sobrado, P. (2011) Biosynthesis of galactofuranose in kinetoplastids: novel therapeutic targets for treating leishmaniasis and chagas' disease, *Enzyme Res. 2011*, 1-13.

10. Almeida, I. C., Ferguson, M. A., Schenkman, S., and Travassos, L. R. (1994) GPI-anchored glycoconjugates from Trypanosoma cruzi trypomastigotes are recognized by lytic anti-alpha-galactosyl antibodies isolated from patients with chronic Chagas' disease, *Braz. J. Med. Biol. Res. 27*, 443-447.

11. Ralton, J. E., Milne, K. G., Guther, M. L., Field, R. A., and Ferguson, M. A. (1993) The mechanism of inhibition of glycosylphosphatidylinositol anchor biosynthesis in Trypanosoma brucei by mannosamine, *J. Biol. Chem. 268*, 24183-24189.

12. Ferguson, M. A. (1997) The surface glycoconjugates of trypanosomatid parasites, *Philos. Trans. R. Soc. Lond. B Biol. Sci. 352*, 1295-1302.

13. Turnock, D. C., and Ferguson, M. A. (2007) Sugar nucleotide pools of Trypanosoma brucei, Trypanosoma cruzi, and Leishmania major, *Eukaryotic cell 6*, 1450-1463.

14. MacRae, J. I., Obado, S. O., Turnock, D. C., Roper, J. R., Kierans, M., Kelly, J. M., and Ferguson, M. A. (2006) The suppression of galactose metabolism in Trypanosoma cruzi epimastigotes causes changes in cell surface molecular architecture and cell morphology, *Mol. Biochem. Parasitol. 147*, 126-136.

15. Turner, C. W., Lima, M. F., and Villalta, F. (2002) Trypanosoma cruzi uses a 45-kDa mucin for adhesion to mammalian cells, *Biochem. Biophys. Res. Commun. 290*, 29-34.

16. Spath, G. F., Garraway, L. A., Turco, S. J., and Beverley, S. M. (2003) The role(s) of lipophosphoglycan (LPG) in the establishment of Leishmania major infections in mammalian hosts, *Proc Natl Acad Sci U S A 100*, 9536-9541.

17. Zhang, K., Barron, T., Turco, S. J., and Beverley, S. M. (2004) The LPG1 gene family of Leishmania major, *Mol. Biochem. Parasitol. 136*, 11-23.

18. Sanders, D. A., Staines, A. G., McMahon, S. A., McNeil, M. R., Whitfield, C., and Naismith, J. H. (2001) UDP-galactopyranose mutase has a novel structure and mechanism, *Nat. Struct. Biol. 8*, 858-863.

19. Beis, K., Srikannathasan, V., Liu, H., Fullerton, S. W., Bamford, V. A., Sanders, D. A., Whitfield, C., McNeil, M. R., and Naismith, J. H. (2005) Crystal structures of Mycobacteria tuberculosis and Klebsiella pneumoniae UDP-galactopyranose mutase in the oxidised state and Klebsiella pneumoniae UDP-galactopyranose mutase in the (active) reduced state, *J. Mol. Biol. 348*, 971-982.

20. Gruber, T. D., Borrok, M. J., Westler, W. M., Forest, K. T., and Kiessling, L. L. (2009) Ligand binding and substrate discrimination by UDP-galactopyranose mutase, *J Mol Biol 391*, 327-340.

21. Gruber, T. D., Westler, W. M., Kiessling, L. L., and Forest, K. T. (2009) X-ray crystallography reveals a reduced substrate complex of UDP-galactopyranose mutase poised for covalent catalysis by flavin, *Biochemistry 48*, 9171-9173.

22. Partha, S. K., van Straaten, K. E., and Sanders, D. A. (2009) Structural basis of substrate binding to UDP-galactopyranose mutase: crystal structures in the reduced and oxidized state complexed with UDP-galactopyranose and UDP, *J. Mol. Biol. 394*, 864-877.

23. Partha, S. K., Sadeghi-Khomami, A., Slowski, K., Kotake, T., Thomas, N. R., Jakeman, D. L., and Sanders, D. A. (2010) Chemoenzymatic synthesis, inhibition studies, and X-ray crystallographic analysis of the phosphono analog of UDP-Galp as an inhibitor and mechanistic probe for UDP-galactopyranose mutase, *J. Mol. Biol. 403*, 578-590.

24. Dhatwalia, R., Singh, H., Oppenheimer, M., Karr, D. B., Nix, J. C., Sobrado, P., and Tanner, J. J. (2012) Crystal structures and small-angle X-ray scattering analysis of UDP-galactopyranose mutase from the pathogenic fungus Aspergillus fumigatus, *J. Biol. Chem. 287* 9041–9051.

25. van Straaten, K. E., Routier, F. H., and Sanders, D. A. (2012) Structural insight into the unique substrate binding mechanism and flavin redox state of UDP-galactopyranose mutase from Aspergillus fumigatus, *J. Biol. Chem. 287*, 10780–10790.

26. Oppenheimer, M., Valenciano, A. L., Kizjakina, K., Qi, J., and Sobrado, P. (2012) Chemical Mechanism of UDP-Galactopyranose Mutase from Trypanosoma cruzi: A Potential Drug Target against Chagas' Disease, *PLoS ONE 7*, e32918.

27. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.

28. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

29. Kabsch, W. (2010) XDS, *Acta Crystallogr. D Biol. Crystallogr. 66*, 125-132.

30. Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr. D Biol. Crystallogr. 62*, 72-82.

31. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite, *Acta Crystallogr D Biol Crystallogr. 59*, 1131-1137.

32. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement, *J. Appl. Cryst. 30*, 1022-1025.

33. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr. 60*, 2126-2132.

34. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-

based system for macromolecular structure solution, *Acta Crystallogr. D Biol. Crystallogr. 66*, 213-221.

35. Oppenheimer, M., Poulin, M. B., Lowary, T. L., Helm, R. F., and Sobrado, P. (2010) Characterization of recombinant UDP-galactopyranose mutase from Aspergillus fumigatus, *Arch. Biochem. Biophys. 502*, 31-38.

36. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

37. Soltero-Higgin, M., Carlson, E. E., Gruber, T. D., and Kiessling, L. L. (2004) A unique catalytic mechanism for UDP-galactopyranose mutase, *Nat Struct. Mol. Biol. 11*, 539-543.

38. Sun, H. G., Ruszczycky, M. W., Chang, W. C., Thibodeaux, C. J., and Liu, H. W. (2011) Nucleophilic participation of the reduced flavin coenzyme in the mechanism of UDP-Galactopyranose mutase, *J. Biol. Chem.*

39. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta crystallogr. Section D, Biological crystallography 66*, 12-21.

40. Li, S., and Hong, M. (2011) Protonation, tautomerization, and rotameric structure of histidine: a comprehensive study by magic-angle-spinning solid-state NMR, *J Am Chem Soc 133*, 1534-1544.

41. Becker, D. F., Zhu, W., and Moxley, M. A. (2011) Flavin redox switching of protein functions, *Antioxid Redox Signal 14*, 1079-1091.

42. Weiss, M. (2001) Global indicators of X-ray data quality, *J. Appl. Cryst. 34*, 130-135.

43. Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for x-ray protein structure refinement, *Acta Cryst. A47*, 392-400.

44. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation, *Proteins 50*, 437-450.

45. DeLano, W. L. (2002) *The PyMOL User's Manual*, DeLano Scientific, Palo Alto, CA, USA.

Chapter 4

Identification of the NAD(P)H Binding Site of Eukaryotic UDP-Galactopyranose Mutase

Richa Dhatwalia, Harkewal Singh, LM Solano, Michelle Oppenheimer, RM Robinson, JF Ellerbrock, Pablo Sobrado, John J. Tanner

Abstract

UDP-galactopyranose mutase (UGM) plays an essential role in galactofuranose biosynthesis in microorganisms by catalyzing the conversion of UDP-galactopyranose to UDP-galactofuranose. The enzyme has gained attention recently as a promising target for the design of new antifungal, antitrypanosomal, and antileishmanial agents. Here we report the first crystal structure of UGM complexed with its redox partner NAD(P)H. Kinetic protein crystallography was used to obtain structures of oxidized *Aspergillus fumigatus* UGM (AfUGM) complexed with NADPH and NADH, as well as reduced AfUGM after dissociation of $NADP^+$. NAD(P)H binds with the nicotinamide near the FAD isoalloxazine and the ADP moiety extending toward the mobile 200s active site flap. The nicotinamide riboside binding site overlaps that of the substrate galactopyranose moiety, thus NADPH and substrate binding are mutually exclusive. On the other hand, the pockets for the adenine of NADPH and uracil of the substrate are distinct and separated by only 6 Å, which raises the possibility of designing novel inhibitors that bind both sites. All twelve residues that contact NADP(H) are conserved among eukaryotic UGMs. Residues that form the AMP pocket are absent in bacterial UGMs, which suggests that eukaryotic and bacterial UGMs have different NADP(H) binding sites. The structures address the longstanding question of how UGM binds NAD(P)H and provide new opportunities for drug discovery.

4.1 Introduction

UDP-galactopyranose mutase (UGM) catalyzes the conversion of UDP-galactopyranose (UDP-Gal*p*) to UDP-galactofuranose (UDP-Gal*f*), which is a central step in Gal*f* biosynthesis (Fig. 4.1A)[1]. The enzyme has attracted interest as a drug target because Gal*f* is an essential building block of the cell wall and extracellular matrix of many bacteria, fungi, and protozoa, but Gal*f*and UGM are absent in humans. In particular, eukaryotic UGMs have emerged recently as targets for the design of antifungal, antitrypanosomal, and antileishmanial agents [2-4]. For example, gene deletion studies have shown that UGM is essential for the virulence of the pathogenic fungus *Aspergillus fumigatus* and the protozoan parasite *Leishmania major* [5,6]. Thus, there is interest in characterizing the ligand binding sites of UGMs to facilitate drug discovery.

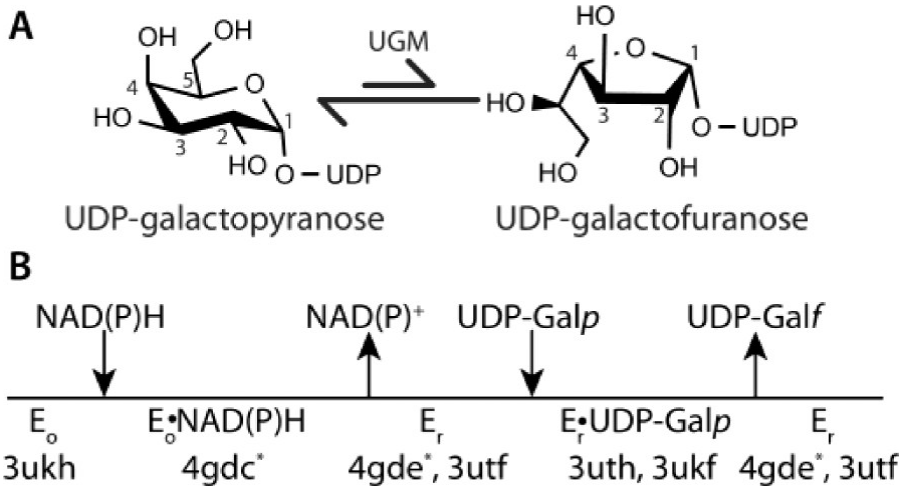

Fig. 4.1 (A) Reaction catalyzed by UGM. (B) Catalytic scheme indicating the PDB codes of AfUGM crystal structures. Asterisks denote the structures reported here.

UGM is an atypical flavoenzyme in that it does not catalyze an oxidation–reduction reaction, and thus the flavin redox state is unchanged by the transformation of substrate to product [7]. Rather than serving as a redox center, the flavin cofactor in UGM is a nucleophile that attacks the anomeric carbon atom of the substrate (C1 in Fig. 4.1A) [(8-10)]. The role as nucleophile requires that the flavin be reduced ($FADH^-$) in the resting state of the enzyme.

The requirement of $FADH^-$ in UGM raises the question of what is the physiological reductant of the enzyme. Reduced nicotinamide adenine dinucleotides (NAD(P)H) are logical candidates because of their ubiquity in biology and known reactivity with a multitude of flavoenzymes, including flavin reductases and flavin monooxygenases[11]. Indeed, early studies showed that NADPH and NADH reduce UGM from *Klebsiella pneumoniae*, albeit relatively slowly [(12,13)]. More recently, we showed that NADPH and NADH are effective redox partners of *Trypanosoma cruzi* UGM (TcUGM) [10]. These studies suggest that reduced nicotinamide adenine dinucleotides may be the physiological reductant of some UGMs. However, an NAD(P)H binding motif is not evident in UGM sequences, and the location of the NAD(P)H binding site has remained elusive despite the availability of dozens of crystal structures of bacterial [(14-19)] and eukaryotic [(20-22)] UGMs. To address this outstanding issue of UGM biochemistry, we have determined the crystal structures of UGM from the pathogenic fungus *Aspergillus fumigatus* (AfUGM) complexed with NADPH and NADH.

4.2 Materials and Methods

4.2.1 Crystal Soaking Experiments

Methods for the crystallization of oxidized AfUGM have been described[20]. Kinetic protein crystallography[26] as used to determine structures of oxidized AfUGM ($AfUGM_o$) complexed with NADPH or NADH and of reduced AfUGM ($AfUGM_r$) after dissociation of $NADP^+$. The soaking time and reductant concentration were varied to find appropriate soaking conditions. The $AfUGM_o$-NAD(P)H complexes were obtained by soaking the crystals for ~1 min in cryobuffer (1.4 M ammonium sulfate, 0.1 M sodium acetate, and 25% ethylene glycol at pH 4.5) containing 120–140 mM NADPH or NADH prior to flash-cooling in liquid nitrogen. The crystals remained yellow during this relatively short soaking time, which is consistent with the oxidized state of the FAD. The structure of $AfUGM_r$ without $NADP^+$ bound was determined from a crystal soaked for 30 min in the cryobuffer containing 100 mM NADPH. Over this time scale, the yellow color was completely bleached, which is indicative of full reduction of the crystalline enzyme.

4.2.2 X-ray Diffraction Data Collection and Refinement

Diffraction data were collected on beamlines 24-ID-C and 24-ID-E of the Advanced Photon Source. The data were integrated using XDS[27] and scaled with SCALA[28]. Data processing statistics are listed in Table 4.1.

Crystallographic refinement calculations were initiated from coordinates derived from the 2.25 Å resolution structure of AfUGM (Protein Data Bank (PDB) code 3UTF). These calculations were performed with PHENIX[29], and the *B*-factor model consisted of an isotropic *B*-factor for each atom and TLS refinement using one TLS group per protein chain. NCS restraints were used in the refinement of the 2.75 Å resolution structures.

Table 4.1 X-ray Diffraction Data Collection and Refinement[a]

	$AfUGM_o$-NADPH	$AfUGM_o$-NADH	$AfUGM_r$
Data collection			
Soaking ligand	NADPH	NADH	NADPH
Soaking time (min.)	1	1	30
Flavin redox state	Oxidized	Oxidized	Reduced
Space group	$P6_522$	$P6_522$	$P6_522$
Unit cell parameters (Å)	a =218.3, c = 319.2	a =218.3, c = 318.1	a = 217.4, c = 319.9
Wavelength (Å)	0.979	0.979	0.979
Resolution (Å)	163-2.75 (2.90-2.75)	162-2.75 (2.90-2.75)	50-2.20 (2.32-2.20)
Observations	851567	982507	984169
Unique reflections	116001	115207	222394
$R_{merge}(I)$[b]	0.137 (0.885)	0.158 (1.216)	0.093 (0.673)
$R_{meas}(I)$[b]	0.147 (0.967)	0.168 (1.297)	0.106 (0.778)
$R_{pim}(I)$[b]	0.054 (0.382)	0.054 (0.429)	0.049 (0.383)
Mean I/σ	11.6 (2.2)	10.9 (2.1)	9.8 (2.0)
Completeness (%)	99.9 (99.9)	99.6 (99.9)	99.7 (99.8)
Multiplicity	7.3 (6.1)	8.5 (7.9)	4.4 (4.0)
Refinement			
Resolution (Å)	163-2.75 (2.78-2.75)	162-2.75 (2.78-2.75)	48-2.20 (2.28-2.20)
No. of protein residues	1987	1989	2017
No. of protein atoms	15008	15164	15526
No. of FAD atoms	212	212	212
No. of NAD(P)H molecules/atoms	4/158	2/88	0/0
NAD(P)H occupancy[c]	0.7-0.9	0.9	-
No. of water molecules	16	26	435
R_{cryst}	0.210 (0.299)	0.212 (0.404)	0.201 (0.293)
R_{free}[d]	0.246 (0.318)	0.245 (0.391)	0.227 (0.343)
rmsd bond lengths (Å)[e]	0.008	0.008	0.007
rmsd bond angles (°)[e]	1.17	1.19	1.05
Ramachandran plot[f]			
Favored (no. residues)	1908	1919	1978
Allowed (no. residues)	59	50	31
Outliers (no. residues)	0	0	0
Average B-factor (Å^2)			
Protein	53	51	41
FAD	52	53	37
NAD(P)H[c]	48-72	50-52	-
Water	36	41	37
Coordinate error (Å)[g]	0.41	0.51	0.36
PDB code	4gdc	4gdd	4gde

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Definitions of R_{merge}, R_{meas}, and R_{pim} can be found in Weiss.{Weiss, 2001 #3255}
[c]Range reflects the ligands bound to different chains of the tetramer.
[d]A common set of test reflections (5 %) was used for refinement of all structures.
[e]Compared to the parameters of Engh and Huber.{Engh, 1991 #980}
[f]The Ramachandran plot was generated with RAMPAGE.{Lovell, 2003 #2719}
[g]Maximum likelihood-based coordinate error estimate from PHENIX.

COOT was used for model building[30]. The ligand occupancy (q) was estimated by inspecting the ligand B-factors after refinements performed with the ligand occupancy fixed at various values. Refinement statistics are listed in Table 4.1.

4.2.3 Mutagenesis and Kinetics

Site-directed mutants of AfUGM were created using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) following the protocol supplied by the manufacturer. All mutant enzymes were expressed and purified following the procedures for AfUGM[31].

AfUGM and AfUGM mutant enzymes were analyzed using stopped-flow kinetics experiments as described previously for TcUGM[10]. In these experiments, the enzyme and NAD(P)H were mixed under anaerobic conditions at 15 °C and pH 7.0, and the reaction was monitored using the flavin absorbance at 452 nm. The time course of the absorbance was fit to a single exponential to obtain an observed rate constant, k_{obs}. Kinetic parameters for the reduction of the enzyme by NAD(P)H were obtained by fitting the k_{obs} values to the equation, $k_{obs} = k_{red}[\text{NAD(P)H}]/(K_d + [\text{NAD(P)H}])$, where k_{red} is the rate constant for flavin reduction, and K_d is the dissociation constant for NAD(P)H.

4.3 Results

4.3.1 Kinetics of Enzyme Activation by NAD(P)H

We first established that NAD(P)H reduces AfUGM (Table 4.2). The rate constant for reduction (k_{red}) by NADPH is 3 s^{-1}, and the dissociation constant for NADPH (K_d) is 25 μM. The catalytic efficiency for NADPH is thus 120,000 M^{-1} s^{-1}, which is 20

times higher than that of TcUGM[10] (k_{red} = 0.6 s^{-1}, K_d = 98 μM, k_{red}/K_d = 6000 M^{-1} s^{-1}). The kinetic constants for the reaction of AfUGM with NADH are k_{red} of 0.2 s^{-1} and K_d of 260 μM (k_{red}/K_d = 670 M^{-1} s^{-1}). These results show that NAD(P)H activates AfUGM *in vitro* and that AfUGM, like TcUGM, exhibits a preference for NADPH.

Table 4.2 Kinetic Parameters for the Reduction of AfUGM Mutant Enzymes by NAD(P)H.

	k_{red} (s^{-1})	K_d (μM)	k_{red}/K_d (M^{-1} s^{-1})	**Variational Effect (k_{red}/K_d of variant)/(k_{red}/K_d of AfUGM)**
		NADPH		
AfUGM	2.98 ± 0.078	25 ± 2	120000 ± 10000	
R447A	0.00240 ± 0.00007	39 ± 6	60 ± 10	0.0005 ± 0.0001
R91A	0.313 ± 0.008	330 ± 20	980 ± 70	0.008 ± 0.001
S93A	0.17 ± 0.01	21 ± 8	8000 ± 3000	0.07 ± 0.03
Y317A	3.36 ± 0.09	100 ± 10	34000 ± 3000	0.28 ± 0.03
Y104A	2.20 ± 0.04	54 ± 4	41000 ± 3000	0.34 ± 0.04
		NADH		
AfUGM	0.172 ± 0.003	260 ± 20	670 ± 50	

The stereochemistry of hydride transfer for AfUGM was determined using KIE measurements (Table 4.3). When the pro-R H atom was substituted with D, KIE values on flavin reduction of 3.0 ± 0.1 and 6.0 ± 0.7 were observed for NADPH and NADH, respectively. In contrast, a KIE value of 0.93 ± 0.03 was measured when pro-S NADPD was used to reduce AfUGM. These results show that the pro-R hydride of NADPH is

transferred to AfUGM. KIE measurements also indicate pro-R stereochemistry for TcUGM (Table 4.3).

Table 4.3 Deuterium Kinetic Isotope Effects for Hydride Transfer to AfUGM and TcUGM[a].

	AfUGM	**TcUGM**
NADPH	3.0 ± 0.1	1.85 ± 0.01
NADH	6.0 ± 0.7	3.9 ± 0.2

[a]expressed as the ratio of k_{red} measured using [4R-4-^{2}H]-NAD(P)H to k_{red} measured using NAD(P)H

4.3.2 NAD(P)H Binding Site

Kinetic protein crystallography[26] was used to obtain structures of AfUGM relevant to enzyme activation by NADP(H) (Fig. 4.1B). Short soaks (<1 min) of crystals in ~100 mM NAD(P)H followed by freeze-trapping in liquid nitrogen yielded 2.75 Å resolution structures of the oxidized enzyme ($AfUGM_o$) complexed with NADPH and NADH. The two structures are very similar and have a root-mean-square deviation (RMSD) of 0.31 Å for the Cα atoms of the tetramer. The pairwise RMSDs for the individual chains of the two structures span the range 0.17–0.51 Å.

Electron density maps clearly indicated the presence of NADPH bound in the active site (Fig. 4.2B). The occupancy of NADPH varied among the protomers of the tetramer, with the strongest density appearing in chains A and B. We note that this pattern of differential ligand occupancy (high in protomers A and B, low in protomers C and D) was observed in our studies of substrate and inhibitor binding to AfUGM, which also involved soaking the $P6_522$ crystal form used here[20]. The maps allowed the building of complete models of NADPH at nearly full occupancy (q = 0.9) in chains A and B

(Fig. 4.2B). In chains C and D of the tetramer, the density was strong for the ADP group but weaker for the nicotinamide riboside group. Therefore, only the ADP part of NADPH (q = 0.7) was modeled in chains C and D.

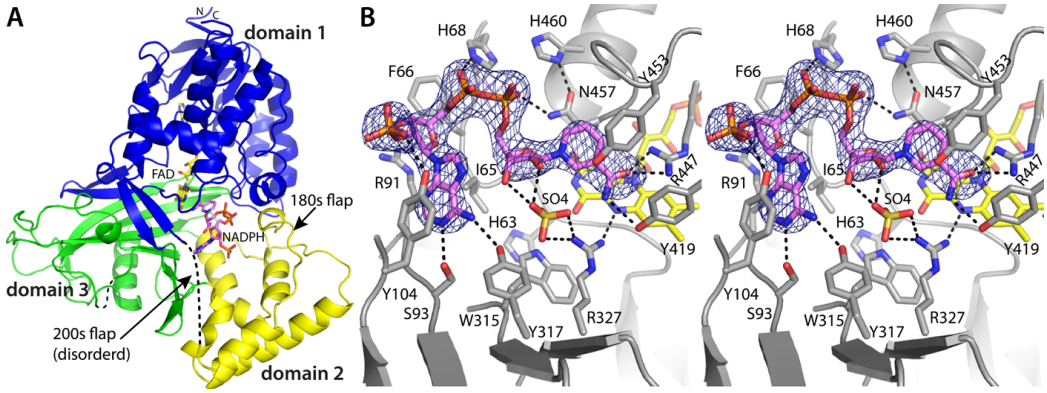

Fig. 4.2 Crystal structure of AfUGMo complexed with NADPH. (A) Protomer structure. FAD and NADPH are colored yellow and pink, respectively. Domains 1, 2, and 3 are colored blue, yellow, and green, respectively. (B) Stereographic view of the NADPH binding site. The cage represents a simulated annealing **σ**A-weighted **F**o – **F**c omit map (3.0 **σ**). Prior to calculating the map, NADPH was omitted and simulated annealing refinement was performed.

Strong density was also observed for NADH (Fig. 4.3). As in the AfUGM$_o$-NADPH structure, differential ligand occupancy was evident. Complete models of NADH were built at q = 0.9 in chains A and B, while NADH was omitted in chains C and D.

NAD(P)H binds to AfUGM$_o$ at the intersection of the three protein domains with the nicotinamide near the *re* face of the isoalloxazine and the rest of the dinucleotide directed toward the 200s substrate-binding flap (Fig. 4.2A). The nicotinamide riboside binding site overlaps that of the Gal*p* moiety of the substrate (Fig. 4.4A), and thus NAD(P)H and substrate binding are mutually exclusive. This result is expected since both NAD(P)H and UDP-Gal*p* require access to the flavin N5 atom at the *re* face of the

isoalloxazine. On the other hand, the AMP group of NAD(P)H and the UMP group of the substrate bind in different pockets on either side of Tyr104 (Fig. 4.4A).

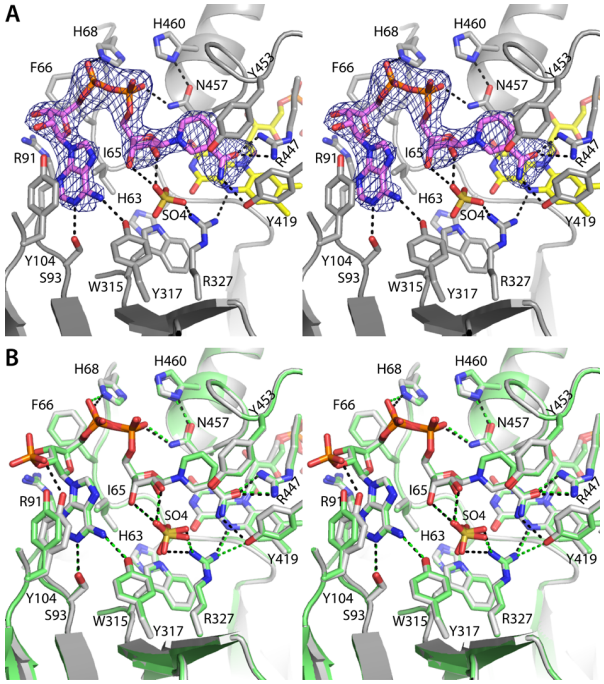

Fig. 4.3 The active site of AfUGM$_o$-NADH (stereographic views). (A) Electron density of NADH. FAD and NADH are colored yellow and pink, respectively, and the cage represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3.0 σ. Prior to calculation of the map, NADH was omitted and simulated annealing refinement was performed. (B) Superposition of AfUGM$_o$-NADH (green) and AfUGM$_o$-NADPH (white).

NAD(P)H forms several interactions with AfUGM$_o$ (Fig. 4.2B and Fig. 4.3). The adenosine moiety occupies a hydrophobic pocket formed by Ile65, Phe66, Tyr104, and the nonpolar part of the side chain of Arg91. The latter residue serves as the floor of the adenosine binding pocket along with the main chain of Ile92 and the side chain of Ser93. The adenine base forms hydrogen bonds with Ser93 and Tyr317. Tyr104 of the pocket makes a hydrogen bond with the 2′-phosphoryl of NADPH. This side chain is rotated slightly in the NADH complex because of the absence of the 2′-phosphoryl in NADH (Fig. 4.3B). Otherwise, the binding sites of the NADPH and NADH complexes are

identical within experimental error (Fig. 4.3B). The pyrophosphate is stabilized by electrostatic interactions with His68 and Asn457. Finally, the nicotinamide ring is wedged between Tyr453 and Asn457, and its carboxamide forms hydrogen bonds with Arg447 and Tyr419.

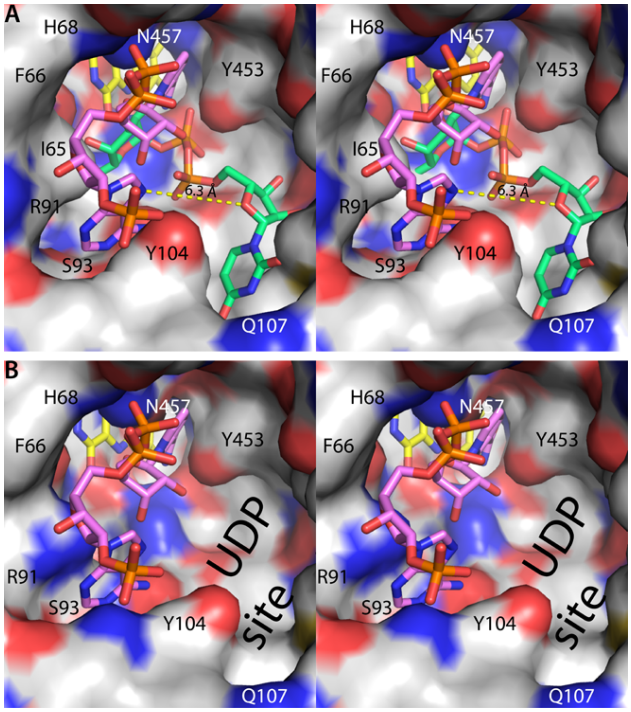

Fig. 4.4 Spatial proximity of the NADPH and substrate binding sites (stereographic views). In both panels, the view is from the outside of the enzyme looking into the active site. (A) Active site surface of AfUGM$_o$-NADPH. FAD and NADPH are colored yellow and pink, respectively. For reference, the AfUGM$_o$-NADPH structure has been overlaid with the AfUGMr-UDP-Galp complex (PDB code 3UTH[20]) and the UDP-Galp (green) included in this image. The dashed yellow line denotes the 6.3 Å distance between the adenine and uracil sites. (B) Active site surface of AfUGMr. For reference, the AfUGMo- NADPH structure has been overlaid with AfUGMr and the NADPH (pink) included in this image.

The nicotinamide is not optimally aligned for hydride transfer in the AfUGM$_o$-NADPH and AfUGM$_o$-NADH complexes. For reference, glutathione reductase, which also binds FAD using a Rossmann fold domain, demonstrates the expected orientation of the nicotinamide relative to the isoalloxazine (PDB code 1GRB)[32]. In glutathione

reductase, the nicotinamide stacks in parallel with the middle ring of the isoalloxazine such that the hydride transfer partners, C4 of NADPH and N5 of FAD, are separated by 3.2 Å. In AfUGM, the nicotinamide and isoalloxazine rings are not parallel, and the C4–N5 distance is 5.8 Å (Fig. 4.2B and Fig. 4.3). The observed arrangement of the hydride transfer partners may result from the adventitious binding of a sulfate ion in the active site (Fig. 4.2B and Fig. 4.3). Alternatively, the observed complex could represent a transient species that precedes the active hydride transfer complex. This possibility is supported by the observation that the 30-min soak in NADPH generated the reduced enzyme with the dinucleotide dissociated (*vide infra*), implying that NADPH moves from the observed conformation to the active one. Baeyer–Villiger monooxygenases provide precedent for this phenomenon, which is known as the "sliding cofactor" model [33-34].

Guided by the constraint of pro-R stereochemistry determined from KIE measurements, we generated a model of the hydride transfer complex. Using only dihedral angle rotations and keeping the ADP fixed, it was possible to bring the *re* face of the nicotinamide in contact with the isoalloxazine such that the hydride transfer partners are separated by 3.1 Å (Fig. 4.5).

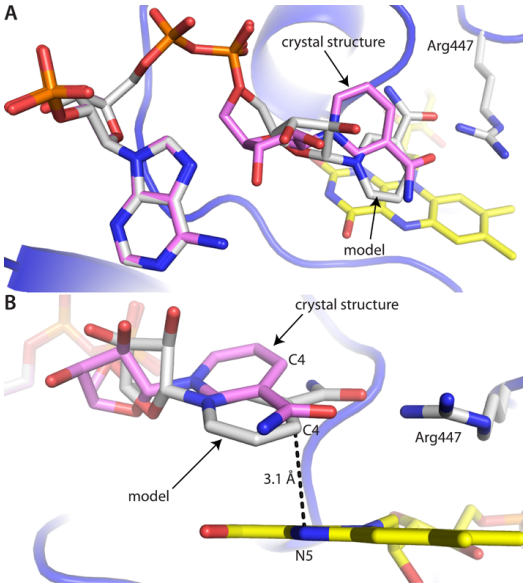

Fig. 4.5 Two views of the model of the hydride transfer complex between NADPH and AfUGM$_o$. NADPH from the crystal structure is colored pink, and the model is shown in white. The model was created by manual torsional adjustment of the nicotinamide riboside moiety in COOT to bring the *re* face of the nicotinamide in contact with the FAD isoalloxazine, which is consistent with kinetic isotope effect data implicating pro-R stereochemistry.

4.3.3 Assignment of the Flavin Redox State

As described previously, the conformations of oxidized and dithionite-reduced AfUGM are dramatically different, and these structural differences can be used to deduce the flavin redox state in lieu of microspectrophotometry [20-22]. Briefly, the oxidation state of the FAD can be deduced from the conformation of the conserved histidine loop (Gly61-Gly62-His63), the identity of the hydrogen-bonding partner of the flavin N5 (Arg327 for FAD, Gly62 for $FADH^-$), the orientation of Trp315, and the curvature of the isoalloxazine (planar for FAD, bent by 7° for $FADH^-$).

To establish the conformation of the NADPH-reduced enzyme, the 2.2 Å resolution structure of AfUGM was determined from a crystal of the oxidized enzyme that had been soaked for 30 min in 100 mM NADPH. The yellow color of the crystal was

completely bleached, which is indicative of full reduction of the crystalline enzyme. The NADPH-reduced and dithionite-reduced (PDB code 3UTF) structures are very similar (RMSD = 0.24 Å). Furthermore, the structural indicators of the flavin redox state in all four chains of the NADPH-reduced enzyme are consistent with the reduced flavin state. In particular, His63 is located on the *si* face of the isoalloxazine, where it forms hydrogen bonds with Gly61 and the ribityl chain (Fig. 4.6A). Furthermore, Gly62 accepts a hydrogen bond from the flavin N5 atom. This interaction is diagnostic of the reduced flavin because the main chain carbonyl is an obligate hydrogen bond acceptor, and N5 is a donor only in the reduced state. Lastly, at 2.2 Å resolution it is possible to discern the 7° bend of the isoalloxazine (Fig. 4.6A). All of these structural features are also observed in the structure of dithionite-reduced AfUGM, as shown by an overlay of the NADPH-reduced and dithionite-reduced active sites (Fig. 4.6A).

The structural indicators of the redox state clearly show that the flavin is oxidized in the NAD(P)H complexes. At 2.75 Å resolution, it is difficult to discern the curvature of the isoalloxazine, so the protein conformation is a better indicator of the flavin redox state. In particular, His63 is near the pyrimidine portion of the isoalloxazine and oriented parallel to Trp315 (Fig. 4.6B). As noted by van Straaten et al.[22], this particular arrangement of these two side chains occurs in $AfUGM_o$. We note that the position of His63 in $AfUGM_o$ and $AfUGM_r$ differs by 7 Å (Fig. 4.6C), which is clearly resolvable at 2.75 Å resolution. Furthermore, Arg327 donates a hydrogen bond to the flavin N5 in $AfUGM_o$-NADPH (Fig. 4.6B). This hydrogen bond is diagnostic of the oxidized enzyme because Arg is an obligate hydrogen bond donor and N5 is an obligate acceptor only in the oxidized state. In summary, the flavin is oxidized in the structures determined from

the short soaks in NAD(P)H, implying that the trapped species represents the oxidized enzyme complexed with the reduced coenzyme.

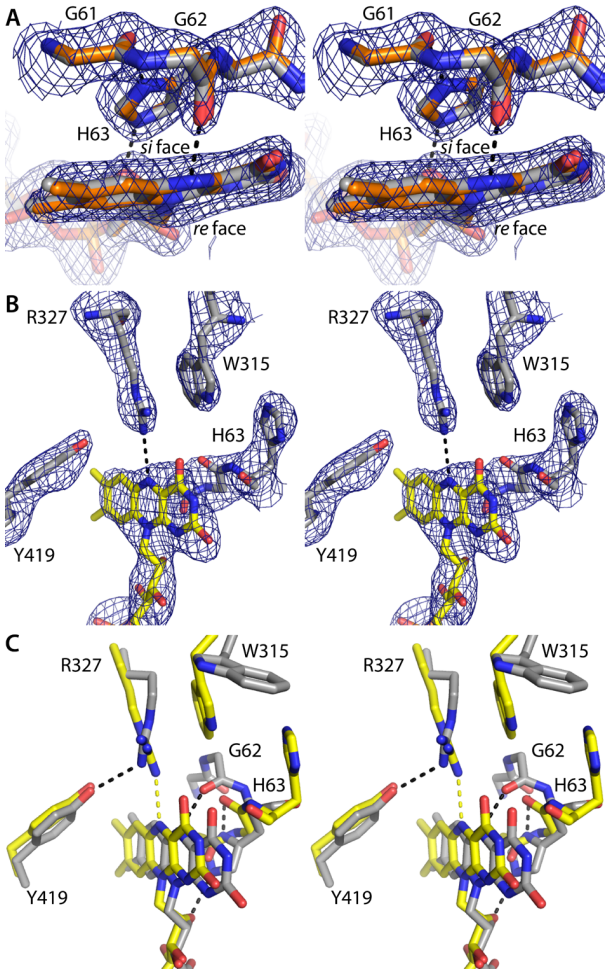

Fig. 4.6 Structural features used to assign the flavin redox state (stereographic views). In panels A and B, the cage represents a simulated annealing σ_A-weighted F_o - F_c omit map (3.0 σ). Prior to calculating the maps, the flavin, His loop, Trp315, Arg327, and Tyr419 were omitted and simulated annealing refinement was performed. (A) Superposition of NADPH-reduced AfUGM (gray) and dithionite-reduced AfUGM (orange, PDB code 3utf). The hydrogen bond between Gly62 and the flavin N5 atom is diagnostic of the reduced state.{Dhatwalia, 2012 #3200;Dhatwalia, 2012 #78} (B) $AfUGM_o$-NADPH. The Arg327-N5 hydrogen bond and the location of His63 near Trp315 are indicative of the oxidized enzyme.{Dhatwalia, 2012 #3200;Dhatwalia, 2012 #78} (C) Superposition of $AfUGM_r$ (gray) and $AfUGM_o$-NADPH (yellow). Black and yellow dashes indicate hydrogen bonds in $AfUGM_r$ and $AfUGM_o$-NADPH, respectively.

4.3.4 Mutagenesis Studies

Site-directed mutagenesis (to Ala) was used to assess the importance of selected residues in the NAD(P)H binding site (Table 4.2). Three mutations have pronounced effects on the kinetics of reduction (R447A, R91A, and S93A). Mutation of Arg447, which interacts with the nicotinamide carboxamide, lowers the catalytic efficiency by a factor of 2000. This result is consistent with Arg447 helping to guide the nicotinamide into position for hydride transfer. Arg91 and Ser93 are on the 90s loop, which forms the base of the adenine binding pocket. The R91A and S93A mutations decrease efficiency 125-fold and 14-fold, respectively. These results are consistent with the 90s loop helping to anchor the adenosine moiety. Mutation of Tyr104 or Tyr317, which interact with the adenine base, decreases the catalytic efficiency 3-fold. This result is consistent with these residues playing ancillary roles in binding the adenine. In summary, Arg447 and the 90s loop, which are located at opposite ends of the NADPH binding site, appear to be essential for efficient activation of AfUGM by NADPH.

4.4 Discussion

The structures reported here provide the first images of a UGM complexed with NAD(P)H. All of the residues within 3.9 Å of NADPH are identically conserved in other eukaryotic UGMs, including TcUGM and *Leishmania major* UGM, two enzymes that are of interest for the design of drugs to treat Chagas disease and leishmaniasis, respectively (Fig. 4.7). This level of sequence conservation suggests that the identified binding site is physiologically relevant and present in other eukaryotic UGMs. Site-directed mutagenesis provides additional validation of the site; mutation of residues contacting

NADPH decreases the catalytic efficiency of FAD reduction by factors of 3–2000 (Table 3.2).

The NAD(P)H site of AfUGM is probably different from that of bacterial UGMs. Notably, none of the residues that contact the AMP half of the dinucleotide are present in the sequences of bacterial UGMs. Furthermore, overlaying NADPH from our structure onto *K. pneumoniae* UGM reveals severe steric clashes with an active site loop (Fig. 4.8). The absence of the high affinity NADPH site described here for AfUGM may explain the low activity of bacterial UGMs[13] with NADPH. For example, we have estimated k_{red} for *Mycobacterium tuberculosis* UGM to be only 0.00002 s^{-1} (Supplementary Figure S5), which is several orders of magnitude slower than AfUGM and TcUGM.

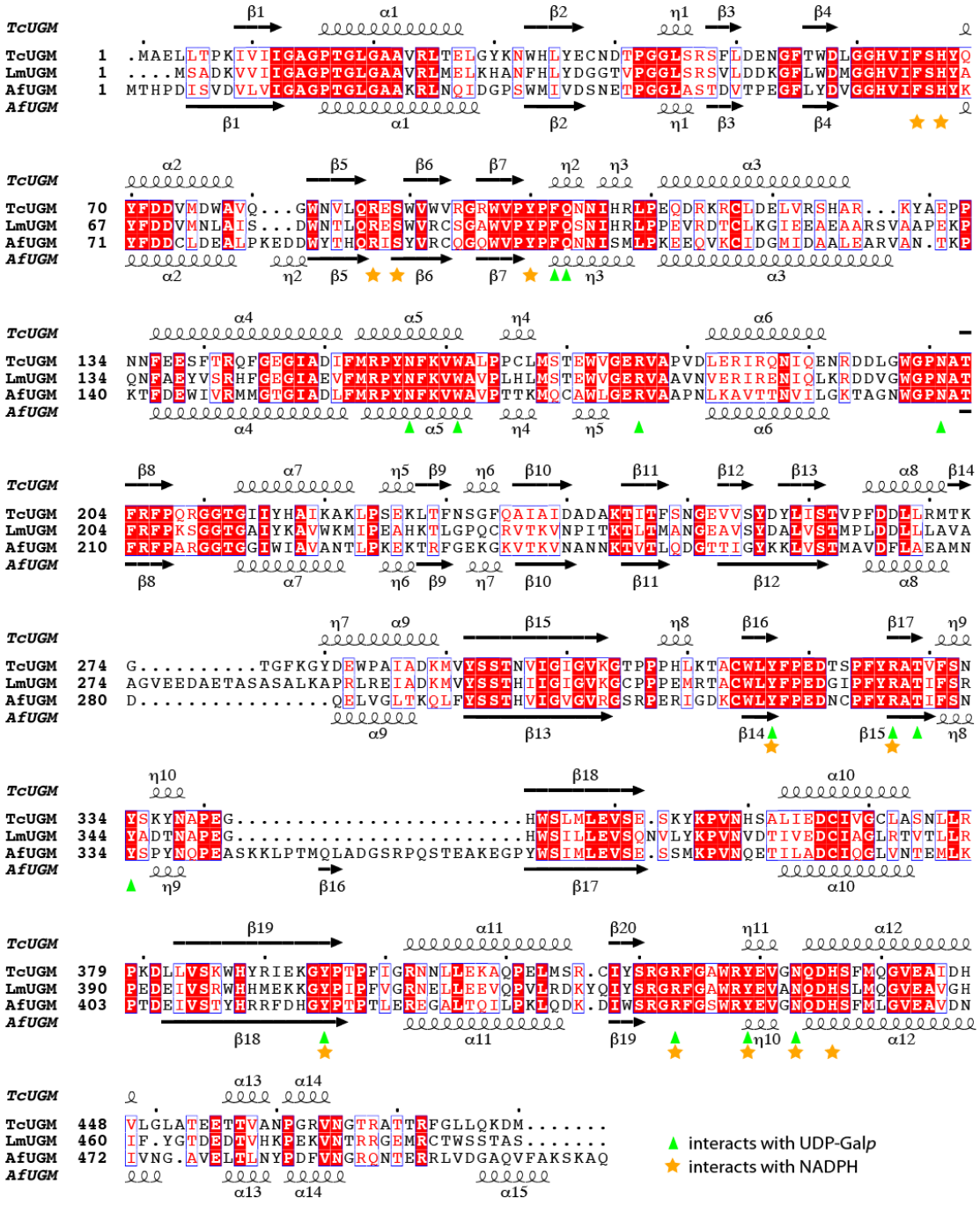

Fig. 4.7 Sequence alignment of the UGMs from *Trypanosoma cruzi*, *Leishmania major*, and *Aspergillus fumigatus*. The secondary structure elements above and below the alignment were derived from the structures of *T. cruzi* UGM and AfUGM, respectively. This figure was created with CLUSTALW2 {Chenna, 2003 #3146} and ESPript {Gouet, 2003 #3144}.

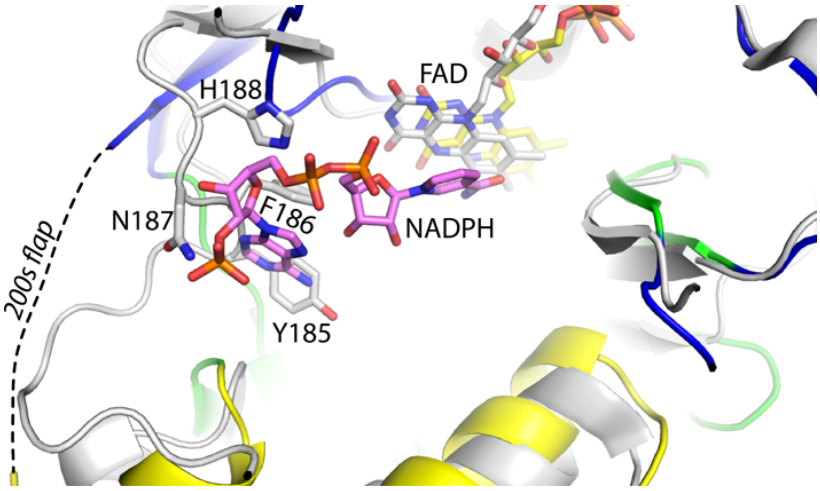

Fig. 4.8 Superposition of the AfUGM$_o$-NADPH complex and oxidized KpUGM (PDB 2BI7). AfUGM$_o$ is shown with domains 1, 2, and 3 colored blue, yellow, and green as in Figure 2A, NADPH is colored pink, and FAD is colored yellow. KpUGM is colored white. Selected side chains of the 180s loop of KpUGM are shown. Note that these residues occupy the space corresponding to the AMP group of NADPH, which suggests that the NADPH binding sites of KpUGM and AfUGM are different.

Finally, the AfUGM$_o$-NADPH structure provides new opportunities for inhibitor design. One possible strategy is to target the NADPH site of AfUGM$_o$ in order to lock the enzyme in the inactive state. Alternatively, the active, reduced enzyme could be targeted because the ADP site is also present in AfUGM$_r$ (Fig. 4.4B). Retention of the ADP site in the reduced enzyme reflects the fact that several residues that contact the ADP moiety have similar conformations in AfUGM$_o$ and AfUGM$_r$, including Ph66, His68, Arg91, Ser93, Tyr104, Tyr317, Asn457, and His460. Both strategies are attractive because one could repurpose existing adenosine analogues. Lastly, although the binding of NADPH and UDP-Gal*p* is mutually exclusive, the adenine and uracil pockets are distinct and separated by 6 Å (Fig. 4.4). Thus, it may be possible to design novel bidentate compounds that simultaneously access both pockets.

4.5 References

1. Richards, M. R.; Lowary, T. L. *Chembiochem* **2009**, *10*, 1920.

2. Kizjakina, K.; Tanner, J. J.; Sobrado, P. *Curr. Pharm. Des.* **2012**, In press.

3. Oppenheimer, M.; Valenciano, A. L.; Sobrado, P. *Enzyme Res.* **2011**, *2011*, 1.

4. Tefsen, B.; Ram, A. F.; van Die, I.; Routier, F. H. *Glycobiology* **2012**, *22*, 456.

5. Schmalhorst, P. S.; Krappmann, S.; Vervecken, W.; Rohde, M.; Muller, M.; Braus, G. H.; Contreras, R.; Braun, A.; Bakker, H.; Routier, F. H. *Eukaryot. Cell.* **2008**, *7*, 1268.

6. Kleczka, B.; Lamerz, A. C.; van Zandbergen, G.; Wenzel, A.; Gerardy-Schahn, R.; Wiese, M.; Routier, F. H. *J Biol Chem* **2007**, *282*, 10498.

7. Bornemann, S. *Nat. Prod. Rep.* **2002**, *19*, 761.

8. Soltero-Higgin, M.; Carlson, E. E.; Gruber, T. D.; Kiessling, L. L. *Nat Struct. Mol. Biol.* **2004**, *11*, 539.

9. Sun, H. G.; Ruszczycky, M. W.; Chang, W. C.; Thibodeaux, C. J.; Liu, H. W. *J Biol Chem* **2011**.

10. Oppenheimer, M.; Valenciano, A. L.; Kizjakina, K.; Qi, J.; Sobrado, P. *PLoS One* **2012**, *7*, e32918.

11. Massey, V. *Biochem. Soc. Trans.* **2000**, *28*, 283.

12. Koplin, R.; Brisson, J. R.; Whitfield, C. *J Biol Chem* **1997**, *272*, 4121.

13. Barlow, J. N.; Marcinkeviciene, J.; Blanchard, J. S. In *Enzymatic Mechanisms*; Frey, P. A., Northrop, D. B., Eds.; IOS Press: Amsterdam, 1999, p 98.

14. Sanders, D. A.; Staines, A. G.; McMahon, S. A.; McNeil, M. R.; Whitfield, C.; Naismith, J. H. *Nat. Struct. Biol.* **2001**, *8*, 858.

15. Beis, K.; Srikannathasan, V.; Liu, H.; Fullerton, S. W.; Bamford, V. A.; Sanders, D. A.; Whitfield, C.; McNeil, M. R.; Naismith, J. H. *J. Mol. Biol.* **2005**, *348*, 971.

16. Gruber, T. D.; Borrok, M. J.; Westler, W. M.; Forest, K. T.; Kiessling, L. L. *J Mol Biol* **2009**, *391*, 327.

17. Gruber, T. D.; Westler, W. M.; Kiessling, L. L.; Forest, K. T. *Biochemistry* **2009**, *48*, 9171.

18. Partha, S. K.; van Straaten, K. E.; Sanders, D. A. *J. Mol. Biol.* **2009**, *394*, 864.

19. Partha, S. K.; Sadeghi-Khomami, A.; Slowski, K.; Kotake, T.; Thomas, N. R.; Jakeman, D. L.; Sanders, D. A. *J. Mol. Biol.* **2010**, *403*, 578.

20. Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Karr, D. B.; Nix, J. C.; Sobrado, P.; Tanner, J. J. *J Biol Chem* **2012**, *287* 9041.

21. Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Sobrado, P.; Tanner, J. J. *Biochemistry* **2012**, *51*, 4968.

22. van Straaten, K. E.; Routier, F. H.; Sanders, D. A. *J Biol Chem* **2012**, *287*, 10780.

23. Weiss, M. *J. Appl. Cryst.* **2001**, *34*, 130.

24. Engh, R. A.; Huber, R. *Acta Cryst.* **1991**, *A47*, 392.

25. Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437.

26. Bourgeois, D.; Royant, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 538.

27. Kabsch, W. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125.

28. Evans, P. *Acta Cryst.* **2006**, *D62*, 72.

29. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr., Sect. D* **2010**, *66*, 213.

30. Emsley, P.; Cowtan, K. *Acta Cryst.* **2004**, *D60*, 2126.

31. Oppenheimer, M.; Poulin, M. B.; Lowary, T. L.; Helm, R. F.; Sobrado, P. *Arch Biochem Biophys* **2010**, *502*, 31.

32. Karplus, P. A.; Schulz, G. E. *J. Mol. Biol.* **1989**, *210*, 163.

33. Mirza, I. A.; Yachnin, B. J.; Wang, S.; Grosse, S.; Bergeron, H.; Imura, A.; Iwaki, H.; Hasegawa, Y.; Lau, P. C.; Berghuis, A. M. *J. Am. Chem. Soc.* **2009**, *131*, 8848.

34. Franceschini, S.; van Beek, H. L.; Pennetta, A.; Martinoli, C.; Fraaije, M. W.; Mattevi, A. *J Biol Chem* **2012**, *287*, 22626.

Chapter 5

Structure determination and small angle X-ray studies of *Leishmania major* UDP-Galactopyranose mutase

5.1 Introduction

Recent advances in targeting neglected tropical diseases have led to the identification of several enzymes involved in the pathogenesis and virulence of protozoan parasite *Leishmania major*[1]. Among these enzymes is UDP-galactopyranose mutase (UGM) that catalyzes the reversible reaction between UDP-galactopyranose (UDP-Galp) and UDP-galactofuranose (UDP-Galf) (Fig. 5.1). *L. major* is the known cause of leishmaniasis (subcutaneous or visceral) affecting mainly inhabitants of tropical and subtropical regions. It is estimated that more than 2 million people are affected worldwide with leishmaniasis resulting in ~50,000 deaths annually[1b]. Current treatments of leishmaniasis are either highly toxic or expensive[1a].

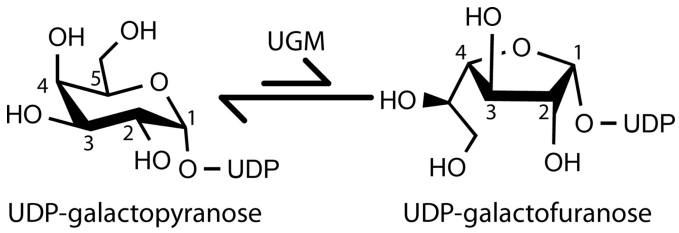

Fig. 5.1 Reaction catalyzed by UGM.

The cell wall of *L. major* is composed of lipophosphoglycans (LPG) and glycoinositolphospholipids (GIPLs) that play a crucial role in the survival of the parasite inside human host[2]. Galactofuranose (Galf) forms a major component of these cell wall and extracellular structures. Deletion of the gene encoding UGM (*glf*) in a mouse model shows a delayed onset of lesions at the site of infection thus, UGM is essential for the virulence of *L. major*[2].

Leishmania major UDP-galactopyranose mutase (LmUGM) shares ~40% sequence homology with *Aspergillus fumigatus* UGM (AfUGM) and ~60% with *Trypanosoma cruzi* UGM (TcUGM). Earlier structures from AfUGM and TcUGM have shown that the active sites of these two eukaryotic UGMs are identical[3]. The determination of the structure of LmUGM would help us identify the similarity in the active sites of eukaryotic UGMs and thus, developing a common strategy for the inhibition of these enzymes.

5.2 Materials and Methods

5.2.1 Crystallization trials

Several crystallization trials were carried out for LmUGM. Prior to setting crystallization trays, protein was incubated with 0.5 mM THP. Equal amounts of protein (at 8 mg/ml in 25 mM HEPES, 125 mM NaCl pH 7.5) and precipitant were used for crystallization trials with and without ligand UDP and substrate UDP-Galp. Plate-shaped crystals were obtained within 3-4 days in 19% PEG 3350, 0.2 M ammonium acetate, 0.1 M Bis-Tris pH 6.5. The crystals of the reduced enzyme were obtained after soaking these crystals with 80 mM dithionite and 50 mM UDP. The crystals diffracted to 2.4 Å resolution at APS 24-ID-C beamline but had severe pathology such that the density of cofactor FAD could not be traced.

Attempts were made to improve the existing crystal form and obtain new crystal forms using techniques such as reductive methylation, limited proteolysis and buffer optimization. Also, several crystallization trials were carried out for homologous UGMs from *Leishmania mexicana* and *Leishmania infantum*.

Table 5.1 X-ray diffraction, data collection and refinement statistics[a].

Enzyme	Native
Active site ligand	UDP
Space group	$P3_121$
Unit cell lengths (Å)	a= 82.20, c= 129.23
Wavelength	1.12711
Resolution (Å)	19.74 – 2.41 (2.54 – 2.41)
No. of observations	71316
No. of unique reflections	20017
$R_{merge}(I)$	0.048 (0.283)
R_{meas}	0.065 (0.380)
R_{pim}	0.042 (0.252)
Average I/σ	17.0 (3.9)
Completeness (%)	99.3 (99.4)
Redundancy	3.6 (3.6)
R_{cryst}	0.295
R_{free}[b]	0.344
No. of protein residues	419
No. of protein atoms	3129
Average B-factor (Å^2)	
Protein	55.65
FDA	43.07
UDP	52.32
rmsd[c]	
Bond lengths (Å)	0.003
Bond angles (deg)	0.783
Ramachandran plot[d]	
Favored (%)	93.8%
Allowed (%)	5.7%
Outliers (%)	0.5%

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]A common set of test reflections (5 %) was used for refinement of all structures.
[c]Compared to the parameters of Engh and Huber[8].
[d]The Ramachandran plot was generated with RAMPAGE[9].

5.2.2 X-ray diffraction, data collection and refinement

Several crystals were screened to obtain a good crystal form without any pathology. The crystals of LmUGM-UDP complex in reduced state diffracted up to 2.4 Å resolution at the APS 24-ID-C beamline (Table 5.1). The structure was solved with

MOLREP[4] using a chainsaw model of TcUGM-UDP reduced as the starting point. Clear density for the cofactor FAD and ligand UDP was observed. The structure was manually built using COOT[5] and then refined using PHENIX[6]. Unlike other eukaryotic UGMs that crystallized with high solvent content of 74%, LmUGM crystallized with one molecule in the asymmetric unit with a solvent content of 44.8% and Vm 2.23 $Å^3$/Da, as suggested by Matthew's probability[7].

5.2.3 Small-angle X-ray scattering (SAXS)

SAXS experiments were performed at beamline 12.3.1 of the Advanced Light Source via the mail-in program[10]. Prior to data collection, the protein sample was subjected to size exclusion chromatography using a Superdex 200 column that had been equilibrated in 50mM Tris, 500mM NaCl, 1mM THP, 5mM UDP pH 8.0 buffer. Scattering intensities were measured at three nominal protein concentrations (2 - 6 mg/mL). For each protein concentration, exposure times of 0.5, 1.0, 3.0, and 6.0 sec were used. Scattering curves collected from the protein samples were corrected for background scattering using intensity data collected from the dialysis buffer. A composite scattering curve was generated with PRIMUS[11] by scaling and merging the low q region from the lowest concentration sample (0.5 sec exposure) with the high q region from the lowest concentration sample (0.5 sec exposure). PRIMUS was also used to perform Guinier analysis. GNOM was used to calculate pair distribution functions.[12] FoXS was used to calculate theoretical scattering profiles from atomic models.[13] MOLEMAN was used to calculate R_g from atomic coordinates. The SASTBX server[14] was used for shape reconstruction calculations.

5.3 Results

5.3.1 SAXS results

The oligomeric state and quaternary structure were determined using SAXS. Guinier analysis indicated a radius of gyration (R_g) of 24.8 ± 0.1 Å (Fig. 5.3, inset). Calculations of the pair distribution function suggest R_g of 24.4 – 24.8 Å and maximum particle dimension of 70 – 80 Å. For reference, a monomer of TcUGM has R_g of 22.8 Å, which suggests that LmUGM is monomeric. Indeed, the SAXS curve calculated from a TcUGM monomer exhibits good agreement with the experimental curve (Fig. 5.3), and the SAXS shape reconstruction is consistent with the size and shape of the TcUGM monomer (Fig. 5.3). It is concluded that LmUGM is monomeric in solution and has the same fold as TcUGM.

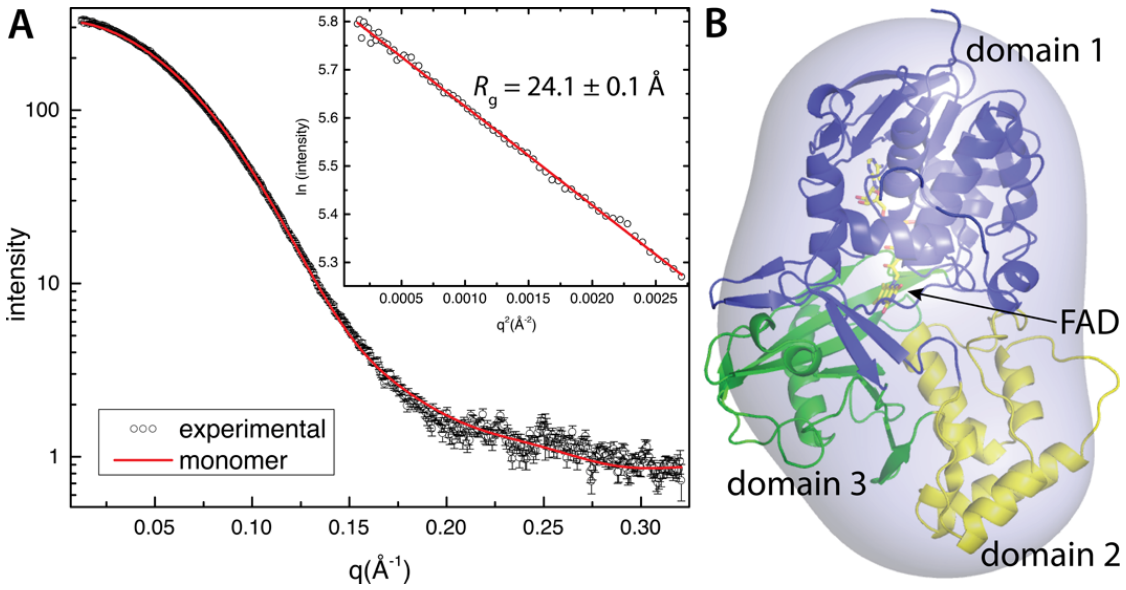

Fig. 5.2 SAXS analysis of LmUGM. (A) Experimental and calculated SAXS curves. The inset shows a Guinier plot spanning the range of qRg of 0.309 – 1.29. The best fit line has R^2 of 0.9977). (B) Superposition of the SAXS shape reconstruction and a monomer of TcUGM. The correlation coefficient between the envelope and the model is 0.94.

5.3.2 Structure of LmUGM complexed with UDP

The structure of LmUGM was solved at 2.4 Å in complex with UDP in the reduced state. LmUGM shares a common fold as other eukaryotic UGMs and is the second structure from a pathogenic protozoan. The binding of inhibitor UDP in the active site of enzyme is identical to that in AfUGM and TcUGM as shown by superposition of the three eukaryotic UGM structures (Fig. 5.4). Thus, LmUGM makes similar interactions with UDP as observed earlier[3] (Fig. 5.4). The 170s flap is in the closed conformation but the residues starting from 180 to 202, including the 200s flap are disordered in the structure. The conformation of histidine loop is the same as observed earlier for eukaryotic UGMs in the reduced state (PDB code: 3UTF[3a] and 4DSH[3b]). The conserved His59 is present on the *si* side of flavin isoalloxazine and the carbonyl oxygen of Gly58 forms a hydrogen bond with N5 of flavin ($FADH^-$).

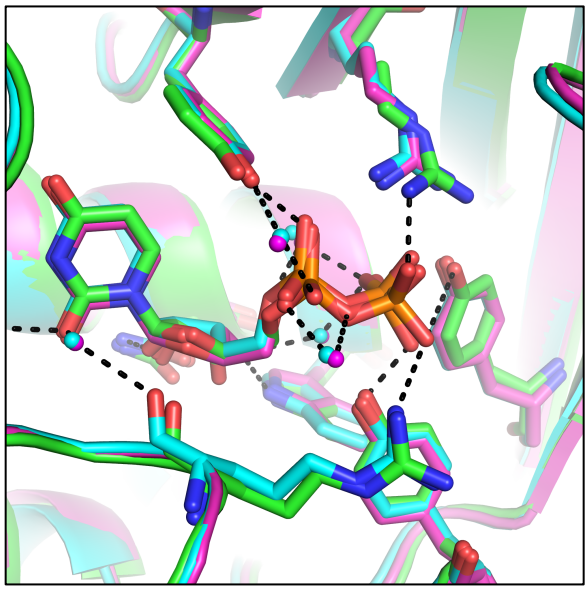

Fig. 5.3 Superposition of the active sites of LmUGM (green), TcUGM (magenta) and AfUGM (cyan). UDP is bound in the active and all the residues involved in interactions are shown as sticks.

5.4 References

1. (a) Kedzierski, L.; Sakthianandeswaren, A.; Curtis, J. M.; Andrews, P. C.; Junk, P. C.; Kedzierska, K., Leishmaniasis: current treatment and prospects for new drugs and vaccines. *Curr Med Chem* **2009,** *16* (5), 599-614; (b) Oppenheimer, M.; Valenciano, A. L.; Sobrado, P., Isolation and characterization of functional Leishmania major virulence factor UDP-galactopyranose mutase. *Biochem Biophys Res Commun* **2011,** *407* (3), 552-6; (c) Damerow, S.; Lamerz, A. C.; Haselhorst, T.; Fuhring, J.; Zarnovican, P.; von Itzstein, M.; Routier, F. H., Leishmania UDP-sugar pyrophosphorylase: the missing link in galactose salvage? *J Biol Chem* **2010,** *285* (2), 878-87.

2. Kleczka, B.; Lamerz, A. C.; van Zandbergen, G.; Wenzel, A.; Gerardy-Schahn, R.; Wiese, M.; Routier, F. H., Targeted gene deletion of Leishmania major UDP-galactopyranose mutase leads to attenuated virulence. *J Biol Chem* **2007,** *282* (14), 10498-505.

3. (a) Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Karr, D. B.; Nix, J. C.; Sobrado, P.; Tanner, J. J., Crystal structures and small-angle x-ray scattering analysis of UDP-galactopyranose mutase from the pathogenic fungus Aspergillus fumigatus. *J Biol Chem* **2012,** *287* (12), 9041-51; (b) Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Sobrado, P.; Tanner, J. J., Crystal structures of Trypanosoma cruzi UDP-galactopyranose mutase implicate flexibility of the histidine loop in enzyme activation. *Biochemistry* **2012,** *51* (24), 4968-79.

4. Vagin, A.; Teplyakov, A., Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 1), 22-5.

5. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.

6. (a) Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **2002,** *58* (Pt 11), 1948-54; (b) Zwart, P. H.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; McKee, E.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C.; Adams, P. D., Automated structure solution with the PHENIX suite. *Methods Mol Biol* **2008,** *426*, 419-35.

7. (a) Matthews, B. W., Solvent content of protein crystals. *J Mol Biol* **1968,** *33* (2), 491-7; (b) Notomista, E.; Lahm, A.; Di Donato, A.; Tramontano, A., Evolution of bacterial and archaeal multicomponent monooxygenases. *J Mol Evol* **2003,** *56* (4), 435-45.

8. Engh, R. A. H., R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst*. **1991,** *A47* (4), 9.

9. Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003,** *50* (3), 437-50.

10. (a) Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, F. L., 2nd; Tsutakawa, S. E.; Jenney, F. E., Jr.; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S. J.; Scott, J. W.; Dillard, B. D.; Adams, M. W.; Tainer, J. A., Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **2009,** *6* (8), 606-612; (b) Classen, S.; Hura, G. L.; Holton, J. M.; Rambo, R. P.; Rodic, I.; McGuire, P. J.; Dyer, K.; Hammel, M.; Meigs, G.; Frankel, K. A.; Tainer, J. A., Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **2013,** *46* (Pt 1), 1-13.

11. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I., PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2003,** *36* (5), 1277-1282.

12. Svergun, D., Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **1992,** *25* (4), 495-503.

13. Schneidman-Duhovny, D.; Hammel, M.; Sali, A., FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **2010,** *38* (Web Server issue), W540-4.

14. Liu, H.; Hexemer, A.; Zwart, P. H., The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. *J. Appl. Crystallogr.* **2012,** *45* (3), 587-593.

Chapter 6

Expression, purification and crystallization of *Legionella pneumophila* histidine acid phosphatase

Richa Dhatwalia, Harkewal Singh, Thomas J. Reilly, John J. Tanner

6.1 Introduction

Histidine acid phosphatases (HAPs) are enzymes that are optimally active at acidic pH and utilize an active site histidine residue for catalysis. A general scheme of the reaction catalyzed by HAPs is shown below (Fig. 6.1).

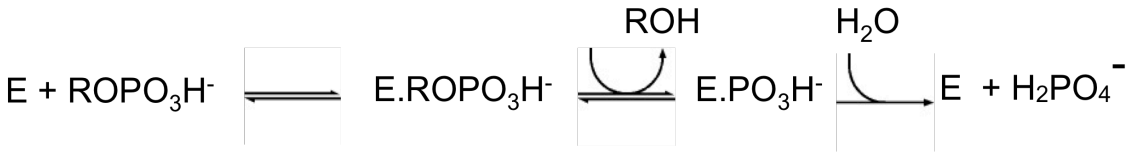

Fig. 6.1 An enzyme- catalyzed reaction with a phosphomonoester substrate.

The active site histidine acts as a nucleophile and is phosphorylated during the reaction (Fig. 6.2). The alkoxide group (RO^-) is then transferred to a nearby proton donor forming a phosphoenzyme intermediate with the further transfer of phosphoryl group to water at acidic pH.

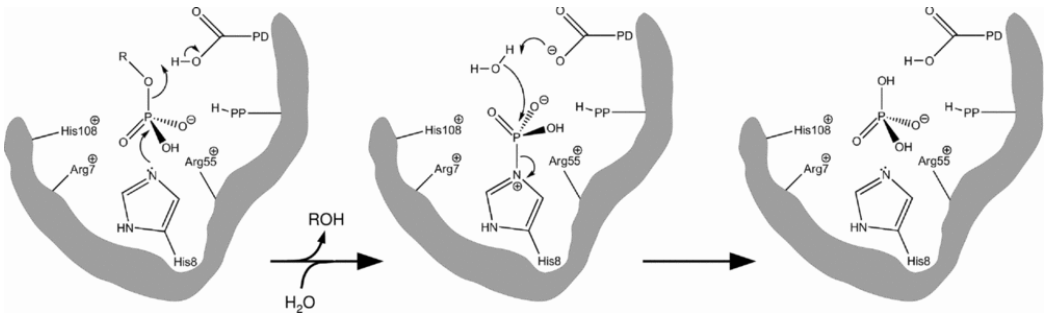

Fig. 6.2 Catalytic mechanism of the histidine phosphatase superfamily[1].

HAPs are present in bacteria as well as humans. An example of HAPs is human prostatic acid phosphatase (PAP) that has recently been discovered to be a pain suppressor[2]. Studies show that PAP posseses potent and long-lasting antinociceptive

properties and is eight times more potent than the opiod analgesic morphine. Furthermore, the antinociceptive properties of PAP have been attributed to its ability to catalyze the dephosphorylatation of 5'-AMP to generate adenosine that activates A_1-adenosine receptors in dorsal spinal cord[2]. An intrathecal injection of hPAP into the wild-type mice showed increased thermal and mechanical sensitivity. The paw-withdrawal latency of mice to a thermal stimulus increased up to six hours and remained elevated up to 3 days[3].

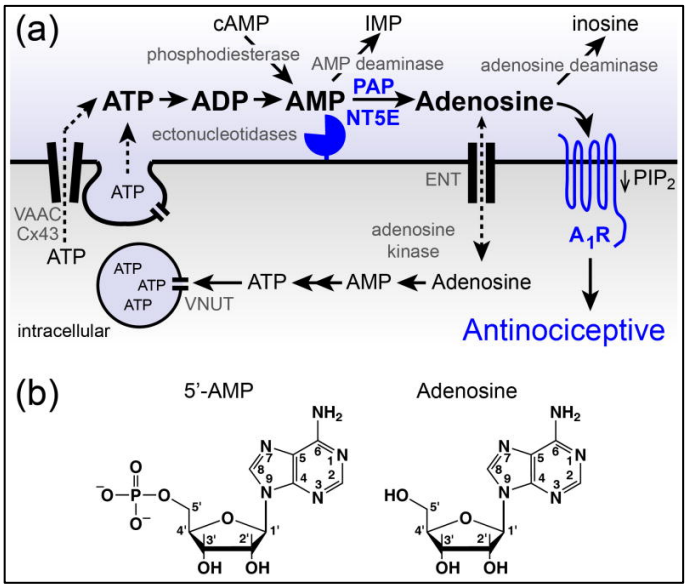

Fig. 6.3 Proteins that regulate extracellular adenosine levels also influence adenosine receptor activation [4].

On the contrary, adenosine itself is also known to have analgesic properties in mammals but has several disadvantages to it. First, although it has antinociceptive effects in humans, it is not currently used to treat chronic pain in patients because of cardiovascular side effects [4]. Second, the basal concentration of adenosine ranges from 50-200 nM and has a short half-life in blood (few seconds) and spinal cerebrospinal fluid (10-20 min)[4]. Thus, due to its rapid metabolism it is not a very efficient analgesic.

Interestingly, ectonucleotidases are stable *in vivo* and have A_1R-dependent antinociceptive effects that last for 2-3 days and might have sustained analgesic effects in patients. Moreover, as enzymes are catalytically restricted, ectonucleotidases would generate adenosine in proportion to the substrate (5'-AMP) availability. Thus, it will have fewer side effects as compared to adenosine itself, which is effective only when given in higher doses. Also, ATP levels are increased post injury providing additional substrate for ectonucleotidases to generate adenosine[4].

These studies provoke the question that whether bacterial HAPs also posses antinociceptive properties and how bacterial HAPs are different structurally from mammalian HAPs. The major focus of this study was to determine the crystal structure of HAP from a human pathogen *Legionella pneumophila (LpHAP)*. LpHAP is an extracellularly secreted enzyme and shares significant sequence homology with the human HAP as compared to other bacterial acid phosphatases[5]. Sequence conservation shows that LpHAP shares ~30 % sequence identity with hPAP and 40% with FtHAP (a previously studied bacterial HAP from *Francisella tularensis*). Thus, LpHAP might serve as a better model to study substrate recognition in this class of enzymes.

6.2 Materials and Methods

6.2.1 Cloning, expression and purification

The gene for LpHAP was cloned into pET-20b vector using NcoI and XhoI restriction sites. Native LpHAP was expressed using a modified autoinduction[6] method. Briefly, the cells were grown in BL21(AI) at 37°C for ~3h and then after the addition of 0.2% arabinose, the temperature was reduced to 18°C. The cells were harvested after 28

hrs. Protein was purified to homogeneity after two-step purification using Ni-affinity and anion-exchange chromatography.

6.2.2 Crystallization of LpHAP

Crystallization trials were carried out using vapor diffusion technique in sitting drop crystal trays. Crystals were grown using protein at ~8.0 mg/ml in the buffer 50 mM Tris, 50 mM NaCl, 1 mM EDTA pH 7.5. Equal amounts of protein and reservoir solution were mixed and plate-shaped crystals of the native enzyme were obtained in 10% PEG 8000, 0.1 M imidazole pH 8.0, 0.2 M calcium acetate (Fig. 6.4A) These crystals diffracted to a low resolution of 3.5 Å. To improve the resolution, crystals were then grown in complex with L(+)-tartrate, which is a known competitive inhibitor for this class of enzymes[7]. Tetragonal-shaped crystals were obtained in 20% PEG 3350 and 0.2 M sodium acetate at pH 4.5 (Fig. 6.4B).

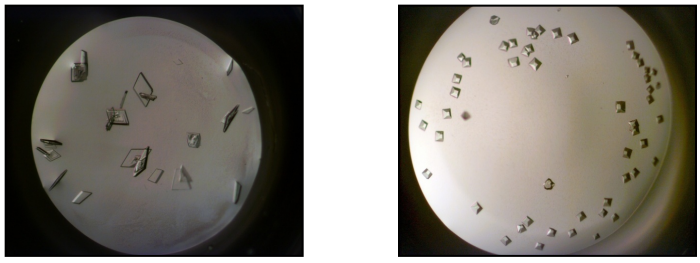

Fig. 6.4 Crystals of (A) LpHAP and (B) LpHAP with L-tartrate.

6.2.3 X-ray Diffraction, Data Collection and Refinement

The crystals of LpHAP were analyzed at APS beamline 24-ID-C using Quantum 315 detector, where they diffracted up to a resolution of 2 Å. A data set of 360 images was collected by a 0.5° rotation of the crystal and the data was processed using d*TREK and

scaled using SCALA[8] via CCP4i[9]. The structure of LpHAP was solved by molecular replacement using MOLREP[10] with FtHAP as the starting model, which shares 40 % sequence identity with LpHAP. The initial solution from MOLREP was then subjected to model building using Phenix.Autobuild[11]. The structure was built using COOT[12] and refined using several iterations of PHENIX[13]. Data collection and refinement statistics are listed in Table 6.1.

Table 6.1 X-ray diffraction, data collection and refinement statistics[a].

Enzyme	Native
Active site ligand	L(+)-tartrate
Space group	C2
Unit cell lengths (Å)	*a*= 217.88, b=134.27, c= 135.780 β = *127.46*
Wavelength	1.12711
Resolution (Å)	50.00 – 2.00 (2.07 – 2.00)
No. of observations	745974
No. of unique reflections	207120
$R_{merge}(I)$	0.079 (0.451)
Average I/σ	20.3 (2.7)
Completeness (%)	99.6 (98.6)
Redundancy	3.6 (3.3)
R_{cryst}	0.204
R_{free}[b]	0.241
No. of protein residues	657
No. of protein atoms	5129
No. of water molecules	413
Average B-factor (Å^2)	
Protein	31.05
Water	27.82
Active site ligand	38.00
rmsd[c]	
Bond lengths (Å)	0.007
Bond angles (deg)	0.975
Ramachandran plot[d]	
Favored (%)	98.3%
Allowed (%)	1.6%
Outliers (%)	0.1%

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]A common set of test reflections (5 %) was used for refinement of all structures.
[c]Compared to the parameters of Engh and Huber[14].
[d]The Ramachandran plot was generated with RAMPAGE[15].

6.2.4 Kinetic characterization of LpHAP

*A*n initial substrate screening for the enzyme was done using malachite green assay[16]. The reaction was carried out for a period of 15 min and the amount of phosphate released was measured spectrophotometrically at 625 nm. The potential substrates identified from this screening (3'-AMP and 5'- AMP) were then studied using steady-state enzyme kinetics. Enzymatic activity was measured at 37 °C using a discontinuous assay in the buffer 0.2 M sodium acetate pH 5.5. Different substrate concentrations were used and the amount of phosphate generated was measured spectrophotometrically at 625 nm. For each substrate concentration, the reaction was stopped at four different time points, i.e, 15, 75, 135, and 195 seconds, using malachite green-ammonium molybdate reagent.

The inhibition constant for LpHAP was determined using L(+)-tartrate as inhibitor. The enzymatic activity was measured using a discontinuous assay and similar buffer conditions were used with L(+)-tartrate. Different concentrations of the substrate *p*-nitrophenyl phosphate (pNPP) were used and for each substrate concentration, the reaction was stopped at four different time points, i.e, 15, 75, 135, and 195 seconds using 0.5 M glycine pH 10. Steady-state kinetic constants were determined by measuring the amount of *p*-nitrophenolate (pNP) formed spectrophotometrically at 405 nm.

6.3 Results and Discussion

6.3.1 Structure of LpHAP complexed with L(+)-tartrate

The structure of LpHAP was determined in complex with the competitive inhibitor L(+)-tartrate. LpHAP shares a similar fold with other members of HAP superfamily. It consists of two domains: the core domain that consists of the catalytic residue (His34) and

the cap domain that helps in substrate recognition. LpHAP crystallized with eight molecules in the asymmetric unit as shown below (Fig.6.5).

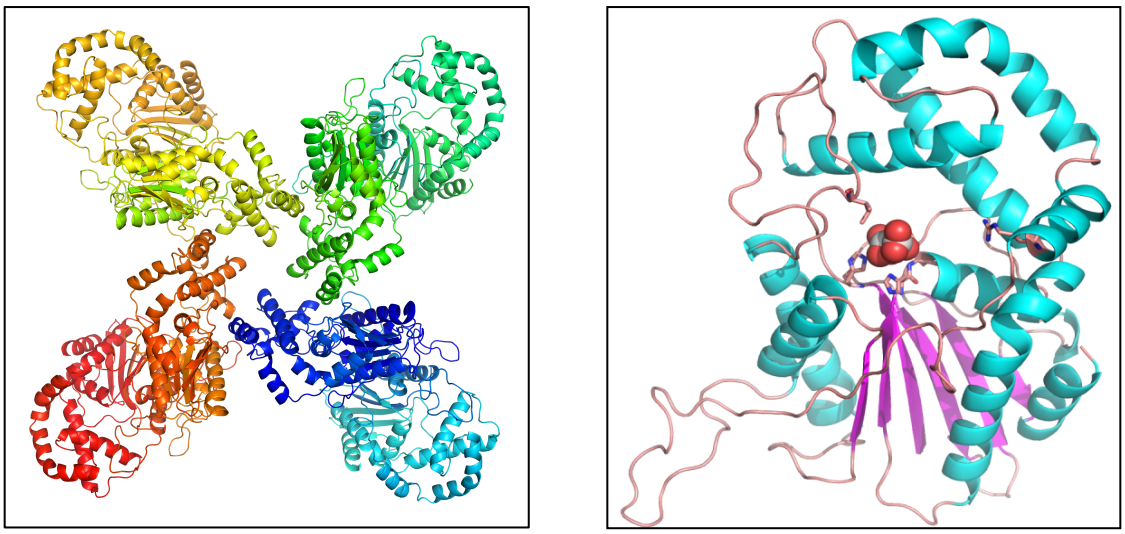

Fig. 6.5 (a) Overall structure of LpHAP with four dimers in the asymmetric unit. (b) LpHAP protomer with bound L-tartrate shown as spheres in red. The active site residues His34, Ile40, Arg 158, His280, and Asp 281 are shown as sticks.

Previous studies have shown that HAPs exist as dimers in solution[17]. Similar interlocking dimer was observed for LpHAP. The protomer structure of LpHAP complexed with L(+)-tartrate is shown (Fig. 6.5b). The hydroxyl groups of L(+)-tartrate are engaged in hydrogen bonding interaction with His34, Arg33, Asp281 and Arg101 residues.

6.3.2 A comparison of LpHAP, FtHAP and hPAP

The superposition of the protomer structures of LpHAP, FtHAP and hPAP reveal a shift in the position of α3 helix (Fig. 6.6). This shift in α3 helix might be responsible for different substrate preferences of this class of enzymes. The position of α3 helix in hPAP

results in a broader active site consistent with the wide substrate preference of this enzyme. hPAP is known to bind large phosphoproteins in addition to nucleoside phosphates, whereas, the narrower active site entrance of FtHAP precludes the binding of large substrates[17]. In LpHAP, α3 helix lies in between that of FtHAP and hPAP which makes it more closer homologue of hPAP as suggested earlier[5]. Thus, the substrate binding preferences of this enzyme might be closer to hPAP.

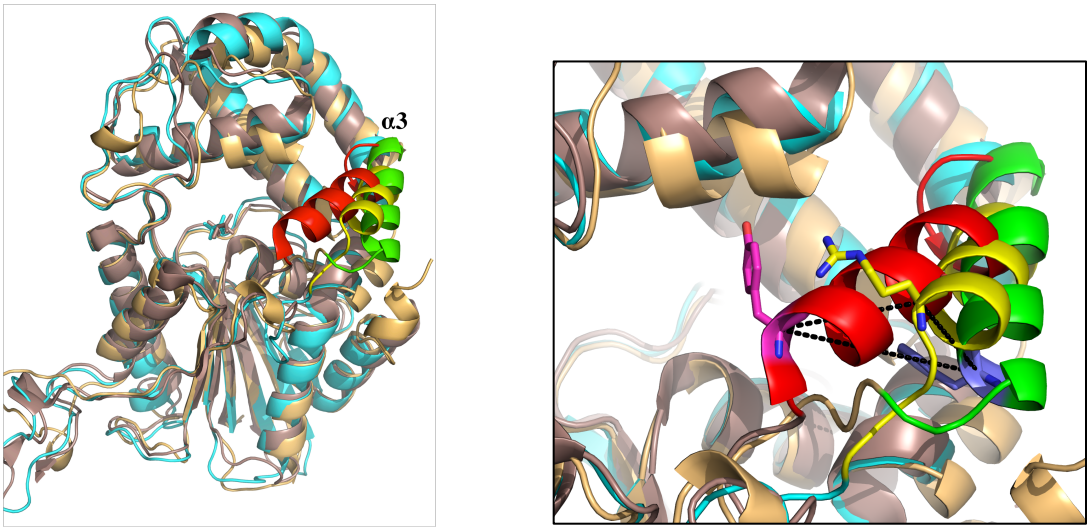

Fig. 6.6 Superposition of protomer structures of LpHAP (yellow), FtHAP (red) and hPAP (green) depicting a shift in α3 helix.

Previously determined structure of FtHAP mutant D261A complexed with 3´-AMP revealed the presence of an aromatic clamp formed by the active site residues Phe23 and Tyr135[17]. Sequence alignment shows that in LpHAP, the substrate clamping residues Phe23 and Tyr135 are substituted by Ile40 and Arg155 respectively.

Both the residues Tyr135 (in FtHAP) and Arg155 (in LpHAP) are present on α3 helix that is shifted in LpHAP by 8 Å as compared to FtHAP (Fig. 6.6). In FtHAP, the distance between the adenine base of 3'-AMP and substrate clamping residues Phe23 and Tyr135 is 3.5 Å and 3.9 Å, respectively, whereas, in LpHAP analogous residues Ile40 and

Arg155 are separated by 15 Å. Thus, Arg155 is too far to form an aromatic clamp analogous to FtHAP.

The shift in α3 helix and the absence of an aromatic clamp in LpHAP suggests that, LpHAP might bind nucleoside monophosphate substrates in a different manner as compared to that in FtHAP. Also, sequence analysis and structural alignment shows that His205 residue is present in the active site of LpHAP, instead of the conserved Asp residue in FtHAP and hPAP. A low resolution structure of D281A mutant of LpHAP at 3.5 Å was determined in complex with 5'-AMP (data not shown) suggesting that His205 might also play a role in substrate clamping, as analogous Arg155 residue is farther away. These results suggest that LpHAP might have a different substrate binding mechanism unique among HAPs.

6.3.2 Kinetic characterization of LpHAP

Steady-state kinetic parameters were measured for the preferred substrates 3'-AMP and 5'-AMP (Fig. 6.8). The Km for 3'-AMP is 10-fold greater as compared to that of FtHAP. This might be due to the broader active site for LpHAP in contrast to FtHAP. The k_{cat} iss ~100-fold less than FtHAP thus, resulting in ~380 fold lower catalytic efficiency as compared to FtHAP (Table 6.2).

Table 6.2 Steady-state kinetic parameters of LpHAP.

Enzyme	Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})
LpHAP	3´-AMP	3.3±0.6	3.0 ± 0.3	0.932
LpHAP	5´-AMP	3.8±0.5	2.40 ± 0.15	0.631
FtHAP	3´-AMP	0.315±0.001	119.9 ± 0.1	381

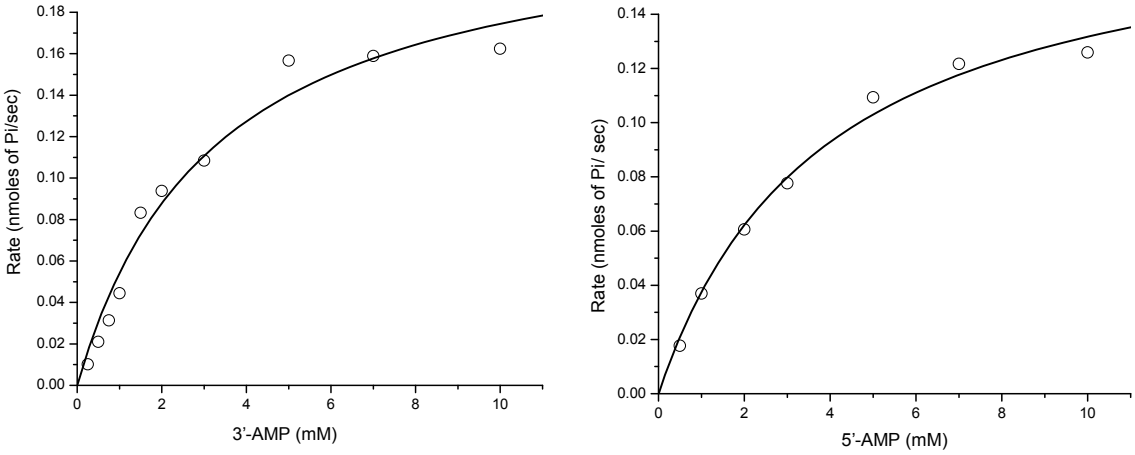

Fig. 6.7 Michaelis-Menten plots of LpHAP using substrates 3'-AMP and 5'-AMP.

The inhibition constant for LpHAP was determined using pNPP as substrate and L(+)-tartrate, which is a known competitive inhibitor for HAPs. The inhibition constants for hPAP and FtHAP were determined previously are 0.15 mM and 0.2 mM respectively[17-18]. The K_i for LpHAP is 0.43 μM that is ~50 times lower than the corresponding values for hPAP and FtHAP. Thus, L(+)-tartrate is a better inhibitor for LpHAP as compared to hPAP and FtHAP.

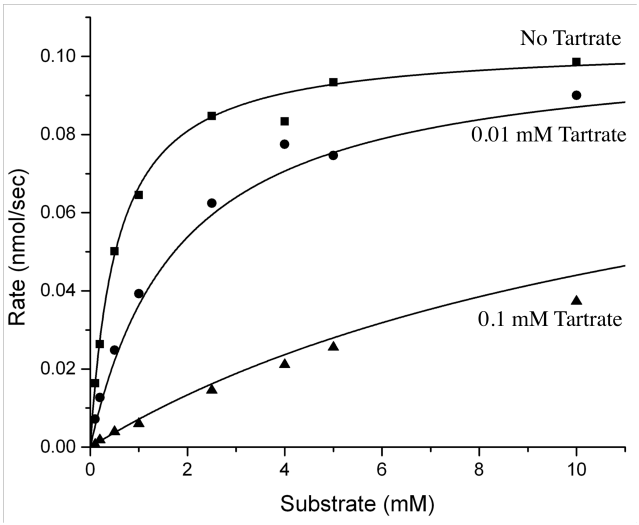

Fig. 6.8 Inhibition of LpHAP using L(+)-tartrate as inhibitor

6.4 References

1. Rigden, D. J., The histidine phosphatase superfamily: structure and function. *Biochem J* **2008,** *409* (2), 333-48.

2. Zylka, M. J.; Sowa, N. A.; Taylor-Blake, B.; Twomey, M. A.; Herrala, A.; Voikar, V.; Vihko, P., Prostatic acid phosphatase is an ectonucleotidase and suppresses pain by generating adenosine. *Neuron* **2008,** *60* (1), 111-22.

3. Sowa, N. A.; Street, S. E.; Vihko, P.; Zylka, M. J., Prostatic acid phosphatase reduces thermal sensitivity and chronic pain sensitization by depleting phosphatidylinositol 4,5-bisphosphate. *J Neurosci* **2010,** *30* (31), 10282-93.

4. Zylka, M. J., Pain-relieving prospects for adenosine receptors and ectonucleotidases. *Trends Mol Med* **2011,** *17* (4), 188-96.

5. Aragon, V.; Kurtz, S.; Cianciotto, N. P., Legionella pneumophila major acid phosphatase and its role in intracellular infection. *Infect Immun* **2001,** *69* (1), 177-85.

6. Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **2005,** *41* (1), 207-34.

7. LaCount, M. W.; Handy, G.; Lebioda, L., Structural origins of L(+)-tartrate inhibition of human prostatic acid phosphatase. *J Biol Chem* **1998,** *273* (46), 30406-9.

8. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 1), 72-82.

9. Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E., A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **2003,** *59* (Pt 7), 1131-7.

10. Vagin, A.; Teplyakov, A., Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 1), 22-5.

11. Zwart, P. H.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; McKee, E.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C.; Adams, P. D., Automated structure solution with the PHENIX suite. *Methods Mol Biol* **2008,** *426*, 419-35.

12. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.

13. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **2002,** *58* (Pt 11), 1948-54.

14. Engh, R. A. H., R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst*. **1991,** *A47* (4), 9.

15. Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003,** *50* (3), 437-50.

16. Geladopoulos, T. P.; Sotiroudis, T. G.; Evangelopoulos, A. E., A malachite green colorimetric assay for protein phosphatase activity. *Anal Biochem* **1991,** *192* (1), 112-6.

17. Singh, H.; Felts, R. L.; Schuermann, J. P.; Reilly, T. J.; Tanner, J. J., Crystal Structures of the histidine acid phosphatase from Francisella tularensis provide insight into substrate recognition. *J Mol Biol* **2009,** *394* (5), 893-904.

18. Ostanin, K.; Saeed, A.; Van Etten, R. L., Heterologous expression of human prostatic acid phosphatase and site-directed mutagenesis of the enzyme active site. *J Biol Chem* **1994,** *269* (12), 8971-8.

A.1 Crystallization and structural analysis of active site mutant G61P of Trypanosoma cruzi UDP-Galactopyranose mutase (TcUGM)

Large conformational changes have been observed in UGMs, both prokaryotic and eukaryotic, upon substrate binding[1]. In contrast, the catalytic histidine loop undergoes conformational changes upon FAD reduction that are unique to eukaryotic UGMs[1a, b]. These changes are substantial in AfUGM upon reduction with dithionite, where catalytic His63 moves from the *re* side of flavin to the *si* face of flavin and stacks parallel to the middle ring of isoalloxazine[1a]. In TcUGM, the histidine loop is retracted in oxidized state but upon FAD reduction, shifts by a distance of 2.3 Å towards isoalloxazine ring[1b]. Interestingly, no such conformational change associated with the histidine loop has been observed before in any of the prokaryotic UGMs.

The histidine loop plays an important role in the catalytic activity of UGMs[1b]. The backbone carbonyl oxygen of second glycine of GGH loop forms a hydrogen bond with N5 of reduced FAD ($FADH^-$). In bacterial UGMs, this residue is replaced by either Pro or Ala that are known for inducing rigidity in protein secondary structures. Steady-state enzyme kinetics shows that mutation of Gly61 residue in TcUGM to either Ala or Pro (analogous to bacterial UGMs) is highly detrimental to its catalytic activity[2]. These results suggest that glycine residue might impart flexibility to histidine loop required for efficient catalysis in eukaryotic UGMs[1b].

To understand the structural basis of histidine loop flexibility, the structures of TcUGM G61P mutant complexed with UDP were determined in the oxidized and reduced states.

A.1.1. Crystallization

The TcUGM histidine loop mutant G61P was crystallized at 7–8 mg/mL in the buffer 50 mM HEPES, 150 mM NaCl at pH 7.5. Prior to crystallization, the enzyme was incubated with 10 mM UDP for 30 min. Crystallization experiments were performed at 20 °C using sitting drop vapor diffusion with the drops formed by mixing 1.5 μL each of the protein and reservoir solutions. Crystals were grown in the same condition as native enzyme with reservoirs containing ammonium sulfate and HEPES buffer[1b]. Small yellow block-shaped crystals appeared after 4-6 weeks. The optimized crystallization reservoir consisted of 1.2 M ammonium sulfate and 0.1 M HEPES at pH 7.5. The crystals were cryoprotected in 1.4 M ammonium sulfate, 0.1 M HEPES pH 7.5, and 25% ethylene glycol before plunging into liquid N_2. The space group is $P3_221$ with unit cell dimensions of a = 143 Å and c = 166 Å and two molecules in the asymmetric unit.

Crystals of G61P-UDP complex with FAD in the reduced state were prepared by soaking the aforementioned crystals in 1.4 M ammonium sulfate, 0.1 M HEPES pH 7.5, 60 mM sodium dithionite, and 25% ethylene glycol. Once the crystals turned from yellow to colorless, they were flash-cooled by plunging into liquid N_2.

A.1.2 X-ray diffraction, data collection, phasing and refinement

The diffraction data for oxidized and reduced G61P–UDP were collected at beamline 24-ID-C. The data were integrated using XDS[3] and scaled with SCALA[4] via CCP4i[5]. Data processing statistics are listed in Table A.1.1.

The phase problem for oxidized G61P–UDP was solved using molecular replacement as implemented in MOLREP[6]. The search model was derived from the

structure of oxidized TcUGM (PDB code 4DSG[1]). Molecular replacement calculations produced a solution having two molecules in the asymmetric unit with *R*-factor of 0.45 and score of 0.60. The structure was built using COOT[7] and refined using PHENIX[8]. Data refinement statistics are listed in Table A.1.1.

Table A.1.1 X-ray Diffraction Data Collection and Refinement[a]

	G61P-Oxidized	G61P-Reduced
Space group	$P3_221$	$P3_221$
Unit cell parameters (Å)	a = 142.9, c = 166.2	a = 143.3, c = 168.0
Wavelength (Å)	0.9791	0.9795
Resolution (Å)	166.23 - 2.77 (2.92 - 2.77)	124.15 - 2.92 (3.08 - 2.92)
Observations	309683	241052
Unique reflections	50464	43797
$R_{merge}(I)$[b]	0.091 (0.778)	0.110 (0.750)
$R_{meas}(I)$[b]	0.099 (0.848)	0.122 (0.830)
$R_{pim}(I)$[b]	0.040 (0.331)	0.051 (0.351)
Mean I/σ	17.0 (2.5)	13.8 (2.4)
Completeness (%)	99.7 (99.4)	99.8 (99.4)
Multiplicity	6.1 (6.1)	5.5 (5.6)
No. of protein residues	944	915
No. of protein atoms	7592	7223
No. of FAD atoms	106	106
No. of ligand atoms	50	50
No. of water molecules	1	0
R_{cryst}	0.190	0.202
R_{free}[c]	0.225	0.254
rmsd bond lengths (Å)[d]	0.009	0.008
rmsd bond angles (°)[d]	1.21	1.19
Ramachandran plot[e]		
Favored (%)	95.9	95.0
Allowed (%)	4.4	4.6
Outliers (%)	0.4	0.3
Average B-factor (Å^2)		
Protein	65	70
FAD	59	66
UDP	54	59
Water	49	0
Coordinate error (Å)[f]	0.44	0.53

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Definitions of R_{merge}, R_{meas}, and R_{pim} can be found in Weiss [9].
[c]A common set of test reflections (5 %) was used for refinement of both structures.
[d]Compared to the parameters of Engh and Huber [10].
[e]The Ramachandran plot was generated with RAMPAGE [11].
[f]Maximum likelihood-based coordinate error estimate from PHENIX[8].

A.1.3 Results and Discussion

The structures of TcUGM mutant G61P were solved in complex with UDP in oxidized and reduced states. In oxidized state, the two structures differ in the conformation of histidine loop (Fig. A.1.1(a)). In $TcUGM_{ox}$, the histidine loop is retracted from FAD isoalloxazine and the backbone carbonyl of Gly61 is 5 Å away from N5 of FAD. In contrast, when proline substitutes Gly61 in G61P structure, the loop shifts by 1.7 Å towards isoalloxazine to form a hydrogen bond between backbone carbonyl of Pro61 and N5 of isoalloxazine. Upon reduction with dithionite, the two structures are identical with similar hydrogen bonding distance of 2.9 Å (Fig. A.1.1(b)).

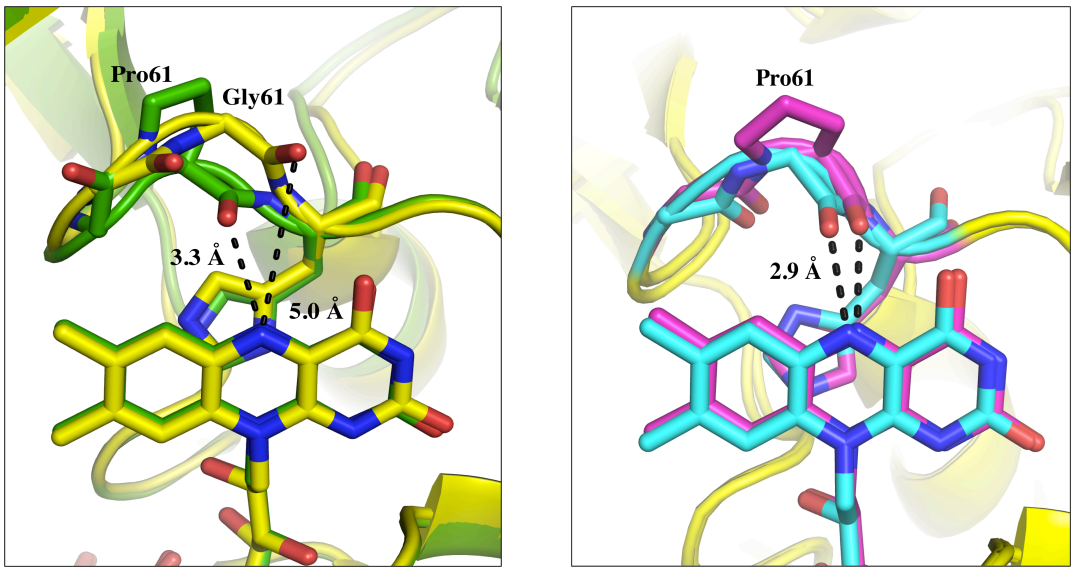

Fig. A.1.1 (a) Superposition of TcUGM (yellow) and G61P mutant (green) in the oxidized state (b) Superposition of TcUGM (cyan) and G61P mutant (magenta) in the reduced state.

The superposition of the oxidized and reduced structures of *Klebsiella pneumoniae* UGM (KpUGM) shows the two structures are exactly identical (Fig. A.1.2). The only difference lies in the planarity of FAD isoalloxazine ring; being planar in the oxidized state and butterfly-shaped bent in the reduced state. The backbone carbonyl oxygen makes a hydrogen bond with FAD isoalloxazine N5 in both oxidized and reduced states. This hydrogen- bonding pattern is similar to that observed in G61P structure, where Gly61 is substituted by Pro, analogous to bacterial UGMs.

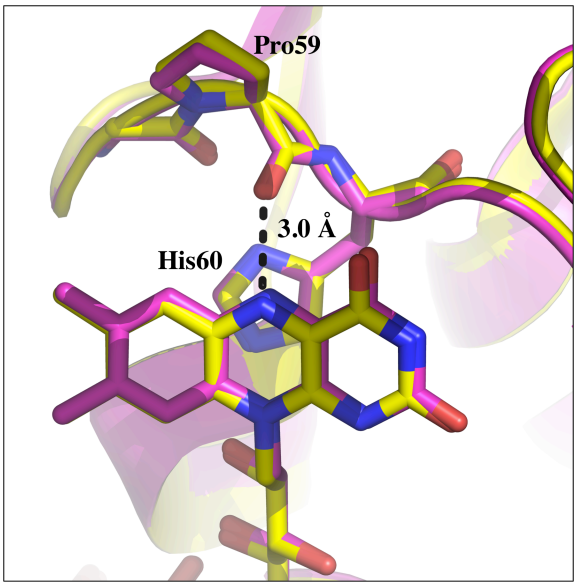

Fig. A.1.2 Superposition of *Klebsiella pneumoniae* UGM (KpUGM) in the oxidized (yellow) and reduced (magenta) states.

Thus, proline induces rigidity in the histidine loop such that, even in oxidized state enzyme is locked in the same conformation as in reduced state. In eukaryotic UGMs, the presence of Gly instead of Pro/Ala imparts flexibility to the histidine loop such that in oxidized state, the loop is farther away from FAD isoalloxazine, whereas, in reduced state it forms a hydrogen bond to stabilize the active form of enzyme.

A.1.4 References

1. (a) Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Karr, D. B.; Nix, J. C.; Sobrado, P.; Tanner, J. J., Crystal structures and small-angle x-ray scattering analysis of UDP-galactopyranose mutase from the pathogenic fungus Aspergillus fumigatus. *J Biol Chem* **2012,** *287* (12), 9041-51; (b) Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Sobrado, P.; Tanner, J. J., Crystal structures of Trypanosoma cruzi UDP-galactopyranose mutase implicate flexibility of the histidine loop in enzyme activation. *Biochemistry* **2012,** *51* (24), 4968-79; (c) Partha, S. K.; van Straaten, K. E.; Sanders, D. A., Structural basis of substrate binding to UDP-galactopyranose mutase: crystal structures in the reduced and oxidized state complexed with UDP-galactopyranose and UDP. *J Mol Biol* **2009,** *394* (5), 864-77.

2. Oppenheimer, M.; Valenciano, A. L.; Kizjakina, K.; Qi, J.; Sobrado, P., Chemical mechanism of UDP-galactopyranose mutase from Trypanosoma cruzi: a potential drug target against Chagas' disease. *PLoS One* **2012,** *7* (3), e32918.

3. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.

4. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 1), 72-82.

5. Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E., A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **2003,** *59* (Pt 7), 1131-7.

6. Vagin, A.; Teplyakov, A., Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 1), 22-5.

7. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.

8. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **2002,** *58* (Pt 11), 1948-54.

9. Weiss, M. S., Global indicators of X-ray data quality. *J. APPL. Cryst.* **2001,** *34*, 6.

10. Engh, R. A. H., R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst.* **1991,** *A47* (4), 9.

11. Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003,** *50* (3), 437-50.

A.2 Structure determination of UDP-galactopyranose mutase in complex with substrate analogs UDP-arabinopyranose and UDP-arabinofuranose

Several enzymes in the Galactofuranose (Galf) biosynthetic pathway have been characterized but not much is known about the metabolism of arabinofuranose (Ara_f). The only structural difference between the two monosaccharides is that L-arabinose has a shorter side chain at C4 as compared to D-galactose.

Previous studies have shown that UDP-L-Arabinopyranose (UDP-Ara_p) can undergo interconversion to UDP-L-Arabinofuranose (UDP-Ara_f) in the presence of *Escherichia coli* UDP-Galactopyranose mutase (EcUGM)[1]. However, the surface glycoconjugate structures of bacteria, fungi and protozoans have D-arabinofuranose, whose precursor is β-D-arabinofuranosyl-1-monophosphoryldecaprenol[2]. Interestingly, L-arabinofuranose forms a major component of plant cell wall structures[3]. Recently, a plant mutase UDP-Arabinopyranose mutase (UAM) was identified that catalyzes the conversion of UDP-Ara_p into UDP-Ara_f[3]. These enzymes belong to the reversibly glycosylated polypeptides (RGP) family and are different as compared to UGMs as they do not require any cofactor and thus, are not FAD-dependent[3].

Structural characterization of AfUGM in complex with substrate analogs UDP-Ara_p and UDP-Ara_f would help us determine how UGM recognizes specifically UDP-galactopyranose and not UDP-arabinopyranose.

A.2.1 Soaking AfUGM with substrate analogs UDP-Ara_p and UDP-Ara_f

Crystals of AfUGM K344A/K345A were grown in the condition consisting of 1.2

M ammonium sulfate and 0.1 M sodium acetate pH 4.5 as discussed previously [4]. Yellow hexagonal crystals were obtained in 3-4 days. The crystals were soaked simultaneously with 80 mM dithionite and 70 mM UDP-L-arabinopyranose for 9 min in the cryoprotectant consisting of 1.6 M ammonium sulfate, 0.1 M sodium acetate pH 4.5 and 25 % ethylene glycol.

Similar strategy was used for the soaking of AfUGM crystals with UDP-L-arabinofuranose. The crystals were soaked with 80 mM dithionite and 6.2 mM UDP-L-arabinofuranose for 4 min in the cryoprotectant consisting of 1.6 M ammonium sulfate, 0.1 M sodium acetate pH 4.5 and 25 % ethylene glycol.

A.2.2 X-ray diffraction, data collection and refinement

The data for AfUGM complexed with UDP-Ara_p and UDP-Ara_f in the reduced state was collected at ALS beamline 4.2.2 using CMOS detector. The data sets were integrated using XDS[5], scaled with SCALA[6] via CCP4i[7]. Data collection and refinement statistics are listed in Table A.2.1.

Table A.2.1 X-ray Diffraction Data Collection and Refinement[a]

	UDP-Ara_p	UDP-Ara_f
Space group	$P6_522$	$P6_522$
Unit cell parameters (Å)	a = 217.8, c = 318.0	a = 218.1, c = 320.5
Wavelength (Å)	0.9791	0.9795
Resolution (Å)	64.21 - 2.95 (2.95 - 2.80)	62.97 – 3.06 (3.06 – 2.90)
Observations	1218850	1108518
Unique reflections	108989	88525
$R_{merge}(I)$[b]	0.306 (1.134)	0.445 (1.272)
$R_{meas}(I)$[b]	0.336 (1.247)	0.481 (1.379)
$R_{pim}(I)$[b]	0.139 (0.517)	0.178 (0.520)
Mean I/σ	8.2 (2.0)	5.8 (1.8)

Completeness (%)	99.7 (98.3)	89.7 (90.3)
Multiplicity	11.2 (11.2)	12.5 (12.3)
No. of protein residues	2019	2019
No. of protein atoms	15499	15698
No. of FAD atoms	212	212
No. of ligand atoms	68	68
No. of water molecules	19	4
R_{cryst}	0.204	0.207
R_{free}[c]	0.244	0.257
rmsd bond lengths (Å)[d]	0.008	0.008
rmsd bond angles (°)[d]	1.15	1.21
Ramachandran plot[e]		
Favored (%)	95.9	95.0
Allowed (%)	4.4	4.6
Outliers (%)	0.4	0.3
Average B-factor (Å^2)		
Protein	41	39
FAD	58	46
UAD	72	54
Water	30	27
Coordinate error (Å)[f]	0.38	0.40

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Definitions of R_{merge}, R_{meas}, and R_{pim} can be found in Weiss [8].
[c]A common set of test reflections (5 %) was used for refinement of both structures.
[d]Compared to the parameters of Engh and Huber [9].
[e]The Ramachandran plot was generated with RAMPAGE [10].
[f]Maximum likelihood-based coordinate error estimate from PHENIX

A.2.4 References

1. Zhang, Q.; Liu, H. W., Chemical synthesis of UDP-beta-L-arabinofuranose and its turnover to UDP-beta-L-arabinopyranose by UDP-galactopyranose mutase. *Bioorg Med Chem Lett* 2001, *11* (2), 145-9.

2. Scherman, M. S.; Kalbe-Bournonville, L.; Bush, D.; Xin, Y.; Deng, L.; McNeil, M., Polyprenylphosphate-pentoses in mycobacteria are synthesized from 5-phosphoribose pyrophosphate. *J Biol Chem* 1996, *271* (47), 29652-8.

3. Konishi, T.; Takeda, T.; Miyazaki, Y.; Ohnishi-Kameyama, M.; Hayashi, T.; O'Neill, M. A.; Ishii, T., A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. *Glycobiology* 2007, *17* (3), 345-54.

4. Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Karr, D. B.; Nix, J. C.; Sobrado, P.; Tanner, J. J., Crystal structures and small-angle x-ray scattering analysis of UDP-galactopyranose mutase from the pathogenic fungus Aspergillus fumigatus. *J Biol Chem* 2012, *287* (12), 9041-51.

5. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* 2010, *66* (Pt 2), 125-32.

6. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 2006, *62* (Pt 1), 72-82.

7. Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E., A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 2003, *59* (Pt 7), 1131-7.

8. Weiss, M. S., Global indicators of X-ray data quality. *J. APPL. Cryst.* 2001, *34*, 6.

9. Engh, R. A. H., R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst.* 1991, *A47* (4), 9.

10. Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 2003, *50* (3), 437-50.

11. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 2002, *58* (Pt 11), 1948-54.

A.3 Crystallization of flavin domain of tetrahydrofuran monooxygenase (ThmD-FD)

Tetrahydrofuran monooxygenase (Thm) is a member of bacterial multicomponent monooxygenases (BMM). The enzyme catalyzes hydroxylation of tetrahydrofuran to 2-hydroxytetrahydrofuran in the presence of NADH and O_2. Thm is composed of three subunits: a hydroxylase subunit with [2Fe-2S], a regulatory subunit, and an oxidoreductase component named ThmD. These enzymes have potential application in bioremediation as they are capable of cleaving C-H bond in hydrocarbons and catalyze the hydroxylation of their substrates[1].

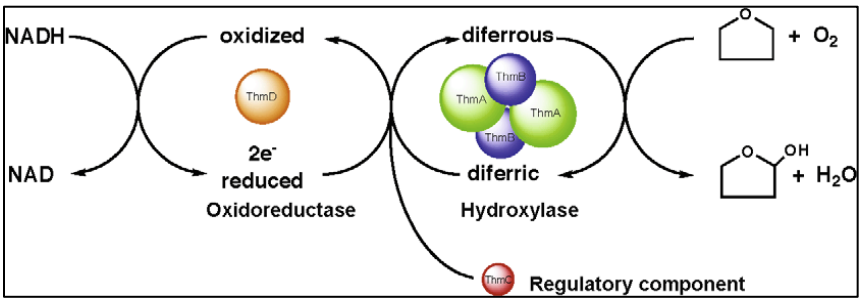

Fig. A.3.1 Catalytic cycle of tetrahydrofuran monooxygenase (Thm)[2].

ThmD from *Pesudonocardia* sp. strain K1 capable of growing on tetrahydrofuran as the sole carbon source, was earlier characterized and purified to homogeneity[2]. ThmD has a molecular weight of 39,845 Da (~40 kDa) and exists as a monomer in solution.

Previous studies on methane monooxygenase (MMO) have shown that both FAD binding domain and [2Fe-2S] domain are capable of folding independently from each other but still are functional[3]. Interestingly, the oxidoreductase component of ThmD is different than other members of the BMM family in that it has a covalently attached

FAD[2]. The typical sites for attachment to FAD are at C6 and C8α- methyl group of the isoalloxazine ring. Mutations and computational studies to determine the site of covalent attachment in ThmD did not reveal any characteristic residues such as His, Cys or Tyr to be involved. Thus, a new site for covalent attachment or a novel fold has been proposed for this enzyme[2]. These findings motivated us to determine the three-dimensional structure of ThmD.

A.3.1 Crystallization of ThmD-FD

Extensive crystallization trials were carried out to crystallize the FAD binding domain of ThmD (ThmD-FD). No crystals were obtained using the commonly available crystallization screens. Thus, the protein was modified using iodoacetic acid[4]. The protein was dialyzed overnight into the buffer 25 mM Tris, 100 mM ammonium sulfate, 0.5% iodoacetatic acid pH 7.8. Commercially available screens were used to set up initial crystallization trials. Few conditions (PEG3350 as the precipitant) yielded dark yellow crystals those were previously never discovered. Few rounds of optimization produced long rods. The final optimized condition consisted of 27% PEG 3350, 0.2 M $MgCl_2$, 0.1 M Tris pH 8.5. Crystals of SeMet derivative were obtained in the same condition. Partial structure was solved but the density of FAD was not completely traceable.

A.3.2 X-ray diffraction, data collection and structure refinement

Several single-wavelength anomalous diffraction data sets were collected at the wavelength corresponding to the experimentally measured peak of f″. The phasing potential of each data set was assessed with the HKL2MAP interface[5] to the SHELXC/D/E programs[5-6]. Strong heavy atom signal was not observed which might be

due to only 60% incorporation of Se, as shown by mass spectrometry results (data not shown). Finally, a data set with 2.43 Å resolution and good anomalous signal was identified. Overall, 12 heavy atom sites were identified with occupancy greater than 0.5. As only five methionine residues were present in the sequence, other heavy atom sites might be due to iodine incorporation by enzyme modification with iodoacetic acid. These heavy atom sites were then input to PHENIX AutoSol[7] for single-wavelength anomalous diffraction phasing, density modification, and automated building.

Partial structure of ThmD-FD was solved but the electron density for FAD cofactor remained elusive. The model from automated building was built and extended manually in COOT[8]. Data collection and refinement statistics are listed in Table A.3.1.

Table A.3.1 X-ray diffraction, data collection and refinement statistics[a].

Parameters	**SeMet –ThmD**
Space group	$P4_3$
Unit cell lengths (Å)	a= 61.0, c= 71.70
Wavelength	0.97918
Resolution (Å)	46.47 – 2.43 (2.56 – 2.43)
$R_{merge}(I)$	0.124 (0.925)
R_{meas}	0.128 (0.958)
R_{pim}	0.034 (0.248)
Average I/σ	14.9 (3.2)
Completeness (%)	99.9 (99.9)
Multiplicity	14.7 (14.9)
Mid-slope of anomalous normal probability	1.159
No. of observations	146836
No. of unique reflections	10009
R_{cryst}	0.372
R_{free}	0.417
rmsd bond lengths (Å)[c]	0.009
rmsd bond angles (°)[c]	1.483

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Definitions of R_{merge}, R_{meas}, and R_{pim} can be found in Weiss[9].
[c]Compared to the parameters of Engh and Huber[10].

A.3.3 References

1. Notomista, E.; Lahm, A.; Di Donato, A.; Tramontano, A., Evolution of bacterial and archaeal multicomponent monooxygenases. *J Mol Evol* **2003,** *56* (4), 435-45.

2. Oppenheimer, M.; Pierce, B. S.; Crawford, J. A.; Ray, K.; Helm, R. F.; Sobrado, P., Recombinant expression, purification, and characterization of ThmD, the oxidoreductase component of tetrahydrofuran monooxygenase. *Arch Biochem Biophys* **2010,** *496* (2), 123-31.

3. Chatwood, L. L.; Muller, J.; Gross, J. D.; Wagner, G.; Lippard, S. J., NMR structure of the flavin domain from soluble methane monooxygenase reductase from Methylococcus capsulatus (Bath). *Biochemistry* **2004,** *43* (38), 11983-91.

4. Bubis, J.; Ortiz, J. O.; Moller, C., Chemical modification of transducin with iodoacetic acid: transducin-alpha carboxymethylated at Cys(347) allows transducin binding to Light-activated rhodopsin but prevents its release in the presence of GTP. *Arch Biochem Biophys* **2001,** *395* (2), 146-57.

5. Schneider, T. P. T. R., HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.* **2004,** *37*, 843-844.

6. Sheldrick, G. M., Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 479-85.

7. Zwart, P. H.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; McKee, E.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C.; Adams, P. D., Automated structure solution with the PHENIX suite. *Methods Mol Biol* **2008,** *426*, 419-35.

8. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.

9. Weiss, M. S., Global indicators of X-ray data quality. *J. APPL. Cryst.* **2001,** *34*, 6.

10. Engh, R. A. H., R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst.* **1991,** *A47* (4), 9.

A.4 Expression, purification and crystallization of 2-Haloacrylate hydratase (2-HAH)

2-Haloacrylate hydratase is a flavoenzyme that catalyzes the conversion of 2-chloroacrylate (2-CAA) to pyruvate (Fig. A.4.1)[1]. The enzyme has an absolute requirement for reduced FAD ($FADH_2$) and reducing partner NAD(P)H for efficient catalysis[1-2]. Due to the interesting chemistry and possible applicability of 2-HAH in the chemical industry, we collaborated with the Sobrado laboratory to unravel the structure and, thereby, function of this novel flavoenzyme. Moreover, the enzyme is similar to UDP-galactopyranose mutase (UGM) as it requires reduced FAD for activity, but catalyzes the reaction without any net transfer of electrons.

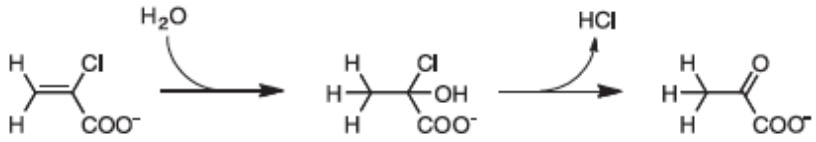

Fig. A.4.1 Reaction catalyzed by 2-haloacrylate hydratase[1].

A.4.1 Cloning, expression and purification of 2-HAH

2-HAH (547 amino acid) from *Pseudomonas* sp. YL ($2\text{-}HAH_{YL}$) was cloned into pVP56K vector by Sobrado laboratory. Briefly, the cloning of $2\text{-}HAH_{YL}$ into pVP56K vector results in a construct with N- terminal 8xHis-maltose-binding-protein(MBP)-TEVP cleavage site $2\text{-}HAH_{YL}$ with kanamycin resistance.

The construct thus obtained was transformed into BL21(DE3)pLysS and the resulting transformant was used for further experiments. $2\text{-}HAH_{YL}$ was expressed using autoinduction in terrific broth (TB) media. Briefly, overnight culture was prepared by inoculating a 50 mL LB broth media supplemented with kanamycin at 40 μg/mL. The TB

culture supplemented with 100 mL phosphate buffer, 2.4 mL of 1 M $MgSO_4$, 30 mL 15% (wt) succinic acid (pH 4.5), 40 mL of 30x80155 (prepared by dissolving 240 mL of glycerol, 4.5 g of glucose, 150 g of lactose per liter in water and sterile filtered) and 40μg/mL kanamycin was inoculated using 1% starter culture. The culture was incubated at 37 °C and 250 rpm for ~ 6 hrs (O.D. 600 ~ 3.0) and then temperature was lowered to 18 °C and cultures were incubated overnight. Following day, the cells were harvested by centrifuging at 3500 rpm at 4 °C for 30 min. The resulting cells were resuspended in 20 mM Tris, 300 mM NaCl, 10% glycerol pH 7.5 (Buffer A) and flash frozen at -80 °C until further purification.

Cells were thawed and ruptured by sonication and centrifuged at 17,000 rpm at 4 °C for 45 min to remove the cell debris. After 45 min, the supernatant was transferred into fresh tubes and centrifuged again at 17,000 rpm for another 30 min. The supernatant thus obtained was loaded onto Ni^{2+} column and the protein was purified using conventional Ni-IDA method in which the 2-HAH_{YL} was eluted using a step gradient of 20 mM Tris, 300 mM NaCl, 10 % glycerol, 1 M imidazole pH 7.5 (buffer B). Bright yellow samples were pooled together and were divided into two different fractions. For crystallization purposes, we used MBP-2-HAH_{YL} fusion protein as well as tag free 2-$HAH_{YL.}$ The MBP- fusion protein was dialyzed into 50 mM Tris, 50 mM NaCl, 10% glycerol, 1mM EDTA, 1 mM THP pH 8.0 (buffer A) and further purified using ion-exchange chromatography. The fusion protein was eluted using linear gradient of NaCl. Fractions were pooled together and further purified using superdex-200 using 50 mM Tris, 150 mM NaCl, 1mM EDTA, 1 mM THP pH 8.3. MBP-2-HAH_{YL} thus obtained was used for crystallization experiments.

For His-MBP cleavage, 1 mg of purified TEVP was added to ~ 40 mg of 2-HAH_{YL} in the presence of 1 mM THP and incubated at 20 °C for 4 hours. After 4 hours, the TEVP mixed 2-HAH_{YL} was dialyzed overnight into buffer A and again loaded onto Ni-IDA column. Tag free 2-HAH_{YL} was collected either in the flow-through or in 3% buffer B. Clean samples (as adjudged by SDS-PAGE) were pooled together and dialyzed overnight into 100 mM Tris, 200 mM NaCl, 1 mM EDTA, 5 % glycerol, 1 mM THP pH 8.0 and further purified using size-exclusion chromatography. All the dialysis steps were performed at 4 °C, unless otherwise stated.

A.4.2 Crystallization of 2-HAH

Several attempts to obtain crystals of 2-HAH_{YL} resulted in almost no success. These attempts included reductive methylation, modification using iodoacetic acid, in situ proteolysis using various commercial available proteases. These results indicated that 2-HAH_{YL} might be recalcitrant to crystallization therefore,\ we decided to use the following approaches:

1)- Domain dissection of 2-HAH_{YL}: 2-HAH_{YL} is FAD- dependent enzyme and is predicted to have multiple domains. To test this hypothesis, we analyzed the protein sequence of 2-HAH_{YL} using InterPro: protein sequence and analysis tool (http://www.ebi.ac.uk/interpro/). InterPro utilizes diverse source repositories to predict domains and important sites in proteins. Based upon our analysis of 2-HAH_{YL} amino acid sequence, we designed 8-394, 5-424, and 1-410 truncation constructs. Unfortunately, all the three constructs resulted in insoluble protein. Perhaps, a more comprehensive design of construct is warranted to obtain soluble and active protein.

2)- Disordered region prediction: Proteins that are recalcitrant to crystallization, sometimes harbor very flexible regions. Therefore, we decided to carry out disordered region prediction using online tools. All the analysis pointed at making a construct that lacks last six amino acids at the C-terminus of the protein. Furthermore, the construct obtained from Dr. Sobrado's laboratory was designed such that MBP was fused to the N-terminus of the 2-HAH_{YL}. Cleaving the fusion protein sometimes renders an unstable target protein. Therefore, we decided to design a construct that does not have MBP as fusion partner.

To achieve this, we cloned $2HAH_{YL}$ into pKA8H using NdeI and BamHI sites such that the last six amino acids (predicted to be disordered by PrDOS[3]) were not included in the final construct. The truncated construct thus designed was cloned into pKA8H using standard methods and several expression tests using autoinduction (LB and TB media at 25 °C and 18 °C) and IPTG (ranging from 100 μM to 0.5 mM at 18 °C and 25 °C) were performed. We were able to express copious amounts of protein but it was mostly insoluble.

3)- Homology screening: In addition to the above-mentioned approaches, we subcloned a related 2-HAH from *Burkholderia* sp. WS (2-HAH_{WS}).

We subcloned 2-HAH_{WS} using NdeI and BamHI restriction sites into pKA8H vector. Clone was verified using DNA sequencing and was transformed into BL21(AI) (for arabinose-induced autoinduction) and BL21(DE3)pLysS (IPTG based expression). Our expression tests concluded that, 2-HAH_{WS} was able to express but was insoluble and therefore could not be used for further studies.

In parallel, to the above, Dr. Sobrado's lab cloned 2-HAH_{WS} into pVP56K vector

(identical to 2-HAH_{YL}). We transformed the MBP-2-HAH_{WS} construct into BL21(DE3)pLysS and protein was expressed using identical method was for 2-$HAH_{YL.}$ Briefly, cells were harvested as mentioned for 2-HAH_{YL} and resuspended in 20 mM Tris-sulfate, 300 mM sodium sulfate, 5 % glycerol pH 7.6 and frozen until further use.

Frozen cells were thawed in the presence of 1 mM PMSF, 0.7 % β-octyl glucoside (BOG), stirred at 4 °C for 15-20 minutes and ruptured using sonicator. The cell debris was separated by centrifugation at 17,000 rpm at 4 °C. The supernatant was loaded onto Ni^{2+} column and protein was purified using identical method as written above for 2-$HAH_{YL.}$ The fusion protein thus obtained was incubated with TEVP (1mg of TEVP/40 mg of MBP-2-HAH_{WS}) in presence of 1 mM THP for three hours at 20 °C and dialyzed overnight against buffer A. Next day, MBP-2-HAH_{WS} plus TEVP mixture was loaded onto Ni^{2+} column and tag free 2-HAH_{WS} was collected either in flow through or eluted using 3 % of buffer B (Buffer A + 1 M imidazole pH 7.6).

Appropriate fractions were pooled together and dialyzed into 75 mM Tris, 50 mM Na_2SO_4, 50 mM glutamic acid, 50 mM L-arginine, 1 mM THP pH 7.6. The dialyzed protein was further purified using size- exclusion chromatography (superdex-200 column) using the same buffer.

A.4.2 References

1. Mowafy, A. M.; Kurihara, T.; Kurata, A.; Uemura, T.; Esaki, N., 2-haloacrylate hydratase, a new class of flavoenzyme that catalyzes the addition of water to the substrate for dehalogenation. *Applied and environmental microbiology* **2010,** *76* (18), 6032-7.

2. Kurihara, T., A mechanistic analysis of enzymatic degradation of organohalogen compounds. *Bioscience, biotechnology, and biochemistry* **2011,** *75* (2), 189-98.

3. Ishida, T.; Kinoshita, K., PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* **2007,** *35* (Web Server issue), W460-4.

A.5 Subcloning and expression of Polyphosphate kinase (PPK2)

Inorganic polyphosphate, poly P is present in all bacteria, fungi, plants and animals. It plays a role in the phosphorylation of nucleoside mono- and di- phosphates. It is needed for bacterial survival under stress conditions and for virulence in some pathogens[1].

Polyphosphate kinases are the enzymes that catalyze the reversible synthesis of inorganic polyphosphate (poly P) from nucleoside di- and tri- phosphates. These can be classified as PPK1 and PPK2 depending upon their substrates. PPK1 catalyzes the reversible synthesis of poly P from ATP using Mg^{2+} as the cofactor, whereas, PPK2 uses poly P as a donor to synthesize both GTP and ATP from their respective nucleoside diphosphates[2]. Thus, in PPK2 the rate of poly P utilization is 100-fold greater than poly P synthesis[2a].

A.5.1 Subcloning of PPK2

The FtPPK gene was previously cloned into pET-21 vector. For the expression of an N-terminus tagged fusion protein, the gene was then subcloned into pSV281 vector using BamHI and XhoI restriction sites. Briefly, the FtPPK gene was amplified using PCR to incorporate BamHI and XhoI restriction sites at N- and C-terminus, respectively.

The PCR product was gel purified (using 0.8% agarose gel and Qiagen PCR clean-up kit). The purified PCR product was then first ligated into P_0 vector at 16 °C overnight. The ligation product was then transformed into DH5α cells and plated onto LB agar supplemented with 40 μg/ml kanamycin. The plates were incubated overnight and the

following day four single colonies were selected and grown overnight at 37 °C and 250 rpm in LB broth using kanamycin. After plasmid extraction, DNA was digested using BamHI and XhoI restriction enzymes and gel purified. The insert thus obtained was ligated into pSV281 vector (previously digested using BamHI and XhoI restriction enzymes and gel purified). The ligation product was transformed into DH5α, plated onto LB agar supplemented with 40 μg/mL Kanamycin and incubated overnight at 37 °C. Four single colonies were picked the following day and were grown in LB broth supplemented with 40 ug/mL Kanamycin. An analytical 0.8% agarose gel was run to confirm whether the gene was incorporated or not. Further, DNA sequencing confirmed the presence of FtPPK gene in pSV281 vector.

A.5.2 References

1. Rao, N. N.; Gomez-Garcia, M. R.; Kornberg, A., Inorganic polyphosphate: essential for growth and survival. *Annu Rev Biochem* **2009,** *78*, 605-47.

2. (a) Zhang, H.; Ishige, K.; Kornberg, A., A polyphosphate kinase (PPK2) widely conserved in bacteria. *Proc Natl Acad Sci U S A* **2002,** *99* (26), 16678-83; (b) Ishige, K.; Zhang, H.; Kornberg, A., Polyphosphate kinase (PPK2), a potent, polyphosphate-driven generator of GTP. *Proc Natl Acad Sci U S A* **2002,** *99* (26), 16684-8.

A.6 Subcloning of *Bdellovibrio bacteriovorous* Proline utilization A (BbPutA).

A synthetic gene encoding PutA from *Bdellovibrio bacteriovorus* HD100 (BbPutA, 982 residues, NCBI RefSeq number NP_968157.1) with codons optimized for expression in *E. coli* was purchased from BIO BASIC Inc. (Markham, Ontario CA). The gene was obtained in pUC57 vector with NdeI and BamHI sites engineered at N- and C- terminus respectively. BbPutA-pUC57 construct was transformed into *E. coli* DH5α and plated on to LB Agar supplemented with ampicillin. The plates were incubated at 37 °C in an incubator. The following day a single transformant was used to inoculate LB broth media supplemented with 50 μg/mL ampicillin and grown for approximately 12 hours at 37 °C/ 250 rpm. Next, the plasmid was extracted using manufacturer's protocol and the resulting plasmid was triple digested using NdeI, BamHI and BsaI at 37 °C for 2 hours. BbPutA free of the parent vector was gel purified using 1 % agarose and ligated into pKA8H vector (already digested with NdeI/BamHI and gel purified). The ligated product was transformed into *E. coli* DH5α, plated on to LB Agar supplemented with 50 μg/mL ampicillin and incubated at 37 °C overnight. Following incubation, four single colonies were picked and grown in LB broth supplemented with 50 μg/mL ampicillin. The resulting cultures were incubated at 37 °C and 250 rpm in an incubator-shaker and further used for plasmid preparation. Finally, the construct was verified by DNA sequencing. The full length BbPutA-pKA8H construct has N- terminal $(His)_8$ tag, cleavable by tobacco etch virus protease (TEVP).

VITA

Richa Dhatwalia was born in India and earned her Bachelors and Masters honors degrees in Chemistry from Panjab University, Chandigarh (India). She joined the Chemistry department at the University of Missouri – Columbia in the Fall 2009 and subsequently joined Prof. John J. Tanner's group in January 2010. Immediately after defending her PhD degree, she was offered a position at Sigma Aldrich Corporations, where she works as Chemist III in the enzyme's quality control group.