This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Structural Basis of Ligand Binding to UDP-galactopyranose Mutase from Mycobacterium tuberculosis Using Substrate and Tetrafluorinated Substrate Analogs.

Journal:	*Journal of the American Chemical Society*
Manuscript ID:	ja-2014-11204p.R1
Manuscript Type:	Article
Date Submitted by the Author:	19-Dec-2014
Complete List of Authors:	van Straaten, Karin; University of Saskatchewan, Chemistry A. Kuttiyatveetil, Jijin; University of Saskatchewan, Chemistry Sevrain, Charlotte; University of Namur (UNamur), Department of Chemistry Villaume, Sydney; University of Namur (UNamur), Department of Chemistry Jimenez-Barbero, Jesus; CSIC, Linclau, Bruno; University of Southampton, Department of Chemistry Vincent, Stéphane; 2University of Namur (UNamur) , Chemistry Sanders, David; University of Saskatchewan, Chemistry

SCHOLARONE™
Manuscripts

Structural Basis of Ligand Binding to UDP-galactopyranose Mutase from *Mycobacterium tuberculosis* Using Substrate and Tetrafluorinated Substrate Analogs

Karin E. van Straaten[1], Jijin R. A. Kuttiyatveetil[1], Charlotte M. Sevrain[2], Sydney A. Villaume[2], Jesús Jiménez-Barbero[3,4,5], Bruno Linclau[6], Stéphane P. Vincent[2] and David A.R. Sanders[1]*

[1]Department of Chemistry, 110 Science Place, University of Saskatchewan, Saskatoon, S7N 5C9, Canada.

[2]Department of Chemistry, University of Namur (UNamur), Rue de Bruxelles 61, 5000 Namur, Belgium.

[3]Department of Chemical and Physical Biology, Centro de Investigaciones Biologicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain

[4]Structural Biology Unit, CIC bioGUNE, Parque Tecnologico de Bizkaia Building 801A, 48160 Derio, Spain

[5] IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

[6]Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Running Title: Structure of TbUGM with substrate and inhibitors

*Address correspondence to:

Dr. David A.R. Sanders, Tel: 1-306-966-6788, Fax: 1-306-966-4730;
E-mail: david.sanders@usask.ca

Abstract

UDP-Galactopyranose mutase (UGM) is a flavin-containing enzyme that catalyses the reversible conversion of UDP-Galactopyranose (UDP-Gal*p*) to UDP-Galactofuranose (UDP-Gal*f*) and plays a key role in the biosynthesis of the mycobacterial cell wall galactofuran. A soluble, active form of UGM from *Mycobacterium tuberculosis* (*Mt*UGM) was obtained from a dual His_6-MBP tagged *Mt*UGM construct. We present the first complex structures of *Mt*UGM with bound substrate UDP-Gal*p* (both oxidized flavin and reduced flavin). In addition, we have determined the complex structures of *Mt*UGM with inhibitors (UDP and the dideoxy-tetrafluorinated analogs of both UDP-Gal*p* (UDP-F_4-Gal*p*) and UDP-Gal*f* (UDP-F_4-Gal*f*)), which represent the first complex structures of UGM with an analogue in the furanose form, as well as the first structures of dideoxy-tetrafluorinated sugar analogs bound to a protein. These structures provide detailed insight into ligand recognition by *Mt*UGM and show a similar overall binding mode as reported for other prokaryotic UGMs. The binding of the ligand induces conformational changes in the enzyme, allowing ligand binding and active site closure. In addition, the complex structure of *Mt*UGM with UDP-F_4-Gal*f* reveals the first detailed insight into how the furanose moiety binds to UGM. In particular, this study confirmed that the furanoside adopts a high energy conformation (^4E) within the catalytic pocket. Moreover, these investigations provide structural insights to the enhanced binding of the dideoxy-tetrafluorinated sugars compared to unmodified analogs. These results will help in the design of carbohydrate mimetics and drug development, and show the enormous possibilities on the use of polyfluorination in the design of carbohydrate mimetics.

Introduction

Mycobacterium tuberculosis is an intracellular human pathogen that targets alveolar macrophages and causes tuberculosis (TB), which kills around 2 million people worldwide every year. *M. tuberculosis* also exists in extracellular environments and causes disseminated disease.[1-3] The occurrence of multidrug (MDR) and extensively drug-resistant (XDR) strains of *M. tuberculosis* have emerged that are resistant to most or all known antibiotics.[4-6] Thus, there is an urgent need to develop new drugs against tuberculosis.

Galactofuranose (Gal*f*) is an essential component of the arabinogalactan that connects the peptidoglycan layer and the mycolic acid layer in the mycobacteria cell wall.[7] One key enzyme involved in Gal*f* metabolism is UDP-Galactopyranose mutase (UGM). UGM is a flavoenzyme that catalyzes the inter-conversion of UDP-Galactopyranose (UDP-Gal*p*) to UDP-Galactofuranose (UDP-Gal*f*), the biosynthetic precursor of all galactofuranose-containing prokaryotic and eukaryotic glycoconjugates (Scheme 1).[8-14] Deletion of the gene encoding for UGM in *M. tuberculosis* demonstrated that this enzyme is essential for mycobacterial growth.[15] Since Gal*f* and UGM are not found in humans, UGM is a validated target for therapeutic intervention.[14,16]

A number of groups have developed inhibitors against UGM, with varying degrees of success.[17-24] These potential inhibitors have included mechanism-based inhibitors[19,23,25-29] heterocyclic molecules obtained from high throughput screening of chemical libraries[18,30] and substrate analogue inhibitors.[17,20,31] UGM inhibitors have been identified that block the growth of mycobacterial cells[18] and a pyrazole-based compound was demonstrated to possess both UGM

inhibitory properties and broad anti-mycobacterial activities.[22] These data validated the pyrazole compound as a promising lead drug candidate against UGM.[22] Nevertheless, a better understanding of the UGM substrate tolerance is required to further assist the development of more potent inhibitors of therapeutic value.

Fluorosugar nucleotides have been synthesised to probe the UGM substrate specificity and the equilibrium position of the enzyme-catalysed reaction (Scheme 1). The 2-, 3- and 6-monofluoro-substituted UDP-Gal*p* analogs, as well as the 2- and 3- and 6-monofluoro-substituted UDP-Gal*f* analogs were found to be substrates for UGM.[32-36] In addition, UDP-D-talopyranose (C2 hydroxyl in axial orientation) and UDP-2-deoxy-2-fluoro-D-talopyranose (2-fluoro in axial orientation) are not substrates of UGM.[34] Recently N'Go *et al.*[37] reported the synthesis of dideoxy-tetrafluorinated analogs of both UDP-Gal*p* (UDP-2,3-dideoxy-2,2,3,3-tetrafluoro-D-*threo*-hexopyranose **3**, abbreviated as UDP-F_4-Gal*p*) and UDP-Gal*f* (UDP-2,3-dideoxy-2,2,3,3-tetrafluoro-D-*threo*-hexofuranose **4**, abbreviated as UDP-F_4-Gal*f*) and shown that these dideoxy-tetrafluorinated analogs are excellent inhibitors, but not substrates of UGM. The rationale for the introduction of this unusual fluorination motif originated from studies[38,39] in which affinity gain is sought through a combination of favourable hydrophobic desolvation energy and atractive multipolar interactions of the C–F groups with protein residues (with such interactions being negligible in aqueous medium). Interestingly, the K_d value of **4** was determined by STD-NMR competition experiments to be in the low micromolar range, around 5-10 μM, depending on the oxidation state of the enzyme, while the corresponding UDP-3-deoxy-3-fluoro-D-galactofuranose **5** showed a much higher K_d value (~400-600 μM).[35]

In 2004, Kiessling and coworkers characterized a covalent adduct of the FAD cofactor and a radiolabeled UDP-Gal*p* by mass spectrometry after having reduced the putative iminium intermediate **8** with sodium cyanoborohydride (Scheme 2).[8] These results strongly suggested that the three FAD-galactose covalent adducts **7**-**9** are intermediates of the reaction. Adducts **7** and **9** can be formed either through a direct attack of the nucleophilic $FADH^-$ cofactor onto the oxycarbenium **6** and **10**, or thanks to a preliminary SET followed by radical coupling. Importantly, a nucleophilic covalent attack by FAD on an anomeric carbon had not been detected previously. Moreover, the iminium adduct **8** could be characterized by UV spectroscopy. These key experiments were further improved[40,41] and applied to other UGMs,[42,43] including eukaryotic enzymes, and always lead to the conclusion that the FAD plays the role of the catalytic nucleophile. This mechanism is currently the most accepted one. However, some important mechanistic aspects are still under debate: a direct S_N2-like substitution of the UDP leaving group of **1** and **2** has also been proposed.[44,45] If verified, the two high-energy intermediates **6** and **10** would not be formed. Moreover, the possibility of SET has been proposed[46-48] but never ruled out.[49]

Detailed structural knowledge of the ligand-enzyme interactions is important for the design of specific UGM inhibitors, as well as yielding insights into the mechanistic aspects of Gal*f* biosynthesis. Crystal structure studies of prokaryotic UGM have been conducted on UGM from *Escherichia coli* (*Ec*UGM),[50] *Klebsiella pneumoniae* (*Kp*UGM),[46] *Deinococcus radiodurans* (*Dr*UGM)[51] and *Mycobacterium tuberculosis (Mt*UGM),[46] though only *Dr*UGM and *Kp*UGM have been crystallized in the presence of pyranosides or inhibitors.[40,41,51,52] Additionnally, structural studies have been conducted on an eukaryotic orthologue of UGM.[53,54] However, in spite of many efforts, none of these UGMs could be crystallized with UDP-Gal*f* or furanoside analogs. Therefore, the binding mode of UGM and its substrate UDP-Gal*f* has remained elusive.

Given the unique mechanism of this flavoenzyme and the fact that Gal*f* has a better affinity than Gal*p* for UGM, such knowledge would aid in confirming the key catalytic principles that govern this rather unusual isomerisation process.

The structures obtained so far have revealed that the 3-dimensional fold is largely conserved among prokaryotic and eukaryotic UGMs. They have also provided detailed insights on the conserved and unique features for ligand recognition and on the structural changes observed upon ligand binding.[55] In an effort to design antimicrobial drugs against the pathogenic organism *M. tuberculosis,* we focus our structural studies on *Mt*UGM. In this article, we present the crystal structures of *Mt*UGM in complex with its substrate, UDP-Gal*p,* and with the inhibitors UDP, UDP-F_4-Gal*p* and UDP-F_4-Gal*f*. These are the first atomic resolution complex structures reported for *Mt*UGM, for a complex involving a substrate with a Gal*f* configuration, and for complexes involving dideoxy-tetrafluorinated sugar analogs. In particular, our structural results provide detailed insights on how the enzyme retains key interactions with the tetrafluorinated ligand **3**, as compared to the natural ligand having a vicinal diol group in these positions **1**. We anticipate that this structural knowledge will not only provide a sound basis for further development of a new generation of potent UGM inhibitors, but is also the cornerstone for the design and employment of polyfluorinated carbohydrate mimetics on a regular basis.

Methods

Cloning of MtUGM- PCR

A pET29b plasmid containing the *MtUGM* gene was used as template in the Gateway cloning.[11] In order to construct the Gateway entry clone the *MtUGM* gene was first PCR amplified according to a modified procedure described by Nalamsetty and Waugh.[56] The PCR was performed in two separate steps using the same program cycling settings: initial melt for 5 min at 95°C, annealing 55°C for 30 s, and elongation 72°C for 2 min; 44 cycles of 95°C for 30 s, 55°C for 45 s, and 72°C for 1 min; 72°C for 10 min, hold at 4°C. The first PCR amplification reaction was carried out with two gene-specific primers, *Mt*UGM-N1 (5'-GAGAACCTGTACTTCCAGGGTATGCAACCGATGACC-3') and *Mt*UGM-C (5'-GGGGACCACTTTG TACAAGAAAGCTGGGTTATTATGCGCCGTCCTGAAGCAGTGG-3') with contain 5'extensions that add an in-frame TEV protease recognition site and an *att*B2 recombination site to the *MtUGM* gene N- and C-termini, respectively. After the first PCR amplification the PCR product was gel purified using the QIAquick gel extraction kit, according to the manufacturer's instructions. The obtained purified PCR product was then used as the template for the final PCR amplification with primers, N2 (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGAGAACCTGTACT TCCAG-3') and *Mt*UGM-C. Primer N2 is a generic primer[56] designed to anneal to the nucleotide sequence encoding the TEV protease recognition site and adds the *att*B1 recombination site to the N-terminus of the amplicon. The final *att*B-*MtUGM* PCR product was gel purified using the QIAquick gel extraction kit to remove residual *att*B primers.

Gateway cloning

Gateway recombinational cloning was performed according to the Gateway® Technology with Clonase®II manufactory manual (Invitrogen). To create the His_6-MBP-*Mt*UGM fusion vector, the final *att*B-*MtUGM* PCR product was first inserted by recombinational cloning using the

standard BP protocol (Invitrogen) into the donor vector pDONR^{TM}221 to yield the entry clone intermediate. 5 µl of the BP reaction product was transformed into Library Efficiency® DH5α™ Competent Cells (Invitrogen) and transformants were selected on LB agar plates containing kanamycin. Plasmid DNA was isolated from saturated cultures grown from individual kanamycin resistant colonies and screened by PCR, using primers *Mt*UGM-N1 and *Mt*UGM-C, to confirm that the clones have the expected structure. DNA sequencing was performed to confirm the correctness of the *Mt*UGM nucleotide sequence. Next, the entry clone intermediate was recombined into the cytoplasmic His_6-MBP destination vector (pDEST-His_6MBP) using the standard RL protocol (Invitrogen) to construct the His_6-MBP-*Mt*UGM overexpression vector. 5 µl of LR reaction product was transformed into Library Efficiency® DH5α™ Competent Cells (Invitrogen) and transformants were selected on LB agar plates containing ampicillin. DNA sequencing was performed to confirm the correctness of the *Mt*UGM nucleotide sequence. The His_6-MBP-*Mt*UGM overexpression vector was transformed into BL21CodonPlus™-RIL (Stratagene) cells and transformants were selected on LB agar plates containing both ampicillin (100 µg/mL) and chloramphenicol (30 µg/mL).

Production of wild-type MtUGM

During the cloning procedure of *Mt*UGM, a G917C point mutation was introduced, resulting in Pro306 to Arg306 mutation. This was initially noticed during structure refinement and later confirmed with DNA sequencing. Additional clones produced during the Gateway cloning step (above) were sequenced and a clone with no mutation was identified. This clone was then also expressed and purified following the procedure outlined below for the P306R mutant.

Expression of MtUGM.

The His_6-MBP–*Mt*UGM fusion protein was overexpressed in *E. coli* BL21(DE3) CodonPlus^{TM}-RIL cells (Stratagene). Briefly, *E. coli* BL21(DE3) CodonPlus^{TM}-RIL cells containing the pDest-His_6-MBP-*Mt*UGM overexpression vector were grown in 100 mL of Luria Broth (LB) supplemented with 100 µg/mL ampicillin, 30 µg/mL chloramphenicol and 0.4% D-(+)-glucose monohydrate (Sigma-Aldrich). Cells were grown overnight at 37 °C and 250 rpm. The next day two times 1L LB media, each supplemented with 100 µg/mL ampicillin, 30 µg/mL chloramphenicol and 0.4% D-(+)-glucose monohydrate, was inoculated with 25 ml of the overnight culture. Cells were grown at 15 °C to increase the solubility of the overexpressed MBP-*Mt*UGM fusion protein. When the cells reached early log phase (OD_{600nm} = 0.3–0.5) isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM. Twenty-six hours later cells were harvest by centrifugation at 8,000 rpm for 20 min and stored at -80 °C. Typically, a 2 L culture yields ~10 g of wet cell pellet.

Purification of MtUGM.

The cell pellet was re-suspended in 70 ml buffer A (25 mM Tris-HCL pH7.4 and 500 mM NaCl). After addition of 2 mM Lysozyme, 1 mM AEPSF, 20 µg/mL DNAse and 200 µl Halt^{TM} EDTA free protease inhibitors (Thermo Scientific), the resulting cell suspension was incubated on ice for 30 min. The cells were ruptured by sonication with a pulse-rest cycle of 10 sec on and 10 sec off, for a total of 3 min. Cell debris was removed by centrifugation at 20,000 rpm for 20 minutes at 4 °C. The supernatant was filtered through a 0.22 µ filter. Per run, 15 ml of the

supernatant was loaded onto three tandem 5 ml MBPTrap HP columns (GE Healthcare) pre-equilibrated with buffer A. The column was washed with 14 column volumes buffer A and bound His_6-MBP-*Mt*UGM fusion protein was eluted in 5 column volumes buffer B (25 mM Tris-HCl pH 7.4 and 500 mM NaCl and 50 mM maltose). Peak fractions containing pure His_6-MBP-*Mt*UGM fusion protein were pooled and concentrated six-fold using a 30KDa Amicon centrifuge filter devise. About 87 mg His_6-MBP-*Mt*UGM fusion protein was obtained from a 2L culture. The His_6-MBP dual tag was digested overnight at 4 °C with 1 mg His_6-tagged TEV protease[57] per 10 mg of fusion protein. The TEV protease-treated sample was filtered through a 0.22 µ filter and loaded onto a 14 ml Ni-Sepharose HP column pre-equilibrated with buffer A. The untagged *Mt*UGM protein passes through the column while residual undigested His_6-MBP-*Mt*UGM fusion protein, His_6-MBP and His_6-TEV protease remained bound. The purity of the *Mt*UGM containing fractions were analyzed by SDS–PAGE. To reduce the concentration of maltose in the *Mt*UGM protein samples, *Mt*UGM was washed with buffer A, through dilution and subsequently re-concentration. Finally, *Mt*UGM was concentrated to 6.5 mg/ml in 25 mM Tris pH 7.5 and 500 mM NaCl, flash frozen in small aliquots and stored at -80 °C.

Activity assay

For kinetic experiments the protein was purified as described above, however the buffer used was 50mM phosphate pH 7.5 and 320 mM NaCl. Kinetic constants for *Mt*UGM were determined as described previously for the kinetic assay of AfUGM[53]. The conversion of UDP-Gal*p* to UDP-Gal*f* was monitored at 262 nm using HPLC (Agilent Technologies, 1100 Infinity). A fixed concentration of *Mt*UGM (100nM) was used to have less than 40% conversion to the product UDP-Gal*p*. Reactions were carried out with varying amounts of UDP-Gal*f* (0 – 600 µM) in a final volume of 100 µL, 50 mM phosphate buffer pH 7.0 containing 20 mM freshly prepared sodium dithionite. The incubations were carried out for 2 minute at 37 °C and quenched with 100 µl *n*-butanol. After centrifugation the aqueous phase was injected on a Carbopac PA1 column. The nucleotide sugars were eluted isocratically with 0.2M ammonium acetate pH 7.0. The amount of conversion was determined by integration of the UDP-Gal*p* and UDP-Gal*f* peaks. The initial velocity was calculated from the substrate concentration and percentage UDP-Gal*p* conversion. Kinetic parameters were determined with GraphPad Prism software (GraphPad Software, San Diego, CA) using nonlinear regression analysis. All experiments were performed in duplicate.

Crystallization of MtUGM

Diffraction quality crystals of *Mt*UGM complexed with the substrate UDP-Gal*p* and inhibitors (UDP, UDP-F_4-Gal*f* and UDP-F_4-Gal*p*) were obtained using the Hanging drop method at 4 °C. For detailed description of the crystallization experiments see Supplemental Text. Briefly substrate/ inhibitors (final concentration of 20 mM) was added to the protein solution (6.5 mg/ml in 25 mM Tris pH 7.5 and 500 mM NaCl) and prior to crystallization was reduced by adding sodium dithionite (final concentration 20 mM). Drops (2.4 µl) were prepared by mixing equal volumes of protein solution (1.2 µl) and reservoir solution (1.2 µl), wells contained 0.5 ml crystallization solution. Plate-like crystals were obtained within 1 week using 20% PEG 3350, 0.1 M Bis-Tris pH 5.5 and additives. Best diffracting crystals were obtained by allowing the crystals to grow for one month. Crystals were cryo-protected in crystallization solution containing 30% ethylene glycol and flash-cooled in liquid nitrogen.

Data collection and processing

Datasets for *Mt*UGM in complex with UDP, UDP-Gal*p*$_{ox}$ (with oxidized FAD), UDP-Gal*p*$_{red}$ (with reduced $FADH_2$), UDP-F_4-Gal*p* and UDP-F_4-Gal*f* were collected at beamline 08ID-1 on a Rayonix MX300HE X-ray detector. The data sets were processed and scaled using autoprocess[58] and d*TREK.[59] The data collection statistics are shown in Table 1A.

Structure determination and refinement

The *Mt*UGM:UDP-Gal*p* complex structure was determined by molecular replacement using MOLREP[60] within the CCP4 package.[61] The structure solution was found by using the unliganded *Mt*UGM structure (PDB code 1VOJ) as the search model.[46] The unliganded *Mt*UGM structure missed the loop residues 138-141. Inspection of the electron density maps showed that these residues could be build. Residues 135-143 were (re)build and the unliganded *Mt*UGM structure was then re-refined with good geometry. This structure was used for the structural comparisons between the unliganded and liganded *Mt*UGM. The *Mt*UGM:UDP-Gal*p* structure was used as the starting model for all other *Mt*UGM complex structures (UDP, UDP-Gal*p*(red), UDP-F_4-Gal*p* and UDP-F_4-Gal*f*). Refinement of all structures was done with PHENIX.[62] Initially, rigid body refinement was done, followed by simulated annealing using Cartesian dynamics at 5000 K to remove model bias. Iterative rebuilding of the model was done in COOT[63] followed by simulated annealing using Cartesian dynamics at 2500 and gradually lowering to 1000K. NCS restraints were used throughout the refinement for all models. Placement of cofactor and ligands was done with ligandFit in PHENIX.[62] The models for UDP-F_4-Gal*p* and UDP-F_4-Gal*f* were generated in SKETCHER as part of CCP4 suite. Libraries for cofactor and ligands were generated with ELBOW in PHENIX.[62] The refinement progress was monitored by following R_{free} and inspecting the electron density maps. When R_{free} dropped below 30%, water molecules were added using water update refinement in PHENIX, and their positions were manually checked using COOT. The final round of refinement was done with optimized refinement target weights for best geometry. Final refinement statistics are shown in Table 1B.

Structural analysis

The stereochemistry of all models was validated with MOLPROBITY [64] as part of PHENIX[62] and the ADIT validation server at RCBS-Rutgers (http://deposit.pdb.org/validate). Superpositions were calculated with DALI-lite[65] and with SUPERPOSE within the CCP4 package.[66] Superpositions of the open and closed *Mt*UGM structures were done with SUPERPOSE by superimposing specific residue selections (residues 4-120 and residues 186-395) and excluding the domain-2 mobile residues 121-185. Conformational changes due to Domain motions were analysed by DynDom.[67] Figures were prepared with PYMOL (http://www.pymol.org), RASMOL[68] and ESPript.[69]

Protein Data Bank accession numbers

Coordinates have been deposited in the PDB with accession codes: *Mt*UGM with UDP-Gal*p*(ox), 4RPG; *Mt*UGM with UDP-Gal*p*(red),4RPH; *Mt*UGM with UDP, 4RPJ; *Mt*UGM with UDP-F_4-Gal*p*, 4RPL and *Mt*UGM with UDP-F_4-Gal*f*, 4RPK.

Results and discussion

MBP enhanced the overexpression and purification of soluble active MtUGM.

E. coli Bl21(DE3) cells harbouring the *Mt*UGM-pet29b expression construct for overexpressing N-terminal His_6-*Mt*UGM resulted in insoluble protein (inclusion bodies). We tried different expression vectors and *E. coli* hosts and used different growth temperatures. However, we were not able to improve the solubility of *Mt*UGM (data not shown). The solubility of a recombinant protein can often be improved when fused to a highly soluble fusion partner. Maltose binding protein (MBP) is known to enhance the solubility of proteins when used as a fusion partner. [70] The His_6-MBP–*Mt*UGM fusion protein could be easily overexpressed in *E. coli* BL21(DE3) CodonPlus^{TM}-RIL cells and purified to homogeneity by means of affinity chromatography. From a 2 L culture about 6 mg of pure *Mt*UGM was obtained. The rate of conversion of UDP-Gal*f* to UDP-Gal*p* was determined at 37°C in the presence of 20 mM sodium dithionite. Although both MBP-*Mt*UGM and *Mt*UGM are active, we have only determined the kinetic parameters for the latter one. The structural studies were conducted using the P306R mutant enzyme, but we also successfully expressed and purified the wild-type enzyme. The kinetic parameters for the wild-type and P306R were obtained and found to be comparable to each other and to other prokaryotic UGMs (Table 2). These results lead us to believe that the P306R mutant will be structurally similar to the wild-type enzyme.

Overall complex structure of MtUGM shows significant structural changes compared to unliganded MtUGM

*Mt*UGM complex structures crystallize in space group C2. The crystals contain three monomers in the asymmetric unit, including one non-crystallographic two-fold dimer and a second crystallographic two-fold dimer, formed by the third monomer and a symmetry-related monomer. Each monomer in the dimer is composed of the 3 characteristic UGM domains (FAD binding domain-1, α-helical domain-2 and β-sheet domain-3) (Figure 1A-B). Crystal structures of *Mt*UGM were determined in complex with the substrate UDP-Gal*p* (both reduced and oxidized flavin), with UDP (oxidized flavin), and with tetra-fluorinated analogs of both UDP-Gal*p* (UDP-F_4-Gal*p*) and UDP-Gal*f* (UDP-F_4-Gal*f*)). All complex structures are highly similar with root mean square deviation (rmsd) for all overlapping Cα atoms of less than 0.3 Å.

The P306R mutatant used in these studies is kinetically similar to the wild-type enzyme. Pro306 is located on the solvent exposed loop (His300-Lys309) connecting the small helix η3 and beta strand β14 of the β-sheet domain (Supplemental Figure S1), over 25 Å from the FAD. The Cδ atom of Pro306 is 3.5 Å from the main chain oxygen of Thr53, located on the small sharp turn (Glu52-Gly56) connecting beta strands β3 and β4. The Pro306Arg mutation releases this clash and results in a 1 Å shift in the position of the two solvent exposed loops without effecting the position of side chains and interaction with the protein. Arg306 forms a salt bridge with the side chain of Asp308 and the main chain oxygen of Gln54 and replaces a salt bridge formed by Lys309. The Lys309 side chain has rotated and forms a new salt bridge with Gln54. In addition, Lys309 is involved in crystal contacts with the side chain of Asp202. The Pro306A mutation therefore may be stabilizing the two loops and promoting the crystallization of *Mt*UGM complex structures.

The comparison of the overall structures between the liganded and the previously reported unliganded *Mt*UGM[46] showed that they are highly similar, with a rmsd value of 0.6 Å for all equivalent Cα atoms, excluding the residues located at the mobile loops. The major difference is the closure of the mobile loops around the active site due to ligand binding. The binding of

ligand results in a large re-orientation of the α-helical domain-2 (residues Phe102-Asp194) (Figure 1C). The mobile loops 1 (Ala131-Asn140) and 2 (Gln167-Arg184) show a similar movement upon ligand binding as seen in other prokaryotic UGMs.[40,41] The structural rearrangement upon ligand binding in *Mt*UGM is far more profound. The structures were analyzed by DynDom[67,71] and showed that mobile loop 1 (with adjacent helices α5 and α6) and mobile loop 2 rotate ~32° and translate 0.7° towards the uridine-diphosphoryl moiety of the ligand resulting in a ~67% closure of the active site. The C-terminus of helix α4 (Ala127-Glu132), helices α6 and α7 (Leu153-Trp166) and residues Ile178-Thr186 therefore function as a flexible hinge (Figure 1D). In addition, the rearrangement of mobile loop 2 results in formation of helix α8 (Ala175-Arg180), adjacent to the uridine-diphosphoryl moiety of the ligand. The closure of the active site will likely protect the reaction intermediates from the environment and potentially prevent the putative UDP intermediate from exiting the active site, as proposed for other UGMs.[51,53,72]

Substrate binding induces local changes in MtUGM active site

The structures of *Mt*UGM$_{ox}$:UDP-Gal*p* and *Mt*UGM$_{red}$:UDP-Gal*p* are nearly identical, with a rmsd value of 0.3 Å for all equivalent Cα atoms. The FAD N5 atom is closer to the C1 atom of the Gal*p* moiety (3.8 Å) compared to the *Mt*UGM$_{ox}$ structure (4.2 Å). The electron density of the Gal*p* moiety is stronger in the *Mt*UGM$_{red}$ complex structure. In the *Mt*UGM$_{ox}$ structure, there is visible sugar density only for molecule B whereas in the reduced structure, there is visible density for all 3 monomers. A similar trend was observed for *Dr*UGM,[51] and suggests the stabilization of the galactose orientation for catalysis.

The crystal structures of both unliganded *Mt*UGM and of the *Mt*UGM:UDP-Gal*p* complex structures enables determination of local structural changes that occur upon substrate binding. Figure 2A and Scheme 3 show the binding mode of *Mt*UGM with UDP-Gal*p* (additional details are included in the Supplemental Information). The positions of the majority of the active site residues remain unchanged when substrate binds. However, several active site residues change position in order for the substrate to bind in a productive binding mode and for the active site to close (Figure 2B). The conserved arginine (Arg180), located on the mobile loop, moves ~9.4 Å (Cα position) upon closure of the active site and interacts with the substrate through the pyrophosphate and Gal*p* 2-OH and 3-OH groups (Figure 2B and 3A). The other conserved active site arginine, Arg292, has a different rotamer conformation in unliganded *Mt*UGM. The guanidinium moiety of Arg292 rotates 180° when the substrate binds and interacts with the UDP β-phosphate and the O5 atom of the Gal*p* moiety. The third strictly conserved arginine, Arg360 at the FAD binding domain, interacts with the FAD phosphate in the unliganded structure and moves to the active site where it interacts, through a conserved H_2O molecule, with both the pyrophosphate and the 2-OH of Gal*p*. In addition, Arg360 helps positioning Tyr366. The side chain of Tyr366 rotates ~90° and shifts 5.4 Å with respect to its hydroxyl group to stabilize the β phosphate group of the UDP moiety. Furthermore, the aromatic ring forms a cation-π interaction with Arg180 and stabilizes the closed conformation of the enzyme.

The uridine-ribose binding region shows the most dramatic structural rearrangements (Figure 2B). Upon binding of the substrate, the side chain of Gln165 changes position to allow proper binding of the ribose moiety and the aromatic ring of Tyr161 rotates ~45° to form the cation-π interaction with the uracil. Ile178 and Leu181 of the newly formed helix α8 (mobile loop 2) are adjacent to the uridine-diphosphate group and aid positioning the uracil-ribose moiety (Figure 2B). Furthermore, the side chain of Trp166 rotates ~45° and interacts with the ribose hydroxyls

(Figure 3A) and with Leu173 of mobile loop 2. The rotation of the Trp166 side chain disrupts the cation-π interaction between Trp166 and Arg261, causing the Arg261 side chain to rotate 180° away from the substrate binding site. Nevertheless, this new orientation is stabilized by interactions with the main chain carbonyl of Gln165, the side chains of Gln167 and Ser317 (Figure 3A and B), and with Phe319, that rotates ~45° to form a new cation-π interaction. In addition, the C-terminus of β14 (Ser317-Phe319) and the connecting loop (Ala320-Pro326) have moved closer to the FAD binding site (Figure 3B). This loop movement triggers repositioning of Arg360 towards the active site to prevent a clash with Pro326. The Ser317-Glu321 segment is also functioning as part of the flexible hinge (Figure 1D).

Potential roles of Trp166 and Arg261 in regulating prokaryotic UGM domain closure

Trp166 and Arg261 are strictly conserved among prokaryotic UGM and are found on the edge of the active site cleft. Figure 3 shows the superposition of open unliganded *Mt*UGM structure with the closed *Mt*UGM:UDP-Gal*p* complex structure. Trp166 is located at the C-terminus of helix α7 and the beginning of mobile loop 2. Arg261 is located above the C-terminus of helix α7 and is part of the FAD binding domain. Trp166 and Arg261 may act as a switch to maintain the active site in a locked and fully open conformation by forming a cation-π interaction between the two residues, as found in unliganded *Mt*UGM and in molecule B of unliganded *Ec*UGM.[46,50] Substrate binding, with concomitant rotation of Trp166 and disruption of the cation-π interaction presumably triggers the closure of the active site. In the open conformation, where Trp166 and Arg261 form the cation-π interaction, the closure of mobile loop 2 would result in a steric clash between Leu173 and Trp166 (Figure 3A). Indeed, the analysis of the semi-closed and closed conformations described for other liganded and unliganded prokaryotic UGM structures show that the cation-π interactions between the strictly conserved tryptophan and arginine residues are broken and that these two residues are found in the same orientation as in the liganded *Mt*UGM.[40,41,46,50,51] This evidence suggests that the conserved tryptophan and arginine residues play key roles in regulating prokaryotic UGM domain closure.

Analysis of the binding modes of the four UGM ligands

MtUGM:UDP complex structure

The crystal structure of *Mt*UGM:UDP was determined to 2.5 Å resolution. All 3 monomers are identical and contain bound UDP and oxidized FAD. The overall structure of *Mt*UGM:UDP complex is highly similar to that of *Mt*UGM:UDP-Gal*p* complex (rmsd less than 0.4 Å). UDP is bound in the closed conformation and makes the same interactions in the active site as seen for the UDP moiety of the substrate UDP-Gal*p* (Figure 5A and B). The galactose binding pocket is filled with water molecules that are nicely accommodated in the absence of the sugar moiety (Figure 5A). The observed closed conformation of the active site indicates that the binding of the UDP moiety is enough to promote closure of the active site, as we previously reported.[51]

UDP-Gal*p *(1) binding is similar to other prokaryotic UGMs

The similarities in the substrate binding between *Mt*UGM and other prokaryotic UGMs reveal a conservation of the binding site. A structure-based sequence alignment of *Mt*UGM with other known prokaryotic UGMs is shown in Supplemental Figure S2.

Superposition of monomer B from the reduced *Mt*UGM:UDP-Gal*p* and *Dr*UGM:UDP-Gal*p* complex structures reveals that 361 residues overlap with a rmsd of 1.3 Å. The substrate is bound in an identical position and orientation in both complexes, with its Gal*p* moiety close to the isoalloxazine N5 and makes identical interactions in the sugar-phosphate-ribose binding regions as observed in the *Dr*UGM:UDP-Gal*p* complex[51] (Figure 4A and B). However, although the interactions in the uridine binding pocket are not strictly conserved (Figure 4B), this binding pocket is predominantly hydrophobic in both structures. On one side of the pocket, the uracil is stacked against a strictly conserved tyrosine residue (Tyr161). In addition, the O4 atom of the uracil moiety is anchored by both conserved Asn282 and non-conserved Asn284. The other side is aligned by the hydrophobic residues Phe102, Val158, Ile178 and Leu181. These residues are not strictly conserved among prokaryotic UGM and help in positioning the uridine moiety when the active site closes.

***MtUGM:UDP-F$_4$-Gal*p** (3) complex structure*

The crystal structure of *Mt*UGM:UDP-F_4-Gal*p* **3** was determined to 2.25Å resolution. The increased affinity of UGM for the deoxyfluorinated UDP-Gal*p* compared to the unfluorinated UDP-Gal*p* is manifested in the improved electron density for the F_4-Gal*p* moiety in this structure. All 3 monomers contain bound UDP-F_4-Gal*p* **3** and oxidized FAD. Monomers A and B display the closed conformation. Monomer C is found in the open conformation, with two alternate conformations of the active site mobile loop (open and closed). Superposition of monomer A from *Mt*UGM:UDP-Gal*p* and monomer A of *Mt*UGM:UDP-F_4-Gal*p* complex structures reveals that all 390 residues nicely overlap with a root mean square deviation of less than 0.3 Å. Thus, the binding mode of UDP-F_4-Gal*p* **3** is strikingly similar to that of UDP-Gal*p* (Figure 6B). As expected, the F_4-Gal*p* conformation only shows minor distortions from a $^4C_1$ conformation.[73] However, the presence of the fluorines at the C2 and C3 positions of the galactopyranose ring induces changes in the pyranose ring orientation (Figure 6A and 6B). Compared to the unfluorinated Gal*p*, the dideoxy-tetrafluorinated Gal*p* is tilted 0.5 Å at its C2, C3 and C4 carbon atom positions. In this presentation mode, the tetrafluorinated sugar moiety is in closer contact with the FAD isoalloxazine ring. As a result, the N5 of the isoalloxazine ring of FAD is pushed away and is buried about 0.9 Å deeper into the FAD binding pocket. Interestingly, although this displacement causes the O4 carbonyl atom of the isoalloxazine ring to shift ~0.8 Å, the hydrogen bond with the C4 hydroxyl group of F_4-Gal*p* is still kept.

Scheme 4 shows the direct contacts between the fluorine atoms and *Mt*UGM (additional details are included in the Supplemental Information). The C2 axial fluorine (F2β) displays a close contact (2.3 Å) with the O4 carbonyl of the isoalloxazine ring and additional interactions with the carbonyl oxygen (3.3 Å) of Ala64. In addition, the C2 equatorial fluorine (F2α) directly interacts with the guanidinium of Arg180 and, through a conserved water molecule, with Arg360 and Asp368 (The F…OH_2 distance is 3.7 Å), similar to the interaction between the C2 hydroxyl group and enzyme in the UDP-Gal*p* structure. The CF_2 group at the 3-position is located within a hydrophobic environment formed by the aromatic rings of Tyr191 and Phe192. The distance of the C3 equatorial fluorine (F3β) to Cϕ of Phe192 (3.7 Å) is similar to that observed for the C3 hydroxyl group of Gal*p* in the corresponding complex. The C3 equatorial fluorine (F3β) largely interacts with the O4 carbonyl of the isoalloxazine ring ($d_{F\ldots O}$ = 2.6 Å), while the F3α group establishes interactions with the α-phosphate moiety, through a conserved water molecule (The

F…O distance is 3.2 Å to the water molecule). Strikingly, three out of the four fluorine atoms of UDP-F_4-Gal*p* **3** display very close distances (from 2.2 to 3.2 Å) to the cofactor O4 carbonyl.

Comparison of molecule C (open form) with the closed UDP-F_4-Gal*p* **3** complex structure reveals that the α-helical domain (domain 2) is in the open conformation, similar to that described for the unliganded *Mt*UGM structure. Given the open conformation of the α-helical domain, the uracil moiety is less tightly bound. As mentioned above, the active site mobile loop (loop 2) is found in two alternative positions (both open and closed). In the open form, this mobile loop makes no contacts with the ligand. In contrast, in the closed conformation, the loop still provides the same interactions as in the closed monomers. The positions of most of the active site residues and UDP-F_4-Gal*p* **3** in this conformation remain largely unchanged and provide analogous interactions to those described above (Figures 6C and 6D). The only two residues that significantly changed their positions are Tyr366 and Arg360, which now display orientations similar to those seen in the unliganded *Mt*UGM structure. Presumably, the additional strain on the FAD isoalloxazine ring described above triggers the domain opening to release the clash introduced by the C2 axial fluorine (F2α, at 2.7 Å in closed conformation). Due to this rearrangement, Tyr366 is now unable to be hydrogen bonded to the β-phosphate, resulting in a rotation of the β-phosphate and the F_4-Gal*p* moieties around the O3 atoms of α-phosphate and β-phosphate, respectively (Figure 6C and 6D). These conformational changes result in a final position of the C2 axial fluorine F2β at 3.4 Å from the N5 of the FAD isoalloxazine ring.

MtUGM:UDP-F_4-Gal*f* complex structure

The crystal structure of *Mt*UGM:UDP-F_4-Gal*f* **4** was also determined to 2.4 Å resolution. Importantly, this is the first reported UGM complex structure with a bound hexafuranose sugar. All three monomers contain bound UDP-F_4-Gal*f* **4** and oxidized FAD. The 3 monomers are basically identical (average root mean square deviation of 0.2 Å for all atoms). The conformation and interactions of the UDP moiety of the inhibitor is identical to that of the other *Mt*UGM complexes already described. The F_4-Gal*f* moiety of **4** is in a similar position and orientation as that observed for the Gal*p* moiety of **1** and establishes major contacts with the *re* face of the FAD isoalloxazine ring (Figure 7A, 7B and Scheme 5). Noteworthy, the anomeric carbon atom (C1) of UDP-F_4-Gal*f* is positioned 4.4 Å from the FAD N5 atom, a distance that is likely too far for covalent catalysis (additional contact details are included in the Supplemental Information). However, it should be noted that these structures have been obtained under non-reducing conditions (oxidation of flavin occurring in the course of crystallization) and we cannot rule out that this distance is an artifact of these conditions and the substrate is positioned closer to the cofactor. In support of this, in all cases where bacterial UGMs have been crystallized with substrate (*Dr*UGM[51], *Kp*UGM[40] and this work), the distance between N5 of FAD and the anomeric carbon decreases upon reduction of the cofactor. Unfortunately, a reduced FAD complex between *Mt*UGM and UDP-F_4-Gal*f* was not successfully determined.

The C2 β-fluorine F2β is fairly close to the N5 (2.7 Å) and O4 carbonyl atoms (3.1 Å) of the FAD moiety. The analysis also indicated that F2β displays additional contacts through structurally conserved water molecules with the polar side chains of Arg360, Asp368 and His68. Indeed, in this presentation mode, in the reduced state of the cofactor, the C2 β-fluorine F2β would clash with the cofactor. Interestingly, the C2 α-fluorine F2α of UDP-F_4-Gal*f* **4** is found in the same position as the C3 axial fluorine F3α of UDP-F_4-Gal*p* **3** (Figure 7C) and provides the same interactions with the enzyme. The C3 β-fluorine F3β of UDP-F_4-Gal*f* **4** is in a similar

position as the C4 axial hydroxyl group of UDP-Gal*p*, although the C3 carbon atom of **4** is tilted about 1.4 Å with respect to the C4 carbon of **1**. The position of the F3β fluorine is shifted by about 1.9 Å with respect to the C4 hydroxyl group, and, as with F2β, makes contacts with the O4 carbonyl of the isoalloxazine ring, as well as with the carbonyl oxygen of Ala64 (as described above for F2β of **3**). This observed shift is likely induced by the electrostatic interaction with the O4 carbonyl of the isoalloxazine ring, and to minimize the possible steric contacts of the C3 α-fluorine F3α with Leu66, His89 and with the aromatic ring of Phe192. Otherwise, the C3 α-fluorine F3α would point in the direction of the C4-equatorial hydrogen in UDP-Gal*p* **1**.

N'Go *et al.*[37] have reported that UDP-F_4-Gal*f* **4** adopts a high energy conformation (^4E) in its *Mt*UGM-bound state, as evidenced by a H1-H4 nOe crosspeak obtained by a TRNOESY experiment, as opposed to the minimum energy conformation adopted in the free state ($^3T_2$). This *in solution* NMR-based observation is now confirmed by the X-ray crystallographic data obtained in this study.

Common binding features of the 3 nucleotide-sugars

Inspection of the structures of UGM bound to **1**, **3** and **4** clearly show that the binding mode of the three molecules is basically the same: 1) the three UDP moieties are essentially identical in orientation, they interact with the same amino-acids, and display the same conformation; 2) in all cases, the endocyclic oxygen of the galactose ring points away from the FAD cofactor, thus exposing the C2-C3 side of the galactose residue to the cofactor. Moreover, in all the oxidized structures, the distance between the anomeric carbon atom and the FAD N5 atom is always above 4.2 Å, a surprisingly long distance for a mechanism involving covalent catalysis. This distance shortens to 3.9 Å for the UDP-Gal*p* structure with reduced cofactor. In fact, for the four structures, the FAD N5 atom is always significantly closer (< 4.0 Å for the oxidized structures and 3.2 Å for the reduced structure) to the C-2 position of the galactose rings. These results are consistent with the previous structures that all show a closer orientation of anomeric carbon with the N5 of FAD after reduction of the cofactor.[40,51,53]

Comparison of the UDP-Galp:UGM and UDP-F_4-Galp:UGM structures.

Despite the extensive hydrogen bonding of the UDP-Gal*p* sugar OH2 and OH3 groups, the drastic change of this vicinal diol moiety to a tetrafluorinated ethylene group clearly does not have a deleterious effect on the interactions with UGM residues in the binding pocket, despite the annihilation of any hydrogen bond donating capacity, and the reduced capacity to accept hydrogen bonds.[74-76] The observed tilting of the pyranose sugar ring is the likely structural consequence from the new set of interactions of the tetrafluorinated moiety with the enzyme.

It is interesting to observe that the fluorine atoms remain engaged in a great number of interactions, including those mediated by water (eg, the C2 equatorial fluorine (F2α) interacts with the guanidinium of Arg180 and, through a conserved water molecule, with Arg360 and Asp368, in an analogous manner to the interactions of the C2 hydroxyl group of Gal*p* (The F…O distance is 3.7 Å).. The interaction of a C-F with water molecules is of great general interest given the ongoing debate regarding its hydrogen bond accepting capacity.[74,76-79] To the best of our knowledge, few crystal structures showing water-mediated hydrogen bonds with a fluorinated carbohydrate have been reported.[80,81] However, the fluorine atoms in these examples were CH-F groups, which are superior hydrogen bond acceptors as compared to the fluorine atoms within our CF_2 fragment.[74,75]

The alcohol groups in UDP-Gal*p* extensively interact with the FAD isoalloxazine ring. For the tetrafluorinated pyranose of **3**, close contacts with three out of the four fluorine atoms are seen. Fittingly, an attractive orthogonal multipolar interaction, as defined by Müller and Diederich,[82,83] can be invoked between the C^2-F^{α} bond with the FAD C^4=O carbonyl (∠ F…C=O 70.7°; $d_{C\text{-}F}$ 3.4Å). In the latter case, the C-F/C=O dihedral angle (θ= -173°) is also consistent with the existence of an attractive dipole-dipole interaction (Table S2).

The C3-F_2 group is located at a hydrophobic environment provided by two contiguous aromatic amino acids. Very probably, the existing interactions between the CF_2 fragment and the enzyme are of stabilising nature, presumably due to hydrophobic desolvation. In fact, when compared to complexed **1**, with a 3-OH group close to Phe192, it is likely that the binding of the CF_2 moiety at this site should be favored.

Comparison of the interactions of the furanoside 4 with pyranosides 1 and 3

In general, based on the gathered structural data, there are fewer strong attractive H-bonds for the two pyranosides **1** and **3** compared to the furanoside **4** (See Schemes 3-5, and ESI). This fact may explain why furanosides (substrates or inhibitors) have always better affinity for UGM than the corresponding pyranoside analogs.[29,31-34] The C5-OH of Gal*f* is located in the same position as the C6-OH of Gal*p* and forms one analogous hydrogen bond to His89. In addition, the C6 hydroxyl group of Gal*f* is anchored by Asn282, Tyr161 and Arg292, which also show one additional interaction with O4, resembling that with O5 of the Gal*p* moiety. The C5 and C6 carbon atoms of Gal*f* are properly positioned to provide hydrophobic interactions with Tyr191, Val91, Val280 and Arg292. Globally, this strong set of attractive contacts at the vicinity of the C4-C5-C6 part of Gal*f* **4** are absent in the pyranosidic substrate **1** and in the inhibitor **3**. These observations likely explain why it was found, in a previous study, that the 6-fluoro-analogue of UDP-Gal*f* was a poor UGM substrate with a dramatic loss of binding affinity.[36] This fact was further verified with 6-deoxy analogs and UDP-L-arabinose (a UDP-Gal*f* analogue without the $^6CH_2OH$ group).[84]

From these three structures, His89 is the common closest residue to the O4 and O5 positions of the sugar rings. This residue could play the role of proton relay during the catalytic process. Indeed, the reversible interconversion between **1** and **2** requires a proton transfer between the O-5 and O-4 of the galactose moiety.

The comparison of the observed interactions involving the fluorine atoms between the furanose and pyranose deserves further attention. The F2β fluorines of **3** and **4** are always the closest atoms to UGM (more precisely, to its cofactor), while the fluorine F3α is precisely the furthest one. However, the two carbohydrate rings are not interacting with the protein in the same way. For the pyranose **3**, the F2α and the two β fluorine atoms are clearly interacting with the FAD moiety. In contrast, for the furanose **4**, only the two β fluorines are in contact with the FAD cofactor. It is also worth to mention that, for both galactose forms, the four fluorine atoms of **3** and **4** are always in close proximity to the oxygen atoms of the C=O carbonyls of FAD and Ala64.

Given the fact that F_4-Gal*p* is bound in a similar conformation as the Gal*p* moiety in UDP-Gal*p,* we may anticipate that the F_4-Gal*f* geometry represents a close representation of the actual binding mode of the Gal*f* ring in the natural compound. However, due to the different interactions caused by substituting the C2 and C3 hydrogens with fluorine atoms, the furanose ring might be slightly tilted, as also observed for F_4-Gal*p*.

Comparison of furanosides 4 and 5

From these structures, one can try to answer why the monofluorinated UDP-Gal*f* analogue **5** (Figure 1) displays poorer affinity for *Kp*UGM than the tetrafluorinated analogue **4**. Assuming both *Galf* analogs bind in the same mode to UGM, it can be safely hypothesized that the increased electron density of the fluorine atom F3β atom in **5** compared to the corresponding atom in **4** would result in strong(er) repulsive interactions with the O4 and N5 FAD atoms. This is in accord with empirical observations made by Dalvit and Vulpetti regarding the correlation of the fluorine chemical shift (shielded CHF fluorines being more electron rich than deshielded CF_2 fluorines) with the magnitude of repulsive dipolar/electrostatic interactions.[85]

Conclusions

The crystal structures reported herein provide the first examples of UGM from the pathogenic organism *M. tuberculosis* complexed with ligands and thus represent the first structures of ligands bound to this key antimicrobial drug target. These structures provide critical insights for further drug design and *in silico* docking studies. Combined with the previously reported unliganded *Mt*UGM structure, these structures have also enabled us to detail the conformational changes that occur upon substrate binding. The previously reported prokaryotic structures complexed with ligands were less than ideal for these comparisons. In the case of *Dr*UGM, there was no unliganded structure and, for *Kp*UGM, the crystals were initially formed using UDP-glucopyranose and then soaked with UDP-galactopyranose after the crystals were formed. As UDP-glucopyranose is a very poor inhibitor of UGM, it is likely that the structural changes noted for the *Kp*UGM complexes did not represent the complete set of changes that take place in the real process. Herein, we have demonstrated that the binding of ligands and inhibitors indeed induces large and local conformational changes in the enzyme to allow the ligand to bind and the active site to close. Although the hinged closure of the active site has been alluded to in previous reports,[50] this research describes the first direct observation of the hinged nature of the closure. The UDP moieties of all the structures bind in nearly-identical manners to *Mt*UGM, with minor differences seen in the sugar binding regions.

We have derived the first 3D-structure of a polyfluorinated carbohydrate analogue complexed with a protein. It is remarkable that for both the pyranoside and furanoside structures, the CF_2-CF_2 motif is involved in multiple interactions with the enzyme, with or without water mediation, as well as multipolar interactions as described by Diederich *et al*.[82] Our observations strongly suggest that the balance of the entropy and entalphy contributions associated to the hydrophobic desolvation effect of the tetrafluorinated system is favourable for the molecular recognition process. Though it is not possible to quantitatively compare all the interactions connected to the ^{2}CH(OH)-^{3}CH(OH) groups of UDP-Gal*p* 1 with those linked to the tetrafluorinated moiety in **3**, these novel interactions are likely at the origin of the favorable binding process of these sugar analogs. Although merely speculative, it appears that the dideoxy-tetrafluorination modification results in both stabilizing enthalpic and entropic contributions to binding, as proposed for the binding of aromatic ligands to human carbonic anhydrase.[86,87]

This study thus validates, in a structural manner, that the polar and hydrophobic character of polyfluorinated groups such as CF_2-CF_2 can improve the contacts between a carbohydrate analogue and its host-protein in a significant manner. The possibilities of polyfluorination of carbohydrates have not been extensively investigated compared to the other main classes of biomolecules.[38,39,88] For instance, polyfluorination of aromatic amino acids has been explored to

enhance protein stability[89,90] and to promote the formation of β-peptides bundles.[91] Polyfluorinated aromatic ligands have also been exploited to discover novel attractive organofluorine-protein interactions[92] and, more recently, to address the key question of the hydrophobic effect that underlies the binding of many ligands to proteins.[86] Trifluoromethylated and polyfluorinated lipids have also been exploited either as direct substrates of lipid- and glycolipid modifying enzymes.[93] Although it has been recently shown than a CF_2 moiety at the pyranose ring may restore the key exo-anomeric effect,[94] to the best of our knowledge, the structural basis to justify the replacement of a carbohydrate vicinal diol for a polyfluorinated system in molecular recognition events has never been reported to date. These observations, along with the discovery of the hexafuranose binding mode to UGM, should now provide the impetus for the development of yet more potent inhibitors.

Acknowledgements

This work was supported by grants to DARS (NSERC and CIHR-RPP) and was partially funded by FNRS (post-doc grant to CS, PDR T.0170.13). BL thanks the University of Southampton and the European Community (INTERREG IVa, IS:CE-Chem, Project 4061) for support. The structural studies described in this paper were performed at the CLS, which is supported by NSERC, the National Research Council of Canada, the Canadian Institutes of Health Research, the Province of Saskatchewan, Western Economic Diversification Canada and the University of Saskatchewan.

Associated Content

Supporting Information. Detailed description of crystallization conditions for all complexes; figure of mutation site; structure-based sequence alignment; detailed schemes and tables describing interactions between ligands and protein. This material is available free of charge via the Internet at http://pubs.acs.org.

Table 1: Crystallographic Data

1A: Data Collection					
Protein	*Mt*UGM UDP-Gal*p* (non-red)	*Mt*UGM UDP-Gal*p* (re-red)	*Mt*UGM-UDP	*Mt*UGM UDP-F4-Gal*p*	*Mt*UGM UDP-F4-Gal*f*
Beamline	08ID-1	08ID-1	08ID-1	08ID-1	08ID-1
Space group	*C2*	*C2*	*C2*	*C2*	*C2*
Unit cell dimensions	a = 172.4 Å, b =99.5 Å , c =100.2 Å α = 90.0° β =110.8° γ = 90.0°	a =171.2 Å, b =99.5 Å , c =100.0 Å α = 90.0° β =110.1° γ = 90.0°	a =176.5 Å, b =101.2 Å , c =102.0 Å α = 90.0° β =109.2° γ = 90.0°	a =171.3 Å, b =98.3 Å , c =100.5 Å α = 90.0° β =109.9° γ = 90.0°	a =173.6 Å, b =100.4 Å , c =101.3 Å α = 90.0° β =108.7° γ = 90.0°
Resolution range (Å)	50.0-2.40 2.46-2.40	50.0-2.60 2.67-2.60	45.5-2.50 2.59-2.50	50.0-2.25 2.31-2.25	50.0-2.55 2.62-2.55
# reflections	256460	170129	372716	295500	184896
# unique reflections redundancy	61216	48236	58458	74331	53750
% completeness	99.0(98.9)	97.8(98.7)	99.5(100)	99.8(99.9)	99.9(99.8)
Rsym (%)	9.4(76.4)	7.3(82.0)	13.0(63.3)	9.7(87.4)	14.5(82.5)
I/σ(I)	9.7(1.6)	11.3(1.7)	7.2(2.2)	9.4(1.7)	6.0(1.8)
#molecules in ASU	3	3	3	3	3
1B: Refinement					
Resolution range (Å)	50.0-2.40Å	50.0-2.60 Å	45.5-2.50Å	50.0-2.25Å	50.0-2.55Å
Rwork/Rfree (%)	19.5/ 23.7	20.7/ 24.9	18.5/ 21.9	21.2/ 25.4	19.2/ 24.1
No. of amino acid residues	3*392	3*391	3*392	3*391	3*391
No. of solvent atoms	404	110	284	427	319
Ligands	3 * FAD 3* UDP-Gal*p*	3 * FADH 3 * UDP-Gal*p*	3 * FAD 3 * UDP	3 * FAD 3 * UDP-F4Galp	3 * FAD 3 * UDP-F4-Galf
Rmds					
Bond length (Å)	0.003	0.003	0.002	0.002	0.004
Bond angles (°)	0.775	0.590	0.623	0.658	0.729
Ramachandran (%)					
Most favored	98.3	96.6	98.2	97.2	97.4
Additionally allowed	1.7	3.4	1.6	2.6	2.6

Table 2: Kinetic data for *Mt*UGM and other bacterial UGM (K_M measured for UDP-Gal*f*).

	k_{cat} (s^{-1})	K_M μM	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)	
*Mt*UGM(P306R)	7.5±0.5	70±8	0.11±0.02	This work
*Mt*UGM(wildtype)	7.8±.2	45±3	0.17±0.06	This work
*Kp*UGM	5.5±0.7	43±6	0.12±0.02	[72]
*Ec*UGM	27	22	1.22	[95]
*Dr*UGM	66±2.4	55±7	1.18	[51]

Scheme 1 – The UGM catalyzed pyranose/furanose interconversion and structures of fluorinated ligands. Kd's have been measured by STD-NMR with UGM from *K. pneumoniae*.[35,37]

Scheme 3 – Schematic interaction map of UDP-Gal*p* **1** and MtUGM (d < 3Å).

Scheme 4 - Schematic interaction map between UDP-F_4-Gal*p* **3** and *Mt*UGM (contacts < 3Å have been represented).

Scheme 5 - Schematic interaction map between UDP-F_4-Gal*f* **4** and *Mt*UGM (contacts < 3Å have been represented).

Figure 1: Structure of *Mt*UGM complexed with UDP-Gal*p*. **A**) Stereodiagram of the monomer from *Mt*UGM, with numbering of the helices and sheets. The numbers correspond to the labels in Supplemental Figure 1, a structure-based sequence alignment. Domain 1 is colored blue, Domain 2 is coloured green and Domain 3 is colored grey. The mobile loops are colored yellow. FAD and UDP-Gal*p* are shown as ball-and-stick representations. **B)** Ribbon representation of reduced *Mt*UGM:UDP-Gal*p* dimer. Individual subunits are colored green and blue. FADH2 and UDP-Gal*p* are shown in stick representation. **C**) Superposition of unliganded *Mt*UGM (blue) and *Mt*UGM:UDP-Gal*p* complex (green). Open conformation of domain 2 with mobile loops I and II shown in yellow. Closed conformation of domain 2 (red) with mobile loops I and II shown in magenta. FAD and UDP-Gal*p* are shown in stick representation. **D**) Stereodiagram representation of the domain 2 movement of *Mt*UGM as it transitions from unliganded (open) and UDP-Gal*p* bound (closed, black) form as identified by Dyndom. The hinge regions (green) are responsible for movement of the α-helical domain (red) upon ligand binding and active site closure with respect to fixed β-sheet and FAD domains (blue).

Figure 2: **A)** *Mt*UGM with bound UDP-Gal*p*. UDP-Gal*p*, FAD and residues within 4Å are shown as sticks. Feature enhanced maps (FEM), 2mFo-DFm-sigmaA weighted electron density maps for the ligands (contoured at 1σ) is shown as a blue wireframe. **B)** Stereodiagram of unliganded *Mt*UGM (blue) superimposed on *Mt*UGM:UDP-Gal*p* (red, yellow). Moving mobile loop 2 residues labelled according to monomer colour.

Figure 3: Trp166 and Arg261 regulate domain closure.
Superposition of unliganded *Mt*UGM (open, blue) and *Mt*UGM:UDP-Gal*p* complex (closed, red). **A)** Open conformation with Trp166 and Arg261 forming the cation-π interaction. Closed conformation Trp166 side chain rotates ~45° and disrupts the cation-π interaction with Arg261. Arg261 side chain rotates 180° around its Cγ atom away from the substrate binding site.
B) The C-terminus of β14 (Ser317-Phe319) and connecting loop (Ala320-Pro326) move closer to the FAD binding site. The loop movement triggers repositioning of Arg360 towards the active site to prevent a clash with Pro326. In addition, Arg360 helps positioning Tyr366.

Figure 4: Comparison of the active sites from *Mt*UGM:UDP-Gal*p* (blue) and *Dr*UGM:UDP-Gal*p* (red). Labeling of active site residues is according to *Mt*UGM sequence. **A)** Overlay of *Mt*UGM:UDP-Gal*p* (blue) and *Dr*UGM:UDP-Gal*p* (red). UDP-Gal*p* is bound in the identical

position and orientation with its Gal*p* moiety close to the isoalloxazine N5 and makes identical interactions in the sugar-phosphate-ribose binding regions. **B)** Overlay of *Mt*UGM:UDP-Gal*p* (blue) on *Dr*UGM:UDP-Gal*p* (red). The interactions in the uridine binding pocket are not strictly conserved between *Mt*UGM and *Dr*UGM.

Figure 5: Binding mode of inhibitor UDP. **A)** *Mt*UGM with bound UDP. Sugar binding pocket is filled with water molecules (purple spheres) to accommodate for the lack of the sugar moiety. **B)** Superposition of *Mt*UGM with bound UDP (B, blue) on *Mt*UGM:UDP-Gal*p* (B, red). The binding mode of UDP is the same as the binding mode of the UDP moiety in UDP-Gal*p* crystal structure. Ligands and residues within 4Å are shown as sticks. Feature enhanced maps (FEM), 2mFo-DFm-sigmaA weighted electron density maps for the ligands (contoured at 1σ) is shown as a blue wireframe.

Figure 6: Binding mode of inhibitor UDP-F_4-Gal*p*. **A)** *Mt*UGM with bound UDP-F_4-Gal*p*. **B)** Superposition of *Mt*UGM with bound UDP-F_4-Gal*p* (B, blue) on *Mt*UGM:UDP-Gal*p* (B, red). Compared to the unfluorinated Gal*p* moiety the dideoxy-tetrafluorinated Gal*p* moiety is tilted about 0.5 Å at its C2, C3 and C4 carbon atoms. In addition, the exchange of the C2 axial hydrogen atom of Gal*p* by fluorine results in steric repulsion of the isoalloxazine ring of FAD. **C)** *Mt*UGM with bound UDP-F_4-Gal*p* with open conformation of domain 2. **D)** Superposition of *Mt*UGM with bound UDP-F_4-Gal*p* (C, red) on *Mt*UGM:UDP-F_4-Gal*p* (B, blue). The position of Tyr366 and Arg360 with are now in a position similar to unliganded *Mt*UGM. The β-phosphate and F_4-Galp moiety are rotated around the 03 atoms of α-phosphate and β-phosphate respectively. Ligands and residues within 4 Å of the ligands are shown as sticks. In panel **A** & **C**, the Feature Enhanced Map (FEM), 2mFo-DFm-sigmaA weighted electron density map for the ligands (contoured at 1σ) is shown as a blue wireframe.

Figure 7: Binding mode of inhibitor UDP-F_4-Galf. **A)** *Mt*UGM with bound UDP-F_4-Gal*f*. The F_4-Gal*f* moiety is in a similar position and orientation as observed for the Gal*p* moiety and binds at the *re* face of the FAD isoalloxazine with the anomeric carbon atom (C1) positioned 4.3 Å from the FAD N5 atom. Ligands and residues within 4 Å are shown as sticks. The Feature Enhanced Map (FEM), 2mFo-DFm-sigmaA weighted electron density map for the ligands (contoured at 1σ) is shown as a blue wireframe. **B)** Superposition of *Mt*UGM with bound UDP-F_4-Gal*f* (B, blue) on *Mt*UGM:UDP-Gal*p* (B, red). **C**) Superposition of *Mt*UGM with bound UDP-F_4-Gal*f* (B, blue) on *Mt*UGM:UDP-F_4-Gal*p* (B, red). The F_4-Gal*f* moiety is in a similar position and orientation as observed for the Gal*p* moiety

REFERENCES

(1) Gandhi, N. R.; Shah, N. S.; Andrews, J. R.; Vella, V.; Moll, A. P.; Scott, M.; Weissman, D.; Marra, C.; Lalloo, U. G.; Friedland, G. H.; Tugela Ferry, C.; Research, C. *Am. J. Resp. Crit. Care Med.* **2010**, *181*, 80.

(2) Gandhi, N. R.; Nunn, P.; Dheda, K.; Schaaf, H. S.; Zignol, M.; van Soolingen, D.; Jensen, P.; Bayona, J. *Lancet* **2010**, *375*, 1830.

(3) *World Health Organization (WHO), Geneva* **2013**.

(4) Marris, E. *Nature* **2006**, *443*, 131.

(5) Young, D. B.; Perkins, M. D.; Duncan, K.; Barry, C. E., 3rd *The J. Clin. Inves.* **2008**, *118*, 1255.

(6) Chan, E. D.; Iseman, M. D. *Curr. Op. Infect. Dis.* **2008**, *21*, 587.

(7) Besra, G. S.; Khoo, K. H.; McNeil, M. R.; Dell, A.; Morris, H. R.; Brennan, P. J. *Biochem.* **1995**, *34*, 4257.

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

(9) Delederkremer, R. M.; Colli, W. *Glycobiol.* **1995**, *5*, 547.

(10) Lamarre, C.; Beau, R.; Balloy, V.; Fontaine, T.; Hoi, J. W. S.; Guadagnini, S.; Berkova, N.; Chignard, M.; Beauvais, A.; Latge, J. P. *Cell. Micro.* **2009**, *11*, 1612.

(11) Weston, A.; Stern, R. J.; Lee, R. E.; Nassau, P. M.; Monsey, D.; Martin, S. L.; Scherman, M. S.; Besra, G. S.; Duncan, K.; McNeil, M. R. *Tubercle and Lung Disease* **1998**, *78*, 123.

(12) Nassau, P. M.; Martin, S. L.; Brown, R. E.; Weston, A.; Monsey, D.; McNeil, M. R.; Duncan, K. *J. Bact.* **1996**, *178*, 1047.

(13) Beverley, S. M.; Owens, K. L.; Showalter, M.; Griffith, C. L.; Doering, T. L.; Jones, V. C.; McNeil, M. R. *Euk. Cell* **2005**, *4*, 1147.

(14) Tefsen, B.; Ram, A. F. J.; van Die, I.; Routier, F. H. *Glycobiol.* **2012**, *22*, 456.

(15) Pan, F.; Jackson, M.; Ma, Y. F.; McNeil, M. *J.f Bact.* **2001**, *183*, 6971.

(16) Pedersen, L. L.; Turco, S. J. *Cell. Mol. Life Sci.* **2003**, *60*, 259.

(17) Dykhuizen, E. C.; Kiessling, L. L. *Org. Lett.* **2009**, *11*, 193.

(18) Dykhuizen, E. C.; May, J. F.; Tongpenyai, A.; Kiessling, L. L. *J. Am. Chem. Soc.* **2008**, *130*, 6706.

(19) Sadeghi-Khomami, A.; Forcada, T. J.; Wilson, C.; Sanders, D. A. R.; Thomas, N. R. *Org. Biomol. Chem.* **2010**, *8*, 1596.

(20) Partha, S. K.; Sadeghi-Khomami, A.; Cren, S.; Robinson, R. I.; Woodward, S.; Slowski, K.; Berast, L.; Zheng, B.; Thomas, N. R.; Sanders, D. A. R. *Mol. Info.* **2011**, *30*, 873.

(21) Carlson, E. E.; May, J. F.; Kiessling, L. L. *Chem. Biol.* **2006**, *13*, 825.

(22) Borrelli, S.; Zandberg, W. F.; Mohan, S.; Ko, M.; Martinez-Gutierrez, F.; Partha, S. K.; Sanders, D. A. R.; Av-Gay, Y.; Pinto, B. M. *Int. J. Antimicro. Ag.* **2010**, *36*, 364.

(23) Itoh, K.; Huang, Z. S.; Liu, H. W. *Org. Lett.* **2007**, *9*, 879.

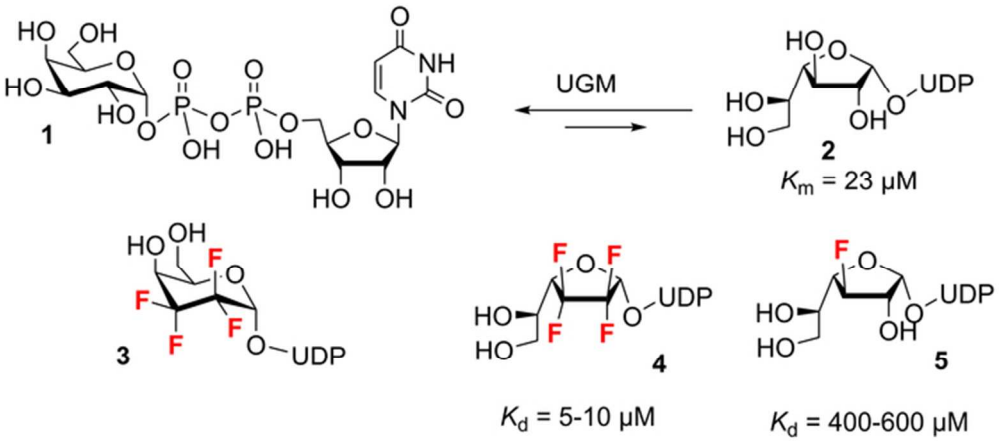
(24) Veerapen, N.; Yuan, Y.; Sanders, D. A. R.; Pinto, B. M. *Carb. Res.* **2004**, *339*, 2205.

(25) El Bkassiny, S.; N'Go, I.; Sevrain, C. M.; Tikad, A.; Vincent, S. P. *Org. Lett.* **2014**, *16*, 2462.

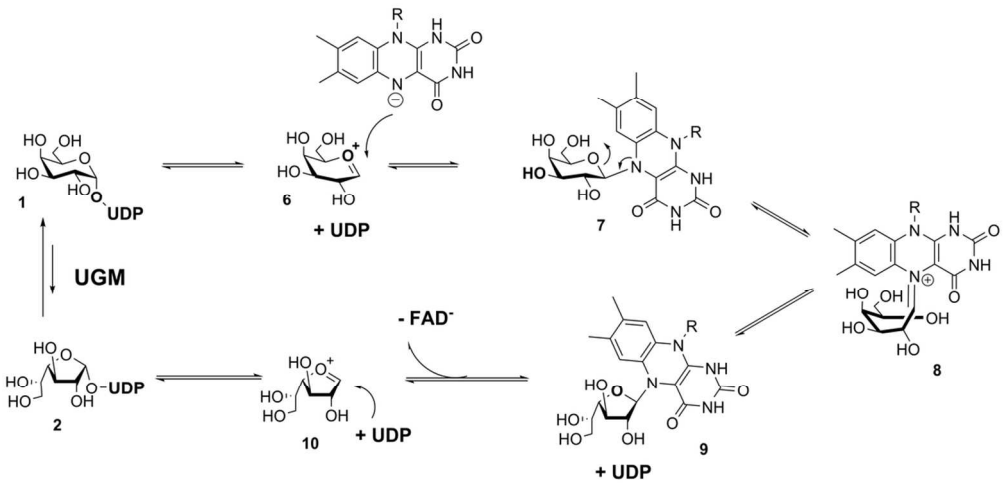
(26) Sadeghi-Khomami, A.; Blake, A. J.; Wilson, C.; Thomas, N. R. *Org. Lett.* **2005**, *7*, 4891.
(27) Pan, W. D.; Ansiaux, C.; Vincent, S. P. *Tet. Lett.* **2007**, *48*, 4353.
(28) Caravano, A.; Dohi, H.; Sinay, P.; Vincent, S. P. *Chem. Eur. J.* **2006**, *12*, 3114.
(29) Caravano, A.; Mengin-Lecreulx, D.; Brondello, J. M.; Vincent, S. P.; Sinay, P. *Chem.* **2003**, *9*, 5888.
(30) Scherman, M. S.; Winans, K. A.; Stern, R. J.; Jones, V.; Bertozzi, C. R.; McNeil, M. R. *Antimicro. Ag. Chemo.* **2003**, *47*, 378.
(31) Caravano, A.; Vincent, S. P. *Eur. J. Org. Chem.* **2009**, 1771.
(32) Barlow, J. N.; Blanchard, J. S. *Carb. Res.* **2000**, *328*, 473.
(33) Zhang, Q. B.; Liu, H. W. *J. Am. Chem. Soc.* **2001**, *123*, 6756.
(34) Errey, J. C.; Mann, M. C.; Fairhurst, S. A.; Hill, L.; McNeil, M. R.; Naismith, J. H.; Percy, J. M.; Whitfield, C.; Field, R. A. *Org. Bio. Chem.* **2009**, *7*, 1009.
(35) Yuan, Y.; Bleile, D. W.; Wen, X.; Sanders, D. A. R.; Itoh, K.; Liu, H. W.; Pinto, B. M. *J. Am. Chem. Soc.* **2008**, *130*, 3157.
(36) Eppe, G.; Peltier, P.; Daniellou, R.; Nugier-Chauvin, C.; Ferrieres, V.; Vincent, S. P. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 814.
(37) N'Go, I.; Golten, S.; Arda, A.; Canada, J.; Jimenez-Barbero, J.; Linclau, B.; Vincent, S. P. *Chem.* **2014**, *20*, 106.
(38) Biffinger, J. C.; Kim, H. W.; DiMagno, S. G. *Chembiochem* **2004**, *5*, 622.
(39) Kim, W. K.; Rossi, P.; Shoemaker, R. K.; DiMagno, S. G. *J. Am. Chem. Soc.* **1998**, *120*, 9082.
(40) Gruber, T. D.; Borrok, M. J.; Westler, W. M.; Forest, K. T.; Kiessling, L. L. *J. Mol. Biol.* **2009**, *391*, 327.
(41) Gruber, T. D.; Westler, W. M.; Kiessling, L. L.; Forest, K. T. *Biochem.* **2009**, *48*, 9171.
(42) Oppenheimer, M.; Poulin, M. B.; Lowary, T. L.; Helm, R. F.; Sobrado, P. *Arch. Biochem. Biophys.* **2010**, *502*, 31.
(43) Oppenheimer, M.; Valenciano, A. L.; Kizjakina, K.; Qi, J.; Sobrado, P. *PloS One* **2012**, *7*, e32918.
(44) Huang, W.; Gauld, J. W. *J. Phys. Chem. B* **2012**, *116*, 14040.
(45) Sun, H. G.; Ruszczycky, M. W.; Chang, W.-C.; Thibodeaux, C. J.; Liu, H.-W. *J. Biol. Chem.* **2012**, *287*, 4602.
(46) Beis, K.; Srikannathasan, V.; Liu, H.; Fullerton, S. W. B.; Bamford, V. A.; Sanders, D. A. R.; Whitfield, C.; McNeil, M. R.; Naismith, J. H. *J. Mol. Biol.* **2005**, *348*, 971.
(47) Fullerton, S. W. B.; Daff, S.; Sanders, D. A. R.; Ingledew, W. J.; Whitfield, C.; Chapman, S. K.; Naismith, J. H. *Biochem.* **2003**, *42*, 2104.
(48) Huang, Z. H.; Zhang, Q. B.; Liu, H. W. *Bioorg. Chem.* **2003**, *31*, 494.
(49) Yuan, Y.; Wen, X.; Sanders, D. A. R.; Pinto, B. M. *Biochem.* **2005**, *44*, 14080.
(50) Sanders, D. A. R.; Staines, A. G.; McMahon, S. A.; McNeil, M. R.; Whitfield, C.; Naismith, J. H. *Nat. Struct. Biol.* **2001**, *8*, 858.
(51) Partha, S. K.; van Straaten, K. E.; Sanders, D. A. R. *J. Mol. Biol.* **2009**, *394*, 864.
(52) Partha, S. K.; Sadeghi-Khomami, A.; Slowski, K.; Kotake, T.; Thomas, N. R.; Jakeman, D. L.; Sanders, D. A. R. *J. Mol. Biol.* **2010**, *403*, 578.

(53) van Straaten, K. E.; Routier, F. H.; Sanders, D. A. R. *J. Biol. Chem.* **2012**, *287*, 10780.
(54) Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Sobrado, P.; Tanner, J. J. *Biochem.* **2012**, *51*, 4968.
(55) Tanner, J. J.; Boechi, L.; Andrew McCammon, J.; Sobrado, P. *Arch. Biochem. Biophys.* **2014**, *544*, 128.
(56) Nallamsetty, S.; Waugh, D. S. *Nat. Protoc.* **2007**, *2*, 383.
(57) Kapust, R. B.; Tozser, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S. *Prot. Eng.* **2001**, *14*, 993.
(58) Fodje, M. N.; Berg, R.; Black, G.; Grochulski, P.; Janzen, K. In *Proc. PCaPAC 2010* Saskatoon, Saskatchewan, 2010, p 130.
(59) Pflugrath, J. W. *Acta Cryst. D.* **1999**, *55*, 1718.
(60) Vagin, A.; Teplyakov, A. *J. Appl. Cryst.* **1997**, *30*, 1022.
(61) Bailey, S. *Acta Cryst. D.* **1994**, *50*, 760.
(62) 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. Cryst. D.* **2010**, *66*, 213.
(63) Emsley, P.; Cowtan, K. *Acta Cryst. D.* **2004**, *60*, 2126.
(64) Davis, I. W.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Nucl. Acids Res.* **2004**, *32*, W615.
(65) Holm, L.; Sander, C. *Tr. Biochem. Sci.* **1995**, *20*, 478.
(66) Collaborative Computational Project, N. *Acta Cryst. D.* **1994**, *50*, 760.
(67) Hayward, S.; Berendsen, H. J. C. *Prot.* **1998**, *30*, 144.
(68) Sayle, R. A.; Milnerwhite, E. J. *Tr. Biochem. Sci.* **1995**, *20*, 374.
(69) Gouet, P.; Courcelle, E.; Stuart, D. I.; Metoz, F. *Bioinf.* **1999**, *15*, 305.
(70) Nallamsetty, S.; Austin, B. P.; Penrose, K. J.; Waugh, D. S. *Prot. Sci.* **2005**, *14*, 2964.
(71) Hayward, S.; Lee, R. A. *J. Mol. Graph. Model.* **2002**, *21*, 181.
(72) Chad, J. M.; Sarathy, K. P.; Gruber, T. D.; Addala, E.; Kiessling, L. L.; Sanders, D. A. R. *Biochem.* **2007**, *46*, 6723.
(73) Linclau, B.; Golten, S.; Light, M.; Sebban, M.; Oulyadi, H. *Carbo. Res.* **2011**, *346*, 1129.
(74) Dalvit, C.; Invernizzi, C.; Vulpetti, A. *Chem.* **2014**, *20*, 11058.
(75) Giuffredi, G. T.; Gouverneur, V.; Bernet, B. *Angew. Chem. Int. Ed.* **2013**, *52*, 10524.
(76) Schneider, H.-J. *Chem. Sci.* **2012**, *3*, 1381.
(77) Muller, K. *Chimia* **2014**, *68*, 356.
(78) Howard, J. A. K.; Hoy, V. J.; O'Hagan, D.; Smith, G. T. *Tetrahedron* **1996**, *52*, 12613.
(79) Dunitz, J. D.; Taylor, R. *Chem. Euro. J.* **1997**, *3*, 89.
(80) Garnett, J. A.; Liu, Y.; Leon, E.; Allman, S. A.; Friedrich, N.; Saouros, S.; Curry, S.; Soldati-Favre, D.; Davis, B. G.; Feizi, T.; Matthews, S. *Prot. Sci.* **2009**, *18*, 1935.
(81) Vermersch, P. S.; Tesmer, J. J.; Quiocho, F. A. *J. Mol. Biol.* **1992**, *226*, 923.
(82) Paulini, R.; Muller, K.; Diederich, F. *Angew. Chem.* **2005**, *44*, 1788.

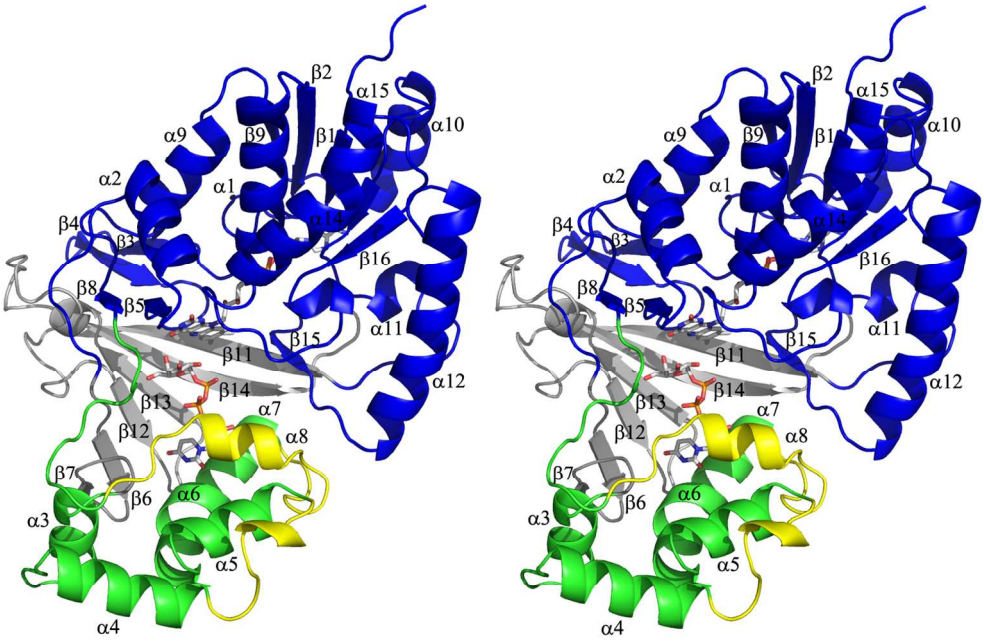
(83) Müller, K.; Faeh, C.; Diederich, F. *Science* **2007**, *317*, 1881.
(84) Zhang, Q. B.; Liu, H. W. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 145.
(85) Dalvit, C.; Vulpetti, A. *ChemMedChem* **2011**, *6*, 104.
(86) Lockett, M. R.; Lange, H.; Breiten, B.; Heroux, A.; Sherman, W.; Rappoport, D.; Yau, P. O.; Snyder, P. W.; Whitesides, G. M. *Angew. Chem.* **2013**, *52*, 7714.
(87) Mecinovic, J.; Snyder, P. W.; Mirica, K. A.; Bai, S.; Mack, E. T.; Kwant, R. L.; Moustakas, D. T.; Heroux, A.; Whitesides, G. M. *J. Am. Chem. Soc.* **2011**, *133*, 14017.
(88) Ioannou, A.; Cini, E.; Timofte, R. S.; Flitsch, S. L.; Turner, N. J.; Linclau, B. *Chem. Comm.* **2011**, *47*, 11228.
(89) Chiu, H. P.; Kokona, B.; Fairman, R.; Cheng, R. P. *J. Am. Chem. Soc.* **2009**, *131*, 13192.
(90) Zheng, H.; Comeforo, K.; Gao, J. *J. Am. Chem. Soc.* **2009**, *131*, 18.
(91) Molski, M. A.; Goodman, J. L.; Craig, C. J.; Meng, H.; Kumar, K.; Schepartz, A. *J. Am. Chem. Soc.* **2010**, *132*, 3658.
(92) Hof, F.; Scofield, D. M.; Schweizer, W. B.; Diederich, F. *Angew. Chem.* **2004**, *43*, 5056.
(93) Chiang, C. H.; Ramu, R.; Tu, Y. J.; Yang, C. L.; Ng, K. Y.; Luo, W. I.; Chen, C. H.; Lu, Y. Y.; Liu, C. L.; Yu, S. S. *Chem.* **2013**, *19*, 13680.
(94) Xu, B.; Unione, L.; Sardinha, J.; Wu, S.; Etheve-Quelquejeu, M.; Pilar Rauter, A.; Bleriot, Y.; Zhang, Y.; Martin-Santamaria, S.; Diaz, D.; Jimenez-Barbero, J.; Sollogoub, M. *Angew. Chem.* **2014**, *53*, 9597.
(95) Zhang, Q. B.; Liu, H. W. *J. Am. Chem. Soc.* **2000**, *122*, 9065.



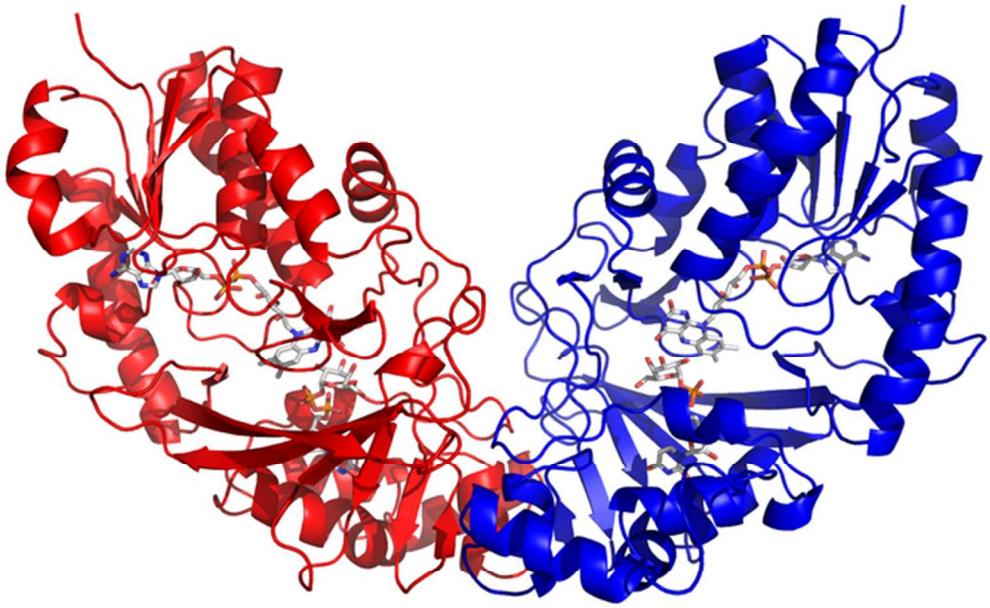
60x27mm (300 x 300 DPI)



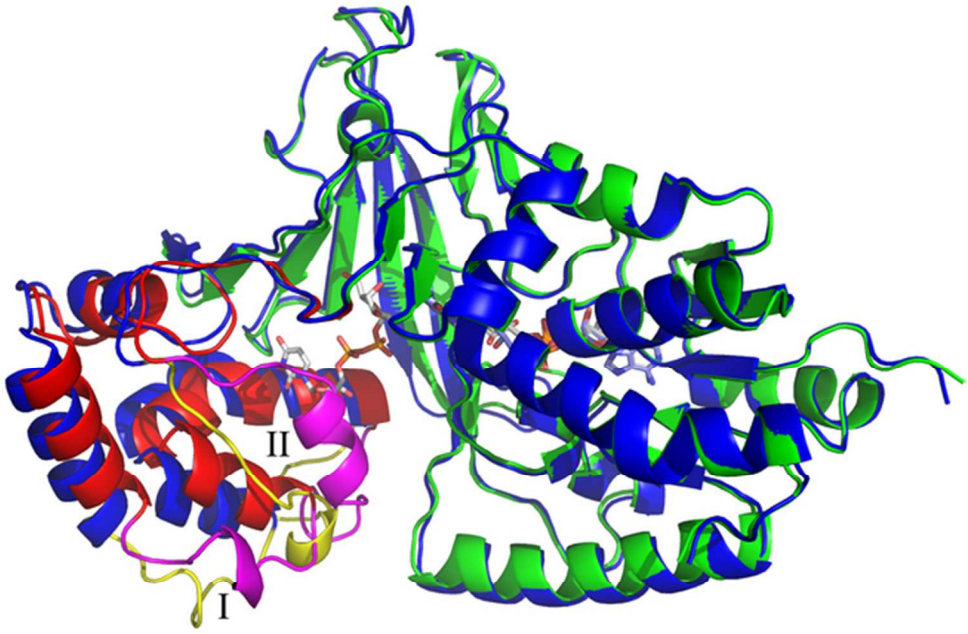
Scheme 2
106x50mm (300 x 300 DPI)



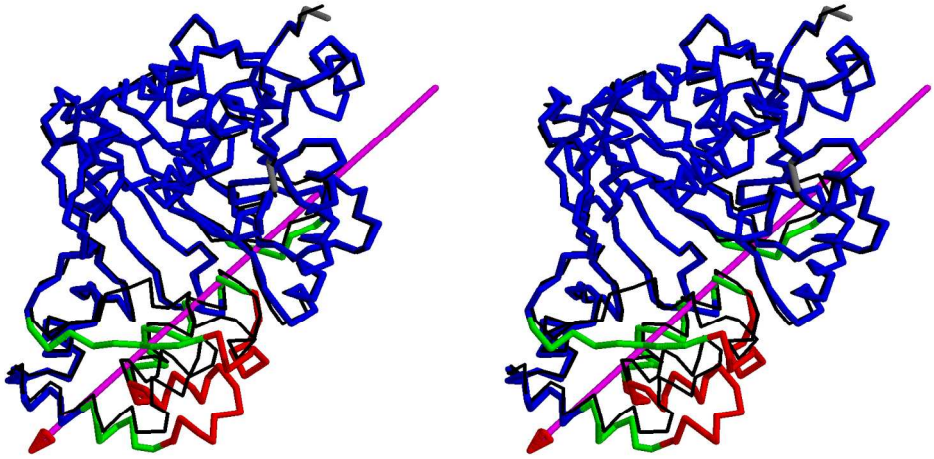
155x116mm (262 x 262 DPI)



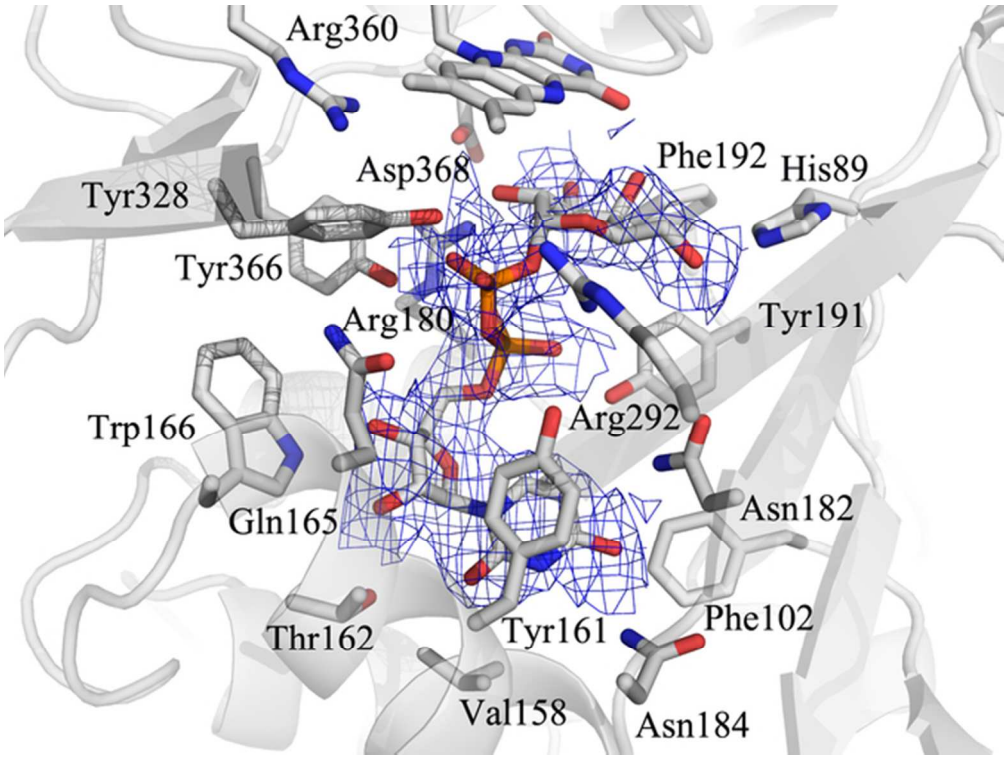
57x42mm (300 x 300 DPI)



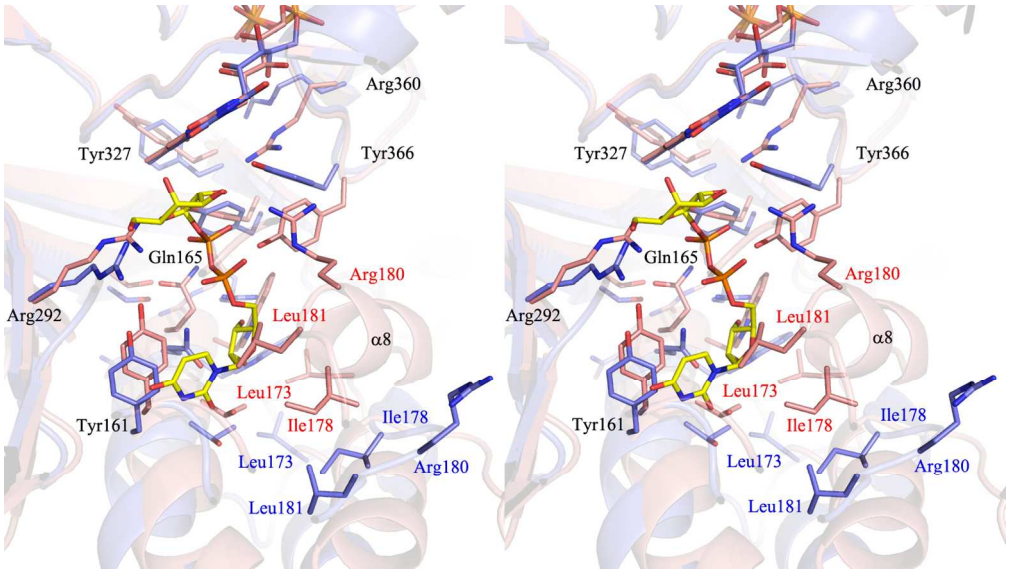
57x42mm (300 x 300 DPI)



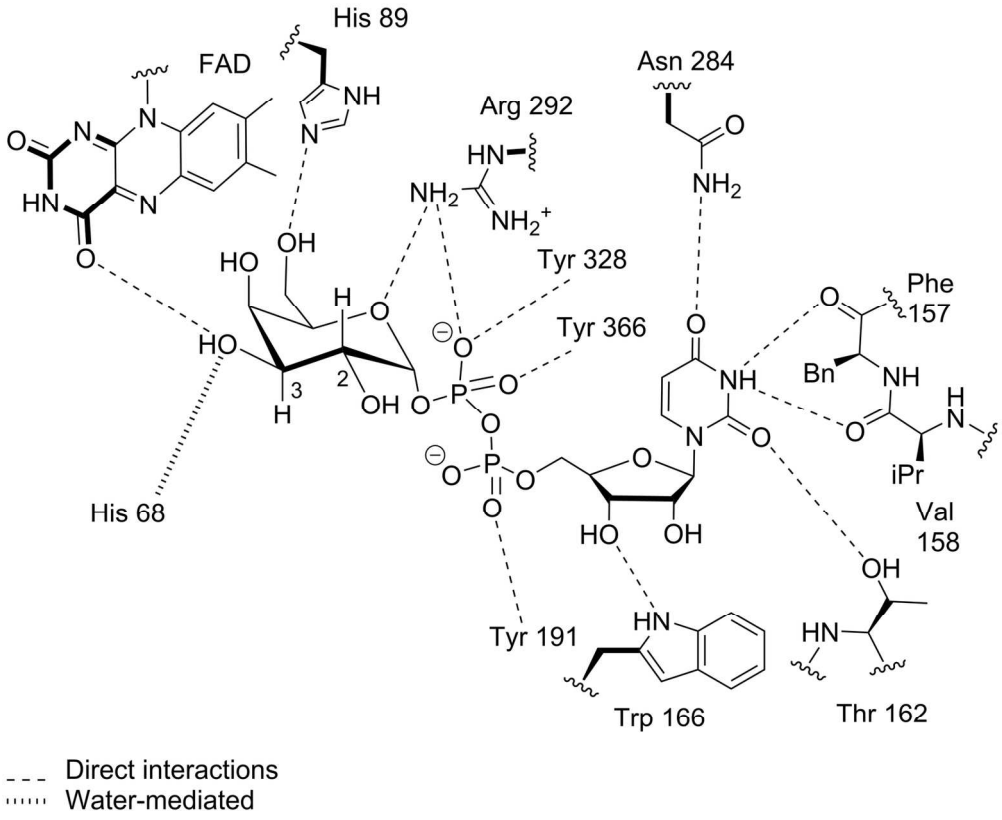
677x372mm (72 x 72 DPI)



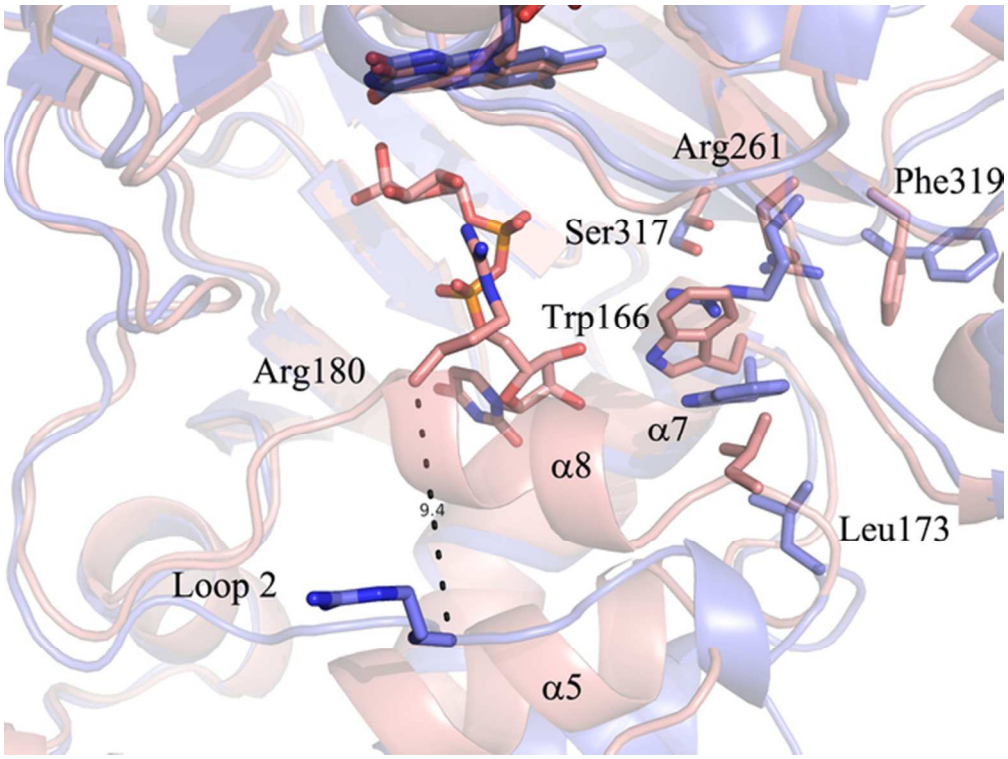
57x42mm (300 x 300 DPI)



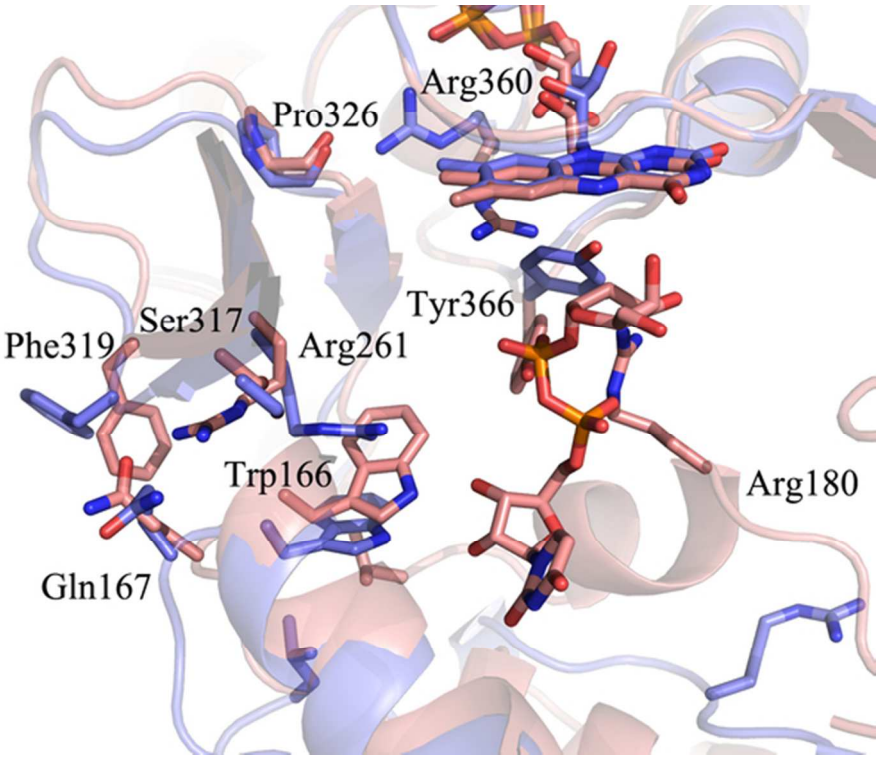
156x87mm (260 x 260 DPI)



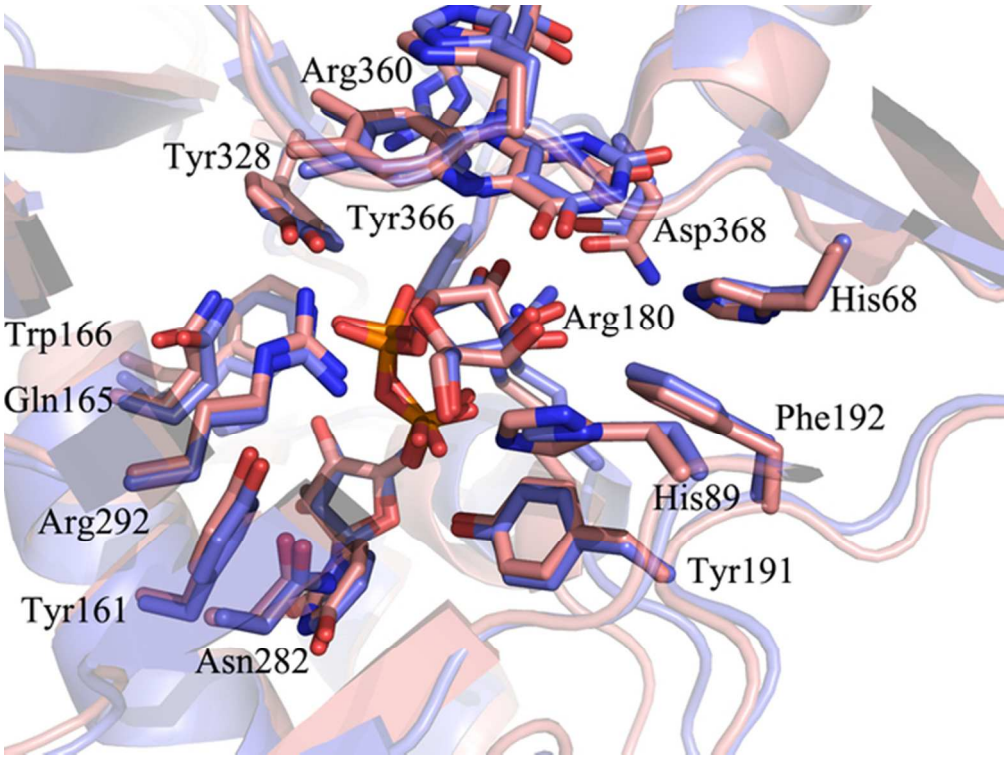
Scheme 3
145x119mm (300 x 300 DPI)



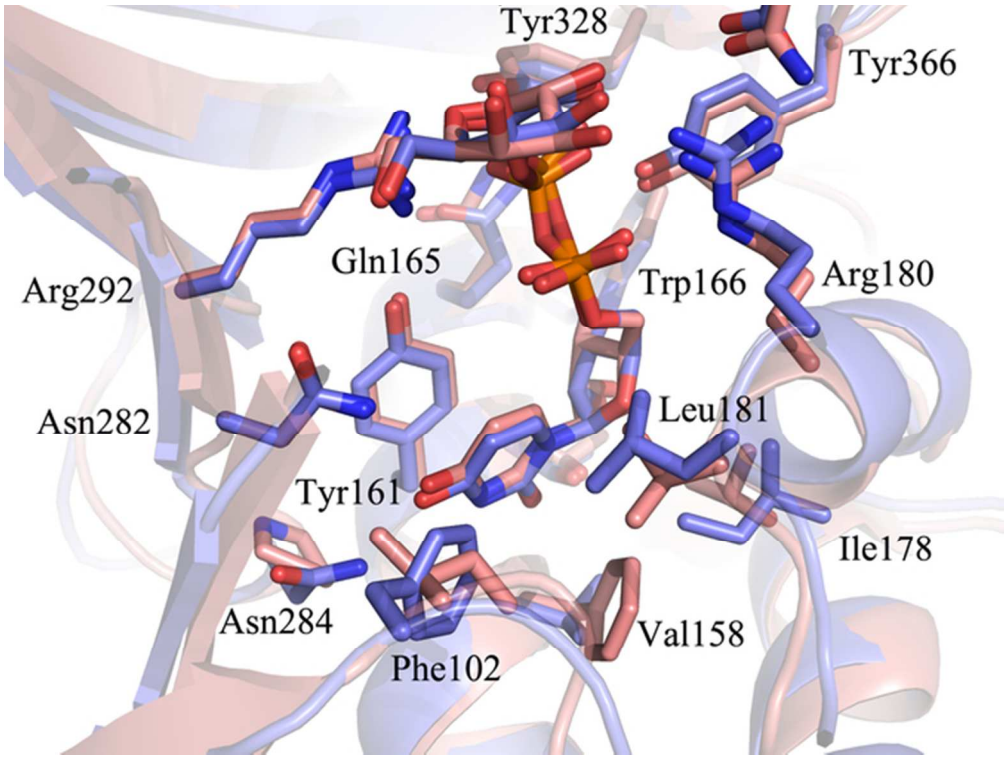
57x42mm (300 x 300 DPI)



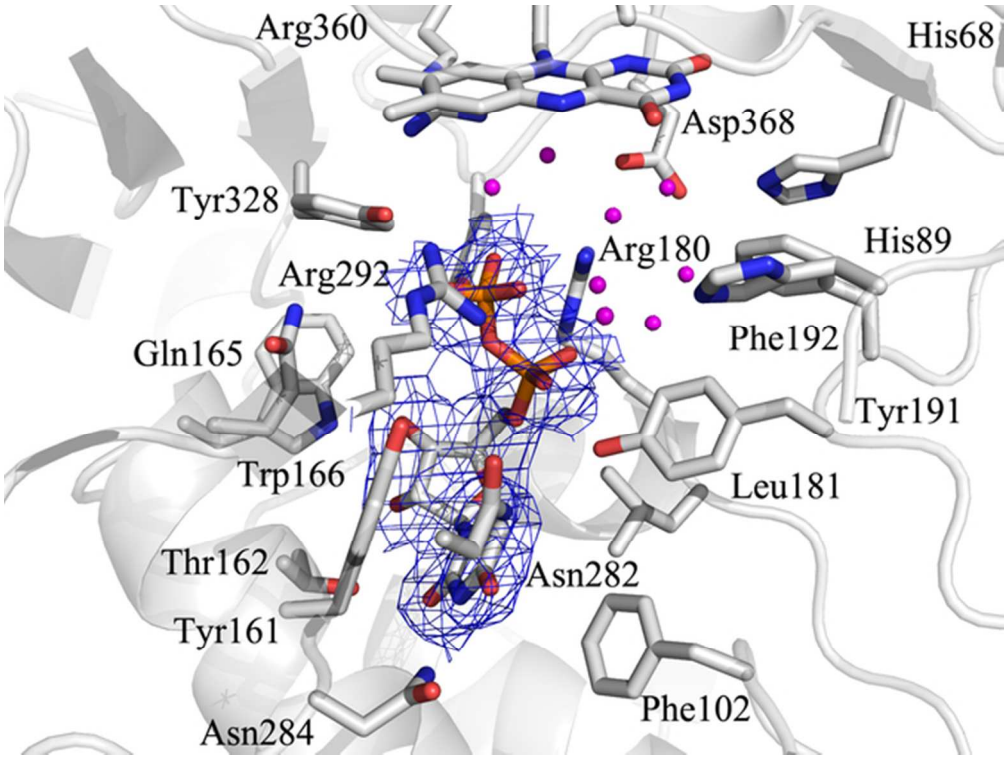
57x42mm (300 x 300 DPI)



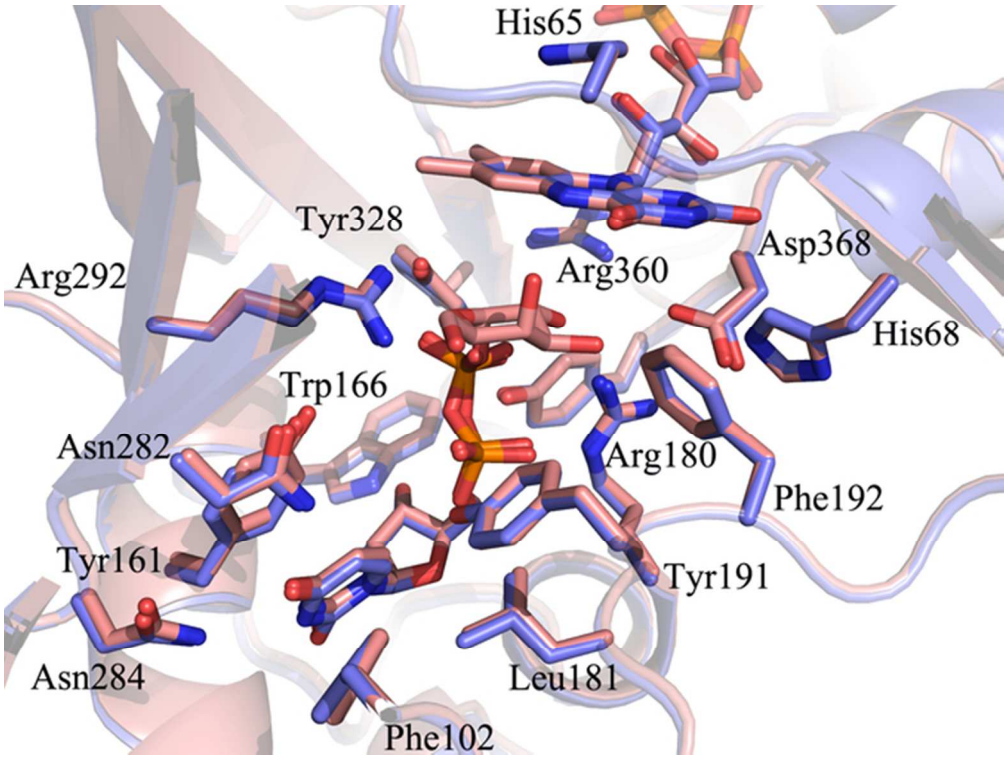
57x42mm (300 x 300 DPI)



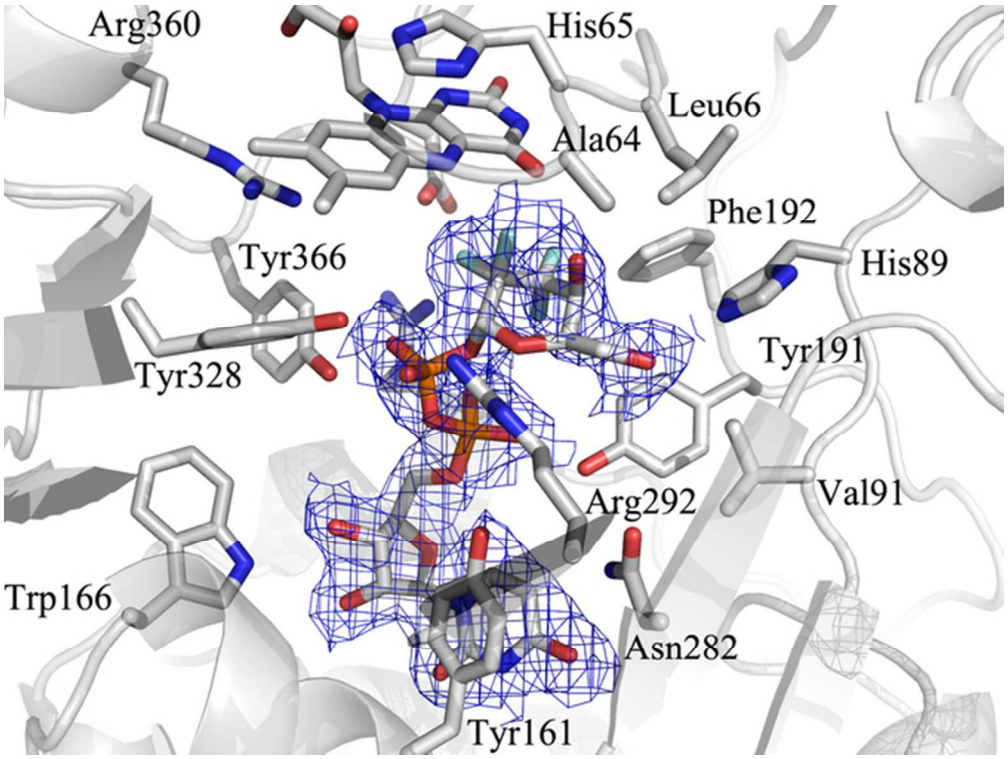
57x42mm (300 x 300 DPI)



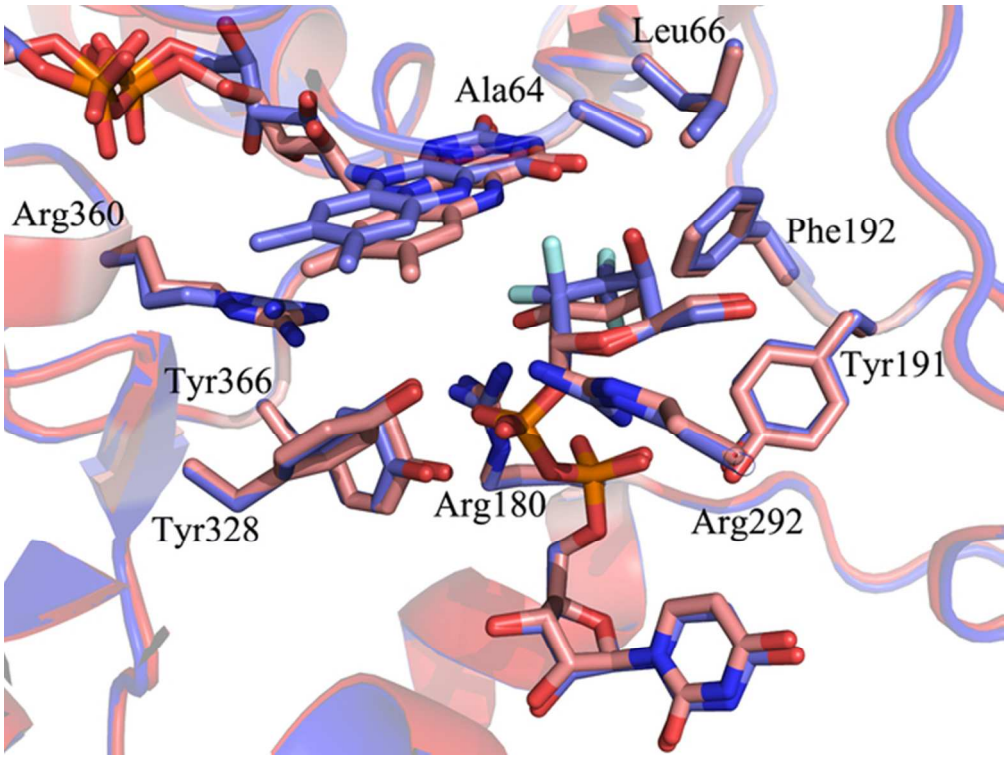
57x42mm (300 x 300 DPI)



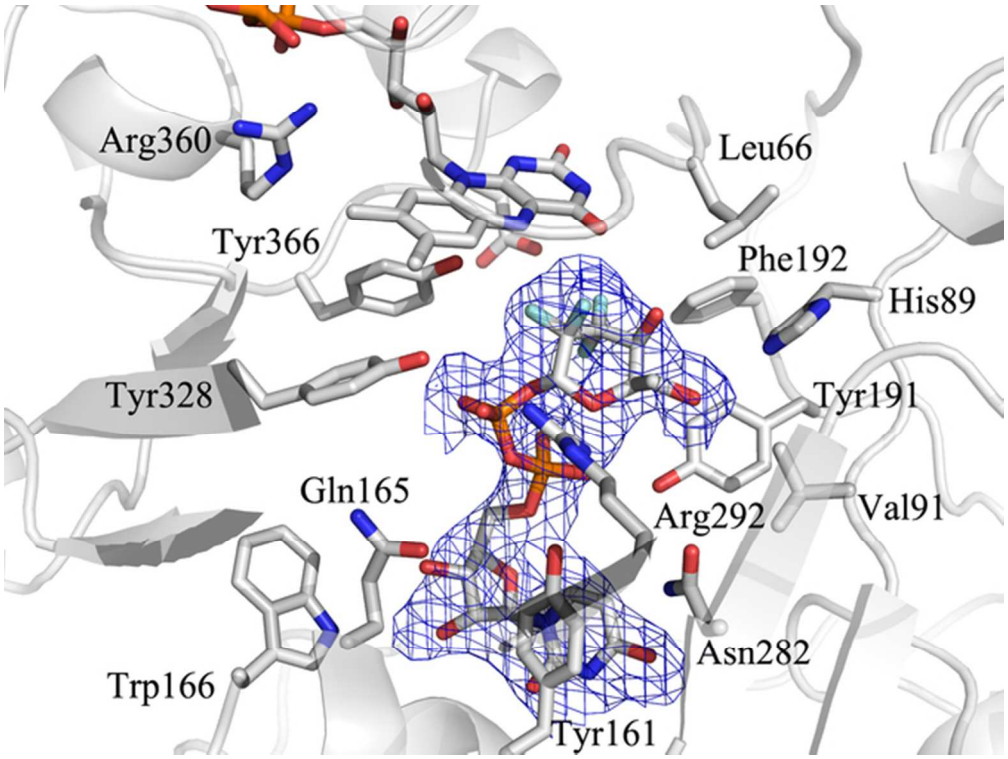
57x42mm (300 x 300 DPI)



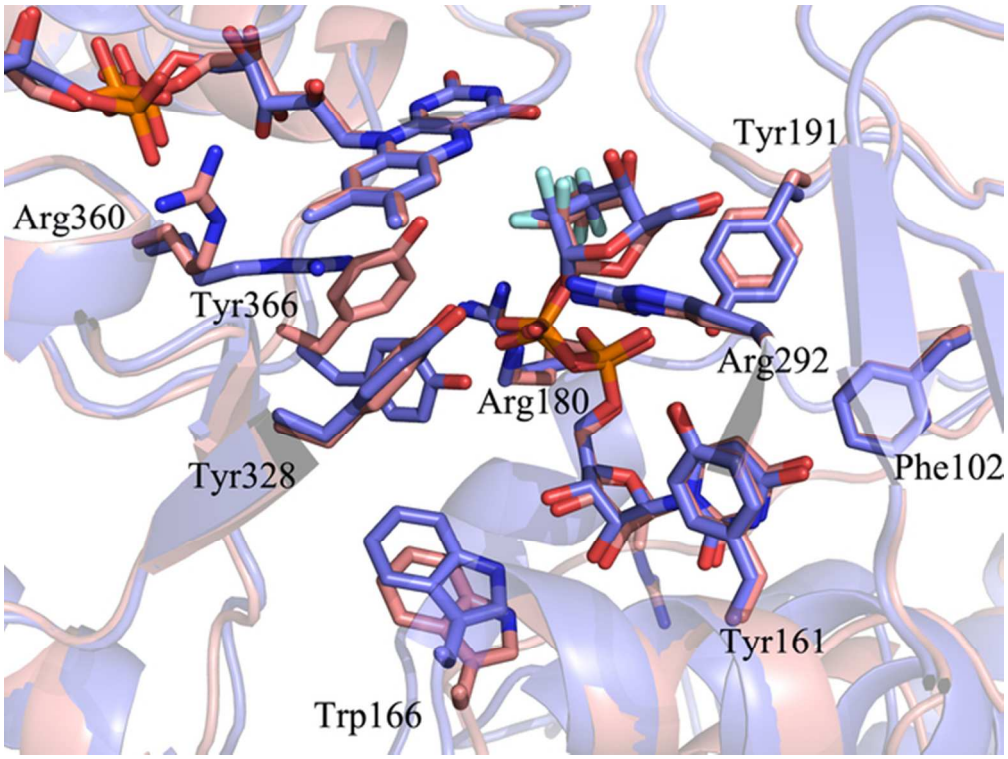
57x42mm (300 x 300 DPI)



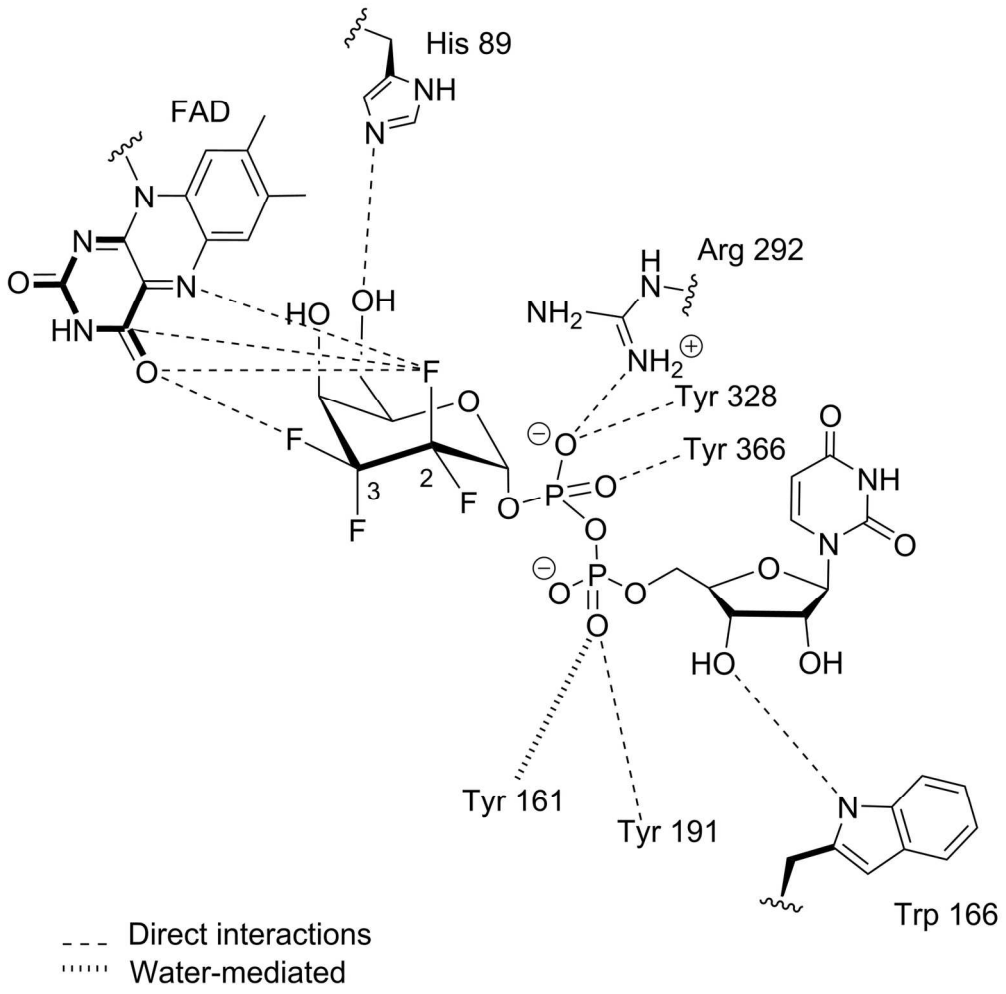
57x42mm (300 x 300 DPI)



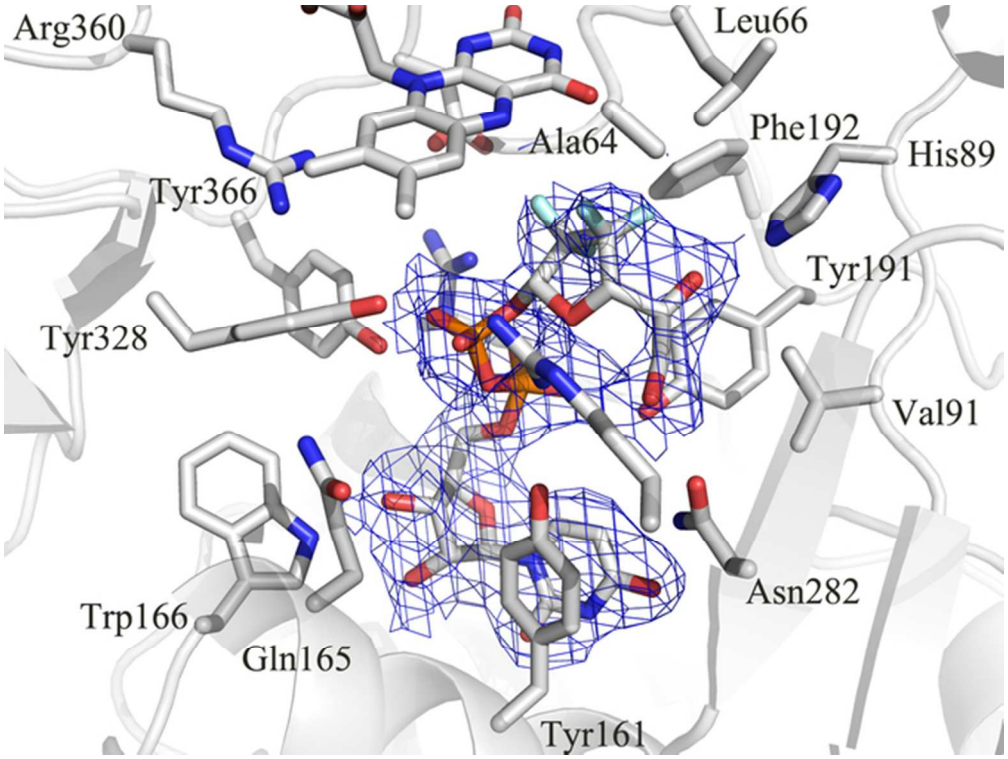
57x42mm (300 x 300 DPI)



57x42mm (300 x 300 DPI)



153x152mm (300 x 300 DPI)



57x42mm (300 x 300 DPI)

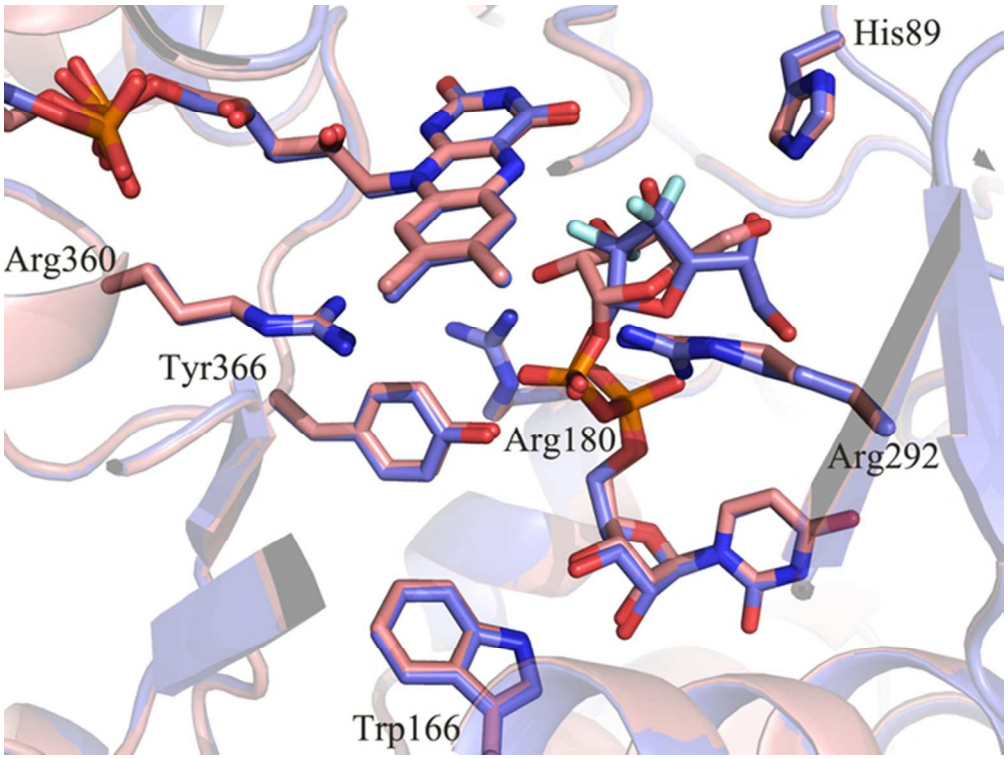
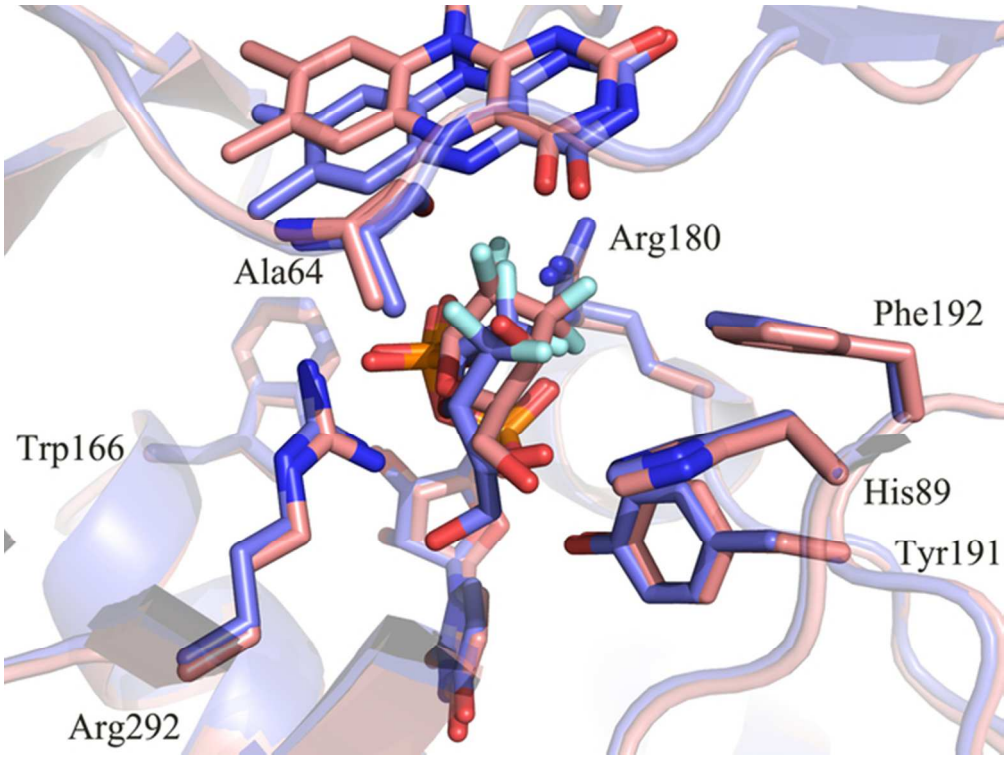
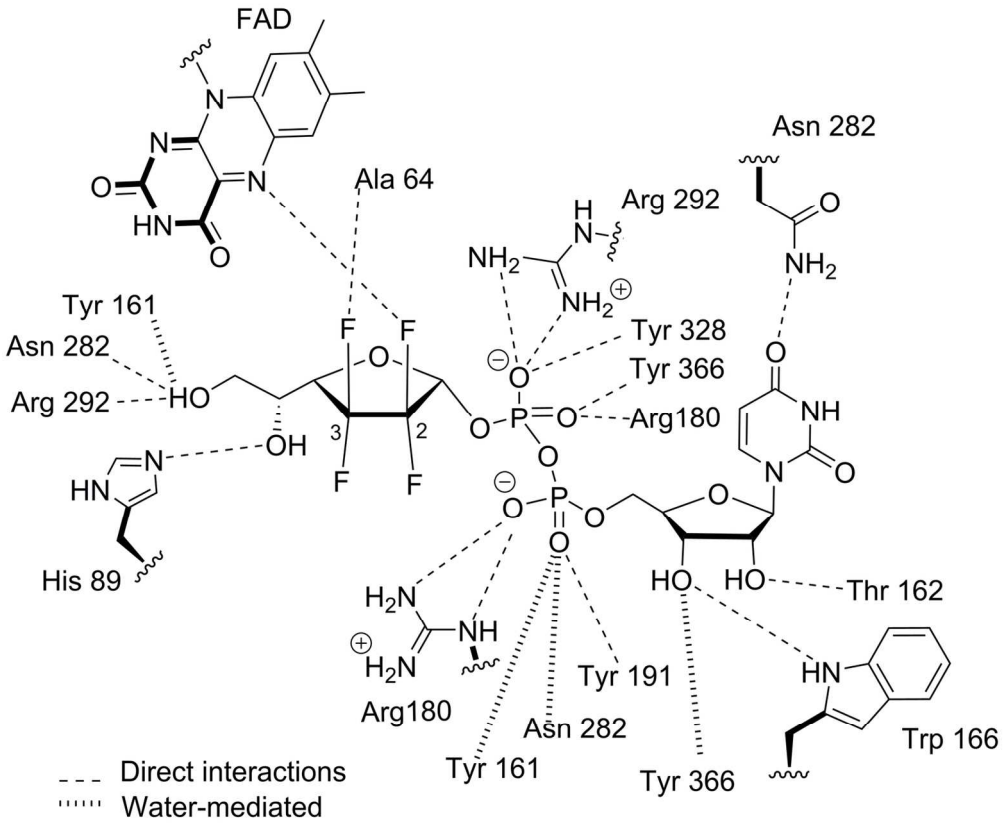


Figure 7B
57x42mm (300 x 300 DPI)



57x42mm (300 x 300 DPI)



Scheme 5
138x114mm (300 x 300 DPI)