Structure, mechanism, and dynamics of UDP-galactopyranose mutase

John J. Tanner, Leonardo Boechi, J. Andrew McCammon, Pablo Sobrado

ABSTRACT

The flavoenzyme UDP-galactopyranose mutase (UGM) is a key enzyme in galactofuranose biosynthesis. The enzyme catalyzes the 6-to-5 ring contraction of UDP-galactopyranose to UDP-galactofuranose. Galactofuranose is absent in humans yet is an essential component of bacterial and fungal cell walls and a cell surface virulence factor in protozoan parasites. Thus, inhibition of galactofuranose biosynthesis is a valid strategy for developing new antimicrobials. UGM is an excellent target in this effort because the product of the UGM reaction represents the first appearance of galactofuranose in the biosynthetic pathway. The UGM reaction is redox neutral, which is atypical for flavoenzymes, motivating intense examination of the chemical mechanism and structural features that tune the flavin for its unique role in catalysis. These studies show that the flavin functions as nucleophile, forming a flavin–sugar adduct that facilitates galactose-ring opening and contraction. The 3-dimensional fold is novel and conserved among all UGMs. However, the larger eukaryotic enzymes have additional secondary structure elements that lead to significant differences in quaternary structure, substrate conformation, and conformational flexibility. Here we present a comprehensive review of UGM three-dimensional structure, provide an update on recent developments in understanding the mechanism of the enzyme, and summarize computational studies of active site flexibility.

Introduction

Galactofuranose (GalF) is the five-membered ring form of the more common 6-membered ring sugar galactopyranose (Galp). GalF is thermodynamically less stable than Galp because of the strain associated with the 5-membered ring. Nevertheless, it has been known for nearly a century that microorganisms produce GalF in several forms. For example, galactocallose, an extracellular β-β(1→5)-linked polygalactofuranose produced by Penicillium charlesii, was the first polysaccharide shown to contain GalF [1], motivating interest in the underlying biosynthetic pathway. Decades later, it was shown that P. charlesii could not use exogenous galactose to produce galactocallose [2], leading to the discovery of a new nucleotide, UDP-GalF, as the precursor in galactocallose biosynthesis [3]. Similarly, investigations into the origins of GalF in the T1 antigen of Salmonella typhimurium identified UDP-Galp in the biosynthetic pathway and suggested the existence of an enzyme that catalyzes the 6-to-5 ring contraction of UDP-Galp to UDP-GalF [4,5]. Interest in GalF biosynthesis stems in part from the observation that this sugar is not present in mammals; however, it is a major component of cell wall and cell surface glycoconjugates in many bacteria and eukaryotic organisms, including the human pathogens Mycobacteria tuberculosis, Klebsiella pneumoniae, Trypanosoma cruzi, Leishmania major, and Aspergillus fumigatus [6–10]. Targeting cell wall biosynthesis is an effective and well-established method for combating bacterial infections. Since GalF is absent in humans, the enzymes involved in the biosynthesis of GalF are potential drug targets. At the center of GalF biosynthesis is the enzyme UDP-galactopyranose mutase (UGM). The gene encoding for a UGM enzyme was first cloned from...
and C1Gal. In this process, the flavin also functions as a moiety to an oxocarbenium galactose intermediate, which must be deprotonated. Although this aspect of the mechanism is enigmatic (Scheme 1A). Several mechanistic routes for the intermediary formation of the flavin iminium ion is very fast (0.003 s⁻¹ for TcUGM). In addition, viscosity effects broken during catalysis using positional isotope effects (PIX) [23], which was later supported by work from Liu’s group [18]. Cleavage of the anomeric bond suggested a number of possible mechanisms for UGMs (Scheme 2). One mechanism predicts the formation of a 1,4-anhydrogalactopyranose (Scheme 2A) [18,23]. However, activity was not detected when reduced UGM was incubated with 1,4-anhydrogalactopyranose in the presence of UDP, eliminating this species from consideration as an intermediate in the UGM reaction [24]. A mechanism involving a single-electron transfer step during catalysis was supported by potentiometric studies that showed that the flavin semiquinone was stabilized in the presence of substrate [25]. In addition, replacement of the FAD with 5-deaza-FAD, a flavin analog restricted to a net 2-electron process, resulted in inactive UGM. The lack of activity of UGM reconstituted with 5-deaza-FAD was initially interpreted as supporting an electron transfer step in the catalytic cycle (Scheme 2B) [26].

A major breakthrough in our understanding of the mechanism of action of UGM came when Kiessling’s group isolated an FAD-galactose covalent intermediate [27]. The covalent intermediate formed between the N5FAD and C1galp was proposed and later validated by isolation and characterization by mass spectrometry and NMR in bacterial UGM (bUGM) and by UV/vis spectrophotometry and mass spectrometry in eukaryotic UGM (eUGM) [28–30]. The flavin–sugar adduct is important in ring opening and activation of the C1galp [27]. In this process, the flavin also functions as a molecular scaffold, providing the structural constraints required for ring contraction (Scheme 2B) [31]. Formation of this intermediate made mechanistic sense and was consistent with the breaking of the anomeric bond determined from PIX studies and the lack of activity observed with the enzyme reconstituted with 5-deaza-FAD.

Although identification of the flavin covalent intermediate was a landmark in UGM research, the mechanistic steps leading to formation of this intermediate were not completely understood. Formation of the flavin–sugar adduct was proposed to occur by direct attack of the N5FAD to the C1galp in an SN2-type mechanism (Scheme 2B) [27,28]. Alternatively, the intermediate could form by attack of the N5FAD to an oxocarbenium galactose intermediate, in an SN1-type mechanism [32]. In addition, it was proposed that a single electron transfer step from the flavin to the oxocarbenium intermediate would lead to the formation of a flavin semiquinone and a sugar radical. In this mechanism, the flavin–sugar adduct forms by recombination of the radical pair (Scheme 2B) [26]. Recently, Liu and coworkers probed the mechanism using flavin analogs with different nucleophilicities at the N5FAD. The kinetic linear free energy relationship resulted in a slope of β = −2.45, consistent with a direct attack of the N5FAD in an SN2-type mechanism [33]. In addition, a flavin-iminium intermediate was observed during time-resolved spectroscopy of the reaction of UDP-Galp with reduced T. cruzi UGM (TcUGM) without the formation of a flavin semiquinone intermediate, inconsistent with a single electron transfer mechanism [30]. Furthermore, the structures of the complex of UGM with UDP-Galp clearly show that the N5FAD is the proper distance for direct attack of the C1galp (described in detailed in the next section).

Recent experimental studies have provided insight into the rate-limiting step. Analysis of the kinetics of eUGMs showed that formation of the flavin iminium ion is very fast (∼300 s⁻¹) compared to the κcat (∼12 s⁻¹) for TcUGM. In addition, viscosity effects studies demonstrated that product release is not rate limiting. This led to the proposal that ring contraction is the rate-determining step (Scheme 3, g and h) [30]. In order to produce the flavin-iminium ion, the N5FAD must be deprotonated. Although this aspect of the reaction has not been elucidated by biochemical approaches, recent quantum mechanical calculations provided insights into this and other steps in the UGM reaction [34]. Hybrid quantum-classical calculations performed at the density functional theory

---

### Chemical mechanisms of UGM

#### Catalytic mechanism of UGM, a noncanonical flavoenzyme

Nassau et al. discovered that E. coli UGM is a flavoenzyme [11], and indeed all UGMs characterized to date contain flavin adenine dinucleotide (FAD). Flavoenzymes typically catalyze oxidation–reduction reactions with the flavin serving as the redox center, and thus, the role of FAD in the redox neutral UGM reaction was enigmatic (Scheme 1A). Several mechanistic routes for the interconversion of UDP-Galp to UDP-Galp′ were initially tested in prokaryotic UGMs using a battery of chemical probes. It was shown that the enzyme was active with UDP-2-F-Galp and UDP-3-F-Galp, eliminating the possibility of oxidation of Galp at the 2-OH or 3-OH moiety [18,19]. Further characterization of the recombinant E. coli protein showed that the enzyme was active when the flavin was in the oxidized form, but significantly more active when the enzyme was chemically reduced with dithionite (Scheme 1B) [20]. Subsequent studies showed that only the reduced enzyme exhibits catalytic activity, and the spurious activity attributed to the oxidized enzyme in fact originated from a subpopulation of reduced protein that had persisted in the enzyme preparation [21]. The presence of reduced UGM during purification was clearly established in the UGM from A. fumigatus, as the recombinant protein purified under aerobic conditions remarkably stabilizes 50% of the FAD in the reduced form [22].

Blanchard’s group demonstrated that the anomeric bond was broken during catalysis using positional isotope effects (PIX) [23], which was later supported by work from Liu’s group [18]. Cleavage of the anomeric bond suggested a number of possible mechanisms for UGMs (Scheme 2). One mechanism predicts the formation of a 1,4-anhydrogalactopyranose (Scheme 2A) [18,23]. However, activity was not detected when reduced UGM was incubated with 1,4-anhydrogalactopyranose in the presence of UDP, eliminating this species from consideration as an intermediate in the UGM reaction [24]. A mechanism involving a single-electron transfer step during catalysis was supported by potentiometric studies that showed that the flavin semiquinone was stabilized in the presence of substrate [25]. In addition, replacement of the FAD with 5-deaza-FAD, a flavin analog restricted to a net 2-electron process, resulted in inactive UGM. The lack of activity of UGM reconstituted with 5-deaza-FAD was initially interpreted as supporting an electron transfer step in the catalytic cycle (Scheme 2B) [26].

A major breakthrough in our understanding of the mechanism of action of UGM came when Kiessling’s group isolated an FAD-galactose covalent intermediate [27]. The covalent intermediate formed between the N5FAD and C1galp was proposed and later validated by isolation and characterization by mass spectrometry and NMR in bacterial UGM (bUGM) and by UV/vis spectrophotometry and mass spectrometry in eukaryotic UGM (eUGM) [28–30]. The flavin–sugar adduct is important in ring opening and activation of the C1galp [27]. In this process, the flavin also functions as a molecular scaffold, providing the structural constraints required for ring contraction (Scheme 2B) [31]. Formation of this intermediate made mechanistic sense and was consistent with the breaking of the anomeric bond determined from PIX studies and the lack of activity observed with the enzyme reconstituted with 5-deaza-FAD.

Although identification of the flavin covalent intermediate was a landmark in UGM research, the mechanistic steps leading to formation of this intermediate were not completely understood. Formation of the flavin–sugar adduct was proposed to occur by direct attack of the N5FAD to the C1galp in an SN2-type mechanism (Scheme 2B) [27,28]. Alternatively, the intermediate could form by attack of the N5FAD to an oxocarbenium galactose intermediate, in an SN1-type mechanism [32]. In addition, it was proposed that a single electron transfer step from the flavin to the oxocarbenium intermediate would lead to the formation of a flavin semiquinone and a sugar radical. In this mechanism, the flavin–sugar adduct forms by recombination of the radical pair (Scheme 2B) [26]. Recently, Liu and coworkers probed the mechanism using flavin analogs with different nucleophilicities at the N5FAD. The kinetic linear free energy relationship resulted in a slope of β = −2.45, consistent with a direct attack of the N5FAD in an SN2-type mechanism [33]. In addition, a flavin-iminium intermediate was observed during time-resolved spectroscopy of the reaction of UDP-Galp with reduced T. cruzi UGM (TcUGM) without the formation of a flavin semiquinone intermediate, inconsistent with a single electron transfer mechanism [30]. Furthermore, the structures of the complex of UGM with UDP-Galp clearly show that the N5FAD is the proper distance for direct attack of the C1galp (described in detailed in the next section).

Recent experimental studies have provided insight into the rate-limiting step. Analysis of the kinetics of eUGMs showed that formation of the flavin iminium ion is very fast (∼300 s⁻¹) compared to the κcat (∼12 s⁻¹) for TcUGM. In addition, viscosity effects studies demonstrated that product release is not rate limiting. This led to the proposal that ring contraction is the rate-determining step (Scheme 3, g and h) [30]. In order to produce the flavin-iminium ion, the N5FAD must be deprotonated. Although this aspect of the reaction has not been elucidated by biochemical approaches, recent quantum mechanical calculations provided insights into this and other steps in the UGM reaction [34]. Hybrid quantum-classical calculations performed at the density functional theory
The simulations also predict that during ring contraction, the proton from the hydroxyl group at C4Galp is transferred to the C4 = O of the flavin (Scheme 3, g and h) [34].

level suggest that the C4 = O of the flavin accepts the proton from the N5FAD (Scheme 3, e and f) and donates it to the O5Galp, facilitating ring opening and formation of the iminium ion (Scheme 3, f and g).
Together, the mechanistic and structural data strongly support a novel role of the flavin cofactor as a nucleophile in the reaction catalyzed by UGMs. Novel roles of flavins in noncanonical redox neutral reactions have been identified in other enzymes [31], however, UGM is the only flavoenzyme that forms a covalent intermediate by acting as a nucleophile [9,35,36].

Mechanism of enzyme activation

Although it was known that reduced UGM is the active enzyme species, detailed characterization of the flavin reduction step was lacking until our studies of eUGMs. Eukaryotic UGMs from the fungus A. fumigatus and from the parasites T. cruzi, L. mexicana, and L. infantum have been utilized to NADPH as a redox partner (Table 1) [30,37,38]. In addition, these enzymes can stabilize the reduced form of the flavin, even under aerobic conditions. The rate of oxidation is 200–1500-fold slower than the reduction step, thus, the enzymes can turn over several hundred times before becoming oxidized (inaactive) [30,38]. The structures of A. fumigatus UGM in complex with NADP(H) have recently been solved and a novel NADPH binding domain was identified [37]. Thus, at least for eUGMs, the mechanism of activation by NAD(P)H has been established.

Bacterial UGMs also can be activated by NAD(P)H; however, activation requires excess coenzyme and extended incubation times, suggesting that NAD(P)H may not be the physiological hydride donor [39]. For example, direct measurement of the rate of flavin reduction of M. tuberculosis UGM was shown to be at least 10,000 times slower than for eUGMs [37]. Therefore bUGMs do not effectively react with NAD(P)H for activation. The mechanism of activation of bUGM thus remains to be discovered. One possibility that we favor is that bUGMs might be activated in vivo by acquiring reducing equivalents from a variety of redox partners in the cell. This hypothesis is supported by reports that have shown that bacterial nitric oxide synthase functions without a specific reductase [44,45]. Nevertheless, as in other Rossmann fold enzymes, the flavin adopts an extended conformation and binds at the C-terminal edge of the β-sheet with the pyrophosphate interacting with the Gly-rich loop and a conserved water molecule [41]. Domain 1 also includes two smaller subdomains, denoted 1A and 1B. Subdomain 1A interrupts the Rossmann fold at the C-terminus of β2 and consists of a two-stranded antiparallel β-sheet. Subdomain 1A is followed by subdomain 1B, which is a 3-stranded antiparallel β-sheet and associated α-helix. The number of strands in the sheet varies from two to three in different UGMs. The connector between 1A and 1B is very important for function, as it provides the conserved histidine loop as well as an aromatic residue that serves as the backstop for the substrate Galp moeity. As described below, the histidine loop of eukaryotic UGMs exhibits large conformational changes when the oxidized enzyme is activated by reducing agents such as dithionite and NAD(P)H. Domain 1 also includes a C-terminal helix that is outside of the Rossmann fold. The connector between Rossmann β5 and the C-terminal helix is functionally important as it provides active site residues that cluster near the substrate Galp.

Domain 2 is the only contiguous domain of UGM and is essentially a bundle of five α-helices. This domain functions in binding the uridine group of the substrate and in dimerization. Domain 2 also contains a mobile active site flap that responds to substrate binding. The flap is located on the loop between the last helix of domain 2 and the final strand of subdomain 1B (Fig. 1) and contains a key Arg residue that interacts with the substrate (described below).

Domain 3 features a twisted, 6-stranded anti-parallel β-sheet. This domain is noncontiguous because strand 2 connects to domain 2 rather than strand 3 of its own sheet. As a consequence, strands 2 and 3 of the 6-stranded sheet are separated by ~150 residues. These intervening 150 residues form several structural elements distributed throughout the fold, including all of domain 2, the final strand of subdomain 1B, and about one-half of the Rossmann fold before returning to domain 3 at β-strand 3. Strand 6 of domain 3 exits to domain 1 to complete the UGM fold.

All bUGMs exhibit essentially the same structure as EcUGM. For example, analysis with SSM [42] shows that 73–96% of the secondary structure elements of EcUGM are preserved in other bUGMs, implying conservation of secondary and tertiary structure despite substantial variation in primary structure (37–44% pairwise identity with EcUGM). Also, the root mean square deviation (RMSD) of the other bUGMs to EcUGM spans the range 1.1–1.5 Å. This range is quite low considering that the two EcUGM protomers in the crystallographic asymmetric unit align with an RMSD of 0.97 Å.
Eukaryotic UGMs – elaboration of the basic UGM fold

Structures of eUGMs first appeared in the literature more than a decade after the EcUGM structure (Table 2). Part of this time lapse was likely due to the many technical challenges encountered during structure determination, including translational pseudosymmetry and/or twinning observed in crystals of A. fumigatus UGM (AFUGM) [43–45] and L. major UGM [46]. The surface mutagenesis strategy employed by Dhatwalia et al. proved to be the most effective solution to these problems [43]. The method involves modifying the protein surface using site-directed mutagenesis focused on long, charged residues, especially Lys and Glu [47]. Since the target structure is unknown, one typically uses bioinformatics methods such as the structural alignment, as described below.

Table 2

<table>
<thead>
<tr>
<th>UGM a</th>
<th>PDB code</th>
<th>Resolution</th>
<th>Redox state(s) b</th>
<th>Active site ligand</th>
<th>Release Date</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ec</td>
<td>1BT</td>
<td>2.40</td>
<td>O</td>
<td></td>
<td>2003</td>
<td>[21]</td>
</tr>
<tr>
<td>Mt</td>
<td>1V0</td>
<td>2.25</td>
<td>O</td>
<td></td>
<td>2005</td>
<td>[64]</td>
</tr>
<tr>
<td>Kp</td>
<td>2BI</td>
<td>2.00</td>
<td>O</td>
<td></td>
<td>2005</td>
<td>[64]</td>
</tr>
<tr>
<td>Kp</td>
<td>2BI8</td>
<td>2.35</td>
<td>R</td>
<td></td>
<td>2005</td>
<td>[64]</td>
</tr>
<tr>
<td>Kp</td>
<td>1WAM</td>
<td>2.35</td>
<td>R</td>
<td></td>
<td>2006</td>
<td>[64]</td>
</tr>
<tr>
<td>Kp</td>
<td>3cF4</td>
<td>2.45</td>
<td>O</td>
<td></td>
<td>2009</td>
<td>[29]</td>
</tr>
<tr>
<td>Kp</td>
<td>3INR</td>
<td>2.30</td>
<td>O</td>
<td>UDP, UDP-glucose</td>
<td>2009</td>
<td>[28]</td>
</tr>
<tr>
<td>Kp</td>
<td>3INT</td>
<td>2.51</td>
<td>R</td>
<td>UDP, UDP-Galp</td>
<td>2009</td>
<td>[28]</td>
</tr>
<tr>
<td>Dr</td>
<td>3HQD</td>
<td>2.36</td>
<td>O</td>
<td>UDP-Galp</td>
<td>2009</td>
<td>[65]</td>
</tr>
<tr>
<td>Dr</td>
<td>3HDY</td>
<td>2.40</td>
<td>O/R</td>
<td>UDP-Galp</td>
<td>2009</td>
<td>[65]</td>
</tr>
<tr>
<td>Dr</td>
<td>3HE1</td>
<td>2.40</td>
<td>O</td>
<td>UDP</td>
<td>2009</td>
<td>[65]</td>
</tr>
<tr>
<td>Dr</td>
<td>3MJ4</td>
<td>2.65</td>
<td>O</td>
<td>UDP, UDP-CH₂-Galp</td>
<td>2010</td>
<td>[66]</td>
</tr>
<tr>
<td>Kp</td>
<td>3KYB</td>
<td>2.30</td>
<td>O</td>
<td>FMN</td>
<td>2010</td>
<td>-</td>
</tr>
<tr>
<td>Af</td>
<td>3UTE</td>
<td>2.35</td>
<td>O</td>
<td>Sulfate</td>
<td>2012</td>
<td>[43]</td>
</tr>
<tr>
<td>Af</td>
<td>3UTF</td>
<td>2.25</td>
<td>R</td>
<td></td>
<td>2012</td>
<td>[43]</td>
</tr>
<tr>
<td>Af</td>
<td>3UTG</td>
<td>2.25</td>
<td>R</td>
<td>UDP</td>
<td>2012</td>
<td>[43]</td>
</tr>
<tr>
<td>Af</td>
<td>3UTH</td>
<td>2.25</td>
<td>R</td>
<td>UDP-Galp</td>
<td>2012</td>
<td>[43]</td>
</tr>
<tr>
<td>Af</td>
<td>3UKA</td>
<td>2.64</td>
<td>O</td>
<td></td>
<td>2012</td>
<td>[45]</td>
</tr>
<tr>
<td>Af</td>
<td>3UKF</td>
<td>2.50</td>
<td>R</td>
<td>UDP-Galp</td>
<td>2012</td>
<td>[45]</td>
</tr>
<tr>
<td>Af</td>
<td>3UKN</td>
<td>2.30</td>
<td>O/R</td>
<td>UDP-Galp</td>
<td>2012</td>
<td>[45]</td>
</tr>
<tr>
<td>Af R182K</td>
<td>3UKK</td>
<td>2.75</td>
<td>O/R</td>
<td>UDP</td>
<td>2012</td>
<td>[45]</td>
</tr>
<tr>
<td>Af</td>
<td>3UKL</td>
<td>2.63</td>
<td>O</td>
<td>UDP</td>
<td>2012</td>
<td>[45]</td>
</tr>
<tr>
<td>Af R327A</td>
<td>3UKP</td>
<td>3.10</td>
<td>O</td>
<td>UDP-Galp</td>
<td>2012</td>
<td>[45]</td>
</tr>
<tr>
<td>Af R327K</td>
<td>3UKQ</td>
<td>3.15</td>
<td>O</td>
<td>UDP-Galp</td>
<td>2012</td>
<td>[45]</td>
</tr>
<tr>
<td>Tc</td>
<td>4DSC</td>
<td>2.25</td>
<td>O</td>
<td>UDP</td>
<td>2012</td>
<td>[54]</td>
</tr>
<tr>
<td>Tc</td>
<td>4DSH</td>
<td>2.25</td>
<td>R</td>
<td>UDP</td>
<td>2012</td>
<td>[54]</td>
</tr>
<tr>
<td>Af</td>
<td>4GDC</td>
<td>2.75</td>
<td>O</td>
<td>NADPH</td>
<td>2012</td>
<td>[37]</td>
</tr>
<tr>
<td>Af</td>
<td>4GDD</td>
<td>2.75</td>
<td>O</td>
<td>NADH</td>
<td>2012</td>
<td>[37]</td>
</tr>
<tr>
<td>Af</td>
<td>4GDE</td>
<td>2.20</td>
<td>R</td>
<td></td>
<td>2012</td>
<td>[37]</td>
</tr>
</tbody>
</table>

a Listed in chronological order of release date.

b Enzyme sources abbreviated as follows: Ec, Escherichia coli; Kp, Klebsiella pneumoniae; Dr, Deinococcus radiodurans; Af, Aspergillus fumigatus; Tc, Trypanosoma cruzi.

c Abbreviations: O, oxidized; R, reduced; O/R, oxidized and reduced forms present in the same crystal.

Oligomeric state and quaternary structure

Despite substantial conservation of protein fold, UGMs display a variety of oligomeric states. Bacterial UGMs form dimers and possibly trimers, whereas eUGMs form monomers and tetramers.
Bacterial UGMs

Dimer seems to be the prominent form of bUGMs. Dynamic light scattering analysis of eUGM suggests a dimer in solution [49]. Analysis of protein–protein interfaces in crystal lattices reveals a common dimeric assembly for UGMs from *E. coli*, *M. tuberculosis*, and *K. pneumoniae*. The observation of the same 2-body assembly in different crystal lattices of different proteins strongly suggests that this unit is formed in solution. The dimer is a semi-circular particle with the interface formed by domain 2 of one protomer packing against the β-sheet of domain 3 of another protomer (Fig. 3A).

The crystal structure of *D. radiodurans* (DrUGM) suggests a different oligomeric state. Inspection of the crystal lattice shows, surprisingly, that the semi-circular dimer is absent. However, analysis of protein–protein interfaces with PDBePISA [42] suggests that a different dimer may be formed in solution. In contrast to the semi-circular dimer, the DrUGM dimer is formed primarily via domain 1 (Fig. 3B). A notable interaction in the dimer interface is a short stretch of intermolecular β-sheet formed by β3 of the Rossmann fold (Fig. 3B, right panel). In the crystal, five such dimers assemble into a pentamer–of–dimers decamer (Fig. 3C). Analysis with PDBePISA suggests that this large assembly is also potentially stable in solution. Biophysical studies are needed to determine the solution oligomeric state of DrUGM.
Eukaryotic UGMs

The oligomeric state and quaternary structure of AfUGM were determined using a combination of small-angle X-ray scattering and X-ray crystallography [43]. These studies showed that AfUGM forms a dimer-of-dimers tetramer in solution (Fig. 3D). The tetramer has 222 point group symmetry and comprises three different 2-body assemblies. Interestingly, none of these assemblies resemble the bUGM dimers, and thus AfUGM exhibits a unique quaternary structure. As described previously [43], the additional structural features of AfUGM, such as the extra helix in domain 2 and the inserts in domain 3, preclude the assembly of bUGM-like dimers.

The interfaces of AfUGM have been analyzed in detail to identify structural features that stabilize the tetramer [43]. This analysis suggests that the unique tetramer results from secondary structure elements of AfUGM that are absent in bUGMs. These include the U-shaped C-terminal extension, the extra helix of domain 2, and the elongated helix of domain 2. All three substructures figure prominently in the tetramer interfaces. In particular, the extra and elongated helices of domain 2 form an intersubunit 4-helix bundle at one of the 2-fold axes (Fig. 3D, right panel). In summary, the unique variation of the UGM fold displayed by AfUGM precludes formation of the classic semicircular dimer and promotes formation of a new tetrameric assembly.

Although the solution structural properties of TcUGM have not been studied as rigorously as AfUGM, the available data suggest a monomeric protein in solution. Size exclusion chromatography data are consistent with a monomer [22]. Also, neither the AfUGM-type tetramer nor any of the previously observed UGM dimers are present in the TcUGM crystal lattice. Curiously, a disulfide-linked dimer is present in the lattice, but this species is likely an artifact of crystal packing.

The monomeric state of TcUGM can be understood in terms of the tertiary structure. As with AfUGM, the extra helix of domain 2 and the inserts in domain 3 likely prevent formation of bUGM-like dimers. Why TcUGM does not form the AfUGM-like tetramer can also be rationalized. As described above, structural differences between TcUGM and AfUGM are evident in domains 1 and 2 (Fig. 2). Because of these differences, TcUGM lacks some of the components of the tetramer interfaces, such as the C-terminal helix of domain 1 and an Arg residue (Arg 133 in AfUGM, Fig. 2B) that forms intersubunit hydrogen bonds in the AfUGM tetramer. Also, long side chains protruding from a helix of domain 2 (Fig. 2B) likely prevent formation of the intersubunit 4-helix bundle that stabilizes the AfUGM tetramer. In summary, the unique fold-level variations exhibited by TcUGM are responsible for the monomeric state. Thus, as with AfUGM, tertiary structure dictates the oligomeric state.

Substrate recognition

The structural basis of substrate recognition has been characterized. Crystal structures of bUGMs and eUGMs complexed with UDP-Galp or UDP have been determined (Table 2), allowing conserved and unique features of substrate recognition to be identified.

Conserved themes of substrate recognition

Some aspects of substrate recognition appear to be shared by all UGMs. For example, the substrate consistently binds with the Galp moiety next to the flavin isoalloxazine such that the N5FAD interacts with the P1 position of the sugars. Although the precise orientations of the Phe and Tyr residues vary between bUGMs and eUGMs, their role in substrate recognition is evident (Fig. 5A and B). Two conserved aspects are evident (Fig. 5A and B). First, a pair of Tyr residues forms hydrogen bonds with the pyrophosphate. These residues are located on the C-terminus of the flavin and AfUGM, their role in shaping the uracil pocket remains a constant (compare Fig. 5A and B). One wall of the uridine pocket is highly conserved. The uracil wall consists of four residues belonging to the fourth helix of domain 2. These residues are conserved at the sequence level and include Trp and Asn/Thr side chains that hydrogen bond to the ribose hydroxyls and two aromatic residues (Tyr and Phe) that stack against the uracil.

Common themes of substrate recognition are also evident in detailed enzyme–substrate noncovalent interactions (Table 3, Fig. 5A and B). One wall of the uridine pocket is highly conserved. The uridine wall consists of four residues belonging to the fourth helix of domain 2. These residues are conserved at the sequence level and include Trp and Asn/Thr side chains that hydrogen bond to the ribose hydroxyls and two aromatic residues (Tyr and Phe) that stack against the uracil. Although the precise orientations of the Phe and Tyr residues vary between bUGMs and eUGMs, their role in shaping the uracil pocket remains a constant (compare Fig. 5A and B). All UGMs appear to share a common theme of pyrophosphate recognition. Two conserved aspects are evident (Fig. 5A and B). First, a pair of Tyr residues forms hydrogen bonds with the pyrophosphate. These residues are located on the C-terminus of the flavin of domain 3 and the loop following Rossmann fold β5. A conserved ion pair also stabilizes the pyrophosphate. The ion pair consists of Arg and Glu residues located on strands 5 and 6 of do-
main 3, respectively (Fig. 1). The Arg side chain forms ionic interactions with the pyrophosphate (Fig. 5A and B).

All UGMs have a dynamic Arg that participates in substrate recognition. This key residue is located in the middle of a mobile active site flap that moves substantially when the active site transitions from the open, ligand-free conformation to the substrate-bound, closed state (Fig. 4). As described above, bUGMs have one active site flap (Fig. 4B), whereas eUGMs have two mobile flaps. The dynamic Arg of eUGMs is part of flap 1 (Fig. 4A). In both UGM classes, the dynamic Arg is solvent exposed in the open state, and interacts with the substrate in the closed state. These interactions vary somewhat in different structures. The dynamic Arg forms a hydrogen bond with the Galp 2-OH in AfUGM (Fig. 5A) and DrUGM (Fig. 5B). This interaction is missing in K. pneumoniae UGM (KpUGM) (Fig. 4B). In bacterial UGMs, the dynamic Arg additionally interacts with the pyrophosphate (Fig. 4B), and/or the Galp 3-OH (Fig. 5B). Despite these variations, the dynamic Arg is always stabilized by a conserved Asn/Asp contributed by the loop following Rossmann β5 and positioned near the pyrimidine ring of the isoalloxazine. The location of the Arg–Asn/Asp link near the Galp moiety and isoalloxazine suggests that this interaction is important for locking down the closed active site conformation and positioning the sugar for catalysis.

Conserved steric interactions are also important for substrate recognition. All UGMs have an aromatic residue that contacts the O3–O4 locus of Galp and appears to serve as a backstop that helps position the sugar for catalysis (Fig. 5A and B). The backstop residue is part of the conserved histidine loop and corresponds to Phe in eUGMs (Phe66 of AfUGM) and His in bUGMs (His88 of DrUGM) (Fig. 6). The backstopping His of bUGM is well outside of hydrogen bonding distance to the Galp hydroxyls (3.9–4.7 Å), consistent with a steric role (Fig. 5B).

The final conserved aspect of substrate recognition involves a “third Arg” residue in the active site (Arg447 of AfUGM, Arg364 of DrUGM). This residue is located on the loop following Rossmann β5 and is positioned near the dimethylbenzene ring of the isoallox-
azine (Fig. 5A and B). The third Arg does not directly contact the substrate, but forms water-mediated hydrogen bonds to the pyrophosphate and Galp in AfUGM.

Unique aspects of substrate recognition

Despite the numerous conserved features described above, eUGMs and bUGMs differ in some aspects of substrate binding. The largest of these involve the UMP moiety. The UMP bound to AfUGM or TcUGM is displaced by 3–5 Å compared to the bacterial enzymes, with the uridine exhibiting the largest shift (Fig. 5C). This variation reflects both sequence and structural differences in domain 2. In AfUGM, the uridine packs against Tyr104 and Tyr317, while forming hydrogen bonds with Gln107 (Fig. 5A). Identical interactions are observed in TcUGM complexed with UDP. These three residues are different in bUGMs. For example, Tyr317 is replaced by Asn in bUGMs (Fig. 5A, Asn296 in DrUGM), and bUGMs lack a residue analogous to Gln107 because of local protein conformational differences. Furthermore, bUGMs have nonpolar residues occupying the space corresponding to the uridine pocket of eUGMs.

The UMP phosphate interactions are also different in the two UCM classes. A Tyr residue from β4 of domain 3 interacts with the UMP phosphate eUGM (Tyr317 in eUGM, Fig. 5A), whereas a different Tyr from the linking peptide between domain 2 and subdomain 1B performs this function in bUGMs (Tyr209 in DrUGM, Fig. 5B). This difference likely contributes to the observed difference in the pyrophosphate dihedral angle.

The two classes of UGM also differ in the region near the O4–O5 locus of Galp. Eukaryotic UGMs have a dynamic Asn residue and a Trp side chain in addition to the Galp backstop (Fig. 5A). The dynamic Asn (Asn207 in AfUGM) is located on mobile flap 2, a region that is stationary in bUGMs. This key residue is solvent exposed in the open state and moves 15 Å to form a hydrogen bond to the O4 hydroxyl in the closed state (Fig. 4A). The Trp residue (Trp315 of AfUGM) forms a steric interaction with C6 of Galp (Fig. 5A). In contrast, bUGMs have a second backstopping residue (Phe210 in DrUGM) and no hydrogen-bonding residue (Fig. 5B). These protein structural differences likely account for the different orientations of the O6 hydroxyl in the two UGM classes.

In summary, bUGMs and eUGMs share many common elements of substrate recognition. However, there are also aspects of substrate binding that are unique to each class of UGM, and the two classes differ most substantially in the region near the UMP. This
The NAD(P)H site of eukaryotic AFUGM

The crystal structures of oxidized AFUGM complexed with NADPH or NADH were recently reported [37]. NAD(P)H interacts exclusively with domains 1 and 3. One face of NAD(P)H packs into a groove at the junction of these domains, while the other face is exposed to the vacant substrate-binding cavity (Fig. 7A). NAD(P)H adopts a compact conformation in which the adenine forms a hydrogen bond with a ribose hydroxyl (Fig. 7B).

The nicotinamide is near the FAD isoalloxazine, and the ADP group of NADPH and the UDP group of the substrate occupy distinct pockets (Fig. 7C), consistent with the observation that NADPH makes no interactions with domain 2. (Recall that domain 2 functions prominently in binding the substrate uridine.) Thus, the NADPH complex reveals a new ligand-binding pocket and provides new opportunities for inhibitor discovery. The new pocket is defined by several residues that form nonpolar and electrostatic interactions with the AMP (Fig. 7B). The adenosine of NADPH binds in a hydrophobic pocket formed by Ile65, Phe66, Tyr104, and the non-polar chain of Arg91. The adenine base forms hydrogen bonds with Ser93 and Tyr317. The pyrophosphate is stabilized by His68 and Asn457. Tyr104 makes a hydrogen bond with the 2’-phosphoryl of NADPH. This interaction is responsible, at least in part, for the 179-fold higher kred/Ki value for NADPH compared to NADH observed with AFUGM [37].

All of the residues of the NADPH site are identically conserved in other eukaryotic UGMs, such as TcUGM and L. major UGM, two enzymes that are of interest for inhibitor design (Fig. S3 of [37]). This high degree of sequence conservation suggests that the identified binding site is present in other eUGMs. In contrast, none of the residues that contact the ADP half of NADPH are present in the sequences of bUGMs, suggesting that the identified NAD(P)H site is not present in bUGMs. We note that the binding site for NADP in AFUGM is entirely different from the Rossman dinucleotide-binding fold, which is the most common structural motif for binding nicotinamide adenine dinucleotide cofactors [41]. Furthermore, the compact NADPH conformation observed in AFUGM differs substantially from the extended conformations typically seen in other enzymes. These observations suggest that compounds targeted at the AFUGM NADPH site should have low affinity for other enzymes that use nicotinamide dinucleotide cofactors. Thus, the NADPH site of AFUGM provides a potential platform for discovering eUGM-specific inhibitors.

Conformational changes associated with enzyme activation – a new flavin switch protein?

Some flavoenzymes are activated by a change in the redox state of the flavin [50]. These flavin switch proteins exhibit large conformational changes upon flavin oxidation or reduction that promotes functions such as light-induced signaling, transcription regulation, and membrane binding. Since UGM is activated by flavin reduction, it is potentially another example of a flavin switch protein. Studies of bUGMs reduced in crystallo showed that reduction induces bending of the isoalloxazine but no perceivable change in the protein structure [28, 51]. However, recent studies of eUGMs have re-

### Table 3
Conserved elements of substrate recognition.

<table>
<thead>
<tr>
<th>Element</th>
<th>Role</th>
<th>DrUGM</th>
<th>AFUGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uridine wall</td>
<td>Four residues of domain 2 that contact the uridine group</td>
<td>Ec 54 GAHIFH 59</td>
<td>Mt 63 GAHLFH 68</td>
</tr>
<tr>
<td>Tyr pair</td>
<td>Two Tyr residues that hydrogen bond with the pyrophosphate</td>
<td>Kp 58 GPHIFH 63</td>
<td>Dr 83 GPHIFH 88</td>
</tr>
<tr>
<td>Ion pair</td>
<td>The Arg of an Arg–Glu ion pair interacts with the pyrophosphate</td>
<td>Lm 57 GGHVIF 62</td>
<td>Tc 60 GGHVIF 65</td>
</tr>
<tr>
<td>Dynamic Arg</td>
<td>Located on a mobile active site flap; interacts with Galp; stabilized by Asn or Asp</td>
<td>Af 61 GGHVIF 66</td>
<td>Ce 90 GGHVF 95</td>
</tr>
<tr>
<td>Galp</td>
<td>Contacts the O3–O4 locus of Galp; last residue of the histidine loop</td>
<td>H88</td>
<td>F66</td>
</tr>
<tr>
<td>3rd Arg</td>
<td>One of three Arg residues in the active site; near dimethylbenzene ring of isoalloxazine; does not directly contact the substrate</td>
<td>R364</td>
<td>R447</td>
</tr>
</tbody>
</table>
revealed that the protein conformation near the flavin changes dramatically in response to flavin reduction, likely heralding a new chapter in UGM biochemistry.

Redox-linked conformational changes observed in AfUGM

Structural studies of AfUGM provide the best evidence for large redox-linked conformational changes. The AfUGM structure has been determined from two different crystal forms, hexagonal [43] and triclinic [45]. Both forms yielded structures of the oxidized and in crystallo reduced enzymes. In addition, the hexagonal form was used to trap a complex of oxidized AfUGM with NADPH (see previous section) [37]. The hexagonal and triclinic reduced enzyme structures are essentially identical. The fact that the two crystal forms give similar structures is reassuring and strongly suggests that the crystalline enzyme conformation is an excellent model for the solution conformation. Likewise, the structures of the NADPH complex and triclinic oxidized enzyme agree well. Again, different crystal forms yielding the same conformation implies a faithful representation of the protein in solution. In contrast, the hexagonal structure of ligand-free oxidized AfUGM displays an unusual conformation of the histidine loop and is likely an artifact of the high sulfate ion concentration and low pH used in crystallization, as we have previously suggested [37,43]. Thus, the hexagonal NADPH complex and triclinic oxidized enzyme structures provide the correct conformation of oxidized AfUGM for comparison to the reduced enzyme.

The AfUGM structures imply that enzyme activation induces profound protein conformational changes (Fig. 8C). In the oxidized enzyme, His63 is near the pyrimidine portion of the isoalloxazine and oriented parallel to Trp315, while Arg327 donates a hydrogen bond to the N5FAD. This hydrogen bond is diagnostic of the oxidized enzyme because Arg is an obligate hydrogen bond donor at physiological pH, and N5FAD is an obligate acceptor only in the oxidized state (Scheme 1B). Note also that the imidazole of His63 is not in contact with the isoalloxazine. Flavin reduction by either dithionite or NAD(P)H induces a dramatic reconfiguration of the conserved histidine loop in which His63 and Gly62 move by 6 and 5 Å, respectively (Fig. 8C). His63 moves to the si face of the isoalloxazine where it forms hydrogen bonds with the carbonyl oxygen of Gly61 and the 2′-OH of the ribityl chain (Fig. 8E). Furthermore, Gly62 changes conformation in order to accept a hydrogen bond from the N5FAD 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 (Scheme 1B). The conformational change of the histidine loop is coordinated with an 8-Å movement of Trp315 to evade Gly62 and rupture of the Arg327-N5 hydrogen bond (Fig. 8C).

Reduction of AfUGM also changes the conformation of the flavin itself. Upon reduction, the isoalloxazine tilts by 2 Å (Fig. 8C). Although smaller in magnitude, this movement is reminiscent of the mobile flavin in p-hydroxybenzoate hydroxylase [52,53]. Also, at the crystallographic resolutions of these structures (~2.2 Å) it is possible to discern the planarity of the isoalloxazine. The maps show that the isoalloxazine is planar in the oxidized state (see Figure S1 of [43] and Fig. 4 of [37]), whereas the reduced flavin exhibits a butterfly-like conformation in which the pyrimidine ring bends 7° out of the plane such that the si face is concave (Fig. 8A).

Redox-linked conformational changes observed in TcUGM

Crystal structures of TcUGM show a different set of conformational changes induced by flavin reduction (Fig. 8D). The structures of oxidized and reduced TcUGM complexed with UDP have been determined [54]. The active site structure of the reduced TcUGM–UDP complex is essentially identical to that of reduced
AfUGM–UDP; even water molecules that bridge UDP to the enzyme are located similarly in the two enzymes. Also, reduction induces bending of the isoalloxazine ring as in AfUGM (Fig. 8B and Fig. 4 of [54]). However, the oxidized TcUGM structure revealed a new conformation of the histidine loop, implying a different set of protein conformational changes associated with enzyme activation. In oxidized TcUGM, the conserved histidine loop is retracted from the FAD isoalloxazine (Fig. 8D). This conformation is stabilized by hydrogen bonds between Asp58 and the backbone of the histidine loop. Upon reduction, the histidine loop is released and shifts 2.3 Å toward the isoalloxazine, allowing Gly61 to accept a hydrogen bond from the flavin N5 (Fig. 8D). As noted above for AfUGM, this hydrogen bond is observed in all reduced UGM structures and is considered essential for stabilizing the reduced flavin. Movement of the histidine loop also triggers rotations of Asp58 and Thr212 so that they engage in a hydrogen bond in the reduced state. Finally, analysis of hydrogen bond interactions suggests that flavin reduction induces protonation of His62 and rotation of its imidazole ring by 180° [54].

**Molecular dynamics studies of active site flexibility**

The dramatically different conformations of the active site flaps observed in various UGM crystal structures indicate that these regions are highly flexible. This flexibility is presumably important for function, since the flaps contain residues that directly contact the substrate (Fig. 4). Furthermore, movement of the flaps is essential for creating the closed active site. As suggested previously [43], flap closure not only assembles the constellation of residues needed for substrate recognition but also creates a protected environment for catalysis and prevents the severed UDP from migrating out of the active site during the catalytic cycle (Scheme 3, e–h). Thus, protein flexibility is a fundamental aspect of UGM catalysis, particularly for eUGMs, which have an additional mobile active site flap and a highly flexible histidine loop.

Molecular dynamics (MD) simulation is used widely to study enzyme flexibility [55] and has been applied to UGM. The first model of a prokaryotic substrate-bound UGM structure was based on short (5 ns) MD simulations in which UDP-Galp and analogs
were docked into the substrate-free active site [56,57]. The simulations predicted that the flexible flap changed from an open conformation to a closed conformation in the presence of the substrate. This prediction would be confirmed by crystal structures of UGMs complexed with UDP and UDP-Galp.

MD simulations have also been used to study substrate-associated active site motion in TcUGM [58]. Classical and accelerated MD simulations agree with the proposed hypothesis that ligand binding influences the open-closed equilibrium of the mobile flaps. After extensive MD simulations of both substrate-free and substrate-bound TcUGM, it was observed that both flaps open in the absence of ligand, thus creating a channel for ligand uptake [58]. When UDP-Galp is bound, intramolecular interactions help maintain both flaps closed. Interestingly, MD simulations show that another flap occasionally moves increasing the channel size. This new mobile loop (residues 462–469 of TcUGM, Fig. 2A) is located in the U-shaped C-terminal region of domain 1, which is absent in bUGMs.

The simulations also suggest a modular mechanism in which each moiety of the substrate controls the flexibility of a distinct flap. When either UDP or UDP-Galp is in the active site, both flaps remain closed, although some distortions appear in the flap 2 in the UDP-bound structure, suggesting that Galp is important in the dynamics of flap 2 (Fig. 9). When the diphosphate moiety is removed from the ligand, and only the uridine moiety was left in the active site, flap 2 displays large distortions similar to those observed in the apo form, although no distortion was observed in flap 1. The dynamic Arg in TcUGM (Arg176), which lies on flap 1, was observed to interact with the bound UDP in the TcUGM structures, however this flap only opens as widely as observed in the apo form when the ligand is completely removed (Fig. 9). The conformation of flap 1, although substrate-dependent, does not seem to be strictly a result of the observed interactions between the dynamic Arg and the bound ligand.

Visual inspection of the TcUGM conformations suggests that the uridine moiety controls flap 1 through a hydrogen bond network that involves Ala178, Gln103, and Phe102. Similar stabilization of the uridine moiety is observed in KpUGM with Phe151, Thr156, and Trp160 (the uridine wall-described previously). On the other hand, the diphosphate and Galp moieties seem to control flap 2 through Asn201 in TcUGM, although no homologue for Asn201 seems to exist in bUGMs, further highlighting the differences between the two UGM classes.

**Summary and Outlook**

The structural and biochemical studies of UGMs have revealed a unique 3-dimensional structure and novel role for flavins in biological processes. Catalysis by UGMs (in particular in eUGMs) is coupled to extensive conformational changes that are modulated by substrate binding and the redox state of the flavin. The aggregate research on UGMs sets the stage for new discoveries.

Exploring the relationship between dynamics and catalytic activity should be a productive area of inquiry. In particular, the MD simulations raise new questions about the mechanism of product release. MD simulations show that UMP is essential for closure of the active site. Since UMP is present in both the substrate and product, the molecular events that promote opening of the active site for product release are unclear. One possibility is that the formation of Galp shifts the equilibrium toward the open state. Determining structures of UGM complexed with UDP-Galp should shed light on this aspect of the mechanism.

The study of redox-linked conformational changes should also be an exciting future area of UGM research. It is clear that activation of eUGMs triggers profound conformational changes. The initial data suggest that the inactive conformations of AfUGM and TcUGM differ, and thus different types of conformational changes bring each enzyme into the active form. However, it should be noted that the structure of ligand-free TcUGM has not yet been determined, and the presence of UDP in the current structures could certainly bias the conformation of the histidine loop. Therefore, a critical goal should be to determine the structure of TcUGM.
and other UGMs in the oxidized, ligand-free state in order to assess conservation of the oxidized enzyme conformation and redox-linked conformational changes. Another goal should be to identify the sequence and structural elements that enable redox-linked conformational changes in UGMs. Sequence analysis suggests one working hypothesis that an additional Gly residue in the histidine loop of eUGMs, which is Ala or Pro in bUGMs, imparts the flexibility to the histidine loop (Fig. 6).

Finally, UGM presents several challenges for inhibitor design that need to be addressed. For the eUGMs, a simple, robust, high-throughput screening assay has not been developed. Application of other methods, such as ThermoFAD and ThermoFluor with focused libraries might be worth pursuing [58,60]. The wealth of structural information on UGM-substrate complexes would seem to bode well for computer-aided drug design. However, the large conformational changes associated with substrate binding must be taken into account during docking, which is challenging. UGM is perhaps a good test case for developing and testing in silico docking methods that incorporate receptor flexibility. Finally, covalent inactivators represent a new potential area of UGM research. The reduced flavin is a reactive nucleophile and could be targeted by inactivators will rely heavily upon the accumulated knowledge of the structure and mechanism of UGM summarized here.

Acknowledgments

This work was supported by a National Institutes of Health Grant GM094469 (P.S. and J.J.T). Work at UCSD is supported in part by NIH, NSF, HHMI, NBCR, and the NSF Supercomputer Centers. L.B. is the recipient of a Pew Latin American Fellowship.

References