Structure of a Sedoheptulose 7-Phosphate Cyclase: ValA from Streptomyces hygroscopicus

Citation	Kean, K. M., Codding, S. J., Asamizu, S., Mahmud, T., & Karplus, P. A. (2014). Structure of a Sedoheptulose 7-Phosphate Cyclase: ValA from Streptomyces hygroscopicus. Biochemistry, 53(26), 4250-4260-4250-4260. doi:10.1021/bi5003508
DOI	10.1021/bi5003508
Publisher	American Chemical Society
Version	Accepted Manuscript
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Structure of a Sedoheptulose 7-Phosphate Cyclase: ValA from *Streptomyces hygroscopicus*

Funding Source Statement: This work was in part supported by an Oregon State University Provost's Distinguished Graduate Fellowship (to KK) and a grant R01AI061528 from the National Institute of Allergy and Infectious Diseases (to TM).

Kelsey M. Kean[1], Sara J. Codding[1], Shumpei Asamizu[2], Taifo Mahmud[2], and P. Andrew Karplus[1,*]

[1] Department of Biochemistry & Biophysics, 2011 ALS Building, Oregon State University, Corvallis, OR 97331-7305, USA

[2] Department of Pharmaceutical Sciences, 203 Pharmacy Building, Oregon State University, Corvallis, OR 97331-7305, USA

[*] Corresponding author: E-mail: karplusp@science.oregonstate.edu;

Contact information for editorial correspondence: P. Andrew Karplus

Tel.: 541-737-3200; Fax: 541-737-0481

Abbreviations:

CG6P: carbaglucose-6-phosphate; CBP: carbaphosphonate; DAHP: 3-deoxy-D-arabino-heptulosonate 7-phosphate; DHQ(S): dehydroquinate (synthase); DDG(S): desmethyl-4-deoxygadusol (synthase); DOI(S): 2-deoxy-*scyllo*-inososе (synthase); EEV(S): 2-epi-5-epi-valiolone (synthase); EV(S): 2-epi-valiolone (synthase); G6P: glucose 6-phosphate; SH7P: sedoheptulose 7-phosphate; SH7PC: sedoheptulose 7-phosphate cyclase; SPC: sugar phosphate cyclase

Abstract

Sedoheptulose 7-phosphate cyclases (SH7PCs) encompass three enzymes involved in producing the core cyclitol structures of pseudoglycosides and similar bioactive natural products. One such enzyme is ValA from *Streptomyces hygroscopicus* subsp. *jinggangensis* 5008 which makes 2-epi-5-epi-valiolone as part of the biosynthesis of the agricultural antifungal agent validamycin A. We present, as the first SH7PC structure, the 2.1 Å resolution crystal structure of ValA in complex with NAD^+ and Zn^{2+} cofactors. ValA has a fold and active site organization resembling the sugar phosphate cyclase dehydroquinate synthase (DHQS) and contains two notable, previously unrecognized interactions between NAD^+ and Asp side chains conserved in all sugar phosphate cyclases that may influence catalysis. Because the domains of ValA adopt a nearly closed conformation even though no sugar substrate is present, comparisons with a ligand-bound DHQS provide a model for aspects of substrate binding. One striking active site difference is a loop that adopts a distinct conformation as a result of an Asp → Asn change with respect to DHQS and alters the identity and orientation of a key Arg residue. This and other active site differences in ValA are mostly localized to areas where the ValA substrate differs from that of DHQS. Sequence comparisons with a second SH7PC making a product with distinct stereochemistry lead us to postulate that the product stereochemistry of a given SH7PC is not the result of events taking place during catalysis, but is accomplished by selective binding of either the α or β pyranose anomer of the substrate.

Natural products have served as a major source of pharmaceuticals and bioactive molecules for centuries and continue to play key roles in guiding the development of new therapeutics today. Among these are pseudooligosaccharides[1], such as the antidiabetic drug acarbose, the crop protectant validamycin A, the antitumor agent cetoniacytone A, and the sunscreen mycosporin-like amino acids which have similar core cyclitol structures (Figure 1A). The core cyclitols of these molecules are generated from the pentose phosphate pathway intermediate sedoheptulose 7-phosphate (SH7P) by one of three enzymes present in some bacteria and fungi that allow SH7P to be used in secondary metabolism. The enzymes – 2-epi-5-epi-valiolone synthase (EEVS), 2-epi-valiolone synthase (EVS), and desmethyl-4-deoxygadosol synthase (DDGS) – each catalyze the cyclization of SH7P to a distinct C7-cyclitol product (Figure 1B). These enzymes, discovered about 15 years ago[2, 3] and known as SH7P cyclases (SH7PCs), are a part of the sugar phosphate cyclase (SPC) family of enzymes, all of which require NAD^+ and a metal ion, either cobalt or zinc, as prosthetic groups[4-7].

The SH7PCs are structurally uncharacterized and current understanding of their enzyme mechanisms are based mostly on studies of two other sugar phosphate cyclases: dehydroquinate synthase (DHQS) and 2-deoxy-*scyllo*-inosose synthase (DOIS). DHQS acts on the substrate 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP; Figure 1B) as part of aromatic amino acid biosynthesis and has been well-studied because it is a target for antimicrobial drug development[4]. Numerous structures of DHQS have been solved from bacteria, fungi, and plants under a variety of conditions, including in the presence and absence of carbaphosphonate (CBP), a substrate analogue inhibitor, and in the presence or absence of NAD^+ [4, 8-12]. DOIS acts on the substrate glucose 6-phosphate (G6P; Figure 1B) and is not as well studied, but structures of it have been

solved both in the presence and absence of carbaglucose 6-phosphate (CG6P), a mechanism-based inhibitor[5].

Both DHQS and DOIS form homodimers. Each individual subunit is composed of two domains connected by a hinge, and in DHQS, the domains move closer together upon substrate binding. The two domains are a N-terminal NAD^+-binding α/β-sandwich and a C-terminal metal-binding α-helical domain. These same two domains are found in class III metal-dependent polyol dehydrogenases (polyol-DHs), that use $NAD(P)^+$ and a bivalent metal to carry out distinct but related chemistry to SPCs[13]. All together, these enzymes form what is known as the DHQS-like superfamily[14]. In SPC structure descriptions, the NAD^+-binding domain has been identified as a Rossmann fold[4, 5, 15], but it has been noted that this assignment is not certain[13].

Extensive studies[4, 16] have established that in converting DAHP to dehydroquinate (DHQ), the DHQS active site coordinates the substrate at its active site metal via two hydroxyls (Figure 1B), and then catalyzes a remarkable five reactions: alcohol oxidation by NAD^+, phosphate β-elimination, carbonyl reduction by the earlier formed NADH, ring opening, and intramolecular aldol condensation. As is common among homologous enzyme pairs with distinct but related chemistry, the first step of the reaction in DHQS and the polyol-DH enzymes, NAD^+ promoted oxidation of an alcohol, is conserved[17]. Based on its crystal structures with and without substrate analogs, DOIS is proposed to have a similar mechanism to DHQS[5], and analogous mechanisms involving the same five steps have been proposed for EEVS, EVS, and DDGS[6]. Based on comparisons of SH7PC sequences with those of DHQS and DOIS, 14 putative active site residues were identified in EEVS, EVS, and DDGS sequences that were mostly identical, but showed characteristic variations in each of the three SH7PC types[6]. These patterns of variation, however, provided no insight into how the active sites produce distinct products from the same

substrate, especially regarding the differing stereochemistry at the C2-position of the EEVS and EVS products (Figure 1B); this remains a major unanswered question. Developing a better understanding of the structure and function of SH7PCs will provide a foundation for their use in generating new bioactive compounds through synthetic biology and semisynthetic production[18].

Here, we present a crystal structure of ValA, the EEVS from *Streptomyces hygroscopicus* 5008 that is involved in the biosynthetic pathway of the agricultural antifungal agent validamycin A[3]. This first structure of a SH7PC fortuitously includes tightly-bound Zn^{2+} and NAD^{+} cofactors and provides an informative view of the residues lining the active site. We combine sequence comparisons with the various SH7PCs sequences and structural comparisons with DHQS and DOIS substrate analog complexes and develop an unexpected hypothesis for how these different S7PCs can use the same substrate to generate different products.

Materials and Methods

Expression, Purification, and Crystallization

The expression of recombinant ValA was done as previously described[6]. For purification, all done at 4 °C, cell pellets from 100 ml cultures were each resuspended in ~5 ml of 40 mM HEPES, 300 mM NaCl, pH 8.0 (buffer A) with 10 mM imidazole, sonicated (13 watts, 4 x 1 min), and centrifuged (14,500 rpm, 30 min). The supernatant was loaded onto a Ni-NTA resin column (5 ml resin; 0.8 ml/min). After washing with 100 ml buffer A with 20 mM imidazole, the protein was eluted using a 200 ml gradient from 20 to 500 mM imidazole in buffer A. Fractions (~6 ml each) containing protein were combined and dialyzed overnight against 2 L 10 mM Tris-HCl, 300 mM NaCl, 5 mM imidazole, pH 8.0. A second phase of purification was done similarly

using a TALON column (~40 ml run at 0.3 ml/min) in buffer B (20 mM Tris-HCl, 300 mM NaCl, pH 8.0) with 5 mM imidazole for column equilibration, 10 mM imidazole for washing, and a 200 ml gradient from 10 mM to 200 mM imidazole for elution. Fractions (~4 ml each) containing pure ValA as judged by SDS gel electrophoresis were combined and dialyzed against 2 L 10 mM Tris-HCl, pH 7.5 (3 x 3 h). The protein solution was concentrated by ultrafiltration (10 K cutoff membrane) to 10 mg/mL, flash frozen in liquid nitrogen, and stored at -80 °C.

The enzyme was crystallized at 4 °C in hanging drops formed from 4 μL protein stock and 1 μL of a 0.6 M succinic acid, pH 6.5 reservoir solution. The resulting crystals were rod-like with dimensions of about 50 x 50 x 200 μm^3.

X-ray diffraction data collection

For diffraction data collection (at –170 °C), crystals were briefly passed through a solution containing 20% PEG 400 and then flash frozen in loops by plunging into liquid nitrogen. Data were collected from two crystals using λ = 1.0 Å X-rays and $\Delta\varphi$ = 1° steps at beamline 5.0.2 at the Advanced Light Source (Berkeley, CA). From both crystals, 120 2.0 s images were collected at detector distance d = 250 mm, and from the second, an additional 200 2.0 s images were collected at d = 350 mm. All these images were integrated using Mosflm[19] and merged using the CCP4 suite of programs[20, 21] to obtain the data set used for structure solution and refinement. The merged data set was usable out to 2.1 Å using $CC_{1/2}$ ~0.2 as the cutoff criteria (Table 1), and a random 5% of reflections were marked for cross-validation. In addition, a third crystal was used for a fluorescence scan and to collect a data set at beamline 5.0.2 using λ = 1.282 Å X-rays to maximize the anomalous signal from the bound zinc. This data set included

two sets of 60 $\Delta\varphi = 1°$, 4.0 s images offset by $\Delta\varphi = 90^{\circ}$ to collect the bijovet pairs, and yielded data useful to 3.5 Å resolution (data not shown).

Structure determination

The phase problem was initially solved by molecular replacement using MR-Rosetta with default settings[22]. As search models, we tried both Chain A of *V. cholerae* DHQS (PDB code 3OKF) and Chain A of *H. pylori* DHQS (PDB code 3CLH) which were the known structures that a BLAST search of the Protein Databank (PDB) showed as having the highest sequence similarity to ValA (33% and 29% identity, respectively). *V. cholerae* DHQS did not yield a solution, but *H. pylori* DHQS gave a result with R/R_{free} of 0.25/0.31 at 2.1 Å resolution and 327 residues built. The electron density map from this solution allowed us to build almost all the side chains, the active-site Zn^{2+}, and the NAD^{+} prosthetic group. In contrast, conventional molecular replacement approaches yielded models with R_{free} near 50% and maps that were very difficult to interpret (data not shown). All manual model building was done in Coot[23]. Refinements at various stages were carried out using Phenix[24] or Buster[25] with TLS refinement, with the final rounds being carried out using Phenix. Water molecules were manually placed based on typical criteria: electron density of $\geq 3\ \rho_{rms}$ in F_o-F_c maps and $\geq 0.8\ \rho_{rms}$ in the $2F_o$-F_c maps and a reasonable potential H-bond partner. Three regions at or near crystallographic two-fold axes were challenging to interpret. The first was a five-residue stretch, residues 46 to 50, that was near and crossing over a crystallographic two-fold axis that was a non-physiological crystal packing interaction. Into this weak helix-like main chain electron density we eventually modeled a portion, residues 48-50 and the side chain of Gln41, at 50% occupancy sharing the space with the same segments from its symmetry mate. As the program would not ignore contact of the side

chain of Gln41 with itself, this side chain position was not allowed to move in the final refinement calculations. We conclude that the segment of only one of the monomers is ordered at a time, and that in solution this part of the protein would be fairly dynamic. The second challenging region centered on a few residues N-terminal to residue 26. These residues had some positive density, but the electron density extended across the same crystallographic two-fold axis noted above and then weakened and we left this small section of density uninterpreted. The third challenging region was a β-hairpin turn (residues 32 to 33) located at the two-fold axis generating the expected physiological dimer interaction. This turn showed weak electron density while the associated β-strands showed strong and clear density. In order to follow the path of the electron density with a single conformation, we modeled the turn with an unfavorable *cis*-peptide bond before Lys32. This model also has a very short non-bonded collision (<2.5 Å) with its own symmetry mate, and so we suspect that it does not represent a true conformation, but only approximates the average chain path associated with a set of multiple conformations that allow reasonable packing at the interface. The final R/R_{free} were 0.179/0.262 with reasonable geometry (Table 1).

Structural comparisons and analyses

Secondary structure assignments were made using DSSP[26, 27] and structure-based sequence alignments were generated using the Dali server[28].

Results and Discussion

Overall structure

The structure of recombinant ValA from *S. hygroscopicus* 5008 presented a challenging molecular replacement problem, as the most similar known structures were only ~30% identical in sequence. The structure determination was greatly facilitated by the MR-Rosetta algorithm[22], which yielded solutions of much higher quality than did conventional molecular replacement (see Materials and Methods). In addition to being aided by MR-Rosetta, the quality of the solution was also enhanced by the inclusion of weak high-resolution data that would have been discarded based on conventional high-resolution cutoff criteria. For the data set used here, the conventional high-resolution cutoff criteria of R_{meas} ~ 60% or $\langle I/\sigma \rangle$ of ~ 2 would lead to limits of 2.85 and 2.3 Å, respectively, whereas the more generous criteria of $CC_{1/2}$ ~ 0.2, shown in recent work to produce better refined models[29-31], leads to a limit of 2.1 Å (Table 1). To test how the inclusion of weak high-resolution data impacted the molecular replacement calculations, we carried out the MR-Rosetta runs using these three justifiable resolution cutoffs. Based on R_{free} values, using the 2.1 Å resolution cutoff yielded the best solution, with the 2.3 Å and 2.85 Å cutoffs being slightly worse and much worse, respectively (Table 2). This example thus shows that weak high-resolution data (out to $CC_{1/2}$ ~0.2 and $\langle I/\sigma \rangle$ ~0.9 in this case) can help with challenging molecular replacement solutions as well as producing better refined models.

Further refinement of the molecular replacement solution yielded a model for the one chain in the asymmetric unit with a final R/R_{free} of 17.9%/26.2% to 2.1 Å resolution (Table 1). The large majority of the main chain as well as an active site NAD^+ and Zn^{2+} are well ordered with strong and clear density, and an absorption scan and anomalous difference map clearly confirm the presence and placement of the active site Zn^{2+} (Figure 2). The final structure

includes 360 of the 414 expected residues, 188 waters, 1 PEG, 1 Zn^{2+}, and 1 NAD^+. The missing residues (1-25, 46-47, 58-62, 244-249, and 399-414) are not modeled due to weak or unclear electron density. Additionally, three sections, including the residues just N-terminal to residue 26, a β-hairpin turn at residues 32-33, and a weakly ordered helix at residues 46-50, laid on or near crystallographic two-fold axes and had weak, ill-formed density, making them challenging to model (see Materials and Methods). A crystallographic two-fold axis brings two ValA chains together to form a dimer that, according to the PISA server[32] buries 4220 $Å^2$ surface area (i.e. 2110 $Å^2$ per monomer). This dimer (Figure 3A) is equivalent to those observed for the homologous enzymes DHQS and DOIS[4, 5, 11, 12, 16, 33] and the dimer interface is well conserved, implying that it is the physiological form of ValA.

Each chain of ValA encompasses the expected N-terminal NAD^+-binding domain and C-terminal metal-binding domain common to the DHQS-like superfamily. We describe here the domain topologies (Figure 3B) using a secondary structure nomenclature that takes into account which elements are conserved among the SPCs (Figure 4). The NAD^+-binding domain has a core 7-stranded β-sheet (with strand order 1-2-9-6-5-3-4) surrounded by five α helices, one β-hairpin (β7 and β8), and two short 3_{10} helices. The metal-binding domain is mainly α helical and includes eight α-helices, one 3_{10}-helix, and one β-hairpin. This domain not only contains the Zn^{2+} binding residues, but based on what has been seen in DHQS and DOIS, it also contains the majority of the residues involved in substrate recognition and so has also been called the substrate-binding domain[5]. However, the sugar-phosphate substrate actually binds in a cleft between the two domains and its recognition involves residues from both domains.

Relationships to other structurally known proteins

A structural similarity search performed using the DALI server[28] showed that ValA is most similar to assorted DHQSs (rmsd ~2.2 – 2.6 Å, Z-scores ~35-45), followed by DOIS (rmsd ~2.2 Å, Z-score 34) and then various polyol-DHs (rmsd ~2.7 – 3.6 Å, Z-scores ~23-28). Interestingly, although the structures used as search models in molecular replacement were the two PDB structures (PDB code 3OKF and PDB code 3CLH) most sequence similar to ValA, in terms of structural similarity these rank only eighth (Z-score = 41) and twenty-first (Z-score =36), respectively. This underscores why it can be useful to try all potential homologs in molecular replacement rather than just the most sequence similar ones[34]. The DALI search further showed that there are no known protein structures outside of the DHQS-like superfamily that share noteworthy structural similarity to ValA or to either of its individual domains.

Based on these results, representative enzymes were chosen for a structure-based sequence alignment (Figure 4): *Aspergillus nidulans* DHQS (*An*DHQS; PDB code 1DQS), the most well-studied DHQS[4]; *Bacillus circulans* DOIS (*Bc*DOIS; PDB code 2D2X), the only structurally known DOIS[5]; and *Bacillus stearothermophilus* glycerol dehydrogenase (*Bs*GlyDH; PDB code 1JQ5), the most structurally similar member of the polyol-DH family[13]. Representatives from the two structurally unknown types of the SH7PCs were also included in this sequence alignment: a DDGS from *Anabaena variabilis* (*Av*DDGS) and an EVS from *Actinosynnema mirum* (*Am*EVS). The structure-based alignment between ValA and DHQS is largely consistent with alignments that led to the previously proposed putative active site residues in ValA [6, 7]. The only change is that Lys356 in *An*DHQS had been previously aligned with Pro370 in ValA, but the structure-based alignment identifies the equivalent residue as His360.

Zinc and NAD^+ binding

Although zinc and NAD^+ were not added during sample preparation or crystallization, the electron density maps showed their unambiguous presence in the crystal structure (Figure 2), presumably meaning that they were bound by ValA already in the *E. coli* cytosol and carried along during the purification. The zinc and NAD^+ are both bound quite similarly to what has been described for DHQS[4] and DOIS[5] and so aside from some novel observations, we will here only briefly summarize the features of the binding. All residues in direct contact with zinc and NAD^+ are designated in Figure 4, and most are conserved among the SPCs. The Zn^{2+} ion is coordinated by Glu213, His284, and His300, all from the C-terminal metal-binding domain, and a water. As inferred from the liganded structures of DHQS and DOIS, this active site water will be displaced upon substrate binding.

The binding of NAD^+ includes characteristic residues conserved among the SPCs such as Asp70 at the end of strand β3 that H-bonds with the adenosine ribose O2' hydroxyl, the glycine-rich turn connecting β5 and H3 providing backbone amides that H-bond to the pyrophosphate oxygens, and Glu101, Lys104, Lys180, and Asn181 that H-bond with the nicotinamide ribose hydroxyls. The nicotinamide amide nitrogen donates H-bonds to the Asp138 side chain and the Lys171 backbone oxygen, and the oxygen forms H-bonds with surrounding waters in this structure. A fascinating pair of interactions that is conserved in known DHQS-like superfamily structures but has not been described before places carboxylate oxygens from Asp138 and Asp165 each roughly in the plane of the nicotinamide ring where they are respectively in position to accept weakly polar H-bonds from the nicotinamide C2 and C4 carbons (see Figure 2). We expect that these interactions will preferentially stabilize the positive charge on the

oxidized form of the nicotinamide ring, which is distributed between the ring carbon atoms 2, 4, and 6 via resonance forms. These interactions may help explain the high affinity of these enzymes for NAD^+, and especially the interaction with the reactive nicotinamide C4 position may play a role in modulating the nicotinamide redox properties during the catalytic cycle.

The ValA sugar phosphate binding site

Although no substrate or substrate analog is bound in this crystal form of ValA, we can still gain insight into its substrate binding by comparisons with the ligand-bound structures of *An*DHQS and *Bc*DOIS structures (henceforth just referred to as DHQS and DOIS, respectively). For convenience, we refer to sequence differences between ValA and DHQS as mutations or changes with respect to DHQS, even though ValA did not evolve from a modern DHQS. As noted in the introduction, DHQS undergoes a conformational change from 'open' in the absence of a sugar-phosphate ligand to 'closed' upon binding the substrate analog CBP via a domain rotation of ~12° that brings the N and C-terminal domains closer together[8]. A recent structure of DHQS from *Actinidia chinensis* reinforces the relevance of the closed conformation seen, as the same closed conformation appears to be stabilized by the binding of inorganic phosphate and glycine in a way that mimics the substrate[35]. DOIS, in contrast, was reported to not undergo such a domain closure based on comparisons of its structures with or without a substrate analog[5].

A set of overlays of ValA with representative liganded and unliganded forms of DHQS and DOIS show that our ValA structure has a conformation between the open and closed DHQS forms but much closer to the closed form, varying by only ~2° (Figure 5). They further show that both the unliganded and liganded DOIS structures do indeed have minimal differences in their

domain orientations, but that the DOIS conformation is ~10° more open in chain A and ~7.5° more open in chain B than the DHQS closed structure (data not shown), suggesting that it might not accurately represent the ligand bound structure. Also supporting this possibility is that the DOIS liganded structure was obtained by soaking crystals of the unliganded enzyme with inhibitor, during which some crystal cracking was observed[5]. This implies that the enzyme could not undergo a complete domain closure without compromising the integrity of the crystal. For this reason, we focus in the following comparisons solely on the DHQS·CBP complex, which by all evidence accurately represents a true inhibitor-bound conformation.

With only this unliganded structure of ValA available, we cannot make any claims about what ligand-induced conformational changes may occur. However, the similarity of ValA to the closed conformation of DHQS is fortunate as it means that the NAD^+, the zinc, and nearly all of the ValA residues equivalent to DHQS active site residues align rather well (Figure 6), giving us confidence that this comparison provides an informative picture of which ValA residues will play a role in substrate binding. Of the ligand-binding residues in the DHQS complex, only one, Arg264, is not in close proximity to its corresponding residue in ValA. The equivalent residue in ValA is Arg277 (Figure 4), and it points in the opposite direction (see green side chains in Figure 5). Interestingly, Arg277 is not conserved among EEVSs (data not shown), suggesting it is not a key residue for this enzyme. Even more interestingly, due to a different nearby loop conformation, the following residue in ValA, Arg278, has its side chain close to that of DHQS Arg264 (Figure 6) and is conserved among EEVSs, suggesting that it may be the functionally equivalent residue. An important question then becomes if the different loop conformation is a robust difference between ValA and DHQS, or if it may be simply due to the ValA structure not having a ligand bound.

A closer look at the loop (residues 257-264 in DHQS and 270-278 in ValA) identifies another key active site position and confirms that the difference in loops is robust (Figure 7). In both unliganded and liganded DHQS structures, the loop wraps around the side chain of Asp257 which accepts multiple backbone amide H-bonds to stabilize the conformation. The equivalent residue in ValA, Asn270, is not compatible with the DHQS loop conformation, but plays an equally central role in stabilizing the alternate less-compact loop path (Figure 7).

Taking ValA Arg278 as the equivalent of DHQS Arg264, Figure 6 compares the DHQS residues surrounding the substrate analog CBP with their ValA equivalents. Among these, just four ValA residues are different types: Met263 replaces a Lys, Asp281 replaces an Asn, Pro288 replaces a His, and His360 replaces a Lys. Using atom numbering for the substrate (see Figure 1) rather than the CBP inhibitor, the essential features of CBP binding to DHQS (clockwise from the top of Figure 6) are a phosphate binding pocket (at two o'clock), bidentate metal coordination by the C5 and C4 hydroxyls (at three to six o'clock) which also serves to point the C5-hydrogen at the nicotinamide C4 atom in good geometry for hydride transfer, and then a pocket for the C2 hydroxyl and carboxylate groups (around ten o'clock). In ValA, the metal and nicotinamide are nearly identically positioned as are key residues interacting with the phosphate (Lys171, Arg149', and Asn181) and the metal-coordinating hydroxyls (Asp165 and Lys216). We take this to mean that the analogous parts of the ValA substrate will be bound similarly to CBP in DHQS. Having a high spatial conservation of these parts of the substrate makes sense, as they are where most of the chemistry takes place.

In contrast, the significantly shifted or mutated residues (Arg278, Met263 and Asp281) are present at the pocket around the C2 hydroxyl and carboxylate groups where the ValA substrate has different substituents. In particular, the Lys→Met change makes sense with the

absence of the substrate carboxylate. The two remaining changes, involving ValA residues Pro288 and His360, create more space around the phosphate group, but we do not understand why that might be. The DHQS residue replaced by Pro288 is His275, which has been proposed to serve as an acid/base during catalysis[4].

Based on the above, we conclude that ValA will bind its substrate, SH7P, with the phosphate group and the C5 and C4 metal-coordinating hydroxyls in similar positions to those in the DHQS·CBP complex. However, other aspects of the binding mode such as ring conformation and/or orientation must differ from CBP, especially because the configuration of the C4 hydroxyl in SH7P differs (see Figure 1B) such that it and the C5 hydroxyl cannot simultaneously be equatorial. This difference makes predicting details of the binding mode of SH7P more challenging.

Variations among the SH7PCs and a proposal for how they catalyze different reactions

As noted in the introduction, a major open question about SH7PC enzymes is how they bind the same substrate and produce different products (Figure 1B), with the most conceptually confusing aspect being how EEVS produces one stereochemistry at the C5 position of the product (derived from the substrate C2 atom; as shown in Figure 1B) while EVS produces the other[6]. Because of an internal symmetry in the DDGS product, it could be produced with either C5 stereochemistry (Figure 1B). Interestingly, EEVS and DDGS are more sequence similar to each other than they are to EVS, with both being reported to vary from DHQS in the identities of four putative active site residues while EVS varied in only one[6, 7] (Figure 8). Our structural results on ValA (an EEVS) strengthen this pattern, in that the Asn residue associated with its distinct loop conformation and the alternate active site Arg residue (Figure 7) are also both

present in DDGSs (Figure 4). In contrast, EVS, like DHQS, has an Asp in the loop and conserves the first Arg (Figure 4). This implies that these two subsets of the SH7PCs (EEVS and DDGS vs. EVS) have distinct binding environments for the substituents of the substrate C2 atom, and leads us to hypothesize that the discrimination related to the stereochemistry at C5 of the product actually occurs upon substrate binding rather than during catalysis.

Specifically, each enzyme would selectively bind either the α- or β-pyranose form of SH7P and the rapid interconversion between the various pyranose and furanose forms would allow these normally less-populated SH7P forms (~16% and <1%, respectively[36]) to accumulate in the enzyme active sites (Figure 9A). In this way, rather than the aldol acceptor geometry variation between EEVS and EVS requiring a 180° rotation of the polar C5 substituents in the middle of the catalytic cycle, each enzyme would start with a substrate preorganized to generate the correct configuration at that position (Figure 9B). As EEV has the configuration at C5 that would be derived from α-pyranose SH7P, this is the form of substrate that EEVS and DDGS would bind, and as EV has the same configuration at C5 as would be derived from β-pyranose SH7P, the EVS active site would select for that form of the substrate. This hypothesis is completely consistent with the behavior of DHQS and DOIS, which both bind their substrate analogs with a fixed configuration corresponding to that of their product. In the case of DOIS, the enzyme's selectivity is directly observed in its preferential binding from a racemic mixture of the inhibitor CG6P only the form that mimics the β-anomer of substrate[5, 37].

In this way, rather than viewing the SH7PCs as a family of enzymes which use one substrate to specifically give two stereochemically distinct types of products, one can then view them as enzymes that bind distinct substrates, either α-pyranose or β-pyranose SH7P, to give products with stereochemistries matching those substrates. In terms of the evolutionary origin of

the SH7PCs, since EVS sequences are more similar to those of DHQSs[6], we suggest that EVS evolved from DHQS with only slight changes to the active site being needed to allow binding of the β-pyranose anomer of SH7P. In contrast, the key differentiating step in EEVS evolution would have been the Asp→Asn mutation that changed which Arg pointed into the active site pocket, as this allowed binding of the α-pyranose anomer of SH7P leading to the production of EEV. Then further mutations, including two active site mutations of Asp →Ala and His→Thr (Figures 4 and 8), gave rise to the mechanistic difference that characterizes the DDGS enzymes. Although we do not understand this transition yet, the more minor nature of differences between the DDGSs and the EEVSs is emphasized by phylogenetic trees showing that EEVS and DDGS are more closely related to each other than to the other SPCs [6].

Although much remains to be learned about ValA and the SH7PCs in general, the ValA structure presented here sheds much light on this enzyme family. Especially seeing the spatial orientation of active site residues (Figure 8) has provided insight into how variation in the active site pocket allows for the different specificities found in the SPC superfamily. Nevertheless, the detailed roles of catalytic residues in the EEVS, DDGS and EVS mechanisms still remain open questions. For example, the residue which takes the catalytic role of an essential histidine in DHQS (His275) and DOIS (His250) has not been identified in these enzymes. We expect that answering such questions for the SH7PCs will best be approached through kinetics and structural studies carried out with stereospecific carbacyclic phosphonate analogs of SH7P that do not yet exist, but that are in concept similar to the DAHP analogs used in informative studies of DHQS [4, 8, 38, 39]. We are now initiating studies in this direction.

Coordinates

The coordinates and the structure factors have been deposited in the Protein Data Bank (www.rcsb.org/pdb/) as entry 4P53.

Acknowledgements

We would like to thank Dale Tronrud and Camden Driggers for help with the crystallographic work, and Murugesh Padmanarayan for growing and analyzing the first crystals.

References

1. T. Mahmud. (2003) The C7N aminocyclitol family of natural products, *Nat. Prod. Rep. 20*, 137-166.

2. A. Stratmann, T. Mahmud, S. Lee, J. Distler, H. G. Floss, and W. Piepersberg. (1999) The AcbC protein from Actinoplanes species is a C-7-cyclitol synthase related to 3-dehydroquinate synthases and is involved in the biosynthesis of the alpha-glucosidase inhibitor acarbose, *J. Biol. Chem. 274*, 10889-10896.

3. Y. Yu, L. Q. Bai, K. Minagawa, X. H. Jian, L. Li, J. L. Li, S. Y. Chen, E. H. Cao, T. Mahmud, H. G. Floss, X. F. Zhou, and Z. X. Deng. (2005) Gene cluster responsible for validamycin biosynthesis in Streptomyces hygroscopicus subsp jinggangensis 5008, *Appl. Environ. Microbiol. 71*, 5066-5076.

4. E. P. Carpenter, A. R. Hawkins, J. W. Frost, and K. A. Brown. (1998) Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis, *Nature 394*, 299-302.

5. E. Nango, T. Kumasaka, T. Hirayama, N. Tanaka, and T. Eguchi. (2008) Structure of 2-deoxy-scyllo-inososе synthase, a key enzyme in the biosynthesis of 2-deoxystreptamine-containing aminoglycoside antibiotics, in complex with a mechanism-based inhibitor and NAD(+), *Proteins: Struct., Funct., Bioinf. 70*, 517-527.

6. S. Asamizu, P. F. Xie, C. J. Brumsted, P. M. Platt, and T. Mahmud. (2012) Evolutionary Divergence of Sedoheptulose 7-Phosphate Cyclases Leads to Several Distinct Cyclic Products, *J. Am. Chem. Soc. 134*, 12219-12229.

7. X. Wu, P. M. Flatt, O. Schlorke, A. Zeeck, T. Dairi, and T. Mahmud. (2007) A comparative analysis of the sugar phosphate cyclase superfamily involved in primary and secondary metabolism, *ChemBioChem 8*, 239-248.

8. C. E. Nichols, J. Ren, H. K. Lamb, A. R. Hawkins, and D. K. Stammers. (2003) Ligand-induced conformational changes and a mechanism for domain closure in Aspergillus nidulans dehydroquinate synthase, *J. Mol. Biol. 327*, 129-144.

9. C. E. Nichols, A. R. Hawkins, and D. K. Stammers. (2004) Structure of the 'open' form of Aspergillus nidulans 3-dehydroquinate synthase at 1.7 A resolution from crystals grown following enzyme turnover, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 971-973.

10. C. E. Nichols, J. Ren, K. Leslie, B. Dhaliwal, M. Lockyer, I. Charles, A. R. Hawkins, and D. K. Stammers. (2004) Comparison of ligand-induced conformational changes and domain closure mechanisms, between prokaryotic and eukaryotic dehydroquinate synthases, *J. Mol. Biol. 343*, 533-546.

11. M. Sugahara, Y. Nodake, M. Sugahara, and N. Kunishima. (2005) Crystal structure of dehydroquinate synthase from Thermus thermophilus HB8 showing functional importance of the dimeric state, *Proteins: Struct., Funct., Bioinf. 58*, 249-252.

12. J. S. Liu, W. C. Cheng, H. J. Wang, Y. C. Chen, and W. C. Wang. (2008) Structure-based inhibitor discovery of Helicobacter pylori dehydroquinate synthase, *Biochem. Biophys. Res. Commun. 373*, 1-7.

13. S. N. Ruzheinikov, J. Burke, S. Sedelnikova, P. J. Baker, R. Taylor, P. A. Bullough, N. M. Muir, M. G. Gore, and D. W. Rice. (2001) Glycerol dehydrogenase. structure, specificity, and mechanism of a family III polyol dehydrogenase, *Structure 9*, 789-802.

14. A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol. 247*, 536-540.

15. I. Sillitoe, A. L. Cuff, B. H. Dessailly, N. L. Dawson, N. Furnham, D. Lee, J. G. Lees, T. E. Lewis, R. A. Studer, R. Rentzsch, C. Yeats, J. M. Thornton, and C. A. Orengo. (2013) New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures, *Nucleic Acids Res. 41*, D490-498.

16. A. Park, H. K. Lamb, C. Nichols, J. D. Moore, K. A. Brown, A. Cooper, I. G. Charles, D. K. Stammers, and A. R. Hawkins. (2004) Biophysical and kinetic analysis of wild-type and site-directed mutants of the isolated and native dehydroquinate synthase domain of the AROM protein, *Protein Sci. 13*, 2108-2119.

17. G. J. Bartlett, N. Borkakoti, and J. M. Thornton. (2003) Catalysing new reactions during evolution: economy of residues and mechanism, *J. Mol. Biol. 331*, 829-860.

18. J. M. Winter, and Y. Tang. (2012) Synthetic biological approaches to natural product biosynthesis, *Curr. Opin. Biotechnol. 23*, 736-743.

19. T. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. Leslie. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 271-281.

20. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 235-242.

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

22. T. C. Terwilliger, F. Dimaio, R. J. Read, D. Baker, G. Bunkoczi, P. D. Adams, R. W. Grosse-Kunstleve, P. V. Afonine, and N. Echols. (2012) phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta, *J. Struct. Funct. Genomics 13*, 81-90.

23. P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. (2010) Features and development of Coot, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 486-501.

24. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 213-221.

25. G. Bricogne, Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P, Sharff, A., Smart, O.S., Vonrhein, C., Womack, T.O. (2011) BUSTER version 2.11.2, Global Phasing Ltd, Cambridge, United Kingdom.

26. W. Kabsch, and C. Sander. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577-2637.

27. R. P. Joosten, T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander, and G. Vriend. (2011) A series of PDB related databases for everyday needs, *Nucleic Acids Res. 39*, D411-419.

28. L. Holm, and P. Rosenstrom. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res. 38*, W545-549.

29. P. A. Karplus, and K. Diederichs. (2012) Linking crystallographic model and data quality, *Science 336*, 1030-1033.

30. N. J. Kershaw, J. M. Murphy, N. P. Liau, L. N. Varghese, A. Laktyushin, E. L. Whitlock, I. S. Lucet, N. A. Nicola, and J. J. Babon. (2013) SOCS3 binds specific receptor-JAK complexes to control cytokine signaling by direct kinase inhibition, *Nat. Struct. Mol. Biol. 20*, 469-476.

31. A. Perkins, M. C. Gretes, K. J. Nelson, L. B. Poole, and P. A. Karplus. (2012) Mapping the active site helix-to-strand conversion of CxxxxC peroxiredoxin Q enzymes, *Biochemistry 51*, 7638-7650.

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

33. K. A. Brown, E. P. Carpenter, K. A. Watson, J. R. Coggins, A. R. Hawkins, M. H. J. Koch, and D. I. Svergun. (2003) Twists and turns: a tale of two shikimate-pathway enzymes, *Biochem. Soc. Trans. 31*, 543-547.

34. I. Stokes-Rees, and P. Sliz. (2010) Protein structure determination by exhaustive search of Protein Data Bank derived databases, *Proc. Natl. Acad. Sci. U. S. A. 107*, 21476-21481.

35. G. Mittelstadt, L. Negron, L. R. Schofield, K. Marsh, and E. J. Parker. (2013) Biochemical and structural characterisation of dehydroquinate synthase from the New Zealand kiwifruit Actinidia chinensis, *Arch. Biochem. Biophys. 537*, 185-191.

36. F. Charmantray, V. Helaine, B. Legeret, and L. Hecquet. (2009) Preparative scale enzymatic synthesis of D-sedoheptulose-7-phosphate from beta-hydroxypyruvate and D-ribose-5-phosphate, *J. Mol. Catal. B: Enzym 57*, 6-9.

37. E. Nango, T. Eguchi, and K. Kakinuma. (2004) Active site mapping of 2-deoxy-scyllo-inososе synthase, the key starter enzyme for the biosynthesis of 2-deoxystreptamine. Mechanism-based inhibition and identification of lysine-141 as the entrapped nucleophile, *J. Org. Chem. 69*, 593-600.

38. S. L. Bender, T. Widlanski, and J. R. Knowles. (1989) Dehydroquinate synthase: the use of substrate analogues to probe the early steps of the catalyzed reaction, *Biochemistry 28*, 7560-7572.

39. J. G. Montchamp. Mechanism-based carbocyclic inhibitors of dehydroquinate synthase. Ph.D. Dissertation, Purdue University, West Lafayette, IN, 1992.

40. J. D. Moore, M. A. Skinner, D. R. Swatman, A. R. Hawkins, and K. A. Brown. (1998) Reactivation of 3-dehydroquinate synthase by lanthanide cations, *J. Am. Chem. Soc. 120*, 7105-7106.

41. F. Kudo, Y. Hosomi, H. Tamegai, and K. Kakinuma. (1999) Purification and characterization of 2-deoxy-scyllo-inososе synthase derived from Bacillus circulans. A crucial

carbocyclization enzyme in the biosynthesis of 2-deoxystreptarnine-containing aminoglycoside antibiotics, *J. Antibiot. 52,* 81-88.

Table 1. Data collection and refinement statistics[a]

A. Data	
Resolution limits (Å)	66.9 - 2.10 (2.21-2.10)
Unique observations	20232 (2875)
Multiplicity	26.8 (19.3)
Completeness (%)	99.4 (98.9)
Average I/σ	11.2 (0.9)
R_{meas} (%)	38 (676)
$CC_{1/2}$ (%)	0.99 (0.22)
B. Refinement	
No. residues	360
No. solvent atoms	188
No. atoms total	3004
<B> protein ($Å^2$)	48
<B> solvent ($Å^2$)	56
R_{cryst} (%)	17.9 (28.1)
R_{free} (%)	26.2 (34.7)
rmsd bonds (Å)	0.010
rmsd angles (˚)	1.28

[a] Numbers in parentheses correspond to values in the highest resolution bin.

Table 2. Resolution dependence of MR-Rosetta results

Resolution (Å)	**R**	**R_{free}**	**Residues Built**[a]
2.85	0.30	0.40	306 (221)
2.30	0.26	0.33	334 (318)
2.10	0.25	0.31	327 (298)

[a] Gives total number of residues built in backbone segments and, in parentheses, the number of these modeled as specific residues in the sequence of the target structure

Figure legends

Figure 1. **Reactions catalyzed by known sugar phosphate cyclases. (A)** Four cyclitol-containing natural products are shown and labeled by name, with their C7-cyclitol units made by SH7PCs highlighted in bold. **(B)** The substrates (above) and products (below) of five sugar phosphate cyclases are shown. A divalent metal cation is drawn next to the two substrate hydroxyls seen (for DHQS and DOIS) to coordinate it. SH7PCs and DHQS may utilize Zn^{2+} or Co^{2+} as the divalent metal cation[6, 40] but are shown here with Zn^{2+}, the metal present in this structure of ValA and in structures of *An*DHQS. DOIS exclusively uses Co^{2+} [41] and is depicted as such. Abbreviations for the substrates and products are introduced, with each enzyme abbreviation being that of its product followed by an additional S for "synthase." Due to resonance, DDG has an internal symmetry so the stereoconfiguration at C5 of the product after it is released into solution is not uniquely defined. We draw it here with the same stereochemistry as EEV anticipating the proposal we make in this work that the cyclization products of EEVS and DDGS have the same C5 configuration.

Figure 2. **Electron density map quality and active site structure.** Stereoview of the ValA active site residues (purple carbons) and a water (red sphere) that are near the NAD^+ (grey carbons) and the Zn^{2+} (silver sphere) cofactors. Coordination bonds (black lines) and select H-bonds (black dashes) are shown along with the $2F_o$- F_c electron density (orange; contoured at 1 ρ_{rms}) and an anomalous difference map (green; contoured at 12 ρ_{rms}).

Figure 3. **Overall structure and topology of ValA. (A)** Ribbon diagrams of the two chains of the ValA dimer are shown in purple and green tones, respectively, with the N-terminal NAD^+-

binding domains in light hues and the C-terminal metal-binding domains in dark hues. Dashed lines indicate internal unmodeled backbone segments. The NAD^+ and the Zn^{2+} with its coordinating ligands are shown (colored as in Figure 1). Secondary structural elements in each domain of one monomer are labeled. **(B)** Topology diagram showing α-helices (cylinders), β-stands (arrows), 3_{10} helices (triangular prisms), and π-helices (wider cylinder) with their respective first and last residues given. The minimal length α- and 3_{10} helices (5 and 3 residues, respectively) are left out of the family secondary structure nomenclature. The domains are colored light and dark purple as indicated, and helices (H) and strands (β) common to the SPCs are named sequentially within each domain. Dashed lines denote unmodeled backbone segments. The three Zn^{2+} binding residues (red asterisks) and the glycine-rich turn and acidic residue (green asterisks) important for NAD^+-binding are indicated.

Figure 4. **Sequence alignment of ValA with representative related enzymes.** The sequence of ValA is listed first and its secondary structure elements are schematically shown above the sequence. Other sequences in order are *Av*DDGS (*A. variabilis* DDGS; Ava_3858), *Am*EVS (*A. mirum* EVS; Amir_2000), 1DQS (*A. nidulans* DHQS), 2D2X (*B. circulans* DOIS), and 1JQ5 (*B. stearothermophilus* glycerol dehydrogenase). For the structurally known proteins, the residues in β-strands (yellow), α-helices (teal), 3_{10} helices (blue), and π helices (orange) are highlighted. Residues involved in metal-binding (m), NAD^+-binding (n), and substrate binding and/or catalysis (*) are indicated below the sequences, and active site residues with notable variation (↓) are indicated above the sequences.

Figure 5. **ValA overlay with closed and open DHQS structures.** Shown are ribbon diagrams of an unliganded, open DHQS (blue; PDB code 1NRX), a CBP-bound, closed DHQS (cyan; PDB code 1DQS) and ValA (purple), all overlaid based on their NAD^+-binding domains. Pale hues are used for the NAD^+-binding domains and NAD^+, and dark hues are used for the metal-binding domains. For clarity, only select secondary structure elements of the metal binding domain are shown along with the three zinc ligating residues (sticks) and the zinc (grey sphere). The active site side chain that does not align well between DHQS (Arg264) and ValA (Arg277) is shown for both structures in green, and the alternate equivalent ValA residue (Arg278) is shown in orange. That other 'open' DHQS structures, such as those of *Tt*DHQS (PDB code 1UJN) and *Hp*DHQS (PDB code 3CLH), are up to 5° different in domain orientation compared with *An*DHQS (PDB code 1NRX) does not alter the conclusions of this analysis.

Figure 6. **Comparing the ValA active site region with the DHQS·CBP complex.** Stereoview of select active site residues in ValA (purple) overlaid on the DHQS (cyan) in complex with CBP (white) shown in roughly the same orientation as DAHP is drawn in Figure 1. H-bonding interactions in the DHQS active site (dashed lines) and coordination bonds with Zn^{2+} (solid lines) are shown. A prime on a residue number means it is from the other subunit of the dimer.

Figure 7. **Active site loop difference relates to presence of Asp257 in DHQS *vs.* Asn270 in ValA.** Shown are residues 269-278 of ValA (purple) and 256-264 of DHQS (cyan; PDB code 1DQS) after the proteins have been overlaid as in Figure 5. H-bonding interactions (dashed lines) involving the loop residues and associated waters are shown. In DHQS, the Asp257 carboxylate receives H-bonds directly or indirectly (via water) from four backbone nitrogens (from Arg259,

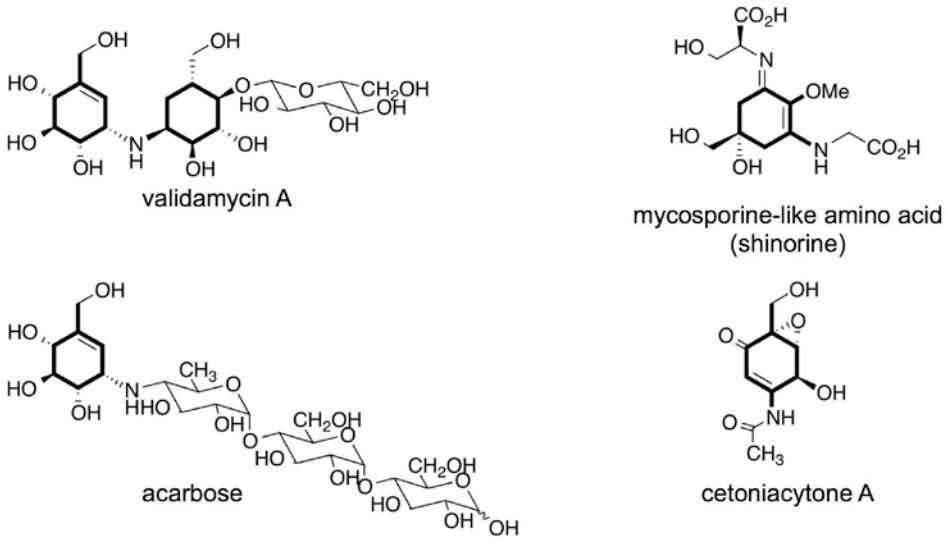
Gly261, Gly262, and Arg264). In ValA, the Asn270 side chain amide directly or indirectly makes H-bonds with two backbone nitrogens (from Trp272 and Glu275) and two backbone oxygens (from Gln275 and Arg277).

Figure 8. **Variation in active site residues among sugar phosphate cyclases.** Schematic drawing of residues lining the substrate binding pocket in the DHQS·CBP complex shown in roughly the same orientation as in Figure 6. Each DHQS residue shown is labeled (cyan) and under that label are listed the corresponding residues found in the structures of DOIS (pink) and EEVS (purple), or based on the alignment in Figure 4, residues expected to be equivalent in DDGS and EVS (black). The CBP ligand is bold. H-bonding interactions (dashed lines) and coordination bonds with Zn^{2+} (solid lines) are shown. Residue numbering corresponds to the representative proteins used in Figure 4 (*An*DHQS, *Bc*DOIS, ValA, *Av*DDGS, *Am*EVS).

Figure 9. **Proposed enzyme-specific selection of forms of sedoheptulose 7-phosphate. (A)** The α and β pyranose and furanose anomers of SH7P with their relative abundance as determined by NMR[36] are shown along with the linear form via which they interconvert. Also shown is our proposal that EEVS and DDGS bind the α-pyranose anomer while EVS binds the β-pyranose anomer. **(B)** The ring opening and intramolecular aldol condensation steps of the proposed reaction mechanisms of EEVS and EVS are shown, emphasizing how the α- and β-pyranose anomers of SH7P are preorganized for the generation of the respective stereochemistries at C5 in the products. "B:" represents an active site base that may aid ring opening.

Figure 1.

(A)


validamycin A
mycosporine-like amino acid
(shinorine)
acarbose
cetoniacytone A

(B)

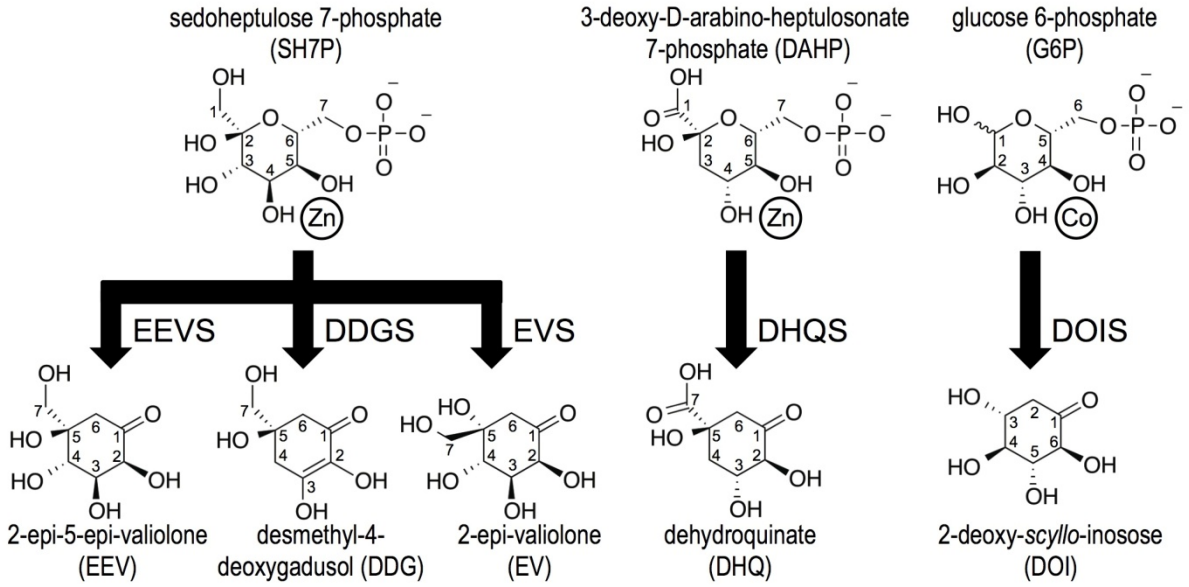

sedoheptulose 7-phosphate
(SH7P)
3-deoxy-D-arabino-heptulosonate
7-phosphate (DAHP)
glucose 6-phosphate
(G6P)
EEVS
DDGS
EVS
DHQS
DOIS
2-epi-5-epi-valiolone
(EEV)
desmethyl-4-
deoxygadusol (DDG)
2-epi-valiolone
(EV)
dehydroquinate
(DHQ)
2-deoxy-scyllo-inosose
(DOI)

Figure 2

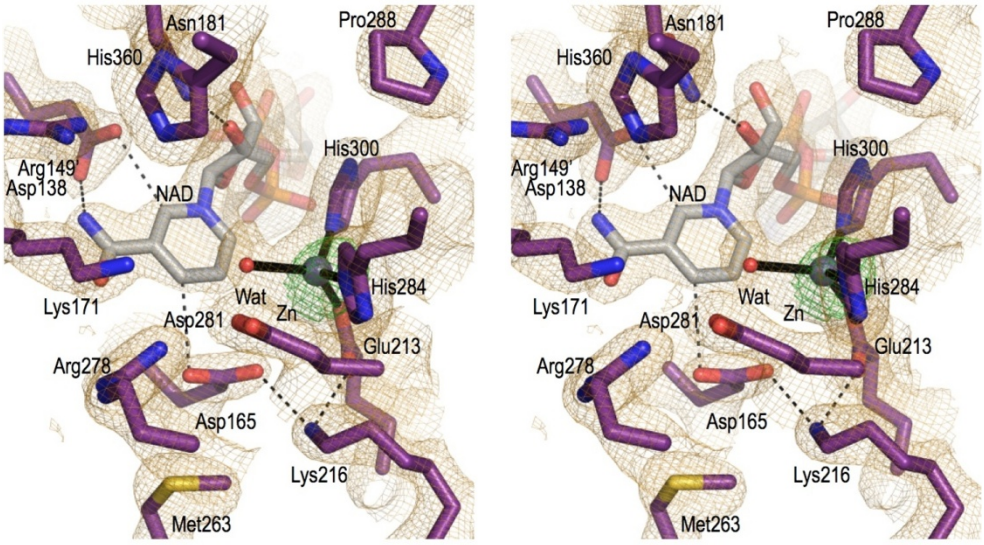
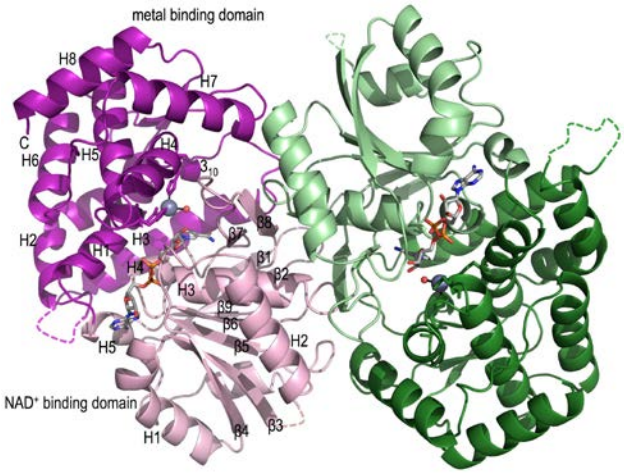

Asn181
Pro288
His360
His300
Arg149'
Asp138
NAD
Lys171
Wat
Zn
His284
Asp281
Glu213
Arg278
Asp165
Lys216
Met263
Asn181
Pro288
His360
His300
Arg149'
Asp138
NAD
Lys171
Wat
Zn
His284
Asp281
Glu213
Arg278
Asp165
Lys216
Met263

Figure 3

(A)


metal binding domain
NAD⁺ binding domain
H8
H7
H6
H5
H4
H3
H2
H1
β1
β2
β3
β4
β5
β6
β7
β8
β9

(B)

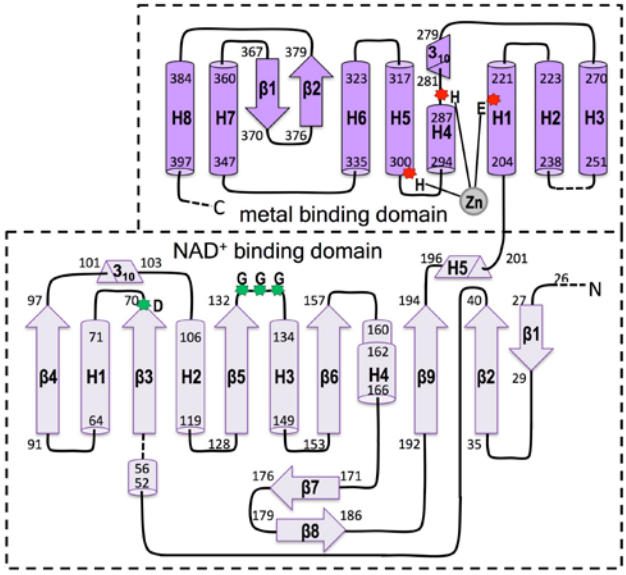

metal binding domain
NAD⁺ binding domain
H8
H7
H6
H5
H4
H3
H2
H1
β1
β2
β3
β4
β5
β6
β7
β8
β9
Zn
C
N

Figure 4

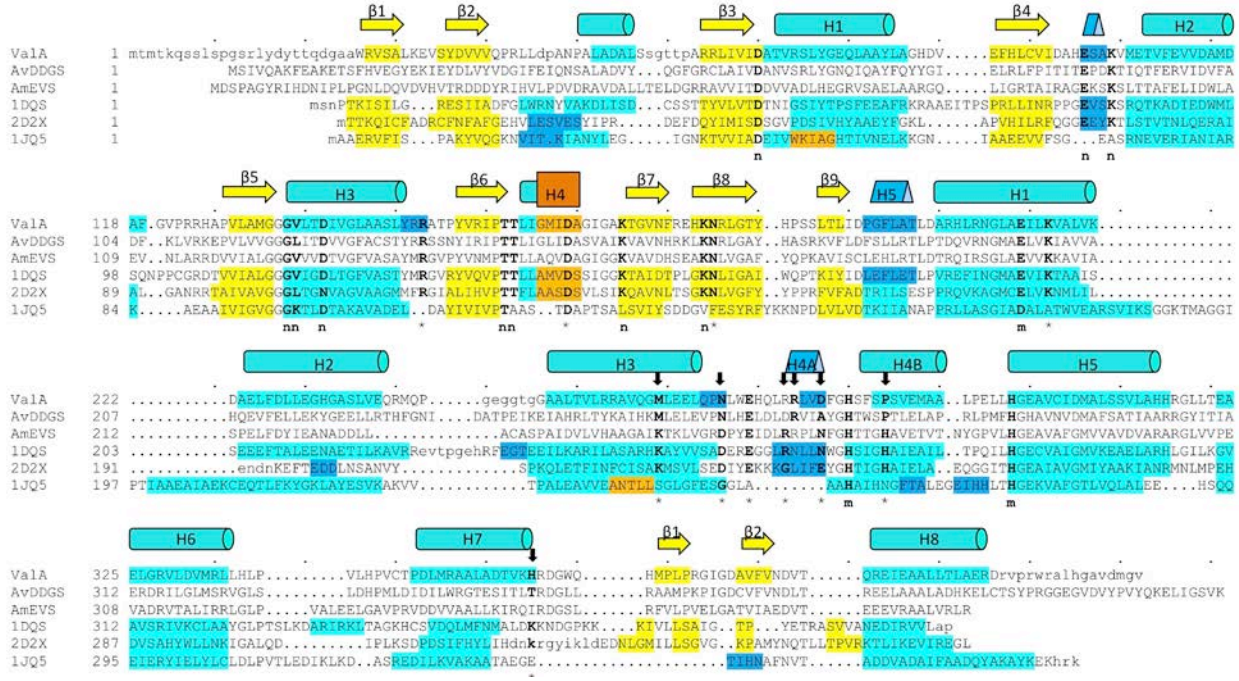


Figure 5

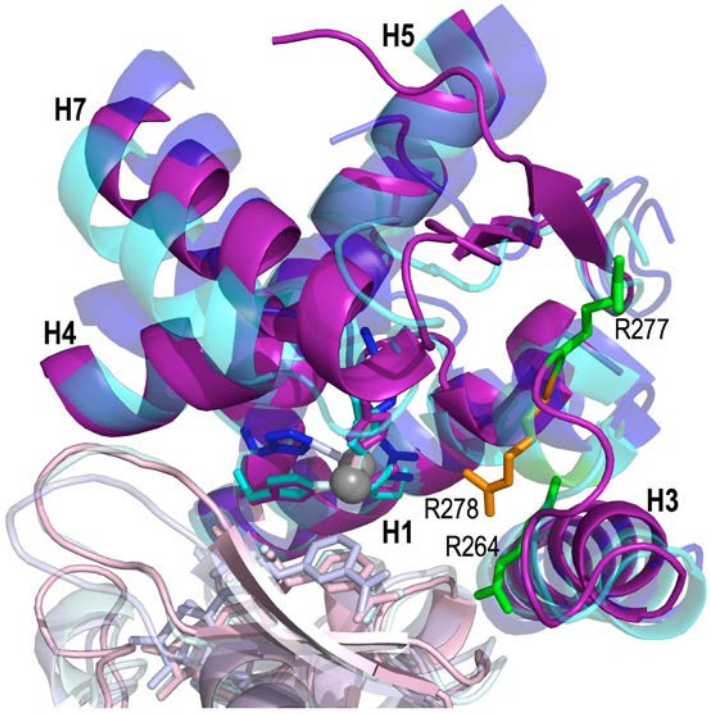


Figure 6

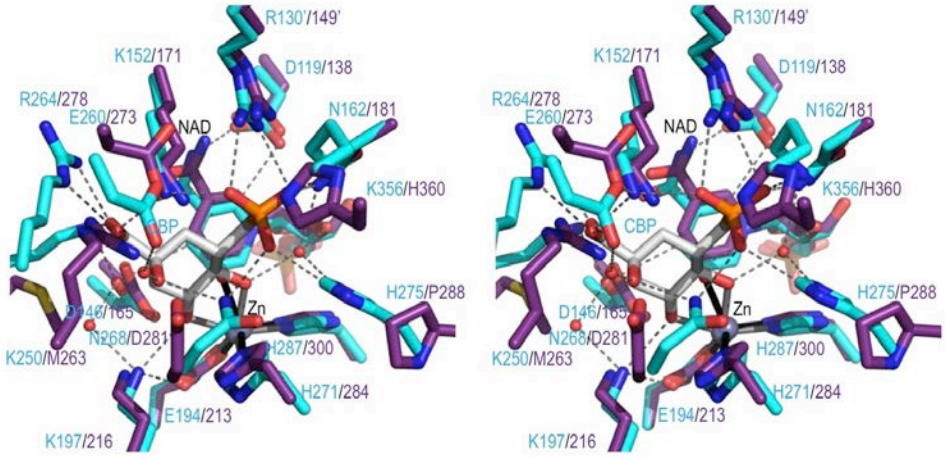


Figure 7

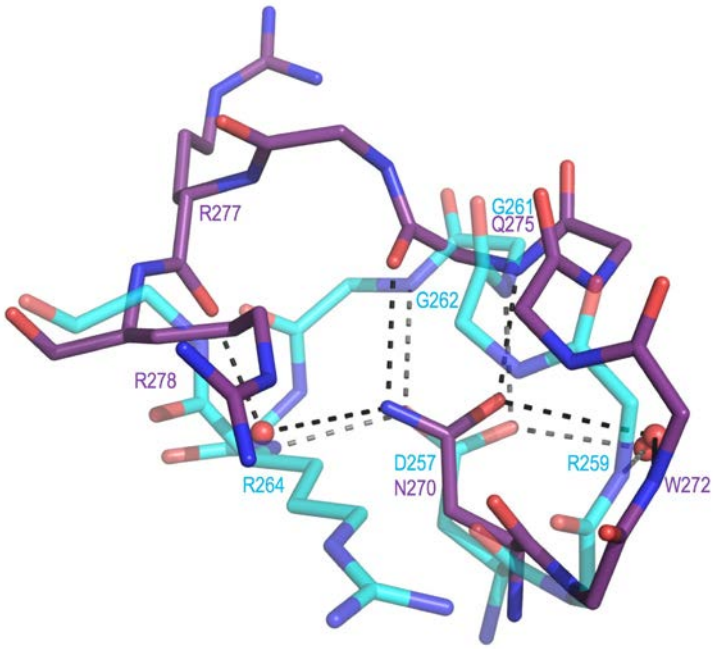


Figure 8

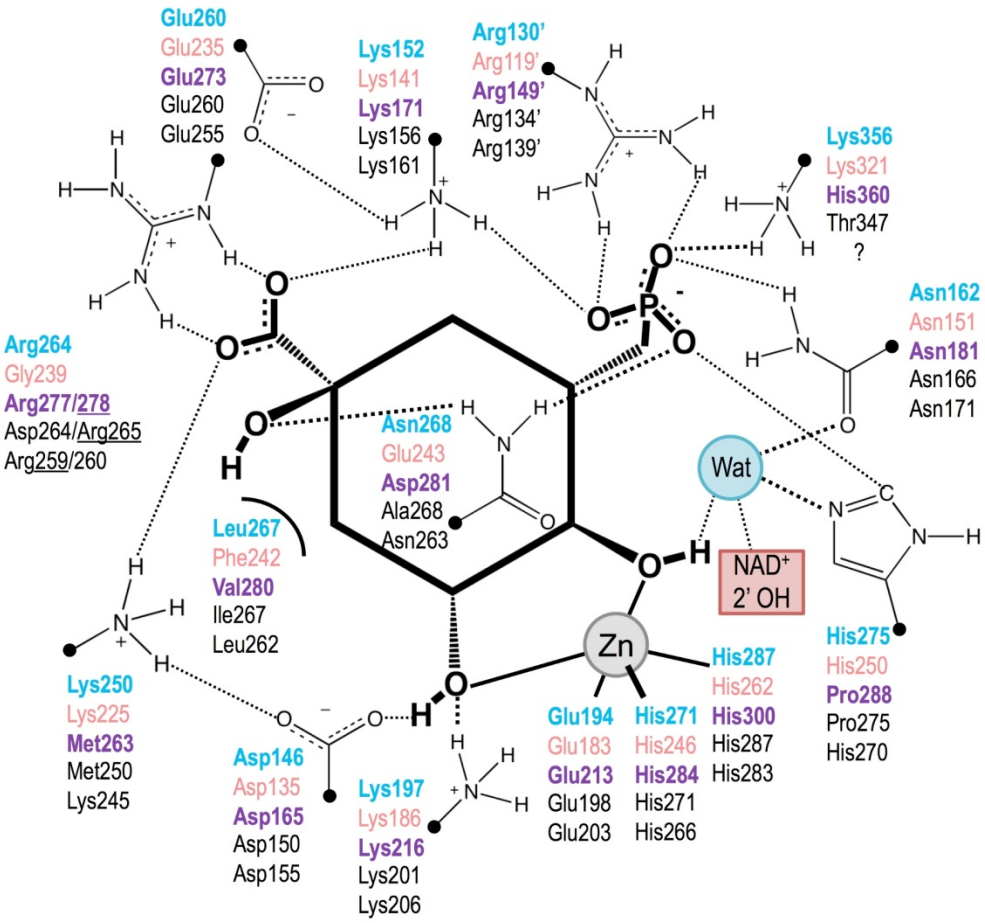
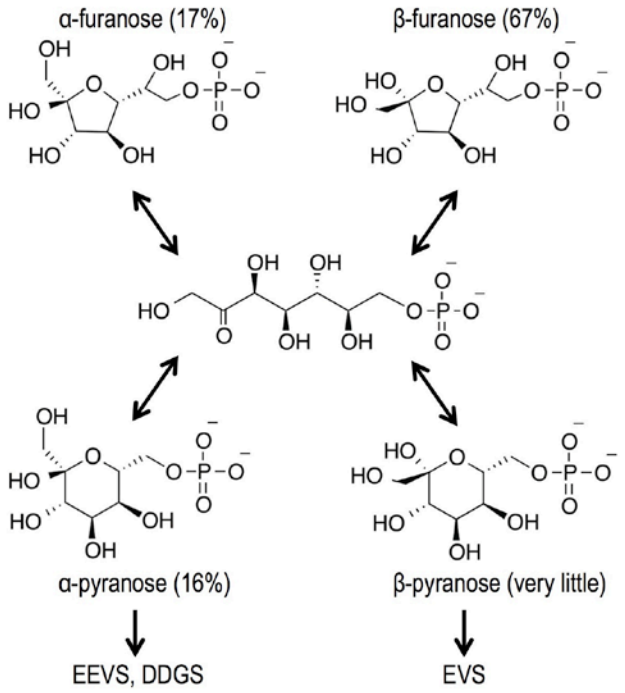
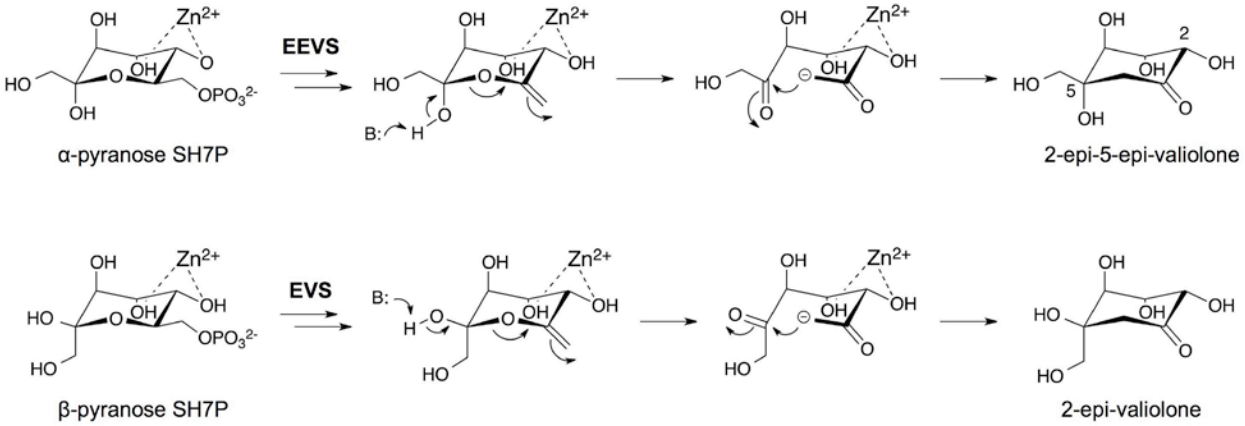

Glu260
Glu235
Glu273
Glu260
Glu255
Lys152
Lys141
Lys171
Lys156
Lys161
Arg130'
Arg119'
Arg149'
Arg134'
Arg139'
Lys356
Lys321
His360
Thr347
?
Asn162
Asn151
Asn181
Asn166
Asn171
Arg264
Gly239
Arg277/278
Asp264/Arg265
Arg259/260
Asn268
Glu243
Asp281
Ala268
Asn263
Wat
NAD+
2' OH
Leu267
Phe242
Val280
Ile267
Leu262
Zn
His275
His250
Pro288
Pro275
His270
His287
His262
His300
His287
His283
Glu194
Glu183
Glu213
Glu198
Glu203
His271
His246
His284
His271
His266
Lys250
Lys225
Met263
Met250
Lys245
Asp146
Asp135
Asp165
Asp150
Asp155
Lys197
Lys186
Lys216
Lys201
Lys206

Figure 9

(A)


α-furanose (17%)
β-furanose (67%)
α-pyranose (16%)
β-pyranose (very little)
EEVS, DDGS
EVS

(B)


EEVS
α-pyranose SH7P
2-epi-5-epi-valiolone
EVS
β-pyranose SH7P
2-epi-valiolone

For Table of Contents Use Only

Structure of a Sedoheptulose 7-Phosphate Cyclase: ValA from *Streptomyces hygroscopicus*

Kelsey M. Kean, Sara J. Codding, Shumpei Asamizu, Taifo Mahmud, and P. Andrew Karplus

Graphic for the Table of Contents