Published in final edited form as:
J Mol Biol. 2010 January 15; 395(2): 390. doi:10.1016/j.jmb.2009.10.045.

The Crystal Structure of the Novobiocin Biosynthetic Enzyme NovP: The First Representative Structure for the TylF *O*-Methyltransferase Superfamily

Inmaculada Gómez García[1,†], **Clare E. M. Stevenson**[1,†], **Isabel Usón**[2], **Caren L. Freel Meyers**[3], **Christopher T. Walsh**[3], and **David M. Lawson**[1,*]

[1]Department of Biological Chemistry, John Innes Centre, Norwich NR4 7UH, UK

[2]Institucio Catalana de Recerca i Estudis Avançats at Instituto de Biologia Molecular de Barcelona (IBMB-CSIC), Barcelona Science Park, Baldiri Reixach 15, 08028 Barcelona, Spain

[3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

Summary

NovP is an *S*-adenosyl-L-methionine-dependent *O*-methyltransferase that catalyses the penultimate step in the biosynthesis of the aminocoumarin antibiotic novobiocin. Specifically, it methylates at the 4-OH of the noviose moiety, and the resultant methoxy group is important for the potency of the mature antibiotic: previous crystallographic studies have shown that this group interacts directly with the target enzyme DNA gyrase, which is a validated drug target. We have determined the high resolution crystal structure of NovP from *Streptomyces spheroides* as a binary complex with its desmethylated co-substrate, *S*-adenosyl-L-homocysteine. The structure displays a typical class I methyltransferase fold, in addition to motifs that are consistent with a divalent metal-dependent mechanism. This is the first representative structure of a methyltransferase from the TylF superfamily, which includes a number of enzymes implicated in the biosynthesis of antibiotics and other therapeutics. The NovP structure reveals a number of distinctive structural features that, based on sequence conservation, are likely to be characteristic of the superfamily. These include a helical 'lid' region that gates access to the co-substrate binding pocket, and an active centre that contains a 3-Asp putative metal-binding site. A further conserved Asp likely acts as the general base that initiates the reaction by deprotonating the 4-OH group of the noviose unit. Using *in silico* docking we have generated models of the enzyme-substrate complex that are consistent with the proposed mechanism. Furthermore, these models suggest that NovP is unlikely to tolerate significant modifications at the noviose moiety, but could show increasing substrate promiscuity as a function of the distance of the modification from the methylation site. These observations could inform future attempts to utilise NovP for methylating a range of glycosylated compounds.

*david.lawson@bbsrc.ac.uk.
Present addresses: I. Gómez García, Structural Biology Unit, CIC-bioGUNE, Bizkaia Technology Park, 48160 Derio, Bizkai, Spain; C. L. Freel Meyers, Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
†IGG and CEMS contributed equally to this work

PDB accession code
Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 2WK1.

Keywords

O-methyltransferase; *Streptomyces spheroides*; novobiocin; TylF superfamily; crystal structure

Introduction

The process of methylation has diverse roles in biology, including chromatin regulation, gene silencing, signal transduction, protein repair, protein partitioning and biosynthesis.[1] Particular interest surrounds the mammalian *O*-methyltransferases (OMTs) as they are the targets of therapeutic agents designed to treat Parkinson's disease.[2] More recently, OMTs have attracted considerable interest on account of their roles in decorating antibiotics3; 4;5 and other therapeutic agents[6] and their potential application for creating chemical diversity in a range of compound classes.

In the majority of methyltransferases, *S*-adenosyl-L-methionine (SAM) is the methyl donor,[7] although there are a number of examples where a cobalamin (vitamin B12) cofactor is used to mediate methyl transfer from a variety of donor molecules to the substrate.8;9 The substrates of both types of enzyme range in size from small molecules through to DNA and proteins, and methylation at carbon, oxygen, nitrogen, sulphur or even halides is possible.10; 11 Equally diverse are the methyltransferases that perform these reactions.1 Analysis of their structures indicates that the majority are fundamentally similar, having core folds reminiscent of the canonical Rossman nucleotide binding fold12 with a central, mainly parallel, β-sheet flanked on both sides by α-helices. Nevertheless, in many cases, these cores have been embellished with additional structural elements tailored to accept a plethora of substrates. The structures seen so far can be loosely grouped by topology into five classes (I – V).[1] However, even within a class, the sequence similarity can be as low as 10%. In the SAM-dependent enzymes, the only residues that are loosely conserved lie in the N-terminal region of the core fold and comprise a glycine-rich sequence (referred to as motif I) that interacts with the amino acid portion of SAM, and an acidic loop (denoted motif II), which interacts with the ribose hydroxyls.7

The aminocoumarin antibiotics target the essential bacterial enzyme DNA gyrase; they bind to the B subunit and are competitive inhibitors of the ATPase reaction.[13] The three main compounds novobiocin, clorobiocin and coumermycin A_1 share common structural features, namely an aromatic acyl component (ring A), a 3-amino-4,7-dihydroxycoumarin moiety (ring B) and an L-noviosyl sugar (ring C). The biosynthesis of novobiocin (Fig. 1) requires three methylation steps which are performed by three distinct SAM-dependent methyltransferases: (1) NovO (subunit molecular weight 25 kDa) methylates at position C8 of the aminocoumarin ring, (2) NovU (subunit molecular weight 46 kDa) introduces a second methyl group at C5 of the noviose sugar prior to reduction of the keto group at C4 and transfer of the sugar moiety to the aglycone moiety, and (3) NovP (subunit molecular weight 30 kDa) methylates the 4-OH of the noviosyl group immediately after the glycosyl transfer step and represents the penultimate step of the whole pathway.[14] This latter methyltransfer is common to both the clorobiocin15 and coumermycin A_1[16] pathways, being catalysed by the enzymes CloP and CouP, respectively. In the case of coumermycin A_1, the reaction occurs twice since it has two noviose sugars, one at each end of the molecule.[4; 17] The final step of novobiocin biosynthesis, which is unique to this pathway, is the transfer of a carbamoyl group to the 3-OH by NovN.[18]

Methylation of an antibiotic can have both indirect and direct effects on its potency. Firstly, methylation will increase the lipophilicity of the antibiotic, facilitating its absorption across membranes and leading to elevated intracellular concentrations. Secondly, methyl substituents may form important hydrophobic interactions with the target. Crystallographic studies on

novobiocin complexes of DNA gyrase[19; 20; 21] have shown that the decorated novoise sugar (ring C) is responsible for the majority of the interactions with the target enzyme and that the 4-methoxy group itself partially occupies a hydrophobic pocket formed by the side-chains of residues Ile78 and Val120 of *E. coli* GyrB. Although no data are available on novobiocin lacking the 4-methoxy group, the equivalent desmethylated form of clorobiocin showed only 5% of the antibacterial activity of the parent compound against *Bacillus subtilis*[22] and a 25-fold increase in the IC_{50} value required to inhibit supercoiling in *E. coli* DNA gyrase.[23]

Previous characterisation of NovP has shown it to have low, but measurable, *in vitro* activity, with $k_{cat} = 0.400 \pm 0.006\ min^{-1}$ and $K_M = 9.5 \pm 3\ \mu m$.[18; 24] Studies with NovP and its orthologue CloP, have shown that both enzymes are capable of methylating unnatural substrates.[25; 26] The crystal structure of NovP presented here may provide explanations for these phenomena and, moreover, identify ways of manipulating the enzyme towards enhancing its *in vitro* activity and substrate promiscuity.

Results

Overall structure of NovP

The crystal structure of NovP was determined at 1.4 Å resolution. Of the 262 residue native sequence, a total of 242 amino acids were present in the final model: the first 13 residues (plus the 20-residue N-terminal His-tag which was not cleaved prior to crystallization), and a surface loop comprising residues 39–45 were not modelled because of insufficient electron density. Only four side-chains were not fitted (Asp14, Arg159, Lys241 and Arg262), and seven were modelled in two alternate conformations (Met22, Asp165, Thr189, Met226, Glu237, Ile249, Arg260).

The NovP monomer consists of a single globular α/β domain with approximate dimensions 53 × 46 × 37 Å (roughly corresponding to the horizontal, vertical and into plane dimensions of the molecule as viewed in Fig. 2). As expected, the topology is reminiscent of the classical Rossmann nucleotide binding fold.[12] The central, predominantly parallel, β-sheet is comprised of seven strands with connectivity −1x, −1x, 3x, 1x, 2x, −1, using the nomenclature of Richardson,[27] which is flanked on both sides and at the C-terminal end by a total of ten α-helices and two 3_{10} helices. This is consistent with a class I methyltransferase fold which is described as “a doubly wound open αβα sandwich”.[1] Crystallization of NovP was strictly dependent on the presence of *S*-adenosyl-L-homocysteine (SAH), the desmethylated co-substrate, and this was clearly visible in the structure, positioned at the C-terminal end of the β-sheet, which is the expected location for the nucleotide binding site in Rossmann fold proteins.

Sequence homologues of NovP

The reaction catalysed by NovP in novobiocin biosynthesis is common to the coumermycin A_1 and clorobiocin biosynthetic pathways. Although there are structural differences in the methyl acceptors, these are not proximal to the 4-OH of the noviosyl moiety (N.B. in all cases the 3-OH is not substituted at this stage in the respective biosynthetic pathways). It is no surprise, therefore, that in a BLAST search (http://www.expasy.ch/tools/blast/) with the NovP protein sequence, CouP and CloP are the top hits with respectively 89.1% and 84.1% amino acid sequence identity to NovP. In fact, all three enzymes belong to the TylF superfamily (PF05711), as defined in the PFAM database (http://pfam.sanger.ac.uk/),[28] which currently contains 69 sequences. These are found across 47 species from 7 phyla and, with the exception of two sequences from the eukaryote *Trypanosoma cruzi*, they are restricted to the bacterial kingdom, with almost two thirds, including TylF and the three aminocoumarin biosynthetic enzymes, being found in Actinobacteria. These latter four sequences, together with

representative sequences from the other six phyla, are aligned in Fig. 3a. TylF itself is responsible for the methylation of macrocin to produce the antibiotic tylosin in *Streptomyces fradiae*[3] and shares 48.9% amino acid sequence identity with NovP. As with NovP, TylF catalyses *O*-methylation of a deoxysugar, but at the 3- rather than the 4-position; this is the final and rate-limiting step in tylosin biosynthesis.[29; 30] To date, NovP provides the only representative crystal structure for the superfamily. Sequence comparisons suggest that a number of features that are characteristic of the NovP structure are found throughout the TylF superfamily (Fig. 3a).

Structural homologues of NovP

The structures of methyltransferases are well represented in the Protein Data Bank (PDB),[31] accounting for approximately 1% of all entries: a simple search of the PDB (as of June 1st 2009) based on the methyltransferase Enzyme Classification number (EC 2.1.1) retrieves 658 structures, which reduces to a non-redundant set of 247 after filtering out multiple copies of those with similar sequences (>90% identity). Close structural homologues were identified by submitting the NovP structure to the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/),[32] which returned 386 hits with Z scores of at least 10.0, reducing to a non-redundant set of 115 after filtering; still a substantial number. For the purposes of the following discussion, comparisons were mostly limited to selected enzymes with Z-scores exceeding 15.0. These are summarised in Table 1 and a structure-based sequence alignment versus NovP is shown in Fig. 3b. These are all OMTs and they belong to the small-molecule methyltransferase family[7] and have catechol-based substrates. Despite their structural similarities to NovP, at the amino acid sequence level, they only share identities in the range 11–14%. On the whole, the structural alignments of these enzymes with NovP are good, and where SAM/SAH ligands are present, they superpose well. However, for the first 60 or so residues of NovP (i.e. prior to α3), the structural alignment is poor (Fig. 4a), but this is not unexpected since these regions are generally quite variable in the small-molecule methyltransferases.[7] In addition, beyond the aminocoumarin biosynthetic enzymes in the TylF superfamily, the sequence conservation in this N-terminal region is poor (Fig. 3a), and therefore it is possible that there is also considerable structural variation for this region even within the superfamily.

The only other sugar OMT for which there is an X-ray structure is RebM (PDB accession code 3BUS), which is involved in the biosynthesis of the antibiotic rebeccamycin in *Lechevalieria aerocolonigenes*.[5] In common with NovP, methylation occurs at the 4-OH of the sugar. However, despite the substrate similarity, RebM is not amongst the closest structural homologues to NovP (Fig. 4b) with a DALI Z score of 10.1; this may be in part because they are likely to function by different mechanisms, as described below.

NovP is a homodimer

There is a single copy of the NovP protomer in the asymmetric unit, although analysis of crystal contacts using the Protein Interfaces, Surfaces and Assemblies server (PISA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)[33] indicates that there may be a biologically relevant dimer, which is generated by the crystallographic 2-fold axis. The resultant dimer is relatively elongated and the interface buries some 856 Å^2 of protein surface per subunit, which amounts to approximately 8% of the monomer solvent accessible surface and mainly involves residues from helices α1 and α7 (Fig. 3 and Fig. 5). This putative dimer is distinct from the relatively compact dimer seen in several of the structurally related OMTs, such as the Synechocystis OMT (SynOMT, PDB code 3CBG)[34] and the alfalfa caffeoyl CoA OMT (CCoAOMT, PDB code 1SUS),35 where the buried surface in these is more extensive (1684 and 1903 Å^2, respectively), and involves a completely different face of the molecule (Fig. 5). The presence of a NovP dimer in solution was confirmed using size exclusion

chromatography. The main NovP fraction eluted at a size of approximately 66 kDa, corresponding closely to the value of 64.2 kDa expected for a His-tagged dimer. There was no peak that could be attributed to monomer. Given that the majority of the interactions are hydrophobic, and therefore non-specific, it is difficult to predict with certainty whether this interface is a common feature of the TylF superfamily.

Nevertheless, with the exception of the amino acid at the equivalent of position 14 in the NovP sequence, the interface residues are all conserved in CouP and CloP, strongly suggesting this dimer is also a feature of these two proteins. TylF has already been shown to exist as a homodimer in solution,[29; 30] but since only around half of the NovP dimer interface residues are conserved in TylF, the existence of an equivalent dimer in TylF would seem less certain. The conservation of the interface residues in the representative sequences from the other six phyla is relatively low and thus NovP-like dimers would seem much less likely for these proteins. Indeed, for MtfB from *Microscilla marina* ATCC 23134, this type of dimer could not exist since it completely lacks the equivalent of helix α1 of NovP (Fig. 3a). To our knowledge, with the exception of NovP and TylF, the oligomeric states of other TylF superfamily members have not been established to date.

The SAM/SAH binding site

In common with other Class I methyltransferases, the SAH ligand in NovP adopts an extended conformation. It is tightly bound, with virtually every potential hydrogen-bonding group of the ligand interacting directly with the protein (Fig. 6); there are also a significant number of non-bonded interactions. The NovP sequence contains recognisable motifs I and II that are typical of SAM-dependent methyltransferases. Motif I, the 'glycine rich sequence' lies between β1 and α4 and forms the expected interactions with the amino acyl portion of SAM via Glu92 and Gly94. There is also a further interaction between Trp96 and the ribose. Motif II, the 'acidic loop', which lies at the C-terminal end of β2, forms the expected interactions with the ribose hydroxyls via Asp122. The sequences of motifs I and II are well conserved in the TylF superfamily (Fig. 3a).

In the close structural homologues of NovP, the nucleotide is bound in a solvent accessible surface cleft with, at most, one or two side-chains gating exchange of the co-substrate with the desmethylated co-product. In NovP, the bound SAH is almost completely buried, such that only the exocyclic N6 amino group of the adenine ring and the sulphur atom are accessible to bulk solvent (Fig. 2). Clearly for turnover, SAH/SAM exchange is required, but this would necessitate significant conformational changes. Inspection of protein backbone indicates that the majority of the polypeptide between β2 and β3, specifically residues 126–151 inclusive, form a largely helical sub-domain that forms a lid over the co-substrate binding site. This region appears as an insertion in the structure-based sequence alignment of NovP with close structural homologues (Fig. 3b). In fact, the residues immediately before and immediately after the lid are close together in space (Cα - Cα separation 6.2Å), such that deletion of the lid would very likely have little effect on the rest of the structure. Although the lid appears to mediate SAH/SAM interchange, only two amino acids, Gly126 and Pro128, interact directly with the co-substrate. They form van der Waals interactions with the ribose and adenine moieties, respectively and thus, presumably, the lid cannot directly sense the methylation status of the co-substrate. Sequence comparisons suggest that the lid is a common feature of all members of the TylF superfamily (Fig. 3a); it is notable that Gly126 and Pro128 are well conserved, suggestive of functional roles, for example, the inherent flexibility of the Gly may be important for conformational changes in the lid. Asp128 is quite well conserved and, in the NovP structure, its side-chain hydrogen bonds to the backbone amides of Tyr200 and Glu201, and thus may have a role in maintaining the closed conformation of the lid. Elsewhere, the lid sequences are poorly conserved, although, in the alignment shown in Fig. 3a, the lid regions

are of a similar length, with the exception of that from *Herpetosiphon aurantiacus*, which is fifteen residues longer than that of NovP.

Analysis of temperature factors reveals that the lid is more mobile than the rest of the NovP structure, consistent with the hypothesis that it gates access to the SAM site (residues 126–151 inclusive have average main-chain and side-chain temperature factors of 20.5 and 24.7 $Å^2$, respectively, as compared to the values of 14.0 and 17.7 $Å^2$, respectively, for the rest of the protein). Moreover, the presence of four Gly residues in this sub-domain may give it increased flexibility. The lid regions of the two monomers within the putative dimer lie on the same face of the assembly. However, since they are not close to the dimer interface itself, they probably function independently.

In the initial characterisation of NovP, the *in vitro* methyltransferase activity of the enzyme was reported to be exceedingly low,[18] taking over 2 minutes per reaction. It is tempting to speculate that turnover is, at least in part, rate-limited by SAH/SAM interchange, which clearly necessitates a large conformational change. However, thermally-driven domain motions generally occur on the micro- to milli-second timescale,[36; 37] and therefore lid movement alone may not fully account for the poor activity. Nevertheless, in these cases, domain motions involve relatively simple and concerted movements of groups of atoms along similar trajectories; perhaps in NovP a more complex, and relatively time-consuming, conformational change is necessitated to expose the co-substrate binding site. It is possible that protein-protein interactions with other pathway enzymes play important roles in upregulating the activity of NovP *in vivo*, perhaps assisting the opening of the lid which would otherwise be firmly closed; in the crystal structure there are nine hydrogen bonds between the lid and the rest of the protein. Likely candidates include the enzymes catalysing the steps immediately before (i.e. the glycosyl transferase NovM) or after (i.e. the carbamoyltransferase NovN) the NovP step. Moreover, such protein-protein interactions may be required to channel the intermediates between successive active sites in the pathway to prevent their loss into the cytoplasm, where they could exert their toxic effects.[38] However, to date, there is no evidence either for or against these protein-protein interactions.

Several of the structural homologues have an extended loop between the equivalents of β5 and α10 in NovP (Fig. 3b, Fig. 4 and Fig. 5). This has been described elsewhere as the 'insertion loop' and plays a role in substrate binding.[35; 39] Superpositions of these structures with that of NovP show that these insertion loops overlap with α5 of the lid region of NovP (Fig. 4 and Fig. 7), suggesting that these two structural features are mutually exclusive in a given enzyme. Indeed, all TylF superfamily members have very short loops here (Fig. 3a). In NovP, this loop contains a CPPC sequence and, in the structure, the two Cys residues are linked by a disulphide bridge (Fig. 2–Fig. 5 and Fig. 7). Disulphide bridges are unusual in cytoplasmic proteins and it is conceivable that this is an artefact brought about by the prolonged exposure of the protein to aerobic conditions. Indeed, the NovP crystals took up to one month to grow.[40] The observation that only the second of the two Cys residues is conserved in the TylF superfamily (Fig. 3a) would suggest that the disulphide bridge is neither functionally nor structurally important. However, given the proximity of the disulphide bridge to the active centre then, perhaps its formation may, for example, distort the metal binding site and thereby to some extent account for low *in vitro* enzyme activity.

The active site and putative mechanism

As mentioned in the previous section, despite being almost completely buried, the SAH sulphur atom is solvent exposed (Fig. 2). In fact, it sits at the end of a well-defined channel leading to bulk solvent; several of the residues lining this channel are well conserved across the TylF superfamily (Fig. 3a). This is undoubtedly the route by which the substrate, desmethyldescarbamoyl novobiocin (DDN), approaches the co-substrate. Indeed, when SAM

is modelled in place of the SAH, the strongly electrophilic methyl group would be the first atom of the co-substrate that an approaching substrate molecule would encounter. The methyl group lies approximately 11 Å from the mouth of the channel. The outer section of the channel is approximately 8–10 Å across and then becomes constricted some 5 Å from the methyl group, being approximately 4–6 Å across thereafter. One side of the channel is delineated by the lid region, and it seems likely that it plays a role in binding the substrate and guiding it into the active site. The residues lining the channel are strictly conserved between NovP, CouP and CloP, with the exception of that at position 146 (NovP sequence numbering), which is in the lid region (Fig. 3a). This is an Arg in both NovP and CouP, and a His in CloP. Rings B and C (see Fig. 1a) of the methyl acceptors for these three enzymes are identical apart from the C8 of ring B, being methylated for NovP and CouP substrates and chlorinated for the CloP substrate. This raises the possibility that the amino acid at position 146 may have some role in binding ring B, in particular the substituent at C8. Nevertheless, there are precedents for both Arg and His interacting with chlorine atoms in halogenated compounds.[41] Beyond the aminocoumarin biosynthetic enzymes, the conservation of the residues lining the active site channel is less marked, for example in TylF, of the 20 residues that line the channel, only 11 are conserved (Fig. 3a), with 4 of these being the Asp residues that are implicated in the mechanism described below (Fig. 8). Although the sugar moieties of the NovP and TylF substrates are both β-linked 6-deoxy sugars, they are quite dissimilar: (i) they have the opposite configuration i.e. L-noviose for NovP, and 6-deoxy-2-*O*-methyl-D-allose for TylF, (ii) they have different substitutions, and (iii) the hydroxyl groups to be methylated are at the 4 position and equatorial for NovP, and at the 3 position and axial for TylF (Fig. 1). Therefore, the relatively poor sequence conservation in the active site channels of NovP and TylF is not altogether unexpected. As the result of the substrate docking experiments described below, the side-chain of Trp58 was identified as a potential hydrogen bonding partner to the ring oxygen of noviose (Fig. 8). This is conserved in roughly half of the TylF sequences and could be indicative of a substrate containing a methyl acceptor sugar with similarities to the β-linked L-noviose of novobiocin. For instance, the sugar moieties in elloramycin[42] and nogalamycin[43] are both β-linked L-sugars and the Trp is conserved in their respective OMTs; the Trp is not conserved in the OMTs for tylosin and chalcomycin,[44] which act on β-linked D-sugars, nor is it conserved in the OMT for chromomycin A_3,[45] which acts on an α-linked D-sugar. Since functional data on many of the TylF superfamily proteins is rather scant, it is difficult to make broader comparisons.

It is conceivable that the disordered loop between α1 and α2 of NovP (residues 39–45) could impinge on the mouth of the substrate binding channel (Fig. 2), and perhaps play a minor role in substrate binding. By contrast, it seems unlikely that the disordered N-terminus would have any effect on substrate binding since it lies on the opposite face of the molecule to the active site. The closest structural homologues to NovP are either proven or putative metal-dependent OMTs and almost exclusively employ the side-chains of two Asp and a single Asn residues to bind the metal, which is usually magnesium. The two Asp residues are structurally conserved in NovP (residues 196 and 223), but the Asn is replaced by a further Asp (residue 224). It would seem highly probable that this triad of residues also represents a metal binding site in NovP; its significance is supported by the juxtaposition of this site to the projected position for the methyl donor group of SAM, as well as the strict conservation of these residues in the TylF superfamily (Fig. 3a). Indeed, TylF itself has previously been shown to be a divalent metal-dependent enzyme, being active with Mg^{2+} ions, as well as with Mn^{2+} and Co^{2+} ions.[29; 30; 46] However, no evidence for metal ions is present in the NovP crystal structure and, moreover, no density for metal ions was apparent after crystallisation or soaking crystals with $MgCl_2$ (data not shown). It is conceivable that metal binding is co-dependent on substrate binding; the majority of the homologous structures that contain metal ions in their active sites also contain bound substrates, products or inhibitors. Although the presence of divalent metal ions was not explicitly reported in the assay conditions used for the biochemical characterisation

of NovP,[18] both 10 mM $MgCl_2$ and $MnCl_2$ were tested in initial experiments, but these did not have a measurable effect on activity (Freel Meyers, personal communication). This suggests that the availability of these metal ions is not rate-limiting *in vitro*, even when they are not specifically added to the assay. Curiously, the addition of 10 mM $CaCl_2$ drastically reduced the level of product formation (Freel Meyers, personal communication). With hindsight, the inhibitory effect of calcium ions is likely due to competition with the presumed trace quantities of the preferred divalent metal ions.

In general, methyltransferase reactions involve the direct transfer of the methyl group to substrate in an S_N2-like mechanism. This process requires that the substrate be deprotonated before, at the same time as, or after methyl transfer. In metal-independent OMTs such as chalcone OMT, isoflavine OMT[47] and RebM,[5] a His residue deprotonates the hydroxyl. The structural data presented above together with comparisons with related structures, and the biochemical data for TylF,[29;30; 46] strongly support the hypothesis that NovP is a metal-dependent enzyme, and that the metal is coordinated via the strictly conserved residues, Asp196, Asp223 and Asp224. The metal ion may have roles in both the binding of the substrate and polarising the 4-OH to facilitate, rather than effect, its deprotonation. The closest structural homologues to NovP (Table 1) appear to employ a nearby Lys residue to deprotonate the substrate hydroxyl. The equivalent residue in the NovP sequence is actually Leu199 (Fig. 3b). However, the protein backbone directs the side-chain of Leu199 away from the active site. Instead, the side-chain of the previous residue, Asp198, points into the active site and would appear to be ideally placed to act as the general base (Fig. 7). This alternative backbone conformation may be a consequence of a single residue insertion in the β4 - α9 loop of NovP relative to its structural homologues. The temperature factors for this loop (residues Asp196-Tyr200) are comparable to the overall average for the rest of the protein (minus the lid), indicating that this is not a particularly mobile region, and is therefore likely to be preconfigured for catalysis. Asp198 is also strictly conserved in the TylF superfamily, thereby supporting its proposed importance. Moreover, in a superposition with RebM, the side-chain of Asp198 in NovP occupies the equivalent position to that of His140, the catalytic base of RebM (Fig. 7b). The proposed NovP mechanism is illustrated in Fig. 8.

Since the lid region constitutes one face of the substrate-binding channel it seems plausible that this would be in place before the substrate binds, in other words, SAM would bind first. After methyltransfer, either product could leave first, although it would seem more likely that SAH would be the last to leave as it is tightly bound in the crystal structure. This is consistent with the ordered Bi Bi mechanism reported for TylF.[29; 30]

It is not possible to draw any firm conclusions from a comparison of the architectures of the NovP and RebM active sites, as it is difficult to predict where the substrate would bind for the latter because there is insufficient space adjacent to His140 to accommodate the substrate in the crystal structure.[5] Access to His140 in RebM is occluded by (i) a 22 residue insertion relative to NovP in the β5 - α10 loop (equivalent position to the insertion loops described above), (ii) a 49 residue insertion in the β6 - β7 loop, and (iii) the N-terminus (which contains a short α-helix in one of the two copies of the monomer in the asymmetric unit) (Fig. 4b). At least one of these elements would need to move to admit the substrate; an analysis of temperature factors in the RebM structure indicates that the N-terminus is the most mobile.

Substrate docking

Attempts were made to obtain crystals of additional NovP-ligand complexes, using the product (descarbamoyl novobiocin) and a substrate analogue (methylrhamnoside), but we did not have sufficient quantities of substrate for these experiments. Co-crystallisation and soaking experiments were performed for these compounds in the presence of either SAH or SAM, both

with and without magnesium ions ($MgCl_2$); unfortunately, we were not successful (data not shown).

In order to generate a model of the quaternary complex formed between NovP, SAM, Mg^{2+} and DDN, we resorted to *in silico* docking using AutoDock4.[48; 49] The 'receptor' consisted of the desolvated protein plus SAM and a Mg^{2+} ion coordinated by the side-chains of Asp196, Asp223 and Asp224 and was prepared as described in Materials and Methods. The ligand was allowed full flexibility and, for each simulation, 100 runs were evaluated. Productive binding modes were identified as poses that placed the 4-OH of the noviose moiety of DDN within 4 Å of the methyl group of SAM. When the situation prior to reaction was simulated, i.e. when the 4-OH is still protonated and Asp 198, the general base, is deprotonated, there were only 3 poses that satisfied this condition, with the best docking energy being −8.68 kJ mol^{-1}. However, when the situation after the deprotonation of the 4-OH by Asp 198 was simulated, there were 98 poses that satisfied this condition, with the best docking energy being −15.88 kJ mol^{-1}. Thus, mimicking an 'activated' enzyme-substrate complex is a much more efficient means of identifying productive binding modes. This approach has been used previously with rat liver catechol OMT.[50] The 98 productive poses were sorted into 12 clusters based on similarity, and the conformation of the lowest energy pose from each cluster is displayed in Fig. 9. It is clear that the conformation of the noviose moiety (ring C) is essentially the same in all 12 poses, with all the interactions shown in Fig. 8 being present. This is not surprising as ring C occupies the relatively constricted inner section of the channel where it has little conformational freedom. The position of the aminocoumarin moiety (ring B) is more variable, but nevertheless the conformations seen fall into roughly 4 clusters. Again, this is to be expected because ring B is located in the outer part of the channel and thus has more conformational freedom. Also shown in Fig. 9 is the position of Arg146 in the lid region. The side-chain of this residue is directed towards the aminocoumarin ring, suggesting that it could have a role in binding this moiety, in particular the substituent at C8, as discussed earlier. The prenylated hydroxybenzoate moiety projects beyond the substrate channel and thus it takes up a variety of conformations and generally makes very little contact with the protein. It would therefore seem likely that NovP could tolerate quite substantial changes to the ring A substituent without compromising its ability to methylate the noviose moiety; this has already been demonstrated experimentally.[25]

Conclusion and Summary

The structure of NovP from *Streptomyces spheroides* is the first representative structure of a SAM-dependent *O*-methyltransferase from the TylF superfamily. Although it has many close structural homologues in the PDB, it displays a number of differences that are likely to be characteristic features of the TylF superfamily. NovP forms a novel elongated dimer mediated largely by hydrophobic interactions. Rougly half of the NovP dimer interface residues are conserved in TylF, so it is possible that the dimer previously observed for TylF[29; 30] may have a similar arrangement. Comparisons with related structures support a divalent metal-dependent mechanism using a 3-Asp metal-binding site and a further Asp as a general base. Moreover, the strict conservation of these four residues in the TylF superfamily, and the observation that TylF itself is also divalent metal-dependent[29; 30; 46] further strengthens this hypothesis. Access to the co-substrate binding pocket is mediated by a 26-residue helical sub-domain, denoted the lid, which also forms one face of a funnel-shaped substrate access channel. Using *in silico* docking we have provided plausible models for the quaternary complex formed between NovP, SAM, Mg^{2+} and the substrate desmethyldescarbamoyl novobiocin.

Although it is not possible to state definitively due to the vagaries of sequence annotations, it seems probable that most, if not all, TylF superfamily members are sugar OMTs and, certainly within the Streptomycetes, they are implicated in the biosynthesis of several glycosylated

therapeutic agents e.g. tylosin,[3] elloramycin[42] and chromomycin A_3;[51] in all of these examples the substrates are sugars attached to bulky aglycone moieties. Thus, the conclusions drawn from the NovP structure could have broader implications for the biosynthesis of glycosylated microbial natural products in general. The lid region, which also appears to be a general feature of the TylF superfamily, could partially account for the poor *in vitro* activity of NovP, since large, potentially rate-limiting, conformational changes would be required for SAM/SAH interchange. It is also possible that NovP is subject to product inhibition, as has been suggested previously;[52] both substrate and product inhibition have also been reported for TylF.[29; 30; 46] The physiological role of the lid is unclear, but it may function to sequester SAM from other competing metabolic pathways in preparation for the arrival of DDN. Alternatively, NovP may represent a control point in the biosynthetic pathway and, by regulating its activity through some mechanism involving the lid, the cell could limit the production of potentially cytotoxic compounds. In this regard, it is notable that in the biosynthesis of tylosin,[3] elloramycin[42] and chromomycin A_3,[51] the TylF superfamily member is either the terminal or penultimate enzyme in the pathway and, therefore, the reactions they catalyse are likely to have a significant effect on the potency of the biosynthetic intermediate. In addition, TylF has been reported to represent the rate-limiting step in tylosin production.[29]

Although the substrate promiscuity of NovP has not been fully explored, it will tolerate modifications to ring A;[25] the same is also true for CloP.[26] These observations are consistent with our models of the enzyme-substrate complex derived from the docking simulations, which show that ring A projects beyond the substrate binding channel, and thus is unlikely to form significant contacts with the protein. It seems probable that modifications to rings B and C are less likely to be tolerated due to steric constraints within the substrate binding channel.

NovP clearly has potential as an enzyme for tailoring a variety of glycosylated compounds, but its utility will be limited by poor *in vitro* activity and restricted substrate specificity. We can potentially address both of these issues simultaneously through mutations to the lid sub-domain; the most drastic mutation being a complete excision of this region (residues 126–151 inclusive). Comparisons with close structural homologues suggest that this deletion should not disrupt the rest of the structure and would leave a solvent accessible SAM binding pocket and a less constrained substrate binding site.

Materials and Methods

X-ray crystallography

Recombinant, N-terminally His-tagged *Streptomyces spheroides* NovP was purified and crystallized, and the structure was solved using a combination of isomorphous replacement with anomalous scattering (SIRAS) with manual building of the initial very noisy map and data extrapolation to 1.0 A, as described previously.[40; 53] The NovP structure was completed by further rounds of restrained refinement and interactive model building. Positional and anisotropic thermal parameters of the model were refined with REFMAC554 against a 1.4 Å resolution native data set that was collected on beamline PX10.1 at the SRS (Daresbury),[40; 53] and solvent molecules were automatically added using the ARP procedure.55 Model building was performed using the program COOT56 with reference to SIGMAA-weighted[57] 2mFobs – DFcalc and mFobs – DFcalc Fourier electron density maps. This yielded a high quality structure with final R_{work} and R_{free} values of 14.5% and 16.2%, respectively, and 96.6% of the residues in the favoured regions of the Ramachandran plot with no outliers, as evaluated by MOLPROBITY.[58] A summary of the X-ray data, together with the contents and the geometrical parameters of the final structure, is given in Table 2.

A difference electron density 'omit' map was generated for the SAH ligand using phases calculated from the final model minus the SAH coordinates after simulated annealing (SA)

refinement (Fig. 6). The SA refinement was performed from a starting temperature of 5000 K after applying random shifts to the model ('shake' term set to 0.3) using PHENIX.[59] All structural figures were produced using PyMOL.60

Size exclusion chromatography

The oligomeric state of a sample of His-tagged NovP was analysed by gel filtration. Protein was purified by nickel affinity chromatography as described previously[40] and subsequently buffer-exchanged into 50 mM Tris-HCl, 200 mM NaCl at pH 8.0. Approximately 2 ml of sample was then loaded onto a Superdex 75 16/60 column (GE Healthcare) at a concentration of approximately 10 mg ml^{-1} and eluted with the same buffer at a flow rate of 1 ml min^{-1}. The column was calibrated using a low molecular weight calibration kit (GE Healthcare) containing the following standards: albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa), ribonuclease A (13.7 kDa) and dextran blue (2,000 kDa).

Substrate docking

Docking simulations were carried out using the program AutoDock4[48] through the AutoDockTools graphical user interface.[49] The ‘receptor’ consisted of the desolvated protein plus SAM and a Mg^{2+} ion. This was prepared by replacing the SAH of the crystal structure with SAM and inserting a Mg^{2+} ion adjacent to the side-chains of Asp196, Asp223 and Asp224 using the superposed structures shown in Table 1 as a guide. The conformations of the three Asps were adjusted manually in order to optimise interactions with the Mg^{2+} ion. The resultant complex was subjected to 300 cycles of maximum likelihood energy minimization using CNS[61] prior to docking (with the 'Wa' term set to zero). During docking, the receptor was treated as a rigid entity with the exception of the side-chains of Asp198, the putative active site base, and Trp58, which appeared to be ideally placed to hydrogen bond to the ring oxygen of the noviose moiety (this had been established from previous trial simulations). A random starting position was chosen for the substrate (DDN) in which all single bonds were allowed to rotate. Searches were performed using a Lamarckian Genetic Algorithm within a cube of edge 26 Å with a grid spacing of 0.375 Å that encompassed the substrate binding channel of NovP. Simulations consisted of 100 runs, each with a population size of 150 and 25,000,000 energy evaluations. A total of 27,000 generations were analyzed, with only a single conformation surviving to the next generation. Three hundred local search iterations were carried out for each run. The resulting substrate conformations were clustered on the basis of their similarity with an rms tolerance of 2 Å. Docked poses were filtered automatically: ‘productive’ binding modes were detected by selecting poses that placed the 4-OH of the noviose moiety of DDN within 4 Å of the methyl group of SAM. The simulation was performed twice with the above parameters and it was assumed that the first step in the reaction is deprotonation at 4-OH of the noviose moiety. For the first simulation, the 4-OH was protonated and Asp198 was deprotonated, in order to model the enzyme-substrate complex prior to the reaction. For the second simulation, the situation after proton transfer from the 4-OH of noviose to Asp198, but prior to the methyltransfer step was modelled. Thus, in this case, the 4-OH was deprotonated and Asp198 was protonated at the start of the simulation.

Abbreviations

SAM	*S*-adenosyl-L-methionine
SAH	*S*-adenosyl-L-homocysteine
DDN	desmethyldescarbamoyl novobiocin
OMT	*O*-methyltransferase
SynOMT	Synechocystis OMT

CCoAOMT	alfalfa caffeoyl CoA OMT
PDB	Protein Data Bank
SA	simulated annealing

Acknowledgments

IGG, CEMS and DML were supported by the BBSRC through responsive mode funding (ref B19400) and the Core Strategic Grant to the John Innes Centre. IU was supported by the European Union Integrated Project BIOXHIT and the Spanish MEC (Grant BIO2006-14139). CLFM and CTW were supported by NIH grants F32 AI054007 and GM 49338, respectively. We are grateful to L. Heide for critically reading this manuscript and to M. Cianci, M. Ellis and R. Strange for assistance with data collection at the SRS (Daresbury).

References

1. Schubert HL, Blumenthal RM, Cheng X. Many paths to methyltransfer: a chronicle of convergence. Trends. Biochem. Sci 2003;28:329–335. [PubMed: 12826405]
2. Mannisto PT, Kaakkola S. New selective COMT inhibitors: useful adjuncts for Parkinson's disease? Trends. Pharmacol. Sci 1989;10:54–56. [PubMed: 2655238]
3. Cundliffe E, Bate N, Butler A, Fish S, Gandecha A, Merson-Davies L. The tylosin-biosynthetic genes of *Streptomyces fradiae*. Antonie Van Leeuwenhoek 2001;79:229–234. [PubMed: 11816964]
4. Li SM, Westrich L, Schmidt J, Kuhnt C, Heide L. Methyltransferase genes in *Streptomyces rishiriensis*: new coumermycin derivatives from gene-inactivation experiments. Microbiology 2002;148:3317–3326. [PubMed: 12368465]
5. Singh S, McCoy JG, Zhang C, Bingman CA, Phillips GN, Thorson JS. Structure and mechanism of the rebeccamycin sugar 4'-*O*-methyltransferase RebM. J. Biol. Chem 2008;283:22628–22636. [PubMed: 18502766]
6. Luo Y, Lin S, Zhang J, Cooke HA, Bruner SD, Shen B. Regiospecific *O*-methylation of naphthoic acids catalyzed by NcsB1, an *O*-methyltransferase involved in the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin. J. Biol. Chem 2008;283:14694–14702. [PubMed: 18387946]
7. Martin JL, McMillan FM. SAM (dependent) I AM: the *S*-adenosylmethionine-dependent methyltransferase fold. Curr. Opin. Struct. Biol 2002;12:783–793. [PubMed: 12504684]
8. Ludwig ML, Matthews RG. Structure-based perspectives on B12-dependent enzymes. Annu. Rev. Biochem 1997;66:269–313. [PubMed: 9242908]
9. Banerjee R, Ragsdale SW. The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. Annu. Rev. Biochem 2003;72:209–247. [PubMed: 14527323]
10. Ohsawa N, Tsujita M, Morikawa S, Itoh N. Purification and characterization of a monohalomethane-producing enzyme *S*-adenosyl-L-methionine: halide ion methyltransferase from a marine microalga, *Pavlova pinguis*. Biosci. Biotechnol. Biochem 2001;65:2397–2404. [PubMed: 11791711]
11. Attieh JM, Hanson AD, Saini HS. Purification and characterization of a novel methyltransferase responsible for biosynthesis of halomethanes and methanethiol in *Brassica oleracea*. J. Biol. Chem 1995;270:9250–9257. [PubMed: 7721844]
12. Rossmann, MG.; Liljas, A.; Branden, C-I.; Bansazak, LJ. Evolutionary relationships among dehydrogenases. In: Boyer, IPD., editor. The Enzymes. Vol. Vol.11. New York, NY: Academic Press; 1975. p. 61-102.
13. Maxwell A, Lawson DM. The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. Curr. Top. Med. Chem 2003;3:283–303. [PubMed: 12570764]
14. Steffensky M, Müuhlenweg A, Wang ZX, Li SM, Heide L. Identification of the novobiocin biosynthetic gene cluster of *Streptomyces spheroides* NCIB 11891. Antimicrob. Agents Chemother 2000;44:1214–1222. [PubMed: 10770754]
15. Pojer F, Li SM, Heide L. Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics. Microbiology 2002;148:3901–3911. [PubMed: 12480894]

16. Wang ZX, Li SM, Heide L. Identification of the coumermycin A(1) biosynthetic gene cluster of *Streptomyces rishiriensis* DSM 40489. Antimicrob. Agents Chemother 2000;44:3040–3048. [PubMed: 11036020]

17. Freel Meyers CL, Oberthur M, Heide L, Kahne D, Walsh CT. Assembly of dimeric variants of coumermycins by tandem action of the four biosynthetic enzymes CouL, CouM, CouP, and NovN. Biochemistry 2004;43:15022–15036. [PubMed: 15554710]

18. Freel Meyers CL, Oberthur M, Xu H, Heide L, Kahne D, Walsh CT. Characterization of NovP and NovN: completion of novobiocin biosynthesis by sequential tailoring of the noviosyl ring. Angew. Chem. Int. Ed. Engl 2004;43:67–70. [PubMed: 14694473]

19. Lewis RJ, Singh OM, Smith CV, Skarzynski T, Maxwell A, Wonacott AJ, Wigley DB. The nature of inhibition of DNA gyrase by the coumarins and the cyclothialidines revealed by X-ray crystallography. EMBO J 1996;15:1412–1420. [PubMed: 8635474]

20. Holdgate GA, Tunnicliffe A, Ward WH, Weston SA, Rosenbrock G, Barth PT, Taylor IW, Pauptit RA, Timms D. The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. Biochemistry 1997;36:9663–9673. [PubMed: 9245398]

21. Lamour V, Hoermann L, Jeltsch JM, Oudet P, Moras D. An open conformation of the *Thermus thermophilus* gyrase B ATP-binding domain. J. Biol. Chem 2002;277:18947–18953. [PubMed: 11850422]

22. Freitag A, Rapp H, Heide L, Li SM. Metabolic engineering of aminocoumarins: inactivation of the methyltransferase gene *cloP* and generation of new clorobiocin derivatives in a heterologous host. Chembiochem 2005;6:1–9.

23. Flatman RH, Eustaquio A, Li SM, Heide L, Maxwell A. Structure-activity relationships of aminocoumarin-type gyrase and topoisomerase IV inhibitors obtained by combinatorial biosynthesis. Antimicrob. Agents Chemother 2006;50:1136–1142. [PubMed: 16569821]

24. Pi N, Meyers CLF, Pacholec M, Walsh CT, Leary JA. Mass spectrometric characterization of a three-enzyme tandem reaction for assembly and modification of the novobiocin skeleton. Proc. Natl. Acad. Sci. USA 2004;101:10036–10041. [PubMed: 15218104]

25. Pacholec M, Freel Meyers CL, Oberthur M, Kahne D, Walsh CT. Characterization of the aminocoumarin ligase SimL from the simocyclinone pathway and tandem incubation with NovM, P, N from the novobiocin pathway. Biochemistry 2005;44:4949–4956. [PubMed: 15779922]

26. Anderle C, Hennig S, Kammerer B, Li S-M, Wessjohann L, Gust B, Heide L. Improved mutasynthetic approaches for the production of modified aminocoumarin antibiotics. Chemistry & Biology 2007;14:955–967. [PubMed: 17719494]

27. Richardson JS. The anatomy and taxonomy of protein structures. Adv. Prot. Chem 1981;34:167–339.

28. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. The Pfam protein families database. Nucleic Acids Res 2008;36:D281–D288. [PubMed: 18039703]

29. Bauer NJ, Kreuzman AJ, Dotzlaf JE, Yeh WK. Purification, characterization, and kinetic mechanism of *S*-adenosyl-L-methionine:macrocin *O*-methyltransferase from *Streptomyces fradiae*. J. Biol. Chem 1988;263:15619–15625. [PubMed: 3170601]

30. Kreuzman AJ, Turner JR, Yeh WK. Two distinctive *O*-methyltransferases catalyzing penultimate and terminal reactions of macrolide antibiotic (tylosin) biosynthesisSubstrate specificity, enzyme inhibition, and kinetic mechanism. J. Biol. Chem 1988;263:15626–15633. [PubMed: 3170602]

31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242. [PubMed: 10592235]

32. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J. Mol. Biol 1993;233:123–138. [PubMed: 8377180]

33. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol 2007;372:774–797. [PubMed: 17681537]

34. Kopycki JG, Stubbs MT, Brandt W, Hagemann M, Porzel A, Schmidt J, Schliemann W, Zenk MH, Vogt T. Functional and structural characterization of a cation-dependent *O*-methyltransferase from the cyanobacterium *Synechocystis sp*. strain PCC 6803. J. Biol. Chem 2008;283:20888–20896. [PubMed: 18502765]

35. Ferrer JL, Zubieta C, Dixon RA, Noel JP. Crystal structures of alfalfa caffeoyl coenzyme A 3-*O*-methyltransferase. Plant Physiol 2005;137:1009–1017. [PubMed: 15734921]
36. Boehr DD, Dyson HJ, Wright PE. An NMR perspective on enzyme dynamics. Chem. Rev 2006;106:3055–3079. [PubMed: 16895318]
37. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 2007;450:913–916. [PubMed: 18026087]
38. Miles EW, Rhee S, Davies DR. The molecular basis of substrate channeling. J. Biol. Chem 1999;274:12193–12196. [PubMed: 10212181]
39. Kopycki JG, Rauh D, Chumanevich AA, Neumann P, Vogt T, Stubbs MT. Biochemical and structural analysis of substrate promiscuity in plant Mg^{2+}-dependent *O*-methyltransferases. J. Mol. Biol 2008;378:154–164. [PubMed: 18342334]
40. Stevenson CEM, Freel Meyers CL, Walsh CT, Lawson DM. Crystallization and preliminary X-ray analysis of the *O*-methyltransferase NovP from the novobiocin-biosynthetic cluster of *Streptomyces spheroides*. Acta Cryst 2007;F63:236–238.
41. Kortagere S, Ekins S, Welsh WJ. Halogenated ligands and their interactions with amino acids: Implications for structure-activity and structure-toxicity relationships. J. Mol. Graphics Modell 2008;27:170–177.
42. Ramos A, Lombo F, Brana AF, Rohr J, Mendez C, Salas JA. Biosynthesis of elloramycin in *Streptomyces olivaceus* requires glycosylation by enzymes encoded outside the aglycon cluster. Microbiology 2008;154:781–788. [PubMed: 18310024]
43. Torkkell S, Kunnari T, Palmu K, Mantsala P, Hakala J, Ylihonko K. The entire nogalamycin biosynthetic gene cluster of *Streptomyces nogalater*: characterization of a 20-kb DNA region and generation of hybrid structures. Mol. Genet. Genomics 2001;266:276–288. [PubMed: 11683270]
44. Ward SL, Hu Z, Schirmer A, Reid R, Revill WP, Reeves CD, Petrakovsky OV, Dong SD, Katz L. Chalcomycin biosynthesis gene cluster from *Streptomyces bikiniensis*: novel features of an unusual ketolide produced through expression of the *chm* polyketide synthase in *Streptomyces fradiae*. Antimicrob. Agents Chemother 2004;48:4703–4712. [PubMed: 15561847]
45. Menéndez N, Nur-e-Alam M, Braña AF, Rohr J, Salas JA, Méndez C. Biosynthesis of the antitumor chromomycin A_3 in *Streptomyces griseus*: analysis of the gene cluster and rational design of novel chromomycin analogs. Chemistry & Biology 2004;11:21–32. [PubMed: 15112992]
46. Seno ET, Baltz RH. Properties of *S*-adenosyl-L-methionine:macrocin *O*-methyltransferase in extracts of *Streptomyces fradiae* strains which produce normal or elevated levels of tylosin and in mutants blocked in specific *O*-methylations. Antimicrob. Agents Chemother 1981;20:370–377. [PubMed: 7305323]
47. Zubieta C, He XZ, Dixon RA, Noel JP. Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant *O*-methyltransferases. Nat. Struct. Biol 2001;8:271–279. [PubMed: 11224575]
48. Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. J. Mol. Recognit 1996;9:1–5. [PubMed: 8723313]
49. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4, AutoDockTools4: Automated docking with selective receptorflexibility. J. Comput. Chem. 2009
50. Palma PN, Rodrigues ML, Archer M, Bonifacio MJ, Loureiro AI, Learmonth DA, Carrondo MA, Soares-da-Silva P. Comparative study of ortho- and meta-nitrated inhibitors of catechol-*O*-methyltransferase: interactions with the active site and regioselectivity of *O*-methylation. Mol. Pharmacol 2006;70:143–153. [PubMed: 16618795]
51. Menéndez N, Nur EAM, Braña AF, Rohr J, Salas JA, Méndez C. Tailoring modification of deoxysugars during biosynthesis of the antitumour drug chromomycin A_3 by *Streptomyces griseus ssp. griseus*. Mol. Microbiol 2004;53:903–915. [PubMed: 15255901]
52. Kominek LA, Sebek OK. Biosynthesis of novobiocin and related coumarin antibiotics. Dev. Ind. Microbiol 1974;15:60–69.
53. Usón I, Stevenson CEM, Lawson DM, Sheldrick GM. Structure determination of the *O*-methyltransferase NovP using the 'free lunch algorithm' as implemented in SHELXE. Acta Cryst 2007;D63:1069–1074.

54. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst 1997;D53:240–255.
55. Lamzin VS, Wilson KS. Automated refinement of protein models. Acta Cryst 1993;D49:129–147.
56. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Cryst 2004;D60:2126–2132.
57. Read RJ. Improved fourier coefficients for maps using phases from partial structures with errors. Acta Cryst 1986;A42:140–149.
58. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:W375–W383. [PubMed: 17452350]
59. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Cryst 2002;D58:1948–1954.
60. DeLano, WL. The PyMOL User's Manual. San Carlos, CA, USA: DeLano Scientific; 2002.
61. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Cryst 1998;D54:905–921.
62. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 1999;15:305–308. [PubMed: 10320398]
63. Hou X, Wang Y, Zhou Z, Bao S, Lin Y, Gong W. Crystal structure of SAM-dependent *O*-methyltransferase from pathogenic bacterium *Leptospira interrogans*. J. Struct. Biol 2007;159:523–528. [PubMed: 17561415]
64. Cho JH, Park Y, Ahn JH, Lim Y, Rhee S. Structural and functional insights into *O*-methyltransferase from *Bacillus cereus*. J. Mol. Biol 2008;382:987–997. [PubMed: 18706426]
65. Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, Notredame C. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D–Coffee. Nucleic Acids Res 2006;34:604–608.

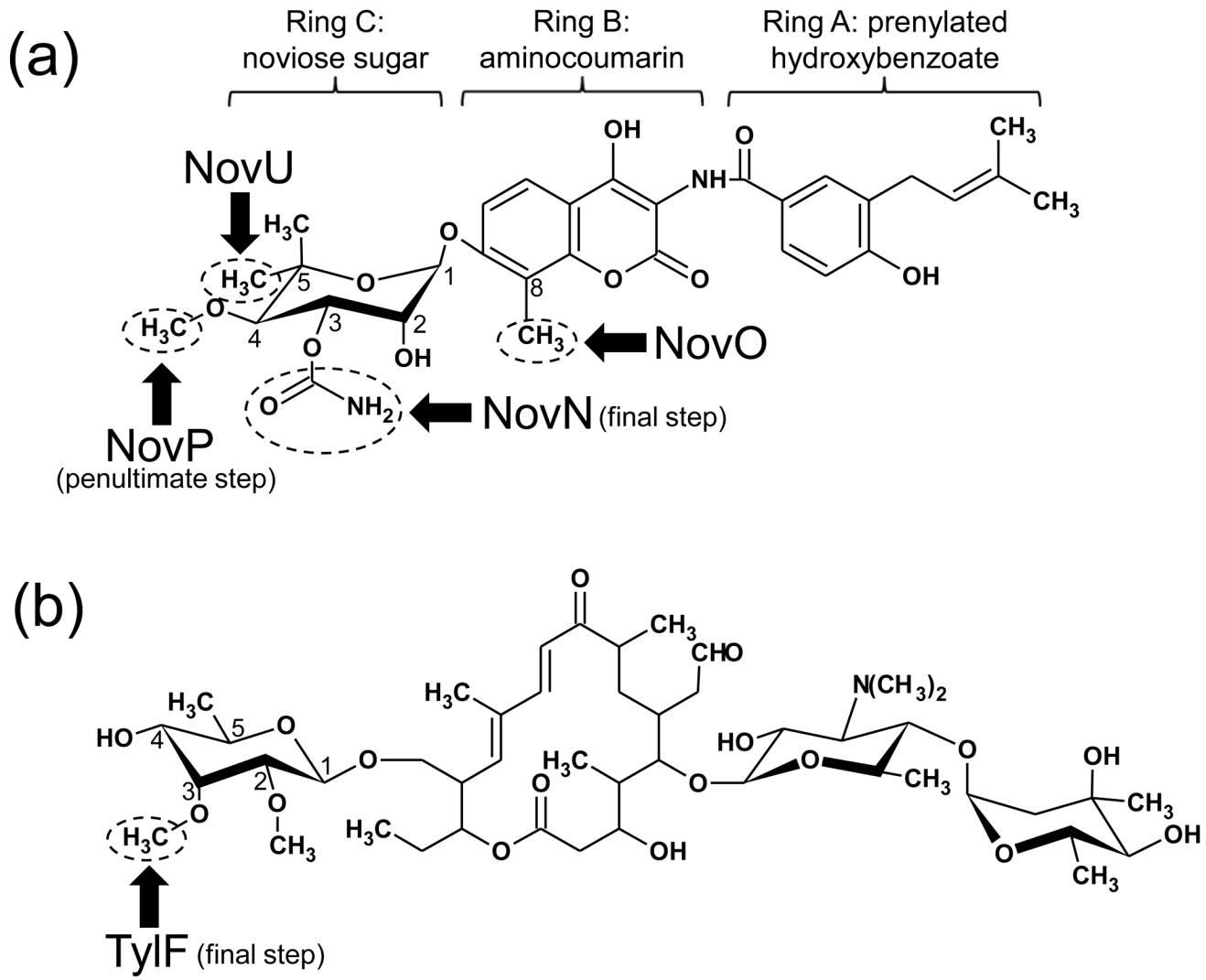

Fig. 1.

The structures of novobiocin and tylosin. (a) Novobiocin is an aminocoumarin antibiotic comprised of three ring systems, which are synthesised separately. The three methylation steps are indicated together with the final *O*-carbamoylation by NovN. NovO and NovU act during the synthesis of rings B and C, respectively. NovP methylates the 4-OH group of the L-noviose moiety when the antibiotic is almost fully formed; this represents the penultimate step of the whole pathway. The substrate for NovP, desmethyldescarbamoyl novobiocin (DDN) bears only hydroxyl groups at positions 3 and 4 of the noviose sugar. Rings B and C are identical in the substrates of the NovP orthologs CouP and CloP, with the exception that in the substrate for CloP, the C8 position of ring B is chlorinated rather than methylated. (b) Tylosin is a macrolide antibiotic. TylF performs the terminal step of the pathway by methylating the 3-OH group of 6-deoxy-2-*O*-methyl-D-allose.

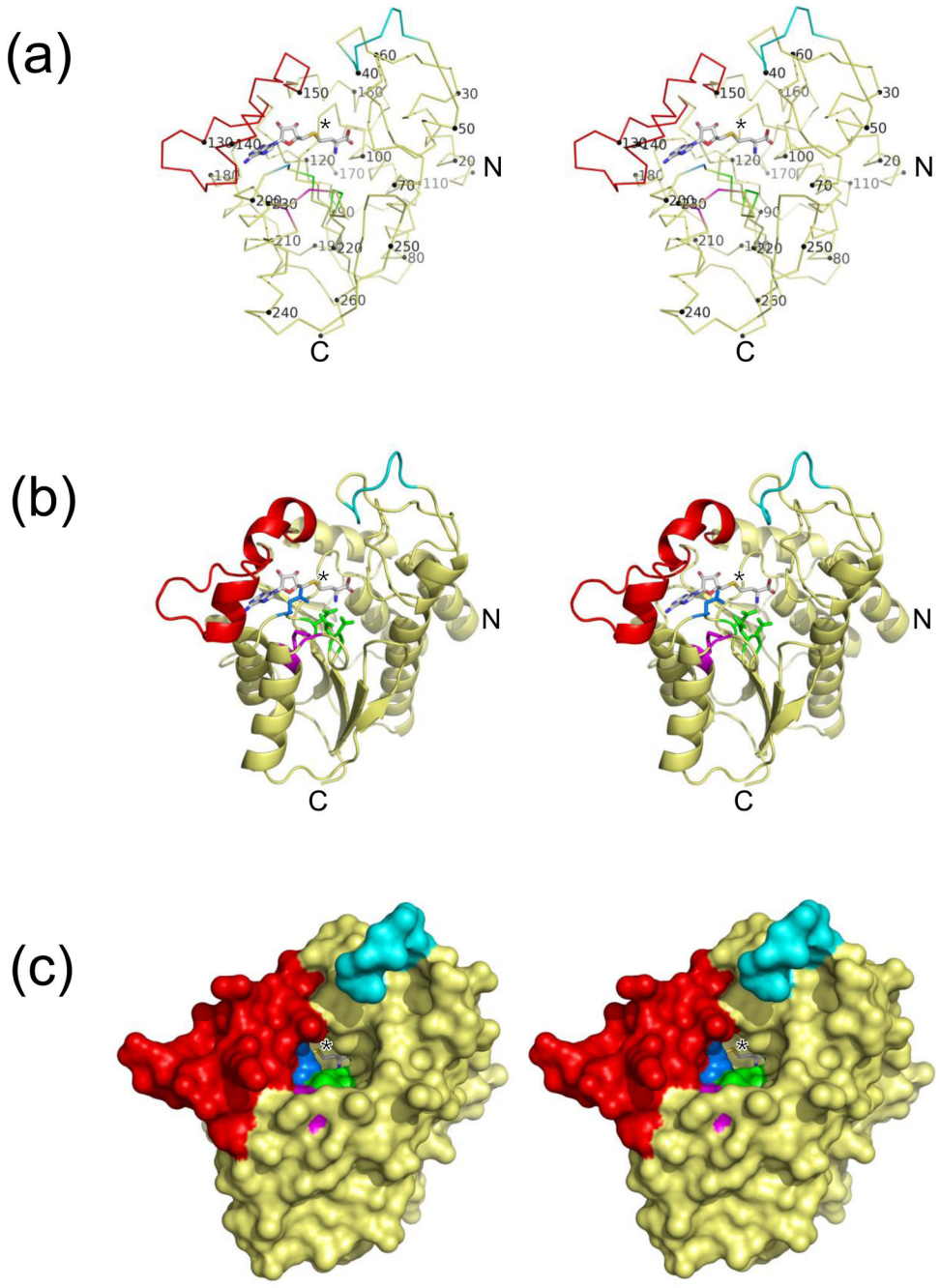

Fig. 2.

Stereoviews showing the overall structure of NovP. (a) Backbone trace with every tenth residue labelled. (b) Cartoon representation with important residues shown in stick mode. (c) Molecular surface representation illustrating the solvent inaccessibility of the SAH and the shape of the active site channel. Throughout, the lid region is coloured red, and the disordered loop (residues 39–45 inclusive, which have been modelled into energetically favourable conformations) is shown in cyan. The 3-Asp putative metal-binding site is coloured green and Asp198, the proposed general base, is shown in blue. The disulphide bridge residues, Cys228 and Cys231, are coloured magenta. The SAH ligand is shown in stick representation and the position of the sulphur atom is indicated by an asterisk.

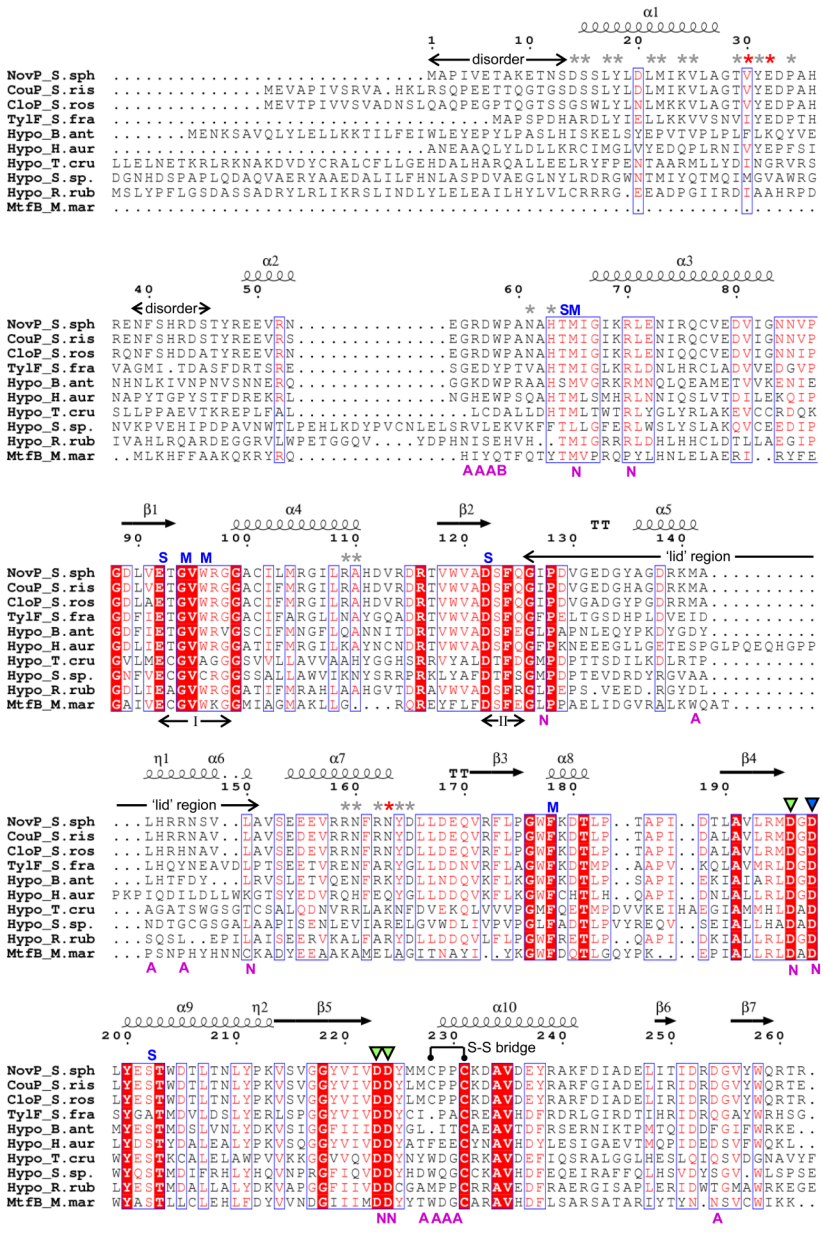

Figure 3a

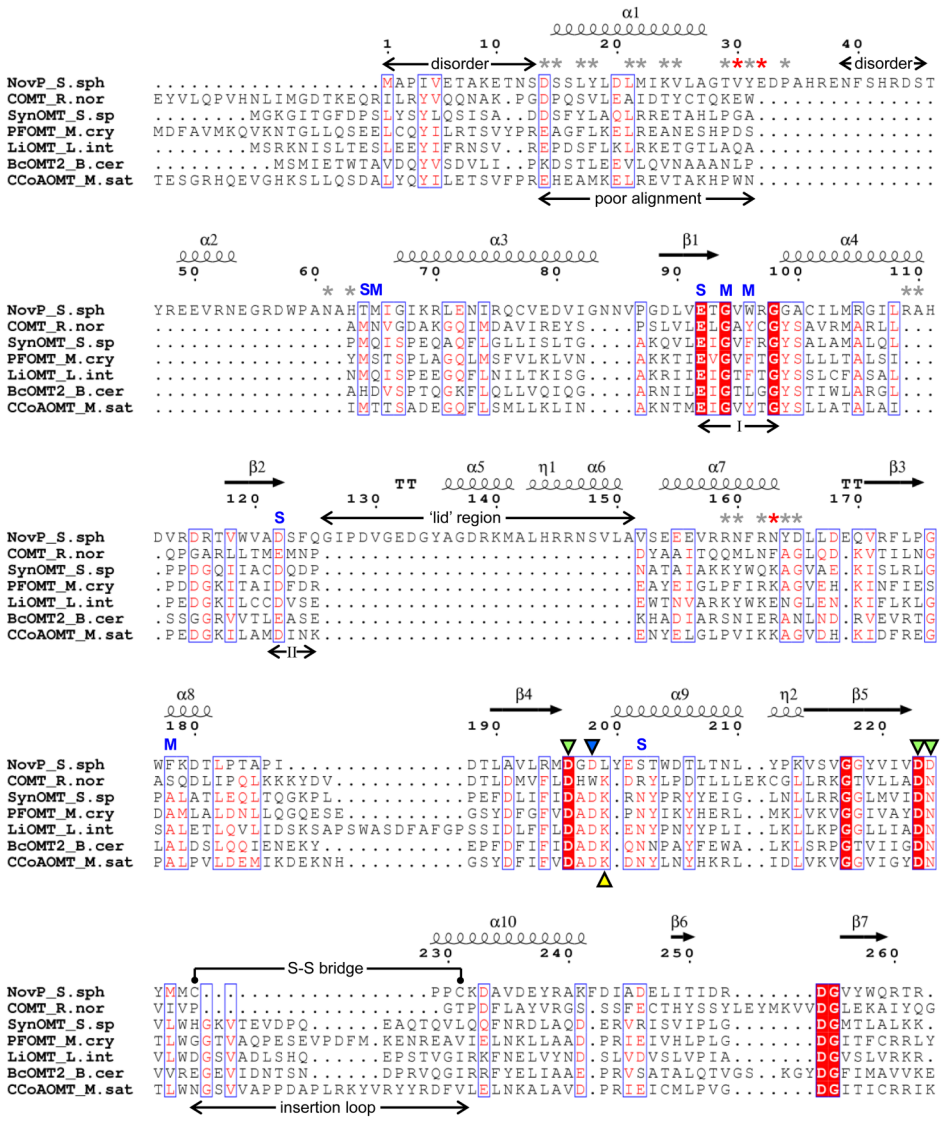

Figure 3b

Fig. 3.

Multiple protein sequence alignments versus NovP. For both panels the alignment is displayed using ESPript.[62] Strictly conserved residues are highlighted with red shaded boxes, and well conserved residues are boxed with the predominant residues coloured red. Secondary structure elements for NovP are shown above the alignment, where α = α helix, β = β strand, η = 3_{10} helix, TT = β turn. Residues that are postulated to be involved in cation binding in NovP are marked by green triangles; Asp198, which is thought to act as a general base during the methyltransfer reaction, is indicated by the single blue triangle. Residues that are hydrogen bonded to SAH are indicated by dark blue letters where "S" refers to a side-chain interaction and "M" refers to a main-chain interaction. Although Asp196 is marked by a green triangle,

indicating a role in cation binding, it also hydrogen bonds to SAH. Residues that form the NovP dimer interface are marked with asterisks, and these are coloured red for those residues that form hydrogen bonds. Other features of the NovP structure are labelled, i.e. the lid region, regions of disorder, the disulphide bridge, and the SAM-dependent methyltransferase motifs I and II. (a) Alignment of NovP with selected sequences from the TylF superfamily. The first four are from Streptomycetes (phylum Actinobacteria). They are: NovP from *S. spheroides* (NovP_S.sph), CouP from *S. rishiriensis* (CouP_S.ris), CloP from *S, roseochromogenes subsp. oscitans* (CloP_S.ros), and TylF from *S. fradiae* (TylF_S.fra). The remainder are representative sequences from the other six phyla which constitute the TylF superfamily, where the prefix "Hypo" indicates an uncharacterised or hypothetical protein. They are from: *Bacillus anthracis* (Hypo_B.ant; phylum Firmicutes), *Herpetosiphon aurantiacus* ATCC 23779 (Hypo_H.aur; phylum Chloroflexi), *Trypanosoma cruzi* (Hypo_T.cru; phylum Euglenozoa), *Synechococcus sp*. (strain JA-3-3Ab) (Hypo_S.sp.; phylum cyanobacteria), *Rhodospirillum rubrum* (Hypo_R.rub; phylum Proteobacteria) and *Microscilla marina* ATCC 23134 (MtfB_M.Mar; phylum Bacteroidetes). The magenta letters below the alignment indicate residues that delineate the substrate access channel, where "N" refers to those adjacent to the expected position of the noviose sugar (as predicted by the docking simulations), "A" refers to those adjacent to the expected position of the aminocoumarin ring, and "B" refers to the single residue, Trp58, that lies adjacent to both moieties. (b) Structure-based multiple sequence alignment of *Streptomyces spheroides* NovP (NovP_S.sph) with selected *O*-methyltransferase sequences for which there are crystal structures. Shown are the sequences from rat (COMT_R.nor),[50] *Synechocystis* sp. strain PCC 6803 (SynOMT_S.sp),[34] *Mesembryanthemum crystallinum* (PFOMT_M.cry),[39] *Leptospira interrogans* (LiOMT_L.int),[63] *Bacillus cereus* (BcOMT2_B.cer)[64] and *Medicago sativa* (CCoAOMT_M.sat).[35] These are also summarised in Table 1. The initial alignment was generated using EXPRESSO[65] and subsequently adjusted manually with reference to the superposed structures. Indicated below the alignment is the N-terminal region of poor structural alignment (which includes helix α1), and the insertion loop that occurs in some of the structural homologues, but not NovP itself. Note that, Asp198, the putative general base (blue triangle), appears to be well conserved in the structural homologues of NovP. However, in these structures the protein backbone directs the Asp side-chain away from the active site. Instead, the side-chain of the next residue, a Lys in all but NovP, points into the active site and functions as the base. This is indicated by the yellow triangle shown below the alignment. This difference in main-chain conformation may be a consequence of the single residue insertion at the N-terminus of helix α9 in NovP.

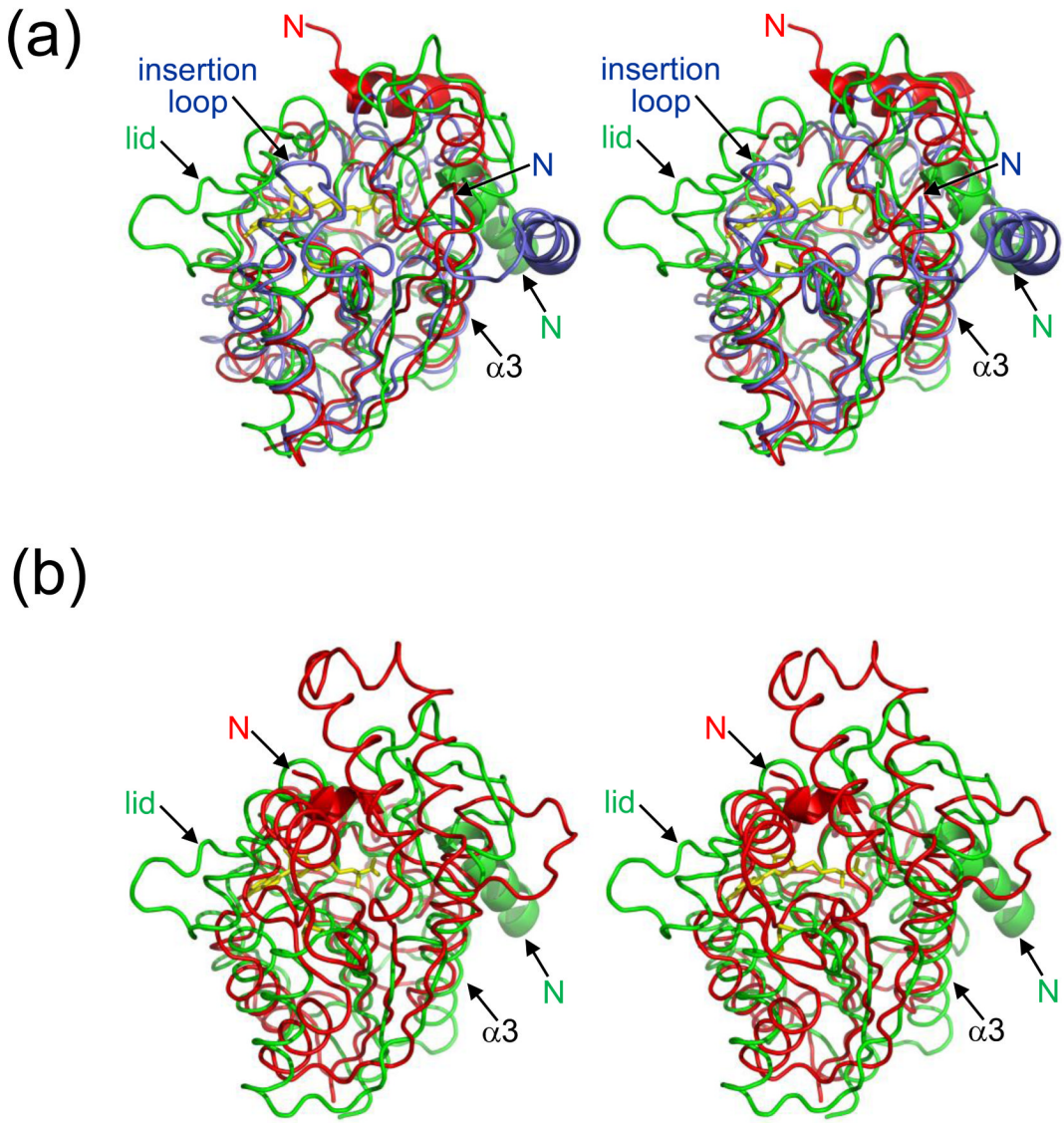

Fig. 4.

Stereoviews showing superpositions of NovP with structural homologues. (a) Comparison of NovP (green) with two metal-dependent OMTs: COMT (red; PDB code 2CL5) and CCoAOMT (blue; PDB code 1SUS). (b) Comparison of NovP (green) with the metal-independent enzyme RebM (red; PDB code 3BUS). The N-termini are highly variable between all four structures shown until α3 of NovP (labelled) where they all converge. The protein backbones are depicted in ribbon representation, except for their N-terminal helices, which are highlighted in cartoon representation; the N-termini are marked with colour-coded labels. Also shown in yellow are the SAH molecule and disulphide bridge of NovP. The view is equivalent to that in Fig. 2.

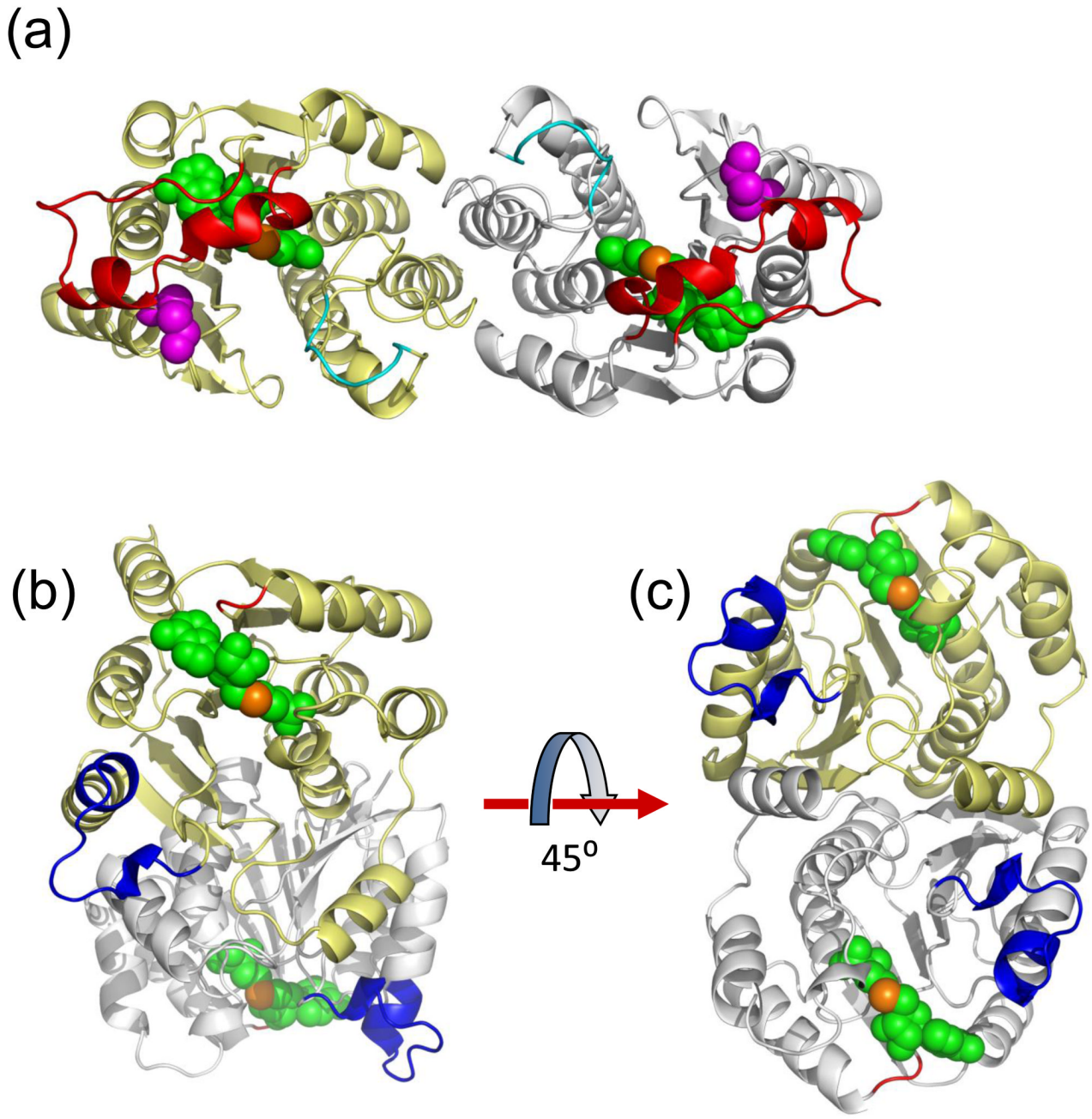

Fig. 5.

NovP forms a new type of dimer. (a) The relatively elongated NovP dimer as viewed down the crystallographic 2-fold axis. The monomers are shown in cartoon representation with the lid region in red and the disordered loop in cyan. Depicted as van der Waals spheres are the bound SAH ligand (green with the sulphur atom in orange) and the disulphide bridge residues (magenta). (b) The more compact SynOMT dimer (PDB code 3CBG). The yellow subunit is shown in the equivalent orientation to the yellow subunit of NovP shown in (a). Note the absence of a lid region; the much shorter loop that follows β2 is shown in red. The 13-residue insertion relative to NovP that includes the so-called 'insertion loop' is shown in blue. This is in the structurally equivalent position to the disulphide bridge of NovP. Rotating the SynOMT

dimer by 45° around the horizontal axis gives a view approximately looking down the dimer 2-fold axis as shown in (c).

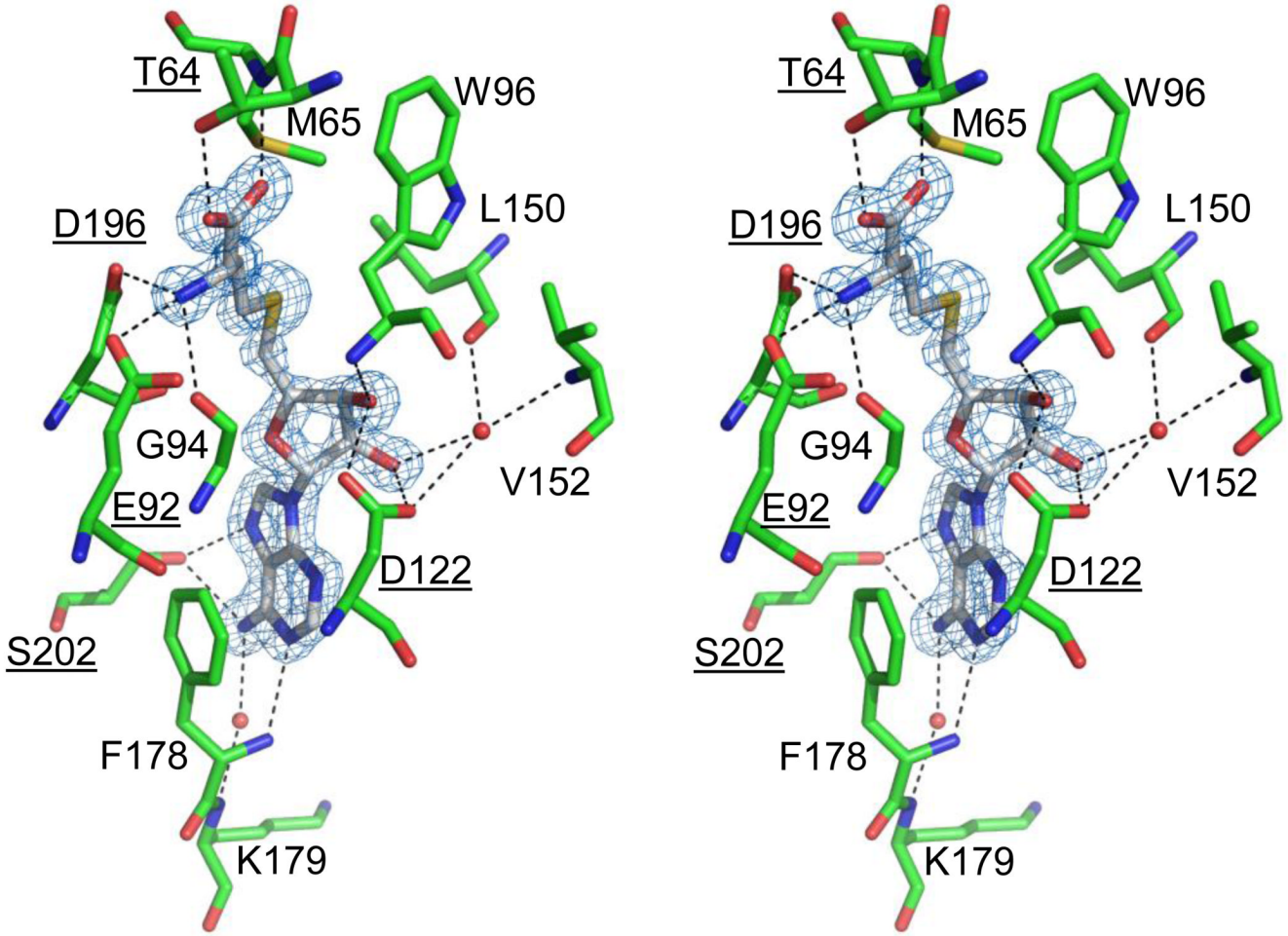

Fig. 6.

Stereoview showing the co-substrate binding site of NovP. A simulated annealing 1.4 Å resolution omit map for SAH contoured at 6σ is shown in blue; the ligand is shown with grey carbon atoms. Also shown, with green carbon atoms, are residues that are hydrogen-bonded to the ligand either directly, or through single buried water molecules. Residue labels are underlined where the interactions involve side-chains and are therefore sequence specific. The SAH forms several non-bonded interactions with the protein, for example, a clear π-stacking interaction is observed between Phe178 and the adenine ring of SAH. However, for clarity, residues that form only hydrophobic interactions with SAH are not shown.

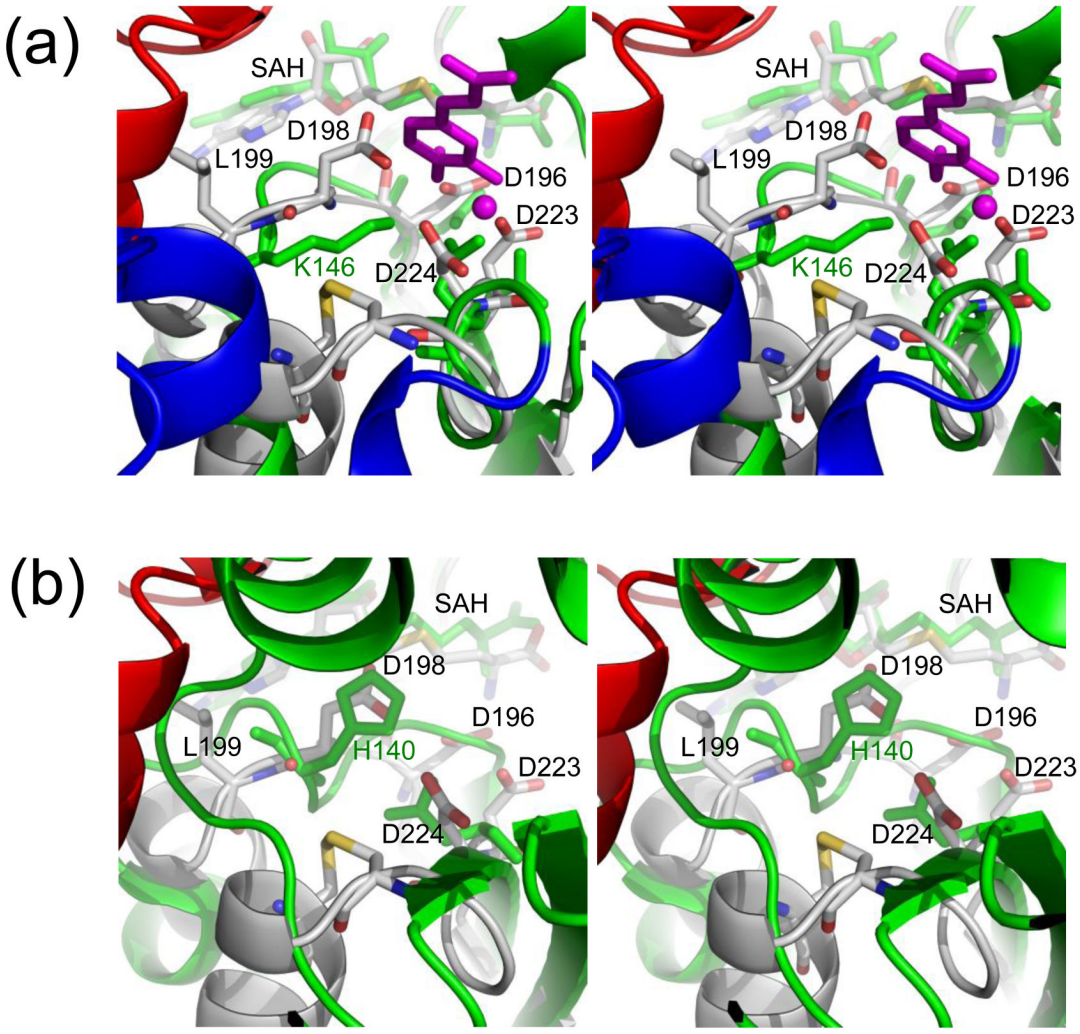

Fig. 7.

Stereoviews showing overlays of the NovP active site with those of other OMTs. In both panels the protein backbones are shown in cartoon representation with NovP mainly in grey and the second protein mainly in green. Key residues and ligands are shown in stick representation. For NovP the atoms are in CPK colours, and for the second protein, all atoms are in green. The labels in black refer to NovP; the green labels refer to the active site general base of the second protein. Also shown is the disulphide bridge of NovP. (a) Comparison of NovP and SynOMT (PDB code 3CBG). Note the clash between the NovP lid region (red) and the insertion loop of SynOMT (blue). Also shown is the bound product (isoferulic acid) and the Mg^{2+} ion, which are both coloured magenta. (b) Comparison of NovP and RebM (PDB code 3BUS).

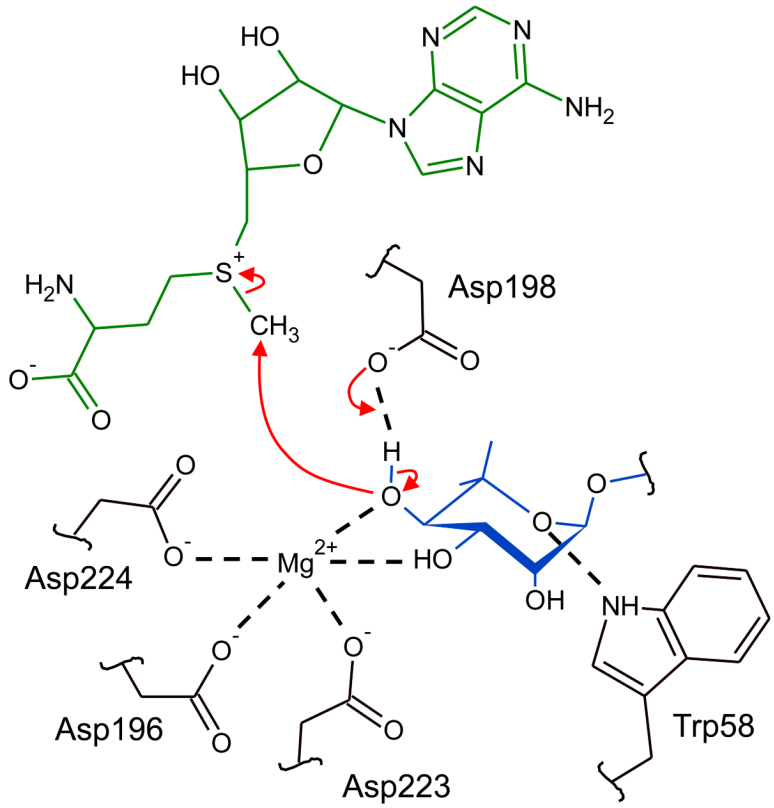

Fig. 8.

Proposed NovP mechanism. Comparisons with NovP structural homologues suggest that Asp198 acts as the active site base which deprotonates the 4-OH of the noviose moiety (shown with blue bonds). The putative 3-Asp cation binding site may serve to assist the deprotonation as well as to correctly orient the substrate for methyltransfer from SAM (shown with green bonds). All the Asp residues are strictly conserved in TylF superfamily members, whilst Trp58 is found in over 50% of the sequences. We would expect the Mg^{2+} to be octahedrally coordinated, with the sixth coordination site most likely occupied by a water molecule; this has been omitted for clarity.

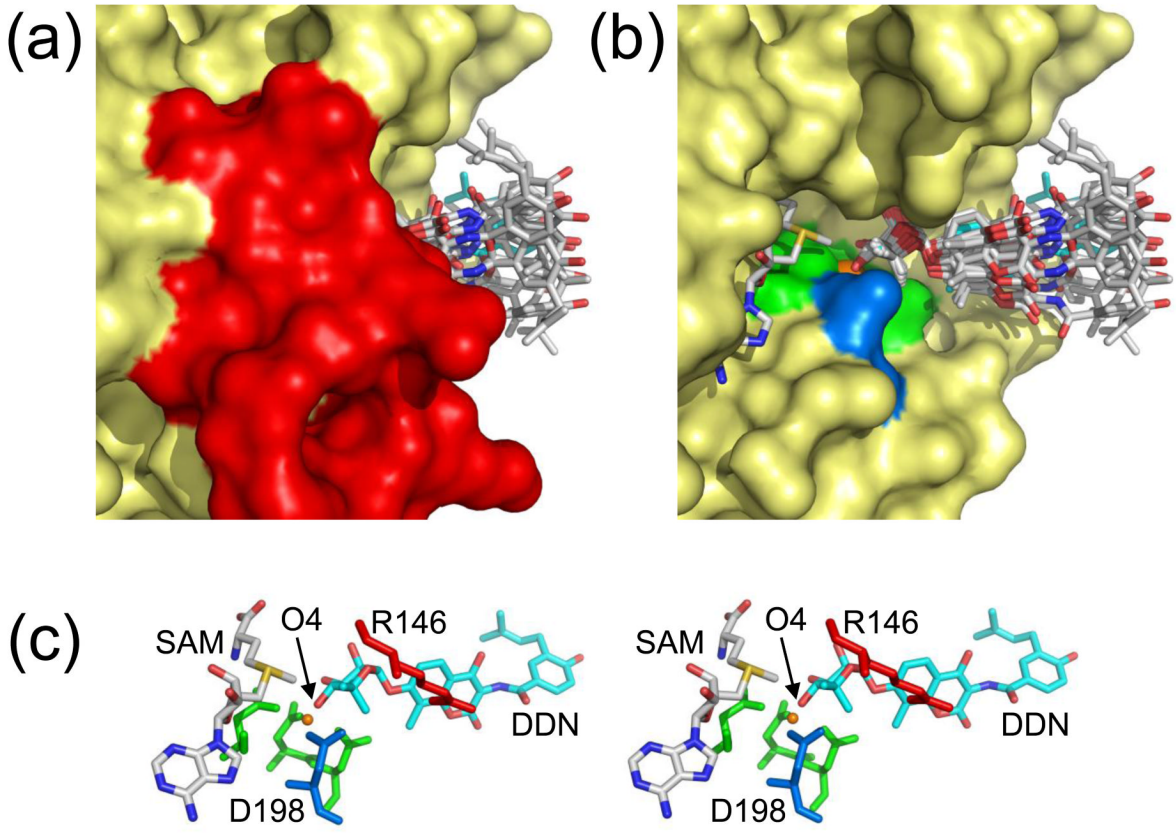

Fig. 9.

Substrate docking simulations. Surface representations of NovP with (a) lid present and (b) lid removed. Colouring is as for Fig. 2. Displayed in stick form are the lowest energy poses from each of the 'productive' 12 clusters from run 2 of the AutoDock4 simulations (i.e. mimicking an activated enzyme-substrate complex). The lowest energy pose for the whole simulation is shown with cyan carbons. (c) Stereoview showing the detail of the quaternary complex involving the lowest energy pose. The Mg^{2+} ion is shown as an orange sphere. Also shown in red is the position of Arg146 which may have a role in interacting with the substituent at C8 of the aminocoumarin ring. Note that the orientation of the noviose moiety is inverted relative to that shown in Fig. 1 and Fig. 8.

Table 1

Summary of selected NovP structural homologues

Protein	Source	PDB code	Resolution (Å)	Biological unit	DALI Z score[a]	Rmsd (Å)[a]	Identity (%)[a]	Co-substrate bound	Metal bound	Ligand bound	Reference
NovP	*S. spheroides*	2WK1	1.40	dimer	n/a	0	100	SAH	none	none	this work
COMT	*R. norvegicus*	2CL5	1.60	monomer	16.7	2.7	13	SAM	Mg^{2+}	inhibitor	50
SynOMT	*Syn. sp. strain PCC 6803*	3CBG	2.00	dimer	16.3	3.1	11	SAH	Mg^{2+}	product	34
PFOMT	*M. crystallinum*	3C3Y	1.37	dimer	16.3	3.1	14	SAH	Ca^{2+}	none	39
LiOMT	*L. interrogans*	2HNK	2.30	dimer	16.3	2.8	13	SAH	none	none	63
BcOMT2	*B. cereus*	3DUW	1.20	dimer	16.1	2.4	14	SAH	none	none	64
CCoAOMT	*M. sativa*	1SUS	2.65	dimer	15.4	2.9	12	SAH	Ca^{2+}	substrate	35
RebM	*L. aerocolonigens*	3BUS	2.65	dimer	10.1	3.1	11	SAH	none[b]	none	5

[a] All comparisons are versus the NovP structure and sequence.

[b] Not a metal-dependent enzyme.

Table 2

Summary of X-ray data and model parameters for NovP

Data collection	
Space Group	P2
Cell parameters (Å/°)	a = 51.81, b = 46.04, c = 61.22, β = 104.97
Wavelength (Å)	1.488
Resolution range[a] (Å)	36.3-1.40 (1.42-1.40)
Unique reflections	54638
Completeness[a] (%)	98.8 (89.5)
Redundancy	3.4
R_{merge}[a,b]	0.057 (0.164)
$\langle I \rangle / \langle \sigma I \rangle$[a]	18.1 (4.2)
Wilson B value (Å^2)	17.6
Refinement	
R_{cryst}[c] (based on 95% of data; %)	14.5
R_{free}[c] (based on 5% of data; %)	16.2
DPI[d] (based on R_{free}; Å)	0.048
Ramachandran favoured/allowed[e] (%)	96.6/100.0
Ramachandran outliers[e]	0
rmsd bond distances (Å)	0.013
rmsd bond angles (º)	1.540
Contents of model (molecules/non-hydrogen atoms)	
Protein (residues/atoms)	242/1965
S-adenosyl-L-homocysteine (molecules/atoms)	1/26
Sulphate	2/10
Ethylene glycol	2/8
Waters	242
Average temperature factors (Å^2)	
Main-chain atoms	14.6
Side-chain atoms	18.1
S-adenosyl-L-homocysteine	12.0
Sulphate	28.4
Ethylene glycol	25.2
Waters	32.3
Overall	18.1

[a] The figures in brackets indicate the values for outer resolution shell.

[b] $R_{merge} = \Sigma_h \Sigma_l |I_{hl} - \langle I_h \rangle| / \Sigma_h \Sigma_l \langle I_h \rangle$, where I_l is the l^{th} observation of reflection h and $\langle I_h \rangle$ is the weighted average intensity for all observations l of reflection h.

[c] The R-factors R_{cryst} and R_{free} are calculated as follows: $R = \Sigma(|F_{obs} - F_{calc}|)/\Sigma|F_{obs}| \times 100$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[d] Estimate of coordinate errors based on R_{free} as calculated in REFMAC5.[54]

[e] As calculated using MOLPROBITY.[58]