Structural and functional dissection of aminocoumarin antibiotic biosynthesis: a review

David M. Lawson[*] & Clare E. M. Stevenson

Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

*Corresponding author email: david.lawson@jic.ac.uk

Abstract   Aminocoumarin antibiotics are natural products of soil-dwelling bacteria called Streptomycetes. They are potent inhibitors of DNA gyrase, an essential bacterial enzyme and validated drug target, and thus have attracted considerable interest as potential templates for drug development. To date, aminocoumarins have not seen widespread clinical application on account of their poor pharmacological properties. Through studying the structures and mechanisms of enzymes from their biosynthetic pathways we will be better informed to redesign these compounds through rational pathway engineering. Novobiocin, the simplest compound, requires at least seventeen gene products to convert primary metabolites into the mature antibiotic. We have solved the crystal structures of four diverse biosynthetic enzymes from the novobiocin pathway, and used these as three-dimensional frameworks for the interpretation of functional and mechanistic data, and to speculate about how they might have evolved. The structure determinations have ranged from the routine to the challenging, necessitating a variety of different approaches.

Summary   This review describes progress towards understanding the structures and functions of enzymes from aminocoumarin antibiotic biosynthesis pathways. We focus on the different approaches used to solve these crystal structures and what they can tell us about their mechanisms and evolution.

Keywords   Antibiotic biosynthesis; Enzyme evolution; Structure determination; Streptomyces; X-ray crystallography

Abbreviations

PDB   Protein Data Bank

ASU   Asymmetric unit

NCS   Non-crystallographic symmetry

SAM   *S*-adenosyl-L-methionine

SAH   *S*-adenosyl-L-homocysteine

SIRAS   Single isomorphous replacement with anomalous scattering

Introduction

The aminocoumarin class of antibiotics are natural products of Streptomycetes, which are ubiquitous soil-dwelling bacteria, and the source of the majority of naturally-derived antibiotics in current clinical use [1]. The interest in aminocoumarins stems from their ability to inhibit DNA gyrase [2], an essential enzyme in bacteria and a validated drug target [3]. They bind to the B subunit of gyrase and competitively inhibit the ATPase activity of the enzyme [4]. The three classical aminocoumarins, novobiocin (Fig. 1), clorobiocin and coumermycin A_1, are biosynthesised by three separate *Streptomyces* species, and 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) [5, 6, 7, 8]. A further type of aminocoumarin, termed the simocyclinones, will not be considered here.

Figure 1

Despite their potency against DNA gyrase, the aminocoumarins have poor pharmacological properties, displaying limited solubility in water, low penetration into Gram negative bacteria and moderate toxicity to humans. Nevertheless, novobiocin is licensed for the treatment of some Gram positive organisms including methicillin-resistant *Staphylococcus aureus* [9, 10] and *Borrelia burgdorferi* [11] (the causative agent of Lyme's disease), and shows synergistic effects with anti-tumour drugs, such as etoposide or teniposide [12], through the inhibition of multidrug resistance proteins. DNA topoisomerase IV is an additional target for the aminocoumarins [13], and the development of agents that specifically target this enzyme is currently unexploited. Moreover, until comparatively recently, DNA gyrase was thought to be restricted to prokaryotes. However, the enzyme has been identified in plants too [14], raising the possibility that aminocoumarins could also be developed for use as herbicides.

Using a combination of structural and mechanistic approaches, we have sought to understand the biosynthetic pathways for the aminocoumarins in sufficient detail to facilitate the rational redesign of the products through the manipulation of the individual enzymes in the pathway. Armed with this knowledge we hope to enable the production of novel aminocoumarin derivatives that retain the ability to inhibit DNA gyrase, but have enhanced pharmacokinetics. We have chosen to focus on novobiocin, as this is the simplest of the three aminocoumarins. Nevertheless, its biosynthesis requires at least seventeen gene products [5]. Towards these goals, we have solved the crystal structures of four diverse enzymes. These structure determinations, ranging from the routine to the challenging, will be reviewed herein, and the functional and mechanistic implications of the resulting models will be discussed.

The sugar nucleotide epimerase NovW

NovW is a sugar nucleotide epimerase from *Streptomyces spheroides* involved in the biosynthesis of the L-noviosyl sugar moiety of novobiocin prior to its attachment to the aminocoumarin ring (Fig. 1) [5]. Whilst the enzyme is a chemically competent 3,5-epimerase *in vitro* [15], we showed by kinetic characterisation and deuterium incorporation analysis, that kinetically it serves only as a 3-epimerase (Fig. 2A) [16]. We have thus functionally assigned it as a dTDP-6-deoxy-D-*xylo*-4-hexulose 3-epimerase.

Of the four structure determinations presented here, NovW was by far the most straightforward. The protein crystallized in the eleventh drop of the very first crystallization trial and, after one round of optimization, yielded crystals of sufficient quality to collect X-ray data in-house to 2.0 Å resolution [17]. The structure was solved by molecular replacement with AMoRe [18] (Fig. 2B) using a poly-Ala template derived from the structure of *Methanobacterium thermoautotrophicum* RmlC (PDB ID 1EP0), a sugar nucleotide 3,5-

5

epimerase, with which NovW shares 44% amino acid sequence identity. There was a single copy of the 24.4 kDa subunit in the asymmetric unit (ASU) of the $P4_32_12$ cell (a = b = 59.4 Å, c = 109.0 Å), giving a solvent content of 37%. The biologically relevant dimer is generated via the application of two-fold crystallographic symmetry. A further native data set collected at the synchrotron enabled the structure to be built with COOT [19] and refined with REFMAC5 [20] to 1.6 Å resolution, giving R_{work} and R_{free} values of 0.150 and 0.195, respectively for the final model (PDB entry 2C0Z; Fig. 2C) [21].

Figure 2

The NovW monomer displays a cupin barrel fold [22] within which resides the active centre (Fig. 2C). By analogy with the closely-related RmlC enzymes [23], the key amino acids are a His-Asp residue pair that together deprotonate at C3 of the sugar from one face of the ring, a Lys residue that stabilises an enolate oxyanion intermediate, and a Tyr residue that reprotonates at C3 from the other face of the ring. It has been speculated that the orientation of the Tyr side-chain may determine whether a particular enzyme is a single or a double epimerase [24], and it adopts distinctly different conformations in NovW to those observed in several RmlC structures that correlate with double epimerase activity [16, 21].

The aromatic prenyltransferase CloQ

CloQ is the prenyltransferase that attaches a dimethylallyl group to the hydroxybenzoate moiety of clorobiocin in *Streptomyces roseochromogenes* (Fig. 3A) [25]. The identical reaction is performed by NovQ in the biosynthesis of novobiocin by *S. spheroides* (Fig. 1) [5]. We were unable to crystallize NovQ, whereas CloQ yielded diffraction quality crystals [26] and thus was selected for further analysis; the two enzymes share 85% amino acid

sequence identity. CloQ belongs to an unusual family of soluble (i.e. not membrane bound) aromatic prenyltransferases and, at the time of this study, there was only one representative structure in the Protein Data Bank (PDB), namely that of NphB from *Streptomyces sp.* CL190 (PDB ID 1ZB6) [27]. However, the latter is only 16% identical to CloQ at the amino acid level, and thus was not a suitable molecular replacement template. X-ray data were collected to resolutions of 2.2 Å and 2.4 Å for native and iodide derivative (1 M potassium iodide) crystals, respectively, using in-house facilities only. The crystals belong to space group $I4_122$ (cell: a = b = 135.2 Å, c = 98.1 Å), and the solvent content was estimated at 60% for a single copy of the 35.6 kDa molecule per ASU. For the iodide derivative, the high symmetry of the space group coupled with the 310° rotation during data collection resulted in a highly redundant dataset (27.3 fold). Experimental phases were determined by the single isomorphous replacement with anomalous scattering method (SIRAS) (Fig. 2B) by combining the native and derivative data in the SHELX suite of programs [28]. SHELXD found twelve significant iodide sites, which gave a figure of merit of 0.552 to 2.5 Å resolution after phasing in SHELXE. This value rose to 0.900 after density modification with DM [29]. Through a combination of autobuilding with ARP/wARP [30], manual building (COOT) and restrained refinement (REFMAC5), a final model was produced consisting of 97% of the expected residues with corresponding R_{work} and R_{free} values of 0.165 and 0.212 to 2.2 Å resolution, respectively (PDB ID 2XLQ; Fig. 3C) [31].

Figure 3

CloQ is currently one of only four representative structures of the soluble aromatic prenyltransferase family in the PDB. In addition to the structure of NphB [27], two further structures, both from the fungus *Aspergillus fumigatus*, have now been deposited, namely that of dimethylallyl tryptophan synthetase (PDB ID 3I4X) [32], and that of brevianamide F prenyltransferase (PDB ID 3O2K) [33]. All four structures share an unusual $(\beta/\alpha)_{10}$ barrel

fold that differs from the ubiquitous $(\beta/\alpha)_8$ TIM-barrel fold [34] in that the β-strands are antiparallel, and the barrel has a central cavity as a consequence of the greater barrel radius (Fig. 3C). The cavity corresponds to the active centre, which has a distinctly hydrophobic region for binding the aromatic substrate, and a polar region for binding the prenyl donor via its pyrophosphate group. In NphB, a magnesium ion is required for prenyl donor binding, whereas CloQ and the two fungal enzymes are magnesium-independent enzymes that make use of a cluster of positively charged side-chains to bind the pyrophosphate group instead of a divalent metal ion. The reaction is thought to proceed via a Friedel-Crafts-type mechanism whereby the polar region of the central cavity promotes the cleavage of the pyrophosphate group from the prenyl donor to form an allylic carbocation, whilst premature quenching of the latter is prevented by the solvent-free environment created by the non-polar region. The carbocation then performs an electrophilic substitution on the aromatic ring of the acceptor substrate to give a σ-complex, which subsequently yields the product upon hydrogen elimination [31]. We have determined the CloQ structure together with its aromatic acceptor substrate 4-hydroxyphenyl pyruvate (Fig. 3C). Intriguingly, this is covalently linked to an active site Cys residue to give a thiohemiketal species. However, substitution of the Cys to either Ser or Ala gave mutant proteins that retained 100% wild-type activity *in vitro*. Thus, the significance of the covalent link remains a mystery [31]. We and others further showed that CloQ [31], and its ortholog NovQ [35], can utilise alternative substrates. In many cases prenylation appears to provide a higher level of bioactivity compared to non-prenylated precursors, often by increasing affinity for biological membranes and favouring interactions with cellular targets [36]. Therefore these enzymes offer promise as versatile catalysts for the prenylation of a range of aromatic substrates to produce a variety of potentially useful compounds.

The non-heme iron oxygenase NovR

Based on its 95% amino acid sequence identity with the characterised enzyme CloR [37], NovR from *S. spheroides* is assumed to be a non-heme iron oxygenase that performs two consecutive oxidative decarboxylations on the product of the NovQ reaction to yield the prenylated hydroxybenzoate moiety of novobiocin (Fig. 4A), prior to its attachment to the aminocoumarin ring (Fig. 1). Native X-ray data to 2.1 Å resolution were recorded at the synchrotron in the monoclinic space group *C*2 (cell: a = 86.7 Å, b = 139.4 Å, c = 100.8 Å, β = 101.2°) [38]. Besides CloR, NovR shares no recognisable sequence homology with other characterised non-heme iron oxygenases. Instead, it displays homology to type-II aldolases, which are zinc-dependent enzymes (PFAM entry PF00596) [39]. Interrogation of the PDB for potential molecular replacement templates retrieved the structure of L-fuculose-1-phosphate aldolase as the closest match (PDB ID 1E4C) [40]. However, the sequence identity was only 21% over just 79% of the NovR sequence, placing it in the so-called "twilight zone" [41], where molecular replacement is notoriously difficult [42]. Solvent content estimations suggested that between two and five copies of the 30.5 kDa NovR protomer could be accommodated in the ASU giving solvent contents in the range 75-37%. Nevertheless, three pieces of evidence all pointed to there being four copies per ASU: (1) dynamic light scattering gave a molecular size of 113 kDa, being close to the value of 122 kDa calculated for four molecules, (2) a self-rotation function calculated using MOLREP [43] showed clear four-fold non-crystallographic symmetry (NCS) [38], (3) the biological unit of L-fuculose-1-phosphate aldolase is a tetramer [40]. A molecular replacement monomer template was constructed from the aldolase structure using the CHAINSAW program [44] based on a careful alignment with the NovR sequence. A tetramer template was also generated by superimposing this monomer template onto each of the monomers in the aldolase tetramer. Both the monomer and tetramer templates were used as input to AMoRe, but only the

tetramer gave a plausible solution. This placed the tetramer in the ASU such that it made few clashes with its symmetry neighbours and, crucially, was orientated such that the NCS four-fold axis was consistent with the axis identified in the self-rotation function. However, the resultant model had R_{work} and R_{free} values of 0.542 and 0.544 to 2.1 Å resolution, respectively, and electron density maps phased on this model were heavily biased and difficult to interpret. Despite these disappointing statistics, we reasoned that the backbone trace was likely to be a fairly good approximation to the true structure and, importantly, that the quaternary structure was essentially correct, because the tetramer model gave a convincing molecular replacement solution. Based on this premise, we developed a density modification procedure involving four-fold averaging, solvent flattening, and very gradual extension from the low resolution phases, which we assumed to be reasonably reliable. This was implemented using DM [29] starting from molecular replacement phases at 5 Å resolution up to the maximum resolution of 2.1 Å in 1000 steps. This yielded a high quality electron density map that enabled 89% of the expected residues to be fitted automatically using ARP/wARP [30] with corresponding R_{work} and R_{free} values of 0.195 and 0.259 to 2.1 Å resolution, respectively (Fig. 4B). Subsequent tests with fewer DM cycles showed that a similar result could be obtained with 500 cycles, whereas ARP/wARP was unable to fit any residues using the density modified phases after only 400 cycles [38].

Figure 4

The NovR structure contains four NCS equivalent metal ions, which lie at subunit interfaces and correspond to the active site zinc atoms in the template aldolase structure (Fig. 4C). However, metal analysis of NovR protein samples indicated that the metal ions were most likely nickel ions derived from the metal affinity purification step [38]. Full details of the final structure and its interpretation will be described elsewhere.

The *O*-methyltransferase NovP

NovP is one of three *S*-adenosyl-L-methionine (SAM)-dependent methyltransferases encoded in the novobiocin biosynthetic cluster of *S. spheroides*. Its role is to methylate the 4-OH of the noviose moiety (Figs. 1 and 5A) as the penultimate step of the pathway [5, 45], and the resultant methoxy group is likely important for the potency of the mature antibiotic, based on the structure of novobiocin bound to its target [46]. Interrogation of the PFAM database [39] places NovP in the TylF family (PFAM entry PF05711), where TylF is a macrocin *O*-methyltransferase [47]. At the time of this study, there were no structures for any TylF family member in the PDB, and the closest match found by the FUGUE server (http://tardis.nibio.go.jp/fugue/prfsearch.html) [48] was that of a catechol *O*-methyltransferase (PDB ID 1VID) giving a Z score of 11.0 and a sequence identity of only 16%. Native NovP crystals diffracted strongly, enabling a dataset to be collected to 1.4 Å resolution at the synchrotron [49]. The space group was *P*2 (cell: a = 51.8 Å, b = 46.0 Å, c = 61.2 Å, β = 105.0°) and the ASU was expected to contain one copy of the 32 kDa molecule giving an estimated solvent content of 44%. This high resolution native dataset was subsequently merged with three further native datasets, two of which were collected at the synchrotron and one in-house, to resolutions in the range 1.9 - 2.1 Å, such that the combined dataset was 100% complete to 1.98 Å resolution excepting just five reflections at very low resolution. For phasing, a 2.45 Å resolution dataset was collected in-house from a crystal soaked in 1 mM mercury acetate [50]. Although the wavelength of the in-house facility (1.542 Å) was not ideal for the measurement of anomalous dispersion from mercury, there was a measurable anomalous signal. Unfortunately, the native and derivative data were poorly isomorphous at higher resolutions. Nevertheless, using the SIRAS approach at 3.5 Å resolution, SHELXD located three heavy atom sites, and SHELXE generated phases

sufficient to yield a low resolution map with reasonable connectivity. Although ARP/wARP was unable to autobuild into this map, it was possible to manually place the structure of catechol *O*-methyltransferase (PDB ID 1VID) into the density using XtalView [51] such that the core of the structure gave a satisfactory fit to the map and relatively few intermolecular clashes with symmetry neighbours. Using this model as a guide, polyalanine secondary structural elements were placed into the map and optimized by real-space refinement to yield a new model that covered just 42% of the expected residues in the native NovP sequence. Next, phases calculated from this model were combined with the SIRAS phases and input to SHELXE, this time employing the so-called "free lunch" algorithm to extrapolate both amplitudes and phases to 1.0 Å. In this method, the missing data are essentially invented: phases for these "new" reflections are calculated by density modification [52], whilst their amplitudes are calculated from a Fourier transformation of the density-modified map normalised to fit an extrapolated Wilson plot. The resultant phases were truncated back to 1.4 Å resolution and used to calculate a very clear electron density map that enabled ARP/wARP to trace 91% of the NovP sequence (Fig. 5B) [50]. Although it is difficult to pinpoint exactly why the free lunch algorithm was so beneficial, one or more of the following could be important factors. Firstly, phases are more important than amplitudes (see: http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html), so provided these are essentially correct, errors in the extrapolated amplitudes will be tolerated. Secondly, extending the resolution will tend to correct Fourier truncation errors, which could otherwise have deleterious effects on the phases. Finally, the availability of a high quality near-atomic resolution data set that was essentially complete to 2 Å resolution (despite the low symmetry space group), was undoubtedly an important factor [50]. Further rounds of rebuilding (COOT) and refinement (REFMAC5) resulted in a model that was 92% complete with R_{work}

and R_{free} values of 0.145 and 0.162 to 1.4 Å resolution, respectively (PDB entry 2WK1; Fig. 5C) [53].

Figure 5

The NovP structure displays a typical class I methyltransferase fold [54] and suggests that the biological unit may be an extended homodimer [53]. The enzyme was co-crystallized with its desmethylated co-substrate, *S*-adenosyl-L-homocysteine (SAH), which is almost completely buried by a helical "lid" region that gates access to the co-substrate binding pocket (Fig. 5C). This lid is characteristic of the TylF superfamily, and it appears as an insertion in sequence alignments with more distantly-related *O*-methyltransferases [53]. The active centre contains a three-Asp putative metal-binding site; a further well conserved Asp likely acts as the general base that initiates the reaction by deprotonating the 4-OH group of the noviose moiety. Thereafter, the strongly electrophilic methyl group of SAM would transfer directly to the 4-OH group in an S_N2-like mechanism to give the product, descarbamoylnovobiocin. *In silico* docking was used to generate models of the enzyme-substrate complex that were consistent with this proposed mechanism. These models suggested that NovP would be intolerant of significant modifications at the noviose moiety, but would show increasing substrate promiscuity as a function of the distance of the modification from the methylation site [53]. The latter conclusion is consistent with experiments using artificial substrates with different groups substituting for the prenylated hydroxybenzoate moiety (ring A; Fig. 1), which are accepted as substrates [55].

Conclusions

Many of the enzymes implicated in secondary metabolite production will most likely have evolved from progenitors in primary metabolism. NovW belongs to the cupin superfamily [21] that is both widespread in all kingdoms of life, and diverse in biological function, and includes primary metabolic enzymes such as phosphomannose isomerase [22]. By contrast, the ancestry of CloQ is difficult to discern, since comparatively few proteins have been recognised as having this fold (PFAM lists only eighteen; PF11468), only a handful have been characterised, and they appear to be restricted to bacteria and fungi [56, 57, 58]. It seems unlikely that these soluble aromatic prenyltransferases have evolved from enzymes with the ubiquitous $(\beta/\alpha)_8$ TIM-barrel fold [34], since they have completely different connectivities. On the other hand, NovR represents a clear example of where secondary metabolism appears to have borrowed a protein fold from primary metabolism. In addition to sharing the same fold with type II aldolases, NovR adopts the same tertiary structure, assembling as a homotetramer with C_4 symmetry [38]. Moreover, NovR has retained the metal binding site, albeit having switched from zinc dependence to iron dependence [37]. Methyltransferase structures are well-represented in the PDB, accounting for roughly 1% of all entries (as of June 2009). The vast majority, including NovP, are SAM-dependent enzymes and have a Rossmann nucleotide binding core fold [59], and therefore NovP could potentially have evolved from any enzyme that shares this fold, such as the NAD-dependent dehydrogenases that are central to primary metabolism.

In this short review, we have presented models of four structurally diverse enzymes. Equally diverse are the methods we have used to determine these structures, with examples of how persistence and creativity can pay off with challenging crystallographic problems. This work has provided deep insights into the structures, mechanisms and evolution of these enzymes, and may offer prospects for rational engineering of aminocoumarin biosynthesis. The final enzyme of the novobiocin pathway, the carbamoyltransferase NovN [45], has also been

crystallized [60]. Details of the structure determination process and the resultant NovN model will be described elsewhere.

References

1. Bentley SD et al (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417(6885): 141-147

2. Gellert M et al (1976) Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. PNAS 73(12): 4474-4478

3. Maxwell A (1997) DNA gyrase as a drug target. Trends Microbiol 5:102-109

4. Maxwell A, Lawson DM (2003) The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. Curr Top Med Chem 3(3): 283-303

5. Steffensky M et al (2000) Identification of the novobiocin biosynthetic gene cluster of *Streptomyces spheroides* NCIB 11891. Antimicrob Agents Chemother 44(5): 1214-1222.

6. Pojer F, Li SM, Heide L (2002) Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics. Microbiology 148(12): 3901-3911

7. Wang ZX, Li SM, Heide L (2000) Identification of the coumermycin A(1) biosynthetic gene cluster of *Streptomyces rishiriensis* DSM 40489. Antimicrob Agents Chemother 44(11): 3040-3048.

8. Heide L (2009) The aminocoumarins: biosynthesis and biology. Nat Prod Rep 26(10): 1241-1250

9. Walsh TJ et al (1993) Randomized double-blinded trial of rifampin with either novobiocin or trimethoprim-sulfamethoxazole against methicillin-resistant *Staphylococcus aureus* colonization: prevention of antimicrobial resistance and effect of host factors on outcome. Antimicrob Agents Chemother 37(6): 1334-1342

10. Montecalvo MA et al (1995) Effect of novobiocin-containing antimicrobial regimens on infection and colonization with vancomycin-resistant *Enterococcus faecium*. Antimicrob Agents Chemother 39(3): 794.

11. Samuels DS, Garon CF (1993) Coumermycin A$_1$ inhibits growth and induces relaxation of supercoiled plasmids in *Borrelia burgdorferi*, the Lyme disease agent. Antimicrob Agents Chemother 37(1): 46-50

12. Lorico A, Rappa G, Sartorelli AC (1992) Novobiocin-induced accumulation of etoposide (VP-16) in WEHI-3B D+ leukemia cells. Int J Cancer 52(6): 903-909

13. Peng H, Marians KJ (1993) *Escherichia coli* topoisomerase IV. J Biol Chem 268:24481-24490

14. Wall MK, Mitchenall LA, Maxwell A (2004) *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. PNAS 101(20): 7821-7826

15. Thu Thuy TT et al (2005) Functional characterizations of *novWUS* involved in novobiocin biosynthesis from *Streptomyces spheroides*. Arch Biochem Biophys 436(1): 161-167

16. Tello M et al (2006) Characterisation of *Streptomyces spheroides* NovW and revision of its functional assignment to a dTDP-6-deoxy-D-*xylo*-4-hexulose 3-epimerase. Chem Commun (10): 1079-1081

17. Jakimowicz P et al (2003) Crystallization and preliminary X-ray studies on the putative dTDP sugar epimerase NovW from the novobiocin biosynthetic cluster of *Streptomyces spheroides*. Acta Crystallogr Sect D 59(Pt 8): 1507-1509

18. Navaza J (1994) AMoRe: an automated package for molecular replacement. Acta Crystallogr Sect A 50(Pt2): 157-163

19. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D 60(Pt 12): 2126-2132

20. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr Sect D 53(Pt3): 240-255

21. Jakimowicz P et al (2006) The 1.6-A resolution crystal structure of NovW: a 4-keto-6-deoxy sugar epimerase from the novobiocin biosynthetic gene cluster of *Streptomyces spheroides*. Proteins 63(1): 261-265

22. Dunwell JM et al (2001) Evolution of functional diversity in the cupin superfamily. Trends Biochem Sci 26(12): 740-746

23. Dong C et al (2003) High-resolution structures of RmlC from *Streptococcus suis* in complex with substrate analogs locate the active site of this class of enzyme. Structure 11(6): 715-723

24. Merkel AB et al (2004) The position of a key tyrosine in dTDP-4-Keto-6-deoxy-D-glucose-5-epimerase (EvaD) alters the substrate profile for this RmlC-like enzyme. J Biol Chem 279(31): 32684-32691

25. Pojer F et al (2003) CloQ, a prenyltransferase involved in clorobiocin biosynthesis. PNAS 100(5): 2316-2321

26. Keller S et al (2006) Crystallization and preliminary X-ray analysis of the aromatic prenyltransferase CloQ from the clorobiocin biosynthetic cluster of *Streptomyces roseochromogenes*. Acta Crystallogr Sect F 62(Pt 11): 1153-1155

27. Kuzuyama T, Noel JP, Richard SB (2005) Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. Nature 435(7044): 983-987

28. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr Sect A 64(Pt 1): 112-122

29. Cowtan K (1994) DM: an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31:34-38

30. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. Nature Struct Biol 6(5): 458-463

31. Metzger U et al (2010) Structure and mechanism of the magnesium-independent aromatic prenyltransferase CloQ from the clorobiocin biosynthetic pathway. J Mol Biol 404(4): 611-626

32. Metzger U et al (2009) The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. PNAS 106(34): 14309-14314

33. Jost M et al (2010) Structure-function analysis of an enzymatic prenyl transfer reaction identifies a reaction chamber with modifiable specificity. J Am Chem Soc 132(50): 17849-17858

34. Gerlt JA, Raushel FM (2003) Evolution of function in (b/a)$_8$-barrel enzymes. Curr Opin Chem Biol 7(2): 252-264

35. Ozaki T et al (2009) NovQ is a prenyltransferase capable of catalyzing the addition of a dimethylallyl group to both phenylpropanoids and flavonoids. J Antibiot (Tokyo) 62(7): 385-392

36. Botta B et al (2005) Novel prenyltransferase enzymes as a tool for flavonoid prenylation. Trends Pharmacol Sci 26(12): 606-608

37. Pojer F et al (2003) CloR, a bifunctional non-heme iron oxygenase involved in clorobiocin biosynthesis. J Biol Chem 278(33): 30661-30668

38. Keller S et al (2006) Molecular replacement in the "twilight zone": structure determination of the non-haem iron oxygenase NovR from *Streptomyces spheroides* through repeated density modification of a poor molecular-replacement solution. Acta Crystallogr Sect D 62(Pt 12): 1564-1570

39. Finn RD et al (2008) The Pfam protein families database. Nucleic Acids Res 36(Database issue): D281-288

40. Joerger AC, Mueller-Dieckmann C, Schulz GE (2000) Structures of L-fuculose-1-phosphate aldolase mutants outlining motions during catalysis. J Mol Biol 303(4): 531-543

41. Rost B (1999) Twilight zone of protein sequence alignments. Protein Eng 12(2): 85-94

42. Claude J-B et al (2004) CaspR: a web server for automated molecular replacement using homology modelling. Nucl Acids Res 32(suppl_2): W606-609

43. Vagin A, Teplyakov A (2000) An approach to multi-copy search in molecular replacement. Acta Crystallogr Sect D 56(Pt 12): 1622-1624

44. Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. J Appl Crystallogr 41(3): 641-643

45. Freel Meyers CL et al (2004) Characterization of NovP and NovN: completion of novobiocin biosynthesis by sequential tailoring of the noviosyl ring. Angew Chem Int Ed Engl 43(1): 67-70

46. Lewis RJ et al (1996) The nature of inhibition of DNA gyrase by the coumarins and the cyclothialidines revealed by X-ray crystallography. EMBO J 15(6): 1412-1420

47. Fouces R et al (1999) The tylosin biosynthetic cluster from *Streptomyces fradiae*: genetic organization of the left region. Microbiology 145(4): 855-868

48. Shi J, Blundell TL, Mizuguchi K (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 310(1): 243-257

49. Stevenson CEM et al (2007) Crystallization and preliminary X-ray analysis of the *O*-methyltransferase NovP from the novobiocin-biosynthetic cluster of *Streptomyces spheroides*. Acta Crystallogr Sect F 63(Pt 3): 236-238

50. Usón I et al (2007) Structure determination of the *O*-methyltransferase NovP using the 'free lunch algorithm' as implemented in SHELXE. Acta Crystallogr Sect D 63(Pt 10): 1069-1074

51. McRee DE (1999) XtalView/Xfit - a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 125(2-3): 156-165

52. Sheldrick GM (2002) Macromolecular phasing with SHELXE. Z Kristallogr 217:644-650

53. Gómez García I et al (2010) The crystal structure of the novobiocin biosynthetic enzyme NovP: the first representative structure for the TylF *O*-methyltransferase superfamily. J Mol Biol 395(2): 390-407

54. Schubert HL, Blumenthal RM, Cheng X (2003) Many paths to methyltransfer: a chronicle of convergence. Trends Biochem Sci 28(6): 329-335

55. Anderle C et al (2007) Improved mutasynthetic approaches for the production of modified aminocoumarin antibiotics. Chem Biol 14(8): 955-967

56. Tello M et al (2008) The ABBA family of aromatic prenyltransferases: broadening natural product diversity. Cell Mol Life Sci 65(10): 1459-1463

57. Saleh O et al (2009) Prenyl transfer to aromatic substrates in the biosynthesis of aminocoumarins, meroterpenoids and phenazines: the ABBA prenyltransferase family. Phytochemistry 70(15-16): 1728-1738

58. Haug-Schifferdecker E et al (2010) A new group of aromatic prenyltransferases in fungi, catalyzing a 2,7-dihydroxynaphthalene 3-dimethylallyl-transferase reaction. J Biol Chem 285(22): 16487-16494

59. Rossmann MG et al (1975) Evolutionary relationships among dehydrogenases. In *The Enzymes* (Boyer, I.P.D., ed), pp. 61-102, Academic Press

60. Gómez García I et al (2008) Crystallization and preliminary X-ray analysis of the *O*-carbamoyltransferase NovN from the novobiocin-biosynthetic cluster of *Streptomyces spheroides*. Acta Crystallogr Sect F 64(Pt 11): 1000-1002
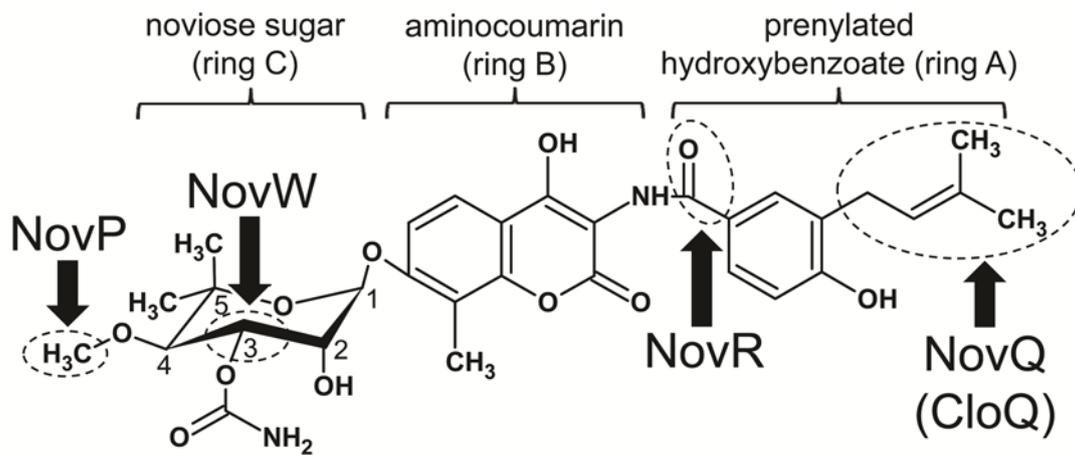
Fig. 1 The structure of the aminocoumarin antibiotic novobiocin. The arrows and dotted ovals indicate where the various enzymes act. With the exception of NovP, which catalyses the penultimate step of the pathway, all the other enzymes act on the precursors of rings A or C prior to their attachment to the aminocoumarin ring.
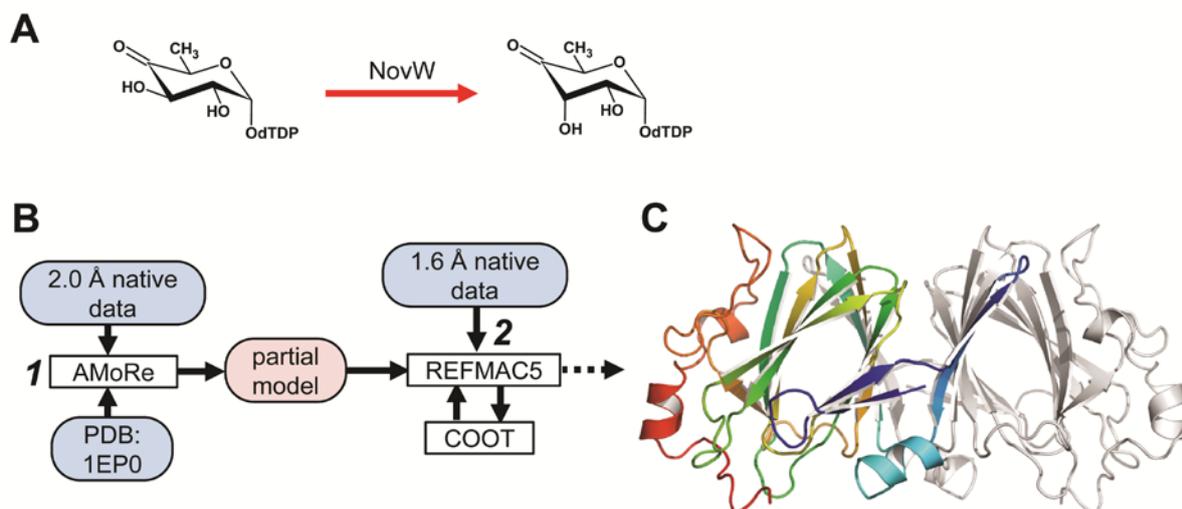
Fig. 2 The sugar nucleotide epimerase NovW. (A) The reaction catalysed by NovW *en route* to the production of the noviose sugar moiety of novobiocin (ring C, Fig. 1). (B) Summary of the structure solution process for NovW. The text in ovals indicates various forms of data, where input data are highlighted with a pale blue background, and intermediate data are highlighted with a pale pink background; programs are represented by text in rectangles, and the italic numbers indicate the order in which the programs were run. For step 2, refinement with REFMAC5 and model building with COOT were performed iteratively until no further improvement was seen. (C) The 1.6 Å resolution structure of NovW in cartoon representation. The asymmetric unit contains only a single NovW subunit; the biologically relevant homodimer shown is generated by the application of two-fold crystallographic symmetry. One subunit is in rainbow colours (N-terminus in dark blue through to C-terminus in red) and the other subunit is coloured in grey.
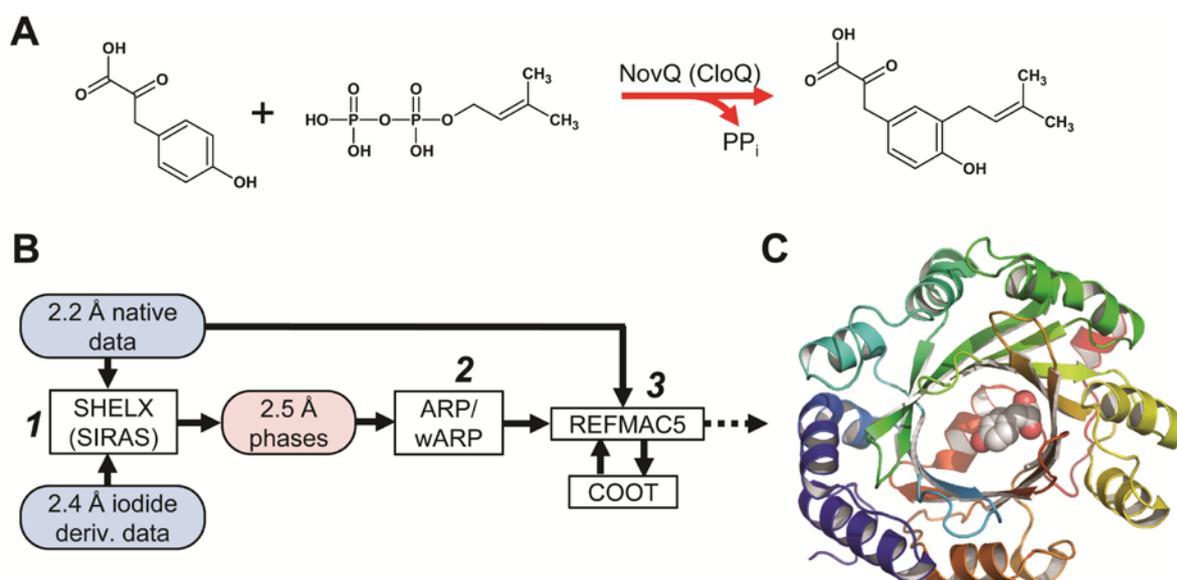
Fig. 3 The aromatic prenyltransferase CloQ. (A) The reaction catalysed by CloQ *en route* to the production of the prenylated hydroxybenzoate moiety of novobiocin (ring A, Fig. 1), where PP$_i$ indicates pyrophosphate. (B) Summary of the structure solution process for CloQ. (C) The 2.2 Å resolution structure of CloQ in cartoon representation. The bound substrate, 4-hydroxyphenyl pyruvate is also shown in space-filling representation in CPK colours. Otherwise the colour schemes for (B) and (C) are as for Fig. 2.
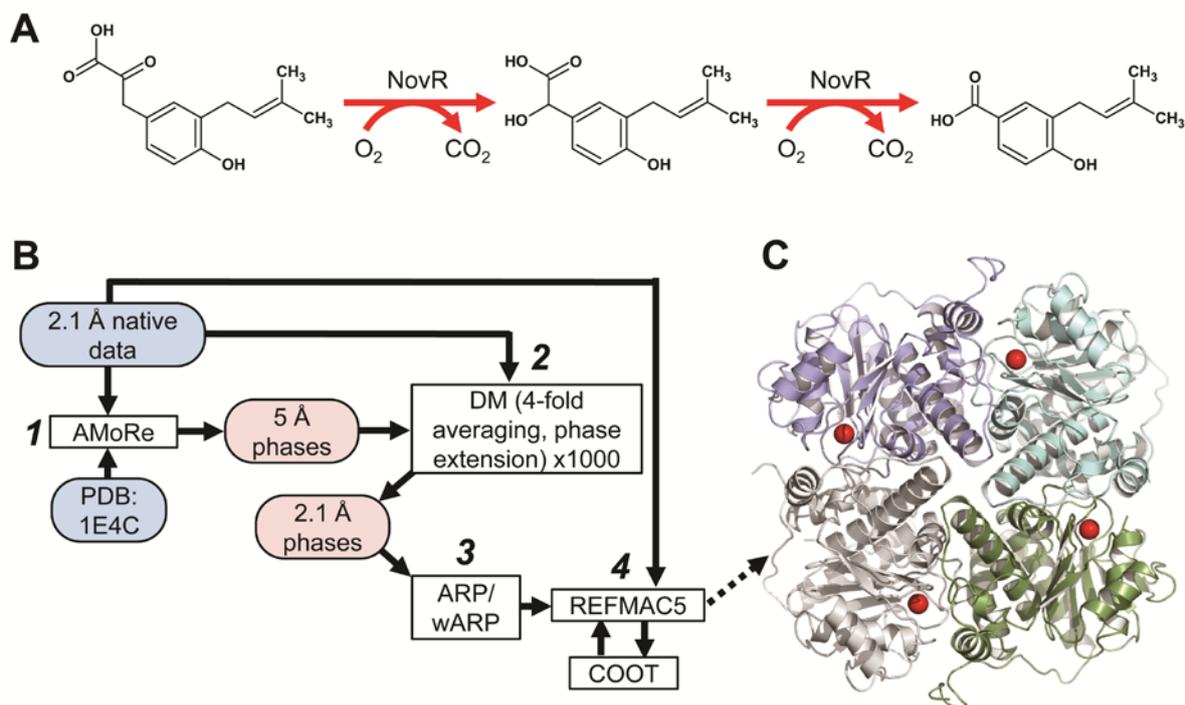
Fig. 4 The non-heme iron oxygenase NovR. (A) The two-step reaction catalysed by NovR that produces the prenylated hydroxybenzoate moiety of novobiocin (ring A, Fig. 1), beginning with the product of the NovQ reaction. (B) Summary of the structure solution process for NovR; the colour scheme is as for Fig. 2. (C) The 2.1 Å resolution structure of NovR in cartoon representation. The asymmetric unit contains a homotetramer, which corresponds to the biological unit. The bound metal ions are represented by red spheres. Each subunit is depicted in a separate colour.
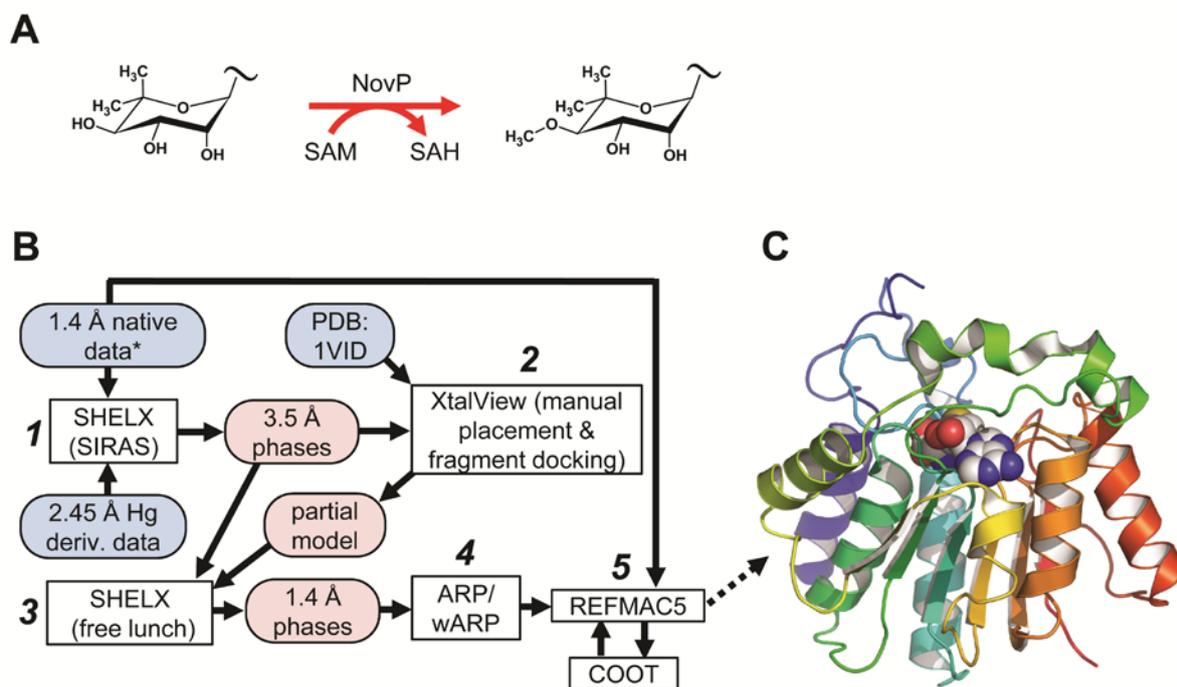
Fig. 5 The *O*-methyltransferase NovP. (A) The reaction catalysed by NovP, which represents the penultimate step in the biosynthesis of novobiocin. At this stage, all three ring systems are conjoined, and rings A and B are complete (Fig. 1); for simplicity, only ring C, the noviose moiety, is shown. (B) Summary of the structure solution process for NovP. (C) The 1.4 Å resolution structure of NovP in cartoon representation. The bound co-substrate, *S*-adenosyl-L-homocysteine (SAH), is also shown in space-filling representation in CPK colours. Otherwise the colour schemes for (B) and (C) are as for Fig. 2.

*The native dataset incorporated data from several crystals (see main text for further details).

Table 1 Summary of data collection and refinement statistics. Data for the highest-resolution shell are given in parentheses. Bijvoet pairs were merged for R_{merge} calculation. Ramachandran plot statistics were calculated by MOLPROBITY

Protein	NovW	CloQ	NovR	NovP
Data collection				
Space Group	$P4_32_12$	$I4_122$	C2	P2
Cell parameters (Å/°)	a = b = 59.4, c = 109.0, α = β = γ = 90.0	a = b = 135.2, c = 98.1, α = β = γ = 90.0	a = 86.7, b = 139.4, c = 100.8, α = γ = 90.0, β = 101.2	a = 51.8, b = 46.0, c = 61.2, α = γ = 90.0, β = 105.0
Wavelength (Å)	0.870	1.542	1.488	1.488
Resolution range (Å)	39.00 - 1.60 (1.64 - 1.60)	31.86 - 2.21 (2.33 - 2.21)	44.46 - 2.10 (2.21 - 2.10)	36.30 - 1.40 (1.42 - 1.40)
Unique reflections	26,449	33,556	65,200	54,638
Completeness (%)	99.9 (99.9)	97.3 (81.6)	95.5 (77.3)	98.8 (89.5)
Multiplicity	5.8	7.7	3.5	3.4
R_{merge}	0.060 (0.362)	0.084 (0.301)	0.062 (0.321)	0.057 (0.164)
Mean $I/\sigma(I)$	25.7 (3.9)	19.6 (7.3)	15.3 (3.0)	18.1 (4.2)
Wilson B value ($\mathring{A}^2$)	15.5	24.4	33.2	17.6
Refinement				
R_{work}/R_{free}	0.150/0.195	0.165/0.212	0.164/0.198	0.145/0.162
Molecules per ASU	1	1	4	1
Solvent content (%)	37	60	50	44
No. of residues	199	316	1025	242
No. of water molecules	200	236	695	242
RMS deviations				
Bond lengths (Å)	0.014	0.016	0.014	0.013
Bond angles (°)	1.49	1.54	1.53	1.54
Ramachandran plot (%)				
Most favoured	99.5	98.4	97.6	96.6
Allowed	0.5	1.6	2.4	3.4
Disallowed	0	0	0	0
Average B factors ($\mathring{A}^2$)				
Main chain atoms	14.6	20.6	31.5	14.6
Side chain atoms	17.0	22.4	35.1	18.1
Water molecules	30.5	26.5	39.0	32.3
Ligands/ions	30.6	28.3	48.2	18.1
Overall	17.6	21.9	33.7	18.1