Structural basis of keto acid utilization in nonribosomal depsipeptide synthesis

Diego A. Alonzo[a,*], Clarisse Chiche-Lapierre[a,*], Michael J. Tarry[a], Jimin Wang[b] and T. Martin Schmeing[a]

[a]Department of Biochemistry and Centre de recherche en biologie structurale, McGill University, Montréal, QC, Canada, H3G 0B1

[b]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 06510

[*]Equal contribution

Correspondence e-mail: martin.schmeing@mcgill.ca

Nonribosomal depsipeptides are natural products composed of amino and hydroxy acid residues. The hydroxy acid residues often derive from α-keto acids, reduced by ketoreductase domains in the depsipeptide synthetases. Biochemistry and structures reveal the mechanism of discrimination for α-keto acids and a remarkable architecture: flanking intact adenylation and ketoreductase domains are sequences separated by >1100 residues that form a split "pseudoA_{sub}" domain, structurally important for the depsipeptide module's synthetic cycle.

Nonribosomal peptide synthetases (NRPSs) are multi-domain enzymes that produce a vast array of biologically-active compounds, including clinically-used therapeutics such as the antibiotic daptomycin, the immunosuppressant cyclosporin and the antifungal caspofungin[1]. NRPSs are composed of a series of modules, sets of domains that work together to add a building block substrate to the growing peptide chain, in a manner analogous to assembly lines.

Depsipeptides contain both amino acid residues and hydroxy acid residues, linked by amide and ester bonds, respectively. Important depsipeptides include the piscicide antimycin, the K+ ionophores cereulide[2,3] and valinomycin[4,5] (**Supplementary Fig. 1**), the anticancer agent cryptophycin[6], the antimicrobial kutzneride[7], the antifungal hectochlorin[8] and the insecticide mycotoxins bassianolide and beauvericin[9]. The hydroxy acid residues in fungal compounds bassianolide and beauvericin originate from direct selection and incorporation of α-hydroxy acids, but the hydroxy acid residues in the bacterial compounds antimycin, hectochlorin, kutzneride, cryptophycin, cereulide and valinomycin are derived from α-keto acid substrates[10].

In bacterial depsipeptide synthetases, modules responsible for adding hydroxy acids include condensation (C) (if an elongation module), adenylation (A), ketoreductase (KR), and peptidyl carrier protein (PCP) domains (**Fig. 1a** and **Supplementary Fig. 1**)[11-14]. These A domains select α-keto acids using a hitherto unknown mechanism to differentiate them from α-amino and α-hydroxy acids. Intriguingly, the aspartate which contacts the α-amino group in amino acid-selecting A domains is altered to a hydrophobic residue in α-keto acid-selecting A domains, not to a positive or polar residue[7,10,11]. Depsipeptide synthetase A domains adenylate the α-keto acid, then transfer it to the PCP domain. The PCP domain transports the α-keto acyl moiety to the KR domain for stereoselective reduction. After reduction, the α-hydroxyl acyl-PCP moves to that module's C domain for condensation, making an ester bond (or goes directly to the downstream C domain in the case of A-KR-PCP initiation modules), and synthesis continues as in the canonical case.

Ketoreducing depsipeptide modules are found commonly in bacteria. Querying databases[15] with A-KR-PCP sequences returns ~1000 non-redundant proteins with high coverage and >27% identity,

mainly from actinobacteria, firmicutes and cyanobacteria. Sequence analysis shows KR domains in depsipeptide synthetases to be similar to polyketide synthase (PKS) KR domains that catalyze β-ketoreduction using NADPH[10,11,14]. NRPS KR domains were proposed to be embedded into the A domain between motifs A8 and A9[11-14]. This was confirmed as the position for some methyltransferase (MT) domains by the A-MT didomain structure of thiocoraline synthetase[16].

To gain insight into depsipeptide synthetases, we undertook crystallographic studies of A-KR-PCP initiation modules. The initiation module of a cereulide synthetase homologue from *Bacillus stratosphericus LAMA 585*, here called stratospherulide synthetase (which has dimodular subunits StsA and StsB, **Supplementary Fig. 1**) produced crystals readily, but initially they diffracted to ~60 Å. Multiple years of optimization improved diffraction and allowed structure determination at ~3.4-3.8 Å resolution (**Fig. 1b**, **Supplementary Table 1**). The mobile PCP domain and some linkers are not visible in electron density maps, but essentially all other components (1193-1209 of 1318 residues) are ordered. The crystal's asymmetric unit contains four copies of the protein with similar overall configurations, which pair as two dimers (**Supplementary Fig. 2**).

The structure reveals a surprising overall architecture for modules which use α-keto acids (**Fig. 1b, Supplementary Fig. 2**). Contrary to predictions, the A domain is intact and uninterrupted, including both the core A domain (A_{core}; residues 82-542) and its small C-terminal subdomain (A_{sub}; residues 543-660). The A_{sub} is in the "closed", adenylation-competent conformation. A short linker joins the A domain to the KR domain (residues 668-1155). This α-KR domain is structurally similar to PKS β-KR domains, but only has a degenerate sequence motif similar to those found within PKS β-KR domains that determine stereospecificity by directing substrate to one of two putative entrance tunnels[6,17]. Unexpectedly, C-terminal to the KR domain is a small domain which possesses the fold of the A_{sub} domain (**Fig. 1b,c**). Residues 1156-1227 assume βαβα secondary structure, ending at a linker to the PCP domain. To assemble a complete A_{sub}-like fold, the βαβα elements require an additional helix and strand, which are provided by the StsA N-terminus (residues 14-43). In this structure, the neighboring molecule donates residues 13-40 to complete this split A_{sub}-like domain (hereafter called pseudoA_{sub}) (**Fig. 1c**). The interruption in pseudoA_{sub} is not the location analogous to where MT domains are inserted (**Supplementary Fig. 3**)[16], but the presence of the A9 and A10 motifs in the pseudoA_{sub}, C-terminal to the KR domain likely led to this assertion.

With guidance of structural information, pseudoA_{sub} domains are evident in all characterized depsipeptide modules (**Supplementary Fig. 4**). C-terminal portions with A10-like motifs (PxxxxGK)[18] are evident in (C)-A-KR-PCP modules from cereulide, valinomycin, kutzneride, cryptophycin,

antimycin and hectochlorin synthetases by detectable (26.4-32.7%) sequence identity to canonical A_{sub} domains and analogous βαβ predicted secondary structures. The separated N-terminal portions of pseudoA_{sub} domains are confirmed by secondary structure prediction and, in some cases, designation as a partial A domain by the conserved domain architecture retrieval tool CDART[15].

Canonical A_{sub} domains contain a catalytic lysine required for adenylation (e.g., Lys517 in gramicidin S synthetase A[19]). The pseudoA_{sub} in cereulide synthetase A has a lysine at the analogous position, but other pseudoA_{sub} sequences do not (**Supplementary Fig. 4**). Nonetheless, because the pseudoA_{sub} is close to the A_{core} in the crystal structure, we asked whether a pseudoA_{sub} could participate in adenylation. In cereulide synthetase A, mutation of the lysine (Lys653) in the canonical A_{sub} abolished adenylation, but mutation of the lysine in the pseudoA_{sub} (Lys1218) did not, demonstrating that the pseudoA_{sub} does not participate in catalysis of adenylation.

The revelation that the architecture of depsipeptide synthetase modules includes an intact A domain allowed us to interrogate the mechanism of α-keto selection with high-resolution structures of the excised A domain. The excised A domain of StsA (residues 83-649) is active in adenylation and preferentially selects α-ketoisocaproic acid (α-kic) (**Supplementary Fig. 5a**). Crystallization trials produced multiple crystals in several space groups. The two co-complexes structures presented here (**Supplementary Table 1**) both show α-kic adenylate, formed *in situ*, bound to A_{core}, and A_{sub} in an "open" conformation (**Fig. 2**, **Supplementary Fig. 5b**). These high-resolution co-complexes demonstrate that side chain specificity is imparted in the same manner as in α-amino acid selective A domains, with the α-kic side chain surrounded by specificity code residues[20,21]. However, the Asp which contacts α-amino groups in amino acid A domains is always replaced by a hydrophobic residue in α-keto A domains[7,10] (here Ile306; **Fig. 2b**) suggesting a different mode of recognition of the α position. Perhaps the simplest solution would be recognition of the α-keto group by a hydrogen bond donor or positive side chain, but there are no such moieties nearby. Instead, the crucial contact in α-keto acid selection is between the α-keto and the carbonyl of the peptide bond joining Gly414 and Met415, through an antiparallel carbonyl-carbonyl interaction (**Fig. 2b**). This type of interaction has approximately the same strength as a hydrogen bond[22]. It is known to occur in small molecule crystals[22] and stabilizes partially-allowed Ramachandran conformations in proteins[23]. The angles and distances of the interaction closely match those predicted by analyses of small molecules[22].

This antiparallel carbonyl-carbonyl interaction relies on the backbone of Gly414-Met415 being flat along the substrate binding pocket. α-Amino acids A domains have a conserved proline in the position equivalent to Met415. Mutation of Met415 to proline in StsA abolished adenylation

(**Supplementary Fig. 5**). Reciprocally, we mutated this proline (Pro483) in the first A domain of linear gramicidin synthetase (LgrA) to methionine or alanine. The resulting mutants still preferred valine, but had a substantial increase in adenylation of the equivalent α-keto acid, α-ketoisovaleric acid (α-kiv). Unexpectedly, the mutation LgrA(Asp396Val) (which occurs in α-kiv specific A domains[6]) in the Pro483Met background abolished that activity. To further investigate the increase in adenylation of α-kiv in mutant LgrA, we determined the co-complex structure of the A domain mutant LgrA(Pro483Met) complexed with α-kiv (**Supplementary Fig. 5g,h, Supplementary Table 1**). We observed a shift in the 482-483 backbone, but it did not form a perfect antiparallel carbonyl-carbonyl interaction, consistent with only a partial gain of α-kiv adenylation.

Having deciphered the mechanism for α-keto selectivity, we interrogated the next steps of the synthetic cycle of α-keto incorporation by depsipeptide synthetase modules. We used three samples that provided insight into the role of the pseudoA_{sub} and the observed strand-swapping within it. StsA A-KR-PCP purifies as a non-interchanging mixture of monomer and dimer, of which only the dimer crystallized. These are the first two samples. The third is a construct (Δlink) with the linker that connects the N-terminal region of the pseudoA_{sub} with the A_{core} shortened. The monomer and dimer were active in each step – adenylation, thiolation and ketoreduction (**Fig. 3a-c, Supplementary Fig. 6**). Modeling shows that the linker that allows the strand swap would also allow the N-terminus to complement its own pseudoA_{sub}, which likely occurs in monomeric A-KR-PCP (**Supplementary Fig. 6b)**. The Δlink construct is active in adenylation, but not in thiolation, indicating that conformational changes which allow proper PCP domain binding require the freedom of movement imparted by a long linker between pseudoA_{sub} and A domains (**Fig. 3d**).

Depsipeptide synthetase modules thus possess their own peculiar architecture. Presumably the architecture arose from a complex genetic insertion between an NRPS gene and a PKS or NRPS-PKS hybrid that split and duplicated the pseudoA_{sub} and led to a functional depsipeptide synthetase. The pseudoA_{sub} is likely maintained because it is not detrimental to peptide synthesis and could help with domain positioning, and because removing either the N or C terminal portion of it could result in unstable protein, while spontaneous, simultaneous removal of both portions would be improbably rare. This architecture is remarkable and highly unusual: the pseudoA_{sub} is folded into a bone fide domain, despite being separated by >1100 residues. Notably, the C-terminal residues of the pseudoA_{sub} (Asp1226) and A_{sub} (Arg622) are only ~10 Å apart, meaning that canonical NRPS module architecture and the depsipeptide module architecture can position the PCP domain in the same area (**Supplementary Fig. 7a**). The PCP domain must transport α-kic from the A domain to the KR

domain, a distance of 70 Å. The transition of the observed adenylation state to a modeled thiolation state and then to a modeled ketoreduction state can seemingly occur without difficulty, with the pseudoA_{sub} domain performing the linker-domain role seen with canonical A_{sub} domains[24,25] (**Supplementary Fig. 7b**).

In all, this study shows a novel architecture of NRPS modules that incorporate non-amino acid monomers and the mode of selection of these monomers, and illustrates that the chemical diversity accomplished by NRPSs involves a rich repertoire of domain arrangements.

Acknowledgements

We thank Claudia Alonso for TEV protease purification and other lab assistance, Janice Reimer for preparing amino coenzyme A, members of the Schmeing lab for helpful advice and discussion, Nancy Rogerson for proofreading, staff at APS (Frank Murphy and Surajit Banarjee) and CLS for support during X-ray data collection, and Nathan Magarvey for discussions and suggesting structural work on depsipeptide synthetases. This work was supported by a Canada Research Chair and NSERC Discovery Grant 418420 to TMS.

Author contributions

TMS, DAA and CCL designed the study and wrote the manuscript. DAA, CCL and MJT performed biochemical experiments. JW performed structure determination and refinement of the A-KR structure using NCS averaging and map sharpening. DAA and CCL performed crystallization, structure determination and refinement.

Corresponding authors

Correspondence to T. Martin Schmeing.

Competing interests

The authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Supplementary information

Supplementary Text and Figures

Supplementary Tables 1-2 and Supplementary Figures 1-8

References

1 Felnagle, E. A. *et al.* Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol Pharm* **5**, 191-211, doi:10.1021/mp700137g (2008).

2 Ekman, J. V. *et al.* Cereulide produced by Bacillus cereus increases the fitness of the producer organism in low-potassium environments. *Microbiology* **158**, 1106-1116, doi:10.1099/mic.0.053520-0 (2012).

3 Alonzo, D. A., Magarvey, N. A. & Schmeing, T. M. Characterization of cereulide synthetase, a toxin-producing macromolecular machine. *PLoS One* **10**, e0128569, doi:10.1371/journal.pone.0128569 (2015).

4 Jaitzig, J., Li, J., Sussmuth, R. D. & Neubauer, P. Reconstituted biosynthesis of the nonribosomal macrolactone antibiotic valinomycin in Escherichia coli. *ACS synthetic biology* **3**, 432-438, doi:10.1021/sb400082j (2014).

5 Huguenin-Dezot, N. *et al.* Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* **565**, 112-117, doi:10.1038/s41586-018-0781-z (2019).

6 Ding, Y., Rath, C. M., Bolduc, K. L., Hakansson, K. & Sherman, D. H. Chemoenzymatic synthesis of cryptophycin anticancer agents by an ester bond-forming non-ribosomal peptide synthetase module. *J Am Chem Soc* **133**, 14492-14495, doi:10.1021/ja204716f (2011).

7 Fujimori, D. G. *et al.* Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc Natl Acad Sci U S A* **104**, 16498-16503, doi:10.1073/pnas.0708242104 (2007).

8 Ramaswamy, A. V., Sorrels, C. M. & Gerwick, W. H. Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium Lyngbya majuscula. *J Nat Prod* **70**, 1977-1986, doi:10.1021/np0704250 (2007).

9 Xu, Y. *et al.* Biosynthesis of the cyclooligomer depsipeptide bassianolide, an insecticidal virulence factor of Beauveria bassiana. *Fungal Genet Biol* **46**, 353-364, doi:10.1016/j.fgb.2009.03.001 (2009).

10 Magarvey, N. A., Ehling-Schulz, M. & Walsh, C. T. Characterization of the cereulide NRPS alpha-hydroxy acid specifying modules: activation of alpha-keto acids and chiral reduction on the assembly line. *J Am Chem Soc* **128**, 10698-10699, doi:10.1021/ja0640187 (2006).

11 Magarvey, N. A. *et al.* Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts. *ACS Chem Biol* **1**, 766-779, doi:10.1021/cb6004307 (2006).

12 Cheng, Y. Q. Deciphering the biosynthetic codes for the potent anti-SARS-CoV cyclodepsipeptide valinomycin in Streptomyces tsusimaensis ATCC 15141. *Chembiochem* **7**, 471-477, doi:10.1002/cbic.200500425 (2006).

13 Ehling-Schulz, M. *et al.* Identification and partial characterization of the nonribosomal peptide synthetase gene responsible for cereulide production in emetic Bacillus cereus. *Appl Environ Microbiol* **71**, 105-113, doi:10.1128/AEM.71.1.105-113.2005 (2005).

14 Ehling-Schulz, M. *et al.* Cereulide synthetase gene cluster from emetic Bacillus cereus: structure and location on a mega virulence plasmid related to Bacillus anthracis toxin plasmid pXO1. *BMC Microbiol* **6**, 20, doi:10.1186/1471-2180-6-20 (2006).

15 Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. CDART: protein homology by domain architecture. *Genome Res* **12**, 1619-1623, doi:10.1101/gr.278202 (2002).

16 Mori, S. *et al.* Structural basis for backbone N-methylation by an interrupted adenylation domain. *Nature Chemical Biology*, doi:10.1038/s41589-018-0014-7 (2018).

17 Zheng, J., Piasecki, S. K. & Keatinge-Clay, A. T. Structural studies of an A2-type modular polyketide synthase ketoreductase reveal features controlling alpha-substituent stereochemistry. *ACS Chem Biol* **8**, 1964-1971, doi:10.1021/cb400161g (2013).

18 Miller, B. R., Sundlov, J. A., Drake, E. J., Makin, T. A. & Gulick, A. M. Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins* **82**, 2691-2702, doi:10.1002/prot.24635 (2014).

19 Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* **16**, 4174-4183 (1997).

20 Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**, 493-505, doi:10.1016/S1074-5521(99)80082-9 (1999).

21 Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**, 211-224 (2000).

22 Allen, F. H., Baalham, C. A., Lommerse, J. P. M. & Raithby, P. R. Carbonyl-carbonyl interactions can be competitive with hydrogen bonds. *Acta Crystallogr. Sect. B-Struct. Sci.* **54**, 320-329, doi:10.1107/S0108768198001463 (1998).

23 Deane, C. M., Allen, F. H., Taylor, R. & Blundell, T. L. Carbonyl-carbonyl interactions stabilize the partially allowed Ramachandran conformations of asparagine and aspartic acid. *Protein Eng* **12**, 1025-1028 (1999).

24 Tanovic, A., Samel, S. A., Essen, L. O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659-663, doi:10.1126/science.1159850 (2008).

25 Reimer, J. M. *et al.* Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, doi:10.1126/science.aaw4388 (2019).

Figures

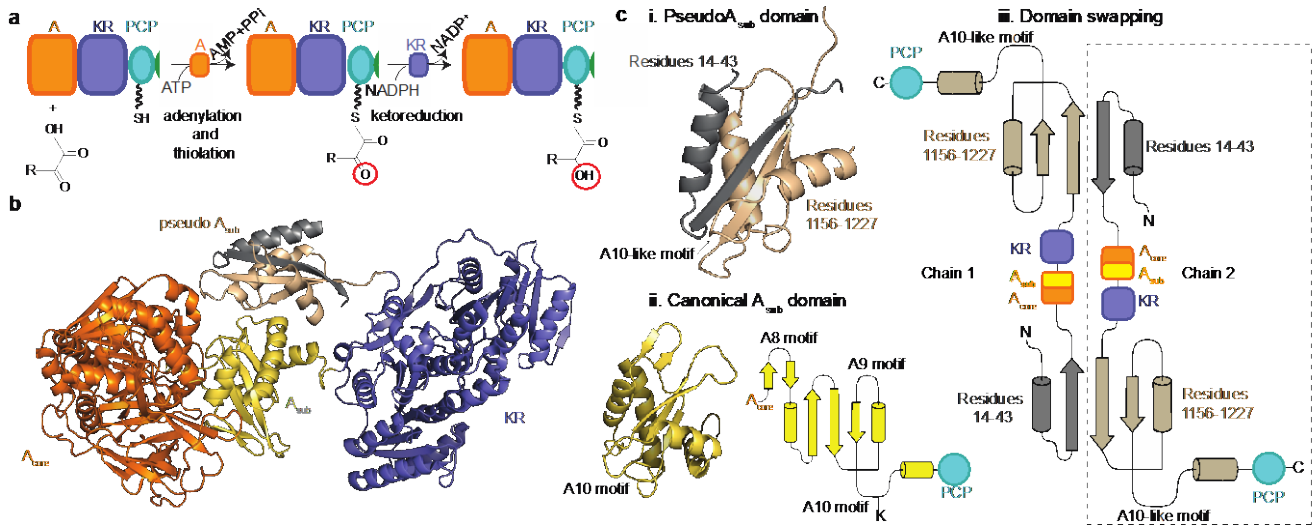

Figure 1. Domain organization and structure of a depsipeptide module

a. Schematic of the reactions depsipeptide synthetases perform to enable incorporation of α-keto acids into depsipeptides. **b.** Overall structure of a depsipeptide module (in which the PCP domain is disordered). **c.** The pseudoA_{sub} domain (i) is formed by structural elements from residues C-terminal to the KR domain (wheat, residues 1156 - 1227) and from the N-terminus of the protein (grey, residues 14 – 43; here donated by the adjacent molecule in the crystal). Its structure is similar to canonical A_{sub} domains (ii). (iii) The N-terminal portion of the pseudoA_{sub} domain (grey) in the crystal is donated by a neighboring molecule.

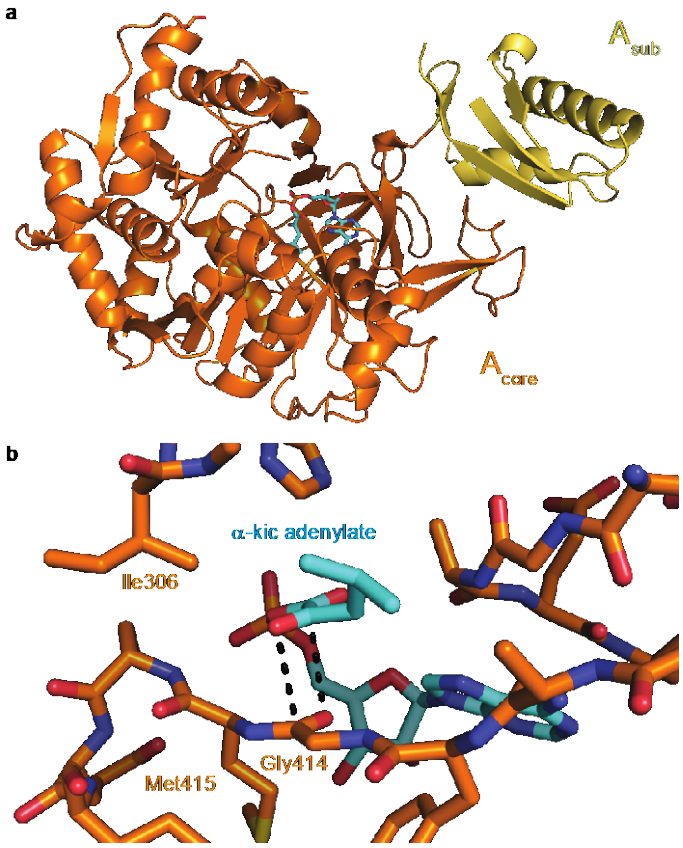

Figure 2. The A domain binds the α-keto acid with an antiparallel carbonyl-carbonyl interaction

a. The structure of an excised α-keto acid-selecting A domain bound to an α-ketoisocaproic (kic) -adenylate. **b.** The α-keto group of the α-kic (cyan) binds through an antiparallel carbonyl-carbonyl interaction with the main-chain carbonyl of the peptide bond between Gly414 and Met415 in the A domain active site. The selectivity code residues for StsA-A1 are I-G-M-W-I-G-A-(M)-S-A-K, where the M in parenthesis is Met415. Other than I306 and M415, which may be useful as indicators of α-keto selectivity, the code is similar to those for leucine.

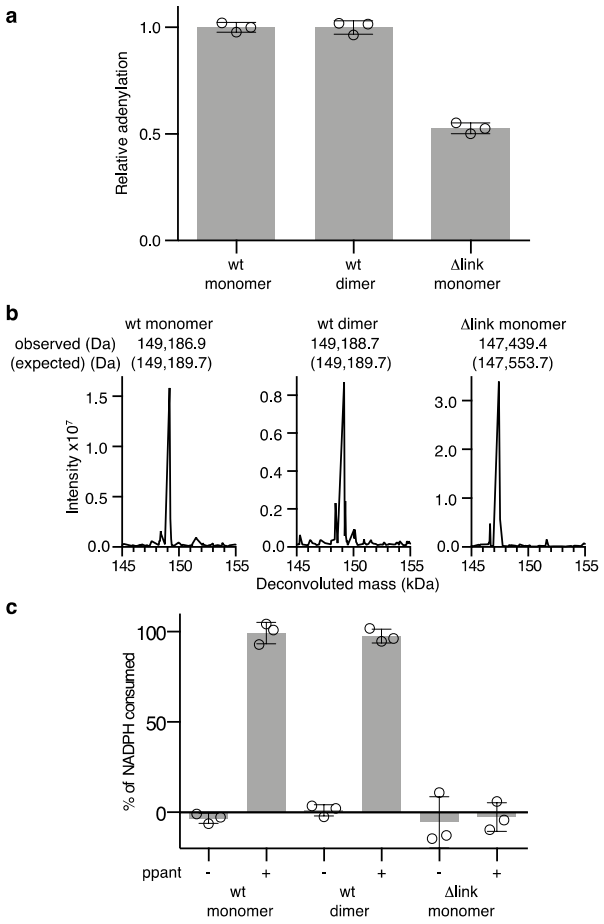

Figure 3. The catalytic activity of the depsipeptide module

a. Adenylation by A-KR-pA$_{sub}$-PCP monomer and dimer samples, and by Δlink (missing residues 57-71 between pseudoA$_{sub}$ and A$_{core}$). Data (n=3 independent reactions) are initial rates expressed as the best-fit values of the slope normalized to the activity of wt monomer. Central values represent means and error bars are ±s.d. of the mean. **b.** Thiolation activity, evaluated by intact protein LC-ESI-MS. Monomer and dimer thiolate α-kic. Thiolation is not detected for Δlink. This experiment was repeated three times with similar results. **c.** Ketoreduction, evaluated by NADPH consumption in a single-turnover fluorescence assay. Monomer and dimer catalyze ketoreduction, Δlink does not catalyze, as expected from the absence of thiolation activity. Data (n=3 independent reactions) are the normalized differences of total fluorescence between controls lacking ATP and α-kic and full reactions. Central values represent means and error bars are ±s.d. of the mean.

Online Methods

Cloning, expression and purification of StsA constructs

A gene construct for the A-KR-PCP module (NCBI: WP_007498213, residues 1-1318) from *B. stratosphericus LAMA 585* was synthesized as an *E. coli* codon-optimized gene and cloned into a pET-11a vector by GenScript. A-KR-PCP (residues 1-1318) and the A domain (residues 83-649, A_{83-649}) were amplified from that plasmid using primers StsA_AKRT_NdeI_F, StsA_AKRT_pspOMI_R, StsA_Ade_83_649_NdeI_F and StsA_Ade_83_649_NotI_R respectively (**Supplementary Table 2**). PCR products were digested with NdeI and NotI (Δlink, A_{83-649}) or NdeI and PspOMI (A-KR-PCP) (New England Biolabs) and ligated into a digested (NdeI and NotI) pJ411-based vector, resulting in constructs which include C-terminal tobacco etch virus (TEV) cleavage sequence and octahistidine tag.

All StsA proteins were overexpressed in *E. coli BL21(DE3)* or *E. coli BL21 entD-* (kindly provided by Dr. Christian Chalut and Dr. Christophe Guilhot)[26] grown in LB media supplemented with kanamycin at 34 μg/L. Cultures were grown at 37 ºC with 220 r.p.m agitation to an O.D. 600 of between 0.5 and 0.8 before being cooled on ice to 16 ºC and induced with 100 μM D-1-thiogalactopyranoside (IPTG), and incubated for an additional 16 hours at 16 ºC until harvest by centrifugation.

For protein purification of A-KR-PCP, cells were resuspended in homogenization buffer (50 mM Tris pH 8.0, 250 mM NaCl, 2 mM $CaCl_2$, 10 % glycerol, 2 mM β-mercaptoethanol (βME), 1 mM PMSF) supplemented with DNaseI and one cOmplete protease inhibitor tablet buffers and sonicated on ice for a total time of 8 minutes, with a cycle set at 10 seconds on, 20 seconds off at 50% amplitude using a model 505 sonic dismembrator (Fisher). Lysates were clarified by centrifugation at 18 000 g and then applied to 2 x 5 mL Hi-Trap Ni-IMAC FF (GE Healthcare) columns connected in tandem, pre-equilibrated in homogenization buffer. Columns were washed to baseline and protein was eluted with buffer B (50 mM Tris pH 8, 500 mM NaCl, 4 mM $CaCl_2$, 150 mM imidazole, 10 % glycerol, 2 mM βME). TEV protease was added to the eluted protein (1:20 w:w ratio) and dialyzed 16 h at 4 ºC wild mild stirring against buffer C (50 mM Tris pH 8, 10 mM NaCl, 4 mM $CaCl_2$, 10 % glycerol, 2 mM βME). Sample was applied to re-equilibrated 2x5 mL Hi-Trap IMAC FF columns. Non-binding protein was recovered and applied to a Mono Q HR 16/10 column pre-equilibrated in buffer D (25 mM HEPES pH 8.0, 10 mM NaCl, 10% glycerol, 2 mM βME). A wash step of 42% buffer E (25 mM HEPES pH 8.0, 500 mM NaCl, 10% glycerol, 2 mM βME) followed by a gradient from 42%-70% buffer E over 400 mL eluted purified protein. A-KR-PCP elutes over the gradient as different oligomeric species, confirmed by native-PAGE and size exclusion chromatography. A-KR-PCP dimers

and monomers were pooled independently and concentrated to at least 10 mg ml^{-1} using a 30 kDa molecular weight cut off Amicon Ultra centrifugal filter (Millipore).

For protein purification of the A domain, the same procedure was followed until the anion exchange step, where the Mono Q HR 16/10 column was pre-equilibrated in buffer D without glycerol and sample was eluted in 22% buffer E without glycerol and concentrated using a 10 kDa molecular weight cut off Amicon Ultra centrifugal filter (Millipore). Sample was applied to a Superdex-75 10/300 column pre-equilibrated in buffer F (25 mM HEPES pH 8.0, 100 mM NaCl, 0.2 mM tris 2-carboxyethyl phosphine (TCEP)), and the fractions of the single peak which contained the A domain concentrated.

The StsA Δlink construct was generated by site-directed mutagenesis using primers del57to71For and del57to71Rev (**Supplementary Table 2**) and was expressed and purified as described for the wild type protein.

All proteins were transferred to solutions containing 20% v/v glycerol, flash frozen in liquid nitrogen and stored at -80 ºC.

Cloning, expression and purification of CesA and LgrA constructs

CesA A-KR-PCP was amplified from previously-described[3] plasmids using primers CesA_AKRT_pBacTrev_ndeI_F and CesA_AKRT_pBacTrevNotI_R. PCR products were digested with NdeI and NotI (New England Biolabs) and ligated into a digested (NdeI and NotI) pJ411-based vector, resulting in a C-terminal TEV-cleavable His8 tagged construct. The constructs for CesA A-KR-T K653A and K1218A were produced by site-directed mutagenesis using the primers CesA_AKrT_K653A_for, CesA_AKrT_K653A_rev, CesA_AKrT_1218A_for and CesA_AKrT_K1218A_rev respectively (**Supplementary Table 2**). Proteins were expressed and purified (**Supplementary Figure 8**) as described for wild type StsA A-KR-PCP until TEV incubation and dialysis, where dialysis buffer G consisted of 20 mM HEPES pH 7.4, 100 mM NaCl, 10% glycerol, 2 mM βME. The dialyzed sample was reapplied to the re-equilibrated HiTrap IMAC FF column and unbound protein was injected into a Mono Q HR 16/10 column pre-equilibrated in buffer G. The protein was eluted in a 200 ml gradient to 100% buffer H (buffer G with 500 mM NaCl), concentrated in a 30 kDa Amicon Ultracentrifugal filter (Millipore) and applied to a S200 16-60 size exclusion column (GE) preequilibrated in buffer I (20 mM HEPES pH 7.4, 200 mM NaCl, 10% glycerol, 0.2 mM TCEP). The sample was concentrated and flash-frozen in liquid nitrogen for storage at -80 ºC.

The constructs for LgrA F-A mutants we produced by site-directed mutagenesis (**Supplementary Table 2**) of the published LgrA A-KR-T construct, and expressed and purified as described[27].

Crystallography of StsA A-KR-PCP

Crystallization of A-KR-PCP was challenging and required several iterations to obtain diffraction quality crystals. Initial crystallization conditions were identified with apo A-KR-PCP dimer sample or A-KR-PCP dimer sample modified with "aminopantetheine" (a derivative of pantetheine in which the terminal thiol is replaced by an amine[27-29]) using sparse-matrix searches with commercially available screens (Qiagen) in 96-well sitting drop format. Even after extensive condition optimization, these crystals did not have desirable morphology, lacking defined edges or appearing gel-like, diffracted very poorly and often displayed twinning. Therefore, we screened for new conditions using the sparse-matrix micro-seeding (SMMS) method[30] with seeds originating from two of the optimized conditions. The number of conditions producing crystals increased dramatically by using SMMS compared to screening without seeding. After iterative rounds of optimization, including streak micro-seeding, additive and detergent screens and switching to a vapor diffusion hanging-drop setup, we obtained a condition that yielded single crystals which diffracted to a maximum resolution of ~7 Å. Screening many cryoprotection methods finally produced crystals which produced reflections up to 3.4 Å in diffraction experiments (**Supplementary Table 1**).

For the crystallization of StsA A-KR-PCP which led to the structure presented here, purified A-KR-PCP dimer protein was incubated for 16 h at room temperature with Sfp (1:1.22 molar ratio), 3 mM amino-CoA (kindly provided by Janice Reimer)[27-29] and 10 mM $MgCl_2$. Phosphopantetheinylation was verified by LC-MS and the sample was incubated for 16 hours with 13 mM ATP and 20 mM α-kic in 200mM HEPES pH 9.2. The sample was analyzed by LC-MS and applied two subsequent times to an S-200 16/60 column pre-equilibrated in buffer F. Relevant fractions were recovered, concentrated to 15.5 mg mL^{-1} and used for crystallization. Crystallization occurred by hanging drop vapor diffusion against 500 μL of reservoir solution (0.167M sodium succinate, 0.25% v/v PEG 2000 MME, 17mM HEPES pH 7.0, 17mM HEPES pH 8.0, 67mM sodium chloride, 6% v/v PEG-3350, and 0.105 M calcium acetate). To set drops, stock solutions of J:K:phenol:protein were mixed sequentially in 0.66:1.33:0.4:1.6 volume proportions in a final volume of 24 uL, mixed thoroughly and then split in two cover slips as 4 μL drops. Stock solutions compositions were: J (1M sodium succinate, 1.5% v/v PEG 2000 MME, 0.1M HEPES pH 7.0, 0.025M HEPES pH 8.0 and 0.1M

NaCl), K (18% v/v PEG-3350 and 0.314 M calcium acetate), phenol (0.1 M phenol solution from Hampton Additive Screen), and protein (15.5mg/ml). Crystals appeared within minutes of setting the drops and grew to their final size in 24 h.

Crystals were cryoprotected by sequential addition of increasing volumes and percentages of ethylene glycol (three times 7 μL of 7%, three times 7 μL of 14%, three times 7 μL of 21%, three times 7 μL of 28% v/v), in cryoprotecting solutions containing all the components and concentrations of the equilibrated crystallization drop, allowing 5 minutes for equilibration after each addition. After the final addition, 86 μL of the drop were replaced with 28% v/v ethylene glycol cryoprotecting solution. The cryo-protected drop was exposed for 30 to 45 minutes to the environment (room temperature) to allow crystal dehydration. Crystals were flash cooled under a liquid nitrogen stream. Data was collected at 100 K at the APS-NECAT 24-ID-C beamline using a Dectris PILATUS 6MF pixel array detector.

A-KR-PCP diffraction data were initially processed at 3.7 Å resolution with a ratio of $\langle I/\sigma_I \rangle$ = 1.0 in the highest resolution shell, and then re-processed at 3.4 Å resolution with $\langle I/\sigma_I \rangle$ = 0.33 in the highest resolution shell from the same diffraction images using the program XDS [31]. Homology models of A_{core} and KR domains were generated using SWISSMODEL[32] and used as ensemble search models (**Supplementary Table 1**) for molecular replacement (MR) in Phaser[33]. With the first data to 3.7 Å resolution, MR was able to locate four KR and 2 A_{core} domains corresponding to the four molecules present in an asymmetric unit. The initial model was missing 2 A_{core}, all A_{sub} and all pseudoA_{sub} domains, for which we generated homology models to aid building (**Supplementary Table 1**). Model refinement statistics for this model were: working R-factor of 37.5%, and free R-factor of 44.4%. The connectivity between A_{core} and KR domains in the atomic model was largely unknown and electron density maps calculated using this model revealed a number of new α-helices that were not present in the atomic model. Next, we applied a combinatory method to attach each of the two A_{core} domains to each of the four KR domains, and applied NCS operations according to KR domains, hoping to establish the valid connectivity between A_{core} and KR domains as well as non-crystallographic symmetry (NCS) among the four molecules in the asymmetric unit. This procedure helped to locate the two missing A_{core} domains. This step was very challenging in part because A_{core}-KR domains in the four molecules do not maintain rigid NCS operations, instead, the NCS of individual domains differs from that of the joined A_{core}-KR domains (**Supplementary Fig. 2**). The same procedure was iteratively carried out for search of all the remaining missing structures in electron density maps. Initial NCS averaging was carried out using Coot[34] according to the connected A_{core}-KR domains from

approximately known atomic coordinates. Once NCS matrices for individual domains were accurately determined, domain-NCS averaging was carried out using RAVE[35] package with sharpened structure factor amplitudes using $\Delta B = -200$ Å^2.

NCS averaging was carried out using the sharpened data set at 3.4-Å resolution, even though model refinement was always carried out using the originally unmodified data set. Amplitude sharpening amplified noise significantly. Upon inclusion of weak data between 3.7 and 3.4 Å resolution (which added 40% independent observations), and through 4-fold domain NCS averaging, the amplified noise was greatly suppressed. Iteration of amplitude sharpening, domain NCS averaging, and model refinement helped to identify all missing structures, and to correct modeling errors made in earlier stages of structure determination. After amplitude sharpening and domain NCS averaging, resulting electron density map unambiguously showed that our initial model that was used to calculate electron density map did not fit the map well. The correct model could be rebuilt in a single step, and refined with Refmac5[36]. The final refined model has working R-factor of 22.8% and free R-factor of 27.5% at 3.4-Å resolution. The mean B-factor for the atomic model is 125.3 Å^2, highlighting weak signals in the data set. Once the higher-resolution A domain structure was determined, it was used to guide improvement of the A-KR atomic coordinates. The final model has been deposited into the protein data bank (PDB) under accession code 6ULW.

Crystallography of StsA A domain

Initial crystallization conditions for the A domain construct were identified by sparse matrix screening using commercial screens (Qiagen). The final optimized crystallization condition was sitting drop vapor diffusion against a 500 µL reservoir solution (1.75 M ($P4_32_12$) or 1.85 M ($P2_12_12_1$) $(NH_4)_2SO_4$ and 22.5% v/v glycerol) and a drop of 1 µL reservoir solution, 0.2 µL 30% w/v trimethylamine N-oxide ($P4_32_12$) and 0.8 µL ($P4_32_12$) or 1.0 µL ($P2_12_12_1$) protein – substrate sample (10 mg ml^{-1} protein, 2 mM ATP or AMPCPP, 10 mM $MgCl_2$, 2 mM α-kic). Crystals appeared between 24 h and 48 hours and reached their maximum size in three days. Crystals were flash cooled into liquid nitrogen. Data was collected at 100 K at the APS-NECAT 24-ID-E beamline using a Dectris EIGER 16M detector.

Data were indexed and integrated to space groups $P4_32_12$ (PDB ID) or $P2_12_12_1$ (PDB ID) with the program DIALS[37]or MOSFLM[38] and scaled with the program AIMLESS[39]. A domain structures were determined using the A_{core} domain fragment from the A-KR structure as a MR search model in Phaser[33], followed by iterative rounds of manual model re-building in Coot[34] and refinement in

Phenix[40]. Crystallization conditions included ATP and α-kic ($P4_32_12$ (PDB ID)), or AMPCPP and α-kic ($P2_12_12_1$ (PDB ID)), and both maps showed density for α-kic adenylate, which was modeled in both structures (**Supplementary Table 1**). Diffraction data and final models have been deposited into the protein data bank (PDB) under accession codes 6ULX and 6ULY.

Crystallography of LgrA mutant Pro483Met

Crystals of LgrA formylation-adenylation didomain were grown using previously-reported conditions[27]. After reaching maximum size, crystals were soaked with a solution consisting of 2 M sodium formate, 100 mM sodium acetate, 150 mM NaCl, 20 mM Tris pH 7.0, 10 mM α-keto isovaleric acid (α-kiv), 2 mM AMPcPP, 10 mM $MgCl_2$. Crystals were cryoprotected with sequential additions of glycerol (5%-20%) in the same solution, incubating during 5 minutes between each addition. Crystals were flash cooled into liquid nitrogen after a final incubation of 30 minutes in the highest glycerol concentration. Data was collected at the Canadian Light Source beamline 08ID-1 using a Pilatus3 S 6M detector. Data was indexed and integrated using DIALS[37] and scaled using AIMLESS[39]. Initial maps and model were obtained by Fourier synthesis using PDB ID: 5es5[27] as a model and further refined in Phenix[40]. Diffraction data and final model has been deposited into the protein data bank (PDB) under accession code 6ULZ.

Biochemical assays

To monitor adenylation, colorimetric pyrophosphate production assays[3,41] were performed in triplicates of 100 μL reactions containing 1 mM (StsA) or 2 mM (CesA and LgrA) acyl substrate, 5 mM (StsA) or 1 mM (CesA and LgrA) ATP, 75 mM TRIS pH 8.0 (StsA), 20 mM HEPES pH 7.4 (CesA) or 20 mM HEPES pH 7.0 (LgrA), 0.2 mM TCEP, 200 mM NaCl, 10% glycerol (CesA and LgrA), 10 mM $MgCl_2$, 0.2 mM 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG), 1 unit/ml purine nucleoside phosphorylase (PNP), 0.03 units/ml of inorganic pyrophosphatase, 0.15 M (StsA and LgrA) or 0.075 M (CesA) hydroxylamine. Reactions were set in clear polystyrene 96-well trays (Corning) and were started by the addition of protein to a final concentration of 1 μM (StsA), 250 nM (CesA) or 1.5 μM (LgrA). Time courses of the total absorbance were recorded at 360 nm in a Molecular Devices SpectraMax M5 plate reader using SoftMax Pro v5.4.1, and baseline-corrected to controls without substrate or enzyme. Initial rates for each replicate reaction were calculated by linear regression and normalized to the activity of wild type controls with native substrate, and means and

standard deviation of the mean calculated. Calculations and plotting were performed in the program GraphPad Prism 6.

To monitor modification with phosphopantetheine and the thiolation reaction, proteins grown in *E. coli BL21 entD-* strain were incubated with Sfp in a 1:1.2 molar ratio with 10 mM $MgCl_2$ and 3 mM CoA. Ppant incorporation and transfer of α-kic were evaluated using previously-described[5] LC-MS methodology using a deconvolution window from 135 to 165 kDa.

To monitor ketoreduction, a single turn-over NADPH fluorescence assay was used. Fluorescence of NADPH (λ_{exc}=340 nm, λ_{emm}=445 nm) is enhanced when bound to the active site of KR domains[3]. Reactions of 100 μL consisting of 25 mM HEPES pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2.5 μM NADPH and 9 μM enzymes were performed in triplicate. In this condition all NADPH is bound to excess protein. Reactions were started by addition of 20 μL of H_2O or a mix of 10 mM α-kic and 10 mM ATP and incubated at room temperature for 1 minute. Total fluorescence intensity (λ_{exc}=340 nm, λ_{emm}=445 nm) was recorded in black polystyrene 96 well plates (Corning) in a Molecular Devices SpectraMax M5 plate reader using SoftMax Pro v5.4.1. NADPH consumption is reported as the differences between the normalized fluorescence of H_2O controls minus the normalized fluorescence of reactions in the presence of ATP+α-kic. Central values represent the mean and standard deviation of n=3 independent reactions. Calculations and graphing were performed in GraphPad Prism 6.

Bioinformatics

Alignments against non-redundant protein sequences database was performed in BLASTP (NCBI, online version)[42] and CDART[15]. Multiple sequence alignments were performed in CLUSTAL OMEGA[43]. Secondary structure prediction was performed using Jpred 4[44].

Full acknowledgement of synchrotrons

This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Pilatus 6M detector on 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). The Eiger 16M detector on 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Data for LgrA

FA was collected using beamline 08ID-1 at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan.

Data availability

Structure coordinates have been deposited in the Protein Data Bank (PDB) under accession codes 6ULW, 6ULX, 6ULY, 6ULZ (**Supplementary Table 1**).

Methods-only References

26 Chalut, C., Botella, L., de Sousa-D'Auria, C., Houssin, C. & Guilhot, C. The nonredundant roles of two 4'-phosphopantetheinyl transferases in vital processes of Mycobacteria. *Proc Natl Acad Sci U S A* **103**, 8511-8516, doi:10.1073/pnas.0511129103 (2006).

27 Reimer, J. M., Aloise, M. N., Harrison, P. M. & Schmeing, T. M. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239-242, doi:10.1038/nature16503 (2016).

28 Liu, Y. & Bruner, S. D. Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases. *Chembiochem* **8**, 617-621, doi:10.1002/cbic.200700010 (2007).

29 Nazi, I., Koteva, K. P. & Wright, G. D. One-pot chemoenzymatic preparation of coenzyme A analogues. *Anal Biochem* **324**, 100-105 (2004).

30 D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W. & Marsh, M. Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Crystallogr F Struct Biol Commun* **70**, 1117-1126, doi:10.1107/S2053230X14015507 (2014).

31 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).

32 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**, W252-258, doi:10.1093/nar/gku340 (2014).

33 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

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

35 Kleywegt, G. J. & Jones, T. A. in *From First Map to Final Model* (eds S. Bailey, R. Hubbard, & D. Waller) 59-66 (SERC Daresbury Laboratory, 1994).

36 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367, doi:10.1107/s0907444911001314 (2011).

37 Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85-97, doi:10.1107/s2059798317017235 (2018).

38 Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271-281, doi:doi:10.1107/S0907444910048675 (2011).

39 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).

40 Adams, P. D. *et al.* Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems. *Annu Rev Biophys* **42**, 265-287, doi:10.1146/annurev-biophys-083012-130253 (2013).

41 Wilson, D. J. & Aldrich, C. C. A continuous kinetic assay for adenylation enzyme activity and inhibition. *Anal Biochem* **404**, 56-63, doi:10.1016/j.ab.2010.04.033 (2010).

42 Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5-9, doi:10.1093/nar/gkn201 (2008).

43 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).

44 Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**, W389-394, doi:10.1093/nar/gkv332 (2015).