Article

Structures and mechanism of condensation in non-ribosomal peptide synthesis

https://doi.org/10.1038/s41586-024-08417-6

Received: 26 April 2024

Accepted: 15 November 2024

Published online: 11 December 2024

Check for updates

Angelos Pistofidis[1], Pengchen Ma[2,3], Zihao Li[4], Kim Munro[1], K. N. Houk[2] & T. Martin Schmeing[1]✉

Non-ribosomal peptide synthetases (NRPSs) are megaenzymes responsible for the biosynthesis of many clinically important natural products, from early modern medicines (penicillin, bacitracin) to current blockbuster drugs (cubicin, vancomycin) and newly approved therapeutics (rezafungin)[1,2]. The key chemical step in these biosyntheses is amide bond formation between aminoacyl building blocks, catalysed by the condensation (C) domain[3]. There has been much debate over the mechanism of this reaction[3–12]. NRPS condensation has been difficult to fully characterize because it is one of many successive reactions in the NRPS synthetic cycle and because the canonical substrates are each attached transiently as thioesters to mobile carrier domains, which are often both contained in the same very flexible protein as the C domain. Here we have produced a dimodular NRPS protein in two parts, modified each with appropriate non-hydrolysable substrate analogues[13,14], assembled the two parts with protein ligation[15], and solved the structures of the substrate- and product-bound states. The structures show the precise orientation of the megaenzyme preparing the nucleophilic attack of its key chemical step, and enable biochemical assays and quantum mechanical simulations to precisely interrogate the reaction. These data suggest that NRPS C domains use a concerted reaction mechanism, whereby the active-site histidine likely functions not as a general base, but as a crucial stabilizing hydrogen bond acceptor for the developing ammonium.

Non-ribosomal peptide synthetases (NRPSs) are large macromolecular machines that biosynthesize a vast variety of natural products, including anti-bacterial, anti-viral, anti-tumour and anti-fungal compounds and siderophores and immunosuppressants[1,2]. NRPSs have greatly shaped modern medicine, with the current arsenal of clinical therapeutics including large numbers of compounds made directly by NRPSs (such as cyclosporin and bacitracin), compounds made by NRPSs with subsequent modification by other enzymes (vancomycin and actinomycin D) or synthetic reactions (rezafungin and cilofungin), and compounds made by NRPS–polyketide synthetase hybrids (bleomycin and pristinamycin II)[1,2] (Extended Data Fig. 1a).

NRPSs have an elegant architecture and synthetic cycle (Extended Data Fig. 1b). They are organized into modules, sets of domains that together incorporate one specific monomer substrate into the growing peptide[1]: the adenylation (A) domain adenylates the cognate amino acid substrate and ligates it to the transport (or thiolation; T) domain, through that domain's 4′-phosphopantetheine (ppant) post-translational modification[16]. The T domain then transports[13,17] the covalently bound amino acid to the C domain, which performs amide bond formation between this acceptor aminoacyl-ppant-T domain and a donor peptidyl-ppant-T domain from the upstream module, elongating and transferring the nascent peptide[3,4,14] (Fig. 1a). Thereafter, the elongated peptidyl-ppant-T domain migrates downstream to act as the donor in next module's condensation.

The C domain has been studied for decades. It is a V-shaped, bi-lobed (N- and C-lobes), pseudodimer, with binding sites for donor and acceptor acyl-ppant-T domains at opposite sides of a long active-site tunnel[3]. The active site is marked by a HHxxxDG motif, and the second histidine is well established to be very important for peptide bond formation[3–11,18]. However, its role in catalysis, and the chemical mechanism of condensation by C domains remains hotly debated, with almost all reasonable mechanisms having been proposed[3–12]. In most NRPSs studied, this histidine is essential, but there are several reports of histidine mutants retaining some, attenuated, activity[3,7,8], and rare cases of C domains with other residues at this position[19].

Structural studies designed to provide insights into C domain catalysis are challenging owing to the unusual nature of the condensation substrates and of NRPSs. The canonical condensation substrates are aminoacyl and peptidyl moieties covalently ligated through thioester bonds to their ppant-T domains, which are attached to the rest of the NRPS (and therefore the C-domain catalyst) by flexible linkers. T domains are necessarily mobile, because they undergo very large movements to transport ligands between reaction centres. Moreover, the thioester linkages are short lived with respect to crystallographic time scale, and NRPSs are typically poorly behaved on grids,

[1]Department of Biochemistry and Centre de Recherche en Biologie Structurale, McGill University, Montréal, Quebec, Canada. [2]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA. [3]Department of Chemistry, School of Chemistry, Xi'an Key Laboratory of Sustainable Energy Material Chemistry and Engineering Research Center of Energy Storage Materials and Devices, Ministry of Education, Xi'an Jiaotong University, Xi'an, China. [4]School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai, China. ✉e-mail: martin.schmeing@mcgill.ca

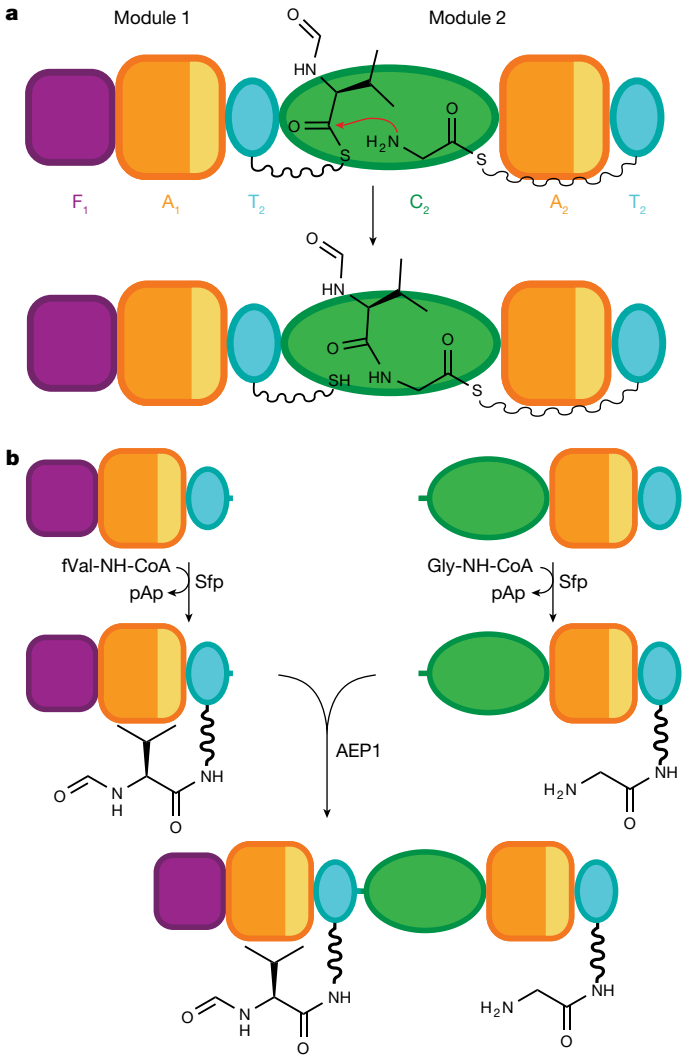


Fig. 1 | Condensation by LgrA and preparation of LgrA condensation complexes. a, The condensation domain in LgrA, C_2, catalyses peptide-bond formation between T_1-bound fVal-ppant and T_2-bound Gly-ppant, resulting in T_2-domain-bound f-Val-Gly-ppant. Domains are coloured consistently through all of the figures (purple, F domain; orange, A_{core}; yellow, A_{sub}; cyan, T; green, C). A_{core} and A_{sub} are the large N-terminal subdomain and small C-terminal subdomain of the A domain, respectively. They are coloured differently from one another because they assume very different relative orientations at various stages of the NRPS synthetic cycle. ppants are represented by squiggles. LgrA is the first subunit of linear gramicidin synthetase, a 16-module NRPS that biosynthesizes the clinically used topical antibiotic. The inactive E domain downstream from T_2 was not used in this study and is not shown. **b**, The strategy to place appropriate condensation reaction substrate analogues onto each module of LgrA. **c**, AEP1-catalysed ligation reactions. Ligation reactions were performed repeatedly with separate protein preparations ($n = 19$) with similar results.

with only atypical, dimeric NRPSs yielding cryo-electron microscopy structures to date[20–22]. Thus, although many structures have been determined for extruded C domains and for multi-domain NRPS constructs containing C domains[3,4,8–12,14,23–29], including some with substrates or analogues[4,8,9,11,14,24], the 'smoking gun' high-resolution canonical condensation complex is lacking. This gap stymies full interpretation of biochemical experiments and the most accurate computational mechanistic interrogations.

Here we present structures of the dimodular linear gramicidin synthetase subunit A bound to non-hydrolysable substrate and product analogues of the full condensation reaction. These complexes were prepared by separate chemoenzymatic loading of substrate or product analogues onto individual modules, followed by protein ligation of these modules. The pre-condensation structure enables informative quantum mechanistic simulations of the catalysis, which—along with the structural, biophysical and biochemical analysis—reveals how condensation is catalysed in NRPSs.

Complex formation

To gain maximum structural insights into peptide bond formation by NRPSs, we constructed samples that represent the condensation state, and that are long-lived. One classical approach to increase the stability of complexes is to substitute labile moieties with similar non-hydrolysable groups. We and others have used amides or thioethers in place of the labile thioesters in acyl-ppants to obtain crystal structures of C domains bound to one of the two condensation substrates[4,13,14,30]. In these studies, non-hydrolysable analogues of acyl-coenzyme A molecules are chemoenzymatically prepared, and the promiscuous ppant transferase Sfp is used to place the acyl-ppant onto the serine attachment point in the T domain[16,31,32]. However, this approach must be modified to obtain the full condensation complex of a typical NRPS, because NRPSs usually contain multiple T domains in the same protein chain. When incubated with several acyl-CoA molecules and an NRPS protein containing two or more T domains, Sfp would randomly ligate the acyl-ppant moieties to the various T domains, giving a heterogeneous, rather than the desired homogeneous, covalent complex. We therefore undertook a divide-and-conquer, multi-step strategy for complex formation whereby two modules are expressed separately and combined with protein ligation.

The first subunit of linear gramicidin synthetase, LgrA, contains domains $F_1A_1T_1C_2A_2T_2E'_2$ (where F is the tailoring formylation domain; A is the adenylation domain; T is the transport domain; C is the condensation domain; E′ is the inactive epimerization domain). Domain C_2 catalyses peptide bond formation between donor formyl-valine-ppant-T_1 and acceptor glycine-ppant-T_2 (Fig. 1a). We have previously generated monomodular and dimodular constructs of LgrA and know them to be robust samples[13,14]. We therefore expressed and purified module 1 ($F_1A_1T_1$) and module 2 ($C_2A_2T_2$) separately, with small C and N terminal extensions that allow them to be fused using protein ligation. Attempts to ligate the modules using sortase A[33] led to only modest amounts of the assembled dimodule and no crystals, so we turned to an engineered variant of *Oldenlandia affinis* asparagine endopeptidase 1 (AEP1)[15]. AEP1 can fuse proteins which have the sequence Asn-Gly-Leu at their C termini with proteins containing Gly-Leu at their N termini, excising a Gly-Leu dipeptide and leaving only a small Asn-Gly-Leu scar. Optimization of the AEP1 ligation conditions led to around a 50% yield of ligated dimodular $F_1A_1T_1$–$C_2A_2T_2$ (Fig. 1), which could be purified by column chromatography. Tripeptide biosynthesis assays, in which a terminating condensation domain from a bacillibactin synthetase is fused to T_2

to enable multiple turnover, show that the small AEP1 scar, located in middle of the flexible linker region between T_1 and C_2, does not affect NRPS activity[14] (Extended Data Fig. 2a). Thus, $F_1A_1T_1$-Asn-Gly-Leu and Gly-Leu-$C_2A_2T_2$ were each expressed and purified in large scale. In parallel, fVal-$_{NH}$CoA (molecule **2**), Gly-$_{NH}$CoA (**3**) and fVal-Gly-$_{NH}$CoA (**4**) were prepared by chemoenzymatic syntheses (Supplementary Methods). To construct the pre-condensation complex, Sfp was used to ligate the fVal-$_{NH}$ppant portion of fVal-$_{NH}$CoA onto T_1 of $F_1A_1T_1$-Asn-Gly-Leu in one reaction, and to ligate Gly-$_{NH}$ppant onto T_2 of Gly-Leu-$C_2A_2T_2$ in a separate reaction (Fig. 1b). The modules were then attached together with AEP1, and the complex was purified. Analogously, for the post-condensation state, ppant was ligated onto $F_1A_1T_1$-Asn-Gly-Leu and fVal-Gly-$_{NH}$ppant was ligated onto Gly-Leu-$C_2A_2T_2$. The modules were attached together and the resulting complex was purified. Crystallization trials and optimizations led to a new crystal form, from which the structures of the pre- and post-condensation complexes could be determined at a resolution of 2.9 and 3.0 Å (Figs. 2 and 3 and Extended Data Table 1).

Structures of the full NRPS condensation state

The pre-condensation $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ complex is a >1,800 residue, single-chain, covalent assembly (Figs. 1 and 2a). The rigid catalytic platform[13,26] portions of each module (F_1A_{core1}, C_2A_{core2}; where A_{core} represents the large N-terminal subdomain of the A domain) lie across one another, with F_1 and C_2 stacked and A_{core1} and A_{core2} splaying out at right angles from each other. A_{sub1} (the small C-terminal subdomain of the A domain) and A_{sub2} each reach back towards the complex's centre, allowing T_1 and T_2 to bind to the opposing donor and acceptor sites of the V-shaped C domain. This crystal structure has two molecules in the asymmetric unit that are very similar, although molecule B displays weaker density for $A_{sub2}T_2$. The post-condensation structure $F_1A_1T_1$–$C_2A_2T_{2\text{-fVal-Gly}}$ adopts the same overall conformation as the pre-condensation structure (Extended Data Fig. 2b), which is similar to the low-resolution substrate-free $F_1A_1T_1C_2A_2T_2$ structure conformation[14] (Extended Data Fig. 3a,b,d and Supplementary Discussion). These structures all have the same relative orientation of the two C domain lobes (C-lobes) as LgrA constructs bound to a single (donor) condensation substrate (Extended Data Fig. 2c). The segments of the N-lobe that reach over to interact with the C-lobe, called the latch loop (residues 1124–1138) and the floor loop (1044–1057)[10], are ordered and engaged in their cross-lobe interactions (Extended Data Fig. 2d). Non-covalent domain–domain interactions between the C and F_1, A_{sub1} and A_{sub2} domains were interrogated by mutagenesis and the mutants were found to not impair activity (Extended Data Fig. 3f–i and Supplementary Discussion).

$T_{1\text{-fVal}}$ is positioned at the donor-binding site of C_2, and $T_{2\text{-Gly}}$ is at the acceptor site (Fig. 2b and Extended Data Fig. 4). The conformations of the T domains are very similar to each other, and to previous observations of LgrA T_1 and T_2, highlighting that their most important structural flexibility is in the conformation of their ppant moieties (Extended Data Fig. 5 and Supplementary Discussion). Notably, the surface with which T_1 binds to the C_2 donor site and the surface with which T_2 binds to the C_2 acceptor site are site largely overlapping (Fig. 2c–g). Thus, T domains do not use one designated surface to bind to the acceptor site, and a separate designated surface to bind to the donor site (Extended Data Fig. 5 and Supplementary Discussion). Also note that the C-domain donor site is located in a depression between the C_2 N-lobe and C-lobe, but T_1 makes contacts only with the C-lobe (Fig. 2e and Extended Data Fig. 5e). At the C-domain acceptor site, located in an opposite depression between the C_2 N-lobe and C-lobe (Fig. 2b), T_2 makes most of its interactions with the N-lobe (Fig. 2f). Thus, the donor T domain, which is N-terminal to the C domain, interacts only with the C-terminal lobe of the C domain, while the acceptor T domain, which is C-terminal to the C domain, interacts mainly with the N-terminal lobe of the C domain (Extended Data Fig. 4f,g). These observations have

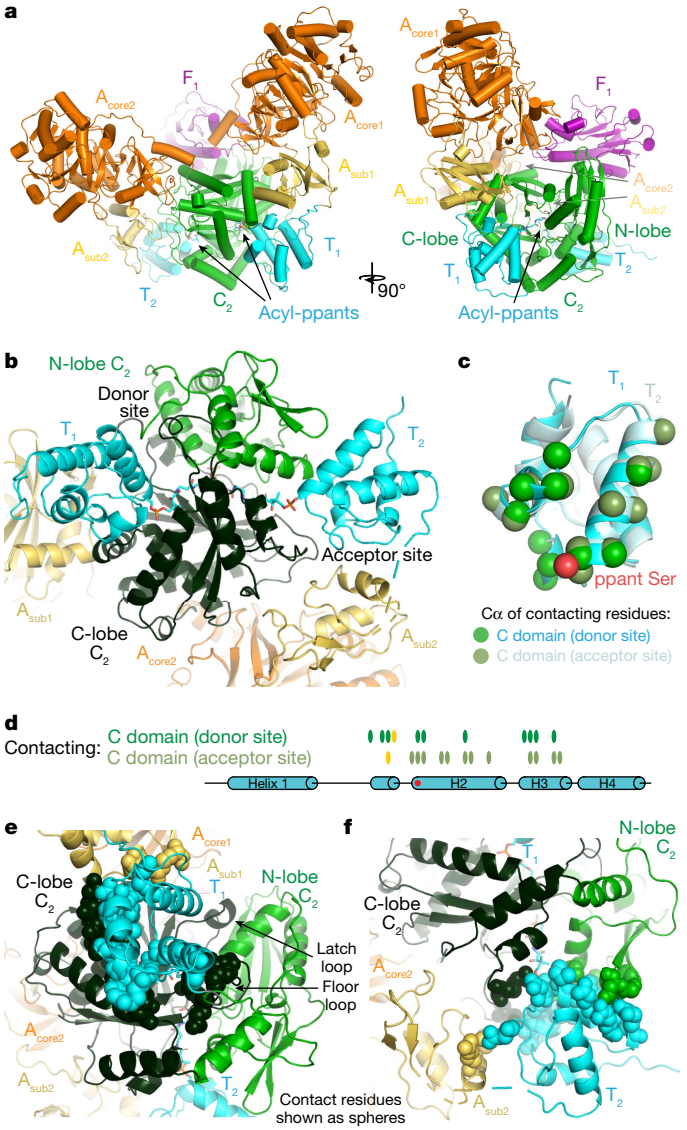


Fig. 2 | The structure of the pre-condensation state. a, Overview of the pre-condensation $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ complex. **b**, A view of the pre-condensation complex highlighting the donor and acceptor binding sites of domain C_2, and the N-lobe (residues 771–952) and C-lobe (residues 952–1199). The lobes each evolved from a duplication and specialization of a chloramphenicol acetyl transferase gene[18]. **c**, Superimposition of T_1 and T_2, indicating which residues are in close proximity (<4 Å) to the C_2 domain in the condensation conformation, by displaying those residues' α carbons. **d**, The residues in **c**, shown on a primary sequence domain diagram. Green dots represent T domain contacts with C domains and yellow dots represent T domain contacts with A_{sub} domains. **e**, Contacting residues at the C_2 donor site shown as spheres. Note that all of the contacting C_2 donor site residues are C-lobe residues. **f**, The contacting residues at the C_2 acceptor site shown as spheres. Note that most of the contacting C_2 acceptor site residues are N-lobe residues.

important consequences for module-swapping type bioengineering experiments—they dictate that there are no cut-points anywhere in an NRPS module that will prevent the introduction of non-native T-domain–partner-domain interactions in such a bioengineered NRPS (Supplementary Discussion).

The pre-condensation complex appears to be well configured for peptide bond formation. Strong density is evident for the fVal-$_{NH}$ppant moiety attached to T_1 at Ser729, and the Gly-$_{NH}$ppant moiety attached to T_2 Ser1760 (Fig. 3a and Extended Data Fig. 4). The ppant moieties each lead from their respective T domains down opposing sides of

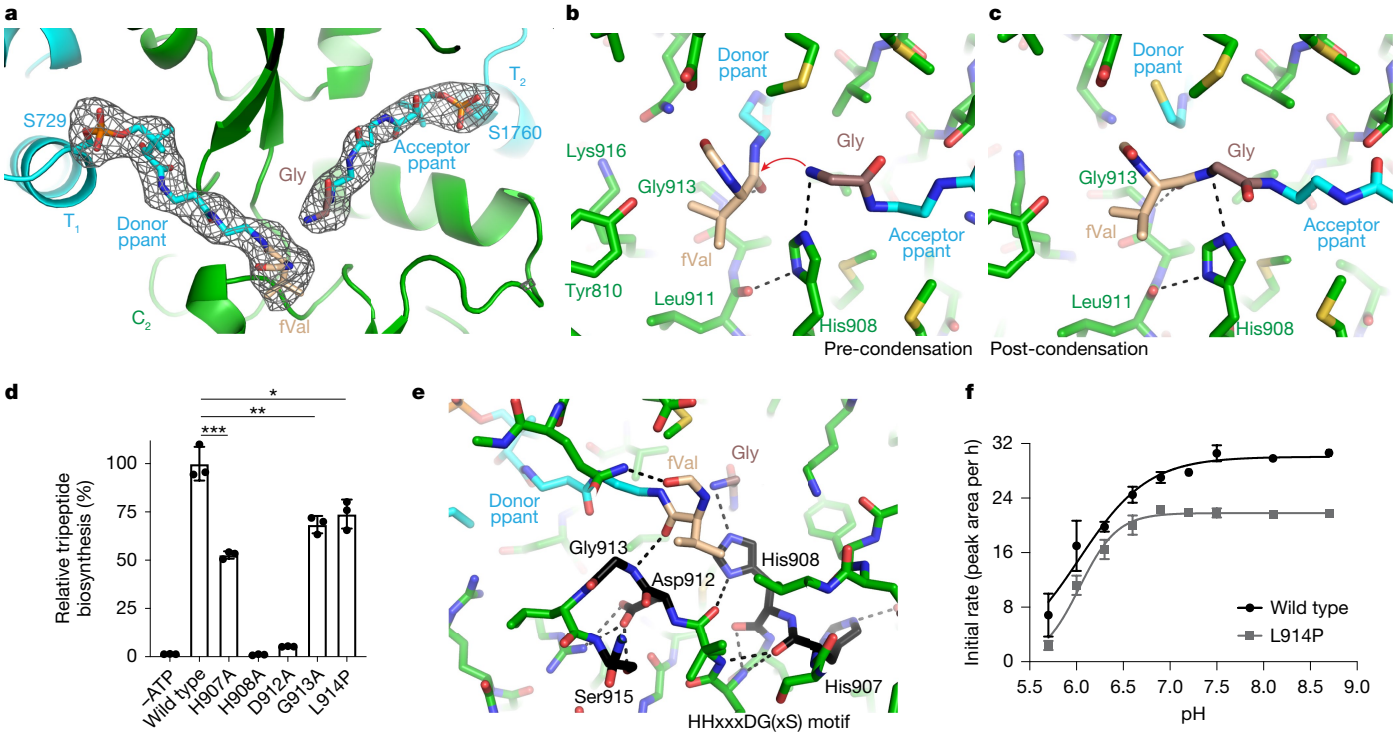


Fig. 3 | Structures and active-site mutagenesis. a, The condensation domain active site of $F_1A_1T_{1\text{-}fVal}$–$C_2A_2T_{2\text{-}Gly}$ and a Polder map (5 σ, no 'carving') of acyl-ppant moieties. **b**, Pre-condensation $F_1A_1T_{1\text{-}fVal}$–$C_2A_2T_{2\text{-}Gly}$; the red arrow is drawn between the nucleophile and electrophile of the reaction. The fVal-$_{NH}$ppant-T_1 carbonyl to Gly913 backbone hydrogen bond is partially obscured. **c**, Post-condensation $F_1A_1T_1$–$C_2A_2T_{2\text{-}fValGly}$. **d**, Tripeptide synthesis by $F_1A_1T_1$–$C_2A_2T_2$–$C_{BmdB\text{-}M3}$ and C_2 mutants thereof. Data are mean ± s.d. Individual points of the triplicate experiments ($n = 3$) are shown. The replicates for each individual bar were performed with one preparation of protein each, but repeated preparations of those proteins repeatedly showed analogous activity. Statistical significance between the wild type and mutants was determined using two-sided Student's t-tests; ***$P = 0.0008$ (wild type versus H907A), **$P = 0.050$ (wild type versus G913A), *$P = 0.0170$ (wild type versus L914P), ****$P = 0.00004$ (wild type versus H908A), ****$P = 0.00005$ (wild type versus D912A). **e**, The $F_1A_1T_{1\text{-}fVal}$–$C_2A_2T_{2\text{-}Gly}$ active-site motif (black) and tight Asp912-Gly913-Leu914-Ser915 turn. Ser915 is 89% conserved in $^{L}C_L$ and $^{D}C_L$ domains (C domains that act on donor substrates where the amino acyl residue attached to the donor T domain ppant has L or D chirality, respectively) and 82% conserved in all of the C domains, higher than Gly913 (83%, 79%), so the active-site motif could be considered to be HHxxxDG or HHxxxDGxS. **f**, Initial tripeptide synthesis rates (area under the high-performance liquid chromatography peak, (mAU_{280nm} min) h^{-1}) at different pH values for $F_1A_1T_1$–$C_2A_2T_2$–$C_{BmdB\text{-}M3}$ and L914P, where C_2 condensation is rate limiting. In the wild type, C_2, $C_{BmdB\text{-}M3}$ or other steps could be rate limiting. Trend lines are visual aids. Datapoints represents the averages of initial rates ($n = 3$ replicates), calculated as the slope of the line plots of peak area over time (Extended Data Fig. 6d,e). The error bars represent the s.d. of the three initial-rate measurements. Statistical significance is reported in Extended Data Fig. 6d,e. The observed pH profiles could be directly related to deprotonation of the His908 ε-nitrogen with increasing pH, allowing the crucial interaction with the amino nucleophile, but they are not definitive.

the substrate tunnel and into the C domain of the active site, where they meet at a near perpendicular orientation (Figs. 2b and Fig. 3 and Extended Data Fig. 4). The density maps (Fig. 3a) show that the nucleophile α-amino nitrogen of the acceptor Gly-$_{NH}$ppant-T_2 substrate is positioned by a 3.1 Å hydrogen bond to the ε nitrogen of His908 (the second His of the HHxxxDG motif). The nucleophile is at 110° and 3.3 Å from the electrophile carbonyl carbon of the donor fVal-$_{NH}$ppant-T_1 substrate—an excellent pre-attack position (Fig. 3b). Moreover, the carbonyl oxygen of the donor fVal-$_{NH}$ppant-T_1 is engaged in a 3.0 Å hydrogen bond to the backbone amine of Gly913 of the catalytic motif, suggesting roles in positioning and stabilizing the TS. No other protein residues are in the vicinity of the reactive atoms. The substrate-bound C domain does not occlude solvent from the active site, but no well-ordered waters are visible within ~4 Å of the nucleophile or electrophile. Important for catalytic considerations, the carbonyl of Leu911 is making a 2.9 Å interaction with the δ nitrogen of His908, indicating that this δ nitrogen is protonated. The side chain of His908 does not interact with any other protein residues. The fVal side chain nestles into a shallow pocket formed between active-site motif residues His908-Leu911-Asp912 and Val808-Tyr810-Lys916, while the acceptor Gly C_α faces a water-filled portion of the active site, contiguous with bulk solvent.

The post-condensation complex also shows strong density for both the T_1 ppant and the T_2 fVal-Gly-$_{NH}$ppant (Extended Data Fig. 4e). There may be several conformations of the T_1 ppant thiol, as some density bridges the acceptor and donor sites but, overall, the density is well fit by a single model (Fig. 3c and Extended Data Fig. 4e). The carbonyl oxygen of the product amide is still accepting a hydrogen bond from the Gly913 backbone amine, and its amine (the former nucleophile) is still donating a hydrogen bonded to the ε nitrogen of His908, once again highlighting their centrality to the C-domain function (Fig. 3c). These snapshots provide an important view of the key catalytic step of non-ribosomal peptide synthesis and enable directed biochemical and quantum dynamic interrogation of C domain catalysis.

Mutational and pH profile analyses

In the presented structures, the side chain of His908 and the main chain of Gly913 are the only two C-domain moieties observed in positions that would allow them to have direct roles in catalysis. We next examined these and other active-site residues using mutational analyses using the tripeptide synthesis assay. As expected, the His908Ala mutation resulted in complete loss of activity[4,6,11] (Fig. 3d). This again confirms the importance of His908 for reaction, but does not clarify

its role. Assessing the importance of the Gly913 amine is challenging, as it is not possible to substitute the backbone amine. Gly913 is central to the tight Asp912-Gly913-Leu914-Ser915 turn, which includes the end of the active site motif (Fig. 3e). A Gly913Ala mutation decreases peptide synthesis by around 35%, probably due to a minor displacement of the donor ppant. We modelled an alanine at position 913 of the pre-condensation structure by maintaining the Gly913 ψ and φ angles, which are within allowed Ramachandran values, and introducing a β carbon. This modelled β carbon projects into the active site tunnel, 3.2 Å from the penultimate donor ppant carbon. As a mutation to introduce a side chain at residue 913 is unlikely to displace its backbone amine, we also evaluated a mutation in the downstream neighbour of Gly913, Leu914Pro, which alters the tight turn (Fig. 3d). Here, an approximately 25% decrease is also seen, which we ascribe to displacement of the Gly913 amine, and is suggestive that the backbone amine aids in catalysis. We also generated the His907Ala and Asp912Ala mutants and, as expected, found them to be impaired in catalysis, retaining 51% and 5% of observed tripeptide biosynthesis activity, respectively[3,6] (Fig. 3d). Circular dichroism (CD) showed that the mutations did not alter the global folding of the enzymes at the temperature used for the peptide synthesis assay (Extended Data Fig. 6a). However, melting temperature analyses using differential scanning fluorimetry showed that the mutants other than His907Ala and Asp912Ala have the same or marginally higher T_m in comparison to the wild type, but that the His907Ala and the Asp912Ala mutations destabilize the protein, each decreasing the protein T_m by around 5 °C (Extended Data Fig. 6b). This is typical behaviour of 'second shell residues', which contribute to fine-tuning of the active-site structure, but are not the catalytic residues[34]. The structures show that His907 and Asp912 reach away from His908 and make three and six hydrogen bonds to the local backbone and their salt-bridging partners, respectively (Extended Data Fig. 6c,d).

We next performed pH profiles using the tripeptide biosynthesis assay. Enzymes that use general base/general acid mechanisms can show a peak in activity around the pKa of their general base/acid[35,36]. If, instead of proton abstraction, the key interaction is acceptance of a hydrogen bond from the reaction's nucleophile, a profile of increasing activity with increasing pH up to above the pKa of the interacting residue, and then no additional increase, might be expected. pH profiles of wild-type $F_1A_1T_1C_2A_2T_2$–C_3 and the Leu914Pro mutant, in which we know that C_2 condensation is rate limiting for tripeptide synthesis from kinetic experiments (Extended Data Fig. 6e,f), both show low activity at the lowest pH assayed, pH 5.7. Activity increases to a maximum near neutral pH, and the rate is maintained up to the highest pH that we assayed, pH 8.7 (Fig. 3g,h). This profile is consistent with (albeit not determinative of) a mechanism that features crucial hydrogen bond donation to a titratable residue such as His908.

Computational study of the condensation mechanism

Next, the pre-condensation structure was used to examine the reaction mechanism computationally. We initially performed molecular dynamics (MD) simulations on the enzyme with substrates bound, but found the substrates separated from the catalytic residues over the time of the MD simulation, which is not unusual for an enzyme with spacious active sites. We next proceeded to perform quantum mechanical DFT calculations on the mechanism of peptide bond formation in the presence of constrained active-site residues, which we call a theozyme[37].

We performed theozyme calculations in which full quantum mechanical calculations are performed on peptide bond formation using the amine nucleophile and thioester electrophile substrates (reactants) and the enzyme residues in contact with these reactants (Fig. 4). The theozyme was created by extracting the reactant and relevant protein residues coordinates of $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ and replacing the stabilizing amides for native thioester moieties. We first started with computations in the gas phase, which represents the extreme of a non-polar hydrophobic binding site. The barriers for formation at the transition state (TS) and intermediate (Int) are only 12.6 and 13.3 kcal mol^{-1} (Fig. 4a and Extended Data Fig. 7). A spontaneous reaction is quite feasible, with a barely bound tetrahedral intermediate with appreciable (0.2 e$^-$) charge separation. The Hirshfeld charges of fragments in each species are listed in Fig. 4a.

Notably, the free energy of Int is higher than that of the TS (Extended Data Fig. 7). The quantum mechanical calculations are performed in the potential energy surface, which shows a very shallow Int. However, when zero-point and thermal energies and the $-T\Delta S$ terms are added to give free energies, an intermediate is no longer present on the free-energy surface. That is, the quantum mechanical calculations predict a concerted pathway. The His908 stabilizes the TS, as reflected in the short N-H distance, but there is no proton transfer, which could have occurred if favourable in the quantum mechanical calculation. Subsequent deprotonation of the developing protonated amide product by the thiolate leaving group completes the reaction.

To mimic a more polar reaction environment in the active-site protein, we also studied the reaction in water. We found that the results are very similar to those in the gas phase (Fig. 4a,b and Extended Data Fig. 7). The highest-energy point on the Gibbs free-energy surface is now higher, 15.8 kcal mol^{-1} versus 12.5 kcal mol^{-1}, but is still very low. The energy in the actual environment is somewhere between these, probably around 14–15 kcal mol^{-1}.

To further examine the roles of His908 and Gly913, we replaced their functional groups and repeated the quantum mechanical calculations. Replacing the Gly913 backbone amine with a methylene interferes with the hydrogen bonding of the reactant carbonyl oxygen. This is unfavourable for reaction, with the highest-energy point on the Gibbs free-energy surface increasing from 15.8 kcal mol^{-1} to 17.3 kcal mol^{-1}, and the corresponding distance of H···O$^-$ increasing from 1.87 Å to 2.11 Å (Extended Data Fig. 7). The energy increase is worth about a factor of ten in rate. Replacing the imidazole group of His908 with a methyl group eliminates the hydrogen bonding acceptor to the nucleophile (Fig. 4c and Extended Data Fig. 8b). The TS for C-N bond formation is now 20.6 kcal mol^{-1}, 4.8 kcal mol^{-1} higher than when the theozyme has His908 intact. The TS is quite early, the forming C-N bond is very long and the C-S bond is shorter. The C-NH_2 moves near the S, presumably to gain some H-bond or electrostatic stabilization, as there is no imidazole to stabilize the developing NH_2^+. This result shows that His908 is important for lowering the barrier of the reaction—5 kcal mol^{-1} is worth 1,000 in rate at room temperature.

These quantum mechanical results predict that His908 functions as a key stabilizing group for the TS by acting as a hydrogen bond acceptor for the developing ammonium as the amine attacks the carbonyl, and not as a general base. This mechanism is a concerted reaction, facilitated by the good leaving group thiolate, which also deprotonates the ammonium.

Discussion

The full NRPS condensation state structure has been challenging to obtain. The first apo structure of a C domain was determined in 2002[3], yet it took almost two more decades for a C-domain co-complex with a single T-domain-delivered substrate analogue[14]. The current literature includes 20 publications describing 52 structures of or including C domains, some of which feature efforts to obtain co-complexes by small-molecule soaking or co-crystallization[8,23,24,38], chemical biology probes and warheads[9,31,39], and ligation by 'spy-catcher' and sortase[4,40], many of which yielded interesting, albeit incomplete, structures. The presented full complexes enable the evaluation of proposals that accompany partial structures from our laboratory and other laboratories, and delineate which components of the 1,800+ residue macromolecular machine are most important for chemical catalysis of condensation. The presented structures provide strong evidence

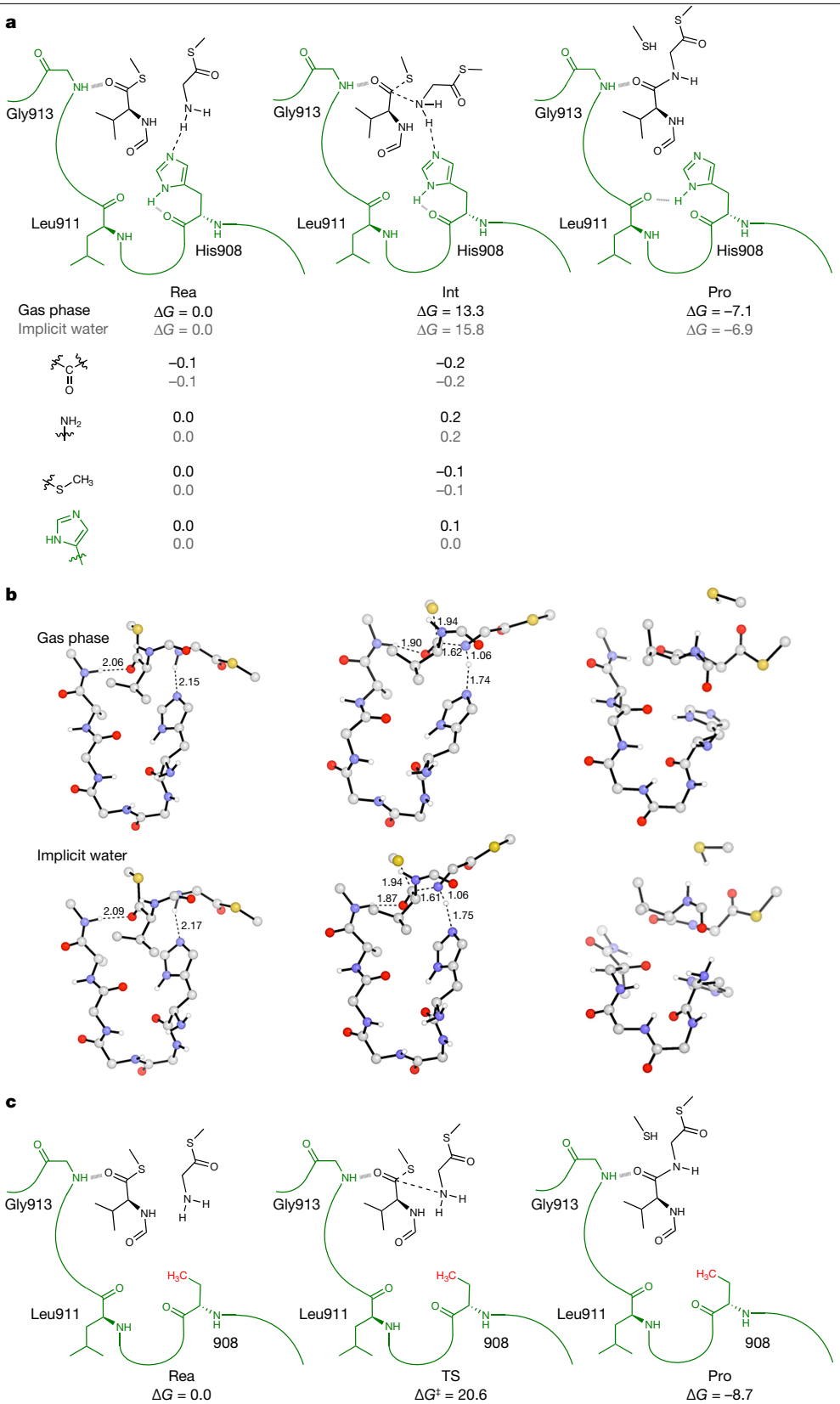


Fig. 4 | Quantum mechanical calculations. a, The theozyme and fragments of the reacting amine and thioester reactants. The curved lines represent groups that were omitted in the computations; we constrained the backbone atoms of the protein chain in the theozyme calculations. The energies are computed free energies for calculations in the absence of solvent (gas phase) or in a CPCM water model. The free energies of stationary points on the potential energy surface for the peptide formation with no solvent (gas phase) and in CPCM water solvent (implicit solvent) are shown below the structures. Hirshfeld charges of each fragment along the reaction pathway are shown. There is no significant difference in the absence or presence of water solvent. Rea, reactants; Pro, products. **b**, The computed structures in gas and water, corresponding to the chemical schematics in **a**. For clarity, non-polar H atoms are omitted except on N and S atoms. The values represent distances in angstrom (Å). **c**, Reaction energetics with imidazole is replaced by CH_3 group. The double dagger symbol refers to the transition state.

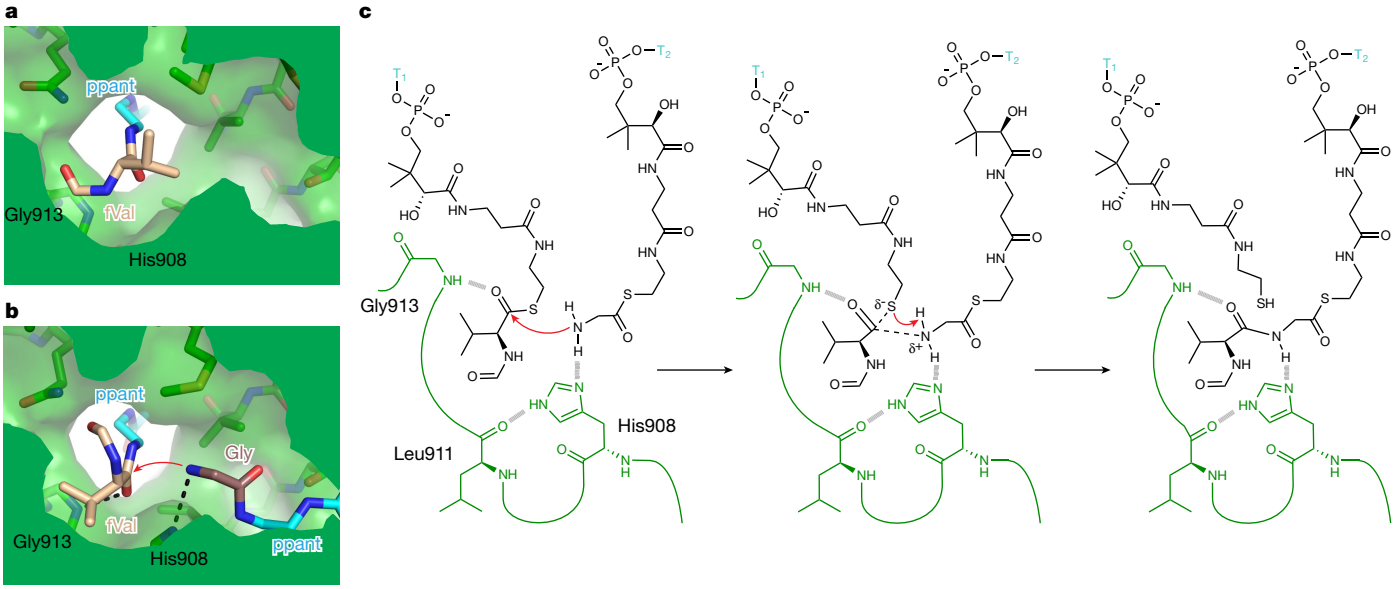


Fig. 5 | Mechanism of peptide bond formation. a, A view of C_2 bound to only fVal-$_{NH}$ppant-T_1[14] shows the carbonyl of the donor inaccessible to attack[50] by water or another nucleophile owing to C domain residues, the Val side chain and the formyl group. The carbonyl is not seen to form a strong H bond with Gly913 in this donor-only complex. **b**, A view of C_2 bound to fVal-$_{NH}$ppant-T_1 and Gly-$_{NH}$ppant-T_2 shows reorientation to an excellent pre-attack position. **c**, The concerted reaction mechanism of NRPS condensation.

against suggestions that there would substantial rearrangement when both substrates are bound for catalysis. With both substrates bound, the latch and floor loop engage in their traditional cross-lobe interactions (Extended Data Fig. 2d), rather than rearranging for a direct role in catalysis[8]. Similarly, the N- and C-lobes of LgrA C_2 do not 'close' when both substrates are present (Extended Data Fig. 2c). This is notable because comparison of structures of C domains from different NRPSs show the two lobes at different relative orientations and it was unclear whether this is related to the size of the native peptidyl substrates and/or whether lobes reorient for catalysis[23,41]. The consistency of conformations of C_2 across functional states, and the excellent pre-reaction position of substrate analogues reported here, is a strong indication that, if the proposed sub-lobe motion does occur, it does not have a critical induced fit role for accommodating T-domain-delivered substrates.

Notably, the structures show that binding of both substrates does not lead to an active-site motif rearrangement, which might be expected if Asp912, His907 or any other residue participated in an active site dyad with His908[6]. In the structures, His908 and Gly913 are the only residues directly interacting with reactive moieties, and they interact in the pre- and post-condensation complexes, leading us to assert that they are the NRPS residues positioned to act in chemical catalysis of peptide bond formation. We therefore used this pre-condensation conformation to create our theozyme, which did not contain side chains of other active-site-motif residues like His907 and Asp912. All computational simulations must include assumptions, and/or limit the number of atoms included owing to limitations in computational power and related approximations that are required to perform the computations, and here we assume that there is not a substantial rearrangement of the active site pre- and mid-condensation involving moieties that are not included in our theozyme. It is possible that such a rearrangement occurs, and that the large defects in peptide synthesis of the H907A and D912A mutants are because these residues act directly in catalysis. However, we believe that the evidence that their roles are important for the local conformation of the active site and domain stability is strong. First, the mutants display classic second-shell residue behaviour, a defect in catalysis accompanied by a decrease in T_m (Extended Data Fig. 6b). Second, although the presented structures are the most complete and representative of the active conformation, if Asp912 and/or His907 were prone to rearranging, one might expect to see such a rearrangement given enough observations of the enzyme. On the contrary, all published C domain, E domain, X domain and Cy domain structures, including all crystallographic independent chains, and all observations by electron microscopy (totalling over 100 independent observations)[42], always have these residues pointing away from the position of His908. The X domain[43] case is especially notable: X domains are relatives of C domains that do not possess any catalytic activity, but have a docking function in glycopeptide synthesis. They have lost the His908 equivalent from their motif, but retain those of His907 and Asp912. Conservation of residues within proteins that have no catalytic activity suggests roles in structure and protein stability rather than catalysis. Third, computational simulations in which the active-site motif could rearrange if it were favourable do not show such a rearrangement. We sought to use quantum mechanics/molecular mechanics, an approach that can include more atoms than the theozyme does, by using more approximate force-fields. Despite multiple attempts, we were unable to obtain convergent simulations. However, we performed several MD simulations informative for this question. In all of our simulations, including those in which the reactive atoms were constrained to 2.2 Å to help to stabilize the simulation, and also in those that started with the highest-energy species Int from theozyme simulation and kept the N-C distance at 1.6 Å, Asp912 and His907 never rearrange to interact with the TS acyl-ppants (Extended Data Fig. 8). We believe that all of these observations are fully consistent with Asp912 and His907 having roles in maintaining the local conformation and stability of the domain, but it is impossible to fully disprove a theoretical rearrangement.

Comparison of the present structure with ligand-bound C domains (Extended Data Fig. 9) highlights the challenges in understanding which feature and conformations of partially bound structures are altered in the C-domain reactive state. Acceptor-bound structures consistently show the interaction between the amino nucleophiles and analogous residues to His908, while the interaction between the carbonyl and Gly913 amine are typically not observed, despite our computation and biochemical analyses showing its high importance for the reaction. Several conformations observed in partially bound

structures clearly require movement of individual residues or the bound substrates themselves to allow the absent substrate to bind productively. One informative comparison is to LgrA with only donor substrate, compared to the pre-condensation LgrA complex (Fig. 5a,b). In the donor-alone complex[14], fVal-$_{NH}$ppant is in the donor site, but its carbonyl oxygen is not within H-bonding distance of Gly913, and is not quite in a reactive position. It is oriented such that the positions in which a nucleophile could attack its carbonyl carbon are occluded by a combination of the C-domain residues, the Val side chain and the formyl group (Fig. 5a). When the acceptor substrate is also bound, an altered, reaction-competent orientation is observed: the donor carbonyl now does interact with the Gly913 backbone amine, and the proximal face of the carbonyl is open to attack by the acceptor α-amine (Fig. 5b).

Our results indicate that, once in a reactive position, the acceptor α-amino group attacks the donor thioester moiety. As delineated by the quantum mechanics, the reaction proceeds through a concerted pathway in which the His908 imidazole and the Gly913 backbone amine stabilize the TS, and ends with proton transfer to the thiolate (Fig. 5c). This concerted pathway is consistent with extensive experimental and theoretical studies of nucleophilic acyl substituents, and the related nucleophilic aromatic substitutions that have been studied and discussed extensively for the past half century. While the two-step process involving a tetrahedral intermediate is often invoked for these reactions, experimental evidence and thermodynamic reasoning[44–46], kinetic isotope effects[47] and computational evidence[48] have shown that concerted mechanisms occur, particularly in cases of good leaving groups, as is the case for the thioesters explored here. Of particular relevance is a recent detailed experimental and theoretical study of nucleophilic aromatic substitutions[49]. Although attack of nucleophiles on aryl halides is often thought to involve Meisenheimer complexes—a type of tetrahedral intermediate—many cases are now known in which the tetrahedral intermediate is really a TS. That is, acyl and aryl substitutions are concerted reactions. We expect that C domains throughout the very large NRPS family, which can share below 20% sequence identity, and catalyse amide-bond formation between diverse acceptor and donor acyl moieties, will have variations in details of substrate binding, but propose that this mechanism is the universal way that NRPSs link together building-block substrates into their important bioactive natural products.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08417-6.

1. Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chem. Rev.* **105**, 715–738 (2005).
2. Hüttel, W. Echinocandins: structural diversity, biosynthesis, and development of antimycotics. *Appl. Microbiol. Biotechnol.* **105**, 55–66 (2020).
3. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526 (2002).
4. Izore, T. et al. Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. *Nat. Commun.* **12**, 2511 (2021).
5. Roche, E. D. & Walsh, C. T. Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry* **42**, 1334–1344 (2003).
6. Bergendahl, V., Linne, U. & Marahiel, M. A. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur. J. Biochem.* **269**, 620–629 (2002).
7. Marshall, C. G., Hillson, N. J. & Walsh, C. T. Catalytic mapping of the vibriobactin biosynthetic enzyme VibF. *Biochemistry* **41**, 244–250 (2002).
8. Zhong, L. et al. Engineering and elucidation of the lipoinitiation process in nonribosomal peptide biosynthesis. *Nat. Commun.* **12**, 296 (2021).
9. Bloudoff, K., Alonzo, D. A. & Schmeing, T. M. Chemical probes allow structural insight into the condensation reaction of nonribosomal peptide synthetases. *Cell Chem. Biol.* **23**, 331–339 (2016).
10. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L. O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792 (2007).
11. Kee, M.-J. C. Y. et al. Structural insights into the substrate-bound condensation domains of non-ribosomal peptide synthetase AmbB. *Sci. Rep.* **12**, 5353 (2022).
12. Zhang, J. et al. Structural basis of nonribosomal peptide macrocyclization in fungi. *Nat. Chem. Biol.* **12**, 1001–1003 (2016).
13. 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 (2016).
14. Reimer, J. M. et al. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, 6466 (2019).
15. Yang, R. et al. Engineering a catalytically efficient recombinant protein ligase. *J. Am. Chem. Soc.* **139**, 5351–5358 (2017).
16. Quadri, L. E. et al. Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585–1595 (1998).
17. Crosby, J. & Crump, M. P. The structural role of the carrier protein—active controller or passive carrier. *Nat. Prod. Rep.* **29**, 1111–1137 (2012).
18. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta* **1865**, 1587–1604 (2017).
19. Maruyama, C. et al. A stand-alone adenylation domain forms amide bonds in streptothricin biosynthesis. *Nat. Chem. Biol.* **8**, 791–797 (2012).
20. Fortinez, C. M. et al. Structures and function of a tailoring oxidase in complex with a nonribosomal peptide synthetase module. *Nat. Commun.* **13**, 548548 (2022).
21. Wang, J. et al. Catalytic trajectory of a dimeric nonribosomal peptide synthetase subunit with an inserted epimerase domain. *Nat. Commun.* **13**, 592 (2022).
22. Katsuyama, Y. et al. Structural and functional analyses of the tridomain-nonribosomal peptide synthetase FmoA3 for 4-methyloxazoline ring formation. *Angew. Chem. Int. Ed. Engl.* **60**, 14554–14562 (2021).
23. Bloudoff, K., Rodionov, D. & Schmeing, T. M. Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J. Mol. Biol.* **425**, 3137–3150 (2013).
24. Chang, C. Y. et al. Structural insights into the free-standing condensation enzyme SgcC5 catalyzing ester-bond formation in the biosynthesis of the enediyne antitumor antibiotic C-1027. *Biochemistry* **57**, 3278–3288 (2018).
25. Drake, E. J. et al. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).
26. 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 (2008).
27. Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C. & Gulick, A. M. Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. *J. Biol. Chem.* **291**, 22559–22571 (2016).
28. Kreitler, D. F., Gemmell, E. M., Schaffer, J. E., Wencewicz, T. A. & Gulick, A. M. The structural basis of *N*-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat. Commun.* **10**, 3432 (2019).
29. Tarry, M. J., Haque, A. S., Bui, K. H. & Schmeing, T. M. X-ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure* **25**, 783–793 (2017).
30. Ho, Y. T. C. et al. Exploring the selectivity and engineering potential of an NRPS condensation domain involved in the biosynthesis of the thermophilic siderophore fuscachelin. *Front. Catal.* **3**, 1184959 (2023).
31. Wheadon, M. J. & Townsend, C. A. Accurate substrate-like probes for trapping late-stage intermediates in nonribosomal peptide synthetase condensation domains. *ACS Chem. Biol.* **17**, 2046–2053 (2022).
32. Worthington, A. S. & Burkart, M. D. One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org. Biomol. Chem.* **4**, 44–46 (2006).
33. Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-mediated protein ligation: a new method for protein engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).
34. Chikunova, A. & Ubbink, M. The roles of highly conserved, non-catalytic residues in class A β-lactamases. *Protein Sci.* **31**, e4328 (2022).
35. Zhang, F. et al. Modulating the pH activity profiles of phenylalanine ammonia lyase from *Anabaena variabilis* by modification of center-near surface residues. *Appl. Biochem. Biotechnol.* **183**, 699–711 (2017).
36. Elfstrom, L. T. & Widersten, M. Catalysis of potato epoxide hydrolase, StEH1. *Biochem. J.* **390**, 633–640 (2005).
37. Tantillo, D. J., Chen, J. & Houk, K. N. Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* **2**, 743–750 (1998).
38. Folger, I. B. et al. High-throughput reprogramming of an NRPS condensation domain. *Nat. Chem. Biol.* **20**, 761–769 (2024).
39. Shi, C., Miller, B. R., Alexander, E. M., Gullick, A. M. & Aldrich, C. C. Design, synthesis, and biophysical evaluation of mechanism-based probes for condensation domains of nonribosomal peptide synthetases. *ACS Chem. Biol.* **15**, 1813–1819 (2020).
40. Ulrich, V. & Cryle, M. J. SNaPe: a versatile method to generate multiplexed protein fusions using synthetic linker peptides for in vitro applications. *J. Pept. Sci.* **23**, 16–27 (2017).
41. Kosol, S. et al. Structural basis for chain release from the enacyloxin polyketide synthase. *Nat. Chem.* **11**, 913–923 (2019).
42. Patel, K. D., MacDonald, M. R., Ahmed, S. F., Singh, J. & Gulick, A. M. Structural advances toward understanding the catalytic activity and conformational dynamics of modular nonribosomal peptide synthetases. *Nat. Prod. Rep.* **40**, 1550–1582 (2023).
43. Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E. & Cryle, M. J. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* **521**, 105–109 (2015).
44. Williams, A. Concerted mechanisms of acyl group transfer-reactions in solution. *Accounts Chem. Res.* **22**, 387–392 (1989).

45. Guthrie, J. P. & Pike, D. C. Hydration of acylimidazoles: tetrahedral intermediates in acylimidazole hydrolysis and nucleophilic-attack by imidazole on esters. The question of concerted mechanisms for acyl transfers. *Can. J. Chem.* **65**, 1951–1969 (1987).
46. Guthrie, J. P. Concerted mechanism for alcoholysis of esters—an examination of the requirements. *J. Am. Chem. Soc.* **113**, 3941–3949 (1991).
47. Hengge, A. C. & Hess, R. A. Concerted or stepwise mechanisms for acyl transfer-reactions of p-nitrophenyl acetate—transition-state structures from isotope effects. *J. Am. Chem. Soc.* **116**, 11256–11263 (1994).
48. Blake, J. F. & Jorgensen, W. L. Ab initio study of the displacement reactions of chloride ion with formyl and acetyl chloride. *J. Am. Chem. Soc.* **109**, 3856–3861 (1987).
49. Kwan, E. E., Zeng, Y. W., Besser, H. A. & Jacobsen, E. N. Concerted nucleophilic aromatic substitutions. *Nat. Chem.* **10**, 917–923 (2018).
50. Bürgi, H. B., Dunitz, J. D., Lehn, J. M. & Wipff, G. Stereochemistry of reaction paths at carbonyl centres. *Tetrahedron* **30**, 1563–1572 (1974).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

Methods

Cloning and mutagenesis

Plasmids encoding modules 1 and 2 of LgrA[14,51] for use in AEP1 ligation (pBacTandem_$F_1A_1T_1$-Asn-Gly-Leu and pBACTandem_Gly-Leu-$C_2A_2T_2$) were constructed by PCR amplification of pBacTandem_FATCAT[14], with the primers pBACt_FAT_NGL_Ins_For and pBACt_FAT_NGL_Ins_Rev, and pBACt_CAT_GL_Ins_For and pBACt_CAT_GL_Ins_Rev, respectively (Supplementary Table 1). Note that we chose the module 1/module 2 split and reassembly point to be between residues Val770 and Leu772, because it is within the flexible linker between modules. The reassembled protein would have only a three-amino acid Asn-Gly-Leu scar in place of Ser771. Reactions were treated with DpnI and then 2 µl of each was transformed in *Escherichia coli* XL10gold for in-cell ligation.

Plasmids for gene expression of mutant LgrA_$F_1A_1T_1C_2A_2T_2$–BmdB_C_{T3} ($F_1A_1T_1C_2A_2T_2$–C_{T3}) for tripeptide synthesis assays were made using site-directed mutagenesis of the parental plasmid (pBacTandem_FATCATC3)[14] with the primers LgrABmdB_H908A_For, LgrABmdB_H908A_Rev, LgrABmdB_D912A_For, LgrABmdB_D912A_Rev, LgrABmdB_G913A_For, LgrABmdB_G913A_Rev, LgrABmdB_L914P_For, LgrABmdB_L914P_Rev, LgrABmdbB_+NGL_For and LgrABmdB_+NGL_Rev (Supplementary Table 1). LgrABmdbB_+NGL_For and LgrABmdB_+NGL_Rev were used to incorporate the short AEP1 ligation scar, Asn-Gly-Leu, between the two modules. LgrABmdb_F1_C2_For and LgrABmdb_F1_C2_Rev were used to introduce the E866A, E867A and E868A mutations into the F_1–C_2 interface, and LgrABmdB_Asub1_C2_For and LgrABmdB_Asub1_C2_Rev were used to introduce the R1001A and K1002A mutations into the A_{sub1}–C_2 interface. LgrABmdB_E1107L_For, LgrABmdB_E1107L_Rev, LgrABmdB_Q974L_For and LgrABmdB_Q974L_Rev were used to introduce the E1107L and Q974L mutations into the A_{sub2}–C2 interface through two rounds of mutagenesis (Supplementary Table 1). For each mutagenesis, each primer was used in a separate amplification reaction for 15 cycles, before combining paired reactions for additional 20 cycles, followed by DpnI digestion and transformation of 2 µl into *E. coli* XL10gold for in-cell ligation.

All plasmid sequences were confirmed by Sanger sequencing (Genome Quebec) using the sequencing primers shown in Supplementary Table 1 and full plasmid sequencing (Plasmidsaurus).

Protein production and purification

LgrA-encoding plasmids were transformed into *E. coli* BL21 (DE3) entD-[52] (pBacTandem_$_{Octahis}$FAT$_{NGL}$, pBacTandem_$_{GL}$CAT) or *E. coli* BAP1[53] (pBacTandem_FATCATC3 and mutants), and colonies were selected on LB-agar supplemented with 50 µg ml^{-1} kanamycin. A single colony was used for growth of an overnight starter culture of 100 ml LB supplemented with 50 µg ml^{-1} kanamycin (LB-kan), grown at 37 °C. Then,10 ml of starter culture was used to inoculate each 1 l of LB-kan in 2.8 l flasks, and the cultures were grown at 37 °C and 220 rpm until reaching an optical density at 600 nm of 0.5–0.7, after which they were placed at 4 °C for 2 h. Protein production was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside and the cultures were incubated at 16 °C with 220 rpm shaking for 16 h. Cells were collected by centrifugation at 3,993*g* for 20 min and used immediately for protein purification, or stored at −20 °C

All protein purifications started with nickel affinity chromatography. Cell pellets were resuspended in buffer IMAC/CBP A (2 mM imidazole, 150 mM NaCl, 3 mM $CaCl_2$, 2 mM β-mercaptoethanol (β-ME), 25 mM HEPES pH 7.5) and cells lysed by sonication. The lysate was clarified by centrifugation at 48,000*g* for 30 min and the supernatant was applied to the 5 ml HiTrap IMAC FF column (Cytiva) equilibrated with buffer IMAC/CBP A. Protein was eluted with buffer IMAC B (buffer IMAC/CBP A with 250 mM imidazole), and the fractions assessed using SDS–PAGE. The fractions containing protein were pooled for further purification.

$F_1A_1T_1$-Asn-Gly-Leu purified by nickel affinity was applied to the MonoQ HR 16/10 column (Cytiva) equilibrated in buffer monoQ A (0.5 mM Tris(2-carboxyethyl)phosphine (TCEP), 25 mM HEPES pH 7.5). Protein was eluted using buffers monoQ A and monoQ B (monoQ A with 1 M NaCl), with a 10% isocratic wash over 2 column volumes (CV), a 22% isocratic wash over 2 CV and a gradient elution of 22–40% over 9 CV. The fractions were analysed by SDS–PAGE, and pure fractions were pooled and concentrated using a 10 kDa molecular weight cut off Amicon Ultra-15 centrifugation concentrator (Millipore-Sigma).

Gly-Leu-$C_2A_2T_2$ purified by nickel affinity was loaded onto a 30 ml calmodulin Sepharose 4B column (Cytiva) equilibrated in buffer IMAC/CBP A. Protein was eluted with buffer CBP B (2 mM imidazole, 150 mM NaCl, 3 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid, 2 mM β-ME, 25 mM HEPES pH 7.5). The fractions were analysed by SDS–PAGE and the cleanest fractions were pooled, concentrated with a 10 kDa molecular weight cut off Amicon Ultra-15 centrifugation concentrator (Millipore-Sigma) and incubated overnight with tobacco etch virus protease[54], while dialysed against buffer IMAC/CBP A, to remove affinity tags. Protein was then applied to HiTrap IMAC FF and calmodulin Sepharose 4B columns to remove tobacco etch virus protease and uncleaved Gly-Leu-$C_2A_2T_2$, and the flow-through was concentrated.

$F_1A_1T_1C_2A_2T_2$–C_{T3} and its mutants were purified by nickel affinity and calmodulin affinity (as described above), concentrated with 100 kDa molecular weight cut off Amicon Ultra-15 centrifugation concentrator (Millipore-Sigma) and further purified with the S200 16/60 column (Cytiva) equilibrated in SEC buffer (0.5 mM TCEP, 150 mM NaCl, 25 mM HEPES pH 7.5).

The plasmid for AEP1 was provided by the W. Bin laboratory from Nanyang Technological University (Singapore)[15]. The expression and purification was performed according to a previously published protocol[15], with an additional size-exclusion chromatography step using the S75 16/60 column (Cytiva) equilibrated in 0.5 mM TCEP, 150 mM NaCl, 50 mM Tris pH 6.5.

All protein that was not used immediately was flash-frozen in liquid nitrogen and stored at −80 °C. Gels of all proteins used in this study are shown in Supplementary Fig. 1.

Covalent condensation complex formation and purification

Amino acyl-coenzyme A analogues **1** ($_{NH2}$CoA) and **2** (fVal-$_{NH2}$CoA) were made as previously described[13,14,32,55] (Supplementary Methods). Compounds **3** (Gly-$_{NH}$CoA) and **4** (fVal-Gly-$_{NH}$CoA) were synthesized using the same methods as those used for **2** (Supplementary Methods). fVal-$_{NH}$ppant or ppant was transferred onto Ser729 of the T domain of LgrA module 1 by reaction of 40 µM $F_1A_1T_1$-Asn-Gly-Leu, 10 µM Sfp and 200 µM of fVal-$_{NH}$CoA or CoA in 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, overnight at room temperature (Supplementary Fig. 11). Gly-$_{NH}$ppant or fVal-Gly-$_{NH}$ppant was analogously transferred to Ser1760 of Gly-Leu-$C_2A_2T_2$. The reactions were monitored through intact liquid chromatography coupled with mass spectrometry (LC–MS), using the Agilent 1260 series high-performance liquid chromatography (HPLC) system, and an amaZon speed EDT (Bruker) ion-trap mass spectrometer. For the HPLC, a PLRP-S 1,000 Å, 5 µm, 2.1 × 50 mm (Agilent) column was used. Holo proteins were then purified by application to the Superdex 200 16/60 column equilibrated in AEP1 reaction buffer (0.5 mM TCEP, 150 mM NaCl, 50 mM Tris pH 6.5). The fractions were analysed using SDS–PAGE and those containing pure protein were concentrated and flash-frozen for later use in the AEP1-mediated ligation.

The optimized AEP1-mediated ligation reaction to join LgrA modules 1 and 2 contained 120 µM $F_1A_1T_1$-Asn-Gly-Leu, 210 µM of Gly-Leu-$C_2A_2T_2$ and 1 µM of AEP1 in AEP1 reaction buffer and was incubated for 1 h at room temperature. The resulting $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ or $F_1A_1T_1$–$C_2A_2T_{2\text{-fVal-Gly}}$ was purified using the monoQ HR 16/60 column (Cytiva) by loading in 90% buffer monoQ A/10% buffer monoQ B, an isocratic wash with 2 CV 10% buffer monoQ B and isocratic wash with 2 CV 20% buffer monoQ B. Protein was eluted with a gradient from 20% to 36% buffer monoQ B over 9 CV. Fractions that were shown by SDS–PAGE to contain purified $F_1A_1T_1$–$C_2A_2T_2$ were concentrated and further purified

using the Superdex 200 16/60 in SEC buffer, followed by a final concentration step.

Crystallography

$F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$ was subject to sparse matrix crystallization screening at 4 mg ml^{-1}, 8 mg ml^{-1} and 12 mg ml^{-1} at 4 °C and 22 °C, using 0.2 µl of crystallization solution and 0.2 µl protein in 96-well trays. Promising crystal hits were optimized by fine screening in 24-well format with drops of 2 µl protein and 2 µl crystallization solution and a 500 µl reservoir. The crystals used here were grown in hanging-drop format at 22 °C, with 12 mg ml^{-1} $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$ or $F_1A_1T_1–C_2A_2T_{2\text{-fVal-Gly}}$ and a crystallization solution of 1 M sodium tartrate pH 7.0 and 10 mM $MnCl_2$ using streak-seeding. Crystals were cryo-protected by incubation for several minutes in a solution of 0.5 mM TCEP, 150 mM NaCl, 25 mM HEPES pH 7.5, 1.1 M sodium tartrate pH 7.0, 10 mM $MnCl_2$ and 20% glycerol. Crystals were looped using 90° Angle-Tip Loops (MiTeGen; to minimize reflection overlap caused by long *b* axis) and vitrified by plunging into liquid nitrogen. X-ray diffraction datasets were collected using the remote-control features of beamline ID-8 (CMCF-ID) with a 0.954 Å wavelength beam and at a temperature of 100 K, at the Canadian Macromolecular Crystallography Facility of the Canadian Light Source. Data were indexed, integrated and scaled using the program HKL2000 (v.721)[56]. Initial phases were calculated by molecular replacement in Phaser v.2.9.0 using the full chain A (with domains $F_1A_1T_1C_2A_2T_2$) of Protein Data Bank (PDB) 6MTZ (ref. 14), followed by iterative refinement in the programs Phenix (v.1.20-4487)[57] and Coot (v.0.9.8.92)[58]. The structure resolutions were set by $CC_{1/2} > 0.3$ (as described previously[59]). The resolution of the pre-condensation structure is 2.9 Å, with $R_{work} = 0.216$ and $R_{free} = 0.258$, while that of the pre-condensation structure is 3.0 Å, with $R_{work} = 0.218$ and $R_{free} = 0.258$. These R_{free} values are better than the averages for structures of similar resolution (see the accompanying PDB validation reports), and are within the target ranges for well refined structures[60]. The *B* factors for the structures vary by domain, and are quite high for mobile domains like A_{sub}, as expected for large flexible, multidomain proteins[61] (Extended Data Table 1 and Supplementary Fig. 12). Ramachandran-favoured, Ramachandran-allowed and Ramachandran-outlier statistics are 95.15%, 4.54% and 0.31%, respectively, for the $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$ (pre-condensation) structure and 94.76%, 4.87% and 0.37%, respectively, for the $F_1A_1T_1–C_2A_2T_{2\text{-fVal-Gly}}$ (post-condensation) structure.

Tripeptide biosynthesis—mutagenic interrogation

To evaluate the activity of LgrA C_2 domain, a tripeptide synthesis assay of $F_1A_1T_1–C_2A_2T_2–C_{BmdB\text{-}M3}$ and C_2 domain mutants thereof was performed. $C_{BmdB\text{-}M3}$ catalyses peptide bond formation of fVal-Gly-ppant-T_2 with free tryptamine, releasing formyl-valinyl-glycyl-tryptamine. $C_{BmdB\text{-}M3}$ catalyses peptide bond formation of fVal-Gly-ppant-T_2 with free tryptamine, releasing formyl-valinyl-glycyl-tryptamine. Note that the decreases observed may represent a greater decrease in the rate of condensation by LgrA C_2 than they seem, as condensation chemistry by wild-type C_2 is probably not rate-limiting for tripeptide synthesis.

To synthesize 10-formytetrahydrofolate (10-fTHF), used by F_1 to formylate valine, 5-formyltetrahydrofolate (Sigma-Aldrich) was first converted to 5,10-methylenetetrahydrofole (5,10-mTHF) using a previously described protocol[14]. For tripeptide biosynthesis assays[14] with $F_1A_1T_1C_2A_2T_2–C_{T3}$ and mutants thereof, 0.2 mM 5,10-mTHF, 2 mM valine, 1 mM glycine, 4 mM tryptamine, 5 mM ATP, 0.7 mM $MgCl_2$, 1 mM TCEP, 150 mM NaCl and 50 mM HEPES pH 7.4 were pre-incubated for 10 min at 23 °C to convert 5,10-mTHF to 10-fTHF before 3.7 µM $F_1A_1T_1C_2A_2T_2–C_{T3}$ was added, and the reactions incubated for 5 h at 23 °C. Negative controls and complete reactions were performed in triplicate. The reactions were quenched with 300 µl of 4:1 butanol:chlorophorm, frozen at −80 °C overnight, and lyophilized to dryness. To dissolve the lyophilate, 50 µl of HPLC-grade methanol was added and vortexed. To pellet the remaining particulates and precipitates, the samples were centrifuged at 15,000*g* for 10 min. The samples were transferred to HPLC tubes and 40 µl was run over the Eclipse XDB-C8 LC column (Agilent) attached to the Agilent 1260 series HPLC system, using 0.1% trifluoroacetic acid (TFA) in water and 0.1% TFA in acetonitrile for buffers A and B, respectively. Using a flowrate of 0.5 ml min^{-1}, the following method was used: the sample was loaded onto the column equilibrated in 98% A and washed for 5 min with 98% A. A gradient was run from 98% A to 30% A over 40 min. Subsequently, a gradient from 30% A to 2% A was performed over 1 min. The column was washed with 2% A for 6 min and then with 98% A for 9 min before injecting the next sample or blank. UV at 220 nm and 280 nm was used to monitor the HPLC run. The fractions containing the product (at a retention time of 27.4–27.8 min) were identified by direct injection on an amaZon speed EDT (Bruker) ion trap mass spectrometer. Product quantification was carried out by integrating the area under the LC peak at 280 nm. Conversion of peak area to relative tripeptide biosynthesis and calculations of average tripeptide biosynthesis were carried out in Microsoft Excel v.16.90.2. Calculations of s.d. and plotting of bar plots was performed in GraphPad Prism v.10.2.3. Two-sided Student's *t*-test analysis was performed in Microsoft Excel v.16.90.2 and GraphPad Prism v.10.2.3.

Activity assays for the variants H908A, D912A and L914P were performed on a different day to those for G913A and the construct harbouring the ligation scar (Asn-Gly). On the 2 days, the wild-type control displayed very similar average LC peak area at 147.803 and 141.97. For each day, a relative yield was calculated for each trial of each variant as the percentage of the average of the three trials of the wild-type for that day. According to this relative yield, the G913A variant was plotted on the same bar graph as H907A, H908A, D912A and L914P (Fig. 3).

The conservation of active site residues Gly913 and Ser915 (Fig. 3 and Extended Data Fig. 9) was determined using a database of condensation domains filtered for maximum pairwise sequence identity of 90% (ref. 62), categorized as $^{L}C_L$ or $^{D}C_L$, or as any C domain[63] by the webserver NAPDOS. We have included a Source Data file containing the accession codes and active-site motifs of each of these sequences (Source Data Extended Data Fig. 9).

Tripeptide biosynthesis—pH profile

Pre-reaction mixes of 150 µl contained 100 mM 2-(*N*-morpholino) ethanesulfonic acid (MES), 100 mM piperazine-*N*,*N*′-bis (2-ethanesulfonic acid), 100 mM 3-[4-(2-hydroxyethyl)piperazin-1-yl] propane-1-sulfonic acid, 2 mM L-valine, 1 mM glycine, 4 mM tryptamine, 5 mM ATP, 0.7 mM magnesium chloride and 0.2 M 10-fTHF at pH 5.7, 6.0, 6.3, 6.6, 6.9, 7.2, 7.5, 8.1 or 8.7. The mTHF had been converted to 10-fTHF, mTHF by incubating in 1 mM TCEP, 150 mM NaCl and 100 mM HEPES pH 7.5 for 10 min at room temperature before addition to above reaction. Negative controls lacked ATP and were at pH 7.5. The reactions were started by addition of 3.7 µM of wild-type or L914P $F_1A_1T_1C_2A_2T_2–C_{T3}$. For the reaction with wild-type protein, 30 µl aliquots were taken 0, 30 60, 120 and 180 min and, for the reactions with L914P protein, 30 µl aliquots were taken at 0, 60, 120, 180 and 240 min. Aliquots were quenched by adding 300 µl of 4:1 butanol:chlorophorm and frozen at −80 °C. Frozen samples were dried by lyophilization and the lyophylate was resuspended in 50 µl of HPLC-grade methanol by vortexing and then centrifuged at 15,000*g* for 10 min. The samples were transferred to HPLC vials and 30 µl samples were injected into the Eclipse XDB-C8 LC column (Agilent) attached to the Agilent 1260 series HPLC system, using 0.1% TFA in water and 0.1% TFA in acetonitrile for buffers A and B, respectively. At 0.7 ml min^{-1}, the sample was applied onto the column equilibrated in 98% A and washed for 4 min with 65% A and 4 min at 64% A. Subsequently the flowrate was increased to 1 ml min^{-1} and the column was washed for 3 min at 2% A. Finally, the flowrate was decreased back to 0.7 ml min^{-1} and the column was washed with 98% A for 4.5 min. Elution of molecules was monitored by UV at 220 nm and 280 nm. By direct injection of fractions onto an amaZon speed EDT (Bruker) ion trap mass spectrometer the tripeptidyl product was confirmed at a

retention time of 5.7–5.8 min. The product of the reaction was quantified by integration of the HPLC peak at 280 nm.

Initial rates were determined by plotting the peak area versus time of each replicate and determining the slope using linear regression analysis in GraphPad Prism v.10.2.3. Profiles of pH versus initial rate were plotted in GraphPad Prism v.10.2.3. The error in the initial rate is calculated as the standard deviation of the three replicates in GraphPad Prism v.10.2.3. Significance was calculated using two-sided Student's *t*-tests in GraphPad Prism v.10.2.3 and Microsoft Excel v.16.90.2 (Extended Data Fig. 6).

CD analysis

CD analysis was carried out on Gly-Leu-$C_2A_2T_2$ wild type and mutants. Before CD analysis, all proteins were processed for gel-filtration chromatography with the S200 10/300 (Cytiva) column equilibrated in SEC buffer and diluted to 8 µM. CD spectra were collected in SEC buffer, between 190 and 260 nm at 1 nm increments and 0.5 s integration times, acquiring 9 spectra per protein sample with the Chirascan CD Spectrometer (Applied Photophysics) using the Chirascan v.4.7.0.194 software. Raw spectra were corrected by subtracting reference spectra collected in the presence of buffer only. Subsequently the spectra were processed with three-point arithmetic smoothing using the Chirascan Pro-Data Viewer software v.4.7.0.194 (Applied Photophysics). Spectral ranges were limited to wavelengths at which the total sample absorbance did not exceed 1.0 unit, as measured by the Chirascan CD Spectrometer.

Preliminary data processing and reference spectra subtraction were performed using the Chirascan Pro-Data Viewer software v.4.7.0.194. Secondary structural deconvolution of the wild-type Gly-Leu-$C_1A_1T_1$ of reference-corrected spectra using the CONTINLL algorithm within OLIS SpectralWorks software v.4.3 (On-Line Instrument Systems) showed very close agreement with the secondary structure shown in the X-ray structures, confirming the correct folding of the protein during the CD experiments. CD spectra were plotted in GraphPad Prism v.10.2.3.

DSF analysis

Samples for differential scanning fluorimetry (DSF) comprised 15 µl of double-deionized water, 5 µl of 5× DSF buffer (5 mM TCEP, 750 mM NaCl, 250 mM HEPES pH 7.5), 2.5 µl of SYPRO Orange dye (Invitrogen) diluted at 1:80 from the commercial stock and 2.5 µl of wild-type or mutant $C_2A_2T_2$, in wells of MicroAmp Fast 96-well Reaction Plates (Applied Biosystems). Experiments were performed in quadruplicate for each $C_2A_2T_2$ variant and a minus protein control. The plates were sealed with a Microseal 'B' seal (BioRad) and centrifuged at 1,000*g* for 1 min at 4 °C, before being transferred to the StepOnePlus real-time PCR system (Applied Biosystems) using the StepOne Software v.2.3. The plates were held for 5 min at 4 °C, followed by a temperature gradient of 4 °C to 95 °C, with a 0.36 °C per minute increase and an incubation for 15 s at 95 °C. Fluorescence was monitored using the FAM target option with excitation at 495 nm and emission at 520 nm. Data were analysed using the StepOne software v.2.3 and the Protein Thermal Shift software v.1.4, and melting temperature (T_m) values were calculated as the inflection point of the increase in fluorescence. Raw data were processed in Microsoft Excel v.16.90.2 to calculate the relative fluorescence versus temperature curves. Curves of relative fluorescence versus temperature were plotted in GraphPad Prism v.10.2.3. Average T_m values and the s.d. of the replicates were calculated and plotted (as dot plots) using GraphPad Prism v.10.2.3 (Extended Data Fig. 6b).

Computational analyses

For quantum mechanical theozyme calculations, constrained geometry optimization was performed at the level of wB97XD/def2-SVP[64]. Harmonic vibrational frequency calculations were performed at the same level of theory for all the stationary points to verify whether they were local minima or transition structures and to derive the thermal energy corrections. The solvent effect of water in the reaction was evaluated using the PCM solvation model[65]. Single-point energy calculations were based on the optimized geometries with wB97XD at a larger basis set of def2-TZVP[64]. All calculations were carried out using the Gaussian 16 program package (Gaussian). All thermodynamic quantities (1 mol l^{-1}, 298.15 K) were computed in the GoodVibes software v.3.1.1 (https://doi.org/10.5281/zenodo.60811). 3D renderings of stationary points were generated using PyMol 2.7 (Schrödinger).

All MD simulations (Extended Data Fig. 7) were performed using the Amber 16 package and the AmberTools 16 codes using a combination of ff14SB[66] force field for protein, TIP3P[67] parameters for water molecules, and the GAFF force field[68] and HF/6-31 G(d) level RESP atomic charges for ligands[69], applying periodic boundary conditions. The charges were calculated according to the Merz–Singh–Kollman scheme[70,71] using the Gaussian 16 program package. MD simulations were performed starting from the X-ray crystal structure including the bound substrate analogues. Chain B was removed, and the amides that had been incorporated to prevent substrate reactivity were replaced with thioesters that are present in the actual substrates in the active site. For the simulations of TSs, the bond-forming distance of the reactants was restrained to 2.2 Å, applying a harmonic constraint with a force constant of 1,000 kcal mol^{-1} $Å^{-2}$.The protein was solvated in a truncated 10 Å octahedron buffer of water box using the tleap module. A minimal number of counterions (Na^+) were added for charge neutralization. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method[72] with periodic boundary conditions. MD simulations were performed in the GPU-accelerated pmemd code[73]. After the preparation, MD simulations were performed according to the following steps: (1) two minimization steps were conducted in serial, each consisting of 2,500 steepest descent steps and 2,500 conjugate gradient steps. Protein and substrate atoms were restrained with a force constant of 500.0 kcal mol^{-1} $Å^{-2}$ in the first minimization step, allowing the solvent to minimize. Side-chain atoms were unrestrained in the second minimization, but backbone and substrate atoms were restrained with a force constant of 2.0 kcal mol^{-1} $Å^{-2}$. (2) The system was then equilibrated at 300 K, 1.0 atm for a total of 4 ns in the NPT ensemble, with the first 2 ns adopting a stronger positional constraint (30.0 kcal mol^{-1} $Å^{-1}$) and the latter 2 ns using a weaker positional constraint (0.5 kcal mol^{-1} $Å^{-1}$), with a constant temperature of 300 K. (3) Finally, production runs were performed in the NPT ensemble for 500 ns (free substrates or constrained substrates at TS bond length). All root mean squared deviation (r.m.s.d.) values were computed for non-H atoms with respect to the first frame of each production trajectory. Coordinate files were visualized, and figures were created, using the program PyMOL.

Analysis of domain–domain interactions in LgrA

The presented structures provide the opportunity to explore domain–domain interactions and compare conformations with previously determined structures[42,74]. The pre-condensation $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ and post-condensation $F_1A_1T_1$–$C_2A_2T_{2\text{-fVal-Gly}}$ structures represent subsequent steps in the catalytic cycle, and these structures are determined from crystals of the same crystal form, so it is not surprising that they have the same domain–domain orientations as each other (Extended Data Fig. 2b). The overall configuration is similar to that in the low-resolution, substrate-free $F_1A_1T_1C_2A_2T_2$ structure[14] (PDB: 6MFZ; Extended Data Fig. 3a,b,d), which was determined from an unrelated crystal form. Superimposition of the F_1A_{core1} catalytic platform of $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ and $F_1A_1T_1C_2A_2T_2$ shows an approximately 12° rotation of module 2 around a pivot point near C domain residue protein Asp898, which propagates to a 22 Å difference in position of the distal portion of A_{core2} (Extended Data Fig. 3d). This similarity is notable because NRPSs are highly flexible, so should not necessarily need to adopt one particular overall conformation for condensation, as many positions of F_1A_1 and C_2A_2 are compatible with the productive T_1–C_2 and C_2–T_2 interactions required in condensation substrate delivery. In contrast to this similarity, the $F_1A_1T_{1\text{-fVal}}C_2A_{core2}$ structure[14] (PDB: 6MFY; Extended Data Fig. 3c) is in a very

different conformation. It is also considered to be 'pre-condensation like', as it features $T_{1\text{-fVal}}$ in a general donation conformation, although the acceptor T_2 domain is absent from this construct. Relative to that of $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$, the second module of $F_1A_1T_{1\text{-fVal}}C_2A_{core2}$ is rotated by around 120°, displacing A_{core2} residues as much as 175 Å.

Notably, similar domain–domain interactions are observed when comparing LgrA structures that are in modestly different overall conformations (for example, $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$ versus low resolution, substrate-free $F_1A_1T_1C_2A_2T_2$), and when comparing LgrA structures that are in a very different overall conformation (such as $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$ versus $F_1A_1T_{1\text{-fVal}}C_2A_{core2}$). As expected, the F–$A_{core1}$ and C–A_{core2} interactions are consistent throughout the structures, as they form the known structural catalytic platforms of module 1 and 2, respectively[13,14,26,75]. Likewise, the T_1–C_2 and C_2–T_2 interactions, which are required for substrate delivery, are consistently observed (see the section below; Extended Data Fig. 5). However, we observe three substantial contacts between domains that do not seem to be necessarily required to interact for LgrA to function. First, in $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$ and $F_1A_1T_1C_2A_2T_2$ (as well as $F_1A_1T_{1\text{-fVal}}C_2$; PDB: 6MFW), one extremity of the C domain N-lobe nestles into the F domain near the formylation active site (Extended Data Fig. 3f). This interaction buries 802 Å^2 of surface area per domain and features C-domain residues Glu866, Glu867 and Glu868 hydrogen bonding with side chains of Arg83 and Asn177 and the backbone amines of Val87 and Phe87, and also hydrophobic packing with Phe10 and Tyr130. Second, in $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$, $F_1A_1T_1C_2A_2T_2$ and $F_1A_1T_{1\text{-fVal}}C_2A_{core2}$, a helix of the C domain C-lobe interacts similarly with A_{sub1}. Specific hydrogen bonding and buried surface area are not identical in each observation of the interaction, but in $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$, salt bridges between Glu599 and Arg1001, and between Asp649 and Lys1002, are central to the interaction, which buries a modest 357 Å^2 of surface area per domain (Extended Data Fig. 3g). Third, in $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$ and $F_1A_1T_1C_2A_2T_2$, a long loop in the C-lobe interacts similarly with A_{sub2}. $F_1A_1T_1C_2A_2T_2$ is too low resolution to observe side chains, but in $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$, Gln974 can be seen donating a hydrogen bond to the backbone amine of Arg1626, while Glu1107 hydrogen bonds to the backbone amine of Glu1684 and Tyr1685 in an interaction that buries 502 Å^2 of surface area (Extended Data Fig. 3h).

As each of these interactions was observed more than once in independent structures, we examined their importance to the LgrA bioactivity through mutagenesis and tripeptide synthesis. The triple point mutation E866A/E867A/E868A and double mutations K1002A/R1001A and E1107A/N974A were each introduced into $F_1A_1T_1C_2A_2T_2–C_{BmdB\text{-}M3}$. The proteins were purified and assayed for tripeptide synthesis (Extended Data Fig. 3i). Each of them had tripeptide synthesis indistinguishable from wild-type $F_1A_1T_1C_2A_2T_2–C_{BmdB\text{-}M3}$, indicating that these observed contacts are unlikely to be crucial for the NRPS catalytic cycle.

Thus, although the observed F_1–C_2, A_{sub1}–C_2 and C_2–A_{sub1} are favourable enough to have been seen multiple times, they are not required for proper substrate delivery, T_{n-1}–C_n and C_n–T_n domain interactions and condensation. These data are consistent with solution scattering data[14] and the idea that NRPSs are very flexible megaenzymes[29] capable of assuming many overall conformations that can allow condensation.

Analysis of T domain conformations

The current $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$ and $F_1A_1T_1–C_2A_2T_{2\text{-fValGly}}$ structures reveal the T domains performing their crucial delivery tasks, and also complete the set of observation of T_1 in all its functionally relevant states, which affords the opportunity to analyse the conformation and binding interactions of T_1, and of T domains generally. Superimposing all observations of LgrA T_1, in thiolation, formylation and donation/condensation states (this and previous studies[13,14]), shows only small variation in conformation of its 4-helix (plus one mini-helix) structure (Extended Data Fig. 5a). The Cα r.m.s.d. values are between 0.2 Å and 0.9 Å for the observations of T_1 binding to C_2. The cross-state comparison shows slightly higher r.m.s.d. values, with r.m.s.d. values of T_1 in a complex with C_2 versus A_1 versus F_1 in the range of 0.8–1.4 Å. The largest differences in positions are found at the very C terminus of the T domain, where the end of helix 4 shifts into slightly different directions (Extended Data Fig. 5a). This could be the result of variations in the direction of the T_1–C_2 linker, rather than a direct function of various partner domains, as helix 4 of the T domain usually does not interact with other domains (Extended Data Fig. 5h). Likewise, T_2 has r.m.s.d. values of 0.5–1.1 Å in its observations all bound to the acceptor site of C_2, also featuring small differences in the end of its helix 4 (Extended Data Fig. 5b).

The superimposition of all observations of LgrA T_1, including those in the thiolation, formylation and condensation donation states (Extended Data Fig. 5a), shows the highest variation to be in the conformation of the ppant arms, which is expected. However, in all states of T_1, the ppant moieties project in a highly similar direction, with only around 30° spread in approximately a single plane, despite the fact that these ppants reach into three completely different partner domain active sites (C_2 versus A_1 versus F_1). The ppant conformation for the C_n–T_n condensation acceptance state, which T domains in canonical initiation modules do not perform but T domains in elongation modules must, is as seen when T_2 binds to C_2 (Extended Data Fig. 5b), and falls outside those ~30°. This flexibility is useful at both large scale (entering different active sites) and at a more subtle scale: superimpositions of the structures of LgrA T_1 interacting with C_2 (Extended Data Fig. 5c (i and ii)) show that the flexible ppant can compensate for small variations in binding position of a T domain to allow successful delivery of the aminoacyl/peptidyl moieties to NRPS domain active sites.

The analysis of T domains and their ppants can be expanded to include other NRPS domains, by inspecting one example each of structurally characterized cognate interaction co-complexes of T domains in with each type of bone fide NRPS domain[12–14,76–79] (according to ref. 42; https://www.acsu.buffalo.edu/~amgulick/NRPSChart.html). This comparison suggests that many of the above-made observations may be more general (Extended Data Fig. 5d): the T domains superimpose well, with the most variation observed in the position of helix 4. The ppant position is variable (a near in-plane ~90° spread), but not as variable as one might imagine for a long, flexible appendage in a sizable set of interactions with diverse domains.

Analysis of T domain interactions with partner domains

Mapping the interaction surfaces of the LgrA T domains in the condensation state is informative. As discussed, the C domain has two T-domain-binding sites: the donor site for an upstream T domain, and the acceptor donor for a downstream T domain. The C domain also has two lobes: the N-terminal N-lobe and the C-terminal C-lobe. From these two facts and from simple domain schematic bubble diagrams (Fig. 1a and Extended Data Fig. 5f), it could be assumed that the upstream T domain binds to the N-lobe, and the downstream T domain binds to the C-lobe. However, it is established that the donor and acceptor sites are each depressions on opposite sides of the C domain structure, both of which are near the interface of the N- and C-lobes. T domains bind to these sites to place their acyl-ppants into the active-site tunnel, which originates at these depressions, and to which residues from both lobes help form[12,14,23,26] (Fig. 2 and Extended Data Fig. 5e).

As expected, $F_1A_1T_{1\text{-fVal}}–C_2A_2T_{2\text{-Gly}}$, $F_1A_1T_1–C_2A_2T_{2\text{-fVal-Gly}}$ and other structures show donor (T_1) and acceptor (T_2) T domains at these known binding sites, each situated between N- and C-lobes. However, donor T_1-domain binding involves contacts only between T_1 and the C-lobe of C_2 (Fig. 2e). The floor loop, a C-lobe element that reaches over and packs with the N-lobe, provides all of the binding residues on the N-lobe side of the donor-binding site. This feature of donor T_{n-1} only interacting with the C-lobe downstream C_n may not have been previously noted, and is certainly not generally appreciated in the field. Furthermore, the situation is roughly mirrored at the acceptor site: acceptor T_2-domain binding to its site in its depression between N- and C-lobes involves

mainly contacts with the N-lobe of C_2 and only two interactions with N-lobe residues (Fig. 2f).

Thus, there is a pattern of interaction where the upstream T_{n-1} domains bind to the C-terminal portion of the C_n domains, and the downstream T_n domains mainly bind to the N-terminal portion of the C_n domains. Transient interactions between each T domain and the other 'side' of their binding site (the less contacted/uncontacted lobe) can be important[14], but the pattern of these observed interactions has at least two consequences. First (and perhaps trivially), the interaction pattern is the opposite of the possible assumption based on the bubble diagrams (Extended Data Fig. 5f,g). Second (and importantly) the described pattern of interaction greatly affects bioengineering strategies: module-swapping type bioengineering experiments where the cut-point is chosen as between the N- and C-lobes[80] would not be expected to maintain native T_{n-1}–C_n or C_n–T_n interactions, but would rather introduce more non-native interactions than those with other cut-points[81–83].

It is also notable that the surface with which a donor T domain (T_1) binds to the C-domain donor site and the surface with which an acceptor T domain (T_2) binds to the C-domain acceptor side site largely overlap (Fig. 2c,d and Extended Data Fig. 5h,i). Both binding events involve the T domain residues from the mini-helix, helix 2 and helix 3, using an overlapping set of residues. The fact that T domains do not use one area to bind to the acceptor site, and another area to bind to the donor site, also has consequences for module-swapping type bioengineering experiments—it dictates that there are no cut-points anywhere in the T domain[84], or anywhere else in an NRPS module (Extended Data Fig. 5i,j), that can avoid introduction of non-native T domain–partner domain interactions in such a bioengineered NRPS. The explanation of successes of such bioengineering endeavours must be for reasons other than preserving native contacts.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

X-ray crystal structures and associated diffraction data for this study are deposited under accession codes 9BE3 and 9BE4 in the Research Collaboratory for Structural Bioinformatics Protein Data Bank. Source data are provided with this paper.

51. Kessler, N., Schuhmann, H., Morneweg, S., Linne, U. & Marahiel, M. A. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J. Biol. Chem.* **279**, 7413–7419 (2004).
52. 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. USA* **103**, 8511–8516 (2006).
53. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
54. Phan, J. et al. Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277**, 50564–50572 (2002).
55. Nazi, I., Koteva, K. P. & Wright, G. D. One-pot chemoenzymatic preparation of coenzyme A analogues. *Anal. Biochem.* **324**, 100–105 (2004).
56. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods. Enzymol.* **276**, 307–326 (1997).
57. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
59. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
60. Bulut, H. et al. Crystal structures of receptors involved in small molecule transport across membranes. *Eur. J. Cell Biol.* **91**, 318–325 (2012).
61. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
62. Bloudoff, K., Fage, C. D., Marahiel, M. A. & Schmeing, T. M. Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc. Natl Acad. Sci. USA* **114**, 95–100 (2017).
63. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78 (2007).
64. Schafer, A., Horn, H. & Ahlrichs, R. Fully optimized contracted Gaussian-basis sets for atoms Li to Kr. *J. Chem. Phys.* **97**, 2571–2577 (1992).
65. Scalmani, G. & Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **132**, 114110 (2010).
66. Maier, J. A. et al. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
67. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
68. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
69. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the Resp model. *J. Phys. Chem.* **97**, 10269–10280 (1993).
70. Besler, B. H., Merz, K. M. & Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **11**, 431–439 (1990).
71. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).
72. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald—an *N*·log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
73. Götz, A. W. et al. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555 (2012).
74. Reimer, J. M., Haque, A. S., Tarry, M. J. & Schmeing, T. M. Piecing together nonribosomal peptide synthesis. *Curr. Opin. Struct. Biol.* **49**, 104–113 (2018).
75. Reimer, J. M., Aloise, M. N., Powell, H. R. & Schmeing, T. M. Manipulation of an existing crystal form unexpectedly results in interwoven packing networks with pseudo-translational symmetry. *Acta Crystallogr. D* **72**, 1130–1136 (2016).
76. Chen, W.-H., Li, K., Guntaka, N. S. & Bruner, S. D. Interdomain and intermodule organization in epimerization domain containing nonribosomal peptide synthetases. *ACS Chem. Biol.* **11**, 2293–2303 (2016).
77. Liu, Y., Zheng, T. & Bruner, S. D. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem. Biol.* **18**, 1482–1488 (2011).
78. Gahloth, D. et al. Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis. *Nat. Chem. Biol.* **13**, 975–981 (2017).
79. Deshpande, S., Altermann, E., Sarojini, V., Lott, J. S. & Lee, T. V. Structural characterization of a PCP–R di-domain from an archaeal non-ribosomal peptide synthetase reveals novel inter-domain interactions. *J. Biol. Chem.* **296**, 100432 (2021).
80. Bozhuyuk, K. A. J. et al. Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661 (2019).
81. Nguyen, K. T. et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl Acad. Sci. USA* **103**, 17462–17467 (2006).
82. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317–347 (2016).
83. Bozhuyuk, K. A. J. et al. De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).
84. Bozhuyuk, K. A. J. et al. Evolution-inspired engineering of nonribosomal peptide synthetases. *Science* **383**, eadg4320 (2024).
85. Mansour, B. & Gauld, J. W. Computational insights into amide bond formation catalyzed by the condensation domain of nonribosomal peptide synthetases. *ACS Omega* **9**, 28556–28563 (2024).

Acknowledgements This research was funded by grants to T.M.S. from Canadian Institutes for Health Research (grant PJT-178084) and to K.N.H. from the US National Science Foundation (CHE-2153972). P.M. acknowledges the National Natural Science Foundation of China (no. 22103060). Infrastructure at the McGill University Centre de Recherche en Biologie Strucutrale, used in this research, is supported by Fonds de Recherche du Québec (Health Sector) Research Centres grant 288558. We thank Q. Zhou for efforts with quantum mechanics/molecular mechanics; M. Eivaskhani, J. Reimer, A. Mantri, N. Frota and the other members of the Schmeing laboratory, K. Auclair, A. Berghuis, T. Sprules, N. Rogerson, G. Challis and A. Guarné for advice and discussions; A. Mittermaier for advice on statistics; and CLS beamline staff for facilitating data collection. The X-ray diffraction data collection described in this paper was performed using beamline CMCF-ID at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation, the Natural Sciences and Engineering Research Council, the National Research Council, the Canadian Institutes of Health Research, the Government of Saskatchewan and the University of Saskatchewan. We thank C. Chalut and C. Guilhot for gift of *E. coli* BL21 (DE3) *entD*- cells. AEP1-encoding plasmid was provided by W. Bin.

Author contributions A.P. and T.M.S. designed, and A.P. performed, all chemistry, biochemistry and crystallography. K.M. performed CD experiments. K.N.H. and P.M. designed, and P.M. and Z.L. performed, computational simulations. T.M.S. supervised A.P.; K.N.H. supervised P.M.; and P.M. supervised Z.L. T.M.S., A.P. and K.N.H. wrote the manuscript with input from all of the authors.

Competing interests The authors declare no competing interests.

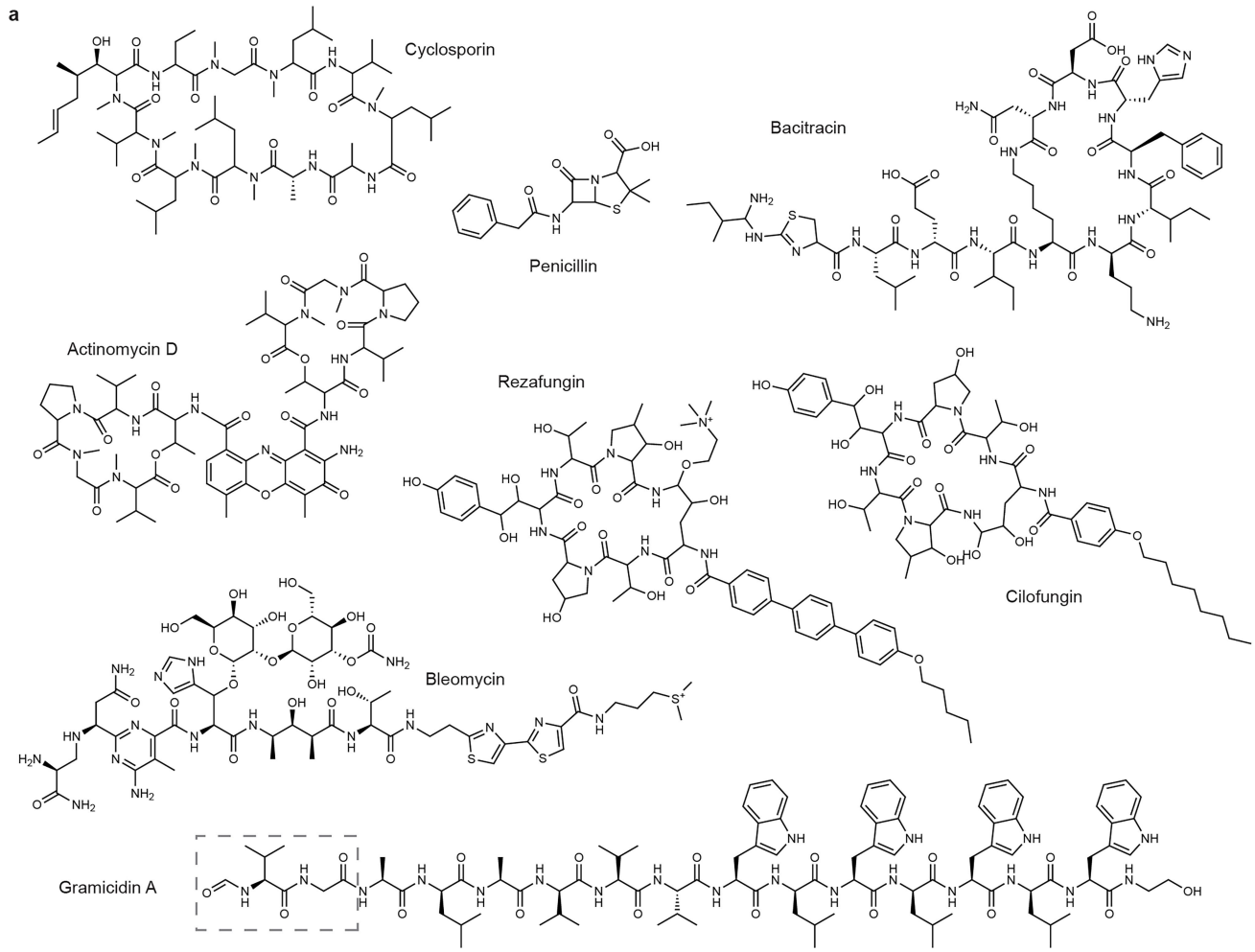
Additional information

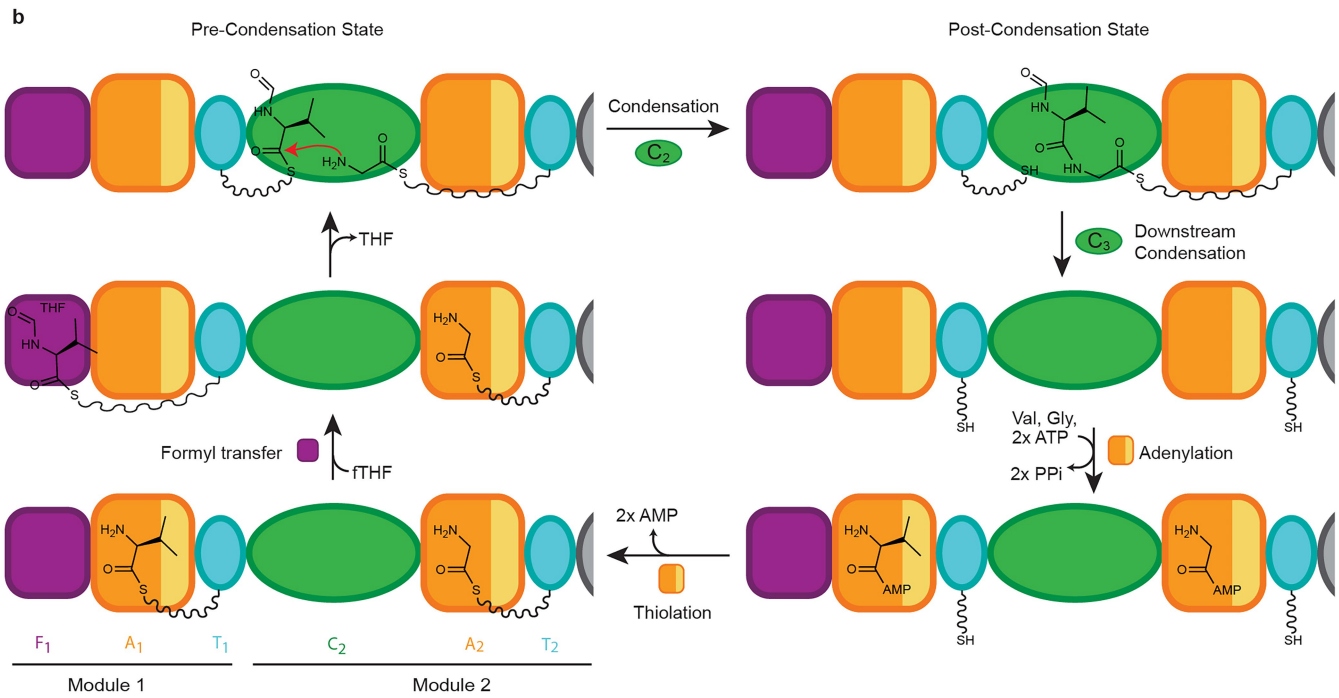
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08417-6.

Correspondence and requests for materials should be addressed to T. Martin Schmeing.

Peer review information *Nature* thanks Adam Balutowski, Andrew Gulick, Shiou-Chuan Tsai and Timothy Wencewicz for their contribution to the peer review of this work. Peer reviewer reports are available.

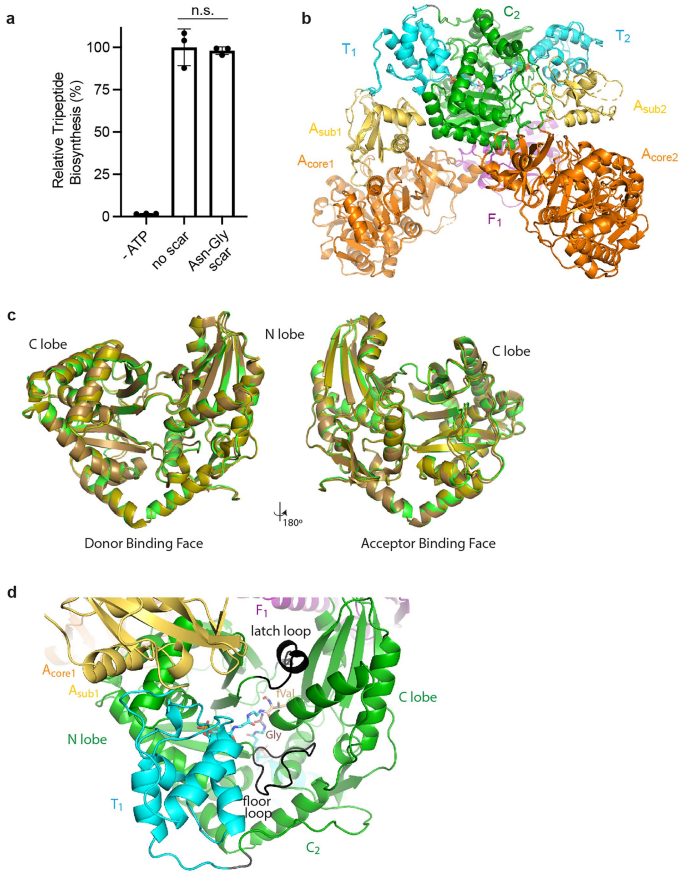
Reprints and permissions information is available at http://www.nature.com/reprints.



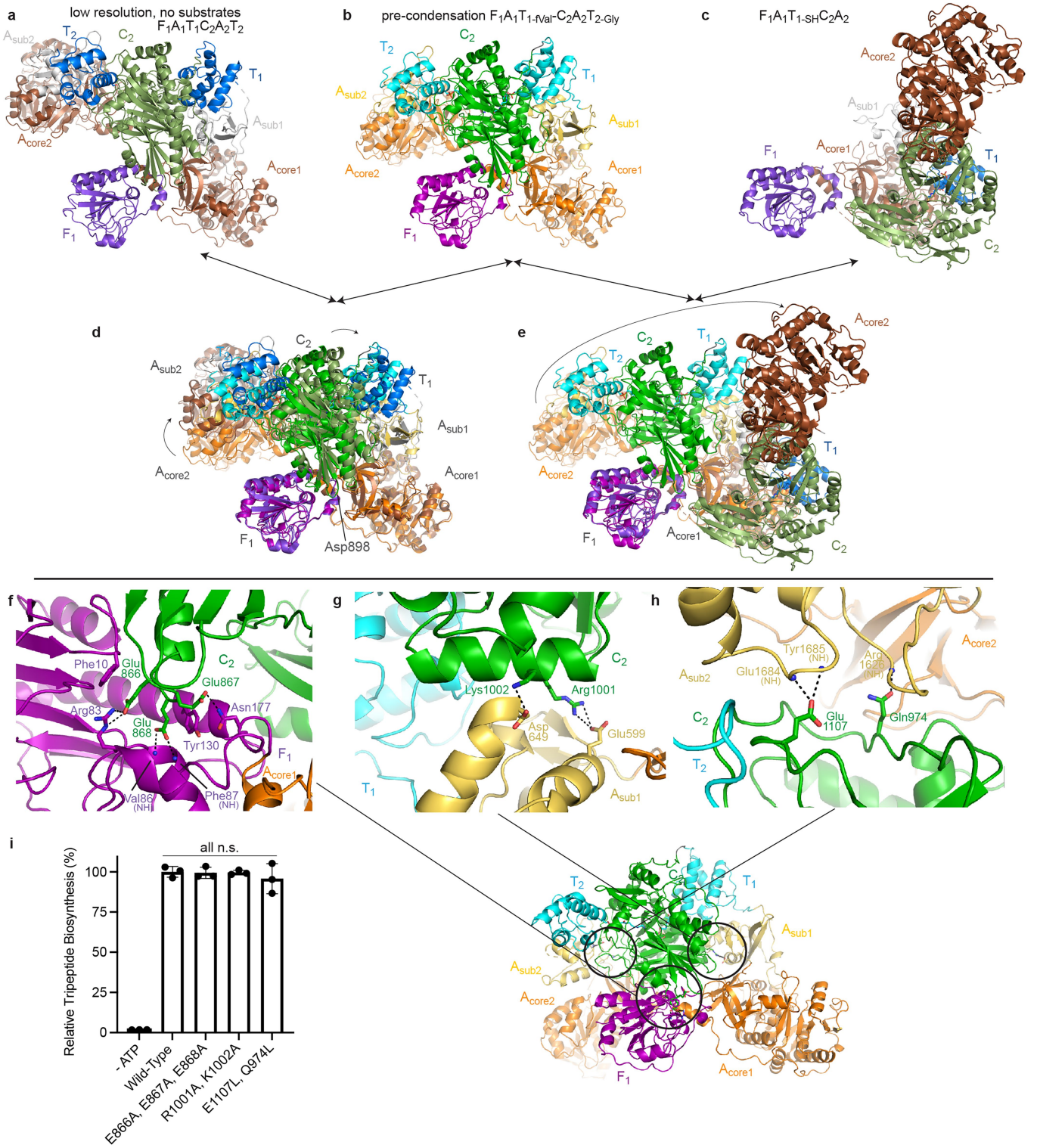


Extended Data Fig. 1 | See next page for caption.

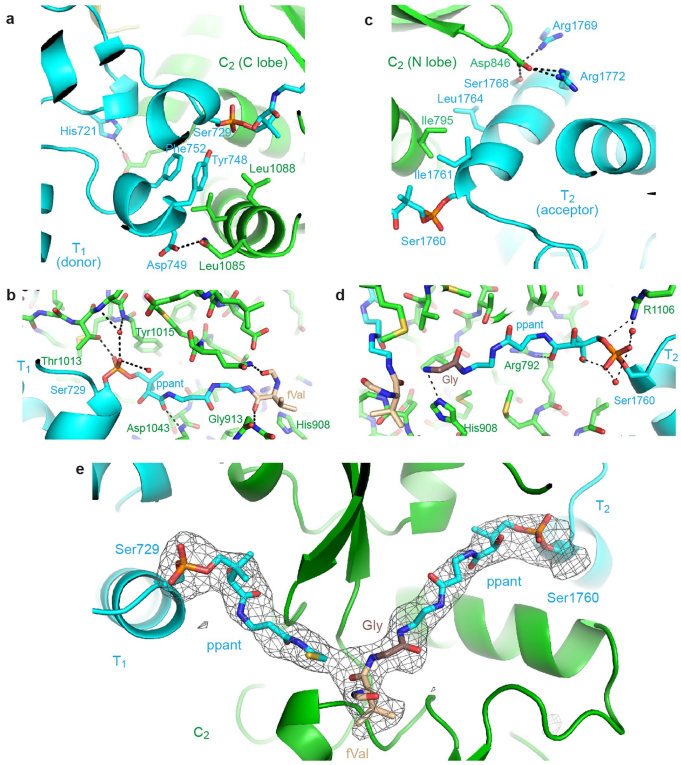
Extended Data Fig. 1 | Nonribosomal peptides and the NRPS synthetic cycle. a. Representative therapeutics fully made by NRPSs, or for which NRPSs play a central synthetic role. The portion of gramicidin A that LgrA assembles is indicated by the grey box. **b**. The synthetic cycle of LgrA. Structures presented here represent the pre-condensation state (top left) and the post-condensation state (top right). Note that the small molecular substrates (fTHF, ATP, amino acids) are shown binding directly before they are required in the synthetic cycle, but the binding (as well as the adenylation reaction) can occur as soon as the relevant domain is free.



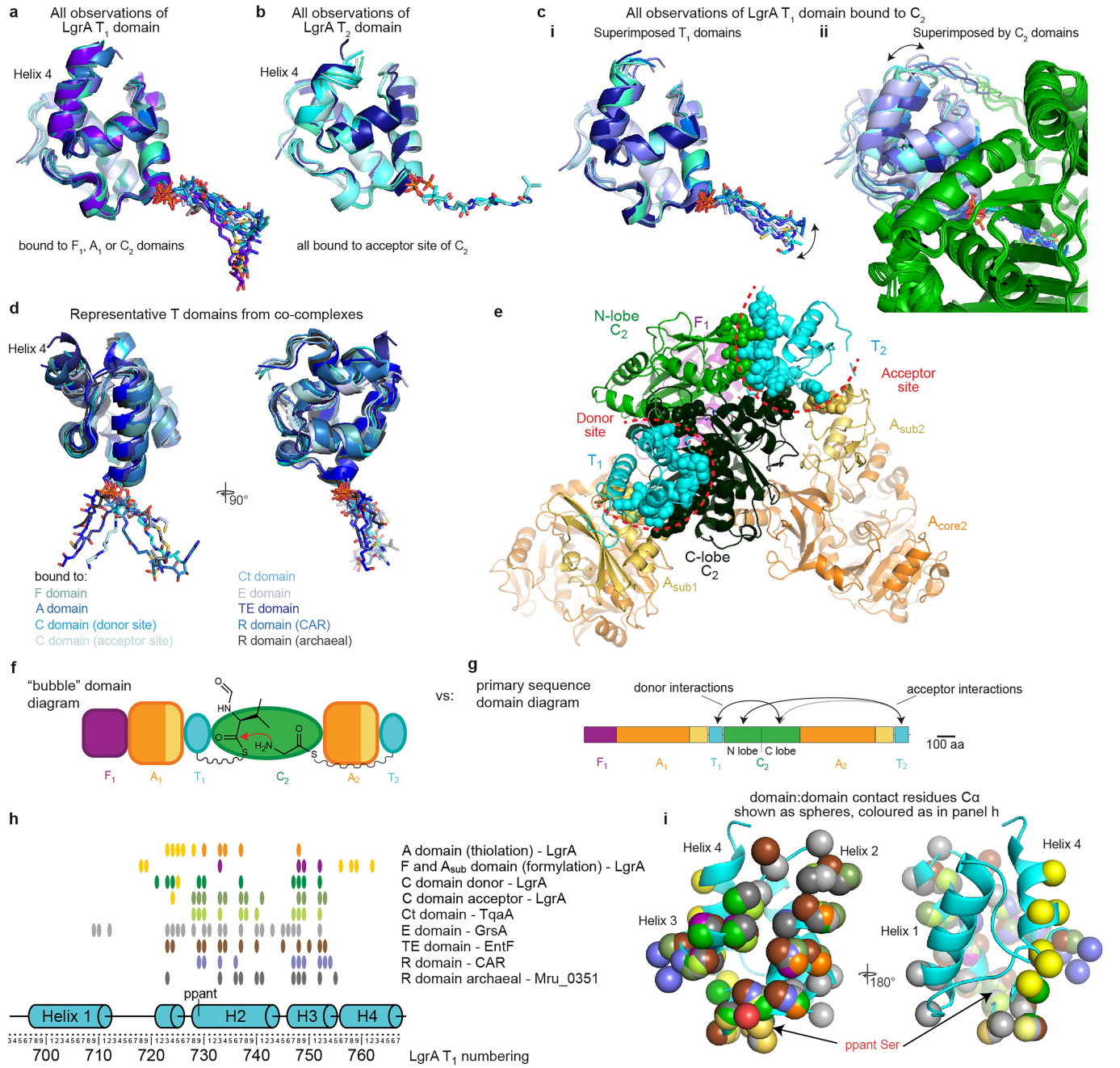
Extended Data Fig. 2 | Activity and structure of FAT–CAT. a. Tripeptide synthesis assays of $F_1A_1T_1C_2A_2T_2$–$C_{BmdB\text{-}M3}$ and $F_1A_1T_1$–$C_2A_2T_2$–$C_{BmdB\text{-}M3}$ shows that the scar residues left by AEP1 ligation does not influence biosynthesis activity. $C_{BmdB\text{-}M3}$ catalyses peptide bond formation of fVal-Gly-ppant-T_2 with free tryptamine, releasing formyl-valinyl-glycyl-tryptamine, quantified by LC-MS. Note that although AEP1 ligation leaves an Asn-Gly-Leu scar, by splitting the dimoldular LgrA before residue Leu771 in the T_1-C_2 linker, only 2 extra residues are added to the AEP1 ligated FAT–CAT. The sequence of the ligated FAT–CAT becomes: ...Val770-Asn770A-Gly770B-Leu771... The central value for the reactions represents the mean, while the standard deviation of the mean is represented by the error bars. Individual points of the triplicates (n = 3) are shown. Replicates presented in the same panel were performed with one preparation of protein each, but repeated preparations of the same proteins repeatedly show the same activity. Statistical significance between the wild-type and mutant variants of $F_1A_1T_1$–$C_2A_2T_2$–$C_{BmdB\text{-}M3}$ was determined by a two-sided Student's t-test. **b**. Comparison between pre-condensation $F_1A_1T_{1\text{-}fVal}$–$C_2A_2T_{2\text{-}Gly}$ and post-condensation $F_1A_1T_{1\text{-}SH}$–$C_2A_2T_{2\text{-}fVal\text{-}Gly}$ crystal structures. **c**. The C domains of $F_1A_1T_{1\text{-}fVal}$–$C_2A_2T_{2\text{-}Gly}$ (green), $F_1A_1T_{1\text{-}SH}$–C_2A_2[14] (light brown), and $F_1A_1T_{1\text{-}fVal}$–C_2[14] (olive) show the same lobe:lobe orientation. Relative motion of C domain lobes has been proposed as important for catalysis and predicated by normal mode analyses, and differential chemical footprinting of C domains observed in presence and absence of substrates[8,10,23,41]. However, different relative orientations of lobes has only been observed across different C domains. **d**. The latch loop (residues 1124–1138) and the floor loop (residues 1044–1057) of C_2 is ordered and interacting with the C lobe, arguing that these elements do not change conformation to act directly in catalysis[8].



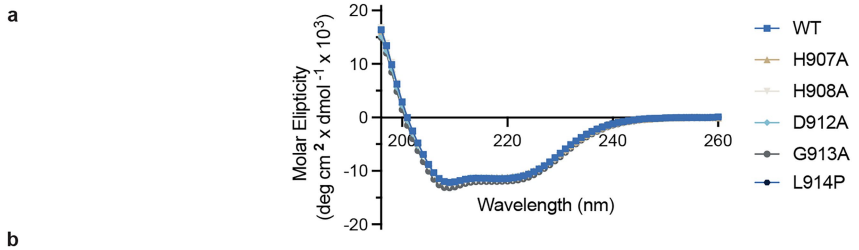
Extended Data Fig. 3 | Domain:domain interactions in dimodular LgrA. Top: Comparison of **a** the low resolution, substrate-free $F_1A_1T_1C_2A_2T_2$ structure (6MFZ[14]), **b** $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ and **c** $F_1A_1T_{1\text{-fVal}}C_2A_{core2}$ (6MFY[14]), shown individually (**a**-**c**) and pairwise (**d**,**e**). The structures are superimposed on F_1A_{core1}. Bottom: Domain:domain interactions in $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ between (**f**) F_1 and C_2, (**g**) A_{sub1} and C_2, and (**h**) C_2 and A_{sub2}. **i**. Tripeptide biosynthesis assay of C_2 mutants E866A-E867A-E868A, K1002A-R1001A and E1107A-N974A. Measurements were carried out over n = 3 individual reactions in one experiment. Replicates presented in the same panel were performed with one preparation of protein each, but repeated preparations of the same proteins repeatedly show analogous activity levels. Individual points of the triplicates are shown. Each bar represents the average of the three measurements and the error bar is indicative of the standard deviation in the three replicates. Statistical significance between the wild-type and linker mutant variants was assessed by a two-sided Student's t-test.

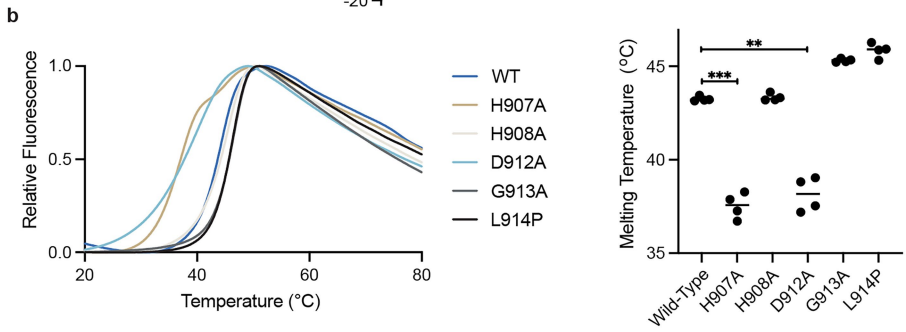


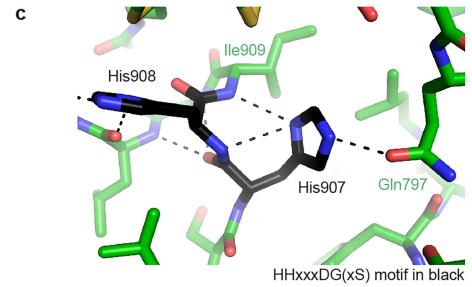
Extended Data Fig. 4 | C domain interactions with aminoacyl-T domains. **a**,**b**. Interactions of the LgrA C_2 domain with donor $_{fVal}$.T_1 domain. At the donor entrance of the C_2 tunnel, the ppant phosphate attached to T_1 Ser729 forms a hydrogen bond with Thr1013, a well-conserved in $^{L}C_L$ domain residue[62]. The ppant dimethyl moiety makes hydrophobic interactions with Tyr1015. This position shows strong conservation for aromatic residues in $^{L}C_L$ domains[62]. The ppant also hydrogen bonds with the backbone amine of Asp1043 from the floor loop. **c**,**d**. At the donor entrance of the C_2 tunnel, the ppant phosphate attached to T_2 Ser1760 forms hydrogen bonds with the highly conserved Arg1106[62]. Arg792 also forms hydrogen bonds with the backbone of the acceptor ppant. Notably, in structures lacking the acceptor T_2 domain[14], the side chains of Arg1106 and Arg792 extend into the tunnel, but are displaced in these full condensation complexes. Similarly, the arginine equivalent to Arg729 in fuscachelin A synthetase FscG-C_3 was seen to move when an amino acyl ppant analogue binds[4]. Likewise, the arginine equivalent to Arg1106 do not obstruct the tunnel when acceptor ppants bind at the active site in FscG-C_3[4] and holo-AB340325, but rather interact with the ppants. The subtle movements of these Arg residues suggest to us that this is a simple sidechain rearrangement upon acceptor binding. **e**. Polder map of the (acyl-)ppant moieties in $F_1A_1T_{1\text{-}SH}$–$C_2A_2T_{2\text{-}fVal\text{-}Gly}$. The map is displayed at 5 σ, and is not "carved" around acyl-ppant moieties.

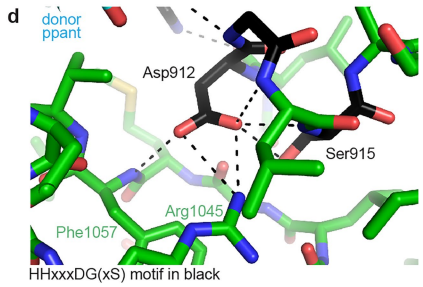


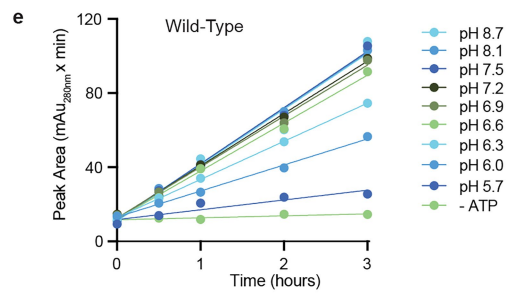
Extended Data Fig. 5 | T domain conformations and interactions. **a**. Superimposition of all observations of LgrA T_1, in thiolation[13,14], formylation[13] and condensation donation state[14]. The C terminus of helix 4 shows minor variability and there is relatively modest differences in ppant moiety positions. **b**. Superimposition of all observations of LgrA T_2, all in the condensation state. **c**. Superimposition of all observations of LgrA T_1:C_2 interactions[14]. (**i**) Comparing superimposition by (**i**) T_1 and by (**ii**) C_2 shows that small variations in C_2 binding by T_1 (mainly minor angle shifts of T_1 around the C_2 donor site interaction surface (highlighted in **ii**)), are compensated for by small changes in ppant conformations (highlighted in **i**), to allow fVal delivery to the active site. **c**. Superimposition of one example each of structurally characterized, cognate interaction co-complexes of T domains with each type of *bonefide* NRPS domain[42]: LgrA A_1[13], LgrA F_1[13], LgrA C_2, TqaA fungal terminal cyclization domain[12], GrsA E domain[76], EntF thioesterase domain[77], CAR reductase domain[78] and Mru_0351 archaeal reductase domain[79]. **e**. $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ with all residues within 4 Å of the partner domains shown as spheres. The donor binding site is exclusively made from residues from the C-lobe of C_2. The acceptor binding site is mainly made from residues from the N-lobe of C_2. One residue from each A_{sub} is close to its neighbouring T domain (yellow). **f**,**g**. The donor binding site being part of the C-lobe of C_2 and the acceptor binding site being mainly part of the N-lobe of C_2 means that useful-but-very-simplified domain bubble diagrams (**f**) can be misleading: They could be taken to imply that upstream and downstream T domains bind at the portions of the C domains that are proximal to them in primary sequence. However, the interactions are mainly with the portions of the C domain which are more distal in primary sequence (**g**). This means any cut point in module-swap bioengineering between T_{n-1} and T_n would disturb all native T_{n-1}: C_n interactions and most native C_n: T_n interactions. **h**,**i**. For structures in **d**, the T domain residues within 4 Å of their partner domain residues, mapped onto (**h**) the primary sequence and (**i**) the structure of LgrA T_1 as Cα spheres. T domains use an extensively overlapping surface for all their known interactions with partner domains.

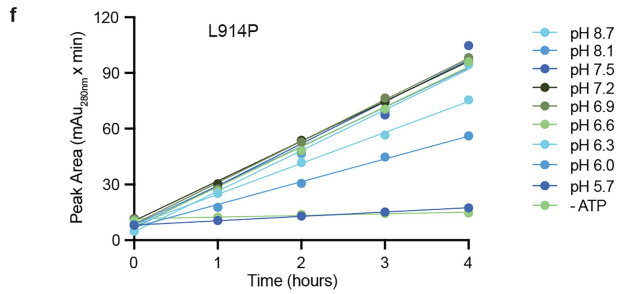


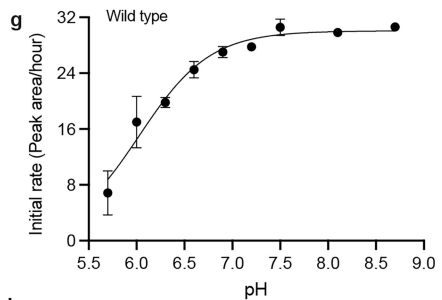






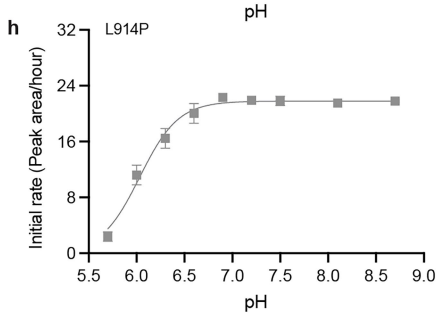






pH	5.7	6	6.3	6.6	6.9	7.2	7.5	8.1
6.0	0.0232 (*)							
6.3	0.0153 (*)	0.3177 (ns)						
6.6	0.0052 (**)	0.0610 (ns)	0.0073 (**)					
6.9	0.0053 (**)	0.0370 (*)	0.0004 (***)	0.0438 (*)				
7.2	0.0065 (**)	0.0354 (*)	0.0002 (***)	0.0279 (*)	0.2604 (ns)			
7.5	0.0024 (**)	0.0165 (*)	0.0006 (***)	0.0032 (**)	0.0167 (*)	0.0413 (*)		
8.1	0.0060 (**)	0.0262 (*)	0.0009 (***)	0.0139 (*)	0.0228 (*)	0.0076 (**)	0.3969 (ns)	
8.7	0.0050 (**)	0.0222 (*)	0.0001 (****)	0.0060 (**)	0.0062 (**)	0.0015 (**)	0.9408 (ns)	0.0755 (ns)

Significance between measurements at different pH points, wild-type

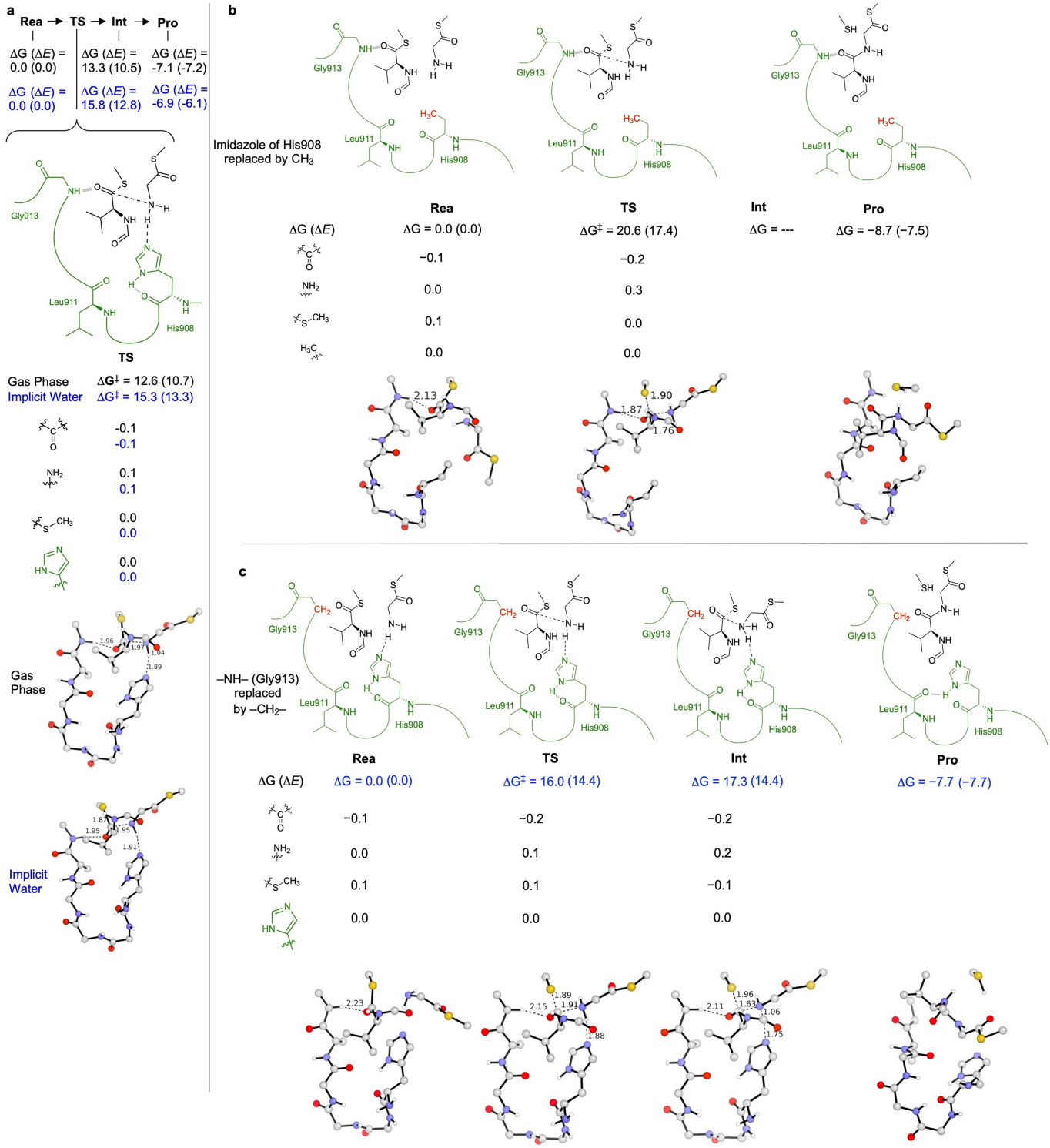


pH	5.7	6	6.3	6.6	6.9	7.2	7.5	8.1
6.0	0.0027 (**)							
6.3	0.0008 (***)	0.0106 (*)						
6.6	0.0004 (***)	0.0016 (**)	0.0392 (*)					
6.9	0.000011 (****)	0.0041 (**)	0.0166 (*)	0.1018 (ns)				
7.2	0.000016 (****)	0.0002 (***)	0.0158 (*)	0.1412 (ns)	0.2373 (ns)			
7.5	0.000002 (****)	0.0017 (**)	0.0117 (*)	0.1460 (ns	0.3254 (ns)	0.8993 (ns)		
8.1	0.000028 (****)	0.0041 (**)	0.0208 (*)	0.2051 (ns)	0.0402 (*)	0.3281 (ns)	0.5304 (ns)	
8.7	0.000012 (****)	0.0045 (**)	0.0202 (*)	0.1583 (ns)	0.0841 (ns)	0.7870 (ns)	0.9538 (ns)	0.3501 (ns)

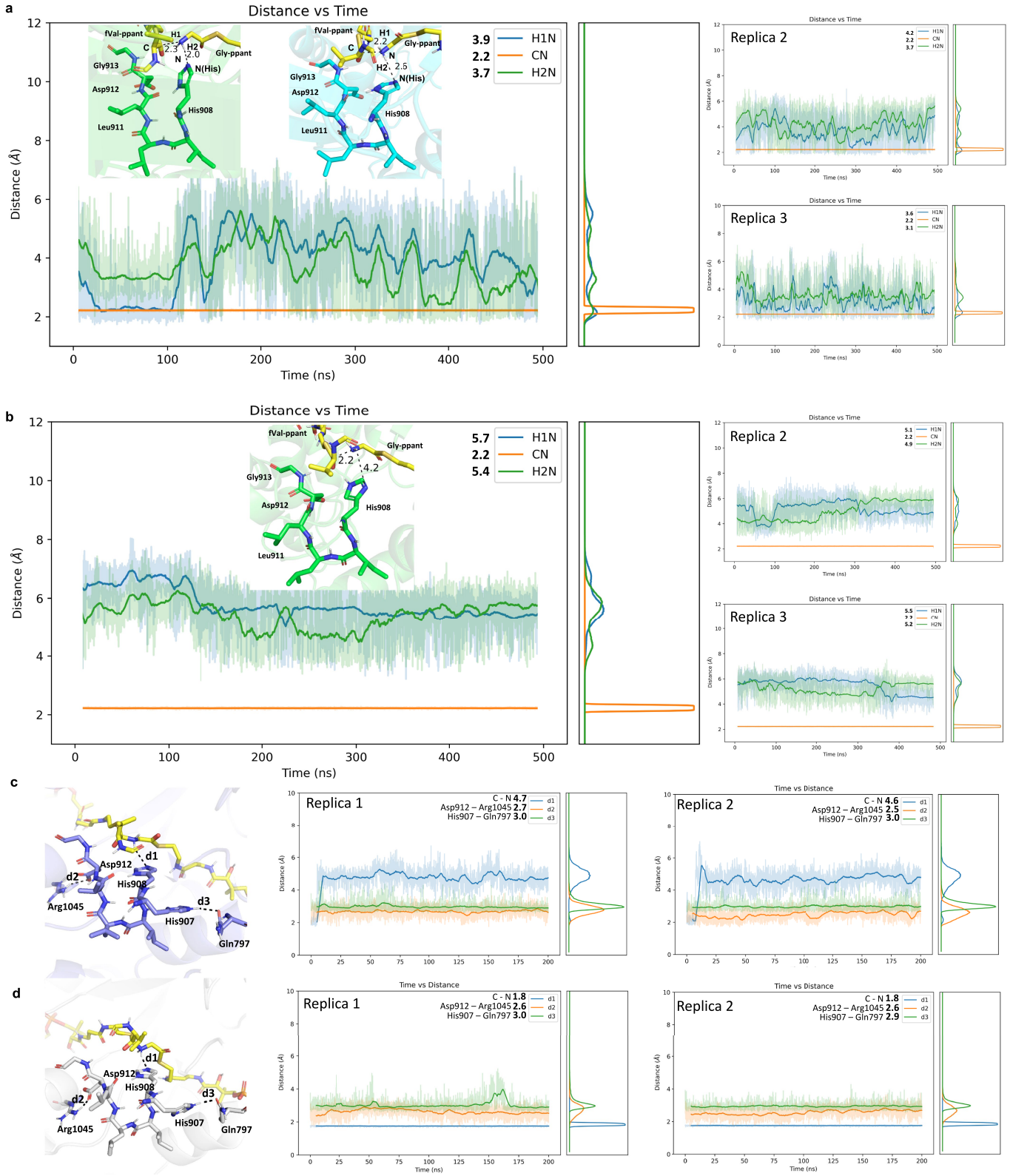
Significance between measurements at different pH points, L914P

Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | C domain mutants and pH profiles. a. CD of wild-type and mutant $C_2A_2T_2$ showing active site mutations not to affect global folding. **b**. DSF in quadruplicate (n = 4) of wild-type and mutant $C_2A_2T_2$ shows decreased Tm for H907A and D912A compared to wild-type and other mutants. Tm values (inflexion point of DSF) are plotted in °C: Wild-type: 43.25 ± 0.14; H907A: 37.54 ± 0.69; H908A: 43.34 ± 0.20; D912A: 38.15 ± 0.91; G913A: 45.32 ± 0.09; L914P: 45.85 ± 0.40. Statistical significance between wild-type and variants were calculated using a two-sided Student's t-test (wild-type versus H907A ($p = 0.0003$; ***), H908A ($p = 0.5067$; n.s), D912A ($p = 0.0013$; **), G913A ($p = 0.000002$; ****), L914P ($p = 0.0002$;***)). Central values represent the mean, and error bars represent the standard deviation of the mean. **c**. His907 points away from His908 and hydrogen bonds with its own backbone amine, the Ile909 backbone amine and the Gln797 sidechain. **d**. Asp912 hydrogen bonds with the backbone amines of Leu914, Ser915 and Phe1057, and the side chains of Ser915 and Arg1045. **e**,**f**. Initial tripeptide synthesis rates at pHs 5.7 to 8.7 for $F_1A_1T_1{-}C_2A_2T_2{-}C_{BmdB\text{-}M3}$ and the L914P mutant, where condensation by C_2 is rate limiting. **g**,**h**. Rates plotted as peak area (mAU_{280nm}·min) per hour. Each data point represents the average of the initial rate of n = 3 replicates collected over three independent experiments. To collect all the data two preparations of protein were necessary, for each variant (wild type and L914P). The central value plotted represents the mean initial rate and the error bars the standard deviation in the initial rates. The error is taken to be the standard deviation between rates determined by individual time course experiments, represented by error bars. Trend lines are shown as a visual aid only. Statistical significance between initial rates at all pH points were determined using two-sided Student's t-tests.

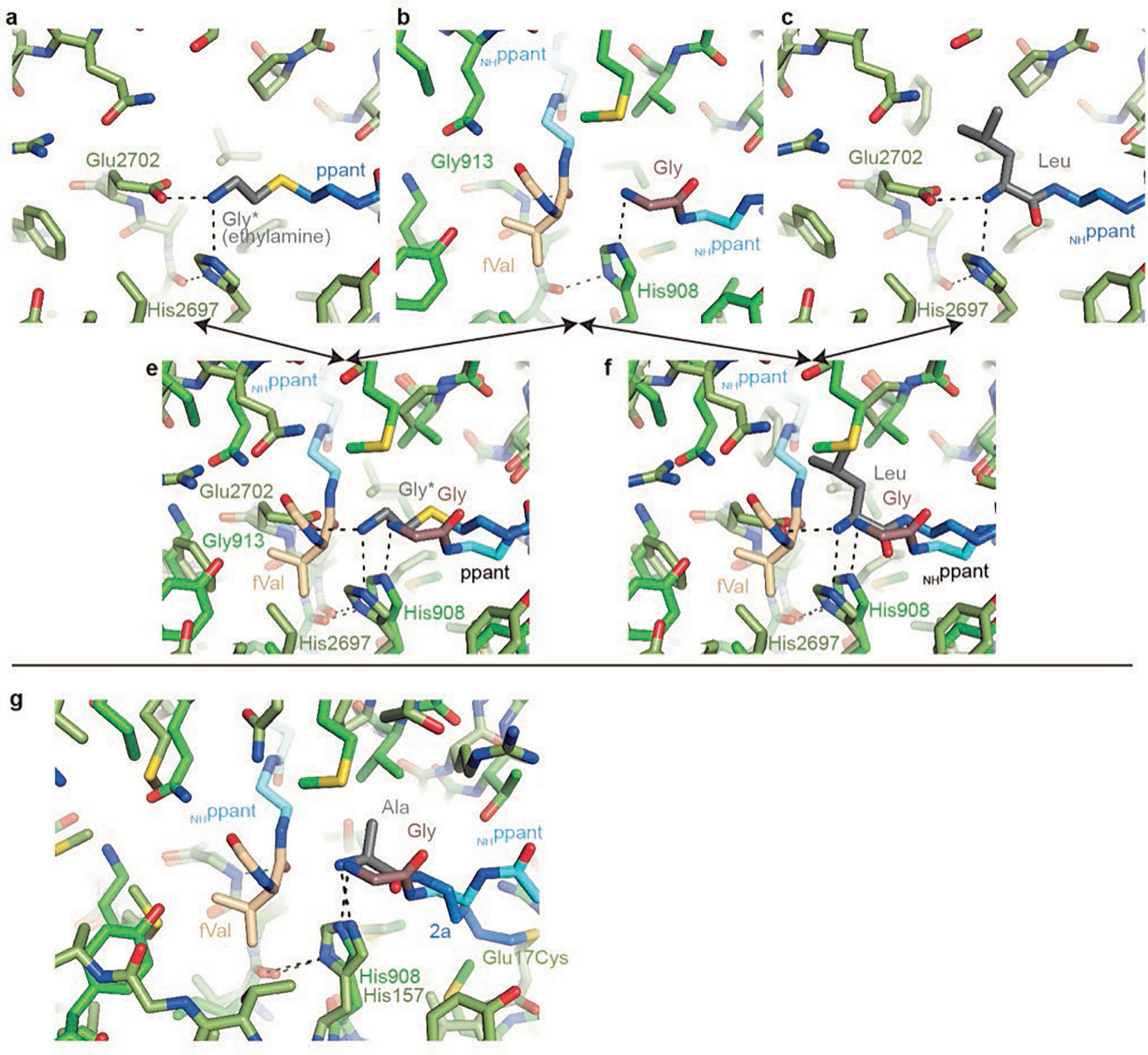


Extended Data Fig. 7 | Quantum mechanical calculations for different conditions and mutants. The theozyme Gibbs free energies and electronic energies of each species are shown for diagram in different conditions and mutants; Hirshfeld Charges of each fragment along the reaction pathway are also shown. 3D structures are shown below. For clarity, nonpolar H atoms are omitted except on Ns and on $-CH_2$ group. Curved lines represent groups omitted in the computations. The backbone atoms of the protein chain were constrained in the theozyme calculations. **a**. The calculated information for no solvent (gas phase) and in CPCM water solvent (implicit solvent), only showing **TS**, which has free energy lower than that of **Int** (Fig. 4a). The other species in these conditions are show in Fig. 4a and not redrawn here. **b**. His908 imidazole replaced by $-CH_3$; **c**. –NH– of Gly913 is replaced by $-CH_2-$.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Molecular dynamics simulations. a,**b**. Distances of H908 δ or ε nitrogen (N) to the hydrogens (H1, H2) on the α-amine nucleophile, during simulations. Large graphs plot distances versus time, small graphs on right show populations of distances over the full simulation, and average distances the simulations are listed left of the caption. In **a**, His908 δ nitrogen is protonated, in **b**, the ε nitrogen is protonated. The C···N distance that eventually leads to an amide bond is constrained to 2.2 Å. The most populated structure in the first 100 ns (green) and in the full 500 ns (cyan) are shown in **a** and the most populated structure in the 500 ns simulation is shown in **b**. In this MD, the substrates and separate from the catalytic residues. This separation is not unusual for enzyme with spacious active sites. A recent study featuring MD of substrates manually placed into tightly packed areas of a C domain showed low root mean square deviation for 30 ns[85], but the manual substrate placement is at odds with known binding sites and T domain substrate delivery, which may compromise analysis. **c**,**d**. The most populated 3D structures (left) and bond distances between the reactants and key residues (right) during 200 ns simulations, starting from **Int1** of the theozyme calculations. Simulation **c** has the H(Acceptor)···N(H908) distance unconstrained during the simulation, while in **d**, the distance is constrained the 2.2 Å distance seen in the theozyme **Int**. In all these simulations, His907 and Asp912 stay pointed away from the reactive atoms and His908. Note the brief increase in d3 in panel **d** replica 1 is from a transient movement of Gln787 away from the active site, not His907 towards His908.



Extended Data Fig. 9 | Comparisons with other C domain structures. This figure shows comparisons with some of the most informative ligand-bound structures of other C domains. **a-f**. Comparison of $F_1A_1T_{1\text{-}fVal}–C_2A_2T_{2\text{-}Gly}$ with acceptor-bound fuscachelin A synthetase C_4 complexes. A thioether analogue of Gly-ppant (ethylamine-ppant)[4] (**a**), and Leu-$_{NH}$ppant[30] (**c**) bind FscG-C_3 similarly to each other and to how LgrA-C_2 binds Gly-$_{NH}$ppant. However, in FscG-C_3, the nucleophile interacts with Glu2702, reaching from the donor side of the active site. This interaction could not be maintained in the presence of a donor substrate as its position overlaps with that of fVal-$_{NH}$ppant in the overlays (**e**,**f**) and the end of the donor tunnel in general. Interestingly, this Glu2702 in FscG-C_3 one of less common cases (along with AmbB[11]) where the final position of the HHxxxDG motif is not a Gly but a side-chain bearing amino acid. Gly913 is 83% conserved in sequences of as $^{L}C_L$ or $^{D}C_L$ domains and 79% conserved in all C domains. Small differences in the approach of the donor ppants compared to that seen in LgrA would be needed to avoid clashing with the sidechain. **g**. A comparison with the a small molecule acceptor fused to the first C domain of calcium dependant antibiotic synthetase[9]. Note that this amide-containing acceptor analogue shown in **g**, and the thioether analogue in **a** are both shown to be competent for reaction[4,9], whereas the amide containing acceptor analogue in **c** is reported to be non-reaction competent[30]. The positions of amide and thioester analogues studied in reference[4] are so similar that both provide equivalent structural insight.

Extended Data Table 1 | Crystallography data and refinement statistics

	LgrA pre-condensation (PDB 9BE3)	LgrA post-condensation (PDB 9BE4)
Data collection		
Space group	$P2_12_12$	$P2_12_12$
Cell dimensions		
a, *b*, *c* (Å)	184.8, 427.9, 77.3	184.4, 426.6, 77.3
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	50.00-2.84 (2.89-2.84)	50.00-3.00 (3.05-3.00)
R_{sym}	0.183 (1.520)	0.254 (1.448)
$I / \sigma I$	9.3 (0.85)	7.8 (1.0)
Completeness (%)	99.9 (99.5)	99.8 (98.1)
Redundancy	7.2 (5.2)	10.1 (7.4)
$CC_{1/2}$	0.990 (0.336)	0.985 (0.348)
Refinement		
Resolution (Å)	48.74-2.90	49.87-3.00
No. reflections	135580	120084
R_{work} / R_{free}	0.2157 / 0.2583	0.2180 / 0.2575
No. atoms	28355	28158
Protein	27747	27713
Ligand/ion	107	95
Water	501	350
B-factors		
Protein	61.23	61.57
Ligand/ion	65.88	67.00
Water	51.25	46.20
R.m.s. deviations		
Bond lengths (Å)	0.0032	0.0028
Bond angles (°)	0.59	0.56

Statistics for the diffraction data, processing and refinement of the LgrA pre-condensation (PDB accession code 9BE3) and post-condensation (9BE4) structures. Values in the parentheses are for the highest-resolution shell. One crystal was used for each structure.

nature portfolio

Corresponding author(s): Thomas Martin Schmeing

Last updated by author(s): Nov 7, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	X-ray data: Synchrotron – Canadian Light Source (CLS) CMCF-ID (08ID) beamline. HPLC – Agilent 1260 series. Mass spectrometry – Mass spectra were collected using a Bruker amazon speed EDT ion trap mass spectrometer.
Data analysis	Crystallography: HKL2000 v721, Phenix v1.20.1-4487 (with Phaser v2.9.0), Coot v0.9.8.92. PyMol v2.4.1. Circular dichroism: Chirascan ProViewer software v.4.7.0.194, OLIS SpectralWorks v4.3. Theozyme calculations: Gaussian 16 program package. Thermodynamic quantities calculated with GoodVibes software 3.1.1; 3D rendering with PyMol v2.7. Molecular dynamics: Amber 16 package and the AmberTools 16 codes. Differential scanning fluorimetry: StepOne software v2.3 and Protein Thermal Shift software 1.4. Kinetic data and statistical analyses: GraphPad Prism version 10.2.3 and Microsoft Excel 16.90.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

X-ray crystal structures and associated diffraction data for this study are deposited under accession codes 9BE3 and 9BE4 in the Research Collaboratory for Structural Bioinformatics Protein Data Bank. Source data for all figures in the main text, Extended Data and Supplementary Information have been supplied with the paper or as Supplementary Information source data files.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All measurements were carried out in triplicate, as three independent experimental measurements. Triplicate measurements were chosen to allow for statistical calculations: Averaging, standard deviation and statistical significance.
Data exclusions	No data has been excluded from the analysis.
Replication	Experiments presented in this study were carried out in triplicate to allow calculation of mean and standard deviation. All replicates were successful.
Randomization	We did not perform randomization as we were not working with population samples. Instead our in vitro experiments relied on known dependent variables (the mutations on the protein or pH of the reaction) that cannot be subjects of randomization.
Blinding	Blinding was not relevant in this study and not carried out for the rational described in 'Randomization'.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A