This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in Science 366, 2019-11-08, doi: 10.1126/science.aaw4388.

Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility

Janice M. Reimer[a], Maximilian Eivaskhani[a], Ingrid Harb[a], Alba Guarné[a], Martin Weigt[b] and T. Martin Schmeing[a]

[a]Department of Biochemistry, McGill University, Montréal, QC, Canada, H3G 0B1

[b]Sorbonne Université, CNRS, Institut de Biologie Paris-Seine, Laboratory of Computational and Quantitative Biology, F-75005, Paris, France.

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

Abstract

Nonribosomal peptide synthetases (NRPSs) are biosynthetic enzymes that synthesize natural product therapeutics using a modular synthetic logic, whereby each module adds one aminoacyl substrate to the nascent peptide. We have determined five crystal structures of constructs of linear gramicidin synthetase subunit A, including a structure of the full core dimodule in a conformation organized for the condensation reaction and a high resolution structure showing intermodular peptidyl substrate delivery. The structures reveal massive differences in the relative positions of adjacent modules which are not strictly coupled to the catalytic cycle, and which are consistent with small angle X-ray scattering (SAXS) data. The structures and covariation analysis of homologs allowed us to create mutants which improve the yield of a peptide from a module-swapped dimodular NRPS.

Introduction

Nonribosomal peptide synthetases (NRPSs) are intricate macromolecular machines that make small molecule products with very high chemical diversity and activity (*1*). NRPS compounds have found widespread clinical use as antitumors, antibiotics, antifungals and immunosuppressants, and are on lists of United Nations-designated essential medicines (*2*) and top-selling pharmaceuticals (*3*).

Nonribosomal peptide synthesis proceeds with thiotemplated, modular assembly-line logic where each multi-domain module adds one amino acid substrate to the growing peptide (Fig. 1) (*4*). A module's adenylation (A) domain selects and activates the amino acid, then covalently attaches it as a thioester to the thiolation (T) domain's phosphopantetheine (ppant) arm. The condensation (C) domain catalyzes peptide bond formation between that aminoacyl-T domain and the donor peptidyl-T domain from the upstream module (*5, 6*). The newly elongated peptidyl-T domain is then the donor substrate for condensation in the downstream module, passing off and further elongating the peptide. Many modules also have tailoring domains integrated within them, which co-synthetically modify the nonribosomal peptide, such as the tailoring formylation (F) domain (*7*) found in the initiation module of the NRPS studied here, linear gramicidin synthetase (Fig. 1).

An excellent structural understanding of the synthetic cycle of isolated modules has been gained from structures of domains, didomains, and individual modules (reviewed in (*4, 8)*). However, modules typically function within the context of the full NRPS. They are physically attached to their neighbors by flexible peptide linkers or through small docking domains (*9*). Adjacent modules must functionally coordinate at least once during the synthetic cycle, when the C domain catalyzes peptide bond formation between aminoacyl and peptidyl moieties attached to T domains of adjacent modules. Little else is known about how modules work with each other in the context of the larger NRPS. There are only two previous high-resolution structures that contain domains from adjacent modules: the T_5C_6 didomain of tyrocidine synthetase is in an unproductive conformation (*10*), while $A_1T_1C_2$ of bacillibactin synthetase (*11*) showed that the sole observed intermodule contact must break in the

course of peptide synthesis. The only 3D data for a multimodular NRPS are our 26-29 Å negative stain electron microscopy reconstructions of two modules of bacillibactin synthetase (CATCAT) which showed heterogeneity in the module:module conformation (*11*). Hypothetical models of multimodular NRPSs can be constructed by consecutively overlapping multi-domain structures from different synthetases, and the models often take the form of rigid superhelices (*4, 12*), but there is no evidence that any of these conformations occur *in vivo*. More data are needed to understand NRPS architecture, organization and intramodular function during the synthetic cycle of an NRPS, and to facilitate their use to make new-to-nature compounds.

Results

LgrA crystallography

Linear gramicidin synthetase is a 16-module, 4-protein NRPS which makes the clinically used eponymous antibiotic (Fig. 1) (*13*). The antibiotic acts by forming dimeric β-helical pores in Gram-positive bacterial membranes, which kills the bacteria by allowing free passage of monovalent cations across the membrane (*14*). To gain insight into outstanding trans- and super-modular questions in NRPS function, we undertook hundreds of thousand crystallography screening trials with constructs of linear gramicidin synthetase subunit A (LgrA, with domains FATCATE^{0}; Fig. 1) complexed with substrates, substrate analogues and dead-end inhibitors. This yielded five structures: two structures of the four-domain construct FATC in peptide donation conformation ($FAT_{fVal}C$ and $FAT_{fVal}C^{*}$; Fig. 2A, fig. S1A), one of FATCA in peptide donation conformation (FATCA; Fig. 2B), one of FATCA in two thiolation conformations ($FAT_{Vad}CA$; Fig. 2C, D), and one of the full dimodule FATCAT in overall condensation conformation (FATCAT; Fig. 2E) (figs. S1-S2, Table S1). In every structure, each domain assumes its canonical forms (*4, 8*): the F domain bears the formyltransferase catalytic domain fold, the C domain is a V-shaped pseudodimer of chloramphenicol acetyltransferase-like lobes with a tunnel to its active site, each A domain has its major portion which include the amino acid binding site

(A_{core}) and its mobile C-terminal subdomain (A_{sub}), and the PCP domains are small 4-helix bundles with prosthetic ppant arms. As examined below, together these structures of LgrA demonstrate three general features of NRPS architecture: 1. The didomain structural unit (FA_{core} or CA_{core}) of an NRPS module largely maintains its overall conformation (*15-19*). 2. The small domains (T and A_{sub}) move according to catalytic state (*15-20*). 3. Observed in detail here, the relative orientations of adjacent modules in an NRPS can vary markedly.

Conformations within modules

The main structural units of the modules are the FA_{core} or CA_{core} didomains. The current structures include 11 crystallographically independent FA_{core} or CA_{core} didomains, more than doubling the number available (*15-19*) (Fig. 2, 3A). These show the didomains in each module as "catalytic platforms" (*15*) that present the binding site for each module's T domains (the F_1 and A_1 active sites for T_1; the C_2 acceptor site and A_2 active site for T_2) on the same face to facilitate substrate delivery. The didomains are fairly rigid, as the F:A or C:A configurations shift by only ~1-12° degrees, propagating to ~10 Å (Fig. 3, figs. S3-S4). In FATCA, there are few crystal contacts at the distal end of A_2, and variation in the C:A_{core} orientation from unit cell to unit cell is evident from progressively increasing B-factors and weaker electron density at the distal end of A_2. However, there is substantially more variation in C:A_{core} conformations between different NRPSs than in a single NRPS: A_{2core} superimposition with modules of enterobactin, AB3403 and surfactin synthetases places some equivalent C domain residues >20 Å apart, because of variations of the C:A interface and "openness" (*21*) of the V shape of the C domain (figs. S4-S5).

The positions of the A_{sub} and T domains do vary depending on catalytic state (Fig. 3A) (*15-20*). As further explored below, T_1 is bound at the donor site of C_2 in four structures, and in three of these ($FAT_{fVal}C$, $FAT_{fVal}C^*$ and FATCAT), A_{1sub} is bound to A_{1core} in adenylation conformation (Fig. 2A, B, F). Simultaneously positioning of T_1 for condensation and A_1 for adenylation reiterates that NRPSs can

start a second synthetic cycle before finishing the first (Fig. 1) (*18*)). In both $FAT_{Vad}CA$ molecules, T_1 and A_{1sub} are bound to A_{1core} in thiolation conformation (Fig. 1, 2D,E) (*17, 22, 23*). This means that $FAT_{fVal}C$ and $FAT_{Vad}CA$ represent consecutive steps in synthesis (Fig. 1). To move between catalytic conformations observed here, A_{sub} rotates up to ~151º and translates up to ~17 Å, and T_1 rotates up to ~153º and translates up to ~47 Å (fig. S5) (*22, 24*). These transitions are as large as those A_{sub} and T_1 require to move between their positions in the rest of the synthetic cycle, for example to and from formylation conformation (*17*).

The condensation and substrate donation conformation

Condensation is the central chemical event of peptide synthesis. It requires that donor T domain (here T_1) and acceptor T domain (here T_2) bind simultaneously to the C domain (Fig. 4). The structure of FATCAT features the full condensation state, with both T_1 and T_2 bound to C_2, and represents a detailed 3D view of a multi-modular NRPS (Fig. 4A). The resolution of this structure is 6 Å, but the high resolution structures of F_1, A_1, T_1, C_2, A_{2core} and homology models of the ~100 residue A_{2sub} and ~90 residue T_2, enabled the building of a high-quality structure for the full, 1800 residue dimodule (fig. S1). T_2 occupies the acceptor binding site on C_2, located near helices α1 and α10 (*15, 18*) (Fig. 4B) and positions the phosphate of its ppant arm at the entrance of the C domain active site tunnel (fig. S2H). This T_2 position agrees with our direct coupling analysis (*25*) (DCA) and that in AB3404 (*18*), but is rotated by ~55° from that in surfactin synthetase (*15*) (fig. S7A-C, Table S2).

Four structures ($FAT_{fVal}C$, $FAT_{fVal}C^*$, FATCA, FATCAT) show T_1 binding to the donor site of C_2 (Fig. 4B,C). This canonical T_n:C_{n+1} interaction is the functional link between modules 1 and 2, which allows the nascent peptide to be elongated and passed downstream in the condensation reaction. The donor site is a shallow depression between helices α4 and α9 on the opposite side of the C domain tunnel from the acceptor site (Fig. 4B). Each LgrA donor structure has slightly different residue-level

T_1:C_2 contacts, all dominated by van der Waals interactions, which shift distal T_1 residues up to ~3.5 Å (Fig 4B, fig. S7D,E). DCA of the T_n:C_{n+1} interaction showed strong co-evolution signal between the areas of T_1 and C_2 we observe in direct contact (Fig. 4C, Table S3). We established a multiple-turnover peptide synthesis assay by fusing FATCAT to the terminal C (C_T) domain of bacillamide synthetase, which catalyzes peptide release by condensation with free tryptamine (Tpm) (*26, 27*). "Wild type" FATCAT-C_T produces fVal-Gly-Tpm tripeptide (Fig. 4D). We then used DCA, whose capacity to predict mutational effects in proteins has recently been established (*28*), to guide mutational analysis of the T_1:C_2 interface. Of four mutations in C_2 predicted to be deleterious for the T_1:C_2 interaction but not for C_2 folding, three showed moderate but significant decrease in tripeptide production (Fig. 4E, fig. S7J-M). The observed binding is thus likely a faithful representation of an important T_1:C_2 interaction.

T domains have previously been observed bound to sites analogous to the donor site in specialized C domain homologues: E domains, found downstream of T domains in some modules, catalyze chirality inversion in the peptide intermediate (*29*). Fungal NRPSs often end with a terminal condensation-like (C_t) domain, which catalyzes peptide release by macrocyclization with an internal nucleophile in the peptide intermediate (*30*). Both E and C_t domains have evolutionarily diverged from canonical C domains (*5, 31*), and each has one only T domain binding site, which is analogous to the donor site. The structure of TqaA didomain T_3C_t (*30*) shows contacts clustered on the α9 side of the donor site depression, similar to our T_1:C_2 interaction, with T_3 shifted by maximally ~7 Å (fig. S7G). However, the position of T_1 in the structure of the GrsA didomain T_1E_1 (*29*) is quite different, rotated by ~37° and translated ~13 Å to the α4 side of the donor site (fig. S7F). Correspondingly, DCA between T_n and E_n domains (within $C_nA_nT_nE_n$ modules) shows signal clustered on that α4 side (fig. S7I, Table S4). T domains thus bind C and E domains in distinct ways, with our structures and TqaA TC_t representing T_n:C_{n+1} binding, and the GrsA didomain representing T_n:E_n binding.

T domain binding to the donor site places the ppant arm into that side of the active site tunnel. The tunnel leads to the C domain's conserved HHxxxDG catalytic motif, with the second histidine (His908) most important for activity (*32-34*). $FAT_{fVal}C$, $FAT_{fVal}C^{*}$ and FATCA all show electron density for the (amino-)ppant (*35*) arm in the C_2 tunnel (Fig. 4F, fig. S2C, E, F). $FAT_{fVal}C$ contains extra density attached to the amino-ppant which fits fVal (fig. S2E, F), placing the formyl group within hydrogen bonding distance of Tyr810 (Fig. 4G). The donor ppant-fVal would require a small shift of the Val to expose the reactive carbonyl carbon to the acceptor site (Fig. 4G), which may only occur when acceptor substrate binds to the active site. A transient "opening" or "closing" of the V shape formed by the C domain's N- and C-lobes (*5, 34*) could also be involved, though C_2 is in very similar conformations in all the structures, not greatly influenced by whether the T domains are interacting with C_2 or what is attached to the ppants (fig. S6). The small shift of a donor substrate to achieve a fully reactive conformation is reminiscent of the large ribosomal subunit, which maintains peptidyl-tRNA in a non-reactive conformation until the aminoacyl-tRNA binds (*36*).

Dimodular conformations and the scale of conformational changes

FATCAT, FATCA and both molecules of $FAT_{Vad}CA$ provide insight into questions of super-modular architecture. In FATCAT and FATCA, T_1 is similarly bound at the donor site of C_2, but A_2 or A_2T_2 is folded back towards the initiation module in two distinctive ways (Fig. 5A): In FATCAT (and also in $FAT_{fVal}C$ and $FAT_{fVal}C^{*}$), C_2 makes contact with F_1 near its active site. In contract, the entire second module in FATCA is rotated ~114° around the F_1A_1 didomain, the C_2A_{2core} didomain center of mass is translocated by ~80 Å (Fig. 5A) and C_2 makes extensive contacts with A_1. Notably, the C_2 acceptor site is not obstructed by any of these interactions, meaning this conformation would allow a full condensation state in solution. Thus the overall conformations of FATCAT and FATCA are very different, but both seem capable of peptide bond formation.

To obtain the crystals of $FAT_{Vad}CA$, we used a valinyl-adenosine-vinylsulfamonamide (Vad) dead-end inhibitor of the thiolation reaction (Fig 1A) (*37*, *38*). Vad binds A_1 and allows the nucleophilic attack of ppant-T_1 on Vad, but is not cleaved by the reaction, tethering T_1 to Vad and stalling T_1 and A_1 in the thiolation state. Fortuitously, the crystals of $FAT_{Vad}CA$ contain two molecules in the asymmetric unit, revealing two different views of the complex. In both molecules, Vad is indeed at the A_1 active site and T_1 and A_{1sub} are in thiolation conformation, but the elongation module has remarkably different orientations. In one molecule, C_2A_2 extends away from $F_1A_1T_1$ at ~45° (Figs. 2E, 5B), and in the other it makes a ~135° angle on a different axis (Figs. 2D, 5B). The transition between the two conformations would require a ~82 Å translation and ~140° rotation of the C_2A_2 (Fig. 5B). The initiation and elongation modules do not form substantial interactions with each other in either conformation, with the T_1-C_2 linker acting as a flexible tether between the two modules.

Comparing the four dimodular conformations observed in our structures shows dramatically the scale of the conformational changes possible in a dimodular NRPS. Residues on the distal side of A_{2core} would move by between 85 Å and 216 Å to transition between observed conformations. This is similar to the length of the full dimodular NRPS, as the longest distance within any structure is 220 Å. The kinds of conformational changes required for these transitions is presented in Movie S1.

To assess whether the conformational variability seen in the crystal structures reflected flexibility in solution, we analyzed the behavior of FATCA, the LgrA construct in three of the four different crystallographic-observed conformations, using small angle X-ray scattering (SAXS). $FAT_{Vad}CA$, $FAT_{fVal}CA$ and apo-FATCA samples behaved well in solution as judged by the initial characterization of their scattering curves (Fig. S8). Notably, their pair-distance distribution functions had limited features and large Dmax values (fig. S8K-M), which is a characteristic of molecules that either adopt extended conformations or are flexible (*39*). Comparison of the experimental scattering curves to theoretical scattering curves calculated from the crystal structures resulted in very poor fits (fig. S9, Table S6), indicating that none of the individual conformations observed in the crystal

structures fully described the conformation of FATCA in solution. Using weighted combinations of the crystal structures improved the χ^2 values somewhat (Table S7) (*40*). Modelling of the scattering curves using the multiple ensemble optimization method (*41*) resulted in ensembles with excellent fits (fig. S9D-F). The ensembles were reminiscent of the series of conformations observed in the crystal structures (fig. S9G-J) and retained a level of flexibility similar to the original pool of 1000 independent models (Table S5), consistent with the interpretation that LgrA is highly flexible.

Structures and sequence statistics enable module swapping bioengineering

The structures and SAXS suggest that there is little constraint on positions of adjacent modules in NRPSs. Other than the intramodular linker, only the condensation reaction and the T_1:C_2 interaction during this reaction strictly couple neighbouring modules. Thus, high-resolution views of the T_1:C_2 interaction should enable module swapping experiments. We fused module 1 of LgrA ($LgrA_{M1}$) with the termination module 4 in cereulide synthetase ($CesB_{M4}$) to produce a chimera which synthesizes fVal-Val (Fig. 6, fig. S10). We then used DCA to guide mutagenesis of the T_1:C_{CesB-4} interface aimed at increasing activity of the chimera. Mutations in C_{CesB-4} rather than T_1 were targeted because T_1 must interact with F_1, A_1 and C_{CesB-4} for synthesis, but the donor site of C_{CesB-4} must interact only with T_1. Saturation mutagenesis *in silico* of 75 residues around the donor site of C_{CesB-4} was performed and five mutations predicted to improve T:C interaction without substantially affecting C domain folding were constructed and tested *in vitro* (Fig. 6B, Table S8). Three of the mutations decreased activity (Fig. 6C), but E925A and H1008Q showed a significant increase in fVal-Val formation (Fig. 6D). DCA scores predicted that effects of C domain mutations should be additive; $LgrA_{M1}$-$CesB_{M4}$(E925A, H1008Q) did indeed show an additive effect, doubling the peptide production of the $LgrA_{M1}$-$CesB_{M4}$ chimera.

Discussion

206 The series of structures and SAXS data presented portray multimodular NRPSs as very flexible.
207 Previous results showed that within a single module, the catalytic state can define the positions of all
208 domains in the module (e.g., the thiolation state has specific positions of all domains in an initiation or
209 elongation module) (*15, 17, 18*). In contrast, our current structures suggest that for multi-modular
210 NRPSs, there is no strict coupling between the catalytic state of a particular module and the overall
211 conformation of the multimodular NRPS. Rather, it appears that many overall conformations can allow
212 the various catalytic states: $FAT_{Vad}CA$ molecule 1, $FAT_{Vad}CA$ molecule 2, and a continuum of
213 unobserved other conformations should allow thiolation; FATCA, FATCAT and a continuum of
214 unobserved conformations should allow condensation. The continuum of conformations need not be
215 equally populated: The observation of the same condensation conformation in multiple crystal forms
216 hints it may be more common than others. Each conformation we observed was fortuitously selected
217 through crystallization in the very different packing environments of the five unrelated crystal forms.
218 Inspection of the crystal packing reveals a myriad of different crystal contacts and few trends, though
219 perhaps predictably, many more contacts are mediated by the larger domains/subdomains (F, A_{cores}, C)
220 than the small, mobile ones (A_{sub}, T), and the lower diffracting crystal forms have more porous packing
221 and spacious solvent channels.
222 The four conformations we observe are all markedly different from each other, and none of the
223 observed positions of module 2 would allow module 1 to perform each step of its catalytic cycle. (For
224 example, module 2 has to move from any observed positions to allow T_1 to reach the F_1 active site.) To
225 test if there is a single overall dimodule conformation that is compatible with the full synthetic cycle,
226 we visualized possible positions of module 2 by drawing spheres with radii of the length of the T_1-C_2
227 linker, at the observed positions of the last residue of T_1. In FATCAT, as well as in an extrapolated
228 CATCAT dimodule, there is little or no available position within the overlapping spheres (fig. S11).
229 These volume constraints, combined with the lack of any detectable C_nA_n : $C_{n+1}A_{n+1}$ co-evolution,
230 make it likely no static module:module conformation exists that accommodates the full NRPS synthetic

cycle. Although it is possible that some NRPSs have a different nature from LgrA, we suggest NRPSs in general do not possess constant and rigid super-modular architecture.

The high flexibility of NRPSs could facilitate synthetic cycles that include co-synthetic modification (tailoring), are non-canonical or are non-linear. Flexibility could be important for NRPSs with tailoring domains inserted at different positions in their architecture (*17*, *42*) or NRPSs with an abnormal order of domains, like heterobactin synthetase (which includes a C-T-A module) (*43*), obafluorin synthetase (which has an A domain C-terminal to its TE domain) (*44*, *45*), or vibriobactin synthetase (which includes VibF: Cy-Cy-A-C-T-C) (*46*). Equally unusual systems include: beauvericin synthetase, which uses tandem T domains to perform iterative condensation (*47*); myxochromides S1-3 synthetase, which performs skipping of module 4, with T_3 donating the peptide directly to module 5 (*48*); thalassospiramide B synthetase, where A_2 is proposed to thiolate T_1, T_2 and T_5 (*49*). Presumably some of these specialized systems have additional interactions that bring domains which are far apart in sequence, close together in space.

Because of their straightforward synthetic logic and important bioactive products, NRPSs have long been the subject of bioengineering attempts to create new-to-nature peptides, with mixed success (*50*, *51*). Strategies include mutation of the A domain substrate binding pocket (*52*, *53*), domain swaps (*54*), module deletion or insertion (*55*, *56*), module swaps (*54*), swaps of module-sized segments (*57-59*) and multi-modular swaps via docking domains (*60*, *61*). This includes interesting recent results using swapped $A_nT_nC_{n+1}$ segments (*57*, *58*), a strategy that conserves the native donor T_n:C_{n+1} interaction but not the acceptor C_n:T_n interaction (*57*, *58*), and using the junction of the N and C lobe of the C domain as the split point (*58*, *59*). Notably, both the N and C lobes contribute to both the donor and the acceptor T domain binding sites.

The lack of strong interactions between the didomain structural units of adjacent modules observed here should facilitate module swapping experiments, but unnatural T_n:C_{n+1} interactions could inhibit synthesis. That we increased the activity of a chimera by improving the unnatural T_1:C_2

interaction indicates that this interaction can be rate-limiting in module-swapped NRPSs. Although it did not produce orders of magnitude higher product quantity, our approach for improving unnatural T_n:C_{n+1} interactions could be combined with other bioengineering strategies, and may help NRPSs fulfill their long-held promise as sources of novel designer bioactive small molecules.

Materials and methods summary

Constructs were cloned with cleavable octa-histidine and calmodulin binding peptide tags and modified by site-directed mutagenesis. Proteins were expressed in *E.coli* and purified for crystallization by calmodulin affinity, nickel affinity, tag removal, anion exchange and gel filtration chromatography. FATCAT-C_T and LgrA$_{M1}$-CesB$_{M4}$ were purified by calmodulin affinity, nickel affinity and gel filtration. Ppants were added to apo T domains using Sfp and fVal-NH-CoA or CoA. Valinyl-adenosine-vinylsulfamonamide was complexed to FATCA by including it in the Sfp reaction.

Initial nano-volume crystallization conditions were optimized in large format to the conditions listed in Supplemental Materials. FATC was phased by MR in Phenix (*62*) with FA_{core} (*17*) and the N-lobe of TycC (PDB 2JGP) (*10*), followed by (re)building in Coot (*63*) and refinement in Phenix. Models of F, A_{1core} and C and homology models A_2 and T_2 were used for MR phasing or as a starting point for (re)building and refinement for other structures.

Small angle scattering data were collected at three concentrations and processed using ATSAS (*40*). Because of evidence of high flexibility, EOM2 (*41*) was used to generate ensembles of FATCA conformations whose theoretical combined scattering matches well the experimentally measured scattering.

For peptide synthesis by FATCAT-C_{T3}, 3.7 µM protein was incubated with 0.2 mM N10-fTHF, 5 mM valine, 2 mM glycine, 1 mM tryptamine, 5 mM ATP at 23 °C for 5 h prior to quenching and LC-

279 MS. For peptide synthesis by $LgrA_{M1}$-$CesB_{M4}$, 5 μM protein was incubated with 5 mM valine, 5 mM
280 ATP and 0.5 mM N10-fTHF for 6 h prior to quenching and LC-MS.
281 For DCA (*64*), we extracted 45,015 T_n:C_{n+1} pairs, 14,506 T_n:E_n pairs and 29,700 C_n:T_n pairs and
282 calculated interdomain contact scores and domain-domain interaction scores. To computationally
283 suggest mutations to improve the T_n:C_{n+1} interaction, we altered all single amino acids in C_{n+1} interface
284 positions to all 19 other amino acids and evaluated the interaction scores, looking for those that
285 improve the T_n:C_{n+1} interaction score and do not affect substantially the C domain score.
286
287 **Supplementary materials**
288 Detailed Materials and Methods
289 Figs. S1 to S11
290 Tables S1 to S10
291 Supplemental References 65-89
292 Movie S1
293 **References**

294 1. C. T. Walsh, Polyketide and nonribosomal peptide antibiotics: modularity and versatility.
295 *Science* **303**, 1805-1810 (2004).
296 2. E. A. Felnagle *et al.*, Nonribosomal peptide synthetases involved in the production of medically
297 relevant natural products. *Mol Pharm* **5**, 191-211 (2008).
298 3. N. A. McGrath, M. Brichacek, J. T. Njardarson, A Graphical Journey of Innovative Organic
299 Architectures That Have Improved Our Lives. *J Chem Educ* **87**, 1348-1349 (2010).
300 4. J. M. Reimer, A. S. Haque, M. J. Tarry, T. M. Schmeing, Piecing together nonribosomal
301 peptide synthesis. *Curr Opin Struct Biol* **49**, 104-113 (2018).

5. K. Bloudoff, T. M. Schmeing, Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochimica et biophysica acta* **1865**, 1587-1604 (2017).
6. G. H. Hur, C. R. Vickery, M. D. Burkart, Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* **29**, 1074-1098 (2012).
7. C. T. Walsh *et al.*, Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol* **5**, 525-534 (2001).
8. K. J. Weissman, The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* **11**, 660-670 (2015).
9. C. Chiocchini, U. Linne, T. Stachelhaus, In vivo biocombinatorial synthesis of lipopeptides by COM domain-mediated reprogramming of the surfactin biosynthetic complex. *Chem Biol* **13**, 899-908 (2006).
10. S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L. O. Essen, Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781-792 (2007).
11. M. J. Tarry, A. S. Haque, K. H. Bui, T. M. Schmeing, X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* **25**, 783-793 e784 (2017).
12. M. A. Marahiel, A structural model for multimodular NRPS assembly lines. *Nat Prod Rep* **33**, 136-140 (2016).
13. G. Schoenafinger, N. Schracke, U. Linne, M. A. Marahiel, Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J Am Chem Soc* **128**, 7406-7407 (2006).
14. B. A. Wallace, Common structural features in gramicidin and other ion channels. *Bioessays* **22**, 227-234 (2000).

15. A. Tanovic, S. A. Samel, L. O. Essen, M. A. Marahiel, Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659-663 (2008).
16. J. M. Reimer, M. N. Aloise, H. R. Powell, T. M. Schmeing, Manipulation of an existing crystal form unexpectedly results in interwoven packing networks with pseudo-translational symmetry. *Acta Crystallogr D Biol Crystallogr* **72**, 1130-1136 (2016).
17. J. M. Reimer, M. N. Aloise, P. M. Harrison, T. M. Schmeing, Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239-242 (2016).
18. E. J. Drake *et al.*, Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235-238 (2016).
19. B. R. Miller, E. J. Drake, C. Shi, C. C. Aldrich, A. M. Gulick, 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).
20. J. Alfermann *et al.*, FRET monitoring of a nonribosomal peptide synthetase. *Nat Chem Biol* **13**, 1009-1015 (2017).
21. K. Bloudoff, D. Rodionov, T. M. Schmeing, 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).
22. A. S. Reger, R. Wu, D. Dunaway-Mariano, A. M. Gulick, Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* **47**, 8016-8025 (2008).
23. C. A. Mitchell, C. Shi, C. C. Aldrich, A. M. Gulick, Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry* **51**, 3252-3263 (2012).
24. A. M. Gulick, Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases. *Curr Opin Chem Biol* **35**, 89-96 (2016).

25. F. Morcos *et al.*, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A* **108**, E1293-1301 (2011).
26. L. Yuwen *et al.*, The role of aromatic L-amino acid decarboxylase in bacillamide C biosynthesis by Bacillus atrophaeus C89. *Sci Rep* **3**, 1753 (2013).
27. K. Bloudoff, C. D. Fage, M. A. Marahiel, T. M. Schmeing, Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc Natl Acad Sci U S A* **114**, 95-100 (2017).
28. M. Figliuzzi, H. Jacquier, A. Schug, O. Tenaillon, M. Weigt, Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1. *Mol Biol Evol* **33**, 268-280 (2016).
29. W.-H. Chen, K. Li, N. S. Guntaka, S. D. Bruner, Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases. *ACS chemical biology* **11**, 2293-2303 (2016).
30. J. Zhang *et al.*, Structural basis of nonribosomal peptide macrocyclization in fungi. *Nat Chem Biol* **12**, 1001-1003 (2016).
31. C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson, Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol* **7**, 78 (2007).
32. C. Y. Chang *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).
33. V. Bergendahl, U. Linne, M. A. Marahiel, Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur J Biochem* **269**, 620-629 (2002).

34. K. Bloudoff, D. A. Alonzo, T. M. Schmeing, Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chem Biol* **23**, 331-339 (2016).
35. Y. Liu, S. D. Bruner, Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases. *Chembiochem* **8**, 617-621 (2007).
36. T. M. Schmeing, K. S. Huang, S. A. Strobel, T. A. Steitz, An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* **438**, 520-524 (2005).
37. C. Qiao, D. J. Wilson, E. M. Bennett, C. C. Aldrich, A mechanism-based aryl carrier protein/thiolation domain affinity probe. *J Am Chem Soc* **129**, 6350-6351 (2007).
38. J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick, Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem Biol* **19**, 188-198 (2012).
39. P. Bernado, Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur Biophys J* **39**, 769-780 (2010).
40. D. Franke *et al.*, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography* **50**, 1212-1225 (2017).
41. G. Tria, H. D. Mertens, M. Kachala, D. I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207-217 (2015).
42. S. Mori *et al.*, Structural basis for backbone N-methylation by an interrupted adenylation domain. *Nature Chemical Biology*, (2018).
43. M. Bosello *et al.*, Structural Characterization of the Heterobactin Siderophores from Rhodococcus erythropolis PR4 and Elucidation of Their Biosynthetic Machinery. *Journal of Natural Products* **76**, 2282-2290 (2013).
44. J. E. Schaffer, M. R. Reck, N. K. Prasad, T. A. Wencewicz, beta-Lactone formation during product release from a nonribosomal peptide synthetase. *Nat Chem Biol* **13**, 737-744 (2017).

45. D. F. Kreitler, E. M. Gemmell, J. E. Schaffer, T. A. Wencewicz, A. M. Gulick, The structural basis of N-acyl-alpha-amino-beta-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nature communications* **10**, 3432 (2019).

46. N. J. Hillson, C. T. Walsh, Dimeric structure of the six-domain VibF subunit of vibriobactin synthetase: mutant domain activity regain and ultracentrifugation studies. *Biochemistry* **42**, 766-775 (2003).

47. Y. Xu *et al.*, Biosynthesis of the cyclooligomer depsipeptide beauvericin, a virulence factor of the entomopathogenic fungus Beauveria bassiana. *Chem Biol* **15**, 898-907 (2008).

48. S. C. Wenzel *et al.*, Structure and biosynthesis of myxochromides S1-3 in Stigmatella aurantiaca: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *Chembiochem* **6**, 375-385 (2005).

49. A. C. Ross *et al.*, Biosynthetic multitasking facilitates thalassospiramide structural diversity in marine bacteria. *J Am Chem Soc* **135**, 1155-1162 (2013).

50. M. Winn, J. K. Fyans, Y. Zhuo, J. Micklefield, Recent advances in engineering nonribosomal peptide assembly lines. *Nat Prod Rep* **33**, 317-347 (2016).

51. A. S. Brown, M. J. Calcott, J. G. Owen, D. F. Ackerley, Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat Prod Rep*, (2018).

52. K. Eppelmann, T. Stachelhaus, M. A. Marahiel, Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718-9726 (2002).

53. J. Thirlway *et al.*, Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew Chem Int Ed Engl* **51**, 7181-7184 (2012).

54. K. T. Nguyen *et al.*, Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc Natl Acad Sci U S A* **103**, 17462-17467 (2006).
55. H. D. Mootz *et al.*, Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J Am Chem Soc* **124**, 10980-10981 (2002).
56. D. Butz *et al.*, Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by Amycolatopsis balhimycina. *Chembiochem* **9**, 1195-1200 (2008).
57. K. A. J. Bozhuyuk *et al.*, De novo design and engineering of non-ribosomal peptide synthetases. *Nat Chem* **10**, 275-281 (2018).
58. C. Steiniger *et al.*, Harnessing fungal nonribosomal cyclodepsipeptide synthetases for mechanistic insights and tailored engineering. *Chem Sci* **8**, 7834-7843 (2017).
59. K. A. J. Bozhüyük *et al.*, Modification and de novo design of non-ribosomal peptide synthetases (NRPS) using specific assembly points within condensation domains. *bioRxiv*, 354670 (2018).
60. M. Hahn, T. Stachelhaus, Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc Natl Acad Sci U S A* **103**, 275-280 (2006).
61. C. Hacker *et al.*, Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. *Nature communications* **9**, 4366 (2018).
62. P. D. Adams *et al.*, PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954 (2002).
63. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

64. M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, T. Hwa, Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A* **106**, 67-72 (2009).

65. D. A. Alonzo, N. A. Magarvey, T. M. Schmeing, Characterization of cereulide synthetase, a toxin-producing macromolecular machine. *PLoS One* **10**, e0128569 (2015).

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

67. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **291**, 1790-1792 (2001).

68. A. S. Worthington, M. D. Burkart, One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org Biomol Chem* **4**, 44-46 (2006).

69. L. E. Quadri *et al.*, Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595 (1998).

70. A. G. W. Leslie, H. R. Powell, Processing diffraction data with MOSFLM. *Nato Sci Ser Ii Math* **245**, 41-51 (2007).

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

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

73. A. J. McCoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

74. G. Bunkoczi, R. J. Read, Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr D Biol Crystallogr* **67**, 303-312 (2011).

75. A. T. Brunger, Version 1.2 of the crystallography and NMR system. *Nat Protoc* **2**, 2728-2733 (2007).

76. M. Biasini *et al.*, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**, W252-258 (2014).
77. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074-1080 (2009).
78. S. D. Breazeale, A. A. Ribeiro, C. R. Raetz, Oxidative decarboxylation of UDP-glucuronic acid in extracts of polymyxin-resistant Escherichia coli. Origin of lipid a species modified with 4-amino-4-deoxy-L-arabinose. *J Biol Chem* **277**, 2886-2896 (2002).
79. G. L. Hura *et al.*, Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods* **6**, 606-612 (2009).
80. K. N. Dyer *et al.*, High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol Biol* **1091**, 245-258 (2014).
81. S. Classen *et al.*, Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J Appl Crystallogr* **46**, 1-13 (2013).
82. P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, D. I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277-1282 (2003).
83. D. Svergun, C. Barberato, M. H. J. Koch, CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768-773 (1995).
84. R. D. Finn *et al.*, The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-285 (2016).
85. S. R. Eddy, Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195 (2011).

86. C. The UniProt, UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169 (2017).
87. N. Ziemert *et al.*, The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **7**, e34064 (2012).
88. M. Ekeberg, C. Lovkvist, Y. Lan, M. Weigt, E. Aurell, Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Phys Rev E Stat Nonlin Soft Matter Phys* **87**, 012707 (2013).
89. A. Procaccini, B. Lunt, H. Szurmant, T. Hwa, M. Weigt, Dissecting the specificity of protein-protein interaction in bacterial two-component signaling: orphans and crosstalks. *PLoS One* **6**, e19729 (2011).

Acknowledgments: We thank Diego Alonzo (MS, ligand building), Angelos Pistofidis and Yael Ripstein (purification help), Kris Bloudoff, Camille Fortinez (*bmdB* plasmid), Jessie Jiang (cloning), Schmeing lab members (discussions), Nancy Rogerson (editing), Greg Hura (SAXS advice and analyses), Kathryn Burnett (SAXS data collection) and Bhushan Nagar (diffraction data discussions). We thank Christian Chalut for BL21(DE3)*entD*-, John Colucci, Mika Guerard and Robert Zamboni (ZCS), staff at CLS 08ID-1 (S. Labiuk, J. Gorin, M. Fodje, K. Janzen, D. Spasyuk, P. Grochulski), APS 24-ID-C (grants GM124165, RR029205, DE-AC02-06CH11357; F. Murphy), SAXS data Advanced Light Source (ALS), SIBYLS beamline (US-DOE-BER Integrated Diffraction Analysis Technologies, NIGMS ALS-ENABLE-P30-GM124169, S10OD018483). **Funding:** This work was funded by CIHR (FDN-148472) and a Canada Research Chair to T.M.S., a European Union H2020 research and innovation programme MSCA-RISE-2016 (#734439 INFERNET) grant to M.W., and studentships from NSERC (J.M.R.) Boehringer Ingelheim Fonds (M.E). and CIHR (I.H.) **Author contributions:** J.M.R. performed the crystallography with assistance of I.H. J.M.R performed activity assays for LgrA-CesB chimeric proteins. M.E. performed activity assays for LgrA-BmdB proteins and

structure refinement. J.M.R and M.E. prepared sample for SAXS and A.G. performed analyses of SAXS data. M.W. performed co-evolution and bioinformatic analyses. T.M.S. directed the project. T.M.S, J.M.R., M.W. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability**: Structural data are available in the RCSB Protein Databank (PDB IDs $FAT_{fVal}C$: 6MFW, $FAT_{fVal}C^{*}$: 6MFX, $FAT_{Vad}CA$: 6MFY, FATCAT: 6MFZ, FATCA: 6MG0). All other data are available in the main text or the supplementary materials.

Figures

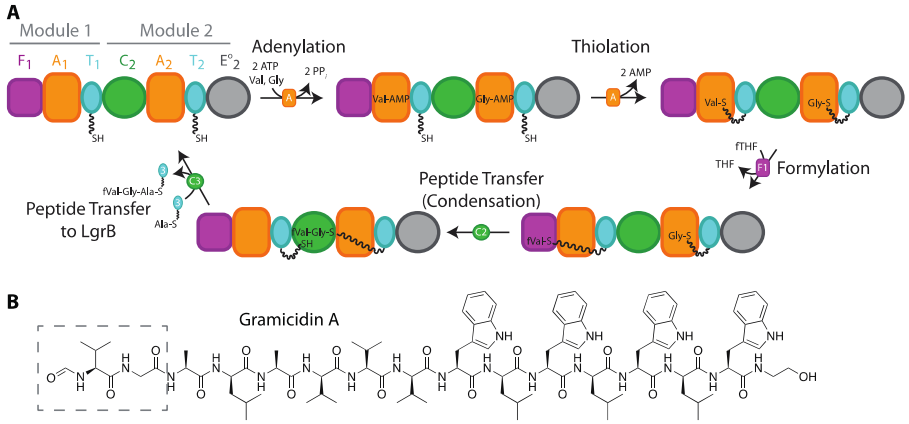


Figure 1

Figure 1. Overview of the biosynthetic steps performed by LgrA.

A, Linear gramicidin synthetase subunit A is a dimodular protein with initiation ($F_1A_1T_1$) and elongation modules ($C_2A_2T_2E^o_2$). Valine and glycine are selected and adenylated by A_1 and A_2, and then transferred to T_1 and T_2. Val-T_1 is formylated by F_1 and then peptide bond formation between fVal-T_1 and Gly-T_2 by C_2 produces fVal-Gly-T_2. This is the donor substrate for peptide bond formation in the C_3 domain of the next NRPS subunit, LgrB. (This schematic is not intended to indicate the timing of rebinding of substrates.) **B**, Chemical structure of linear gramicidin A with a box highlighting the fVal-Gly portion assembled by LgrA. F, formylation domain; A, adenylation domain; T, thiolation domain; C, condensation domain; E^o, inactive epimerization domain.

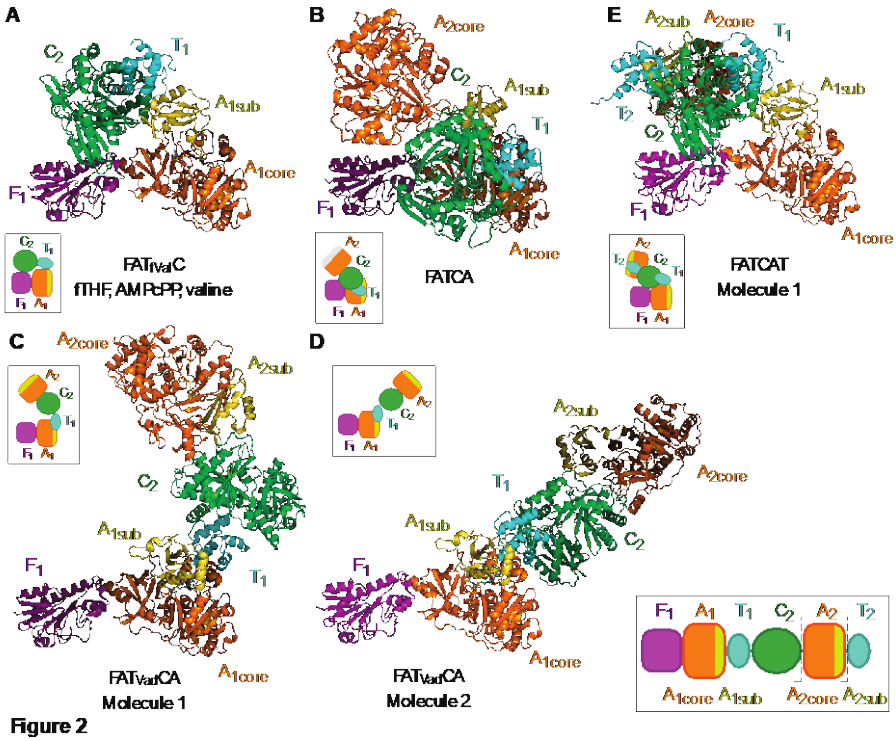
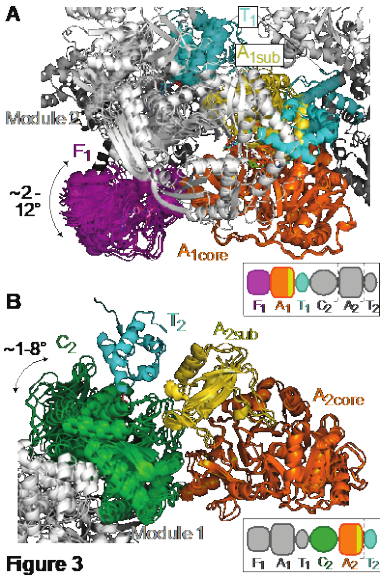


Figure 2. Crystal structures of dimodular LgrA.

Structures of LgrA constructs **A**, $FAT_{fVal}C$, which has fVal-aminoppant on T_1 and is bound with valine, AMPcPP and fTHF (space group $P2_12_12_1$; 2.5 Å resolution), **B**, FATCA ($P2_12_12_1$; 2.5 Å resolution), **D, E**, two crystallographically-independent molecules of $FAT_{Vad}CA$, for which dead-end valinyl-adenosine-vinylsulfamonamide (Vad) were used to stall T_1 at A_1 during thiolation ($P2_12_12_1$; 6 Å resolution) **F**, **G**, FATCAT, which has ppant on both T domains ($C222_1$; 6 Å resolution). See fig S1A-B for additional structures. Cartoon insets: schematic depicting protein constructs crystallized. Domains which are greyed in the labels are disordered in the crystal structure



549 Figure 3

550 **Figure 3. Comparison of intramodular conformations.**

551 **A**, The structural unit for the initiation module (F_1A_{1core}) and **B,** elongation module (C_2A_{2core}) are in

552 similar conformations in all structures. The F_1:A_{1core} interface buries 773-860 $Å^2$ and the C_2:A_{2core}

553 interface buries 565-770 $Å^2$. Superimposing based A_{core} shows a ~1-12° shift of F_1 or C_2. Previously

554 EntF C_1:A_1 was seen to shift ~15° (*18*) between structures.

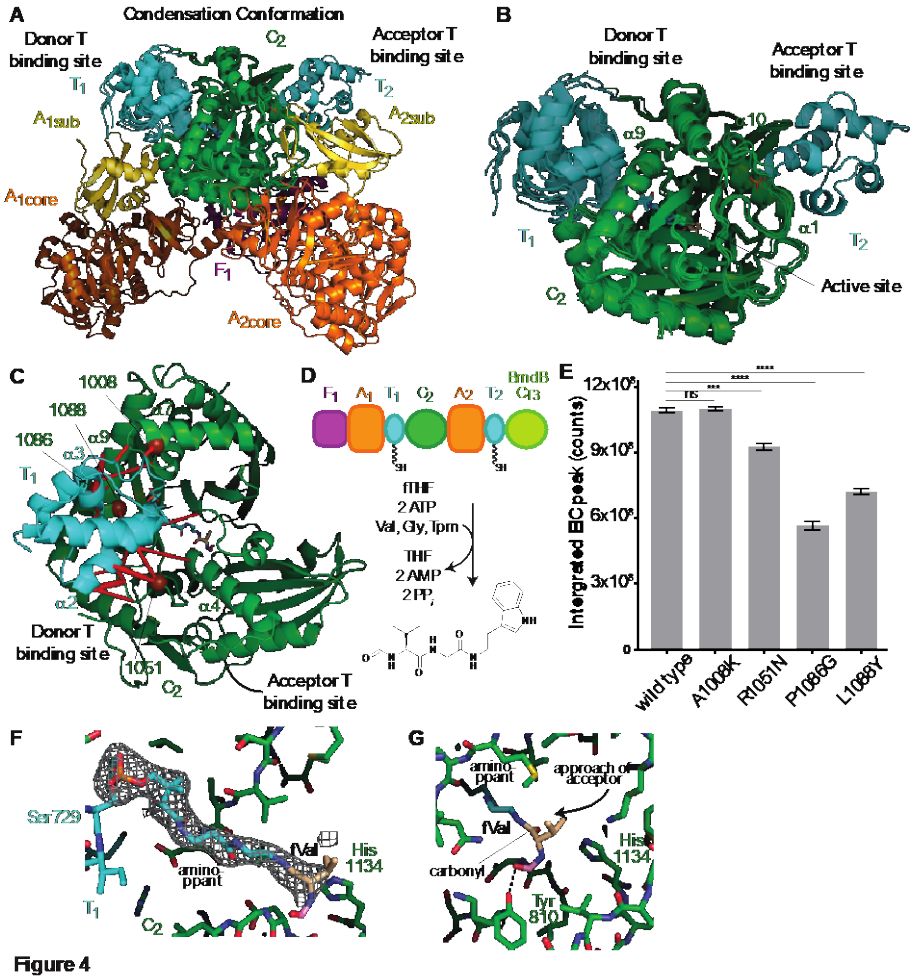


Figure 4

Figure 4. Condensation in LgrA.

A, FATCAT shows a full condensation state, with T_1 and T_2 docked at the donor and acceptor binding sites, respectively. $FAT_{fVal}C$ and $FAT_{fVal}C^*$ are overlayed to show that the first four domains are in analogous positions regardless of space group or resolution of the structures. **B**, Overlay of LgrA structures with T_1 at C_2. **C**, DCA between T_n and C_{n+1} displayed on LgrA; co-evolution signal (red lines) between residues in close proximity in the LgrA structures. Residues selected for mutation are indicated with brown α-carbon spheres. **D**, Schematic of the LgrA-BmdB chimera protein FATCAT-C_{T3} and its product fVal-Gly-Tpm. **E**, LC-MS peptide synthesis assay of mutations near the T_1 binding site of C_2. All reactions were performed in triplicate (n=3) and are shown as mean values from integration of extracted-ion-chromatogram (EIC) peaks, normalized against the average wild type value. Statistical significance was determined by two-sided student's t-test (ns: $p > 0.05$; *: $p \leq 0.05$;

: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$). **F, Unbiased F_O-F_C (3σ) simulated annealing Polder omit electron density map of the C_2 active site of $FAT_{fVal}C$, calculated with phases from model that never included fVal-ppant ligand. **G**, Zoom in to the $FAT_{fVal}C$ C_2 active site which shows that the Val would rotate for condensation.

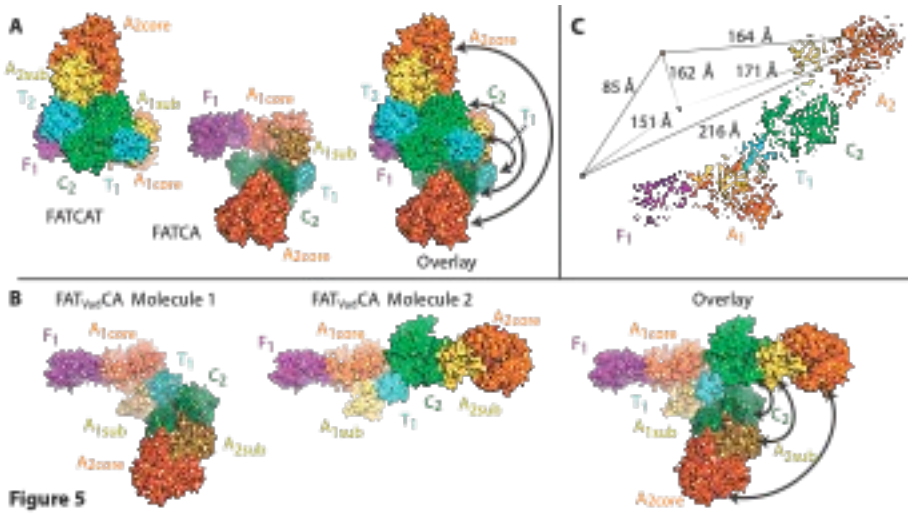


Figure 5. Different dimodular conformations for the same catalytic states.

A, FATCAT and FATCA both show T_1 binding to the donor site of C_2 but have very different overall conformations. **B**, The two crystallographically-independent molecules of $FAT_{Vad}CA$ both show module 1 in thiolation conformation, but have very different positions of module 2. **C**, The distances between positions of residue D1236 in the four dimodular conformations. The structures are superimposed by their A_{1core}.

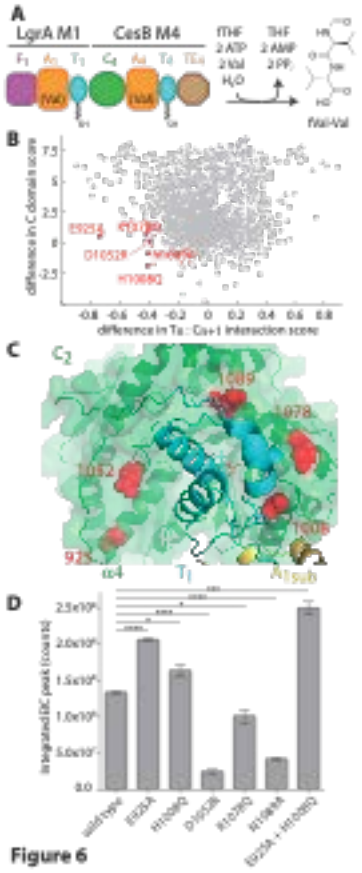


Figure 6. Module swapping bioengineering using structural and direct coevolution analysis. **A**, Schematic of the $LgrA_{M1}$-$CesB_{M4}$ chimera and its product fVal-Val. **B**, Scatter diagram of the difference in C domain score against the difference in T_n:C_{n+1} interaction score for the DCA analysis of 75 residues around the donor T domain binding site of CesB C_4 mutated to each of the other 19 amino acids. CesB C_4 mutations in red were analyzed biochemically. **C**, The corresponding positions of these five mutations in LgrA C_2. **D**, Wild type $LgrA_{M1}$-$CesB_{M4}$ and mutant proteins were assayed for peptide production by LC-MS. All reactions were performed in triplicate (n=3) and are shown as mean values from integration of extracted-ion-chromatogram (EIC) peaks, normalized against the average wild type value. Statistical significance was determined by two-sided student's t-test (ns: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$).