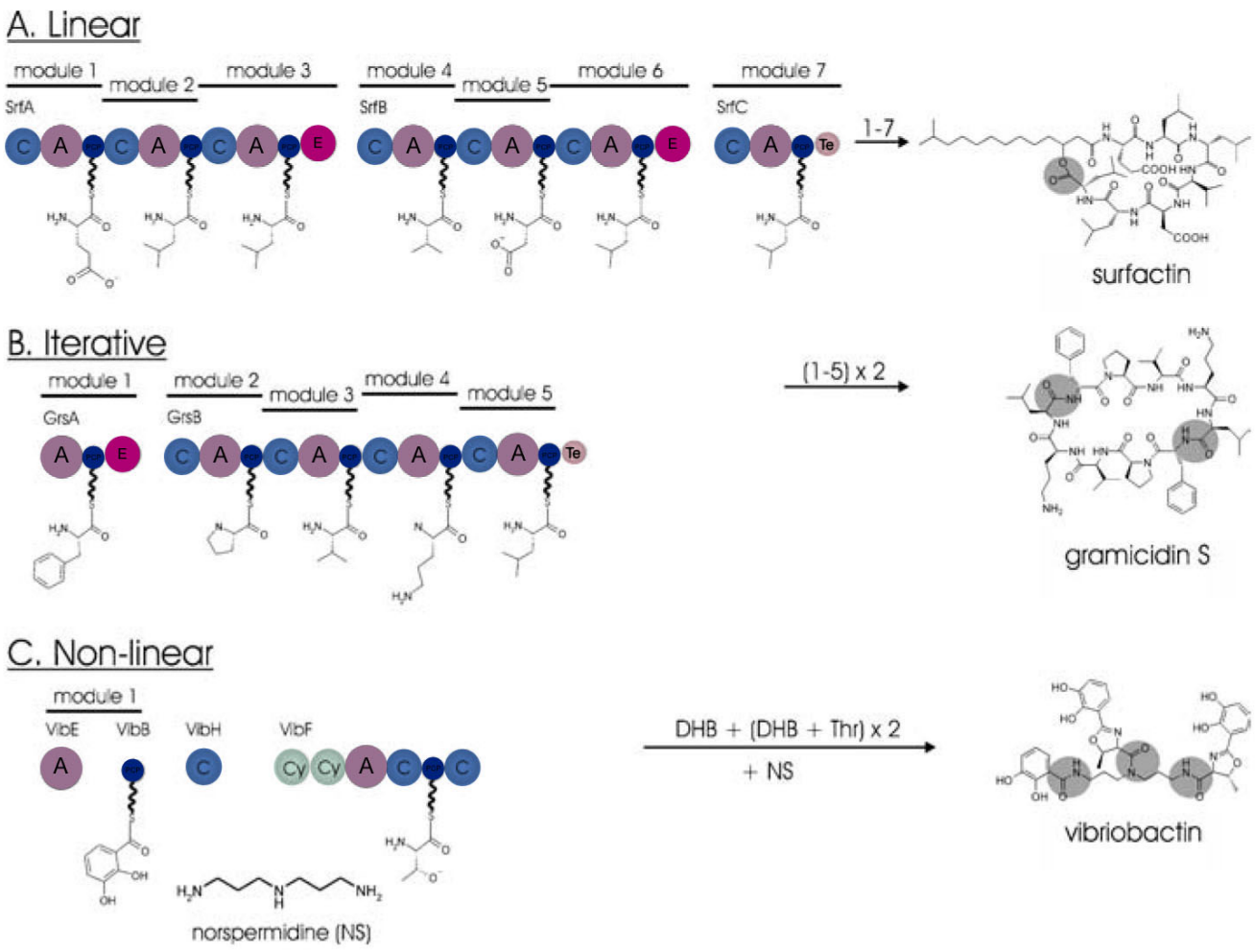
Strategies for analysis and modification of NRPS

Classification of NRPS



Thio-template mechanism of NRPS

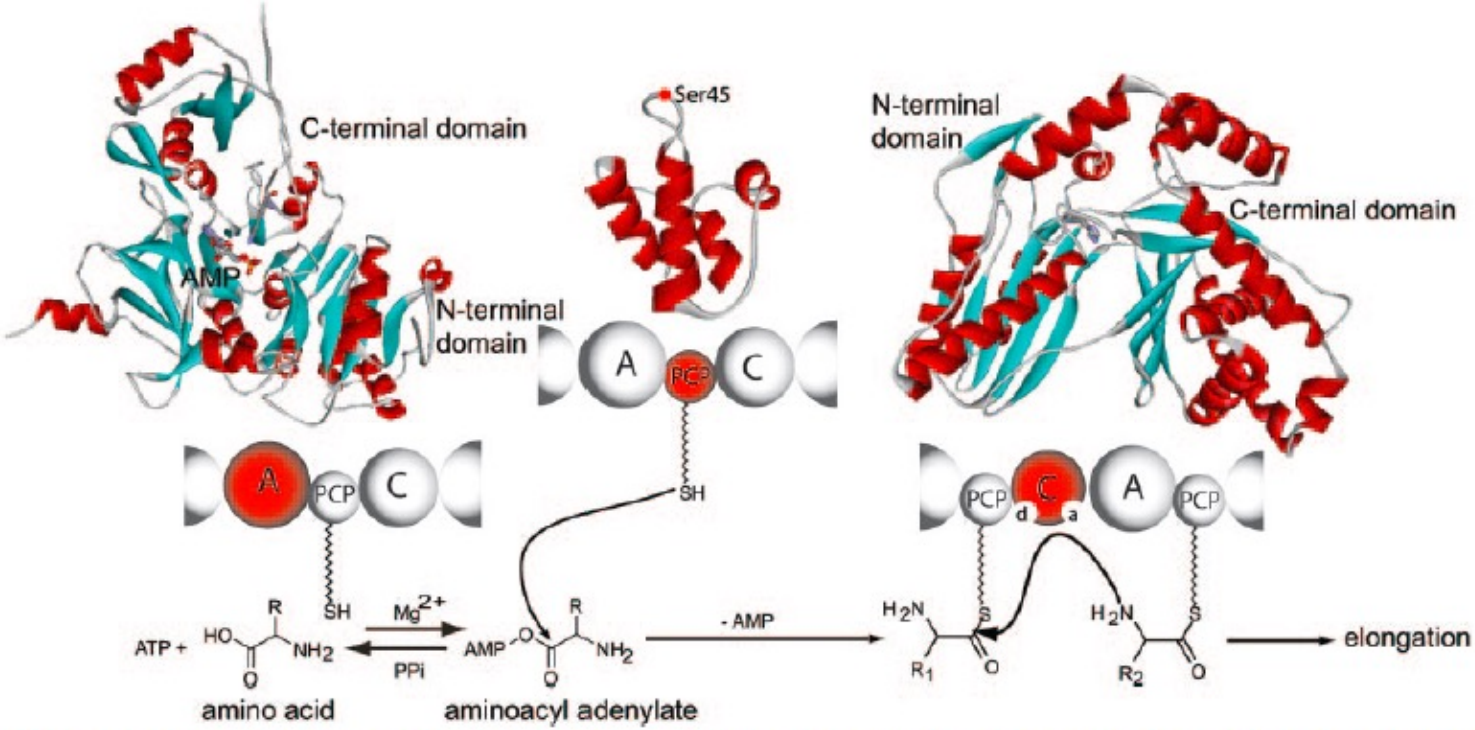
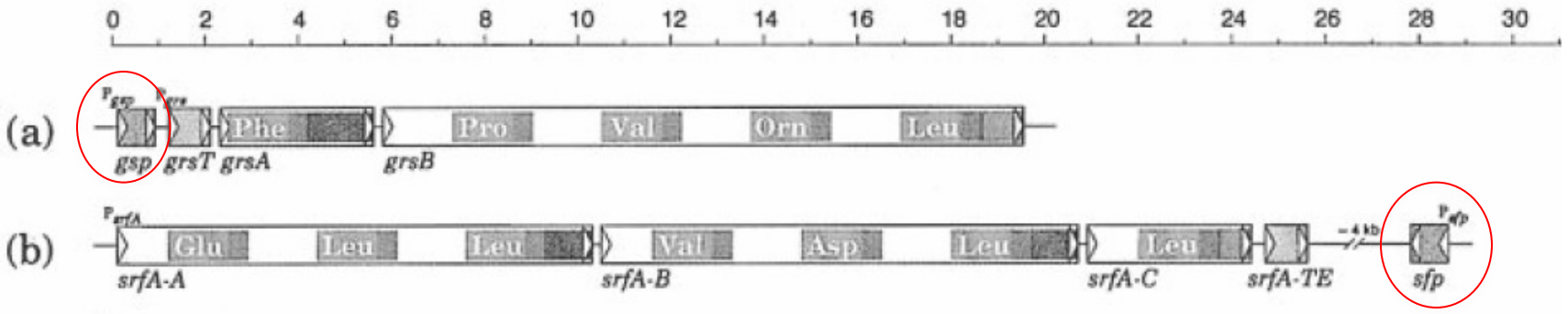


FIG. 4. Chemical principles of nonribosomal peptide synthesis. Domains in action are indicated in red and the respective crystal structures are shown above. First, the A-domain specifically recognizes a dedicated amino acid and catalyzes formation of the aminoacyl adenylate under consumption of ATP. Second, the activated aminoacyl adenylate is tethered to the free thiol group of the PCP-bound phosphopantetheine (ppan) cofactor. Third, the C-domain catalyzes peptide elongation. Here, the nucleophilic amine of the acceptor substrate nucleophilically attacks the electrophilic thioester of the donor substrate (a, acceptor site; d, donor site). The crystal structure of the A-domain is derived from the Phe-activating A-domain (PheA) of the first module of gramicidin S synthetase of *B. brevis* (22). The NMR-structure of the PCP is derived from the third module of the *B. brevis* tyrocidine synthetase (141), and the C-domain is derived from the crystal structure of VibH, a stand alone C-domain of the *V. cholerea* vibriobactin synthetase (60).

Thio-template mechanism of NRPS

Addition of 4-phosphopantetheine on the active site serine of the thiolation domain is required for NRPS activity. This post-translational modification is catalyzed by 4-phosphopantetheine transferases (4-PPTases), accessory enzymes encoded by genes belonging to the NRPS cluster.



Sfp is the 4-PPTase most widely used for the studies of NRPS.

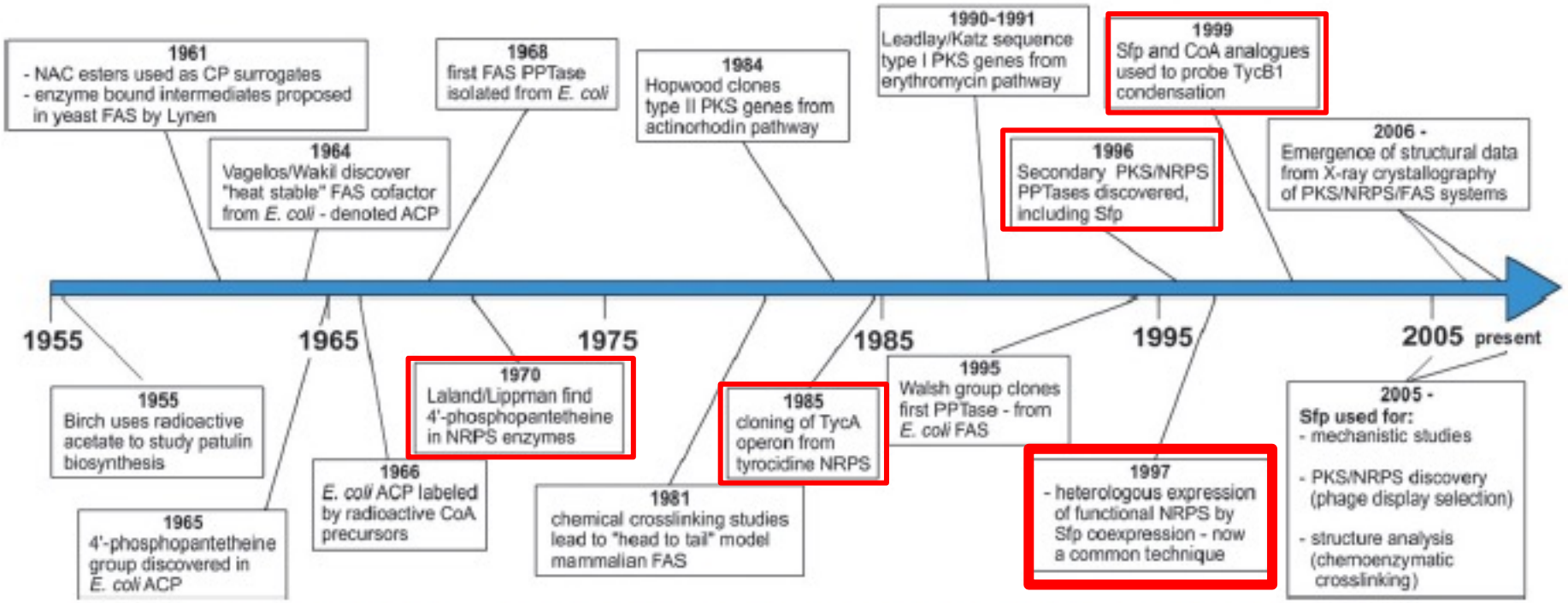
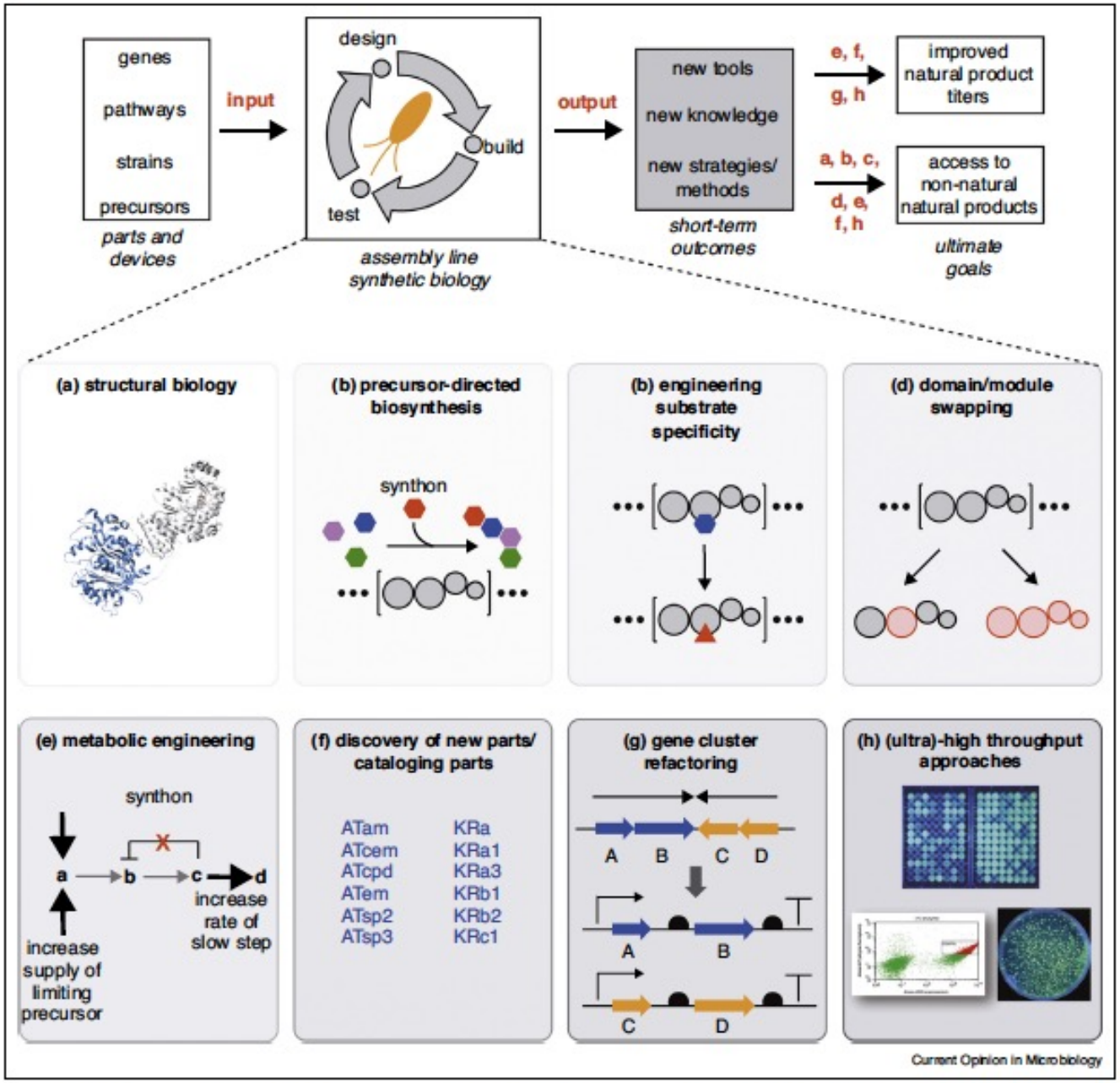


Fig. 2 Timeline of some notable events in the study of carrier-protein-mediated FAS, PKS, and NRPS biosynthetic enzymes.

Recombinant expression of NRPS or of single modules and domains allows to set up the studies of the structure and function of these enzymes at the molecular level.

Strategies for production of structural analogues of bioactive peptides

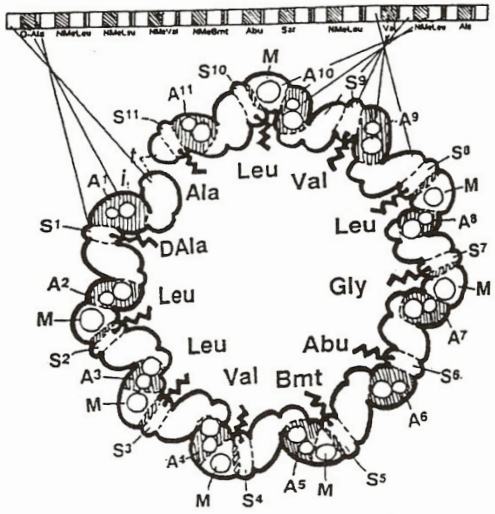
- Alteration of growth conditions of the producing microorganism
  - Precursor-directed biosynthesis
- Reprogramming of NRPS
  - Targeted alteration of recognition specificity (adenylation domain)
  - Construction of hybrid NRPS by module or domain exchange
    - A domain substitution (A-T)
    - C-A domain substitution (C-A-T)



Summary of strategies and tools employed to engineering assembly lines. Most assembly line engineering efforts leverage the dogma of synthetic biology and build on the concept of a 'Design-Build-Test' cycle. With each new advancement towards a different aspect of assembly line biosynthesis, novel strategies and tools (e.g. a–h, arranged in approximate order of typical throughput) are being developed that feed back into our ability to achieve the ultimate goals: improved titers and access to novel natural product analogues.

Production of cyclosporin A analogues

Cyclosporin A is an immunosuppressor produced by the fungus *Tolypocladium niveum*. Cyclosporin synthetase is constituted by about 15000 amino acids organized in 11 modules.



Numerous structural analogues of cyclosporin A have been isolated from cultures in media containing different precursors.

➔ NRPS is able to insert different amino acids in some positions of the peptide.

peptide	1	2	3	4	5	6	7	8	9	10	11
CyA	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	Abu	Sar	MeLeu	Val	MeLeu	Ala
CyB	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	**Ala**	Sar	MeLeu	Val	MeLeu	Ala
CyC	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	**Thr**	Sar	MeLeu	Val	MeLeu	Ala
CyD	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	**Val**	Sar	MeLeu	Val	MeLeu	Ala
CyE	D-Ala	MeLeu	MeLeu	**Val**	MeBmt	Abu	Sar	MeLeu	Val	MeLeu	Ala
CyF	D-Ala	MeLeu	MeLeu	MeVal	**MeBmt**	**Nva**	Sar	MeLeu	Val	MeLeu	Ala
CyG	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	Abu	Sar	MeLeu	Val	MeLeu	Ala
CyH	D-Ala	MeLeu	MeLeu	**D-MeVal**	MeBmt	Abu	Sar	MeLeu	Val	MeLeu	Ala
CyI	D-Ala	**Leu**	MeLeu	MeVal	MeBmt	**Val**	Sar	MeLeu	Val	MeLeu	Ala
CyK	D-Ala	MeLeu	MeLeu	MeVal	**MedBmt**	**Val**	Sar	MeLeu	Val	MeLeu	Ala
CyL	D-Ala	MeLeu	MeLeu	MeVal	**Bmt**	Abu	Sar	MeLeu	Val	MeLeu	Ala
CyM	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	**Nva**	Sar	MeLeu	**Nva**	MeLeu	Ala
CyN	D-Ala	**Leu**	MeLeu	MeVal	MeBmt	**Nva**	Sar	MeLeu	Val	MeLeu	Ala
CyO	D-Ala	MeLeu	MeLeu	MeVal	**MeLeu**	**Nva**	Sar	MeLeu	Val	MeLeu	Ala
CyP	D-Ala	MeLeu	MeLeu	MeVal	**Bmt**	**Thr**	Sar	MeLeu	Val	MeLeu	Ala
CyQ	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	Abu	Sar	**Val**	Val	**MeLeu**	Ala
CyR	D-Ala	**Leu?**	MeLeu	MeVal	MeBmt	Abu	Sar	MeLeu	Val	**Leu?**	Ala
CyS	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	**Thr**	Sar	**Val**	Val	MeLeu	Ala
CyU	D-Ala	**Leu**	MeLeu	MeVal	MeBmt	Abu	Sar	MeLeu	Val	**Leu**	Ala
CyV	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	Abu	Sar	MeLeu	Val	MeLeu	**Abu**
CyW	D-Ala	MeLeu	MeLeu	**Val**	MeBmt	Thr	Sar	MeLeu	Val	MeLeu	Ala
CyX	D-Ala	**Leu**	MeLeu	MeVal	MeBmt	**Nva**	Sar	MeLeu	Val	MeLeu	Ala
CyY	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	**Nva**	Sar	MeLeu	Val	**Leu**	Ala
CyZ	D-Ala	MeLeu	MeLeu	MeVal	MeAOA	Abu	Sar	MeLeu	Val	MeLeu	Ala
Cy26	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	**Nva**	Sar	MeLeu	**Leu**	MeLeu	Ala
Cy27	D-Ala	MeLeu	MeLeu	MeVal	**Bmt**	**Val**	Sar	MeLeu	Val	MeLeu	Ala
Cy28	D-Ala	MeLeu	MeLeu	MeVal	**MeLeu**	Abu	Sar	MeLeu	Val	MeLeu	Ala
Cy29	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	Abu	Sar	**MeIle**	Val	MeLeu	Ala
Cy30	D-Ala	MeLeu	MeLeu	MeVal	**MeLeu**	**Val**	Sar	MeLeu	Val	MeLeu	Ala
Cy31	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	Abu	Sar	**Ile**	Val	MeLeu	Ala
Cy32	D-Ala	MeLeu	MeLeu	MeVal	MeBmt	Abu	Sar	MeLeu	Val	MeLeu	Ala

Strategies for NRPS reprogramming

- It is necessary to have information on the structure and the molecular details of the mechanism of NRPS
  - Which is the molecular basis of substrate-specificity of the adenylation domain?
  - Thiolation and condensation domains are substrate-specific?
  - Which is the molecular basis of recognition between domains and modules?

- ➔ Biochemical studies *in vitro* on purified proteins (simplified versions: isolated domains or modules)

Molecular basis of substrate binding in the adenylation domain

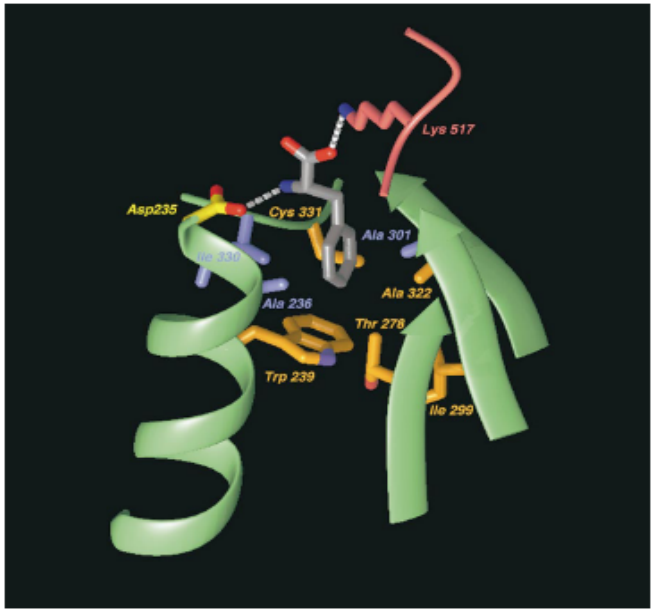
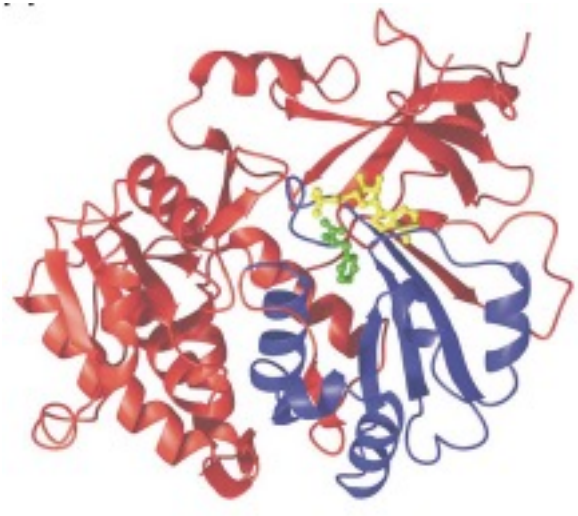


Figure 3. The Phenylalanine binding Pocket of PheA as a Prototype Substrate binding Pocket for the Superfamily of Adenylate-Forming Enzymes

The green ribbon elements are located in the larger N-terminal domain and comprise helix 8 and β sheets B5, B6, and B7 (left to right, blue in Figure 2a). The red loop coming from the top belongs to the smaller C-terminal domain of PheA. For amino acid activating A domains, the Asp235 (yellow) and Lys517 (red) residues are invariant. The sidechains in blue vary only slightly between different A domains and are mostly hydrophobic, whereas residues in orange are highly variant and their composition confers the greatest part of the substrate recognition. The phenylalanine substrate of PheA is shown in gray. (The figure was prepared using the program MOLMOL [38].)



Structure determination of the adenylation domain PheA from Gramicidin Synthetase A in the presence of Phe

Identification of amino acid residues that form the Phe substrate binding pocket

Molecular basis of substrate binding in the adenylation domain

The selectivity-conferring code of A domains.

	Position											
Domain	235	236	239	278	299	301	322	330	331	517	Biosynthetic template	Similarity
Aad	E	P	R	N	I	V	E	F	V	K	AcvA	94%
Ala	D	L	L	F	G	I	A	V	L	K	CssA, Hts1	55%
Asn	D	L	T	K	L	G	E	V	G	K	BacA, CepA, Dae, Glg1, TycC	90%
Asp	D	L	T	K	V	G	H	I	G	K	BacC, SrfAB, LicB, LchAB	100%
Cys(1)	D	H	E	S	D	V	G	I	T	K	AcvA	96%
Cys(2)	D	L	Y	N	L	S	L	I	W	K	BacA, HMWP2	88%
Dab	D	L	E	H	N	T	T	V	S	K	SyrE	100%
Dhb/Sal	P	L	P	A	Q	G	V	V	N	K	EntE, DhbE, MbtA, PchD, VibE, YbtE	83%
Gln	D	A	Q	D	L	G	V	V	D	K	LicA, LchAA	100%
Glu(1)	D	A	W	H	F	G	G	V	D	K	FenA, FenC, FenE, PPS1, PPS3, PPS4	95%
Glu(2)	D	A	K	D	L	G	V	V	D	K	BacC, SrfAA	95%
Ile (1)	D	G	F	F	L	G	V	V	Y	K	BacA, BacC, LicC, LchAC	92%
Ile (2)	D	A	F	F	Y	G	I	T	F	K	FenB, PPS5	100%
Leu(1)	D	A	W	F	L	G	N	V	V	K	BacA, LicA, LchAA, LicB, LchAB, SrfAA, SrfAB	99%
Leu(2)	D	A	W	L	Y	G	A	V	M	K	CssA	100%
Leu(3)	D	G	A	Y	T	G	E	V	V	K	GrsB, TycC	100%
Leu(4)	D	A	F	M	L	G	M	V	F	K	LicA, LchAA, SrfAA	97%
Orn(1)	D	M	E	N	L	G	L	I	N	K	FxbC	100%
Orn(2)	D	V	G	E	I	G	S	I	D	K	BacB, FenC, GrsB, PPS1, TycC	98%
Phe	D	A	W	T	I	A	A	V	C	K	GrsA, SnbDE, TycA, TycB	88%
Phg/hPhg	D	I	F	L	L	G	L	L	C	K	CepB, CepC, SnbDE	80%
Pip/Pip-@	D	F	Q	L	L	G	V	A	V	K	FkbP, RapP, SnbA, SnbDE	75%
Pro	D	V	Q	L	I	A	H	V	V	K	GrsB, FenA, PPS4, SnbDE, TycB	87%
Ser	D	V	W	H	L	S	L	I	D	K	EntF, SyrE	90%
Thr/Dht	D	F	W	N	I	G	M	V	H	K	AcmB, Fxb, PPS2, PyoD, SnbC, SyrB, SyrE	91%
Tyr(1)	D	G	T	I	T	A	E	V	A	K	FenA, PPS2, PPS4	100%
Tyr(2)	D	A	L	V	T	G	A	V	V	K	TycB, TycC	80%
Tyr(3)	D	A	S	T	V	A	A	V	C	K	BacC, CepA, CepB	78%
Val(1)	D	A	F	W	I	G	G	T	F	K	GrsB, FenE, LicB, LchAB, PPS3, SrfAB, TycC	96%
Val(2)	D	F	E	S	T	A	A	V	Y	K	AcvA	94%
Val(3)	D	A	W	M	F	A	A	V	L	K	CssA	95%
Variability	3%	16%	16%	39%	52%	13%	26%	23%	26%	0%	Wobble-like positions	

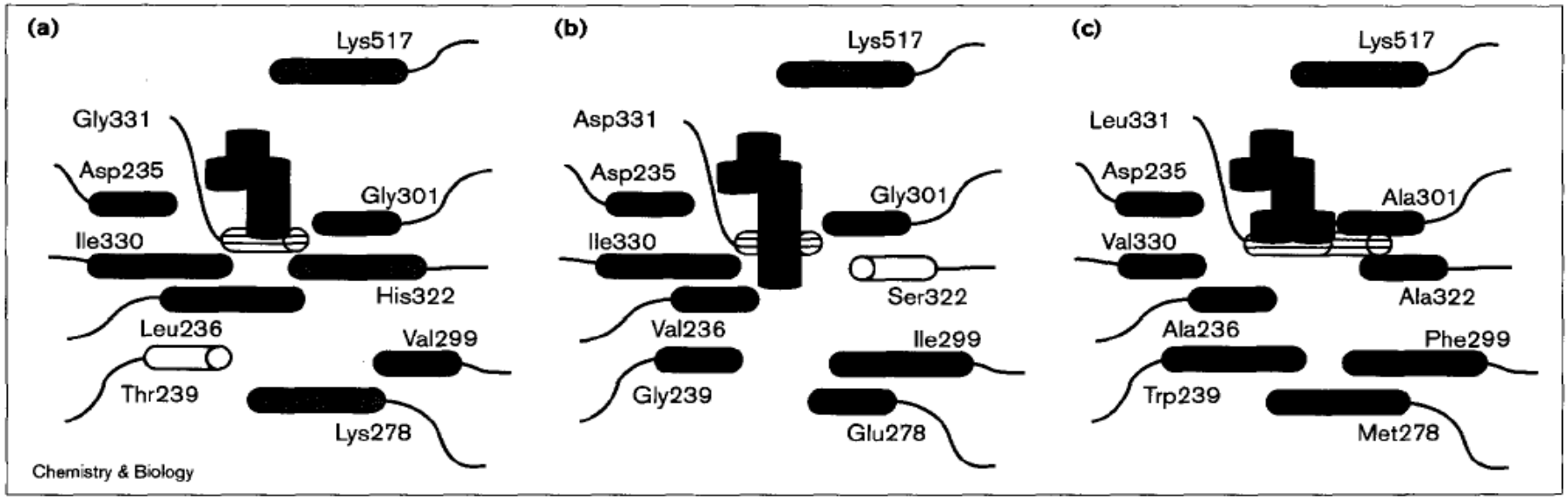
Sequence alignment of adenylation domains with known substrate binding specificity

Identification of residues present in positions corresponding to PheA binding site

Is there a correlation ?

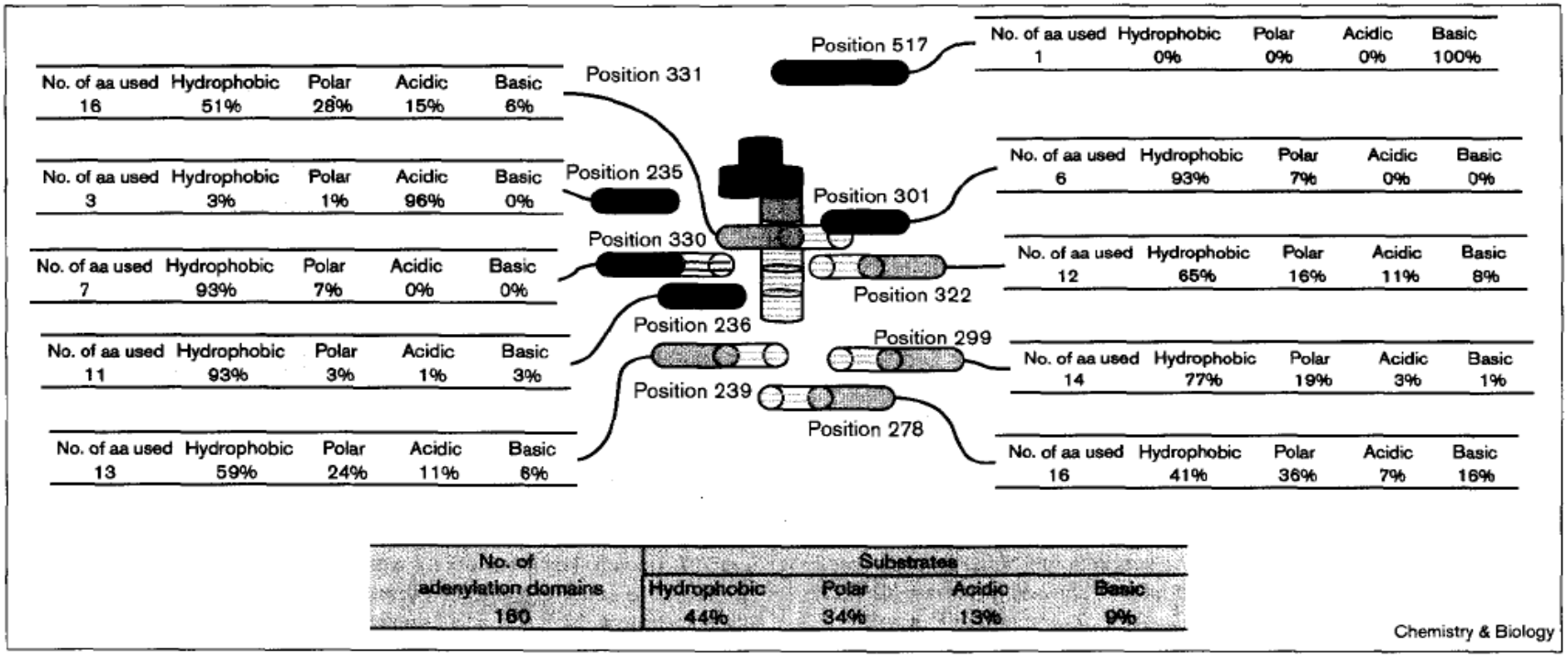
Substrate binding pockets in different adenylation domains

Aspartate Ornithine Valine



A simplified representation of the proposed binding pockets of three A domains. The putative binding pocket constituents of **(a)** an aspartate-activating domain (SrfAB2), **(b)** an ornithine-activating domain (GrsB3) and **(c)** a valine-activating domain (CssA9), determined in Figure 2, were projected onto the binding pocket of PheA shown in Figure 1b. The assessed aliphatic (blue), polar (yellow), acidic (red) and basic (green) sidechains are shown schematically. In all cases, Asp235 and Lys517 mediate key interactions with the α-amino and α-carboxylate group of the substrate, and all other residues facilitate recognition of the substrate sidechains and (ideally) complement the polarity of the recognized substrate.

The non-ribosomal code



No. of aa used	Hydrophobic	Polar	Acidic	Basic
16	51%	28%	15%	6%

No. of aa used	Hydrophobic	Polar	Acidic	Basic
3	3%	1%	96%	0%

No. of aa used	Hydrophobic	Polar	Acidic	Basic
7	93%	7%	0%	0%

No. of aa used	Hydrophobic	Polar	Acidic	Basic
11	93%	3%	1%	3%

No. of aa used	Hydrophobic	Polar	Acidic	Basic
13	59%	24%	11%	6%

No. of aa used	Hydrophobic	Polar	Acidic	Basic
1	0%	0%	0%	100%

No. of aa used	Hydrophobic	Polar	Acidic	Basic
6	93%	7%	0%	0%

No. of aa used	Hydrophobic	Polar	Acidic	Basic
12	65%	16%	11%	8%

No. of aa used	Hydrophobic	Polar	Acidic	Basic
14	77%	19%	3%	1%

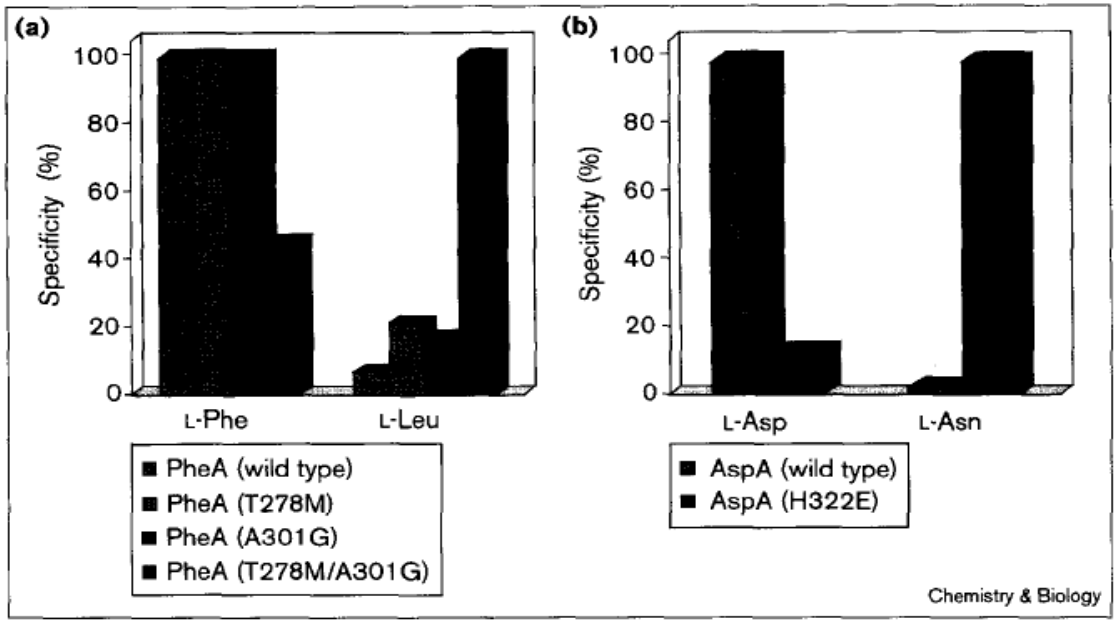
No. of aa used	Hydrophobic	Polar	Acidic	Basic
16	41%	36%	7%	16%

No. of adenylation domains	Substrates			
	Hydrophobic	Polar	Acidic	Basic
160	44%	34%	13%	9%

Observed variations of amino acids that constitute substrate-binding pockets. The proposed signature sequences from 160 different A domains were investigated (compare with Figure 2). The proportional distribution of the nature of their substrates is shown in the lower table. The yellow table linked to each position displays the number of different amino acids found in that position, and the proportional occurrence of hydrophobic, polar, acidic and basic sidechains. According to these data, the ten constituent amino acids can be classified into three subgroups. Positions 235 (aspartate: acidic, red) and 517 (lysine: basic, green) are considered 'invariant'. Positions 236, 301 and 330 are only 'moderately variant'. The vast majority (93%) of the A domains examined use hydrophobic sidechains in these positions (blue). 'Highly variant' are the residues at positions 239, 278, 299, 322 and 331 (gray), which reveal the highest variability of amino-acid usage.

Targeted alteration of substrate specificity of PheA and AspA

Targeted alteration of the substrate specificity of PheA and AspA. The specificity of wild-type (green) and mutant (red, orange and blue) proteins was investigated using the ATP–pyrophosphate exchange assay. The applied substrates are shown on the x axis, and the maximum value obtained for each protein was set to 100%. **(a)** In PheA, single substitutions towards the 'Leu(4)' codon (Thr278→Met and Ala301→Gly) modestly increased the specificity for leucine, although the preferred substrate was still phenylalanine. The corresponding double mutant, however, preferentially activated leucine with a catalytic efficiency approaching that of wild-type PheA. **(b)** A single His322→Glu mutation in AspA was sufficient to completely alter the substrate specificity of the mutant protein from aspartate to asparagine. The observed activation pattern of the mutants coincides with the appearance of their signature sequences in the phylogenetic tree shown in Figure 3 (green boxes).



Expression in E. coli of isolated A domain and ATP/PPi assay

Catalytic efficiency of wild-type and mutant enzymes.

Protein	Substrate	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m (mM^{-1} min^{-1})
PheA (wt)	L-Phe	0.9	62	69
	D-Phe	0.9	65	72
	L-Leu	nd	nd	nd
PheA (T278M/A301G)	L-Phe	5.2	58	11
	D-Phe	1.4	57	41
	L-Leu	0.7	60	86
Asp (wt)	L-Asp	2.9	29	10
	L-Asn	nd	nd	nd
AspA(H322E)	L-Asp	nd	nd	nd
	L-Asn	27.6	30	1

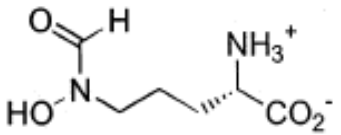
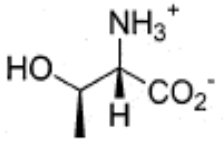
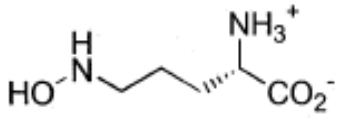
nd, not determined.

Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase

Gregory L. Challis [a,*], Jacques Ravel [b]

[a] *Department of Genetics, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK*
[b] *Department of Chemistry, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA*

Table 1
Prediction of CchH A-domain specificity determining residues and amino acid substrates

A-domain[a]	Residue (according to GrsA numbering)								Substrate
	235	236	239	278	299	301	322	330	
FxbB-M1-5hfOrn	D	I	N	Y	W	G	G	I	
CchH-M1-5hfOrn	D	I	N	Y	W	G	G	I	
Cda1-M2-Thr	D	F	W	N	V	G	M	V	
AcmB-M1-Thr	D	F	W	N	V	G	M	V	
SnbC-M1-Thr	D	F	W	N	V	G	M	V	
FenD-M2-Thr	D	F	W	N	I	G	M	V	
SyrB-M1-Thr	D	F	W	N	L	G	M	V	
SyrE-M7-Thr	D	F	W	N	V	G	M	V	
PvdD-M1-Thr	D	F	W	N	I	G	M	V	
PvdD-M2-Thr	D	F	W	N	I	G	M	V	
FxbC-M2-Thr	D	F	W	N	I	G	M	V	
CchH-M2-Thr	D	F	W	N	I	G	M	V	
FxbC-M1-5hOrn	D	M	E	N	L	G	L	I	
FxbC-M3-5hOrn	D	M	E	N	L	G	L	I	
CchH-M3-5hOrn	D	M	E	N	L	G	L	I	

[a] Nomenclature is as follows: protein name-module number-amino acid substrate (5hfOrn: 5-hydroxy-5-formylornithine; 5hOrn: 5-hydroxyornithine). The Genbank accession numbers are as in [7].

Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase

Gregory L. Challis [a,*], Jacques Ravel [b]

[a] *Department of Genetics, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK*
[b] *Department of Chemistry, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA*

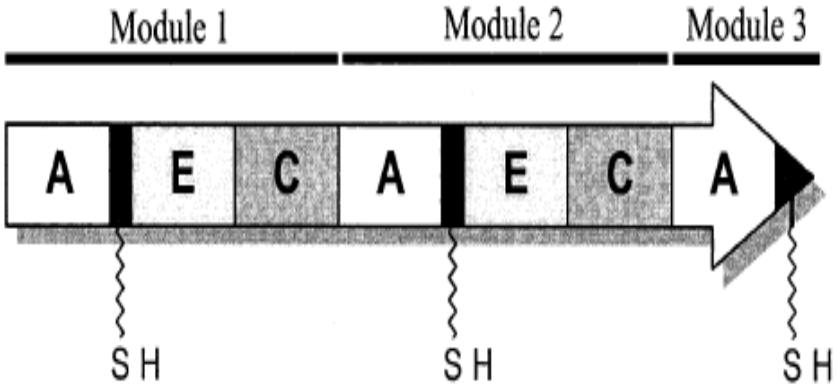
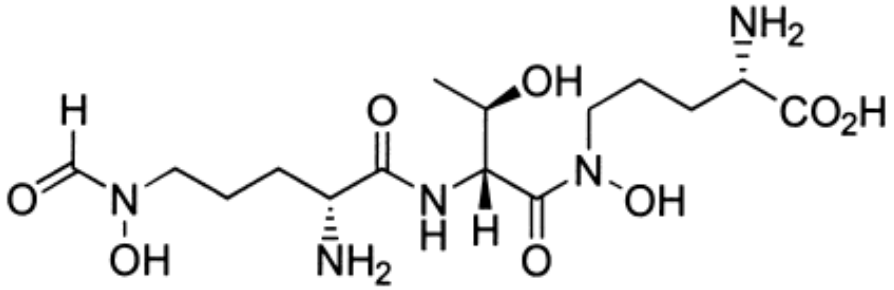
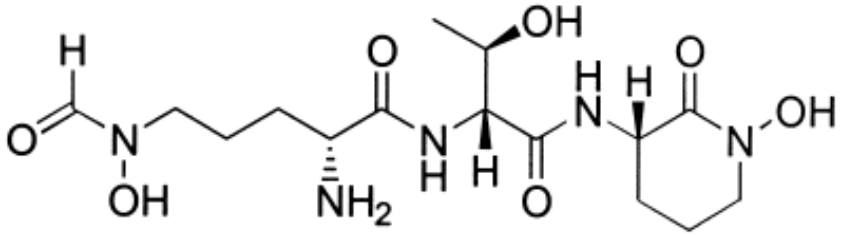


Fig. 1. Domain and module organization of the NRPS (CchH) encoded on cosmid SCF-34 of the ordered *S. coelicolor* genomic library. Domains are indicated as follows: adenylation: A, white; PCP: black, ~SH; epimerization: E, light gray; and condensation: C, dark gray.

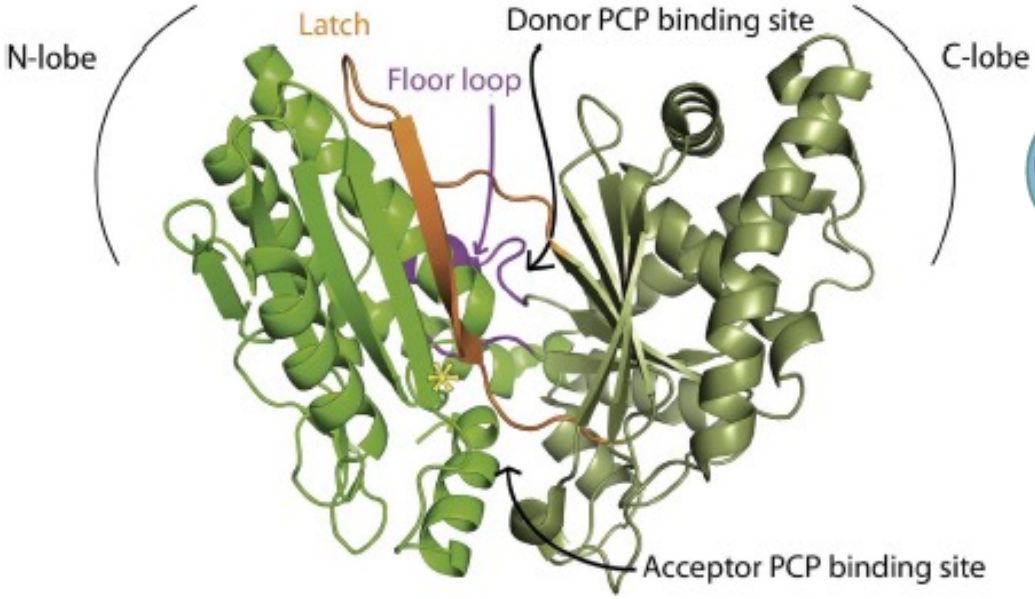


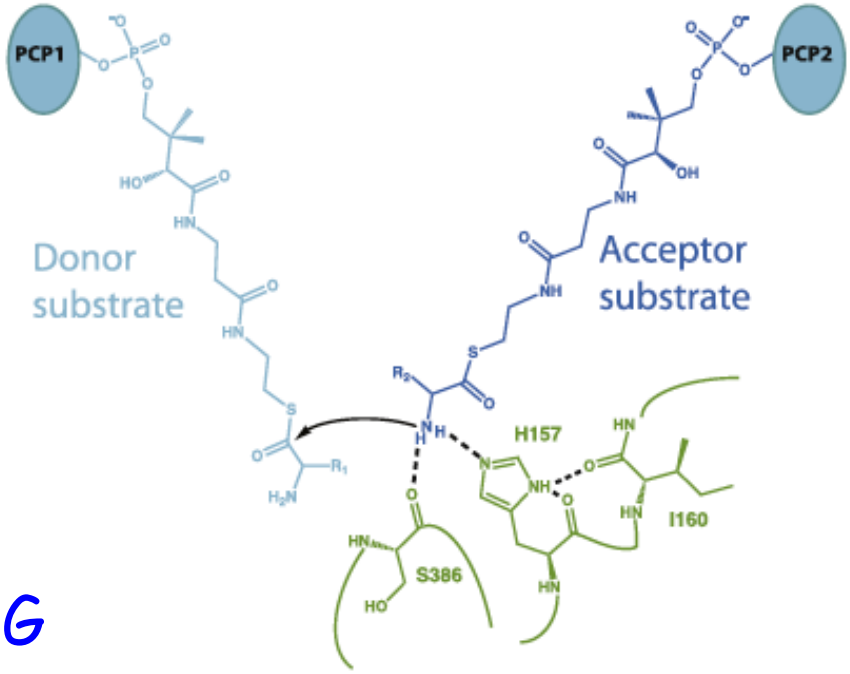
1



2

The condensation domain





Consensus sequence HHxxxDG

Role of His147 in the condensation domain

To confirm the role of the condensation domain an NRPS formed by **two** modules: PheATE and ProCAT. The two modules have been expressed in *E. coli* and **activated** by 4'-PP transferase sfp.
Dipeptide synthesis was studied by supplying labeled Phe and Pro.

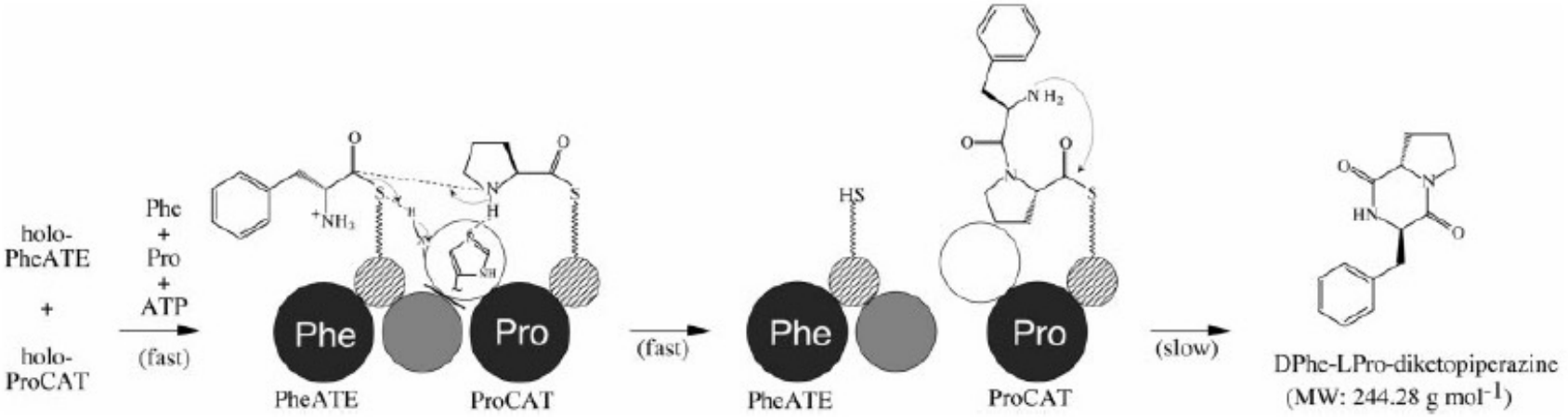


FIG. 7. **Reaction sequence of D-Phe-L-Pro and diketopiperazine formation.** A proposed simple model for peptide bond and diketopiperazine formation is shown. The single domains of the modules PheATE and ProCAT are illustrated as *black* (adenylation), *striped* (thiolation), *gray* (epimerization), and *white* (condensation) *circles*. PheATE and ProCAT are loaded with their cognate amino acids. His^{147} of ProCAT, a strictly conserved residue of the condensation domain, acts as a base to promote the nucleophilic attack of the imino group of L-Pro onto the carboxy carbon atom of D-Phe. Another residue should reprotonate the thiol moiety of the liberated P-pant. The *black line* indicates protein-protein interactions between ProCAT and the C-terminal region of PheATE. In a second step, diketopiperazine formation is believed to occur by an intramolecular, noncatalyzed cyclization reaction.

Role of His147 in the condensation domain

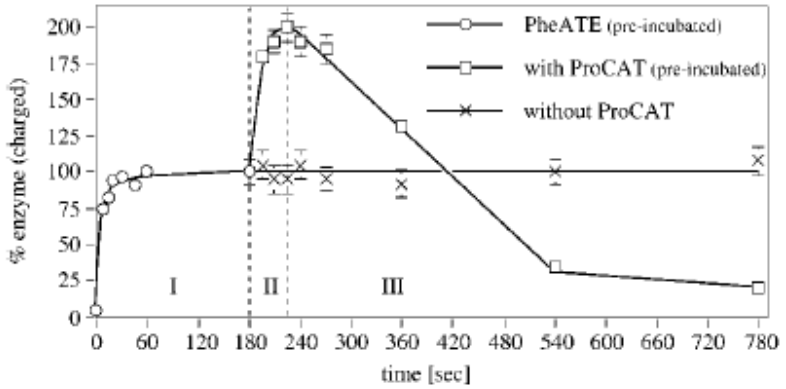
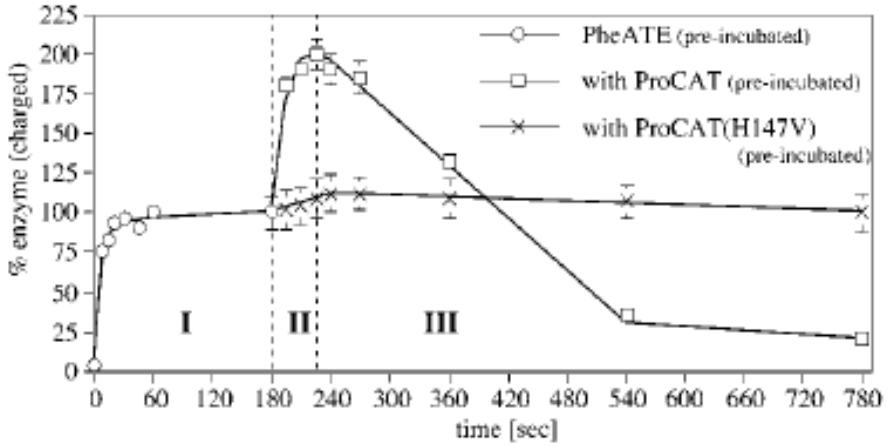
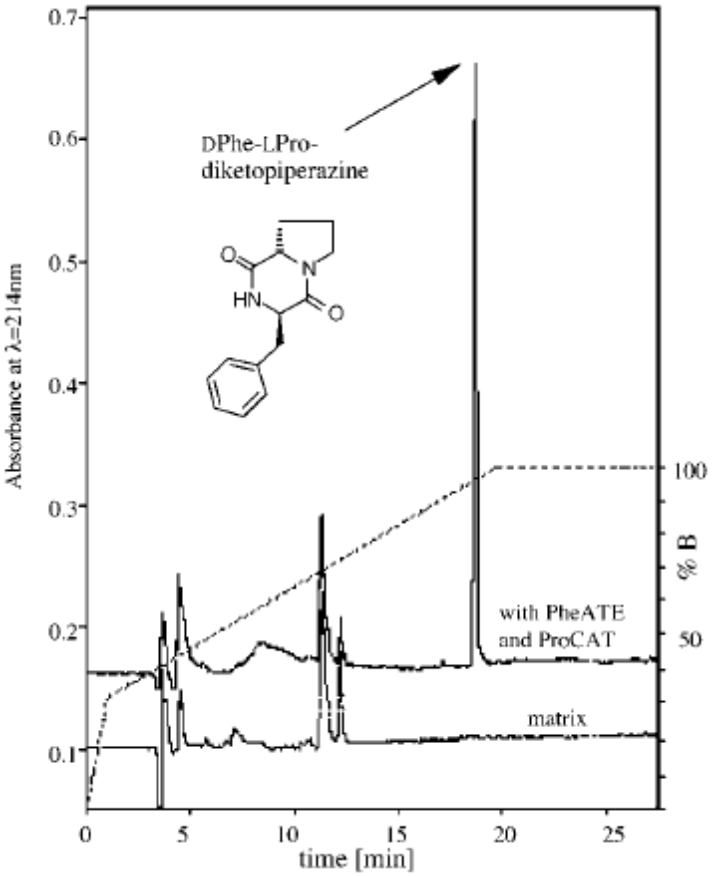


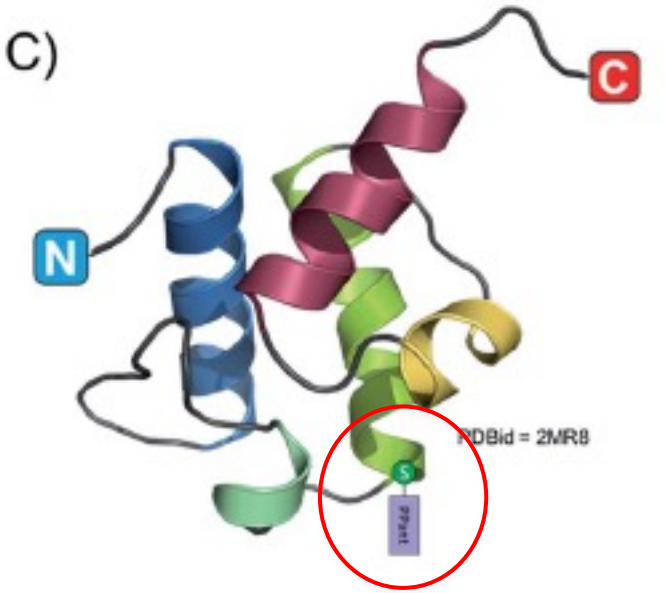
FIG. 4. **Dipeptide formation is catalyzed by PheATE and ProCAT.** PheATE and ProCAT were separately incubated with their cognate amino acids (^{3}H-labeled L-Phe and L-Pro) in order to achieve a complete charging of both enzymes (stage I). At t = 180 s, equivalent amounts of both enzymes were combined. Dipeptide formation was monitored by taking samples at various time points, precipitation by the addition of 10% trichloroacetic acid (w/v), and quantification of the acid-stable label by LSC. Stage II corresponds to the formation of the D-Phe-L-Pro dipeptide bound to ProCAT. Stage III is explained with the cleavage of the cyclic D-Phe-L-Pro-diketopiperazine.



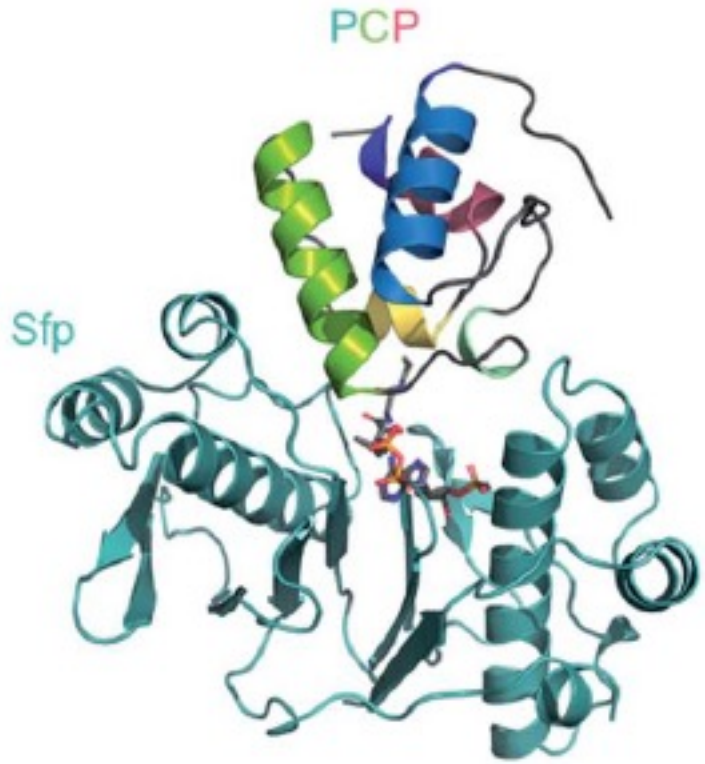


The Phe-Pro dipeptide does not form in the H147V mutant.

The thiolation domain



Structure of the T (PCP) domain
The active site serine and the 4-PP cofactor are solvent-exposed



Structure of the Sfp-PCP complex

Activation of the thiolation domain

- 4'-PP transferases have low substrate specificity and can activate thiolation domains with acyl-CoA, that are normally present in the cell
- This process can block NRPS
- Thiolation domains can be reactivated by thioesterases that catalyze hydrolysis of undesired acyl groups and could act as selectivity filters between correct and uncorrect amino acids *in vivo*

- ***In vitro*** it is possible to activate the thiolation domain with aminoacyl-CoA or functionalized derivatives of CoA

B. subtilis 4-phosphopantetheine transferase Sfp has low substrate specificity and can utilize CoA analogues

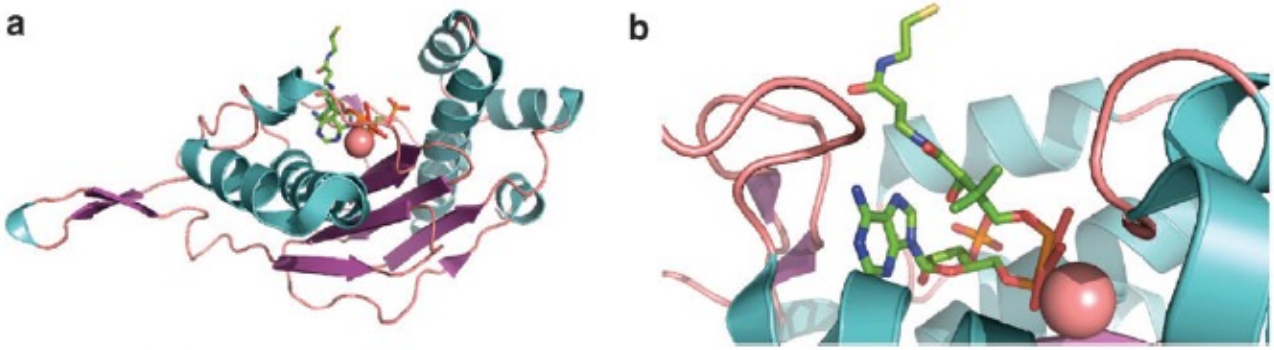
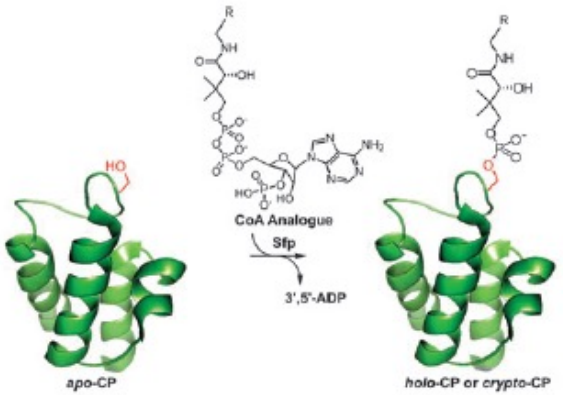


Fig. 6 Crystal structure of Sfp, the promiscuous PPTase from the *B. subtilis* surfactin biosynthetic pathway (PDB:1QR0). (a) Structure of Sfp with CoA substrate and Mg^{2+} cofactor bound in active site. (b) Close-up of CoA bound to Sfp. The terminal β-alanine and cystamine portions of CoA extend into the solvent and make no clear interactions with the enzyme, helping explain the permissivity of Sfp for CoA analogues modified at the thiol-terminus.



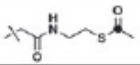
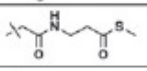
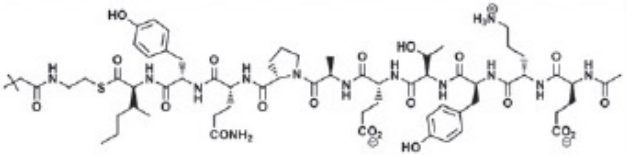
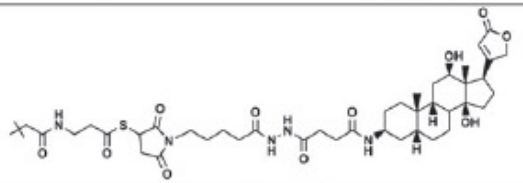
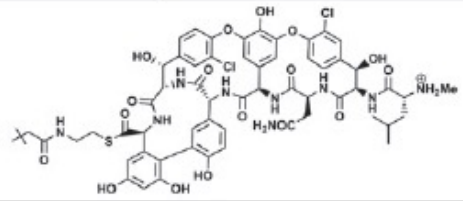
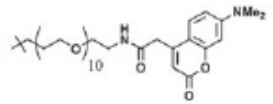
CoA Analogue (R =)	Mol. weight	Ref.
5 	727	46
6 	727	48
7 	623	75
8 	1948	76
9 	1425	77
10 	1810	78
11 	1338	75

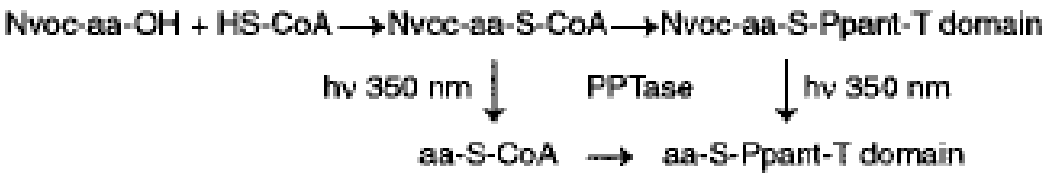
Fig. 7 Notable examples of CoA analogues transferred to carrier proteins using Sfp.

Selectivity of condensation domains

To evaluate condensation domain selectivity it is necessary to overcome substrate specificity control exerted by adenylation domains.

The thiolation domain is activated *in vitro* by sfp with aminoacyl-CoA

Bimodular NRPS PheATE and ProCAT: the two modules can be charged independently



The Nvoc group blocks the amino acid NH_2 group to stabilize aminoacyl-CoA without interfering with binding by sfp, and it can be easily removed by UV exposure.

Selectivity of condensation domains

Selectivity of the donor site (electrophilic)

PheATE + aa-CoA L,D-Phe L,D-Ala L-Leu

ProCAT + [^{3}H]Pro

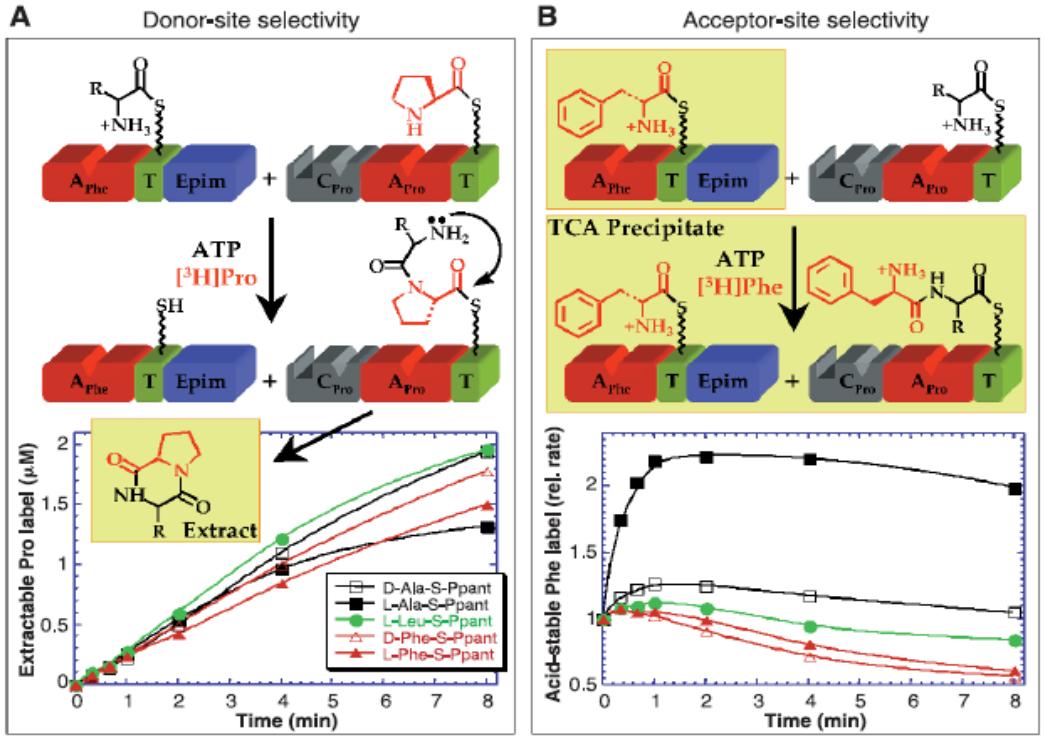
Low selectivity: the product is formed with all amino acids

Selectivity of the acceptor site (nucleophilic)

PheATE + [^{3}H]Phe

ProCAT + aa-CoA

High selectivity: the product is formed only with L-Ala



Selectivity of condensation domains

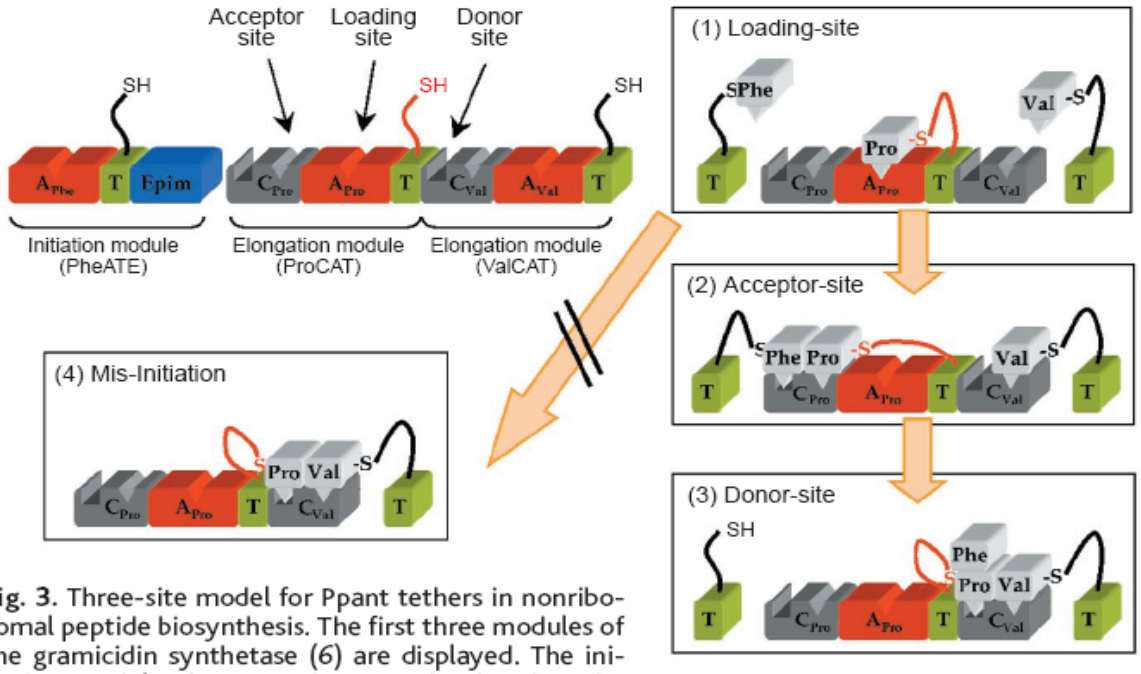


Fig. 3. Three-site model for Ppant tethers in nonribosomal peptide biosynthesis. The first three modules of the gramicidin synthetase (*6*) are displayed. The initiation module PheATE activates and epimerizes the first amino acid (Phe) as a separate subunit from the modules that activate the second (ProCAT) and third (ValCAT) amino acids. To account for the multiple tasks of acceptor recognition, donor recognition, and chain translocation, the Ppant arm of a T domain is postulated to adopt three distinct configurations (*20*), detailed here for the T domain of ProCAT. In the loading site (panel 1), the free thiol of Ppant is acylated by Pro-O-AMP that was formed by the adjacent A_{Pro} domain. The Pro-S-Ppant arm must reach upstream to the acceptor site (panel 2) on the C_{Pro} domain and wait for the donor Ppant from the upstream T domain (here Phe-S-Ppant-T) to fill the donor site. At this juncture the C domain catalyzes peptide bond formation and the dipeptidyl (D-Phe-L-Pro)-S-Ppant arm can exit the C_{Pro} domain and swing downstream to the donor site (panel 3) in the C_{Val} domain. For an orderly progression, to avoid incomplete chains and chain misinitiation at internal sites (panel 4), there must be carefully orchestrated kinetic control of occupancy of the C domain donor and acceptor sites for the Ppant arms of every T domain. The PheATE initiation module has no C domain to trap its aminoacylated T domain and presumably can only occupy the donor site on the C_{Pro} domain to initiate chain growth.

Selectivity of condensation domains: epimerization precedes condensation

The condensation domain is stereospecific

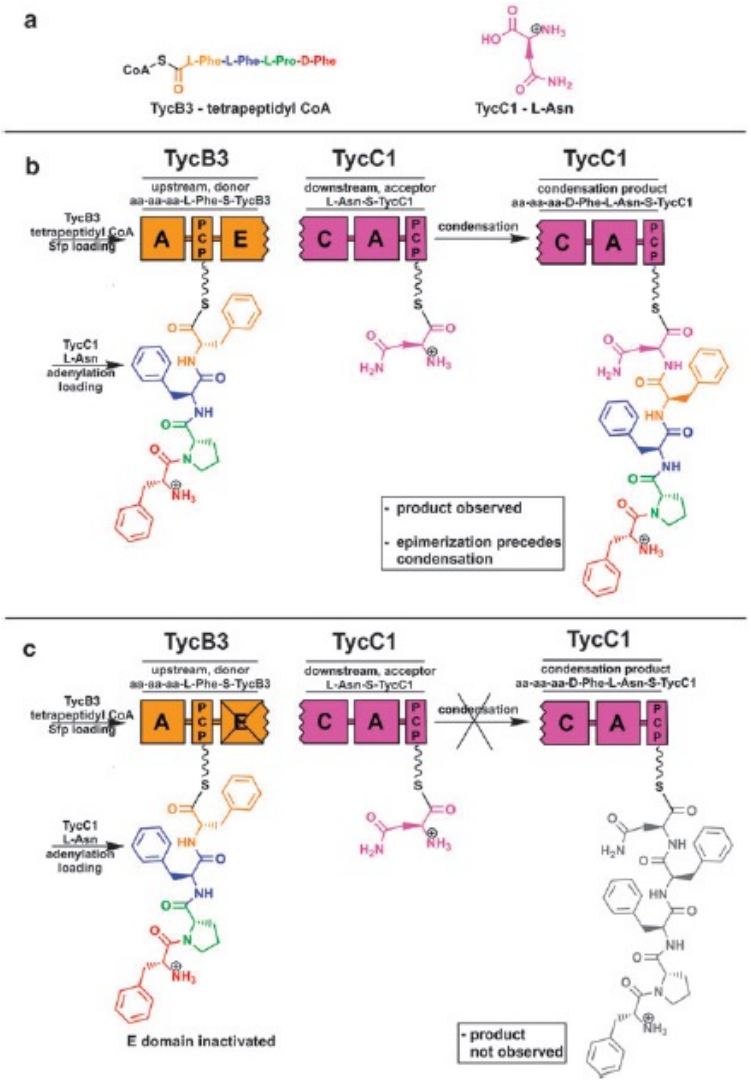
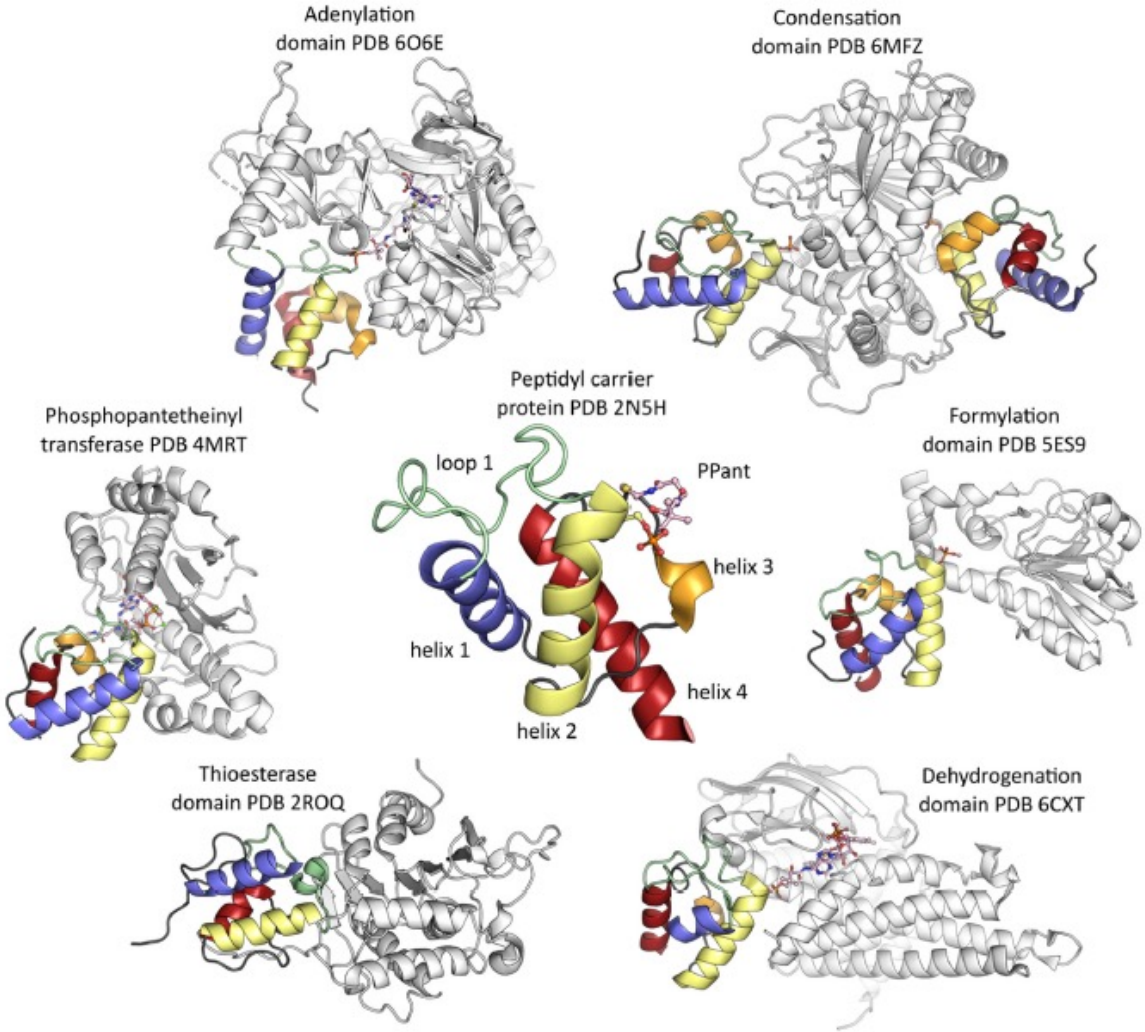
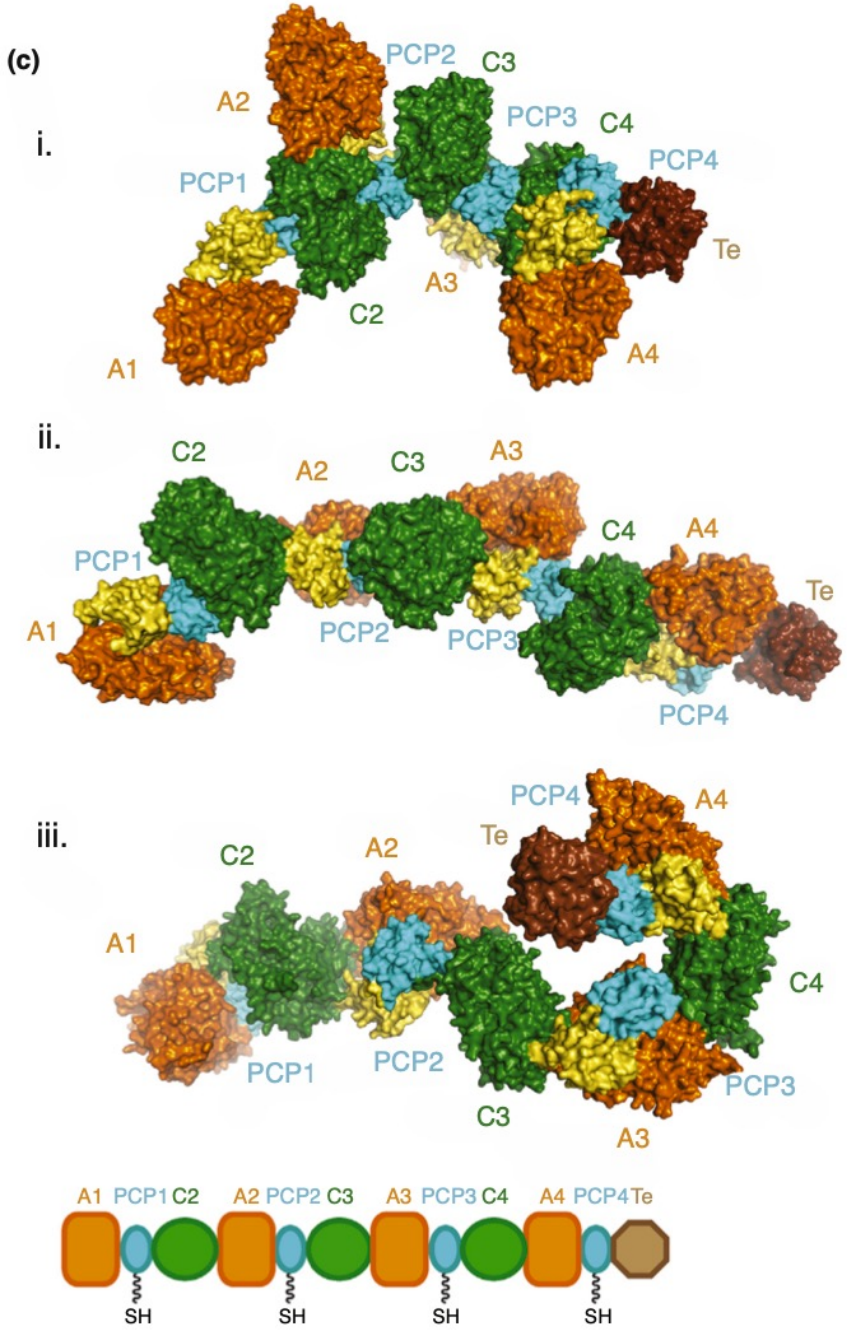


Fig. 9 Peptidyl CoAs as probes of NRPS condensation mechanism. (a) Tetrapeptidyl CoA analogues of the natural TycB4 chain elongation intermediate and L-Asn, the natural substrate for the A domain of TycC1. (b) Upon loading of TycB4 with a L-Phe tetrapeptidyl CoA, and TycC1 with Asn by auto-adenylation, only the epimerized pentapeptide product is observed, implying epimerization precedes condensation in NRPS biosynthesis. (c) A similar experiment using a TycB4 E domain mutant shows no formation of the pentapeptide, implying the non-epimerized peptide is not a substrate for the TycC1 C domain.

The central role of the thiolation domain



Structural models of the module and domain organization of NRPS

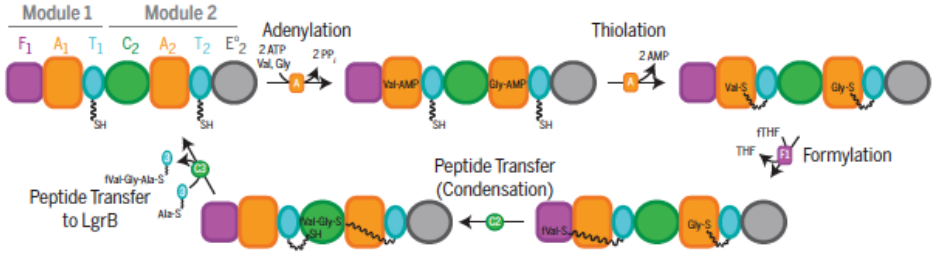


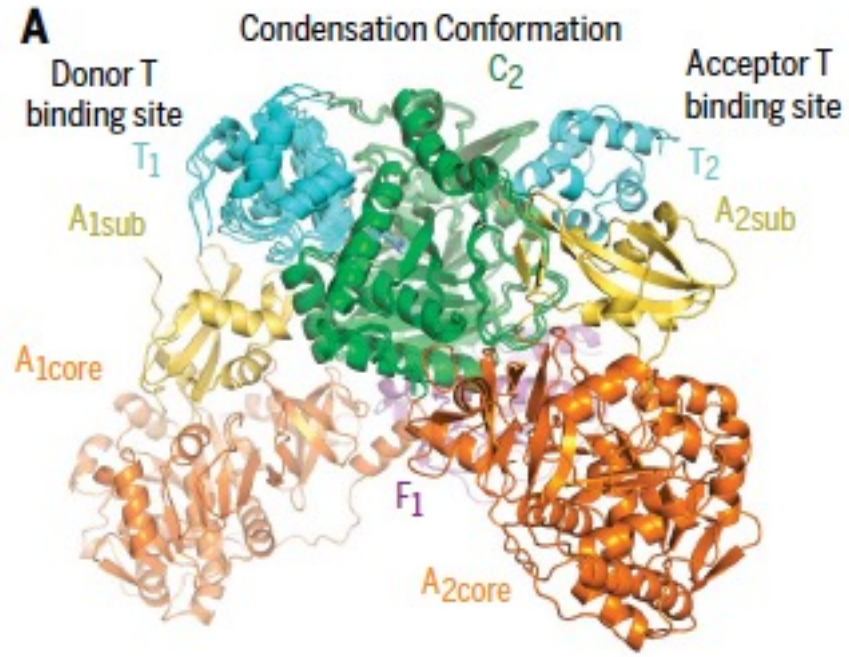
Current Opinion in Structural Biology

BIOSYNTHETIC ENZYMES

Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility

Janice M. Reimer[1]*, Maximilian Eivaskhani[1]*, Ingrid Harb[1], Alba Guarné[1], Martin Weigt[2], T. Martin Schmeing[1]†





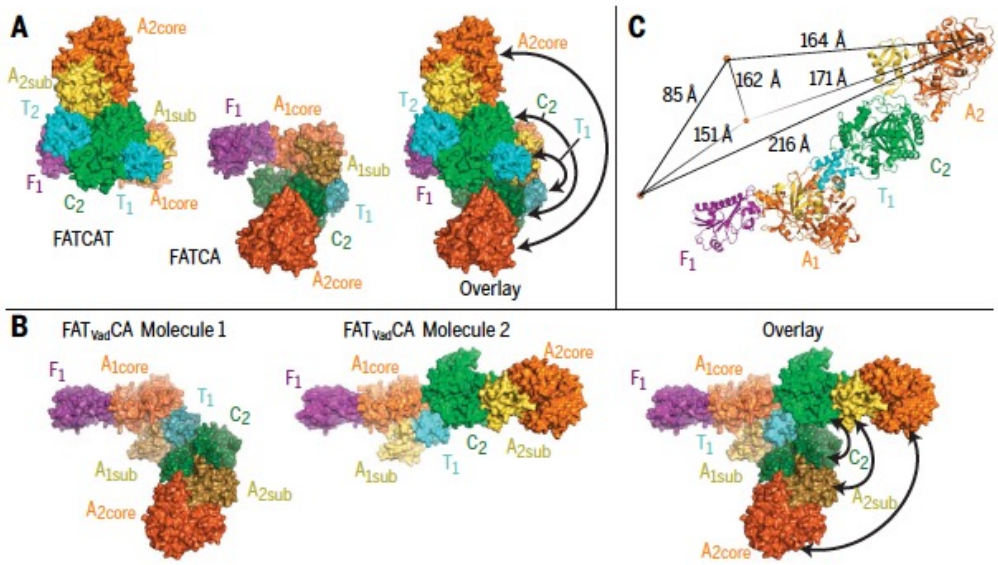
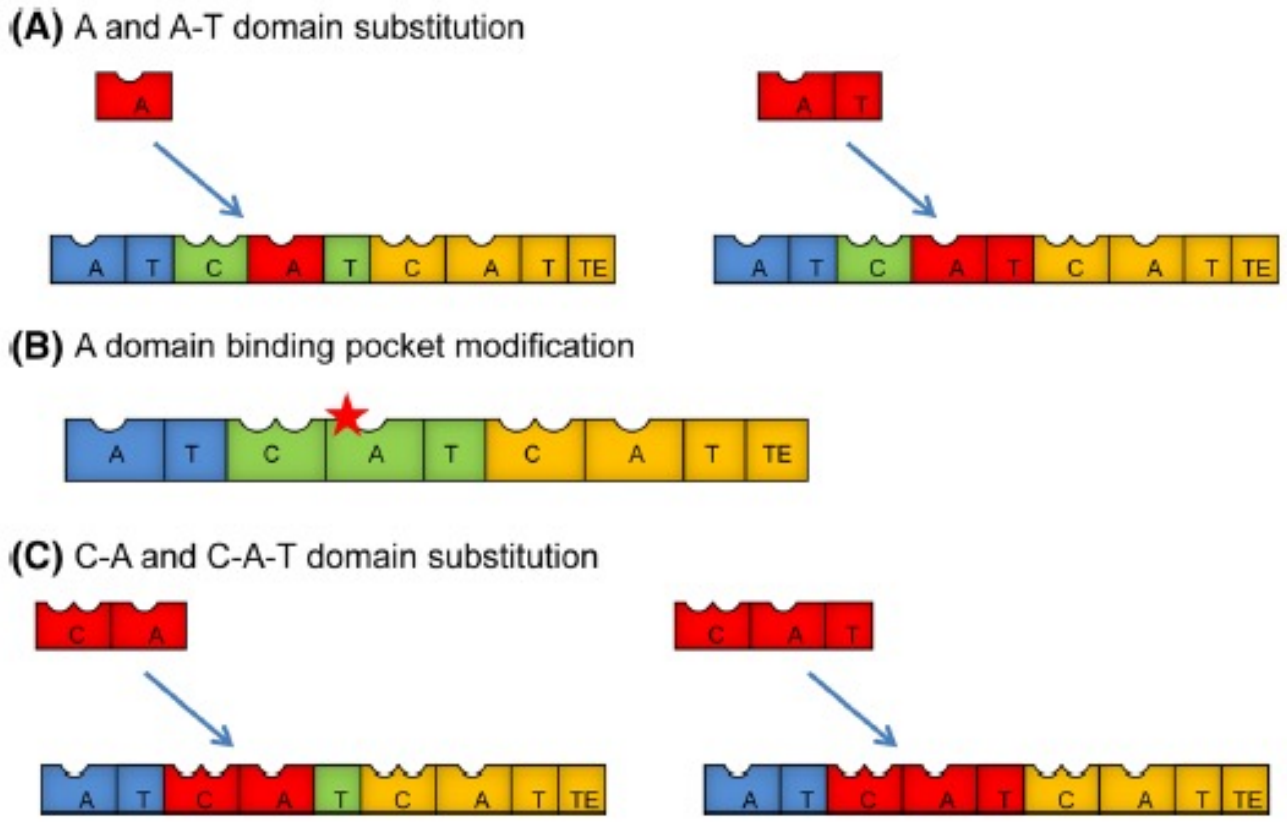


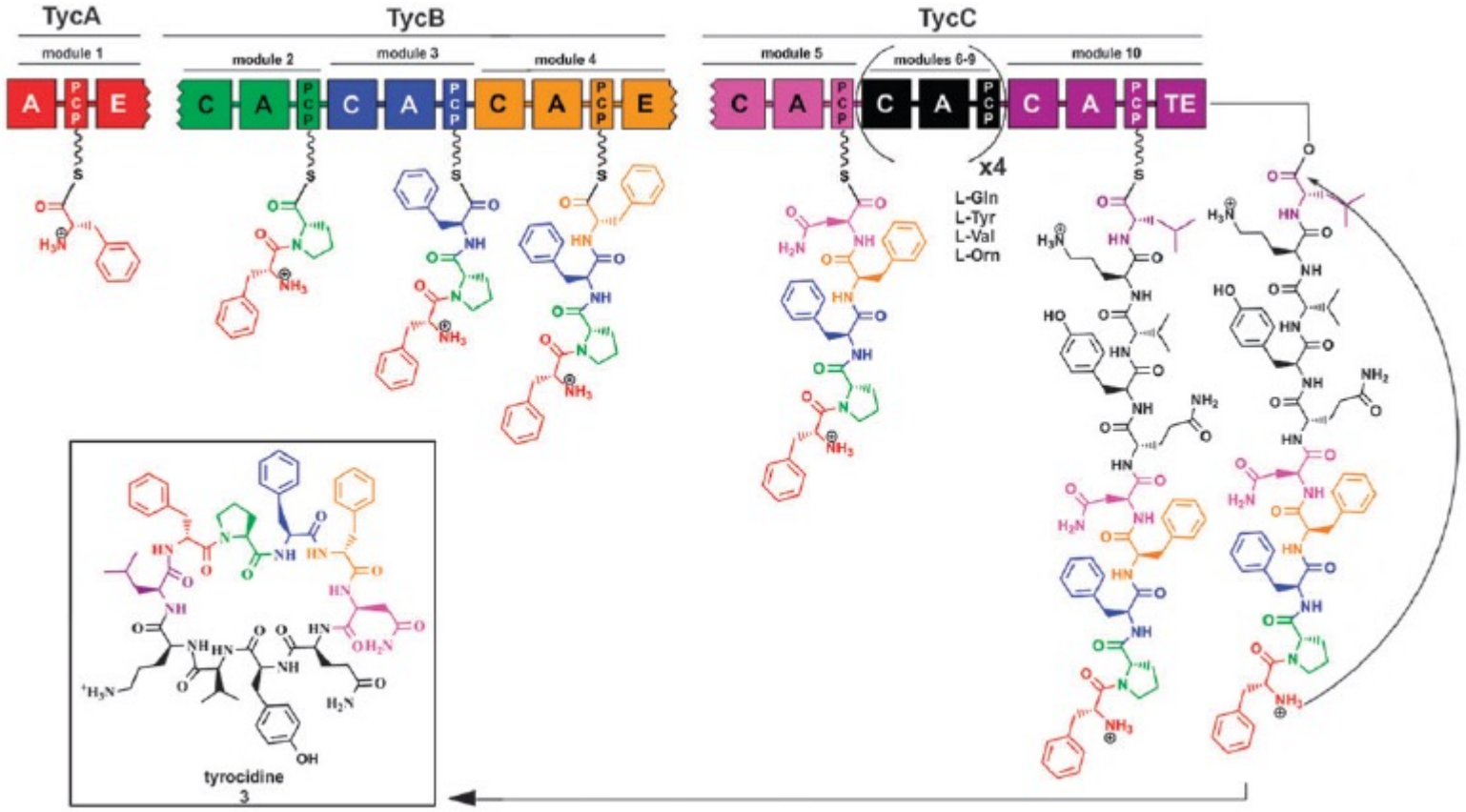
Fig. 5. Different dimodular conformations for the same catalytic states. (**A**) FATCAT and FATCA both show T_1 binding to the donor site of C_2 but have very different overall conformations. (**B**) The two crystallographically independent molecules of $FAT_{Vad}CA$ both show module 1 in thiolation conformation but have very different positions of module 2. (**C**) The distances between positions of residue Asp^{1236} in the four dimodular conformations. The structures are superimposed by their A_{1core}.

Strategies for NRPS reprogramming



Strategies for NRPS reprogramming *In cis* and *in trans* module and domain communication

Tyrocidine synthetase TycA (loading module), TycB (3 modules) and TycC (6 modules and Te domain for cyclization and release).



Tyrocidine synthetase:construction of hybrid NRPS by module fusion

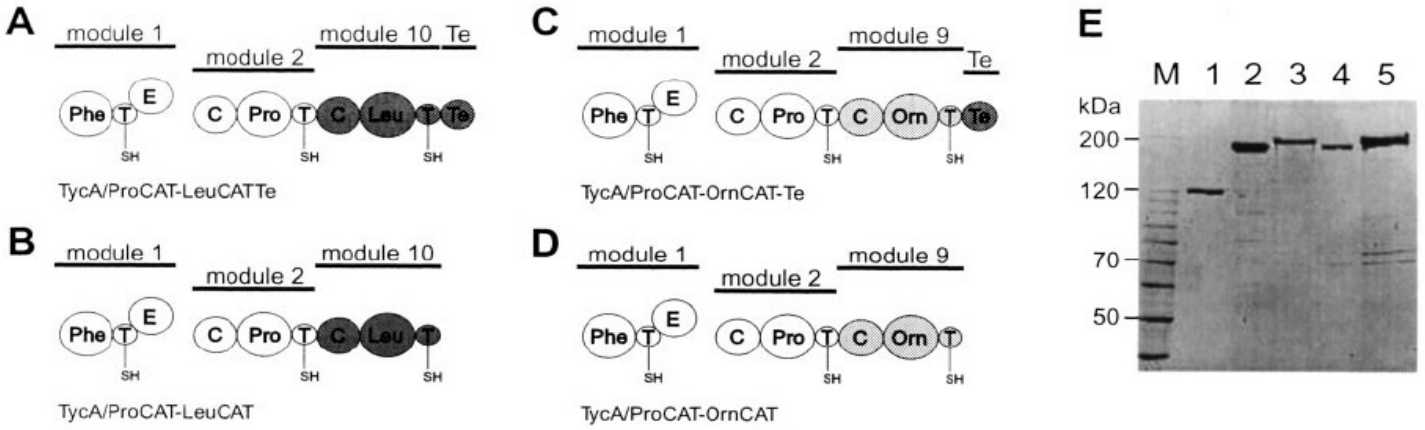


Fig. 3. Construction of artificial trimodular NRPS systems. Artificial trimodular NRPS systems were obtained by extension of the first two modules of tyrocidine synthetases, TycA and TycB1 (here referred to as ProCAT). Fusion of the last module of TycC, containing the Te-domain, to ProCAT yielded TycA/ProCAT-LeuCATTe (*A*). TycA/ProCAT-LeuCAT is devoid of the Te-domain (*B*). Fusion of the second last module of TycC, OrnCAT, resulted in the system TycA/ProCAT-OrnCAT (*D*), which was extended in a second fusion step to TycA/ProCAT-OrnCAT-Te (*C*). The purified His_6-tagged enzymes are shown on a Coomassie blue-stained SDS/PAGE (7.5% polyacrylamide) in *E*. M = marker. Lanes 1–5: TycA, ProCAT-OrnCAT, ProCAT-OrnCAT-Te, ProCAT-LeuCAT, and ProCAT-LeuCATTe, respectively. The gel could not sufficiently resolve the dimodular enzymes, which are 234–265 kDa in size.

Analysis of sequences between T-C and T-Te: identification of linker regions with low conservation employed to construct fusion proteins

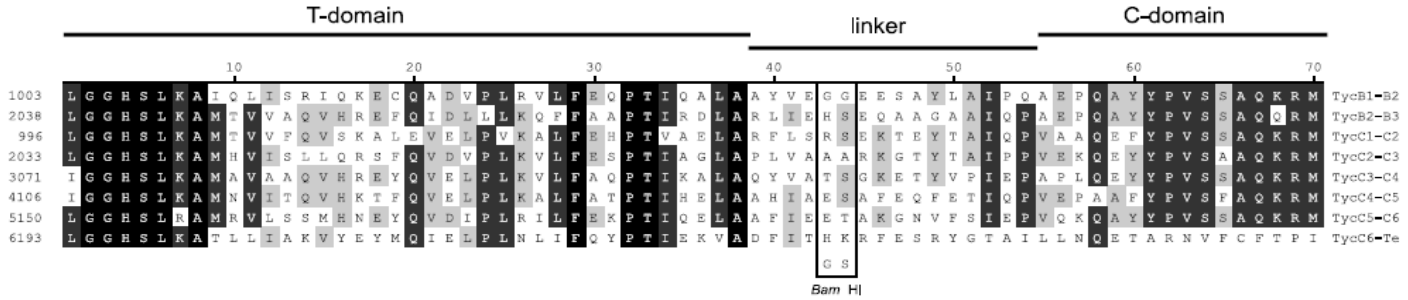


Fig. 2. Linker sequence between T- and C-domains. An alignment of the seven sequences in tyrocidine synthetases TycB and TycC that connect T- and C-domains, as well as the terminal T- and the Te-domains, reveals a stretch with very little conservation. This stretch was defined as linker between the domains. All fusions were performed at the site indicated by introducing a *Bam*HI restriction site in the corresponding gene fragments, which itself codes for the amino acids glycine and serine.

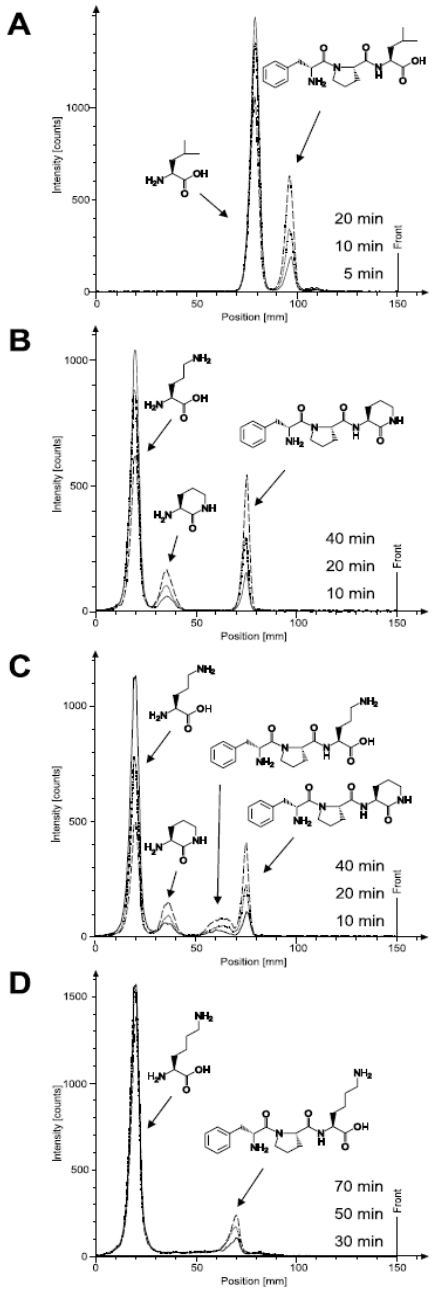


Fig. 4. Formation of the predicted peptides. The product pattern of the trimodular NRPS systems was monitored by using Phe, Pro, and the third amino acid in radiolabeled form, TLC separation of the products and subsequent radioactivity scanning. (*A*) TycA/ProCAT-LeuCATTe with [^{14}C]-Leu. (*B*) TycA/ProCAT-OrnCAT with [^{14}C]-Orn. (*C*) TycA/ProCAT-OrnCAT-Te with [^{14}C]-Orn and (*D*) [^{14}C]-Lys.

Tyrocidine synthetase

Catalytic efficiency is similar to wild type NRPS: communication is retained and C and Te domains are able to accept different substrates.

Table 1. Electrospray ionization-MS analysis of the tripeptide products

Product	Mass calculated $[M + H]^+$	Mass found $[M + H]^+$
	TycA/ProCAT-LeuCATTe	
DPhe-Pro-Leu	376	376
DPhe-Pro-Ile	376	376
DPhe-Pro-NVal	362	362
DPhe-Sar-Leu	350	350
DPhe-Abu-Leu	364	364
DTrp-Pro-Leu	415	415
	TycA/ProCAT-OrnCAT	
DPhe-Pro-Orn_{cyc}	359	359
	TycA/ProCAT-OrnCAT-Te	
DPhe-Pro-Orn_{cyc}	359	359
DPhe-Pro-Orn	377	377
DPhe-Pro-Lys	391	391

Strategies for NRPS reprogramming Comunication between modules and domains *in trans*

Identification of sequences at the extremes of E domains (C-terminal) and C domains (N-terminal) defined COM^D and COM^A.

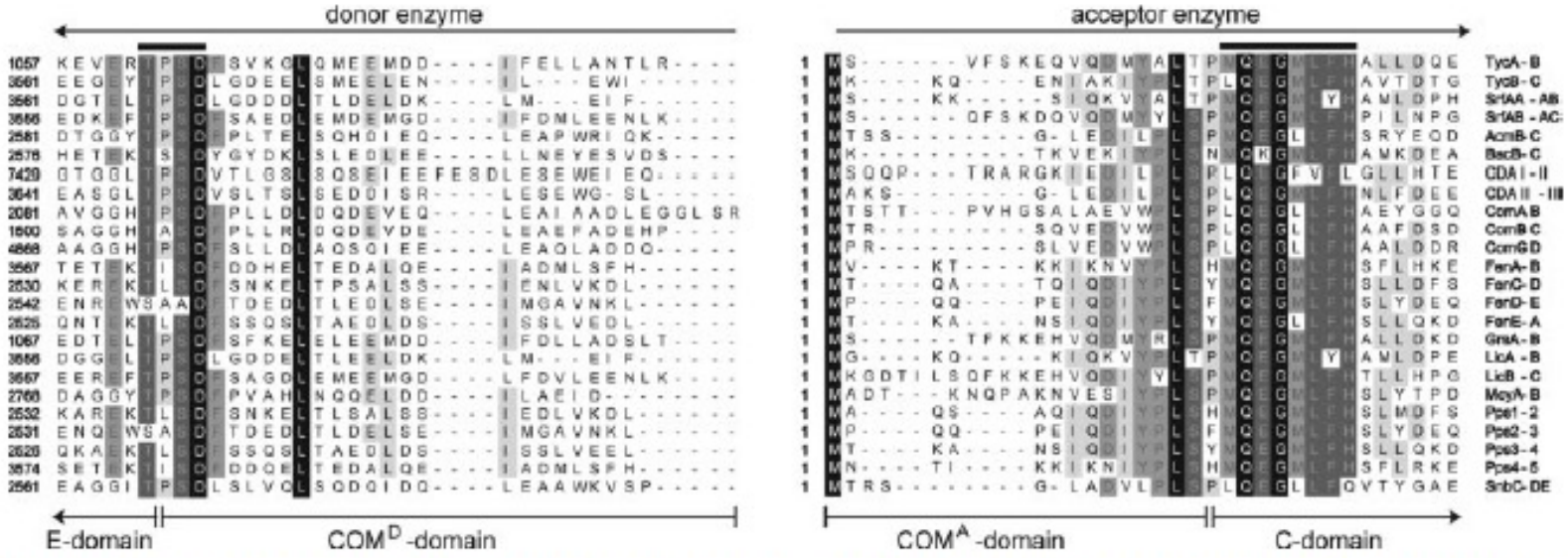


Fig. 2. Sequence comparison of proposed COM^D and COM^A domains derived from 12 biosynthetic systems: actinomycin (Acm), bacitracin (Bac), complestatin (Com), calcium-dependent antibiotic (CDA), fengycin (Fen), Grs, lichenysin (Lic), microcystin (Mcy), pliplastatin (Pps), pristinamycin (Snb), SrfA, and Tyc. Invariant residues are shown in black and conserved residues in gray.

'The Universal Communication System'

Hybrid trimodular system formed by TycA, BacB2 and SrfA engineered to have the same COM^D and COM^A domains.

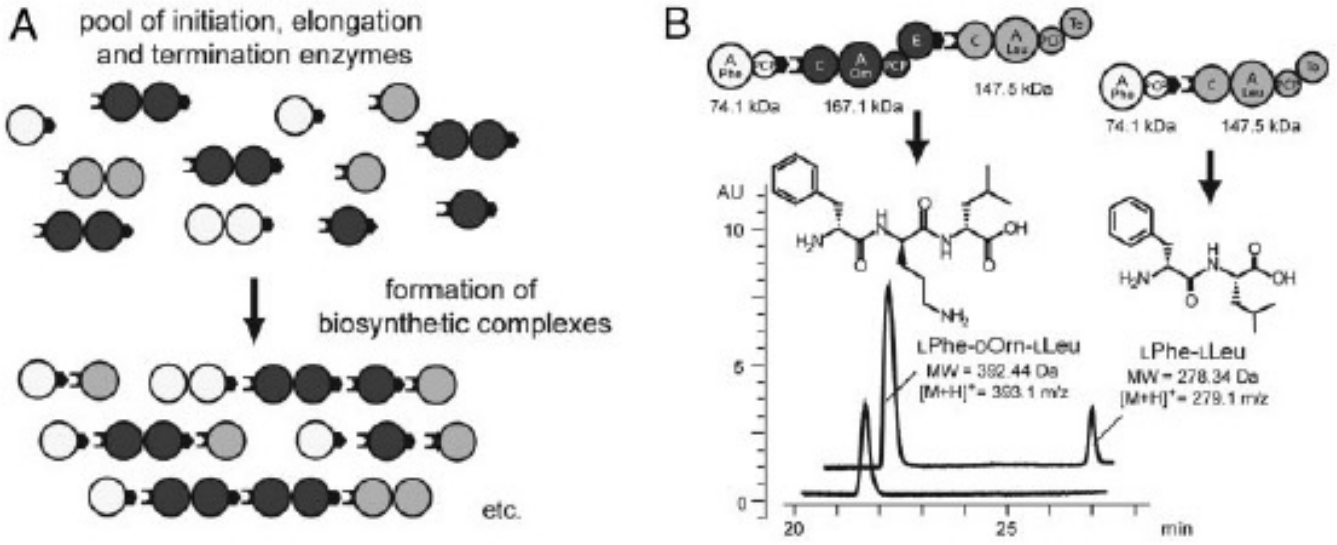
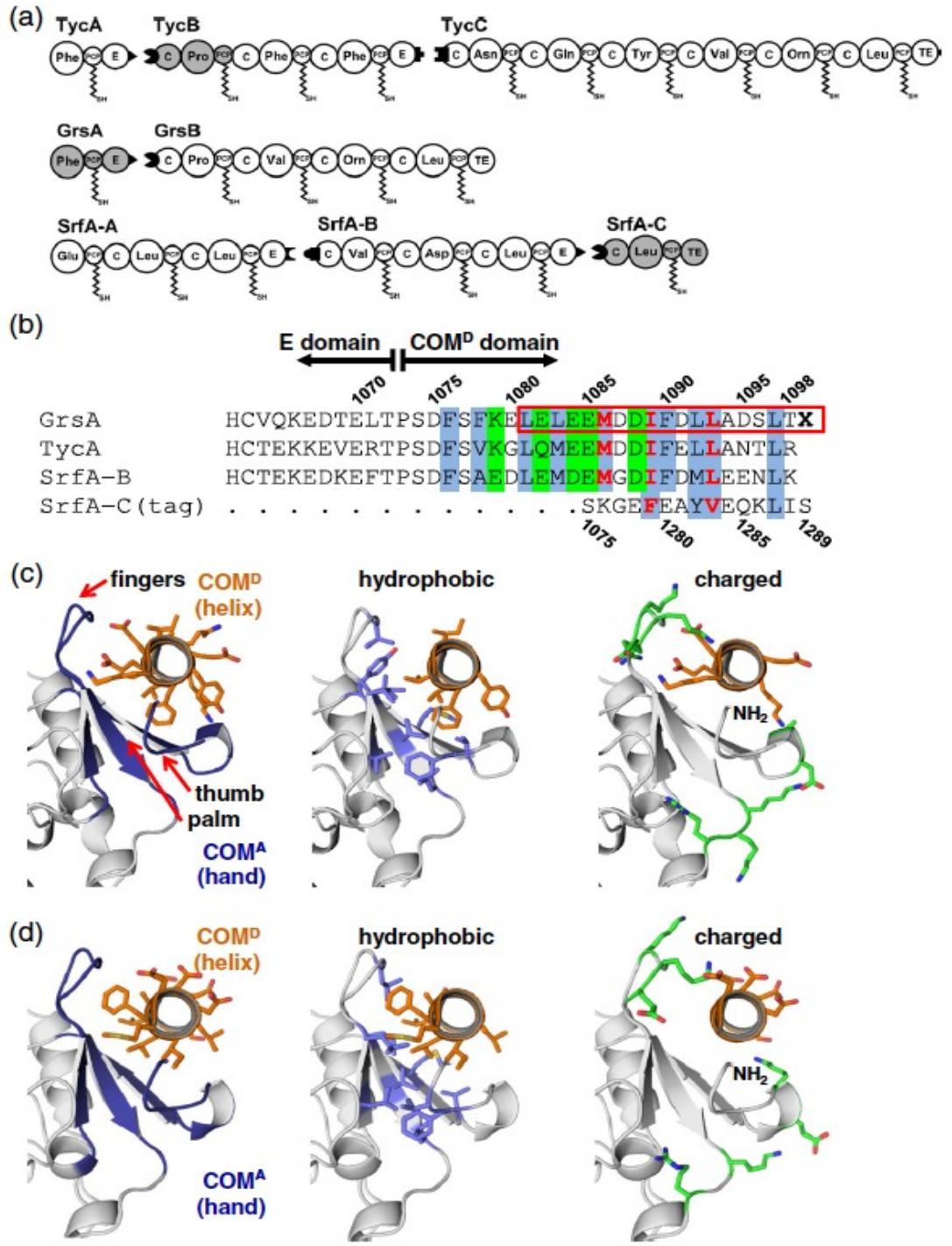


Fig. 5. The UCS. (*A*) General principle of UCS. Note, all protein–protein interactions are mediated by the same set of COM domains. (*B*) Possible NRPS complexes formed by the initiation enzyme TycA::$COM^D\Delta E$, the elongation enzyme $COM^A{}_{TycB1}$-(C-A_{Orn}-PCP-E)$_{BacB2}$-$COM^D{}_{TycA}$-His, and the termination enzyme SrfAC. HPLC/MS confirmed the simultaneous formation of the expected tripeptide LPhe-DOrn-LLeu ($[M+H]^+ = 393.1\ m/z$, expected mass, 393 m/z; retention time: 21.8 min), and the dipeptide LPhe-LLeu ($[M+H]^+ = 279.1\ m/z$, expected mass: 279 m/z, retention time: 26.1 min).

Communication between modules *in trans*

Structural analysis of COM^D and COM^A regions.

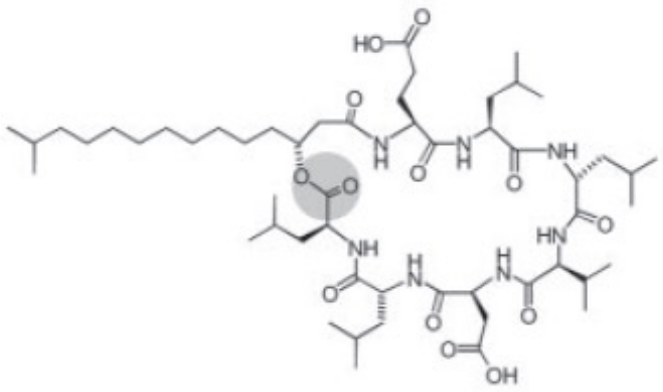
Charged residues and hydrophobic residues contribute to binding specificity



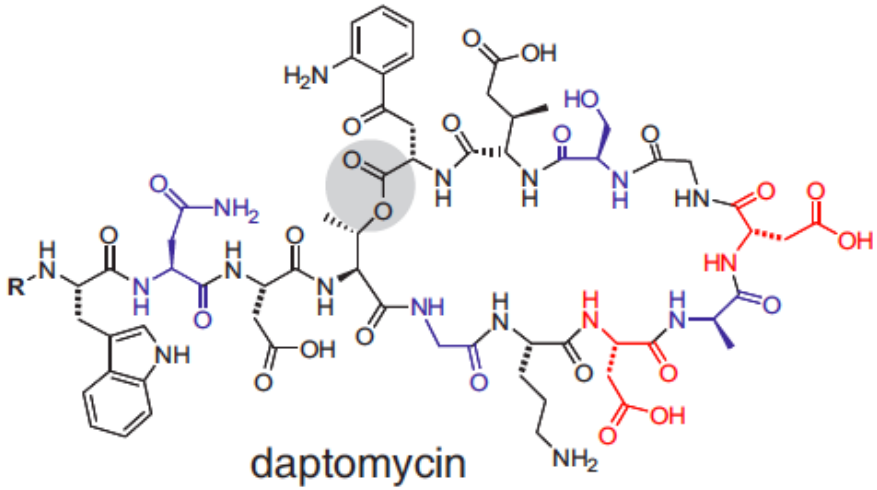
Lipopeptides

Lipopeptides possess antibiotic activity: they interfere with membrane integrity because they are able to insert in the lipid bilayer.

Lipopeptides are usually found as a mixture of variants that may differ in the peptide or lipid moiety depending on external factors (e.g., precursor availability).



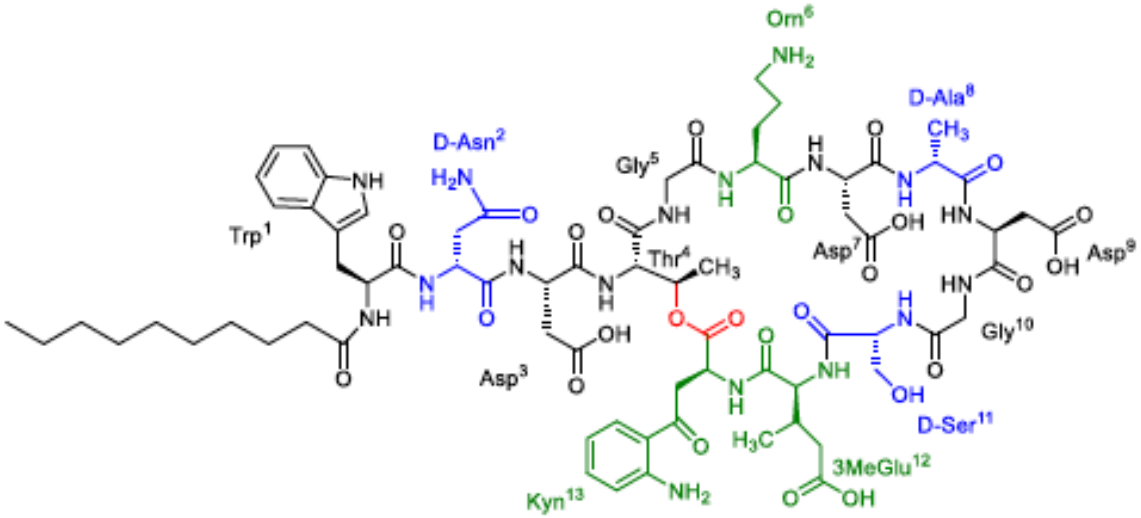
11: surfactin A



daptomycin

Daptomycin

Daptomycin is an acidic lipopeptide produced by *Streptomyces roseosporus*. The antibacterial activity of daptomycin is mediated by calcium.

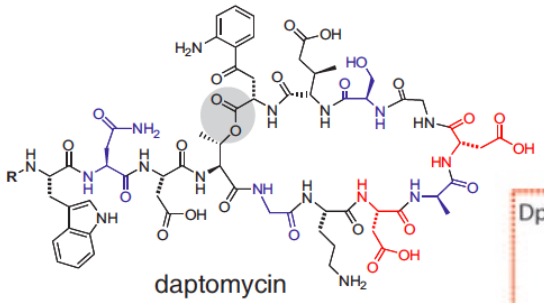


Daptomycin

Daptomycin is used for treatment of multidrug resistant pathogens (vancomycin-resistant enterococci, penicillin-resistant *S. aureus* and *S. pneumoniae*)

Daptomycin synthetase

Daptomycin is produced by NRPS DptA-BC-D. DptE and DptF catalyze activation and incorporation of the fatty acid chain.



factor	R
A21978C-C1	8-methyldecanoyl (*a*C11)
A21978C-C2	10-methylundecanoyl (*i*C12)
A21978C-C3	10-methyldodecanoyl (*a*C13)
daptomycin	*n*-decanoyl (*n*C10)

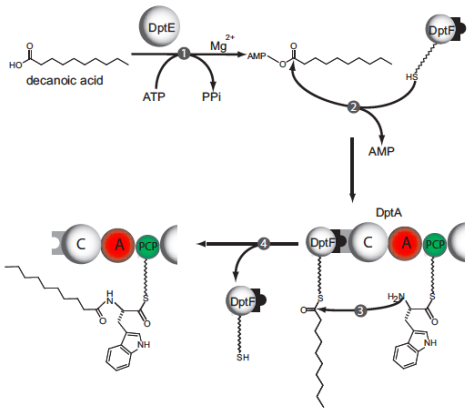
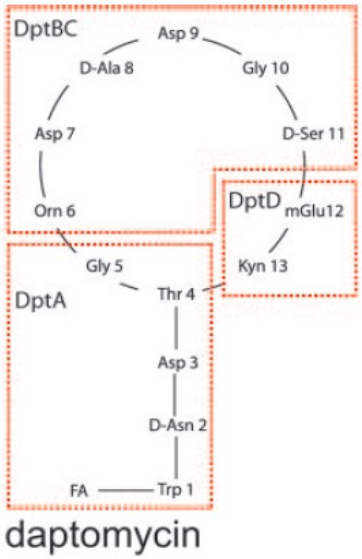
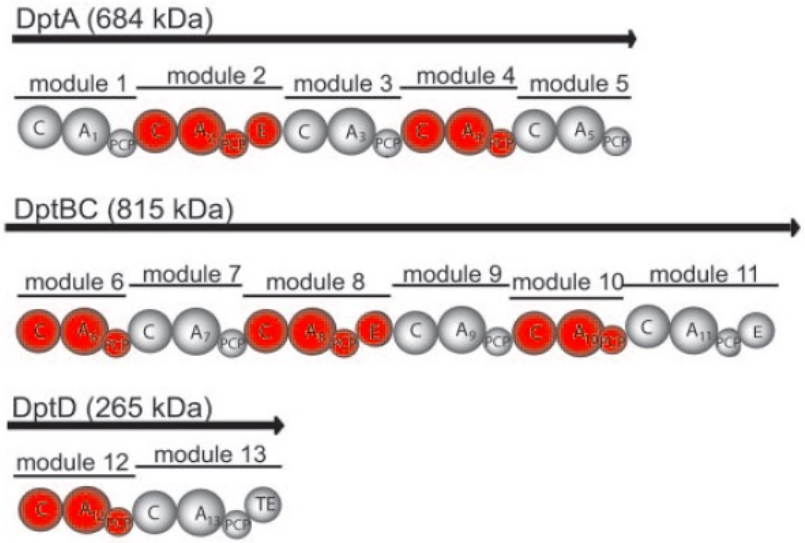


FIG. 5. Proposed mechanism of the lipidation of daptomycin. Decanoic acid is activated as decanoyl-adenylate under the consumption of ATP (1). This step is catalyzed by DptE. The fatty acid is transferred on to the ppan cofactor of the putative acyl-carrier protein DptF (2). DptF interacts with the starter C-domain of DptA, which catalyzes the subsequent acylation of Trp_1 (3). Finally, DptF is released (4).





Mechanism of Daptomycin

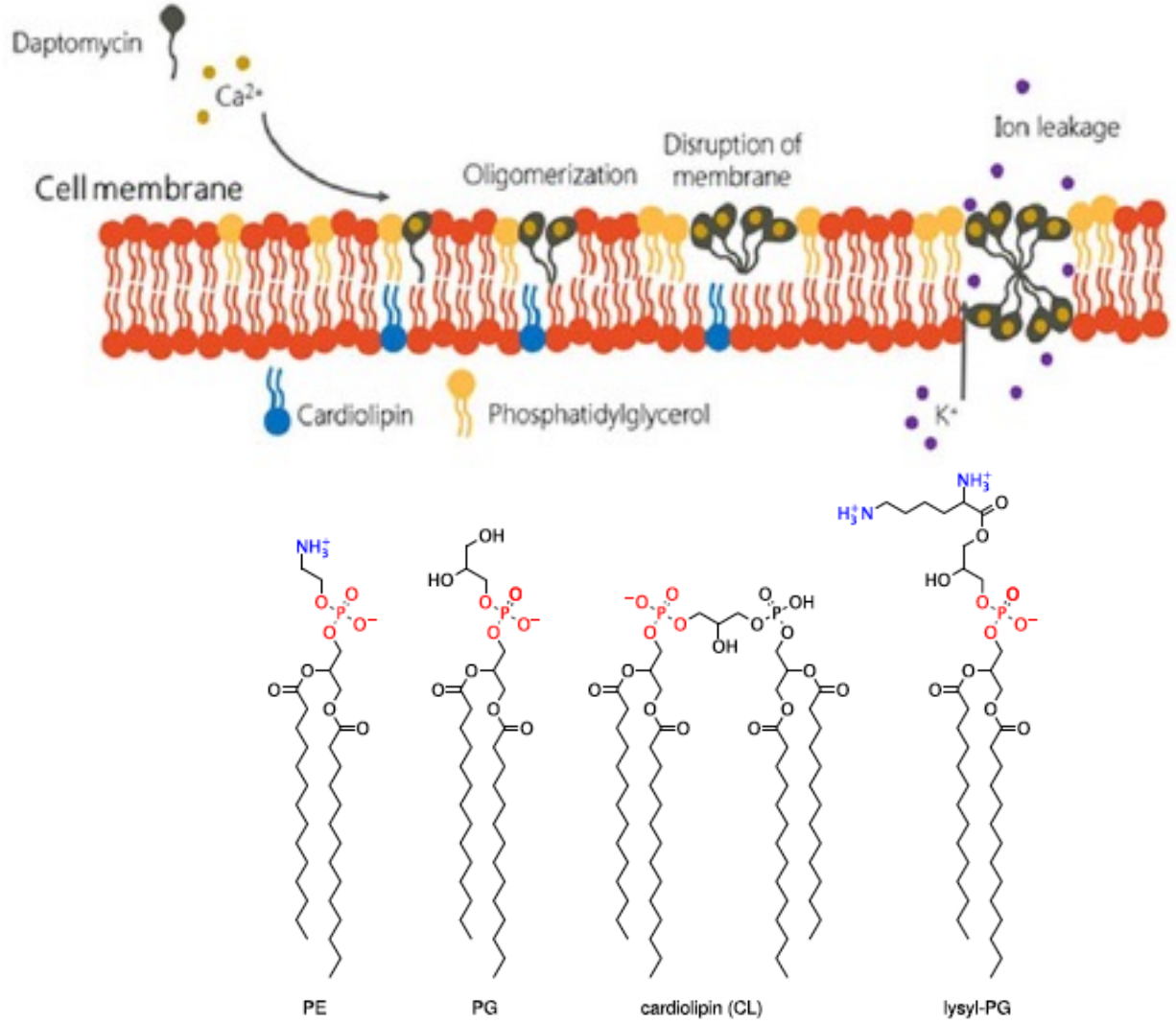


Figure 5. Major phospholipids of Gram-positive bacteria. Phosphatidylethanolamine (PE) is the major neutral phospholipid, whereas phosphatidylglycerol (PG) and cardiolipin (CL) are acidic; with CL, only one of the phosphate groups is usually deprotonated. Lysyl-PG carries a net positive charge. While the figure shows the myristoyl variants of all lipids, the fatty acyl residues in vivo are subject to variation in length and degree of unsaturation.

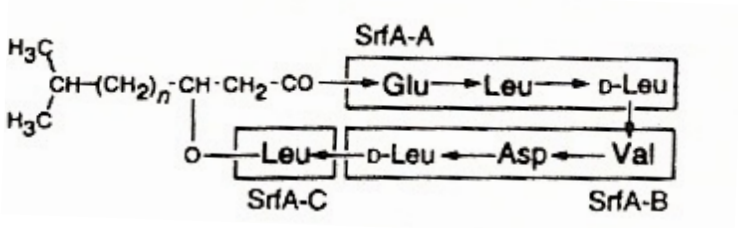
Surfactin synthetase

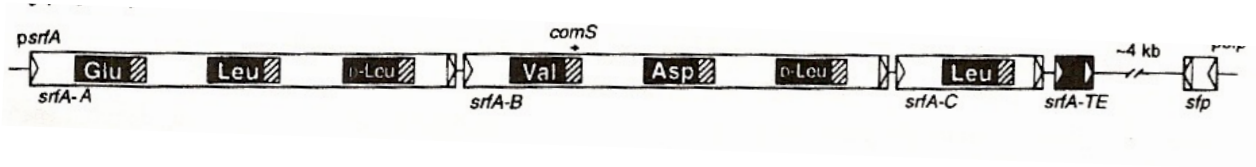
Surfactin is lipopeptide produced by *Bacillus subtilis* that contains a 13-15 C atom β-hydroxyacid. The most frequent is 3-hydroxy-13-methyl-tetradecanoic acid

Genes for surfactin synthetase

SrfA operon

SrfA-A	⇨ E1A	400 kDa
SrfA-B	⇨ E1B	400 kDa
SrfA-C	⇨ E2	150 kDa
SrfA-D	thioesterase	25 kDa





Sfp associated to SrfA

necessary for surfaction biosynthesis

4-phosphopantheteine transferase 26 kDa

Surfactin synthetase

Three fractions with enzymatic activity were purified by gel-filtration

- SrfA-A E1A Glu-Leu-Leu (D)
- SrfA-B E1B Val-Asp-Leu (D)
- SrfA-C E2 Leu
- E3 acyltransferase activity: starts synthesis by transferring β-HA-CoA to Glu on E1A. In the absence of other components it has thioesterase activity. It may be related fatty acid metabolism genes because it is not found in the SrfA cluster

  acyltransferase 40 kDa

Surfactin synthetase

SrfA-A Srf-B Srf-C

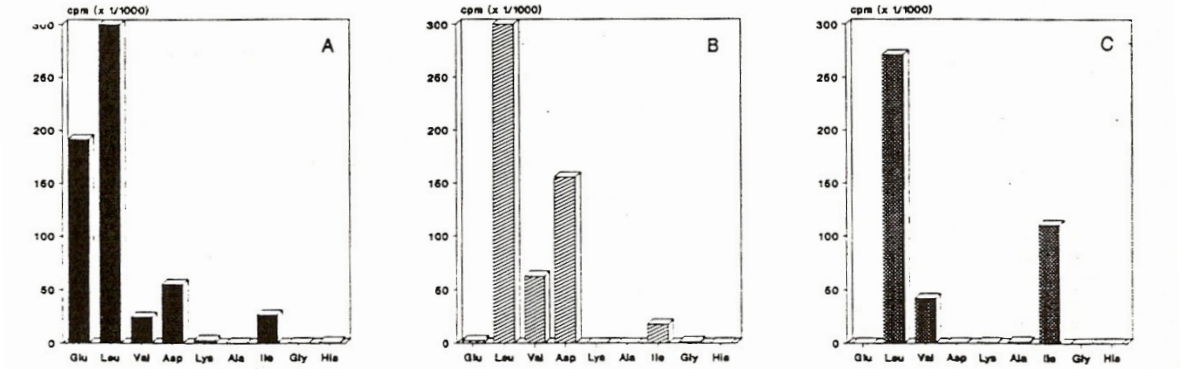


Fig. 4. ATP/PP_i exchange activity of srfAORF1, srfAORF2 and srfAORF3. The partially purified subunits were incubated in the presence of ATP, [^{32}P]PP_i and one amino acid and, after 10 min incubation at 37°C, the [^{32}P]ATP was adsorbed on Norite A (activated coal, Sigma) and counted.

Assignment of specificity of SrfA-A, -B and -C: ATP/PP_i exchange assay and activity assay with labeled amino acids

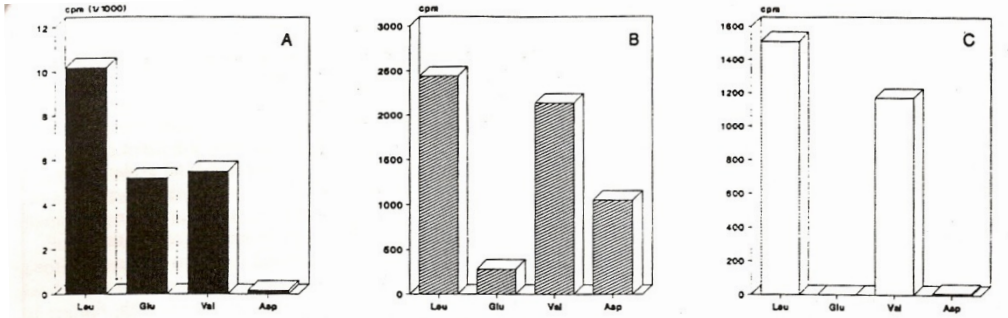


Fig. 5. Amino-acid binding activity of (A) srfAORF1, (B) srfAORF2 and (C) srfAORF3. Each enzyme subunit was incubated in the presence of ATP and one [^{14}C]-labelled amino acid. After 30 min incubation at 37°C, proteins were precipitated with 5% TCA, filtered through cellulose nitrate filters (MFS, USA) and the TCA precipitated material was counted.

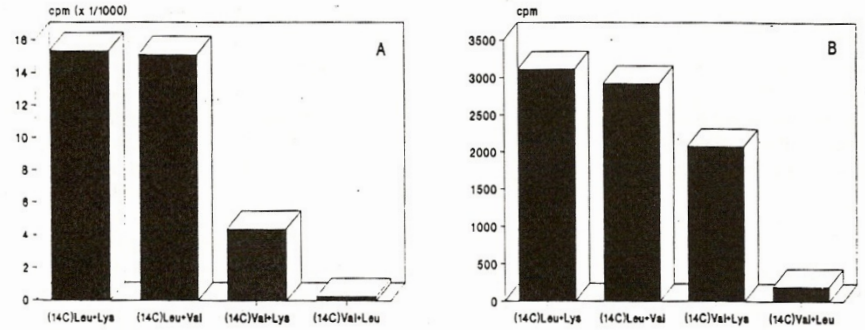


Fig. 6. Amino-acid binding competition experiments. Either srfAORF1 (A) or srfAORF3 (B) were incubated simultaneously with two amino acids, one of which ^{14}C-labelled. Samples were processed as described in Fig. 5. See text for details.

Surfactin synthetase

Construction of hybrid synthetases by module exchange (AT domains) with GrsA, GrsB and ACV

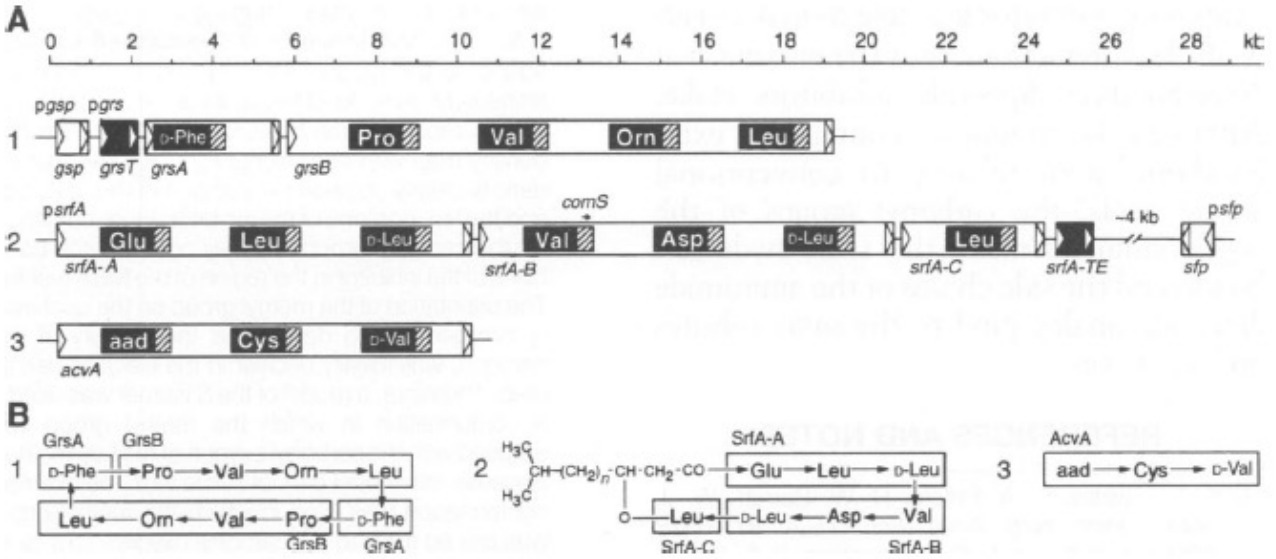


Fig. 1. Multidomain structure of peptide synthetases. (**A**) Schematic diagram of the highly conserved and ordered domain organization of peptide synthetases encoded by the bacterial operons *grs* (row 1) and *srfA* (row 2), and the fungal gene *acvA* (row 3) [aad, δ-(L-α-aminoadipyl)] (*3, 5, 6*). The homologous domains are each about 650 amino acid residues in length and contain modules involved in amino acid–specific adenylation (black boxes) and thioester formation (shaded boxes) (*7*). They are separated by nonhomologous regions (white areas). The locations of promoters (p) and genes associated with antibiotic production (*gsp/sfp*, *grsT/srfA-TE*, and *comS*) are indicated (*3, 5, 17, 19*). (**B**) Primary structure of the nonribosomally synthesized peptides gramicidin S (row 1), surfactin (row 2), and the tripeptide δ-(L-α-aminoadipyl)-cysteinyl-D-valine (ACV), an intermediate of penicillins and cephalosporins (row 3). The amino acid sequences and enzymes that catalyze peptide synthesis are shown.

Surfactin synthetase

Strategy for production of hybrid Srf synthetases *in vivo* in *Bacillus subtilis* to alter specificity for the last amino acid in the last module and replace Leucine in position 7 of the peptide.

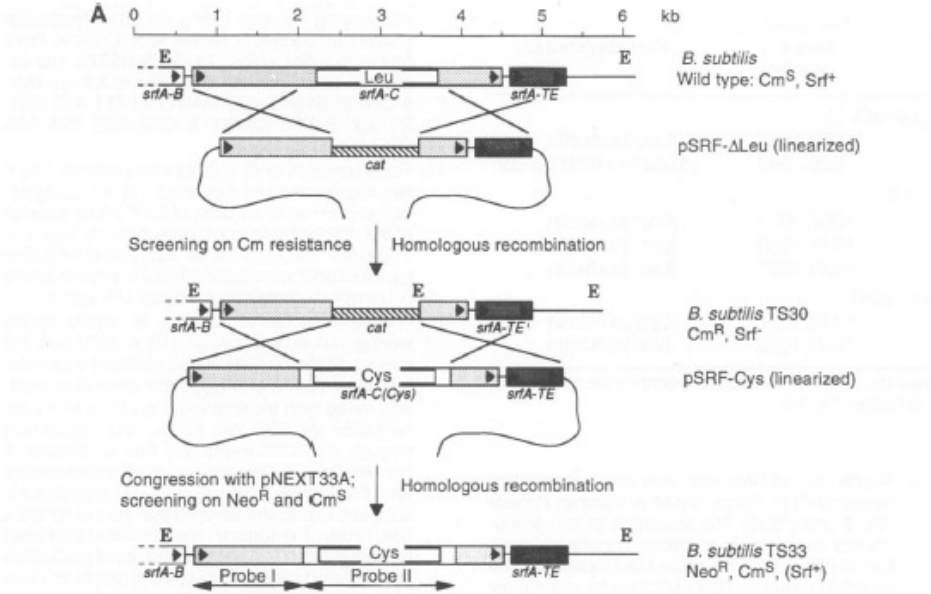


Fig. 3. Targeted *srfA-C* replacement in *B. subtilis*. (**A**) The integrative plasmids constructed (Fig. 2B) were used for in vivo *srfA-C* disruption and domain substitution within the chromosome of *B. subtilis*. Two steps were required: first, the *srfA-C* gene of *B. subtilis* (ATCC 21332) was disrupted by homologous recombination with the linearized plasmid pSRF-ΔLeu (*11*). Subsequently, a domain substitution was achieved by a marker exchange reaction, mediated by the desired domain substitution plasmid (for instance, pSRF-Cys) bearing the gene of a hybrid synthetase (*11*). Loss of the *cat* cassette and the unlinked cotransformation (congression) with pNEXT33A, mediating neomycin resistance (Neo^R), were used to identify strains carrying domain replacements (*11*). Cleavage sites of the restriction enzyme Eco RI within the chromosomal location of *srfA-C* are shown (E). Cm^S, chloramphenicol-sensitive. (**B**) Southern blots of Eco RI–digested genomic DNA were probed with the 5′-linker fragment (left panel) and the substituted Cys domain-coding area of *acvA* (right panel) [shown as double-arrows in (A)], respectively (*16*). For a positive control, we used the unlabeled 5′-*srfA-C* fragment (left) and *acvA* (right) (lanes 1). Genomic DNA was prepared from the wild-type strain *B. subtilis* (ATCC 21332) (lanes 2), TS30 (lanes 3), and TS33 (lanes 4) [compare (A)]. Lanes 5 show Hind III–digested λ DNA as a negative control. The patterns confirmed the disruption (TS30) and domain exchange (TS33) by homologous integration, as shown above. Size markers are indicated on the sides of the gels in kilobases.

Surfactin synthetase

Table 1. Summary of domain exchanges in *srfA-C*. The numbers in the domain indicate the location of the domain-coding regions within the DNA sequence of the corresponding gene, as reported for *srfA* (*5*), *grs* (*3*), and *acvA* (*6*). Hemolytic activity (Hem. act.) was monitored on blood agar plates as described (*13*). MS, mass spectra; *E. c.*, *E. coli*; *B. s.*, *B. subtilis*. The Phe domain originates in the *grsA* gene; the Orn and Leu domains originate in the *grsB* genes.

Domain	Expression* in		Hem. act.	MS data† (major peaks)	Peptide product
	E. c.	*B. s.*			
			B. subtilis (*wild-type* srfA-C)		
Leu (22960–24850)	+	+	+	1023,1037	[Leu^7]surfactin
ΔLeu7 (22960–24850)	–	–	–	928, 942	[ΔLeu^7]surfactin, linear
			B. brevis (grs)		
Phe (1890–3760)	+	+	+	1057,1071	[Phe^7]surfactin
Orn (12770–14670)	+	+	+	1024,1038	[Orn^7]surfactin
Leu (15920–17790)	+	+	+	1023,1037	[Leu^7]surfactin
			P. chrysogenum (acvA)		
Cys (5280–7180)	+	+	+	1013,1027	[Cys^7]surfactin
Val (8490–10460)	+	+	+	1009,1023	[Val^7]surfactin

*Expression was tested by protein immunoblot analysis as described (*20*). †Molecular weights have been determined from mass spectra derived for the various lipopeptides extracted (*12, 14*).

Expected products are obtained but yields are low

- *C* domain selectivity?
- Module/domain communication?

Hu et al. Microb Cell Fact (2019) 18:42
https://doi.org/10.1186/s12934-019-1089-x

Microbial Cell Factories

REVIEW Open Access

Rational strain improvement for surfactin production: enhancing the yield and generating novel structures

Fangxiang Hu, Yuyue Liu and Shuang Li*

Abstract

Surfactin, one of the most powerful microbial surfactants, is a lipopeptide-type biosurfactant which combines interesting physicochemical properties and biological activities. However, the high cost caused by its low productivity largely limits the commercial application of surfactin. Hence, many engineered bacterium have also been used to enhance surfactin biosynthesis. This review briefly summarizes the mechanism of surfactin biosynthesis, highlighting the synthesis pathway of N-terminally attached fatty acids, and outlines the main genetic engineering strategies for improving the yield and generating novel structures of surfactin, including promoter engineering, enhancing efflux systems, modifying the transcriptional regulatory genes of surfactin synthase (*srfA*), genomics and transcriptomics analysis, non ribosomal peptide synthetase (NRPS) domain and combinatorial biosynthesis. Finally, we discuss the future prospects of the research on surfactin.

Keywords: Surfactin, Branched chain fatty acids, Biosynthesis, Structure, NRPS