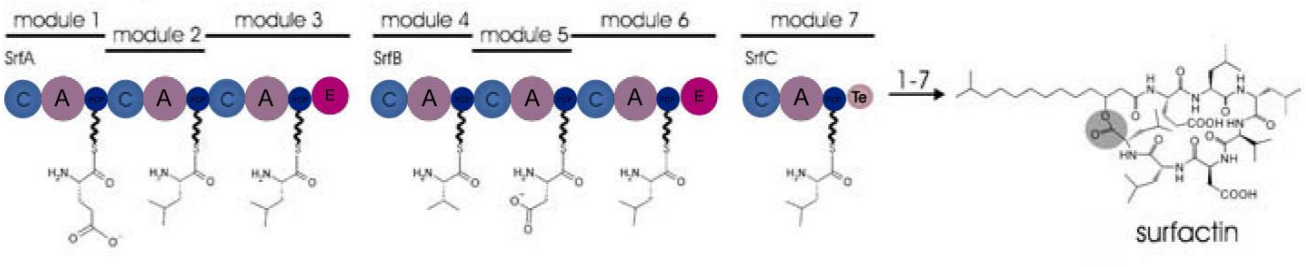
Strategie di analisi e modificazione delle NRPS



Meccanismo tio-templato delle NRPS

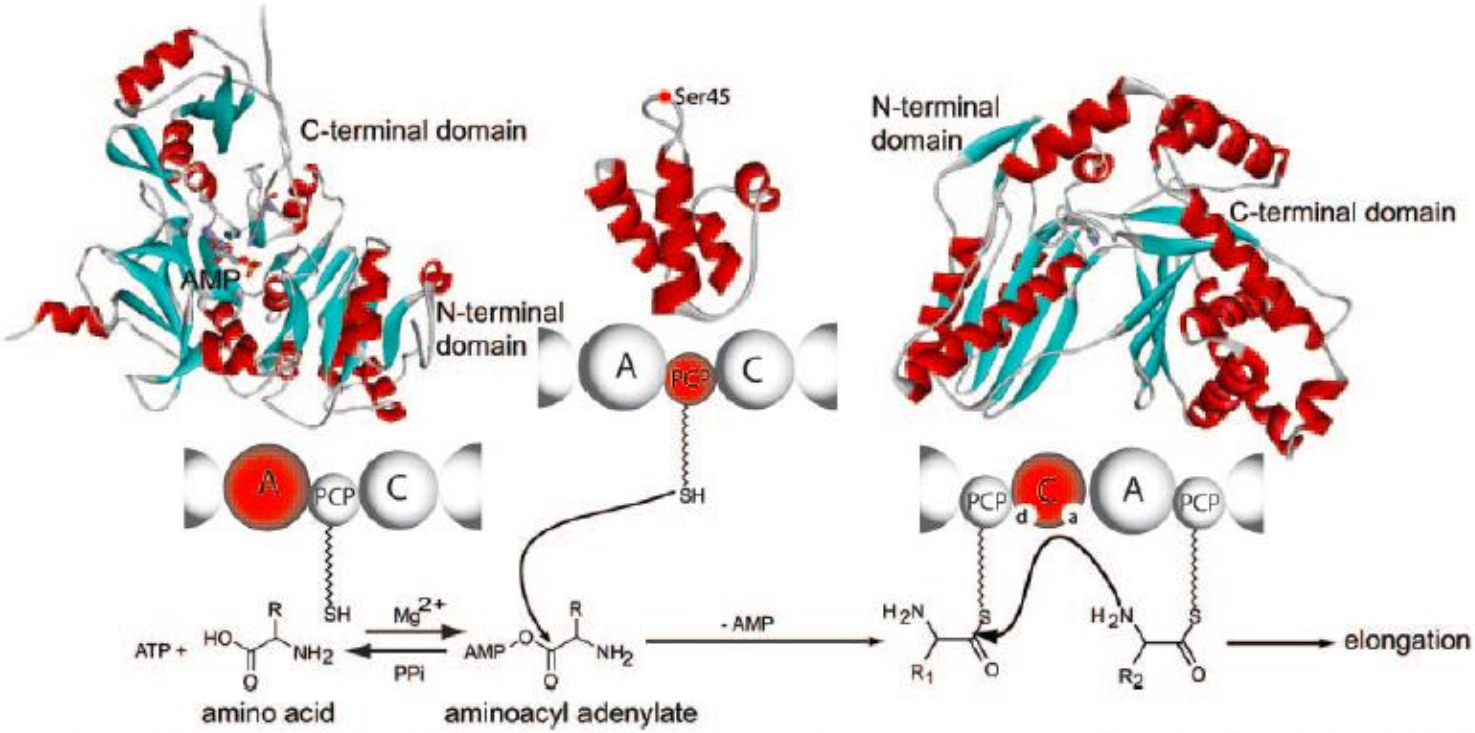
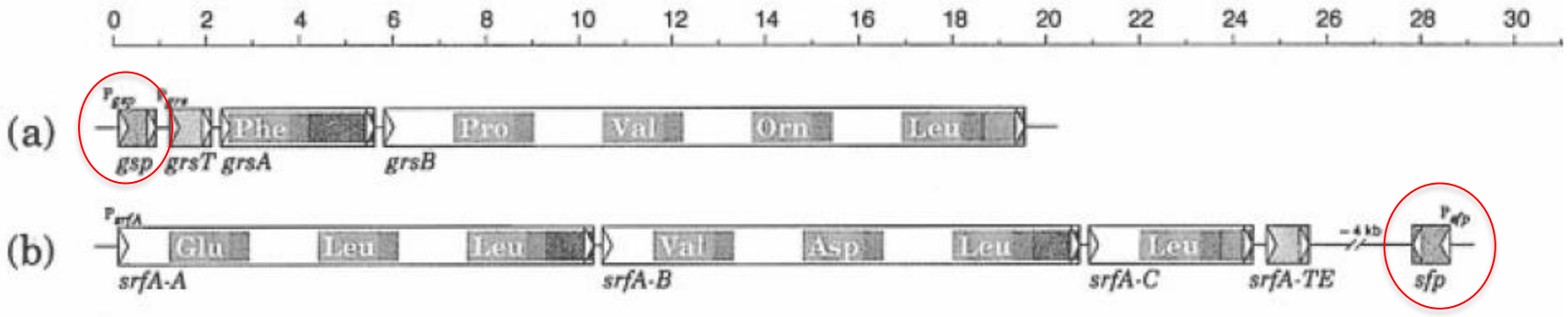


FIG. 4. Chemical principles of nonribosomal peptide synthesis. Domains in action are indicated in red and the respective crystal structures are shown above. First, the A-domain specifically recognizes a dedicated amino acid and catalyzes formation of the aminoacyl adenylate under consumption of ATP. 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).

Meccanismo tio-templato delle NRPS

Per l'attività delle NRPS è necessaria l'aggiunta della 4-fosfopanteteina sulla serina del sito attivo del dominio di tiolazione. Questa modifica post-sintetica è catalizzata dalle 4-fosfopanteteina trasferasi (4-PPTasi), enzimi codificati da geni appartenenti al cluster della NRPS.



Sfp è la 4-PPTasi più utilizzata negli studi sulle NRPS

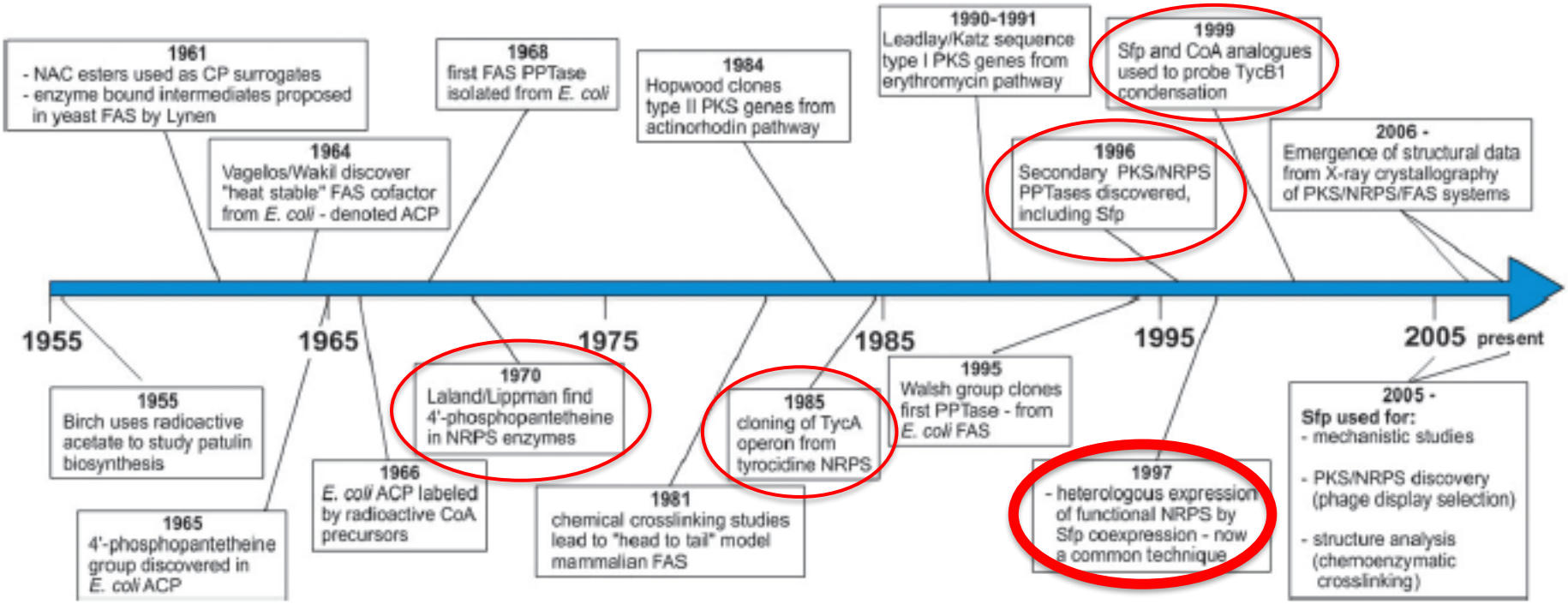
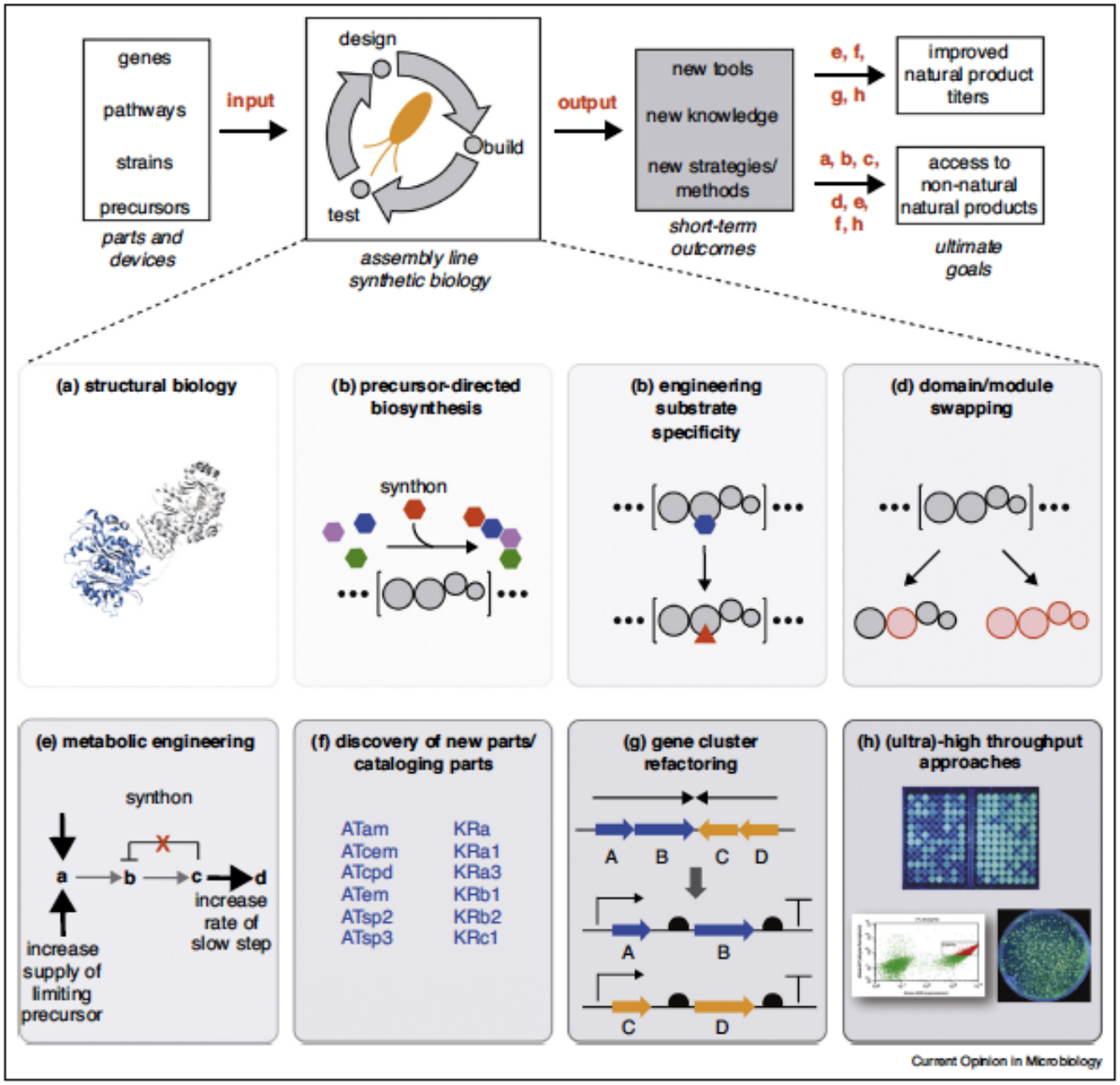


Fig. 2 Timeline of some notable events in the study of carrier-protein-mediated FAS, PKS, and NRPS biosynthetic enzymes.

L'espressione delle NRPS o di singoli moduli o domini ricombinanti permette di avviare lo studio a livello molecolare della struttura e del meccanismo d'azione di questi enzimi.

Strategie per la produzione di analoghi strutturali dei peptidi bioattivi

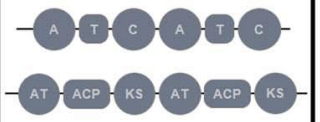
- Alterazione delle condizioni di crescita del microrganismo produttore
  - Biosintesi diretta da precursore
- Riprogrammazione delle NRPS
  - Alterazione mirata della specificità di riconoscimento (dominio adenilazione)
  - Costruzione NRPS ibride per scambio di moduli o domini
    - Sostituzione domini *A* (*A*-T)
    - Sostituzione domini *C*-*A* (*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.

Structure-based applications

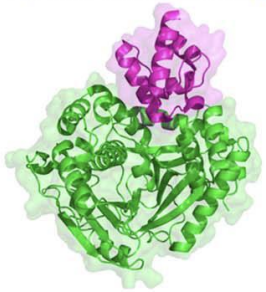
1) Sequence



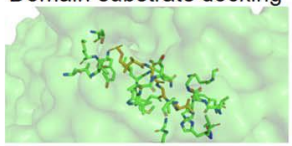
...APGEDAFARQAYQAOQGEUEU
AKATUWREKKBVEQVGRGDSFFA
LGGHSLLAVRMIERLRRIGLGLSV
QTLFQHPTLSVLAQSLVPHREISVP
DNGITADTTVLTPAMLPLIDLT...

2) Homology modeling and docking

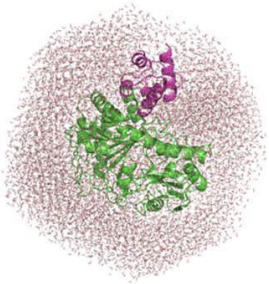
Domain-domain docking



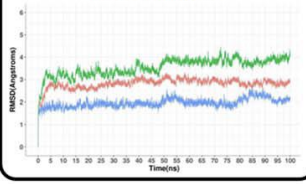
Domain-substrate docking



3) Molecular dynamics simulations

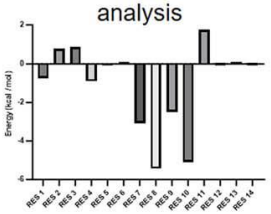


MD analysis

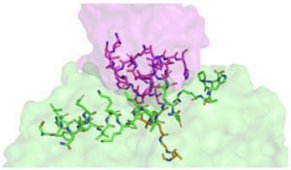


4) Binding free energy calculations

Decomposition per residue analysis

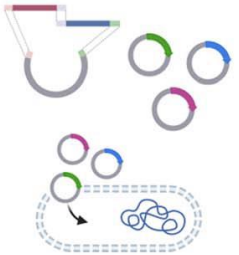


Visual analysis of MD frames

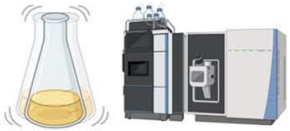


5) Experimental validation

Bioengineering



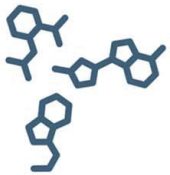
Extraction and analysis



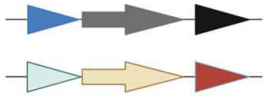
AI-based applications

1) Databases

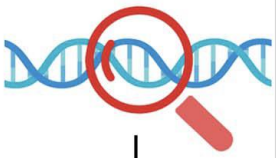
Structure databases



Genome databases



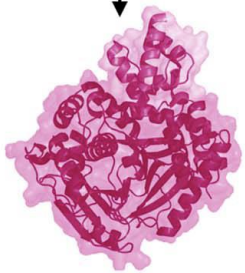
2) Genome mining and BGC prediction



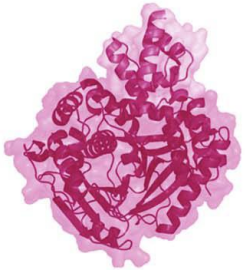


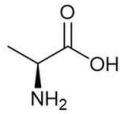
3) Domain structure prediction

...APGEDAFARQAYQAOQGEUEU
AKATUWREKKBVEQVGRGDSFFA
LGGHSLLAVRMIERLRRIGLGLSV
QTLFQHPTLSVLAQSLVPHREISVP
DNGITADTTVLTPAMLPLIDLT...



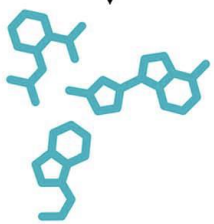
4) Adenylation domain substrate prediction





5) Compound prediction



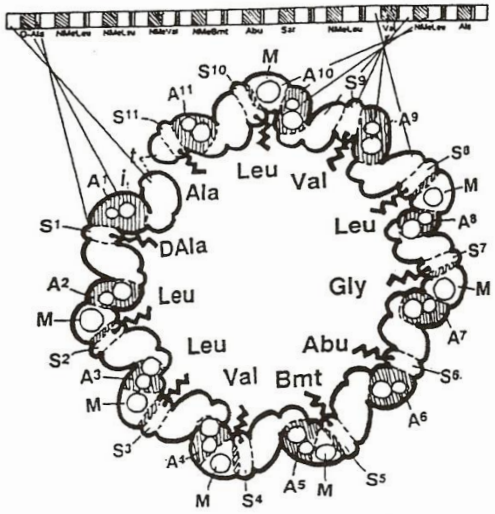


Produzione di analoghi della ciclosporina *A*

La ciclosporina *A* è un immunosoppressore prodotto dal fungo *Tolypocladium niveum*. La ciclosporina sintetasi è costituita da circa 15000 amminoacidi organizzati in 11 moduli.

Sono stati isolati numerosi analoghi strutturali della ciclosporina *A* da colture in terreni contenenti precursori diversi.

➔ la NRPS è in grado di inserire amminoacidi diversi in alcune posizioni del 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

Strategie per la riprogrammazione delle NRPS

- È necessario avere informazioni sulla struttura e i dettagli molecolari del meccanismo d'azione delle NRPS
  - Quali sono le basi della specificità di riconoscimento nel dominio di adenilazione?
  - I domini di tiolazione e di condensazione sono specifici per il substrato?
  - Quali sono le basi del riconoscimento tra domini e moduli?

➔ Studi biochimici *in vitro* sulle proteine purificate

Basi molecolari del riconoscimento del substrato nel dominio di adenilazione

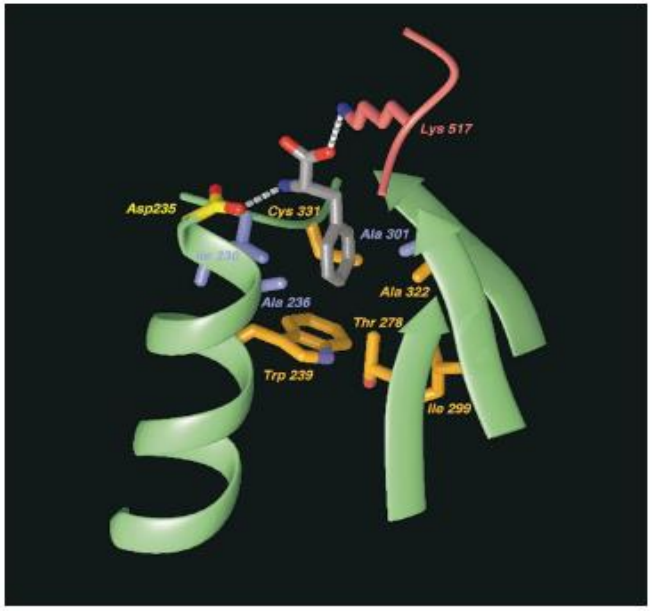


Figure 3. The Phenylalanine binding Pocket of PheA as a Prototype Substrate binding Pocket for the Superfamily of Adenylate-Forming Enzymes

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].)



1. Determinazione della struttura del dominio di adenilazione PheA della Gramicidina Sintetasi A in presenza di Phe

2. Identificazione dei residui amminoacidici che fanno parte della tasca di riconoscimento del substrato Phe

Basi molecolari del riconoscimento del substrato nel dominio di adenilazione

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	

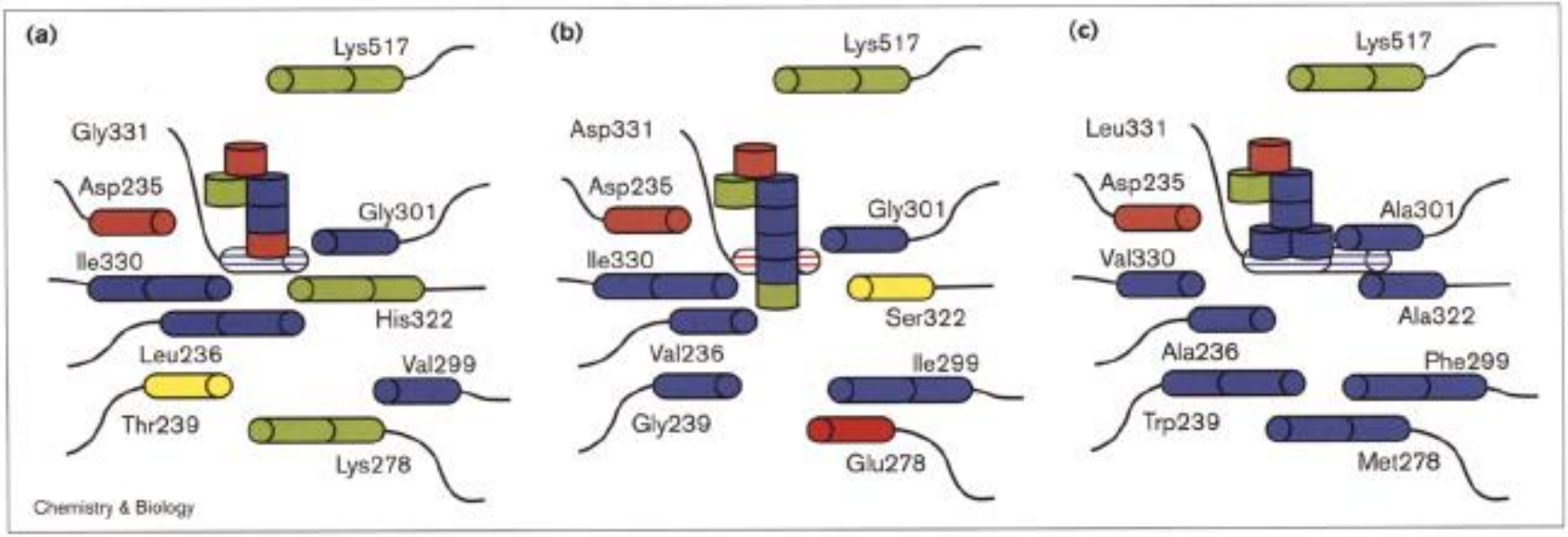
3. Allineamento di sequenze di domini di adenilazione con specificità di riconoscimento nota

4. Identificazione dei residui presenti nelle posizioni corrispondenti a quelle che in PheA legano il substrato

Correlazione ?

Tasche di riconoscimento del substrato in diversi domini di adenilazione

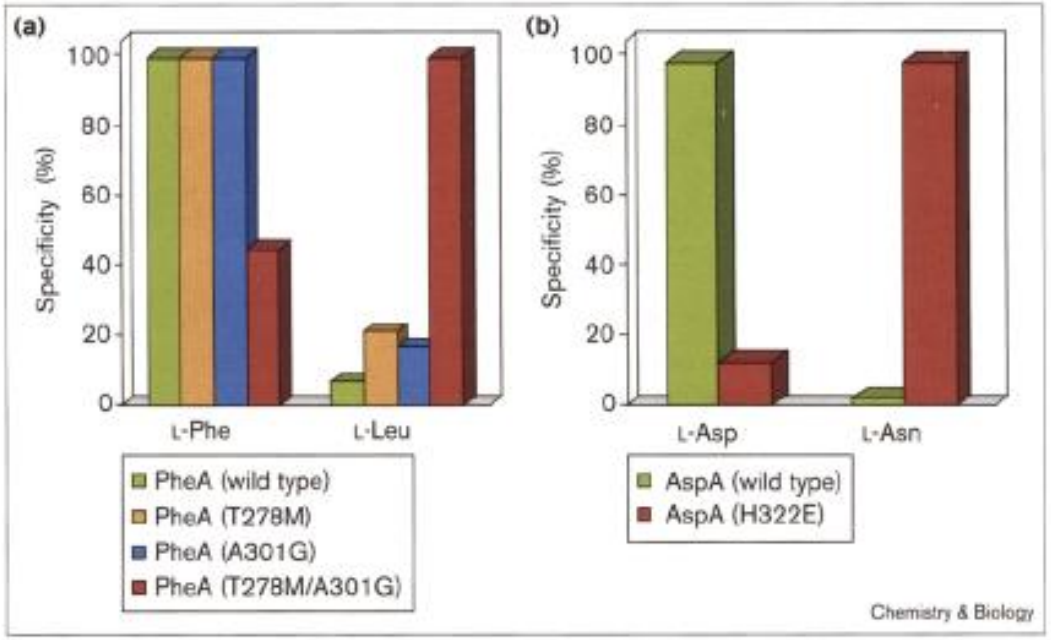
Aspartato Ornitina Valina



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.

Alterazione mirata della specificità di riconoscimento di PheA e 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).



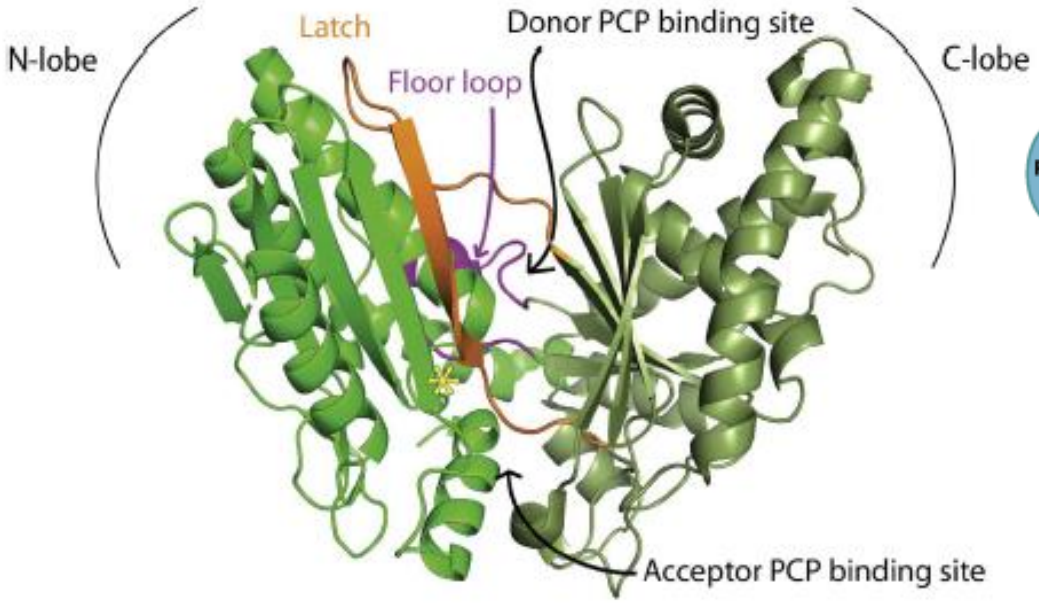
Espressione in *E. coli* del dominio A isolato e saggio di scambio ATP/PPi

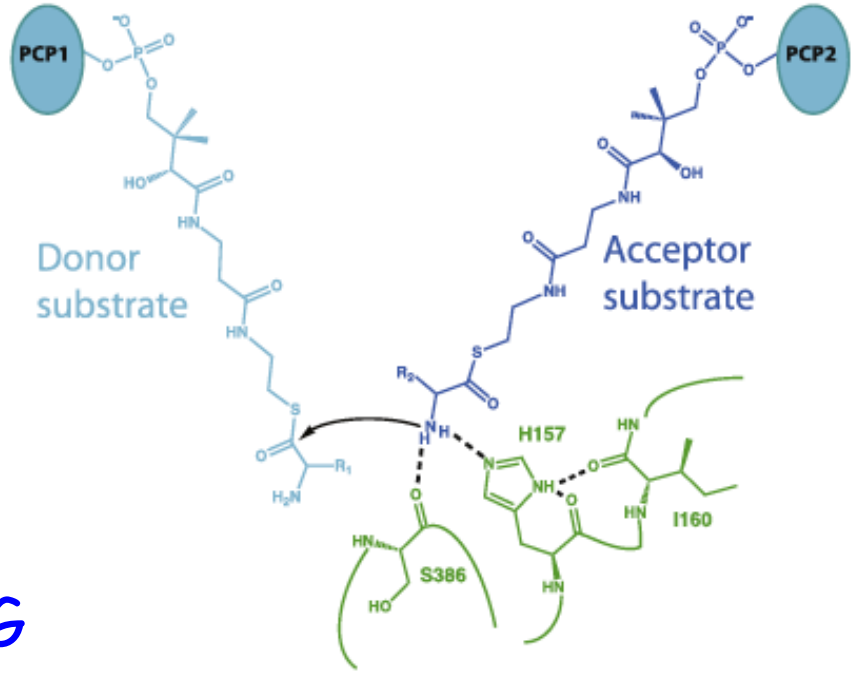
Catalytic efficiency of wild-type and mutant enzymes.

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.

Il dominio di condensazione





Sequenza consenso HHxxxDG

Ruolo di His147 nel dominio di condensazione

Per confermare il ruolo del dominio di condensazione è stata costruita una NRPS formata da **due** moduli: PheATE e ProCAT. I due moduli sono stati espressi in *E. coli* e **attivati** dalla 4'-PP trasferasi sfp.
Fornendo Phe e Pro marcati è stata studiata la sintesi del dipeptide.

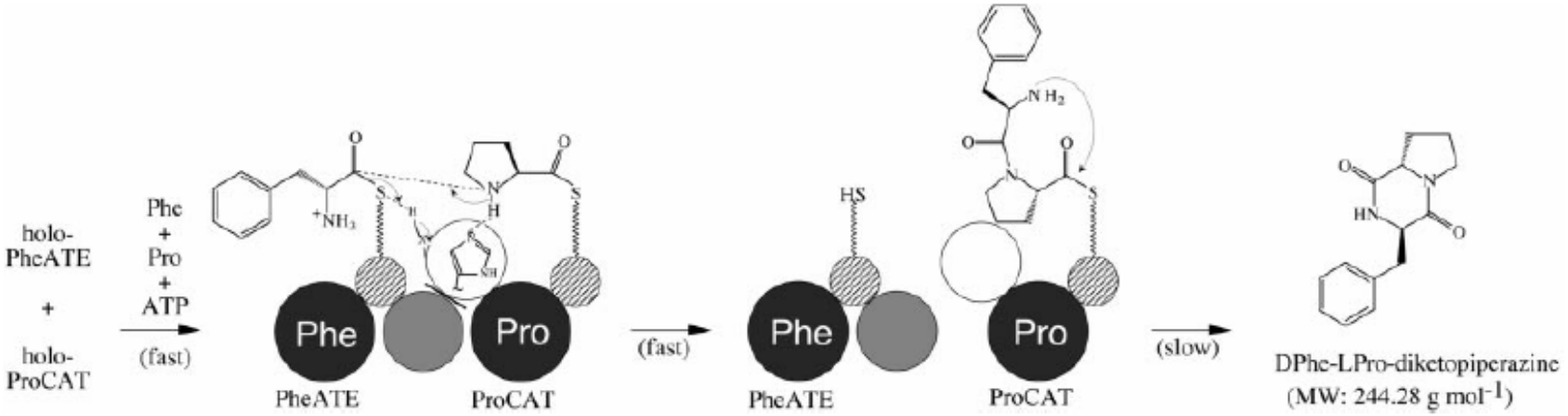


FIG. 7. **Reaction sequence of D-Phe-L-Pro and diketopiperazine formation.** A proposed simple model for peptide bond and diketopiperazine formation is shown. The single domains of the modules PheATE and ProCAT are illustrated as *black* (adenylation), *striped* (thiolation), *gray* (epimerization), 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.

Ruolo di His147 nel dominio di condensazione

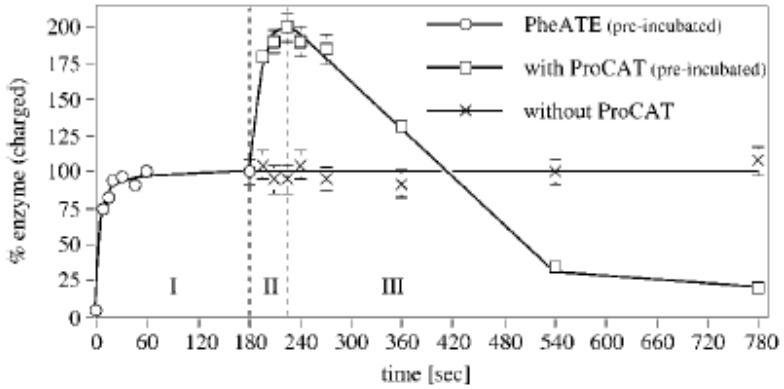
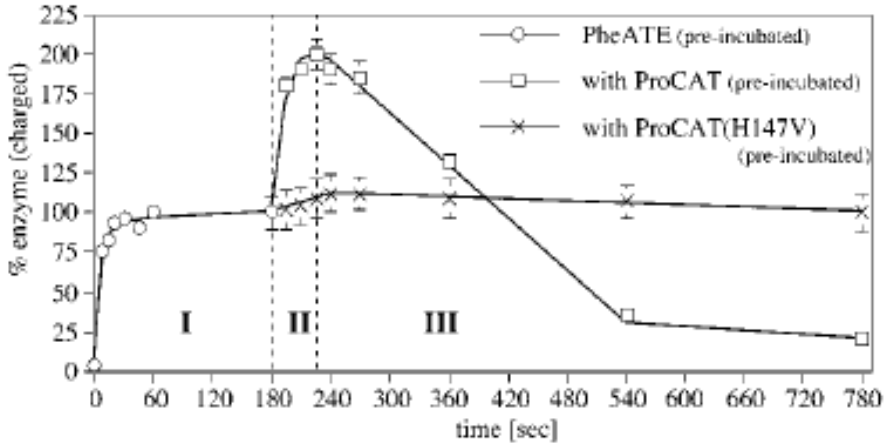
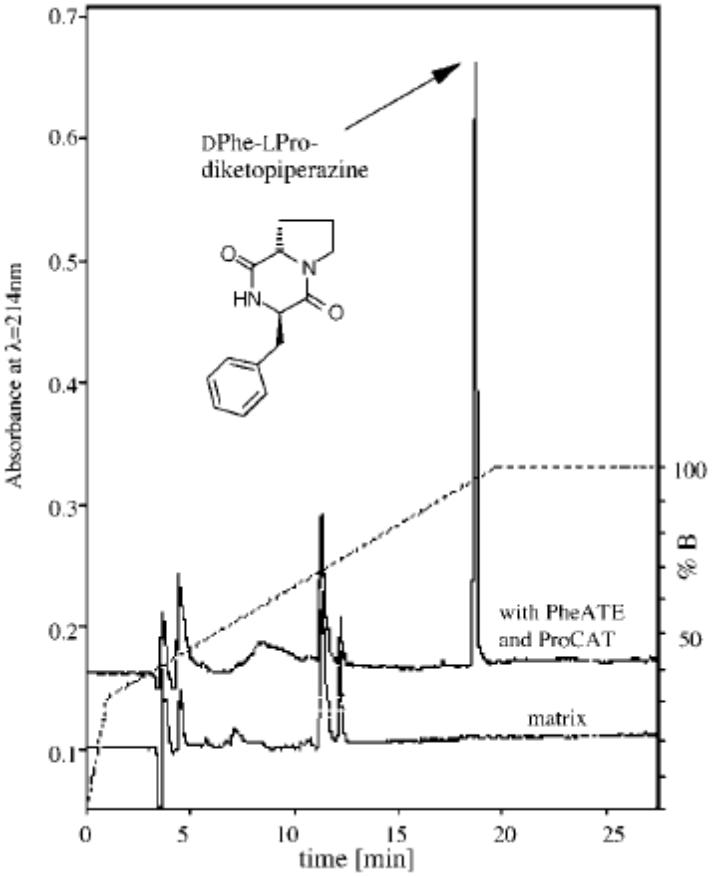


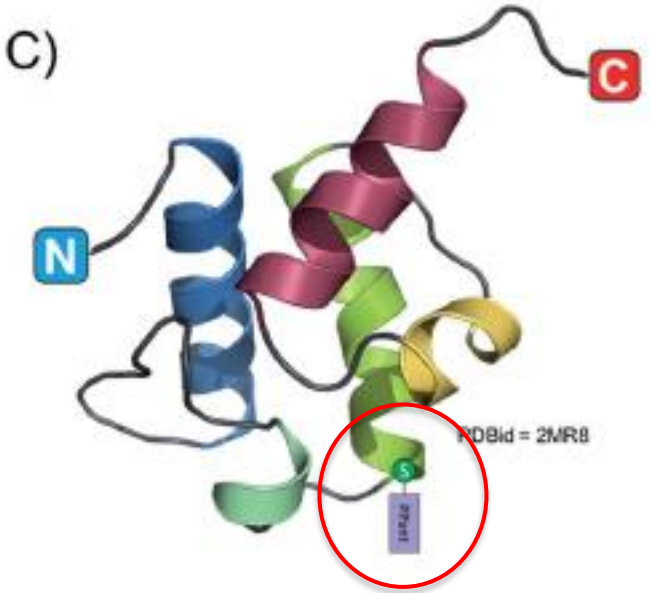
FIG. 4. **Dipeptide formation is catalyzed by PheATE and ProCAT.** PheATE and ProCAT were separately incubated with their cognate amino acids (^{3}H-labeled L-Phe and L-Pro) in order to achieve a complete charging of both enzymes (stage I). At t = 180 s, equivalent amounts of both enzymes were combined. 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.



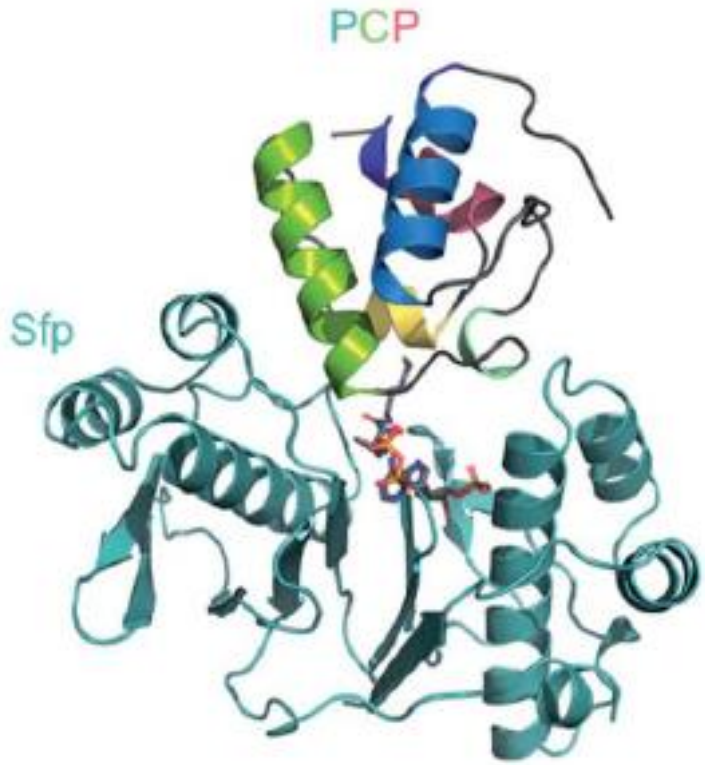


Il dipeptide Phe-Pro non si forma nel mutante H147V.

Il dominio di tiolazione



Struttura del dominio T (PCP)
La serina del sito attivo e il cofattore 4-PP sono esposti al solvente



Struttura del complesso Sfp-PCP

Attivazione del dominio di tiolazione

- Le 4'-PP trasferasi hanno bassa specificità di substrato e quindi possono attivare i domini di tiolazione con acil-CoA, che sono normalmente presenti nella cellula
- Questo processo può causare il blocco della NRPS
- I domini di tiolazione possono essere riattivati da tioesterasi che catalizzano l'idrolisi dei gruppi acilici indesiderati e potrebbero agire anche da filtri di selettività discriminando fra amminoacidi corretti e non corretti *in vivo*

- ***In vitro*** è possibile attivare il dominio di tiolazione con amminoacil-CoA o derivati funzionalizzati del CoA

Sfp, la 4-fosfopanteteina trasferasi di *B. subtilis* possiede bassa specificità di substrato e può utilizzare analoghi del CoA

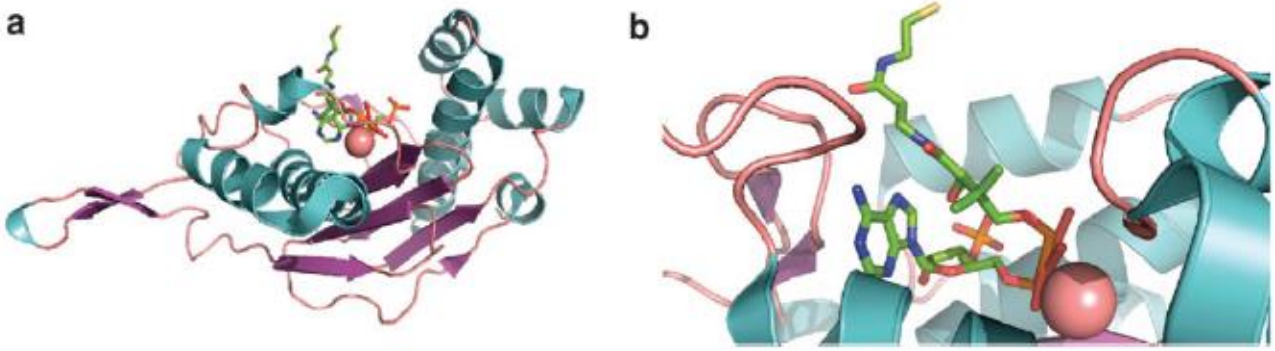
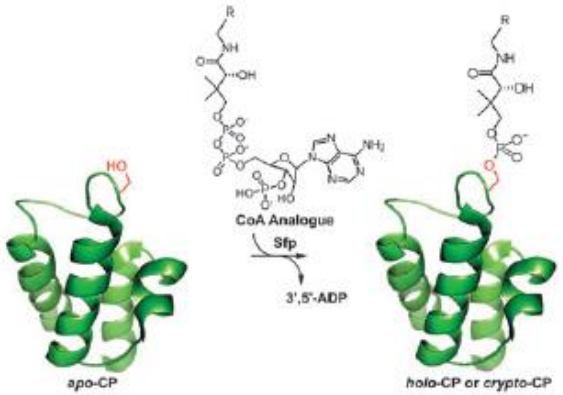


Fig. 6 Crystal structure of Sfp, the promiscuous PPTase from the *B. subtilis* surfactin biosynthetic pathway (PDB:1QR0). (a) Structure of Sfp with CoA substrate and Mg^{2+} cofactor bound in active site. (b) Close-up of CoA bound to Sfp. The terminal β-alanine and cystamine portions of CoA extend into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



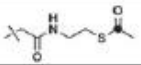
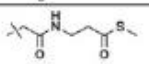
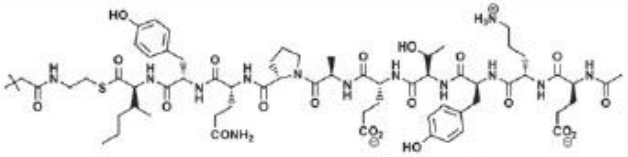
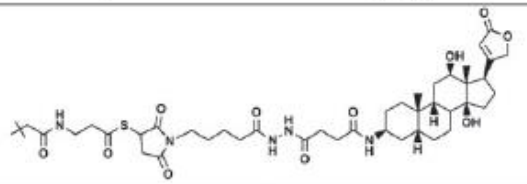
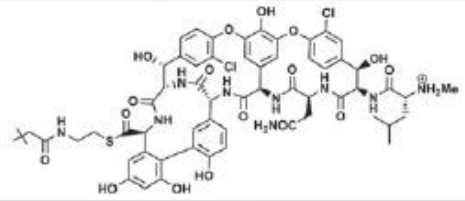
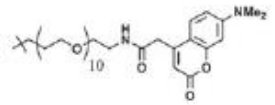
CoA Analogue (R =)	Mol. weight	Ref.
5 	727	46
6 	727	48
7 	623	75
8 	1948	76
9 	1425	77
10 	1810	78
11 	1338	75

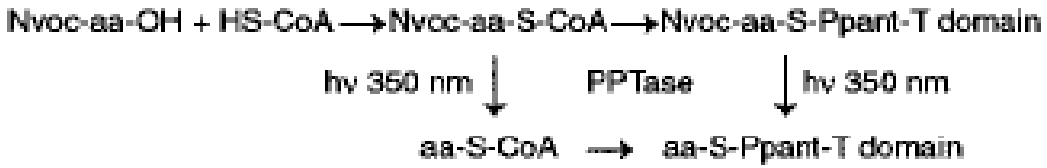
Fig. 7 Notable examples of CoA analogues transferred to carrier proteins using Sfp.

Selettività dei domini di condensazione

Per poter valutare la selettività dei domini di condensazione è necessario superare il controllo esercitato dai domini di adenilazione.

Il dominio di tiolazione viene attivato *in vitro* da sfp con amminoacil-CoA

NRPS bimodulare PheATE e ProCAT: i due moduli possono essere caricati indipendentemente



Il gruppo Nvoc blocca il gruppo NH_2 dell'amminoacido rendendo l'amminoacil-CoA più stabile senza interferire con il riconoscimento da parte di sfp e può essere facilmente rimosso irradiando con luce UV.

Selettività dei domini di condensazione

Selettività del sito donatore (elettrofilo)

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

ProCAT + [^{3}H]Pro

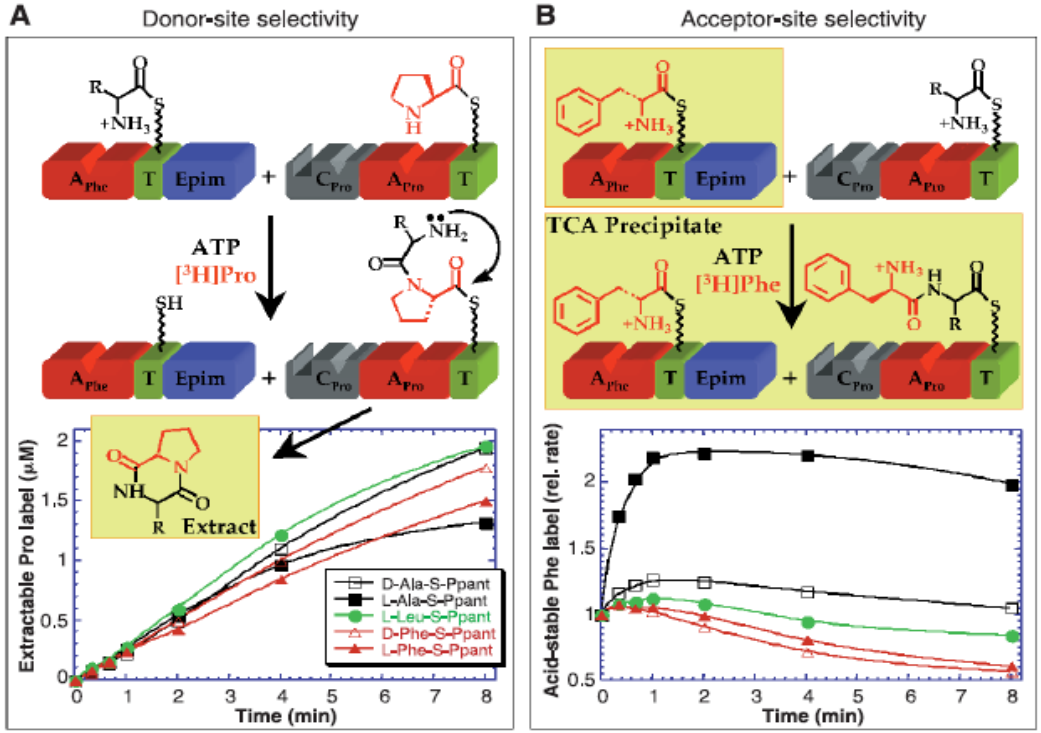
Bassa selettività: il prodotto si forma con tutti gli amminoacidi

Selettività del sito accettore (nucleofilo)

PheATE + [^{3}H]Phe

ProCAT + aa-CoA

Alta selettività: il prodotto si forma solo con L-Ala



Selettività dei domini di condensazione

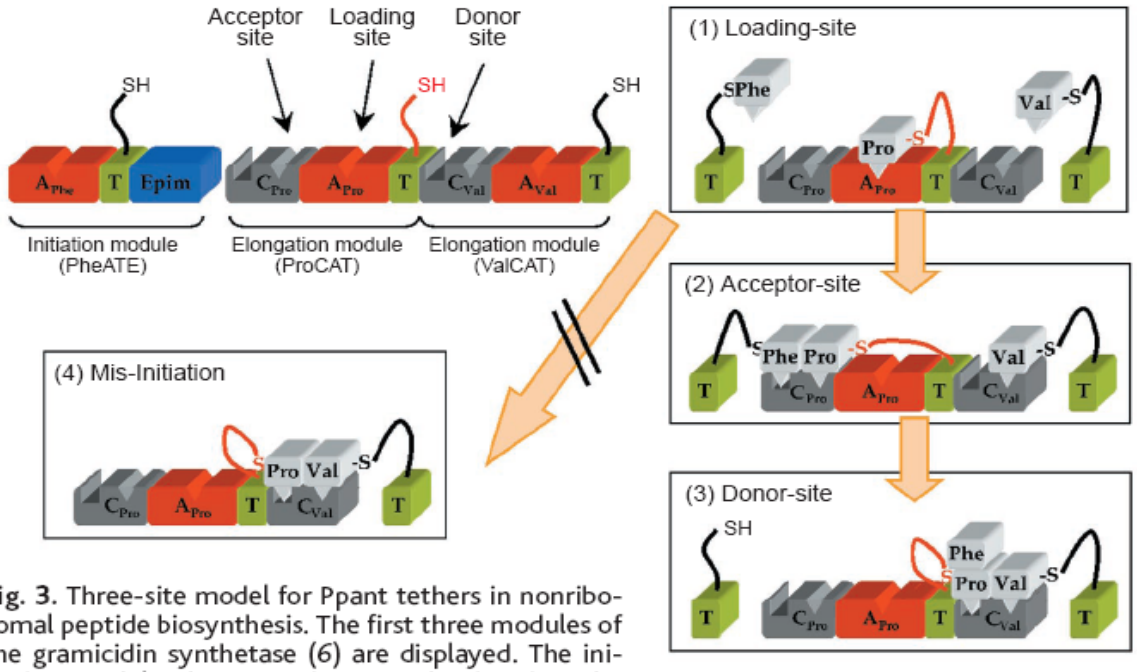


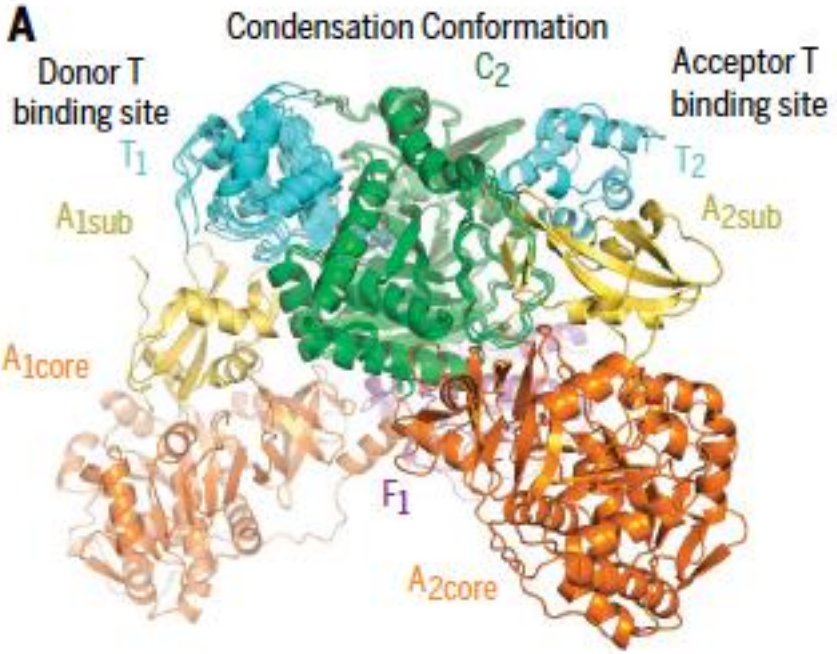
Fig. 3. Three-site model for Ppant tethers in nonribosomal peptide biosynthesis. The first three modules of the gramicidin synthetase (*6*) are displayed. The initiation module PheATE activates and epimerizes the first amino acid (Phe) as a separate subunit from the modules that activate the second (ProCAT) 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.

BIOSYNTHETIC ENZYMES

2019

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]†



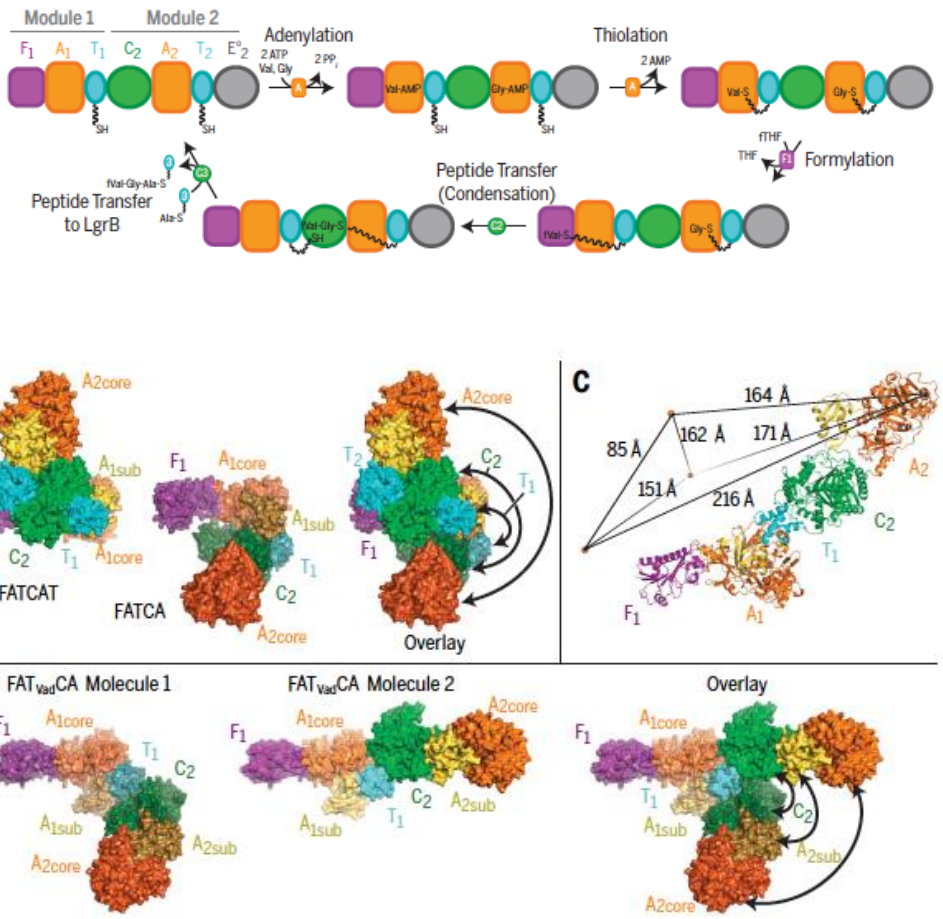
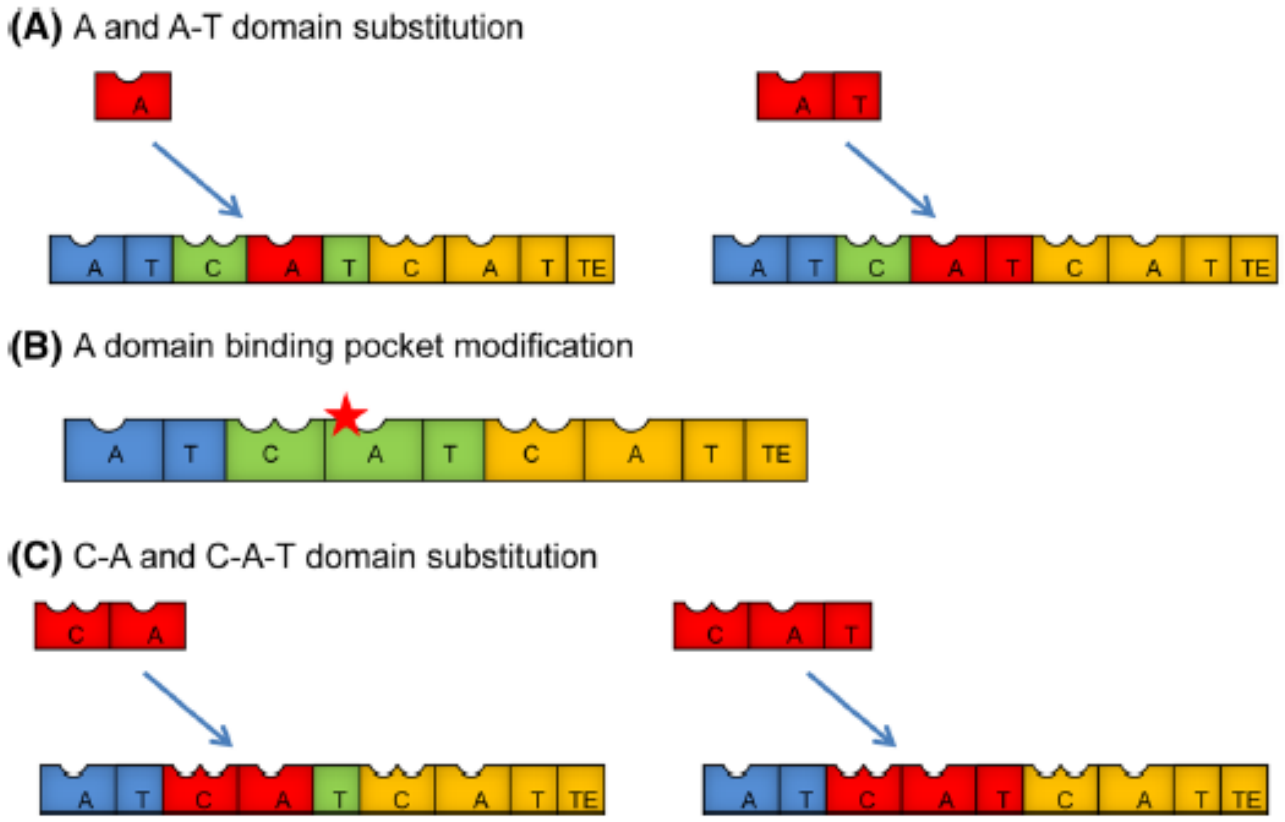


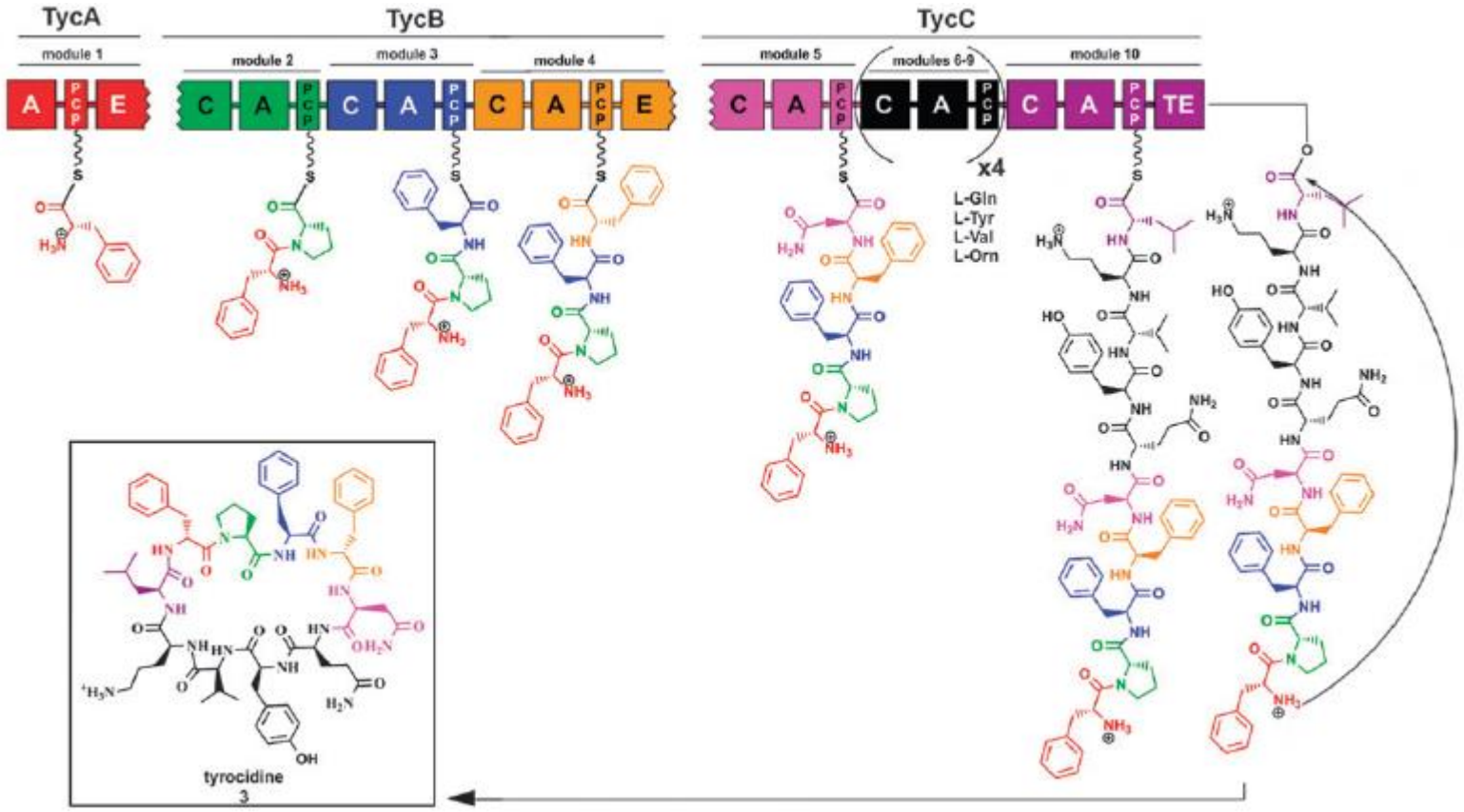
Fig. 5. Different dimodular conformations for the same catalytic states. (**A**) FATCAT and FATCA both show T_1 binding to the donor site of C_2 but have very different overall conformations. (**B**) The two crystallographically independent molecules of $FAT_{Vad}CA$ both show module 1 in thiolation 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}.

Strategie per la riprogrammazione delle NRPS



Strategie per la riprogrammazione delle NRPS
Comunicazione tra moduli e domini *in cis* e *in trans*

Sistema modello Tirocidina sintetasi TycA (modulo di caricamento), TycB (3 moduli) e TycC (6 moduli e dominio Te per la ciclizzazione e il rilascio).



Tirocidina sintetasi:costruzione NRPS ibride per fusione di moduli

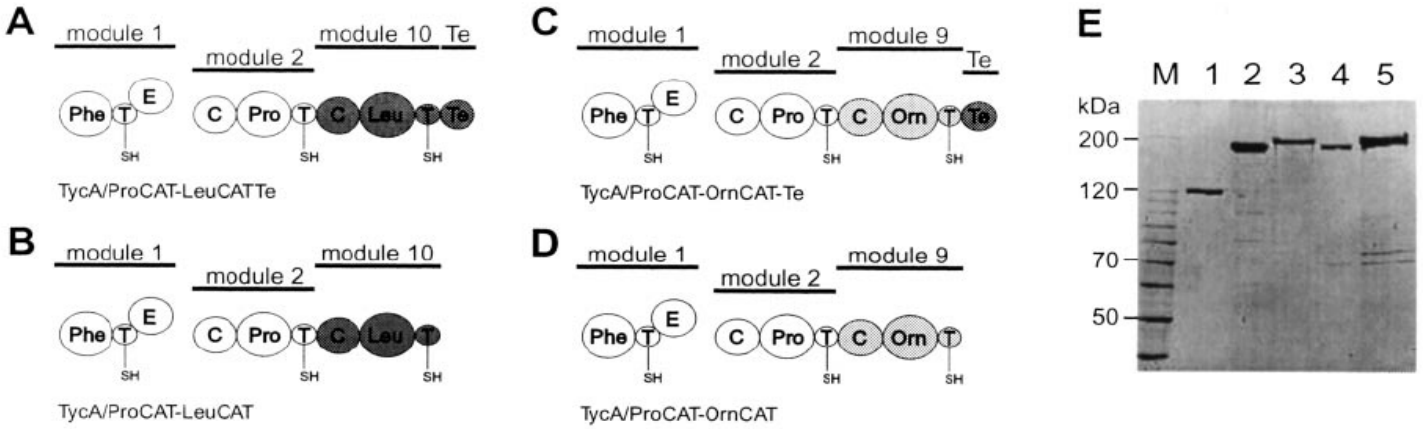


Fig. 3. Construction of artificial trimodular NRPS systems. Artificial trimodular NRPS systems were obtained by extension of the first two modules of tyrocidine synthetases, TycA and TycB1 (here referred to as ProCAT). Fusion of the last module of TycC, containing the Te-domain, to ProCAT yielded TycA/ProCAT-LeuCATTe (*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.

Analisi delle sequenze tra domini T-C e T-Te: individuazione di regioni linker poco conservate sfruttate per costruire le proteine di fusione

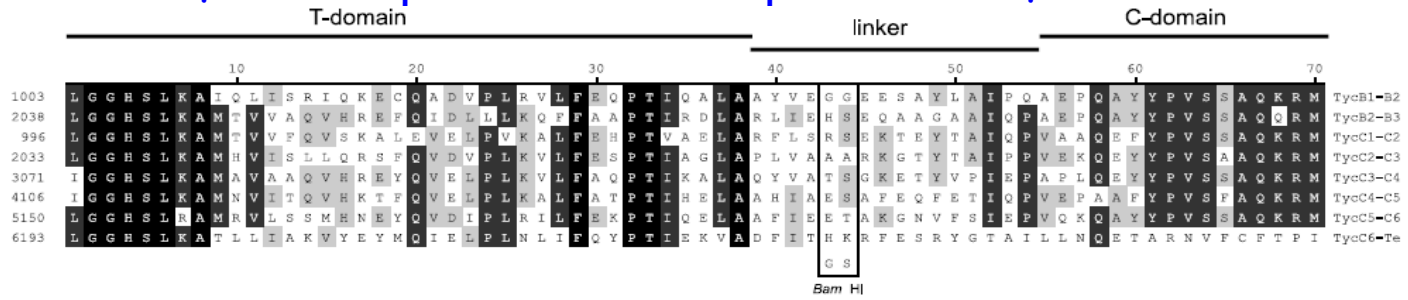


Fig. 2. Linker sequence between T- and C-domains. An alignment of the seven sequences in tyrocidine synthetases TycB and TycC that connect T- and C-domains, as well as the terminal T- and the Te-domains, reveals a stretch with very little conservation. This stretch was defined as linker between the domains. 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.

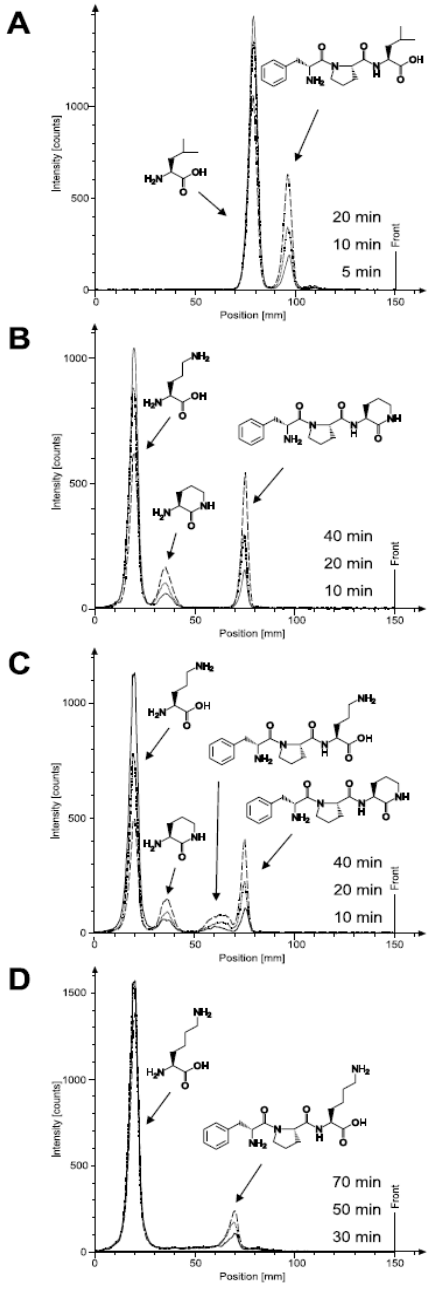


Fig. 4. Formation of the predicted peptides. The product pattern of the trimodular NRPS systems was monitored by using Phe, Pro, and the third amino acid in radiolabeled form, TLC separation of the products and subsequent radioactivity scanning. (*A*) TycA/ProCAT-LeuCATTe with [^{14}C]-Leu. (*B*) TycA/ProCAT-OrnCAT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

Tirocidina sintetasi

Efficienza catalitica nella formazione dei prodotti simile a quella di NRPS wild type: la comunicazione è mantenuta e i domini C e Te possono accettare substrati diversi.

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

Strategie per la riprogrammazione delle NRPS
Comunicazione tra moduli e domini *in trans*

Identificazione di sequenze alle estremità di domini E (C-terminali) e domini C (N-terminali) definite COM^D e COM^A.

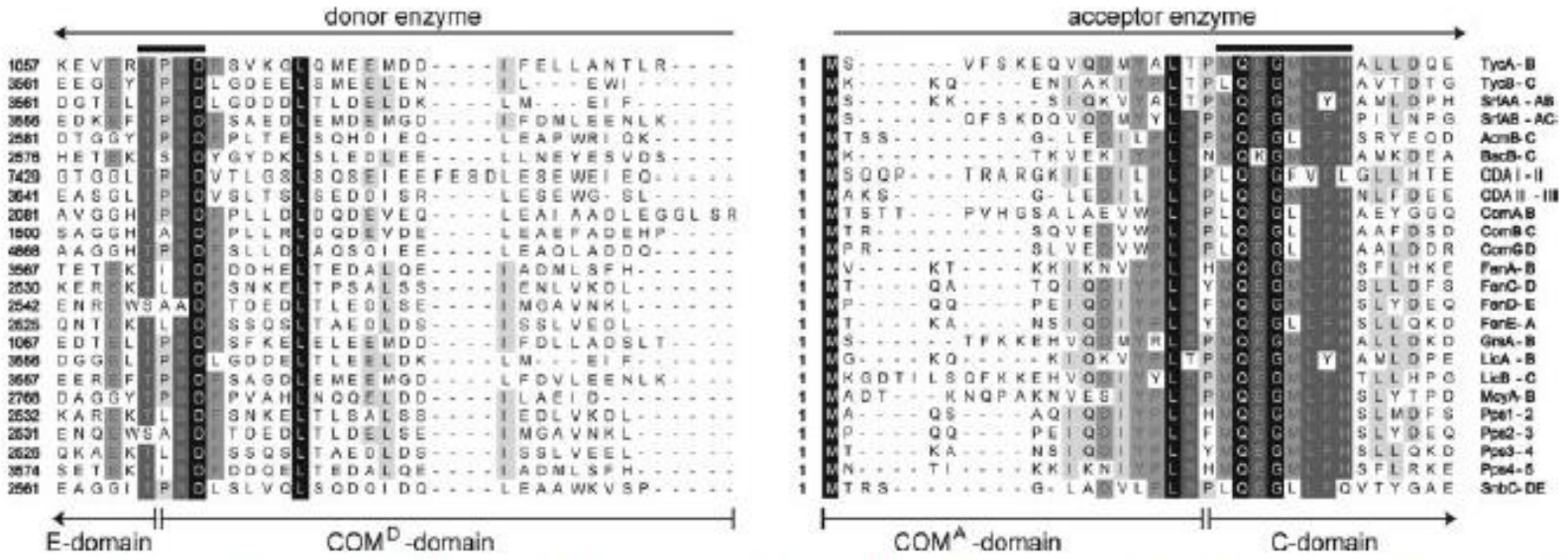
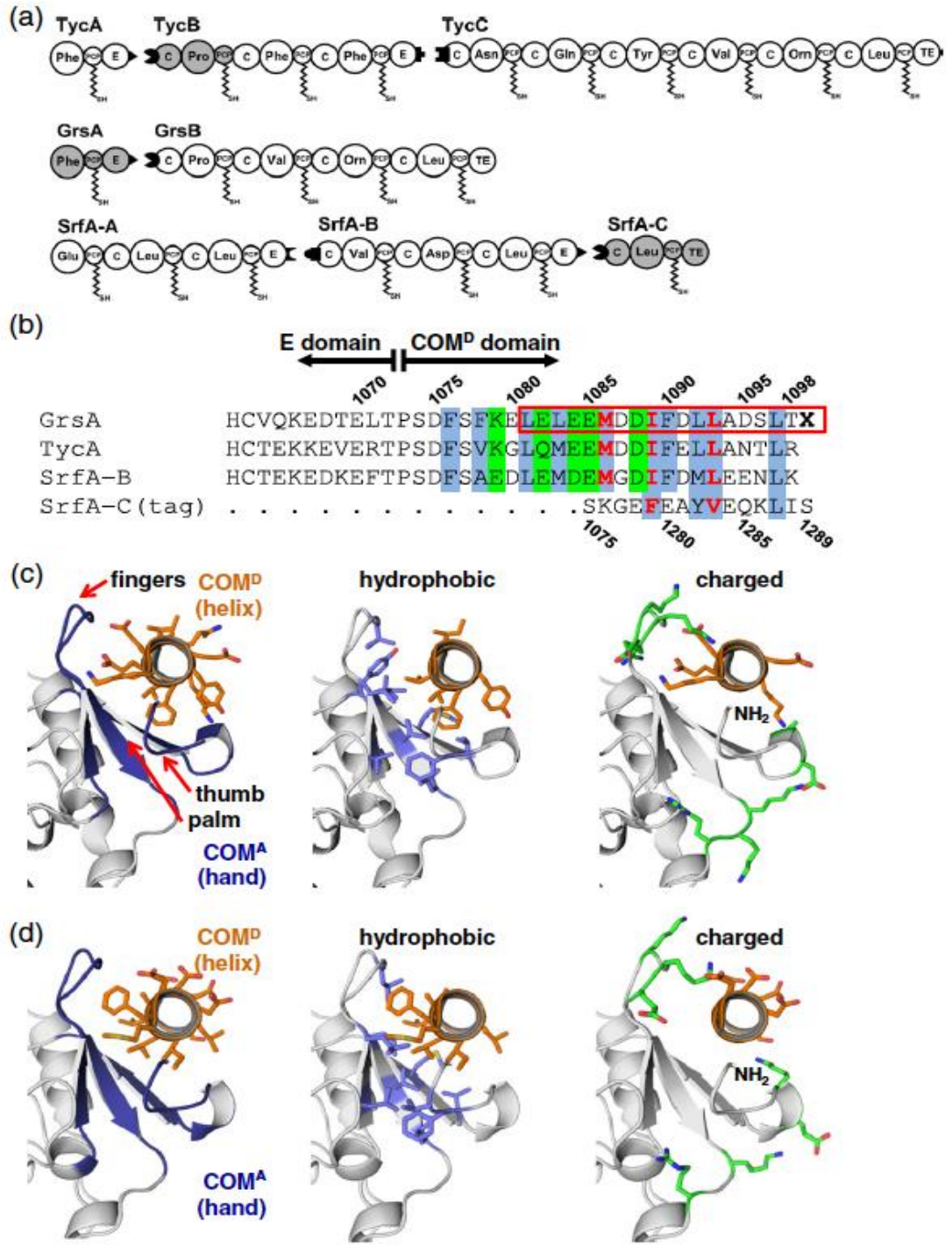


Fig. 2. Sequence comparison of proposed COM^D and COM^A domains derived from 12 biosynthetic systems: actinomycin (Acm), bacitracin (Bac), complestatin (Com), calcium-dependent antibiotic (CDA), fengycin (Fen), Grs, lichenysin (Lic), microcystin (Mcy), pliplastatin (Pps), pristinamycin (Snb), SrfA, and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

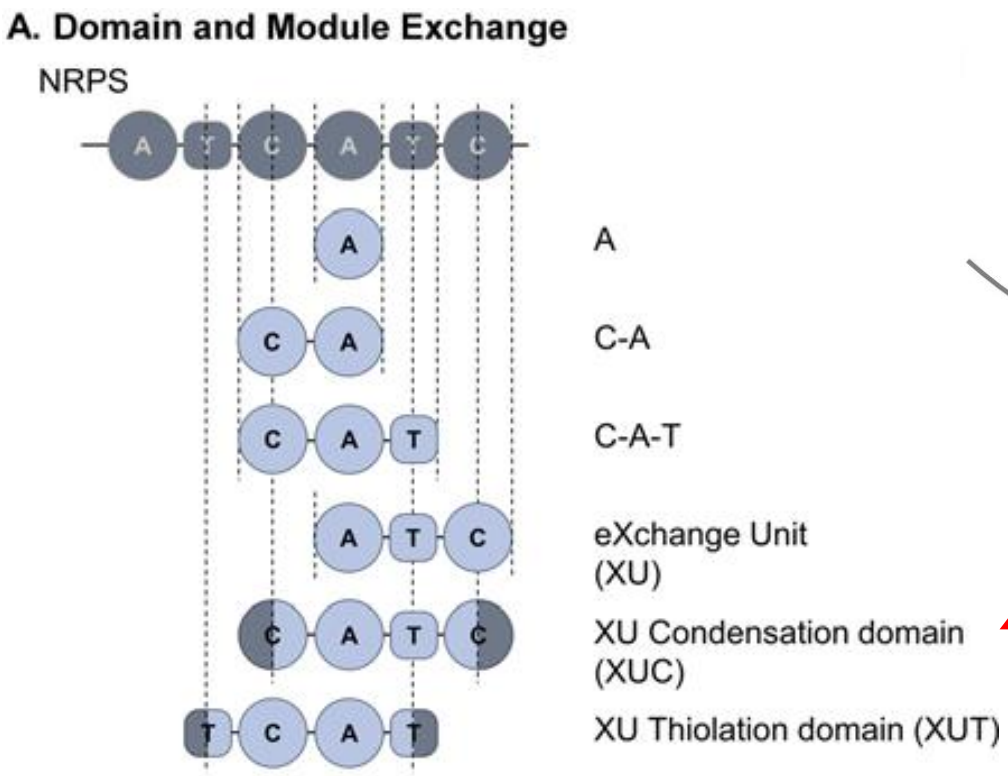
Comunicazione tra moduli *in trans*

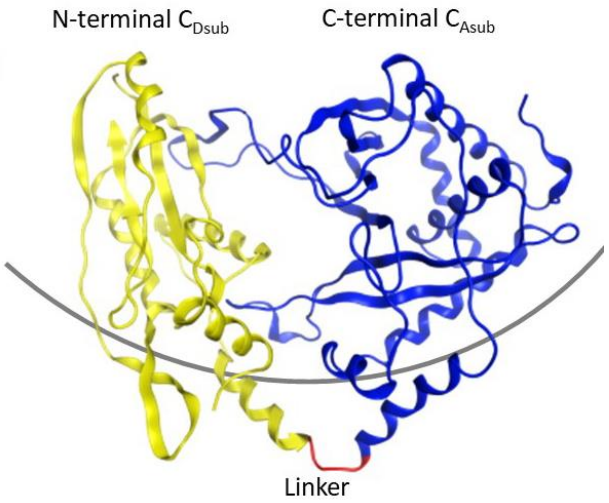
Analisi strutturali delle regioni COM^D e COM^A.

Residui carichi e residui idrofobici contribuiscono alla specificità di riconoscimento

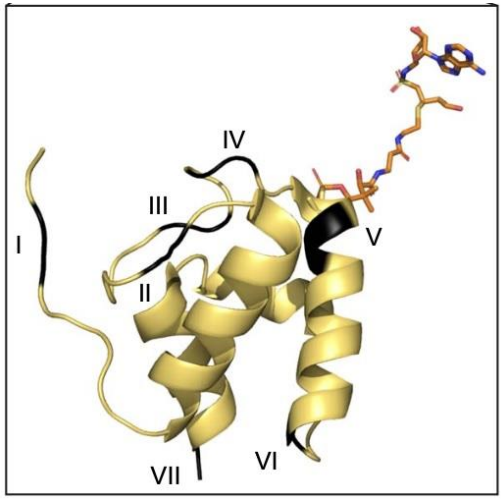


Diverse fusioni per lo scambio di domini e moduli

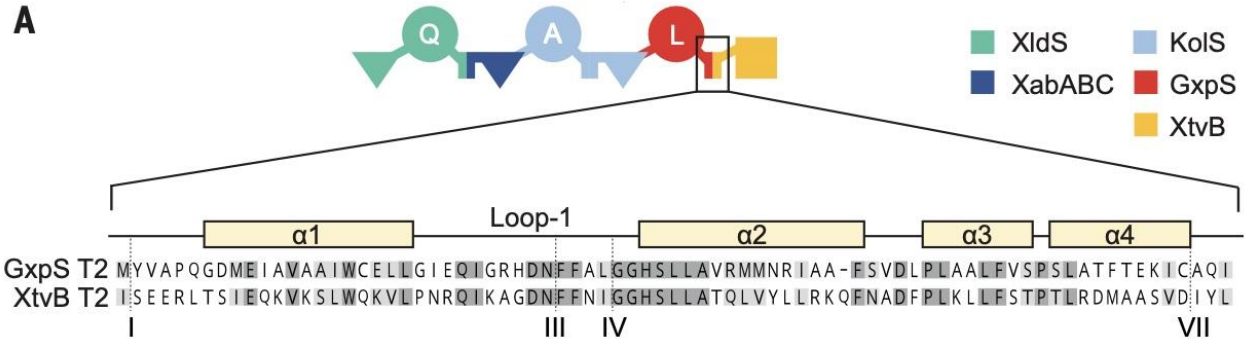




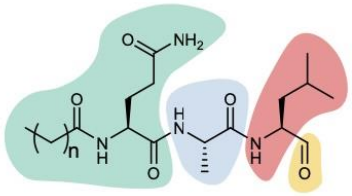
Costruzione di una NRPS ibrida usando lo schema XUT



A



B



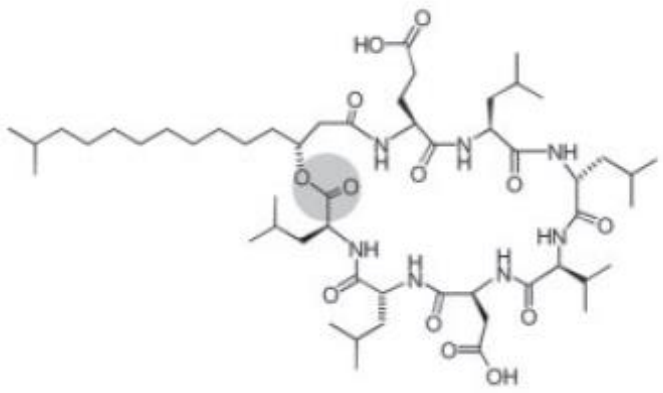
Products	n
46	11
47	12
48	13

Fusion site	NRPS	Products	Production [mg/l]
I	29	**46**	1.2 ± 0.1
		47	3.5 ± 0.2
		48	1.5 ± 0.03
III	30	**46**	7.2 ± 0.4
		47	21.6 ± 0.5
		48	5.7 ± 0.02
IV	31	**46**	7 ± 0.4
		47	20.9 ± 0.7
		48	5.7 ± 0.1
VII	32	**46**	1.7 ± 0.2
		47	4.1 ± 0.1
		48	1.8 ± 0.1

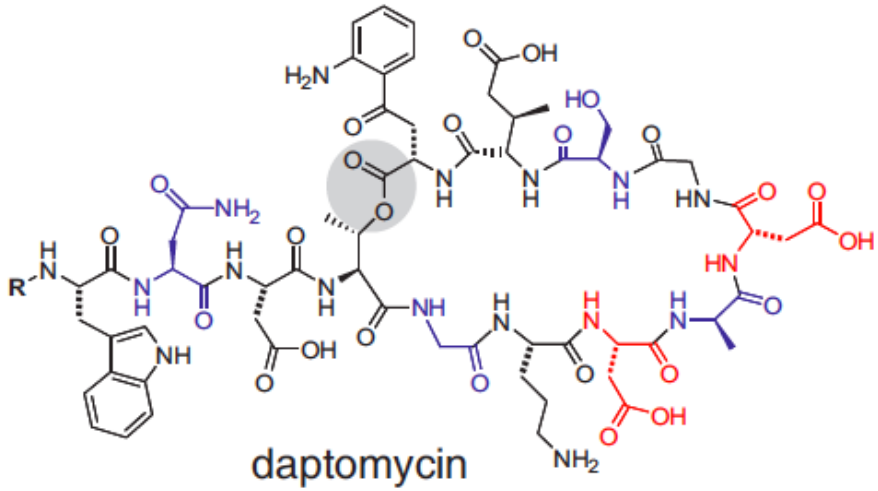
Lipopeptidi

I lipopeptidi hanno attività antibiotica: interferiscono con l'integrità della membrana cellulare perché sono in grado di inserirsi nel doppio strato lipidico.

Di solito sono presenti come una miscela di varianti strettamente affini che differiscono nella componente peptidica o lipidica in dipendenza da fattori esterni (es. disponibilità dei precursori).



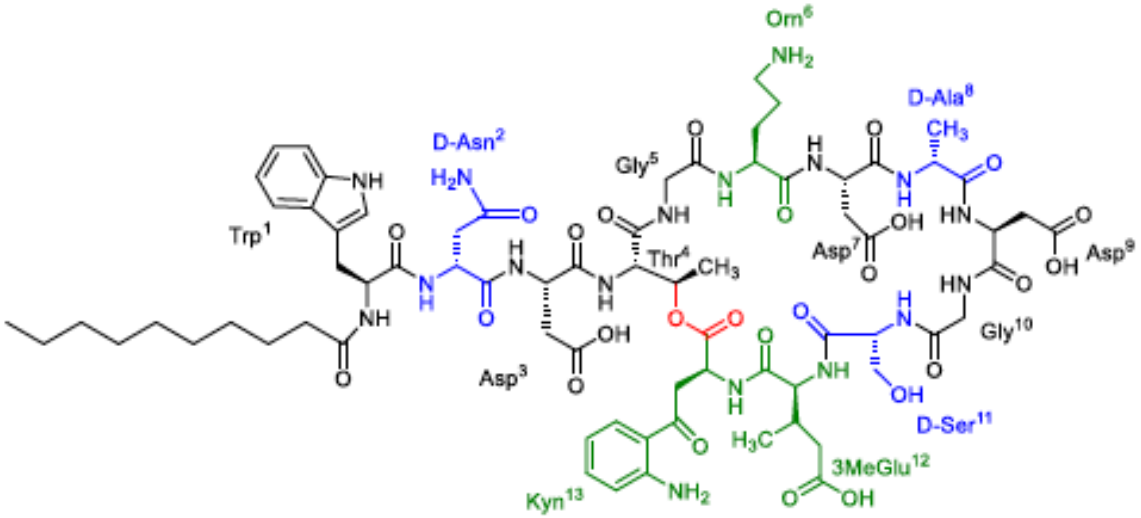
11: surfactin A



daptomycin

Daptomicina

La daptomicina è un lipopeptide acido prodotto da *Streptomyces roseosporus*. L'attività antibatterica della daptomicina è mediata dal calcio.

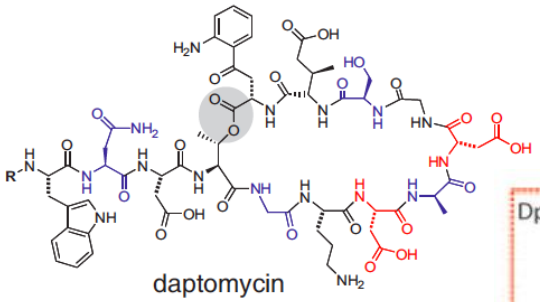


Daptomycin

La daptomicina è usata nel trattamento di patogeni resistenti per i quali ci sono poche alternative terapeutiche (enterococci resistenti alla vancomicina, *S. aureus* e *S. pneumoniae* resistenti alle penicilline)

Daptomicina sintetasi

La daptomicina è sintetizzata dalla NRPS DptA-BC-D. DptE e DptF catalizzano l'attivazione e l'incorporazione della catena di acido grasso.



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)

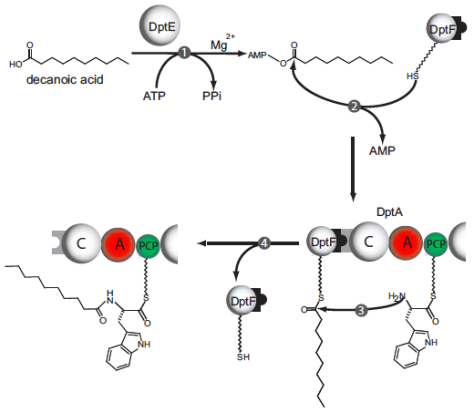
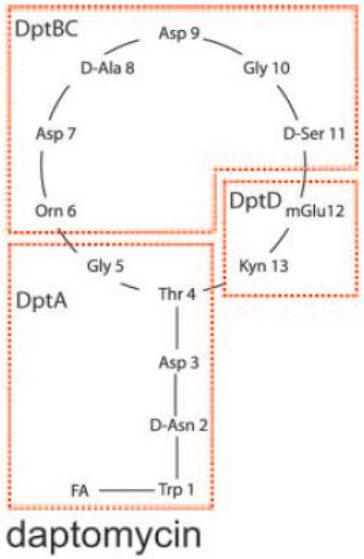
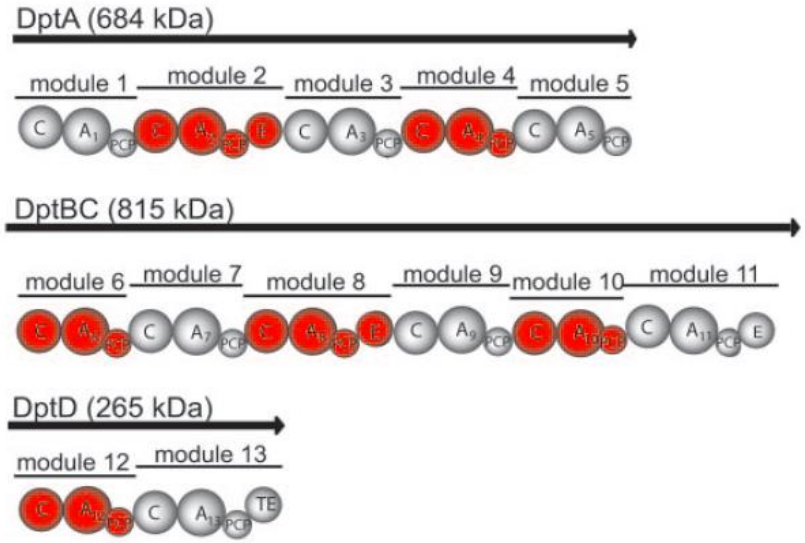


FIG. 5. Proposed mechanism of the lipidation of daptomycin. Decanoic acid is activated as decanoyl-adenylate under the consumption of ATP (1). This step is catalyzed by DptE. The fatty acid is transferred on to the ppan cofactor of the putative acyl-carrier protein DptF (2). DptF interacts with the starter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





Meccanismo d'azione della Daptomicina

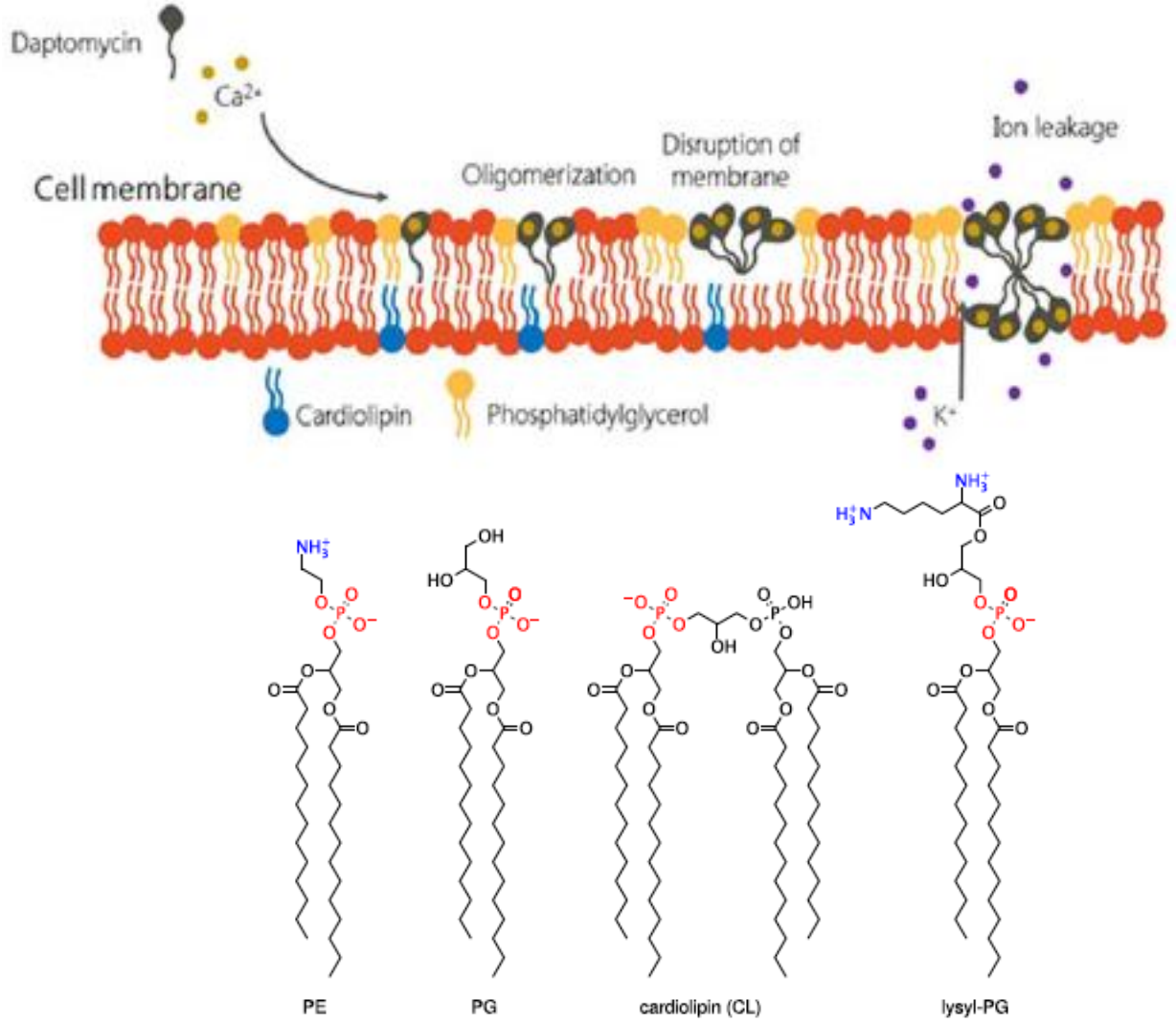


Figure 5. Major phospholipids of Gram-positive bacteria. Phosphatidylethanolamine (PE) is the major neutral phospholipid, whereas phosphatidylglycerol (PG) and cardiolipin (CL) are acidic; with CL, only one of the phosphate groups is usually deprotonated. Lysyl-PG carries a net positive charge. While 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.

Surfactina sintetasi

La Surfactina è un lipopeptide prodotto da *Bacillus subtilis* che contiene un β-idrossiacido di 13-15 atomi di C. Il più comune è l'acido 3-idrossi-13-metil-tetradecanoico

Geni che codificano per la surfactina sintetasi

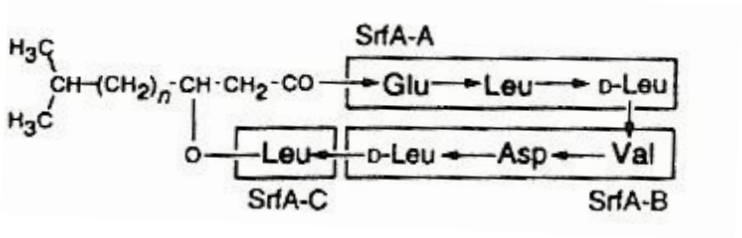
Operone SrfA

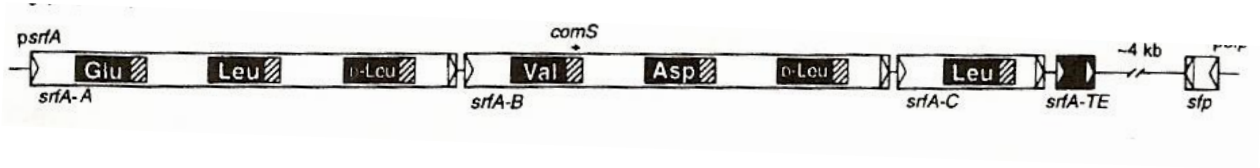
SrfA-A ⇨ E1A 400 kDa

SrfA-B ⇨ E1B 400 kDa

SrfA-C ⇨ E2 150 kDa

SrfA-D tioesterasi 25 kDa





Sfp associata a SrfA

indispensabile per la biosintesi della surfactina

4-fosfopanteteina trasferasi 26 kDa

Surfactina sintetasi

Purificate per gel-filtrazione tre frazioni con attività enzimatica

- SrfA-A E1A Glu-Leu-Leu (D)
- SrfA-B E1B Val-Asp-Leu (D)
- SrfA-C E2 Leu
- E3 attività aciltrasferasica: inizia la sintesi trasferendo β-HA- CoA al residuo di Glu su E1A. In assenza degli altri componenti funziona da tioesterasi. Non sembra essere codificato in SrfA, forse è correlato ai geni del metabolismo degli acidi grassi

  aciltrasferasi 40 kDa

Ingegnerizzazione della surfactina sintetasi

Spostamento modulo tioesterasico

Scambio di moduli con altre sintetasi

Surfactina sintetasi

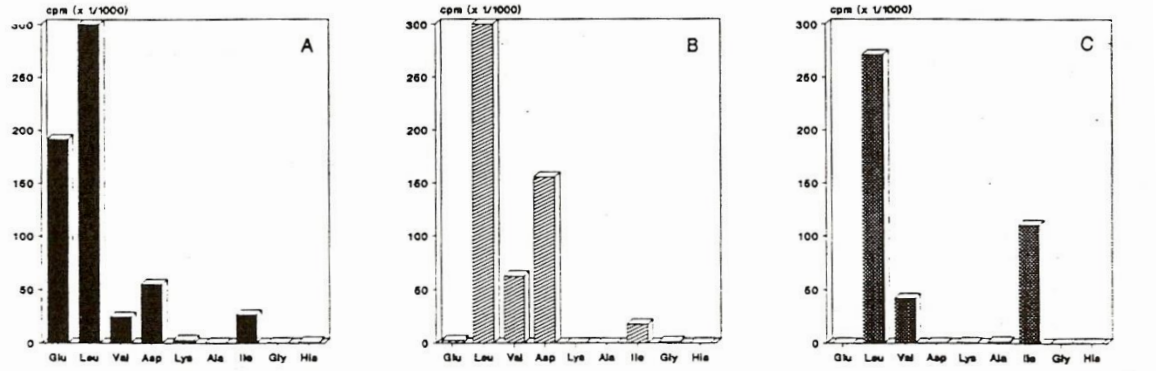


Fig. 4. ATP/PP_i exchange activity of srfAORF1, srfAORF2 and srfAORF3. The partially purified subunits were incubated in the presence of ATP, [^{32}P]PP_i and one amino acid and, after 10 min incubation at 37°C, the [^{32}P]ATP was adsorbed on Norite A (activated coal, Sigma) and counted.

Assegnazione specificità di SrfA-A, -B e -C: saggio di scambio ATP/PP_i e saggio di legame con amminoacidi marcati

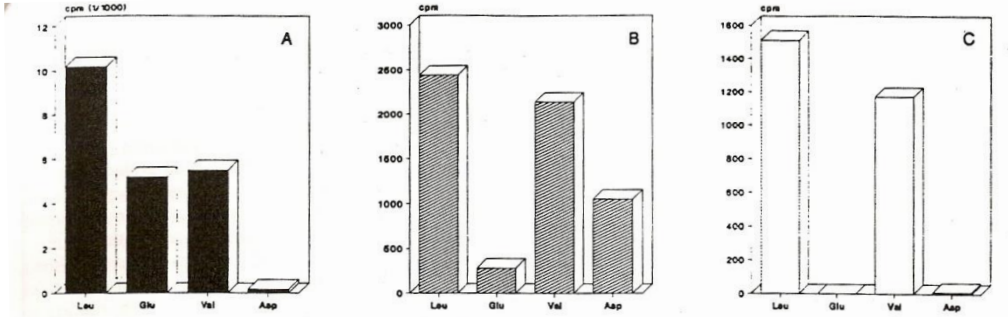


Fig. 5. Amino-acid binding activity of (A) srfAORF1, (B) srfAORF2 and (C) srfAORF3. Each enzyme subunit was incubated in the presence of ATP and one [^{14}C]-labelled amino acid. After 30 min incubation at 37°C, proteins were precipitated with 5% TCA, filtered through cellulose nitrate filters (MFS,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

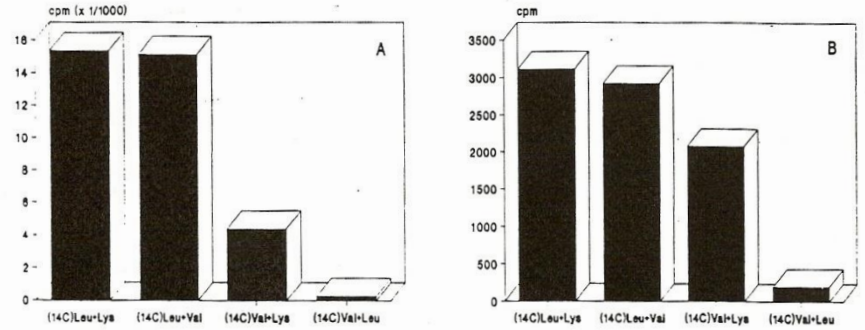


Fig. 6. Amino-acid binding competition experiments. Either srfAORF1 (A) or srfAORF3 (B) were incubated simultaneously with two amino acids, one of which ^{14}C-labelled. Samples were processed as described in Fig. 5. See text for details.

Surfactina sintetasi

Costruzione sintetasi ibride per scambio di moduli (domini AT) con GrsA, GrsB e ACV

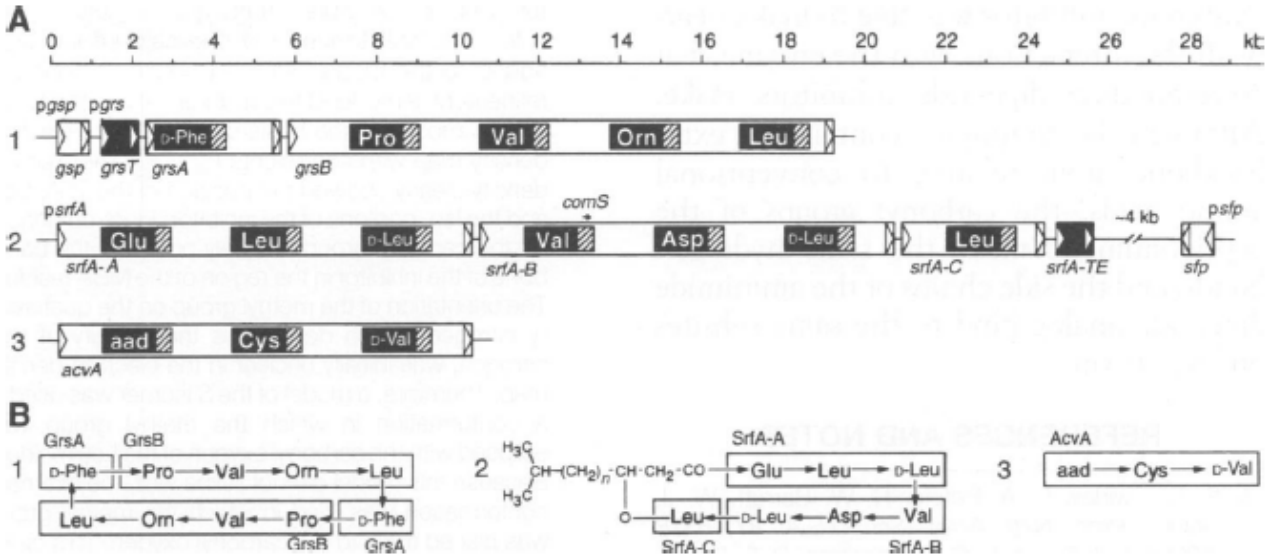


Fig. 1. Multidomain structure of peptide synthetases. (**A**) Schematic diagram of the highly conserved and ordered domain organization of peptide synthetases encoded by the bacterial operons *grs* (row 1) and *srfA* (row 2), and the fungal gene *acvA* (row 3) [aad, δ-(L-α-aminoadipyl)] (*3, 5, 6*). 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.

Surfactina sintetasi

Strategia per la produzione di Srf sintetasi ibride *in vivo* in *Bacillus subtilis* per alterare la specificità dell'ultimo modulo e inserire amminoacidi diversi da Leucina in posizione 7 nel peptide.

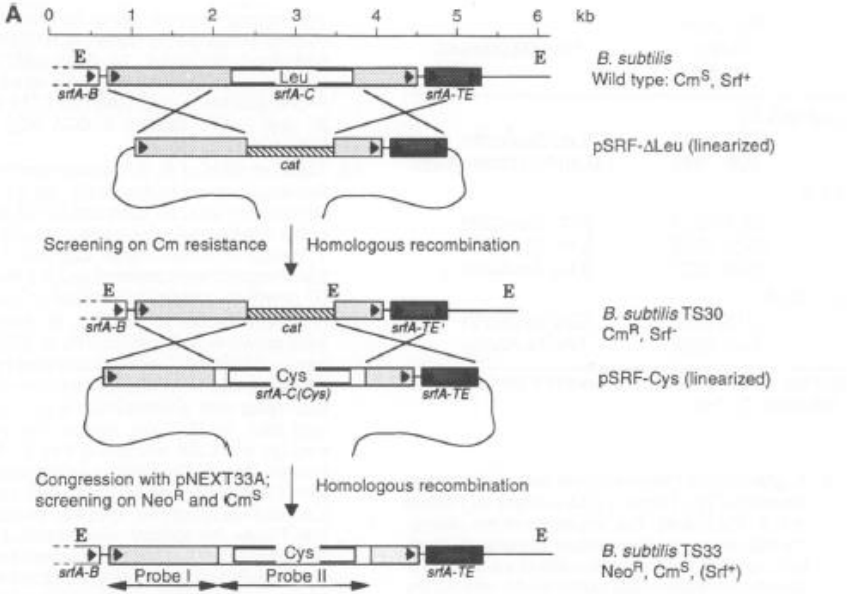


Fig. 3. Targeted *srfA-C* replacement in *B. subtilis*. (**A**) The integrative plasmids constructed (Fig. 2B) were used for in vivo *srfA-C* disruption and domain substitution within the chromosome of *B. subtilis*. Two steps were required: first, the *srfA-C* gene of *B. subtilis* (ATCC 21332) was 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.

Surfactina sintetasi

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*).

Si ottengono i prodotti previsti, ma le rese sono molto basse

- Selettività dominio C?
- Comunicazione tra moduli/domini?

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