Non-ribosomal biosynthesis of bioactive peptides

Main antibiotic targets

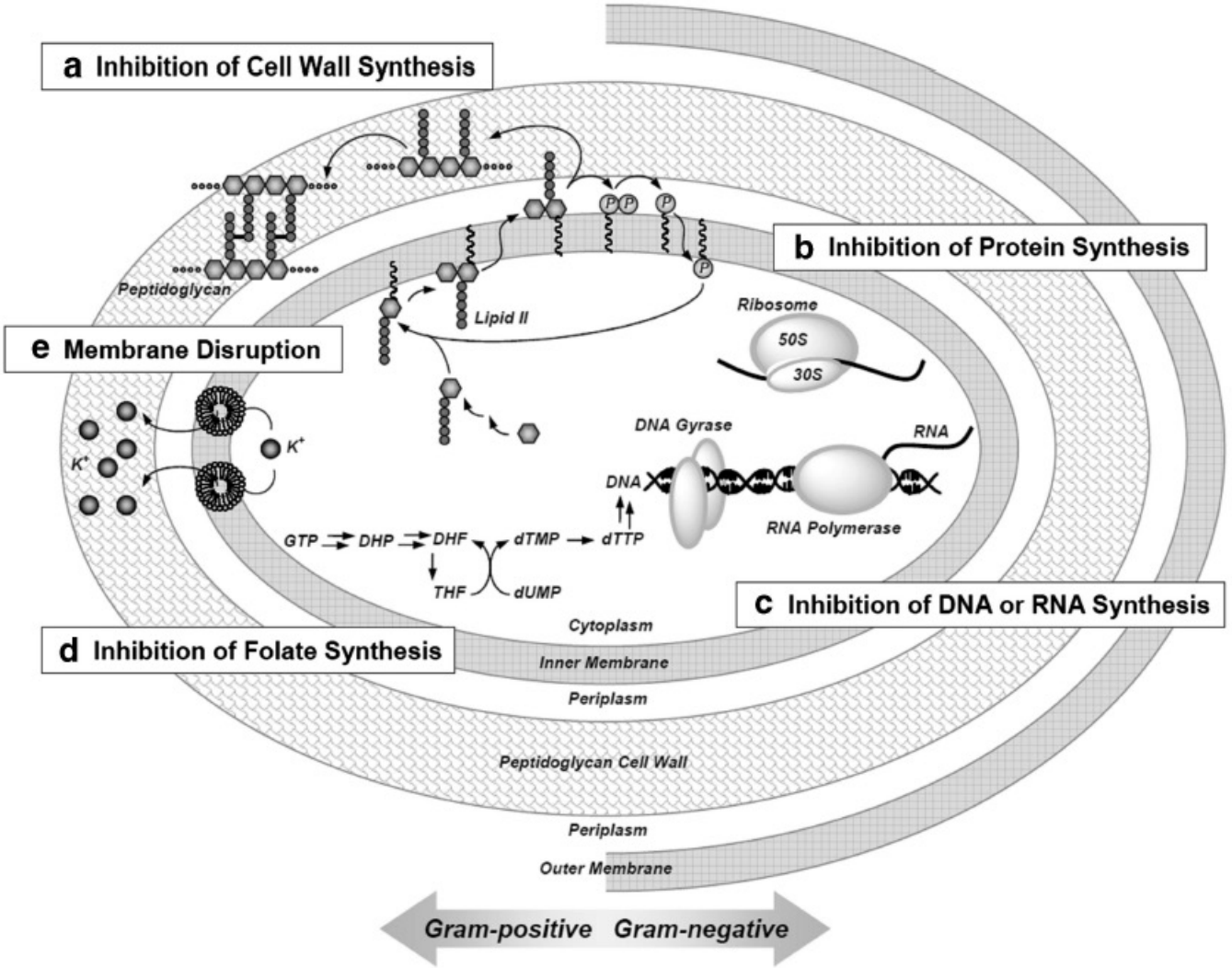


Figure 5 The five major clinically validated antibacterial targets/pathways. (**a**) Inhibition of cell wall biosynthesis. (**b**) Inhibition of protein synthesis. (**c**) Inhibition of DNA or RNA synthesis. (**d**) Inhibition of folate biosynthesis. (**e**) Disruption of membrane integrity. Modified from the cover of Walsh.[1]

Most antibiotics derive from natural compounds

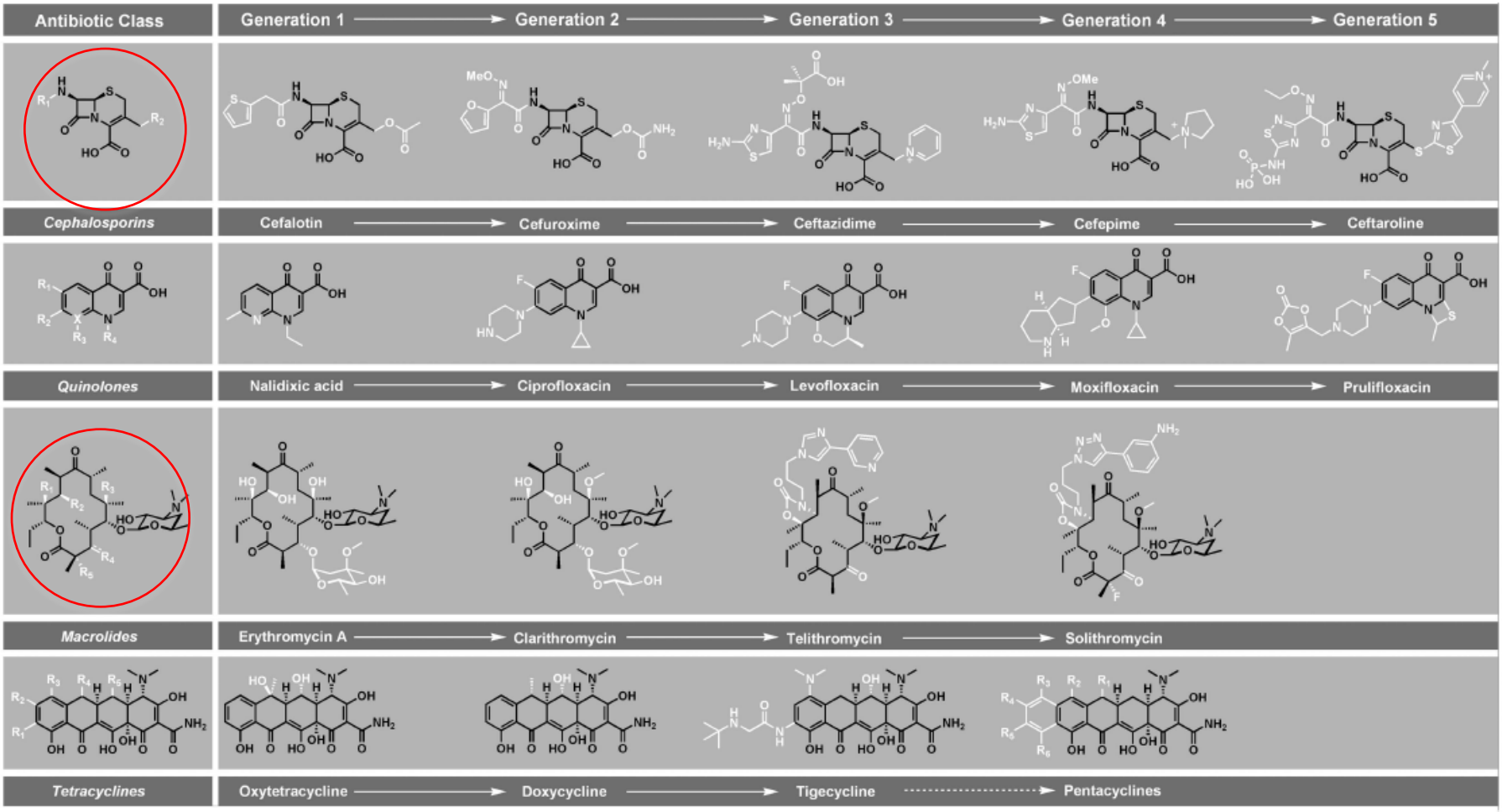
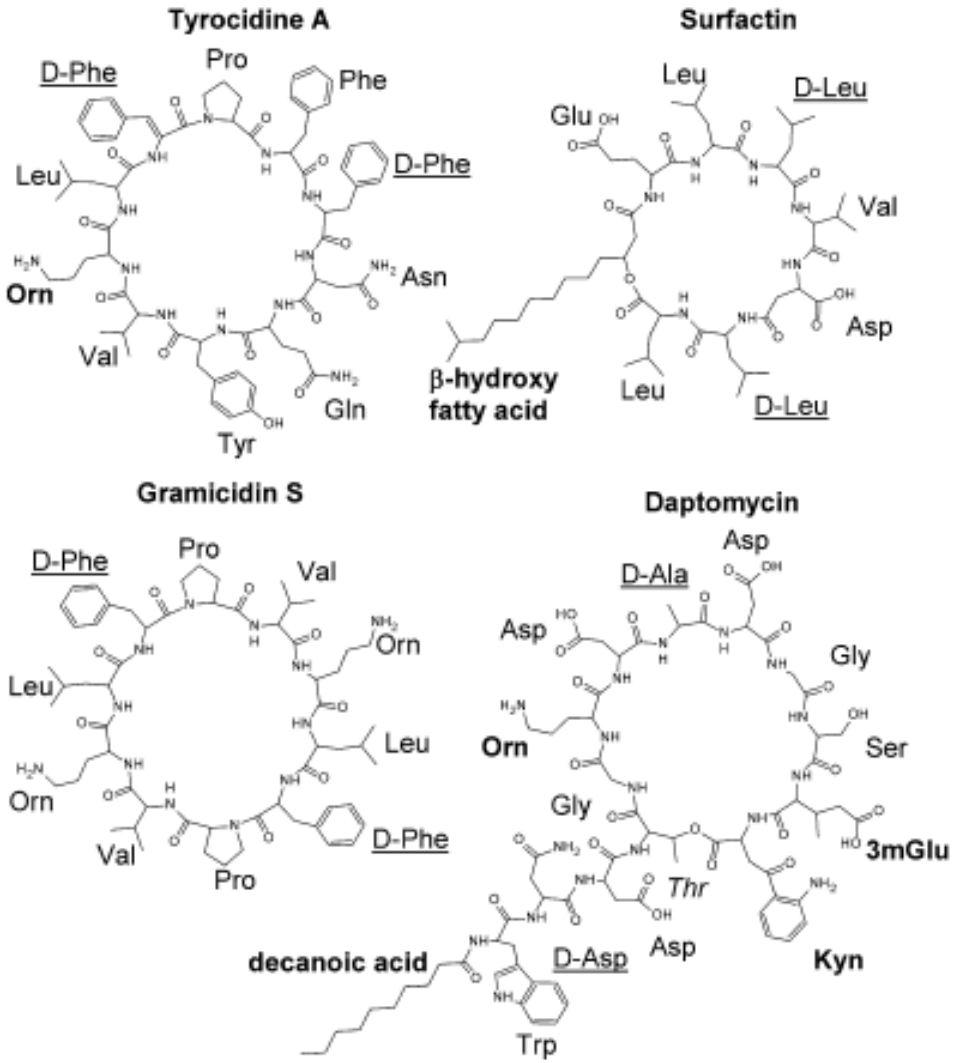


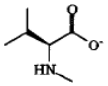
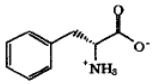
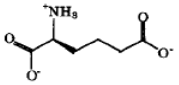
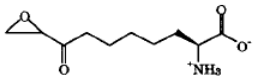
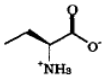
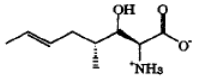
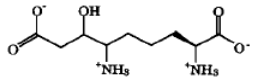
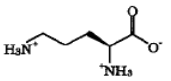
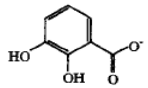
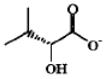
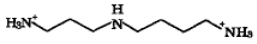
Figure 3 Synthetic tailoring of antibiotic core structures leading to successive generations of antibiotic classes has been the mainstay of antibiotic drug development for the past 50 years. New antibiotic core scaffolds are desperately needed. Core scaffolds are shown in black and peripheral chemical derivatizations are shown in white. The quinolone/fluoroquinolone scaffolds are of synthetic origin, while all other scaffolds are natural products or their semisynthetic derivatives.

Structures of bioactive peptides produced by bacteria and fungi



Many bioactive peptides contain non-canonical amino acids

Table 1. Nonproteinogenic Constituents of Peptide Antibiotics (Examples for Some Unusual Moieties)

Name	Structure	Abbreviation	System(s)	Organism(s)
modified, proteinogenic amino acids				
N-methyl aa (e.g. *N*-methyl valine)		MeVal	cyclosporin enniatin	*Tolypocladium niveum* *Fusarium scirpi*
D-aa (e.g. D-phenylalanine)		D-Phe	e.g. bacitracin gramicidin S tyrocidine	*Bacillus licheniformis* *Bacillus brevis* *Bacillus brevis*
non-proteinogenic amino acids				
δ-(L-α-amino adipic acid)		Aad	ACV-tripeptide (precursor of penicillin and cephalosporin)	*Penicillium chrysogenum* *Aspergillus nidulans* *Streptomyces clavuligerus*
2-amino-9,10-epoxy-8-oxodecanoic acid		Aeo	HC-toxin	*Cochliobolus carbonum*
L-α-amino butyric acid		Abu	cyclosporin	*Tolypocladium niveum*
(4R)-4[(E)-2-butenyl-4-methyl-L-threonine]		Bmt	cyclosporin	*Tolypocladium niveum*
2,6-diamino-7-hydroxy-azealic acid		Dha	edeine	*Bacillus brevis*
ornithine		Orn	e.g. bacitracin gramicidin S tyrocidine	*Bacillus licheniformis* *Bacillus brevis* *Bacillus brevis*
carboxy acids				
2,3-dihydroxy benzoic acid		Dhb	enterobactin	*Escherichia coli*
D-α-hydroxyisovaleric acid		Hiv	enniatin B	*Fusarium scirpi*
amines				
spermidine		Sperm	edeine	*Bacillus brevis*

How are these peptides produced?

- They have complex (often cyclic) structures and they contain modified or non-canonical amino acids.
- Their synthesis requires ATP, but it proceeds also in the absence of nucleic acids (DNA/RNA) or in the presence of protein synthesis inhibitors.

Non-ribosomal peptide synthesis is catalyzed by large multifunctional enzymes: non-ribosomal peptide synthetases (NRPS)

Strategies and methods for the study of secondary metabolites

- Which are the biosynthetic precursors?
- Which is the sequence of reactions and which are the biosynthetic intermediates?
- Which enzymes are involved in biosynthesis and which is their reaction mechanism?
- Which genes are important for production and which is their role?

Protein ⇨ Gene

Gene ⇨ Protein

- Feeding of labeled hypothetical precursors (e.g. with radioactive isotopes) to growing cells or 'cell-free' systems, isolation of the product and analysis of position of the label
- Production of 'blocked' mutants that accumulate the intermediate.
- A labeled intermediate is produced and the incorporation in the metabolite is measured (in the wild type strain or in 'cell-free' systems)
- Purification of enzymes and activity assay
- Structural and functional characterization of enzymes
- Identification of genes coding for enzymes involved in biosynthesis
- Cloning and expression of proteins involved in biosynthesis
- Regulation?

Peptide synthetases (NRPS)

- NRPS are large multienzymatic complexes
- NRPS have a repetitive structure
- Genes involved in the synthesis of the peptide are organized in clusters
- The reaction mechanism of NRPS requires SH groups and ATP
- Informations obtained on the protein together with identification of the NRPS genes have allowed to understand the mechanism of action

Gene organization of NRPS

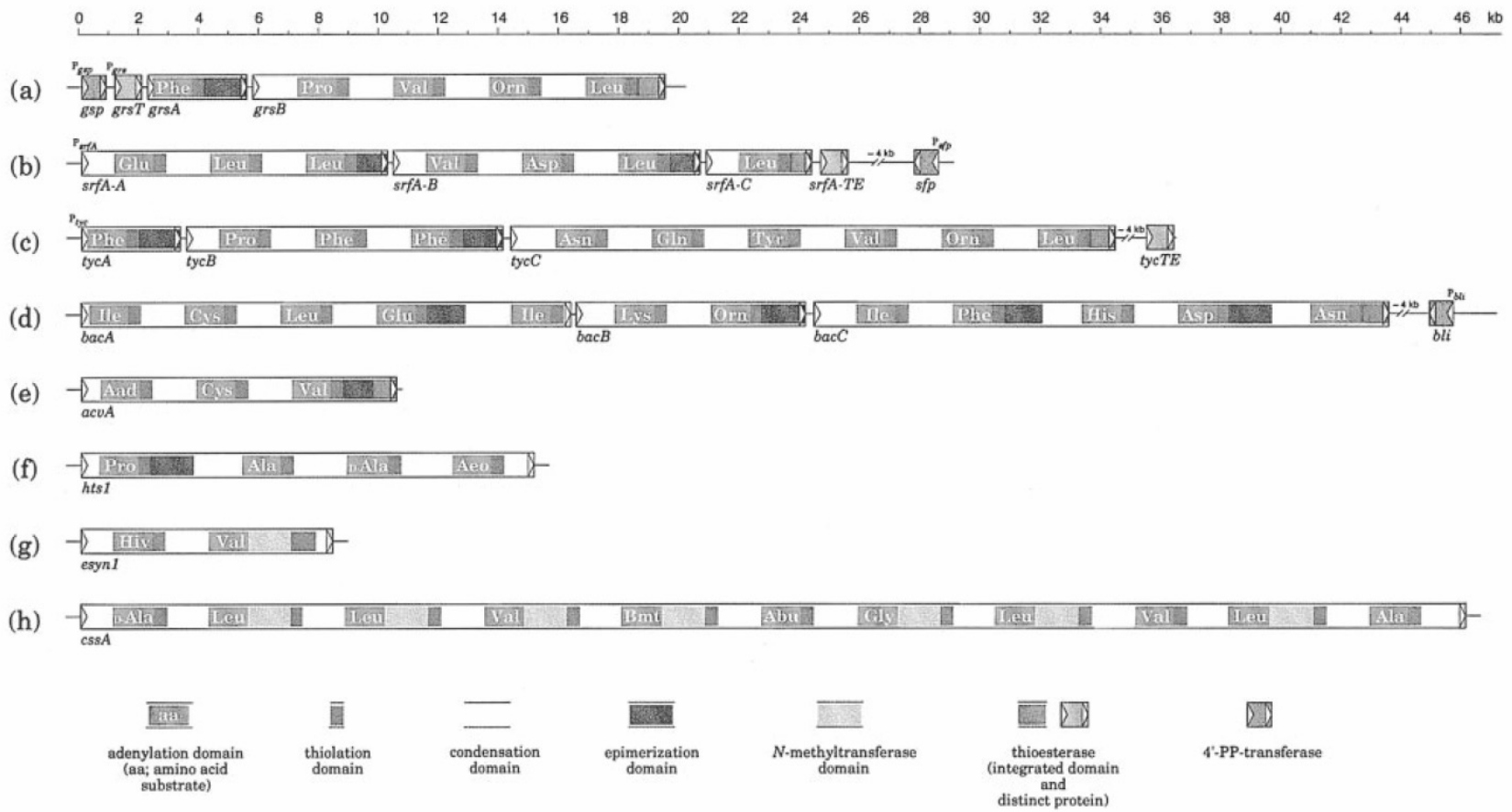
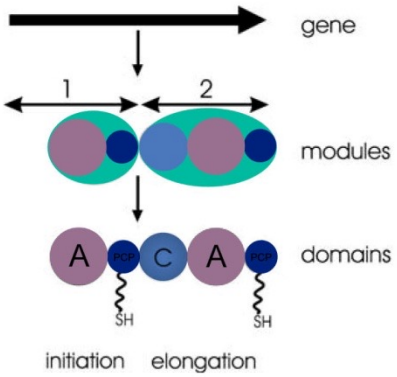


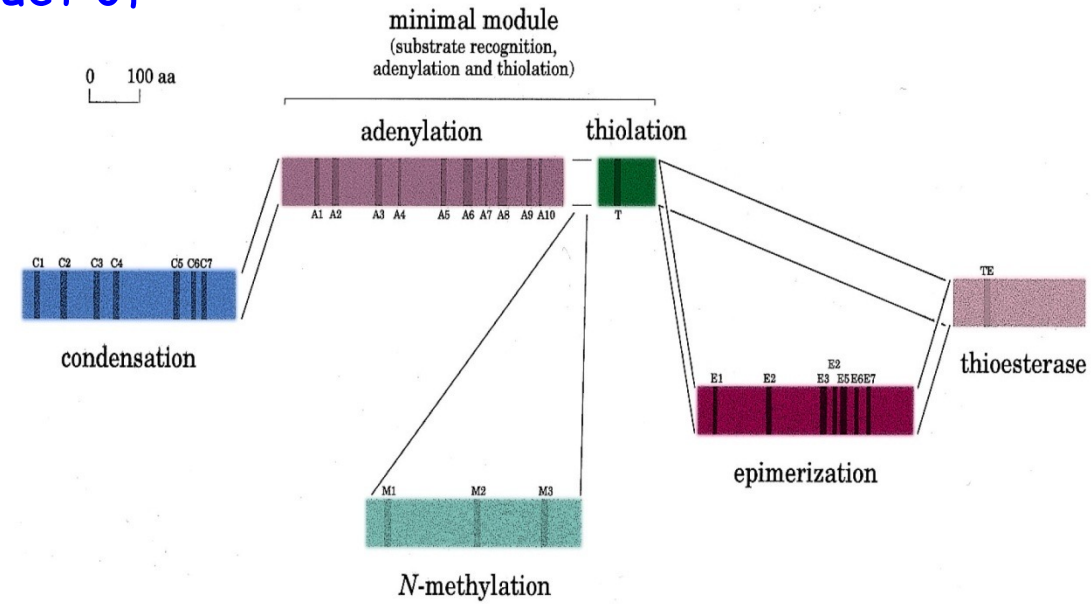
Figure 3. Schematic diagram showing the modular organization of peptide synthetases encoded by the bacterial operons *grs* (gramicidin S; row a), *srfA* (surfactin; row b), *tyc* (tyrocidine; row c) and *bac* (bacitracin; row d) as well as the fungal genes *acvA* (δ-(L-α-aminoadipyl)-L-Cys-D-Val; row e), *hts1* (HC-toxin; row f), *esyn1* (enniatin; row g) and *cssA* (cyclosporin A; row h). Red boxes indicate the amino acid adenylation domains, and green strips show the location of the thiolation domain, the site of cofactor 4′-PP-binding. Modules associated with the activation and incorporation of D-amino acids (GrsA, TycA, SrfA-A-Leu, SrfA-B-Leu, TycB-Phe, BacA-Glu, BacB-Orn, BacC-Phe and BacC-Asp, AcvA-Val, and Hts1-Pro) have an epimerization domain (blue regions) located downstream of the thiolation domain. The first module of the *cssA* gene (D-Ala) and the third one in the *hts1* gene (D-Ala) have no epimerization modules; however, it incorporates D isomers that are supplied by external racemases. A fourth type of domain (yellow regions) is only found in peptide synthetases of fungal origin (*esyn1* and *cssA*) that are associated with the activation and incorporation of *N*-methylated amino acids. These putative *N*-methyltransferases have a cosubstrate SAM-binding site and are inserted between the adenylation and the thiolation domains. The intermodular regions shown as white boxes represent putative elongation domains. Also shown is the location of gene-(fragment)s encoding the thioesterase-like proteins (pale pink) and thioesterase-like domains (pink) found associated with terminal modules in bacterial operons (*grs*, *srfA*, *tyc*, and *bac*). Genes that encode the 4′-PP transferase, located outside the operons *grs*, *srfA*, and *bac*, are also displayed (gray regions).

Modular organization of NRPS

The amino acid sequence of the peptide is determined by the sequence of modules in the enzyme structure, which depends on the gene sequence encoding the different modules. There is a colinearity between the product of the gene and the product of the enzyme.



Each module is organized in domains and it catalyzes insertion of a single amino acid in the peptide



Analogies between ribosomal and non-ribosomal biosynthesis

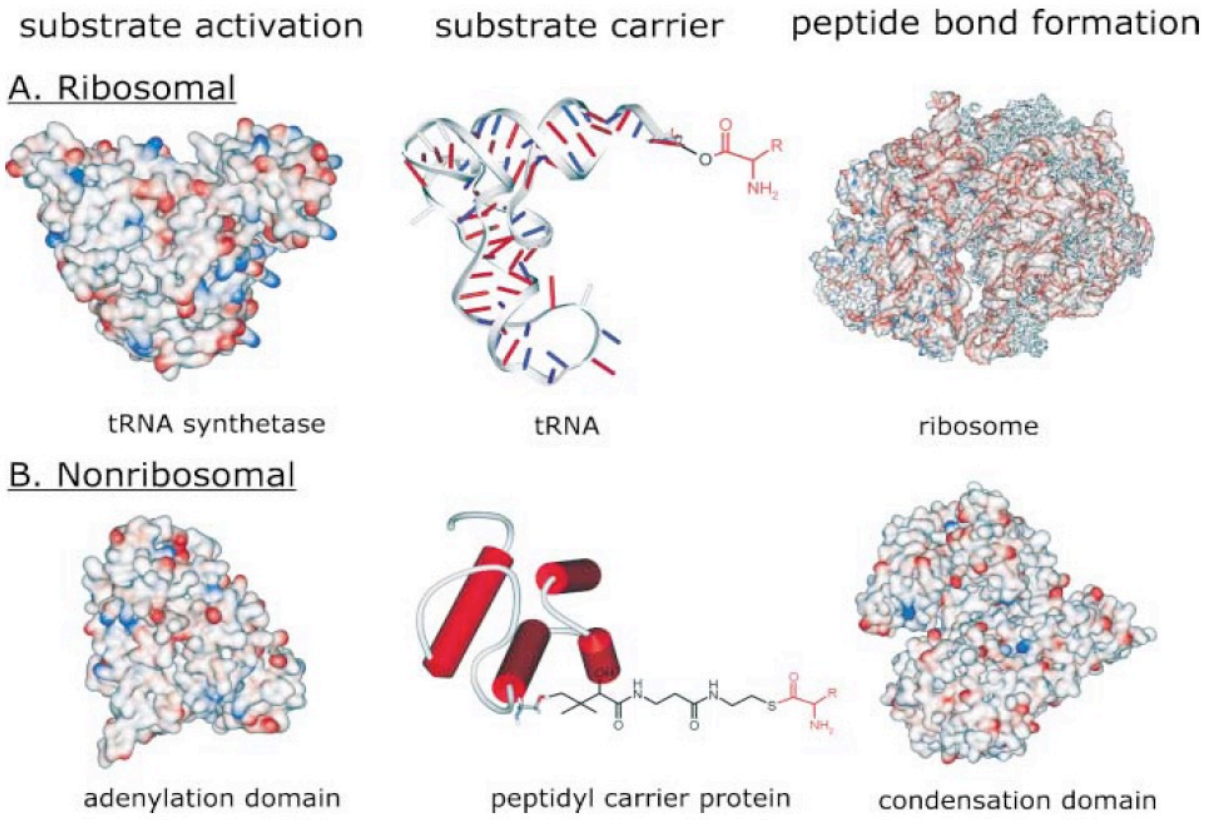
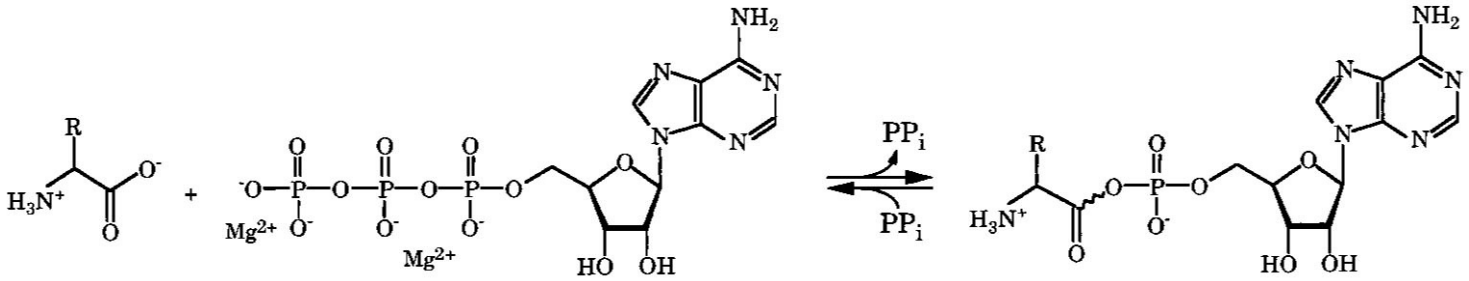


Figure 1 (*A*) Enzymes of the ribosomal machinery for the production of peptides and proteins. t-RNA synthetases activate amino acids as acyl adenylate and load it onto the t-RNA, which is subsequently directed to the ribosome for peptide bond formation with an amino acid tethered to a second t-RNA. (*B*) The basic enzymes of the nonribosomal system. The A-domain, although structurally unrelated to the t-RNA synthetase, activates amino acids as acyl adenylate. The activated amino acid is transferred to the cofactor of the PCP, and the C-domain forms peptide bonds between amino acids loaded onto two different PCPs.

Adenylation domain (*A*)

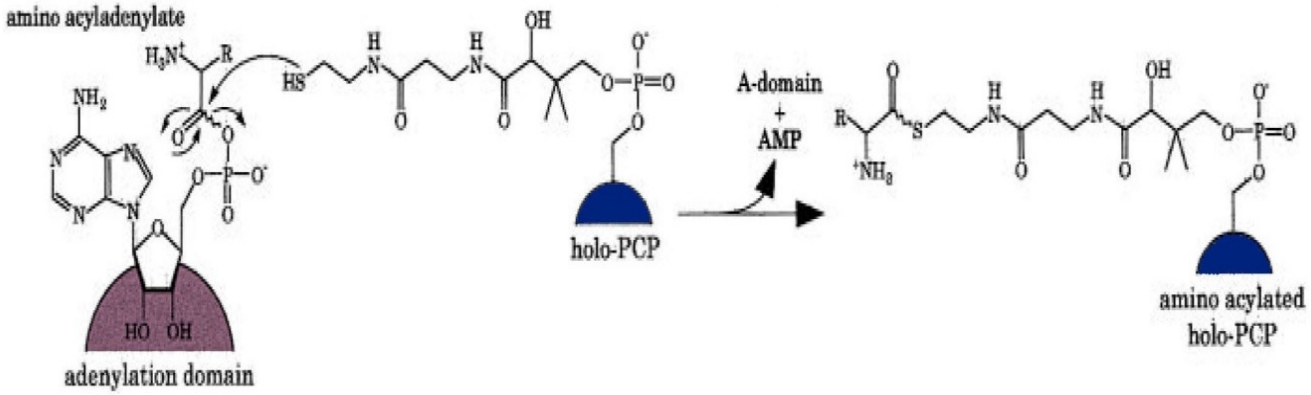
The adenylation domain is formed by about 550 amino acids and it performs recognition and activation of the amino acid substrate in an aminoacyl-adenilation reaction, with ATP hydrolysis and formation of aminoacyl-AMP.



amino acid ATP aminoacyl adenylate

Thiolation domain (T or PCP)

The thiolation domain (T) or peptidyl carrier protein (PCP) is formed by about 100 amino acids and it transports the amino acid by covalently binding aminoacyl-AMP to 4'-phosphopantetheine (4'-PP) via formation of a thioester bond and release of AMP.



The thiolation domain must be activated by phosphopantetheinyl-transferases that use Coenzyme A as 4'-PP donor and catalyze binding of the cofactor on a conserved serine residue.

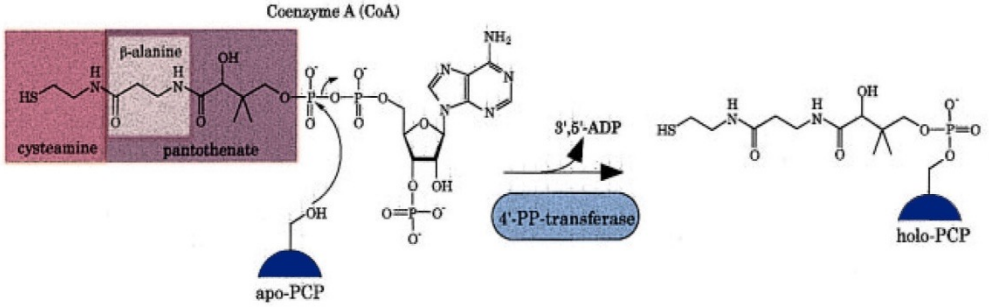


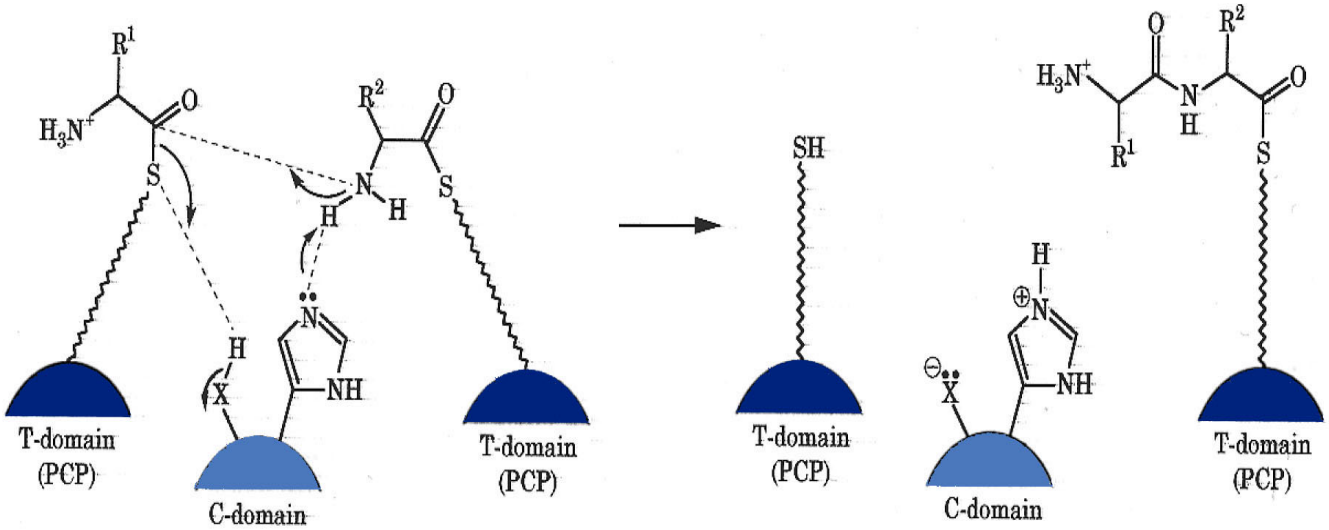
Table 4. Enzyme Superfamily of Acyl/Peptidyl Carrier Proteins (ACPs/PCPs): Sequence Alignment around the Highly Conserved Cofactor 4'-PP Binding Site

Enzyme	Organism	Position (aa)	Sequence [#]
A) Peptide synthetases			
TycA	*Bacillus brevis*	553	D N F Y S L G G H **S** I Q A I Q V
GrsB	*Bacillus brevis*	2033	D N F F E L G G H **S** L R A M T M
SrfA-B	*Bacillus subtilis*	990	D N F F M I G G H **S** L K A M M M
AcvA	*Penicillium chrysogenum*	3049	D D L F K L G G D **S** I T S L H L
Hts1	*Cochliobolus carbonum*	2405	S D F F S S G G N **S** M A A I A L
CssA	*Tolypocladium niveum*	13645	D N F F E L G G H **S** L L A T K L
B) Acyl carrier proteins			
Polyketide synthases			
Act-ACP	*Streptomyces coelicolor*	33	L R F E D I G Y D **S** L A L M E T
Gra-ACP	*Streptomyces violaceoruber*	33	I T F E E L G Y D **S** L A L M E S
Fatty acid synthases			
FAS-ACP	*Escherichia coli*	28	S F V E D L G A D **S** L D T V E L
FAS-ACP	*Saccharomyces cervisiae*	73	Q F H K D L G L D **S** L D T V E L
		consensus	L G x(HD)S L

[#] Sequence data are derived from: TycA [54], GrsB [32], SrfA-B [33]; AcvA [29], CssA [35], Hts [36], Act-ACP [104], Gra-ACP [105], FAS-ACP from *E. coli* [103], FAS-ACP from yeast [102].

Condensation domain (C)

The condensation domain is formed by about 450 amino acids, it is localized upstream of the adenylation domain and it catalyzes formation of the peptide bond between amino acids tethered on two adjacent thiolation domains.



Thio-template mechanism of NRPS

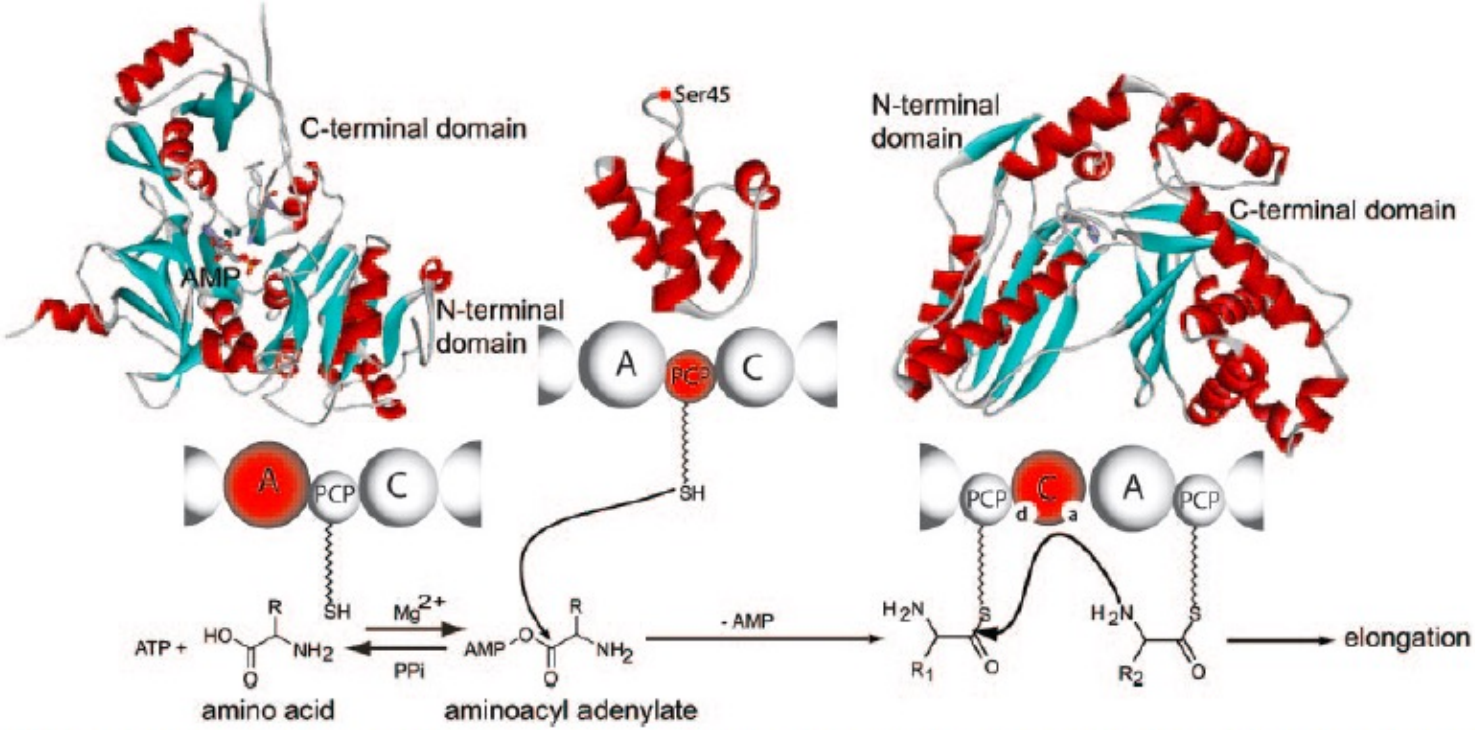


FIG. 4. Chemical principles of nonribosomal peptide synthesis. Domains in action are indicated in red and the respective crystal structures are shown above. First, the A-domain specifically recognizes a dedicated amino acid and catalyzes formation of the aminoacyl adenylate under consumption of ATP.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

How was the NRPS thio-template mechanism demonstrated?

Identification of the covalent binding site for the amino acid substrate(s).

- It is necessary to purify the enzyme.
- It is necessary to have an enzymatic activity assay to identify fractions that contain the enzyme.

Enzyme activity assays for NRPS

1. Amino acid activation

Amino acid + ATP ↔ amino acyl-AMP + $^{32}PP_i$

The assay is based on ATP-$^{32}PP_i$ exchange: formation of ^{32}P-labeled ATP takes place in the presence of the enzyme and of amino acids recognized as substrates.

The enzyme is incubated with substrates and then excess $^{32}PP_i$ is added. ATP that forms is separated from the reaction mixture by adsorption on activated charcoal and radioactivity is measured.

Enzyme activity assays for NRPS

2. Formation of the peptide/Covalent bond of the amino acid to the enzyme

Precursor amino acids are supplied, at least one is radioactively labeled (^{14}C or ^{3}H) and radioactivity incorporated in the peptide is measured.

The peptide and enzyme are separated from the reaction mixture by precipitation with trichloroacetic acid (TCA). TCA makes protein and peptide precipitate, but not free amino acids so radioactivity associated with non-incorporated amino acids will remain in the supernatant.

How is radioactivity measured?

Tabella 5.5 Principali vantaggi e svantaggi degli emettitori β più comunemente usati

Isotopo	Vantaggi	Svantaggi
^{3}H	Sicurezza Possibilità di avere una attività specifica elevata Ampia scelta di posizioni nei composti organici Risoluzione molto elevata nell'autoradiografia	Bassa efficienza di rilevazione Scambio isotopico con l'ambiente Effetto isotopico
^{14}C	Sicurezza Ampia scelta di atomi sostituiti nei composti organici Buona risoluzione nell'autoradiografia	Bassa attività specifica
^{35}S	Alta attività specifica Buona risoluzione nell'autoradiografia	Tempo di dimezzamento corto Tempo di dimezzamento biologico relativamente lungo
^{32}P	Facilità di rivelazione Alta attività specifica Tempo di dimezzamento breve che semplifica l'eliminazione degli scarti Conteggio Cerenkov	Tempo di dimezzamento breve che alza i costi e rende complessi i protocolli sperimentali Pericoloso per le radiazioni esterne Bassa risoluzione nell'autoradiografia

Tratto da R.J. Slater, *Radioisotopes in Biology. A Practical Approach*. IRL Press, Oxford

How is radioactivity measured?

In liquid scintillation counters the sample containing β-emitting radioisotopes is placed in liquid scintillation cocktails (a mixture of solvent and molecules that are excited by electrons emitted by the radioisotope). Light emission is measured (counted) by the instrument, and it will be proportional to the radioisotope concentration.

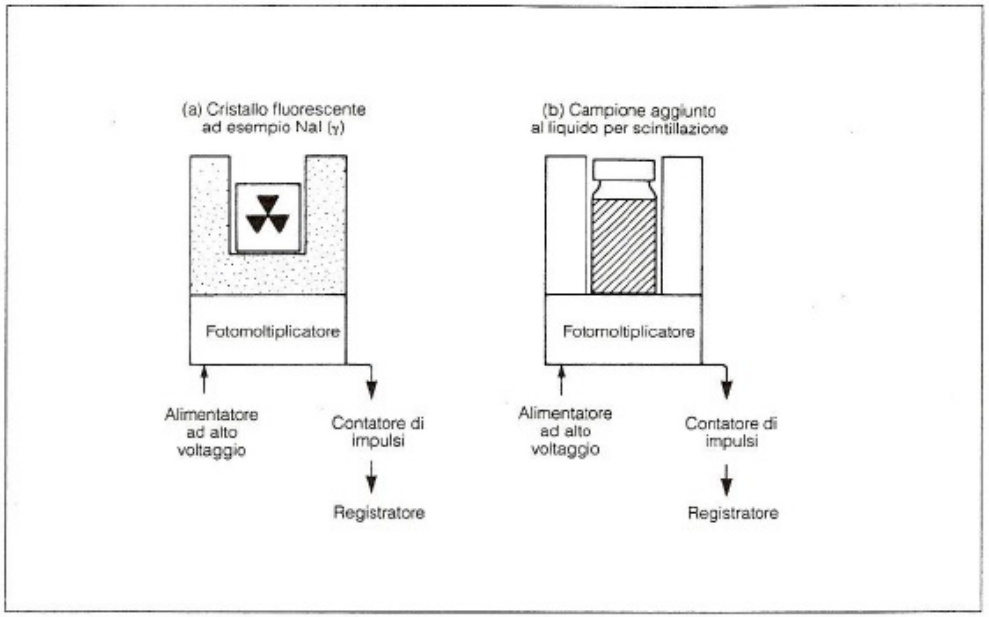


Figura 5.6 Schema dei metodi di conteggio (a) a scintillazione solida e (b) a scintillazione liquida.

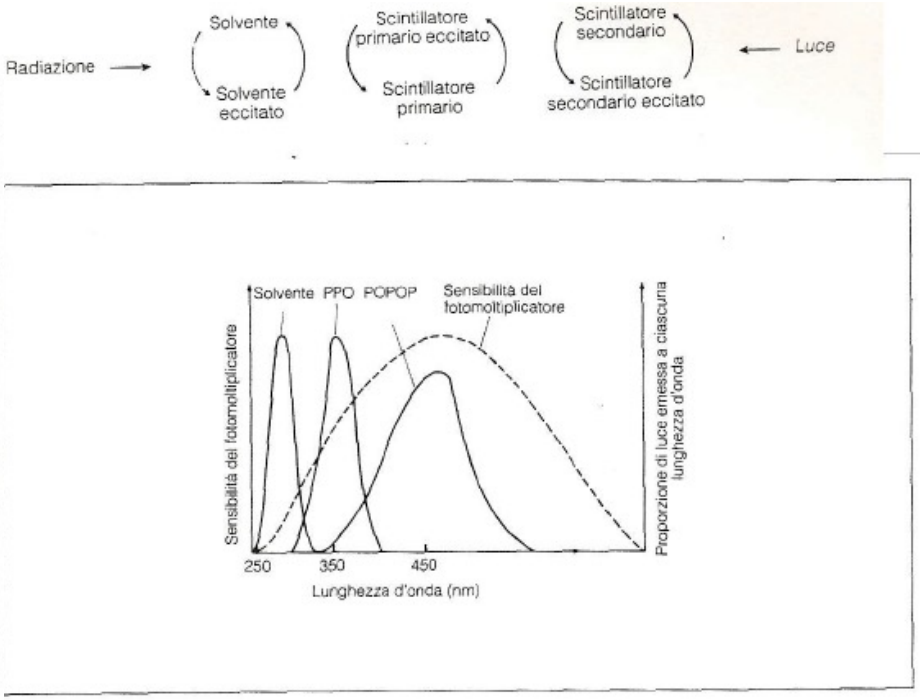
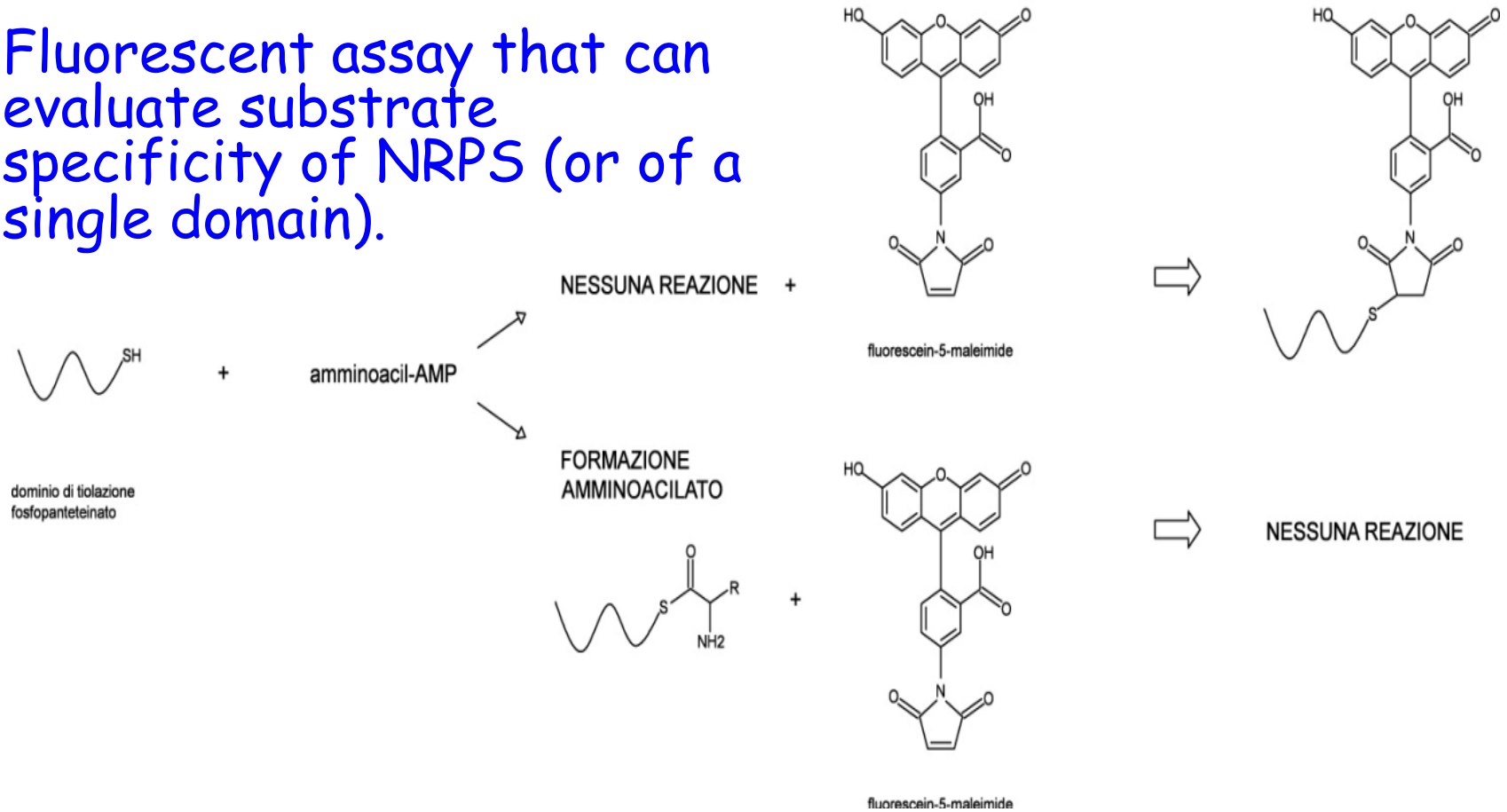


Figura 5.7 Spettri di emissione di alcuni scintillatori e sensibilità dei fotomoltiplicatori.

Enzyme activity assays for NRPS

3. Covalent bond of the amino acid to the enzyme

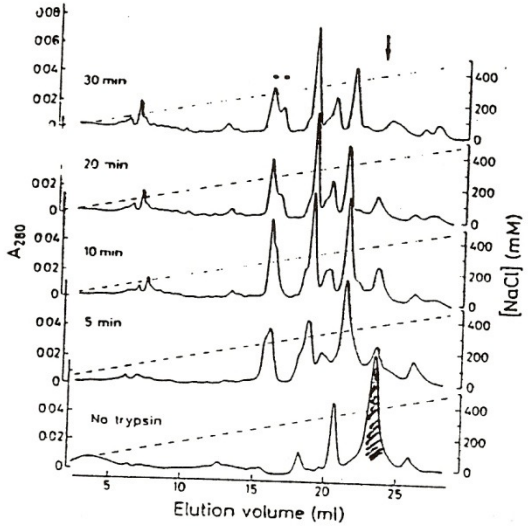
Fluorescent assay that can evaluate substrate specificity of NRPS (or of a single domain).



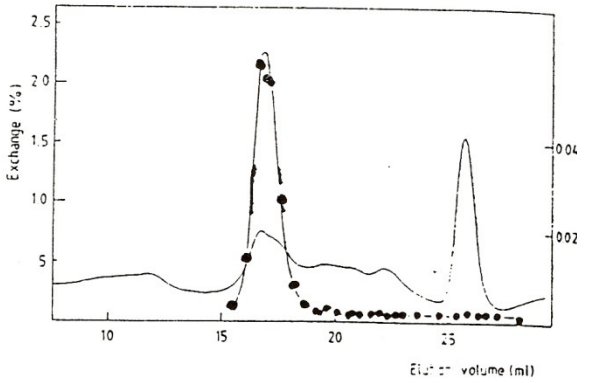
How was the NRPS thio-template mechanism demonstrated?

Identification of the covalent binding site for the amino acid substrate.

NRPS are large enzymes so it is necessary to employ limited proteolysis in native conditions to obtain single modules or domains that can be chromatographically separated and further characterized.



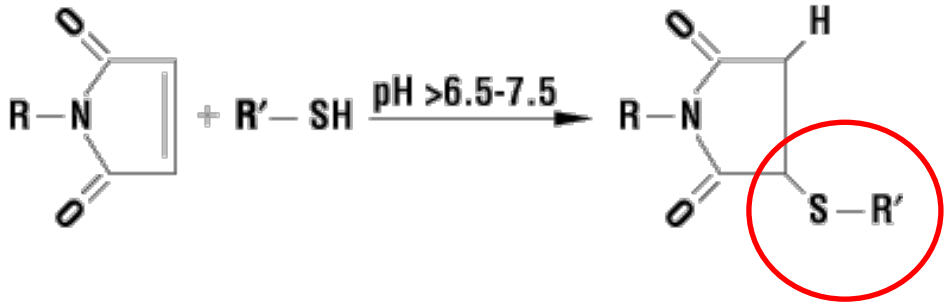
Limited proteolysis of Grs2 with trypsin to isolate the valine-binding domain. Ion-exchange chromatography



Gel-filtration on Superose •-• valine-dependent ATP/PP_i exchange

Identification of the covalent binding site for the amino acid substrate:
affinity labeling of the valine binding site on gramicidin S synthetase 2

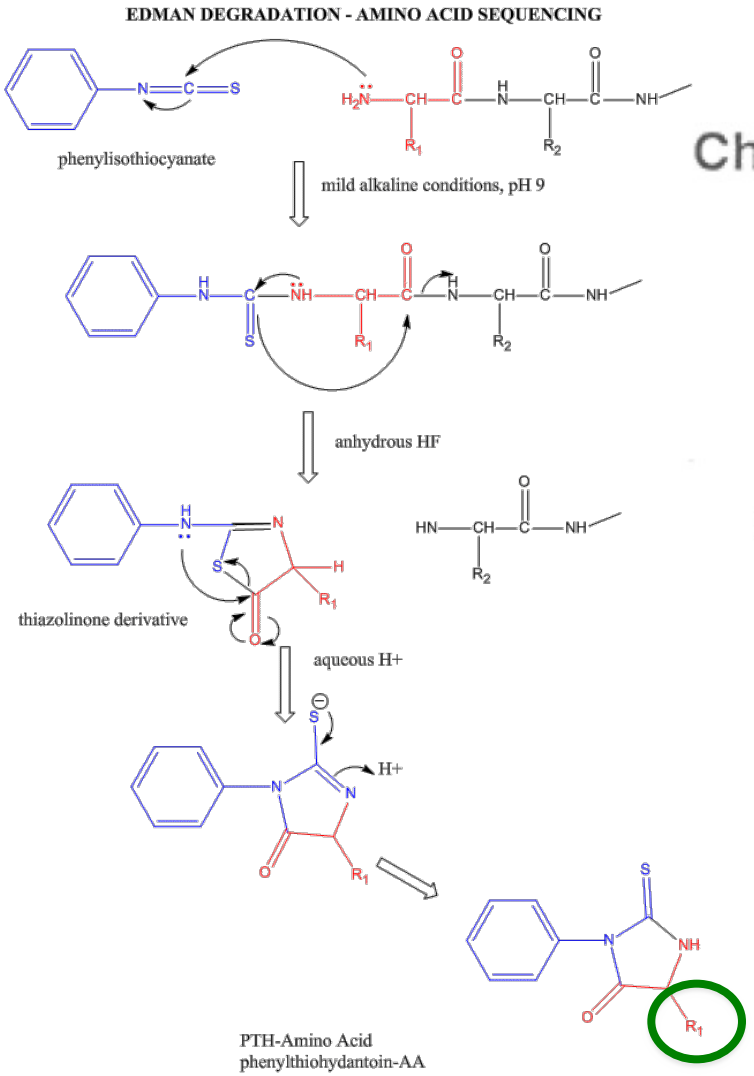
- Protection of the active site –SH group with valine
- Block of all other -SH groups with N-ethyl-maleimide (NEM)
- Gel-filtration to remove excess NEM
- Reduction with dithiotreitol (DTT) and gel-filtration to remove valine
- Labeling of the active site with [^{3}H] NEM

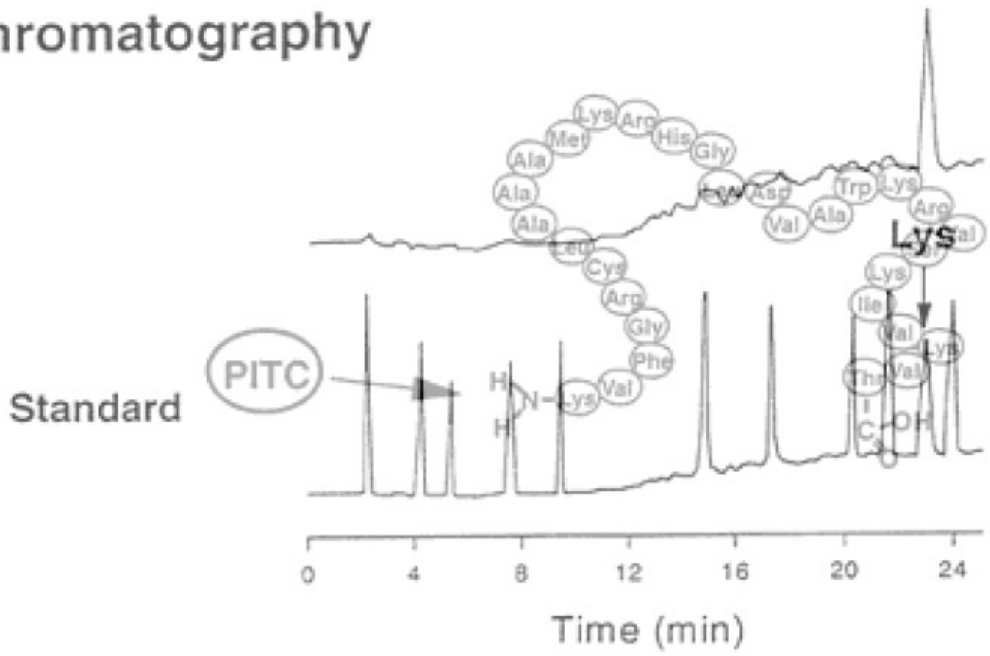


Identification of the covalent binding site for the amino acid substrate:
affinity labeling of the valine binding site on gramicidin S synthetase 2

- Digestion with trypsin to obtain protein fragments that can be analyzed
- Purification by chromatography of the [^{3}H] labeled peptide and sequencing by Edman degradation

Determination of amino acid sequences: Edman degradation





Identification of the covalent binding site for the amino acid substrate: affinity labeling of the valine binding site on gramicidin S synthetase 2

- Sequence obtained by Edman degradation of the labeled peptide:

  LGGH ΔA LR

  ΔA: dehydroalanine, it is derived from Cysteine or Serine

- But…. NEM is not directly bound to the enzyme otherwise it would have been identified because it forms a stable adduct with R-SH groups

Thio-template mechanism of NRPS

Experimental evidence in favour of the thio-template mechanism:

- -SH groups are required

  -SH inhibitors block activity

  a thioester bond is formed between the amino acid and the enzyme

  the thioester is unstable in the conditions of Edman degradation (the label is lost also if labeled valine is used)

- Information obtained on the **protein**: the amino acidic sequence of the active site peptide indicates the presence of cysteine or serine

  dehydroalanine

- Information obtained on the **gene**: the nucleotidic sequence of the active site peptide indicates the presence of serine

  codon for serine

Where does the SH group that forms the thioester come from???

Mass spectrometry reveals the presence of phosphopantetheine bound to the active site serine residue

TABLE I

Structure of the radiolabeled thiotemplate site peptide fragments of gramicidin S synthetase

Enzyme	Thio-template	Position of the fragment within the multienzyme	Thiotemplate site peptide fragments	Molecular mass (Dalton) Calculation from gene[a]	Calculation from Pan-adduct[b]	Results of ESI-MS[c]
GS1	Phe	D 564 - K 575	DNFYALGGDSIK └Pan-[^{3}H]NES	1299.4	1764.9	1764.4
GS2	L-Pro	I 983 - K1008	IWEEVLGISQIGIQDNFFSLGGHSLK └Pan-[^{3}H]NES	2888.4	3353.8	3352.4
	L-Val	I2029 - R2044	IGVLDNFFELGGHSLR └Pan-[^{3}H]NES	1774.0	2239.5	2239.3
	L-Orn	V3075 - K3090	VGIHDDFFTIGGHSLK └Pan-[^{3}H]NES	1742.9	2208.3	2208.0
	L-Leu	F4120 - L4132	FELGGHSLKATLL └Pan-[^{14}C]Leu	1385.6	1848.9	1849.1

[a] The molecular masses of the peptides were calculated from the gene-derived sequence (7, 8).

[b] Mass was calculated as the sum of the molecular masses of the peptide moiety, the 4′-phosphopantetheine substituent that is covalently attached to the serine residue, and the radioactively labeled tracers (shaded boxes, NES: *N*-ethylsuccinimido and L-leucine, respectively) bound to the reactive thiol group of the Pan cofactors.

[c] Results of the investigation of the active site peptide fragments by electrospray mass spectrometry (ESI-MS).

Strategy for labeling and purification of gramicidin S synthetase 1 and 2 active site peptides

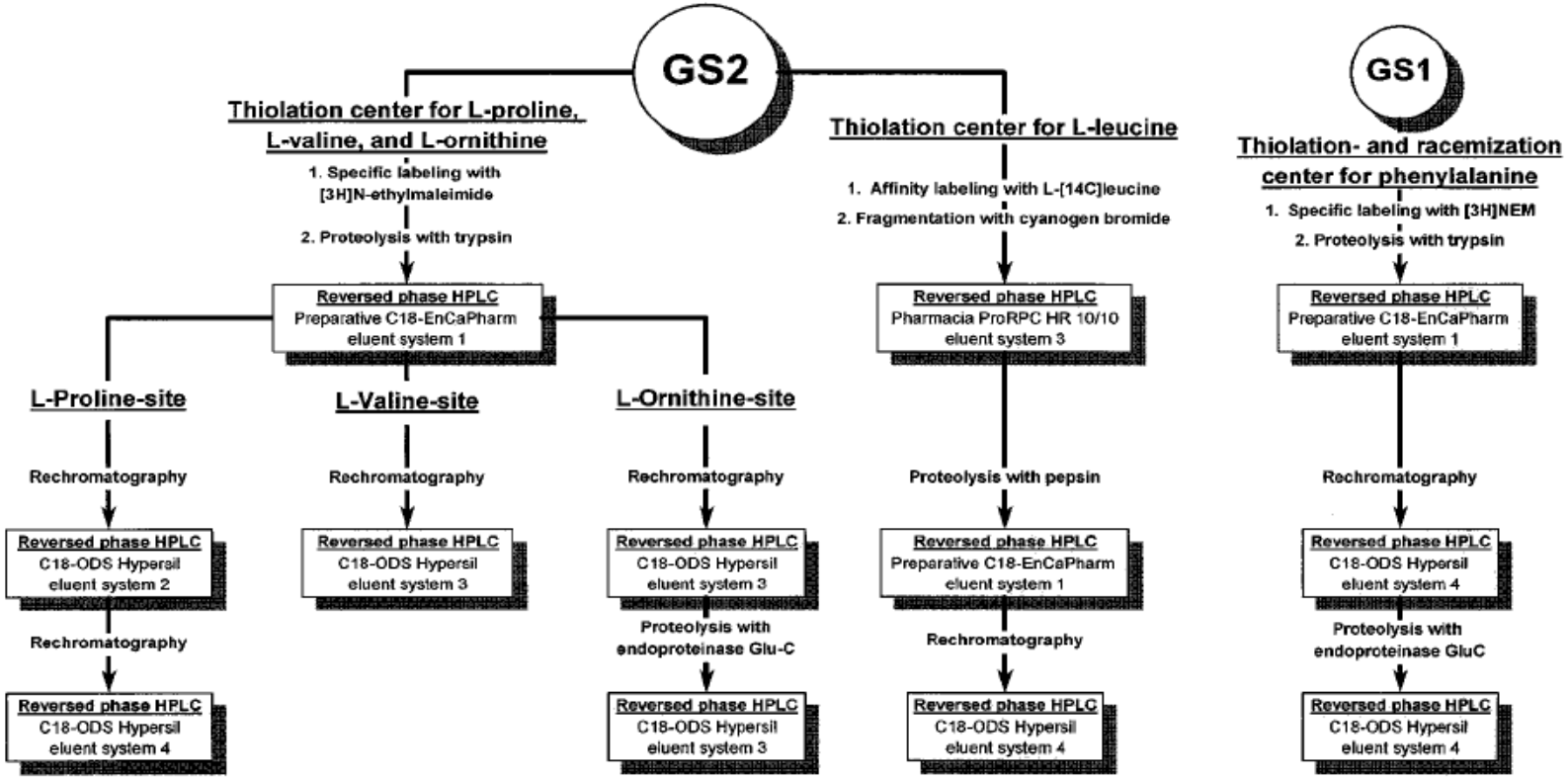
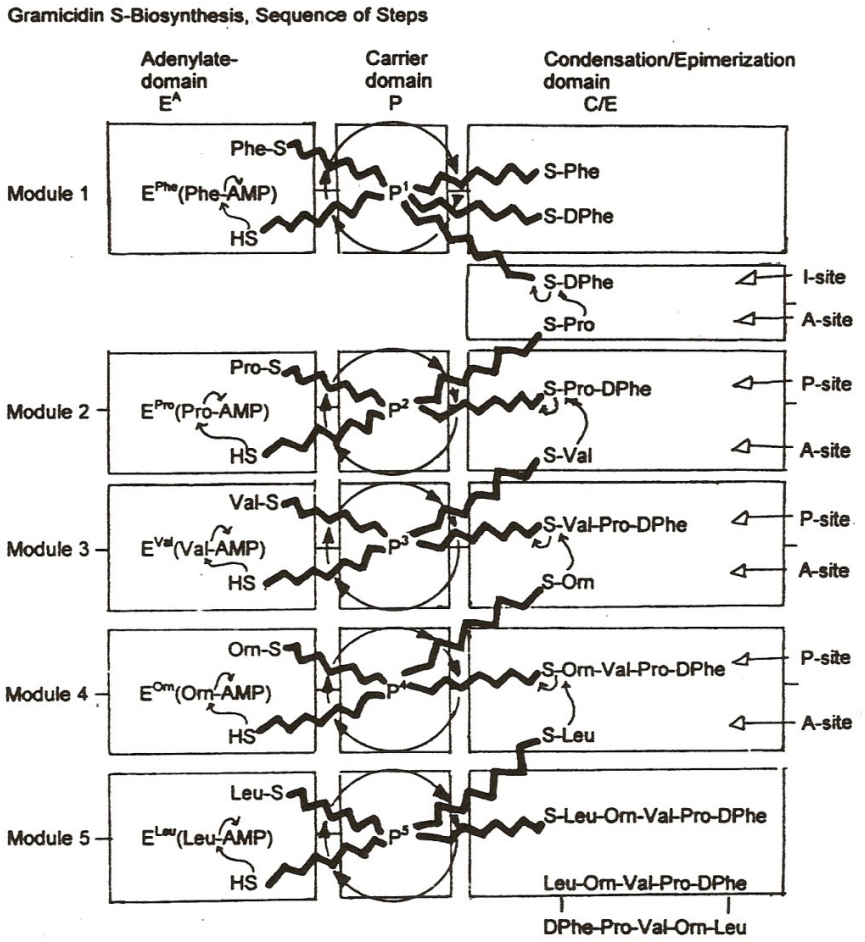
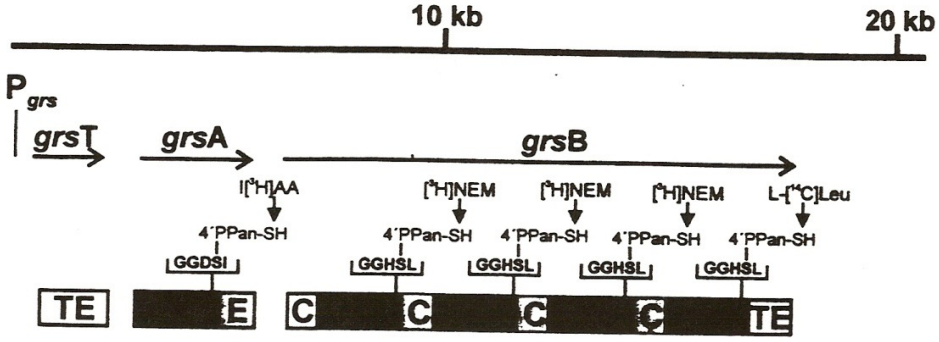


FIG. 1. Summary of the steps in the course of the purification of all thiotemplate site peptide fragments of gramicidin S synthetase 1 and 2.

Model of the reaction mechanism of NRPS: multitransporter thio-template mechanism. The 4'-PP arms facilitate transport of the substrate and of the various intermediates to the catalytic centers of the different domains.





Modifications of the peptide

- N-methylation
- Oxidation
- Halogenation
- Heterocyclization
- Epimerization

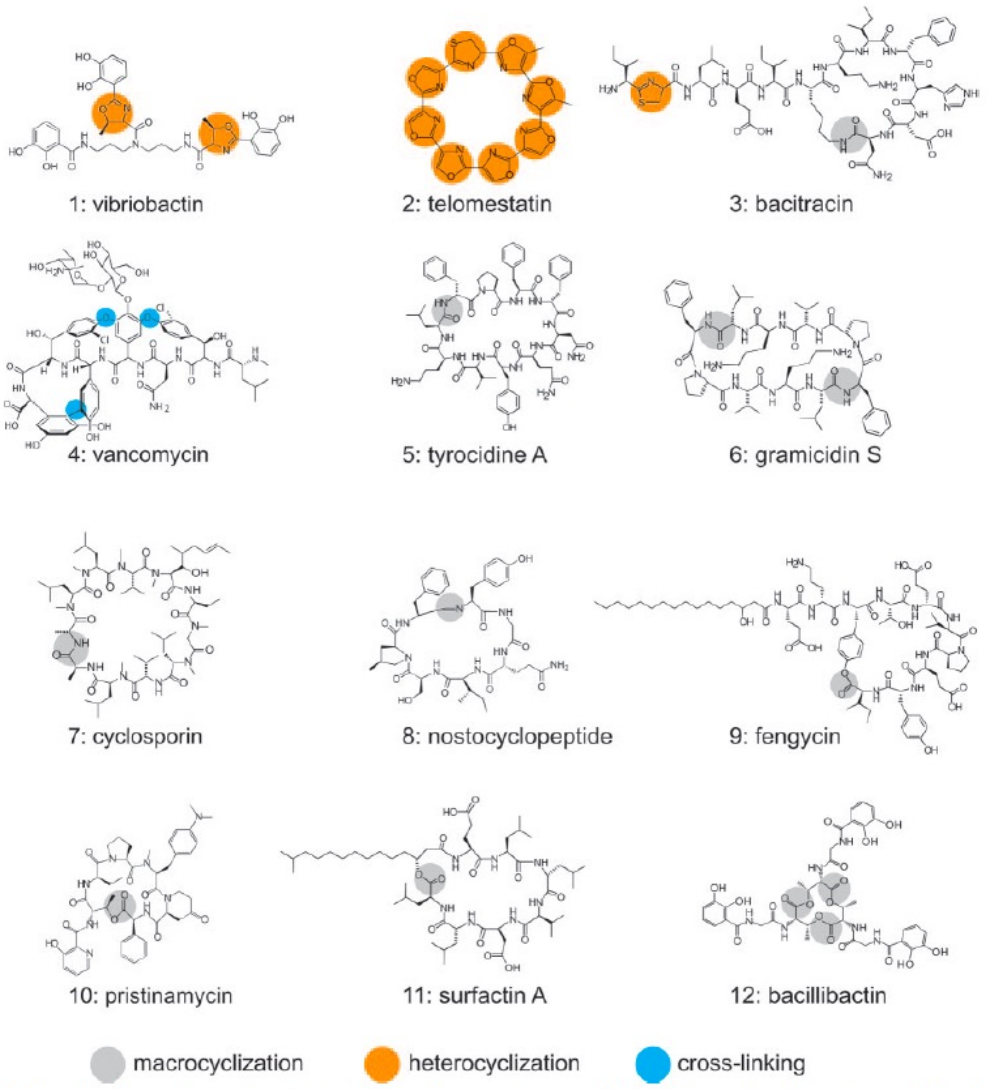


FIG. 1. A selection of non-ribosomally synthesized peptides. Characteristic structural features that confer rigidity to the peptide backbone are highlighted.

Epimerization

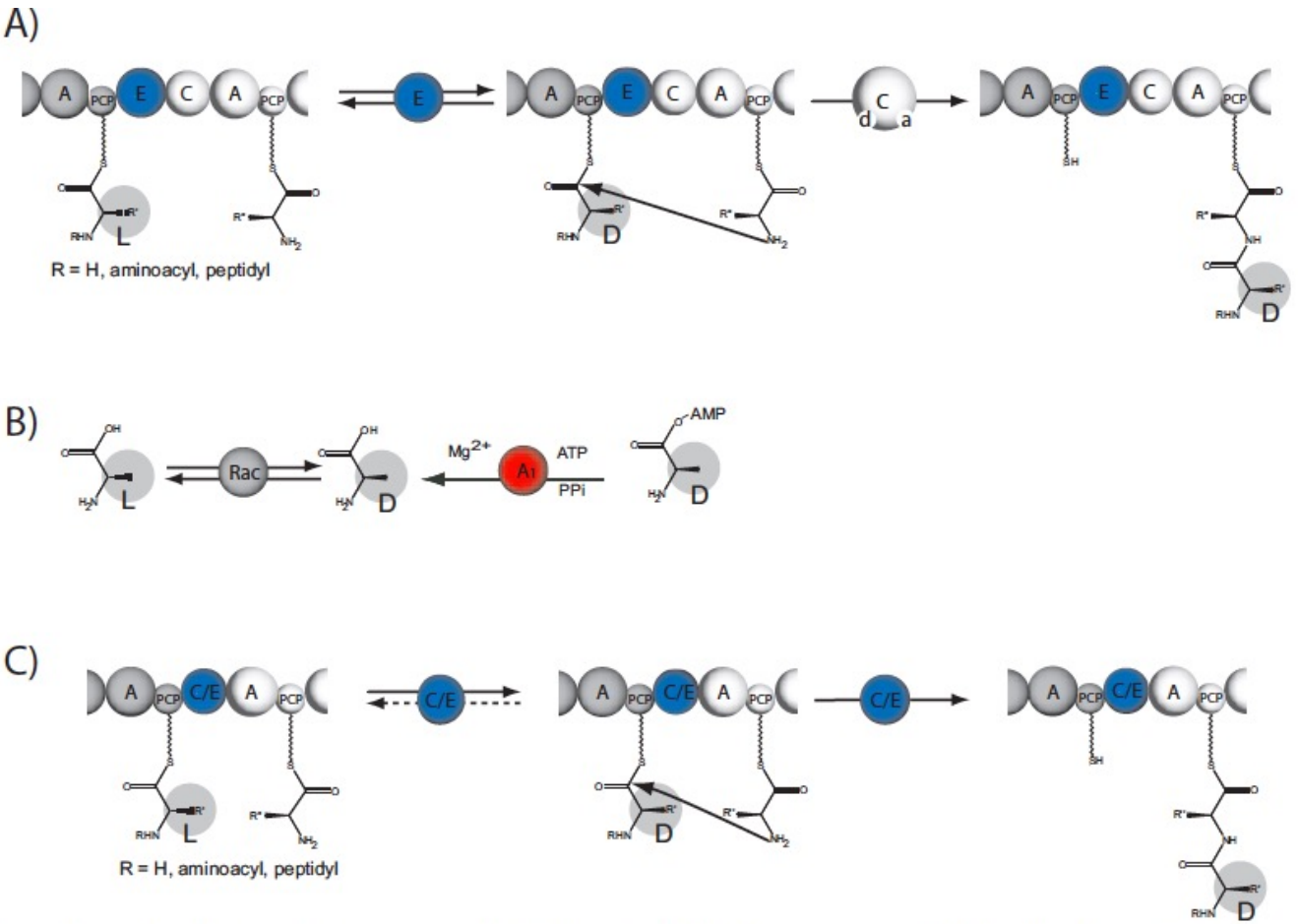
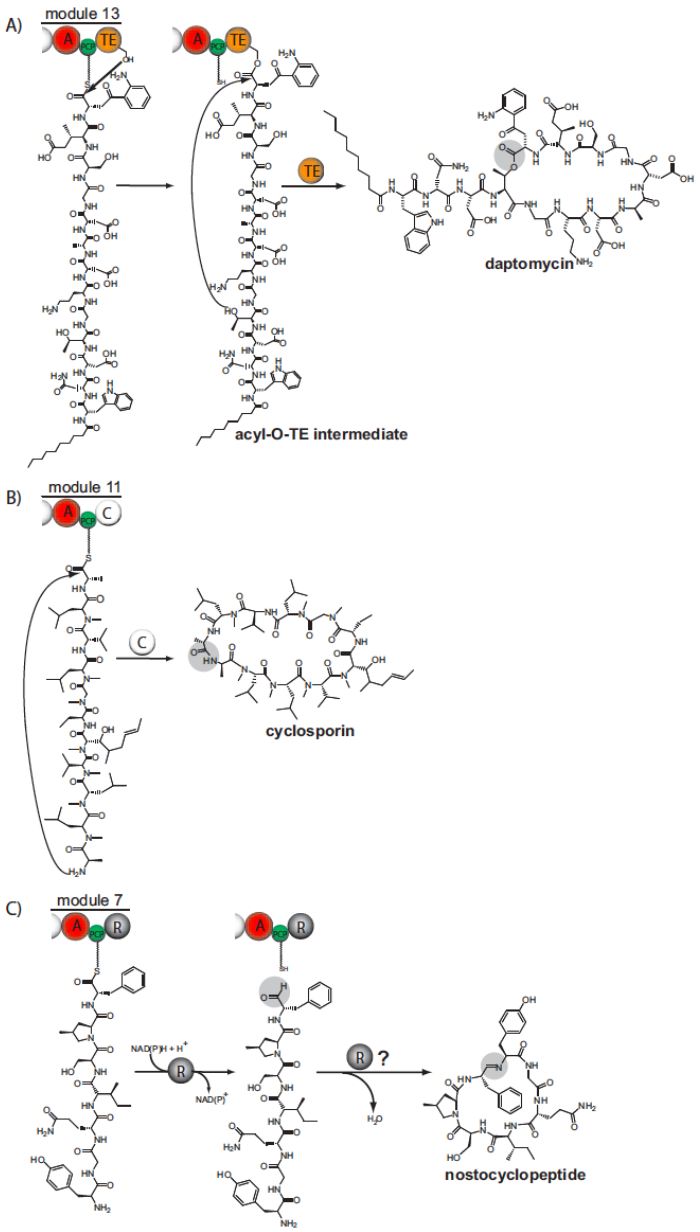


FIG. 6. Proposed mechanisms underlying amino acid epimerization. (A) The E-domain converts the PCP-tethered aminoacyl substrate into a D/L equilibrium. The stereoselective donor site (d) of the C-domain of the downstream module uses only the D-configured amino acid for subsequent peptide elongation. (B) In some cases, an external racemase (Rac) catalyzes the racemization of a freely diffusible amino acid. Here, a stereoselective A-domain is the determinant that activates solely the corresponding D-enantiomer. (C) D-Amino acid incorporation into arthrofactin, syringomycin, and syringopeptin is catalyzed by a new type of condensation domain (C/E-domain). Epimerization does not take place unless the PCP downstream of this C/E-domain is loaded with the dedicated amino acid. It is not yet known whether the epimerization reaction is reversible or not. After epimerization of the upstream aminoacyl/peptidyl thioester, the C/E-domain mediates the elongation of the peptidyl chain with $^{D}C_{L}$ chirality.

Release of the product

Product release is catalyzed by a thioesterase domain (TE) of about 250 amino acids, found at the end of the last module.
The catalytic mechanism is similar to serine proteases and the reaction generates an acyl-O-TE intermediate.
Release of the peptide is due to intramolecular nucleophilic attack that generates a cyclic product or by hydrolysis in the case of linear products.
Product release can also be mediated by a condensation domain (cyclosporin) or by a reductase domain (nostocyclopeptides).



Classification of NRPS