REVIEW ARTICLE

Linking Peptide and Polyketide Biosynthesis†

HORST KLEINKAUF and HANS VON DÖHREN

Institute of Biochemistry and Molecular Biology, Technical University Berlin, Franklinstr. 29, D-10587 Berlin, Federal Republic of Germany

(Received for publication April 24, 1995)

I. Levels of Organization
II. Biosynthetic Modules
III. Modules in Peptide Biosynthesis
IV. Modules in Polyketide Biosynthesis
V. The Unclear Fate of the Precursor
VI. Competing with the Natural Design?
VII. Integration
VIII. Spatial Structures
IX. Transport of Intermediates
X. Direction of Transport
XI. Mixed Biosynthetic Systems
XII. Conclusion

The biosynthesis of natural products involves many similar reactions. In contrast to the art of the classical organic chemist, natural chemistry within living organisms operates with protein catalysts. These catalysts assemble the reactants and perform their tasks in general in neutral aqueous media at ambient temperatures. To perform the synthesis of a complex metabolite, the proteins may have to interact in various ways. The most simple is the reversible interaction of a set of proteins. More often, and especially in the formation of complex peptides, a covalent assembly of enzymes has been found, poly- or multienzymes. This assembly seems at first sight to simplify complex processes. One protein may catalyse a complete set of reactions. It is not required to analyse each step of a multistep pathway in detail to finally be able to perform a multistep synthesis with many different enzymes. One protein may take up simple precursors, and deliver the final product, as all intermediates remain attached, just as in solid phase synthesis[1].

During the past years many biosynthetic systems have been analysed in some detail, especially at the gene level. This analysis has brought many unexpected results. At the moment, we even cannot fully appreciate the data, since most of it we cannot interpret. However, it tells us already many details of how the biosynthesis of natural products is organized[2].

I. Levels of Organization

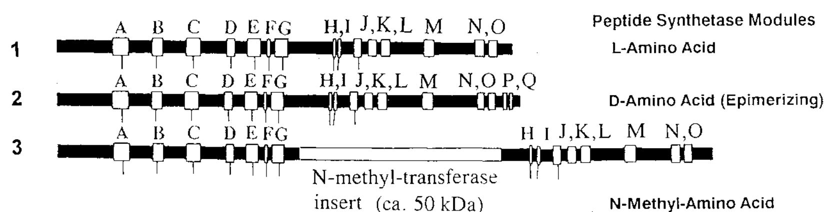
There are several levels of organization. From the view of the chemist, the prime interest in proteins is their function as biocatalysts. This is the view of the precise function of each part of the machinery, of each chemist performing his special task in the total synthesis. But proteins are much more than this. They have also been optimized to interact, and only those are supposed to interact, that serve the purpose of producing the one special product. Only the good team of chemists will be able to compete, and the standards in nature appear to be very high. Finally, the whole system has to be controlled within the compartment, and serve its special purpose. It has to be called upon to act, and this signal has to come from outside. For the view of biosynthesis, we may stop at this level, and not progress further to the level of the changing community. For our present state of knowledge, the discussion will be restricted to the first level, and at best, we can suggest how perhaps to approach the next level.

II. Biosynthetic Modules

The analysis of biosynthetic systems has revealed a modular organization of biocatalysts. Each gene of a biocatalyst is composed of structural elements, which by their sequence can often be identified to encode a certain

† This article is a special contribution in honour of Professor SATOSHI ŌMURA's 60th birthday.

Fig. 1. Modules of peptide synthetases.



By sequence alignment three types of modules can be identified: (1) nonepimerizing modules, generally adding an L-amino acid; occasionally, if it is the direct precursor, a residue in D-configuration may also be used; (2) epimerizing modules, which lead to a racemate of the aminoacyl- or peptidyl-intermediate, from which selectively the D-isomer is selected for the following reaction; (3) *N*-methylating modules, which catalyse the transfer of a methyl group from S-adenosyl-methioonine to the amino group of an activated amino acid. The letters indicate conserved motifs, which have been identified. Note, that the *N*-methyl-transferase segment is inserted between the G- and H-motif.

function. Biocatalysts, which assemble peptides or polyketides perform several condensation reactions in a defined order. This means that *e.g.* activated amino acids or activated acetates are connected step by step. The information encoding the catalysis of the condensations reaction is contained in highly similar gene segments, respectively. For simplicity, the term module has been introduced, to describe not only this condensation segment, but all segments required to complete a single condensation step[3]. A biosynthetic system can thus be recognized already at the gene level by the number of modules present[3,4]. A tripeptide synthetase will be identified by three adjacent amino acid adding modules. A triketide synthase gene will show three modules for condensing activated carboxylic acids. Note that peptide synthetases require ATP for carboxyl activation, while polyketide synthases use activated compounds as coenzyme A-thioesters.

III. Modules in Peptide Biosynthesis

Although we cannot yet identify all structural elements of biosynthetic modules, we still can draw much information from the gene sequences. In peptide forming systems, we can identify modules adding an amino acid, or adding an epimerized amino acid (Fig. 1)[5,6]. We cannot yet differentiate modules adding L- or D-forms of amino acids. But if the structure of the peptide under investigation is known, we can safely predict, if the L- or D-isomer is the precursor in that particular position. We can further identify modules adding an *N*-methylated amino acid by the pesence of an *N*-methyl-transferase segment[7]. This could tell what positions in a peptide are *N*-methylated. We can make very faint predictions on the specificity of a module, this means on the code of the enzymatic peptide forming system. In very special cases, when comparing *e.g.* modules of peptides formed by strains of Bacillus, we can sometimes make a good guess if a module adds proline, or leucine, or ornithine[8]. If we compare synthetases forming the same product from various sources, we have a good chance of predicting what the module does. So considering fungi and bacteria, both producing β-lactam antibiotics starting from the tripeptide precursor ACV, we can by sequence comparison identify each module, regardless of its origin[9].

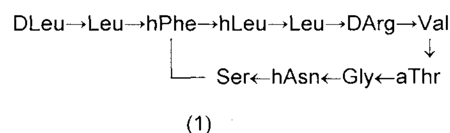
IV. Modules in Polyketide Biosynthesis

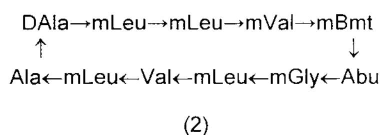
In the polyketide field, things seem to look more simple. If elements like acetate, propionate and butyrate are added *via* their carboxylated coenzyme A-thioesters, this can be identified without problem. For other precursors only insufficient information is available, but these cases are rare. When condensation has occurred, the resulting keto-compounds may undergo reduction, dehydration, and the second reduction. The structure of each respective module will indicate if none, one, two, or all three transformations occur[10]. Thus the set of modules identified permits the prediction of the structure of the polyketide precursor.

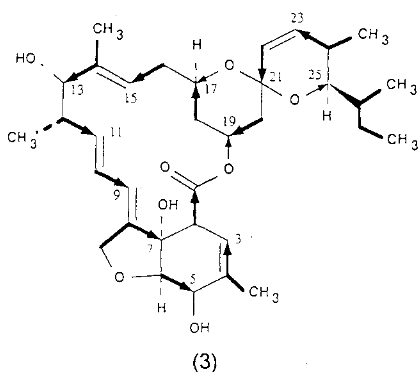
V. The Unclear Fate of the Precursor

But here our understanding of the data stops. We arrive at a linear peptide precursor and at a linear polyketide precursor, attached to a protein, but the protein structure may leave us without any clue of the fate of this precursor. We cannot even decide yet if it is released by action of a thioesterase or will release itself by a cylization reaction, since both types of reaction seem to involve similar modules, catalysing transfer reactions. As a rule however, we do know the structure of the compound under investigation, and thus may gain considerable knowledge of its step by step growth and of the organization of the biocatalysts.

Fig. 2. Structures of lysobactin (1), cyclosporin (2) and avermectin (3).







Each amino acid in the peptides is contributed by a peptide biosynthetic module. While all eleven modules are integrated into a single structure in the case of cyclosporin, two multienzymes contributing three and eight residues, respectively, have been detected in the lysobactin system. In the avermectin system twelve successive condensations are performed, starting at position 25. The respective modules are integrated in at least four multienzyme structures.

VI. Competing with the Natural Design?

The module story, which is the very first level of the path to a natural product, already seems to open up roads to new products. As we can engineer genes, we should be able to engineer modules, and thus create biosynthetic systems. We could have these systems operate in organisms of our choice, or even more appealing to the chemist, take the biocatalysts out of the cell, and have them perform *in vitro* with even more opportunities of manipulation. But do we have sufficient information at the very first level for such tasks? The answer could be yes, which means that in case of extreme luck a new biocatalyst could be made and work as intended. However, its performance would for sure be disappointing, when compared to the operating systems having evolved in a yet unconceivable history.

VII. Integration

A closer look to the modular organization of genes already points to many unsolved issues. Modules may be integrated to different extents. Generally, eukaryotic systems tend to higher degrees of integration. Prokaryotic systems, however, have always been catching up. The most complex assembly of modules known so far is the cyclosporin synthetase gene comprising eleven amino acid adding modules (Fig. 2), with a size of 45.7 kilobases[11]. This fungal gene, which has no introns, encodes a protein of 1.7 Megadaltons in size. The largest prokaryotic protein known so far is a peptide synthetase from the gliding bacterium *Lysobacter* involved in the synthesis of the antibiotic lysobactin or katanosin. The synthetase adds eight amino acids to a tripeptide precursor and has an estimated size of 1 Megadalton[12,13]. The largest polyketide synthase known so far is found within a set of at least four synthases forming the powerful antihelminthic avermectin, and combines three modules[4]. The size can be estimated to about 750 kilodaltons.

Here already ends our understanding of the modular organization. Why are eleven modules fused in the fungus *Beauveria nivea*, and organized into two synthetase genes of three and eight modules respectively in *Lysobacter*? And why are the twelve modules required for the basic avermectin structure fused into entities containing at most three modules? There is yet no answer to this question. We do also not recognize from the gene level where gene products will interact. Analysis of protein-protein interactions has to come from experimental work, which is still missing in this field.

VIII. Spatial Structures

As we enter the second level of protein structure, we again have to admit, that our understanding is fairly small. So far, no three dimensional structure of a synthase or synthetase is known at the molecular level. We do have some electron micrographs of fatty acid synthases and gramicidin S synthetase, which show somewhat typical globular proteins. But no clue emerges concerning their functions, except, that the protein domains originating from the assembly of modules are closely interacting. However, we are still completely ignorant about this interaction, and cannot explain, how the intermediates are transferred to their respective reaction sites.

IX. Transport of Intermediates

The key element in this transport is a structural element contained in all modules, regardless if peptide or polyketide forming systems. It is called either peptidyl or acyl carrier protein, and this small protein of about 10 kilodaltons can be free or integrated within an adding element. The best studied acyl carrier protein in bacterial

and plant systems fulfills a variety of functions transporting acyl groups to sites of transforming enzymes. A rough estimate suggest that this ACP has to interact with at least ten other proteins, and the processes of recognition are yet unknown. In peptide synthetases in general, the carrier protein is integrated in each amino acid adding unit. Still, as careful sequence comparisons show, its structural properties are retained, and the prediction of its secondary structure within the multienzyme context is essentially unchanged (R. DIECKMANN and H. VON DÖHREN, unpublished). As the carrier protein domain is now linked with two adjacent domains, the interactions with other proteins are restricted. We observe less charged amino acid residues, and can conclude that surface charges play a major role in protein-protein interactions.

X. Direction of Transport

The carrier protein with its cofactor 4′-phosphopantetheine performs a directed transport of the respective intermediates. Although the domain is positionally linked between adjacent domains, and thus the transport function could be thought to be spatially restricted to these adjacent sites, still a direction of transport has to be provided. The cofactor is thus not a swinging arm, as envisioned in the earlier descriptions of fatty acid and peptide forming multienzymes. Its action is rather limited and directed by specific binding sites within the adjacent protein domains. The carrier protein is the obvious link between the biosynthetic systems performing condensation reactions. Integrated synthases and synthetases may carry several carrier protein domains, and these correspond to at least the number of condensation reactions catalysed, but may even be larger, if modifications of intermediates require transport to auxiliary enzymes or enzyme sites.

XI. Mixed Biosynthetic Systems

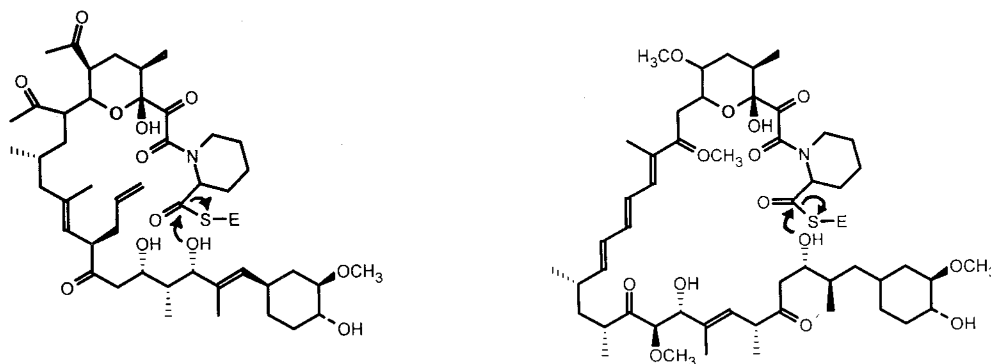
Many compounds are known to originate from both amino acid and carboxylic acid precursors. The adenylate module is used in the activation of many types of simple carboxyls, including also imino acids, hydroxy acids, aromatic acids to then enter the condensation step, each adenylate has to reach the thioester stage at a cysteamine attached to coenzyme A or a carrier protein domain containing 4′-phosphopantetheine.

Aminoacylation reactions have so far not been analysed in detail at the gene level, with the exception of enterobactin biosynthesis. Although not completely unravelled, the transfer and condensation of dihydroxybenzoate as adenylated precursor to the thioester-bound seryl-residue between EntE and EntF requires pantetheine[14]. This evidence has also been obtained for the similar transfer of 4-methyl-hydroxy-anthranilic acid in the initiation of actinomycin biosynthesis[15].

In some cases modification of amino acids suggests the interplay with an acetate introducing module. Amino acids would be extended by one or two carbons. The modified amino acid would then function as precursors in the respective peptide biosynthetic system.

Some data has been obtained in the biosynthesis of macrolactones containing aminoacyl- or peptidylresidues. Presumably here, a peptide synthetase is interacting directly with a polyketide synthase in transferring the polyketide intermediate to the thioester-bound imino acid. The termination reaction would then be a lactone formation of the poylketide-iminoacyl-intermediate (Fig.3).

Fig. 3. Postulated termination reaction in the biosynthesis of the peptidylmacrolactones FK-506 (left) and rapamycin.



The polyketide-precursor is transferred to an activated imino acid on a peptide synthetase (E), which then directs cyclization by lactone formation. This final intermediate transfer requires the interaction of a polyketide synthase with a peptide synthetase.

Table 1. Peptide synthetase structural elements and their occurence in related enzymes and enzyme systems.

Element	Enzymes	Functions
Adenylate formation	Acyl-CoA ligases	Providing acyl-CoA for a variety of biosynthetic reactions
	Insect luciferases	Luminescence
	Adenylate forming, like	Siderophore
	EntE	Biosynthesis
4'-Phosphopante-theine binding site (carrier protein)	Peptide synthetases Polyketide synthases PHB-synthase	Covalent transport of acyl-intermediates
Epimerase	Peptide synthetases Polyketide synthases?	Epimerization of acyl-intermediates
Thioesterase	Thioesterases Polyketide synthases	Providing free carboxylates for a variety of bio-synthetic routes
	Peptide synthetases	Possible transferase function (cyclization)
N-Methyltransferase	Methyl-transferases	Various methyl transfers

XII. Conclusion

Analysis of biosynthetic reactions at the gene level has provided evidence for the similarity of modules encoding related steps. Such modules are assembled from substructures coding for catalytic and transport steps. Substructures and modules can be followed to evolutionarily originate from several common ancestors, and the generation of biosynthetic pathways is accomplished by module construction, module assembly, and clustering of genes including modification and export functions, as well as regulatory signals. Many elements detected are found in polyketide and polypeptide forming systems (Table 1). The central element linking both systems is the acyl carrier protein, directing the transport of intermediates.

References

1) KLEINKAUF, H. & H. VON DÖHREN: Nonribosomal biosynthesis of peptide antibiotics. Eur. J. Biochem. 19: 1~15, 1990
2) VINING, L. C. & C. STUTTARD (*Eds.*): Genetics and Biochemistry of Antibiotic Production. Butterworth-Heinemann, Boston 1995
3) DONADIO, S.; M. J. STAVER, J. B. MCALPINE, S. J. SWANSON & L. KATZ: Modular organization of the genes required for complex polyketide biosynthesis. Science 252: 675~679, 1991
4) MACNEIL, D. J.; J. L. OCCI, K. M. GEWAIN & T. MACNEIL: Correlation of the avermectin polyketide synthase genes to the avermectin structure. Ann. N. Y. Acad. Sci. 721: 123~132, 1994
5) PFEIFER, E.: Struktur-Funktionsanalyse der Tyrocidin Synthetase 1. Ph. D. Thesis, TU Berlin 1994
6) FUMA, S.; Y. FUJISHIMA, N. CORBELL, C. D'SOUZA, M. M. NAKANO, P. ZUBER & K. YAMANE: Nucleotide sequence of 5' portion of *srfA* that contains the region required for competence establishment in *Bacillus subtilis*. Nucleic Acids Res. 21: 93~97, 1993
7) HAESE, A.; M. SCHUBERT, M. HERRMANN & R. ZOCHER: Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. Mol. Microbiol. 7: 905~914, 1993
8) COSMINA, P.; F. RODRIGUEZ, F. DE FERRA, G. GRANDI, M. PEREGO, G. VENEMA & D. VAN SINDEREN: Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. Mol. Microbiol. 8: 821~831, 1993
9) VON DÖHREN, H.; E. PFEIFER, H. VAN LIEMPT, Y.-O. LEE, M. PAVELA-VRANCIC & T. SCHWECKE: The nonribosomal system: What we learn from the genes encoding protein templates, In *Industrial Microorganisms: Basic and Applied Molecular Genetics* (R. BALTZ, G. D. HEGEMANN & P. SKATRUD, *Eds.*), S. 159~167, Am. Soc. Microbiol., Washington D. C., 1993
10) ROBINSON, J. A.: Enzymes of secondary metabolism in microorganisms. Chem. Soc. Rev. 17: 383~452, 1988
11) WEBER, G.; K. SCHÖRGENDORFER, E. SCHNEIDER-SCHERZER & E. LEITNER: The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8 kiolbase open reading frame. Currr. Genet. 26: 120~125, 1994
12) KOPIEZ, H.: Untersuchungen zur Biosynthese von Lysobactin in Lysobacter sp SC 14,067. Ph. D. Thesis, TU Berlin 1994
13) LÜCK-KLATTE, B.: Untersuchungen zur Biosynthese von Lysobactin in Lysobacter sp SC 14,067 und Charakterisierung von beteiligten Multienzymen. Ph. D. Thesis, TU Berlin 1994
14) RUSNAK, F.; M. SAKAITANI, D. DRUECKHAMMER, J. REICHERT & C. T. WALSH: Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the *EntF* gene, expression and purification of EntF, and analysis of covalent phophopantetheine. Biochem. 30: 2916~2927, 1991
15) KELLER, U.: Peptidolactones in *Genetics and Biochemistry of Antibiotic Production* (L. C. VINING & C. STUTTARD, *Eds.*) pp. 173~196, Butterworth-Heinemann, Boston 1995