Molecular Diversity (2005) 9: 171–186

Perspective

What can a chemist learn from nature's macrocycles? – A brief, conceptual view

Ludger A. Wessjohann*, Eelco Ruijter, Daniel Garcia-Rivera & Wolfgang Brandt
Leibniz-Institute of Plant Biochemistry, Department of Bioorganic Chemistry, Weinberg 3, D-06120 Halle (Saale), Germany
*(*Author for correspondence, E-mail: wessjohann@ipb-halle.de, Tel: +49 (345) 5582 1300, Fax: +49 (345) 5582 1309)*

Received 23 February 2004; Accepted 14 May 2004

Key words: cheminformatics, cyclases, cyclopeptides, macrocycles, natural products, polyketides, terpenoids

Summary

Macrocyclic natural products often display remarkable biological activities, and many of these compounds (or their derivatives) are used as drugs. The chemical diversity of these compounds is immense and may provide inspiration for innovative drug design. Therefore, a database of naturally occurring macrocycles was analyzed for ring size, molecular weight distribution, and the frequency of some common substructural motifs. The underlying principles of the chemical diversity are reviewed in terms of biosynthetic origin and nature's strategies for diversity and complexity generation in relation to the structural diversity and similarities found in the macrocycle database. Finally, it is suggested that synthetic chemists should use not only nature's molecules, but also nature's strategies as a source of inspiration. To illustrate this, the biosynthesis of macrocycles by non-ribosomal peptide synthetases and terpene and polyketide cyclases, as well as recent advances of these strategies in an integrated synthesis/biotechnology approach are briefly reviewed.

Abbreviations: ACE, angiotensin converting enzyme; ACP, acyl carrier protein; AT, acyl transferase; C, condensation; 6-dEB, 6-deoxyerythronolide B; DEBS, deoxyerythronolide B synthase; E, epimerization (domain); GGPP, geranylgeranyl diphosphate; KR, ketoreductase (domain); KS, ketosynthase (domain); M, *N*-methylation (domain); Nic, nicotinoyl; NRPS, non-ribosomal peptide synthetase; PCP, peptidyl carrier protein; PKS, polyketide synthase; QSAR, quantitative structure-activity relationships; TE, thioesterase

Introduction

Natural products, predominantly secondary metabolites, play a key role in modern drug development. Their disproportionally high success rate was confirmed by recent reports indicating that natural products, together with their derivatives, mimics, and other compounds derived or inspired by them, comprise approximately 35–40% of all current trade drugs [1–6]. The proportion of these compounds in screening efforts, however, constitutes less than 1%. In some areas, such as antibiotic and anticancer agent research, up to 3/4 of all drugs are natural product-derived. One obvious reason for this success is the million years of adaptation and natural selection of these products within the biological realm, i.e. our current natural products inherited an evolutionary advantage.

Moreover, the premise that natural products are generally polar with an inherited low bioavailability profile may be unfounded. Lee and Schneider estimated that known natural products follow Lipinski's "rule of five" [7] for bioavailability equally well as current trade drugs do [8]. This is hardly surprising considering that these compounds are required to exert a certain biological effect in nature (which will, however, not necessarily be the same as that on the pharmacological target) and must be able to reach a specific site of action.

Accordingly, natural products have been suggested to be excellent starting points for library construction in drug development [9, 10]. Many approaches led to the discovery of active compounds [9].

Since natural products are not produced by any organism to solve human health problems, their extraordinary success in drug development must have an additional component beyond the evolutionary advantage and bioavailability. They must exhibit an increased biological activity probability against certain targets, which do not necessarily coincide with the "native" target in their biotope. A good question to illustrate this point is why a yew tree bark compound like Taxol®, which originally maybe directed against bark fungi or parasites, proved useful for human purposes as an anticancer agent?

The simplest and obvious explanation is that certain compounds or compound groups bind to specific and highly conserved or otherwise structurally related protein domains

[11, 12] regardless of the organism expressing such proteins. Proteins are built up in a modular fashion and have evolved through variation and recombination of modules (domains), which fold in a certain way more or less independent of the rest of the protein. Whereas the total number of proteins in humans is thought to be in the 100,000–450,000 range, the number of folding types is presumably only 600–8,000 and the number of sequence families in these protein domains is estimated to be 4,000–60,000. Therefore, proteins with similar folding types (though not necessarily functionally related) may be inhibited by the same class or rather type of small molecule compounds. This was demonstrated by the selective inhibition of two structurally, but not functionally, related proteins with derivatives of a natural product [13, 14]. This theory also explains why highly active natural products often have a native biological target different from their pharmaceutical (human use) target. The theory also validates the use of natural products as starting points (leads) for diversity-oriented synthesis, even when the native target of the natural product is not the desired one. In most cases, the native target of many successful natural products is not even known, or was discovered years after they were used by humans for other purposes.

In order to optimize the compound to fit non-native, but pharmacologically useful targets, often minor variations are sufficient to improve the lead compound. Thus, some derivatives of natural products were found to have a very different biological target than the natural compound itself [13–15]. Slight structural changes, artificial or natural, to some basic pharmacophores cause very different receptor activities (e.g. histamine derivatives, steroid derivatives, morphine versus codeine, etc.). The high probability of finding natural products with desired biological activities, however, must not be confused with the readiness of such compounds for commercial use. It must be clearly stated that, in contrast to the beliefs of the general public, natural products or their derivatives are by no means "better" drugs than artificially designed or synthetic ones. The latter often have the advantage of providing better availability and easier access to libraries for QSAR-studies. Thus, natural products make good lead compounds, but not necessarily good medicines. For an eager synthetic chemist, this raises the question "what information will I need to be able to design lead compounds with a success rate similar to the one that natural products provide?"

The evolutionary and (especially in the context of protein domains) co-evolutionary advantage of natural products answers part of the question about the reasons why their odds of showing activity against a given target are superior to random, artificially designed compounds. A recent survey shows that natural products generally have fewer nitrogen atoms and fewer five- and six-membered rings than synthetic drugs [8]. This finding is contrary to the general belief that successful drugs should ideally have small, nitrogen-rich heterocycle-based pharmacophores. We think that this classical pharmacophore model is biased by the easier synthetic availability of such five- and six-membered ring structures in the past. On the other hand, natural products often possess complex structures. So far, the cheminformatic analysis of natural products focused on a comparison with established drugs [4–6], on general diversity parameters, and on Lipinski's rule of five [7] and other medicinal chemistry properties, most of which are macroscopic like log P. The complex structural parameters of these compounds are rarely analyzed with cheminformatic methods, and are almost never compared with the underlying metabolic processes and natural building blocks. Also, rather than looking for similarities to current synthetic drugs, we consider it worthwhile to also concentrate on properties of successful natural lead compounds not mirrored in synthetic drugs, and on nature's principles to produce them. This should pave the way for faster, successful lead design off the historical beaten track. If we are able to extract enough information from natural products, ultimately, the physical or even virtual presence of natural compounds in drug discovery may be replaced by the underlying first principles of success and lead to synthetic lead compounds designed directly from these first principles. In order to garner the necessary information, we initiated an in-depth cheminformatic study into the structural and metabolic patterns of natural products. Our hope is to support the design of improved synthetic (really natural product-like) lead compounds, independent of natural scaffolds, but using principles or tendencies taught by the natural products' successes. Here we present some preliminary results.

Natural products commonly contain medium to large ring systems [8, 16–18]. These are more frequently encountered in natural than in artificial drugs. Indeed, macrocycles constitute a significant part of natural product(-derived) drugs and this article will focus on them. We decided to have a closer look into the principles of macrocyclic natural products, structurally and conceptually, based on some basic cheminformatic and metabolic observations. Naturally occurring macrocycles often have very complex structures and remarkably high activities (Figure 1). The variety in ring sizes and chemical constitution is overwhelming. The relatively high occurrence of macrocycles in nature may be rationalized by the fact that they constitute an equilibrium or compromise between conformational preorganization (i.e. less entropy loss upon docking) and flexibility to achieve optimal binding properties to their biological target, which also explains their often unrivaled activities. The variety in structure and biological activity of naturally occurring macrocycles is demonstrated by comparing FK-506 (**1**, immunosuppressant), erythromycin A (**2**) and vancomycin (**3**, antibiotic), epothilone B (**4**, anticancer), amphotericin B (**5**, antifungal), and K-13 (**6**, ACE inhibitor) (Figure 1).

Cheminformatic analysis of natural macrocycles – the structural concept

In order to gain insight into the structural variety and distribution of common substructures within naturally occurring macrocycles, we analyzed as many natural product entries as

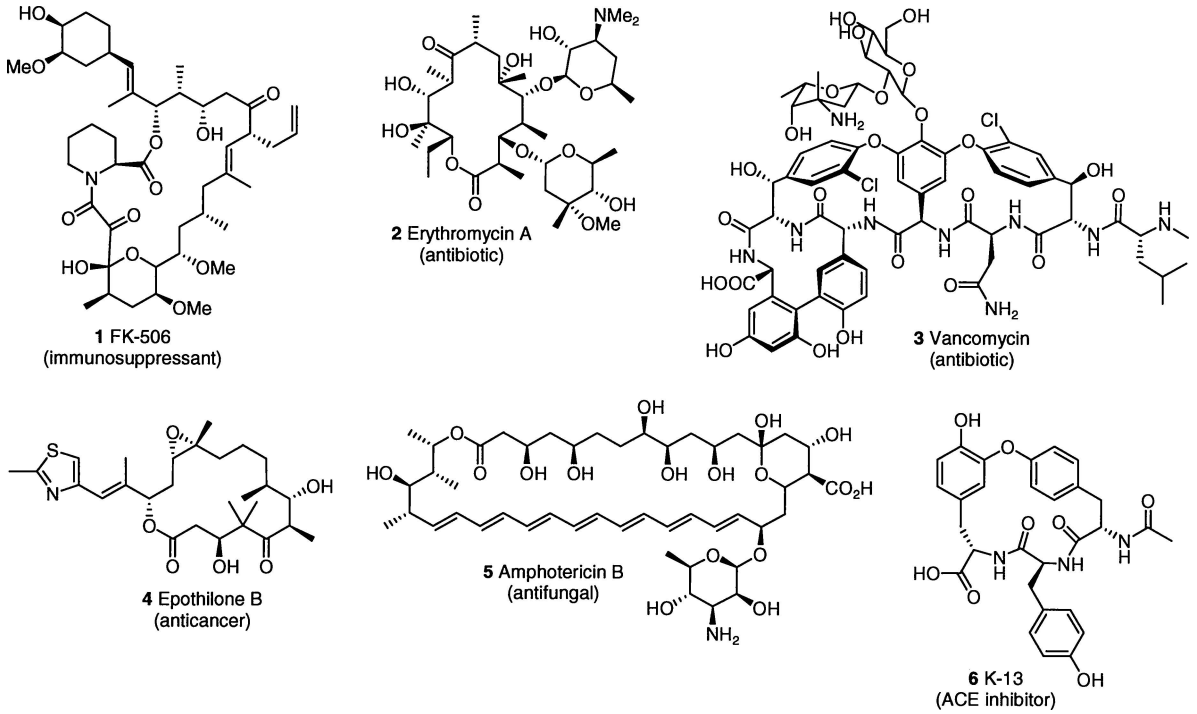

Figure 1. Some naturally occurring macrocycles displaying a variety of medicinally important biological activities.

possible, listed in reagent catalogs, the Dictionary of Natural Products [19], Beilstein database [20], and Phytobase® [21], using SciDex [22] and self-programmed subroutines for the cheminformatic analysis. Of these sources, Beilstein is the most complete one, suitable for this analysis. Altogether, we analyzed 132,522 compounds designated as natural products.

Some remarks need to be made with respect to the dataset, since most readers will not be aware of the problems and limitations inherent to such data mining and cheminformatic analyses. The effort may best be compared to the task of finding relevant literature for a given reaction. Here it is well accepted, that often only a percentage of the relevant literature will be discovered by sending a query to a database, and sometimes whole groups of relevant authors are not discovered, and only will be found by further refinement and secondary checks of references cited. Similar discrepancies apply here, only the data numbers in this case are several orders of magnitude larger.

The assignment of a compound as a natural product was taken as listed in the database sources. In view of the numbers involved, this could not be verified individually. Next, we tried to eliminate duplicates by comparing CAS-numbers of the various sources, or simply by hand. In conclusion, this semi-automatic approach did avoid almost all double entries, but did not allow us to verify that each macrocycle was a natural product, nor does it fully preclude the erroneous inclusion of some non-macrocyclic compounds and residual duplicate entries. Also, many members of unusually well-studied classes of macrocycles (e.g. substance classes studied in depth because of outstanding medicinal importance, like erythromycin- and vancomycin-type compounds [23]) do not occur in these databases in all natural variations known. The same applies to compounds isolated from genetically modified organisms, or from wild-type organisms cultivated under abnormal conditions, for example by the One Strain MAny Compounds (OSMAC)-procedure or from mutasynthesis [24, 25]. Thus, although more than 140 glycopeptide antibiotics such as vancomycin are known, only a few are discovered by automatic data mining and are included in the analysis. Other structures had to be eliminated because they lacked chemical integrity (meaning they were missing bonds, had carbons with five bonds, had inconsistencies between the structural and molecular formulas, etc.). The individual biological activity information had to be excluded from our analysis altogether, because such data are often not reported, or cannot be evaluated automatically with our current tools, or because only "soft" information is given (for example "... use was reported against abdominal pain ..." without quantifiable data or citation).

What is the consequence for our dataset? Completeness as an ultimate goal currently is not possible. Apart from the fact that a complete, verified database is not yet available, such a database would be distorted because many organisms and compound classes are much more thoroughly studied than others. Microorganisms allow rapid variation of metabolite production [24, 25] and strains can produce many derivatives, which are more easily isolated and characterized than compounds produced by more complex or challenging systems, such as tree bark or non-cultivable marine organisms. In other cases, selective in-depth isolation

efforts (e.g. of glycopeptides) sparked by an important application, distort the results further and will lead to a disproportional abundance of these pre-selected structures. Even metabolomics as a more universal approach is not able to fill this data gap within foreseeable time since it is currently limited to very few organisms and, despite its hypothetical claim of completeness, is limited to compounds amenable to current rapid separation and identification methods [26].

Thus, the numbers given here should not be taken as absolutes. They are distorted by selective or erroneous data production, as well as, incomplete, selective, or even incorrect database entries. However, we believe the error rate from data mining and selection is negligible for most basic analyses, based on random spot checks, especially for statistical considerations of a relative character. This set of compounds is one of the largest to be used for such a purpose so far and we think it is a good representation overall of structural variability in naturally occurring macrocycles. Also, some compound properties had limitations, which forced us to refrain from statements or correlations we were eager to draw. For example, a correlation of the presented macrocycle data with the mostly very "soft" or unavailable biological activity data would generate results with a relevance factor too low to be acceptable. Thus, the reader will have to wait for further refinement of our analyses for correlations to biological activity.

For this analysis, macrocycles were defined as compounds containing at least one non-bridged ring with $\geq$13 covalently connected atoms. This arrangement was found in 3,747 compounds, that is ca. 3% of all natural products are macrocycles.

This subset of naturally occurring macrocycles was analyzed for ring size and molecular weight distribution. The largest ring size considered consisted of 72 atoms and was found in eight compounds, ether lipids of archeae bacteria (e.g. 2,3-di-*O*-biphytanyl-*sn*-glycerols). The compound with the highest molecular weight (2559 g mol^{-1}) is merremin. Of course, even larger cyclic compounds are known from nature (e.g. cyclic DNA in plasmids), but these are usually not part of secondary metabolism and behave like linear compounds rather than macrocycles, except for the missing terminals. From the 3,747 compounds, only 36 (<1%) macrocycles have ring sizes larger than 40 atoms. These compounds were removed from our subsequent, more detailed analyses, since they represent a negligible proportion of the total set, and their omission resulted in an improved graphical presentation of the more relevant small and medium macrocycles. The total number of compounds per ring size is represented by the connected dots in Figure 2. Please note that this number is not absolute since it is limited by the dataset available for automated analysis as discussed above, and real numbers will be higher.

The general trend is a decrease in the number of compounds with increasing ring size, which is logical with respect to entropy and cyclization statistics. Also, there is a ring size-dependent oscillation in the abundance. Smaller macrocycles oscillate with every second atom added to the ring, with even-numbered macrocycles being much more frequent than odd ones, whereas for medium range macrocycles the oscillation period is stretched to have a peak after roughly three more atoms. This is a preliminary indication that smaller macrocycles may be formed by building blocks of two (or four) connecting atoms, such as acetate, malonate, or propionate in polyketides and lipids (or isoprenoids for four ring atom distance), whereas in larger rings, three-atom building blocks such as amino acids for peptides dominate. The latter also show a propensity for higher molecular weights, something also expected for peptides with their extensive side chains (compared to polyketides or terpenoids).

Fourteen-membered macrocycles are by far the most common with 806 compounds represented. The molecular weight distribution of the 13–40-membered compounds (all 3,711) with respect to ring size is represented as bars in Figure 2. Only two 13-membered ring compounds had a molecular weight <200 g mol^{-1}. These two were omitted for better clarity in the graphical representation. By far, the highest frequency of compounds (414) was found for 14-membered rings with a molecular weight between 300 and 400 g mol^{-1}. Other relatively frequent ring sizes are 16, 18, 19, 21, 24, and 25. For 25-membered macrocycles, the high frequency of high molecular weight compounds (>1000 g mol^{-1}) is remarkable. Please note that in this analysis, the high numbers of selectively isolated derivatives of erythromycin (14-membered) and glycopeptide (16-membered) antibiotics are not fully represented for the reasons given previously.

Further analysis highlights the distribution of macrocycles relative to elements and functional groups, or their omission (Figure 3). For each size, the proportion of rings that consist of carbon atoms only, or contain at least one ring nitrogen or oxygen atom, and at least one lactam or lactone functionality are compiled. Quite unexpectedly, 14-membered rings that contain only carbon atoms, but no heteroatoms (463 compounds) are most abundant. The majority of these compounds (320) have a total number of 20 carbon atoms, and it is therefore plausible that the majority of these compounds are diterpenoids, i.e. cyclized geranylgeranyl diphosphate, with three exocyclic methyl and one isopropyl(idene) or dimethylcyclopropane group [27]. An additional 72 C_{14}-macrocyles possess 22 carbon atoms. They may be the corresponding acetates. This interpretation is supported by the extraordinarily high number of isolated double bonds in 13- and 14-membered rings (see Figure 4). Thirteen-, 15-, and 16-membered derivatives may either be 1,2- or allyl-shifted diterpenoids, or internally cyclized ones, or have a different (polyketide) metabolic origin. Accordingly, another peak of all-carbon macrocycles would be expected for 22-membered terpenoids, which might be attributed to macrocyclic triterpenes or cyclized squalene, but such compounds are not of statistical importance, and C_7-bridged *para*-cyclophanes dominate this group. A more detailed discussion of terpene macrocyclization is given below.

It is evident from Figure 3 that the majority of nitrogen-containing macrocycles are also lactams. Some 25- and 30-membered nitrogen-containing rings are not lactams. These

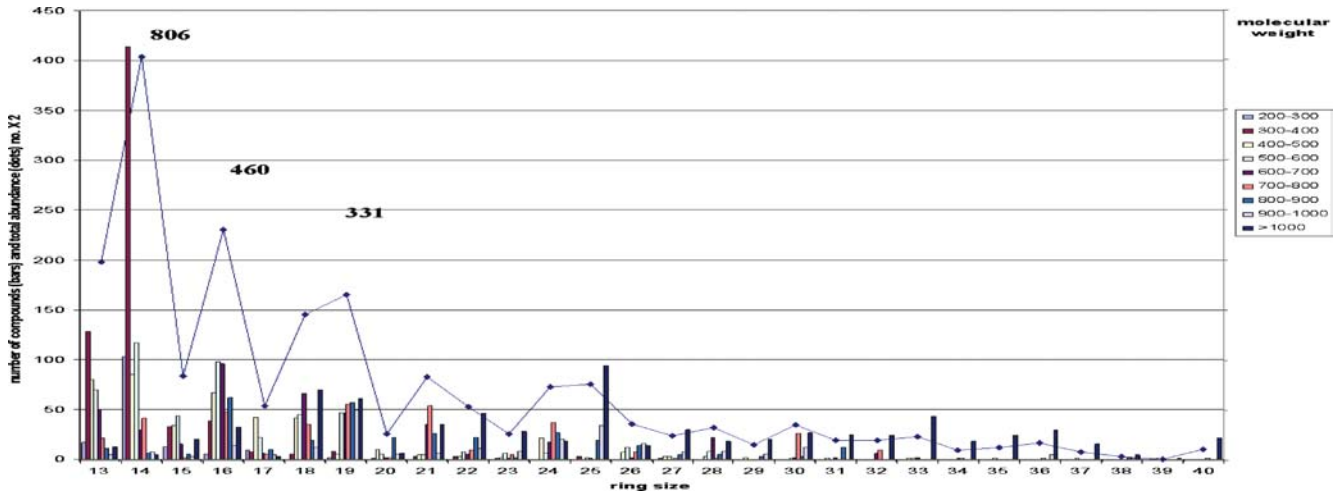

Figure 2. Distribution (number of compounds) of 100 D molecular weight ranges (bars) and total number of natural macrocycles (line connected dots) of a given ring size. The scale for the total count (abundance) is twice the scale for the weight distribution given at the ordinate.

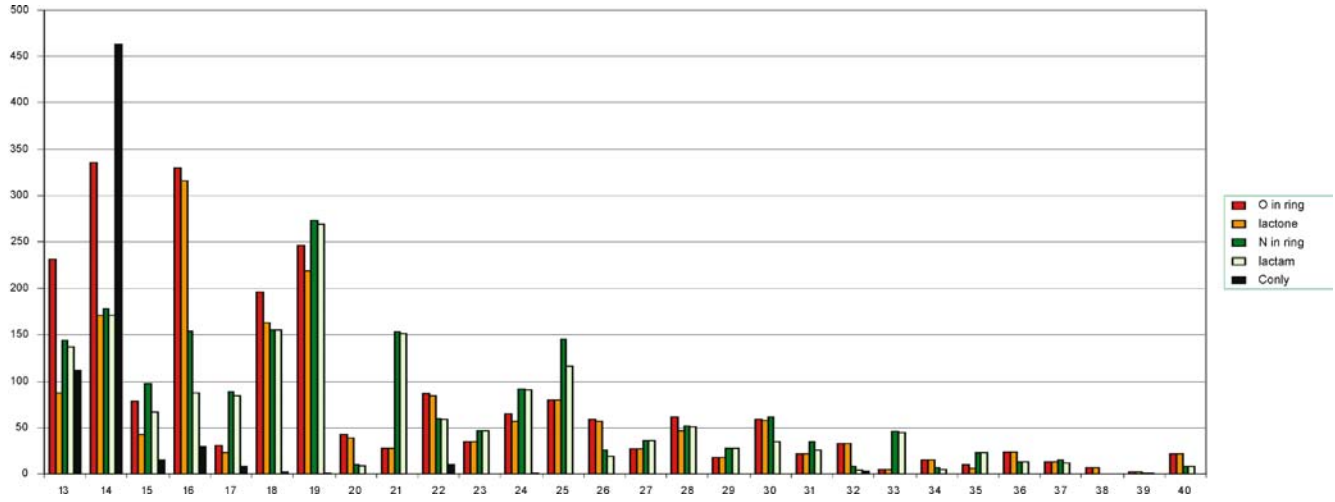

Figure 3. Distribution of only carbon (black) or at least one structural heteroatom moiety within the rings.

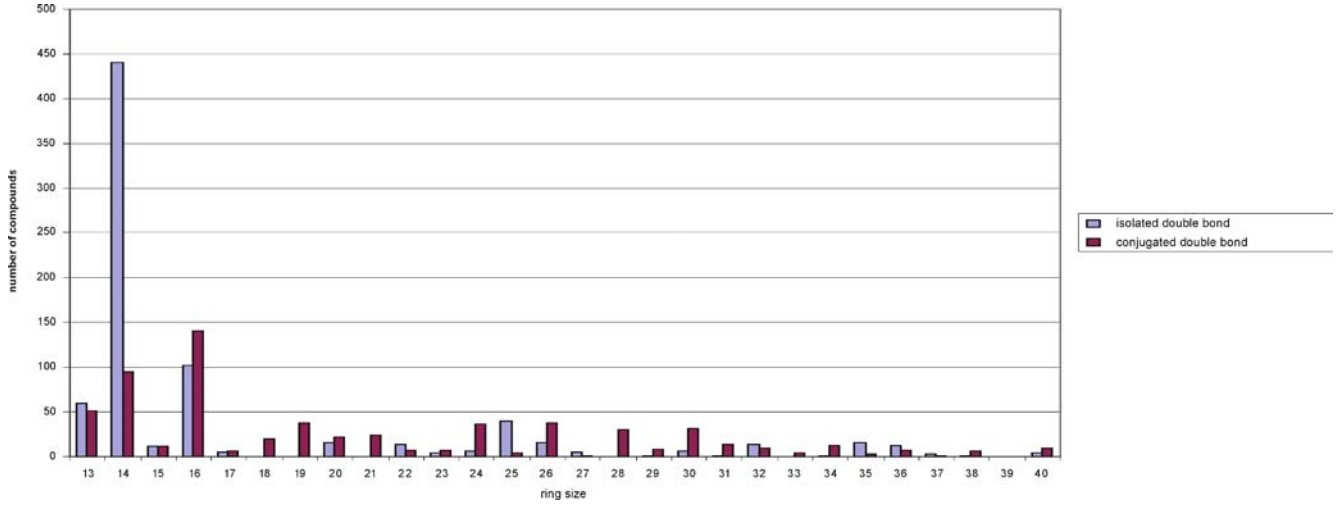

Figure 4. Distribution of at least one C═C double bond or at least two non-aromatic conjugated C═C double bonds within the ring structures.

are mainly polyketide-type compounds with typically two or three oxazole residues in the macrocyclic framework, such as the halichondramides and the disorazoles. Although most oxygen-containing macrocycles with ring size >16 are lactones, significant portions of 13–15-membered oxygen-containing macrocycles do not contain a lactone moiety, but are mainly ethers. For 13- and 14-membered macrocycles, the amount of non-lactone, oxygen-containing rings corresponds quite well to the abundance of lactams, and this suggests that most of these compounds are cyclopeptide ethers of the *ansa*-cyclopeptide alkaloid type [28]. For the underrepresented 16-membered glycopeptides such as vancomycin (**3**) and derivatives, the same connection, ether–oxygen + lactam applies, but in contrast to the 14-membered rings, the 16-membered ones contain many lactams, which do not simultaneously have an ether oxygen.

Also, for 16-membered rings, an especially high frequency of oxygen-containing rings and corresponding lactone-containing rings was observed. These compounds are mainly of polyketide origin.

Cyclopeptides, having ring sizes of $3n$ (n being the number of amino acids in the cycle) can be expected to constitute a significant proportion of naturally occurring macrocycles. Indeed, Figure 2 shows relatively high occurrence of 18-, 21-, 24-, 30-, and 33-membered macrocycles. In some cases, however, ring sizes $3n + 1$ prevail, particularly ring sizes 19, 25, 28. This might be attributed to the occurrence of cyclodepsipeptides, compounds where one peptide bond is replaced by a lactone bond connecting the peptide C-terminus and a β-oxidized amino acid, such as the L-*allo*-threonine residue in pristinamycin I_A (**7**) (Figure 5). Some classes of cyclodepsipeptides may also be responsible for the predominant occurrence of high molecular weight compounds (>1000 g mol^{-1}) in 25-membered macrocycles.

In Figure 4, the occurrence of at least one isolated carbon–carbon double bond in comparison to non-benzenoid conjugated double bonds is shown. Isolated double bonds occur particularly frequently in 14-membered rings. This observation is consistent with the assumption that most of the all-carbon 14-membered ring compounds are macrocyclic diterpenes. In contrast, conjugated double bonds predominate

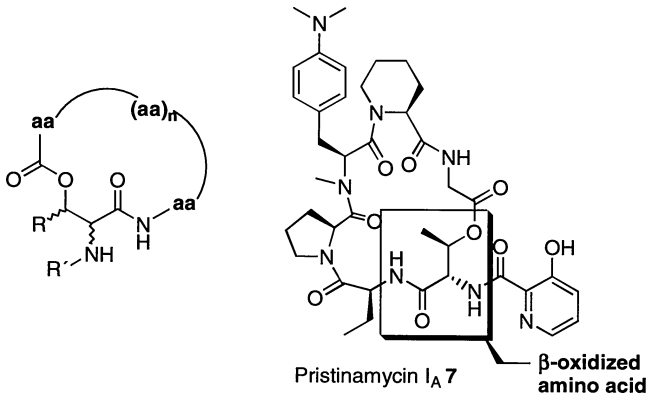

Figure 5. Cyclodepsipeptides (aa = amino acid).

in 16-membered macrocycles (also constituting a significant part of all-carbon ring compounds). Presumably, a different class of cyclic diterpenes or polyketides is involved here. The relatively high frequency of conjugated double bonds in 24-, 26-, 28-, and 30-membered macrocycles is also remarkable. For these ring sizes, hardly any all-carbon rings were found and, presumably, many of these compounds are lactones of the polyketide type.

The observation that macrocycles frequently contain benzene rings incorporated into the macrocycle, especially in the form of aryl ether and biphenyl units, spurred us to analyze the occurrence of both phenylene groups (Figure 6) and phenyl ethers (Figure 7) in macrocyclic natural products. For both series, 1,2- (*ortho*), 1,3- (*meta*), or 1,4- (*para*) linked benzene rings were distinguished. The main conclusion is that these moieties occur primarily in the smaller macrocycles (<20). The occurrence of 1,4-linked benzene rings (158) in 14-membered cycles is especially high and all of these benzene rings are a part of *para*-substituted phenyl ethers. This is in accordance with the previously stated suggestion that *ansa*-cyclopeptide alkaloid-type compounds represent a significant part (some 160 compounds) of naturally occurring 14-membered macrocycles [28, 29]. In the smaller 13-membered rings, 1,2- and 1,3-linked structures are common (a 1,4-linkage would probably have too much ring strain, cf. **11**). Indeed, a predominance of 1,4-arylethers can be expected in cyclopeptides with tyrosine as building block. A different attack during the cyclization mechanism of the same (oxidized) intermediate or a rearrangement can lead to the corresponding 13-membered 1,3-*ansa*-derivatives instead [28].

Whereas more than half of the 1,3-linked benzene rings are part of the corresponding phenyl ethers, this is not the case for the 1,2-linked compounds. In general, 1,2-linked phenyl ethers are quite rare, although 20 of these compounds, probably salicylic acid derivatives, were found among 18-membered macrocycles. In 16-membered macrocycles, benzene rings are quite common, but only a minority is part of a phenyl ether moiety. However, the number of basic structures is accompanied by a considerable number of derivatives (some 140 glycopeptides) that are not included in this analysis for the reasons stated previously.

Of the larger macrocycles, 24- and 28-membered rings most frequently contain benzene rings, and most of the 1,3- and 1,4-linked structures simultaneously form part of the corresponding phenyl ethers. In these cases, 1,3- and 1,4-linked phenyl ethers usually occur in identical amounts if 3,4′-substituted diphenyl ether units are involved. The latter also applies to 18-membered ring substructures.

Overall, the majority of the phenyl ether and the phenylene units are derived from tyrosine metabolism in various oxidation and substitution states. Salicylic acid derivatives form an exception to this rule.

In order to estimate the biosynthetic origin of macrocyclic natural products, we analyzed the occurrence of pyrrolidine (mostly originating from proline, as a measure for peptide

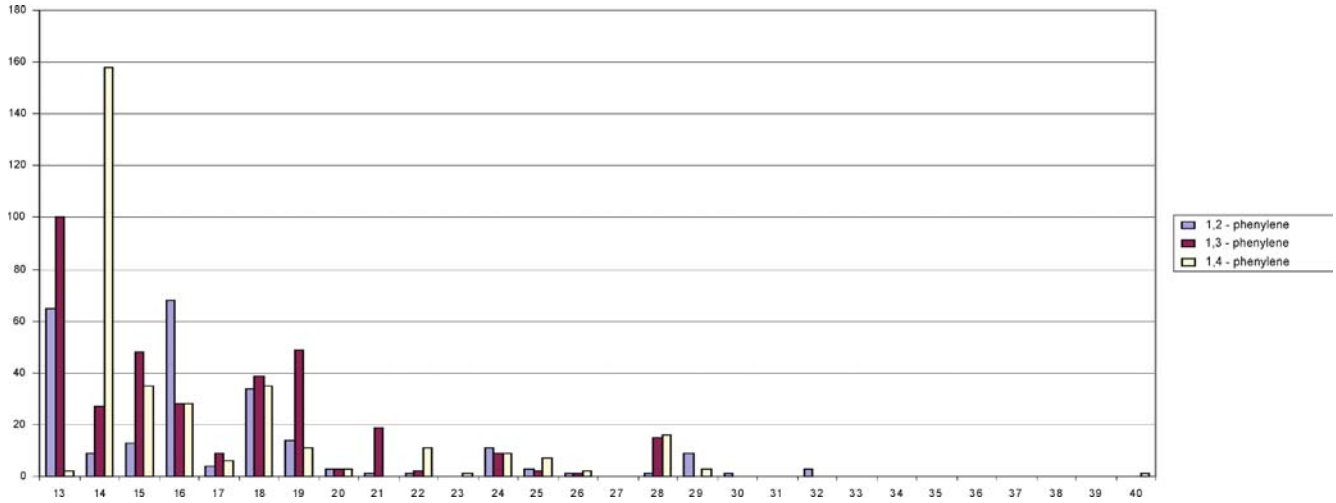

Figure 6. Distribution of phenylene rings within the macrocycles.

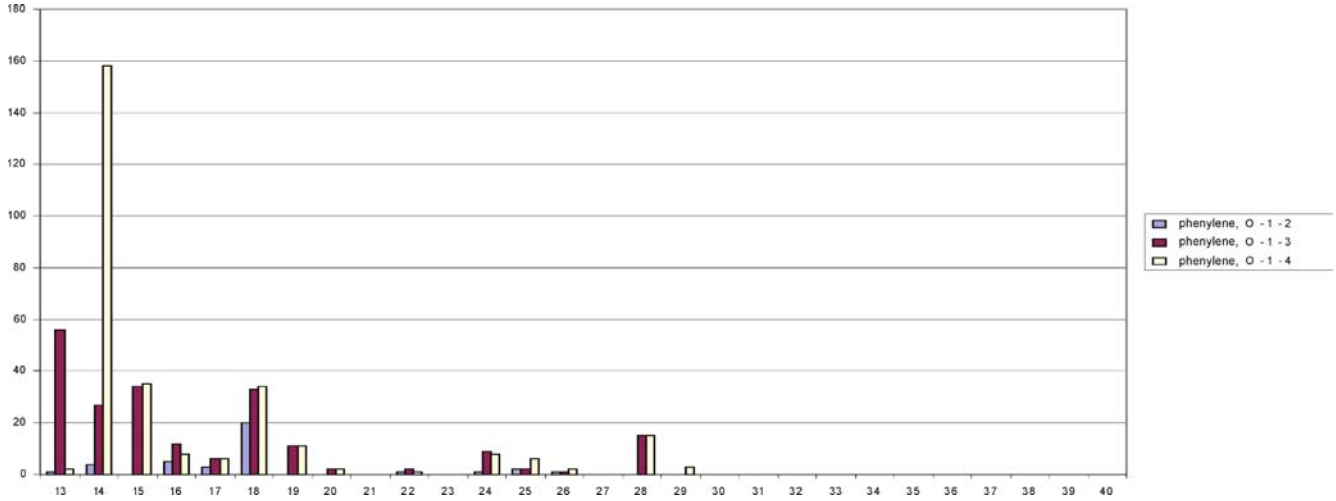

Figure 7. Distribution of phenolic rings including the phenolic oxygen atom inside the macrocyclic ring.

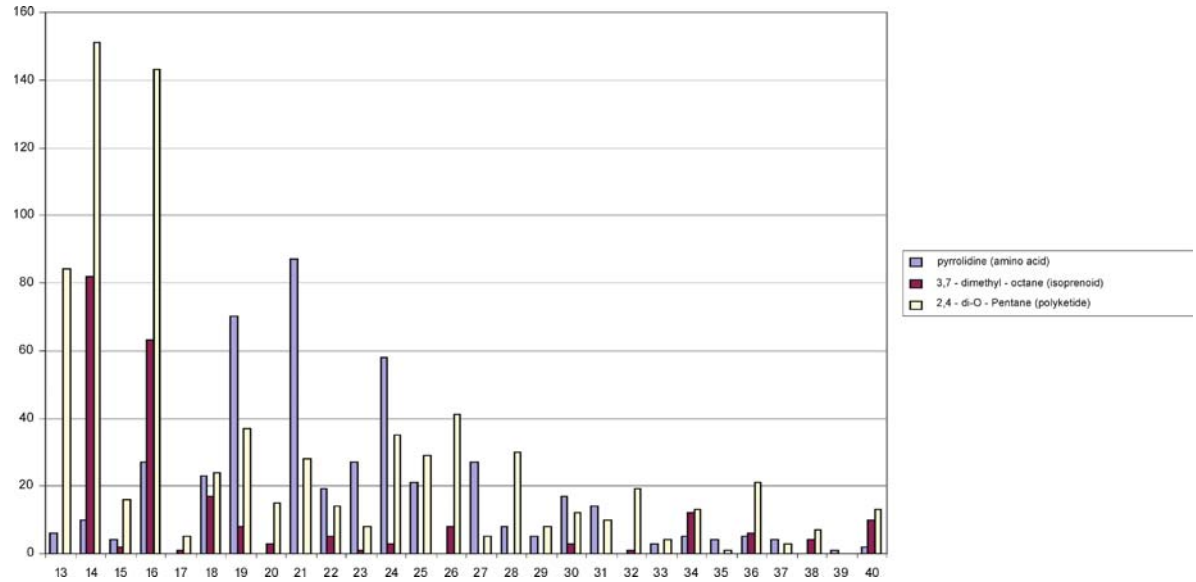

Figure 8. Distribution of pyrrolidine rings including the nitrogen atom inside the macrocycles, and of typical isoprene and polyketide fragments within the ring systems.

(-like) compounds) fused to the macrocycle at the C–N bond, 3,7-dimethyloctane units (with the longest linear chain in the ring, as a measure for isoprenoid abundance) and 2,4-di-*O*-substituted pentane units (with all carbon atoms in the ring, as an indication for polyketide compounds). These moieties were selected because they were deemed the most typical structural elements likely to result from the respective metabolic pathways. These indicator elements have a limited absolute value, but we expected them to be representative in a relative sense. Of course, proline-derived pyrrolidine is not present in all small cyclic peptides, but the relative value of this indicator function is not lost by such known limitations. The results are given in Figure 8.

Pyrrolidine rings occur most frequently in 19-, 21- and 24-membered macrocycles, confirming the previously stated suggestion that cyclo(depsi) peptides are the main sources for naturally occurring macrocycles of these ring sizes. The 3,7-dimethyloctane unit is most common in 14- and 16-membered macrocycles, and is also significant in 13-membered ones. Regarding the high number of all-carbon 13- and 14-membered macrocycles (see Figure 2), it can be assumed that such compounds are cyclized diterpenes. All-carbon rings are rare in 16-membered structures, but lactones are quite common and diterpene-derived lactones may contribute some compounds. However, 3,7-dimethyloctane units (i.e. the geranyl/neryl skeleton) may also be derived from polyketide metabolism, as is the case in the biosynthesis of epothilone B (**4**) [30–32], and this is most likely the case here. In support of this explanation, the isoprenoid indicator unit can also be found in several 34- and 40-membered macrocycles. These compounds are all polyketide (di)lactones of the niphimycin (34), swinholide (40), and mathemycin (40) families.

Finally, 2,4-di-*O*-pentane units are most common in 13-, 14-, and 16-membered rings. Furthermore, this unit occurs relatively often in larger macrocycles, especially in those with an even number of atoms in the ring. This makes sense from a biosynthetic perspective, as the polyketide chain is always elongated by two carbon atoms, and there is typically an odd number of carbons between two oxygen functionalities. Since lactonization is the main macrocyclization method of polyketides, an even-membered macrocycle results.

Another aspect of many macrocycles is a repeating pattern of charge/polarity distribution within the molecule. Many structures show an apolar, lipophilic side and a polar, hydrophilic side. In many cases, they will also have an additional small heterocyclic (e.g. imidazole/oxazole/thiazole [33], sugar, etc.) or substituted aromatic moiety, which maybe endo- or exo-cyclic, resulting in a general structure **8**, shown with an endocyclic small heterocycle. This empirical observation was not yet backed by cheminformatic analyses nor is it related to ring size, since it would be necessary to calculate minimized 3D-structures, charge distribution and heterocyclic moiety distributions for all macrocycles and then compare and cluster these three-dimensionally. This information is not available at this time. However, this frequently observed phenomenon is not limited to membrane-active polyene macrocycles, such as amphothericin B (**5**) or nystatin. It is also found in many smaller polyketides (for example see FK-506, **1**; epothilones like **4**; and kendomycin [36–40], **12**), in macrocyclic glycolipids (such as sophorolipid lactone, **10**) [34], to some degree in cyclo(depsi)peptides (see arylether-cyclopeptides such as **3**, the *ansa*-cyclopeptide alkaloids from plants, see **11**, or K-13, **6**) [35] and hybrid polyketide/peptide compounds (such as the clinically applied anticancer agent maytansin, **9**). Charge and polarity distribution are not necessarily located "along the ring" as in the examples given, but may occur with respect to top and bottom face (β- and α-face) as seen in cyclic peptides [41] and in the anticancer taxanes (Taxol®, Taxotere®, v.i.) [42–44], derived from the 12-membered verticillene, which possess a polar α-surface and an apolar β-surface. Finally, differentially polar inner and outer ring sides are possible as seen in some cyclodextrins, but are rare [3, 45–49].

Biosynthetic origin of macrocyclic diversity – the production concept

Although the variety in structures is large, the variability in substructures is essentially limited to the available biosynthetic pathways. The complexity arises from the versatility of the separate metabolic pathways and the possibility to combine them (Figure 10) [3, 4].

The basis of these metabolic pathways is formed by simple, repetitive processes, mostly iterative in nature. These include (1) polyketide pathways related to fatty acid metabolism, connecting primarily acetate, malonate, and methylmalonate/propionate units in aldol (Claisen) type reactions; (2) isoprenoid metabolism, connecting isopentenyl and dimethylallyl diphosphate building blocks by Friedel-Crafts-like reactions; (3) peptide biosynthesis condensing amino acids to form amide bonds, and (4) sugar condensation to give oligo- and polysaccharides by acetal formation. The structural diversity resulting from these primarily linear processes is limited by the discrete number of monomers, but the diversity increases dramatically through subsequent cyclization reactions leading to (1) aromatic or macrocyclic polyketides; (2) terpenoids and steroids, or (3) alkaloids or cyclopeptides, respectively. Sugars are rarely directly incorporated into macrocyclic rings of $\leq$40 members (exceptions are cyclodextrins [45–48] and macrocyclic sugar lipids [34]), but sugars and amino sugars are of extraordinary importance in the periphery of other macrocycles.

The structural diversity is further increased by connecting products derived from the major metabolic pathways to form conjugates. Such conjugations often involve sugars to produce glycosylated compounds including polyketide glycosides such as erythromycin, steroid glycosides such as digitoxin, and glycopeptides such as vancomycin.

A final possibility to enhance diversity involves post-modification reactions, including oxidation and reduction,

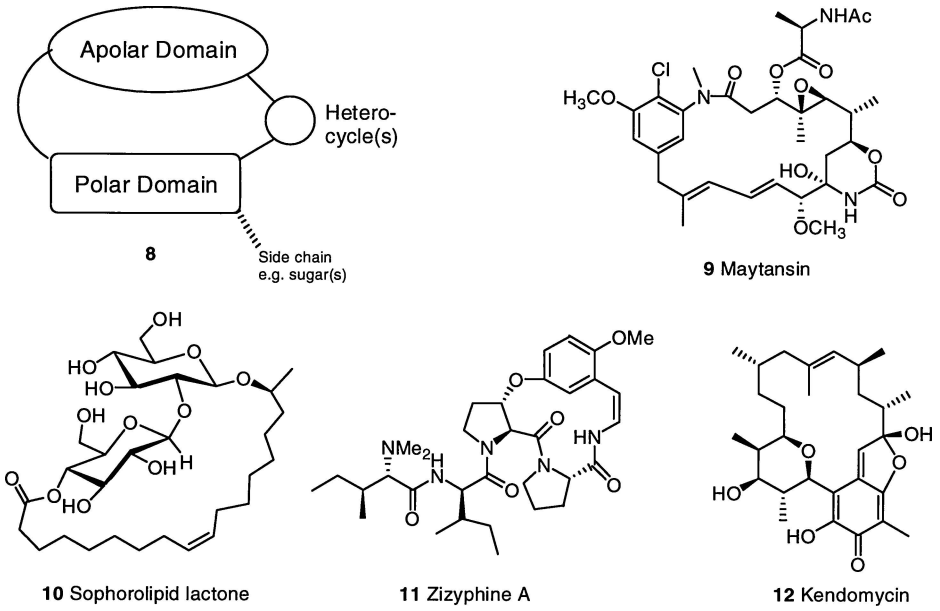

Figure 9. A polarity domain distribution found in many natural macrocycles. Polar domains are usually rich in hydroxy, keto, amide, and sometimes amino groups, apolar domains often have alkyl (methyl), aryl (phenylene), ether, halogen, or alkene groups. The heterocycle(s), mostly 5- or 6-membered rings, can be endocyclic or exocyclic. Some examples are given in Figure 1 (**1**, **3**–**6**), and here with compounds **9**–**12**.

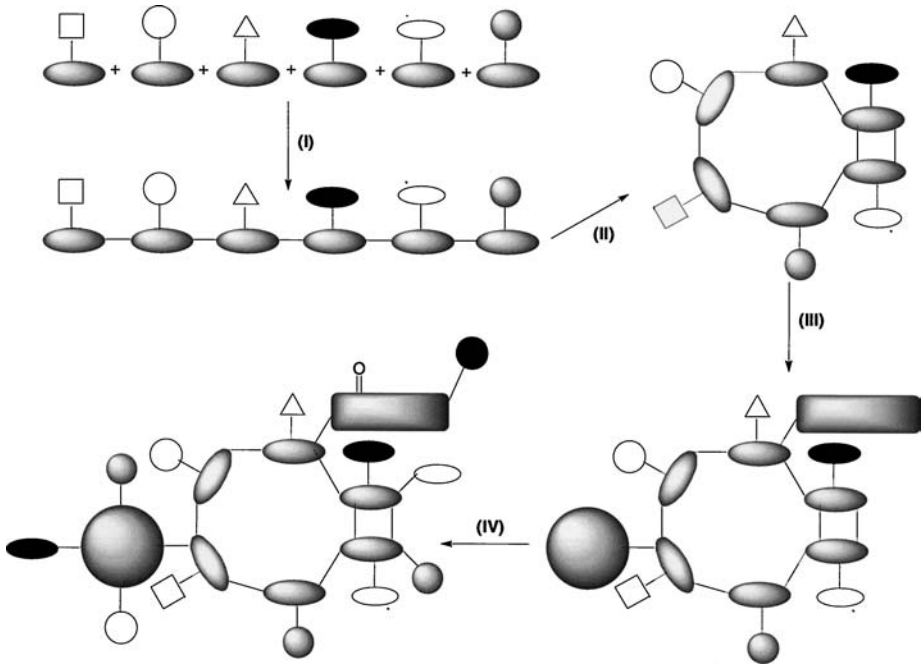

Figure 10. Natures strategy to macrocycle diversity: a typical biosynthesis of a complex natural product: (I) iterative oligomerization, (II) cyclization, (III) combination/conjugation, and/or (IV) (post-)modification.

heterocycle formation, eliminations, and/or side chain attachment reactions. The course of steps III and IV in Figure 10 may be inverted or repeated.

The macrolide antibiotic erythromycin (Figure 11) provides a good example for diversity enhancement by macrocyclization, polyketide and sugar metabolism, and post-modification. In its biosynthesis, a linear polyketide is cyclized to form a macrocyclic aglycon precursor, which is subsequently modified by two oxidation reactions. Two uncommon sugars are synthesized by post-modification of common sugars, and subsequently combined with the polyketide-derived macrocycle to form a complex natural product with a remarkable biological activity.

The question that arises now is: How does nature achieve such complex processes? Or, more precisely, how is such a complex product synthesized alongside primary metabolism?

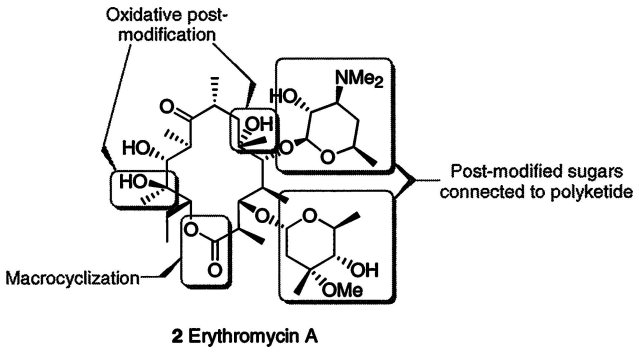

Figure 11. Complexity enhancement in erythromycin biosynthesis after the linear polyketide formation.

These questions are best answered by regarding well-studied examples. Here, of course, we will focus on the macrocyclization reaction. The biosynthesis of the antibiotic erythromycin B has been studied in great detail [50–53]. The biosynthesis of the complete aglycon 6-deoxyerythronolide B (6-dEB, **13**) takes place in a polyketide synthase (PKS), a protein complex often compared to an assembly line. This specific PKS, 6-deoxyerythronolide B synthase (DEBS) consists of three polypeptide chains (DEBS 1, 2, and 3). Each chain comprises two modules, and each module consists of several domains (see Figure 12). Each module adds an acyl residue to the growing polyketide chain by the combined action of its domains.

All six modules have an acyl carrier protein (ACP) domain that binds one of the six acyl residues of 6-dEB, selected by the adjacent acyl transferase (AT) domain, as a thioester. A ketosynthase (KS) domain then transfers the ACP-bound acyl residue to the next module in a Claisen-type reaction. Often, the resulting ketone is reduced by a ketoreductase (KR) domain. The growing polyketide chain progresses to the final module, which has an additional thioesterase (TE) domain. This module breaks the thioester bond, presumably by nucleophilic attack of a serine residue OH group. The protein-bound acyl chain then undergoes transesterification to give the macrolide 6-dEB. Substrate specificity of the individual modules has been the subject of research. Generally, the main determinants of specificity are the α- and β-stereocenters and/or a specific chain length (range). The functional tolerance in the rest of the substrate is quite high. Therefore, the modules function more or less independently from each other. This has been exploited by recombining modules, even from different organisms, to generate novel macrolides. This combinatorial biotechnology is a promising concept, and it has already led to the identification of new lead structures [50–52], although its applicability is limited to cases where substrate specificity is sufficiently high, so that the modules (or single enzymes) are more or less independent of each other and the activity of one module (enzyme) does not depend on the previous one.

Despite these achievements, the mechanistic details of PKSs are still quite unclear. Perhaps the most intriguing function of the synthase complex is the cyclase, which forms a specific macrocycle by intramolecular nucleophilic attack in the presence of several other nucleophiles. In synthetic chemistry, this would avoid the need for protective groups, and the frequently problematic protection/deprotection steps involved. However, the mechanism of these cyclases is largely unclear. Studies on DEBS TE suggest that the TE domain alone does not cause cyclization, and that the adjacent ACP domain may be involved in this process [53].

Certain cyclopeptides are synthesized by non-ribosomal peptide synthetases (NRPSs), in a manner similar to PKSs [54–58]. The peptide residues are connected to peptidyl

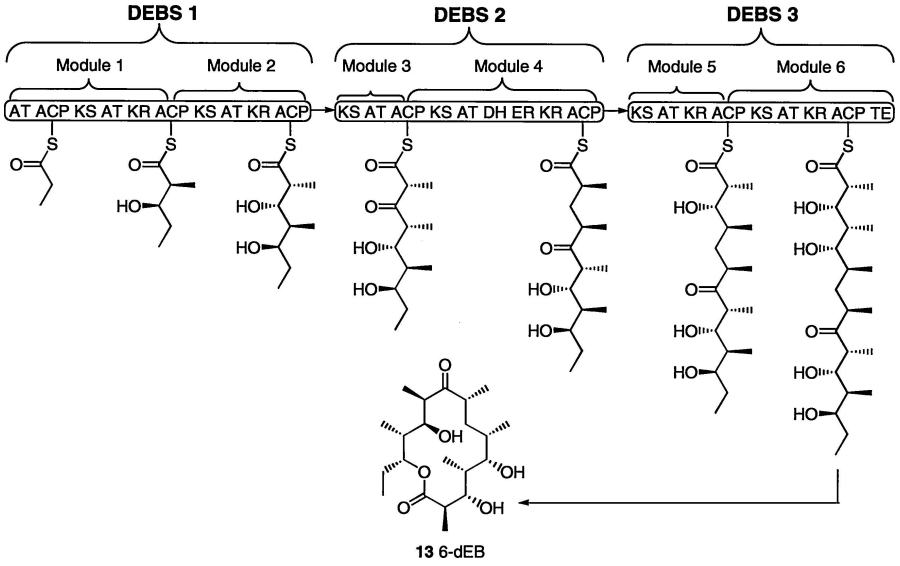

Figure 12. Schematic representation of 6-deoxyerythronolide B synthase (DEBS).

carrier protein (PCP) domains as thioesters, and transferred to the next module by condensation (C) domains, forming a new peptide [54–56] or depsipeptide bond [59]. In some cases, additional domains are included, such as epimerase (E), which transforms amino acid residues to their D-form, or *N*-methylation (M) domains. Again at the end of the terminal module, a TE domain is located, which directs and catalyses the intramolecular attack of the *N*-terminal amino group or a side chain nucleophile to give the cyclo(depsi)peptide (cf. the TE in PKS-Clusters). Chemical diversity is even increased by combining peptide and polyketide modules to form hybrid synthases, which are responsible for a wide variety of interesting natural products, including many macrocycles [54, 56].

The linear C_{10}, C_{15}, C_{20}, and C_{30}-isoprenoids are formed by sequential addition of carbocation intermediates derived from dimethylallyl or oligoprenyl diphosphates to isopentenyl diphosphate, derived from either the mevalonate (MEV) or the methylerythrosediyl cyclodiphosphate (MEP or DXP) pathways [60, 61]. The iterative generation of linear terpenoid precursors is different from that of peptides and polyketides in that the process is much less complex and does not involve stereocenters. The only variation found is that head-to-head cyclization of two oligoprenyl diphosphates forms squalene or phytoene. Nevertheless, this limited diversity generation in the linear precursors is compensated for by the many cyclization modes of isoprenoids, forming an almost endless number of steroids (from squalene), carotenoids (from phytoene), and terpenoids (from oligoprenyl diphosphates). The cyclization is followed by rich post-cyclization modification chemistry, especially oxidations and rearrangements. In contrast to polyketides and cyclopeptides, the stereocenters in terpenoids are introduced by selective cyclization and post-modification reactions. The divergent synthetic routes of terpenoid biosynthesis are a consequence of the evolution of diversity in terpene cyclases [62, 63] such that different active site templates evolved as chaperones for reactive carbocation intermediates forcing different cyclization pathways.

Terpene cyclases require the presence of a Lewis-acidic divalent metal ion such as Mg^{2+} or Mn^{2+} to aid the formation of an intermediate prenyl diphosphate cyclase complex with activation of the diphosphate as leaving group [27, 64–69]. A variety of isoprenoid synthases (cyclases), most of which are from plants and fungi, have been cloned, purified, and extensively studied [62, 63, 67, 68, 70, 71]. X-ray crystal structures of some monoterpene [72] and sesquiterpene cyclases [67, 68, 70] have also been recently reported. Since the vast majority of isoprenoids is produced by eukaryotes, there have been few reports of eubacterial isoprenoid cyclases and their genes, such as squalene-hopene cyclases [73]. Almost all of the naturally occurring terpenoid macrocycles or derivatives fit into the geranylgeranyl diphosphate (GGPP) derived diterpene family with C_{20} (Figure 13), supporting our preliminary idea that the high numbers of 13- and 14-membered, as well as some 15- and 16-membered, all-carbon macrocycles could be derived from this precursor rather than a polyketide one.

Diterpene cyclases are classified into two major types with respect to their modes of cyclizations [74]. The first type of diterpene cyclases catalyzes a reaction initiated by protonation at the 14,15-double bond of GGPP, followed by consecutive cyclization steps, which furnish polycyclic skeletons, but not macrocycles. The polycycle formation probably occurs in a similar manner to that of the triterpene and squalene-hopene cyclases [64, 65, 73, 75, 76]. Biosynthetic evidence and X-ray structures of triterpene cyclases [77, 78] explain why polycyclic triterpenoid skeletons are generally found in nature [71, 79] rather than macrocyclic ones. The available crystal structures are from plant enzymes. However, a high degree of sequence identity, the presence of aromatic residues to stabilize and retain the formed carbocations within

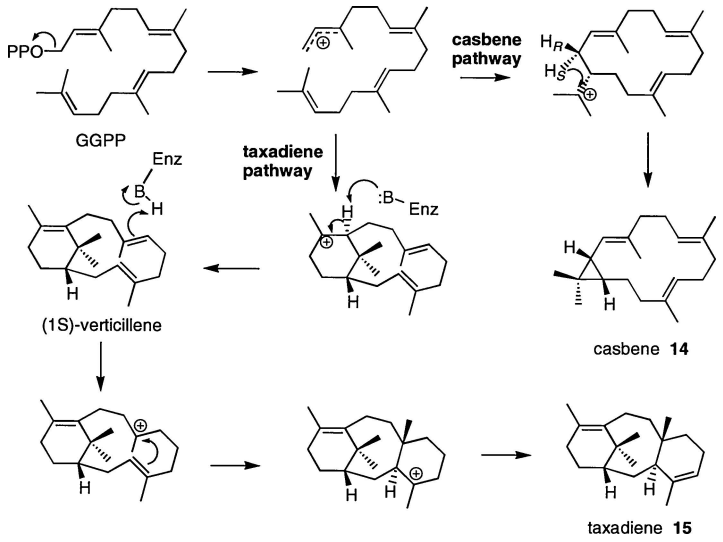

Figure 13. Biosynthetic pathways to casbene and taxadiene.

the cavity, the highly conserved and essential aspartate-rich motif(s) as initial proton donor residues, and M^{2+} binder found in different organisms suggest similar, evolutionary preserved, catalytic mechanisms for all prenyl transferases, including terpene cyclases. Moreover, eukaryotic as well as prokaryotic oligoprenyl transferases, synthases, or cyclases present similar elements. These similarities support the idea of a widespread mechanism to protect the intermediate carbocations against the addition of water or deprotonation by a base. This protected system allows the shift of the hydride and methyl groups along a thermodynamically and kinetically favorable cascade [27, 79] culminating in the (cyclizing) addition to a double bond of an aromatic ring (aromatic prenyl transferases [69, 80, 81]), another molecule of isopentenyl diphosphate (synthases, elongases), or an intramolecular one as required for cyclases. In accordance with Baldwin's cyclization rules of cations to alkenes (*n*-endo or *n*-exo-trig reactions), the most common ring sizes are 5- and 6-membered ones. Strained three-membered [27], and medium sized (8–12-membered) rings are not rare, but only few macrocycles are formed. The latter will require an excellent control of folding of a long lipophilic tail without allowing subsequent transannular reactions.

Casbene [82, 83] and taxadiene [84, 85] synthases are representative of the second class of diterpene cyclases. Casbene (**14**) is a macrocyclic diterpene phytoalexin with antibacterial and antifungal activity [64, 65], whereas taxadiene (**15**) represents the diterpenoid portion of the important anticancer agent Taxol® (Paclitaxel) [31, 42–44]. Both compounds are formed from geranylgeranyl diphosphate by the action of their corresponding cyclases through ionization of the diphosphate group to an allylic carbocation [69] followed by cyclization and deprotonation to the olefins (Figure 13). The formation of the macrocyclic casbene involves the simultaneous formation of a cyclopropane ring [27].

Although the taxadiene skeleton is not macrocyclic, but rather a polycyclic medium-sized ring derivative, its biosynthetic precursor (1*S*)-verticillene is already 12-membered. The taxanes have continued to attract considerable interest and there are estimated to be about 400 known naturally occurring taxoids. Recent updated reviews of the chemistry and biology of these compounds as well as major advances in the genetics of Taxol® biosynthesis have appeared [42, 43], including the details of the cyclization pathway to taxadiene and the post-cyclization functionalization toward taxol-type structures.

The huge amount of bioactive macrocyclic diterpenoids, which have been reported in the recent years [71, 86–88] (see Figures 13 and 14), as well as the high number of 14-membered macrocycles, suggest that this biosynthetic

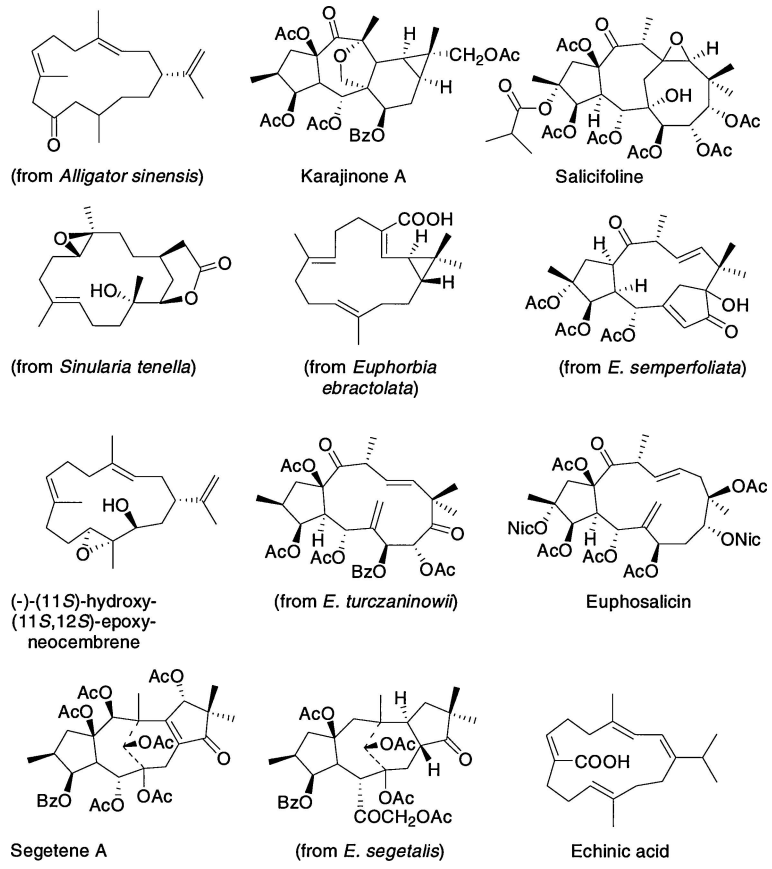

Figure 14. Structures of some macrocyclic diterpenoids and subsequent transannular cyclization products [86–88].

mechanism is widespread in a large variety of organisms, and this dominance is reflected in our statistical results. Examination of their striking structures confirms that most of them possess a casbene-type skeleton and therefore their chemical diversity is primarily due to post-cyclization modifications, which include transannular cyclization, lactonization, oxidation and epoxidation, alkyl migration, and acylation steps.

For the synthetic chemist, the most remarkable essence of these systems is the striking similarity to solid phase synthesis, in this case, combined with a cyclative cleavage strategy. Indeed, this is a very elegant and effective strategy for the solid phase synthesis of cyclic compounds [3, 89–94]. However, nature has taken the strategy to the next level, where each of the reactants is "solid-phase"-bound, and is connected to a growing "solid-phase"-bound chain in an ordered, predetermined fashion, followed by cleavage/cyclization at exactly the desired moment with prefolded chains to achieve defined ring sizes and even stereocenters where required. Although some iterative solid-phase based strategies have been developed (Merrifield peptide synthesis, DNA-synthesizer, etc.), a similarly perfect modular 'molecular assembly line' for laboratory chemistry to make products of secondary metabolism seems far away. However, other concepts from nature have already found widespread application. Cyclative cleavage strategies, for example, are well established in solid-phase synthesis, and have contributed to the synthesis of many cyclic compounds [3, 90–92, 95].

Recently, cyclases from NRPSs have been employed in the construction of cyclopeptides [41, 57, 59, 96]. The use of peptidyl phenylthioesters, which were synthesized using standard solid-phase synthesis [96], was especially effective. When these were treated with a recombinant protein consisting of the terminal PCP and TE domain of the corresponding NRPS, the peptidyl chains were spontaneously transacylated to the PCP domain, subsequently transferred to the TE domain, and finally cyclized to give the natural cyclopeptides. In some cases, hydrolysis and/or spontaneous aspecific cyclization occurred in competition to the transacylation. In these cases, however, results were dramatically improved by the use of less reactive, substituted phenylthioesters. Even solid-phase bound peptidyl esters can be substrates for an isolated TE domain, not only for the natural peptide substrate, but also for sufficiently similar derivatives [41, 58]. This study demonstrates that TEs have a certain degree of flexibility in substrate acceptance, and the compatibility with solid-phase bound substrates makes this a very promising strategy, elegantly integrating chemical synthesis and biotechnology [41]. Recent, dramatic improvements of enzyme reactions with polymer bound substrates have been achieved [97–99], and these may be combined with macrocyclizing enzyme reactions other than peptide formation, for example, macrolactonization and lactam formation [100–103].

In conclusion, it is evident that nature's strategies for the synthesis of complex macrocycles are a great source of inspiration, providing opportunities both for synthetic chemistry and biotechnology, as well as integrated approaches. With the current dataset, it became evident that each metabolic family (e.g. polyketides, cyclopeptides) has different preferred ring sizes and structural elements that dominate. Molecular weight tendencies and some common structural units could be extracted, and polarity distribution trends were discussed. Further studies are under way to refine these analyses, to improve the database, and to correlate and integrate this with biological and other data.

Acknowledgements

The authors are grateful to Prof. Dr. Volkmar Vill and Dr. Pavel Weigner for their help and consultancy with cheminformatics and data mining, and the Deutsche Forschungsgemeinschaft (CERC3 and GRK 894) and the European Union (Combiocat) for financial support.

References

1. Newman, D.J., Cragg, G.M. and Snader, K.M., *Natural products as sources of new drugs over the period 1981–2002*, J. Nat. Prod., 66 (2003) 1022–1037.
2. Rouhi, A.M., *Rediscovering natural products*, Chem. Eng., (2003) 77–107.
3. Wessjohann, L.A. and Ruijter, E., *Total and diversity-oriented synthesis of natural product(-like) macrocycles*, in Mulzer, J.H. (Ed.) Topics in Current Chemistry, Vol. 243: Springer, Heidelberg, 2004, pp. 137–184.
4. Henkel, T., Brunne, R.M., Mueller, H. and Reichel, F., *Statistical investigation into the structural complementarity of natural products and synthetic compounds*, Angew. Chem., 111 (1999) 688–691; Angew. Chem. Int. Ed., 38 (1999) 643–647.
5. Feher, M. and Schmidt, J.M., *Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry*, J. Chem. Inf. Comput. Sci., 43 (2003) 218–227.
6. Link, A., *Maßgeschneiderte Nadeln aus dem kombinatorischen Heuhaufen*, Pharmazeutische Zeitung, (2001) 3760–3766.
7. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*, Adv. Drug Delivery Rev., 46 (1997) 3–25.
8. Lee, M.-L. and Schneider, G., *Scaffold architecture and pharmacophoric properties of natural products and trade drugs: Application in the design of natural product-based combinatorial libraries*, J. Comb. Chem., 3 (2001) 284–289.
9. Nielsen, J., *Combinatorial synthesis of natural products*, Curr. Opin. Chem. Biol., 6 (2002) 297–305.
10. Wessjohann, L.A., *Synthesis of natural product-based compound libraries*, Curr. Opin. Chem. Biol., 4 (2000) 303–309.
11. Breinbauer, R., Manger, M., Scheck, M. and Waldmann, H., *Natural product guided compound library development*, Curr. Med. Chem., 9 (2002) 2129–2145.
12. Breinbauer, R., Vetter, I.R. and Waldmann, H., *Von Proteindomänen zu Wirkstoffkandidaten–Naturstoffe als Leitstrukturen für das Design und die Synthese von Substanzbibliotheken*, Angew. Chem., 114 (2002) 3002–3015; Angew. Chem. Int. Ed., 41 (2002) 2878–2890.

13. Stahl, P., Kissau, L., Mazitschek, R., Huwe, A., Furet, P., Giannis, A. and Waldmann, H., *Total synthesis and biological evaluation of the nakajiquinones*, J. Am. Chem. Soc., 123 (2001) 11586–11593.
14. Stahl, P., Kissau, L., Mazitschek, R., Giannis, A. and Waldmann, H., *Naturstoffanaloga als Tyrosinkinaseinhibitoren: Identifikation von IGF1R-, Tie-2- und VEGFR-3-Inhibitoren*, Angew. Chem., 114 (2002) 1222–1226; Angew. Chem. Int. Ed., 41 (2002) 1174–1178.
15. Lindsley, C.W., Chan, L.K., Goess, B.C., Joseph, R. and Shair, M.D., *Solid-phase biomimetic synthesis of carpanone-like molecules*, J. Am. Chem. Soc., 122 (2000) 422–423.
16. Corey, E.J. and Cheng, X.-M., The Logic of Chemical Synthesis, 1st edn., John Wiley & Sons, NewYork, 1989.
17. Nuhn, P., Naturstoffchemie, 3rd edn., S. Hirzel, Leipzig, 1997.
18. Nuhn, P. and Wessjohann, L.A., Naturstoffchemie – Mikrobielle, Pflanzliche und Tierische Naturstoffe, 4th edn., S. Hirzel, Stuttgart, 2005.
19. Dictionary of Natural Products on CD-ROM, Version 11:2, Chapman & Hall/CRC Press, 2003.
20. MDL Information Systems GmbH, http://www.beilstein.com/.
21. Schmidt, H., Porzel, A., Wessjohann, L. and Vill, V., Phytobase – Trademark of the Leibniz-Institute of Plant Biochemistry, Halle (D), http://www.phytobase.de/.
22. SciDex—Volkmar Vill, LCI-Publisher GmbH, Hamburg (D), http://www.lci-publisher.com/.
23. Nicolaou, K.C., Boddy, C.N.C., Bräse, S. and Winssinger, N., *Chemistry, biology, and medicine of the glycopeptide antibiotics*, Angew. Chem., 111 (1999) 2230–2287; Angew. Chem. Int. Ed., 38 (1999) 2096–2152.
24. Bode, H.B., Bethe, B., Hofs, R. and Zeeck, A., *Big effects from small changes: Possible ways to explore nature's chemical diversity*, ChemBioChem, 3 (2002) 619–627.
25. Galm, U., Dessoy, M.-A., Schmidt, J., Wessjohann, L.A. and Heide, L., *In vitro and in vivo production of new aminocoumarins by a combined biochemical, genetic, and synthetic approach*, Chem. Biol., 11 (2004) 173–183.
26. von Roepenack-Lahaye, E., Degenkolb, T., Zerjeski, M., Franz, M., Roth, U., Wessjohann, L., Schmidt, J., Scheel, D. and Clemens, S., *Profiling of arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry*, Plant Physiol., 134 (2004) 548–559.
27. Wessjohann, L.A., Brandt, W. and Thiemann, T., *Biosynthesis and metabolism of cyclopropane-rings in natural compounds*, Chem. Rev., 103 (2003) 1625–1647.
28. Gournelis, D.C., Laskaris, G.G. and Verpoorte, R., *Cyclopeptide alkaloids*, Nat. Prod. Rep., 14 (1997) 75–82.
29. Morel, A.F., Machado, E.C.S., Moreira, J.J., Menezes, A.S., Mostar deiro, M.A. and Wessjohann, L.A., *Cyclopeptide alkaloids of Scutia buxifolia (Rhamnaceae)*, Phytochemistry, 47 (1998) 125–129.
30. Hardt, I.H., Steinmetz, H., Gerth, K., Sasse, F., Reichenbach, H. and Höfle, G., *New natural epothilones from Sorangium cellulosum, strains So ce90/B2 and So ce90/D13: Isolation, structure elucidation, and SAR studies*, J. Nat. Prod., 64 (2001) 847–856.
31. Wessjohann, L.A., *Epothilones: Promising natural products with taxol-like activity*, Angew. Chem., 109 (1997) 739–742; Angew. Chem. Int. Ed., (1997) 715–718.
32. Wessjohann, L.A., Scheid, G.O., Bornscheuer, U., Henke, E., Kuit, W. and Orru, R.V.A., *Epothilone synthesis components III and IV: Asymmetrically substituted acyloins and acyloin derivatives, method for the production thereof and method for the production of epothilone B, D, and epothilone derivatives*, In *CAS 136:340534*. 2002: Int. Patent Application WO2002/032844A2 (25.04.2002), EP 200111992 (16.10.2001).
33. Jin, Z., *Muscarine, imidazol, oxazole and thiazole alkaloids*, Nat. Prod. Rep., 20 (2003) 584–605.
34. Fürstner, A., *Total syntheses and biological assessment of macrocyclic glycolipids*, Eur. J. Org. Chem., (2004) 943–958.
35. Bigot, A., Bois-Choussy, M. and Zhu, J., *An efficient total synthesis of K-13, a non-competitive inhibitor of ACE I*, Tetrahedron Lett., 41 (2000) 4573–4578.
36. Marques, M.M.B., Pichlmair, S., Martin, H.J. and Mulzer, J., *Stereocontrolled synthesis of C-arylglycosides applied to the south west fragment of the antibiotic kendomycin*, Synthesis, (2002) 2766–2770.
37. Martin, H.J., Drescher, M., Kaehlig, H., Schneider, S. and Mulzer, J., *Synthesis of the C1–C13 fragment of kendomycin: Atropisomerism around a C-aryl glycoside bond*, Angew. Chem., 131 (2001) 3287–3289; Angew. Chem. Int. Ed., 40 (2001) 3186–3188.
38. Pichlmair, S., Marques, M.M.B., Green, M.P., Martin, H.J. and Mulzer, J., *A novel approach toward the synthesis of kendomycin: Selective synthesis of a C-aryl glycoside as a single atropisomer*, Org. Lett., 5 (2003) 4657–4659.
39. Bode, H.B. and Zeeck, A., *Biosynthesis of kendomycin: Origin of the oxygen atoms and further investigations*, J. Chem. Soc. Perkin Trans., 1 (2000) 2665–2670.
40. Bode, H.B. and Zeeck, A., *Structure and biosynthesis of kendomycin, a carbocyclic ansa-compound from Streptomyces*, J. Chem. Soc. Perkin Trans., 1 (2000) 323–328.
41. Kohli, R.M., Walsh, C.T. and Burkart, M.D., *Biomimetic synthesis and optimization of cyclic peptide antibiotics*, Nature, 418 (2002) 658–661.
42. Kingston, D.G.I., *Taxol, a molecule for all seasons*, Chem. Commun., (2001) 867–878.
43. Walker, F. and Croteau, R.B., *The taxol biosynthetic genes*, Phytochemistry, 58 (2001) 1–7.
44. Wessjohann, L.A., *The first total syntheses of taxol*, Angew. Chem., 106 (1994) 1011–1013; Angew. Chem. Int. Ed. Engl., 33 (1994) 959–961.
45. D'Souza, V.T. and Lipkowitz, K.B., *Cyclodextrins: Introduction*, Chem. Rev., 98 (1998) 1741–1742.
46. Saenger, W., Jacob, J., Gessler, K., Steiner, T., Hoffmann, D., Sanbe, H., Koizumi, K., Smith, S.M. and Takaha, T., *Structures of the common cyclodextrins and their larger analogues-beyond the doughnut*, Chem. Rev., 98 (1998) 1787–1802.
47. Szejtli, J., *Introduction and general overview of cyclodextrin chemistry*, Chem. Rev., 98 (1998) 1743–1754.
48. Uekama, K., Hirayama, F. and Irie, T., *Cyclodextrin drug carrier systems*, Chem. Rev., 98 (1998) 2045–2076.
49. Voigt, B., Schaks, A., Bögel, M., Brandt, W. and Wessjohann, L.A., *Synthesis of steroid containing macrocycles via multi component reactions*, In 2nd International Conference on Multi Component Reactions, Combinatorial and Related Chemistry (MCR 2003), 2003, Genua, Italy.
50. Katz, L., *Manipulation of modular polyketide synthases*, Chem. Rev., 97 (1997) 2557–2575.
51. Khosla, C., *Harnessing the biosynthetic potential of modular polyketide synthases*, Chem. Rev., 97 (1997) 2577–2590.
52. Staunton, J. and Wilkinson, B., *Biosynthesis of erythromyin and rapamycin*, Chem. Rev., 97 (1997) 2611–2629.
53. Gokhale, R.S., Hunziker, D., Cane, D.E. and Khosla, C., *Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase*, Chem. Biol., 6 (1999) 117–125.
54. Moffitt, M.C. and Neilan, B.A., *The expansion of mechanistic and organismic diversity associated with non-ribosomal peptides*, FEBS Microbiol. Lett., 191 (2000) 159–167.
55. Mootz, H.D. and Marahiel, M.A., *Biosynthetic systems for nonribosomal peptide antibiotic assembly*, Curr. Opin. Chem. Biol., 1 (1997) 543–551.
56. Schwarzer, D., Mootz, H.D. and Marahiel, M.A., *Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases*, Chem. Biol., 8 (2001) 997–1010.
57. Trauger, J.W., Kohli, R.M. and Walsh, C.T., *Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the*

tyrocidine nonribosomal peptide synthetase, Biochemistry, 40 (2001) 7092–7098.
58. Bordusa, F., *Enzymes for peptide cyclization*, ChemBioChem, 2 (2001) 405–409.
59. Tseng, C.C., Bruner, S.D., Kohli, R.M., Marahiel, M.A., Walsh, C.T. and Sieber, S.A., *Characterization of the surfactin synthase C-terminal thioesterase domain as a cyclic depsipeptide synthase*, Biochemistry, 41 (2002) 13350–13359.
60. Brandt, W., Dessoy, M.A., Fulhorst, M., Gao, W., Zenk, M.H. and Wessjohann, L.A., *Structural basis for the reductive ring opening of the cyclodiphosphate MEcPP, a crucial transformation in the new DXP/MEP-pathway to isoprenoids*, ChemBioChem, 5 (2004) 311–323.
61. Gao, W., Loeser, R., Raschke, M., Dessoy, M.A., Fulhorst, M., Alpermann, H., Wessjohann, L.A. and Zenk, M.H., *(E)-4-Hydroxy-3-methylbut-2-enyl diphosphate: An intermediate in the formation of terpenoids in plant chromoplasts*, Angew. Chem., 114 (2002) 2716–2719; Angew. Chem. Int. Ed., (2002) 2604–2607.
62. Bohlman, J., Meyer-Gauen, G. and Croteau, R.B., *Plant terpenoid synthases: Molecular biology and phylogenetic analysis*, Proc. Nat. Acad. Sci. U.S.A., 95 (1998) 4126–4133.
63. Trapp, S. and Croteau, R.B., *Genomic organization of plant terpene synthases and molecular evolutionary implications*, Genetics, 158 (2001) 811–832.
64. Dewick, P.M., *The biosynthesis of C5–C25 terpenoids compounds*, Nat. Prod. Rep., 16 (1999) 97–130.
65. Dewick, P.M., *The biosynthesis of C5–C25 terpenoids compounds*, Nat. Prod. Rep., 19 (2002) 181–222.
66. Pyun, H.-J., Wagschal, K.C., Jung, D.-I., Coates, R.M. and Croteau, R., *Monoterpene biosynthesis: Isotope effects associated with bicyclic olefin formation catalyzed by pinene synthases from sage (Salvia officinalis)*, Arch. Biochem. Biophys., (1994) 488–496.
67. Caruthers, J.M., Kang, I., Rynkiewicz, M.J., Cane, D.E. and Christianson, D.W., *Crystal structure determination of aristolochene synthase from the blue cheese mold, Penicillium roqueforti*, J. Biol. Chem., 275 (2000) 25533–25539.
68. Rynkiewicz, M.J., Cane, D.E. and Christianson, D.W., *From the cover: Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade*, Proc. Nat. Acad. Sci. U.S.A., 98 (2001) 13543–13548.
69. Bräuer, L., Brandt, W. and Wessjohann, L.A., *Modeling of the E. coli 4-hydroxybenzoic acid oligoprenyl transferase (ubiA-transferase) and characterization of potential active sites*, J. Mol. Mod., 10 (2004) 317–327.
70. Starks, C.M., Back, K., Chappel, J. and Noel, J.P., *Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase*, Science, 277 (1997) 1815–1820.
71. Conolly, J.D. and Hill, R.A., Dictionary of Terpenoids, Chapman & Hall, New York, 1992.
72. Whittington, D.A., Wise, M.L., Urbansky, M., Coates, R.M., Croteau, R.B. and Christianson, D.W., Biochemistry, 24 (2002) 15375–15380.
73. Brown, G.D., *The biosynthesis of steroids and triterpenoids*, Nat. Prod. Rep., 16 (1998) 653–696.
74. MacMillan, J. and Beale, M.H., Comprehensive Natural Products Chemistry, Elsevier Science Publishing Co., Inc., New York, 1999.
75. Ait-Ali, T., Swain, S.M., Reid, J.B., Sun, T.P. and Kamiya, Y., *The LS locus of pea encodes the gibberellin biosynthesis enzyme ent-kaurene synthase A*, Plant J., 11 (1997) 443–454.
76. Hedden, P. and Kamiya, Y., *Gibbelerin biosynthesis: Enzymes, genes and their regulation*, Ann. Rev. Plant Physiol. Plant Mol. Biol., 48 (1997) 431–460.
77. Wendt, K.U., Poralra, K. and Schulz, E., *Structure and function of a squalene cyclase*, Science, 277 (1997) 1811–1815.
78. Wendt, K.U., Lenhart, A. and Schulz, E., *The structure of the membrane protein squalene-hopene cyclase at 2.0 A resolution*, J. Mol. Biol., 286 (1999) 175–187.
79. Wendt, K.U., Schulz, E., Corey, E.J. and Liu, D.R., *Enzyme mechanisms for polycyclic triterpene formation*, Angew. Chem., 112 (2000) 2930–2952; Angew. Chem. Int. Ed., 39 (2000) 2812–2833.
80. Wessjohann, L.A. and Sontag, B., *Prenylation of benzoic acid derivatives catalyzed by a transferase from Escherichia coli overproduction: Method development and substrate specificity*, Angew. Chem., 108 (1996) 1821–1823; Angew. Chem. Int. Ed. Engl., (1996) 1697–1699.
81. Wessjohann, L.A., Sontag, B. and Dessoy, M.-A., *Enzymatic C–C coupling: The development of aromatic prenylation for organic synthesis*, In Diederichsen, U., Lindhorst, T.L., Wessjohann, L.A. and Westermann, B. (Eds.) Bioorganic Chemistry – Highlights and New Aspects, Wiley-VCH, Weinheim, 1999, pp. 79–88.
82. Huang, K., Huang, Q. and Scott, A.I., *Overexpression, single-step purification, and site-directed mutagenetic analysis of casbene synthase*, Arch. Biochem. Biophys., 352 (1998) 144–152.
83. Hill, A.M., Cane, D.E., Mau, C.J. and West, C.A., *High level expression of Ricinus communis casbene synthase in Escherichia coli and characterization of the recombinant enzyme*, Arch. Biochem. Biophys., 336 (1996) 283–289.
84. Wildung, M. and Croteau, R.B., *A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis*, J. Biol. Chem., 271 (1996) 9201–9204.
85. Hamano, Y., Kuzumaya, T., Itoh, N., Furihata, K., Seto, H. and Dairi, T., *Functional analysis of eubacterial diterpene cyclases responsible for biosynthesis of a diterpene antibiotic, terpentecin*, J. Biol. Chem., 277 (2002) 37098–37104.
86. Hanson, J.R., *Diterpenoids*, Nat. Prod. Rep., 19 (2002) 125–132.
87. Hanson, J.R., *Diterpenoids*, Nat. Prod. Rep., 17 (2000) 165–174.
88. Hanson, J.R., *Diterpenoids*, Nat. Prod. Rep., 20 (2003) 70–78.
89. Lambert, J.N., Mitchell, J.P. and Roberts, K.D., *The synthesis of cyclic peptides*, J. Chem. Soc. Perkin Trans. 1, (2001) 471–484.
90. van Maarseveen, J.H., den Hartog, J.A.J., Engelen, V., Finner, E., Visser, G. and Kruse, C.G., *Solid phase ring-closing metathesis: Cyclization/cleavage approach towards a seven membered cycloolefin*, Tetrahedron Lett., 37 (1996) 8249–8252.
91. van Maarseveen, J.H., *Solid phase synthesis of heterocycles by cyclization/cleavage methodologies*, Comb. Chem. High Thr. Scr., 1 (1998) 185–214.
92. Park, K.-H. and Kurth, M.J., *Cyclo-elimination release strategies applied to solid-phase organic synthesis in drug discovery*, Drugs Fut., 25 (2000) 1265–1294.
93. Akritopoulou-Zanze, I. and Sowin, T.J., *Solid-phase synthesis of macrolide analogues*, J. Comb. Chem., 3 (2001) 301–311.
94. Cristau, P., Vors, J.-P. and Zhu, J., *Solid-phase synthesis of natural product-like macrocycles by a sequence of Ugi-4CR and S_NAr-based cycloetherification*, Tetrahedron Lett., 44 (2003) 5575–5578.
95. Brohm, D., Philipe, N., Metzger, S., Bhargava, A., Müller, O., Lieb, F. and Waldmann, H., *Solid-phase synthesis of dysidiolide-derived protein phosphatase inhibitors*, J. Am. Chem. Soc., 124 (2002) 13171–13178.
96. Sieber, S.A., Tao, J., Walsh, C.T. and Marahiel, M.A., *Peptidyl thiophenols as substrates for nonribosomal peptide cyclases*, Angew. Chem., 116 (2004) 499–504; Angew. Chem. Int. Ed., 43 (2004) 493–498.
97. Ulijn, R.V., Baragana, B., Halling, P.J. and Flitsch, S.L., *Protease-catalyzed peptide synthesis on solid support*, J. Am. Chem. Soc., 124 (2002) 10988–10989.
98. Basso, A., De Martin, L., Gardossi, L., Margetts, G., Brazendale, I., Bosma, A.Y., Ulijn, R.V. and Flitsch, S.L., *Improved biotransformations on charged PEGA supports*, Chem. Commun., (2003) 1296–1297.
99. Zhu, M., Ruijter, E. and Wessjohann, L.A., *A new scavenger resin for the reversible linking and monoprotection of functionalized aromatic aldehydes*, Org. Lett., 6 (2004) 3921–3924.

100. Makita, A., Nihira, T. and Yamada, Y., *Lipase catalyzed synthesis of macrocyclic lactones in organic solvents*, Tetrahedron Lett., 28 (1987) 805–808.
101. Antczak, U., Gora, J., Antczak, T. and Galas, E., *Enzymatic lactonization of 15-dihydroxypentadecanoic acid and 16-hydroxyhexadecanoic acids to macrocyclic lactones*, Enzyme Microb. Technol., 13 (1991) 589–593.
102. Mori, K. and Tomioka, H., *Synthesis of four macrolide pheromones to define the scope and limitation of enzymatic macrolactonization*, Liebigs Ann., (1992) 1011–1018.
103. Lu, W. and Sih, C.J., *Chemoenzymatic synthesis of macrocyclic polyamines*, Tetrahedron Lett., 40 (1999) 4965–4968.