Journal of Computer-Aided Molecular Design, **16:** 431–439, 2002.
KLUWER/ESCOM

Chemoinformatics methods for systematic comparison of molecules from natural and synthetic sources and design of hybrid libraries

Jürgen Bajorath
Albany Molecular Research, Inc., Bothell Research Center (AMRI-BRC), 18804 North Creek Pkwy, Bothell, Washington 98011, and Department of Biological Structure, University of Washington, Seattle, WA 98195

Key words: Computational analysis, molecular descriptors, natural and synthetic molecules, database analysis, similarity searching, library design, compound arrays, hybrid libraries

Summary

Until recently, the field of diversity and library design has more or less ignored natural products as a compound source. This is probably due to at least two reasons. First, combinatorial and reaction-based approaches have been major focal points in the early days of computational library design. In addition, a widespread view is that natural products are often highly complex and not amenable to medicinal chemistry efforts. This contribution introduces recent computational approaches to systematically analyze natural molecules and bridge the gap between natural products and synthetic chemistry programs. Large scale comparisons of natural and synthetic molecules are discussed as well as studies designed to identify 'synthetic mimics' of natural products with specific activity. In addition, a concept for the design of natural/synthetic hybrid libraries is introduced. Although research in this area is still in its early stages, an important lesson to be learned from computational analyses is that there is no need to a priori 'shy away' from natural products as a source for molecular design.

Natural products in drug discovery: an introductory perspective

The majority of efforts discussed in this contribution focus on analyzing naturally occurring molecules and relating them to synthetic compounds, using different methodologies. Thus, initially, it seems appropriate to address the question why one should at all be interested in exploring and exploiting natural product diversity by computational means, for example, in the context of library design.

Background

Natural products from microbial, plant, marine, or even mammalian sources have traditionally been a major drug source and continue to play a significant role in today's drug discovery environments [1–3]. For example, approximately 40% of new drugs introduced between the mid '80s and '90s originated from natural molecules [1]. Furthermore, nine of the top 20 small molecule drugs in 1999 were based on natural products [3]. In fact, in some therapeutic areas, for example, oncology, the majority of currently available drugs are derived from natural products.

Although existing biodiversity is widely recognized as a highly attractive and only partly explored source of pharmaceutically active substances [3], natural products have not always been as popular in drug discovery research as one might expect. This has several reasons including the significant efforts required to efficiently access natural products and isolate active compounds [4] and the difficulties to synthesize many of these molecules de novo, an instructive example being the important anti-cancer drug paclitaxel [5–7]. In fact, in parallel with the rapidly increasing interest in combinatorial chemistry and high-throughput screening (HTS) a number of years ago, focus on natural products as a lead source rapidly diminished in many pharmaceutical settings.

*To whom correspondence should be addressed. E-mail: jurgen.bajorath@albmolecular.com.

Recent trends

However, as experience shows, initial expectations after introduction of new technologies are often too high and, accordingly, many R&D trends are cyclical in nature. Therefore, while the number of newly introduced drugs per year remains more or less constant [8], despite massive application of high-throughput technologies, and while the immediate impact of combinatorial and HTS methodologies on drug discovery is beginning to be questioned [9], interest in natural products-based discovery is on the rise again [3, 10]. For example, as more and more protein-protein interactions are being targeted [11], the view that complex natural molecules are preferred candidates to interfere with such 'complicated' interactions has become quite popular, although solid scientific support for such assumptions is as of yet limited at best. In addition, it is increasingly being recognized that the intrinsic diversity of natural products by far exceeds the degree of molecular diversity that can be created by synthetic means and that the vast majority of biodiversity is yet to be explored [3, 4]. Such views or insights are in part responsible for the again increasing interest in natural products in pharmaceutical research [10].

Natural products chemistry

Natural products extracts, either crude or fractionated (partially purified), have long provided the basis for isolation of novel compounds and elucidation of their structures [12]. Total synthesis of natural molecules of pharmaceutical interest has been, and continues to be, an equally important and challenging area of research in synthetic chemistry [13]. The advent of combinatorial reaction methodologies has also influenced natural products chemistry, and some emphasis is being put on the generation of natural product-focused libraries [13, 14], mostly based on total synthesis concepts [15] but also on what is called a semi-synthetic approach [16]. In the latter case, a natural molecule is directly subjected to derivatization chemistry at selected functional groups. This usually requires that the compound has been scaled up and purified in sufficient quantities. In addition, conceptually distinct biocatalytic approaches are being pursued whereby enzymatic reactions are utilized to modify and diversify metabolites [17]. However, in contrast to analytical and synthetic efforts, computational investigations have not yet played a major role in the natural products arena. In fact, while many studies in computational chemistry and chemoinformatics over the past few years have focused on various concepts for design of combinatorial libraries or analysis of large compound databases [18, 19], natural products and their characteristics have only recently been studied computationally, as discussed in the following.

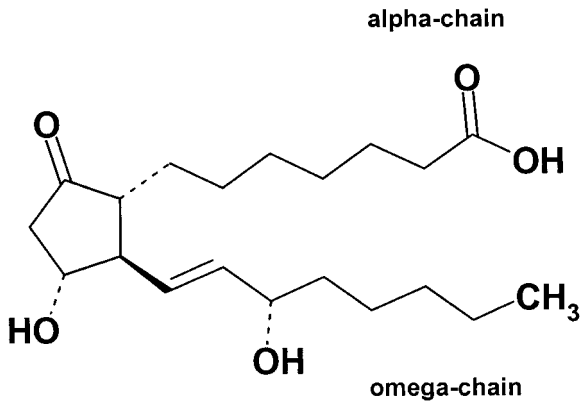

Figure 1. Structure of prostaglandin E1. A small and relatively simple natural molecule that is difficult to synthesize (example adapted from ref. 13).

Computational approaches

Our laboratory became initially interested in the computational analysis of natural products and their comparison with synthetic molecules since we had access to in-house natural product libraries and knowledge of a number of hits obtained by screening of natural products. In several cases, these hits could not be transformed into solid leads because it was very difficult to synthesize the compounds de novo and subject them to medicinal chemistry. Figure 1 shows a representative example of a small natural molecule, prostaglandin E1, a potent mediator of inflammatory and related immune responses. Despite apparent structural simplicity, the prostaglandin family has presented great challenges for total synthesis and has only recently been used as a template for library design [13]. In our experience, examples like this are quite common. Often a natural product with desired activity is difficult to develop not because it is very large or complex, although this certainly occurs, but because it is not amenable to total synthesis, at least via established synthetic routes. From our point of view, these cases opened the door to computational investigations attempting to relate natural and synthetic molecules to each other. However, only a few computational studies have thus far addressed these or related questions.

Statistical analyses

A widely recognized initial contribution to this field has been made by Henkel and colleagues [20] who compared the distributions of molecular building blocks, functional groups, and molecular properties in libraries of natural, synthetic, and drug-like compounds. The study revealed some systematic chemical differences between natural and synthetic molecules. For example, as expected, the molecular weight of natural molecules is generally higher than of synthetics or drugs and they are richer in oxygen-containing groups (alcohols, ethers, esters), whereas synthetic and drug-like molecules have more nitrogen-containing groups (amines, arenes, amides) and halogen atoms. Importantly, Henkel et al. reported that approximately 40% of core structures or scaffolds of natural products are not found in synthetic compounds, which very well correlates with the difficulties to synthesize many natural molecules, regardless of their complexity.

Similar findings have recently been reported by Lee and Schneider [21] who also analyzed statistical distributions of core structures, functional groups, and molecular properties of natural products and drugs and, in addition, studied natural scaffolds and their suitability for library design. Furthermore, 'pharmacophore patterns' in natural products and drugs were compared using self-organizing maps [22]. Lee and Schneider reported that roughly 35% of complex ring systems found in trade drugs also occur in natural products but that only 17% of natural ring structures are also seen in drugs. These ratios reflect not only greater diversity of natural core structures but also confirm that there is little overlap between naturally occurring and synthetic scaffolds. On the other hand, it was found that natural products and drugs share similar pharmacophores, which has some implications for library design, as further discussed below.

Chemoinformatics

Informatics-type investigations to study natural products are also rare. When introducing substructure-based calculation of biological activity profiles to distinguish between active and inactive compounds, Gillett et al. used known active and presumed inactive natural products as one of their test cases, among others [23]. Taking the uncertainty of database activity assignments into account and combining genetic algorithm calculations with cluster analysis, the authors were able to predict active natural molecules approximately seven times better than by random selection.

Our approach to study natural molecules was initially based on the analysis and comparison of molecular descriptor distributions in databases containing natural or synthetic compounds, as part of larger scale descriptor evaluations [24]. Thus, rather than studying structural features and differences directly, we projected the comparison into chemical descriptor space(s). As our knowledge base, we used the Available Chemicals Directory (ACD) [25] ($\sim$ 199,000 entries) as a source of synthetic compounds and the Chapman and Hall compendium of natural products (C&H) [26] ($\sim$ 116,000 entries). We calculated and compared distributions of approximately 100 property descriptors using our histogram-based implementation of Shannon entropy (SE) analysis [27] and, later on, its extension, termed Differential Shannon entropy (DSE) [28]. SE and SE-DSE analysis make it possible to reduce descriptor distributions to their relative information content and identify descriptors that are most responsive to systematic differences between compound databases [27,28]. Table 1 lists examples of property descriptors that are very responsive to systematic differences between synthetic and natural molecules. Combinations of such descriptors were thus considered likely to distinguish between these compound classes.

In a study designed to evaluate this hypothesis, we attempted to systematically distinguish between natural and synthetic molecules [29]. Therefore, we selected descriptors with desired SE characteristics, complemented them with structural keys [30] that are variably set in synthetic and natural molecules, and derived binary QSAR models [31] from preferred descriptor combinations. We then used these models to predict the chemical source of molecules (i.e., 'synthetic' or 'natural') in different test databases. In these calculations, we achieved greater than 80% prediction accuracy when synthetic and natural molecules were randomly assembled. The accuracy was further improved, to greater than 90%, when these molecules belonged to specific biological activity classes (including, for example, synthetic tyrosine kinase inhibitors or naturally occurring protein kinase C inhibitors) [29]. Table 1 also shows the descriptor combination that produced our best-performing binary QSAR model. The results of this investigation provided some support for our descriptor selection methods and catalyzed further studies in our laboratory.

Table 1. Molecular descriptors that are sensitive to systematic differences between synthetic and natural molecules and perform well in binary QSAR analysis.

Descriptor	Definition
a_ICM*	Entropy of element distribution in a molecule
bpol	Sum of the absolute value of the difference between atomic polarizabilities of bonded atoms in a molecule
chi0v_C	Sum of the inverse square roots of the van der Waals surface area of carbon atoms in a molecule
b_double	Number of non-aromatic double bonds
chi1v	Sum of the inverse square roots of the product of the van der Waals surface of all bonded heavy atoms i and j
a_nH*	Number of hydrogen atoms
b_single*	Number of single bonds
b_ar*	Number of aromatic bonds
vsa_hyd	Approximate sum of the van der Waals surface area of hydrophobic atoms
apol	Sum of atomic polarizabilities including implicit hydrogen atoms
99*	Carbon-carbon double bond
127*	Fragment with more than one non-ring oxygen atom bound to a ring
139*	Hydroxyl group

The top ten property descriptors were identified by Shannon entropy analysis. The bottom three descriptors are structural keys that are among those with most different frequency of occurrence in the ACD and C&H databases [29]. Asterisks indicate descriptors forming the combination that achieved overall highest accuracy in distinguishing between synthetic and natural molecules in binary QSAR calculations [29]. For methodological details concerning Shannon entropy or binary QSAR analysis, the reader is referred to references 28 and 31, respectively, and citations therein.

Library design

While combinatorial chemistry makes some use of natural products as templates for library synthesis [13,32], computational design strategies to support such efforts are not yet established. Natural product libraries relying on total synthesis or semi-synthetic approaches, as mentioned above, usually target a selected or a few closely related natural products, for example, libraries derived from prostaglandins [13,33]. In essence, such libraries often represent (combinatorial) analog collections. By contrast, only a few libraries have been reported that more systematically capture, and extend, structural features of natural products [14].

Hybrid libraries

For us, an interesting question has been how to best make use of chemical information encoded in natural products in a library format. A conceptually straightforward but laborious and time-consuming possibility is to build libraries directly from purified natural products. However, another idea is to deliberately combine natural and synthetic components. As discussed above, Schneider and Lee [21] found that ring structures in natural and synthetic molecules differ substantially, whereas pharmacophore patterns are much more similar. This would suggest to combine natural and synthetic building blocks and thereby explore natural product diversity by combinatorial means [3, 21]. However, a major challenge for this approach is to chemically access natural scaffolds, reflecting the bottleneck of total synthesis. Our laboratory has pursued a different approach to hybrid library design that does not, or only little, depend on de novo synthesis of natural products. By contrast, the strategy relies to a large extent on molecular similarity calculations to combine information from natural and synthetic molecules, which is described below.

Design tools: similarity searching

In this context, one of our major interests has been the identification of 'synthetic mimics' of 'difficult' natural products with specific activity. Ideally, mimics and their analogs should be capable of replacing natural product leads in the course of lead optimization projects. As a starting point, we focused on similarity search calculations in compound databases and systematically tested in-house developed 'mini-fingerprints' (MFPs) for their ability to recognize natural or synthetic compounds with similar activity and distinguish them from others. MFPs were originally designed as short binary bit strings encoding combinations of molecular descriptors that performed well in compound partitioning according to biological activity [34]. Thus, the focal point of MFP design has been the recognition of molecules with similar activity. Our MFPs capture combinations of a few selected property descriptors (e.g., accounting for molecular rigidity, aromatic character, or hydrogen bond acceptors) and 20-40 structural keys and consist of only 50–70 bit positions [34]. They are thus much smaller and less complex than other more widely used fingerprint designs [24].

In our hands, MFPs have been quite successful in recognizing remote similarity relationships, probably because they balance the detection of fine chemical details and key molecular features that determine a specific activity. Although MFPs were not specifically designed to recognize natural and synthetic compounds, we have been able, in at least two cases, to replace natural product inhibitors of specific protein-protein interactions with MFP-identified synthetic compounds (these data can not be disclosed at present). In systematic search calculations, a prototypic MFP correctly recognized approximately 35% of molecules having similar activity (with a false positive rate of less than 1%) in a test database consisting of eight synthetic and eight natural compound classes [35]. These initial results were encouraging and we thus applied MFP-based similarity searches also in the context of hybrid library design.

Design tools: similarity/diversity metric

As another component of our library design efforts, we have developed a dual fingerprint-based metric that makes it possible to combine similarity searching and diversity sampling in database mining [36]. The algorithm is summarized in Figure 2. In the initial 'similarity step', compounds are selected that reach a specified threshold value of the Tanimoto coefficient [37] (e.g., 0.75) in MFP search calculations using template molecule(s) of interest. These templates are often known inhibitors or antagonists. In a subsequent 'diversity step', pre-selected molecules are compared to each other and a new compound is only accepted, if it does not exceed a second threshold value (e.g., 0.85) in any pair-wise comparison. For the second step, a different fingerprint (e.g., MACCS keys [30]) is used. The diversity step ensures that compounds are reasonably different from each other and prevents, for example, the selection of series of closely related analogs. The filtering proceeds until a specified number of molecules has passed the subsequent similarity and diversity tests.

Original compound pools from which selections are made may vary significantly. They can consist of various compound databases or randomly sampled scaffold/R-group combinations [38]. In the latter case, molecular scaffolds can be obtained automatically based on hierarchical descriptions of molecules [38] or by making use of reaction information [35, 39]. For the design of compound arrays, we often generate pools of scaffold/R-group combinations using database scaffolds that are modified to represent intermediates of selected chemical reactions. This approach is probably best understood as a combination of product- and reaction based design. The fingerprint-based metric was originally developed for focusing of libraries on specific template molecules but can also be used for fingerprint-based diversity sampling by simply omitting the similarity step. Similarly, compound pools of limited size generated by random sampling of combinations of 'reactive' scaffolds and lists of ACD reagents can be screened by applying the diversity step only. For convenient use, MFPs, routines for similarity searching, and the library design algorithm have been implemented in the Molecular Operating Environment [40] and make use of some of its built-in functions.

Compound arrays

Our approach to the design of hybrid libraries containing natural and synthetic molecules is based on similarity searching and complemented by generation of diverse derivatives. The underlying concept of what we call the 'MetaFocus' (Metabolite Focused) library [35] is illustrated in Figure 3. Distinct compound arrays represent the building blocks of this library, and each array is focused on a natural product of therapeutic interest [35]. In many cases, natural molecules have

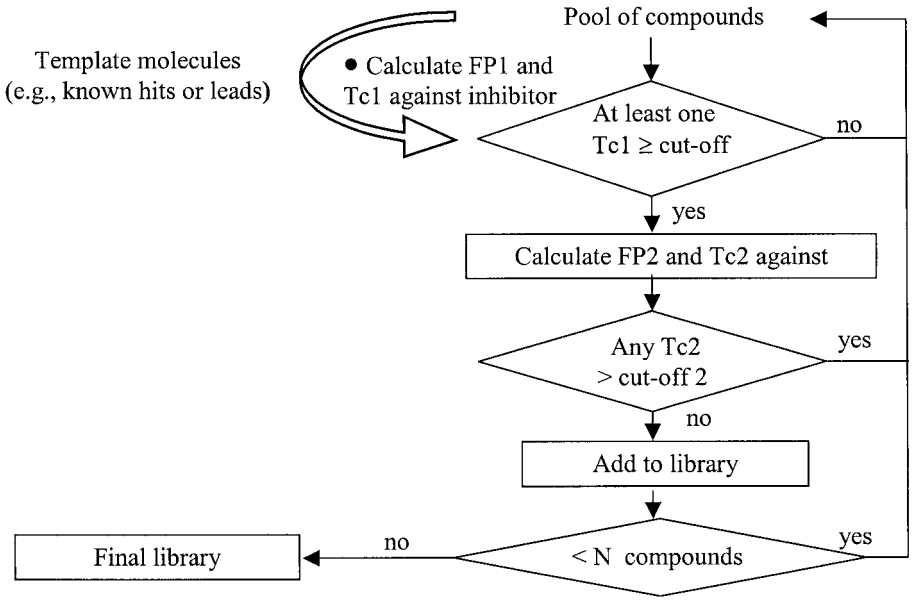

Figure 2. Flow-chart representation of the dual fingerprint-based algorithm for library design. 'FP' means fingerprint and 'Tc' stands for Tanimoto coefficient. Calculation of FP1 (for compounds of the pool and comparison with templates) represents the 'similarity step' and calculation of FP2 (for pair-wise comparisons of molecules passing the first test) the 'diversity step'.

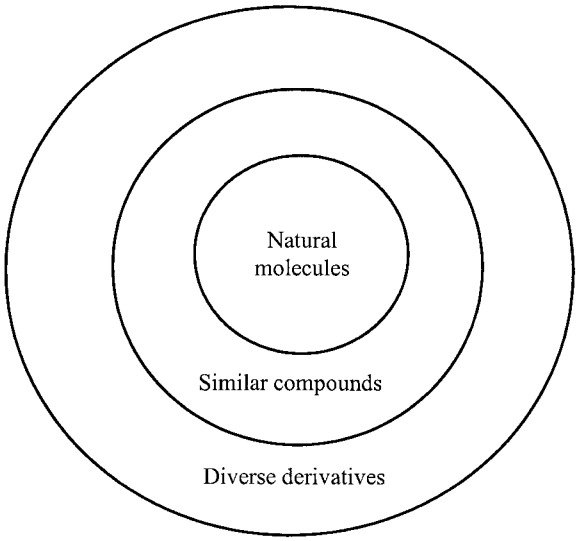

Figure 3. Illustration of the 'MetaFocus' concept. Each array is based on a natural molecule with attractive biological activity. Similar molecules from natural or synthetic sources, identified by database searching, represent the first layer of expansion. The collection is further diversified by semi-synthetic or combinatorial derivatives of natural and similar molecules. Generation of the second layer goes beyond database mining requires chemistry efforts.

known therapeutic effects (e.g., anti-bacterial activity) but their mechanism of action is little explored. Building arrays around compounds with interesting therapeutic properties, be their molecular targets known or not, aims at increasing the probability of finding other candidate molecules with enhanced, modulated, or easier to improve activity. Following selection of a specific natural product, the initial step is to search for similar natural and synthetic molecules in databases. Identified candidates are then subjected to a 'diversity expansion' that involves scaffold-based design of compounds. If a selected natural product is available in sufficient quantities for chemistry, the array is further 'filled' with semi-synthetic derivatives.

As an example, a Metafocus array built around anisomycin is discussed. Anisomycin, shown in Figure 4a, is a small metabolite from *Streptomyces griseolus* that inhibits protein synthesis [41] and acts as an activator of MAP kinase [42]. Figure 4a also shows two similar synthetic molecules identified by MFP searching and four scaffolds that were directly derived from anisomycin. Figure 4b shows examples of the different components of the array. These include similar synthetic and natural molecules, diverse compounds designed from them, and semi-synthetic derivatives designed based on scaffolds that incorporate reaction information. Figure 4c illustrates how the different parts of the anisomycin array have been populated during the design stage, containing a total of 8,033 compounds and 52 scaffolds. Similarity searching, using a Tanimoto coefficient threshold value of

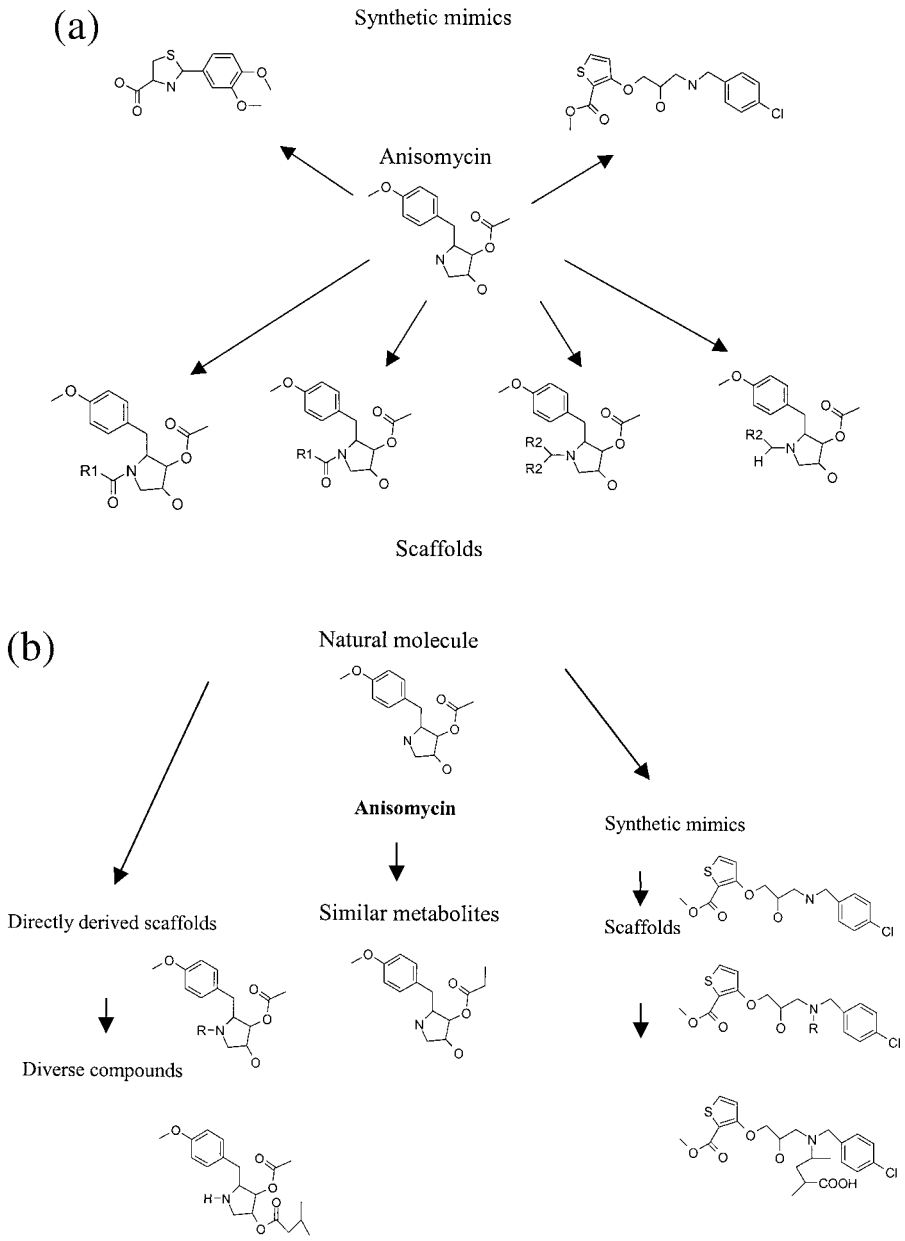

Figure 4. Anisomycin array. In (**a**), the structure of anisomycin is shown in the center. At the top, two synthetic molecules are depicted that were identified by MFP similarity searching using a Tc cut-off value of 0.80. At the bottom, four molecular scaffolds are shown that were directly derived from anisomycin (for semi-synthetic chemistry). Each scaffold represents an intermediate of a different chemical reaction at the amine. For compound design based on these scaffolds, R1 substitutions were computed using R-groups with only carbon atoms as substitution points, whereas R2 substitutions were designed using different R-group libraries with either carbon, nitrogen, or oxygen atoms as substitution points. In (**b**), major components of the anisomycin array are shown, consistent with the MetaFocus concept. These include semi-synthetic scaffolds, similar compounds, scaffolds derived from these molecules, and compounds computed on the basis of these scaffolds. Molecular scaffolds and compounds obtained from similar metabolites are omitted from this representation to emphasize that synthesis of these components is often problematic, mostly due to the difficulties involved in obtaining many of these metabolites. Figure (**c**) illustrates how the anisomycin array is populated with designed compounds. Here '23 from 2' means that a total of 23 scaffolds were designed from two of eleven predicted synthetic mimics.

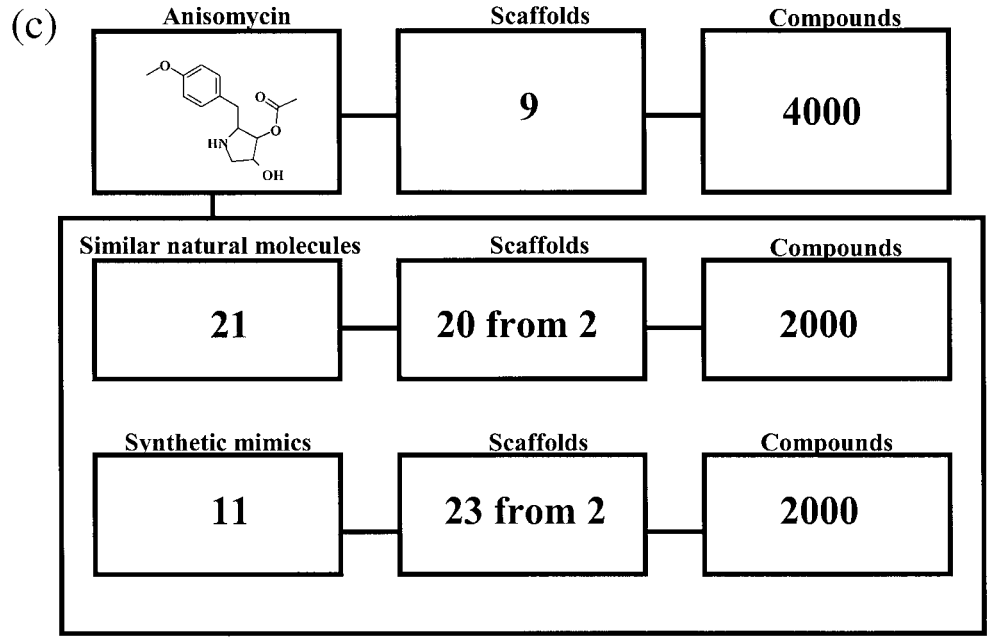

Figure 4. Continued.

0.8, identified 21 similar metabolites and 11 potential synthetic mimics. For initial diversification, scaffolds were designed from two natural and two synthetic molecules only. Based on an evaluation of this design by our combinatorial chemistry group, the major difficulty was to obtain or synthesize natural products similar to anisomycin. By contrast, scaffold-based transformations usually involved only one- or two-step reactions and were thus relatively straightforward. Anisomycin provides a representative example of a number of hybrid arrays designed thus far.

Summary and outlook

We have begun to develop and apply chemoinformatics-type methods to systematically analyze natural molecules and compare them to synthetic compounds. To a significant extent, these efforts rely on computational analysis of molecular properties in large natural and synthetic compound databases. The computational study of natural products is still a rather underdeveloped area. However, a few noteworthy contributions have recently been made in the literature and provide a basis for further investigations.

The size and complexity of many natural products renders them difficult, if not impossible, to study by synthetic and also computational means. However, we have learned that, in many cases, active natural products are small and not more complex than many synthetic compounds. Our design efforts were initially spurred on by recurrent problems associated with de novo chemical synthesis of active natural molecules, regardless of their size. In these cases, searching for similar synthetic molecules becomes particularly attractive. In our experience, chemoinformatics approaches in this area benefit from the situation that differences between natural and synthetic molecules are often much less distinct in carefully chosen theoretical 'descriptor spaces' than at the synthetic level.

Library design strategies such as the MetaFocus approach are quite new and still at a conceptual stage. However, MetaFocus arrays are indexed and represent an easily searchable and expandable data structure that is useful even if only parts of the designed arrays are finally translated into actual compound collections. However, since synthesis of natural products-based libraries has become more popular, it is anticipated that research and development in this area will benefit from more intense computational support in the near future.

There is little doubt that natural products will continue to play an important role in drug discovery, despite periodic changes in views about preferred compound sources and discovery strategies. However, although biodiversity continues to be recognized as a largely unexplored source of novel drug candidates,

probably more so than ever, it would be premature to predict that interest in natural molecules will steadily increase. In this regard, much will depend on medium term success of combinatorial and high-throughput technologies in increasing the number of novel drugs. However, helping to bridge synthetic and natural products programs and perhaps remove some of the traditional barriers between medicinal and natural products chemistry should be an exciting new challenge for scientists in chemoinformatics and drug design.

References

1. Cragg, G.M., Newman, D.J., Snader, K.M., J. Nat. Prod., 60 (1997) 52–60.
2. Grabley, S. and Thiericke, R., Adv. Biochem. Eng. Biotechnol., 64 (1999) 101–154.
3. Harvey, A., Drug Discov. Today, 5 (2000) 294–300.
4. Strohl, W.R., Drug Discov. Today, 5 (2000), 39–41.
5. Wessjohann, L., Angew. Chem. Intl. Ed., 106 (1994), 1011–1013.
6. Nicolaou, K.C., Dai, W.-M., Guy, R.K., Angew. Chem. Intl. Ed., 106 (1994), 38–69.
7. Cragg, G.M., Med. Res. Rev., 18 (1998), 315–331.
8. Drews, J., Science, 287 (2000), 1960–1964.
9. Lahana, R., Drug Discov. Today, 4 (1999), 447–448.
10. Lawrence, R.N., Drug Discov. Today, 4 (1999), 449–451.
11. Boulton, S.J., Vincent, S. and Vidal, M., Curr. Opin. Chem. Biol., 5 (2001), 57-62.
12. Cragg, G.M. and Newman, D., Chem. Br., 37 (2001), 22–26.
13. Wessjohann, L.A., Curr. Opin. Chem. Biol., 4 (2000), 303–309.
14. Tan, D.S., Foley, M.A., Shair, M.D. and Schreiber, S.L., J. Am. Chem. Soc., 120 (1998), 8565–8566.
15. Nicolaou, K.C., Winssinger, N., Pastor, J., Ninkovic, S., Sarabia, F., He, Y., Vourloumis, D., Yang, Z., Li, T., Giannakakou, P. and Hamel, E., Nature, 387 (1997), 268–272.
16. Höfle, G. and Sefkow, M., Heterocycles, 48 (1998), 2485–2488.
17. Turner, N. and Schneider, M., Curr. Opin. Chem. Biol., 4 (2000), 65–67.
18. Mason, J.S. and Hermsmeier, M.A., Diversity assessment, Curr. Opin. Chem. Biol., 3 (1999), 342–349.
19. Martin, Y.C., J. Comb. Chem., 3 (2001), 231–250.
20. Henkel, T., Brunne, R. M., Müller, H. and Reichel, F., Angew. Chemie. Intl. Ed., 38 (1999), 643–647.
21. Lee, M.-L. and Schneider, G., J. Comb. Chem., 3 (2001), 284–289.
22. Kohonen, T., Biol. Cybern., 43 (1982), 59–69.
23. Gillet, V.J., Willett, P. and Bradshaw, J., J. Chem. Inf. Comput. Sci., 38 (1998), 165–179.
24. Bajorath, J., J. Chem. Inf. Comput. Sci., 41 (2001), 233–245.
25. ACD (Available Chemicals Directory), MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577.
26. Chapman & Hall, Dictionary of Natural Products, CD-ROM 1999 version, CRC Press LLC, NW Corporate Blvd, Boca Raton, FL 33431, USA.
27. Godden, J.W. and Bajorath, J., J. Mol. Graph. Mod., 18 (2000), 73–76.
28. Godden, J.W. and Bajorath, J., J. Chem. Inf. Comput. Sci., 41 (2001), 1060–1066.
29. Stahura, F.L., Godden, J.W., Xue, L. and Bajorath, J., J. Chem. Inf. Comput. Sci., 40 (2000), 1245–1252.
30. MACCS keys, MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577.
31. Labute, P., Pac. Symp. Biocomput., 7 (1999), 444–455.
32. Weber, L., Curr. Opin. Chem. Biol. 4 (2000), 295–302.
33. Dragoli, D.R., Thompson, L.A., O'Brien, J., and Ellman, J.A., J. Comb. Chem., 1 (1999), 534–539.
34. Xue, L., Godden, J.W. and Bajorath, J., J. Chem. Inf. Comput. Sci., 40 (2000), 1227–1234.
35. Stahura, F.L., Xue, L., Godden, J.W. and Bajorath, J. J. Mol. Model., 6 (2000), 550–562.
36. Xue, L., Godden, J.W., Stahura, F.L. and Bajorath, J., J. Mol. Model., 7 (2001), 125–131.
37. Willett, P., Barnard, J.M. and Downs, G.M., J. Chem. Inf. Comput. Sci., 38 (1998), 983–996.
38. Xue, L. and Bajorath, J., J. Mol. Model., 5 (1999), 97-102.
39. Lewell, X.Q., Judd, D.B., Watson, S.P. and Hann, M.M., J. Chem. Inf. Comput. Sci., 38 (1998), 511–522.
40. MOE (Molecular Operating Environment), Chemical Computing Group Inc., 1255 University Street, Montreal, Quebec, Canada, H3B 3X3.
41. Martinez, J.L. Jr., Jensen, R.A. and McGaugh, J.L., Prog. Neurobiol., 16 (1981), 155–186.
42. Kato, K., Ito, H., Iwamoto, I., Lida, K. and Inaguma, Y., Cell Stress Chaperones, 6 (2001), 16–20.