Minireview

QSAR & Combinatorial Science QCS

Charting Biological and Chemical Space: PSSC and SCONP as Guiding Principles for the Development of Compound Collections Based on Natural Product Scaffolds

Lars Arve[a,b], Tobias Voigt[a,b] and Herbert Waldmann[a,b]*

[a] Department of Chemical Biology, Max Planck Institute of Molecular Physiology, Otto-Hahn-Straße 11, 44227 Dortmund, Germany
[b] Fachbereich 3, Chemische Biologie, Universität Dortmund, Otto-Hahn-Straße 6, 44227 Dortmund, Germany
E-mail: herbert.waldmann@mpi-dortmund.mpg.de

Keywords: Chemical biology – Combinatorial chemistry – Cheminformatics – Library design – Natural products

Received: December 15, 2005; Accepted: January 10, 2005

DOI: 10.1002/qsar.200540213

Abstract

Finding small molecules that selectively modulate protein function remains one of the major challenges in the fields of chemogenomics and drug development. Natural products interact with proteins in their biosynthesis and when displaying biological activity, and can therefore be considered as "privileged structures". Thus, natural products may serve as biologically validated starting points for the design of focused compound collections. In recent years, effective synthesis methods for natural-product-based focused compound collections have been developed, including multi-component reactions. This review discusses recent advances in the synthesis of natural-product-inspired compound collections as well as the application of Protein Structure Similarity Clustering (PSSC) and a Structural Classification Of Natural Products (SCONP) as guiding principles for the design of focused compound collections based on natural products.

1 Introduction

Recent knowledge of protein networks that regulate biological processes increased rapidly due to modern approaches addressing DNA sequence (genomics), protein structure (structural biology) as well as protein expression and protein interactions (proteomics). As a result, new proteins are being discovered whereas the functions of these proteins often remain to be elucidated. One approach to unravel protein functions is chemogenomics where small molecules are used to alter the functions of either gene or protein families [1]. Chemogenomics can be regarded as part of the drug discovery process since the discovery of small molecules that modulate gene or protein function might ultimately lead to the identification of new drug candidates.

The most challenging task in chemogenomics and drug discovery is to find those small-molecule modulators that specifically modulate the gene or protein of interest. Combinatorial chemistry has emerged in the last decade as a powerful tool to address this problem by generation of large compound libraries, which are subsequently evaluated for bioactivity [2]. However, initial expectations that large compound libraries would result in the discovery of many new hit and lead structures were not met. These efforts made clear that the design principles of combinatorial libraries need to be relevant in a biological context and that quality not quantity of the library members determines the hit rates. In this process diversity [3], drug-likeness [4] and biological relevance [5–7] have emerged as the most relevant parameters in the choice of library scaffolds from the almost infinite number of thinkable small-molecule modulators of protein function. In this context, the development of natural-product-guided compound collections deserves particular attention.

2 Natural Products as Biologically Pre-validated Modulators of Protein Function

Since natural products interact with more than one protein during their biosynthesis and show multiple biological activities (toxins, neurotransmitters, *etc.*) mediated by interactions with different proteins, they fulfill the requirement for biological relevance. Thus, these natural products can be considered as embodying privileged structures that confer the ability to bind to various proteinaceous receptors [8]. Those privileged structures might have been evolutionarily selected by nature; this is supported by the observation that natural-product-based compound collections of-

ten show significantly higher hit rates in biochemical and biological screens than compound libraries obtained exclusively on the basis of chemical feasibility [6]. Notably, a major fraction of the new chemical entities introduced into the market in the last twenty years are natural products or derived therefrom. Moreover, nearly 60% of the anticancer drugs and 75% of the infectious disease drugs are natural products in origin [9].

Thus, libraries and compound collections, whose design is inspired by natural products, are expected to yield biologically pre-validated modulators of protein function in relatively high frequency at considerably reduced library size. This natural-product-structure-inspired strategy does not neglect the issue of chemical diversity; rather it builds on the diversity created by nature itself. Hence, this approach offers an opportunity to identify "privileged structures" from nature and explore their possible applications in chemical biology and drug discovery.

3 Solid-Phase Synthesis of Natural-Product-Derived Compound Libraries

Solid-phase synthesis is exceedingly useful in compound library synthesis [10]. Generation of libraries of natural-product analogs using solid-phase chemistry has received increasing attention in recent years and the synthesis of new libraries is in steady progress. In designing a multi-step synthesis of a natural-product-based library on a solid support, versatile and high-yielding reactions that employ a wide range of reactants with different electronic and steric properties are most desirable. Alternatively, solution-phase combinatorial synthesis allows for a rapid transfer of established synthesis methods whereas method development is mandatory in almost all cases, if solid-phase synthesis is applied. However, solution-phase synthesis poses the difficulty of isolating and purifying the library members rendering solid-phase synthesis the perfect tool for multi-step syntheses, which are typical for the preparation of natural-product-based libraries. Methodologies for solution-phase synthesis of compound libraries have been recently reviewed by Boger et al. [11]. In an intermediate strategy natural-product cores are synthesized in solution phase and attached to a solid support to proceed further with the synthetic sequences to introduce diversity. Hall et al. have reviewed a number of such libraries [12].

Multi-component reactions have become another powerful tool to construct complex molecular frameworks in an atom-economic fashion and very short synthetic sequences [13]. This concept was also applied in natural-product synthesis as well as the solid-phase synthesis of compound collections [14]. Recently, de Meijere and co-workers showed the construction of a collection of compounds with spiro[2.5]octene scaffold by a four-component reaction involving Heck coupling and Diels–Alder cycloaddition in solution [15]. A collection of benzodiazepines was synthesized by Hulme et al. applying an Ugi reaction to solid-support bound isonitriles [16].

Another approach is the application of polymer-supported reagents and scavengers. For example, Ley and co-workers have extensively used polymer-supported reagents and scavengers in the total synthesis of epothilone C and other natural products [17]. A variety of libraries derived from natural products such as carbohydrates, steroids, fatty acid derivatives, polyketides, peptides, terpenoids, flavonoids and alkaloids has been designed and synthesized in recent years (for representative examples, see Figure 1) [18, 19]. Strategies for the exploitation of natural products in compound library synthesis on solid supports have been reviewed recently [19, 20]. Therefore, in this article we focus on the most recent developments in the area.

4 Natural-Product-Inspired Compound Collections Based on a 6,6′-Spiroacetal Framework

6,6′-Spiroacetals are structural motifs found in many natural products such as spongistatins, integramycin, tautomycin and okadaic acid. These natural products display a variety of biological activities such as microtubuli polymerization, HIV-1 integrase and serine-threonine phosphatase inhibition. Interestingly, the basic spiroacetal structure may retain the biological activity of some of the parent natural products. Thus, these motifs may be considered as privileged structures, which constitute a biologically validated starting point for compound library development. Recently, asymmetric synthesis of 6,6′-spiroacetals has been carried out on solid support (Scheme 1) [21]. The synthesis essentially involves up to twelve linear steps including two asymmetric boron-mediated aldol reactions, which proceed with high enantio- and diastereoselectivity. The salient feature of the synthesis is the use of polymer-bound and soluble chiral boron enolates for the asymmetric induction. The synthesis is also attractive in the sense that the cleavage from the polymeric support (Merrifield resin with Wang linker) is accompanied by PMB (*p*-methoxybenzyl) deprotection to result in the formation of the spiroacetal in an overall yield comparable with that of the corresponding solution-phase synthesis. The developed synthetic sequence was employed successfully for the generation of a small compound collection.

A different route to a 6,6′-spiroacetal-based compound collection is shown in Scheme 2 [22]. In this case, the spiroacetal precursor is built up by the addition of different terminal alkynes to polymer-bound aldehydes. By selecting an acid-labile linker – such as the Wang linker-release from the solid support, removal of the protecting groups and acetal formation *via* conjugate addition could be accomplished in a single step. After cleavage from the solid support, the obtained spiroacetal ketones were reduced by a polymer-bound borohydride reagent to yield a collection of 146 spiroacetals.

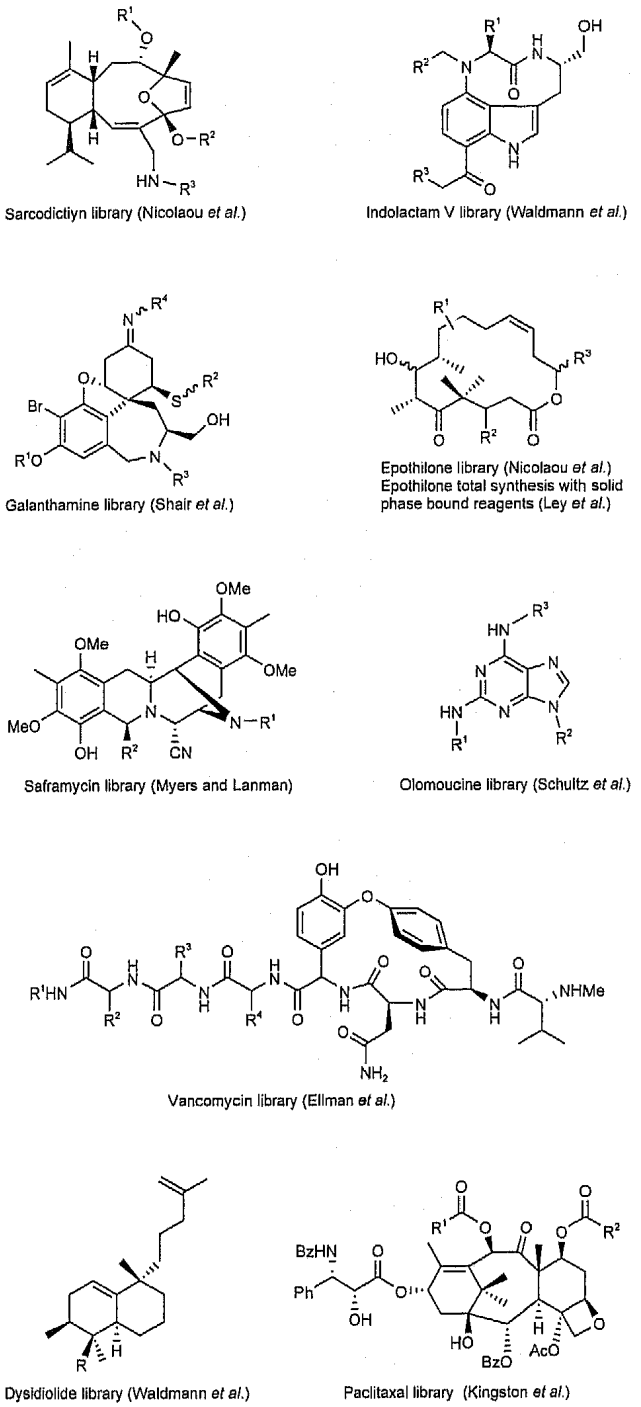
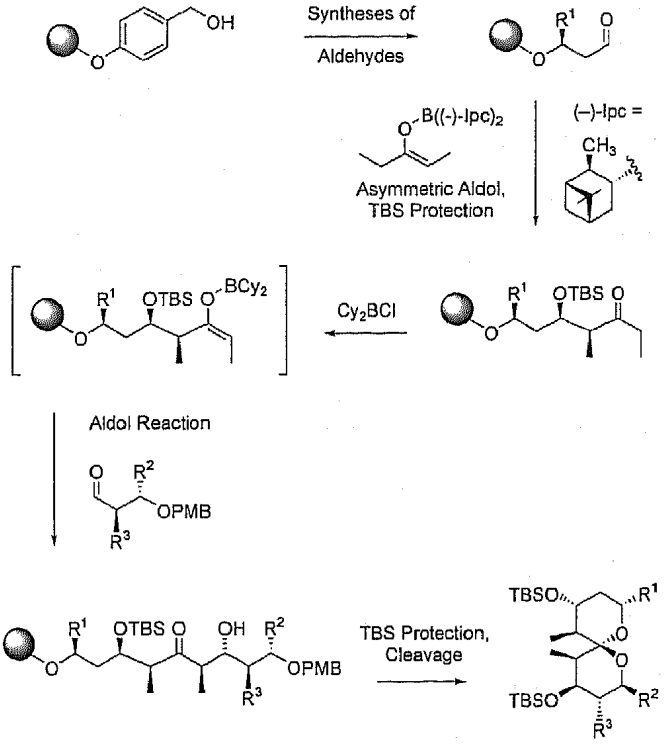
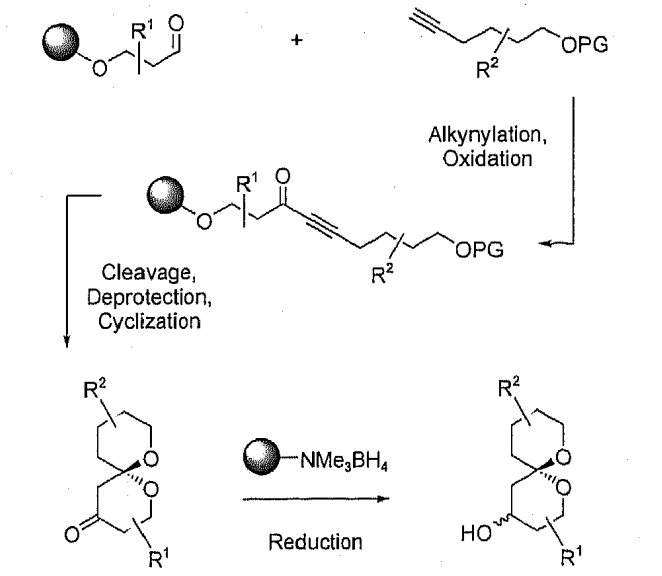


Figure 1. Representative examples of natural-product-derived compound libraries.



Scheme 1. Asymmetric solid-phase synthesis of a 6,6′-spiroacetal compound collection.



Scheme 2. Different synthetic approach to a collection of 6,6′-spiroacetals.

5 Protein Structure Similarity Clustering (PSSC)

The crucial question in library synthesis is where to find biologically validated starting points in vast structural space. Nature itself might provide a convincing answer to this question, namely the use of natural-product scaffolds as outlined earlier. Another answer lies in the conserved structures of proteins. Thus, ligand-sensing cores of proteins with similar three-dimensional structures can be grouped into protein structure similarity clusters. The structural similarity among the members of these clusters can be employed to guide compound library design aiming at inhibitors for all members of the cluster [7].

Usually, only a small part of a domain, the binding site, is directly involved in ligand binding. Therefore, it can be

assumed that structural similarity in this subset is far more decisive for the binding event than fold-type similarity. For this reason, in PSSC spherical cutouts of the protein structure centered on the ligand-binding site are used, which are called "ligand-sensing cores". These are then subjected to structural alignments to yield the protein similarity clusters (Figure 2A). The clusters holding proteins with a high structural similarity and a low sequence similarity are particularly relevant cases for PSSC, because they represent more distant relationships, which classical approaches might not consider.

In library design, the first step is the search for known ligands of the cluster members. Subsequently, one or more of those which are biologically prevalidated, *e.g.*, natural products, are chosen. Then, a compound collection around the chosen scaffold is synthesized.

In a first example, it was proven that this approach yields clusters stretching over different fold types and functional classes [23]. Moreover, it was also demonstrated that the synthesized compound collection was significantly enriched in bioactivity and yielded several potent ligands for each protein in the identified cluster.

6 Structural Classification of Natural Products (SCONP)

Nature's structural conservatism in proteins inspires the idea of a restricted and conserved structural space in natural products. An organizing principle for natural-product scaffolds combined with biological data would chart the regions of chemical space used by nature. The resulting map of this structural space would provide guidance for the development of natural-product-derived libraries. Thus, it would be most helpful in designing new modulators of protein function, especially if this classification could be used in combination with the PSSC.

As a basis for the analysis of natural-product space [24] the *Dictionary of Natural Products* (DNP) [25] was used since it represents the largest available database of natural products. The database was subjected to a procedure whose sequential steps are outlined in Figure 2B. In a first step, faulty records and small molecules associated with the natural products (*e.g.*, water, salts) were removed. Subsequently, the resulting dataset was subjected to an *in silico* deglycosylation since the influence of glycosidic residues on the biological function of natural products is still matter of debate. In a last step, only scaffolds that contained rings were selected for most pharmacological important compounds also incorporate ring-based scaffolds for rigidity.

For categorization of these scaffolds, a structure-based approach was chosen since chemical structure is the foundation of chemical thinking and very instructive to chemists. The guiding idea is that every multi-ring structure can be reduced to a simpler one with a smaller ring system. This way each "parent scaffold" represents a substructure of a more complex "child scaffold". For the assignment of only one "parent scaffold" to each child, a set of rules was applied [24]. These rules were designed in such a way that

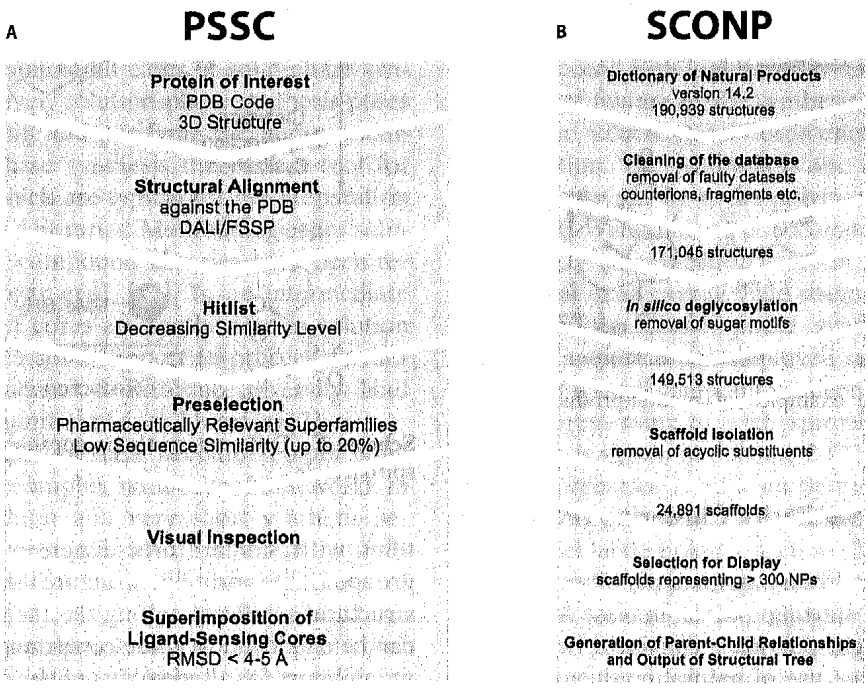


Figure 2. Flowchart of structural analysis approaches. A) Protein Structure Similarity Clustering (PSSC); B) Structural Classification of Natural Products (SCONP).

the result reflects the rational of synthetic organic and medicinal chemists.

The results are displayed in a tree diagram (see Figure 3), which clearly reflects the relationships between individual scaffolds as well as those between scaffold classes. The diagram also introduces a structural hierarchy generated by the growing size of the scaffolds. The resulting picture provides a scheme which can accommodate every conceivable natural product as long as it contains a ring-based scaffold.

The quantitative analysis of the natural-product scaffold tree shows that compounds with two to four rings make up more than half of all naturals products. The calculated volume for this part of the natural products lies between 100 and 500 Å^3 with a maximum at 250 Å^3. Thus, these molecules fit perfectly into the cavities of most proteins whose binding pockets were estimated to have a volume of 300 to 800 Å^3, regarding the fact that ligands often do not fill out the whole cavity [26].

7 Application of PSSC and SCONP: Discovery of a New Class of 11βHSD Inhibitors

The viability of PSSC and SCONP was investigated by the development of 11β-hydroxysteroid dehydrogenase 1 (11βHSD1) inhibitors. This enzyme converts inactive cortisone into active cortisol, is a promising target for the development of new drugs [27, 28], and isoenzyme-selective 11βHSD1 inhibitors are very actively being sought. 11βHSD1 belongs to the same protein structure similarity cluster as protein phosphatase Cdc25A and acetylcholine esterase (AChE, Figure 4A). Thus, according to the reasoning for PSSC, inhibitors of either enzyme should also

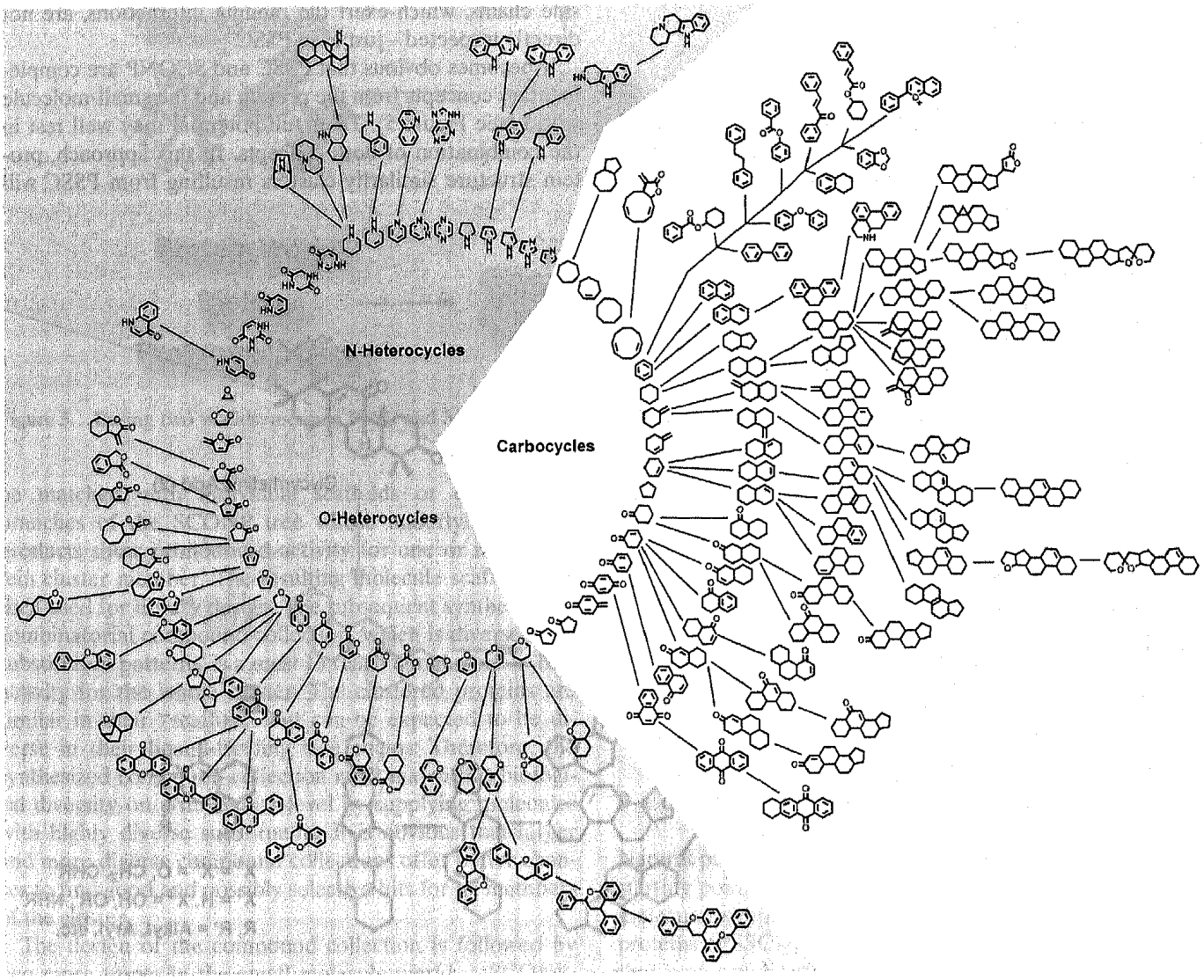


Figure 3. Tree-like graphical representation of natural-product scaffolds. For clarity, only scaffolds that cumulatively represent at least 0.2% of the natural products are shown.

display activity on 11βHSD1. On the other hand, glycyrrhetinic acid (GA, **2**, Figure 4A) is a naturally occurring ligand of 11βHSD1. Analysis of the pentacyclic core of GA, following the rules for SCONP, leads to an octahydronaphtalene scaffold, which is also present in the natural product dysidiolide (**1**), a potent inhibitor of Cdc25A [29], Thus, a compound collection with an octahydronaphtalene core should yield hits on both the enzymes 11βHSD1 and Cdc25A. Such a compound collection was synthesized on solid support yielding 162 compounds. These compounds were evaluated *in vitro* for inhibition of 11βHSD1 and 2 [24, 27]. 30 compounds inhibited 11βHSD1 with IC_{50} values in the low micromolar range and four 11βHSD1 were nanomolar inhibitors. Three compounds inhibited 11βHSD2 in the micromolar range. Most remarkably, even at this comparably small size of compound collection, the hits indicated a pronounced degree of selectivity for the isoenzymes 11βHSD1 and 2: 28 compounds were selective inhibitors of 11βHSD1.

8 Combining PSSC and SCONP

It could successfully be shown that the application of either PSSC or SCONP alone, leads to compound collections enriched in biological activity. Thus, each concept has proven its value although this could possibly be increased by the combination of PSSC and SCONP. Both methods have been designed with a similar underlying concept in mind: the idea that similar scaffolds exhibit similar – not identical! – binding properties while minor deviations are due to the ligand-binding residues involved.

In PSSC, the peptidic backbone of a ligand sensing core forms the scaffold, which is similar among all members of one cluster. The side chains of the individual residues, however, can differ a lot as documented by the low sequence similarities. The side chains bind the substrate while the rigid backbone, the scaffold, is responsible for their spatial arrangement. The comparison focuses on the backbone structure and the location of the actual binding side chains but does not take into account their chemical properties.

The molecules considered in SCONP also consist of scaffolds and their substituents. While the scaffolds are analyzed and grouped in the SCONP tree-like diagram, the side chains, which exert the binding interactions, are not directly inspected – just as in PSSC.

It becomes obvious that PSSC and SCONP are complementary concepts from the protein and the small-molecule world (see Figure 5). Their full potential may well rest in the combination of both concepts. In this approach, protein structure similarity clusters resulting from PSSC will

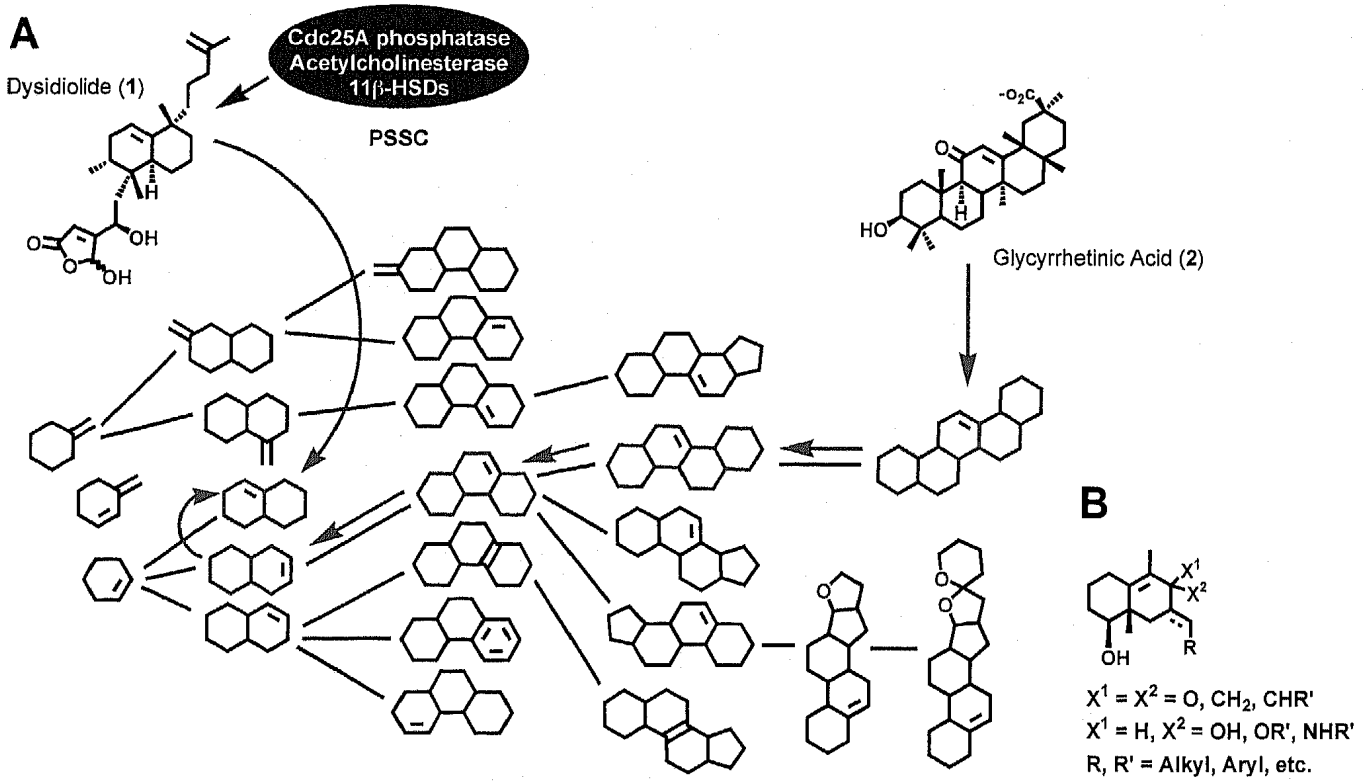


Figure 4. Strategic use of PSSC and SCONP. A) The scaffold of glycyrrhetinic acid (**2**) is analyzed according to SCONP rules leading to an octahydronaphtalene scaffold, which is a substructure of dysidiolide (**1**). B) General structure of the octahydronaphtalene collection synthesized.

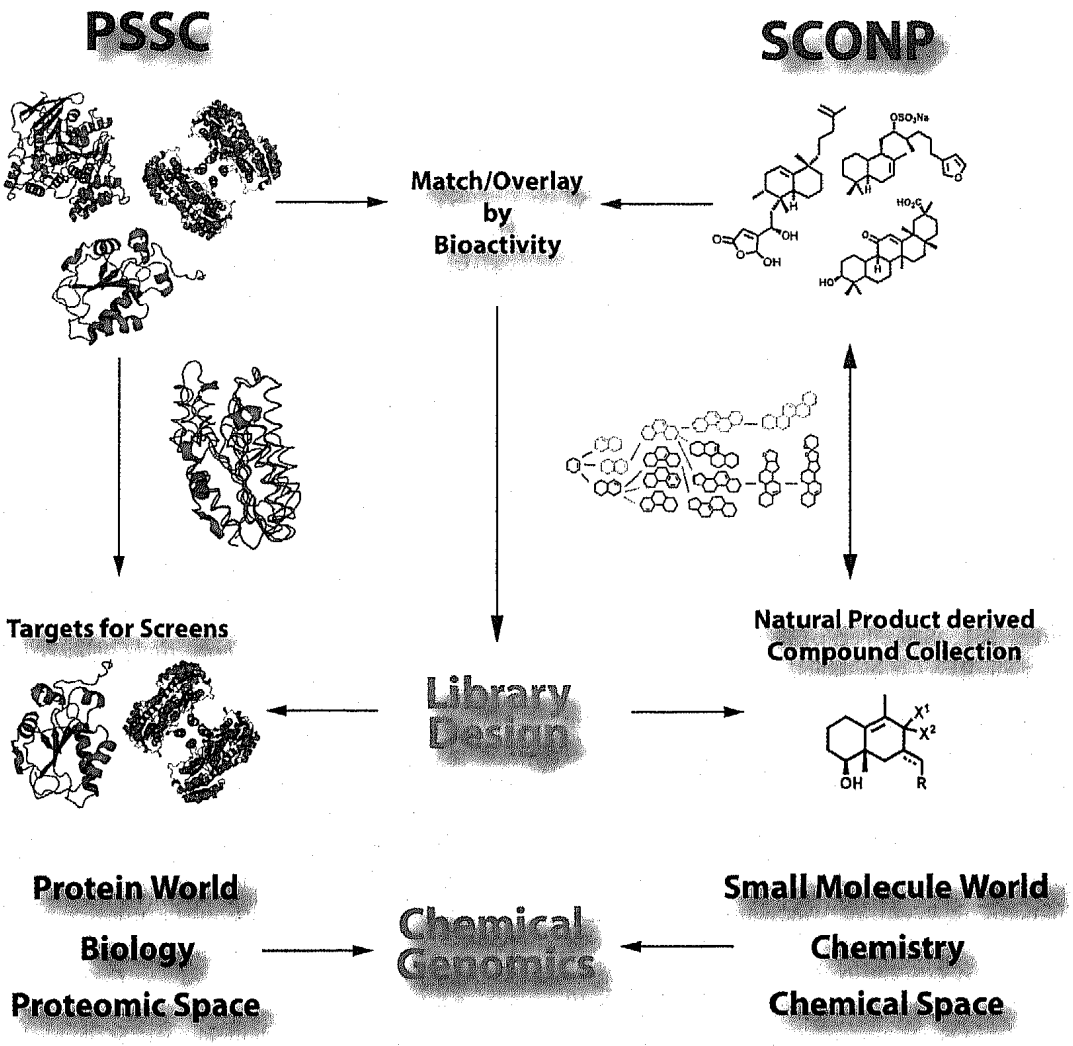


Figure 5. Fusing two worlds: merging PSSC and SCONP to a powerful tool for library design.

be matched with individual scaffolds or even whole branches of the SCONP tree whose underlying natural products show pronounced activity for one or more protein cluster member. The resulting molecule scaffolds are then used for library design. The subsequent synthesis of a combinatorial compound collection, which is diverse in its substitution patterns, is a must to reach enrichment in bioactivity for the whole cluster. The clustered proteins are similar in their "scaffold" but can be expected to be diverse in their ligand-binding side chains. Therefore, the synthesized compound collection must match this biological diversity on a molecular level by supplying molecules with highly diverse substituents. It is obvious that larger and more diverse compound collections offer better chances to find good and possibly selective hits for all members of the cluster.

The design of the compound collection is followed by two more steps: In the small-molecule world, a SCONP-guided compound collection has to be synthesized. In the protein world, the clustered proteins pose as molecular targets against which the compound collection will be screened. The results of the biological evaluation then feed back into SCONP and broaden the database on which it works.

Fusing PSSC and SCONP means matching proteomic space with biologically pre-validated chemical space. This approach could prove to be a most valuable tool in chemical genomics and the development of small-molecule modulators of protein function.

9 Conclusions

Natural products can be used as biologically pre-validated starting points in combinatorial library design. Employing molecular similarity in the design process as described for proteins (PSSC) and small molecules (SCONP) can further enhance hit rates of biological screens. Even more valuable could be the combined approach described in which protein space and chemical space are matched to

generate focused natural-product-based compound collections enriched in bioactivity.

Acknowledgement

The authors would like to thank Stefan Wetzel and Dr. Frank Dekker for providing generous assistance and suggestions during the preparation of this article. This work was financially supported by the Max-Planck-Gesellschaft. L. A. and T. V. are indebted to the Fonds der Chemischen Industrie for a Kekulé Scholarship. T. V. is a fellow of the Studienstiftung des Deutschen Volkes.

References

[1] a) B. R. Stockwell, *Trends Biotechnol.* **2000**, *18*, 449–455; b) B. R. Stockwell, *Nature* **2004**, *432*, 846–854.
[2] R. E. Dolle, *J. Comb. Chem.* **2004**, *6*, 623–679.
[3] a) M. D. Burke, S. L. Schreiber, *Angew. Chem. Int. Ed.* **2004**, *43*, 46–58; b) S. L. Schreiber, *Science* **2000**, *287*, 1964–1969.
[4] a) W. P. Walters, M. A. Murcko, *Adv. Drug. Deliv. Rev.* **2002**, *54*, 255–271; b) J. R. Proudfoot, *Bioorg. & Med. Chem. Lett.* **2002**, *12*, 1647–1650; c) D. E. Clark, S. D. Pickett, *Drug. Discov. Today* **2000**, *5*, 49–58; d) W. P. Walters, A. A. Murcko, M. A. Murcko, *Curr. Opin. Chem. Biol.* **1999**, *3*, 384–387; e) A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, *J. Comb. Chem.* **1999**, *1*, 55–68.
[5] a) R. Breinbauer, I. R. Vetter, H. Waldmann, *Angew. Chem. Int. Ed.* **2002**, *41*, 2878–2890; b) F. J. Dekker, M. A. Koch, H. Waldmann, *Curr. Opin. Chem. Biol.* **2005**, *9*, 232–239.
[6] M. A. Koch, H. Waldmann, *Drug. Discov. Today* **2005**, *10*, 471–483.
[7] M. A. Koch, R. Breinbauer, H. Waldmann, *Biol. Chem.* **2003**, *384*, 1265–1272.
[8] B. E. Evans, K. E. Rittle, M. G. Bock, R. M. DiPardo, R. M. Freidinger, W. L. Whitter, G. F. Lundell, D. F. Veber, P. S. Anderson, R. S. Chang, *J. Med. Chem.* **1988**, *31*, 2235–2246.
[9] a) D. J. Newman, G. M. Cragg, K. M. Snader, *Nat. Prod. Rep.* **2000**, *17*, 215–234; b) D. J. Newman, G. M. Cragg, K. M. Snader, *J. Nat. Prod.* **2003**, *66*, 1022–1037.
[10] a) R. E. Sammelson, M. J. Kurth, *Chem. Rev.* **2001**, *101*, 137–202; b) R. G. Franzen, *J. Comb. Chem.* **2000**, *2*, 195–214; c) K. Burgess, *Solid Phase Organic Synthesis*, Wiley-VCH, Weinheim, **2000**; d) F. Z. Dörwald, *Organic Synthesis on Solid Phase*, Wiley-VCH, Weinheim, **2000**; e) B. A. Lorbach, M. J. Kurth, *Chem. Rev.* **1999**, *99*, 1549–1582.
[11] D. L. Boger, J. Desharnais, K. Capps, *Angew. Chem. Int. Ed.* **2003**, *42*, 4138–4176.
[12] D. G. Hall, S. Manku, F. Wang, *J. Comb. Chem.* **2001**, *3*, 125–150.
[13] a) D. J. Ramon, M. Yus, *Angew. Chem. Int. Ed.* **2005**, *44*, 1602–1634; b) C. Hulme, T. Nixey, *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 921–929.
[14] A. Dömling, *Chem. Rev.* **2005**, *105*, doi: 10.1021/cr0505728.
[15] B. Yüçel, L. Arve, A. de Meijere, *Tetrahedron* **2005**, *61*, 11355–11373.
[16] C. Hulme, J. Peng, G. Morton, J. M. Salvino, T. Herpin, R. Labaudiniere, *Tetrahedron Lett.* **1998**, 7227.
[17] a) I. R. Baxendale, S. V. Ley, M. Nessi, C. Piutti, *Tetrahedron* **2002**, *58*, 6285–6304; b) R. I. Storer, T. Takemoto, P. S. Jackson, D. S. Brown, I. R. Baxendale, S. V. Ley, *Chem. Eur. J.* **2004**, *10*, 2529–2547.
[18] P. M. Abreu, P. S. Branco, *J. Braz. Chem. Soc.* **2003**, *14*, 675–712.
[19] A. M. Boldi, *Curr. Opin. Chem. Biol.* **2004**, *8*, 281–286.
[20] a) J.-Y. Ortholand, A. Ganesan, *Curr. Opin. Chem. Biol.* **2004**, *8*, 271–280; b) R. Breinbauer, M. Manger, M. Scheck, H. Waldmann, *Curr. Med. Chem.* **2002**, *9*, 2129–2145; c) U. Abel, C. Koch, M. Speitling, F. G. Hansske, *Curr. Opin. Chem. Biol.* **2002**, *6*, 453–458; d) K. C. Nicolaou, J. A. Pfefferkorn, *Biopolymers* **2001**, *60*, 171–193; e) P. Arya, R. Joseph, D. T. H. Chou, *Chem. Biol.* **2002**, *9*, 145–156; A. Ganesan, *Pure Appl. Chem.* **2001**, *73*, 1033–1039; f) L. A. Wessjohann, *Curr. Opin. Chem. Biol.* **2000**, *4*, 303–309; g) P. Arya, R. Joseph, Z. Gan, B. Rakic, *Chem. Biol.* **2005**, *12*, 163–180.
[21] a) O. Barun, K. Kumar, S. Sommer, A. Langerak, T. Mayer, U. Müller, H. Waldmann, *Eur. J. Org. Chem.* **2005**, *22*, 4773–4788; b) O. Barun, S. Sommer, H. Waldmann, *Angew. Chem. Int. Ed.* **2004**, *43*, 3195–3199.
[22] S. Sommer, H. Waldmann, *Chem. Comm.* **2005**, *45*, 5684–5686.
[23] M. A. Koch, L.-O. Wittenberg, S. Basu, D. A. Jeyaraj, E. Gourzoulidou, K. Reinecke, A. Odermatt, H. Waldmann, *Proc. Nat. Acad. Sci. USA* **2004**, *101*, 16721–16726.
[24] M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl, H. Waldmann, *Proc. Nat. Acad. Sci. USA* **2005**, *102*, 17272–17277.
[25] *Dictionary of Natural Products* (Chapman & Hall/CRC Informa, London), Version 14.1, **2005.**
[26] S. Schmitt, D. Kuhn, G. Klebe, *J. Mol. Biol.* **2002**, *323*, 387–406.
[27] R. A. S. Schweizer, A. G. Atanasov, B. M. Frey, A. Odermatt, *Mol. Cell. Endocrinol.* **2003**, *212*, 41–49.
[28] T. C. Sandeep, J. L. W. Yau, A. M. J. MacLullich, J. Noble, I. J. Deary, B. R. Walker, J. R. Seckl, *Proc. Nat. Acad. Sci. USA* **2004**, *101*, 6734–6739.
[29] S. P. Gunasekera, P. J. McCarthy, M. Kelly-Borges, E. Lobkovsky, J. Clardy, *J. Am. Chem. Soc.* **1996**, *118*, 8759–8760.

