

## Review

# Biology-inspired synthesis of compound libraries

M. Kaiser<sup>a</sup>, S. Wetzel<sup>b,c</sup>, K. Kumar<sup>b,c</sup> and H. Waldmann<sup>b,c,\*</sup>

<sup>a</sup> Chemical Genomics Centre of the Max Planck Society, Otto-Hahn-Str. 15, 44227 Dortmund (Germany)

<sup>b</sup> Max Planck Institute of Molecular Physiology, Department of Chemical Biology, Otto-Hahn-Str. 11, 44227 Dortmund (Germany), Fax: +49-231-133-2499, e-mail: herbert.waldmann@mpi-dortmund.mpg.de

<sup>c</sup> Universität Dortmund, Chemische Biologie, Otto-Hahn-Str. 6, 44227 Dortmund (Germany)

Received 23 October 2007; received after revision 26 November 2007; accepted 28 November 2007  
Online First 15 January 2008

**Abstract.** Biologically active small molecules represent the basis for chemical biology applications in which small molecules are used as chemical tools to probe biological processes. In this report, we review

two approaches to design and synthesize compound libraries for biological screenings, i.e., diversity-oriented synthesis (DOS) and biology-oriented synthesis (BIOS).

**Keywords.** Chemical biology, biology-oriented synthesis (BIOS), diversity-oriented synthesis (DOS), compound libraries.

## Introduction

Since the isolation of morphine by Friedrich Sertürner in 1805 and its subsequent validation as the active principle of opium, the synthesis of small molecules with biological activity has been one of the prime aims of chemistry. Although Paul Ehrlich already at the beginning of the 20th century noticed that the action of drugs is explicable in terms of conventional chemical interactions, drug discovery in its early times was mainly based on trial and error, as the molecular principles of drug action were not then understood. With the advances in biological sciences and especially the rise of structural biology allowing a study of drug-protein interactions at an atomic level during the second half of the 20th century, our molecular understanding of the underlying principles of drug action has increased dramatically. It is now clear that small molecules exert their biological activity by specific binding to defined regions on

their molecular targets, thereby modulating their biological function.

This small molecule-induced modulation of protein function represents the basis for medicinal chemistry and chemical biology. However, while in medicinal chemistry small molecules are sought to chemically modify diseased states, e.g., by small molecule inhibition of pathologically overexpressed enzymes, they are used in chemical biology as chemical probes for introducing an artificial and directed perturbation in a biological system. A comparison of the perturbed and non-perturbed state of the biological system then allows an analysis of the functioning of the biological system. This approach is complementary to existing biological methods.

Importantly, although the specifications of small compounds for medicinal chemistry or chemical biology applications differ in detail, both approaches rely on the availability of suitable small molecules. These small molecules are usually identified by a screening process in which a collection of compounds, also referred to as “compound libraries”, are systematically surveyed for a proper small molecule probe.

\* Corresponding author.

Consequently, the quality of the compound library is a major determinant for successful screening.

### Concepts for small molecule synthesis

The need for suitable chemical probes has led to several concepts that try to give an answer to one of the fundamental questions of chemical biology, i.e., which type of compounds should be synthesized and employed in biochemical or biological screenings. Unfortunately, this question cannot be answered generally taking into account the various applications for which small molecules can be used. The selection for a certain type of compound is especially complicated if the subsequent screening, e.g., in a phenotypic screen, should be performed unbiased.

As a consequence, chemical biologists and medicinal chemists have resorted to the synthesis of compound collections or libraries to provide a defined set of compounds for initial screening of various targets. However, this approach is hampered by the enormous range of chemical substances that in principle can be synthesized, also referred to as “chemical space”. Calculations have led to the proposition that  $10^{63}$  different compounds with drug-like properties exist in chemical space [1, 2]. Surely, such a gigantic number of compounds is no longer comprehensively or systematically amenable to synthesis. This dilemma has spurred several different approaches to cover extended regions of chemical structure space or to identify regions of chemical space for which the likelihood of obtaining biologically active compounds is given or at least enhanced.

In this review, we want to present the underlying concepts of two approaches for compound collection development, i.e., diversity-oriented synthesis (DOS) and biology-oriented synthesis (BIOS). While DOS focuses on the generation of structurally diverse and complex compound libraries, BIOS centers on the generation of small compound libraries based on scaffolds of proven biological relevance. Although both concepts therefore rely on different design concepts, they are not mutually exclusive, as either DOS libraries of scaffolds of proven biological relevance or BIOS libraries of a scaffold derived from DOS screening can be generated. Consequently, compound libraries made under these guiding principles have been extensively studied in various screening campaigns and their potential has been well proven as a source for biologically active small molecules. Other well-validated approaches such as fragment-based design or computational methods are not covered by this review. The reader is referred to other authoritative reviews [3–6].

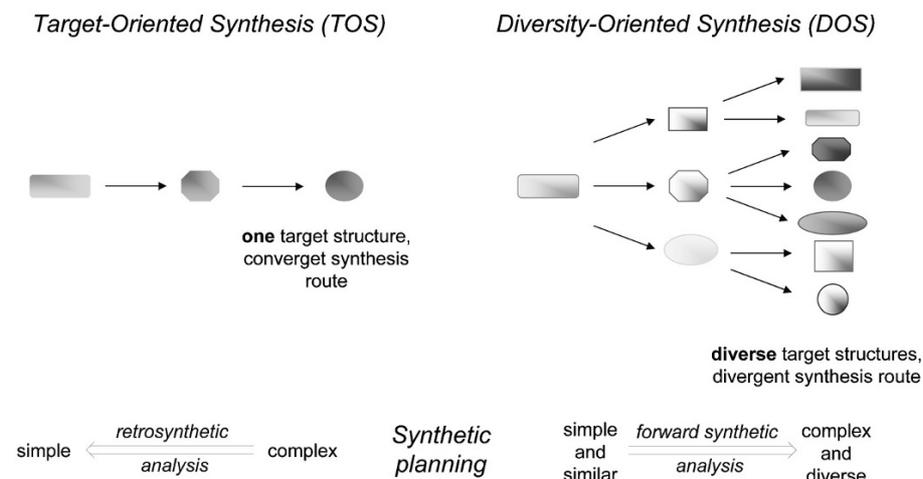
### Diversity-oriented synthesis

#### DOS as a principle for library synthesis

DOS has been introduced by Schreiber et al. [7] as a new concept for compound library generation. The term DOS has been coined to distinguish this approach from “traditional” synthesis that aims at the synthesis of one defined compound also referred to as the “target”. Such a target is often a natural product of complex structure and the aim of the synthesis is to obtain this compound in high efficiency and yield. Consequently, this “classical” textbook synthesis approach has been described as target-oriented synthesis (TOS). Contrary to TOS, DOS targets the facile preparation of collections of structurally complex and also diverse compounds from simple starting materials.

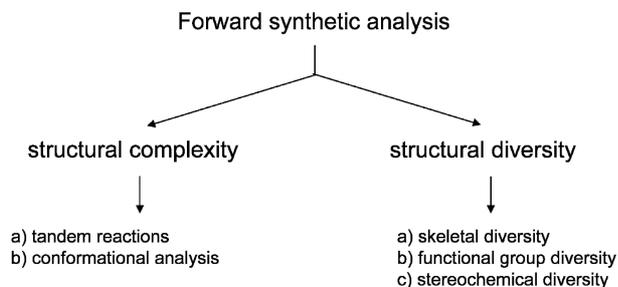
Towards this end, a new synthetic planning strategy is required. In TOS a “retrosynthetic analysis” of the desired target structure is employed (which means a step-wise analysis of chemical transformations, starting from the complex target structure towards more simple starting materials). DOS requires a forward “synthetic analysis” of the different reaction steps, initiating from simple and similar starting materials towards complex and diverse products (Fig. 1) [8, 9]. In this context, the two terms “structural complexity” and “diversity” should be closer examined. Structurally complex molecules exert a highly folded, preferably rigid three-dimensional structure. As a rule of thumb, structurally complex molecules often consist of annulated ring systems, while linear molecules show significantly lower complexity. Diversity on the other hand is a measure for the structural difference of the compounds in a given library. However, for library assessment it is usually difficult to exactly quantify these two terms. Instead, they are normally used in a more intuitive manner.

The importance of these two characteristics of compound libraries started to become clear in the 1990s as combinatorial chemistry for the first time allowed the synthesis of large compound collections. The parallel establishment of high-throughput screening (HTS) made a rapid and reliable screening of these huge libraries amenable. Although expectations in this approach were high in the beginning, the results of the screening campaigns were usually not very satisfactory. However, a thorough statistical analysis of the hit rates of different libraries revealed that compound collections which contained more structurally complex and diverse compounds usually yielded much higher hit rates than libraries consisting of “simple” compounds, highlighting their importance for biological activity. In fact, the two quality factors, library complexity and diversity, have to be regarded



**Figure 1.** Comparison of “traditional” target-oriented synthesis (TOS) and diversity-oriented synthesis (DOS).

as even more important if compound collections are used for phenotypic screens as the phenotypic readout can be caused by a multitude of proteins derived from different protein classes within a cell [10]. Therefore, the aim of the synthetic planning of a DOS library as performed during the forward synthetic analysis is to incorporate reaction pathways that lead to an optimization of these two important factors (Fig. 2).



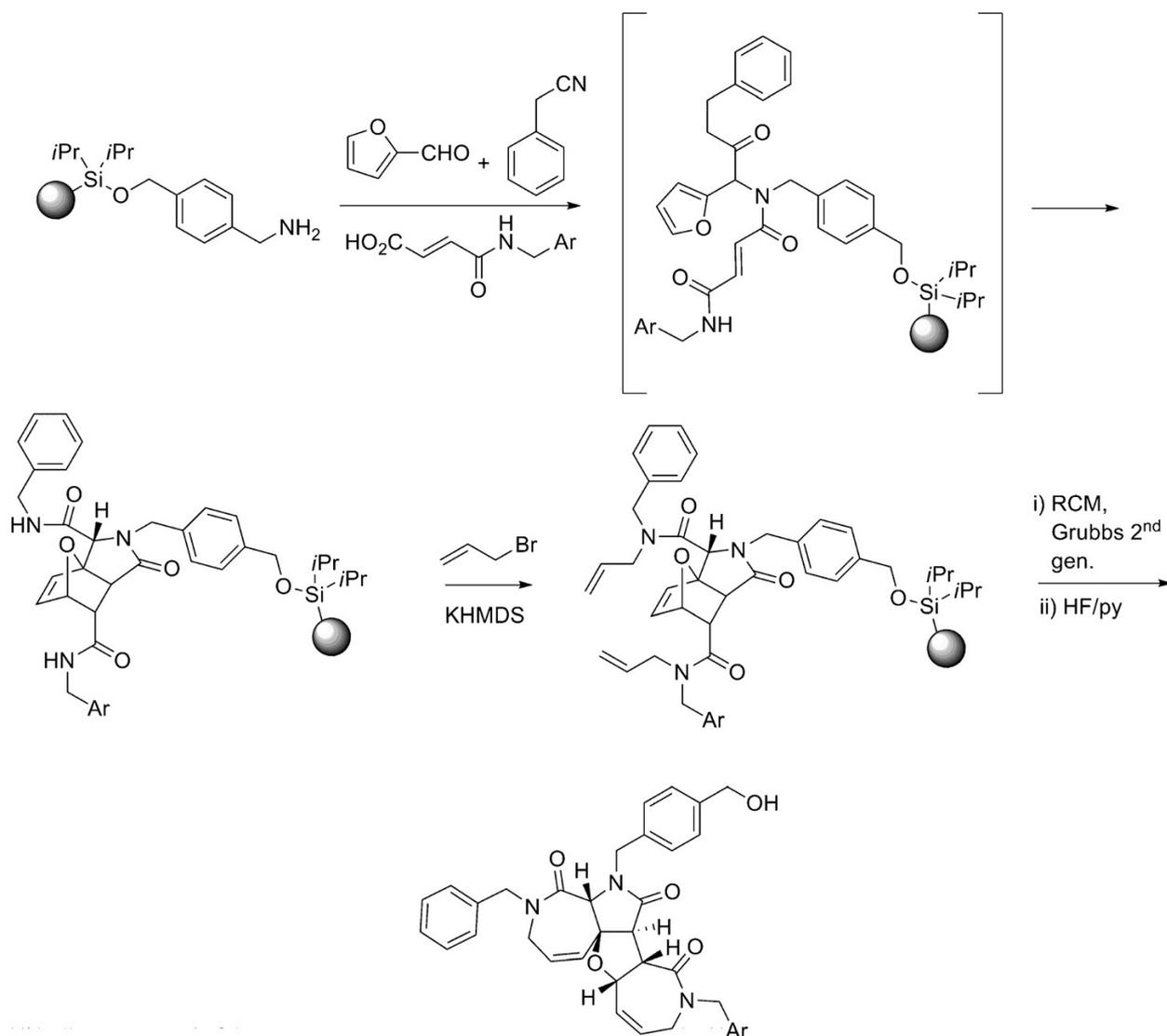
**Figure 2.** Forward synthetic analysis incorporates analysis of reaction pathways that lead to structural complexity and diversity.

Structural complexity for example has been introduced in DOS libraries through the use of tandem- or domino reactions. In these reactions, two or more transformations are intimately coupled with each other so that the product of the first reaction is used as a substrate for the subsequent reaction (Fig. 3) [11]. Additionally, conformational analysis of macrocyclic ring structures can be used to suggest suitable scaffolds with high structural complexity. Using this method, some medium-ring sized systems have been the focus of DOS [12].

Structural diversity on the other hand can be implemented by (a) generating different scaffolds or skeletons by specialized reactions – a concept which

leads to “skeletal diversity”; (b) introducing “functional group diversity”, i.e., changing the functional groups at a given scaffold structure, or (c) synthesizing different stereoisomers to access different binding patterns with the binding partners – an approach leading to “stereochemical diversity”. In general, skeletal diversity can be achieved in several ways, e.g., by building block diversity or by allowing one intermediate to react according to different protocols addressing distinct functional groups, which gives access to structurally diverse products. Also substituent diversity can be used, i.e., the generation of functional groups at different positions, which are then further converted in the same way. One example for such a reaction cascade is the introduction of terminal olefins with different chain lengths and subsequent ring closing metathesis. Other reactions used include for instance ring expansion and aldol reactions [13, 14]. Another synthetic approach towards skeletal diversity as an example for implementing structural diversity in compound libraries is illustrated in Figure 4 [15].

The impact of structural complexity and diversity on exploratory biological screening was probed and demonstrated by Schreiber et al. [16] in another study. In an extensive biological screening campaign using 40 different cell-based assays, the influence of stereochemical and skeletal diversity on biological activity was tested. For this purpose, a library consisting of 122 macrolides (Fig. 5A, structurally more complex) and their 122 linear precursors (Fig. 5B, structurally less complex) was employed. In addition, the stereochemistry (marked with an asterisk in Fig. 5) of the compounds was systematically permuted to examine the influence of stereochemical diversity. Statistical analysis of the screening results revealed that structurally more complex macrocycles were significantly more active than linear derivatives. In



**Figure 3.** An example of a synthetic strategy to obtain a compound library with high structural complexity. To achieve this, a tandem Ugi four component condensation-Diels-Alder reaction was used. Further complexity was then introduced through a subsequent ring-opening/ring-closing metathesis reaction to provide structurally complex products containing two five-membered and two seven-membered rings.

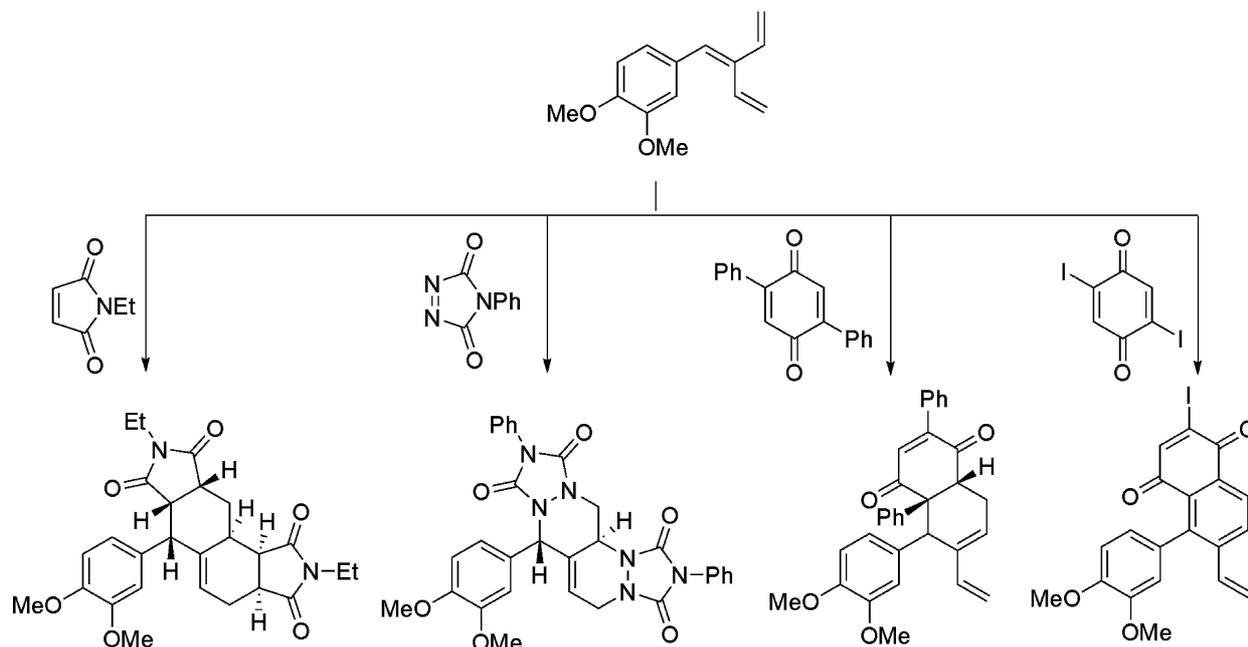
addition, the macrocyclic compounds were also more likely to exhibit activity in only one assay, indicating that structural complexity also results in higher target selectivity. The linear analogues (if active at all) exerted activity in several assays, pinpointing towards a connection between structural complexity and biological specificity. In addition, a hierarchical clustering of the data also identified stereochemistry (i.e., stereochemical diversity) as a second dominant factor in global activity patterns.

To achieve maximum efficiency in the library generation, the synthesis pathways in DOS should be limited to three to five reaction steps. Thus, the challenge in devising DOS libraries by forward synthetic analysis lies in counterbalancing the conflict between aiming at structurally complex and diverse

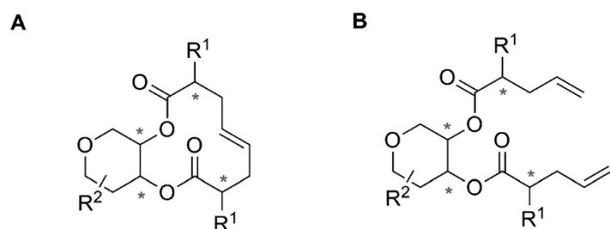
compounds on the one hand and synthesizing them in as few steps as possible with high yields and efficiency on the other hand. To achieve this difficult task, special type of chemical reactions, so-called “complexity-generating reactions”, have to be identified, optimized and implemented into the synthetic pathway.

#### Biological applications of DOS libraries

DOS libraries have been employed in manifold different screens, ranging from cell-free protein binding, enzyme-linked immunosorbent assays (ELISA) and fluorescence resonance energy transfer (FRET) assays to cell-based reporter-gene, cytoblot and phenotype assays [17, 18]. In these screens, libraries generated under the DOS principle proved to yield



**Figure 4.** An example of a DOS approach towards skeletal diversity. Four different skeletons in a one-step synthesis were generated from only one starting structure.



**Figure 5.** General chemical structures of compounds employed in the screen (permuted chiral centers have been marked with an asterisk). (A) Structures of macrocyclic library members (structurally more complex). (B) Structures of linear library members (structurally less complex).

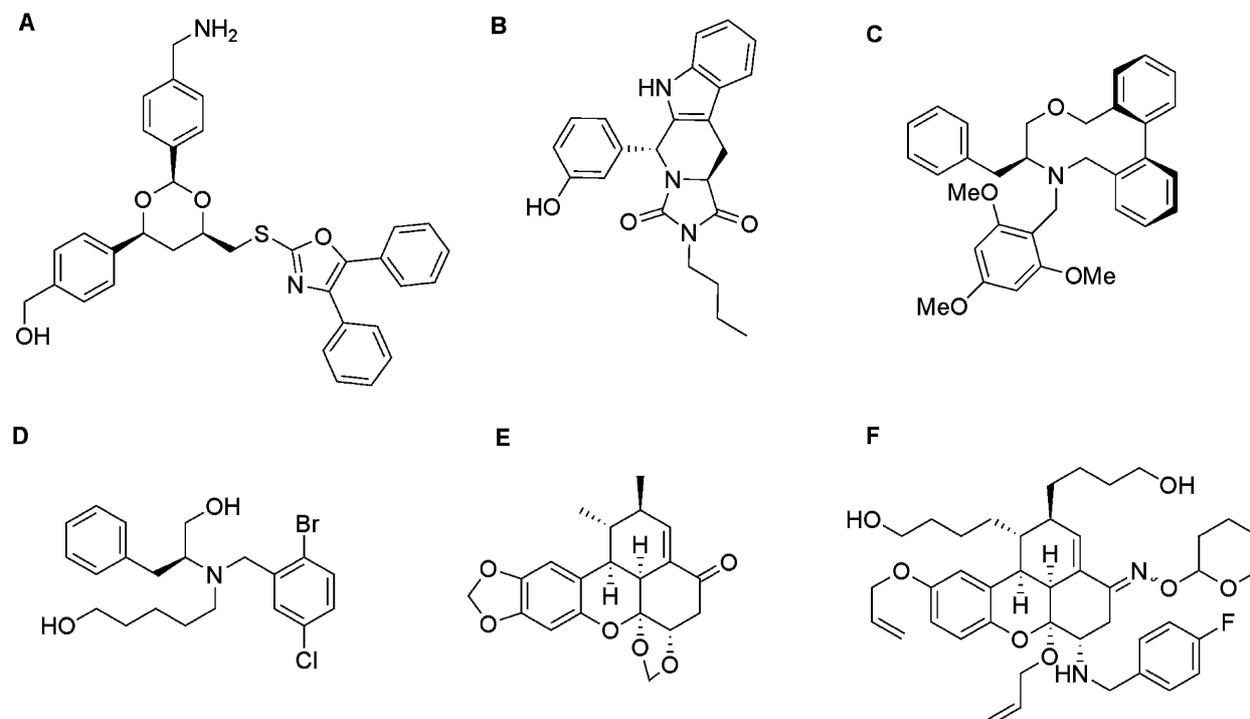
versatile small molecule probes for studying biology with chemical genomics applications.

For example, a small molecule named Uretupamine B (Fig. 6A) was found as a function-selective suppressor of the yeast glucose signaling protein Ure2p from a DOS library [19]. Ure2p regulates cellular responses to different carbon and nitrogen nutrients by repressing two transcription factors Nil1p and Gln3p. In response to the nutrient-type, differential signaling pathways are initiated. Consequently, a Ure2p knockout (*ure2D*) could not be used to study the Ure2p-mediated regulation of the signaling pathways; however, a function-selective small molecule inhibitor such as Uretupamine B was able to assist in this task. In another study, a cell permeable small molecule inhibitor HR22C16 (Fig. 6B) of the molecular motor protein Eg5 was discovered from a DOS library [20].

A DOS library of medium ring-sized biaryls was used in a plant phenotype screen. One compound, (*P*)-4k (Fig. 6C) was found to severely affect plant development, causing rigorous pigment loss and eventually death to plants. In another screen with the same library, a derivative named (*S*)-13ab (Fig. 6D) was identified as a small molecule probe for *Danio rerio* (zebra fish) development. This compound induced a strongly retarded development even at a concentration of 100 nM, leading to lower than normal pigmentation, weak hearts, abnormal brains and misshapen jaws by day 2 after fertilization [21].

DOS approaches can also be used to generate analogs of natural products, an approach which combines the strengths of DOS and BIOS [22]. For example, a 10,000-member library of carpanone (Fig. 6E) was synthesized. Screening of this library revealed various vesicular traffic inhibitors by preventing exocytosis from the Golgi apparatus, of which CLL-19 (Fig. 6F) was the most potent one. Interestingly, the starting structure carpanone did not exhibit any vesicular traffic inhibition [23].

Clearly, these examples demonstrate the potential of DOS libraries as a source for interesting biologically active small molecule probes. In view of the ongoing further developments and increasing availability of DOS based libraries, a new toolbox for the “small molecule biologist” is now available and will influence future research directions.



**Figure 6.** Structures of small molecule probes derived from DOS libraries. (A) Uretupamine B; (B) HR22C16 analogue; (C) (*P*)-4k; (D) (*S*)-13ab, (E) carpanone (no probe); (F) CLL-19.

## Biology-oriented synthesis

### BIOS as a principle for library synthesis

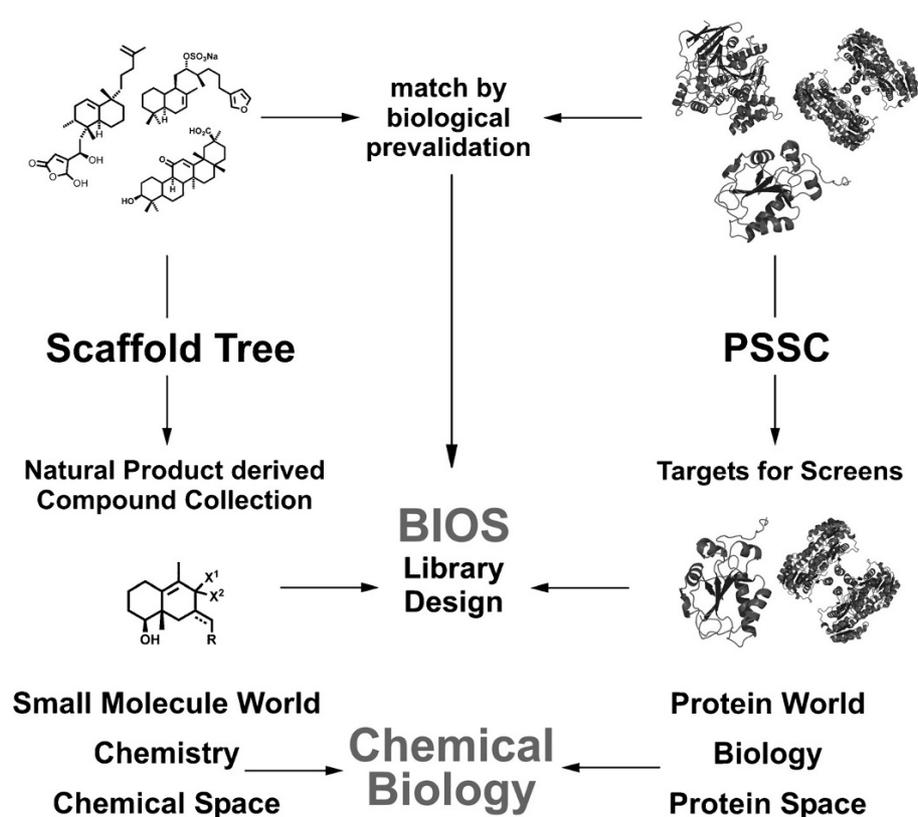
While DOS seeks to limit chemical space by focusing on structurally complex and diverse small molecules, BIOS builds on the inspiration given by nature through natural products. Natural products are defined chemical entities that often display biological activity. However, on its venture for low molecular weight natural products with such biological activity, nature itself explored only a tiny fraction of the available chemical space during evolution. Interestingly, the same holds true for the targets of those natural products, which are mainly proteins. During their evolution also only a tiny fraction of all possible amino acid combinations could have been probed by biosynthesis [24].

In addition, the three-dimensional folds of proteins have been shown to be even more conserved during evolution than their underlying sequence since similar three-dimensional structures can be formed by different sequences. This leads to the conclusion that the natural product space as well as the protein structure space explored by nature during evolution are strongly limited in size and highly conserved [25, 26].

As a structural matching between the protein binding site and its ligand is strictly required for a high-affinity interaction, both the structure space of the proteins

and of their ligands have to be highly complementary. The chemical space explored by natural products is certainly not the only region compatible with protein structure space but number and size of such regions in chemical space can be expected to be limited. The natural product space therefore is enriched with bioactive structures and thus contains promising starting points for the search of new bioactive molecules.

Consequently, Waldmann and coworkers [27] forged these complementary properties of bioactive small molecules (either natural products or known drugs) and their protein targets (receptors, proteins, transporters, etc.) into a new concept for library synthesis, termed BIOS. In this approach, library generation is focused on compound classes from “biologically relevant space”, e.g., the natural product or drug space, and only scaffolds from this area of proven relevance are selected as starting points for the design and synthesis of small focused libraries with limited diversity. BIOS, therefore, represents a conceptually alternative approach and is complementary to other library design concepts such as DOS. Moreover, with BIOS, previously unrecognized relations between the natural product and the protein world can be elucidated, such as predicting target proteins of a natural product-based library by bio- and cheminformatic methods (Fig. 7).



**Figure 7.** BIOS builds on biologically prevalidated structures, either natural products or drugs, and uses them as scaffolds for library generation. In addition, it allows an educated guess of protein targets of natural products using protein similarity analysis (see below).

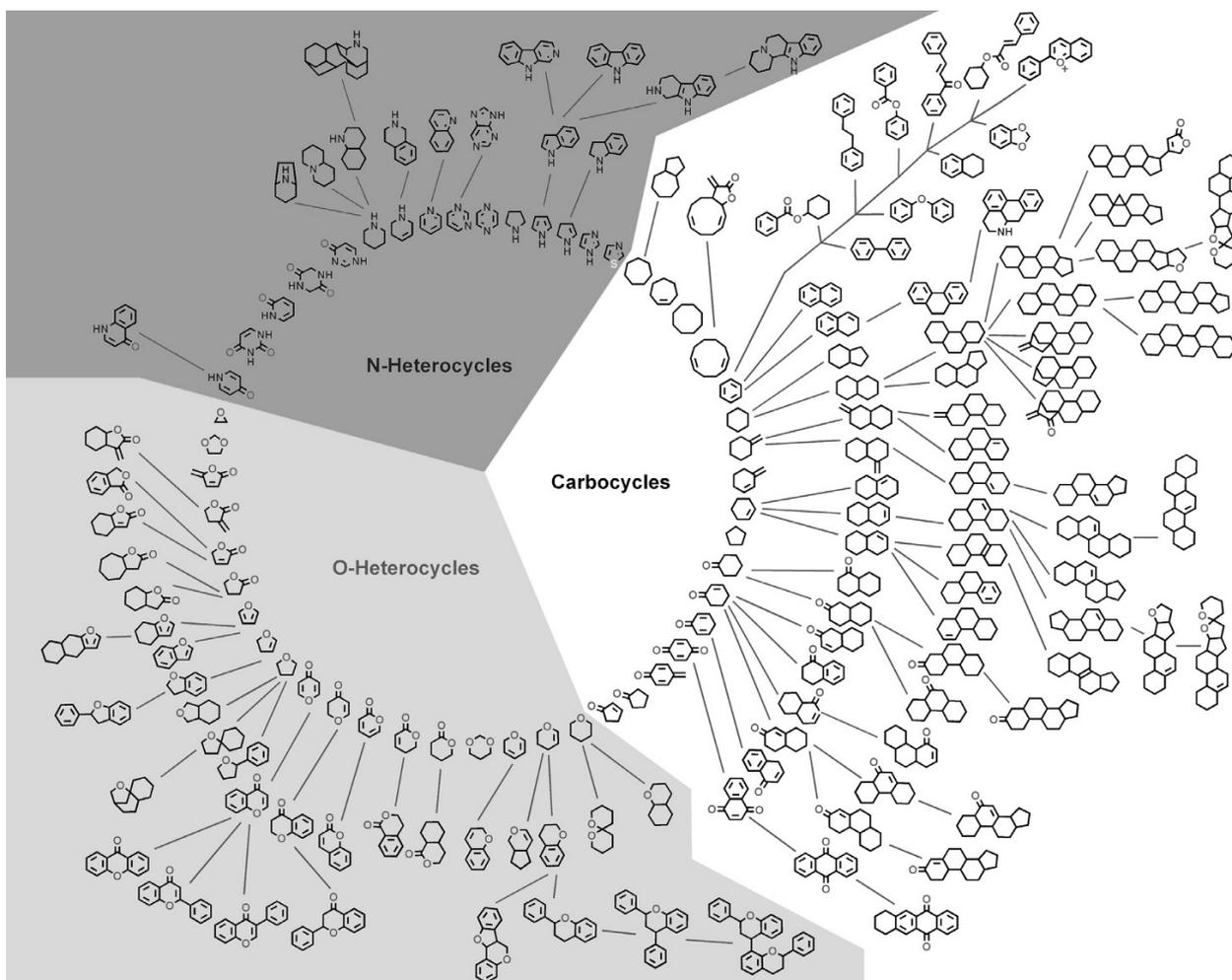
BIOS bases on and extends two concepts developed earlier by Waldmann and colleagues [28]. In the first concept, a “scaffold tree” that maps and correlates the scaffolds of natural products in a hierarchical manner was created. The resulting diagram results in a “structural classification of natural products” (SCONP) and allows charting of and navigation in chemical space. In the second concept, “protein structure similarity clustering” (PSSC), the biological targets of small molecules, i.e., the proteins, are classified into clusters employing structural similarity around the ligand binding sites [29]. Both approaches and the application of BIOS to design small molecules are explained in more detail in the next sections.

### SCONP in a scaffold tree

Natural products have a long history as starting points for drug design, as evidenced by almost half of the marketed drugs over the last 30 years having been derived from natural products [30]. However, a detailed analysis of the mode of action of natural products revealed that their overall biological activity is often a consequence of targeting multiple proteins. Natural products usually interact with multiple proteins during their biosynthesis as well as when they exert their mode of action, which indicates that structural information for protein recognition has

been encoded into their molecular structure during evolution.

Consequently, the underlying scaffolds of natural products have been recognized as privileged structures for library design for several years and explored in library design [31]. To chart and visualize the chemical space populated by natural products, Waldmann and coworkers [28] introduced a hierarchical classification of their underlying scaffolds that is based on their cyclic frameworks and linkers. The focus on molecules containing ring structures is justified as most small molecule inhibitors and drugs are based on cyclic systems, which leads to a stiffening of the molecule, resulting in enhanced target affinity due to less entropy loss upon binding. Starting from complex natural products, the scaffold was isolated based on rings, linker chains and ring-based double bonds. A set of defined rules derived from organic and medicinal chemistry was subsequently used to step-wise deconstruct the derived natural product scaffolds into smaller parent structures. This reductionist approach led to a unique scaffold tree, in which each scaffold in the hierarchy represents a well defined chemical entity and substructure of the original natural product. Moreover, the scaffold tree puts more complex scaffolds into relation with simpler framework, thus representing an efficient procedure to reduce molec-

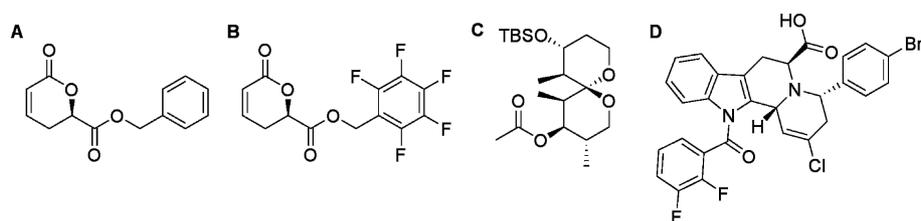


**Figure 8.** A scaffold tree generated in the structural classification of natural products (SCONP). For clarity, only scaffolds that represent at least 0.2% of the natural products described in the Dictionary of Natural Products (DNP) database have been sketched.

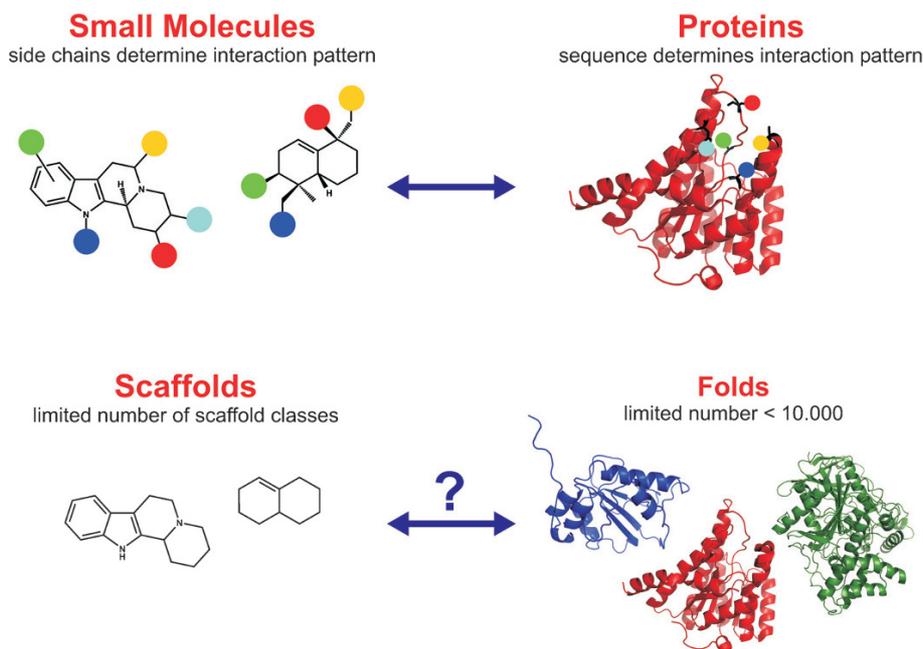
ular complexity from multiple annulated rings at the outer rims of the scaffold tree to monocyclic systems at the most inner circle of the scaffold and allowing a categorization of structural diversity in an intuitive and chemically meaningful way (Fig. 8).

Within the scaffold tree, three scaffold classes can be distinguished: nitrogen heterocycles, carbocycles and oxygen heterocycles of which two to four ring systems are most common. The significance of the natural product tree lies in its descriptive representation of biologically relevant scaffolds for the design and synthesis of focused compound collections. Moreover, it allows a correlation of different scaffold classes as performed during BIOS (see the BIOS section for more details). Based on this tree, Waldmann and coworkers synthesized libraries featuring spiroketal [32–34],  $\alpha,\beta$ -unsaturated lactone [35–40], tetrahydropyran [41, 42], indolactam [43, 44], decaline [45–48], indole scaffolds [49, 50] and indoloquinolizidine [27, 51], and proved their biological activity in various

screening campaigns, thereby validating the scaffold tree approach as a powerful method for generating libraries with high contents of biologically active molecules. For example, two cell-based screens of a library consisting of only 50  $\alpha,\beta$ -unsaturated  $\delta$ -lactones yielded small molecule modulators of cell cycle progression (Fig. 9A) and of viral entry via the secretory pathway (Fig. 9B) [40]. Using a library based on [5.5]spiroketals a modulator of the tubulin cytoskeleton in breast cancer cells was identified (Fig. 9C) [33]. In addition, several libraries based on natural product scaffolds have been screened for enzyme inhibition. Figure 9D shows a representative example that was derived from an indoloquinolizidine compound library and exerts potent inhibition of the phosphatase MtpB of *Mycobacterium tuberculosis* [27].



**Figure 9.** Structures of biologically active small molecules derived from SCONP-based compound libraries.



**Figure 10.** A complementary interaction between proteins and small molecules is the basis for binding. A similar complementary interaction might exist on the level of protein folds and natural product scaffolds.

### PSSC for identifying conserved protein-binding sites

Proteins are built from one or several domains that can be classified by their fold types. These folds describe the spatial arrangement of secondary structure elements such as  $\alpha$ -helices and  $\beta$ -sheets into defined motifs that have been found as conserved in nature and during evolution [52–55]. Consequently, different fold types were summarized in a structural classification of proteins (SCOP) database in which about 1000 distinct folds have been described up to now [56, 57]. Although computational genome analysis predicts that between 1000 and 10,000 different folds should exist in nature, the vast majority of proteins are expected to be built up from roughly 1000 of the most abundant folds [58–60]. The fold defines the three-dimensional structure of the protein backbone, as well as the shape and size of the active site and the spatial arrangement of the catalytic residues, while the different amino acid sequence of the proteins fine-tunes the exact shape of the binding pockets within the active site. As a consequence, only the combination of fold and sequence determines the binding properties of a protein and thus enables the enormous number of

different functions to be carried out by a limited number of folds [25, 61–63].

As revealed by SCONP, the number of scaffolds explored by natural products is also limited and conserved (see previous section) [64]. However, the implementation of different substitution patterns of the scaffold leads in the end to a vast number of possible compounds. This analogy between the protein and natural product world has been diagnosed by Waldmann and colleagues and has led to the hypothesis of a complementary relation between different protein folds and the scaffolds of natural products, similar to the existing complementarity of small molecules and their binding sites (Fig. 10) [29, 65–70]. If this proposition holds true, it would allow a prediction of suitable small molecule scaffolds for targeting a whole group of structurally similar proteins. Importantly, this prediction would be exclusively based on structure and nature's conservatism during evolution.

However, as described earlier, the binding pocket is shaped by the individual side chains of the amino acids constituting the fold. Consequently, library synthesis of an identified scaffold is necessary to optimize its

substitution pattern for the distinct binding site. To illustrate this in a more comprehensive example, one could imagine two nearly identical binding sites as generated by a similar fold. However, in the first case, a negative charge resides in a binding pocket of a protein while in the second case a positive charge is located there. Consequently, as the small molecules have to be complementary to the binding site, the scaffold structure has to be decorated with a positively charged side chain to target the protein with a negative charged pocket. However, the same compound would not bind to the second protein, as the two positive charges would induce repelling. This example clearly demonstrates that, while similar folds bind similar scaffolds, it is also necessary to generate sufficient chemical diversity to match biological diversity in the quest for biologically active molecules/ligands for proteins.

PSSC relies on bio- and cheminformatic algorithms and analysis of existing data, e.g., from the Protein Data Bank (PDB) to identify proteins in which the subfold around the ligand-binding site, the so-called “ligand-sensing core”, is similar. Structures that show structural similarity of the ligand-sensing cores are assigned to one protein structure similarity cluster. If a ligand for one member of the cluster is known, the scaffold of this ligand can then be expected to be a prevalidated starting point for the synthesis of ligands for the other members of the protein cluster. Consequently, members of a library of synthesized derivatives based on the scaffold should target several members of the protein cluster.

This concept was proven by a recent study with the natural product dysidiolide. Dysidiolide [71] targets the Cdc25A phosphatase that is involved in cell cycle regulation and represents a promising target for the development of novel anti-cancer agents [72–74]. A protein structure similarity analysis revealed that similar ligand binding sites are found in acetylcholine esterase (AChE) with a root mean square deviation (RMSD) of the aligned  $C_{\alpha}$  traces of 2.74 Å over 49 residues and a sequence identity of 8.2% [75]. Another member of the protein similarity cluster is 11 $\beta$ -hydroxysteroid dehydrogenase with an RMSD of 4.13 Å to Cdc25A over 80 aligned residues and a sequence identity of 5.0% [76–79].

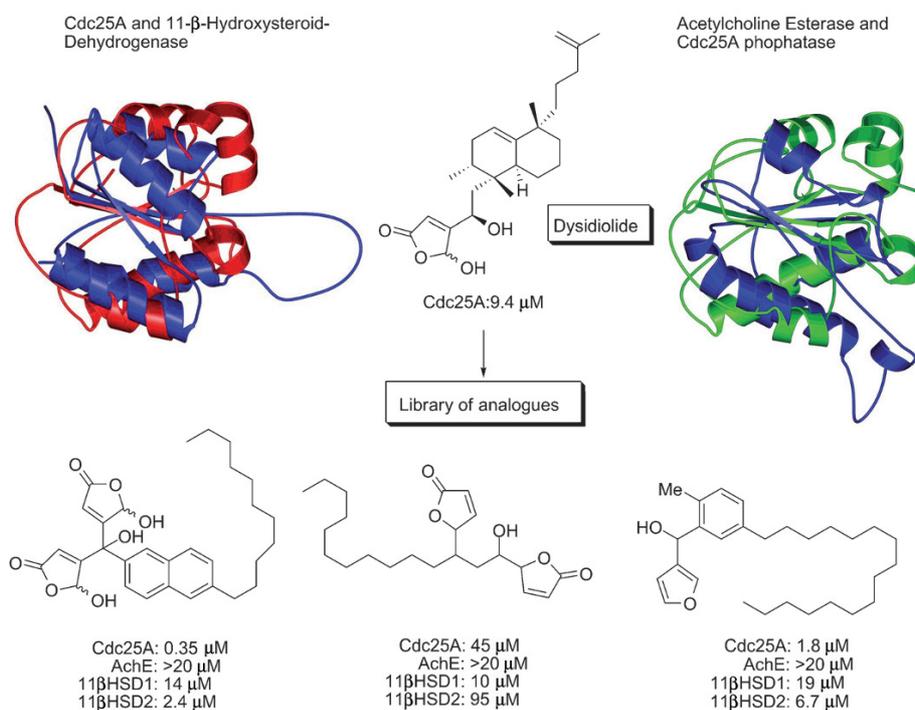
Waldmann and coworkers therefore assumed that dysidiolide might represent a starting point for the development of inhibitors of the other two members of the protein cluster. Consequently, the hydroxybutenolide part of the natural product was used as a scaffold for library development as it was assumed to be the active motif for phosphatase inhibition. A library of 150 compounds built around this scaffold was screened and revealed three hits for AChE in the

low micromolar range ( $IC_{50}$  values of 1.7–6.9  $\mu$ M). For 11 $\beta$ -hydroxysteroid dehydrogenase, seven hits with  $IC_{50}$  values from 2.4 to 10  $\mu$ M were found. Moreover, for one compound a pronounced selectivity between the two isoforms of 11 $\beta$ -hydroxysteroid dehydrogenase was determined (Fig. 11). These results demonstrated unequivocally the potential of the PSSC approach as a guiding principle for compound development [29].

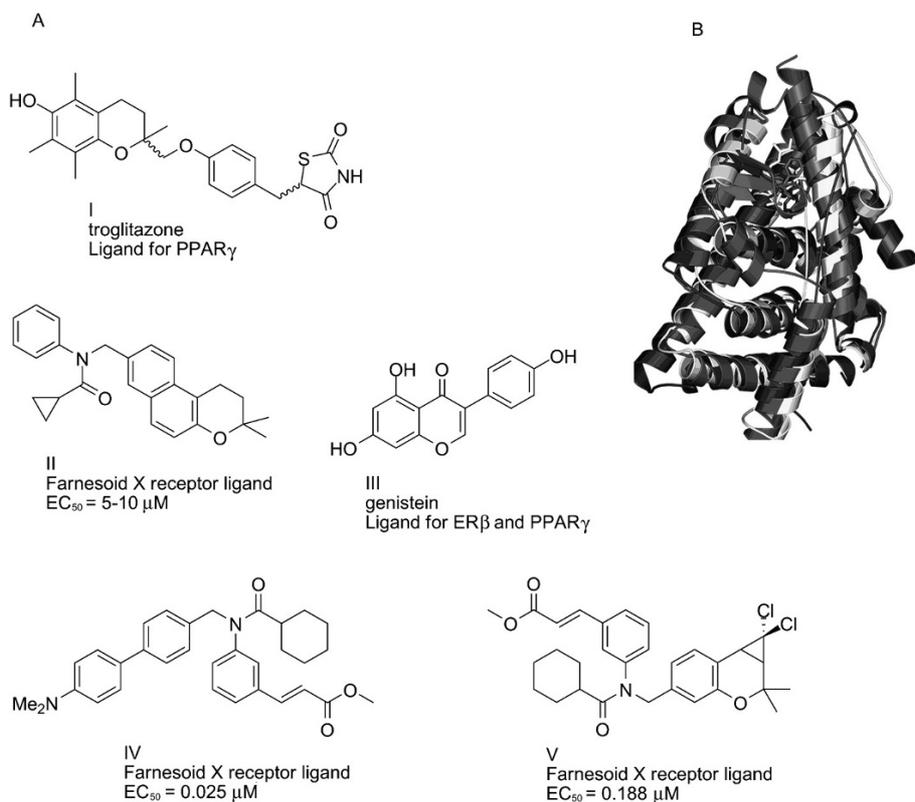
Another demonstration of the impact of PSSC can be made from a survey of literature data. The farnesoid X receptor (FXR) is a nuclear hormone receptor involved in development and homeostasis and plays key roles in diseases such as obesity, diabetes and cancer [80, 81]. A PSSC analysis of FXR disclosed a structural similarity to estrogen receptor  $\beta$  (ER $\beta$ ) [82] and the peroxisome proliferation-activated receptor  $\gamma$  (PPAR $\gamma$ ) [83]. Although exhibiting a similar fold pattern, their sequence similarities are below 20% (Fig. 12). However, as all three proteins were located in the same PSSC cluster, it should be possible to find a common scaffold class acting as general modulators of FXR, ER $\beta$  and PPAR $\gamma$ . A literature search revealed that the natural product genistein, which is based on a benzopyran scaffold, is a known inhibitor of ER $\beta$  and PPAR $\gamma$  [84]. The drug troglitazone also based on a benzopyran scaffold is a known PPAR $\gamma$  modulator [85]. Interestingly, a high-throughput screening for FXR inhibitors with a 10,000-member benzopyran library resulted in several hits, as would have been predicted by PSSC [86, 87]. In addition, the benzopyran library also yielded ligands for other members of the PSSC cluster, thereby further supporting the application of PSSC in library design.

Besides using X-ray or NMR protein structures, high-quality homology models can also be employed for PSSC, further extending its use. In a recent study, an additional method that also considers protein flexibility was introduced to the PSSC procedure. Molecular dynamics calculations generated different protein conformations, which are then tested for possible binding sites; using this modified procedure the ligand binding sites previously undetected by “static” PSSC were identified [88].

Moreover, the scope of PSSC can be extended by moving beyond structural similarity between targets, e.g., by including similarities between biosynthetic enzymes and other proteins. Quinn et al. [89, 90] proposed that the structural similarity between biosynthetic enzymes and other proteins could provide a link to identify new inhibitors. In their approach, proteins were grouped according to known inhibition data and subsequently searched for common structural motifs that may encode for the binding proper-



**Figure 11.** Protein structure similarity clustering (PSSC) of Cdc25A phosphatase (red in left PSSC and green in right PSSC, respectively) and acetylcholine esterase (blue, right PSSC) and 11 $\beta$ -hydroxysteroid dehydrogenase (red, left PSSC). Synthesis of a library of analogues yielded several hits for these two proteins with a similar fold in the binding pocket.



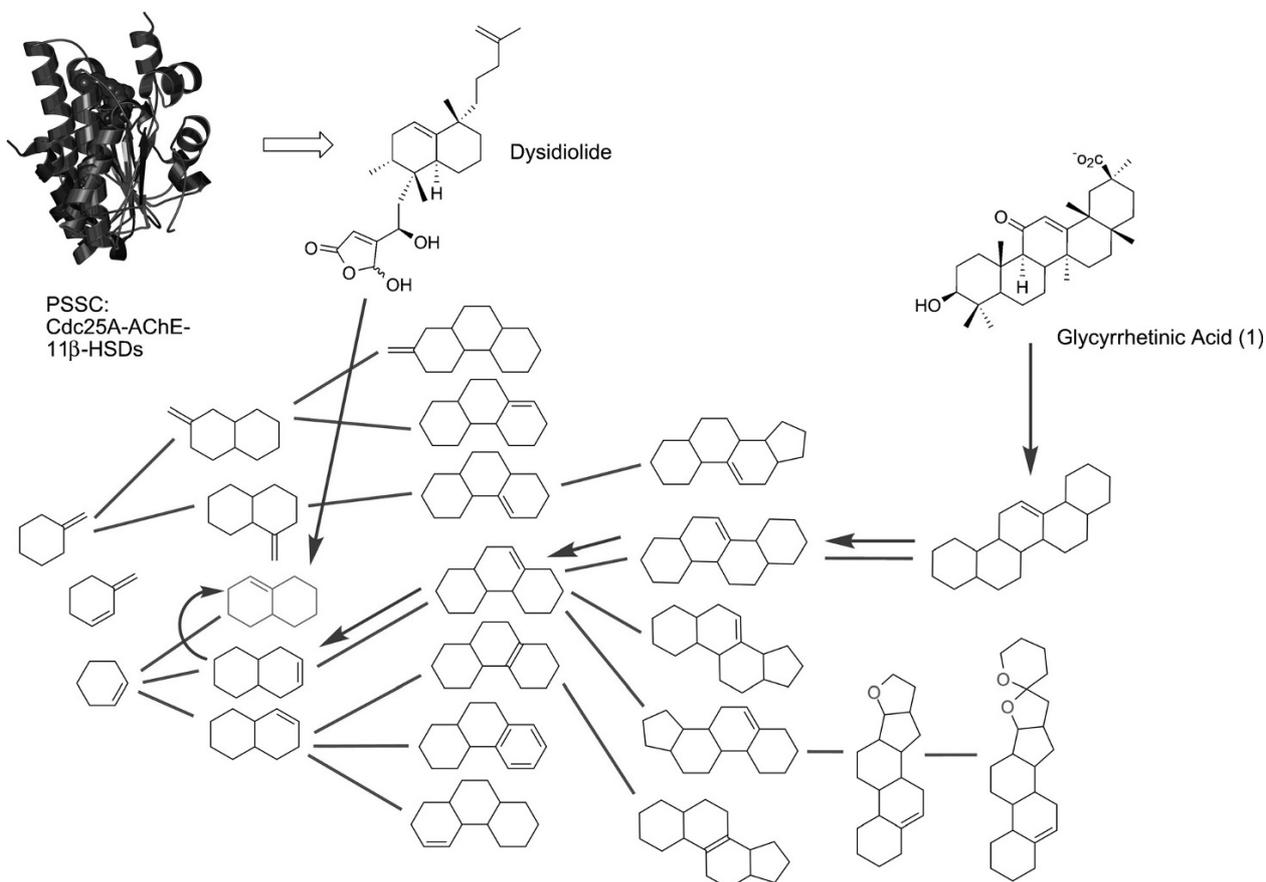
**Figure 12.** (A) Structures of benzopyran-based ligands for ER $\beta$ , PPAR $\gamma$  and FXR. (B) Superimposition of the X-ray structures of ER $\beta$  complexed with genistein (III), PPAR $\gamma$  with rosiglitazone and FXR

ties. Although this procedure is quite different from PSSC, which starts from structural comparisons, the underlying logic should also be applicable in the PSSC approach, which could then be used to cluster

biosynthetic enzymes and potential targets based on a structural similarity.

In summary, PSSC represents an alternative approach for compound library development. The grouping of

## Tree Segment - Carbocycles



**Figure 13.** BIOS is based on the merging of PSSC and SCOMP. The scaffold tree of glycyrrhetic acid was analyzed by SCOMP rules, leading to an octahydronaphthalene scaffold that represents a substructure of dysidiolide, a known Cdc25A inhibitor.

proteins opens new routes to small molecule inhibitor design as it allows the transfer of knowledge about the interaction of a small molecule with its protein target to a whole cluster of proteins. Moreover, it can be used to predict possible target proteins of small molecules. Besides its use for screening applications during drug discovery, it might also find application for identifying cross-inhibitions of proteins to foresee unwanted drug side effects during drug development.

### Merging SCOMP and PSSC leads to BIOS

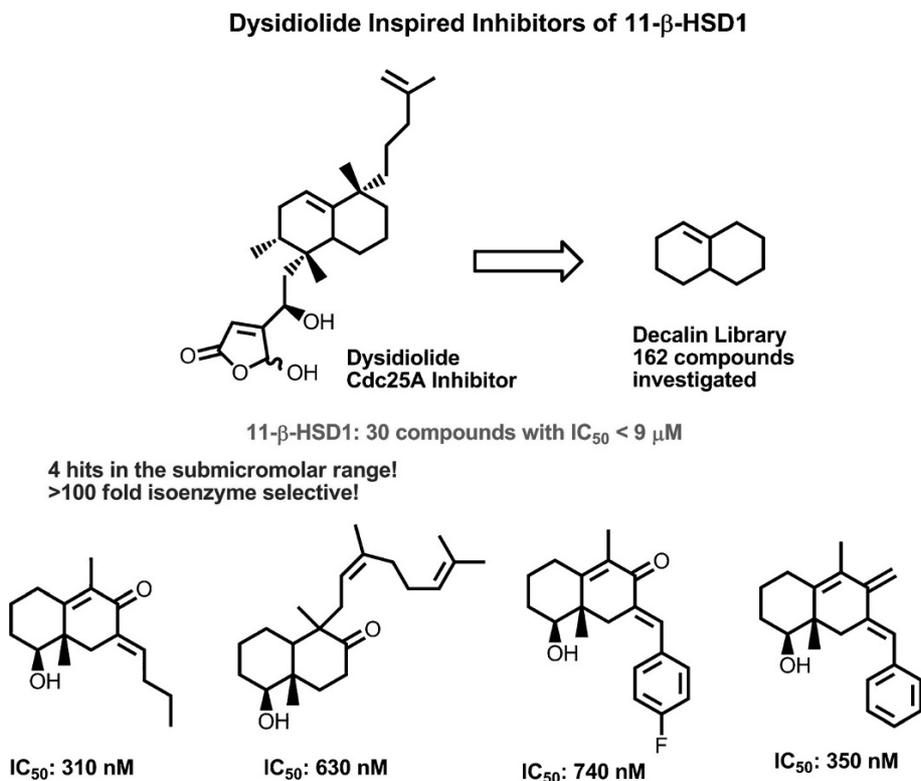
SCOMP and PSSC are novel principles for the design of biologically prevalidated compound collections. A merging of these two approaches into one concept, i.e., BIOS, leads to synergistic effects, providing, for example, an approach to simplify structures of known inhibitors while retaining their basic biological activity [27, 37, 40, 51].

The scaffold tree approach arranges scaffolds of natural products in a tree-like structure with the inner segments of the tree representing less complex

scaffolds. Consequently, moving inward along the branches (brachiation) of the tree to reach regions populated by less complex scaffolds leads to a structural simplification of the scaffolds. While structural simplifications of scaffold structures have been tried many times before with varying success, the scaffold tree offers a reduced set of simplifications pre-selected on the basis of organic and medicinal chemistry knowledge. This may lead to situations in which it may not be possible to choose the obvious retrosynthetic disconnection but rather the solution suggested by the scaffold tree.

This novel concept was first applied in a study involving the Cdc25A phosphatase protein cluster derived from PSSC, consisting of Cdc25A phosphatase, AChE and 11 $\beta$ -hydroxysteroid dehydrogenase (11 $\beta$ HSD) and the scaffold tree of the complex natural product glycyrrhetic acid, a known ligand of 11 $\beta$ HSD (Fig. 13).

SCOMP analysis of glycyrrhetic acid enabled a stepwise reduction of complexity from its pentacyclic



**Figure 14.** Merging PSSC and SCONP to BIOS for the development of 11 $\beta$ HSD inhibitors.

scaffold to monocyclic systems, suggesting several possible scaffolds for library synthesis. Therefore, the final decision about which ring scaffold should be explored in library synthesis was based by a second criteria, which was derived from PSSC. Glycrrhetic acid is a known 11 $\beta$ HSD modulator and is part of a PSSC cluster with Cdc25A and AChE. Dysidiolide contains a particular dehydrodecaline system and Cdc25A, its natural receptor, is also part of the cluster, which renders the dysidiolide dehydrodecaline core scaffold a viable starting point for the design of a library of molecules targeting 11 $\beta$ HSD [72, 76]. Using this concept, a library of *ca.* 500 dehydrodecaline compounds was synthesized, of which 162 compounds were tested in a biochemical 11 $\beta$ HSD inhibition assay. Remarkably, among these, 30 inhibitors with  $IC_{50}$  values below 10  $\mu M$  were found, of which four of them even displayed inhibitory activities between 310 and 740 nM. Moreover, the most potent inhibitors also proved bioactivity in cellular assays (Fig. 14) [28]. In a second example, structurally complex alkaloids, namely yohimibine and ajmalicine, were identified as moderate Cdc25A phosphatase inhibitors. The SCONP analysis of these natural products led to a scaffold tree, ranging from the parent pentacyclic over tricyclic indole-based scaffolds to monocyclic rings. Library synthesis of indoloquinolizidines (tetracyclic systems) revealed two compounds with  $IC_{50}$  values comparable to the natural products. Moreover, ex-

tending the screen to other phosphatases allowed the elucidation of structurally new inhibitors of the *M. tuberculosis* protein tyrosine phosphatase B (MPTPB) from the same library. Structurally simpler compounds with three- and two-membered rings also yielded another inhibitor of Cdc25A with a similar  $IC_{50}$  value as the natural products and also contained an inhibitor of the protein tyrosine phosphatase 1B (PTP1B) and eight inhibitors of MPTPB in the submicromolar range [27, 51].

A more elaborate set of rules was developed by Schuffenhauer et al. [91], which consists of 13 rules and was designed to reflect more the thinking of a medicinal chemist. Moreover, this rule set produces more transparent parent-child assignments, i.e., a given parent scaffold will always be deconstructed to the same child scaffold independent of the underlying dataset. This characteristic allows easy extension of scaffold tree to new molecules and integration of virtually any complex molecule at the outer fringes of the scaffold tree. Schuffenhauer et al. [91] also showed a correlation of the scaffold hierarchy with bioactivity indicating that brachiation may indeed yield simpler molecules with comparable bioactivity.

The separation of scaffolds and side chains is a common paradigm in medicinal chemistry processes, e.g., in hit-to-lead optimization. BIOS, however, facilitates hit generation, and it does so in two ways. It provides scaffolds prone to bind the target protein

and structurally simpler relatives as a starting point for library design. In design of the compound collection, side chain diversity is the key to increasing the probability of finding good hits by matching the distinct interaction pattern of the target selectively. BIOS also generates a target cluster that can be used to identify cross-inhibition and selectivity profiles very early on. In general, BIOS is well suited to provide starting points in the design of small compound collections for the discovery of initial hits.

In summary, these examples demonstrated that a combined use of SCONP and PSSC as proposed in BIOS represents an intriguing approach for the structural simplification of natural products, required for focused library synthesis. Although generality of the concept cannot yet be claimed by only two examples, and the brachiation approach cannot be expected to work for all scaffolds and proteins, it has been shown in principle that a combined use of SCONP and PSSC provides an advantageous opportunity for the design of structurally simplified protein ligands based on complex natural product structures. Therefore, BIOS represents a versatile approach for library synthesis as it significantly enhances the chance of devising and finding modulators of proteins, a challenge which lies at the heart of chemical biology.

### Conclusions and future prospects

The need for suitable small molecules for the future advancement of chemical biology has spurred the development of novel approaches for small molecule library design. From these, the two concepts of DOS and BIOS are discussed in this review. While DOS builds on general considerations of small molecule properties and tries to forge them into the target molecules, BIOS relies on biological prevalidation and conservatism during evolution in protein and natural product structures to identify biologically relevant substructures of vast chemical space. As a consequence, DOS usually leads to the generation of large libraries for exploring larger areas of chemical space. BIOS, however, prefers small focused libraries exploring a well-defined part of chemical space previously identified as of relevance to the particular problem. Key to this concept is the postulate of structural complementarity between similar proteins and their ligands.

Both approaches have the potential to deliver high-quality small molecule libraries, thereby laying the foundation for future advances in the field of chemical biology research and drug discovery. Libraries generated under one of these two principles are becoming

increasingly available to the scientific community and their contribution to the exploration of biology will become measurable within the coming years.

- 1 Dobson, C. M. (2004) Chemical space and biology. *Nature* 432, 824–828.
- 2 Bohacek, R. S., McMartin, C. and Guida, W. C. (1996) The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* 16, 3–50.
- 3 Augen, J. (2002) The evolving role of information technology in the drug discovery process. *Drug Discov. Today* 7, 315–323.
- 4 Blundell, T. L., Jhoti, H. and Abell, C. (2002) High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* 1, 45–54.
- 5 Carr, R. and Hann, M. (2002) The right road to drug discovery? *Modern Drug Discov.* 5, 45–48.
- 6 Patani, G. A. and LaVoie, E. J. (1996) Bioisosterism: A rational approach in drug design. *Chem. Rev.* 96, 3147–3176.
- 7 Schreiber, S. L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* 287, 1964–1969.
- 8 Burke, M. D. and Schreiber, S. L. (2004) A planning strategy for diversity-oriented synthesis. *Angew. Chem. Int. Ed.* 43, 46–58.
- 9 Spring, D. R. (2003) Diversity-oriented synthesis; a challenge for synthetic chemists. *Org. Biomol. Chem.* 1, 3867–3870.
- 10 Mayer, T. U., Kapoor, T. M., Haggarty, S. J., King, R. W., Schreiber, S. L. and Mitchinson, T. J. (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* 286, 971–974.
- 11 Lee, D., Sello, J. K. and Schreiber, S. L. (2000) Pairwise use of complexity-generating reactions in diversity-oriented organic synthesis. *Org. Lett.* 2, 709–712.
- 12 Lee, D., Sello, J. K. and Schreiber, S. L. (1999) A strategy for macrocyclic ring closure and functionalization aimed toward split-pool syntheses. *J. Am. Chem. Soc.* 121, 10648–10649.
- 13 Burke, M. D., Berger, E. M. and Schreiber, S. L. (2003) Generating diverse skeletons of small molecules combinatorially. *Science* 302, 613–618.
- 14 Burke, M. D., Berger, E. M. and Schreiber, S. L. (2004) A synthesis strategy yielding skeletally diverse small molecules combinatorially. *J. Am. Chem. Soc.* 126, 14095–14114.
- 15 Kwon, O., Park, S. B. and Schreiber, S. L. (2002) Skeletal diversity via a branched pathway: Efficient synthesis of 29 400 discrete, polycyclic compounds and their arraying into stock solutions. *J. Am. Chem. Soc.* 124, 13402–13404.
- 16 Kim, Y.-K., Arai, M. A., Arai, T., Lamenzo, J. O., Dean, E. F. III, Patterson, N., Clemons, P. A. and Schreiber, S. L. (2004) Relationship of stereochemical and skeletal diversity of small molecules to cellular measurement space. *J. Am. Chem. Soc.* 126, 14740–14745.
- 17 Walters, W. P. and Namchuk, M. (2003) Designing screens: How to make your hits a hit. *Nat. Rev. Drug Discov.* 2, 259–266.
- 18 Stockwell, B. R., Haggarty, S. J. and Schreiber, S. L. (1999) High-throughput screening of small molecules in miniaturized mammalian cell-based assays involving post-translational modifications. *Chem. Biol.* 6, 71–83.
- 19 Kuruvilla, F. G., Shamji, A. F., Sternson, S. M., Hergenrother, P. J. and Schreiber, S. L. (2002) Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature* 416, 653–657.
- 20 Hotha, S., Yarrow, J. C., Yang, J. G., Garrett, S., Renduchintala, K. V., Mayer, T. U. and Kapoor, T. M. (2003) HR22C16: A potent small-molecule probe for the dynamics of cell division. *Angew. Chem. Int. Ed.* 42, 2379–2382.
- 21 Spring, D. R., Krishnan, S., Blackwell, H. E. and Schreiber, S. L. (2002) Diversity-oriented synthesis of biaryl-containing medium rings using a one bead/one stock solution platform. *J. Am. Chem. Soc.* 124, 1354–1363.

- 22 Shang, S. and Tan, D. S. (2005) Advancing chemistry and biology through diversity-oriented synthesis of natural product-like libraries. *Curr. Opin. Chem. Biol.* 9, 248–258.
- 23 Goess, B. C., Hannoush, R. N., Chan, L. K., Kirchhausen, T. and Shair, M. D. (2006) Synthesis of a 10,000-membered library of molecules resembling carpanone and discovery of vesicular traffic inhibitors. *J. Am. Chem. Soc.* 128, 5391–5403.
- 24 Smith, J. M. (1970) Natural selection and the concept of a protein space. *Nature* 225, 563–564.
- 25 Russell, R. B., Sasiemi, P. D. and Sternberg, M. J. E. (1998) Supersites within superfolds. Binding site similarity in the absence of homology. *J. Mol. Biol.* 282, 903–918.
- 26 Wilks, H. M., Cortes, A., Emery, D. C., Halsall, D. J., Clarke, A. R. and Holbrook, J. J. (1992) Opportunities and limits in creating new enzymes: Experiences with the NAD-dependent lactate dehydrogenase frameworks of humans and bacteria. *Ann. N. Y. Acad. Sci.* 672, 80–93.
- 27 Nören-Müller, A., Reis-Corrêa Jr. I., Prinz, H., Rosenbaum, C., Saxena, K., Schwalbe, H. J., Vestweber, D., Cagna, G., Schunk, S., Schwarz, O., Schiewe, H. and Waldmann, H. (2006) Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis. *Proc. Natl. Acad. Sci. USA* 103, 10606–10611.
- 28 Koch, M. A., Schuffenhauer, A., Scheck, M., Wetzler, S., Casaulta, M., Odermatt, A., Ertl, P. and Waldmann, H. (2005) Charting biologically relevant chemical space: A structural classification of natural products (SCONP). *Proc. Natl. Acad. Sci. USA* 102, 17272–17277.
- 29 Koch, M. A., Wittenberg, L.-O., Basu, S., Jeyaraj, D. A., Gourzoulidou, E., Reinecke, K., Odermatt, A. and Waldmann, H. (2004) Compound library development guided by protein structure similarity clustering and natural product structure. *Proc. Natl. Acad. Sci. USA* 101, 16721–16726.
- 30 Newman, D. J. and Cragg, G. M. (2007) Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* 70, 461–477.
- 31 Evans, B. E., Rittle, K. E., Bock, M. G., DiPrado, R. M., Freidinger, R. M., Whitter, W. L., Lundell, G. F., Veber, D. F., Anderson, P. S., Chang, R. S. L., Lott, V. J., Cerino, D. J., Chen, T. B., Kling, P. J., Kunkel, K. A., Springer, J. P. and Hirshfield, J. (1988) Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* 31, 2235–2246.
- 32 Barun, O., Sommer, S. and Waldmann, H. (2004) Asymmetric solid-phase synthesis of 6,6-spiroketal. *Angew. Chem. Int. Ed.* 43, 3195–3199.
- 33 Barun, O., Kumar, K., Sommer, S., Langerak, A., Mayer, T. U., Müller, O. and Waldmann, H. (2005) Natural product-guided synthesis of a spiroacetal collection reveals modulators of tubulin cytoskeleton integrity. *Eur. J. Org. Chem.*, 4773–4788.
- 34 Sommer, S. and Waldmann, H. (2005) Solid phase synthesis of a spiro[5.5]ketal library. *Chem. Commun.* 5684–5686.
- 35 Mamane, V., Garcia, A. B., Umarye, J. D., Lessmann, T., Sommer, S. and Waldmann, H. (2007) Stereoselective allylation of aldehydes on solid support and its application in biology-oriented synthesis (BIOS). *Tetrahedron* 63, 5754–5767.
- 36 Garcia, A. B., Lessmann, T., Umarye, J. D., Mamane, V., Sommer, S. and Waldmann, H. (2006) Stereocomplementary synthesis of a natural product-derived compound collection on a solid phase. *Chem. Commun.* 3868–3870.
- 37 Umarye, J. D., Lessmann, T., Garcia, A. B., Mamane, V., Sommer, S. and Waldmann, H. (2007) Biology-oriented synthesis of stereochemically diverse natural-product-derived compound collections by iterative allylations on a solid support. *Chem. Eur. J.* 13, 3305–3319.
- 38 Bialy, L. and Waldmann, H. (2002) Synthesis of the protein phosphatase 2A inhibitor (4S,5S,6S,10S,11S,12S)-cytostatin. *Angew. Chem. Int. Ed.* 41, 1748–1751.
- 39 Bialy, L. and Waldmann, H. (2004) Total synthesis and biological evaluation of the protein phosphatase 2A inhibitor cytostatin and analogues. *Chem. Eur. J.* 10, 2759–2780.
- 40 Lessmann, T., Leuenberger, M. G., Menninger, S., Lopez-Canet, M., Müller, O., Hümmer, S., Bormann, J., Korn, K., Fava, E., Zerial, M., Mayer, T. U. and Waldmann, H. (2007) Natural product-derived modulators of cell cycle progression and viral entry by enantioselective oxa-Diels-Alder reactions on the solid phase. *Chem. Biol.* 14, 443–451.
- 41 Sanz, M. A., Voigt, T. and Waldmann, H. (2006) Enantioselective catalysis on the solid phase: Synthesis of natural product-derived tetrahydropyrans employing the enantioselective oxa-Diels-Alder reaction. *Adv. Synth. Cat.* 348, 1511–1515.
- 42 Broussy, S. and Waldmann, H. (2007) Solid phase synthesis of highly substituted tetrahydropyrans by tandem ene-reaction/intramolecular Sakurai cyclization. *J. Comb. Chem.* 9, 1138–1148.
- 43 Meseguer, B., Alonso-Díaz, D., Griebenow, N., Herget, T. and Waldmann, H. (1999) Natural product synthesis on polymeric supports – Synthesis and biological evaluation of an indolactam library. *Angew. Chem. Int. Ed.* 38, 2902–2906.
- 44 Meseguer, B., Alonso-Díaz, D., Griebenow, N., Herget, T. and Waldmann, H. (2000) Solid-phase synthesis and biological evaluation of a teleocidin library – Discovery of a selective PKC $\delta$  down regulator. *Chem. Eur. J.* 6, 3943–3957.
- 45 Brohm, D., Metzger, S., Bhargava, A., Müller, O., Lieb, F. and Waldmann, H. (2002) Natural products are biologically validated starting points in structural space for compound library development: Solid-phase synthesis of dysidiolide-derived phosphatase inhibitors. *Angew. Chem. Int. Ed.* 41, 307–311.
- 46 Brohm, D., Philippe, N., Metzger, S., Bhargava, A., Müller, O., Lieb, F. and Waldmann, H. (2002) Solid-phase synthesis of dysidiolide-derived protein phosphatase inhibitors. *J. Am. Chem. Soc.* 124, 13171–13178.
- 47 Stahl, P., Kissau, L., Mazitschek, R., Huwe, A., Furet, P., Giannis, A. and Waldmann, H. (2001) Total synthesis and biological evaluation of the nakijiquinones. *J. Am. Chem. Soc.* 123, 11586–11593.
- 48 Stahl, P., Kissau, L., Mazitschek, R., Giannis, A. and Waldmann, H. (2002) Natural product derived receptor tyrosine kinase inhibitors: Identification of IGF1R, Tie-2, and VEGFR-3 inhibitors. *Angew. Chem. Int. Ed.* 41, 1174–1178.
- 49 Rosenbaum, C., Baumhof, P., Mazitschek, R., Müller, O., Giannis, A. and Waldmann, H. (2004) Synthesis and biological evaluation of an indomethacin library reveals a new class of angiogenesis-related kinase inhibitors. *Angew. Chem. Int. Ed.* 43, 224–228.
- 50 Rosenbaum, C., Katzka, C., Marzinzik, A. and Waldmann, H. (2003) Traceless Fischer indole synthesis on the solid phase. *Chem. Commun.* 1822–1823.
- 51 Corrêa, I. R. Jr., Nören-Müller, A., Ambrosi, H.-D., Jakupovic, S., Saxena, K., Schwalbe, H., Kaiser, M. and Waldmann, H. (2007) Identification of inhibitors for mycobacterial protein tyrosine phosphatase B (MptpB) by biology-oriented synthesis (BIOS). *Chem. Asian J.* 2, 1109–1126.
- 52 Grant, A., Lee, D. and Orengo, C. (2004) Progress towards mapping the universe of protein folds. *Genome Biol.* 5, 107.
- 53 Koonin, E. V., Wolf, Y. I. and Karev, G. P. (2002) The structure of the protein universe and genome evolution. *Nature* 420, 218–223.
- 54 Leonov, H., Mitchell, J. S. B. and Arkin, I. T. (2003) Monte Carlo estimation of the number of possible protein folds: Effects of sampling bias and folds distributions. *Proteins* 51, 352–359.
- 55 Coulson, A. F. W. and Moulton, J. (2002) A unfold, mesofold, and superfold model of protein fold use. *Proteins* 46, 61–71.
- 56 Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.
- 57 Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J. P., Chothia, C. and Murzin, A. G. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.* 32, D226–D229.

- 58 Chothia, C., Gough, J., Vogel, C. and Teichmann, S. A. (2003) Evolution of the protein repertoire. *Science* 300, 1701–1703.
- 59 Liu, J. and Rost, B. (2003) Domains, motifs and clusters in the protein universe. *Curr. Opin. Chem. Biol.* 7, 5–11.
- 60 Lee, D., Grant, A., Buchan, D. and Orengo, C. (2003) A structural perspective on genome evolution. *Curr. Opin. Struct. Biol.* 13, 359–369.
- 61 Stark, A., Shkumatov, A. and Russell, R. B. (2004) Finding functional sites in structural genomics. *Structure* 12, 1405–1412.
- 62 Jones, S. and Thornton, J. M. (2004) Searching for functional sites in protein structures. *Curr. Opin. Chem. Biol.* 8, 3–7.
- 63 Anantharaman, V., Aravind, L. and Koonin, E. V. (2003) Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr. Opin. Chem. Biol.* 7, 12–20.
- 64 Lamb, S. S. and Wright, G. D. (2005) Accessorizing natural products: Adding to nature's toolbox. *Proc. Natl. Acad. Sci. USA* 102, 519–520.
- 65 Koch, M. A., Breinbauer, R. and Waldmann, H. (2003) Protein structure similarity as guiding principle for combinatorial library design. *Biol. Chem.* 384, 1265–1272.
- 66 Koch, M. A. and Waldmann, H. (2004) Natural product-derived compound libraries and protein structure similarity as guiding principles for the discovery of drug candidates. In: *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective*, pp. 377–403, Kubinyi, H. and Müller, G. (eds.) Wiley-VCH, Weinheim.
- 67 Koch, M. A. and Waldmann, H. (2005) Protein structure similarity clustering and natural product structure as guiding principles in drug discovery. *Drug Discov. Today* 10, 471–483.
- 68 Balamurugan, R., Dekker, F. J. and Waldmann, H. (2005) Design of compound libraries based on natural product scaffolds and protein structure similarity clustering (PSSC). *Mol. Biosyst.* 1, 36–45.
- 69 Dekker, F. J., Koch, M. A. and Waldmann, H. (2005) Protein structure similarity clustering (PSSC) and natural product structure as inspiration sources for drug development and chemical genomics. *Curr. Opin. Chem. Biol.* 9, 232–239.
- 70 Dekker, F. J., Wetzel, S. and Waldmann, H. (2006) Natural product scaffolds and protein structure similarity clustering (PSSC) as inspiration sources for compound library design in chemogenomics and drug development. In: *Chemogenomics: Knowledge-based Approaches to Drug Discovery*, pp. 59–84, Jacoby, E. (ed.), World Scientific Publishing, New Jersey.
- 71 Gunasekera, S. P., McCarthy, P. J., Kelly-Borges, M., Lobkovsky, E. and Clardy, J. (1996) Dysidiolide: A novel protein phosphatase inhibitor from the Caribbean sponge *Dysidea etheria* de Laubenfels. *J. Am. Chem. Soc.* 118, 8759–8760.
- 72 Lyon, M. A., Ducruet, A. P., Wipf, P. and Lazo, J. S. (2002) Dual-specificity phosphatases as targets for antineoplastic agents. *Nat. Rev. Drug Discov.* 1, 961–976.
- 73 Fauman, E. B., Cogswell, J. P., Lovejoy, B., Rocque, W. J., Holmes, W., Montana, V. G., Pivnicka-Worms, H., Rink, M. J. and Saper, M. A. (1998) Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell* 93, 617–625.
- 74 Bialy, L. and Waldmann, H. (2005) Inhibitors of protein tyrosine phosphatases: Next-generation drugs? *Angew. Chem. Int. Ed.* 44, 3814–3839.
- 75 Racchi, M., Mazzucchelli, M., Porrello, E., Lanni, C. and Covoni, S. (2004) Acetyl-cholinesterase inhibitors: Novel activities of old molecules. *Pharmacol. Res.* 50, 441–451.
- 76 Walker, B. R. and Seckl, J. R. (2003) 11- $\beta$ -Hydroxysteroid dehydrogenase type 1 as a novel therapeutic target in metabolic and neurodegenerative disease. *Expert Opin. Ther. Targets* 7, 771–783.
- 77 Chrousos, G. P. (2004) Is 11-beta-hydroxysteroid dehydrogenase type 1 a good therapeutic target for blockade of glucocorticoid actions? *Proc. Natl. Acad. Sci. USA* 101, 6329–6330.
- 78 Ross, S. A., Gulve, E. A. and Wang, M. (2004) Chemistry and biochemistry of type 2 diabetes. *Chem. Rev.* 104, 1255–1282.
- 79 Sandeep, T. C., Yau, J. L. W., MacLulich, A. M. J., Noble, J., Deary, I. J., Walker, B. R. and Seckl, J. R. (2004) 11- $\beta$ -Hydroxysteroid dehydrogenase inhibition improves cognitive function in healthy elderly men and type 2 diabetics. *Proc. Natl. Acad. Sci. USA* 101, 6734–6739.
- 80 Robinson-Rechavi, M., Garcia, H. E. and Laudet, V. (2003) The nuclear receptor superfamily. *J. Cell Sci.* 116, 585–586.
- 81 Gronemeyer, H., Gustafsson, J.-A. and Laudet, V. (2004) Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discov.* 3, 950–964.
- 82 Jordan, V. C. (2004) Selective estrogen receptor modulation: Concept and consequences in cancer. *Cancer Cell* 5, 207–213.
- 83 Alarcón de la Lastra, C., Sánchez-Fidalgo, S., Villegas, I. and Motilva, V. (2004) New pharmacological perspectives and therapeutic potential of PPAR-gamma agonists. *Curr. Pharm. Des.* 10, 3505–3524.
- 84 Dang, Z.-C., Audinot, V., Papapoulos, S. E., Boutin, J. A. and Löwik, C. W. G. M. (2003) Peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ ) as a molecular target for the soy phytoestrogen genistein. *J. Biol. Chem.* 278, 962–967.
- 85 Van Gaal, L. and Scheen, A. J. (2002) Are all glitazones the same? *Diabetes Metab. Res. Rev.* 18, S1-S4.
- 86 Nicolaou, K. C., Evans, R. M., Roecker, A. J., Hughes, R., Downes, M. and Pfefferkorn, J. A. (2003) Discovery and optimization of non-steroidal FXR agonists from natural product-like libraries. *Org. Biomol. Chem.* 1, 908–920.
- 87 Downes, M., Verdecia, M. A., Roecker, A. J., Hughes, R., Hogenesch, J. B., Kast-Woelbern, H. R., Bowman, M. E., Ferrer, J.-L., Anisfeld, A. M., Edwards, P. A., Rosenfeld, J. M., Alvarez, J. G. A., Noel, J. P., Nicolaou, K. C. and Evans, R. M. (2003) A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. *Mol. Cell* 11, 1079–1092.
- 88 Charette, B. D., MacDonald, R. G., Wetzel, S., Berkowitz, D. B. and Waldmann, H. (2006) Protein structure similarity clustering: Dynamic treatment of PDB structures facilitates clustering. *Angew. Chem. Int. Ed.* 45, 7766–7770.
- 89 McArdle, B. M. and Quinn, R. J. (2007) Identification of protein fold topology shared between different folds inhibited by natural products. *Chembiochem* 8, 788–798.
- 90 McArdle, B. M., Campitelli, M. R. and Quinn, R. J. (2006) A common protein fold topology shared by flavonoid biosynthetic enzymes and therapeutic targets. *J. Nat. Prod.* 69, 14–17.
- 91 Schuffenhauer, A., Ertl, P., Roggo, S., Wetzel, S., Koch, M. A. and Waldmann, H. (2007) The scaffold tree – Visualization of the scaffold universe by hierarchical scaffold classification. *J. Chem. Inf. Model.* 47, 47–58.

---

To access this journal online:

<http://www.birkhauser.ch/CMLS>

---