ACCOUNTS
of chemical research

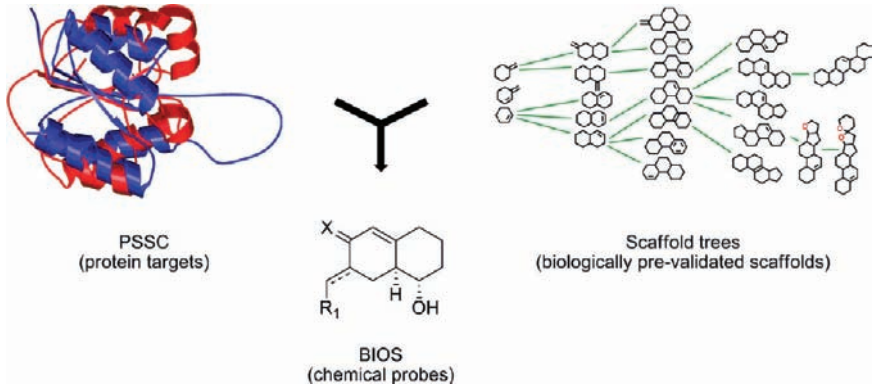
Bioactivity-Guided Navigation of Chemical Space

ROBIN S. BON† AND HERBERT WALDMANN*

Department of Chemical Biology, Max Planck Institute of Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany, and Faculty of Chemistry, Technische Universität Dortmund, Otto-Hahn-Strasse 6, 44227 Dortmund, Germany

RECEIVED ON JANUARY 14, 2010

CONSPECTUS



A central aim of biological research is to elucidate the many roles of proteins in complex, dynamic living systems; the selective perturbation of protein function is an important tool in achieving this goal. Because chemical perturbations offer opportunities often not accessible with genetic methods, the development of small-molecule modulators of protein function is at the heart of chemical biology research. In this endeavor, the identification of biologically relevant starting points within the vast chemical space available for the design of compound collections is a particularly relevant, yet difficult, task. In this Account, we present our research aimed at linking chemical and biological space to define suitable starting points that guide the synthesis of compound collections with biological relevance.

Both protein folds and natural product (NP) scaffolds are highly conserved in nature. Whereas different amino acid sequences can make up ligand-binding sites in proteins with highly similar fold types, differently substituted NPs characterized by particular scaffold classes often display diverse biological activities. Therefore, we hypothesized that (i) ligand-binding sites with similar ligand-sensing cores embedded in their folds would bind NPs with similar scaffolds and (ii) selectivity is ensured by variation of both amino acid side chains and NP substituents. To investigate this notion in compound library design, we developed an approach termed biology-oriented synthesis (BIOS). BIOS employs chem- and bioinformatic methods for mapping biologically relevant chemical space and protein space to generate hypotheses for compound collection design and synthesis. BIOS also provides hypotheses for potential bioactivity of compound library members. On the one hand, protein structure similarity clustering (PSSC) is used to identify ligand binding sites with high subfold similarity, that is, high structural similarity in their ligand-sensing cores. On the other hand, structural classification by scaffold trees (for example, structural classification of natural products or SCONP), when combined with software tools like "Scaffold Hunter", enables the hierarchical structural classification of small-molecule collections in tree-like arrangements, their annotation with bioactivity data, and the intuitive navigation of chemical space. Brachiation (in a manner analogous to tree-swinging primates) within the scaffold trees serves to identify new starting points for the design and synthesis of small-molecule libraries, and PSSC may be used to select potential protein targets.

The introduction of chemical diversity in compound collections designed according to the logic of BIOS is essential for the frequent identification of small molecules with diverse biological activities. The continuing development of synthetic methodology, both on solid phase and in solution, enables the generation of focused small-molecule collections with sufficient substituent, stereochemical, and scaffold diversity to yield comparatively high hit rates in biochemical and biological screens from relatively small libraries. BIOS has also allowed the identification of new ligand classes for several different proteins and chemical probes for the study of protein function in cells.

Published on the Web 05/19/2010 www.pubs.acs.org/acr
10.1021/ar100014h

Introduction

In order to close the genotype−phenotype gap biological research has to reach beyond genomics, proteomics, and dissection of biological systems into their prime constituents. Protein function in space and time is regulated in complex networks with other biomacromolecules, small molecules and supramolecular structures like membranes. Systematic perturbation of individual network components reveals network wiring, topology, and dynamics.[1] Whereas genetic methods are chronic and change the native structure of the network, chemical perturbations with small molecule modulators of protein function are acute, that is, they do not change the system and are rapid, reversible, conditional, and easily tunable by varying concentration.

In chemical biology research, the choice of the compound class for the study of biological phenomena and systems is of utmost importance. Chemical space, which encompasses all possible small organic molecules, is vast. Current estimates of the number of small molecules in drug-like chemical space exceed 10^{60}, and there is probably not enough matter in the universe and not sufficient time to make them all.[2] Also nature has not fully explored chemical space complementary to the binding sites of proteins. For proteins with an average size of 300 residues, made from 20 different amino acids, more than 10^{390} unique combinations are possible.[2] However, the human genome, for example, encodes a mere 25 000 proteins.[3] The challenge is to identify biologically relevant areas in chemical space that are likely to contain biologically active compounds.[4] For this purpose, various strategies, in particular, ones employing mechanistic considerations, evolutionary arguments (sequence homology), and the generation of chemical diversity (diversity-oriented synthesis, DOS), have been developed, and new in silico methods have been introduced for virtual screening, de novo compound design, and the structural classification of small molecules.[5−9]

As a complementary approach, we have devised a structure-based, systematic analysis of the biological and chemical space exploited by nature during the evolution of proteins and natural products (NPs). To this end, we have developed cheminformatic and bioinformatic approaches to chart chemical and biological space defined by given compound collections and databases and the known bioactivities of analyzed compounds. These approaches focus on hierarchical structural relationships between compound classes and structural similarities in ligand-sensing protein cores and lead to the prospective identification of new targets of known biologically active compound classes and the design of compound libraries. In this Account, we discuss the development of these hypothesis-generating tools and their combination with novel organic synthesis methods. We refer to this concept as biology-oriented synthesis (BIOS).

The Link between Natural Product Space and Protein Space

Natural products (NPs) remain a major source of inspiration for the development of new drugs.[10,11] NPs have evolved to interact with multiple proteins; that is, they have to be recognized, bound, and modified by the biosynthetic machinery and they need to interact with proteins when exerting their biological functions. Often, their multiple biological activities, for example, a combination of taste and toxicity, reflect the interaction with multiple proteins. This is especially true for classes of NPs with related structures. On the level of NP scaffolds, nature has exploited only a tiny fraction of chemical space. However, NPs that share a common scaffold but differ only in their substituent patterns may display a range of well-defined biological activities. Therefore NP scaffolds are encoded with structural information for recognition by proteins. They define "privileged structures", that is, molecular scaffolds that confer the ability to interact with multiple protein targets to the entire compound class and that, therefore, are capable of interacting with a variety of cellular targets.[12]

Selective substrate recognition by biosynthetic enzymes and interaction with appropriate receptors requires tight interaction and matching structures between NPs and proteins. The spatial arrangement of protein secondary structure elements like α-helices and β-sheets (i.e., folds)[13] is highly conserved in nature, and current bioinformatics knowledge revealed that the total number of fold types is in the range of 1000−8000.[14] However, this conservation is complemented by a level of diversity resulting from the different amino acid sequences defining similar folds. The fold of a protein shapes its scaffold; that is, the 3D structure of the amino acid backbone and the structure of the subfold defining the ligand-sensing core determines the size and shape of ligand binding sites and the spatial orientation of catalytic and ligand-recognizing residues. Diversity of the amino acid side chains accounts for the diversity of the ligands that can be bound. Similarly, a limited number of NP scaffolds (possibly resulting from a limited number of subfold types present in their biosynthetic enzymes) serves to spatially position a diverse range of substituents for selective protein binding (Figure 1). This suggests a relationship between the subfold types of ligand-sensing cores of protein domains and scaffolds of NPs, as well as a

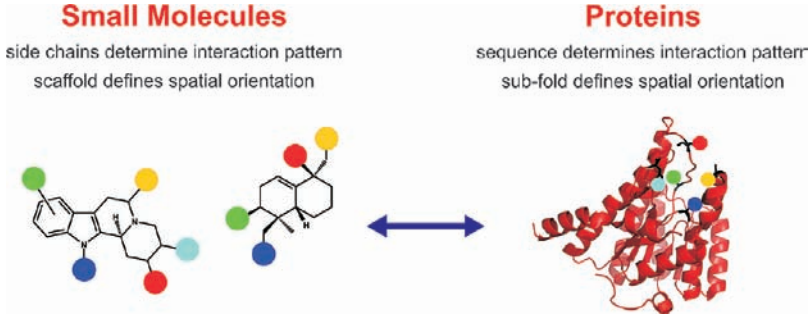
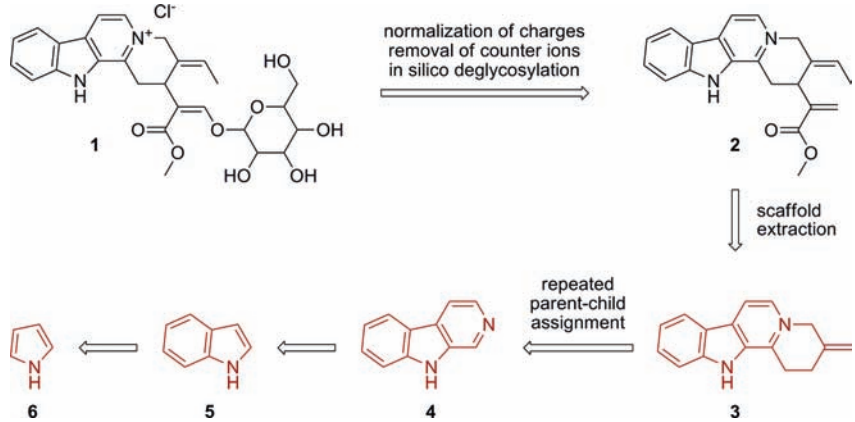


FIGURE 1. Scaffold−substituent analogy between small molecules and proteins. Small molecule substituents and amino acid residues are shown as colored circles. The small molecule scaffold determines the spatial orientation of the substituents, whereas the protein subfold arranges the amino acid side chains in space. Binding occurs when compatible substituents (circles with identical color) match in their spatial position so they can interact.

SCHEME 1. Deconstruction of the Tetrahydrogeissoschizine Derivative **1** According to the SCONP Rule Set[a]



[a] Structures depicted in red are scaffolds of NPs in the DNP, which were used to construct a SCONP tree branch.

relationship between the amino acid side chains in ligand binding sites and NP substituents. Based on the hypothesis that small molecules with similar scaffolds will bind to protein domains with similar ligand-sensing subfolds, we have developed structure-based approaches to chart chemical and biological space that allow us to exploit structural links between proteins and NPs. These approaches are used to identify biologically prevalidated starting points in chemical space for the generation of small molecule libraries in order to find new ligands for particular protein domains.

Charting Natural Product Space by Scaffold Trees: Structural Classification of Natural Products (SCONP)

In order to chart NP space by a hierarchical, structural classification of NP scaffolds,[15] the CRC dictionary of natural products (DNP, version 02/05), with 190 939 entries, was initially processed to remove empty entries, counterions, and stereochemical information, leading to 171 045 structures. Since the majority of historically developed small molecule inhibitors and drugs are based on cyclic structures, further analysis focused on NPs containing rings (154 428 molecules). Because glycosidic moieties often primarily serve to modify the solubility or pharmacokinetic properties of NPs, the resulting NPs were deglycosylated in silico, leading to 149 513 ring-containing aglycons. Subsequently, the NP scaffolds, defined as the ring systems including linker chains between rings and exocyclic double bonds, were extracted by removal of all noncyclic ring substituents (The extraction of scaffold **3** from NP **1** is depicted in Scheme 1 as a representative example). This resulted in the generation of 24 891 unique NP scaffolds.

Hierarchical organization according to ring number included the stepwise deconstruction of larger "child" scaffolds into smaller "parents". Assignment of only one parent scaffold to each more complex "child" was assured; for example, parent scaffolds are always substructures of child scaffolds and parents were always selected to retain the larger scaffolds with a maximum number of heteroatoms. Furthermore, only scaffolds were retained that actually represent entries in the DNP. The resulting parent−child relationships were displayed in a tree diagram (Figure 2), clearly depicting the structural genealogies of NP scaffolds. Scheme 1 shows the deconstruction of NP scaffold **3** into parent scaffolds **4**−**6**.

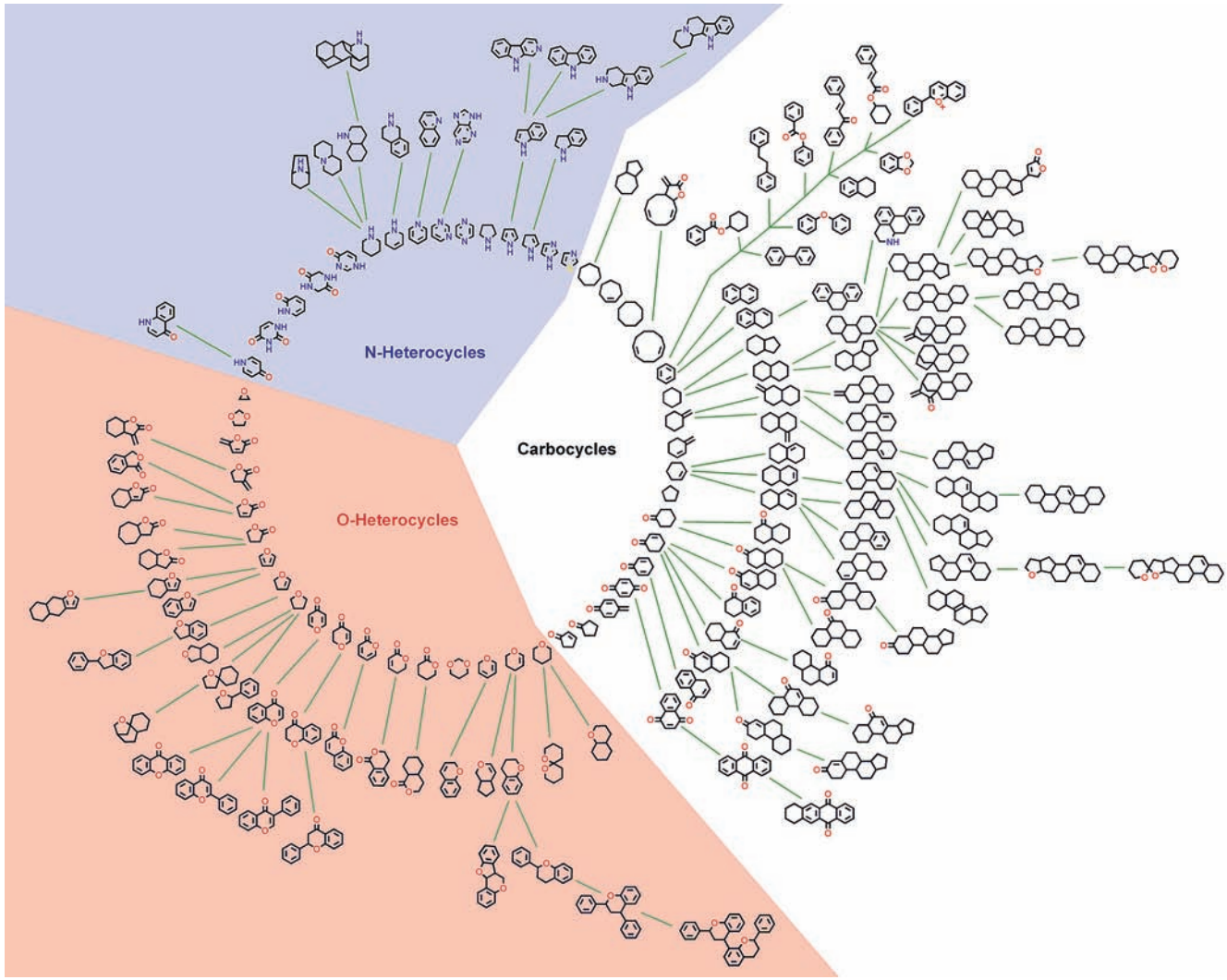


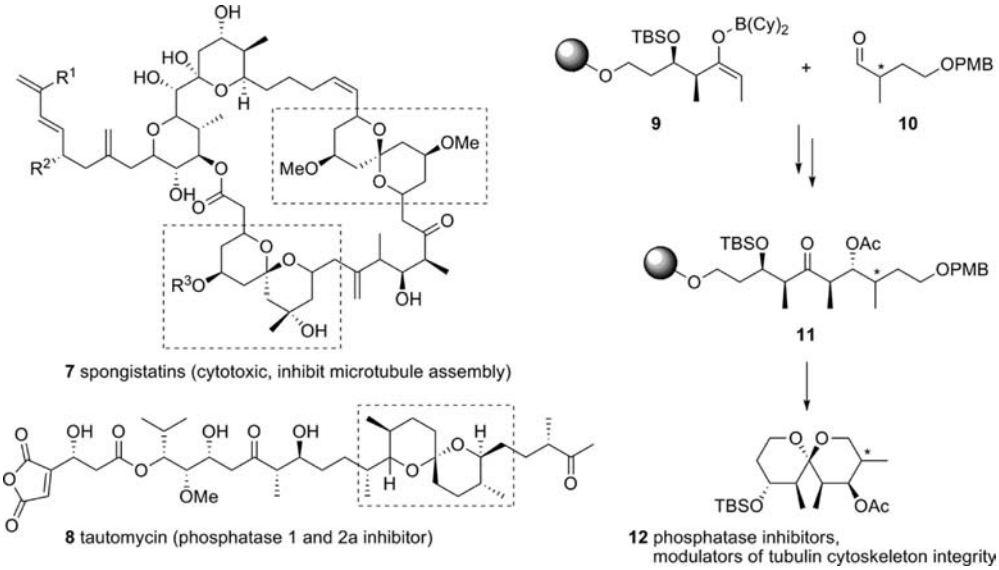
FIGURE 2. The SCONP tree, generated from the Dictionary of Natural Products. For clarity, only scaffolds that represent at least 300 (0.2%) of the analyzed structures are shown. Figure was reproduced from ref 15. Copyright 2005 National Academy of Sciences.

Analysis of the scaffold tree showed that more than half of the NPs in the DNP contain two to four rings. The calculated volumes of the NP scaffolds containing two to four rings range between 150 and 500 Å^3. By comparison, a statistical evaluation of ca. 18 000 protein cavities by Klebe et al. revealed that their volumes are in the range of 300 to 800 Å^3.[16] Thus, compounds containing two to four ring systems have the right size to further accommodate substituents and still match the size of protein cavities. In addition, an analysis of ca. 30 000 drugs from the World Drug Index (WDI) revealed that these molecules have volumes comparable to the scaffolds of two- to four-ring-containing NPs. Analysis of the removed substituents allows the delineation of substituents chosen most often by nature.[17] These results can be used to generate hypotheses for the synthesis of NP-inspired compound collections enriched with bioactive molecules.

Brachiating through the Natural Product Tree: Natural Product Inspired Compound Collections

Scaffolds in the SCONP tree have been selected by nature in evolution and can be considered biologically prevalidated. Consequently, libraries designed on the basis of scaffolds commonly found in NPs most likely display biological activity at comparatively high frequency. In our experience, corresponding medium-sized compound collections of 200−500 members often suffice to find novel bioactive compounds typically with hit rates of 0.1−1.5%. However, given the diversity of amino acid side chains in ligand-sensing cores of protein binding domains, natural product-inspired compound libraries with sufficient substituent and stereochemical diversity are needed to delineate structure−activity relationships (SAR) and binding modes. The synthesis of such diverse compound libraries

SCHEME 2. Biologically Active Spiro[5.5]ketal-Based NPs and the Synthesis of Biologically Active Simplified Analogs **12**



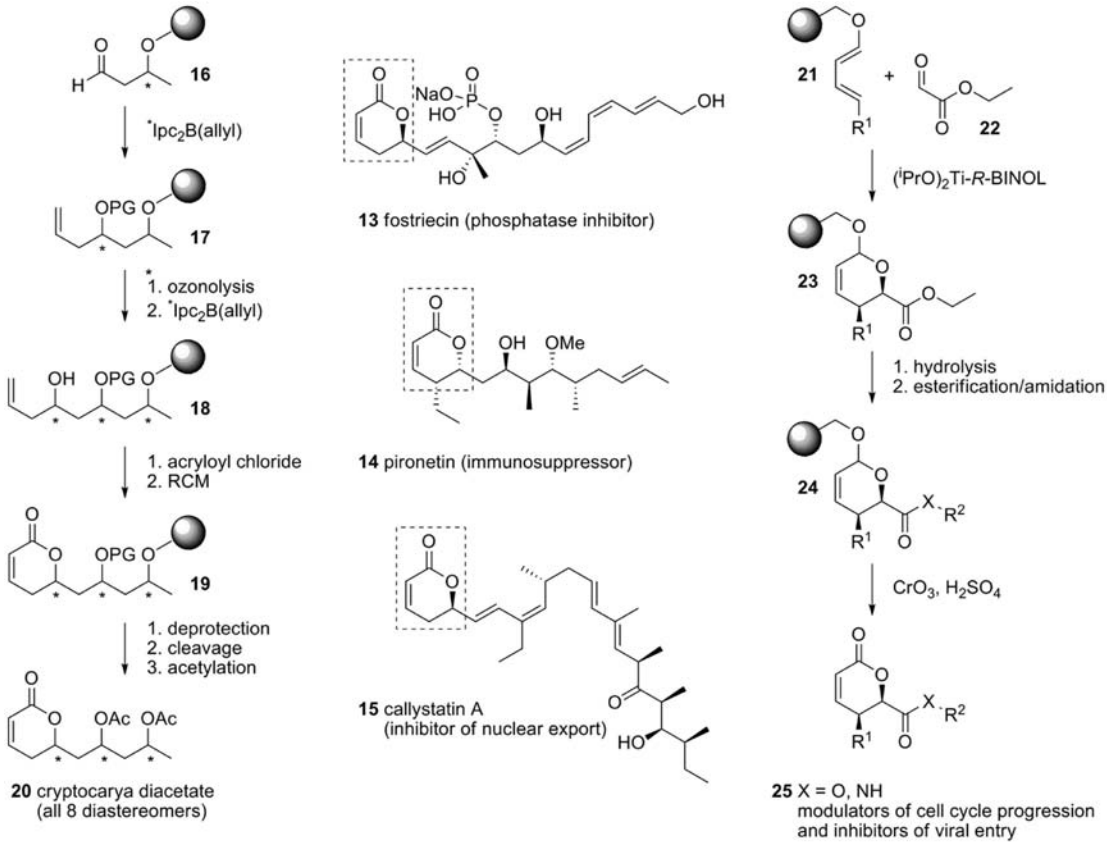
requires the availability of versatile reactions, preferably with full control of chemo-, regio-, and stereochemistry during multistep reaction sequences.

Since spiro[5.5]ketals form a class of naturally abundant oxygen heterocycles with various biological activities (e.g., spongistatins **7** and tautomycin **8**, Scheme 2), we generated a focused compound library employing an asymmetric aldol reaction with immobilized boron enolates (like **9**) to obtain a set of 33 spiroketals.[18,19] Another 251 spiroketals were synthesized using a double intramolecular hetero-Michael reaction on solid support.[20] Biochemical and cell-based assays revealed that the spiroketal library contained new phosphatase inhibitors and modulators of the tubulin cytoskeleton.[18] The biological prevalidation of spiroketal scaffolds was further proven by Ley et al., who designed and synthesized NP-inspired spiroketals as inducers of apoptosis.[21]

Since NPs containing α,β-unsaturated δ-lactones display a wide range of biological activities (see **13**−**15** in Scheme 3 for selected examples), we synthesized a set of 50 NP-like α,β-unsaturated δ-lactones with 1,3-polyol side chains using a sequential ozonolysis/asymmetric allyl boration strategy on solid phase followed by acryloylation and ring-closing metathesis.[22] To demonstrate the stereochemical versatility of the methodology, all eight stereoisomers of the NP cryptocarya diacetate **20** were synthesized (Scheme 3). An additional set of 50 α,β-unsaturated δ-lactones **25** were made using an oxa-Diels−Alder reaction between ethyl glyoxylate **22** and immobilized electron-rich 1-alkoxydienes **21** as the key step.[23] Through the evaluation of these compounds in cell-based assays new modulators of cell cycle progression and inhibitors of viral entry into cells were identified.[23]

Brachiation is a form of arboreal locomotion in which primates swing from branch to branch using only their arms. By analogy, we investigated whether brachiation along the lines of biological prevalidation represented by the branches of the SCONP tree can be used for compound library design via structure simplification, while conserving the kind of bioactivity. This hypothesis was initially investigated by the synthesis and evaluation of a set of decalins based on the natural 11β-hydroxysteroid dehydrogenase (11βHSD) ligand glycyrrhetinic acid **42** (vide infra). In another example, simplified analogues of yohimbine alkaloids were identified as inhibitors of the dual specificity phosphatase Cdc25A (Figure 3).[24] Based on SCONP tree analysis, a diverse set of 450 indoloquinolizidines **27**[25] and 188 2,3-disubstituted indole derivatives based on indomethacin **28** were synthesized on the solid phase[26] and screened for inhibition of a panel of phosphatases, including Cdc25A. The screen revealed that bioactivity can indeed be conserved within branches of the SCONP tree: both libraries contained Cdc25A inhibitors with inhibitory activities comparable to the yohimbine alkaloids.[24] Moreover, the libraries also contained the first potent inhibitors of the mycobacterial tyrosine phosphatase MptpB, a possible target in the treatment of tuberculosis infections. Further biological evaluation revealed potent inducers of apoptosis in cancer cell lines[27] and inhibitors of angiogenesis related kinases.[26] Inspired by the macrolines, a NP family of cycloocta[*b*]indoles with multiple biological activities, an additional library of 120 bridged tetrahydro-β-carbolines was synthesized stereoselectively on the solid phase.[28,29] The fact that this library also contains several selective MptpB inhibitors further validates the brachiation approach.

SCHEME 3. Biologically Active α,β-Unsaturated δ-Lactone-Based NPs and the Synthesis of Bioactive Simplified Analogs



Charting Protein Space: Protein Structure Similarity Clustering (PSSC)

The SCONP approach provides a purely structure-based classification of natural protein ligands. To simultaneously chart the structure space defined by the ligand-sensing cores of proteins, protein structure similarity clustering (PSSC) was developed.[30] PSSC calls for complementarity between protein scaffold (subfold) and chemical scaffold for binding, such that ligand binding sites with structurally similar subfolds should also bind ligands with similar chemical scaffolds. While a single given compound might not bind all members of a PSSC with given subfold because of possible mismatches between small molecule substituents and individual amino acid side chains, a compound collection based on a prevalidated scaffold with sufficient substituent diversity could very well contain binders for several cluster members. The PSSC approach is related to Sternberg's concept of protein supersites, reflecting binding site similarity in the absence of sequence homol-

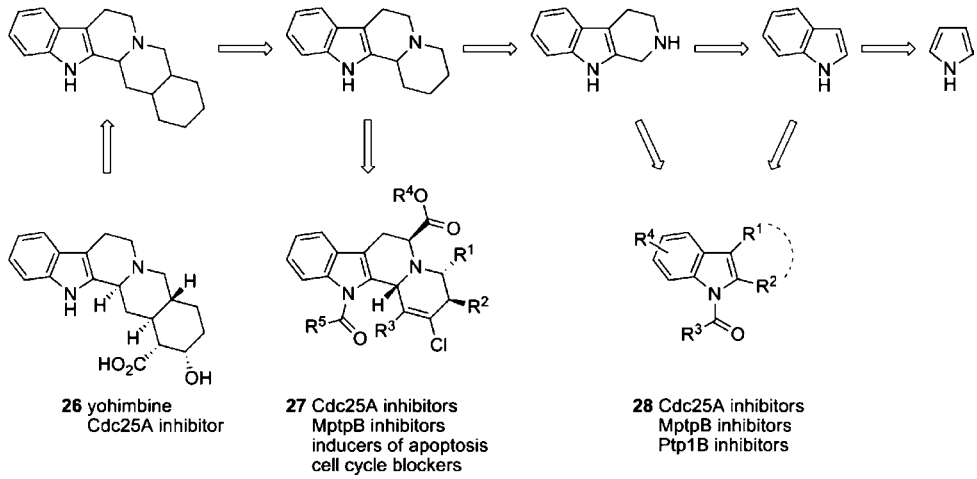


FIGURE 3. Brachiation along the yohimbine branch of the SCONP tree.

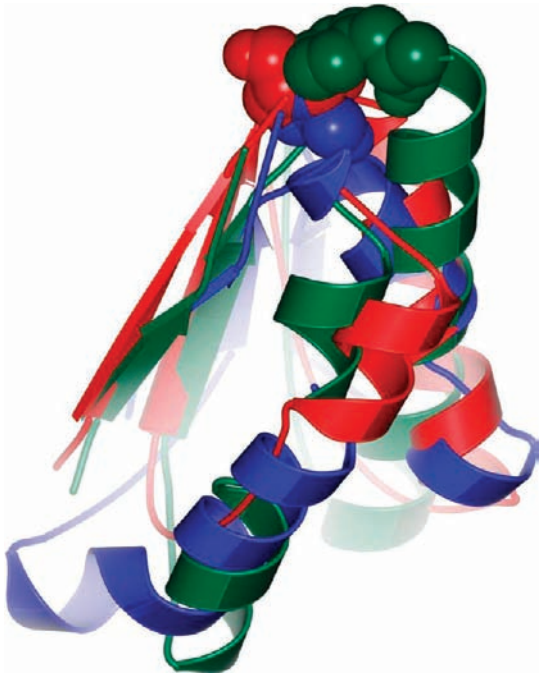


FIGURE 4. Superimposed catalytic sites of Cdc25A (red), 11βHSD1 (green), and AChE (blue). The key catalytic residues, Cys-430 (Cdc25A), Tyr-183 (11βHSD1), and Ser-200 (AChE), are shown in space-filling representation.

ogy.[31] Quinn et al. pointed out that biosynthetic enzymes that bind similar NP intermediates might also be targeted by similar inhibitors and that this observation can be translated into the identification of inhibitors of structurally related proteins.[32]

In the initial PSSC approach, the C_α trace of a protein of interest was compared with the C_α traces of all other proteins in the Protein Data Bank (PDB). From the list of similar proteins, pharmaceutically relevant superfamilies with only low sequence similarity (up to 20% sequence identitiy) were selected. Subsequently, the catalytic cores of the hits and their superfamily members were defined, superimposed, and visually inspected to define protein clusters.

For instance, dual specificity phosphatase Cdc25A, acetylcholine esterase (AChE), and the two isoforms of 11β-hydroxysteroid dehydogenase (11βHSD1 and 11βHSD2) were assigned to one cluster (Figure 4), and a library of 147 γ-hydroxybutenolides and α,β-unsaturated γ-lactones inspired by the natural Cdc25A inhibitor dysidiolide **37** were synthesized. Apart from new Cdc25A inhibitors **38** that were significantly more potent than dysidiolide, evaluation of this compound collection afforded several 11βHSD1 and 11βHSD2 inhibitors **39**.

Subsequently, the PSSC approach was combined with brachiation through the SCONP tree. Stepwise simplification of the pentacyclic scaffold of natural 11βHSD1and 11βHSD2 ligand glycyrrhetinic acid **42** led to the bicyclic dehydrodecalines. Since dysidiolide incorporates a similar dehydrodecaline core, a diverse collection of 483 dehydrodecalines was generated using a combination of solution-phase and solid-phase synthesis.[33] This library included several AChE inhibitors and three highly potent, isoenzyme-specific 11βHSD1 inhibitors endowed with cellular activity (**40** and **41**; Figure 5).[15]

BIOS beyond Natural Products: Hunting in the Scaffold Tree

NPs populate areas of biologically relevant chemical space selected in evolution, yet they do not cover these areas completely, and numerous non-natural biologically relevant compounds have been found in pharmaceutical research. The BIOS concept calls for biological relevance, not necessarily occurrence in nature. Notably, the successful structural simplification of the morphine scaffold[34] cannot be delineated from the SCONP tree because there is no NP with a four-ring combination derived from morphine. However, sequential simplification of the natural alkaloid has led to bicyclic (e.g., tramadol), tricyclic (e.g., pentazocine), and tetracyclic (e.g., levorphanol) morphine analogues, some of which are marketed drugs. Thus, it proved necessary to develop methods for complementation of the SCONP tree with further biologically active compounds and to analyze any data set of molecules in order to identify and explore uncharted parts of chemical space.

To this end, a set of 13 new rules for the stepwise deconstruction of complex scaffolds into less complex structures was introduced.[35] The new classification is independent of the used data set, such that a particular scaffold will always lead to the same branch (in the SCONP tree, assignment of parent scaffolds is restricted to scaffolds representing molecules in the analyzed database). Because such scaffold trees should ultimately inspire the design and synthesis of new compound libraries, the 13 rules also reflect synthetic and medicinal chemistry rationales. According to rule 1, for example, three-membered heterocyclic rings like epoxides are pruned first, since they are usually installed at the end of a synthetic sequence, for example by oxidation of a double bond. These rules should lead to parents with maximum rigidity, since these are more likely to possess unique interaction patterns.

Scaffold trees as, for example, depicted in Figure 2 are static, do not allow the analysis of the whole data set, do not code for different bioactivity, and need experts for creation in a labor-intensive, time-consuming process. To overcome these drawbacks, a highly intuitive, interactive, and easily applicable tool for the exploration and navigation of data sets, called Scaffold Hunter (SH), was developed.[36] This program, which is freely available from www.scaffoldhunter.com, reads compound data sets

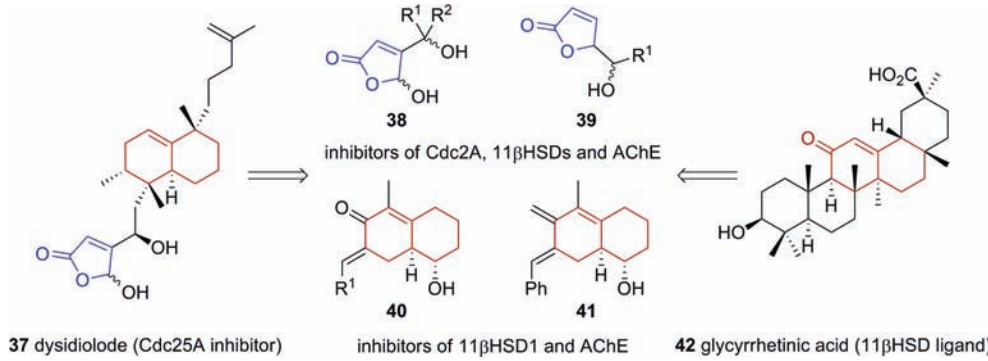


FIGURE 5. NP-inspired inhibitors of the Cdc25A PSSC cluster members.

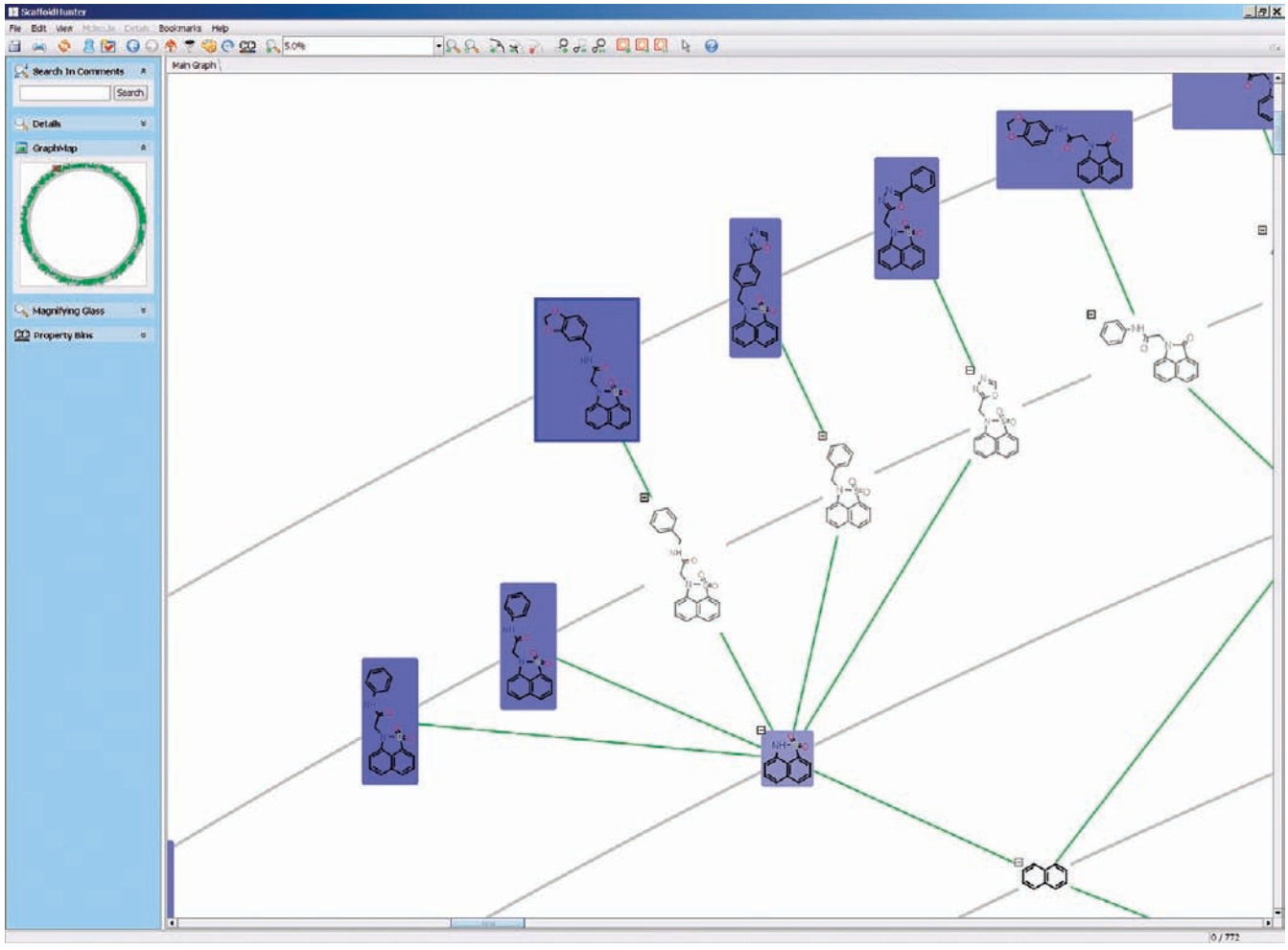


FIGURE 6. Scaffold Hunter screenshot showing a subtree with nodes colored according to the median activity of the molecules represented by each scaffold. A darker color indicates higher activity.

and creates parent scaffolds according to the 13 predefined rules mentioned above. It generates and displays scaffold trees and maps the chemical structures on the nodes, while concentric circles mark the different hierarchy levels (Figure 6). SH can be used to annotate properties like bioactivity values to the scaffolds, to filter the data according to all properties stored in the database, to navigate through the scaffold tree, and to export data, including high-resolution images. The use of Scaffold Hunter does not require expert training in computer science, runs on average PC platforms, and allows the addition or deletion of rules for scaffold deconstruction.

Notably, the Scaffold Hunter identifies virtual scaffolds that do not represent molecules in the used data set but are required to complete the tree. Since compounds representing these scaffolds could very well share bioactivity properties with their parent or child scaffolds, analysis with Scaffold Hunter may result in the

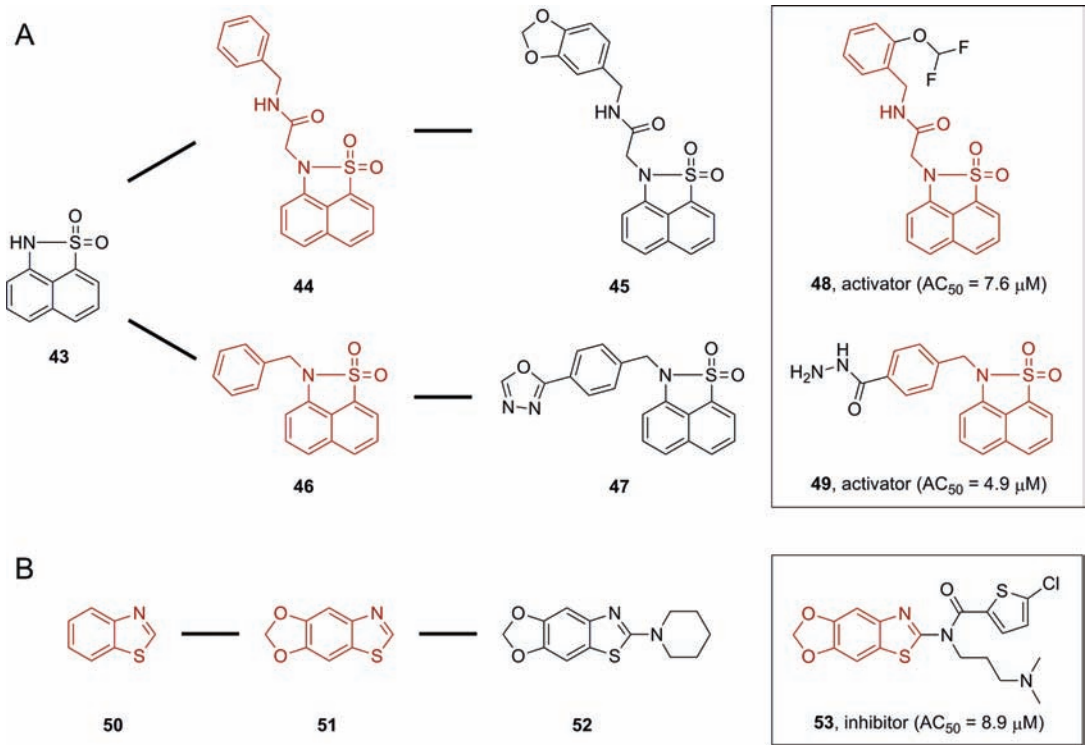


FIGURE 7. Scaffold tree branches containing PK activators and inhibitors. Virtual scaffolds are depicted in red. (A) Compounds **48** and **49** are new PK activators. (B) Compound **53** is a new PK inhibitor.

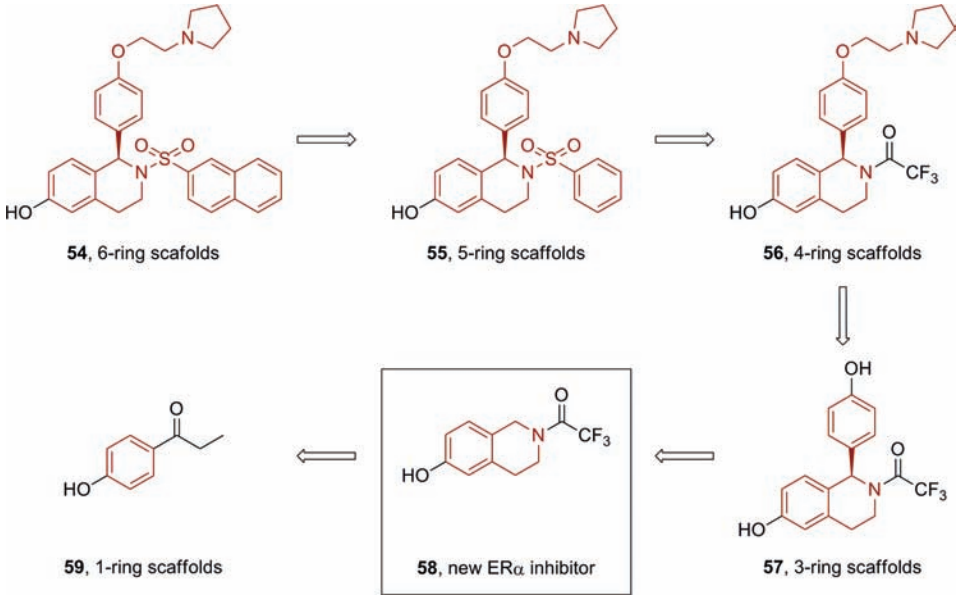


FIGURE 8. Scaffold tree resulting from bioactivity-guided scaffold deconstruction. The scaffolds of ERα agonists **54**−**59** are depicted in red. Compound **58**, which was based on a nonannotated scaffold, was identified as a new ERα agonist.

identification of new scaffold classes. To demonstrate this possibility, we analyzed the PubChem pyruvate kinase (PK) screen of 51 415 unique molecules, resulting in a scaffold tree consisting of 35 868 scaffolds, including 8684 virtual scaffolds distributed over 767 branches. Subsequent filtering for scaffolds displaying activation or inhibition of PK resulted in a focused scaffold tree. A collection of 107 compounds, representing four virtual scaffolds on branches with actives (e.g., **43** and **50**), was purchased and tested, leading to the identification of new low micromolar PK activators and inhibitors. Notably, virtual scaffolds from branches with inhibitors gave new inhibitors (e.g., **53**, Figure 7), whereas virtual scaffolds from branches with activators gave new activators (e.g., **48** and **49**). Furthermore, screening of an additional 88 compounds representing six virtual scaffolds from

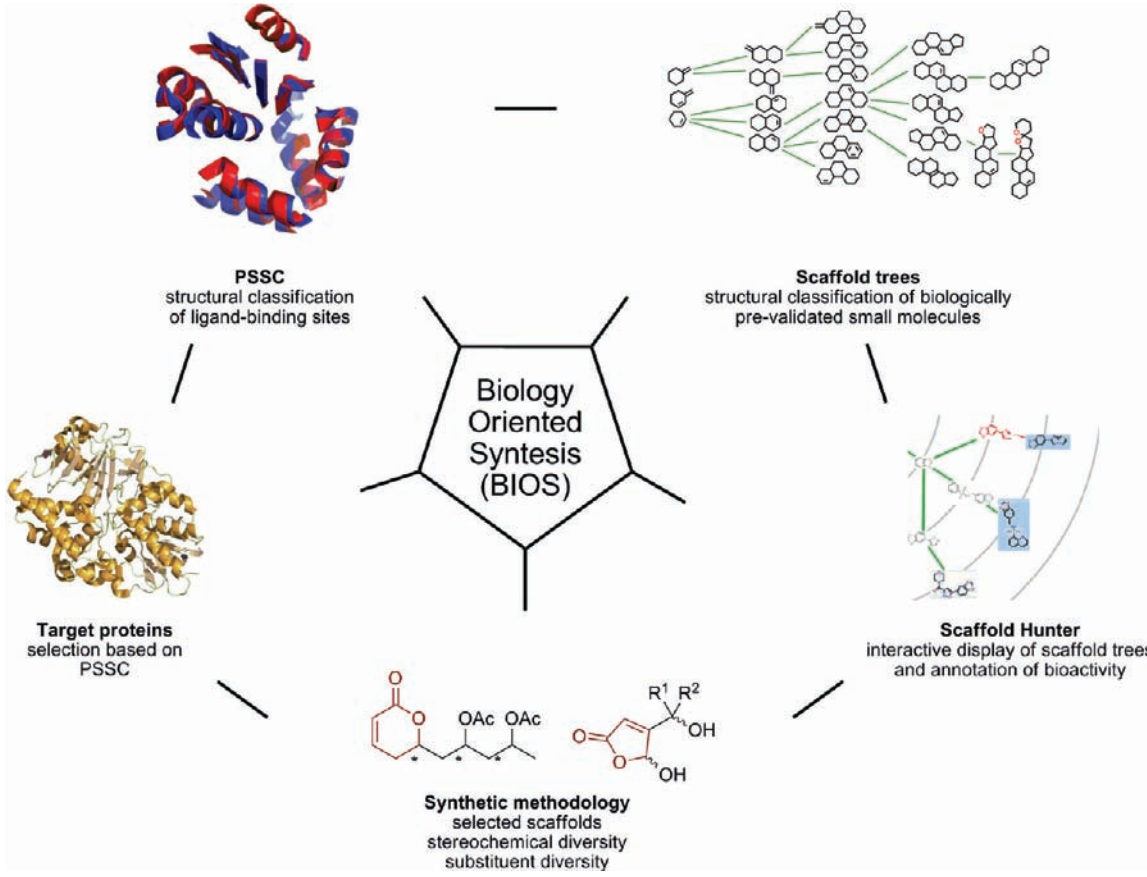


FIGURE 9. Biology-oriented synthesis integrates cheminformatic, bioinformatic, and synthetic tools to find new bioactive small molecules.

branches containing mainly inactives did not give any modulators of PK activity.[36]

The virtual scaffolds represent "holes" in chemical space that may not be shared by the corresponding biological space (see the morphine example above, suggesting the development of a method that simultaneously maps chemical space and associated biological space). In order to map chemical space guided by bioactivity, molecules were hierarchically arranged along lines of biological relevance, using type of bioactivity rather than a chemistry-driven rule set as key selection criterion for structural simplification during tree construction. Scaffolds were defined as parent−child pairs if they displayed a common type of bioactivity. The analysis revealed that brachiation along the resulting bioactivity-guided branches is a fairly widespread phenomenon: for all major protein target classes of current pharmaceutical interest, branches of at least five hierarchy levels were identified. In a prospective application of the bioactivity-guided scaffold trees, analysis of the branches containing 5-lipoxygenase (5-LOX) inhibitors and estrogen receptor α (ERα) activity modulators (**54**−**57**, **59**) revealed scaffolds not annotated with biological activity against the protein of interest in both branches. Design and synthesis of compounds based on these nonannotated scaffolds gave new classes of 5-LOX inhibitors and ERα agonists (e.g., **58**, Figure 8).[37]

Conclusions and Outlook

BIOS combines structure-based hypothesis-generating tools for the mapping of chemical and biological space in order to identify and explore biologically relevant chemical space (Figure 9). On the one hand, PSSC is used for the mapping of protein ligand-sensing cores and the identification of ligand binding sites with similar subfolds. On the other hand, scaffold trees of natural products, for example, the SCONP tree, classify the scaffolds found in the evolutionarily selected small molecule protein ligands. The classification of NP scaffolds in the original SCONP tree is based on occurrence in nature. In an extension of this initial approach, the unbiased generation of scaffold trees with a new rule set implemented in Scaffold Hunter and the introduction of virtual scaffolds render the method independent of the data set. Notably, virtual scaffolds provide novel opportunities to find new ligand types for a particular protein target.

The rule set currently implemented in Scaffold Hunter is based on 40 different chemical properties. However, parent−child relationships of scaffolds may be assigned on the basis of a particular biological activity, which allows navigation through chemical

space along lines of biological prevalidation. Future implementation of the bioactivity-guided methods for scaffold tree generation in Scaffold Hunter might enable the identification of compound classes that target multiple proteins and therefore the development of multitargeted bioactive molecules.

Since both protein folds and NP scaffolds are highly conserved in nature, ligand binding sites with similar subfolds should bind small molecules with similar scaffolds. Selectivity would then be ensured by the diversity of amino acid sequences in ligand binding sites that has to be matched by the substituent patterns of NP-inspired compound collections. Therefore, BIOS-derived small molecule scaffold trees provide hypothesis-generating tools for the design of new compound collections for chemical biology and medicinal chemistry research. Scaffold tree generation identifies new scaffold types for bioactive small molecules, whereas PSSC allows the identification of new potential biological targets of small molecules. Because ligand binding ultimately depends on interactions of amino acid residues in proteins and substituents of small molecule scaffolds, focused small molecule libraries containing 200−500 members are typically required to ensure sufficient stereochemical and substituent diversity. The development of versatile synthetic methodology in solution and on solid phase is instrumental in the synthesis of these compound collections. Various syntheses of natural product inspired compound collections synthesized by us as well as others have demonstrated that currently available synthetic methodology in principle is powerful enough to develop such libraries with high scaffold, substituent, and stereochemical diversity.[38−40] Notably, multicomponent and domino reactions have also proven highly efficient for the generation of diverse libraries of complex NP-inspired compounds.[41,42]

Evaluation of NP-inspired compound collections in biochemical and biological assays in our experience has yielded relatively high hit rates (typically 0.1−1.5%), and we have identified new small molecule classes that target a range of proteins and display diverse biological activities. We have also shown that brachiation through scaffold trees can lead to simplified NP analogues with retained biological activity. Although in some cases bioactivity is retained over many levels of a scaffold tree (e.g., morphine analogues), it should be mentioned that this is not automatically true for all scaffold trees. Many NP-inspired compound libraries also contained selective modulators of protein function. However, initial BIOS libraries may contain compounds of limited selectivity. Thus, it will often be necessary to improve selectivity and potency of protein ligands resulting from BIOS in subsequent optimization studies. Small molecule probes resulting from BIOS have allowed us to identify cellular targets of NPs and to selectively perturb proteins involved in signal transduction pathways in order to study their function.[43,44]

Possible future applications of BIOS in the search for new protein−protein interaction (PPI) inhibitors might be enabled by the increasing structural knowledge about PPIs through X-ray crystallography and NMR studies. Since PPIs often occur through an induced fit mechanism, the combination of PSSC and dynamic simulation of protein structure might be instrumental in such an approach.[45]

A major goal of Chemical Biology is to identify chemical tools for the study of biological phenomena such as the role of proteins in complex, dynamic biological systems through the selective perturbation of their function. High hit rates within NP-inspired compound collections, the identification of small molecules endowed with cellular activity, and the development of novel chemical probes for target identification and selective modulation of cellular protein function underline the potential of BIOS to contribute to this ambitious goal.

The BIOS concept was conceptualized and realized with the help of many (former) Ph.D. students, postdoctoral researchers, group leaders, and collaborators whose names can be found in the reference section. Dr. Stefan Wetzel is gratefully acknowledged for stimulating discussions and for providing graphics.

BIOGRAPHICAL INFORMATION

Robin S. Bon completed his Ph.D. in organic chemistry at the Vrije Universiteit Amsterdam in 2007 with Prof. Romano Orru and carried out postdoctoral research, supported by an Alexander von Humboldt fellowship, with Prof. Herbert Waldmann at the MPI of Molecular Physiology, Dortmund. Since November 2009, Robin is a senior research fellow at the University of Leeds. His research focuses on the development of small molecule modulators of protein function and tools for biochemical assays.

Herbert Waldmann completed his Ph.D. in organic chemistry at the University of Mainz in 1985 with Prof. Kunz and carried out postdoctoral research with Prof. G. Whitesides at Harvard University. Following positions as Professor of Organic Chemistry at the Universities of Bonn (from 1991) and Karlsruhe (from 1993), he became Director at the MPI of Molecular Physiology in Dortmund and Professor of Organic Chemistry at the University of Dortmund in 1999. His research interests lie in the study of chemical biology with small-molecule and protein probes and the integration of cheminformatics, bioinformatics, and synthetic methodology to explore biologically relevant chemical space.

FOOTNOTES

*To whom correspondence should be addressed. E-mail: herbert.waldmann@mpi-dortmund.mpg.de.

†Current address: School of Chemistry, University of Leeds, Leeds LS2 9JT, U.K.

REFERENCES

1 Zamir, E.; Bastiaens, P. I. Reverse engineering intracellular biochemical networks. *Nat. Chem. Biol.* **2008**, *4*, 643–647.

2 Dobson, C. M. Chemical space and biology. *Nature* **2004**, *432*, 824–828, and references cited therein.

3 International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945.

4 Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **2004**, *432*, 855–861.

5 Klebe, G. Virtual ligand screening: Strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.

6 Hermann, J. C.; Marti-Arbona, R.; Fedorov, A. A.; Fedorov, E.; Almo, S. C.; Shoichet, B. K.; Raushel, F. M. Structure-based activity prediction for an enzyme of unknown function. *Nature* **2007**, *448*, 775–779.

7 Rupp, M.; Schroeter, T.; Steri, R.; Zettl, H.; Proschak, E.; Hansen, K.; Rau, O.; Schwarz, O.; Müller-Kuhrt, L.; Schubert-Zsilavecz, M.; Müller, K.-R.; Schneider, G. From machine learning to natural product derivatives that selectively activate transcription factor PPAR γ. *ChemMedChem* **2009**, *5*, 191–194.

8 Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today* **2006**, *11*, 1046–1053.

9 Rosén, J.; Gottfries, J.; Muresan, S.; Backlund, A.; Oprea, T. I. Novel chemical space exploration via natural products. *J. Med. Chem.* **2009**, *52*, 1953–1962.

10 Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* **2007**, *70*, 461–477.

11 Hert, J.; Irwin, J. J.; Laggner, C.; Keiser, M. J.; Shoichet, B. K. Quantifying biogenic bias in screening libraries. *Nat. Chem. Biol.* **2009**, *5*, 479–483.

12 Evans, B. E.; Rittle, K. E.; Bock, M. G.; DiPardo, R. M.; Freidinger, R. M.; Whitter, W. L.; Lundell, G. F.; Veber, D. F.; Anderson, P. S.; Chang, R. S.; Lotti, V. J.; Cerino, D. J.; Chen, T. B.; Kling, P. J.; Kunkel, K. A.; Springer, J. P.; Hirshfield, J. Methods for drug discovery: Development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* **1988**, *31*, 2235–2246.

13 Yeats, C. A., Orengo, C. A. *In* Handbook of Proteins; Cox, M. M., Philips, G. N. J., Eds.; John Wiley & Sons: Chichester, U.K., 2007; pp 23−32.

14 Sadreyev, R. I.; Grishin, N. V. Exploring dynamics of protein structure determination and homology-based prediction to estimate the number of superfamilies and folds. *BMC Struct. Biol.* **2006**, *6*, 6.

15 Koch, M. A.; Schuffenhauer, A.; Scheck, M.; Wetzel, S.; Casaulta, M.; Odermatt, A.; Ertl, P.; Waldmann, H. Charting biologically relevant chemical space: A structural classification of natural products (SCONP). *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17272–17277.

16 Kuhn, D.; Weskamp, N.; Schmitt, S.; Hullermeier, E.; Klebe, G. From the similarity analysis of protein cavities to the functional classification of protein families using cavbase. *J. Mol. Biol.* **2006**, *359*, 1023–1044.

17 Wetzel, S.; Schuffenhauer, A.; Roggo, S.; Ertl, P.; Waldmann, H. Cheminformatic analysis of natural products and their chemical space. *Chimia* **2007**, *61*, 355–360.

18 Barun, O.; Kumar, K.; Sommer, S.; Langerak, A.; Mayer, T. U.; Muller, O.; Waldmann, H. Natural product-guided synthesis of a spiroacetal collection reveals modulators of tubulin cytoskeleton integrity. *Eur. J. Org. Chem.* **2005**, 4773–4788.

19 Barun, O.; Sommer, S.; Waldmann, H. Asymmetric solid-phase synthesis of 6,6-spiroketals. *Angew. Chem., Int. Ed.* **2004**, *43*, 3195–3199.

20 Sommer, S.; Kühn, M.; Waldmann, H. Solid-phase synthesis of [5.5]-spiroketals. *Adv. Synth. Catal.* **2008**, *350*, 1736–1750.

21 Milroy, L. G.; Zinzalla, G.; Loiseau, F.; Qian, Z. Z.; Prencipe, G.; Pepper, C.; Fegan, C.; Ley, S. V. Natural-product-like spiroketals and fused bicyclic acetals as potential therapeutic agents for B-cell chronic lymphocytic leukaemia. *ChemMedChem* **2008**, *3*, 1922–1935.

22 Umarye, J. D.; Lessmann, T.; Garcia, A. B.; Mamane, V.; Sommer, S.; Waldmann, H. Biology-oriented synthesis of stereochemically diverse natural-product-derived compound collections by iterative allylations on a solid support. *Chem.—Eur. J.* **2007**, *13*, 3305–3319.

23 Lessmann, T.; Leuenberger, M. G.; Menninger, S.; Lopez-Canet, M.; Muller, O.; Hummer, S.; Bormann, J.; Korn, K.; Fava, E.; Zerial, M.; Mayer, T. U.; Waldmann, H. Natural product-derived modulators of cell cycle progression and viral entry by enantioselective oxa Diels-Alder reactions on the solid phase. *Chem. Biol.* **2007**, *14*, 443–451.

24 Noren-Muller, A.; Reis-Correa, I., Jr.; Prinz, H.; Rosenbaum, C.; Saxena, K.; Schwalbe, H. J.; Vestweber, D.; Cagna, G.; Schunk, S.; Schwarz, O.; Schiewe, H.; Waldmann, H. Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10606–10611.

25 Corrêa, I. R. J.; Nören-Müller, A.; Ambrosi, H.-D.; Jakupovic, S.; Saxena, K.; Schwalbe, H.; Kaiser, M.; Waldmann, H. Identification of inhibitors for mycobacterial protein tyrosine phosphatase B (MptpB) by biology-oriented synthesis (BIOS). *Chem.—Asian. J.* **2007**, *2*, 1109–1126.

26 Rosenbaum, C.; Baumhof, P.; Mazitschek, R.; Muller, O.; Giannis, A.; Waldmann, H. Synthesis and biological evaluation of an indomethacin library reveals a new class of angiogenesis-related kinase inhibitors. *Angew. Chem., Int. Ed.* **2004**, *43*, 224–228.

27 Wehner, F.; Noren-Muller, A.; Muller, O.; Reis-Correa, I., Jr.; Giannis, A.; Waldmann, H. Indoloquinolizidine derivatives as novel and potent apoptosis inducers and cell-cycle blockers. *ChemBioChem* **2008**, *9*, 401–405.

28 Nören-Müller, A.; Wilk, W.; Saxena, K.; Schwalbe, H.; Kaiser, M.; Waldmann, H. Discovery of a new class of inhibitors of *Mycobacterium tuberculosis* protein tyrosine phosphatase B by biology-oriented synthesis. *Angew. Chem., Int. Ed.* **2008**, *47*, 5973–5977.

29 Wilk, W.; Noren-Muller, A.; Kaiser, M.; Waldmann, H. Biology-oriented combined solid- and solution-phase synthesis of a macroline-like compound collection. *Chem.—Eur. J.* **2009**, *15*, 11976–11984.

30 Koch, M. A.; Wittenberg, L.-O.; Basu, S.; Jeyaraj, D. A.; Gourzoulidou, E.; Reinecke, K.; Odermatt, A.; Waldmann, H. Compound library development guided by protein structure similarity clustering and natural product structure. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16721–16726.

31 Russell, R. B.; Sasieni, P. D.; Sternberg, M. J. E. Supersites within superfolds. Binding site similarity in the absence of homology. *J. Mol. Biol.* **1998**, *282*, 903–918.

32 McArdle, B. M.; Campitelli, M. R.; Quinn, R. J. A common protein fold topology shared by flavonoid biosynthetic enzymes and therapeutic targets. *J. Nat. Prod.* **2006**, *69*, 14–17.

33 Scheck, M.; Koch, M. A.; Waldmann, H. Synthesis of a dysidiolide-inspired compound library and discovery of acetylcholinesterase inhibitors based on protein structure similarity clustering (PSSC). *Tetrahedron* **2008**, *64*, 4792–4802.

34 Eguchi, M. Recent advances in selective opioid receptor agonists and antagonists. *Med. Res. Rev.* **2003**, *24*, 182–212.

35 Schuffenhauer, A.; Ertl, P.; Roggo, S.; Wetzel, S.; Koch, M. A.; Waldmann, H. The scaffold tree - Visualization of the scaffold universe by hierarchical scaffold classification. *J. Chem. Inf. Model.* **2007**, *47*, 47–58.

36 Wetzel, S.; Klein, K.; Renner, S.; Rauh, D.; Oprea, T. I.; Mutzel, P.; Waldmann, H. Interactive exploration of chemical space with Scaffold Hunter. *Nat. Chem. Biol.* **2009**, *5*, 581–583.

37 Renner, S.; van Otterlo, W. A.; Dominguez Seoane, M.; Mocklinghoff, S.; Hofmann, B.; Wetzel, S.; Schuffenhauer, A.; Ertl, P.; Oprea, T. I.; Steinhilber, D.; Brunsveld, L.; Rauh, D.; Waldmann, H. Bioactivity-guided mapping and navigation of chemical space. *Nat. Chem. Biol.* **2009**, *5*, 585–592.

38 Kumar, K.; Waldmann, H. Synthesis of natural product inspired compound collections. *Angew. Chem., Int. Ed.* **2009**, *48*, 3224–3242.

39 Nandy, J. P.; Prakesch, M.; Khadem, S.; Reddy, P. T.; Sharma, U.; Arya, P. Advances in solution- and solid-phase synthesis toward the generation of natural product-like libraries. *Chem. Rev* **2009**, *109*, 1999–2060.

40 Nielsen, T. E.; Schreiber, S. L. Diversity-oriented synthesis - towards the optimal screening collection: A synthesis strategy. *Angew. Chem., Int. Ed.* **2008**, *47*, 48–56.

41 Waldmann, H.; Kühn, M.; Lui, W.; Kumar, K. Reagent-controlled domino synthesis of skeletally-diverse compound collections. *Chem. Commun.* **2008**, *10*, 1211–1213.

42 Ulaczyk-Lesanko, A.; Hall, D. G. Wanted, new multicomponent reactions for generating libraries of polycyclic natural products. *Curr. Opin. Chem. Biol.* **2005**, *9*, 266–276.

43 Knoth, T.; Warburg, K.; Katzka, C.; Rai, A.; Wolf, A.; Brockmeyer, A.; Janning, P.; Reubold, T. F.; Eschenburg, S.; Manstein, D. J.; Hubel, K.; Kaiser, M.; Waldmann, H. The Ras pathway modulator melophlin A targets dynamins. *Angew. Chem., Int. Ed.* **2009**, *48*, 7240–7245.

44 Dekker, F. J.; Rocks, O.; Vartak, N.; Menninger, S.; Hedberg, C.; Balamurugan, R.; Wetzel, S.; Renner, S.; Gerauer, M.; Schölermann, B.; Rusch, M.; Kramer, J. W.; Rauh, D.; Coates, G. W.; Brunsveld, L.; Bastiaens, P. I.; Waldmann, H. Small-molecule inhibition of APT1 affects Ras localization and signaling. *Nat. Chem. Biol.* **2010**, *6*, 449−456.

45 Charette, B. D.; Macdonald, R. G.; Wetzel, S.; Berkowitz, D. B.; Waldmann, H. Protein structure similarity clustering: dynamic treatment of PDB structures facilitates clustering. *Angew. Chem., Int. Ed.* **2006**, *45*, 7766–7770.