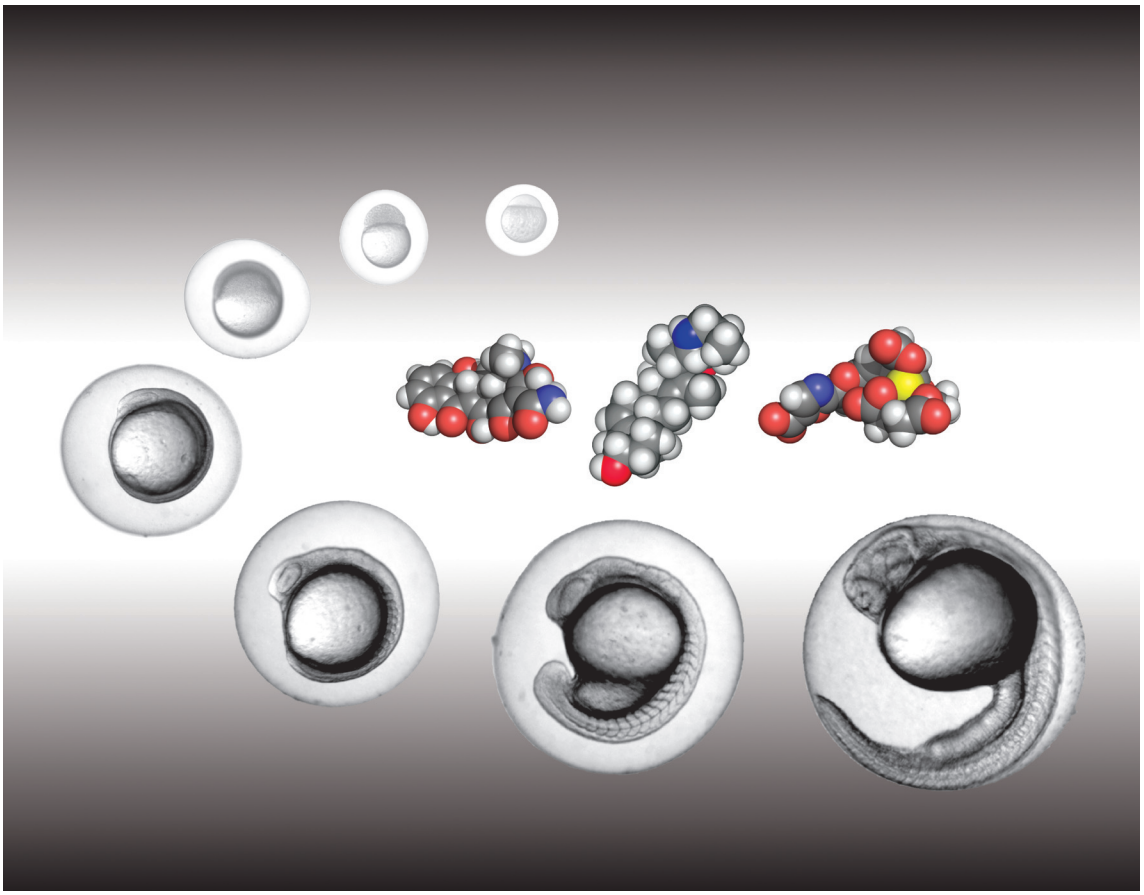
Chem Soc Rev

This article was published as part of the

2008 Chemistry–Biology Interface Issue

Reviewing research at the interface where chemistry meets biology

Please take a look at the full table of contents to access the other papers in this issue



Chemical biology—identification of small molecule modulators of cellular activity by natural product inspired synthesis†

Katja Hübel,[a] Torben Leßmann[a] and Herbert Waldmann*[ab]

Received 31st March 2008
First published as an Advance Article on the web 14th May 2008
DOI: 10.1039/b704729k

The aim of this *tutorial review* is to introduce the reader to the concept, synthesis and application of natural product-inspired compound collections as an important field in chemical biology. This review will discuss how potentially interesting scaffolds can be identified (structural classification of natural products), synthesized in an appropriate manner (including stereoselective transformations for solid phase-bound compounds) and tested in biological assays (cell-based screening as well as biochemical *in vitro* assays). These approaches will provide the opportunity to identify new and interesting compounds as well as new targets for chemical biology and medicinal chemistry research.

Introduction: small molecules in chemical biology research

The completion of the Human Genome Project in 2003 provided researchers with the basic information needed to analyze and ultimately understand human biology, and in the process, 1800 disease genes were uncovered (National Institutes of Health, http://www.genome.gov/10001772). The task lying before us is to group the letters which make up the genome sequences into words, sentences, paragraphs and chapters, ultimately assembling the book of life.[1]

Besides classical genetic methods, the use of small molecules can help to decipher this book of life. This so called chemical genetic approach can be used to investigate complex pathways, the interaction of proteins with each other or cell fate. Furthermore, these small molecules may have a powerful therapeutic value in disease treatment. Thus, the identification of small molecules that are able to modulate cellular functions has become one of the main areas of interest in chemical biology and medicinal chemistry research. Small molecules are low-molecular weight compounds *e.g.* natural products as the reader will learn in the next paragraph.

Compared to classical genetic approaches, the use of small molecules is advantageous for several reasons. Gene deletion of a multifunctional protein will completely remove a specific function from the cell, remove protein–protein interactions and may even lead to the death of the organism under investigation. Alternatively, a selective small molecule can perturb specific protein features without total removal of the target from the cell. Temporal control and reversibility of the inhibition of protein function are other benefits of using small molecule modulators. Effects can be induced by adding a small molecule, and subsequently washing off the modulating molecule results in a return to the non-perturbed state. The

[a] *Max-Planck-Institute of Molecular Physiology, Otto Hahn-Str. 11, 44227 Dortmund, Germany. E-mail: herbert.waldmann@mpi-dortmund.mpg.de; Fax: 0049 231-1332499; Tel: 0049 231-1332401*
[b] *Fachbereich Chemische Biologie, Technische Universität Dortmund, Germany*

† Part of a thematic issue examining the interface of chemistry with biology.

Katja Hübel

Katja Hübel studied biochemistry at the Ruhr-Universität Bochum, Germany. Later, she continued her studies at the Ecole Européenne de Chimie, Polymères et Materiaux in Strasbourg, France. She obtained her PhD in 2005 under the supervision of Prof. Dr K. Bauer at the Max-Planck-Institute of Experimental Endocrinology in Hannover, Germany. Currently, she coordinates the field of "Cellular Assays and Target Identification" in the group of Chemical Biology with Prof. Dr. H. Waldmann at the MPI in Dortmund.

Torben Leßmann

Torben Leßmann was born in Hamburg, Germany in 1976. He studied chemistry at the universities of Tübingen and Heidelberg, where he received his diploma in 2002. He then joined the group of Herbert Waldmann at the Max-Planck-Institute of Molecular Physiology in Dortmund, where he worked on the development of asymmetric reactions on polymer-bound substrates and obtained his PhD in 2007. Currently, he is undertaking postdoctoral studies in the group of Craig J. Forsyth at the Ohio State University.

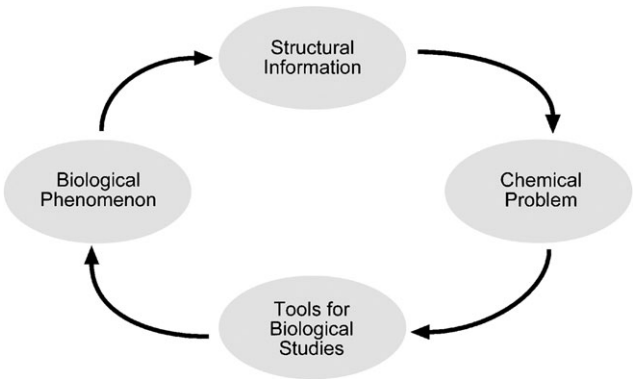


Fig. 1 Interplay between organic synthesis and biology in chemical biology research.[1]

disadvantages may be lack of specificity, cytotoxicity and solubility.[2] A lack of specificity can slow down drug development and can lead to misinterpretation of the observed effects.[3] Small molecules may be toxic to the organism under investigation in a dose-dependent manner. Compounds with low solubility may precipitate when applied to cells or may not penetrate into the tissues of the organism to their site of action. Using diverse experimental and theoretical techniques, chemical biologists have solved challenging problems in biology, ranging from cellular signaling to cell division and neurobiology.[4]

Chemical biological approaches often start by analyzing a biological system or phenomenon of interest (Fig. 1) to deduce information concerning the structure of biomacromolecules (*e.g.* proteins, mini-proteins and peptides) involved in a particular biological phenomenon or the chemical structure of small molecules which interact with these macromolecules. This structural information is then employed to identify unsolved chemical problems, for example, the development of new methods for synthesizing biomacromolecules and low molecular weight compounds, such as natural products and their analogs. Also, the design and synthesis of inhibitors that can be used to perturb and probe biological systems is of major interest. Once methods of accessing the desired compounds have been devised and developed, the newly prepared compounds are employed in appropriately designed biological and/or biochemical experiments to assess their activity. The results gleaned may then give rise to a better understanding of the biological problem(s) in hand and may also highlight new structural features, thus forming the basis for a new round of investigations.[1]

It is important to point out that all criteria of the outlined cycle can be applied and also that the cycle can be accessed at any starting point.

To summarize, molecular interactions and chemical transformations that are at the heart of biology and all biological phenomena which can be analyzed today can be traced back to chemical processes: after all, biology is fundamentally molecular.[1]

In this review, we will give a case study to explain the principles of chemical biology, including (i) an introduction to the concept and advantages of natural product-inspired combinatorial synthesis and (ii) hit identification using biochemical *in vitro* screening or cell-based assays. A short section will also explain how promising small molecules can be prepared for target identification.

Natural products in compound collections/natural product-inspired compound collections

Throughout history, natural products have been used as the active agents in many traditional medical preparations from plants or other biological sources and continue to play an important role as pharmaceuticals. Additionally, many drugs on the market either are or have been derived from natural products, allowing this compound class to have an overwhelming influence on the success of the pharmaceutical industry. In fact, roughly half of the drugs sold today are derived from natural products.[5] In this context, natural products are defined as low-molecular-weight chemical compounds (small molecules) that are naturally synthesized by biological organisms.[6]

The interest in collections of natural compounds as a source of pharmaceutical innovation diminished with the introduction of combinatorial techniques to synthesize compounds. Large compound collections became available to the pharmaceutical industry which promised to yield many new lead structures and drug candidates. Combinatorial methods also seemed to be best suited to meet the demand of steadily growing compound numbers, which could be tested by high-throughput analysis techniques that were developed in parallel. Innovation in automation technology and protein purification methods were the driving forces that made the high-throughput enzyme assay the standard of rapid biological evaluation, allowing thousands or even millions of compounds to be tested in a short time.[7]

However, despite major investments, this concept has not fulfilled its expectations. Although combinatorial chemistry has yielded compounds that not only display biological activity, but could also be successfully developed into drugs, the overall number of newly approved "new chemical entities" has decreased in the past 10 years and does not reflect the increase in compound collection size, research effort or financial investments. In fact, of all the new chemical entities which were brought into the market as drugs between 2000 and 2006, roughly 50% were still natural products or derivatives of natural products.[8]

It became clear that quantity, *i.e.* the sheer number of compounds, does not determine the value of a compound collection. Instead, in order to improve the quality of the compound set (and obtain higher hit rates in biological screens), the attention turned to smaller libraries that are focused around a promising scaffold. The assumption is that when relevance to the natural target is given in the underlying scaffold architecture, considerably higher success in biological screens can be expected. New concepts in scaffold design in combinatorial chemistry focus, for example, on drug-likeness[9] or scaffold diversity. The latter concept was explored by the group of Schreiber, which also coined the term diversity oriented synthesis.[10]

Other approaches have reanimated the use of natural products for biomedical research.[11] Important contributions

to this area were made by the group of Danishefsky, which varied the natural product parent scaffolds (diverted organic synthesis[12]) and the group of Wender, which systematically reduced scaffold complexity (function-oriented synthesis[13]).

Natural products can be regarded as biologically pre-validated since they were selected by evolution to interact with their biological target(s). Moreover, in the course of their biosynthesis, their precursors were recognized by, bound to and modified by multiple enzymes. This means that the molecular properties of natural products encode preferred architectures for substrate recognition by enzymes and proteins.

Compound collections that are built around natural product scaffolds could therefore benefit from this biological preselection, exploit the relevant structural diversity present in nature and consequently give better results in biological screens. Synthetic efforts that are based on these principles are referred to as biology oriented synthesis (BIOS).[14]

Besides the focus on binding sites that are evolutionary selected (a concept that implicitly assumes a limited set of ligand binding site structures in proteins[15] which has been discussed in detail elsewhere[6]), statistical comparisons between natural products and other compound classes also reveal relevant differences.

Statistical comparison between natural products, drugs and combinatorial libraries

In order to understand the low success of large combinatorial compound collections in producing new candidates for drug development, chemoinformatic methods were applied to more closely examine the similarities and differences between the available drugs, combinatorial libraries and natural product compound collections. Feher and Schmidt reported an analysis of property distributions of the three different groups of small molecules.[16] Their results show that natural products differ considerably from combinatorial compounds with respect to scaffold architecture (fewer rotatable bonds, fewer unsaturated/aromatic rings, more stereocenters), molecular composition (fewer nitrogen and sulfur atoms, more oxygen atoms) and hydrogen bonding capabilities (more acceptors and substantially more donors). In general, natural products covered a wider range of different property parameter distributions than combinatorial compounds. This finding can be taken as support for the idea that biological diversity was not represented in the combinatorial libraries available at the time, *i.e.* the chemical space occupied by these collections is only of limited relevance for the discovery of bioactive compounds.

In a recent study, Shelat and Guy compared scaffold compositions of drugs, natural products, diversity-oriented compound collections, fragment collections and ''drug-like'' compound collections (defined by compliance with Lipinsky's ''rule of five'' for bioavailability).[17] The chemical space occupied by existing drugs was taken as a representation for the biologically relevant regions. It turned out that natural product scaffolds show a higher probability of overlapping with drug scaffolds than any other compound collection (when subcollections with the same number of members were compared). In conclusion, this results in a higher probability for natural products to score in biological screens (as long as the known biological targets are regarded).[18–21]

These results from statistical studies support the hypothesis that compound collections from BIOS closer approximate properties that are necessary for interaction with biological targets.

Such endeavours, in which the basic scaffold of the compound collection only approximates the original product, are called natural product inspired syntheses. They are in contrast to syntheses in which diversity is built around an existing natural product scaffold, which might itself be either the product of a synthesis or purified from biological extracts. Compound collections obtained from such syntheses can be referred to as natural product derived.[14]

Structural classification of natural products, SCONP

In order to gain deeper insight into the chemical space occupied by compounds from biological sources, Waldmann *et al.* undertook a statistical analysis which resulted in the structural classification of natural products, SCONP.[22] It provides an overview of the scaffolds that are present in natural products and can help to identify structures that are most relevant to nature and assist one in finding starting points for compound collection synthesis. The procedure by which scaffold architectures were identified, extracted from molecular structures and graphically displayed was designed to approximate and support a synthetic chemist's intuitive approach and is briefly described in the following section.

The basis for this analysis was set by all entries in the *dictionary of natural products* (DNP), the largest available collection of natural product structures. In order to make the information of this database amenable to chemoinformatic analysis, the records were automatically brought to a standard format. Most of these compounds ($>90\%$) contained at least one cyclic system, as do most of the molecules in biomedical research. Substances that did not contain any cyclic system were therefore excluded from the study. Notably, despite the presence of stereogenic centers in many natural products and the often drastically different bioactivities of enantiomeric compounds, the absolute configuration of stereocenters was neglected. This reduction of 3D structures to their 2D projections was necessitated by the limitations of the algorithm. However, other studies corroborated the fact that 2D representations generally provide a sufficient base for meaningful results in chemoinformatics. Moreover, absolute and relative configurations of stereocenters still remain to be determined for many natural products. Stereoisomers were thus not treated as different entries, but counted individually to add statistical weight to the underlying scaffolds. Synthesis projects based on this analysis can compensate for this limitation by addressing stereoisomers separately.

Additionally, compounds that bear sugar moieties were deglycosylated *in silico* prior to further systematization. Often, appending sugars serve as solubilising auxiliaries and do not contribute to the biological action of the aglycons.

After these operations, the record set was ready for further systematization. The guiding principle was to reduce structural

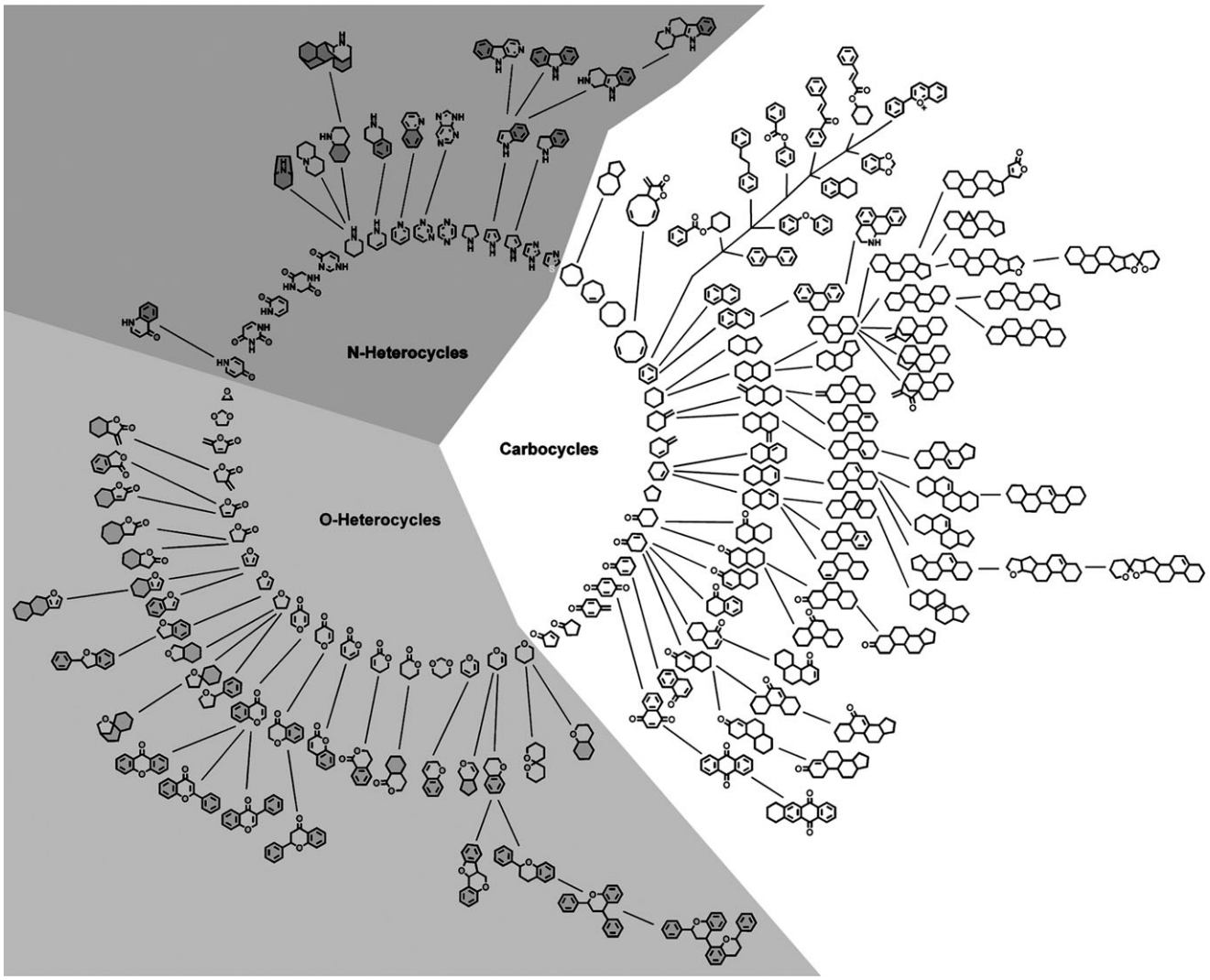


Fig. 2 SCONP: graphical representation of scaffolds of natural products. Reprinted with permission of PNAS.[22]

complexity of multi-ring systems and find less complex substructures that were also representatives of natural products. In this way, every multi-ring system (devoid of any substituents) could be assigned to a less complex parent scaffold. In order to follow a coherent methodology, a set of priority rules for the identification of the parent scaffold was set up and translated into an algorithm. The algorithm was designed to follow a chemist's intuition; for example, heterocyclic substructures were chosen to constitute the parent scaffold rather than carbocycles.

In doing so, all records of natural products could be represented in the form of a genealogical tree, the simplest and highest ranking elements of which were (necessarily) monocyclic compounds. Fig. 2 shows a graphical representation of all natural product scaffolds that represent at least 0.2% of the records in the DNP. In this arrangement, any information about natural sources or biosynthetic origins of the natural product scaffolds is omitted, which gives the chemist a quick orientation on structurally related scaffolds.

The diagram shown in Fig. 2 can be used as a guide and hypothesis generating tool when it comes to deciding where to start a compound collection synthesis project. Not only does it give an overview about scaffolds represented in natural products, but also gives suggestions about the simplifications or deviations associated with a given structure.

Synthesis of natural product-inspired compound collections

Natural product-like compound collections have been synthesized in a number of cases and these efforts are collected within a number of review articles.[6,23–25] It was shown that such compound collections score very well in biological screens.[26] Diversity-oriented approaches, reviewed by Tan and Shang[27] and exemplified by the work of Shair *et al.*,[28] were also successful. The synthesis of a natural product-like compound collection corresponds to the "chemical problem" depicted in the cycle in Fig. 1.

For the synthesis of natural product-inspired and -derived compound collections, solid phase organic synthesis is a viable technology. Immobilization of the substrate on a polymeric carrier enables efficient and straightforward removal of all

surplus reagents required in the multi-step sequences typical for the synthesis of such compound collections and, thereby, facilitates purification of the desired compounds.

The material obtained in such synthesis projects is usually sufficient for an initial evaluation. For example, if one takes 200 mg of a resin with a loading capacity of 0.9 mmol g^{-1} through a multi-step synthesis to generate a compound with a molecular weight of 300 g mol^{-1}, an overall yield of 50% would furnish a total of 27 mg. In several cases (especially when large collections are generated), the amount of resin can be considerably reduced.

It is evident that natural product-derived and -inspired syntheses on the solid support, which usually require multi-step sequences, pose a specific challenge to the synthetic chemist. Most prominently, the introduction of stereocenters must occur under strict control in order to avoid a mixture of isomers. Although this has been achieved in a couple of compound collection synthesis projects,[29] the established "toolbox" for reactions on immobilized substrates is considerably smaller than that known for solution phase chemistry. Thus, the continuous development of asymmetric reaction protocols for application on immobilized substrates is of great importance. Full stereocontrol in a multi-step sequence can be used to address each stereoisomer individually, which adds information to structure–activity relationships (SAR).

Although the development of such multi-step sequences requires more time and experimentation, this investment is paid off by the higher biological relevance of the compound collection, which should result in better hit rates in biological screens. An alternative approach towards natural product-like collections makes use of solid support-bound reagents.[30] This area of research, however, will be not covered by this review.

The next section presents two case studies, in which the successful solid-support syntheses of different compound classes are presented.

Case study I: α,β-unsaturated δ-lactones

Natural products which contain α,β-unsaturated δ-lactones display a wide range of bioactivities. Fostriecin (Fig. 3) inhibits the protein phosphatase PP2A with an IC_{50} of 1 nM and was also shown to inhibit topoisomerase II.[31] Leptomycin B inhibits nucleocytoplasmatic transport and was studied as a possible anti-HIV agent.[32] Pironetin was first reported to be an immunosuppressant, and was later found to inhibit tubulin polymerization.[33] Interestingly, the structurally less complex goniothalamin has cytotoxic activity.[34]

The Michael acceptor system of this structural motif could act as a nucleophile scavenger and thus represents the mechanistic basis for the biological activity of the different compounds. The nature of the nucleophile might be determined by the respective targets. Such mechanisms have been discussed[32] or shown to be operative[35] in a number of cases. The covalent linkage to the target proteins makes this compound class an illustrative example of protein reactive natural products.[32] Selectivity for the respective targets of the natural products might arise from the side chain of the lactone.

The fact that fostriecin and its relative cytostatin are potent inhibitors of the serine/threonine protein phosphatase PP2A,

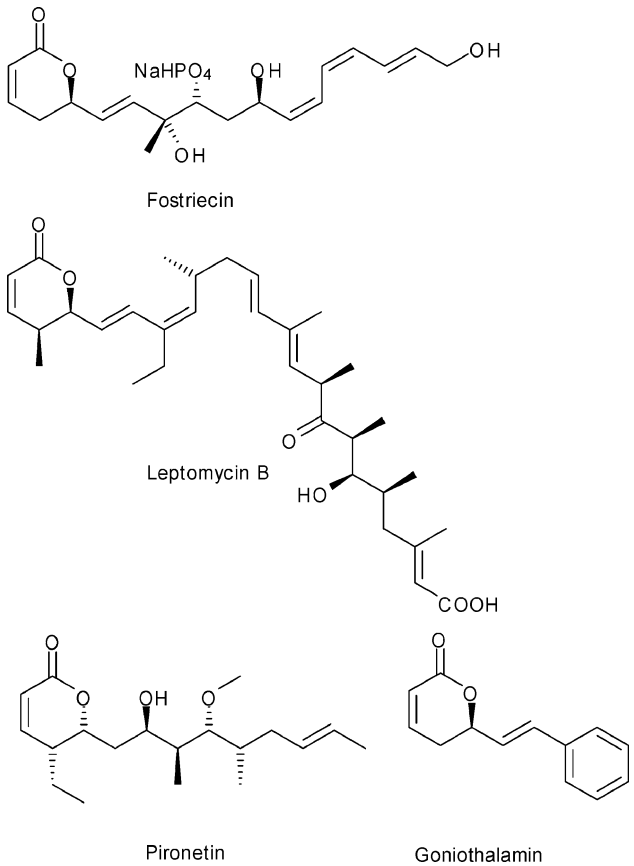


Fig. 3 Bioactive compounds that embody an α,β-unsaturated δ-lactone moiety.

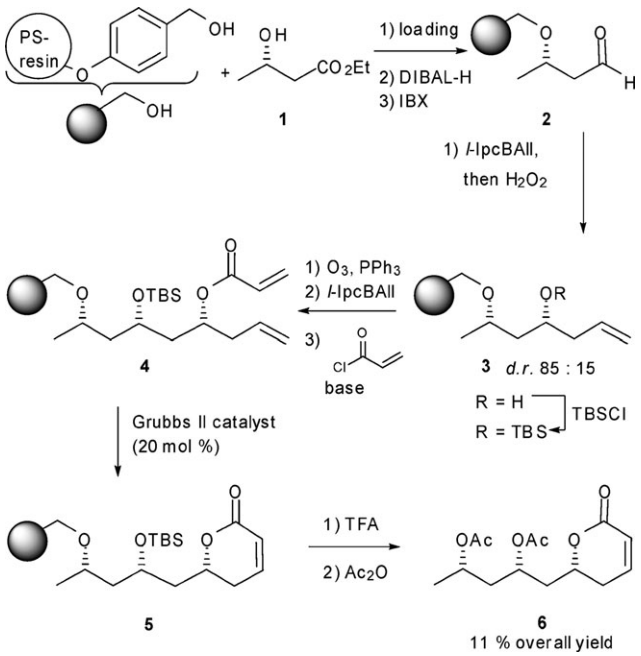
while other members of this family of natural products exhibit a variety of different biological activities led to the question of whether compounds that were built around the central α,β-unsaturated δ-lactone motif would display activity in phosphatase inhibition or beyond.

In order to answer this question, solid-phase syntheses for this class of molecules were developed that fulfilled the following criteria: (i) the lactone moiety should be installed in a stereoselective way, thereby enabling the independent biological evaluation of both enantiomers of a given structure and reducing the number of isomers from further stereocenters in the molecules and (ii) the δ-lactone moiety should be installed in the late steps of the synthesis since the α,β-unsaturated carbonyl system represents a reactive electrophile moiety.

Two different synthetic ways were developed in which the desired compounds were obtained stereoselectively by using solid-phase bound substrates.

In the first approach (Scheme 1), multiple stereoselective allylations were used on polymer-bound aldehydes to build up 1,3-polyols which are frequently found as substructures in natural products.[36] Acylation of the intermediate homoallylic alcohols with acryloyl chloride and ring closing metathesis should establish the electrophilic lactone in the last step before release from the solid support.

The synthesis started with enantiomerically pure β-hydroxy ester **1** (Scheme 1), which was immobilized on Wang resin and further converted to the corresponding aldehyde **2** (loading



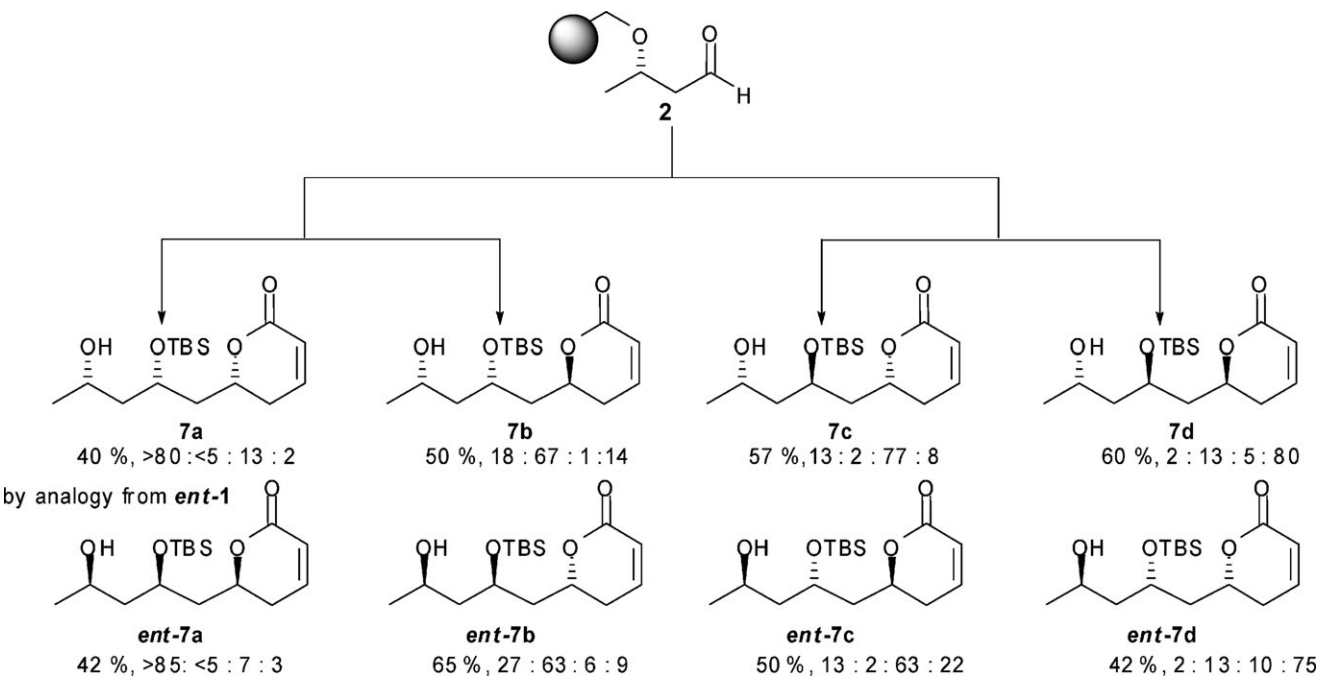
Scheme 1 Synthesis of cryptocarya diacetate on a solid support. IpcBAll: di(isopinocampheyl)allylborane; IBX: *o*-iodoxybenzoic acid; TFA: trifluoroacetic acid.

0.5 mmol g^{-1}). Allylation with Brown's diisopinocampheylborane gave rise to a homoallylic alcohol **3**, which was formed in a diastereomeric ratio of 85 : 15 (determined after release from the resin). The secondary hydroxy function was protected as a silyl ether, and the terminal alkene was cleaved with ozone, which generated a new aldehyde functionality upon reductive workup. A second Brown allylation, followed by acylation with acryloyl chloride led to cyclization precursor **4**, and ring closing metathesis with 20 mol% Grubbs II catalyst (2 cycles) furnished the α,β-unsaturated δ-lactone **5**. Release from the resin with concomitant deprotection and final acylation yielded the final product **6**, a natural product called cryptocarya diacetate, which was isolated from the African tree *Cryptocarya latifolia*.[37]

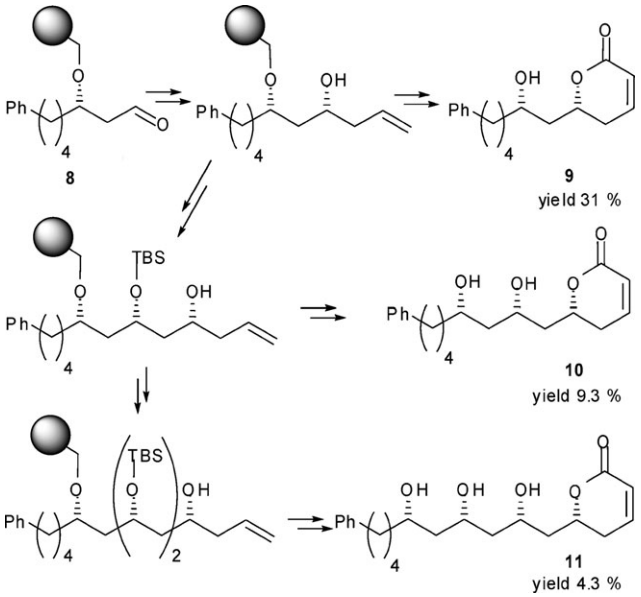
This synthetic sequence could be used to generate a collection of all diastereomers **7a–d** and their enantiomers of the natural product scaffold (Scheme 2). The combination of both enantiomers of the β-hydroxy ester and the two enantiomeric allylation reagents furnished all isomers in a 40–60% overall yield for the 7-step sequence. NMR analysis of the cleavage products also showed the high degree of stereoselectivity in the allylation steps.

A second series of experiments (Scheme 3) showed that this method can also be applied in longer sequences: up to four consecutive asymmetric allylations were performed on the immobilized aldehyde **8** to give compounds **9–11**. The compound collection obtained in this way could be used to probe the individual stereoisomers and compare the effects of the different homologs.

The second approach made use of a titanium-catalyzed, enantioselective hetero Diels–Alder reaction to build up the lactone ring (Scheme 4).[38] Bromo-Wang resin was loaded with the sodium enolate of malonic aldehyde **12**, and the resulting immobilized aldehyde **13** could be transformed into dienes **13a,b** with the respective Wittig-reagents. The polymer-bound dienes reacted with ethyl glyoxylate in the presence of 50 mol% of a catalyst formed from $Ti(OiPr)_4$ and *R*-BINOL, and dihydropyrans **14** were formed on the solid support in up to 95% ee, which was comparable to analogous reactions in solution. When the methyl-substituted diene **13b** was employed in the reaction, only the 5,6-*syn*-diastereomer was detected.



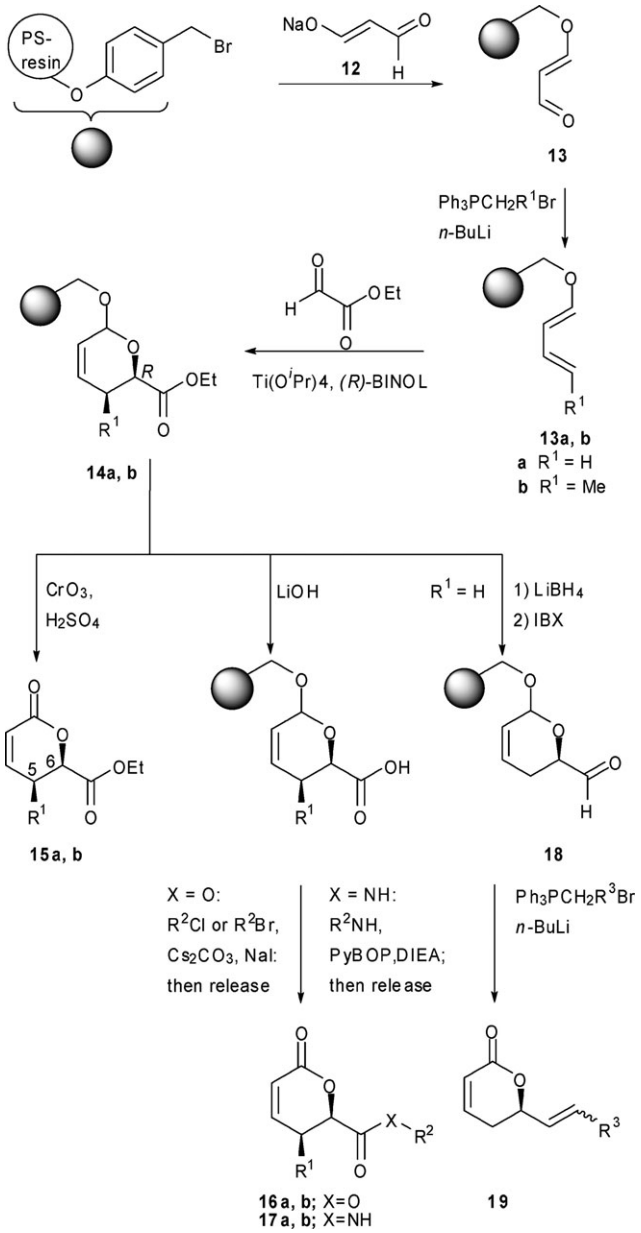
Scheme 2 Stereocomplementary synthesis of all diastereomers of the cryptocarya diacetate scaffold. See Scheme 1 for synthetic details, compounds were released with DDQ.



Scheme 3 Iterative asymmetric allylations on a solid support. See ref. 35 for details.

Treatment of the immobilized acetals with Jones' reagent led to an acid-mediated release and *in situ* oxidation to the α,β-unsaturated δ-lactones **15**, thereby generating the sensitive Michael acceptor system only in the very last step.

Further diversification of the esters **14** was achieved by transformations at the carboxyl functionality. Hydrolysis of the ethyl ester and re-esterification with alkyl- or benzyl halides furnished a collection of esters **16**, which were released from the solid support in 15–40% yield. Amidation of the free acid led to amides **17**, which were obtained in *ca*. 60% yield after cleavage and purification. Alternatively, ester **14a** was converted to the corresponding aldehyde **18** in a two-step sequence. Condensation with Wittig-ylides yielded a collection of olefinic compounds **19**, which were released from the solid support and obtained in 50–60% yield per step. In total, a collection of 50 enantioenriched compounds was obtained and used in cellular screens (see below).



Scheme 4 Synthesis of a collection of α,β-unsaturated δ-lactones using an oxa-Diels–Alder reaction as the key step. PyBOP: (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate; DIEA: *N*,*N*-diisopropylethylamine.

Case study II: indoloquinolizidines

The second case study demonstrates the use of SCONP in the field of BIOS. It shows how natural product scaffolds of bioactive compounds were reduced to less complex structures according to the graph displayed in Fig. 2.

In an initial screen of a natural product compound collection which examined inhibition of different protein phosphatases, several members of the family of yohimbine alkaloids (Fig. 4A) were identified as micromolar inhibitors of the protein phosphatase Cdc25A.[14] Cdc25A is a cell cycle regulator and a putative anticancer target, which makes selective inhibitors of this enzyme of great interest. Compounds with the pentacyclic yohimbine scaffold had not been reported to inhibit this phosphatase before, so a synthetic program directed towards the identification of related compounds with higher potency was initiated.

As the pentacyclic framework found in this type of alkaloid poses quite a synthetic challenge, especially in solid-phase synthesis, the SCONP tree was taken as a base for the so-called ''brachiation'' approach to identify related scaffolds (Fig. 4B). The structural ''parent'' scaffold of the yohimbine alkaloids **20**, which are part of the ''indole branch'' of natural products, was identified by removal of the E-ring, which led to the indoloquinolizidine scaffold **21**. Further retroannellation with respect to the SCONP tree identified tetrahydro-β-carbolines as the corresponding class of parent compounds, before the bicyclic indole scaffold **22** was reached. Scaffolds **21** and **22** were chosen as synthetic targets for a natural product-inspired compound collection. Both scaffold structures represent a large group of natural products themselves, which suggests

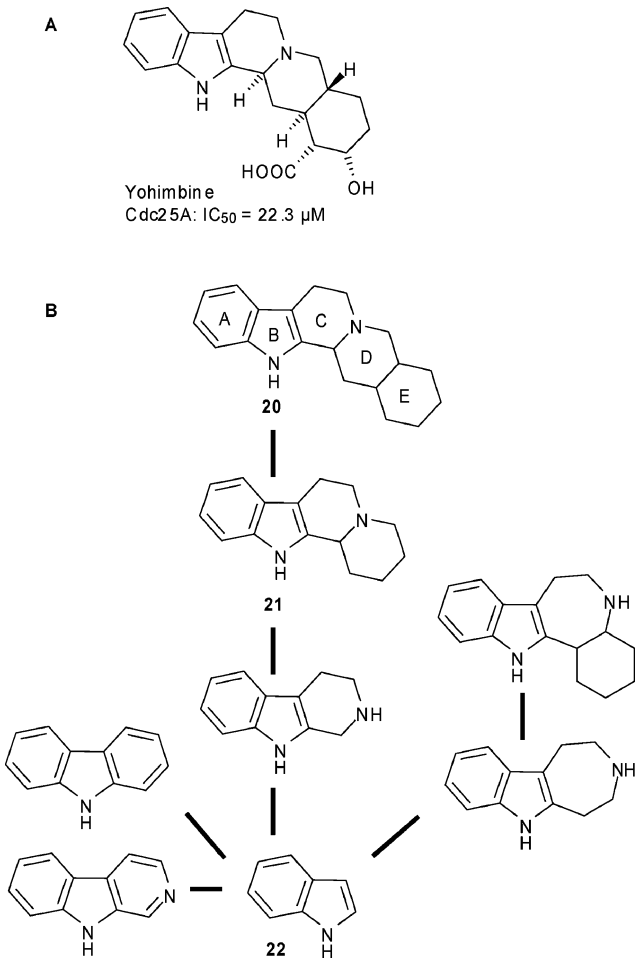


Fig. 4 **A** Yohimbine; **B** excerpt from the SCONP tree.

that biological relevance is retained during the "brachiation" process.

Again, all scaffold-building transformations were executed on polymer-bound substrates.[39] The synthesis started with Fmoc-L-tryptophan attached to Wang or 4-(hydroxymethyl)-benzoic acid amide (HMBA) resins **23** (Scheme 5). After removal of the Fmoc group, the amino group was condensed with a collection of aldehydes to give imines **24**. A Lewis acid catalyzed Mannich–Michael tandem addition yielded compounds **25**, in which the R^1 substituent was directed into the equatorial position with modest to good selectivity (dr 65 : 35 to 90 : 10). Another cyclization step was applied to obtain the desired scaffolds. In the case of HMBA resin, trifluoroacetic acid and TMSCl were used to give ketones that were readily released with sodium methoxide to give compounds **26** in 17–43% overall yield for the six step sequence. The main diastereomer as depicted in Scheme 5 was obtained with high selectivity (up to >99 : 1). Wang resin-bound dihydropyridinones **25** were cyclized with phosgene in the presence of TMSCl leading to vinyl chlorides **27** (selectivity ~ 7 : 3), which were either directly released with trifluoroacetic acid to give compounds **28** or further acylated with benzoyl chlorides prior to their cleavage from the polymeric support. The released compounds **28**, **29** and **30** were obtained in 55–76% overall yield as a mixture of four diastereomers, which could be readily separated by HPLC. Scheme 5 shows the diastereomer that was found to be the major product of the sequence.

Taken together, the synthesis furnished a collection of 450 tetracyclic compounds. Notably, compounds with both natural product-like and -unlike absolute configurations at C-12b (C–D-ring junction) were synthesized and tested for phosphatase inhibition independently.

Indoles

In the SCONP hierarchy, indoles represent the next parent scaffold. In order to probe the reduction approach further, a collection of indole derivatives that were originally designed as indomethacin analogs were also taken into consideration as potential phosphatase inhibitors.

The synthesis of these compounds was based on a "catch-and-release" strategy, incorporating the Fischer indole synthesis as the main step (Scheme 6).[40] Eight phenylhydrazines **33** were attached to a polystyrene aldehyde resin (loading 0.9 mmol g^{-1}) and acylated with nine different acyl chlorides. Release of compounds **34** from the resin under acidic conditions in the presence of ketones **35** led to *in situ* formation of hydrazones and their rearrangement to *N*-acyl indoles **36**. In this way, a collection of 197 compounds were synthesized with varying overall yields (4 to 99%). For further examples of solid phase synthesis yielding indole derivatives, please refer to ref. 41–43.

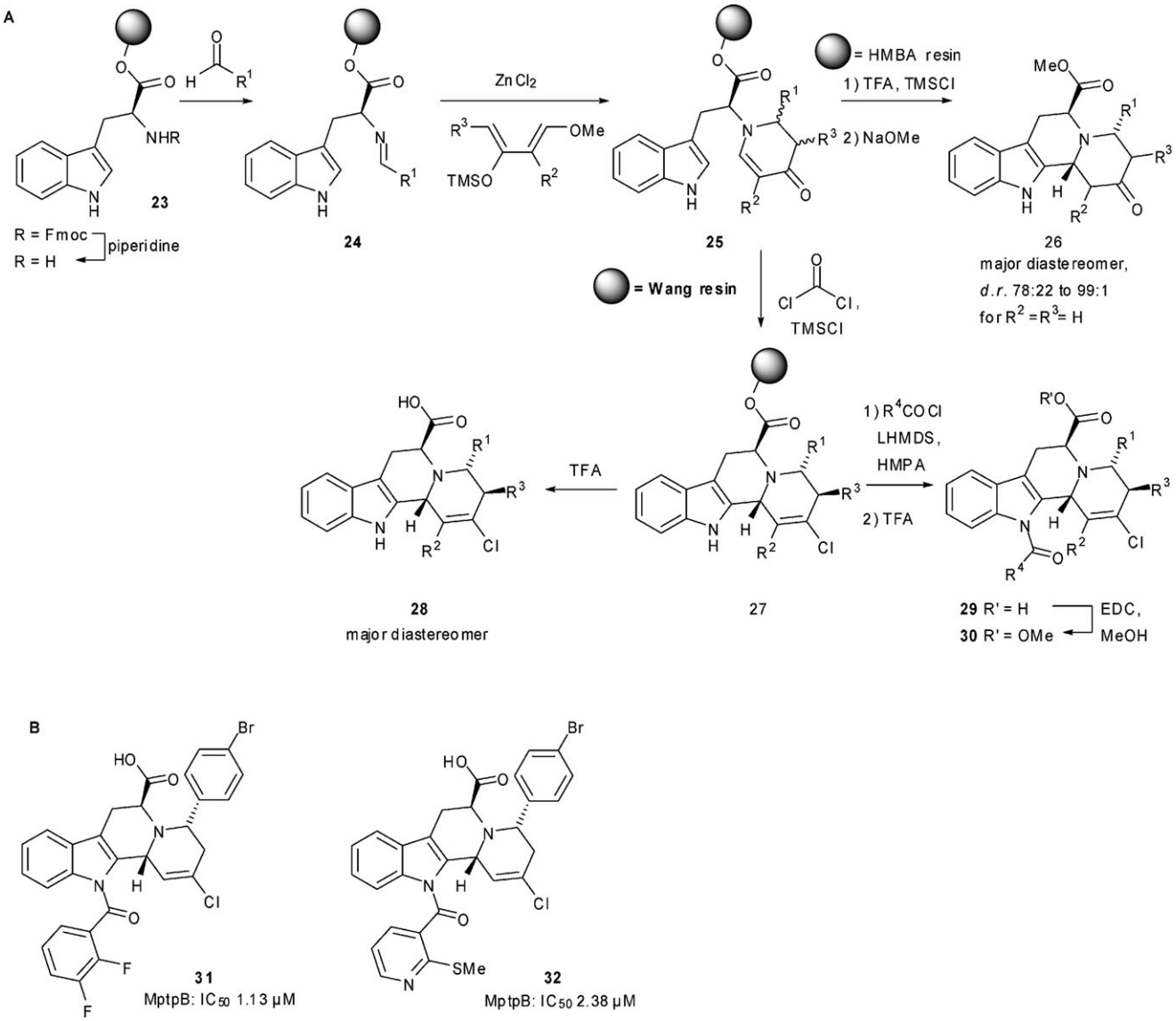
Biochemical *in vitro* assays and cell-based screening of the compound collections

Chemical biology aims to identify soluble and cell permeable small molecules that modulate cellular function in order to study biological signaling, regulation of gene transcription, protein–protein interactions or to identify compounds that have a potential therapeutic use. Thus, those small molecules can be used as research tools to investigate the set-up of the book of life. As depicted in Fig. 1, a starting point in the cycle between organic synthesis and biology is the analysis of a biological system or phenomenon of interest. This is accomplished by perturbing a given biological system by adding a small molecule and then analyzing the response, in whatever form it may be.[1] There are two ways to perform primary screens of large compound libraries: biochemical *in vitro* assays (reverse chemical genetics) and cell-based screening (forward chemical genetics). Potentially interesting small molecules or whole compound classes identified in such screens will thereby serve as starting points for the subsequent identification and validation of protein targets.

Before explaining the biochemical *in vitro* assays and cell-based assays, it is necessary to introduce the reader to the fundamentals of screening larger compound collections, particularly in an automated platform using 384-well microtiter plates. The use of the compound collections of α,β-unsaturated-δ-lactones and of the indoloquinolizidine derivatives in such a screen is described below.

Screening of compound collections

In compound collection screening, small molecules are tested for activity in biochemical *in vitro* assays as well as in cell-based



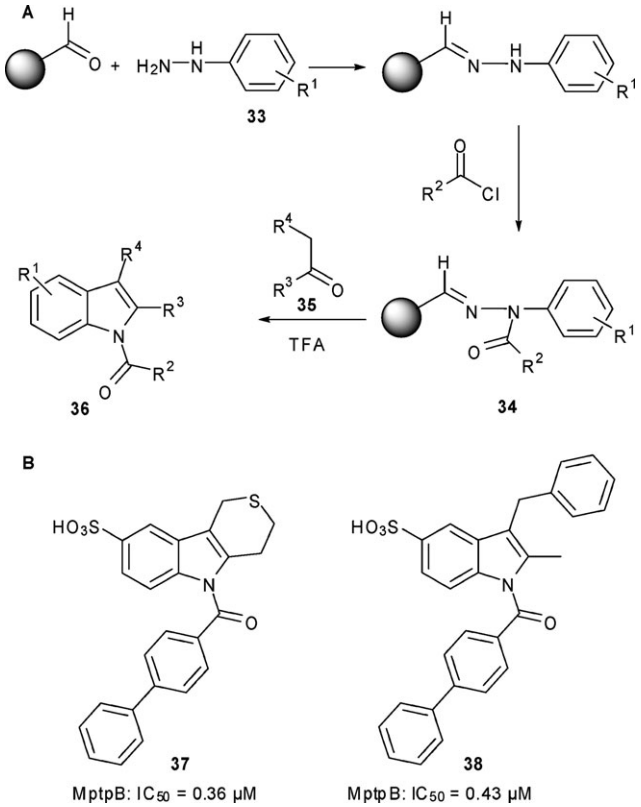
Scheme 5 **A** Synthesis of a collection of indoloquinolizidines, HMBA: 4-(hydroxymethyl)benzoic acid amide; **B** compounds identified as hits in the *in vitro* phosphatase screen.

assay systems. Major advantages of screening large collections are that more samples can be screened per unit time and less sample volume is needed. The assay design has to consider many issues. Is the nature of the response clearly defined? Is the stimulus dependent only on the compound being tested? Furthermore follow up screens (also called secondary or counter screens) should be included in assay design to validate the activity of identified hits, to eliminate false positives and also to determine the specificity of the hits. Often the follow up screen is performed in lower throughput but is more complex.[44]

During screening, several parameters must be optimized: The *Z* factor, a statistic parameter, is the most reliable parameter for predicting assay performance.[45] Furthermore, one needs to improve the sensitivity to allow identification of compounds with low potency, to achieve reproducibility, increase the stability of the measured response between different screening plates and wells, and to guarantee economic feasibility.[43]

Another important issue is the choice of the right positive and negative controls to be used in every screening plate. Once the right assay conditions are established, the small molecules are applied in a solubilized form. DMSO solubilizes most compounds and is compatible with biochemical *in vitro* assay screens and cell-based assays. *In vitro* biochemical assays are tolerant to 1–5% DMSO and most cell lines are unaffected by 0.1% DMSO. The right compound concentration for the primary screen must then be tested. Usually concentrations up to 20–40 μM are chosen for initial screening to identify also less active small molecules. Often it is also useful to include a concentration series of the identified hits.

To bring the assay format from cuvettes, tubes or dishes to higher throughput, one uses suitable microtiter plates. In 96-well microtiter plates, an overall assay volume of ~100 μL is used. Whereas the assay volume in 384-well microtiter plates ranges from 10–50 μL.[46] Advantages of the 384-well plate



Scheme 6 **A** Synthesis of a collection of indoles; **B** compounds that were identified as hits in the phosphatase screen.

format are the small amount of compound needed combined with the higher throughput.

Biochemical *in vitro* screens (reverse chemical genetics)

Compound collection screening can be used to efficiently investigate large numbers of compounds in biochemical *in vitro* assays to identify small molecules that are able to modulate the biological target of interest. In this reverse chemical genetic approach, a chosen target protein is screened against a compound collection. Since the target protein is a gene product, the gene sequence of interest is first cloned and then the target protein is expressed. Small molecule ligands able to bind to the target and either activate or inhibit its function are then identified. In the next step, the phenotypic response to the addition of these small molecules to a cellular system is studied. The reverse chemical genetic approach starts at the target (protein) and progresses to the analysis of the phenotype (Fig. 5).[46,47]

Setting up *in vitro* screens in a precise and fast manner often calls for automation, including appropriate and accurate liquid handling. Automation includes all necessary steps from mixing the components, incubation to a suitable end-point or equilibrium and measuring the read-out in the form of a detectable signal. This requires devices for liquid-handling in the micro- and nanolitre range since the total assay volume may be 1–10 μL after the addition of multiple reagents.[48] The assay readout should be simple and is, therefore, often based on a color change reaction.

Although in such screens the target of an active small molecule identified is already known, a detailed biological evaluation must follow the *in vitro* screen. Therefore, sufficient amounts of the target proteins are required for performing such screens. The expression of a correctly folded, functional and fully equipped protein can be a difficult and a time consuming task.[47]

The automated screening for inhibitors of protein phosphatases provides an example of a biochemical *in vitro* assay screen. Protein phosphatases have become important targets for chemical biology and medicinal chemistry research and there is a high demand for new phosphatase inhibitor classes. Protein phosphatases regulate innumerable biological processes and small molecule modulators of these enzymes have been useful in studying the chemical biology of the phosphatases. Recently, protein tyrosine phosphatases (PTPs) and dual-specificity phosphatases have moved into the focus of drug discovery programs in the fields of diabetes and anticancer research.

However, although important progress has been made, the development of potent and selective phosphatase inhibitors is still in its early stages and structurally new phosphatase inhibitor classes are in high demand. A recent screening investigated the inhibition of the tyrosine phosphatases VE-PTP, Shp-2, PTP1B, MptpA, and MptpB and the dual-specificity phosphatases Cdc25A and VHR. The compound collection of indoloquinolizidines and indoles described above was used in this *in vitro* phosphatase screen since members of

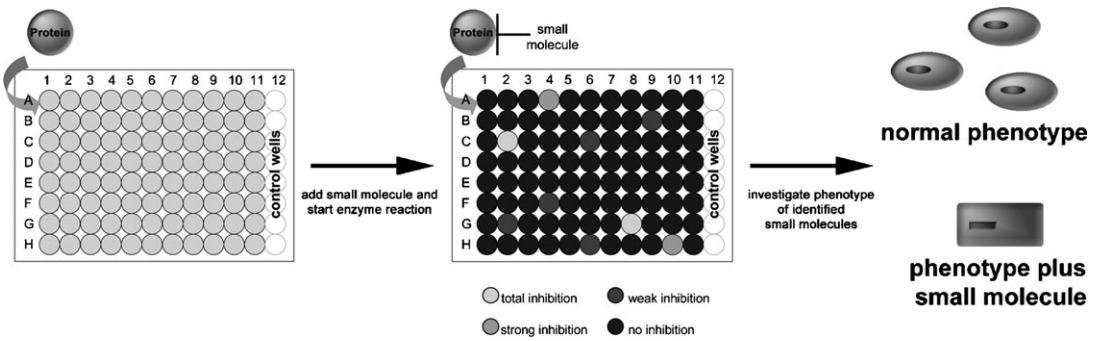


Fig. 5 Biochemical *in vitro* screen. A target protein is screened against a compound collection. The readout is done by a color reaction produced by an enzyme-catalyzed reaction involving the target protein. Small molecules that inhibit the enzyme reaction (light grey) are then investigated in a cellular system to study their phenotypic effects.

the yohimbine alkaloids were identified as micromolar inhibitors of the protein phosphatase Cdc25A. Besides four new selective phosphatase inhibitor classes, the screen discovered the first potent inhibitors for VE-PTP and MptpB (Schemes 5B and 6B). The identification of structurally less complex inhibitors by means of brachiation from the pentacyclic yohimbine scaffold to the scaffolds of substituted indoles *via* indoloquinolizidines provides a convincing example and argument for the notion that brachiation along the *N*-heterocyclic branches of the SCONP tree is a viable approach to phosphatase inhibitor development.[14]

Cell-based screening (forward chemical genetics)

High-content cell-based assays are a complementary approach to the biochemical *in vitro* identification of compounds by binding *e.g.* to an isolated protein.[38] In this forward chemical genetic approach, small molecules are used to screen for a desired phenotypic effect in the biological system under investigation. Once the screen reveals a suitable compound, the gene product which is modulated by the small molecule must be identified.[47] The biological system under investigation can be a prokaryotic or eukaryotic single cell organism (bacteria, fungi), physiological or pathological cells from complex multicellular vertebrate or mammalian organisms, or whole higher organisms, such as a fly, worm, zebrafish or mouse.[49] Thus, the forward chemical genetic approach identifies potential compounds on the basis of their conditional phenotypic effect on a whole biological system and not on the basis of their inhibition of a specific protein target. The approach goes from phenotype to protein (target) (Fig. 6).[47,49]

Selecting compounds that can act at multiple steps along a given biological pathway and monitoring them in a cellular context is a major advantage of cell-based screens.[38] The approach can be used to hit unknown proteins or to screen targets which cannot be assessed in pure form. Cell-based assays require the compounds to be cell-permeable. In a non-high throughput approach, this can be circumvented by microinjecting the small molecule into the cell.[50] The success of a phenotypic primary screen is dependent on the subsequent identification of the protein target which can be a demanding task as the reader will learn in the next paragraph. Often the amount of miniaturization is limited due to the fact that cell viability is poor in automated liquid handling.[48]

Pironetin (Fig. 3) arrests cell cycle progression in the M phase and is a potent inhibitor of tubulin assembly indicating a possible link between this biological process and α,β-unsaturated-δ-lactones. Thus, the collections of the α,β-unsaturated-δ-lactones and of the indoloquinolizidine derivatives described above were subjected to a high-content cell-based screen monitoring cell cycle progression.[38] To monitor cell cycle progression, BSC-1 cells (from African green monkey) were cultured in 384-well plates and treated with various concentrations of the synthesized compounds. BSC-1 cells are advantageous in assessing the influence of chemical compounds on the cell cycle. After formaldehyde treatment, cells were stained for chromatin and the actin and microtubule cytoskeletons. Microscopic analysis revealed that several of the investigated compounds influenced the microtubule cytoskeleton in dividing and/or non-dividing cells (Fig. 7). The screen identified modulators of biological processes at a high hit rate, supporting the notion that the previously described compound collections are enriched in biological activity.

Target identification and validation

Research projects starting with a cell-based phenotypic screen are usually followed by subsequent identification of the protein target(s) responsible for the observed phenotype. Current methods in target identification use either genomic approaches or affinity-based proteomics methods.

In the following section the reader will get a glimpse into these approaches. As target identification is a vast field with various aspects to consider, the reader is referred to several reviews and references therein which discuss these issues.[10,51–53]

Genomic approaches to target identification

In genomic approaches, loss-of-function mutants are generated by deletion, transposon insertion or RNAi in a genetic model organism (*e.g. S. cerevisae* (yeast), *C. elegans* (worm), *D. melangonaster* (fly) or mammalian cells).[54] This global approach is often conducted in large throughput and assists in elucidating the mode of action of small molecules with an unknown target or can help to reveal the mode of action of compounds whose direct targets are known but whose cellular consequences are not fully understood (Fig. 8).

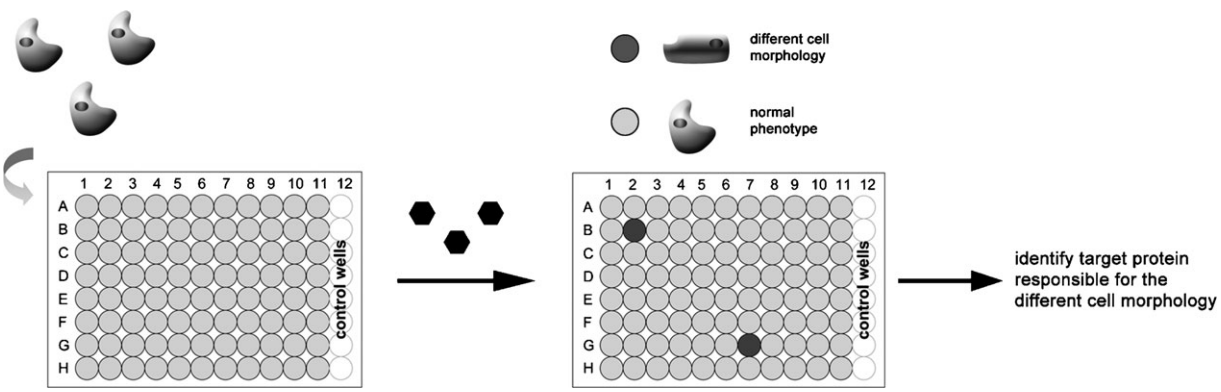


Fig. 6 Cell-based screening. Small molecules are used to screen for a desired phenotypic effect in the biological system under investigation. Afterwards the gene product (protein) responsible for the observed effect must be identified.

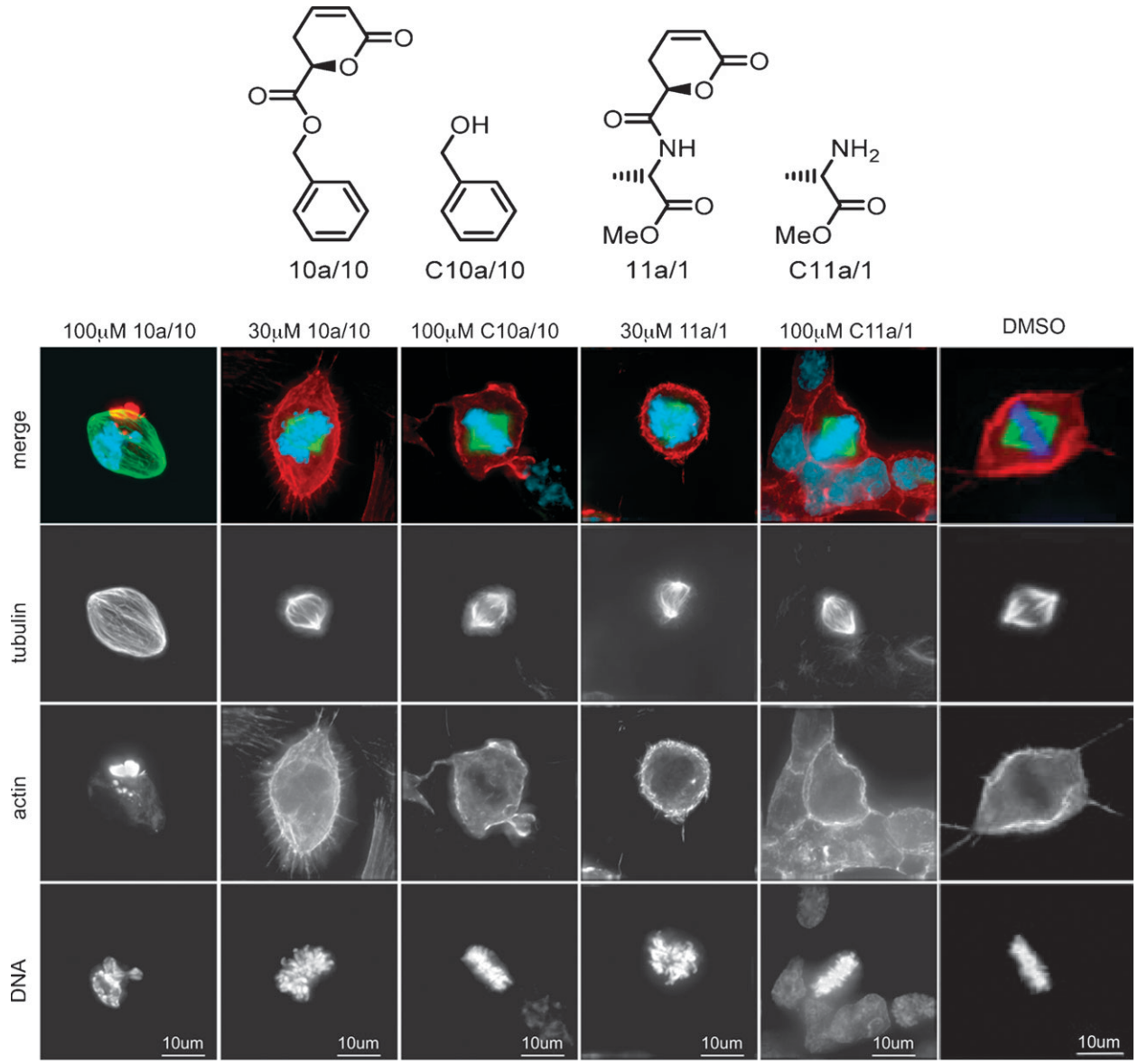


Fig. 7 Results of the assay for influence on the progression of the cell cycle. The pole-to-pole distance of spindles observed in cells treated with 100 μM **10a/10** seems to be greater than in DMSO-treated cells, suggesting that these compounds might affect proteins involved in regulating spindle morphology. At this concentration, **10a/10** also affected the actin cytoskeleton. While **10a/10** induced the formation of larger spindles, **11a/1** did not affect spindle length, but caused severe defects in chromosome alignment. Additionally, **11a/1**-treated cells displayed a bent spindle phenotype with curvy and disorganized microtubules, indicating that **11a/1**, like **10a/10**, might affect microtubule dynamics. Compound **11a/1**, unlike **10a/10**, did not affect the actin cytoskeleton. To confirm that the observed cellular phenotypes were mediated by **10a/10** and **11a/1** rather than by possible decomposition products, cells were treated with benzyl alcohol and alanine methyl ester, the products of eventual intracellular esterase or peptidase cleavage. These compounds had no effect on either the actin or microtubule cytoskeleton, confirming that **10a/10** and **11a/1** account for the observed cellular effects. Reprinted with permission from Chemistry and Biology.[38]

Most of those genetic approaches do not require any chemical modification. Apart from that, many of the methods require minute quantities of material.

All in all, this global approach is useful in only indirectly determining small molecule targets while providing the molecular basis for bioactivity.[55] Genomic approaches to target identification, including a number of success stories, are reviewed by Luesch[55] as well as by Zheng, Chan and Zhou.[54] To give one example, the drug target of tunicamycin has been identified using a subset of barcoded heterozygous deletion strains of *S. cerevisae*.[56] Methods in yeast-based genomic approaches are reviewed by Suter, Auerbach and Stagljar.[57]

Affinity based approaches to target identification

In the affinity based approach, a small molecule is covalently immobilised on a resin followed by affinity chromatography with protein lysates. Alternatively, biotinylated small molecules allow immobilization on streptavidin beads. Other approaches use photoaffinity probes in combination with, *e.g.* biotinylation, to covalently attach the small molecule

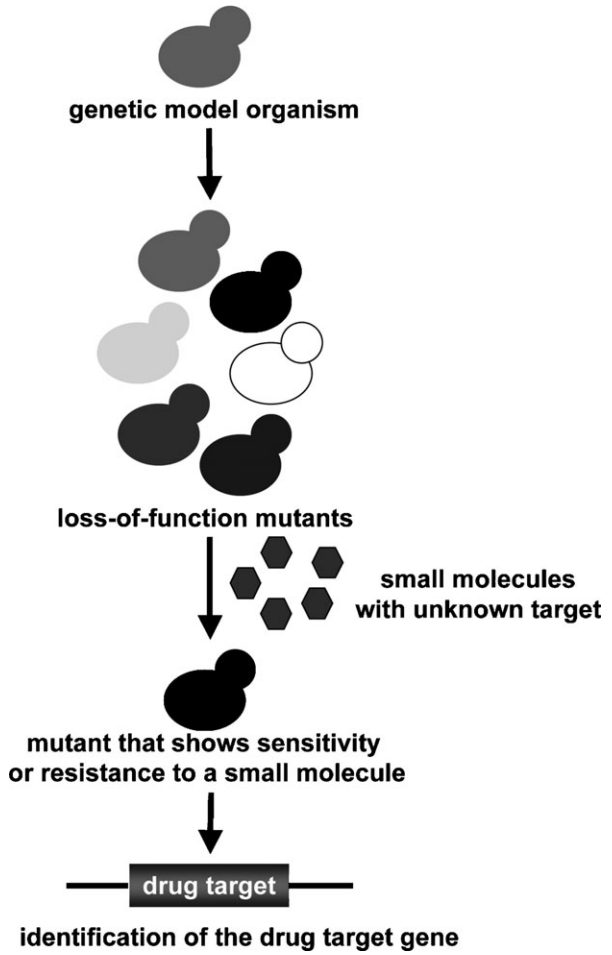


Fig. 8 Genomic approaches to target identification. Loss-of-function-mutants are generated in a genetic model organism (yeast, worm, fly or mammalian cells). Mutants that exhibit sensitivity or resistance towards a small molecule are further investigated to identify the gene target responsible for this alteration.

to the protein target. The target proteins are pulled down and are subsequently identified by either 1D or 2D SDS-PAGE followed by an in-gel tryptic digestion or by directly doing an on-bead tryptic digestion. Both methods finish with using diverse mass spectrometry methods (Fig. 9).

In the affinity based method, one must add an affinity tag using the correct linker to the small molecule. Other problems include the sensitivity of protein detection by mass spectrometry, non-specific binding and the abundance of proteins in the analyzed cell lysate.[51,53,55]

One example of a successful affinity based approach is the discovery of the cellular target of the natural product parthenolide.[58] Other examples of success stories in target identification using the affinity-based method are summarized by Burdine and Kodadek[51] and by the group of Gaviraghi.[52]

Neither the genomic nor the affinity based proteomics approach is widely applicable. The afore described methods to target identification merely give a general impression. Many approaches differ in subtleties from the described procedures while others start new basic approaches to target identification. Recent diverse approaches in target identification are reviewed in ref. 47, 52 and 59. Notable target identification approaches by the groups of Forsyth,[60] Romo,[61] Osada,[62] Golub,[63]and Licitra and Liu[64] are mentioned.

Once a potential protein target is identified, it has to be validated to confirm that it is the physiologically relevant binding partner of the small molecule *in vivo*. Methods to do so include, among others, the use of gene expression profiling, RNAi and yeast three hybrid studies. Many more experimental strategies exist,[65] however, they are not described here since a detailed analysis of these methods is beyond the scope of this review.

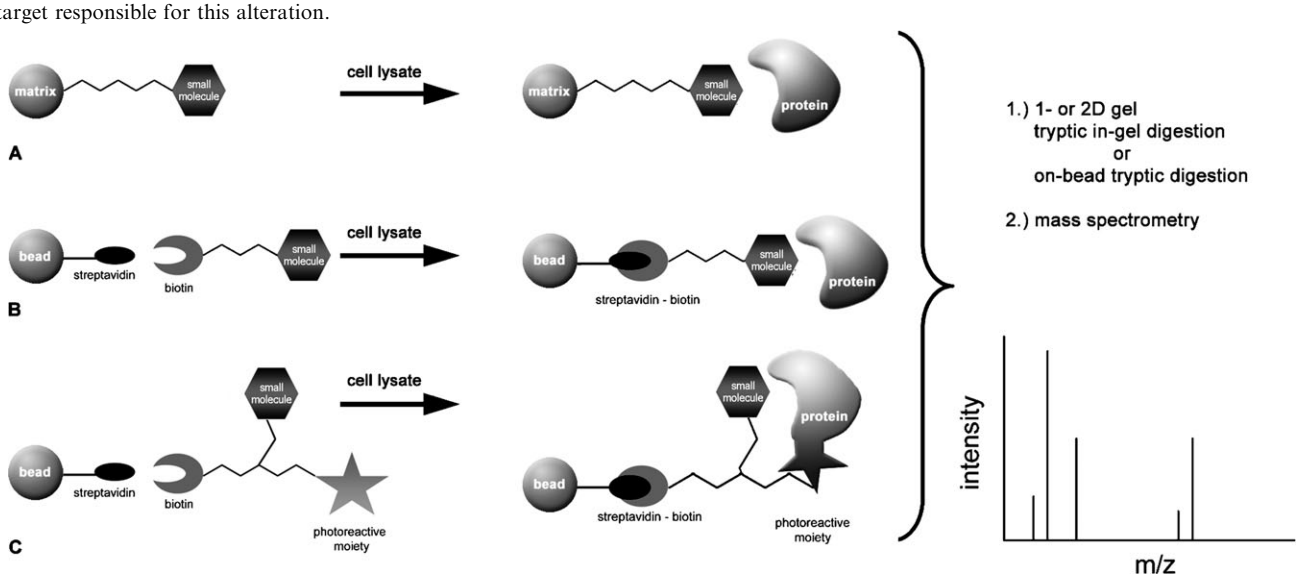


Fig. 9 Affinity based approaches to target identification. In the affinity based approach **A** a small molecule is covalently immobilized on a resin followed by affinity chromatography with protein lysates; **B** biotinylated small molecules allow immobilization on streptavidin beads; **C** other approaches use photoaffinity probes in combination with *e.g.* biotinylation to covalently attach the small molecule to the protein target. The target proteins are identified by either doing a 1D or 2D SDS-PAGE followed by an in-gel tryptic digestion or by doing an on-bead tryptic digestion. Both methods finish with using diverse mass spectrometry methods.

Conclusions

This review introduces the reader to the identification and the importance of small molecules, especially natural product-inspired compounds, as research tools to study biological questions. An efficient way to get suitable small molecules for this task is the concept of synthesizing and employing natural product-inspired compound collections. In this review, this concept was illustrated by the examples of solid phase synthesis of α,β-unsaturated-δ-lactones and indoloquinolizidine derivatives. Biochemical assays for phosphatase inhibition proved that brachiation along the *N*-heterocyclic branches of the SCONP tree is a viable approach to phosphatase inhibitor development. A cell-based phenotypic screen monitoring cell cycle progression identified modulators of this biological process at a high hit rate, supporting the notion that the previously described compound collections are enriched in biological activity. The next step will be to identify and validate the protein targets of the identified small molecules before they may be used to study biological questions.

References

1. K. Hinterding, D. Alonso-Diaz and H. Waldmann, *Angew. Chem., Int. Ed.*, 1998, **37**, 688–749.
2. B. R. Stockwell, *Nature*, 2004, **432**, 846–854.
3. D. Guiffant, D. Tribouillard, F. Gug, H. Galons, L. Meijer, M. Blondel and S. Bach, *Biotechnol. J.*, 2007, **2**, 68–75.
4. Editorial, *Nat. Chem. Biol.*, 2005, **1**(1), 3.
5. D. J. Newman, G. M. Cragg and K. M. Snader, *J. Nat. Prod.*, 2003, **66**, 1022–1037.
6. R. Breinbauer, I. Vetter and H. Waldmann, *Angew. Chem., Int. Ed.*, 2002, **41**, 2878–2890.
7. F. E. Koehn and G. T. Carter, *Nat. Rev. Drug Discovery*, 2005, **4**, 206–220.
8. D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461–477.
9. W. P. Walters and M. A. Murcko, *Adv. Drug Delivery Rev.*, 2002, **54**, 255–271.
10. M. D. Burke and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2004, **43**, 46–58.
11. J. Clardy and C. T. Walsh, *Nature*, 2004, **432**, 829–837.
12. R. M. Wilson and S. J. Danishefsky, *J. Org. Chem.*, 2006, **71**, 8329–8351.
13. P. A. Wender, V. A. Verma, T. J. Paxton and T. H. Pillow, *Acc. Chem. Res.*, 2008, **41**, 40–49.
14. A. Nören-Müller, I. Reis-Corrêa Jr., H. Prinz, C. Rosenbaum, K. Saxena, H. J. Schwalbe, D. Vestweber, G. Cagna, S. Schunk, O. Schwarz, H. Schiewe and H. Waldmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10606–10611.
15. M. A. Koch, L. O. Wittenberg, S. Basu, D. A. Jeyaraj, E. Gourzoulidou, K. Reinecke, A. Odermatt and H. Waldmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16721–16726.
16. M. Feher and J. M. Schmidt, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 218–227.
17. A. A. Shelat and R. K. Guy, *Nat. Chem. Biol.*, 2007, **3**, 442–446.
18. S. Wetzel, A. Schuffenhauer, S. Roggo, P. Ertl and H. Waldmann, *Chimia*, 2007, **61**, 355–360.
19. A. Schuffenhauer, P. Ertl, S. Roggo, S. Wetzel, M. A. Koch and H. Waldmann, *J. Chem. Inf. Model.*, 2007, **47**, 47–58.
20. K. Grabowski and G. Scheider, *Curr. Chem. Biol.*, 2007, **1**, 115–127.
21. M.-L. Lee and G. Schneider, *J. Comb. Chem.*, 2001, **3**, 284–289.
22. M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl and H. Waldmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 17272–17277.
23. D. G. Hall, S. Manku and F. Wang, *J. Comb. Chem.*, 2001, **3**, 125–150.
24. P. Arya, R. Joseph and D. T. H. Chou, *Chem. Biol.*, 2002, **9**, 145–156.
25. A. Ganesan, *Curr. Opin. Biotechnol.*, 2004, **15**, 584–590.
26. D. Brohm, S. Metzger, A. Bhargava, O. Müller, F. Lieb and H. Waldmann, *Angew. Chem., Int. Ed.*, 2002, **41**, 307–311.
27. S. Shang and D. S. Tan, *Curr. Opin. Chem. Biol.*, 2005, **9**, 248–258.
28. H. E. Pelish, N. J. Westwood, Y. Feng, T. Kirchhausen and M. D. Shair, *J. Am. Chem. Soc.*, 2001, **123**, 6740–6741.
29. T. Leßmann and H. Waldmann, *Chem. Commun.*, 2006, 3380–3389.
30. S. V. Ley and I. R. Baxendale, *Nat. Rev. Drug Discovery*, 2002, **1**, 573–586.
31. D. S. Lewy, C. M. Gauss, D. R. Soenen and D. L. Boger, *Curr. Med. Chem.*, 2002, **9**, 2005–2032.
32. C. Drahl, B. F. Cravatt and E. J. Sorensen, *Angew. Chem., Int. Ed.*, 2005, **44**, 5788–5809.
33. S. Kobayashi, K. Tsuchiya, T. Harada, M. Nishide, T. Kurokawa, T. Nakagawa, N. Shimada and K. Kobayashi, *J. Antibiot.*, 1994, **47**, 697–702.
34. A. de Fatima, L. K. Kohn, M. A. Antonio, J. E. De Carvalho and R. A. Pilli, *Bioorg. Med. Chem.*, 2005, **13**, 2927–2933.
35. T. Usui, H. Watanabe, H. Nakayama, Y. Tada, N. Kanoh, M. Kondoh, T. Asao, K. Takio, H. Watanabe, K. Nishikawa, T. Kitahara and H. Osada, *Chem. Biol.*, 2004, **11**, 799–801.
36. A. B. García, T. Leßmann, J. D. Umarye, V. Mamane, S. Sommer and Herbert Waldmann, *Chem. Commun.*, 2006, 3868–3870.
37. S. E. Drewes, M. M. Horn and R. S. Shaw, *Phytochemistry*, 1995, **40**, 321–323.
38. T. Leßmann, M. G. Leuenberger, S. Menninger, M. Lopez-Canet, O. Müller, S. Hümmer, J. Bormann, K. Korn, E. Fava, M. Zerial, T. U. Mayer and H. Waldmann, *Chem. Biol.*, 2007, **14**, 443–451.
39. I. R. Corrêa Jr., A. Nören-Müller, H. D. Ambrosi, S. Jakupovic, K. Saxena, H. Schwalbe, M. Kaiser and H. Waldmann, *Chem.–Asian J.*, 2007, **2**, 1109–1126.
40. C. Rosenbaum, P. Baumhof, R. Mazitscheck, O. Müller, A. Giannis and H. Waldmann, *Angew. Chem., Int. Ed.*, 2004, **43**, 224–228.
41. B. Meseguer, D. Alonso-Díaz, N. Griebenow, T. Herget and H. Waldmann, *Chem.–Eur. J.*, 2000, **6**, 3943–3957.
42. B. Meseguer, D. Alonso-Díaz, N. Griebenow, T. Herget and H. Waldmann, *Angew. Chem., Int. Ed.*, 1999, **38**, 2902–2906.
43. B. Sauerbrei, V. Jungmann and H. Waldmann, *Angew. Chem., Int. Ed.*, 1998, **37**, 1143–1146.
44. J. I. Inglese, R. L. Johnson, A. Simeonov, M. Xia, W. Zheng, C. P. Austin and D. S. Auld, *Nat. Chem. Biol.*, 2007, **3**, 466–479.
45. P. W. Iversen, B. J. Eastwood, G. S. Sittampalam and K. L. A. Cox, *J. Biomol. Screening*, 2006, **11**, 247–252.
46. D. A. Preira and J. A. Williams, *Br. J. Pharmacol.*, 2007, **152**, 53–61.
47. D. R. Spring, *Chem. Soc. Rev.*, 2005, **34**, 472–482.
48. J. Wölke and D. Ullmann, *Drug Discovery Today*, 2001, **6**, 637–646.
49. M. Bredel and E. Jacoby, *Nat. Rev. Genet.*, 2004, **5**, 262–275.
50. T. J. Mitchison, *ChemBioChem*, 2005, **6**, 33–39.
51. L. Burdine and T. Kodadek, *Chem. Biol.*, 2004, **11**, 593–597.
52. G. C. Terstappen, C. Schlüpen, R. Raggiaschi and G. Gaviraghi, *Nat. Rev. Drug Discovery*, 2007, **6**, 891–903.
53. R. Aebersold and M. Mann, *Nature*, 2003, **422**, 198–207.
54. X. S. Zheng, T.-F. Chan and H. H. Zhou, *Chem. Biol.*, 2004, **11**, 609–618.
55. H. Luesch, *Mol. BioSyst.*, 2006, **2**, 609–620.
56. G. Giaever, D. D. Shoemaker, T. W. Jones, H. Liang, E. A. Winzeler, A. Astromoff and R. W. Davis, *Nat. Genet.*, 1999, **21**, 278–283.
57. B. Suter, D. Auerbach and I. Stagljar, *BioTechniques*, 2006, **40**, 625–644.
58. B. H. B. Kwok, B. Koh, M. I. Ndubuisi, M. Elofsson and C. M. Crews, *Chem. Biol.*, 2001, **8**, 759–766.
59. S.-Y. Han and S. H. Kim, *Arch. Pharm. Chem. Life Sci.*, 2007, **340**, 169–177.
60. C. J. Forsyth, L. Ying, J. Chen and J. J. La Clair, *J. Am. Chem. Soc.*, 2006, **128**, 3858–3859.
61. S. Peddibhotla, Y. Dang, J. O. Liu and D. Romo, *J. Am. Chem. Soc.*, 2007, **129**, 12222–12231.
62. N. Kanoh, K. Honda, S. Simizu, M. Muroi and H. Osada, *Angew. Chem., Int. Ed.*, 2005, **44**, 3559–3562.
63. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander and T. R. Golub, *Science*, 2006, **313**, 1929–1935.
64. E. J. Licitra and J. O. Liu, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 12817–12821.
65. R. Kramer and D. Cohen, *Nat. Rev. Drug Discovery*, 2004, **3**, 965–972.