Virtual screening for the discovery of bioactive natural products

By Judith M. Rollinger[1],
Hermann Stuppner[1,2] and
Thierry Langer[2,3]

[1]Institute of Pharmacy/Pharmacognosy, Leopold-Franzens University of Innsbruck, Innrain 52c, 6020 Innsbruck, Austria

[2]Inte:Ligand GmbH, Software Engineering and Consulting, Clemens Maria Hofbauergasse 6, 2344 Maria Enzersdorf, Austria

[3]Institute of Pharmacy/Pharmaceutical Chemistry/Computer Aided Molecular Design Group, Leopold-Franzens University of Innsbruck, Innrain 52c, 6020 Innsbruck, Austria
<Thierry.Langer@uibk.ac.at>

Abstract

In this survey the impact of the virtual screening concept is discussed in the field of drug discovery from nature. Confronted by a steadily increasing number of secondary metabolites and a growing number of molecular targets relevant in the therapy of human disorders, the huge amount of information needs to be handled. Virtual screening filtering experiments already showed great promise for dealing with large libraries of potential bioactive molecules. It can be utilized for browsing databases for molecules fitting either an established pharmacophore model or a three dimensional (3D) structure of a macromolecular target. However, for the discovery of natural lead candidates the application of this *in silico* tool has so far almost been neglected. There are several reasons for that. One concerns the scarce availability of natural product (NP) 3D databases in contrast to synthetic libraries; another reason is the problematic compatibility of NPs with modern robotized high throughput screening (HTS) technologies. Further arguments deal with the incalculable availability of pure natural compounds and their often too complex chemistry. Thus research in this field is time-consuming, highly complex, expensive and ineffective. Nevertheless, naturally derived compounds are among the most favorable source of drug candidates. A more rational and economic search for new lead structures from nature must therefore be a priority in order to overcome these problems.

Here we demonstrate some basic principles, requirements and limitations of virtual screening strategies and support their applicability in NP research with already performed studies. A sensible exploitation of the molecular diversity of secondary metabolites however asks for virtual screening concepts that are interfaced with well-established strategies from classical pharmacognosy that are used in an effort to maximize their efficacy in drug discovery. Such integrated virtual screening workflows are outlined here and shall help to motivate NP researchers to dare a step towards this powerful *in silico* tool.

1   Introduction

In the field of drug discovery we are confronted by a paradox situation: highly efficient tools and advanced technological and molecular know-how, e.g., in the area of genomics, combinatorial chemistry, high through-put screening (HTS), robotized and miniaturized process cycles, could find entrance in big pharmaceutical industries. These costly procedures were expected to raise the number of launched drug substances; however the results were disappointing [1, 2]. In 2002, Adam Smith, the chief-editor of *Nature* presented the sobering data of research and development expenses of the 20 leading pharma companies *versus* new drugs on the market. They have steadily fallen in recent years despite the increasing financial efforts [3].

On the other side we are faced by a high traditional impact of naturally derived medicines and incredible success stories of natural products (NPs)

213

as potent remedies from the beginnings of human therapeutic activity to modern research and drug development. Nevertheless, most large pharmaceutical companies scaled down or terminated their work in NPs operations. The reasons behind this are that the drug discovery process starting from natural sources is hardly compatible with the today's highly automated drug discovery technologies. Thus, the pre-eminence of combinatorial chemistry as the preferred method for generating new drug leads has led to the comparative neglect of this valuable resource. William Strohl from Merck Research Laboratories summarized the difficulties of NP programs *versus* synthetic chemicals in his editorial remarks in *Drug Discovery Today* [4]. These include (i) the existence of already found potent antimicrobic and antitumor NPs and the lack of sufficient dereplication programs which prevent their repeated discovery; (ii) the fact that – in contrast to the highly sophisticated molecular targets – NP extracts are generally regarded as too 'dirty', too difficult to assay and too time-consuming; (iii) obtaining an assay hit resulting from a bio-guided fractionation, the NPs' structure still has to be elucidated compared with synthetic chemicals; (iv) NPs are often deemed as too structurally complex, possessing multiple hydroxyl moieties, ketones and chiral centers. Strohl nevertheless concluded by listing a number of advantages applying an active NP program, which he finally described as an '*expensive endeavor*' which, however, is '*well worth the cost*'.

The use of NPs has been the single-most successful strategy for the discovery of new drug leads, which is clearly shown by different statistics [5, 6]. With increased calls in recent years for further research on NPs [7, 8] there are again signs that they may play a more active role in the future drug discovery process, since their reintroducing may help to *re-discover the sweet spot in drug discovery* [1].

2   Status of NPs

To date some 200,000 natural compounds [9–11] have been published. The terrestrial flora has been intensively investigated over the last decades; the potential in finding new NPs slumbering in untapped biota is however nearly inconceivable. It is estimated that only 5–15% of the approximately 250,000 described high plant species have ever been in the focus of phytochemical and pharmacological investigations [12]. More sobering is the

percentage in the field of bacterial (less than 1%) and fungal species (less than 5%) [13].

The main part of known NPs belongs to secondary metabolites. These compounds provide living systems with their characteristic features mandatory for surviving. They contain an inherently large-scale of structural diversity. About 40% of the chemical scaffolds of published NPs are unique and have not been made by any chemist [14].

In the past 100 years researchers have discovered many potential therapeutic targets. Since the completion of the human genome, 30,000 to 40,000 genes and at least the same number of proteins are assumed [15]. Thus, we are up against an increasing number of macromolecular targets, like proteins, receptors, enzymes, and ion channels – that might be of pathological concern for humankind. Among them, proteins continue to attract significant attention from pharmaceutical technology as a valuable source of drugable targets [16]. Proteins provide the critical link between genes and disease, and thus are the key to understanding the basic biological processes. Up to now drug discovery has been performed against only approximately 500 targets [17], though the number of potential targets are estimated to be in the range of 2,000 to 5,000 [2, 15].

Taken together, it can be assumed that a large number of drug leads and hits are conserved in the inexhaustible pool of NPs pre-screened by evolution. But how to dig out and to recognize the respective drug leads is a challenging task for both industry and academia, for medicinal chemists, pharmacognosists and pharmacologists. NP research is affected with a wealth of time-consuming and cost intensive investigations. Collection of the natural material, phytochemical analysis, isolation and identification of the constituents is just the basic procedure. A biological screening of extracts or even the arbitrary testing of isolated metabolites is feasible and often performed, though is not at all a focused procedure, thus unpractical and too expensive. The NPs' diversity has to be accessed in a more rational way.

3   Holistic *versus* molecular approaches in drug discovery from nature

During the last century and even today the discovery of bioactive NPs and their development into potential drug candidates are mainly covered by
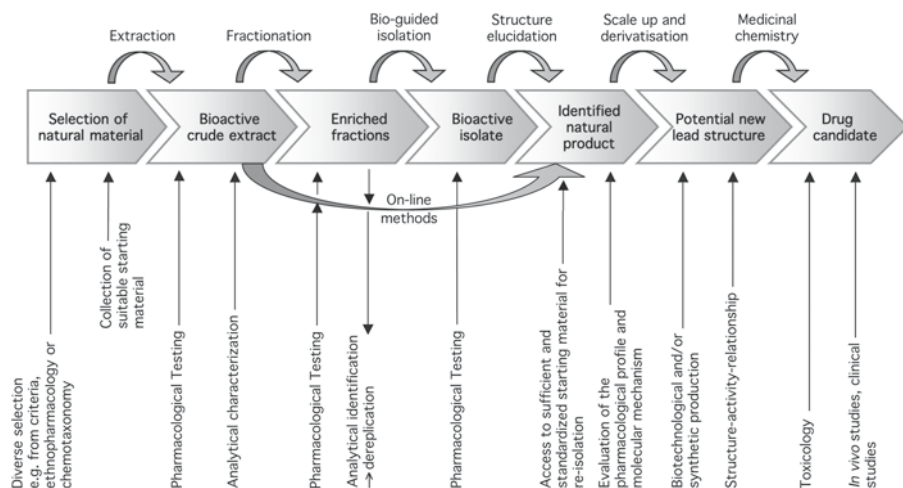
Figure 1.
Traditional early drug development of a nature based drug candidate

a holistic approach. A characteristic workflow of this procedure is given in Figure 1. Starting from the knowledge or assumption about a biological effect the natural material is selected and adequately extracted. If a positive effect in the obtained multi-component extract is recorded, it is attempted to trace it back to the active principle/s by intense phytochemical and analytical investigations (Fig. 1). This can for instance be achieved by a bioactivity-guided fractionation. A more targeted approach focuses on innovative technological tools combining analytical and biological information. An overview of recent developments in this area and successful examples thereof are presented by Potterat and Hamburger [18, 19].

As soon as the constituent regarded to be responsible for the overall effect is isolated, further research focuses on a molecular level including structure elucidation and pharmacological profiling. Synthesis and testing of series of derivatives enable an insight into a structure-activity-relationship and pharmacokinetic aspects. Finally, potential drug leads become drug candidates after some intense toxicological studies and after the verified effectiveness *in vivo* (Fig. 1).

Recent advances in lead identification from nature work on a molecular base more than on a holistic one. A first prerequisite for that is on bioin-
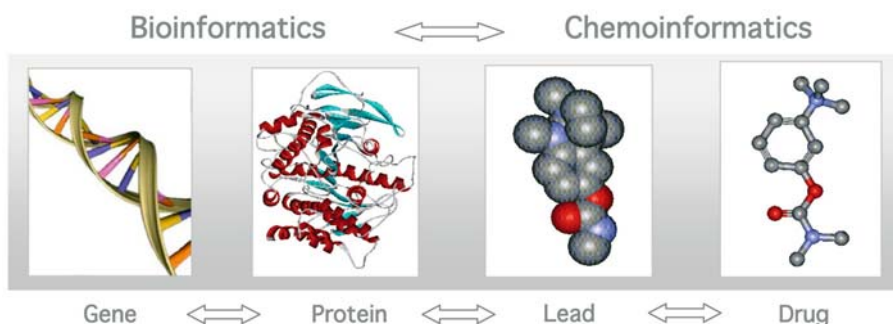
Figure 2.
Schematic correlation between bioinformatics and chemoinformatics

formatics comprising 3D structures from genes and proteins (bioinformatics), substantial knowledge on molecular target functions with accurate structural information and protein–ligand interactions. Secondly, it is essential to refer to unambiguously characterized structures of secondary metabolites preferably with some information to their biological effect. Based on available structural as well as biological knowledge from both sides, information can be deduced from chemoinformatics to bridge the gap between known ligands and the discovery of new lead structures (Fig. 2).

4   Computational approaches for the discovery of lead structures from nature

The increasing understanding of fundamental principles of protein–ligand interactions and the steadily growing number of 3D-structures of potential and experimentally proved ligands provide undreamed of possibilities towards more rationalized concepts in drug discovery. However, too much is expected of the human brain to profit from the already published information. Thus, efficient and effective approaches benefit from today's knowledge about NPs. In the area of medicinal chemistry, computational methods, like virtual screening experiments, have already proved to satisfy these requirements. They are needed to exploit the available structural

information, to understand specific molecular recognition events, and to clarify the function of the target macromolecule. Though rationalized procedures in the search for bioactive natural products are in great demand to find the 'needles in a haystack', computational assistance could hardly break into natural product research.

The common idea of all computational approaches within the early drug discovery process is to mine more or less large compound databases *in silico* and to select a limited number of candidates proposed to have the desired biological activity. For this process the term 'data mining' was coined in 1996 [20], which was concisely defined by Gasteiger and co-authors: '*to extract knowledge from a large set of data in order to make predictions of new events*' [21].

Within the lead discovery process, virtual screening technologies have largely enhanced the impact of computational chemistry and nowadays chemoinformatics plays a predominant role in early phase drug research [22, 23]. The key goal of the use of such methods is to reduce the overall cost associated to the discovery and development of a new drug, by identifying the most promising candidates to focus the experimental efforts on. Recently published books and reviews on the impact of computational chemistry for lead structure determination highlight these efforts [24–27].

If the 3D structure of the biological target is known, high through-put docking turned out to be a valuable structure-based virtual screening method to be used [28–31]. Within this context, the scoring of hits retrieved still remains a question that is often discussed. In fact, currently the major weakness of docking programs lies not in the docking algorithms themselves but still in the inaccuracy of the functions that are used to estimate the affinity between ligand and target, the so-called scoring functions. Previously, Stahl and Rarey analyzed scoring functions for virtual screening [32], giving valuable insight into strengths and weaknesses of currently used models for affinity estimation. The combination of several different scoring functions termed as consensus scoring turns out to be one of the possible answers to the question raised previously. In fact, several authors recently described their efforts in this area; an example is given in reference [33]. In a theoretical study, other authors demonstrate that consensus scoring outperforms any single scoring for simple statistical reasons and that a moderate number of scoring functions (i.e., three or

four) are sufficient for the purpose of consensus scoring [34]. However, it has been shown that consensus scoring alone is not suitable for all cases of docking, and, as highlighted in a recent review by Krovat and co-authors, considerable efforts are still devoted to the optimization of scoring functions [28].

Because of the restricted free access to NP 3D libraries (see below), the number of virtual screening studies published for the rational access to bioactive NPs is limited. Some examples using high throughput docking as a structure-based virtual screening tool will be given here: Liu and Zhou applied a theoretical approach to find natural ligands as potential inhibitors of the SARS-CoV protease, a virus target of the severe acute respiratory syndrome [35]. They used a docking-based virtual screening cycle and applied drug-like filters to finally propose 18 drug candidates out of two 3D databases comprising metabolites from marine organism and compounds from traditional Chinese Medicine. The same virus organism was the main interest in the study performed by Toney et al., who focused on its main proteinase, 3CLpro. The crystal structure of this attractive target was used as the starting point for the virtual docking screening of the NCI database. Searching for non-peptidyl inhibitors, the authors identified the naturally occurring terpenoid alkaloid sabadinine (i.e., cevine; **1**) as potential anti-SARS agent [36].

The author group around Stefano Moro could identify ellagic acid (**2**) as inhibitor of the protein kinase CK2 screening an in-house generated database with almost 2,000 structures of natural compounds [37]. A combination of four docking protocols and five scoring functions has been utilized to dock and rank the molecules in the database. The consensus docking suggested ellagic acid to be one of the most promising candidates. This assumption could be verified by experimental studies revealing this NP as highly potent CK2 inhibitor (K_i = 20 nM).

Estrogen receptor-β plays a key role in regulating brain development and estrogen-induced promotion of neurogenesis and memory. Using the 3D coordinates of the co-crystal structure of human estrogen receptor-β bound with genistein as starting point, Zhao and Brinton pursued a receptor-based molecular docking approach [38]. They focused on the search for natural estrogen receptor-β-selective ligands. Twelve candidate molecules, which had been suggested by the database screening, were selected. The authors determined their binding affinity and selectivity; three of the com-

pounds belonging to the flavanoid family (**3–5**) displayed over 100-fold binding selectivity to the estrogen receptor-β over α. A similar approach was employed by Liu and co-authors. Applying a docking virtual screening filtering experiment, the authors discovered potent inhibitors of the potassium ion channel from a Chinese NP database [39].

5 Pharmacophore concept in NP research

The pharmacophore concept has proven to be extremely successful, not only in rationalizing structure-activity relationships, but also by its large impact in developing the appropriate 3D-tools for efficient virtual screening [40]. Profiling of combinatorial libraries and compound classification are other often-used applications of this concept. Although well established in combinatorial chemistry, it has to be pointed out that the tools described in this section have likewise a considerable impact on the rational finding of new potential lead compounds originating from the immense source of secondary metabolites. The prior use of pharmacophore models in biological screening of NPs is an efficient procedure since it quickly eliminates molecules that do not possess the required features thus leading to a dramatic increase of enrichment, when compared to a purely random screening experiment. In a previous study conducted by Doman and co-authors [41], only 85 molecules or 0.021% revealed as protein tyrosine phosphatase-1B inhibitors ($IC_{50} < 100$ μM) by a HTS of approximately 400,000 compounds. On the other hand, of 365 molecules suggested by molecular docking, 127 or 34.8% were found to be active. Thus, docking-based virtual screening enriched the hit rate by almost 1,700-fold over random screening.

One should not forget, however, that additional molecular characteristics not reflected by pharmacophore models (physicochemical properties relevant for ADME and toxicological properties) must be taken into account when deciding upon which compounds should be further developed [42]. A rapid identification and elimination of compounds with unsuitable physicochemical and pharmacokinetic properties is a pivotal step in the early drug discovery process [43, 44]. They can be evaluated traditionally or by high throughput screening, which are discussed in detail by Avdeef and Testa [45]. This must be considered for synthetics

as well as NPs, though studies revealed secondary metabolites not only high scaffold diversity; biosynthesized molecules also show structural and spatial characteristics that are closer to drug leads than those of synthetic molecules [46, 47]. Typically, NPs include more chiral centers and their stereochemical architecture is much more complex than that of synthetic molecules. Furthermore, they usually contain more carbons, hydrogen and oxygen, however, less nitrogen and other atoms compared to synthetics. Surprisingly, NPs often show a molecular weight higher than 500 Da combined with a high polarity [7], which is in clear contrast to Lipinski's rule of five [48]. Nevertheless only about 10% of NPs contain two or more violations of Lipinski's rules [47]. In summary, natural chemistry can be seen as highly diverse scaffolds endowed with potential drugable pharmacophores.

6   Structure-based pharmacophore model

An inevitable prerequisite for generating a structure-based model is the knowledge about the ligand-target interaction [49] including the availability of the 3D structure of the target either by X-ray crystallography or NMR or constructed on the basis of the structure of homologous proteins. A unique platform containing 3D coordinates of experimentally solved protein structures is the Brookhaven Protein Data Bank (PDB [50]). A crystalline complex with a ligand bound to a protein's active site is the best requirement to start the construction of a structure-based 3D model. In this case, one may profit from the exact information of the ligand's bioactive conformation which is preserved in the binding site of the crystalline complex. The building of a structure-based pharmacophore is depicted in a step by step way in Figure 3.

A new software tool has recently been described for the successful generation of such chemical features-based models: The software LIGAND-SCOUT [51] is a program for ligand interpretation and data mining in the PDB. The performance of this program allows the detection of relevant interaction points between ligand and protein. The binding mode of the ligand in the active site of a protein can be visualized in a sophisticated way. LIGANDSCOUT's algorithms perform a stepwise interpretation of the ligand molecules: Planar ring detection, assignment of functional
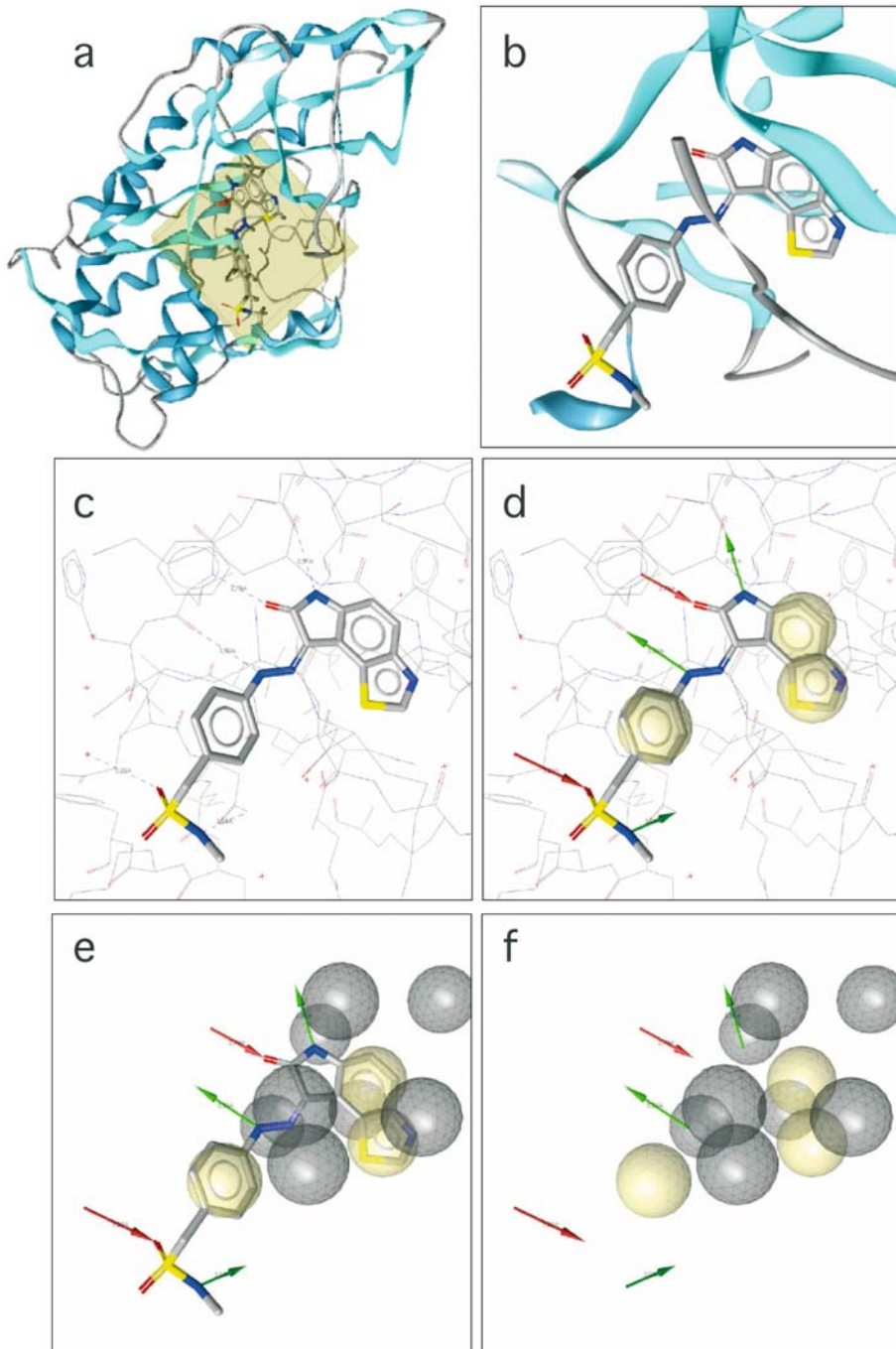
group patterns, determination of the hybridization state and finally the assignment of Kekulé pattern. The interpretation of the ligand molecules is the basis for the next step, an automated generation of pharmacophore models, derived from the data provided by a crystalline complex of the PDB. An automatic detection and classification of protein–ligand interactions into hydrogen bonds, charge transfer, and lipophilic regions leads to a collection of chemical features in a pharmacophore model. The graphical user-interface can provide an integrated view of protein, ligand, pharmacophore model, and interaction lines. In a previously published study, LIGANDSCOUT was used for the detection and interpretation of crucial interaction patterns between ligands and the factor Xa protein structure [52]. In a second step, the program CATALYST, a state of the art virtual screening platform, was used for rapid virtual screening of multi-conformational 3D structure databases. The information for the pharmacophore pattern (i.e., 3D coordinates of interaction points) was obtained by the interpretation of LIGANDSCOUT pharmacophore definitions and resulted in specific interaction models that were able to map the ligand in their bioactive conformation and to retrieve selectively a 78% fraction of the known factor Xa inhibitors within a small subset of the large Derwent World Drug Index library. A further application of the LIGANDSCOUT pharmacophore definitions covers the rationalized search for angiotensin converting enzyme (ACE)-2 inhibitors by virtual screening of approximately 3.8 million compounds from various commercial databases [53].

Figure 3

Concept for generating a structure-based pharmacophore model; Visualizing and calculation of chemical features using LigandScout [51]:

a. Protein (e.g., CDK2) complexed with a ligand {shown in ball-and-stick mode; N-methyl-{4-[2-(7-oxo-6,7-dihydro-8H-[1,3]thiazolo[5,4-E]indol-8-ylidene)-hydrazino]-phenyl}-methane-sulfonamide} in the active binding site (highlighted in the yellow cube)

b. Zoom up of the binding site with the ligand

c. Ligand with calculated distances to the interacting amino acid residues of the protein

d. Determination of interactions between the ligand and the target; evaluation and setting of chemical features (yellow sphere, hydrophobic feature; green arrow, hydrogen bond donor function; red arrow, hydrogen bond acceptor function)

e. Subtraction of the protein; the ligand, the chemical features and exclusion volumes (= grey spheres; representing areas not to be occupied by the ligand) are left

f. Subtraction of the ligand; the pharmacophore model remains comprising chemical features and exclusion volumes

Hit reduction and selection was achieved using a five feature hypothesis based on a recently resolved inhibitor-bound ACE2 crystal structure. Seventeen virtual hits were selected for their experimental validation in a bioassay; the concept was confirmed since all of them were revealed as ACE-2 inhibitors.

Barreca and co-authors developed a 3D structure-based pharmacophore model with LIGANDSCOUT for the discovery of new scaffolds acting as HIV-1 non-nucleoside reverse transcriptase inhibitors by virtual screening of large chemical databases. Six virtual hits were finally selected for determination of their inhibitory effects. Those belonging to the new scaffold class of the quinolin-2(1H)-one family exhibited reverse transcriptase inhibitory activity at sub-micromolar concentrations [54].

In a recently published work, Schuster et al. presented a so-called cytochrome P450 profiler [55]. Several structure-based (generated with LIGANDSCOUT) and ligand-based pharmacophore models (using CATALYST) for substrates and inhibitors of five cytochrome P450 isoenzymes (1A2, P450 2C9, P450 2C19, P450 2D6, and P450 3A4) were created and validated by the authors' group. Their results showed that the models were suitable for fast pharmacokinetic profiling of large drug-like databases.

In this context the **parallel screening** is of particular interest. Whereas in usual virtual screening cycles interactions of *thousands* or even *millions* of 3D database entries are browsed against *one* pharmacophore model, it is contrary in the case of parallel screening; low-energetic conformers of *one* structure are screened for their potential interactions against *numerous* models. The basics of parallel screening have just recently been presented by Steindl and co-authors [56, 57]. Furthermore, the authors exemplified this strategy for the activity profiling using a set of HIV protease pharmacophore models [58]. This *in silico* concept is of particular interest to virtually scrutinize drug candidates for their preliminary activity profiling relevant to putative side effects and toxicity [40]. According to the obtained interactions to virtually screened antitargets (e.g., hERG, sigma-1, sigma-2, alpha-1A, alpha-1B, alpha-1D, alpha-2A, alpha-2B, alpha-2C, D2L, D3, D4.2, 5-HT$_{1A}$, 5-HT$_{2A}$, 5-HT$_7$, H1, I2, A2A, A2B, cytochrome P 450) a first insight to potentially risky affinities is provided before time and cost intensive toxicological studies are performed.

The virtual screening approach using a structure-based pharmacophore model has revealed some first application examples in NP research: Niko-

lovska-Coleska and co-authors successfully pursued this *in silico* strategy in the area of X-linked inhibitors of apoptosis (XIAP) [59]. A high resolution 3D structure of the XIAP BIR3 domain complexed with the N-terminal end of the Smac/Diablo protein [60], which is an endogenous ligand of the respective XIAP binding pocket, was used as the starting point to virtually screen an in-house 3D-NP database. Embelin (**6**) from the Japanese Ardisia herb emerged as virtual small molecule weight hit, which was found to be a fairly potent inhibitor of XIAP using a fluorescence polarization binding assay.

In our group, we previously focused on acetylcholinesterase (AChE) [61]; according to the cholinergic hypothesis of the pathogenesis of Alzheimer's disease, inhibitors of the AChE are successfully used as therapeutic strategy. Based on the co-crystal structure of AChE with its ligand galanthamine, a structure-based pharmacophore model was generated and used for an *in silico* screening of a multi-conformational database consisting of more than 110,000 NPs. From the obtained hit list, promising, virtually active candidates were selected, namely scopoletin (**7**) and its glucoside scopolin (**8**). Their AChE inhibitory effect was first verified from the crude extract of *Scopolia carniolica* roots using a bioautographic TLC assay. The isolated coumarins showed a significant and dose-dependent inhibition of the AChE in the microplate enzyme assay as well as in the *in vivo* test. The i.c.v. application of both coumarins on rats resulted in a long-lasting, pronounced and – in case of the glucoside – even in a two-fold higher increase of the neurotransmitter's concentration than the one caused by the positive control galanthamine.

7   Ligand-based pharmacophore model

Very often, however, lead discovery projects have reached a well-advanced stage before detailed structural data on the protein target has become available, even though it is well recognized that modern methods of molecular biology together with biophysics and computational approaches enhance the likelihood of successfully obtaining detailed atomic structure information. A possible consequence is that often scientists identify and develop novel compounds for a target using preliminary structure-activity information, together with theoretical models of interaction. Only responses that
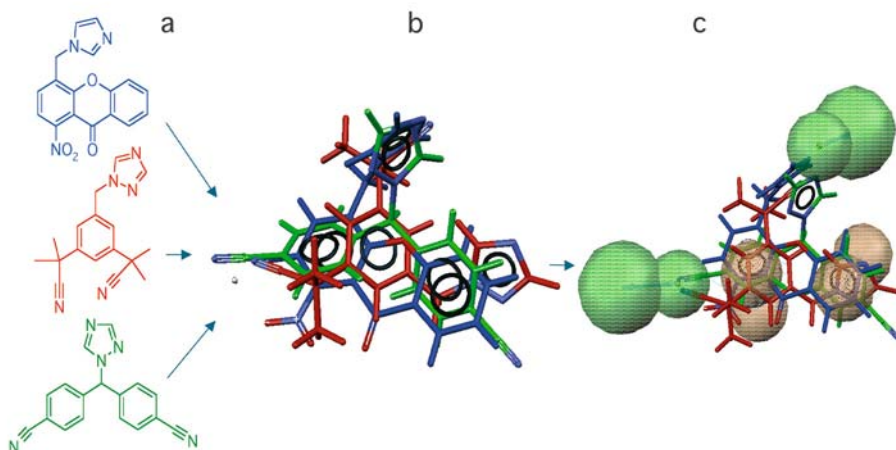
Figure 4.
Concept for generating a ligand-based pharmacophore model using the Catalyst program (Accelrys Inc., CA)
a. critical selection of active ligands
b. alignment of low-energetic conformers of the selected ligands
c. derivation and determination of common features

are consistent with the working hypotheses contribute to an evolution of the used models. Within this framework, the chemical feature-based pharmacophore approach has proven to be successful [62] allowing the perception and understanding of key interactions between a receptor and a ligand on a generalized level. A function-based pharmacophore represents the common ensemble of steric and electrostatic features of different compounds which are necessary for their interaction with a specific biological target structure (Fig. 4).

Such pharmacophore models together with large 3D structure databases originating either from in-house compound collections, from commercial vendors, or from natural products databases have proven to be extremely useful *in silico* screening experiments. When using ligand-based pharmacophore models as screening filters instead of protein 3D structures, affinity estimation is only based on geometric fit of compound atoms or groups to features of the model. In these cases, the values calculated are often far away from reality, however, still are useful for filtering possible

hits from non-binding molecules. Additionally, in pharmacophore fitting procedures, calculation demands are considerably lower than in docking algorithms allowing the number of compounds to be processed in the same time to be by far higher than even in high throughput docking.

Since in most of the studies no experimental information on either the biological conformation of the ligand or the target protein are currently available, the ligand-based chemical feature pharmacophore approach can provide essential information for medicinal chemists. Several successful applications within this subject have been performed using the CATA-LYST program, one of the leading software packages in chemical feature-based pharmacophore modeling. Schuster and co-workers succeeded in the identification of 11β-hydroxysteroid dehydrogenase type 1 inhibitors applying a common feature-based pharmacophore model for their virtual screening filtering experiments [63]. Similarly, the authors preceded by suggesting compounds with a proposed inhibition to the cytochrome P450 19 isoenzyme [64]. Several reviews covering successful applications of such feature-based methods have been published by Kurogi et al. [65], by Krovat et al. [28] and by Güner et al. [66]. They outline the theoretical background and describe several significant studies including 3D database search strategies.

In the field of NPs only a very limited number of studies report from the rationalized access to bioactive compounds *via* ligand-based virtual screening. For example, this method was pursued for the discovery of inhibitors of the COP9 signalosome (CNS) associated kinases CK2 and PKD [67]. Using NPs curcumin and emodin as lead structures, a virtual screening of an in-house database was carried out. Among the virtual hits seven NPs, e.g., anthraquinone (**9**) and piceatannol (**10**), were found to significantly induce apoptosis by inhibition of the CSN-associated kinases using *in vitro* and cell culture experiments. A further study has demonstrated the power of the ligand-based approach applied to pharmacophore modeling of sigma-1 ligands [68]. Therein, some reliable pharmacophore models could be extracted solely from ligand structure information. Compounds with potent affinities to the sigma-1 receptor known from literature were structurally aligned to derive distinct common features. Their 3D arrangement in combination with a spatial restriction was then used for the generation of a pharmacophore model, which was able to retrieve compounds with high affinity values, among them also NPs, like solanidine (**11**).

8   Discriminant analysis

Further ligand-based approaches use various forms of discriminant analysis, e.g., artificial **neural network** simulations. They are based on collections of mathematical models that are interconnected and organized in different layers. They are analogous to an adaptive human learning process and usually trained with learning sets applying one or more molecular descriptors in order to form clusters that enable to distinguish between different objects and their properties. The resulting models are then applied to make predictions on test sets, until the validated models may be used to derive a QSAR of chemically related structures or to mine larger datasets. One may distinguish between supervised and unsupervised (e.g., Kohonen network) learning methods as discussed in detail by Zupan and Gasteiger [69]. A successful application example within the field of NPs was published by Wagner et al. [70]. The authors used a dataset of 103 structurally diverse sesquiterpene lactones with known NF-κB inhibitory activity to derive a QSAR. By the application of multiple 3D structure representations as descriptors, a single model was achieved which provided detailed information on the structural influence of the investigated biological activity. Sangma and co-authors pursued a combination of two approaches to predict new inhibitors of the HIV-1 RT and HIV-1 PR from a NP database comprising metabolites from Thai medicinal plants. After a high throughput docking of the molecules into the target enzymes, self-organizing maps were generated to reduce the number of promising candidates to be tested [71].

A set of different *in silico* methodologies was previously applied by Cherkasov and co-authors to aid in the discovery of natural non-steroidal ligands for human sex hormone binding globulin [72]. Therein, a rigorously cross-validated neural network based QSAR model identified 105 prospective compounds from a structure collection of 23,836 commercial natural substances. This stringent QSAR ranking was combined with docking studies and pharmacophore-aided database search. The integrated computational methods resulted in a convincing predictive tool which identified a set of 29 structurally diverse NPs, of which every fourth compound was able to inhibit the target protein in a micromolar range.

Compounds of arbitrary structural diversity and with known activity against a target are particularly suitable not only for generating a ligand-

based pharmacophore model (as described before), but also for structure similarity studies using a **decision tree**. The object is to find as good a distinction as possible on the basis of a set of molecular descriptors, which identify molecular features shared by different subsets of active compounds and accordingly filter out compounds within the dataset in which these combinations are lacking. Using not only a simple logical description of one model, but an ensemble of decision trees tend to be the preferred option, since the consensus voting among trees give the approach higher predictive accuracy. One form of multiple decision trees well performed to virtually screen large 3D databases is Random Forest [73]. This chemoinformatic method was recently applied in a theoretical work performed by Ehrman and co-workers to predict ligands of multiple targets, like cyclooxygenase (COX), lipoxygenase (LOX), aldose reductase, HIV-1 enzymes etc., from a large dataset of Chinese herbs [74].

9   Databases

The advent of structure databases has provided a basis for the development and feasibility of automatic methods for the search of new lead structures. Conceptually, all the virtual screening concepts presented above have their origins in synthetic chemistry. Their application, however, is just as well adaptable to NPs' chemistry. Prior to the *in silico* filtering experiment, a 3D structure database requires an efficient generation of reasonable, energetically minimized conformations assumed to meet approximately those conformations that might be of biological relevance [75]. The underlying algorithms for 3D structure generation and conformation analysis are implemented in commercial software tools, e.g., in CORINA [21] or the CATALYST program (CATALYST, available from Accelrys Inc., San Diego, CA, USA; www.accelrys.com).

In the field of NPs the virtual screening application is mainly restricted due to the lack of searchable resources for structurally well defined natural compounds. In general, molecular databases with free access on the internet may comprise a high number of molecules, e.g., ChemBank (>1,100,000, http://chembank.broad.harvard.edu) or PubChem (>5,000,000; http://pubchem.ncbi.nlm.nih.gov); however, information about the number of contained natural molecules is rarely available. The library of the National

Cancer Institute (NCI) stores structural information of more than half a million compounds from both synthetic and natural origin that have been collected and tested by the NCI since 1955. About half of the synthetic compounds, which represent the large majority of the samples, may be used for free and are thus in the public domain. It is called the 'Open NCI Database' (Development Therapeutics Program NCI/NIH; http://dtp.nci.nih.gov/webdata.html). An interesting property prediction approach to the more than 250,000 compounds contained in this open database was provided by Poroikov and co-authors [76]. By use of the program PASS (Prediction of Activity Spectra for Substances) an *in silico* tool for complex searches of 565 different types of activities is provided; e.g., in the case of antineoplastic effects, the authors could demonstrate a substantial dataset enrichment over random selection by the use of PASS-predicted probabilities.

Libraries covering a major part of entities from nature (at least some thousands) or consisting of structural information exclusively from natural origin are not free of charge, e.g., the Traditional Chinese Medicinal Database (TCMD; http://tcm3d.com/services.htm [77]) or the Dictionary of Natural Product Database launched by Chapman & Hall (DNP; http://www.chemnetbase.com) providing chemical and physical data on some 200,000 natural compounds gathered from the world's chemical literature.

An excellent survey of public and commercial databases focusing on NPs has recently been published by Füllbeck and co-authors [78]. The authors provide information as to storing characteristics of the databases, web-addresses, total number of compounds and – if given – number of natural ones. In addition, a selection of suppliers and manufacturers of natural compounds and extracts are given. A new database is introduced by the authors (Super Natural Database [79]) storing information on available NPs, thus allowing the selection of compounds that can be purchased.

Moreover a number of non-commercial in-house created databases have been used from different groups for their virtual screening studies on NPs, e.g., a marine natural product database (MNDP [80]), a natural product database (NPD [61, 81]), a database based on the antique source '*de materia medica*' by Pedanius Dioscurides (DIOS [81]), or a database fed with metabolites of ethnopharmacologically known plants [82]. Recently, Ehrman and co-authors generated a 3D multiconformational database of

Chinese herbal constituents containing a total of more than 8,000 compounds from 230 Chinese herbs [83].

10 Integrated strategies for the discovery of bioactive NPs

The more or less accurate prediction of potentially active compounds by virtual screening has doubtlessly rationalized the early drug discovery process. These filtering experiments definitely assist in saving costly and time-intensive pharmacological assays, since the pool of predicted ligands (i.e., virtual hits) is usually drastically reduced compared to the initial amount of compounds (i.e., 3D-database). Demands to be made on a good model are selectivity and target-specificity on the one hand, but it is also seminal not to lose too many valuable ligands during the filtering process.

How far all of these demands can be fulfilled strongly depends on the quality of information used as the basis for generating the model and the algorithm underlying the virtual screening process. In medicinal chemistry, an activity prediction of 10–30% is usually regarded as satisfying enrichment. In NP research, however, this percentage may be too scarce. It is rarely found that a large set of natural compounds can be acquired so easily. Only a minority of secondary metabolites are commercially available – usually at incredibly high prices. Thus, extraordinary charges and efforts are typically necessary before a virtual hit from nature is available for pharmacological testing. This process embraces the acquisition of the natural material described to contain the desired metabolite to the point of phytochemical analysis and isolation. Though advanced separation techniques, analytical instrumentation, and innovative tools for structure identification are at the phytochemists' disposal, it remains a complex and sometimes uncertain endeavor. This is why the results obtained from *in silico* predictions may nevertheless be too vague for a NP researcher.

Methods are asked to further increase the probability of following the straight tip. There is the possibility to hyphenate sundry computational approaches, e.g., pharmacophore-based virtual screening combined with docking of the resulting virtual hits, or to consider only the consensus hits applying two or more screening concepts. Nevertheless all these strate-

gies remain virtual and speculative. The combination of two approaches, which are completely divergent in nature, like a computational and an empirical one may however offer a more deepened access to bioactive NPs and may sometimes help to avoid a distorted view.

Thus, the computer-aided molecular selection is best combined with further discovery methods, labeled as integrated approaches, to increase the probability in finding a real hit. In traditional pharmacognosy there are some well established methods in targeting this aim starting from a holistic level. These include (i) hints from ethnopharmacology, (ii) phenomenological effects registered after application of naturally derived preparations, (iii) guidance of chemotaxonomy, (iv) phylogenetic selection criteria, or (v) simply information gathered from a high/medium throughput screening of extracts. In a recently published review from our group, different strategies in the field of NPs have been presented with special emphasis on anti-inflammatory NPs interacting within the arachidonic cascade [84]. Integrated computational strategies for the discovery of natural bioactive compounds have been introduced elsewhere concentrating on their scope, strengths and limits [85].

Some strategies and examples from literature combining virtual screening approaches and classical methods for activity exploitation are outlined below.

10.1 Strategy A (Fig. 5)

As soon as a sensitive data-mining tool has been developed and has proved itself by more or less selectively finding the active compounds within a test set, it can be applied for screening a 3D multi-conformational database. The subsequent procedure consists of the evaluation of the virtual hits considering physicochemical properties, toxicity and pharmacokinetics. In this stage additional virtual filtering tools for the profiling of ADME parameters [86] might have an invaluable impact to aid a refined selection of compounds. Then, a sensible choice of natural materials known to contain the focused metabolites and worth investigating in detail is a crucial step. It requires a comprehensive study in literature considering the hit content in the natural source, its availability and maybe hints from ethnopharmacology.
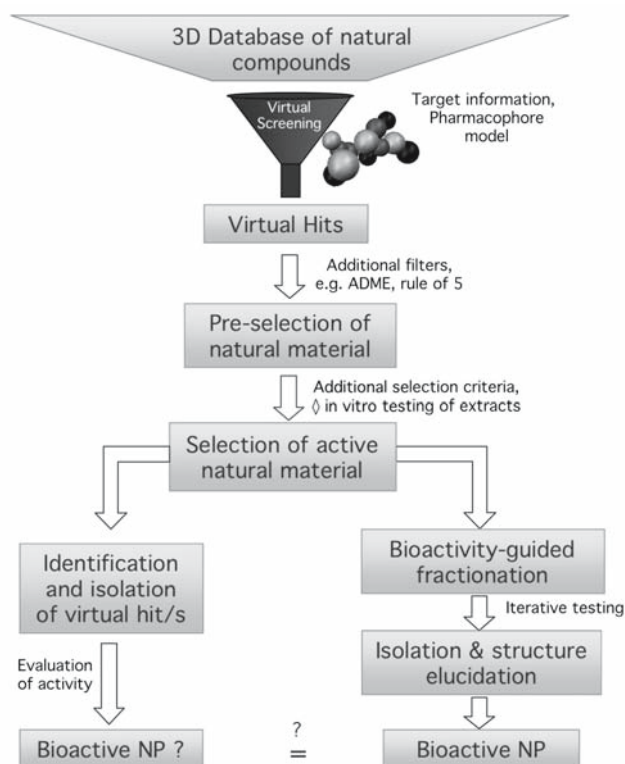
Figure 5.
Strategy A for the discovery of bioactive NPs using an integrated virtual screening approach

Once some natural materials are selected, it is advisable to perform a preliminary assay with those crude extracts and fractions assumed to contain the promising metabolite/s. Though being aware that in case of small hit amounts present in the natural material the activity may be overseen. Therefore, it is advisable to first identify the promising constituent and to possibly enrich it in the extract to be tested. Those samples that scored well are then subjected to phytochemical investigations. In this way, the tricky selection of the natural material turns from a bold venture to a more rationalized endeavor. As soon as a promising (i.e., active) starting material is found, there are in principle two possible strategies to embark on: The first one relies more on the *in silico* approach and focuses directly on

the identification of the initially obtained virtual hits within the natural matrix applying analytical tools, like LC-MS or LC-NMR, GC-MS etc. In a straightforward manner the hits are isolated using different chromatographic separation steps. After structural confirmation the compounds are then tested to hopefully verify the predicted activities. This strategy is very goal-oriented, since only pharmacological assays for the finally isolated virtual hits are necessary. On the other side, one may run the risk of ignoring further active NPs not necessarily fitting into the pharmacophore model.

The second strategy focuses on a bioactivity-guided fractionation irrespective of the virtual hits used for the selection of the starting material. Following the concept, the finally isolated active ingredients should correspond to the predicted virtual hits. This approach is usually associated with higher phytochemical efforts and costs, because it requires an iterative testing of all arising fractions and sub-fractions. For the evaluation of all the bioactive constituents in detail and for the discovery of possibly unknown metabolites this procedure is however indispensable.

The decision, which of the presented ways is the more appropriate for the investigation at hand, strongly depends on the reliability and selectivity of the used pharmacophore model, and the costliness of the used assay.

The strategy schematized in Figure 5 was recently applied to a medicinal plant with anti-inflammatory potential known from ethnopharmacological sources [87]. From the pharmacophore based virtual screening filtering experiment a number of secondary metabolites known from the mulberry tree complied with all the models' requirements, thus revealed as virtual hits. Indeed, *in vitro* tests attested extracts of *Morus* root bark a distinct COX inhibitory potential. The objective was to find the active principles from this plant material applying both different methods for their discovery. First, the computer-aided approach was used to identify the virtually active compounds able to interact with the pharmacophore models for COX-1 and -2. Second, the bioactivity-guided fractionation was conducted for the isolation of the COX-inhibiting constituents. This resulted in the isolation of nine compounds belonging to the chemical classes of sanggenons and moracins. In the enzyme assay, all the isolates showed moderate to potent inhibitory effects on COX-1 and -2. When comparing the hits of the virtual screening with the experimental data, a good correlation between predictions provided by the computer assisted method and *in vitro* data

could be obtained in the case of the isolated sanggenons (e.g., sanggenon C; **12**). However, this agreement could not be achieved with the moracins (e.g., moracin M; **13**). In any case the virtual screening was particularly helpful for the decision regarding which plant material is worth extensive study. Furthermore, the disclosed interactions of the sanggenons with the pharmacophore model – miming the binding site of the target – provided us with some essential information about the molecular requirements of COX-ligands.

10.2 Strategy B

A different integrated procedure is schematized in Figure 6. Applying this approach, the pre-selection of the natural material is not guided by virtual prediction; but a number of extracts is roughly screened with a bioassay to identify the active ones. A similar strategy is to collect information about the traditional application of natural preparations in the field of the focused pharmacological target. A 3D database is then generated consisting of all the metabolites known from literature to be included in that extract/s that came off well. Likewise, ethnopharmacological knowledge about useful preparations from nature may guide the selection of NPs. The resulting biased database is virtually screened with an established pharmacophore model of the aiming target.

The impact of ethnopharmacology has been analyzed in a previous study from our group; there we investigated the statistical evidence considering hints from folk medicine for the discovery of anti-inflammatory NPs utilizing pharmacophore-based virtual screening techniques [81]. COX-1 and -2 were used as preferential targets, since they are key enzymes in the inflammation process. Dioscorides' *de materia medica*, which was written in the 1st Century AD, was used as the ethno-pharmacological source. Secondary metabolites of those medicinal plants, which Dioscorides described as active against fever, rheumatism, pain and pus were stored in a multi-conformational 3D database. This was virtually screened against the validated pharmacophore models. The resulted hit list was analyzed and compared with those obtained by screening unbiased databases of natural as well as of synthetic origin. The effectiveness of an ethnopharmacological approach could be statistically demonstrated by obtaining a significantly
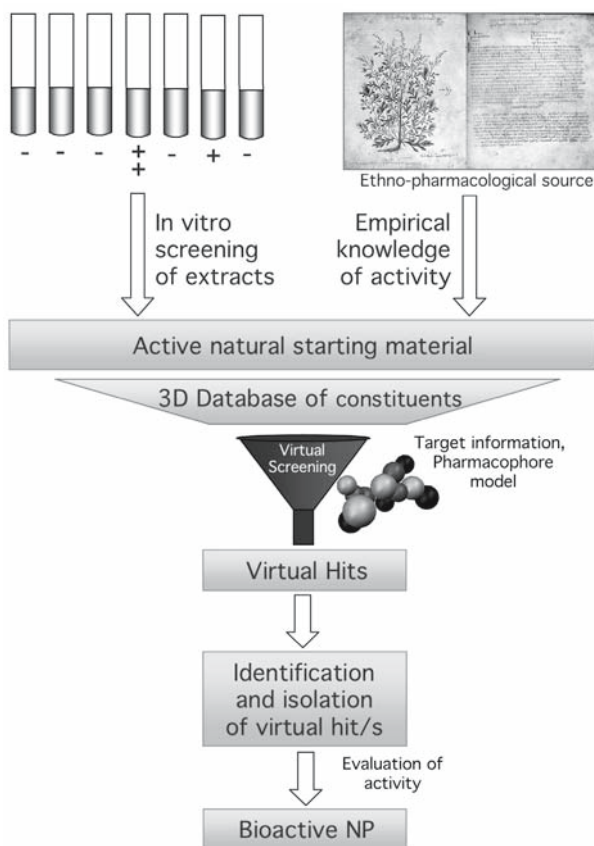
Figure 6.
Strategy B for the discovery of bioactive NPs using an integrated virtual screening approach

higher hit rate compared to the hit rates of the unbiased natural as well as synthetic databases.

Following this strategy the putative hits may then be identified by modern analytical tools like LC-MS or LC-NMR to isolate them from the natural matrix in a target-oriented way for pharmacological testing. This approach is especially helpful for intricate pharmacological assays, which would turn a bioguided fractionation into an unrealistic endeavor.

A combination of an ethnopharmacologically based pre-selection of plant material and a computational approach was reported by Bernard

and co-workers, who used this strategy to rationalize a phytochemical lead discovery [88]. Starting with an *in vitro* screening on phospholipase A_2 performed with traditionally used anti-inflammatory plant extracts, a focused structural database was generated and virtually screened on an established ligand-based pharmacophore model for human non-pancreatic phospholipase A_2. The combination of experimental data with database exploitation and molecular modeling resulted in the efficient identification of betulin (**14**) and betulinic acid (**15**) as extract ingredients with distinct anti-inflammatory *in vitro* effects.

The combination of the two different, but complementary strategies consisting of *in vitro* screens and *in silico* assessment has recently been described by van de Waterbeemd [89]. He labeled this method as '*in combo*' approach and used it for the straight forward access of various ADME properties. The application of the '*in combo*' approach for the discovery of NPs has recently been tested in our group by the search of natural acetylcholinesterase inhibitors [90]. In a medium-sized throughput screening about 100 plant extracts were investigated using an acetylcholinesterase enzyme test. From the sample showing the best inhibitory activity, all the known secondary metabolites were fed into a small 3D multiconformational database and subsequently subjected to a virtual screening on a generated pharmacophore model. The efficacy of this procedure could be confirmed by the isolation of the obtained virtual hits, i.e., 8-deoxylactucin (**16**) and lactucopicrine (**17**). They showed a significant and dose-dependent inhibitory effect in the enzyme assay.

Methods and expectations of this integrated virtual screening concept have previously been discussed in detail by J. Bajorath [91, 92] with the author's final statement that '*a meaningful integration of virtual and experimental screening programs, together with lessons to be learned from structural genomics, holds great promise for more rapid and consistent identification of high quality hits or leads across divers classes of therapeutic targets*'. Though this conclusion was not particularly coined to NPs, it comes especially true in the rich world of secondary metabolites.

Further hybridized computational strategies are quite sensible to get an improved understanding of ligand-target interactions. In the following two examples docking protocols helped enlighten the molecular mechanism of bioactive natural compounds. Chimenti and co-authors isolated quercetin (**18**) among other secondary metabolites from the Mediterra-

nean shrub *Hypericum hircinum* and identified this flavonol as selective inhibitor of the MAO-A with an activity in the nanomolar range (IC_{50} = 10 nM) [93]. For a more comprehensive understanding of the underlying molecular selectivity, conformation analysis and docking simulations were performed using the most recent crystallographic structures of both human isoforms MAO-A and MAO-B. This enabled the authors to identify the most important interactions between the residues and the cofactor within the enzymatic cleft. The estimated free energies of complexation were in agreement with experimental data and confirmed the distinct preference for the MAO-A cleft with more intermolecular hydrogen bonds and π-π interactions.

The goal of a recent in-house study was to rationalize the binding interaction of the protoalkaloid taspine (**19**) within acetylcholinesterase. Taspine was isolated in a bioactivity-guided manner from *Magnolia x sou-langiana* and revealed as selective inhibitor of acetylcholinesterase with a significantly higher effect than the positive control galanthamine (**20**; IC_{50} = 0.33 ± 0.07 μM). Extensive molecular docking studies were performed with human and Torpedo californica-acetylcholinesterase employing Gold software (Vers. 3.1; www.ccdc.cam.ac.uk/products/life_sciences/gold/). The results suggested taspine to bind in an alternative binding orientation than galanthamine [94]. While this is located in close vicinity to the catalytic amino acid triad, taspine was found to be mainly stabilized by sandwich-like π-stacking interactions in the aromatic gorge of the enzyme.

In both case studies the active natural compound was already identified. Thus, the *in silico* tool was not employed for data mining, but to elicit the putative binding mode in the macromolecular target. Docking simulations turned out to be excellent tools to get an idea about the assumed molecular ligand target interaction.

10.3   Strategy C

Another approach capitalizes exactly on the just-mentioned observation that computational predictions may reveal an idea about the interaction to a specific target's binding site. Thus, it is possible to start with one compound of unknown activity and to mine it against a number of structurally disclosed targets in terms of elaborated pharmacophore models (Fig. 7), i.e.,
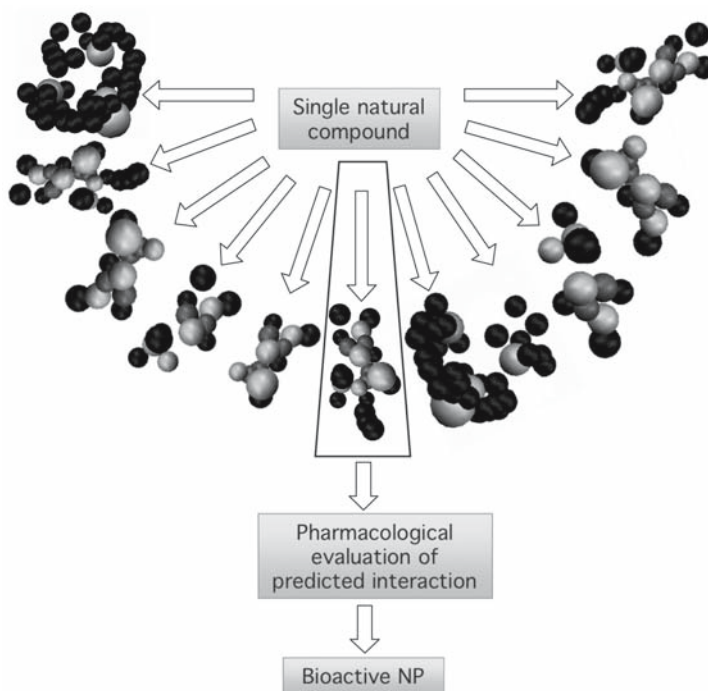
Figure 7.
Parallel screening for the discovery of bioactive NPs

parallel *in silico* screening (see previous). As soon as the orphaned molecule is able to comply with all the requirements and restrictions imposed by any model, it can be assessed as rational hint. Consequently, the focused compound will be subjected to a pharmacological testing on the predicted target/s. In this way, the parallel screening is not only helpful to estimate the interactions of a drug candidate with diverse antitargets; or to canvass its interactions to related targets as is performed for an activity profiling. In this approach, the parallel screening is a computational tool for *target fishing* to get a rational idea about any potential target interaction and to prioritize a few targets for experimental evaluation by applying simple ligand-based or target-based queries. The potential of virtual screening of target libraries was recently discussed by Didier Rognan [95]. In his group a structure-based method for target screening was pursued applying inverse

docking [96]. The authors used 2,148 structurally well-defined PDB entries to build a 3D protein library. The virtual screening of this protein library with four unrelated ligands was suitable for recovering the true targets of specific ligands and may as well be used for virtual selectivity profiling of any ligand of interest.

Nettles and co-authors performed the *target fishing* approach using a ligand-based procedure [97]. The potential of both 2D and 3D chemical descriptors were compared as tools for predicting the biological targets of ligand probes on the basis of their similarity to reference molecules in a chemical database comprising 46,000 biologically annotated compounds. The ligand-based 3D tool FEPOPS (FEature POint PharmacophoreS), which provides pharmacophoric alignment of the small molecules' chemical features consistent with those seen in experimental ligand/receptor complexes, was used for scaffold jumping within the screened database. Using ATP the authors were able to identify the natural compound balanol (21) as ligand of CDK2.

The highest effort applying this strategy is the availability of a representative amount of reliable pharmacophore models covering a wide range of relevant targets (Fig. 7). Thus, it may be of particular interest to focus on one pathological syndrome, e.g., obesity, inflammation, apoptosis etc., where a phenomenological activity of a NP is already evident. Applying this approach the disposition of pharmacophore models for targets involved in the respective pathological complex is easier to manage. In this way, a goal-oriented strategy may help to bridge the gap between a phenomenological effect and the underlying molecular mode of action.

11 Conclusion

Pertaining to the drug discovery from nature we are facing two facts: (i) statistics show that the myriad of structurally diverse natural compounds are the most favored source of new drugs for clinical use [5]; (ii) the drug discovery process has moved towards more rational concepts based on the increasing understanding of the molecular principles of protein–ligand interactions. Spurred on by economic interest fundamental advances have been made in research applying data mining strategies, like virtual screening.
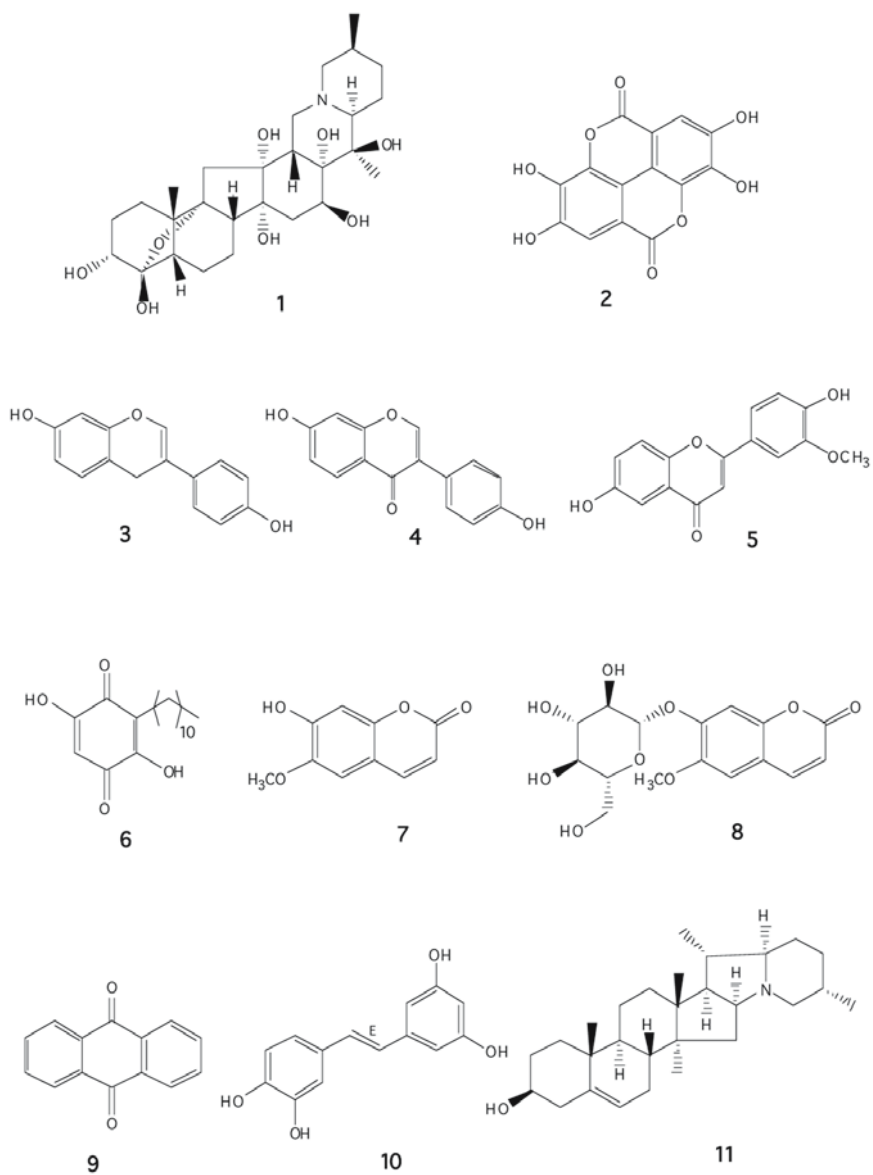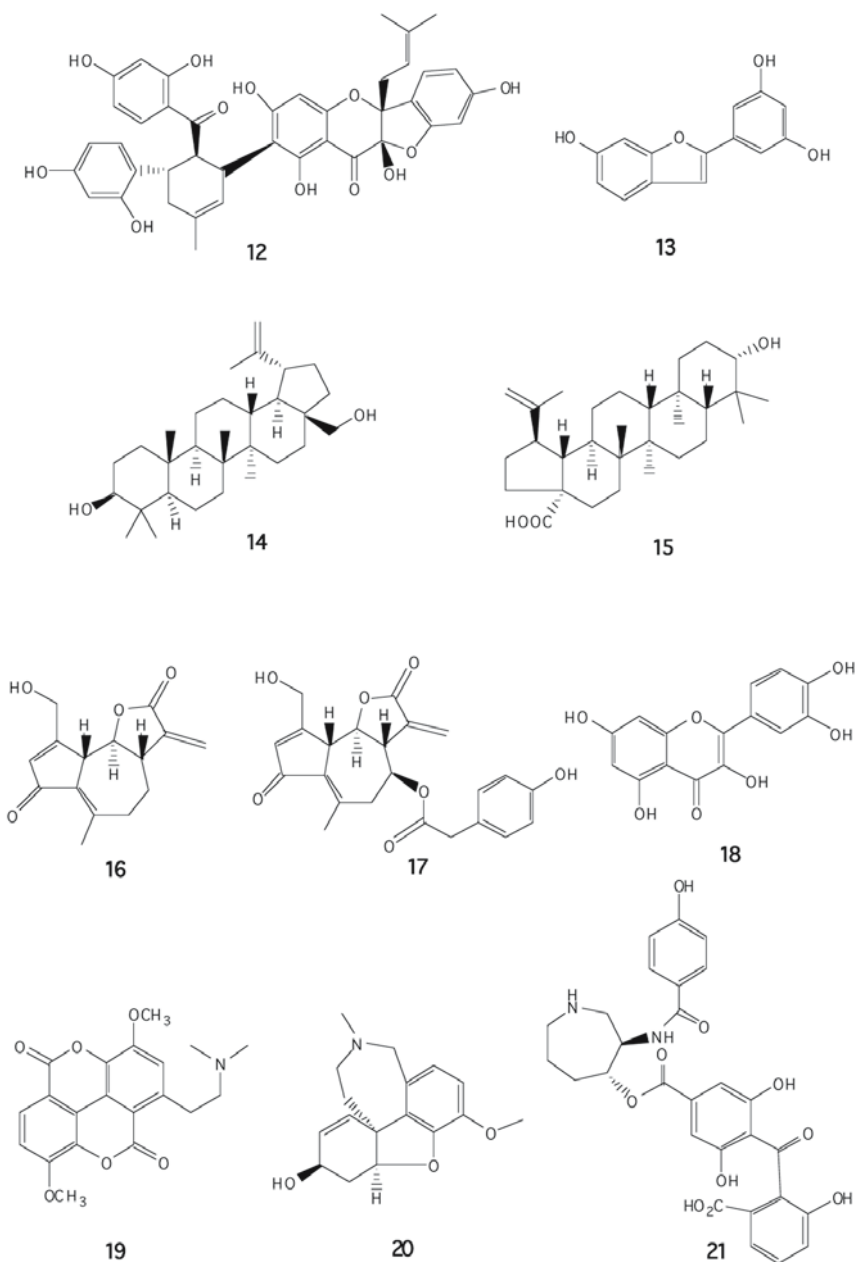
Chart 1.
Structures 1–11

Chart 2.
Structures 12–21

Though being aware of both potentials, their combined benefit could only rudimentary be savored. Only limited attempts applying innovative *in silico* tools in NP research are pursued so far, because the search for bioactive compounds is a complex and multidisciplinary challenge. Thus, a sensible adaptation of computational strategies is in demand to profit in an economic way from the unique chemical and biological diversity associated with NPs. Virtual screening techniques, however, must not be used exclusively as activity-predicting tools, since the results provide merely an indication for a putative activity: it is only by the creation of interfaces between computational tools and well-established methods from pharmacognosy that a reasonable standard of success can be achieved. The search for the most effective strategy is best performed by a drug discovery process that involves the exploitation of all the information which can be gathered from bioactivity-guided fractionation, on-line analytical activity profiling, ethnopharmacological screening, chemoinformatics, virtual and *in vitro* screening studies. In the first instance it behoves modern pharmacognosy to skillfully exploit knowledge from all these fields because it is of paramount importance to sift through the enormous wealth of NPs.

Examples underlining the impact of virtual screening on the identification of active NPs have been presented in this survey. Though the full potential in this field is by far untapped, these early results indicate that the integrated virtual screening approaches are target-oriented and trendsetting strategies. However, as any computer-based technique, the successful use of virtual screening will entirely depend on the way it is utilized and the quality of its underlying experimental data. The advantages implemented to a virtual screening cycle compared to a conventional *in vitro* screening are obvious: (i) higher capacity, (ii) no need for isolated compounds, (iii) less experimental efforts for testing; (iv) theoretically, interactions of all known NPs to all structurally defined targets can be calculated and predicted, (v) the quality of hit compounds can be increased by additional drug-like filters and virtually restricted ADME properties; thus diminishing failures in the early drug development.

Nevertheless experimental investigations are seminal, but can be focused in a more effective fashion. A cautious handling of virtual hits together with lessons learned from traditional pharmacognosy seems to be crucial for a successful exploitation of treasures from nature. In this area,

virtual screening will most likely play an essential role in accelerating the early stage of drug discovery by efficiently digging out lead compounds from nature.

References

1   Brown D, Superti-Furga G (2003) Rediscovering the sweet spot in drug discovery. *Drug Discov Today* 8: 1067–1077

2   Drews J (2003) Strategic trends in the drug industry. *Drug Discov Today* 8: 411–420

3   Smith A (2002) Screening for drug discovery: the leading question. *Nature* 418: 453–459

4   Strohl WR (2000) The role of natural products in a modern drug discovery program. *Drug Disc Today* 5: 39

5   Newman DJ, Cragg GM (2007) Natural products as sources of new drugs over the last 25 years. *J Nat Prod* 70: 461–477

6   Harvey A (2000) Strategies for discovering drugs from previously unexplored natural products. *Drug Disc Today* 5: 294–300

7   Clardy J, Walsh C (2004) Lessons from natural molecules. *Nature* 432: 829–837

8   Jones WP, Chin YW, Kinghorn AD (2006) The role of pharmacognosy in modern medicine and pharmacy. *Curr Drug Targets* 7: 247–264

9   Dictionary of Natural Products provided by Chapman &Hall/CRC. Available at: http://www.crcpress.com/shopping_cart/products/product_detail.asp?sku=C9150 (accessed in January 2007)

10   Tulp M, Bohlin L (2005) Rediscovery of known natural compounds: Nuisance or goldmine? *Bioorg Med Chem* 13: 5274–5282

11   Sitte P, Weiler EW, Kadereit JW, Bresinsky A, Körner C (2002) *Strasburger – Lehrbuch der Botanik*, 35. Auflage, Spektrum Akademischer Verlag, Heidelberg, Berlin, 339–351

12   Balandrin M (1993) Plant-derived natural products in drug discovery and development: an overview. *Amer Chem Soc Symposium Series* No 534: 2–12

13   Cragg GM, Newman DJ, Snader KM (1997) Natural products in drug discovery and development. *J Nat Prod* 60: 52–60

14   Henkel T, Brunne RM, Muller H, Reichel F (1999) Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew Chem Int Ed* 38: 643–647

15   Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar M, Doyle M, FitzHugh W et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921

16   Deisenhofer J, Smith JL (2001) Proteins. *Curr Opin Struc Biol* 11: 701–702

17   Maggio ET, Ramnarayan K (2001) Recent developments in computational proteomics. *Drug Disc Today* 6: 996–1004

18   Potterat O, Hamburger M (2006) Natural products in drug discovery – concepts and approaches for tracking bioactivity. *Curr Org Chem* 10: 899–920

19   Pottereat O (2006) Targeted approaches in natural product lead discovery. *Chimia* 60: 19–22

20   Morrell J (1996) Mining information from databases for drug discovery. *Book Abstr 211th ACS National Meeting* CINF–052

21   Gasteiger J, Teckentrup A, Terfloth L, Spycher S (2003) Neural networks as data mining tools in drug design. *J Phys Org Chem* 16: 232–245

22   Xu J, Hagler A (2002) Chemoinformatics and drug discovery. *Molecules* 7: 566–600

23   Alvarez AJ, Shoichet B (2005) *Virtual screening in drug discovery*. Taylor & Francis, CRC-Press, Boca Raton

24   Langer T, Hoffmann RD (2006) Pharmacophores and pharmacophore searches. In: *Methods and principles in medicinal chemistry* Vol. 32, Wiley-VCH, Weinheim

25   Manly CJ, Louise-May S, Hammer JD (2001) The impact of informatics and computational chemistry on synthesis and screening. *Drug Discov Today* 6: 1101–1110

26   Langer T, Hoffmann RD (2001) Virtual screening: an effective tool for lead structure discovery? *Curr Pharm Des* 7: 509–527

27   Böhm HJ, Schneider G (2000) *Virtual screening for bioactive molecules*. Wiley, New York

28   Krovat EM, Steindl T, Langer T (2005) Recent advances in docking and scoring. *Curr Comput-Aided Drug Des* 1: 93–102

29   Abagyan R, Totrov M (2001) High-throughput docking for lead generation. *Curr Opin Chem Biol* 5: 375–382

30   Schneider G, Böhm HJ (2002) Virtual screening and fast automated docking methods. *Drug Disc Today* 7: 64–70

31   Shoichet BK, McGovern SL, Wei B, Irwin JJ (2002) Lead discovery using molecular docking. *Curr Opin Chem Biol* 6: 439–446

32   Stahl M, Rarey M (2001) Detailed analysis of scoring functions for virtual screening. *J Med Chem* 44: 1035–1042

33   Charifson PS, Corkery JJ, Murcko MA, Walters WP (1999) Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 42: 5100–5109

34   Wang R, Wang S (2001) How does consensus scoring work for virtual library screening? An idealized computer experiment. *J Chem Inf Comput Sci* 41: 1422–1426

35   Liu B, Zhou J (2005) SARS-CoV protease inhibitors design using virtual screening method from natural products libraries. *J Comp Chem* 26: 484–490

36   Toney JH, Navas-Martin S, Weiss SR, Koeller A (2004) Sabadinine: a potential non-peptide anti-severe acute-respiratory-syndrome agent identified using structure-aided design. *J Med Chem* 47: 1079–1080

37   Cozza G, Bonvini P, Zorzi E, Poletto G, Pagano MA, Sarno S, Donella-Deana A, Zagotto G, Rosolen A, Pinna LA et al (2006) Identification of ellagic acid as potent inhibitor of protein kinase CK2: a successful example of a virtual screening application. *J Med Chem* 49: 2363–2366

38   Zhao L, Brinton RD (2005) Structure-based virtual screening for plant-based ERβ-selective ligands as potential preventative therapy against age-related neuro-degenerative diseases. *J Med Chem* 48: 3463–3466

39   Liu H, Li Y, Song M, Tan X, Cheng F, Zheng S, Shen J, Luo X, Ji R, Yue J et al (2003) Structure-based discovery of potassium channel blockers from natural products virtual screening and electrophysiological assay testing. *Chem Biol* 10: 1103–1113

40   Langer T, Hoffmann RD (2006) Pharmacophore modelling: applications in drug discovery. *Exp Opin Drug Discov* 1: 261–267

41   Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, Stallings WC, Connolly DT, Shoichet BK (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* 45: 2213–2221

42   Samiulla DS, Vaidyanathan VV, Arun PC, Balan G, Blaze M, Bondre S, Chandrasekhar G, Gadakh A, Kumar R, Kharvi G et al (2005) Rational selection of structurally diverse natural product scaffolds with favorable ADME properties for drug discovery. *Mol Divers* 9: 131–139

43   Testa B (1984) Drugs? Drug research? Advances in drug research? Musings of a medicinal chemist. *Adv Drug Res* 13: 1–58

44   Smith DA, van de Waterbeemd H, Walker DK (2006) *Pharmacokinetics and metabolism in drug design*. 2nd Ed., Wiley-VCH, Weinheim

45   Avdeef A, Testa B (2002) Physicochemical profiling in drug research: a brief survey of the state-of-the-art of experimental techniques. *Cell Mol Life Sci* 59: 1681–1689

46   Feher M, Schmidt JM (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci* 43: 218–227

47   Stahura F, Godden JW, Ling X, Bajorath J (2002) Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations. *J Chem Inf Comput Sci* 40: 1254–1252

48   Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23: 3–25

49   Congreve M, Murray CW, Blundell TL (2005) Structural biology and drug discovery. *Drug Disc Today* 10: 895–907

50   Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P (2000) The protein data bank. *Nucleic Acids Res* 28: 235–242

51   Wolber G, Langer T (2000) LigandScout: 3D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model* 45: 160–169

52   Krovat EM, Frühwirth KH, Langer T (2005) Pharmacophore identification, *in silico* screening, and virtual library design for inhibitors of the human factor Xa. *J Chem Inf Model* 45: 146–159

53   Rella M, Rushworth C, Guy JL, Turner AJ, Langer T, Jackson RM (2006) Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *J Chem Inf Model* 46: 708–716

54   Barreca ML, De Luca L, Iraci N, Rao A, Ferro S, Maga G, Chimirri A (2007) Structure-based pharmacophore identification of new chemical scaffolds as non-nucleoside reverse transcriptase inhibitors. *J Chem Inf Model* 47: 557–562

55   Schuster D, Laggner C, Steindl TM, Langer T (2006) Development and validation of an *in silico* P450 profiler based on pharmacophore models. *Curr Drug Discov Technol* 3: 1–48

56   Steindl TM, Schuster D, Laggner C, Langer T (2006) Parallel screening: a novel concept in pharmacophore modelling and virtual screening. *J Chem Inf Model* 46: 2146–2157

57   Steindl TM, Schuster D, Wolber G, Laggner C, Langer T (2007) High throughput structure-based pharmacophore modeling as a basis for successful parallel virtual screening. *J Comput-aided Mol Des* ASAP doi 10.1007/s10822–006–9066–y

58   Steindl TM, Schuster D, Laggner C, Chuang K, Hoffmann R, Langer T (2007) Parallel screening and activity profiling with HIV protease inhibitor pharmacophore models. *J Chem Inf Model* 47: 563–571

59   Nikolovska-Coleska Z, Xu L, Hu Z, Tomita Y, Li P, Roller PP, Wang R, Fang X, Guo R, Zhang M et al (2004) Discovery of embelin as a cell-permeable, small-molecular weight inhibitor of XIAP through structure-based computational screening of a traditional herbal medicine three-dimensional structure database. *J Med Chem* 47: 2430–2440

60   Wu G, Chai J, Suber TL, Wu JW, Du C, Wang X, Shi Y (2000) Structural basis of IAP recognition by Smac/DIABLO. *Nature* 408: 1008–1012

61   Rollinger JM, Hornick A, Langer T, Stuppner H, Prast H (2004) Acetylcholinesterase inhibitory activity of scopolin and scopoletin discovered by virtual screening of natural products. *J Med Chem* 47: 6248–6254

62   Langer T, Krovat EM (2003) Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr Opin Drug Discov Dev* 6: 370–376

63   Schuster D, Maurer E, Laggner C, Nashev L, Wilckens T, Langer T, Odermatt A (2006) The discovery of new 11β-hydroxysteroid dehydrogenase Type 1 inhibitors by common feature pharmacophore modeling and virtual screening. *J Med Chem* 49: 3454–3466

64   Schuster D, Laggner C, Steindl TM, Palusczak A, Hartmann RW, Langer T (2006) Pharmacophore modeling and *in silico* screening for new P450 19 (aromatase) inhibitors. *J Chem Inf Model* 46: 1301–1311

65   Kurogi Y, Güner OF (2001) Pharmacophore modeling and three-dimensional database searching for drug design using Catalyst. *Curr Med Chem* 8: 1035–1055

66   Güner O, Clement O, Kurogi Y (2004) Pharmacophore modeling and three dimensional database searching for drug design using CATALYST: Recent advances. *Curr Med Chem* 11: 763–771

67   Füllbeck M, Huang X, Dumdey R, Frommel C, Dubiel W, Preissner R (2005) Novel curcumin- and emodin-related compounds identified by *in silico* 2D/3D conformer screening induce apoptosis in tumor cells. *BMC Cancer* 5: 97

68   Laggner C, Schieferer C, Fiechtner B, Poles G, Hoffmann RD, Glossmann H, Langer T, Moebius F (2005) Feature based pharmacophore models for sigma1 receptor, ERG2 and EBP. *J Med Chem* 48: 4754 –4764

69   Zupan J, Gasteiger J (1999) *Neural networks in chemistry and drug design*, 2nd Ed. Wiley-VCH, Weinheim

70   Wagner S, Hofmann A, Siedle B, Terfloth L, Merfort I, Gasteiger J (2006) Development of a structural model for NF-κB inhibition of sesquiterpene lactones using self-organizing neural networks. *J Med Chem* 49: 2241–2252

71   Sangma C, Chuakheaw D, Jongkon N, Saenbandit K, Nunrium P, Uthayopas P, Hannongbua S (2005) Virtual screening for anti-HIV-1 RT and anti-HIV-1 PR inhibitors from the Thai Medicinal Plants Database: A combined docking with neural networks approach. *Comb Chem HTS* 8: 417–429

72   Cherkasov A, Shi Z, Fallahi M, Hammond GL (2005) Successful *in silico* discovery of novel nonsteroidal ligands for human sex hormone binding globulin. *J Med Chem* 48: 3203–3213

73   Svetnik V, Liaw A, Tong D, Culberson C, Sheridan RP, Feuston BP (2003) Random Forest: a classification and regression tool for compound classification and QSAR Modeling. *J Chem Inf Comput Sci* 43: 1947–1958

74   Ehrman TM, Barlow DJ, Hylands PJ (2007) Virtual screening of Chinese herbs with random forest. *J Chem Inf Model* ASAP 10.1021/ci600289v

75   Kirchmair J, Laggner C, Wolber G, Langer T (2005) Comparative analysis of protein-

bound ligand conformations with respect to catalyst's conformational space subsampling algorithms. *J Chem Inf Model* 45: 422–430

76   Poroikov VV, Filimonov DM, Ihlenfeldt WD, Gloriozova TA, Lagunin AA, Borodina YV, Stepanchikova AV, Nicklaus MC (2003) PASS Biological activity predictions in the enhanced open NCI database browser. *J Chem Inf Comput Sci* 43: 228–236

77   Lu A, Liu B, Liu H, Zhou J, Xie G (2004) A traditional Chinese medicine plant-compound database aid its application for searching. *Int Electron J Mol Des* 3: 672–683

78   Füllbeck M, Michalsky E, Dunkel M, Preissner R (2006) Natural products: sources and databases. *Nat Prod Rep* 23: 347–356

79   Dunkel M, Füllbeck M, Neumann S, Preissner R (2006) SuperNatural: a searchable database of available natural compounds. *Nucleic Acid Res* 34: D678–683

80   Lei J, Zhou J (2002) A marine natural product database. *J Chem Inf Comp Sci* 42: 742–748

81   Rollinger JM, Haupt S, Stuppner H, Langer T (2004) Combining ethnopharmacology and virtual screening for lead structure discovery: COX-inhibitors as application example. *J Chem Inf Comp Sci* 44: 480–488

82   Bernard P, Berton JY, Chrétien JR (1999) Computer-aided molecular selection and design of natural bioactive molecules. *Curr Opin Drug Disc Dev* 2: 213–223

83   Ehrman TM, Barlow DJ, Hylands PJ (2007) Phytochemical databases of Chinese herbal constituents and bioactive plant compounds with known target specifities. *J Chem Inf Model* ASAP 10.1021/ci600288m

84   Rollinger JM, Langer T, Stuppner H (2006) Strategies for efficient lead structure discovery from natural products. *Curr Med Chem* 13: 1491–1507

85   Rollinger JM, Langer T, Stuppner H (2006) Integrated *in silico* tools to exploit the natural products' bioactivity. *Planta Med* 72: 671–678

86   Van de Waterbeemd H, Gifford E (2003) ADMET *in silico* modelling: towards prediction paradise? *Nature Rev Drug Disc* 2: 192–204

87   Rollinger JM, Bodensieck A, Seger C, Ellmerer EP, Bauer R, Langer T, Stuppner H (2005) Discovering COX-inhibiting constituents of *Morus* root bark: Activity-guided *versus* computer-aided methods. *Planta Med* 71: 399–405

88   Bernard P, Scior T, Didier B, Hibert M, Berthon JY (2001) Ethnopharmacology and bio-informatic combination for leads discovery: application to phospholipase A2 inhibitors. *Phytochemistry* 58: 865–874

89   Van de Waterbeemd H (2005) Which *in vitro* screens guide the prediction of oral absorption and volume of distribution? *Bas Clin Pharmacol Toxicol* 96: 162–166

90   Rollinger JM, Mock P, Zidorn C, Ellmerer EP, Langer T, Stuppner H (2005) Application of the *in combo* screening approach for the discovery of non-alkaloid acetylcholinesterase inhibitors from *Cichorium intybus*. *Curr Drug Discov Techn* 2: 185–193; Erratum (2006) 3: 89

91   Bajorath J (2002) Virtual screening in drug discovery: methods, expectations and reality. *Curr Drug Disc* March: 24–28

92   Bajorath J (2002) Integration of virtual and high-throughput screening. *Nature Rev Drug Discovery* 1 Nov: 882–894

93   Chimenti F, Cottiglia F, Bonsignore L, Casu L, Casu M, Floris C, Secci D, Bolasco A, Cimenti P, Granese A et al (2006) Quercetin as the active principle of *Hypericum hircinum*

exerts a selective inhibitory activity against MAO-A: extraction, biological analysis, and computational study. *J Nat Prod* 69: 945–949

94 Rollinger JM, Schuster D, Baier E, Ellmerer EP, Langer T, Stuppner H (2006) Taspine: Bioactivity-guided isolation and molecular ligand-target insight of a potent acetyl-cholinesterase inhibitor from *Magnolia x soulangiana*. *J Nat Prod* 69: 1341–1346

95 Rognan D (2006) *In silico* screening of the protein structure repertoire and of protein families. *Chemogenomics* 109–131

96 Paul N, Kellenberger E, Bret G, Mueller P, Rognan D (2004) Recovering the true targets of specific ligands by virtual screening of the Protein Data Bank. Proteins 54 : 671–680

97 Nettles JH, Jenkins JL, Bender A, Deng Z, Davies JW, Glick M (2006) Bridging chemical and biological space: "target fishing" using 2D and 3D molecular descriptors. *J Med Chem* 49: 6802–6810