Top Med Chem (2009) 5: 85–124
DOI: 10.1007/7355_2009_3

Published online: 6 Aug 2009

Lead Discovery Using Virtual Screening

Jack Andrew Bikker and Lakshmi S. Narasimhan

Abstract The practice of virtual screening (VS) to identify chemical leads to known or novel targets is becoming a core function of the computational chemist within industry. By employing a range of techniques, when attempting to identify compounds with activity against a biological target, a small focused subset of a larger collection of compounds can be identified and tested, often with results much better than selecting a similar number of compounds at random. We will review the key methods available, their relative success, and provide practical insights into best practices and key gaps. We will also argue that the capability of VS methods has grown to a point where fuller integration with experimental methods, including HTS, could increase the effectiveness of both.

Keywords VS, Virtual Screening, Lead discovery, lead, HTS, Pharmacophore-Based, Structure-Based, Fragment-based, Ligand-based, Docking, Scoring, hybrid workflows, VS strategy, Benchmarking VS

Contents

J.A. Bikker
Department of Chemistry and Screening Sciences, Wyeth Research, 401 North Middletown Road, Pearl River, NY, 10965, USA
e-mail: bikkerj@wyeth.com

L.S. Narasimhan(✉)
Pfizer Global Research and Development, La Jolla Laboratories, La Jolla, CA, 92121, USA
e-mail: Lakshmi.narasimhan@pfizer.com

Abbreviations

ADME	Absorption, distribution, metabolism, excretion
FBVS	Fragment-based virtual screening
LBVS	Ligand-based virtual screening
PHBVS	Pharmacophore-based virtual screening
SBVS	(Protein) Structure-based virtual screening
SSR	Selection to superset ratio
TMVS	Text-mining based virtual screening
VS	Virtual screening

1 Introduction

Over the last decade, improvements in algorithms for molecular comparison and in docking and scoring, in conjunction with the advent of affordable yet fast computing through clusters of relatively inexpensive processors have made VS a promising strategy to identify novel leads to known and new targets. It is a highly cost-efficient and relatively fast way to leverage limited information on a biological target, namely a small number of compounds that are active against it, or its structure determined to atomic resolution or both, to find additional leads. When successful, this method can often identify leads that are of interest, as defined by the key characteristics of potency, novelty, exploitability, selectivity, and ADME. As a

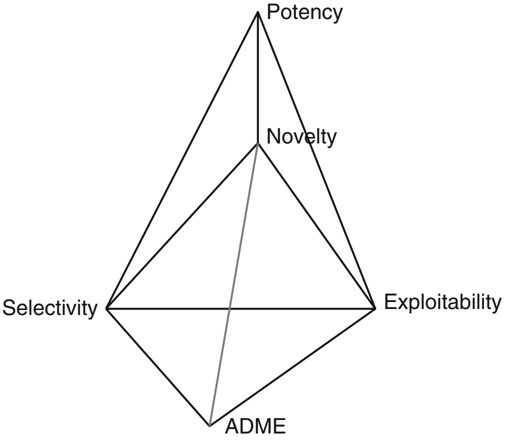

Fig. 1 The desired attributes of a lead molecule. Often, molecules identified by any screening strategy might satisfy optimal criteria for only a subset of these attributes and most laboratories would proceed with a medicinal chemistry campaign banking on improving the rest in a subsequent lead optimization phase

strategy, the software and methods available can be used alone, or in combination with more common high throughput screening (HTS) or fragment screening technologies. As this review will demonstrate, there is ample evidence to show that this technology has been applied to targets in gene families for which inhibitors are known, and to recently identified targets. Success, dependent on a variety of factors and defined in as many ways, is variable, and controlled by what we have termed the zeroth law of screening: *If the compound is not in the screening collection, it can't be found.* However, experience and the literature suggest that, if there are inhibitors of modest potency or better present in the screening collection, a subset may well be found by applied VS methods (Fig. 1).

VS refers to any computational filtering or statistical prediction applied to cherry-pick compounds from a large database. The logical next step is to acquire these compounds for experimental testing. An operational definition of VS, that it is the exercise of ranking molecules by descending order of likelihood of relevant biological activity, regardless of how that ranking is performed, captures the essence of VS ([1], quoting [2]). The choice of ranking algorithm generally depends on the information known on the target, knowledge of compounds active in the relevant biological assay, how dissimilar the desired ligands need to be from known bioactive molecules, and what percentage of the ranked database would be selected for experimental testing. The smaller the percentage of compounds to be tested, the more efficient the ranking algorithm needs to be to result in successful hits from VS.

The most common VS method is a similarity-based (almost always executed through the use of a fingerprint) or substructure-based search. These are so integrated into medicinal chemistry practice that they are often overlooked as being amongst the most common and effective VS methods. However, given one or more active compounds, chemists invariably attempt to identify similar molecules using substructure and similarity queries. Substructure and similarity searching is often the unexciting but highly effective follow-up of a more complex virtual screen that attempts to find new lead matter. Even when only a few analogs turn up as initial hits, substantial structure–activity relationships of an entire series can sometimes be gathered without requiring new synthesis. The continuing interest in fingerprint-based methods is covered in more depth in the LBVS section of this review and in several reviews in the literature [3–7].

The more challenging scenario arises when the need to identify new scaffolds or series becomes the driver of the VS experiment. Often, the ratio of the number of compounds selected for testing to the size of the database of compounds screened, SSR (selection to superset ratio), is in the range of a thousandth or less. Success (which could be 1% or more of selected compounds having relevant biological activity) while selecting in such low SSR situations (small number of molecules selected from a very large collection) has won VS the recognition as a distinct function of computational chemistry that can deliver new leads to a drug discovery effort complementing experimental methods like high-throughput screening (HTS). Mostly such VS is done to select compounds from databases typically present in medium to large pharmaceutical companies or compendia of commercially available compounds or combinatorially synthesized collections provided by vendors or combinations thereof. A characteristic of such databases is the variable extent to which different segments of chemical space are over or under represented.

VS methods to identify new chemical series can be broadly classified into three classes:

1. Methods that rank compounds based on some measure of similarity to known actives, based on 2D or 3D structure of the molecule (LBVS).
2. Methods that deduce a pharmacophore, an arrangement in 3D space of features that contribute or detract from binding and look for its presence in the database that is searched. This method places emphasis on features like hydrogen bond donors, hydrogen bond acceptors, acidic or basic units and hydrophobic fragments and opens the possibility of identifying unexpected scaffolds with required features (pharmacophore-based VS or PHBVS).
3. Methods that utilize structural data of the target, generally identified by protein crystallography, to look for molecules that complement the “binding site” through favorable protein–ligand interactions (protein structure-based VS or SBVS).

The choice of method used is often facilitated or constrained by the information available. In the absence of structural information on target, if one or more active small molecules are known, LBVS or PHBVS are feasible. If no active compounds are known, but an experimental or computational model of the protein structure is available, SBVS can be considered. If both active compounds and target structure are available, one or more appropriate methods can be applied, or multiple methods combined.

There have been a number of very helpful reviews of aspects of VS in the past few years. These have focused on either specific methods, or on the field as a whole. Cramer has provided an interesting review of methods of lead-hopping, concentrating on technologies applicable to find scaffolds very different from the initial scaffold known to be active [8]. This is especially useful if there is some doubt as to whether a molecule from the current series can be developed with sufficient chemical novelty to allow it to be patented. Jalaie and Shanmugasundaram have reviewed the state of the art prior to 2005 [9] as have Reddy et al. [10]. A review focused on LBVS has been published by Hert et al. [11]. Finally, a broad and characteristically trenchant review has been provided by Klebe [12]. In this review we will focus on advances and successes reported in the past 2 years, with the perspective of practitioners of the art in two large pharmaceutical companies.

1.1 Benchmarking Virtual Screening Methods

Numerous researchers in academia and industry have worked to advance the performance of VS methods. Many sets containing molecules active at a given target mixed in with known or presumed inactives (better referenced as decoys) have been created and have been used to demonstrate the performance of individual methods, or compare the performance of multiple methods. Table 1 provides a summary of many of these data sets, most of which are publicly available. A key consideration is the choice of inactives/decoys present in these datasets. Ideally, the physicochemical profile of the inactives/decoys should be matched to those of the actives, thereby preventing the observed enrichment from being a surrogate for property differences between active and inactive members. This is a consideration because many scoring functions are somewhat correlated with the molecular weight and lipophilicity of the ligands docked and scored.

Generally, performance of a method is often judged in one of two ways. The first is the *enrichment factor*, *enrichment* for short, which is the ratio of the cumulative number of actives in the top N% of the total number in the dataset to random retrieval rate. Many early studies focused on the enrichment obtained when the top 10% of the dataset was screened. However, this is operationally unrealistic if compound collections exceed 100,000 compounds, which is common in mid- to large-sized companies. A more realistic test is the enrichment obtained in the top

Table 1 Reference data sets, and location as of 2008

Data set(s)	Actives	Link
Cox2	128	http://www.ncbi.nlm.nih.gov
Estrogen receptor	55	
Gyrase B	55	
Neuramidase	83	
P38 kinase	55	
Thrombin	67	
DHFR	100	
Factor Xa	100	
ZINC	Variable	http://zinc.docking.org/
DUD	2,539 actives against 40 different targets	http://dud.docking.org/

0.1–1% of the compounds ranked. Many papers now include 10%, 5%, and 1%. This method unfortunately depends on the ratio of actives to decoys present in the dataset and makes comparisons across datasets difficult.

Another measure, that is independent of the ratio of actives to decoys, is the more comprehensive receiver operating curve and an enhanced version [13, 14] and reduces the dependence of the success measure on the number of decoys in the set. This also graphically demonstrates the enrichment as a continuum, plotting the fraction of the actives retrieved (true positive retrieval) against the fraction of the inactives retrieved (false positive retrieval) [15]. The first ratio is the sensitivity of the method (fraction of compounds that are predicted to be active out of the total true actives present in the sample) and the second is the specificity of the method (the fraction obtained by subtracting from 1 the ratio of the true negatives to total negatives which would be the ratio of compounds falsely predicted to be positive out of all the inactives). The area under the curve, AUROC, is a measure of the efficiency of the method. As the VS method gets better the area under the ROC curve will approach 1. The method is better than random if the area is >0.5 and this method allows comparison across datasets since the curve shape and area are independent of the size of the dataset. Figure 2 shows a ROC curve reproduced from a PubMedCentral Article, which also has a lucid account of this widely used statistical method [16, 17].

The variety of test data sets has helped to broaden our understanding of how different methods perform under a variety of circumstances. Notably, the enrichment

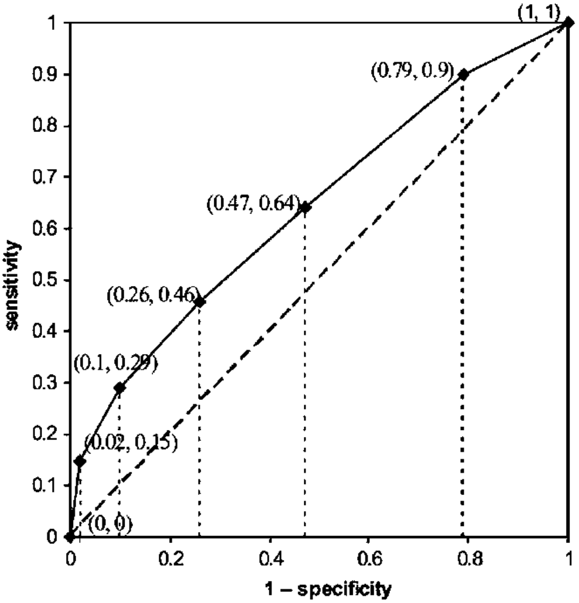

Fig. 2 A receiver operator characteristic curve reproduced from PubMedCentral http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1065080 and [16]

in structure-based methods that utilize docking and scoring can be highly dependent on the quality of the scoring function employed and how sensitive it is to small errors which creep in at the docking step [18, 19]. Therefore, by testing performance against a variety of targets, a more realistic assessment of a given technique can be produced. This information allows the experienced drug hunter to better tailor their VS experiment to the protein class. In recent years, there has been a significant advance in the ease of availability of curated datasets publicly available ZINC and DUD being notable examples [19, 20] that allow comparison of the performance of different methods using active/decoy combinations available to all practitioners of the art.

A key issue that arises in practice, is whether the experiment is biased toward enrichment (most actives for the number of compounds screened) or novelty (compounds identified are drawn from different chemical classes). Generally, any protocol leading to a small number of compounds tested is biased toward enrichment. In this case, the strategy can often involve multiple serial VS methods with aggressive property filtering. With fewer experimental constraints, the computational chemist is afforded the luxury of fewer assumptions. For example, by performing a pharmacophore-based VS, but not refining the data set by docking into a known protein complex, compounds that might otherwise be excluded due to a change in the active site might be found. Furthermore, the practice of inspection generally aids in eliminating unrealistic binding modes or undesirable chemical functionality [21]. This is done with the risk that the unexpected turns out to be true.

1.2 Database Creation

One big advantage of the VS experiment is that the compound screened need not be available physically and possibly not exist at all. Generally, the enrichment available through VS is thought to be insufficient alone to justify de novo synthesis. An exception to this rule is the practice of computationally screening a large combinatorial virtual library, identifying potential actives and following up with combinatorial synthesis of a smaller subset of the library [22–24]. As the enrichment offered by VS methods improves and with sufficient synthetic capacity, this is an assumption that could be challenged.

There are a number of commercial vendors who have made their compound collections available in formats amenable to translation into databases that can be screened with relevant software. Generally, starting from any standard format (MDL Mol or SDF, SMILES), compounds can be converted into a format required for database searching. For docking or 3D database searching, this also requires the creation of one or more 3D conformations of the molecule, which are stored for using by the screening software. Software such as CORINA [25], CONCORD [26], OMEGA [27] and Catalyst [28] have procedures to convert 2D to 3D coordinates and to generate a family of minimized conformers.

One element of database generation that is a key consideration is whether to expand the representative compounds to include alternative tautomers, protonated and deprotonated forms of the molecule, and also to enumerate stereochemistry fully if not specified in the input. Depending on the molecules in question and the options considered, these can lead to a 10-fold increase in the size of the database to be explored. However, such an expansion is necessary if methods are used that are sensitive to such chemical precision (e.g., docking). For 3D similarity searching, it is sometimes more efficient to consider various modifications to the query, leading to multiple searches against a smaller database.

A further consideration when combining databases from multiple suppliers is how to identify and deal with redundant compounds. Here, some method of mapping multiple supplier information onto a single compound is needed for efficiency. Generally, all information can be mapped, although some consideration of cost and supplier reliability may allow a hierarchy of supplier information to be applied.

1.3 Database Filtering

For many practical reasons, some element of filtering is often applied either at the point of creating a subset of “chosen” compounds for VS, or to a VS hit list before ordering compounds for testing. A number of property-based filtering criteria are available. By far the most famous are the Lipinski rule of five criteria [29–31]. By reviewing the computed properties of known oral drugs, a pattern emerged that suggested that an orally absorbed drug had a molecular weight less than 500, fewer than five hydrogen bond donating groups, fewer than 10 hydrogen bond accepting groups, and a clogP less than five. Veber et al. [32] further proposed that intestinal permeability decreased as the polar surface area (PSA) exceeded 140 Å^2. Inspired by these studies, a number of researchers demonstrated that the impetus of most lead optimization efforts tended to add size and lipophilicity to the molecule, and that the desired lead should be smaller and less hydrophobic than the eventual drug [33, 34]. This work has caused a number of researchers to limit the size of the ligands that are introduced into the “collections” of potential leads. From a practical perspective, these rules have considerable operational benefit, because they limit the size and conformational complexity of the molecules assessed. This leads to smaller databases, and less search time per molecule. Another filter which addresses size complexity from an operational perspective is the number of rotational bonds present in the molecule [35]. Prefiltering the collection to remove molecules with more than 10 rotatable bonds is a common practice except in specialized situations where there is prior knowledge that a long linear chain is a necessity for biological activity.

Another assessment of drug-likeness is afforded by Jorgensen’s Qikprop family of ADME models [36]. In addition to the individual predictions, Jorgensen has proposed a rule-of-three. This proposes that a successful drug will have predicted solubility (log S) >-5.7, predicted pCaCO $>22\,\mathrm{nm\,s^{-1}}$, and less than seven predicted primary metabolites. Qikprop will also provide an assessment of whether the compound is considered similar to molecules in the training set for these models. Given the uncertainties about applying any general model to a diverse set of molecules, this might be a consideration later in the assessment phase of the VS. It might also be used to prioritize the eventual experimental hits for experimental ADME assessment.

Additional filtering options arise from considerations of potential toxicity. Davis et al. [37] published an extensive list of chemical fragments that were proposed by medicinal chemists to be either reactive or that might be linked to toxicity. Such filters can remove unwanted or suspect functionality prior to testing to increase the likelihood of a hit being attractive as a chemical lead. Hit lists can also be filtered by any number of general QSAR models of ADME properties. While effective at further reducing the numbers of compounds in a database or list of virtual hits, the applicability of a general model on a compound from a series on which the model was not trained is suspect, at best. Such models are best applied late in the process, when some critical assessment of their validity might be attempted based on known data or by comparison of the compounds in the hitlist to the training set of the model.

2 Ligand-Based Methods

2.1 Introduction

Probably the most efficient ligand-based search method devised to date is the similarity search based on chemical fingerprints. There is a wide range of ways of defining "features" that can be mapped as part of a fingerprint: atoms, atom pairs, chemical functional group fragments and connected bond fragments [5, 38, 39]. These can then be further generalized, either by atomic properties, atom type, interactions afforded by the chemical features (e.g., hydrogen bond donor/acceptor/both), or various topological and graph theory indices [40, 41]. The choice of information encoded and the degree of generalization or abstraction can be tuned in an attempt to bias the "similarity" to match molecules with desired common attributes.

Clearly, within the conceptual framework described above, there is extensive room for exploration in creating fingerprints and similarity measures to retrieve molecules based on varying conceptions of "similarity" [42–44]. The simplest types of fingerprint consist simply of features indices that map the presence or absence of a small library of functional groups. The most well known and effective are the MACCS keys. These were initially chemical feature indices, that we later used successfully as a similarity metric.

A richer fingerprint description is provided by the Daylight [45, 46] or UNITY (Tripos Inc., St. Louis) fingerprints. These incorporate a much broader range of features, notably including connected bond path fragments up to seven bonds long.

Additional commonly-used fingerprints offer alternative ways to encode path lengths. The ECFP [47, 48] series of fingerprints used in Pipeline Pilot use a different algorithm to code path lengths of four bonds (ECFP4) or six bonds (ECFP6) or higher in length. If the atoms are genericized to a small number of roles (e.g., hydrogen bond donors and acceptors), the topologically related family of FCFP fingerprints [49] can be generated. These fingerprints have proven useful in multiple roles including similarity searching, complexity analysis, and QSAR model generation using Bayesian learning machines [50].

Another family of fingerprints available are the MOE pharmacophore fingerprints accessible through software from the Chemical Computing group [51]. In this system, the atoms are generalized into a smaller vocabulary of pharmacophore features, after which the fingerprint is constructed based on connected paths.

Feature-based fingerprints should be noted for their inclusion of pharmacophore feature types, and counts along with structural and property data into a single fingerprint for VS [52–56]. One of these arises from the Leadscope hierarchical classification of 64,000 scaffolds which has been converted into a fingerprint and used in similarity analysis [57]. Another more customizable set is one put together by Digital Chemistry software [58]. Here, a wide number of feature, path, and generalized features can be created as a huge dictionary, and then a subset of bits with the best characteristics for a given task can be chosen. Unlike many folded fingerprints, this approach has the dual advantages of being able to tailor the fingerprint to the task, and to map back the features set to the molecule.

The pragmatic beauty of the chemical fingerprint is that the more common features of two molecules that there are, the more common bits are set. The mathematic approach used to translate the fingerprint comparison data into a measure of similarity tunes the molecular comparison [5]. The Tanimoto similarity index works well when a relatively sparse fingerprint is used and when the molecules to be compared are broadly comparable in size and complexity [5]. If the nature of the molecules or the comparison desired is not adequately met by the Tanimoto index, multiple other indices are available to the researcher. For example, the Daylight software offers the user over ten similarity metrics, and the Pipeline Pilot as distributed offers at least three. Some of these metrics (e.g., Tversky, Cosine) offer better behavior if the query molecule is significantly smaller than the molecule compared to it.

When used in the VS context, the fingerprints of both query molecules and the database of molecules probed must all be computed. Generally, the fingerprints of the database compounds are often precomputed and held as additional attributes for each molecule. For each type of fingerprint and similarity metric, some similarity threshold is often applied to limit the number of hits achieved. Because of a fair amount of work early in the 1990s, the 85% similarity threshold is often applied ($T_c = 0.85$). However, this was first done in the context of Daylight fingerprints and Tanimoto indices, and should not be extended to other systems without further validation. For example, our own work suggests that similar molecules will still be retrieved using a 70% similarity threshold with UNITY fingerprints. Researchers at Leadscope [59] applied a 45% similarity threshold to comparisons using their proprietary fingerprints. Some validation is generally needed when considering a new fingerprint and similarity metric combination.

A different approach to molecular similarity is offered by various descriptor sets generated either from calculated physical properties (e.g., molecular weight, cLogP) or more complex metrics derived from graph theory. An example of the latter are BCUT descriptors developed by Dr. Robert Pearlman [60, 61]. This is currently available as part of DiverseSolutions (Tripos Inc., St. Louis). These descriptors are generally understood to encode the molecular hydrogen bond donating or accepting nature, charge, or polarizability. Operationally, this metric has the advantage of scaffold hopping in practice [62–64]. A variant of this approach is available from the CCG MOE software as QSAR descriptors [51].

2.2 Case Studies

Given the relative simplicity of ligand-based methods, it is interesting to note that in only comparatively few published reports of VS successes do the authors rely primarily on ligand-based methods. Of these studies, most appear to combine an interest in a given target with an interest in providing proof-of-concept for some extension of chemoinformatic theory.

In Table 2 we highlight pertinent information from a number of studies. We did not aim to be exhaustive, but rather to provide enough examples to provide a flavor for the type of studies performed. Of the studies in Table 2, one element to note is the small number of compounds tested in five cases. Despite starting with databases that range from 37 K to 2.5 million compounds, most researchers end up actually testing less than 100 in most cases and several hundreds at most.

An example of the value of VS based on descriptors alone is that of the identification of inhibitors of 5-lipoxygenase by Franke et al. [73]. 5-Lipoxygenase catalyzes the first transformation of arachidonic acid to leukotrienes that mediates many inflammatory responses. It has also been proposed as a contributor to atherosclerosis, cancer and osteoporosis. To seed their study, 43 known 5-lipoxygenase pathway inhibitors were used. The investigators chose the AnalytiCon Megx library of purified natural products as their database, which contained 1,298 compounds at the time of testing and the Nat-X library containing 7,839 compounds [74]. The CATS-2D topological pharmacophore-pair descriptors [75, 76] were used, and 430 hits (10/query) were assessed visually for the novelty of their scaffolds. Just 18 were tested, of which two showed activity in a cell-based assay. Both hits were from a library of natural products derived from α-santonin. For each hit, several close neighbors were selected for screening from the NAT-5 library. Additional hits were obtained for both series. Since cell-based screening was performed first, followed by receptor-based screens, some ambiguity remains as to whether the hits – especially related to series 2, are genuine 5-lipoxygenase inhibitors or act elsewhere in the pathway probed by the cellular assay.

Table 2 Examples of ligand-based VS workflows

Target	Notes	Outcome	Reference
Dopamine D2, D3	SPECS db (230 K), NN, clustering, SOM	9 D3 antagonists, 6 D2 antagonists of 190 tested	[65]
Kir6.2/SUR1 K ATP channels	ZINC db (65 K), FLAP screening to 1,913 compounds	3 hits of 32 tested	[66, 67]
L-type calcium channels (voltage-gated calcium channels L-subtype)	Similarity to Diltiazem and a second ligand. ZINC db (~50 K commercially available subset screened but most filtered to achieve desired PK profile using VolSurf). SHOP similarity, and feature-presence filtering down to 36 compounds	7 hits 18 tested. active in a vasorelaxant assay and some had novel structures.	[67]
5-Lipoxygenase	AnalytiCon Discovery db, Similarity based on 2D CATS descriptors	18 hits/430 tested	[68]
ICAM-1/LFA-1	Database of 2,500 K, custom minifingerprints based on pharmacophore pairs	1 hit/25 tested	[69]
Na/K ATPase	ICB natural product database (37 K), QSAR, Chemfinder similarity search based on ouabain	4 hits/10 tested	[70]
PDE1, PDE5	SPECS (88 K), CART regression trees based on 2-point pharmacophores	7 hits/19 tested	[71]
Mycobacterium tuberculosis	Recursive partitioning, similarity to conceptual virtual libraries	1 hit/4 tested	[72]

A second example of a VS exercise that was largely fingerprint-based was that of Boecker et al., in search of novel series for dopamine D2 and dopamine D3 blockers [65]. A set of known actives consisting of 472 dopamine D2 and D3 ligands was assembled from the literature. The SPECS database of 230,000 compounds was chosen from which to identify compounds. Two descriptor sets were calculated: MOE2D [51] and CATS3D [77] for both query and database molecules. Neighbors of the known actives were then identified using NIPALSTREE hierarchical clustering, hierarchical *k*-Nearest Neighbor analysis, and a self-organizing map analysis. These analyses yielded 37, 144, and 52 neighbors respectively. These hitlists were culled by considering druglike properties, the presence of an ionizable nitrogen (a key pharmacophoric element) and novelty. Of 17 compounds eventually purchased and tested, nine had potent (K_i <1 μM) D2 binding and six had potent D2 binding. The most interesting had dopamine D3 binding of 65 nM and was 13-fold selective over D2. As a follow-on study, a pharmacophore model was built using the MOE [51] software and the dataset of literature and recently identified molecules. This was applied to the SPECS database, and four additional compounds were ordered. The best had dopamine D3 binding of 65 nM and was mildly selective over D2. All four had binding of <10 μM at either the D2 or D3 receptors.

An interesting example of the use of novel fingerprints developed by the cheminformatics group at the University of Perugia and marketed by Molecular Discovery Inc. is afforded by a paper describing the search for novel potassium channel openers reported by Carosati et al. [66]. Compounds that open pancreatic ATP-dependent potassium channels may help regulate insulin secretion in diabetes. The ZINC database [16, 19] of 65,208 compounds (in 2005) was reduced to 1,913 compounds by applying pharmacokinetic filters. Molecular weight was restricted to between 200 and 600 amu, and clogP to between 1 and 5. In addition, three Volsurf [78] ligand-based models were applied to select compounds predicted to have good absorption, limited blood-brain barrier penetration, and adequate cell permeation. From this smaller pool of compounds, molecules were chosen that were similar to six known potassium channel openers. This was accomplished by principal components analysis of the GRIND [79] (grid-independent pharmacophore descriptors), multivariate similarity of TOPP [80] (three-point pharmacophore-based fingerprints) and pairwise superposition and scoring of FLAP [80] (four point pharmacophore-based descriptors) were calculated for query and target molecules and similar compounds identified. After inspection and selection, 3 compounds of 32 eventually purchased demonstrated $E_{\max}$ >100% when tested in channel preparations. The paper highlights that each different type of descriptor used identified different compounds, which were combined into the final set that was ordered.

A fourth example highlights the value of generating a predictive model of activity from known SAR and then applying this model to a database of compounds. Yamazaki and coworkers undertook this analysis to identify new classes of PDE1 and PDE5 inhibitors for development as potential cardiovascular therapeutics. Existing SAR for 130 compounds was initially used to train a CART recursive partitioning model, with 10,000 diverse compounds selected from 88,000 SPECS compounds used as an inactive background. One hundred and sixty eight descriptors were calculated based on binned distances between pharmacophore pairs, and an additional 12 physical property descriptors were added. The SPECS database of compounds was searched using the derived model, although filtered to ca. 50,000 compounds by comparing the latest version of the catalogue with a 1998 version and removing common (older) compounds. This was done to bias the compounds to those likely to be available. One thousand eight hundred and twenty one putative inhibitors were identified using the CART model, of which 100 were selected by diversity analysis. From these, 19 compounds were tested, of which 11 showed >50% at 10 μM and 7 were of interest as dual PDE1 and PDE5 inhibitors.

3 Pharmacophore-Based Methods

3.1 Introduction to Methods

The notion of the "pharmacophore" has a long and successful tradition within medicinal chemistry. Before the visualization of protein–ligand interaction brought on by crystal structures, chemists working within a given series would – by trial and error – identify those parts of the molecule most associated with a desired biological activity [81-83]. Provided the pharmacophore remained constant, changes elsewhere in the molecule might modulate activity but often ensured that potency was retained with exceptions arising only when additional molecular fragments caused serious disruption. This idea can be further generalized; if a pharmacophore is satisfied by other functional groups, or by comparable groups or atoms arranged in a spatially comparable way on another scaffold, then the two classes of molecules might share similar biological activity. This precept – that even when 2D topology might not suggest a common pattern of features, the presence of required pharmacophoric elements in desired spatial geometry is sufficient to provide relevant biological activity – has powered and continues to power the contributions of computational modeling and VS to drug discovery and design, and is well reviewed in the literature and a few are included [84–89]. These ideas were then extended to searching a database of 3D structures for ligands that matched a 3D pharmacophore [90, 91]. These are the methods that are generally referred to by the "pharmacophore-based VS" shorthand. Implicit in some of the discussion about pharmacophore-based fingerprints above is that another use of the term "pharmacophore" is for any scheme that refers to a collection several atoms or functional groups to pharmacophore features without the 3D geometry being included. However, in this section, we will tend to focus on methods and case studies in which a 3D pharmacophore method was applied.

The 3D pharmacophore, in its simplest form, is the presence and geometric arrangement of a combination key elements, usually selected from hydrogen bond donor/s, hydrogen bond acceptor/s, aromatic ring/s, and hydrophobic group/s. In the absence of 3D structures of receptors complexed to ligands, the pharmacophore was considered the major biologically relevant metric [88] that related molecular structure to biological activity. However, as one could easily perceive, a collection of descriptors, which capture the characteristic elements, the charges, hydrophobic character and shape, can readily describe a 3D pharmacophore in finer detail. Such descriptors were deployed in modeling and design under the general umbrella of 3D QSAR and VS experiments were accomplished with a spectrum of variations that ranged from a simple collection of pharmacophoric binding elements to multidimensional QSAR. These have been covered in many recent and almost recent reviews and we include a large selection of them for the benefit of the reader who wishes to explore applying those methods [7, 8, 92–103].

A number of very useful tools have emerged using methods that rely on shape matching or surface similarity matching. These include the ROCS method from Open Eye (www.eyesopen.com, [1]) and the Surflex-Sim [104] surface-matching method developed by Jain and currently marketed through Tripos. The shape-based method from Open Eye, called ROCS [105] has emerged as a frequently used tool in the hands of industrial chemists [1]. ROCS relies on the conversion of a single molecule in a putative bioactive conformation into a series of Gaussian grid functions representing shape or atomic character. This probe is compared to similar information coming from a precomputed database of stored conformations, and a scaled similarity function is generated from either shape overlap or similarity of atomic character. Recent publications highlight the need to employ both types of information to ensure enriched screening lists [106]. This method is distinguished by its speed, reasonably simple command-line interface, parallelization, and robust behavior across multiple ligand classes.

The Surflex-Sim method operates significantly differently [104]. Each of the molecules is surrounded by a set of "observer" points that characterizes the local character of the surface and potential interactions. Two similar molecules will have a common subset of comparable observer points. A optimal alignment occurs when the differences in pharmacophore character and molecular surface inferred from the observer points are minimized between two molecules. To speed up the algorithm, large molecules can be fragmented into parts which are then compared, and then tested for consistency. This feature also makes the method capable of identifying alignments when one molecule is much smaller than the other.

An older but effective and widely used method is the Catalyst program from Accelrys. Like ROCS, it operates as a VS tool against a database that contains a precomputed conformational expansion for all ligands. Multiple conformations of every compound are stored. It is distinguished by the ability to generate a 3D pharmacophore based on hydrogen bond donating and accepting elements, hydrophobes, and optionally positively and negatively ionizable functional groups. If trained on known ligands with three or more orders of magnitude of biological data, a robust activity prediction equation can often be generated. This function can be used as a scoring function in the subsequent VS experiment. Unlike a similarity function, this type of function can penalize for features that are already known to detract from biological activity. However, in the absence of such a scoring function, Catalyst can operate in a similarity scoring mode. Effective variations of Catalyst like functionality are also available from Computational Chemistry software from Chemical Computing Group and Schrodinger.

A third and slightly older method available is the UNITY package from Tripos Inc. This also relies on the user to identify pharmacophore features and spatial arrangement. When multiple compounds and biological activity is known, this can be used to focus on a limited number of features or to exclude specific volume from the molecule. The compounds in the database are then compared to the query pharmacophore using a flexible directed tweak algorithm. In practice, some tuning of tolerances and features are often necessary to achieve reasonable recall of actives. Validation with known actives against a small, diverse background of inactives is often recommended prior to a large-scale database search.

A complementary method that derives pharmacophores from a protein crystallographic complex is Ligand Scout from Inte:Ligand, [107] (www.inteligand.com). This method has a limited vocabulary of pharmacophore features that includes hydrogen bond donors and acceptors (and extension points), normals to aromatic rings, and hydrophobes. In practice, it has been used to convert the putative or known binding sites into pharmacophore search queries, after which the pharmacophore information is transferred to software such as Catalyst or MOE. In validation studies, it is effective at reproducing relevant binding modes.

Most of the pharmacophore methods employ a set of features that include hydrogen bond donors and acceptors, hydrophobic volume, sometimes excluded volume, and also positive and negative ionizable groups. An alternative pharmacophore description is that of the Cresset software [108, 109] [www.cressetbmd.com]. This software relies on using the extrema of the electrostatic potential, as well as a description of hydrophobic regions, to create a database query. To improve the quality of the electrostatic potential around the molecule, additional charge-bearing features are included in the force field representation to reproduce delocalized pi electrons better. The field pattern is then compared to a database of precomputed field representations based on multiple conformations for each molecule. The software offers options to generate a consensus pharmacophore from multiple ligands and to align the molecules retrieve for visual inspection.

3.2 Case Studies

A number of recent examples of the use of pharmacophores as a primary VS method have appeared in the literature. Table 3 provides a selection of these studies, with outcomes listed. The databases searched range in size from 630 molecules to 1.7 million molecules. Of the studies shown in Table 3, the Catalyst software is the method most often used, followed by UNITY and the FlexS superposition tool.

An excellent example of the ability of pharmacophore methods to search a large database rapidly is afforded by the VS done by Schuster [110] and coworkers to find antagonists of 11-β HSD. This enzyme catalyzes the conversion of 11 ketosteroids to 11-β hydroxysteroids. Inhibition of glucocorticoid overexpression may be effective in treating metabolic syndrome, and inhibition may also have a role in treating diabetes and muscle atrophy. Known selective 11-β HSD1 inhibitors were used to train a model in Catalyst that contained a donor location, an acceptor location, and four hydrophobes. The ability to retrieve inhibitors was tested both using the known inhibitors against a random set of molecules (presumed inactive) and from the WDI database of approximately 63,000 compounds. A second model was generated from inhibitors that bound to both 11-β HSD1 and 11-β HSD2 and was tested in a similar way.

Table 3 Examples of VS using primarily pharmacophore methods

Target	Notes	Outcomes	Reference
11-β HSD	Database of 1,700 K, Catalyst –>31 hits	7 hits/30 tested	[110]
AR downregulating agents (ARDA)	Maybridge (60 K) and NCI (239 K) Catalyst –>41 hits	6 hits of 17 tested	[111]
Alzheimer's tau protein	Maybridge database, 136 identified	2 hits of 19 tested	[112]
CoX-2	Maybridge database (12.5 K) Catalyst search followed by GOLD docking	5 hits of 8 tested	[113]
Chloroquine-resistance reversal agents	3D QSAR		[114]
Fetal Hb transcription inducers	TFIT pseudoreceptor, Similarity search of 630 candidate molecules	2 of 26 active	[115]
Ginkgolides as GABA modulators	Pharmacophore search of 300 K structures	No hits of 31 tested	[116]
GR-Glucocorticoid receptor	Commercial db (718 K) filtered to 862, searched by FlexS	One series	[117]
Chalcones	Chemical library probed with pharmacophore	Ligands active in vitro and in vivo	[118]
Mycobacterium tuberculosis H37Rv	Pharmacophore selection of 95 compounds from a database of 15 K, docking to further reduce candidates	4 potent hits	[119]
PPARγ	Maybridge db (62 K), Catalyst	Novel series	[120]
Pfmrk: plasmodium falciparum	3D QSAR		[121]
SIRT-2	Maybridge, Leadquest dbs, UNITY search	4 of 11 tested	[122]
T-type calcium channel	Maybridge (55 K) and ion channel inhibitor db (8 K), Catalyst search	3 hits of 25 tested	[123]

The 2 pharmacophore models were used to search a database of about 1.8 million compounds assembled from 12 commercial databases. Hypothesis 1 returned approximately 20,000 hits, which were aggressively filtered using the Catalyst scoring function, lack of hit to a hERG pharmacophore, clogP <5, fewer than five donors and ten acceptors. Fifteen compounds remained and were available for purchase after filtering. The second hypothesis returned 107 hits, of which 15 were chosen for testing. Seven of the 30 compounds eventually purchased inhibited the activity of cell lysates by at least 70% at 10 μM.

A second study points out the need to develop a strategy consistent with the computational tools being used. Ray et al. [117] performed VS based on 3D similarity to three glucocorticoid receptor blockers. High glucocorticoid levels may be linked to the psychotic symptoms of psychotic major depression. A commercial database of 718,000 compounds was aggressively clustered and filtered to 862 compounds. FlexS [124], a 3D similarity program, was then used to assess the similarity of these molecules to three known glucocorticoid receptor blockers. The filtering was needed as FlexS performs a flexible superposition and is comparatively slow. Because one of the query compounds was racemic, both enantiomers were built and used as query molecules. Conformational searches of the query molecules identified low energy conformations for each. From these searches, 123 compounds were identified, which were further narrowed to 18 by inspection and supplier considerations. Of these, one compound was reported to block the glucocorticoid receptor with a K_i of 4.5 μM. Two rounds of similarity searching identified more potent analogues, the best of which had a K_i of 16 nM in in vitro screening. This demonstrates the need to match the database to the computational capacity availability, the implicit value of inspection, and the value of follow-up similarity searching to rapidly fill out SAR.

A third study demonstrates the value of using a pharmacophore obtained from the binding site of a protein complex. Tervo and coworkers [122] used the UNITY software to create two pharmacophore hypotheses based on the docking of three known sirtuin-2 histone deacetylase inhibitors. Sirtuin-2 is believed to be essential to the mitosis of some cells and may play a role in fat storage, some cancers, and possibly Alzheimer's disease. Based on the docked poses, a pharmacophore containing two hydrophobic locations, a donor atom, and one of two possible acceptor atom sites was defined, as well as regions of excluded volume. Lipinski filtering was applied, with the donor atom limit reduced to three and the acceptor atom limit reduced to seven. Flexible searches of the Maybridge and LeadQuest libraries were performed, which resulted in 34 compounds. These were reduced to 32 compounds by applying the Volsurf [125] permeability model. Further inspection led to the purchase of 11 molecules. Of these, four showed IC50 inhibition of <200 μM in in vitro testing.

4 Receptor Structure-Based Methods (SBVS)

4.1 Introduction to Methods

The genomic era continues to transform itself into the proteomic era [126]. A number of entities ranging from pharmaceutical companies to publicly funded academic research groups have been solving the crystal structures of many genes, and tackling ever more complex crystallographic challenges [127, 128]. For many families of drug targets there is now one or more crystal structures available of the target itself or a close homolog or ortholog.

The elegance and promise of the availability of structure for ligand discovery – that once we have an apo site in atomic resolution, we can find molecules that bind tightly to it by generating a very large number of virtual complexes, followed by scoring, ranking and selecting the very best – has been a holy grail of structure-based discovery ever since Irwin (Tack) Kuntz and colleagues came up with a program called Dock roughly two decades ago [129] that could identify molecules from the Cambridge Crystallographic Database that could fill a given protein site. Much has happened in the last two decades and a recent review, interestingly with the same researcher being the first author, gives a picture of the state of the art [130]. In the intervening 20 or so years, at least 50 docking programs and their variants have been developed. Docking has come of age and docking software available can in most cases reliably and quickly reproduce observed crystallographic binding modes of protein–ligand complexes with RMS variations approaching the experimental error in the crystallographic experiment that characterized them. With robust docking tools and fast, cheap and plentiful computing power, it is a surprise that SBVS has not replaced experimental screening. In practice, however, several published and unpublished success stories notwithstanding, this still stays a challenge, to the point that successful SBVS is not as routine as one could have expected it to be per our outlook from a decade ago [131]. This is despite vigorous development of docking methods and scoring functions by the computational chemistry community for well over a decade chronicled in the representative set of citations here [21, 36, 132–142].

Part of this disconnect between expectations and performance in SBVS originates from the way protein–ligand interactions are quantitated to arrive at selecting the best pose of the small molecule in the receptor site rapidly, or the way the "docking problem is solved." These make approximations in correctly describing the entropy change upon binding, and free energy components such as free energy of solvation, in order to sample and evaluate rapidly a substantial number of conformations including multiple poses for each conformation of the small molecule in the receptor, assuming the receptor is held rigid, which is common in SBVS applications. When applied to the problem of choosing the correct docking pose for the same molecule, the changes in solvation and entropy tend to become negligible from pose to pose, leading to substantial success in selecting the best pose from amongst a set likely of the docked poses. This is only a generalization and can be influenced by the nature of the binding site, in terms of whether it is a site that deviates to the extremes of hydrophobic or hydrophilic character, whether there is potential for less or more protein movement, and whether the ligand in question has more or less rotatable bonds [135] resulting in one or more docking programs being better than another for a particular combination of receptor and ligand. In addition, most docking programs generate a series of poses that are relatively closely spaced in their "docking score", the pseudoenergy function used by the docking software

Table 4 Virtual Screening examples using primarily SBVS

Target	Database	Protocol	Outcome	Reference
E. coli primase	Commercial (500 K)	Glide, filtering	4 hits of 68 tested	[143]
RNA methyltransferase	SPECS, Maybridge (300 K)	FlexX, filtering and inspection	2 series, 8 hits of 33 tested	[23]
LXR	Proprietary 135 K	Glide	1,295 tested, one hit series highlighted	[144]
S-Adenosylmethionine carboxylase	NCI diversity (2 K)	Glide	18/133 tested	[145]
Sex-hormone binding globulin	90 K, 52 K remaining after filters applied	Glide, top 16 tested	4 hits of 16 tested	[146]
CYP 2D6	6 K	GOLD,	11 hits of 16 tested	[147]
Falcipain 2 and 3	355 K	GOLD	22 of 100 tested	[148]
SARS CoV	361 K	EuDOC	1 of 12 tested	[149]
Chk1 kinase	700 K	RDOCK	9 hits	[150]
Heme oxygenase	Commercial databases (800 K)	DOCK	8 hits of 27 tested	[151]
Ubiquitin C-terminal hydrolase	Chembridge (33 K)	DOCK/GOLD	3 hits	[152]
12-Lipoxygenase and 15-Lipoxygenase	Chembridge diversity (50 K)	GLIDE SP, XP	3 hits of 20 tested	[153]
ER alpha	SPECS 202 K	LIGIN, Chemgauss	3 hits of 7 tested	[154]
NkkB	Proprietary 5,000 K	4SCAN/ProPose	1 hit shown of 236 tested	[155]
Cyclophilin A	SPECS 280 K, 85 K after filtering	DOCK, CSCORE, FlexX	15 hits of 82 tested	[156]
Potassium K+ channel	DOCK 4.0/homology model	ACD (200 K)	6 hits of 20 tested	[157]
HCV 3D polymerase NS5B	GLIDE	2,000 K commercial	1 series from 50 tested	[158]
SARS Coronavirus main protease	Maybridge 59 K	GOLD	2 hits of 50 tested	[159]
A1-antitrypsin aggregation	Combined commercial 1,200 K	ICM	10 hits of 68 tested	[160]

CDC25 phosphatase	Chembridge 413 K (313 K after filtering)	FRED/Surflex	99 hits of 1,500 tested	[161]
trans-Sialidase	Asinex GOLD and Platinum (300 K)	DOCK	5 hits of 32 tested	[162]
CDK2	975 K	RDOCK	38 of 1,121 tested	[163]
Protein phosphatase 2C	NCI diversity set (2 K)	Autodock	4 hits of ca. 100 tested	[164]
Protein arginine methyltransferase	Proprietary 9 K	GOLD		[165]
Protein arginine methyltransferase	NCI diversity set (2 K)	GOLD	7 hits of 30 tested	[166]
Arylalkylamine *N*-acetyltransferase	Commercial 1,200 K	GOLD	5 confirmed hits of 188 tested	[167]
SARS Cov 3C-like protease	Maybridge 59 K	Dock 4.0	23 hits of 93 tested	[168]
AHAS	NCI (164 K)	DOCK, Autodock	3 hits of 14 tested	[169]
Integrin avb3	SPECS 89 K	DOCK	14% hit rate of 50 tested	[170]

to differentiate between poses of the same molecule in a given receptor site. The top ranked pose, the pose with the best "docking score" might often actually be less similar to a crystallographically observed binding pose compared to a lower ranked pose and could be seen to be one with less binding energy when evaluated with a more accurate scoring function. This seemingly small error in choice of the best pose of a single molecule gets magnified and becomes substantial when the best docked poses of different molecules are compared, partly due to the breakdown in the comparability of the approximations in solvation and entropic terms across different ligands which can lead to incorrect rankings [171–174]. One could venture to say that a consensus SBVS view today would be that the inability to rank a database of ligands in order of their potential for binding to a given receptor site has more to do with our inability to score the binding affinity of a series of ligands in their predicted pose reliably than our ability to predict a reasonably accurate binding pose [175]. To that extent, one of the strategies proposed for effective SBVS is to generate a set of poses for a large collection of molecules rapidly using a well approximated but fast "docking function" and then rank with a more thorough but slower energy evaluation to rank molecules [176]. Extensive and continuing effort has focused on generating better scoring functions that better capture free energy differences between molecules, but can still operate fast enough to be of use to a high throughput docking experiment [175–178]. Results of head to head comparisons of docking and scoring using multiple docking scoring software frequently suggested that different scoring functions could be more effective for different receptors and this led to the drive towards consensus scoring functions [179–182].

Given that the focus of this chapter is SBVS and not a treatise on docking protocols, we give here a very brief and less than comprehensive coverage of docking algorithms and some of the commonly used docking software. For SBVS applications, the two most relevant pieces of information on the docking software would be the speed of the docking software and quality of pose(s) obtained. A number of packages are available, many of which have been applied to structure-based VS experiments and with success. Among the earliest attempted were incremental construction approaches, wherein the program attempts to exhaustively position the largest fragment in as many locations as possible with the active site, followed by adding subsequent fragments with suitable torsions. DOCK, FlexX, Hammerhead, and eHits are amongst the software that use this approach. Monte Carlo approaches to sampling the pose and conformation of the ligand are used by QXP, ICM, and PRODOCK and these tend to be slower. Evolutionary algorithms that improvise on preferred poses are used by docking software like GOLD, EP Dock and FITTER. GLIDE software uses a rough sampling initially and follows with a refined search using a more sophisticated scoring scheme. Surflex-Dock also performs a rough docking simulation to obtain seed poses which are refined further with a more rigorous scoring scheme. By using sequential docking simulations of varying rigor, the sampling approximately mimics the outcome of a more intensive search. Most of the docking software mentioned could be used to dock anywhere from a few hundred to 10,000 molecules in reasonable time and depending upon availability of processing power and parallelizability of the application, could screen up to 100,000 molecules within days. If the task is to dock millions of molecules, then it becomes faster to precompute a set of conformations for each molecule and limit docking to positioning the rigid ligand in the active site of the receptor. The FRED software from OpenEye used in conjunction with Omega, the conformer generator also available from OpenEye, takes this approach. Flexibase/FLOG also share the precomputed database approach.

With so much docking software to choose from, the SBVS practitioner is left with limited guidance in choice of docking and scoring options not to mention the critical postprocessing that has to bring the followed up hitlist to less than a hundred if the ligands or to be acquired through purchase for testing, and possibly a few thousand in a pharma setting. To that end, several studies have been published over the years that compare a subset, usually the most commonly available, docking and scoring applications, in a head-to-head comparison using datasets containing known hits and decoys for receptors where structural information is available and the enrichment could be studied carefully [106, 135, 173, 174, 176, 183–185]. If nothing else, these highlight the significant variation in performance of any package based on subtle variations in decisions about database construction, choice of data sets (both of active molecules and inactive decoys), program settings, and protein systems. In practice, most users rely on docking and scoring packages readily available to them, rather than try to find/use THAT ONE package that always works. The enrichment or the efficiency of the VS effort becomes more and more stringent as the proportion of compounds screened approaches 1% or less of the database of compounds screened. In these instances, for increasing the chances of success, one needs more than a protein structure, computing power, and software. Additional knowledge of the binding preferences of the active site in question can easily outstrip incremental advantages provided by one software over another and visual inspection and/or consensus approaches can aid in weeding out false positives effectively [21, 134].

The convenient shorthand in the community is that SBVS approaches are limited by “inadequate scoring functions”. This is partly true, because in an SBVS experiment, due to the need for speed, the scoring functions do not perform a rigorous free energy calculation, and they will remain limited. Sampling, especially in situations where the ligand in question is increasingly complex with multiple rotatable bonds, can also be an issue. A significant other limitation is that generally only one protein structure is used, and held rigid. Movements of side chains or entire domains will not be modeled correctly and the extent an active site moves in response to a given ligand can be largely dependent on receptor mobility and ligand-dependent in addition to that. The resulting scores, whether fortuitously good or bad, will not be of the correct docked mode. In the latter case, even if appearing highly ranked in a hit list, this may be filtered out when inspected. Fortunately, there have been several attempts at addressing limited side chain mobility at least for situations of medium throughput and SBVS is becoming practical with inclusion of protein mobility [176, 186–190].

4.2 Case Studies

The study by Agrawal and coworkers [191] demonstrates key features in performing a structure-based VS. They searched for inhibitors of DNA primase from *E. coli*, an essential enzyme for bacterial reproduction with distant human homologs. Although a crystal structure of the DNA primase catalytic domain is available (1DDE) [192], this alone provides little insight into the most productive binding site. The GRID [193] software was used to identify three putative binding sites. The database to be searched was constructed from the catalogues of 20 vendors, and filtered to remove reactive functional groups, compounds with more than eight rotatable bonds, cLogP less than 5, and MW between 275 and 500. This resulted in a database of approximately 500,000 molecules. Representative 3D conformations were generated, protonated, and minimized. For each of the three sites identified, grids were generated with a 16-Å bounding box and a 20-Å enclosing box. Glide docking was performed, and the top 2,500 compounds as defined by the Glide SP score were inspected individually for feature complementarity, and correct ionization. A short list of 79 inhibitors was created, of which 68 were available for purchase. Of these, four inhibitors inhibited primase with an IC50 less than 100 μM.

A study by Alvesalo [194] and coworkers provides an interesting contrast in terms of methods and highlights some subtle considerations. In this case, they attempted to develop antimicrobial agents against *Chlamydia pneumoniae*. They chose to target dimethyladenosine transferase, but, because no crystal structure was available, chose to screen the structure of *Bacillus subtilis* RNA methyltransferase (1QAO) [195] as a surrogate. A database was constructed of molecules available from Specs and Maybridge, and contained 300,000 compounds after filtering for undesirable chemical groups. The database was docked into the protein binding site using FlexX [196], after which the top 2,000 molecules were inspected. From this set, 33 molecules were purchased. Of these, eight demonstrated >50% inhibition at 50 μM in a cell assay and represented two series of interest. This demonstrates that the use of a surrogate protein is viable if no exact crystal structure to the target of interest is available.

A VS study by Furci [197] and coworkers highlights the use of a third docking program, DOCK, against heme oxygenase from *Neisseria meningitides*, a Gram-negative pathogen. Heme oxygenase is an essential enzyme for heme utilization by the bacteria and blocking its function should arrest bacterial growth. The protein complex including heme (1P3T) [198] was subjected to molecular dynamics simulations with the heme removed to identify four suitable *apo* structures into which to dock the ligands. A database of 800,000 molecules was assembled from the supplier catalogues of Chembridge, Chemdiv, Maybridge, and SPECS. Compounds were docked into a single protein conformation to identify 50,000 molecules using the DOCK software. A second round of docking into all four representative protein conformations obtained from the molecular dynamics simulation allowed narrowing to the top 1,000 compounds (based on best docking score to any of the 4 protein conformations). This list was further narrowed by clustering and inspection, with 153 compounds being purchased for testing. Of the 153 compounds obtained, only 37 were soluble in DMSO or buffer, and of these, 10 interfered with the fluorescence polarization-based assay. Of the 27 tested, 8 exhibited inhibition of heme oxygenase with K_d values ranging from 12 to 240 μM. This study demonstrates the value of a sequential VS strategy, and also the way in which experimental considerations (i.e., compound solubility) can limit the overall impact of a VS study.

An example of a sequential docking strategy using different software is provided by a VS for CDC25 phosphatase inhibitors by Montes and coworkers [199]. CDC25 phosphatases play an important role in initiating cell cycle events; blockade may lead to useful anticancer effects. The structure of CDC25B (pdb code 1CWT) [200] was prepared for VS by adjusting the protonation states of various residues in the putative binding region. The 2005 release of the Chembridge database was filtered to remove compounds with undesirable reactive groups, leaving approximately 313,000 compounds. Up to 50 conformations per molecule were generated and the FRED software was used to dock the database into CDC25. The top 50,000 compounds were then redocked with full ligand flexibility using Surflex. The docked poses were then scored using either a receptor-specific Surflex function or with a receptor-specific function generated by LigScout. The top 450 molecules from each list, and the molecules that appeared in the top 3,000 molecules of both lists (total 1,500) were tested for enzyme inhibition. Of these, 99 showed at least 20% IC50 at 100 μM, with the most potent having an IC50 of 13 μM and showing inhibition in a HeLa cell assay. Overall, a number of interesting series were obtained, and the authors note the importance of consensus scoring in choosing their most active molecule.

5 Hybrid Workflows

As seen from the case studies described in the previous sections, many investigators use multiple complementary methods to reduce and refine their hit lists to manageable numbers. Often, an inspection step is included, which places a de facto upper bound to the size of the hitlists that are reviewed.

In this section, a number of case studies (Table 5) in which different types of VS methods are combined into a hybrid workflow. Often these combine a fast, ligand or pharmacophore-based method with a later docking method. The latter is useful at the inspection stage as it allows the molecule to be reviewed within the context of the protein binding site. A poor binding pose can be an indicator of a poor fit. Furthermore, possible interactions outside the scope of the molecules used to train the ligand-based method can be identified.

Table 5 Workflows that provide examples of the combined use of methods

Target	Database	Protocol	Outcome	Reference
Chorismate mutase	Proprietary 15 K	(a) UNITY, (b) FlexX	4 of 15 tested	[119]
ALK: anaplastic lymphoma kinase	Chembridge		'Numerous' hits of 2,677	[201]
HIV reverse transcriptase NNRTI	Derwent WDI, CAP complete, (67 K + 1,670 K)	(a) LigandScout, (b) Catalyst, c. Glide	5 of 6 tested	[202]
Cyclophilin A	ACD 2004 (296 K)	(a). ISISBASE, (b) FlexX, (c) Surflex	9 hits of 31 tested	[203]
DDAH	308 K	(a) rNN, (b) FlexX, (c) Sim searches	2 series of 109 purchased	[204]
Protein Kinase A, *Yersinia*	Proprietary (2,000 K)	(a) SVM kinase-like model, (b) FlexE	7 hits of 45 tested	[205]
CCR5	1,600 K	(a) pharmacophore (to 44 k), (b) GOLD, Surflex (homology model),	10 hits of 59 tested, 6 with functional response	[206]
HIV integrase	Chemnavigator (13,500 K)	(a) Catalyst (to 235 K), (b) Glide	9 hits of 88	[207]
MDM2/p53	NCI-3d (250 K filtered to 110 K)	(a)Web-pharmacophore (to 2.6 K), (b) GOLD	10 hits of 67 tested	[208]
ACE-2	3,800 K	(a) LigandScout/Catalyst, (b) ehits	6 hits of 17 tested	[209]
MGL/FAAH	Maybridge, Leadquest	(a) UNITY, (b) GOLD	No hits of 62 tested, 5 hits when retested in FAAH	[210]
HIV intergrase	Asinex GOLD (200 K)	(a) EIIP pharmacophore, (b) Catalyst, (c) Autodock	1 hit of 12 tested	[211]

5.1 Case Studies

An informative example of a hybrid workflow applied to HIV reverse transcriptase is provided by Barreca and coworkers [212]. Nonnucleoside reverse transcriptase inhibitors (NNRTI) bind to HIV reverse transcriptase and block viral replication. In this study, the Ligand Scout software was used to create a Catalyst pharmacophore from the protein complex of reverse transcriptase and Janssen R185545 [213] (1SUQ). This pharmacophore was used to search the World Drug Index (WDI, 67,000 molecules) and the Chemicals Available for Purchase (CAP, 1.7 million compounds). The molecules retrieved by the Catalyst pharmacophore with a fitness score greater than 3.0 included 521 from the WDI and 11,273 from the CAP. After filtering using Lipinski conditions, 9,345 remained. These were docked using Glide with SP scoring, and the best 1,000 hits were inspected individually. Interesting, novel compounds were evaluated for availability using the substructure capabilities in the Scifinder software, and six compounds were ordered and tested. Of these, five showed significant activity, with potency ranging from 0.2 to 4 μM.

A second study by Hartzoulakis et al. [204] also provides an example where the use of multiple methods facilitates an efficient search strategy. The target in this case was dimethylarginine dimethylaminohydrolase, an enzyme that modulates the nitric acid pathway in endothelial function, and may also control a cardiac risk factor. A bacterial ortholog from *Pseudomonas aeruginosa* may also contribute to pathogenicity in cystic fibrosis. A database of 308,000 commercial compounds was filtered to keep compounds with cLogP between −2 and 5, molecular weight less than 650, five or fewer hydrogen bond donors, ten or fewer acceptors, and ten or fewer rotatable bonds. This removed about 43,000 compounds. A reciprocal Near Neighbor clustering was used to select 35,000 compounds. These were docked into the active site of the DDAH enzyme from *Pseudomonas aeruginosa* [214] (1H70) using the FlexX software. The top 1,000 compounds were rescored using a combination of scoring methods, and the top 200 were inspected. Of the 109 selected, 90 were available and tested, of which three were interesting molecules, the most potent of which had an IC50 of 17 μM. This is an example in which clustering was used to reduce the number of compounds that were docked to a number consistent with the capacity of the FlexX program and their computing resources.

As an example of the utility of combining pharmacophore models and docking to select ligands from very large databases, the VS of Liao and coworkers [207] of HIV integrase offers an excellent example. In this case, the ChemNavigator database containing approximately 13.5 million compounds was searched. Thirty Catalyst pharmacophores were generated from known HIV integrase inhibitors, and all were used to search the database, resulting in about 235,000 hits. After filtering using Lipinski conditions and deduplication, the resulting 167,000 compounds were docked into a model of HIV integrase. The docked poses of the 1,500 top scoring compounds were inspected visually. After additional ADME models were applied and availability assessed, 88 compounds were obtained for testing.

Of these, eight compounds were assessed as active, with IC50 in their primary assay ranging from 37 to 780 μM.

6 Fragment-Based Virtual Screening

Many a pharmaceutical scientist would have at one time or another looked at a competitor's patent compound and looked for ways to find a lead that retains the activity of the competitor's compound but looks different enough not to infringe on the competitor's patent. A common strategy in such situations is to replace fragments in the molecule with isosteric fragments. These fragments could be small amounting to a few atoms or pieces that are over 100 Da or more in molecular weight. The FBVS discussion here is of fragments/substructures and does not pertain to fragments that are composed of five atoms or less. With increasing need in pharmaceutical research to have leads derived from more than one chemical class for a given target, to serve as a backup in case of unexpected failure of the lead candidate in the clinic which is attributable to compound class, researchers are sometimes looking to imitate their own compounds with a sufficiently different scaffold. FBVS is very similar to this strategy with a small twist.

FBVS presumes that all fragments of a tight binding ligand do not bind with the same ligand efficiency. While this is nothing new, in that computing properties of molecules using properties of their components is a very common occurrence in computational chemistry, fragment-based design successes in the recent literature [215–217] have given strong support to the notion that tight binding ligands can be obtained by starting from very ligand efficient albeit weak binding fragments and growing to larger ligands with high affinity when the added fragments are chosen with care so as not to compromise ligand efficiency significantly. When two fragments with affinity for a receptor are linked without restraining the ability of the fragments to bind to their respective preferred site on the receptor, the combined affinity is the sum of their binding affinities [218, 219].

For this to work, one has to have one or more seed ligands with at least moderate to high potency against the receptor. The more potent the seed ligand, the better. The molecule is then logically broken to fragments, typically at retrosynthetic bonds or if synthetic issues are not a key criteria, at rotatable bonds. Automated methods that take advantage of such fragmentation followed by piecewise similarity based retrieval followed by assembly have been reported [42, 220]. In cases where structural information is available for how the ligand binds to the target receptor, one could run energy computations to find the receptor affinity of the various fragments and weight the substructures and get better retrievals [221]. The rest of the VS is very straightforward. Two-dimensional similarity, pharmacophoric similarity or shape and electrostatic similarity could be used to find new fragments. The new fragments are linked together in an n × n matrix and tested for relevance by passing through a 2D similarity filter (to the seed molecule) or pharmacophore or protein–ligand interaction energy scoring filter (where structure is available, using docking and scoring) and other relevant filters. The resulting hits that look attractive enough could be synthesized or ordered for testing based on availability and synthesizability considerations and could also be used as idea generators.

The applicability of such VS in combination with tools available include situations where portions of any molecule need replacement with bioisosteric fragments. In this regard, BROOD software [105] and MOE [222] provide automated tools for fragment removal, replacement, and minimization to relieve any strain in the molecular assembly step and provide a database of fragments(isosteres) that could be enhanced in custom fashion by an enterprise as well. These software allow facile FBVS in 3D. Since this software has become available within the last 2 years, there seem to be a dearth of use cases in the published literature. However, anecdotal reports indicate that these are being used regularly in industry and the Websites of these two vendors provide adequate information for the inquisitive reader.

6.1 Case Study

Rummey et al. [223] searched replacements for the pyrrolidine present in their DPP-IV inhibitor searching a 10,000-molecule subset of small primary aliphatic amines extracted from the available chemical directory and visually inspected the top 500 of them. Four were selected for testing and two of them were novel hits.

Considering the power of these methods to retrieve novel molecules, it is only a matter of time before more successful reports are available.

7 Text-Mining as a Novel Virtual Screening Tool

“Can I use Google to find other molecules that have similar properties as my molecule?” could be an innocent question posed by someone new to computational chemistry. The irony of it is that all information about molecules are present in publications that are predominantly text, yet, the most powerful text-mining tool cannot retrieve it for us, at least at the present time, unless the molecular query is a simple name like glucose or pyrrolidine. To the present-day scientist, this might look something of an impossibility only if the person does not stop to think for a moment that the question would have hardly been comprehended by the average person only a decade ago. Text mining and natural language processing (NLP) a decade ago was not what it is today [224].

To a computer scientist, VS is nothing but another text mining, only the bits and bytes stored that contain molecular information adopt a format quite different from natural language and without adequate warning cannot be quickly interpreted. It is not that modern day text does not contain text that is not natural language, but that they are adequately flagged and do not stop the NLP software. For example, hyperlinks do not read like natural language but they are adequately flagged and are properly processed. In the case of chemical structures, the material to be searched, the algorithms, used and retrieval techniques are geared towards structure perception and manipulation although the information is still stored and operated on as bits and bytes. This limitation exists because molecular information is not expressed in natural language in an easily perceptible form, and where we do express them, in patents for example, it is so convoluted that very few people attempt to read and decipher the chemical structure or composition by reading the IUPAC name detailed in a patent. Everyone reaches for a translator, nowadays inevitably the appropriate software, that could translate the name into the familiar chemical structure form. Unfortunately the one line smiles representation of a molecule did not come into vogue soon enough and computing facilities did not exist to encourage the broad range of scientists to represent every structure to be associated with its smiles in written documents with appropriate flags to enable software to interpret it correctly.

7.1 Current Limitations

One of the greatest limitations of searching for molecules is the fact that the database is finite. Several forms of text similarity are a part of the strategies used by people not trained in science and those easy similarity search strategies are not available to the scientist searching through molecules. Unless the database is prepared in a specific format and made available, searching cannot proceed. Search results are curation dependent and associations are limited by curation capabilities and subject to errors and biases introduced at the point of curation [224]. To give a simple example, if the curator errs and associates a wrong number with a molecule structure in the main database, regardless of how many other documents carry the correct information, people will repeatedly extract the wrong information because the association cannot be deciphered using NLP from other corporate documents. The rate of publishing is exploding, and curation is limiting. Imagine entering the smiles string for a molecular fragment in Google and get 300 references all discussing various pieces of information about it! Imagine replacing one of the carbon with an asterisk and seeing many analogs and information about them as well.

7.2 The Rewards of Storing Molecular Structures in NLP Searchable Form

Screening brings back a rich variety of information, not just what the curator put in the database. Suddenly a chemist can read everything about a molecule ever printed, not just what someone decided to associate it with. Distant associations – A related to B and B related to C might mean A related to C– will become apparent.

Chemical structural information is one of the missing pieces in the great effort to bring biomedical research into the realm of twenty-first century information extraction and knowledge discovery paradigms. Proteins, genes, diseases, and chemical compounds constitute the major entities extracted in the biomedical domain. The ability to read structure information and substructure information and their association to other entities could have a major impact on toxicity information in particular and ADMET data in general.

7.3 Potential Long Term Solutions

How do we do it? Every 2D structure reference created in the future should have a hyperlink to a canonical smiles string. Smiles readers should be freeware so when mousing over the molecule reference, the structure pops up. Start representing structures today and 15 years from now, our 2D VS efforts will look very different. The main added advantage will be that the data associated with every structure will be available for natural language processing software from which to process and extract information. Structures themselves can be searched in unforeseen ways. This will bring information about molecules in an unprecedented fashion to the average reader.

7.4 Potential Short Term Solutions

There are few short term solutions that we can think of. The technology for accessing publications underwent a dramatic makeover in the last decade, moving from predominantly paper to predominantly electronic through a coordinated set of efforts from publishers and consumers (in this case scientific research users) alike. A study of how such a transition was successfully handled could provide clues on how to make it happen.

8 Summary

VS continues to be a growing area, fueled by the dramatic increase in affordably priced computing capability, and the development of better algorithms and software. Its position as a cost-effective alternative to high-throughput screening, the traditional engine for lead identification for pharmaceutical discovery, is bound to rise, despite the technology advancements in screening through ultrahigh throughput methods, miniaturization, and automation. This is partly due to the high cost of personnel and reagents, both of which are needed in larger supply for HTS compared to VS. However, as the field stands today, one would be very justified in stating that VS is nowhere near replacing experimental screening methods and this is mostly due to the inconsistency of success in finding leads using VS. Many factors, the target itself and the information available to prime the VS effort being the major ones, the technique, the software and the expertise of the screener being the minor ones, influence the success rate. This continues to be a fast growing field, and the recent trends and progress in identifying the major challenges and addressing them effectively both at the scientific and algorithmic levels bodes well for the future of this method.

8.1 Virtual Screening Strategy

There has been considerable debate within the community and in the literature about the relative merits of ligand-based vs protein structure-based screening. In principle, the protein-based screen should provide the broadest access to novel chemotypes that could interact with the relevant binding site. The 2D ligand-based methods are often best at retrieving hits chemically similar (same or highly related scaffold, comparable pendant groups) to the query molecules. There have been efforts to develop measures of chemical similarity based on 2D graphs alone that better generalize the hits retrieved to compounds that include dissimilar but acceptable alternative scaffolds. However, these approaches tend to retrieve a large number of false positives; setting a similarity threshold to include these more dissimilar-but-acceptable hits often leads to the inclusion of far more dissimilar-but-unacceptable hits, leading to less enrichment. A third option has been the emergence of 3D similarity methods. These appear to provide a compromise leading to a balanced retrieval of both analogues and compounds containing alternative chemical scaffolds [1, 106].

Optimal strategy rests in balancing a mix of techniques and shaping the workflow for a given VS based on the information available, the perceived strengths and limitations of various techniques, and the time and effort needed. Clearly, in the absence of a protein crystal structure or acceptable homology model, ligand-based screening is the obvious option. At the other extreme, in the absence of known ligands, a protein-based screen could be contemplated. When one or more protein crystal structures are available, as well as a number of ligands that have been identified either from the literature or by some previous experimental effort, a priori, the all-out approach would be to bring to bear all available techniques to the problem. However, ligand-based screening often requires less preparation and less analysis of results, thus being sparing of the computational chemist's time and first one to get results out. Protein-based screening generally requires more time to prepare and validate the simulation, and to analyze the results, often including visual inspection to ensure that docked modes are acceptable. The choice of strategy then requires a balance between the enrichment that is expected, the anticipated novelty of the hits, and the time and effort available to invest in the effort.

As general guidance, we would suggest the following guidelines:

1.

Include VS in a lead discovery strategy whenever possible. Computational VS is low cost. It is typically performed by a single scientist who employs multiple processors, typically LINUX clusters now available at commodity prices. Of the many resources needed in the drug discovery process, processor time belongs in the inexpensive category. VS also brings considerable benefit. Many of the methods available offer some enrichment over purely random screening, and often offer significant enrichment.

2. *Test a substantial number of compounds*. VS methods generally offer enrichment, but most ranked hit lists contain a significant proportion of false positives. Hitlists should be scaled to 1–5% of the compounds in the virtual library screened. In many real world situations, the computational chemist is being asked to choose lists of compounds representing 0.1% or less of the compounds screened (e.g., the "best 100" of 100,000 compounds). Typically, VS methods have been validated considering 1%, 5%, or 10% of the total number of compounds in the VS collection. By following up on more compounds, one increases the probability of impact from VS.
3. *Include a 3D ligand-based method*. In our internal efforts across two companies, we arrived at the same conclusion as the Merck researchers [106] that a 3D similarity method appears to offer a good balance between effort expended and the number and novelty of hits generated.
4. *Automate*. Much of the human effort in VS arises at the point of combining various hit lists, followed by scoring and selection. The more this can be automated, the more efficient the VS experiment becomes.
5. *Integrate*. An effective strategy is to view VS as an approach to identifying chemical matter that is complementary to wet methods. This opens up potential symbiosis between the VS benefiting from the HTS, or alternatively, HTS benefiting from early hits identified by VS. Such a complementary view cannot be overemphasized given that the role of VS in drug discovery is often looked upon as competitive with high throughput screening or focused subset screening. However, the lower cost and faster completion times should make VS acceptable even with lower enrichment numbers. The savings in cost and time to obtain a hitlist of active compounds can be significant if additional factors like the cost and time of adaptation of an assay for HTS purposes, compound depletion in the collection due to HTS, level of false positives from HTS created by mechanical and measurement errors are considered.
6. *Whenever possible, inspect the hitlist*. Within the literature, there is a surprising number of instances in which small numbers of compounds were ultimately ordered. This inevitably requires individual inspection of compounds. In this situation, applying all relevant simulations and any hypotheses based on prior knowledge about key features are key contributors to higher enrichment. Where it is possible to order a larger VS hitlist for testing, some additional tolerance in favor of serendipity is beneficial. (For example, lowering the VDW radius of ligands or proteins to allow for possible protein motion or just ignoring small steric clashes.)

References

1. Hawkins PCD, Skillman AG, Nicholls A (2007) J Med Chem 50:74
2. Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, Schuffenhauer A (2005) J Med Chem 48:7049
3. Xie X-Q, Chen J-Z (2008) J Chem Inf Model 48:465
4. Perekhodtsev GD (2007) QSAR Comb Sci 26:346
5. Willett P (2006) Drug Disc Today 11:1046
6. Eckert H, Bajorath J (2007) Drug Disc Today 12:225
7. Cramer RD, Jilek RJ, Guessregen S, Clark SJ, Wendt B, Clark RD (2004) J Med Chem 47:6777
8. Cramer RD (2006) Expert Opin Drug Disc 1:311
9. Jalaie M, Shanmugasundaram V (2006) Mini-Rev Med Chem 6:1159
10. Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN (2007) Curr Protein Pept Sci 8:329
11. Hert J, Willett P, Wilton DJ (2006) Chemogenomics 133
12. Klebe G (2006) Drug Disc Today 11:580
13. Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O (2005) J Med Chem 48:2534
14. Bayly CI, Truchon J-F (2007) Abstracts of Papers, 234th ACS National Meeting, Boston, MA, United States, August 19–23
15. Wikipedia www.wikipedia.org
16. Bewick V, Cheek L, Ball J (2004) Critical Care 8:508
17. Bewick V, Cheek L, Ball J (2004) http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1065080
18. Seifert MHJ (2008) J Chem Inf Model 48:602
19. Irwin JJ, Shoichet BK (2005) J Chem Inf Comput Sci 45:177
20. Huang N, Shoichet BK, Irwin JJ (2006) J Med Chem 49:6789
21. Perola E (2006) Proteins Struct Funct Bioinf 64:422
22. Al-Gharabli SI, Shah STA, Weik S, Schmidt MF, Mesters JR, Kuhn D, Klebe G, Hilgenfeld R, Rademann J (2006) ChemBioChem 7:1048
23. Alvesalo JKO, Siiskonen A, Vainio MJ, Tammela PSM, Vuorela PM (2006) J Med Chem 49:2353
24. Boehm M, Wu T-Y, Claussen H, Lemmen C (2008) J Med Chem 51:2468
25. Gasteiger J, Rudolph C, Sadowski J (1990) Tetrahedron Comput Methodol 3:537
26. Pearlman RS (1993) 3D QSAR Drug Des 41
27. OpenEyeScientificSoftware (2006). http://www.eyesopen.com/
28. AccelrysInc (2008) http://www.accelrys.com
29. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Adv Drug Deliv Rev 23:3
30. Wenlock MC, Austin RP, Barton P, Davis AM, Leeson PD (2003) J Med Chem 46:1250
31. Lipinski CA (2004) Drug Disc Today Technol 1:337
32. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) J Med Chem 45:2615
33. Hann MM, Leach AR, Harper G (2001) J Chem Inf Comput Sci 41:856
34. Oprea TI, Davis AM, Teague SJ, Leeson PD (2001) J Chem Inf Comput Sci 41:1308
35. Vainio Mikko J, Johnson Mark S (2007) J Chem Inf Model 47:2462
36. Jorgensen WL (2004) Science 303:1813
37. Davis AM, Keeling DJ, Steele J, Tomkinson NP, Tinker AC (2005) Curr Top Med Chem 5:421
38. Chen B, Harrison RF, Papadatos G, Willett P, Wood DJ, Lewell XQ, Greenidge P, Stiefl N (2007) J Comput Aided Mol Des 21:53
39. Yeap SK, Snarey M, Federico C (2003) Abstracts of Papers, 225th ACS National Meeting, New Orleans, LA, United States, March 23–27
40. Bender A, Glen RC (2005) J Chem Inf Model 45:1369

41. Kogej T, Engkvist O, Blomberg N, Muresan S (2006) J Chem Inf Model 46:1201
42. Joergensen AMM, Langgaard M, Gundertofte K, Pedersen JT (2006) QSAR Comb Sci 25:221
43. Muchmore SW, Debe DA, Metz JT, Brown SP, Martin YC, Hajduk PJ (2008) J Chem Inf Model 48:941
44. Fechner U, Franke L, Renner S, Schneider P, Schneider G (2004) J Comput Aided Mol Des 17:687
45. Weininger D (1994) Spec Publ R Soc Chem 142:67
46. Shemetulskis NE, Weininger D, Blankley CJ, Yang JJ, Humblet C (1996) J Chem Inf Comput Sci 36:862
47. Young SS, Gombar VK, Emptage MR, Cariello NF, Lambert C (2002) Chemom Intell Lab Syst 60:5
48. Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, Schuffenhauer A (2004) Org Biomol Chem 2:3256
49. PipeLinePilot (2008) http://accelrys.com/products/scitegic/
50. Sun H (2005) J Med Chem 48:4031
51. Chemical Computing Group M, Quebec, Canada (2005) www.chemcomp.com
52. Ewing T, Baber JC, Feher M (2006) J Chem Inf Model 46:2423
53. Bender A, Mussa HY, Gill GS, Glen RC (2004) J Med Chem 47:6569
54. Bender A, Jenkins JL, Glick M, Deng Z, Nettles JH, Davies JW (2006) J Chem Inf Model 46:2445
55. Zimmermann M, Hindle SA, Naumann T, Matter H, Hessler G, Baringhaus K-H, Lemmen C, Gastreich M, Rarey M (2003) Abstracts of Papers, 226th ACS National Meeting, New York, NY, United States, September 7–11
56. Hessler G, Zimmermann M, Matter H, Evers A, Naumann T, Lengauer T, Rarey M (2005) J Med Chem 48:6575
57. Blower PE Jr, Cross K, Fligner M, Verducci J (2002) Abstracts of Papers, 223rd ACS National Meeting, Orlando, FL, United States, April 7–11
58. Digital_Chemistry http://www.digitalchemistry.co.uk/
59. Roberts G, Myatt GJ, Johnson WP, Cross KP, Blower PE Jr (2000) J Chem Inf Comput Sci 40:1302
60. Pearlman RS, Smith KM (1998) Perspect Drug Disc Des 9/11:339
61. Pearlman RS, Smith KM (1999) J Chem Inf Comput Sci 39:28
62. Beno BR, Mason JS (2001) Drug Disc Today 6:251
63. Pirard B, Pickett SD (2000) J Chem Inf Comput Sci 40:1431
64. Shanmugasundaram V, Maggiora GM, Lajiness MS (2005) J Med Chem 48:240
65. Boecker A, Sasse BC, Nietert M, Stark H, Schneider G (2007) ChemMedChem 2:1000
66. Carosati E, Mannhold R, Wahl P, Hansen JB, Fremming T, Zamora I, Cianchetta G, Baroni M (2007) J Med Chem 50:2117
67. Carosati E, Budriesi R, Ioan P, Ugenti MP, Frosini M, Fusi F, Corda G, Cosimelli B, Spinelli D, Chiarini A, Cruciani G (2008) J Med Chem 51:5552
68. Franke L, Schwarz O, Mueller-Kuhrt L, Hoernig C, Fischer L, George S, Tanrikulu Y, Schneider P, Werz O, Steinhilber D, Schneider G (2007) J Med Chem 50:2640
69. Shoda M, Harada T, Yano K, Stahura FL, Himeno T, Shiojiri S, Kogami Y, Kouji H, Bajorath J (2007) ChemMedChem 2:515
70. Stanton DT, Ankenbauer J, Rothgeb D, Draper M, Paula S (2007) Bioorg Med Chem 15:6062
71. Yamazaki K, Kusunose N, Fujita K, Sato H, Asano S, Dan A, Kanaoka M (2006) Bioorg Med Chem Lett 16:1371
72. Manetti F, Magnani M, Castagnolo D, Passalacqua L, Botta M, Corelli F, Saddi M, Deidda D, De Logu A (2006) ChemMedChem 1:973
73. Franke L, Schwarz O, Mueller-Kuhrt L, Hoernig C, Fischer L, George S, Tanrikulu Y, Schneider P, Werz O, Steinhilber D, Schneider G (2007) J Med Chem 50:2640

74. AnalytiCon Drug Discovery http://www.ac-discovery.com
75. Schneider G, Neidhart W, Giller T, Schmid G (1999) Angew Chem Int Ed 38:2894
76. Fechner U, Franke L, Renner S, Schneider P, Schneider G (2003) J Comput Aided Mol Des 17:687
77. Fechner U, Franke L, Renner S, Schneider P, Schneider G (2003) J Comput Aided Mol Des 17:687
78. Mannhold R, Berellini G, Carosati E, Benedetti P (2005) In: Cruciani G (ed) Molecular interaction fields in drug discovery (Methods and Principles in Medicinal Chemistry), vol 27. Wiley, Weinheim, Germany, p 173
79. Pastor M, Cruciani G, McLay I, Pickett SD, Clementi S (2000) J Med Chem 43:3233
80. Baroni M, Cruciani G, Sciabola S, Perruccio F, Mason JS (2007) J Chem Inf Model 47:279
81. Van Drie JH (2007) Internet Electron J Mol Des 6:271
82. Kubinyi H (1999) J Recept Signal Transduct Res 19:15
83. Kubinyi H (2003) Nat Rev Drug Disc 2:665
84. Van Drie J (2006) Abstracts of Papers, 231st ACS National Meeting, Atlanta, GA, United States, March 26–30, 2006
85. van Drie JH (2005) Drug Disc Ser 1:157
86. Martin YC (2006) Compr Med Chem II 4:515
87. Martin YC, Bures MG, Danaher EA, DeLazzer J, Kim KH, Lico I, Pavlik PA (1993) Comput Aided Innovation New Mater 2, Proc Int Conf Exhib Comput Appl Mater Mol Sci Eng 1117
88. Guner OF (2002) Curr Top Med Chem 2:1321
89. Wermuth CG (2006) Methods Princ Med Chem 32:3
90. Cohen C, Fischel O, Cohen E (2006) Chem Biol Drug Des 67:182
91. Kurogi Y, Guner OF (2001) Curr Med Chem 8:1035
92. Leitao A, Andricopulo AD, Montanari CA (2007) Curr Methods Med Chem Biol Phys 1:61
93. Cramer RD, Wendt B (2007) J Comput Aided Mol Des 21:23
94. Jilek R, Cramer RD (2006) Abstracts of Papers, 232nd ACS National Meeting, San Francisco, CA, United States, Sept. 10–14
95. Polanski J (2006) Expert Opin Drug Disc 1:693
96. Dixon SL, Smondyrev AM, Rao SN (2006) Chem Biol Drug Des 67:370
97. Oprea TI (2004) Comput Med Chem Drug Disc 571
98. Akamatsu M (2002) Curr Top Med Chem 2:1381
99. Kubinyi H (2008) Comput Struct Approaches Drug Disc 24
100. Debnath AK (2001) Mini-Rev Med Chem 1:187
101. Kubinyi H (1997) Drug Disc Today 2:457
102. Kubinyi H (1997) Drug Disc Today 2:538
103. Connolly Martin Y (1998) Perspect Drug Disc Des 12/14:3
104. Jain AN (2000) J Comput Aided Mol Des 14:199
105. Openeye_Scientific_Software http://en.wikipedia.org/wiki/OpenEye_Scientific_Software
106. McGaughey GB, Sheridan RP, Bayly CI, Culberson JC, Kreatsoulas C, Lindsley S, Maiorov V, Truchon J-F, Cornell WD (2007) J Chem Inf Model 47:1504
107. Wolber G, Langer T (2005) J Chem Inf Comput Sci 45:160
108. Cheeseright T, Mackey M, Rose S, Vinter A (2007) Expert Opin Drug Disc 2:131
109. Cheeseright T, Mackey M, Rose S, Vinter A (2006) J Chem Inf Model 46:665
110. Schuster D, Maurer EM, Laggner C, Nashev LG, Wilckens T, Langer T, Odermatt A (2006) J Med Chem 49:3454
111. Purushottamachar P, Khandelwal A, Chopra P, Maheshwari N, Gediya LK, Vasaitis TS, Bruno RD, Clement OO, Njar VCO (2007) Bioorg Med Chem 15:3413
112. Larbig G, Pickhardt M, Lloyd DG, Schmidt B, Mandelkow E (2007) Curr Alzheimer Res 4:315
113. Michaux C, de Leval X, Julemont F, Dogne J-M, Pirotte B, Durant F (2006) Eur J Med Chem 41:1446
114. Bhattacharjee AK (2006) Lett Drug Des Disc 3:219

115. Bohacek R, Boosalis MS, McMartin C, Faller DV, Perrine SP (2006) Chem Biol Drug Des 67:318
116. Jensen AA, Begum N, Vogensen SB, Knapp KM, Gundertofte K, Dzyuba SV, Ishii H, Nakanishi K, Kristiansen U, Stromgaard K (2007) J Med Chem 50:1610
117. Ray NC, Clark RD, Clark DE, Williams K, Hickin HG, Crackett PH, Dyke HJ, Lockey PM, Wong M, Devos R, White A, Belanoff JK (2007) Bioorg Med Chem Lett 17:4901
118. Bhattacharjee AK, Nichols DA, Gerena L, Roncal N, Gutteridge CE (2007) Med Chem 3:317
119. Agrawal H, Kumar A, Bal NC, Siddiqi MI, Arora A (2007) Bioorg Med Chem Lett 17:3053
120. Lu IL, Huang C-F, Peng Y-H, Lin Y-T, Hsieh H-P, Chen C-T, Lien T-W, Lee H-J, Mahindroo N, Prakash E, Yueh A, Chen H-Y, Goparaju CMV, Chen X, Liao C-C, Chao Y-S, Hsu JTA, Wu S-Y (2006) J Med Chem 49:2703
121. Bhattacharjee AK (2007) Expert Opin Drug Disc 2:1115
122. Tervo AJ, Suuronen T, Kyrylenko S, Kuusisto E, Kiviranta PH, Salminen A, Leppaenen J, Poso A (2006) J Med Chem 49:7239
123. Doddareddy MR, Choo H, Cho YS, Rhim H, Koh HY, Lee J-H, Jeong S-W, Pae AN (2007) Bioorg Med Chem 15:1091
124. Lemmen C, Lengauer T, Klebe G (1998) J Med Chem 41:4502
125. Cruciani G, Pastor M, Guba W (2000) Eur J Pharm Sci 11:29
126. Norin M, Sundstrom M (2002) Trends Biotechnol 20:79
127. Chayen NE (2004) Curr Opin Struct Biol 14:577
128. Chayen NE (2002) Trends Biotechnol 20:98
129. DesJarlais RL, Sheridan RP, Dixon JS, Kuntz ID, Venkataraghavan R (1986) J Med Chem 29:2149
130. DesJarlais RL, Cummings MD, Gibbs AC (2007) Front Drug Des Disc 3:81
131. Kubinyi H, Boehm HJ (1997) Book of Abstracts, 214th ACS National Meeting, Las Vegas, NV, September 7–11 CINF
132. Waszkowycz B (2008) Drug Disc Today 13:219
133. Polgar T, Keseru GM (2007) Front Drug Des Disc 3:477
134. Brenk R, Klebe G (2006) Methods Princ Med Chem 32:171
135. Perola E, Walters WP, Charifson PS (2004) Proteins Struct Funct Bioinf 56:235
136. Cross JB, Jalaie M, Gantt SL, Joshi NA, Wild DJ, Snow ME, Narasimhan LS (2004) Abstracts of Papers, 227th ACS National Meeting, Anaheim, CA, United States, March 28–April 1, 2004 COMP
137. Lemmen C, Rarey M, Schellhammer I (2007) (Biosolveit GmbH, Germany) Application: WO 2007071411, p 27
138. Schellhammer I, Rarey M (2004) Proteins Struct Funct Bioinf 57:504
139. Jansen JM, Martin EJ (2004) Curr Opin Chem Biol 8:359
140. Smith R, Hubbard RE, Gschwend DA, Leach AR, Good AC (2003) J Mol Graphics Modell 22:41
141. Good AC, Cheney DL, Sitkoff DF, Tokarski JS, Stouch TR, Bassolino DA, Krystek SR, Li Y, Mason JS, Perkins TDJ (2003) J Mol Graphics Modell 22:31
142. Good AC, Cheney DL (2003) J Mol Graphics Modell 22:23
143. Agarwal A, Louise-May S, Thanassi JA, Podos SD, Cheng J, Thoma C, Liu C, Wiles JA, Nelson DM, Phadke AS, Bradbury BJ, Deshpande MS, Pucci MJ (2007) Bioorg Med Chem Lett 17:2807
144. Bakir F, Kher S, Pannala M, Wilson N, Nguyen T, Sircar I, Takedomi K, Fukushima C, Zapf J, Xu K, Zhang S-H, Liu J, Morera L, Schneider L, Sakurai N, Jack R, Cheng J-F (2007) Bioorg Med Chem Lett 17:3473
145. Brooks WH, McCloskey DE, Daniel KG, Ealick SE, Secrist JA III, Waud WR, Pegg AE, Guida WC (2007) J Chem Inf Model 47:1897
146. Cherkasov A, Ban F, Li Y, Fallahi M, Hammond GL (2006) J Med Chem 49:7466

147. de Graaf C, Oostenbrink C, Keizers PHJ, van der Wijst T, Jongejan A, Vermeulen NPE (2006) J Med Chem 49:2417
148. Desai PV, Patny A, Gut J, Rosenthal PJ, Tekwani B, Srivastava A, Avery M (2006) J Med Chem 49:1576
149. Dooley AJ, Shindo N, Taggart B, Park J-G, Pang Y-P (2006) Bioorg Med Chem Lett 16:830
150. Foloppe N, Fisher LM, Howes R, Potter A, Robertson AGS, Surgenor AE (2006) Bioorg Med Chem 14:4792
151. Furci LM, Lopes P, Eakanunkul S, Zhong S, MacKerell AD Jr, Wilks A (2007) J Med Chem 50:3804
152. Hirayama K, Aoki S, Nishikawa K, Matsumoto T, Wada K (2007) Bioorg Med Chem 15:6810
153. Kenyon V, Chorny I, Carvajal WJ, Holman TR, Jacobson MP (2006) J Med Chem 49:1356
154. Knox AJS, Meegan MJ, Sobolev V, Frost D, Zisterer DM, Williams DC, Lloyd DG (2007) J Med Chem 50:5301
155. Leban J, Baierl M, Mies J, Trentinaglia V, Rath S, Kronthaler K, Wolf K, Gotschlich A, Seifert MHJ (2007) Bioorg Med Chem Lett 17:5858
156. Li J, Chen J, Gui C, Zhang L, Qin Y, Xu Q, Zhang J, Liu H, Shen X, Jiang H (2006) Bioorg Med Chem 14:2209
157. Liu H, Gao Z-B, Yao Z, Zheng S, Li Y, Zhu W, Tan X, Luo X, Shen J, Chen K, Hu G-Y, Jiang H (2007) J Med Chem 50:83
158. Louise-May S, Yang W, Nie X, Liu D, Deshpande MS, Phadke AS, Huang M, Agarwal A (2007) Bioorg Med Chem Lett 17:3905
159. Lu IL, Mahindroo N, Liang P-H, Peng Y-H, Kuo C-J, Tsai K-C, Hsieh H-P, Chao Y-S, Wu S-Y (2006) J Med Chem 49:5154
160. Mallya M, Phillips RL, Saldanha SA, Gooptu B, Brown SCL, Termine DJ, Shirvani AM, Wu Y, Sifers RN, Abagyan R, Lomas DA (2007) J Med Chem 50:5357
161. Montes M, Braud E, Miteva MA, Goddard M-L, Mondesert O, Kolb S, Brun M-P, Ducommun B, Garbay C, Villoutreix BO (2008) J Chem Inf Model 48:157
162. Neres J, Bonnet P, Edwards PN, Kotian PL, Buschiazzo A, Alzari PM, Bryce RA, Douglas KT (2007) Bioorg Med Chem 15:2106
163. Richardson CM, Nunns CL, Williamson DS, Parratt MJ, Dokurno P, Howes R, Borgognoni J, Drysdale MJ, Finch H, Hubbard RE, Jackson PS, Kierstan P, Lentzen G, Moore JD, Murray JB, Simmonite H, Surgenor AE, Torrance CJ (2007) Bioorg Med Chem Lett 17:3880
164. Rogers JP, Beuscher AE, Flajolet M, McAvoy T, Nairn AC, Olson AJ, Greengard P (2006) J Med Chem 49:1658
165. Spannhoff A, Machmur R, Heinke R, Trojer P, Bauer I, Brosch G, Schuele R, Hanefeld W, Sippl W, Jung M (2007) Bioorg Med Chem Lett 17:4150
166. Spannhoff A, Heinke R, Bauer I, Trojer P, Metzger E, Gust R, Schuele R, Brosch G, Sippl W, Jung M (2007) J Med Chem 50:2319
167. Szewczuk LM, Saldanha SA, Ganguly S, Bowers EM, Javoroncov M, Karanam B, Culhane JC, Holbert MA, Klein DC, Abagyan R, Cole PA (2007) J Med Chem 50:5330
168. Tsai K-C, Chen S-Y, Liang P-H, Lu IL, Mahindroo N, Hsieh H-P, Chao Y-S, Liu L, Liu D, Lien W, Lin T-H, Wu S-Y (2006) J Med Chem 49:3485
169. Wang JG, Xiao YJ, Li YH, Liu XH, Li ZM (2006) Chin Chem Lett 17:1555
170. Zhou Y, Peng H, Ji Q, Qi J, Zhu Z, Yang C (2006) Bioorg Med Chem Lett 16:5878
171. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M (2005) J Med Chem 48:4040
172. Schulz-Gasch T, Stahl M (2004) Drug Disc Today Technol 1:231
173. Schulz-Gasch T, Stahl M (2003) J Mol Model 9:47
174. Stahl M, Rarey M (2001) J Med Chem 44:1035
175. Spyrakis F, Kellogg GE, Amadasi A, Cozzini P (2007) Front Drug Des Disc 3:317
176. Zhou Z, Felts AK, Friesner RA, Levy RM (2007) J Chem Inf Model 47:1599
177. Spyrakis F, Amadasi A, Fornabaio M, Abraham DJ, Mozzarelli A, Kellogg GE, Cozzini P (2007) Eur J Med Chem 42:921

178. Montes M, Miteva MA, Villoutreix BO (2007) Proteins Struct Funct Bioinf 68:712
179. Teramoto R, Fukunishi H (2007) J Chem Inf Model 47:526
180. Betzi S, Suhre K, Chetrit B, Guerlesquin F, Morelli X (2006) J Chem Inf Model 46:1704
181. Feher M (2006) Drug Disc Today 11:421
182. Baber JC, Shirley WA, Gao Y, Feher M (2006) J Chem Inf Model 46:277
183. Joseph-McCarthy D, Baber JC, Feyfant E, Thompson DC, Humblet C (2007) Curr Opin Drug Disc Dev 10:264
184. Onodera K, Satou K, Hirota H (2007) J Chem Inf Model 47:1609
185. Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Proteins Struct Funct Bioinf 57:225
186. Bowman AL, Lerner MG, Carlson HA (2007) J Am Chem Soc 129:3634
187. Lerner MG, Bowman AL, Carlson HA (2007) J Chem Inf Model 47:2358
188. Koska J, Spassov VZ, Maynard AJ, Yan L, Austin N, Flook PK, Venkatachalam CM (2008) J Chem Inf Model 48:1965
189. Venkatachalam CM (2007) Abstracts, 41st Western Regional Meeting of the American Chemical Society, San Diego, CA, United States, October 9–13 GEN
190. Cavasotto CN, Orry AJW (2007) Curr Top Med Chem 7:1006
191. Agrawal H, Kumar A, Bal NC, Siddiqi MI, Arora A (2007) Bioorg Med Chem Lett 17:3053
192. Keck JL, Roche DD, Lynch AS, Berger JM (2000) Science 287:2482
193. Goodford P (2006) In: Cruciani G, Mannhold R, Kubinyi H, Folkers G (eds) Molecular interaction fields: applications in drug discovery and ADME prediction. Wiley, New York, p 3
194. Alvesalo JKO, Siiskonen A, Vainio MJ, Tammela PSM, Vuorela PM (2006) J Med Chem 49:2353
195. Schluckebier G, Zhong P, Stewart KD, Kavanaugh TJ, Abad-Zapatero C (1999) J Mol Biol 2:277
196. Rarey M, Kramer B, Lengauer T, Klebe G (1996) J Mol Biol 3:470
197. Furci LM, Lopes P, Eakanunkul S, Zhong S, MacKerell AD Jr, Wilks A (2007) J Med Chem 50:3804
198. Friedman J, Lad L, Deshmukh R, Li H, Wilks A, Poulos TL (2003) J Biol Chem 278:34654
199. Montes M, Braud E, Miteva MA, Goddard M-L, Mondésert O, Kolb S, Brun M-P, Ducommun B, Garbay C, Villoutreix BO (2008) J Chem Inf Model 48:157
200. Reynolds RA, Yem AW, Wolfe CL, Deibel MR Jr, Chidester CG, Watenpaugh KD (1999) J Mol Biol 293:559
201. Li R, Xue L, Zhu T, Jiang Q, Cui X, Yan Z, McGee D, Wang J, Gantla VR, Pickens JC, McGrath D, Chucholowski A, Morris SW, Webb TR (2006) J Med Chem 49:1006
202. Barreiro G, Guimaraes CRW, Tubert-Brohman I, Lyons TM, Tirado-Rives J, Jorgensen WL (2007) J Chem Inf Model 47:2416
203. Guichou J-F, Viaud J, Mettling C, Subra G, Lin Y-L, Chavanieu A (2006) J Med Chem 49:900
204. Hartzoulakis B, Rossiter S, Gill H, O’Hara B, Steinke E, Gane PJ, Hurtado-Guerrero R, Leiper JM, Vallance P, Rust JM, Selwood DL (2007) Bioorg Med Chem Lett 17:3953
205. Hu X, Prehna G, Stebbins CE (2007) J Med Chem 50:3980
206. Kellenberger E, Springael J-Y, Parmentier M, Hachet-Haas M, Galzi J-L, Rognan D (2007) J Med Chem 50:1294
207. Liao C, Karki RG, Marchand C, Pommier Y, Nicklaus MC (2007) Bioorg Med Chem Lett 17:5361
208. Lu Y, Nikolovska-Coleska Z, Fang X, Gao W, Shangary S, Qiu S, Qin D, Wang S (2006) J Med Chem 49:3759
209. Rella M, Rushworth CA, Guy JL, Turner AJ, Langer T, Jackson RM (2006) J Chem Inf Model 46:708
210. Saario SM, Poso A, Juvonen RO, Jaervinen T, Salo-Ahen OMH (2006) J Med Chem 49:4650
211. Tintori C, Manetti F, Veljkovic N, Perovic V, Vercammen J, Hayes S, Massa S, Witvrouw M, Debyser Z, Veljkovic V, Botta M (2007) J Chem Inf Model 47:1536

212. Barreca ML, De Luca L, Iraci N, Rao A, Ferro S, Maga G, Chimirri A (2007) J Chem Inf Model 47:557
213. Das K, Clark AD Jr, Lewi PJ, Heeres J, De Jonge MR, Koymans LM, Vinkers HM, Daeyaert F, Ludovici DW, Kukla MJ, De Corte B, Kavash RW, Ho CY, Ye H, Lichtenstein MA, Andries K, Pauwles R, Debethune M-P, Boyer PL, Clark P, Hughes SH, Janssen PA, Arnold E (2004) J Med Chem 47:2550
214. Murray-Rust J, Leiper J, McAlister M, Phelan J, Tilley S, Santa Maria J, Vallance P, McDonald N (2001) Nat Struct Biol 8:679
215. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) J Med Chem 51:3661
216. Congreve M, Murray CW, Carr R, Rees DC (2007) Annu Rep Med Chem 42:431
217. Hajduk PJ, Greer J (2007) Nat Rev Drug Disc 6:211
218. Hajduk PJ, Huth JR, Sun C (2006) Methods Princ Med Chem 34:181
219. Hajduk PJ (2006) Mol Interventions 6:266
220. Batista J, Bajorath J (2007) J Chem Inf Model 47:59
221. Crisman TJ, Sisay MT, Bajorath J (2008) J Chem Inf Model 48:1955
222. Chemical_Computing_Group http://www.chemcomp.com
223. Rummey C, Nordhoff S, Thiemann M, Metz G (2006) Bioorg Med Chem Lett Letters 16:1405
224. Krallinger M, Erhardt RA-A, Valencia A (2005) Drug Disc Today 10:439