Integration of Virtual Screening into the Drug Discovery Process

D.N. Chin, C.E. Chuaqui and J. Singh*

Computational Drug Design Group, Biogen Idec, Inc, 14 Cambridge Center, Cambridge, MA 02142, USA

Abstract: Advances in high-throughput virtual screening using docking, predictive ADME methods and their integration with informatics and high-performance computing are reviewed. Docking approaches have led to the identification of novel active compounds. Predictive ADME methods have improved on selective test sets with broader training sets, though extensive validation is lacking.

Keywords: Virtual Screening, Docking, Predictive ADME, Informatics.

VIRTUAL SCREENING AND DRUG DISCOVERY

Pharmaceutical companies are adopting a new paradigm of integrating computational and experimental technologies earlier in the drug discovery process [1-4]. This new paradigm is a response to the problem that the development of new drugs is not getting more efficient despite significant investment in research and technologies by the pharmaceutical industry [5, 6]. The failure or attrition rates of compounds during the course of drug development is typically high, [6] and previous major technologies such as combinatorial chemistry and high throughput screening (HTS) by themselves have not delivered as promised [3-6].

The current solution to problems in the drug-discovery process is seen as better utilization of genomic, chemical, biological, structural, and molecular property information earlier, [7] and improved management of procedural bottlenecks [8, 9]. Better utilization of information, in particular, plays an important role in designing chemical libraries that have more desirable chemical and biological properties, [5, 10] and are therefore more likely to contain compounds that can survive primary *in vitro* and secondary *in vivo* screens [11, 12]. A current approach to this problem is the integration of information from many areas such as gene or protein family analysis for target selection, virtual screening (VS), structural biology, medicinal chemistry, and ADME as shown in (Fig. **1**).

VS in particular is emerging as a key strategy to help filter out compounds with poor potency, and biological or pharmacological properties [7, 13, 14]. There are two key factors driving the importance of using and developing VS methods to decide on which compounds to progress to *in vitro* and *in vivo* screening. First, chemical space is massive and chemical library size quickly becomes an issue. For example, a three-component reaction with 100 reagents per component results in a combinatorial library with 100x100x100 (10^6) products. There have been estimates that the size of chemical space appropriate for drugs is theoretically > 10^{18}, [13] and only computational methods are able to address this problem. Second, experimental screening is expensive and time consuming. Predicting chemical and biological properties with VS methods can in principle identify compounds that are likely to fail in primary, secondary and further downstream screens at a significantly lower cost than experiments.

VS methods, though far from perfect, can be used to reduce the possibilities in chemical space to a manageable number of potent and "drug-like" candidates for lead discovery and optimization [13, 15]. These methods have grown to include 2D searches of databases, [16] structure-based screening (receptor docking, [17-19] pharmacophore screening [20]), and predictive ADME [21]. Receptor docking, in particular, has grown in importance not only due to advances in methodology, but also due to the availability of many new 3D protein structures of drug targets [13, 17-20, 22-26]. The advent of high performance computing and fast docking algorithms, in addition, have allowed database screening to be more routine. Other important pharmacophore or fragment (*de novo*) based methods will not be discussed in this review, but have been reviewed elsewhere in detail [16, 27].

A key element in an integrated approach to drug discovery is the concept of property-based drug design, [28-30] where efforts are made earlier to optimize for fundamental drug properties such as aqueous solubility and permeability. The goal is that by incorporating a variety of predictive ADME filters into the virtual screening workflow, it will be possible to address ADME issues early on and reduce the downstream failure rates of drug candidates due to poor pharmacokinetic properties.

Drug-discovery programs produce a significant amount of information (for example, structure-activity relationships, NMR, crystallography, *in vitro* and *in vivo* data) that can be leveraged by VS for more effective predictions. Conversely, data from VS such as predicted binding mode patterns can also be "mined" for useful clues that can be used to design libraries with improved hit rates. The informatics infrastructure to support and analyze experimental and virtual data will clearly be important and enable faster and better decisions for the next library design cycle (Fig. **1**). In the new drug-discovery paradigm, VS is a complement to experimental technologies, which is useful in influencing the direction of the drug-design program towards the desired goal in addition to predicting actives from non-actives in a given screen [3, 7, 9, 14, 31]. Our review will therefore highlight recent work in target selection, fast library enumeration, predictive ADME, structure-based docking

*Address correspondence to this author at the Computational Drug Design Group, Biogen Idec, Inc, 14 Cambridge Center, Cambridge, MA 02142, USA; E-mail: Juswinder.Singh@biogenidec.com

1389-5575/04 $45.00+.00

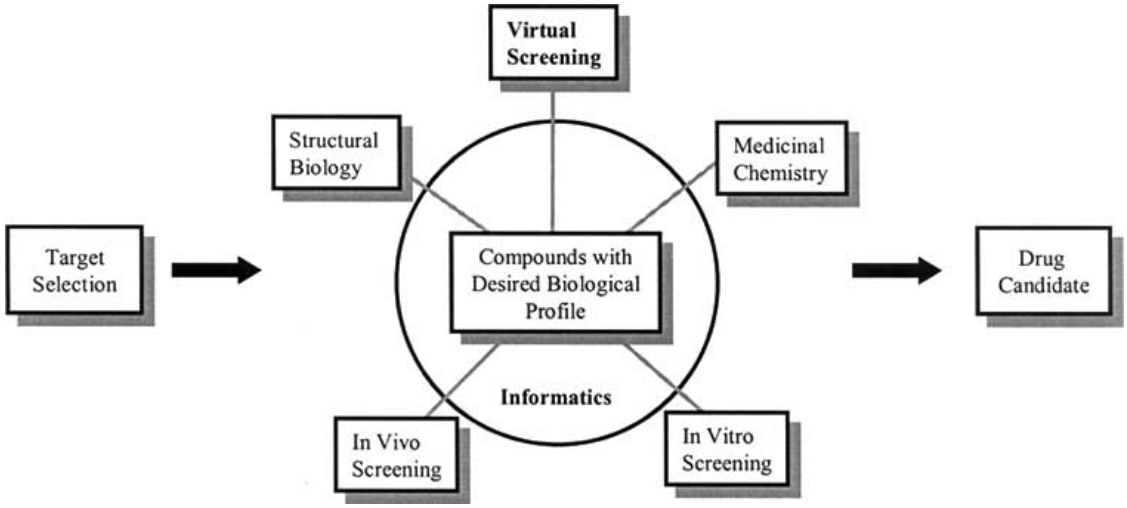


Fig. (1). A schematic showing the integration of VS with a parallel drug-discovery paradigm.

(including scoring functions), and informatics infrastructure that support high throughput VS.

TARGET SELECTION AND GENE-FAMILIES

The Boston Consulting Group (BCG) has estimated that the current number of known protein targets amenable to VS using structure-based drug design is 30% [32]. This number is expected to rise dramatically with the advent of structural genomics and homology modeling initiatives [33]. To date, the drug industry has focused on only 483 drug targets, [5] and the sequencing of the 30,000 genes in the human genome [34, 35] has generated a wealth of potential new drug targets for the pharmaceutical industry. It is not clear, however, how many of these new targets are amenable to small molecule intervention ("druggable"). Bailey *et al.* [36] estimate the number of druggable targets in the human genome to be >10,000 by assessing the number of proteins with ligand binding domains, however, estimates using other criteria are much lower [5, 37]. Several groups are developing databases of protein-binding sites to be used to prioritize the druggability of the target [38]. The gene-family concept is emerging as a powerful paradigm to accelerate drug discovery through leveraging structural insights into druggability and transferability of chemotypes amongst related drug targets [39]. VS methods are therefore poised to play an important role in addressing the increased number of targets with structure.

BUILDING AND FAST ENUMERATION OF VIRTUAL LIBRARIES

Electronic methods for building *in silico* chemical libraries by 'reaction' or 'Markush' representations can be performed with standard chemical drawing packages such as IsisDraw [40] and ChemDraw [41]. For example, a Markush library can be drawn in ISISDraw as a core and R-groups (including nested R-groups) descriptions [40]. The Markush description can then be enumerated or assembled into compounds by algorithms [10]. The massive number of compounds that can result from virtual libraries, however, present particular problems for enumeration and molecular property calculations [10]. Fast methods for enumerating very large virtual libraries and computing molecular properties have therefore emerged to compliment the throughput of VS methods.

An example of a fast enumeration method of Markush library definitions has been developed by Barnard, [42] and is based on string concatenation. The method quickly assembles molecules and computes limited properties (MW, LogP, number of hydrogen-bond donors and acceptors) by adding the values of the fragments (computed on the fly) in the Markush definition to the final enumerated compound. The speed of the enumeration for Markush libraries is very fast. We tested the method by enumerating and calculating the Lipinski properties of a 1 million compound library and found that it took only 20 seconds using an SGI R10000 workstation.

Another approach for fast library enumeration called combinatorial networks involves first building a diverse representation of a library definition, second developing a neural network model on the initial set, and third using the model to forecast the structures and properties of the remaining molecules in the library [43]. The authors compared the results from their approach to that from fully enumerated methods and showed that the chemical and property space spanned by the systematic and predicted methods were very similar. The method, however, works best when the library is relatively similar chemically—as might be found in combinatorial libraries—and may not be appropriate for libraries with very different structures of molecules.

PREDICTIVE DRUG LIKENESS AND *IN VIVO* EFFECTS

This section will discuss filters that may be used to incorporate good ADME properties into compound libraries at the outset of an integrated drug-discovery program. For ADME filters to be most effective (predictive) for VS they should optimally meet the following criteria: 1) be rapid enough to deal effectively with millions of structures; 2) be appropriate for relevant drug compounds (that is, as opposed

to models that are based on non-drug data); 3) be sufficiently flexible so that they can be quickly updated as new data become available. Ideally a "toolbox" of filters would be at our disposal, ranging from very general filters to select "drug-like" molecules (which we will describe below) for initial lead discovery applications, to custom models designed to identify molecules with specific properties, e.g., central nervous system (CNS) activity. In order to satisfy the first criterion, this section will focus on methods that do not require 3D molecular coordinates, but rather, depend only on 1D (based on molecular composition) and 2D (based on atomic connectivity) descriptors. In contrast, two examples of well-known approaches that rely on 3D structure are the VolSurf methodology developed by Cruciani *et al.* [44] and modeling of P450 mediated drug metabolism by Ekins and co-workers [45]. VolSurf requires the computation of 3D interaction energy grids with a set of atomic probes to arrive at a set of descriptors used for model building whereas the P450 models may involve homology modeling, pharmacophore modeling, and the generation of 3D-quantitative structure-activity relationships. Although both approaches have produced good results in a variety of applications, [46-49] the computational overhead associated with descriptor generation renders them unsuitable to profile large virtual libraries fast.

Prediction of Drug-Likeness

As pointed out by several authors, "drug like" is not a precisely defined term [11, 50]. In general, however, "drug like" refers to the similarity between a molecule and a set of known drugs, quantified using descriptors based on molecular structure (atom types, functional groups, molecular topology, and common drug frameworks) and global physicochemical properties. Recent reviews have appeared on the prediction of drug-likeness, [50, 51] and the generation of drug-like libraries [11]. The most widely adopted rules for filtering compound libraries were derived by Lipinski, based on counting criteria of physicochemical properties correlated to poor absorption [52]. More recently, Wenlock *et al.* [53] have carried out an analysis of the physicochemical property profiles of oral drugs. The mean molecular weight of drugs was found to decrease in successive phases of clinical development converging to the mean value of those already in the market. In addition to molecular weight, lipophilicity was also found to be a limiting factor in development, supporting the work of Lipinski.

A variety of methods have been employed to tackle the drug-like recognition problem including classification models based on structural decision trees, [54] pharmacophore point filters based on simple structural rules, [55] as well as various rule-based filters (counting methods) based on the properties of known drugs [52, 56, 57]. Some of the more successful approaches are based on neural networks, which are able to recognize approximately 77-90% of compounds in drug databases [58-60] depending on the method, while having a mis-classification rate of approximately 10-18% [61-64]. An interesting alternative to the various neural network based approaches employs a pharmacophore point filter based on simple structural rules [55]. At the expense of accuracy, the approach provides a structural rationale for the drug/non-drug classification. As pointed out by Walters and Murcko, [50] the generation of "local" filters optimized to recognize on particular sub-classes of drugs may offer improved performance over global drug-like filters.

Prediction of Intestinal Absorption and Oral Bioavailability

Computational methods to predict oral absorption have been reviewed recently [30, 65-69]. The challenges in predicting oral absorption from molecular structure have been discussed by Burton *et al.* [70]. In particular, they highlight the importance in discriminating between intestinal absorption, permeability, fraction absorbed (FA), and oral bioavailability (OBA) when constructing structure-based predictive models. OBA indicates the fraction of the oral dose that reaches systemic circulation and is therefore influenced by intestinal absorption and metabolism. The drug's intestinal absorption, in turn, depends on its chemical stability, solubility, and membrane permeability. Consequently, high aqueous solubility can compensate for poor permeability and lead to acceptable intestinal absorption, complicating efforts to use permeability alone as a surrogate for intestinal absorption.

Prediction of drug solubility has recently attracted the attention of several reviewers [11, 68, 71]. The problematic nature of experimental solubility data must be addressed in efforts to derive predictive models as variations in pH, crystalline form, and experimental conditions can all lead to significant uncertainty in reported values [71]. Recently, Enqkvist *et al.* [72] have built a solubility model using a neural network trained on 3042 molecules from the PHYSPROP [73] database. The final model utilized 63 1D and 2D descriptors and yielded an r^2=0.86 for an external validation set of 307 molecules. Also using a set of 1D and 2D descriptors, Liu *et al.* [74] generated a predictive solubility model based on 1312 organic compounds compiled from the AQUASOL [75] and PHYSPROP [73] databases. The model achieved a correlation coefficient of 0.93 when tested against a validation set of 258 randomly selected compounds set aside from the initial training set.

Particular attention has been devoted to the use of molecular surface properties alone or in combination with other descriptors to model permeability and its relationship to intestinal absorption [30, 76-83]. Results obtained from rapid fragment based methods for computation of polar surface area (PSA) has been shown to be comparable to methods requiring 3D structure generation [84-86]. Approaches not based primarily on molecular surface properties to model permeability include quantitative structure property relationship (QSPR) models based on 3D descriptors, [87] topological descriptors, [88] and Abraham's free energy descriptors to model solvation. [89-91]

In addition to approaching the prediction of absorption via permeability, models have been derived to yield a direct prediction of human intestinal absorption (HIA) from molecular structure. Wessel *et al.* [92] have used literature values of FA of 86 drug and drug-like molecules to derive and validate a QSPR model. The dataset was divided into 76 compounds for model development and 10 compounds

for validation, where the validation set was not used in the model development process and was chosen to span the range of 5-100% HIA. The final model was based on six 2D and 3D descriptors and had a root mean square (RMS) error 16% HIA for the validation set. This same data set was modeled by Niwa [93] using 2D topological descriptors rather than the 3D molecular surface descriptors used by Wessel. Although the RMS error of 22.8% obtained using the same validation set is poorer than that obtained by Wessel, greater throughput is afforded by eliminating the need for 3D structure generation.

The molecular property profiles for over 1100 drug candidates studied at SmithKline Beecham Pharmaceuticals (now GlaxoSmithKline) have been analyzed in order to gauge their relative importance in influencing OBA [94]. Low molecular flexibility (number of rotatable bonds ≤ 10) and PSA ≤ 140) were found to be the most significant determinants of OBA. Yoshida and Topliss [95] have derived a QSPR model, based on 1D and 2D descriptors and counting rules, for human OBA from literature data for 232 drugs. In order to account for experimental variability, the OBA data was divided into four classes based on activity. The final model was able to correctly classify 60% of a 40 compound validation set not included in model development. Andrews *et al.* [96] have also developed a QSPR model for human OBA applying 85 substructure descriptors to model data (literature and internal) for 591 compounds. The model was found to have a leave-one-out cross-validated r^2=0.63, with an RMS error of 20.4% averaged over 2000 random selections of a prediction set (a 20% subset of the total dataset). In comparison with classification using the Lipinski "Rule of 5", the QSPR model returns a significantly lower number of false positive classifications.

Prediction of CNS activity

Computational methods developed to predict blood brain barrier (BBB) penetration have been reviewed recently [51, 67, 97-99]. In general, the descriptors that play an important role in reported QSPR models describe the lipophilicity, hydrogen-bonding potential, and molecular bulkiness. [100-102] Particularly appealing for its simplicity is Clark's QSPR model for the log of the blood-brain partition coefficient (logBB) derived from a set of 55 diverse organic compounds using only PSA and calculated logP. [100] Ajay *et al.* [103] have reported a filter based on 1D and 2D descriptors aimed to design CNS-active libraries. Moreover, Engkvist *et al.* [102] have demonstrated that a substructure analysis approach, based on exhaustive enumeration of 2D chemical fragments up to a prescribed size, is able to perform as well as (80% accuracy) neural network based approaches.

STRUCTURE-BASED VIRTUAL SCREENING

Recent and Popular Methods for Docking

For the purpose of this review, we will refer to methods appropriate for high throughput receptor-based VS simply as 'docking'. These docking methods have increased in popularity as a response to the increased throughput of combinatorial chemistry technologies and the availability of 3D structures of drug targets [13, 17, 18, 26, 31]. The goal of docking is to predict quickly (1-2 min/molecule) the binding mode and binding affinity of molecules to a receptor target. The primary problem in docking is that the scoring functions used with these methods for ranking ligand binding do not properly capture all of the possible events (or the relative contributions among these events) that can contribute to binding. The hope is, nevertheless, that these functions capture most of the important events (e.g. complimentary fit, hydrogen bonds) for the system studied, and that there is a significant probability that the analyst will find more true actives than false positives or negatives.

Table **1** lists some, but not all, of the more popular or recently developed methods. As details of the algorithms used in these methods can be found in the literature, we have

Table 1. Examples of New or Popular High Throughput Docking Methods for VS

Method	[a]Receptor	Ligand Conformers	Automatic [b]Filter or [c]Constraint	Reference
DOCK	Rigid	On the Fly/Precompute	Filter/Constraint	[106]
AutoDock	Rigid	On the Fly	Filter	[107]
Gold	Rigid	On the Fly	Constraint	[108]
ICM	Rigid/Flexible	On the Fly	Filter/Constraint	[109]
FlexX	Rigid	On the Fly	None	[110]
FlexPharm	Rigid	On the Fly	Filter	[111]
FlexE	Ensemble	On the Fly	None	[112]
Fred	Rigid	Precompute	Filter	[113]
Glide	Rigid	On the Fly	Filter	[114]
LigandFit	Rigid	On the Fly	None	[115]
EUDOC	Rigid	Single 3D	None	[116]

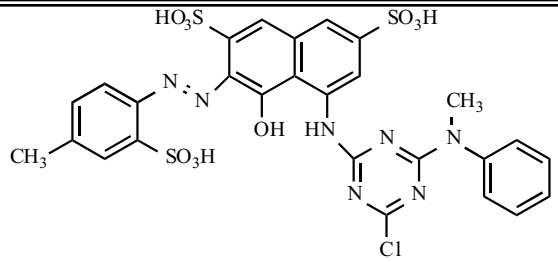
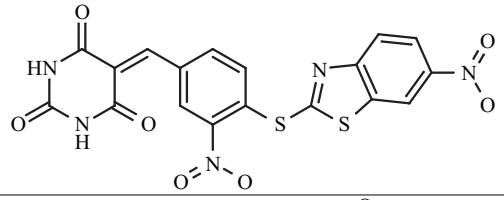
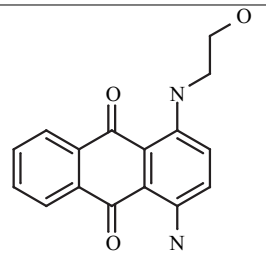
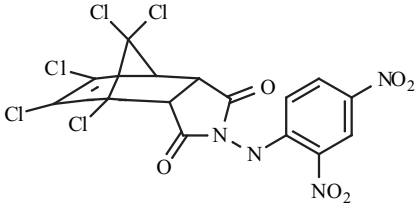
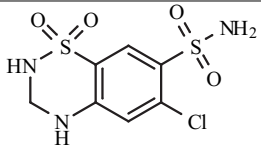
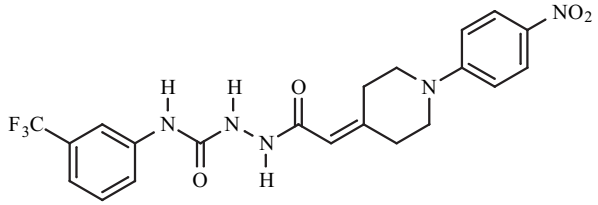
[a]A receptor is rigid or flexible in that discrete conformations of the receptor are explored automatically by the docking algorithm. In principle, grid-based methods such as AutoDock, DOCK, and LigandFit can approximate flexible receptors by weighting the grid locations by the average effect of multiple conformations. [b]We define a filter as using a criteria to eliminate compounds before or after the compound has been optimized by the docking algorithm. [c]We define a constraint as a term that is a part of the main function being optimized by the docking algorithm.

chosen to categorize these methods based on how the receptor and ligands are treated, and if any constraints [104] or filters [105] are used during the screening, as we feel that these attributes may help the reader choose a method for a particular system.

As can be seen in Table **1**, the receptor is usually described as fixed during the screening, but methods such as ICM and FlexE introduce receptor flexibility during a docking run (although flexibility can be addressed in part by the other methods by multiple docking runs of different receptor conformation). ICM, however, can introduce side-chain flexibility as another variable during ligand docking. We note that molecular mechanics methods such as CHARMM [117] and AMBER [118] can also in principle introduce full flexibility into the docking process, but the suitability of these programs to VS has not yet fully matured.

Most methods in Table **1** can introduce ligand flexibility during the docking process, while FRED and DOCK can use pre-computed ligand conformers (usually computed without the receptor present). The primary advantage of pre-computed conformers is speed, as the conformers are generated once, and only translational and rotational degrees of freedom are explored during docking. The disadvantage is that the true conformation of the binding mode in the receptor may not be contained within the pre-computed conformers (which are

Table 2. Examples of where Docking has Produced Previously Unknown Leads from Database Screens

Ligand Structure	Target	Method	Assay	Reference
	Anti-HIV-1	DOCK	3.18 μg/ml (IC_{50})	[119]
	Tyrosine Phosphatase-1B	NW DOCK	4.1 μM (IC_{50})	[121]
	TAR RNA	DOCK ICM	1 μM (CD_{50})	[122]
	Farnesyltransferase	EUDOC	25 μM (IC_{50})	[123]
	Carbonic Anhydrase	FlexS FlexX	5.26 nM (IC_{50})	[124]
	Thyroid Hormone Receptor	ICM	1.5 μM (IC_{50})	[125]

typically done in the gas phase), and conformer pre-generation introduces another computational complexity in the VS process. Using experimental knowledge during VS can be an important advantage during an integrated drug-discovery program. To this end, Glide, FlexPharm, Fred, ICM, and DOCK allow the user to automatically filter out compounds either before or after they are scored that do not, for example, satisfy particular geometrical arrangements of ligand and receptor atoms. ICM, Gold, and DOCK can also include a constraint as part of the scoring function that is optimized during the docking. For example, particular hydrogen bond-formations or ligand conformations can be biased to occur during the docking process.

Database Screening Applications

More examples of identifying previously unknown active compounds from the docking of databases are being published. In addition, several publications have compared multiple, but not all, methods for their ability to discriminate known binders from among randomly selected and assumed to be non-binders. A review of the types of targets, docking procedure, and validated predictions of the docking applications follows, and representative examples of the types of novel compounds found from docking are shown in Table **2**.

DOCK was used to identify μM leads for the HIV-1 envelope glycoprotein (gp41) from a commercially available database of 20,000 compounds from ComGenex Inc. (Table **2**) [119]. The target of the structure-based docking strategy was a deep hydrophobic pocket identified within the X-ray structure of gp41. CONCORD [120] was used to precompute the global minimum 3D conformation (as judged by CONCORD) for each molecule in the database. The 3D database was then docked into a rigid receptor and analyzed.

The authors then selected the top 200 scoring ligands from DOCK, and used 3D stereo visualization to further filter those compounds that had appropriate interactions with the receptor (e.g. atomic shape similarity and hydrogen bonds). This visual analysis led the authors to select 16 compounds for inhibition activity by ELISA. Two compounds were found to have low (0.73 and 3.18 μg/ml) IC_{50} inhibition (Table **2**), while the other 14 were >100 μg/ml. An interesting observation is that the two low μM compounds had MW of 1177 and 802 respectively, while the other non-binders were no greater than 660 MW. This observation suggests the scoring function in DOCK, a sum of van der Waals and Coulombic terms, is suitable for this system that seems to discriminate ligands based roughly on size; a result that is not always expected due to the potential bias in this function to molecular size.

A direct comparison of HTS and docking screening against tyrosine phosphatase-1B (PTP1B) was done by Doman and coworkers [121]. PTP1B is an enzyme that hydrolyzes phosphotyrosines on the insulin receptor resulting in deactivation. The Northwestern version [126] of DOCK 3.5 [127] was used to identify 127 compounds with <100 μM IC_{50} from a database comprising ACD, BioSpecs, and Maybridge with a hit rate of 34.8% (Table **2**).

The HTS screen yielded 85 compounds <100 μM with a hit rate of 0.021% using the corporate database of Pharmacia. The conformations of the ligands for docking were precomputed (an average of 345 conformations per molecule) and stored in a database. The 19.5 billion docked complexes were filtered by steric fit followed by scoring with van der Waals and electrostatic terms that were corrected for apolar and polar desolvation energies. An interesting finding by the authors is that the hits from docking were more "drug like" than those from HTS as judged by Lipinski parameters. This observation, however, may have resulted from the fact that corporate databases are more likely to be "drug like" as compared to the commercially available compounds used above for docking, which are more "lead like" and therefore less likely to violate the Lipinski rules [15]. The results from docking were nevertheless different chemically and structurally than those from HTS, which highlights the complementarities of the methods.

Perola and coworkers used the program EUDOC to perform a database screen to identify 21 compounds that had μM inhibition to Farnesyltransferase (FT) (four in the range of 25-100 μM) [123]. FT is characterized by a Zn atom that is the center of catalytic activity in its binding site. Classes of proteins that contain Zn or other metals in the active site that interact with a ligand such as FT are a challenging case for docking as few scoring functions adequately account for metal-ligand interactions. Their approach was to use a series of filtering steps to remove compounds that are reactive, receptor specific (e.g. remove zwitterionic compounds) and outside the molecular weight range of most drugs (300-700) from the ACD 3D database. The remaining compounds were docked with their program EUDOC into the Farnesyltransferase (FT) receptor including the Zn atom in the active site. The scoring function they used was the Cornell *et al.* version of the AMBER force field, [128] which contains parameters for Zn binding and other metals. Compounds that had docked scores < -35 kcal/mol (AMBER) were selected for more detailed docking studies; compounds with subsequent scores < -45 kcal/mol were selected for further studies. These remaining compounds were then filtered using AMSOL [129] to remove compounds that were too hydrophobic (solvation energies greater than that of neopentane, 2.5 kcal/mol). The final binding energy of the docked compounds was estimated as the EUDOC interaction energy (van der Waals and Coulombic terms from AMBER) minus the AMSOL solvation energy. The final set of 128 compounds from docking were those below the estimated binding energy value of –33 kcal/mol, and 21 were tested experimentally. An unusual feature in their approach was to use only the existing 3D conformer in the database (no generation of alternative conformers), with the hope that the database was sufficiently large enough to identify different subsets of compounds that have, among other things, the right binding conformation. The authors attempted to compare the hit rate from docking to that from random screening by randomly selecting 21 compounds from the same ACD dataset used in docking. Their random screen yielded no compounds with < 100 μM IC50s (although 9 and 5 compounds were 500 and 100 μM, respectively).

A study by Bissantz [130] compared different docking methods (DOCK, FlexX, Gold) with several scoring functions (Chemscore, DOCK, FlexX, Gold, Fresno, PMF, Score) to determine the optimal combination for finding known hits in a database for estrogen receptor (ER) and thymidine kinase (TK). The authors found the best combination of docking method-scoring function to be Gold-DOCK for ER and Gold-Gold for TK. The two databases for each target were created by appending 10 known active compounds each for TK and ER to a set of 990 randomly selected compounds from the ACD that were filtered to remove chemical reagents, inorganics, and molecular weights between 250 and 500. TK is considered by the authors to be a "hard case" for docking as the binding site is accessible to water, the side chains can adopt different rotamer states depending on the ligand, and participation of a water upon binding differs for purine and pyrimidine substructures. ER, on the other hand, is considered by the authors to be more suitable for docking because the binding site is less exposed to solvent than TK, is less prone to conformational changes upon binding, and is selective for compounds with low nM affinity. The best docking method-scoring function found in this study, Gold-DOCK, was able to identify 9 of 10 true hits among the top 2% of scorers. An observation by the authors was that the scoring functions that were more robust in selecting TK compounds, FlexX and PMF, were among the poorer discriminators of active ER compounds. They also noted that FlexX and PMF tended to perform well for the highly polar binding site of TK, but DOCK was better for apolar active site of ER. While they also found that consensus scoring [131] generally outperforms single scoring functions, they proposed a two step process of first identifying the optimal method-scoring function combination on a small representative set of known binders, and second, using this new combination for database screening.

A recently published example used ICM to find 14 novel antagonists of thyroid hormone nuclear receptor (TR) ranging from 1.5 to 30 μM [125]. The authors docked ~190,000 compounds from the ACD that passed a Lipinksi filter into a predicted model of the antagonist-bound conformation of the TR ligand–binding region. A selection of 1000 favorably docked structures that could interfere with the active state of the receptor were selected and further refined with energy minimization. A final 100 compounds were selected based on shape complementarity, hydrogen-bond network and flexibility for characterization by *in vitro* assays, which resulted in the 14 novel antagonists that appear structurally diverse in their paper. The authors extended their work in this paper by designing a focused virtual library of 101 compounds based on the top hit (Table **2**) from the initial database screen. This virtual library was screened with ICM against the same receptor as before, and resulted in several new compounds with increased antagonist activity over the initial compound.

Flexible Receptor Docking

There are a few papers describing recent methods for flexible receptor docking [19]. One approach extends the popular FlexX docking method to model the receptor as a "united protein" representation of structures [112]. The united protein approach takes a collection of receptor structures (resulting from, for example, crystallography, NMR, or molecular dynamics) and superimposes rigid regions into one feature, but treats flexible regions as discrete conformations from the collection. FlexE was designed to work in cases where only relatively localized motion occurs in the binding site, and was tested on 10 protein targets that included 105 crystal structures with mostly bound and a few unbound ligands. The authors showed that FlexE resulted in 67% of the ligands having an RMSD of < 2.0 Å from the known crystallographic solution (as compared to 63% for FlexX), and was faster than the accumulated time from individual docking with FlexX. The method is fully automated for allowing the user to identify the most favorable docked ligand-receptor configuration.

Another flexible receptor method, ICM [109] was used by Stigler and coworkers [132] to correctly predict the three-dimensional structure of a complex between the monoclonal antibody (mAb) TE33 and its cholera-toxin-derived peptide epitope VPGSQHID [132]. The coordinates of TE33 used for docking came from a co-crystal of TE33 and another peptide. The binding sites of mAbs are typically shallow in shape with a significant amount of the bound ligand exposed to solvent, which can present a challenge for docking. The authors allowed full flexibility during docking of dihedral angles for backbone and side chains in the VPGSQHID peptide, and side chains within 6Å of the bound ligand of the initial co-crystal. Starting with a fully extended peptide, ICM was able to produce a docked conformation that had a backbone root-mean-square deviation of 1.9 Å to the crystal structure. The ICM scoring function contains more terms than are typically found in docking methods such as conformational entropy, surface-based solvation, and electrostatics desolvation via Poisson-Boltzmann solutions. The increased complexity of the ICM scoring function may be more beneficial for selecting actives from non-actives from docking of databases.

Using Experimental Data to Enhance Docking

Exploiting experimental information during docking is a useful approach for limiting docking solutions to those with prerequisite attributes (e.g. binding mode or pharmacophore). Relying only on docking for a lead is challenging because methods do not systematically explore all possible degrees of freedom and the scoring functions are imperfect. Information that becomes available during drug-design cycles, however, can be exploited by docking for more plausible solutions. For example, Last-Barney *et al*., [133] used information from photoaffinity labeling of the lymphocyte-function associated antigen-1 (LFA-1) integrin to restrict the docking site to that centered on the residue shown to be important in binding. In this paper, a combination of docking with AutoDock, molecular dynamics, and minimization led to a predicted binding mode that was later confirmed by crystallography.

One approach that we will review in more detail for exploiting experimental information during VS is by Gruneberg and coworkers, who described an approach for identifying nanomolar IC_{50} inhibitors of human carbonic anhydrase II (hCAII) [124]. The hCAII enzyme is characterized by a deep conical binding site containing a

coordinated Zn at the catalytic site. Their approach was the following series of steps. First, filter a combined database from Maybridge and LeadQuest using Lipinski rules (98, 850 compounds). Second, filter by functional groups known to be important for Zn binding in other Zn proteases (5904 compounds). Third, determine pharmacophore sites in the receptor by taking the geometric centers of various interaction probes, and then filter by a flexible search for those compounds that fit the pharmacophore (3314 compounds). Fourth, rank the compounds by 3D similarity to two known tight binding reference ligands using FlexS [134] (top 100 compounds chosen). Fifth, dock the remaining 100 compounds into hCAII using FlexX. A final set of 13 compounds, which were mostly from LeadQuest, were selected based on visual inspections and tested experimentally for IC_{50} inhibition of hCAII; three compounds were sub-nM, one was nM, and seven were μM. The authors noted that most of the hits in their final set came from LeadQuest. While the preponderance of hit ligands contained the known hCAII Zn-binding sulfonamide moiety, the authors noted that side chains of this moiety can still have a significant impact on binding potency with hCAII, and that several of their leads were not previously described in the patent literature. The predicted binding modes of the ligands were in good agreement with those of the crystallographic structures.

Scoring Functions

Docking is essentially made up of two components. First, the algorithm used by the program for generating trial binding modes, or poses, for each ligand. Second, the scoring function used during docking for ranking these poses. A post-processing step can occur where one or more scoring functions are used to re-rank the poses, and select the best pose among the ligands from a database. Results from studies suggest that the search algorithms used in docking will probably produce the true, or nearly true, binding mode among the poses that are generated [19, 135-137]. It is the *ranking* of these compounds with respect to binding constants by scoring functions, however, that is typically where much of the problem with accuracy occurs [136, 137]. Scoring functions in docking are a trade-off between the fast evaluation of ligand-receptor interactions needed for database screening and accuracy. The limited accuracy is due in large part to an incomplete understanding of complex ligand-receptor interactions — and what is understood can be computationally expensive — to an incomplete training set for parameterization, and to a highly simplified approximation (or omission) of important variables in binding such as solvation.

There are three main classes of scoring functions. First, molecular mechanics functions, such as AMBER [128], OPLS [138], CHARMM [115, 139] and MMFF [140] are the closest to first principles that are used in docking. These functions are based on additive atomic parameters (van der Waals, electrostatics, bonds, angles and torsions), usually derived from quantum mechanical calculations, which are assumed to be transferable between molecules containing similar atom types. Second, empirical functions that are calibrated to associate numbers of atomic or molecular features (e.g. rotatable bonds, hydrogen-bond donors and acceptors) with known binding data from training sets. Third, knowledge-based functions, [141] which apply a Boltzmann weighting to distributions of atom-atom distances observed in crystal structures.

Each of the three classes of scoring functions has their strengths and weaknesses in docking, [130, 136] and a few relatively recent developments have attempted to address the problems. For example, the popular consensus scoring approach attempts to make up for deficiencies in any one scoring function by using several different scoring functions (e.g. up to five) at a time [131]. The method "ranks" a compound by counting the number of individual scoring functions that scored the compound favorably relative to a threshold. For example, a compound that scored well in three or more scoring functions is presumed to be "better" than one that scored well in only two or less. Some of the newer developments in scoring functions are to customize the scoring functions for the system under study [13]. These functions can be based either on classification or regression algorithms, or be "hybrids" of others classes of scoring functions that are parameterized by statistically fitting the terms to available binding data (e.g. a binary output such as active non-active, or a continuous value such as free energy) [142]. For example, Stahl built an optimized combination of FlexX and PLP in an attempt to balance the localized and hydrogen bond features of FlexX and the lipophilic features of PLP [136]. They found that this new combination resulted in better enrichment of known hits among the top scoring compounds than the individual scoring functions for five out the seven target protein cases studied [136]. Another example is the use of binary kernel discrimination (BKD) for classifying actives and non-actives in virtual screening [143]. In this paper, the authors studied several combinations of classification scoring functions and molecular fingerprints and found that the BKD algorithm with UNITY [120] fingerprints best classified actives and non-actives for the NCI [144] and Syngenta [143] data sets.

Finally, we mention that more accurate methods of ranking compounds exist such as binding free energy perturbation methods [145]. Poisson-Boltzmann solvation terms with molecular mechanics (MM/PBSA), [146] and the less theoretically rigorous linear response free energy methods [147]. These methods, however, are not yet common for high throughput docking (but can be considered for lower throughput) currently due to the high computational cost, the unvalidated atomic parameters for the types of diverse ligands found in drug databases, and the increased complexity in setting up and using some of these methods.

Perhaps the most important emerging trend in the recent literature involving scoring functions is that they are clearly target specific, and dependent on which physical factors are important in binding for a given ligand-receptor interaction. For example, Perez found molecular mechanics (AMBER) to outperform PMF for 34 examples when a ligand from a co-crystal is docked into its co-crystalline receptor, but similar performance between these scoring functions when docking close analogues to this ligand from a database into the same receptor [148]. Muegge *et al*., however, found PMF to outperform molecular mechanics (DOCK) when docking weak ligands into the FK506 binding protein [149]. The

discrepancy between these two studies probably reflects the sensitivity of molecular mechanics functions to steric clashes, which would favor cases where the ligand is docked into its native receptor and reject analogue compounds; PMF, a function that is more tolerant of these clashes, would favor analogue compounds docked into a non-native receptor. Stahl has reviewed FlexX, PLP, DrugScore and PMF on seven protein targets finding that some scoring functions are better suited for polar targets (FlexX) while others work for lipophillic targets (DrugScore, PLP) [136].

INFORMATICS AND COMPUTING INFRASTRUCTURE

The integration of VS into drug discovery depends on robust informatics infrastructure with the goal being to convey information among VS and experimental groups rapidly so that key decisions can be made in meaningful ways. For example, the informatics system should minimize any latency time between a chemist's virtual library design, the modelers's use of high performance computing for VS and analysis, and finally the reporting of the predicted hits and 3D patterns back to the chemist. There have been a few publications that have talked about this problem and described several types of solutions, [13, 150] and there are several companies heading in this direction with products [40, 113, 115, 120, 151].

At Biogen Idec, we have also recently implemented a solution in collaboration with Accelrys [115] of an integrated VS environment (Fig. **2a**) for increasing the quality of compounds in drug-design cycles (Fig. **2b**). The main points are the following. First, a chemist can use a desktop client and server architecture that is designed to handle millions of compounds easily. From the desktop, the user can design chemical libraries in flexible ways (e.g. *in silico* chemistry or database queries), review molecular properties, and filter compounds based on molecular structure or properties, and submit the library for VS. Second, the modeler can retrieve the submitted library from the database and perform the necessary tasks for VS using a 75-node/150 CPU cluster running Linux. Many custom "scripts" and data analyses are used by the modeler to minimize time spent on virtual screening, and to quickly determine rankings and ligand-receptor binding patterns, which are then submitted back to the central database. Third, the chemist can then review the VS results for their library using the same desktop interface as before to browse quickly through large number of docked compounds with 3D graphics and interact with high dimensional plotting and chemical spreadsheets.

High Performance Computing

The ability to routinely screen and filter massive virtual libraries on the order of 10^6 compounds with the VS methods described depends on the availability of high performance computer hardware and supporting software infrastructure. Screening large databases computationally can be addressed due to the inherently parallel nature of the VS problem. That is, computer processors can treat each molecule progressing through enumeration, filtering, conformer generation, and receptor docking independently.

Two hardware approaches have emerged to support distributed parallel computing. The first employs dedicated clusters consisting of a few to several hundred computational nodes typically running Linux [152, 153]. The second, termed Grid computing, [154, 155] exploits underutilized or idle workstations and PCs in an organization to provide computing power, potentially leveraging thousands of computers for large database problems. Software vendors are now providing functionality to take advantage of distributed parallel computing, including software for 3D conformer generation (for example Omega [113], Catalyst [115] and Corina [156]), and docking (FlexX, Fred, Glide, DOCK, ICM, and LigandFit), and filtering.

CONCLUSIONS

It remains to be seen what impact the new paradigm of integrated drug discovery will have on the production of new drugs. What is clear, though, is that highly publicized technologies of the past 15 years such as computational chemistry, combinatorial chemistry, and HTS have not by themselves increased the new number of compounds entering preclinical or Phase I development [6]. In fact the number of new chemical entities (NCE) have declined over this time despite skyrocketing investment in R&D [6]. The new pharmaceutical strategy places a greater emphasis on the quality of compounds entering experimental screening than existed before by way of more intelligent utilization of the confluence of available experimental and computational information. The current direction of VS methods and informatics approaches is to play a key role in processing the large number of possibilities and predicting which compounds are likely to succeed.

As more information about drug targets (including structures), and drug or compound profiles become increasingly available, it is likely that there will be more emphasis on lead optimization than lead discovery. That is, it may become routine to exploit existing data on a particular target, gene-family of targets, or class of compounds in predictive models or constraints in ways that rapidly progress to new biologically relevant compounds instead starting from scratch.

We have shown in this review that VS using docking methods are becoming more validated for identifying active compounds using a variety of docking algorithms, library preparation techniques, and experimental information as constraints. Published examples of the practical application of *in silico* ADME methods in a pharmaceutical setting, however, are rare despite the vast literature describing model development and approaches. Until validated application examples are available, it is difficult to gauge the actual cost savings and efficiencies from the use of predictive ADME in VS.

The availability of cluster and grid computing solutions will make it possible to consider broader VS applications involving sophisticated computational approaches that are currently intractable for VS. With the computing power available from a corporate-wide grid, it will be possible to expand the VS workflow to include receptor flexibility, [157, 158] entire protein target families, [159] anti-targets,

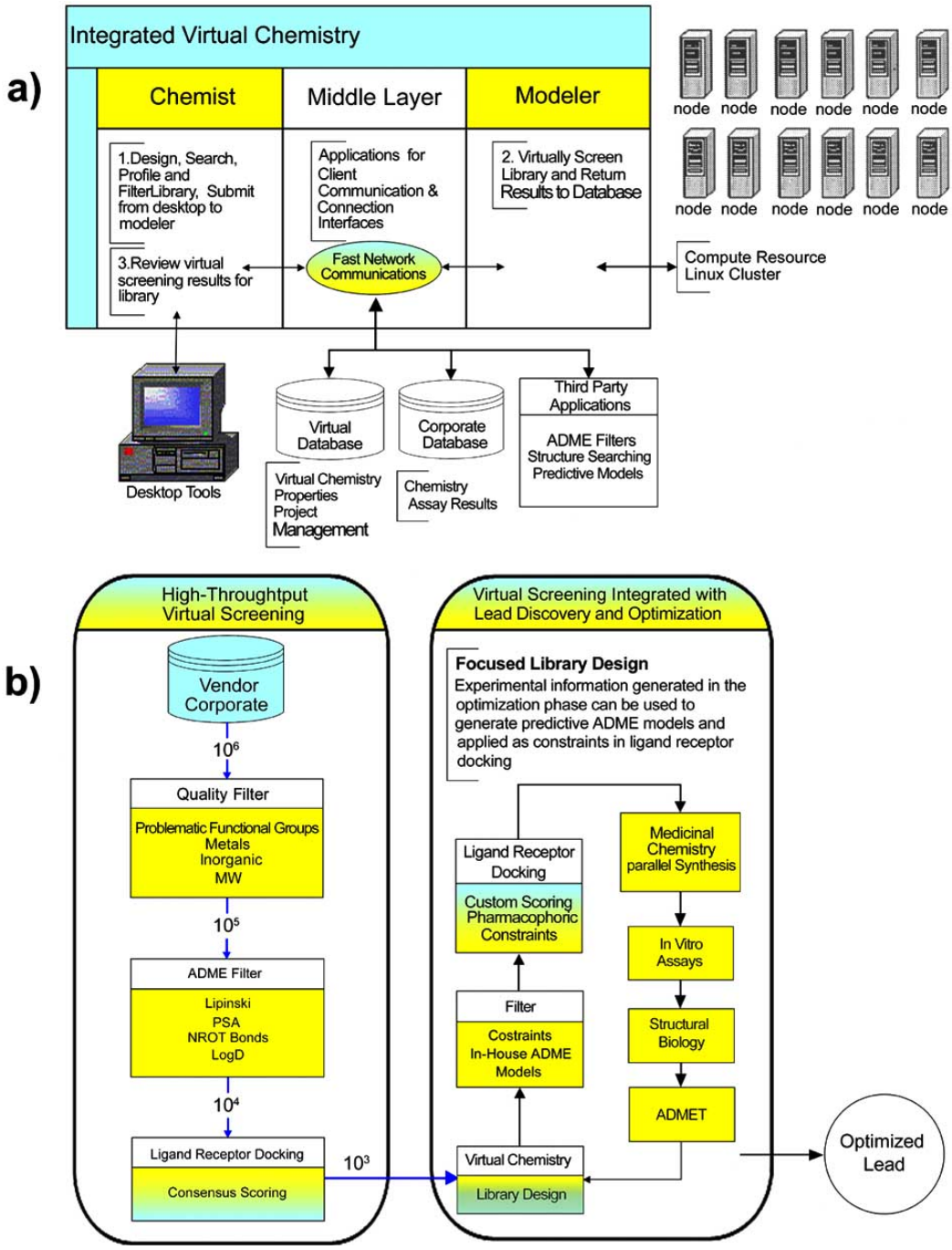


Fig. (2). a). A schematic for an integrated virtual chemistry and screening environment at Biogen Idec. The desktop tools are designed to allow users to easily and quickly design, filter, and review the results of virtual libraries containing up to millions of compounds. The interface also permits the review and sharing of chemistries among users. b). The workflow of virtual chemistry and screening at Biogen. The numbers of compounds are estimates to highlight the removal of compounds that do not pass a particular filter or virtual screen.

[142] and more accurate estimates of ligand binding, [160-166] or of possible sites for metabolism [167, 168].

Finally, a new definition for informatics in drug discovery in the future will likely merge areas of chemo-, bio-, and structural-informatics. Desktop systems that allow seamless integration and interrogation of virtual chemistry, predictive models, pattern analysis, VS results, and biological data will permit better decisions about which compounds should be made, and hopefully increase the chances of designing more successful drugs.

ACKNOWLEDGEMENTS

We are grateful to Herman Van Vlijmen, Alexey Lugovskoy, Paul Lyne, and Max Totrov for reading the manuscript.

REFERENCES

[1] Beresford, A. P.; Selick, H. E.; Tarbit, M. H. *Drug Discovery Today* **2002**, *7*, 109-116.

[2] Johnson, D. E.; Grushenka, H. I. W. *Drug Discovery Today* **2000**, *5*, 445-454.

[3] Bajorath, J. *Drug Discovery Today* **2001**, *6*, 989-995.

[4] Terstappen, G. C.; Reggiani, A. *Trends in Pharmacological Sciences* **2001**, *22*, 23-26.

[5] Drews, J. *Science* **2000**, *287*, 1960-1964.

[6] Handen, J. S. *Drug Discovery Today* **2002**, *7*, 83-85.

[7] Toledo-Sherman, L. M.; Chen, D. *Current Opinion in Drug Discovery & Development* **2002**, *5*, 414-421.

[8] Han, H. N. *Current Drug Discovery* **2001**, 25-28.

[9] Stahura, F. L.; Bajorath, J. *Drug Discovery Today* **2002**, *7*, S41-S47.

[10] Leach, A. R.; Hann, M. M. *Drug Discovery Today* **2000**, *5*, 326-336.

[11] Matter, H.; Baringhaus, K.-H.; Naumann, T.; Klabunde, T.; Pirard, B. *Combinatorial Chemistry and High Throughput Screening* **2001**, *4*, 453-475.

[12] Van de Waterbeemd, H. *Current Opinion in Drug Discovery & Development* **2002**, *5*, 33-43.

[13] Walters, W. P.; Stahl, M. T.; Murcko, M. A. *Drug Discovery Today* **1998**, *3*, 160-178.

[14] Oprea, T. I. *Molecules* **2002**, *7*, 51-62.

[15] Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. *Journal of Chemical Information and Computer Sciences* **2001**, *41*, 1308-1315.

[16] Merlot, C.; Domine, D.; Church, D. J. *Current Opinion in Drug Discovery and Development* **2002**, *5*, 391-399.

[17] Lyne, P. D. *Drug Discovery Today* **2002**, *7*, 1047-1054.

[18] Schneider, G.; Bohm, H.-J. *Drug Discovery Today* **2002**, *7*, 64-70.

[19] Abagyan, R.; Totrov, M. *Current Opinion in Chemical Biology* **2001**, *5*, 375-382.

[20] Langer, T.; Hoffmann, R. D. *Current Pharmaceutical Design* **2001**, *7*, 509-527.

[21] Butina, D.; Segall, M. D.; Frankcombe, K. *Drug Discovery Today* **2002**, *7*, S83-S88.

[22] Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. *Current Opinion in Chemical Biology* **2002**, *6*, 439-446.

[23] Kubinyi, H. *Current Opinion in Drug Discovery & Development* **1998**, *1*, 16-27.

[24] Good, A. *Current Opinion in Drug Discovery & Development* **2001**, *4*, 301-307.

[25] Gane, P. J.; Dean, P. M. *Current Opinion in Structural Biology* **2000**, *10*, 401-404.

[26] Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comp. Chem.* **2002**, *16*, 151-166.

[27] Good, A. C.; Krystek, S. R.; Mason, J. S. *Drug Discovery Today* **2000**, *5*, S61-S69.

[28] Eddershaw, P. J.; Beresford, A. P.; Bayliss, M. K. *Drug Discovery Today* **2000**, *5*, 409-414.

[29] Ekins, S.; Boulanger, B.; Swann, P. W.; Hupcey, M. A. Z. *Journal of Computer Aided-Molecular Design* **2002**, *16*, 381-401.

[30] van der Waterbeemd, H.; Smith, D. A.; Beaumont, K.; Walker, D. K. *J. Med. Chem.* **2001**, *44*, 1313-1333.

[31] Waszkowycz, B. *Current Opinion in Drug Discovery & Development* **2002**, *5*, 407-413.

[32] Group, B. C. "A Revolution in R and D, " Boston Consulting Group, 2001.

[33] Stevens, R.; Yokoyama, S.; Wilson, I. A. *Science* **2001**, *294*, 89-92.

[34] Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang J.; Gabor, Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. *Science* **2001**, *291*, 1304-1351.

[35] Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D. R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Myers, R. M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Schultz, R.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W.

R.; de la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D. G.; Burge, C. B.; Cerutti, L.; Chen, H. C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; Furey, T. S.; Galagan, J.; Gilbert, J. G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L. S.; Jones, T. A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A. F.; Stupka, E.; Szustakowski, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Yang, S. P.; Yeh, R. F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Patrinos, A.; Morgan, M. J.; Szustakowki, J.; de, Jong, P.; Catanese, J. J.; Osoegawa, K.; Shizuya, H.; Choi, S.; Chen, Y. J.; International, Human, Genome, Sequencing Consortium. *Nature* **2001**, *409*, 860-921.

[36] Bailey, D.; Zanders, E.; Dean, P. *Nature Biotechnology* **2001**, *19*, 207-209.

[37] Hopkins, A. L.; Groom, C. R. *Nature Reviews Drug Discovery* **2002**, *1*, 727-730.

[38] Chan, E.; Weir, M. *Chemical Innovation* **2001**, *31*.

[39] Schuffenhauer, A.; Zimmermann, J.; Stoop, R.; van der Vyver, J. J.; Lecchini, S.; Jacoby, E. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 947-955.

[40] www.mdli.com.

[41] www.cambridgesoft.com.

[42] Barnard, J. M.; Downs, G. M.; von Scholley-Pfab, A.; Brown, R. D. *Journal of Molecular Graphics and Modeling* **2000**, *18*, 452-463.

[43] Lobanov, V. S.; Agrafiotis, D. K. *Journal of Molecular Graphics & Modelling* **2001**, *19*, 571-578.

[44] Cruciani, G.; Pastor, M.; Guba, W. *Eur. J. Pharm. Sci.* **2000**, *11*, S29-39.

[45] de Groot, M. J.; Ekins, S. *Adv. Drug Deliv. Rev.* **2002**, *54*, 367-383.

[46] Filipponi, E.; Cruciani, G.; Tabarrini, O.; Cecchetti, V.; Fravolini, A. *J. Comput. Aided Mol. Des.* **2001**, *15*, 203-217.

[47] Crivori, P.; Cruciani, G.; Carrupt, P. A.; Testa, B. *J. Med. Chem.* **2000**, *43*, 2204-2216.

[48] Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. *Journal of Pharmacology and Experimental Techniques* **1999**, *290*, 429-438.

[49] Ekins, S.; de Groot, M. J.; Jones, J. P. *Drug Metab. Dispos.* **2001**, *29*, 936-944.

[50] Walters, W. P.; Murcko, M. A. *Adv. Drug Deliv. Rev.* **2002**, *54*, 255-271.

[51] Clark, D. E.; Pickett, S. D. *Drug Discovery Today* **2000**, *5*, 49-58.

[52] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Advanced Drug Delivery Reviews* **1997**, *23*, 3-25.

[53] Wenlock, M. C.; Austin, R. P.; Barton, P.; Dacis, A. M.; Leeson, P. D. *J. Med. Chem.* **2003**, *46*, 1250-1256.

[54] Wagener, M.; van Geerestein, V. J. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 280-292.

[55] Muegge, I.; Heald, S. L.; Brittelli, D. *J. Med. Chem.* **2001**, *44*, 1-6.

[56] Oprea, T. *Journal of Computer Aided-Molecular Design* **2000**, *14*, 251-264.

[57] Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. *Agnew. Chem., Int. Ed. Engl.* **1999**, *38*, 3743-3748.

[58] MACCS-II Drug Data Report (MDDR). Molecular Design Limited, San Leandro, CA.

[59] World Drug Index (WDI). Derwent Information, London, UK.

[60] Comprehensive Medicinal Chemistry (CMC). Molecular Design Limited, San Leandro, CA.

[61] Sadowski, J.; Kubinyi, H. *J. Med. Chem.* **1998**, *41*, 3325-3329.

[62] Ajay; Walters, W. P.; Murko, M. A. *J. Med. Chem.* **1998**, *41*, 3314-3324.

[63] Walters, W. P.; Ajay; Murcko, M. A. *Curr. Opin. Chem. Biol.* **1999**, *3*, 384-387.

[64] Frimurer, T. M.; Bywater, R.; Naerum, L.; Lauritsen, L. N.; Brunak, S. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1315-1324.

[65] Egan, W. J.; Lauri, G. *Adv. Drug Deliv. Rev.* **2002**, *54*, 273-289.

[66] Stenberg, P.; Bergstrom, C. A.; Luthman, K.; Artursson, P. *Clin. Pharmacokinet.* **2002**, *41*, 877-899.

[67] Clark, D. E. *Comb. Chem. High Throughput. Screen.* **2001**, *4*, 477-496.

[68] Blake, J. F. *Current Opinion in Biotechnology* **2000**, *11*, 104-207.

[69] Stenberg, P.; Luthman, K.; Artursson, P. *J. Control Release* **2000**, *65*, 231-243.

[70] Burton, P. S.; Goodwin, J. T.; Vidmar, T. J.; Amore, B. M. *The Journal of Pharmacology and Experimental Therapeutics* **2002**, *303*, 889-895.

[71] Jorgensen, W. L.; Duffy, E. M. *Adv. Drug Deliv. Rev.* **2002**, *54*, 355-366.

[72] Engkvist, O.; Wrede, P. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1247-1249.

[73] Physical/Chemical Property Database (PHYSPROP). Syracuse Research Corporation, Syracuse, NY.

[74] Liu, R.; So, S.-S. *J. Chem. Inf. Comp. Sci.* **2001**, *41*, 1633-1639.

[75] Yalkowsky, S. H.; Dannelfelser, R. M. In; College of Pharmacy, University of Arizona: Tuscon, AZ, 1990.

[76] Palm, K.; Luthman, K.; Ungell, A. L.; Strandlund, G.; Artursson, P. *J. Pharm. Sci.* **1996**, *85*, 32-39.

[77] Palm, K.; Luthman, K.; Ungell, A. L.; Strandlund, G.; Beigi, F.; Lundahl, P.; Artursson, P. *J. Med. Chem.* **1998**, *41*, 5382-5392.

[78] Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. *Pharm. Res.* **1997**, *14*, 568-571.

[79] Bergstrom, C. A. S.; Strafford, M.; Lazorova, L.; Avdeef, A.; Luthman, K.; Artusson, P. *J. Med. Chem.* **2003**, *46*, 558-570.

[80] Stenberg, P.; Norinder, U.; Luthman, K.; Arturrson, P. *J. Med. Chem.* **2001**, *44*, 1927-1937.

[81] Winiwarter, S.; Bonham, N. M.; Ax, F.; Hallberg, A.; Lennernais, H.; Larlen, A. *J. Med. Chem.* **1998**, *41*, 4939-4949.

[82] Egan, W. J.; Merz, K. M. J.; Baldwin, J. *J. Med. Chem.* **2000**, *43*, 3867-3877.

[83] Norinder, U.; Osterberg, T. *J. Pharm. Sci.* **2001**, *90*, 1076-1085.

[84] Clark, D. E. *J. Pharm. Sci.* **1999**, *88*, 807-814.

[85] Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, *43*, 3714-3717.

[86] Osterberg, T.; Norinder, U. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1408-1411.

[87] Ekins, S.; Durst, G. L.; Stratford, R. E.; Thorner, D. A.; Lewis, R.; Loncharich, R. J.; Wikel, J. H. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1578-1586.

[88] Yamashita, F.; Wanchana, S.; Hashida, M. *J. Pharm. Sci.* **2002**, *91*, 2230-2239.

[89] Abraham, M. H.; Ibrahim, A.; Zissimos, A. M.; Zhao, Y. H.; Comer, J.; Reynolds, D. P. *Drug Discov. Today* **2002**, *7*, 1056-1063.

[90] Platts, J. A.; Abraham, M. H.; Hersey, A.; Butina, D. *Pharmaceutical Research* **2000**, *17*, 1013-1018.

[91] Zhao, Y. H.; Le, J.; Abraham, M. H.; Hersey, A.; Eddershaw, P. J.; Luscombe, C. N.; Boutina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J. A. *J. Pharm. Sci.* **2001**, *90*, 749-784.

[92] Wessel, M. D.; Mente, S. *Annual Reports in Medicinal Chemistry* **2001**, *36*, 257-266.

[93] Niwa, T. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 113-119.

[94] Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615-2623.

[95] Yoshida, F.; Topliss, J. G. *J. Med. Chem.* **2000**, *43*, 2575-2585.

[96] Andrews, C. W.; Bennett, L.; Yu, L. X. *Pharmaceutical Research* **2000**, *17*.

[97] George, P. *J. Neurochem.* **2003**, *85*, 7.

[98] Norinder, U.; Haeberlein, M. *Adv. Drug Deliv. Rev.* **2002**, *54*, 291-313.

[99] Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A.; Wikel, J. H. *Journal of Pharmacological and Toxicological Methods* **2001**, *44*, 251-272.

[100] Clark, D. E. *J. Pharm. Sci.* **1999**, *88*, 815-821.

[101] Luco, J. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 396-404.

[102] Engkvist, O.; Wrede, P.; Rester, U. *J. Chem. Inf. Comp. Sci.* **2003**, *43*, 155-160.

[103] Ajay *J. Med. Chem.* **1999**, *42*, 4942-4951.

[104] We define a constraint as a term that is added to the main function being optimized by the docking algorithm. The purpose of the constaint, for example, is to force a particular event to occur as part of the ligand-receptor interaction.

[105] We define a filter as using a criteria to eliminate compounds before or after the compound has been optimized by the docking algorithm.

[106] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269-288.
[107] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comp. Chem.* **1998**, *19*, 1639-1662.
[108] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727-748.
[109] Abagyan, R.; Totrov, M.; Kuznetsov, D. *J. Comp. Chem.* **1994**, *15*, 488-506.
[110] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470-489.
[111] www.biosolveit.com.
[112] Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.* **2001**, *308*, 377-395.
[113] www.eyesopen.com.
[114] www.schrodinger.com.
[115] www.accelrys.com.
[116] Pang, Y.-P.; Perola, E.; Xu, K.; Prendergast, F. G. *J. Comp. Chem.* **2001**, *22*, 1750-1771.
[117] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemsitry* **1983**, *4*, 187-217.
[118] Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; DeBolt, S.; Ferguson, D. M.; Seibel, G. L.; Kollman, P. A. *Computer Physics Communications* **1995**, *91*, 1-41.
[119] Debnath, A. K.; Radigan, L.; Jiang, S. *J. Med. Chem.* **1999**, *42*, 3203-3209.
[120] www.tripos.com.
[121] Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. *J. Med. Chem.* **2002**, *45*, 2213-2221.
[122] Filikov, A. V.; Mohan, V.; Vickers, T. A.; Griffey, R. H.; Cook, P. D.; Abagyan, R. A.; James, T. L. *Journal of Computer-Aided Molecular Design* **2000**, *14*, 593-610.
[123] Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y.-P. *J. Med. Chem.* **2000**, *43*, 401-408.
[124] Gruneberg, S.; Stubbs, M. T.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 3588-3602.
[125] Schapira, M.; Raaka, B. M.; Das, S.; Fan, L.; Totrov, M.; Zhou, Z.; Wilson, S. R.; Abagyan, R.; Samuels, H. H. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 7354-7359.
[126] Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Prot. Struc. Func. Gen.* **1999**, *34*, 4-16.
[127] Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comp. Chem.* **1992**, *13*, 505-524.
[128] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, J. K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
[129] Hawkins, G. D.; Cramer, C. J.; Truhar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824-19839.
[130] Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759-4767.
[131] Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. *J. Med. Chem.* **1999**, *42*, 5100-5109.
[132] Stigler, R.-D.; Hoffmann, B.; Abagyan, R.; Schneider-Mergener, J. *Structure* **1999**, *7*, 663-670.
[133] Last-Barney, K.; Davidson, W.; Cardoza, M.; Frye, L. L.; Grygon, C. A.; Hopkins, J. L.; Jeanfavre, D. D.; Pav, S.; Qian, C.; Stevenson, J. M.; Tong, L.; Zindell, R.; Kelly, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 5643-5650.
[134] Lemmen, C.; Lengauer, T.; Klebe, G. *J. Med. Chem.* **1998**, *41*, 4502-4520.
[135] Kramer, B.; Rarey, M.; Lengauer, T. *Prot. Struc. Func. Gen.* **1999**, *37*, 228-241.
[136] Stahl, M.; Rarey, M. *J. Med. Chem.* **2001**, *44*, 1035-1042.
[137] Wang, R.; Lu, Y.; Wang, S. *J. Med. Chem.* **2003**, *46*, 2287-2303.
[138] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
[139] MacKerell, A. D., Jr.; Bashford, D.; Bellott, R. L.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B.* **1998**, 3586-3616.
[140] Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730-748.
[141] Gohlke, H.; Klebe, G. *Current Opinion in Structural Biology* **2001**, *11*, 231-235.
[142] Keseru, G. M. *Journal of Computer-Aided Molecular Design* **2001**, *15*, 649-657.
[143] Wilton, D.; Willett, P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 469-474.
[144] http:\\Cancer.gov.
[145] Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395-2417.
[146] Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. *J Med Chem* **2002**, *45*, 1412-1419.
[147] Pierce, A. C.; Jorgensen, W. L. *J. Med. Chem.* **2001**, *44*, 1043-1050.
[148] Perez, C.; Ortiz, A. R. *J. Med. Chem.* **2001**, *44*, 3768-3785.
[149] Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. *J. Med. Chem.* **1999**, *42*, 2498-2503.
[150] Leach, A. R.; Bradshaw, J.; Green, D. V. S.; Hann, M. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1161-1172.
[151] www.cambios.com.
[152] Kinghorn, D. B. *J. Mol. Graph Model.* **1999**, *17*, 148-150.
[153] Bitner, M.; Skelton, G. *Biomed. Sci. Instrum.* **2003**, *39*, 42-47.
[154] Richards, W. G. *Nat. Rev. Drug Discov.* **2002**, *1*, 551-555.
[155] Chien, A.; Foster, I.; Goddette, D. *Drug Discov. Today* **2002**, *7*, S176-180.
[156] www.mol-net.de.
[157] Hodoscek, M.; Borstnik, U.; Janezic, D. *Cell Mol. Biol. Lett.* **2002**, *7*, 118-119.
[158] Fersht, A. R. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14122-14125.
[159] Lamb, M. L.; Burdick, K. W.; Toba, S.; Young, M. M.; Skillman, A. G.; Zou, X.; Arnold, J. R.; Kuntz, I. D. *Prot. Struc. Func. Gen.* **2001**, *42*, 296-318.
[160] Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *Biopolymers* **2003**, *68*, 47-62.
[161] Hou, T.; Zhu, L.; Chen, L.; Xu, X. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 273-287.
[162] Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. *J. Med. Chem.* **2002**, *45*, 1412-1419.
[163] Wang, W.; Kollman, P. A. *J. Mol. Biol.* **2000**, *303*, 567-582.
[164] Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, *44*, 3417-3423.
[165] McCarrick, M. A.; Kollman, P. A. *J. Comput. Aided Mol. Des.* **1999**, *13*, 109-121.
[166] Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, *44*, 502-511.
[167] Singh, S. B.; Shen, L. Q.; Walker, M. J.; Sheridan, R. P. *J. Med. Chem.* **2003**, *46*, 1330-1336.
[168] Lamarche, O.; Platts, J. A. *Chemistry* **2002**, *8*, 457-466.