ELSEVIER

Towards improving compound selection in structure-based virtual screening

Bohdan Waszkowycz

Argenta Discovery Ltd., 8/9 Spire Green Centre, Flex Meadow, Harlow, Essex CM19 5TR, UK

Structure-based virtual screening is now an established technology for supporting hit finding and lead optimisation in drug discovery. Recent validation studies have highlighted the poor performance of currently used scoring functions in estimating binding affinity and hence in ranking large datasets of docked ligands. Progress in the analysis of large datasets can be made through the use of appropriate data mining techniques and the derivation of a broader range of descriptors relevant to receptor–ligand binding. In addition, simple scoring functions can be supplemented by simulation-based scoring protocols. Developments in workflow design allow the automation of repetitive tasks, and also encourage the routine use of simulation-based methods and the rapid prototyping of novel modelling and analysis procedures.

Introduction

Virtual screening is now well established within drug discovery as an effective paradigm for prioritising compounds for high-throughput or fragment-based physical screening, as well as for guiding compound progression through lead optimisation [1,2]. Many diverse *in silico* techniques, incorporating a broad range of 2D and 3D descriptors of chemical structure and predicted physicochemical properties, have been found effective for the identification of compounds that are similar to known ligands for a particular receptor or complementary to the structure of the binding site [3,4].

Structure-based virtual screening is most commonly implemented as the prediction of binding modes and binding affinities of each compound in the dataset by means of high-throughput docking to an X-ray structure or model of the target receptor [5,6]. The popularity of the technique is because of its potential for identifying novel chemotypes outside the scope of known ligands, as well as the mechanistic insight that it affords into the likely binding modes of active compounds. Although molecular docking remains a computationally demanding method compared to molecular similarity techniques, its widespread uptake has been boosted by the affordability of modern PC workstations and clusters, as well as the recent trend among software vendors to develop products to support the automation of large-scale virtual screening protocols. In this article we highlight some of the technical challenges of high-throughput docking and scoring, and discuss recent progress in analysis protocols and the increasing popularity of workflow design.

Corresponding author: Waszkowycz, B. (bohdan.waszkowycz@argentadiscovery.com)

Success and reliability of available virtual screening protocols

Docking programs were originally evaluated primarily in terms of their accuracy in reproducing the crystallographic binding modes of known ligands. More recently the focus has shifted towards assessing their effectiveness in simulated virtual screening experiments – that is, the recovery of known ligands from a large dataset of decoy compounds. Despite a plethora of published studies, there is increasing recognition of the technical difficulties in conducting an objective and statistically robust assessment of docking protocols and, as a result, insufficient progress has been made towards fully benchmarking their general applicability and reliability [7].

Some recent validation studies, most notably those by Warren *et al.* [8] and Chen *et al.* [9a], have focussed on a more thorough evaluation of docking programs against a wide range of targets, with due attention paid to crucial factors such as the size and diversity of the ligand dataset, the representation of the receptor and the fine-tuning of docking procedures. (The technical difficulties of implementing a robust validation study are highlighted by the recent critique of Chen *et al.*'s study by Perola *et al.* [9b] and

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2007.12.002

the subsequent revision of some of the original findings [9c].) These studies raise a number of issues concerning the performance of available software. Of particular note is the wide variability in the accuracy of docking and scoring – either within a single docking package when applied across a set of targets, or between several packages applied to a single target. Although predicted binding modes may be considered generally reliable, current scoring functions are found to have very limited accuracy in ranking compounds from diverse series or indeed even from homologous series. As a result, virtual screening of large datasets typically gives rise to a large number of false positives (high-scoring decoy compounds), with known ligands buried well down the rankings.

For the molecular modeller, this raises practical problems as to how to select the most appropriate combination of docking protocol and scoring function for a particular project without having to perform extensive validation studies, which may not be feasible for a target with few known ligands. While a consensus approach, combining the results of several docking and scoring protocols, has often proven successful in practice, the large variability in the performance of different methods suggests that this approach is itself not without problems and also requires validation [10].

In the light of the technical difficulties outlined above, several authors have found docking methods to be no more effective, or indeed less effective, than ligand-based virtual screening methods based on 3D shape matching or 2D descriptors [11–13]. Even when used optimally, it should not be expected that docking methods can retrieve all possible solutions, and therefore complementary technologies should be used wherever possible in order to broaden the selection of compounds for screening [4]. Given all the above caveats, it is encouraging to see many success stories of structure-based virtual screening both from industrial and academic groups, as well as the application of docking methods to particularly challenging problems such as virtual screening of GPCR targets and the assessment of CYP-mediated metabolism. Table 1 summarises a selection of recent papers to exemplify the range of applications in drug discovery.

The problem of achieving accurate scoring

As they are designed to be rapidly evaluated and applicable across heterogeneous receptors and ligands, scoring functions rely on relatively simple descriptors of intermolecular interactions, for example, hydrogen bonding and hydrophobic interactions [14,15]. These energy functions need to be fairly 'soft' so that ligands are not heavily penalised for small errors in the binding geometry. Other significant contributors to binding affinity are poorly described, if at all: these include ligand strain, desolvation, polarisation, pi-electron-mediated interactions and conformational entropy [16–18]. A significant error in scoring can result from discrepancies in the representation of either the receptor or the ligand – for example, the incorrect choice of protonation or tautomerisation states.

A further source of error is in the lack of a proper treatment of protein flexibility and solvation. Predicted ligand binding modes are usually highly sensitive to subtle changes of sidechain and backbone conformation, as well as interactions with bound water molecules [19,20]. These effects are not adequately handled by current docking programs, which treat the receptor as rigid or with limited sidechain flexibility and, as a result, known ligands may be mis-docked or scored more poorly than expected. Recent developments to overcome these problems include a model of water displacement and rotation in GOLD [21], docking to ensemble structures [22,23] and the induced-fit docking protocol in Glide [24]. Even so, these methods are not always appropriate for large-scale virtual screening, and the energetic cost of receptor deformation remains extremely difficult to estimate.

On a positive note, there have been continuous improvements over recent years to many of the mature scoring functions, examples being the re-parameterisation of PMF [25], DrugScore [26] and Glide XP [27,28]. The development of Glide XP highlights the benefit of introducing more novel terms into the scoring function,

TABLE 1

Recent applications of structure-based virtual screening in drug discovery

Method	Refs	Comments
Docking of ~200,000 compounds using FlexX-Pharm, selected by 3D pharmacophore searching	Lyne *et al.* [54]	Identification of four novel classes of inhibitor for Chk1 kinase
Docking of ~700,000 druglike compounds using rDock	Barril *et al.* [55]	Discovery of a new class of Hsp90 inhibitor with low to submicromolar potency
Docking of ~700,000 druglike compounds using rDock	Foloppe *et al.* [56]	Identification of 10 diverse inhibitors of Chk1 kinase. Binding modes subsequently confirmed by X-ray crystallography
Docking of 10,000 fragment-sized compounds using FlexX-Pharm	Rummey *et al.* [57]	Identification of novel S1-binding fragments for DPP-IV
Docking of ~315,000 compounds using ICM	Cavasotto *et al.* [58]	Identification of novel lead series of EGFR kinase inhibitor
Docking of ~44,000 compounds using GOLD and Surflex, selected by 2D pharmacophore searching	Kellenberger *et al.* [59]	Identification of novel non-peptide ligands for the GPCR CCR5, using a homology model based on bovine rhodopsin
Docking of virtual libraries using FlexX	Price *et al.* [60]	Example of docking focused virtual libraries to support hit finding and lead optimisation of HDAC inhibitors
Docking with AutoDock, FlexX and GOLD and multiple scoring functions	de Graaf *et al.* [61]	Example of a docking procedure for metabolic site prediction of CYP2D6 substrates

including desolvation, hydrophobic packing, aromatic interactions and weighting of specific binding motifs. New scoring functions have appeared, including the Astex Statistical Potential [29] which also emphasises the value of deriving receptor-specific scoring functions. Other researchers have reported the benefits of including polarisation of the ligand's partial atomic charges by the receptor using a quantum mechanics approach [30,31].

A number of groups have sought to improve the accuracy of consensus scoring schemes, recent examples including the systematic evaluation of multiple consensus strategies [32] and the application of machine learning methods to assist rapid elimination of poor binders [33] or supervised learning of correctly bound poses [34].

An additional positive development is the improved design of datasets for validation of virtual screening protocols and parameterisation of scoring functions, including initiatives to make such datasets more widely available. Of particular note are the Astex test set of diverse and well-resolved public domain X-ray complexes [35] and the DUD virtual screening dataset featuring some 3000 ligands for 40 different targets, with each ligand matched by 36 decoy molecules with similar physical properties [36]. Pham and Jain [37] have highlighted the importance of including negative data in training sets so that unfavourable features may be parameterised as accurately as possible in scoring functions.

The role of data-mining in analysing large datasets

In practice, it is generally accepted by modellers that it is important to visualise the docked poses of high-scoring compounds, rather than to make a selection based on score alone, as it is often readily apparent that many high-scoring ligands are docked in unrealistic conformations or miss interactions that are known to be important for the target receptor. This approach clearly becomes more difficult as the size of the dataset increases: with the availability of faster docking software and the relative affordability of PC clusters, it is now feasible to dock datasets of the order of 10^5 to 10^6 compounds (i.e. the size of a large pharma's corporate collection, or all commercially available druglike compounds) [38].

One solution for managing large datasets is to eliminate unpromising compounds before docking. Routinely, modellers will restrict the dataset to druglike or leadlike compounds by application of appropriate property and substructural filters, and perhaps perform diversity analysis to remove highly similar compounds [39]. A further option is to select compounds likely to possess the right features for binding in terms of similarity to known ligands or pharmacophores generated from the receptor structure [5]. These various approaches can be highly effective in reducing the dataset to be docked to the order of 10^3 to 10^4 compounds. These approaches are less suited when little is known about the likely preferences of the receptor, or where the objective is to identify compounds with novel chemistry and binding modes. In addition, similarity and pharmacophore searches can introduce undesired biases in the final dataset, as it is often difficult for the modeller to keep track of which classes of chemistry are being discarded in the process.

Is it at all feasible to analyse very large docked datasets in a meaningful way? As mentioned earlier, it is unwise to make a selection based on a single predicted binding score. It is more useful to apply diverse scoring functions and to examine the component terms of each individual scoring function – the balance of, say, hydrogen bonding and lipophilic terms may give some indication of the nature of the binding mode, or there may be excessive ligand–receptor clashes or internal ligand strain. These various component energies may therefore be applied as filters to include or exclude selections of ligands. Energy terms may be usefully normalised with regard to molecular size or functionality (particularly important as many studies aim to recover relatively low molecular weight hits that will tend to be ranked poorly compared to larger compounds) or averaged over multiple receptors [40,41]. For many receptors it is possible to identify key interaction sites which, even if not essential for activity, may at least offer a mechanism for classification of ligands in terms of their binding modes [42]. Indeed, the ability to cluster a dataset in terms of 2D ligand structure and predicted physicochemical properties, as well as binding scores and 3D binding modes, is a very powerful technique for mining high volume datasets [38]. It allows the modeller to eliminate rapidly clusters of little interest and to focus on more promising clusters, with the goal of making a final selection of compounds demonstrating diversity in both structure and binding interactions. A notable example of a technique for clustering by binding mode is the structural interaction fingerprint (SiFT) developed by Deng *et al.* [43].

One practical difficulty in performing this type of large-scale data mining is the limited availability of appropriate commercial software. The modeller requires spreadsheet functionality for the interactive browsing and graphing of descriptors, as well as the calculation of new descriptors and selection of subsets, with immediate visualisation of 3D docked binding modes of any ligand or cluster of ligands of interest [38]. Visualisation is essential in order to gauge the physical meaningfulness of any particular descriptor and to enable the modeller to derive new descriptors that capture features of interest. Although at first sight many of the established modelling packages appear to offer excellent visualisation and spreadsheet capabilities, there are often problems with handling very large datasets in real time, or the software may generate only a limited range of descriptors that are relevant to the characterisation of ligand–receptor interactions. The molecular modeller's demands on the software are likely to increase in the future as it becomes more commonplace for virtual screening studies to retain multiple poses per ligand or to compare docking modes across multiple receptor models, all of which further complicate the analysis stage.

While in the past, many industrial users have relied on in-house software development, it is encouraging to see that the commercial software vendors have begun to address these issues. GoldMine from CCDC (http://www.ccdc.cam.ac.uk) features a wide range of descriptors to characterise receptor–ligand binding contacts, spreadsheet functionality for data analysis and subset selection, and visualisation of multiple ligands/receptors. Similarly, the Docking Database from BioSolveIT GmbH (http://www.biosolveit.de/ddb) [44] manages very large datasets through an Oracle database and includes a range of utilities for the analysis and characterisation of virtual screening runs, for example, the generation of tailored scoring functions and the automated plotting of enrichment curves (Figure 1).

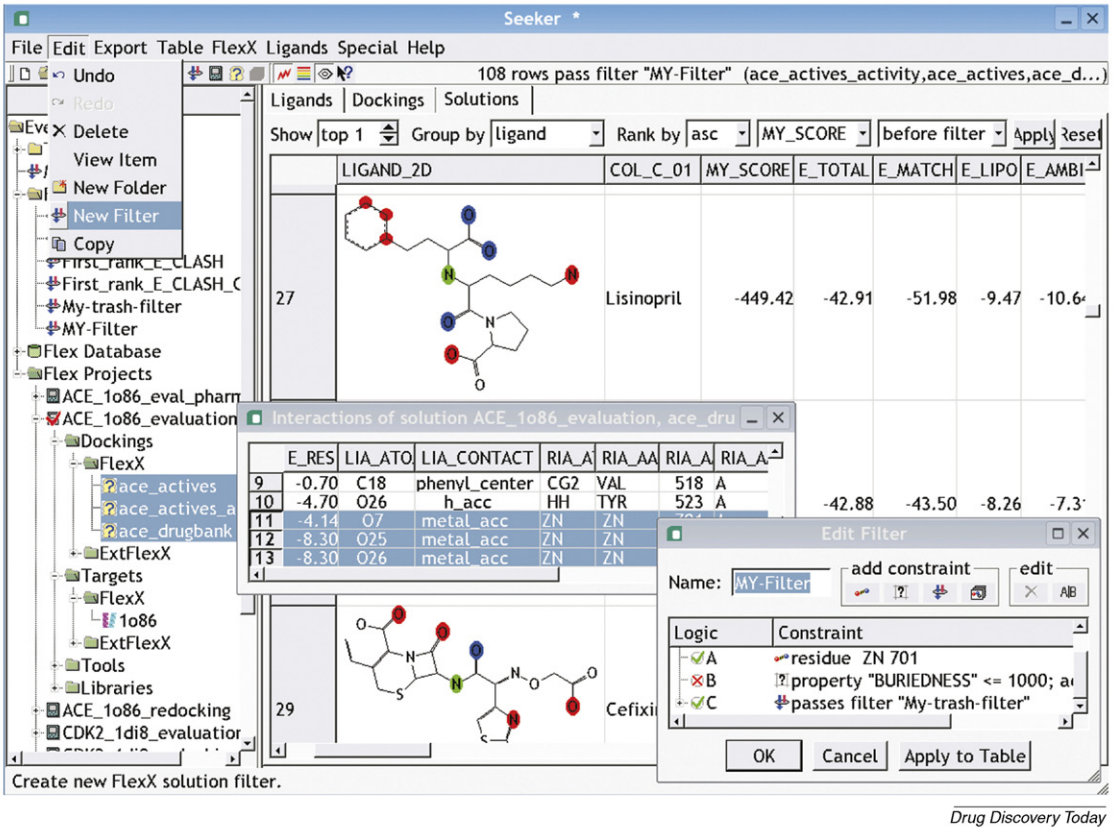

FIGURE 1

Graphical analysis of large virtual screening datasets. A screenshot of the Docking Database (DDB) graphical interface from BioSolveIT GmbH, which allows the integration of spreadsheet functionality for property analysis with the visualisation of 2D ligand structure and 3D ligand–receptor binding modes. In this view, atoms of the docked ligand that make favourable interactions with the receptor are highlighted on the 2D ligand structure. Image courtesy of BioSolveIT GmbH (http://www.biosolveit.de).

The challenge of accurate simulation-based scoring

Given the intrinsic limitations of simple scoring functions, it may be necessary to consider simulation-based (or physics-based) methods that yield a more physically accurate, if computationally expensive, estimate of binding affinity. Even if processing times are slower by several orders of magnitude, this is not necessarily a problem if the calculation is reserved as a final scoring stage to rank a modest number of the most promising compounds. This scenario is frequently encountered during lead optimisation where it is desired to estimate binding affinity as accurately as possible, typically between very similar ligands, for the prioritisation of synthetic candidates [45].

Rigorous statistical mechanics methods for the direct calculation of binding free energies, as typified by the free energy perturbation (FEP) and thermodynamics integration (TI) techniques, remain too computationally intensive for routine use in drug discovery, particularly in the industrial setting where a fast turnaround is important [46,47]. These methods require a thorough

TABLE 2

Recent examples of simulation-based scoring

Method	Refs	Comments
FEP	Jorgensen *et al.* [62]	Design and analysis of HIV reverse transcriptase inhibitors
LIE	Stjernschantz *et al.* [63]	Evaluation of LIE across four receptors
LIE	Su *et al.* [64]	Extensions to the original method applied to HIV reverse transcriptase inhibitors
LIE	Bortolato *et al.* [65]	Application to casein kinase 2 inhibitors
MM-PBSA	Kuhn *et al.* [51]	Extensive validation of MM-PBSA across eight targets
MM-PBSA, TI	Steinbrecher *et al.* [66]	Detailed analysis of human neutrophil elastase inhibitors
MM-PBSA	Brown and Muchmore [50]	Application to urokinase inhibitors
MM-GBSA	Lyne *et al.* [67]	Application to p38, Aurora A, CDK2 and Jnk3 kinases
MM-GBSA	Lee and Sun [49]	Use of MM-GBSA scoring to prioritise docking poses

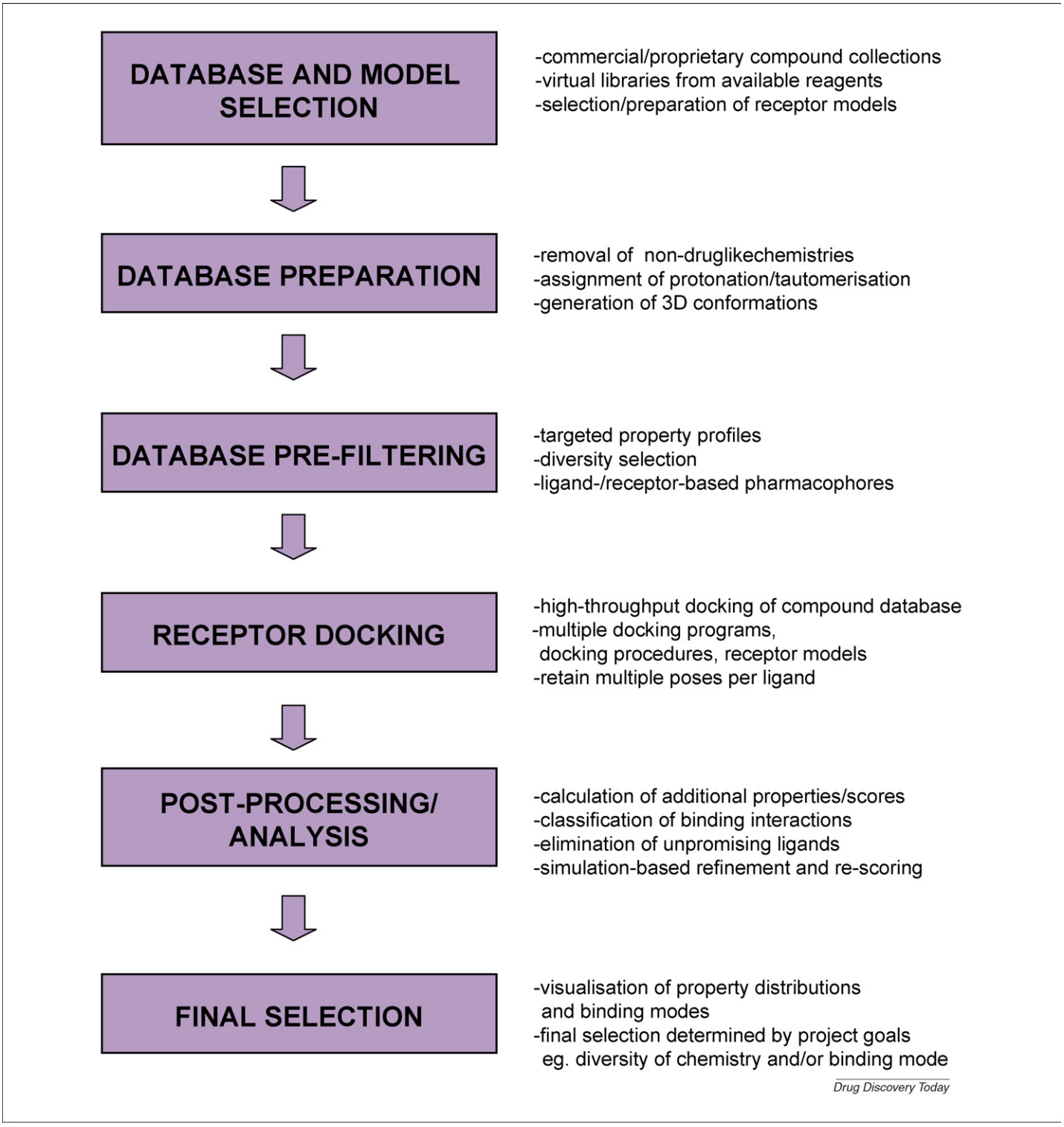

FIGURE 2

Typical workflow of tasks in structure-based virtual screening. Representative modelling and chemoinformatics tasks are listed, usually requiring integration of a wide range of commercial and/or proprietary software packages.

sampling of the conformational landscape of the receptor–ligand complex as one ligand is gradually mutated into another, typically implemented within an extensive molecular dynamics or Monte Carlo simulation with an explicit solvation model. Although theoretically robust, these methods are limited to assessing small structural changes and are notoriously prone to sampling and convergence errors.

More popular are methods based on the linear response approximation that are less computationally demanding and are more applicable to structurally diverse ligands. The original implementation of the linear interaction energy (LIE) method focussed on simulation of only the endpoints of the perturbation pathway, with the resulting interaction energies weighted by a regression procedure to reproduce experimental binding affinities [46,47].

The current trend is towards analysing a small number of snapshots, or even a single snapshot, of each receptor–ligand complex in a molecular mechanics (MM) framework, using an implicit continuum solvent model to estimate desolvation effects, most commonly the Poisson–Boltzmann Surface Area (PBSA) or Generalised Born Surface Area (GBSA) models. Together, these modifications result in a reduction in computational expense by several orders of magnitude and are routinely used for processing tens of compounds in a lead optimisation study or, given sufficient processing power, for 10^2 to 10^3 compounds in a virtual screening scenario [48–50]. Table 2 summarises a number of successful applications of these various physics-based scoring methods.

Although gaining in popularity, these methods are not without their limitations. While they are typically reported to be significantly more robust and more accurate than scoring functions, the accuracy is still below that usually required in typical lead optimisation applications to differentiate highly similar compounds [51]. Success is very dependent on the nature of the system in question, which is not surprising given the many approximations inherent in the methods, and therefore they are likely to work best where the binding mode is known or can be estimated reasonably accurately, and where strain, desolvation and entropic changes are likely to be fairly constant across a series of ligands. In addition, as with more traditional QSAR approaches, the quality and utility of regression-based models such as LIE depend heavily on the composition of the training set. Even so, these methods hold promise for the further development of fast and accurate scoring protocols. There is also much to be said for revisiting receptor-based QSAR methods such as COMBINE [52] which allow for variable weightings to be applied to different subsites within the binding pocket.

Pulling together multiple methodologies: workflow design

While the concept of workflow design has grown in popularity in recent years, it is not new to molecular modellers: it has long been common practice to write scripts to link together diverse programs and utilities in order to automate complex or repetitive tasks. More recently, however, attention has turned towards graphical interfaces specifically designed for facilitating workflow design and execution, with well-known examples including PipelinePilot (http://www.scitegic.com), InforSense ChemSense (http://www.inforsense.com), Knime (http://www.knime.org) and Taverna (http://taverna.sourceforge.net). These programs allow easy integration of diverse applications running on different hardware platforms, including web services and distributed computing

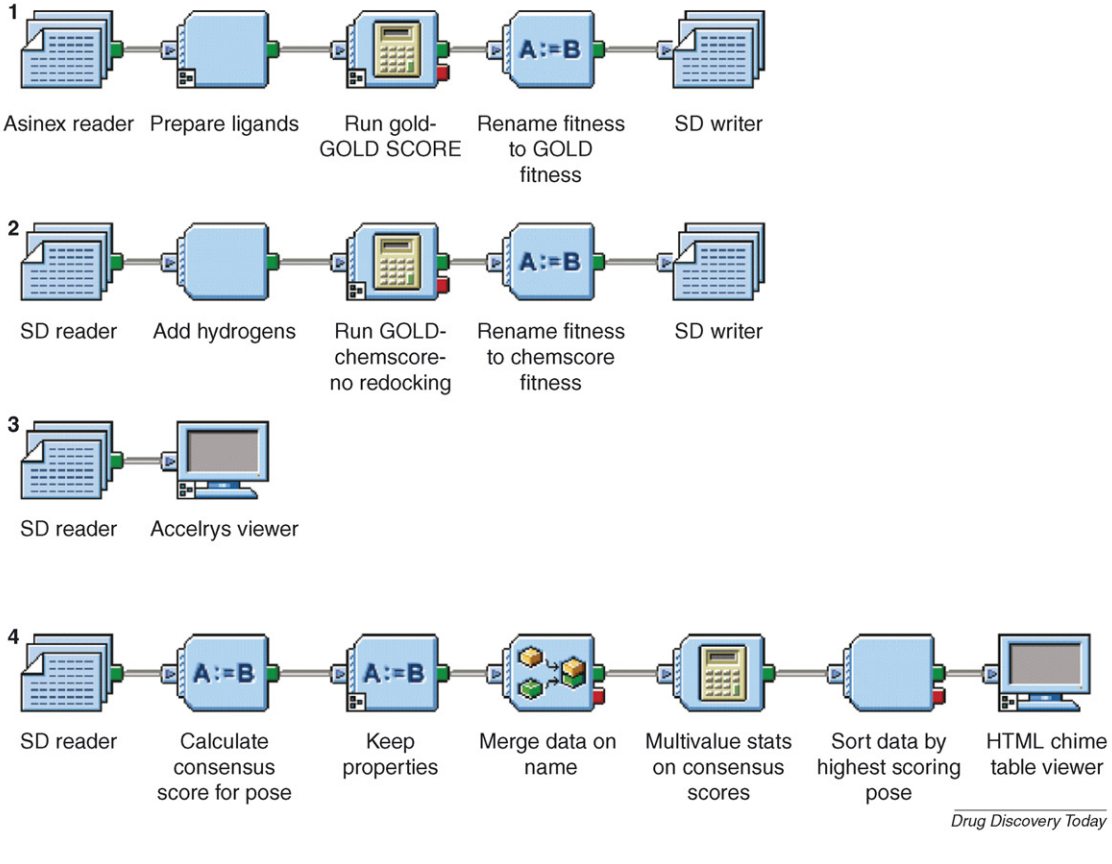

FIGURE 3

Automated virtual screening workflow. Screenshot of a typical Pipeline Pilot workflow for virtual screening. Briefly, the individual components represent the reading of an initial ligand dataset (*Asinex Reader*), docking with Gold, re-scoring with the ChemScore scoring function, visualisation of the docked poses (*Accelrys Viewer*), calculation of consensus scores and presentation of data in a web-based report (*HTML Table Viewer*). Image courtesy of Accelrys Inc. (http://www.accelrys.com/products/scitegic).

of compute-intensive tasks. They provide utilities for interconversion of a wide range of data formats and user-friendly interfaces to tasks that traditionally require an expert user, and allow rapid prototyping and sharing of novel modelling protocols with the automatic archiving of workflows and data. Although these software systems were designed for general data management tasks, virtual screening is a good example of a compute-intensive task requiring the integration of many applications in chemoinformatics (e.g. database processing), modelling (e.g. receptor preparation, docking, simulation) and analysis (e.g. data-mining, statistical analysis and report generation) (Figure 2).

The *virtual screening workflow* in the Maestro graphical user interface (Schrödinger, Inc., http://www.schrodinger.com) offers the user a unified interface for compound database processing and the submission of a series of large-scale docking runs using a hierarchy of Glide docking protocols. While this is a straightforward example of automation, other Maestro workflows offer more novel functionality: for example, *induced fit docking* iterates between Glide docking and Prime homology modelling functionality in order to incorporate some degree of protein backbone and sidechain refinement [24]; *quantum-polarised ligand docking* iterates between Glide docking and Jaguar quantum chemical calculations in order to derive partial atomic charges that reflect polarisation of the ligand by the protein [31]. In PipelinePilot, virtual screening workflows can be tailored from a variety of chemoinformatics and molecular modelling modules supplied by Accelrys Inc. and other software vendors, including a wide range of molecular descriptors and statistical tools to enable the rapid derivation of models for QSAR or druglikeness (Figure 3) [53]. Such examples illustrate the value of a broad and modular software platform for the design of novel functionality.

There are obvious drawbacks in seeking to automate every modelling task. Virtual screening is not a mature enough process to be thought of as a 'black-box' in which one protocol is adequate for every system. Most modellers would emphasise the importance of tailoring the docking and analysis procedures for the specific characteristics of the particular target or dataset and ensuring that they are validated as objectively as possible. However, workflows enable complex virtual screening and analysis protocols to become more routinely accessible and, when well-designed, to be adapted to the case in hand. Areas of interest for future development might include:

- iterative docking procedures utilising multiple docking programs and receptor models (either multiple conformations of the target or a set of anti-targets), with appropriate analysis of the resulting docking output;
- routine use of simulation-based refinement and scoring procedures for larger datasets, incorporating checks that binding modes and ligand chemistries are within the applicability domain of previous training sets;
- integration of ligand-based and structure-based virtual screening techniques as a way to navigate very large datasets rapidly and to make maximum use of available SAR data.

Perhaps current workflows should be seen as first generation protocols with an emphasis on simple automation of core tasks; second generation workflows may see increasing sophistication in order to incorporate sufficient checks that the modeller can fully monitor the quality of the process at all stages of execution and can intervene where appropriate.

Conclusion

Despite many technical challenges, structure-based virtual screening has an important role in drug discovery. The underlying docking and scoring technologies have steadily improved and the importance of the adequate validation of pragmatic virtual screening protocols is now well recognized. While current software has demonstrated its worth in identification of novel lead compounds, the challenge for the future is to improve the predictive accuracy of scoring functions, particularly to enable these methods to have a greater impact in guiding lead optimisation. In the meantime, two important ways forward include improved methods for data mining of large datasets and the use of robust and well-designed workflows to bring a wider range of analysis and scoring procedures into daily use by the modeller.

Acknowledgements

The author would like to thank Dr David Clark and Dr Hazel Hunt for helpful discussions and proof-reading of the text.

References

1 Jalaie, M. and Shanmugasundaram, V. (2006) Virtual screening: are we there yet? *Mini Rev. Med. Chem.* 6, 1159–1167

2 Fara, D.C. *et al.* (2006) Integration of virtual and physical screening. *Drug Discov. Today Technol.* 3, 377–385

3 Eckert, H. and Bajorath, J. (2006) Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov. Today* 12, 225–233

4 Muegge, I. and Oloff, S. (2006) Advances in virtual screening. *Drug Discov. Today Technol.* 3, 405–411

5 Klebe, G. (2006) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today* 11, 580–594

6 Seifert, M.H.J. *et al.* (2007) Virtual high-throughput screening of molecular databases. *Curr. Opin. Drug Discov. Dev.* 10, 298–307

7 Cole, J.C. *et al.* (2005) Comparing protein–ligand docking programs is difficult. *Proteins* 60, 325–332

8 Warren, G.L. *et al.* (2006) A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49, 5912–5913

9 (a) Chen, H. *et al.* (2006) On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.* 46, 401–415;
(b) Perola, E. *et al.* (2007) Comments on the article 'On evaluating molecular-docking methods for pose prediction and enrichment factors'. *J. Chem. Inf. Comput. Sci.* 47, 251–253;
(c) Lovell, T. *et al.* (2007) On evaluating molecular-docking methods for pose prediction and enrichment factors (Erratum). *J. Chem. Inf. Model* ASAP Web Release Data 18 Oct. 2007

10 Feher, M. (2006) Consensus scoring for protein–ligand interactions. *Drug Discov. Today* 11, 421–428

11 Zhang, Q. and Muegge, I. (2006) Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting and consensus scoring. *J. Med. Chem.* 49, 1536–1548

12 Hawkins, P.C.D. *et al.* (2007) Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* 50, 74–82

13 McGaughey, G.B. *et al.* (2007) Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* 47, 1504–1519

14 Jain, A.N. (2006) Scoring functions for protein–ligand docking. *Curr. Protein Pept. Sci.* 7, 407–420

15 Rajamani, R. and Good, A.C. (2007) Ranking poses in structure-based lead discovery and optimization: current trends in scoring function development. *Curr. Opin. Drug Discov. Dev.* 10, 308–315

16 Tirado-Rives, J. and Jorgensen, W.L. (2006) Contribution of conformer focusing to the uncertainty in predicting free energies for protein–ligand binding. *J. Med. Chem.* 49, 5880–5884

17 Mobley, D.L. *et al.* (2007) Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* 3, 1231–1235

18 Chang, C.A. *et al.* (2007) Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1534–1539

19 Rockey, W.M. and Elcock, A.H. (2006) Structure selection for protein kinase docking and virtual screening: homology models or crystal structures? *Curr. Protein Pep. Sci.* 7, 437–457

20 Thomas, M.P. *et al.* (2006) Protein structures in virtual screening: a case study with CDK2. *J. Med. Chem.* 49, 92–104

21 Verdonk, M.L. *et al.* (2005) Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* 48, 6504–6515

22 Cavasotto, C.N. and Abagyan, R.A. (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* 337, 209–225

23 Cavasotto, C.N. *et al.* (2005) The challenge of considering receptor flexibility in ligand docking and virtual screening. *Curr. Comput. Aid. Drug Des.* 1, 423–440

24 Sherman, W. *et al.* (2006) Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* 49, 534–553

25 Muegge, I. (2006) PMF scoring revisited. *J. Med. Chem.* 49, 5895–5902

26 Velec, H.F.G. *et al.* (2005) DrugScoreCSD – knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native poses and better affinity prediction. *J. Med. Chem.* 48, 6296–6303

27 Friesner, R.A. *et al.* (2006) Extra precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* 49, 6177–6198

28 Zhou, Z. *et al.* (2007) Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets. *J. Chem. Inf. Model.* 47, 1599–1608

29 Mooij, W.T.M. and Verdonk, M.L. (2005) General and targeted statistical potentials for protein–ligand interactions. *Proteins* 61, 272–287

30 Raha, K. and Merz, K.M. (2005) Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein–ligand complexes. *J. Med. Chem.* 48, 4558–4575

31 Cho, A.E. *et al.* (2005) Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* 26, 915–931

32 Oda, A. *et al.* (2006) Comparison of consensus scoring strategies for evaluating computational models of protein–ligand complexes. *J. Chem. Inf. Model.* 46, 380–391

33 Betzi, S. *et al.* (2006) GFscore: a general nonlinear consensus scoring function for high-throughput docking. *J. Chem. Inf. Model.* 46, 1704–1712

34 Teramoto, R. and Fukunishi, H. (2007) Supervised consensus scoring for docking and virtual screening. *J. Chem. Inf. Model.* 47, 526–534

35 Hartshorn, M.J. *et al.* (2007) Diverse, high-quality test set for the validation of protein–ligand docking performance. *J. Med. Chem.* 50, 726–741

36 Huang, N. *et al.* (2006) Benchmarking sets for molecular docking. *J. Med. Chem.* 49, 6789–6801

37 Pham, T.A. and Jain, A.N. (2006) Parameter estimation for scoring protein–ligand interactions using negative training data. *J. Med. Chem.* 49, 5856–5868

38 Waszkowycz, B. *et al.* (2001) Large-scale virtual screening for discovering leads in the postgenomic era. *IBM Syst. J.* 40, 360–376

39 Oprea, T.I. *et al.* (2007) Lead-like, drug-like or 'Pub-like': how different are they? *J. Comput. Aid. Mol. Des.* 21, 113–119

40 Jacobsson, M. and Karlén, A. (2006) Ligand bias of scoring functions in structure-based virtual screening. *J. Chem. Inf. Model.* 46, 1334–1343

41 Carta, G. *et al.* (2007) Unbiasing scoring functions: a new normalization and rescoring strategy. *J. Chem. Inf. Model.* 47, 1564–1571

42 Perola, E. (2006) Minimizing false positives in kinase virtual screens. *Proteins* 64, 422–435

43 Deng, Z. *et al.* (2004) Structural Interaction Fingerprint (SiFT): a novel method for analyzing three-dimensional protein–ligand binding interactions. *J. Med. Chem.* 47, 337–344

44 Claussen, H. *et al.* (2004) The FlexX database docking environment – rational extraction of receptor based pharmacophores. *Curr. Drug Discov. Technol.* 1, 49–60

45 Joseph-McCarthy, D. *et al.* (2007) Lead optimization via high-throughput molecular docking. *Curr. Opin. Drug Discov. Dev.* 10, 264–274

46 Foloppe, N. *et al.* (2006) Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* 13, 3583–3608

47 Huang, N. and Jacobson, M.P. (2007) Physics-based methods for studying protein–ligand interactions. *Curr. Opin. Drug Discov. Dev.* 10, 325–331

48 Huang, N. *et al.* (2006) Physics-based scoring of protein–ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* 46, 243–253

49 Lee, M.R. and Sun, Y. (2007) Improving docking accuracy through molecular mechanics generalized Born optimization and scoring. *J. Chem. Theory Comput.* 3, 1106–1119

50 Brown, S.P. and Muchmore, S.W. (2007) Rapid estimation of relative protein–ligand binding affinities using a high-throughput version of MM-PBSA. *J. Chem. Inf. Model.* 47, 1493–1503

51 Kuhn, B. *et al.* (2005) Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* 48, 4040–4048

52 Lushington, G.H. *et al.* (2007) Whither Combine? New opportunities for receptor-based QSAR. *Curr. Med. Chem.* 14, 1863

53 Hassan, M. *et al.* (2006) Cheminformatics analysis and learning in a data pipelining environment. *Mol. Divers.* 10, 283–299

54 Lyne, P.D. *et al.* (2004) Identification of compounds with nanomolar binding affinity for Checkpoint Kinase-1 using knowledge-based virtual screening. *J. Med. Chem.* 47, 1962–1968

55 Barril, X. *et al.* (2005) Structure-based discovery of a new class of Hsp90 inhibitors. *Bioorg. Med. Chem. Lett.* 15, 5187–5191

56 Foloppe, N. *et al.* (2006) Identification of chemically diverse Chk1 inhibitors by receptor-based virtual screening. *Bioorg. Med. Chem.* 14, 4792–4802

57 Rummey, C. *et al.* (2006) In silico fragment-based discovery of DPP-IV S1 pocket binders. *Bioorg. Med. Chem. Lett.* 16, 1405–1409

58 Cavasotto, C.N. *et al.* (2006) In silico identification of novel EGFR inhibitors with antiproliferative activity against cancer cells. *Bioorg. Med. Chem. Lett.* 16, 1969–1974

59 Kellenberger, E. *et al.* (2007) Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. *J. Med. Chem.* 50, 1294–1303

60 Price, S. *et al.* (2007) Identification and optimisation of a series of substituted 5-(1*H*-pyrazol-3-yl)-thiophene-2-hydroxamic acids as potent histone deacetylase (HDAC) inhibitors. *Bioorg. Med. Chem. Lett.* 17, 370–375

61 De Graaf, C. *et al.* (2006) Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* 49, 2417–2430

62 Jorgensen, W.L. *et al.* (2006) Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* 16, 663–667

63 Stjernschantz, E. *et al.* (2006) Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method. *J. Chem. Inf. Model.* 46, 1972–1983

64 Su, Y. *et al.* (2007) Linear interaction energy (LIE) models for ligand binding in implicit solvent: theory and application to the binding of NNRTIs to HIV-1 reverse transcriptase. *J. Chem. Theory Comput.* 3, 256–277

65 Bortolato, A. and Moro, S. (2007) In silico binding free energy predictability by using the linear interaction energy (LIE) method: bromobenzimidazole CK2 inhibitors as a case study. *J. Chem. Inf. Model.* 47, 572–582

66 Steinbrecher, T. *et al.* (2006) A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. *J. Med. Chem.* 49, 1837–1844

67 Lyne, P.D. *et al.* (2006) Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* 49, 4805–4808