Statistical potentials and scoring functions applied to protein–ligand binding

Holger Gohlke and Gerhard Klebe*

In virtual screening, small-molecule ligands are docked into protein binding sites and their binding affinity is predicted. Knowledge-based, regression-based and first-principle-based methods have been developed to rank computer-generated binding modes. As a result of still existing deficiencies, a best compromise might be the combination of several scoring schemes into a consensus scoring approach.

Addresses
Department of Pharmaceutical Chemistry, Philipps University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany
*e-mail: klebe@mailer.uni-marburg.de

Current Opinion in Structural Biology 2001, 11:231–235

0959-440X/01/$ – see front matter

Introduction

As we approach the post-sequencing phase of many genome projects, it is estimated that the number of potential drug targets will increase from about 500 at present to around 5000–10,000 in the next few years [1]. Methodological advances in protein structure determination will provide a tremendous number of new protein structures. Accordingly, the scope and importance of structure-based drug design projects (for reviews, see [2–6]) involving virtual screening [7,8] will increase significantly. A key prerequisite for the successful application of these techniques is a profound understanding of the criteria that determine whether a ligand binds to a particular biological target. A move towards a limited set of prospective leads, discovered *in silico*, can only be considered as a practical alternative to the high-throughput screening of large compound libraries if the binding modes and affinities of structurally distinct ligands for the protein under investigation can be predicted correctly.

For a virtual screening strategy to be reliable, both relevant binding geometries must be generated and relevant binding affinities must be predicted correctly. Substantial progress has been made in the latter area in 1999 and 2000, resulting in a broad spectrum of methods for the estimation of binding affinities. Here, we summarize and compare methods that have recently emerged in the field, primarily newly developed knowledge-based potentials and, furthermore, improved empirical scoring functions or techniques based on first principles. Further information can be found in other recent reviews [6,9,10,11•,12–15].

Knowledge-based potentials

Knowledge-based approaches evaluate the increasing number of experimentally determined protein–ligand complexes by statistical means to extract rules on preferred interaction geometries. These rules are converted into 'pseudo-potentials', which, in turn, can be applied to score computer-generated ligand binding modes. Compared with force-field potentials, these knowledge-based potentials tend to be 'softer', allowing better handling of the uncertainties and deficiencies of computed interaction geometries. Furthermore, such a statistical approach implicitly incorporates physical effects not yet fully understood from a theoretical point of view, for example, (de)solvation and polarization. A major concern with statistical potentials is that they might only model geometric properties close to experimental (near-native) situations previously observed in the crystal structures used to derive these potentials. Might they also provide a meaningful description of the entire effective energy surface of a generalized protein–ligand complex? The application of a knowledge-based potential to describe protein unfolding of the GCN4 leucine zipper [16•] suggests they may approach a more general validity, as similar results were obtained to those using the CHARMM all-atom force-field. Three knowledge-based approaches to the scoring of protein–ligand interactions have been developed during the past two years [17••,18••,19•].

Muegge and Martin [18••] have extracted distance-dependent Helmholtz free energies for pairs of 16 protein-atom and 34 ligand-atom types from 697 crystallographically determined protein–ligand complexes to develop a potential of mean force score ('PMFScore'). Using the crystal structures of 77 protein–ligand complexes, a correlation with a squared regression coefficient of $r^2 = 0.61$ was obtained for predicting their experimental binding energies. Implemented into the docking program DOCK4, PMFScore was used to screen 3247 small molecules for binding to the FK506-binding protein [20]. Compared with the standard force-field scoring in DOCK4, PMFScore achieved a significantly improved enrichment of particularly weakly binding ligands. The DOCK4/PMFScore combination was also used to predict the binding modes of neuraminidase inhibitors [21] and stromelysin-1 (MMP-3) inhibitors [22].

Thornton and co-workers [19•] have also developed a potential of mean force to describe protein–ligand interactions. Following Sippl's approach [23], 820 pair-potentials were derived for 40 nonmetal atom types, including polar hydrogens, based on 351 carefully selected Protein Data Bank (PDB) complexes. Two sets of potentials were reported, one excluding ('BLEEP-1') and one including ('BLEEP-2') interactions with water. Applying BLEEP-2 to a set of 90 crystallographically

determined complexes of known binding affinities, a correlation coefficient of 0.74 was obtained [24]. As a test for the recognition of near-native binding modes, BLEEP-2 also successfully identified the crystal geometry of heparin bound to bFGF among a set of decoy structures generated by FTDock.

Gohlke *et al.* [17••] have developed a knowledge-based scoring function ('DrugScore') by converting structural information for 1376 protein–ligand complexes, extracted from Relibase [25], into distance-dependent pair-potentials and solvent-accessible surface-dependent singlet-potentials using 17 different atom types. The sum of both terms is used to score protein–ligand binding modes. For two test sets of 91 and 68 complexes, DrugScore recognizes a near-native binding mode out of multiple solutions generated by the docking program FlexX and scored as best solution (rank 1) in 75% of all cases. For ligand geometries generated by DOCK4, DrugScore is superior to DOCK's 'chemical scoring'. This underlines its discriminative power to extract near-native binding modes. For a set of 56 crystallographically determined protein–ligand complexes, an r^2 value of 0.56 is obtained for predicting binding affinities. DrugScore also successfully ranked data sets of docked thrombin, trypsin and thermolysin inhibitors [26]. Finally, the pair-potentials of DrugScore were used to identify 'hot spots' in binding pockets. For a set of 159 crystal structures, the observed atom type of a ligand corresponds to an atom type predicted to be most favorable in that region in 74% of all cases [26].

All three approaches are based on the same formalism and relate observed frequency or probability distributions to pair-(pseudo-)potentials $\Delta W_{ij}(r)$:

$$\Delta W_{ij}(r) \propto -\ln \frac{g_{ij}(r)}{g_{ref}}.$$

Here, $g_{ij}(r)$ is a frequency or probability distribution of atom pairs of types i and j at a distance r from each other, and g_{ref} corresponds to a reference distribution. The choice of the latter is crucial to the form of the potential and all three approaches differ in this aspect [27,28]. To include implicit solvation effects, Muegge and Martin [18••], and Mitchell *et al.* [19•] chose a relatively large limit of 12 Å and 8 Å, respectively, up to which contact data have been sampled for compiling the pair-potentials. In contrast, Gohlke *et al.* [17••] introduced additional solvent-accessible surface-dependent singlet-potentials and data for the pair-potentials were sampled up to 6 Å, emphasizing specific interactions between ligand and protein atoms. Applying an extended cutoff, Muegge and Martin [18••] stressed the need for a volume correction term to compensate for intraligand interactions; intraprotein interactions, however, are not considered. Interestingly, for two test sets that are limited to distinct protein classes, the influence of the ligand volume correction factor was found to be weak, whereas for test sets containing diverse protein–ligand complexes, an improved ranking is described once the volume correction is applied [29]. In this context, Shimada *et al.* [30] proposed a self-consistent approach to analyze knowledge-based atom-pair-potentials. Protein–ligand complexes modeled using a particular input potential were selected to construct databases from which statistical protein–ligand potentials were derived. This strategy allows both the consequences of modifications to such potentials to be explored and the study of their performance in reproducing the input potential.

Regression-based scoring functions

Empirical scoring functions estimate the binding affinity of protein–ligand complexes by adding up interaction terms derived from weighted structural parameters of the complexes. The weights are assigned by regression methods — fitting predicted and experimentally determined affinities to a given set of training complexes. Although the decomposition of the binding free energy into single portions is not valid from a theoretical point of view [31,32], these weights are usually regarded as free energy contributions from specific interactions, for example, hydrogen bonding, ionic interactions, hydrophobic interactions and entropic contributions. Apart from their computational efficiency, the individual terms of such a function reflect the way a medicinal chemist would decompose contributions to protein–ligand binding [33•].

A major drawback of any regression-based scoring function is the dependence on the size, composition and generality of the training set used to derive the weights. As yet, thermodynamic data for protein–ligand complexes are relatively sparse and more research is required to assemble such information. Another cause for concern is the implicit assumption that each occurrence of a particular interaction contributes equivalently [11•]. This may result in an overestimation of polar interactions with respect to nonpolar ones [10]. Stahl [34] addressed this problem for hydrogen bonds in the regression-based scoring function of FlexX by introducing a weighting scheme that scales these contributions according to the solvent accessibility of the hydrogen bond partners. Retrieval of thrombin inhibitors from the World Drug Index (WDI) yielded a higher percentage of compounds known to be active at thrombin on the best-scored ranks (improved enrichment factor) using the latter solvent-accessibility-dependent scoring scheme.

A clear trend away from generally applicable functions towards tailor-made scoring functions specifically adapted to predict binding affinities for one protein has been seen in the past two years [33•]. Although better (cross-validated) correlations between predicted and experimental data are achieved with the more recent developments, one has to bear in mind that regression-based approaches can only interpolate. Thus, it remains to be seen whether these tailor-made functions are capable of identifying new

molecular scaffolds that are not present in the training set. Nevertheless, an encouraging example is given by the study of Rognan *et al.* [35], who reliably predicted the binding affinities of 26 nonapeptides for a class I MHC protein using only five data points to adapt a six parameter regression equation ('Fresno') similar in form to SCORE [36] or ChemScore [37]. Transferring the adapted equation to a set of octapeptides binding to a related class I MHC protein, however, achieved no correlation, demonstrating the limited scope of such individually calibrated scoring functions that must be reparameterized for every new application. Similar concepts based on the formalism of 3D-QSAR analyses try to consider additional information about the specific binding pocket [38,39].

Kasper *et al.* [40] decomposed the measured free energy of binding of peptides to the molecular chaperone DnaK into nonpolar, electrostatic and conformational entropy contributions. They related the nonpolar interactions to the change in solvent-accessible surface, used the empirical scale of Pickett and Sternberg [41] to account for changes in the conformational entropy of the solutes, and included Coulombic interactions and solvation effects following a scheme derived by Gilson and Honig [42]. Finally, the individual energy contributions were scaled to the given binding affinities of 11 peptides of equal length. In this case, the entropic term showed very high fluctuations and contributed little to the total binding energy, so it was discarded from the final equation. This is almost certainly a result of using the restricted training set: contributions with little or no variance across the data set will not be recognized.

An interesting alternative to partially circumvent such shortcomings has been presented by Murray *et al.* [43]. They adapted their previously published general scoring function [37] to a set of thrombin complexes under the restraints of the old (diverse) training set using Bayesian regression, allowing fine-tuning of their function to either general or specific predictive power.

First-principle-based approaches

First-principle-based approaches also approximate the binding free energy of protein–ligand complexes by adding up the individual contributions of different types of interaction. In contrast to regression-based scoring functions, however, the individual terms of such a 'master equation' are derived from physico-chemical theory and are not determined by fitting to experimental affinities. In most cases, gas-phase molecular mechanical energies are combined with solvation free energies and contributions from vibrational, rotational and translational entropies. While comparing structurally related ligands, the latter contributions to entropy are frequently identified to be of negligible influence. Evaluation of the solvent contributions still represents a major challenge in view of the computational demands and accuracy [44].

Honig and co-workers [45] calculated binding free energies for four inhibitors of bovine β-trypsin. They considered changes in Coulombic and solvation energies, as derived from the solution of the finite-difference Poisson–Boltzmann equation, together with contributions resulting from nonpolar interactions and losses of conformational entropy either of the protein or of the ligands. Interestingly, electrostatic interactions turn out to be adverse to binding in all cases. Although the calculated binding free energies exceed the experimental values by about 10 kcal/mol in all cases, the relative order of binding is correctly predicted. Hoffmann *et al.* [46] combined CHARMM22 force-field energies with electrostatic and nonpolar portions of the solvation energy to rerank ligand binding modes generated by FlexX. Because of the roughness of the energy surface provided by the all-atom force-field, a time-consuming energy minimization of the ligand conformations (several hours per 100 conformations) inside the protein binding site was necessary.

The Kuntz group [47] investigated the impact of ligand solvation on docking. Electrostatic and nonpolar solvation free energies for putative ligands taken from the Available Chemicals Directory (ACD) were precalculated and stored in look-up tables. These energies were subsequently subtracted from the interaction energies experienced by the ligands inside the protein (calculated using electrostatic energies derived from DelPhi potentials and van der Waals contributions). Upon docking ligands into thymidylate synthase, T4 lysozyme mutants and dihydrofolate reductase (DHFR), it was observed that the scores of known ligands improved and that the number of molecules predicted too well because of high formal charges or large molecular weights was reduced. In another study, Zou *et al.* [48] applied a modified 'generalized Born' (GBSA) model to ligand–receptor binding and optimized the parameters with respect to DHFR and trypsin, such that the predicted binding energies of known inhibitors agree well with experimental values and the number of false-positive predictions is reduced.

Extending their 'MM-PBSA/GBSA' approach [44,49], which combines molecular mechanics gas-phase energies with a continuum solvation approach to handle solvent effects and with normal mode analysis, Kuhn and Kollman [50] described a 'computational fluorine scanning' approach to probe possible hydrogen-to-fluorine replacements on ligands with respect to an improvement of ligand binding affinity. Based on a single molecular dynamics simulation of a reference ligand, this reference is substituted in a subsequent analysis of the trajectory by a partially fluorinated ligand. The free energy is estimated for the complex, protein and ligand by applying the MM-PBSA energy analysis. In a similar manner, 'computational alanine scanning' was used to probe protein–protein interactions [51]. A major limitation of this technique is the fact that changes in binding can only be

estimated if a smaller group replaces a larger one; mutations resulting in a larger group replacing a smaller group will create steric clashes that must be relieved before any binding energy calculation.

Consensus scoring and comparison of approaches

In an intersection-based consensus scoring approach, Charifson *et al.* [52•] combined up to three scoring functions to rank docked ligand–protein geometries. Compared with the performance of the single functions, improved hit rates were found. However, it has been questioned by the authors themselves [52•] whether such an approach is of general use to predict binding free energies of small sets of compounds. Similarly, So and Karplus [53] combined the affinity predictions of five QSAR-based methods and observed a general improvement of the predictivity with an increasing number of models considered.

Independent comparative studies of docking/scoring approaches are still rather rare. In a recent publication, Bissantz *et al.* [54] assessed the accuracy of several methods using three different docking programs (DOCK4, FlexX, Gold) in combination with seven scoring functions applied to two protein targets. Because of the restricted test scenario, the identification of the 'best' docking/scoring combination is difficult, especially as the combination of best predictive power varied with the target selected.

Conclusions

The past two years have seen substantial progress in the development of tools to score protein–ligand interactions, particularly in the area of knowledge-based potentials. Knowledge-based potentials have benefited from the steadily increasing amount of structural data on protein–ligand complexes. As no preselected training sets are used, these methods should be generally applicable, in particular compared with the specifically adapted regression-based approaches. The incorporation of computationally feasible and yet sufficiently accurate (continuum) solvation models dominated the developments in the field of first-principle-based techniques.

The increasing number of successful applications using these scoring schemes (see [2,55] for references) indicates that currently available methods allow discovery of active molecules in a virtual screening approach. The improved performance of consensus scoring approaches suggests that, at present, no individual scoring function adequately treats all of the effects important for protein–ligand binding. Thus, more effort is required to improve the existing functions; in particular, better experimental, structural and energetic data must be made available. In addition, competitions such as CATFEE (Critical Assessment of Techniques for Free Energy Evaluation; http://uqbar.ncifcrf.gov/~catfee) will help elucidate the scope and reliability of these techniques.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Drews J: Drug discovery: a historical perspective. *Science* 2000, 287:1960-1964.
2. Murcko MA, Caron PR, Charifson PS: Structure-based drug design. *Annu Rep Med Chem* 1999, 34:297-306.
3. Kirkpatrick DL, Watson S, Ulhaq S: Structure-based drug design: combinatorial chemistry and molecular modeling. *Comb Chem High Throughput Screen* 1999, 2:211-221.
4. Joseph-McCarthy D: Computational approaches to structure-based ligand design. *Pharm Therap* 1999, 84:179-191.
5. Gane PG, Dean PM: Recent advances in structure-based rational drug design. *Curr Opin Struct Biol* 2000, 10:401-404.
6. Böhm HJ, Stahl M: Structure-based library design: molecular modelling merges with combinatorial chemistry. *Curr Opin Chem Biol* 2000, 4:283-286.
7. Böhm HJ, Schneider G (Eds): *Virtual Screening for Bioactive Molecules.* Weinheim: Wiley-VCH; 2000. [Mannhold R, Kubinyi H, Timmermann H (Series Editors): Methods and Principles in Medicinal Chemistry, vol 10.]
8. Klebe G (Ed): *Virtual Screening: An Alternative or Complement to High Throughput Screening?* Dordrecht: Kluwer/Escom; 2000. [Martin YC (Series Editor): Perspectives in Drug Discovery and Design, vol 20.]
9. Müller-Dethlefs K, Hobza P: Noncovalent interactions: a challenge for experiment and theory. *Chem Rev* 2000, 100:143-167.
10. Davis AM, Teague SJ: Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew Chem Int Ed Engl* 1999, 38:736-749.
11. • Tame JRH: Scoring functions: a view from the bench. *J Comput Aided Mol Des* 1999, 13:99-108.

A critical review, taken from an experimentalist's point of view, of the complexity of physical chemistry that underlies noncovalent interactions between protein and ligands. Shortcomings of recent scoring schemes are revealed and related to experimental results.

12. Hirst JD: Predicting ligand binding energies. *Curr Opin Drug Discov Development* 1998, 1:28-33.
13. Oprea TI, Marshall GR: Receptor-based prediction of binding activities. In *3D Qsar in Drug Design: Ligand Protein Interactions and Molecular Similarity*. Edited by Kubinyi H, Folkers G, Martin YC. Dordrecht: Kluwer/Escom; 1998:3-17.
14. Reddy MR, Viswanadhan VN, Erion MD: Rapid estimation of relative binding affinities of enzyme inhibitors. In *3D QSAR in Drug Design: Ligand Protein Interactions and Molecular Similarity*. Edited by Kubinyi H, Folkers G, Martin YC. Dordrecht: Kluwer/Escom; 1998:85-98.
15. Knegtel RMA, Grootenhuis PDJ: Binding-affinities and non-bonded interaction energies. In *3D QSAR in Drug Design: Ligand Protein Interactions And Molecular Similarity*. Edited by Kubinyi H, Folkers G, Martin YC. Dordrecht: Kluwer/Escom; 1998:99-114.
16. • Mohanty D, Dominy BN, Kolinski A, Brooks CL III, Skolnick J: Correlation between knowledge-based and detailed atomic potentials. *Proteins* 1999, 35:447-452.

This work investigates the relationship between pseudo free energies for unfolding the GCN4 leucine zipper calculated either from knowledge-based potentials or from the CHARMM all-atom potential in combination with a generalized Born solvation component. Both energies were found to be highly correlated, even for largely unfolded states of the protein. The authors conclude that, although the knowledge-based potentials were derived from native structures of proteins, they can also describe non-native states.

17. •• Gohlke H, Hendlich M, Klebe G: Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 2000, 295:337-356.

The development of the knowledge-based scoring function DrugScore is described. The combination of a distance-dependent pair-potential and a solvent-accessible surface-dependent singlet-potential discriminates efficiently between well-docked ligand binding modes and those largely deviating from the native structure. The new function implicitly regards solvation and entropy contributions, and is independent of assumptions concerning protonation states.

18. Muegge I, Martin YC: A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* 1999, 42:791-804.
•• The authors present the development of a potential of mean force score (PMFScore) consisting of a distance-dependent pair-potential. They introduced a volume correction term in connection with a rather large sampling distance to account for solvent effects. For the majority of the eight different test sets of protein–ligand complexes, the score performs better than established scoring schemes in ranking the protein–ligand complexes.

19. Mitchell JBO, Laskowski RA, Alex A, Thornton JM: BLEEP-potential of mean force describing protein-ligand interactions: I. Generating potential. *J Comput Chem* 1999, 20:1165-1176.
• A new knowledge-based scoring function is introduced.

20. Muegge I, Martin YC, Hajduk PJ, Fesik SW: Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. *J Med Chem* 1999, 42:2498-2503.

21. Muegge I: The effect of small changes in protein structure on predicted binding modes of known inhibitors of influenza virus neuraminidase: PMF-scoring in DOCK4. *Med Chem Res* 1999, 9:490-500.

22. Sookhee H, Andreani R, Robbins A, Muegge I: Evaluation of docking/scoring approaches: a comparative study based on MMP3 inhibitors. *J Comput Aided Mol Des* 2000, 14:435-448.

23. Sippl MJ: Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J Mol Biol* 1990, 213:859-883.

24. Mitchell JBO, Laskowski RA, Alex A, Forster MJ, Thornton JM: BLEEP-potential of mean force describing protein-ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J Comput Chem* 1999, 20:1177-1185.

25. Hendlich M: Databases for protein-ligand complexes. *Acta Crystallogr* 1998, D54:1178-1182.

26. Gohlke H, Hendlich M, Klebe G: Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function. *Persp Drug Discov Des* 2000, 20:115-144.

27. Muegge I: A knowledge-based scoring function for protein-ligand interactions: probing the reference state. *Persp Drug Discov Des* 2000, 20:99-114.

28. Stahl M: Structure-based library design. In *Virtual Screening for Bioactive Molecules*. Edited by Böhm HJ, Schneider G. Weinheim: Wiley-VCH; 2000:229-264.

29. Muegge I: Effect of ligand volume correction on PMF-scoring. *J Comput Chem* 2001, in press.

30. Shimada J, Ishchenko AV, Shakhnovich EI: Analysis of knowledge-based protein-ligand potentials using a self-consistent method. *Protein Sci* 2000, 9:765-775.

31. Mark AE, van Gunsteren WF: Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J Mol Biol* 1994, 240:167-176.

32. Dill KA: Additivity principles in biochemistry. *J Biol Chem* 1997, 272:701-704.

33. Böhm HJ, Stahl M: Rapid empirical scoring functions in virtual screening applications. *Med Chem Res* 1999, 9:445-462.
• A comprehensive overview and comparison of current state-of-the-art regression-based scoring functions are presented.

34. Stahl M: Modifications of the scoring function in FlexX for virtual screening applications. *Persp Drug Discov Des* 2000, 20:83-98.

35. Rognan D, Lauemoller SL, Holm A, Buus S, Tschinke V: Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J Med Chem* 1999, 42:4650-4658.

36. Böhm HJ: The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* 1994, 8:243-256.

37. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP: Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 1997, 11:425-445.

38. Venkatarangan P, Hopfinger AJ: Prediction of ligand-receptor binding thermodynamics by free energy force field three-dimensional quantitative structure-activity relationship analysis: application to a set of glucose analogue inhibitors of glycogen phosphorylase. *J Med Chem* 1999, 42:2169-2179.

39. Kulkarni SS, Kulkarni VM: Structure based prediction of binding affinity of human immunodeficiency virus-1 protease inhibitors. *J Chem Inf Comput Sci* 1999, 39:1128-1140.

40. Kasper P, Christen P, Gehring H: Empirical calculation of the relative free energies of peptide binding to the molecular chaperone DnaK. *Proteins* 2000, 40:185-192.

41. Pickett SD, Sternberg MJ: Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol* 1993, 231:825-839.

42. Gilson MK, Honig B: Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* 1988, 4:7-18.

43. Murray CW, Auton TR, Eldridge MD: Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *J Comput Aided Mol Des* 1998, 12:503-519.

44. Massova I, Kollman PA: Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Persp Drug Discov Des* 2000, 18:113-135.

45. Polticelli F, Ascenzi P, Bolognesi M, Honig B: Structural determinants of trypsin affinity and specificity for cationic inhibitors. *Protein Sci* 1999, 8:2621-2629.

46. Hoffmann D, Kramer B, Washio T, Steinmetzer T, Rarey M, Lengauer T: Two-stage method for protein-ligand docking. *J Med Chem* 1999, 42:4422-4433.

47. Shoichet BK, Leach AR, Kuntz ID: Ligand solvation in molecular docking. *Proteins* 1999, 34:4-16.

48. Zou X, Sun Y, Kuntz ID: Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. *J Am Chem Soc* 1999, 121:8033-8043.

49. Chong LT, Duan Y, Wang L, Massova I, Kollman PA: Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc Natl Acad Sci USA* 1999, 96:14330-14335.

50. Kuhn B, Kollman PA: A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. *J Am Chem Soc* 2000, 122:3909-3916.

51. Massova I, Kollman PA: Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J Am Chem Soc* 1999, 121:8133-8143.

52. Charifson PS, Corkerey JJ, Murcko MA, Walters WP: Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 1999, 42:5100-5109.
• This work compares several scoring functions for their performance in virtual screening applications and introduces an intersection-based consensus approach to enhance the reliability of their predictions.

53. So SS, Karplus M: A comparative study of ligand-receptor complex binding affinity prediction methods based on glycogen phosphorylase inhibitors. *J Comput Aided Mol Des* 1999, 13:243-258.

54. Bissantz C, Folkers G, Rognan D: Protein-based virtual screening of chemical databases. 1. Evaluations of different docking/scoring combinations. *J Med Chem* 2000, 43:4759-4767.

55. Kubinyi H: Structure-based design of enzyme inhibitors and receptor ligands. *Curr Opin Drug Discov Develop* 1998, 1:4-15.