Journal of Medicinal Chemistry

Volume 49, Number 20 October 5, 2006

Docking and Scoring

Perspective

Prediction of Protein−Ligand Interactions. Docking and Scoring: Successes and Gaps

Andrew R. Leach,[†] Brian K. Shoichet,[‡] and Catherine E. Peishoff*,[§]

GlaxoSmithKline Pharmaceuticals, 1250 South Collegeville Road, Collegeville, Pennsylvania 19426, GlaxoSmithKline Pharmaceuticals, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, U.K., and Department of Pharmaceutical Chemistry, University of California—San Francisco, QB3 Building, 1700 4th Street, Box 2550, San Francisco, California 94148

Received August 18, 2006

Computational methods have become standard in today's medicinal chemistry tool kit. Like any tool, it is important to periodically evaluate utility and ask how function can be improved. In this section of the Journal, we call attention to the area of calculating molecular interactions, specifically docking, the positioning of a ligand in a protein binding site, and scoring, the quality assessment of docked ligands. As several recent reviews have made clear,[1−3] the technology has been productive for both finding and elaborating bioactive molecules. But has docking and scoring delivered on the promises first made over 20 years ago? To consider that question, we follow up on an extensive symposium held in Philadelphia during the 2004 Fall National Meeting of the American Chemistry Society and on subsequent meetings sponsored by the National Institutes of Health (NIH) and the National Institute of Standards and Technology (NIST) in 2005 and 2006 to address the outcomes of the American Chemical Society symposium. Speakers at the symposium were invited to contribute original manuscripts to be published with this overview to highlight the area of docking and scoring and to identify some of the major gaps yet to be addressed.[4−10]

In this overview, we first summarize the current state of docking and scoring technology and highlight key areas to address. Next, we turn to the topic of benchmarks and measurements to take up the data infrastructure on which new algorithm development depends. Finally, we open the topic of possible industrial, academic, and governmental partnerships that may be needed to fully deliver on the docking and scoring promise.

State of the Art

Docking and scoring technology is applied at different stages of the drug discovery process for three main purposes: (1) predicting the binding mode of a known active ligand; (2) identifying new ligands using virtual screening; (3) predicting the binding affinities of related compounds from a known active series. Of these three challenges, successful prediction of a ligand binding mode in a protein active site is perhaps the most straightforward and is the area where most success has been achieved. Docking a ligand into a binding site models several degrees of freedom. These are the six degrees of translational and rotational freedom of one body relative to another and then the conformational degrees of freedom of the ligand and of the protein. The first docking algorithms only considered translation and orientation, with both ligand and protein treated as rigid bodies. Increases in computer performance and new algorithms enabled the ligand conformational degrees of freedom to be explored. Most docking programs today treat the ligand as flexible with a rigid (or nearly rigid) receptor structure. Receptor flexibility remains one of the major challenges for the field.

Once the configurations of a system are sampled, docking programs must then score these to identify the most likely

* To whom correspondence should be addressed. Phone: (610) 917-6584. Fax: (610) 917-7393. E-mail: Catherine.E.Peishoff@gsk.com.
[†] GlaxoSmithKline, Hertfordshire.
[‡] University of California—San Francisco.
[§] GlaxoSmithKline, Pennsylvania.

10.1021/jm060999m CCC: $33.50
Published on Web 09/28/2006

candidate for the true structure. By rough count, about 30 docking programs have appeared in a public venue. The best of these predict the experimental pose about 70% of the time, although selecting the program that will give the best result for any given target is not straightforward. Kontoyianni et al.[11] demonstrated these points in a study of 69 diverse protein−ligand complexes with five docking programs where the overall results ranged from 38% to 69%, using a 2 Å root-mean-square deviation (rmsd) as being acceptable. Though not clear-cut, some docking algorithms typically do better with certain flavors of protein active sites; e.g., GOLD performs better with more hydrophilic sites.[11] Warren et al[10] tested 10 docking programs against 7 protein classes and compound series totaling 1303 molecules. Results varied from some algorithms with some targets predicting 0% of the poses satisfactorily and other combinations where >90% of the poses were within the 2 Å rmsd cutoff. It is reassuring that for five of the seven targets at least one docking algorithm produced ≥50% accurate poses, but the number of times the answer was 0% underscores the dilemma faced by computational and medicinal chemists today. Clearly, good poses can be produced, but how does one pick the program that will do so reliably for the targets of interest?

For the second challenge, virtual screens to identify new lead molecules may involve searching databases containing hundreds-of-thousands, if not millions, of molecules. Such molecules may come from several sources, including in-house compound collections, external databases of commercial compounds,[12] virtual libraries, and de novo designs. Each library molecule must be docked into the protein to predict a "pose" of the ligand in the active site. The best scoring pose of each molecule is then compared to arrive at a top-ranking "hit list." In principle, the functions used to calculate these scores predict the free energies of binding of every molecule being screened. In practice, however, one cannot hope for better than monotonic ranking of the molecules, and even this is typically beyond current methods. Indeed, docking results are often judged by "enrichment" of true hits among a larger number of molecules tested; by this criterion, even correct ranking is not expected. On the more positive side, virtual screening has become increasingly automated. This has allowed many structures to be processed with relatively little manual intervention. Access to inexpensive Linux clusters and other forms of both intra- and internet grid computing has been indispensable.

There are many published examples of successful virtual screens to identify new hit molecules. On the basis of such successes and unpublished internal studies, many pharmaceutical companies use virtual screening as part of the standard processes for choosing compounds to screen and for enhancing their compound collections. But does such enrichment happen for the right reasons; i.e., do algorithms both reproduce experimental binding positions and score these correct poses appropriately to the compounds' affinity? A recent publication by Cummings et al.[13] addressed this question. In this study, 49 known ligands across 5 proteins were seeded into a database of 1000 compounds from the MDL drug data report. Four docking programs were tested for their ability to select a protein's own ligands at rates better than random and to fit these ligands in the correct geometry. The screening results confirmed previous observations that the programs demonstrated substantial enrichment, albeit not consistently, across the targets. In one case, at least 80% of the known actives were found for two proteins, but the results were consistently worse than random for another target. The results of Warren et al.[10] demonstrate that the situation is similar when the compound set is a closely related set of bioactive molecules. In neither study was there a strong correlation between the ability of a program to produce a correct pose and its success in a virtual screen. Some of this can be attributed to the inherent danger of using a single metric such as rmsd, as poses can be fundamentally correct despite a large deviation in one part of the molecule. More worrisome are those cases where the poses are barely in the binding site, and yet good enrichment is observed. Enrichment may be due to screening out compounds that are wrong for the target rather than selecting those that are right. This does not imply that enrichment through virtual screening is a failure; clearly, it is not. Indeed, docking hit lists are often full of interesting molecules that look sensible to trained medicinal chemists and structural biologists. Most of these "sensible" molecules will not, however, bind to the target at reasonable concentrations. Virtual screening has, then, advanced to the point where it can distinguish between sensible and unlikely candidate ligands, but distinguishing among the former is a challenge it has yet to meet.

The most stringent test of docking, and its most useful possible outcome, is the accurate prediction of the binding affinities of a series of related compounds. This goal is essentially beyond all of the current docking methods, as the scatter plots in Warren et al.[10] make clear. Even the more limited goal of rank-ordering a list of compounds is beyond the field. Occasionally, a scoring function will work for a particular compound series for a particular target and advances can be made with that series. The β-secretase inhibitors studied by Moitessier et al.[7] are a good example, and when such examples are found, they should be tested aggressively. It is tempting to believe that the results of these relatively narrow studies suggest applicability of the scoring function to a broader set of data, but on the whole, reliable rank-ordering of hits from a diverse library remains inaccessible and is a key weakness with today's docking and scoring algorithms.

The Problem with Scoring

Why does docking remain so primitive that it is unable to even rank-order a hit list? Accurate prediction of binding affinities for a diverse set of molecules turns out to be genuinely difficult. At its simplest level, this is a problem of subtraction of large numbers, inaccurately calculated, to arrive at a small number. The large numbers are the interaction energy between the ligand and protein on one hand and the cost of bringing the two molecules out of solvent and into an intimate complex on the other hand. The result of this subtraction is the free energy of binding, the small number we most want to know.

Why are we unable to calculate interaction and solvation energies more accurately? The problem arises from the condensed phases in which biology occurs and the many degrees of freedom of biomolecules.[14] Were we concerned only with molecules in the gas phase, without the complexities of water, our calculations would be much simpler. Indeed, accurate calculations have been conducted in this phase for many years. The high symmetries of crystalline phases would also afford more accurate calculations, but this simplification is also denied us. In water, and with highly flexible proteins and ligands, accurate calculations are much more costly and error prone. Additionally, as pointed out by Tirado-Rives and Jorgensen,[6] the "window of activity" is very small. Thus, the free energy difference between the best ligand that one might reasonably expect to identify using virtual screening (potency, ∼50 nM) and the experimental detection limit (potency, ∼100 μM) is only about 4.5 kcal/mol. The free energy contributions due to

conformational factors alone for typical druglike ligands (which are usually neglected in most scoring functions) can be as large as this.[6] The authors conclude that with such a high uncertainty arising from only one of many terms, "consistent, successful ranking of diverse library members is inconceivable." Of course, more accurate methods may be considered. Among the most accurate today are thermodynamic integration/free energy perturbation methods, which can sometimes calculate the differences in affinities between related molecules to within "chemical accuracy," about 1 kcal/mol.[15,16] But even these methods only compare close analogues; they do not predict absolute binding affinities nor can they compare affinities among the diverse, unrelated molecules found in a typical screening library. They also demand so much computation time as to be infeasible for a large library. Recent reports suggest that progress is being made in calculating absolute binding affinities,[17,18] but these methods, too, remain too slow for docking screens, though they may be useful in rescoring docking hit lists.

For these reasons, much effort has been devoted to more empirical scoring functions. "Knowledge-based" scoring functions derive statistical potentials of mean force from large sets of protein−ligand complex structures. Such approaches were initially restricted by the limited amounts of data available, but the growth in the numbers of protein−ligand complex structures has enabled more accurate functions to be derived. Two papers in this section (from Muegge[8] and Yang et al.[9]) describe recent developments in this field that take advantage of this growth. Other groups have used regression and other QSAR techniques[19,20] to construct equations for predicting binding affinities. Traditionally, such models were derived solely from experimentally observed binding modes, but more recently "negative" data have been incorporated in an attempt to improve the predictability of such approaches.[4,21] The underlying rationale here, as pointed out by Pham and Jain,[4] is that the ligands in experimentally determined structures invariably have a good fit into the active site and contain few unfavorable interactions (such as steric clashes, same charge close contacts, and the burial of hydrophobic surfaces against hydrophilic ones). Consequently, the coefficients associated with such terms may not be weighted appropriately. In addition, rapid though the empirical scoring functions are, for certain applications such as de novo structure-based design, it may be more effective first to assess the output using scoring schemes that focus on the chemical connectivity, such as the method for complexity analysis described by Boda and Johnson.[5] Incremental progress is also being made in sampling molecular degrees of freedom. Protein flexibility is an active area of development, though most methods are restricted to side chain sampling or local relaxation. Some efforts to model the role of ordered water molecules have been made,[22,23] but this work, too, is at an early stage. As for more physics-based scoring functions, models for solvation energy costs and better relaxation of ligand−protein complex energies continue to advance[24−27] and will no doubt improve docking enrichment, though they will only take us so far toward the goal of rank-ordering hit lists.

What, then, is to be done?[28] Whereas we do not pretend to anticipate the progress of the field (it is the unexpected and fundamental advance to which we look forward most hopefully), there are certain obvious directions that might be usefully sketched.

Measurements and Benchmarks

Without major advances in fundamentals, docking must fall back on benchmarks and training sets for evaluating new algorithms that we know, from first principles, will be imperfect. Here, at least, one may expect progress. In its early days, the field was hampered by a lack of appropriate test sets of protein−ligand complexes. Even today, too few methods are tested extensively. At a minimum, new methods should be evaluated against a large enough number of targets to be representative of the available data as a whole. With the dramatic growth in the Protein Data Bank (PDB), current docking programs can be tested against hundreds of diverse protein−ligand complexes. Indeed, there are several curated test sets now available for this purpose (e.g., PDBbind,[29] BindingDB,[30] Binding MOAD,[31] CCDC/Astex[32]).

Unalloyed with affinities, structural data sets are insufficient for more robust algorithm development. The 1303-compound set in Warren et al.[10] comprised of groups of related compounds assayed using standard protocols begins to address the issue. Systematic, rigorous measurement of large compound sets is needed to test the strengths and weaknesses of docking methods. The best measurements will reflect K_d values. To get these, affinity must be measured directly or, if using substrate or ligand competition, the mechanism of binding must be determined. Without mechanism or direct binding, we are left with IC_{50} values that typically differ from K_d values and, in the worst case, can reflect artifacts.[33,34] Isothermal calorimetry measurements to separate contributions from enthalpy and entropy would also be valuable, though this technique remains costly in reagent. Preferably, at least some of these data sets would be "living" such that new compounds could be predicted, acquired (e.g., using fee-for-service chemistry arrangements), and tested, thus removing any inherent bias in how training and test sets are selected. It is also important to include micromolar affinity ligands and a good selection of likely decoys in such data sets. The better docking programs can distinguish nanomolar inhibitors from inactive molecules, but they cannot reliably do the same for micromolar inhibitors.

More fundamental compound measurements are also needed. Examples include transfer energies (for solvation and charges), dipoles/quadrupoles (for charge models), and pK_a values for both small molecules and proteins (for solvation and binding energy). Such measurements are certainly not as exciting as protein−ligand evaluations and will not uncover the next blockbuster drug, but their importance should not be overlooked. Building accurate scoring functions demands rigorous physical chemistry measurements for a wide variety of chemotypes.

Case Studies

Most docking studies continue to be retrospective, demonstrating that the particular technique can reproduce observed results in a published data set. Too few studies are published that describe prospective studies, particularly from industry where the opportunity for such work is the greatest. Whereas we understand the need to protect intellectual property, publication of these studies would help the field, and there certainly must be examples where release of data would not endanger active projects. Moreover, unsuccessful studies are rarely published and this type of negative data could be equally informative. Mechanisms to make the results of both positive and negative studies more widely available would facilitate the development of new and improved algorithms.

Methods Development

One can always hope that incremental improvements in current techniques will gradually lead to major advances. Such efforts are sensible, but they cannot be the only strategy; there

is a call for more adventurous departures than are being published. For scoring in particular, the gap between what is required and the current methods is large. We do not presume to recommend specific research strategies except to encourage bold leaps into new areas, as well as the small hops that many, ourselves included, have typically taken.

Moving Forward

In August 2005, the NIH sponsored a workshop where participants from academics, industry, and government met to identify the impediments to faster progress in docking and scoring.[35] One recommendation that emerged from this group was to foster alliances among pharmaceutical, government, and academic researchers at a higher level than currently experienced. A second NIH sponsored meeting was held in February 2006 to consider what each sector could offer and how they might collaborate.[36] Investigators subsequently met at NIST in April 2006 to address experimental measurements that could contribute to docking/scoring technology.[37]

The outcomes of these meetings were presented to the National Institute of General Medical Sciences Council (NIGMS) in May 2006,[38] and mechanisms for supporting these recommendations are now being explored. A key contribution from industrial groups would be large SAR series made up of multiple ligand structures and affinities for multiple targets. Whereas much of these data are currently proprietary, many of the targets for which they were generated are no longer of therapeutic interest and could be releasable under the right structure. Federal agencies, and NIST in particular, are interested in accurate compound measurements. Academic and software company researchers are committed to improving the technology. With all of this in mind, the primary recommendation to the NIGMS Council is the development of a national data resource with which all researchers can interact. A second recommendation was to empower such a center to make new experimental measurements, including compound characterization, protein−ligand affinity, and protein−ligand complex crystal structures. Such measurements are critical for testing new methods prospectively. Finally, well-developed testing sets must be evaluated with all available technology, without barriers, if we are to see forward rather than lateral growth in the field.

Twenty years ago, docking screens were widely considered at best a gentlemanly pursuit, unlikely to affect real drug discovery, and at worst a fool's errand. No ligand had then been predicted; getting geometries correct was difficult; scoring functions used steric fit alone; molecular flexibility was not modeled; and there were hardly any interesting targets to dock against. Today, multiple novel ligands have been predicted and confirmed by experiment, even to atomic resolution. Docking routinely treats ligand flexibility and typically includes some receptor plasticity, and scoring functions include most of the terms in molecular mechanics force fields. We are confronted by an embarrassment of riches in protein structures. Docking is now used by almost every major pharmaceutical company. But it is also true that docking seems to have reached a plateau and is waiting for an important breakthrough. Like most scientific breakthroughs, such an advance is hard to predict and might well result from idiosyncratic efforts in small research groups. This is not the only way forward in docking, however. There is also room for more of an engineering approach, and like many engineering efforts there is a call for a higher level of coordinated effort in the field. Such efforts must partner industry, which can contribute data sets, academics and commercial developers who are working on the underlying methods, and government. Indeed, the curated data and benchmarking sets, and ideally prospective measurements, that would emerge would underpin both the engineering efforts and fundamental advances, in a mixed model of individual and collective research. From such efforts we might expect significant advances in docking and scoring.

Acknowledgment. We thank the authors who have contributed manuscripts for this section of the Journal: Ajay Jain, Peter Johnson, Bill Jorgensen, Nicolas Moitessier, Ingo Muegge, Shaomeng Wang, and Greg Warren. We also acknowledge the efforts of those involved in the discussions with NIH and NIST. B.K.S. thanks NIH Grant GM59957 for support.

Biographies

Andrew R. Leach received his Ph.D. in 1989 at Oxford University under the direction of Keith Prout in the field of computational approaches to conformational analysis. Following postdoctoral studies in protein−ligand interactions and macromolecular simulations with Bob Langridge, Tack Kuntz, and Peter Kollman at University of California−San Francisco, he accepted an advanced fellowship at Southampton University in 1991. In 1994 he joined Glaxo Group Research and is currently Director of Computational Chemistry U.K. for GlaxoSmithKline. Dr. Leach is an Editor-in-Chief for the *Journal of Computer-Aided Molecular Design.* He has a long-standing interest in computational chemistry, cheminformatics, and scientific education in these fields.

Brian K. Shoichet received his Ph.D. for work with Tack Kuntz on molecular docking in 1991 from University of California−San Francisco. His postdoctoral research was largely experimental, focusing on protein structure and stability with Brian Matthews in Eugene, OR. Shoichet joined the faculty at Northwestern University in the Department of Molecular Pharmacology in 1996 and received tenure in 2002, only one year after his younger sister, Molly Shoichet. (Shoichet denies any sensitivity around this issue.) Around that time he was recruited back to University of California−San Francisco, where he is now a Professor of Pharmaceutical Chemistry. "We confused him with Kevan Shokat", admits a member of the recruiting committee. Research in the Shoichet laboratory uses computational and experimental techniques to investigate enzyme inhibition, structure, and function. It is supported by the NIH.

Catherine E. Peishoff received her Ph.D. in 1984 from Purdue University under the direction of William L. Jorgensen. Immediately after receiving her degree, she started her career in the pharmaceutical industry, first at Lederle Laboratories (now part of Wyeth) and beginning in 1987 at SmithKline Beckman (now GlaxoSmithKline Pharmaceuticals). She is currently the Philadelphia, PA, Director for Computational, Analytical, and Structural Sciences. Dr. Peishoff is a Senior Editor for the *Journal of Medicinal Chemistry* and has interests in the development and application of computational chemistry methods for medicinal chemistry.

References

(1) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorth, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.
(2) Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. Docking: Success and challenges. *Curr. Pharm. Des.* **2005**, *11*, 323−333.
(3) Krovat, E. M.; Steindl, T.; Langer, T. Recent advances in docking and scoring. *Curr. Computer-Aided Drug Des.* **2005**, *1*, 93−102.
(4) Pham, T. A.; Jain, A. J. Parameter estimation for scoring protein−ligand interactions using negative training data. *J. Med. Chem.* **2006**, *49*, 5856−5868.
(5) Boda, K.; Johnson, A. P. Molecular complexity analysis of de novo designed ligands. *J. Med. Chem.* **2006**, *49*, 5869−5879.
(6) Tirado-Rives, J.; Jorgensen, W. L. Contribution of conformer focusing to the uncertainty in predicting free energies for protein−ligand binding. *J. Med. Chem.* **2006**, *49*, 5880−5884.
(7) Moitessier, N,; Therrien, E.; Hanessian, S. A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptide and pseudopeptidic *β*-secretase (BACE 1) inhibitors *J. Med. Chem.* **2006**, *49*, 5885−5894.

(8) Muegge, I. PMF scoring revisited. *J. Med. Chem.* **2006**, *49*, 5895−5902.
(9) Yang, C.-Y.; Wang, R.; Wang, S. M-Score: A knowledge-based potential scoring function accounting for protein atom mobility. *J. Med. Chem.* **2006**, *49*, 5903−5911.
(10) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−5931.
(11) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: Comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.
(12) Irwin, J. J.; Shoichet, B. K. ZINC, a free database of commercially available compounds for virtual screening *J. Chem. Inf. Model.* **2005**, *45*, 177−182.
(13) Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **2005**, *48*, 962−976.
(14) van Gunsteren, W. F.; Berendsen, H. J. C. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992−1023.
(15) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167−4180.
(16) Pearlman, D. A. Evaluating the molecular mechanics Poisson−Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796−7807.
(17) Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.
(18) Chang, C. E.; Gilson, M. K. Free energy, entropy, and induced fit in host−guest recognition: calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.
(19) Bohm, H.-J. 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.
(20) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. 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.
(21) Smith, R.; Hubbard, R. E.; Gschwend, D. A.; Leach, A. R.; Good, A. C. Analysis and optimization of structure-based virtual screening protocols (3). New methods and old problems in scoring function design. *J. Mol. Graphics Modell.* **2003**, *22*, 41−53.
(22) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein−ligand docking predictions. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 17−28.
(23) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.
(24) Cerutti, D. S,; Jain, T,; McCammon, J. A. CIRSE: a solvation energy estimator compatible with flexible protein docking and design applications. *Protein Sci.* **2006**, *15*, 1579−1596.
(25) Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. Virtual screening against highly charged active sites: identifying substrates of alpha-beta barrel enzymes. *Biochemistry* **2005**, *44*, 2059−2071.
(26) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., 3rd. Assessing scoring functions for protein−ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.
(27) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **2002**, *322*, 339−355.
(28) Lenin, V. I. What Is To Be Done? *Lenin: Collected Works*, 4th ed. (English); Progress Publishers: Moscow, 1973; Vol. V, pp 375−376, 451−453, 464−467.
(29) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein−ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977−2980. Also, see www.pdbbind.org.
(30) Chen, X.; Lin, Y.; Gilson, M. K. The Binding database: overview and user's guide. *Biopolym. Nucleic Acid Sci.* **2002**, *61*, 127−141. Also, see www.bindingdb.org/bind/index.jsp.
(31) Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (mother of all databases). *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 333−340. Also, see www.BindingMOAD.org.
(32) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein−ligand interaction. *Proteins* **2002**, *49*, 457−471. Also, see www.c-cdc.cam.ac.uk/products/life_sciences/validate/astex/.
(33) McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **2002**, *45*, 1712−1722.
(34) Shoichet, B. K. Screening in a spirit haunted world. *Drug Discovery Today*. **2006**, 607−615.
(35) A workshop report is available at http://www.nigms.nih.gov/News/Reports/DockingWorkshop082505.htm.
(36) A workshop report is available at http://www.nigms.nih.gov/News/Reports/DockingMeeting022406.htm.
(37) A workshop report will be posted at http://www.boulder.nist.gov/div838/.
(38) Meeting minutes are available at http://www.nigms.nih.gov/About/Council/Minutes/May18−19_2006.htm.

JM060999M