Available online at www.sciencedirect.com
ScienceDirect
Focus Article, *Mendeleev Commun.*, 2009, **19**, 237–242

Mendeleev Communications

Molecular docking: theoretical background, practical applications and perspectives

Fedor N. Novikov[a,b] and Ghermes G. Chilov*[a,c]

[a] *MolTech Ltd., 119992 Moscow, Russian Federation*
[b] *Department of Bioengineering and Bioinformatics, M. V. Lomonosov Moscow State University, 119991 Moscow, Russian Federation*
[c] *N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation. Fax: +7 499 135 5328; e-mail: Ghermes@moltech.ru*

DOI: 10.1016/j.mencom.2009.09.001

Molecular docking is one of the key computational chemistry techniques that are routinely applied to drug discovery. The holy grail of molecular docking is to replace experimental studies of protein-ligand complexes by modeling their structures and binding affinities *in silico*. However, current practical achievements of docking suggest that approaching experimental accuracy with computations is a big challenge for theoretical chemistry.

From the practical point of view, molecular docking is a computational chemistry tool that has a clear intuitive definition of finding the structure and binding energy of a protein-ligand complex when the spatial structures of the protein and the ligand are known (Figure 1). With an instrument that can correctly and reliably predict protein-ligand structure, one could easily identify the right molecules to modulate functions of the desired protein, for example, a therapeutic target. Obviously, the world of drug discovery and biotechnological applications would have been very different if such instruments were available.

Pioneering research works in the area of molecular docking date back into the early 1980s.[1] However, it took at least a decade for this technology to become popular among computational chemists and pharmaceutical researchers. In the early 1990s, several factors contributed to the mainstream acceptance of molecular docking.[2] First, large molecular libraries were synthesized as a result of the development of combinatorial chemistry methods. Second, the target-based drug discovery paradigm was widely accepted. Molecular entities were screened in search of those selectively interacting with characterized protein targets important in therapeutic treatment of a disease. Third, computational resources had become a widely available commodity.

However, by the middle 1990s, it had become evident that the molecular modeling methods were associated with a high degree of uncertainty and they could not substitute or even reduce the need for experimental studies.[2] As a result of the shifting focus, the development of novel molecular modeling methods and molecular docking methods in particular had lost steam and by and large migrated into the academic domain. At the same time, it had become clear that the development of robust molecular modeling techniques is a task that involves significantly higher than expected degree of complexity and the one that requires breakthroughs in chemical and computational science.

Theoretical background and challenges of docking

Molecular docking typically involves approaching two interrelated tasks: (1) sampling of possible conformational states of the protein-ligand complex and (2) calculation of the free energy of such complexes or producing a score that correlates with biological activity or other function, also known as scoring. Mathematically, the sampling algorithm returns protein-ligand conformations populating the search space for the global minimum of a protein-ligand binding energy, while the scoring function defines the potential energy surface on which the optimization takes place. Both sampling and scoring are crucial since deficient sampling might overlook the best candidates in the search space, while deficient scoring would distort the potential energy surface and lead to wrong structure and energy estimations.

Correct sampling of the search space remains a very significant challenge despite major advances in the mathematical science and the development of global optimization methods. The main issue is that the search space is so extremely large that it is not currently possible to even estimate the order of

Ghermes G. Chilov received his M.S. in Chemistry (2000) and Ph.D. in Chemistry (2003) from M. V. Lomonosov Moscow State University. Since 2007 he heads MolTech Ltd. – a small innovative company developing molecular modeling software for pharmaceutical and biotechnology applications. His scientific interests include molecular modeling, biocatalysis and drug discovery.

Fedor N. Novikov received his M.S. in Chemistry (2006) from M. V. Lomonosov Moscow State University. Since 2007 he works at MolTech Ltd. His scientific interests include molecular modeling, computational sciences, biocatalysis and drug discovery.

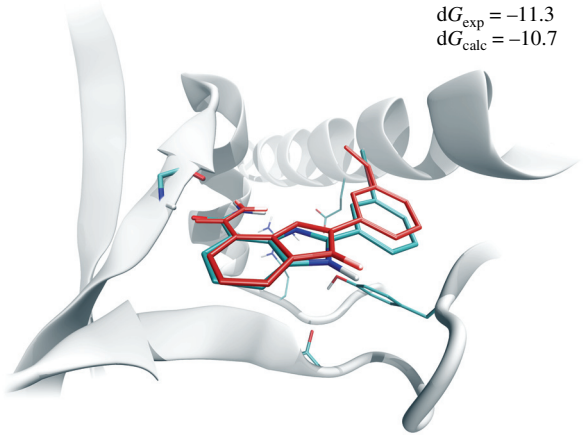

Figure 1 An illustration of the molecular docking applications: modeling structure and binding energy of the enzyme [poly-(ADP-ribose)-polymerase] complex with one of its inhibitors. The inhibitor position obtained by docking is coloured by atom name in gray scale. The crystallographic position (PDB structure 1efy) is coloured in dark gray. Calculations were performed with Lead Finder docking software.

magnitude of the number of points in the conformational search space. The search space is multi-dimensional where the additional degrees of freedom originating from the internal rotation along freely rotatable bonds and flexibility of ring structures in the ligand molecule, as well as the dislocation of certain side chains and the flexibility of protein structure[3] must be considered along with the classical translational and rotational degrees of freedom. However, these are not the only issues that require consideration in molecular docking. Water molecules, solvation and desolvation effects play an important role in the interaction between protein and ligand structures and so far, there have been few studies of the role of water molecules and the additional degrees of freedom brought by the presence or absence of water molecules in the active center and the surrounding space. Clearly, intelligent handling of the effects induced by water molecules is crucial for the accuracy of molecular docking methods.[4]

So far, a number of modern mathematical approaches have been implemented in molecular docking software programs, such as: Monte Carlo algorithms (ICM,[5] GlamDock[6]), genetic algorithms (AutoDock,[7] Fitted,[8] Gold[9,10] MolDock[11]), incremental construction of an optimal ligand pose (DOCK,[12] FlexX,[13,14] Surflex[15]), systematic analysis of possible minima using graph searches (eHits[16,17]), algorithms using hierarchical scoring functions for crude shape fitting and finer optimization of ligand pose (QXP,[18] LigandFit,[19] Glide[20,21]).

Our docking software, Lead Finder, attempts to achieve a fine balance between the genetic algorithm and the multilevel local optimizations.[22] While analyzing the applied methods implementing the classical genetic algorithm we have identified a significant gap between the mathematical theories and their specific applications, including molecular docking. The classical genetic algorithm stipulates that the conformation of the protein-ligand complex with the lowest energy value in a given round of evolution must produce the greatest number of offspring for the next round of evolution. As a result, one 'good' conformation can endanger the survival and sometimes make extinct the populations of individuals and that, in turn, can dramatically reduce the search space and result in a premature termination of the search for a global minimum. This weakness of the genetic algorithm has long been known. What has been for the most part overlooked, however, is that the excessive accumulation of 'good' conformations leads to the reduction of sampling efficiency. New rounds of evolution bring about large numbers of individuals that are quite similar and, therefore, the computing power is used less efficiently in processing large sets of individuals with reduced diversity. In other words, there is a degree of redundancy in calculations when new rounds of evolution produce similar individuals.

We have developed two methods to alleviate this weakness of the algorithm. First, similar conformations are grouped into clusters of a certain limited size. If the size of a cluster exceeds the allowable limit, the redundant conformations are discarded. This approach helps maintain the genetic diversity of a population. Second, as soon as new rounds of evolution stop bringing conformations with substantially better properties for a given cluster, such cluster gets excluded from further rounds of evolution and other clusters with less favourable characteristics take the vacated space. The excluded clusters can still carry their genes such as torsional angles and translational and rotational positions with some degree of probability to the individuals who still participate in the evolutionary process. This approach, however astonishing or not logical it may seem initially, allowed us to substantially improve conformational sampling by increasing the probability of finding the global minimum in a shorter time.

In addition to these adaptations of the genetic algorithm, we have developed a number of local optimization techniques that allowed us to optimize ligand poses obtained from the genetic evolution with either higher accuracy but slower method, or lower accuracy but faster method. Varying and balancing the accuracy *versus* the speed of calculations at different stages of the genetic evolution proved to be an important and useful technique to achieve a higher degree of genetic diversity, a higher degree of optimization and a faster speed of calculations.

Correct scoring of generated ligand poses also carries very significant scientific challenges. What the ideal scoring function must be able to do? First, it must produce correct ranking of the free energy of protein-ligand complexes. The correct conformation, *i.e.*, the one that is experimentally observed in a protein-ligand complex by X-ray or other techniques, must have the lowest energy ranking. Second, the predicted free energy of the best-ranked conformation must be close to the experimentally observed value, such as the one derived from an experimentally measured binding constant. Finally, the scoring function must be able to tell apart the non-binding ligands whose binding constant cannot be measured experimentally. This last requirement by itself has a substantial practical value. Experimental measurement of the binding constant is usually a laborious and costly procedure that is performed on small sets of compounds while screening can be performed on hundreds of thousand and millions of compounds. Therefore, it is important to be able to predict which compounds will likely bind to a target and which will not. However, with the typical binding energy of 8–12 kcal mol^{-1} in the case of non-covalently bound ligands,[23] the errors in quantum chemistry single point energy calculations of 1–2 kcal mol^{-1} can make a dramatic impact on the ability of a molecular modeling method to separate binding ligands from non-binding. We should mention that the high level *ab initio* calculations remain not computationally feasible for protein-ligand complexes and will remain such for a long time.

A retrospective look in the previous research works reveals one notable fact: the development of various approaches in scoring algorithms does not necessarily addresses the main practical applications of this algorithm, such as: pose ranking, binding energy calculation and discrimination between active and inactive ligands. As far as we are aware, to this date Lead Finder remains the only software program that employs scoring functions specifically optimized for the practical applications listed above. Lead Finder's scoring functions are based on a semi-empiric molecular mechanical functional and they explicitly

account for a common set of different types of interactions (*e.g.*, van der Waals, electrostatic, hydrogen bonds, *etc.*). Then, three distinct sets of energy-scaling coefficients are used to optimize predictions for a particular purpose. The three sets of scaling coefficients have been fine-tuned for: (1) correct ranking of docked ligand poses, (2) accurate binding energy prediction, and (3) correct rank-ordering of active and inactive compounds in virtual screening experiments. For each task, a special training data set was used: for ligand pose ranking – a vast set of protein–ligand complexes with known 3D structures; for binding energy calculations – a set of protein ligand complexes with known 3D structures and experimentally measured ligand affinity; for discriminating active ligands from decoys – a set of proteins with their known ligands and a common set of decoy compounds. We are hopeful that the approach of constructing specialized scoring functions for solving particular problems will gain wider acceptance in the future.

If we examine the types of energy considered in the predictions of molecular interactions, the existing scoring functions can be divided into three main classes. The first class is the empirical scoring functions that employ various descriptors, often derived from the molecular geometry and the common chemical sense, to estimate properties of a protein-ligand complex. The empirical scoring function methods resemble QSAR in their algorithmic approaches and applications. The empirical scoring functions were the first to be developed, and it is not surprising that a large variety of such functions are currently available[24] since the introduction of new descriptors is only limited by the imagination of authors.

Unlike the case with empirical scoring functions, the scoring functions that are based on the molecular mechanics force field interactions forming the second class of scoring functions are strictly limited by the types of energy terms that constitute the particular force field. The development and validation of a force field interaction is a costly and time-consuming procedure, and as a result there are only a few force field interactions in use today and even fewer scoring functions that are based on those.[4]

Finally, the third class is formed by the hybrid scoring functions. The hybrid functions combine the energy terms characteristic of the molecular mechanics force field interactions with arbitrary descriptors. The hybrid functions frequently employ scaling coefficients for faster and more flexible accounting of the energy terms, albeit deviating from the first principles.

The hybrid scoring functions, while being simpler and more computationally feasible than the force-field based functions, often produce more accurate results. This observation suggests that obtaining reliable results in docking by the molecular mechanics methods is not possible without the explicit accounting for all participants of the protein-ligand binding process, including water molecules. Addressing this deficiency is the primary objective of the empirical components of the force-field based scoring functions. For example, Lead Finder uses a scoring function that employs a number of empirical adjustments to account for the solvation effects. One of such adjustments is the screened coulomb potential[25] model that accounts for the electrostatic interactions relative to their microenvironment, *i.e.*, hydrophilicity and the degree of exposure to the solvent environment. Also, we introduced explicit penalty terms for the cases when the protein and ligand atoms that are capable of forming hydrogen bonds in solvent do not form such bonds between themselves in the bound state and when they are sterically not available for forming hydrogen bonds with water molecules. Lastly, in addition to the standard model of volume-based solvation[26] that provides a computationally effective mechanism of calculation of the degree of overlap between the hydrophobic protein and ligand atoms, we introduced four additional components in order to account for the contributions of the contact areas between polar/non-polar protein and ligand atoms with the surrounding water. Those components allowed us to make a necessary correction for the solvation state of the ligand in the protein-ligand complex *versus* its non-bound state in water.

In which areas one can now expect further breakthroughs in the pursuit of accuracy in docking predictions? In our opinion, the solvation models are on the verge of significant innovations that can be achieved either by more extensive sampling that would include sampling of water molecules and, therefore, would necessitate greater computing power, or by the appearance of new statistical mechanical theories that would describe changes in a surrounding aqueous environment without the need to account for the individual water molecules. At present, attempts to account for the individual structurally conserved water molecules in the docking process have not yet resulted in a notable increase in the accuracy of docking predictions.[4,27–29]

The protein flexibility deserves special attention. While some docking software, such as Glide combined with Prime,[30] AutoDock,[31] Fitted[24] support the protein flexibility model, one must be cautious with expectations that such a model would result in the increased docking accuracy. The protein flexibility model brings the need for additional conformational sampling and that in the end may lead to the reduction of docking accuracy, especially when amplified by the increase in complexity of the global search algorithm.

The need to increase accuracy of the force field calculations used in docking methods remains an open question. While it is obvious that the increase of accuracy of force field interactions would be beneficial, deficiencies of other components of the process may render the degree of accuracy of force field calculations irrelevant. The efforts to optimize partial atomic charges on ligand atoms with QMMM have demonstrated that, for the majority of studied structures, the optimization of charges does not result in any notable increase in the scoring accuracy while the computational cost increased dramatically.[34]

Consequently, it is important to look for additional empirical techniques that could reduce the possibility of errors at various stages of the docking process. For example, one of the most nagging problems in scoring has been the correct discrimination between active and inactive ligands, since true active compounds typically constitute a very small percentage of a large library of compounds. One obvious approach to deal with this is to construct target-specific scoring functions that will be trained to discriminate active ligands from inactive ones for a given target protein.[32] Customization of the scoring function can be achieved by optimizing its energy scaling coefficients with respect to the virtual screening efficacy and can be routinely performed for different proteins.[33] However, note that customization is rather a scoring function-centered approach and its capacity of improvement is heavily limited by the quality of original scoring function. Recently, we have identified an alternative approach to screen out inactive compounds that is based on the detection of certain specific protein-ligand interactions characteristic of a given protein. If a set of specific interactions (*e.g.*, H-bonds) is observed for all known complexes of a particular protein with its bound ligands, then we anticipate the same interactions must be present in a ligand recovered in virtual screening (Figure 2). If those interactions are not present, the ligand is discarded. Obviously, structural filtration is a structure-centered approach and it does not depend on the scoring function and even the docking program used. We successfully applied the post-docking structural filtration of docked ligand poses to achieve multi-fold increase of virtual screening efficiency for such enzyme as poly-(ADP-ribose)-polymerase.[35] Similar ideas, such as Interaction Fingerprint (IFP) that looks at the entire list

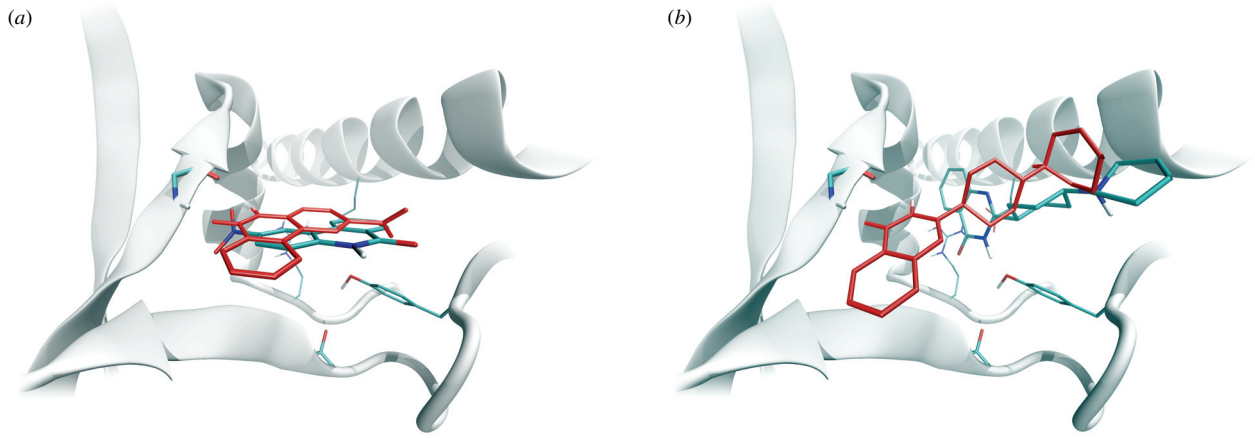

Figure 2 Illustrations of (*a*) scoring and (*b*) sampling errors in docking. (*a*) The top-scored ligand pose (coloured by atom name) received a better energy than the correct pose (coloured in red). (*b*) The crystallographic ligand pose (coloured in red) received a better energy than the top-scored pose produced by docking (coloured by atom name). Calculations were performed with the Lead Finder docking software.[35]

of all inter-atom contacts instead of certain specific interactions, have also been successfully applied to increase enrichment in virtual screening experiments.[36,37]

Performance of docking approaches

Benchmarking studies of the performance of docking software programs, calculation protocols, sets of protein-ligand complexes used for testing and training, *etc.* are abundant. However, the results and conclusions of such studies have not been universally consistent. In many cases, benchmarking studies were performed in incompatible environments or with incompatible criteria, therefore making it impossible to compare the results of such studies side-by-side.

In our opinion, at least three criteria are universally applicable to the evaluation of accuracy of docking methods. First, it is the accuracy in predicting the ligand pose. Second, it is the calculation of the free energy of protein-ligand complex. Third, it is the ability to distinguish between active and inactive compounds in virtual screening experiments. The third criterion is distinctly different from the second one, since the free energy of binding can be experimentally verified only for those protein-ligand complexes where a considerable degree of ligand binding exists. In virtual screening experiments, the vast majority of ligands usually do not demonstrate any noticeable degree of binding; therefore, the free energy of binding cannot be evaluated for such ligands.

The commonly accepted criterion of docking success rate is the percentage of protein-ligand complexes that differ from the experimentally measured structures by no more than a certain value of rmsd, typically 2 Å. Obviously, such benchmark heavily relies upon the input data – the set of protein-ligand complexes with known 3D structure. One can easily find about ten different test sets of protein-ligand complexes used by the docking software developers to benchmark their programs, and even more sets from end users. The differences between these test sets make the side-by-side comparison of docking programs impossible. In order to correctly compare the performance of Lead Finder side-by-side with other docking software programs, we took the overlapping structures from the published test sets of the most recognized programs such as FlexX, Glide SP, Glide XP, Gold, LigandFit, MolDock, and Surflex, where docking success rate data were provided in original publications.[25] The use of such an overlapping set of structures allowed us to compare the performance of Lead Finder with those programs side-by-side (Table 1). We hope that, in the future, we will see the appearance of standard test sets to facilitate consistent benchmarking studies. Data sharing and online resources, including protein and ligand structure data that is ready for docking, should help this process.

The classification of docking errors is another aspect that is important for molecular modelers who want to make the fullest use of the available technology, and for technology developers who aim at increasing accuracy of their software. If the docked protein-ligand complex with the lowest free energy has the ligand pose that is significantly different (by more than 2 Å) from the correct ligand pose, then it is a scoring error [Figure 2(*a*)]. In other words, scoring errors exemplify situations when the correct ligand pose does not receive the lowest energy. When the software does not find the correct ligand pose but correctly ranks the poses it finds by their free energy, this type of error is called a sampling error [Figure 2(*b*)]. From the practical viewpoint, such analysis of docking errors allows a user to understand strengths and weaknesses of a particular docking method and make appropriate decisions regarding the choice of a particular docking method in various practical situations. For example, a user may choose to run a method that provides the best sampling technology to find correct docked poses and then choose a method with the best scoring technology for ranking the docked poses. Also, knowing the strengths and weaknesses of a docking method allows the user to fine-tuning the algorithm when the software allows such fine-tuning.

Unfortunately, few attempts have been made to analyze and classify docking errors.[38] In our view, the analysis of docking

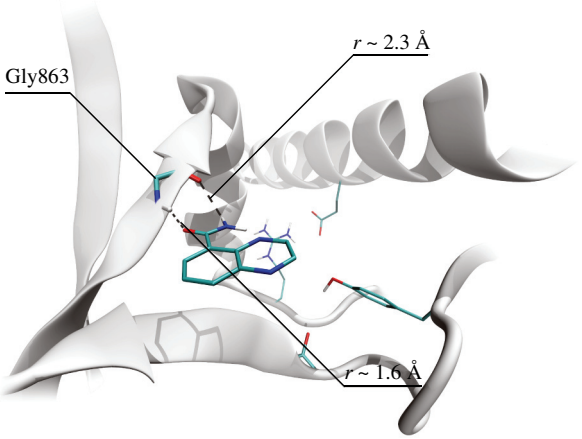

Figure 3 Illustration of the principle of structural filtration of the docked ligand poses to improve recognition of the true binding ligands. We have proven the two structurally conservative hydrogen bonds are crucial for the inhibitor recognition by poly-(ADP-ribose)-polymerase.[35]

Table 1 Docking success rates (%) of different programs obtained on their native test sets.

Program	Original data	Lead Finder	Number of structures in the test set
FlexX[14]	46.5	85.0	200
Glide SP[20]	70.2	82.3	282
Glide XP[21]	69.4	81.3	268
Gold[9]	72.4	87.3	134
Gold[10]	76.5	90.6	85
MolDock[11]	87.0	96.1	77
Surflex[15]	70.4	96.3	81
All test sets	—	85.0	407

errors brings substantial value to both users and developers of docking software. In our case, the analysis of errors produced by Lead Finder allowed us not only to identify the priority directions for further development of our docking algorithms, but also provided us with valuable insights into performing reliable benchmarking studies.

The selection of an appropriate test set is one of the most considerable challenges in defining the test environment for evaluating the accuracy in prediction of the free energy of protein-ligand complex. While Protein Data Bank (PDB) provides extensive data on the structure of protein-ligand complexes, the experimental data on the protein-ligand binding constant is very scarce. A few research groups have attempted compilation of the data on the protein-ligand binding constants.[39–41] However, (*a*) the data collection protocols have not been standardized and in some cases are simply absent, and (*b*) the data are not being curated and frequently contain errors and omissions. The standard test sets for assessing performance in prediction of the free energy of protein-ligand binding do not exist at present.

We should note that while working on the selection of data for a test set for benchmarking Lead Finder performance we tried to verify the experimental data before adding it to the test set. In total, we collected 330 protein-ligand complexes with experimentally determined binding constants and 3D structures. The RMSD between the calculated binding energies and the experimentally determined values was found to be 1.50 kcal mol^{-1} for the entire set of 330 protein-ligand complexes.[22] Another interesting initiative of compiling an extensive test set for benchmarking scoring function accuracy is carried out by the Scoring Function Consortium, a collaboration body formed by several pharmaceutical companies providing access to their internal data on the binding constants for about 855 protein-ligand complexes for the purposes of development and validation of novel empirical scoring functions.[42] Unfortunately, access to the data is limited by the descriptor sets for protein and ligand atoms and the coordinates of atoms are not yet provided. We hope that in the future, such kind of information will be made available to the public, and that eventually there will be a curated analogue of the PDB for the protein-ligand affinity data.

Finally, the accuracy of virtual screening is probably one of the most practically important characteristics of docking methods; however, a variety of factors make comparisons and performance benchmarking of different docking methods in virtual screening anything but straightforward. In a typical virtual screening benchmarking study, a set of active ligands for a given protein is mixed with a relatively large set, also known as a library, of inactive (decoy) ligands. Then the docking software assigns numeric score values to each ligand in the entire set and the ligands are ranked by their score values. Finally, one or another quantitative metric is applied to the rank-ordered list of ligands in order to assess the ability of the docking program to accumulate active ligands near the top of the rank-ordered list, *i.e.,* discriminate active ligands from decoys. Such benchmarking studies are usually quite sensitive to the protein models used and the specific characteristics of ligand libraries. By taking a different protein model and a different set of ligands, one may obtain significantly different benchmarking results. Recognizing this issue, the methodological aspects of virtual screening benchmarking studies have been recently considered,[43] however the acceptance of these ideas by software developers and users has yet to come.

Traditional applications of docking in drug discovery

Traditionally, docking is being viewed as a tool to search for active ligands, also known as hits, for a given protein from a particular library of compounds.[44] The process of hit finding is followed by the optimization of their structures in order to obtain more potent compounds, also known as leads. At this stage, hundreds or even thousands of novel compounds are synthesized using the core chemical scaffolds of the found hits. Obviously, docking studies may provide clearer directions in the hit-to-lead optimization. However, lead optimization demands higher accuracy in scoring, since in this case the difference between active and 'very' active compounds is more subtle than the difference between active and inactive compounds. In their present state, the docking methods appear to be better suited for qualitative rather quantitative assessments in the lead optimization process. One of the recent studies conducted by pharmaceutical industry researchers has demonstrated the absence of a correlation between the experimental and calculated binding affinities for a set of well studied therapeutic targets.[45]

The *de novo* drug design is a relatively new application of molecular docking. Novel active compounds are not looked up in the databases of chemical structures but rather constructed from scratch using atoms or fragments. While the idea of the *de novo* drug design is a several decades-old idea, it has found practical applications relatively recently following the development of the fragment-based drug design approaches.[46] In the fragment-based drug design, a new compound is assembled from relatively small molecular entities ($M_r < 250$) rather than from atoms or small chemical groups. Fragments are designed to occupy distinct non-overlapping sub-sites on the targeted patch of the protein exterior. By cross-linking such fragments, in the ideal situation one can obtain a molecule whose affinity is a sum of the affinities of its fragment components. It is worth mentioning the reasons why the fragment-based approaches became so widely popular (according to recent studies, about 30% of all drug discovery projects are fragment-based).[47] On the one hand, advances in the experimental techniques such as NMR spectroscopy facilitated the development of fast high-throughput determination of the ligand interaction with a particular protein site. On the other hand, the number of molecules representing the same diversity of molecular properties (*e.g.*, the 3D-shape) is dramatically smaller for fragments than it is for the bigger molecules. That is why a relatively small library of fragments can yield hits with a higher probability. Obviously, successful docking and scoring of fragment-like molecules, whose affinities are in the milli- to submicromolar range, is not possible without a major revision of the scoring approaches that are presently designed to simulate binding of the drug-like molecules ($M_r \sim 500$) and the calculation of binding constants in the micro- to nanomolar range. This step in the computational chemistry science is yet to be done.

Novel applications of docking

With the impressive progress in the structural and functional genomics, it now becomes possible to solve the reverse task of virtual screening, namely, finding a protein that will specifically bind a given compound.[48] The *in silico* discovery of a target protein gives medicinal chemists a clue for the structure-based

drug design. Second, finding a therapeutic target can help in deciphering the molecular mechanism of drug action when it is unknown (which is not a rare situation). Finally, finding novel targets to the well studied compounds (*e.g.*, generic drugs) provides an opportunity for the development of novel therapeutic applications to the known drugs. The latter option is of high interest for the pharmaceutical industry, since clinical trials of a well studied drug are much less risky than those for a new compound.

Another class of practically important applications for molecular docking is the rational design of proteins with given properties: design of an enzyme selective for a particular substrate,[49] or design of a protein to achieve a more potent interaction with another protein (ligand-receptor, antigen-antibody, or another type of interaction).[50] The optimization of such properties of proteins is crucially important for the biopharmaceutical industry. In principle, the task of protein design can be solved by molecular docking methods: for each possible mutant protein, conformations of the mutated aminoacids can be sampled to find the energy minimum of the system (for example, an enzyme-substrate or a protein-protein complex), and then the optimum solution can be chosen from those mutant proteins. The existing computational approaches in the biopharmaceutical industry often use the simplified scoring functions that consider only the steric overlaps and the statistical weights of various amino acid rotamers taken from a database of experimentally observed rotamers.[51] Furthermore, in real life the application of these computational approaches is limited to such tasks as modeling protein structure by homology or the refinement of protein structures obtained through the low resolution X-ray diffraction methods. Hopefully, the increase in accuracy of scoring functions will bring new important practical applications for the computational methods.

References

1 I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge and T. E. Ferrin, *J. Mol. Biol.*, 1982, **161**, 269.
2 A. B. Richon, *Drug Discovery Today*, 2008, **13**, 665.
3 S. J. Teague, *Nat. Rev. Drug Discovery*, 2003, **2**, 527.
4 N. Huang and B. K. Shoichet, *J. Med. Chem.*, 2008, **51**, 4862.
5 M. Totrov and R. Abagyan, *Proteins: Struct. Funct. Genet.*, 1997, **29**, 215.
6 S. T. Apostolakis and J. Glam, *J. Chem. Inf. Model.*, 2007, **47**, 1657.
7 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, 1998, **19**, 1639.
8 C. R. Corbeil, P. Englebienne and N. Moitessier, *J. Chem. Inf. Model.*, 2007, **47**, 435.
9 G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol.*, 1997, **267**, 727.
10 M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson and C. W. Murray, *J. Med. Chem.*, 2007, **50**, 726.
11 R. Thomsen and M. H. Christensen, *J. Med. Chem.*, 2006, **49**, 3315.
12 C. M. Oshiro, I. D. Kuntz and J. S. Dixon, *J. Comput. Aided Mol. Des.*, 1995, **9**, 113.
13 M. Rarey, B. Kramer and T. Lengauer, *J. Comput. Aided Mol. Des.*, 1997, **11**, 369.
14 B. Kramer, M. Rarey and T. Lengauer, *Proteins: Struct. Funct. Genet.*, 1999, **37**, 228.
15 A. N. Jain, *J. Med. Chem.*, 2003, **46**, 499.
16 Z. Zsoldos, D. Reid, A. Simon, B. S. Sadjad and A. P. Johnson, *Curr. Protein Pept. Sci.*, 2006, **7**, 421.
17 Z. Zsoldos, D. Reid, A. Simon, S. B. Sadjad and A. P. Johnson, *J. Mol. Graphics Model.*, 2007, **26**, 198.
18 C. McMartin and R. S. Bohacek, *J. Comput. Aided Mol. Des.*, 1997, **11**, 333.
19 C. M. Venkatachalam, X. Jiang, T. Oldfield and M. Waldman, *J. Mol. Graphics Model.*, 2003, **21**, 289.
20 R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J. Med. Chem.*, 2004, **47**, 1739.
21 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2006, **49**, 6177.
22 O. V. Stroganov, F. N. Novikov, V. S. Stroylov, V. Kulkov and G. G. Chilov, *J. Chem. Inf. Model.*, 2008, **48**, 2371.
23 J. P. Overington, B. Al-Lazikani and A. L. Hopkins, *Nat. Rev. Drug Discovery*, 2006, **5**, 993.
24 N. Moitessier, P. Englebienne, D. Lee, J. Lawandi and C. R. Corbeil, *Br. J. Pharmacol.*, 2008, **153**, S7.
25 E. L. Mehler, *J. Phys. Chem.*, 1996, **100**, 16006.
26 P. F. W. Stouten, C. Frommel, H. Nakamura and C. Sander, *Mol. Simul.*, 1993, **10**, 97.
27 M. Rarey, B. Kramer and T. Lengauer, *Proteins: Struct. Funct. Genet.*, 1999, **34**, 17.
28 M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor and R. Taylor, *J. Med. Chem.*, 2005, **48**, 6504.
29 C. R. Corbeil and N. Moitessier, *J. Chem. Inf. Model.*, 2009, **49**, 997.
30 W. Sherman, T. Day, M. P. Jacobson, R. A. Friesner and R. Farid, *J. Med. Chem.*, 2006, **49**, 534.
31 R. Huey, G. M. Morris, A. J. Olson and D. S. Goodsell, *J. Comput. Chem.*, 2007, **28**, 1145.
32 T. A. Pham and A. N. Jain, *J. Comput. Aided Mol. Des.*, 2008, **22**, 269.
33 M. H. Seifert, *J. Comput. Aided Mol. Des.*, 2009, **23**, 633.
34 A. E. Cho, V. Guallar, B. J. Berne and R. Friesner, *J. Comput. Chem.*, 2005, **26**, 915.
35 F. N. Novikov, V. S. Stroylov, O. V. Stroganov, V. Kulkov and G. G. Chilov, *J. Mol. Model.*, 2009, in press.
36 G. Marcou and D. Rognan, *J. Chem. Inf. Model.*, 2007, **47**, 195.
37 R. K. Nandigam, S. Kim, J. Singh and C. Chuaqui, *J. Chem. Inf. Model.*, 2009, **49**, 1185.
38 K. Zhu, M. R. Shirts and R. A. Friesner, *J. Chem. Theory Comput.*, 2007, **3**, 2108.
39 P. Block, C. A. Sotriffer, I. Dramburg and G. Klebe, *Nucl. Acids Res.*, 2006, **34**, D522.
40 R. Wang, X. Fang, Y. Lu, C. Y. Yang and S. Wang, *J. Med. Chem.*, 2005, **48**, 4111.
41 T. Liu, Y. Lin, X. Wen, R. N. Jorrisen and M. K. Gilson, *Nucl. Acids Res.*, 2007, **35**, D198.
42 C. A. Sotriffer, P. Sanschagrin, H. Matter and G. Klebe, *Proteins*, 2008, **73**, 395.
43 A. N. Jain, *J. Comput. Aided Mol. Des.*, 2008, **22**, 201.
44 C. McInnes, *Curr. Opin. Chem. Biol.*, 2007, **11**, 494.
45 G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff and M. S. Head, *J. Med. Chem.*, 2006, **49**, 5912.
46 G. E. de Kloe, D. Bailey, R. Leurs and I. J. P. de Esch, *Drug Discovery Today*, 2009, **14**, 630.
47 P. J. Hajduk and J. Greer, *Nat. Rev. Drug Discovery*, 2007, **6**, 211.
48 E. Kellenberger, N. Foata and D. Rognan, *J. Chem. Inf. Model.*, 2008, **48**, 1014.
49 K. Hult and P. Berglund, *Curr. Opin. Biotechnol.*, 2003, **14**, 395.
50 G. A. Lazar, W. Dang, S. Karki, O. Vafa, J. S. Peng, L. Hyun, C. Chan, H. S. Chung, A. Eivazi, S. C. Yoder, J. Vielmetter, D. F. Carmichael, R. J. Hayes and B. I. Dahiyat, *Proc. Natl. Acad. Sci.*, 2006, **113**, 4005.
51 A. A. Canutescu, A. A. Shelenkov and R. L. Dunbrak, *Protein Sci.*, 2003, **12**, 2001.

Received: 3rd August 2009; Com. 09/3370