Assessment of Multiple Binding Modes in Ligand−Protein Docking

Per Källblad,* Ricardo L. Mancera, and Nikolay P. Todorov

De Novo Pharmaceuticals, Compass House, Vision Park, Histon, Cambridge CB4 9ZR, U.K.

Received March 5, 2004

Abstract: Computational ligand−protein docking is routinely used for binding mode prediction. We have quantified the effect of considering multiple docking solutions on the success rate of obtaining the crystallographic binding mode. By selection of a small set of representatives, the experimentally observed binding mode can be predicted with a higher probability after a ligand−protein docking simulation. The proportion of correctly predicted complexes improved from 69% to 87% when five distinct binding modes were considered.

The prediction of the correct mode of ligand−protein binding is extremely important not only as an essential molecular recognition problem but also for its implications for drug discovery. In the absence of experimental structural data on binding modes, computational protein−ligand docking methods are now routinely used to predict the binding mode of biologically active small molecules when complexed to their protein receptors. Observed binding affinities and biological effects can be rationalized in terms of specific interactions with the protein binding site.[1−3]

A variety of different search and optimization methods have been developed for ligand−protein docking applications, such as genetic algorithms,[4] energy minimization,[5] molecular dynamics,[6] simulated annealing,[7] and more recently parallel tempering[8] and stochastic tunneling.[8,9] To be of practical importance, docking methods have to address the balance between efficiency and accuracy in the search for the global minimum of the energy landscape. It has been suggested that the energy landscape that is characteristic of ligand−protein binding is funnel-shaped and, in that respect, resembles the energy landscape of protein folding.[10,11] Native ligands exhibit minimally frustrated pathways to the global minimum leading to a stable binding mode, while nonbinding ligands will have a frustrated energy landscape leading to multiple binding modes.[12] Native ligands thus fulfill both thermodynamic stability and kinetic accessibility criteria due to a funnel-shaped binding energy landscape.

The search for the binding modes of ligands in protein binding sites also depends crucially on the quality of the approximations employed to characterize the exact ligand−protein binding energy landscape. Approximate functions for describing ligand−protein interactions have been extensively investigated.[13−15] The performance of these functions is limited by the difficulties in accurately treating the effects of solvation and the entropic changes due to hydrophobic interactions.[16] The reliability of scoring functions in molecular docking has been addressed by the use of consensus scoring, where a number of scoring functions are used to rank ligand-binding modes,[17] and by the use of different scoring functions for the ligand-binding mode search and for the scoring of ligand−protein interactions.[2]

Approximate scoring functions aim to preserve the native energy landscape to correctly identify the native binding mode as the one with the lowest energy. The results of molecular docking simulations suggest that the native binding mode corresponds to a low-energy structure but not necessarily with the lowest energy.[13−15] This observation suggests that the success rate for finding the native binding mode might be significantly increased if a small set of distinct binding modes within a certain energy threshold above the lowest energy binding mode is considered.[18] Other methods have dealt with multiple binding modes using clustering techniques and ranking based on population pairs.[19]

The accurate prediction of the mode of binding of a ligand to a protein can be critical in a drug discovery program. We have examined and statistically quantified the effect of retaining a small number of low-energy solutions on the success rate of finding the binding mode observed in the crystal structure. First, we apply a recently introduced quantum stochastic tunneling (QSTUN) docking method[20] to generate multiple docking solutions for each of the 305 ligand−protein complexes in the CCDC/Astex data set,[21] and then we select binding modes within a certain energy threshold above the lowest energy solution. We have found that by selection of a small set of distinct binding mode representatives the accuracy of the prediction of the experimentally observed binding mode improves substantially. This finding can add significant value to drug design applications, where the subsequent rational selection of one or more ligand-binding modes from a list of a few distinct docking poses could be carried out on the basis of a further more rigorous theoretical analysis and on the basis of the known binding mode of other ligands.

There are two main stages in the methodology: (i) docking and generation of ligand-binding modes and (ii) analysis and selection of distinct binding modes.

Quantum stochastic tunneling[20] has been reported as an efficient method for flexible ligand−protein docking. This is a hybrid optimization method that combines stochastic tunneling[8,9,22,23] with a path-integral Monte Carlo[24] method to allow for the quantum and stochastic tunneling through high-energy barriers and the nonlocal exploration of the potential energy surface.[20] Two-hundred simulations were performed for each complex using the PLP scoring function[4] to represent the potential energy surface. Several binding modes were generated with each docking simulation. This yielded a large number of ligand-binding modes that were then rescored using the ScreenScore scoring function.[13]

An analysis of the ligand-binding modes was performed for each ligand−protein complex. The objective was to find out whether, by taking the first few best energy-ranking solutions, the success rate of correctly

* To whom correspondence should be addressed. Phone: +44 1223 238000. Fax: +44 1223 238088. E-mail: per.kallblad@denovopharma.com.

10.1021/jm0498147 CCC: $27.50
Published on Web 05/26/2004

predicting the native binding mode of the ligand increased significantly. Three types of analysis were performed.

First, the best root-mean-square deviation (rmsd) from the crystal was found from solutions within a fixed energy threshold above the lowest binding energy. Different energy thresholds above the lowest-energy solution were tried in steps of 1.0 kJ/mol, and different numbers of binding modes were collected. These thresholds account for the uncertainty in the binding energy values inherent in the scoring function used.

Second, since only a limited number of solutions can be investigated in detail in a drug design program, a more relevant question to ask is how the accuracy improves if only a fixed number of binding modes is selected. Does the success rate of predicting the correct crystallographic binding mode increase significantly by considering a small number of distinct binding modes? For this purpose the ligand-binding modes were ordered according to their binding energies. Then, starting from the lowest-energy solution, the rmsd of each pose with respect to all the previous lower-energy poses was computed. If the rmsd was higher than a given threshold (2.0 Å), then that pose was added to a list of distinct solutions. The result of this procedure was a list of distinct ligand-binding mode poses that had rmsd's between one another of more than the predefined threshold. Such a selection procedure can add significant value to drug design applications, where the subsequent selection of one or more ligand-binding modes from a list of a few distinct docking poses can be carried out on the basis of a further full energy minimization, the position and orientation of chemical substitution points, and the known binding mode of other ligands or simply by intuition.

Third, all docking solutions were assigned to the nearest distinct binding mode and the distinct binding modes were reranked according to the number of docking solutions they attract, or their occupancy. This ranking procedure relates to concepts from the energy landscape theory describing the ligand−protein interaction.[12] Consequently, it is possible to assume that many random initial conformations of the native ligand regularly dock in the stable (and dominant) binding mode, although because of the frustrated nature of the energy landscape and the limited search, alternative binding modes will also be found. Therefore, it is reasonable to expect that the most occupied binding mode would correspond to the native binding mode as observed experimentally.

All 305 ligand−protein complexes in the CCDC/Astex validation set[21] were used. This data set has entries that cover the largest populated structural classes of proteins and their homologues. Furthermore, the complexes have protonation and tautomeric states that have been assigned manually. The original validation set was further "cleaned" by removing structures that had (i) factual and structural errors in their PDB files, (ii) ligands with unlikely conformations or determined inconsistently with regard to their electron density, (iii) severe clashes between protein and ligand atoms, and (iv) ligand contacts to crystallographically related protein residues. This "clean" subset contained 224 structures and was then filtered further on the basis of the crystallographic

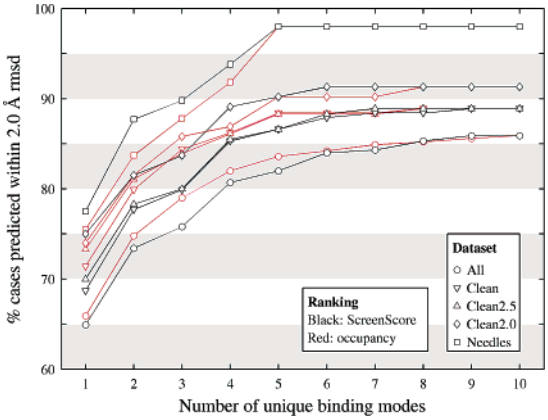

Figure 1. Changes in the success rate of finding the correct crystallographic binding mode when using the distinct binding mode search method. Data for five subsets of the CCDC/Astex data set is shown. Results where binding modes were ranked on the basis of occupancy are shown in red.

resolution (R) of the protein structures. Two further subsets were used: one with protein structures that have a resolution of 2.5 Å or better (containing 180 structures) and another that had protein structures that have a resolution of 2.0 Å or better (containing 92 structures). Full details of the above subsets can be found elsewhere.[21] We also defined a new subset of small structures from the "clean" CCDC/Astex data set in a fashion similar to that in a previous study.[25] Structures included in this set, which we have named "needles",[26] were those with a molecular weight below 300 and with three or fewer rotatable bonds. The needle subset consisted of 67 structures. These five different subsets are referred to as All, Clean, Clean2.5, Clean2.0, and Needles throughout the paper.

The results for the different sets of complexes are summarized in Figures 1−3. Each figure shows the percentage of the complexes for which a docking solution was found within a given rmsd from the crystal structure at a specified threshold (energy or number of distinct binding modes). First, we analyze the results when only the best-ranking solution is considered for each complex, and then we investigate the effect of different thresholds above the lowest-energy solution.

Our results show (Figure 1) that by using ScreenScore to evaluate the energy of binding, 65% of the cases in the whole data set (All) are predicted correctly (rmsd < 2.0 Å). We can see that there is a significant improvement in the success rate as the quality of the subset improves. The success rate improves to 69% in the Clean subset, to 70% in the Clean2.5 subset, and to 75% in the Clean2.0 subset. These are all good figures but clearly indicate that a docking simulation run to predict the binding mode of a ligand could fail between 35% (for any protein structure) and 25% (for a "clean" high-resolution protein structure) of the cases. The highest level of success was observed for the Needles subset, with a 78% success rate.

The real test of a docking method is for it to be able to identify the true crystallographic binding mode of a ligand on the basis of its energy of binding. It is well established that it is the scoring function (ScreenScore in this case) that eventually becomes the limiting factor for the success of a docking method. When docking solutions are considered within a small threshold above

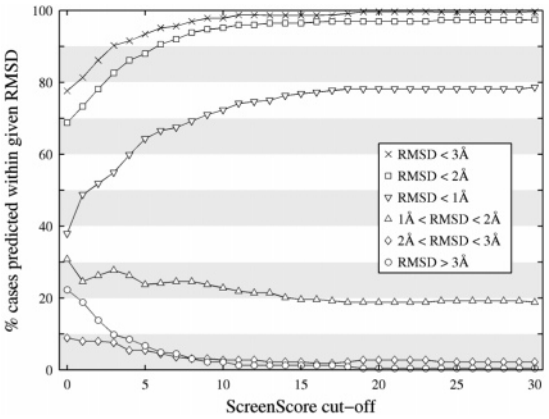

Figure 2. Changes in the success rate of finding the correct crystallographic binding mode as a function of binding energy tolerance. Results shown are for the "clean" CCDC/Astex data set.

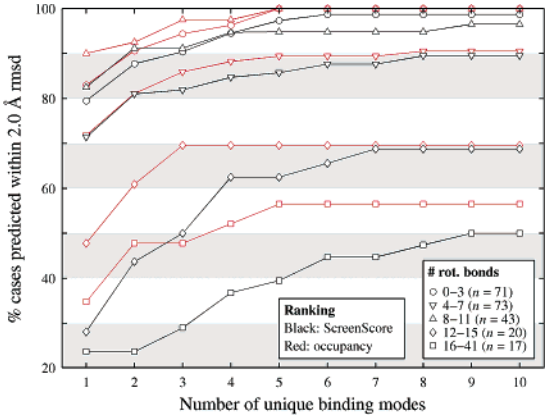

Figure 3. Effect of subdividing the "clean" CCDC/Astex data set according to the number of rotatable bonds. Results where binding modes were ranked on the basis of occupancy are shown in red.

the lowest energy, the number of cases with good solutions increases substantially. This effect is shown in Figure 2 for the Clean data set. For example, the percentage of complexes with solutions within a 2.0 Å rmsd from the crystallographic binding mode increases from 69% to 88% within a 5.0 kJ/mol threshold and to 95% within a 10.0 kJ/mol threshold.

The effect of applying the distinct binding mode search to each of the subsets of the CCDC/Astex data set can also be seen in Figure 1. For the Clean subset the success rate increased from 69% with only one distinct binding mode to 80% for three distinct binding modes and to 87% for five distinct binding modes. The equivalent numbers for the Clean2.0 subset were 75%, 84%, and 90% for one, three and five distinct binding modes, respectively. Significant improvement was also seen for the Needles subset, starting at 78% with one binding mode and reaching success rates of 90% and 98% for three and five binding modes, respectively. Increasing the number of binding modes beyond five does not seem justified in terms of the marginal further success rate improvement and also the added complexity of having more than five possible solutions to a docking simulation. We have observed that the increase in the success rate is mostly due to a large decrease in the number of test cases with an rmsd greater than 3.0 Å and also, to a lesser extent, to a decrease in the number of test cases with rmsd values between 2.0 and 3.0 Å (data not shown). This is accompanied by a large increase in the number of test cases with an rmsd of less than 1.0 Å and also, to a lesser extent, by an increase in the number of test cases with an rmsd between 1.0 and 2.0 Å.

The largest improvements of the success rates are seen for ligands with an intermediate number of rotatable bonds (Figure 3). Complexes containing ligands with up to 11 rotatable bonds show a modest improvement of up to 18% points (with five binding modes). For complexes containing ligands with 12−15 rotatable bonds the increase is around 34% points. For ligands with more than 16 rotatable bonds the increase is around 16% points due to the fact that the native binding mode was not always found. Similar trends are observed if the binding modes are ranked according to their occupancy (Figure 3, in red). When ranked by occupancy, ligands with 12 or more rotatable bonds had higher success rates compared to energy ranking. On the other hand, energy ranking is a better predictor for the Needles data set (Figure 1). Therefore, it should be valuable to consider energy and occupancy ranking when assessing docking results.

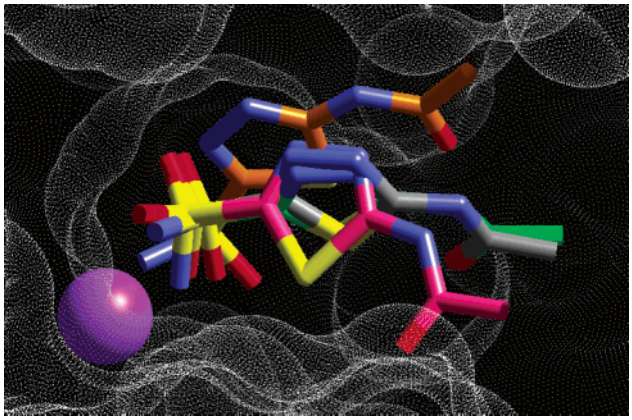

Figure 4. Example (1AZM) where the lowest-energy docking solution is incorrectly predicted but where subsequently ranked distinct binding modes provide superior solutions. Carbon atoms represent the following: (green) rmsd < 1.0 Å; (orange) 2.0 Å < rmsd < 3.0 Å; (pink) rmsd < 3.0 Å. The crystallographic binding mode is shown in gray, and the Zn atom is shown as a purple sphere.

These results demonstrate that selecting a few distinct binding modes, on the basis of their pairwise rmsd or their occupancy, can provide a small number of alternative binding modes that can increase significantly the chances of finding the native binding mode. Consideration of multiple binding modes selected on the basis of rmsd differences should be applicable to any docking method and/or combination of scoring functions. Furthermore, a benefit of using a stochastic docking approach to obtain alternative binding modes is that they reflect the shape of the ligand−protein energy landscape and their relative occupancy can be taken as an indicator of the "near-nativeness" of the binding mode of a given ligand.

Figure 4 illustrates the effect of considering multiple distinct binding modes in one of the test cases, carbonic anhydrase I (1AZM). Three distinct docking solutions of 5-acetamido-1,3,4-thiadiazole-2-sulfonamide are shown. The lowest-energy solution (in pink) has an rmsd of 3.5

Å from the experimental binding mode (in gray). The subsequent distinct binding mode (in orange) has an energy that is 1.2 kJ/mol higher and shows a slight improvement with an rmsd of 2.7 Å. The third distinct binding mode (in green) is a further 2.5 kJ/mol higher in energy and shows good agreement with the crystallographic binding mode having an rmsd of only 0.8 Å.

In conclusion, we have evaluated statistically a distinct binding mode search as a means of predicting the native ligand−protein complex. The usefulness of the distinct binding mode search method was assessed with the CCDC/Astex data set of ligand−protein complexes using the QSTUN docking method and the ScreenScore function. We have found that the success rate of docking simulations improves significantly when a number of distinct binding mode poses are considered, the main effect arising from an improvement in the success rates of those ligand−protein complexes having ligands with a large number of rotatable bonds. We have been able to find the correct crystallographic ligand-binding mode in up to 87% of the cases when up to five distinct binding modes are considered, compared to a success rate of 69% if only the top-ranking mode is considered. Such distinct binding modes ranked by occupancy reflect the shape of the ligand−protein energy landscape, which dictates the higher thermodynamic and kinetic accessibility of near-native binding modes.

This result can add significant value to drug design applications, where the subsequent judicious selection of one or more ligand-binding modes from a list of a few distinct docking poses could be carried out on the basis of a further more rigorous theoretical analysis or on the basis of the known binding mode of other ligands.

Acknowledgment. R.L.M. is also a Research Fellow of Hughes Hall, Cambridge, U.K.

References

(1) Glen, R. C.; Allen, S. C. Ligand−Protein Docking: Cancer Research at the Interface between Biology and Chemistry. *Curr. Med. Chem.* **2003**, *10*, 767−782.
(2) Stahl, M.; Schulz-Gasch, T. Practical Database Screening with Docking Tools. *Ernst Schering Res. Found. Workshop* **2003**, *42*, 127−151.
(3) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A Review of Protein−Small Molecule Docking Methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151−166.
(4) Gehlaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317−324.
(5) McMartin, C.; Bohacek, R. S. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333−344.
(6) Wang, J.; Dixon, R.; Kollman, P. A. Ranking ligand binding affinities with avidin: A molecular dynamics-based interaction energy study. *Proteins* **1999**, *34*, 69−81.
(7) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293−304.
(8) Merlitz, H.; Wenzel, W. Comparison of stochastic optimization methods for receptor−ligand docking. *Chem. Phys. Lett*. **2002**, *362*, 271−277.
(9) Merlitz, H.; Burghardt, B.; Wenzel, W. Application of the stochastic tunneling method to high throughput database screening. *Chem. Phys. Lett.* **2003**, *370*, 68−73.
(10) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. Complexity and simplicity of ligand−macromolecule interactions: the energy landscape perspective. *Curr. Opin. Struct. Biol.* **2002**, *12*, 197−203.
(11) Wang, J.; Verkhivker, G. M. Energy landscape theory, funnels, specificity and optimal criterion of biomolecular binding. *Phys. Rev. Lett.* **2003**, *90*, 188101.
(12) Rejto, P. A.; Verkhivker, G. M. Unraveling principles of lead discovery: From unfrustrated energy landscapes to novel molecular anchors. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8945−8950.
(13) Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(14) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(15) Pérez, C.; Ortiz, A. R. Evaluation of Docking Functions for Protein−Ligand Docking. *J. Med. Chem*. **2001**, *44*, 3768−3785.
(16) Ajay; Murcko, M. A. Computational Methods To Predict Binding Free Energy in Ligand−Receptor Complexes. *J. Med. Chem.* **1995**, *38*, 4953−4967.
(17) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. 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.
(18) Verkhivker, G. M.; Bouzida, D.; Gehlaar, D. K.; Rejto, P. A.; Schaffer, L.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Hierarchy of Simulation Models in Predicting Structure and Energetics of the Src SH2 Domain Binding to Tyrosyl Phosphopeptides. *J. Med. Chem*. **2002**, *45*, 72−89.
(19) Paul, N.; Rognan, D. ConsDock: A new program for the consensus analysis of protein−ligand interactions. *Proteins* **2002**, *47*, 521−533.
(20) Todorov, N. P.; Mancera, R. L.; Monthoux, P. H. A new quantum stochastic tunneling optimisation method for protein−ligand docking. *Chem. Phys. Lett*. **2003**, *369*, 257−263.
(21) 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 Interactions. *Proteins* **2002**, *49*, 457−471.
(22) Hansmann, U. H. E. Protein folding simulations in a deformed energy landscape. *Eur. Phys. J. B* **1999**, *12*, 607−611.
(23) Wenzel, W.; Hamacher, K. Stochastic tunneling approach for global minimization of complex potential energy landscapes. *Phys. Rev. Lett*. **1999**, *82*, 3003−3007.
(24) Feynman, R. P.; Hibbs, A. R. *Quantum Mechanics and Path Integrals*; McGraw-Hill: New York, 1965.
(25) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.
(26) Boehm, H. J.; Boehringer, M.; Bur, D.; Gmuender, H.; Huber, W.; Klaus, W.; Kostrewa, D.; Kuehne, H.; Luebbers, T.; Meunier-Keller, N.; Mueller, F. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* **2000**, *43*, 2664−2674.

JM0498147