An Alternative Method for the Evaluation of Docking Performance: RSR vs RMSD

Dilmurat Yusuf,[§] Andrew M. Davis,[†] Gerard J. Kleywegt,[‡] and Stefan Schmitt*

AstraZeneca R&D Mölndal, S-431 83 Mölndal, Sweden

Received March 10, 2008

A new assessment criterion for docking poses is proposed in which experimental electron density is taken into account when evaluating the ability of docking programs to reproduce experimentally observed binding modes. Three docking programs (Gold, Glide, and Fred) were used to generate poses for a set of 88 protein−ligand complexes for which the crystal structure is known. The new criterion is based on the real space R-factor (RSR), which measures how well a group of atoms−in our case the ligand−fits the experimental electron density by comparing that density to the expected density, calculated from the model (i.e., the predicted ligand pose). The RSR-based measure is compared to the traditional criterion, the root-mean-square distance (RMSD) between the docking pose and the binding configuration in the crystallographic model. The results highlight several shortcomings of the RMSD criterion that do not affect the RSR-based measure. Examples illustrate that the RSR-derived approach allows a more meaningful a posteriori assessment of docking methods and results. Practical implications for docking evaluations and for methodological development work in this field are discussed.

INTRODUCTION

Knowledge of the three-dimensional (3D) structure of target proteins in the form of high-resolution crystal structures forms the basis for structure-based design (SBD)−the rational design of ligands with increased, ability to occupy, and affinity for the binding site of a particular target protein. While structure-based design can be an exercise in "manual" interpretation of the 3D target information, the use of computational tools and technologies to exploit structures has become routine in the pharmaceutical industry.[1,2] Among these tools, ligand-docking programs are particularly popular.[3,4]

Docking comprises two distinct tasks namely the prediction of favorable binding geometries for a small molecule in the binding site of a target protein and the estimation of the binding free energy of the complex so formed. The latter task, which is usually referred to as scoring, is one of the major challenges in the development of docking programs.[5,6] Docking scores evaluate the complementary of a small molecule and the target structure in terms of shape, electrostatics, or the presence of key stabilizing interactions such as hydrogen bonds, salt-bridges, and hydrophobic contacts. The prediction of correct binding geometries is therefore an essential prerequisite for the calculation of meaningful docking scores. An effective scoring function in fact serves multiple purposes. It has to guide the docking search algorithm toward reasonable placements, orientations, and conformations of a ligand (= generation of a docking *pose*), it must distinguish the best or "correct" pose from alternative docking solutions of the *same* ligand that the search algorithm suggests, and it should rank *different* ligands in the order of their binding affinity.

In the last few decades a wide range of docking programs has been developed with different search algorithms to generate poses and with a variety of scoring functions.[7] The search algorithms can be divided into

1. Those that decompose a ligand into fragments and incrementally build it up again in the environment of the binding site after placing a base fragment[8,9]
2. Those that relax the entire ligand in the binding site[10–13]
3. Those that rigidly position pregenerated ligand conformers[14] into the protein area of interest.

Each docking program usually has its own scoring function or uses different functions for the various steps of the docking procedure, but postscoring,[15–19] consensus scoring,[20–23] or using a combination of docking techniques[24–26] are also common methods for the prediction of binding modes and energies.

In the wake of the impressive amount of methodological development work in this field, numerous docking program evaluations and comparisons have appeared in the literature. These attempt to assess which program or combination of techniques is the best choice for a particular SBD-driven drug target.[27–38] These evaluations typically apply two performance criteria: enrichment rates and comparison to protein−ligand crystal structures. Enrichment rates are used to measure how well docking programs perform in the context of virtual screening,[39] where a database of many thousands of small molecules is docked against a particular target. The database includes a small subset of known binders against the target, and the performance of the docking procedure is assessed by determining the retrieval rate of these known binders among molecules with the highest docking scores. The second performance criterion is the ability of docking programs to reproduce the binding mode of ligands as

* Corresponding author e-mail: Stefan.Schmitt@astrazeneca.com.
[§] Current address: Institute for Theoretical Chemistry, University of Vienna, Währingerstrasse 17, 1090 Vienna, Austria.
[†] Current address: AstraZeneca R&D Charnwood, Bakewell Road, Loughborough, Leicestershire LE11 5RH, U.K.
[‡] Current address: Department of Cell and Molecular Biology, Uppsala University, Biomedical Centre, Box 596, S-75124 Uppsala, Sweden.

10.1021/ci800084x CCC: $40.75
Published on Web 07/04/2008

observed in the corresponding crystal structure which is usually measured by calculation of the root-mean-square distance (RMSD) between the non-hydrogen atoms of the ligand in the crystal structure and the corresponding atoms in the docked pose.

It is tempting to consider enrichment rates as the more relevant performance criterion since the retrieval of true positives is of paramount importance in virtual screening. However, this criterion emphasizes the ability of the scoring function to discriminate between different ligands. It does not necessarily address the question whether a molecule that has been identified as a good binder to a particular target has obtained this ranking for the right reasons, i.e., because the search engine of the docking program has found the correct placement, orientation, and conformation and which the scoring function has then appropriately evaluated and thus ranked that particular molecule above others. Consequently, the robustness of binding mode prediction is as important to docking programs as the scoring function and is crucial for a correct interpretation of the docking scores. The RMSD measure used in docking evaluations can accomplish this if a threshold value is defined below which a docked pose is considered to be correct. The frequency of such successful docking poses in a data set is taken as a measure of the docking performance. An RMSD threshold value of 2.0 Å is widely accepted as distinguishing between success and failure in reproducing a known binding mode.

In some efforts to improve scoring functions it was found that the scores of the ligand poses in the crystal structures are generally better than those obtained for the closest corresponding docking poses, i.e., the poses that have the lowest RMSD compared to the ligand in the crystal structure.[40] Such findings have led to the use of lower RMSD threshold values, which means that improvements to scoring functions are tailored to favor poses that better match the ligand coordinates as deposited in the PDB.

However, the indiscriminate use of the RMSD measure has a number of serious drawbacks. First, the protein environment of the ligand is completely ignored. Depending on the size and complexity of the ligand, this may lead to deceptively low RMSD values even when key interactions with the protein are absent in the docked pose. On the other hand, unreasonably high RMSD values may be obtained even if essential binding features have been identified correctly in a docked pose. This may happen if a moiety that is irrelevant to the binding mode (e.g., a solvent-exposed group) deviates substantially from the crystal structure. Kroemer et al. have attempted to alleviate such intrinsic limitations by introducing a new interaction-based accuracy classification scheme (IBAC) as an alternative to RMSD values to assess pose-prediction accuracy.[41] Second, the use of RMSD values contravenes the fundamental scientific tenet that predicted data should be compared to experimental data in order to evaluate the predictive power of a method. The ligand coordinates from the crystal structure are not the primary experimental data but a subjective interpretation of the electron density.[10,42,43] In the same way that a docking algorithm is guided by a scoring function to position the ligand into the binding site, crystallographers are governed by experience, prejudice, expectation, and local practice when building and refining a model based on crystallographic data.[42] Furthermore, even for a correctly built model the experimental uncertainty of the individual non-hydrogen atom positions can still be 0.1−0.5 Å, depending on the resolution of the data.[44] To overcome such crystallographic uncertainties, the data sets used for docking evaluations therefore tend to contain mostly structures determined at relatively high resolution. An increasing awareness of the potential pitfalls in interpreting protein−ligand complex structures[44–46] has also led to a closer inspection of docking evaluation data sets and to the proposal of high-quality standard sets.[47] However such weeding further limits the size and diversity of docking evaluation data sets.

In this paper we propose a novel metric for assessing the success and failure of pose predictions for docking programs that overcomes many of the limitations of the use of RMSD values. Instead of comparing the non-hydrogen atom coordinates from a docking pose against those of the crystal structure we use crystallographic electron density maps for the evaluation. Although there may be issues with residual bias (see Methods section), properly calculated electron density maps constitute a much more faithful representation of the experimental data than any model. By comparing a pose directly with electron density, any ambiguous features of a particular ligand orientation in a binding site environment and alternative placements or conformations can be taken into account implicitly. As the quantitative measure for the comparison of docking poses against electron densities we make use of the real space R-factor (RSR),[48] which is routinely used in crystallography and measures the compatibility of an experimental electron density map and a group of atoms (here, the ligand atoms). The definition of RSR and how it is used to evaluate binding mode predictions are detailed in the methods section. To generate docking poses we chose three docking programs, Gold (CCDC),[49] Glide (Schroedinger),[50] and Fred (OpenEye),[51] that are commonly used in our own modeling work. The aim of this study was not to compare different docking programs with each other but to investigate the utility of an alternative to the RMSD measure for assessing the correctness of docking poses.

METHODS

We recognize that docking programs may give different results if used in different versions or with different input parameters and that users modify their docking protocols depending on the target structure or a particular modeling task. The primary objective for the docking protocols used in this study was to generate sets of docking poses that were then subjected to RSR- and RMSD-based performance tests for evaluating binding mode predictions. The programs used for our calculations were applied in the standard way recommended by the vendors' manuals and according to our own experience with these programs. We will only briefly describe the preparations and computations for the docking calculations since we followed standard docking protocols. The calculation of RSR values will be described at the end of this section in more detail since it is key to the novel assessment criterion and the readers of this journal may not be familiar with the crystallographic niceties involved.

Data Set. A data set of 88 crystal structure complexes from the Protein Data Bank (PDB)[52,53] was compiled from data sets used in earlier docking evaluations.[28,29,31,33,35–37,47,54] Structures for which no experimental data (structure factor

Table 1. PDB Codes of the Crystal Structures of the 88 Protein−Ligand Complexes Used in This Work

PDB codes
13gs 1a28 1ai5 1ajv 1ajx 1aq1 1azm 1b9t 1b9v 1bcu 1bhx 1bju 1bjv 1br6 1bzm 1cet 1ctr 1cvu 1d0l 1di9 1eoc 1ere 1err 1eve 1exa 1ezq 1f0r 1f0s 1f0t 1fjs 1fl3 1fm6 1fm9 1frp 1fvt 1fvv 1gwx 1h1p 1h1q 1h1s 1hck 1hdy 1hgg 1hgh 1ibg 1icn 1ivc 1ive 1k1j 1k7e 1k7f 1ki4 1ki8 1kim 1l2s 1lqd 1m7q 1mbi 1ml1 1mq5 1mq6 1mrg 1mrk 1nhu 1nhv 1qf0 1qf1 1qh7 1rt5 1rt7 1thy 1tpp 1ukz 1vtk 1ydr 1yds 1z1h 25c8 2ack 2ada 2cbs 2fgi 2pcp 3ptb 3vtk 4aah 4cox 5cna

files) had been deposited were rejected, as were structures for which the electron-density map calculations at the Uppsala Electron Density Server (EDS)[55] did not reproduce the published R-value to within 5 percentage points. Complex structures in which the ligand is close to, or interacts with, a protein molecule due to crystal packing were removed from the data set and so were structures in which the ligand is covalently bound to the protein. We also excluded PDB structures whose ligands have more than 15 rotatable bonds and structures in which the ligand coordinates specified in the PDB file represent only a fragment of the actual bioactive molecule that was used for crystallization. The PDB codes of the remaining structures used in this study are listed in Table 1.

Each of the 88 protein complexes was prepared for docking following the protocol recommended by the respective vendors of the docking programs. Hydrogen atoms were added and oriented where appropriate, protonation states were generated corresponding to an aqueous environment at pH 7, and ligand and water molecules were subsequently removed from the structure. Cofactors and metal ions were preserved in the binding sites. Preparation of the protein structures for docking with Gold and Fred were done using Sybyl,[56] and those for Glide were performed via Maestro.[57]

The structure of the ligand from each PDB entry was retrieved in mol2-format from Relibase+[58] and checked for correct bond orders and protonation states. Corina[59] was used to add hydrogen atoms to each ligand. Corina also generated new conformations, randomly translated and rotated the ligands, and thereby eliminated possible bias for the docking.

Gold. Gold[49] was used with "Standard default settings" with binding sites defined as 10 Å spheres around the centroids of the ligand coordinates of the corresponding crystal structures and the cavity detection mode turned on.

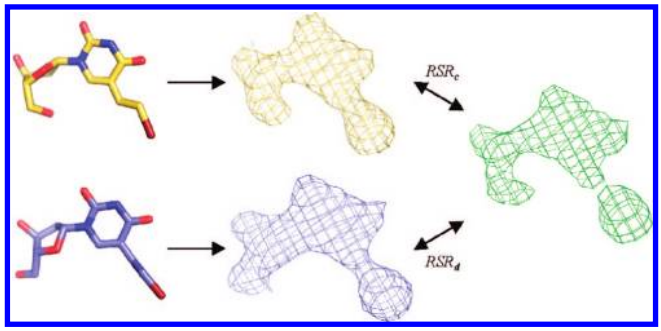

Figure 1. The calculation of RSR values. Theoretical electron density maps (yellow and blue meshes) for a ligand are calculated using the atom coordinates from the crystallographic model (yellow) and for the pose that was produced via docking (blue). These calculated electron density maps are then correlated with the experimental electron density (green), yielding RSR_c and RSR_d, for the crystallographic and docking pose, respectively.

Ligands were used as mol2-files and prepared as described above. A maximum of 10 docking poses per ligand were generated, and early termination was allowed. For each PDB entry two sets of Gold docking poses were generated using the Goldscore and the Chemscore scoring functions, respectively.

Glide. Glide dockings were performed with Impact[50] using Glide-input files that were generated via Maestro.[57] Default settings were applied for the setup steps "Protein Preparation", "Receptor Grid Generation", and "Ligand Docking", respectively. The centers of the grid-enclosing boxes that delineate the binding sites were defined by the centers of the ligands from the underlying PDB files. For Glide set-ups via Maestro the enclosing box dimensions are automatically deduced from the ligand size and fit the entire active site. Ligands were prepared as described above and converted to SD-format via Corina to enable use with the Maestro software. For each PDB structure a set of at most ten docking poses was generated using Glidescore in the Standard Precision (SP) mode. Since Glide requires an energy minimization of the entire protein during the protein preparation step, the atomic coordinates can change slightly from the reference coordinate frame of the original PDB file. However, it is important for the calculation of the electron density maps that all calculations for a particular structure are carried out in the same reference frame. Therefore, the transformation matrix that puts the coordinates back into the original reference frame was calculated and applied to all docking poses resulting from Glide.

Fred. For Fred[51]dockings the boxes delineating the binding site were defined by the ligands from the crystal structures with 4 Å added to the box edges. The protein files were the same as the ones prepared for Gold. For each ligand a multiconformer database in oeb-format was prepared using an in-house protocol for multiconformer generations with Omega.[60] Dockings were done with default parameters, and the number of alternate poses was set to 9, so that sets of at most ten docking poses per Fred run were produced. Only the output corresponding to the Gaussian shape scoring function was considered in this study.

Real-Space R Values (RSR). RSR values were introduced by Jones[43,48] as a way of measuring the fit or compatibility of parts of a crystallographic model to experimental electron density. Whereas a conventional R-value provides a measure of the fit of an entire model to all of the crystallographic data, RSR values can be used to assess the quality of small parts of a model, e.g. individual residues or ligands. RSR values compare an experimental density (ρ_{obs}) with a theoretical density (ρ_{calc}) calculated directly from the atomic model. Both densities are calculated on a grid with a spacing that is typically set to one-third of the resolution. To calculate the RSR values, only grid points in the vicinity of the

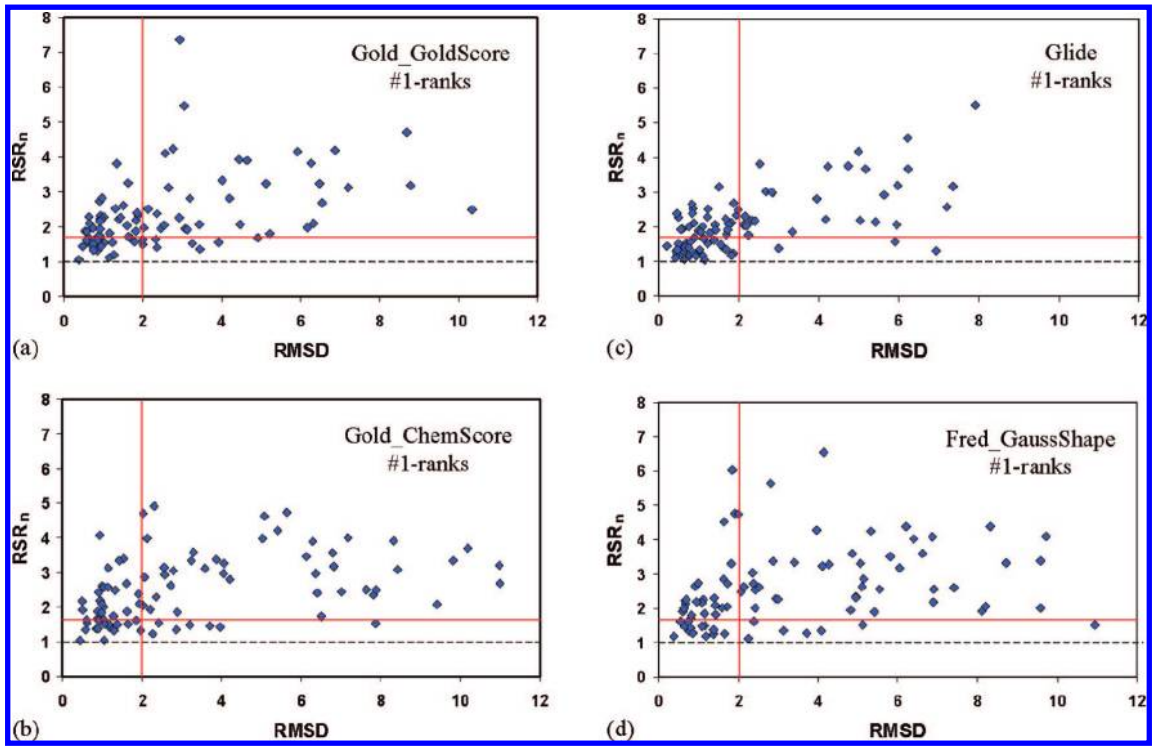

Figure 2. Scatter plots of RSR_n and RMSD values for the top-ranked docking poses of the 88 complex structures in the data set generated using (a) Gold with Goldscore, (b) Gold with Chemscore, (c) Glide, and (d) Fred. The vertical red line in each plot shows the RMSD cutoff of 2.0 Å that is typically used to separate apparent docking successes (left, RMSD < 2.0 Å) and failures (right, RMSD ≥ 2.0 Å). Similarly, red horizontal lines show the empirically determined RSR_n cutoff value of 1.7 used in this work that separates apparent successes (below, $RSR_n < 1.7$) and failures (above, $RSR_n \geq 1.7$) by this criterion. A horizontal dashed line is drawn at $RSR_n = 1$ and illustrates that docking poses rarely fit the electron density better than the model "hand-crafted" during the structure refinement process.

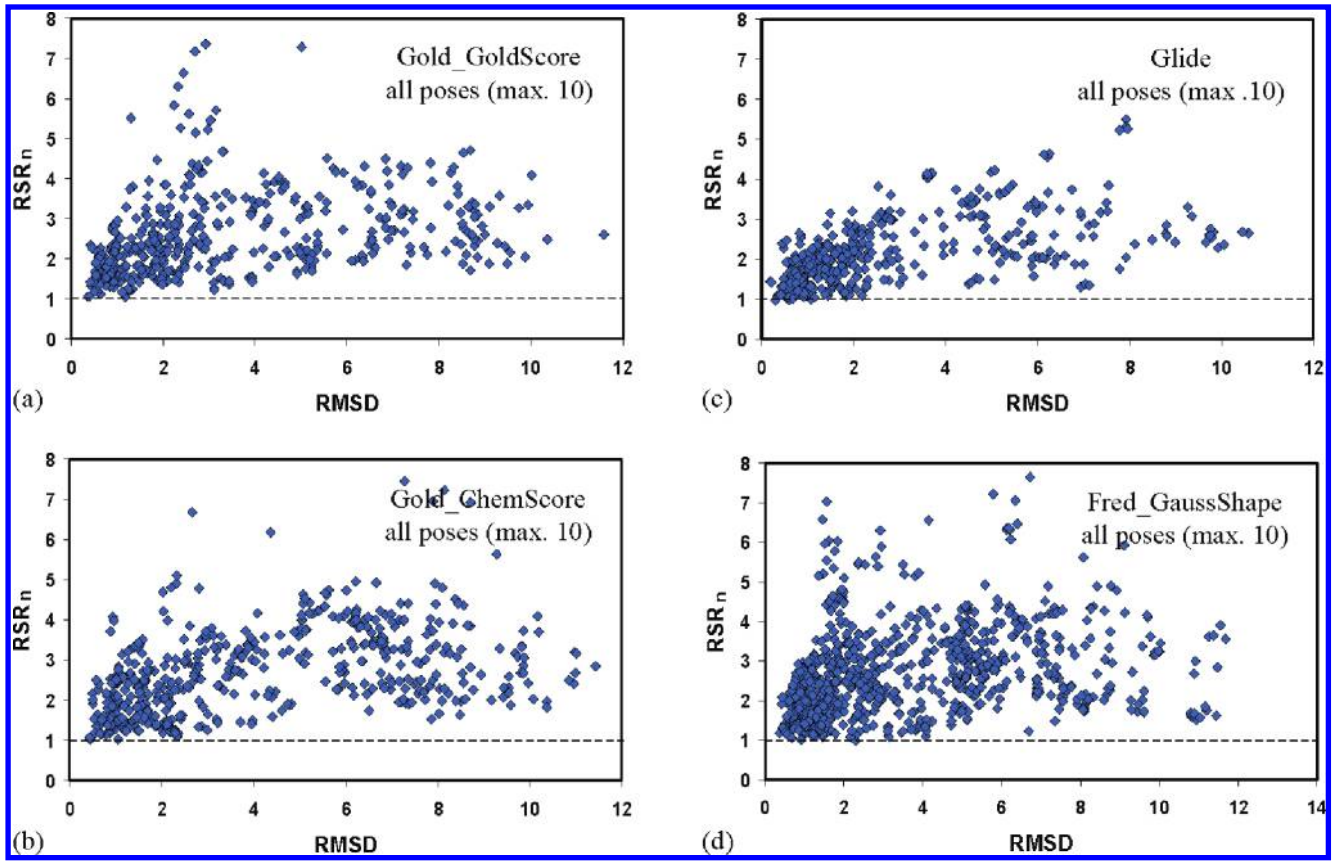

Figure 3. Similar scatter plots to those in Figure 2 but now for all docking poses (with a maximum of 10 per ligand and program/scoring function) generated by (a) Gold with Goldscore, (b) Gold with Chemscore, (c) Glide, and (d) Fred. The horizontal dashed line is drawn at $RSR_n = 1$ and illustrates that only very few docking poses fit the electron density better than the model "hand-crafted" during the structure refinement process.

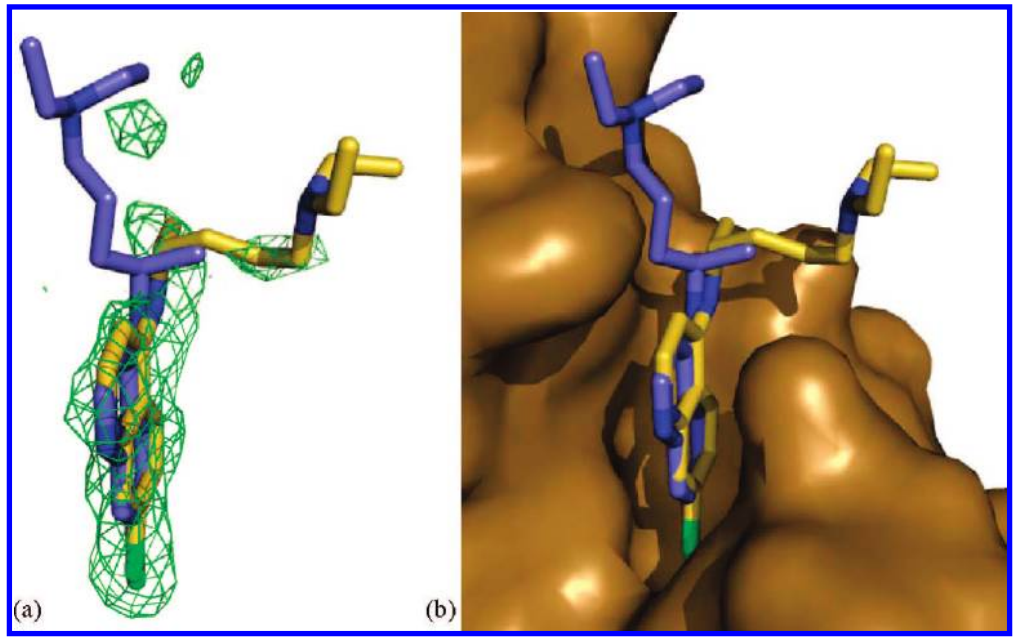

Figure 4. Example of a case in which a docking pose would be incorrectly classified as a docking failure based on the RMSD criterion. (a) The crystal structure of the ligand in PDB entry 1CET is shown in yellow, and the top-ranked docking pose produced by Gold with ChemScore is shown in blue. The experimental electron density is shown as a green mesh, contoured at the rms level ("1 σ"; the same coloring scheme and contour level is also used for the following figures). (b) View of the two ligand configurations in the context of the binding site (brown surface). The high value of the RMSD (3.7 Å) is caused chiefly by a moiety for which there is no clear electron density. The RSR_n value (1.46), on the other hand, is dominated by the part of the ligand for which electron density is observed.

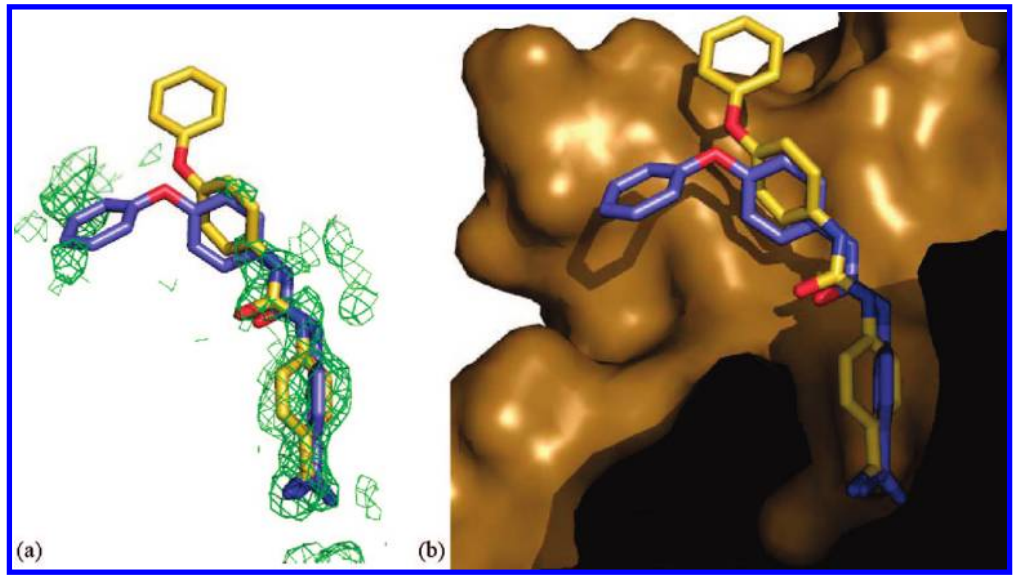

Figure 5. Another example of a case in which a docking pose would be incorrectly classified as a docking failure based on the RMSD criterion. (a) The figure shows a superimposition of the crystal structure (PDB code 1BJV, yellow), docking pose (Gold with Goldscore, blue) and electron density (green mesh). (b) View of the two ligand configurations in the context of the binding site (brown surface). The docking pose has a high RMSD value (3.3 Å) but a moderate RSR_n value (1.52).

scrutinized entity (e.g., a residue or ligand) are taken into account. The RSR value is defined as

$$\mathrm{RSR} = \frac{\sum |\rho_{\mathrm{obs}} - \rho_{\mathrm{calc}}|}{\sum |\rho_{\mathrm{obs}} + \rho_{\mathrm{calc}}|} \qquad (1)$$

where both sums extend over the grid points in the vicinity of the residue or ligand. If the model is a faithful interpretation of the experimental density, the observed and calculated density around the residue or ligand will be very similar, and the RSR value will be low. If the model is a poor interpretation of the density, on the other hand, the RSR will be high. Here, we used a variation of the methods used to calculate EDS maps,[55] in that the ligand whose docking mode had to be predicted was initially omitted from the structure that was used to calculate the electron density maps (ρ_{obs}). In addition, we used minimally biased maps (so-called σ_A-weighted maps[61]). This process was applied to remove some model bias from the experimental electron-density maps. There may still be some residual model bias ("memory" of the ligand as built by the crystallographer) because the crystallographer may have had to build the binding-site residues in a particular way to accommodate the ligand. For the purposes of this study, however, any such secondary bias is ignored. Various CCP4 programs[62] and EDS scripts were used to produce such maps for all 88 structures. RSR values

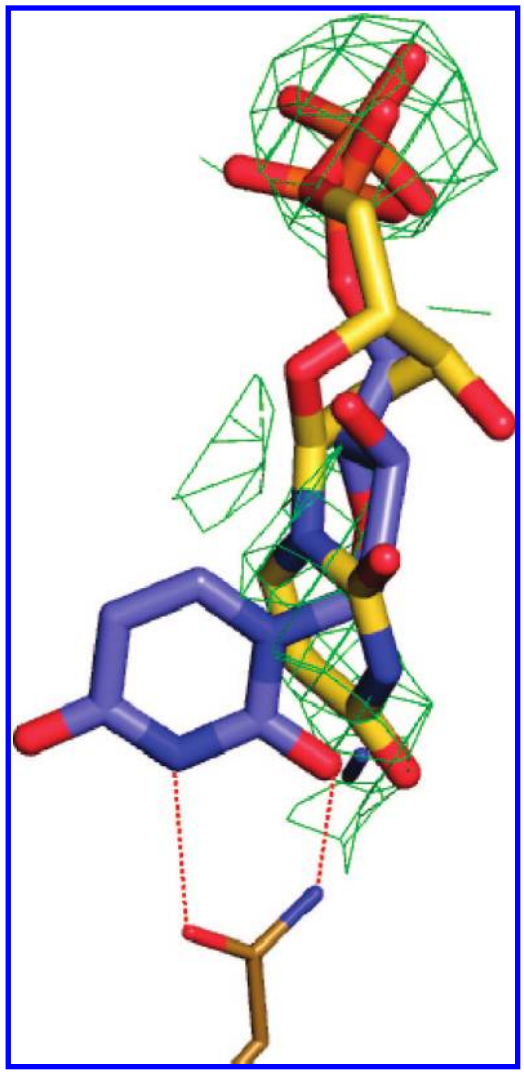

Figure 6. Example of a case where the ligand density (green mesh) is rather poor leading to alternative interpretations having high RMSD values but acceptable RSR_n values. The ligand orientation from the crystal structure (PDB code 1THY, yellow) fits the density slightly better than the docking pose (Gold with Chemscore, blue) which orients the pyrimidine-dione ring so as to form a double hydrogen bond with an asparagine residue of the protein (RMSD = 3.4 Å and $RSR_n = 1.36$).

can be calculated by a variety of programs. For this study we calculated the actual RSR values with the crystallographic modeling program O.[48]

RSR values were subsequently calculated both for the ligand pose that was determined during the crystal-structure refinement process (denoted RSR_c) and for each of the docking poses (denoted RSR_d) (Figure 1).

Since RSR values are strongly dependent on the resolution of the underlying crystallographic data, the value of RSR_d alone cannot be used to differentiate between success and failure. However, the values of RSR_d and RSR_c can be compared directly in every case, and their ratio is expected to reflect the relative quality of the docking pose. Hence, we used this ratio, RSR_n, as the criterion to assess how well a docking pose fits the experimental electron density:

$$RSR_n = \frac{RSR_d}{RSR_c} \qquad (2)$$

The unlikely case of $RSR_n < 1$ would indicate that the ligand pose resulting from the docking fits better into the electron density map than the ligand model that was built by the crystallographer. $RSR_n = 1$ implies that both models fit the electron density map equally well (although the actual atomic coordinates of the docked and the crystallographer's model can still be different). In cases where $RSR_n > 1$, the docking model fits the experimental data less well than the "hand-crafted" model from the structure refinement process. The determination of an appropriate RSR_n cutoff value to assess the binding mode prediction performance is discussed in the Results and Discussion section by comparing RSR_n values with the classic RMSD-based criterion.

RESULTS AND DISCUSSION

Using four different docking methods, we generated ligand docking poses for a set of 88 experimentally determined protein−ligand complex crystal structures. Each predicted ligand pose was then assessed by both the RMSD- and RSR-based criterion. It is important to determine the extent to which RMSD and RSR_n are complementary or redundant measures of the quality of a ligand pose. Figure 2 shows scatter plots of the values of the two measures for the top-ranked poses, and Figure 3 shows the same data for all poses (with a limit of 10 poses per complex as described in the Methods section). Spearman's rank correlation coefficient between the two quantities varies from $r^2 = 0.21$ to $r^2 = 0.35$ (for data in Figure 2) and from $r^2 = 0.17$ to $r^2 = 0.39$ (for data in Figure 3) which demonstrates that RMSD and RSR_n are not equivalent. As expected, poses with low RMSD values cluster near an RSR_n value of 1, but at higher values of RMSD the data points show a considerable spread. This suggests that RMSD and RSR_n often agree in their assessment when a docking program predicts a good pose but that the level of agreement diminishes as the RMSD increases. The fact that high and low RMSD and RSR_n values occur in any combination shows that they capture different kinds of information.

Having established that RMSD and RSR_n are non-redundant measures of docking pose quality, a cutoff to separate apparent docking failures and successes needs to be defined. When RMSD values are used for assessing docking success, a cutoff of 2.0 Å is typically used. In order to define a suitable cutoff for the RSR_n criterion, all the top-ranked poses (Figure 2) were inspected visually by simultaneously looking at docking pose, ligand configuration from the crystal structure, and experimental electron density. Inspection of randomly selected alternative poses from the data shown in Figure 3 subsequently confirmed the findings of the analysis of the top-ranked poses. In the following discussion we will primarily focus on the data from the top-ranked poses.

Of particular interest in this analysis are poses with low RSR_n values (i.e., close to 1) but high RMSD values as well as poses with high RSR_n values but low RMSD values. It was found that most docking poses with RMSD ≥ 2.0 Å and $RSR_n < 1.5$ can still be classified as correct when judged by the electron density (i.e., they are false negatives by the RMSD criterion). In many cases, the high RMSD value is caused by a substantial conformational difference in part of the ligand for which there is no convincing electron density. In other words, the absence of density for such a moiety means that there is no experimental support for either the

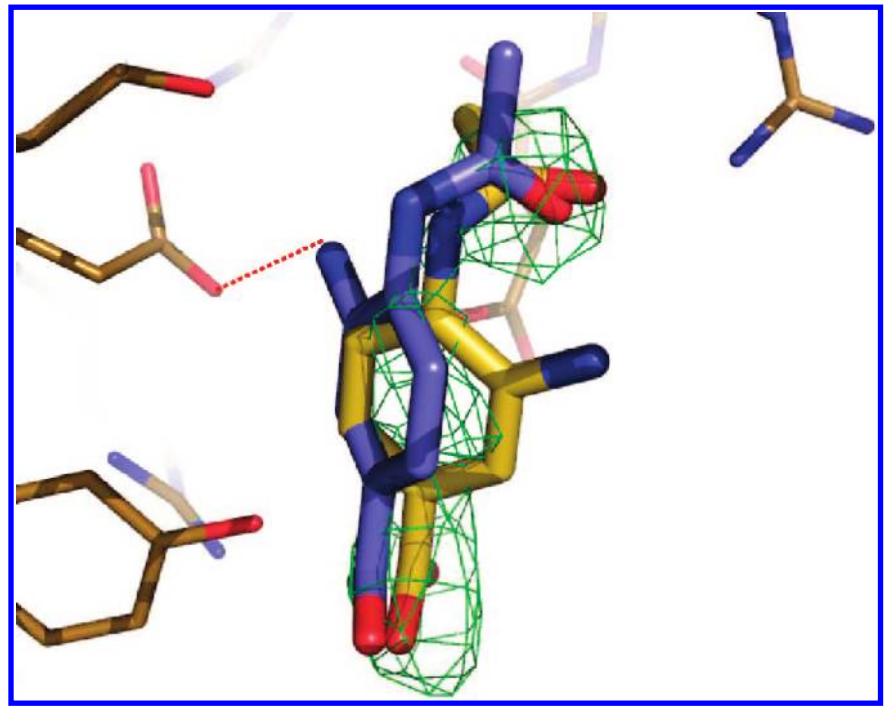

Figure 7. Example of a case where the ligand density is ambiguous, leading to alternative interpretations having high RMSD values but acceptable RSR_n values. The ligand orientation from the crystal structure (PDB code 1IVE, yellow) fits the electron density (green mesh) slightly better than the docking pose (Gold with Goldscore, blue) which has a different orientation of the aromatic ring so as to form a hydrogen bond with a protein residue (RMSD = 2.0 Å, RSR_n = 1.64).

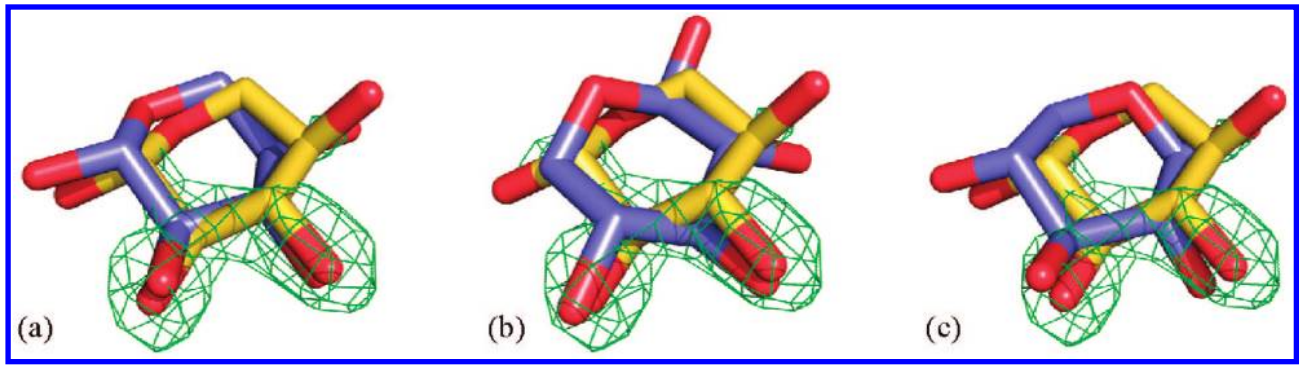

Figure 8. Crystal structure (yellow carbon atoms) and experimental electron density (green mesh) for PDB entry 1QH7, overlaid with the top-ranked docking poses (blue) from (a) Gold with Goldscore (RMSD = 0.9 Å, RSR_n = 1.44), (b) Glide (RMSD = 3.0 Å, RSR_n = 1.38), and (c) Fred (RMSD = 3.1 Å, RSR_n = 1.35). All three poses match the (poor) electron density equally as well as the crystal structure. However, the poses in (b) and (c) have RMSD values in excess of 2.0 Å and would therefore have been considered docking failures. The pose in (a) differs from the crystal structure by a slight shift in orientation, the one in (b) is flipped 180°, and the pose in (c) is flipped and rotated. The REMARK records in the PDB entry suggest that the crystallographers have considered alternative interpretations of the ligand density, but no alternative coordinates were deposited.

crystallographer's or the docking program's conformation. Simultaneously, the pose matches the crystal structure very well for the remainder of the ligand, for which there is convincing electron density. Examples of such cases are shown in Figures 4 and 5.

If there is hardly any electron density for a substantial part of a ligand, the crystal structure and the docking pose will fit equally poorly for that part, and hence RSR_d and RSR_c will be somewhat elevated. However, if the docking pose is a good match for the part of the ligand for which there is density, the ratio of RSR_c and RSR_d (i.e., RSR_n) will remain relatively low. The RMSD measure on the other hand is dominated by the largest structural deviations, which in cases like these happen to be the least important ones.

In the regions of the plots in Figure 2 (and Figure 3) where the RMSD exceeds 2.0 Å but the RSR_n values are fairly low, there are also cases in which the ligand density is ambiguous. In such cases, the crystallographer has made a particular interpretation of the density in terms of the atomic model of the ligand. The docking pose then shows an alternative binding mode that explains the density just as well (or just as poor) as the crystal structure but emphasizes alternative binding features. Examples of such cases are shown in Figures 6, 7, and 8.

The fraction of genuine docking failures increases markedly for cases with RMSD $\geq$ 2.0 Å and $RSR_n \geq 1.7$. For this reason, an RSR_n value of 1.7 was selected as an empirical cutoff to classify poses as successes ($RSR_n < 1.7$) or failures ($RSR_n \geq 1.7$). Somewhat surprisingly, there are quite a few cases where poses with RMSD < 2.0 Å have RSR_n values in excess of 1.7. This suggests that there are cases (false positives) in which poses are classified as docking successes by the RMSD criterion even though they apparently do not fit the experimental electron density particularly well. Even

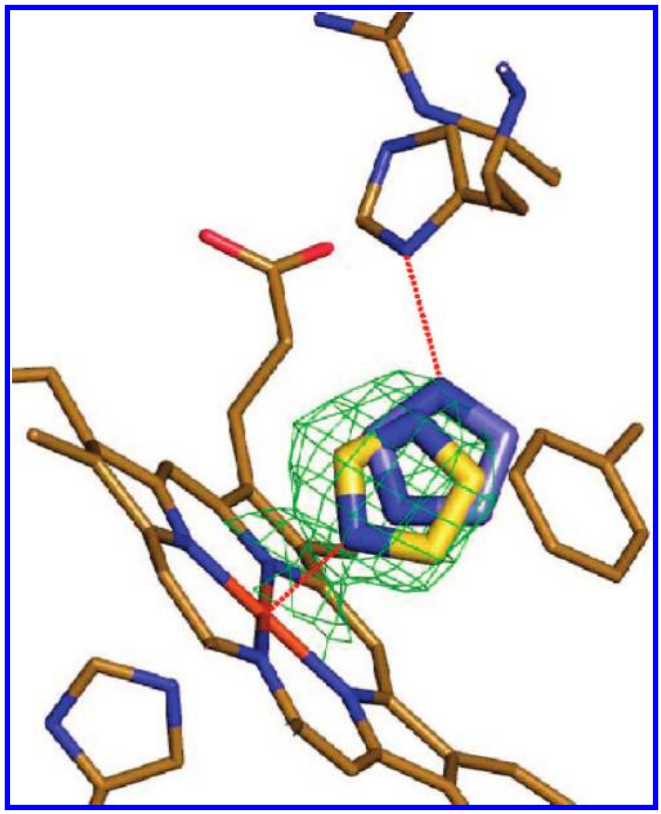

Figure 9. Example of a case in which a pose has a low RMSD value but fits the density poorly. The crystal structure (ligand in yellow) and density (green mesh) for PDB entry 1MBI are shown together with the top-ranked pose from Glide (blue). The docking program favors a hydrogen bond from the ligand to a nearby histidine ring, which leads to a deceptively small displacement from the crystal structure (RMSD = 0.8 Å), but causes a poor fit to the experimental density. The RSR_n value is 2.82, which correctly reveals the pose to be a docking failure. In the crystal structure, the ligand interacts with the iron ion of the cofactor, and this interpretation is supported by the crystallographic data.

some docking poses with RMSD < 1.0 Å can have rather high RSR_n values. Some of the cases involve ligands that are quite small so that even a moderate displacement may lead to a poor fit with the density (see Figure 9 for an example).

The majority of cases in which the docking solutions have low RMSD values but high RSR_n values are due to differences in the ligand conformations (in particular, of ring systems). The sampling of ligand conformations in docking calculations is sometimes sparse due to speed considerations and usually aims to produce low-energy conformations. During crystal-structure determination, on the other hand, energetic considerations usually take a lower priority than the need to explain the observed electron density in terms of an atomic model. This often leads to strained ligand conformations that explain the density and that may or may not be accurate representations of the true structures.[63–65] In such cases, docking programs are unlikely to sample conformations identical or similar to that of the model in the crystal structure. Even if the docking procedure places the ligand in the binding site so that there is good overall agreement with the crystal structure in terms of atomic positions, interactions, and molecular shape, this does not guarantee that it will fit the electron density as well as the "hand-crafted" model produced by the crystallographer. In such cases, ostensibly small differences can result in high RSR_n values. This effect is amplified if the pose differs from the crystal structure in the conformation of one or more rings, as the differences will be propagated to the substituents. Examples of this phenomenon are shown in Figures 10 and 11.

It is interesting to note that none of the top-ranked poses and only a few of the alternative docking solutions give an RSR_n below 1.0. Such a case indicates that the docking pose yields a better fit to the electron density than the ligand placement and conformation chosen and refined by the crystallographer during structure determination. An example of such a case is shown in Figure 12.

In the exhaustive manual analysis of the 4 × 88 top-ranked docking poses one case surfaced in which both RMSD and RSR_n failed to recognize an incorrect docking. The docking pose, shown in Figure 13, passes the RMSD and RSR_n tests even though it fails to reproduce an important hydrogen bond to a protein residue. To recognize cases like these, assessment methods that capture information about protein−ligand interactions, such as IBAC,[41] might be more appropriate or could be used as an additional filter.

The performance of a docking program in terms of its ability to reproduce the ligand placement and conformation of crystal structures is typically assessed by determining the fraction of top-ranked docking poses that display an RMSD of less than 2.0 Å to the crystal structure for a collection of experimental structures. The higher this fraction, the better the docking engine is judged to be. To assess the effect of using the experimental data (electron density) in a docking evaluation, the fraction of docking poses judged to be correct by the RSR_n criterion (i.e., RSR_n < 1.7) can be determined accordingly (Table 2). Clearly, the use of RSR_n (with the empirically determined cutoff value of 1.7) paints a less bright picture of the docking accuracy of all four methods than the use of RMSD.

Although the use of RSR_n values is a more objective measure of docking accuracy, one example showed that the use of RSR_n alone is no panacea. It appears that no single performance measure can capture all aspects that need to be considered when judging the quality of a docking pose: satisfactory explanation of the experimental data (captured by RSR_n) and satisfactory reproduction of the pivotal interactions between the ligand and its receptor. The latter point is captured only to some extent by the RMSD (and not always reliably so). The IBAC measure goes some way to addressing the issue of key-interaction recognition, but it seems nontrivial to automate its calculation. Although RMSD and RSR_n calculations can be automated, the sensitivity of the RSR-based measure toward small conformational differences is a limitation in the practical iterative improvement processes for docking programs and scoring functions.

In order to make a fair assessment of docking robustness that reflects the true performance of docking engines we suggest using combinations of assessment criteria with RSR_n as the primary objective measure.

CONCLUSIONS

The robustness of binding mode prediction in docking programs is an essential prerequisite for the calculation of meaningful docking scores and is therefore of major significance in docking evaluation routines applied during

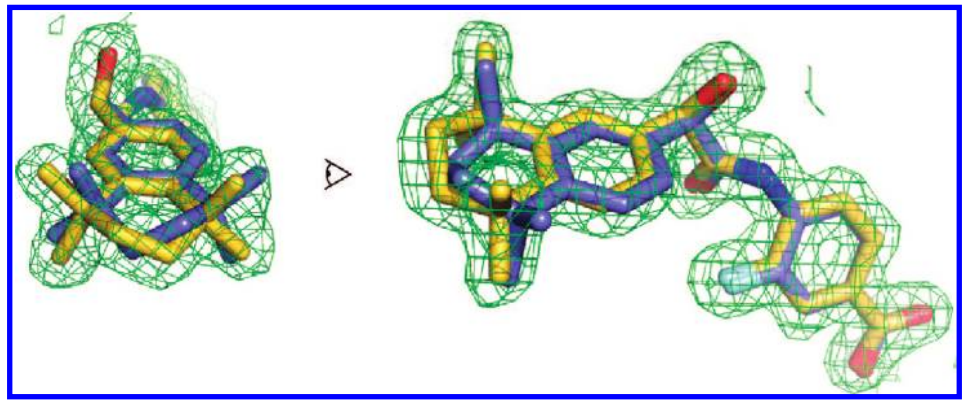

Figure 10. Example of a case in which a different ring conformation in a pose leads to a high RSR_n value. The crystal structure (yellow) and density (green mesh) for the PDB entry 1EXA is shown with the top-ranked pose from Glide (blue). The overall pose is in good agreement with the crystal structure (RMSD =0.4 Å). However, the comparatively small conformational differences in the methyl-substituted cyclohexenyl ring cause a mismatch of the ring and substituent atoms with the density, and this results in a high RSR_n value of 2.27.

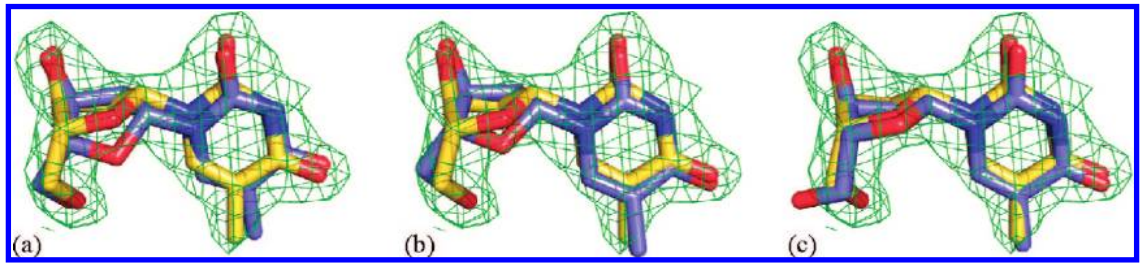

Figure 11. Example of a case in which different furanose-ring conformations in a pose lead to high RSR_n values. The ligand model from the crystal structure (yellow) and density (green mesh) of PDB entry 1KIM is shown together with the top-ranked poses (blue) suggested by the three docking programs used in this study. (a) The pose from Gold with Goldscore shows a significant mismatch of the furanose ring and minor structural differences (RMSD = 0.6 Å, RSR_n = 2.28). (b) The furanose ring conformation in the Glide pose is somewhat better but still does not match the density very well (RMSD = 0.5 Å, RSR_n = 1.93). (c) The top-ranked solution from Fred displays a ring conformation close to that of the crystal structure that fits the density well. However, the methylhydroxyl substituent does not fit very well leading to an overall RMSD of 0.6 Å and an RSR_n value of 1.91.

methodological development work in this field. Although the RMSD is a common choice for the assessment of docking robustness, it bears intrinsic limitations that can lead to misclassifications of both correct and incorrect poses. It further disregards the fundamental scientific principle that *predicted* data should be compared to *experimental* data when evaluating the predictive power of an approach. Our novel measure for docking pose prediction accuracy, the RSR_n value, abides by this scientific tenet by evaluating docking poses directly against the experimental electron density data from crystal structures. Although the ligand coordinates from the crystal structure, from which the RMSD values are calculated, are an interpretation of this electron density information, our results show that there is no clear correlation between the RSR-based assessment and the RMSD-based one. This emphasizes that RMSD and RSR_n calculations assess different aspects of docking results. After careful inspection of an exhaustive data set of results from different docking engines we determined an empirical cutoff value for the RSR-based assessment criterion which prevents many of the misclassifications of docking poses that are due to the intrinsic limitations of the RMSD measure. The RSR_n criterion also allows for alternative binding modes that fit the experimental electron density as well as the crystallographic model. There are only a few examples where a docking program generated a ligand configuration in the binding site that fits the electron density as well as or better than the "hand-crafted" model of the crystallographer. This can be attributed to the different drivers for the conformational sampling during structure determination and docking. In docking there is usually a bias toward low-energy conformations, whereas during crystal structure determination there is a bias toward fitting the electron density, which can lead to internal strain and high-energy ligand conformations.

It is important to realize that, even though RSR_n is a relative measure, it is still intrinsically resolution dependent. With low-resolution data, the density for a puckered ring, for instance, will be blob-like, and many different poses may fit it more or less equally well and, hence, yield similar (low) RSR_n values (see, *e.g.,* Figures 7 and 8). This is an inherent reflection of the experimental uncertainty associated with the placement, orientation, and conformation of the ligand. In an RMSD calculation, however, the experimental uncertainty is not taken into account leading to unjust rejection of poses that provide an alternative, but valid, interpretation of the density. At high resolution, on the other hand, the electron-density maps begin to show atomic detail, and the RSR_n and RMSD criteria will tend to give the same results. In some cases (for example the one shown in Figure 10), the RSR_n criterion may in fact become too sensitive and reject reasonable poses. This problem could be alleviated by making the cutoff value resolution-dependent, but this would require a much larger test data set. It could, however, also be taken as a challenge to the docking field to improve the methods so they produce "higher resolution" docking poses that look beyond the confines of the nearest local energy minimum.

Our results suggest that the use of electron density information considerably impacts the results of pose-prediction accuracy assessments, irrespective of the choice of

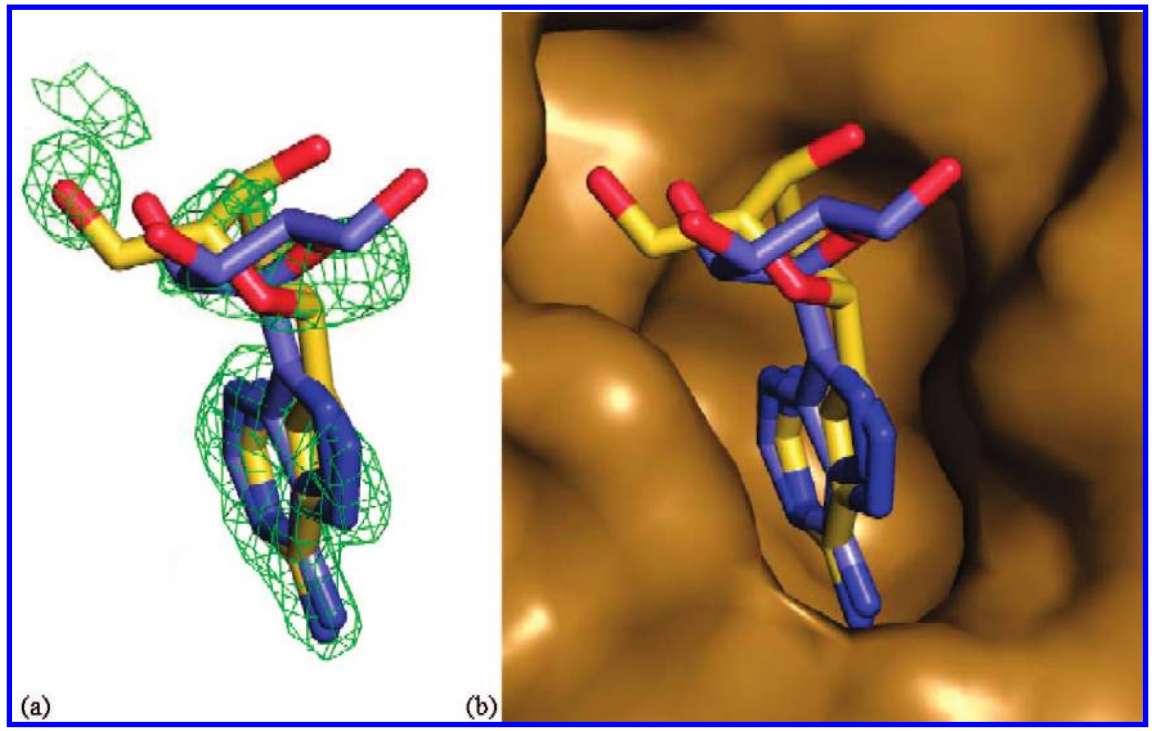

Figure 12. Example of a case with an RSR_n value below 1.0 ($RSR_n = 0.99$, RMSD = 2.3 Å). The ligand from the crystal structure of PDB entry 1MRK (yellow) is shown together with the fourth-ranked pose from the docking program Fred (blue), (a) together with the electron density (green mesh) and (b) in the context of the binding site (brown surface). The ligand model from the crystal structure and the docking pose differ in the conformation and orientation of the furanose ring. Both conformations are reasonable in the context of the binding site, but the electron density fit for the docking pose yields a slightly better RSR value.

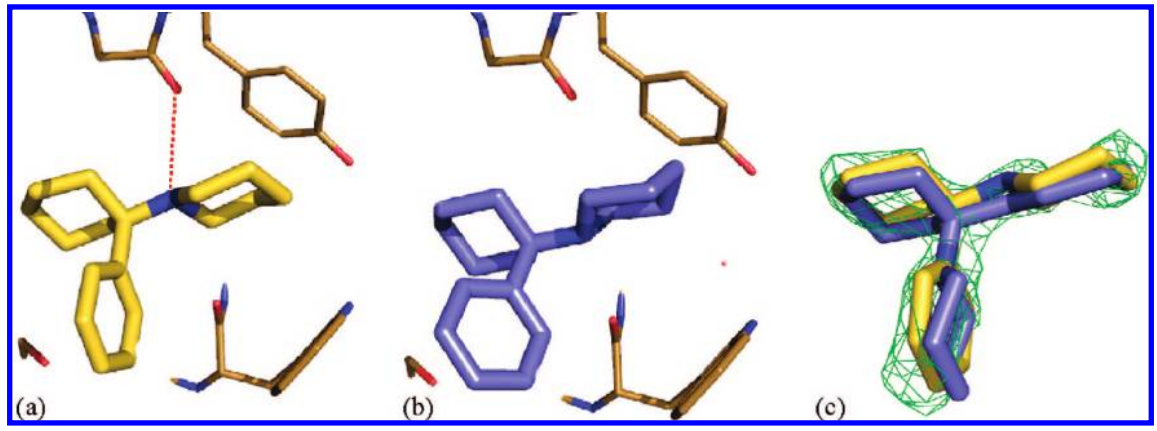

Figure 13. Example of a case in which both the RMSD and RSR_n criteria fail to recognize an incorrect docking pose. (a) Binding site and ligand model as present in the crystal structure (PDB entry 2PCP). The piperidine nitrogen atom forms a hydrogen bond with the protein (red dotted line). (b) Top-ranked solution from Glide. This pose has an RMSD of only 0.8 Å and an RSR_n value of 1.53. Hence, by both standards the docking solution is judged correct, even though the reported polar interaction of the piperidine-nitrogen[66] is not possible in this pose. (c) Top-ranked solution obtained with Fred (blue), crystal structure (yellow), and electron density (green mesh). The Fred docking places the ligand into the binding site rotated by ~120° (clockwise) resulting in a high RMSD (4.1 Å). This alternative placement is in good agreement with the density ($RSR_n = 1.34$). The top-ranked poses from Gold (with both Goldscore and Chemscore) are similar to this pose. Note that this alternative pose also precludes the formation of the reported polar interaction between the ligand and the protein.

Table 2. Fraction of Docking Poses Classified as Successful Using the Two Performance Measures RMSD and RSR_n for the Data Set of 88 Complex Structures Used in This Work

docking program	RMSD < 2.0 Å	RSR_n < 1.7
Gold (Goldscore)	53%	30%
Gold (Chemscore)	44%	28%
Glide (SP)	66%	40%
FRED (ShapeGauss)	43%	24%

docking engine. Nevertheless, given its considerable sensitivity in high resolution cases RSR_n should probably not be used as the *only* criterion for the assessment of pose prediction accuracy. We encountered one case of a top-ranked pose that both RMSD and RSR_n failed to recognize as incorrect. In that case, the use of IBAC[41] might have been more appropriate. A combination or consensus of different assessment criteria appears to be key to developing accurate and sensible docking evaluations. As docking and scoring methods become more and more sophisticated, there is a corresponding need for better evaluation benchmarks.

The use of RMSD is appealing for the assessment of pose prediction accuracy due to its ease of calculation and because it can be determined in an automated fashion when working with large data sets. The calculation of RSR_n values is

equally manageable in an automated fashion once the required raw data for an evaluation set is gathered. As of February 1, 2008, deposition of experimental data at the PDB is mandatory. This means that crystallographic data will be routinely available, and the derived electron-density maps can be readily retrieved from services such as EDS.[55,67] We strongly recommend the use of all of this information.

ACKNOWLEDGMENT

The authors thank Dr. Tove Sjögren (AstraZeneca, R&D Mölndal) for helpful discussions and Prof. T. Alwyn Jones (Uppsala University) for help with the RSR calculations in O as well as for modifying O. G.J.K. was funded by Uppsala University as well as through a Royal Swedish Academy of Sciences (KVA) Research Fellowship (supported through the Knut and Alice Wallenberg Foundation). Workflow and instructions on how to obtain the RSR_n values via the program O are available from the authors on request.

Supporting Information Available: Excel table with all raw data for the plots shown in Figures 2 and 3 and corresponding docking poses, PDB files, electron density files, and a script to browse these data using the program Pymol.[68] This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Congreve, M.; Murray, C. W.; Blundell, T. L. Keynote review: Structural biology and drug discovery. *Drug Discovery Today* **2005**, *10* (13), 895–907.
(2) Hardy, L. W.; Malikayil, A. The impact of structure-guided drug design on clinical agents. *Curr. Drug Discovery* **2003**, 15–20.
(3) Sotriffer, C.; Stahl, M.; Böhm, H.-J.; Klebe, G. Docking and Scoring Functions/Virtual Screening. In *Burger's Medicinal Chemistry and Drug Discovery*, 6th ed.; Wiley: New York, 2003; Vol. 1, pp 281−333.
(4) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7* (20), 1055.
(5) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem., Int. Ed.* **2002**, *41* (15), 2644–2676.
(6) Tame, J. R. H. Scoring functions: A view from the bench. *J. Comput.-Aided Mol. Des.* **1999**, *13* (2), 99–108.
(7) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Struct., Funct., Genet.* **2002**, *47* (4), 409–443.
(8) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261* (3), 470–489.
(9) Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15* (5), 411–428.
(10) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
(11) Totrov, M.; Abagyan, R. A. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins: Struct., Funct., Genet.* **1997**, *29* (S1), 215–220.
(12) McMartin, C.; Bohacek, R. S. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11* (4), 333–344.
(13) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19* (14), 1639–1662.
(14) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, A. J.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68* (1), 76–90.
(15) Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* **2001**, *44* (7), 1035–1042.
(16) Feher, M.; Deretey, E.; Roy, S. BHB: A Simple Knowledge-Based Scoring Function to Improve the Efficiency of Database Screening. *J. Chem. Inf. Model.* **2003**, *43* (4), 1316–1327.
(17) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295* (2), 337–356.
(18) 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* (5), 425–445.
(19) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42* (5), 791–804.
(20) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, P. W. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42* (25), 5100–5109.
(21) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Modell.* **2002**, *20* (4), 281–295.
(22) Zhang, Q.; Muegge, I. Scaffold Hopping through Virtual Screening Using 2D and 3D Similarity Descriptors: Ranking, Voting, and Consensus Scoring. *J. Med. Chem.* **2006**, *49* (5), 1536–1548.
(23) Lyne, P. D.; Kenny, P. W.; Cosgrove, D. A.; Deng, C.; Zabludoff, S.; Ashwell, S.; Wendoloski, J. J. Identification of Compounds with Nanomolar Binding Affinity for Checkpoint Kinase-1 Using Knowledge-Based Virtual Screening. *J. Med. Chem.* **2004**, *47* (8), 1962–1968.
(24) Nicodème, P.; Rognan, D. ConsDock: A new program for the consensus analysis of protein-ligand interactions. *Proteins: Struct., Funct., Genet.* **2002**, *47* (4), 521–533.
(25) Miteva, M. A.; Lee, W. H.; Montes, M. O.; Villoutreix, B. O. Fast Structure-Based Virtual Ligand Screening Combining FRED, DOCK, and Surflex. *J. Med. Chem.* **2005**, *48* (19), 6012–6022.
(26) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins: Struct., Funct., Genet.* **1998**, *33* (3), 367–382.
(27) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759.
(28) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.
(29) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: Pharmacophore-Based Protein-Ligand Docking. *J. Med. Chem.* **2004**, *47* (27), 6804–6811.
(30) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing Scoring Functions for Protein-Ligand Interactions. *J. Med. Chem.* **2004**, *47* (12), 3032–3047.
(31) Perez, C.; Ortiz, A. R. Evaluation of Docking Functions for Protein-Ligand Docking. *J. Med. Chem.* **2001**, *44* (23), 3768–3785.
(32) 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* (4), 962–976.
(33) Perola, E.; Walters, P. W.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins: Struct. Funct. Bioinform.* **2004**, *56* (2), 235–249.
(34) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins: Struct., Funct., Genet.* **2003**, *52* (4), 609–623.
(35) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. *J. Chem. Inf. Model.* **2006**, *46* (1), 401–415.
(36) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.* **2004**, *47* (3), 558–565.
(37) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46* (12), 2287–2303.
(38) 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* (20), 5912–5931.
(39) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11* (13−14), 580–594.
(40) Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScore^{CSD}-Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *J. Med. Chem.* **2005**, *48* (20), 6296–6303.
(41) Kroemer, R. T.; Vulpetti, A.; McDonald, J. J.; Rohrer, D. C.; Trosset, J.-Y.; Giordanetto, F.; Cotesta, S.; McMartin, C.; Kihlen, M.; Stouten, P. F. W. Assessment of Docking Poses: Interactions-Based Accuracy

Classification (IBAC) versus Crystal Structure Deviations. *J. Chem. Inf. Model.* **2004**, *44* (3), 871–881.
(42) Kleywegt, G. J. Validation of protein crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56* (3), 249–265.
(43) Brändén, C.-I.; Jones, A. T. Between objectivity and subjectivity. *Nature* **1990**, *343* (6260), 687–689.
(44) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem., Int. Ed.* **2003**, *42* (24), 2718–2736.
(45) Kleywegt, G. J.; Henrick, K.; Dodson, E. J.; van Aalten, D. M. F. Pound-Wise but Penny-Foolish: How Well Do Micromolecules Fare in Macromolecular Refinement. *Structure* **2003**, *11* (9), 1051–1059.
(46) Kleywegt, G. J. Crystallographic refinement of ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63* (1), 94–100.
(47) 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: Struct., Funct., Genet.* **2002**, *49* (4), 457–471.
(48) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. Improved methods for building protein modles in electron density maps and the locations of errors in these models. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, *47*, 110–119.
(49) GOLD, 2.0; CCDC: Cambridge, U.K., 2005.
(50) GLIDE; Schrödinger LLC: New York, 2005.
(51) Fred, 2.1; Openeye Scientific Software: Santa Fe, NM, 2005.
(52) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
(53) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
(54) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337* (1), 209–225.
(55) Kleywegt, G. J.; Harris, M. R.; Zou, J.-Y.; Taylor, T. C.; Wahlby, A.; Jones, T. A. The Uppsala Electron-Density Server. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60* (12−1), 2240–2249.
(56) Sybyl, 7.1; Tripos Inc.: St. Louis, MO, 2002.
(57) Maestro, 7.0.113; Schrödinger LLC: New York, 2005.
(58) Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. Relibase: Design and Development of a Database for Comprehensive Analysis of Protein-Ligand Interactions. *J. Mol. Biol.* **2003**, *326* (2), 607–620.
(59) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodology* **1990**, *3* (6, Part 3), 537–547.
(60) Omega, 1.8.1; Openeye Scientific Software: Santa Fe, NM, 2005.
(61) Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1986**, *42*, 140–149.
(62) Collaborative Computational Project Nr 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **1994**, *50* (5), 760–763.
(63) Boström, J.; Norrby, P.-O.; Liljefors, T. Conformational energy penalties of protein-bound ligands. *J. Comput.-Aided Mol. Des.* **1998**, *12* (4), 383–396.
(64) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. A. Conformational changes of small molecules binding to proteins. *Biorg. Med. Chem.* **1995**, *3* (4), 411–428.
(65) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.* **2004**, *47* (10), 2499–2510.
(66) Lim, K.; Owens, S. M.; Arnold, L.; Sacchettini, J. C.; Linthicum, D. S. Crystal Structure of Monoclonal 6B5 Fab Complexed with Phencyclidine. *J. Biol. Chem.* **1998**, *273* (44), 28576–28582.
(67) Electron Density Server. http://eds.bmc.uu.se/(accessed Apr 1, 2006).
(68) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: Palo Alto, CA, 2002.

CI800084X