Published in final edited form as:
J Comput Aided Mol Des. 2012 June ; 26(6): 675–686. doi:10.1007/s10822-012-9547-0.

Docking and scoring with ICM: the benchmarking results and strategies for improvement

Marco A. C. Neves,
Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. Centro de Neurociências, Lab. Química Farmacêutica, Faculdade de Farmácia, Universidade de Coimbra, Pólo das Ciências da Saúde, 3000-548 Coimbra, Portugal

Maxim Totrov, and
Molsoft L.L.C. 11199 Sorrento Valley Road, S209, San Diego, CA 92121, USA

Ruben Abagyan
Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. Molsoft L.L.C. 11199 Sorrento Valley Road, S209, San Diego, CA 92121, USA

Ruben Abagyan: rabagyan@ucsd.edu

Abstract

Flexible docking and scoring using the Internal Coordinate Mechanics software (ICM) was benchmarked for ligand binding mode prediction against the 85 co-crystal structures in the modified Astex data set. The ICM virtual ligand screening was tested against the 40 DUD target benchmarks and 11-target WOMBAT sets. The self-docking accuracy was evaluated for the top 1 and top 3 scoring poses at each ligand binding site with near native conformations below 2 Å RMSD found in 91% and 95% of the predictions, respectively. The virtual ligand screening using single rigid pocket conformations provided the median area under the ROC curves equal to 69.4 with 22.0% true positives recovered at 2% false positive rate. Significant improvements up to ROC AUC= 82.2 and $ROC_{(2\%)}$= 45.2 were achieved following our best practices for flexible pocket refinement and out-of-pocket binding rescore. The virtual screening can be further improved by considering multiple conformations of the target.

Keywords

Docking; Scoring; Virtual ligand screening; Structure-based drug design; ICM; Internal coordinate mechanics

Introduction

Over the past decade the number of protein structures solved by X-ray crystallography increased by more than four times allowing the use of atomic details derived from high resolution target structures as a standard procedure in many drug discovery projects. Structure-based drug design is typically used at several stages of the drug discovery and development pipeline, such as for the virtual ligand screening (VLS) of large electronic compound databases and initial hit identification, and for the lead optimization of the most promising drug candidates into new potent, selective and drug-like chemical entities [1]. Docking-based methods rely on computer algorithms to generate possible small molecule binding modes within a protein pocket and appropriate scoring functions to estimate the strength of the protein-ligand interaction. Binding poses generated at each docking run

sample translational, rotational and conformational degrees of freedom of the ligand (semi-flexible docking), and in some cases, additional conformational degrees of freedom of protein residues within the ligand binding site (fully-flexible docking). The multiple ligand poses are then ranked by a scoring function in an attempt to identify the most energetically favorable protein-bound conformation. Virtual screening of multiple ligands to the same protein site uses docking and scoring to generate a ranked list of compounds. While most software packages are able to predict experimental poses with reasonable accuracy, binding affinity predictions for a diverse set of molecules followed by ranking their predicted potencies is a much more challenging endeavor. The reasons behind this scoring problem have been well described in recent reviews [2, 3], and typically pointed out to protein flexibility and induced-fit upon ligand binding [4], oversimplification of energy terms commonly used in the scoring functions such as solvation [5] and entropy contributions [6], and specific non-covalent interactions not commonly present in scoring functions such as cation-π [7], $n \rightarrow \pi^*$ [8] and weak hydrogen bonds [9]. Prediction of the most stable tautomeric states for the ligands [10], water-mediated hydrogen-bonds [11], metal coordination [12, 13] and identification of pKa-based protonation microspecies [14] are additional challenges that need to be solved before further progress can be made in the field.

Because of the multiple approximations used in docking and scoring, benchmarking software packages is critically important in structure-based drug discovery. New algorithms are typically trained against a large number of high quality co-crystal structures collected from the PDB database [15] and evaluated for their ability to reproduce known binding modes. On the other hand, retrospective virtual screening benchmarks use test sets with annotated known binders for particular targets and presumed inactive decoys. Graphical representation of the true positive rate versus the false positive rate in receiver operating characteristic (ROC) plots is assumed to provide a measure of both the software performance and the ligand binding pocket quality for virtual screening purposes. Because the top scoring compounds will most likely be selected for further testing, early enrichment measures are important to identify successful VLS runs where active compounds are ranked ahead of decoys on a large score list [16].

The Internal Coordinate Mechanics (ICM) method has been extensively validated in bioinformatics and drug discovery projects. Prospective and retrospective studies demonstrate that ICM is able not only to reconstitute the most critical protein-ligand contacts, but also to successfully identify high affinity ligands for the most important classes of targets in drug discovery, such as enzymes [17–24], receptors [25–30], ion channels [31, 32] and transport proteins [33–35]. As part of the programmatic theme, “Docking and Scoring: A Review of Docking Programs”, ICM was benchmarked for self-docking accuracy and virtual ligand screening using the Astex [36], DUD [37] and WOMBAT [38] test sets, and the benchmarking results, current developments and future prospects were presented during a symposium at the 241st ACS National Meeting held in March 2011 in Anaheim, CA. This article summarizes the results discussed in Anaheim, adds calculation details, and proposes how to improve the performance further using our best practices.

Computational methods

The ICM method

Flexible ligand docking with the ICM software uses Monte Carlo simulations to globally optimize a set of ligand internal coordinates in the space of grid potential maps calculated for the protein pocket [39], according to the following procedure: (1) a random move is introduced to one of the rotational, translational or conformational variables of the ligand within the binding pocket; (2) differentiable terms of the energy function are minimized; (3) desolvation energy is calculated; (4) the Metropolis selection criterion is used to accept or

reject the final minimized conformation [40] and the procedure is repeated until the maximal number of steps is achieved. An adaptive algorithm is used to determine the maximal number of steps, according to the number of rotatable bonds in the ligand multiplied by a user-defined thoroughness value. The thoroughness value represents the user-defined multiplier by which the automatically determined Monte Carlo run length is extended. Ligand sampling in a set of pre-calculated grid maps accounting for hydrogen bonding potential, van der Waals potential with carbon-, sulphur- and hydrogen-like probes, hydrophobic potential and electrostatic potential, significantly reduces the time required for calculation. The maps are generated in a rectangular box with 0.5 Å grid spacing centered at the ligand binding site. Each molecule is first submitted to a conformational analysis outside of the protein pocket and a stack of low energy conformations is collected and used as starting geometries for the grid docking. Ligand binding modes are scored according to the quality of the complex and a user-defined number of the top scoring poses is re-ranked using the full ICM scoring function. The predicted score is calculated as the weighted (α_1 to α_5) sum of ligand-target van der Waals interactions and internal force field energy of the ligand ($\Delta E_{\mathrm{IntFF}}$), free energy changes due to conformational energy loss upon ligand binding ($T\Delta S_{\mathrm{Tor}}$), hydrogen bonding interactions ($\Delta E_{\mathrm{HBond}}$), hydrogen bond donor-acceptor desolvation energy ($\Delta E_{\mathrm{HBDesol}}$), solvation electrostatic energy upon ligand binding ($\Delta E_{\mathrm{SolEl}}$), hydrophobic free energy gain ($\Delta E_{\mathrm{HPhob}}$) and a size correction term proportional to the number of ligand atoms (Q_{Size}) [28, 41, 42]:

$$\Delta G=\Delta E_{\mathrm{IntFF}}+T\Delta S_{\mathrm{Tor}}+\alpha_1\Delta E_{\mathrm{HBond}}+\alpha_2\Delta E_{\mathrm{HBDesol}}+\alpha_3\Delta E_{\mathrm{SolEl}}+\alpha_4\Delta E_{\mathrm{HPhob}}+\alpha_5 Q_{\mathrm{Size}}$$

Binding mode prediction

The Astex diverse set provides a collection of 85 high-resolution protein structures from representative target families co-crystalized with drug-like small molecules [36]. The files used in this study were provided by the thematic ACS meeting organizers and prepared for docking with ICM version 3.7-2b using a small script written in the ICM language. For each target:

1. A protein structure with co-factors, metals and other co-crystal molecules (excluding crystallographic waters) was imported into ICM and ligand binding sites were enumerated using the co-crystal ligands as reference, provided in a separate file. Multiple sites per ligand were found in several X-ray structures due to either oligomeric nature of the target or multiple copies of the protein within the asymmetric crystallographic unit.
2. Alternative side chain conformations for residues within a 5 Å cut-off distance from the co-crystal ligands were enumerated and saved. Alternatives A and B were saved for structures with two alternative side chain conformations of a single binding site residue. Alternatives AA, BB, AB and BA were saved when two binding site residues with alternative side chain conformations were found. Because of the multiple sites per ligand and alternative side chain conformations, the total number of pockets considered for docking was larger than the initial number of protein structures.
3. Amino acid protonation states were used as provided. Partial charges of protein atoms were taken from a library of ECEPP/3 residue templates (icm.res).
4. MMFF atom types and partial charges were assigned to the ligands [43] in a separate file, using the starting conformations independent of the co-crystal geometries as provided by the organizers.

5. For each site, the ligand binding pocket for docking was defined as a rectangular box centered at the co-crystal molecule and extending additional 4 Å in any direction.

6. Grid potential energy maps accounting for hydrophobic interactions, van de Walls interactions, hydrogen bonding and electrostatic potential were calculated within the ligand binding sites.

7. Ten independent docking runs were performed with the length of the docking simulation adjusted by the default thoroughness value (thoroughness=1). Five top quality poses were rescored at the end of each run using the default ICM scoring function. Calculations took an average of 10 seconds per run using a Linux workstation (Intel Core i7 Processor 3.07 GHz, 12 GB RAM) running Fedora 12.

   Alternative docking solutions for each molecule were loaded into ICM, sorted by predicted binding score and clustered to remove redundancy, i.e. identical conformations where the ligand was bound with less than a 0.5 Å symmetry corrected RMSD cutoff calculated for heavy atoms. Briefly, the second conformation in the ranked score list was considered unique if the RMSD value was less than 0.5 Å, as compared to the first conformation, otherwise it was deleted. This procedure was repeated for the third and consecutive conformations of the same molecule, using the pool of unique conformations as reference, until all similar conformations were removed. Symmetry-corrected RMSD was used to account for topological symmetries of chemical groups (e.g. equivalent atoms in benzene rings or negatively charged carboxylic acids), as well as three-dimensional symmetries generated by rotation.

8. As a measure of binding mode prediction accuracy, symmetry-corrected RMSD was calculated between each unique binding pose found with the ICM software and the co-crystal ligand. Statistical analysis was performed using RMSD values found for the top scoring poses in the ranked score lists (top 1) and the lowest RMSD value among the top 3 unique scoring poses for each site (top 3). Both symmetry corrected RMSD and contact measures were also used in the GPCRdock evaluation [44] and are publically accessible at the laboratory web server Simicon, http://abagyan.ucsd.edu/SimiCon/[45].

Virtual ligand screening

The directory of useful decoys (DUD), a benchmarking test set freely available on the internet (http://dud.docking.org/) [37], covers 40 different targets with 3950 active ligands and 36 decoys per active. Decoys have similar molecular weight, number of hydrogen bonding groups, logP and number of rotatable bonds compared to the active compounds, but their molecular topologies are different. DUD is currently the largest, and one of the most challenging test sets for virtual ligand screening benchmarks, however, because a significant number of the true binders are very close analogs, it has limited usefulness for scaffold hopping evaluations [38]. The WOMBAT data sets derived from DUD by filtering non-lead-like compounds, clustering unique chemotypes and expanding the original test sets with additional active compounds, provide an alternative benchmark for a sub-set of 11 DUD targets [38].

Virtual ligand screening performance using the ICM method was evaluated with modified versions of the DUD and WOMBAT test sets provided by the ACS meeting organizers. Protein targets were prepared for docking using a similar procedure described above. Ligand binding pockets were initially defined as a rectangular box centered on a known co-crystal ligand, provided with the test set, and extended additional 4 Å in any direction. However,

given the large diversity of ligands and protein pockets, box dimensions were subsequently adjusted manually after visual inspection of the protein-ligand complex and the list of known binders. For each target, active compounds and decoys were grouped into an annotated sdf file and used as provided. Semi-randomized pairings of small molecules and targets were used to detect potential chemical bias on the DUD test sets that might contribute for discrimination based on anything other than protein-ligand contacts. Target sites of approximately the same size but different families were matched as described by McGann [46] and their ligands included in the annotated sdf file for docking. For example, cyclooxygenase 2 actives and decoys were docked against neuraminidase, and vice-versa. WOMBAT ligands for a subset of 11 DUD targets were also included in this study. Three independent docking runs were performed with a thoroughness value set to 1 and the top 5 best quality complexes rescored. The top scoring pose was selected for each compound.

For each target the true positive rate was plotted as a function of the false positive rate for all positions of the ranked score list. Binding pockets with perfect discrimination, i.e. scoring all true positives at the top ranked positions, have ROC plots that pass through the upper left corner and area under the curve (AUC) equal to 100. Therefore the higher the AUC value in a ROC curve, the better the discrimination. Because successful VLS ranks active compounds early on a large score list, the fraction of actives recovered at 0.1%, 1% and 2% decoys recovered (abbreviated to $ROC_{(0.1\%)}$, $ROC_{(1\%)}$ and $ROC_{(2\%)}$) were used in this study as early recognition metrics. AUCs, early enrichment values and statistical analysis were calculated with appropriate ICM macros. Because there are no decoys available with the WOMBAT test sets, enrichment studies were performed using DUD decoys derived for the same targets. This approach allows comparison between benchmarks of different software packages but introduces some chemical bias to the results because, contrarily to DUD, WOMBAT actives are predominately lead-like molecules with lower molecular weights and logP, as well as less hydrogen bonding groups.

Best practices: soft out-of-pocket penalty function and ligand binding site optimization

Virtual ligand screening with ICM was further improved using our best practices for binding rescore and induced fit analysis. The default ICM scores were rescored using predicted out-of-pocket protein-ligand contacts according to the following penalty function:

$$Score' = Score + \left(1 - \frac{C_{top10} \cap C_{ligand}}{C_{ligand}}\right) \times \left(\frac{\overline{A}_{actives}}{A_{ligand}}\right)^2 \times 100$$

Near-native interactions important for productive binding were defined as the superset of contacts between the top 10 scoring actives (C_{top10}) and protein atoms within a 3.5 Å cutoff. Contacts beyond this distance were considered to be out-of-pocket. For each ligand in the test set, the total number of contacts within a 3.5 Å cutoff (C_{ligand}) was compared with C_{top10} to derive a subset of common protein-ligand interactions ($C_{top10} \cap C_{ligand}$). The fraction between common contacts and total contacts was then used in a function to increase the original ICM score whenever the molecule binds extensively to a new surrounding area. However, the penalty applied gradually decreases as the overlap with binding modes of known actives increases. Because large compounds have higher propensity to occupy additional surrounding areas of the pocket, a scaling factor was introduced based on the fraction between the average number of non-hydrogen atoms of active compounds in the test set ($\bar{A}_{actives}$) and the number of non-hydrogen atoms of each ligand in the test set (A_{ligand}). Therefore, the function assigns a larger penalty for out-of-pocket binding of small ligands than compounds with more heavy atoms.

Induced fit upon small molecule binding was simulated by fully-flexible ligand and protein docking. To this purpose, crystallographic coordinates of a single co-crystal ligand per target, provided with the DUD test sets, were used. For each target, residues with side chain atoms within a 5 Å cutoff from the seed ligand were allowed to randomly move using the ICM Biased Probability Monte Carlo algorithm, followed by full local energy minimization [42]. Geometrically diverse low-energy conformations were saved in a conformational stack. The number of alternative conformations generated during this procedure was largely dependent on the pocket plasticity. The maximal number of structures in the conformational stack was set to 300 by adjusting the maximum angular RMSD per variable when two structures are still considered belonging to the same cluster. Calculations took less than 60 minutes per pocket in all cases using a Linux workstation (Intel Core i7 Processor 3.07 GHz, 12 GB RAM) running Fedora 12. The files were prepared for docking using the VLS procedure described before. For each new conformation, grid potential maps were calculated and three independent docking runs were performed with a thoroughness value set to 1 and the top 5 best quality complexes rescored.

Results and Discussion

Ligand binding mode prediction

The 85 X-ray co-crystal structures used in this study provide a representative ensemble of non-redundant ligand binding pockets for self-dock benchmarks. Detailed statistical analysis on ligand binding accuracy using the ICM method is provided in Table 1, whereas individual RMSD values for predictions on every ligand binding site are shown in Fig. 1.

ICM was found to predict the top 1 scoring poses below 2 Å RMSD in 91% of the sites with an average RMSD of 0.91 Å (median= 0.54 Å). Predictions below 1 Å and below 0.5 Å were found in 78% and 43% of the cases, respectively. When the lowest RMSD value among the top 3 solutions in the ranked score list was considered, near native conformations below 2 Å RMSD were found in 95% of the sites with an average RMSD value of 0.67 Å (median= 0.48 Å). In all cases, the highest RMSD prediction among the top 3 scoring poses was below 3.8 Å.

The top 1 scoring solutions provided the lowest RMSD among the top 3 in 76% of the sites. However, with PDB entries 1jje, 1gpk, 1sq5, 1gm8 and 1meh, illustrated in Fig. 2, predicted binding modes considering the top 3 ensemble provided geometries significantly closer to the co-crystal ligand. PDB entry 1jje (top 1 RMSD= 8.2 Å, top 3 RMSD= 0.4 Å, Fig. 2A) provided the highest RMSD value for the top 1 predictions in the whole set. Visual inspection of the protein-ligand complex revealed a 180° flip where most native contacts are captured. Because the ligand is highly symmetric, benzyl moieties and the succinyl acid group are perfectly aligned with the co-crystal structure. Furthermore, analysis of crystal packing interactions revealed the presence of charged residues Lys8 and Asp10 in the ligand binding pocket vicinity with stabilizing effect over the co-crystal binding mode. Highly symmetric protein-ligand hydrophobic interactions are also involved in a 180° flip found with the top 1 prediction of PDB entry 1gpk (top 1 RMSD= 3.6 Å, top 3 RMSD= 0.3 Å, Fig. 2B). The co-crystal ligand is additionally stabilized by a water-mediated hydrogen bond involving the amide nitrogen, Glu199 and Gly117. Water mediated hydrogen bonds between co-crystal ligands and at least one protein residue are also found in PDB entries 1gm8 (top 1 RMSD= 3.0 Å, top 3 RMSD= 1.5 Å, Fig. 2C) and 1meh (top 1 RMSD= 2.4 Å, top 3 RMSD= 0.6 Å, Fig. 2D). The β-lactam ring in PDB entry 1gm8 interacts with Ser386 through a water-mediated hydrogen bond and the contact is correctly predicted with the top 1 scoring pose. However, because water molecules were excluded from the calculation, the penicillin core is translated by approximately 3 Å (Fig. 2C). On the other hand, the heterocyclic moiety of mycophenolic acid in PDB entry 1meh is correctly predicted but the

highly flexible aliphatic chain adopts an alternative conformation. The native binding mode is stabilized by hydrogen bonding contacts between the carboxylic acid and Ser263, and a water-mediated hydrogen bond involving Asp261. For this ligand ICM predicts a non-native charged interaction with Arg414 (Fig. 2D).

In a few cases ICM failed to predict the co-crystal binding mode below 2 Å RMSD within the top 3 scoring poses. PDB entry 1jd0 is predicted at 3.8 Å RMSD because of an alternative bi-dentate metal coordination geometry with zinc between the nitrogens of the sulfonamide and the thiadiazole ring (Fig. 2E). Furthermore, the native binding mode is stabilized by a water-mediated hydrogen bond between the thiadiazole ring and Pro201. Binding to multiple sites of PDB entry 1hvy is predicted at RMSD values between 2.2 Å and 2.8 Å. ICM provides good predictions for the heterocyclic ring moieties but the highly flexible aliphatic chain, stabilized in the X-ray co-crystal structure by a complex network of water-mediated hydrogen bonds between the terminal carboxylic acids and Lys77, Leu221, Ile307 and Met309, adopts an alternative conformation (Fig. 2F). Similar self-docking inaccuracies have been reported with PDB entries 1jje, 1hvy, 1gm8, 1jd0 and 1sq5 using alternative docking and scoring methods [36, 47–49].

ICM predicted similar binding poses for multiple sites of the same ligand, providing final docked conformations independent of the initial site selection. The average difference between the highest and the lowest RMSD predictions for multiple sites of the same ligand was 0.4 Å, and only 3 cases of RMSD differences above 2 Å were found, i.e. PDB entries 1w1p, 1tz8 and 1sq5 (Fig. 2). Out of two sites in PDB entry 1w1p, only one ICM prediction matches to the co-crystal binding mode. Visual inspection revealed the presence of glycerol molecules in only one of the pockets, establishing additional favorable interactions that correctly align the ligand, such as a hydrogen bond between the amide nitrogen and one of the primary glycerol hydroxyls and hydrophobic contacts between the ligand and the aliphatic chain of the second glycerol. The alternative binding mode predicted for the second site corresponds to a 180° flip (RMSD= 2.9 Å, Fig. 2G) that is fully compatible with the available electron density. Ligand binding to PBD entry 1tz8 was predicted at 2.9 Å RMSD in one out of 3 possible sites. However, analysis of symmetry related neighbors revealed an alternative ligand binding conformation that is predicted as top 1 scoring pose (RMSD= 1.3 Å, Fig. 2H). One out of four sites in PDB entry 1sq5 is incorrectly predicted at the top 1 scoring pose (top 1 RMSD= 3.3 Å, top 3 RMSD= 0.7 Å). Residue conformations are well conserved among different sites but a slight rotation of the Cα-Cβ bond in His177 causes a small displacement of the imidazole ring (0.4 Å translation of the ε^2 nitrogen) that favors a strong non-native hydrogen bond with the top 1 prediction.

In all cases ICM provided identical binding mode predictions below 1.5 Å RMSD when alternative side chain conformations for residues located within the binding pocket were considered. On average, selecting the alternative A provided slightly better predictions than alternative B or any combination of alternatives A and B for cases where two variable residues were present.

Alternative binding modes available for ligands in PDB entries 1ig3, 1sg0 and 1tz8 were predicted below 2 Å RMSD within the top 3 scoring poses, as described previously for 1tz8 (Fig. 2H). This result highlights the good sampling performance of the ICM docking algorithm.

Virtual ligand screening

Virtual screening with the ICM method was benchmarked against 40 protein targets using single ligand binding site conformations and the DUD and WOMBAT test sets. Statistical analyses of the results obtained are provided in Table 2, Table 3 and Table S1. Individual

values of ROC AUC, $ROC_{(0.1\%)}$, $ROC_{(1\%)}$ and $ROC_{(2\%)}$ are reported in Fig. 3, Fig. 4 and Fig. S1.

Docking DUD test sets against the original ligand binding site coordinates provided by the ACS meeting organizers and scoring with the default ICM function provided good results for a significant number of targets. On average, the AUC in linear ROC plots was 71.6 (median= 69.4, Table 2A). However, individual performances were largely target-dependent (Fig. 3A.). Neuraminidase (na, AUC= 95.8), peroxisome proliferator activated receptor γ (ppar, AUC= 94.5), epidermal growth factor receptor kinase (egfr, AUC= 92.9), trypsin (AUC= 92.6) and glycinamide ribonucleotide transformylase (gart, AUC= 92.1) were among the best performing targets, whereas platelet derived growth factor receptor kinase (pdgfrb, AUC= 27.0), P38 mitogen activated protein kinase (p38, AUC= 41.4), glucocorticoid receptor (gr, AUC= 41.8) and fibroblast growth factor receptor kinase (fgfr1, AUC= 46.7) performed close to, or worse than random (Fig. 3A). As expected, semi-randomized target/ligand test sets for the null hypothesis testing provided ROC AUC values close to random (AUC_{null}= 46.7 on average, median= 43.4) suggesting minimal chemical bias for most of the test sets. However, exceptions were found when dihydrofolate reductase (dhfr) was used to dock glycinamide ribonucleotide transformylase molecules (gart), when the mineralocorticoid receptor (mr) was used to dock estrogen receptor agonists (er_agonist) and when thrombin was used to dock factor Xa compounds (fxa). These 3 cases provided abnormally high ROC AUC_{null} values of 85.1, 83.1 and 87.4, respectively (Fig. 3A). Factor Xa and thrombin are trypsin-like serine proteases with similar catalytic domains and many ligands are cross-active indeed [50, 51]. On the other hand, the nuclear receptors for estrogens and mineralocorticoids bind to a similar steroid scaffold.

True positive rates $ROC_{(0.1\%)}$, $ROC_{(1\%)}$ and $ROC_{(2\%)}$ at 0.1%, 1% and 2% false positive rates were calculated as a metric of early enrichment. On average the ICM method provided good early enrichments identifying 7.3%, 21.0% and 26.6% of true positives (median= 3.8%, 14.8% and 22.0%), respectively, using the original pocket coordinates and the default scoring method (Table 2A). Also in this case, individual values were highly target-dependent (Fig. 3A), being maximal with the peroxisome proliferator activated receptor γ (ppar, $ROC_{(2\%)}$= 77.7%), neuraminidase (na, $ROC_{(2\%)}$= 77.6%) and glycogen phosphorylase β (gpb, $ROC_{(2\%)}$= 65.2%) and minimal with thymidine kinase (tk, $ROC_{(2\%)}$= 0%), human heat shock protein 90 (hsp90, $ROC_{(2\%)}$= 2.7%) and acetylcholine esterase (ache, $ROC_{(2\%)}$= 2.8%).

Given the fast computation time, rescoring default docking results is a popular approach to improve chemical recognition. Here we propose a simple rescoring approach based on a soft out-of-pocket penalty described in detail in the experimental section. Improvements in ROC AUC and early enrichment measures were found with most targets (Table S1A and Fig. S1A), such as with acetylcholine esterase (ache, $AUC_{initial}$= 67.8, $AUC_{rescored}$= 77.8, $ROC_{(2\%)initial}$= 2.8%, $ROC_{(2\%)rescored}$= 49.5%), adenosine deaminase (ada, $AUC_{initial}$= 50.3, $AUC_{rescored}$= 72.3, $ROC_{(2\%)initial}$= 7.7%, $ROC_{(2\%)rescored}$= 25.6%), poly(ADP-ribose) polymerase (parp, $AUC_{initial}$= 84.6, $AUC_{rescored}$= 90.5, $ROC_{(2\%)initial}$= 17.1%, $ROC_{(2\%)rescored}$= 57.1%) and thymidine kinase (tk, $AUC_{initial}$= 74.7, $AUC_{rescored}$= 80.1, $ROC_{(2\%)initial}$= 0%, $ROC_{(2\%)rescored}$= 36.4%).

Energy-based refinement accounting for pocket plasticity upon ligand binding provided consistent improvements for most DUD targets in terms of ROC AUC and early enrichments (Table S1B and Fig. S1B). Examples include catechol-*O*-methyltransferase (comt, $AUC_{initial}$= 83.0, $AUC_{refined}$= 85.7, $ROC_{(2\%)initial}$= 9.1%, $ROC_{(2\%)refined}$= 54.6%), cyclooxygenase 2 (cox2, $AUC_{initial}$= 75.1, $AUC_{refined}$= 82.1, $ROC_{(2\%)initial}$= 7.0%, $ROC_{(2\%)refined}$= 38.5%), fibroblast growth factor receptor kinase (fgfr1, $AUC_{initial}$= 46.7,

$AUC_{refined}$= 73.2, $ROC_{(2\%)initial}$= 7.5%, $ROC_{(2\%)refined}$= 30.0%), poly(ADP-ribose) polymerase (parp, $AUC_{initial}$= 84.6, $AUC_{refined}$= 92.9, $ROC_{(2\%)initial}$= 17.1%, $ROC_{(2\%)refined}$= 77.1%), progesterone receptor (pr, $AUC_{initial}$= 62.0, $AUC_{refined}$= 68.7, $ROC_{(2\%)initial}$= 11.1%, $ROC_{(2\%)refined}$= 37.0%), tyrosine kinase SRC (src, $AUC_{initial}$= 69.3, $AUC_{refined}$= 82.7, $ROC_{(2\%)initial}$= 13.8%, $ROC_{(2\%)refined}$= 44.0%) and thymidine kinase (tk, $AUC_{initial}$= 74.7, $AUC_{refined}$= 79.7, $ROC_{(2\%)initial}$= 0%, $ROC_{(2\%)refined}$= 54.6%). Because a single seed ligand was used for the induced fit modeling of each target (co-crystal ligand provided by the ACS meeting organizers), the conformational sampling was somehow limited. Better performances are expected for this method if more co-crystal ligands are used.

Optimal virtual ligand screening performance was obtained combining pocket refinement with out-of-pocket rescoring (Table 2B, Fig. 3B). An average ROC AUC of 79.4 (median= 82.2) with 24.7%, 39.5% and 44.3% of true positives (median= 20.4%, 37.0% and 45.2%) identified at 0.1%, 1% and 2% false positive rates, respectively, was obtained using this combined approach. Moreover, ROC AUC values above 70 and $ROC_{(2\%)}$ above 20 were found in 78% of the targets, being potential good candidates for virtual screening. While for most targets ligand recognition was improved after pocket refinement and out-of-pocket penalty, worse than random models such as pdgfrb, p38 and gr did not change significantly with any of the approaches used.

Virtual screening with actives from WOMBAT (combined with decoys for the same targets derived from DUD), provided an average ROC AUC value of 63.1 (median= 64.8) with 12.8% (median= 12.5%) true positives identified at a 2% false positive rate (Table 3A and Fig. 4A). Pocket refinement followed by out-of-pocket rescoring led to an average $AUC_{refined/rescored}$= 72.1 (median=69.8) and $ROC_{(2\%)refined/rescored}$= 28.7% (median= 22.8% Table 3B). Despite more conservative than DUD, these results confirm the good overall virtual ligand screening performance of ICM. Peroxisome proliferator activated receptor γ (ppar, $AUC_{refined/rescored}$= 92.7, $ROC_{(2\%)refined/rescored}$= 69.8%) and estrogen receptor antagonists (er_antag, $AUC_{refined/rescored}$= 89.0, $ROC_{(2\%)refined/rescored}$= 53.0%) were among the best performing targets (Fig. 4B).

Virtual ligand screening benchmarks using the WOMBAT test sets of lead-like, chemically diverse small molecules is inherently more challenging than DUD. This is particularly critical when using a single ligand binding pocket conformation because of potential induced fit mechanisms upon binding of different chemotypes. On the other hand, chemical bias was introduced combining WOMBAT actives with DUD decoys. Compared to WOMBAT actives, DUD decoys have higher molecular weights and more functional groups than WOMBAT, which in turn leads to additional favorable contacts and better scores.

Conclusions

In this study, docking and scoring with ICM was benchmarked for ligand binding mode accuracy and virtual screening against publicly available test sets. ICM was highly successful in generating multiple poses that include experimentally solved near native conformations. Known co-crystal binding modes were reproduced as the top 1 scoring pose in most sites. Visual inspection of a few cases with larger RMSD predictions revealed that the ligands were correctly positioned within the protein pocket and that the most critical contacts were captured. Difficult cases included PDB entries with protein-ligand water-mediated hydrogen bonds, metal coordination, highly symmetrical and flexible molecules. Our results demonstrate that ICM performs remarkably well in self-docking experiments; however, cross-docking tests where the extracted ligands are docked into protein structures from different complexes, would provide a more challenging and realistic assessment.

Cross-docking pose prediction test sets using experimentally solved conformational ensembles of druggable binding pockets will be important for future software comparison benchmarks [52].

Virtual ligand screening using the default ICM scoring method, single pocket conformations per target and no additional modeling steps to account for protein plasticity, showed good overall performance, yet highly dependent on the target of interest. This finding highlights the importance of target and software validation before undergoing time- and cost-consuming screening of large compound databases followed by experimental evaluation of new molecules. Rescoring based on out-of-pocket contacts was an efficient and CPU-time inexpensive method to improve recognition in virtual ligand screening. On the other hand, an intrinsic complexity in the structure-based identification of new active compounds is related to the flexibility of the protein binding pockets. Co-crystal structures of identical ligand-binding domains bound to different chemotypes reveal more or less pronounced conformational changes to accommodate binding of ligands. Our results using energy-based pocket refinement, accounting for induced fit and protein plasticity, dramatically improved the overall discrimination and early enrichments using a single binding site model per target. However, because different chemotypes can induce alternative conformational changes to the ligand binding pocket, such as rotation of side chains and small loop rearrangements, each of them represents only a fraction of the total molecular chemical recognition properties and has somewhat limited potential for virtual screening of novel chemotypes. Increasing evidence, including our own results, suggests that pocket ensembles and 4D (multi-conformational) docking are more effective to recognize ligands in virtual screening and predict their binding geometries [53–55].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

M.A.C. Neves thanks Fundação para a Ciência e a Tecnologia (FCT), Portugal, for a Post Doctoral grant (SFRH/BPD/64216/2009). The authors thank Irina Kufareva, Manuel Rueda, Winston Chen, Chayan Acharya and Chris Edwards for useful discussions and comments. This work was supported by National Institutes of Health [grant numbers R01 GM071872, U01 GM094612 and U54 GM094618].

References

1. Andricopulo AD, Salum LB, Abraham DJ. Structure-based drug design strategies in medicinal chemistry. Curr Top Med Chem. 2009; 9:771–790. [PubMed: 19754394]
2. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol. 2008; 153:S7–S26. [PubMed: 18037925]
3. Kroemer RT. Structure-based drug design: Docking and scoring. Curr Protein Peptide Sci. 2007; 8:312–328. [PubMed: 17696866]
4. Morra G, Genoni A, Neves MAC, Merz KM, Colombo G. Molecular recognition and drug-lead identification: What can molecular simulations tell us? Curr Med Chem. 2010; 17:25–41. [PubMed: 19941480]
5. Zou XQ, Sun YX, Kuntz ID. Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. J Am Chem Soc. 1999; 121:8033–8043.
6. Ruvinsky AM. Role of binding entropy in the refinement of protein-ligand docking predictions: Analysis based on the use of 11 scoring functions. J Comput Chem. 2007; 28:1364–1372. [PubMed: 17342720]

7. Scharer K, Morgenthaler M, Paulini R, Obst-Sander U, Banner DW, Schlatter D, Benz J, Stihle M, Diederich F. Quantification of cation-π interactions in protein-ligand complexes: Crystal-structure analysis of factor Xa bound to a quaternary ammonium ion ligand. Angew Chem-Int Edit. 2005; 44:4400–4404.
8. Bartlett GJ, Choudhary A, Raines RT, Woolfson DN. n->π* interactions in proteins. Nat Chem Biol. 2010; 6:615–620. [PubMed: 20622857]
9. Takahashi O, Kohno Y, Nishio M. Relevance of weak hydrogen bonds in the conformation of organic compounds and bioconjugates: Evidence from recent experimental data and high-level ab initio MO calculations. Chem Rev. 2010; 110:6049–6076. [PubMed: 20550180]
10. Milletti F, Vulpetti A. Tautomer preference in PDB complexes and its impact on structure-based drug discovery. J Chem Inf Model. 2010; 50:1062–1074. [PubMed: 20515065]
11. Robeits BC, Mancera RL. Ligand-protein docking with water molecules. J Chem Inf Model. 2008; 48:397–408. [PubMed: 18211049]
12. Kirton SB, Murray CW, Verdonk ML, Taylor RD. Prediction of binding modes for ligands in the cytochromes p450 and other heme-containing proteins. Proteins. 2005; 58:836–844. [PubMed: 15651036]
13. Irwin JJ, Raushel FM, Shoichet BK. Virtual screening against metalloenzymes for inhibitors and substrates. Biochemistry. 2005; 44:12316–12328. [PubMed: 16156645]
14. ten Brink T, Exner TE. pK_a based protonation states and microspecies for protein-ligand docking. J Comput Aided Mol Des. 2010; 24:935–942. [PubMed: 20882397]
15. Rose PW, Beran B, Bi CX, Bluhm WF, Dimitropoulos D, Goodsell DS, Prlic A, Quesada M, Quinn GB, Westbrook JD, Young J, Yukich B, Zardecki C, Berman HM, Bourne PE. The RCSB Protein Data Bank: redesigned web site and web services. Nucleic Acids Res. 2011; 39:D392–D401. [PubMed: 21036868]
16. Truchon JF, Bayly CI. Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. J Chem Inf Model. 2007; 47:488–508. [PubMed: 17288412]
17. Kufareva I, Abagyan R. Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. J Med Chem. 2008; 51:7921–7932. [PubMed: 19053777]
18. Monceaux CJ, Hirata-Fukae C, Lam PCH, Totrov MM, Matsuoka Y, Carlier PR. Triazole-linked reduced amide isosteres: An approach for the fragment-based drug discovery of anti-Alzheimer's BACE1 inhibitors. Bioorg Med Chem Lett. 2011; 21:3992–3996. [PubMed: 21621412]
19. Bowers EM, Yan G, Mukherjee C, Orry A, Wang L, Holbert MA, Crump NT, Hazzalin CA, Liszczak G, Yuan H, Larocca C, Saldanha SA, Abagyan R, Sun Y, Meyers DJ, Marmorstein R, Mahadevan LC, Alani RM, Cole PA. Virtual ligand screening of the p300/CBP histone acetyltransferase: Identification of a selective small molecule inhibitor. Chem Biol. 2010; 17:471–482. [PubMed: 20534345]
20. Endo S, Matsunaga T, Kuwata K, Zhao HT, El-Kabbani O, Kitade Y, Hara A. Chromene-3-carboxamide derivatives discovered from virtual screening as potent inhibitors of the tumour maker, AKR1B10. Bioorg Med Chem. 2010; 18:2485–2490. [PubMed: 20304656]
21. Odell LR, Howan D, Gordon CP, Robertson MJ, Chau N, Mariana A, Whiting AE, Abagyan R, Daniel JA, Gorgani NN, Robinson PJ, McCluskey A. The pthaladyns: GTP competitive inhibitors of dynamin I and II GTPase derived from virtual screening. J Med Chem. 2010; 53:5267–5280. [PubMed: 20575553]
22. Khan MTH, Fuskevag OM, Sylte I. Discovery of potent thermolysin inhibitors using structure based virtual screening and binding assays. J Med Chem. 2009; 52:48–61. [PubMed: 19072688]
23. Wu SD, Bottini M, Rickert RC, Mustelin T, Tautz L. In silico screening for PTPN22 inhibitors: Active hits from an inactive phosphatase conformation. Chemmedchem. 2009; 4:440–444. [PubMed: 19177473]
24. An JH, Lee DCW, Law AHY, Yang CLH, Poon LLM, Lau ASY, Jones SJM. A novel small-molecule inhibitor of the avian influenza H5N1 virus determined through computational screening against the neuraminidase. J Med Chem. 2009; 52:2667–2672. [PubMed: 19419201]
25. Bisson WH, Cheltsov AV, Bruey-Sedano N, Lin B, Chen J, Goldberger N, May LT, Christopoulos A, Dalton JT, Sexton PM, Zhang XK, Abagyan R. Discovery of antiandrogen activity of

nonsteroidal scaffolds of marketed drugs. Proc Natl Acad Sci U S A. 2007; 104:11927–11932. [PubMed: 17606915]

26. Cavasotto CN, Orry AJW, Murgolo NJ, Czarniecki MF, Kocsi SA, Hawes BE, O'Neill KA, Hine H, Burton MS, Voigt JH, Abagyan RA, Bayne ML, Monsma FJ. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. J Med Chem. 2008; 51:581–588. [PubMed: 18198821]
27. Katritch V, Jaakola VP, Lane JR, Lin J, IJzerman AP, Yeager M, Kufareva I, Stevens RC, Abagyan R. Structure-based discovery of novel chemotypes for adenosine A_{2A} receptor antagonists. J Med Chem. 2010; 53:1799–1809. [PubMed: 20095623]
28. Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. J Med Chem. 2003; 46:3045–3059. [PubMed: 12825943]
29. Schapira M, Raaka BM, Das S, Fan L, Totrov M, Zhou ZG, Wilson S, Abagyan R, Samuels HH. Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. Proc Natl Acad Sci U S A. 2003; 100:7354–7359. [PubMed: 12777627]
30. Schapira M, Raaka BM, Samuels HH, Abagyan R. In silico discovery of novel Retinoic Acid Receptor agonist structures. BMC Struct Biol. 2001; 1:1–7. [PubMed: 11405897]
31. Dey R, Chen L. In search of allosteric modulators of alpha 7-nAChR by solvent density guided virtual screening. J Biomol Struct Dyn. 2011; 28:695–715. [PubMed: 21294583]
32. Schapira M, Abagyan R, Totrov M. Structural model of nicotinic acetylcholine receptor isotypes bound to acetylcholine and nicotine. BMC Struct Biol. 2002; 2:1–8. [PubMed: 11860617]
33. Ravna AW, Sylte I, Sager G. Binding site of ABC transporter homology models confirmed by ABCB1 crystal structure. Theor Biol Med Model. 2009:6. [PubMed: 19416527]
34. Ravna AW, Sylte I, Dahl SG. Molecular mechanism of citalopram and cocaine interactions with neurotransmitter transporters. J Pharmacol Exp Ther. 2003; 307:34–41. [PubMed: 12944499]
35. Ravna AW, Sylte I, Dahl SG. Molecular model of the neural dopamine transporter. J Comput Aided Mol Des. 2003; 17:367–382. [PubMed: 14635728]
36. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WTM, Mortenson PN, Murray CW. Diverse, high-quality test set for the validation of protein-ligand docking performance. J Med Chem. 2007; 50:726–741. [PubMed: 17300160]
37. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. J Med Chem. 2006; 49:6789–6801. [PubMed: 17154509]
38. Good AC, Oprea TI. Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection? J Comput Aided Mol Des. 2008; 22:169–178. [PubMed: 18188508]
39. Abagyan R, Totrov M, Kuznetsov D. ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. J Comput Chem. 1994; 15:488–506.
40. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys. 1953; 21:1087–1092.
41. Schapira M, Totrov M, Abagyan R. Prediction of the binding energy for small molecules, peptides and proteins. J Mol Recognit. 1999; 12:177–190. [PubMed: 10398408]
42. Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins. 1997:215–220. [PubMed: 9485515]
43. Halgren TA. Merck molecular force field.1. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem. 1996; 17:490–519.
44. Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R. Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment. Structure. 2011; 19:1108–1126. [PubMed: 21827947]
45. Rueda M, Katritch V, Raush E, Abagyan R. SimiCon: a web tool for protein-ligand model comparison through calculation of equivalent atomic contacts. Bioinformatics. 2010; 26:2784–2785. [PubMed: 20871105]
46. McGann M. FRED pose prediction and virtual screening accuracy. J Chem Inf Model. 2011; 51:578–596. [PubMed: 21323318]

47. Davis IW, Baker D. ROSETTALIGAND docking with full ligand and receptor flexibility. J Mol Biol. 2009; 385:381–392. [PubMed: 19041878]

48. Olsen L, Pettersson I, Hemmingsen L, Adolph HW, Jorgensen FS. Docking and scoring of metallo-β-lactamases inhibitors. J Comput Aided Mol Des. 2004; 18:287–302. [PubMed: 15562992]

49. Korb O, Stutzle T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with PLANTS. J Chem Inf Model. 2009; 49:84–96. [PubMed: 19125657]

50. Donnecke D, Schweinitz A, Sturzebecher A, Steinmetzer P, Schuster M, Sturzebecher U, Nicklisch S, Sturzebecher J, Steinmetzer T. From selective substrate analogue factor Xa inhibitors to dual inhibitors of thrombin and factor Xa. Part 3. Bioorganic & Medicinal Chemistry Letters. 2007; 17:3322–3329. [PubMed: 17462889]

51. Nar H, Bauer M, Schmid A, Stassen JM, Wienen W, Priepke HWM, Kauffmann IK, Ries UJ, Hauel NH. Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. Structure. 2001; 9:29–37. [PubMed: 11342132]

52. Kufareva I, Ilatovskiy AV, Abagyan R. Pocketome: an encyclopedia of small-molecule binding sites in 4D. Nucleic Acids Res. 2011

53. Bottegoni G, Kufareva I, Totrov M, Abagyan R. Four-dimensional docking: A fast and accurate account of discrete receptor flexibility in ligand docking. J Med Chem. 2009; 52:397–406. [PubMed: 19090659]

54. Neves MAC, Simoes S, Melo MLSE. Ligand-guided optimization of CXCR4 homology models for virtual screening using a multiple chemotype approach. J Comput Aided Mol Des. 2010; 24:1023–1033. [PubMed: 20960031]

55. Park SJ, Kufareva I, Abagyan R. Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles. J Comput Aided Mol Des. 2010; 24:459–471. [PubMed: 20455005]

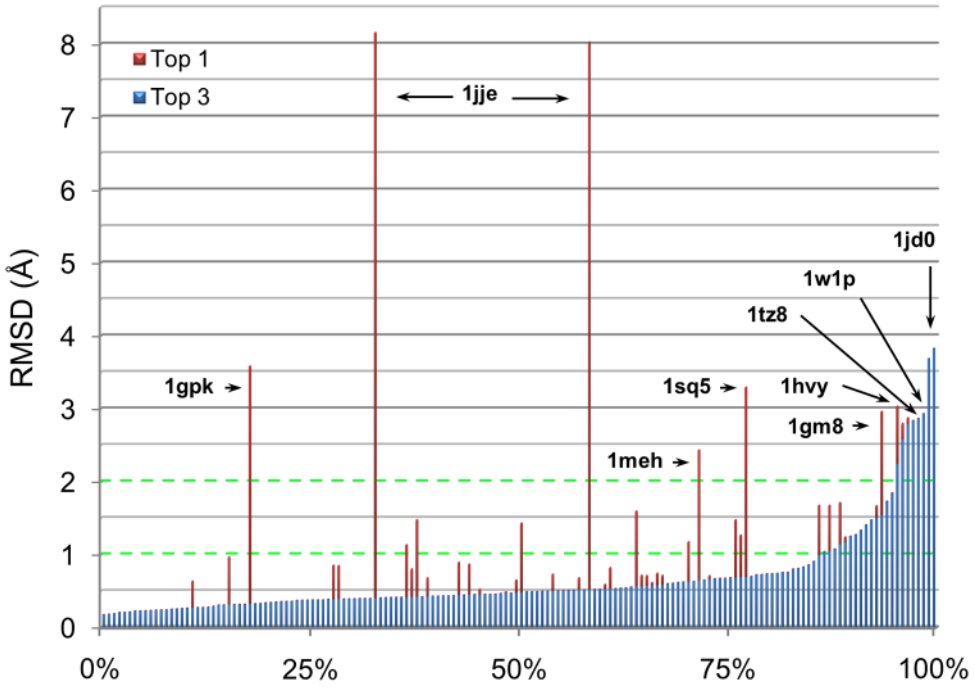

Fig. 1.

RMSD values for ICM predictions on all ligand binding sites. Red bars indicate RMSD values for the top 1 scoring poses, whereas blue bars indicate the lowest RMSD among the top 3 scoring poses. PBD entries with binding mode predictions above 2 Å RMSD are labeled.

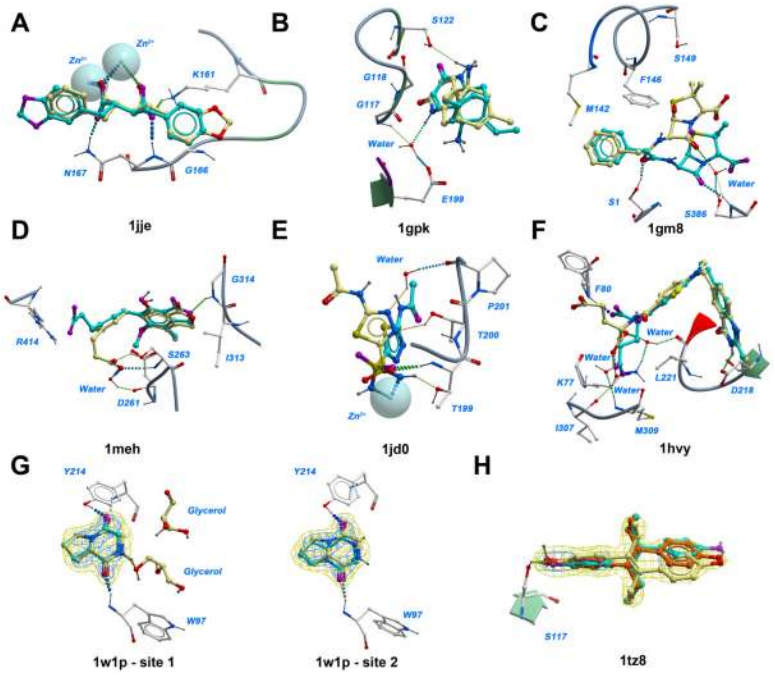

Fig. 2.

Ligand binding mode predictions above 2 Å RMSD. Top 1 scoring poses are represented with cyan carbons and magenta oxygens whereas co-crystal ligands are represented with yellow carbons and red oxygens. Crystallographic residues, water molecules, metal ions and other small molecules found in the binding site are labeled and numbered as they appear in the PDB files. Water molecules were excluded for docking purposes. Hydrogen bonds are represented with spheres and colored according to the estimated energy (blue – strong interaction, red – weak interaction). Fig. G shows the predicted binding modes for two sites in PDB entry 1w1p. Fig. H displays an alternative co-crystal binding mode represented with orange carbons. Electron density maps at 1.0 and 2.5 sigma levels are represented with yellow and blue meshes, respectively, around the co-crystal ligands of PDB entries 1w1p and 1tz8.

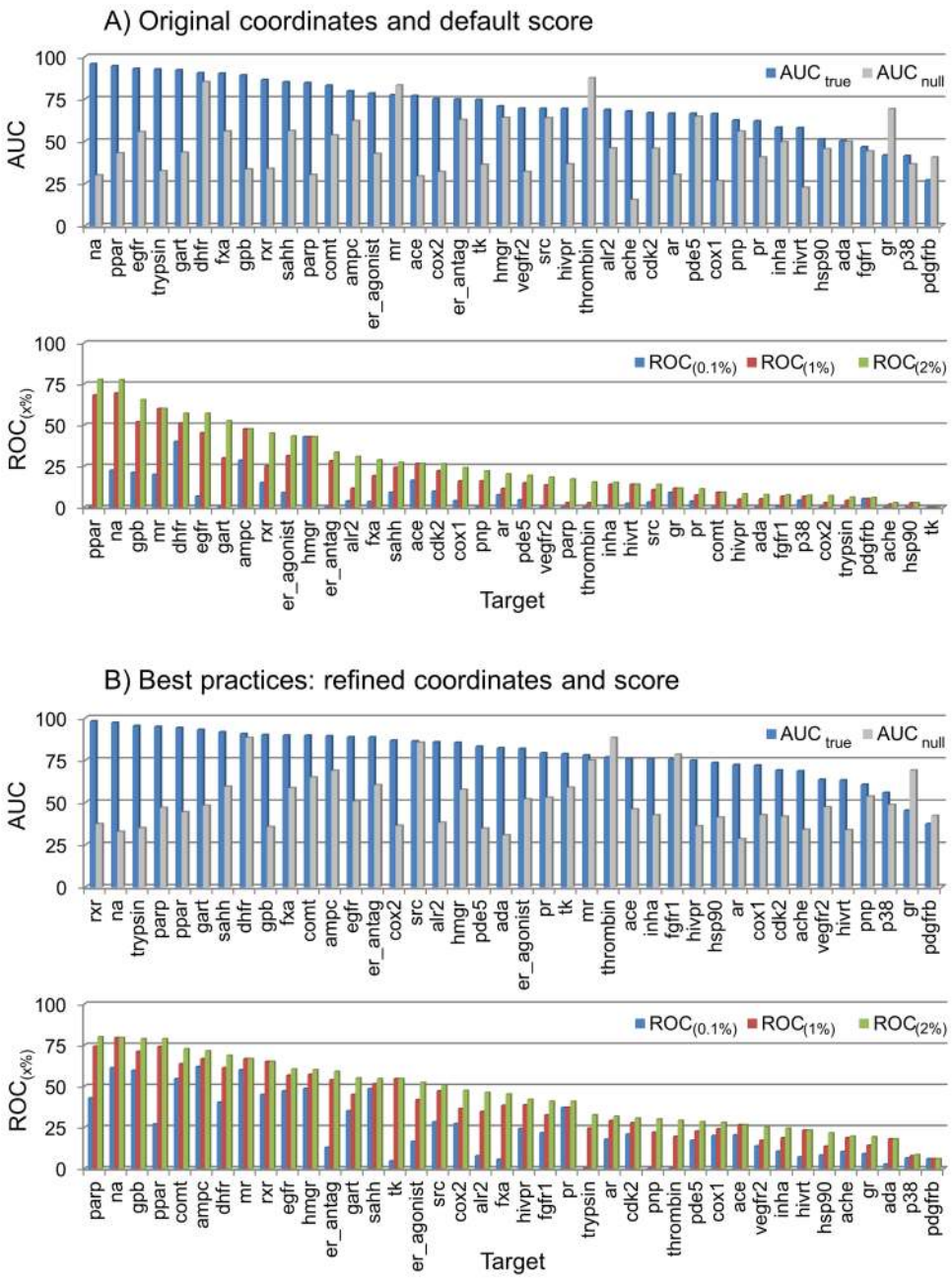

Fig. 3.

Virtual ligand screening benchmark of 40 DUD test sets docked against the original target coordinates using the default ICM scoring method (A), or docked against refined induced fit models followed by out-of-pocket rescoring (B). ROC AUC values were calculated for the true ligands (blue) and null hypothesis (gray). Early enrichments are reported as the fraction of true positives recovered at 0.1% (blue), 1% (red) and 2% (green) false positive rates. A list of abbreviations is provided as Supporting Information.

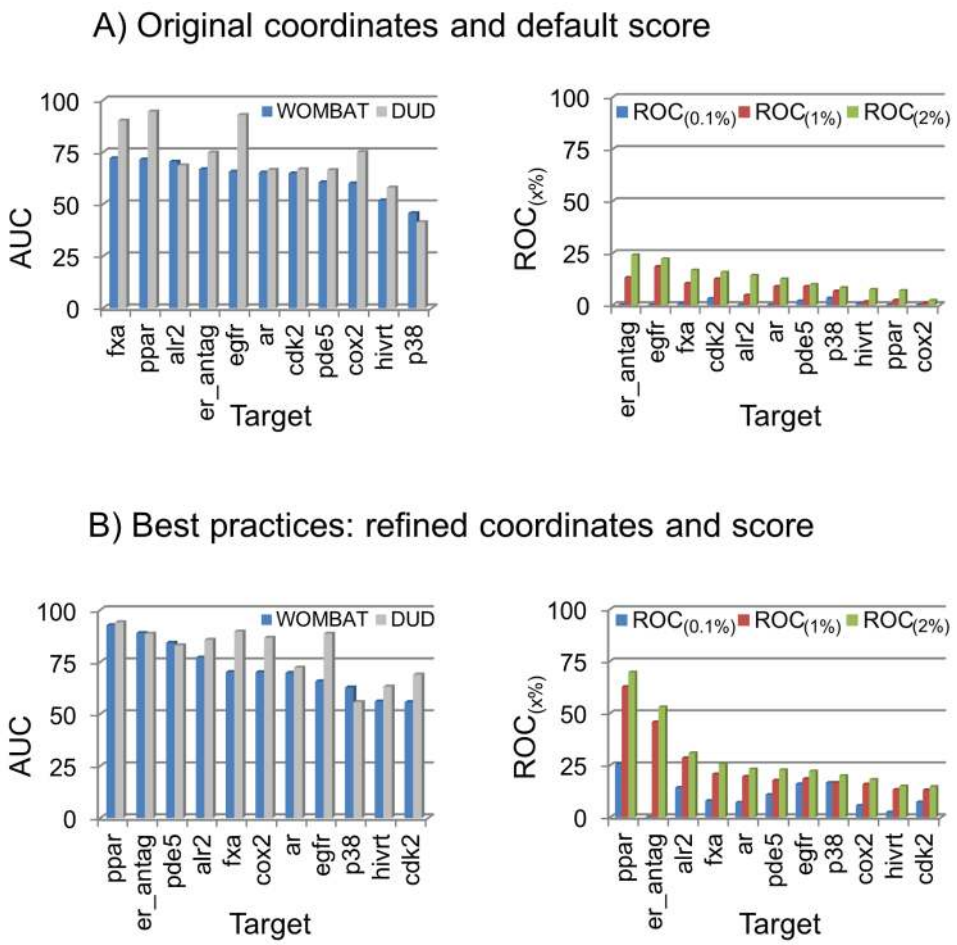

Fig. 4.

Virtual ligand screening benchmark of 11 WOMBAT test sets docked against the original target coordinates using the default ICM scoring method (A), or docked against refined induced fit models followed by out-of-pocket rescoring (B). ROC AUC values obtained with the WOMBAT test sets (blue) are compared with the corresponding DUD results (gray). Early enrichments for the WOMBAT test sets are reported as the fraction of true positives recovered at 0.1% (blue), 1% (red) and 2% (green) false positive rates. A list of abbreviations is provided as Supporting Information.

Table 1

Mean, standard deviation, median, range and percentage of RMSD values within 0.5 Å, 1 Å and 2 Å, calculated between the top 1 and top 3 ICM predictions and the co-crystal ligands.

Rank	Mean	SD	Median	Min	Max	≤0.5 Å	≤1 Å	≤2 Å
Top 1	0.91 Å	1.10 Å	0.54 Å	0.18 Å	8.2 Å	43 %	78 %	91 %
Top 3	0.67 Å	0.62 Å	0.48 Å	0.18 Å	3.8 Å	53 %	86 %	95 %

Table 2

Statistics for the virtual ligand screening benchmark of 40 DUD test sets docked against the original target coordinates using the default ICM scoring method (A), or against refined induced fit models followed by out-of-pocket rescoring (B). ROC AUC values were calculated using the true test sets (AUC_{true}) and the null hypothesis (AUC_{null}). The rate of true positives at x=0.1%, x=1% and x=2% are provided.

	A) Original coordinates and default score					**B) Best practices: refined coordinates and score**				
	Mean	**SD**	**Median**	**Min**	**Max**	**Mean**	**SD**	**Median**	**Min**	**Max**
AUC_{true}	71.6	16.3	69.4	27.1	95.8	79.4	13.8	82.2	37.2	98.1
AUC_{null}	46.7	17.0	43.4	15.4	87.4	50.6	16.3	46.9	28.4	88.3
$ROC_{(0.1\%)}$	7.3	10.8	3.8	0	42.9	24.7	19.5	20.4	0	61.9
$ROC_{(1\%)}$	21.0	19.3	14.8	0	69.4	39.5	20.9	37.0	5.9	79.6
$ROC_{(2\%)}$	26.6	21.4	22.0	0	77.8	44.3	21.1	45.2	5.9	80.0

J Comput Aided Mol Des. Author manuscript; available in PMC 2012 July 17.

J Comput Aided Mol Des. Author manuscript; available in PMC 2012 July 17.

Table 3

Statistics for the virtual ligand screening benchmark of 11 WOMBAT test sets docked against the original target coordinates using the default ICM scoring method (A), or docked against refined induced fit models followed by out-of-pocket rescoring (B). ROC AUC values were calculated using WOMBAT actives and DUD decoys. The rate of true positives at x=0.1%, x=1% and x=2% are provided.

	A) Original coordinates and default score					**B) Best practices: refined coordinates and score**				
	Mean	**SD**	**Median**	**Min**	**Max**	**Mean**	**SD**	**Median**	**Min**	**Max**
AUC_{true}	63.1	8.2	64.8	45.7	72.1	72.1	12.4	69.8	55.7	92.7
$ROC_{(0.1\%)}$	0.9	1.3	0	0	3.3	10.4	7.3	8.0	0	25.6
$ROC_{(1\%)}$	8.1	5.5	8.9	1.1	18.5	24.8	15.6	18.5	13.2	62.8
$ROC_{(2\%)}$	12.8	6.7	12.5	2.3	24.1	28.7	17.2	22.8	14.7	69.8