Evaluation of the Wilma-SIE virtual screening method in community structure–activity resource 2013 and 2014 blind challenges

Hogues, Hervé; Sulea, Traian; Purisima, Enrico O.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1021/acs.jcim.5b00278
Journal of Chemical Information and Modeling, 2015-08-17

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

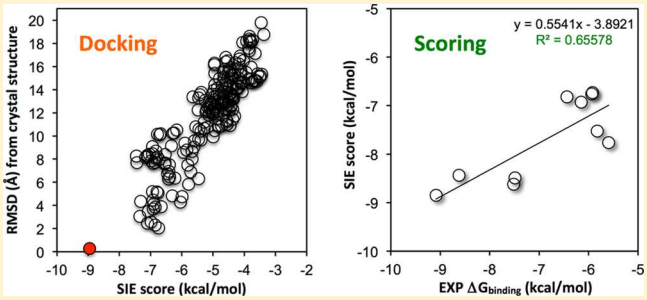
Evaluation of the Wilma-SIE Virtual Screening Method in Community Structure−Activity Resource 2013 and 2014 Blind Challenges

Hervé Hogues, Traian Sulea, and Enrico O. Purisima*

Human Health Therapeutics, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada

Supporting Information

ABSTRACT: Prospective assessments of the Wilma-SIE (solvated interaction energy) platform for ligand docking and ranking were performed during the 2013 and 2014 editions of the Community Structure−Activity Resource (CSAR) blind challenge. Diverse targets like a steroid-binding protein, a serine protease (factor Xa), a tyrosine kinase (Syk), and a nucleotide methyltransferase (TrmD) were included. Pose selection was achieved with high precision; in all 24 tests Wilma-SIE top-ranked the native pose among carefully generated sets of decoy conformations. Good separation for the native pose was also observed indicating robustness in pose scoring. Cross-docking was also accomplished with high accuracy for the various systems, with ligand median-RMSD values around 1 Å from the crystal structures. Larger deviations were occasionally obtained due to the rigid-target approach even if multiple target structures were used. Affinity ranking of congeneric ligands after cross-docking was reasonable for three of the four systems, with Spearman ranking coefficients around 0.6. Poor affinity ranking for FXa is possibly due to missing structural domains, which are present during measurements. Assignment of protonation states is critical for affinity scoring with the SIE function, as shown here for the Syk system. Including the FiSH model improved cross-docking but worsened affinity predictions, pointing to the need for further fine-tuning of this newer solvation model. The consistently strong performance of the Wilma-SIE platform in recent CSAR and SAMPL blind challenges validates its applicability for virtual screening on a broad range of molecular targets.



INTRODUCTION

Modern drug discovery, design, and optimization is an application-driven field of active research that has matured tremendously during the past decade due to the increasing health and social demands of a growing and aging demographics. This has fuelled the diversification of a plethora of experimental techniques complemented by data manipulation algorithms, from cell-based screening to medicinal and combinatorial chemistry, from in vitro assay development to crystallography and molecular modeling. Within this scope, the screening approach has proven to be one of the most productive methods in identifying early hit molecules to be considered for progression toward next phases of drug development. High-throughput assay development and large physical collections of compounds are very expensive; hence a computational alternative called virtual screening has emerged as a promising alternative for hit discovery.[1−3] However, virtual screening requires computational speed without compromising on prediction accuracy, the optimal balance of which is difficult to achieve but critical to success.

The first ingredient required in order to carry out virtual screening is an effective docking procedure. To this end we developed Wilma, an exhaustive docking program that has the required efficiency for large-scale virtual screening of drug-like compound libraries.[4,5] Owing to its exhaustive nature as well as to its fast empirical pose-ranking function calibrated on crystal structures of protein−ligand complexes, the top-ranked pose produced by Wilma has been shown to be consistently close to the experimental pose for drug-like ligands.

The second ingredient required in a virtual screening method is accurate scoring, so Wilma was coupled with the solvated interaction energy (SIE) scoring function.[6−8] SIE is an end-point force-field-based scoring function similar in spirit with another popular method, MM-PB(GB)/SA,[9−11] which combines molecular mechanics and continuum solvation terms. Calibrated on a diverse data set of 99 protein−ligand complexes,[6] SIE achieves a reasonable transferability across a wide variety of protein−ligand systems for which it predicts absolute binding affinities within the experimental range as shown by various test cases reported in the literature.[8,12] External testing of the standard SIE parametrization on the CSAR-2010 data set of 343 protein−ligand complexes diverse with respect to ligands and targets predicted absolute binding affinities with a mean-unsigned-error of about 2 kcal/mol, a Pearson squared correlation coefficient (R^2) of 0.38, and a Spearman rank-order correlation coefficient (ρ) of 0.62.[12]

Special Issue: Community Structure Activity Resource (CSAR)

Received: May 13, 2015

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

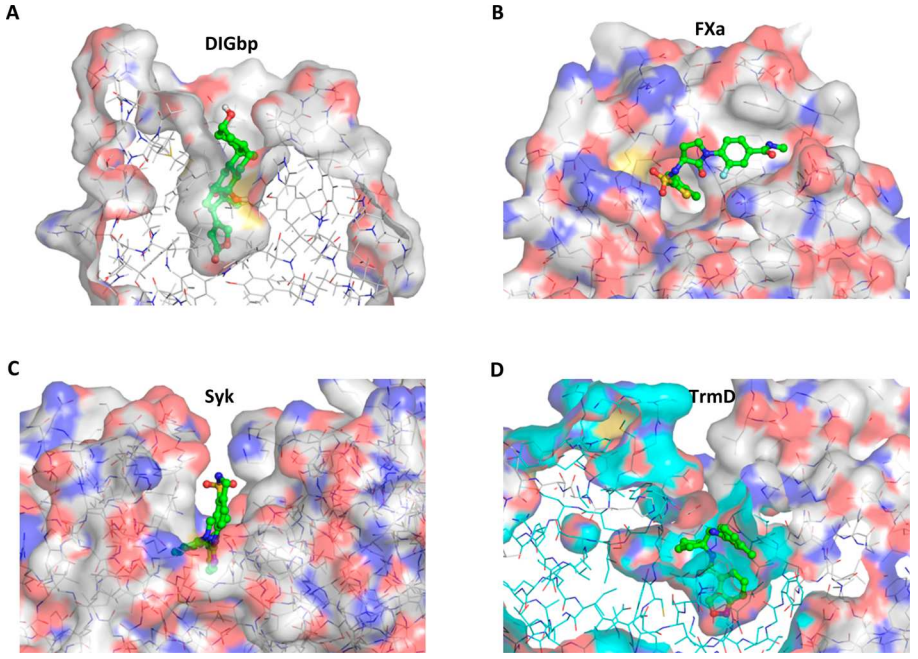


Figure 1. Diversity of protein−ligand systems included in the 2013 and 2014 editions of the CSAR challenge. Proteins are represented by molecular surfaces and line bond-models with standard coloring of atom types. If required, Z-slicing is applied in order to reveal deeply buried binding pockets (panels A and D). Ligands are shown with ball-and-stick models and standard coloring of atoms except for carbon atoms shown in green. (A) DIGbp binds hydrophobic steroids at a deeply buried location. (B) FXa binds polar analogs that remain largely solvent-exposed. (C) Syk binds ATP-competitive inhibitors between the N- and C-terminal domains of the kinase fold. (D) TrmD binds drug-like inhibitors in a cleft formed between the subunits of the homodimeric enzyme (cyan versus white carbon-atom-surface patches for the two subunits).

Both SIE and Wilma have been employed for blind testing in SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) challenges (www.eyesopen.com/sampl). A reasonable performance of SIE in binding affinity prediction for the SAMPL1 set of kinase inhibitors with available cognate crystal structures had been noted.[13] In SAMPL3, the Wilma-SIE virtual screening platform achieved good enrichment of true positives from a data set of fragment-size ligands against trypsin, with an AUC of about 0.7 for a ROC curve characterized by an excellent early enrichment performance.[4] Binding affinity predictions for trypsin-ligand and host−guest complexes in SAMPL3 were generally within 2 kcal/mol of the experimental values but rank ordering of affinities within 2 kcal/mol was not well predicted. This was reiterated in SAMPL4 for a data set of weak inhibitors of HIV-integrase spanning only 1.2 kcal/mol of binding affinity, whereas Wilma-SIE achieved good performance on host−guest systems having wider affinity ranges.[5] Interestingly, despite the narrow affinity range that precluded accurate ranking within the HIV-integrase ligands in SAMPL4, the absolute affinity predictions of Wilma-SIE were improved by employing our newer solvation model, FiSH.[14]

In order to continue prospective assessment of the Wilma-SIE virtual screening platform in terms of ligand docking and scoring predictions, we tested it on all four molecular systems proposed in the 2013 and 2014 editions of the CSAR (Community Structure−Activity Resource) challenge (http://csardock.org). These systems cover a wide range in terms of binding site topology, ligand polarity, and overall protein−ligand interactions. As exemplified in Figure 1, the engineered digoxigenin steroid-binding protein (DIGbp) has a deep hydrophobic pocket for binding relatively hydrophobic steroid analogs, factor Xa (FXa), as a serine protease, binds polar ligands that remain largely solvent-exposed, the spleen tyrosine kinase Syk binds hydrogen-bond capable heterocyclic analogs competitively at the ATP site, whereas the tRNA guanine-methyltransferase (TrmD) has a deeply buried but polar active site nested between the two subunits of this homodimeric enzyme.

In addition to providing molecular diversity of the systems used in method testing, these CSAR challenges were also devised to test the method performance at various levels: (1) starting with decoupling pose scoring from pose sampling in pose-selection tests, (2) testing cross-docking accuracy, and (3) culminating with assessing ranking performance against measured protein−ligand binding data that span several orders of magnitude and reflect the real-life practice. Hence, the composition and the format of these latest CSAR blind tests are well suited to help assess the strengths and limitations of the Wilma-SIE method, complementing previous prospective tests.

METHODS

Wilma-SIE Docking Pipeline. The docking software Wilma uses a brute-force searching approach where the interaction with the rigid protein of all the discrete rotational and translational states of every ligand conformation generated by Omega (OpenEye Scientific Software, Inc., Santa Fe, NM) is examined.[4,5] Using an efficient filtering method, the program exhaustively enumerates, scores and ranks all the ligand poses that do not overlap with the protein. Docking is done within one or several predefined rectangular volumes with a translation step size of 0.5 Å. The discrete rotation of the ligand is adjusted to ensure that the maximum movement of any atom between adjacent orientations is less than 1 Å. The ligand conformations generated by Omega are controlled by setting the internal energy cutoff to 20 kcal/mol and adjusting the root-mean-

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

square deviation (RMSD) clustering parameter to produce at most 5000 conformations.

The newer 4-term scoring function,[5] which was trained to recover the most native states using 320 protein−ligand complexes from the curated CSAR-2010 data set,[15] was used

$$\text{WilmaScore2} = w_1E_{\text{coul}} + w_2E_{\text{vdw}} + w_3E_{\text{HB}} + w_4E_{\text{flaws}} \quad (1)$$

This scoring function includes a Coulombic interaction term, E_{coul}, a van der Waals 6−12 Lennard-Jones potential, E_{vdw}, an explicit H-bond term, E_{HB}, which considers donor and acceptor orientations, and a term, E_{flaws}, which introduces an energetic penalty for flaws present in the docked pose in terms of protein−ligand polar-atom complementarity.[5] These flaws account for the obstruction of polar groups by nonpolar or like-charged polar groups. Introduction of the E_{flaws} model is an attempt to reduce occasional top-ranked poses and false-positive ligands that are "flawed" due to the presence of buried partially charged atoms without formation of electrostatically complementary interactions in the bound state, which were still observed when using surface complementarity terms. This empirical geometrical model poses a more stringent electrostatic desolvation penalty on such unfavorable interactions (flaws), in addition to addressing the charge polarity of interactions better than the surface-based model (which does not distinguish opposite-charge from like-charge interactions). The Wilma scoring function was used exclusively for structure prediction, i.e., to select a reduced set of realistic docked poses (e.g., top-200 ranked poses) that will be further rescored by the higher-accuracy scoring functions described below (SIE and SIE + FiSH, vide infra).

Scoring of binding affinities was carried out using the solvated interaction energy (SIE) end-point force-field based method,[6−8,12] which approximates the binding free energy from the electrostatic and nonpolar components of the interaction energy and the desolvation free energy

$$\text{SIE} = \alpha(E_{\text{coul}} + E_{\text{vdw}} + E_{\text{RF(BEM)}} + E_{\text{npsolv}}) + C \quad (2)$$

where E_{coul} and E_{vdw} describe solute−solute interactions by intermolecular Coulombic and van der Waals interaction energies in the bound state calculated with AMBER and GAFF molecular mechanics force fields.[16−18] Desolvation effects are described by $E_{\text{RF(BEM)}}$, the change in the reaction field energy between the bound and free states calculated with a continuum model based on a boundary element solution to the Poisson equation using the BRI BEM program,[19,20] and a solute dielectric constant D_{in} = 2.25, and E_{npsolv}, the nonpolar desolvation approximated from a linear proportionality with the change in solute molecular surface area.[21−23] The free state of the system is obtained by rigid separation of the interacting molecules from the bound state. Partial atomic charges for protein atoms are taken from the AMBER force field, whereas organic solutes are assigned AM1-BCC partial charges.[24,25] α is a global scaling factor of the total raw solvated interaction energy relating to the scaling of the binding free energy due to configurational entropy effects.[26,27] The standard parameters of the SIE function in eq 2 are α = 0.1048 and C = −2.89 kcal/mol calibrated against a protein−ligand training data set of 99 complexes refined by restrained energy minimization.[6]

We also tested prospectively a different SIE function in which the solvation terms are replaced by our latest continuum solvation model FiSH that captures some of the properties of the first shell of hydration.[14,28] For example, the electrostatic desolvation in the FiSH model, $E_{\text{RF(FISH)}}$, can account for charge asymmetry effects. Also, instead of a single surface-area-based term for all nonelectrostatic components of solvation, FiSH includes an additional continuum van der Waals term, E_{cvdw}, to more accurately describe the solute−solvent nonpolar interactions, and a separate surface-area based cavity term, E_{cav}. Unlike the default solvation model within SIE, which uses a solute dielectric of 2.25, the FiSH model uses a solute dielectric of 1.0. The modified SIE + FiSH scoring function then has the form

$$\text{SIE + FiSH} = \alpha(E_{\text{coul}} + E_{\text{vdw}} + E_{\text{RF(FiSH)}} + E_{\text{cvdw}} + E_{\text{cav}}) + C \quad (3)$$

where the parameters α = 0.1232 and C = 1.46 were obtained by training against the same 99 protein−ligand data set used for the original SIE function.[6]

Operationally, only a subset of representative structures taken from the top-200 poses ranked by WilmaScore2, are rescored using SIE or SIE + FiSH. The main reason is to reduce the number of redundant SIE calculations since Wilma's very fine spatial sampling typically returns groups of adjacent poses that, after minimization, fall into the same minimum state. Starting from the best WilmaScore2 pose, poses are retained if their heavy-atom RMSD is greater than 2.0 Å to every selected pose so far. Otherwise the poses are dropped. Thus, the top WilmaScore2 pose is always selected, alongside typically less than 10 alternate pose representatives.

In order to partially mitigate the rigid-target docking approach adopted in the Wilma-SIE platform, docking was carried out on multiple conformations of the target protein (described in the next subsection). Hence, the top SIE-scored pose over all these target conformations was taken as the docked pose, and also for prediction of binding affinity. The same approach was used for the Wilma-SIE + FiSH method. Prior to the SIE and SIE + FiSH calculations, all complexes were refined by constrained energy minimization as described previously.[4,12] Briefly, the ligand, and the protein residues within 4 Å from the ligand, were energy-minimized using the AMBER and GAFF force-field parameters together with a distance-dependent dielectric constant ($4r$) to crudely mimic solvent screening, down to a gradient of 0.01 kcal/(mol Å). Harmonic restraints with force constants of 3 and 20 kcal/(mol Å^2) were applied to the heavy atoms of the ligand and protein, respectively, in this region.

Structural Preparation. *Protein Targets.* CSAR 2013 organizers provided crystal structures for three proteins designed to bind the steroid digoxigenin (DIG), which we call DIG binding proteins (DIGbp). These included two proteins, DIG18 and DIG20 for pose-selection testing, and a third structure, DIG10.2, for cross-docking and subsequent affinity ranking of analog steroids. Several protein conformations of FXa, Syk, and TrmD proteins were provided as part of the CSAR 2014 edition for pose-selection (phase 1) and cross-docking and affinity ranking (phase 2). These protein conformations were determined as complexes with various ligands by X-ray crystallography, from which the ligands' coordinates were removed by the CSAR organizers. These included 3 conformations for FXa, 5 conformations for Syk, and 14 conformations for TrmD. We supplemented the FXa set with an additional conformation corresponding to the crystal structure with PDB code 1F0R.

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Structural preparation of these protein targets was largely done by the CSAR organizers. Their preparation included adjustment of all His residues within active site for the most appropriate tautomeric state, flipping of Asn and Gln side chains if required in order to improve the H-bonding network, removal of all water molecules and crystallographic additives, capping of N- and C-termini with acetyl and methylamine groups, reconstruction of residues with missing side-chain atoms using MOE 2011.10 (Chemical Computing Group, Inc., Montreal, QC), manual examination of any residue within the active site with alternate conformations and retaining of the most appropriate conformation (for any residue outside of the active site, conformation A was retained), and addition of hydrogen atoms followed by minimization with the MMFF94x force field,[29,30] in MOE 2011.10.

After verification of these structures, we performed the following structural changes. For the DIG20 structure, the originally missing hydrogen atoms of four side-chain hydroxyl groups of reconstructed residues were added. For the FXa structures, the protonation of the active-site His57 residue from the catalytic triad was changed to neutral and the orientation of the hydroxyl H atom of the active site Ser195 was directed for H-bonding to the NE atom of His57 according to the catalytic mechanism of this serine proteinase; the N-terminus of the α-chain was charged (first residue, also engaged in a salt-bridge with Asp194). For the Syk structures, two phosphotyrosine (Ptr) residues were dephosphorylated to Tyr residues; the solvent-exposed His506 was deprotonated. For the TrmD structures, the N-terminus was adjusted according to the protein sequence by removal of the upstream RGSH sequence (cloning artifact) and protonation of the chain terminus at Met1 (according to the UniProtKB database entry P43912); the last residue of the protein, Ser246, was built if missing in some of the structures using as template a TrmD structure containing this residue, and then negatively charged (C-terminus) in all structures. After these modifications, for all target structures the positions of polar hydrogen atoms were refined to improve the H-bonding network. This optimization procedure systematically explores alternate discrete orientations of all polar hydrogen atoms (e.g., in hydroxyl groups), scoring the H-bond network configuration with the WilmaScore2 E_{HB} term. Atomic partial charges were assigned according to the AMBER all-atom force field.[16,17]

In the case of FXa, we used four conformations of the target; however, the side-chain of Gln192 residue near the active site can adopt several conformation as seen in these structures, and it can obviously impact ligand docking and scoring. Hence, we constructed and then used for cross-docking and affinity ranking 16 structures of FXa by importing these four Gln192 conformations into the four FXa structures. We also noted that there are missing loops in some of the Syk, TrmD, and FXa structures—we did not attempt to reconstruct these segments; hence, the number of atoms varies between the structures used for the same target.

Ligands. Salt counterions were filtered from the original SMILES-formatted files supplied for the ligands. Protonation states of the ligands were generated with Epik 2.3 (Schrödinger, Inc., New York, NY).[31,32] We used the *scan* option for sequential pK_a estimation of microscopic pK_a values at pH 7.3 and then retrieved the adjusted protonation for the most populated state at this pH. Partial charges were then calculated for this state using the AM1-BCC method,[24,25] as implemented in Molcharge (OpenEye Scientific Software, Inc., Santa Fe, NM), using the lowest-energy conformation generated by Omega (OpenEye Scientific Software, Inc.) as the input conformation. For simplicity, we ignored the dependence of partial charges on ligand conformations.

Data Analysis. Analysis of ligand cross-docking accuracy was carried out by the CSAR organizers (see CSAR overview papers in this special issue). Briefly, a Gaussian-weighted RMSD tool was used for aligning the protein of the submitted complex to the experimental X-ray structure in order to ensure the structures are in the correct frame of ref 33. This tool only uses the Cα atoms for alignment, no fitting is done on the ligand. A symmetry-corrected RMSD algorithm (Chemical Computing Group, Inc., Montreal, QC), which utilizes the Morgan algorithm to determine the appropriate ligand atom correspondences, was then used for calculating the RMSD of the ligand.

Outlier analysis was informed by chemical clustering of the analogs in each congeneric set of ligands provided for cross-docking and binding affinity prediction. This was done by complete hierarchical clustering in Sybyl X.2.0 (Tripos, Inc., St-Louis, MO), using 2D-FINGERPRINTS and ATOM_-PAIR_FP, as 2D-descriptors of molecular structure.

For crossreference with the overview CSAR articles appearing in this special issue, the group identifiers of our prospective submissions in the 2013 and 2014 editions are O and G, respectively.

RESULTS AND DISCUSSION

Pose-Selection Accuracy. The simplest test for evaluating the pose-scoring performance of a docking method is pose selection. In this case, the conformation of the target protein corresponds to its bound state for the given ligand as determined experimentally (e.g., cocrystal structure) and is termed the cognate target conformation. In CSAR 2013 and 2014 editions, prospective pose selection was tested by the ability to identify a single near-native pose of the ligand (<1 Å RMSD from the crystallographic pose) within a set of 200 ligand poses. Hence, 199 challenging decoy poses were carefully pregenerated by the CSAR organizers for each test system (see accompanying paper in this special issue). In this way, ligand pose sampling and pose scoring are decoupled, hence various methods for scoring ligand poses can be evaluated comparatively without being affected by variations of different ligand pose-sampling algorithms. To locate the near-native binding pose means that this pose has to be scored best (i.e., top-1-ranked). As can be seen in Table 1, the SIE scoring function achieved 100% success across 24 complexes spanning all 4 targets: DIGbp (2 complexes), FXa (3 complexes), Syk (5 complexes), and TrmD (14 complexes). The Wilma-SIE method was among the 7 out of 52 submissions (13%) that correctly best-scored the native pose

Table 1. Performance of Pose Selection Based on Top-SIE-Ranked Pose

CSAR edition/phase	target protein	success[a]
2013/2	DIGbp	2 out of 2
2014/1	FXa	3 out of 3
	Syk	5 out of 5
	TrmD	14 out of 14

[a]Number of complexes with the near-native pose top-1-scored by SIE among a set of 200 pregenerated decoy poses.

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

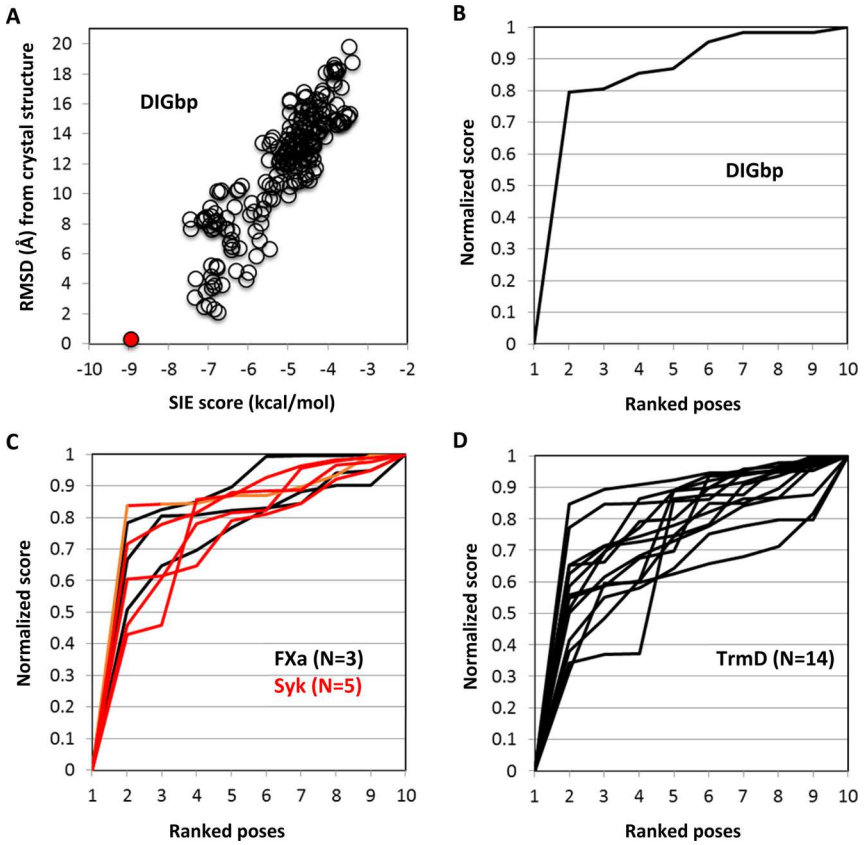


Figure 2. Pose-selection performance. (A) Illustrative example showing that the top-scored pose with the SIE function has the lowest RMSD to native structure (red-filled circle). (B) Pose separation by SIE scores for the top-10 ranked poses. Same system as in panel A shown. Top-10 SIE scores are normalized, with the lowest score assigned a value of 0, and the highest a value of 1. (C) Score separation of the top-1-ranked pose for the FXa (3 complexes) and Syk (5 complexes) systems. (D) Score separation of the top-1-ranked pose for the TrmD system (14 complexes).

Table 2. Performance of Cross-Docking Based on Top-Ranked Pose (CSAR 2014, Phase 2)

		RMSD (Å)									
		median		average		SD		min		max	
target	*n*	SIE + FiSH	SIE	SIE + FiSH	SIE	SIE + FiSH	SIE	SIE + FiSH	SIE	SIE + FiSH	SIE
FXa	3	1.310	1.320	1.307	3.224	0.013	2.699	1.290	1.310	1.322	7.041
Syk	8	0.613	1.248	0.733	1.376	0.430	0.820	0.358	0.358	1.793	2.785
TrmD	31	1.207	1.207	1.965	1.986	1.954	2.064	0.217	0.224	7.143	7.580
all	42	1.081	1.269	1.683	1.958	1.759	1.993	0.217	0.224	7.143	7.580
cognate in[a]	20	0.680	0.875	0.973	1.222	0.634	0.968	0.217	0.224	2.370	4.352
no cognate[b]	22	1.451	1.471	2.329	2.628	2.160	2.406	0.562	0.470	7.143	7.580

[a]Cognate protein conformation present in the set of target structures used. [b]Cognate protein conformation not present within the set of target structures used.

for all 22 complexes over the three systems from CSAR 2014 (Syk, TrmD, and FXa).

The near-native pose, having the lowest RMSD from the actual one, was in all cases scored with the best SIE score, hence at the top of the list. A plot of the SIE scores versus RMSD from the native pose is shown in Figure 2A for the DIGbp target, a behavior that is representative for all other systems tested here. Importantly, the native poses were not only scored best, but they are also separated well from the next, incorrectly placed, poses in terms of scores. This is exemplified in Figure 2B, where we plot the normalized scores for the top 10 ranked poses of the same system analyzed in Figure 2A (the top-1 pose is given a normalized score of 0, and the top-10 pose a normalized score of 1). We see that, on this normalized scale, the score of the second-ranked pose is 80% weaker than that of the first-ranked pose. A large score gap between the first and second pose is a desirable characteristic of a robust scoring function for docking. It was observed that the SIE function consistently displays this feature across the other systems tested prospectively here (Figure 2C and D). Moreover, we noted that the WilmaScore2, which is the scoring function used for providing reasonable poses for SIE rescoring in typical applications of the Wilma-SIE pipeline, was also capable of correctly top-1 ranking the native pose for all 24 systems, and generally did so with a significant gap to the second-ranked pose.

Cross-docking Accuracy. The next test in the CSAR 2014 challenge was cross-docking. This is different from pose selection because the participants had to use their pose-sampling algorithms to generate ligand poses during docking

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

prior to scoring them, and then locate the best docking solution. A total of 42 ligands were included in this docking assessment exercise (3 for FXa, 8 for Syk, and 31 for TrmD), for which cocrystal structures were solved but withheld by the CSAR organizers. Importantly, for more than half of these ligands (22), the cognate conformation of the target had not been previously disclosed, which represents a true cross-docking testing. For the rest of the set (20 ligands), cognate conformations of the target could be found in the set of structures provided for the pose-selection challenge (described in the previous subsection). However, we did not limit docking of those ligands only to their cognate target conformations. Instead, our general approach was to dock all ligands to the same set of multiple conformations available for each target. We used 4 conformations for FXa (one conformation was added to the 3 provided by the organizers), the 5 conformations provided for Syk, and all 14 conformations provided for TrmD. In this way, we extended the cross-docking treatment to the entire set of 42 ligands. There are substantial conformational variations within each set of employed target structures (Figure S1).

In Table 2 we report statistics of the RMSD values between the Wilma-docked poses and the crystallographic ones. These values are based on the top-1-ranked poses by SIE and SIE + FiSH scoring functions, over all multiple target conformations used. It can be seen that the median RMSD values are close to 1 Å across all systems. Average RMSD values are slightly larger because they are more sensitive to outliers. Wilma docking can reach very high accuracy, with RMSD values as low as 0.2 Å. Only for a few isolated cases in the TrmD system did the top-1-scored poses have large deviations from the native poses of about 6−7 Å.

We also see that using SIE + FiSH for pose scoring leads to somewhat better docking accuracy overall as well as for individual targets. For example, the median RMSD in the case of docking to Syk with Wilma-SIE is about 1.2 Å, but about 0.6 Å when SIE + FiSH is employed for pose scoring. This is seen more clearly in Figure 3A where the RMSD values between native poses and the poses top-ranked by the two scoring functions are compared. Notably, the FXa ligand ID 102, and the TrmD ligand ID 446, which are misdocked by Wilma-SIE (RMSD values of 7.0 and 4.4 Å, respectively) are well-docked by Wilma-SIE + FiSH (RMSD values of 1.3 and 0.5 Å, respectively). For the other ligands, the discrepancy between the results obtained with the two scoring functions is less, and in the majority of cases the two methods arrive at the same docking solution. These data may suggest that the improvements introduced by FiSH in modeling desolvation effects lead to better ranking among poses of a given ligand, although a wider study will be needed in order to confidently draw this conclusion.

Both Wilma-SIE and Wilma-SIE + FiSH generated docking predictions that were among the top-7% to top-28%-performing submissions depending on the particular system. Importantly, in terms of consistency and system transferability, they were the only submissions that achieved median RMSD values below 1.5 Å on every system (out of 28 submissions on all three systems).

Only four TrmD ligands (ID 455, 461, 470, and 471) were misdocked by Wilma-SIE. These cases could not be improved by SIE + FiSH pose scoring either. Chemical cluster analysis indicates that these ligands have as distinct feature a bulky substituent (Figure S2), so it is possible that the space required to accommodate this branched group was not sufficient in any of the TrmD target conformations used for docking. That is, even if multiple target conformations can be used when available, Wilma is essentially a rigid-target docking approach.

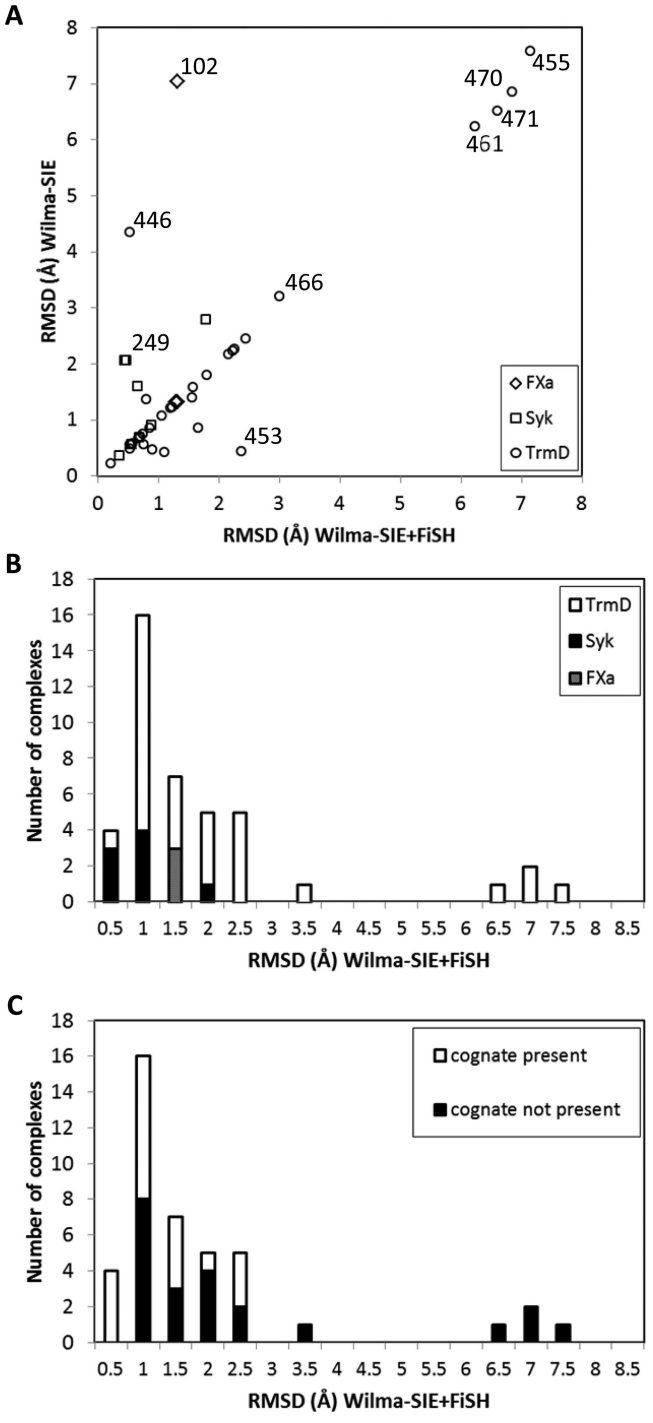


Figure 3. Cross-docking performance. (A) Correlation between the accuracies of top-1 pose predictions with the SIE and SIE + FiSH scoring functions for the three systems from CSAR 2014 edition, phase 2: FXa 3 complexes, Syk 8 complexes, and TrmD 31 complexes. Selected ligands discussed in the text are labeled by CSAR ID codes. (B) Distribution of top-1 pose prediction accuracy (SIE + FiSH scoring) over all systems shown in panel A and listed in Table 2. (C) Dependence of top-1 pose prediction accuracy on the presence of the cognate protein conformation within the set of multiple protein structures used for each system.

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Inevitably, there are cases when the method will fail due to large conformational changes that are induced upon ligand binding, which are not explored by the method and may not be present in the collection of available structures. Despite this limitation, this CSAR blind testing demonstrated that the Wilma docking method achieved good performance for each target as well as across multiple diverse targets. The docking accuracy can be better appreciated in Figure 3B showing the distribution of RMSD values in the case of the Wilma-SIE + FiSH approach, indicating that 48% of the ligands were docked with an RMSD below 1 Å, and 74% below 2 Å.

These results validate the general applicability of our docking procedure since they were obtained in the real-life cross-docking scenario. More than half of the ligands (22 out of 42) did not have the cognate target conformation available, whereas for the remaining 20 ligands the cognate target conformation was used alongside other available target conformations. In Figure 3C we further decompose the RMSD distribution of the 42 ligands into those two classes. It can be seen that docking was slightly more accurate when the cognate target conformation was present. However, the performance of true cross-docking for the 22 ligands that did not have the cognate target conformation available is still very good. Eight of these ligands (36%) were docked with an RMSD below 1 Å, and 15 (68%) below 2 Å.

Affinity Ranking after Cross-Docking. The last test proposed by CSAR 2013 and 2014 editions was ranking and scoring of binding affinities following cross-docking. This is relevant for many important drug discovery applications including virtual screening and lead optimization. Hence, protein−ligand binding or enzyme inhibition measurements were carried out for ligands of the four targets described earlier: DIGbp, Syk, TrmD, and FXa, and these data were again withheld by the CSAR organizers. The compounds tested included those from the cross-docking test, for which crystal structures were solved, but with the exceptions of the TrmD set (31 ligands), they also included many other analogs. Thus, the Syk set contains 276 ligands, FXa ligands were split into three sets of 45, 67, and 51 ligands (for 106 unique ligands after excluding common compounds among these sets), and 10 analogs were provided in the DIGbp set. The performance of SIE and SIE + FiSH scoring function after Wilma cross-docking of these protein−ligand sets is listed in Table 3 in terms of Spearman rank-order correlation coefficient, ρ, and squared Pearson linear regression correlation coefficient R^2. As can be seen, reasonable correlations could be obtained after cross-docking for two of the four systems (DIGbp and TrmD) with at least one scoring function. Although the initial prospective affinity predictions for the Syk set showed no correlation with experimental data, retrospective correction of errors in ligand protonation led to significantly improved results (discussed below). No acceptable correlation could be obtained for any of the three sets of ligands against FXa. In general, the SIE scoring function faired better than the SIE + FiSH function. Below we will discuss each system in some detail, with a focus on outlier analysis when applicable.

Table 3. Performance of Ligand Binding Affinity Ranking after Cross-Docking[a]

CSAR edition/ phase	target protein	n	SIE		SIE + FiSH	
			ρ	R^2	ρ	R^2
2013/3	DIGbp	10	0.649	0.656	0.576	0.610
2014/2	Syk	276	0.120	0.000	0.127	0.001
	Syk[b]	276	0.585	0.306	0.514	0.235
	TrmD	31	0.514	0.295	0.058	0.008
	FXa set1	45	0.263	0.022	0.139	0.000
	FXa set2	67	0.088	0.010	0.057	0.003
	FXa set3	51	0.019	0.003	0.091	0.001

[a] ρ is the Spearman-ρ rank-order correlation coefficient; R^2 is the squared Pearson linear-regression correlation coefficient. [b] Retrospective results after correction of ligand protonation.

DIGbp Binding Data Set. In the case of the 10 steroid analogs binding to the DIGbp, good ranking (ρ = 0.65) and correlation (R^2 = 0.66) were obtained with the SIE function, and only slightly poorer after the inclusion of the FiSH solvation model. In Figure 4A we also see that not only relative binding affinities but also the absolute ones are of the same magnitude with the SIE values (kcal/mol). This is in agreement with the calibration of the SIE function and the various tests that we have reported previously.[6,8,12]

TrmD Binding Data Set. The set of 31 ligands of TrmD is essentially the same set used in the cross-docking exercise, so both experimental structures and affinities were available in this case, but withheld. A reasonable ranking (ρ = 0.51) was obtained with the SIE function after Wilma-SIE cross-docking into multiple TrmD structures, and a linear correlation could be detected (R^2 = 0.30). Interestingly, this performance vanishes when the SIE + FiSH function was employed (see Table 3). This is in striking contrast with the results obtained in the cross-docking testing, especially in the case of TrmD where pose prediction was significantly improved by employing the more advanced FiSH solvation model (see Table 2). There are two lessons that can be learned from these results. First, a scoring function that properly selects poses of the same ligand does not necessarily perform in ranking top-1-poses of different ligands. Second, the FiSH model likely needs some technical refinements and additional parametrization. For example, there is a certain level of noise arising from the surface integral associated with the continuum van der Waals term, and the dielectric constant of 1 used in the calibration of the first-shell-hydration electrostatic term may not be entirely appropriate for organic molecules and the protein interior with expected dielectric values of 2−4.

Outlier analysis for this set highlighted the somewhat weaker predicted binding affinities for the two most potent inhibitors (ID 449 and 450), which have pIC50 values by about 1.5 log units lower than the next strongest inhibitor (ID 451). These three compounds are highlighted in red in the plot shown in Figure 4B. Chemical cluster analysis indicated that these three ligands form a tight cluster of analogs (Figure S3), with the stronger inhibitors (449 and 450) possessing a t-butyl substituent that is replaced by a methyl group in the weaker analog (451). We could trace back the weaker SIE predictions for the two stronger inhibitors to a weakening of electrostatic interactions, which more than offsets the stronger van der Waals interaction energies for these larger analogs (449 and 450) relative to the smaller one (451). In turn, this predicted weakening of electrostatic interactions was incurred by a flip in the amide carbonyl from the middle part of the chemical scaffold, a necessary distortion of the ligand conformation due to the limited space available for the bulkier t-butyl substituent during cross-docking (Figure S3). We could confirm that indeed, while cross-docking was fairly accurate, it was more

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

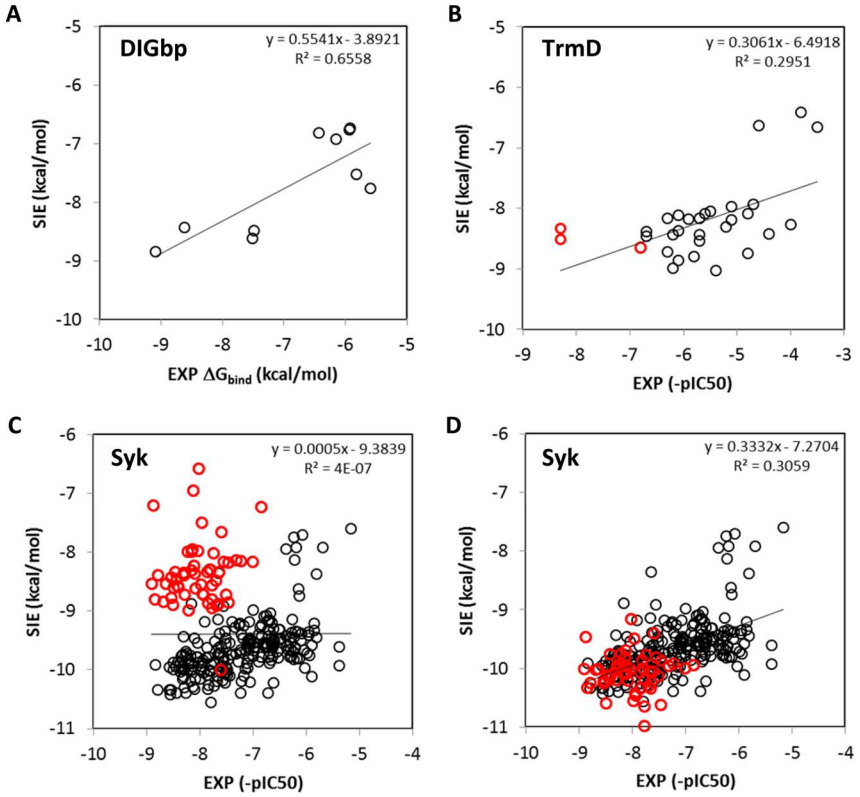


Figure 4. Ligand binding affinity predictions based on the SIE scoring function after cross docking (see Figure S5 for the corresponding plots based on the SIE + FiSH scoring function). (A) DIGbp-ligand affinities. (B) TrmD-ligand affinities. Red circles mark a chemical cluster of three analogs discussed in the text. (C) Syk-ligand affinities. Prospectively submitted data. Red circles are outliers discussed in the text. (D) Syk-ligand affinities. Corrected data. Red circles are outliers corrected as discussed in the text.

precise in the case of the smaller inhibitor (RMSD of 1.2 Å) than for the two larger analogs appearing as outliers (RMSDs of 2.3 and 2.4 Å). This result illustrates that some of the inaccuracies in binding affinity predictions are not necessarily related only to inaccuracies of the scoring functions employed but may also arise due to the rigid-target docking approximation adopted for cross-docking in Wilma and many other virtual screening docking algorithms, even if multiple structures of the target are used as input.

Syk Binding Data Set. This was the largest set for binding affinity prediction testing proposed in CSAR 2014 and it contains 276 analogs. Prospectively submitted Wilma-SIE and Wilma-SIE + FiSH predictions had poor rankings and showed no correlations (ρ = 0.12 and R^2 = 0). However, during a retrospective analysis of the data we noted that there are clearly two groups of ligands, a larger group showing some correlation trend, and a smaller group that can be defined as outliers relative to the first group (Figure 4C). Further structural inspection revealed that the differentiating feature between these two groups is the different protonation of the 2-amino-1,3-pyrimidine moiety that is common in this class of kinase inhibitors (Figure S4). The larger group of ligands that shows a correlation between predicted and actual binding data has a neutral 2-amino-1,3-pyrimidine, whereas the 55 ligands in the outlier group (red circles in Figure 4C) were assigned as positively charged 2-amino-1,3-pyrimidinium ions, hence a difference in the net charge between these groups. Based on Syk-bound structures known for several kinase inhibitors from this class, one can ascertain that the correct protonation of the 2-amino-1,3-pyrimidine moiety is the neutral state, since it is required to act as an H-bond acceptor from the backbone amide group of Syk residue Ala451, whereas the protonated state would create an "electrostatic clash" (Figure S4). After deprotonating the 2-amino-1,3-pyrimidine moiety for the 55 ligands in the outlier subset, the ranking and correlations between predictions and actual values were much improved (see Table 3), for both SIE (Figure 4D) and SIE + FiSH (Figure S5) functions. With this retrospective correction, the Wilma-SIE predictions have Spearman-rho rank order correlation coefficients to experiment above 0.5 for both the TrmD and Syk sets, which is worth nothing since only one other method achieved this performance. A few other methods did better on one set, but poorer on the other one.

SIE and SIE + FiSH scoring functions contain electrostatic contributions to binding describing direct Coulomb interaction and electrostatic desolvation, and are sensitive to changes in electrostatics, especially to the net charge of the ligand. Since the ligands from the outlier group bury a charged moiety upon binding, the desolvation cost is high without being compensated by newly established favorable electrostatic interactions (in fact, unfavorable electrostatic interactions are established by the protonated group). Hence, their predicted affinities are weaker than for the ligands in the correlating set (Figure 4C). Charge removal at the 2-amino-1,3-pyrimidine group reduces the desolvation penalty leading to stronger predicted binding affinities and improved agreement with experimental data. This observation highlights the fact that correct assignment of protonation states is of paramount importance to the success of scoring protein−ligand interactions with physics-based scoring functions like the SIE.

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

This reiterates our previous findings on the CSAR-2010 data set.[12] Here, we relied on an automated assignment of ligand protonation in the free state with the Epik software, in order to predict the most abundant protonation state at the physiological pH. However, care should be exercised for cases where there are groups with pK_a values similar to the targeted pH value (as it is the case of 2-amino-1,3-pyrimidine group), since the standard errors for pK_a prediction with empirical methods like Epik are relatively large. Prediction of pK_a values and protonation states with several alternate methods, as well as visual inspection, are strongly advisible in such cases.

FXa Binding Data Set. The only system where we obtained a poor performance of affinity prediction with both SIE and SIE + FiSH scoring functions after Wilma cross-docking is FXa (Table 3, Figure S6). Only in set 1 do we obtain some detectable ranking predictability with Wilma-SIE (ρ = 0.26). It should be noted that we do not report the combined set due to redundancies between sets for a large number of ligands and complications arising from having different measured affinities for the same ligand in many cases.

The problem with poor prediction of FXa binding affinities has been highlighted and discussed before as part of the CSAR 2010 edition.[34] It may relate to the fact that the form of FXa used in the crystal structures is not the form used in the assays. The crystal structure is a truncated form that is missing several regulatory domains (e.g., the N-terminal domain which is required for calcium activation of FXa in vivo), and their effect on ligand binding is unclear. A large number of FXa ligands exhibit high levels of binding affinity, but the pockets are largely solvent-exposed and the complementarity appears rather poor. It is possible that some of these domains provide parts of the pocket for the inhibitors or affect the electrostatic or conformational behavior of the catalytic domain. This is a reasonable possibility since docking to the truncated form can be done quite accurately (Table 2), but the docked poses do not describe the affinities measured using the full protein. All of the domains of FXa are present in the assays for the inhibitors, but missing in the crystal structures, so the data could be mismatched and misleading. Whether retaining some water molecules in the binding site can lead to improvements in affinity prediction in this system is an idea worthwhile testing, but this has not been done here as it is not the standard operating mode of the Wilma-SIE platform.

CONCLUSIONS

The CSAR 2013 and 2014 blind challenges stringently tested the Wilma-SIE docking-scoring platform with congeneric series of ligands against diverse targets. The intention was to test the performance at various levels, from decoupling pose scoring from pose sampling in pose-selection tests, then testing cross-docking accuracy, based on which affinity ranking tests were finally carried out in order to approach a real-life practice. Wilma-SIE passed the majority of these tests with excellent (pose selection) to good (cross-docking) to reasonable (affinity ranking) prediction accuracies.

Clearly, ligand docking was more tractable than affinity scoring, which is not new to this active field of research. In terms of pose selection, it was interesting to note that good separation of the native pose at the top of the ranked list was obtained, which indicates robust pose scoring.

The message reinforced by the cross-docking tests was that the rigid-target approach underlying Wilma docking will inevitably sometimes fail to provide the correct ligand pose, which also has consequences to affinity ranking. Fortunately for large-scale virtual screening applications, these events are seldom and the screened compound libraries are typically large enough that one can afford the ensuing false negatives.

Another lesson that reiterates a finding from our previous study that followed the CSAR-2010 edition is that care should be exercised during assignment of protonation states when employing a physics-based scoring function like SIE. Multiple pK_a/protonation predictors as well as human-based visual inspection of the structure are highly advisible. Finally, the improved cross-docking results obtained with our most advanced solvation model FiSH prompt toward its further refinements in order to improve affinity scoring too.

In conclusion, the present performance of Wilma-SIE remains consistent with our own experience with this method on various systems, as well as with previous CSAR and SAMPL blind challenges.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00278.

Figures showing the conformational flexibility of the targets (S1), outlier analysis (S2−S4), and binding affinity prediction scatter plots (S5 and S6) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: Enrico.Purisima@nrc-cnrc.gc.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This is NRC Canada publication number 53301.

REFERENCES

(1) McInnes, C. Virtual Screening Strategies in Drug Discovery. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494−502.

(2) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.

(3) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Jr. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2013**, *66*, 334−395.

(4) Sulea, T.; Hogues, H.; Purisima, E. O. Exhaustive Search and Solvated Interaction Energy (SIE) for Virtual Screening and Affinity Prediction. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 617−633.

(5) Hogues, H.; Sulea, T.; Purisima, E. O. Exhaustive Docking and Solvated Interaction Energy Scoring: Lessons Learned from the SAMPL4 Challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 417−427.

(6) Naim, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; Purisima, E. O. Solvated Interaction Energy (SIE) for Scoring Protein-Ligand Binding Affinities. 1. Exploring the Parameter Space. *J. Chem. Inf. Model.* **2007**, *47*, 122−133.

(7) Cui, Q.; Sulea, T.; Schrag, J. D.; Munger, C.; Hung, M. N.; Naim, M.; Cygler, M.; Purisima, E. O. Molecular Dynamics–Solvated Interaction Energy Studies of Protein-Protein Interactions: The Mp1-P14 Scaffolding Complex. *J. Mol. Biol.* **2008**, *379*, 787−802.

(8) Sulea, T.; Purisima, E. O. The Solvated Interaction Energy Method for Scoring Binding Affinities. *Methods Mol. Biol.* **2012**, *819*, 295−303.

(9) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.;

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(10) Brown, S. P.; Muchmore, S. W. Large-Scale Application of High-Throughput Molecular Mechanics with Poisson-Boltzmann Surface Area for Routine Physics-Based Scoring of Protein-Ligand Complexes. *J. Med. Chem.* **2009**, *52*, 3159−3165.
(11) Gohlke, H.; Kiel, C.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*, 238−250.
(12) Sulea, T.; Cui, Q.; Purisima, E. O. Solvated Interaction Energy (SIE) for Scoring Protein-Ligand Binding Affinities. 2. Benchmark in the CSAR-2010 Scoring Exercise. *J. Chem. Inf. Model.* **2011**, *51*, 2066−2081.
(13) Skillman, A. SAMPL1 at First Glance. *CUP IX meeting*, Santa Fe, NM, March 19, 2008.
(14) Corbeil, C. R.; Sulea, T.; Purisima, E. O. Rapid Prediction of Solvation Free Energy. 2. The First-Shell Hydration (FiSH) Continuum Model. *J. Chem. Theory Comput.* **2010**, *6*, 1622−1637.
(15) Dunbar, J. B., Jr.; Smith, R. D.; Yang, C. Y.; Ung, P. M.; Lexa, K. W.; Khazanov, N. A.; Stuckey, J. A.; Wang, S.; Carlson, H. A. CSAR Benchmark Exercise of 2010: Selection of the Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2011**, *51*, 2036−2046.
(16) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(17) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(18) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(19) Purisima, E. O. Fast Summation Boundary Element Method for Calculating Solvation Free Energies of Macromolecules. *J. Comput. Chem.* **1998**, *19*, 1494−1504.
(20) Purisima, E. O.; Nilar, S. H. A Simple yet Accurate Boundary Element Method for Continuum Dielectric Calculations. *J. Comput. Chem.* **1995**, *16*, 681−689.
(21) Chan, S. L.; Purisima, E. O. Molecular Surface Generation Using Marching Tetrahedra. *J. Comput. Chem.* **1998**, *19*, 1268−1277.
(22) Chan, S. L.; Purisima, E. O. A New Tetrahedral Tesselation Scheme for Isosurface Generation. *Comput. Graph.* **1998**, *22*, 83−90.
(23) Bhat, S.; Purisima, E. O. Molecular Surface Generation Using a Variable-Radius Solvent Probe. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 244−261.
(24) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(25) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(26) Chang, C. E.; Gilson, M. K. Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.
(27) Chen, W.; Chang, C. E.; Gilson, M. K. Calculation of Cyclodextrin Binding Affinities: Energy, Entropy, and Implications for Drug Design. *Biophys. J.* **2004**, *87*, 3035−3049.
(28) Purisima, E.; Corbeil, C.; Sulea, T. Rapid Prediction of Solvation Free Energy. 3. Application to the SAMPL2 Challenge. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 373−383.
(29) Halgren, T. A. Merk Molecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(30) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interaction Energies and Geometries. *J. Comput. Chem.* **1999**, *20*, 730−748.
(31) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-Like Molecules in Aqueous Solution. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 591−604.
(32) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for pK(a) Prediction and Protonation State Generation for Drug-Like Molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681−691.
(33) Damm, K. L.; Carlson, H. A. Gaussian-Weighted RMSD Superposition of Proteins: A Structural Comparison for Flexible Proteins and Predicted Protein Structures. *Biophys. J.* **2006**, *90*, 4558−4573.
(34) Smith, R. D.; Dunbar, J. B., Jr.; Ung, P. M.; Esposito, E. X.; Yang, C. Y.; Wang, S.; Carlson, H. A. CSAR Benchmark Exercise of 2010: Combined Evaluation across All Submitted Scoring Functions. *J. Chem. Inf. Model.* **2011**, *51*, 2115−2131.

DOI: 10.1021/acs.jcim.5b00278
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX