Docking Ligands into Flexible and Solvated Macromolecules. 8. Forming New Bonds – Challenges and Opportunities.

Anne Labarre,[1,#] Julia K. Stille,[1,#] Mihai Burai Patrascu [1, #] Andrew Martins,[1,] Joshua Pottel[2,] and Nicolas Moitessier[1,2]*

[1] Department of Chemistry, McGill University, 801 Sherbrooke St W, Montreal, QC, Canada H3A 0B8. [2] Molecular Forecaster Inc. 7171, rue Frederick-Banting, Montreal, QC, Canada H4S 1Z9

nicolas.moitessier@mcgill.ca

ORCIDs:
AL: 0000-0003-4939-3928
JS: 0000-0003-4815-1732
MBP: 0000-0001-9289-7887
AM: 0000-0002-7078-7625
JP: 0000-0002-4455-4713
NM: 0000-0001-6933-2079

Abstract: Over the years, structure-based design programs and specifically docking small molecules to proteins have become prominent in drug discovery. However, many of these computational tools have been developed to primarily dock enzyme inhibitors (and ligand to other protein classes) relying heavily on hydrogen bonds, electrostatic and hydrophobic interactions. In reality, many drug targets either feature metal ions, can be targeted covalently, or are simply not even proteins (e.g., nucleic acids). Herein, we describe several new features that we have implemented into FITTED to broaden its applicability to a wide range of covalent enzyme inhibitors, and to metalloenzymes, where metal coordination is essential for drug binding. We also report new datasets that were essential demonstrate areas of success and those where additional efforts are required. This resource could be used by other program developers to assess their own.

INTRODUCTION

Drug discovery is an expensive endeavor with a high failure rate. It takes on average over 10 years to bring a drug to the market with a total development cost estimated from $300M to nearly $3B,[1,2] with a large fraction of the costs being attributed to failures.[3,4] It is now fairly common to find computer-aided drug design (CADD) techniques applied at various stages of the preclinical drug discovery pipeline in an attempt to reduce costs and accelerate the time required to reach clinical trials. Indeed, it has been proposed that using CADD tools in the drug discovery process could lead to an overall cost reduction of 50%.[5] One such technique is structure-based drug design (SBDD), more specifically docking methods, with its primary application in the hit identification stage. Over the past three decades, several research and development teams in both academia and industry have developed and improved docking software, leading to programs like AutoDock,[6,7] FITTED,[8,9] FlexX,[10,11] Glide,[12,13] GOLD,[14,15] ICM,[16-18] and Surflex.[19,20]

Docking methods aim to predict the binding mode of potential ligands and to rank their binding affinity for a target—ideally applicable to vast types of ligands and targets. Docking programs generally perform well in predicting traditional non-bonding ligand-protein interactions (e.g., using previously trained electrostatic, van der Waals, and hydrogen bonding terms from molecular mechanics). However, cross-docking comparative studies, more indicative of a real prospective study than self-docking, are much less common and less accurate than self-docking. In addition, retrospective virtual screening studies rarely demonstrate area under the receiver operating curve (AUROC) values exceeding 0.8, although machine learning techniques have been proposed to improve this accuracy.[21,22] Among the challenges noted are the large chemical diversity in ligands, numerous classes of proteins with varying structural features, as well as biases in available testing sets.[23,24]

When looking at the applicability domain of existing docking programs, some limitations are apparent. For example, covalent drugs have featured in recent reports, demonstrating their potential.[25,26] Unfortunately, many computational platforms overlook automated recognition of covalent warheads, approximation of covalent group reactivity, and simultaneous consideration of covalent and non-covalent binding modes, making docking difficult to apply. Similarly, drug discovery campaigns are frequently targeting metalloenzymes, like matrix metalloproteinase[27] and histone deacetylases.[28] In most docking environments, coordination to a metal is often modeled using non-bonding interactions omits any proton transfers to a neighboring basic residue observed in several metalloprotein-inhibitor complexes. For example, a hydroxamic acid becomes highly acidic upon zinc coordination and will transfer a hydrogen to a neighboring glutamate residue.[29] Several metals play key structural or functional roles in enzymes; iron is found in nearly 100 enzymes,[30,31] while magnesium and zinc are found in over 300 enzymes each.[32] Other important metals include copper (present in over 25 enzymes),[33,34] nickel (in 9 enzymes identified to date),[35] and manganese (present in over 100 enzymes).[36] Each of these metals present specific coordination geometries and binding energies, and therefore challenges for docking methods.

Since the initial report of our docking program FITTED in 2007,[8] we have been extending its scope to address gaps in computational modeling.[9] More specifically, our interests in the covalent inhibition of prolyl oligopeptidase[37,38] and in zinc-containing histone deacetylases (HDACs)[39] led us to implement an accurate covalent docking method[40] and a zinc coordination method.[41]

We report herein our latest development of customized functionalities, their retrospective benchmarking on various datasets, as well as the datasets themselves. We believe that these sets can be useful for other research groups developing similar routines into their docking programs. Finally, to streamline the usage of FITTED by the medicinal chemistry community at large, several aspects have been improved to ease the access to our platform, including the development of a user-friendly graphical user interface (GUI) which can be used to quickly set up docking experiments (Figure S1).

STATUS, THEORY, AND IMPLEMENTATIONS

Covalent docking. Reactions observed in covalent inhibition and encoded into FITTED are now classified into five classes (examples of PDB files featuring these reactive groups and reactive residues are given in brackets):

Class 1. The reactive protein residue (e.g., a catalytic serine) reacts with an electrophilic warhead, and the bond order of the electrophile changes, hence the hybridization of the atoms (e.g., sp^2 to sp^3). The electrophile then abstracts the residue's acidic proton (e.g., to become a hemiacetal). Among these reactive groups are aldehydes (PBD 1ME3), ketones (PDB 1BGO), Michael acceptors (PDB 1F29), isothiocyanates (PDB 6PQP) and terminal alkynes (PDB 6QBS). Michael acceptors are not limited to α,β-unsaturated amides and esters but also include α,β-unsaturated sulfonamides and nitro alkenes (PDB 1ZK5).

Class 2. The reactive residue reacts with the electrophilic warhead, but there is no change in the hybridization of the electrophile. A basic residue in the binding site abstracts the acidic proton of the residue. This is observed with boronic acids (PDB 1VGC) reacting with catalytic serine residues.

Class 3. The reactive residue attacks an electrophilic warhead, a leaving group escapes, and a basic residue abstracts the proton. This is observed with activated chloro phenyls (*p*-nitro in PDB 3B0R), α-halo ketones (PDB 1AIM), disulfide exchanges with cysteine residues (PDB 1F4C), phosphonyl chloride (PDB 1LPM), sulfonyl fluoride (PDB 5U8L), fluoro-amidines (PDB 2DW5), aryl sulfones (PDB 4ONM) and thionyl esters (PDB 6CL0).

Class 4. The reactive residue attacks the electrophilic warhead, a bond is broken opening a ring (no leaving group escapes) and the ligand abstracts the residue's acidic proton. Among these groups are epoxides (PDB 1CVZ), aziridines (PDB 5V6V), β-lactams (PDB 1B12), β-lactones (PDB 1AWH), and hydroxy-isothiazoles (PDB 2IJN).

Class 5. A negatively charged reactive residue (with no proton) attacks the electrophilic warhead, a bond is broken, and a stable ion is released. As an example, we can cite the opening of a β-lactone by glutamic acid (PDB 4WV7) through attack of a sp^3 carbon of the lactone with release of a carboxylate.

The encoding first required a complete revision of our functional group recognition algorithm which now not only identifies reactive groups, but also leaving groups and various tautomeric states of each warhead. This identification is performed by SMART, a pre-processing tool which prepares small molecules for docking.

To enable proton exchanges, up to three dummy hydrogen atoms are added to the system by FITTED prior to the actual docking (Figure 1): one to the neighboring basic residue (if any), one to the warhead (if required) and one to the reactive residue (if required). The status (ON or OFF) of these protons will depend on the class of reaction under investigation. If FITTED identifies a candidate pose as non-covalent (warhead too far from the reactive residue), the proton will remain on the reactive residue as shown in Figure 1A, unless the reaction is from class #5 (negatively charged residue). If from classes #2 or #3, the proton is virtually moved to the basic residue (Figure 1B). If the pose is identified as covalent and if the reaction is from class #1 or class #4, the proton on the warhead is turned ON and the one on the active residue is turned OFF to represent a proton transfer (Figure 1C).
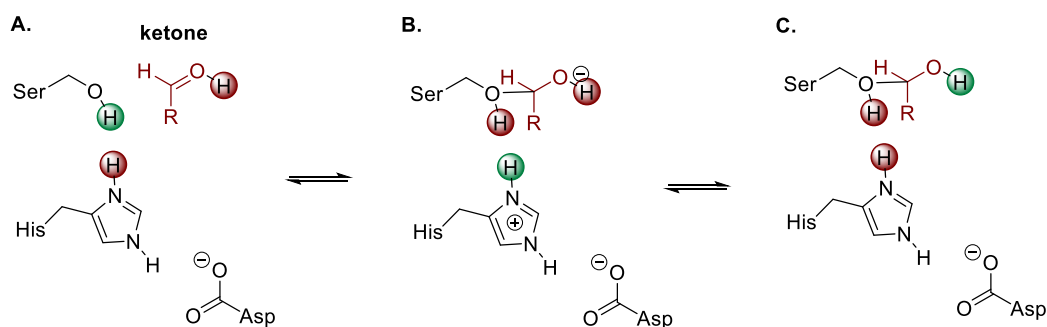
Figure 1. Proton exchanges between groups. This example shows a serine-histidine-aspartate triad that can be found in the active site of POP. In green, protons in the ON state; in burgundy, protons in the OFF state.

Similarly, to form bonds, dummy bonds are added when the reaction class is identified and turned ON or OFF depending on the covalent or non-covalent nature of the binding. Meanwhile, bond order and hybridization are adjusted as the reaction proceeds (sp^2 aldehyde into sp^3 acetal in Figure 1) and leaving groups "disappear" after the reaction is complete (and should be restored if a non-covalent pose is deemed better). This is achieved by assigning each bond and atom a covalent and a non-covalent form (ON or OFF, single, double or triple bond hybridization). If the reactive group is found within a user-defined cutoff distance of the reactive residue (default: covalent bond length +1.8 Å), the covalent form is selected, otherwise the non-covalent form is used.

We have implemented seven residues that can react covalently. While cysteine residues have historically been the major target for covalent drugs (in particular for targeting kinases),[42] catalytic (i.e., activated) serines,[43,44]

lysine,[45] aspartic acids,[46] glutamic acids,[47] tyrosines,[48] and histidines[49] have also been used to bind covalent drugs or reactive probes. These different residues have now been implemented into FITTED. The user must instruct the program which reactive residue and basic neighboring residue (if any) are to be considered. FITTED will artificially duplicate this residue, add the copy (additional bonds and atoms) to the small molecule, and annotate it as ON when covalent and OFF when non-covalent. In tandem, the same residue will be kept as part of the protein and will be annotated as ON when non-covalent and OFF when covalent. The use of dummy atoms and annotated bonds and atoms facilitates the use of a constant number of bonds and atoms throughout the entire docking process. When the list of interactions (angles, torsions, non-bonded) is created, they are annotated as "covalent only" or "non-covalent only" and "always" to ensure that the molecular mechanics calculations are considering only the interactions related to each of the poses, whether covalent or non-covalent.

Reaction mechanism specificities. Several challenges arose when developing this new covalent docking functionality. First, some groups are prone to undergo different reactions depending on their environments. For example, aldehydes react with catalytic serine residues to form the corresponding hemiacetals. However, in many serine proteases (e.g., PDB 1QFS), an oxyanion hole stabilizes the developing negative charge, and the proton is not transferred from the serine to the acetal but rather to the catalytic histidine (Figure 2A). Alternatively, aldehydes can also form imines with lysines. Second, as shown in Figure 2B, the hydroxy-isothiazole has a resonance structure which is more electrophilic and will react with a cysteine to form a disulfide bridge (PDB 2IJN). However, a much more complex mechanism with inversion of double bond configuration has also been proposed which may explain the possible (E) double bond configuration observed in the crystal structure. Third, when reacting with an epoxide, the reaction is stereospecific with inversion of configuration.[50] Thus, when docking an epoxide covalently, the forming bond should consider this stereochemical characteristic (Figure 2C – PDB 1CVZ). Fourth, more complex and less usual mechanisms such as the S_N2' adopted by Moxalactam (PDB 1I5Q),[51,52] the furan opening of Wortmannin (PDB 1E7U) likely resulting from a conjugate addition-retro addition sequence[53] or the tosyl displacement of BAY11-7082 also resulting from a similar sequence (PDB 4ONN),[54] are more difficult to encode and deemed too specific to be of interest for most users (Figure 3). However, some complex mechanisms appearing in the literature have been included into our program. For example, catechol[55] (Figure 2D – PDB 3V4J) or activated alkyl fluoride[56] (Figure 2E – PDB 2GH5) can be converted into Michael acceptors under specific redox conditions.
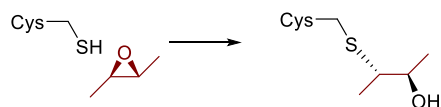
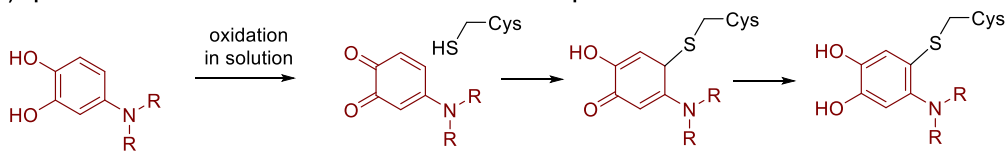**A) Oxyanion stabilization of developing negative charge - PDB: 1QFS**
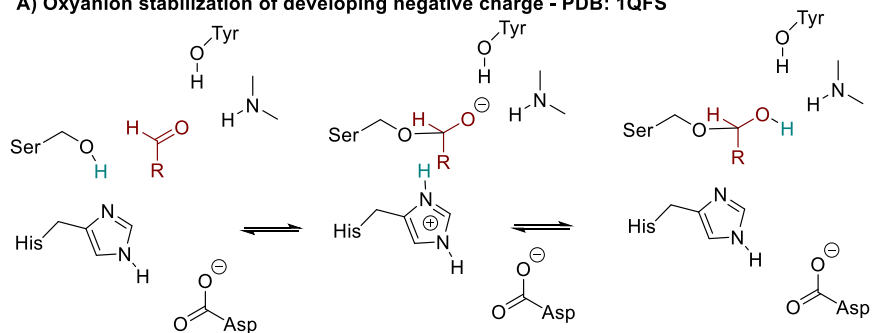
B) Multiple resonance structures and mechanisms possible - PDB: 2IJN
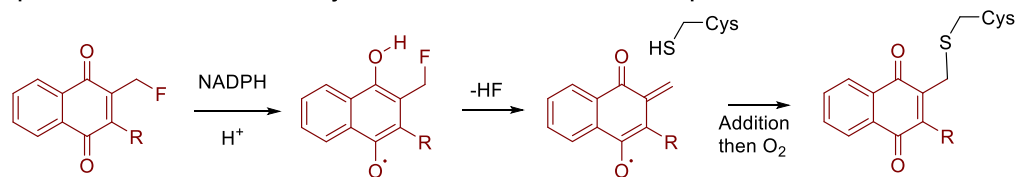
alternative mechanism
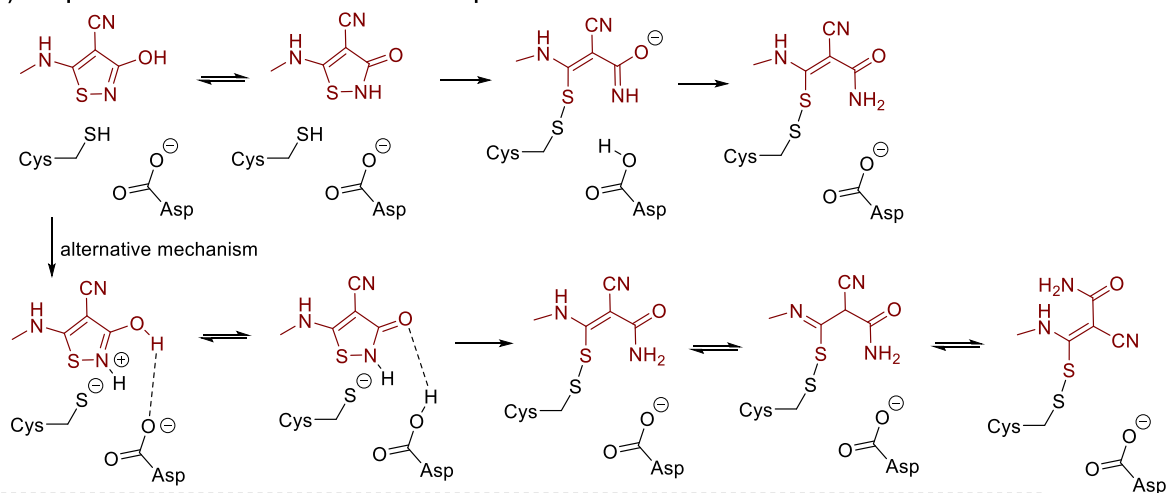
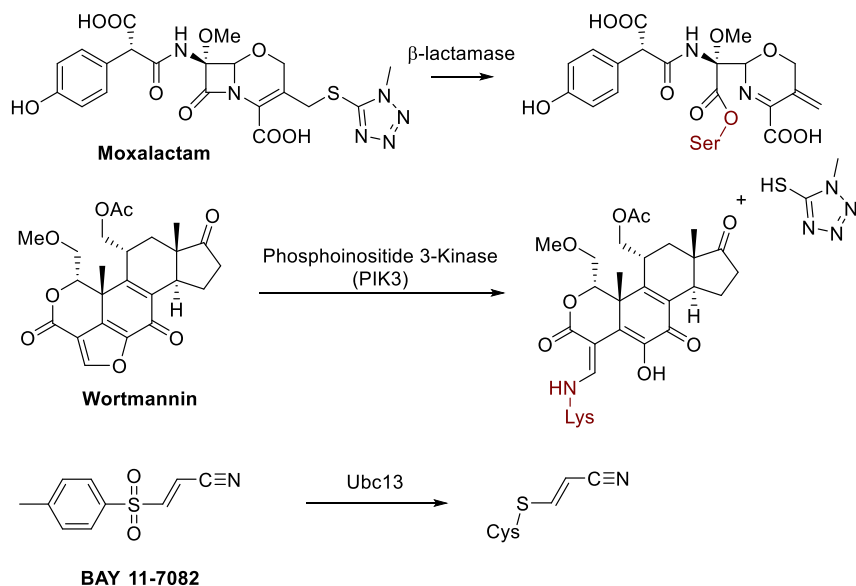**C) Stereochemistry characteristics: PDB: 1CVZ**

D) Specific mechanism - catechol conversion to Michael acceptor - PDB: 3V4J

oxidation in solution

E) Specific mechanism - activated alkyl fluoride conversion to Michael acceptor - PDB: 2GH5

NADPH

H$^+$

-HF

Addition then O$_2$

Figure 2. Examples of covalent binding modes in less common mechanisms.

6

Figure 3. Drug-specific mechanisms.

To address these issues, additional implementations were required. First, to identify what the product should be (aldehyde into hemiacetal and/or negatively charged hemiacetal or imine), the program is informed of the reactive residue and of the presence of the oxyanion hole using specific parameters. The product will then be automatically identified depending on the reactive group and residue (aldehyde + serine → hemiacetal; aldehyde + lysine → imine). Second, the various possible protonation states of the warheads have been implemented to ensure the docking will not fail as a result of an improper protonation assignment. Third, the stereochemistry inversion is automatically identified (reaction specific) and a routine ensuring proper stereochemistry in the covalently bound ligand has been added.

Docking to metalloproteins. Since the development of the initial zinc parameterization proved successful, we followed the same recipe for three other classes of proteins: enzymes featuring neutral Zn complexes, heme-Fe, and one or two Mg ion systems with strict coordination geometry. Our approach began with obtaining representative crystal structures to be used for quantum mechanics (QM) evaluation of iron, zinc, and magnesium coordination (see Supporting Information for complete set). To reduce the computational cost of the calculations, we truncated the crystal structures' active sites such that only the metal ion and the essential amino acids for enzymatic reactivity are present. For iron-containing enzymes, we focused exclusively on cytochrome P450s (CYPs), where the iron ion is usually coordinated by a heme and a cysteine residue. Then, for each class of proteins, we assembled training sets containing small molecules and fragments known to bind these metals. For example, the zinc-cysteine training set is comprised of generic thiol, sulfonamide, hydroxamic acid, and water. The CYP training set is comprised of nitrogen containing aromatic and aliphatic five and six membered rings. The rings contain a basic nitrogen atom and other heteroatoms present at different positions, as well as a wide variety of electron withdrawing and donating

groups at the *ortho*, *para* and *meta* positions (Chart S1). For magnesium, phosphate, alcohol and carbonyl containing groups were selected.

For each metal ion, we explored the potential energy surface (PES) of ligand binding, starting from a distance of 10.0Å, where there is little interaction between the ligand and the metal, and going to 1.6Å, where there is high repulsion between the ligand and the metal. We used increments of 0.2Å, thus obtaining 43 unique structures across the PES. At every point on the PES, the ligand was optimized while the metal ion and surrounding residues were frozen in the crystal structure geometries. An example of a representative PES for iron-ligand coordination is given in Figure 4, along with snapshots of the binding process at 10.0 Å (no interaction), 2.0 Å (attractive interactions) and 1.6 Å (repulsive interactions).
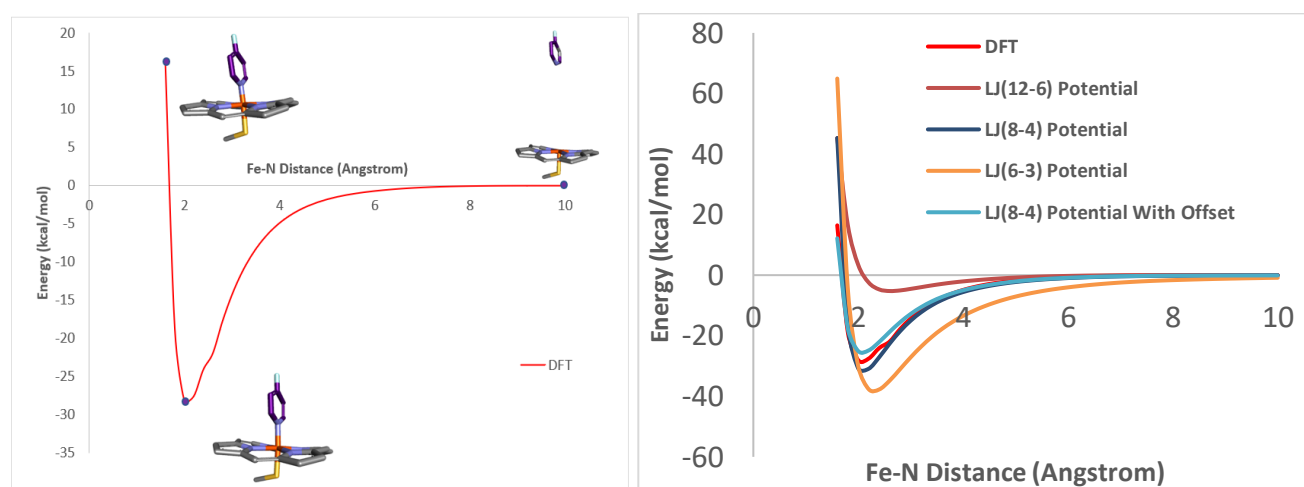
Figure 4. Left: PES scan example for binding of ***p*-fluoro-pyridine** to heme. Level of theory: PBE0(D3BJ)/def2-SVP/ LANLDZ(Fe).[57] Right: Overlay of QM and FITTED energy profiles obtained for compound *p*-fluoro-pyridine (mean unsigned error—MUE— is 0.63 kcal/mol).

A new MM Lennard-Jones (LJ) potential was fit to the acquired QM data and implemented into FITTED for each metal ion. An overlay of the energy profiles obtained with QM and FITTED for a representative ligand for iron-ligand coordination is given in Figure 4 (orange curves). Compared to iron coordination, very high binding energies (greater than 200 kcal/mol) were computed for phosphate-magnesium complexes consistent with previous reports.[58] As a result, greater MUE values were obtained, ranging from 1—20 kcal/mol in most cases. However, these large binding energies are balanced by the displacement of a water molecule from the magnesium coordination sphere (10-13 kcal/mol[58]) and the desolvation of phosphate oxygens (ca. 20 kcal/mol[58]).

As can be seen in Figure 4, the existing FITTED LJ (12-6) potential, as well as a LJ (6-3) potential, were unable to describe the iron-nitrogen coordination process for the ion coordination. In contrast, the energy minimum described by the LJ (8-4) potential (dark blue curve) is very close in energy to the QM one. In contrast, coordination to zinc and magnesium was optimally modelled with a LJ (6-3) potential. As previously observed,[41] the minimum

described by the LJ (8-4) (iron) and LJ (6-3) (zinc and magnesium) potentials was offset by 0.2 to 0.25Å when compared to the QM minimum. To verify whether incorporating the offset in the computation of the LJ potentials would improve the energy profile described by FITTED, we plotted the potential containing the offset (Figure 4). This led to an optimal overlap between the QM and FITTED energy profiles. Our analysis showed that an offset of 0.2Å was required with iron while 0.25Å was preferred for the other two metals. We then implemented the modified potential shown in Eq. 1 into FITTED for describing the iron-nitrogen binding process (the LJ (6-3) potential previously implemented for zinc was extended to magnesium). Our study also showed that the zinc binding groups behave similarly whether binding to cysteine-coordinated zinc ion or aspartate/glutamate coordinated zinc ions. As a result, no special parameters were needed for this class of zinc-containing enzymes.

$$E_{vdW(8-4)-\text{FITTED}} = \frac{4\varepsilon\sigma^8}{(r-0.2)^8} - \frac{4\varepsilon\sigma^4}{(r-0.2)^4} \qquad (1)$$

In the case of magnesium, we needed to also compute the effect of the coordination geometry around magnesium. We used 10 of the 20 model systems mentioned above, moved ligands in the three perpendicular directions as shown in Figure 5, derived and potential energies using B3LYP-D3BJ/6-31G*. The angle and torsion parameters around the O–Mg bond were then derived to optimize the match between the MM and QM-derived relative energies. A difficulty in modeling coordination geometry using force fields is the consideration of multiple possible angles for the same group types. For example, magnesium often has six ligands in an octahedral geometry and may have two glutamates both axial (ca. 180 °) or both equatorial (90º). In other words, traditional polynomial potentials with a single energy minimum cannot be used. Rather, we choose a periodic function similar to the one used for torsion angles (Equation 2). This function with a non-zero V4 value would have minima at 0 °, 90 ° and 180 °. More advanced effects such as the Jahn-Teller effects will not be considered here. Unfortunately, some systems appeared more dependent on coordination geometry than others and average values were ultimately selected.

$$E = V_n/2[1 + \cos(n\phi - \phi_0] \qquad (2)$$

The coordination energy was computed at various positions and the corresponding FF parameters were derived (Figure 5).
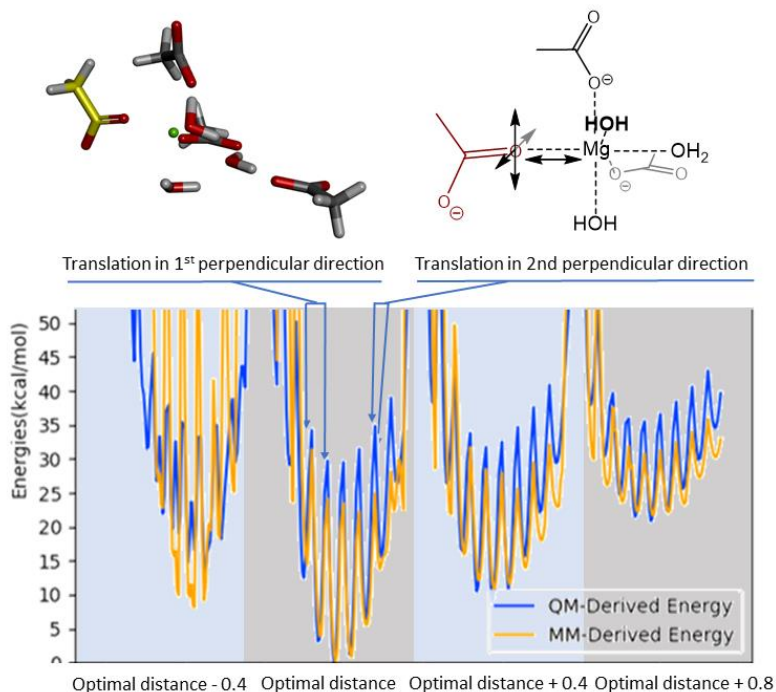
Figure 5. Modeling the metal coordination geometry. Blue: QM derived relative energy; orange: FITTED energy.

Binding to zinc, heme or magnesium is only possible if water present in the resting state is displaced. Thus, as was done previously with zinc, a routine has been implemented that adds water molecules to unoccupied coordination sites of iron and magnesium ions considering an octahedral geometry, with a free energy of displacement of 20.6 (iron), and 24.6 kcal/mol (magnesium).[59] As FITTED considers water molecules displaceable, it will be possible to simulate the displacement of one (or more) water molecules by a ligand. In addition, considering these resting states is critical for optimal binding as these water molecules may be tightly bound and hardly displaced by weak ligands. Also, water-bridged complexes may be common with CYPs as shown by EPR studies of CYP-ligand complexes in solution.[60] With these displaceable water molecules, both directly coordinated ligands and water-bound ligands may be predicted.

BENCHMARKING DATASETS.

To evaluate FITTED's ability to dock covalent drugs and metalloprotein inhibitors, we sought large, diverse testing sets of high-quality structures from the PDB. This led to a refinement of our existing datasets, which enabled us to benchmark our programs.

Covalent docking set. In 2018, Scarpino, Ferenczy, and Keserű reported a comparative study of docking programs applied to covalent docking.[61] The dataset of 208 complexes used in this study was selected and was

expanded to include targeted amino acids beyond cysteine and serine and to include a large variety of warheads. This enhanced set is comprised of nearly 300 structures (Table S2).

Metalloproteins. In 2014, we reported our initial validation set for zinc metalloenzymes, containing 123 protein/ligand complexes.[41] For iron-containing enzymes, we assembled a representative set of 85 CYP protein-ligand complexes from the PDB, with and without iron-coordinating nitrogen ligands. For magnesium, we assembled a set containing 29 diverse crystal structures. As a note, the MetalPDB database was very useful in identifying metal-containing enzymes.[62]

RESULTS AND DISCUSSION

Covalent docking. As reviewed recently, many docking programs now enable the docking of covalent inhibitors.[26] However, the process often requires the manual identification of reactive functional groups, or covalent warheads, and precludes the use of these programs in large virtual screening campaigns. In addition, when the covalent mode is selected, all the molecules are docked covalently, regardless of whether the covalent mode should be preferred or not. In our original covalent docking development, FITTED was used to discover prolyl oligopeptidase (POP) inhibitors.[40] This class of serine proteases (S9 proteins also including DPP-IV and fibroblast activation protein-α) has primarily been targeted using nitrile, aldehyde, and boronic acid derivatives. To customize our software and facilitate the discovery of POP covalent inhibitors, these three functional groups were hard-coded into the program and this version of FITTED was successfully used to design potent inhibitors.[40] In addition, a mode was developed to enable both covalent and non-covalent binding modes to be considered simultaneously, enabling the program to identify the most likely mode of binding. Since then, we have started a large medicinal chemistry program to identify covalent inhibitors of SARS-CoV-2 3CL^{pro} and PL^{pro}, two enzymes featuring a catalytic cysteine residue.[63] To respond to this need, we decided to redesign our program and rewrite the entire covalent docking routine such that it includes many different warheads and reactive residues.

Docking to metalloproteins. Several drug targets include a metal ion in their catalytic site. Previously, our software FITTED was designed to accurately compute the energetics and geometrical considerations of a ligand in the active site of a zinc metalloprotein and to model the exchange of a proton between a ligand and a neighboring residue.[41] This methodology was developed for zinc ions with histidine and glutamic acid residues in their coordination sphere. However, our recent attempts to design inhibitors of AID has shown that the binding of ligands to zinc ions that feature cysteine residues in the coordination sphere (a neutral zinc complex) is not accurately described with the current parametrization.

In addition to zinc, iron is present in many important proteins such as CYPs, which are involved in the metabolism of most drugs on the market.[3,64-66] The co-administration of drugs, one metabolized by CYP and the other inhibiting

CYP, leads to drug-drug interactions where there is accumulation of the unmetabolized drug which can lead to severe toxicity. One strategy to prevent drug-drug interactions is to computationally predict CYP inhibition and include this routine in drug development software like FITTED. Reversible CYP inhibition is observed with several molecules termed Type II ligands that contain a basic nitrogen atom which may coordinate the heme iron, (Figure 6a).[67] An example of such ligand is ketoconazole, an antifungal compound that inhibits CYP3A4. To correctly model Type II ligands and thus better predict CYP inhibition, we implemented nitrogen-iron coordination in FITTED. This functionality is lacking in the majority of SBDD software, with some exceptions, of which we mention EADock. This docking program has been trained to predict heme coordination, with a self-docking accuracy of 62% on a small set of complexes (31/50 heme structures that exhibited iron-nitrogen coordination).[68]

While zinc is often used to cleave peptide or protein amide bonds and iron is used for its redox properties, magnesium (and manganese) ions are often used by enzymes involved in nucleic acids[69] biochemistry, as well as by human ubiquitin hydrolase[70] and pyruvate carboxylase.[71] Complicating matters, some proteins feature more than one magnesium and/or manganese ion in their catalytic site. For example, the Flap Endonuclease 1, involved in DNA replication and repair and a potential target for cancer therapeutics, and the restriction endonuclease (EcoVR, Figure 6b) require two magnesium ions for phosphate activation.[72] In contrast to soft zinc ions, which do not have a very strict preference for a specific geometry, magnesium has a strong preference for six ligands (preferentially oxygen containing) with very well defined coordination geometry.[73,74] For optimal docking, this factor should be considered.
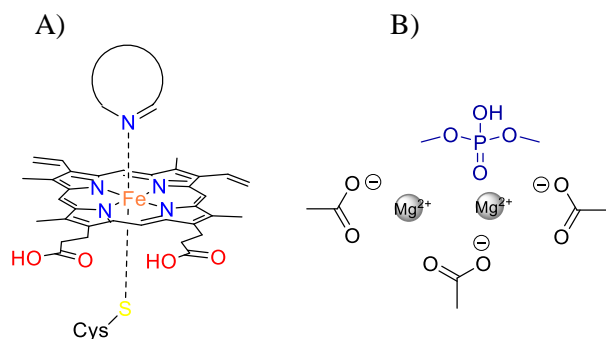

Figure 6. A) Heme bound to a nitrogen-containing inhibitor; B) Phosphate hydrolysis by EcoVR adapted from [72].

Considering these different requirements, special consideration for ligand-metal binding (both energy- and geometry-wise) must be given to these classes of proteins and specific implementations must be devised. As described in the previous section, specific force field terms and parameters were developed.

Evaluating the software. Before looking at the new features described above, we tested our modified programs on existing benchmark sets and assessed whether the previously implemented routines were still as effective. To perform this evaluation, the lowest energy pose over 10 runs of docking for each ligand-protein complex was

compared to the cognate ligand pose. As has been widely used in most docking evaluations, a root mean square deviation (RMSD) of 2.0 Å was considered as the threshold to distinguish poor and good predictions. The Astex set, previously used by numerous docking program users, our large self-docking/cross-docking set reported previously and expanded over time and our zinc metalloenzyme set were selected for this preliminary validation.

A)

B)

C)
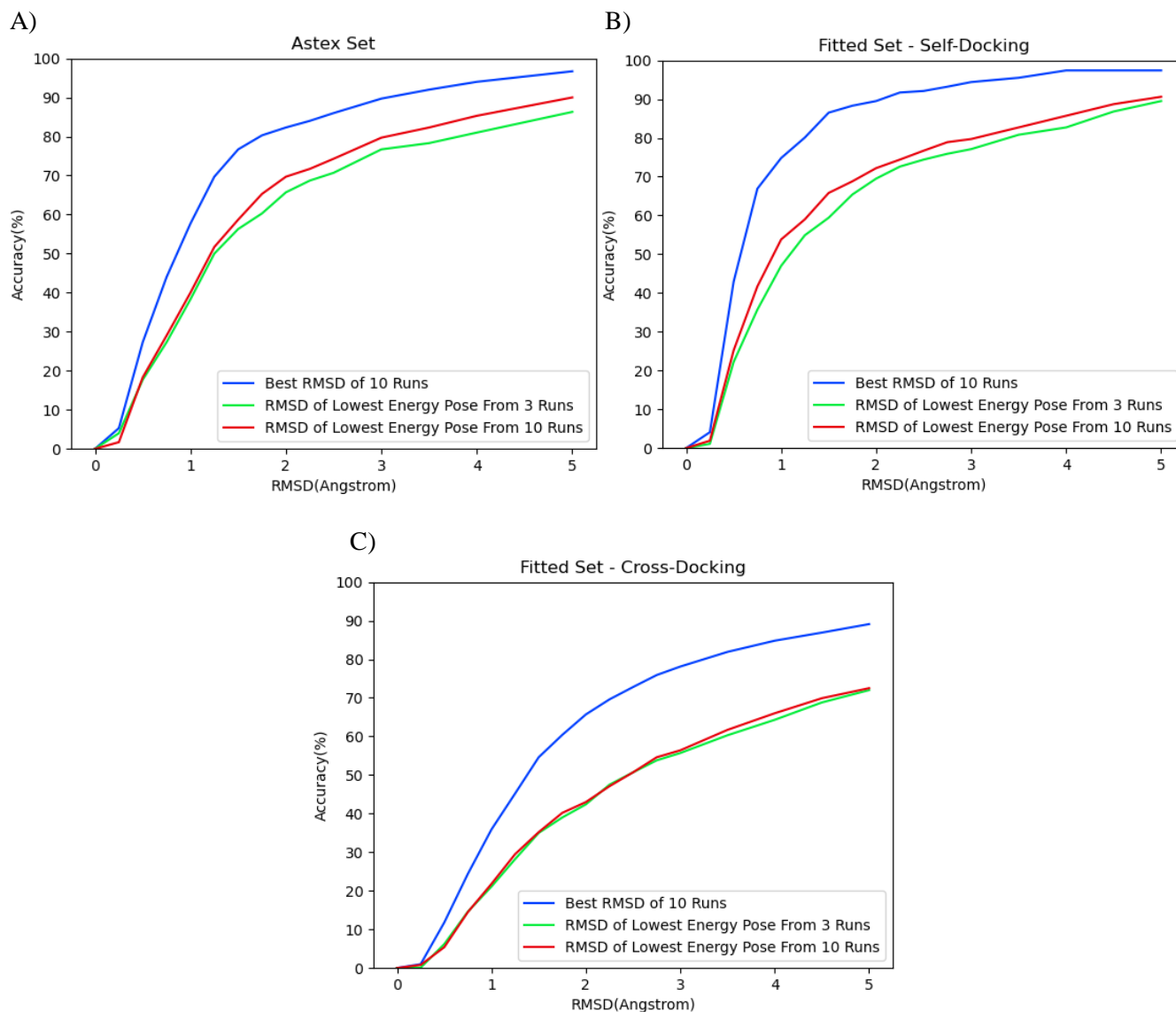
Figure 7. Validation on previous sets. A) Astex set (299 protein-ligand complexes, set #1); B) FITTED set – self docking (306 protein-ligand complexes, set #2) [75]; C) FITTED set – cross-docking (1,210 protein-ligand complexes, set #3) [75]. See Supporting Information for more complete data.

Docking of these three sets ran uneventfully confirming that the programs were stable and pose prediction accuracy was evaluated (Figure 6). The measured accuracies were similar to those reported earlier (2014: set #2 and 3: 73%

(3 runs), 40% (3 runs); current work: 72% (3 runs), 47% (3 runs)). We were content to see that our programs provided accurate results with the Astex set (70% with 10 runs). We then moved to virtual screening to evaluate whether our program still performs as previously reported (ca. 75% AUROC over the 102 proteins of the DUDe set with AUROC over 80% with 39 proteins and over 90% with 12 proteins).[75] This first set of evaluation experiments confirmed that our programs are stable and that our new implementation did not affect the accuracy.

Covalent docking. Significant changes have been made to our programs SMART (identification of covalent warhead) and FITTED (docking covalently). To evaluate the updated modules, we first started with a set reported by Scarpino *et al.*[61] In their study, they showed that FITTED was a top-performer although 15% of the cases could not be evaluated as the reaction mechanisms and/or warheads were not implemented. The automated warhead recognition and pre-defined mechanisms implemented into FITTED enable its use in large screenings, but limit it to those implemented groups and mechanisms. To overcome this limitation, we have rewritten the covalent docking routines to facilitate the addition of novel mechanisms and warheads. Using these new routines, we implemented ca. 20 mechanisms and over 50 common warheads (Table S1). With these new mechanisms and warheads, FITTED's accuracy (reported to be 56% on 175 cases by Keserű and co-workers) increased to 71% (on the entire set of 208 complexes, Figure 8A). As noted by Keserű, a number of failures comes from the scoring as a pose with RMSD below 2Å is found in nearly 90% of the cases in at least one of the 10 runs (but not necessarily identified as the most favoured). However, the convergence of the conformational search contributes to the problem as the accuracy substantially increased (64 to 71%) when going from 3 runs to 10 runs.
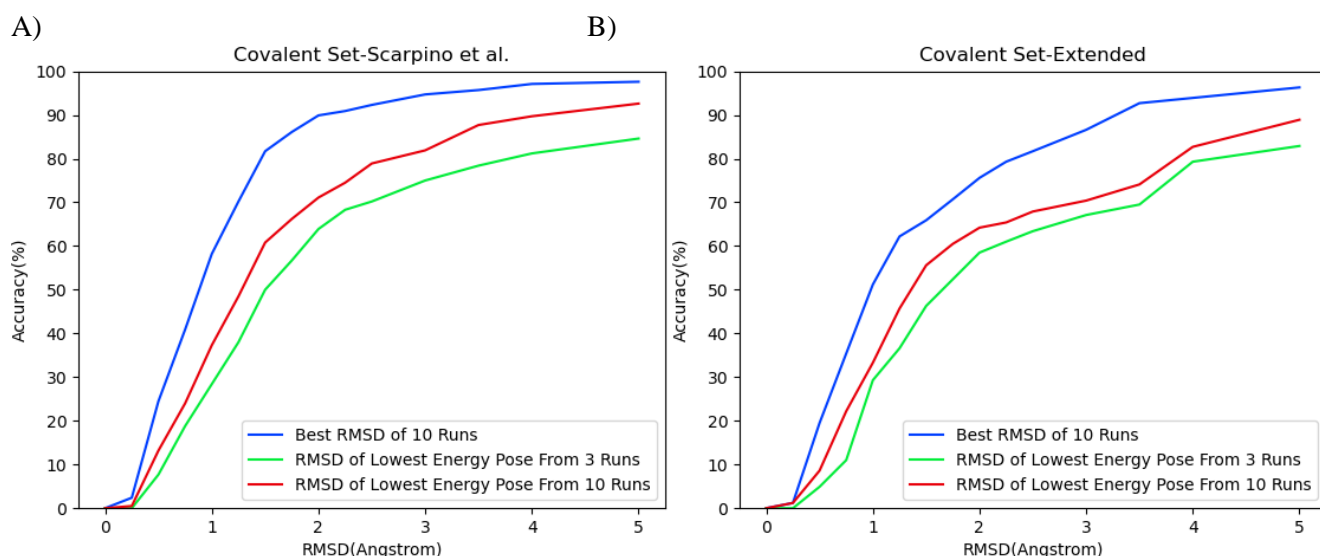

Figure 8. Accuracy of covalent docking. A) set reported by Scarpino *et al.*[61]; B) extended set.

Next, we evaluated our new covalent docking on the entire extended set which covers a much larger diversity of warheads and reacting residues. While the measured accuracy is lower (Figure 8B), the difference between 3 and

10 runs remained approximately the same. Interestingly, pose prediction when considering 10 poses (lowest RMSD out of 10 runs) is also less accurate for these less common warheads, although still acceptable (64% below 2 Å). As discussed below, the scoring functions in FITTED are likely the cause of this observation.

Metalloproteins. The sets made of 123 zinc-containing protein-ligand complexes (Figure 9A),[41] 85 heme-containing systems and 29 magnesium-containing systems (Figure 9B) were used for a total of 237 metalloproteins. To our knowledge, no dataset of this magnitude has previously been compiled to widely test docking programs and their applicability to this target class.

A)

B)
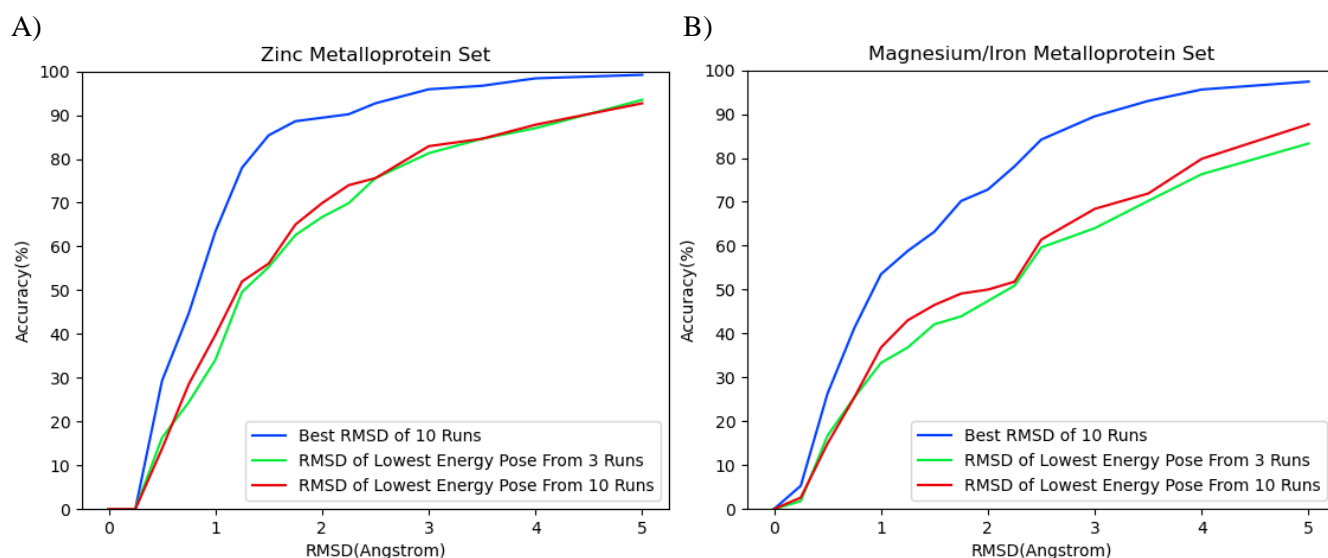
Figure 9. A) Accuracy of metalloprotein mode in the self-docking of ligands to zinc; B) Accuracy of metalloprotein mode in the self-docking of ligands to magnesium or heme-containing proteins.

We first reported FITTED for zinc-containing metalloproteins in 2014 and validation on the 123 complexes revealed an accuracy of 64% (10 runs).[41] Our investigations on proteins including a neutral zinc-complex (e.g., zinc ion coordinated by 2 cysteine thiolates) revealed that our parameters previously developed with charged zinc complexes (e.g, zinc coordinated by 3 histidine residues) were accurately modeling the coordination energy profiles and no further parameters were needed. We were pleased to see that the accuracy of the current version has not been affected by the latest changes and even improved (70%) thanks to changes we made throughout the years on the preparation of the ligands for docking. We were, however, a little disappointed by the observed accuracy for heme and magnesium containing enzymes with an accuracy of about 50% (top pose) and 73% (best pose out of 10 runs) which is lower than with the other protein families.

From these studies, we have identified two areas to focus for improvement with scoring functions as the first. Whereas FITTED often finds a correct pose out of 10 runs, it is not always identified as the most favored. Better selection of the best pose candidate would obviously increase the pose prediction accuracy and the accuracy of

virtual screening campaigns should, in turn, improve. Second, our evaluation of FITTED in virtual screens revealed an AUROC below 65% in 17/102 proteins from the DUD-e set. Whereas some improvement is expected from better pose prediction/selection, predicting ligand-protein affinity needs additional consideration.

CONCLUSION

We have implemented many new features in FITTED that will streamline the drug discovery process and allow the automated screening of large libraries of compounds. First, we expanded the scope of our covalent docking methods. To do so, we modified our functional group recognition algorithm to handle several additional reactive warheads and we implemented eight residues that are known to react covalently with ligands. The covalent bond formation between a warhead and a residue is now modelled via the use of dummy hydrogen atoms and bonds that are turned ON or OFF. Furthermore, FITTED identifies if a molecule preferentially binds covalently or non-covalently.

Second, we expanded the scope of our metalloenzyme docking routines. Previously, only docking to zinc bound to aspartate or glutamate was implemented; FITTED is now suited to dock on metalloenzymes containing a neutral (cysteine) or charged (aspartate/glutamate) zinc complex, a heme-iron complex, and one or two magnesium ions. To describe the ligand-metal coordination process, we derived Lennard-Jones energy profiles from high level calculations. We also implemented the addition of water molecules to unoccupied coordination sites of the metal centers.

Benchmarking on a number of datasets (new and previously reported) demonstrated the accuracy of this current version of FITTED with pose prediction accuracy above 70% for most systems in self-docking (threshold RMSD < 2Å) and of about 50% in cross-docking (threshold RMSD < 2.5Å) whether covalent or non-covalent, the accuracy being a little lower for magnesium and iron-containing metalloproteins (ca. 50% in self-docking). Interestingly, a correct pose is found in about 90% of the test cases when 10 poses are generated. This points to our scoring functions as a main area for improvement.

SUPPORTING INFORMATION

Detailed methodology, datasets, description of the user interface. The complete datasets used for pose prediction evaluation (protein and ligand structures in mol2 format) are available on our website (http://moitessier-group.mcgill.ca/software.html).

CONFLICT OF INTEREST

The authors declare the following competing financial interest: The developed tool is distributed by Molecular Forecaster co-founded by N.M. (available for free to academia). The program is available on request (free license for academic purposes, license with purchase for commercial purposes) from the authors (www.molecularforecaster.com).

ACKNOWLEDGMENT

REFERENCES

(1) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Health Econ.* **2016**, *47*, 20-33.

(2) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **2020**, *323*, 844-853.

(3) Dalvie, D.; Kalgutkar, A. S.; Chen, W. Practical approaches to resolving reactive metabolite liabilities in early discovery. *Drug Metab. Rev.* **2015**, *47*, 56-70.

(4) Kaitin, K. I. Deconstructing the drug development process: the new face of innovation. *Clin. Pharmacol. Ther.* **2010**, *87*, 356-361.

(5) Taft, C. A.; da Silva, V. B.; da Silva, C. H. T. d. P. Current topics in computer-aided drug design. *J. Pharma. Sci.* **2008**, *97*, 1089-1098.

(6) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* **1998**, *19*, 1639-1662.

(7) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein–ligand docking: Current status and future challenges. *Prot. Struct. Funct. Genet.* **2006**, *65*, 15-26.

(8) Corbeil, C. R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47*, 435-449

(9) Moitessier, N.; Pottel, J.; Therrien, E.; Englebienne, P.; Liu, Z.; Tomberg, A.; Corbeil, C. R. Medicinal chemistry projects requiring imaginative structure-based drug design methods. *Acc. Chem. Res.* **2016**, *49*, 1646-1657.

(10) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470-489.

(11) Rarey, M.; Kramer, B.; Lengauer, T. Multiple automatic base selection: Protein-ligand docking based on incremental construction without manual intervention. *J. Comput. Aided Mol. Des.* **1997**, *11*, 369-384.

(12) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid,

Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739-1749

(13) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750-1759

(14) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727-748.

(15) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Prot. Struct. Funct. Genet.* **2003**, *52*, 609-623.

(16) Totrov, M.; Abagyan, R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Prot. Struct. Funct. Genet.* **1997**, *29*, 215-220.

(17) Schapira, M.; Totrov, M.; Abagyan, R. Prediction of the binding energy for small molecules, peptides and proteins. *J. Mol. Recogn.* **1999**, *12*, 177-190.

(18) 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. Comp. Chem.* **1994**, *15*, 488-506

(19) Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499-511

(20) Spitzer, R.; Jain, A. N. Surflex-Dock: Docking benchmarks and real-world application. *J. Comput. Aided Mol. Des.* **2012**, *26*, 687-699.

(21) Li, J.; Liu, W.; Song, Y.; Xia, J. Improved method of structure-based virtual screening based on ensemble learning. *RSC Adv.* **2020**, *10*, 7609-7618.

(22) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein–Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model.* **2017**, *57*, 942-957.

(23) Chen, L.; Cruz, A.; Ramsey, S.; Dickson, C. J.; Duca, J. S.; Hornak, V.; Koes, D. R.; Kurtzman, T. Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. *PLoS ONE* **2019**, *14*, e0220113.

(24) Sieg, J.; Flachsenberg, F.; Rarey, M. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2019**, *59*, 947-961.

(25) Sutanto, F.; Konstantinidou, M.; Dömling, A. Covalent inhibitors: a rational approach to drug discovery. *RSC Med. Chem.* **2020**, *11*, 876-884.

(26) De Cesco, S.; Kurian, J.; Dufresne, C.; Mittermaier, A.; Moitessier, N. Covalent inhibitors design and discovery. *Eur. J. Med. Chem.* **2017**, *138*, 96-114.

(27) Fields, G. B. The Rebirth of Matrix Metalloproteinase Inhibitors: Moving Beyond the Dogma. *Cells* **2019**, *8*, 984.

(28) Behera, J.; Jayaprakash, V.; Sinha, B. N. Histone deacetylase inhibitors: a review on class-I specific inhibition. *Mini Rev Med Chem* **2015**, *15*, 731-750.

(29) Cross, J. B.; Duca, J. S.; Kaminski, J. J.; Madison, V. S. The active site of a zinc-dependent metalloproteinase influences the computed pKa of ligands coordinated to the catalytic zinc ion. *J. Am. Chem. Soc.* **2002**, *124*, 11004-11007.

(30) *Iron-Containing Enzymes: Versatile Catalysts of Hydroxylation Reactions in Nature*; RSC, 2011.

(31) Roat-Malone, R. M.: Iron-Containing Proteins and Enzymes. In *Bioinorganic Chemistry*, 2007; pp 343-476.

(32) Swaminathan, R. Magnesium metabolism and its disorders. *Clin. Biochem. Rev.* **2003**, *24*, 47-66.

(33) Roat-Malone, R. M.: Copper Enzymes. In *Bioinorganic Chemistry*, 2002; pp 187-230.

(34) Tishchenko, K. I.; Beloglazkina, E. K.; Mazhuga, A. G.; Zyk, N. V. Copper-containing enzymes: Site types and low-molecular-weight model compounds. *Rev. J. Chem.* **2016**, *6*, 49-82.

(35) Alfano, M.; Cavazza, C. Structure, function, and biosynthesis of nickel-dependent enzymes. *Prot. Sci.* **2020**, *29*, 1071-1089.

(36) Crowley, J. D.; Traynor, D. A.; Weatherburn, D. C. Enzymes and proteins containing manganese: an overview. *Met. Ions Biol. Syst.* **2000**, *37*, 209-278.

(37) Plescia, J.; Dufresne, C.; Janmamode, N.; Wahba, A. S.; Mittermaier, A. K.; Moitessier, N. Discovery of covalent prolyl oligopeptidase boronic ester inhibitors. *Eur. J. Med. Chem.* **2020**, *185*, 111783.

(38) Plescia, J.; De Cesco, S.; Patrascu, M. B.; Kurian, J.; Di Trani, J.; Dufresne, C.; Wahba, A. S.; Janmamode, N.; Mittermaier, A. K.; Moitessier, N. Integrated Synthetic, Biophysical, and Computational Investigations of Covalent Inhibitors of Prolyl Oligopeptidase and Fibroblast Activation Protein α. *J. Med. Chem.* **2019**, *62*, 7874-7884.

(39) Mendoza-Sanchez, R.; Cotnoir-White, D.; Kulpa, J.; Jutras, I.; Pottel, J.; Moitessier, N.; Mader, S.; Gleason, J. L. Design, synthesis and evaluation of antiestrogen and histone deacetylase inhibitor molecular hybrids. *Bioorg. Med. Chem.* **2015**, *23*, 7597-7606.

(40) Mariaule, G.; De Cesco, S.; Airaghi, F.; Kurian, J.; Schiavini, P.; Rocheleau, S.; Huskić, I.; Auclair, K.; Mittermaier, A.; Moitessier, N. 3-Oxo-hexahydro-1H-isoindole-4-carboxylic acid as a drug chiral bicyclic scaffold: structure-based design and preparation of conformationally constrained covalent and noncovalent prolyl oligopeptidase inhibitors. *J. Med. Chem.* **2016**, *59*, 4221–4234.

(41) Pottel, J.; Therrien, E.; Gleason, J. L.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 6. Development and application to the docking of HDACs and other zinc metalloenzymes inhibitors. *J. Chem. Inf. Model.* **2014**, *54*, 254-265.

(42) Abdeldayem, A.; Raouf, Y. S.; Constantinescu, S. N.; Moriggl, R.; Gunning, P. T. Advances in covalent kinase inhibitors. *Chem. Soc. Rev.* **2020**, *49*, 2617-2687.

(43) Powers, J. C.; Asgian, J. L.; Ekici, Ö. D.; James, K. E. Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chem. Rev.* **2002**, *102*, 4639-4750.

(44) Johnson, D. S.; Weerapana, E.; Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Fut. Med. Chem.* **2010**, *2*, 949-964.

(45) Pettinger, J.; Jones, K.; Cheeseman, M. D. Lysine-Targeting Covalent Inhibitors. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 15200-15209.

(46) Jones, A.; Zhang, X.; Lei, X. Covalent Probe Finds Carboxylic Acid. *Cell Chem. Biol.* **2017**, *24*, 537-539.

(47) Martín-Gago, P.; Fansa, E. K.; Winzker, M.; Murarka, S.; Janning, P.; Schultz-Fademrecht, C.; Baumann, M.; Wittinghofer, A.; Waldmann, H. Covalent Protein Labeling at Glutamic Acids. *Cell Chem. Biol.* **2017**, *24*, 589-597.e585.

(48) Bum-Erdene, K.; Liu, D.; Gonzalez-Gutierrez, G.; Ghozayel, M. K.; Xu, D.; Meroueh, S. O. Small-molecule covalent bond formation at tyrosine creates a binding site and inhibits activation of Ral GTPases. *Proc. Natl Acad. Sci. USA* **2020**, *117*, 7131-7139.

(49) Gambini, L.; Baggio, C.; Udompholkul, P.; Jossart, J.; Salem, A. F.; Perry, J. J. P.; Pellecchia, M. Covalent Inhibitors of Protein–Protein Interactions Targeting Lysine, Tyrosine, or Histidine Residues. *J. Med. Chem.* **2019**, *62*, 5616-5627.

(50) Yan, S.; Appleby, T.; Gunic, E.; Shim, J. H.; Tasu, T.; Kim, H.; Rong, F.; Chen, H.; Hamatake, R.; Wu, J. Z.; Hong, Z.; Yao, N. Isothiazoles as active-site inhibitors of HCV NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 28-33.

(51) Trehan, I.; Beadle, B. M.; Shoichet, B. K. Inhibition of AmpC β-Lactamase through a Destabilizing Interaction in the Active Site. *Biochemistry* **2001**, *40*, 7992-7999.

(52) Khrenova, M. G.; Kulakova, A. M.; Grigorenko, B. L.; Nemukhin, A. V. Mechanism of Metallo-β-Lactamase Inhibition by Oxacephalosporin Antibiotic. *Moscow Univ. Chem. Bull.* **2018**, *73*, 155-157.

(53) Walker, E. H.; Pacold, M. E.; Perisic, O.; Hawkins, P. T.; Wymann, M. P.; Williams, R. L. Structural Determinants of Phosphoinositide 3-Kinase Inhibition by Wortmannin, LY294002, Quercetin, Myricetin, and Staurosporine. *Mol. Cell* **2000**, *6*, 909-919.

(54) Hodge, C. D.; Edwards, R. A.; Markin, C. J.; McDonald, D.; Pulvino, M.; Huen, M. S. Y.; Zhao, J.; Spyracopoulos, L.; Hendzel, M. J.; Glover, J. N. M. Covalent Inhibition of Ubc13 Affects Ubiquitin Signaling and Reveals Active Site Elements Important for Targeting. *ACS Chem. Biol.* **2015**, *10*, 1718-1728.

(55) Li, M.; Shandilya, S. M. D.; Carpenter, M. A.; Rathore, A.; Brown, W. L.; Perkins, A. L.; Harki, D. A.; Solberg, J.; Hook, D. J.; Pandey, K. K.; Parniak, M. A.; Johnson, J. R.; Krogan, N. J.; Somasundaran, M.;

Ali, A.; Schiffer, C. A.; Harris, R. S. First-In-Class Small Molecule Inhibitors of the Single-Strand DNA Cytosine Deaminase APOBEC3G. *ACS Chem. Biol.* **2012**, *7*, 506-517.

(56) Bauer, H.; Fritz-Wolf, K.; Winzer, A.; Kühner, S.; Little, S.; Yardley, V.; Vezin, H.; Palfey, B.; Schirmer, R. H.; Davioud-Charvet, E. A Fluoro Analogue of the Menadione Derivative 6-[2'-(3'-Methyl)-1',4'-naphthoquinolyl]hexanoic Acid Is a Suicide Substrate of Glutathione Reductase. Crystal Structure of the Alkylated Human Enzyme. *J. Am. Chem. Soc.* **2006**, *128*, 10784-10794.

(57) Rydberg, P.; Olsen, L. The Accuracy of Geometries for Iron Porphyrin Complexes from Density Functional Theory. *J. Phys. Chem. A* **2009**, *113*, 11949-11953.

(58) Villa, F.; MacKerell, A. D.; Roux, B.; Simonson, T. Classical Drude Polarizable Force Field Model for Methyl Phosphate and Its Interactions with Mg2+. *J. Phys. Chem. A* **2018**, *122*, 6147-6155.

(59) Peschke, M.; Blades, A. T.; Kebarle, P. Hydration Energies and Entropies for Mg2+, Ca2+, Sr2+, and Ba2+ from Gas-Phase Ion−Water Molecule Equilibria Determinations. *J. Phys. Chem. A* **1998**, *102*, 9978-9985.

(60) Lockart, M. M.; Rodriguez, C. A.; Atkins, W. M.; Bowman, M. K. CW EPR parameters reveal cytochrome P450 ligand binding modes. *J. Inorg. Chem.* **2018**, *183*, 157-164.

(61) Scarpino, A.; Ferenczy, G. G.; Keserü, G. M. Comparative Evaluation of Covalent Docking Tools. *J. Chem. Inf. Model.* **2018**, *58*, 1441-1458.

(62) Putignano, V.; Rosato, A.; Banci, L.; Andreini, C. MetalPDB in 2018: a database of metal sites in biological macromolecular structures. *Nucl. Acids Res.* **2017**, *46*, D459-D464.

(63) Stille, J.; Tjutrins, J.; Wang, G.; Venegas, F. A.; Hennecker, C.; Rueda, A. M.; Miron, C. E.; Pinus, S.; Labarre, A.; Plescia, J.; Burai Patrascu, M.; Vlaho, D.; Huot, M.; Mittermaier, A. K.; Moitessier, N. Design, Synthesis and Biological Evaluation of Novel SARS-CoV-2 3CLpro Covalent Inhibitors. *ChemRxiv* **2020**, https://chemrxiv.org/articles/preprint/Design_Synthesis_and_Biological_Evaluation_of_Novel_SARS-CoV-2_3CLpro_Covalent_Inhibitors/13087742.

(64) Guengerich, F. P. Cytochrome P450 and chemical toxicology. *Chem. Res. Toxicol.* **2008**, *21*, 70-83.

(65) Stjernschantz, E.; Vermeulen, N. P. E.; Oostenbrink, C. Computational prediction of drug binding and rationalisation of selectivity towards cytochromes P450. *Exp. Opin. Drug Metab. Toxicol.* **2008**, *4*, 513-527.

(66) Lin, J. H.; Lu, A. Y. H. Inhibition and Induction of Cytochrome P450 and the Clinical Implications. *Clin. Pharmacokin.* **1998**, *35*, 361-390.

(67) Ahlström, M. M.; Zamora, I. Characterization of Type II Ligands in CYP2C9 and CYP3A4. *J. Med. Chem.* **2008**, *51*, 1755-1763.

(68) Röhrig, U. F.; Grosdidier, A.; Zoete, V.; Michielin, O. Docking to heme proteins. *J. Comp. Chem.* **2009**, *30*, 2305-2315.

(69) Cowan, J. A.: Magnesium Enzymes and Models. In *Encyclopedia of Catalysis*.

(70) Piovesan, D.; Profiti, G.; Martelli, P. L.; Casadio, R. The human "magnesome": detecting magnesium binding sites on human proteins. *BMC Bioinf.* **2012**, *13 Suppl 14*, S10-S10.

(71) Chen, P.; Bornhorst, J.; Aschner, M. Manganese metabolism in humans.

(72) Zahran, M.; Berezniak, T.; Imhof, P.; Smith, J. C. Role of magnesium ions in DNA recognition by the EcoRV restriction endonuclease. *FEBS Lett.* **2011**, *585*, 2739-2743.

(73) Bock, C. W.; Kaufman, A.; Glusker, J. P. Coordination of water to magnesium cations. *Inorg. Chem.* **1994**, *33*, 419-427.

(74) Dudev, T.; Grauffel, C.; Lim, C. How Native and Alien Metal Cations Bind ATP: Implications for Lithium as a Therapeutic Agent. *Sci. Rep.* **2017**, *7*, 42377.

(75) Therrien, E.; Weill, N.; Tomberg, A.; Corbeil, C. R.; Lee, D.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 7. Impact of Protein Flexibility and Water Molecules on Docking-Based Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2014**, *54*, 3198-3210.