JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

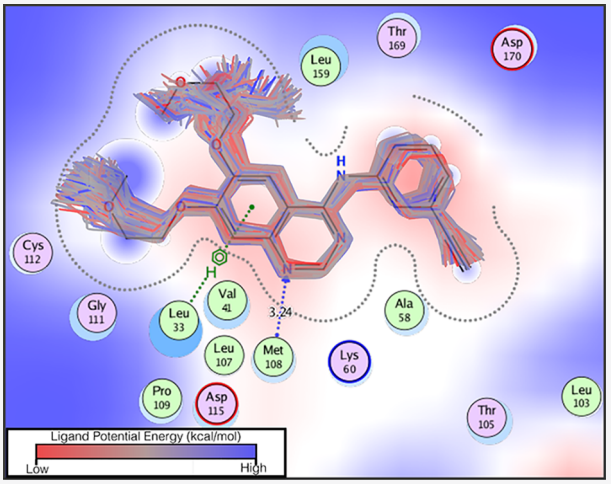
Flexible Fitting of Small Molecules into Electron Microscopy Maps Using Molecular Dynamics Simulations with Neural Network Potentials

John W. Vant, Shae-Lynn J. Lahey, Kalyanashis Jana, Mrinal Shekhar, Daipayan Sarkar, Barbara H. Munk, Ulrich Kleinekathöfer, Sumit Mittal, Christopher Rowley,* and Abhishek Singharoy*

Cite This: *J. Chem. Inf. Model.* 2020, 60, 2591−2604 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Despite significant advances in resolution, the potential for cryo-electron microscopy (EM) to be used in determining the structures of protein−drug complexes remains unrealized. Determination of accurate structures and coordination of bound ligands necessitates simultaneous fitting of the models into the density envelopes, exhaustive sampling of the ligand geometries, and, most importantly, concomitant rearrangements in the side chains to optimize the binding energy changes. In this article, we present a flexible-fitting pipeline where molecular dynamics flexible fitting (MDFF) is used to refine structures of protein−ligand complexes from 3 to 5 Å electron density data. Enhanced sampling is employed to explore the binding pocket rearrangements. To provide a model that can accurately describe the conformational dynamics of the chemically diverse set of small-molecule drugs inside MDFF, we use QM/MM and neural-network potential (NNP)/MM models of protein−ligand complexes, where the ligand is represented using the QM or NNP model, and the protein is represented using established molecular mechanical force fields (e.g., CHARMM). This pipeline offers structures commensurate to or better than recently submitted high-resolution cryo-EM or X-ray models, even when given medium to low-resolution data as input. The use of the NNPs makes the algorithm more robust to the choice of search models, offering a radius of convergence of 6.5 Å for ligand structure determination. The quality of the predicted structures was also judged by density functional theory calculations of ligand strain energy. This strain potential energy is found to systematically decrease with better fitting to density and improved ligand coordination, indicating correct binding interactions. A computationally inexpensive protocol for computing strain energy is reported as part of the model analysis protocol that monitors both the ligand fit as well as model quality.



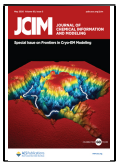
INTRODUCTION

Cryo-electron microscopy (EM) has emerged as one of the most powerful tools in structural biology, providing molecular models that approach the resolution commonly achieved in X-ray crystallography and NMR spectroscopy.[1] Some recent examples of high-resolution cryo-EM structures include GPCR-G protein complexes,[2] rabbit ryanodine receptor,[3] and ligand-bound ribosomal subunits,[4] all of which are resolved at near-3 Å resolution. Furthermore, cryo-EM based structure determination overcomes two key limitations faced in traditional X-ray crystallography, namely, the arduous task of preparing well-ordered crystals of macromolecules,[5] and the more fundamental problem with capturing these molecules in unphysiologically relevant states as a result of crystal contacts. Thus, Cryo-EM allows structure determination of large and/or dynamic macromolecular assemblies in native-like environments.

In view of the methodological advantages and milestone discoveries made by leveraging cryo-EM, it holds strong potential for monitoring the interactions between proteins and RNA with small molecules.[1−6] However, data processing from cryo-EM is slow—EM would require about half a year of data collection (assuming 8 h per data set) and at least a year of computation and model building. In contrast, once the arduous crystallization protocol is established, the postprocessing of the X-ray diffraction takes little time. Around 2000 crystallization

Special Issue: Frontiers in Cryo-EM Modeling

Received: December 17, 2019
Published: March 24, 2020



https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

conditions can be set up in 1 h, with a further 2−3 h needed to evaluate all the images for identifying suitable crystallization conditions (which can be reduced with image recognition software).[6] Thus, following the sample preparation stage, cryo-EM is not nearly as high-throughput as X-ray crystallography.

The second drawback of cryo-EM stems from the popular choices of systems relevant to drug discovery. Many drug targets are small proteins or their complexes, but cryo-EM is mostly employed to determine structures of larger biomolecules with molecular mass >500 kDa. The poor contrast transfer function seen in defocused single-particle images makes analysis of smaller complexes challenging.[7] Finally, there are issues of inhomogeneous local resolution,[8] radiation damage of acidic side chains[9] and inherent disorder of large macromolecular complexes, which makes *ab initio* determination of accurate protein−ligand interactions intractable within the routinely available 3−5 Å EM models.[6,10] Therefore, applications of cryo-EM in the mainstream of drug discovery is still limited. New methods need to be developed to transform this structural biology paradigm into a physical chemistry tool for drug discovery, drawing parallels to a journey that NMR has successfully undertaken *albeit* for small proteins.

One approach to overcome the protein-size limitation of cryo-EM is to form a rigid complex with the antigen-binding fragment. This strategy improves the quality of the data collected by increasing the size of the complex, while also facilitating alignment of the single-particle images. Combined with the so-called energy filtering methods,[11] the overall performance of cryo-EM is being gradually improved both in terms of size and resolution to suit drug discovery applications. Development of 4D cryo-electron microscopy by integrating the fourth dimension, time, into this powerful technique is also posited to offer information on protein dynamics.[12]

A positive impact of the aforementioned advances in cryo-EM instrumentation is seen on the completeness of the reported models, and growth in the automated structure building protocols in EMBD competitions.[13] Contingent on data resolution, information-driven modeling tools trace large pieces of the protein backbone and several side chains in minutes of computer time.[14−16] However, resolution of the binding pocket remains a stiff challenge as, unlike standard protein structure prediction, most of these tools are not trained *a priori* on the physics of protein-small molecule interactions. For low-resolution data, particularly in transmembrane systems, such structure determination of protein−ligand complexes is even more difficult,[17] given the extensive sampling of the structures that needs to be undertaken for finding the correct protein−ligand conformation.

Molecular modeling of the ligand, guided by the density of the protein, offers a concrete solution to resolving the drug−protein problems in cryo-EM. Already performed with high-resolution X-ray data, quantum mechanics/molecular mechanics (QM/MM) methods have resolved several protein−ligand complexes.[18] Key to the success of QM/MM, however, is the initial docking of the ligand to the proteins. To this end, a molecular docking protocol can be employed.[19] Very recently, a hybrid pipeline composed of GLIDE-docking (built within the Schrödinger package),[20] QM/MM geometry optimizations using the Gaussian QM package[21] and real-space refinement with the structure determination toolkit, PHENIX,[22] was proposed.[23] For the midresolution models derived from cryo-EM, the docking predictions e.g. from GLIDE, AUTODOCK, or ROSETTA-DOCK are often fraught with false-positives born out of multiple local minima within which the ligands can fit.[24] Therefore, an enhanced sampling tool needs to be coupled to the docking and the QM/MM protocol, so the false-positives can be avoided. Monte Carlo (MC) methods have been employed together with QM/MM protocols and classical force fields to perform enhanced sampling and resolve ion-binding to channels.[25] Despite these plausible MC remedies, QM/MM or the development of new force fields for novel drugs is prohibitively time-consuming for biological systems.[26] In the past, we combined classical CHARMM force fields on NAMD and density functional theory (DFT) on Gaussian iteratively with the ELBOW module of PHENIX to determine the structure of organic macrocycle-ligand complexes from 1 Å-resolved X-ray diffraction data.[27] The investigation of lower-resolution data for biological systems, however, requires a much larger number of such iterative re-refinements, slowing the model-building step. The deployment of a single parallelizable refinement platform that combines both quantum and classical descriptions of proteins with molecular docking and flexible fitting of ligands is therefore needed to address the iterative structure determination of the protein−ligand systems.

In this article, we use electron density data with resolutions between 3−5 Å to determine the structure of protein−ligand complexes by combining the GLIDE-docking protocol with the popular molecular dynamics (MD)-based real-space refinement tool molecular dynamics flexible fitting (MDFF).[8] We demonstrate this method on protein−ligand complexes of horse liver alcohol dehydrogenase, EGFR tyrosine kinase, and the kinase domain of the insulin receptor. The initial ligand-docked model derived from GLIDE is refined either with a hybrid QM/MM-MD platform[28] or by combining the traditional CHARMM force fields of the proteins with the neural network potentials (NNPs)[29,30] to describe the intramolecular forces of the ligands, all in the presence of an additional biasing force that conforms this refinement to the EM density data. The outcome is a computational protocol, implemented in the MD platform NAMD,[31] that docks and optimizes the geometry of the ligand within the protein, while simultaneously fitting to the density. The uncertainties of lower-resolution data are addressed using the built-in enhanced sampling capabilities of MD that are available on NAMD.[32]

A key advantage of using an NNP to represent the intramolecular interactions of the ligand stems from avoiding the reparameterization of a new force field for approximating these interactions. This is particularly significant for novel ligands, where a well-validated force field is not available and where the chosen force field does not define optimal parameters for all torsional barriers. The efficacy of the NNP is validated here by comparison with the QM/MM refinements on the same systems. Furthermore, discrepancies in the GLIDE-docked structures are removed by the use of either traditional or accelerated MD to equilibrate the structures; these accelerations are derived using the collective variable or COLVAR module of NAMD.[32] Taken together, for the first time, we combine GLIDE docking with refinement using QM/MM or NNP/MM models of the protein−ligand interactions and with MDFF refinements of cryo-EM data on the protein−ligand complexes accelerated by COLVAR. The entire refinement protocol is available for free on the latest builds of NAMD. As indicators of model quality, we monitor the (i) local geometry in the vicinity of the protein−ligand complex, (ii) consistency of the ligand coordination states with those

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

derived from high-resolution X-ray or cryo-EM structures in the PDB, and (iii) the interplay between protein−ligand interaction and ligand strain potential energies. These indicators are found to be robust to the resolution of the data, yet sensitive to the quality of the model and that of the fitting. Thus, we can credibly determine protein−ligand models by respecting the high-level molecular physics at play, while being concomitantly constrained by the information extractable from the data at hand.

METHODS

In what follows, we describe the individual components of our hybrid MD-based structure determination workflow, namely (i) MDFF and converting experimental data to synthetic potentials within MD; (ii) the QM/MM implementation with associated spatial embedding and COLVAR schemes available on NAMD,[31] which provided the basic framework for the refinement of ligand structures from cryo-EM; (iii) the scheme for hybridizing the NNP force fields for ligand with CHARMM force fields for proteins; (iv) the docking scheme by employing Schrodinger's GLIDE; and (v) computations of ligand coordination and strain energies for analysis of model quality. Finally, we integrate these five individual methodologies into one protocol for the structure determination and refinement of ligand geometries embedded in protein pockets.

Molecular Dynamics Flexible Fitting. MDFF works by supplementing an MD force field ($V_{\text{force-field}}$) with an electrostatic-like potential derived from the cryo-EM density (V_{EM}).[8] The $V_{\text{force-field}}$ can be split into three terms describing the potential energies of the protein (V_{protein}), that of the ligand (V_{ligand}) and the protein−ligand interactions ($V_{\text{protein-ligand}}$). The V_{EM} biases MD simulations toward structures that are consistent with the cryo-EM electron density maps. Structural models are refined against the EM density by determining atomic positions that minimize the weighted sum of $V_{\text{force-field}}$ and V_{EM}. A detailed explanation of the terms in the MDFF potential can be found in Supplementary Note 1.

There are a number of force fields for calculating V_{protein}, notably CHARMM, AMBER, and OPLS. However, constructing the potential energy functions V_{ligand} and $V_{\text{protein-ligand}}$ is challenging due to the enormous chemical diversity of known ligands. This process can require a tedious parametrization of the ligand force field based on experimental or quantum chemical data. Several force field construction protocols have been developed to capture the ligand interactions.[26,33,34] Yet, even the most elaborate versions of these force fields cannot provide universally accurate models for the structure and relative stability of the conformations of all possible ligands. A truly general strategy for resolving the structures of cryo-EM protein−ligand complexes requires an accurate method for calculating V_{ligand} for all possible ligands. Here, we explore two alternatives to force fields for calculating V_{ligand} and $V_{\text{protein-ligand}}$ within MDFF simulations: employing quantum mechanics and neural network potentials.

Data Integration. In this subsection, we detail the method to generate low-resolution synthetic densities from experimentally determined high-resolution densities for the three macromolecules (PDB: 4HJO, 3ETA, and 6NBB). X-ray crystallographic structure and intensity files of the inactive conformation of the EGFR tyrosine kinase domain bound with chemotherapy drug erlotinib (4HJO, resolution: 2.75 Å)[35] and of the kinase domain of an insulin receptor with a pyrrolo pyridine inhibitor (3ETA, resolution: 2.6 Å)[36] were downloaded from the Protein Databank for MDFF refinements. The cryo-EM structure reported by Herzik et al. was also used in the MDFF simulations of nicotinamide adenine dinucleotide-bound horse liver alcohol dehydrogenase (6NBB, resolution: 2.9 Å).[17]

For the high-resolution experimental densities determined from X-ray crystallography, we truncated the diffraction data at 3 and 5 Å using the *phenix.maps* tool from the Phenix software suite.[37] After truncating the high-resolution data to 5 Å, the data were smoothed using a B factor of 35 Å^2, in a fashion similar to the technique used in ref 38. This smoothing procedure further reduces the signal-to-noise ratio of the 5 Å resolution diffraction data, posing perhaps a more realistic refinement scenario. To generate synthetic densities of lower resolution from a high-resolution cryo-EM density, the experimental map is blurred by adding Gaussian functions with increasing half-widths, σ, using the *volutils* plugin of VMD. For σ = 0 Å, high-resolution experimental density is recovered, while for values of σ > 0 Å, synthetic densities of lower resolution are achieved.[8] After these density files are created in a .ccp4 format, they are projected on a 3D grid space in the .dx format using the *MDFF* plugin in VMD. Using Supplementary eq 3, this density is converted into potential V_{EM}, for driving MDFF.

NAMD QM/MM. The QM/MM interface of NAMD allows partitioning of a system into quantum and classical levels of description. The ligand is described quantum chemically, while the protein and the solvent are probed classically. The energies from the protein are computed using molecular mechanics force fields, such as CHARMM. The ligand energies can be calculated with an external QM software, such as MOPAC and ORCA, as well as a number of other programs amendable to the NAMD interface.[28] Forces are passed from the QM program into NAMD, which are then integrated using r-RESPA[31] multiple time step scheme to evolve the system. More details on NAMD's QM/MM interface are provided in ref 28 and Supplementary Note 2.

Normally, higher levels of electronic structure theory are prescribed for accurate QM/MM single point energy computations.[39] However, the semiempirical level theories (PM6 or AM1) are found applicable for geometry optimization, as evidenced by a number of protein−ligand systems involving atoms from main-group elements.[40] The constraints imposed by the experimental data during QM/MM-MDFF also prevents the ligand geometry from deviating toward unphysical structures, and yet, resolving negatively charged ligands or ones coordinated with transition metal ions remains a seminal challenge in cryo-EM.[13] These computations take ∼1.3 s of wall-clock time to perform MD steps of 1 fs, allowing MDFF to sample for ∼65 ps per day on a 2-CPU node in a typical system with 60−70 QM atoms and 500 000 MM atoms. For proteins alcohol dehydrogenase, EGFR kinase, and insulin receptor, the QM and MM regions are demarcated in Table S1 of the Supporting Information. The choice of the partitioning scheme follows from high-level QM/MM studies of these proteins by Zhu et al.[39] and independent simulations by Lahey et al.[30] Hydrogen was employed as the link-atom connecting the QM and MM regions, and an electrostatic embedding is employed. A NAMD input file for running QM/MM-MDFF is provided with the Supporting Information.

NAMD COLVAR. The CoordNum collective variable defines a coordination number (or the number of contacts),

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

represented by C.[31,32] It is calculated using the following expression

$$C(\text{group}_1, \text{group}_2) = \sum_{i \in \text{group}_1} \sum_{j \in \text{group}_2} \frac{1 - (|\mathbf{x}_i - \mathbf{x}_j|/d_o)^n}{1 - (|\mathbf{x}_i - \mathbf{x}_j|/d_o)^m} \quad (1)$$

Here, we have considered the ligand as one group (group_1) and the key residues in the binding pocket of the protein as another group (group_2). Table S2 presents the details of the ligands and the protein key residues used in defining groups 1 and 2 to compute the CoordNum (C) for each of the three complexes.

Neural Network Potentials Embedded in MM Models. Recently, neural-network potentials (NNPs) have emerged as an alternative to quantum mechanical calculations of molecular interactions.[29,30] NNPs like the ANI-1ccX are trained to reproduce the QM-calculated molecular energies of a large ensemble of molecules in a range of conformations. These potentials are almost as accurate as the high-level *ab initio* calculations they are trained to reproduce but are far less computationally expensive. This allows nanosecond length simulations to be performed readily. The ANI-1ccX NNP was trained to reproduce highly accurate CCSD(T*)/CBS calculations, so it avoids some of the limitations of low-cost semiempirical quantum methods that are commonly used in QM/MM MD. In MDFF simulations of protein−ligand complexes, NNPs can be used to represent the ligand embedded within a MM model for the protein. The protein−ligand interactions are calculated by pairwise additive Lennard-Jones and electrostatic potentials, as in a mechanically embedded QM/MM model.

$$V_{\text{protein-ligand}} = \sum_{i}^{\text{protein}} \sum_{j}^{\text{ligand}} \frac{C q_i q_j}{r_{ij}} + 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (2)$$

In this work, the ligand charges and Lennard-Jones parameters from CGenFF are used without modification for the calculation of the protein−ligand interactions.

The TorchANI implementation of the ANI-1ccX was interfaced with NAMD through its Custom QM/MM interface. Interactions between the ligand and the protein/solvent were calculated by conventional QM/MM MD interactions defined by the CGenFF model. We note that the ANI-1ccX model was not trained to model charged systems, so it is not designed to model a complex like 3ETA, which has an ammonium group. Nevertheless, we found that this potential provides reasonable descriptions of the intramolecular interactions of the ligand in the binding site. The NNP performance in NAMD is 10-folds faster than QM/MM MD.

GLIDE Docking Procedure. *Ligand Preparation.* The first step in the docking protocol entails ligand preparation and conformer generation. All ligands used in the study were generated using Maestro in the sdf format. Thereafter, the LigPrep utility of the Schrodinger molecular modeling toolkit was employed to generate 3D conformers. The 3D conformers were optimized utilizing the OPLS3e force field (default settings). Furthermore, Ligprep was employed to assign correct bond order with correct chirality along with ionization states and tautomers.

Protein Preparation. The protein structures for ligand docking were prepared and optimized using the Maestro molecular modeling software. The protonation state of the ionizable residues: ASP, GLU, ARG, and LYS were assigned based on the pK_a calculation using PROPKA. Crystal structure water molecules were removed except close (5.0 Å) to the binding site. Thereafter, restrained minimization was performed on the protein−ligand system using the impact refinement module and the OPLS3e force field. This allows for the relaxation of steric clashes that may have been present in the deposited PDB structure. The minimization protocol was terminated upon energy convergence or when the root-mean-square deviation (RMSD) between the minimization steps reached a tolerance of 0.30 Å.

Generation of Grid and Docking Protocol. The receptor grid was generated by defining the grids centered on the bound ligands in the crystal structure employing default settings for the box size in Glide. The van der Waals radii of the protein atoms with atomic charges ≤0.25 were scaled by 1.0. On the other hand for the ligand atoms, the van der Waals radii of the atoms with atomic charge ≤0.15 were scaled by 0.8. Docking was performed in the unconstrained mode while allowing for 5- and 6-membered ring flips and penalizing for nonplanar amide bonds. The van der Waals radii of ligand atoms with a partial atomic charge less than 0.15 were scaled by 0.8. All the docking used in this work was performed using GLIDE in the "extra precision" XP mode.[20] Subsequently, the docked ligand poses were minimized and rescored using the Glide (Gscore) scoring scheme. The Glide score (Gscore) is defined as

$$\text{Gscore} = a\text{*vdW} + b\text{*Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{BuryP} + \text{RotB} + \text{Site} \quad (3)$$

Here vdW represents the van der Waals energy, Coul = Coulomb energy, Lipo = lipophilic contact term, Hbond = hydrogen-bonding term, Metal = metal-binding term, BuryP = penalty for buried polar groups, RotB = penalty for freezing rotatable bonds, Site = polar interactions at the active site, and the coefficients of vdW and Coul are a = 0.065 and b = 0.130.[20]

Strain Energy. The internal conformational energy of bound ligands changes because the protein interactions allow a conformation that may be higher in energy than the conformation of the ligand in solution. This extra energy of the bound ligands is termed strain potential energy. To this end, following,[18] we used the MOPAC NAMD-QM/MM optimized geometries of the ligand in the protein−ligand complex and that of the ligand in water and performed high-level quantum chemistry calculations. Illustrated in Figure S1 of the Supporting Information, these protein-bound and solvated ligand geometries represent local energy minima in the PM6/CHARMM computations. The energy difference between the ligand conformations in the protein complex vs in the water yields the strain potential energy

$$\Delta E = E_{(\text{ligand conformation in complex})} - E_{(\text{ligand conformation in water})} \quad (4)$$

All single-point calculations were performed using the density functional theory (DFT) employing the Minnesota 06 functional (M06)[41] in conjunction with the 6-311+G(d,p)[42,46] basis set. The calculations denoted $\text{SMD}_{\text{water}}$M06/6-311+G(d,p) in the following were performed using the SMD[43] solvation model, i.e., a self-consistent reaction field method (ε

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

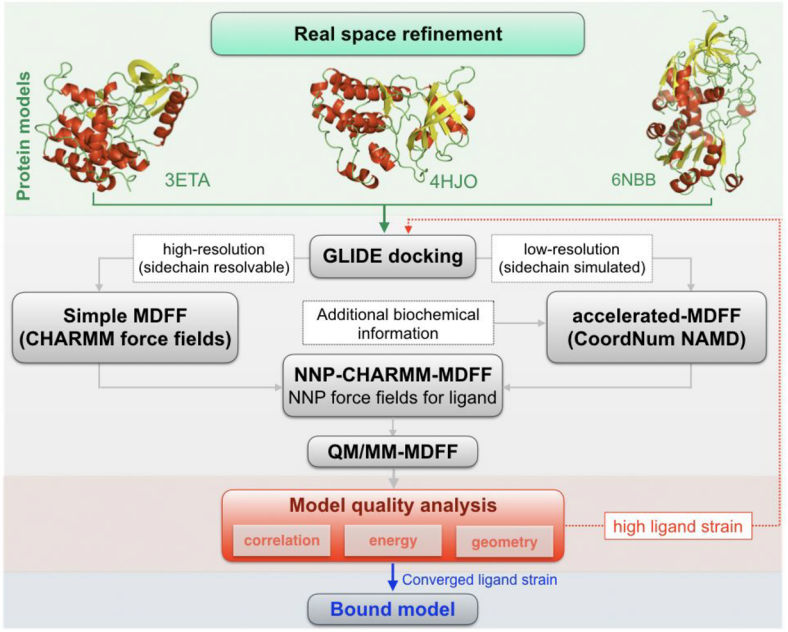


Figure 1. Workflow for molecular docking and refinement of protein−ligand complex using hybrid algorithms, integrating QM/MM, NNPs, and MDFF. The initial step is a real-space refinement of protein structure performed using popular *ab initio* structure determination tools (Phenix or Rosetta). Second, rigid-body docking using GLIDE (Schrödinger) is performed to get a starting model for the protein−ligand complex. Next, based on EM map resolution, high (<3 Å) and low (>3 Å), either simple MDFF or accelerated MDFF using CoordNum is adopted as the desired refinement protocol. For low-resolution EM maps, additional biochemical information may be included to improve the quality of side chain refinement. Third, further refinement of the structure is performed using MDFF with CHARMM force field for the protein and a neural network trained potential (NNP) for the ligand. The final step of the refinement involves another round of refinement of the NNP-derived model by QM/MM-MDFF. Next, the analysis stage includes calculating cross-correlation, strain energy and geometric parameters (MolProbity statistics), which helps determine both the quality of the fit and the quality of the bound model.

= 78.4) in the Gaussian 16 software package.[21] The present combination of DFT functional and basis set has resulted in quite reasonable results compared to MP2 calculations in an earlier calculation of the strain energy.[18] Moreover, we have performed strain potential energy computations using the MMFF94s force fields with the same geometries as employed in the DFT studies. Subsequently, the strain potential energy trends are compared between these two levels of theory.[18,41−43]

Hybrid Workflow. Illustrated in Figure 1, the first step of the protocol is to construct a preliminary protein model using either *ab initio* methods or rigid and flexible fitting.[44,45] This step can be achieved using real-space refinement tools such as Phenix or Rosetta,[46,47] manual building on Coot,[48] or a combination of backbone tracing and flexible fitting of side chains.[14] Direct fitting of available homology models to density maps has also been successful.[49] CHARMM36 force fields have been the most successful for MDFF refinements. Second, when the density has a high resolution of 2−4 Å, the key side chains are resolvable. Subsequently, this model can be subjected to GLIDE, followed by NNP-MDFF and/or a final QM/MM refinement step. NNP-MDFF can be performed both in implicit and explicit solvents with comparable accuracy.[50] When NNP force fields are unavailable, QM/MM-MDFF will be performed directly. Modeling of the solvent environment explicitly is essential to maintaining adequate solvation of the ligand-binding pocket during QM calculations and provide enhanced sampling through thermal fluctuations.

The radius of convergence of MDFF is defined in terms of the RMSD of the final refined model relative to the initial search model. When the density has a low resolution or the local resolution in the vicinity of the binding pocket is larger than 4 Å, NNP-MDFF is more useful as it has a higher radius of convergence than QM/MM-MDFF (see Results). To avoid the sampling of unphysical local minima, the CoordNum collective variable of NAMD is employed to improve the coordination between the protein and ligand in increments of 10%. When available, prior biochemical knowledge, e.g., from mutational assays on key residues engaging in protein−ligand interactions can also be introduced at this stage. This fitting step will continue until the local correlation coefficient of the binding pocket change converges. Normally it takes around 200 000−1 000 000 MDFF steps (200−1000 ps of simulation time) for the ligand correlation and coordination measures to converge.

In the third step, the low-energy structures are isolated and ligand strain energy is computed. Molecular mechanics methods are sufficiently accurate to offer a qualitative trend over a large ensemble.[51] For an accurate estimation of strain energy, higher level single point computations are recommended. Noting that lower strain correlates with higher binding affinity, the model with the lowest strain will be reported. For cross-validation, the ligand will be removed from this refined model to create an apo structure. Then, GLIDE will be used to reintroduce the ligand into the apo structure. If the GLIDE-docked apo structure and the MDFF prediction are commensurate in terms of model quality (to be determined using MolProbity[52]) within a chosen tolerance, the hybrid protocol will be deemed complete. In case these models are different, the next round of MDFF will begin with the new GLIDE-docked apo structure.

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

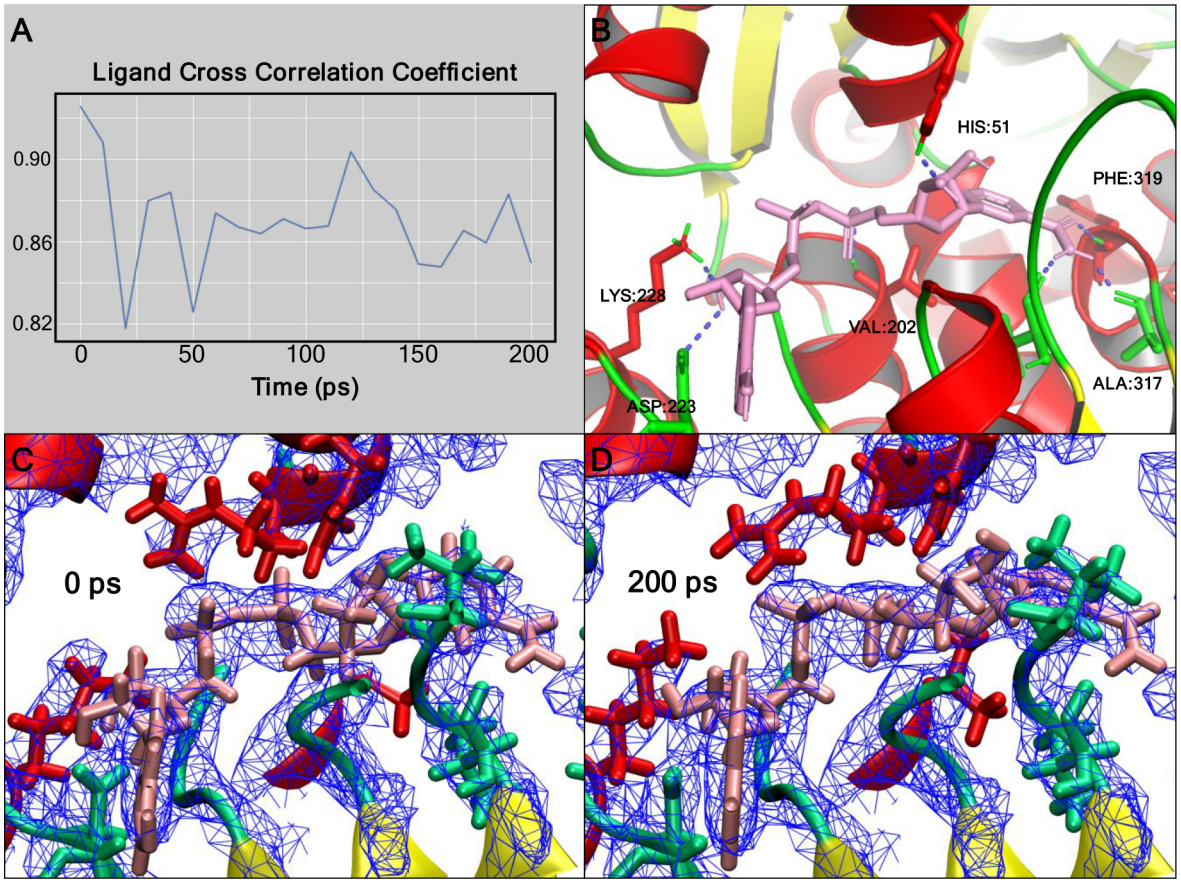


Figure 2. QM/MM-MDFF refinement of horse liver alcohol dehydrogenase. (A) Local cross-correlation (CC) changes of the NAD ligand during 200 ps of MDFF with the 2.9 Å cryo-EM density (EMD0406), where the ligand is described using the PM6 level of theory and the protein by CHARMM force fields. Starting with a GLIDE-docked model of the apo−6NBB structure, the CC stabilized around 0.86. (B) Key residues of the binding pocket from the refined model are labeled, together with ligand-stabilizing side chain contacts. This conformation is comparable to the one in the crystal (compare Figures 3C and S3−6NBB). Fitted atomic models of the NAD-binding pocket at the beginning (C) and end (D) of the MDFF refinement, illustrating comparable conformations of the ligand (pink) and the side chains (red and green).

RESULTS

In the following, we will describe how the QM/MM and NNP description of ligands is achieved within MDFF to refine the three example protein−ligand complexes. This hybrid flexible fitting procedure is repeated across (i) synthetic density maps of multiple resolutions and (ii) GLIDE-docked initial ligand poses of varying confidence to monitor the accuracy of the structure determination. The QM/MM and NNP models of the ligands are compared in terms of their geometric quality and binding-pocket energetics. Finally, to enhance the radius of convergence of ligand refinement within MDFF, particularly for determining binding pocket structures from poorly resolved side chains and small-molecule conformations, we resort to combining known biochemical information with NNP-MDFF via the enhanced sampling and COLVAR module of NAMD.

QM/MM-Modified MDFF has a Low Radius of Convergence. The PM6/CHARMM interface was employed with MDFF to study the structure of nicotinamide adenine dinucleotide (NAD)-bound horse liver alcohol dehydrogenase (6NBB) in explicit solvent. Two different search models were prepared to test the radius of convergence of this MDFF protocol. The first search model is the best GLIDE-docked model, where the NAD ligand is docked into the apo form of 6NBB. The second model is a GLIDE-docked conformation of an MD-modified apo 6NBB (see Methods). During this MD simulation, the side chains of the binding pocket are displaced by an RMSD of 5.5 Å using 300 K heating for 10 ns, while still maintaining the backbone in place. Effectively, this treatment produces an apo model with the backbone resolved but side chains uncertain, as is often the case with real low-resolution models. We examine the quality of GLIDE predictions when the binding-pocket side chains are thermally randomized and determine whether a subsequent MDFF with QM/MM refinement resurrects this binding conformation commensurate to the submitted PDB structure.

In the first QM/MM-MDFF, we refine the high-quality GLIDE docked dehydrogenase model. The NAD ligand has been reported in nine distinct binding poses in the 6NBB structure derived from cryo-EM at 2.9 Å resolution. The positional variance between these poses ranges from 0.4 to 0.7 Å (Figure S2A). Starting with the GLIDE-docked model and the experimental density, a similar diversity in conformation is captured by 200 ps of QM/MM MDFF, and even with a lower-resolution density blurred by a Gaussian function of half-width σ = 5 Å (Figure S2B−E). The ligand cross-correlation was found to be 0.86 and 0.90 at 2.9 and 5 Å, respectively (Figure 2). The local binding-pocket and global cross-correlation values converged to ∼0.20 and ∼0.88, respectively, at 2.9 Å and ∼0.21 and ∼0.81, respectively, at 5.0 Å (Figures 3 and S4). The ligand coordination measured in terms of the CoordNum variable (described in Methods) changed between

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

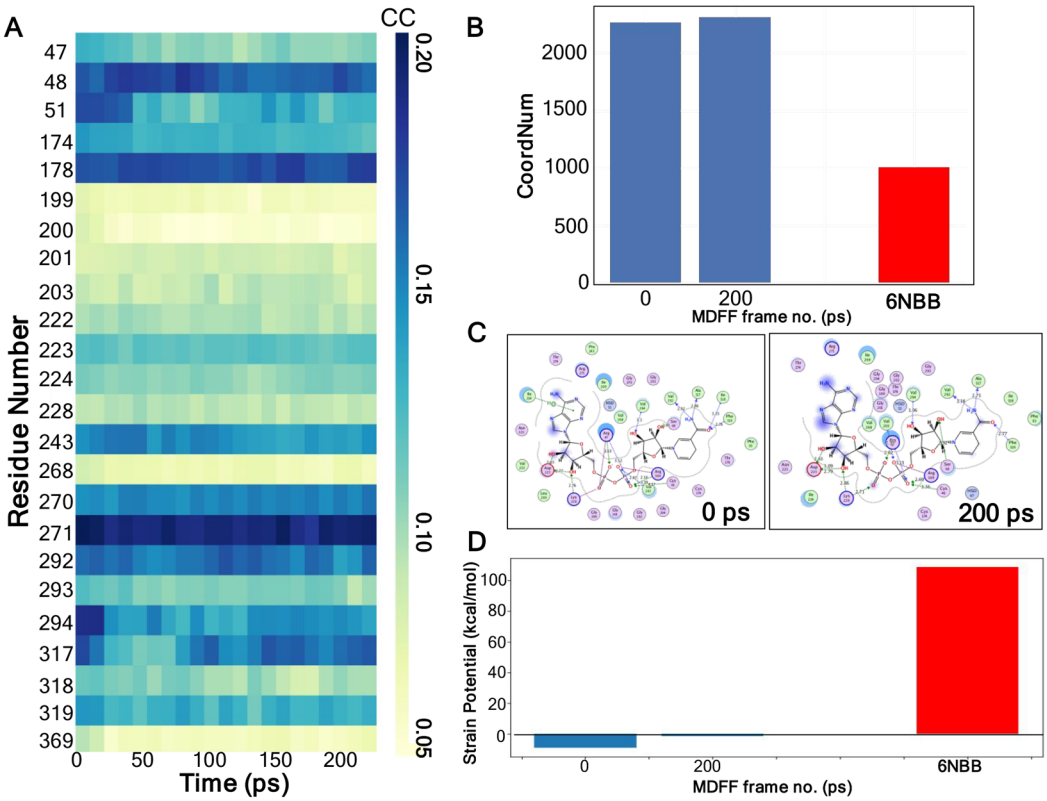


Figure 3. Model quality of the NAD binding pocket in horse liver alcohol dehydrogenase. (A) Per-residue cross-correlation coefficient of the NAD binding pocket side chains showing minimal changes across 200 ps of MDFF. (B) Binding coordination number, monitored in terms of the CoordNum collective variable,[53] improves by almost 2-fold over the submitted 6NBB model after GLIDE docking (t = 0 ps). These improved coordination interactions are maintained by PM6/CHARMM MDFF for 200 ps. (C) Details of the NAD binding pocket conformation (presented as a 2D projection or ligplot[54]) remain conserved across the MDFF refinement. (D) Consistent with the gain in ligand-coordinating interactions, strain potential energy on the NAD ligand is an order of magnitude lesser than the GLIDE-docked and MDFF-refined models than in the submitted 6NBB model.

2150 to 2350 and 2125 to 2220 at low resolution (Figure 3B). All the results are within the margin of error of 6NBB models reported from the 2.9 Å EM density—the ligand cross-correlation is 0.94, and the local and global correlations are approximately 0.2 and 0.91. Remarkably, the NAD coordination in this reported model is almost half of that determined here by PM6/CHARMM-MDFF, suggesting improvements in binding interactions derived from our QM treatment.

Despite strong model bias, the alcohol dehydrogenase example demonstrates that our protocol does not degrade the quality of the structure even at low resolution (Figures 3C, S3−6NBB, and S4). The overall MolProbity score is 0.82, an improvement over 1.90 from the reported 6NBB model (Table S3): Ramachandran and rotamer statistics are comparable (97% and 95% favored), while all the unphysical clashes are removed by the PM6/CHARMM level of protein−ligand description. This result confirms that the physical accuracy of the ligand's quantum chemical description supersedes biases from the uncertainty in EM density, providing an accurate model of the NAD coordination geometry (Figuress 2−3) and local interaction energy minimum (Figure S1A).

The PM6/CHARMM MDFF computations were repeated with GLIDE-docked NAD into the MD-modified apo alcohol dehydrogenase structure, our second search model. The system has a low global cross-correlation of 0.66 with respect to the 6NBB models, which improved to 0.80 (Figure S5) following the same protocol as with the first model (200 ps of flexible fitting with the 2.9 Å density determined structures). The RMSD and ligand cross-correlations improved from 5.5 Å and 0.30 relative to 6NBB to only 3.2 Å and 0.32 (Figure S6). Thus, consistent with the previous studies,[55] the quality of QM/MM refinement depends completely on the accuracy of the docked model, which in turn, is biased by the side chain assignments in the original density. With the first search model, PM6 maintained the correct binding pocket conformation, while with the second model starting from rearranged side chain conformations, the pocket was never recovered even from the 2.9 Å data. In effect, the introduction of QM/MM with MDFF has limited the radius of convergence to sub-1 Å RMSDs. This discrepancy will be addressed below with enhanced sampling simulations.

Strain potential energies of the ligand are computed to monitor how the quality of the model evolves across the 200 ps of the fitting procedure (Figure 3D). Fitting to the 2.9 Å EM density produces structures with a 10-fold decrease in the strain energy values with respect to the submitted 6NBB model (see Table S4 for details on the protein−ligand and water−ligand interactions derived using DFT). Consistent with the increase in CoordNum of Figure 3B, the lowering of the strain energy confirms improvement in the ligand-binding affinity.[56] Therefore, our PM6/CHARMM-MDFF starting with GLIDE-docked NAD into the 6NBB protein structure refines the ligand with comparable statistics as those reported in the Protein Data Bank, but now we achieve much stronger

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

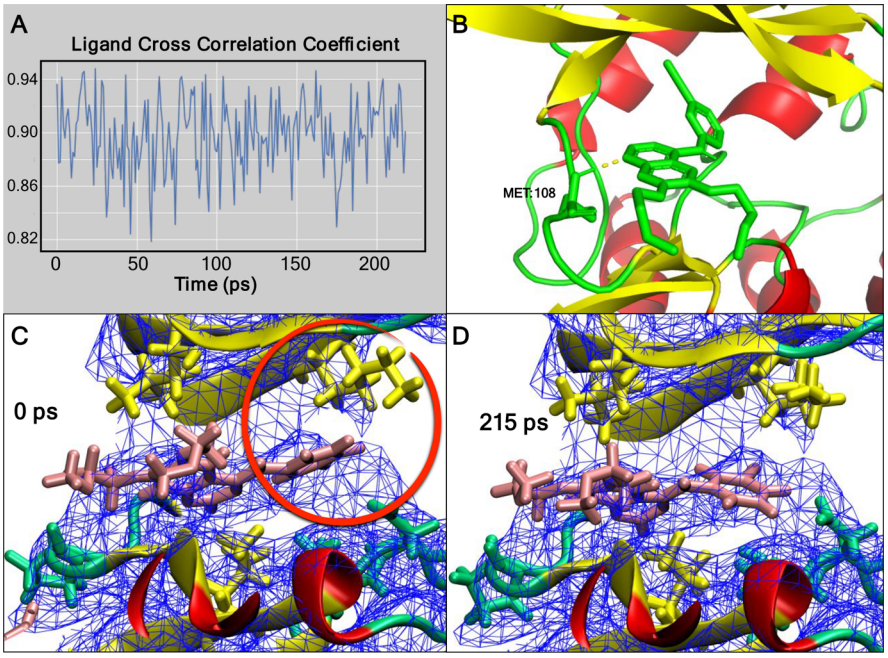


Figure 4. NNP-MDFF refinement of the erlotinib-bound EGFR tyrosine kinase domain. (A) Local CC changes of the erlotinib ligand during 215 ps of MDFF with synthetic X-ray density truncated at 4 Å, where the ligand is described using the NNP force field and the protein is described by CHARMM force fields. Starting with a GLIDE-docked model of the apo-4HJO structure, the CC stabilized around 0.90. (B) Key residues of the binding pocket from the refined kinase model are labeled, together with ligand-stabilizing side chain contacts. Fitted atomic models of the NAD-binding pocket at the beginning (C) and end (D) of the MDFF refinement, illustrating that the ligand conformation at t = 215 ps is more embedded into the density than at t = 0 ps (highlighted by the red circle).

protein−ligand interactions due to minor sub-1 Å changes brought forth by the QM/MM treatment of the pocket.

NNP Force Fields Improve the Radius of Convergence of Ligand Determination. The QM computations employed in the QM/MM-MDFF refinements of the previous section require simulations that may take a few days to complete. Additional computing resources can only improve this performance in a limited way because the Semiempirical QM (SQM) code used here cannot be efficiently parallelized beyond four nodes. Furthermore, the approximations inherent in the low-cost SQM methods that are amenable to MDFF can be inaccurate for some types of intramolecular interactions. NNPs have the potential to overcome the limitations of SQM because they are more computationally efficient, and they can also be more accurate due to their training to reproduce the results of high-level *ab initio* methods (e.g., CCSD(T*)/CBS). Following this assumption, we employ NNPs to describe the intramolecular interactions of the ligand instead of an SQM method. Because the NNPs are yet to be trained for phosphate moieties of NAD, a second example with a binding pocket structure that is comparable in complexity to alcohol dehydrogenase was chosen, namely the EGFR kinase binding to erlotinib.[35]

Similar to the QM/MM MDFF simulations, two of the GLIDE-docked structures were used as the starting points for NNP-MDFF simulations. The first structure had erlotinib docked to the apo form of the reported 4HJO model. Again, in the second structure, the side chains in the binding pockets were first rearranged using 10 ns of MD, and thereafter erlotinib was docked using GLIDE. Simulations were attempted using synthetic electron density maps truncated at a resolution of 4 and 5 Å resolution. With the 4 Å data, the bound-pose was recovered with both the starting structures after a maximum of 215 ps of MDFF in implicit solvent (Figure 4 and Table S5). The RMSD values of the refined models with respect to the crystallographic structures are ∼1.6 Å, starting with models that are deviated from 4HJO by 2.0 and 3.5 Å, respectively. The local cross-correlation of erlotinib is 0.90, comparable to 0.94 from the crystal structure (Figure S7A). Similarly, the strain energy of the refined erlotinib pose is 35.1 kcal/mol, 3.6 kcal/mol more stable than the initial pose, and comparable to that of 4HJO conformation (Figure S7D). Thus, the application of NNP enhances the radius of convergence of MDFF to 3.5 Å, at least 3-fold higher than the sub-1 Å radius of convergence achieved in QM-MM/MDFF.

Figure S8 compares the extent of conformational sampling achieved by the QM/MM-MDFF and NNP-MDFF methods, starting refinements from the 4HJO crystal structure. The NNP-MDFF modeled ligand explores 3-fold more conformational space than those from the QM/MM-MDFF refinements. The conformations sampled by the NNP method are also lower in potential energy. Taken together, the higher radius of convergence observed in the NNP-guided refinements is attributed to a more exhaustive sampling of the potential energy surface enabled by these force fields, which cannot be accessed by the SQM methods. The computational cost of this model is considerably lower than the QM/MM model. Even with the rudimentary NNP/MM interface used in the refinements, the simulations were completed at a rate of 0.5 ns/day using a single Titan Xp GPU, enabling structure determination with MDFF.

Simulations were also attempted where the resolution of the synthetic electron density was set to 5 Å instead of 4 Å. These simulations failed to reproduce the high-resolution crystallographic binding pose starting from the GLIDE-docked model

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

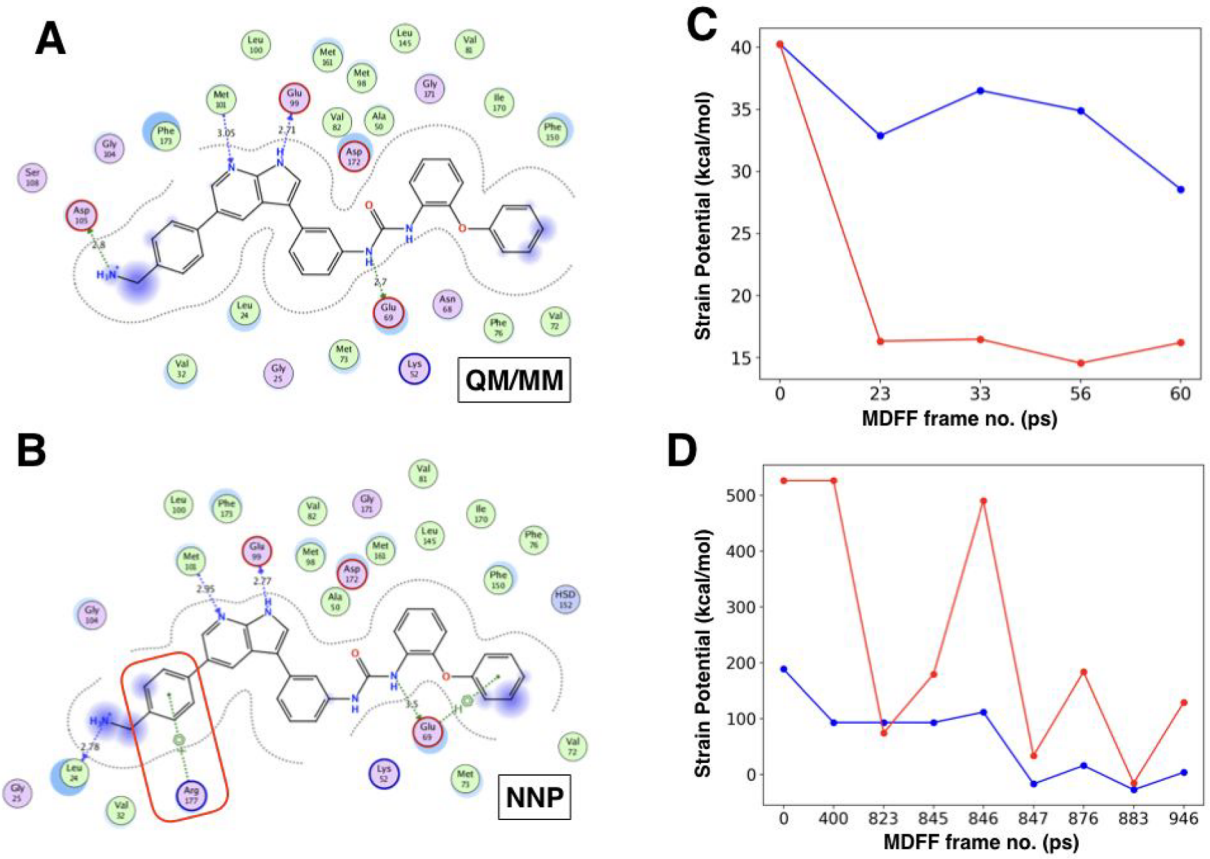


Figure 5. PM6 vs NNP refinement of pyrrolo pyridine inhibitor-bound insulin receptor. Ligplots of the pyrrolo pyridine inhibitor binding pocket in the 3ETA complex rerefined at 5 Å resolution using 60 ps of MDFF with PM6 level of theory (A) and NNP force field (B). The $\pi-\pi$ protein−inhibitor interaction captured by the NNP-driven refinements is highlighted in a red block. A similar interaction is present in the 2.9 Å-resolved structure reported in the PDB (Figure S3−4HJO). (C) Strain energies computed on the structures optimized by NNP (red) are almost half of that derived by PM6 (blue). These structures are selected from potential energy minima lying along the MDFF refinement path, similar to Figure S1. (D) Strain energies computed from DFT [SMD_{water}M06/6-311+G(d,p)—presented in blue] and MM [MMFF94s—presented in red] show commensurate trends.

of the rearranged pocket. At these coarser resolutions, the electron density associated with the protein backbone and the ligand becomes more ambiguous. So the simulation begins to adopt poses where the ligand overlaps with the electron density that corresponds to the protein. We will overcome this limitation in the subsequent example.

NNP Refinements Offer Higher-Quality Models than PM6. MDFF simulations of the first two examples clearly establish the computational challenges of resolving protein−ligand interactions from mid- to low-resolution electron density maps. Based on two separate examples we have established that the radius of convergence of NNP-driven refinements is higher than that of PM6/CHARMM-MDFF. For another comparison between PM6 and NNP inside MDFF, we perform both the refinements on a third system, namely the insulin receptor bound to a pyrrolo pyridine inhibitor (PDB: 3ETA) in explicit solvent. Unlike erlotinib in 4HJO that is charge neutral, this inhibitor ligand in the reported 3ETA model has a net −1 charge and is composed of 68−52 = 16 more atoms, making the insulin receptor a more challenging system for ligand refinement with MDFF.

We start with GLIDE docking of the high-resolution apo model. MDFF simulations were repeated at two different resolutions: the inherent resolution of the data set (2.6 Å), and the other truncated to 5 Å. Remarkably, strain energy of the ligands from the NNP-driven refinements is only 25−50% of that derived using PM6, both fitted for 60 ps with MDFF (Figure 5). This decrease in strain energy implies that NNP is better able to predict stable bound conformations of the ligand than the SQM methods.[57] In these conformations, stronger protein−ligand interactions can be obtained. For instance, additional $\pi-\pi$ stacking interactions are observed in the NNP models that were missing from the PM6 description (Figure 5A−B). These interactions are indeed present in the 2.6 Å-resolved 3ETA model (Figure S3−4HJO).[35] In addition, MolProbity statistics of the NNP results (0.61 overall score, 97% favored Ramachandran and 92% favored rotamer, presented in Table S6) are an improvement over the QM/MM results (0.95 overall score, 95% favored Ramachandran and 93% favored rotamer) and that in the reported X-ray model of 3ETA (1.34 overall score, 96% favored Ramachandran and 94% favored rotamer). Altogether, the NNP refinement of the inhibited insulin receptor structure offers better geometries and lower energies than the PM6 scheme.

Enhanced Sampling with NNPs Vastly Improves Ligand Predictions at 3−5 Å Resolution. The productive binding of small molecules to proteins ensues from specific side chain conformations. Consequently, the binding energy surface represents a funnel accommodating a large number of closely lying higher energy states, and a few, often unique,

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

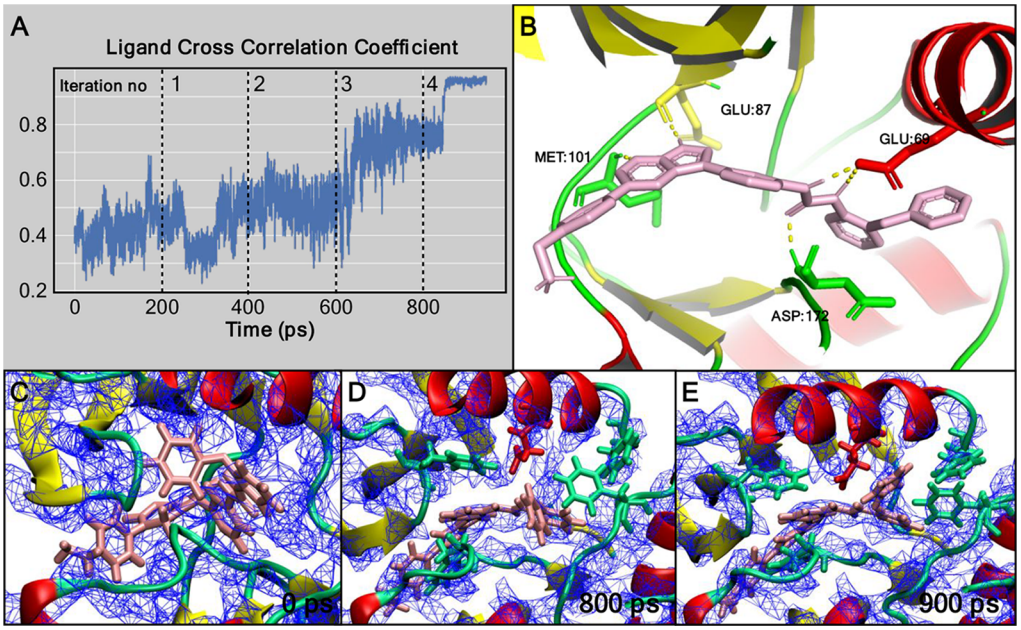


Figure 6. Enhanced NNP-MDFF refinement of pyrrolo pyridine inhibitor-bound insulin receptor. (A) Local CC changes of the pyrrolo pyridine inhibitor ligand during four iterative 200 ps steps of CoordNum driven NNP-MDFF refinements. In the last 200 ps, forces from the CoordNum bias are switched off. Starting with a GLIDE-docked structure of the annealed-3ETA model, the CC improved from 0.4 and stabilized around 0.9. (B) Key residues of the insulin receptor binding pocket from the refined model are labeled, together with ligand-stabilizing side chain contacts. Fitted atomic models of the binding pocket at the beginning (C) and end (D) of the MDFF with CoordNum refinement, as well as (E) the end of the last 200 ps with CoordNum switched off, illustrating that the inhibitor conformation at t = 800 ps is more akin to the density than at t = 0 ps. The convergence of the fitted model is illustrated by minimal changes in the cross-correlation of the inhibitor after 900 ps of MDFF.

lower energy states. Starting from a poorly docked model, one major limitation of finite-time MD stems from the entrapment of simulated structures in the higher energy states. This limitation precludes the sampling of the most relevant binding poses in MD.[56] Despite the ability of NNP force fields in recovering structures from initial models that are deviated by 3.5 Å (demonstrated for the 4HJO example), such results are applicable up to 4 Å data. At even lower resolutions, the assignment of side chains conflicts with those of the ligand heavy atoms, pushing the ligand into unphysical pockets away from the primary docking site. Here, we increase the radius of convergence of MDFF and recover ligand-binding to the correct site by imposing additional knowledge of the binding pocket through the CoordNum variable in the COLVAR module of NAMD.[32]

First, the 3ETA complex was heated at 800 K for 10 ns in explicit solvent employing a protocol described in ref 8, and then the system was cooled back to room temperature in another 10 ns. This simulated annealing procedure disengaged the inhibitor and shifted its center of mass by ∼3 Å from the primary pocket into a distinct coordination environment (Figures 6 and 7) and with ligand-RMSD of 6.5 Å relative to the 3ETA conformation. The use of NNP force fields inside MDFF marginally improved the binding, wherein the coordination to the correct side chains improved by 20%. Yet, the aromatic ring of the pyrrolo pyridine ligand was inversely oriented, reproducing a classic problem of MD binding aromatic rings in an inverted pose seen in case studies of benzene binding to lysozyme.[58,59]

From prior biochemical studies, residue MET 101 was known to be involved in binding the inhibitor.[60] To recover the correct binding-site during MDFF, we enhanced the contact between this methionine with the ligand heavy atoms using the CoordNum COLVAR (see Methods—eq 1). While performing this step, no specific part of the ligand was over weighted to define the contact with the particular residue. The methionine—ligand contact was raised by increments of 10% on the instantaneous values, while the ligand was simultaneously refitted into the 5 Å density for 200 ps. After four such incremental steps, the ligand coordination number converged. Remarkably, the inverted orientation of the ligand's aromatic ring is corrected over the cumulative time of 4 × 200 ps = 800 ps CoordNum-accelerated MDFF—an observation that normally takes microseconds of conventional MD.[59] Following this convergence, 200 ps more of simple NNP-MDFF was performed without any further bias from the known contact of the methionine residue. This simulation recovers the inhibitor-bound insulin receptor commensurate with the 2.6 Å-resolved 3ETA model. The strain energies rapidly decreased by 10-fold and the MolProbity score improved from 1.92 to 0.91 across the refinement (Figure 7 and Table S6).

Surprisingly, the strain energy trend derived using molecular mechanical force fields (MMFF94S) is comparable to that derived from higher-level DFT (Figure 5D and Table S7). Since the MMFF94S computations take only seconds of compute-time using the Avogadro software,[61] we suggest that qualitative strain energy trends from these MM calculations can be employed as an efficient measure for probing model quality. Taken together, with the enhanced sampling of protein—ligand contacts inside MDFF using NNP force fields, the correct binding site and pose are resolved at 5 Å refining, beginning from a completely different binding site and coordination. Therefore, this procedure increases the radius of convergence for the determination of the ligand structure to 6.5 Å. Also, the strain energy of the inhibitor derived from MDFF is lower than that of the 3ETA model. This result

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

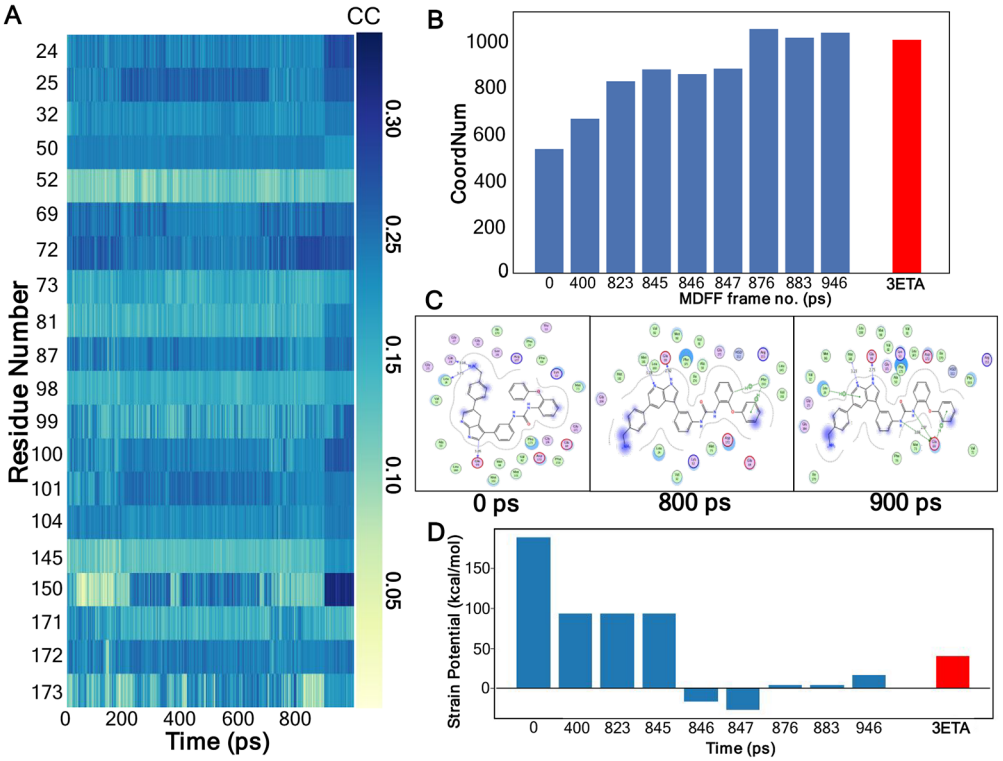


Figure 7. Model quality of the insulin receptor. (A) The per-residue cross-correlation coefficient of the insulin receptor binding pocket shows a significant rearrangement for residues involved in ligand coordination over 900 ps. (B) The CoordNum variable measured for nine structures from the 900 ps simulation as well as the crystal structure. Over the course of the iterative simulations, the CoordNum variable approximately converged to the value measured for the crystal structure. (C) Ligplots of pyrrolo pyridine inhibitor and surrounding residues. At 0 ps, the molecular environment of the pyridine inhibitor is shifted from its original docked position. At 800 ps, the pyridine inhibitor recovers key interactions with the residues originally strongly interacting in the 3ETA crystal structure, e.g. GLU 69, GLU 87, and ASP 172. The recovery of ligand interactions follows from the gain in MET 101−inhibitor coordination, which is imposed by using prior biochemical knowledge within the CoordNum collective variable and NNP-driven MDFF. All the protein−ligand interactions remain stable at 900 ps even after removing the additional CoordNum constraints. (D) Strain potential energies of the pyridine inhibitor decrease throughout the simulations and converge to a value commensurate with the crystal structure.

suggests that the NNP force field-derived models guided by prior biochemical knowledge capture the physics of protein−ligand interactions better than traditional real-space refinement or manual methods, even when the fitting is performed into low-resolution data.

DISCUSSION

In this article, we present a hybrid MDFF methodology for the refinement of protein-embedded ligand structure, geometry and coordination from electron density of 3−5 Å resolution. The quality of MDFF results heavily depends on the quality of the initial docked model, derived here using Schrödinger's GLIDE software. Lower-quality docked models offer poorly refined structures, even with high-resolution density data, particularly when the ligand is described with an SQM. This limitation is overcome by employing NNP force fields for the ligand in concert with traditional CHARMM force fields for the protein within MDFF. Finally, the use of prior biochemical information via the CoordNum collective variable enhances the sampling of the ligand geometry within the solvated protein environment. The deployment of such enhanced sampling schemes increases the radius of convergence of ligand structures to 6.5 Å.

Structures determined using the SQM and NNP models were validated with DFT calculations of strain potential energy. Besides having a 6-fold higher radius of convergence over SQM refinements, NNPs offer higher quality models of the ligands than PM6. MDFF simulations of the protein-pocket refined by the NNP/CHARMM interface converge on improved MolProbity scores relative to the use of PM6/CHARMM interface for structure determination, starting from the same initial model. The strain energies of ligands are systematically lower for the former, suggesting access to stronger binding interactions. The overall better quality of the NNP results reflects its very high fidelity to the high-level *ab initio* training data over the on-the-fly application of semi-empirical models.

The entire methodology, including the use of QM/MM, NNP, and CoordNum collective variables inside MDFF is scripted within NAMD and can be employed as a single structure-determination platform. Ligand determination from low-resolution data will further benefit from the built-in simulated annealing and resolution-exchange schemes in MDFF.[8] The docking and analysis tools form stand-alone segments of the pipeline. Surprisingly, the strain energy trends for monitoring model quality show a remarkable similarity between MM and DFT methods. In congruence with cross-correlation measures, MolProbity statistics, and ligand coordination values, these MM-based energy computations serve as a list of high-throughput model quality criteria.

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

The NNPs are trained using results from the coupled cluster level of theory. This training allows the NNP force fields to mimic a number of ground-state properties of small molecules that are captured by the high-level QM methods. In contrast to the higher-level theories, the use of SQM in QM/MM results in less accurate representation of ligand. Using a higher level of theory directly in QM/MM is expected to result in a more accurate description of the ligand. However, the QM/MM method, which is already slow at the SQM level of theory, will become even more computationally expensive for the use of any post-SCF QM method, making it intractable to resolve the conformationally diverse ligand ensembles. These practical differences between the NNP and QM approaches make the former a method of choice within MDFF. Nonetheless, with high-resolution X-ray data, close to 1 Å resolution, quantum chemical methods have been employed successfully for structure refinement.[18,27] With further improvements in resolution, which is definitely forthcoming in cryo-EM,[1−4] higher-level QM methods will certainly find more applications in structural biology.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b01167.

MolProbity tables, raw data for strain energy computations and example NAMD input files for MDFF simulations, and MDFF potential energy surfaces (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Christopher Rowley − *Department of Chemistry, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada;* orcid.org/0000-0002-0205-952X; Email: crowley@mun.ca

Abhishek Singharoy − *School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, United States;* orcid.org/0000-0002-9000-2397; Email: asginhar@asu.edu

Authors

John W. Vant − *School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, United States*

Shae-Lynn J. Lahey − *Department of Chemistry, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada*

Kalyanashis Jana − *Department of Physics and Earth Sciences, Jacobs University Bremen, 28759 Bremen, Germany*

Mrinal Shekhar − *School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, United States; Center for Development of Therapeutics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141-2023, United States*

Daipayan Sarkar − *School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, United States*

Barbara H. Munk − *School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, United States*

Ulrich Kleinekathöfer − *Department of Physics and Earth Sciences, Jacobs University Bremen, 28759 Bremen, Germany;* orcid.org/0000-0002-6114-7431

Sumit Mittal − *School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, United States; School of Advanced Sciences and Languages, VIT Bhopal University, Bhopal 466114, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.9b01167

Author Contributions

○These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.S. acknowledges National Science Foundation for a CAREER grant (MCB-1942763), start-up funds from the School of Molecular Sciences and Center for Applied Structure Discovery at Arizona State University, and the resources of the OLCF at the Oak Ridge National Laboratory, which is supported by the Office of Science at DOE under Contract No. DE-AC05-00OR22725, made available via the INCITE program. J.W.V. and S.M. acknowledge Research Computing at Arizona State University for providing HPC resources that have contributed to the research results reported within this paper. C.R. and S.-L.J.L. thank NSERC of Canada for funding through the Discovery Grants program (RGPIN-05795-2016). S.-L.J.L. thanks the School of Graduate Studies at Memorial University for a graduate fellowship. C.R.'s computational resources were provided by Compute Canada (RAPI: djk-615-ab). C.R. and S.-L.J.L. acknowledge the support of NVIDIA Corporation through the donation of the Titan Xp GPU used for this research. K.J. acknowledges the Alexander von Humboldt (AvH) foundation for an AvH postdoctoral research fellowship. We also acknowledge NAMD and VMD developments supported by NIH (P41GM104601) and R01GM098243-02 for supporting our study of membrane proteins.

■ REFERENCES

(1) Liao, M.; Cao, E.; Julius, D.; Cheng, Y. Structure of the TRPV1 Ion Channel Determined by Electron Cryo-Microscopy. *Nature* **2013**, *504* (7478), 107−112.

(2) Liang, Y. L.; Khoshouei, M.; Radjainia, M.; Zhang, Y.; Glukhova, A.; Tarrasch, J.; Thal, D. M.; Furness, S. G. B.; Christopoulos, G.; Coudrat, T.; Danev, R.; Baumeister, W.; Miller, L. J.; Christopoulos, A.; Kobilka, B. K.; Wootten, D.; Skiniotis, G.; Sexton, P. M. Phase-Plate Cryo-EM Structure of a Class B GPCR-G-Protein Complex. *Nature* **2017**, *546* (7656), 118−123.

(3) Yan, Z.; Bai, X. C.; Yan, C.; Wu, J.; Li, Z.; Xie, T.; Peng, W.; Yin, C. C.; Li, X.; Scheres, S. H. W.; Shi, Y.; Yan, N. Structure of the Rabbit Ryanodine Receptor RyR1 at Near-Atomic Resolution. *Nature* **2015**, *517* (7532), 50−55.

(4) Wong, W.; Bai, X. C.; Brown, A.; Fernandez, I. S.; Hanssen, E.; Condron, M.; Tan, Y. H.; Baum, J.; Scheres, S. H. W. Cryo-EM Structure of the Plasmodium Falciparum 80S Ribosome Bound to the Anti-Protozoan Drug Emetine. *eLife* **2014**, *3*, e03080.

(5) Chapman, H. N.; Fromme, P.; Barty, A.; White, T. A.; Kirian, R. A.; Aquila, A.; Hunter, M. S.; Schulz, J.; Deponte, D. P.; Weierstall, U.; Doak, R. B.; Maia, F. R. N. C.; Martin, A. V.; Schlichting, I.; Lomb, L.; Coppola, N.; Shoeman, R. L.; Epp, S. W.; Hartmann, R.; Rolles, D.; Rudenko, A.; Foucar, L.; Kimmel, N.; Weidenspointner, G.; Holl, P.; Liang, M.; Barthelmess, M.; Caleman, C.; Boutet, S.; Bogan, M. J.; Krzywinski, J.; Bostedt, C.; Bajt, S.; Gumprecht, L.; Rudek, B.; Erk, B.; Schmidt, C.; Hömke, A.; Reich, C.; Pietschner, D.; Ströder, L.; Hauser, G.; Gorke, H.; Ullrich, J.; Herrmann, S.; Schaller, G.; Schopper, F.; Soltau, H.; Kühnel, K. U.; Messerschmidt, M.; Bozek, J. D.; Hau-Riege, S. P.; Frank, M.; Hampton, C. Y.; Sierra, R. G.; Starodub, D.; Williams, G. J.; Hajdu, J.; Timneanu, N.; Seibert, M.

M.; Andreasson, J.; Rocker, A.; Jönsson, O.; Svenda, M.; Stern, S.; Nass, K.; Andritschke, R.; Schröter, C. D.; Krasniqi, F.; Bott, M.; Schmidt, K. E.; Wang, X.; Grotjohann, I.; Holton, J. M.; Barends, T. R. M.; Neutze, R.; Marchesini, S.; Fromme, R.; Schorb, S.; Rupp, D.; Adolph, M.; Gorkhover, T.; Andersson, I.; Hirsemann, H.; Potdevin, G.; Graafsma, H.; Nilsson, B.; Spence, J. C. H. Femtosecond X-Ray Protein Nanocrystallography. *Nature* **2011**, *470* (7332), 73−78.

(6) Ceska, T.; Chung, C. W.; Cooke, R.; Phillips, C.; Williams, P. A. Cryo-EM in Drug Discovery. *Biochem. Soc. Trans.* **2019**, *47*, 281−293.

(7) Glaeser, R. M.; Hall, R. J. Reaching the Information Limit in Cryo-EM of Biological Macromolecules: Experimental Aspects. *Biophys. J.* **2011**, *100*, 2331−2337.

(8) Singharoy, A.; Teo, I.; McGreevy, R.; Stone, J. E.; Zhao, J.; Schulten, K. Molecular Dynamics-Based Refinement and Validation for Sub-5 Å Cryo-Electron Microscopy Maps. *eLife* **2016**, *5*, e16105.

(9) Karuppasamy, M.; Karimi Nejadasl, F.; Vulovic, M.; Koster, A. J.; Ravelli, R. B. G. Radiation Damage in Single-Particle Cryo-Electron Microscopy: Effects of Dose and Dose Rate. *J. Synchrotron Radiat.* **2011**, *18* (3), 398−412.

(10) Renaud, J. P.; Chari, A.; Ciferri, C.; Liu, W. T.; Rémigy, H. W.; Stark, H.; Wiesmann, C. Cryo-EM in Drug Discovery: Achievements, Limitations and Prospects. *Nat. Rev. Drug Discovery* **2018**, *17*, 471−492.

(11) Thompson, R. F.; Walker, M.; Siebert, C. A.; Muench, S. P.; Ranson, N. A. An Introduction to Sample Preparation and Imaging by Cryo-Electron Microscopy for Structural Biology. *Methods* **2016**, *100*, 3−15.

(12) Fitzpatrick, A. W. P.; Lorenz, U. J.; Vanacore, G. M.; Zewail, A. H. 4D Cryo-Electron Microscopy of Proteins. *J. Am. Chem. Soc.* **2013**, *135* (51), 19123−19126.

(13) Lawson, C. L.; Chiu, W. Comparing Cryo-EM Structures. *J. Struct. Biol.* **2018**, *204* (3), 523−526.

(14) Terashi, G.; Kihara, D. De Novo Main-Chain Modeling for Em Maps Using MAINMAST. *Nat. Commun.* **2018**, *9* (1618), 1−11.

(15) Terwilliger, T. C.; Adams, P. D.; Afonine, P. V.; Sobolev, O. V. A Fully Automatic Method Yielding Initial Models from High-Resolution Cryo-Electron Microscopy Maps. *Nat. Methods* **2018**, *15* (11), 905−908.

(16) Moritz, S. A.; Pfab, J.; Wu, T.; Hou, J.; Cheng, J.; Cao, R.; Wang, L.; Si, D. Cascaded-CNN: Deep Learning to Predict Protein Backbone Structure from High-Resolution Cryo-EM Density Maps. *bioRxiv* **2019**, 572990.

(17) Herzik, M. A.; Wu, M.; Lander, G. C. High-Resolution Structure Determination of Sub-100 KDa Complexes Using Conventional Cryo-EM. *Nat. Commun.* **2019**, *10* (1032), 1−9.

(18) Fu, Z.; Li, X.; Merz, K. M. Accurate Assessment of the Strain Energy in a Protein-Bound Drug Using QM/MM X-Ray Refinement and Converged Quantum Chemistry. *J. Comput. Chem.* **2011**, *32* (12), 2587−2597.

(19) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational Protein-Ligand Docking and Virtual Drug Screening with the AutoDock Suite. *Nat. Protoc.* **2016**, *11* (5), 905−919.

(20) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177−6196.

(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratch, D. J. *Gaussian 16, Revision C.01*; Wallingford CT, 2016.

(22) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (2), 213−221.

(23) Robertson, M. J.; van Zundert, G. C. P.; Borrelli, K. W.; Skiniotis, G. GemSpot: A Pipeline for Robust Modeling of Ligands into CryoEM Maps. *bioRxiv* **2019**, 750778.

(24) Deng, N.; Forli, S.; He, P.; Perryman, A.; Wickstrom, L.; Vijayan, R. S. K.; Tiefenbrunn, T.; Stout, D.; Gallicchio, E.; Olson, A. J.; Levy, R. M. Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening against the Flap Site of HIV Protease. *J. Phys. Chem. B* **2015**, *119* (3), 976−988.

(25) Kotev, M.; Pascual, R.; Almansa, C.; Guallar, V.; Soliva, R. Pushing the Limits of Computational Structure-Based Drug Design with a Cryo-EM Structure: The Ca $^{2+}$ Channel $\alpha 2\delta$-1 Subunit as a Test Case. *J. Chem. Inf. Model.* **2018**, *58* (8), 1707−1715.

(26) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J. Comput. Chem.* **2013**, *34* (32), 2757−2770.

(27) Singharoy, A.; Venkatakrishnan, B.; Liu, Y.; Mayne, C. G.; Lee, S.; Chen, C. H.; Zlotnick, A.; Schulten, K.; Flood, A. H. Macromolecular Crystallography for Synthetic Abiological Molecules: Combining XMDFF and PHENIX for Structure Determination of Cyanostar Macrocycles. *J. Am. Chem. Soc.* **2015**, *137* (27), 8810−8818.

(28) Melo, M. C. R.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D. C.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z. NAMD Goes Quantum: An Integrative Suite for Hybrid Simulations. *Nat. Methods* **2018**, *15* (5), 351−354.

(29) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10* (1038), 1−8.

(30) Lahey, S.-L.; Rowley, C. Simulating Protein-Ligand Binding with Neural Network Potentials. *Chem. Sci.* **2020**, *11*, 2362−2368.

(31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.

(32) Fiorin, G.; Klein, M. L.; Hénin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Mol. Phys.* **2013**, *111* (22−23), 3345−3362.

(33) 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* (9), 1157−1174.

(34) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *31* (4), 671−690.

(35) Park, J. H.; Liu, Y.; Lemmon, M. A.; Radhakrishnan, R. Erlotinib Binds Both Inactive and Active Conformations of the EGFR Tyrosine Kinase Domain. *Biochem. J.* **2012**, *448* (3), 417−423.

(36) Patnaik, S.; Stevens, K. L.; Gerding, R.; Deanda, F.; Shotwell, J. B.; Tang, J.; Hamajima, T.; Nakamura, H.; Leesnitzer, M. A.; Hassell, A. M.; Shewchuck, L. M.; Kumar, R.; Lei, H.; Chamberlain, S. D. Discovery of 3,5-Disubstituted-1H-Pyrrolo[2,3-b] Pyridines as Potent Inhibitors of the Insulin-like Growth Factor-1 Receptor (IGF-1R) Tyrosine Kinase. *Bioorg. Med. Chem. Lett.* **2009**, *19* (11), 3136−3140.

(37) Pražnikar, J.; Afonine, P. V.; Gunčar, G.; Adams, P. D.; Turk, D. Averaged Kick Maps: Less Noise, More Signal and Probably Less Bias. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65* (9), 921−931.

(38) Schröder, G. F.; Levitt, M.; Brunger, A. T. Super-Resolution Biomolecular Crystallography with Low-Resolution Data. *Nature* **2010**, *464* (7292), 1218−1222.

(39) Zhu, K.; Day, T.; Warshaviak, D.; Murrett, C.; Friesner, R.; Pearlman, D. Antibody Structure Determination Using a Combina-

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604

tion of Homology Modeling, Energy-Based Refinement, and Loop Prediction. *Proteins: Struct., Funct., Genet.* **2014**, *82* (8), 1646−1655.
(40) Bikadi, Z.; Hazai, E. Application of the PM6 Semi-Empirical Method to Modeling Proteins Enhances Docking Accuracy of AutoDock. *J. Cheminf.* **2009**, *1* (15), 1−16.
(41) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120* (1−3), 215−241.
(42) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta* **1973**, *28* (3), 213−222.
(43) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378−6396.
(44) Wriggers, W. Conventions and Workflows for Using Situs. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68* (4), 344−351.
(45) Lindert, S.; McCammon, J. A. Improved CryoEM-Guided Iterative Molecular Dynamics-Rosetta Protein Structure Refinement Protocol for High Precision Protein Structure Prediction. *J. Chem. Theory Comput.* **2015**, *11* (3), 1337−1346.
(46) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkoczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **2019**, *75*, 861−877.
(47) Wang, R. Y. R.; Song, Y.; Barad, B. A.; Cheng, Y.; Fraser, J. S.; DiMaio, F. Automated Structure Refinement of Macromolecular Assemblies from Cryo-EM Maps Using Rosetta. *eLife* **2016**, *5*, e17219.
(48) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (4), 486−501.
(49) Goh, B. C.; Hadden, J. A.; Bernardi, R. C.; Singharoy, A.; McGreevy, R.; Rudack, T.; Cassidy, C. K.; Schulten, K. Computational Methodologies for Real-Space Structural Refinement of Large Macromolecular Complexes. *Annu. Rev. Biophys.* **2016**, *45* (1), 253−278.
(50) Qi, Y.; Lee, J.; Singharoy, A.; McGreevy, R.; Schulten, K.; Im, W. CHARMM-GUI MDFF/XMDFF Utilizer for Molecular Dynamics Flexible Fitting Simulations in Various Environments. *J. Phys. Chem. B* **2017**, *121* (15), 3718−3723.
(51) Singharoy, A.; Maffeo, C.; Delgado-Magnero, K. H.; Swainsbury, D. J. K.; Sener, M.; Kleinekathöfer, U.; Vant, J. W.; Nguyen, J.; Hitchcock, A.; Isralewitz, B.; Teo, I.; Chandler, D. E.; Stone, J. E.; Phillips, J. C.; Pogorelov, T. V.; Mallus, M. I.; Chipot, C.; Luthey-Schulten, Z.; Tieleman, D. P.; Hunter, C. N.; Tajkhorshid, E.; Aksimentiev, A.; Schulten, K. Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism. *Cell* **2019**, *179* (5), 1098−1111.
(52) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (1), 12−21.
(53) Jiang, W.; Phillips, J. C.; Huang, L.; Fajer, M.; Meng, Y.; Gumbart, J. C.; Luo, Y.; Schulten, K.; Roux, B. Generalized Scalable Multiple Copy Algorithms for Molecular Dynamics Simulations in NAMD. *Comput. Phys. Commun.* **2014**, *185* (3), 908−916.
(54) Vilar, S.; Cozza, G.; Moro, S. Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. *Curr. Top. Med. Chem.* **2008**, *8* (18), 1555−1572.
(55) Jakobi, A. J.; Wilmanns, M.; Sachse, C. Model-Based Local Density Sharpening of Cryo-EM Maps. *eLife* **2017**, *6*, e27131.
(56) Mobley, D. L.; Dill, K. A. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get. *Structure* **2009**, *17*, 489−498.
(57) Kanal, I. Y.; Keith, J. A.; Hutchison, G. R. A Sobering Assessment of Small-Molecule Force Field Methods for Low Energy Conformer Predictions. *Int. J. Quantum Chem.* **2018**, *118* (5), e25512.
(58) Nunes-Alves, A.; Zuckerman, D. M.; Arantes, G. M. Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways. *Biophys. J.* **2018**, *114* (5), 1058−1066.
(59) Feher, V. A.; Schiffer, J. M.; Mermelstein, D. J.; Mih, N.; Pierce, L. C. T.; McCammon, J. A.; Amaro, R. E. Mechanisms for Benzene Dissociation through the Excited State of T4 Lysozyme L99A Mutant. *Biophys. J.* **2019**, *116* (2), 205−214.
(60) Menting, J. G.; Whittaker, J.; Margetts, M. B.; Whittaker, L. J.; Kong, G. K. W.; Smith, B. J.; Watson, C. J.; Žáková, L.; Kletvíková, E.; Jiráček, J.; Chan, S. J.; Steiner, D. F.; Dodson, G. G.; Brzozowski, A. M.; Weiss, M. A.; Ward, C. W.; Lawrence, M. C. How Insulin Engages Its Primary Binding Site on the Insulin Receptor. *Nature* **2013**, *493* (7431), 241−245.
(61) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4*, 1−17.

https://dx.doi.org/10.1021/acs.jcim.9b01167
J. Chem. Inf. Model. 2020, 60, 2591−2604