Glycomics and Lipidopics

Extended Abstract

GemSpot allows modeling of ligands in cryo-EM maps

Arunima Singh

Abstract:

Recent developments in cryoelectron microscopy (cryo-EM) now routinely allow near-atomic and atomic-level resolution for biomolecules. Determining structures of protein–ligand complexes, however, remains a challenge, with the resolution of the bound ligand being significantly lower than that of the protein. Thus, understanding ligand binding position necessitates creative solutions, such as using computational chemistry methods in conjunction with experimental data. Robertson et al. have developed GemSpot, an automated computational docking pipeline for modeling and evaluating ligand binding poses in cryo-EM maps. It uses the tool GlideEM to for docking, taking into account the EM map potentials as restraints. Model refinement is done using the OPLS3e force field and quantum mechanics calculations, and water molecule sites are predicted using JAWS. The authors demonstrate the pipeline on a several proteins at varying levels of modeling complexity and EM map resolution.

Introduction:

Electron Cryo-Microscopy (cryoEM) has emerged as a major methodology for high-resolution structure determination of macromolecules and their complexes. The number of deposited cryoEM structures in the PDB[1] with resolution better than 4 Å has increased from 48 in 2015 to almost 1,500 to date. These structures include many macromolecular complexes and membrane proteins that have generally proven very challenging or intractable for traditional structural techniques, particularly X-ray crystallography. For example, GPCRs and ion channels are very important classes of drug targets representing 33% and 18% of FDA approved pharmaceuticals respectively, where structural studies have been historically limited by difficulties associated with their crystallization, although recent advances have been made CryoEM, on the other hand, offers increasingly robust workflows for the structural determination of these types of macromolecules[4]. As a result, cryoEM is opening unprecedented opportunities for structure-based drug discovery on a large variety of targets that were up to recently intractable. Structure-based drug discovery is a rational drug design approach that takes into account the three-dimensional structure of the biomolecular target[5]. The unliganded structure can be employed for large-scale virtual screening to get initial hit compounds with the desired biochemical activity. With lead compound(s) in hand, an experimental structure of the liganded complex is necessary to verify the exact binding mode, and often assists in identifying modifications to improve potency. The correct pose is particularly important for methods such as free energy perturbation calculations, where a compound is alchemically mutated to an analogue over the course of a molecular simulation and the relative free energy of binding is calculated, as was recently successfully shown on a cryoEM derived structure for human ATP-citrate lyase Furthermore, a sufficiently high-resolution (usually <2.5 Å) structure will allow for the identification of bound water molecules, which can play crucial roles in drug design.For example, development of successful HIV protease inhibitors often involves the replacement of a key structural water, Given the recent remarkable progress of cryoEM, the methodology will become an invaluable tool for drug discovery efforts, especially for challenging macromolecular complexes. Underlined by continuous advancements in sample preparation, automated data collection, and improved availability of microscopes capable of achieving high resolution, cryoEM will inevitably be employed in the lead optimization phase to obtain structures of intermediate compounds bound to their targets. Decades of crystallography have led to robust methods of modeling and validating protein-ligand crystal structures. While both X-ray crystallography and single-particle cryoEM are in principle scattering techniques based on the interaction of radiation with a biological specimen, there are key differences that complicate modeling in cryoEM maps and prevent the usage of the metrics developed for crystal structures. In crystallography, the phase information of the scattered radiation that is measured is lost and needs to be recovered with either additional experimental information (*e.g.,* Multi-wavelength anomalous dispersion (MAD), isomorphous displacement) or comparison to known structures (molecular replacement) The initial phase values are then improved during model building by comparing calculated scattering from the current model to the experimental scattering. Thus, X-ray crystallography structure determination involves a continuous cross-talk between

Arunima Singh

Assistant Professor, Arizona State University

model and experimental data with simultaneous feedback on the quality of the model. By contrast, in cryoEM the phases are readily available as they are embedded in the specimen images, which are directly used for the calculation of 3D maps. Once a final threedimensional map has been determined from thousands of experimental projections, the model is built into the map with no further feedback from the raw EM data. Furthermore, the maps obtained in crystallography correspond to the electron density, while in cryoEM they represent the coulombic potential of the molecule under investigation. Thus, using the tools developed for crystallography directly for cryoEM structure modeling can be inherently problematic. While the number of cryoEM maps of macromolecular complexes determined to date is relatively low, the existing structures suggest that there are some fundamental challenges associated with modeling protein-ligand complexes. Even with very high-resolution data for a biomolecule, the resolution of the map for a bound ligand is often significantly lower than its surrounding environment[14]. Given that cryoEM structures derive from flash-frozen macromolecules in aqueous solution, it is perhaps not surprising to observe additional mobility for some ligands within protein active sites. In addition, cryoEM reconstructions are vulnerable to spurious map features, currently evident with different software yielding noticeably different maps from the same dataset. This characteristic may arise from inaccuracies in image defocus estimation and correction of the contrast transfer function at high resolution, as well as variability in masking and weighting schemes employed in different software platforms for processing cryoEM data. Notably, in some cases, even different settings with the same software will yield map deviations that may have significant effects in ligand modeling. This problem is compounded by the fact that ligands lack the structural constraints adopted by proteins, *e.g.*, secondary structure constraints that facilitate more robust modeling. Such caveats present the modeler with the challenge of identifying the bound pose of a ligand within a relatively high-resolution cryoEM map, resulting in often incorrect ligand poses and interpretations with significant implications for molecular mechanism and drug discovery efforts. Parallel to developments in cryoEM, computational chemistry methods for modeling protein-ligand complexes have improved significantly over time. Computational force fields have been successfully used for decades to describe the energy and forces of various conformations of proteins[15]. These force fields have been expanded to accurately describe the energy and force of a large variety of ligands, and can easily be expanded by users to cover ligands of interest or even be automatically extended to cover ligands outside of those used in the initial parameterization. Such force field parameters have been used for a variety of applications including dynamics and for enumerating the conformations of proteins and ligands that will be accessible in biologically relevant conditions. Molecular docking is an approach that uses force fields, in conjunction with highly optimized sampling and refinement algorithms, to predict protein-ligand binding modes given only the conformation of the protein and the identity of the ligand. This methodology has been extensively applied to both identify ligands that bind to specific proteins with high affinity and to predict their protein-ligand binding conformations It should be noted however that, in the absence of experimental data, these purely computational methods are often hampered by significant false positive and false negative rates. For structure-based drug design, significant emphasis has also been put on predicting the location of water molecules, which often coordinate ligand binding in pockets and have profound effects in pharmacological activities. Several approaches for predicting hydration sites, including grid-based approaches like JAWS and dynamics approaches like WATERMAP are now capable of predicting the location of bound water molecules. These computational predictions yield impressive agreement with experimentally derived structures and further highlight the role of hydration in lead optimization It thus becomes apparent that an array of well-established computational tools can be employed in combination with cryoEM to address the challenge of modeling ligands into cryoEM maps. To this end, we have developed and validated 'GemSpot', a pipeline of computational chemistry methods that assists in obtaining the most probable bound pose using a combination of ligand docking coupled with refinement, quantum mechanical (QM) calculations, automatic water placement and additional external information, all while taking into account the experimental cryoEM data. The GemSpot pipeline has been validated against a varied set of 19 structures obtained from cryoEM data ranging from 1.9-4.3 Å resolution, consisting of both protein and RNA, together with a diverse selection of ligands, including small molecules and peptides (Scheme of all ligands in Supplementry. In the first step using GemSpot ,the ligand is docked with the popular software GLIDEby employing a novel combination of the traditional GLIDE docking score function and a real space cross-correlation score to the map. This software, called GlideEM, generates several candidate poses for the ligand that are then subjected to real space refinement with PHENIX including the state-of-the-art OPLS3e / VSGB2.1 force field A combination of real space correlation coefficient and pre-refinement docking scores

Arunima Singh

Assistant Professor, Arizona State University

are used to eliminate any poses that make little chemical sense or fit poorly into the experimental map. Once the top poses are identified, further computational techniques can be used to generate enhanced confidence in the lead candidate pose, when necessary. For high-resolution EM maps, a free energy approach to hydrate the active site using JAWScan be used to help differentiate potential water molecules from noise in the map and gain insight into ligand interactions. When there are still doubts about the conformation of the molecule, one can leverage quantum chemistry to examine the conformational strain associated with any bound poses, e.g. with GAUSSIANor Jaguar. In situations where these computational methods alone may be unable to determine a single pose that unambiguously fits all of the data, it may be necessary to determine which of the top poses are also consistent with data from other experiments. Particularly valuable is comparison to structure-activity relationship (SAR) data, i.e., whether the prospective pose can effectively explain the changes in binding affinity for analogues of that molecule. By combining the resulting data, a high degree of confidence can often be obtained even with a low resolution or problematic density for the ligand.

Arunima Singh

Assistant Professor, Arizona State University