Inverse Drug Design through Optimal Electrostatics, Molecular Moment Alignment, and Combinatorial Fragment Assembly

Michael D. Altman & Bruce Tidor

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: A fundamental challenge in drug design is how to go directly from a macromolecule’s structure to a small molecule binds tightly and specifically. Although this question is very difficult to answer, one can often look at a crystal structure of a tight-binding macromolecule–drug complex and justify, in retrospect, why the drug binds so tightly. In general, the ligands of tight binding complexes share many of the same features, including a large burial of surface area, shape complimentarity to the binding site, satisfaction of almost all hydrogen bond donors or acceptors, and a near optimal charge distribution [2]. If this is true, then it should be possible, given a particular binding site, to use all available knowledge to define a hypothetical ideal ligand that optimizes all of these observed properties. The next, and more challenging task, is to find real molecules that resemble these ideal ligands, with the hope that they will bind almost as tightly and specifically. Building based on a hypothetical, optimal template is known as inverse or top-down design and represents a novel way of computationally developing drugs with predicted high affinity and specificity.

Previous Work: Existing approaches in computational drug development have almost exclusively been based on forward, bottom-up design. These methods start with no template, and with successive application of chemical rules for tight binding, try to fit whole or partial molecules into a binding site. These existing design methodologies fall mainly into two categories. The first involves trying to fit a library of whole molecules into an active site, usually referred to as docking [1]. For every orientation of each ligand, a low resolution energy function is used to evaluate the fit. The second class of ligand design procedures falls in the category of functional group fitting [3]. Instead of fitting whole molecules, these methods take a library of organic functional groups and try to find places in the macromolecule binding site where they belong. Again, each functional group position is scored with an inexpensive energy function. Both of these methods have several major disadvantages. One is that they rely on extremely fast but inaccurate energy functions that do not adequately describe all important contributions to binding, such as solvation. Another is that the molecule or functional group libraries are fixed and cannot accommodate combinatorial expansion. Functional group fitting has an additional disadvantage in that it is unclear how to connect the distinctly placed groups together into a single molecule without changing their relative orientations. Although these bottom-up ligand-design methodologies have met with some limited success, their general usefulness has been quite limited.

Approach: The developed procedure for inverse drug design begins by completely packing a macromolecular binding site with either spheres or a lattice of points to determine an optimal ligand shape. The centers of the spheres or the lattice points themselves are then subjected to charge optimization [2] in order to determine the optimal charges for binding at these positions. The combination of optimal shape and electrostatics defines a hypothetical ideal ligand which can then be used as a target for combinatorial molecular design. First, a combinatorial library of molecular scaffolds attached to one polar or charged organic functional group are compared to the target using techniques borrowed from computer vision. Both the target ideal ligand and the real molecules are discretized into spherical regions. For each spherical region, a multipole expansion of electrostatics is computed to provide a three-dimensional reference frame [4]. Each molecule is then compared to the target by computing the affine transforms required to map every molecule sphere onto every compatible ideal target sphere. These transforms are then clustered to determine the most likely alignment that maps the real molecule onto the target. This idea is known in the computer vision literature as pose clustering [5]. The alignment is accepted for the next round if two conditions are met. Firstly, the aligned molecule must not bump into the macromolecular receptor. Secondly, the alignment
must have been generated from spheres containing all highly charged regions of the molecule. This serves to prevent alignments that match one region well at the expense of matching another region poorly. If these conditions are met, the molecule is subjected to another round of combinatorial generation where a second functional group is added. The alignment/growth process is then repeated on this library of molecules and so on until no further growth is permitted. At the end of the procedure, many real molecules are produced where each functional group is matching a compatible region of the ideal target. However, many of these molecules may still fail to reproduce the electrostatic properties of the ideal target as a whole. Due to indirect effects in electrostatics, even if all functional groups match, the failure to have matched an additional region may lead to poor predicted binding. In addition, it may be better to not match a region of the target if matching that region precludes the matching of another. As a result, this problem is challenging because more correspondence may not mean a better match. To deal with this problem, the current implementation screens all completed ligands with a low-resolution energy function to remove molecules with these complications.

**Impact:** The inverse drug design methodology has been implemented and rigorously tested on simple test cases such as spherical charge distributions and molecules designed against themselves or other molecules. The method is currently being tested on the *E. coli* chorismate mutase binding site in an effort to design tight binding molecules against a hypothetical ideal ligand built into its active site. Initial results look promising as ligands are being generated that reproduce features found on known binders (Figure 1).

![Figure 1: Two molecules designed against an optimal hypothetical ligand generated from the E. coli chorismate mutase active site (thick lines). They are shown superimposed upon a molecule known to bind (thin lines).](image)

**Future Work:** After further validation and optimization of the inverse design methodology, this tool will be used on two types of systems. Firstly, it will be applied to systems with a large number of known tight binding ligands such as HIV protease or thrombin to test reproduction of known drug molecules. It will subsequently be used to design drug molecules against receptors without or with only weakly binding ligands in order to test predictive power.

**Research Support:** This work is supported by the National Institutes of Health and the MIT-Merck Fellowship Program.

**References:**


