The role of computational simulation in drug discovery

Incorporation of simulations into the drug design process can optimise the potency of existing lead compounds, and uncover the activity of new ones - thus guiding the synthesis of more effective therapeutics.

Nancy Ogihara, Accelrys Inc

More than ever before, simulation methods are enabling pharmaceutical researchers to study the properties of larger and more complex protein systems. Simulations provide a tremendous amount of information that is not available from a static picture alone. For example, a protein might ‘open up’ to reveal an apparently inaccessible active site cleft that enables a ligand or potential drug candidate to bind. Simulating the movements and behaviours of molecules provides a more thorough understanding of intermolecular interactions, and can reveal the motions responsible for controlling biological and chemical activity.

Force-fields

Forming the foundation of simulations are force-fields, which combine equations to compute the energy of a molecule with parameters determined by fitting those equations to various molecular properties. Force-fields commonly employ a combination of internal coordinates and terms – including bond distances, bond angles, torsion angles and so on – to describe part of the potential energy surface due to interactions between bonded atoms. Force-fields also incorporate van der Waals and electrostatic interactions to describe non-bonded interactions between atoms. The functional forms of the equations range from simple quadratic forms to Morse functions, Fourier expansions, Lennard-Jones potentials and many others. At the heart of a force-field-based simulation is the calculation of the potential energy for a given configuration of atoms. The calculation of this energy, along with its first and second derivatives with respect to the atomic coordinates, yields the information necessary for energy minimisation, harmonic vibrational analysis and dynamics simulations. This calculation is actually performed by the simulation engine or force-field-based program.

Simulation engines are the computational packages that handle the application of force-fields in minimisation, dynamics and other molecular mechanics simulations. These engines are used to perform energy minimisation for optimising model geometry; molecular dynamics for searching conformational space of residues in active sites,
and calculating free energies of binding (including solvation and entropy effects); and other force-field-based calculations, such as analysing intramolecular or intermolecular interactions in terms of residue-residue or molecule-molecule interactions, energy per residue, or interactions within a radial distance.

To the protein modeller and structure-based designer, the simulations and the force-fields behind them play an important role in molecular modelling. The quality of the force-field used, its applicability to the model at hand, and its ability to predict the particular properties measured in a simulation directly determine the validity of the results. This has a direct bearing on accurately modelling drug-receptor interactions and predicting drug-binding affinities – where inaccurate models can lead to erroneous hypotheses regarding structure-activity relationships.

Not to be left out of the discussion of advances in simulation technology is the increasing speed of computers; this enables faster simulations spanning longer periods of time. Two decades ago, the first simulation covered an elapsed time of 8 picoseconds ($8 \times 10^{-12}$ seconds). Now researchers are running simulations in the multi-nanosecond time-frame. Since changes in molecular conformations of proteins often take milliseconds, faster and longer simulations bring improved and realistic insights into these complicated chemical and biological problems.

The drug design process
Structure-guided drug design has become the preferred technique for the discovery of new pharmaceuticals. Computer simulation of the binding of potential drugs to target receptors is an essential step in this process.

Analysing 3D structures of proteins gives us an unprecedented insight into the behaviour and mechanics of drug binding and biological activity. The study of biomolecular interactions using computational techniques – including modelling and simulation – provides a microscopic picture of events that may be difficult to investigate experimentally in the laboratory, and can help reveal topological features that are important in determining molecular recognition. This information can then be used for predicting ligand-receptor complex formation, and designing ligands and protein mutations that produce desired ligand-receptor interactions.

Why use simulation? The traditional lock and key model for protein-ligand docking (1) is useful as a first approximation of an interaction; however, highly accurate docking requires the consideration of conformational changes in the protein receptor. In addition, solvent effects may play a crucial role in the binding of ligands.

In today’s pharmaceutical laboratory, modelling and simulation methods are an integral part of developing drugs that optimally bind to protein targets with greater selectivity and improved efficacy. Having an accurate picture of what happens at the molecular level includes an understanding of how atoms move. By incorporating simulations into the drug design process, these strategies can help optimise the potency of existing lead compounds as well as uncover novel ones – thus guiding the synthesis of more effective therapeutics.

Many simulation strategies for drug discovery are available today; simulation tools available from Accelrys include CHARMm (2), C2.LigandFit (6) and Affinity (6). Described below is how each can be used to speed up and facilitate the drug discovery process.

Tools of the trade
CHARMm (Chemistry at HARvard Macromolecular Mechanics) from Professor Martin Karplus’ group

Figure 1. Using CHARMm, a researcher can study the dynamics and thermodynamics of macromolecules such as prion proteins to understand structure-function relationships.

Figure 2. Active site identification for HIV protease using the LigandFit flood-filling technique.
at Harvard University has capabilities for energy minimisations, dynamics simulations, free energy perturbation (FEP), combined quantum and molecular mechanics (QM/MM) methods and analysis. It can perform calculations on a variety of small or large molecular systems – including proteins, nucleic acids, carbohydrates, lipids and small molecules – and computes molecular behaviour based on empirical descriptions of the system’s energy.

In conjunction with a graphical interface such as Accelrys’ Insight II (3) or QUANTA (5), the system provides scientists with a powerful set of tools to explore in detail the dynamics and thermodynamics of a protein’s active site, and its ability to bind a ligand. The variety of calculations it can perform to hone in on potential drug compounds include interaction and conformational energies, local minima, barriers to rotation and time-dependent dynamic behaviour.

Whether it is protein dynamics, protein-drug interactions or many other aspects of biomolecular structure, CHARMm’s simulation engine provides information concerning molecular-level structure, interactions and energetics, and thus helps provide a complete picture of the structure and dynamics of protein-ligand interactions.

**Affinity** Docking a flexible ligand to a protein active site is a critical step in the process of structure-based drug design. As part of the Insight II modelling environment, the Affinity module automatically docks potential drug molecules to receptors and identifies low energy orientations of the molecule bound to the receptor. Using force-field-based methods, Affinity automatically finds the best binding modes. This energy-driven method is especially useful in structure-based drug design where the experimentally determined structure of a protein-ligand complex is unavailable.

The system is especially useful for cases where ligand binding is accompanied by some level of conformational change in the protein receptor. These changes range from side-chain movements to displacements of whole loop regions, giving the advantage that orientations do not have to be predetermined prior to docking. Affinity allows the exploration of these potential movements by enabling the definition of flexible areas in the protein’s binding site as well as in the ligand. Using a two-step process to dock the ligands, first the ligand is placed within the receptor using a ‘Monte Carlo’-type procedure to search both conformational and Cartesian space. Second, simulated annealing optimises the location of each ligand placement. During this phase, the bulk of the receptor (that is, atoms not in the user-specified binding site) is held rigid, while the binding site atoms and ligand atoms are allowed to move.

**Figure 3.** Docking of benzamidine onto trypsin using Affinity. The crystallographically-determined position for benzamidine is shown with green carbons; the other coloured structures represent four of the lowest energy docked conformations calculated by Affinity. The green ribbons represent the protein backbone of trypsin (1BIT.PDB).
With Affinity, the flexibility of the receptor is inherent in the calculations. The grid-based methodology allows the efficient and accurate calculation of solvation effects on the ligand. In addition, the system allows the user to incorporate experimental data, such as known hydrogen bonds, into the calculations to help refine the docking study.

C2.LigandFit is a structure-based design application that operates in the Cerius2 modelling and simulation environment (4). It identifies the binding sites on a protein target, performs flexible docking and has various scoring capabilities to evaluate energies of the docked ligand-protein complexes. Several possible sites are proposed so as to select and edit the most probable binding site. This is especially useful when a binding site has not been identified for a theoretical model or experimental structure of a protein.

C2.LigandFit uses the energy of the ligand-receptor complex to automatically find the best binding modes of the ligand to the receptor. Using a stochastic conformational search technique, it retains the best results from the conformational sampling. For non-bonded interactions, Ligandfit uses a grid method to evaluate the interactions between a rigid protein and the movable atoms of a flexible ligand. Grid methods have proven to be very effective for fast and accurate approximations of protein-ligand interactions compared with full force-field representations (6,7). The conformers resulting from a C2.Ligandfit calculation can then be scored and ranked using several scoring functions. The system's fast flexible docking ultimately allows the quick evaluation and prioritisation of series of compounds with respect to their ability to fit and bind to a target receptor.

**Conclusion**

In spite of the advances in force-field and computational development, and the tools that are now available, it should be noted that researchers are still a far way off from a comprehensive description of the interactions that contribute to molecular properties. Not every arrangement of atoms has been parameterised, and new potential drug candidates can conceivably have arrangements of functional groups not previously envisioned. Additionally, partial charges, atom types and van der Waals parameters – each necessary to obtain an accurate picture of a compound – may not be accurately determined.

Notwithstanding these drawbacks, the latest generation of simulation technologies provides the pharmaceutical industry – more than ever before – with the flexibility and power to focus on the best drug candidates, while minimising the loss of valuable time throughout the many stages of drug discovery. The challenge still remains, however, to accurately simulate macromolecular systems and events in a biological time-scale. Although advances in computational speed have propelled the field forward, the greatest advances will come with the continued development and application of new algorithms, better force-fields and a realistic understanding of the system at hand.

**References**


