Chapter 19
Molecular Docking

Garrett M. Morris and Marguerita Lim-Wilby

Summary Molecular docking is a key tool in structural molecular biology and computer-assisted drug design. The goal of ligand–protein docking is to predict the predominant binding mode(s) of a ligand with a protein of known three-dimensional structure. Successful docking methods search high-dimensional spaces effectively and use a scoring function that correctly ranks candidate dockings. Docking can be used to perform virtual screening on large libraries of compounds, rank the results, and propose structural hypotheses of how the ligands inhibit the target, which is invaluable in lead optimization. The setting up of the input structures for the docking is just as important as the docking itself, and analyzing the results of stochastic search methods can sometimes be unclear. This chapter discusses the background and theory of molecular docking software, and covers the usage of some of the most-cited docking software.

Keywords: AutoDock · Computer-assisted drug design · DOCK · FlexX · GOLD · ICM · Molecular recognition · Protein–ligand docking

1 Introduction

The field of molecular docking has emerged during the last three decades driven by the needs of structural molecular biology and structure-based drug discovery. It has been greatly facilitated by the dramatic growth in availability and power of computers, and the growing ease of access to small molecule and protein databases [1–4]. The goal of automated molecular docking software is to understand and predict molecular recognition, both structurally, finding likely binding modes, and energetically, predicting binding affinity. Molecular docking is usually performed between a small molecule and a target macromolecule. This is often referred to as ligand–protein docking, but there is growing interest in protein–protein docking. In this chapter, we will focus on ligand–protein docking, and use the more generic term

From: Methods in Molecular Biology, vol. 443, Molecular Modeling of Proteins
Edited by Andreas Kukol

"target" to refer to the protein, DNA, or RNA macromolecule to which a much smaller molecule (or "ligand") is being docked.

Molecular docking has a wide variety of uses and applications in drug discovery, including structure–activity studies, lead optimization, finding potential leads by virtual screening, providing binding hypotheses to facilitate predictions for mutagenesis studies, assisting x-ray crystallography in the fitting of substrates and inhibitors to electron density, chemical mechanism studies, and combinatorial library design.

Virtual screening on the basis of molecular descriptors and physicochemical properties of (in)active ligands has great usefulness in finding hits and leads through library enrichment for screening [5], a strategy that is also well-used for reducing and enriching the library of ligands for molecular docking; there are recent reports that ligand shape-matching does as well as, if not better than, docking [6]. However, molecular docking when used as the final stage in virtual screening helps to provide a three-dimensional (3D), structural hypotheses of how a ligand interacts with its target.

Given the limitations of space, and in the interests of fairness, we are not able to survey the details of specific docking tool, except where illustrative, and we, therefore, refer the reader to the documentation provided with each of these tools. Instead, we aim to provide an overview comparing and contrasting the methodologies of the most cited [7] docking tools, namely AutoDock [8–10] http://autodock.scripps.edu; DOCK [11, 12] http://dock. compbio.ucsf.edu/; FlexX [13] http://www.biosolveit.de/FlexX; GOLD [14, 15] http://www.ccdc.cam.ac.uk/products/life_sciences/gold/; and ICM [16] http://www.molsoft.com/docking.html.

2 Theory

There are a number of excellent reviews of molecular docking methods [7, 17] and a large number of publications comparing the performance of a variety of molecular docking tools [18–29], often for virtual screening. It should be stressed that comparing docking methods is difficult [28], and because there is evidence that some docking methods do better with certain classes of target than others, the reader is encouraged to try several docking methods to determine which one(s) work best for their target of interest. The process of taking a known crystal structure of a complex of the target of interest, separating the ligand, and then docking the ligand back into the *apo*-form of the target is known as "re-docking." The reader should compare the ability of their chosen docking methods and parameters to re-dock a variety of ligands to the target of interest. Success is often measured in terms of root mean square deviation (RMSD) of the Cartesian coordinates of the atoms of the ligand in the docked and crystallographic conformations; a docking is generally regarded as successful if this is less than the somewhat arbitrary threshold of 2 Å; there are alternative measures of success, such as whether the correct ligand–target interactions are recovered.

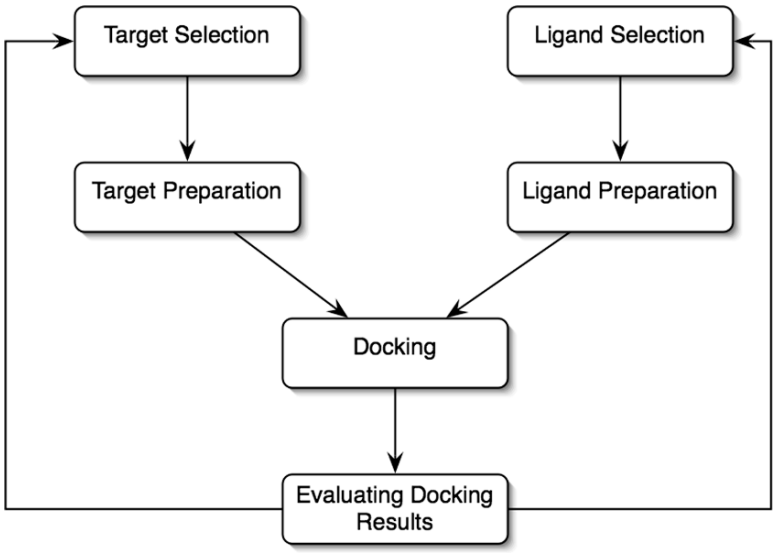

Fig. 1 A typical docking workflow. This flowchart shows the key steps common to all docking protocols. The 3D structures for the target macromolecule and the small molecule must first be chosen, and then each structure must be prepared in accordance with the requirements of the docking method being used. Following the docking, the results must be analyzed, selecting the binding modes with the best scores

Figure 1 shows the key steps in docking that are common to all protocols. Docking involves finding the most favorable binding mode(s) of a ligand to the target of interest. The binding mode of a ligand with respect to the receptor can be uniquely defined by its state variables. These consist of its position (x-, y-, and z-translations), orientation (Euler angles, axis-angle, or a quaternion), and, if the ligand is flexible, its conformation (the torsion angles for each rotatable bond). Each of these state variables describes one degree of freedom in a multidimensional search space, and their bounds describe the extent of the search. Rigid body docking is faster than treating the ligand as flexible, because the size of the search space is much smaller, but if the conformation of the ligand is not correct, then there will be a lower probability of finding a complementary fit.

All docking methods require a scoring function to rank the various candidate binding modes, and a search method to explore the state variables. Scoring functions can be empirical, force field based, or knowledge based, whereas search methods fall into two major categories: *systematic* and *stochastic*. Systematic search methods sample the search space at predefined intervals, and are deterministic. Stochastic search methods iteratively make random changes to the state variables until a user-defined termination criterion is met, so the outcome of the search varies; Sousa et al. discuss these classes of algorithms in more detail [7]. Search methods can also be classified by how broadly they explore the search space, as either *local* or *global*. Local search methods tend to find the nearest or local minimum energy

to the current conformation, whereas global methods search for the best or global minimum energy within the defined search space. Hybrid global–local search methods have been shown to perform even better than global methods alone, being more efficient and able to find lower energies [8].

In AutoDock 4, for example, there is the choice of two local search methods (Solis and Wets [30] and Pattern Search [31]); two global search methods: Monte Carlo (MC) simulated annealing (SA) [32], and the genetic algorithm (GA) [33–35]; and one hybrid global–local search method, the Lamarckian GA (LGA) [8]. DOCK uses a systematic search method to match chemical features between the ligand and the negative image of the binding site. FlexX matches ligand features with complementary interaction sites. GOLD's global search method is a GA. The search method of ICM combines a biased MC procedure and a local energy minimization.

2.1 Target Selection and Preparation

Ideally, the target structure should be experimentally determined, usually by either x-ray crystallography or nuclear magnetic resonance. Docking has been performed successfully against homology models [36–39], although the reliability of the docking results depends heavily on the quality and bias of the homology model.

In some cases, the biologically relevant form of the target structure—the biological unit—is a multimer, which means that the appropriate symmetry-related molecules must also be included in the target structure. The online database Binding MOAD, for example, provides target structures as a biological unit [2] suitable for docking studies.

Many docking tools do not allow the target to be flexible, although this is a very important aspect of molecular recognition [40]. A target may adopt different conformations in the unbound and bound states, and with different classes of ligands; examples of different degrees of structural change on ligand binding are given in [41]. To tackle this and other problems, molecular dynamics has found an increasing number of applications in conjunction with molecular docking. These range from preparing the target before docking, to accounting for receptor flexibility, solvent effects, and induced fit, to calculating binding free energies and ranking docked ligands [42]. The so-called "relaxed complex method" developed in the laboratory of McCammon [43] generates snapshots from molecular dynamics simulations [44–46] of the *apo* form of the target, and then applies AutoDock to dock the ligand of interest; the technique effectively takes into account induced fit, and has been applied to develop novel inhibitors of HIV integrase [47].

In all x-ray crystal structures, there is a range of certainty with which atomic positions are defined. This is quantified by the temperature factors (also known as B-values) assigned to each atom in the PDB file. It is possible in some molecule viewers to color the atoms by B-value, which can visually indicate regions with more structural ambiguity. For a given x-ray crystal structure, atomic positions that may be suspect are those with 1) B-values higher than their surroundings, or 2)

incomplete side chains (some atom positions are not assigned by the crystallographer). Furthermore, in certain crystal structures, alternate locations of atoms may be observed: in such cases, both alternatives must be tested.

To accelerate the scoring calculation, some docking methods precalculate grid maps to represent the receptor when calculating interaction energies with a ligand. A set of grid maps for a given receptor can be reused for docking of a library of ligands, also saving time. In general, grid maps are not transferable from one docking tool to another. For AutoDock, a grid map needs to be computed for each atom type in the ligand or set of ligands being docked, in addition to electrostatic potential and desolvation grid maps.

2.2 Ligand Selection and Preparation

The type of ligands chosen for docking will depend on the goal: for lead discovery, crude filters such as net charge, molecular weight, polar surface area, solubility, commercial availability, and price-per-compound can be applied to reduce the number of molecules to be docked. For lead optimization, filters such as similarity thresholds, pharmacophores, synthetic accessibility, and absorption, distribution, metabolism, excretion, and toxicology (ADME-Tox) properties are additionally applied. For focused lead optimization, a custom library of analogs that are related to the lead compound(s) is often constructed for docking, to inform and prioritize medicinal chemistry efforts [48]. Refer to Chap. 17 for more information regarding filtering libraries of compounds.

AutoDock uses a united-atom model for the ligand and receptor, in which only polar hydrogens are present. It also requires partial atomic charges to be assigned to the ligand. The AutoDock scoring functions were calibrated using Gasteiger charges [49] on the ligand, thus, to use the scoring functions correctly, the ligand must be assigned Gasteiger partial charges. It should be noted that alternative charge calculation methods for ligands have been successfully used in AutoDock [24].

Most docking tools treat ligands flexibly, with the exception of ring conformations. In general, the more rotatable bonds in a ligand, the more difficult and time consuming the docking will tend to be. This is because the size of the search space increases exponentially with the number of torsions. More highly branched torsion trees lead to more difficult searches than do linear torsion trees. Rotation of conjugated bonds, such as in amides, carbamates, ureas, etc., should be limited. One strategy to explore ring flexibility is to perform conformational analysis on any ring-containing ligands before docking [50–52]; another strategy is to compute the conformations of flexible ring systems during the docking, as can be done using FlexX [53] with either CORINA [54, 55] or Confort [56].

2.3 Docking

Molecular docking involves computationally exploring a search space that is defined by the molecular representation used by the method, and ranking candidate solutions to determine the best binding mode. Thus, docking requires both a search method and a scoring function.

Search methods can be divided into two main categories: *systematic* and *stochastic*. In the former case, the outcome of the search is deterministic, but the quality of the solution depends on the granularity of sampling of the search space. Stochastic methods rely on an element of randomness, therefore, the outcome varies. Systematic search methods are commonly used in rigid protein–rigid protein docking, where there are only six degrees of freedom, in programs such as DOT [57], GRAMM [58, 59], and ZDOCK [60]. Stochastic search methods are more suitable for higher-dimensional problems, such as flexible ligand–protein docking. Stochastic search methods include MCSA [10], GAs [8, 14, 15], and hybrid global–local search methods [8].

Scoring functions can be empirical, knowledge based, or molecular mechanics-based, see [17] for a review. In addition, some docking strategies use one scoring function during the docking, and a different one postdocking to rerank the results; such retrospective scoring, however, cannot affect the efficiency and accuracy of the primary scoring function [61].

The AutoDock scoring function is based on the molecular mechanics force field AMBER [62], with two additional terms: one to model the desolvation free energy change on binding, which is based on atomic solvation parameters [63]; and one empirical term to model the loss of conformational entropy on binding [8, 63]. The AutoDock scoring function in version 3 and later was inspired by the work of Böhm [64]. The individual contributions to the total energy of binding, namely van der Waals, hydrogen bonding, electrostatic, desolvation, and number of rotatable bonds in the ligand, were treated as independent variables. These were used to train a linear regression model given the observed free energy of binding, using a training set that included 30 protein–ligand complexes for AutoDock 3 [8] and 188 complexes for AutoDock 4 [63].

2.4 Evaluating Docking Results

Regardless of the ligand–protein docking tool used, docking results should be evaluated by considering the chemical complementarity between ligand and protein. Are all possible hydrogen bond donors and acceptors in the ligand satisfied? Are the charged groups in the ligand interacting with oppositely charged side chains in the receptor, or are they buried in hydrophobic pockets? Are hydrophobic groups in the ligand buried in hydrophobic pockets in the receptor?

Furthermore, the parameters chosen for the docking can be judged by the docking tool's ability to reproduce the binding mode of a ligand to protein, when the structure

of the ligand–protein complex is known. The criterion usually used is the all-atom RMSD between the docked position and the crystallographically observed binding position of the ligand, and success is typically regarded as being less than 2 Å.

When docking using stochastic methods, it is recommended that the experiment be run at least 50 times with different initial conditions. The similarity of the predicted binding modes can be assessed by computing a matrix of pairwise RMSD values, and clustering docked conformations according to an RMSD threshold, typically 2 Å. If all of the dockings cluster into one family, this indicates that the search parameters were sufficient for each docking to converge. If there is no clustering at all, then the dockings should be repeated but with increased sampling: either increasing the number of iterations per search, increasing the number of searches, or, if the method is population based, increasing the population size.

If the scoring function were perfect, the docked conformation with the lowest energy would always correspond to the crystallographically observed binding mode, assuming that there are no bad contacts in the crystal structure. This is not always the case, and sometimes a different binding mode is observed significantly more often than the lowest energy binding mode. Furthermore, current docking methods will tend to find the binding mode with the lowest possible interaction energy for a given ligand: this score does not necessarily indicate whether the ligand even binds. There has been growing interest in developing methods to distinguish binders from nonbinders. One of the earliest reports that used docking to successfully discriminate binders from nonbinders [65] considered a simple metric that combined the mean binding energy for all of the conformations in the cluster, and the total number of conformationally distinct clusters found out of 100 dockings. The more clusters and the weaker the mean energy, the less likely the ligand was to bind. By building on statistical mechanical foundations, new methods are emerging that estimate the contributions of translational and rotational entropy to binding affinity, by approximating the configurational entropy using the sizes of the clusters [66, 67].

3 Methods

No matter which docking method is selected, the user needs to prepare the appropriate input files. This will depend on the docking method used, and in particular, on the molecular representation used in that method. To assist the user in setting up and in postdocking analysis, many docking programs include auxiliary tools, scripts, and graphical user interfaces (GUI); Table 1 summarizes some of these.

Docking methods that do not use a force field, such as FlexX and GOLD, do not require partial charges to be assigned to the atoms in the ligand and receptor molecules. AutoDock and UCSF DOCK, on the other hand, use an AMBER-derived force field and, therefore, require partial atomic charges. The AutoDock 3 scoring function was calibrated using Kollman united-atom partial charges on the macromolecule, unlike AutoDock 4, which uses Gasteiger PEOE charges for both ligand and macromolecule. It is important to note that other AutoDock users

Table 1 Ligand input requirements for the most commonly cited docking software

Docking tool	Auxiliary tools	File format	Hydrogen atoms	Partial charges
AutoDock 4	AutoGrid, ADT, BDT	mol2, PDBQT	United atom	Gasteiger PEOE[1]
DOCK 6	Chimera, Grid, Docktools, Nchemgrids, Sphgen, ANTECHAMBER	mol2	Explicit or united atom	AM1-BCC, Gasteiger
FlexX 2	FlexV	mol2, SD	United atom	Formal charge only
GOLD 3	GOLD Front End, SILVER	mol2, SD[2]	Explicit	None
ICM 3.4	ICM-Pro, ICM-VLS	mol2, SD	Explicit	MMFF, ICM

[1] Alternative partial charge calculations can be used (for example, AMSOL [68, 69] with the AM1-CM2 Hamiltonian; see [24])
[2] PDB format is also possible, but not recommended

have investigated the use of alternative partial charges on the ligand: e.g., Evans and Neidle concluded that the best charges to be used in AutoDock 3 for virtual screening of DNA minor groove binders came from calculations using AMSOL [68, 69] with the AM1-CM2 Hamiltonian for nonpolar organic solvent [24].

AutoDock is distributed with a GUI called AutoDockTools (ADT; see http://autodock.scripps.edu/resources/adt). ADT helps to prepare the ligand and receptor input files, and to set up the AutoGrid and AutoDock calculations. BDT [70] is an alternative preparatory tool to ADT (see http://www.quimica.urv.cat/∼pujadas/BDT/) that helps in setting up virtual screening runs with AutoDock, and in setting up collections of AutoGrid maps for blind docking and also in combining grid maps to incorporate structural variability in the receptor. AutoDock's AutoGrid program and DOCK's Grid program precompute the necessary grid maps that describe the chemical potential at regular intervals around the target. In addition, DOCK's Sphgen program uses spheres to create the required "negative image" of the binding site.

The key stages in docking are: 1) target selection and preparation, 2) ligand selection and preparation, 3) docking setup, and 4) evaluating docking results; these are discussed in the following sections.

3.1 Target Selection and Preparation

1. Gather structures of the target, ideally with bound ligands, from internal and external sources. Good publicly available sources include the Protein Data Bank [4], http://www.rcsb.org/pdb; ReLiBase [1], http://relibase.ccdc.cam.ac.uk; and Binding MOAD [2, 71] http://www.bindingmoad.org. See **Note 1**.

2. Discard any structures that lack the biologically necessary cofactors, if any are required for biological activity. Structures that are incomplete or missing side chains should also be disregarded.
3. If there is more than one target structure, overlay them by superimposing the key residues in the binding site or region of interest using a least-squares superimposition method. SwissPdbViewer [72], a freely available tool from http://www.expasy.org/spdbv, offers several superimposition options under its "Fit" menu, such as "Magic Fit" and "Fit molecules (from selection)". Note also that SwissPdbViewer can also automatically reconstruct incomplete side chains.
4. Identify the extent of the structural variability and select a representative structure (see **Note 2**).
5. Add all hydrogen atoms to the target at the desired pH; under physiological conditions at pH 7.2, the following residues have ionized side chains: arginine, lysine, aspartic acid, and glutamic acid. This defines the formal charges (see **Note 3**). Each histidine side chain can be either neutral or positively charged at physiological pH. If it is neutral, either the delta or the epsilon nitrogen can be protonated (see **Note 4**).
6. The atomic assignments of imidazole rings in histidine and amido groups in asparagine and glutamine side chains can be ambiguous; tools such as REDUCE and its web interface, MOLPROBITY [73, 74] can evaluate 180° flips of these groups to optimize the hydrogen-bond network, and add hydrogen atoms appropriately.
7. Remove all water molecules, except those that are integral to your binding hypothesis(see **Note 5**).
8. If the representative target structure is complexed with a ligand, remove the ligand.
9. Calculate the partial charges, if required by the docking tool (see Table 1). Some tools may use a dictionary of amino acid partial charges to simply assign the charges. If there are any cofactors in the target structure, it will be necessary to compute the appropriate partial charges if required by the docking method.
10. When using AutoDock, merge nonpolar hydrogens, because it uses a united-atom representation (see Table 1 and **Note 6**).
11. AutoDock uses grid maps that must be calculated using AutoGrid. Each map describes a 3D grid of interaction energies with the target, one for each atom type in the ligand (see **Note 7**).

3.2 Ligand Selection and Preparation

Most docking tools require a 3D structure for each ligand, including explicit hydrogens. Depending on the source of the ligands—real molecules, molecules that have yet to be synthesized, or vendor libraries—the steps required to process the

molecules will vary. The following steps exemplify how to obtain these structures, and how to process them for use in AutoDock.

1. ZINC is one of the largest collections of commercially available compounds; it is well curated and has 4.6 million compounds (http://blaster.docking.org/zinc; see also [3]). It is particularly useful for molecular docking because it provides 3D structures in SYBYL MOL2 formats, and is also free of charge. Subsets of compounds can be created by composing a query that specifies constraints on both molecular properties and two-dimensional (2D) molecular topology.
2. Ligands in the form of SMILES strings [75] can be converted into full 3D atomic coordinates, including hydrogens, using tools such as CORINA [54, 55] or ZINC [3].
3. Ligands in 2D SD format [76] can be converted into full 3D atomic coordinates using CORINA [54, 55] or Ghemical [77–79]. Ghemical can be used to sketch the ligand in 3D and then perform energy minimization, molecular dynamics, or conformational search to identify low energy conformations. PRODRG [80–82] can take PDB format, MDL MOL files, or even ASCII–text drawings of the molecule, instead of SD format. PRODRG is available as a standalone executable or as a web service, where the user can sketch the molecule in 2D and then convert the molecule into 3D; PRODRG is convenient for AutoDock 2.4 and 3, because it outputs PDBQ format.
4. It is important that the protonation, tautomeric, and stereoisomeric forms of the ligand be correct, otherwise subsequent calculations will be highly suspect. The enumeration of all possible ligand tautomers can be achieved with such programs as QUACPAC (Open Eye) [83], TAUTOMER (Molecular Networks), and LigPrep (Schrödinger).
5. When preparing ligands for AutoDock, the GUI AutoDockTools (ADT) can be used to set up the necessary input files. The first step for AutoDock is to calculate Gasteiger partial charges [49] and assign AutoDock atom types to each atom in the ligand (see **Notes 8–10**).
6. Define the "root" of the torsion tree and the rotatable bonds interactively using ADT. The "Ligand > Torsion Tree > Detect Root…" option automatically examines all the rotatable bonds in the ligand and chooses the atom that is nearest to the center of the torsion tree. The "Ligand > Torsion Tree > Choose Torsions…" option displays all rotatable bonds as green or magenta, indicating that they are active or inactive, respectively. Clicking on these bonds toggles whether they are active or not. Make sure any conjugated bonds are not rotatable.
7. AutoDock 4 requires the ligand to be in PDBQT format, which is very similar to PDB format, but also includes the partial atomic charge and the AutoDock atom type for each atom. The ligand should be saved using the "Ligand > Output > Save as PDBQT…" option in ADT.

3.3 Docking

1. Define the search space. There are two possibilities, depending on how much is known about the binding site:
   (a) If there is no previous information regarding the location of the binding site, then the translational search space should encompass the entire surface of the receptor. This is known as "blind docking," and is possible with AutoDock [84]. If the docking tool cannot encompass the whole target, then probable sites such as cavities large enough to contain the ligand(s) should be investigated separately; the third-party tool BDT [70] can be used to set up staggered grid boxes for AutoGrid.
   (b) If there is previous information, such as ligands with known binding modes, active site residues, or mutagenesis data, then the search space can be reduced to focus on the region of interest, thus, simplifying the search problem.
2. Set the target to be docked to, using the ADT menu item "Docking > Macromolecule > Set Rigid Filename…"
3. Select the search method (if there is more than one), and set the appropriate parameters. AutoDock offers MCSA, a traditional GA, and a hybrid global–local search method called LGA. The best search algorithm was shown to be LGA [8], therefore, we recommend this for most dockings (see **Note 11**).
4. Save the input parameter file for the docking tool, if necessary. For AutoDock, use the "Docking > Output > Lamarckian GA…" option in ADT to save an AutoDock docking parameter file (DPF) set up to perform LGA dockings.

3.4 Evaluating Docking Results

When evaluating the results of dockings, there are two main criteria to consider: 1) how well did the binding mode predicted by the docking match known structural data, where available; and 2) how well did the docking rank the ligands? If the method's scoring function is designed to predict binding affinities, how well did it match experimental binding data?

To answer the first criteria, a crystal structure of the complex of the ligand bound to the target must be known, and then the RMSD between the docked and the "reference" crystallographic binding mode of the ligand can be calculated; success is usually counted as RMSD less than 2 Å. To answer the second criteria, inhibition constants, or K_i values, must be known for the ligands and the target system.

When the search method used is stochastic, it is important to consider how often a given binding mode was predicted across all the dockings that were run. This is usually achieved using conformational clustering, building families of related conformations using RMSD tolerances to decide whether two conformations belong in the same cluster.

1. Read in all of the docked conformations into the docking analysis tool. For AutoDock, use ADT with the menu option "Analyze > Dockings > Open..." for one docking log (DLG) (see **Note 12**).
2. It is useful to view the dockings in the context of the target, therefore, if necessary, load the structure of the target. In ADT, use "Analyze > Macromolecule > Open..." to read in the target PDBQT structure used to compute the AutoGrid maps.
3. Perform conformational cluster analysis on the dockings to assess the level of agreement in the results. In ADT, use "Analyze > Clusterings > Recluster...", and type in a list of RMSD tolerances in angstroms separated by spaces. This performs clustering for each RMSD tolerance value, grouping the docked conformations accordingly.
4. Display the conformational clustering as a histogram, and visually inspect each cluster. In ADT, use "Analyze > Clusterings > Show..." and then choose the RMSD tolerance value. This displays a histogram of number of docked conformations in the cluster, versus the energy of the most tightly binding conformation in that cluster. The histogram is interactive, thus, clicking on a histogram bar sets up the "play" buttons in the "Conformation Player" window to play through the conformations in that cluster. This window has buttons to play forward and backward, and to step through the conformations one at a time.
5. It is possible to examine AutoDock-docked conformations in more detail using the "Conformation Player" in ADT, by clicking on the "&" button. This displays a panel in which it is possible to show more information regarding the current docking, by clicking on the "Show Info" check-button. It is also possible to monitor which hydrogen bonds are formed between the ligand and the target using the "Build H-bonds" check-button. The atoms in the ligand can be colored by a color scale that goes from dark blue to green to yellow to orange to red, indicating more favorable to less favorable interaction energies, using the "Color by" option; "vdw" colors by van der Waals or H-bond plus desolvation free energy, "elec_stat" colors by electrostatic interaction energy, and "total" colors by the total interaction energy; and "atom" returns to the default color-by-atom coloring.
6. If the docking results do not cluster into at least one significantly populated cluster, with an RMSD tolerance of between 2 and 3 Å, this is an indication that the dockings did not search for long enough. In AutoDock and ADT, increase the number of energy evaluations used in the LGA, and rerun the dockings. To get decent statistics, it is advisable to repeat the docking at for at least 50 runs. See Sect. 3.3, Step 3.
7. If the docked conformations are too far from the target structure, make sure that the AutoGrid grid box is centered on or near the target; the grid box can be visualized in ADT using the "Analyze > Grids > Open...", then choosing one of the grid map files. The x-, y-, and z-axes are color-coded red, green, and blue, respectively. The energy values in the grid map can be isocontoured by dragging the blue solid triangle on the "IsoValue" slider, with lower energy values indicating pockets of tighter binding affinity, and higher-resolution isocontours can be plotted using a "Sampling" value of 1 instead of the default value, 3.

4 Notes

1. It is preferable to use only high-resolution structures where available, ideally better than 2.5 Å.
2. As an alternative to Steps 1 to 3 in Sect. 3.1, *Target Selection and Preparation*, a representative structure or "leader" for a 90% homology family of structures is already precalculated and available from Binding MOAD [2] http://www.bindingmoad.org.
3. In AutoDockTools, use the "Edit > Hydrogens > Add" then choose "All Hydrogens"; all hydrogens are required for the initial Gasteiger partial charge calculation, but the nonpolar hydrogens will be merged later on.
4. AutoDockTools offers a tool to help set the desired protonation state of each His side chain, under the "Edit > Hydrogens > Edit Histidine Sidechains" menu. Which protonation state a His adopts will depend on its environment in the target.
5. Consolv [85], freely available from http://www.bch.msu.edu/labs/kuhn/software.html, "predicts whether water molecules bound to the surface of a protein are likely to be conserved or displaced in other, independently solved crystallographic structures of the same protein."
6. AutoDockTools calculates the partial charges and merges the nonpolar hydrogens automatically when the user selects the "Grid > Macromolecule > Choose…" menu items.
7. AutoGrid requires the target to be saved in PDBQT format, and it requires a Grid Parameter File (GPF). To save the receptor, use the "File > Save > Write PDBQT…" option. Set which types of grid maps should be calculated using either "Grid > Set Map Types > Directly…" or "Grid > Set Map Types > Choose Ligand…". To set up the location and grid spacing of the grid maps, use "Grid > Grid Box…". Finally, to save the GPF, use "Grid > Output > Save GPF…".
8. Note that ADT can read in a ligand with partial charges using SYBYL mol2 format. Use "File > Read Molecule…" and change the "Files of type" button to "MOL2 files (*.mol2)".
9. If the ligand is missing hydrogen atoms, then they must be added before calculating the Gasteiger charges. After selecting the ligand in ADT, use the menu option "Edit > Hydrogens > Add". It is very important to consider the tautomeric and ionization states when adding hydrogens, or use one of the tools in Step 4 in Sect. 3.2.
10. AutoDock atom types are assigned automatically in ADT by choosing the "Ligand > Input > Choose…" option. This command will also merge the nonpolar hydrogens, making the ligand suitable for use with the united-atom force field in AutoDock.
11. It is important to make sure that the number of energy evaluations is increased from the default value of 250,000 if the ligand has any rotatable bonds. Use the ADT menu option "Docking > Search Parameters > Genetic Algorithm…" and change the "Maximum number of energy evaluations" to at least 2,500,000.

It is also possible to increase the "Number of GA runs" in the panel from the default value 10. One other important parameter is the "Population Size"; the default is 150, although Hetenyi et al. showed that larger values up to 300 can improve the efficiency of the search [84]. Note also that this panel works for both the traditional and LGA search methods.

12. Alternatively, if the same ligand has been docked to the same target, but separate runs of AutoDock have produced uniquely named DLG files, as is the case when running dockings in parallel on computational clusters, use the "Analyze > Dockings > Open All..." option to read in all the dockings.

Acknowledgements This manuscript is TSRI publication number 18829. This work was supported by the National Institutes of Health (NIH) grant R01-GM069832.

References

1. Hendlich, M. (1998) Databases for protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, **54**(Pt 6 Pt 1): 1178–1182.
2. Hu, L., Benson, M.L., Smith, R.D., Lerner, M.G., and Carlson, H.A. (2005) Binding MOAD (Mother Of All Databases). *Proteins*, **60**(3): 333–340.
3. Irwin, J.J. and Shoichet, B.K. (2005) ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model*, **45**(1): 177–182.
4. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**(1): 235–242.
5. Pozzan, A. (2006) Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery. *Curr Pharm Des*, **12**(17): 2099–2110.
6. Hawkins, P.C., Skillman, A.G., and Nicholls, A. (2007) Comparison of shape-matching and docking as virtual screening tools. *J Med Chem*, **50**(1): 74–82.
7. Sousa, S.F., Fernandes, P.A., and Ramos, M.J. (2006) Protein-ligand docking: Current status and future challenges. *Proteins*, **65**(1): 15–26.
8. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*, **19**: 1639–1662.
9. Morris, G.M., Goodsell, D.S., Huey, R., and Olson, A.J. (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des*, **10**(4): 293–304.
10. Goodsell, D.S. and Olson, A.J. (1990) Automated docking of substrates to proteins by simulated annealing. *Proteins*, **8**(3): 195–202.
11. Ewing, T.J.A. and Kuntz, I.D. (1997) Critical evaluation of search algorithms for automated molecular docking and database screening. *J Comput Chem*, **18**(9): 1175–1189.
12. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T.E. (1982) A geometric approach to macromolecule-ligand interactions. *J Mol Biol*, **161**(2): 269–288.
13. Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996) A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, **261**(3): 470–489.
14. Jones, G., Willett, P., Glen, R.C., Leach, A.R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*, **267**(3): 727–748.
15. Jones, G., Willett, P., and Glen, R.C. (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol*, **245**(1): 43–53.
16. Abagyan, R.A., Totrov, M.M., and Kuznetzov, D.A. (1994) ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry*, **15**: 488–506.

17. Taylor, R.D., Jewsbury, P.J., and Essex, J.W. (2002) A review of protein-small molecule docking methods. *J Comput Aided Mol Des*, **16**(3): 151–166.
18. Bissantz, C., Folkers, G., and Rognan, D. (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem*, **43**(25): 4759–4767.
19. Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., and Shenkin, P.S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*, **47**(7): 1739–1749.
20. Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T., and Banks, J.L. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*, **47**(7): 1750–1759.
21. Kellenberger, E., Rodrigo, J., Muller, P., and Rognan, D. (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins*, **57**(2): 225–242.
22. Kontoyianni, M., McClellan, L.M., and Sokol, G.S. (2004) Evaluation of docking performance: comparative data on docking algorithms. *J Med Chem*, **47**(3): 558–565.
23. Perola, E., Walters, W.P., and Charifson, P.S. (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins*, **56**(2): 235–249.
24. Evans, D.A. and Neidle, S. (2006) Virtual screening of DNA minor groove binders. *J Med Chem*, **49**(14): 4232–4238.
25. Evers, A., Hessler, G., Matter, H., and Klabunde, T. (2005) Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J Med Chem*, **48**(17): 5448–5465.
26. Cummings, M.D., DesJarlais, R.L., Gibbs, A.C., Mohan, V., and Jaeger, E.P. (2005) Comparison of automated docking programs as virtual screening tools. *J Med Chem*, **48**(4): 962–976.
27. Cotesta, S., Giordanetto, F., Trosset, J.Y., Crivori, P., Kroemer, R.T., Stouten, P.F., and Vulpetti, A. (2005) Virtual screening to enrich a compound collection with CDK2 inhibitors using docking, scoring, and composite scoring models. *Proteins*, **60**(4): 629–643.
28. Cole, J.C., Murray, C.W., Nissink, J.W., Taylor, R.D., and Taylor, R. (2005) Comparing protein-ligand docking programs is difficult. *Proteins*, **60**(3): 325–332.
29. Vigers, G.P. and Rizzi, J.P. (2004) Multiple active site corrections for docking and virtual screening. *J Med Chem*, **47**(1): 80–89.
30. Solis, F.J. and Wets, R.J.-B. (1981) Minimization by random search techniques. *Mathematical Operations Research*, **6**: 19–30.
31. Conn, A.R., Gould, N.I.M., and Toint, P.L. (1991) A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, **28**(2): 545–572.
32. Kirkpatrick, S., C. D. Gelatt, J., and Vecchi, M.P. (1983) Optimization by simulated annealing. *Science*, **220**(4598): 671–680.
33. Holland, J.H., *Adaptation in natural and artificial systems*. 1992, Cambridge, MA: The MIT Press. 211.
34. Goldberg, D.E., *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st ed. 1989, Boston, MA: Addison-Wesley Longman Publishing Co., Inc. 372.
35. Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Program*. 3rd ed. 1996, London, UK: Springer-Verlag. 387.
36. de Graaf, C., Oostenbrink, C., Keizers, P.H., van der Wijst, T., Jongejan, A., and Vermeulen, N.P. (2006) Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J Med Chem*, **49**(8): 2417–2430.
37. Diller, D.J. and Li, R. (2003) Kinases, homology models, and high throughput docking. *J Med Chem*, **46**(22): 4638–4647.
38. Evers, A. and Klabunde, T. (2005) Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J Med Chem*, **48**(4): 1088–1097.

39. Shoichet, B.K., McGovern, S.L., Wei, B., and Irwin, J.J. (2002) Lead discovery using molecular docking. *Curr Opin Chem Biol*, **6**(4): 439–446.
40. Murray, C.W., Baxter, C.A., and Frenkel, A.D. (1999) The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J Comput Aided Mol Des*, **13**(6): 547–562.
41. Gunasekaran, K. and Nussinov, R. (2007) How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J Mol Biol*, **365**(1): 257–273.
42. Alonso, H., Bliznyuk, A.A., and Gready, J.E. (2006) Combining docking and molecular dynamic simulations in drug design. *Med Res Rev*, **26**(5): 531–568.
43. Lin, J.H., Perryman, A.L., Schames, J.R., and McCammon, J.A. (2002) Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J Am Chem Soc*, **124**(20): 5632–5633.
44. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., III, Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Ferguson, D.M., Radmer, R.J., Seibel, G.L., Singh, U.C., Weiner, P.K., and Kollman, P.A. (1997) *AMBER 5* University of California: San Francisco.
45. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., III, J., W., Ross, W.S., C., S., Darden, T., Merz, K.M., Stanton, R.V., Cheng, A., Vincent, J.J., Crowley, M., V., T., Gohlke, R.R., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, C., Weiner, P., and Kollman, P.A. (2002) *AMBER 7* University of California: San Francisco.
46. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem*, **26**: 1781–1802.
47. Schames, J.R., Henchman, R.H., Siegel, J.S., Sotriffer, C.A., Ni, H., and McCammon, J.A. (2004) Discovery of a novel binding trench in HIV integrase. *J Med Chem*, **47**(8): 1879–1881.
48. Gastreich, M., Lilienthal, M., Briem, H., and Claussen, H. (2006) Ultrafast *de novo* docking: combining pharmacophores and combinatorics. *J Comput Aided Mol Des*, in press.
49. Gasteiger, J. and Marsili, M. (1978) A new model for calculating atomic charges in molecules. *Tetrahedron Lett.*, **34**: 3181–3184.
50. Mulakala, C., Nerinckx, W., and Reilly, P.J. (2006) Docking studies on glycoside hydrolase Family 47 endoplasmic reticulum alpha-(1 → 2)-mannosidase I to elucidate the pathway to the substrate transition state. *Carbohydrate Research*, **341**(13): 2233–2245.
51. Laederach, A. and Reilly, P.J. (2005) Modeling protein recognition of carbohydrates. *Proteins-Structure Function and Bioinformatics*, **60**(4): 591–597.
52. Rockey, W.M., Laederach, A., and Reilly, P.J. (2000) Automated docking of alpha-(1 → 4)- and alpha-(1 → 6)-linked glucosyl trisaccharides and maltopentaose into the soybean beta-amylase active site. *Proteins-Structure Function and Genetics*, **40**(2): 299–309.
53. Rarey, M., Kramer, B., and Lengauer, T. (1995) Time-efficient docking of flexible ligands into active sites of proteins. *Proc Int Conf Intell Syst Mol Biol*, **3**: 300–308.
54. Gasteiger, J. and Sadowski, J. (1992) *CORINA* 3.4, Molecular Networks GmbH: Erlangen, Germany, http://www.molecular-networks.com/online_demos/corina_demo.html.
55. Gasteiger, J., Rudolph, C., and Sadowski, J. (1992) Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.*, **3**: 537–547.
56. Pearlman, R.S. and Balducci, R. *Confort: A Novel Algorithm For Conformational Analysis*. in *National Meeting of the American Chemical Society*. 1998. New Orleans, LA.
57. Ten Eyck, L.F., Mandell, J., Roberts, V.A., and Pique, M.E., *Surveying Molecular Interactions With DOT*. 1995.
58. Vakser, I.A. (1997) Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins*, **Suppl 1**: 226–230.
59. Tovchigrechko, A. and Vakser, I.A. (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res*, **34**(Web Server issue): W310–314.
60. Chen, R. and Weng, Z. (2002) Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins*, **47**(3): 281–294.

61. Mohan, V., Gibbs, A.C., Cummings, M.D., Jaeger, E.P., and DesJarlais, R.L. (2005) Docking: successes and challenges. *Curr Pharm Des*, **11**(3): 323–333.
62. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, J., Kenneth M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**: 5179–5197.
63. Huey, R., Morris, G.M., Olson, A.J., and Goodsell, D.S. (2007) A semiempirical free energy force field with charge-based desolvation. *J Comput Chem*, **28**(6): 1145–1152.
64. Böhm, H.-J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des*, **8**(3): 243–256.
65. Rosenfeld, R.J., Goodsell, D.S., Musah, R.A., Morris, G.M., Goodin, D.B., and Olson, A.J. (2003) Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. *Journal of Computer-Aided Molecular Design*, **17**(8): 525–536.
66. Ruvinsky, A.M. and Kozintsev, A.V. (2005) New and fast statistical-thermodynamic method for computation of protein-ligand binding entropy substantially improves docking accuracy. *Journal of Computational Chemistry*, **26**(11): 1089–1095.
67. Ruvinsky, A.M. and Kozintsev, A.V. (2006) Novel statistical-thermodynamic methods to predict protein-ligand binding positions using probability distribution functions. *Proteins-Structure Function and Bioinformatics*, **62**(1): 202–208.
68. Cramer, C.J. and Truhlar, D.G. (1992) AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution. *J Comput Aided Mol Des*, **6**(6): 629–666.
69. Hawkins, G.D., Giesen, D.J., Lynch, G.C., Chambers, C.C., Rossi, I., Storer, J.W., Li, J., Zhu, T., Thompson, J.D., Winget, P., Lynch, B.J., Rinaldi, D., Liotard, D.A., Cramer, C.J., and Truhlar, D.G. (2007) *AMSOL* 7.1, Department of Chemistry and Supercomputer Institute, University of Minnesota: Minneapolis, Minnesota, http://comp.chem.umn.edu/amsol/.
70. Vaque, M., Arola, A., Aliagas, C., and Pujadas, G. (2006) BDT: an easy-to-use front-end application for automation of massive docking tasks and complex docking strategies with AutoDock. *Bioinformatics*, **22**(14): 1803–1804.
71. Smith, R.D., Hu, L., Falkner, J.A., Benson, M.L., Nerothin, J.P., and Carlson, H.A. (2006) Exploring protein-ligand recognition with Binding MOAD. *J Mol Graph Model*, **24**(6): 414–425.
72. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**(15): 2714–2723.
73. Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res*, **32**(Web Server issue): W615–619.
74. Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, **50**(3): 437–450.
75. Weininger, D., *Daylight Theory Manual*. 2006, Daylight Chemical Information Systems, Inc.
76. Dalby, A., Nourse, J.G., Hounshell, W.D., Gushurst, A.K.I., Grier, D.L., Leland, B.A., and Laufer, J. (1992) Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J Chem Inf Comput Sci*, **32**: 244–255.
77. Acton, A., Banck, M., Bréfort, J., Cruz, M., Curtis, D., Hassinen, T., Heikkilä, V., Hutchison, G., Huuskonen, J., Jensen, J., Liboska, R., and Rowley, C. (2006) *Ghemical* 2.00, Department of Chemistry, University of Kuopio: Kuopio, Finland, http://www.uku.fi/∼thassine/projects/ghemical/.
78. Hassinen, T. and Peräkylä, M. (2001) New energy terms for reduced protein models implemented in an off-lattice force field. *J Comput Chem*, **22**(12): 1229–1242.
79. Hassinen, T., Hutchison, G., Cruz, M., Banck, M., Rowley, C., and Curtis, D. (2007) *Ghemical-GMS* 2.10, Department of Chemistry, University of Iowa.: Iowa City, IA, http://www.uiowa.edu/∼ghemical/ghemical.shtml.
80. van Aalten, D. and Oswald, S. (2007) *PRODRG* 2, University of Dundee: Dundee, Scotland, http://davapc1.bioch.dundee.ac.uk/programs/prodrg/.

81. van Aalten, D.M., Bywater, R., Findlay, J.B., Hendlich, M., Hooft, R.W., and Vriend, G. (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J Comput Aided Mol Des*, **10**(3): 255–262.
82. Schuttelkopf, A.W. and van Aalten, D.M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, **60**(Pt 8): 1355–1363.
83. Skillman, A.G. *QUACPAC* OpenEye Scientific Software: Santa Fe, NM, http://www.eyesopen.com/products/applications/quacpac.html.
84. Hetenyi, C. and van der Spoel, D. (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci*, **11**(7): 1729–1737.
85. Raymer, M.L., Sanschagrin, P.C., Punch, W.F., Venkataraman, S., Goodman, E.D., and Kuhn, L.A. (1997) Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J Mol Biol*, **265**(4): 445–464.