The Demonstration of AutoDock as an Educational Tool for Drug Discovery

Travis R. Helgren, Timothy J. Hagen*

Department of Chemistry and Biochemistry, Northern Illinois University, 1425 W. Lincoln Hwy, DeKalb, IL 60115, USA

Purpose: This experiment is intended demonstrate how computational modeling can be used in drug discovery. Students will dock a series of known CDK2 inhibitors discovered by fragment growth. Following completion of the experiment, students will understand how to use the docking software and will be able to apply fragment growth to hit molecules.

Materials: The following software packages are necessary to perform molecular docking: AutoDockTools, ChemSketch (or ChemDraw), and OpenBabel. They are open source for academic purposes with the exception of ChemDraw.

Introduction

The practice of drug discovery can be broken into specific phases, each involving multiple years of research and scrupulous cataloging of data to bring a novel therapeutic agent to market.[1] This process begins with the identification of a biological target (usually a protein or class of proteins) which can be manipulated to elicit a desired biological response. Chemical agents capable of causing such a response are then discovered and modified to maximize interactions with the target (Figure 1). The resulting compounds are then subject to numerous biological screening assays (toxicological, animal models, whole-cell assays, etc.) to elucidate both the safety and efficacy of the therapeutics. Finally, the appropriate compounds are monitored in clinical trials and the results are sent to the FDA for approval. If the potential drug molecule fails to meet industry and regulatory standards at any point along this discovery and development path, the compound fails to become a new drug. Because the requirements to bring a novel therapeutic to clinical trials include years of work and the involvement of many teams of scientists, compounds that fail in clinical trials are highly discouraging and costly. Therefore, a large number of tools have been developed to guide scientists' research efforts, especially in the early stages of drug discovery.

As shown in Figure 1, potential chemotherapies against a specific target are first discovered and then modified to afford species of maximum potency. At this stage in drug discovery, computational methods can be applied to aid in the design of more potent compounds.[2-6] The exact steric and electrostatic composition of potential binding sites can be determined by elucidation of the protein structure via crystallographic methods; protein structures determined in this manner are placed in the protein databank (PDB; www.pdb.org) for open source applications. Additionally, if potent inhibitory species are co-crystallized within the active site of the protein, the exact binding mechanism can be discerned. This information is paramount, as the inhibitor molecule can be computationally removed from the protein crystal structure leaving behind an unoccupied binding site. Molecular docking programs can then be employed to discern potential binding interactions between new species and the protein crystal structure. The output of this process is generally a predicted binding affinity and a docking score. Compounds can then be ranked based upon these indicators to determine which species

are predicted to bind most favorably; the most potent species can then be synthesized and screened against the target to discover the actual inhibitory activity.

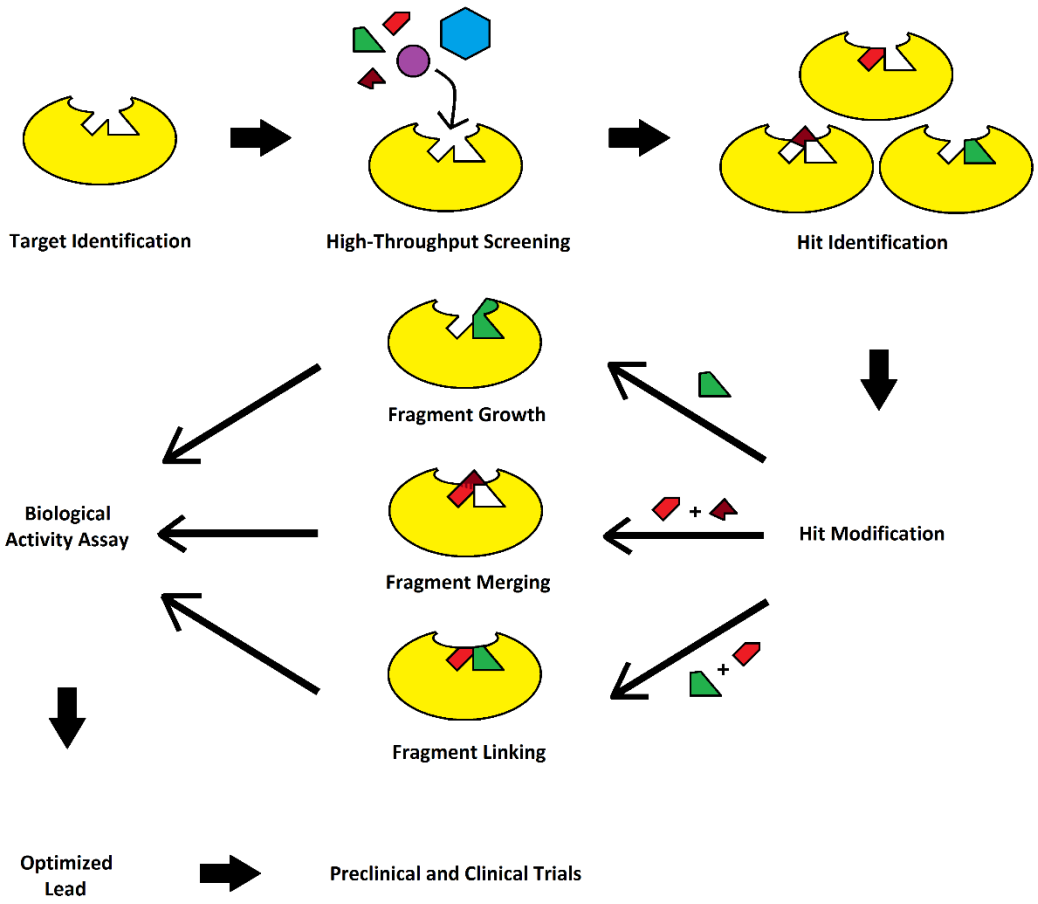


Figure 1: Flow chart demonstrating the early stages of target-based drug discovery. A target is first identified. Fragment libraries are then screened against the target and the hit molecules are crystallized within the active site. Utilizing the binding interactions revealed via the crystal structure of hit molecules bound to the target, the hits are modified, usually according to three approaches: fragment growth, where fragments are combined with groups designed to exploit potential binding interactions; fragment merging, where a fragment hit is conjoined with an optimized lead (olive oval); and fragment linking, where two or more fragments shown to bind in differing regions of the active site are conjoined. The resulting compounds are screened in activity assays and modified according to the observed activities. Once an optimal molecule is obtained, the compound is subjected to preclinical development and sent to clinical trials to determine efficacy.

The AutoDock suite of programs (ADT) determines docking poses of ligands by first generating a model of the binding site.[7] The model results from the combination of various maps of a defined region of interest which includes the binding site, where each individual map corresponds to a specific atom type within the ligand. The maps are generated by placing an sp^3 hybridized carbon atom with a +1

charge at fixed intervals within the predefined map region and the associated interaction energies are calculated. These maps are then used in combination to determine the location and poses of the ligand corresponding to the lowest energy conformation.

However, prior to mapping the protein active site and obtaining docking results, the ligand and receptor files must be generated (Figure 2). Ligands can be drawn with most chemical illustration software (ChemDraw, ChemSketch, etc.) and converted to file types (.pdb and .mol2) readable by ADT through a variety of methods (OpenBabel[8], Chem3D, etc.). Once the ligand file is readable, the rotatable angles (torsions) are set and Gasteiger charges are assigned to each atom. Protein structural files can be downloaded from the PDB and can be read directly with ADT. Most protein crystal structures contain crystallographic artifacts (e.g. water molecules, bound inhibitor or substrate species, buffer constituents or additional protein units) that must be removed prior to docking. Upon their removal, hydrogen atoms must be added to the protein structure. To prepare this structure for map generation, all non-polar hydrogen atoms are removed and Gasteiger charges are assigned to each atom. In ADT, only one map type was written for all hydrogen atoms (H-bond donors) and only polar hydrogens should be used for map generation. The necessary maps are then generated and compounds are docked within the resulting model. ADT allows for visualization of the docking results and provides docking scores and predicted binding affinities for each docking pose. The combination of these two pieces of information (predicted interactions and predicted affinity) allows the user to determine the optimal docking pose.

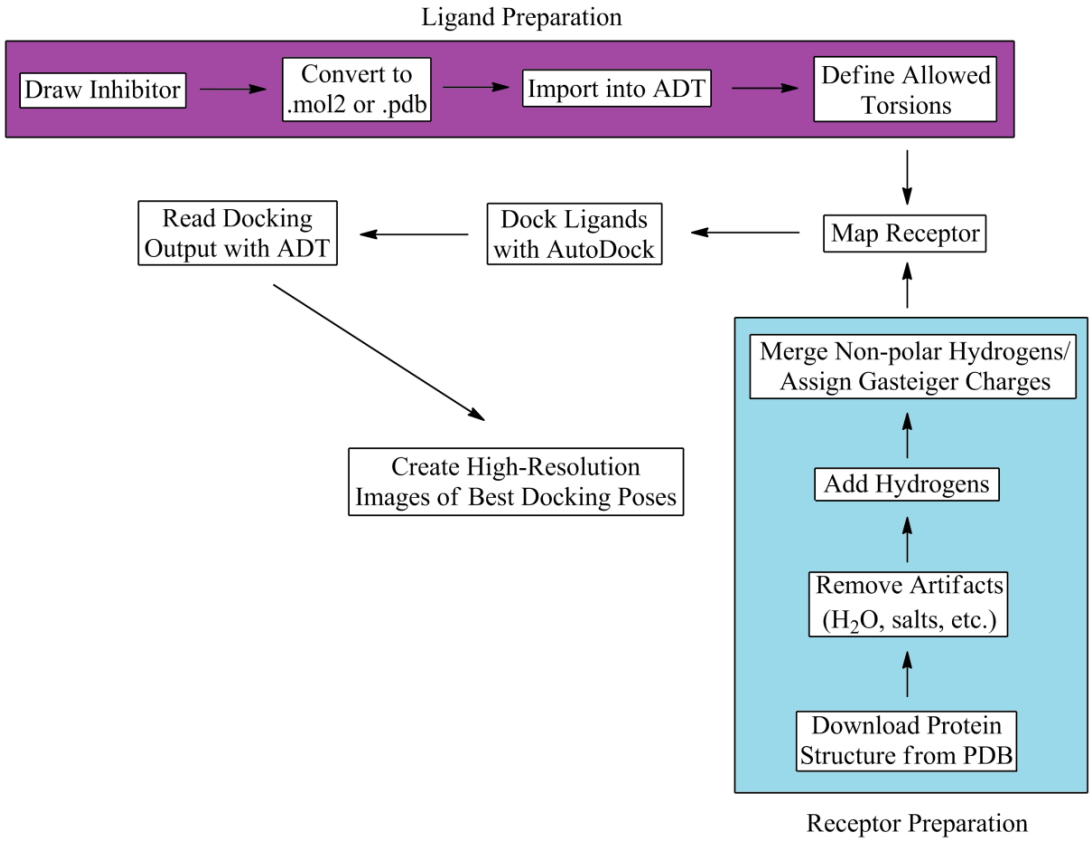


Figure 2: General AutoDock workflow including the steps for ligand and receptor preparation.

To demonstrate both fragment growth methodology and the use of computational modeling in drug discovery, the exercise below describes the docking of a series of indazole (**1**) based inhibitors against cyclin-dependent kinase 2 (CDK2), a promising cancer target (Figure 3). Indazole (**1**) was first identified as a hit molecule (IC_{50} = 185 µM) through high-throughput screening of a library of small molecules.[9] Amide (**2**) was then designed to increase binding to the receptor (IC_{50} = 3 µM) via H-bonding through the amide N-H to the backbone carbonyl of Leu83. Encouraged by the resulting potency increase, sulfonamide (**3**) was designed to further increase the activity (IC_{50} = 0.66 µM) of the original fragment hit through H-bonding to the amide N-H of backbone residue Asp86. Finally, pyrazole (**4**) was screened to determine the necessity of the fused ring system of indazole (**1**). Compound (**4**) exhibited a dramatic decrease in activity (IC_{50} = 97 µM), suggesting the indazole ring system is necessary for potent inhibition of the enzyme.

The procedure below details the docking of indazole (**1**) and the evaluation of the docking results. Using the receptor file (2VTA.pdbqt) you'll prepare when docking (**1**), you can dock the remaining three compounds (**2**-**4**) and compare the output docking poses and scores to those of (**1**). Additionally, compounds (**1**), (**3**) and (**4**) have been crystallized with CDK2, revealing the experimental binding orientations (PDB: 2VTA,[9] 2VTI,[9] and 2VTL,[9] respectively). You can then compare your best docking pose with the actual crystal structure to determine if this docking method produces meaningful output.

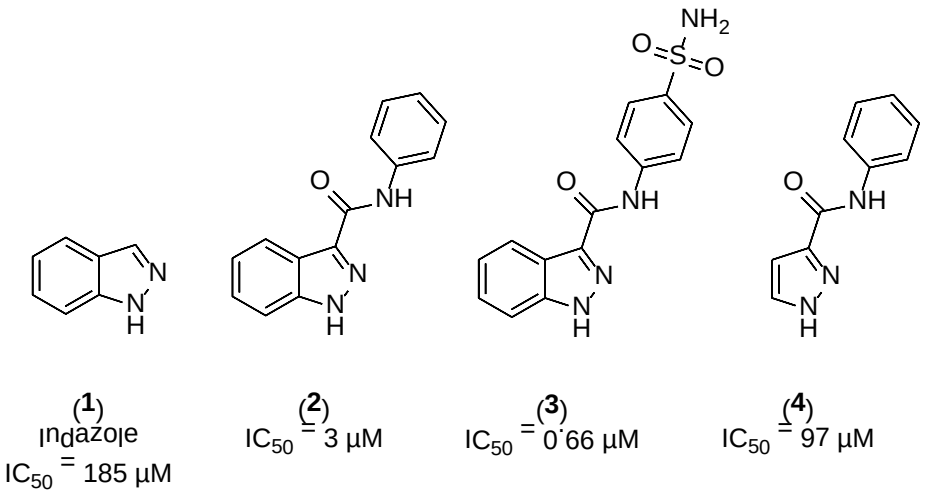


Figure 3: Indazole series of CDK2 inhibitors designed by fragment growth. The listed inhibitory activities correspond to activity against CDK2[9]

Downloading Necessary Software

Various different programs are necessary to perform the actions necessary to go from drawn ligands to output docking poses. Follow the supplied URL's to download any of the programs (with the exception of ChemDraw). Most programs are available on Windows, MacOS and Linux platforms. These include software packages for the following applications:

Structure Drawing:

Programs utilized to construct ligand files from chemical structures:

- The ChemDraw suite of programs: this is the best, as the drawn ligand structures can be imported into Chem3D where energy minimization can take place. The resulting structures can be saved as a Sybyl .mol2 file, which is directly readable by ADT. This eliminates the need for OpenBabel[8], as all files are already in a format readable without conversion.

- ChemSketch: this program is open source, available as a free download for educational and personal applications from the developers (ACD Labs). To download and install, follow the URL below. Click on the "Download Free Version" tab and create an account to register with ACD Labs. Once you have created an account, sign in and click the "Download" tab adjacent to the ACD/ChemSketch Freeware heading. Once the installer downloads, open the file and follow the instructions regarding installation on your machine.
  URL: http://www.acdlabs.com/resources/freeware/chemsketch/

Format Conversion:

- OpenBabel[8]: Highly versatile program that converts structural files between a volume of different formats. To download and install OpenBabel, follow the link below and choose the link to download the appropriate version of the software according to the operating system on your computer. Once downloaded, open the file and follow the installation instructions presented within the Setup window.
  URL: http://openbabel.org/wiki/Category:Installation

Molecular Docking:

Programs utilized to generate docking output to determine computed binding:

- AutoDock4 and AutoGrid4: These are the applications required to run AutoGrid, which generates a model of a defined region of the receptor for the docking of ligands, and AutoDock, which docks ligands into the receptor model. These files need to be copied into your working directory (i.e. the folder that you will be working from). To download these files, follow the URL below. Open the executable file following the download and

follow the instructions within the Setup window. The default directory is a folder called "The Scripps Research Institute". Copy the two files within this folder (entitled "autodock4" and "autogrid4") into your working directory.
URL: http://autodock.scripps.edu/downloads/autodock-registration/autodock-4-2-download-page/

- AutoDockTools (ADT): This is the graphical user interface that is used to prepare ligands and receptors for docking, and to visualize docking results. To download and install ADT, follow the URL below. Select the link corresponding to the correct version of ADT according to the operating system on your computer. Once downloaded, open the file and follow the installation instructions. Once complete, you'll be able to open ADT as you would any other program.
  URL: http://mgltools.scripps.edu/downloads/

Molecular Viewing:

Although ADT can be used for this purpose, other software have been developed specifically for molecular viewing and are more refined:

- Pymol: Very popular program suite used to visualize macromolecules with or without bound structures. Available as a free download for educational applications from the developers. To download and install Pymol, follow the URL below and register with Schrödinger. You'll then receive an email in the address you provided. Open the email and follow the link at the bottom of the page to download the Pymol installer. Once downloaded, open the file and follow the installation instructions.
  URL: http://pymol.org/edu/?q=educational/

- Chimera: Can be used for a number of applications other than visualization. Available as a free download for educational and personal applications. To download and install Chimera, follow the URL below. Select the appropriate link corresponding to the operating system on your computer and open the installation file once downloaded. Follow the instructions within the installation window.
  URL: http://www.cgl.ucsf.edu/chimera/download.html

Procedure

Videos detailing these instructions can be found on the web at:

1. Drawing and Importing Ligands: https://drive.google.com/open?id=0B129zXihqI9iVXZ1a3VJZktZSE0
2. Importing and Cleaning Proteins: https://drive.google.com/open?id=0B129zXihqI9ibUwzQzkzVU9lREU
3. Generating Grid Maps and Running AutoDock: https://drive.google.com/file/d/0B129zXihqI9ibkdlRjdyU3RnZWM/view?usp=sharing
4. Viewing Docking Results: https://drive.google.com/open?id=0B129zXihqI9iQWlyM2h6LU93dWM

NOTE: It is highly recommended that you save all files in a single folder (which we'll call the working directory). We generally save all files associated with a specific docking run in a dedicated folder on the desktop. It will be much easier find and catalog your results if you set the default directory in AutoDockTools to your working directory for each docking run. To do this, click File → Preferences → Set and change the field under "Startup Directory" to the address of your working directory.

1. Drawing Ligands in Chemical Illustration Software

ChemSketch

Download and install the ChemSketch suite of programs from the URL described above. Following installation, open ChemSketch (the default directory is "ACDLabs"). Draw indazole (**1**) using the drawing tools. You must draw the molecule to proportion, as ADT does not change the bond lengths of non-rotatable bonds. Use templates to construct ring systems; when drawing bonds, click on the atom from which you would like to generate a new bond. Once you've drawn the correct structure, save it as a MDL Molefiles [V3000](*.mol) file. *Note: Save all files you generate in your working directory.

ChemDraw

Draw the correct structure in ChemDraw and save it as a .mol file. You can then find the minimum energy conformation of the ligand using Chem3D, an additional program included with most ChemDraw licenses. Open Chem3D; then open the .mol file you just prepared. On the version of Chem3D we used (Version 12.0), the program automatically changes the tautomerism of (**1**) from 1*H*-indazole to the corresponding 2*H*-indazole. You can redraw the structure using the drawing tools built-into Chem3D.

To perform an energy minimization, click "Calculations → MM2 → Minimize Energy". Keep all of the default settings selected and click "Run" to perform the minimization. You'll see the molecule contort until a structure of the minimum energy is found. Save the resulting compound as a .mol2 file. Because this is a file format compatible with ADT, you will not need to use OpenBabel to convert the file, as described below for structures generated with ChemSketch.

**Note: Save all of your inhibitor files with the same file name (e.g. indazole.xxx for indazole files of different extensions). Additionally, save all of these files in a specific working directory titled appropriately (e.g. "Indazole"). This will make cataloging easier.*

2. Converting Between File Types with OpenBabel

Once you've downloaded and installed OpenBabel, open the program (the default directory is "OpenBabel"). The left side of the program interface shows information regarding the input file type and location under the heading "----INPUT FORMAT----". Change the file extension to "mol -- MDL MOL format". Search your computer for the indazole.mol file prepared using ChemSketch and select the appropriate file by clicking on the "…" button underneath the input file address.

The right side of the interface shows information regarding the output file. AutoDockTools (ADT) reads only a few file types, with .pdb, .pdbqt, and .mol2 being the most utilized. We've found that the use of .mol2 files for imported ligands results in fewer errors than when .pdb files are employed; we therefore reserved the use of .pdb file extensions for receptor files imported from the PDB. Change the output file extension under "----OUTPUT FORMAT----" to mol2 -- Sybyl Mol2 format". Define the name of the output file by typing the desired file name into the box under "Output file". Make sure to type the correct file extension along with the file name. For example, the output file name should be "indazole.mol2" to keep the file naming consistent. To determine binding interactions between the ligand and receptor, polar hydrogen atoms must be included in the structure of both the ligand and receptor. To incorporate hydrogen atoms into the structure of the ligand, click the "Add hydrogens (make explicit)" box. Finally, 2D and 3D coordinates must be assigned to the output structural file. If 3D coordinates are not retained, stereochemistry of the ligand will be lost. Click the boxes entitled "Generate 2D coordinates" and "Generate 3D coordinates". Click "Convert" to generate the .mol2 structural file. After a few seconds, you should see content of the output file appear in the previously empty box below the output file name. The output file is conveniently automatically written to the same directory as the input file.

3. Importing the Ligand and Defining Torsions in ADT

The .mol2 file you generated can now be imported into ADT and prepared for docking. After downloading and installing ADT, launch the program (the default directory is "MGLTools"). It is important to set the default directory that the program uses to fetch files as your working directory. To do this, click File → Preferences → Set (Figure 4). A window entitled "Set User Preferences" will appear. Under the heading "Startup Directory", paste the address of your working directory; click "Make Default", then "Dismiss". ADT will save this as the default directory every time you open a new session.

Therefore, change the default directory to your working directory each time you dock a new ligand into a receptor.

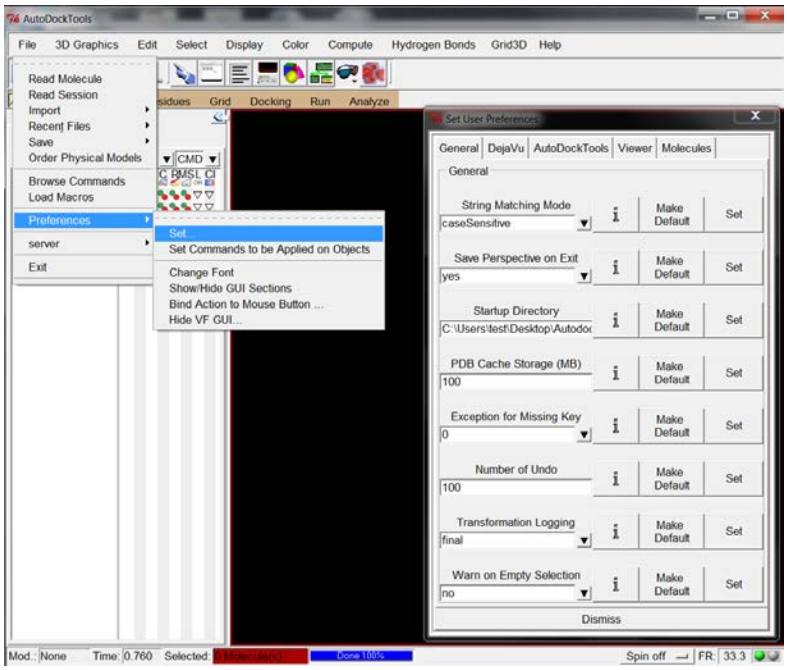


Figure 4: Setting the default directory in ADT.

Now that the directory has been set, you can import your ligand. To do this, click Ligand → Input → Open; change the file extension to "MOL2 files: (*.mol2)", select indazole.mol2, and click open. The structure of indazole (**1**) will appear in the display on ADT and a popup window entitled "summary for indazole" will appear. This window describes that gasteiger charges were added, all non-polar hydrogen atoms were merged, and goes on to discuss the number of aromatic carbons and rotatable bonds found. Click OK to close the window. To prepare the ligand for docking, you must define the allowable torsions. Click Ligand → Torsion Tree → Detect Root; a green sphere will appear over the indazole 1-N atom. This is the torsion root, or the "center of rotation" within the molecule. To visualize which bonds are allowed to be rotatable, click Ligand → Torsion Tree → Choose Torsions. This will change the color of all the bonds to demonstrate which are rotatable. Because the inhibitor does not have any rotatable bonds, all are highlighted in red. However, there will usually be a number of rotatable bonds (green) and non-rotatable bonds (magenta), such as amide bonds, displayed. These parameters can be changed to fix specific bonds or allow additional degrees of freedom in ligand molecules (Figure 5). Click "Done" to close the popup window; the image of (**1**) will change back to colored by atom type. Finally, you need to save the ligand file as a .pdbqt file. Files with this extension contain information regarding partial charges of each atom, information that is necessary for accurate docking predictions, as well as 3D coordinates of each atom. Click Ligand → Output → Save as PDBQT. Title the file "indazole.pdbqt" and save it in the working directory. Hide the ligand from view by clicking the red circle next to "indazole" on the dashboard. The circle is under the heading "L" for "Lines"; these headings designate different viewing options and are: L = Lines, B = Ball and Stick, C = Space Fill Model, R = Ribbon, MS = Molecular Surface. The different options can be selected by clicking the corresponding circle adjacent to the molecule of which you wish to change the display.

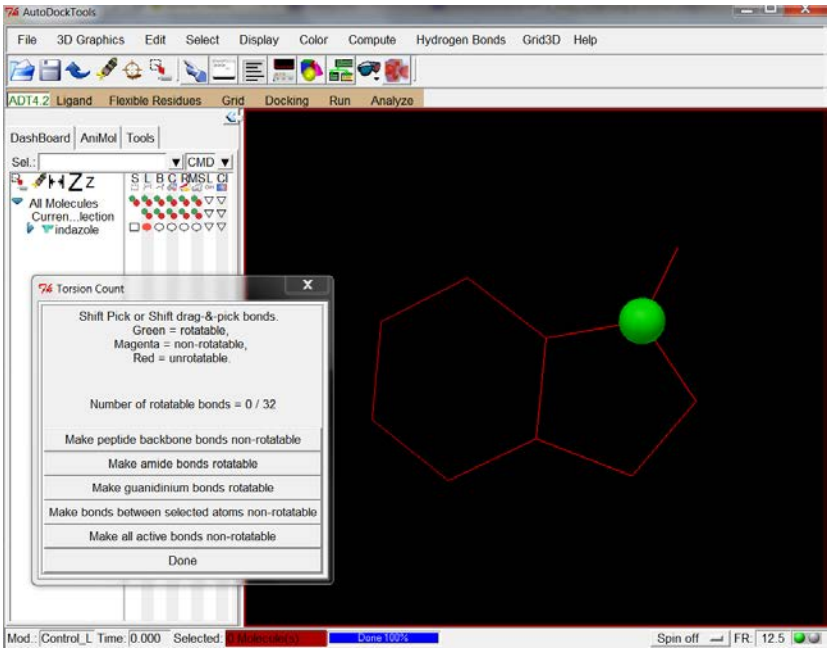


Figure 5: Demonstration of the torsion root of indazole (**1**) in ADT.

4. Importing and Cleaning-Up the Receptor

AutoDockTools (ADT) has a built-in fetch function that can be used to retrieve structures from the PDB. Click File → Import → Fetch from web. A popup entitled “Fetch PDB” will appear. You can enter any 4-digit PDB code to retrieve the corresponding structure. To gretrieve the CKD2 structure with bound indazole (**1**), type “2VTA” and click OK. The protein structure will then appear in the display. Additionally, the structure will be pinned to the dashboard under the name “2VTA”. To change the display to color by atom type, click the inverted triangle under “Cl” and select “By atom type” (Figure 6). All carbon atoms will then change from green to gray.

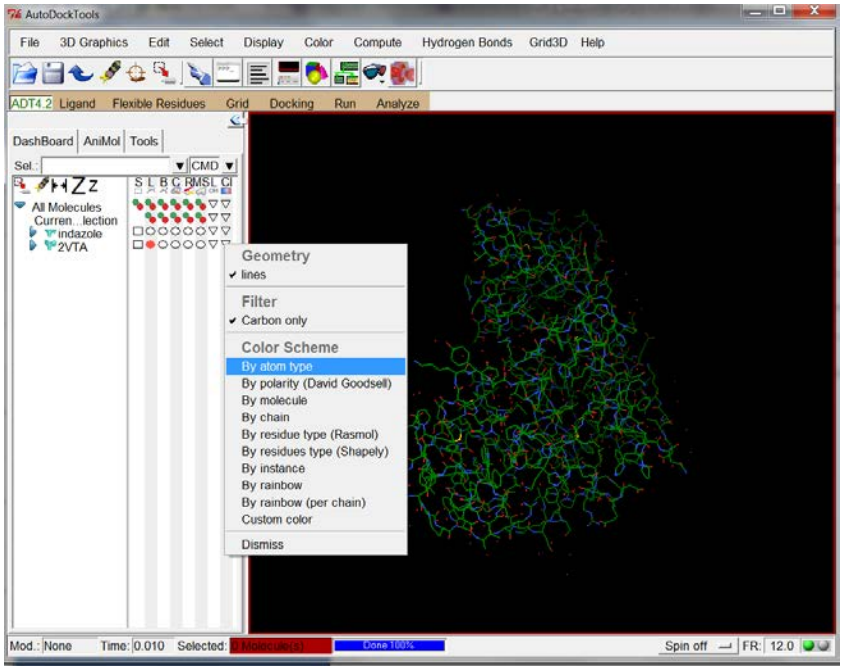


Figure 6: Changing the color scheme of the receptor in ADT.

The structure must now be cleaned up prior to docking; water molecules must be removed, hydrogens and gasteiger charges added, and crystallography fragments or bound inhibitor molecules must be removed. To remove the water molecules, click Edit → Delete Water. You should see the red spheres corresponding water O atoms disappear. Hydrogen atoms must now be added to the receptor. To select only the protein, click the box next to "2VTA" on the dashboard (the box will turn yellow, indicating selection). You should see yellow crosses highlight all of the atoms of the protein. To add hydrogens, click Edit → Hydrogens → Add. Click OK on the popup to add all hydrogens. Click on the selection square again to deselect the protein.

**Note: make sure hydrogen atoms were not added to the ligand structure. If they were, you must merge all non-polar hydrogens (Select the ligand structure and click* Edit → Hydrogens → Merge nonpolar*)*

Finally, you must locate the active site of the protein and determine if any bound fragments are close as these may interfere with docking. The PDB lists 3-digit codes for any bound fragments; for example, there are two molecules of glycerol co-crystallized with the receptor. These can be selected individually with the selection tool; click Select → Select from String. A popup will appear; this is the selection tool and can be used to select individual residues or atoms (atoms can also be selected by holding Shift and clicking on the specific atom within the graphical interface). The code for the glycerol molecules is "GOL"; type "GOL*" under Residue and "*" under Atom. Click "Select" to select the two molecules. At the bottom of the screen is a counter that shows the number of items selected for a specified selection level. You have selected the atoms composing the glycerol molecules; as such, you have selected 28 atoms, 14 for each glycerol. An important feature for of ADT is the storage of specific selections for later viewing. To post this selection to the dashboard under a new heading, click Select → Add selection to dashboard. A popup will appear requesting a title for the selection. Type "Glycerol" and

click OK. A new heading entitled “Glycerol” will appear on the dashboard and your selection will disappear. Notice that the visualization options are new for this type of selection; you now have red and green circles under each visualization category. These are toggled by clicking the green circle to enable a feature and the red circle to disable the feature. Hide the enzyme (2VTA) structure by clicking the red circle under the “L” heading on the dashboard. You can now click a visualization option (such as “L” for lines) to view only the glycerol molecules (Figure 7).

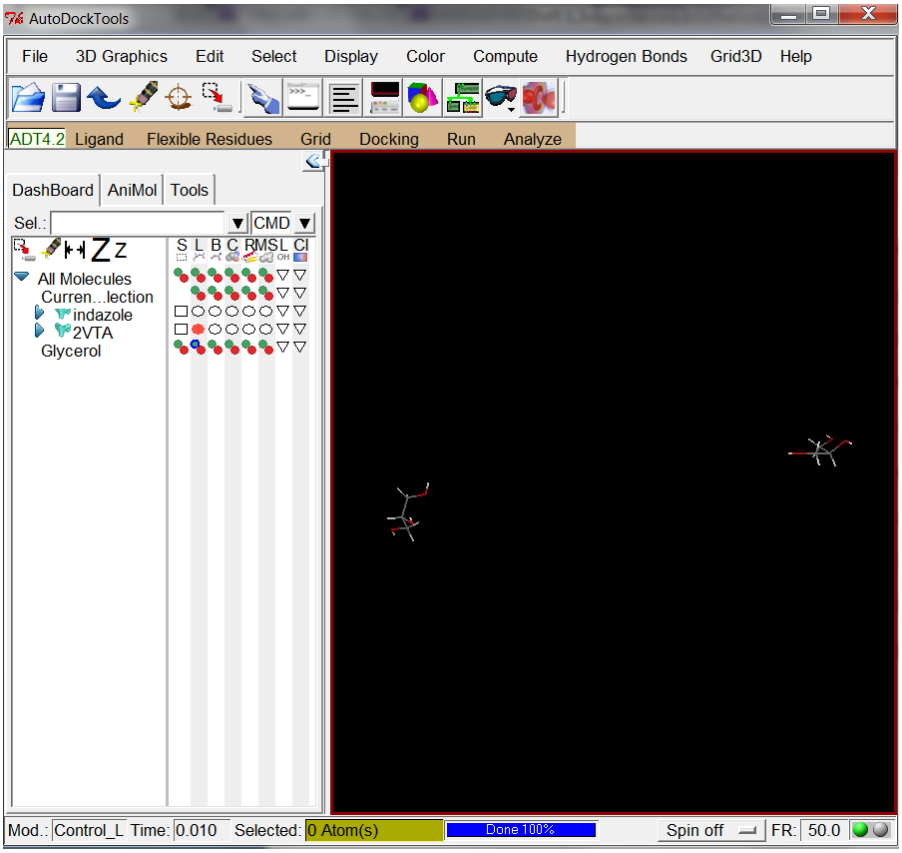


Figure 7: Display of only the glycerol molecules found bound to the protein in the crystal structure of (**1**) with CKD2 (PDB: 2VTA[9])

An easy way to find the active site is to find the bound inhibitor molecule (indazole (**1**)). Redisplay the protein as lines; open the selection tool (Select → Select from String) and enter “LZ1*” under residue and “*” under atom and click select. Pin this selection to the dashboard under the title “LZ1” as shown previously for the glycerol molecules (Select → Add selection to dashboard). You can now visualize only the glycerol molecules and the inhibitor to see that the glycerols are not near the active site (Figure 8). You can therefore leave the glycerol molecules within the protein for docking. Click the green circle under “S” next to the heading “LZ1” to select only the ligand. Make sure only 14 atoms are selected. To delete the ligand, click Edit → Delete → Delete selected atoms. Click “Continue” on the popup to delete the selection. The inhibitor will be removed from the display and the selection you pinned to the dashboard (under the heading “LZ1”) will also be removed.

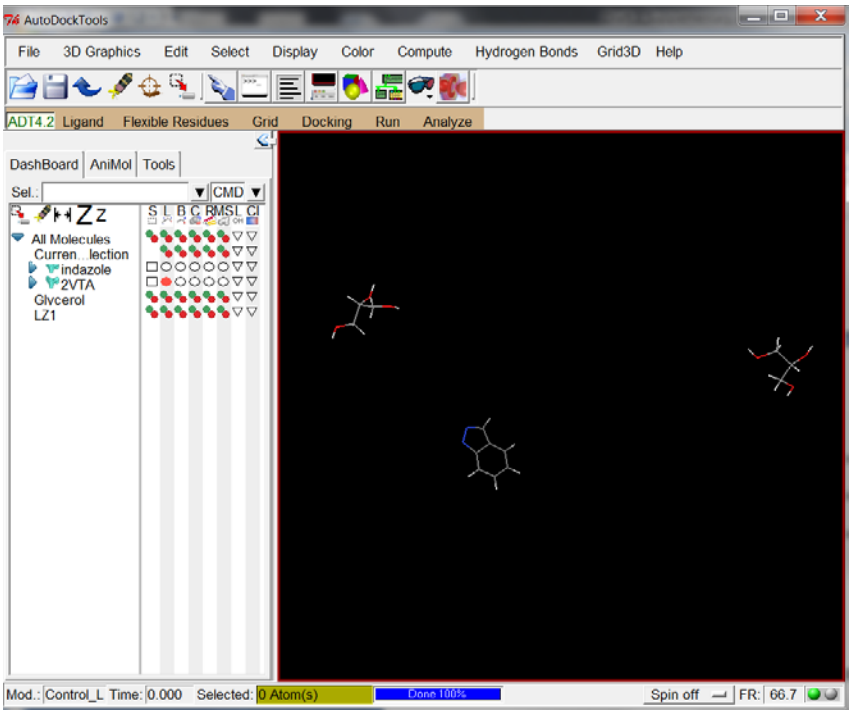


Figure 8: Display of the glycerol and ligand molecules found bound to the protein in the crystal structure of (**1**) with CKD2 (PDB: 2VTA[9]). The bound indazole must be removed prior to docking.

5. Generating Receptor Maps with AutoGrid

To dock ligands into the receptor, AutoDock uses pre-generated grids of the receptor. A grid map is prepared corresponding to each atom type found within the ligand to be docked. These maps are generated computationally, by placing atoms at lattice points within the predefined docking region and calculating the associated energy. The output of this process is a contour map describing regions of favorable and unfavorable interactions with the receptor. In addition to the production of maps corresponding to each atom within the ligand, AutoGrid calculates two additional maps: one corresponding to electrostatic interactions (labeled e.map) and one corresponding to desolvation effects (labeled d.map). The docking calculation could be performed without pre-generating these maps; however, the maps reduce the computation time significantly (see this post on the AutoDock site for more information: http://autodock.scripps.edu/wiki/AutoGrid).

Before AutoGrid can prepare the required receptor maps, a file containing information regarding the docking region, the types of atoms in the ligand, and the "density" of each map must be prepared. This file is called a grid parameter file and has the extension .gpf. First, click Grid → Macromolecule → Choose. Select the receptor (2VTA) and click "Select Molecule". After a short time, a popup will appear describing the selected macromolecule. The window will describe the number of non-bonded atoms (ideally zero), the number of non-polar hydrogen atoms found, and will note that the non-polar hydrogens (nphs) have been merged or removed from the receptor structure. Click "OK" to close the popup. You will be asked to save the receptor as a .pdbqt file. Title the file after the receptor

name or PDB code (ex: 2VTA.pdbqt) and click “Save”. The appearance of the receptor will change, as all non-polar hydrogen atoms have been removed. This process also calculates gasteiger charges for each atom (which is included in the .pdbqt file), as this information is required for both grid generation and docking.

While map types can be set manually, ADT has an incorporated function to automatically detect the atom types of a specific ligand. Click Grid → Set Map Types → Choose Ligand. In the popup, select the ligand (indazole) and click “Select Ligand”. To make sure the correct atom types have been selected, click Grid → Set Map Types → Directly. A popup entitled “AutoGpf4 Ligand” will appear. Because the structure of indazole (**1**) is simple, only a few map types are necessary to prepare grid maps. As such, only the following map types should be displayed: A, NA, HD and N, where A = aromatic carbons, NA = nitrogen atoms capable of being H-bond acceptors, HD = hydrogen atoms capable of being H-bond donors (AutoDock treats all H-atoms as donors, thus, non-polar hydrogens are merged) and N = general nitrogen atoms. If these and only these map types do not appear in the popup, change the text to read: A NA HD N. All atom types must be in caps and have a single space between each. Click “Accept” when the correct atom types are displayed.

Next, you must set the grid box parameters, which describe the region of the protein where docking will take place. However, you must first know the region of the protein containing the binding site. Therefore, you should select some residues residing within the binding pocket for easy visualization. Open the selection tool (Select → Select from string). According to the accompanying manuscript, Glu81 and Phe82 are two residues involved in substrate binding. Type “GLU81*” under Residue and “*” under Atom in the selection tool. Click “Add”. Then type “PHE82*” under Residue and “*” under Atom. Click “Add”. 22 atoms should be selected. Pin this selection to the dashboard (Select → Add selection to dashboard) under the heading “AS” for active site. Display the residues as sticks by clicking the green circle under “B” on the dashboard. The residues within the binding site are now easily discernible from the total protein structure. This region should be the center of the grid box. To set the grid box, click Grid → Grid Box. A popup will appear detailing the parameters of the grid box (Figure 9). You must make sure the grid box is large enough to encompass the entirety of the binding site. Increase the number of points in each direction (x,y and z) to 50. You can either click and drag each “wheel” to 50, or you can hover your cursor over the selector wheel, type the desired interval (50) and press enter. The dimensions of the grid box should increase within the graphical interface of ADT and the “Current Total Grid Pts per map” should read 132651. Next, you must center the grid box over the active site. Spin the wheel for each direction until the box is centered over the two residues (Glu81 and Phe82) that you highlighted earlier (Figure 10). We found the following coordinates are approximately the center of the binding site: (27.86, 0.69, 66.23). Alternatively, if the coordinates of the binding site are known, they can be entered manually by typing the appropriate value and pressing enter. To save the grid box properties, click File → Close saving current in the popup (Grid Options). The grid parameters you just set now need to be saved as a .gpf file. Click Grid → Output → Save GPF. Save the file as 2VTA.gpf. You now have all of the necessary information and files to prepare grid maps of the receptor.

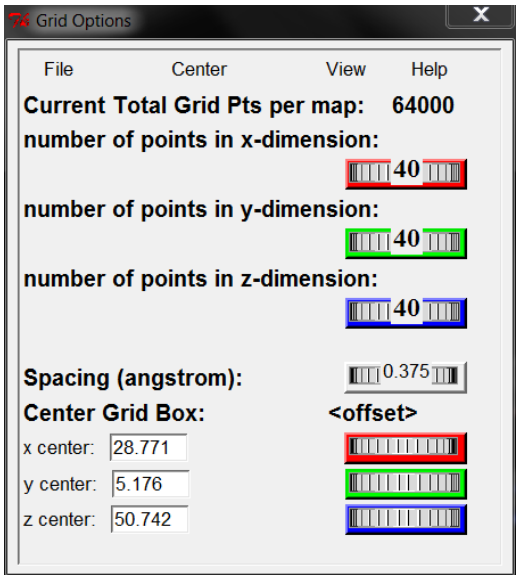


Figure 9: "Grid Options" popup in ADT used to set the grid box size and location for modeling the receptor.

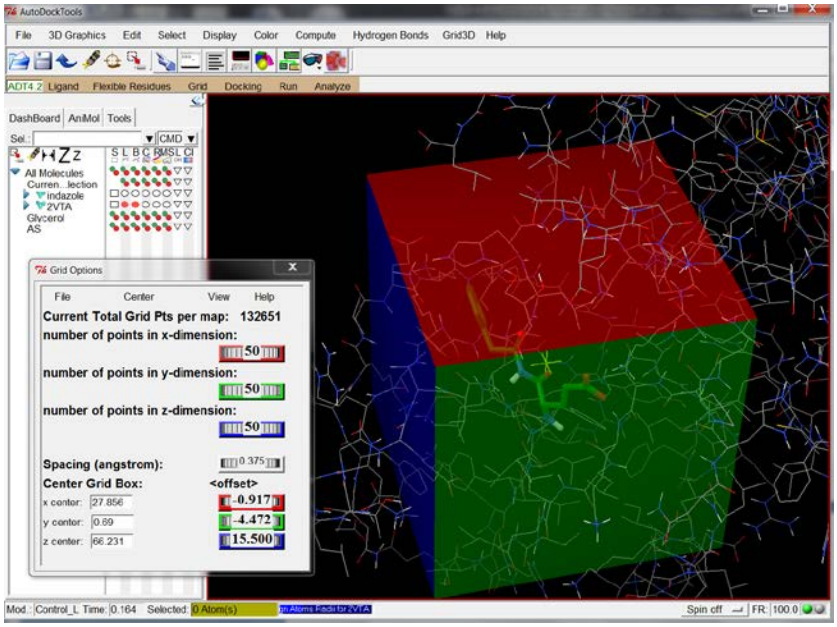


Figure 10: Demonstration of the grid parameters used for docking against CDK2. The box length should be set to 50 in all dimensions. The grid box should be centered on the active site residues involved in binding the ligand. The following coordinates should center the grid box in the correct position (27.86, 0.69, 66.23).

To run AutoGrid, click Run → Run AutoGrid. A popup will appear; you will need to set the "Program Pathname" and the "Parameter Filename". The program you will run is entitled Autogrid4.exe. This is downloaded from the AutoDock website and is included in the AutoDock 4.2.6 download described previously (the default location on Windows is Program Files (x86) → The Scripps Research

Institute → Autodock → 4.2.6; copy the two files, autodock4.exe and autogrid4.exe, into your working directory). Click "Browse" next to "Program Pathname" and find autogrid4.exe (you will ONLY find this program by the method described if it is in your working directory). Select this file. Next, set the parameter filename by clicking "Browse" and selecting the .gpf file you just prepared. The empty field next to "Log Filename" will be automatically written. Leave the rest of the options on the defaults and click "Launch". A popup entitled "Autodock Process Manager" will appear. This popup will automatically close when the process ends (about 30 seconds). If the popup appears and closes quickly (1 or 2 seconds), an error has likely occurred. You can double check the terminal that automatically opens with each session of ADT for error messages. Many errors are due to input and can easily be corrected. Some errors are more extensive and require you to close and reopen ADT to correct. When the popup closes, AutoGrid has completed the generation of receptor maps. You can double check by looking in your working directory. A number of files with the extension .map should have been written.

6. Docking Ligands with AutoDock

You now have all of the necessary files to dock the ligand into the binding site of the receptor. You now need to set the docking parameters and save them as a docking parameter file or .dpf file. To begin, you need to set the receptor. Click Docking → Macromolecule → Set Rigid Filename; select the receptor PDBQT file (2VTA.pdbqt) and click OPEN. Next, you must select the ligand file to be docked. Click Docking → Ligand → Choose; in the popup, select the ligand (indazole) and click "select ligand". A popup will appear describing the ligand according to atom types, initial ligand position in 3D space and information regarding the torsions and degrees of freedom in the ligand. Leave all options on the default settings and click "Accept". You then must select the search algorithm that is used to determine the nature and number of docking evaluations; click Docking → Search Parameters → Genetic Algorithm. A popup will appear; leave all options on the default settings and click "Accept". Next, you must set the docking run options; click Docking → Docking Parameters. A popup will appear; leave all settings on the default and click "Accept." Finally, you must save all of this information as a single docking file; click Docking → Output → Lamarckian GA(4.2); for easy cataloging, save DPF files as XXX_Docked.dpf, where XXX is the ligand title (in this case: indazole_Docked.dpf). Type the file name in the appropriate field and click "Save". You can now run AutoDock to determine the optimum docking poses for the ligand.

Click Run → Run AutoDock; a popup will appear and you will again need to set the "Program Pathname" and "Parameter Filename". Click "Browse next to "Program Pathname"; select the file titled autodock4.exe and click "Open". Next, set the parameter filename by clicking "Browse", select the docking parameter file titled indazole_Docked.dpf, and click "Open". The empty field next to "Log Filename" will automatically be written. Finally, click "Launch". A popup entitled "Autodock Process Manager" will appear. This popup will automatically close when the process ends (about 30 seconds). If the popup appears and closes quickly (1 or 2 seconds), an error has likely occurred. When the popup closes, AutoDock has completed the docking run (usually less than 10 minutes).

7. Analyzing Docking Runs

The default docking parameters used to dock indazole (**1**) produce free energy values with corresponding docking poses for 10 iterations. These can be viewed and analyzed with ADT. When you import docking results, a molecule of the ligand is also imported. To make things less complicated, we should first delete the original indazole molecule from the current ADT session. While hovering your cursor over the heading "indazole" on the dashboard, right click and select "Delete". This will remove the indazole molecule. To import the docking results, click Analyze → Dockings → Open; select "indazole_Docked.dlg" and click "Open". A popup will appear; click "OK" to close. Additionally, viewing options for the ligand molecule will appear on the dashboard under the heading "indazole". To visualize the docking results, click Analyze → Conformations → Load. A popup will appear that shows a table of the docking results for the 10 most favorable docking poses (Figure 11).

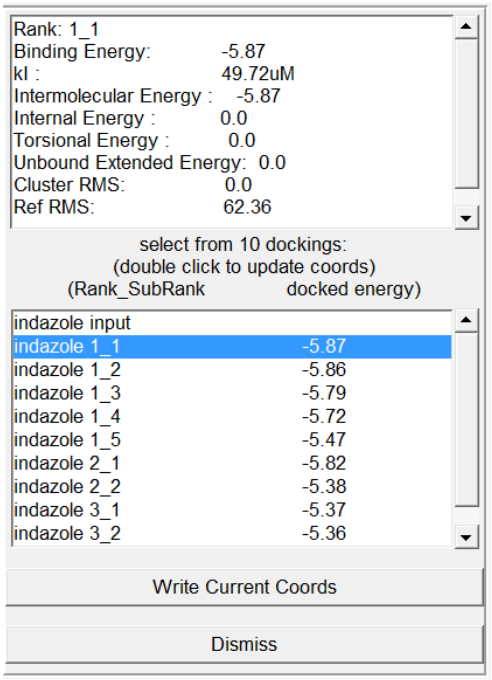


Figure 11: Example of the output docking results in ADT. The top window displays the calculated binding energy and predicted binding affinity (K_i) of the ligand to the receptor. The bottom window displays the binding energies of the ten most favorable docking poses, ranked best to worst (top to bottom).

To visualize the docking results, click Analyze → Conformations → Play. A popup will appear; the buttons allow for the cycling of the docking results in either direction. The pose number displayed at the center corresponds to the numerical entry in the table shown in Figure 11. The poses can be changed incrementally by clicking the buttons to the right and left of the pose number.

The current viewing settings make visualization of each docking pose difficult. We will therefore select the residues composing the binding site and selectively display those. Open the selection tool (Select → Select from string). The following residues form much of the ligand binding pocket and can be used to evaluate the docking results: Lys33, Phe80, Glu81, Phe82, Leu83, His84, Gln85, Asp86, Leu134, and Asp145. To select these residues, enter each in all caps into the "Residue" field, followed by "*" in the "Atom" field; click "Add" to select the residue. For example, use "LYS33*" to select the corresponding residue. When you've selected all of the residues, 10 Residue(s) should appear in green at the bottom of the ADT window. Pin this selection to the dashboard under the heading AS_2 (Select → Add selection to dashboard). Hide all of the molecules from display using the viewing options within the dashboard. View the active site as lines by clicking the green circle under "L", next to the AS_2 heading. Additionally, set the ligand to view as sticks by clicking the circle under "B", next to the heading "indazole". You can then cycle through your available docking poses until you find the pose of the lowest energy that is within the binding site. For our docking run, this was pose 2, corresponding to a predicted binding energy of -5.86 and binding constant of 51 µM; attempt to replicate Figure 12 to determine your optimum docking pose and score.

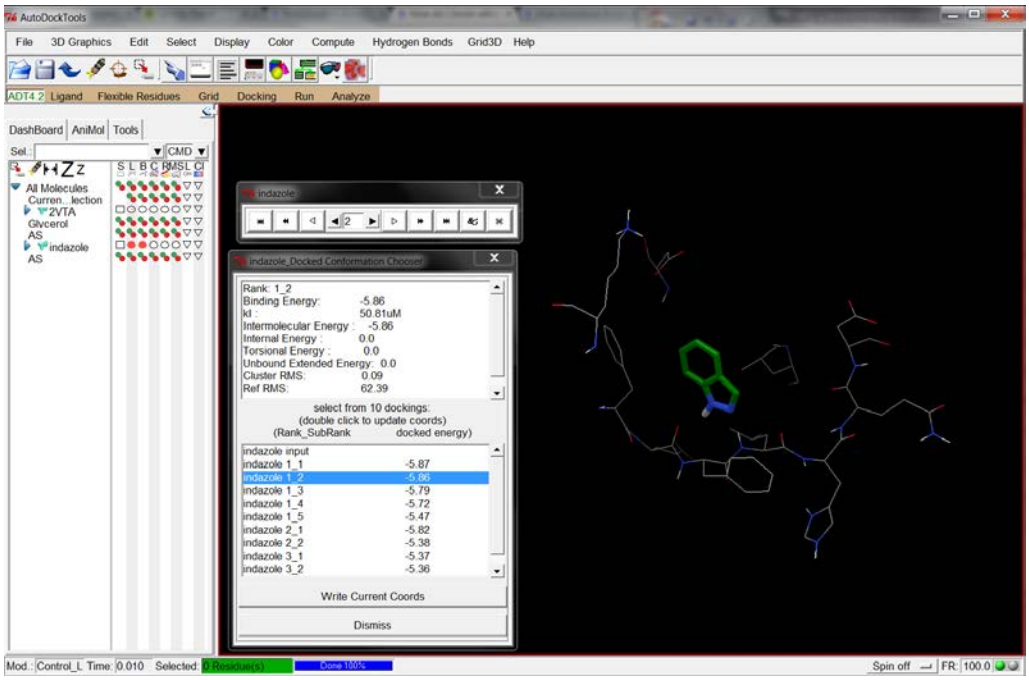


Figure 12: Best docking pose of (**1**) with CDK2. The residues lining the binding pocket are displayed.

**Note that a variety of factors contribute to the calculation of docking poses and energies. As such, duplicate docking runs will unpredictably produce different results; the AutoDock site references an error of +/- 2.5 kcal/mol for docking energies (See "How do I know which docking results are "hits"?" under the "faqs & help tab" on the AutoDock site).*

Because there is a large degree of error associated with the calculation of binding energies, docking poses are generally utilized more extensively for the determination of the best prediction of binding affinity. For example, the highest scored docking pose for our docking run is shown in Figure 13. As you can see, the ligand lies outside of the binding pocket. However, the program predicted almost exactly the same docked energy (-5.87 and -5.86 for poses 1 and 2, respectively). Clearly, pose 2 is the lowest energy pose exhibiting the predicted binding mechanism and we therefore consider the docking energy associated with this pose as the predicted binding energy for the ligand.

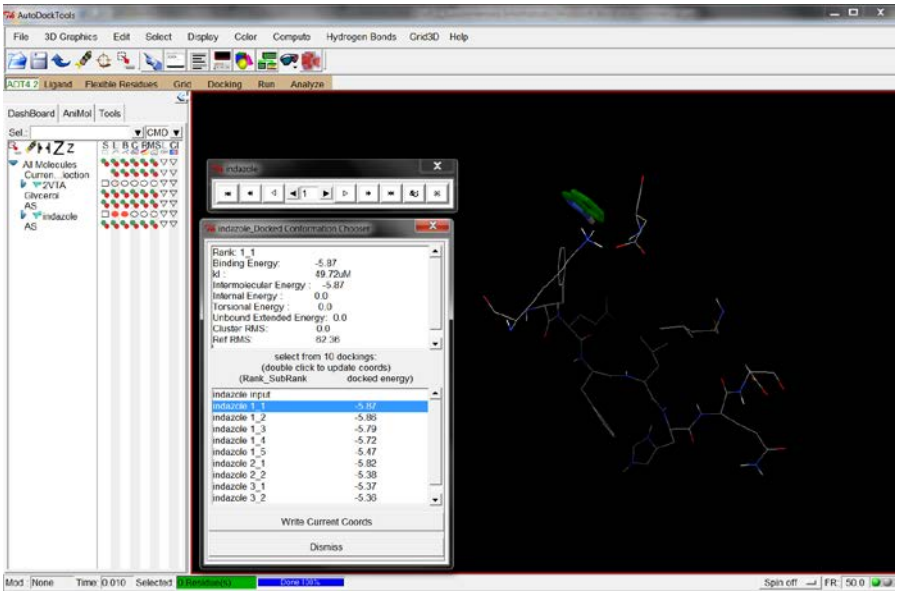


Figure 13: Example of a docking pose that does not resemble the known binding mechanism revealed by crystallography.

ADT includes a display setting for viewing molecular surfaces. This is an extremely useful setting, as you can visualize the ligand in the binding pocket and can use this information to design new compounds. To view the receptor as a surface: on the dashboard, click the circle under “MS” next to the “2VTA” entry. In ADT the representation of the receptor will change to a white surface. With your optimum docking pose selected, rotate the surface to duplicate Figure 14. As you can see, the ligand fits nicely into a deep pocket extending within the exterior of the receptor. This viewing option is also useful for the determination of acceptable binding poses. For example, docked pose 1 lies within the exterior protein surface, representing a computationally reasonable binding mechanism that is physically impossible. This demonstrates that care must be exercised when evaluating the docking output.

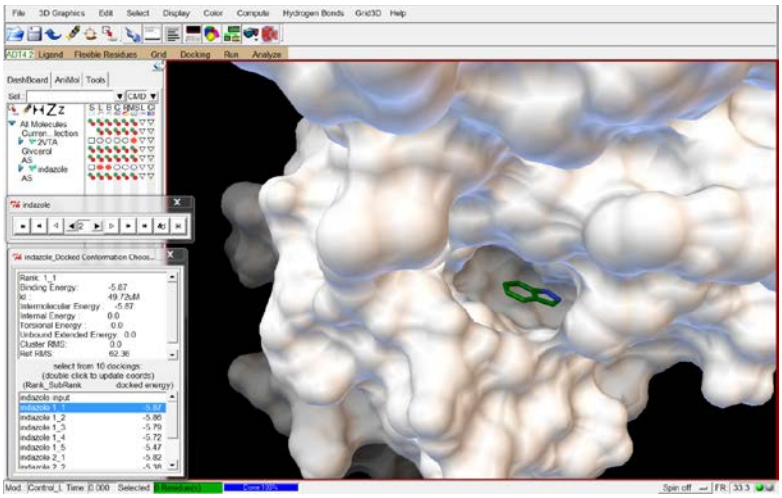


Figure 14: Docked pose of indazole (**1**) with the protein surface displayed. The protein surface display can be used to demonstrate how ligands bind within the pockets formed by protein tertiary structure.

ADT can also be used to measure bond distances between atoms of the receptor and docked ligand. First, change the display back to that of Figure 12 by hiding the molecular surface view; click the red circle under "MS" next to the "2VTA" heading. To measure bond distances, click Display → Measure → Distance; you can now hold shift and click two atoms to select. A yellow line will be drawn between the two atoms and the distance will be labeled (Figure 15). Replicate the bond distances for the two H-bonding interactions displayed in Figure 15; measure the distance between the heteroatoms involved in the H-bond, rather than from the hydrogen to the adjacent heteroatom.

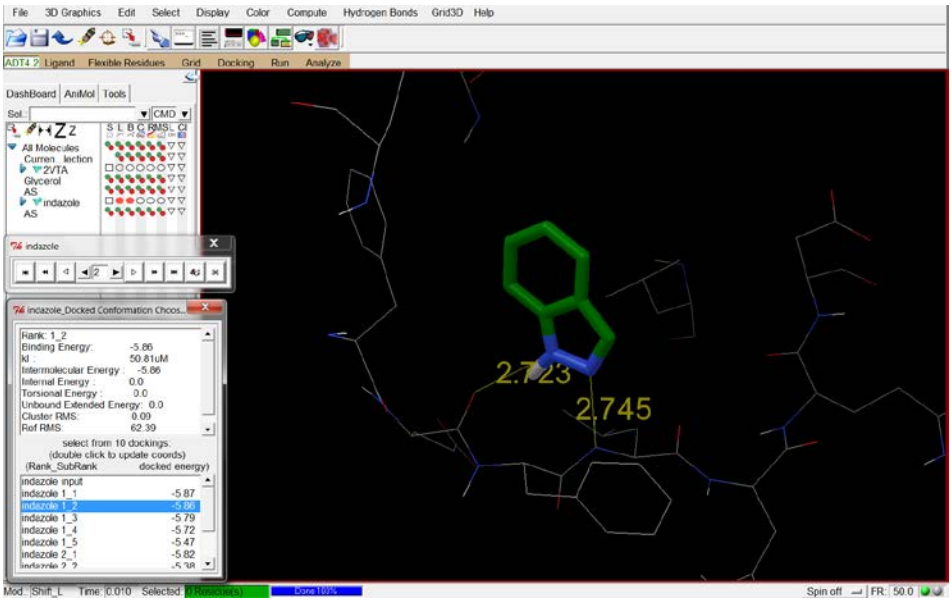


Figure 15: Docked pose of (**1**) with predicted H-bond distances between the protein and ligand displayed. Distances are measured between heteroatoms.

8. Exporting Docking Poses

You can export your most favorable docking pose as a .pdb file for image generation using Pymol or Chimera. The easiest way to accomplish this task is to reopen your docking results in a new session of ADT. You can save your current ADT session by clicking → File → Save → Current session; after saving your progress, close the ADT session. Then, open a new ADT session and import your docking results by clicking Analyze → Dockings → Open. Select indazole_docked.dlg and click "Open"; click "OK" on the popup. To cycle the docking poses, click Analyze → Conformations → Play; cycle to your most favorable docking pose. Save this docking pose as a .pdb file by clicking File → Save → Write PDB. Change the file title by clicking the field adjacent to "Filename" and scrolling all the way to the right. You'll see that the default file title is indazole.pdb. Change this to "indazole_docked_p2.pdb", where pX refers to the optimum docking pose. Click "OK". You've now generated a .pdb file of your best docking pose. This file can be read by Pymol or Chimera for the generation of high-resolution images.

**Note: This file contains structural information for the ligand only. You will need to superimpose the receptor structure over that of the docked ligand pose to view the ligand docked within the protein.*

9. Docking Additional Ligands

Now that you've docked compound (**1**), dock the three remaining indazole based inhibitors (**2**-**4**) of CDK2 (Figure 3, page 4). You can draw the compounds in ChemDraw or ChemSketch and prepare them in the same manner described for compound (**1**). Additionally, you have already prepared an acceptable receptor model; copy the file 2VTA.pdbqt into each new working directory and use this file as the receptor. The only major difference when docking the other compounds is different map types will be required by AutoDock to successfully generate docking poses. The following are the required map types for each compound: (**2**): A C NA OA N S HD; (**3**) and (**4**): A C NA OA N HD.

Post Docking Considerations

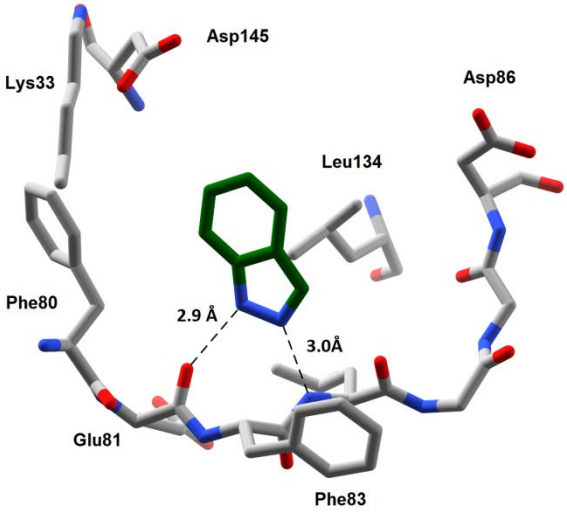
1. Before docking results can effectively be analyzed, one must have information regarding how ligands of a specific class bind to the receptor. How is this information obtained?

2. The default docking parameters for AutoDock are set to provide 10 output docking poses. For each of the 4 compounds shown in Figure 3, how many of these 10 iterations correspond to reasonable docking poses for your docking runs?

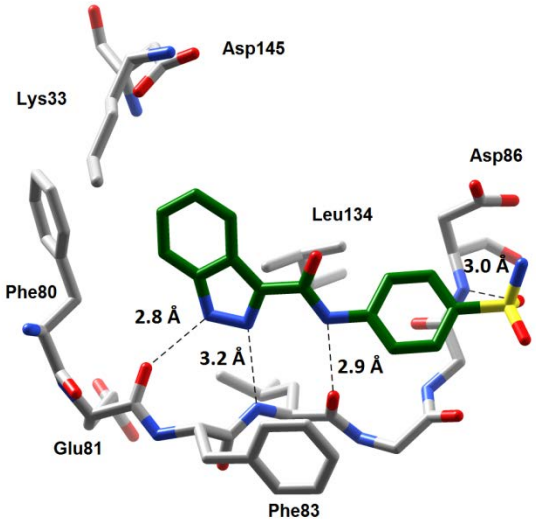
3. The actual inhibitory activities of each compound are shown in Figure 3. How do your predicted binding affinities (K_i) correlate with the experimentally determined inhibitory

activities (IC_{50})? Although you cannot directly relate the numerical values of IC_{50} and K_i, the relative activities should be the same (i.e. the lowest K_i should be for the most potent compound, which was (**3**)). How does this series of compounds demonstrate fragment growth?

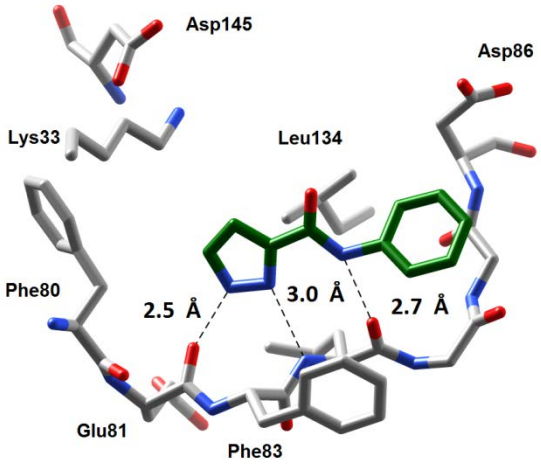
4. The experimental binding modes of compounds (**1**), (**3**), and (**4**) have been determined by X-ray crystallography and are shown below. How does your best docking pose for each of these compounds compare to the experimentally determined binding mechanism?



(**1**) (PDB: 2VTA[9])



(**3**) (PDB: 2VTI[9])



(**4**) (PDB: 2VTL[9])

5. Measure the H-bond distances between the ligand and receptor for the optimum docking pose of each of the 4 compounds. How do these distances compare to those observed for the crystalline structures of the bound inhibitors displayed in question 4? Note that compound (**3**) was not crystallized with CDK2 and the actual binding mechanism is unconfirmed.

6. Using the docking method outlined in the procedure, you used a rigid model of the receptor. However, proteins are continually changing conformation in solution. Can you think of a way to improve the docking method used in this exercise? How would your results change if the protein or some part of the protein was allowed to be flexible?

Troubleshooting

Although ADT does an excellent job of determining minimum energy docking poses of ligands against drug targets, we have encountered a number of program errors during our docking studies. Therefore, we suggest the following steps when error messages appear:

- Attempt to initiate the same process that caused the error after double checking for user input inaccuracies.

- Check the terminal for a description of the error. If the error is due to inaccurate user input, the appropriate correction can easily be made (for example; if you try to dock a ligand against a target but have not generated the correct map types, the error message will say 'I'm sorry; I can't find or open "___.X.map"', where ___ is a file name and X is an atom type. You can then rerun AutoGrid with the correct maps selected to generate the missing map type.)

- Opening many different molecules within the same ADT session can often lead to graphical errors. Try to avoid having more than one receptor and ligand open within the same session.

- If error messages continue to appear in a given ADT session, close the session and reopen an ADT session; retry the same process.

- If the display becomes distorted at any time when using ADT, place your cursor in the graphical display and type "r", "n", "c" on the keyboard to return the display to the defaults

These instructions guide you through the steps necessary to replicate the experiment described in the accompanying manuscript. Follow all steps carefully to avoid the generation of error messages. Additionally, refer to the following support materials for troubleshooting. All sites were accessed on (1/12/16):

OpenBabel:

- Official Site: http://openbabel.org/wiki/Main_Page
- FAQ's on the OpenBabel wiki: http://openbabel.org/wiki/Frequently_Asked_Questions
- Subscribe to the mailing list and report bugs here: http://openbabel.org/wiki/Help

AutoDockTools (ADT):

- Official Site: http://autodock.scripps.edu/
- ADT Manuals: http://autodock.scripps.edu/faqs-help/manual
- FAQ's on the AutoDock site: http://autodock.scripps.edu/faqs-help/faq
- How-to's for a variety of topics: http://autodock.scripps.edu/faqs-help/how-to
- Tutorials for preparing and docking ligand/receptor files: http://autodock.scripps.edu/faqs-help/tutorial

References

1. Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem.* **2004**, *47* (14), 3463-3482.

2. Joseph-McCarthy, D.; Campbell, A. J.; Kern, G.; Moustakas, D. Fragment-Based Lead Discovery and Design. *J. Chem. Inf. Model.* **2014**, *54* (3), 693-704.

3. Shang, E.; Yuan, Y.; Chen, X.; Liu, Y.; Pei, J.; Lai, L. De Novo Design of Multitarget Ligands with an Iterative Fragment-Growing Strategy. *J. Chem. Inf. Model.* **2014**, *54* (4), 1235-1241.

4. Chéron, N.; Jasty, N.; Shakhnovich, E. I., OpenGrowth: An Automated and Rational Algorithm for Finding New Protein Ligands. *J. Med. Chem.* **2015**, *59* (9), 4171-4188.

5. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A powerful approach for structure-based drug discovery. *Curr. Comput.-Aided Drug Des.* **2011**, *7* (2), 146-157.

6. Macalino, S.; Gosu, V.; Hong, S.; Choi, S. Role of computer-aided drug design in modern drug discovery. *Arch. Pharm. Sci. Res.* **2015**, *38* (9), 1686-1701.

7. Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual screening with AutoDock: theory and practice. *Expert Opin. Drug Discovery* **2010**, *5* (6), 597-607.

8. O'Boyle, N.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchinson, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, 3:33.

9. Wyatt, P. G.; Woodhead, A. J.; Berdini, V.; Boulstridge, J. A.; Carr, M. G.; Cross, D. M.; Davis, D. J.; Devine, L. A.; Early, T. R.; Feltell, R. E.; Lewis, E. J.; McMenamin, R. L.; Navarro, E. F.; O'Brien, M. A.; O'Reilly, M.; Reule, M.; Saxty, G.; Seavers, L. C. A.; Smith, D.-M.; Squires, M. S.; Trewartha, G.; Walker, M. T.; Woolford, A. J. A. Identification of N-(4-Piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a Novel Cyclin Dependent Kinase Inhibitor Using Fragment-Based X-Ray Crystallography and Structure Based Drug Design. *J. Med. Chem.* **2008**, *51* (16), 4986-4999.