ACEMFS FUT Minna Bioinformatics Workshop

Ligand Preparation and Protein-Ligand Binding

Itunoluwa Isewon PhD
Covenant University

Docking & Interactions Overview

Introduction to Molecular Docking

Molecular docking is a computational method that predicts the preferred orientation of one molecule (the **ligand**) when bound to another (the **receptor** or protein).

Think of it as a **3D lock-and-key simulation**.

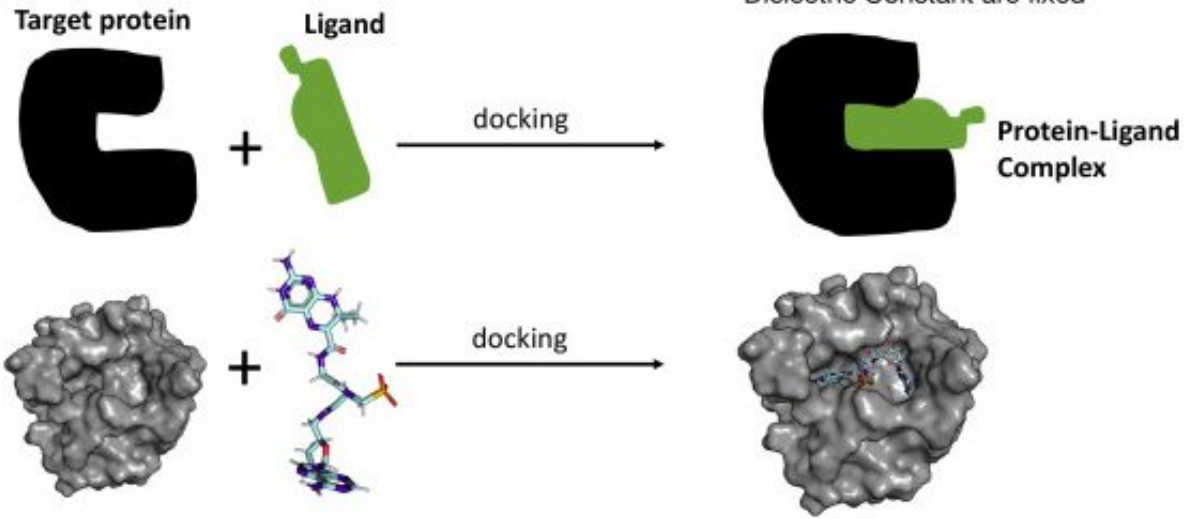
- **The Lock:** Your protein of interest (e.g., a fungal Cytochrome P450). It has a specifically shaped keyhole, the **binding site** or **active site**.
- **The Key:** Your small molecule or ligand (e.g., a fungicide, mycotoxin, or enzyme substrate).

Assumptions for Protein

- Rigid and receptive for binding
- Simple Side Chain Rotations

Assumptions for Ligand

- Size is small
- Flexibility is allowed
- Polarity, Atomic Charge, and Dielectric Constant are fixed



Advantages

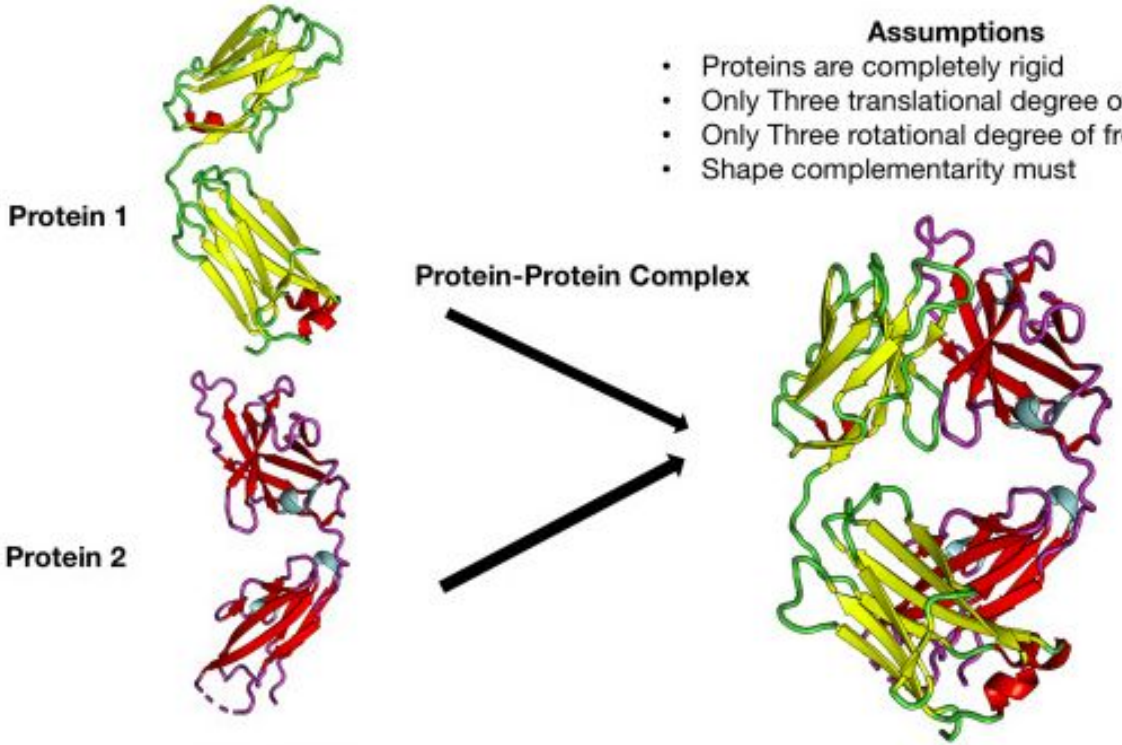
- Identification of good binders is fast
- Selection of lead based on scoring function
- Protein interactions crucial for binding
- Virtual screening of compounds

Limitations

- Accuracy depends on scoring function
- Estimation of absolute energy associated with intermolecular interaction.
- Thousands of degree of freedom
- Multiple number of conformations

Assumptions

- Proteins are completely rigid
- Only Three translational degree of freedom
- Only Three rotational degree of freedom
- Shape complementarity must



Advantages

- Understanding complexes
- Predicting Interactions that are crucial
- Development of therapeutics
- High quality docking models

Limitations

- Prediction Accuracy
- Weak or transient interaction

Relevance to Fungal/Mycotoxin Research

Mechanism of Inhibition: Visualize how an azole fungicide fits perfectly into the active site of its target, CYP51, blocking its function.

Substrate Specificity: Predict whether a novel compound produced by a fungus could be a substrate for a known enzyme in its metabolic pathway.

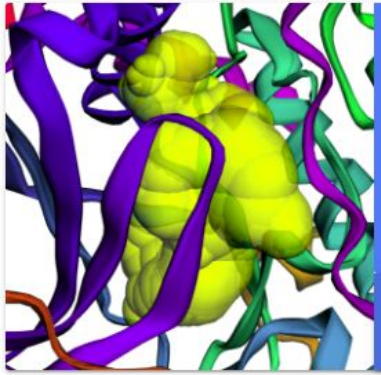
Lead Discovery: Screen a library of natural compounds to see which might be a potential inhibitor for a mycotoxin-producing enzyme.

Finding the Target

Binding Site Prediction

Key Tool: CASTp 3.0

- **What it is:** A highly-cited, web-based tool for identifying and measuring pockets and cavities on protein surfaces.
- **Link:** http://sts.bioe.uic.edu/castp/
- **Why we use it:** It's incredibly simple, fast, and provides excellent visualizations.



CASTp

Computed Atlas of Surface Topography of proteins

CASTp | Calculation | Background | Plugin | FAQ

Please cite this paper if you publish or present results using CASTp analysis: Tian et al., Nucleic Acids Res. 2018. PMID: 29860391 DOI: 10.1093/nar/gky473.

The enhanced CASTp(CASTpFold) server with new features for the whole protein universe is online, check here: CASTpFold

For questions and bugs, please contact uic.lianglab{at}gmail.com .

SHOW POCKETS | DOWNLOAD | PDB or job ID

3IGG

Novel CDK-5 inhibitors - crystal structure of inhibitor EFQ with CDK-2

DEMO 1: Finding Pockets on a Fungal Enzyme

Let's find the active site pocket of **lanosterol 14-alpha demethylase (CYP51B)** from *Aspergillus fumigatus*.

1. **Get the PDB ID:** We will use the PDB code **4UYL**.
2. Navigate to the **CASTp 3.0** website.
3. In the input bar, enter 4UYL.
4. Click "Search".
5. The job usually finishes in a minute or two. You'll be taken to the results page.

Getting Ready

Protein Preparation

- **Removing non-essential molecules:** Water molecules, buffers, or other ions are usually removed.
- **Adding hydrogen atoms:** X-ray crystallography often doesn't resolve the positions of hydrogen atoms, but they are essential for calculating interactions like hydrogen bonds.
- **Assigning atomic charges:** The docking software needs to know the partial charges on each atom to calculate electrostatic interactions.
- **Handling missing residues/loops:** Sometimes, a crystal structure has gaps. These may need to be fixed or noted.

Ligand Preparation

- **Generating a 3D structure:** You might start with just a name or a 2D drawing. This needs to be converted to an accurate 3D conformation.
- **Assigning correct protonation states and charges:** Is that amine group charged (NH3+) or neutral (NH2) at physiological pH? This is crucial.

Key Tool: SwissDock & SwissTargetPrediction

SwissDock: A web server that automates protein/ligand preparation and performs docking using a powerful engine called EADock.

- **Link:** http://www.swissdock.ch/

SwissTargetPrediction: Predicts the likely protein targets of a small molecule. Great for finding a 3D structure of a ligand.

- **Link:** http://www.swisstargetprediction.ch/

DEMO 2: Preparing our Molecules

Scenario: We want to "re-dock" the fungicide **voriconazole** back into the *A. fumigatus* CYP51B protein (4UYL) to see if the software can replicate the experimental binding pose.

1. Prepare the Protein:

- Open the PDB file 4UYL.pdb in UCSF ChimeraX.
- The structure contains the protein, the heme group (HEM), and the voriconazole ligand (VCZ). We need a "clean" protein file.
- In the ChimeraX command line, type: delete :HOH (removes water).
- Type: delete :VCZ (removes the ligand we want to dock).
- Go to File > Save and save the cleaned structure as 4UYL_protein.pdb. This is our "receptor."

DEMO 2: Preparing our Molecules

2. Prepare the Ligand:

- We need the 3D structure of voriconazole. We can get this from the original PDB file.
- Re-open the original 4UYL.pdb file in a new ChimeraX session.
- Select the voriconazole ligand. Command: select :VCZ.
- Invert the selection and delete everything else. Command: select invert; then delete sel.
- You should be left with only the voriconazole molecule.
- Go to File > Save and save it as voriconazole.mol2. This is our "ligand."

Advanced Note: Preparing molecules locally gives you more control. Tools like **AutoDock Tools (ADT)** or command-line programs like **OpenBabel** are used for this. They allow you to manually set protonation states and rotatable bonds, which can be critical for difficult docking problems.

Running a Docking Simulation

DEMO 3: Docking with SwissDock

Let's use our prepared files to run the docking simulation.

1. Navigate to the **SwissDock** website.
2. Under **"Target Selection,"** click "Choose File" and upload your 4UYL_protein.pdb file.
3. Under **"Ligand Selection,"** click "Choose File" and upload your voriconazole.mol2 file.
4. You can give your job a name, e.g., "Afum_CYP51B_Vori_Docking".
5. Enter your email address.
6. **Crucially:** Since we know the binding site (it's around the heme group), we can specify it to make the search more efficient and accurate. In the original 4UYL structure, the voriconazole was near the center. We can leave the X, Y, Z coordinates blank for a "blind" dock, or specify them if we know them. For this exercise, we will proceed with a blind dock to see if it finds the correct pocket.
7. Click **"Start Docking."**

Post-Docking Analysis in ChimeraX

The real analysis happens when you visualize the results.

1. From the SwissDock results page, download the PDB file for the **top-ranked pose** (the one with the most negative ΔG). Let's call it best_pose.pdb.
2. Open your prepared protein 4UYL_protein.pdb in ChimeraX.
3. Open the docked ligand best_pose.pdb in the same session (File > Open).
4. You will now see your ligand docked in the protein.
5. **Key questions to ask:**
    - Did it dock in the correct active site pocket? (For our re-docking, it should be right where the original was).
    - What amino acid residues is it interacting with? Use ChimeraX's selection tools to find residues within 4-5 Ångstroms of the ligand.
    - Are there key hydrogen bonds or hydrophobic interactions? (Use the FindHBond tool in ChimeraX).
    - Does the pose make chemical sense?

Fungal/Mycotoxin Use-Cases & Hands-on Exercise I

Exercise: A known mutation, G54W, in fungal CYP51 enzymes confers azole resistance. Find residue 54 in the structure. How might this bulky Tryptophan (W) residue block the binding of voriconazole compared to the small Glycine (G)? Use the structure to explain the resistance mechanism.

Scenario: You have isolated a new antifungal compound from a bacterium. Dock it to the CYP51B active site. Does it bind as well as voriconazole (compare ΔG scores)? What interactions does it make? This could suggest if it's a promising lead compound.

DEMO 4: Docking with PyRx (A More Advanced, Local Alternative)

1. Setup & Loading Molecules

- **Open PyRx.** You'll see several panes. The main ones are the *Navigator* on the left and the *3D Viewer* on the right.
- **Load the Receptor (Protein):**
  - In the *Navigator* pane, right-click on **Molecules** and select **Load Molecule**.
  - Find and open your cleaned protein file, 4UYL_protein.pdb.
  - The protein will appear in the 3D viewer.
- **Load the Ligand:**
  - In the same way, right-click **Molecules** -> **Load Molecule**.
  - Find and open your voriconazole.mol2 file. You will now see both molecules loaded under the Molecules tab.

DEMO 4: Docking with PyRx (A More Advanced, Local Alternative)

2. Molecule Preparation (Conversion)

- **AutoDock Vina** requires molecules in a specific format (.pdbqt). PyRx handles this conversion.
- **Convert the Ligand:**
  - In the *Navigator* pane, under the **Molecules** tab, right-click your ligand (voriconazole).
  - Select **Make Ligand**. PyRx will process it and move it to the **Ligands** tab.
- **Convert the Receptor:**
  - Right-click your protein (4UYL_protein).
  - Select **Make Macromolecule**. PyRx will add hydrogens, assign charges, and move it to the **Macromolecules** tab. This step confirms the protein is ready for docking.

DEMO 4: Docking with PyRx (A More Advanced, Local Alternative)

3. Defining the Search Space (The Grid Box)

- This is the most critical step. You must tell Vina *where* to search for a binding pose.
- In the *Navigator*, switch to the **Vina** tab at the very bottom. You'll see your macromolecule and ligand listed.
- Select the macromolecule (4UYL_protein). The **Grid Box** controls will appear below the 3D viewer.
- **Adjust the Grid Box:** Use the X, Y, and Z center and size controls to create a box that completely encloses the known active site (the pocket where the heme group is). Make it snug but ensure it covers the entire binding cavity.

4. Running the Docking Simulation

- Once your grid box is set, click the **"Run Vina"** button (often looks like a "Forward" or "Play" icon) at the top of the Vina tab.
- The docking process will start, running locally on your computer. For a small ligand like this, it should only take a few minutes. A progress bar will appear.

DEMO 4: Docking with PyRx (A More Advanced, Local Alternative)

5. Analyzing the Results

- When finished, PyRx automatically opens a table showing the predicted binding poses.
- **The Table:** It lists each pose (Mode 1, 2, 3...) and its corresponding **Binding Affinity (kcal/mol)**. Just like with SwissDock, this is the docking score. **More negative values are better.**
- **Visualize Poses:** Click on any row in the table. The 3D viewer will automatically update to show that specific binding pose. You can click through the top poses to see how they differ.
- The top pose (Mode 1) is Vina's best prediction for how voriconazole binds to the CYP51B enzyme. You can then save this pose and analyze the specific amino acid interactions in ChimeraX, just as you would with the SwissDock results.

Protein-Protein Interaction (PPI) Networks

PPI Networks

Proteins don't work in isolation. They form complex networks to carry out biological processes.

Mycotoxin biosynthesis, for example, is not the result of one enzyme but a whole metabolic assembly line of interacting proteins, regulated by transcription factors that also interact with other proteins.

PPI databases collect and curate evidence of these interactions from multiple sources:

- Direct experimental evidence (e.g., yeast two-hybrid, co-immunoprecipitation).
- Co-expression data (genes that are always turned on/off together).
- Genomic context (genes that are neighbors in the genome).
- Scientific literature (text mining).

Visualizing these networks helps us understand the functional context of our protein of interest.

Key Tool: STRING Database

- **What it is:** A comprehensive, user-friendly database for known and predicted protein-protein interactions.
- **Link:** https://string-db.org/
- **Why we use it:** It's the gold standard. It provides a confidence score for every interaction and links back to the original evidence.

DEMO 5: Building a Fungal PPI Network

Scenario: Let's explore the interaction network of **AflR**, the master transcriptional regulator of the aflatoxin biosynthesis cluster in *Aspergillus flavus*.

1. Navigate to the **STRING** website.
2. In the "Protein Name" search box, type AflR.
3. In the "Organism" box, start typing Aspergillus flavus and select it from the list. Click "Search."
4. STRING will identify the correct protein and take you to the network page.

Fungal/Mycotoxin Use-Cases & Hands-on Exercise II

Exercise: Search for one of the enzymes from the aflatoxin pathway (e.g., Ver-1) in STRING for *Aspergillus flavus*. What does its network look like? How does it connect back to AflR?

Scenario: You have a new, uncharacterized protein from a fungal genome. Running it through STRING might show that it interacts with a known pathway (e.g., ergosterol biosynthesis). This provides an instant, powerful clue to its potential function!

DEMO 6: Building a Fungal PPI Network with BioGRID

1. Searching the Database

- Navigate to the **BioGRID** website.
- In the main search bar, type Hsp90.
- Beside the search bar, click on **"All Organisms"** and start typing *Candida albicans*. Select it from the list.
- Click the **"Search"** button.

DEMO 6: Building a Fungal PPI Network with BioGRID

2. Interpreting the Results Page

- BioGRID will return a list of all proteins matching your search term. Find the correct one (often HSP90). Click on its official Gene Name.
- This takes you to the main interaction page for that protein. You'll see a table where **each row is a single, documented interaction.**

DEMO 6: Building a Fungal PPI Network with BioGRID

3. Understanding the Interaction Data

- The results table is very detailed. Here are the key columns:
  - **Interactor A/B:** The two proteins that are interacting.
  - **Experimental System:** The type of experiment used to detect the interaction (e.g., "Two-hybrid," "Affinity Capture-MS"). This tells you the *type* of evidence.
  - **Source:** A direct link to the PubMed ID of the publication where this interaction was reported.
- This is BioGRID's power: every claimed interaction is tied directly to a specific experiment in a peer-reviewed paper.

DEMO 6: Building a Fungal PPI Network with BioGRID

4. Exporting Data for Visualization

- Above the results table, look for an "**Export**" or "**Download**" button.
- You can download the interaction data in various formats. For visualization in Cytoscape, a simple **Tab-delimited format (TAB2)** is excellent.
- Download the file. It will be a simple text file with columns detailing the interacting proteins and the evidence.

DEMO 7: Visualization with Cytoscape

1. Importing the Network Data

- **Open Cytoscape.**
- From the main menu, go to File > Import > Network from File...
- Select the TAB2 file you downloaded from BioGRID.
- **The Import Dialog:** Cytoscape is smart and will try to guess which columns contain the interacting proteins.
    - Designate the column with the first protein as the **"Source Node"**.
    - Designate the column with the second protein as the **"Target Node"**.
    - You can ignore other columns for now or import them as "Edge Attributes." For instance, you could import the "Experimental System" column as an attribute to know how each interaction was found.
- Click **"OK."** You will see a raw, messy-looking network.

DEMO 7: Visualization with Cytoscape

2. Applying a Layout

- The initial network is a "hairball." We need a layout algorithm to arrange the nodes logically.
- Go to the menu Layout > Prefuse Force Directed Layout > (default).
- This will automatically rearrange the nodes so that connected proteins are pulled closer together and the network structure becomes clear.

DEMO 7: Visualization with Cytoscape

3. Customizing Styles (Making it Pretty & Informative)

- The **"Style"** panel is Cytoscape's control center for visuals.
- **Node Styles:**
  - **Fill Color:** You can change the color of the nodes. For example, you can select your main protein (Hsp90) and make it a different color (e.g., red) to make it stand out.
  - **Label:** Change the "Label" property to use the "name" column, so you can see the names of the proteins on the nodes.
  - **Size:** You can map the size of a node to a data value. For example, nodes with more connections (higher degree) could be made larger.
- **Edge Styles:**
  - **Stroke Color/Width:** You can change the color or thickness of the lines connecting the proteins. You could, for example, make the lines representing high-confidence interactions thicker.

DEMO 7: Visualization with Cytoscape

4. Final Interpretation

- After applying layouts and styles, you have a clear, interpretable map of your protein's neighborhood.
- You can immediately see which proteins are "hubs" (highly connected) and identify functional modules or clusters. For our Hsp90 example, you would see it connected to various client proteins and co-chaperones, visually representing its role as a central hub in protein folding and stability.