BT 305 Lab session 9 and 10

BT 305 Lab Session 9: Docking

Docking is a method which predicts the preferred orientation of one molecule with respect to a second when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using **for example-** scoring functions. Scoring functions are fast approximate mathematical methods used to predict the strength of the non-covalent interaction (also referred to as binding affinity) between two molecules after they have been docked.

The associations between biologically relevant molecules such as proteins, nucleic acids, carbohydrates, and lipids play a central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (e.g., agonism vs antagonism). Therefore docking is useful for predicting both the strength and type of signal produced. Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs.

Exercise 3 : Docking:

Overview: This exercise is a learning procedure for studying the binding interactions of different drugs to the same target and is essential for the identification of a lead compound in the process of drug discovery.

Steps:

1. Open following link, in your web browser : http://bioinfo3d.cs.tau.ac.il/PatchDock/
2. Browse to select the ligand from Ex_2 folder in pdb format and write one of the following target pdb IDs for their respective drugs: 1J2E (Dipeptidyl peptidase IV), 2HJW (Acetyl CoA carboxylase II) , 2QLY (alpha-glucosidase).
3. Leave all settings to default. Click submit form and download result files.
4. Repeat the docking process for all targets and their corresponding drugs.
5. Analyze the results and try to predict the most suitable drug against the given targets.

OPENBABEL

Introduction

Open Babel is a chemical toolbox designed to speak the many languages of chemical data. It's an open, collaborative project allowing anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas.

Fingerprinting- screening and similarity

Fingerprint is an abstract representation of certain structural features of a molecule. The most common type of fingerprint is a series of binary digits (bits) that represent the presence or absence of particular substructures in the molecule. Comparing fingerprints allows determining the similarity between two molecules, to find matches to a query substructure, etc.

1. Structural keys: Structural keys were the first type of screen employed for high-speed screening of chemical databases. A structural key is usually represented as a boolean array, an array in which each element is TRUE or FALSE. Boolean arrays in turn are usually represented as bitmaps.

To make a structural key, one decides which structural features (patterns) are important, assigns a part of the bitmap to each and then generates a complete bitmap for each molecule in the database. Generating a structural key is time-consuming; each time we have to do a substructure search for each pattern represented in the bitmap and repeat this for each molecule present in the database.

Example of patterns:

- The presence/absence of each element, or if an element is common (nitrogen, for example), several bits might represent "at least 1 N", "at least 2 N", "at least 4 N", and so forth.
- Unusual or important electronic configurations, such as "sp^3 carbon" or "triple-bonded nitrogen."
- Aromatic rings and cyclic structures.
- Common functional groups, such as alcohols, amines, hydrocarbons, and so forth.

The structural keys suffer from a lack of generality. The choice of patterns included in the key has a critical effect on the search speed across the database

2. Fingerprints: Fingerprints address this lack of generality by eliminating the idea of pre-defined patterns. A fingerprint is a boolean array, or bitmap, but unlike a structural key there is no assigned meaning to each bit. Once own fingerprint is very characteristic of oneself, yet there is no meaning to any particular feature. Similarly, a pattern's fingerprint characterizes the pattern, but the meaning of any particular bit is not well defined.

For example, the molecule **OC=CN** would generate the following patterns:

<table>
<tr><td>0-bond paths:</td><td>C</td><td>O</td><td>N</td></tr>
<tr><td>1-bond paths:</td><td>O-C</td><td>C=C</td><td>C=N</td></tr>
<tr><td>2-bond paths:</td><td colspan="3">O-C=C C=C=N</td></tr>
<tr><td>3-bond paths:</td><td colspan="3">O-C=C=N</td></tr>
</table>

Jaccard / Tanimoto's Coefficient for similarity and distance

Jaccard / Tanimoto Coefficient is a statistic used for comparing the similarity and diversity of sample sets. The similarity ratio of Tanimoto coefficient,is equivalent to Jaccard similarity, but the distance function is not the same as Jaccard Distance. The Jaccard distance, which measures dissimilarity between sample sets, is complementary to the Jaccard coefficient and is obtained by subtracting the Jaccard coefficient from 1

For similarity, Tanimoto's Coefficient (T_C) uses the ratio of the intersecting set to the union set as the measure of similarity. Represented as a mathematical equation:

where, N represents the number of attributes in each object (a,b). C in this case is the intersection set.

For distance, Tanimoto coefficient (T_d)

Lab session 10: Exercise 1: File conversion

Overview: This exercise explains how diferent file formats of chemical structures can be inter-converted.

Steps:

1. Download mol and SMILES files from Therapeutic Target Database (TTD)
    - Download the complete database in sdf format from following link:

      http://bidd.nus.edu.sg/group/ttd/filedownload.asp?file=All.sdf
    - Type targets "Dipeptidyl Peptidase IV", "Acetyl CoA Carboxylase 2" and "Alpha Glucosidase" in target search field (Third from the top).
    - Open the Target info link from the Left column for your target of interest (e.g. Dipeptidyl peptidase IV) in the results page.
    - Browse the drugs for the target and select one drug of your own choice, for example: Saxagliptin, Sitagliptin, Vildagliptin etc.
    - Make a separate folder for each target and save the 3D structure of the drug in the SDF file format (.sdf) and also their Canonical SMILES string (*.smi) in text file accordingly.
2. Open terminal, make a directory with the name "QIP" and navigate to this folder by using the following commands:

   mkdir QIP (enter)

   cd QIP (enter)
3. Make a working directory with the name Ex_1 with the following command:

   mkdir Ex_1 (enter)
4. Copy-paste the downloaded files (sdf and SMILES) to the Ex_1 folder and access it with the following command:

   cd Ex_1 (enter)
5. Convert smiles format (*.smi) to mol format(*.sdf) by writing the following command in the terminal, which converts the input .smi file to an output SDF file named out.sdf:

   babel -ismi name.smi -osdf out.sdf (enter)
6. View the output files by typing "**gedit** outputfilename.sdf (enter)" in the terminal and compare the output SDF file with the respective downloaded SDF files. What do you observe?
7. Convert the mol format (*.sdf) to fingerprint format(*.fpt) by writing the following command in the terminal, which converts the input SDF file to an output FPT file named out.fpt :

   babel -isdf name.sdf -ofpt out.fpt (enter)
8. Now convert the mol format (*.sdf) to pdb format(*.pdb) by writing the following command in the terminal, which converts the input SDF file to an output pdb file named out.pdb

   babel -sdf name.sdf -opdb out.pdb (enter)

Exercise 2: Tanimoto's coefficient calculation:

Overview: This exercise aims at calculation of structural similarity for identification of the three most similar drugs for the drug of your choice.

Steps:

1. Go back to the previous folder(QIP), create a new folder- Ex_2 and access this folder with the following set of commands:

   cd .. (enter)

   mkdir Ex_2 (enter)

   cd Ex_2 (enter)

2. Copy the contents of all the downloaded SDF files in one file and save it as "all.sdf" in the same folder Ex_2. Also, copy the downloaded SDF files to this folder.
3. Select one of the downloaded SDF files as subject say "test1.sdf".
4. Calculate Tanimoto's coefficient (T_c) by writing the following command in the terminal:

   a. First convert "all.sdf" to "all.fs" by following command:

   obabel all.sdf -ofs -o all.fs (enter)

   b. Calculate Tanimoto's coefficient (T_c) for identifying the 3 most similar structures to your query structure by following command:

   obabel all.fs -s test1.sdf -ofpt -at 3 -O test1.fpt (enter)

   c. Get their corresponding structure file by using following command:

   obabel all.fs -s test1.sdf -osdf -at 3 -O testing1.sdf (enter)

   d. "testing1.sdf" file will have 3 different mol files. Save these mol files in three different sdf files.

   e. Convert these sdf files to pdb file format as mentioned in step exercise 1 (step 8) and use them for the next exercise.