

Developing A Tool For Interactive *In Silico* Analysis of Medicinal Plant Extracts From In House Medicinal Plant Database

A Thesis Submitted in Partial Fulfillment

of the Requirement for the Degree of

Bachelor of Technology

Biotechnology

By

Siddharth Biswal

108BT008

Under guidance of

Prof. / Dr. Bibhukalyan Prasad Nayak



**Department of Biotechnology and Medical Engineering
National Institute of Technology, Rourkela**



National Institute of Technology Rourkela Certificate

This is to certify that the thesis entitled, “Developing a tool for interactive *in silico* analysis of medicinal plant extracts from in house medicinal plant database” submitted by Siddharth Biswal in partial fulfillment of the requirement for the award of Bachelor of technology degree in Biotechnology Engineering at National Institute of Technology, Rourkela is an authentic work carried out by him under my supervision and guidance. To the best of my knowledge the matter embodied in the thesis has not been submitted to any other University/Institute for award of any Degree/Diploma.

Date:

Prof. / Dr. B.P Nayak

Dept. of Biotechnology and Medical Engineering

National Institute of Technology, Rourkela

ACKNOWLEDGEMENT

I wish to express my sincere thanks and gratitude to Prof. B.P Nayak, Department of Biotechnology and Medical engineering, National Institute of Technology, Rourkela for invaluable guidance and inspiration during each step of this project. With his guidance and help this project has come to a successful completion.

I would like to express my thanks to everybody associated with Department of biotechnology and medical engineering for inculcating an interest in me for Biotechnology. I would also like to express my heartily thanks to my batch mates and others who helped me a lot during this project

This project has been a great learning experience for me and it will definitely help me in my future endeavors. Finally I pay regards to my parents for being my internal motivation and source of energy.

Siddharth Biswal

Contents

Chapters	Title	Page No.
	Abstract	
1	Introduction	1
	1.1 Objective	3
2	Literature review	
	2.1 Pubchem	5
	2.2 Drugbank	6
	2.3 Ligand	7
	2.4 Superdrug	8
3	Materials and methods	
	3.1.1 MySQL	11
	3.1.2 PHP	11
	3.1.3 Python	13
	3.1.4 Pybel	13
	3.1.5 SMILE format	14
	3.1.6 HEX server for docking	15
	3.2 Methods	
	3.2.1 Similarity calculation between two small molecules	16
	3.2.2 Molecular weight calculation for small	17

	molecules	
	3.2.4 Conversion of different	17
	molecular formats	
4	Results	
	4.1 Similarity score tool	20
	4.2 Molecular weight tool	21
	4.3 Conversion tool	22
	4.4 Validation of application	23
	by docking	
5	Conclusion and future work	29
	References	30
	Appendix	32
	Similarity score database	

Abstract

Bioinformatics plays key role in creating useful information from raw biological data. So in this project a bioinformatics tool has been designed and linked to the in house developed database for analysis of medicinal plant extracts. Various small molecules i.e. alkaloids, flavonoid, glycosides are the main extracts from plants which are widely used as established therapeutics for an array of human diseases. The current work has focused mainly on those molecules. So we have designed an application which is capable of finding similarity among two small molecules based on their structure. This similarity tool combined with other tools such as molecular format conversion tool are designed to make the research process easy for end user. Finally we have created a tool for automated docking of selected similar molecules to a protein of interest. This process would identify new drug molecules. In addition, the target protein of interest can be sent for homology modeling directly from the application to get proteins with similar 3D structure and folding. Various permutations and combinations can be applied between ligand (small molecules) and the whole range of proteins. In nut a shell, the application utilizes a versatile algorithm for discovering newer ligands as well as newer target proteins to intervene various pathways leading to disease.

This application would definitely help the researchers to a great extent in finding new small molecules since the need of similar molecule finding is of great importance in drug discovery process.

Key Words: Bioinformatics, SMILE format, Homology modeling, docking, ligands

List of Figures

Figure No	Figure name	Page No
2.1	Pubchem Database	6
2.2	Drugbank database of chemical compounds and pathways	7
3.1	A compound represented in SMILE format	15
4.1	Similarity toolbox	20
4.2	Result page of similarity toolbox Similarity toolbox	20
4.3	Molecular weight toolbox	21
4.4	Result page of molecular weight tool	21
4.5	Conversion tool	22
4.6	Python server running the conversion tool	22
4.7	Curcumin & Piceatannol in 3D view	23
4.8	Docking of Piceatannol with COX-2	25
4.9	Docked Curcumin with COX-2	26

4.10	Homology modeling of COX-2 protein	27
4.11	Docking of Curcumin with new homology modeled protein	27

Chapter 1

Introduction

Chapter 1

Introduction:

In the post genomic era there has been a huge eruption of data from various sources. Starting from genomic analysis, gene sequencing, microarray experiments, and protein structures from experiments there is huge creation of data daily in the labs around the world. These data are needed to be processed to create truly useful information. After all these information are supposed to discover new drugs and therapeutic materials which can be used for humans. But the problem is that all the information extracted from DNA and genes do not have information about chemical compounds which are needed to be used as drugs against some pathway or some receptor. These small molecules are very important for curing many diseases as they are very useful to act against many targets. So information about those chemical compounds is also very important and they need to be presented to the scientific world in a proper manner.

Information about small molecules or ligands which act in a key and lock pattern to inhibit the action of various proteins are becoming more and more valuable. Since the drug discovery trend is quite highly increasing these days, competition is increasing to find new novel compounds to be used as drug.

So these days there is increasing activities among different scientific labs to collect data about chemical compounds and make database. This new trend is called “chemical genomics” emerging these days.(Lipinski and Hopkins 2004)

Few example databases are Pubchem (<http://pubchem.ncbi.nlm.nih.gov/>), KEGG/LIGAND (<http://www.genome.jp/kegg/ligand.html>), ChEBI (<http://www.ebi.ac.uk/chebi/>) have been developed as databases about chemical compounds.

There are many types of use of this information obtained about chemical compounds. This will enable selection of compounds for probing target pathways, finding side effect of old compound or suggest new compounds, assessment of ADMET (absorption, distribution, metabolism, excretion, toxicity) properties of drugs, quantification of structure activity relationships for small compounds. This information can be used for defining many new pathway models which can be used for prediction works.

These ligand databases are useful for screening a lot of compounds for creating a novel compound for its use as drug. If a researcher has some information about what type of small compound is used as inhibitor then he can use that information to search for other similar molecules which can be used as drug or inhibitor. But to date there is no exhaustive database which allows the search for similar compounds for a small molecule.

1.1 Objective:

- A similarity finding tool has to be designed for finding similarity score between small molecules using their SMILE string. The similarity score is based on Tanimoto coefficient which is calculated using molecular fingerprints.
- A tool box for showing the molecular weight of the compounds has to be created.
- A conversion tool box which will convert different format of molecule to another format has to be designed. This is created using Pybel which is python module of OpenBabel API.

This will take user input which will be a molecule in some definite format, then that will be converted to another format using this tool. User will have option to save the molecule in the new format in their local disk.

- Finally a tool has to be designed for giving users an option to get the molecule which is similar to another molecule and do docking with an inhibitory protein which binds to the first protein. This will give an easy method to find alternative ligand tool for users.
- An option is also given to the user to go to homology modeling tools and find a similar protein structure with the inhibitory protein for which they will be able to do docking with similar ligands.

Chapter 2

Literature review

Chapter 3

Literature review

There are many databases available for information regarding chemical compounds and their various features.

2.1 Pubchem (<http://pubchem.ncbi.nlm.nih.gov/>):

It is a repository of small molecules detailing their structure and activities. The Pubchem project is created and maintained by National center of biotechnological information which is a part of National institute of health. Pubchem contains three main databases. Substance database (primary accession-SID) contains contributed sample descriptions provided by depositors, whereas the Compound database (primary accession-CID) contains unique chemical structures derived from the substance depositions.(Wang, Xiao et al. 2009)

There are around 70 million substances in the substances database in Pubchem which contains mixtures, extracts, complexes provided by depositors. In the compound database of Pubchem there are around 30 million compounds deposited. The bioassay database contains bioactivity screens of substances deposited in the other substance database.

After searching for a compound in Pubchem it shows the summary of compounds in which a figure of the compound is also shown. Then its source is listed, its compound id, substance ids are also given.

Pubchem gives option to download the structure in sdf, xml and ASN format.

The screenshot shows the PubChem search interface. At the top, the search bar contains 'abraxane' and a 'Search' button. Below the search bar, there are links for 'Save search', 'Limits', and 'Advanced'. The main content area displays 'Results: 2' and lists two entries. The first entry is 'TAXOL: paclitaxel: Paxceed ...' with a molecular structure icon, MW: 853.906140 g/mol, MF: C₄₇H₅₁NO₁₄, CID: 36314, and links for 'Active in 96 BioAssays', 'Tested in 440 BioAssays', 'Similar Compounds (MeSH Keyword)', 'Same Parent Connectivity', 'Mixture/Component Compounds', 'BioAssays_activity ≤ 1 μM', and 'PubMed'. The second entry is 'paclitaxel: Abraxane (TN): S1150_Selleck ...' with a molecular structure icon, MW: 853.906140 g/mol, MF: C₄₇H₅₁NO₁₄, CID: 441276, and links for 'Active in 1536 BioAssays', 'Tested in 2789 BioAssays', 'Similar Compounds (MeSH Keyword)', 'Same Parent Connectivity', 'Mixture/Component Compounds', 'BioAssays_activity ≤ 1 μM', and 'PubMed'. On the right side, there are sections for 'Filters: Manage Filters', 'Actions on your results' (including BioActivity Analysis, Structure Clustering, Structure Download, and Pathways), and 'Refine your results' (including BioActivity Experiments and BioMedical Annotation).

Figure 1.1

Pubchem Database

2.2. Drug bank: DrugBank is a diverse bio and cheminformatics database quantitative, molecular- information about drugs and drug targets molecules. Drug bank database contains the information which is designed by mixing the information about molecular biology information from Swiss-prot, NCBI, from different chemistry text books. It has around more than 4100 drug molecules matching to more than 12,000 different brand name and synonym.(Wishart, Knox et al. 2006)

To create and curate this DrugBank database many textbooks, journal articles, more than 30 electronic databases and around 20 in house web based programs were searched, compared to compile all the data in the database. It had taken more than four years,

Now it is a completely searchable resource available on the internet with many in built tools and features for accessing information, sorting, searching those data.


Identification	
Name	Vincristine
Accession Number	DB00541 (APRD00495)
Type	small molecule
Groups	approved
Description	Antitumor alkaloid isolated from Vinca Rosea. (Merck, 11th ed.)
Structure	 <p>Download: MOL SDF SMILES InChI Display: 2D Structure 3D Structure</p>
Synonyms	<ul style="list-style-type: none"> • 22-Oxovincal leukoblastine • Indole alkaloid • LCR • Leurocristine • VCR • VIN • Vincristina [DCIT] • Vincristine Sulfate • Vincristinum [INN-Latin] • Vincristine • Vincristine • Z-D-Val-Lys(Z)-OH
	22-Oxovincal leukoblastine Indole alkaloid LCR Leurocristine

Figure 1.2

DrugBank database of chemical compounds and pathways

2.3 Ligand: It is a database and chemical compounds and pathways that occur in biological reactions. This database is mainly composed of three main database or sections namely “compound” section for information about metabolites and small molecule chemical compounds.”Reaction” section for collection of substrate and product reactions representing metabolic and other reactions. There is another section “Enzyme” which contains information about many different enzyme molecules.(Goto, Okuno et al. 2002)

In this database the COMPOUND and ENZYME sections are based on flat-files for information storing and the data format of each section is equivalent to that of GenBank. Compound and reaction sections are managed as MDB and RXN formats.

The COMPOUND section of this LIGAND database was originally made by extracting different chemical compounds from the metabolic pathways of the KEGG, PATHWAY

database. They are also trying to add many drug related information to the chemicals. This database can be accessed at <http://www.genome.ad.jp/ligand/>.

2.4 Superdrug: This is a database designed for essential marketed ugs. This contains around 2500 structure of active ingredients of many important drugs. In this they have given the three dimensional structures of the compounds. For structural flexibility there are around 10^5 structural conformers given in the database.(Goede, Dunkel et al. 2005)

In this database there is facility for searching which queries for drug name, synonyms, brand names, trivial name, CAS number. There is a similarity based searching facility in this database. The results can be obtained in the molfile format in graphical user interface.

This database has been designed by collecting information from various sources for example the CAS numbers are obtained from “The Chemical Abstracts (<http://www.cas.org>)”

Chapter 3

Materials and Methods

Chapter 3

Materials and Methods

Technologies used:

Back-end platforms:

1. PHP 5.3.8
2. MySQL 5.5.16
3. Python 2.7
4. Pybel module of Open Babel
5. XAMPP (Apache, MySQL, PHP)

Front-end platforms:

1. HTML
2. CSS

To build a web application there is requirement of few building blocks. First is the web server. Server is the software which delivers web pages after getting a request from client and return a web page using many different protocols.

In this project Apache web server has been used as sever. Basically Apache is the software package which runs the server on a local-host and serves web pages after getting a request. It uses HTTP protocol to transfer web pages on request. HTTP protocol is Hypertext transfer protocol which is a method to deliver web pages.

To use the Apache web server XAMPP has been used. XAMPP is a software bundle which consists of Apache, MySQL and interpreters for PHP, PERL and scripting languages. This XAMPP is a free software distributed under GNU general public license. After starting

XAMPP it runs a Apache web server on the localhost. With this MySQL and FileZilla (FTP) also comes bundled, so they also start working. With XAMPP PHPMyAdmin also gets installed which is an web interface to manage the MySQL database.

MySQL: It is one of the best relational database management systems available. Using MySQL we can store various types of data such as integers, boolean, text, images, URL, variable character etc. it uses structural query language for doing query on the database. MySQL is written in C and C++. It runs as a server providing multiple user access to the database. Usually MySQL has no GUI and runs from command prompt, but in our development we have used PHPMyAdmin which is web graphical interface to connect MySQL database.

MySQL consists of many databases in it. Then each database has tables inside them. Each table is made of columns and rows. So while creating a table one has to define the elements inside the table. The elements in the column can have many different types as mentioned earlier like text, integers, images, URLs, date, time etc. Data from MySQL database can be retrieved by structural query language or SQL which is has a query followed by an action to be done. Some examples of SQL are CREATE, DROP, ALTER, SHOW, INSERT, UPDATE, DELETE. An example of this is like: SELECT alk-name from Mplant; where alk-name is the column header from the Mplant table.

PHP: It is a server side scripting language which is mainly used for web development. PHP stands for PHP: Hypertext processor which used to stand for personal home page. PHP code can be embedded in any HTML page. So when the HTML is being requested it will

execute the PHP code and give results. PHP takes input data from file or stream having text and outputs another stream of text. Since PHP works on the server on the browser; so developers do not have to worry about the different browsers and their running environment. Every time a PHP is embedded in the HTML code it will be executed and results will be again formatted and shown as HTML pages.

PHP is very popular because of its clean syntax, object oriented fundamentals, extensible architecture, portability. Another important thing about PHP is that it is very easy to connect to different types of databases and querying them to get results or modifying the databases. PHP's ease of use in the web, together with its tight integration with MySQL, has made it the most used programming language for web-based data-driven applications.

So the Apache web server is first set up to intercept HTTP requests from clients and then it either serves them directly or passes them to PHP interceptors to execute and give the result. Then The MySQL serves as the data store accepting connections from the PHP layer and inserting, modifying, or retrieving data and this data is displayed using HTML.

Example of PHP codes:

```
<? php  
echo "<p>Hello World!"  
?>
```

Here <?php is the opening php tag and ?> is the closing php tag. PHP uses "echo" command to print anything.

To store variables in php \$ sign is used before the name of the variable. For example
\$value = 12;

for connecting to MySQL database there are many functions in PHP such as

```
mysql_connect($connection);
```

```
mysql_query("SQL statement",$connect);
```

In this project we have used local-host as our web server. So every database connection first connects to local host before making any request for data retrieval.

The command for connection to MySQL using PHP is

```
$connection=mysql_connect("localhost","root","password")
```

```
or die ("message");
```

This gives a handle called connection which can be further used for many data retrieval steps.

Python:

Python is scripting language having many different kinds of use. In this project we have used python as the web interfacing language for the cgi based conversion program. Because python is really simple and easy to use, so Python is used in many projects these days. Using common gateway interface; python can be connected to the web application. Python has many modules which simplify the use of it.

Pybel: Pybel is a python module that simplifies the access to OpenBabel API.(O'Boyle, Morley et al. 2008)

OpenBabel is a c++ toolkit with extensive capabilities of reading and writing molecular format. More than 80 formats are supported by OpenBabel. We can do many types of manipulations with OpenBabel. Many tasks such as determination set of smallest rings,

bond order perception, addition of hydrogens atoms and calculating and assignment of Gasteiger charges can be done using OpenBabel API.

So by using the OpenBabel Python module Pybel we can easily do many tasks in few lines of code. Simplified Wrapper and Interface Generator or SWIG is a tool that automatically generates bindings to libraries written in C or C++ such as OpenBabel. So Pybel module was created using this SWIG tool.

To use Pybel first one has to import Pybel module by invoking the call:

- `import Pybel`
- `mol = pybel.readfile("mol", "inputfile.mol").next()`
- `mol1 = Pybel.readstring("smi", "CCCC")`

This command is written to read the string and store that variable in mol1. Then we can do a number of things from it. Pybel provides replacements for two of the main classes in the OpenBabel library, OBMol and OBAtom.

- **Make3D:** This function is used to generate 3D co-ordinate for the molecules.
- **calcfp:** This is the function used in Pybel to calculate the molecular fingerprint.
- **Calcdesc:** This function is used to calculate the descriptor values.

SMILE format: To describe molecular structure of chemical compounds using ASCII strings is known as SMILE format. It is also known as simplified molecular input line entry system. Developed by David Weininger at Daylight chemical information systems it is very useful for many computational tasks which require strings to be processed. It has the facility to describe

atoms, bonds, benzene rings, stereochemistry, and isotopes. Here is an example molecule represented in SMILE format.

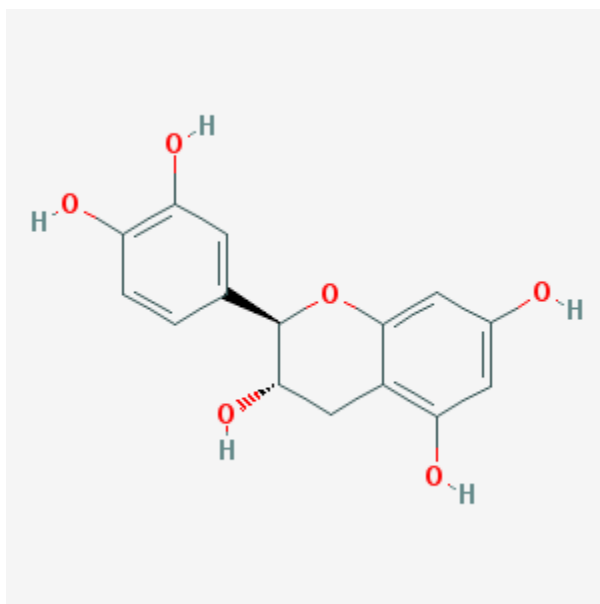


Figure 3.1 A compound represented in SMILE format

C1C(C(OC2=CC(=CC(=C21)O)O)O)C3=CC(=C(C=C3)O)O)O

HEX server for docking: Hex is an interactive molecular graphics docking suite for calculating and displaying feasible docking modes of pairs of protein and DNA molecules. Hex docking algorithms begin by assuming the proteins to be docked are rigid, and they employ geometric hashing or (fast Fourier transform (FFT) correlation techniques to find putative initial docking poses, which are then re-scored and refined.

- Input: we give 2 protein molecules as input as receptor and ligand.
Different docking parameters can be controlled such as grid dimensions, steric search etc.
- Output: As output it gives 100 different conformations and log files which arrange the complexes in order of their energy values.

3.2 Methods:

The web application designed for analysis of information from medicinal plants has the following features:

1. Similarity calculation between two small molecules
2. Molecular weight calculation for small molecules
3. Conversion of different molecular formats

3.2.1 Similarity calculation between two small molecules:

To get the similarity between two small molecules a specific term “fingerprint” is used. Fingerprint of a small molecule is representation of molecule’s different features in a measurable manner. Since if we convert the structure of a molecule into a bit string then it becomes very easy to compare two bit strings. One of the common type of fingerprint used is a series of binary digits or bits to represent certain substructure in the small molecule. In OpenBabel there few different type of fingerprint formats available. One if FP2, FP3, FP4, MACCS. FP2 is path based fingerprint format which stores the small molecule fragments up to 7 atoms. FP2, FP3, MACCS use smarts pattern to store molecular fingerprint.

Then Tanimoto coefficient was used as find similarity between molecular fingerprints. This coefficient is one the most widely used formula for calculating similarity. Suppose we have 2 molecules and BitA, BitB, BitAB are bits calculated for molecule A, B.

Then Tanimoto coefficient is

$$TC = \frac{BitAB}{(BitA + BitB - BitAB)}$$

The value of this Tanimoto coefficient varies in between 0 to 1. If it is 1 then two molecules being compared are totally same and 0 if they are totally different.

To get the value of similarity between two molecules the following code was implemented:

```
>>import Pybel  
  
>> mol1 =Pybel.readstring(“smi”,”smile format first molecule ”)  
  
>> mol2 = Pybel.readstring(“smi”,“smile format second molecule”)  
  
>> print mol1.calcfp() | mol2.calcfp()
```

This gives the similarity value between two molecules.

3.2.2 Molecular weight calculation for small molecules

For molecular weight calculation the following command is used as it comes in built in Pybel. Mol.molwt () is the predefined function used in Pybel. Molwt() function is able to calculate the molecular weight of any molecular format.

3.2.3. Conversion of different molecular formats

To convert from one molecular format to another format e.g. from SDF to SMI, can be done by OpenBabel. Usually this is needed because from Pubchem only SDF format can be downloaded. So to make the process easier for the user to convert formats a conversion box has been designed. Using this it will be quite easy to upload the molecule file in one format and it will be converted into another format as chosen by the user. After that he can save the molecule and use it for many other purposes. The main aim behind this was that when we are giving the similarity user should have the flexibility to upload his molecule and compare then to other molecule. So giving a conversion box it makes

the process easier if user has downloaded the molecule from Pubchem or drawn it using chemsketch.

To design this conversion tool we have used the OpenBabel API and combine that with HTML and JavaScript. Since python come within built server which can be accessed from HTTPServer module, we have used that to run a server which works on GET request. The tool box is able to read molecular formats and convert and save them in another format as chosen by the user.

Chapter 4

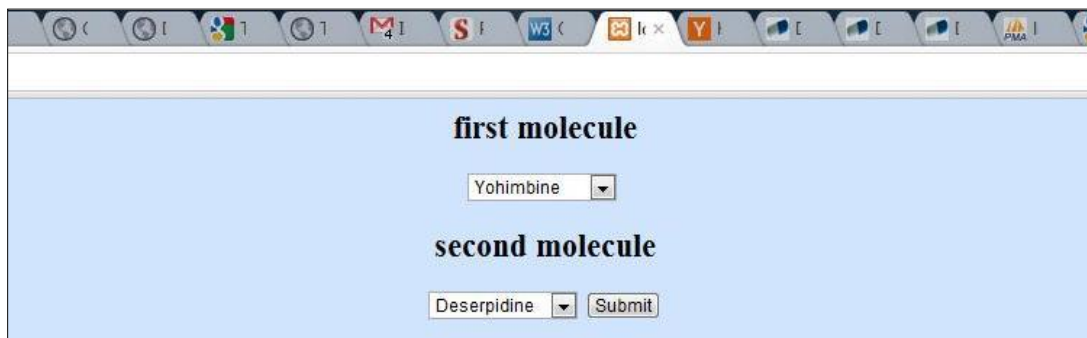
Results

Chapter 4

Results

4.1 Similarity score tool:

The final similarity search tool box looks like as shown here below in the figure.



The screenshot shows a web browser window with a light blue background. At the top, there is a navigation bar with several icons. Below this, the text "first molecule" is centered. Underneath it is a dropdown menu with "Yohimbine" selected. Below that, the text "second molecule" is centered. Underneath it is another dropdown menu with "Deserpidine" selected, followed by a "Submit" button.

Figure 4.1 Similarity toolbox

This has option of selecting the first molecule and second molecule between which similarity has to be calculated.

After selecting the two molecules user clicks on the submit button which takes him to the result page. Then the similarity score is shown in another page. It shows the similarity between two molecules in terms of a scoring value between 0-1. Also it shows the images of 2 molecules which are being compared according to their order.

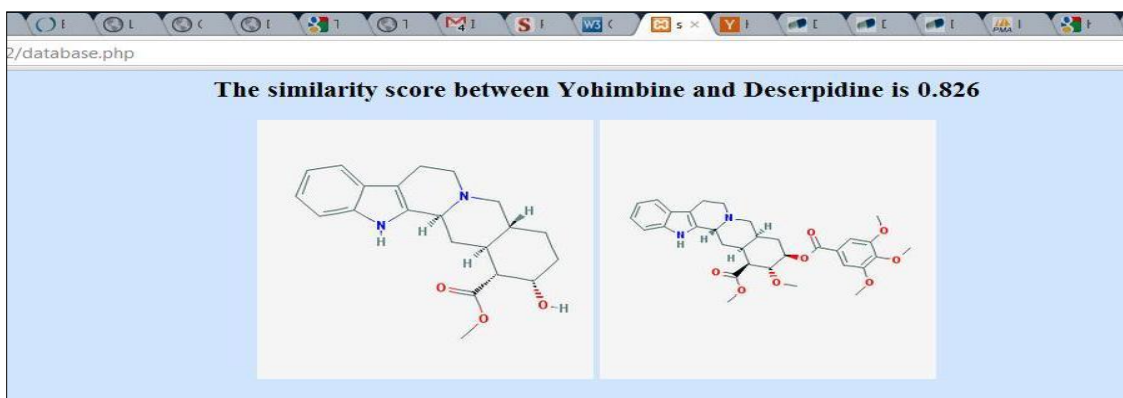


Figure 4.2 Result page of similarity toolbox

4.2 Molecular weight tool:

For calculation of molecular weight the molecular weight tool was made. Actually this has been made using PHP, MySQL. The values for different molecular weight are stored in the database.



Get molecular weight of the compound

Abraxane

Figure 4.3 Molecular weight toolbox

After selecting the molecule for which user want the molecular weight information, they submit the option which shows them the result.

The image of the molecule is also be shown in the result page.

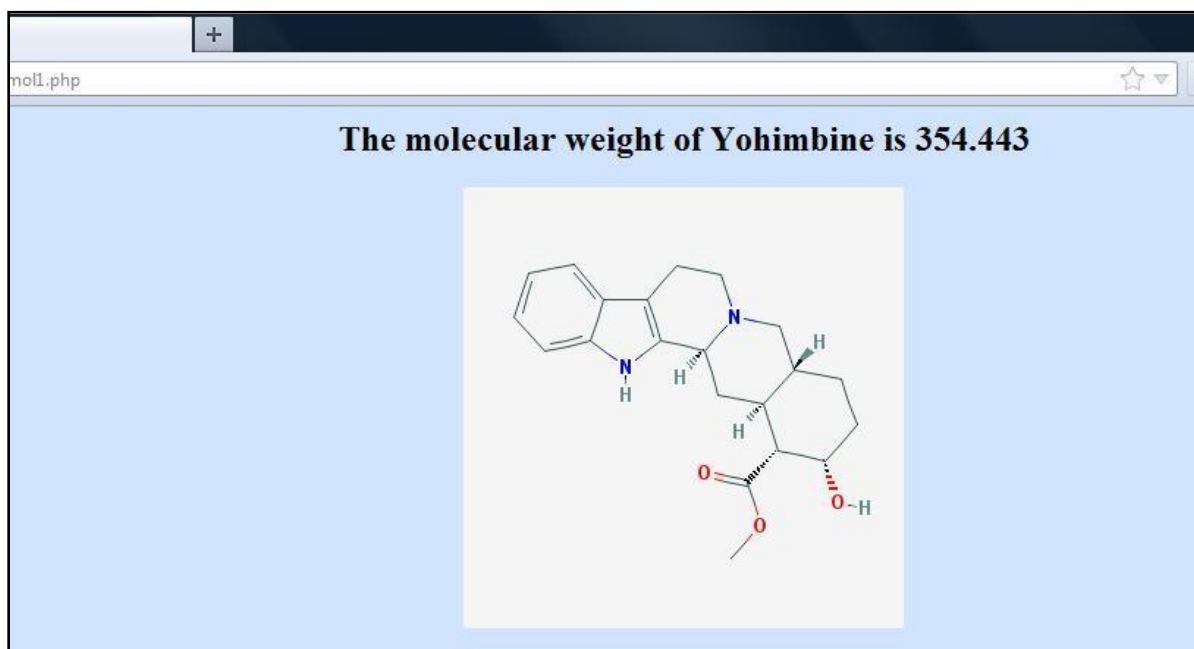


Figure 4.4 Result page of molecular weight tool

4.3 Conversion tool:

This conversion tool is made using a python server and OpenBabel API or python module of OpenBabel Pybel.

Here user gives input first by selecting which format the input is in and which format they want the output to be in. Then they have to upload the molecule in that format and click the convert button. This will convert the molecule in the output format.

There is also option to save the molecule.

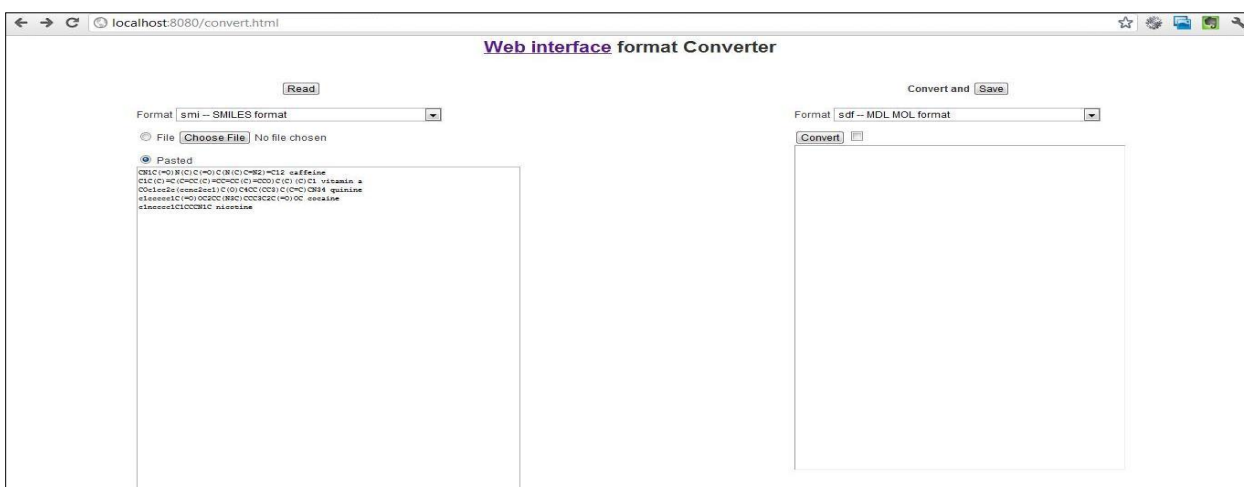


Figure 4.5 conversion tool

Using the box on the left user can give the input and output box on the right will give the output in the desired format. Below is the figure of the server running which takes client request and gives output.

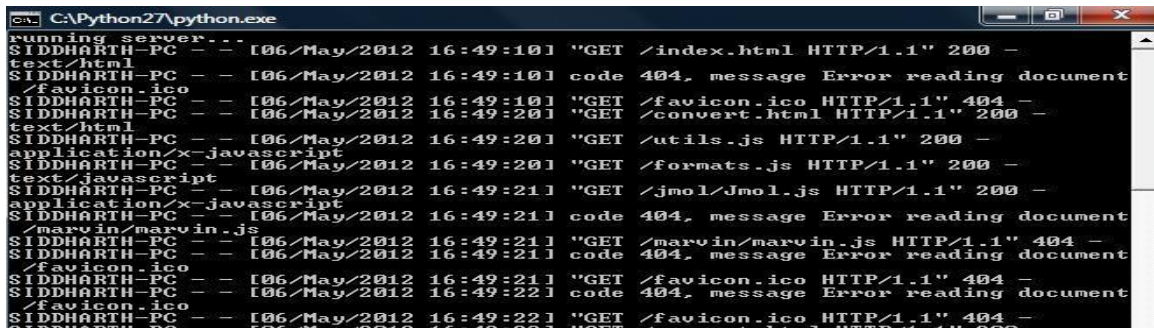


Figure 4.6 Python server running the conversion tool

4.4 Validation of application by docking:

Since we have the similarity tool we wanted to verify its usefulness for researchers by showing that two molecules which one user can find similar by the similarity tool are able to dock to same protein. So if the user has one ligand then he can find ligand which is similar and use them for docking.

We have taken two molecules

1. Curcumin ((1E,4Z,6E)-1,7-bis(3-methoxy-4-oxidanyl-phenyl)-5-oxidanyl-hepta-1,4,6-trien-3-one)
2. Piceatannol(3,3',4,5'-Tetrahydroxy-trans-stilbene,(E)4[2(3,5Dihydroxyphenyl)ethenyl] 1,2-benzenediol)

These two have similarity score of 0.53. So we have taken them and tried to show that they are able to dock to same inhibitors.

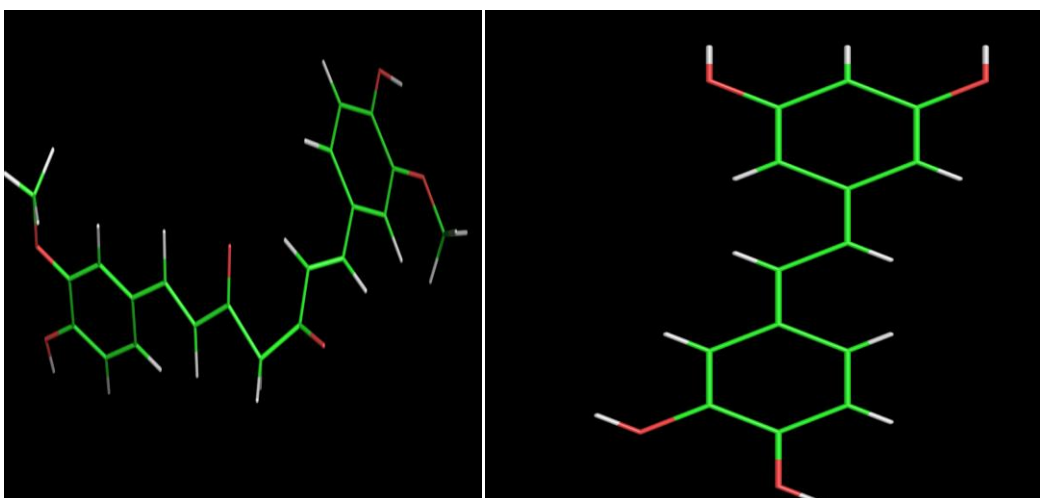


Figure 4.7

Curcumin in 3D structure

Piceatannol in 3D view

After getting two molecules in PDB format we did docking using hex and Patchdock server.(Schneidman-Duhovny, Inbar et al. 2005)

Results show that both Curcumin and piceatannol were able to bind to COX-2 Protein which is known to be inhibited by Curcumin. Curcumin has anti-inflammatory and anti-cancer activities and since piceatannol was similar to Curcumin it is expected that it also show the same effect.

COX-2 has effect in colon carcinogenesis. So binding of Curcumin with COX-2 was able to inhibit its action.

So after docking it was found that both Curcumin and piceatannol bind to COX-2. The interaction energy was found to be -224.45kcal/mol for the piceatannol and COX-2 interaction.

While the interaction energy for the interaction between Curcumin and COX-2 was found to be -244.60.

This shows that the interaction between Curcumin and piceatannol is quite similar. Since they both have negative interaction energy this means the interaction is stable.

This docking tool which will enable users to do automated docking among the molecule. Users can select which ligands are similar using the similarity finding tool, then they will be shown a option to do docking with protein based on information collected from literature. This will automatically connect to docking server where they can do the docking,

So in this way user of this tool box can have many benefits by easily finding the similar ligands and seeing whether they are able to inhibit the protein which their similar protein does.

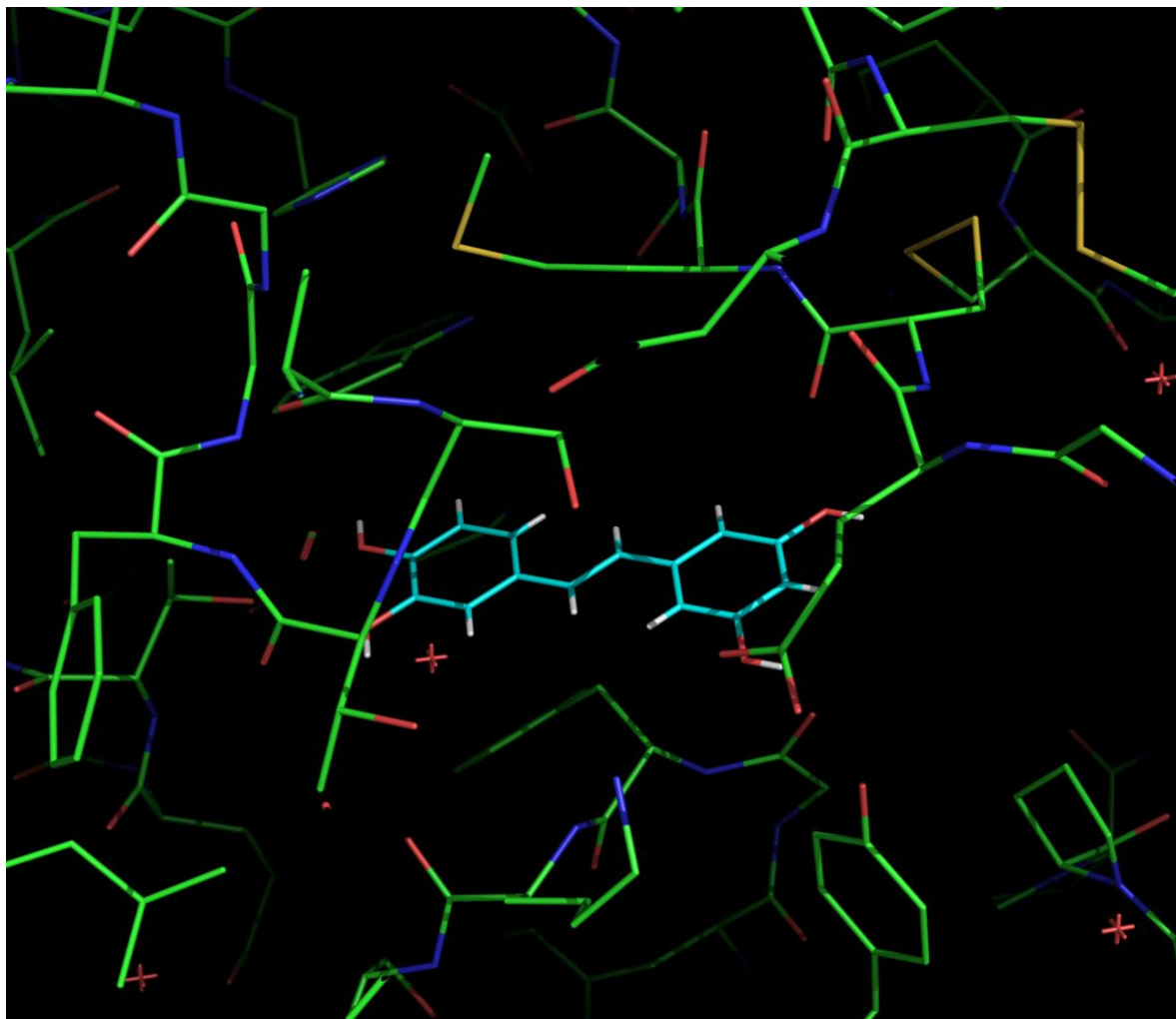


Figure 4.8

Docked piceatannol with COX-2

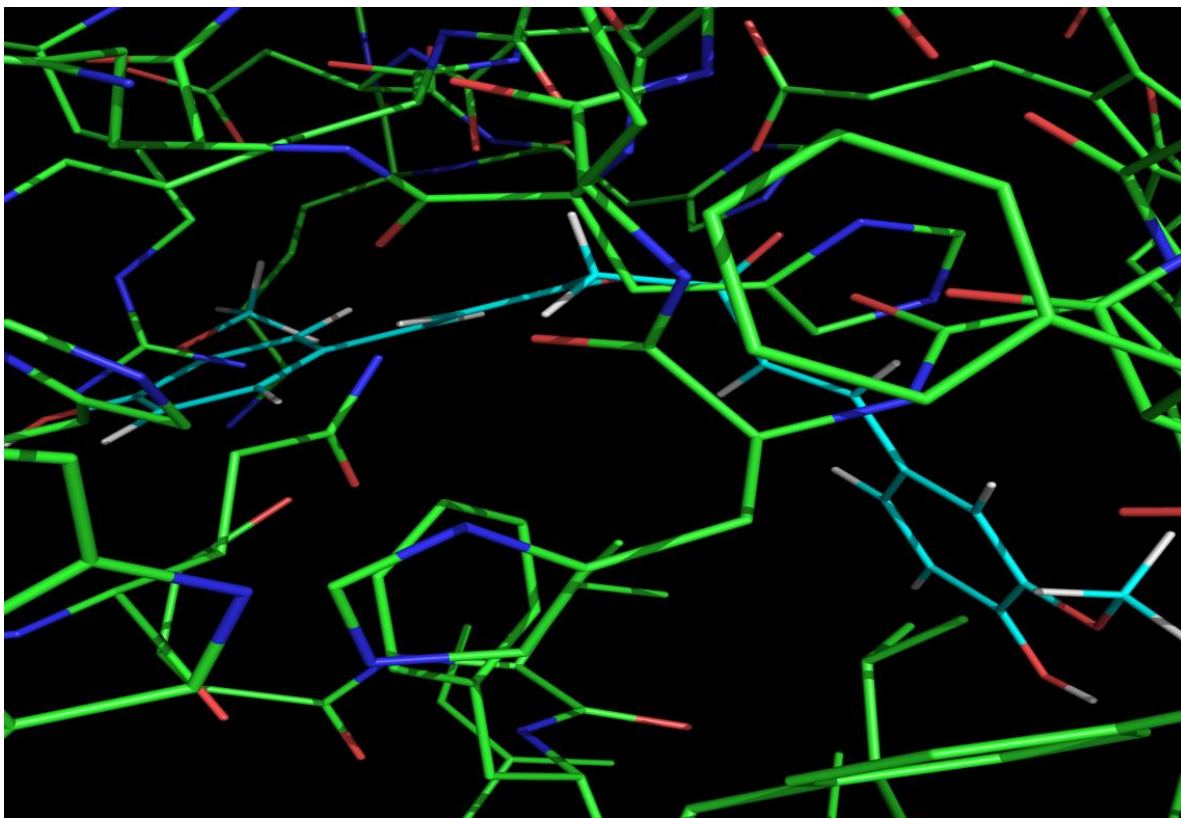


Figure 4.9

Docking of Curcumin with COX-2

Combining homology modeling (Schwede, Kopp et al. 2003) with the tool box has given the user another option to find or design another new protein molecule which is similar to initial one taken. So we have done the homology modeling for the protein COX-2 and another new protein model is created.

In the following page the images of newly formed protein and its docking with Curcumin is shown.

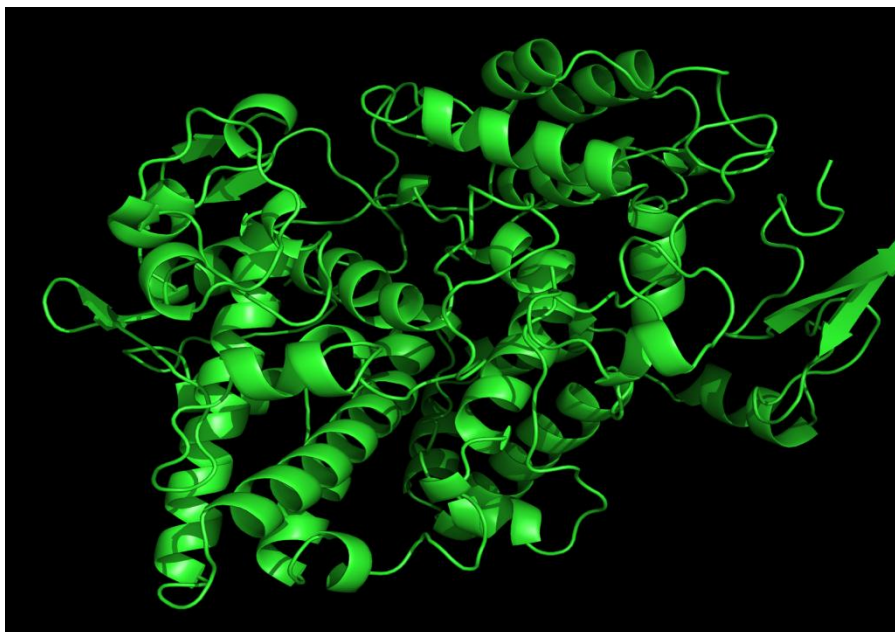


Figure 4.10

After getting homology modeled structure we again did docking among the protein which is homologous to COX-2 and Curcumin and piceatannol.

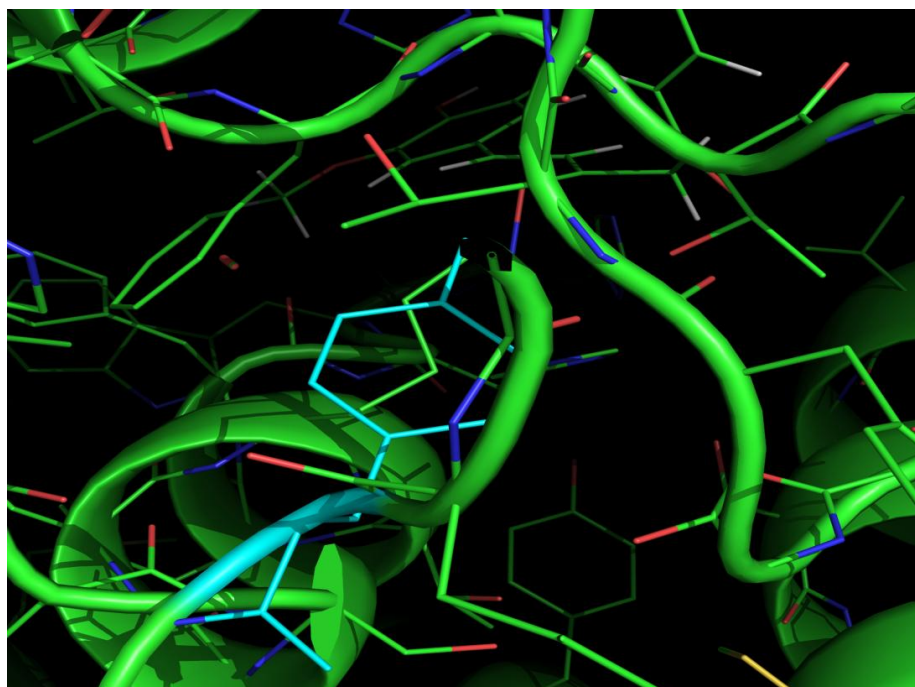


Figure 4.11

Docked structure among homologous protein of COX-2 and Curcumin

Chapter 5

Conclusion and future work

Chapter 5

Conclusion and future work

- As we have seen that the similarity tool box will give users option to find out similar ligands based on a similarity score. Users will be able to find similarity score among all molecules in a simplified manner. So it will be useful for users to find alternative ligands which can be used for drug discovery.

Future work for this tool box is that if a new sketching tool can be built in the application then the user will be able to draw new molecules and save them in any format they want. Then they will be able to find the similarity among other molecules in the database.

- Molecular weight tool box is useful for getting more information about the molecule. In future many more molecular information extraction features will be added to this tool.
- From the conversion tool box users can easily convert molecules into different format. This tool is helpful to find other similar molecules since similarity score finding tool can find the similarity between molecules which are in SMILE format.

In future this tool has to be incorporated with a visualizing tool which will be able to show the molecule in 3D format in the web browser.

References

1. Goede, A., M. Dunkel, et al. (2005). "SuperDrug: a conformational drug database." Bioinformatics **21**(9): 1751-1753.
2. Goto, S., Y. Okuno, et al. (2002). "LIGAND: database of chemical compounds and reactions in biological pathways." Nucleic Acids Res **30**(1): 402-404.
3. Lipinski, C. and A. Hopkins (2004). "Navigating chemical space for biology and medicine." Nature **432**(7019): 855-861.
4. O'Boyle, N. M., C. Morley, et al. (2008). "Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit." Chem Cent J **2**: 5.
5. Schneidman-Duhovny, D., Y. Inbar, et al. (2005). "PatchDock and SymmDock: servers for rigid and symmetric docking." Nucleic Acids Res **33**(Web Server issue): W363-367.
6. Schwede, T., J. Kopp, et al. (2003). "SWISS-MODEL: An automated protein homology-modeling server." Nucleic Acids Res **31**(13): 3381-3385.

7. Wang, Y., J. Xiao, et al. (2009). "PubChem: a public information system for analyzing bioactivities of small molecules." Nucleic Acids Res **37**(Web Server issue): W623-633.

8. Wishart, D. S., C. Knox, et al. (2006). "DrugBank: a comprehensive resource for in silico drug discovery and exploration." Nucleic Acids Res **34**(Database issue): D668-672.

Appendix

Similarity score database

id	mol_name 1	mol name 2	mol_fp
11	Abraxane	Abraxane	1
12	Abraxane	Acarbose	0.3798
13	Abraxane	acetyldigitoxin	0.345744
14	Abraxane	Acitritin	0.15942
15	Abraxane	Aclarubicin	0.38697
16	Abraxane	Acyclovir	0.18303
17	Abraxane	lomefloxacin	0.17142
18	Abraxane	Nafcillin	0.27016
19	Abraxane	Trifluralin	0.4219
21	Acarbose	Abraxane	0.3798
22	Acarbose	Acarbose	1
23	Acarbose	Acetyldigitoxin	0.32168
24	Acarbose	Acitritin	0.4215
25	Yohimbine	Deserpidine	0.826
26	Acarbose	Dihydro-Acarbose	0.686
27	Acarbose	H TYPE I TRISACCHARIDE	0.665
28	Acarbose	Netilmicin	0.641
29	Acarbose	Amikacin	0.636
30	Acarbose	Apramycin	0.633
31	Acarbose	Arbekacin	0.621
32	Acarbose	Lactose Sialic Acid	0.62
33	Acarbose	Aspartate Semialdehyde	0.613
34	Acarbose	Cis-tetracosenoyl sulfatide	0.594
35	Acarbose	Di(N-Acetyl-D-Glucosamine)	0.591
36	Acarbose	Chitotriose	0.591
37	Acarbose	Alpha-Methyl-N-Acetyl-D-Glucosamine	0.591
38	Acarbose	Spectinomycin	0.58

id	mol_name 1	mol name 2	mol_fp
40	Acarbose	O-Sialic Acid	0.578
41	Acarbose	Gentamicin	0.578
42	Acarbose	Lividomycin A	0.574
45	Acarbose	Paromomycin	0.574
46	Acarbose	Framycetin	0.574
47	Acarbose	Ribostamycin	0.574
50	Acarbose	Neomycin	0.574
52	Acarbose	Kanamycin	0.569
54	Acarbose	Streptomycin	0.569
55	Acarbose	N-Acetyl-D-Allosamine	0.563
56	Acarbose	Neamine	0.558
57	Acarbose	Azithromycin	0.555
58	Acarbose	Tobramycin	0.553
59	Acarbose	Josamycin	0.549
61	Azithromycin	Dirithromycin	0.831
62	Azithromycin	Erythromycin	0.829
64	Azithromycin	Clarithromycin	0.825
65	Azithromycin	Troleandomycin	0.769
66	Azithromycin	Roxithromycin	0.732
67	Azithromycin	Josamycin	0.717
69	Azithromycin	Dihydro-Acarbose	0.693
70	Azithromycin	Gentamicin	0.687
71	Azithromycin	Spectinomycin	0.644
72	Azithromycin	Amphotericin B	0.643
77	Azithromycin	Arbekacin	0.616
78	Azithromycin	Netilmicin	0.613
81	Azithromycin	Streptomycin	0.603
84	Spectinomycin	Gentamicin	0.735
88	Spectinomycin	Amikacin	0.702
89	Spectinomycin	Dihydro-Acarbose	0.698
91	Digoxin	Digitoxin	1

id	mol_name 1	mol name 2	mol_fp
92	Digoxin	Acetyldigitoxin	0.978
93	Digoxin	Deslanoside	0.978
94	Digoxin	Ouabain	0.963
95	Digoxin	ginsenoside Rb1	0.699
96	Digoxin	Fusicoccin	0.684
97	Digoxin	Ciclesonide	0.671
98	Digoxin	Ivermectin	0.663
99	Digoxin	Budesonide	0.655
100	Digoxin	Fusidic Acid	0.653
101	Digoxin	Desonide	0.627
102	Digoxin	Prednicarbate	0.608
103	Digoxin	Amcinonide	0.608
104	Digoxin	Natamycin	0.604
105	Hesperetin	Naringenin	0.932
106	Hesperetin	5-Deoxyflavanone	0.925
107	Rescinnamine	Rescinnamine	1
108	Rescinnamine	Reserpine	0.911
109	Rescinnamine	Deserpidine	0.891
110	Rescinnamine	Yohimbine	0.823
111	Rescinnamine	Vincristine	0.685
112	Rescinnamine	Vinblastine	0.682
113	Rescinnamine	Tadalafil	0.678
114	Rescinnamine	Vindesine	0.662
115	Rescinnamine	Voacamine	0.658
116	Rescinnamine	Vinorelbine	0.655
117	Rescinnamine	Dihydroergotamine	0.601
118	Rescinnamine	Ergoloid mesylate	0.59
119	Rescinnamine	Bromocriptine	0.59
120	Rescinnamine	Quinupristin	0.589
121	Rescinnamine	Balhimycin	0.584
122	Rescinnamine	Trabectedin	0.58

id	mol_name 1	mol name 2	mol_fp
123	Rescinnamine	Virginiamycin factor S1	0.578
124	Rescinnamine	Desvancosaminyl Vancomycin	0.578
125	Rescinnamine	Nicergoline	0.576
127	Rescinnamine	Acetyldihydrocodeine	0.575
128	Rescinnamine	Vancomycin	0.573
129	Rescinnamine	Sinapoyl Coenzyme A	0.571
130	Rescinnamine	Deglucobalhimycin	0.57
131	Rescinnamine	Irinotecan	0.568
132	Rescinnamine	N-Methylnaloxonium	0.566
133	Rescinnamine	Benzquinamide	0.566
134	Rescinnamine	Feruloyl Coenzyme A	0.564
135	Rescinnamine	Feruloyl Coenzyme A	0.563
136	Yohimbine	Deserpidine	0.826
137	Yohimbine	Rescinnamine	0.823
138	Yohimbine	Tadalafil	0.678
139	Yohimbine	Voacamine	0.63
140	Ginsenoside Rg1	ginsenoside C	0.98
141	Ginsenoside Rg1	ginsenoside Rb1	0.98
142	Ginsenoside Rg1	Ouabain	0.716
143	Ginsenoside Rg1	Cyclohexyl-pentyl-maltoside	0.71
144	Ginsenoside Rg1	Cyclohexyl-Hexyl-Beta-D-Maltoside	0.71
145	Ginsenoside Rg1	Deslanoside	0.706
146	Ginsenoside Rg1	Digoxin	0.684
149	Ginsenoside Rg1	Digitoxin	0.684
150	Ginsenoside Rg1	Octyl alpha-L-altropyranoside	0.673
151	Ginsenoside Rg1	Acetyldigitoxin	0.669
152	Ginsenoside Rg1	Inulin	0.654

This list is not exhaustive and more data is available in the database.