Laboratory of Natural Information Processing
DA-IICT Gandhinagar

DrugCalc

User Manual

DrugCalc

User Manual

Laboratory of Natural Information Processing
DA-IICT, Gandhinagar, Gujarat 382007
http://www.guptalab.org/drugcalc

Documentation version 1.0

This file was last modified on May 18, 2013.

Credits & Team

Principle Investigator: Manish K. Gupta, PhD.
Graduate Advisor: Vandana Ravindran
Key Developers: Jinay Kothari* and Hritik Bharadwaj*
Software Logo: Himanshu Agarwal

* Key developers contributed equally to the project

Acknowledgments

We thank Ms. Vandana Ravindran and Ms. Dixita Limbachia for useful discussion in every stage of project. For fetching molecule files we will thank PDB, Uniprot and Drug Bank databases which are free to access on internet. For the Amber Force Field algorithm we will thank Peter Kollman's group at the University of California, San Francisco whose open source code has been used with modifications.

Table of Contents

1. Introduction 4

2. Fetching of Protein Molecule 5

    2.1. Fetching from Uniprot Database 6

    2.2. Fetching from PDB database 8

3. Fetching of Drug Molecule 9

4. View Genes from PharmGKB 11

5. JMOL 12

6. Force Field Calculation of Bonded Atoms 13

7. Delete History 14

8. Help Menu 15

9. About 16

10. Support and Feedback 17

11.References 17

1. Introduction

In the fields of medicine, biotechnology and pharmacology, drug discovery is the process by which new candidate medications are discovered. Historically, drugs were discovered through identifying the active ingredient from traditional remedies or by serendipitous discovery. Later chemical libraries of synthetic small molecules, natural products or extracts were screened in intact cells or whole organisms to identify substances that have a desirable therapeutic effect in a process known as classical pharmacology. Since sequencing of the human genome which allowed rapid cloning and synthesis of large quantities of purified proteins, it has become common practice to use high throughput screening of large compounds libraries against isolated biological targets which are hypothesized to be disease modifying in a process known as reverse pharmacology. Hits from these screens are then tested in cells and then in animals for efficacy. Even more recently, scientists have been able to understand the shape of biological molecules at the atomic level, and to use that knowledge to design drug candidates. DrugCalc is an open source software which will be useful in drug discovery by simplifying some of the basic chores which are done in the initial phase of drug discovery. It integrates three of the major Drug/Protein databases (Drug Bank, Uni Prot, PDB). So from a single platform you can view description about any Drug/ Protein molecule in all the available formats. Using the user choice protein molecule it helps in calculating the force field (potential energy of system of particles). This document is prepared to give users an overview of DrugCalc software, utilities available in DrugCalc and motive behind the development of the software.

Entire software can be well understood by understanding its four major functionalities.

- Fetching of Protein Molecule
- Fetching of Drug Molecule
- Gene Molecule
- Force Field calculation

Fetching of Protein Molecule fetches the data from Uniprot Database as well as Protein Data Bank. Fetching of Drug Molecule fetches the data from Drug Bank. Force Field Calculation calculates the force field (the potential energy of system of particles) for the given protein molecule and thus gives information regarding the binding site, potential energy of the system which is very essential in the process of drug discovery. The Force field calculation is available through AMBER algorithm.

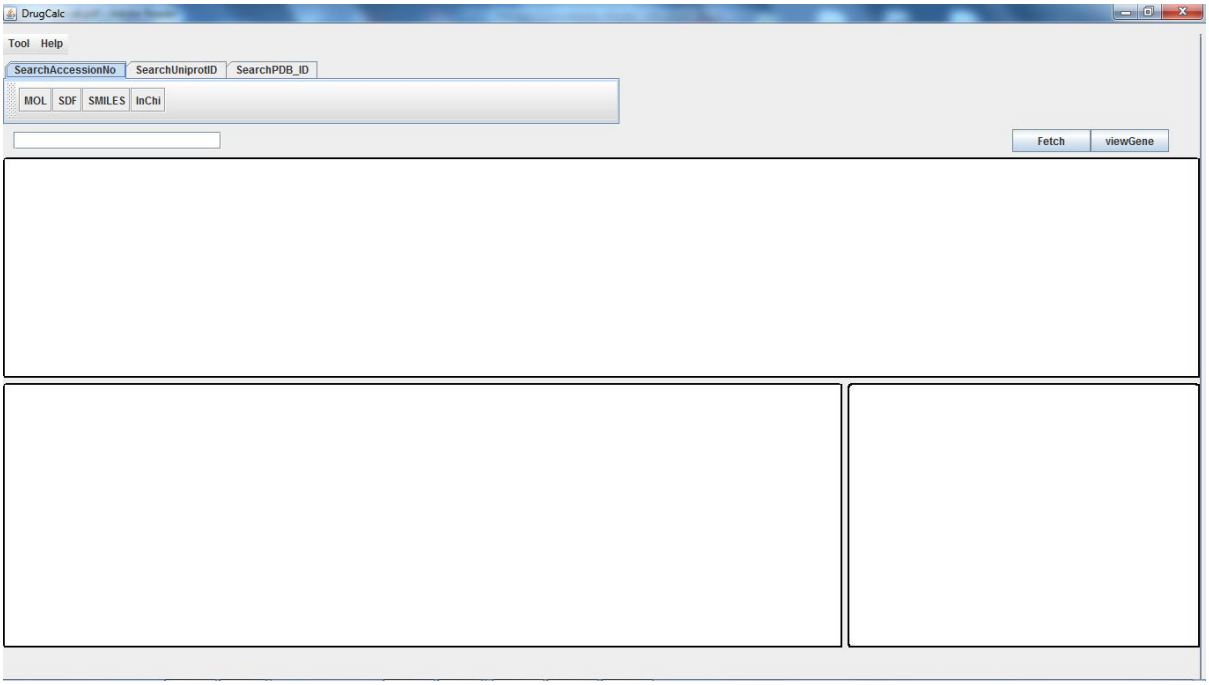

Figure 1: Actual snapshot of the software showing the main window.

Fig 1 shows us the actual snapshot of the software showing the main window. As we can see there are various options available at the top. The two drop down menus provides options to users regarding the functionalities available as well as help or a user can get any information regarding the software. The About button will give a user information about the software DrugCalc. In case a user needs any help he can get basic help from the help button at the top. In case user wants to give a feedback or wants to contact us he can get all the information regarding that from contact us button.

There are 3 options available regarding the fetching of molecule. Search Accession no provides a user the functionality to fetch a drug molecule through its access number from the Drug Bank database. A user can download the drug molecule file in MOL, SDF, SMILES as well as InChi format. Search UniprotID provides a user the functionality to fetch a protein molecule through its UniprotID from the Uniprot database. A user can download the protein molecule file fetched from Uniprot database in FASTA , XML as well as text format. Search PDB_ID provides a user the functionality to fetch a protein molecule through its PDB ID from the Protein Data Bank. A user can download the protein molecule fetched from PDB database in PDB, FASTA, CIF as well as XML format. The text box below them is the place where a user can write the ID of the molecule he wants to fetch.The main central window panel below them displays the molecule file in the file format user has asked for. The Left bottom window displays Accession Numbers of drug molecules for the corresponding protein molecule. The Centre bottom window calculates and displays the force field of system of particles. The right bottom window displays the 2D structure of a given molecule.

2. Fetching of Protein Molecule

“Fetching of protein molecule” forms very fundamental block of the software. The protein molecules are fetched from the online databases of protein molecule banks. Uniprot and PDB

databases are used in fetching of protein molecule. A user can easily select one of the two databases to fetch the protein molecule.

2.1. Fetching from Uniprot Database

Uniprot database provides details about the protein molecules which can be accessed by the unique protein Uniprot ID which is associated with each protein molecule. Each and every record of this Uniprot database is given unique Uniprot ID. Very important attribute is the Uniprot ID where we provide the ID of particular protein molecule.

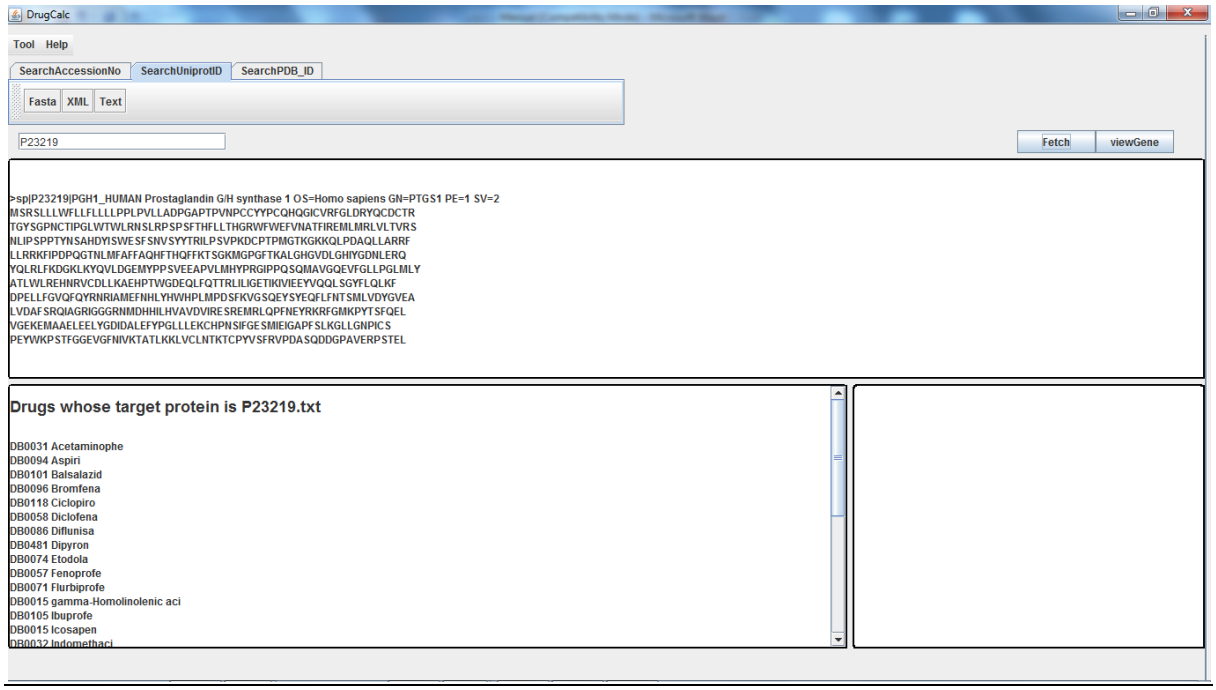

Figure 2: Actual snapshot of the software fetching P23219 protein molecule.

Figure 2 shows the screen shot of DrugCalc fetching a particular protein molecule. A user can fetch protein molecule in FASTA, XML as well as in Text file format from Uniprot Database. When the application starts it shows the most recent fetched file in the upper pane. If you are using for the first time or if there are no files in your history then you will see the blank pane. Figure 2 shows the fetching of protein molecule in FASTA file format having Uniprot ID P23219. Here the upper panel displays the FASTA file for the protein molecule having Uniprot Id P23219. As Uniprot provides the details of protein molecules in terms of sequences of amino acids the 2D structure as well as 3D structure view for the molecule is not available. The left bottom panel displays the information about the Accession number of Drug molecules. These drug molecules are the one which can target the particular protein molecule. Thus a user gets the information about the drugs for the corresponding protein molecule.

The user can access this facility by pressing the tab SearchUniprotID and then selecting the file format in which he wants the protein molecule file. The user has to enter the Uniprot ID in the text box given below them. After writing the Uniprot ID a user has to click on Fetch button. Before downloading any sequence from internet, our application checks user's internet connectivity and validity of Uniprot ID.

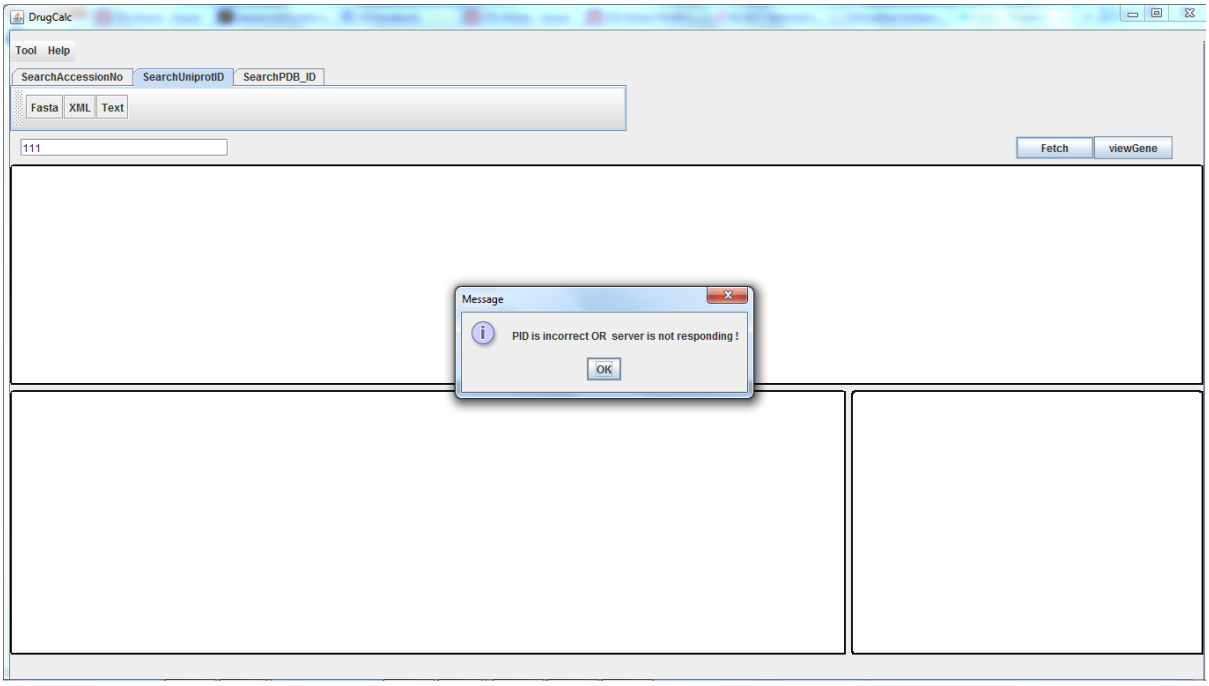

Figure 3: DrugCalc showing the dialog box when there is internet problem or ID is not correct.

In case of user providing wrong Uniprot ID or user having internet issues, an error dialog box opens up as shown in figure 3 asking user to check internet as well as to write a valid Uniprot ID. Once all the validation is done, software starts downloading the protein molecule file in the file format user has asked for. On successful download of the file, software shows the prompt acknowledging that operation has been successful as shown in figure 4.

As soon as download is over, newly downloaded file is shown in the upper pane, Accession numbers of drug molecules which can target particular protein molecule are displayed in left bottom panel.

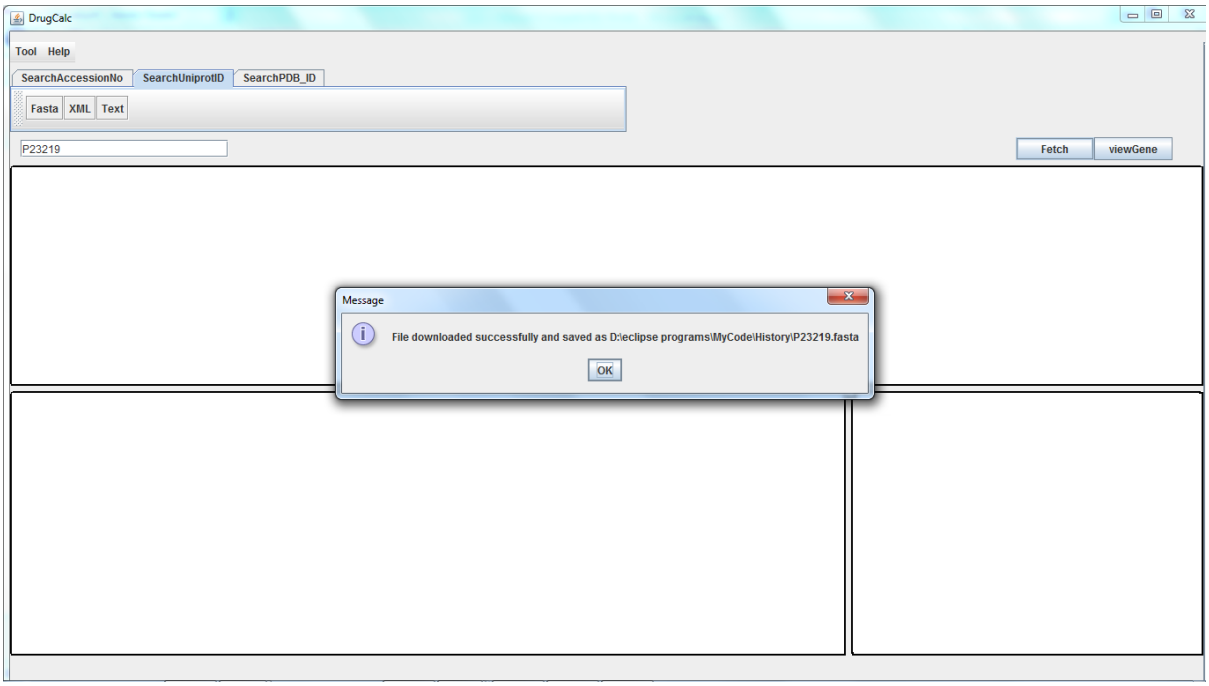

Figure 4: DrugCalc showing the dialog box when the file from Uniprot has been downloaded.

2.2. Fetching from PDB database

PDB database provides details about the protein molecules which can be accessed by the unique PDB ID which is associated with each protein molecule. Each and every record of this PDB database is given unique PDB ID. Very important attribute here is the PDB ID where we provide the ID of particular protein molecule.

Figure 5 shows the screen shot of DrugCalc fetching a particular protein molecule. A user can fetch protein molecule in PDB, FASTA, CIF as well as in XML file format from PDB Database. When the application starts it shows the most recent fetched file in the upper pane. If you are using for the first time or if there are no files in your history then you will see the blank pane.

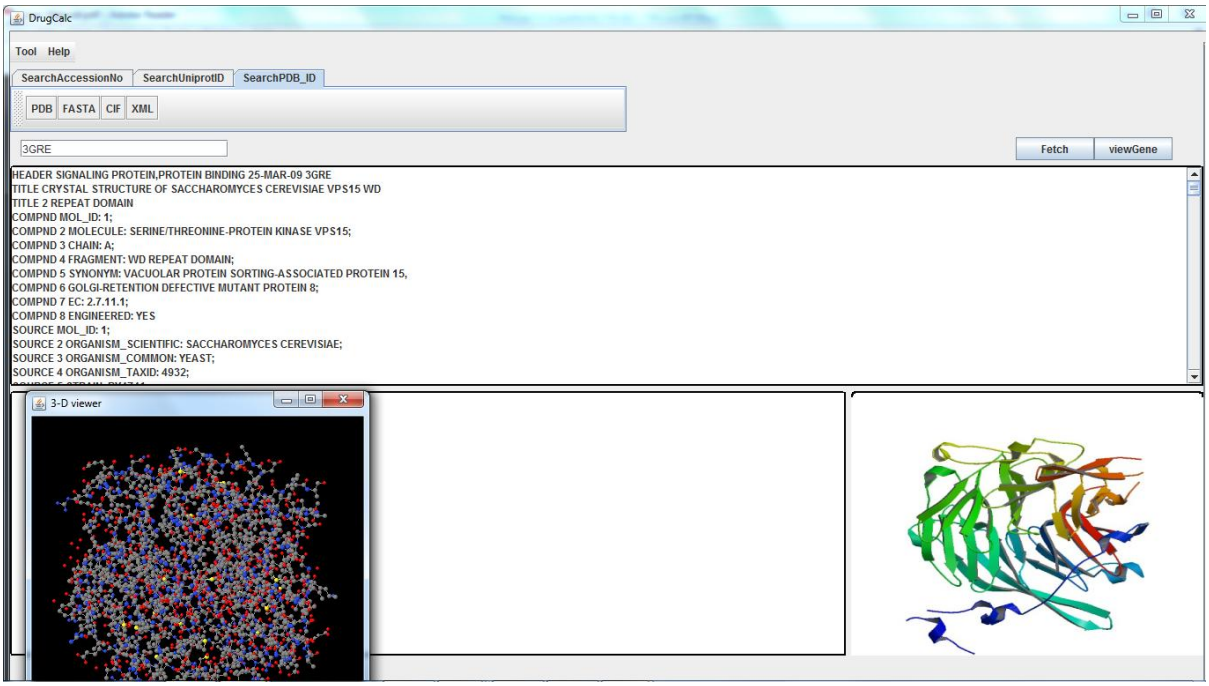

Figure 5: Actual screenshot of the software fetching 3GRE protein molecule.

Figure 5 shows the fetching of protein molecule in PDB file format having PDB ID 3GRE. Here the upper panel displays the PDB file for the protein molecule having PDB ID 3GRE. As PDB provides the details of protein molecules in terms of the structure of atoms their coordinates, it is possible to create 2D as well as 3D structure from the given file. The right bottom panel displays the 2D string structure of the given protein molecules. As the file fetched from PDB database gives information about x, y as well as z co-ordinates of the atoms present in molecule, it is possible to create the 3D structure view of the molecule. So as shown in the figure 5 a new window pops up showing 3D structural view of the given protein molecule.

The user can access this facility by pressing the tab SearchPDBID and then selecting the file format in which he wants the protein molecule file. The user has to enter the PDB ID in the text box given below them. After writing the PDB ID a user has to click on Fetch button. Before downloading any sequence from internet, our application checks user's internet connectivity and validity of PDB ID.

In case of user providing wrong PDB ID or user having internet issues, an error dialog box opens up asking user to check internet as well as to write a valid PDB ID. Once all the validation is done, software starts downloading the protein molecule file in the file format user has asked for. On successful download of the file, software shows the prompt acknowledging that operation has been successful as shown in figure 6.

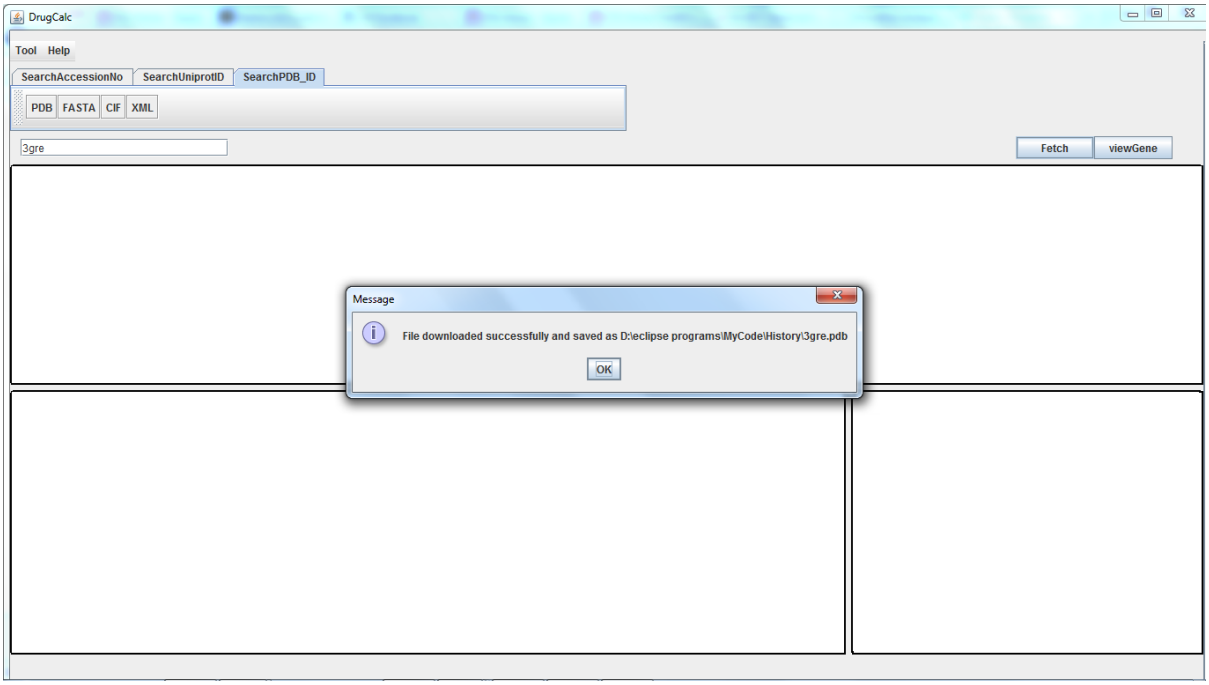

Figure 6: DrugCalc showing the dialog box when the file from PDB has been downloaded.

As soon as download is over, newly downloaded file is shown in the upper pane, 2D string structure of the molecule is shown in the right bottom panel while the 3D structure of the molecule is shown in the new window.

3. Fetching of Drug Molecule

"Fetching of drug molecule" is also one of the very fundamental blocks of the software. The drug molecules are fetched from the online databases of drug molecule banks. Drug Bank database is used in fetching of drug molecule.

Drug Bank database provides details about the drug molecules which can be accessed by the unique Accession No. which is associated with each drug molecule. Each and every record of this Drug Bank database is given unique Accession No. Very important attribute here is the Accession No. where we provide the ID of particular drug molecule.

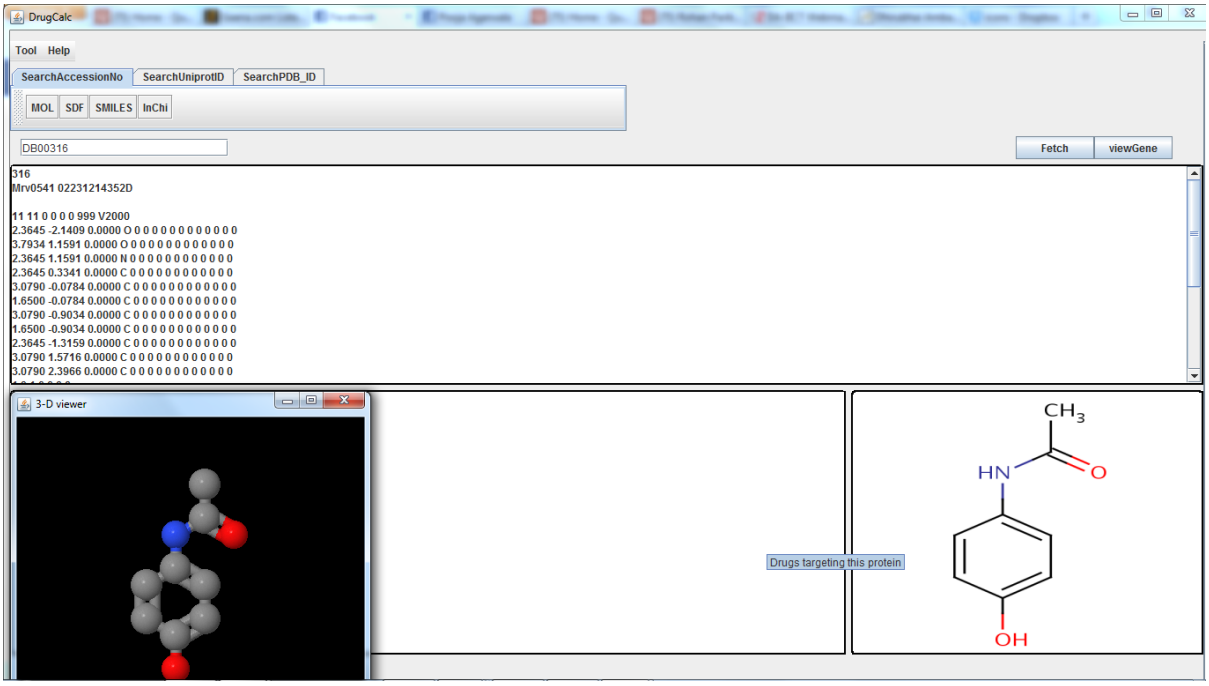

Figure 7: Actual screenshot of the software fetching DB00316 drug molecule.

Figure 7 shows the screen shot of DrugCalc fetching a particular drug molecule. A user can fetch drug molecules in MOL, SDF, SMILES as well as in InChi file format from Drug Bank Database. When the application starts it shows the most recent fetched file in the upper pane. If you are using for the first time or if there are no files in your history then you will see the blank pane.

Figure 7 shows the fetching of drug molecule in MOL file format having Accession No. DB00316 . Here the upper panel displays the MOL file for the drug molecule having Accession No. DB00316. As Drug Bank provides the details of drug molecules in terms of the structure of atoms their coordinates, it is possible to create 2D as well as 3D structure from the given file. The right bottom panel displays the 2D stick structure of the given drug molecules. As the file fetched from Drug Bank database gives information about x, y as well as z co-ordinates of the atoms present in molecule, it is possible to create the 3D structure view of the molecule. So as shown in the figure a new window pops up showing 3D structural view of the given drug molecule.

The user can access this facility by pressing the tab SearchAccessionNo and then selecting the file format in which he wants the protein molecule file. The user has to enter the Accession No. in the text box given below them. After writing the Accession No. a user has to click on Fetch button. Before downloading any sequence from internet, our application checks user's internet connectivity and validity of Accession No. In case of user providing wrong Accession No. or user having internet issues, an error dialog box opens up asking user to check internet as well as to write a valid Accession No. Once all the validation is done, software starts downloading the drug molecule file in the file format user has asked for. On successful download of the file, software shows the prompt acknowledging that operation has been successful as shown in figure 8.

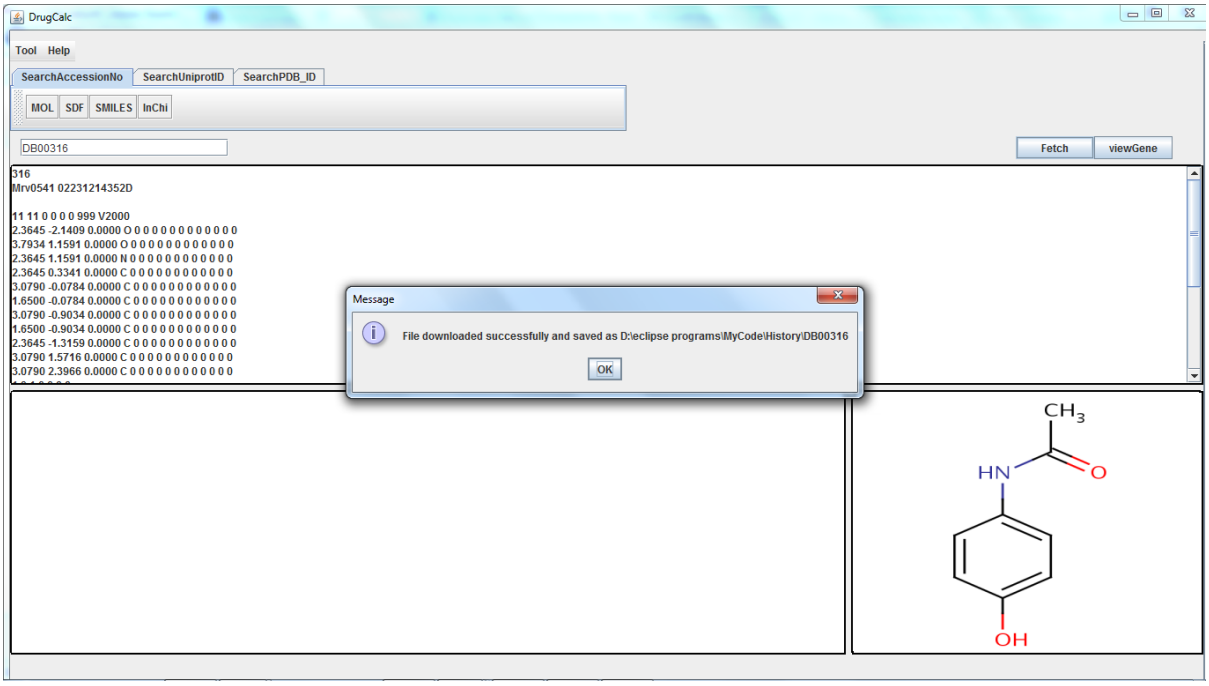

Figure 8: DrugCalc showing the dialog box when the file from Drug Bank has been downloaded.

As soon as download is over, newly downloaded file is shown in the upper panel, 2D stick structure of the molecule is shown in the right bottom panel while the 3D structure of the molecule is shown in the new window.

4. View Genes from PharmGKB

PharmGKB is an online database site which provides information about genes, drugs and diseases. It does not give permission to fetch any molecule file. So to integrate Gene bank with DrugCalc we redirect our user to the webpage of PharmGKB for a particular Gene PharmGKB ID. As showin in figure 9 user wants to view details of gene having PharmGKB ID PA 267. DrugCalc will redirect user to www.pharmgkb.com/genes/PA267. Here user can check details about the gene. Also user is provided information about the accession number of related drug molecules.

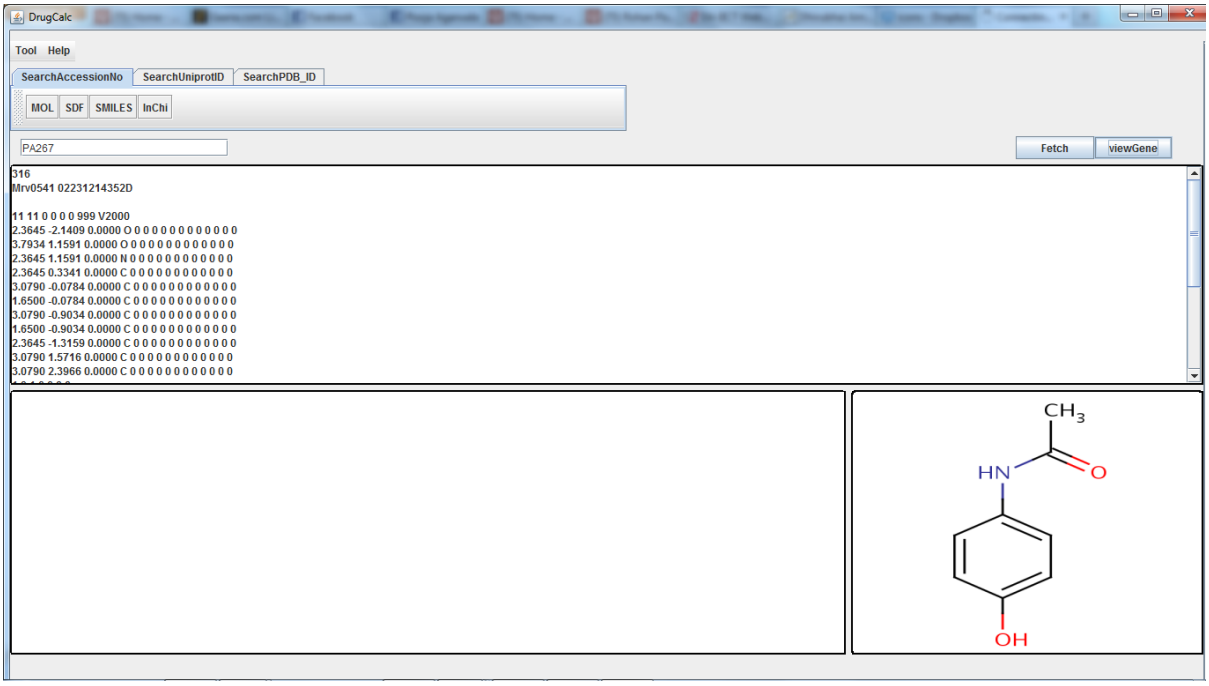

Figure 9: DrugCalc showing PharmGKBID of gene which a user wants to view

5. JMOL

JMOL is an open source Java viewer for three-dimensional chemical structures, with features for chemicals, crystals, materials and bio-molecules. Features include reading a variety of file types and output from quantum chemistry program and animation of multi-frame files and computed normal modes from quantum programs.

A user can perform a lot of function in JMOL viewer which can be available by executing commands on the console which is given below the 3D image. Some of the commands such as Zoom IN, Zoom Out, Left View, Centre View can be available by right clicking on the image of 3D structure.

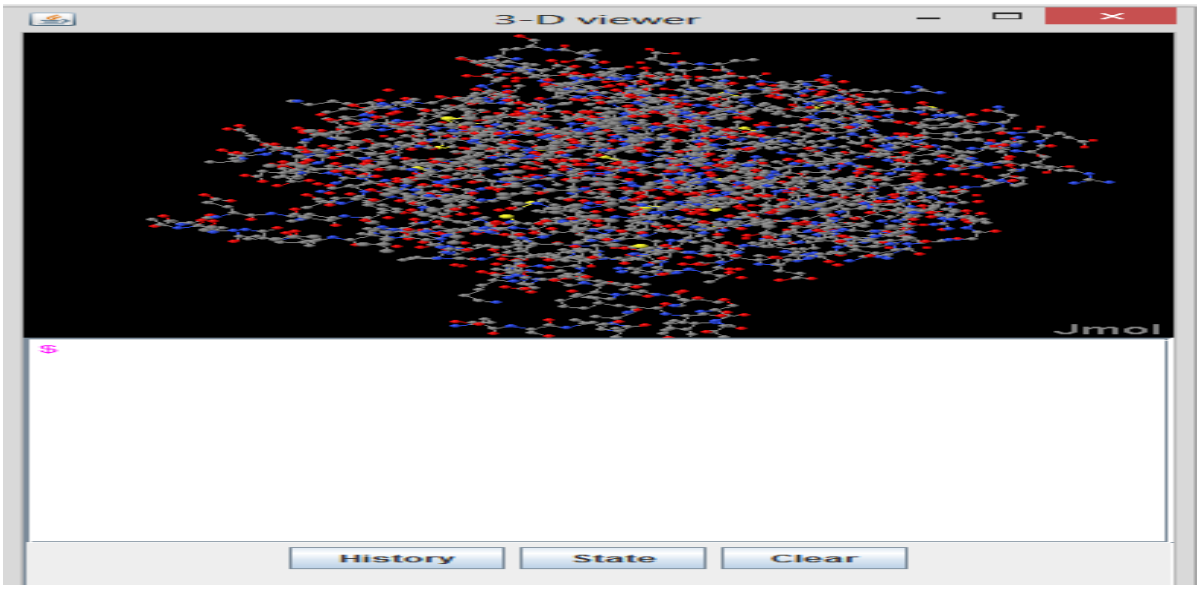

Figure 10: Actual screenshot from the software showing the 3D structure of the molecule using JMOL viewer

The three buttons at the end

History- IT helps to recall the previous commands used.
State-It gives information regarding the present state of 3D structure image.
Clear-This command clears the console window.

Many more functionalities can be performed in JMOL. The detailed user manual of JMOL can be access through

http://jmol.sourceforge.net/docs/JmolUserGuide/

6. Force Field Calculation of Bonded Atoms

In the context of molecular modeling, a force field refers to the form and parameters of mathematical functions used to describe the potential energy of a system of particles.

Force field functions and parameter sets are derived from both experimental work and high-level quantum mechanical calculations.

The force field calculation option Is available under the tool drop down menu.

The basic functional form of a force field encapsulates both bonded terms relating to atoms that are linked by covalent bonds, and non-bonded (also called "non-covalent") terms. DrugCalc helps in calculating the Force Field caused due to Bonded atoms. The Force Field calculation is done by using AMBER algorithm. The equation used for force field calculation is as below

$$V(r^N) = \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2$$
$$+ \sum_{\text{torsions}} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{i,j} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

Here V (r^N) is the AMBER Force Field. The first summation calculates the energy between covalently bonded atoms where kb is a constant. A user needs to give experimental value of kb while the other two parameters will be fetched from the molecule file. The second summation represents the energy due to the geometry of electron orbitals involved in covalent bonding where ka is a constant. Here too the user needs to give experimental value of ka while the other two parameters will be fetched from molecule file. The third summation represents the energy for twisting a bond due to bond order and neighboring bonds or lone pairs of electrons. Here n is the coefficient of symmetry while V_n depends on value of n. The advantage of force field method - mainly computational simplicity and speed allows Molecular Dynamic runs and Monte Carlo simulations on large systems.

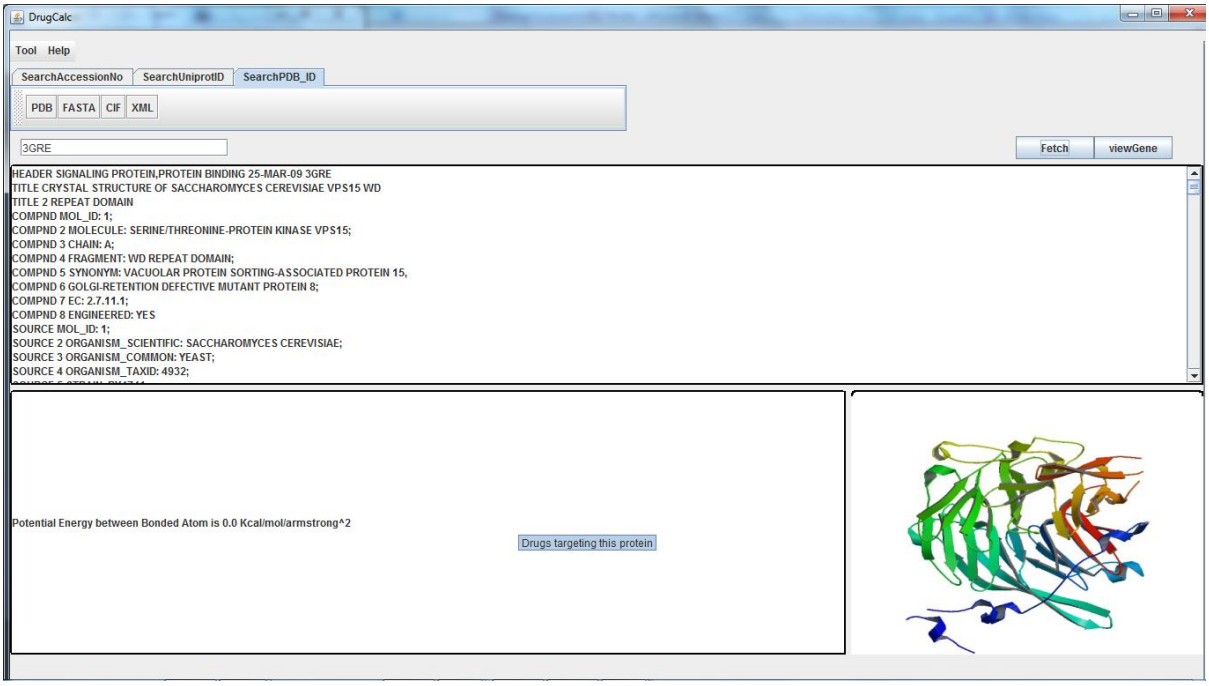

Figure 11: DrugCalc calculating force field due to bonded atoms.

The user can access this facility by providing the experimental value for Kb as well as Ka. Whenever a user fetches protein molecule file using PDB ID the software application will ask for the value of constants. If user is not interested in calculating the force field he can simply select the cancel button. By providing the value of constants, the software application will calculate Force Field due to Bonded atoms for a given protein molecule. The calculated force field can be seen in the bottom left panel in figure 11.

7. Delete History

This is a very simple function. As the name suggests, clear history button simply clears all the files in all the directories. It is like resetting application for the first time use. When clutter of different fetched files of different molecules as well as the 2D and 3D structure view and force field calculation reaches a situation which is out of control for user, this feature really comes very handy. Delete option is available under the tool menu option.

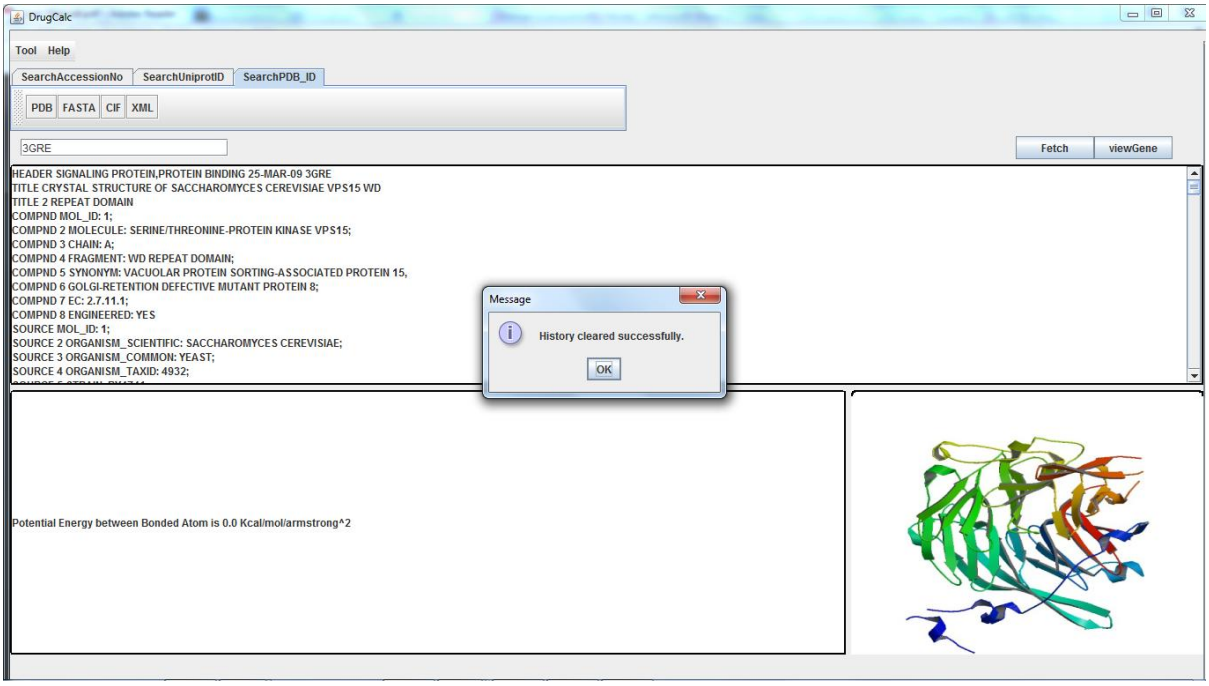

Figure 12: Screenshot showing clear history dialog

A dialog box will appear giving information to user about the report of clear history. Figure 12 shows the message to a user giving information to the user that History cleared command has been executed successfully and history is deleted.

8. Help Menu

"Help" menu is very common in any software. There are five options in the Help Menu.

1. User Manual
2. Software Update
3. Product Demo
4. Product Feedback
5. About

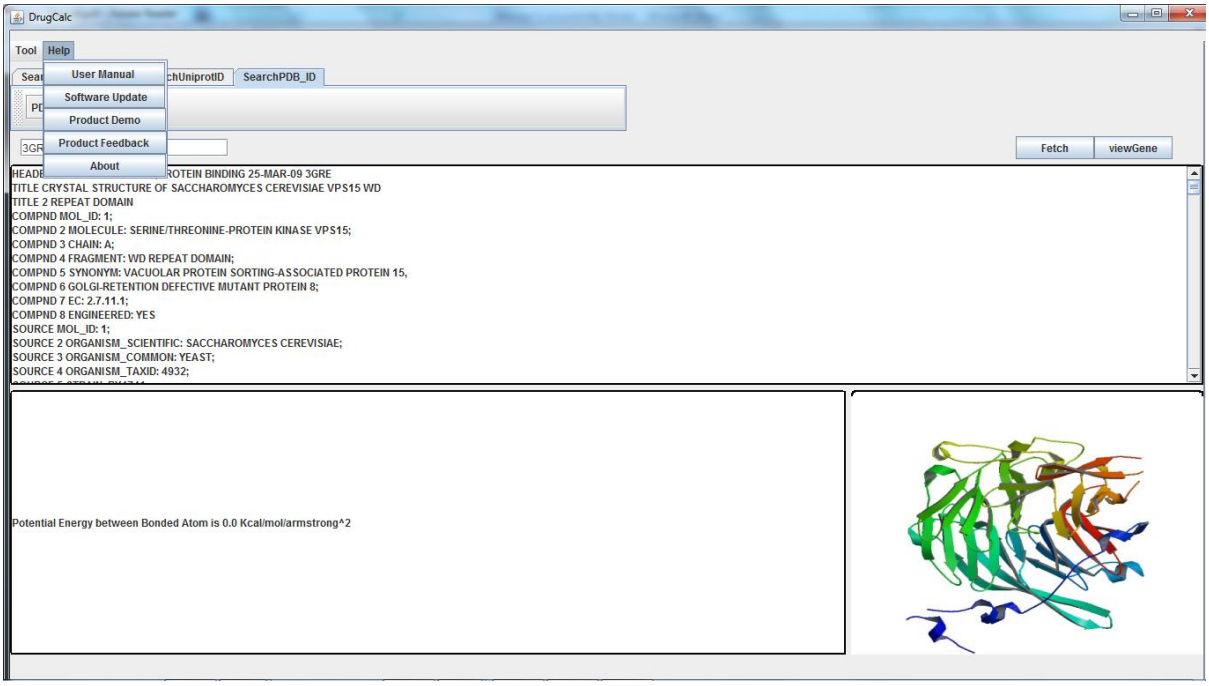

Figure 13: Screenshot showing help menu bar

First option, user manual should open this user manual in the default pdf reader of user's system. Second option is right now not very relevant but it will be once there are more versions released. It opens a default browser in the user's PC and opens a webpage of our application's download page. If user is using older version then he should upgrade to the newer version which can be learned from that webpage. Finally Product Demo option redirects a user to the Product Video demo of the software application which is also available on the DrugCalc Download page.

9. About

About option recognize the contributors in this project. There is a dialog box that opens up as shown below, which contains information like logo of the software, version of the software, name of the software, Credits button and home page of the software.

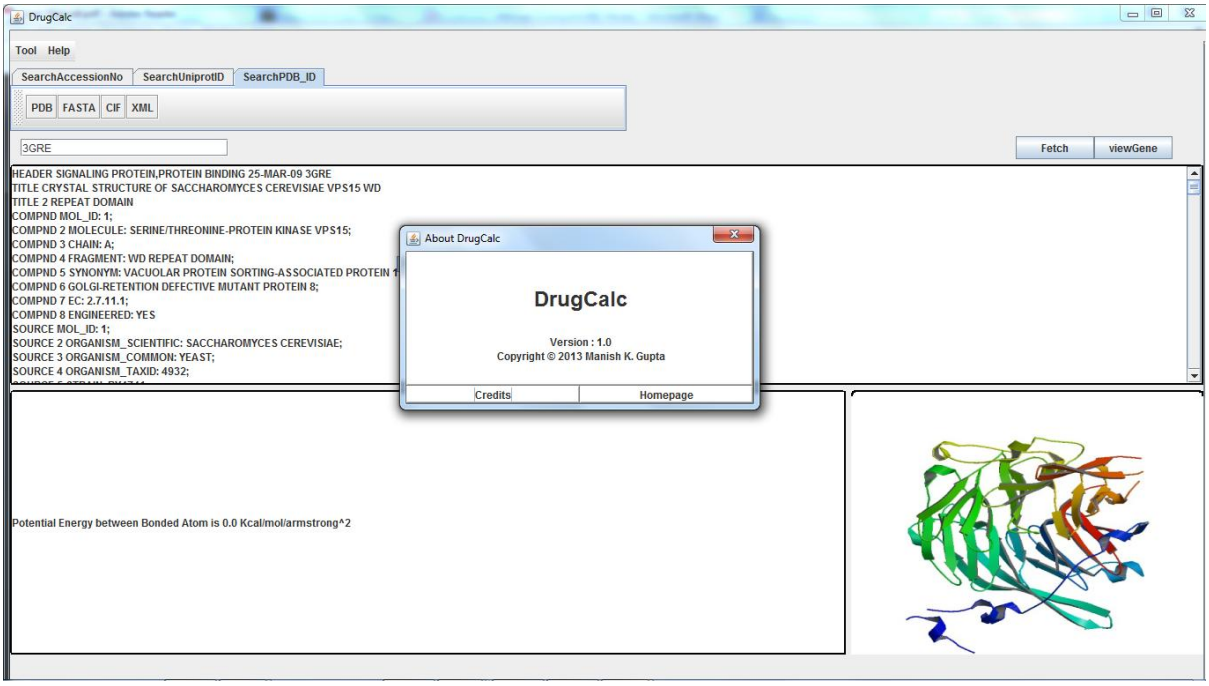

Figure 14: Screenshot showing about menu.

On pressing "Credits", it opens a PDF document in user's default pdf reader. On pressing the home page button a user is directed to the home page of the software where he can check regarding the software updates.

10. Support and Feedback

Product demo video is available at the homepage of DrugCalc as well as on YouTube. Users are requested to contact Manish K. Gupta at the email: mankg@computer.org for feedback and any other issues with the software. Two platform specific installers (Windows and Mac) are available on the project home page along with source code with open source license agreement. We plan to have discussion forums etc. for users at the home page. Enjoy the software!

11.References

[1] alcovelifesciences. (2013) Molecular modeling and drug designing. [Online]. Available: http://www.alcovelifesciences.com/MolecularmodellingandDrug designing.html

[2] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications," Nature Reviews Drug Discovery, vol. 3, no. 11, pp. 935-949, Nov. 2004. [Online]. Available: http://dx.doi.org/10.1038/nrd1549

[3] A. Leach, Molecular Modelling: Principles and Applications (2nd Edition), 2nd ed. Prentice Hall, Apr. 2001.

[4] Scripps. (2013) Autodock. [Online]. Available: http://autodock.scripps. edu/.

[5] Ucsf. (2013) Vega zz manual. [Online]. Available:

http://www.msg.ucsf.edu/local/programs/Vega/pages/glindex.htm.

[6] Wikipedia. (2013) Wikipedia amber. [Online]. Available: http://en. wikipedia.org/wiki/

[7] Gromacs. (2013) Gromacs. [Online]. Available: http://www.gromacs.org/

[8] Wikipedia. (2013) Protein docking. [Online]. Available: http://en.wikipedia.org/wiki/Macromoleculardocking.