CHEMREAD – A CHEMICAL INFORMANT

Jyoti G Angadi[1], Rakesh N.R[2]

[1]Mtech 4th Sem, (Bio Informatics, Department of Biotechnology, KLE Dr M S Sheshgiri College of Engineering and Technology Udyambag, Belgaum, India
[2]Assistant Professor, Department of Bio Technology, KLE Dr M S Sheshgiri College of Engineering and Technology Udyambag, Belgaum, India

Abstract

Recent years have seen explosions in the amount of chemical information are freely accessible in the databases. While these databases has a straightforward web-based user interface for searching the chemical and it is quite unlimited in its facilities where the analysts and inconvenient to process the accessed data CHEMREAD tool constructs information retrieval system that help analysts to access the chemical information which could be a applicable source for validating a compound in the field of Drug Discovery.

Keywords: PubChem, SDF file format, Drug Likeness Properties, .Net Framework.

---***---

1. INTRODUCTION

There is an explosion in the amount of chemical and related biological information in freely accessible databases. The most widely known of these is Pubchem, a repository of over 40 million chemical substances with associated property, literature reference and biological activity information. Pubchem is a huge database for accessing chemical compounds while Pubchem has a straightforward web-based user interface for searching. In addition to being a resource of information about compounds, PubChem database is the primary repository for High Throughput Screening results generated by the Molecular Libraries Screening Centers Network [1]. There is a strong need for tools designed for scientists that permit easy use of these data. We present one such tool, CHEMREAD.

We access a compound by its ID from Pubchem. The downloaded chemical file from the Pubchem database has the chemical information in the text format with .sdf extension. This file has all the information regarding that compound. Researchers find difficult to analyze the accessed data for further analysis like finding drug likeness or properties due to the large amount of the data in chemical files. All the information in SDF file is not required and hence we want to select the fields or queries which will be frequently used by the user. Thus provide the user to select the required data for further work. Hence it helps in easy analysis of the chemical properties of a compound by interactive refinement of searches. The program supports a searching technique for compounds from Pubchem database.

2. OBJECTIVES

- To build a tool which makes easy access of compound properties from compound data file.
- To study the chemical compound files from Pubchem database.
- To analyze the necessary and important data in the file.
- To generate the set of queries.
- To develop a GUI server with an analyzed information.

3. MATERIALS AND METHODS

3.1 Materials

3.1.1 PubChem Compound Database

Pubchem Compound database is a searchable database of chemical structures with validated chemical properties information provided to describe substances in Pubchem Substance. Structures stored within Pubchem Compounds are pre-clustered and cross-referenced by identity and similarity groups. Pubchem Compound includes over 5Million compounds [2].

- Molecular Name Searches (e.g., Aspirin, Benzene) allow searching with a variety of chemical synonyms.
- Chemical Property Range Searches (e.g., Molecular Weight of 100 and 200, Hydrogen Bond Acceptor Count between 3 and 5) allow searching for compounds with a variety of physical/chemical properties, and descriptors.
- Simple Elemental Searches (all compounds containing hydrogen) allow searching with specific element restrictions.

3.1.2 PubChem Compound File Format

The document defines a simple data format (Structure Data Format) for publishing and sharing tabular-style data. Standard database format is used for storing data in a structured format; uses fields of fixed lengths to store values; can be imported and exported by various database

and data analysis programs; commonly used for transferring data between different programs [3].

Chemical data file created in the Structure Data File (SDF) format, which was developed by Molecular Design Limited (MDL) is saved in plain text and contains chemical structure records which is used as a standard exchange format for chemicals information. Molecular Design Limited was renamed to MDL Information Systems and then later was acquired by Symyx Technologies, the organization that now maintains the SDF format.The format's focus is on simplicity and ease of use over the web – that is, usage online with access and transmission over HTTP. In addition the format is focused on data that can be presented in a tabular structure and in making it easy to produce this format from spreadsheets and relational databases[4].

3.2 Methodology

3.2.1 Visual Studio IDE

Visual Studio is an integrated development environment for writing, compiling and debugging the code i.e it provide the complete tool for all process. It provides the editor for the writing code, debugger for debug the errors and compiler to compile that code. In short visual studio provides the complete set for all activities. Visual Studio .NET is a complete set of development tools for building ASP Web applications, XML Web services, desktop applications, and mobile applications. Visual Basic .NET, Visual C++ .NET, Visual C# .NET, and Visual J# .NET all use the same integrated development environment (IDE), which allows them to share tools and facilitates in the creation of mixed-language solutions. In addition, these languages leverage the functionality of the .NET Framework, which provides access to key technologies that simplify the development of ASP Web applications and XML Web services. The Visual C# integrated development environment (IDE) is a collection of development tools exposed through a common user interface. Some of the tools are shared with other Visual Studio languages, and some, such as the C# compiler, are unique to Visual C#[5][6].

3.2.2 ASP.NET with C# in Building Application

Asp.net with C# programming language is used to build CHEMREAD. ASP.NET Web server controls are objects on ASP.NET Web pages that run when the page is requested and that render markup to a browser. Many Web server controls resemble familiar HTML elements, such as buttons and text boxes. Other controls encompass complex behavior, such as a calendar controls, and controls that manage data connections[7].

Server controls are tags that are understood by the server. There are three kinds of server controls:
- HTML Server Controls - Traditional HTML tags
- Web Server Controls - New ASP.NET tags
- Validation Server Controls - For input validation

3.2.3 JavaScript

JavaScript is a client side scripting language. Programs will be passed to the computer that the browser is on, and that computer runs them[8]. JavaScript is used in this tool for validation purpose in client side.

3.2.4 CSS

CSS, abbreviated as Cascading Style Sheets is used to style and control the layout of WebPages. This is used to define the presentation of HTML elements on a webpage. CSS language can be used to define colors, fonts, borders, and the layout of a webpage [9]. CSS is used in this application for designing the web pages with HTML.

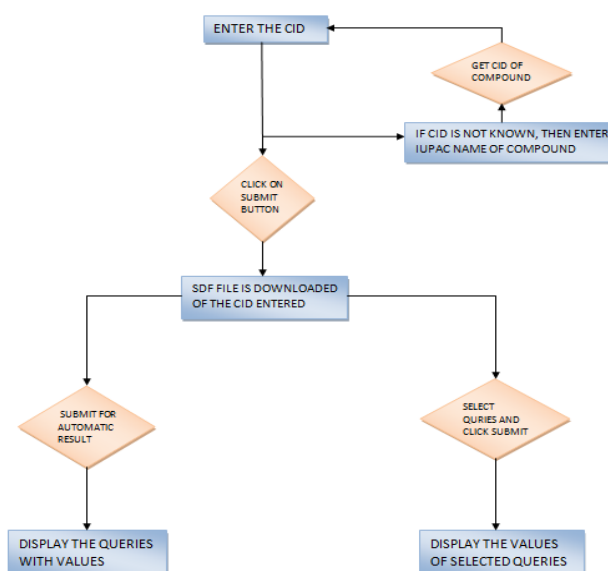## 4. IMPLEMENTATION AND RESULTS

4.1 Work Flow

Fig 1 Work Flow diagram

THE ALGORITHM OF THE FLOW OF THE TOOL WORKS AS FOLLOWS.

Step1. The user will specify the Pubchem CID in the first textbox and clicks on the submit button.

Step2. The application downloads the particular sdf file format of the compound of the specified CID.

Step3. If the user does not knows the CID of the compound, ten we are providing a textbox where he enters the Compound name and it is directed to Pubchem database where user can get the CID of that compound.

Step4. The user gets CID and enters it in textbox and click submit button.

Step5. Then user will be directed to next web page where he is provided to select his queries of interest which is manual method or can view the Automatic result.

Step6. In Automatic method, when user clicks on submit button we are providing the set of queries with their values as result.

Step7. In Manual method, user manually selects the queries of interest and clicks on submit button.

Step8. Then user will be directed to next web page with the results of selected queries.

Step9. User can analyze his results.

Step10. In the next step analysis of chemical information which could be applicable source for validating a compound in the field of drug discovery for example checking the compound whether it can be a lead molecule by checking Lipinski rule of five.

4.2 Implementation Details

In the Implementation chapter, we will be going to give the brief idea about all the pages designed and working of our project. Graphical User Interface (GUI) or Front Page of CHEMREAD is designed using different ASP.Net HTML controls, CSS and JavaScript.Front Page of CHEMREAD have Home link, About Us link, References link, Contact Us link in the Menu Bar and Tool in Home Page.

4.2.1 Home Page

Home page of CHEMREAD is designed by using ASP.Net HTML controls such as image tags, link tags, button tags, div tags, label tags, tables, forms, list tags and different attributes of these control tags which makes a user friendly GUI.

When the user opens CHEMREAD tool, a user friendly GUI is seen where the user is given options to see the links and know about CHEMREAD Tool. User can navigate to all other links to know about the tool and its working.

- Home page link gives the information about the tool.
- User can click the link button TRY OUR TOOL to open the tool.
- User can go to About Us page to view details of CHEMREAD.
- User can view the References page to search for references.
- User can click on Contact Us for the contact details.
- User can view some of the drug likeness properties in the right side of the page. When user moves the cursor over a particular rule ten it gives the definition of that drug likeness rule.
- Lipinski's Rule, Veber Rule, CMC 50 like rule and MDDR like Rule are stated. These will give the idea of drug likeness properties to the user.

Similarly the other pages of the tool such as About Us, References, Contact us and Working of tool is designed and operated.

4.2.2 Try the Tool

This link is available in the Home Page of CHEMREAD; user can click the link to use the tool. CHEMREAD is a tool from which user can download sdf file of a particular compound and fetch the values from sdf file which are further useful to the user.

- User should enter the CID of compound in the textbox provided and Click Submit button.
- If the user does not enter CID and clicks submit then an error message is generated using validation control stating appropriate error message.
- If the user does not know the CID of compound then he can enter the compound IUPAC name in next textbox provided and click on link, it is redirected to Pubchem compound database where user can get the CID of that compound and can copy and enter it and click submit.
- After pressing submit button, it is linked to the Pubchem Compound database and searches the compound of entered CID in Pubchem compound database and downloads the SDF file of the compound.
- Used JavaScript to writ the code for downloading SDF file of the entered CID.
- The SDF file is downloaded in some part of hard disk on the system.
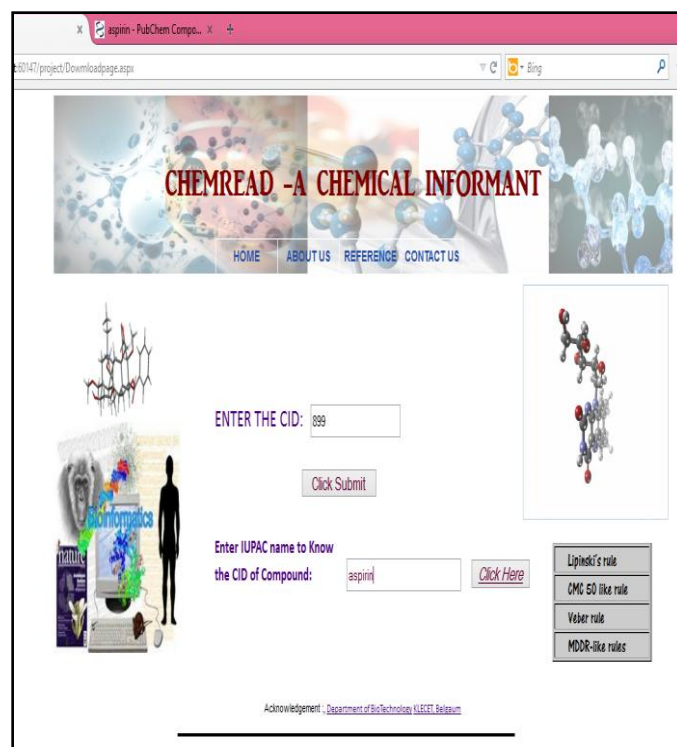- And simultaneously user is redirected to the next page of tool.

Fig2: Screenshot Downloading SDF file by entering CID

After the SDF file is downloaded, we provide user two options:

1) Automatic Result.
2) Manual Result.

We have selected 16 set of queries from the huge SDF file for our tool which we find common and important chemical properties of compound which researchers or users will be commonly using in their work. Following 16 set of queries are selected.

Automatic Result:

- In Automatic Result method, user can view some pre selected queries with their values as result.
- When the user clicks on Automatic Result option, the code is redirected to the downloaded file and fetches the queries line by line and their value with the chemical properties is displayed in the page.
- Used ASP.Net with C# code to read lines from SDF file and fetch the values to display result.
- We have selected some 16 set of queries from the huge query list for automatic result method from SDF file.

Fig 3: Clicking Automatic Result to view the result snapshot

Checking whether the compound satisfies Lipinski,s Rule:
Lipinski's rule states that
1. Not more than 5 hydrogen bond donors.
2. Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms)
3. A molecular mass less than 500 Daltons.
4. An octanol-water partition coefficient log P not greater than 5

- User can check whether the compound downloaded satisfies Lipinski's Rule or not.
- Click on the button provided check for Lipinski's Rule.
- If the compound satisfies the properties it will display as "The Compound satisfies Lipinski's Rule".
- Else it will display "The Compound does not satisfy Lipinski's Rule.
- By checking this rule user can find whether the compound can become a lead molecule in drug discovery.

Fig 4: Check for Lipinski's Rule snapshot

Manual Result:

- In Manual Result method, user is given to choose his queries of interest through checkboxes.
- We have given 16 set of queries from the SDF file to user selection.
- User can check the chemical properties i.e. queries from the checkboxes and click on the submit button.
- When clicked, code is redirected to the downloaded SDF file and searches line by line to get the name of the selected queries and fetches their values and gives the result.
- Used ASP.Net with C# code to read checked lines from SDF file and fetch the values to display result.
- User can check as many queries from 16 sets and all the values of checked queries are displayed.

Fig 5: Select Queries and view their values snapshot

5. CONCLUSIONS

CHEMREAD supports a searching technique for compounds from Chemical Databases like Pubchem. This tool has been built with an application that helps in easy analysis of the chemical properties from a chemical file of Pubchem file format by interactive refinement of searches.

CHEMREAD constructs information retrieval system that helps analysts especially for compounds which are in the queue to become a Ligand. The tool will make to access the chemical information which could be an applicable source for validating a compound in the field of drug designing.

REFERENCES

[1]. Baykoucheva S: A New Era in Chemical Information: Pubchem, DiscoveryGate, and Chemistry Central. ONLINE Sep/Oct edition. 2007.

[2]. Cheminformatics - Basics: Searching for Chemicals on the Internet (Pubchem, eMolecules)Oleg Ursu, Dan C. Fara and Tudor I. Oprea

[3]. Brooks, B.M., et al. (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations". J Comp Chem 4: 187–217.

[4]. Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. (1992). "Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited". Journal of Chemical Information and Modeling.

[5]. Visual Basic 2008 Programming Black Book, Platinum Edition. By Kogent Solutions Inc

[6]. C# Programming Language (Covering C# 4.0), The (4th Edition)
Anders Hejlsberg, Mads Torgersen, Scott Wiltamuth, and Peter Golde.

[7]. Developing Microsoft® ASP.NET Server Controls and Components (Developer Reference) by Nikhil Kothari , Vandana Datye.

[8]. The Good Parts by Douglas Crockford.

[9]. HTML and CSS: Design and Build Websites Jon Duckett