Chem-Bio Informatics Journal, Vol. 3, No. 4, pp.157-162(2003)

Construction of Three-Dimensional Pharmaceutical Structure Database and Its Posting to the Internet

Yoichi Yamamoto, Kaori Kurata, Masaki Hamada and Akira Dobashi*

Tokyo University of Pharmacy and Life Science, School of Pharmacy, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392 Japan

**E-mail: dobashi@ps.toyaku.ac.jp*

(Received September 11, 2003; accepted October 15, 2003; published online December 31, 2003)

Abstract

Our database contains three-dimensional pharmaceutical structures simulated and optimized with the molecular modeling software Macromodel. The database has been compiled from the compounds contained in "Drug in Japan: Ethical Drugs" , edited by the Japan pharmaceutical information center (Tokyo, Japan) in the initial stage of its development. The three-dimensional models were stored on the Web server where they can be viewed, rotated, and animated by the client with either the plug-in, Chemscape Chime, or the Java Applet, Chemis3D. We applied a molecular dynamics simulation to the molecular motion trajectories; three-dimensional animation demonstrates the actual movement and rotation of molecules on their bond axes. This database provides both zipped coordination files indicating conformational distribution for each drug and zipped virtual reality modeling language (VRML) files indicating its electrostatic potential mapped on the electron density surface. A search of the pharmaceutical structures can be conducted through drawing of partial plane structures or input of general names with JChem software to the SQL server containing fully optimized structural data.

Key Words: pharmaceutical structure database, optimized structure, conformation distribution, molecular motion, electrostatic potential map

Area of Interest: Information and Computing Infrastructure for Drug Design and Toxicology

1. Introduction

Three-dimensional structures of drugs composed of single organic compounds, published in "Drugs in Japan: Ethical Drugs", have been calculated using molecular mechanics, molecular dynamics, and molecular orbital methods. These structural data were stored into a Web-enabled server, and opened to the Internet with no limitations when accessed for academic purposes. Our structural database, named the three-dimensional structure database (3DPSD,

http://www.ps.toyaku.ac.jp/dobashi/), allows browser-based Web access to the structural database. By applying Web technology such as Java applets and plug-in software to the Web browser to provide a unique GUI interface, the database becomes independent of any platform. The database provides a fundamental resource for drug development because pharmaceuticals already offered commercially are among the most important lead compounds for developing novel drugs and improving their clinical efficacy.

3DPSD contains the fully optimized structures, molecular motion trajectories, conformational distributions, and electrostatic potentials mapped on the electron density surfaces for Japanese ethical pharmaceuticals. This paper reports the procedures for providing Web access to these structural data and their calculation procedures.

2. Contents of 3DPSD and Molecular Modeling Procedures

2.1 Database Structure of 3DPSD

3DPSD offers two database components to Internet users. One is a hierarchical database arranged alphabetically by pharmaceutical general name, with each file stored into different directories. Another is a database searchable by exact and partial structures on a SQL server.

For the hierarchical database, the information on each pharmaceutical is described in one HTML document. This document is accessible by the first letter of the pharmaceutical general name and can provide the fully optimized, most stable structure in MDLmol file format and molecular motion animation in multi XYZ file format. Along with these structural representations, other information is published in the HTML documents including: general English name, molecular weight, molecular formula, IUPAC name, Japanese trade name, manufacturer, individual pharmaceutical approval number (called YJ code in Japan), logP defined by the logarithm of the distribution coefficient between octanol and water of the drug. LogP is an aspect of Quantitative Structure-Activity Relationship (QSAR), which indicates hydrophobic-lipophilic balance, calculated by the fragment method using the software Pallas PrologP 3.0 (CompuDrug International, Inc., South San Francisco, CA).

These HTML documents contain planar structures capable of being downloaded as two-dimensional MDLmol files, CAS registry numbers corresponding to the general name, the six-letter code for the drug registered in the Cambridge X-ray crystallographic database, and the fully optimized structural data described by the sxf file format capable of being visualized with SPARTAN View software (Wavefunction, Inc.).

2.2 Window Design of 3DPSD and General Operating Feature

The window is divided into three sections: a display of the alphabetic letters, list of general names beginning with the letter selected, and the main section for display of the structure, as shown in Figure 1.

The alphabet is always displayed in an upper part of the window, allowing users to select a drug by choosing the first letter of the drug's general name. Clicking on a particular letter allows the user to select from the list of pharmaceutical general names beginning with this letter, which is displayed on the left side of the window. Upon clicking on the particular name, information about the pharmaceutical is displayed on the right side. This operation initiates the HTML file assigned by the general name inside the local directory containing various structural files. This HTML file displays the optimized structure in MDLmol file format as the default during the first access.

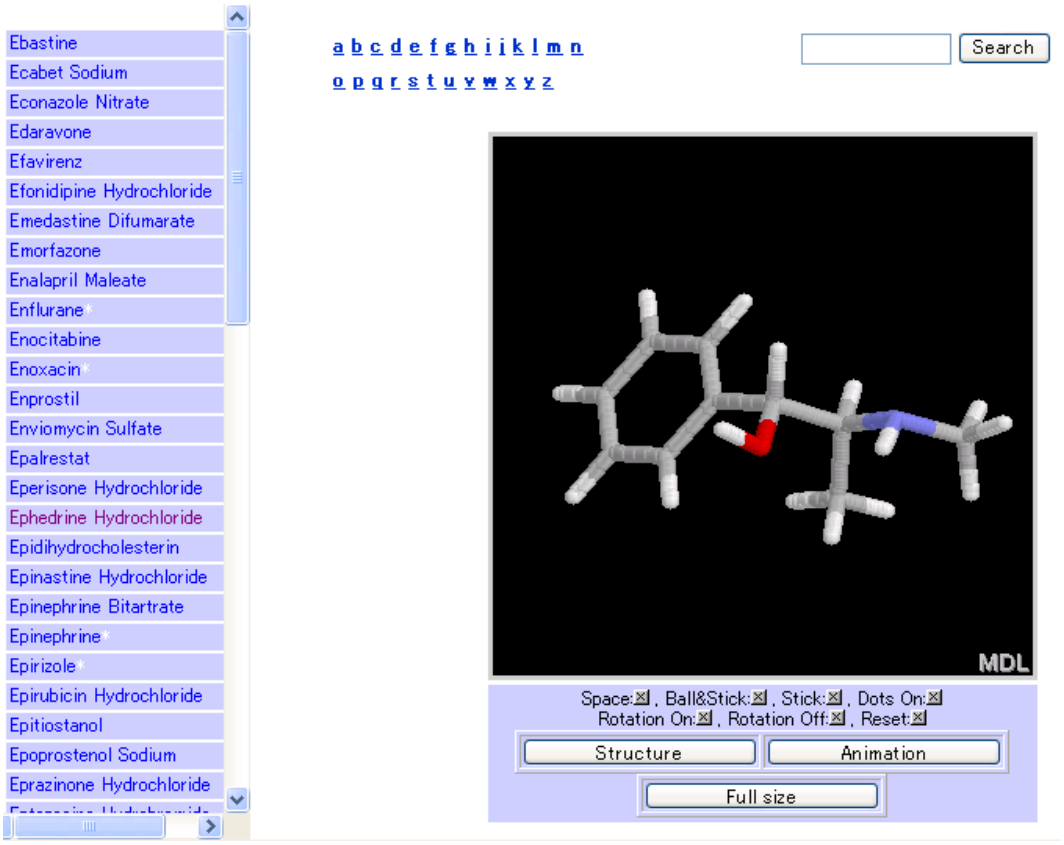

Figure 1. Structure viewing window in a hierarchical database of 3DPSD.

2.3 Structural Data Published in 3DPSD and Related Calculations

3DPSD contains four main pieces of structural information for each pharmaceutical. In the database, pharmaceuticals containing counter ions by salt formation were treated as either free base or free acid with no counter ion.

2.3.1 The fully optimized, lowest energy structure obtained by conformational search

All conformational searching is conducted by the batch-mode molecular modeling facility implemented in MacroModel, Batchmin 4.5 (software provided from the Department of Chemistry, Columbia University, New York. Higher versions are available from Schrödinger, Inc.). MM2* is used as the empirical force field for almost all molecular mechanics and dynamics calculations. MM2* implemented in MacroModel 4.5 is based on Allinger`s 1987 parameter set with many additional parameters. Where MM2* lacks complete force field parameters for the conformational search, AMBER* is used as a substitute in MacroModel 4.5. For all molecular mechanics calculations, a dielectric constant of 1.0 is assumed, steric energies are minimized with a PR conjugate gradient minimizer, and convergence is obtained when the gradient root mean square is less than 0.001 kJ.

To predict the fully optimized, lowest energy structure with flexible pharmaceutical structures, a systematic pseudo Monte Carlo conformation search was conducted using the calculation steps not

less than 1000. This Monte Carlo steps causes an input structure to be modified by random changes in its torsion angles. The most stable conformer obtained by this technique was further subjected to molecular dynamics calculations for 100 ps at 300K under vacuum. A total of 100 structures obtained over the course of this dynamic run were minimized and then the duplicate conformers deleted. The coordinates of conformers obtained in this quick quenching technique were connected into a multifile named sd file, which was compressed into zip format as the conformational distribution. The most stable conformer from among the structures is retrieved as the fully optimized, most stable structure.

Various chemistry software packages have been developed to edit molecular structures and calculate their properties based on molecular mechanics and molecular orbital calculations. Each modeling software package supports several coordination file formats describing molecular structure within more than 50 file formats. MacroModel 4.5 used in this study produces two different types of structural files: Protein Data Bank (PDB) and original macromodel file formats. The coordination file usually is transferred to the plug-in Chemscape Chime distributed by MDL Inc [1]. Thus, at the initial stage of the database development, we applied the MDLmol file format, originated by MDL Inc. itself, to the plug-in. MacroModel file format was converted to MDLmol file format using BABEL2.0 [2], which is public domain software capable of mutual conversion between various coordination file formats. BABEL2.0 was run on the same UNIX operating system (SGI IRIX5.4) on which MacroModel4.5 was running.

2.3.2 Animation of molecular motion

Animation is accomplished by displaying the fifty structures obtained from the time course of the first 50 ps in the molecular dynamics run. The structure coordinates obtained from the dynamic trajectories were converted to XYZ file format using BABEL, connected to one multifile without a separator, and displayed continuously with the Chemscape Chime.

2.3.3 Conformational distribution

he conformational distribution of each drug was presented by MDL sd file, in which the MDLmol files are combined and partitioned by a specific code of "$$$$" in order of its steric energy. These files describe all conformers collected with the quick-quenching technique during molecular dynamics simulation. This multifile was condensed by zip format and stored on the Web page corresponding to the drug.

2.3.4 Electrostatic Potential Mapped on the Electron Density Surface

The electrostatic potential was calculated using PM3 semi-empirical molecular orbital calculations implemented in SPARTAN software (Wavefunction, Inc.) for the fully optimized structure and then exported in virtual reality modeling language (VRML) file format. This format is the descriptive language for utilizing virtual space and is interpreted by the CosmoPlayer plug-in [3] on the browser. This method does not allow secondary utilization of the VRML data by other molecular modeling software. Since the VRML data are not simple numeric tables expressing atomic locations and types, they cannot be utilized by other molecular modeling software or modified according to the end-user`s requirement. Using VRML for the three-dimensional electrostatic potential map provides large data capacity compared to that of the MDLmol file corresponding to the optimized structure. Therefore, the electrostatic potential map data is provided to the end-users as the compressed file in zip format.

The uncompressed VRML data is interpreted by the Cosmoplayer plug-in and displayed on the

client side. The electrostatic potential map showing nuclear charge around the electron density surface is a contour map based on molecular orbital calculations. Energy information on the electrostatic potential map is necessary to examine interactions between the drug and its receptors, and to compare pharmaceuticals with similar or identical physiological action. The concrete numerical value of potential energy is not presented in the electrostatic potential map. Instead, an energy bar divided into 13 different colors is illustrated in the Web page along with the VRML data files. The red zone indicates the maximum value, the blue zone represents the minimum value, and the energy values in the middle energy region are designated by 11 different colors according to potential energy.

2.4 Dispatch of the Three-dimensional Molecular Structure to the Internet

Combining the Web browser and the plug-in expressing the three dimensional molecular structures is one of the most sophisticated techniques for posting them to the Internet, at least during times of minimum traffic along the Internet pathway. Improvements in Internet technology and computer performance have allowed expression of molecular structures with the Java applet as an alternative method for posting the molecular structure to the Internet without the aid of any particular software. In our Web-based database, three-dimensional presentation of molecular structures is accomplished with the following three Web techniques.

1) Installation of the Chemscape Chime plug-in to the Web browser followed by downloading of the coordination data. 2) Downloading of Java applet, Chemis3D 1.94 [4], and the coordination data to the browser, eliminating the need for software on the client side. This applet permits the browser to rotate and magnify the molecular structure. The most recent version of Chemis3D allows the molecular motion animation similar to that provided by Chemscape Chime. 3) Installation of the Cosmoplayer plug-in to the browser followed by downloading of the zip-compressed VRML file and its decompression.

The method using the Chemscape Chime plug-in has been adapted to a number of Web sites for viewing molecular structures, such as the Protein Data Bank. Chemscape Chime can distinguish PDB, MDLmol, and XYZ file formats and automatically display three-dimensional structures on the browsers. When our Web-based database is used with the plug-in Chemscape Chime, the end-user must install it onto the Web browser. Depending on such plug-ins for displaying specific information is not always ideal for sharing information on the Internet. However, since the data transmitted to the client is only the MDLmol file for viewing the optimized structure, this combination provides end-users with the fastest access to the Web-based database.

We used Chemis3D as the Java applet to display the three-dimensional molecular structure to the Web browser. Development of a Web-enabled client/server system with this Java applet was accomplished with the primary aim of requiring that the client only needed to provide a Web browser. This Web technology should improve data access and fulfill the main objective of providing end-users with an easy and efficient access to our Web-based database server.

2.5 Drug Search by General Name and Partial Plane Structure

The alternative structure database constructed with a combination of JChem software [5] and SQL server system allows a search of exact or partial structures through a Web browser. JChem, made by Chemaxon, Inc., is a tool for creating Web applications that access chemical structures and corporate data in a database over the Internet. The MDLmol files corresponding to the fully optimized, most stable drug structures were imported to a SQL server. To construct the SQL server we used the free database software, MySQL [6].

The Web application of JChem provides the structure-description window querying the partial

plane structure and transfers the structural query to the SQL server. Not only the structure, but also the general name and molecular formula, are applied to the query. This search system can be used with both Active Server Pages (ASP) on Microsoft Internet Information Service (IIS) and Java Server Pages (JSP). All optimized structural data stored in the SQL server were converted to Simplified Molecular Input Line Entry Specification (SMILES) file format, which displays the molecular structure by character strings. On the basis of their conversion of the molecular fingerprint, searching among pharmaceutical structures can be conducted as well as the substructure searching through the Web browser.

The compilation of 3DPSD and its posting to the Internet have been supported by the Science Research Fund of the Promotion and Mutual Aid Corporation for Private School in Japan (1997-2003). CD-ROMs containing 3DPSD have been prepared through the financial support of the Foundation for In-service Training and Welfare of the Private School Personnel (Project number, 19: 2002) and Grant-in-Aid for Publication of Scientific Research Results (Project number, 148086:2003).

In 3DPSD, some CA registry numbers and all of the two-dimensional structural data were provided by the National Institute of Health Science, Division of Chem-Bio Informatics, and Division of Organic Chemistry.

References

[1] http://www.mdli.com/.
[2] http://www.eyesopen.com/babel/.
[3] http://www.cai.com/cosmo/.
[4] http://www.schrenzerschule.de/lernen_pc/JAVA-SCHULE/SONSTIGES/Chemis3D164/Chemis3D.htm.
[5] Ferenc Csizmadia, *J. Chem. Inf. Comput. Sci.*, **40,** 323-324(2000).
[6] http://www.softagency.co.jp/MySQL/downloads/index.html.