

MOLDA for Protein Modeling: A Molecular Modeling Program for Biological Molecules

Hiroshi YOSHIDA

Center for Quantum Life Sciences
and
Department of Chemistry, Graduate School of Science
Hiroshima University, Kagamiyama 1-chome, Higashi-Hiroshima 739-8526, Japan
e-mail: yoshida@molda.org

(Received: July 22, 2003; Accepted for publication: August 28, 2003; Published on Web: November 27, 2003)

A molecular modeling program for biological molecules has been developed as a 3D data-interaction tool for computational chemistry and structural bioinformatics. This program named *MOLDA for Protein Modeling* is written in the Java language and works multiplatform environment. MOLDA for Protein Modeling has been implemented with various functions useful for studying biological molecules by structural bioinformatics. The functions are as follows: (1) the importing function of the PDB file is implemented, (2) 3D structure of polypeptides can be generated by inputting the sequences, (3) the conformation of the polypeptide can be changed by reading the dihedral angle matrix, and (4) an amino acid residue can be changed by the point mutation operation. MOLDA for Protein Modeling is expected to be a useful tool for drug design.

Keywords: Molecular modeling, Protein modeling, Protein Data Bank, Point mutation, Bioinformatics

1 Introduction

Quantum theory has brought many breakthroughs in the sciences since the beginning of the 20th century. It has been a driving force for elucidating laws of the molecular world from a theoretical viewpoint. As computers were developed, quantum theory became an important tool for predicting various properties of materials. Along with the development of software for molecular dynamics and molecular orbital calculations in the 1970's, software that allows creation of data for use in calculation and visualization programs has also become important for developing new scientific technologies. MOLDA, developed at the beginning of the 1980's, is the world's first computational chemistry oriented molecular modeling program for PC's [1–3]. Since MOLDA's first release, many molecular modeling applications have been developed, but most of these aim at promoting affiliated computational chemistry programs. In contrast, MOLDA has worked for 20 years to create an interface for computational chemistry programs (freeware programs in particular). It has always changed with the times, and has avoided concentrating on any particular methods in favor of flexibility. We have always striven to keep abreast

of computational chemistry trends and methods, including molecular mechanics calculations, semi-empirical molecular orbital calculations, *ab initio* molecular orbital calculations, and density functional theory. In 1998, MOLDA for Java, a multi-platform molecular modeling program which operates on various platforms without re-compiling, was released [4, 5]. MOLDA is now widely used as a “Hub Program” in the field of computational chemistry as a platform-independent graphical user interface. In the meantime, the new field of bioinformatics has been expanding rapidly along with recent developments in genomics, making the computer a much more important tool in biology than previously. The mysteries of the life sciences are now being revealed using new computer techniques for information processing.

In the present work, *MOLDA for Protein Modeling* is introduced as a molecular modeling program for computer chemistry and structural bioinformatics.

2 Program Overview

MOLDA for Protein Modeling (hereafter “MOLDA”) is written in Java. It is a molecular modeling program, which works on any operating system, including Win-

dows, Macintosh, and Linux, on which the Java Virtual Machine is installed. Here are some of MOLDA's capabilities:

1. Handling proteins of up to 20,000 atoms and 2,000 residues.
2. Loading and saving coordinate data in common formats.
3. Easily creating a molecular model by mouse operations.
4. Loading Protein Data Bank (PDB) [6] format data and displaying sequence information.
5. Performing point mutations for PDB data.
6. Creating polypeptides with structures such as α -helices or β -sheets by entering the amino acid sequence.
7. Exporting data from molecular models created for use in molecular dynamics calculations (TINKER [7], Amber [8]), semi-empirical molecular orbital calculations (MOPAC [9]), and *ab initio* molecular orbital calculations (GAMESS [10], Gaussian [11]), as well as displaying molecular models structurally optimized by these programs.
8. Saving molecular models in VRML created for viewing on the Internet using a Web browser.

As described above, MOLDA contains molecular modeling functions, such as protein modeling, and provides an interface with other major computational chemistry programs. The following section describes MOLDA's functions in detail and gives instructions for their use.

3 Results and Discussion

3.1 Menus

3.1.1 File Menu

In the File menu, one can load and save molecular models in MOLDA format, import and export data in common chemistry and biology formats (PDB, CML [5, 12], MolFile [13]), save data for use in various computational chemistry programs (TINKER, Amber, MOPAC, GAMESS, Gaussian), and use the output of these. The molecular models in VRML can also be saved, so that they may be viewed on the WWW browser. In addition, one may load sequencing files in FASTA format and build three-dimensional polypeptide chain structures. This function allows one to specify the backbone conformation of polypeptides. Commonly used compounds, including fundamental organic compounds,

amino acids, nucleotides, and sugars, are built in as template molecules. The molecular models that have been created may also be saved as substituents with the File menu.

3.1.2 View Menu

In the View menu, one may display or hide the coordinate axes or atomic numbers of selected items. When loading molecular models, this menu allows one to adjust automatically the size and make connections between molecular models. Moreover, when PDB files are loaded, their sequence information and hydrogen bonds can be viewed. For a hydrogen bond to be made, however, the following two criteria must be met: the distance between the hydrogen donor and acceptor (currently oxygen only) must be within 1.2 – 2.76Å, and the angle of X–H...O must be greater than 120 degrees (if hydrogen atoms are not displayed in the PDB file, no hydrogen bond will be displayed).

3.1.3 Model Menu

In the Model menu, one may create alkanes, add hydrogen to skeletal atoms, connect substituents with molecules, replace atoms with other elements, and change bond lengths, bond angles, and dihedral angles. Also, the atoms, atom groups, and bonds may be deleted. Protein modeling functions available from the Model menu allow one to create α -helix polypeptides by inputting one-letter amino acid codes, as well as to perform point mutations for specified amino acid residues.

3.1.4 Display Menu

In the Display menu, one may display molecular graphics using Java3D. This function is enabled only when the Java3D API is installed.

3.1.5 Analyze Menu

In the Analyze menu, one may display coordinate values, interatomic distances, angles among three atoms, and dihedral angles of atoms specified with the mouse in the status bar. For PDB files, matrices for dihedral angles of amino acids (ϕ, ψ) will be output. An animation of molecular dynamics simulations is also available by importing Amber output files.

3.1.6 Help Menu

The Help menu provides program version information.

3.2 Modeling Proteins

MOLDA has functions that allow one to load a PDB file and edit protein structure to build proteins from scratch,

as well as functions for modeling of common organic compounds.

3.2.1 Loading PDB Files and Displaying Sequences

When loading a PDB file, select [File]-[Import]-[Protein Data Bank (*.pdb)]. To view the sequence of the protein, enable [View]-[Sequence Viewer]. After clicking one of the amino acid residues listed in the sequence viewer and then pressing the OK button, corresponding amino acid residues will change color (Figure 1). In order to cancel, click the Cancel button. When clicking an atom in the protein model, the name of the amino acid residue in which that atom is included will appear in the status bar.

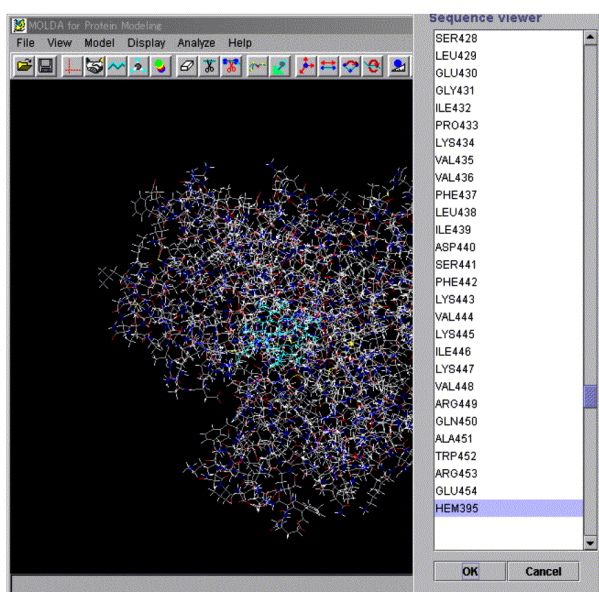


Figure 1. Heme in Cytochrome P450 (2C17)

3.2.2 Building Polypeptides

Any polypeptide chain can be generated in MOLDA by entering its amino acid sequence. The easiest way is to use [Model]-[Input]-[Peptide]. Selecting [Model]-[Input]-[Peptide] will display a dialog box for one letter amino acid codes. If "AAAAAAAAA" is entered, for example, in this box, an alanine nonamer will be created in α -helix structure. Polypeptides may also be built by loading files containing one letter codes of amino acid sequences with [File]-[Import]-[Sequence]. In this method, if the top line of the file contains ">ALPHA", a polypeptide will be created in α -helix structure; similarly, a β -sheet will be generated if the first line contains ">BETA". For example, if the following text file is imported, the polypeptide will be created as shown in Figure 2.

```
>ALPHA
```

```
EEIDQNV
```

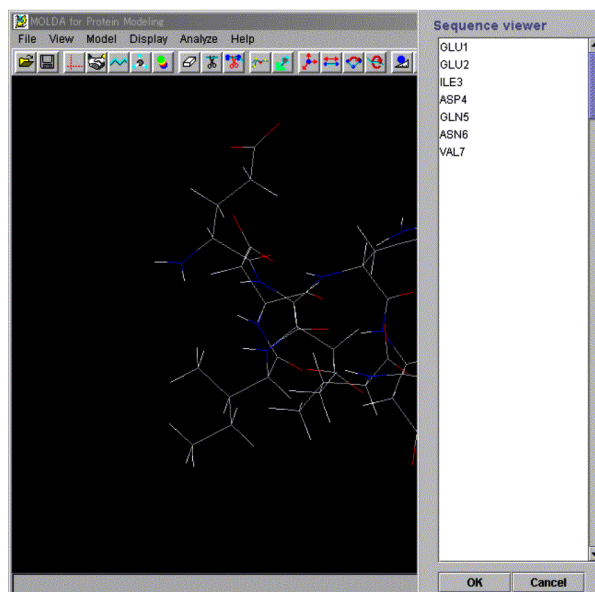


Figure 2. EEIDQNV

A polypeptide in a specified conformation may also be constructed by loading the amino acid sequence together with matrices for dihedral angles (ϕ, ψ). PDB files may be converted to this format by selecting [Analyze]-[Ramachandran]. The created data file in PDB may also be saved.

3.2.3 Point Mutations

MOLDA allows point mutations for a loaded PDB file and amino acid residues of generated polypeptides. To substitute the glutamine (Q/GLN) of EEIDQNV for a histidine (H/HIS) as shown in Figure 3, select GLN5 in the Sequence Viewer and click the OK (GLN5 should be highlighted). When selecting [Model]-[Point Mutation], a dialog box will appear. Then enter "HIS" (the three letter code for the target amino acid residue) to replace the glutamine with a histidine (Figure 4). Note that this function is applicable only to PDB files which include amino acid residues. It may not be applied to data files processed by some MOLDA functions, such as those that attach amino acid residues or remove atom groups (if the PDB file is edited with a MOLDA function that does not support PDB format, it cannot be exported correctly. In that case, it must be saved in MOLDA format). It is also possible to bind the molecules saved in PDB format, and the resulting file may be a PDB data file. The following subsection gives an example of connecting polypeptides with a protein chain while leaving sequence information unchanged.

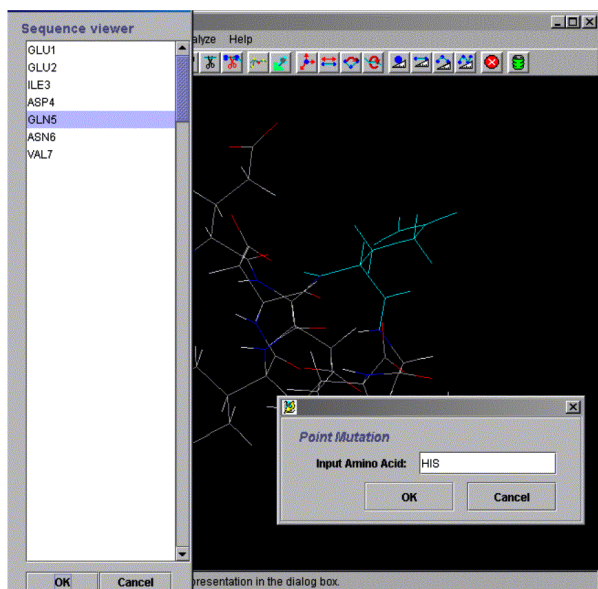


Figure 3. Point mutation

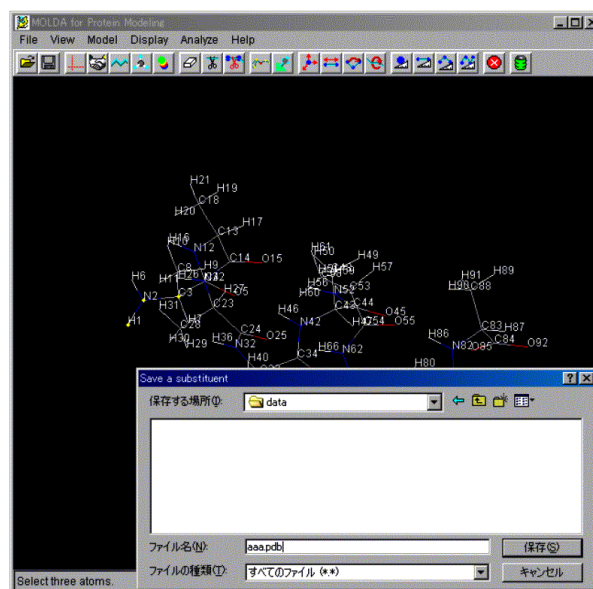


Figure 5. Saving alanine nonamer as a substituent

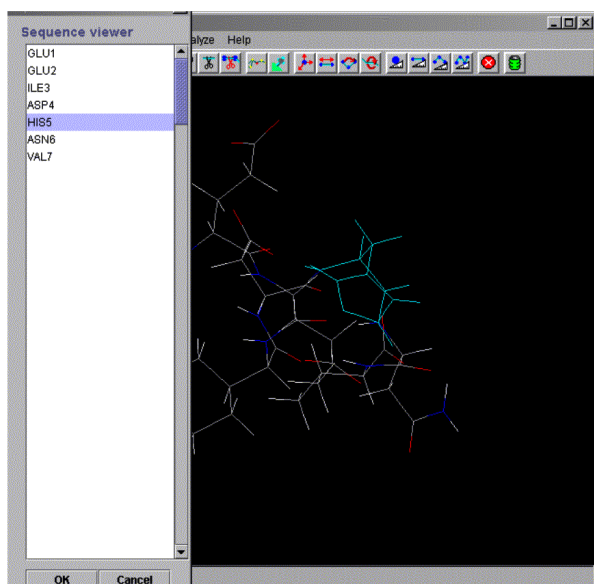


Figure 4. Results screen

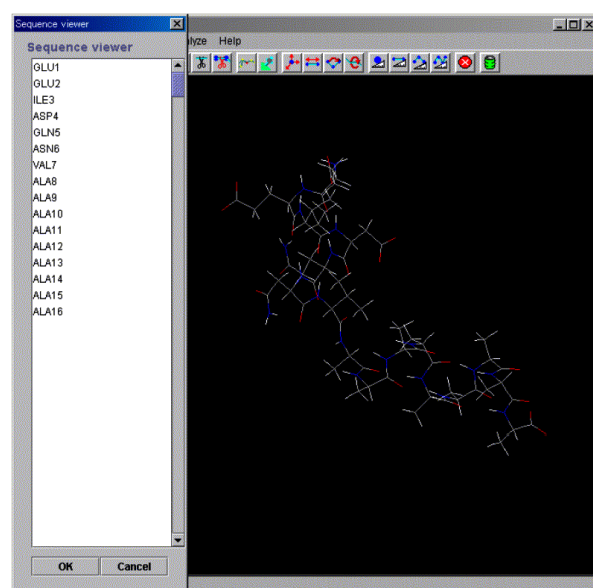


Figure 6. Alanine nonamer is connected

3.2.4 Connecting a Polypeptide Chain

In [File]-[Save as a Substituent] and [Model]-[Merge]-[Substituent], one may use substituents data in PDB format, allowing one to model proteins without loss of sequence information. In this example, the polypeptide chain EEIDQNV in Figure 2 will be connected with an alanine nonamer in α -helix formation. The first step is to create an alanine nonamer as described previously. To view the atomic numbers and the sequence information, enable [View]-[Show Atom Number] and [View]-

[Sequence Viewer] menu. After selecting [File]-[Save as a substituent], use the mouse to select H1, N2, and C3 as target atoms for forming bonds. The three selected atoms should be highlighted in yellow, and a dialog box will appear allowing one to save the alanine as a substituent (Figure 5). If one wishes to save substituent data in PDB format, the file name should include “.pdb” extension. In this example, save the file as “a9.pdb”, clear the data on the screen by clicking the button, then load the EEIDQNV data saved in PDB format with [File]-[Export]-[Protein Data Bank (*.pdb)]. To load “a9.pdb”, select

[Model]-[Merge]-[Substituent] and click the Browse button. Finally, use the mouse to select the O110 atom of the valine in the EEIDQNV chain. Now a polypeptide molecule has been created by connecting the polyalanine molecule to the EEIDQNV chain. One can see 9 alanine added in the Sequence Viewer (Figure 6).

4 Conclusions

The explosive development of computer chemistry and bioinformatics has changed the material and life sciences. MOLDA for Protein Modeling aims to help to unlock the secrets of various materials and life. It will contribute to developing nanotechnology and biotechnology on the basis of first principles by providing protein modeling functions, where computational chemistry and bioinformatics methods are effectively combined: the former allows the predicting of structures and reactivity of molecules based on quantum theory while the latter helps understanding of the program of life written in the molecular language.

The author thanks the Information-technology Promotion Agency, Japan for financial support. MOLDA for Protein Modeling will be distributed from <http://www.molda.org/>.

References

- [1] K. Ogawa, H. Yoshida, and H. Suzuki, *Journal of the Association of Personal Computers for Chemists*, **6**, 21 (1984).
- [2] K. Ogawa, H. Yoshida, and H. Suzuki, *J. Mol. Graphics*, **2**, 113 (1984).
- [3] K. Ogawa, H. Yoshida, and H. Suzuki, *Introduction to Molecular Modeling and Molecular Mechanics Calculations on Personal Computers*, Science House Inc., Tokyo (1986).
- [4] H. Yoshida and H. Matsuura, *J. Chem. Software*, **4**, 81 (1998).
- [5] H. Yoshida, H. S. Rzepa, and A. P. Tonge, *J. Mol. Graph. Mod.*, **16**, 144 (1998).
- [6] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
<http://www.rcsb.org/pdb/>
- [7] J. W. Ponder, *TINKER - Software Tools for Molecular Design version 4.0* (2003).
<http://dasher.wustl.edu/tinker/>
- [8] P. K. Weiner and P.A. Kollman, *J. Comp. Chem.*, **2**, 287 (1981).
<http://amber.scripps.edu/>
- [9] J. J. P. Stewart, *QCPE Bull.*, **9**, 10 (1989).
- [10] M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.*, **14**, 1347 (1993).
- [11] M. J. Frisch et al., *Gaussian 98*, Gaussian, Inc., Pittsburgh, PA (1998).
- [12] P. Murray-Rust and H. S. Rzepa, *J. Chem. Inf. Comp. Sci.*, **39**, 928 (1999).
- [13] <http://www.mdli.com/>

MOLDA for Protein Modeling: 生体分子構築のための 分子モデリングプログラム

吉田 弘

広島大学大学院理学研究科化学専攻
量子生命科学プロジェクト研究センター
〒 739-8526 東広島市鏡山一丁目 3-1
e-mail: yoshida@molda.org

タンパク質の立体構造を予測し設計するためには、計算化学的な手法とバイオインフォマティクス的な手法を効果的に組み合わせて利用することが重要である。本研究では、各種計算化学プログラムのインターフェースとして開発してきた分子モデリングプログラム MOLDA を、構造バイオインフォマティクスによる生体分子の研究に対応できるように改良した。この MOLDA for Protein Modeling の機能として次のものがある。(1) Protein Data Bank 形式ファイルの読み込みが強化された。(2) シーケンス情報を入力することによりポリペプチド鎖の三次元構造を生成することができる。(3) 指定した二面体角をもつポリペプチド鎖の三次構造を生成することができる。(4) 指定したアミノ酸残基に対してポイントミューテーションを行うことができる。以上のような機能が追加されたことにより、MOLDA は今後、計算化学に基づいた創薬を行うための有用なプログラムとなることが期待される。

キーワード: 分子モデリング, タンパク質モデリング, Protein Data Bank, ポイントミューテーション, バイオインフォマティクス