Apr 14th, 12:00 AM - 12:00 AM

A Preliminary Comparative Study of Molecular Visualization Software for Education

Ruoming Shen
Ocean Lakes High School

A PRELIMINARY COMPARATIVE STUDY OF MOLECULAR VISUALIZATION SOFTWARE FOR EDUCATION

Ruoming Shen
Math and Science Academy
Ocean Lakes High School
855 Schumann Drive, Virginia Beach, VA, USA
ruomingshen@gmail.com

ABSTRACT

Chemistry and biology are sciences vital for understanding metabolic processes, developing disease treatments, and improving environmental conditions. With extensive knowledge of biochemistry, we can take advantage of a material's unique chemical composition and properties in various applications. Visualization software is essential for analyzing complex chemical and biological structures and predicting their interactions with each other. This paper presents a preliminary study of three open source molecular visualization software tools - Visual Molecular Dynamics (VMD), Jmol, and Mol*, and evaluates their strengths and deficiencies. This paper utilizes the March Molecule of the Month, Vascular Endothelial Growth Factor (VEGF), from the Protein Data Bank as a case study. The VEGF is a complex signaling protein that stimulates angiogenesis for building new blood vessels. All three tools handle VEGF well by providing a variety of visualization and analytical capabilities, with Jmol and Mol* better suited for educational purposes and VMD for research purposes. The paper also discusses the author's interest and intent to contribute to the development of molecular visualization software in the future.

Keywords: visualization, molecular, chemistry, biology, education

1 INTRODUCTION

As two of the four basic sciences, chemistry and biology play a pivotal role in medicine development, as they provide insights into a drug's reaction withs the human body. As existing diseases persist and new ones emerge, analyzing biochemical structures is more important than ever. Molecular biology allows scientists to tailor medical treatments to a patient's unique genetic composition. Based on the patient's DNA, molecular biologists can predict the patient's response to a drug, thus personalizing the dosages and types of administered drugs. Chemistry forms the basis of the medical field, for chemical reactions govern metabolic processes and drug synthesis. Analytical, organic, and computational chemistry are especially crucial in diagnosing diseases and developing treatments.

A molecule is a group of atoms chemically bonded together that exhibit distinctive physical and chemical properties, ranging from simple diatomic molecules, such as hydrogen gas, to complex macromolecules, such as lipids. Molecules are a fundamental concept in both chemistry and biology; biochemistry analyzes the structures, bonding, functions, and interactions of biological macromolecules, such as proteins, carbohydrates, and nucleic acids. Molecular geometry refers to the 3D arrangement of constituent atoms and chemical bonds in a molecule, dictated by the molecule's electron geometry. Molecular visualization software enables researchers to model, visualize, and analyze complex macromolecules, which would

otherwise be unfeasible to render due to their lengthy chains of sometimes over 100,000 atoms and their intertwined arrangements. Molecular modeling software depicts chemical and biological processes in environments selected by the user, making it the optimal platform to model and experiment with pharmaceuticals. Recent advances in computing hardware and software have magnified the volume and diversity of chemical data and thus have revolutionized the study of molecular biology.

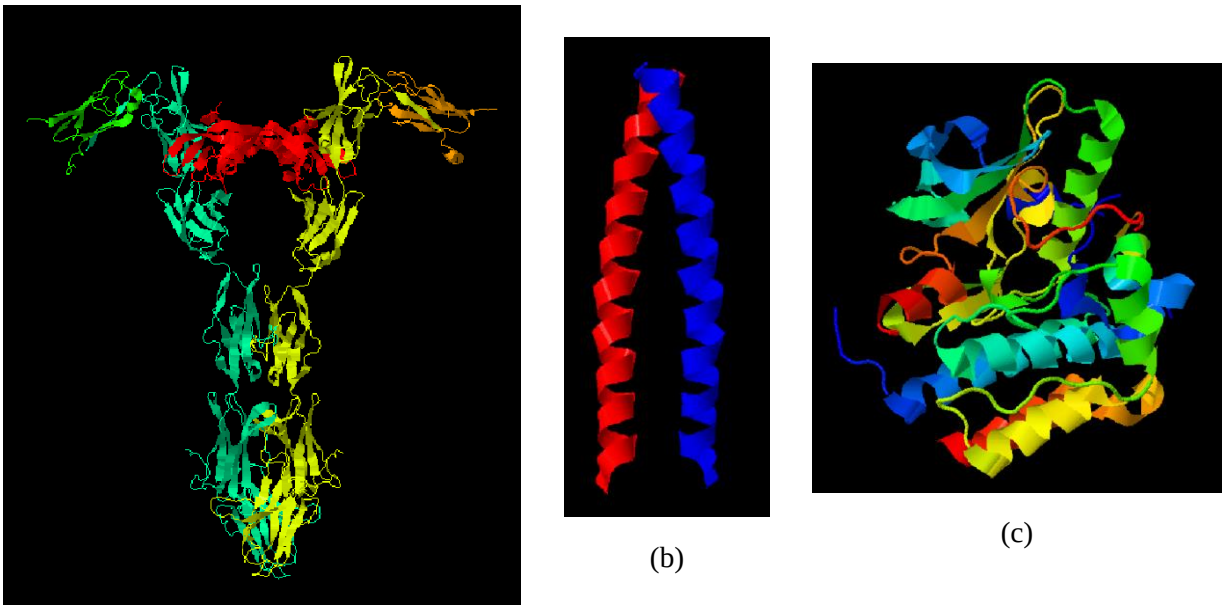
This paper studies three open source molecular visualization software tools–Visual Molecular Dynamics (VMD), Jmol, and Mol*–and evaluates their strengths and deficiencies. The macromolecule Vascular Endothelial Growth Factor (VEGF) from the Protein Data Bank is utilized for a comparative study of VMD, Jmol, and Mol*. All three software handle VEGF well by providing a variety of visualization and analytical capabilities, with Jmol and Mol* better suited for educational purposes. The paper also discusses the author's future plans to improve these software tools.

2 VASCULAR ENDOTHELIAL GROWTH FACTOR

A growth factor is a secreted biologically active molecule that influences the growth of cells (Stone et al., 2021). Vascular endothelial growth factor, or VEGF, is a growth factor for angiogenesis, the formation of blood vessels. Human bodies constantly form new blood vessels to supply sufficient nutrients to cells. Therefore, VEGF is crucial for the proper functioning of human bodies. Normal quantities of VEGF promote wound healing, nutrient replenishment, and blood circulation. However, when overproduced, VEGF can be manipulated and exploited by cancer cells, diverting nutrients from tissues and organs to a malignant tumor. Therefore, regulation of VEGF production is crucial for healthy metabolism. The purpose of this paper is to understand angiogenesis and the role of VEGF in cancer treatment using VMD, Jmol, and Mol*.

VEGF is a signal protein that stimulates the formation of blood vessels. Its corresponding cellular receptor is the VEGF receptor tyrosine kinase (VEGFR), a complex transmembrane protein consisting of three domains: the extracellular ectodomain, the transmembrane domain, and the intracellular tyrosine kinase domain, as illustrated in Figure 1. It should be noted that Figure 1 contains a pair of two VEGFR that is activated by receiving the signal emitted by VEGF. During its inactive state, the VEGFR exists as a single unit (Cartagena et al., 2022). The VEGF, shown in red color in Figure 1(a), initiates the angiogenesis process. After VEGFR binds with VEGF, the VEGFR pairs off with another VEGFR, creating an active dimer, as shown Figure 1(a). The extracellular ectodomain shown in Figure 1(a) consists of 11,357 atoms, 11,616 bonds, 6 chains, and 14 polymers. The dimerization process continues by passing through the transmembrane domain, as shown in Figure 1(b), and joining the two tyrosine kinases in the intracellular domain, as shown in Figure 1(c) (only one tyrosine kinase is shown in Figure 1(c)). The two tyrosine kinases then activate each other and other signaling proteins inside the cell. Some of the signaling proteins travel to the nucleus and change the expression of genes involved in angiogenesis. Other actions of the signaling proteins, such as phosphorylation of the interior portion of cadherin proteins, generate openings between two cells, which attract precursor cells that form the lining of a new blood vessel (Cartagena et al., 2022).

The VEGF and VEGFR were featured as the Molecule of the Month for March 2022 by PDB-101, an online protein and nucleic acid portal available to teachers, students, and the general public (Protein Data Bank, 2022a). PDB-101 was developed by the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). The RCSB PDB operates the US data center for the global PDB archive and makes PDB data accessible to all users without limitations on usage (Protein Data Bank, 2022b).



(a) (b) (c)

Figure 1. Active complex of VEGF and VEGFR. (a) The extracellular ectodomain: VEGF (red) and VEGFR (green and yellow) (PDB ID 5T89). (b) Transmembrane domain (PDB ID 2M59). (c) Intracellular tyrosine kinase domain (PDB ID 3HNG). All visualizations were generated by the author using Jmol.

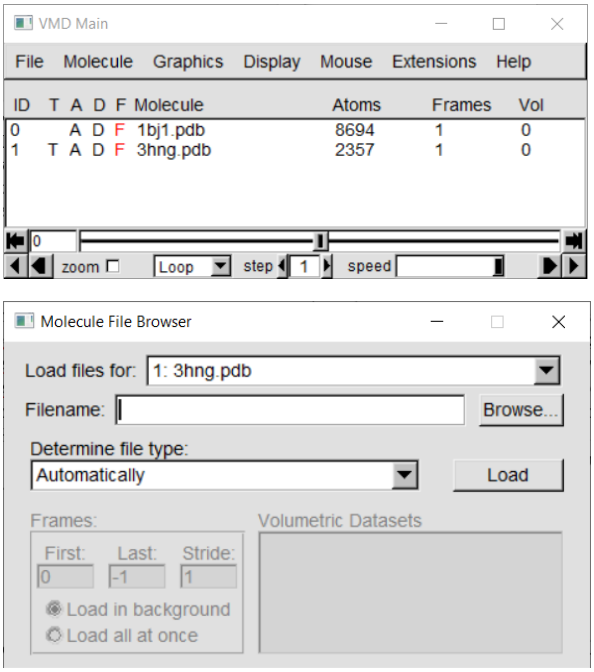
3 VISUAL MOLECULAR DYNAMICS

Visual Molecular Dynamics (VMD) is an open source molecular visualization program displaying, animating, and analyzing biomolecular systems using 3-D graphics and built-in scripting (Theoretical and Computational Biophysics Group, 2008). It was developed by the Theoretical and Computational Biophysics Group at the University of Illinois at Urbana-Champaign (UIUC), a leading university in high performance computing (Theoretical and Computational Biophysics Group, 2022). VMD can be used to visualize standard PDB files and more general molecules. VMD can model and analyze both individual molecules and biological systems with intermolecular interactions. It is compatible with PDB files and can display and manipulate various chemical structures. The VMD molecular drawing styles, such as van der Waals, Licorice, Tube, and Backbone, reveal different properties of a selected molecule. The NewCartoon representation, for instance, distinguishes the protein secondary structures of alpha helices and beta sheets, which appear as coiled ribbons and arrows respectively, from the remainder of the chemical structure, which appear as solid tubes. Further, this software can animate chemical pathways and molecular dynamics, producing reaction simulations. Users can alter chemical structures and analyze properties through either Tcl/Tk scripting or a graphical interface. VMD offers the following main features:

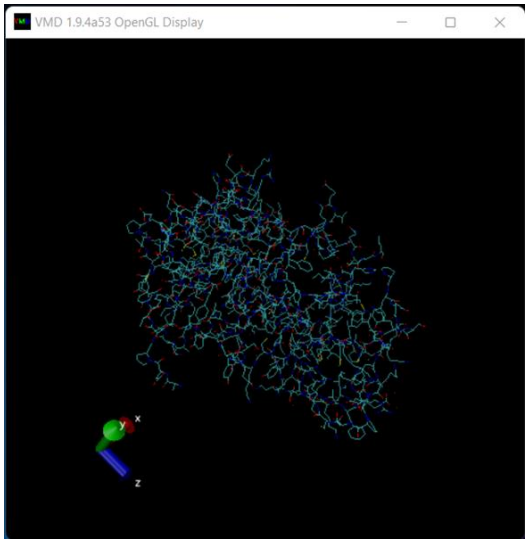
- VMD is open source software written in C++ that supports all major operating systems, such as Windows, Linux, and MacOS.
- VMD supports multicore processors and GPU accelerated computation.
- It has no limits on the number of molecules, atoms, residues, and number of trajectory frames, except available computer memory.
- It provides a plethora of molecular visualization capabilities, such as numerous molecular rendering and coloring methods, stereoscopic displays, and multiple sequence alignment.
- It supports more than 60 molecular file formats and provides an extensive library of built-in file readers, writers, and translators.

- It provides both graphical user interfaces and text-based user interfaces that can be extended by the user using standard scripting languages such as Tcl/Tk and Python.
- It can be used to concurrently display and interact with the NAMD Interactive Molecular Dynamics Simulation developed by the Theoretical and Computational Biophysics Group at UIUC.

While VMD is an extremely powerful molecular visualization program, there are issues with its usability. In particular, its visualization window is separate from its main window and other control windows, as shown in Figure 2 and Figure 3. The separate VMD windows are scattered at different locations on the screen, complicating the user interface. The user may struggle to find and navigate to the control window. Furthermore, when implementing different visualization parameters in the control window, the user must switch quite frequently between the control window and the visualization window to examine the impacts of these parameters, making the user interactions very cumbersome. VMD does not support some standard user interactions that are common in computer software, such as a context menu activated by a right mouse button click to display various options. Many commands provided in the VMD Main window (shown on the top of Figure 2(a)) should be available on a context menu.



(a)



(b) 3HNG in VMD

Figure 2. Screen capture of VMD. (a) Control windows. (b) Visualization window.

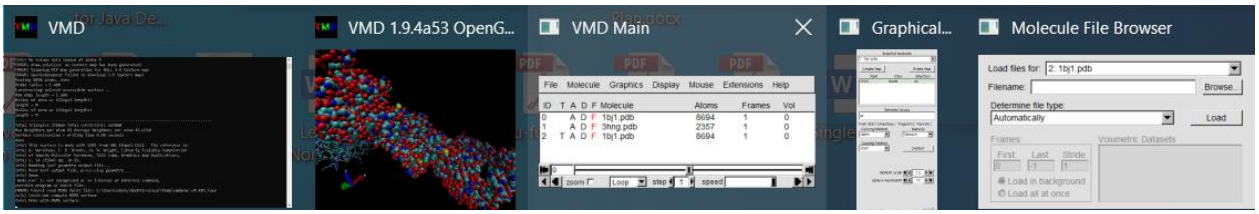


Figure 3. Icons for different windows of VMD on the Task Bar on Windows 10. The actual VMD windows are scattered at different locations on the screen.

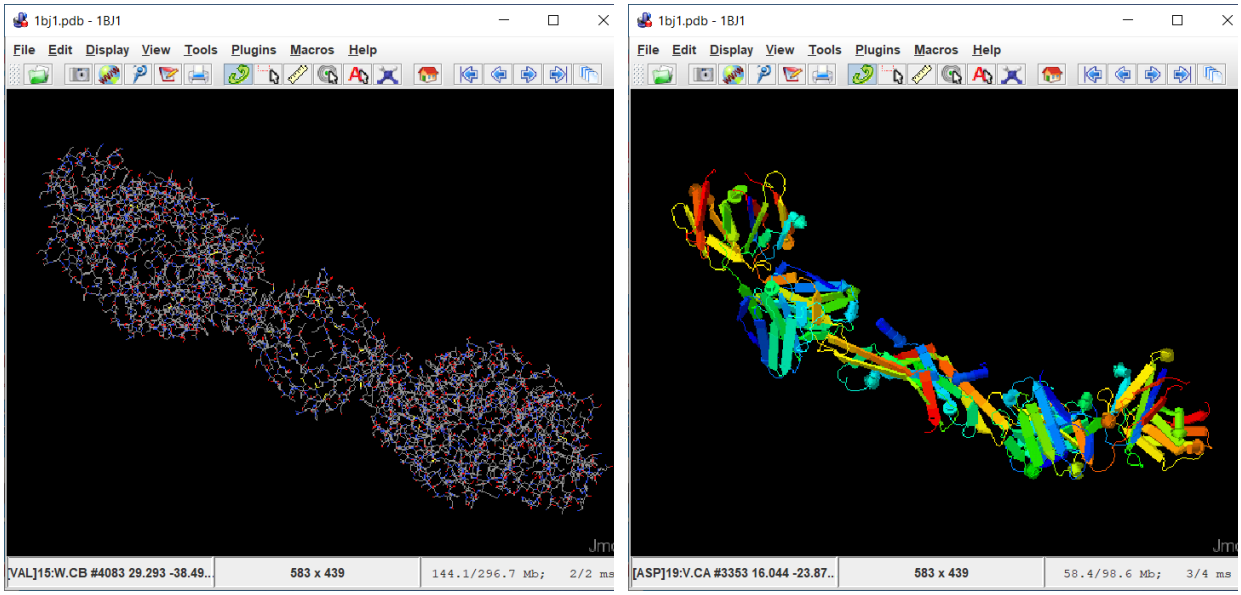
Three common user interactions in interactive 3D modeling are zoom, orbit (or rotate), and pan (or translate). The user uses zoom to focus on an area of interest, orbit to exam different parts (orientations) of the area of interest, and pan to move to a new area of interest. Usually, these three user interaction modes are associated with different mouse buttons (left, middle, and right) or combinations of mouse buttons and key presses, e.g., ALT + left mouse button. However, in VMD all the user interaction modes are associated with the left mouse button and the user has to go to the VMD Main window to change user interaction mode. In addition, when the user pans to a new area of interest, the pivot for orbit should be centered about the new area, usually at the center of the screen. However, the pivot for orbit in VMD is fixed, yielding unexpected orbit results and frustrating user experience. Given the importance of usability in interactive visualization, the author feels that VMD is not well suited for novice users, such as students learning chemistry or biology. The overall user interface design and usability of VMD must be vastly improved.

4 JMOL

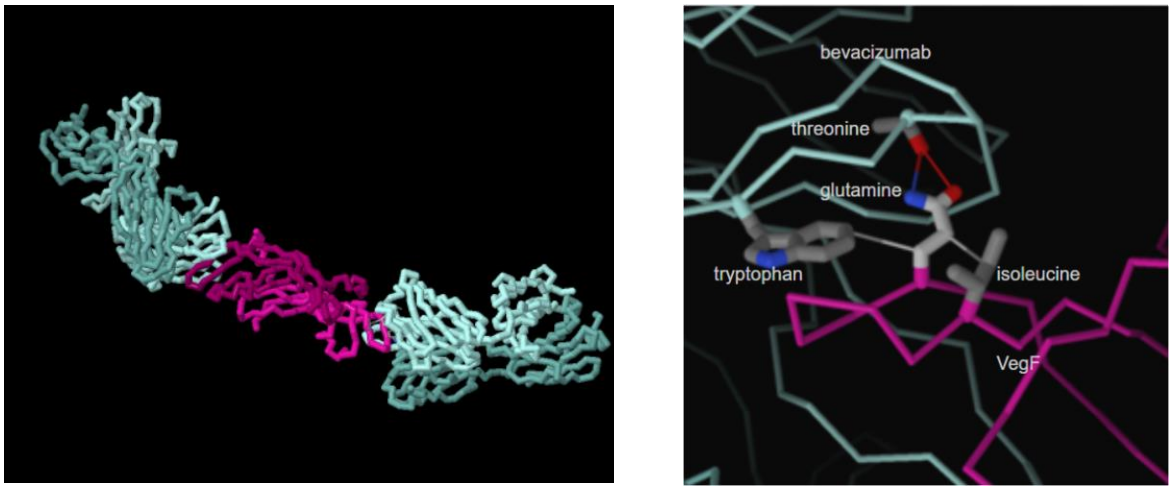
Jmol is an interactive viewer for 3D chemical structures. It is an open source software project developed by a team of about 40 software developers and hosted on SourceForge (Herraez et al., 2022). Jmol can run in four modes: a stand-alone Java program, an HTML5-only web application, a Java applet, and a "headless" server-side component. Jmol displays various representations of chemical structures, such as Cartoon and Backbone. Jmol supports various file types, such as PDB, Spartan, and PSE, and users can directly upload files from chemical databases, such as PubChem, into Jmol. Interactive animations, linear morphing, and measurements are prominent capabilities of Jmol. This software employs the RasMol/Chime scripting language. In addition, Jmol allows for quality spectroscopy, 2D to 3D conversion, and 3D printing through its compatibility with JSpecView, JSME, and CAD programs, respectively. Its major features include the following:

- Jmol is open source software written in Java that supports all major operating systems such as Windows, Linux and MacOS, web browsers, and mobile devices.
- It offers an active, world-wide user group with approximately 100,000 downloads per year (Herraez et al., 2022). It is used by many universities such as University of Pennsylvania (University of Pennsylvania X-Ray Crystallography Facility), Miami University (Miami University), University of Kent (University of Kent), and University of Liverpool (University of Liverpool, 2004).
- It supports HTML5 for all modern browsers, including these on mobile platforms such as iOS and Android. It uses customized, optimized Java-to-JavaScript compilation for pure HTML5 app.
- It provides a rich scripting language and web API for easy customization of the user interface.
- It supports more than 60 file formats, and can load and compare multiple files.

The standalone application of Jmol provides an integrated user interface that consists of major controls (in the forms of a menu bar, a tool bar, and a context menu activated by right mouse button click) and the visualization window, as illustrated in Figure 4(a) and (b), which depict different representations of VEGF in complex with a neutralizing antibody (PDB ID: 1BJ1). Bevacizumab is a cancer treatment drug and is also known by its brand name, Avastin. Bevacizumab is an antibody that binds to VEGF and thus prevents VEGF from binding to the VEGF receptor tyrosine kinase (VEGFR). Consequently, VEGFR hinders the process of angiogenesis, for it cannot activate another VEGFR and other signaling proteins inside the cell. As a result, the inhibition of blood vessel formation deprives the tumor of sufficient oxygen and nutrients, preventing its proliferation to other parts of the body. Figure 4(c) shows that one VEGF molecule binds with two Bevacizumab molecules rather than VEGFR. Figure 4(d) shows the interactions around a glutamine necessary for the binding between VEGF and Bevacizumab including hydrogen bonding with a neighboring threonine, and hydrophobic interactions with tryptophan and isoleucine (Cartagena et al., 2022). It should be noted that while Bevacizumab effectively inhibits cancer cell proliferation, it can also cause serious side effects, such as slowing down wound healing (Cartagena et al., 2022).



(a) (b)



(c) (d)

Figure 4. Jmol illustration of Bevacizumab (Avastin) for cancer treatment. (a) Ball and stick representation of VEGF in complex with a neutralizing antibody (PDB ID: 1BJ1). (b) Monomer representation within Jmol. (c) Binding of two Bevacizumab molecules and one VEGF molecule, tube representation. (d) Three critical interactions for the binding.

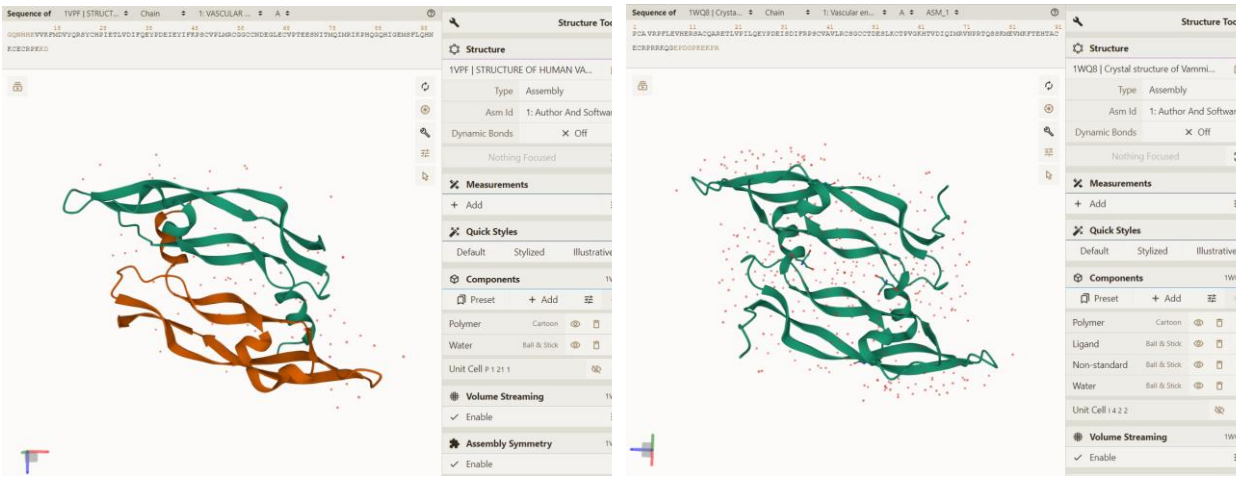
It should also be noted that Jmol does not support pan (or translate) using mouse directly. Instead, the user must pick a point in the new area of interest. The selected point becomes the new center of the window, thus effectively "moving" the new area of interest to the center of the window. In addition, there are multiple modes for the picking operation, including Center, Identify, Label, among others. Jmol also demonstrates some issues refreshing menus and the main visualization window. The user must manually refresh the window by resizing it frequently.

5 MOL*

Mol* is a modern web-based open-source toolkit for visualisation and analysis of large-scale molecular data. It was developed jointly by PDBe (Protein Data Bank in Europe) and RCSB PDB based on their

previous molecular visualization software, namely, LiteMol developed by PDBe and NGL by RCSB PDB (Sehnal et al., 2021). It is hosted on GitHub.com (Protein Data Bank in Europe & RCSB Protein Data Bank, 2022). Mol* affords high-performance graphics and data handling and analytics capabilities that allow users to simultaneously visualize protein structures, render cell-level models at atomic level, and play molecular dynamics simulations. Mol* offers the following major features (Sehnal et al., 2021):

- Mol* is open source software written in JavaScript, HTML, WebGL, Node.js, and TypeScript that runs inside a web browser on all software and hardware platforms, including mobile devices.
- It provides a sleek user interface, as shown in Figure **5**. It offers high-quality rendering with numerous advanced capabilities.
- It has unique sequence view and molecular component focus tools to facilitate navigating the structure and making selections.
- It allows alignment of molecules with sequence-guided pairwise alignment and alignment by manual selection of corresponding atoms.
- It enables various geometric measurements and labels and generates high quality screenshots and animation exports.
- It supports a large collection of file formats that represent structures, volumes, trajectories, and generic triangle geometries.



(a) (b)

Figure 5. Visualization in Mol* (Protein Data Bank in Europe & RCSB Protein Data Bank, 2022). (a) Human VEGF (PDB ID: 1VPF). (b) VEGF from a snake venom (PDB ID: 1WQ8).

As discussed in previous sections, VEGF is essential for angiogenesis. While VEGF's functioning should be restricted to prevent tumor progression, it should be promoted for other bodily processes, such as wound healing (Cartagena et al., 2022). Figure **5**(a) shows a human VEGF (PDB ID: 1VPF) while Figure **5**(b) shows the VEGF from a snake venom (PDB ID: 1WQ8). It can be seen that these two VEGF exhibit highly similar molecular structures and sequences. Scientists are investigating the possibility of using the latter for wound healing by inducing similar activation of VEGFR (Cartagena et al., 2022).

Mol* provides a visually appealing and responsive user interface with carefully designed text styles, icons, and layouts. It supports pan by the mouse right button directly as well as pan on mobile devices by two fingers. RCSB PDB and PDBe use variants of Mol* as their standard viewers for proteins.

6 DISCUSSIONS AND CONCLUSION

This paper presented a preliminary study of three leading software tools for molecular visualization: Visual Molecular Dynamics (VMD), Jmol, and Mol*. The Molecule of the Month published by PDB-101 in March 2022 (Cartagena et al., 2022) was selected in this study to discuss its molecular structure, the angiogenesis process, and its applications for cancer treatment and wound healing. Table **1** summarizes the major features offered by these software tools and presents a high-level comparison of them. The following discussions provide further insights regarding their usability and applicability for educational use.

While all three software tools are open source software, VMD is written in C++, Jmol is written in Java, and Mol* is written in JavaScript, HTML, and WebGL in addition to a few other software libraries. The installation files for VMD vary with the computer platform used, i.e., different platforms (Windows, Linux, and MacOS) require different installation files. In contrast, as a Java program, Jmol only provides one installation file. Mol* does not require installation at all, as it runs inside a web browser directly. As a result, Mol* requires Internet connection, while both VMD and Jmol can run as standalone applications without network connection. Mol* can therefore pose issues for students without reliable Internet connection.

VMD provides the most extensive and advanced capabilities, with 26 extensions for analysis and 17 extensions for modeling. VMD promotes greater customization of molecular models, as it allows the user to combine multiple drawing methods, such as points and ribbons, in one structure. With its myriad of analysis operations, VMD is optimal for research and high performance computing. VMD is utilized by various universities and laboratories in the world. However, VMD does not have a well-designed user interface; thus, it is not well suited for educational use. In addition, since fully exploiting VMD's features requires an advanced understanding of quantum chemistry and organic chemistry, VMD is inappropriate for basic chemistry instruction. A simpler interface and limited capabilities render Jmol more user friendly than VMD. Jmol is ideal for educational purposes, for it specializes in molecular viewing rather than analysis. Students can explore the structures of chemicals more easily in Jmol than in VMD. Conversely, users glean more knowledge about chemical properties from VMD compared to Jmol. Mol* offers the latest technology and the best user interaction among the three software tools explored in this study. However, its user interface is relatively complex and may be excessive for studying relatively simple chemical structures. Mol* and Jmol are the molecular visualizers recommended by RCSB PDB (Protein Data Bank, 2021). Overall, the author concludes Jmol is best suited for visualizing relatively simple molecular structures, while Mol* is ideal for more complex structures.

In addition to exploring chemistry and biochemistry through these software, the author is also interested in further development and expansion of such software tools. Jmol provides the best option for novice software developers like the author given that Java is the only programming language involved in its development and it has the same cross-platform deployment. The author intends to read the source code of Jmol and attempt to address some existing issues in Jmol.

Table 1: Comparison of VMD, Jmol, and Mol*.

Features	VMD	Jmol	Mol*
Cross-platform	Yes	Yes	Yes
Programming Language(s)	C++	Java	JavaScript, HTML, WebGL, Node.js, TypeScript, and React
Open Source Code	Yes	Yes	Yes
Standalone Execution	Yes	Yes	No
Software Complexity	High	Moderate	High
Developer Community	Very Active	Very Active	Very Active
Web Support	No	Yes	Yes
Mobile Support	No	Yes	Yes
Usability	Poor	Good	Excellent
Rendering	Good	Good	Excellent
Analysis	Excellent	Good	Good
Play Molecular Dynamics Simulation	Yes	Yes	Yes
Overall Quality	Good	Good	Excellent
Suitable for Educational Use	No	Yes for less complex molecules	Yes for complex molecules

7 ACKNOWLEDGMENT

The author would like to thank Dr. Michel Audette for his guidance and help on this work. The author was mentored by Dr. Audette in the summer of 2021 to explore various molecular visualization software.

REFERENCES

Cartagena, E., Gelashvili, M., Keyes, J., Rosenzweig, E., Goodsell, D. S., & Burley, S. K. (2022). *Vascular endothelial growth factor (VegF) and angiogenesis*.

Herraez, A., Hanson, B., Lancashire, R., & Kanters, R. (2022). *Jmol: An interactive viewer for three-dimensional chemical structures*. Retrieved March 23 from https://sourceforge.net/projects/jmol/

Miami University. *Research Computing Cluster Data*. https://sites.miamioh.edu/researchcomputing/hpc-software/jmol/

Muller, Y. A., Christinger, H. W., & De Vos, A. M. (1998). *Vascular endothelial growth factor in complex with a neutralizing antibody*. Protein Data Bank. http://doi.org/10.2210/pdb1BJ1/pdb

Muller, Y. A., & De Vos, A. M. (1997). *Structure of human vascular endothelial growth factor*. Protein Data Bank. http://dx.doi.org/10.2210/pdb1bj1/pdb

Protein Data Bank. (2021). *Jmol*. https://www.rcsb.org/docs/3d-viewers/jmol

Protein Data Bank. (2022a). *About PDB-101*. https://pdb101.rcsb.org/more/about-pdb-101

Protein Data Bank. (2022b). *About RCSB PDB: Enabling Breakthroughs in Scientific and Biomedical Research and Education*. https://www.rcsb.org/pages/about-us/index

Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S. K., Koča, J., & Rose, A. S. (2021). Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research, 49*(W1). https://doi.org/10.1093/nar/gkab314

Stone, W. L., Leavitt, L., & Varacallo, M. (2021). *Physiology, growth factor*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK442024/

Theoretical and Computational Biophysics Group. (2008). *What is VMD?* https://www.ks.uiuc.edu/Research/vmd/allversions/what_is_vmd.html

Theoretical and Computational Biophysics Group. (2022). *VMD: Visual Molecular Dynamics*. Retrieved March 27 from https://www.ks.uiuc.edu/Research/vmd/

University of Kent. *Jmol*. https://www.kent.ac.uk/software/jmol

University of Liverpool. (2004). *Jmol animations*. https://www.liverpool.ac.uk/~ngreeves/ltfdemo/jmol/EpoxideOpen.html

University of Pennsylvania X-Ray Crystallography Facility. *About the Jmol applet*. http://macxray.chem.upenn.edu/Jmol.html

AUTHOR BIOGRAPHIES

RUOMING A. SHEN is a junior at the Math and Science Academy of Ocean Lakes High School in Virginia Beach, Virginia. Ms. Shen has been on the Principal's List every year since 2016. She was selected to attend the prestigious Virginia Summer Residential Governor's School in the summer of 2021. She was the recipient of the STAR Scholarship from National Math + Science Initiative and the 2021 Youth of the Year award from the Organization of Chinese Americans Eastern Virginia Chapter. She was recognized for her Perfect Score in the SAT test by Virginia Beach City School Board in March 2022. Ms Shen's research and career interests include computer science, modeling and simulation, chemistry, and bioinformatics. Her email address is ruomingshen@gmail.com.