Visualizing protein structures - tools and trends
Xavier Martinez, Matthieu Chavent, Marc Baaden

To cite this version:
Xavier Martinez, Matthieu Chavent, Marc Baaden. Visualizing protein structures - tools and trends. Biochemical Society Transactions, 2020, 48 (2), pp.499-506. 10.1042/BST20190621. hal-02946006

HAL Id: hal-02946006
https://hal.science/hal-02946006
Submitted on 22 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Visualizing protein structures - tools and trends

X. Martinez¹², M. Chavent³, M. Baaden¹²

1) CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, F-75005, Paris, France
2) Institut de Biologie Physico-Chimique-Fondation Edmond de Rothschild, PSL Research University, Paris, France
3) Institut de Pharmacologie et de Biologie Structurale IPBS, Université de Toulouse, CNRS, UPS, Toulouse, France

Abstract
Molecular visualisation is fundamental in the current scientific literature, textbooks and dissemination materials, forming an essential support for presenting results, reasoning on and formulating hypotheses related to molecular structure. Visual exploration has become easily accessible on a broad variety of platforms thanks to advanced software tools that render a great service to the scientific community. These tools are often developed across disciplines bridging computer science, biology and chemistry. Here we first describe a few Swiss Army knives geared towards protein visualisation for everyday use with an existing large user base, then focus on more specialised tools for peculiar needs that are not yet as broadly known. Our selection is by no means exhaustive, but reflects a diverse snapshot of scenarios that we consider informative for the reader. We end with an account of future trends and perspectives.

Keywords
Molecular Graphics, Protein visualization, Software tools, Virtual reality

Introduction
Many parts of science rely on a visualization-driven cycle of experimentation, reasoning, conjecture and validation, even more so in relation with structural biology and biophysics. Molecular visualization (1) in particular is now broadly used in many contexts, with the purpose of illustration in the scientific literature or the aim to gain insight about primary research data. A broad interest in and need for these methods exists for many decades as described in (2). A key challenge is to pass information from intrinsically three-dimensional objects onto a 2D support such as a sheet of paper or common computer displays. Significant contributions originate in the computer science field of computer graphics. These contributions may only slowly transfer to the field of structural biology (3) as this process requires available end-user oriented software tools for efficient dissemination. Such tools are at the core of this survey.

Tools for visualization of macromolecular structures have emerged from the longstanding need for molecular graphics and are accessible to scientists with broadly varying backgrounds as previously reviewed in (4). In this mini-review we focus on a few essential and generic tools along with some novelties and more singular solutions for specific needs. We will not focus on commercial closed source software, although there are definitely tools available, amongst which we may mention Yasara (5) and Maestro (6), both freely accessible. Commercial tools often integrate or interface with modelling features and may
preferably be used in industry settings or for occasional lay usage because of their user guidance, support and commercial ecosystem. The scope of applications we target with this mini-review ranges from very generic needs to specific usages. For instance a first generic application concerns the illustration of scientific publications and the scrutiny of hypotheses (7), as for example relating mutational data to structural representations (8). A further level of application may involve a more in-depth visual analysis of macromolecular structures and their properties (2), possibly related to the (spatial) distribution of charges, electrostatic properties, pockets and surface complementarities (4,9,10). An even more specialized usage may refer to depicting and analyzing data from theoretical chemistry (9), computational biology (10) and bioinformatics approaches (11). The need to analyze increasingly complex molecular dynamics simulations is one prominent example that has driven the field forward. This usage naturally leads into the field of data analytics and in particular visual analytics (12–14), which is beyond the scope of this review, although we will briefly discuss some aspects thereof.

**Swiss army knives for general needs**

A few molecular visualization packages have been used by the community for many decades and provide robust visualization capacities suitable for a large audience. Among these tools we will briefly discuss the Chimera (latest version called ChimeraX), Jmol (and more recently JSmol), PyMol and VMD softwares, which all serve a large user base. Such legacy packages appear as a safe choice for visualization-based projects, with a stable codebase and API to interact with. Their usability has been refined over time, and although not necessarily easy to use for newcomers, they are well adapted to the tasks related to molecular visualization. Each tool may feature many functionalities and provide tutorials, scripts and extensions that enrich the user’s experience. Table 1 provides a first overview of this core set of software tools, whereas Figure 1 provides a pictorial comparison of a typical work session and each software’s user interface.

Table 1: Feature overview for four broadly used tools for protein visualization

<table>
<thead>
<tr>
<th>Features</th>
<th>Tools</th>
<th>VMD</th>
<th>Pymol</th>
<th>Chimera</th>
<th>Jmol/JsMol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scripting</td>
<td></td>
<td>TCL</td>
<td>Python</td>
<td>Python</td>
<td>Javascript</td>
</tr>
<tr>
<td>OS</td>
<td></td>
<td>Windows/Mac/Linux</td>
<td>Windows/Mac/Linux</td>
<td>Windows/Mac/Linux</td>
<td>Web, Windows/Mac/Linux</td>
</tr>
<tr>
<td>Selection Language</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Raytracing / Illustrations</td>
<td>✔</td>
<td>✔</td>
<td>External POVRay</td>
<td>External POVRay</td>
<td></td>
</tr>
<tr>
<td>Plugins</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Cost</td>
<td></td>
<td>Free&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Free&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Free&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Free</td>
</tr>
</tbody>
</table>

*Some limitations apply to its redistribution. †Both an open-source and a commercial version are distributed. ‡For non-commercial use.*
The Chimera software (15) dates back to 1976, then developed under the name Midas, and provides functionalities for the visual exploration and analysis of molecular structures. Related descriptive data can be examined, in particular cryo-EM datasets with density maps, molecular dynamics trajectories and more. Creating animations of the visualized systems is made easy and the repertoire of visual depictions is rich.

Jmol was started ca. 1999 (16,17). It is a generic platform usable either as a standalone viewer or within a web context. Many educational applications feature Jmol for structure depiction as it is versatile and easy to embed in courseware or to use as graphical frontend for the exploration of structural databases.

PyMol (18) was launched at the end of 1999 and, as its name suggests, is built around a python scripting environment. This tool is very popular with experimentalists as it provides useful features for crystallographic and NMR-derived structures. It generates publication quality figures and provides convenient molecular editing and atom selection functionalities.

Since its inception in 1993, the VMD software focused on the graphical analysis of molecular dynamics data, hence its acronym Visual Molecular Dynamics. It manages even quite large systems smoothly and provides many extensions, in particular for advanced visual analysis, through plugins.

An important feature common to all these packages is the possibility to automate and script tasks for easy re-use and batch deployment.

Despite their long history, all packages still provide regularly novel functionalities and adapt to the progressing hardware technologies, for instance in the graphics card market. Sometimes such extensions may require an extensive re-design of the underlying code as has been documented for instance for ChimeraX (19). Such extensions may address more specific needs and usages adapting to an ever-evolving field as described in the subsequent section.

**Specific needs may be met by these (or additional) tools**

**Using molecular visualization as a “service”** in a broad sense is one of the more recent evolutions. This concept may relate to web integration, interactive visual notebooks such as with Jupyter (20,21) and web-based exploration with novel technologies such as accelerated graphics rendering through WebGL or virtual reality (VR) through WebVR. The NGLviewer (22) and more recently Mol* viewer (molstar.org)(23) address many of these particular aspects. Mol* for instance is now implemented as the default viewer on the PDBe web pages. JSmol and EzMol (24) provide easy web integration as well. WebVR-capable molecular viewers are less widespread, possibly due to the currently limited browser-support for this technology. VRMol (26) is a complete web-based VR tool available at vrmol.net which has recently been released. UnityMol experiments with WebVR have been reported as well (25).

Another wide-spread and very common need is to generate illustrative and compelling images for publications, which has been reviewed by Goodsell and Jenkinson (27). David Goodsell has provided for many years the captivating images of the molecule of the month (28), which can now be reproduced by everyone with the Illustrate software (29). Previously, the introduction of ambient occlusion lighting with the Qutemol software (30) had already led to a significant improvement in the depth perception from static illustrative images. The possibility to continuously vary the abstraction level added further possibilities (31). For particularly eye-catching results, tools from the cinema and animation industry can be used.
as well (32–36). Such publication quality images can be produced with the majority of packages described here, either directly within the package or through a subsequent ray-tracing step. Using interactive raytracing (37,38), very high quality images can be produced almost on the fly and explored interactively. Real-time, often GPU-powered, path-tracing approaches become more widely accessible, for instance through the Blender game engine (30). Alternative approaches attempt to produce visually similar results (39).

The visual exploration of specific data sets represents an important aspect of protein visualization. Historically this aspect is illustrated by the need to visualize experimental electron density maps in crystallography, spawning many tools, among which O (40,41), Coot (42), CCP4mg (43) and ugl ylim (44) just to mention a few. More generally the advanced exploration of specific experimental data may represent a specific feature of some packages, such as Chimera, with a focus on CryoEM data. Similarly, through its tight integration with the protein data bank, Mol* provides many useful features for crystallographic data. A particular challenge may be the joint analysis of data spawning different data spaces. This task is nicely illustrated by the visual analytics approach implemented in the Aquaria software (45,46) enabling a user to link the evolutionary dimension of sequence-space with structure space. The MolArt structure annotation and visualization tool pursues similar avenues (47). Molecular dynamics simulations - and more generally computational simulation approaches including normal modes, Poisson-Boltzmann electrostatics, finite element and elastic network models - represent another source of specific data to be visualized. Historically, the VMD package was the first to focus on molecular simulations, but many others now provide such possibilities, most recently for instance MDsrv (38) or HTMoL (39) on the web (40) and Unitymol in VR. Such virtual reality exploration (48) of protein structures addresses the need to render the complex spatial features of macromolecules very efficiently in immersive 3D thereby helping shape perception and benefiting from advanced interaction devices. Several packages have reported VR capacities such as Molecular Rift (49), ChimeraX (50), UnityMol (25,51), Narupa (52), Nanome (53) and CootVR.

Future trends?

Virtual reality was mentioned in the previous section, featuring currently available software tool implementations for protein visualization. In the future, much is to be expected from augmented reality (AR) headsets or glasses, which promise a great potential for multi-user interactions without isolating the scientists in a virtual world. These collaborative and social aspects are important for efficient scientific exchange as well as to lower the barrier to use such novel technology. AR is currently limited by three main barriers related to the available hardware. First of all, the characteristics of the devices need to improve for instance in terms of field of view, second, better graphics performance and processing power are required when they come embedded in the device such as is for instance the case with the 1st generation of the HoloLens (54), and finally a good human-computer interaction interface with the device needs to be available. Current interaction paradigms through finger gestures and/or voice recognition are a useful step to get a glimpse of the possibilities, but are in the long run not satisfactory for a continued regular usage of these devices as their current error rates build up user frustration rather quickly.

Another trend lies in social features relating to annotating protein visualizations, sharing them, collaborating with other scientists and disseminating findings community-wide with
technologies akin to current social network tools. The jolecule tool (jolecule.com) introduced easily shareable annotated views using a web-based approach. Mo* also supports shareable remote states. To some extent the collaborative social features may be targeted by current VR approaches, such as (50,53,55), whereas foundations were laid a long time ago (56–59).

A particular challenge for future molecular visualization will be to **embrace the next scale** of huge biomacromolecular assemblies and models. With experimental developments such as the resolution revolution of CryoEM (60) and the rise of integrative methods (61), the foreseeable future is in the representation of not only a single small protein but the representation of full organelles. The molecular modeling field significantly contributes to this expansion of the scale of biological objects that can be investigated down to (quasi)-atomistic resolutions (62–64). Exploration of such complex models would definitely benefit from immersive technologies to better comprehend the global molecular system, which is not possible to apprehend on a standard 2D display. This task may also necessitate new ways to interact with such complex systems. This interaction will be possible by using technologies such as virtual reality CAVEs, sophisticated stereoscopic display walls or possibly future generations of VR gear. At the same time these objects and simulations will even further increase the big amount of data to analyze, and immersive visual analytics approaches may contribute significantly to our ability to analyze such data sets and identify non-obvious complex patterns therein. Better tools to generate the analysis data, for instance through advanced analysis of molecular dynamics trajectories, may lead the way into this data-driven future. Some tools already exist (65,66) and the constitution of intuitively accessible data-bases including access to the raw data and visualizations thereof may be a way to delve into the wealth of information generated by such approaches. Several of these databases already exist, such as (67–69), to mention only a few, however interactive visual analysis (and visual analytics in particular) is currently rather limited.

Whereas the overwhelming part of protein visualization so far may have focused on displaying existing models, the future may focus more intensely on methods to visually **create complex molecular landscapes** from scratch. Novel technology may be instrumental to this task (70–73).

A final general and more philosophical comment about the progress of molecular graphics will conclude our collection of thoughts on future trends and perspectives. There still exists a huge potential for substantial improvements of visualization methods by better **integration** of different scientific communities, in particular computer scientists involved in the field of computer graphics and biologists with visualization needs. Much untapped potential can be found in exciting new methodologies developed in computer science that do not or only slowly transfer into tools accessible to structural biologists as described in more detail in (3).

Among the upcoming new challenges in visualization that would benefit from such integration, we may mention for instance the depiction of uncertainty, see (74) for a recent exploration of this topic, or how to capture the dynamics governing macromolecules in static pictures (briefly discussed in (3) as well).

**Perspectives**

- software tools enable protein visualization, an essential part of scientific reasoning
- knowing such tools and their range of applicability is key to efficiently coping with molecular structures and gaining relevant insights
ongoing revolutions in technology such as augmented and virtual reality will drive even further the possibilities and level of visualizations that are possible

• collaborative aspects may be more natively integrated in visualization tools as heralded by trends in multi-user VR applications

• the complexity of visualizations is increasing and new concepts may require new visual metaphors; increasing the transfer of visualization methods from computer science to structural biology bears great potential to address these

Conclusion
Protein visualization is nowadays a vibrant and striving field, with an extraordinary longevity since the first drawings by hand that date back before the 1960s. It has reached a broad audience in diverse subfields of modern biology and is ubiquitous on the web, in publications and on a broad variety of display devices used by researchers, from mobile phones to 3D theaters. This success has been driven by the availability of high-quality software tools, often developed and refined collectively by the scientific community. Beyond the classical axiom: “bigger (structure), larger (dataset), faster (visualization)”, future new developments related to molecular visualization of protein structures will serve our increasing need to integrate heterogeneous data and to efficiently use new hardware devices (from VR/AR headsets to smartphones).

Acknowledgments. This work was supported by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX- 0011 and grant "CACSICE", ANR-11-EQPX-0008) and received funding from the ANR-16-CE05-0024 ENZYMOR project. X.M. and M.B. thank Sesame Ile-de-France for co-funding the display wall used to assess several software packages described in this review. M.C. acknowledges support from CNRS-MITI grants PEPS MPI 2018 and "Modélisation du vivant" 2019. X.M. and M.B. thank UCB Biopharma for support.


67. van der Kamp MW, Schaeffer RD, Jonsson AL, Scouras AD, Simms AM, Tooafanny RD,


Figure 1: A visual comparison of user interfaces and typical rendering features for VMD (top left), Chimera (top right), Jmol (bottom left) and Pymol (bottom right) software tools. A largely alpha-helical protein channel is used as example, highlighting features related to its secondary structure.