Trends in Biochemical Sciences

CellPress
REVIEWS

Opinion

Bringing Molecular Dynamics Simulation Data into View

Peter W. Hildebrand,[1,2,3,*] Alexander S. Rose,[4,@] and Johanna K.S. Tiemann[1,@]

Molecular dynamics (MD) simulations monitor time-resolved motions of macromolecules. While visualization of MD trajectories allows an instant and intuitive understanding of dynamics and function, so far mainly static representations are provided in the published literature. Recent advances in browser technology may allow for the sharing of trajectories through interactive visualization on the web. We believe that providing intuitive and interactive visualization, along with related protocols and analysis data, promotes understanding, reliability, and reusability of MD simulations. Existing barriers for sharing MD simulations are discussed and emerging solutions are highlighted. We predict that interactive visualization of MD trajectories will quickly be adopted by researchers, research consortiums, journals, and funding agencies to gather and distribute results from MD simulations via the web.

Highlights

MD simulations have become routinely used over the past years to investigate dynamic motions of macromolecules at the atomic level.

Interactive visualization of MD trajectories may provide an instant, transparent, and intuitive understanding of complex dynamics.

Sharing of MD trajectories may generate transparency and trust, allowing collaboration, knowledge exchange, and data reuse.

Recent technological developments now allow visual sharing of MD trajectories over the web using tools such as the MDsrv and HTMoL.

GPCRmd presents the first centralized special-purpose MD deposition platform featuring powerful trajectory visualization.

Dynamics Leads to Understanding Macromolecule Function

Molecular dynamics (**MD**, see Glossary) simulations are a well-established technique to investigate time-resolved motions of biological macromolecules at atomic resolution [1,2]. Traditionally, macromolecules such as enzymes, channels, transporters, or receptors have been perceived as being rigid entities mainly because structures obtained by X-ray crystallography are fixed in crystal lattices and are therefore resolved as single, static snapshots. However, when compiled together, these snapshots often reveal that macromolecules exist in different substates and states, hinting that there is dynamic fluctuations between substates and states [3]. Prominent examples are the G protein-coupled receptors (GPCRs), which exist in multiple inactive and active states with different signaling properties [4,5]. Complete understanding of the structural background of GPCR signaling and pharmacological applications requires an in-depth knowledge of receptor and G protein dynamics [6–8].

To overcome these limitations, MD has been developed over several decades, becoming a cutting-edge technology in the life sciences. Formally, the award of the Nobel Prize in Chemistry in 2013 to Martin Karplus, Michael Levitt, and Ariel Warshel for the development of multiscale models for complex chemical systems (https://www.nobelprize.org/prizes/chemistry/2013/press-release/) has affirmed the relevance and importance of MD [9]. Methodological advances, continued software optimization, and hardware developments have broadened the applicability of MD simulations with respect to feasible system size, runtime, and overall quality [10]. These developments are also reflected by the rising number of standardized and intuitive usable tools for automated setup and analysis, facilitating usage and enhancing **reproducibility/replicability** of MD simulations [11,12].

To bring MD into view for a broader audience, we believe the results of MD simulations now have to be presented in a comprehensive and feasible way. So far, even accessing, viewing, and sharing of MD **trajectories** has been hindered by large file sizes for macromolecular structures and the need for specialized software, limiting the audience to which this technology has been available. However, recent developments in web-based technology have allowed for efficient visualization of even large macromolecules such as macromolecular machines and virus capsids in web browsers [13,14]. We believe these advances are opening up new possibilities for sharing the visualization of MD trajectories on the web and will foster interactive collaborations, accessibility, and transparency (Box 1, Figure 1, Key Figure).

Visualization Is Timely and Necessary

Visualization facilitates and guides data analysis, and the complementary strengths of human and machine analysis are potentiated when led by interactive visualization [15]. Advanced visual interfaces that fuse analysis and visualization can combine human flexibility, creativity, and background

[1]Institute for Medical Physics and Biophysics, Leipzig University, 04107 Leipzig, Germany

[2]Institute for Medical Physics and Biophysics, Charité - Universitätsmedizin, 10117 Berlin, Germany

[3]Berlin Insitute of Health (BIH), 10178 Berlin, Germany

[4]RCSB Protein Data Bank, San Diego Supercomputer Center, University of California San Diego, CA, USA

@Twitter: @arose (A.S. Rose), @j0kaso (J.K.S. Tiemann).

*Correspondence: peter.hildebrand@medizin.uni-leipzig.de

Box 1. FAIR Principles

To deal with the growing amounts of large scientific datasets, discovery and innovation increasingly rely on automation and computational support. In 2016, The FAIR Guiding Principles for Scientific Data Management and Stewardship were established to endorse the reliability of data [52]. A special emphasis was, and is, put on enhancing automated searchability and reusability of data. Reusability applies not only to deposition of (raw) data, but also of algorithms, tools, and workflows for data generation and analysis. According to Wilkinson *et al.* [52], using FAIR principles is beneficial for:

- researchers willing to share, credit, and reuse data;
- software/tool developers and data scientists providing processing, integrating, and analysis workflows to enhance discoveries;
- professional data publishers selling their capabilities; and
- funding agencies in minimizing their rising concerns in long-term data handling.

To ensure that data are findable, a unique and persistent idem has to be assigned, connected to descriptive keywords and registered in a searchable resource. Accessibility is obtained when (meta)data can be retrieved by their identifiers using an open, free, and universally implementable protocol. This protocol allows for an authentication and authorization procedure and keeps metadata accessible even when the data are no longer available. To be interoperable, (meta)data have to use a specific, general, and consistent language for knowledge representation including linked references. By ensuring an accurate and rich description with relevant attributes and community standards including usage license and detailed provenance, (meta)data become reusable.

knowledge with storage and processing capacities of computers to gain insights into complex problems [16]. Taking into consideration that scientists trust results presented by **visual analytics** more than other nonvisual media [17], we feel it is time to enable a comprehensive and detailed view on MD trajectories.

Visual analysis has always guided pattern recognition along with other types of MD analysis [18]. For analysis and visualization, tools installed on local machines such as Visual Molecular Dynamics (VMD) [19] or Chimera [20] are used by experts to manipulate the perspective, zoom in on a certain structural detail, or change the display mode of a trajectory. When MD data (Box 2) are prepared for publication, they are commonly translated into tables and static figures showing running averages of trajectories with related statistics to describe the dynamic properties of a specific biophysical system. Figures visualize a predefined setting of those systems focusing the view of the reader to selected aspects like the binding pocket of a receptor while hiding other features or regions due to their static property. Videos provide a more dynamic but still limited glance on the simulations as they are also predefined without allowing manipulation of the perspective.

The full potential of MD simulations, however, is only exploited through interactive visualization of trajectory files; interactive visualization may strengthen and deepen the understanding of a characteristic finding described in a publication [18]. For example, by zooming in on a certain structural and dynamic feature of a trajectory that was not within the scope of the original analysis, novel ideas and hypothesis can be generated and new lines of analysis may be triggered. This approach will thus add new aspects to a published analysis, detect or explain unresolved issues, and complement or update previous findings. In our opinion, interactive visualization of trajectories will, therefore, ultimately strengthen the perception of MD simulations as a reliable technique to monitor the dynamic motion of macromolecules.

Visualizing MD Trajectories on the Web

The web has been developed into the primary resource to gather and distribute information in nearly all areas of life [21,22]. When information is stored and presented in a sustainable way, the web promotes discussion, education, and reputation [23]. To maximize its potential in science, data

Glossary

Application Programming Interface (API): set of protocols/routines/software packages to facilitate development of programs.
Docker images: program that runs software packages through containers within an operating system similar to virtual machines. Docker images can contain complete program environments/servers and can be shared.
DOI/idem: digital object identifier, unique and permanent URL/sequence/phrase linked to data. Often used to find and access the assigned data.
Findable, Accessible, Interoperable, Reusable (FAIR): see Box 1.
Graphics processing unit (GPU): single-chip processor often embedded on a video card, motherboard, or mobile phone to efficiently perform graphical rendering.
Graphical user interface (GUI): allowing the user to interact through graphical icons, menus, or other types of objects in contrast to texts with electronic devices.
Hosted services: provided over the Internet where a computer is configured to handle applications, IT infrastructure components, and functions, covering offerings, including web hosting, infrastructure services, off-site backup, and virtual desktops customized to the needs of the user.
Interoperability: allows data or tools to work with resources other than those they have been originally generated or implemented, respectively, with minimal effort.
Molecular dynamics (MD): (computer) simulations to study the movement of atoms and molecules by solving Newton's equations of motion.
Replicability/reproducibility: refer to the generation of the same/original study results by (i) using exactly the same protocol such as the same input, original source code, techniques, software, settings; or (ii) following/reimplementing the (often less-detailed/automated) protocol or algorithmic/workflow description as given in the published study (e.g., often different tools, software, clusters are used). The association/paring of the terms

management should follow the **Findable, Accessible, Interoperable, and Reusable** (**FAIR**) principles to maximize its potential (Boxes 1 and 2). For example, the **worldwide Protein Data Bank** (**wwPDB**) consortium [24] is a prominent data resource that has obtained a central role in structural biology and life science by following the FAIR principles. As such, all wwPDB members offer web-based molecular graphics to present the curated and deposited 3D structural coordinates in an intuitive and comprehensive way. The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) [25,26], for instance, uses the NGL viewer for a powerful web visualization [13,14] (Box 3). By providing analysis and visualization in an instant and easy-to-use fashion, the wwPDB has the ability to reach out to a wider audience. Further, with predefined settings, the NGL viewer can highlight prominent structural features such as a ligand-binding pocket of a receptor. Optionally, users can visualize experimental structural data such as density maps, orientations, and contacts within the assembly, or they can view B factors, thus facilitating critical reviewing of the structures even before downloading them. Embedded visualization, online analytics, and application of cross-referenced tools have consequently changed the habits of wwPDB users from just downloading structures towards instant online visualization and analysis [27].

Advances in browser technologies have only recently opened up possibilities for web molecular graphics and rendering techniques to allow interactive visualization of macromolecules as discussed above [28]. As a result of a series of these technical developments (Box 3), even web-based visualization of MD trajectories has become feasible. The MegaMol framework for particle-based visualization was the first to show the capability of web visualization of MD simulations; its client-server approach for viewing MD trajectories mainly focuses on visualization research and prototyping [29,30]. The JSmol viewer (https://sourceforge.net/projects/jsmol/) can visualize some trajectory files (e.g., .xyz files) and can interpret structure files with multiple models (e.g., in .pdb files) as dynamic trajectories. Similar to MegaMol, JSmol is, however, only capable of reading very few trajectory file formats and demands expert knowledge for usage.

More recently, the MDsrv was the first user-friendly tool to stream and visualize MD trajectories interactively within web browsers without requiring expert knowledge and specialized software [31]. To interact with many different (currently 22) trajectory formats, the MDsrv utilizes MDTraj [32] and the MDAnalysis [33] software packages. By integrating the NGL viewer for web-visualization like RSCB PDB [13], the MDsrv can display even huge trajectory files in various representations. The MDsrv enables efficient handling of requests for any trajectory frame by not reading the whole file into memory but by transferring only the information requested for visualization [29]. MD trajectory frames rendered into a suitable representation editable by the user, for example, a cartoon representation for the secondary structure and a licorice representation to highlight specific residues or a ligand, can then be shared and accessed via any modern web browser.

About this time, HTMoL, another web-based MD visualization tool, was presented which also facilitates streaming and visualization of MD trajectories on the web [34]. Unique to HTMoL is its implementation of a fast visualization by direct calling of the **graphics processing unit** (**GPU**) that enables rendering and parsing of trajectories from the three most commonly used trajectory formats – Gromacs [35], Charmm [36], and Amber [37]. This is performed by utilizing the Node.js runtime engine to transfer MD trajectory binary data through a WebSocket connection (Box 3). An Apache HTTP server is then applied to share and access trajectories via any modern web browser.

In comparison to MDsrv, HTMoL has a higher frame rate for small to medium-sized systems, most likely due to its direct GPU usage. This means a smoother and faster loading of the trajectory frames of those sized systems prior to their visualization. HTMoL is currently limited to only three file types and to a reduced set of structural representations [34], while MDsrv can currently utilize 22 trajectory file types and more than 20 structural representations [13,31–33]. HTMoL and MDsrv are alike in that they both require the users to set up their own servers, either within a private network or the public cloud. We believe that for wider adoption of these tools, **hosted services** that can access data from scientific data sharing resources (Box 2) and do not require individual researchers to set up servers would be highly beneficial. Moreover, web-based MD visualization tools are still far behind the

reproducibility/replicability with their definition (i) or (ii) depends on the reference, author, or field as a consistent and universally accepted terminology is still missing and its controversy is key to ongoing discussions [72,73].
Trajectory: time evolution of an object moving under the action of given forces.
Visual analytics: scientific field focusing on analytical reasoning supported by interactive visual interfaces.
Web Workers API: allows browsers to run heavy calculations without blocking the user interface.
Web Graphics Library (WebGL): JavaScript API for interactive graphical rendering within a web browser without the use of plug-ins.
Worldwide Protein Data Bank (wwPDB): archive of macromolecular structures determined by experimental techniques such as NMR, cryo-electron microscopy or X-ray crystallography.

Key Figure

Visualization and Sharing of Molecular Dynamics (MD) Simulations on the Web

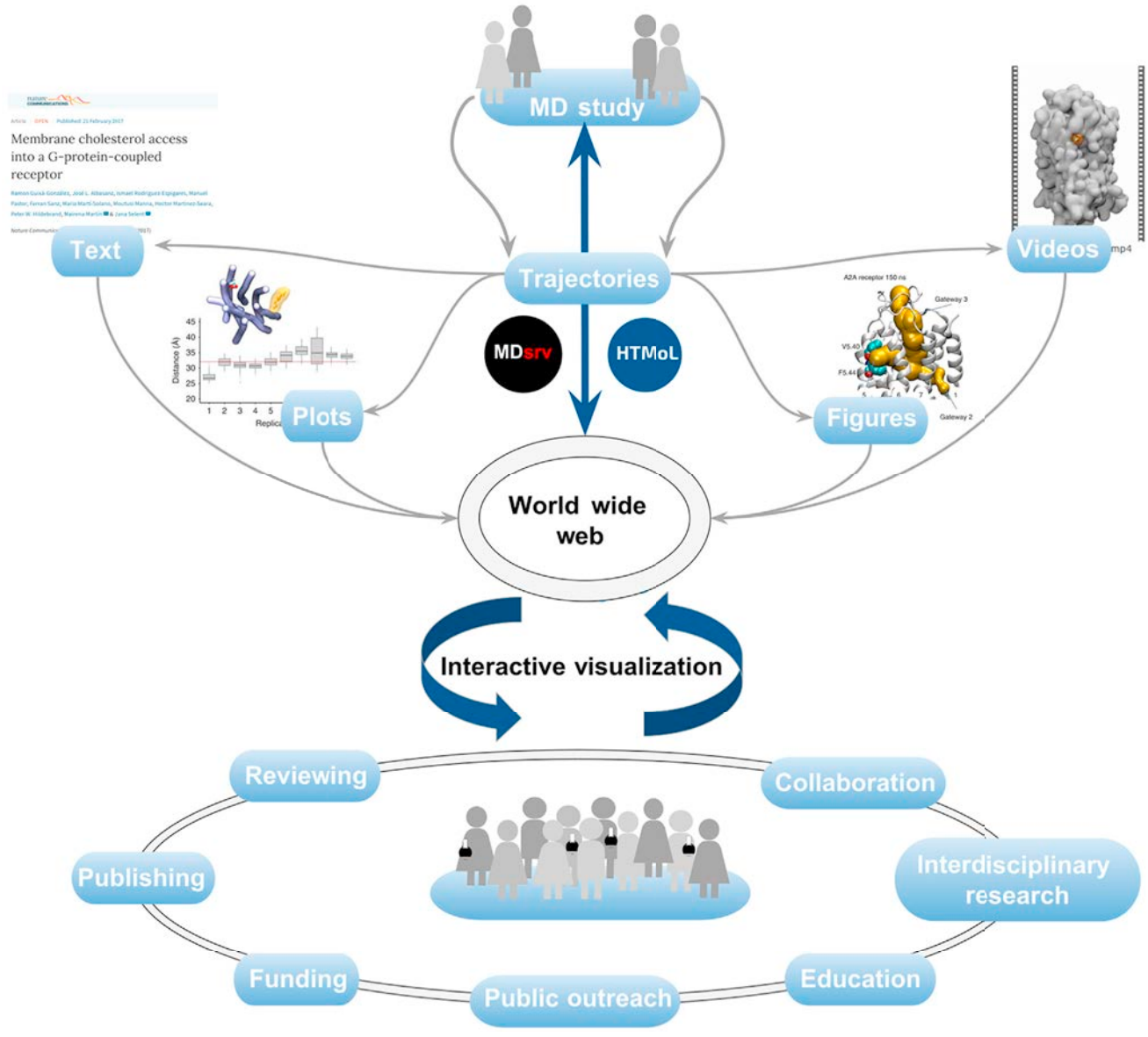


Figure 1. MD trajectories have so far been translated into text, plots, figures, and videos for presentation and publication. Interactive visualization of MD trajectories on the World Wide Web by MDsrv and HTMol, for example, opens up new possibilities for sharing. Sharing of MD simulations with colleagues through interactive visualization promotes discussions and triggers new lines of investigations ultimately strengthening interdisciplinary research projects. Moreover, interactive visualization may support the reviewing process of journals and funding agencies, ultimately contributing to a better understanding of published MD simulation data and analyses. Interactive visualization may also broaden the outreach of MD simulations in society when used in teaching and education. Adapted, with permission, from [76].

Box 2. MD Data and Scientific Data Sharing Resources

Files from MD simulations can be divided into input files, used for production, and output files. Depending on the software used such as GROMACS [35], NAMD [53], AMBER [37], DESMOND [54], or ESPResSo [55], many different input and output file formats are required and generated which are often incompatible. Input files are typically composed of starting structures, core force-field parameters, additional ligands, or other nonstandard atomic parameters and finally operated by detailed, ideally automated, and shareable protocols. For reproducibility/replicability, additional information about used software versions and computer/cluster systems has to be provided in addition to metadata, like configuration files and workflow scripts. However, so far, executable and easily reproducible protocols and metadata to generate and analyze MD studies are only rarely made available. The output of MD simulations consists of raw unprocessed trajectories including coordinates and velocities, energy, and log files. MD simulations belong to data-intensive sciences (also called e-sciences) like proteomics, genomics, oceanography, astrophysics, engineering, web sciences, and more. Sharing of raw or processed MD trajectory files is, accordingly, hampered by large file sizes. Nevertheless, scientific data sharing resources offer the possibility to deposit data even as huge as trajectories [56].

In response to open science movements [57], numerous scientific general-purpose data repositories at scales ranging from institutional to open globally scoped repositories emerged. Scientific data sharing resources such as Dataverse [58], FigShare (http://figshare.com), Dryad [59], Mendeley Data (https://data.mendeley.com), Zenodo (http://zenodo.org), DataHube (http://datahub.io), OSF (https://osf.io), or NOMAD (https://nomad-coe.eu) accept a wide range of data types in a large variety of formats. The deposited data can be accessed globally through a unique, preserved DOI. Compared with well-curated, special-purpose archives like the wwPDB, most scientific data sharing resources do not aim to integrate, harmonize, validate, or standardize the deposited data. Since appropriate references and metadata are often not available, it is difficult to filter suitable data from the huge variety of deposited files and databases. To overcome this challenge, the establishment of information platforms summarizing and integrating available data should be instrumental in providing confidence in open source deposition concepts until global data standards are established. Websites such as https://fairsharing.org/ or https://okfn.org/ are helpful resources recommending repositories for scientific data, standards, and policies [60].

capabilities of stand-alone viewer programs such as VMD [19] or others highlighted here [38] that integrate manifold built-in analysis tools like root mean square deviation (RMSD), radius of gyration, extracting min/max coordinates/distances, heat maps, contact maps, energy plots and many more.

Despite their differences, MDsrv and HTMoL allow instant and interactive dynamic web-based visualization of MD trajectories through intuitive **graphical user interfaces** (**GUIs**). After set up of the server by the MD simulation distributer, the data can be streamed directly through the web without the requirement to download any trajectory data or to install software by the user/observer. The trajectories become immediately available in the browser as soon as a link served by the creator of MD simulations (also known as MD simulation distributer) is clicked by the user/observer. In addition, the MDsrv and HTMoL visualization can be embedded into any website to be combined with other user-specified analysis tools.

Towards Common Platforms for Sharing and Visualization

Scientific data can be shared through centralized or decentralized solutions. Centralized sharing options provide a highly efficient way to gather and distribute research results via the web. The wwPDB serves as a blueprint for centralized solutions to share structural information via the web [24]. Precisely curated data are accurately and richly annotated by data-depositing authors with easy on-the-point keywords to ensure findability by the broader community. Additionally, all data deposited to the wwPDB are freely and openly accessible and reusable. Furthermore, a major reason for the success of the wwPDB is that most journals require structural biologists to upload their structure information to the database prior to publication of their manuscripts. This synergistic effect has led to exceptionally high citation rates of publications, especially on novel structures deposited in the wwPDB [39].

A similar positive effect could also be expected for MD simulations if trajectories and related data would become accessible and interoperable at the moment of publication (Box 1). In contrast to

Box 3. Technical Developments Enabling Web Molecular Graphics

Molecular graphics have been a part of the web since its early days (Figure I). Initially, browsers could not display 3D content by themselves but relied on plugins [e.g., Chime (MDL Information Systems, Inc. https://www.umass.edu/microbio/chime/abtchime.htm)] or extensions (Java as used by Jmol [61] and OpenAstexViewer [62]) that had to be installed in addition to the browser. Lack of access to hardware accelerated graphics and inflexible data streaming have been major limitations for web molecular graphics. In recent years, advances in browser technology have opened up new possibilities for web molecular graphics and rendering techniques [28].

Modern browsers now include native support for GPU hardware acceleration through the **Web Graphics Library (WebGL) Application Programming Interface (API)** (https://www.khronos.org/registry/webgl/specs/latest/). This enables web molecular graphics tools to provide fast 3D graphics on par with desktop programs. In addition to GPU rendering, WebGL also includes support for handling binary data natively in the browser which makes working with numerical and binary data much more efficient. Starting with GLmol (https://github.com/biochem-fan/GLmol), the new browser capabilities have led to an explosion of web molecular graphics tools, including PV (https://github.com/biasmv/pv), 3Dmol.js [63], NGL [13], Molmil [64], LiteMol [65], and more (Figure I).

Nowadays, nearly all computers have multicore processors and the **Web Workers API** runs computations in worker threads in parallel, leveraging the many cores of processors. As an additional benefit, the worker threads perform tasks in the background without blocking the user interface. This is helpful for computationally intensive tasks like calculating molecular surfaces and is employed, for example, by NGL and 3Dmol.js.

To provide real-time access to large amounts of data on demand, an efficient, low overhead data streaming approach is often used. The WebSocket protocol (https://tools.ietf.org/html/rfc6455) enables this interaction between a web browser and a web server and is much more efficient than the traditional HTTP protocol (as used by MDsrv [31]). WebSockets are used by HTMol [34] and MegaMol [29,30] to stream MD trajectories to a web browser.

A new generation of web applications is leveraging these tools enabled by the advances in browser technologies for modern computer-aided drug design [66] and general molecular sciences [22]. Additionally, a new collaborative project, Mol* [67], was started to develop a common tool for web molecular graphics to better meet visualization challenges.

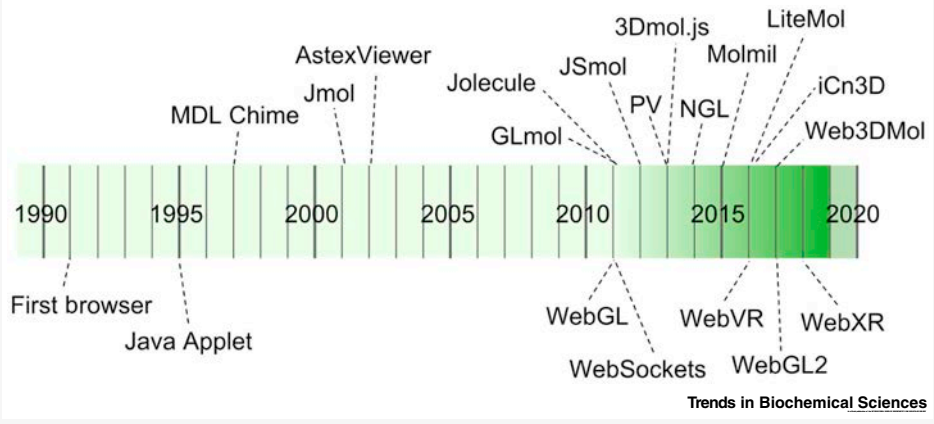


Figure I. Timeline (Center) of Relevant Web Browser Technologies (Bottom) and Popular Web Molecular Graphics Tools (Top).

The listed web molecular graphics tools can be divided into a pre and post WebGL era (light green to increasingly dark green). Pre WebGL: MDL Chime (MDL Information Systems, Inc. https://www.umass.edu/microbio/chime/abtchime.htm), Jmol/JSmol [61], AstexViewer [62]. Post WebGL: GLmol (https://github.com/biochem-fan/GLmol), Jolecule (http://jolecule.com/), PV (https://github.com/biasmv/pv), 3Dmol.js [63], NGL Viewer [13], LiteMol [65], Molmil [64], and iCn3D [74] and Web3DMol [75].

data uploaded to the wwPDB, MD simulations can accumulate huge trajectory files ranging from gigabytes to terabytes, which may raise concerns about a sharable solution for storage and handling (see Outstanding Questions). However, sharing of huge amounts of data is feasible as demonstrated by other data-intensive science fields. The Galaxy Project [40] is a prominent example showing how community fostered standards and guidelines promote efficient and accurate sharing and reuse of huge genomic data [41].

Nevertheless, the establishment of centralized sharing platforms requires many efforts, guidelines, community engagement, and a central dedicated and sustainably funded organization [42]. Community engagement is crucial as any centralized organization has to be guided by the community. The wwPDB has, for example, various task forces for validation, format development, and regarding new experimental techniques [43,44]. Centralized, special-purpose MD platforms may accordingly be considered a first step on the way towards the establishment of global MD platforms. Several such platforms have already been created (Table 1), but only BIGNASim [45] and GPCRmd (http://www.gpcrmd.org/) feature interactive visualization of trajectories. BIGNASim is a platform containing a database system and analysis portal for MD simulations of nucleotides [45]. It includes a comprehensive analysis tool package and a spartan interactive visualization of trajectories by JSmol, limited to play and pause. Curated simulations by users can be up- and downloaded. Moreover, a reduced version of BIGNASim can be installed as a local platform to serve sensitive or unpublished data via **Docker images** by individual researchers/MD simulation distributers.

The GPCRmd is a comprehensive database and web platform for MD simulations of GPCRs and related analysis (http://www.gpcrmd.org/). It uses the MDsrv for interactive visualization of trajectories with predefined settings to highlight GPCR specific features (Figure 2). Linked to the MDsrv are pharmacological, biochemical, or biophysical data like mutations or X-ray crystallography density maps and an interactive, topic-specific Flareplot analysis (https://gpcrviz.github.io/flareplot/). The platform has a descriptive, easily searchable framework and includes an application for simulation setup and deposition, also available with unique **idem/DOI**. Similar to the Galaxy Project [40], GPCRmd is discussed, designed, and influenced by a research community which makes the platform frequently used, widely supported, and sustainable.

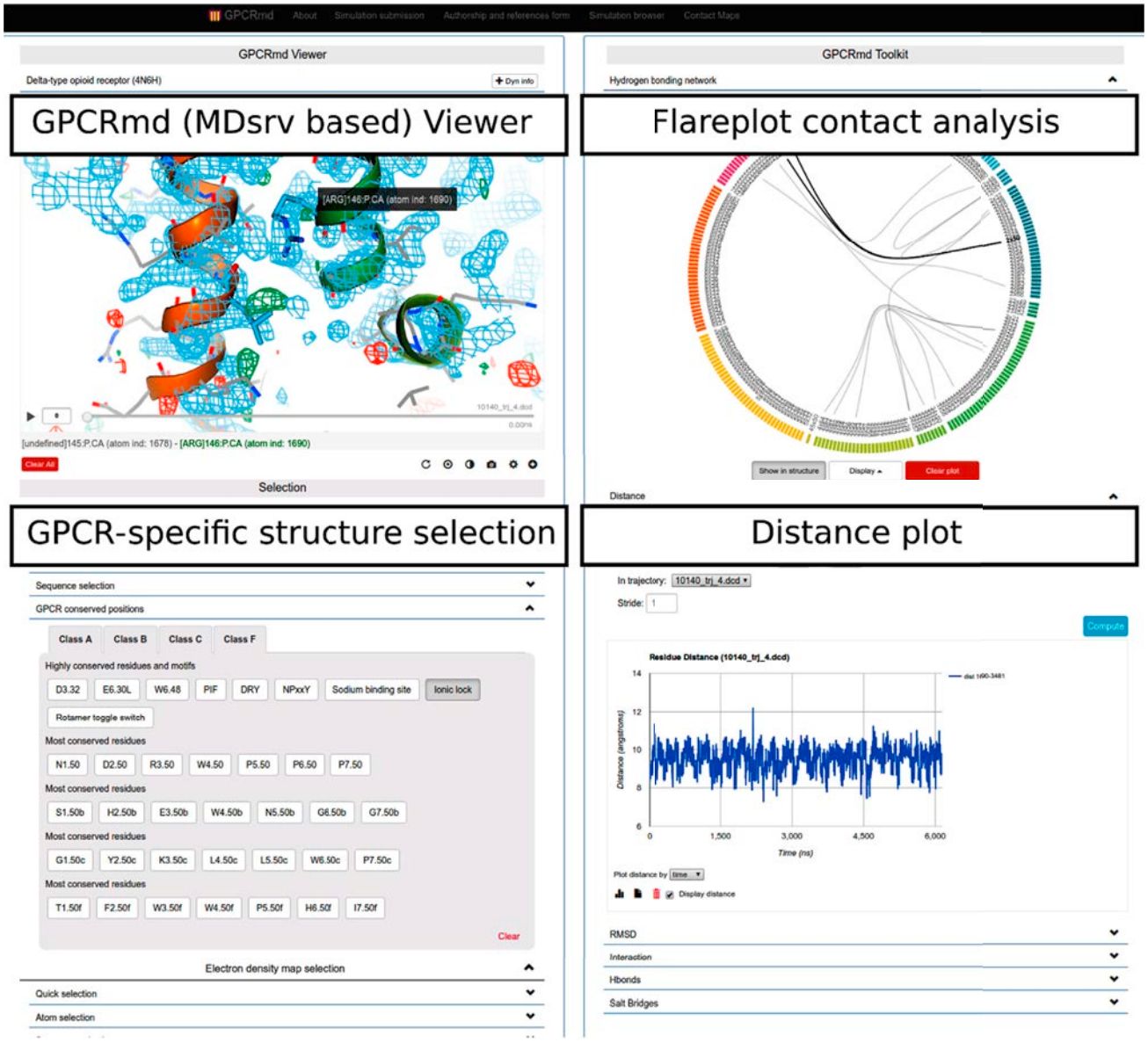
For research areas and fields where no centralized sharing platforms are available, individual researchers or groups can enrich their own websites with embedded simulations (e.g., http://nglviewer.org/mdsrv/examples). In contrast to the default settings of centralized options, sharing through decentralized solutions allows the creator of MD simulations to define study specific pre-settings. Analogously, user-specified analysis tools can be combined with interactive visualization in a modular fashion. As an example, iBIOMES facilitates searchability and findability by providing a metadata schema for indexing and summarizing MD data in a web-searchable representation [46,47]. In summary, decentralized options may be used by scientists preferring instant, independent and creative solutions for sharing and publication of MD trajectories. Moreover, it may be the primary way to share trajectories with collaboration partners fostering critical discussions already on an early stage of the project.

Concluding Remarks and Future Perspectives

Visualization and sharing of MD trajectories can be considered initial steps towards a community following the FAIR principles. We expect several lines of developments guiding this process, though several issues remain to be resolved (see Outstanding Questions). Notably regarding shared visualization in context of the FAIR principles, the data must be findable and accessible and sufficiently interoperable to allow for visualization. Especially, findability suffers from poor description while reusability suffers from missing verification for completeness and coherence of all MD data. In the broader context of making MD simulations reproducible, the **interoperability** and reusability aspects become more important. Currently, there are no widely accepted guidelines for making MD data globally findable or accessible. A comprehensive and constantly developing discussion and description is needed to examine if full FAIR compliance can be obtained for all MD data, whether it is needed and how much it would cost.

Name	Topic/ specificity	Access/ usage	Collection available	Deposition/ upload	Setup	Analysis tools	Data mining	Structural visualization	Trajectory visualization	Searchable	Shareable	Status	Refs	Accessed
BIGNASIM	Nucleic acids	Open access	Yes	Yes	Yes	Yes	Yes	Yes	Limited	Yes	Yes	Active	[45]	2019/01/29
Cyclo-lib	Cyclodextrins	Public access	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Limited	Active	[68]	2019/01/29
Dynameomics	native state and (un) folding dynamics	Not stated	Yes	No	No	Yes	Yes	Yes	No	Limited	Limited	Inoperative	[69]	2019/01/29
GPCRmd	GPCR	Open access	Yes	Yes	Yes	Yes	Possible	Yes	Yes	Yes	Yes	Active	http://www.gpcrmd.org	2019/01/29
Mdbox	all	Open access	No	Yes	No	Yes	Possible	No	No	Yes	Yes	Prototype	http://www.mdbox.org	2019/01/29
MemProtMD	Membrane proteins	Open access	Yes	No	No	Yes	Possible	Yes	No	Yes	Yes	Active	[70]	2019/01/29
MoDEL/ MDWeb	Monomeric soluble proteins	Academic use	Yes	No	Yes	Yes	No	Yes	No	Yes	Limited	Not updated	[71]	2019/01/29
NMRlipids Database	Lipid bilayers	Open access	Yes	Yes	No	No	Possible	No	No	Yes	Yes	Active	http://www.nmrlipids.fi/	2019/01/29

Table 1. List of (Selected) MD Databases



Trends in Biochemical Sciences

Figure 2. Screenshot from GPCRmd (http://www.gpcrmd.org/).
The interactive visualization is shown on the left panel: in the upper part, the embedded MDsrv is showing the molecular dynamics (MD) trajectory together with X-ray crystallography density maps. Below, a G protein-coupled receptor (GPCR)-specific selection mode can be used to visualize GPCR-specific structural features. The analysis (right panel) is interconnected to the interactive visualization by the MDsrv. By clicking on a link within the Flareplot (right upper part), the hydrogen bonding network between those residues is directly visualized in the MDsrv (left upper part). On the right lower panel, the time-dependent development of a distance between two atoms clicked within the MDsrv is plotted.

The availability of online tools for interactive visualization can set a novel standard for sharing of MD trajectories: first, raw and unprocessed trajectories prior to extensive analysis can be shared within workgroups to guide initial research, before sharing the MD outcome with collaborators and a broader audience (Figure 1). This development accelerates as user-friendliness and acceptance of tools for interactive visualization of MD trajectories improve and scientific data repositories become standardized. Consciousness for interactive visualization and sharing options of MD simulations will

grow through appropriate workshops, social media discussions, scientific meetings, or simply by successful integration into centralized MD platforms such as GPCRmd.

We believe a particularly important step will be the integration of interactive visualization of MD trajectories into reviewing processes and ultimately journal policies that have a major impact on the research behavior of individual scientists [48]. Recently established journal formats such as ActivePapers [49] or the Living Journal of Computational Molecular Science (https://www.livecomsjournal.org/) already foster interactive web-based publications. Nevertheless, journals will gradually adapt their policies with authors increasingly sharing their trajectories and reviewers asking for interactive visualization. A similar trend can be predicted for funding agencies exploring appropriate and effective ways to publish MD simulation projects. Specifically, centers for high-performance computing would significantly gain visibility if results of funded research projects are presented in an interactive and intuitive way. First steps in adopting journal policies for MD are presented by https://fairsharing.org or [50].

The benefits of interactive visualization and sharing of MD trajectories may have to overcome common resistance and reluctance in publishing complete data sets, specifically of raw data [41,51]. We believe that publishing complete datasets and trajectories will help the community to reinterpret data, reevaluate outliers, and provide a more realistic picture of the complex and often nonergodic nature of MD simulations, because we realize that although automatization of setups and analysis of MD simulations facilitates application and analysis, it also entails the risk of misinterpretations by newcomers. Adopting the FAIR principles by the MD simulation community may help to identify setup issues and false interpretations already during the reviewing process and promote critical discussions.

In summary, visualization and sharing of MD simulations may increase the reliability and understanding of this technique and foster a more direct understanding of molecular dynamics. This development will facilitate access and ultimately broaden the outreach of MD simulations in structural biology, education, and society.

Outstanding Questions

How can we address concerns to overcome the common resistance and reluctance in sharing and publishing complete data sets of MD simulations? Will shared visualization help in this process?

How could general guidelines and standards for sharing of MD simulations be defined and how should they be controlled?

What requirements should scientific journals demand for MD simulation data? How should they update their guidelines?

What are the (minimum) requirements to ensure a rich annotation, description, and verification of completeness and coherence of all MD simulation files for full or partial compliance with the FAIR principles?

Should all replicates or only representative MD simulation trajectories be deposited and published?

Is deposition of complete MD trajectories required, or will future hardware acceleration make (re-) production of simulations (potentially from snapshots) even on extended timescales more feasible?

To what extend shall analysis data become standardized and visualized?

How should MD-derived transient state models, transition paths, or binding energies requiring elaborate curation be published and shared?

Would it be helpful to link raw data and protocols of integrative/hybrid method models from MD simulations to the wwPDB?

Should MD data in general be curated and archived like experimentally derived structural data by the wwPDB? Who or which organization will be responsible for that effort?

Are general data sharing platforms sufficient for deposition and sharing of MD simulations or are

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Sfb740/B6, HI1502/1-2, by Stiftung Charité, Berlin Institute of Health (BIH) and the Einstein Center Digital Future (P.W.H.). The RCSB PDB was jointly funded by the National Science Foundation (NSF), the National Institutes of Health (NIH), and the US Department of Energy (NSF DBI-1338415; PI: S.K. Burley). We further would like to thank the reviewer and especially the editor Sannie Culbertson for their helpful comments and suggestions.

References

1. Dror, R.O. *et al.* (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* 41, 429–452
2. Perilla, J.R. *et al.* (2015) Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* 31, 64–74
3. Elgeti, M. *et al.* (2013) Precision vs flexibility in GPCR signaling. *J. Am. Chem. Soc.* 135, 12305–12312
4. Manglik, A. and Kobilka, B. (2014) The role of protein dynamics in GPCR function: insights from the β2AR and rhodopsin. *Curr. Opin. Cell Biol.* 27, 136–143
5. Rose, A.S. *et al.* (2014) Position of transmembrane helix 6 determines receptor G protein coupling specificity. *J. Am. Chem. Soc.* 136, 11244–11247
6. Velgy, N. *et al.* (2018) GPCRs: What can we learn from molecular dynamics simulations? In *Methods in Molecular Biology*, pp. 133–158, Humana Press
7. Wingler, L.M. *et al.* (2019) Angiotensin analogs with divergent bias stabilize distinct receptor conformations. *Cell* 176, 468–478.e11
8. Liu, X. *et al.* (2019) Structural insights into the process of GPCR-G protein complex formation. *Cell* 177, 1243–1251.e12
9. Levitt, M. (2014) Birth and future of multiscale modeling for macromolecular systems (Nobel Lecture). *Angew. Chemie Int. Ed.* 53, 10006–10018
10. Vlachakis, D. *et al.* (2014) Current state-of-the-art molecular dynamics methods and applications. *Adv. Protein Chem. Struct. Biol.* 94, 269–313
11. Lee, J. *et al.* (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* 12, 405–413
12. Doerr, S. *et al.* (2017) High-throughput automated preparation and simulation of membrane proteins with HTMD. *J. Chem. Theory Comput.* 13, 4003–4011
13. Rose, A.S. and Hildebrand, P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.* 43, W576–W579

14. Rose, A.S. *et al.* (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* 34, 3755–3758
15. Crouser, R.J. *et al.* (2017) Toward theoretical techniques for measuring the use of human effort in visual analytic systems. *IEEE Trans. Vis. Comput. Graph.* 23, 121–130
16. Keim, D.A. *et al.* (2009) Visual analytics: scope and challenges. In *Visual Data Mining*, pp. 76–90, Springer
17. Dasgupta, A. *et al.* (2017) Familiarity vs trust: a comparative study of domain scientists' trust in visual analytics and conventional analysis methods. *IEEE Trans. Vis. Comput. Graph.* 23, 271–280
18. Hirst, J.D. *et al.* (2014) Molecular simulations and visualization: introduction and overview. *Faraday Discuss.* 169, 9–22
19. Humphrey, W. *et al.* (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38
20. Pettersen, E.F. *et al.* (2004) UCSF Chimera? A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612
21. Schroeder, R. (2018) The internet in everyday life II. In *Social Theory after the Internet*, pp. 101–125, UCL Press
22. Abriata, L. (2017) Web apps come of age for molecular sciences. *Informatics* 4, 28
23. Abriata, L.A. *et al.* (2018) Augmenting research, education, and outreach with client-side web programming. *Trends Biotechnol.* 36, 473–476
24. Berman, H. *et al.* (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* 10, 980
25. Berman, H.M. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
26. Burley, S.K. *et al.* (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 47, D464–D474
27. Berman, H.M. *et al.* (2012) The Protein Data Bank at 40: reflecting on the past to prepare for the future. *Structure* 20, 391–396
28. Mwalongo, F. *et al.* (2016) State-of-the-art report in web-based visualization. *Comput. Graph. Forum* 35, 553–575
29. Mwalongo, F. *et al.* (2016) GPU-based remote visualization of dynamic molecular data on the web. *Graph. Models* 88, 57–65
30. Grottel, S. *et al.* (2015) MegaMol – a prototyping framework for particle-based visualization. *IEEE Trans. Vis. Comput. Graph.* 21, 201–214
31. Tiemann, J.K.S. *et al.* (2017) MDsrv: viewing and sharing molecular dynamics simulations on the web. *Nat. Methods* 14, 1123–1124
32. McGibbon, R.T. *et al.* (2015) MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109, 1528–1532
33. Michaud-Agrawal, N. *et al.* (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* 32, 2319–2327
34. Carrillo-Tripp, M. *et al.* (2018) HTMoL: full-stack solution for remote access, visualization, and analysis of molecular dynamics trajectory data. *J. Comput. Aided Mol. Des.* 32, 869–876
35. Berendsen, H.J.C. *et al.* (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91, 43–56
36. Brooks, B.R. *et al.* (2009) CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614
37. Case, D.A. *et al.* (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688
38. Dubbeldam, D. *et al.* (2019) Highlights of (bio-) chemical tools and visualization software for computational science. *Curr. Opin. Chem. Eng.* 23, 1–13
39. Markosian, C. *et al.* (2018) Analysis of impact metrics for the Protein Data Bank. *Sci. Data* 5, 180212
40. Afgan, E. *et al.* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544
41. Martens, L. and Vizcaíno, J.A. (2017) A golden age for working with public proteomics data. *Trends Biochem. Sci.* 42, 333–341
42. Anderson, W.P. (2017) A global coalition to sustain core data. *Nature* 543, 179
43. Sali, A. *et al.* (2015) Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure* 23, 1156–1167
44. Read, R.J. *et al.* (2011) A new generation of crystallographic validation tools for the Protein Data Bank. *Structure* 19, 1395–1412
45. Hospital, A. *et al.* (2016) BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.* 44, D272–D278
46. Thibault, J.C. *et al.* (2014) iBIOMES Lite: summarizing biomolecular simulation data in limited settings. *J. Chem. Inf. Model.* 54, 1810–1819
47. Thibault, J.C. *et al.* (2014) Data model, dictionaries, and desiderata for biomolecular simulation data indexing and sharing. *J. Cheminform.* 6, 4
48. Stodden, V. *et al.* (2018) An empirical analysis of journal policy effectiveness for computational reproducibility. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2584–2589
49. Hinsen, K. (2015) ActivePapers: a platform for publishing and archiving computer-aided research. *F1000Research* 3, 289
50. Elofsson, A. *et al.* (2019) Ten simple rules on how to create open access and reproducible molecular simulations of biological systems. *PLoS Comput. Biol.* 15, e1006649
51. Longo, D.L. and Drazen, J.M. (2016) Data sharing. *N. Engl. J. Med.* 374, 276–277
52. Wilkinson, M.D. *et al.* (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018
53. Phillips, J.C. *et al.* (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802
54. Bowers, K.J. *et al.* (2006) Molecular dynamics – scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing - SC '06*, p. 84, IEEE
55. Arnold, A. *et al.* (2013) ESPResSo 3.1: molecular dynamics software for coarse-grained models. In *Lecture Notes in Computational Science and Engineering 89 LNCSE*, pp. 1–23, Springer
56. Lotz, S.D. and Dickson, A. (2018) Unbiased molecular dynamics of 11 min timescale drug unbinding reveals transition state stabilizing interactions. *J. Am. Chem. Soc.* 140, 618–628
57. Editorial. (2018) Data sharing and the future of science. *Nat. Commun.* 9, 2817
58. Crosas, M. (2011) The Dataverse Network®: an open-source application for sharing, discovering and preserving data. *D-Lib Mag.* 17, https://doi.org/10.1045/january2011-crosas
59. Klas, W. and Greenberg, J. (2008) *Metadata for Semantic and Social Applications DC-2008 Berlin*, Göttingen University Press
60. Data, S. (2018) *Scientific Data Recommended Repositories.* https://doi.org/10.6084/m9.figshare.1434640.v12
61. Hanson, R.M. (2010) Jmol – a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.* 43, 1250–1260

special-purpose MD platforms, centralized or distributed, a more appropriate solution?

Where and how should scientists deposit their simulation data at the moment?

How can we handle large data storage transfer bandwidth requirements?

Will deposition of MD simulation data at scientific data sharing resources in combination with interactive visualization prove useful; for example, for interdisciplinary research, reusage, teaching, and publishing?

Which features should the next generation of tools for interactive visualization of MD trajectories integrate to improve usefulness?

62. Hartshorn, M.J. (2002) AstexViewer: a visualisation aid for structure-based drug design. *J. Comput. Aided Mol. Des.* 16, 871–881
63. Rego, N. and Koes, D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* 31, 1322–1324
64. Bekker, G.-J. *et al.* (2016) Molmil: a molecular viewer for the PDB and beyond. *J. Cheminform.* 8, 42
65. Sehnal, D. *et al.* (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat. Methods* 14, 1121–1122
66. Yuan, S. *et al.* (2017) Implementing WebGL and HTML5 in macromolecular visualization and modern computer-aided drug design. *Trends Biotechnol.* 35, 559–571
67. Sehnal, D. *et al.* (2018) Mol*: towards a common library and tools for web molecular graphics. In *Proceedings of the Workshop on Molecular Graphics and Visual Analysis of Molecular Data (MolVA '18)*, pp. 29–33, Eurographics Association Aire-la-Ville
68. Mixcoha, E. *et al.* (2016) Cyclo-lib: a database of computational molecular dynamics simulations of cyclodextrins. *Bioinformatics* 32, 3371–3373
69. van der Kamp, M.W. *et al.* (2010) Dynameomics: a comprehensive database of protein dynamics. *Structure* 18, 423–435
70. Newport, T.D. *et al.* (2019) The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res.* 47, D390–D397
71. Meyer, T. *et al.* (2010) MoDEL (molecular dynamics extended library): a database of atomistic molecular dynamics trajectories. *Structure* 18, 1399–1409
72. Plesser, H.E. (2018) Reproducibility vs. replicability: a brief history of a confused terminology. *Front. Neuroinform.* 11, 76
73. Barba, L.A. (2018) *Terminologies for reproducible research*. https://arxiv.org/pdf/1802.03311.pdf
74. Wang, J. *et al.* (2018) iCn3D, a web-based 3D Viewer for the visualization of biomolecular structure and sequence annotation. *bioRxiv*. https://doi.org/10.1101/501692
75. Shi, M. *et al.* (2017) Web3DMol: interactive protein structure visualization based on WebGL. *Nucleic Acids Res.* 45, W523–W527
76. Guixà-González, R. *et al.* (2017) Membrane cholesterol access into a G-protein-coupled receptor. *Nat. Commun.* 8, 14505