ISSN 2186-7437

NII Shonan Meeting Report

No.142

Web Molecular Graphics: Emerging Technologies & Standards

Marc Baaden
Seán O'Donoghue
Masakazu Sekijima

December 03–06, 2018

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

1 Overview of the meeting

1.1 Background

Molecular Graphics (MG) unites computer graphics and software engineering to provide modelling, simulation, and visualization essential to a wide range of scientific and commercial research — from the life sciences to chemistry and material science. MG is a mature discipline with a wealth of methods and tools, originally focused on stand-alone software. Recently, however, rapid new developments are exploring the use of MG within web browsers, opening exciting new possibilities, such as large-scale collaboration. At the same time, advances in experimental methods now make available molecular structures of unprecedented size and complexity, giving new insights into the molecular machinery of life, but also requiring much increased computational performance, and sometimes the use of new display technologies (e.g., distributed and multiscale displays).

In response, the first Shonan seminar on Web-based Molecular Graphics was held in 2016 [60]. This exciting meeting brought together, for the first time, key international players in this newly emerging field. The meeting featured broad-ranging discussions that identified key common issues and created new collaborations; this has had significant impact in our field, helping consolidate ongoing efforts, and resulting in several publications.

This second, more focused meeting was set up to explore four specific topic areas: (1) Standards (de facto and emerging); (2) Big molecular data; (3) Emerging technologies (e.g., VR/AR, WebVR, HMDs); and (4) Community building. In the 2016 meeting, these issues were identified as high-priority next steps for our community. Since then, each of these areas has been impacted by rapid technological advances. Thus, for our next proposed meeting, we have selected invitees with expertise matching to these topics, and thereby spanning computer science, molecular graphics, and bioinformatics. The focus will be on how advances in these topics can be leveraged to create the next generation of molecular graphics methods and tools, which, in turn, will give scientists the power to explore the molecular realm in unprecedented detail and clarity.

1.2 Aims of the meeting

Standards (de facto and emerging): This will be a core topic for discussion, focusing on clarifying standards specifically related to (1) web-based graphics (e.g., WebGL, OpenGL, three.js), (2) molecular graphics (e.g., PDBx, MMTF), (3) related biomolecular information (e.g., 2D images or 3D envelopes from microscopy experiments), (4) a common query grammar for specifying parts of molecular structures (e.g., MolQL), and (5) improving visualization conventions for web molecular graphics (e.g., CPK).

Big molecular data: Increasingly more molecular dynamics trajectories are available, and for increasingly larger molecular systems (especially due to advances in CryoEM). Managing these unprecedented amounts of data will require new approaches in the data structure, compression, transport, and storage, as well as new client-server paradigms and on-demand, multi-scale display methods.

Emerging technologies (VR/AR, WebVR, HMDs): Molecular graphics have always been an area where new graphic technologies were tested. Today, new

display technologies such as Virtual Reality (VR) and Augmented Reality (AR) are being developed within the browser (e.g. WebVR). These technologies are very promising for exploring large spatial structures. Although the use of VR and AR in molecular graphics is not new, their use in a web context provides exciting new possibilities that we plan to explore, along with a consideration of user-experience design perspectives.

Community building: Perhaps most importantly we will focus on defining and initiating new community practises that will facilitate global, pre-competitive collaborations that will help accelerate the development of web molecular graphics. We will do this by defining our needs, primarily focusing on creating new strategies to enable collaborative work on the development of molecular graphics. Increasingly, the power of the web is being leveraged to enable collaborative development; but standards and best-practises are still evolving. Thus, we will invite key players who have been involved in successful, long-term, and online collaborative developments (e.g., PDB, SBML) —- including some outside of molecular graphics (e.g., Khronos Group) —- so our field can benefit from their experience. The result will be a clear set of new, very interesting challenges for computer science, and eventually, many benefits to users of molecular graphics across many scientific fields.

Publication plan: We plan to publish several focused papers on the discussion topics in venues that will best reach specific target communities. A core focus will be greatly extending existing guidance documentation and accompanying illustrations at the online repository created in GitHub during the 2016 meeting. This repository will host ongoing conversations, helping to orchestrate the growth and outreach of the web molecular graphics community.

2 Meeting schedule

Check-in Day: December 2nd (Sun)

- Welcome Banquet

Day1: December 3rd (Mon)

- Lightning talks
- Plenary: Seán O'Donoghue
- Agreement on breakout topics
- Group photo shooting
- Breakout sessions
- Breakout check-in and round up
- Agreement on further breakout topics
- Plenary: J.B. Brown

Day2: September 4th (Tue)

- Plenary: Barbora Kozlikova
- Plenary: Ivan Viola
- Plenary: Alexander S Rose, David Sehnal
- Breakout sessions
- Breakout summary
- Agreement on further breakout topics
- Plenary: Monica Zoppè

Day3: September 5th (Wed)

- Plenary: Matthieu Chavent
- Plenary: Michael Krone
- Breakout session
- Excursion and Main Banquet

Day4: September 6th (Thu)

- Breakout session
- Breakout summary
- Wrap up

3 Plenary talks

Visualizing biomolecular structures for research & outreach

Seán O'Donoghue, CSIRO & Garvan Institute, Australia

As a scientific field, structural biology has been extraordinarily impactful. The elucidation of the double helix structure of DNA [70] provided immediate insight into how biological information is stored and replicated; this insight, in turn, launched modern biology. In the intervening decades, structural biologists have systematically revealed atomic-level detail for tens of thousands of other molecular processes occurring within living cells; these insights, in turn, have led to numerous breakthroughs across many fields, including healthcare and material sciences.

Today, at the opening of this Shonan meeting on web molecular graphics, it may be worth pondering the well-known quotation attributed to Albert Einstein: "What does a fish know about the water in which he swims all his life?". Applying this to our own field, we can ask: do we understand the reasons underlying the extraordinary success of structural biology? Certainly one reason is because our tools are built upon a solid bedrock of exceptionally well-managed data. Created in 1972, the Protein Data Bank [7] has exemplary practices and stability that facilitate reproducibility and substantially simplify the difficult task of creating and maintaining tailored visualization tools. However, I believe an even more important reason may be the primary role played by computational visualization.

In many fields of life science, much of data landscapes being explored can be observed directly and are therefore usually familiar — this applies to ecosystem data on geographical-scales, to tissue-scale data, down to the organization of sub-cellular organelles accessible by microscopy. Computational visualization in these fields is often more straightforward: the terrain is already known, thus mistakes can often be immediately recognized and corrected. In contrast, much of the landscape for biomolecular structures remains unobserved, unfamiliar, and unknown — as a result, computational visualization plays a more primary role; visualization has been an essential tool in our struggle to understand and explain biological phenomena on the molecular scale — it has also always been integral to how 3D structural models of biomolecules are derived from experimental data.

As a result, structural biologists have long been early-adopters for cutting-edge visual methods, starting with physical models used in solving the first protein structures [35], stereoscopic imaging [23], and virtual reality (VR) [16]. More recently, this has continued with adoption of low-cost VR (e.g., Oculus Rift is supported by VMD, [31]), very low-cost VR (e.g., Google Cardboard [6]), 3D printing [28], commodity interaction devices (e.g., Leap motion & Kinect, [59]), augmented reality (AR) [27], crowd-sourced evaluations [30], and concepts from computer gaming [46].

Thus, driven by necessity and to some extent desperation, visualization tools for macromolecular structures have been more advanced compared with many other scientific disciplines [51]. This has had at least one unintended, positive consequence: more than other fields, we have been taking advantage of the extraordinary capacities of the human visual system, which can easily man-

age much greater data density and complexity than found in most published scientific figures [36, 29, 67]. This Shonan meeting very much continues this long-standing tradition: once again, molecular graphics is an early adopter for some of today's cutting edge visualization technologies, such as WebGL as well as web-based augmented and virtual reality.

However, we should remember that being early comes with a price: for example, consider earlier efforts to bring scientific visualization to the web. A large investment in time and effort was spent by the scientific community on creating sophisticated Java applets; this has now been squandered due to decisions made by companies such as Oracle, Google, and Apple. As we now plan a similar scale of development based on web technologies such webGL, we should learn from these past difficulties. A practical step would be for us to lobby with browser companies directly, or with organizations responsible for maintaining web standards, such the ECMA technical committees and the Khronos group.

The need to future-proof the work that we are now planning and undertaking is more pressing than with previous efforts; while much of the past focus of molecular graphics was on the effective display of single PDB structures, we are now focused on much larger, more ambitious goals. Rapid advances in experimental methods such as especially cryo-electron microscopy (cryo-EM) are making accessible vastly larger structures and molecular assemblies than ever before [5]; this has promoted improvements in methods for visual exploration of multiscale molecular data (e.g., LiteMol, [62]). In addition, advances in high-throughput computational modeling now make it feasible to systematically calculate comparative 3D models across all known protein sequences, resulting in databases of $>$ 100 million 3D models [53]. Similarly, high-throughput computational approaches are also being applied to molecular dynamics [58], generating increasingly large, complex trajectories, requiring further innovations to create very specific, tailored visualization tools [31, 52, 54]. Finally, high-throughput computing is also being used to integrate structural data in the construction of atomic-scale models of viruses, sub-cellular compartments, or even whole cells [34]. The scale and complexity of these models requires the development of radically new visualization methods, bridging 3D structures with molecular data (from genomics, transcriptomics, proteomics, and systems biology), as well as data from higher-level scales, such as the cells, tissues, even populations [26].

Pre-modeling prior expectation in binary task machine learning, and visualizing 'what' machine learning is doing

J.B. Brown, Kyoto University, Japan

Machine learning (ML), also known as artificial intelligence (AI) more recently, is being applied to countless areas of research and industry. Yet, the general public and even those who are developing ML models fail to understand when and why a model that seems good during internal validation fails to work in novel, prospective situations.

As a direction toward solving this problem, a visualization method for understanding the metric values possible in a given ML situation has been proposed, known as the metric surface method [9]. More specifically the method employs

a combination of (1) the ratio of positive/negative data available or to be predicted, and (2) a binary classification metric (e.g., Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$) to yield a visual image which conveys the challenge (or ease) of a binary classification task.

In addition, recent classification research has demonstrated that large molecular datasets can in fact be efficiently classified using only 5% to 20% of data [55]. The method which has resulted in this discovery, known as active learning, typically provides a plot of the iteration or fraction of data on the independent axis and the prediction performance on the dependent axis. Yet, metrics such as Accuracy make it difficult to understand the reason for the gains in performance; that is, the 'what' aspect of ML's success cannot be answered by such a typical time-series plot. As a way to resolve this issue, Brown has proposed the 'active projection' method [8]. In this method, the True Negative Rate ($\frac{TN}{TN+FP}$) and True Positive Rate ($\frac{TP}{TP+FN}$) are used as axes, and a third metric such as Accuracy is used for the background metric surface which characterizes the ratio-metric setting.

By systematically plotting the evolution of the continually-updating model performance obtained during active learning (e.g., the TPR and TNR against the backdrop of ratio-metric), the visualization provides an interpretation that leads to significant insights for the research team. This method will become critical as automation of synthesis and assay becomes commonplace, and its adaptation to a web-based framework will allow drug discovery and chemical biology scientists to execute wet-bench assays in one location while confirming the results in another.

Analysis of dynamic protein structure

Barbora Kozlikova, Masaryk University, Czech Republic

Molecular dynamic (MD) simulations continue to play an ever more important role in the analysis and understanding of protein behavior and function. The current computational power enables to capture very long MD simulations, consisting of hundreds of thousands of time steps, which cannot be explored using the traditionally used animation of protein movements anymore. Here, visualization can help substantially to explore such large data in a more efficient way, enabling to reveal potentially interesting parts of the simulations. For this purpose, we have developed several visualization methods and visual analysis tools, which help the protein engineers to visually explore MD simulations. These methods were presented within the talk.

Our collaboration partners from the protein engineering field are focusing mostly on detection and exploration of tunnels in proteins, which can serve as potential transportation paths for ligand molecules to protein active site. Here a chemical reaction between the protein and ligand can undergo and the product of such a reaction can, for example, serve as a basis of new medication. Among many existing tunnels in a protein, detected using computational geometry methods, only several can be used as transportation paths. Among those belong some tunnels which are the most stable over time. Therefore, our first visualization methods focused on the exploration of tunnel evolution over time [12, 11]. In the following phase, we started to design visualization

techniques suitable for analysis of MD simulations containing the ligand transportation. The very complex ligand trajectory can be explored using our visual analysis tools [24, 20]. In [68], we extended the concept of ligand trajectory exploration to water molecules. In this case, the complex task is to explore the behavior of thousands of water molecules interacting with the protein at once. For that purpose, we came with a set of interactively linked views, enabling to explore the trends in water molecules behavior and focus only on a subset of interesting ones.

Currently we are interested in designing techniques for visual exploration and filtering of large ensembles of long MD simulations, where we are incorporating different properties (e.g., ligand distance to the active site, energetic profile of the transportation path) in order to reveal interesting events happening in such complex data sets. These efforts are supported mainly by the Czech Republic national grant agency (GA17-07690S), enabling us to cooperate with experts in computational geometry and robotics on this complex problem.

In terms of visualization of MD simulations in web environments, there are already existing pioneering solutions for small molecules and short simulations. However, complex data, as described above, cannot be processed interactively using current capabilities. Therefore, future research possibilities lie in designing web-based techniques that will enable researchers to visually explore very long MD simulations. This spans from techniques for visual abstraction of simulations, to conveying the information about individual time steps.

Whole-cell visualization and modeling

Ivan Viola, KAUST, Saudi Arabia

Computer graphics technology can nowadays interactively display billions of atoms forming structures up to entire biological organisms such as bacteria or protista. The key insight that allows it is the following: while biology is immensely complex, it is also very repetitive, which could be exploited in coping with the complexity. Life forms are internally composed of evolutionary successful patterns that are frequently repeated. This repetition, or multi-instancing in our terminology, can be observed on every level of spatial organization. Thanks to such patterns, model construction of the entire life form can be efficiently parallelized and consequently displayed using fast rendering routines, where both stages are executed on the graphics hardware.

In terms of complexity, resulting scenes are of multi-scale, multi-instance, crowded, and dense three-dimensional nature. To effectively convey such complex structural arrangement, visualization algorithms need to cope with all of these structural characteristics simultaneously. This need triggers the necessity of visualization algorithms that handle novel problems in 3D occlusion management, color assignment, shading, or textual labeling. New algorithms that can again take advantage of structural hierarchy, and repetitiveness. Another successful strategy is to tame the complexity with efficient view-guided image-space algorithms. Instead of computationally-demanding algorithms that compute the solution globally for the entire model, these techniques focus on a local solution that is perfectly tailored for limited viewpoint settings, but can be calculated for each image without notably penalizing the overall visualization performance. All these new algorithms lead to gradual democratization of computer graphics

and visualization technology for structural biology. Ultimately, the advances lead into new exciting ways how biology can be explored, understood, and communicated in the future.

The Mol* project

Alexander S Rose, University of California, San Diego/ San Diego Supercomputer Center/ RCSB Protein Data Bank, USA; David Sehnal, PDBe, CEITEC, Czech Republic

Rapidly evolving experimental methods (X-ray free electron laser crystallography, XFEL; cryo-electron microscopy, cryo-EM) as well as emerging Integrative/Hybrid methods pose immense challenges of growing data size and complexity for (not only) web-based data visualization and delivery. XFEL investigations can produces hundreds of individual macromolecular structures corresponding to complex chemical reactions and biological processes. In cryo-EM, experimental density maps for large high resolution structures are multiple GB large. Integrative/hybrid experimental methods for determining three-dimensional structures of biomolecules provide the means for studying large molecular complexes. These structures typically consist of multiple components depicted using models of varying resolution and length scale (e.g., all atom representations, gaussian shapes).

Web-based visualization and analyses of macromolecular structures and associated data represents a critical step in enabling access and gaining knowledge from these data. Embracing advances in browser technology provides the means for creating scalable molecular graphics and analysis tools with near-instant data access. To meet the challenges posed by evolving and emerging methods we initiated an international collaborative open source project, called Mol* (/'mol-star'/, https://molstar.org), to develop the next generation web molecular graphics, analysis and data delivery services. We present herein the development and features of the project and its common library for macromolecular visualization and analysis to facilitate building tools and services for the scientific community [63].

Tools and tricks for visualizing proteins in cells: BioBlender and the making of SciVis movies

Monica Zoppè, Institute of Clinical Physiology (IFC), CNR, Italy

The Scientific Visualization Unit of IFC, CNR, in Pisa Italy, has been dedicated to the creation and development of cellular animations for several years. During this time we have produced BioBlender, a tool that allows the elaboration of molecular data directly into one of the most sophisticated Computer Graphics packages, Blender, as well as a series of short videos, describing some cellular processes. These videos are intended to combine the highest available scientific accuracy, with the most advanced tools for representing molecules in their cellular environment. With BioBlender, atomic data are imported from PDB or other databanks, and are treated using a series of scientific methods [71] to calculate their motions, and to display chemical and physical features following a specific visual code, developed in the lab. Among the most important

forces driving molecular interactions in the aqueous environment of the cell, are the hydrophobic and the electrostatic potentials. These are calculated and converted onto visual features using a combination of scientific and CG tools, thus showing hydrophobicity as a texture feature displayed on the surface, going from white-smooth-shiny for the most lipophilic, to dark-rough-opaque for the most hydrophilic [3]. In parallel, as the electrostatic force exerts its effect at some distance from the surface of the molecule, its presence, force and directions are visually expressed as series of particles travelling along the field lines [73]. The direct integration with Blender allows the creation of the cellular environment, by applying the techniques of CG to reproduce the best knowledge available from literature. For example, membranes (which are made of lipids, and separate specific compartments), are shown in the video as structured surfaces, obtained by applying both bump (mesh) and texture displacements, made to display the 'heads' of lipids, with randomly varied sizes of Voronoi cells, changing in time.

The entire process is only partially automated, and much personal artistic input is required at many steps, in particular those related to film direction, such as camera views (photography), lighting, movements and the accompanying titles and sound track. The main steps for each video are reported in 'Explanatory Notes' that can be downloaded from the www.scivis.it website. The key factor for the success of the making of the movies was the building of a team composed of single persons very expert in their own field, but also willing to explore and learn the general (and sometimes detailed) aspects of other disciplines involved, which ranged from biology, chemistry, physics and math, to computer graphics, programming, and including artistic and musical inputs. During the talk, the major challenges of scientific, technical and perceptive nature are described and commented.

Molecular visualization: From molecular questions to technical advances — and vice versa

Matthieu Chavent, IPBS, CNRS, France

Visualizations for computational biology have been developing for over 50 years. With recent advances in both computational biology and computer graphics techniques, these fields have witnessed rapid technological advances in the last decade. Nevertheless, there remains a gap between the two communities of visualization and computational biology, resulting in additional challenges to bridge the divide [1].

Working at the interface in between computer graphics and modeling resulted in developments in both fields in order to better describe molecular objects. From Molecular surface depiction [14] to rendering of lipid flow in large membrane models [15], these methodologies were designed to help computational biologists. This resulted in the creation of different tools accessible through modules in the well known VMD molecular viewer such as Bendix, to analyze alpha helices bending [18] or cavities volume in proteins [43].

Nevertheless, it is now necessary to go further and tightly collaborate with computer graphics in order to develop new ways of rendering molecular models especially in the context of larger ad larger systems [13]. We recently applied

computer graphics techniques to better describing membrane lipid-lipid [2] and protein-lipid interactions [50].

Technical aspects of web-based immersive molecular visualization

Michael Krone, University of Tübingen, Germany

Web Molecular Graphics and Visualization shares a lot of technical challenges with other application domains; however, it also poses challenges that are specific to this area. In my talk, I will discuss recent technical advances in the field and their potential for web-based molecular visualization.

Based on the recent state of the art report on web-based visualization by Mwalongo et al. [49], I will first give an overview of current methods and technical possibilities, and provide examples for their applicability to web-based molecular visualization. This includes for example remote visualization methods, compression techniques for fast data transfer, and GPU-accelerated rendering via modern WebGL. I will also present technological advances that are not yet available for web-based applications and which would greatly increase the possibility for web molecular graphics. Examples are support for GPU computing in the browser, WebGL support for features of modern desktop graphics hardware like geometry/tessellation shaders, or faster data transfer.

In the second part of my talk, I will discuss the emerging topic of web-based immersive analytics for molecular data. Immersive display technologies like large displays and head-mounted displays for Virtual or Augmented Reality (e.g., HTC Vive, Google Cardboard, or Microsoft HoloLens) open up new possibilities for molecular visualization [39]. However, web-based applications that make use of these methods are yet rarely found, despite the availability of software frameworks like WebVR that provide convenient access to these technologies. I will give examples for immersive molecular visualizations and pose open challenges and questions for this field of research.

4 Breakout group discussions

Review of the state of the art

Marc Baaden, J.B. Brown, Bob Hanson, Michael Krone, Andrea Schafferhans, Ian Sillitoe, Masakazu Sekijima

The background of this breakout group is an initiative for a review article about web representation of molecular graphics in the browser, which had been started VIZBI 2016 with Seán O'Donoghue and Björn Sommer. In this breakout session, the focus and structure of the manuscript was discussed and improved. Subgroups then defined the content of specific subsections in more detail.

- Introduction: [[AUTHORS: Seán O'Donoghue, J.B. Brown]]
  - Experimental methods yielding more molecular data, but key is how to deliver it to biologists and chemists in an interpretable format, regardless of device and technical skill level.

  - Currently can perform homology search systematically, but a visual interpretation of that result is often still not automated.
  - Emerging field of Web Molecular Graphics to address these needs.
  - How does WebMolGraph differ from focus of HTML5+WebGL (Cell 2017 review).
  - How does WebMolGraph differ from standalone applications?
  - FIGURE: overview on molecular representation and knowledge discovery from it.
  - What is capable? We showcase example web applications available for driving insight from biological, chemical, and pharmaceutical data.

- Tool introductions and concepts [[AUTHORS: Sameer Verlanker, Ian Sillitoe, Andrea Schafferhans]]
  - MASTER TABLE: Table of tools selected. Issues include open source and maintenance efforts.
  - Critically, what is available and what is not available/done.
  - FIGURE: Multi-scale and large-scale (Web) visualization.
  - Supplementary data: Box on molecular visualization.
  - BOX: components/libraries versus applications/tools. "Brushing and linking (M. Krone)"
  - Fully managed data (e.g., from a DB) versus Drag-and-Drop applications (JSMol for 3D structure, sequence copy-and-paste)
  - Educational applications of the web molecular graphics applications (transit to emerging issues).

- Emerging issues [[AUTHORS: Michael Krone, Marc Baaden]]
  - External data and annotation above basic visualization. (Either from DB or manual annotation).
  - FIGURE: Visualization of visualization + associated annotation. 1D vs. 3D, with 2D intermediate as well. 2 to 3 examples
  - Visual analytics w/ FIGURE: beyond pretty images, web-driven pipelines, visualization entropy (protein versus graphene), evaluation of information value in visualization.
  - BOX discussion: new data as a web address, de-coupling of data versus tools, re-use of existing tools that retrieve data from web address prior to developing a new visualization, RDF/SPARQL (future data re-usage) [AUTHOR: J.B.] .
  - Java/Chime fade-out as a case-in-point, 3D printing accessibility.
  - BOX: integration of VR/AR to interact with molecules – inspection of pocket depth, follow a ligand into a pocket based on simulation or resolved 3D structure stored in web databases [Marc/Bjorn/MKrone].

- Perspectives [[AUTHORS: Seán O'Donoghue, J.B. Brown]]

- Support and funding (inter-governmental, etc)
- Standards declaration body for web molecular graphics
- Preparing for upcoming eras (and challenges) in data modalities, and ensuring that existing components are available to create custom visualizations/applications with less repetition of innovation.
- Learning curves and investments required to leverage open components and develop new applications.
- The demarcation between for-fee industry and for-free open standards – clarifying where it does make sense for companies to pay for software and services.

Universal molecular scene representation

Bob Hanson, Alexander S Rose, Marc Baaden, David Sehnal, Radka Svobodova

The discussion focused on the development of what we are preliminarily calling 'Universal Molecular Scene Representation'. It is an outgrowth of discussions from the 2016 Shonan meeting. The key is the Mol-* method of saving the state as a network of transformation steps (a finite acyclic digraph) developed by David Sehnal and Alexander S Rose.

Principal aspects of UMSR:

- UMSR defines a set of interface methods (functions that have well-defined typed inputs and outputs that, when stringed together in a nodal tree graph, define the state of a molecular visualization.
- The standard is not about implementation. That is still totally in the hands of a developer. The standard defines only what could be developed, not what must be developed. The result is the ability of different programs to selectively implement the methods in order to reproduce the desired molecular scene state as best they can. It is likely that no program will be able to fully implement the final/living standard.
- The form of the saved representation of UMSR will be a JSON structure that uses standard JavaScript types directly serializable using JSON.

Community aspects of UMSR:

- UMSR will be open-source
- The developing standard will be housed at GitHub, including descriptions of transformations, examples of implementation, and test suites.
- The scheme allows for extensive 'plug-in' capability, since any transformation can in principle be 'overwritten' (in the Java vernacular) to suite, and any additional transformational methods could be proposed, preferably with implementation examples.
- Participation by any and all developers of web-based molecular visualization software are encouraged to participate.

- Any developer is free to propose or implement any number of well-defined transformations.
- Semantic versioning will be used.

Development plan:

- Start small, with very simple visualizations.
- Build a suite of tests starting with what Mol-* creates as "UMSR v. 0.0.0
- Provide links to specific implementations of the tests.
- Develop a validation scheme.

Tasks and strategies:

- David Sehnal and Alexander S Rose will develop a set of descriptions of primary transformations.
- David and Alexander will review current Mol* processes with an eye toward generalization.
- Bob Hanson will build into JSmol first the capability to read UMSR JSON, then to write it as well.
- Bob Hanson will leverage JSmol's limited capability to read Pymol session files and write them to UMSR. This will provide a first case and template for interoperability.
- Seán O'Donoghue has expressed interest in involvement in relation to Jolecule.
- Marc Baaden has expressed interest in involvement in relation to UnityMol (also together with Xavier Martinez, main developer).

Release and publication plan: ISMB/ECCB 2019 joint announcement (submission deadline Jan 31, 2019).

Massive, mesoscale, and multiscale data

Matthieu Chavent, Martin Falk, Barbora Kozlikova, Peter Mindek, Alexander S Rose, David Sehnal, Sameer Velanker, Radka Varekova, Ivan Viola, Monica Zoppè

With the advances in structural biology (with the rise of CryoEM and Cryo tomography techniques) as well as in computational biology, it is now possible to visualize larger and larger molecular systems up to the cellular level [33]. This will require new ways of rendering such large systems in function of the structural resolution available as well as the viewer position. Furthermore, it is now important to not only represent a static 3D object but include the intrinsic dynamics of these molecular 'machines'.

The goal of the session was discussing how to perform a tool that can handle a multiscale view and deal with heterogeneous input data. While it seems that computer graphics techniques exist at each scale the real challenge is now to integrate these different techniques in order to seamlessly pass from one to the other.

During this discussion numerous questions appeared: "What happens when we do not know the time scale ?" (S. Velanker), "How to display multi-dimensionality data while a human can only handle objects spanning 2 to 3 log of dimensions at all?" (M. Zoppè) [72]. The issues related to large datasets and multiscale rendering still remain very hard to tackle but we identified some strategies which may aim towards some solutions: first, we need to design a Level of Detail (LoD) algorithm taking into account both time and scales. In this respect, renderings used for astrophysics constitute a good example to follow. Then, it will be necessary to unify existing schemes commonly used (recipes to quickly create simple primitives like lines, spheres, surfaces, etc ..) and develop a hierarchical tree/graph to pass from one scheme to the other.

Methods for analysis of multi-ligand collections

J.B. Brown, Bob Hanson, Andrea Schafferhans, Masakazu Sekijima, Sameer Valenkar

The pharmaceutical industry and chemical biology research fields are concerned with screening large numbers of chemical compounds against protein targets involved in biology and medicine. As a result, they have the digital data for hundreds to millions of ligand (compound) structures, and visualizing these data intuitively is a challenge. More importantly, the processes of synthesis and evaluation are becoming increasingly automated [61], with the possibility of synthesis and screening in one location yet the visualization of the results to be done in a second, remote location.

While protein visualization over the web has made steady progress, corresponding compound visualization has yet to be addressed sufficiently. One method towards the creation of visual approaches to aid chemists is the SAR matrix method [69]. In this approach, a compound structure with two defined locations for substitutions (e.g., replacement of $-CH_3$ by $-CF_3$ or $-COOH$) is transformed to a grid with one axis for the R_1 substitutions and another for the R_2 substitutions, and the resulting grid resembles a heatmap, where colors in the heatmap correspond to a property from a specific pair of R_1 and R_2 substituents. For example, a steroid scaffold might be decorated by a methyl group on R_1 and a hydroxyl group on R_2 with a corresponding EC_{50} concentration of 38nM. As there is a possibility that not all pairs of R_1 and R_2 are tested, it is possible to have missing values in a matrix.

The challenges discussed in this breakout group pertained to the following major topics:

- Prior to SAR matrix display, how to efficiently organize a multi-ligand dataset in a web browser when many compound scaffolds are present?
- How to efficiently present a SAR matrix when three or more replacement R-groups are possible.

- What are the server- and client-side roles of such a framework? Should this depend on the dataset size?
- How to keep the technology deliverable in a standalone fashion where network firewalls are present to prevent data transmission (e.g., data theft at a pharmaceutical company or imposed restrictions between research teams).

A collection of solutions was also considered:

- Ligand data could be organized by framework or scaffold. To this end, the Matched Molecular Pair framework [32] could be used to create a network among scaffolds such that pairs of scaffolds are linked if they differ by a single transformation. In the web viewer, each scaffold would then be annotated by statistics related to a measured value (e.g., min/max/mean of bioactivity on a specific receptor), where the use of color could be used to accelerate analysis. The browser view could be initially positioned, for example, on the scaffold with the largest variance in bioactivity; that is, where changes in R-groups yield large shifts in bioactivity.
- When scaffolds contain three or more R-groups, a tabular view can be presented, organized by R group columns and bioactivities or other endpoints. This table should be interactive and immediately re-sortable by a specific R-group or endpoint, including a statistical value derived from raw values.
- One solution to the client- and server-side issue would be to use server-side rendering of flat 2D structures (e.g., by the open source RDkit package) prior to the data being transmitted for visualization in a client. In cases of large (10,000+ compound) libraries, this would alleviate strain on the browser device.
- Even for network firewall environments, a solution which can retrieve the requested data via a URL would provide a single interface to data retrieval and analysis. Additional design considerations are necessary to systematically obtain related compounds across multiple databases; efficient server-side retrieval and filtering before transmission to a client will benefit from a discussion from the primary web-based data providers (e.g., ChEMBL, PubChem, DrugBank, etc).

Visual recipes for molecular graphics

Seán O'Donoghue, Marc Baaden, Martin Falk, Barbora Kozlikova, Michael Krone, Alexander S Rose, David Sehnal, Radka Svobodova, Ivan Viola, Monica Zoppè

This breakout discussion initially centered on finding the best ways to visually represent biomolecular information, especially data on properties related to, but distinct from, structure, such as pH, chemical gradient, forces, electrostatics, energy flow, lipophilicity, and hydrophilicity. The discussion quickly broadened to consider additional properties, such as flexibility, dynamics, molecular paths,

disorder, uncertainty, and missing data. When scientists need to visually represent such properties, the task is often challenging, and there are currently few generally agreed upon standards to follow.

In some cases, the properties can be suitably encoded using one of the generic, well-described visual channels (e.g., color luminance, saturation, or hue) — this, in turn, has the advantage that these channels have been ranked in order of visual effectiveness, thus helping scientists choose the best encoding (e.g., see Figure 2 from O'Donoghue et al., 2018 [51]). Often, however, less standard visual encodings may be needed, such as lighting, texture, material properties — possibly also non-visual channels such as data sonification. In such cases, finding an effective solution can be difficult, and may require studying literature from very different scientific domains. For example, to find good methods for showing ligands occupying binding sites on proteins, a structural biologist might need to read publications on how void space surfaces are used in medical imaging (e.g., Kreiser et al., 2018 [40]). Or, to improve the visualization of biomolecular electrostatics, it may be useful to examine how contour lines are used to show air pressure in meteorology or elevation in cartography.

In many cases, however, the use of generic visual encodings is inadequate, and tailored visual strategies are required [51] — while such strategies may be described in previous literature, this literature is not always easily available or accessible. For example, Krone et al. recently published a comprehensive review on the visualization of biomolecular cavities [42]; unfortunately, the article is not indexed in PubMed and the journal that it appears in (Computer Graphics Forum) is not accessible to many biologists or biomedical researchers.

The use of specifically tailored representations can be powerful, sometimes enabling concise visualization of multiple properties; for example, simplified biomolecular surfaces can be used to show simultaneously key spatial details, molecular interaction sites, as well electrostatics [17]. While this strategy can be very useful, there is generally a limit on how many properties can be shown before the visual channels used begin to interfere with each other, resulting in representations that are overcrowded and confusing. This limit can be extended, to some extent, using visual strategies that enable the viewer to interactively highlight or focus on specific properties of interest (e.g., using the visual metaphor of 'semantic depth of field' [37]).

One of the major challenges is visualizing dynamic properties — an unavoidable issue since dynamics are an essential feature of biomolecular function. Here again, structural biologists can draw upon generic visual strategies recently developed for depicting visualization in other fields — such as mechanical engineering [47] — as well as strategies specifically tailored for depicting biomolecular motion [10].

An important issue that arose in our discussion was that, for some of the above visual challenges, an effective strategy may be to fundamentally re-frame the problem — for example, by escaping the familiar 3D molecular landscape and re-casting the problem using a 2D projection method (e.g., Krone et al., 2017 [41]).

After discussing a range of strategies that could help life scientists facing the above challenges, we concluded that a useful step would be a publication articulating this problem, and calling for a discussion focused on articulating visual recipes for molecular graphics problems. To be successful, this publication would need to be co-authored by specialists in molecular graphics, as well as

researchers focused on more general computer graphics methods. Many of the participants in this breakout felt this would be a worthwhile goal, and that — together with some additional colleagues — we would have the appropriate range of expertise to make such a publication successful.

VR & AR for molecular graphics and visual analysis

Michael Krone, Marc Baaden, Martin Falk, Bara Kozlikova, Andrea Schafferhans, David Sehnal, Masakazu Sekijima

Virtual Reality (VR) and Augmented Reality (AR) have recently gained a lot of attention besides the obvious use for entertainment, mainly due to the availability of affordable consumer-off-the-shelf hardware like Oculus Rift, HTC Vive, or Microsoft HoloLens. The emerging field of research called *immersive analytics* tries to answer the open question of how these immersive technologies can be used to enhance the visual analysis or exploration of data [64]. We discussed this question with respect to web molecular graphics and defined a list of open challenges that have to be solved for an effective application.

We first focused on the question: Which application scenarios concerning molecular data would benefit from immersive visualization? In molecular graphics, stereoscopic rendering for the visual inspection of complex, three-dimensional molecular structures has a long tradition. Therefore, we envision that immersive data exploration in VR using head-mounts displays (HMDs) is just the next logical step in this application area. VR HMDs provide immersive, stereoscopic images in combination with natural user interaction, for example, through head tracking or gestures. From our point of view, AR offers basically the same immersive experiences as VR; however, the additional benefit of modern see-though AR HMDs is that the user is not completely shut off from the environment. That is, he or she can still communicate with other people, take notes, or look at an additional computer screen[38]. The drawback of current AR HMDs is that the imaging is not as clear as in VR due to the display technology. Immersive data exploration can also be used to analyse molecular simulation results. An example is UnityMol [46], which can not only render 3D structures in VR, but also additional 2D panels showing plots with additional information about the simulation for a comprehensive visual analysis (http://bit.ly/Baaden2018).

Probably the most promising application for immersive environments—which has not yet been extensively explored—should be molecular modelling, for example, fitting atomic structures into CryoEM [48] or X-ray maps, folding proteins [65], or arranging molecular structures in a cellular environment [25]. Similar to the immersive data exploration, users will benefit from the stereoscopic vision, which facilitates the modelling task. Furthermore, the user interaction would be more natural and intuitive when using a tracked controller or hand tracking/gestures.

A third prospective application area can be summarized as dissemination, outreach, and education. Here, high-quality visuals and smooth interaction together with annotations providing additional information are important. In this scenario, immersive molecular visualization can be used to increase the engagement for lay viewers as well as students on all levels of education [66]. Serious

games may be an interesting target where the visual aspect is of primary importance for instance [4]. While there are already some applications that make use of AR and VR, we still see a high potential for future developments and improvements in this area. Low-cost solutions like Google Cardboard that enable a VR experience on the users' mobile phones make this application scenario very appealing for outreach projects (e.g., in a museum or scientific exhibition) or in the classroom. Especially in this context, web-based visualization is an ideal choice, since interested users or students can experience the content without installing apps on their laptops or mobile devices.

As user interaction is an important factor for immersive environments, we also discussed possible devices that would enable users to interact more naturally. While the tracked game controllers, which are usually part of a VR HMD kit, are good for games and low-precision input (e.g., as a virtual pointing device), many immersive analytics tasks require high-precision input. A tracked computer mouse could be a simple solution for this, as it would allow users to select small items in the virtual environment (just like a regular computer mouse), for example, on the 2D plot panels in the previously mentioned UnityMol application for molecular simulation analysis [46, 19]. Another useful device would be a haptic glove that would allow users to feel and grab virtual objects in VR or AR. Currently, only prototypes of such gloves exist. One example of such a prototype is the HaptX glove (https://haptx.com/), which can exert strong forces on the users hands and even emit heat and cold, but it is still relatively big and tethered to a base station that provides the necessary pneumatic power. In contrast to haptic gloves, 3D printing can — already today — provide tangible models of molecular data, which could be tracked and used as proxies for interaction. In AR, the 3D printed models could additionally be overlaid with visualizations, for example, showing the electrostatic field around the molecule. Other modes of input that could improve immersive molecular visualization environments are brain interfaces (like the Emotiv Insight: https://www.emotiv.com/insight/) or eye tracking in HMDs.

The above-mentioned application scenarios for molecular visualization using immersive environments could also benefit from using sound as an additional means to convey information. This so-called sonification could lead to a higher immersion and, consequently, engagement. Especially for large, multidimensional data, using the auditory channel could also provide an intuitive way for users to get more information at the same time compared to using only visual output. One caveat for this technique that we discussed was that the sounds should probably be mostly pleasant for the user, since an unpleasant auditory environment will probably drive users away.

In summary, we see a high potential for web-based immersive visualization applications for molecular data. While we did not focus so much on the technical challenges that have to be solved for web-based immersive applications, we rather discussed potential use cases and hardware developments that could drive this technology.

Visualising protein families

Ian Sillitoe, Bob Hanson, Alexander S Rose, Radka Svobodova, Sameer Valenkar

A number of online resources exist (Pfam, CATH, SCOP, InterPro, etc)

that cluster proteins, or more specifically protein domains, that are related by evolution. These collections, often termed 'families' or 'superfamilies', can provide clues on the biological features that are important to function since these features will be preferentially conserved during evolution. Many tools already exist that help to visualise and analyse the conservation of these features (e.g., phylogentic trees, multiple sequence alignments). However the recent advances in web-based molecular graphics provide opportunities for the development of novel analytical visualisations and tools.

The group discussed the types of biological features that often appear in analysis of conservation (listed in order of scale):

- Amino acid identity / property
- Secondary structure
- Protein fold (arrangement in 3D space)
- Domain organisation (sequential arrangement of domains within proteins)

We also summarised the visualisation tools that already exist (categorised by dimensionality):

1D:

- Protein amino acid sequence
- Residue-based features (e.g., active sites)
- Domain organisation (sequential assignment of domains within a protein)

2D:

- Multiple alignments (based on sequence or structure)
- Phylogenetic trees
- Residue-residue contact maps
- Secondary structure topology diagrams (e.g., HERA)

3D:

- Visualisation of 3D structure

The mother of all demos

Seán O'Donoghue, Marc Baaden, Matthieu Chavent, Alexander S Rose, David Sehnal, Ivan Viola, Monica Zoppè

This breakout was initially motivated by Douglas Engelbart's famous 'Mother of all Demos' presentation [22], given in 1968 at the Fall Joint Computer Conference in San Francisco. This inspired us to contemplate how a 'mother of all demos' could look for web molecular graphics.

We first defined overall goals of such a demo: it would be designed to be especially memorable and visually impressive, thus inspiring enthusiasm for — and awareness of — the emerging field of web molecular graphics. Additionally, we would aim to use the demo to help ensure that our field is visible to — and supported by — key stake-holders and organizations in web, such as the KRONOS group, the W3C consortium, as well as key browser makers, such as Apple, Google, and Microsoft.

During the discussion, we considered how the demo could be composed by a mashup comprising existing web-based resources, such as NGL viewer [57], LiteMol [62], Mol*, Aquaria [53], and Genome3D [45]. Ideally, the demo would include molecular landscapes, such as those constructed in cellVIEW [44], and would allow interactive visual exploration using VR and AR devices, as is done in UnityMol [46, 19].

A key question that we discussed concerned the possible scenarios or stories that the demo should focus on. We concluded that an important virus (e.g., HIV) may be a good central subject, as has been done in the existing NGL Viewer demo [56], as well as the Mol* 'Capsing Capsids' demo, first presented as this workshop by David Sehnal and Alexander S Rose. In addition, inspired by the influential 'Powers of Ten' documentary film from 1977 [21], we discussed using zoom as a central feature of the demo, spanning the molecular and cellular scales.

Finally we considered upcoming scientific meetings where the demo could be presented — suitable venues could include the annual meeting on Visualizing Biological Data (VIZBI), or the 3D-SIG and BioVis tracks at the annual ISMB conference.

Community building and funding

Sameer Velanker, Marc Baaden, J.B. Brown, Matthieu Chavent, Martin Falk, Bob Hanson, Barbora Kozlikova, Michael Krone, Peter Mindek, Seán O'Donoghue, Alexander S Rose, Andrea Schafferhans, David Sehnal, Masakazu Sekijima, Ian Sillitoe, Radka Varekova, Ivan Viola, Monica Zoppè

To realize maximum impact, this breakout session discussed building a community by bringing people together for mid-term collaborative work through workshops and meetings. A number of possible funding resources for such workshops and meetings were identified — BBSRC International Workshops calls (similar to https://bbsrc.ukri.org/funding/filter/international-workshops/), Wellcome trust collaborative grants, and the Pistoia alliance which has already held workshops on visualisation. The possibility of a EC COST Action grant (https://www.cost.eu/actions/CA18127) was discussed. EC COST Action programme specifically targets building international community to develop standards and policies. This would be a good mechanism to coordinate efforts to further develop the nascent web-based molecular graphics community. The next COST Action proposal submission deadline seems to be Sept 2019.

An EU project Visionair (http://www.infra-visionair.eu/index.php) looks relevant and it may help to establish contact with this community. Visionair calls for the creation of a European infrastructure for high level visualisation facilities that will be open to research communities across Europe and around the world.

There was also discussion about having close interaction with ELIXIR and trying to establish an ELIXIR community. This will require contacting ELIXIR and finding out the requirements for such an application.

5 List of participants

- Marc Baaden*, IBPC, CNRS, France
- J.B. Brown, Kyoto University, Japan
- Matthieu Chavent, IPBS, CNRS, France
- Martin Falk, Linköping University, Sweden
- Bob Hanson, St. Olaf College, USA
- Barbora Kozlikova, Masaryk University, Czech Republic
- Michael Krone, University of Tübingen, Germany
- Peter Mindek, TU Vienna, Austria
- Seán O'Donoghue*, CSIRO & Garvan Institute, Australia
- Alexander S Rose, University of California, San Diego/ San Diego Supercomputer Center/ RCSB Protein Data Bank, USA
- Andrea Schafferhans, Hochschule Weihenstephan-Triesdorf, Germany
- David Sehnal, PDBe CEITEC, Czech Republic
- Masakazu Sekijima*, Tokyo Institute of Technology, Japan
- Ian Sillitoe, University College London, UK
- Sameer Velanker, PDBe, EMBL-EBI, UK
- Radka Varekova, Masaryk University, Czech Republic
- Ivan Viola, KAUST, Saudi Arabia
- Monica Zoppè, Institute of Clinical Physiology, CNR, Italy

* Organizers

References

[1] Naif Alharbi, Mohammad Alharbi, Xavier Martinez, Michael Krone, A Rose, Marc Baaden, Robert S Laramee, and Matthieu Chavent. Molecular visualization of computational biology data: A survey of surveys. *Eurovis Short Papers*, pages 133–137, 2017.

[2] Naif Alharbi, Robert S Laramee, and Matthieu Chavent. Molpathfinder: interactive multi-dimensional path filtering of molecular dynamics simulation data. In *The Computer Graphics and Visual Computing (CGVC) Conference 2016*, 2016.

[3] Raluca Mihaela Andrei, Marco Callieri, Maria Francesca Zini, Tiziana Loni, Giuseppe Maraziti, Mike Chen Pan, and Monica Zoppè. Intuitive representation of surface properties of biomolecules using BioBlender. *BMC Bioinformatics*, 13(4):S16, 2012.

[4] Marc Baaden, Olivier Delalande, Nicolas Ferey, Samuela Pasquali, Jérôme Waldispühl, and Antoine Taly. Ten simple rules to create a serious game, illustrated with examples from structural biology. *PLoS Computational Biology*, 14(3):e1005955, 2018.

[5] Xiao-Chen Bai, Greg McMullan, and Sjors HW Scheres. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40(1):49–57, 2015.

[6] Aidin R Balo, Merry Wang, and Oliver P Ernst. Accessible virtual reality of biomolecular structural models using the Autodesk Molecule Viewer. *Nature Methods*, 14(12):1122–1123, 2017.

[7] Helen Berman, Kim Henrick, and Haruki Nakamura. Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, 10(12):980, 2003.

[8] J.B. Brown. Adaptive mining and model building of medicinal chemistry data with a multi-metric perspective. *Future Medicinal Chemistry*, 10(16):1885–1887, 2018.

[9] J.B. Brown. Classifiers and their metrics quantified. *Molecular Informatics*, 37(1-2):1700127, 2018.

[10] Aaron Bryden, George Phillips, and Michael Gleicher. Automated illustration of molecular flexibility. *IEEE Transactions on Visualization and Computer Graphics*, 18(1):132–145, 2010.

[11] Jan Byška, Mathieu Le Muzic, M Eduard Gröller, Ivan Viola, and Barbora Kozlikova. AnimoAminoMiner: Exploration of protein tunnels and their properties in molecular dynamics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):747–756, 2015.

[12] Jan Byška, Adam Jurčík, M. Eduard Gröller, Ivan Viola, and Barbora Kozlíková. MoleCollar and Tunnel Heat Map Visualizations for Conveying Spatio-Temporo-Chemical Properties Across and Along Protein Voids. *Computer Graphics Forum*, 34(3), 2015.

[13] Matthieu Chavent, Anna L Duncan, and Mark SP Sansom. Molecular dynamics simulations of membrane proteins and their interactions: From nanoscale to mesoscale. *Current Opinion in Structural Biology*, 40:8–16, October 2016.

[14] Matthieu Chavent, Bruno Lévy, and Bernard Maigret. MetaMol: High quality visualization of Molecular Skin Surface. *Journal of Molecular Graphics and Modelling*, 27(2):209–216, 2008.

[15] Matthieu Chavent, Tyler Reddy, Joseph Goose, Anna Caroline E. Dahl, John E. Stone, Bruno Jobard, and Mark S. P. Sansom. Methodologies for the analysis of instantaneous lipid diffusion in MD simulations of large membrane systems. *Faraday Discussions*, 169:455–475, 2014.

[16] James C Chung, Mark R Harris, Fredrick P Brooks, Henry Fuchs, Michael T Kelley, John Hughes, Ming Ouh-Young, Clement Cheung, Richard L Holloway, and Michael Pique. Exploring virtual worlds with head-mounted displays. In *Proceedings of the Non-Holographic Three-Dimensional Visualization and Display Technologies*, volume 1083 of *Proceedings of SPIE*, pages 15–20, Bellingham, Wash., 1989.

[17] Gregory Cipriano and Michael Gleicher. Molecular surface abstraction. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1608–1615, 2007.

[18] Anna Caroline E Dahl, Matthieu Chavent, and Mark S P Sansom. Bendix: Intuitive helix geometry analysis and abstraction. *Bioinformatics*, 28(16):2193–2194, August 2012.

[19] Sébastien Doutreligne, Tristan Cragnolini, Samuela Pasquali, Philippe Derreumaux, and Marc Baaden. UnityMol: Interactive scientific visualization for integrative biology. In *2014 IEEE 4th Symposium on Large Data Analysis and Visualization (LDAV)*, pages 109–110. IEEE, 2014.

[20] David Duran, Pedro Hermosilla, Timo Ropinski, Barbora Kozlíková, Álvar Vinacua, and Pere-Pau Vázquez. Visualization of large molecular trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 2018.

[21] Charles Eames and Ray Eames. Powers of ten. https://youtu.be/0fKBhvDjuy0, 1977.

[22] Doug Engelbart. A research center for augmenting human intellect (a.k.a. 'The Mother of All Demos'). https://youtu.be/M5PgQS3ZBWA, 1968.

[23] Louis Farrugia. WinGX and ORTEP for Windows: An update. *Journal of Applied Crystallography*, 45(4):849–854, 2012.

[24] Katarína Furmanová, Miroslava Jarešová, Jan Byška, Adam Jurčík, Július Parulek, Helwig Hauser, and Barbora Kozlíková. Interactive exploration of ligand transportation through protein tunnels. *BMC Bioinformatics*, 18(2):22, 2017.

[25] Adam Gardner, Ludovic Autin, Brett Barbaro, Arthur J Olson, and David S Goodsell. Cellpaint: Interactive illustration of dynamic mesoscale cellular environments. *IEEE computer graphics and applications*, 38(6):51–66, 2018.

[26] Samik Ghosh, Yukiko Matsuoka, Yoshiyuki Asai, Kun-Yi Hsin, and Hiroaki Kitano. Software for systems biology: From tools to integrated platforms. *Nature Reviews Genetics*, 12(12):821, 2011.

[27] Alexandre Gillet, Michel Sanner, Daniel Stoffler, David Goodsell, and Arthur Olson. Augmented reality with tangible auto-fabricated models for molecular biology applications. In *IEEE Visualization*, pages 235–241. IEEE, 2004.

[28] Alexandre Gillet, Michel Sanner, Daniel Stoffler, and Arthur Olson. Tangible interfaces for structural molecular biology. *Structure*, 13(3):483–491, 2005.

[29] Christopher Healey and James Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2011.

[30] J. Heinrich, J. Vuong, C. J. Hammang, A. Wu, M. Rittenbruch, J. Hogan, M. Brereton, and Seán I. O'Donoghue. Evaluating viewpoint entropy for ribbon representation of protein structure. *Computer Graphics Forum*, 35(3):181–190, 2016.

[31] W. Humphrey, A. Dalke, and K. Schulten. Vmd: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

[32] Jameed Hussain and Ceara Rea. Computationally efficient algorithm to identify matched molecular pairs (mmps) in large data sets. *Journal of Chemical Information and Modeling*, 50(3):339–348, 2010.

[33] Wonpil Im, Jie Liang, Arthur Olson, Huan-Xiang Zhou, Sandor Vajda, and Ilya A Vakser. Challenges in structural approaches to cell modeling. *Journal of Molecular Biology*, 2016.

[34] Graham T Johnson, Ludovic Autin, Mostafa Al-Alusi, David S Goodsell, Michel F Sanner, and Arthur J Olson. cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nature Methods*, 12(1):85–91, 2015.

[35] John C Kendrew, Richard E Dickerson, Bror E Strandberg, Roger G Hart, David R Davies, David C Phillips, and VC Shore. Structure of myoglobin: A three-dimensional fourier synthesis at 2 Å resolution. *Nature*, 185(4711):422, 1960.

[36] Kristin Koch, Judith McLean, Ronen Segev, Michael A Freed, Michael J Berry II, Vijay Balasubramanian, and Peter Sterling. How much the eye tells the brain. *Current Biology*, 16(14):1428–1434, 2006.

[37] Robert Kosara, Silvia Miksch, and Helwig Hauser. Semantic depth of field. In *Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, page 97. IEEE, 2001.

[38] Atsushi Koyama, Nobuyaki Yasuo, and Masakazu Sekijima. HoloMol: Mixed reality-based protein and ligand tertiary structure visualization system. https://www.youtube.com/watch?v=NipdOcna76Y, 2018.

[39] B. Kozlíková, M. Krone, M. Falk, N. Lindow, M. Baaden, D. Baum, I. Viola, J. Parulek, and H.-C. Hege. Visualization of biomolecular structures: State of the art revisited. *Computer Graphics Forum*, 36(8):178–204, 2017.

[40] Julian Kreiser, Pedro Hermosilla, and Timo Ropinski. Void space surfaces to convey depth in vessel visualizations. *arXiv e-prints*, page arXiv:1806.07729, Jun 2018.

[41] Michael Krone, Florian Friess, Katrin Scharnowski, Guido Reina, Silvia Fademrecht, Tobias Kulschewski, Jürgen Pleiss, and Thomas Ertl. Molecular surface maps. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):701–710, 2016.

[42] Michael Krone, Barbora Kozlíková, Norbert Lindow, Marc Baaden, Daniel Baum, Julius Parulek, H-C Hege, and Ivan Viola. Visual analysis of biomolecular cavities: State of the art. In *Computer Graphics Forum*, volume 35, pages 527–551. Wiley Online Library, 2016.

[43] Benoist Laurent, Matthieu Chavent, Tristan Cragnolini, Anna Caroline E Dahl, Samuela Pasquali, Philippe Derreumaux, Mark S P Sansom, and Marc Baaden. Epock: rapid analysis of protein pocket dynamics. *Bioinformatics*, December 2014.

[44] Mathieu Le Muzic, Ludovic Autin, Julius Parulek, and Ivan Viola. cellVIEW: A tool for illustrative and multi-scale rendering of large biomolecular datasets. In *Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine*, VCBM '15, pages 61–70, Aire-la-Ville, Switzerland, Switzerland, 2015. Eurographics Association.

[45] Tony E Lewis, Ian Sillitoe, Antonina Andreeva, Tom L Blundell, Daniel WA Buchan, Cyrus Chothia, Alison Cuff, Jose M Dana, Ioannis Filippis, Julian Gough, Lawrence A. Kelley David T. Jones, Gerard J. Kleywegt, Federico Minneci, Jaina Mistry, Alexey G. Murzin, Bernardo Ochoa-Montaño, Matt E. Oates, Marco Punta, Owen J. L. Rackham, Jonathan Stahlhacke, Sameer Velankar Michael J.E. Sternberg, and Christine Orengo. Genome3D: A UK collaborative project to annotate genomic sequences with predicted 3d structures based on SCOP and CATH domains. *Nucleic Acids Research*, 41(D1):D499–D507, 2012.

[46] Zhihan Lv, Alex Tek, Franck Da Silva, Charly Empereur-Mot, Matthieu Chavent, and Marc Baaden. Game on, science-how video game technology may help biologists tackle visualization challenges. *PLOS One*, 8(3):e57990, 2013.

[47] Niloy J. Mitra, Yong-Liang Yang, Dong-Ming Yan, Wilmot Li, and Maneesh Agrawala. Illustrating how mechanical assemblies work. *ACM Transactions on Graphics*, 29(4):58:1–58:12, July 2010.

[48] A-E Molza, Nicolas Férey, Mirjam Czjzek, Elisabeth Le Rumeur, J-F Hubert, Alex Tek, B Laurent, Marc Baaden, and Olivier Delalande. Innovative interactive flexible docking method for multi-scale reconstruction elucidates dystrophin molecular assembly. *Faraday discussions*, 169:45–62, 2014.

[49] Finian Mwalongo, Michael Krone, Guido Reina, and Thomas Ertl. State-of-the-art report in web-based visualization. *Computer Graphics Forum*, 35(3):553–575, 2016.

[50] Michael Krone Robert S. Laramee Naif Alharbi, Matthieu Chavent. Vapli: Novel visual abstraction for protein-lipid interactions. In *IEEE VIS 2018 – SciVis Short Papers*, 2018.

[51] Seán I. O'Donoghue, Benedetta Frida Baldi, Susan J. Clark, Aaron E. Darling, James M. Hogan, Sandeep Kaur, Lena Maier-Hein, Davis J. McCarthy, William J. Moore, Esther Stenau, Jason R. Swedlow, Jenny Vuong, and James B. Procter. Visualization of biomedical data. *Annual Review of Biomedical Data Science*, 1:275–304, 2018.

[52] Seán I. O'Donoghue, David S. Goodsell, Achilleas S. Frangakis, Fabrice Jossinet, Roman Laskowski, Michael Nilges, Helen R. Saibil, Andrea Schafferhans, Rebecca Wade, Eric Westhof, and Arthur J. Olson. Visualization of macromolecular structures. *Nature Methods*, 7(3s):S42–S55, 2010.

[53] Seán I. O'Donoghue, Kenneth S. Sabir, Maria Kalemanov, C. Stolte, Benjamin Wellmann, Vivian Ho, Manfred Roos, Nelson Perdigão, Fabian Buske, Julian Heinrich, B. Rost, and Andrea Schafferhans. Aquaria: Simplifying discovery and insight from protein structures. *Nature Methods*, 12:98–99, 2015.

[54] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF Chimera — a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[55] Daniel Reker, Petra Schneider, Gisbert Schneider, and J.B. Brown. Active learning for computational chemogenomics. *Future Medicinal Chemistry*, 9(4):381–402, 2017.

[56] Alex Rose and Peter Rose. 3D visualization of large virus structures using NGL viewer powered by MMTF. https://youtu.be/A8MxTRP5t4c, 2017.

[57] Alexander S Rose, Anthony R Bradley, Yana Valasatava, Jose M Duarte, Andreas Prlić, and Peter W Rose. NGL viewer: Web-based molecular graphics for large complexes. *Bioinformatics*, 34(21):3755–3758, 2018.

[58] Steven J Rysavy, David AC Beck, and Valerie Daggett. Dynameomics: Data-driven methods and models for utilizing large-scale protein structure repositories for improving fragment-based loop prediction. *Protein Science*, 23(11):1584–1595, 2014.

[59] Kenneth S. Sabir, Christian Stolte, Bruce Tabor, and Seán I. O'Donoghue. The Molecular Control Toolkit: Controlling 3D molecular graphics via gesture and voice. In *IEEE Symposium on Biological Data Visualization (BioVis)*, pages 49–56. IEEE, 2013.

[60] Andrea Schafferhans, Seán I. O'Donoghue, and Haruki Nakamura, editors. *Web-based molecular graphics*, NII Shonan Meeting Report No. 2016-12. National Institute of Informatics, Japan, 2016.

[61] Gisbert Schneider. Automating drug discovery. *Nature Reviews Drug Discovery*, 17:97–113, 2018.

[62] David Sehnal, Mandar Deshpande, Radka Svobodová Vařeková, Saqib Mir, Karel Berka, Adam Midlik, Lukáš Pravda, Sameer Velankar, and Jaroslav Koča. LiteMol suite: Interactive web-based visualization of large-scale macromolecular structure data. *Nature Methods*, 14(12):1121, 2017.

[63] David Sehnal, Alexander Rose, Jaroslav Koca, Stephen Burley, and Sameer Velankar. Mol*: Towards a Common Library and Tools for Web Molecular Graphics. In Jan Byska, Michael Krone, and Björn Sommer, editors, *Workshop on Molecular Graphics and Visual Analysis of Molecular Data*. The Eurographics Association, 2018.

[64] Björn Sommer, Marc Baaden, Michael Krone, and Andrew Woods. From virtual reality to immersive analytics in bioinformatics. *Journal of integrative bioinformatics*, 15(2), 2018.

[65] Fabio Sterpone, Sébastien Doutreligne, Thanh Thuy Tran, Simone Melchionna, Marc Baaden, Phuong H Nguyen, and Philippe Derreumaux. Multi-scale simulations of biological systems using the opep coarse-grained model. *Biochemical and biophysical research communications*, 498(2):296–304, 2018.

[66] Antoine Taly, Francesco Nitti, Marc Baaden, and S Pasquali. Molecular modelling as the spark for active learning approaches for interdisciplinary biology teaching. *Interface Focus*, 9(3):20180065, 2019.

[67] Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, 2nd edition, 2009.

[68] Viktor Vad, Jan Byška, Adam Jurcík, Ivan Viola, Eduard Gröller, Helwig Hauser, Sérgio M. Marques, Jiří Damborský, and Barbora Kozlíková. Watergate: Visual exploration of water trajectories in protein dynamics. In Stefan Bruckner, Anja Hennemuth, Bernhard Kainz, Ingrid Hotz, Dorit Merhof, and Christian Rieder, editors, *Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2017.

[69] Anne Mai Wassermann and Jürgen Bajorath. Chemical substitutions that introduce activity cliffs across different compound classes and biological targets. *Journal of Chemical Information and Modeling*, 50(7):1248–1256, 2010.

[70] James D. Watson and Francis H. C. Crick. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738, 1953.

[71] Maria Francesca Zini, Yuri Porozov, Raluca Mihaela Andrei, Tiziana Loni, Claudia Caudai, and Monica Zoppè. BioBlender: Fast and efficient all atom morphing of proteins using Blender game engine. *arXiv e-prints*, page arXiv:1009.4801, 2010.

[72] Monica Zoppè. For a perceptive understanding of size in cellular biology. *Nature Methods*, 14(7):662–665, 2017.

[73] Monica Zoppè and Tiziana Loni. The representation of electrostatics for biological molecules. In Walter Rocchia and Michela Spagnuolo, editors, *Computational Electrostatics for Biological Applications: Geometric and Numerical Approaches to the Description of Electrostatic Interaction Between Macromolecules*, pages 215–225. Springer International Publishing, Cham, 2015.