NII Shonan Meeting Report

No. 2016-12

Web-based Molecular Graphics

Andrea Schafferhans
Seán O'Donoghue
Haruki Nakamura

September 05–08, 2016

Web-based Molecular Graphics

Organizers:
Andrea Schafferhans (TUM & HSWT, Germany)
Seán O'Donoghue (CSIRO & Garvan Institute, Australia)
Haruki Nakamura (Osaka University, Japan)

September 05–08, 2016

1 Overview of the meeting

1.1 Summary

Molecular graphics is a well established discipline used in many scientific fields (e.g., life sciences, biomedical research, chemistry, material sciences), including commercial research (e.g., pharmaceutical industry). A wealth of methods and tools have been developed, and are still actively developed – however until recently activity has focused on stand-alone software. This meeting will explore an emerging new frontier for molecular graphics, namely deployment in a server-client environment, which promises to make molecular structural information much more easily accessible to scientists. In addition, the recent hybrid approaches have revealed structures, functions and dynamics of very large molecules and cellular machineries, which are displayed as both atomic structures and molecular/cellular images, greatly extending the roles of molecular graphics. This meeting has brought together key players in this emerging field to explore the challenges and opportunities raised by web-based molecular graphics and new applications.

1.2 Background

There is a rapidly growing wealth of molecular structure data that has been derived from experimental methods (e.g., X-ray crystallography, NMR spectroscopy, electron microscopy) and are freely provided from the international database PDB, Protein Data Bank, as well as computational methods, such as molecular dynamics. These data provide scientists today with unprecedented insight into a wide range of molecular phenomena, for example protein function, biological membrane transport, or material properties. The amount of data is steadily growing in size as well as in complexity: more protein structures, larger protein structures (big complexes) and time-resolved data, such as molecular dynamics trajectories. In the near future, these data may increase in scale to include, for example, complex cellular machineries and whole living cells. A key challenge is to make all of these data accessible to scientists, in ways that enable interpretation and the generation of new research insights. Fortunately, nowadays almost all scientists – even students – have easy access to computer

graphics capabilities powerful enough to show a wide variety of molecular structures. With modern bandwidth speeds, it is even feasible to stream complex, interactive molecular graphics into simple devices such as an iPad or iPhone, via GPU-rendering on a remote server. Thus, it is now possible to depict and explore 3D models in regular browsers, such as Mozilla Firefox, Internet Explorer or Google Chrome. This allows to bring 3D molecular visualisation tools to non-expert end users without the need to install specialised software. However, in times of a steadily growing amount of data, improved visualization methods and tools are urgently needed, which can make the more complex data accessible and interpretable.

1.3   Aims of the meeting

We believe that web molecular graphics is likely to soon become very widely adopted. Therefore, we feel it was very timely to host a Shonan meeting on this topic. Overall, the seminar aimed to help give this emerging field direction and clarity, by bringing together global efforts in making molecular structure information available to scientists via the web. We discussed strategies for effectively managing very large molecular structures (e.g., millions of atoms or more). Most current browsers fail to visualize these larger molecular structures for various reasons. In some cases (e.g., whole organelles, entire genomes, or materials, such as metal-organic frameworks), the molecular assembly becomes so large that it requires the use of multi-scale, often hierarchical representations that change automatically at different scales, somewhat analogously to Google Maps. Similar strategies can also be needed for large molecular dynamic simulations. All of these cases are a particular challenge for web molecular graphics, which is limited by network bandwidth. In addition, many of the scientific problems being investigated cannot be solved with molecular graphics alone. Instead, data on molecular structures needs to be exchanged with other resources, for example resources that integrate data derived from structures with other information, and display the result using network graphs, or using other visualization paradigms. Therefore, an approach is required which is compatible to different modeling environments. Nowadays, the most widely used scripting language is JavaScript which is moreover used by all recent web browsers. This language seems best suited to be used to interactively combine the website's content with the molecular 3D visualisation via WebGL. The seminar will discussed pros and cons of this architecture and defined desirable interaction paradigms and communication methods.

2   Meeting Schedule

Check-in Day: September 4th (Sun)

- Welcome Banquet

Day1: September 5th (Mon)

- Lightning talks
- Keynote: Peter Rose
- Agreement on breakout topics
- Group photo shooting
- Breakout sessions
- Keynote: Mike Goodstadt, Jim Zheng
- Breakout check-in and round up

Day2: September 6th (Tue)

- Keynote: Haruki Nakamura
- Keynote: Sameer Velankar
- Tech talk: Sean O'Donoghue, Bosco Ho
- Breakout session
- Breakout summary
- Agreement on further breakout topics
- Breakout session

Day3: September 7th (Wed)

- Remote Keynote: Graham Johnson
- Tech talk: Ian Silitoe
- Breakout session
- Excursion and Main Banquet

Day4: September 8th (Thu)

- Breakout session
- Breakout summary
- Wrap up

3 Keynote and Technical Talks

Compressive Structural Bioinformatics: High Efficiency 3D Structure Compression

Peter Rose, Peter Rose, RCSB Protein Data Bank, San Diego Supercomputer Center, University of California, San Diego, USA

Technologies in structural biology continue to improve and help determine ever-larger 3D macromolecular complexes. Interactive visualization of large complex structures and large scale queries or structural comparisons across the entire Protein Data Bank (PDB) archive are becoming a bottleneck in terms of network bandwidth, I/O, parsing, and memory consumption. We have developed a compact and extensible file format (MacroMolecular Transmission Format, MMTF, http://mmtf.rcsb.org) for 3D molecules to overcome these challenges. This compact representation enables efficient data transfer and parsing for interactive visualization on the web that is one to two orders of magnitudes fast than using standard PDB or mmCIF file formats. MMTF files also contain bonds and bond orders and consistently calculated secondary structure information (DSSP). Both the MMTF specification and encoders and decoders are available open source in GitHub (https://github.com/rcsb/mmtf). MMTF Decoders are available in Java, JavaScript, Python, and a C/C++ version is under development. Early adopters of MMTF include NGL Viewer, Jmol/JSmol, 3Dmol.js, iCn3D, ICM-Browser, PyMol, BioJava, and Biopython.

Visualizing Human Genome in Time and Space

Jim Zheng, University of Texas Health Science Center at Houston, USA

The high order nuclear organization of mammalian genome plays significant roles in important cellular functions such as gene regulation and cell state determination. The influx of new details about the higher-level structure and dynamics of the genome from Hi-C and CHIA-PET technology requires new techniques to model, visualize and analyze the full extent of genomic information in three dimensions. While existing genome browsers have been proven as successful genome information management and visualization tools, these browsers are based on two-dimensional visual interface with limited capacity to represent structural hierarchies and long-range chromosome interactions, particularly across non-contiguous genomic segments.

We created the first model-view framework of eukaryotic genomes, Genome3D (http://genome3d.org), to enable integration and visualization of genomic and epigenomic data in a three-dimensional space. Our physical genome model implicitly contains all levels of structure and hierarchy, and provides an underlying platform for integrating multi-scale genomic information within three dimensions. Our viewer uses a hierarchical model of the relative positions of all nucleotide atoms in the cell nucleus.

We further developed a game engine based Genome3D browser that has better performance, is platform independent and can be configured to allow users to access and visualize 3D genome models on a remote server. In this works, we integrated various genomic databases with Genome3D software, providing

a multi-scale genome information visualization system to explore and navigate eukaryotic genome in 3-dimension.

The new system, iGenome3D, provides a wide spectrum of tools, ranging from model construction to spatial analysis, to decipher the relationships between 3D conformation of the genome and its functional implication. The incorporation of literature allows users to quickly identify key features from PubMed abstracts for genes in the displayed 3D genome structure. The seamless integration of UCSC Genome Browser allows genetic and epigenetic features from the 2D browser to be visualized in 3D genome structure. iGenome3D can also output 3D genome model in various forms, including one that allows these models to be explored in modern virtual reality environments such as Oculus. Eukaryotic genomes can be analyzed from a completely new angle in iGenome3D that enables researchers to make new discoveries from a truly multi-scale exploration

Visualizing Genomes in Three Dimensions

Mike Goodstadt, CNAG (Centro Nacional de Anàlisis Genòmica), Spain

Genomic data is encoded and represented in linear sequence. Yet when contained within nucleus this 2m long strand is not merely efficiently packed but it is organized in a non-random structure. This spatial conformation has been shown to serve as a key part of gene expression and regulation of cell function. Recent research also reveals that this is dynamic in nature.

This structure has been resolved by modeling at a molecular level and by microscopy at cellular scale. However it has yet to be determined at an intermediate resolution. To fill this gap in knowledge, researchers combine distinct techniques (light microscopy (FISH) and 3C-based experiments) to detect interactions across the genome which indicate proximity. The resulting contact matrices can be computationally analyzed to build ensembles of probable conformations.

However there are still significant challenges given the increasing scale of computational demands (calculation, storage, manipulation), the need for cohesive and coordinated research, and the aim for an integrated and reproducible model of the whole genome. Web-based molecular graphics can assist in tackling key aspects of these. The multiple scales of structure, organization and function need discrete visual description and inter-connection.

Yet diverse sources and types of data from which the graphics are derived need to be typified and standard methods of integration agreed. Therefore identification of shared properties is important, and development of a new grammar will help optimize visualization of structure and dynamics. There are also more complex unresolved requirements, for example the display and navigation of multiple genomes, networks, data confidence, or phylogenetics.

These new representations must be developed inline with expertise and best-practice in visualization and with awareness of the obfuscation inherent in 3D images (see Tamara Munzner. Visualization Analysis and Design. A K Peters *Visualization Series, CRC Press*, 2014). Existing popular genome browsers do not comfortably integrate interaction or 3D data if at all. However there are a number of new browsers based on these types of data which demonstrate various form of achieving such an comprehensive genomics tool: Juicebox, 3DBG, Foldsynth, Genome3D, and our own TADkit.

Activities of wwPDB and PDBj towards multiscale structural biology

Haruki Nakamura & Gert-Jan Bekker, Osaka University, Japan

In recent years, the structures of very large macromolecular machines in cells have been determined by combining observations from multiple, complementary experimental methods, such as X-ray crystallography, NMR spectroscopy, cryo-EM, small-angle scattering (SAS), FRET, crosslinking, and many others. Thus, so called "Hybrid method" is going to provide us the biological systems in the multiscale manner. Currently, many large structures (421 entries on 31 August 2016) determined by such hybrid methods appear in high-impact-factor journals, and are being deposited in the PDB (Protein Data Bank), which is managed by an international organization, the wwPDB (worldwide PDB: http://wwpdb.org/)). We will introduce the activities of the wwPDB and PDBj to view those large structures, in particular using the web-based molecular graphics, Molmil (Bekker et al. *J Cheminform* 8:42, 2016 DOI 10.1186/s13321-016-0155-1), and will discuss several issues to be overcome.

From data to knowledge – exploiting macromolecular structure data

Sameer Velankar, EMBL-EBI, UK

The recent advances in structure determination technology have led to researchers being able to tackle large and complex problems at cellular and molecular level. As a result the Protein Data Bank (PDB) and Electron Microscopy Data Bank (EMDB) have increasing number of depositions of large macromolecular complexes and it is expected that this trend will continue in the future. The data archives and structural bioinformatics community, as a result, has the challenging task of delivering these large data sets to the user community for data mining, data integration efforts and/or to the browser based user interfaces for interactive visualisation. Providing seamless access to datasets at cellular level and being able to investigate the molecules represented in those data sets at atomic resolution is a difficult task both in terms of user interface design and delivery of data for interactive visualisation. At PDBe we have addressed the data delivery challenge by implementing a server (https://www.ebi.ac.uk/pdbe/coordinates/) that allows users to dynamically select the subset of data they are interested in to ensure fast and interactive delivery of molecular structure information. A prototype where in addition to dynamic selection, compression of data to reduce the size of data packet is also developed and is undergoing testing. The coordinate server makes these data available in mmCIF format, a stadard developed for representation of macromolecular structure data. Additionally we have developed REST API (http://www.ebi.ac.uk/pdbe/api/doc/) and web components (http://www.ebi.ac.uk/pdbe/pdb-component-library/) to display the biological context for the macromolecular structures. We have also developed LiteMol - a WebGL based, interactive 3D viewer which is available of the PDBe pages (e.g. pdbe.org/1cbs) and is able to show both macromolecular structure data

and associated experimental results (Electron density maps for both diffraction experiments and Electron microscopy structures).

Aquaria – Insight from protein structures

Seán O'Donoghue, CSIRO & Garvan Institute, Australia

Aquaria (http://aquaria.ws; O'Donoghue et al., Nature Methods 2015) is a new web resource that streamlines and simplifies the generation of insight from protein structures. It has been designed to make structural information more accessible to life scientists who find existing resources too time-consuming to use. The Aquaria interface is organized primarily by protein sequence, not structure. For any protein of interest, Aquaria generates a concise visual summary of all related PDB structures. Aquaria also allows mapping of domains, single-nucleotide polymorphisms, post-translational modifications and other sequence features onto 3D structures. Users can also define and map their own sequence features. Aquaria uses a minimal set of mouse-based controls that are intuitive yet powerful. Aquaria can also be controlled via hand gestures using the Leap Motion. Currently, Aquaria contains over 100 million pre-calculated sequence-to-structure alignments, resulting in at least one matching structure for 87% of SwissProt proteins and a median of 35 structures per protein; this provides a depth of sequence-to-structure information currently not available from other resources. Recently, we used this database of structural alignments to map the 'dark proteome', i.e., the regions of proteins where molecular conformation is completely unknown (Perdigao et al., PNAS 2015).

Jolecule – protein viewer in the cloud

Bosco Ho, Monash University, Australia

Introduction to Jolecule (http://jolecule.com), a web-based viewer that focuses on annotations. It focuses on user interactions for structural biology, such as a arrow embellishment to indicate chain direction, and depiction of the backbone peptide bond plane. There are several ways to find residues, and it's easy to add atom labels and distance measurements. The most distinguishing aspect of Jolecule is the ability to store annotations: specific views of a protein structure with explanatory text and labels. This is stored on a central server and is easily available on any device.

Using CATH to predict the function of novel protein structures

Ian Silitoe, UCL, UK

CATH is a classification of protein structures: it groups together protein structural domains that are related by evolution. This body of work (built over 20 years) has been used to predict the location of structural domains on over 50 million known protein sequences. When combined with functional annotations (such as UniProtKB, Gene Ontology, Enzyme Commission, etc), this information can be used to recognise sequence patterns (HMMs) observed during protein

evolution. In turn, these evolutionary 'fingerprints' can be used to provide predictions of the structure and even function for new protein sequences. Allowing scientists to interrogate these sequence, structure and functional annotations in a clear and meaningful way provides a number of interesting visualisation challenges.

SSBD: a database of quantitative data of spatiotemporal dynamics of biological phenomena

Shuichi Onami, RIKEN, Japan

Recent advances in live-cell imaging analysis and mathematical modeling have produced a large amount of quantitative data on spatiotemporal dynamics of biological objects ranging from molecules to organisms. In this talk, Shuichi discussed how these data are revolutionizing biological researches by showing his group's works on a large collection of quantitative data on nuclear division dynamics in early C. elegans embryos. Computational phenotype analysis and inference of causal relationship of phenotype expressions were discussed in addition to the power of interactive 3D visualization. Shuichi then gave an overview of the SSBD database, which his group launched for storing and sharing quantitative biological dynamics data. Current technologies and problems of web-based visualization of biological objects ranging from single molecules to organisms were discussed Finally, Shuichi discussed future plans of the SSBD database and possible collaborations with the web-based molecular graphics research activities.

4 Breakout group discussions

4.1 Augmented Reality and Virtual Reality

This session investigated the current state of augmented reality (AR) and virtual reality (VR) in web molecular graphics. We considered challenges, opportunities, use cases, and perspectives of these technologies for education, outreach, and research.

Recent capabilities and advancements in AR/VR hardware and experiences have significantly revolutionized gaming and entertainment industries. Within the past year, we believe this technology has reached several 'tipping points', evident in the worldwide craze in applications such as Pokemon Go and Snapchat. Although AR/VR technologies have been available for decades, the technology has often been labelled as 'gimmicky', or one-off experiences with high wow-factor. With the introduction of affordable and effective low-end (Google Cardboard, mobile AR) and high-end hardware (Oculus, HTV Vive, Hololens), we discussed the importance of experience creation, delivery, and the power of content-driven adoption. Authoring and designing experiences for molecular data include use cases in remote collaboration, education, outreach, and research.

In an effort to accelerate the 'tipping point' for web molecular graphics, we present a review of available web AR/VR authoring tools, successful content to

date, and available hardware[1].

4.2 Building Mash-Ups – Interoperability Requirements

While a plethora of excellent web-based molecule viewers exist, it is still a challenge to allow a user to dynamically annotate the viewer with additional annotations that are relevant to the molecule viewing task at hand and to the user's field of expertise. The Web Molecular Graphics 2016 Shonan meeting "mash-up" subgroup has proposed a preliminary conceptual outline and logical flow for establishment of a standard by which web-based molecular viewers can dynamically enrich a molecule display with relevant annotations.

We considered the following motivation example: Next-generation sequencing (NGS) technology has matured to a stage in which it will be mass deployed in clinics and hospitals around the world. Two common NGS results are genotyping for genetic variants in DNA (DNAseq) and expression of genes under a specific physiological condition (RNAseq). A clinician, researcher, or patient could benefit from a web-based experience in which the variants detected are presented from several views:

- A view of variants as mapped onto a view of the collection of chromosomes (chromosome global view).

- A highlighted view of the chromosome and gene region containing a specific variant (chromosome global-local view).

- A view of which organs in the body express the gene (organism view).

- A view of drugs available that specifically target the variant gene being viewed (pharmacogenomic action view).

- A gene subset view of patient gene expression in tissue samples analyzed (expression view) and comparison to case/control tissue (expression comparison view).

- An alignment of standard gene protein/nucleotide sequence versus patient sequence (sequence comparison view).

A key challenge is how to present this information to a clinician in a format that is compact enough that a clinical treatment decision can be made in minutes, as well as how to allow interaction with one view to update the remaining views.

Using this example, we discussed the required technical building blocks to build such a mash-up. We envisaged a three-tier architecture:

1. The *Data layer* uses REST services to provide the required dat

2. The *Mashup layer* retrieves data from the data layer through the respective APIs and provides components for coordinate mapping and communication (e.g. about events) between different components. Coordinate systems to be mapped are genomic, protein and structure coordinates.

---

[1]https://docs.google.com/document/d/1Zf0kB4amYvnSkEAo2bTBSpMKjpvgcYXrqMkNtky
MKq0

3. The *Visualization layer* uses BioJS or specialised Custom Widgets to display the information and communicates with the Mashup layer through an Event API.

In order to encourage cooperation between different components we recommend the establishment of a registry for the Mashup layer. Such a registry should not only give a detailed (ontology-enhanced) description of the service, but also provide means to ensure that the component is valid and responsive at the time of query.

4.3 Mesoscale and Intermediate Representations

This group looked into the challenge of representing mesoscopic biological structures. Mesoscopic structures are biological objects which are too large to be resolved by methods used to determine protein structures, but also too small to be resolved by standard microscopy techniques. These can include large protein complexes and small organelles. Objects in this range present a general problem for visualisation because there are few datasets available, and there is also a lack of visualisation paradigms and solutions. Examples of mesoscopic structures can include viruses, nuclear pores, ribosomes, focal adhesion complexes, etc.

Mesoscopic visualisation was of interest to the group for a variety of reasons. This includes -

- Whole cell visualisation

- Combining structure views with pathways

- Combining mesoscopic and molecular visualizations/Viewing protein interactions in biological contexts

- Drug development

- Representations for audiences using storytelling

- Moving beyond molecular views

- Maximising use from electron microscopy data

This group discussed the current state of mesoscopic structure visualisation paradigms, and addressed the challenges for visualising these objects. The main challenges we addressed can be summarised as follows:

1. Mesoscopic structures can occupy a vast range of physical sizes - this can present a significant computational challenge since mesoscopic assemblies could require very large numbers of data points to be represented, which may place computational constraints on the possibilities for visualisation.

2. Mesoscopic data can be very heterogeneous - coming from a variety of experiment types and representing a range of structure types. This may present technical challenges for visualisation.

3. Interactive, real time visualisation are highly desirable for mesoscopic visualisation, to allow exploration of complex datasets. This interactivity also presents specific challenges.

The discussions investigated strategies to tackle these challenges. The primary focus was on data formats that could accommodate interactive visualisations of large scale biological data sets. Some key themes emerged from these discussions:

- Lossy data formats are probably acceptable for mesoscopic visualization, as they prioritize scale, interactivity and dynamic range over accuracy.

- Formats which contain a range of representations for different zoom levels are desirable, as they provide on demand detail when zoomed in on small structures, but should allow rapid transitions to zoomed out views containing large numbers of objects.

Some basic file reduction procedures are of interest to make mesoscopic visualisation more practical.

1. Representations which feature only backbone trace or 1 or 2 point residue use can still represent overall protein structures while reducing file sizes, but sacrificing some molecular detail

2. Gaussian mixture models may be particularly interesting to vastly reduce the complexity of displayed objects, while still representing the overall shapes of objects[2].

3. Creating surfaces and then reducing the surface count using decimation procedures may be necessary for surface style representations Volumetrics could be interesting, but may be more computationally expensive. Perhaps looking at representations of astronomical datasets could help inform this sort of practice.

4.4 3D Rendering and Representation of Molecular Structures

In our breakout group about rendering and representing molecular structures we identified a set of challenges common to most molecular graphics tools. Related to providing interactive rendering are: perception of smooth interaction; leveraging increased parallelism of modern computers; providing scalable performance given highly variable resources on desktop and mobile machines. Scalability becomes an issue with a growing number of objects to be rendered and when data is dynamic in nature. In terms of visual quality, transparency, anti-aliasing, shadows, and ambient occlusion are important factors. The web lacks reusable (standard) implementations for many common algorithms used in molecular graphics, including electron density handling, structure/sequence alignment and generally calculation of derived features such as molecular surfaces and volumes. Apart from those more technical challenges, there are many open question regarding representing variability and ambiguity: how to show comparisons, highlight differences; how make the user aware of incomplete data and uncertainty therein; how to make alternative ensemble and dynamic data accessible. Another question is how to efficiently annotate 3D structures with data from other sources to show spatial relationships.

---

[2]See http://www.sciencedirect.com/science/article/pii/S0006349508786041

We discussed potential solutions or at least approaches to/for some of those challenges. If data gets too large to download or render all at once one can move to a client-server architecture. This can for instance allow access to large molecular dynamics trajectories by only loading and rendering parts the user wants see at the same time. With respect to representing data in appropriate ways we found that there are a number of non-standard representations that are very useful but not widely available. Those include Bendix, Hyperballs, Molecular Surface Maps, 2D Maps of Channels, Abstractions for Sugars and Lipids, Volume Rendering, Directional Arrows on Backbone Atoms (e.g., showing the direction of the amino acid chain).

We find that there is a large gap between computer graphics research and its application in life sciences. This is especially true with respect to the technical rendering challenges, but also innovative representations. In the world of desktop applications it took 10 years and more for techniques developed in computer graphics research to find their way into tools used by life scientist. For the web, we don't want to wait that long. Therefore we are proposing to create Modular Molecular Graphics (MMG.js), a low level library for modular molecular graphics. The idea for that library is that it would provide a common platform for joint development and collaborations between computer graphics researchers and application developers from life sciences to bring state of the art rendering to the user. Such a project could evolve around a github project and raise awareness through a homepage, mentioning in articles and posters as well as through personal invitations of potential collaborators from both fields.

4.5 User Experience Best Practise

Over the past decades, molecular graphics systems have been developed to address a wide variety of user scenarios, and using a diversity of languages, operating systems, and user interface paradigms. The emerging web molecular graphics systems are likely to be used for an even more diverse range of applications, many not possible previously (e.g., multiscale interactive navigation from molecular to cellular or even tissue scales).

However, in contrast to previous work, web molecular graphics will have a common language (JavaScript) and platform (browser) – and are therefore more likely to share commonalities in the user experience (UX) than previous. There is therefore an opportunity for de facto standards in user experience to emerge, blending existing web standards and the best of breed pre-existing conventions from existing stand-alone browsers. That will enable users unfamiliar with molecular graphics to be more easily 'onboarded', and to learn to use and gain insight from 3D structural information. For expert users, good de facto standards will enable them to more quickly learn new applications, thus broadening the scope of analysis tools in their arsenal. Our goal in this breakout was to explore these possibilities, with the aim of articulating a set of user experience best practises.

As a preliminary step towards a planned publication that articulates key elements of such a de facto graphical user interface, we discussed that it could include:

- A default 'minimal' mode with only the most essential GUI elements.

- Specialized modes for specific types of exploration, e.g., dynamics trajectories, or small molecule interactions.

- Enhancing orientation awareness using graphical metaphors such as, for example, the 'View Cube' from 3D Studio Max.

- Easy access to sets of pre-calculated canonical orientations.

- Use 3D controls familiar to 3D gamers, e.g., via keyboard 'numpad' controls.

- Enhancing perception of depth cues, e.g., by y-axis rocking.

- Intuitive yet sophisticated selections using familiar GUI conventions.

- Color conventions that account for color blindness, as well as for limitations in normal color perception.

Another key element of molecular graphics tools is the selection and scripting language; as a preliminary step towards a Common Molecular Modeling Scripting and Selection Language', we identified key goals and features such a language would have, including:

- Compact

- Expressive

- Easily parsable

- Encodable (URLs, filenames, etc.)

- Human readable

- Extensible to large, multi-scale

- Efficient

- Hierarchical

- Supports boolean combinations - e.g., A and (B or not C)

- Pre-defined keywords (e.g., 'protein', 'ligand', etc.) built from primitives.

- Operators

- Wildcards.

Finally, we anticipate that web molecular graphics systems will often be part of an ecosystem of 'mash-ups'; this will greatly facilitated by defining a rich set common of inter-application communication. The challenge of defining and communicating this context can be broken down:

- defining a set of predefined common scientific contexts (e.g. "small molecule chemistry")

- deciding on a common ontology

- defining a data structure that describes context (minimum amount of information, optional extras)

- allowing this context to be easily integrated with existing tools.

A number of factors need to be taking into account when communicating information about the scientific context in which a particular user is interested:

- privacy

- user may be interested in more than one context

- user may not want suggested views imposed upon them (always allow "default settings")

- the definition of "context" would need to grow (without breaking existing tools).

Expecting developers to implement and maintain all these different views into their individual resources would represent an unrealistic amount of work. However, rather than developing their own pages, we would encourage resource developers to reuse existing web components to provide views such as 3D rendering of protein structure. This way the task of understanding and communicating scientific context would be left to a small number of developers on an opt-in basis. This would represent an entirely tractable problem and could provide an extremely useful tool for the community.

4.6 Chemoinformatics

The "Chemoinformatics" breakout group discussed visualisation challenges specific to interactions between proteins and small molecules – which range from obtaining an overview of large dataset of protein-ligand interactions over visualising specific physico-chemical interactions within the protein structure to comparing different ligand poses, but also include the challenge of displaying and building chemical compounds in a 2D sketch as well as a probable 3D conformation.

As a unifying scenario that includes many of the above challenges, we proposed to build a framework that can create an interesting, educational, and interactive experience for students to understand the basic concept of a ligand successfully docking into a receptor. Students interested in clinical medicine and medical science would benefit from understanding the biophyiscal, pharmacological, and other related aspects of a small molecule binding into a large receptor. An interactive experience should allow a student to design a ligand interactively and test its spatial and electrostatic compatibility.

As a result of the discussions we developed an outline for how to build such a system as well as a plan how to fund this joint project.

We also produced a wish-list of features molecular viewers should implement in order to enhance the usability for more detailed analyses of chemical interactions. For some of these features we were able to cases we suggest possible solutions (based on Bob Hanson's implementations in Jmol), others would benefit from suggestions of visualisation experts on how to achieve the intended aim:

- Visualisation of surfaces:

  - show iso-surfaces for energy maps

  - simultaneously visualise multiple surfaces with quick on/off switch

  - see pockets and surfaces without obscuring chemical detail – *suggestion:* use separate slab for surfaces

- Visualisation of clashes – *suggestion:* colour the molecular surface by the amount of the van-der-Waals overlap

- Visualise H-bonds so they are distinguishable from covalent bonds – *suggestion:* show H-bonds as disks or small surfaces in the areas of the van-der-Waals overlap

- Visualisation of electrostatic interactions: difficult because that would require displaying electrostatic fields of the interaction partners – *suggestion:* coordinate with a 2D view (in a mash-up display) that shows electrostatics

4.7 Review of the State of the Art

Participants: Mike Goodstadt, Shuichi Onami, Björn Sommer, Jenny Vuong, Merry Wang, W. Jim Zheng

Visualization in molecular biology has played an essential role in revealing mechanisms of biological function. Our ability to visualize molecular interactions at the atomic level is astonishing. For many cases, 2D visualization will suffice, but there remain many biological systems where 3D visualization is essential: the exploration of 3D genomes, proteins, membrane patches and simulations, vesicles, metabolic reactions on the atomic scale, etc.

In the past, the dissemination and exploration of three-dimensional models was limited, as visualizations of models were prepared on specialist equipment, and then distilled into publication figures. For complex operations, specialized tools were required for the visualization of these models.

Today, we are poised to enter an era of unprecedented access to 3D visualizations of biological molecular systems. This is due to the maturation of a number of technologies: 1) commodity graphics cards, 2) JavaScript in web browsers, 3) the WebGL standard and 4) distributed data delivery. In this breakout session, the initial manuscript was discussed which focuses at the web representation of molecular graphics in the browser. The initial idea for this manuscript goes back to an initiative starting at the VizBi 2016 with Seán O'Donoghue and Bjorn Sommer, and later extended by Bosco Ho and Xavier Ho.

With the break out session group, the structure of the paper was discussed and improved. Whereas the initial manuscript was focusing only on web tools visualized with WebGL and JavaScript, it was decided to discuss also relevant tools that are still based on older technologies, such as Java applets. A good example is Jmol, which was used in the past as a Java applet, but nowadays is also available for the web as the web version JSmol. Still, tools like Aquaria are not fully implemented in WebGL.

The basic chapters for the manuscript were fixed as follows:

- Introduction (with history, from Java Applets to WebGL)

- Methods (Technical Background: OpenGL, JavaScript, WebGL)

- Frameworks (such as ThreeJS and Unity3D; with table)

- Applications (Jmol, JSmol, Aquaria, Jolecule, etc.; with table)

- Data Formats

- Conclusion

- Outlook

A big advantage of the workshop was the fact that many developers of the different tools were available on site, so we decided to prepare the tables required for the final publications by directly confronting the developers with our questions. The result is an initial table that will be later revised by all co-authors.

The actual plan is to involve all attendees of the workshop to use the paper as a global outcome of the workshop, with a future outlook based on the work of the other breakout sessions.

One basic outcome will be the idea that there should be a basic framework, such as three.js, which is optimized for molecular graphics and which can be used by all other programmers as a base, which is then extended by their custom functionalities.

The initial manuscript should be finished by middle of October, and the idea is to submit the full manuscript until the end of this year.

5   List of Participants

- Marc Baaden, CNRS, France

- Gert-Jan Bekker, Osaka University, Japan

- Spencer Bliven, National Center for Biotechnology Information, USA

- J.B. Brown, Kyoto University, Japan

- Matthieu Chavent, University of Oxford, UK

- Mike Goodstadt, Centre Nacional d'Anàlisi Genòmica - Centre de Regulació Genòmica (CNAG-CRG) , Spain

- Christopher Hammang, Garvan Institute, Australia, Australia

- Bob Hanson, St. Olaf College Chemistry, USA

- Bosco Ho, Burnet Institute, Australia

- Michael Krone, VISUS, University of Stuttgart, Germany

- Haruki Nakamura, Osaka University, Japan

- Shuichi Onami, Riken Quantitative Biology Center, Japan

- Seán O'Donoghue, CSIRO & Garvan Institute, Australia

- Alexander Rose, UC San Diego/ RCSB Protein Data Bank, USA

- Peter Rose, UC San Diego/ RCSB Protein Data Bank, USA

- Andrea Schafferhans, Technische Universität München, Germany

- Ian Silitoe, University College London, UK

- Bjorn Sommer, University of Konstanz, Germany

- Christian Stolte, New York Genome Center, USA

- Baichuan Sun, CSIRO, Australia

- Hirofumi Suzuki , IPR, Osaka University, Japan

- Sandhya Tiwari, Riken, Japan

- Sameer Velanker, PDBe, EBI, UK

- Jenny Vuong, CSIRO Australia, Australia

- Merry Wang, Autodesk Research, Canada

- Jon Wedell, BMRB, Wisconsin-Madison, USA

- Philippe Youkharibache, National Institute of Health, USA, USA

- Jim Zheng, University of Texas Health Science Center at Houston, USA