On the Use of 1D, 2D, and 3D Visualisation for Molecular Graphics

Julian Heinrich*
CSIRO

Michael Burch†
Visualization Research Institute

Sean I. O'Donoghue‡
CSIRO
Garvan Institute of Medical Research

ABSTRACT

Macromolecular structures have become easily accessible by means of large databases, for example the World Wide Protein Data Bank. Several years ago, existing data in this field of research was rare and demanded for specialists and experts which wasted a lot of time to explore the data. Today, various visualisation techniques, tools, and systems have been designed and developed to support biochemists, chemists, or molecular biologists at certain tasks such as analysing protein functions, exploring ligand-binding sites, or understanding RNA signal and message processing. The intent of this paper is to discuss how 1D, 2D, and 3D representations are typically employed to facilitate some of these tasks. We describe common tasks that involve molecular structures and illustrate how 1D, 2D, and 3D visualisations are currently being used to address these. We discuss the benefits and drawbacks of these concepts in the context of these tasks and propose methods to provide empirical evidence by means of user evaluations.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology; I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms

1 INTRODUCTION

Understanding molecular structures is pertinent to our current knowledge of the basic principles of life. The most famous example of a structure leading to insight about its purpose was the discovery of the DNA double helix by Watson and Crick [34]. In recent years, structural biology has accumulated a wealth of data about proteins, RNA, DNA, binding sites and ligands derived from crystallography, NMR, and electron microscopy [19]. At the time of writing this manuscript, the protein data bank [1] contains over 100.000 entries for experimentally determined structures. In addition to that, modelling and simulation techniques generate structural information about many more proteins.

Ideally, a researcher working in this field is able to easily and quickly derive key insights from 3D structures by a visual depiction of the data, e.g. from the DNA double helix or a model of a protein. Finding those insights more or less reliably are due to the ability of the human visual system to recognize patterns very fast, which is one of the key benefits of visualisation in general - not only of 3D visualisation approaches.

While structural biologists are familiar with analysing macromolecular structures using three-dimensional representations, chemists might prefer to look at the interaction of a ligand with a protein using 2D drawings such as Lewis plots [16]. Although three-dimensional diagrams seem to be the best choice to represent spatial data, it has been found that a corresponding 2D representation of the same data might perform better in terms of accuracy and completion times for a given task [30]. Similarly, all information about the spatial configuration of a molecular structure can be captured in a distance matrix of all contributing atoms, and the patterns that emerge from visualising such a distance matrix can be meaningful to structural biologists.

In this work, we discuss the expected benefits and drawbacks of 1D, 2D, and 3D visualisations that are being used to analyse and communicate macromolecular structures. To this end, we present a list of common tasks in the analysis of macromolecular structures, indicate whether each task can be solved with each visualisation, and discuss the benefits and drawbacks of the various representations with respect to the tasks. We intend to use this list as a basis for a controlled user study in the future, where tasks must typically be answered by study participants using different visualisations as independent variables and time and accuracy as dependent variables. Such a study could then be used to drive the development of new visualisations to better support common tasks in structural biology.

Note that, for the purpose of this paper, we use the term 'dimensionality' to describe the amount of information a visualisation conveys about a molecular structure instead of the actual dimensionality of the visual representation. For example, sequence-based protein visualisations are one-dimensional, as they contain only information about the sequence of amino acids, whereas 3D models of macromolecules add spatial locations to each of the atoms and thus contain a full description of the data.

2 RELATED WORK

The visualisation of macromolecular structures has become a standard tool in molecular biology with many tools available to support a variety of tasks using 1D, 2D, and 3D representations [19, 13]. As it is out of the scope of this work to review rendering techniques and algorithms, we will focus on related work that is relevant to the 2D versus 3D debate in this section.

To some extent, molecules can be modeled as 3D point clouds depicting the spatial position of individual atoms. For these, various studies have shown that estimating distances and navigation [18], search, density estimation [30], memorizability [31], and usability [4] perform worse than 2D point clouds. In a slightly different setting, Tavanti and Lind [27] found that the spatial memory plays a crucial role for making a visualisation memorable and hence easier to handle for data analysis. However, these findings could not be confirmed in a subsequent study by Cockburn [5]. Others suggest that combined displays can be superior in some circumstances [29]. The question arises if this is also the case for macromolecular structure visualisations and if it is depending on a specific task.

One of the most frequently reported drawbacks for 3D graphics is the occlusion problem: objects located closer to the current viewpoint might hinder the view onto objects further behind. Elmquist and Tsigas define a taxonomy for occlusion management in 3D [8] and propose a set of design patterns that cover similar solutions to the occlusion problem, such as multiple viewports or route planning.

Taking the connectivity of atoms into account, molecular structures can also be interpreted as node-link representations of a graph with a fixed 3D layout [6]. While this model can be useful to ben-

*e-mail: julian.heinrich@csiro.au
†e-mail:michael.burch@visus.uni-stuttgart.de
‡e-mail:sean.o'donoghue@csiro.au

efit from the results of the graph drawing community with respect to the visualisation and evaluation of three-dimensional diagrams, other rendering modes such as cartoon or solvent-excluded surface can not be modelled as such and thus have to be investigated separately. Neglecting the inherent three-dimensional nature of the structure, traditional 2D graph drawing approaches might also be applied. These do not suffer from occlusion problems, but produce vast numbers of edge crossings that may lead to visual clutter and a degradation of performance at some tasks [22].

Molecular structures can further be enriched with a wealth of data of several types and complexity, ranging from free text of relevant scientific publications over all the information provided by the associated gene (if available) to organism-specific data. This generally allows for various visual metaphors - also in combined forms - to be applicable to molecular structures.

Interaction is a key concept in visualisation and visual analytics and an active area of research within the broader HCI community. Manipulating the data and/or the current visual representation on a display enables the user to select items, explore the data space, etc. [36]. For 3D graphics, typical low-level interactions include rotation, translation, and zoom. While these interactions are intended to support the user in overcoming occlusion and distortion effects, some tasks such as the Vandenberg test for mental rotation were found to take significantly longer when participants were allowed to interact with the stimuli [3]. Others argue that the display of task-relevant information is much more important than interaction for certain tasks to be completed effectively [14].

Only recently, three-dimensional interaction devices have become affordable as they found their way into commodity hardware. However, some initial studies that involved controlling molecular structures did not find significant differences between traditional mouse and keyboard controls and 3D interaction devices such as the Leap Motion and the Microsoft Kinect [23].

While there are various studies investigating different aspects of the 2D versus 3D debate, we are not aware of any user studies that compare 3D molecular graphics visualisations with their corresponding 1D and 2D representations. Consequently, in this paper we will discuss some ideas towards closing this gap.

3 1D Techniques

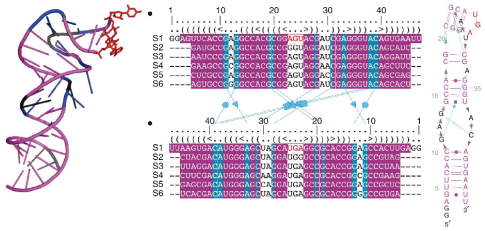

Figure 1: The structure of a RNA molecule in 1D (center), 2D (right), and 3D (left) visualisations [19].

Visualisations of sequences are typically rendered using a series of one-letter codes with colour or glyphs encoding residue features, such as secondary structure or single-nucleotide polymorphisms. Figure 1 (center) shows an example of coloured sequences of amino acids. Sequence views can easily be stacked (as in Figure 1) to compare multiple sequences and are particularly useful in a linked-view setup with 3D molecular graphics [20, 17]. Glyphs are frequently used to enhance the encoding of a particular feature. While sequence representations are very common, they obviously lack spatial positions of atoms and can therefore not be used to complete those tasks that require a spatial context of the structure. Note that for the tasks discussed below, we are not considering sequence views enhanced with visual links between residues.

4 2D Techniques

2D visualisations for molecular structures include abstract representations such as the distance matrix as well as 2D projections of 3D models such as chemical drawings or topology diagrams. In addition to the order of components or amino acids in a molecule, each component may be assigned two coordinates on a plane. While this allows one to convey some spatial information about the data, 2D visualisations can not provide the complete information in most cases. Here, we briefly review three of the most common 2D visualisations used for analysis and communication.

4.1 Lewis Representation

Two-dimensional drawings of molecules have a long tradition in chemistry to communicate the structure of a molecule in terms of the bonding between atoms. While Lewis drawings (see Figure 2, left and Figure 3) provide much detail with respect to the state of individual electrons (bound versus unbound), more compact forms may leave out covalent bonds altogether or introduce symbols for common compounds such as benzene. These drawings can be

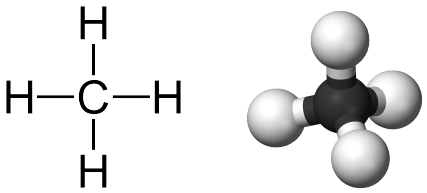

Figure 2: The molecular structure of Methane in 2D and 3D. While the 2D structure provides a simple and easy to understand depiction of the atomic composition, it fails to convey the 3D shape: the water molecules and the central carbon atom do not define a plane as the 2D image suggests.

used to effectively communicate the composition of molecules, but they are limited by the plane when it comes to conveying the three-dimensional structure, as illustrated in Figure 2. To address this issue, additional symbols may be used to indicate bonds with negative or positive z-direction, typically relative to a central carbon atom. In order for 2D drawings not to suffer from occlusion, a layout has to be computed that allows one to render atomic details without overlap and that preserves distances as much as possible. As Lewis plots provide a very detailed view on the structure, they are most frequently used to inspect protein/ligand interfaces (Figure 3) or protein/protein interfaces, where only parts of the structure are involved.

4.2 Distance Matrix

Another common technique to look at 3D molecular structures is via distance matrices. Here, the Euclidean distance for each residue (i.e. the α-carbon of each residue) or atom to every other residue or atom is used to construct a distance matrix, typically mapped to a colour scale (see Figure 4). Atomic distances are also obtained from Nuclear Overhauser Effect Spectroscopy (NOESY) experiments, which are then used to infer the 3D structure of a molecule via distance geometry computations. Hence, many structural biologists are used to inspecting distance matrices and are familiar with most of the patterns that emerge in this visualisation.

4.3 Topology Diagram

Topology diagrams are a schematic representation of protein secondary structure, as illustrated in Figure 4. Similar to the sequence

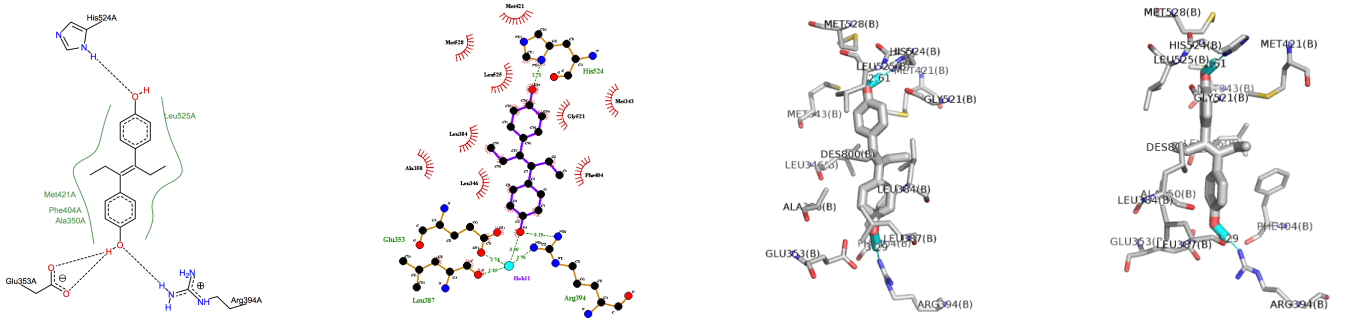

Figure 3: 2D and 3D representations of a protein/ligand complex (Estrogen receptor, 3ERD). From left to right: produced with PoseView [25], LigPlot+ [15] and PyMOL [24]. All images convey the structure of the ligand and amino acids in close proximity. Note that both 2D drawings on the left are aligned with respect to amino acids Glu353A (left) and Arg394A (right), while the ligand and His524 are rotated by 180 degrees on the right drawing. The 3D structure reveals that residues Arg394 and Glu353 are not located in the same plane as the ligand. The rightmost image is a rotated version of the left 3D image without labels. Note the phenylalanine (with its characteristic benzene ring in the side chain) is almost completely occluded in the left image and only becomes visible after rotation.

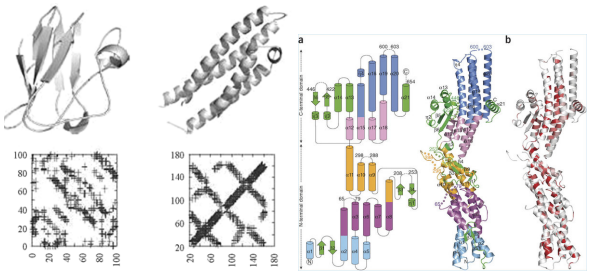

Figure 4: Protein families (top left) and the corresponding distance matrices (bottom left), as well as topology diagram (center) and 3D ribbon representation of a molecule (right). Figures reproduced with permission from [9] and [10].

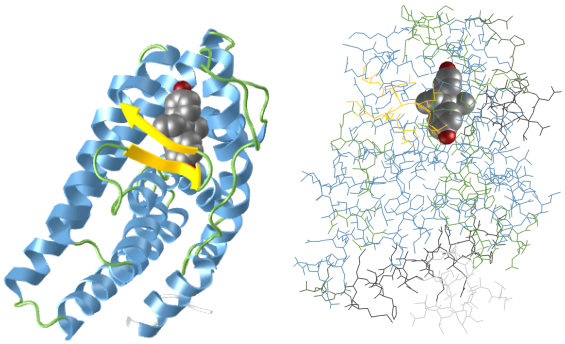

Figure 5: Different 3D representations of a model of the human estrogen receptor (3ERD) in complex with a ligand (Diethylstilbestrol). Left: ribbon rendering of the receptor molecule with spacefilling rendering for the smaller ligand. Right: Wireframe for the receptor.

representation, this technique uses glyphs in a 2D layout to depict secondary structure elements as they appear on the sequence. The topology diagram is a trade-off between its 1D and 3D counterparts, as it does not show the structure in 3D, but achieves to convey some of the spatial arrangement of the molecule without occlusion.

5 3D Techniques

3D visualisations are created from a 3D model of a molecule which contains the complete spatial information of the data in addition to the sequence of chemical compounds or amino acids. For visualisation, however, these models have to be projected to two dimensions and may lose important information about its spatial arrangement.

There are various approaches to rendering molecules in 3D [19], each emphasizing different aspects of the molecule [13]. Figure 5 (left) shows a ribbon diagram of the estrogen receptor (3ERD) in complex with a ligand rendered as spacefilling. The right image in the same figure shows the entire molecule in wireframe and the ligand as a space-filling model. Ribbon diagrams are intended to convey the spatial layout of a protein in terms of its secondary structure elements with a glyph-based approach. Secondary structure is thereby represented only with C_α atoms, visually encoded using geometry only. Space filling and wireframe representations include visual elements for every atom in the molecule, allowing for a more detailed view on the structure. Since the ball and stick representation is the one that most closely resembles the 2D drawings shown in the previous section, it might be the best option to compare with. Figure 3 shows the same molecule/ligand complex as Figure 5 using a 3D ball and stick representation.

6 Interaction

Interactive molecular graphics systems typically support a wide range of controls, in particular for the 3D view of a molecular structure or complex. These include rotating, translating, zooming, panning, picking, brushing, filtering, colouring, and many more. In addition to traditional mouse and keyboard controls, haptic devices [35] as well as a range of 3D interaction devices [23] have been proposed to facilitate controlling molecular structures. Among others, interaction facilitates occlusion and clutter management [7, 8], e.g. by viewing a scene from varying viewpoints or zooming into the view, which eventually reveals aspects of the data that were hidden before. This makes interaction key to exploration and enables the user to switch between overview and detail views. On the other hand, interaction adds complexity for the very same reason: multiple views of the same data have to be matched and combined by the user to create a mental map of the data. While interaction is generally considered to be essential in visual analytics [28], showing task-relevant information in static images can be of more importance than providing interaction [14].

Interaction allows the user to change parameters of a visualisation over time. As the parameter space may vary substantially between different visualisations, care must be taken to achieve a fair comparison between them in a comparative study. One option

would be to allow a user to control every single parameter for each visualisation. For a user evaluation, this ensures a maximal degree of freedom for every usage scenario and inherently captures the parameter space of each visualisation in the comparison. However, it also leads to a number of new problems: typical input devices such as the mouse and keyboard are not equally well suited to control all kinds of visualisation parameters. As a result, interaction parameters have to be categorized and possibly ranked according to their importance and feasibility for the available input devices. This would have to be done in such a manner that does not have a high impact on the comparability of the different techniques in order to keep it fair. Finally, software has to be written or adopted for all visualisations to support the control of all parameters. Altogether, these issues would render a comparative study of interactive visualisations impractical. Another option would be to reduce the number of interaction parameters to those that are applicable to all visualisations equally well and thus create a common ground for comparison. This raises the question of which interactions are common to all visualisations and how to find a fair set of controls. Rotation, for example, is essential for users to understand 3D models of proteins, but is probably less important for 1D and 2D visualisations.

7 Common Tasks

The visualisation of molecules can support both exploratory and well defined analysis settings. However, in order to compare the effectiveness of using 2D and 3D to render molecular structures, we assume that significant results are more likely to be obtained with a confined set of tasks, provided that a solution exists for each. In that case, measurements such as response times and accuracy are much easier to assess statistically than data gathered by observational and 'think aloud' studies.

Due to the defined hierarchy of molecules, potential tasks can be carried out at various levels, ranging from atoms and residues to domains and polymer chains. Depending on the level of detail and the purpose of a structure visualisation, 1D, 2D, and 3D representations are being used in the literature: 1D sequence-based representations readily support tasks based on a residue basis, 2D drawings of molecules are frequently used for detailed atomic scale representations, in particular for smaller organic molecules, while some of the most popular 3D visualisations show the locations of α-carbons only (ribbon, trace) and can be used to visualise large molecule complexes. On a higher level, i.e. to communicate protein interactions or cellular processes that involve proteins, symbolic 2D representations are also very common. Here, we limit our investigation to the highest level supported by the PDB and list only tasks that involve the biological assembly (or biounit) of a single PDB entry, which is believed to be the functional form of the molecule.

We compiled a (non-exhaustive) list of tasks in molecular biology that require a certain level of insight into the composition and/or the structure of a biological assembly. Note that biology is complex and proteins play a key role in many biological processes, rendering it impossible to list all tasks a biologist might need to complete in order to answer his particular question. The reason for this is the wealth of information that can be linked to every single molecular structure, such as interaction networks, gene expression profiles, gene locations, etc. The following tasks were chosen to include information that is contained in or can be derived from (i) the sequence of amino acids and the (ii) spatial configuration of the molecule:

1. Obtain an overview of the composition of the molecule, based on sequence features: the number and type of molecules (RNA/DNA, protein, ligand), the number of chains and domains, the oligomerization state, secondary structures, etc.

2. Find and characterize structural motifs, such as transmembrane regions or binding pockets without annotation.

3. Characterize the spatial context of single residue features, e.g. find spatially clustered post-translational modifications (PTMs), single-nucleotide polymorphisms (SNPs), etc.

4. Characterize volume or surface features.

5. Identify which atoms and residues are involved in binding a ligand given a distance threshold and annotations for residues/atoms that are closer to each other than the threshold.

6. Identify and characterize macromolecular (e.g. protein-protein) interfaces: Similar to task 5, but for large molecules.

7. Identify two different conformational states of a molecule. For this task, animation or a sequence of static images may be used.

8. Compare molecule ensembles: determine different sub conformations. For this task, animation or a sequence of static images may be used.

Some of these tasks can be solved using various representations, while others require a particular type of visualisation in order to obtain a solution. Table 1 shows our hypotheses whether a task can be solved for each representation type, along with the total number of citations from some key papers about each visualisation as a measure of usage in the scientific community. Note that for the purpose of this table, we assume a single static image for every task.

8 Discussion

For each task in Table 1, we state our hypotheses about whether it is solvable by visually inspecting any of the visualisation techniques discussed above. The null hypothesis for each cell is expressed in terms of three possible levels: n = the task cannot be solved, p = the task may be solved partially if it allows for a partial solution (e.g. counting secondary structures in a 3D view where occlusion may occur), y = the task can be solved entirely. Note that we based our decisions on the amount of information any of the views is capable of conveying with respect to the fully annotated 3D model of a structure. If these tasks were to be used in a user study, our hypotheses would have to be refined (e.g. with a pilot study) and are likely to change slightly. In addition, a precise set of constraints would have to be defined for each task, visualisation type, and interaction technique that is allowed in a controlled study to solve a task. For example, choosing an initial viewpoint is critical for a 3D view if no interaction is allowed. This raises the question which viewpoint to chose and whether to allow participants to control it or not.

Until confirmed by a formal user study, we can only discuss our expectations of the performance of 1D, 2D, and 3D visualisations with respect to the list of tasks given above:

We would expect that simple, sequence-based representations will perform better both in terms of accuracy and time than 2D and 3D representations for all tasks that are solvable with it. For example, task 1 involves finding and counting a number of features represented by shape (glyphs), colour, or both in the different visualisations. As these features are all based on the sequence of amino acids in the molecule, sequence representations seem to be tailored for this task. However, tasks that require a spatial embedding, such as the discovery of functional motifs (task 2) or the analysis of hydrophobic interactions cannot be solved using the sequence representation. Given a meaningful annotation, some tasks can be solved partially (i.e. on the residue level), as for example the identification of binding sites or residues involved in protein-protein interfaces.

The distance matrix contains all the information required to reconstruct a 3D model of a molecule. In theory, all tasks (except 4, which requires information beyond the topology of the molecule)

Task	Sequence	Distance Matrix	Lewis Plots	Topology Diagram	3D
1 Sequence Features	y	y	y	y	p
2 Structural Motifs	n	y	n	p	p
3 Single Residue Features	n	y	p	p	p
4 Volume and Surface Features	n	n	p	n	p
5 Ligand Binding	p	y	y	p	p
6 Macromolecular Interface	p	y	y	p	p
7 Molecular Dynamics	n	y	p	p	p
8 Ensembles	n	y	n	p	p
References	[33, 20, 17]	N/A	[16, 32, 25]	[2, 26]	[21, 12, 24]
Citations	1813	N/A	3578	204	31277

Table 1: Which tasks can be solved with different technologies? y = yes, n = no, p = partial. For each visualisation, we list up to three of the most prominent references in terms of the number of citations and provide the total number of citations (as per Google Scholar on September 25, 2014) as an indication of their usage in the scientific literature.

are therefore solvable using this 2D representation. However, since distance matrices represent a very abstract way of conveying this information, we assume that only experts (such as structural biologists who are used to working with distance matrices) are able to accurately solve most of the tasks. Inexperienced participants, to the contrary, will struggle with recognising even simple motifs such as alpha helices and beta sheets from the patterns emerging in a distance matrix. We would therefore consider the overall performance of the distance matrix to be at the lower end of the spectrum for most tasks.

Lewis drawings (and its variations) have a long tradition in chemistry and biochemistry. Participants with a background in either of these fields will certainly perform better than others, though some tasks require good spatial orientation skills as molecular surface features or 3D configurations are not easy to extract from these drawings. Furthermore, we are not aware of any software that supports drawing large molecules (e.g. proteins with more than 20 residues) using 2D representations and are thus unable to assess their usability for macromolecular structures. However, assuming we could draw even large structures, we would expect the performance to decrease with the size of the data more rapidly than with other representations, as Lewis drawings don't scale well due to their high level of detail. The actual 3D configuration of a molecule is expected to have a high impact on the performance, as more complex structures are more likely to result in a distorted view when projected to 2D. Note that although Lewis drawings are frequently used to investigate ligand/ligand or protein/ligand interfaces, we marked this task to be only partially solvable in 2D, as it may require the 3D configuration of the interface atoms to be solved.

Topology diagrams are an interesting variation between sequence views using glyphs for secondary structure and 3D ribbons, as they allow for some spatial information to be displayed, such as clusters of helices or sheets, but do so without occlusion. Hence we assume that all tasks (except 4) are at least partially solvable using this representation. Tasks that require the 3D context of the molecule might benefit from the occlusion-free layout of a topology view, but are most probably not fully solvable, depending on the data at hand.

3D interactive molecular graphics are probably the most widely applied visualisation technique for macromolecular structure (cf. number of citations in Table 1. As 3D views enhanced with interaction to rotate molecules conveys all the information contained in the data, we assume that this type of representation can be used to solve all tasks accurately. However, due to the occlusion problem, all tasks will only be partially solvable using static 3D approaches. For this type of representation, we expect the overall performance to be highly dependent on whether interaction is allowed or not: with interaction, it is probably above the others for tasks that are likely to benefit from a 3D spatial embedding (such as 2, 3, 7, and 8). Due to the complexity added by interaction, completion times may be slower than for other views. Without interaction, though, occlusion will have a high impact on the accuracy of results, but completion times might still be faster than others.

Finally, we assume that there would be large variation in task performance depending on the data and the background of a participant: Molecular structures differ in size and complexity with respect to their spatial arrangement and the contained features. Sampling structures randomly from the PDB will result in many unsolvable tasks, as most tasks require some feature of a structure to be present. A more practical approach is to manually compile a set of datasets for the study. In this case, care must be taken to find a sample that is a good approximation to the population of structures (in terms of the features they contain) in order to get results that do not depend on the data. The same is true for selecting potential participants: structural biologists that work in the field for a long time are used to inspecting 3D structures and distance matrices, whereas biologists might perform better on sequence views.

9 Conclusion

This work is intended to start a discussion about (i) how to evaluate different visual representations for macromolecular structures and (ii) how to improve on existing techniques based on what is learned from such an evaluation. To this end, we presented a list of tasks that could be used as a starting point in a formal user study to measure completion time and accuracy with respect to different visualisation techniques and discussed our hypotheses and expectations towards the performance of the different representation types. It is important to note that the list of possible visualisations as well as the list of tasks are not complete with respect to the huge space of possible tasks and ways of solving them. Similarly, many tasks can be solved using other visualisations or automatic approaches that we did not consider in this work. Examples are statistical plots such as the Ramachandran plot or parallel coordinates [11], graphs and networks [6] as well as visualisation techniques that make use of stereoscopic 3D displays, or hybrid approaches. The discussion of the role of interaction was also restricted to very basic operations, though 3D input devices such as the Leap Motion controller, 3D mice, or the phantom haptic device lend themselves to control macromolecular structures.

Traditional visualisation evaluations record completion time and accuracy in order to compare the performance of different techniques with each other. The next step towards improving on existing techniques is to understand why some representations perform better than others. A promising method to investigate this question is to track the participants eyes during the test. Eye-tracking devices record the time and position of fixations on the screen, which can be used, e.g., to investigate the strategies users employ to solve

a task [3]. This data, in turn, can be used to reveal the limitations of a visualisation by analysing where participants spent most of their time looking at while solving a task.

While interactive 3D molecular graphics are ubiquitous in structural biology, we propose that there is scope for conducting research towards developing more sophisticated 1D and 2D visualisations to aid molecular biologists in understanding their data. Given that both interaction and 3D were previously found to add more complexity instead of helping to reduce it, applications such as molecular dynamics simulations or the analysis of protein ensembles could greatly benefit from fresh visualisation ideas.

Acknowledgements

This work was supported by CSIROs OCE Science Leader program and its Computational and Simulation Sciences platform.

References

[1] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[2] C. S. Bond. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics*, 19(2):311–312, 2003.

[3] J.-M. Boucheix and M. Chevret. Alternative strategies in processing 3d objects diagrams: Static, animated and interactive presentation of a mental rotation test in an eye movements cued retrospective study. In *Diagrammatic Representation and Inference*, number 8578 in Lecture Notes in Computer Science, pages 138–145. Springer Berlin Heidelberg, 2014.

[4] M. Chalmers. Using a landscape metaphor to represent a corpus of documents. In *Spatial Information Theory A Theoretical Basis for GIS*, number 716 in Lecture Notes in Computer Science, pages 377–390. Springer Berlin Heidelberg, 1993.

[5] A. Cockburn. Revisiting 2d vs 3d implications on spatial memory. In *Proceedings of the Fifth Conference on Australasian User Interface - Volume 28*, pages 25–31, Darlinghurst, Australia, Australia, 2004. Australian Computer Society, Inc.

[6] N. T. Doncheva, K. Klein, F. S. Domingues, and M. Albrecht. Analyzing and visualizing residue networks of protein structures. *Trends in Biochemical Sciences*, 36(4):179–182, 2011.

[7] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216–1223, 2007.

[8] N. Elmqvist and P. Tsigas. A taxonomy of 3d occlusion management for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):1095–1109, 2008.

[9] J. Gomide, R. Melo-Minardi, M. A. d. Santos, G. Neshich, W. Meira Jr., J. C. Lopes, and M. Santoro. Using linear algebra for protein structural comparison and classification. *Genetics and Molecular Biology*, 32(3):645–651, 2009.

[10] V. Guillet, M. Knibiehler, L. Gregory-Pauron, M.-H. Remy, C. Chemin, B. Raynaud-Messina, C. Bon, J. M. Kollman, D. A. Agard, A. Merdes, and L. Mourey. Crystal structure of -tubulin complex protein GCP4 provides insight into microtubule nucleation. *Nature Structural & Molecular Biology*, 18(8):915–919, 2011.

[11] J. Heinrich, M. Krone, S. O'Donoghue, and D. Weiskopf. Visualising intrinsic disorder and conformational variation in protein ensembles. *Faraday Discussions*, 2014.

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

[13] G. T. Johnson and S. Hertig. A guide to the visual analysis and communication of biomolecular structural data. *Nature Reviews Molecular Cell Biology*, 15(10):690–698, 2014.

[14] M. Keehner, M. Hegarty, C. Cohen, P. Khooshabeh, and D. R. Montello. Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science*, 32(7):1099–1132, 2008.

[15] R. A. Laskowski and M. B. Swindells. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*, 51(10):2778–2786, 2011.

[16] G. N. Lewis. THE ATOM AND THE MOLECULE. *Journal of the American Chemical Society*, 38(4):762–785, 1916.

[17] E. C. Meng, E. F. Pettersen, G. S. Couch, C. C. Huang, and T. E. Ferrin. Tools for integrated sequence-structure analysis with UCSF chimera. *BMC Bioinformatics*, 7:339, 2006.

[18] G. B. Newby. Empirical study of a 3d visualization for information retrieval tasks. *Journal of Intelligent Information Systems*, 18(1):31–53, 2002.

[19] S. I. O'Donoghue, D. S. Goodsell, A. S. Frangakis, F. Jossinet, R. A. Laskowski, M. Nilges, H. R. Saibil, A. Schafferhans, R. C. Wade, E. Westhof, and A. J. Olson. Visualization of macromolecular structures. *Nature Methods*, 7:S42–S55, 2010.

[20] S. I. O'Donoghue, J. E. W. Meyer, A. Schafferhans, and K. Fries. The SRS 3d module: integrating structures, sequences and features. *Bioinformatics*, 20(15):2476–2478, 2004.

[21] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF chimera–a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.

[22] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin. Feature congestion: a measure of display clutter. In *Proceedings of the 2005 Conference on Human Factors in Computing Systems, CHI 2005, Portland, Oregon, USA, April 2-7, 2005*, pages 761–770, 2005.

[23] K. Sabir, C. Stolte, B. Tabor, and S. O'Donoghue. The molecular control toolkit: Controlling 3d molecular graphics via gesture and voice. In *2013 IEEE Symposium on Biological Data Visualization (BioVis)*, pages 49–56, 2013.

[24] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1. 2010.

[25] K. Stierand and M. Rarey. Drawing the PDB: Proteinligand complexes in two dimensions. *ACS Medicinal Chemistry Letters*, 1(9):540–545, 2010.

[26] A. Stivala, M. Wybrow, A. Wirth, J. C. Whisstock, and P. J. Stuckey. Automatic generation of protein structure cartoons with pro-origami. *Bioinformatics*, 27(23):3315–3316, 2011.

[27] M. Tavanti and M. Lind. 2d vs 3d, implications on spatial memory. In *IEEE Symposium on Information Visualization*, pages 139–145, 2001.

[28] J. J. Thomas and K. A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–3, 2006.

[29] M. Tory, A. Kirkpatrick, M. Atkins, and T. Moller. Visualization task performance with 2d, 3d, and combination displays. *IEEE Transactions on Visualization and Computer Graphics*, 12(1):2–13, 2006.

[30] M. Tory, D. Sprague, F. Wu, W. Y. So, and T. Munzner. Spatialization design: Comparing points and landscapes. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1262 –1269, 2007.

[31] M. Tory, C. Swindells, and R. Dreezer. Comparing dot and landscape spatializations for visual memory differences. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1033–1040, 2009.

[32] A. C. Wallace, R. A. Laskowski, and J. M. Thornton. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering*, 8(2):127–134, 1995.

[33] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. Jalview version 2a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191, 2009.

[34] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[35] A. M. Wollacott and K. M. Merz Jr. Haptic applications for molecular structure manipulation. *Journal of Molecular Graphics and Modelling*, 25(6):801–805, 2007.

[36] J. S. Yi, Y. ah Kang, J. Stasko, and J. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, 2007.