Eurographics Workshop on Visual Computing for Biology and Medicine (2019)
B. Kozlíková, L. Linsen, and P.-P. Vázquez (Editors)

Hybrid Visualization of Protein-Lipid and Protein-Protein Interaction

Naif Alharbi[1], Michael Krone[2], Matthieu Chavent[3], Robert S. Laramee[1]

[1]Swansea University, UK
[2]University of Tübingen, Germany
[3]Institute of Pharmacology and Structural Biology, Toulouse, France

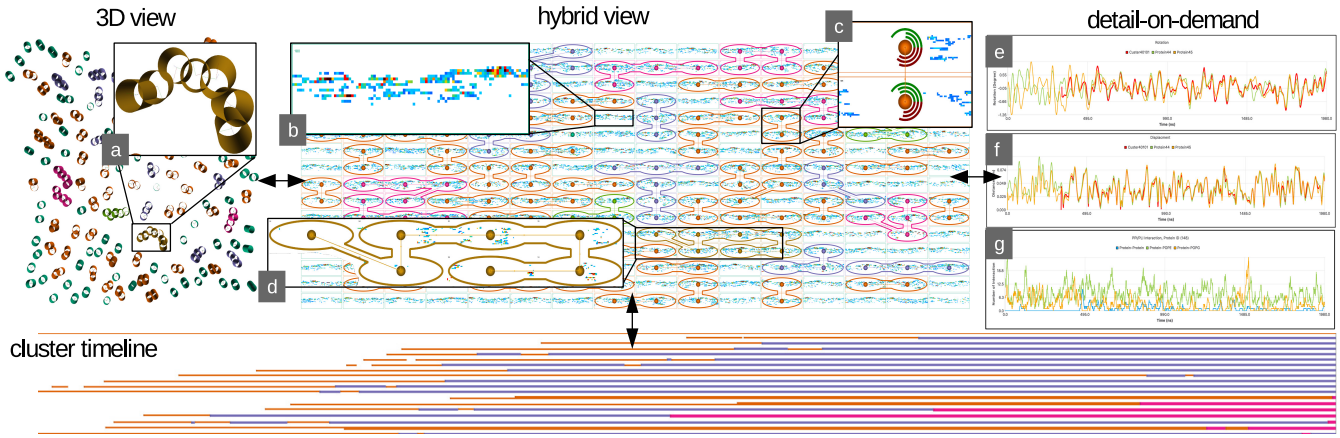

Figure 1: *Four views of Protein-Lipid Interaction (PLI) and Protein-Protein Interaction (PPI). The 3D view (top left), a) zoomed in snapshot of a cluster in the 3D view. The 2D hybrid view (top middle). Abstraction is used to explore PLI and PPI in 2D space. b) PLI depicted with a tiled heat-map style visual design. c) A depiction of PPI, and rotation. The PPI is represented by a tiled visual design (same as PLI) and a glyph is used to convey the protein behavior. d) A large cluster representation (the same selected cluster in the 3D view). The hybrid 2D view enables the user to couple and decouple the PLI and PPI imagery. Details-on-demand (top right), e) and f) interactive time series graphs display the amount of rotation and displacement of a cluster and its proteins respectively. g) an interactive time series that displays the frequency of interactions between a given protein and other molecules (a protein and two lipid types). The cluster timeline (bottom), an interactive timeline by which the user can track the history of each cluster. The framework can be used to study the historical evolution of a cluster. Any cluster can be selected in the 3D view to be analyzed in the hybrid view. The user can observe details and investigate a particular protein from the cluster. More detail about the protein such as PLI, PPI, and amount of rotation can be conveyed in the form of glyphs and graphs.*

Abstract

In the Molecular Dynamics (MD) visualization literature, different approaches are utilized to study protein-lipid interactions (PLI) and protein-protein interaction (PPI) in decoupled contexts. However, the two types of interaction occur in the same space-time domain. It is beneficial to study the PLI and PPI in a unified context. Nevertheless, the simulation's size, length, and complexity increase the challenge of understanding the dynamic behavior. We propose a novel framework consisting of four linked views, a time-dependent 3D view, a novel hybrid view, a clustering timeline, and a details-on-demand window. We introduce a selection of visual designs to convey the behavior of PLI and PPI through a unified coordinate system. Abstraction is used to present proteins in hybrid 2D space, a projected tiled space is used to present both PLI and PPI at the particle level in a heat-map style visual design while glyphs are used to represent PPI at the molecular level. We couple visually separable visual designs in a unified coordinate space. The result lets the user study both PLI and PPI separately or together in a unified visual analysis framework. We also exemplify its use with case studies focusing on protein clustering and we report domain expert feedback.

CCS Concepts

• ***Human-centered computing*** → *Visualization;*

1. Introduction and Motivation

In computational biology, scientists rely on simulation to study life cells and the behavior of molecules. Monte Carlo (MC) sampling and Molecular Dynamics Simulation (MD) are widely used to simulate large molecular systems (Frenkel and Smit [FS01]). With the recent advances in GPU computing and acceleration techniques, MC and MD software can produce very large time-dependent simulations. Simulation involves the evolution of molecular systems in the form of sequential position-snapshots that represent the trajectory of particles. The nature of dynamical systems is influenced by many molecular phenomena such as molecular interactions and molecular clustering.

Molecular interactions are studied in different fields such as protein folding, drug design, and origins of life study. Spatio-temporal protein clustering also plays an essential role in studying the spatio-temporal dynamics of membranes [CDR*18]. The molecular interactions occur at the atomic level resulting from attractive or repulsive forces between molecules and between non-bonded atoms. However, a molecular interaction can involve molecules of different types and sizes, e.g. protein-lipid interaction (PLI) or molecules from the same classes and sizes, e.g. protein-protein interaction (PPI). Protein spatio-temporal clustering occurs when two proteins (or more) are in close proximity to make an interaction.

Molecular interactions are often classified into two interaction types; PLI and PPI both of which are interesting. The two types of interaction occur in the same physical space which increases the challenge of understanding the behavior of the simulation. The interplay between PLI and PPI is expected to be a determinant factor in the assembly and dynamics of such membrane complexes [MGP*15]. Nevertheless, it is helpful to study the PLI and PPI in a unified context. Figure 2 shows the visualization of PLI and PPI in complex simulation data utilizing a naive approach. It can be seen, in addition to the view-dependency problem in the 3D visualizations, many of the interaction details are obscured due to particle overlap and occlusion. If we try to use color-mapping to visualize both PLI and PPI in the traditional 4D coordinate system, we encounter occlusion, visual complexity, and overlap. However, domain experts are still interested in analyzing and visualizing these two types of interactions both together and separately. Thus, we propose a novel, hybrid unfolded view that enables this. The PLI is depicted as a projected tiled space in a heat-map style visual design, while the PPI can be represented in the same coordinate space using tiles or glyphs which are easily separable from the tiled PLI visual design. By exploiting visually separable visual designs in a unified coordinate space we can provide the domain expert with a unified view of PLI and PPI with reduced occlusion and overlap while at the same time enable focusing on either PLI or PPI. We propose a novel visual design utilizing a tiled map of both PLI and PPI based on the abstract interaction space of proteins described in Alharbi *et al* [ACKL18]. Similarly, protein clustering and PPI occur in the same physical space and are associated with each other. It makes sense that they are studied in a unified visual framework. We employ abstract glyphs to convey the dynamic behavior of PPI and the different dynamic properties of proteins and their clusters.

In this work we propose a novel molecular interaction and protein clustering visualization tool including a linked 3D, visual abstract, and novel, projected hybrid view to convey a unified representation of PLI, PPI, and protein clustering. The tool enables the user to explore PLI, PPI and protein clustering interactively in a unified system. Additionally, we enable the user to analyze the details of an individual protein molecule or cluster by applying selection, focus-and-context, zooming, and details-on-demand techniques. Computational biologists can investigate both PLI and PPI, and protein clustering using a hybrid 2D and 3D view in a unified tool. We make the following contributions in this paper:

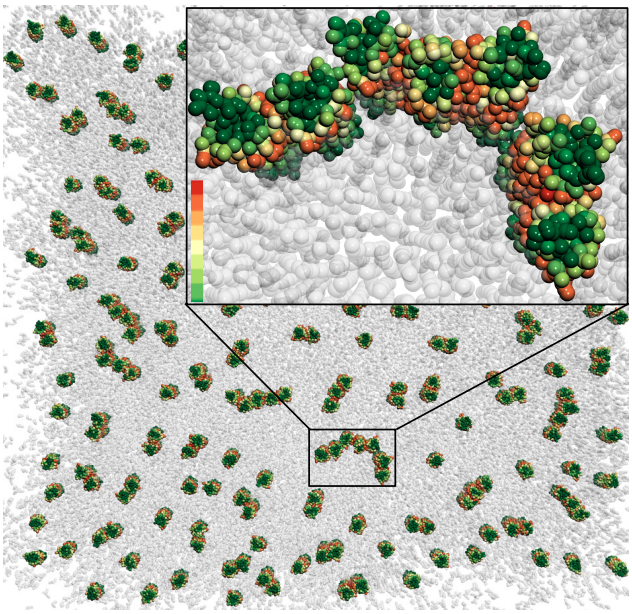

Figure 2: *Visualizing PLI and PPI using a naive approach. Protein particles are color mapped to the number of interactions they received. Lipid particles are rendered as context to reduce occlusion, overlap and visual complexity.*

- The first tool to both visually separate and combine time-dependent PLI and PPI imagery in a unified visual design. Our novel design is based on projecting the cylindrical interaction-space of each protein into a tiled, time-dependent 2D space.
- A novel, glyph-based representation of PPI that also encodes rotational properties. We utilize abstraction to depict proteins and clusters in a hybrid 2D view side-by-side the 3D view. Dynamic glyphs are used to visualize time-dependent properties such as angle of cluster rotation, displacement, and radius.
- Case studies focused on domain expert hypotheses related to protein clustering, rotation and translation including feedback from a domain expert in computational biology.

The rest of the paper is structured as follows. After discussing related work in Section 2, we describe the simulation and our methodology on computing the properties of PLI, PPI and protein clusters in Section 3. The molecular interaction and protein clustering visualization approaches are described in Section 4. In Section 5 we provide two case studies and report domain-expert feedback. The paper concludes in Section 6, where we also discuss future work.

		Ref.	PLI		PPI		Clustering		Vis. Capacity	
			2D	3D	2D	3D	2D	3D	N	1
Molecular Interaction	PLI	Falk *et al.* [FKRE09]		T						✓
		Le Muzic *et al.* [LMPSV14]		T						✓
		Skaňberg et al. [SVGR16]		T						✓
		ZigCell3D [dHCKMK13]	T	T					✓	✓
		Khazanov and Carlson [KC13]	S							✓
		Hermosilla *et al.* [HEG*17]	T	T						✓
		Vázquez *et al* [VHG*18]	T	T						✓
		Alharbi *et al* [ACKL18]		T					✓	✓
	PPI	Finn *et al.* [FMB04]			S	S				✓
		PocketQuery [KC12]				S				✓
		Furmanova *et al.* [FBG*18]			S	S				✓
		Alharbi *et al* [ACKL18]				T			✓	✓
Clustering		Le Muzic *et al.* [LMWPV15]						T		✓
		Chavent *et al.* [CDR*18]					T	T	✓	
		this work	T	T	T	T	T	T	✓	✓

Table 1: *A classification of related work based on features. (T) visualizing time-dependent simulation. (S) visualizing static data. (N) indicates a focus on a complex molecular system. (1) indicates a focus on a single molecule.*

2. Related Work

The field of molecular visualization has received significant attention in the last decades. For an overview, we recommend the following recent survey papers on molecular visualization. O'Donoghue *et al.* [OGF*10] review visualization methods and tools that enable the community of life scientists to obtain insight into their molecular data. Gehlenborg *et al.* [GOB*10] review stand-alone and web-based applications that are used to visualize PPI with a focus on 2D visualization techniques such as interaction networks, scatter plots, heat maps, dendrograms and graphs. In two recent state-of-the-art reports, Kozlíková *et al.* [KKF*17] review and classify visualization techniques developed to render molecular structures and Krone et al. [KKL*16] review the visual analysis of biomolecular cavities. Alharbi *et al.* [AAM*17] present a survey of surveys (SoS) on molecular dynamics visualization (11 surveys in total). The SoS focuses on challenges for computational biologists which can be addressed by computer graphics approaches. Most recently Miao *et al.* [MKK*18] provide a high-level survey of multi-scale molecular visualization techniques focusing on application domain questions, challenges, and tasks.

We present the related work based on our contributions, 1) the visualization of two molecular classes of interactions (PLI and PPI) and 2) visualization of protein clustering. Table 1 classifies the related work based on the primary focus of each paper.

2.1. Visualization of Molecular Interactions

Here we classify the work on visualizing molecular interactions into 1) PLI visualization, and 2) PPI visualization. In each sub-section, the related work is, also, classified based on the main visualization theme (spatial 3D visualization or information visualization).

Protein-lipid Interaction (PLI) Visualization The following papers focus mainly on 3D and 4D visualization. Two approaches have been used to highlight PLI in 3D, 1) glyphs, and 2) color-coding interaction on the surface of individual particles (a sphere) or molecular surfaces. Falk *et al.* [FKRE09] develop a visualization to show molecules of interest as well as their trajectory and interactions. Their method enables the user to zoom in and observe protein interaction by following the trajectory of the protein in 3D. The interactions along the trajectory of a protein are highlighted by glyphs. Le Muzic *et al.* [LMPSV14] design a framework to convey the spatial aspects of molecular interaction. The user can focus on a single molecule such that interactions between the given molecule and other are brought into focus in front of the user in a 3D view. The interacting protein is color mapped to a user-selected attribute. Skaňberg et al. [SVGR16] visualize pairwise interaction strengths between molecules utilizing mutual interreflections as a new visual communication channel. The interaction strength is encoded on the surface of atoms in 3D view. ZigCell3D [dHCKMK13] provides 3D visualization of molecular systems while the interaction between particles is visually highlighted. The visualization includes an interaction network by which the user can interactively select a molecule of interest in the 3D visualization. The network stores the interaction state and represents the relationship between molecules. Alharbi *et al* [ACKL18] focus on the PLI and PPI interaction space. They propose a cylindrical geometry surrounding each protein to capture PLI and PPI. The interaction frequency is stored in the tiles of the cylinder. The proposed solutions lack details of the interactions. With the exception of ZigCell3D [dHCKMK13] and Alharbi *et al* [ACKL18], no overview concerning the system is provided.

In the field of information visualization, Khazanov and Carlson [KC13] visualize the contacts with ligands and their binding sites by modifying the atoms' Van der Waals radii and color. They provide several visualization elements in the form of tables and different types of 2D graphs. Hermosilla *et al.* [HEG*17] present a visualization approach addressing protein interaction by considering the interaction forces. The visualization utilizes two linked-views, the first view conveys the 3D structure of the protein and the ligand, while the other depicts the spatial arrangement of the atoms. Vázquez *et al* [VHG*18] present a compact 2D visualization of molecular simulations. They utilize 2D information visualization tools with coordinated views to present physical properties of single molecular components and their pairwise interactions. The ligand is shown at the center of the display, and the user can hover the mouse over the secondary structure or the residue to obtain the details.
In summary, these researches provide great detail concerning molecular interaction, however, most of them are limited to the interaction of either a single molecule or a single timestep. Our work differs from the others in two aspects. First, our framework enables the user to explore the entire time-dependent simulation of a complex membrane consisting of 256 proteins and more than 18K lipids. Second, it is capable of visualizing two types of interaction, PLI, and PPI. In addition, it provides historical information about the PPI at the molecular level, i.e. clustering. This is indicated in Table 1.

Protein-Protein Interaction (PPI) Visualization Finn *et al.* [FMB04] propose three visual designs to investigate PPI, 1) a visualization of a PPI network, 2) a 3D representation of the protein, and 3) a schematic diagram that links to the images of residue-level interactions. PocketQuery [KC12] is a graphical interface which is designed to investigate the properties of PPI. A receptor protein is displayed by its surface, and a ligand molecule is represented by a stick. The receptor surface is color mapped by the partial charge of the residues while the ligand can be color mapped to different properties such as energy estimates. The above tools are limited to individual molecules and do not support dynamic simulations.

Most recently Furmanova *et al.* [FBG*18] propose a novel set of visual designs focussing on PPI configuration space. They utilize different interactive techniques summarized in a PPI matrix view to enable the user to visually filter configurations that contain a desired combination of interacting amino acids. A 3D representation of the filtering is depicted in a 3D view. To address overlap, occlusion, and visual clutter they enlarge the distance between the interacting proteins. The tool is limited to exploring the PPI configuration space of two proteins.

The visualization of time-dependent PPI has received less attention than PLI which suggests that the study of PPI is more of an open research area. In general, the previous work is dedicated to studying a particular class of interaction either PLI or PPI. With the exception of ZigCell3D [dHCKMK13] and Alharbi *et al* [ACKL18], the previous work is generally not capable of visualizing interactions of time-dependent (MD) interaction for the entire lifespan of molecular systems. Our tool enables the user to visualize PLI and PPI simultaneously for the time-dependent simulation of a membrane including 256 proteins and 19K lipid molecules. The tool provides an overview of the dynamical system, and also the user can focus on a specific protein molecule.

2.2. Visualization of Protein Clustering

In the molecular visualization literature, the term 'clustering' is often associated with visualizing the result of various analysis routines which are categorized in the visual analytics field. In this work, we focus on the spatial-temporal clustering of proteins.

The term 'spatial clustering' appears more often in the imaging and spectroscopic analysis literature [ORR*10, OWM*12, LKWS13, PBI*15, CZA*16] while it is seen less often in molecular dynamics visualization. Le Muzic *et al.* [LMWPV15] propose an illustrative time-lapse method that slows down the movement of proteins while they are involved in a reaction. The tool is designed to address the challenge of tracking molecules in consecutive time steps. This visualization can be used to identify protein clusters with the ability to zoom in and out of time. The limitation of this work is that it cannot provide a comprehensive view of the system clusters. Most recently Chavent *et al.* [CDR*18] provide a set of 2D and 3D visualizations employing VMD [HDS96] and matplotlib [Hun07]. The proposed visualizations are used to study and compare large-scale simulation to experimental data of outer membrane protein behavior. However, the visualizations lack interactivity, hence, two separate tools are used.

Most of the previous work focuses on a single aspect of the molecular visualizations, i.e. either scientific or information visualization. The 3D and 4D visualizations alone might be less informative especially when the underlying information about the interactions is desired. The information visualizations are sometimes the best choice to convey details on demand concerning interaction. However, the information visualizations represent abstract data which is useful to link to its 3D source. A combination of 3D visualization, visual abstraction, and information visualization can enable a better understanding of the molecular system, especially if they are interactively linked. Furthermore, uniting these features in one tool accelerates the process of investigating the system. PLI, PPI, and protein clustering are three different phenomena that occur in a molecular dynamics system. However, in general there is still no means by which the three phenomena are integrated together.

In this work, we propose a hybrid and 4D space-time interactive visualization tool to investigate a time-dependent molecular dynamics system. The tool employs 2D and 3D time-dependent visualizations and enables the user to explore two types of molecular interactions (PLI and PPI) both separately or together simultaneously along with the clustering of proteins and their properties. The 2D and 3D views are complementary to each other and the user can observe them in combination. The tool is designed to help the user interactively explore the three aspects in detail utilizing a novel 2D projection and a protein abstraction containing dynamic properties of proteins and protein clusters.

3. Simulation Data and Computational Methodology

In this section we describe the simulation data and briefly discuss the methodologies we use to compute meta-data of interaction properties.

3.1. Simulation Data Description

The dataset we study represents the behavior of a membrane constituted by lipids and proteins resulting from a molecular dynamics simulation. The simulation is used to study the dynamic behavior of both proteins and lipids in the context of a membrane aligned primarily in the *xy* plane [CRG*14]. The dimensions of the system are $116 \times 116 \times 10$ nanometers (x, y, and z respectively) and the simulation represents individual trajectories of 336,260 particles over almost 2 microseconds (stored in c.a. 2000 frames). The system consists of three molecule types: one protein, and two lipid types (POPE and POPG). There are 256 protein molecules while the number of POPE and the POPG lipid molecules are 14,354 and 4,738 respectively. The models presented are based on the MARTINI coarse-grained forcefield [MT13] which does not represent all the atoms but simplifies four heavy atoms into one coarse-grained (CG) particle. Thus, the hierarchy for each type of molecule is simplified:

Each protein is constituted by 171 residues, and each residue is composed of 1 to 3 particles. The total number of particles per protein is 344 resulting in 88,064 particles in total (256 proteins $\times$ 344 particles).

The total number of lipids (POPG and POPE) is 19,092. Each lipid molecule has 13 particles which result in (19,092 molecules $\times$ 13 particles) 248,196 particles in total. For this representation, we divide each lipid into three groups: a head group consisting of 2 particles, a tail group consisting of 6 particles, and a second tail group consisting of 5 particles. In terms of the data size, the simulation contains more than 666 million space-time positions that occupy 8 Gigabytes of memory.

3.2. Computation

Deriving Rotational Data In the model of the membrane studied here, each protein molecule is constituted by a finite number of particles (344 particles). These particles are linked together through constraints defined by the forcefield which limits the overall displacement of one particle with respect to the others belonging to the

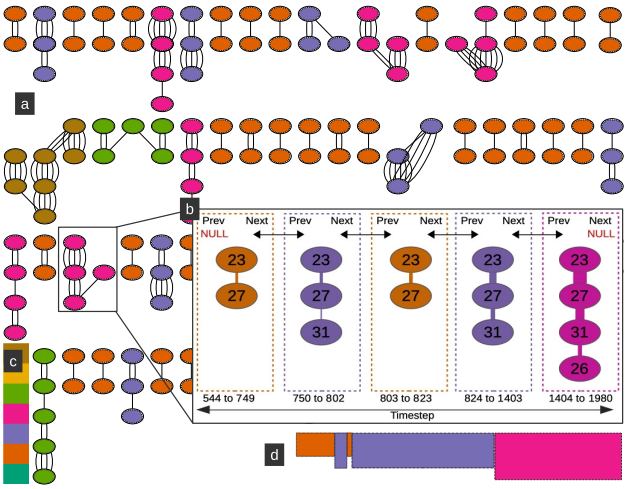

Figure 3: *a) A static, disconnected graph consisting of 172 connected sub-graphs (based on PPI). Each sub-graph represents a cluster (based on the final timestep). Cluster color encodes the size of the cluster. A node represents a protein. Edges between a protein pair represent PPI at the molecule level (i.e. the protein pair are attached). In the static graph, the number of edges between a protein pair indicates how many times the pair participates in forming a cluster. b) A resultant Time Varying Graph (TVG) shows the evolution of a cluster of 4 proteins. The thickness of the edge between protein pair encodes the time-span of the connection. Each dashed box represents a cluster with respect to the time span of its PPI. The arrows show the relationship between clusters within the TVG. c) A color map legend, the color mapps to the the cluster size between 1 to 7 (bottom to top). d) A cluster timeline of the relevant TVG. The height of the clusters in the timeline also encodes the cluster size to make the small cluster more visible. A cluster size of 1 is not considered.*

same protein. The dynamics of protein particles play a significant role in defining the geometry of a protein at a given time step. Some protein properties (e.g. global position, Center of Mass (CoM), principal axes, spatial extent, and rotation) are influenced by the system dynamics. As a result, these properties should be updated for every time step. We use Singular-Value Decomposition (SVD) [KL80] to calculate the main properties of a protein space i.e., the principal axes (*eigenvectors*), dimension sizes (*eigenvalues*) and rotation matrix as described in Alharbi *et al* [ACKL18]. In the case of clusters, proteins that form a cluster are addressed as a unified entity that has its own properties (e.g. global position, CoM, principal axes, and rotation).

Protein Radius The size and shape of a protein are often estimated, and the result varies depending on the mass of the protein. A sphere is often used to approximate the shape of a protein. The radius of the sphere can be calculated utilizing a number of parameters as described in Erickson [Eri09]. In this work, instead of the spherical shape, we employ a cylindrical shape dedicated specifically to the protein interaction space as described and computed in Alharbi *et al* [ACKL18]. The calculation results in a CoM and three principal axes (eigenvectors). The major axis is aligned with the height of the cylinder and cuts through the membrane while the other two axes are aligned with the cylinder base. Therefore, the radius of the cylinder can be estimated by the largest distance between the major axis and particles that lie in the *xy* plane. For each time-step, we calculate the radius of every individual protein. The result is utilized in spatio-temporal protein clustering.

Spatio-temporal Clustering Spatio-temporal clustering is a result of PPI. Proteins within 5 nm of one another are considered as interacting Chavent *et al.* [CDR*18]. The usual threshold represents a static diameter of a protein. Based on our calculations, the size of a protein's radius (the radius of a cylinder) varies between 1.7 nm to 2.5 nm, which implies that utilizing a static threshold might result in a less accurate PPI. In this work, we utilize a dynamic protein radius to determine PPI. We redefine the threshold distance of PPI based on a protein pair within a distance less than or equal to the sum of their respective radii.With this definition, we ensure that protein pair is always close enough to each other to initiate an interaction.

Cluster Timeline Construction In order to construct a cluster timeline we utilize a custom time-varying graph (TVG). A node represents a protein and an edge represents a connection (a PPI at the molecular level) between a protein pair. The lifespan of the connection between a protein pair is stored in the corresponding edge. The construction of the timeline is a three step process: 1) constructing a static graph of the PPI (for the whole simulation Figure 3), 2) deriving a time varying graph (TVG) from the sub-graphs of the static graph (Figure 3 - a), 3) building the cluster timeline by using the TVG (Figure 3 - b). First, the static graph is constructed by inputting all existing protein pairs including time steps at which they interact. The result is a large, disconnected graph consisting of a number of sub-graphs. Each sub-graph represents a sub-set of a cluster's lifespan. We define a sub-graph by its connectivity i.e. a sub-graph is said to be connected if there is a path between every pair of nodes. A node pair might have more than one edge. Each edge represents an interaction over a given time span (from the first step). The number of edges between two proteins is mapped to how often they interacted (with splits and merges). Second, for each sub-graph, we iterate through its edges and nodes to build a TVG. With respect to time, each sub-graph is linked to its previous and subsequent sub-graph in the time series. The conjunction of two edges that have a node in common results in a new sub-graph until no conjunction is found in the TVG. Finally, the TVGs are used to visualize the cluster timeline by converting the sub-graphs into timelines. See Figure 3. It is important to note that the final layout is determined by each protein's seed position in the simulation. The positions in this view are fixed so that the user does not need to track their position overtime. The clustering timeline provides historical information on protein clustering. It is useful to understand the behavior of proteins prior to and after forming a cluster.

4. Hybrid Interactive MD Visualization

Our compact MD visualization enables the user to explore three MD interaction types in a unified context. In this section, we provide an overview of the system design followed by a description of the PLI, PPI, and clustering visualization.

4.1. Design Overview

The primary aim of our design is to enable the user to explore and analyze PLI, PPI and protein clustering interactively in the same context. The system is designed to show the history of the behavior within a molecular dynamics system. The first step in the workflow starts in the 3D view (Figure 1 - top left). The user observes the spatio-temporal clustering of proteins and selects a desired cluster in the 3D view. Then, the hybrid view (Figure 1 - top middle) can be used to obtain additional information about the system and the selected cluster. Finally, details concerning the molecular interactions and the properties of proteins and clusters can be visualized in the form of graphs (Figure 1 - top right). Our visual design includes an interactive clustering timeline view (Figure 1 - bottom) to enable the user to investigate the clusters over time. To realize this, the design should fulfill the following requirements: 1) the user must be able to investigate PLI, PPI, and protein clustering simultaneously, and 2) the three visualizations (PLI, PPI, protein clustering) must be linked to each other. To do so, we propose three resizable views, 1) a 4D overview, 2) a hybrid overview, 3) a clustering timeline view, and 4) a details-on-demand view. See Figure 1. The hybrid, 2D time-dependent view and 4D view provide the user with an overview of the system with spatial information. Both views offer a zooming feature which can be used to explore a single molecule or a cluster in both hybrid 2D and 4D. The clustering timeline enables the user to investigate any cluster at any time step in the simulation. The details-on-demand view provides information concerning the behavior of the system in the form of graphs. The four views are linked to each other in such a way the user can interact with the system using any of them. For example, if the user observes an interesting pattern in the line-chart graph he/she can steer the simulation directly to the corresponding timestep by clicking on the pattern to update the other views. The justification of the design for each view is described in the corresponding subsections.

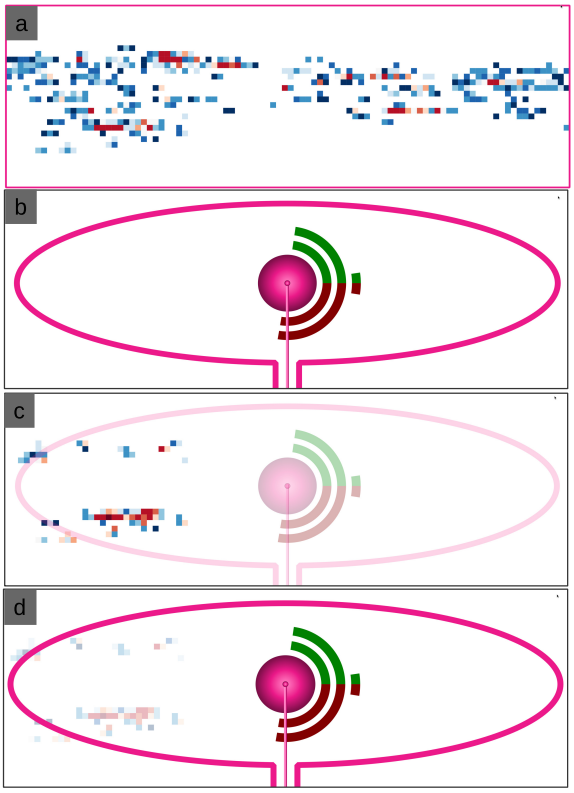

Figure 4: *Four different visualizations of the same protein at time step 1980 utilizing the hybrid view. The protein is part of a cluster. a) Visualization of the projected, tiled space of both PLI and PPI. b) Visualization of the clustering and protein behavior. c) Coupling PPI tiled space and PPI behavior with a focus on the tiled space. d) Coupling PPI tiled space and PPI behavior with a focus on clustering and protein behavior.*

4.2. Hybrid Visualization of PLI and PPI

Protein surfaces are often used to depict molecular interactions in 3D. The density of the interaction, for example, can be color-coded on the surface of a protein to indicate its accessibility. However, this approach poses three challenges. Keeping track of the motion of more than 8 objects, even at slow speed, over 4D space-time is beyond general human perception [AF07]. Hence, he first challenge is posed by the complex motion of multiple proteins in 4D space-time. The second challenge is caused by the complex geometry of protein surfaces. The third challenge is associated with the nature of 3D visualizations in general (view-dependency and occlusion). We exploit abstraction and projection to address these challenges. In the hybrid view, each protein is projected onto a fixed position in 2D space (the initial 3D seeding point) and is depicted via a glyph. See Figure 1 - top middle. Each protein is the center of its own local coordinate system in this new grid-style layout (Figure 5 - right). This new protein coordinate and abstraction visual design addresses the challenges posed by the complex motion of protein and the general 4D visualization challenges. To address the challenge posed by the complexity of the geometry we utilize the protein abstract interaction space as described in Alharbi *et al* [ACKL18]. It simplifies the problem by representing PLI and PPI (at the particle level) using a dedicated cylindrical shape. The cylinder, in 3D, is responsible for capturing and storing PLI in its tiles. In the hybrid view, the interaction space is projected onto a 2D plane to reduce occlusion. The projected space can be used to obtain an overview and identify areas of interest. It also provides a continuous map of the interaction space. See Figure 5.

Protein Abstraction In the 3D view, we use two types of abstraction, spheres and cylinders. They are used to visualize proteins and their interaction space respectively. A protein is depicted by a group of spheres. Each sphere represents a residue or particle. A protein interaction space is abstracted and visualized by a cylindrical shape (Figure 5). In the hybrid 2D view, proteins are abstracted via glyphs. The initial seeding point of each protein is used to indicate the location of the relevant abstract protein in the 2D view. The abstract protein in the 2D view is considered as a hybrid because PPI is represented in two different ways, one way using tiles, the other, using glyphs. Using two modes of visual design: a rasterized, tiled space, and a discrete glyph-based presentation makes them visually

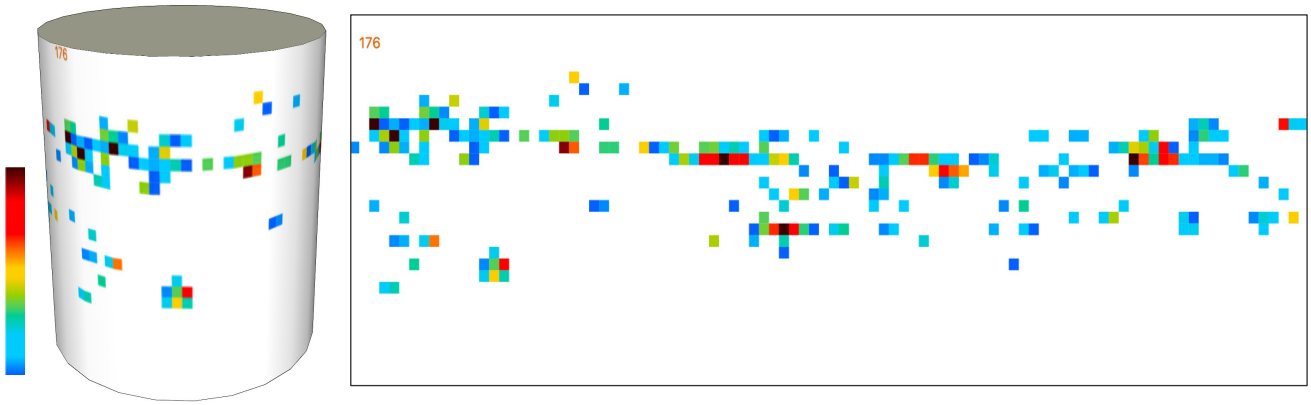

Figure 5: *The PLI and PPI tiled visual design. (left) A protein interaction space in 3D space-time. The image based on Alharbi et al [ACKL18]. (right) The corresponding projected, tiled space in the hybrid view. Color is mapped to the frequency of interaction. The image is produced with this tool.*

separable (exploiting focus and context rendering as well). However, laying the glyphs on top of the tiled space also enables the user to study their behavior in a unified, complementary context. The hybrid 2D view is used to visualize PLI and PPI both together and separately. The next two subsections discuss the visualization of the two types of interaction.

Interactive Visualization of PLI A PLI occurs when a lipid particle interacts with a protein particle. The frequency and the duration of the PLI are stored in the tiles of the protein interaction space (in the 3D space-time Figure 5 - left). In the hybrid 2D view, the PLI is visualized by a rectangle placed on the 2D abstraction of the corresponding proteins. See Figure 6. Each rectangle represents an unfolded view of the protein interaction space. The color of the tiles can be mapped to either the frequency of interaction with the relevant tile or mapped to the donor molecule type (POPE type and POPG type). See Figure 6. By mapping the tiles' color to the frequency of interaction (Figure 6 - top), the user can easily identify the accessible areas of the protein. The other representation, mapping the tiles' color to the donor type (Figure 6 - bottom), enables the user visually to understand which type has more interactions with a protein in the areas of interest. Behavior that underlies an area of interest, such as the frequency of interaction between an amino acid and the donor particles, can be revealed by clicking on the desired tile.

Interactive Visualization of PPI A PPI occurs when a protein particle interacts with a particle from another protein. PPI is a special case of molecular interaction. In the case of PLI, the two interacting molecules (I.e the lipid and protein molecule) are not studied to form clusters. The PPI, on the other hand, involves two aspects: the interaction between the particles of a protein pair, and the spatio-temporal clustering of protein pairs and their properties. Similar to PLI, the tiled visual design can be used to visualize behavior of the first aspect. The PPI tiles also can be mapped either to the frequency of PPI, or the type of the donor. The second aspect is associated with a number of properties (rotation, radius and distance to seed position of a protein). These protein attributes are visualized via

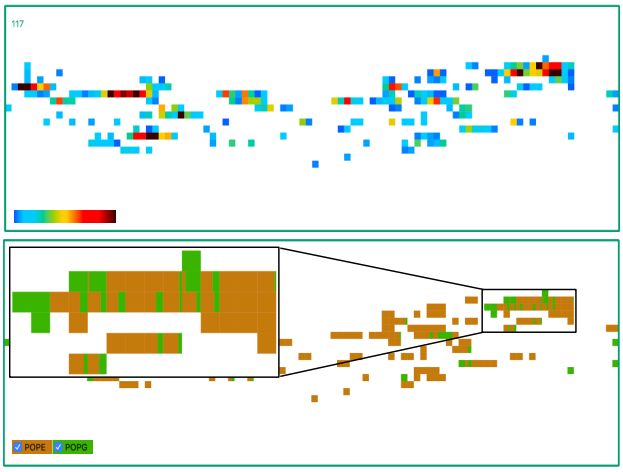

Figure 6: *2D images of the same protein focusing on PLI. (top) the tiles' color is mapped to the number of interactions. (bottom) the color of tiles is mapped to molecule types. See Section 4.2.*

glyphs. Figure 7 illustrates the visual design of the glyph properties. Arcs surrounding the glyph are used to visualize the amount of rotation. See Figure 7. Our domain expert collaborator is interested in studying the rotational behavior of proteins before and after joining a cluster. See the case studies. We use three separate arc segments to present three aspects of rotation. 1) The amount of rotation before a protein joins a cluster (inner-most arc). 2) The amount of rotation after it joins a cluster (middle arc). 3) The amount of rotation of the parent cluster (outer arc). This design enables the user to compare the collective rotation of the cluster with the proteins that belong to it. The rotation here focuses on an axis at the center of each protein. The axis is orthogonal to the membrane layer. Thus, each axis pierces (conceptually) the membrane. The changes to the protein's radius size are reflected by a circle-glyph (Figure 7 - d). The cluster membership of each protein in a cluster

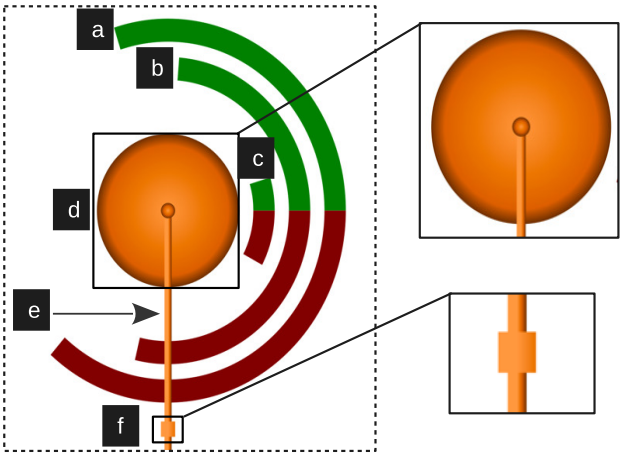

Figure 7: *Visualization of protein properties with respect to PPI and rotation. Glyphs show three rotational aspects. The green and red arcs show the positive and negative gross rotation. a) The outer arc represents the rotation of the entire cluster. b) The middle arc represents the amount of rotation after joining the cluster. c) The closest arc to the protein ellipse represents the amount of rotation of a protein before forming a cluster. d) A dynamic glyph represents the change in the radius of a protein. e) A glyph connects protein pair in a cluster. f) A dynamic indicator shows how far a protein is from its original seed point in 3D relative to the pair. See Section 4.2.*

is indicated using an outline that encompasses the proteins (Figure 7 - e). The portion of the distance between protein pair and their original seed point in 3D is visualized by an indicator placed on the connector-glyph (Figure 7 - f). A parent cluster is simply the sum of its individual proteins. It's principle components are derived by the sum of individual protein particles.

Focus and Context The initial design of the hybrid view enables the user to toggle the visualization of either tiled space or PPI behavior individually (Figure 4 - a and b). The focus and context option is introduced to enable the user to couple and decouple the visualization of tiled space and PPI behavior (Figure 4 - c and d). The user can visualize the two aspects at the same time, one in focus while the other in context. This feature is useful in studying the relation between the two aspects and understanding how PLI and PPI influence protein behavior.

4.3. Interactive Visualization of Protein Clusters

In the 4D view, a protein initially is depicted by cylindrical shape. The cylinder is color-mapped to the size of the cluster it belongs to. See Figure 1 - a. In the 3D view, the position of a protein changes over time. Therefore, a cluster can be easily identified when we find a group of cylinders that have the same color and are in close proximity. In the hybrid view, each protein is visualized via a glyph and is projected onto a fixed position in the 2D space (the initial seeding point in the simulation). To identify protein clusters in the hybrid 2D view we utilized three techniques. Firstly, a group of proteins that form a cluster are bounded in such a way that they represent one object (a cluster). Secondly, for each protein pair in a cluster, an edge is used to connect them. The edge features a dynamic indicator that shows how far a protein is from its original seed point in 3D relative to the pair. Thirdly, glyphs that represent a protein are color mapped based on the size of the cluster (Figure 1 - d).

Interactive Timeline The timeline is responsible for tracking the history of clusters. The x-axis represents the time and the y-axis represents each cluster. A simple cluster consists of two proteins, while a more complex cluster is formed by the combination of three or more proteins. The timeline shows the evolution of the clusters over time and the size of a cluster is mapped to color. See Figure 1 - bottom. The longer a cluster lifespan, the stronger the stability of the cluster. The timeline is linked to the other views in such a way the user can click on a cluster in the timeline, to move the simulation pointer to the timestep at which the cluster is formed. Also, the user can hover-over on the timeline to highlight the corresponding cluster in the hybrid view.

4.4. Detail-on-Demand

We provide the user with a set of interactive graphs in the form of line-charts. See Figure 1 - top right. The graphs are linked to the other views. The user can click on any interesting pattern in a graph to investigate them in the other views. The graphs also feature a time indicator pointing to the current simulation timestep. The graphs can be displayed by clicking on either a cluster or an individual protein. The cluster graph displays the rotation and translation of the cluster and the proteins that belong to it (Figure 1 -e and f). The protein graph displays the frequency of three types of interaction over time (Figure 1 -g). PPI, POPE type protein interaction, and POPG type protein interaction. The framework also enables the user to analyze the PLI and PPI in more detail by revealing the underlying interaction details. i.e. the user can click on areas of interest to reveal details concerning the particles or residues that are involved in the interaction. The visualization classifies the interaction based on the interacting particle pair and the frequency of interaction between them. See the accompanying video.

5. Case Studies and Domain Expert Feedback

In this section, we provide case studies and the domain expert feedback. This work has been developed in close collaboration with Dr. Matthieu Chavent. Dr. Chavent has more than 15 years of experience in the field of visualizing and analyzing molecular dynamics systems. In additional to his Ph.D. in computational biology, Dr. Chavent held the following positions: a post-doc (2009-2010) at CNRS laboratory IBPC (Paris France), post-doc 2010-2011) at CEA (Arpajon, France), and research associate position (2011-2017) at the Structural Bioinformatics & Computational Biochemistry group at the University of Oxford (UK). Since 2017, he is a CNRS researcher at the Institute of Pharmacology and Structural Biology (IPBS), Toulouse, France. He uses multiscale modeling and NMR experiments to study lipids, membrane proteins, and their interactions. He is especially interested in deciphering how lipids can modify the physical properties of membranes and the functions of

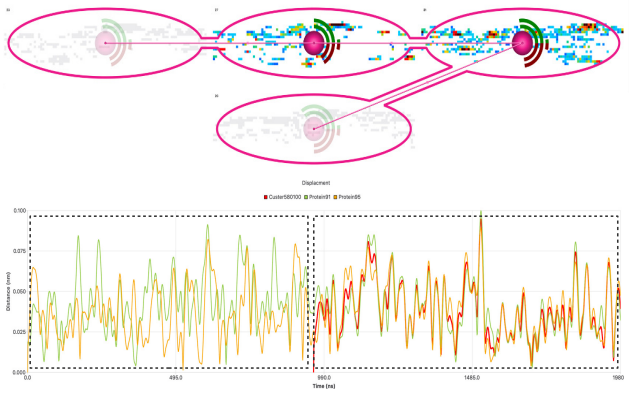

Figure 8: *Two snapshots of the visualization result illustrate the two observations. (top image) Proteins at the extremity of a cluster have interaction less than the average frequency. PPI less than the average is rendered as context. (bottom image) A cluster unifies the translation behavior of its members. (left dashed box) The translation of proteins before forming the cluster. (right dashed box) The translation of proteins after forming the cluster. The red line displays the cluster translation.*

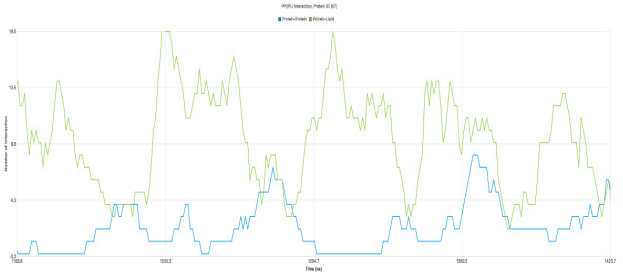

Figure 9: *Image shows an inverse relationship between the PLI and PPI. The blue line represents the frequency of PPI overtime. The green line represents the frequency of PLI over time.*

membrane proteins. All of the features were guided by and developed together with him. Since 2015, we conducted 11 meetings (60-90 minutes each), demonstrating and receiving feedback about the software during its development stages. Each meeting was video recorded. For each case hypothesis, we start with an overview of PPI and PLI depicted in the hybrid and 4D views. From there we can identify sub-set of interest and interesting behavior.

Case study 1 hypothesis: If a protein is interacting with lipids then it is interacting less with other proteins and vice versa. This hypothesis can be tested by the hybrid view and the details-on-demand view. We utilize the protein interaction graph to examine the hypothesis. We invoke the graph by selecting a protein that is involved in a cluster. See Figure 1 (g). The graph displays the PLI and PPI of the selected protein. From the graph, it can often be seen that the peak of PPI is associated with the lowest point of PLI. The visualization using the detailed interaction line-chart view supports the hypothesis. See Figure 9. Furthermore, by examining the hybrid view we observe that the proteins at the extremity exhibit less than average interaction. See Figure 8 (top image). However, in exceptional cases, proteins at the extremity of a cluster might exhibit greater than average interaction. For example, consider two clusters that are well-established at the beginning of the simulation. Then afterward, a third cluster is formed. If the third cluster forms a new, larger cluster by linking the two clusters together, it is likely to have extreme proteins exhibiting greater than average interaction. In order to understand this further, we need to investigate the history of the cluster and the proteins in question with respect to PPI. The clustering timeline view plays a significant role in understanding the behavior of these two clusters and helps us to investigate this issue. See the accompanying video for more details.

Case study 2 Hypothesis: The larger the size of a cluster, the less the rotation and translation. We examined clusters of varying sizes and use the details-on-demand feature to compare the translation and rotation of each. From the line-chart, we found that the amount of rotation of individual proteins decreases immediately after joining a cluster. See Figure 10. Also, we find that the size of a cluster plays a significant role in reducing the amount of rotation, which provides strong evidence supporting the rotation hypothesis. We do not witness a considerable change in the translation speed. Yet, we observe that a cluster unifies the translation behavior of its members in such a way the proteins behave similarly after entering a cluster. See Figure 8 (bottom image).

Domain Expert Feedback:
The following feedback was produced directly by the domain expert: “Visualizing easily both PLI and PPI is especially novel as, currently, no other molecular viewer are designed to perform this type of analysis concurrently. Going back and forth in between the different views help to better understand the formation of the protein clusters and if lipids surrounding the proteins may modulate such cluster formation. The 2D tiled space is especially useful to quickly compare in a large systems interactions in between proteins and in between protein and lipids and see how these interactions evolve in function of the time."

6. Conclusions and Future Work

In this work, we propose a novel PLI and PPI visualization framework. The framework utilizes four visual designs that enable the user to study a time-dependent membrane simulation. The framework employs abstraction and space projection to address a number of visualization challenges. We also propose a novel hybrid view to enable the user to study PLI and PPI, and the behavior of the PPI and clusters. A details-on-demand view is used to provide the user with desired information about proteins, clusters, and their behavior. In future work, we will use a longer membrane simulation consisting of different types of proteins. The visual analysis will be our focus in the future work including the correlation between lipid and protein types.

ACKNOWLEDGEMENTS

This work was partially funded by the Ministry of Education of Saudi Arabia, the Saudi Cultural Bureau in London, the Department

Figure 10: *Three line-charts illustrate the change in the rotation of a cluster during its evolution. The amount of cluster rotation is displayed by the thick, red lines. (top) A line-chart displays the rotation of the cluster and its members (the first formation of the cluster). The cluster's accumulative rotation ranges between 0.70 to -0.75 radian. (middle) The second formation of the cluster (three proteins). The line-chart shows a small decrease in the counterclockwise rotation. (bottom) The cluster rotation after the cluster formed by four proteins. The line-chart shows a decrease by approximately 0.50% in both the clockwise and counterclockwise rotation.*

of Computer Science at Swansea University, and the DFG as part of project PROLINT. Finally, we would like to thank Rhodri Fabbro for proof-reading the paper.

References

[AAM*17] ALHARBI N., ALHARBI M., MARTINEZ X., KRONE M., ROSE A., BAADEN M., LARAMEE R. S., CHAVENT M.: Molecular visualization of computational biology data: A survey of surveys. *Eurovis short papers* (2017), 133–137. 3

[ACKL18] ALHARBI N., CHAVENT M., KRONE M., LARAMEE R. S.: VAPLI: Novel visual abstraction for protein-lipid interactions. *IEEE SciVis Short Papers* (2018), 133–137. 2, 3, 4, 5, 6, 7

[AF07] ALVAREZ G. A., FRANCONERI S. L.: How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of vision 7*, 13 (2007), 14–14. 6

[CDR*18] CHAVENT M., DUNCAN A. L., RASSAM P., BIRKHOLZ O., HÉLIE J., REDDY T., BELIAEV D., HAMBLY B., PIEHLER J., KLEANTHOUS C., ET AL.: How nanoscale protein interactions determine the mesoscale dynamic organisation of bacterial outer membrane proteins. *Nature communications 9* (2018). 2, 3, 4, 5

[CRG*14] CHAVENT M., REDDY T., GOOSE J., DAHL A. C. E., STONE J. E., JOBARD B., SANSOM M. S.: Methodologies for the analysis of instantaneous lipid diffusion in md simulations of large membrane systems. *Faraday discussions 169* (2014), 455–475. 4

[CZA*16] CROISSANT J. G., ZHANG D., ALSAIARI S., LU J., DENG L., TAMANOI F., ALMALIK A. M., ZINK J. I., KHASHAB N. M.: Protein-gold clusters-capped mesoporous silica nanoparticles for high drug loading, autonomous gemcitabine/doxorubicin co-delivery, and in-vivo tumor imaging. *Journal of Controlled Release 229* (2016), 183–191. 4

[dHCKMK13] DE HERAS CIECHOMSKI P., KLANN M., MANGE R., KOEPPL H.: From biochemical reaction networks to 3d dynamics in the cell: The zigcell3d modeling, simulation and visualisation framework. 41–48. 3, 4

[Eri09] ERICKSON H. P.: Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological procedures online 11*, 1 (2009), 32. 5

[FBG*18] FURMANOVÁ K., BYŠKA J., GRÖLLER E. M., VIOLA I., PALEČEK J. J., KOZLÍKOVÁ B.: Cozoid: contact zone identifier for visual analysis of protein-protein interactions. *BMC bioinformatics 19*, 1 (2018), 125. 3, 4

[FKRE09] FALK M., KLANN M., REUSS M., ERTL T.: Visualization of signal transduction processes in the crowded environment of the cell. 169–176. 3

[FMB04] FINN R. D., MARSHALL M., BATEMAN A.: i Pfam: visualization of protein–protein interactions in pdb at domain and amino acid resolutions. *Bioinformatics 21*, 3 (2004), 410–412. 3

[FS01] FRENKEL D., SMIT B.: *Understanding molecular simulation: from algorithms to applications*, vol. 1. Elsevier, 2001. 2

[GOB*10] GEHLENBORG N., O'DONOGHUE S. I., BALIGA N. S., GOESMANN A., HIBBS M. A., KITANO H., KOHLBACHER O., NEUWEGER H., SCHNEIDER R., TENENBAUM D., ET AL.: Visualization of omics data for systems biology. *Nature methods 7*, 3s (2010), S56. 3

[HDS96] HUMPHREY W., DALKE A., SCHULTEN K.: VMD: visual molecular dynamics. *Journal of Molecular Graphics 14*, 1 (1996), 33–38. 4

[HEG*17] HERMOSILLA P., ESTRADA J., GUALLAR V., ROPINSKI T., VINACUA A., VÁZQUEZ P.-P.: Physics-based visual characterization of molecular interaction forces. *IEEE transactions on visualization and computer graphics 23*, 1 (2017), 731–740. 3

[Hun07] HUNTER J. D.: Matplotlib: A 2d graphics environment. *Computing in science & engineering 9*, 3 (2007), 90. 4

[KC12] KOES D. R., CAMACHO C. J.: Pocketquery: protein–protein interaction inhibitor starting points from protein–protein interaction structure. *Nucleic acids research 40*, W1 (2012), W387–W392. 3

[KC13] KHAZANOV N. A., CARLSON H. A.: Exploring the composition of protein-ligand binding sites on a large scale. *PLoS computational biology 9*, 11 (2013), e1003321. 3

[KKF*17] KOZLÍKOVÁ B., KRONE M., FALK M., LINDOW N., BAADEN M., BAUM D., VIOLA I., PARULEK J., HEGE H.-C.: Visualization of biomolecular structures: State of the art revisited. *Computer Graphics Forum 36*, 8 (2017), 178–204. 3

[KKL*16] KRONE M., KOZLÍKOVÁ B., LINDOW N., BAADEN M., BAUM D., PARULEK J., HEGE H.-C., VIOLA I.: Visual analysis of biomolecular cavities: State of the art. *Computer Graphics Forum 35*, 3 (2016), 527–551. URL: http://dx.doi.org/10.1111/cgf.12928, doi:10.1111/cgf.12928. 3

[KL80] KLEMA V., LAUB A.: The singular value decomposition: Its computation and some applications. *IEEE Transactions on Automatic Control 25*, 2 (1980), 164–176. 5

[LKWS13] LANGDON B. B., KASTANTIN M., WALDER R., SCHWARTZ D. K.: Interfacial protein–protein associations. *Biomacromolecules 15*, 1 (2013), 66–74. 4

[LMPSV14] LE MUZIC M., PARULEK J., STAVRUM A.-K., VIOLA I.: Illustrative visualization of molecular reactions using omniscient intelligence and passive agents. In *Computer Graphics Forum* (2014), vol. 33, pp. 141–150. 3

[LMWPV15] LE MUZIC M., WALDNER M., PARULEK J., VIOLA I.: Illustrative timelapse: A technique for illustrative visualization of particle-based simulations. In *2015 IEEE Pacific Visualization Symposium (PacificVis)* (2015), IEEE, pp. 247–254. 3, 4

[MGP*15] MOLINA M. L., GIUDICI A. M., POVEDA J. A., FERNÁNDEZ-BALLESTER G., MONTOYA E., RENART M. L., FERNÁNDEZ A. M., ENCINAR J. A., RIQUELME G., MORALES A., ET AL.: Competing lipid-protein and protein-protein interactions determine clustering and gating patterns in the potassium channel from streptomyces lividans (kcsa). *Journal of Biological Chemistry 290*, 42 (2015), 25745–25755. 2

[MKK*18] MIAO H., KLEIN T., KOŘIL D., MINDEK P., SCHATZ K., GRÖLLER M. E., KOZLÍKOVÁ B., ISENBERG T., VIOLA I.: Multiscale molecular visualization. *Journal of molecular biology 431*, 6 (2018), 1049–1070. 3

[MT13] MARRINK S. J., TIELEMAN D. P.: Perspective on the martini model. *Chemical Society Reviews 42*, 16 (2013), 6801–6822. 4

[OGF*10] O'DONOGHUE S. I., GOODSELL D. S., FRANGAKIS A. S., JOSSINET F., LASKOWSKI R. A., NILGES M., SAIBIL H. R., SCHAFFERHANS A., WADE R. C., WESTHOF E., ET AL.: Visualization of macromolecular structures. *Nature Methods 7* (2010), S42–S55. 3

[ORR*10] OWEN D. M., RENTERO C., ROSSY J., MAGENAU A., WILLIAMSON D., RODRIGUEZ M., GAUS K.: Palm imaging and cluster analysis of protein heterogeneity at the cell surface. *Journal of biophotonics 3*, 7 (2010), 446–454. 4

[OWM*12] OWEN D. M., WILLIAMSON D., MAGENAU A., ROSSY J., GAUS K.: Optical techniques for imaging membrane domains in live cells (live-cell palm of protein clustering). In *Methods in enzymology*, vol. 504. Elsevier, 2012, pp. 221–235. 4

[PBI*15] PRESCHER J., BAUMGÄRTEL V., IVANCHENKO S., TORRANO A. A., BRÄUCHLE C., MÜLLER B., LAMB D. C.: Super-resolution imaging of escrt-proteins at hiv-1 assembly sites. *PLoS pathogens 11*, 2 (2015), e1004677. 4

[SVGR16] SKANBERG R., VÁZQUEZ P.-P., GUALLAR V., ROPINSKI T.: Real-time molecular visualization supporting diffuse interreflections and ambient occlusion. *IEEE Trans. Vis. Comput. Graph 22*, 1 (2016), 718–727. 3

[VHG*18] VÁZQUEZ P., HERMOSILLA P., GUALLAR V., ESTRADA J., VINACUA A.: Visual analysis of protein-ligand interactions. *Computer Graphics Forum 37*, 3 (2018), 391–402. 3