GazeMolVR: Sharing Eye-Gaze Cues in a Collaborative VR Environment for Molecular Visualization

Rajkumar Darbar
Laboratoire de Biochimie Théorique, UPR 9080 CNRS
Paris, France
rajdarbar.r@gmail.com

Hubert Santuz
Laboratoire de Biochimie Théorique, UPR 9080 CNRS
Paris, France
santuz@ibpc.fr

Antoine Taly
Laboratoire de Biochimie Théorique, UPR 9080 CNRS
Paris, France
taly@ibpc.fr

Marc Baaden
Laboratoire de Biochimie Théorique, UPR 9080 CNRS
Paris, France
baaden@smplinux.de

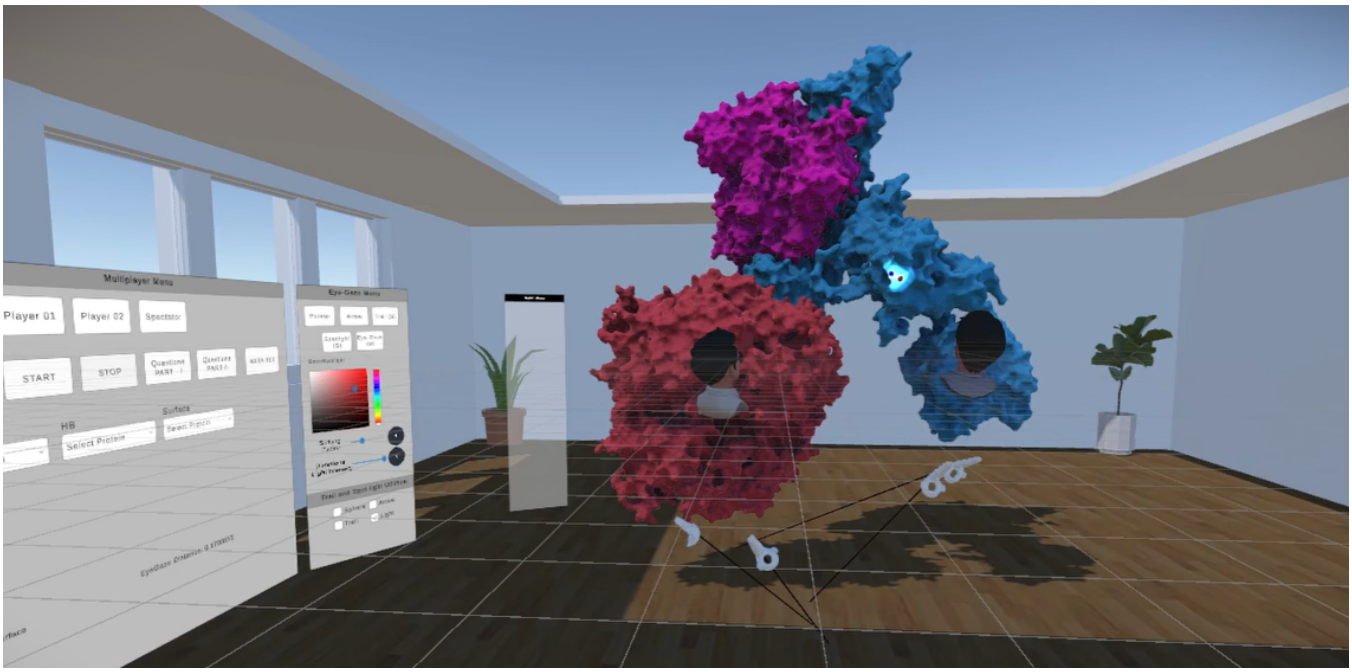

Figure 1: Our GazeMolVR setup where a remote pair, depicted as avatars, engages in real-time discussion about a protein represented in a surface model. They utilize GazeSpotlight to share their mutual visual focus.

Abstract

Virtual Reality (VR) has significantly enhanced the visualization of molecular structures, offering an intuitive and immersive experience. However, immersive collaborative virtual environments, despite their benefits that can come close to physical co-location, often lack crucial non-verbal communication cues such as gaze awareness, essential for enriching face-to-face collaboration. This research introduces GazeMolVR, a tool based on the UnityMol software that enables a remote pair to collaboratively explore and discuss a protein's structure and function within a VR environment. It incorporates bi-directional eye-gaze cues through four distinct representations—GazePoint, GazeArrow, GazeSpotlight, and GazeTrail—to enhance mutual awareness of visual focus during discussions. We conducted two user studies to evaluate GazeMolVR. The first aimed to identify the most suitable gaze visualization for discussing proteins depicted in cartoon, ball-and-stick, and surface models. The second compared the effects of bi-directional gaze sharing during collaborative discussions to a scenario without gaze sharing, especially in the field of structural biology. Study results showed a preference for GazeTrail with cartoon and ball-and-stick models, and GazeSpotlight for the surface model. Additionally, sharing bi-directional eye-gaze cues significantly enhanced collaborative discussions compared to not using gaze cues.

Author Preprint, December 2024, Location Withheld

CCS Concepts

• **Human-centered computing → User studies**.

Keywords

Molecular Visualization, Virtual Reality (VR), Augmented Reality (AR), Remote Collaboration, Eye-Gaze, Scientific Data Visualization.

ACM Reference Format:

Rajkumar Darbar, Hubert Santuz, Antoine Taly, and Marc Baaden. 2024. GazeMolVR: Sharing Eye-Gaze Cues in a Collaborative VR Environment for Molecular Visualization. In *Open-Source Author Version, December 2024.* ACM, New York, NY, USA, 17 pages.

1 Introduction

Interactive visualization of nanoscale molecular objects is pivotal in the fields of molecular and structural biology [58, 68]. It assists domain experts in tasks ranging from structural analysis to interactive drug design. Several widely used desktop-based tools for molecular visualization exist, including VMD [48], PyMOL [21], Chimera [76], UnityMol [67], and JSmol [44, 91]. However, viewing 3D biomolecular structures on a 2D display often lacks spatial perception regarding depth, distance, and scale. This issue is further compounded when interaction with molecular models is confined to mouse and keyboard inputs, not only making the process less intuitive but also being limited by the desktop flat screen's restricted field of view (FoV), which hinders a comprehensive understanding of complex molecular landscapes. To overcome these challenges, researchers have been exploring advanced display technologies for many decades. These technologies include CAVEs [14] and Stereoscopic 3D Display Walls [93] that were targeted for molecular visualization since their inception. Although these systems offer a co-located multi-user immersive experience, they are expensive and require specialized hardware and space for installation.

Recent advancements in virtual reality (VR) technology - including improvements in portability, computational power, field of view (FoV), resolution, tracking, and reduced weight - have revolutionized the visualization of molecular structures. As a result, several molecular visualization tools for VR head-mounted displays (HMDs) were developed [41, 50, 57, 64, 67], enabling structural biologists to immerse themselves in a virtual environment, where direct interaction with biomolecules is not only possible but also intuitive. This immersive approach, facilitated by modern VR HMDs, significantly enhances comprehension of complex molecular structures and their biological properties by providing a unique, three-dimensional perspective. Furthermore, these tools foster remote collaboration, allowing experts across different disciplines to engage in meaningful discussions about complex molecular models within a shared virtual space.

In immersive collaborative virtual environments, which offer benefits comparable to physical co-location [38, 79], certain nonverbal communication cues, such as gaze awareness, are often lacking. These cues are fundamental to enriching face-to-face collaboration by aiding in the coordination of attention. Thanks to advancements in eye-tracking technology, researchers can now share eye-gaze cues in both co-located and remote task spaces through immersive AR/VR displays [5, 78]. Previous studies have introduced various gaze visualization techniques to depict eye movements, including saccades, fixations, and joint gaze. For example, Jing et al. [52] evaluated three bi-directional eye-gaze visualizations—Cursor Donut (CD), Laser Eye (LE), and Trail Path (TP)—in co-located tasks such as visual searching, matching, and puzzle-solving. They found that all three significantly enhanced user engagement compared to a no gaze-cue condition, with Laser Eye (LE) being the most preferred. In another work [53], Jing et al. developed a 360° panoramic mixed reality remote collaboration system for physical tasks, sharing various gaze behavior visualizations between a local AR user and a remote VR collaborator. In the educational domain, Rahman et al. [80] displayed student eye gaze in a teacher's VR environment to identify distracted students. They evaluated six gaze visualizations—Gaze Ring, Gaze Disk, Gaze Arrow, Gaze Trail, Gaze Trail with Arrow, and Gaze Heatmap—finding that a short particle trail representing eye trajectory was promising, while 3D heatmaps were problematic for short-term visualization. Delgado and Ruiz [22] examined virtual assembly tasks with two co-located collaborators using AR HMDs. They evaluated three gaze visualizations—Constant Ray, Gaze Trigger, and Gaze Hover—finding no clear preference among them. Furthermore, Ichino et al. [49] investigated how gaze visualizations in virtual spaces facilitate the initiation of informal communication among multiple co-located users. They evaluated three types of gaze visualizations—Arrow, Bubble, and Miniavatar—for both one-sided and joint gaze behaviors. Their findings indicated that Bubbles were effective for one-sided gaze, while all three were suitable for joint gaze.

It is evident from previous research that no single eye-gaze visualization technique is suitable for all types of collaborative tasks (see Table 1); the appropriate representation truly depends on the nature of the task itself [16–18, 66]. Additionally, the way collaborators coordinate and focus their attention during a task is significantly influenced by the design of gaze visualizations. To our knowledge, no research has specifically evaluated the impact of sharing mutual eye-gaze cues during a collaborative task where a remote pair simultaneously views, manipulates, and discusses the structural and functional aspects of a biomolecular entity, such as a protein, and its complex interactions within a VR environment. Notably, a 3D protein structure is a dense three-dimensional information space. Collaboratively exploring biomolecular structures and their interactions in a VR environment significantly differs from the collaborative tasks explored in previous studies, such as visual searching and puzzle solving [52, 53], initiation of informal communication [49], virtual assembly [22], and identifying distracted students [80]. For instance, in a virtual reality environment, collaborators might begin by examining the entire protein structure to understand its overall shape and topology. They may then zoom in on specific regions, such as active sites or binding pockets, for closer inspection. As the discussion progresses, they could delve into atomic-level details, examining individual atoms and interactions like hydrogen bonds and van der Waals forces. Additionally, they might switch between visualization modes—using cartoon representations to highlight secondary structures, ball-and-stick models for atomic interactions, surface models to illustrate the exterior, and electrostatic potential maps to discuss charge distributions.

This research gap underscores the need to study mutual eye-gaze sharing in the context of complex biomolecular discussions.

It raises several interesting questions: How should eye-gaze cues be represented during interactive molecular discussions? Is there a preference for gaze visualizations depending on the protein representations? Does sharing gaze cues enhance collaborative discussions when analyzing specific aspects of protein structures, such as active sites or binding interactions?

In this work, we introduce GazeMolVR, based on the UnityMol framework [67], to enable remote pairs to collaboratively explore and discuss protein structures and functions within a VR environment, while simultaneously sharing bi-directional eye-gaze cues through four distinct gaze representations—GazePoint, GazeArrow, GazeSpotlight, and GazeTrail—to enhance mutual awareness of visual focus, as shown in Figure 1. It is important to note that while these gaze visualizations are not entirely novel, they have been specifically adapted for molecular graphics, drawing on established techniques from the literature [49, 52, 53, 80, 90]. Given that proteins can be depicted using various models to illustrate their complex structures and functions, our work specifically focuses on the three most widely used representations for collaborative discussion: cartoon, ball-and-stick, and surface models [65, 85], ensuring relevance and applicability to common scientific practices. To evaluate GazeMolVR, we conducted two user studies. The first user study aimed to determine the most suitable gaze visualization for discussing proteins in cartoon, ball-and-stick, and surface representations, respectively. In our second user study, we compared the effects of bi-directional gaze sharing with no gaze sharing during collaborative discussion.

The main contributions of this paper are:

- Introducing the sharing of eye-gaze cues in a collaborative virtual reality environment for molecular visualization.
- Presenting the results of two formal user studies. The first study explores the interaction between protein representations and gaze visualizations, whereas the second user study assesses whether gaze cues are beneficial during collaborative discussions about a protein.

2 Related Work

Our work draws inspiration from the literature on molecular visualization using AR/VR HMDs and the sharing of eye-gaze cues in collaborative tasks, briefly reviewed in this section.

2.1 Molecular Visualization using AR/VR HMDs

Over the last decade, virtual reality has garnered significant attention for its interactive and immersive capabilities in visualizing biomolecular structures. To date, several VR systems have been developed, which are briefly summarized herein. For a comprehensive overview of this evolving field, interested readers are advised to refer to the recent review article by Kut'ák et al. [61].

One of the pioneering VR applications, Molecular Rift [74], emphasizes the manipulation of biomolecules in 3D space using hand tracking instead of VR controllers. However, when users interact with molecules in VR using their hands, they do not receive any haptic feedback. Moreover, it remains unclear which types of feedback would be realistic or intuitively useful for manipulating these objects. Addressing this gap, Roebuck Williams et al. explored the potential of pseudo-haptic feedback in molecular simulations [82].

Specialized VR tools have also emerged to tackle specific challenges in molecular visualization. For instance, BioVR [97] assists researchers in integrating and visualizing DNA/RNA sequences alongside their protein structures. Similarly, Kut'ák et al. et al. [60] developed Vivern, a tool specifically for modeling and examining DNA nanostructures in virtual reality, employing abstract visual representations and varied color schemes to navigate the spatial complexity of DNA origami structures. Additionally, Laureanti et al. [63] enhanced the visualization of electrostatic potential fields at specific protein sites by integrating Adaptive Poisson-Boltzmann Solver (APBS) tool [34] with UnityMol's VR interface [25]. These systems provide powerful tools for immersive molecular visualization, but they often require specific hardware setups and software installations. In contrast, ProteinVR [8] and VRmol [95] are web-based implementations accessible on a broad range of devices without requiring third-party programs or plugins, offering users the convenience of easy access.

While the previously mentioned tools focus on the visualization and exploration of static biomolecular structures, recent advancements in VR have pushed the boundaries further by integrating real-time molecular dynamics simulations. In this context, O'Connor et al. introduced Narupa, enabling the interactive visualization and manipulation of molecular dynamics with atomic-level precision, a significant leap from static structures or prerecorded trajectories [75]. Deeks et al. [20] combined interactive molecular dynamics in VR with free energy (FE) calculations to study protein-ligand interactions at the molecular level. Similarly, Juárez-Jiménez et al. developed a framework for ensemble molecular dynamics simulation in VR, allowing for the real-time exploration of protein conformational changes over millisecond timescales [56].

Visualization of scientific data is crucial not only for scientific discovery but also for communicating science to the general audience. Bearing this in mind, researchers explored interactive molecular illustrations in virtual reality, such as CellPAINT-VR [9, 37], immersive guided tours through dense molecular environments [3], journeys to the center of the cell [55], and LifeBrush [19]. Moreover, Brůža et al. [7] introduced the VRdeo tool, which enables tutors to prepare and record a VR scene with educational content that students can later enter and explore interactively.

Beyond virtual reality, researchers also investigated the potential of augmented reality (AR) for molecular visualization [47, 81, 99]. Müller et al. [70] evaluated the performance of various methods for rendering the space-filling representation of molecules using HoloLens [12]. Noizet et al. [73] augmented 3D printed molecules with additional visual representations through the use of a HoloLens device. A user study demonstrated that their setup significantly facilitated co-located collaboration in an intuitive manner, as users remained fully aware of their surroundings and could communicate with others naturally. Although AR HMDs are excellent for co-located collaborations, their current limitations mainly include the field of view and graphics performance.

Most of the work described above often lacks robust support for effective remote collaboration, with the exception of Narupa [50, 75], which enables multiple users to manipulate molecular dynamics simultaneously in real-time. Other works, such as AMMP-Vis [10], the multi-user VR version of ChimeraX [41, 42] and UnityMol [67],

MolecularWebXR [13], and Nanome [57], have also attempted to address this significant issue.

While AR/VR systems for collaborative molecular visualization exist, none have incorporated eye-gaze cues into immersive molecular discussions. GazeMolVR, presented in this paper, fills this gap by integrating mutual eye-gaze sharing—a key non-verbal communication cue—enhancing collaboration and adding a new dimension to molecular interaction and understanding.

2.2 Sharing Eye-Gaze Cues in Collaborative Task Environments

Researchers explored sharing eye-gaze cues in both co-located and remote task spaces using traditional 2D screens (e.g., desktop, projector, large public display) and immersive technologies (AR/VR HMDs). Table 1 provides a summary of eye-gaze cue visualizations in various collaborative task environments.

In traditional collaborative settings with 2D displays, mutual gaze awareness was used for various purposes: enhancing co-located collaborative search tasks on a large shared display [98], inferring remote players' intentions in competitive strategic games [71, 72], increasing social presence in an online cooperative game [69], commanding and controlling Unmanned Aerial Vehicles (UAV) [4], improving communication in pair programming [15] and writing [62], helping students achieve higher learning gains in remote teaching [83, 84, 86, 96], and enhancing physical task performance [2, 46]. In these works, the authors proposed several gaze visualization techniques (e.g., dot, cursor, spotlight, heatmap, scan path, trail, etc.) to represent different characteristics of eye movements (i.e., fixations, saccades, and joint gaze) on the collaborator's screen. The design of gaze visualizations and the attributes of the task significantly influenced how pairs coordinated their actions and allocated their attention effectively [16, 17, 66].

Compared to traditional 2D displays, AR/VR HMDs enable room-scale collaboration and offer unique capabilities for conveying spatial information. With the recent availability of AR/VR headsets with eye-tracking capabilities, there has been a growing number of studies exploring different eye-gaze visualizations to enhance collaboration in various tasks, including visual searching [52, 53], initiation of informal communication [49], virtual assembly [22], and identifying distracted students [80]. Most of these prior works share eye-gaze in a unidirectional manner (from a local to a remote user or vice versa), except for the works done by Jing et al. [51–54], where the authors enabled bi-directional sharing of gaze cues in co-located and remote mixed reality collaboration tasks, allowing participants to see both their own and their partner's gaze points. Eye-gaze cues were utilized in mixed reality collaborative settings to facilitate the communication of intentions, act as pointers for deictic references, and enhance the sense of co-presence among collaborators [5, 43, 78, 89].

Although prior work on mixed reality collaboration has explored sharing eye-gaze cues for various tasks, no studies have investigated the influence of exchanging eye-gaze cues in the context of immersive molecular discussions in virtual reality. Conventional tasks, such as visual search or virtual object assembly in mixed reality environments, often involve objects with natural, easily identifiable reference points, such as a wall or table. In contrast, proteins and other biomolecular structures have complex three-dimensional geometries that lack intuitive features, making it difficult for collaborators to establish common visual references. Additionally, molecular discussions are grounded in scientific inquiry, requiring the interpretation of the biological significance of these structures, which differs from traditional collaborative tasks. These challenges make immersive molecular discussion a unique domain. This study extends previous research by introducing GazeMolVR, which uses bi-directional eye-gaze cues through four distinct representations to enhance visual focus during collaborative discussions of protein structure and function.

3 System Design

GazeMolVR is a symmetric virtual reality system designed for remote collaboration between dyads of structural biologists, enabling them to discuss protein structures, including their folding patterns, functional sites, secondary structures, active domains, molecular interactions, and their biological functions, without being physically co-located. In this work, we considered the three most widely used protein representations: cartoon, ball-and-stick, and surface, for collaborative discussion (see 3.1 for details).

Our system leverages built-in eye-trackers in VR headsets to share gaze cues bi-directionally, allowing participants to observe both their own and their partner's gaze simultaneously. Users can represent their eye-gaze through four distinct gaze visualizations: GazePoint, GazeArrow, GazeSpotlight, and GazeTrail (see 3.2 for details). By being mutually aware of each other's visual focus, collaborators can synchronously and efficiently investigate complex protein structures, thereby achieving a unified understanding of diverse biological properties through collaborative interpretation. Previous studies showed that a higher mutual awareness of visual focus correlates with higher measures of perceived collaboration quality and visual coordination [15, 83]. Furthermore, when users see their own gaze, they experience confidence and certainty that their gaze location is being accurately communicated [52–54].

3.1 Protein Representations

Protein structures are complex, consisting of thousands or tens of thousands of atoms, sometimes even an order of magnitude more, bonded together in specific arrangements. Given their nanoscopic scale, researchers developed a variety of molecular graphics methods over the years to visualize protein structures, making it easier to study their properties [58, 68]. Each protein representation serves a unique purpose. In GazeMolVR, we considered three such representations: cartoon, ball-and-stick, and surface models [65, 85]. These three representations are considered the most common primarily due to their balance of simplicity, clarity, and the breadth of information they provide for a wide range of scientific tasks in structural biology. They each cover a broad set of needs—structural overview, detailed analysis, and interaction focus—making them go-to choices in most contexts, compared to other more specialized representations, such as space-filling, wireframe, electrostatic potential map, or density map, which are used for specific tasks. A brief overview of these three representations is given below.

Cartoon: It describes a protein's secondary structures, such as alpha-helices, represented as coils or spirals, and beta-sheets,

Table 1: Summary of sharing eye-gaze cues in collaborative tasks.

Author(s)	Task Type	Setup	Platforms	Eye-Gaze Visualizations	Gaze Direction	Preferred Visualizations
Zhang et al. [98]	Collaborative visual search	Co-located with 2 collaborators	Large shared display	Cursor, Trajectory, Highlight, and Spotlight	Uni-directional	Highlight and Spotlight conditions
Atweh et al. [4]	UAV search and rescue command-and-control tasks	Co-located with 2 collaborators	Desktop - Desktop	Fixation Dot and Fixation Trail	Uni-directional	Fixation Trail
Newn et al. [72]	Competitive strategy games	Remote with 2 collaborators	Desktop - Desktop	Dot, Cursor, Spotlight, Fixation, Scanpath, Fixation Trail, Heatmap, Convex Hull, and Bee Swarm	Uni-directional	Heatmap
D'Angelo and Begel [15]	Pair programming	Remote with 2 collaborators	Desktop - Desktop	A vertical bar on the left margin	Uni-directional	Proposed gaze visualization than no gaze condition
Kütt et al. [62]	Collaborative writing	Remote with 2 collaborators	Desktop - Desktop	Circle, Highlighted Block, Vertical Bar, and Gradient Visualization	Uni-directional	Gradient Visualization
Akkil et al. [2]	Collaborative physical tasks	Remote with 2 collaborators	Desktop - Projector	Spotlight	Uni-directional	Gaze Spotlight than camera-based interface
Rahman et al. [80]	Identifying distracted students in educational VR	Pre-recorded VR scene with 5 Students and 1 Teacher	VR HMD	Gaze Ring, Gaze Disk, Gaze Arrow, Gaze Trail, Gaze Trail with Arrow, and Gaze Heatmap	Uni-directional	Gaze Trail
Delgado and Ruiz [22]	Virtual assembly tasks	Co-located with 2 collaborators	AR HMD - AR HMD	Constant Ray, Gaze Trigger, and Gaze Hover	Uni-directional	No preference for any gaze condition
Ichino et al. [49]	Initiation of informal communication in 3D virtual spaces	Co-located with 2 collaborators	VR HMD - VR HMD	Arrow, Bubble, and Miniavatar with one-sided and joint gaze behaviors	Uni-directional	Bubbles for one-sided gaze; all three for joint gaze.
Jing et al. [52]	Visual searching and matching of pictographic symbols and puzzle solving	Co-located with 2 collaborators	AR HMD - AR HMD	Cursor Donut, Laser Eye, and Trail Path augmented with gaze behavioural states	Bi-directional	Laser Eye condition
Jing et al. [53], Jing et al. [54]	Visual-searching of abstract symbols in a physical workspace	Remote with 2 collaborators	AR HMD - VR HMD	Gaze Cursor and Gaze Ray augmented with gaze behavioural states	Bi-directional	Proposed gaze visualization than no gaze condition
GazeMolVR	Discussing molecular structures and functions in VR	Remote with 2 collaborators	VR HMD - VR HMD	GazePoint, GazeArrow, GazeSpotlight, and GazeTrail	Bi-directional	GazeTrail for cartoon and ball-and-stick models; GazeSpotlight for surface model.

depicted as arrows or flat strands (see Figure 2). This method offers a simplified overview of the protein's overall folding pattern and structural motifs. It effectively highlights the protein's backbone and three-dimensional shape, enhancing understanding of its structure-function relationship.

Ball-and-Stick: In this model, atoms are represented as balls, and covalent bonds between these atoms are represented as sticks connecting the balls (see Figure 3). It provides detailed insights into molecular geometry and atom-level connectivity, offering a visually complex but information-rich representation. The size of the balls representing the atoms is scaled in order to ensure the visibility of the bonds (sticks) and prevent the model from becoming overcrowded, particularly with larger molecules. This representation is preferred for in-depth molecular studies, including chemical reactions, bonding arrangements, and precise spatial relationships.

Surface: This model highlights the protein's exterior surface, emphasizing features such as grooves and pockets vital for understanding molecular interactions (see Figure 4). By color-coding the surface to denote properties like hydrophobicity and electrostatic potential, it aids in identifying potential binding sites for various molecules, including ligands, ions, and small proteins. This is key for insights into protein-ligand interactions and docking processes.

It's important to note that molecular visualization software effectively scales protein sizes from nanometers to a visible scale. This adjustment allows for the detailed examination and interactive analysis of structures that would otherwise be too small to see.

3.2 Eye-Gaze Visualizations

This section outlines the techniques employed to visualize eye-gaze cues in GazeMolVR, along with the design principles that informed their development.

3.2.1 Design Requirements for Visualizing Gaze Cues. Drawing on prior research [52, 53, 90] and user feedback gathered through iterative prototyping of GazeMolVR, we considered the following design requirements for visualizing eye-gaze cues during molecular discussions in a collaborative virtual environment.

Design Requirement 1: Subtlety and Precision. Given that gaze is fast-moving and never entirely still, eye-gaze cues in molecular graphics should be both subtle and precise. Subtlety ensures visual cues remain small and unobtrusive, guiding user attention without overwhelming the visual field. Precision is essential for accurately representing the user's gaze location, which is critical for identifying specific atoms or bonds within the molecular structure. This combination allows a cleaner, more intuitive interface, enhancing collaborative discussions by effectively communicating visual focus without distraction.

Design Requirement 2: Real-time Immediate Referencing of Visual Focus. To facilitate synchronized and fluid molecular discussions in VR, it is essential to implement eye-gaze cues that allow for real-time, immediate referencing of visual focus on the 3D protein structure without delay. For instance, while real-time heatmap visualization is favored in competitive gaming [72], our prototyping revealed that using heatmap visualization on a protein structure in a dynamic VR discussion session takes some time to 'heat up' (show higher focus areas) and 'cool down' (show less focus) based on user attention. This could lead to delays in real-time interactions and might not provide the instant feedback needed in such a collaborative setting. Furthermore, we noticed that the heatmap can be distracting, as its colors continuously change from cool to warm hues to indicate gaze focus. Consequently, we excluded heatmap-like visualization from our system.

Design Requirement 3: Minimize Visual Clutter. Ensuring a clear and focused interface requires minimizing visual clutter in eye-gaze visualizations. Excessive visual elements can overwhelm users, making it difficult to filter out irrelevant information, which increases cognitive load and leads to confusion and frustration. For example, the virtual gaze ray, another common method for representing eye-gaze cues [52–54], was sometimes confused with rays from handheld controllers during our pilot testing. The gaze ray indicates attention and focus, while the controller ray is used for pointing, selecting, or manipulating the protein model. Users felt that an excess of rays led to visual clutter, overwhelming the interface and diminishing the collaborative experience's effectiveness. To maintain a clean, user-friendly interface in GazeMolVR, we made the design decision to exclude the gaze ray visualization.

Design Requirement 4: Consistency Across Representations and Distinct Color Coding. To ensure effective eye-gaze visualization in collaborative sessions, it is important to maintain consistency in the design of gaze cues across different protein representations, such as cartoon, ball-and-stick, and surface models. This consistency ensures that each visualization technique functions effectively regardless of the protein model being used. Additionally, distinct color coding should be employed to differentiate between the gaze cues of different collaborators. By assigning unique colors to each collaborator's gaze visualization, users can easily identify and follow the focus of their peers, enhancing the collaborative experience.

3.2.2 Gaze Visualization Techniques for Molecular Structures. Based on the design criteria described in 3.2.1, we developed four distinct styles for bi-directional eye-gaze cue visualizations: GazePoint, GazeArrow, GazeSpotlight, and GazeTrail. These styles are specifically adapted for molecular graphics by incorporating established techniques from the literature [49, 52, 53, 80, 90]. The first three techniques utilize point-based representations, while the last employs a trajectory-based approach. Each visualization is relatively small within the context of the scaled-up protein model (see Figure 1). During collaborative discussions in VR, each pair uses the same visualization technique but with distinct color coding—one in red and the other in blue. Below are descriptions of all four eye-gaze visualization techniques.

GazePoint: In this technique, a small sphere with a radius of 3 cm appears at the eye-gaze location on a 3D protein structure (see Figure 2(A), Figure 3(A), and Figure 4(A)). This design offers a simple and minimalist representation of a gaze cue, similar to an on-screen cursor pointer on a 2D display.

GazeArrow: This method is similar to GazePoint, with the key difference being the use of a three-dimensional arrow, to point downwards, to indicate the current location of the eye-gaze (see Figure 2(B), Figure 3(B), and Figure 4(B)). The arrow's overall dimensions are a length of 7.5 cm, a width of 2.5 cm, and a thickness of 1 cm. In virtual 3D environments, an arrow is often used to pinpoint the position of an object of interest. Therefore, we anticipate that its use will naturally and effectively highlight a user's current gaze location. Additionally, the arrow is comparatively larger in size than the sphere used in GazePoint.

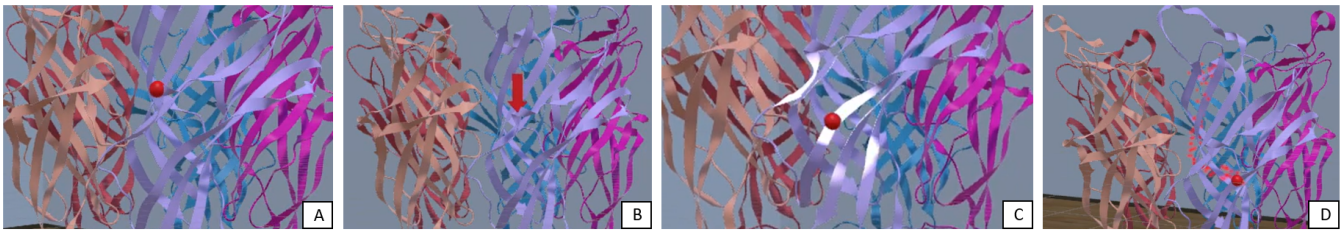

Figure 2: Eye-gaze visualizations for protein in cartoon model: (A) GazePoint, (B) GazeArrow, (C) GazeSpotlight, and (D) GazeTrail.

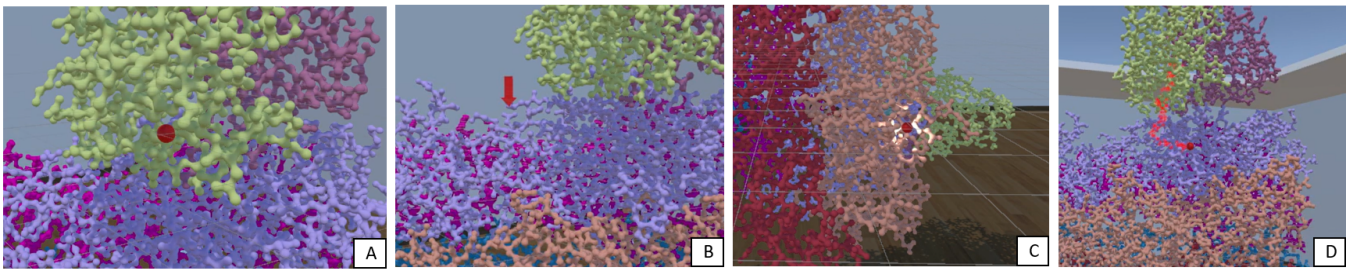

Figure 3: Eye-gaze visualizations for protein in ball-and-stick model: (A) GazePoint, (B) GazeArrow, (C) GazeSpotlight, and (D) GazeTrail.

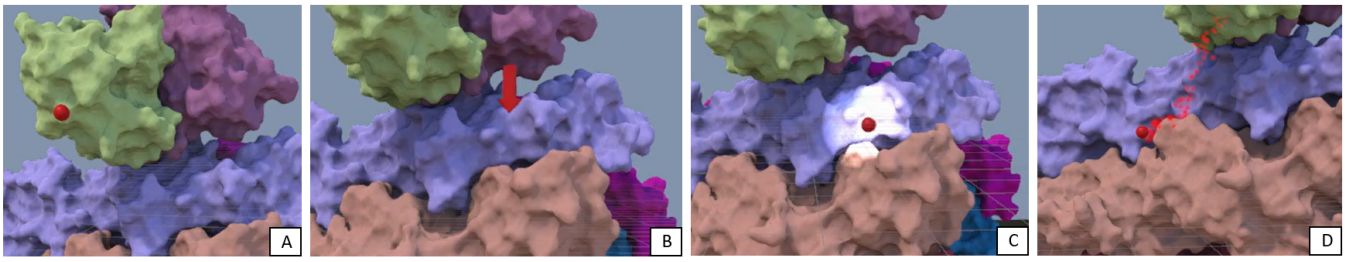

Figure 4: Eye-gaze visualizations for protein in surface model: (A) GazePoint, (B) GazeArrow, (C) GazeSpotlight, and (D) GazeTrail.

GazeSpotlight: This visualization technique combines GazePoint with additional lighting features to enhance focus on specific areas of a 3D protein structure (see Figure 2(C), Figure 3(C), and Figure 4(C)). For proteins depicted using cartoon and ball-and-stick representations, GazeSpotlight uses GazePoint as the origin for a point light source. This light source emits white light uniformly in all directions up to a radius of 12 cm. The intensity decreases with distance from the source and eventually becomes zero at the outer limit of its range, following the inverse square law where intensity is inversely proportional to the square of the distance. For proteins with surface representations, GazeSpotlight incorporates GazePoint with a custom shader. This setup illuminates the protein's surface with white light within a 8 cm radius from GazePoint, simulating a realistic light fall-off effect, thereby spotlighting the area under the user's gaze effectively.

GazeTrail: This method visualizes gaze history by aggregating gaze points into a trail over a specific period (see Figure 2(D), Figure 3(D), and Figure 4(D)). To achieve this, we utilized a particle system where the emitter dynamically relocates to each new gaze point, producing particles with a lifespan of 2 seconds. Additionally, we explored representing the gaze trail using a volumetric line. However, through pilot testing, we realized that the particle-based trail presents a more aesthetically pleasing effect compared to the volumetric line approach in our molecular visualization context. To highlight the trail's leading edge, we employed the GazePoint technique to denote the trail head. In our particle system, the size of each particle diminishes progressively throughout its lifespan. GazeTrail effectively captures and represents the spatial and temporal dynamics of a user's focus.

The initial values for the eye-gaze visualization attributes were determined empirically to ensure that the gaze cues are clearly visible on the protein structure from approximately 2 meters away and remain unaffected by protein scaling. These values were used consistently across both of our user studies. However, users can customize these settings through the system menu to meet their specific needs. While it is possible for the color of the gaze cues

to occasionally match parts of the protein, as the protein is color-coded by chain, the dynamic nature of eye-gaze behavior ensures that users can still locate each other's gaze cues.

4 Implementation

The system prototype, as illustrated in Figure 1, was built using two HTC VIVE Pro Eye [11] headsets, each connected to a separate PC equipped with an Intel Core i7-11800H CPU, 32GB RAM, and an NVIDIA RTX 3080 GPU, all running on Windows 11 OS. Our collaborative VR application was developed using the Unity 3D game engine (version 2019.4.40f1). In our application, we utilized the VIVE Input Utility (VIU) toolkit [24] for VR interactions, the SRanipal SDK [23] (version 1.3.6.8) to capture eye-tracking data from the headset, the Photon PUN 2 library [39] and Photon Voice 2 library [40] to integrate multiplayer features and high-quality, low-latency voice chat, respectively, and UnityMol [33] for biomolecule visualization. Participants, represented by avatars in the virtual environment, could join remotely to engage in collaborative molecular discussions, navigating the space through teleportation. Furthermore, we applied a moving average filter on top of the eye tracker's already lightly filtered data to reduce the jittery motions that naturally arise from the participant's eye movements. This approach gave all four eye-gaze representations a hovering effect when a user's gaze shifts across different parts of the 3D protein structure.

5 User Study - I

Our first study aims to identify the most suitable eye-gaze visualization method—GazePoint, GazeArrow, GazeSpotlight, or GazeTrail—for discussing proteins depicted with cartoon, ball-and-stick, and surface models, respectively.

5.1 Participants

In our experiment, we recruited 20 unpaid participants (P1-P20), comprising 13 males and 7 females, with ages ranging from 24 to 40 years (mean = 29.55, SD = 5.29). All participants were engaged in advanced studies, either as PhD students or postdoctoral researchers, specializing in biochemistry, with a focus on structural, theoretical, and computational aspects. They regularly used various molecular visualization tools such as VMD [48], Chimera [76], and PyMOL [21] on their computers for visualizing molecular structures, density maps, and trajectories from molecular dynamics simulations. Regarding their experience with AR/VR technology, 15 participants were beginners, having played some games in virtual reality, while the remaining five had no prior experience. All had either normal or corrected-to-normal vision.

5.2 Study Design

The study utilized a within-subjects design with 12 conditions across 3 types of protein representations and 4 eye-gaze visualizations. We pre-recorded GazeMolVR scenes for each condition, with each recording lasting approximately 2 minutes. In these recordings, an instructor discussed the structure and biological function of various proteins, emphasizing how their unique structural features facilitate diverse biological processes. Given that there were 4 types of eye-gaze visualizations for each type of protein representation, we selected four different proteins for each category for our recordings. For the cartoon representation, our focus was on the following proteins: Multiple C2 Domains and Transmembrane Region Proteins (MCTPs), consisting of 37,881 atoms [87]; Mitofusin Fzo1, with 24,650 atoms [92]; NOX2-p22phox complex, containing 6,098 atoms [1]; and the PvdRT-OpmQ Efflux Pump, comprising 55,542 atoms [88]. The ball-and-stick representation featured the GABA-A alpha1-beta2-gamma2 receptor (17365 atoms) [32] and its interactions in various complexes with bicuculline (17407 atoms) [26], GABA-flumazenil (17387 atoms) [28], and the Erwinia chrysanthemi bromoflurazepam complex (25120 atoms) [35]. Lastly, for the surface representation, we considered the GABA-A alpha1-beta2-gamma2 receptor's interaction with GABA and four specific molecules: propofol (17415 atoms) [27], etomidate (17426 atoms) [29], phenobarbital (17399 atoms) [30], and diazepam (17470 atoms) [31]. These proteins were selected to match the area of work and expertise of the researcher who acted as a teacher in our recordings, ensuring a realistic use case, but there were no restrictions on which proteins could be considered for this experiment. Before the start of each recording, the instructor adjusted the scale, orientation, and position of the protein in VR using handheld controllers. Each protein representation was color-coded by chain. During the recording, the instructor looked at different parts of the 3D protein structure using eye-gaze while talking about its various biological properties and functions, with no controller-based pointing being used. As the instructor described a protein, the transformation of the VR headset and controllers, the eye-gaze position, the verbal description, and the protein's transformation were recorded. It is important to note that the instructor's eye-gaze was visualized using GazePoint during all 12 pre-recordings. Since the eye-gaze position was recorded, the gaze visualizations could be interactively changed during the replay of those pre-recordings. This approach resulted in a total of 48 pre-recordings, as we applied each of the 4 eye-gaze visualizations to the 12 proteins, thus creating 4 unique recordings for each protein.

In our study, each participant experienced 3 x 4 = 12 conditions, combining three protein representations (cartoon, ball-and-stick, and surface) with four eye-gaze visualizations (GazePoint, GazeArrow, GazeSpotlight, and GazeTrail). During the experiment, the order of specific proteins within each protein representation category was consistent for all participants. To ensure an unbiased assessment, we counter-balanced the two independent variables (protein representation types and eye-gaze visualization techniques) using a Balanced Latin Square method. There was one trial per condition, resulting in a total of 12 trials per participant.

Overall, our pre-recorded VR sessions guaranteed that each participant received the same information and visual stimuli, essential to compare the impact of the four eye-gaze visualization techniques on protein structure comprehension. By selecting different proteins for each representation type, we guaranteed variety in the study material, further minimizing variables that could affect the study's outcome and ensuring a controlled evaluation environment.

5.3 Study Procedure, Task, and Measures

Participants were welcomed upon arrival at our lab, where they were asked to read and sign a consent form, and fill out a pre-study questionnaire, to gather demographic information and their prior

experience with AR/VR technology. They were then introduced to the VR setup and the objectives of the experiment. We assisted participants in wearing the HTC VIVE Pro Eye headset comfortably and guided them through the eye-tracking calibration process to adjust for their personal interpupillary distance. Each participant went through a training phase before starting the actual experiment. In both phases, they replayed pre-recorded 3D scenes in which the instructor was represented using an avatar with two controllers. Participants could see their own eye-gaze represented in blue and the instructor's in red. They were instructed to stand side-by-side with the avatar to ensure they shared a similar viewpoint, making it easier to follow the instructor's focus on the complex 3D protein structure. Without this arrangement, participants might not only focus on different parts of the protein during the tutorials but also struggle to align their perspective with the instructor's, potentially leading to misunderstandings or missed details in the explanation. Their task was to attentively listen to the instructor's verbal description of the protein while simultaneously following the instructor's gaze on the 3D protein structure to fully comprehend the content. The instructor's dynamic gaze cue referred to points of interest on the protein, directly linked to the verbal explanation. Since pausing or rewinding the tutorial was not allowed, participants couldn't revisit missed sections if they didn't pay attention. Therefore, closely following the instructor's gaze from the beginning was crucial to understanding the content being discussed. Participants were encouraged to imagine themselves as students in a remote collaborative VR discussion session with their instructor, aiming to understand a protein's structure and its biological functions. Participants were asked to evaluate how eye-gaze visualizations might help them easily understand the content and follow the instructor's gaze. Both eye gazes were visualized using the same representation technique in each replay session. We also informed them that their eye-gaze data would be recorded during the actual experiment to measure the similarity of gaze paths between theirs and the instructor's. Separate pre-recorded scenes were used for the practice phase, lasting about 15 minutes per participant.

Once participants felt ready, we began the main experiment. Upon completion of each condition, they were asked to fill out a NASA-TLX questionnaire [45] to assess subjective task workload. Following the completion of all four conditions for each protein representation type, they ranked the eye-gaze visualization techniques based on their effectiveness in facilitating the task of following the instructor's tutorial and took a short break. This break allowed us to discuss the reasons behind their preferences for gaze techniques. Once all 12 conditions were completed, we conducted an informal post-study interview. The entire study took an average of 80 minutes to complete. It was conducted in accordance with the rules of the local ethics committee of our institute, which does not require formal approval for this type of experiment.

5.4 Results

Since none of the dependent variables (i.e., gaze path similarity, NASA-TLX, and eye-gaze visualization rankings) met the ANOVA assumptions of normality and equal variances, we applied a 3x4 Aligned Rank Transform (ART) for nonparametric factorial analysis [94]. The independent variables in our analysis were protein representations and eye-gaze visualizations. When the ART analysis revealed a significant main effect of these independent variables, or an interaction effect between them, we conducted post hoc pairwise comparisons using the Holm correction method. The ART analysis was performed using the ARTool[1] package in RStudio[2]. For all significance tests, we set a threshold of $\alpha = 0.05$, the standard for indicating statistical significance.

5.4.1 Gaze Path Similarity. Gaze path similarity plays a pivotal role in our study as it assesses how closely participants' gaze paths aligned with the instructor's during the VR tutorial. This metric reflects how effectively participants followed and focused on the protein structures emphasized by the instructor. A high similarity suggests successful tracking of the instructor's attention, which is crucial for understanding the protein's structural and functional characteristics. Conversely, divergence in gaze paths indicates that the eye-gaze visualizations may not have sufficiently supported participants in following the instructor's gaze, potentially causing them to miss key information.

To quantify this similarity, we employed the dynamic time warping (DTW) algorithm, a common technique used in the literature for finding similarities among eye-gaze scanpaths [36, 59]. A lower DTW distance indicates a higher degree of alignment between the participant's and the instructor's gaze paths, capturing both spatial and temporal similarities. The ART analysis revealed a significant main effect of protein representation on gaze path similarity ($F_{2,209} = 252.91$, $p < 0.001$), as depicted in Figure 5. Pairwise comparisons revealed that the cartoon representation exhibited a significantly higher DTW distance (Mean = 1555.47, SD = 54.41) compared to both the ball-and-stick (Mean = 832.67, SD = 79.86) and surface (Mean = 382.24, SD = 49.72) models. Furthermore, the ball-and-stick model also showed a significantly higher DTW distance than the surface model. The results showed no main effect of eye-gaze visualizations ($p = 0.25$) and no significant interaction effects between protein representations and eye-gaze visualizations ($p = 0.67$).

5.4.2 NASA-TLX. The overall NASA-TLX scores for all eye-gaze visualizations under each protein representation are presented in Figure 7, 8, and 9 respectively.

The ART analysis showed that performance was significantly influenced by protein representations ($F_{2,209} = 4.64$, $p = 0.01$) and eye-gaze visualizations ($F_{3,209} = 3.93$, $p < 0.01$), with no interaction ($p = 0.14$). The surface model significantly outperformed both cartoon ($p = 0.04$) and ball-and-stick ($p = 0.02$) models. GazeTrail was significantly more effective than GazePoint ($p < 0.01$).

No significant effects were observed for physical demand regarding protein representations ($p = 0.25$), eye-gaze visualizations ($p = 0.36$), or their interaction ($p = 0.36$).

Temporal demand was significantly affected by protein representations ($F_{2,209} = 3.27$, $p < 0.05$) and eye-gaze visualizations ($F_{3,209} = 3.91$, $p < 0.01$), with no interaction effect ($p = 0.72$). The ball-and-stick model was found to increase temporal demand over the surface model ($p = 0.03$), and GazePoint was more demanding than GazeTrail ($p < 0.01$).

[1] https://depts.washington.edu/acelab/proj/art/
[2] https://posit.co/products/open-source/rstudio/

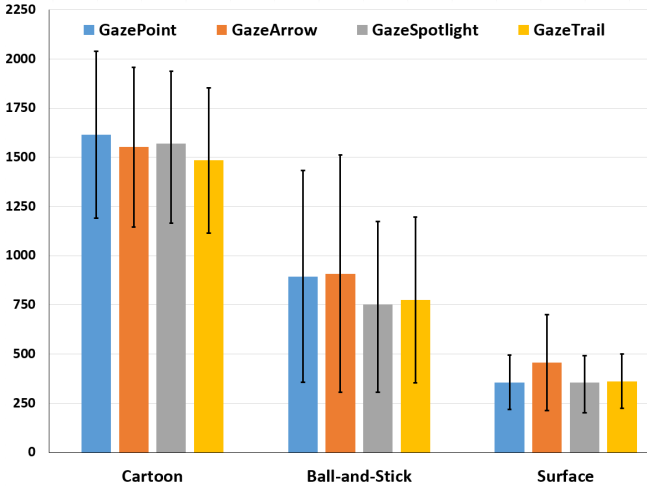

Figure 5: Mean DTW distance values for each protein representation type across all four eye-gaze visualizations. Error bars represent standard deviation.

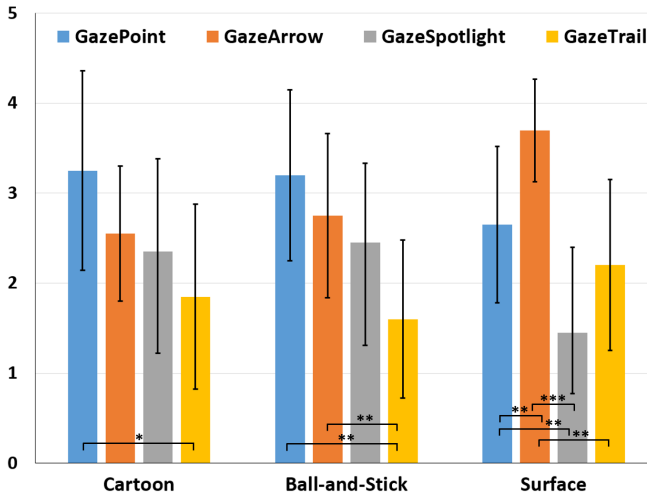

Figure 6: Mean ranking score on a scale of 1 to 4 for all eye-gaze visualizations for each protein representation; the lower the score, the higher the preference. Statistical significances are marked with stars (*: $p < 0.001$, **: $p < 0.01$, and *: $p < 0.05$). Error bars represent standard deviation.**

Protein representations had a significant impact on mental demand ($F_{2,209} = 11.24$, $p < 0.001$), while eye-gaze visualizations did not show a significant effect ($p = 0.13$) nor did their interaction ($p = 0.41$). Further post hoc analysis demonstrated that the ball-and-stick representation demanded more mental effort than both the cartoon ($p = 0.03$) and surface ($p < 0.0001$) models. However, there was no significant difference in mental demand between the cartoon and surface models ($p = 0.08$).

The ART analysis for frustration revealed significant effects of protein representations ($F_{2,209} = 6.36$, $p < 0.01$) and eye-gaze visualizations ($F_{3,209} = 5.18$, $p = 0.001$), with no interaction ($p = 0.15$). Post hoc tests indicated ball-and-stick significantly increased frustration over cartoon ($p = 0.01$) and surface ($p < 0.01$). GazeTrail was significantly less frustrating than GazePoint ($p < 0.01$) and GazeArrow ($p < 0.01$), with no other differences.

Effort was significantly influenced by protein representations ($F_{2,209} = 6.53$, $p < 0.01$) but not by eye-gaze visualizations ($p = 0.13$) or their interaction ($p = 0.42$). Ball-and-stick demanded more effort than cartoon ($p = 0.01$) and surface ($p < 0.01$), with no significant difference between the latter two.

5.4.3 Eye-Gaze Visualization Rankings. The ART analysis for overall ranking (1: most preferred, 4: least preferred) showed no significant main effects of protein representations ($p = 0.99$), but significant main effects of eye-gaze visualizations ($F_{3,209} = 25.45$, $p < 0.0001$) and a significant interaction between protein representations and eye-gaze visualizations ($F_{6,209} = 7.83$, $p < 0.0001$).

Post hoc pairwise comparisons revealed that GazePoint was significantly less preferred compared to GazeTrail and GazeSpotlight across protein representations ($p < 0.0001$ for both). GazeArrow was also less preferred compared to GazeTrail and GazeSpotlight ($p < 0.0001$ for both). No significant preference difference was found between GazeTrail and GazeSpotlight ($p = 0.65$).

Given the significant interaction effect, the Friedman test was used to assess the differences in gaze visualization rankings within each protein representation type. This test helps determine if the ranking preferences vary significantly among different visualizations within each specific protein representation. This analysis found significant differences for cartoon ($\chi^2(3) = 12.12$, $p < 0.01$), ball-and-stick ($\chi^2(3) = 16.38$, $p < 0.001$), and surface ($\chi^2(3) = 31.86$, $p < 0.0001$) models (see Figure 6). Specifically, for the cartoon representation, GazeTrail was preferred over GazePoint ($p < 0.05$). For ball-and-stick, GazeTrail was favored more than GazeArrow and GazePoint ($p < 0.01$ each). Within the surface category, GazeArrow was less preferred than GazePoint ($p < 0.01$), GazeSpotlight ($p < 0.001$), and GazeTrail ($p < 0.01$); additionally, GazePoint was less favored than GazeSpotlight ($p < 0.01$).

5.5 Discussion

In this section, we reflect on the key findings of our user study. Through the analysis of gaze path similarity, NASA-TLX scores, and participant preferences, we aim to identify which eye-gaze visualizations are most effective, depending on the protein representations, for facilitating discussions in a VR environment centered on complex molecular structures.

5.5.1 Reflecting on Gaze Path Similarity Metric. Overall, participants were able to track the instructor's gaze path using all four eye-gaze visualizations across each type of protein representation, as illustrated in Figure 5. They found the surface model to be the easiest to follow because it presents a seamless, smooth contour that encapsulates the molecule's volume (see Figure 4).

In contrast, the cartoon model simplifies the protein's structure into distinct geometric shapes, such as alpha-helices and beta-sheets. However, as shown in Figure 2, this simplification introduces

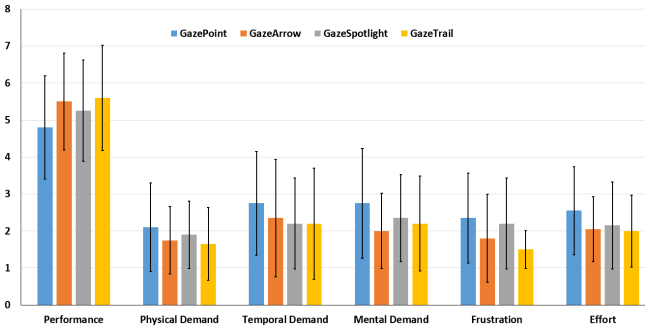

Figure 7: NASA-TLX scores for cartoon representation, indicating mean task load values on a scale of 1 to 7, where lower scores are favorable, except for the performance metric. Error bars represent standard deviation.

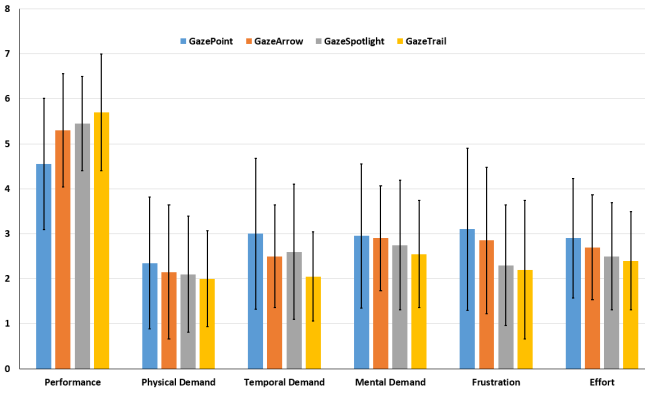

Figure 8: NASA-TLX scores for ball-and-stick representation, indicating mean task load values on a scale of 1 to 7, where lower scores are favorable, except for the performance metric. Error bars represent standard deviation.

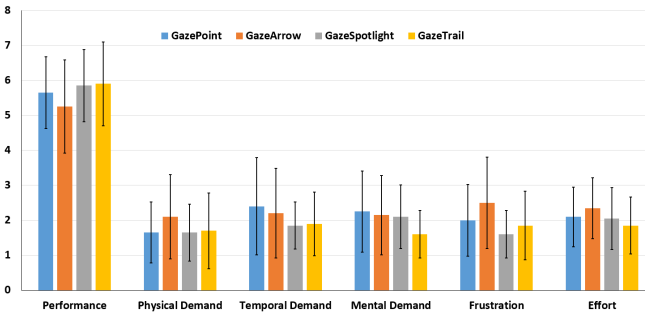

Figure 9: NASA-TLX scores for surface representation, indicating mean task load values on a scale of 1 to 7, where lower scores are favorable, except for the performance metric. Error bars represent standard deviation.

greater variability in spatial depth, creating gaps and spaces between elements, which makes it more difficult for participants to track the instructor's dynamic gaze movements. This difficulty is reflected in the DTW distance measurements, where the mean DTW distance for the cartoon representation is 306.94% higher compared to the surface representation.

The ball-and-stick model, which represents atoms as spheres and bonds as sticks, is more detailed than the cartoon model and less continuous than the surface representation. Since this model is denser compared to the cartoon, it offers reduced depth variability (see Figure 3), leading to more consistent and aligned gaze paths between the instructor and participants. Consequently, it yields an intermediate DTW distance that is 46.47% lower than the cartoon representation but 117.84% higher than the surface representation.

5.5.2 Reflecting on NASA-TLX Measures and Eye-Gaze Visualization Rankings. For each protein representation, NASA-TLX scores frequently varied across different eye-gaze visualizations, but gaze path similarity did not. This suggests that while participants tracked the instructor's gaze equally well with all visualizations, the effort and cognitive load required differed. Certain visualizations made it easier for participants to process and comprehend the information.

In cartoon representations of complex proteins, participants occasionally found it difficult to quickly locate the instructor's gaze with simple visualizations like GazePoint and GazeArrow, as these markers could temporarily get lost within the structure, particularly when their colors matched those of the protein chains. In this context, GazeArrow was slightly preferred over GazePoint due to its larger size. Participants noted that GazeSpotlight's lighting effectively highlighted the instructor's current focus within the cartoon model. However, gaps between the structure's elements sometimes diminished its effectiveness by affecting light reflection. Most participants found GazeTrail the most helpful, as its continuous trail made it easier to anticipate the instructor's gaze direction, especially when they momentarily lost track of it within the protein structure while focusing on the verbal description. This continuous feature distinguished GazeTrail from the other, more discrete visualizations. During the interview, several participants commented, *"For smaller proteins like the NOX2-p22phox complex, GazePoint or GazeArrow works well for following the instructor's gaze throughout the tutorial. However, for larger proteins (i.e., MCTPs, Fzo1, PvdRT-OpmQ) represented in cartoons, I would personally prefer GazeTrail as it provides a continuous, easier-to-follow path that accommodates the greater variability in spatial depth in the structure".*

The ball-and-stick model's detailed atomic-level interactions significantly increased participants' mental and temporal demands, frustration, and effort compared to the cartoon and surface models. This increase was due to the need to rapidly process fine-grained information, track multiple interactions simultaneously, and navigate a densely packed visual field, making the task more cognitively taxing. Multiple participants (P2, P3, P8, P11, P19) remarked, *"During the tutorial, as the instructor explained how GABA binds to the receptor using the ball-and-stick model, I had to closely follow the instructor's eye-gaze to pinpoint specific atoms in GABA interacting with residues in the receptor, navigating through a dense network of atoms. But it wasn't just about seeing where the instructor was*

looking—I also needed to mentally piece together how these atoms interacted, like forming hydrogen bonds or salt bridges, and understand how these interactions led to biological functions, such as channel opening and chloride ion flow, based on the verbal description. Balancing this dual task of following the gaze references and processing the verbal explanation, all while shifting focus between different atomic groups, made the experience more cognitively demanding than the cartoon and surface models".

Regarding eye-gaze visualization preferences for the ball-and-stick model, participants' feedback was similar to that for the cartoon model. They reported that GazePoint and GazeArrow could easily become lost within the protein structure due to the model's density. Tracking the instructor's gaze was more difficult with GazePoint than with GazeArrow, as its shape blended with the atoms. A common observation among participants was, *"Although I appreciated the subtle design of GazePoint, I found it a bit frustrating to follow the instructor's gaze with it"*. GazeTrail was the most effective, offering continuous tracking. Interestingly, GazeSpotlight's lighting worked better in this dense model, which led to it being the second most preferred visualization among participants.

In terms of performance scores on the NASA-TLX, the surface model outperformed both the cartoon and ball-and-stick models. Most participants mentioned, *"Following the instructor's gaze in the surface model of the GABA-A receptor was easier due to the clear, continuous view of the protein's exterior. The smooth contours allowed me to quickly identify where the instructor was focusing, without getting into the detailed internal structures like in the cartoon or ball-and-stick models. The simplicity of the surface model made it straightforward to understand how different ligands (i.e., propofol, etomidate, phenobarbital, and diazepam) interact with the protein"*.

In the surface model, GazePoint was generally effective, though it occasionally became less visible due to its size or when its color resembled that of the protein chains. Most participants did not prefer GazeArrow as it frequently became occluded or hidden behind surface contours due to its three-dimensional shape and vertical orientation. Unlike GazePoint, which lies directly on the surface, the GazeArrow extends above the surface at the participant's gaze point (see Figure 4(A) and Figure 4(B)). This positioning makes it susceptible to being partially or fully obscured by the protein's surface features, especially in concave regions, pockets, or grooves. Similar to the cartoon and ball-and-stick models, GazeTrail facilitated tracking of the instructor's dynamic gaze on the surface. However, some participants pointed out that the appearance of floating particles was visually unappealing and generated visual clutter, particularly when examining protein pockets. GazeSpotlight emerged as the most favored visualization; its lighting blended nicely with the surface, providing an aesthetically pleasing effect. Participants were particularly impressed by how GazeSpotlight illuminated the pockets during examination.

Moreover, it was observed that none of the participants paid attention to the instructor avatar during the sessions. This was expected because each of the 12 conditions was very short (approximately 2 minutes), requiring participants to follow the instructor's gaze on the protein structure while simultaneously paying attention to the verbal description. There was no opportunity to look away from the protein, as they were all focused on following the pre-recorded tutorial. Nevertheless, a few participants expressed sensitivity to GazeSpotlight's bright illumination.

6 User Study - II

In our first user study, the use of pre-recorded sessions ensured that all participants received identical visual and verbal stimuli, eliminating variations that could arise in live, interactive settings. This standardized approach allowed for reliable comparisons and generalizable conclusions on preferred eye-gaze visualization across different protein representation styles.

The purpose of the second study is to explore whether sharing bi-directional eye-gaze cues enhances real-time collaborative discussions about proteins, compared to a baseline condition without eye-gaze sharing. In a live, interactive setting, participants and instructors can dynamically interact, ask questions, and adjust their focus based on the discussion's flow, leading to a more tailored and responsive learning experience.

6.1 Study Design and Procedure

The study involved 20 unpaid participants (P1-P20), comprising 12 males and 8 females, aged between 23 to 36 years (mean age: 27.21), who were randomly paired. Four of the participants had not taken part in our first user study. All were structural biologists who regularly used molecular visualization tools like VMD [48], Chimera [76], and PyMOL [21] in their research. Most participants were novices in AR/VR technology, and all had either normal or corrected-to-normal vision.

To facilitate engaging collaborative sessions, we asked each pair to discuss the protein they regularly use in their molecular dynamics simulations. Each dyad participated in four discussion sessions, each lasting about seven minutes. In the first two sessions, one participant acted as the 'instructor', explaining the protein's properties while the other followed along and asked questions. Roles were switched for the last two sessions. During their turns, participants first described their proteins using the cartoon representation, followed by the surface representation. The ball-and-stick model was not used separately, as GazeTrail was preferred for both the cartoon and ball-and-stick representations in the first user study. To avoid redundancy, we selected the cartoon representation. Additionally, most participants incorporated ligands in the ball-and-stick style within their protein structures, making a separate emphasis on this model less necessary. In each dyad, the eye-gaze condition was randomly assigned to both sessions of one participant. In the gaze condition sessions, participants used handheld controllers to manipulate the protein (e.g., changing its position, rotation, and zoom) while relying on eye-gaze for pointing. GazeTrail was employed for the cartoon representation and GazeSpotlight for the surface representation. This design aimed to assess the effectiveness of eye-gaze as a mutual awareness cue during molecular discussions in VR, isolating its impact by minimizing the use of manual pointing. In the no eye-gaze condition, they utilized controllers for pointing and manipulating the protein. In both conditions, only one participant at a time could manipulate the protein.

Upon arrival at our lab, participants completed a pre-study questionnaire on their demographic details and AR/VR experience, and signed a consent form. They were briefed on the study objectives

and tested the GazeMolVR system in pairs after calibrating their eyes with the HTC VIVE Pro Eye headset. During the actual experiment, each participant's individual protein was loaded into our system and color-coded by chains. The study was conducted in a room divided into two sections, each measuring 3 meters in length, 3 meters in width, and 2.5 meters in height, where participants were physically separated but could freely speak to each other. Each section was calibrated using the HTC VIVE Lighthouse system. Participants used teleportation to navigate in the VR environment. They were embodied in avatars that stood side by side during the discussions, ensuring a shared viewpoint. Unlike in the first user study, participants were allowed to move independently within the scene to explore the protein from different perspectives, depending on the specific aspects they were discussing. Additionally, their eye-gaze visualizations were represented in red and blue, respectively. After each condition, a short interview was conducted to assess their collaborative experience, focusing on mutual and self-awareness of visual focus, mutual understanding of spatial references, deictic pointing, avatar embodiment, and preferences. The study took about an hour and was conducted in line with our institute's ethics committee guidelines, which do not require formal approval for this type of experiment.

6.2 Discussion

The presence of bi-directional eye-gaze cues enabled participants to seamlessly track each other's visual attention, thus aligning their focus and enhancing collaboration efficiency compared to the baseline condition. Throughout the interview, many participants mentioned, *"Eye-gaze cues served as an implicit signal and required no effort. It feels more connected. By seeing where my partner was looking on the protein, I could dynamically adjust my explanations, making them easier to understand"*. This mutual awareness of visual focus facilitated deictic pointing (e.g., 'this', 'that', 'here', 'there'), streamlining discussions about proteins by allowing participants to directly observe and follow their partner's gaze, pinpointing areas of interest without complex verbalizations. Similar findings were reported in previous studies [15, 52–54]. Conversely, in the no eye-gaze condition, pairs often sought confirmation by asking, *"Do you see where I am pointing with the controller?"*. This frequent use of explicit verbal confirmations to synchronize their spatial references slowed the flow of discussion. Additionally, some participants noted that they occasionally observed the other participant's avatar to discern their focus within the virtual environment based on the direction of their head.

In the gaze condition, participants observed that during general conversational phases—such as discussing the protein's broader biological role, research context, or experimental findings—there was often no need to focus on specific parts of the protein structure. In these situations, gaze visualizations felt redundant and sometimes distracting, especially during longer discussions. They suggested that controller ray-based pointing would be more appropriate for these broader conversations, if needed. This distraction issue was also pointed out by Yang et al. in their remote tutoring experiment with eye-gaze [96]. Furthermore, participants highlighted specific contexts during molecular discussions where sharing bi-directional eye-gaze cues significantly enhanced collaboration. For example, when a question focused on a particular area of the protein, bi-directional gaze cues helped confirm that collaborators were paying attention to the same detail. P7 and P9 commented, *"When I asked about the active site, the bi-directional gaze cues made it obvious that we were both focused on the same spot, so I could jump straight into my question without having to double-check where we were looking"*. Similarly, when one collaborator explained interactions between a protein region and ligands, these gaze cues ensured that their partner was following along. Another context where bi-directional gaze cues were particularly useful was during the exploration of the protein's interior in surface representation. The interior surfaces often appeared darker due to rendering techniques, making it difficult to see details. In these cases, the GazeSpotlight proved invaluable by enhancing visibility and facilitating the discussion of biological interactions within the protein structure. Participants also suggested implementing an easily accessible toggle button to activate or deactivate gaze cues as needed, allowing for smoother transitions between detailed structural discussions and broader, context-driven conversations—similar to the suggestion by Jing et al. [51] to use contextual speech input for visualizing shared gaze cues between remote collaborators.

Participants found it slightly challenging to maintain their attention on a specific area of the cartoon model while explaining, compared to the surface model, due to the cartoon's inherent spatial depth variability (e.g., gaps and spaces between alpha-helices and beta-sheets). Additionally, they remarked that using hand gestures during explanations felt instinctive, which sometimes led to spontaneous controller pointing, even in the gaze condition. Lastly, participants occasionally confused their eye-gaze cue with their partner's and had to either adjust their focus or rely on memorizing their assigned cue color to distinguish between them.

Overall, all participants considered sharing bi-directional eye-gaze cues to be very useful and interesting. They also mentioned that this was their first experience with such a collaborative interaction space and emphasized that further practice would be necessary to fully synchronize all actions, including manipulating the protein, verbal descriptions, eye-gaze, and controller pointing.

7 Design Implications

Our research provides valuable insights into the effective use of eye-gaze visualizations for facilitating remote collaborative discussions about proteins in virtual reality. The design implications derived from our findings are outlined below:

- **Availability of All Eye-Gaze Visualizations:** All four eye-gaze visualizations (GazePoint, GazeArrow, GazeSpotlight, and GazeTrail) should be available for each type of protein representation (cartoon, ball-and-stick, and surface). This ensures the system remains flexible and user-centered, allowing users to select the gaze visualization that best suits their needs and preferences. As demonstrated in our study, participants were able to successfully follow the instructor's tutorial using any of the eye-gaze visualizations across all types of protein representations (see 5.5.1 for details).
- **Default Visualization Settings:** While all visualizations should be available, the default eye-gaze visualization settings can be optimized based on the preferences observed

in the study (see 5.5.2 for details). GazeTrail should be set as the default for cartoon and ball-and-stick models due to its effectiveness in tracking gaze across complex and densely packed structures, owing to its continuous path. Conversely, GazeSpotlight should be the default choice for surface representations, as it enhances visibility and focus by effectively illuminating specific areas during discussions about proteins with smooth, continuous surfaces.

- **Bi-Directional Eye-Gaze Cues for Real-Time Collaboration:** Sharing bi-directional eye-gaze cues enhances real-time collaborative discussions about proteins in VR by improving focus alignment and reducing the need for verbal confirmations, especially when discussing specific protein areas. However, these cues can become distracting during broader conversations, suggesting the need for a toggle option to turn them on or off as needed. Integrating controller-based pointing with eye-gaze cues could offer a more versatile approach, allowing users to select the most appropriate tool depending on the specific context and requirements of their discussion (see 6.2 for details).

8 Limitations and Future Work

Our studies shed light on the potential of eye-gaze visualizations in collaborative molecular discussions; however, this section addresses the main limitations of our research and offers suggestions for future exploration.

In our first user study, we focused on the three most commonly used protein representations: cartoon, ball-and-stick, and surface. However, it would be valuable to explore the effectiveness of eye-gaze visualizations for other representations, such as wireframe, space-filling, van der Waals surface, and electrostatic potential maps. Broadening this investigation could offer important insights into optimizing gaze-based interactions across a wider range of molecular visualization techniques. Additionally, our study evaluated eye-gaze visualizations using isolated proteins. In contrast, real molecular dynamics simulations often include proteins visualized alongside solvents or lipids. Future research should explore how these gaze visualizations support collaborative discussions in such dynamic and complex environments.

Currently, our second study is limited to collaborative molecular discussion tasks and relies solely on subjective feedback, with no quantitative evaluation. While focusing on protein discussions is valuable for understanding the impact of sharing bi-directional eye-gaze cues, this approach doesn't fully reflect real-world use cases. Tasks like collaborative searching (e.g., identifying AlphaFold[3] prediction errors, locating ligand binding sites, or finding docking sites) are more aligned with the practical needs of researchers, instructors, and students. Inspired by previous work [17, 52–54], expanding our study to include these tasks could provide a more comprehensive evaluation of eye-gaze cues in various collaborative scenarios, especially in educational contexts.

Our both studies focused on dyadic interactions, using symmetric eye-gaze visualizations to share attention cues between two participants. While this approach effectively explored bi-directional eye-gaze cues, it does not address the complexities of multi-user settings, where managing multiple gaze paths and attention cues becomes more challenging. Further research is needed to develop strategies for visualizing and coordinating mutual gaze awareness in group discussions, where attention dynamics are more complex.

Another limitation of our studies is that all participants were experienced researchers, either doctoral students or postdoctoral researchers, with significant expertise in structural biology. To better understand the potential of eye-gaze visualizations in collaborative settings, especially in educational contexts, it would be valuable to conduct similar studies with participants of varying levels of expertise, such as both novices and experts.

Furthermore, eye-tracking in low-cost VR HMDs is often limited, leading to the use of head-tracking-based FoV frustum visualization as a proxy for eye movements [6, 77, 78]. Although this method is less precise, it remains functional for identifying user focus. A significant challenge arises in hybrid setups where one user utilizes eye-tracking and another relies solely on head-tracking. Research into how these distinct tracking methods can be synchronized to effectively share mutual gaze awareness cues during collaborative molecular discussions would be valuable.

9 Conclusion

In this work, we developed GazeMolVR to facilitate collaborative exploration and discussion of protein structures and biological functions within a VR environment. By integrating four distinct gaze representations—GazePoint, GazeArrow, GazeSpotlight, and GazeTrail—we enabled remote pairs to share bi-directional eye-gaze cues, enhancing mutual awareness of visual focus. Our findings from the first study indicate that users preferred GazeTrail for discussing proteins depicted in cartoon and ball-and-stick models, while GazeSpotlight was preferred for surface representations. Additionally, our second study confirmed that sharing bi-directional gaze cues significantly enriched collaborative interactions by aligning visual attention and promoting coordinated discussions compared to a baseline condition with no gaze cues.

10 Supplementary Material

The code and demo videos of the GazeMolVR prototype are available at https://github.com/collabmolviz/GazeMolVR.

References

[1] Sana Zineb Aimeur, Burcu Aykac Fas, Xavier Serfaty, Hubert Santuz, Sophie Sacquin-Mora, Tania Bizouarn, Antoine Taly, and Laura Baciou. 2024. Combining Computational Biology and Experimental Knowledge to Draw a Structural Profile of the Active Membrane-Assembled NADPH Oxidase Complex. *bioRxiv* (2024), 2024–02. https://doi.org/10.1101/2024.02.19.579638

[2] Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. 2016. GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1151–1158. https://doi.org/10.1145/2851581.2892459

[3] Ruwayda Alharbi, Ondrej Strnad, Laura R Luidolt, Manuela Waldner, David Kouril, Ciril Bohak, Tobias Klein, Eduard Gröller, and Ivan Viola. 2021. Nanotilus: Generator of Immersive Guided-Tours in Crowded 3D Environments. *IEEE Transactions on Visualization and Computer Graphics* (2021). https://doi.org/10.1109/TVCG.2021.3133592

[4] Jad A Atweh, Rana Hazimeh, and Sara L Riggs. 2023. Can Real-Time Gaze Sharing Help Team Collaboration? A Preliminary Examination of Its Effectiveness with Pairs. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 67. SAGE Publications Sage CA: Los Angeles, CA, 716–721. https://doi.org/10.1177/21695067231193659

[5] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. 2020. A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand

[3] https://deepmind.google/technologies/alphafold/

Gesture Sharing. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13. https://doi.org/10.1145/3313831.3376550

[6] Riccardo Bovo, Daniele Giunchi, Muna Alebri, Anthony Steed, Enrico Costanza, and Thomas Heinis. 2022. Cone of Vision as a Behavioural Cue for VR Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27. https://doi.org/10.1145/3555615

[7] Vojtěch Brůža, Jan Byška, Jan Mičan, and Barbora Kozlíková. 2021. VRdeo: Creating Engaging Educational Material for Asynchronous Student-Teacher Exchange using Virtual Reality. *Computers & Graphics* 98 (2021), 280–292. https://doi.org/10.1016/j.cag.2021.06.009

[8] Kevin C Cassidy, Jan Šefčík, Yogindra Raghav, Alexander Chang, and Jacob D Durrant. 2020. ProteinVR: Web-based Molecular Visualization in Virtual Reality. *PLoS computational biology* 16, 3 (2020), e1007747. https://doi.org/10.1371/journal.pcbi.1007747

[9] Cell-PaintVR 2012 2012. CENTER FOR COMPUTATIONAL STRUCTURAL BIOLOGY: *Cell-PaintVR*. https://ccsb.scripps.edu/cellpaint/cellpaint-vr/. Last Accessed: March 03, 2024.

[10] Jeffrey W Chastine, Jeremy C Brooks, Ying Zhu, G Scott Owen, Robert W Harrison, and Irene T Weber. 2005. AMMP-Vis: A Collaborative Virtual Environment for Molecular Modeling. In *Proceedings of the ACM symposium on Virtual reality software and technology*. 8–15. https://doi.org/10.1145/1101616.1101620

[11] HTC Corporation. 2019–present. Vive Pro Eye. https://www.vive.com/sea/product/vive-pro-eye/overview/. Accessed: 2024-08-29.

[12] Microsoft Corporation. 2019–present. Microsoft HoloLens 2. https://www.microsoft.com/en-us/hololens. Accessed: 2024-08-29.

[13] Fabio J Cortés Rodríguez, Gianfranco Frattini, Sittha Phloi-Montri, Fernando Teixeira Pinto Meireles, Danaé A Terrien, Sergio Cruz-León, Matteo Dal Peraro, Eva Schier, Kresten Lindorff-Larsen, Taweetham Limpanuparb, et al. 2023. MolecularWebXR: Multiuser Discussions about Chemistry and Biology in Immersive and Inclusive VR. *bioRxiv* (2023), 2023–11. https://doi.org/10.1101/2023.11.01.564623

[14] Carolina Cruz-Neira, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. 1992. The CAVE: Audio visual experience automatic virtual environment. *Commun. ACM* 35, 6 (1992), 64–73. https://doi.org/10.1145/129888.129892

[15] Sarah D'Angelo and Andrew Begel. 2017. Improving Communication Between Pair Programmers using Shared Gaze Awareness. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 6245–6290. https://doi.org/10.1145/3025453.3025573

[16] Sarah D'Angelo, Jeff Brewer, and Darren Gergle. 2019. Iris: A Tool for Designing Contextually Relevant Gaze Visualizations. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–5. https://doi.org/10.1145/3317958.3318228

[17] Sarah D'Angelo and Darren Gergle. 2018. An Eye for Design: Gaze Visualizations for Remote Collaborative Work. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3173574.3173923

[18] Sarah D'angelo and Bertrand Schneider. 2021. Shared Gaze Visualizations in Collaborative Interactions: Past, Present and Future. *Interacting with Computers* 33, 2 (2021), 115–133. https://doi.org/10.1093/iwcomp/iwab015

[19] Timothy Davison, Faramarz Samavati, and Christian Jacob. 2019. LifeBrush: Painting, Simulating, and Visualizing Dense Biomolecular Environments. *Computers & Graphics* 82 (2019), 232–242. https://doi.org/10.1016/j.cag.2019.05.006

[20] Helen M Deeks, Kirill Zinovjev, Jonathan Barnoud, Adrian J Mulholland, Marc W Van der Kamp, and David R Glowacki. 2023. Free Energy Along Drug-Protein Binding Pathways Interactively Sampled in Virtual Reality. *Scientific Reports* 13, 1 (2023), 16665. https://doi.org/10.1038/s41598-023-43523-x

[21] Warren L DeLano et al. 2002. PyMOL: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr* 40, 1 (2002), 82–92. https://doi.org/newsletters/newsletter40/11_pymol.pdf

[22] Daniel A Delgado and Jaime Ruiz. 2024. Evaluation of Shared-Gaze Visualizations for Virtual Assembly Tasks. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 765–766. https://doi.org/10.1109/VRW62533.2024.00179

[23] Vive Developers. 2018–present. Vive SRanipal SDK. https://developer-express.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/. Accessed: 2024-08-29.

[24] Vive Software Developers. 2016–present. Vive Input Utility (VIU). https://github.com/ViveSoftware/ViveInputUtility-Unity. Accessed: 2024-08-29.

[25] Sébastien Doutreligne, Tristan Cragnolini, Samuela Pasquali, Philippe Derreumaux, and Marc Baaden. 2014. UnityMol: Interactive Scientific Visualization for Integrative Biology. In *2014 IEEE 4th Symposium on Large Data Analysis and Visualization (LDAV)*. IEEE, 109–110. https://doi.org/10.1109/LDAV.2014.7013213

[26] Jeong Joo Kim et al. 2020. Structure of 6X3S: Human GABAA Receptor Alpha1-Beta2-Gamma2 Subtype in Complex with Bicuculline Methbromide. https://www.rcsb.org/structure/6X3S. Accessed: 2024-08-29.

[27] Jeong Joo Kim et al. 2020. Structure of 6X3T: Human GABAA Receptor Alpha1-Beta2-Gamma2 Subtype in Complex with GABA and Propofol. https://www.rcsb.org/structure/6X3T. Accessed: 2024-08-29.

[28] Jeong Joo Kim et al. 2020. Structure of 6X3U: Human GABAA Receptor Alpha1-Beta2-Gamma2 Subtype in Complex with GABA Plus Flumazenil. https://www.rcsb.org/structure/6X3U. Accessed: 2024-08-29.

[29] Jeong Joo Kim et al. 2020. Structure of 6X3V: Human GABAA Receptor Alpha1-Beta2-Gamma2 Subtype in Complex with GABA and Etomidate. https://www.rcsb.org/structure/6X3V. Accessed: 2024-08-29.

[30] Jeong Joo Kim et al. 2020. Structure of 6X3W: Human GABAA Receptor Alpha1-Beta2-Gamma2 Subtype in Complex with GABA and Phenobarbital. https://www.rcsb.org/structure/6X3W. Accessed: 2024-08-29.

[31] Jeong Joo Kim et al. 2020. Structure of 6X3X: Human GABAA Receptor Alpha1-Beta2-Gamma2 Subtype in Complex with GABA and Diazepam. https://www.rcsb.org/structure/6X3X. Accessed: 2024-08-29.

[32] Jeong Joo Kim et al. 2020. Structure of 6X3Z: Human GABAA Receptor Alpha1-Beta2-Gamma2 Subtype in Complex with GABA. https://www.rcsb.org/structure/6X3Z. Accessed: 2024-08-29.

[33] Marc Baaden et al. 2013–present. UnityMol. https://unity.mol3d.tech/. Accessed: 2024-08-29.

[34] Nathan A. Baker et al. 2018. Poisson-Boltzmann Solver. https://www.poissonboltzmann.org/. Accessed: 2024-08-29.

[35] Radovan Spurny et al. 2012. Structure of 4A98: X-ray Structure of a Pentameric Ligand-Gated Ion Channel from Erwinia chrysanthemi in Complex with Bromoflurazepam. https://www.rcsb.org/structure/4A98. Accessed: 2024-08-29.

[36] Ramin Fahimi and Neil DB Bruce. 2021. On Metrics for Measuring Scanpath Similarity. *Behavior Research Methods* 53 (2021), 609–628. https://doi.org/10.3758/s13428-020-01441-0

[37] Adam Gardner, Ludovic Autin, Brett Barbaro, Arthur J. Olson, and David S. Goodsell. 2018. CellPAINT: Interactive Illustration of Dynamic Mesoscale Cellular Environments. *IEEE Computer Graphics and Applications* 38, 6 (2018), 51–66. https://doi.org/10.1109/MCG.2018.2877076

[38] Danilo Gasques, Janet G Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, et al. 2021. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14. https://doi.org/10.1145/3411764.3445576

[39] Exit Games GmbH. 2011–present. Photon Unity Networking (PUN). https://www.photonengine.com/pun. Accessed: 2024-08-29.

[40] Exit Games GmbH. 2014–present. Photon Voice. https://doc.photonengine.com/voice/current/getting-started/voice-intro. Accessed: 2024-08-29.

[41] Thomas D Goddard, Alan A Brilliant, Thomas L Skillman, Steven Vergenz, James Tyrwhitt-Drake, Elaine C Meng, and Thomas E Ferrin. 2018. Molecular Visualization on the Holodeck. *Journal of molecular biology* 430, 21 (2018), 3982–3996. https://doi.org/10.1016/j.jmb.2018.06.040

[42] Thomas D Goddard, Conrad C Huang, Elaine C Meng, Eric F Pettersen, Gregory S Couch, John H Morris, and Thomas E Ferrin. 2018. UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis. *Protein Science* 27, 1 (2018), 14–25. https://doi.org/10.1002/pro.3235

[43] Kunal Gupta, Gun A Lee, and Mark Billinghurst. 2016. Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE transactions on visualization and computer graphics* 22, 11 (2016), 2413–2422. https://doi.org/10.1109/TVCG.2016.2593778

[44] Robert M Hanson. 2010. Jmol – A Paradigm Shift in Crystallographic Visualization. *Journal of Applied Crystallography* 43, 5 (2010), 1250–1260. https://doi.org/10.1107/S0021889810030256

[45] Sandra G Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908. https://doi.org/10.1177/154193120605000909

[46] Keita Higuch, Ryo Yonetani, and Yoichi Sato. 2016. Can Eye Help You? Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5180–5190. https://doi.org/10.1145/2858036.2858438

[47] MA Hoffman and JB Provance. 2017. Visualization of Molecular Structures using HoloLens-based Augmented Reality. *AMIA Summits on Translational Science Proceedings* 2017 (2017), 68. https://doi.org/article/med/28815109

[48] William Humphrey, Andrew Dalke, and Klaus Schulten. 1996. VMD: Visual Molecular Dynamics. *Journal of molecular graphics* 14, 1 (1996), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

[49] Junko Ichino, Masahiro Ide, Takehito Yoshiki, Hitomi Yokoyama, Hirotoshi Asano, Hideo Miyachi, and Daisuke Okabe. 2023. How Gaze Visualization Facilitates Initiation of Informal Communication in 3D Virtual Spaces. *ACM Transactions on Computer-Human Interaction* 31, 1 (2023), 1–32. https://doi.org/10.1145/3617368

[50] Alexander D Jamieson-Binnie, Michael B O'Connor, Jonathan Barnoud, Mark D Wonnacott, Simon J Bennie, and David R Glowacki. 2020. Narupa iMD: A VR-Enabled Multiplayer Framework for Streaming Interactive Molecular Simulations. In *ACM SIGGRAPH 2020 Immersive Pavilion*. ACM, 1–2. https://doi.org/10.1145/3388536.3407891

[51] Allison Jing, Gun Lee, and Mark Billinghurst. 2022. Using Speech to Visualise Shared Gaze Cues in MR Remote Collaboration. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 250–259. https://doi.org/10.1109/VR51125.2022.00044

[52] Allison Jing, Kieran May, Gun Lee, and Mark Billinghurst. 2021. Eye See What You See: Exploring How Bi-Directional Augmented Reality Gaze Visualisation Influences Co-Located Symmetric Collaboration. *Frontiers in Virtual Reality* 2 (2021), 79. https://doi.org/10.3389/frvir.2021.697367

[53] Allison Jing, Kieran May, Brandon Matthews, Gun Lee, and Mark Billinghurst. 2022. The Impact of Sharing Gaze Behaviours in Collaborative Mixed Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27. https://doi.org/10.1145/3555564

[54] Allison Jing, Kieran William May, Mahnoor Naeem, Gun Lee, and Mark Billinghurst. 2021. eyemR-Vis: Using Bi-Directional Gaze Behavioural Cues to Improve Mixed Reality Remote Collaboration. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7. https://doi.org/10.1145/3411763.3451844

[55] Angus PR Johnston, James Rae, Nicholas Ariotti, Benjamin Bailey, Andrew Lilja, Robyn Webb, Charles Ferguson, Sheryl Maher, Thomas P Davis, Richard I Webb, et al. 2018. Journey to the Centre of the Cell: Virtual Reality Immersion into Scientific Data. *Traffic* 19, 2 (2018), 105–110. https://doi.org/10.1111/tra.12538

[56] Jordi Juárez-Jiménez, Philip Tew, Michael O' Connor, Salome Llabrés, Rebecca Sage, David Glowacki, and Julien Michel. 2020. Combining Virtual Reality Visualization with Ensemble Molecular Dynamics to Study Complex Protein Conformational Changes. *Journal of Chemical Information and Modeling* 60, 12 (2020), 6344–6354. https://doi.org/10.1021/acs.jcim.0c00221

[57] Laura J Kingsley, Vincent Brunet, Gerald Lelais, Steve McCloskey, Kelly Milliken, Edgardo Leija, Stephen R Fuhs, Kai Wang, Edward Zhou, and Glen Spraggon. 2019. Development of a Virtual Reality Platform for Effective Communication of Structural Data in Drug Discovery. *Journal of Molecular Graphics and Modelling* 89 (2019), 234–241. https://doi.org/10.1016/j.jmgm.2019.03.010

[58] Barbora Kozlíková, Michael Krone, Martin Falk, Norbert Lindow, Marc Baaden, Daniel Baum, Ivan Viola, Julius Parulek, and H-C Hege. 2017. Visualization of Biomolecular Structures: State of the Art Revisited. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 178–204. https://doi.org/10.1111/cgf.13072

[59] Ayush Kumar, Neil Timmermans, Michael Burch, and Klaus Mueller. 2019. Clustered Eye Movement Similarity Matrices. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–9. https://doi.org/10.1145/3317958.3319811

[60] David Kut'ák, Matias Nicolas Selzer, Jan Byska, María Luján Ganuza, Ivan Barisic, Barbora Kozlikova, and Haichao Miao. 2021. Vivern - A Virtual Environment for Multiscale Visualization and Modeling of DNA Nanostructures. *IEEE Transactions on Visualization and Computer Graphics* (2021). https://doi.org/10.1109/TVCG.2021.3106328

[61] David Kut'ák, Pere-Pau Vázquez, Tobias Isenberg, Michael Krone, Marc Baaden, Jan Byška, Barbora Kozlíková, and Haichao Miao. 2023. State of the Art of Molecular Visualization in Immersive Virtual Environments. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library. https://doi.org/10.1111/cgf.14738

[62] Grete Helena Kütt, Kevin Lee, Ethan Hardacre, and Alexandra Papoutsaki. 2019. Eye-Write: Gaze Sharing for Collaborative Writing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3290605.3300727

[63] Joseph Laureanti, Juan Brandi, Elvis Offor, David Engel, Robert Rallo, Bojana Ginovska, Xavier Martinez, Marc Baaden, and Nathan A Baker. 2020. Visualizing Biomolecular Electrostatics in Virtual Reality with UnityMol-APBS. *Protein Science* 29, 1 (2020), 237–246. https://doi.org/10.1002/pro.3773

[64] Oscar Legetth, Johan Rodhe, Stefan Lang, Parashar Dhapola, Mattias Wallergård, and Shamit Soneji. 2021. CellexalVR: A Virtual Reality Platform to Visualize and Analyze Single-Cell Omics Data. *IScience* 24, 11 (2021), 103251. https://doi.org/10.1016/j.isci.2021.103251

[65] Hui Li and Xinru Wei. 2024. A Concise Review of Biomolecule Visualization. *Current Issues in Molecular Biology* 46, 2 (2024), 1318–1334. https://doi.org/10.3390/cimb46020084

[66] Jerry Li, Mia Manavalan, Sarah D'Angelo, and Darren Gergle. 2016. Designing Shared Gaze Awareness for Remote Collaboration. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. 325–328. https://doi.org/10.1145/2818052.2869097

[67] Xavier Martinez and Marc Baaden. 2021. UnityMol Prototype for FAIR Sharing of Molecular-Visualization Experiences: From Pictures in the Cloud to Collaborative Virtual Reality Exploration in Immersive 3D Environments. *Acta Crystallographica Section D: Structural Biology* 77, 6 (2021), 746–754. https://doi.org/10.1107/S2059798321002941

[68] Xavier Martinez, Michael Krone, Naif Alharbi, Alexander S Rose, Robert S Laramee, Sean O'Donoghue, Marc Baaden, and Matthieu Chavent. 2019. Molecular Graphics: Bridging Structural Biologists and Computer Scientists. *Structure* 27, 11 (2019), 1617–1623. https://doi.org/10.1016/j.str.2019.09.001

[69] Bernhard Maurer, Michael Lankes, Barbara Stiglbauer, and Manfred Tscheligi. 2016. EyeCo: Effects of Shared Gaze on Social Presence in an Online Cooperative Game. In *International Conference on Entertainment Computing*. Springer, 102–114. https://doi.org/10.1007/978-3-319-46100-7_9

[70] Christoph Müller, Michael Krone, Markus Huber, Verena Biener, Dominik Herr, Steffen Koch, Guido Reina, Daniel Weiskopf, and Thomas Ertl. 2018. Interactive Molecular Graphics for Augmented Reality using Hololens. *Journal of integrative bioinformatics* 15, 2 (2018). https://doi.org/10.1515/jib-2018-0005

[71] Joshua Newn, Fraser Allison, Eduardo Velloso, and Frank Vetere. 2018. Looks Can Be Deceiving: Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3173574.3173835

[72] Joshua Newn, Eduardo Velloso, Fraser Allison, Yomna Abdelrahman, and Frank Vetere. 2017. Evaluating Real-time Gaze Representations to Infer Intentions in Competitive Turn-based Strategy Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 541–552. https://doi.org/10.1145/3116595.3116624

[73] Mathieu Noizet, Valentine Peltier, Hervé Deleau, Manuel Dauchez, Stéphanie Prévost, and Jessica Jonquet-Prévoteau. 2021. A Collaborative Molecular Graphics Tool for Knowledge Dissemination with Augmented Reality and 3D Printing. In *Workshop MolVA EuroVis2021*. https://doi.org/10.2312/molva.20211071

[74] Magnus Norrby, Christoph Grebner, Joakim Eriksson, and Jonas Bostrom. 2015. Molecular Rift: Virtual Reality for Drug Designers. *Journal of chemical information and modeling* 55, 11 (2015), 2475–2484. https://doi.org/10.1021/acs.jcim.5b00544

[75] Michael B O'Connor, Simon J Bennie, Helen M Deeks, Alexander Jamieson-Binnie, Alex J Jones, Robin J Shannon, Rebecca Walters, Thomas J Mitchell, Adrian J Mulholland, and David R Glowacki. 2019. Interactive Molecular Dynamics in Virtual Reality from Quantum Chemistry to Drug Binding: An Open-Source Multi-Person Framework. *The Journal of chemical physics* 150, 22 (2019), 220901. https://doi.org/10.1063/1.5092590

[76] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. 2004. UCSF Chimera — A Visualization System for Exploratory Research and Analysis. *Journal of computational chemistry* 25, 13 (2004), 1605–1612. https://doi.org/10.1002/jcc.20084

[77] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2017. CoVAR: Mixed-Platform Remote Collaborative Augmented and Virtual Realities System with Shared Collaboration Cues. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, 218–219. https://doi.org/10.1109/ISMAR-Adjunct.2017.72

[78] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2019. The Effects of Sharing Awareness Cues in Collaborative Mixed Reality. *Frontiers in Robotics and AI* 6 (2019), 5. https://doi.org/10.3389/frobt.2019.00005

[79] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13. https://doi.org/10.1145/3173574.3173620

[80] Yitoshee Rahman, Sarker M Asish, Nicholas P Fisher, Ethan C Bruce, Arun K Kulshreshth, and Christoph W Borst. 2020. Exploring Eye Gaze Visualization Techniques for Identifying Distracted Students in Educational VR. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 868–877. https://doi.org/10.1109/VR46266.2020.00009

[81] Fabio Cortés Rodríguez, Matteo Dal Peraro, and Luciano A Abriata. 2022. Online Tools to Easily Build Virtual Molecular Models for Display in Augmented and Virtual Reality on the Web. *Journal of Molecular Graphics and Modelling* 114 (2022), 108164. https://doi.org/10.1016/j.jmgm.2022.108164

[82] Rhoslyn Roebuck Williams, Xan Varcoe, Becca R Glowacki, Ella M Gale, Alexander Jamieson-Binnie, and David R Glowacki. 2020. Subtle Sensing: Detecting Differences in the Flexibility of Virtually Simulated Molecular Objects. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8. https://doi.org/10.1145/3334480.3383026

[83] Bertrand Schneider and Roy Pea. 2017. Real-time Mutual Gaze Perception Enhances Collaborative Learning and Collaboration Quality. In *Educational media and technology yearbook*. Springer, 99–125. https://doi.org/10.1007/978-3-319-45001-8_7

[84] Kshitij Sharma, Sarah D'Angelo, Darren Gergle, and Pierre Dillenbourg. 2016. Visual Augmentation of Deictic Gestures in MOOC Videos. In *Proceedings of the 12th International Conference of the Learning Sciences, ICLS 2016*. Singapore: International Society of the Learning Sciences, 202–209. https://doi.org/bitstream/1/117/1/28.pdf

[85] Jitin Singla, Kylie Burdsall, Brian Cantrell, Jordan R Halsey, Alex McDowell, Colleen McGregor, Sanraj Mittal, Raymond C Stevens, Shaoyu Su, Alexandra Thomopoulos, et al. 2022. A New Visual Design Language for Biological Structures in a Cell. *Structure* 30, 4 (2022), 485–497. https://doi.org/10.1016/j.str.2022.01.002

[86] Oleg Špakov, Diederick Niehorster, Howell Istance, Kari-Jouko Räihä, and Harri Siirtola. 2019. Two-way Gaze Sharing in Remote Teaching. In *IFIP Conference on Human-Computer Interaction*. Springer, 242–251. https://doi.org/10.1007/978-3-030-29384-0_16

[87] Sujith Sritharan, Raphaelle Versini, Jules Petit, Emmanuelle Bayer, and Antoine Taly. 2023. Prediction of A. Thaliana's MCTP4 Structure using Deep Learning-based Tools and Exploration of Transmembrane Domain Dynamics using Coarse-Grained Molecular Dynamics Simulations. *bioRxiv* (2023), 2023–08. https://doi.org/10.1101/2023.08.04.552001

[88] Nicola Victoria Maria Stein. 2023. *Navigating Pyoverdine and Beyond: The Role of Tripartite Efflux Pumps in Pseudomonas Putida KT2440.* Ph. D. Dissertation. LMU. https://edoc.ub.uni-muenchen.de/32605/2/Stein_Nicola.pdf.

[89] William Steptoe, Oyewole Oyekoya, Alessio Murgia, Robin Wolff, John Rae, Estefania Guimaraes, David Roberts, and Anthony Steed. 2009. Eye Tracking for Avatar Eye Gaze Control During Object-focused Multiparty Interaction in Immersive Collaborative Virtual Environments. In *2009 IEEE virtual reality conference*. IEEE, 83–90. https://doi.org/10.1109/VR.2009.4811003

[90] Veronica Sundstedt and Valeria Garro. 2022. A Systematic Review of Visualization Techniques and Analysis Tools for Eye-Tracking in 3D Environments. *Frontiers in neuroergonomics* 3 (2022), 910019. https://doi.org/10.3389/fnrgo.2022.910019

[91] Jmol Development Team. 2012–present. JSmol. https://jmol.sourceforge.net/. Accessed: 2024-08-29.

[92] Raphaëlle Versini, Marc Baaden, Mickael M Cohen, Laetitia Cavellini, Antoine Taly, and Patrick FJ Fuchs. 2023. Lys716 in the Transmembrane Domain of Yeast Mitofusin Fzo1 Modulates Anchoring and Fusion. *bioRxiv* (2023), 2023–09. https://doi.org/10.1101/2023.09.22.559002

[93] Michael Wiebrands, Chris J Malajczuk, Andrew J Woods, Andrew L Rohl, and Ricardo L Mancera. 2018. Molecular Dynamics Visualization (MDV): Stereoscopic 3D Display of Biomolecular Structure and Interactions Using the Unity Game Engine. *Journal of integrative bioinformatics* 15, 2 (2018), 20180010. https://doi.org/10.1515/jib-2018-0010

[94] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses using only ANOVA Procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146. https://doi.org/10.1145/1978942.1978963

[95] Kui Xu, Nan Liu, Jingle Xu, Chunlong Guo, Lingyun Zhao, Hong-Wei Wang, and Qiangfeng Cliff Zhang. 2021. VRmol: An Integrative Web-based Virtual Reality System to Explore Macromolecular Structure. *Bioinformatics* 37, 7 (2021), 1029–1031. https://doi.org/10.1093/bioinformatics/btaa696

[96] Stephanie Yang, Amreen Amin Poonawala, Tian-Shun Allan Jiang, and Bertrand Schneider. 2023. Can Synchronous Code Editing and Awareness Tools Support Remote Tutoring? Effects on Learning and Teaching. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–30. https://doi.org/10.1145/3610177

[97] Jimmy F Zhang, Alex R Paciorkowski, Paul A Craig, and Feng Cui. 2019. BioVR: A Platform for Virtual Reality Assisted Biological Data Integration and Visualization. *BMC bioinformatics* 20, 1 (2019), 1–10. https://doi.org/10.1186/s12859-019-2666-z

[98] Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017. Look Together: Using Gaze for Assisting Co-located Collaborative Search. *Personal and Ubiquitous Computing* 21, 1 (2017), 173–186. https://doi.org/10.1007/s00779-016-0969-x

[99] Min Zheng and Mark P Waller. 2017. ChemPreview: An Augmented Reality-based Molecular Interface. *Journal of Molecular Graphics and Modelling* 73 (2017), 18–23. https://doi.org/10.1016/j.jmgm.2017.01.019