Eurographics Workshop on Visual Computing for Biology and Medicine (2022)
J. Byška and T. Torsney-Weir (Poster Chairs)

Comparative Visual Analysis of Molecular Dynamics

M. Bok[1], M. Schäfer[1,2], N. Brich[1,2], K. Schreiner[3], V. Fäßler[3], M. Keckeisen[3], O. Kohlbacher[1,2,4,5], and M. Krone[1,2]

[1]Department of Computer Science, University of Tübingen, Tübingen, Germany
[2]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany
[3]TWT GmbH Science & Innovation, Stuttgart, Germany
[4]Quantitative Biology Center, University of Tübingen, Tübingen, Germany
[5]Institute for Translational Bioinformatics, University Hospital Tübingen, Tübingen, Germany

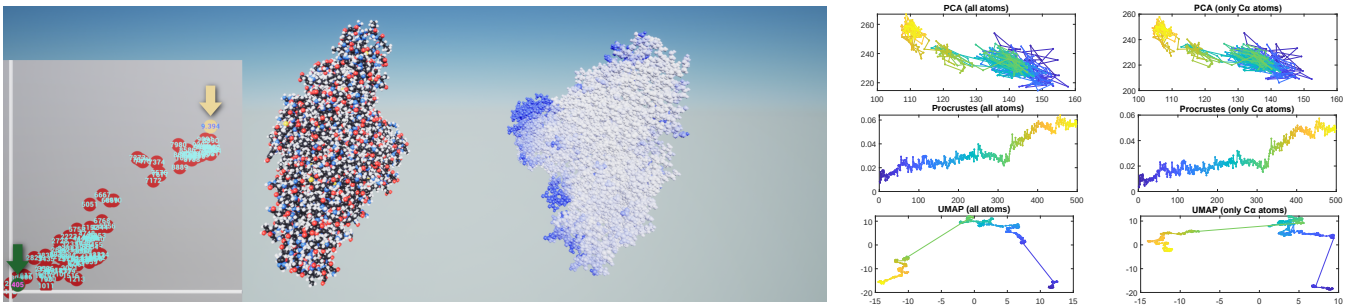

Figure 1: *Left: Example of our comparative visualization using a molecular simulations of the SARS-CoV-2 spike protein. The scatterplot shows a PCA-based dimensionality reduction for all time steps of the simulation, where each dot is one time step. Two clusters are observable, reflecting the initially closed state and the open state at the end of the simulation. From each of the clusters, one time step was selected (green and yellow dots/arrows). The two protein conformations corresponding to these time steps are visualized as spacefilling models. One of these 3D models is colored by the chemical element, the other one by the difference using a white-to-blue gradient (white: no diff., blue: high diff.). The 3D views can be zoomed and rotated individually or in a coordinated fashion. Right: Comparison of different methods (rows: PCA, Procrustes, UMAP) applied to all atoms (left column) and only the Cα atom of each amino acid (colored by timestep: 1 n).*

Abstract

We present a visual analysis application for Molecular Dynamics simulations, which facilitates the analysis of conformational changes of a protein and allows users to visually compare different time points. We employ dimensionality reduction to project the protein to a 2D domain for each time step based on its atom positions. A scatter plot shows the change of conformation during the simulation: the distance in the embedding domain reflects the conformational difference. Thus, users can easily assess whether there are one or more semi-stable conformations by looking for clusters. For an in-depth analysis, users can pick two time points and compare the 3D structures of the proteins in a side-by-side view. Our application was built using the TRONIS® environment, which is based on Unreal Engine 4. We tested it using multiple simulation data sets.

CCS Concepts

• ***Human-centered computing*** → ***Visualization systems and tools;*** • ***Applied computing*** → ***Molecular structural biology;***

1. Introduction

Understanding how proteins and other biological molecules behave under different conditions is a crucial step in many applications like protein engineering or drug design. This includes environmental parameters like different temperatures or solvent mixtures, but also mutations of the protein itself. Molecular Dynamics (MD) simulation is an established method to investigate this behavior, which can give insights about functional aspects like the flexibility of a protein or the catalytic rate of an enzyme. The output of a MD simulation is a so-called trajectory data set, which provides the position of each atom for each time step. The output data can be in the order of tens to hundreds of thousands of time steps, resulting in gigabytes of binary data. Analyzing these simulation data is, thus, a challenging task. Visualization is an established means to facilitate the analysis. Many different molecular representations like ball-and-stick models, molecular surfaces, or abstract "cartoon" representations that show different properties of the simulated proteins have been proposed [KKF*17]. More advanced visual analysis methods can extract and visualize certain features—for example, cavities containing binding sites [KKL*16]—to further support the analysts.

However, the length of the simulations still makes it difficult to assess the dynamic behavior of the simulated molecules. Therefore, several methods have been proposed that summarize the simulation or highlight interesting parts (e.g., [KRS*13, BTM*19]).

We present an application that gives an overview of the simulation by showing how the so-called conformation of a protein changes during the simulation. The conformation is the three-dimensional arrangement of the amino acid chain of the protein. Changes in the conformation can have wide-ranging effects on the function of a protein. Consider, for example, a protein that is embedded into a cell membrane and forms an ion channel. A deformation of the protein that leads to an opening or closing of the channel would be such a conformational change. While such conformational changes are often rather subtle even though they can have a large effect on the function, an extreme case would be the complete unfolding of a protein chain due to high temperature.

2. Methods and Implementation

Our goal was to facilitate the analysis of conformational changes of a protein during the simulation and to allow users to identify and visually compare interesting time points on demand. We employ dimensionality reduction (DR) approaches that project the protein for each time steps to a 2D domain based on its atomic positions. While this approach has been applied before, e.g., using established methods like PCA, t-SNE, or UMAP [How01, TG19, TWT21], it is not yet a standard procedure for analyzing protein simulations in the available molecular visualization tools. We chose and compared different DR methods. The resulting scatter plot effectively shows the evolution of the protein conformation over time. The distance in the embedding domain reflects the conformational difference, that is, clusters in the plot highlight similar conformations. Consequently, users can easily assess whether there are one or more semi-stable conformations over time. Thus, DR has an advantage over simple methods like Procrustes analysis, which is offered by most molecular visualization tools using RMSD. For an in-depth analysis, users can pick two time points and compare the 3D structures of the proteins in a side-by-side view. This protein visualization currently only supports the spacefilling model, where each atom is represented as a sphere that is usually colored by its chemical element. As shown in Figure 1, the second protein is colored by the difference (Euclidean distance between the atom position of the two time steps) using a white-to-blue gradient, where white denotes no difference and blue a high difference. The 3D views can be zoomed and rotated individually or in a coordinated fashion.

Our application was built using TRONIS®† [KD17], an environment for virtual prototyping and validation of driver assistance systems developed by TWT GmbH, which is based on the Unreal Engine 4‡. TRONIS® is designed to be easily extensible through plugins that communicate via socket connections. The data loading and DR is implemented as such a plugin and then streams the data to the main TRONIS® application for visualization on demand.

† https://twt-gmbh.de/en/produkte/, accessed: 08/19/2022.
‡ https://www.unrealengine.com, accessed: 08/19/2022.

3. Summary and Outlook

Our current implementation is still work in progress, however, all described functionality is already available. It also shows that TRONIS®, which was initially designed for automotive simulations, was easily adaptable to molecular data. We tested our application using various MD simulations of different proteins, e.g., the SARS-CoV-2 spike protein simulation [D. 20] shown in Figure 1. With the combination of 2D and 3D views, users can easily analyze how the conformation of the protein changes over time and the overall stability. This goes beyond the capabilities of common molecular visualization tools [KKF*17] and is a promising basis for future developments.

On our poster, we will describe the technical details as well as the capabilities of our current, prototypical implementation. We will also show the comparison results for the different DR methods (PCA, UMAP, Procrustes analysis; see Figure 1 right) and the various data sets that we used to evaluate our application. Furthermore, we will discuss directions for planned future work such as supporting further molecular representations for the 3D view, improving the user interface, and extending the analysis capabilities (e.g., by adding filtering and clustering for the scatter plot).

Acknowledgments: This work was partially funded by Carl Zeiss Stiftung and by German Research Foundation (DFG) as part of project PROLINT (project ID 391088465).

References

[BTM*19] Byška J., Trautner T., Marques S., Damborský J., Kozlíková B., Waldner M.: Analysis of Long Molecular Dynamics Simulations Using Interactive Focus+Context Visualization. *CGF 38*, 3 (2019), 441–453. doi:10.1111/cgf.13701. 2

[D. 20] D. E. Shaw Research: Molecular Dynamics Simulations Related to SARS-CoV-2, 2020. Technical Data. URL: http://www.deshawresearch.com/resources_sarscov2.html. 2

[How01] Howe P. W.: Principal components analysis of protein structure ensembles calculated using NMR data. *J. Biomol. NMR 20* (2001), 61–70. doi:10.1023/A:1011210009067. 2

[KD17] Kufieta K., Ditze M.: A virtual environment for the development and validation of highly automated driving systems. In *17. Internationales Stuttgarter Symposium* (2017), Springer Fachmedien, pp. 1391–1401. doi:10.1007/978-3-658-16988-6_106. 2

[KKF*17] Kozlíková B., Krone M., Falk M., Lindow N., Baaden M., Baum D., Viola I., Parulek J., Hege H.-C.: Visualization of Biomolecular Structures: State of the Art Revisited. *CGF 36*, 8 (2017), 178–204. doi:10.1111/cgf.13072. 1, 2

[KKL*16] Krone M., Kozlíkova B., Lindow N., Baaden M., Baum D., Parulek J., Hege H.-C., Viola I.: Visual Analysis of Biomolecular Cavities: State of the Art. *CGF 35*, 3 (2016), 527–551. doi:10.1111/cgf.12928. 1

[KRS*13] Krone M., Reina G., Schulz C., Kulschewski T., Pleiss J., Ertl T.: Interactive Extraction and Tracking of Biomolecular Surface Features. *CGF 32*, 3 (2013), 331–340. doi:10.1111/cgf.12120. 2

[TG19] Tribello G. A., Gasparotto P.: Using Dimensionality Reduction to Analyze Protein Trajectories. *Front. Mol. Biosci. 6* (2019). doi:10.3389/fmolb.2019.00046. 2

[TWT21] Trozzi F., Wang X., Tao P.: UMAP as Dimensionality Reduction Tool for Molecular Dynamics Simulations of Biomacromolecules: A Comparison Study. *J. Phys. Chem. B 125*, 19 (2021), 5022–5034. doi:10.1021/acs.jpcb.1c02081. 2