RESEARCH

Comparative Visualization of Protein Secondary Structures

Lucia Kocincová[1*], Miroslava Jarešová[1], Jan Byška[1,2], Július Parulek[2], Helwig Hauser[2] and Barbora Kozlíková[1]

Abstract

Background: Protein function is determined by many factors, namely by its constitution, spatial arrangement, and dynamic behavior. Studying these factors helps the biochemists and biologists to better understand the protein behavior and to design proteins with modified properties. One of the most common approaches to these studies is to compare the protein structure with other molecules and to reveal similarities and differences in their polypeptide chains.

Results: We support the comparison process by proposing a new visualization technique that bridges the gap between traditionally used 1D and 3D representations. By introducing the information about mutual positions of protein chains into the 1D sequential representation the users are able to observe the spatial differences between the proteins without any occlusion commonly present in 3D view. Our representation is designed to serve namely for comparison of multiple proteins or a set of time steps of molecular dynamics simulation.

Conclusions: The novel representation is demonstrated on two case studies. The first study aims to compare a set of proteins from the family of cytochromes P450 where the position of the secondary structures has a significant impact on the substrate channeling. The second study focuses on the protein flexibility when by comparing a set of time steps our representation helps to reveal the most dynamically changing parts of the protein chain.

Keywords: Molecular Sequence Analysis; Molecular Structural Biology; Computational Proteomics

Background

Our method aims to cover the gap between two commonly used representations of molecular structure. Moreover, the proposed solution serves for the comparison of multiple protein structures. We focus on structure alignment which finds the best fitting of protein chains and align them with respect to the root mean square deviation (we are using the Combinatorial Extensions algorithm). When studying the similarity between more proteins or the differences in whole protein families, the 3D representation is inapplicable because of the visual clutter caused by many intersections (image). This is very true even for small sets of proteins to compare (for very similar structures 5 and more, for structures with lot of differences 3 and more). Thus, in these cases the traditional sequential 1D information is used (image). Here the individual amino acids are aligned (TODO describe structure and sequence alignment and their difference?) and when each row of the representation represents one protein structure, the user can observe the similarities by studying the columns. Some methods allow to integrate the information about the secondary structures in these chains. However, the information about the spatial orientation of the corresponding secondary structures is completely lost. In our representation we utilize the fact that the users are trained to understand the sequential representation well but we add another crucial information about the mutual position of corresponding secondary structures. So our representation helps to reveal not only the differences in the sequence and length of corresponding secondary structures but also the differences in the spatial orientation of the corresponding SS. This is impossible to reach using the traditionally used 1D representation and reaching it in 3D representation is very complex task for large datasets because the user has to study the mutual alignment of proteins from this dataset in smaller portions and merge the gained information manually. Strong points of our representation:

- the domain experts are accustomed with the sequential representation as well as with secondary structures

*Correspondence: lucia.koc@gmail.com

[1]Masaryk University, Brno, Czech Republic
Full list of author information is available at the end of the article

- interactive manipulation
- predefined ranking according to RMSD
- changing superimposition and juxtaposition
- linking with 3D
- automatic highlighting of the most similar and the most different secondary structures
- filtering (removing the most different proteins)
- coloring of SS in the abstracted representation with respect to the physico-chemical properties of the amino acids

Related work: [**?**], maybe also [**?**]

Methodology

TODO

Design

Implementation

With the use of the Combinatorial Extension (CE) algorithm for structure alignment [**?**] and force layout algorithm, we let the structures to align in 3D, so there is no deformation caused by choosing a particular projection or distortion. At the end, we visualize the flattened molecules as if they were stretched out from 3D by pulling the chosen reference molecule at its ends into a straight line, so the actual length of secondary structures is preserved as well as the position of near structures of other molecules which are "locked" to the reference molecule.

Algorithm

Interaction

The 3D view and 2D view are both interactive – basic information about the secondary structure is shown when mouse is moving over the visualizations and a structure is highlighted in green when a structure is selected by a mouse click. Moreover, the 3D and 2D views are interconnected creating a unique way to explore the molecules – when any of the structure is clicked on, the highlight is visible in both views. This feature gives the user very important context of the actual spatial positions of the selected structures and enables him or her to interact with both views independtly, yet still in context.

Results and Discussion

TODO

Case Study

Conclusions

TODO

Future work

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Masaryk University, Brno, Czech Republic. [2]University of Bergen, Norway.

References

1. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D.: The human genome browser at ucsc. Genome Research **12**(6), 996–1006 (2002)

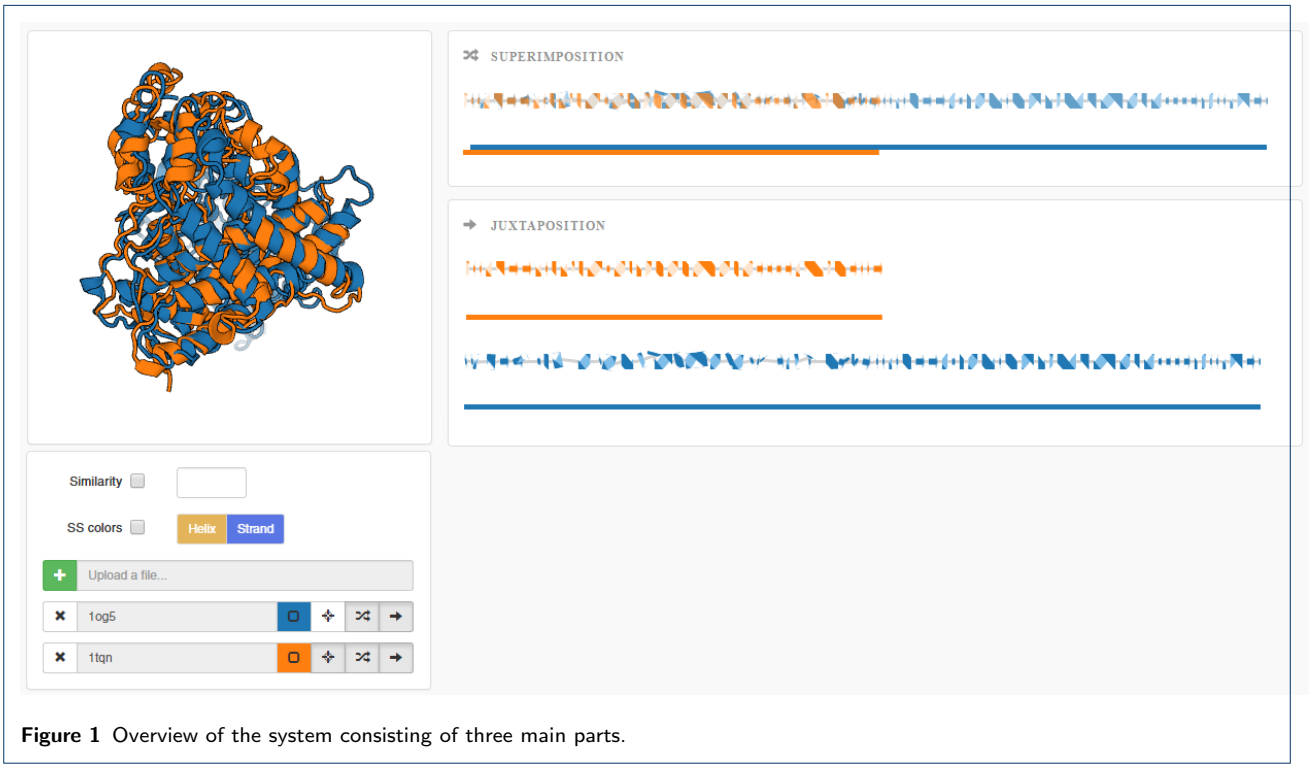

Figure 1 Overview of the system consisting of three main parts.