VERMONT: A tool for mutation visualization

Sócrates S. Araújo Jr.[1], Samuel S. Guimarães[1], Alexandre V. Fassio[2], Raquel C. de Melo-Minardi[2], Sabrina de A. Silveira[1]

[1]*Universidade Federal de Viçosa,* [2]*Universidade Federal de Minas Gerais*

Mutations are events that occur naturally due the evolution, changing the sequence of residues, which can possibly affect protein structure and function. Thus, an important open problem in Bioinformatics is to understand how these specific mutations in protein residues can affect or not on protein function. To tackle this problem, VERMONT (Visualization Mutation Tool) was proposed in the contest of IEEE BioVis 2013, where it received the Biology Experts Pick award and its paper was published on BMC Proceedings journal. At that time, VERMONT was a static tool that allowed users to study the impacts of a mutation on a specific dataset provided in the contest. Now, we present a generic version of VERMONT, which allows users to set up their own set of proteins to be analyzed. To start the analysis user can provide his/her own files in .pdb format, or the user can choose to get files directly from the Protein Data Bank (PDB). Next, a structural alignment of all protein family is computed using MultiProt and then the similar sequences are grouped to help visualization of the residue conservation on the dataset with the Expectation Maximization algorithm. Then, we modeled the protein structures as graphs in two different levels of granularity to compute contacts through the Delaunay triangulation: (i) as atomic graphs, where each node is a protein atom and each edge represents the interactions between atoms and (ii) as residue graphs, where each node is a residue and each edge represents the interactions among a residue pair. In both levels, the nodes and edges are labeled with their physicochemical properties. To analyze the protein networks, we used some measures of complex networks (degree, closeness and betweenness) that helps us to determine how central and connected a node is in the network. Highly connected nodes can potentially cause more damage in the protein's function in case they suffer a mutation. We propose interactive visualization strategies to show all computed data, especially the contacts conservation all over the dataset, coupled with centrality measures and solvent accessibility for all residues, allowing users to get details on demand by clicking or passing the mouse over each residue.

This work has been supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).