Mutation Explorer: Understanding the Effect of Mutations in Proteins by Connecting Sequence and Structure

Sebastian Bittrich[1,2] and Dirk Labudde[1,*]

[1]University of Applied Sciences Mittweida, Technikumplatz 17, 09648, Mittweida, Germany and

[2]Biotechnology Center (BIOTEC), TU Dresden, Tatzberg 47-49, 01307, Dresden, Germany

[*]Corresponding author: dirk.labudde@hs-mittweida.de

The genome is prone to constant change. Some mutations will cause no difference on phenotypic level, while some lead to diseases such as hypertension [1] or phenylketonuria [2]. Understanding if and how some single-nucleotide polymorphism effects the phenotype is one of biology's central challenges [3].

Protein folding describes the arrangement of a nascent protein chain into a native, functional 3D structure. However, the possible structure space is inconceivably large and hard to explore. Evidently, the protein's sequence and its surroundings provide all necessary information to arrive at a native conformation [4]. In order to assess the influence of certain mutations, knowledge of the folded structure can contribute to unveiling effects on protein stability, ligand binding and catalysis [1].

Strategies such as homology modeling [5] or direct-coupling analysis [3] were designed to predict 3D structures for given, peculiar sequences. By their nature, they differ in accuracy, speed and how well they can model the influence of delicate sequence changes. Despite these tools at hand, the exact connection of sequence and structure remains enigmatic.

We present *Mutation Explorer* as web server to pinpoint the influence of changes in protein sequence on the 3D structure. Today, the PDB offers a sufficient number of data on groups of homologous protein chains which provide evidence on the structural impact of mutations. For these clusters the respective sequences are aligned using Clustal Omega [6]. An interactive 3D visualization of selected structures is provided [7], so the user can align complete chains or hover over mutated sequence regions to highlight them and compare small structural fragments of interest. That way, one can inspect whether a sequence change leads to a conformational shift.

All these sequence and structure information is enriched by annotated phenotypic effects (and supporting evidence) of mutations and sequence variants taken from the UniProt [8]. This data facilitates interpretation of previous findings as structural change may not manifest in the phenotype. In contrast, there are also cases where the correct fold is preserved after an amino acid substitution, yet the protein still proofs to be dysfunctional. The Protein-Ligand Interaction Profiler [9] provides another perspective to assess the influence of mutations as it annotates non-covalent interactions (e.g. hydrogen bonds or pi-stacking) regarding protein and potential ligands as well as interactions between amino acids. Disruptions in these interaction networks could either explain misfolded proteins or loss of function as substrates cannot be bound or processed like in the wild-type structure [1]. Thus, the user can assess mutations on sequence and structure level while getting more abstract feedback on how certain (mutated) amino acids interact with surrounding residues or ligands.

Mutation Explorer helps scientists and students alike to study how certain variations in protein sequence changes the fold by comparing their 3D structures to known homologous while integrating disease and amino acid interaction data.

References

[1] Zhen Wang and John Moult. Snps, protein structure, and disease. *Human mutation*, 17(4):263–270, 2001.

[2] E Bjørgo, P M Knappskog, A Martinez, R C Stevens, and T Flatmark. Partial characterization and three-dimensional-structural localization of eight mutations in exon 7 of the human phenylalanine hydroxylase gene associated with phenylketonuria. *European journal of biochemistry*, 257:1–10, October 1998.

[3] Thomas A Hopf, John B Ingraham, Frank J Poelwijk, Charlotta PI Schärfe, Michael Springer, Chris Sander, and Debora S Marks. Mutation effects predicted from sequence co-variation. *Nature Biotechnology*, 2017.

[4] C.B. Anfinsen and H.A. Scheraga. Experimental and theoretical aspects of protein folding. In *Advances in Protein Chemistry*, pages 205–300. Elsevier BV, 1975.

[5] Tianyun Liu, Grace W. Tang, and Emidio Capriotti. Comparative modeling: The state of the art and protein drug target structure prediction. *Combinatorial Chemistry & High Throughput Screening*, 14(6):532–547, jul 2011.

[6] Fabian Sievers, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology*, 7:539, October 2011.

[7] Marco Biasini. Pv - webgl-based protein viewer. Zenodo, 2015.

[8] UniProt Consortium et al. Uniprot: a hub for protein information. *Nucleic acids research*, page gku989, 2014.

[9] Sebastian Salentin, Sven Schreiber, V Joachim Haupt, Melissa F Adasme, and Michael Schroeder. Plip: fully automated protein–ligand interaction profiler. *Nucleic acids research*, 43(W1):W443–W447, 2015.