Automated homology modelling for genetic variant assessment with Rosetta

Igor Petrik, Min-Sun Park, Adam Chamberlin, and R. Bryn Fenwick

Structural and functional disruption of proteins is a major mechanism of known human genetic disease. While the structures of only ~12% of the 10 million residue positions in the human proteome have been determined experimentally, ~50% are covered by a structure with at least 30% local identity. To enable structure based variant assessment of proteins we have developed an automated pipeline for the construction of homology models. Here we describe our progress in this endeavor and highlight some ongoing challenges. We find that ~90% of classified pathogenic alterations lie at positions covered by templates with at least 30% identity. Furthermore, over 90% of alterations covered at this level are unclassified or of unknown significance, and these alterations account for 50% of clinically observed alterations. Therefore, the availability of accurate homology models provides a powerful advantage for assessing the impact of genetic alterations.