High-Precision Pathogenicity Prediction of PKLR Mutations Integrating 3D Structural Centroids and Evolutionary Conservation

Mahnoor Aziz Khan
Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

Abstract

Objective and Scope: The primary objective of this study was to develop a computational predictive tool to assess the pathogenicity of single amino acid substitutions within the human erythrocyte pyruvate kinase (PK-R) enzyme. The research specifically targeted missense mutations, which constitute the majority of pathogenic alleles in PKLR-associated Pyruvate Kinase Deficiency (PKD). A comprehensive literature review identified three critical path mechanisms driving enzyme dysfunction: thermodynamic instability leading to rapid protein degradation, structural disruption of the catalytic active site or allosteric regulatory interfaces, and impaired kinetic efficiency.

Methodology and Feature Engineering: To quantify the functional importance of specific residues, evolutionary conservation scores were generated by performing homology searches via BLAST, followed by Multiple Sequence Alignment (MSA) using MEGA software. To model the potential for functional site disruption, a structural feature set was engineered to calculate the 3D Euclidean distances between target residues and the centroids of the enzyme's active and allosteric binding sites, derived from the high-resolution crystal structure (PDB ID: 2VGB). While thermodynamic stability changes provide valuable mechanistic insights (particularly for stability mutants like p.R510Q), the computational resources required to simulate all possible substitutions across the 574-residue protein were prohibitive for this study; consequently, evolutionary conservation profiles were utilized as a high-throughput proxy for structural stability.

Implementation and Validation: A Python-based predictive algorithm was implemented to integrate these features. The tool accepts a specific amino acid position as input and generates a comprehensive pathogenicity profile for all 19 potential amino acid substitutions at that locus. For each substitution, the program reports the associated conservation score and structural distance metrics, culminating in a binary classification of "Pathogenic" or "Benign." The tool's performance was validated against a curated "ground truth" dataset derived from ClinVar and LOVD and benchmarked against established generalist predictors (SIFT and PolyPhen-2). Comparative analysis was conducted using robust statistical metrics for imbalanced datasets, specifically the Matthews Correlation Coefficient (MCC) and the F1 score, to demonstrate the superior specificity of this mechanism-aware approach. Comparative analysis showed that the tool significantly outperformed generalist predictors, which yielded lower specificities, SIFT (0.40) and PolyPhen-2 (0.60), due to their tendency to misclassify rare benign variants located in flexible domains. Our results confirm that structural proximity to functional centroids is a decisive feature for filtering false positives.

Conclusion: The integration of 3D structural context with high-throughput evolutionary conservation scores provided a robust framework for the functional annotation of PKLR mutations. This mechanism-aware approach not only enhances the reliability of molecular diagnostics for PKD but also serves as a scalable template for predicting pathogenicity in other structurally well-characterized enzymopathies.