3D-protein C mutation database: integration of structural, functional and clinical data of natural protein C mutants

D'Ursi P (1), Marino F (2), Caprera A (2), Milanesi L (2), Faioni E (3), Rovida E (2)

(1) Department of Science and Biomedical Technologies, University of Milano, Italy
(2) Institute of Biomedical Technologies, National Research Council, Segrate (Mi), Italy
(3) Hematology and Thrombosis Unit, Ospedale San Paolo, University of Milano, Italy

Motivation

Protein C (PC) is a vitamin K-dependent anticoagulant plasma serine protease that exerts its action through the inactivation of factors Va and VIIIa. In addition it plays an important role in inflammation and cell proliferation. Several mutations of the PC gene have been found in patients with protein C deficiency, a condition that is associated to the risk of developing venous thrombosis. PC deficiency determine a reduced plasma concentration and/or catalytic activity of the protein. In previous work, we have identified 33 mutations (18 novel) in the promoter and coding regions of the PC gene by PCR and sequencing in 46 patients reporting venous thromboembolic events. We have constructed the molecular models of missense mutations localized in the structurally resolved regions of PC, starting from PDB x-ray coordinates (pdb ID: 1aut) and have performed detailed computational analysis of those presenting aminoacidic substitution in critical positions for structure and function. The availability of structural models can be useful in the research and clinical fields to elucidate how a mutation may interfere with enzymatic activity, ligand binding and cofactors interaction and relate the effect to patient phenotype. Thus, we started to collect our data in a database that was conceived to be freely available for consultation and data retrieval. The database was then enriched with already described variants extracted from other sources (literature, SwissProt, HGMD). Our effort was to create an updated interactive tool to integrate clinical and phenotypical descriptions with functional and structural data obtained by computational approaches to help elucidate the chain of events leading from a molecular defect to pathology.

Methods

MySQL database management system and PHP web programming language were used to construct database and graphical interface. The database consists of 4 tables where all the informations about protein variants are stored. It contains 197 entries that include 184 missense and 13 stop mutations. Multiple alignments were obtained with CLUSTALW using a set of orthologous and paralogous sequences to show the residue conservation between species and within serine protease family. A in-house developed Perl script was used for process the CLUSTALW output, automatically producing for each variant an HTML-format multi-alignment file with the mutation highlighted in red. Molecular models were constructed manually in InsightII (Accelrys) or automatically with a modified script in Python of Modeller (Andrej Sali), both methods consisting in residue replacement and the conformation of the mutant side chain is optimized by coniugate gradient and refined using molecular dynamics for Python script. Structural models can be visualized as realtime 3D images by using RasMol and VRML (Virtual Reality Modeling Language). The functional and mutation sites are automatically mapped on the structure and highlighted with different colors. The output format for VRML was generated by MolScript v2.1.2 from the stored PDB coordinates.

Results

We have realized a specialized relational database and a search tool for natural mutants of Protein C. A query page allows the user to retrieve entries by position in sequence of a mutated residue, by aminoacid substitution, by keyword and by domain localization. The query results are listed in a table where each entry is linked to a details page. This page resumes data about gene, secondary structure and domain localization of the mutated residue. A structural data section reports multiple alignment files which highlights the substituted position and help to evaluate the degree of conservation, the coordinates of modelled variants, and a gallery of 3D images that illustrates the structural implications of the aminoacidic substitution as long as the results of computational

analysis, when available, like electrostatic potential representations and molecular dynamics trajectories. Clinical, phenotypic and functional annotations, manually extracted from the literature, are also reported. Direct links to relevant literature references and, when present, to the corresponding SwissProt Variant Page are included. The site provides tools to analyze and mutate new variants, not included in database. Given a position in sequence, it is possible to generate a multiple alignment highligthing the residue of interest in homologs and to mutate the protein using the program Modeller (A. Sali & T.L. Blundell.J. Mol. Biol. 234, 779-815, 1993) through an implemented Phyton script. 3D-protein C mutation database can be accessed at the site http://www.itb.cnr.it/procmd/ Acknowledgments This work was supported by European Project BioinfoGRID (Bioinformatics Application for Life Science).

Availability: http://www.itb.cnr.it/procmd/

Contact email: ermanna.rovida@itb.cnr.it