Poster I-48

Serine Proteases analysis based on phylogenetic trees constructed from the sequence and structure alignments

Authors:
Cristina Ribeiro (*UFMG*)
Paula Kuser (*EMBRAPA CNPTIA*)
Goran Neshich (*EMBRAPA CNPTIA*)
Marcelo Santoro (*UFMG*)

Short Abstract: We compared two approaches to construct phylogenetic trees from a set of serine proteases with deciphered 3D_structure. The first approach: sequence alignment generates a phylogenetic tree; the seccond approach is based on a sequence alignment strictly following the structure alignment. The differences b/w the two is discussed.

Long Abstract:

Almost one-third of all proteases can be classified as serine proteases, named for the nucleophilic Ser residue at the active site. This mechanistic class was originally distinguished by the presence of the Asp-His-Ser “charge relay” system or “catalytic triad”.The Asp-His-Ser triad can be found in at least four different structural contexts, indicating that this catalytic machinery has evolved on at least four separate occasions. These four clans of serine proteases are typified by chymotrypsin, subtilisin, carboxypeptidase Y, and Clp protease (MEROPS nomenclature). Serine proteases with the classic Asp-His-Ser triad are the largest class of proteases, including digestive enzymes with minimal especificity and processing enzymes with exquisite substrate recognition. These proteases can be found in eukaryotes, prokaryotes, archae, and viruses. Chymotrypsin-like proteases are involved in many critical physiological processes, including digestion, hemostasis, apoptosis, signal transduction, reproduction, and the immune response.

Objectives:

The main objective of this work is to compare two different approaches to construct phylogenetic trees from a set of non-redundant serine proteases with deciphered 3D structure. The first one the sequence alignment generates a phylogenetic tree, and the other one is based on a sequence alignment strictly following the structure alignment. The differences b/w the two is discussed.

Work interest

To our knowledge, there are very few papers where the authors have tried to generate phylogenetic approaches from both sequence alignment and structural homology and compared them in a systematic way to evaluate the results.

Methodology:

Database construction:

The first step to the construction of a non-redundant database of crystallized serine proteases was a query for all serine proteases in Protein Data Bank (http://www.rcsb.org/pdb/Welcome.do), followed by extensive manual curation to remove duplicates and single-amino acid mutants. After that we removed all sequences that didn't have the classical two barrel folding of serine proteases and the subtilisin subfamily, since it has a great sequence divergence from the traditional members of this family. After that we ended up with the list of 82 non-redundant serine protease structures ready for further analysis.

3D conserved residues based on structural alignment

The PrISM (Protein Informatics System for Modeling, Yang, 1999) is a protein analysis tool that allows, among other things, for a structural alignment. We analyzed our set of non-redundant serine proteases with PrISM software, generating a subset of the original full-length sequences containing only the regions structurally shared by them.
Resulting alignments (the sequence and structure based one) are then used to generate the two phylogenetic trees and those are discussed in details in this work.

Yang AS, Honig B. Sequence to structure alignment in comparative modeling using PrISM. Proteins. 1999;Suppl 3:66-72.