Protein Dossier and Protein Fingerprints produce Surface Signatures: A simple and powerful technique for coding and comparing compound and receptor shape information

Cristina Ribeiro[1], cris.lino@gmail.com
Goran Neshich[2] , neshich@cbi.cnptia.embrapa.br
Marcelo Matos Santoro[1], santoro@icb.ufmg.br

1 – Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil;
2 – National Agricultural Information Technology Research Center (CNPTIA - EMBRAPA), Campinas - SP, Brazil;

Abstract

Serine proteases are among the most intensively studied enzymes. Hans Neurath (Neurath, 1989) was among the first scientists to recognize that proteinases act not only as digestive enzymes, but also fulfill numerous other functions in the organism. The protein processing turns on or off numerous activities, which are, in turn, responsible for a large range of biological phenomena such as blood clotting, clot dissolution, protein hormone action, emergence of the silk moth from cocoons, penetration of outer layers of ova by sperm, differentiation, cell death and apoptosis.

Almost one-third of all proteases can be classified as serine proteases, named after 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 specificity 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. Chymotrypsin-like proteases are the most abundant in nature, with over 240 proteases recognized in the MEROPS database. The objective of this work is to analyze the structural interaction between serine proteases and theirs inhibitors using amino acids residues present in the interface of this molecules to discovery complex patterns of recognition and specificity.

In this work, we are mapping the interactive potential of these proteins. The differences between the serine protease complexes were discussed. After that we are mapping contacts in the interface of protein-protein and protein-ligant complexes.

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 did not 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, we had a list of 81 non-redundant serine proteases structures ready for further analysis. Therefore, we used the PrISM (Protein Informatics System for Modeling, Yang, 1999) that is a protein analysis tool that allows one to extract from a set of related sequences, the regions and individual amino acids that are conserved in the 3D structures of them, and structure alignment at all 81 proteins. So we analyzed our set of non-redundant serine proteases with PrISM software, generating a subset of the original structure containing only the regions structurally shared by them.

References

Murzin, a. G., brenner, s. E., hubbard, t., chothia, c. Scop: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536-540, 1995.

Berman, h. M., westbrook, j., feng, z., gilliland, g., bhat, t. N., weissig, h., shindyalov, i. N., bourne, p.e. The protein data bank. Nucleic acids research, 28 pp. 235-242 , 2000.

Yang, a.s., honig, b. Sequence to structure alignment in comparative modeling using prism. Proteins.;suppl 3:66-72, 1999.

Hedstrom, l. Serine protease mechanism and specificity. Chem rev. Dec;102(12):4501-24, 2002.