Topic: **Structural Bioinformatics and Molecular Dynamics**

USING SINGULAR VALUE DECOMPOSITION IN THE IDENTIFICATION OF PATTERNS IN SUBTILASE-INHIBITOR COMPLEXES

V Almeida[1], S Silveira[1], M Crepalde[2], M Santos[2], A Gomes[2], W Meira[2], D Pires[1], C Silveira[3], J Lopes[4], M Santoro[5]

[1]*Federal University of Minas Gerais - Bioinformatics PhD Program*
[2]*Federal University of Minas Gerais - Department of Computer Science*
[3]*Federal University of Itajuba - Department of Mathematics and Computer Science*
[4]*Federal University of Minas Gerais - Department of Chemistry*
[5]*Federal University of Minas Gerais - Department of Biochemistry- Immunology*

Proteins are usually divided in families according to their topology organization. In some cases, the biologic functions is also taken in account. For example, trypsins and subtilases are grouped in the same family, denominated serine-protease, although they do not share a similar tridimensional structure. In this work we describe a method to analyze the interactions between serine-proteases and their inhibitors, with the objective of finding invariant properties in enzymes that are functionally equivalent but have diferent folding patterns. We based on the hypothesis that the enzyme residues that form the protein-inhibitor interface would store information that could be retrievel through techniques of Data Mining. For this, the methodology employed was the division of the 20 types of amino acids in six diferent groups, in accordance with their physico-chemical properties. The interface residues have been grouped as triplets in the primary sequence, where at least one enzyme residue belongs the protein-inhibitor complex. The arrangements of six physico-chemical properties (forming six-tuples) in triplets produced a vector of 56 positions (or dimensions). Thus, we first characterized 11 subtilase-inhibitor complexes, creating a 56 six-tuple frequency matrix. The SVD (Singular Value Decomposition) method then was used in an attempt to extract invariant elements of this protein family. After extracting the main components through SVD we established, using cosines values, a similarity measure of a mix of 42 proteins of other families (including some other subtilases) in an attempt to find proteins that could have similar pattern with subtilases. Our results shown that the frequencies of physico-chemical properties involving the residues in the enzyme-inhibitor complexes were insufficient to aggregate, with high cosine scores, even others subtilases not submitted to the initial training data set. Despite the apparent failure, these negative results seem to imply an important conclusion: only the information regarding the frequencies of the enzyme interface residues is not enough to discriminate eventual patterns in serine-protease-inhibitor complexes. Possibly, we should have to aggregate some topology information like the relative residues distribution on the space. We are now working to review and test our methodology in face of this new hypothesis.