Experimental-theoretic study of peptide fingerprints in *Leishmania* parasites

M. Auxiliadora Dea-Ayuela[1], Riccardo Concu[2], Lazaro G. Perez-Montoto[3], Eugenio Uriarte[3], Francisco Bolás-Fernández[4], Florencia Ubeira[2], Humberto González-Díaz[2]

[1]Faculty of Experimental and Health Sciences, Department of Chemistry, Biochemistry and Molecular Biology, Cardenal Herrera University, 46113, Moncada, Valencia, Spain. [2]Faculty of Pharmacy, Department of Microbiology and Parasitology, [3]Faculty of Pharmacy, Department of Organic Chemistry University of Santiago de Compostela, 15782 Santiago de Compostela, Spain. [4]Faculty of Pharmacy, Department of Parasitology, Complutense University, 28040, Madrid, Spain

Abstract

The number of protein and peptide structures parasites, their hosts, and other organisms, released to Protein Data Bank (PDB) and Gen Bank without functional annotation has increased. Using Proteomics tools such as MASCOT, we can perform function annotation of these proteins if we have Mass Spectra (MS) outcomes and there are in databases known template proteins with known function and high similarity scores if not these methods may not succeed. Consequently, there is a high demand for Quantitative Structure-Activity Relationships (QSAR) equations to predict these functions from protein 3D structure and/or sequence. We trained and validated a new 3D-QSAR equation based on a Markov Chain Model (MCM) that classifies protein by their possible mechanism of action according to Enzyme Classification (EC) number [1] using a *Leishmania* peptide mass fingerprints model. The methodology proposed here is essentially new, and enables prediction of all EC classes with a single equation without the need for an equation for each class or nonlinear models with multiple outputs. In addition, the model may be used to predict whether one peptide presents a positive or negative contribution of the activity of the same EC class.

Materials and methods

Stationary promastigotes proteins of the *Leishmania infantum* strain LEM75 were analyzed by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) employing an immobilized linear pH 4-7 gradient for isoelectric focusing. Proteins were visualized by silver stain and spots of interest were excised from preparative 2D gels, and then analyzed by MALDI-TOF and/or MALDI-TOF/TOF mass spectrometry. [2]

In order to predict protein function we used the Electrostatic k and HINT Potentials μk. These values were used as inputs to construct the QSAR model. The average general potentials depend on the absolute probabilities *p*k(j) and the total potential with which the amino acid *j*-th interact with the rest of amino acids. All calculations were carried out with our in-house software MARCH-INSIDE [3].

Results

The model predicts EC for 106 out of 151 (70.2%) oxidoreductases, 178/178 (100%) transferases, 223/223 (100%) hydrolases, 64/85 (75.3%) lyases, 74/74 (100%) isomerases and 100/100 (100%) ligases, as well as 745/811 (91.9%) non-enzymes. We also give an example illustrating the study Peptide Mass Fingerprints (PMFs) of new protein sequences of parasites proteins when traditional alignment search procedures fail. First, we obtained the 2D-Electrophoresis (2DE) localization of the spot for a new protein from *Leishmania infantum*. We also give MALDI TOF MS characterization and results of MS MASCOT search and detected that all templates with possible enzyme or any other action have low similarity score and/or unknown function. Then, we decided to carry out prediction with our QSAR-MCM model of the contribution to specific enzyme action of 29 peptides found in the PMF of the new query protein. As the QSAR bases on 3D information, we had to predict previously possible 3D structures of these peptides using Molecular Dynamic (MD) methods. Some of the more interesting peptides were predicted with positive contribution to some specific enzyme actions; which coincided with additional BLAST alignment findings.

Conclusion

This combined strategy may be used to identify and predict peptides of proteins from parasite and/or their hosts when classic proteomics tools fails; which may be of interest in drug or vaccine development and/or drug target identification.

References

[1] C.o.B. Nomenclature, Enzyme Nomenclature.

[2] Dea-Ayuela MA, Rama-Iñiguez S, Bolás-Fernández F. Proteomic analysis of antigens from *Leishmania infantum* promastigotes. *Proteomics* 2006;6:4187-94.

[3] González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. *Proteomics* 2008;8:750-78.