GASS-WEB: a web server for identifying enzyme active sites based on genetic algorithms

João P. A. Moraes[1], Douglas E. V. Pires[2], Raquel C. de Melo-Minardi[3,4], Gisele L. Pappa[3,4], Sandro Carvalho Izidoro[1].

[1]*Advanced Campus at Itabira, Universidade Federal de Itajubá, MG,* [2]*Fiocruz Minas, MG,* [3]*Computer Science and* [4]*Department of Biochemistry and Immunology, Universidade Federal de Minas Gerais, MG.*

Structure-guided methods have been proposed over the years to infer protein function based on active site similarity. Given an active site template, these methods use different mathematical modeling and searching procedures to match the template to a given set of proteins. Many of the current available methods present, however, limitations such as performing only exact matches on template residues (not accounting for conservative changes), pruning the search space using ad-hoc procedures, and not being able to find inter-domain active sites. In order to tackle these problems, we have proposed GASS (Genetic Active Site Search), a search method based on genetic algorithms that aims to cope with the aforementioned issues. Here we propose a user-friendly web server implementing the method's capabilities, called GASS-WEB. GASS-WEB can be used under two different scenarios: (a) given a protein of interest, to try to match a set of specific templates (*i.e.,* known active sites); or (b) given an active site template, looking for it in a database of protein structures. The method has shown to be very effective on a range of experiments. Based on the Catalytic Site Atlas (CSA) annotation, it was able to correctly identify >90% of the cataloged active sites. It also managed to achieve a MCC of 0.63 on the Critical Assessment of protein Structure Prediction (CASP 10) data set, ranking fourth among 18 methods. We believe GASS-WEB would be an invaluable tool to aid in active site search and protein function prediction of newly discovered proteins.