Conserved biochemical features among four protease molecules with no structural or sequential homology.

Russ B. Altman & Steven C. Bagley
Section on Medical Informatics, SUMC, MSOB X215
Stanford, CA 94305-5479, {altman, bagley}@camis.stanford.edu

We have created a system for automatically identifying conserved biochemical and biophysical features in a set of aligned three-dimensional structural sites [1,2]. Our system takes as input a set of sites that are hypothesized to share critical features, and a set of control non-sites that are used to establish the background distributions for these features. The system then looks at small volumes within the three-dimensional sites and searches for features that are significantly abundant (or deficient) in the sites compared with the nonsites, using statistical significance tests. The program currently has a set of about thirty properties including chemical groups, mobility, charge, electronegativity, hydrophobicity, atom type, amino acid type, and secondary structure type. The volumes examined are cubes with a side dimension of 2.8 Å. The output of the program is a triplet of property-volume-significance which specifies the property found to be significantly different, the location in three-dimensions of the volume in which it is significant, and the level of significance.

In order to understand the features that are shared by successful protease molecules, and therefore may be critical for their common function, we have used our system to study the environment around the catalytic histidine in a set of serine protease enzymes. The major challenge in using our system on these sites is to control for features that may be conserved because of structural or sequential homology, but that are not actually critical for the protease activity. In order to address this problem, we selected four proteases with apparently similar catalytic mechanisms, but with no significant structural or sequential homology: a trypsin-like protease (1TLD), a subtilisin-like protease (1ST3), a carboxypepsidase-like protease (1WHS), and an engineered immunoglobin domain with some protease activity (0EAP). We superimposed the histidine and serine atoms of these four enzymes (OEAP needed to have the serine sidechain rotated, as described in the initial report of its structure). As control non-sites we selected 48 histidines that are not part of a protease catalytic site. We anticipated, therefore, that the properties that would be found significant would not be attributable to the general chemical milieu of histidine, since we had controlled for this. Instead, significant properties should be attributable to the common function of the four sites.

Our results show that there are indeed some strongly conserved features. Some of these features are well known features of these sites and include the abundance of serine (and particularly the hydroxyl group) in the region of S195 in the trypsin molecule, an abundance of negative-charge around the catalytic ASP102, and an abundance of polar groups in the binding site. Other significant features include differences in the abundance of VAL, GLY, LEU and TYR sidechains, location of ring systems, the position of amine and amide groups, and the distribution of charge. Some of these findings are consistent with location of the socalled “oxyanion hole.” This poster presents the detailed differences between the four aligned protease sites and the forty eight control histidine environments.

[1] Bagley, S.C. and Altman, R., "Characterizing the microenvironment surrounding protein sites." Protein Science, 4, 622-635, 1995.

[2] Bagley, S.C., Wei, L., Cheng, C., and Altman, R.B., Characterizing oriented protein structural sites using biochemical properties, Third International Conference on Intelligent Systems in Molecular Biology, Cambridge, London, July, 1995, AAAI Press, Menlo Park, p. 12-20.

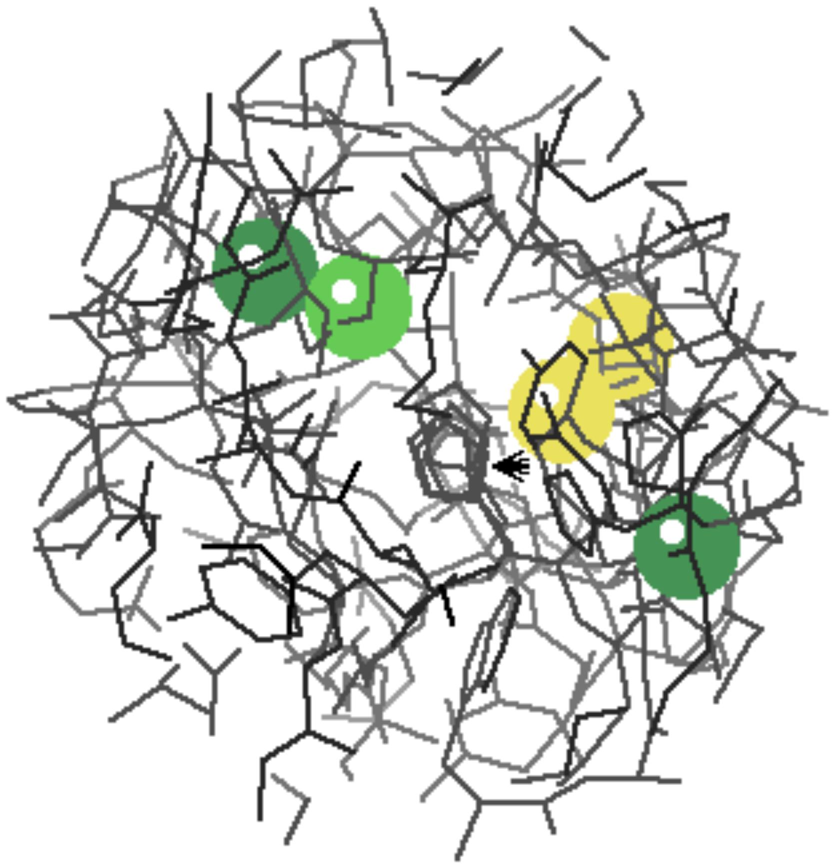

Figure 1. Four serine proteases superimposed such that the active sites are coincident. The central HIS residues are shown in the center of the figure, with an arrowhead pointing at them. Shaded balls are used to mark volumes in which statistically significant properties are observed when the four sites are superimposed in this manner. The two light (yellow) balls mark the volumes associated with high abundance of acidic residues in these structures, the two dark (green) balls in the upper left and lower right mark the volumes with high abundance of amide groups. Finally, the single light (green, second from left) ball marks the location of high abundance of hydroxyl groups—in the region of the catalytic SER.