Available online at www.sciencedirect.com
ScienceDirect
Mendeleev Commun., 2006, **16(6)**, 290–292

Mendeleev Communications

Serine hydrolase catalytic sites: geometry invariants and modeling catalytic activity

Alexander V. Nemukhin,*[a,b] Igor A. Gariev,*[a] Alexander V. Rogov[b] and Sergei D. Varfolomeev[a,b]

[a] *N. M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, 119991 Moscow, Russian Federation. Fax: +7 495 137 4101; e-mail: gariev@hotmail.com*
[b] *Department of Chemistry, M. V. Lomonosov Moscow State University, 119992 Moscow, Russian Federation. E-mail: anem@lcc.chem.msu.ru*

DOI: 10.1070/MC2006v016n06ABEH002418

An analysis of the catalytic site structures and quantum mechanical – molecular mechanical modeling of the reaction mechanism of serine hydrolases provide direct credentials to the Koshland's induced fit concept of enzymatic catalysis.

A knowledge on the evolution of enzymes and their catalytic sites[1] is dramatically influenced by an exponentially growing amount of data on protein sequences and three-dimensional structures at atomic resolution. In particular, the current archive of the Protein Data Bank[2] (PDB) includes almost 38000 structures available for immediate analysis. However, the amount of known enzymes far exceeds that of different catalytic sites, by which we mean the sets of amino acid residues (and metal ions) directly involved in chemical transformations in catalysis. An important task is to create automated procedures to recognize catalytically active sites among large arrays of protein structures and to attribute essential features of reaction mechanisms to their properties.

Here, we consider serine hydrolases, the enzymes with the catalytic triad serine–histidine–aspartic acid in the active site. Presently, PDB includes more than a thousand of protein structures referring to serine hydrolases from different enzymes obtained in X-ray and NMR studies. Although geometry configurations of the amino acid residues forming the catalytic triad (Ser, His, Asp) may be fairly different in those structures we succeed in generating a computer template the use of which allows us to recognize computationally such active sites in proteins from the PBD archive.

For computational purposes, we define an enzymatically active site as a rigid configuration of atoms characterised by sets of interatomic distances, valence and dihedral angles, which are allowed to vary within certain ranges in different enzymes. In the template, we take into account local identities of atomic configurations in protein structures based on geometry invariants. As such, we use distances between pairs of atoms,

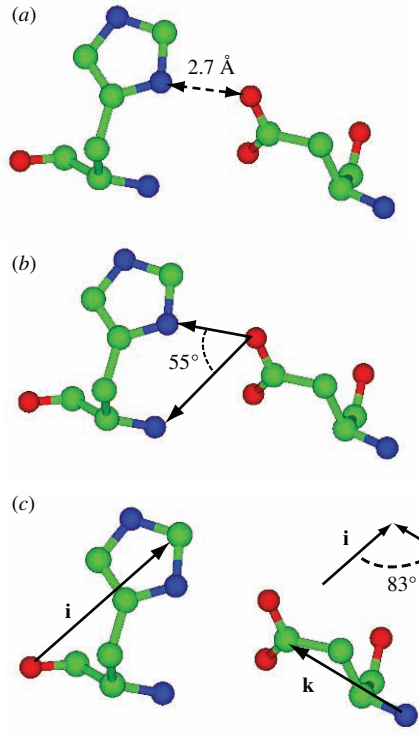

Figure 1 Definition of geometry invariants for the His–Asp pair showing typical values for (*a*) an interatomic distance and (*b*), (*c*) angles between His and Asp groups.

planar angles related to three atoms from two or three catalytic groups (Ser, His, Asp), and planar angles formed by two vectors, each of which is constructed on the same catalytic group. Figure 1 illustrates the definition of (*a*) distances and (*b*) angles for the His and Asp groups; Figure 1(*c*) illustrates the angle between two vectors **i** and **k** constructed on His and Asp, respectively.

Parameters of the template were tuned by considering enzymes of the definitely known structure, and the restraints were imposed by using the maximal and minimal values of each parameters estimated in preliminary analyses. Table 1 illustrates typical parameters of the template for the catalytic centres containing Ser–His–Asp showing five important factors: three distances and two angles.

The developed automated procedure allows us to identify the enzymatic catalytic sites by scanning all available three-dimensional PDB structures by using the template Ser–His–Asp. A statistical filter in the form of the Schennon entropy can be efficiently applied as an additional criterion when the data on threading of amino acid sequences are employed.[1] In particular, we could identify all 224 enzymes annotated in the original papers as serine hydrolases by searching among 8761 structures containing Ser, His and Asp residues.

It is important to analyse a distribution of geometry invariants for the entire array of PDB structures in order to gain better understanding of the origin of catalytic activity of enzymes. Figure 2 illustrates the distribution of enzymes by the distances between the hydroxyl oxygen (O^{γ}) from serine and nitrogen (N^{ε}) from histidine. These two atoms are hydrogen bonded, and activation of the hydroxyl group in serine by histidine is the most important step in a serine hydrolase catalytic cycle.[3]

Table 1 The most significant restraints on the structure of the catalytic triad Ser–His–Asp in active sites of serine hydrolases.

Parameter	Minimal value	Maximal value
Distance $HisN^{\varepsilon}$–$SerC^{\beta}$ (Å)	2.9	4.1
Distance $HisN^{\delta}$–$AspO^{\delta 1}$ (Å)	2.5	3.8
Distance $HisN^{\varepsilon}$–$SerO^{\gamma}$ (Å)	2.4	4.2
Angle $HisC^{\beta}$–$AspO^{\delta 1}$–$HisC^{\varepsilon}$ (°)	47.8	61.6
Angle $HisN^{\varepsilon}$–$HisC^{\beta}$–$SerO^{\gamma}$ (°)	0.9	3.6

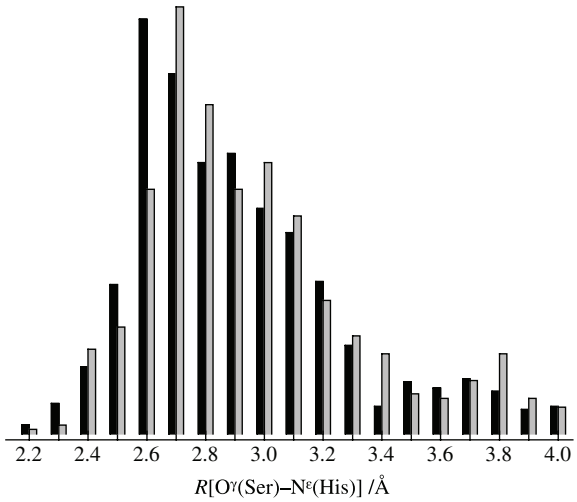

Figure 2 Histogram of the distribution of enzymes by distances between the hydroxyl oxygen (O^{γ}) from serine and nitrogen (N^{ε}) from histidine. The data referred to free enzymes are in grey; the data for enzymes with inhibitors or substrate analogues are shown in black.

We analysed 1180 structures including 291 structures of free enzymes, 515 enzyme–inhibitor (enzyme–substrate) complexes and 374 acylenzyme complexes.

One of the most important conclusions drawn from this analysis is that the O^{γ}–N^{ε} distance for free enzymes, as shown in grey in Figure 2, is systematically longer by about 0.1–0.2 Å than that in enzyme–inhibitor (enzyme–substrate) complexes shown in black. This observation can be interpreted as follows: the binding of a substrate causes a noticeable restraint on the active site.

Motivated by this analysis of the structures, we performed estimates of possible role of this effect on the catalytic activity of enzymes. More specifically, we estimated how important could be effects of the compression of an enzyme due to substrate binding. For this goal, we compared activation properties of the catalytic site of trypsin for different distances between oxygen (O^{γ}) from Ser195 and nitrogen (N^{ε}) from His57. Calculations of energies were performed in the quantum mechanical–molecular mechanical (QM/MM) approach[4,5] for the trypsin-substrate model described in detail previously.[6] The molecular model was based on the crystal structure of the trypsin–inhibitor complex,[7] including atoms of the enzyme within approximately 15 Å from the oxygen centre O^{γ} of the catalytic serine residue. A substrate was mimicked by a short peptide chain. The QM subsystem was composed of the fragments of the catalytic triad (serine, histidine and aspartic acid), a central fragment of the model substrate and an explicit water molecule.

Firstly, the O^{γ}–N^{ε} distance was set to the value typical of the protein without substrate, 2.7 Å (see Figure 2, grey). When freezing this distance, all other geometry parameters of the enzyme–substrate complex were optimised in QM/MM energy minimization. The Hartree-Fock/6-31G quantum chemistry method in the QM part and the OPLS-AA set of force field parameters in the MM part were applied. Figure 3 illustrates an arrangement of the Ser and His side chains in the active site. We note the 1.76 Å distance between N^{ε} from His and the proton from Ser.

Next, a shorter O^{γ}–N^{ε} distance (2.60 Å) was considered corresponding to a restrained structure of the active site as a result of a substrate binding (see Figure 2, black). Again all other geometry parameters of the molecular model were optimised in QM/MM energy minimization. In this model structure, the distance between N^{ε} from His and the proton from Ser was estimated as 1.65 Å. This proton transfer from Ser to His, accompanied by creation of the first tetrahedral intermediate, presents the initial rate-limiting stage of the entire catalytic cycle of serine proteases.[3] Therefore, the energy required to stretch the O^{γ}–H bond of Ser from the minimum energy configuration, shown in Figure 3, to the structure with two imposed

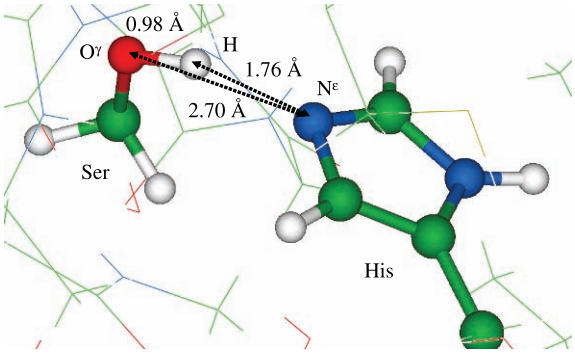

Figure 3 Relative position of Ser and His residues in the active site of trypsin at the O^{γ}–N^{ε} distance set as 2.70 Å.

constraints, $R(O^{\gamma}–N^{\varepsilon}) = 2.70$ Å, and $R(O^{\gamma}–H) = 1.65$ Å, serves as an estimate of the fraction of activation energy, which could be gained upon compression of the enzyme due to substrate binding. The QM/MM calculations resulted in the 3 kcal mol^{-1} estimate for such quantity. Taking into account that the total activation energy barrier on the reaction segment between the enzyme–substrate complex and the first tetrahedral intermediate for trypsin is computed within the same approximation as 9.6 kcal mol^{-1},[6] our analysis shows to what extent an enzymatic active site may be getting more reactive upon complexation with a substrate due to geometry restraints. The 3 kcal mol^{-1} estimate for the activation energy difference approximately corresponds to 3 orders in the rate constant.

According to the induced fit concept by Koshland,[8] the active site of an enzyme can be modified as the substrate interacts with the enzyme: the amino acid side chains, which make up the active site, are assembled into a precise shape, which enables the enzyme to perform its catalytic function. We demonstrate straightforwardly how this theory works by taking serine hydrolases as an important example.

Note that this theory is also consistent with the relaxation concept of enzymatic catalysis considered by Blumenfeld and co-authors.[9,10] Correlations between the structures of the active sites and their physiological properties in serine hydrolases are discussed in detail by Hedstrom.[3]

This study was partially supported by the Russian Foundation for Basic Research (grant no. 04-03-32007) and by programme no. 10 from the Division of Chemistry and Material Sciences of the Russian Academy of Sciences.

References

1 S. D. Varfolomeev, *Khimicheskaya enzimologiya* (*Chemical Enzymology*), Academiya, Moscow, 2005 (in Russian).
2 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. H. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucl. Acids Res.*, 2000, **28**, 235.
3 L. Hedstrom, *Chem. Rev.*, 2002, **102**, 4501.
4 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
5 A. V. Nemukhin, B. L. Grigorenko, I. A. Topol and S. K. Burt, *J. Comput. Chem.*, 2003, **24**, 1410.
6 A. V. Nemukhin, B. L. Grigorenko, A. V. Rogov, I. A. Topol and S. K. Burt, *Theor. Chem. Acc.*, 2004, **111**, 36.
7 M. Marquart, J. Walter, J. Deisenhofer, W. Bode and R. Huber, *Acta Crystallogr.*, 1983, **B39**, 480.
8 D. E. Koshland, *Proc. Natl. Acad. Sci. USA*, 1958, **44**, 98.
9 L. A. Blumenfeld, *Problemy Biologicheskoi Fiziki* (*Problems of Biological Physics*), Nauka, Moscow, 1974 (in Russian).
10 L. A. Blumenfeld and A. N. Tikhonov, *Biophysical Thermodynamics of Intracellular Process, Molecular Machines of the Living Cell*, Springer-Verlag, New York, 1994.

Received: 3rd August 2006; Com. 06/2763