3

On the Structural Characteristics of the Protein Active Sites and Their Relation to Thermal Fluctuations

Shao-Wei Huang and Jenn-Kang Hwang
Institute of Bioinformatics and Systems Biology, National Chiao Tung University, HsinChu, Taiwan, R. O. C.

1. Introduction

Due to the advances in structural biology research, a large number of protein structures have been solved in the last decade. In the same time, we also witness a rapidly growing number of structures of *unknown function* being deposited in the PDB. As a result, the ability to predict protein function from its structure becomes increasingly important in computational biology.

The conventional comparative methods(for example, Laskowski, Watson et al. 2005; Watson, Sanderson et al. 2007) for identifying functional sites rely on evolutional information like homologous structures of known function or the known catalytic templates. However, these approaches are not applicable to these novel structures. One needs to develop ingenious approaches that do not rely on evolutionary information.

Recently, several groups(Amitai, Shemesh et al. 2004; Ben-Shimon & Eisenstein 2005; Sacquin-Mora, Laforet et al. 2007; Huang, Yu et al. 2011) developed novel approaches to predict the active sites of enzymes *from a single structure* without using any homologous structures or known catalytic templates. The basic idea of their approaches is simple: they first identify certain structural or dynamical features that are unique to the active sites; they then further refine this relationship such that it can be used to accurately predict the enzyme catalytic sites. For example, Pietrokovski and co-workers(Amitai, Shemesh et al. 2004) transformed the protein structure into residue interaction graphs, with each amino acid residue represented as a graph node and the interaction between them as a graph edge. They then computed the network closeness of each residue. They found that most catalytic residues are associated with the network centrality. Ben-Shimon and Eisenstein(Ben-Shimon & Eisenstein 2005), analyzing 175 enzymes, observed that most catalytic residues are near the enzyme centroid. Based on these results, they developed novel methods to predict catalytic sites from a single structure.

The aim of this review is to show that this peculiar relationship between catalytic sites and the structure centroid or its network centrality can be accounted for by the dynamical properties of the catalytic residues, which are in general more *rigid* than other residues. In addition, we will discuss the recent studies (Halle 2002; Shih, Huang et al. 2007; Huang, Shih et al. 2008; Lin, Huang et al. 2008; Lu, Huang et al. 2008) on the surprisingly close link

between protein structure and its thermal fluctuations. These studies showed that the atomic thermal fluctuations and motional correlations can be extracted directly from protein structures without using any mechanical models.

The outline of this review is as follows: first, we will cover the recent works on extracting average dynamical properties directly from protein structures. This part occupies a significant portion of the text, since it provides not only theoretical foundations for but also physical insights on what we will introduce later – the prediction of the active sites from a single structure, which is a straightforward application of the results discussed in the first part. Finally, we will show that the rigidity of the active site residues can be inferred from the generally accepted theory of the mechanism of enzyme catalysis.

2. Atomic thermal fluctuations and motional correlations in protein

The motional correlation between atom i and j is given by

$$C_{ij} \sim \left\langle \delta\mathbf{r}_i \cdot \delta\mathbf{r}_j \right\rangle \tag{1}$$

where $\delta\mathbf{r}_i = \mathbf{r}_i - \langle \mathbf{r}_i \rangle$ is the displacement of the instantaneous position $\mathbf{r}_i$ of the atom i from its equilibrium position $\langle \mathbf{r}_i \rangle$. Knowledge of the motional correlations in protein provides valuable information about the relationship between protein dynamics and its function. For example, dynamical correlation networks may account for the long-range effects of faraway mutation sites on the functional site (Saen-Oon, Ghanem et al. 2008; Ishida 2010) or protein allostery(Fidelak, Ferrer et al. 2010; Amaro, Sethi et al. 2007; Tsai, del Sol et al. 2008). The diagonal terms or the auto-correlation terms in Eq. 1 describe atomic fluctuations. They can be obtained as B-factors from X-ray crystallography refinement or order parameters from NMR. The B-factor in its isotropic form is given formally as $B = (8\pi^2 / 3)\langle \delta\mathbf{r} \cdot \delta\mathbf{r} \rangle$.

2.1 Normal mode analysis

To compute the correlation matrix, one needs to evaluate the second derivatives of the potential energy of the protein structure. In molecular mechanics, the protein structure is usually modelled by analytical potential functions(Warshel 2002). In general, the bonding energy is approximated by a harmonic function, i.e., $\frac{1}{2}K_b(b-b_0)^2$, where K_b is the force constant, b and b_0 are the bond length and the equilibrium bond length, respectively. The bending interaction is approximated by $\frac{1}{2}K_\theta(\theta-\theta_0)^2$, where K_θ is the bending force constant, θ and θ_0 are the bending angle and the equilibrium bending angle, respectively. The torsional interaction is modelled by a sinusoidal functions: $K_\phi\left[1-\cos(n\phi+\delta)\right]^2$, where K_ϕ is the torsional force constant, ϕ and δ are the torsional angle and the reference torsional angles, respectively, and n is the torsional periodicity. The non-bonded van der Waals (VDW) interaction is approximated by the Lennard-Jones function (also referred to as a 6-12 function): $\varepsilon(r_0/r)^{12} - 2\varepsilon(r_0/r)^6$, where r is the distance between atoms, and ε and r_0 are the VDW parameters related to the potential depth and its minimum distance. The electrostatic interaction is described by the Coulomb equation: $332q_iq_j / r$, where q_i and q_j are the charges of the atom pair and r their separation. The complete potential function is usually referred to as a *force field*:

$$U = \sum_{\text{All Bonds}} \tfrac{1}{2}K_b(b-b_0)^2 + \sum_{\text{All Angles}} \tfrac{1}{2}K_\theta(\theta-\theta_0)^2 + \sum_{\text{All Torsional Angles}} K_\phi\left[1-\cos(n\phi+\delta)\right]$$
$$+ \sum_{\text{All nonbonded pairs}} \varepsilon\left[\left(r_0/r\right)^{12} - 2\left(r_0/r\right)^6\right] + \sum_{\text{All partial charges}} 332 q_i q_j / r \quad (2)$$

The Hessain matrix **H** is the square matrix of the second derivatives of the potential energy $\partial U / \partial x_i \partial x_j$, where x_i and x_j are the Cartesian coordinates of the atoms of the protein structure. Diagonalization of the Hessian matrix, i.e., $\mathbf{U}^{-1}\mathbf{H}\mathbf{U} = \mathbf{L}$, gives the normal mode frequencies **L** and normal mode vectors **U**, which describe harmonic vibrational motions of the structure. The overall molecular motion can be described as a linear combination of these normal modes. These frequencies and vectors define different modes of harmonic motions occurring in a protein structure. The above procedure is referred to as normal mode analysis (NMA) (Brooks & Karplus 1983; Levitt, Sander et al. 1985; Go 1990; Ma 2004). The correlations between atomic fluctuations given by Eq. 1 can be calculated from the normal modes. Mathematically, the correlation matrix is the pseudo-inverse of the Hessian matrix.

One limitation of NMA is its assumption that motions are harmonic. This may not be valid in the case of large-amplitude conformational dynamics in protein which are presumably highly anharmonic(Ma 2005). Furthermore, NMA needs to be carried out in an energy-minimized structure. This is to avoid the occurrence of the unphysical complex-valued normal mode frequencies. However, minimization may cause significant conformational deformations due to the inherent deficiencies of the force field.

2.2 Elastic network model

Elastic network model (ENM) (Tirion 1996), a simplified version of NMA, can be directly applied to the protein structures without energy minimization. In ENM, every atom is connected to any atoms (except itself, of course) as long as they are within a certain threshold distance (usually 12 Å) of the given atom. ENM does not distinguish between bonded interactions (such as bonding, bending or torsional interactions) and nonbonded interactions (such as VDW or electrostatic interactions). All interactions are represented by a harmonic function with a uniform force constant. Since all interactions in ENM are assumed to be of covalent nature, i.e., the atom pairs are connected by a harmonic force, this is equivalent to assuming a relatively rigid protein structure. We will comment more on that later.

A number of ENM variants has been developed: the Gaussian Network Model (GNM) (Bahar, Atilgan et al. 1997), Anisotropic Network Model (ANM) (Atilgan, Durell et al. 2001) and Quantized Elastic Deformational Model (Ming, Kong et al. 2002). Despite the simplicity of ENM, it predicts relatively accurate correlated motions in proteins. ENM has recently become a popular tool to analyze protein dynamics (Zheng, Brooks et al. 2007; Yang, Song et al. 2009; Zheng & Thirumalai 2009).

Recently, Hwang and co-workers (Shih, Huang et al. 2007; Huang, Shih et al. 2008; Lin, Huang et al. 2008; Lu, Huang et al. 2008) developed even simpler models to calculate the correlations between the atomic fluctuations in proteins directly from their structures without using mechanical models, or performing energy minimization or matrix

diagonalization. They are the Protein Fixed-Point Model and the Weighted Contact Number Model, which will be discussed in the next sections.

2.3 The protein fixed-point model

In the Protein Fixed-Point (PFP) model, the protein structure is characterized by the PFP profile $\mathbf{R}$

$$\mathbf{R} = (\mathbf{r}_1 - \mathbf{r}_0, \mathbf{r}_2 - \mathbf{r}_0, \ldots, \mathbf{r}_N - \mathbf{r}_0) \tag{3}$$

where $\mathbf{r}_i$ is the coordinate of Cα atom of the i^{th} residue, $\mathbf{r}_0$ is the fixed point and N is the number of residues. The fixed point $\mathbf{r}_0$ is identified with the centroid of the protein chain, $\mathbf{r}_0 = \sum_i \mathbf{r}_i / N$. Hwang and co-workers (Shih, Huang et al. 2007; Lu, Huang et al. 2008) showed that the correlation matrix $\mathbf{C}$ is well approximated by $\mathbf{R}^T\mathbf{R}$, where $\mathbf{R}^T$ is the transpose of $\mathbf{R}$, i.e.,

$$C_{ij} = \Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j \tag{4}$$

where $\Delta\mathbf{r}_i = \mathbf{r}_i - \mathbf{r}_0$ and $\Delta\mathbf{r}_j = \mathbf{r}_j - \mathbf{r}_0$. The motional correlation between atoms is the inner product of the vectors radiating from the centroid to the respective atoms. The B-factor, i.e., the diagonal element, is proportional to the square of the distance to the centroid(Kundu, Melton et al. 2002).

In the PFP model, the tricky issue is to determine the fixed points. In the case of a single domain protein, the fixed point is simply the centroid (or the center of mass) of the whole structure. However, in the case of multidomain protein or a protein complex, the fixed point is defined as the centroid of the *structure module*. There may be more than one fixed point since the protein structure may be composed of a number of modules. The structure modules are identified with either the structure domains or the biological units. The coordinates of biological units can be retrieved from either PDB or PQS(Henrick & Thornton 1998). Though the biological units are not uniquely defined, the PDB and PQS biological units agree on 82% of entries(Xu, Canutescu et al. 2006).

The general procedures to determine the structural modules of a protein structures go as follow: each protein chain is checked for its domains through the use of the Protein Domain Parser (PDP)(Alexandrov & Shindyalov 2003); if the PDP domain is not defined, the SCOP database(Murzin, Brenner et al. 1995) will be searched; if not found, the CATH database(Orengo, Michie et al. 1997) will be searched. If the chain is not a multidomain chain, it will be checked whether it is a part of a protein complex or a biological from PDB or PQS.

In Figure 1, we compare the computed PFP profiles with the experimental X-ray B-factors. Both the PFP and the B-factor profiles are expressed in terms of the Z-scores. In the case of the B-factor, the Z-score is defined as: $Z_B = (B - \overline{B}) / \sigma_B$, where $\overline{B}$ and σ_B are the mean and standard deviation of the B-factor. The Z-score of the PFP profile is defined similarly. For a dataset comprising 972 high-resolution X-ray structures with pairwise sequence identity ≤ 25%, the correlation coefficient between the computed and the X-ray B-factors is 0.59. There are 727 out of 972 of proteins (around 75%) having a correlation coefficient ≥ 0.5. In

comparison, GNM yields a correlation coefficient of 0.56 and the fraction of proteins with a correlation coefficient ≥ 0.5 is 69% for the same data set(Lu, Huang et al. 2008).
Figure 2 compares the PFP correlations with the NMA maps computed by GROMAC(Van Der Spoel, Lindahl et al. 2005). The agreements are excellent.

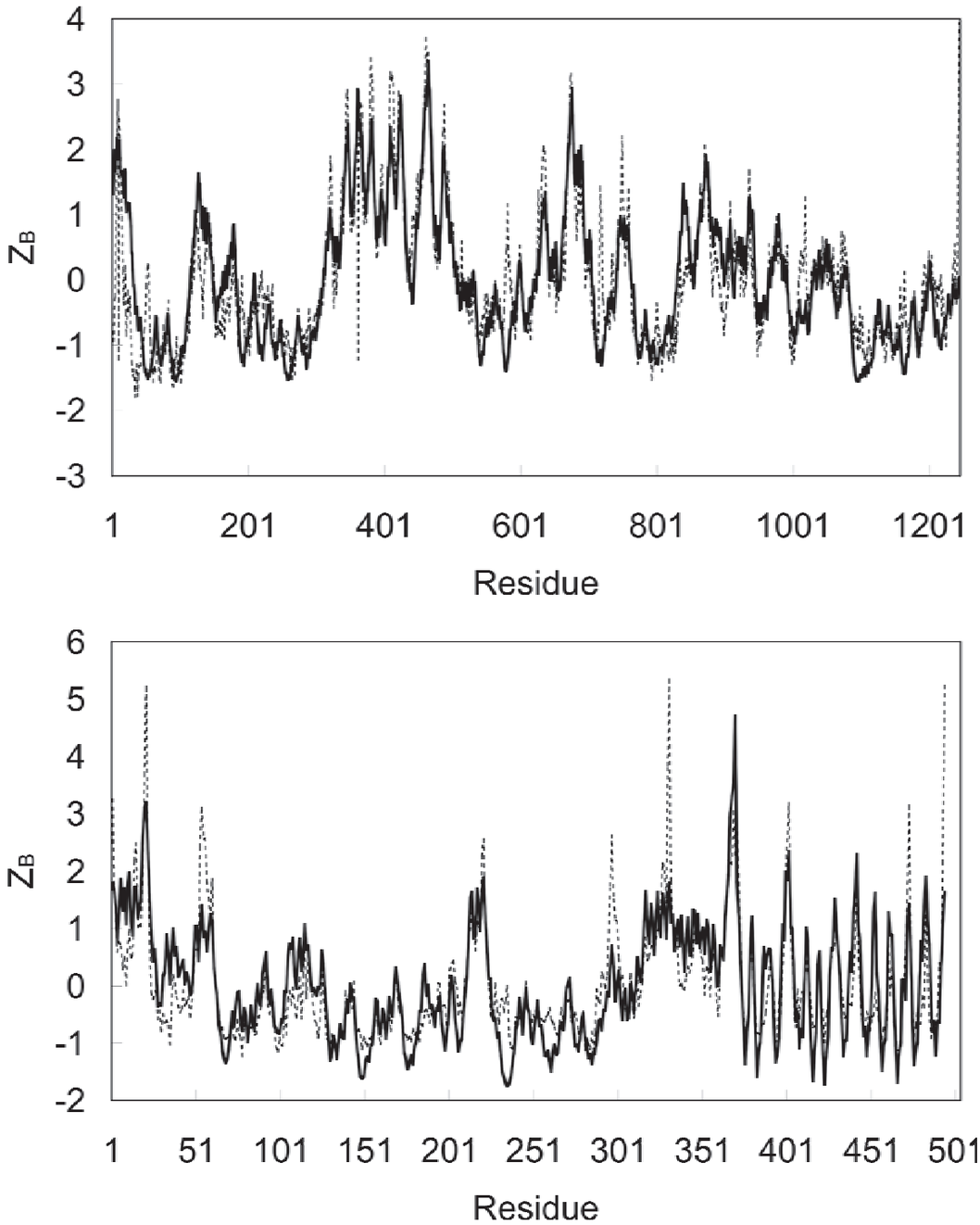


Fig. 1. Comparison of the computed PFP B-factor profile (solid line) and the X-ray B-factor profile (dotted line) of 1q16:A (top) and 2ffu (bottom). The vertical axis Z_b is the normalized B-factor.

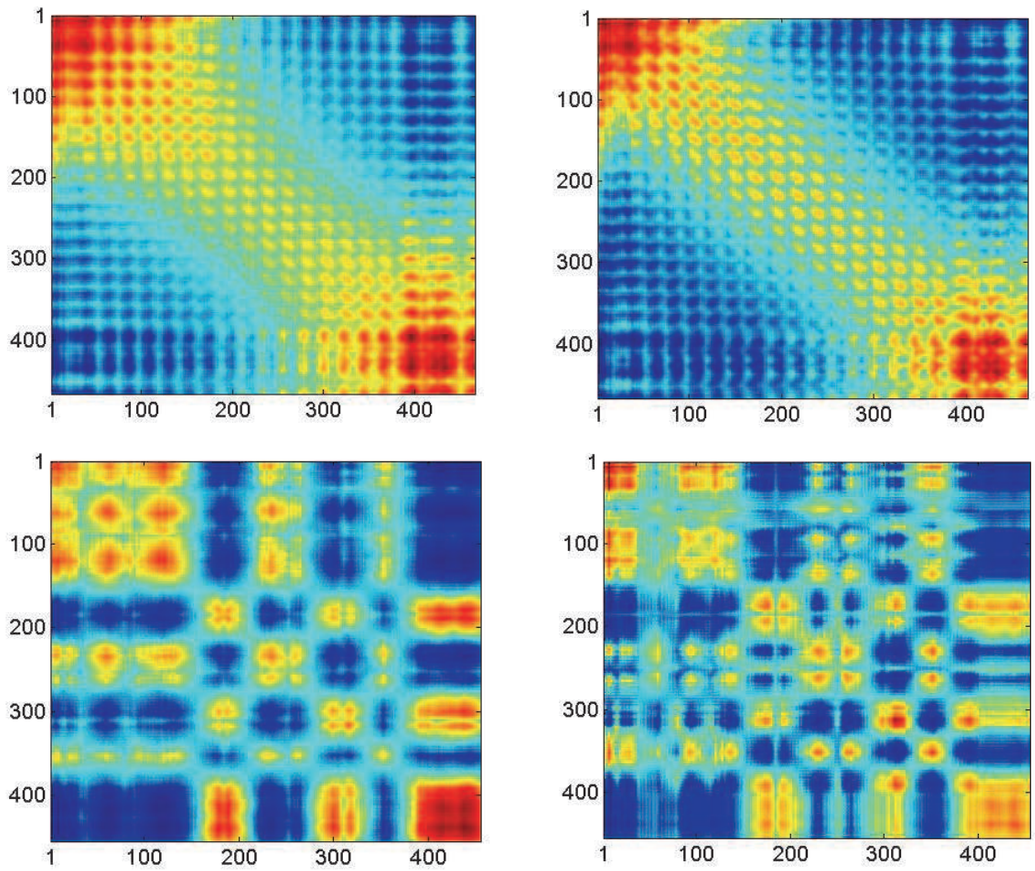


Fig. 2. Comparison of the PFP correlation map (left) and NMA correlation map computed by Gromacs (right) of 1o6v:a (top row) (this figure is adapted from Lu, Huang et al. 2008) and 1fup:a (bottom row). The colors are coded according to the rainbow spectrum. A more reddish color indicates a more negative correlation, while a more bluish color a more positive correlation.

2.4 The weighted contact number methods

The contact number (CN) of the residue i is defined as $n_i = \sum_{j \neq i}^{N} H(r_0 - r_{ij})$, where r_0 is the cutoff distance, r_{ij} is the distance between the Cα atoms of residue i and j, and $H(r)$ is the Heaviside step function defined as: $H(r) = 1$ if $r > 0$ and $H(r) = 0$ otherwise. Despite its popular uses in computational structural biology (Miyazawa & Jernigan 1984; Halle 2002), the CN has one major shortcoming: it treats the contribution *equally* of every contact atom, regardless of its distance to the center atom.

To take distance into consideration, Hwang and co-workers(Lin, Huang et al. 2008) define a weighted contact number (WCN) as

$$\nu_i = \sum_{j \neq i}^{N} \frac{1}{r_{ij}^2} \tag{5}$$

They showed that the reciprocal WCN profiles better reproduce the B-factor profiles and that the correlation between residue i and residue j is well approximated by

$$C_{ij} = \left(\sum_{k}^{N} \frac{1}{r_{ik} r_{jk}} \right)^{-1} \hat{\mathbf{x}}_i \cdot \hat{\mathbf{x}}_j \tag{6}$$

and $\hat{\mathbf{x}}_i$ is given by $(\mathbf{r}_i - \bar{\mathbf{r}}) / |\mathbf{r}_i - \bar{\mathbf{r}}|$, where $\bar{\mathbf{r}}$ is $\sum_k^N \mathbf{r}_k / N$, and $\hat{\mathbf{x}}_j$ is defined similarly.

The WCN model may be a little more computationally expensive than the PFP model, but it can be complemented as an automatic procedure. This is in contrast to the PFP model, which requires manual intervention to define the centroids.

Figure 3 compares the computed WCN B-factor profiles with the experimental X-ray B-factors. The WCN model predicts better B-factors than the PFP model, the original CN model (Halle 2002) and the GNM. For the same dataset stated before, the WCN model yields a correlation coefficient of 0.61 with 79% of proteins having a correlation coefficient ≥ 0.5.

Figure 4 compares the WCN and the NMA correlation maps. The agreements are excellent.

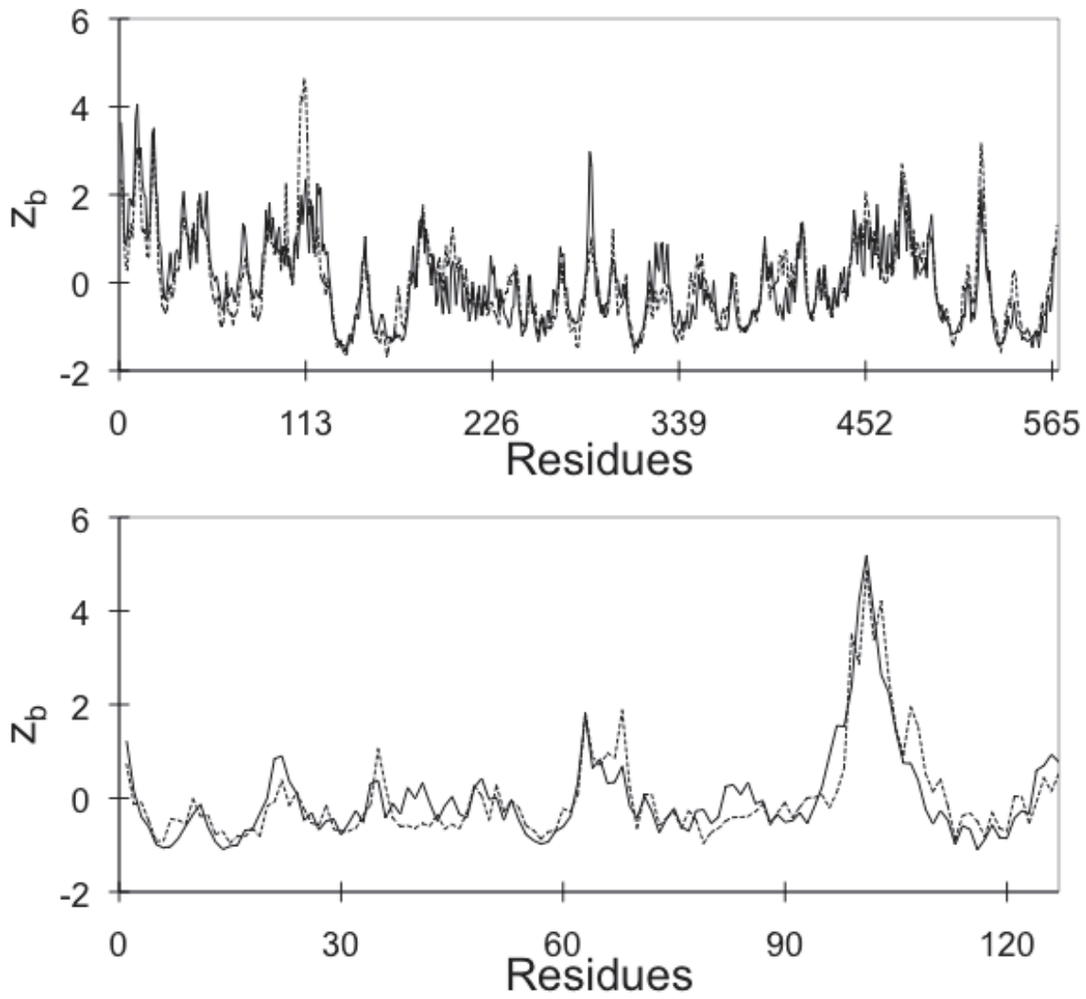


Fig. 3. Comparison of the computed WCN B-factor profile (solid line) and the X-ray B-factor profile (dotted line) of 1y0p:A (top) (this figure is adapted from Lin, Huang et al. 2008) and 1nu0 (bottom).

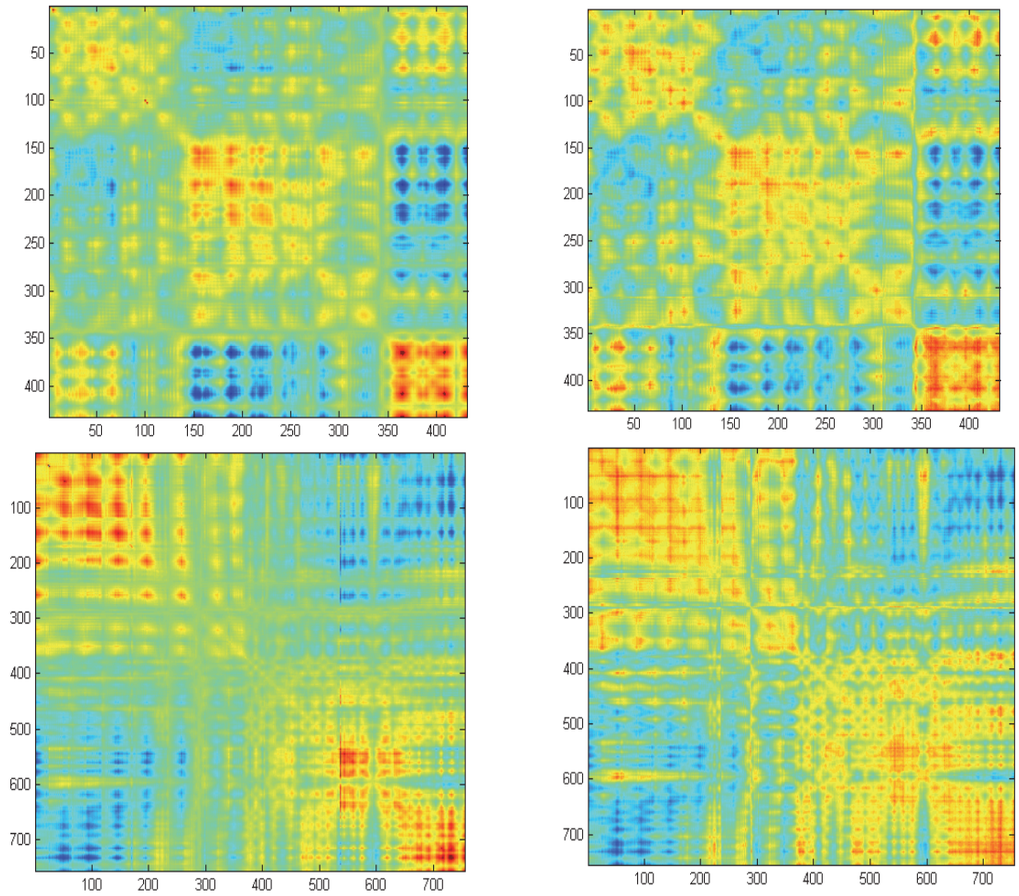


Fig. 4. Comparison of the WCN correlation map (left) and NMA correlation map computed by Gromacs (right) of 1cvr:a (top row) and 1rwh:a (bottom row) (this figure is adapted from Lin, Huang et al. 2008). The colors are coded according to the rainbow spectrum.

2.5 More on the structural and B-factor profiles

It is hardly surprising that an atom's thermal fluctuation as quantified by its B-factor should be somehow related in a qualitative way to its structural features, such as its packing environment or its position relative to the structure centroid. However, it is unexpected that the B-factor profiles and the structural profiles (i.e., the PFP and the WCN profiles) are so similar that, in some cases, one profile can be put on the top of the other with little discernible differences. It is even more surprising that this was done without information of the amino acid sequence.

Figure 5 compares the distributions of the correlation of B-factors with the original CN model, the WCN model and the PFP model for a dataset comprising 972 non-homologous structures. The B-factors distributions are computed using only Cα atoms.

The good agreement between the computed and the X-ray B-factors indicates that the B-factors can be derived solely from the protein backbone to a relatively high accuracy. A study(Lenin, Parthasarathy et al. 2000) showed that the variation in the atomic fluctuations, quantified by the B-factors, of a given protein segment only weakly depends on that of its amino acid. Note that the ENM (Tirion 1996; Bahar, Atilgan et al. 1997; Ming, Kong et al. 2002) also computes accurate B-factors without using the amino acid sequence.

While the amino acid sequence completely determines the 3-dimeniosnal structure of a protein, the thermal fluctuations and the motional correlations in a protein can be determined from its structure without its side-chain groups. The physical meaning of this is not clear and further study is definitely needed to clarify this issue.

3. The rigidity of enzyme catalytic sites

Enzymes accelerate chemical reactions by reducing the activation barrier. To achieve this, the enzyme structures are optimized through evolution to partially pre-organize their catalytic residues such that their charge distributions will stabilize the transition site complex, thus reducing the reorganization energy required for reaching the transition state(Warshel 1978; Warshel, Naray-Szabo et al. 1989; Warshel, Sharma et al. 2006). The reorganization energy is related to the activation free energy through the Marcus equation(Marcus 1956; Marcus 1956; Marcus 1957; Sumi & Marcus 1986):

$$\Delta G^{\neq} = \frac{(\lambda + \Delta G_0)^2}{4\lambda} \tag{7}$$

where $\Delta G^{\neq}$ is the activation free energy, ΔG_0 the free energy difference between the reactant and the product states and λ the reorganization energy. Eq. 7 states that the reduction of the reorganization energy λ will result in smaller activation free energy $\Delta G^{\neq}$.

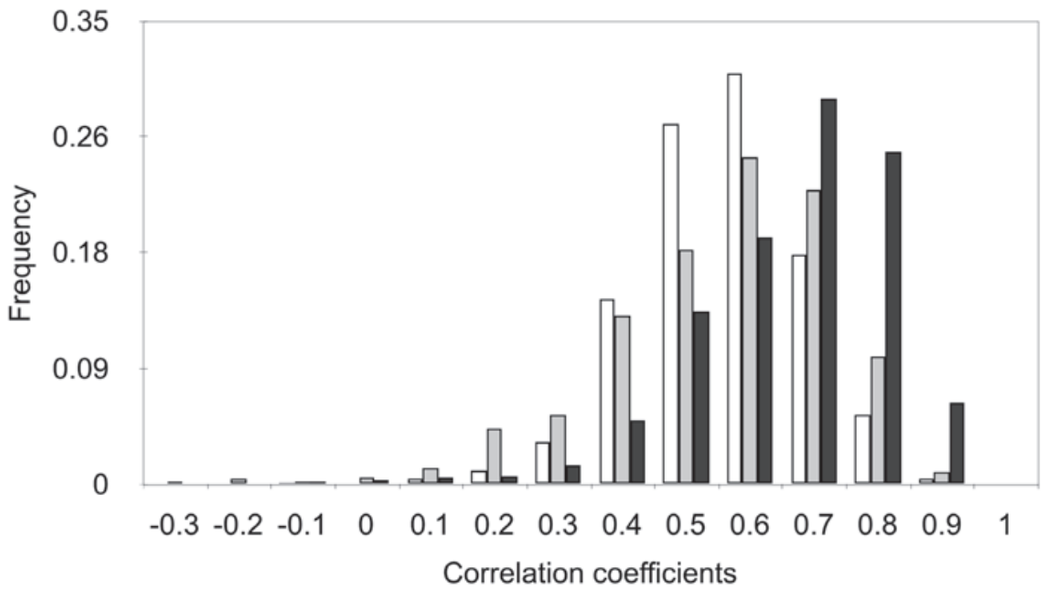


Fig. 5. Comparison of the correlation coefficients between experimental and the computed B-factor profiles based on the original CN model (white), the PFP (grey) and the WCN model (black) for the nonhomologous data set comprising 972 protein structures of length > 60. All of them are high-resolution X-ray structures with resolution ≤ 2.0 Å and R-factors ≤ 0.2. All chains are of pair-wise sequence identity ≤ 25%. This figure is adapted from Lin, Huang et al. 2008.

Hence, the active site residues are expected to be more rigid than the non-active site residues. And this has been verified by a number of studies: Yuan and co-workers (Yuan, Zhao et al. 2003) compared the B-factors of the active site sites with those of other non-active site residues of 69 apo-enzymes. They found that the active site residues indeed have lower B-factors. Analyzing a set of a set of 98 enzymes, Yang and Bahar (Yang & Bahar 2005) found that the catalytic sites usually occur in the global hinge centers and have low translational mobility. Recently, Lavery and co-workers(Sacquin-Mora, Laforet et al. 2007) showed that the force needed to displace a catalytic residue is usually larger than that to displace a non-catalytic residue.

3.1 Prediction of active site residue from a single structure

Since B-factors quantify structure flexibility, it may be tempting to use B-factors to distinguish between the active site residues and other residues. However, The B-factor values are affected by various experimental conditions such as temperature, crystallization and structural refinement.

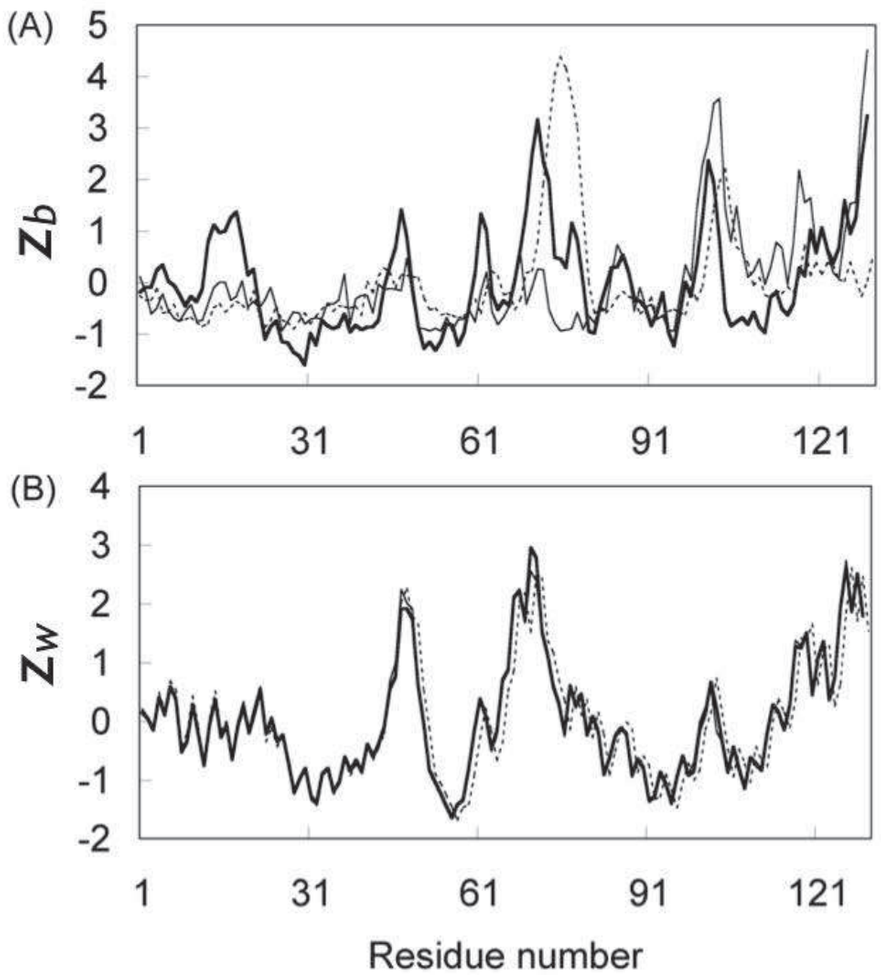


Fig. 6. (A) The Z_b profiles and (B) the Z_w profiles of 3 lysozymes: 6lyt (thick solid), 2bqo (dotted) and 2lzt (thin solid). Here Z_w is the normalized WCN. This figure is adapted from Huang, Yu et al. (2011).

Figure 6 compares the B-factor and the WCN profiles of several X-ray structures of lysozyme. Despite their almost identical structures -- their root-mean-square-deviations are within 0.6-0.8 Å, their B-factor profiles are very different.
On the other hand, their WCN profiles look almost identical. The B-factor is obviously not robust enough for predicting active site residues. The WCN profile, which correlates well with the B-factors but depends only on structure *per se*, is a much better discriminator for the catalytic residues.
Hwang and co-workers showed a straightforward application of the WCN profile to predicting the catalytic residues(Huang, Yu et al. 2011). We will illustrate this with an example: S-adenosylmethionine decarboxylase (AdoMetDC) (Ekstrom, Mathews et al. 1999), a critical regulatory enzyme of the polyamine synthetic pathway, has 5 catalytic residues. They are located in two chains: C82, S229 and H243 in chain A, and E11 and E67 in chain B. The active site of AdoMetDC is shown in Figure 7.

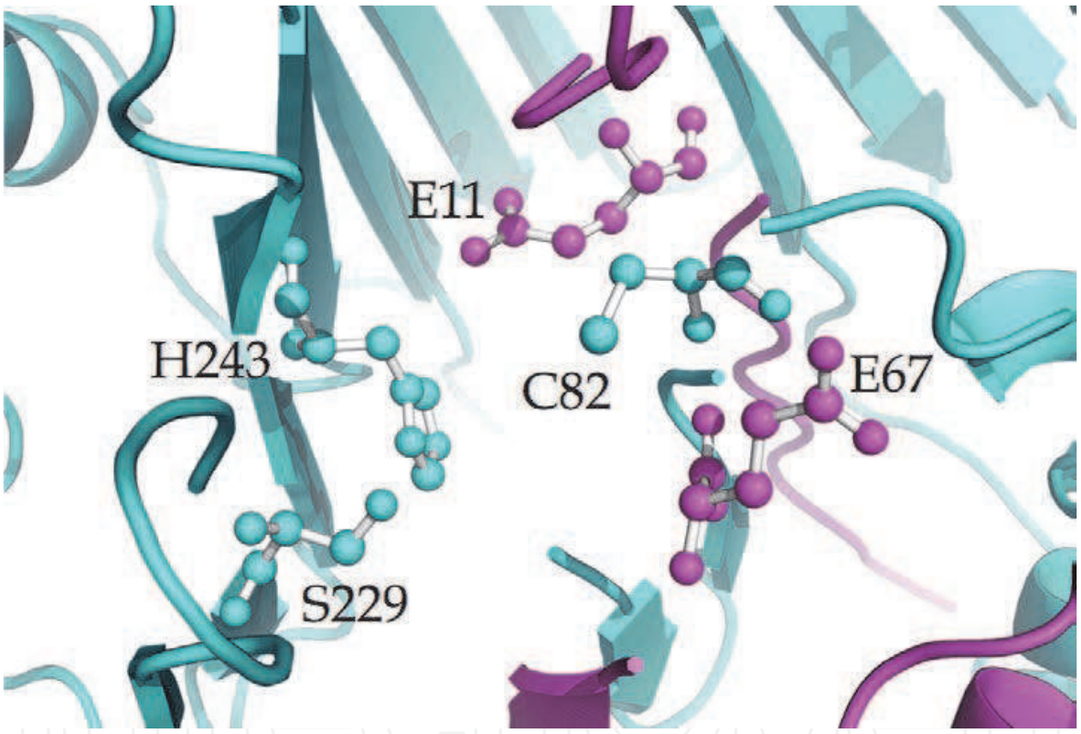


Fig. 7. The active site of AdoMetDC. Chain A is shown in cyan and chain B in magenta. The catalytic residues are labelled.

The WCN profile of AdoMetDC is shown in Figure 8. It should be noted that the WCN profile is computed for an incomplete structure -- many residues are missing in the X-ray structure of AdoMetDC. Despite this, most catalytic residues but one are located near the local minima of the WCN profile, as shown in Figure 8. The basic idea of the approach is simple: to obtain a threshold value for the WCN profile such that the residues whose Z-values are below the threshold are predicted to be catalytic residues. The threshold Z-value is determined by minimizing both the false positives and the false negatives of the predicted catalytic residues(Huang, Yu et al. 2011). The threshold for the WCN profile is -0.9.

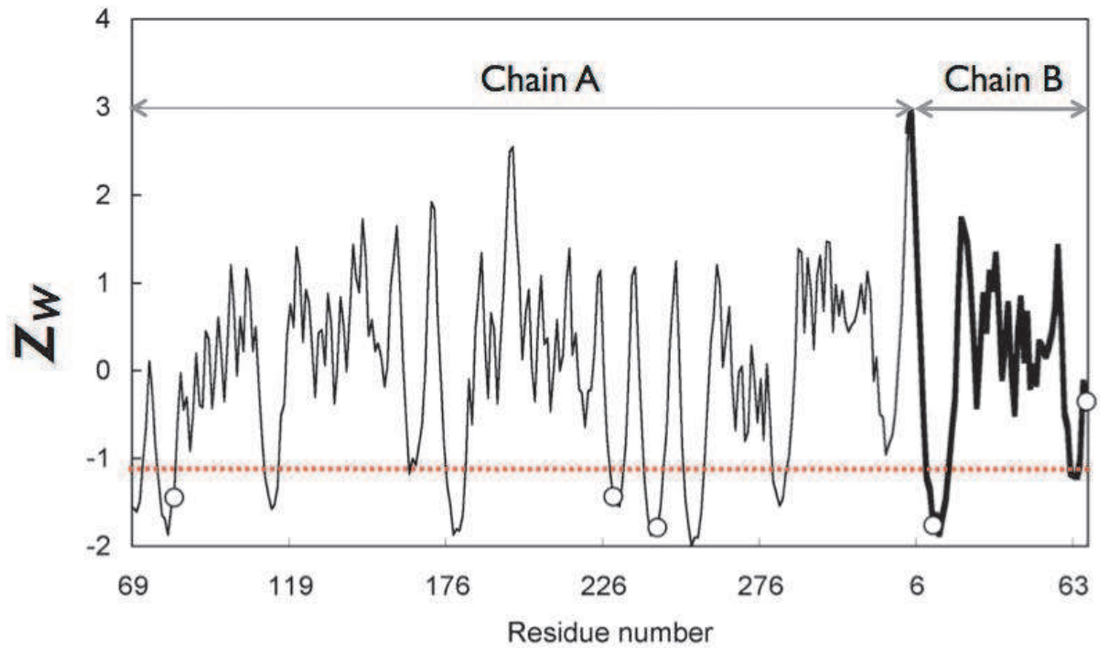


Fig. 8. The WCN profile of AdoMetDC (PDB ID: 1jen). The catalytic residues are marked in empty circles. The red dotted line indicates the threshold value.

3.2 Comparison of structural profiles

Figure 9 compares the receiver operating characteristics (ROC) curves of 4 structural profiles for their prediction of the active site residues for a data set comprising 760 X-ray nonhomologous enzyme structures(Huang, Yu et al. 2011). These profiles are the WCN profile, the PFP profile, the CN profile and the B-factor profile. The ROC curve is obtained by plotting the true positive rate (TPR) as a function of the false positive rate (FPR). TPR is defined as TPR = TP/P, where TP is true positives, i.e., number of correctly predicted catalytic residues, and P is the positive examples, i.e., the total number of catalytic residues. FPR is defined as FPR = FP/N, where FP is the false positives, i.e., the number of incorrectly predicted catalytic residues, and N is the negative examples, i.e., the number of non-catalytic residues. TPR is the *sensitivity*, while 1 – TPR is the *specificity*.

The WCN model performs the best among all the models. The PFP model comes in second. The relatively poor performance of the CN mode underlines the importance of attenuating the contributions of the neighboring atoms that depend on the distance between the interacting pair. The B-factor profile performs the worst. To have a feeling for the different performance between the WCN and the B-factor profiles, we can examine the case of highly specific predictions, i.e., at 95% specificity, the WCN model gives 52% sensitivity but the B-factor model gives only 11% sensitivity.

3.3 Other prediction methods

Here we will compare the WCN profile method with other methods in their prediction of the catalytic residues from a single structure.

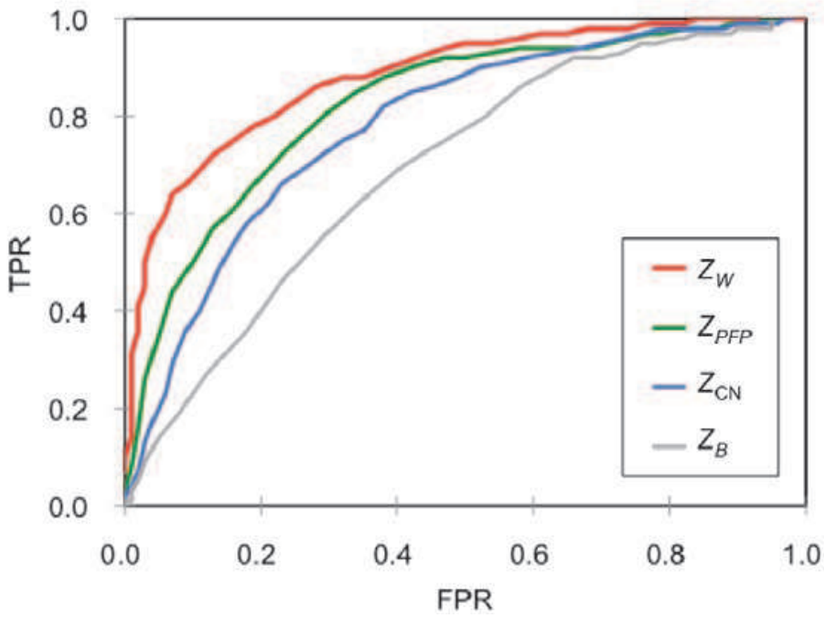


Fig. 9. The ROC curves of the WCN (Z_W), the PPF (Z_{PFP}), the CN (Z_{CN}) and the B-factor (Z_B) profiles for the prediction of catalytic residues.

Lavery and co-workers(Sacquin-Mora, Laforet et al. 2007), using Brownian dynamics simulation, computed the force needed to move any given amino acid residue with respect to other residues. They found that catalytic residues are invariably associated with high force constants. Their method gave 78% sensitivity and 74% specificity for predicting catalytic residues for a dataset of 98 non-homologous enzymes. In comparison, the WCN profile method yields 84% sensitivity and 82% specificity (Huang, Yu et al. 2011) for the same data set. This is quite remarkable, considering the simplicity of the WCN profile method as compared with the complexity of the Brownian simulation.

Analyzing the normal modes of a set of enzymes, Yang and Bahar (Yang & Bahar 2005) found that that the catalytic sites are usually of lower translational mobility than other residues, and that these catalytic sites generally coincide with the global hinge centers predicted by the GNM. Their method gave a sensitivity of 56% for a dataset of 24 enzymes. For the same data set, the WCN profile method gives 73% sensitivity and 82% specificity(Huang, Yu et al. 2011).

Ben-Shimon and Eisenstein(Ben-Shimon & Eisenstein 2005) developed a prediction algorithm called EnSite, based on the finding that the catalytic residues are often found among the 5% of residues closest to the enzyme centroid. The algorithm of EnSite is straightforward: it computes only the molecular surface that is close to the centroid, identifies continuous surface patches and finally ranks them by their area size. EnSite gives the result in terms of the rank of the correct prediction. This makes it difficult to compare the EnSite with the WCN profile method. The Pietrokovski's network centrality approach(Amitai, Shemesh et al. 2004) is similar to EnSite: both are based on the idea that the catalytic sites are near the *center* of the protein – be it represented by a structure or a network. Pietrokovski's approach gives 46.5% sensitivity and 9.4% specificity for a dataset of 178 structures. For a much larger dataset of 760 enzymes, the WCN profile gave 78% sensitivity and 80% specificity(Huang, Yu et al. 2011).

4. Conclusion

Catalytic residues are associated with a variety of structural or dynamic properties -- they are closer to the structure centroid, have higher packing density, or have smaller B-factors. Since the packing density and the centroid distances are closely related to the B-factors(Shih, Huang et al. 2007; Huang, Shih et al. 2008; Lin, Huang et al. 2008; Lu, Huang et al. 2008), all these seemingly diversified relationships can be reduced to one conclusion: the catalytic residues are more rigid. This is consistent with the present theory of enzyme catalysis(Warshel 1978; Warshel, Naray-Szabo et al. 1989; Warshel, Sharma et al. 2006; Sigala, Kraut et al. 2008): to accelerate the chemical reactions, the enzyme structures are optimized through evolution to partially pre-organize their catalytic residues to stabilize the transition state. As a result, catalytic residues tend to maintain similar conformations in both the reactant and the transition states. The catalytic residues will be more rigid than other non-catalytic residues.

5. Acknowledgements

This research was supported by the National Science Council and ATU from the Ministry of Education, Taiwan, R.O.C.

6. References

Alexandrov, N. and I. Shindyalov (2003). PDP: protein domain parser. *Bioinformatics*, Vol.19, No.3, pp. 429-30.

Amaro, R. E., A. Sethi, et al. (2007). A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry*, Vol.46, No.8, pp. 2156-73.

Amitai, G., A. Shemesh, et al. (2004). Network analysis of protein structures identifies functional residues. *Journal of Molecular Biology*, Vol.344, No.4, pp. 1135-1146.

Amitai, G., A. Shemesh, et al. (2004). Network analysis of protein structures identifies functional residues. *J Mol Biol*, Vol.344, No.4, pp. 1135-46.

Atilgan, A. R., S. R. Durell, et al. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J*, Vol.80, No.1, pp. 505-15.

Bahar, I., A. R. Atilgan, et al. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des*, Vol.2, No.3, pp. 173-81.

Ben-Shimon, A. and M. Eisenstein (2005). Looking at enzymes from the inside out: the proximity of catalytic residues to the molecular centroid can be used for detection of active sites and enzyme-ligand interfaces. *J Mol Biol*, Vol.351, No.2, pp. 309-26.

Brooks, B. and M. Karplus (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A*, Vol.80, No.21, pp. 6571-5.

Ekstrom, J. L., Mathews, II, et al. (1999). The crystal structure of human S-adenosylmethionine decarboxylase at 2.25 A resolution reveals a novel fold. *Structure*, Vol.7, No.5, pp. 583-95.

Fidelak, J., S. Ferrer, et al. (2010). Dynamic correlation networks in human peroxisome proliferator-activated receptor-gamma nuclear receptor protein. *Eur Biophys J*, Vol.39, No.11, pp. 1503-12.

Go, N. (1990). A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis. *Biophys Chem*, Vol.35, pp. 105-112.

Halle, B. (2002). Flexibility and packing in proteins. *Proc Natl Acad Sci U S A*, Vol.99, No.3, pp. 1274-9.

Henrick, K. and J. M. Thornton (1998). PQS: a protein quaternary structure file server. *Trends Biochem Sci*, Vol.23, No.9, pp. 358-61.

Huang, S. W., C. H. Shih, et al. (2008). Prediction of NMR order parameters in proteins using weighted protein contact number model. *Theo. Chem. Acc.*, Vol.121, pp. 197-200.

Huang, S. W., S. H. Yu, et al. (2011). On the relationship between catalytic residues and their protein contact number. *Curr Protein Pept Sci*, Vol.In Press, pp.

Ishida, T. (2010). Effects of Point Mutation on Enzymatic Activity: Correlation between Protein Electronic Structure and Motion in Chorismate Mutase Reaction. *Journal of the American Chemical Society*, Vol.132, No.20, pp. 7104-7118.

Kundu, S., J. S. Melton, et al. (2002). Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophys J*, Vol.83, No.2, pp. 723-32.

Laskowski, R. A., J. D. Watson, et al. (2005). Protein function prediction using local 3D templates. *J Mol Biol*, Vol.351, No.3, pp. 614-26.

Lenin, V. M. S., S. Parthasarathy, et al. (2000). Atomic displacement parameters of homologous proteins: Conservation of dynamics. *Current Science*, Vol.78, No.9, pp. 1098-1105.

Levitt, M., C. Sander, et al. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J Mol Biol*, Vol.181, No.3, pp. 423-47.

Lin, C. P., S. W. Huang, et al. (2008). Deriving protein dynamical properties from weighted protein contact number. *Proteins*, Vol.72, No.3, pp. 929-35.

Lu, C. H., S. W. Huang, et al. (2008). On the relationship between the protein structure and protein dynamics. *Proteins*, Vol.72, No.2, pp. 625-34.

Ma, J. (2004). New advances in normal mode analysis of supermolecular complexes and applications to structural refinement. *Curr Protein Pept Sci*, Vol.5, No.2, pp. 119-23.

Ma, J. (2005). Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, Vol.13, No.3, pp. 373-80.

Marcus, R. A. (1956). Electrostatic Free Energy and Other Properties of States Having Nonequilibrium Polarization. I. *J Chem Phys*, No.24, pp. 979.

Marcus, R. A. (1956). On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I. *J Chem Phys*, Vol.24, No.966, pp.

Marcus, R. A. (1957). On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. II. Applications to Data on the Rates of Isotopic Exchange Reactions. *J Chem Phys*, Vol.26, pp. 867.

Ming, D., Y. Kong, et al. (2002). How to describe protein motion without amino acid sequence and atomic coordinates. *Proc Natl Acad Sci U S A*, Vol.99, No.13, pp. 8620-5.

Ming, D., Y. Kong, et al. (2002). Domain movements in human fatty acid synthase by quantized elastic deformational model. *Proc Natl Acad Sci U S A*, Vol.99, No.12, pp. 7895-9.

Miyazawa, S. and R. L. Jernigan (1984). Effective Inter-Residue Contact Energies from Protein Crystal-Structures. *Biophysical Journal*, Vol.45, No.2, pp. A130-A130.

Murzin, A. G., S. E. Brenner, et al. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, Vol.247, pp. 536-540.

Orengo, C. A., A. D. Michie, et al. (1997). CATH- A Hierarchic Classification of Protein Domain Structures. *Structures*, Vol.5, No.8, pp. 1093-1108.

Sacquin-Mora, S., E. Laforet, et al. (2007). Locating the active sites of enzymes using mechanical properties. *Proteins*, Vol.67, No.2, pp. 350-9.

Sacquin-Mora, S., E. Laforet, et al. (2007). Locating the active sites of enzymes using mechanical properties. *Proteins-Structure Function and Bioinformatics*, Vol.67, No.2, pp. 350-359.

Saen-Oon, S., M. Ghanem, et al. (2008). Remote mutations and active site dynamics correlate with catalytic properties of purine nucleoside phosphorylase. *Biophys J*, Vol.94, No.10, pp. 4078-88.

Shih, C. H., S. W. Huang, et al. (2007). A simple way to compute protein dynamics without a mechanical model. *Proteins*, Vol.68, No.1, pp. 34-38.

Sigala, P. A., D. A. Kraut, et al. (2008). Testing geometrical discrimination within an enzyme active site: Constrained hydrogen bonding in the ketosterold isomerase oxyanion hole. *Journal of the American Chemical Society*, Vol.130, No.41, pp. 13696-13708.

Sumi, H. and R. A. Marcus (1986). Dyanmics effects in electron-transfer reactions. *J Chem Phys*, Vol.84, No.9, pp. 4894-4914.

Tirion, M. M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters*, Vol.77, No.9, pp. 1905-1908.

Tsai, C. J., A. del Sol, et al. (2008). Allostery: absence of a change in shape does not imply that allostery is not at play. *J Mol Biol*, Vol.378, No.1, pp. 1-11.

Van Der Spoel, D., E. Lindahl, et al. (2005). GROMACS: fast, flexible, and free. *J Comput Chem*, Vol.26, No.16, pp. 1701-18.

Warshel, A. (1978). Energetics of enzyme catalysis. *Proc Natl Acad Sci U S A*, Vol.75, No.11, pp. 5250-4.

Warshel, A. (2002). Molecular dynamics simulations of biological reactions. *Acc Chem Res*, Vol.35, No.6, pp. 385-95.

Warshel, A., G. Naray-Szabo, et al. (1989). How do serine proteases really work? *Biochemistry*, Vol.28, No.9, pp. 3629-37.

Warshel, A., P. K. Sharma, et al. (2006). Electrostatic basis for enzyme catalysis. *Chem Rev*, Vol.106, No.8, pp. 3210-35.

Watson, J. D., S. Sanderson, et al. (2007). Towards fully automated structure-based function prediction in structural genomics: a case study. *J Mol Biol*, Vol.367, No.5, pp. 1511-22.

Xu, Q., A. Canutescu, et al. (2006). ProtBuD: a database of biological unit structures of protein families and superfamilies. *Bioinformatics*, Vol.22, No.23, pp. 2876-82.

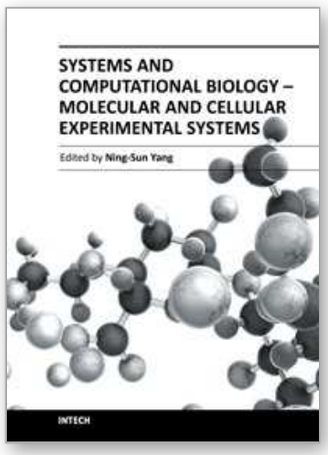
Yang, L., G. Song, et al. (2009). Protein elastic network models and the ranges of cooperativity. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.106, No.30, pp. 12347-12352.

Yang, L. W. and I. Bahar (2005). Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure*, Vol.13, No.6, pp. 893-904.

Yuan, Z., J. Zhao, et al. (2003). Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng*, Vol.16, No.2, pp. 109-14.

Zheng, W., B. R. Brooks, et al. (2007). Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. *Biophys J*, Vol.93, No.7, pp. 2289-99.

Zheng, W. and D. Thirumalai (2009). Coupling between normal modes drives protein conformational dynamics: illustrations using allosteric transitions in myosin II. *Biophys J*, Vol.96, No.6, pp. 2128-37.



Systems and Computational Biology - Molecular and Cellular Experimental Systems

Edited by Prof. Ning-Sun Yang

ISBN 978-953-307-280-7
Hard cover, 332 pages
Publisher InTech
Published online 15, September, 2011
Published in print edition September, 2011

Whereas some â€œmicroarrayâ€ or â€œbioinformaticsâ€ scientists among us may have been criticized as doing â€œcataloging researchâ€, the majority of us believe that we are sincerely exploring new scientific and technological systems to benefit human health, human food and animal feed production, and environmental protections. Indeed, we are humbled by the complexity, extent and beauty of cross-talks in various biological systems; on the other hand, we are becoming more educated and are able to start addressing honestly and skillfully the various important issues concerning translational medicine, global agriculture, and the environment. The two volumes of this book presents a series of high-quality research or review articles in a timely fashion to this emerging research field of our scientific community.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Shao-Wei Huang and Jenn-Kang Hwang (2011). On the Structural Characteristics of the Protein Active Sites and Their Relation to Thermal Fluctuations, Systems and Computational Biology - Molecular and Cellular Experimental Systems, Prof. Ning-Sun Yang (Ed.), ISBN: 978-953-307-280-7, InTech, Available from: http://www.intechopen.com/books/systems-and-computational-biology-molecular-and-cellular-experimental-systems/on-the-structural-characteristics-of-the-protein-active-sites-and-their-relation-to-thermal-fluctuat

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen