

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# On the Relationship Between Residue Solvent Exposure and Thermal Fluctuations in Proteins

Yu-Tung Chien<sup>1</sup>, Jenn-Kang Hwang<sup>2</sup> and Shao-Wei Huang<sup>1</sup>

<sup>1</sup>*Department of Medical Informatics,  
Tzu Chi University, Hualien, Taiwan,*

<sup>2</sup>*Institute of Bioinformatics and Systems Biology,  
National Chiao Tung University, Hsinchu, Taiwan,  
R.O.C.*

## 1. Introduction

The dynamic properties such as temperature factors or B-factors of a crystallographic protein structure result from a complex network of various interactions. To compute temperature factors, the straightforward way is to integrate the long time trajectory of the protein structure using molecular dynamics (Levitt & Warshel, 1975; Warshel, 1976; McCammon et al., 1977) and to compute the root mean square fluctuations of the trajectory, which takes into account all possible interactions in terms of empirical force field. Recently, the elastic network model (Tirion, 1996; Bahar et al., 1997; Ming et al., 2002) has been shown to be quite successful in computing temperature factors. Compared with the usual molecular mechanical force field, the elastic network model is a much simpler model. It is based on a mechanical model that each C $\alpha$  atom is connected through a single-parameter harmonic potential function to its surrounding atoms that are within a cut-off distance usually in the range of 7 to 10 Å. The mathematical operations in the elastic network model are simply the matrix inversion and diagonalization, no trajectory computation required. It is somewhat surprising that a mechanical model as simple as the elastic network model, which uses only single parameter, i.e., the cut-off distance, can compute dynamics properties such as B-factors with accuracy comparable to the more complex molecular dynamics method.

The study of solvent accessible surface area (ASA) has been one of the most important topics in computation biology due to the fact that the residues interacting with other biological molecules are located on protein surface (Connolly, 1983). ASA has been used in the studies of protein stability (Gromiha et al., 1999), protein folding (Eisenberg & McLachlan, 1986), and fold recognition (Liu et al., 2007). The relationships between thermal fluctuation and some concepts related to ASA have been studied. B-factor has shown to correlate to local packing densities (Halle, 2002), atomic distance to protein center-of-mass (Shih et al., 2007; Lu et al., 2008) and residue depth (Zhang et al., 2009). Residue flexibility was known to be correlated to its ASA in single protein (Sheriff et al., 1985) or small dataset (Carugo & Argos,

1997). In a related study, the impact of the ASA of immediate and further neighbors of investigated residues was also noted (Zhang et al., 2009). In this study, we further discuss the relationship between ASA and theoretical thermal fluctuations which are not reported in previous studies.

Since the elastic network model is mainly based on information of local packing of each C $\alpha$  atom, one may expect that relative atomic solvent accessibility may qualitatively reflect atomic fluctuation. In this work, we showed that this relation goes beyond just a qualitative one - the profiles of the temperature factors of crystallographic structures are very similar to those of the smoothed amino acid solvent accessibility. Our results show that protein dynamical properties can be inferred directly from the static structural properties without assuming an additional mechanical model. Another interesting corollary from our results is that one may predict temperatures factors from protein sequences with a prediction program of solvent accessibility.

## 2. Comparison between B-factor and actual RSA value

In this section, we compare B-factor and relative solvent accessibility (RSA) derived from protein structure. First, we discuss B-RSA relationships when different window sizes are used in smoothing RSA. Second, we show in detail how smoothed RSA (sRSA) are better correlated to B-factor than unsmoothed RSA (uRSA) in several example proteins and the phenomenon is generally true in a large dataset. Third, B-RSA correlations are calculated separately for residues located on protein surface, in the core, and in between.

### 2.1 Methods

#### 2.1.1 The B-factor and the relative solvent accessibility profiles

The amino acid solvent accessible areas of the proteins are obtained from DSSP (Kabsch & Sander, 1983). For each protein, we can compute its relative solvent accessibility (RSA) profile. The RSA of the  $i^{\text{th}}$  amino acid of type  $x$ ,  $a_i^x$ , is computed from

$$a_i^x = 100A_i / A_0^x \quad (1)$$

where  $A_i$  is the solvent accessible area of the  $i^{\text{th}}$  amino acid and  $A_0^x$  is the maximal solvent accessible area of amino acid of type  $x$ . The RSA profile of a protein of  $N$  residues is denoted as  $A_u = (a_1, a_2, \dots, a_N)$ .

The *smoothed* RSA profile is computed by averaging the RSA of each amino acid together with those of its  $n$  flanking amino acids. In the case of the terminal amino acid, its one-sided flanking amino acids are used twice in averaging. The smoothed RSA profile of a protein of  $N$  residues is denoted as  $A'_u = (a'_1, a'_2, \dots, a'_N)$ , where  $a'_i$  is the smoothed relative solvent accessibility. For convenience, we will refer to the original RSA as uRSA and the smoothed RSA as sRSA. The predicted RSA is denoted as pRSA and the smoothed pRSA as spRSA. The spRSA is computed from pRSA by the smoothing method mentioned above. The B-factors are extracted from the PDB files of x-ray protein structures. The B-factor profile is given as  $B = (b_1, b_2, \dots, b_N)$ , where  $b_i$  is the B-factor of the C $\alpha$  atom of the  $i^{\text{th}}$  residue.

### 2.1.2 The z-scores of the B-factor and the relative solvent accessibility

For easy comparison, the B-value, the uRSA, the sRSA, the pRSA and the spRSA profiles are normalized to their corresponding z-scores

$$z_x = \frac{x - \bar{x}}{\sigma_x} \quad (2)$$

where  $\bar{x}$  and  $\sigma_x$  are the mean and standard deviation of  $x$ , respectively. Here  $x$  is  $b$ ,  $a$  or  $a'$ . We will use the notations  $z_B$ ,  $z_{uRSA}$ ,  $z_{sRSA}$ ,  $z_{pRSA}$  and  $z_{spRSA}$  and to denote the z values of B, uRSA, sRSA, pRSA and spRSA, respectively. Their corresponding profiles are denoted by the vectors  $\mathbf{z}_B$ ,  $\mathbf{z}_{uRSA}$ ,  $\mathbf{z}_{sRSA}$ ,  $\mathbf{z}_{pRSA}$  and  $\mathbf{z}_{spRSA}$ . In terms of  $\mathbf{z}$ , the correlation coefficient  $c$  between, for example,  $\mathbf{z}_{sRSA}$  and  $\mathbf{z}_B$  is computed by  $\mathbf{z}_{sRSA} \cdot \mathbf{z}_B$ . If  $c = 1$ , these profiles are perfectly correlated; if  $c = 0$ , they are completely independent of each other;  $c = -1$ , they are perfectly anti-correlated.

### 2.1.3 Data set

The data set (PR972) is selected from PDB-REPRDB (Noguchi & Akiyama, 2003), including 972 protein chains with sequence length larger than 60 residues and pair wise sequence identity smaller than 25%. All structures are solved by x-ray crystallography with resolution better than 2.0 Å and R-factors smaller than 0.2. Detail list of the data set can be found in our previous study on protein structure-dynamics relationship (Lu et al., 2008).

### 2.2 Selection of smoothing window size of sRSA profiles

The thermal fluctuation of a single residue is affected by its flanking amino acids because of various interactions between their atoms, for example, hydrogen bonds, van der Waals interaction, electrostatic force, etc. Here, we show that the correlation between B-factor and RSA increases obviously with proper smoothing window size.

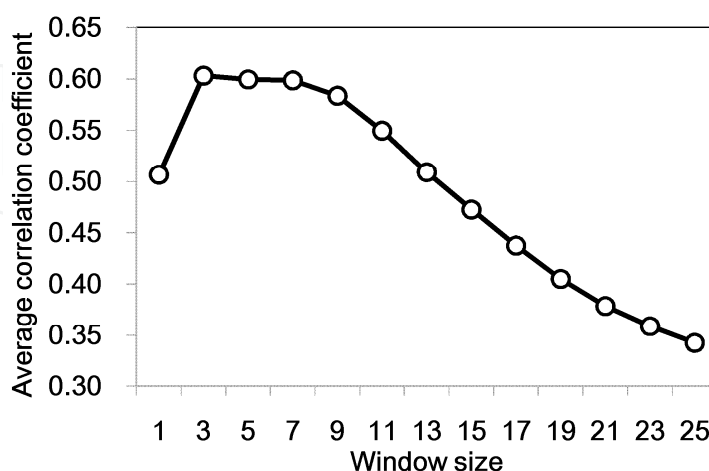


Fig. 1. Average correlation coefficients between normalized B-factor and RSA profiles based on the PR972 dataset. The RSA profiles are smoothed by different window sizes, from 0 (unsmoothed) to 25.

Figure 1 shows the average correlation coefficients between B-factor and RSA profiles smoothed by different window sizes, ranging from 1 (unsmoothed) to 25. The calculations are done based on the PR972 dataset. The average correlation coefficient between B-factor and RSA increases greatly from 0.51 (unsmoothed) to 0.60 (window size = 3). It is interesting that the results are similar when the window sizes are between 3 to 9 and become worse when the window size is larger than 9. It suggests that the thermal fluctuation of a residue is mostly affected by the neighbouring residues in this range (window size from 3 to 9). When the window size is too large, it seems that the information from distant residues has little correlation with the thermal fluctuation.

Based on the calculations shown in Figure 1, the window size is set to 3 (with the highest average correlation coefficient of 0.60) in the following sections.

### 2.3 The effects of smoothing on the correlation between B-factor and RSA

Figure 2 shows the B-factor, unsmoothed RSA (uRSA) and smoothed RSA (sRSA) profiles of a typical example: 5-carboxymethyl-2-hydroxymuconate isomerase (1OTG:C). The uRSA profile has a more rugged shape when compared with the B-value profile. If the rugged fine structures of uRSA profile are smoothed out, the global shapes of the B-factor and the sRSA profiles are seen to be quite similar (Figure 1B). The ruggedness of the uRSA profile is due to that the solvent accessibility of an amino acid may be quite different from its immediate flanking amino acids, but the B values of an amino acid and its immediate flanking amino acids appear to be more correlated.

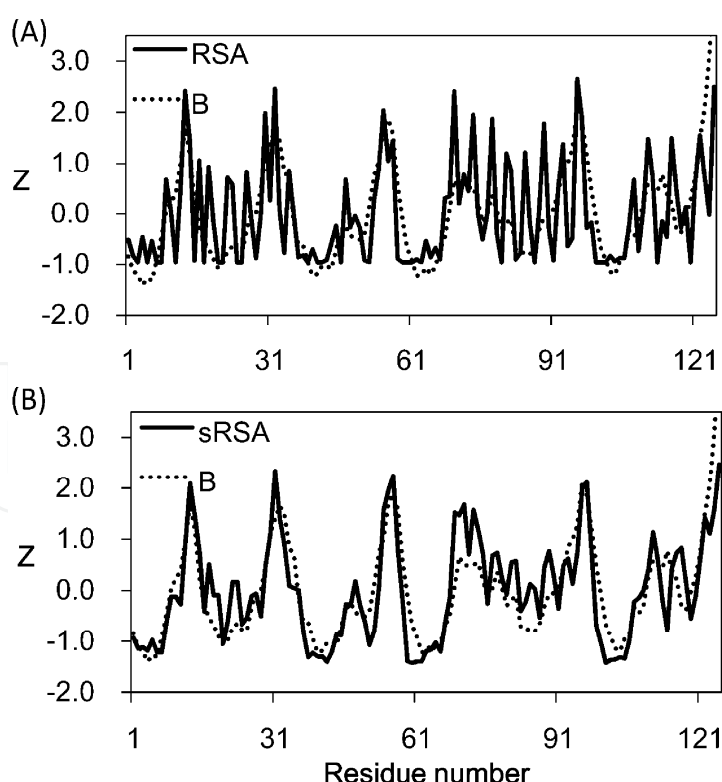


Fig. 2. Comparison of the B-value, uRSA and sRSA profiles of 5-carboxymethyl-2-hydroxymuconate isomerase (1OTG:C): (A) the uRSA (solid line) and the B-factors (dotted line) profiles; (B) the sRSA (solid line) and the B-factors (dotted line) profiles.

We computed the correlation coefficient between the B-factor, uRSA and sRSA profiles for the PR972 data set. Figure 3A shows the distribution of correlation coefficient between the B- factor and the uRSA profiles. The median of the correlation coefficients is 0.52 and 62% of the proteins have a correlation coefficient  $\geq 0.5$ . Figure 3B shows the distribution of correlation coefficient between the B-factor and the sRSA profiles. The median of the correlation coefficients is now improved to 0.65 and 86% of the proteins have a correlation coefficient  $\geq 0.5$ .

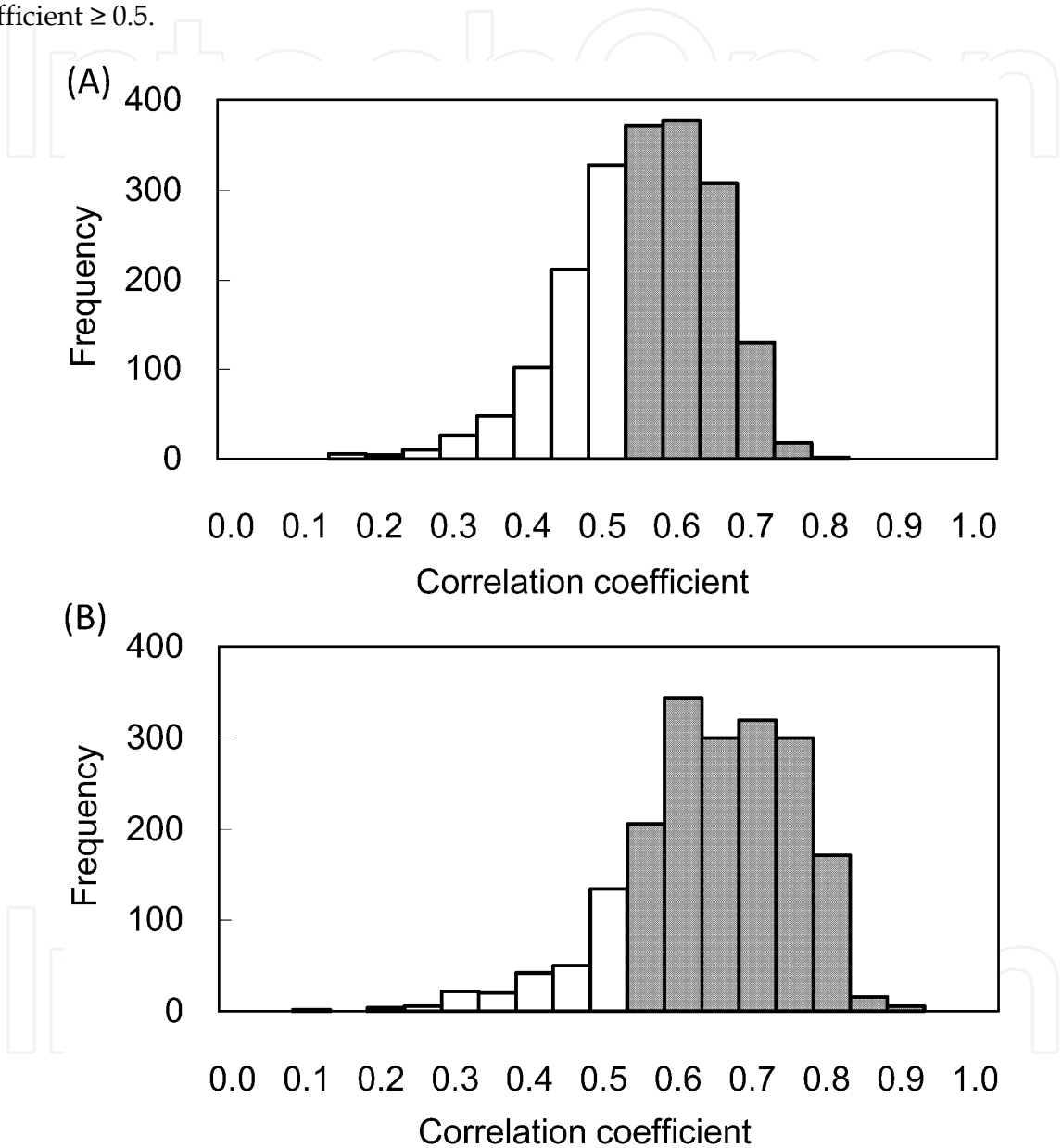


Fig. 3. The distribution of correlation coefficients of (A) the uRSA and (B) the sRSA profiles for the nonhomologous data set of 972 chains (PR972). The shaded areas indicate the sequences that have correlation coefficients  $\geq 0.5$ .

## 2.4 Examples

Figure 4 shows some examples of the uRSA and sRSA profiles together with the B-value profiles: iron (II) superoxide dismutase (1ISA:B), dimethylsulfoxide reductase (1EU1:A) and

the  $\alpha/\beta$  domain of 6-phosphogluconate dehydrogenase (2PGD), *Serratia* endonuclease (1SMN:B). The B-factor profiles are very similar to sRSA profiles. Figure 4 shows the correlation between  $z_{sRSA}$  and  $z_B$  of the corresponding proteins.

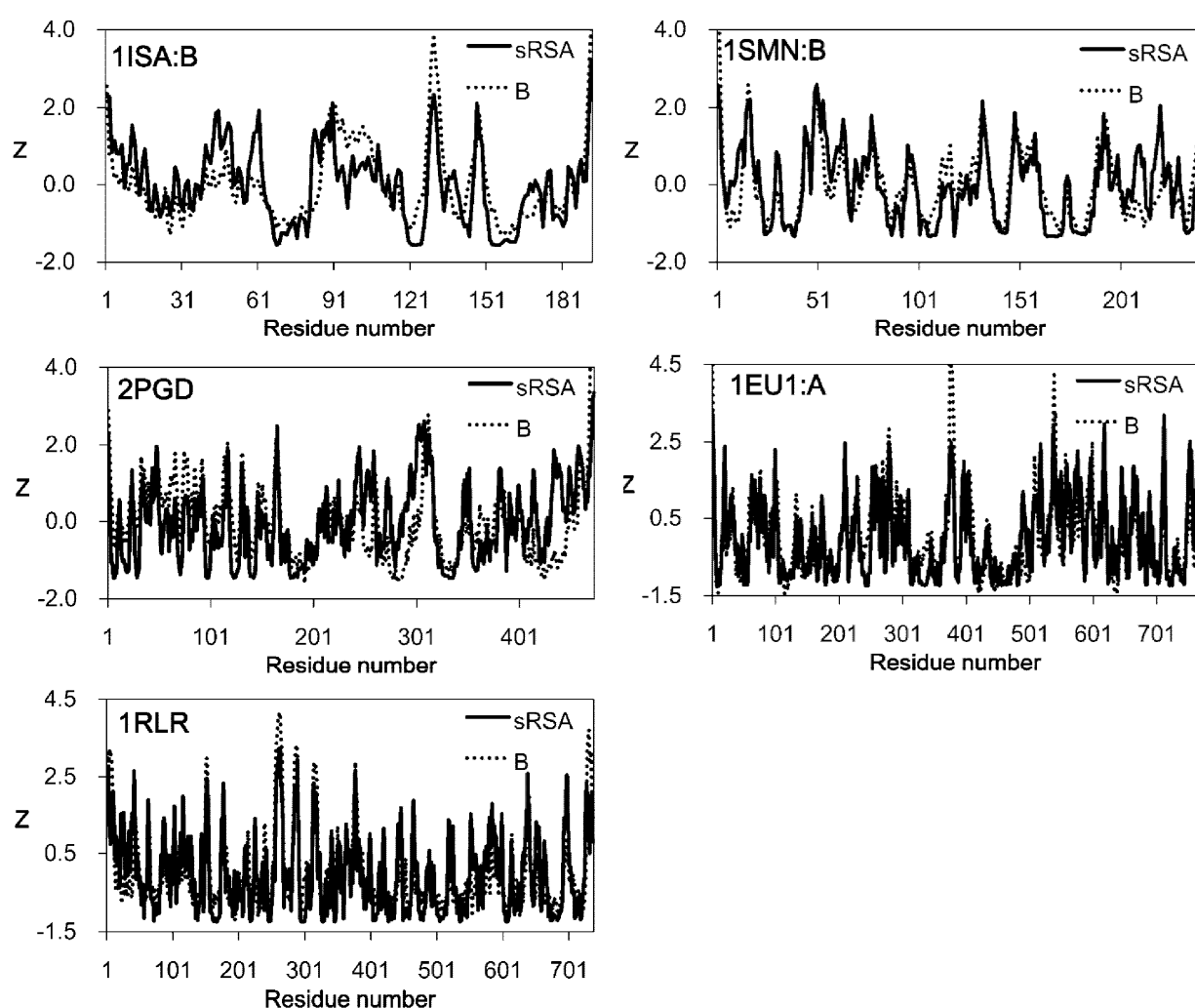


Fig. 4. Comparison of the crystallographic B-factors (dotted lines) of C $\alpha$  atoms and the sRSA profiles (solid lines) of the following proteins: 1ISA:B, iron (II) superoxide dismutase; 1EU1:A, a dimethylsulfoxide reductase; 2PGD, the  $\alpha/\beta$  domain of 6-phosphogluconate dehydrogenase; 1SMN:B, a *Serratia* endonuclease; 1RLR, the  $\alpha/\beta$  domain of ribonucleotide reductase protein R1. The sRSA profiles are normalized to the scale of the crystallographic B-factors.

## 2.5 Comparison of B-RSA relationship for residues in different structural contexts

Proteins are strongly interacting with their environment and protein dynamics have been shown to be affected or “slaved” to the solvent around them (Fenimore et al., 2004). Since we have observed close relationship between B-factor and RSA, it is interesting to examine the relationships separately for the surface residues, the partially buried residues and the residues deeply buried in protein core.



The residues in the PR972 dataset are grouped based on the local rigidity of each residue. The rigidity score is computed using the WCN model (Lin et al., 2008) which is a cutoff-free contact number model. The rigidity scores range from 0 to 1 and we use two bin sizes, bin size=0.1 and bin size=0.2, to separate the residues into 10 and 5 groups respectively. Figure 5 shows the correlation coefficients between B values and sRSA for each group (bin size=0.1 shown in filled circle, bin size=0.2 shown in rhomb). The results clearly show that the correlations between B values and sRSA decrease as the rigidity scores increase (correlation coefficient=-0.77). In other words, B values are better correlated to sRSA for residues located in less crowded environment or on the surface of the protein.

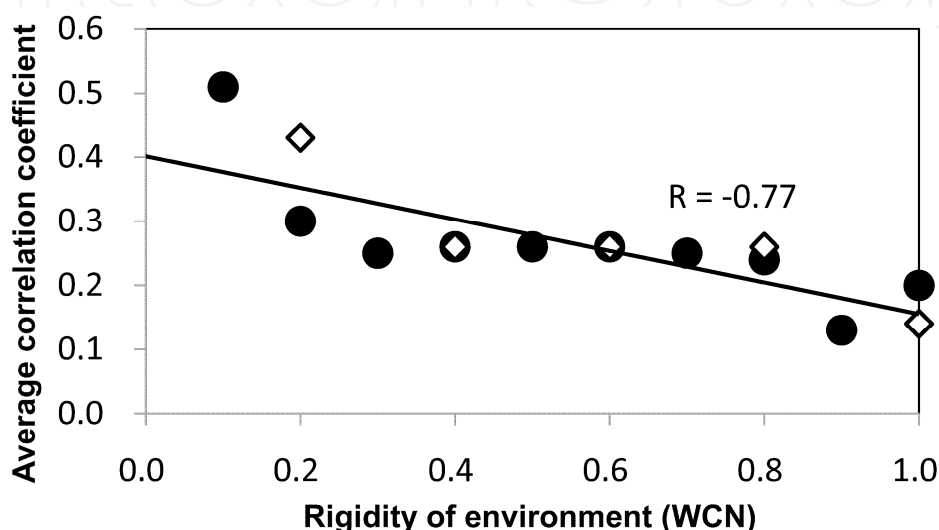


Fig. 5. The correlation coefficients between B and sRSA profiles for residue groups with increasing rigidity scores. The bin size=0.1 (filled circle) and bin size = 0.2 (rhomb) are used to separate the residues into groups. The linear regression is based on the groups of bin size=0.1. The statistics are done using the PR972 dataset.

### 3. Comparison between theoretical thermal fluctuation and actual RSA value

Thermal fluctuations in protein can be computed theoretically from its structure. Here, we further compare RSA profiles with theoretical thermal fluctuations computed by two methods, the Protein-fixed-point (PFP) model (Shih et al., 2007; Lu et al., 2008) and Weighted-contact number (WCN) model (Lin et al., 2008). These two models are shown to be able to reproduce B-factor from protein structure. The WCN model was further applied to the prediction of NMR order parameters (Huang et al., 2008) and the identification of enzyme active sites (Huang et al., 2011). The basic idea of the PFP model is assume that thermal fluctuation of a residue is related to the squared distance between the residue and the center-of-mass of the protein:

$$B_i \propto r_i^2 \quad (3)$$

where  $B_i$  is the theoretical thermal fluctuation of residue  $i$  and  $r_i$  is the distance between residue  $i$  and protein center-of-mass. The WCN model is a cutoff-free contact number model. Unlike traditional contact number calculation, the WCN model assumes that all



residues interact with each other in a protein. The contact number of a residue is contributed by all other residues but each contact term is weighted inversely by the squared distance between them:

$$w_i = \sum_{j \neq i}^N \frac{1}{r_{ij}^2} \quad (4)$$

where  $w_i$  is the weighted contact number score of residue  $i$ ,  $N$  is the total residue number of the protein, and  $r_{ij}$  is the distance between residue  $i$  and  $j$ .

Experimental thermal fluctuations are affected by different experimental conditions. For example, two structures of rubredoxin, 1CAA and 1IRO, have almost identical structure (RMSD = 0.44Å) but their B-factor profiles are dissimilar. Figure 6 compares their B-factor and WCN profiles, clearly showing that similar X-ray structures may have very different B-factor profiles.

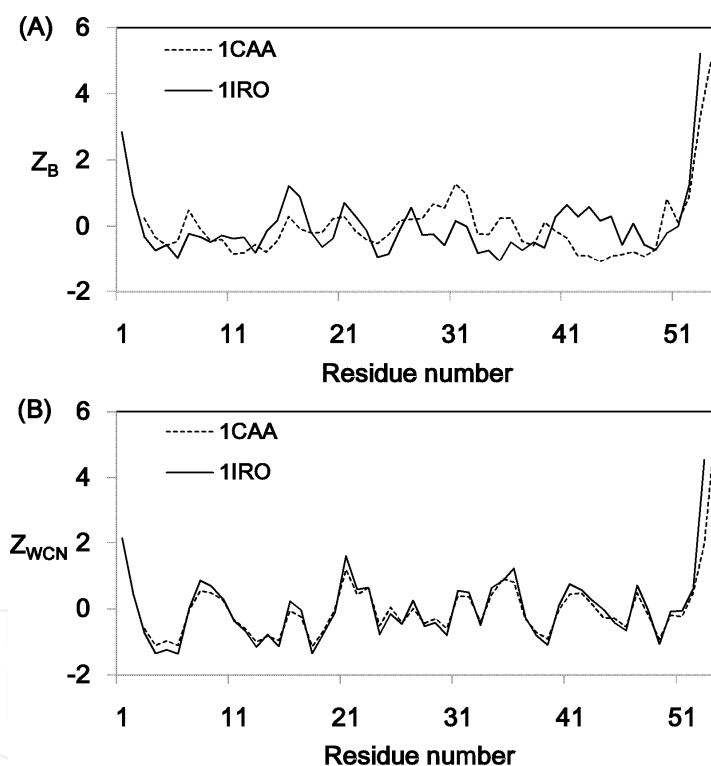


Fig. 6. Comparison of Z-scores of (A) B-factor profiles and (B) WCN profiles of two structures of rubredoxin (PDB id: 1CAA and 1IRO)

Figure 7 shows the distributions of correlation coefficients between sRSA and thermal fluctuations derived from different methods, including experimental B-factor, PFP model, and WCN model. The average correlation coefficients based on the PR972 dataset are 0.72 (WCN) and 0.55 (PFP) which are higher than that of B-factor (0.52). The correlations between sRSA and theoretical thermal fluctuations are more obvious than that of sRSA and B-factor. One reason may be that B-factor is affected by experimental conditions as shown in the example illustrated in figure 6.

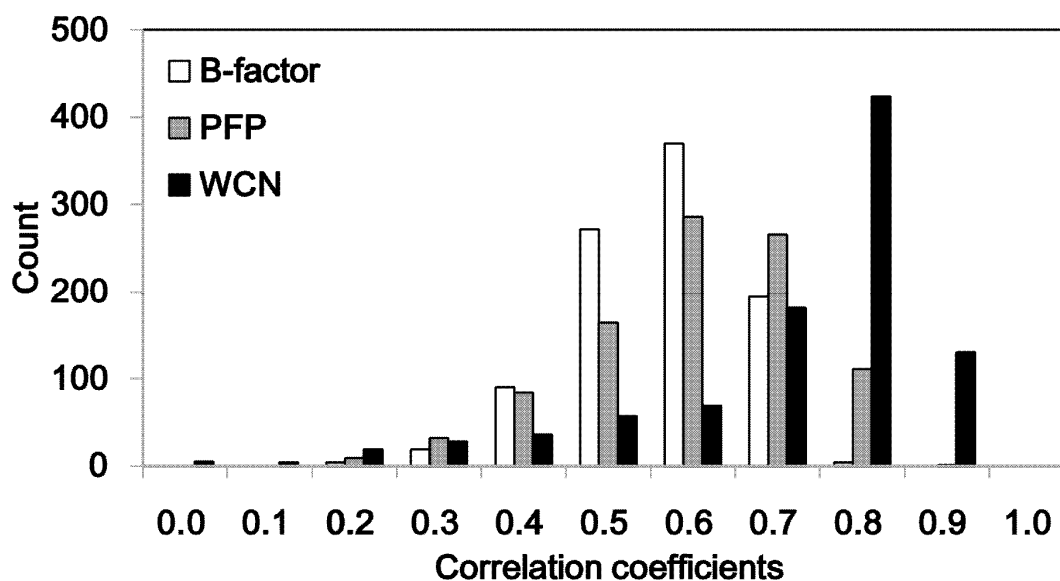


Fig. 7. The distributions of correlation coefficients between sRSA and B-factor (white), PFP (grey), and WCN (black) profiles based on the PR972 dataset

#### 4. Prediction of B-factors from sequences based on predicted RSA profiles

Since the smoothed solvent accessible surface profiles are quite similar to the B-factor profiles, we can indirectly predict B-factors from sequence using the smoothed RSA profile predicted from sequence.

##### 4.1 Methods

##### 4.1.1 Prediction of relative solvent accessibility

The real value prediction of RSA (pRSA) is computed by the support vector machine (SVM) regression model. The inputs to the SVM are position-specific substitution matrix (PSSM) obtained from PSI-BLAST (Altschul et al., 1997), secondary structure profile predicted by PSIPRED (McGuffin et al., 2000) and hydropathy index. The PSSM profile was obtained after three iterations ( $E$ -value = 0.003) against the non-redundant protein sequence database. A  $N \times 24$  scoring matrix is used as an input to the SVM, which including the PSSM profiles, secondary structure profiles and hydropathy index. Here, the size of the sliding window of the sequence,  $N$ , is set to 15. We train the SVM regression model by the commonly used RS126 dataset (Rost & Sander, 1993), a non-homologous data set with pair wise sequence identity less than 25% over a length of more than 80 residues.

##### 4.1.2 The support vector machine

The support vector machine method (SVM) (Vapnik, 1995) has been successfully applied to secondary structure prediction (Hua & Sun, 2001; Kim & Park, 2003; Ward et al., 2003), subcellular localization prediction (Hua & Sun, 2001; Yu et al., 2004), protein fold assignment (Ding & Dubchak, 2001; Yuan et al., 2003) and other biological pattern classification problems (Dobson & Doig, 2003; Chen et al., 2004; Kim & Park, 2004). The original idea of the SVM is to find the separating hyperplane with the largest distance between two classes.

However, because the data to be classified may not always be linearly separable, the SVM overcomes this difficulty by using the kernel function to nonlinearly transform the original input space into a higher dimensional feature space, so that the data may be effectively separated in the higher dimensional space. SVMs perform well compared with other machine-learning methods because of convenient classifier's capacity control and effective avoidance of overfitting. In this work, the software package LIBSVM (Chang & Lin, 2001) was used.

4.2 Prediction results and comparison with other methods

The 2-state model is used to evaluate the performance of RSA prediction. Each residue is assigned buried or exposed by comparing its RSA value with a threshold and the prediction accuracy is defined by the percentage of correctly predicted residues. The prediction accuracies using different thresholds are listed in Table 1. The accuracies of two related prediction methods, SVMpsi and Fuzzy K-NN, are also listed for comparison with our results. The SVMpsi (Kim & Park, 2004) used SVM and the position-specific scoring matrix (PSSM) generated from PSI-BLAST and the Fuzzy K-NN (Sim et al., 2005) used fuzzy k-nearest neighbor method and PSSM as feature vectors.

Method	State threshold			
	25%	16%	5%	0%
SVMpsi	76.8	77.8	79.8	86.2
Fuzzy K-NN	78.3	79.0	82.2	87.2
This work	77.8	78.1	80.0	88.8

Table 1. The accuracies of predicting RSA from sequence in 2-state model based on the RS126 dataset

The correlation coefficients between the B-factor profiles and the predicted RSA (pRSA) and smoothed pRSA (spRSA) profiles are computed for the PR972 data set. The correlation coefficient between the B-factor and pRSA profiles is 0.44 and 31% of the proteins have correlation coefficient  $\geq 0.5$ . After smoothing the pRSA profiles, the correlation coefficient between the B-factor and spRSA profiles increases to 0.53 and 55% of the proteins have correlation coefficient  $\geq 0.5$ . Since spRSA and B-factor are well correlated, we assume that the spRSA of each residue is its predicted B-factor value. We found that the results of utilizing spRSA are comparable to one of the current best B-factor prediction results (Yuan et al., 2005). Yuan used support vector regression approach and PSSM information to predict the B-factor distribution from sequence and reported a correlation coefficient of 0.53 on a dataset of 766 high-resolution proteins. For fair comparison, we tested our method on the same dataset. The average correlation coefficient between spRSA and B-factor is 0.52 which is comparable to theirs (0.53).

Figure 8 shows the B-factor, pRSA and spRSA profiles of human uracil-DNA glycosylase inhibitor (1UGH:I). The correlation coefficient between the B-factor and pRSA profiles of the protein is 0.68 (Figure 8A). After smoothing, the correlation coefficient between the B-factor and spRSA profiles is 0.83. The B-factor profiles are better correlated to the spRSA profiles than the unsmoothed profiles, especially in the N-terminal and the P26-S39 residue regions (Figure 8B).

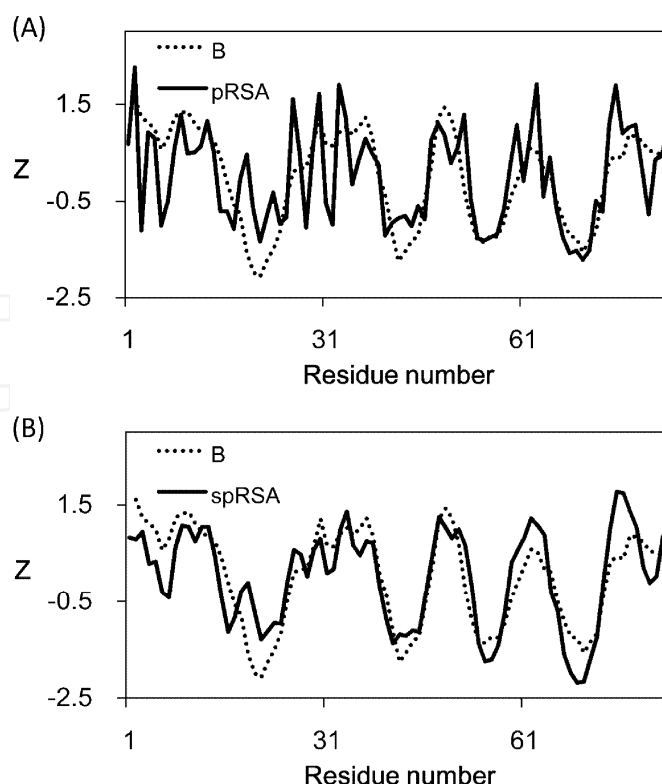


Fig. 8. Comparison of the B-value, pRSA and spRSA profiles of human uracil-DNA glycosylase inhibitor (1UGH:I): (A) the pRSA (solid line) and the B-factors (dotted line) profiles; (B) the spRSA (solid line) and the B-factors (dotted line) profiles.

Several methods have been developed to predict B-factor from sequence using different methodologies and testing datasets. The results of these methods have been compared using a small dataset (290 protein chains) (Radivojac et al., 2004). The average correlation coefficients are 0.32 (Vihinen et al., 1994) by a sliding window averaging technique and B-factor propensities, 0.43 by using neural network and multiple sequence alignment information (Radivojac et al., 2004). There were also structure-based prediction methods, for example, it was reported that the elastic network model has an average correlation coefficient of 0.59 between B-factor and their residue flexibility score (Kundu et al., 2002).

## 5. Conclusions

Though the dynamic properties of a protein result from a complex network of various interactions, we found that the B-factors of crystallographic structures are closely related to solvent accessibility directly derived from the protein structure. Our results indicate that dynamic properties such as B-factors can be computed directly from the protein's geometrical shape without assuming any mechanical models. Furthermore, we found that the smoothed solvent accessibility profiles are very similar to the B-factor profiles, and in some proteins, these profiles can overlap with each other almost perfectly. In a dataset comprising 972 non-homologous protein sequences, 86% of the proteins have a correlation coefficient between  $Z_B$  and  $Z_{sRSA}$  larger than 0.5. The results are consistent with the research by previous work (Zhang et al., 2009) showing the linear relationship between RSA and B-value profiles. In addition, we show that the relationship is not equal for residues in

the environments of different rigidity. The correlations are higher for the residues located in the loose regions than those in the rigid environments.

Our results suggest that protein structure and protein dynamics are so closely related that the relative solvent accessibility profile allow one to directly infer the complete B-factor profiles. It will be interesting to investigate further whether there exists similar relationship between B-factors and other structural properties other than solvent accessibility. Recent studies showed a close relationship between the sites of low B-factor values and the active or the binding sites (Yuan et al., 2003; Yang & Bahar, 2005). Our results suggest a potential way to identify active sites from sequence without structure information. Figure 9 shows the spRSA and B-value profiles of type 2 rhinovirus 3C protease (1CQQ). The catalytic residues, H40, E71 and C147 (shown in open circle), are located at the low-mobility regions in the spRSA profile.

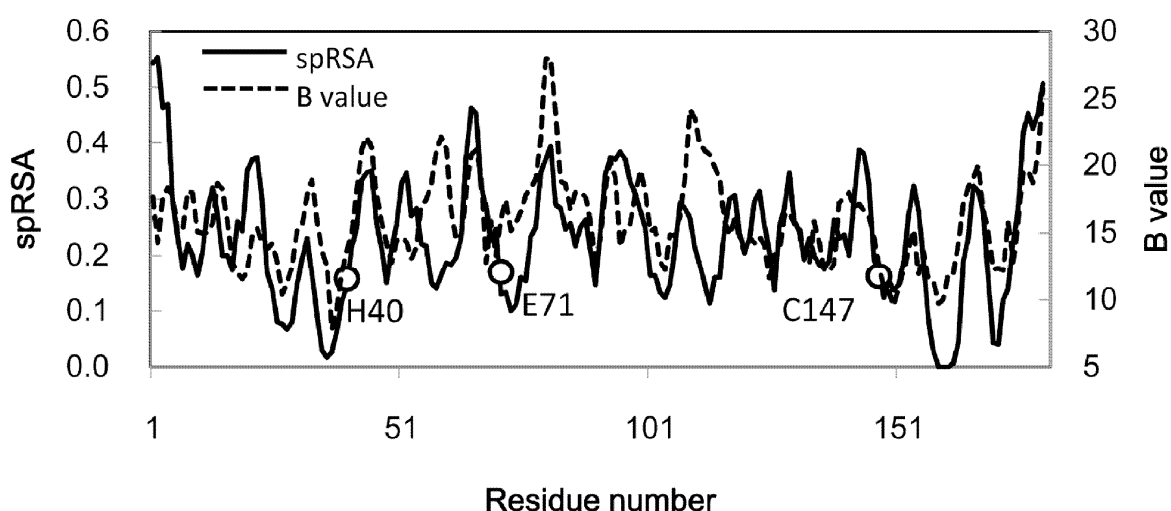


Fig. 9. B-value (dashed line) and spRSA (solid line) profiles of type 2 rhinovirus 3c protease (1CQQ). The catalytic residues, H40, E71 and C147, are labelled as open circle.

Since the prediction of solvent accessibility (Ahmad et al., 2003; Kim & Park, 2004) as well as the prediction of secondary structure elements (Qian & Sejnowski, 1988; Rost & Sander, 1993; Jones, 1999) from protein sequences are among the earliest and the best developed methodologies in computational biology, our findings suggest an indirect way of predicting B-factors from sequences - approaches based on various machine-learning techniques such as the support vector machines or the neural networks have been quite successful in the prediction of real value solvent accessibility, and we can borrow these methods and directly apply them to the prediction of B-factors by the use of the smoothed relative solvent accessibility. We predict the solvent accessibility from sequence using the support vector machine method. The results are comparable to that of the current best methods, with an average correlation coefficient of 0.53 between the B-factor and spRSA profiles over a data set of 972 proteins.

## 6. Acknowledgements

Thanks to Tzu-Hui Yang for special supports. This research was supported by the National Science Council and ATU from the Ministry of Education, Taiwan, R.O.C.



## 7. References

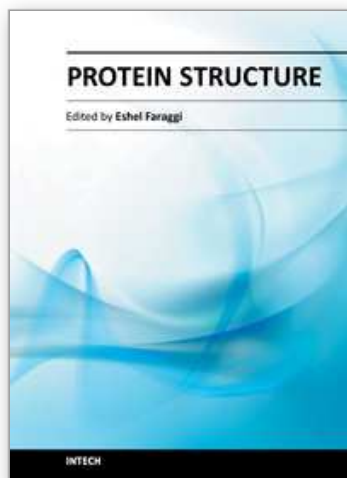
- Ahmad, S., Gromiha, M.M. & Sarai, A. (2003). Real value prediction of solvent accessibility from amino acid sequence. *Proteins*. Vol. 50, No. 4, (March 2003), pp 629-635, ISSN 1097-0134
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. Vol. 25, No. 17, (September 1997), pp 3389-3402, ISSN 0305-1048
- Bahar, I., Atilgan, A.R. & Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*. Vol. 2, No. 3, (January 1997), pp 173-181, ISSN 1359-0278
- Carugo, O. & Argos, P. (1997). Correlation between side chain mobility and conformation in protein structures. *Protein Engineering*. Vol. 10, No. 7, (July 1997), pp 777-787, ISSN 1741-0126
- Chang, C.C. & Lin, C.J. (2001). LIBSVM: a library for support vector machines. Available from: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, No.,
- Chen, Y.C., Lin, S.C., Lin, C.J. & Hwang, J.K. (2004). Prediction of the bonding states of cysteines using the support vector machines based on multiple feature vectors and cysteine state sequences. *Proteins-Structure Function and Bioinformatics*. Vol. 55, No. 4, (June 2004), pp 1036-1042, ISSN 0887-3585
- Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*. Vol. 221, No. 4612, (August 1983), pp 709-713, ISSN 1095-9203
- Ding, C.H.Q. & Dubchak, I. (2001). Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics*. Vol. 17, No. 4, (April 2001), pp 349-358, ISSN 1367-4803
- Dobson, P.D. & Doig, A.J. (2003). Distinguishing enzyme structures from non-enzymes without alignments. *Journal of Molecular Biology*. Vol. 330, No. 4, (July 2003), pp 771-783, ISSN 0022-2836
- Eisenberg, D. & McLachlan, A.D. (1986). Solvation energy in protein folding and binding. *Nature*. Vol. 319, No. 6050, (January 1986), pp 199-203, ISSN 1476-4687
- Fenimore, P.W., Frauenfelder, H., McMahon, B.H. & Young, R.D. (2004). Bulk-solvent and hydration-shell fluctuations, similar to alpha- and beta-fluctuations in glasses, control protein motions and functions. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 101, No. 40, (October 2004), pp 14408-14413, ISSN 0027-8424
- Gromiha, M.M., Oobatake, M., Kono, H., Uedaira, H. & Sarai, A. (1999). Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations. *Protein Engineering*. Vol. 12, No. 7, (July 1999), pp 549-555, ISSN 0269-2139
- Halle, B. (2002). Flexibility and packing in proteins. *Proceedings of the National Academy of Sciences*. Vol. 99, No. 3, (February 2002), pp 1274-1279, ISSN 1091-6490
- Hua, S.J. & Sun, Z.R. (2001). A novel method of protein secondary structure prediction with high segment overlap measure: Support vector machine approach. *Journal of Molecular Biology*. Vol. 308, No. 2, (April 2001), pp 397-407, ISSN 0022-2836



- Hua, S.J. & Sun, Z.R. (2001). Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*. Vol. 17, No. 8, (August 2001), pp 721-728, ISSN 1367-4803
- Huang, S.W., Shih, C.H., Lin, C.P. & Hwang, J.K. (2008). Prediction of NMR order parameters in proteins using weighted protein contact-number model. *Theoretical Chemistry Accounts*. Vol. 121, No. 3-4, (October 2008), pp 197-200, ISSN 1432-881X
- Huang, S.W., Yu, S.H., Shih, C.H., Guan, H.W., Huang, T.T. & Hwang, J.K. (2011). On the Relationship Between Catalytic Residues and their Protein Contact Number. *Current Protein and Peptide Science*. Vol. 12, No. 6, (September 2011), pp 574-579, ISSN 1389-2037
- Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*. Vol. 292, No. 2, (September 1999), pp 195-202, ISSN 0022-2836
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. Vol. 22, No. 12, (December 1983), pp 2577-2637, ISSN 1097-0282
- Kim, H. & Park, H. (2003). Protein secondary structure prediction based on an improved support vector machines approach. *Protein Engineering*. Vol. 16, No. 8, (August 2003), pp 553-560, ISSN 0269-2139
- Kim, H. & Park, H. (2004). Prediction of protein relative solvent accessibility with support vector machines and long-range interaction 3D local descriptor. *Proteins: Structure Function and Genetics*. Vol. 54, No. 3, (February 2004), pp 557-562, ISSN 0887-3585
- Kundu, S., Melton, J.S., Sorensen, D.C. & Phillips, G.N. (2002). Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models. *Biophysical journal*. Vol. 83, No. 2, (August 2002), pp 723-732, ISSN 0006-3495
- Levitt, M. & Warshel, A. (1975). Computer simulation of protein folding. *Nature*. Vol. 253, No. 5494, (February 1975), pp 694-8, ISSN 0028-0836
- Lin, C.P., Huang, S.W., Lai, Y.L., Yen, S.C., Shih, C.H., Lu, C.H., Huang, C.C. & Hwang, J.K. (2008). Deriving protein dynamical properties from weighted protein contact number. *Proteins-Structure Function and Bioinformatics*. Vol. 72, No. 3, (August 2008), pp 929-935, ISSN 0887-3585
- Liu, S., Zhang, C., Liang, S. & Zhou, Y. (2007). Fold recognition by concurrent use of solvent accessibility and residue depth. *Proteins: Structure, Function, and Bioinformatics*. Vol. 68, No. 3, (August 2007), pp 636-645, ISSN 1097-0134
- Lu, C.H., Huang, S.W., Lai, Y.L., Lin, C.P., Shih, C.H., Huang, C.C., Hsu, W.L. & Hwang, J.K. (2008). On the relationship between the protein structure and protein dynamics. *Proteins: Structure, Function, and Bioinformatics*. Vol. 72, No. 2, (August 2008), pp 625-634, ISSN 1097-0134
- McCammon, J.A., Gelin, B.R. & Karplus, M. (1977). Dynamics of folded proteins. *Nature*. Vol. 267, No. 5612, (June 1977), pp 585-90, ISSN 0028-0836
- McGuffin, L.J., Bryson, K. & Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*. Vol. 16, No. 4, (April 2000), pp 404-405, ISSN 1367-4803
- Ming, D., Kong, Y., Lambert, M.A., Huang, Z. & Ma, J. (2002). How to describe protein motion without amino acid sequence and atomic coordinates. *Proceedings of the National Academy of Sciences*. Vol. 99, No. 13, (June 2002), pp 8620-5, ISSN 1091-6490

- Noguchi, T. & Akiyama, Y. (2003). PDB-REPRDB: a database of representative protein chains from the Protein Data Bank (PDB) in 2003. *Nucleic Acids Research*. Vol. 31, No. 1, (January 2003), pp 492-493, ISSN 0305-1048
- Qian, N. & Sejnowski, T.J. (1988). Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology*. Vol. 202, No. 4, (August 1988), pp 865-884, ISSN 0022-2836
- Radivojac, P., Obradovic, Z., Smith, D.K., Zhu, G., Vucetic, S., Brown, C.J., Lawson, J.D. & Dunker, A.K. (2004). Protein flexibility and intrinsic disorder. *Protein Science*. Vol. 13, No. 1, (January 2004), pp 71-80, ISSN 1469-896X
- Rost, B. & Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *Journal of Molecular Biology*. Vol. 232, No. 2, (July 1993), pp 584-599, ISSN 0022-2836
- Sheriff, S., Hendrickson, W.A., Stenkamp, R.E., Sieker, L.C. & Jensen, L.H. (1985). Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrins. *Proceedings of the National Academy of Sciences*. Vol. 82, No. 4, (February 1985), pp 1104-1107, ISSN 1091-6490
- Shih, C.H., Huang, S.W., Yen, S.C., Lai, Y.L., Yu, S.H. & Hwang, J.K. (2007). A simple way to compute protein dynamics without a mechanical model. *Proteins-Structure Function and Bioinformatics*. Vol. 68, No. 1, (July 2007), pp 34-38, ISSN 1097-0134
- Sim, J., Kim, S.Y. & Lee, J. (2005). Prediction of protein solvent accessibility using fuzzy k-nearest neighbor method. *Bioinformatics*. Vol. 21, No. 12, (June 15), pp 2844-2849, ISSN 1460-2059
- Tirion, M.M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters*. Vol. 77, No. 9, (August 1996), pp 1905-1908, ISSN 0031-9007
- Vapnik, V. (1995). *The Nature of statistical learning theory.*, Springer, ISBN 0-387-94559-8, New York
- Vihinen, M., Torkkila, E. & Riikonen, P. (1994). Accuracy of protein flexibility predictions. *Proteins-Structure Function and Bioinformatics*. Vol. 19, No. 2, (June 1994), pp 141-149, ISSN 0887-3585
- Ward, J.J., McGuffin, L.J., Buxton, B.F. & Jones, D.T. (2003). Secondary structure prediction with support vector machines. *Bioinformatics*. Vol. 19, No. 13, (September 2003), pp 1650-1655, ISSN 1367-4803
- Warshel, A. (1976). Bicycle-pedal model for the first step in the vision process. *Nature*. Vol. 260, No. 5553, (April 1976), pp 679-683, ISSN 0028-0836
- Yang, L.W. & Bahar, I. (2005). Coupling between catalytic site and collective dynamics: A requirement for mechanochemical activity of enzymes. *Structure*. Vol. 13, No. 6, (June 2005), pp 893-904, ISSN 0969-2126
- Yu, C.S., Lin, C.J. & Hwang, J.K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Science*. Vol. 13, No. 5, (May 2004), pp 1402-1406, ISSN 0961-8368
- Yu, C.S., Wang, J.Y., Yang, J.M., Lyu, P.C., Lin, C.J. & Hwang, J.K. (2003). Fine-grained protein fold assignment by support vector machines using generalized npeptide coding schemes and jury voting from multiple-parameter sets. *Proteins-Structure Function and Bioinformatics*. Vol. 50, No. 4, (March 2003), pp 531-536, ISSN 0887-3585

- Yuan, Z., Bailey, T.L. & Teasdale, R.D. (2005). Prediction of protein B-factor profiles. *Proteins-Structure Function and Bioinformatics*. Vol. 58, No. 4, (March 2005), pp 905-912, ISSN 0887-3585
- Yuan, Z., Zhao, J. & Wang, Z.X. (2003). Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Engineering*. Vol. 16, No. 2, (February 2003), pp 109-114, ISSN 0269-2139
- Zhang, H., Zhang, T., Chen, K., Shen, S., Ruan, J. & Kurgan, L. (2009). On the relation between residue flexibility and local solvent accessibility in proteins. *Proteins: Structure, Function, and Bioinformatics*. Vol. 76, No. 3, (August 2009), pp 617-636, ISSN 1097-0134



## **Protein Structure**

Edited by Dr. Eshel Faraggi

ISBN 978-953-51-0555-8

Hard cover, 396 pages

**Publisher** InTech

**Published online** 20, April, 2012

**Published in print edition** April, 2012

Since the dawn of recorded history, and probably even before, men and women have been grasping at the mechanisms by which they themselves exist. Only relatively recently, did this grasp yield anything of substance, and only within the last several decades did the proteins play a pivotal role in this existence. In this expose on the topic of protein structure some of the current issues in this scientific field are discussed. The aim is that a non-expert can gain some appreciation for the intricacies involved, and in the current state of affairs. The expert meanwhile, we hope, can gain a deeper understanding of the topic.

### **How to reference**

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

Yu-Tung Chien, Jenn-Kang Hwang and Shao-Wei Huang (2012). On the Relationship Between Residue Solvent Exposure and Thermal Fluctuations in Proteins, Protein Structure, Dr. Eshel Faraggi (Ed.), ISBN: 978-953-51-0555-8, InTech, Available from: <http://www.intechopen.com/books/protein-structure/on-the-relationship-between-residue-solvent-exposure-and-thermal-fluctuations-in-proteins>

**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](http://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

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen