國立交通大學

生物資訊所

博士論文

利用加權蛋白質接觸數目推導蛋白質動力學特性

Deriving protein dynamical properties from weighted protein contact number

研究生：林志鵬

指導教授：黃鎮剛　教授

中華民國九十八年七月

利用加權蛋白質接觸數目推導蛋白質動力學特性

Deriving protein dynamical properties from weighted protein contact number

研 究 生：林志鵬　　Student：Chih-Peng Lin

指導教授：黃鎮剛　　Advisor：Jenn-Kang Hwang

國 立 交 通 大 學

生 物 資 訊 所

博 士 論 文

A Dissertation

Submitted to Institute of Bioinformatics

College of Biological Science and Technology

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

PhD

in

Bioinformatics

November 2008

Hsinchu, Taiwan, Republic of China

中華民國九十八年七月

利用加權接觸數目衍生出蛋白質的動力學特性

研 究 生：林志鵬　　指導教授：黃鎮剛博士

國立交通大學 生物資訊所 博士班

摘　　要

最近的研究顯示了蛋白質上的原子的平均平方位移(或稱之為 B 因子, B-factor)與該原子周圍接觸的原子數目(contact number)以及其到蛋白質的質量重心(center of mass)的平方距離有關。這樣的關係允許我們可以直接的藉由原子接觸數目和原子與質量重心的平方距離去計算 B 因子側寫(B-factor profile)。這兩種方法我們稱之為蛋白質接觸數目模型(contact number model)和蛋白質固定點模型(protein fixed-point model)，因為此兩種方法不僅不需要使用分子模擬來獲得原子軌跡整合，也不需要針對矩陣去做對角化，而使得這兩種方法變得相當具有吸引力。因此，蛋白質接觸數目模型和蛋白質固定點模型能夠應用在大體積且結構複雜的蛋白質，並且可作為一種高通量(high throughput)的計算工具去計算蛋白質內原子的擾動。在這個研究裡，我們展示藉由融合這兩種模型的特性所產生的加權蛋白質接觸數目模型，來使得計算蛋白質內部的原子運動的準確性能夠被更進一步的提昇。我們藉由除以兩個原子接觸的平方距離來作為加權的方法，並且分析將蛋白質的接觸距離加上不同的切斷半徑(cut-off radius)之後的計算結果。此外我們也展示了這樣的模型也可以被用來計算原子運動的交互相關性(cross-correlation)。實際上，B 因子就是原子運動的自我相關性(auto-correlation)。我們使用了一個序列相關性小於 25%的非同源性的資料組，其包含了 972 條具有解析度小於 2.0 Å 的高解析度

X 射線結晶結構。針對這樣的資料組，經由 X 射線所得到的 B 因子與藉由加權蛋白質接觸數目模型去計算所得到的 B 因子，兩者之間的平均相關係數可達 0.61，比起原始的蛋白質接觸數目模型(0.51)或是其他的方法要來得更好。針對一些例子，我們的研究結果也顯示了經由加權蛋白質接觸數目模型計算所得到的交互關係圖像以及特徵值和特徵向量，與藉由正模分析法(normal mode analysis)和高斯網路模型(Gaussian Network Model)所得到結果，具有全面性的相似度。我們的結果不僅強調了蛋白質的動力特性與蛋白質集結特性有關之外，我們也相信這樣的方法對於研究蛋白質的結構和動力學之間的相關性有相當大的幫助。

Deriving protein dynamical properties from weighted protein contact number

Student: Chih-Peng Lin Advisor: Dr. Jenn-Kang Hwang

Institute of Bioinformatics

National Chiao Tung University

ABSTRACT

It has recently been shown that in proteins the atomic mean-square displacement (or B-factor) can be related to the number of the neighboring atoms (or protein contact number) and the square distance from the center of mass of a protein. This relationship allows one to compute the B-factor profiles directly from protein contact number and the square distance from center of mass. The two methods, referred to as the protein contact number model and the protein fixed-point model, are appealing, since they require neither trajectory integration nor matrix diagonalization. As a result, the protein contact number model and the protein fixed-point model can be applied to very large proteins and can be implemented as a high-throughput computational tool to compute atomic fluctuations in proteins. Here, we show that the properties of the two models can be integrated and further refined to that between the atomic mean-square displacement and the *weighted* protein contact-number, the weight being the square of the reciprocal distance between the contacting pair. In addition, we show that this relationship can be utilized to compute the cross-correlation of atomic motion (the B-factor is essentially the auto-correlation of atomic motion). For a nonhomologous dataset comprising 972 high-resolution X-ray protein structures (resolution < 2.0 Å and sequence identity < 25%), the mean correlation coefficient between the experimental and computed B-

factors based on the weighted protein contact-number model is 0.61, which is better than those of the original contact-number model (0.51) and other methods. We also show that the computed cross-correlation maps, eigenvalues and eigenvectors based on the weighted contact-number model are globally similar to those computed through Gaussian Network Model and normal model analysis for some selected cases. Our results underscore the relationship between protein dynamics and protein packing. We believe that our method will be useful in the study of the protein structure-dynamics relationship.

ACKNOWLEDGEMENT

一轉眼，就在交大博愛校區待了七年。目前為止，人生中最精彩的時刻幾乎都發生在這段期間。從一開始一起進來的好同學：志豪、力彰、志杰，學長姐：勇欣、玉菁、景盛，以及學弟妹：蔚倫、小操、少偉、建華、世瑜、士中、彥龍、啟文、瓊文、松桓、子琳、慧雯、乃文、人維。很開心在這段人生重要的過程當中與你們相遇。不論是在討論研究的腦力激盪，抑或是國內外的旅行踏青，甚至是日常生活中的打鬧鬥嘴，有了你們著實讓我成長不少，也快樂不少。

特別要感謝黃鎮剛老師在我研究的路上，不停給予諄諄教誨。在與您討論的過程中，總是讓我發現自己思考不夠縝密的地方。今後，我會帶著你所給予的信念及指導，繼續朝著無窮無盡的研究路上大步邁進。同時也要感謝生資所的其他老師們，你們在課堂所所給予的教導還有整個系所的研究設備，讓我有著最佳的學習環境。

最後感謝我的父母，漫長的學生生涯即將告一個段落，謝謝你們陪著我從小到大，度過一個一個的求學階段，你們是最辛苦的。謝謝你們，也謝謝所有曾經幫助我的，可是未能在致謝中提及的人們。

CONTENT

INTRODUCTION

Protein dynamics is dictated by protein structure. The dynamic properties of proteins result from a network of complex interactions like covalent bonding and nonbonded electrostatic or van der Waals interactions. To compute the dynamical properties of proteins, one usually resorts to molecular dynamics simulation[1-5], which consists of integrating long time trajectories of protein structure using empirical force field. Though molecular dynamics is a powerful method, it is computationally expensive. A recent study[3] show that a massive molecular dynamics simulations of 30 proteins using four different force fields in aqueous solution reportedly took computational time equivalent to around 50 years of CPU.

Due to the recent progress of structural biology research, the number of protein structures deposited in Protein Data Bank has nearly quadrupled since 2000. Hence, there is an increasing interest in developing efficient methods to compute protein dynamic properties from protein structures in a high-throughput fashion. For example, some studies have shown that simple energy functions that include the two most important characteristics of amino acids, mainly hydrogen bond formation capacity and hydrophobicity, may well account for the prediction of some supersecondary structures or tertiary folds in small proteins [6,7]. From the above point of view, the question "Is it possible to find a novel characteristic of a given protein to study protein dynamics with or even without the development of a new simple energy function based on the characteristics?" is interesting and valuable.

In 1996, Tirion developed a single-parameter Hookean potential and made it suitable for the contact interaction of all atoms in X-ray protein structures in a normal

mode analysis (NMA) of large-amplitude (low frequency) vibrations in elastic motions[8]. This approximation is based on a Gaussian distribution of interatomic distances about their equilibrium values. The simplicity of the postulated single-parameter Gaussian model and its success in predicting results for a complex system may have far-reaching consequences in understanding protein structures. Based on this theory, Bahar *et al.*[9] analyzed the connectivity matrix of nonbonded interactions in proteins and developed the elastic network model. The Elastic Network Model (ENM) or Gaussian Network Model (GNM)[8-10] provides an alternative for molecular dynamics in computing average dynamical properties. In the GNM, each Cα atom is connected through a single-parameter harmonic potential to its neighboring atoms that are within a certain cut-off distance, usually in the range of 7 to 10 Å. The elastic network model then builds a connectivity matrix (or called the Kirchhoff matrix), from which cross-correlations and auto-correlations of fluctuations of residues can be obtained through matrix diagonalization. Micheletti and co-workers[11] have developed a model based on a mean field theory[11-13] to study the dynamics of a protein. The shape of the protein is specified by the locations of the Cα atoms with 2 types of interactions: simple harmonic potential functions describing bonded interactions and Go-like functions describing nonbonded interactions. This model was applied to protein-protein interactions[12]. Zhou and co-workers[13] later extended this Cα-based model protein to a model based on all heavy atoms and they were able to make a more accurate prediction of the atomic mean-square displacement (or B-factor) using this extended model.

All of these methods mentioned above all supposed a complex or simple energy function in interresidue. But recent studies[14,15] showed that the B-factor is closely related to the number of noncovalent neighboring atoms or the atoms lying on the same

shell centered at the fixed point tend to have similar thermal fluctuations. For convenience, we will refer to the two methods as the protein contact number (CN) model and protein fixed point (PFP) model. The CN model and PFP model are appealing, since they predict the B-factor profiles directly form protein structures without either trajectory integration or matrix diagonalization. However, despite theirs simplicity, the CN model has been shown to be superior to the GNM for a small set of 38 structures[14] and the PFP model has been comparable with GNM for a large set of 972 structures[16]. Besides, when observing the contact numbers of residues and centers of mass in the proteins for the nonhomologous dataset (see Methods), we found the frequencies of residue contact numbers tend to have a normal distribution and the frequencies of center of mass tend to have larger contact numbers (a more crowded space environment). Figure 1 shows the frequencies of the residue contact numbers and figure 2 shows the frequencies of center of mass contact numbers of the centers of mass in proteins. This implies we may can combine the characteristics of the CN model and the PFP model to develop a new model which calculate the contact numbers within the effect of distances for every point (atom) of a protein instead of just using one point (the center of mass) of a protein to predict the B-factor. Here, we show that the CN model, which relates the B-factors to protein contact number, can be further improved if the protein contact number is scaled down by the square of the distance between the contacting pair and we call this model as weight contact number (WCN) model.

Large conformational transitions contain the relative movement of almost rigid structural regions in many proteins. The domain motions are important for various protein functions such as catalysis and regulation of activity. Besides, some proteins have at least two different conformations in functionally distinct states. The transition

from one to another form must be a major mode of internal motion. For example, in citrate synthase, this is a two-domain protein and have been shown[17,18] that coenzyme A binding induces the small domain to rotate around an axis which is close to the residue 274 and get a conclusion of this motion is the closure of the cleft between two domains in which the substrate binding site puts in. One of the best suited theoretical methods for studying motions in proteins is normal mode analysis [19-23](NMA) and the other one is GNM. Both methods can decompose the protein dynamics into a collection of motions which include large amplitude (low frequency) and small amplitude (high frequency) motions. Here, we show that cross-correlation between residues, vibration frequency and amplitudes with phases of collective motions can also be computed in the framework of the weighted contact number model. By diagonalizing the cross-correlation matrix computed by WCN, we can get the similar inverse of eigenvalues (vibration frequencies) and eigenvectors (vibration amplitudes and phases) within 3 lowest modes to those from NMA or GNM. It is surprising to directly link not only protein structures and thermal fluctuations but also protein structures and collective motions.

METHODS

Protein contact number model and weighted contact number model

The contact number ν_i of the i^{th} residue is defined as the number of the neighboring residues whose Cα atoms are within a cut-off radius r_0 of that of the i^{th} residue.

$$\nu_i = \sum_{j \neq i}^{N} H(r_0 - r_{ij}) \tag{1}$$

where r_{ij} is the distance between Cα atoms of residue i and j, and $H(x) = 1$ if $x \geq 0$ and $H(x) = 0$ if $x < 0$. Eq. 1 defines an integral contact number and gives an equal unitary weight to every contacting atom regardless of its distance to the central atom. Figure 3 schematically illustrates the contact number (or CN) model.

To improve the CN model, we must think about PFP model first as what was mentioned in the introduction section. In the PFP model, the predicted B-factor values will increase by the main effect of the square of the distance between a Cα atom and the center of mass. In other words, the ability of the center of mass to stabilize atomic fluctuations will decrease squarely by the distance. So, we make an assumption that the contact effect between atoms may have similar property to affect atomic fluctuations as PFP model. In this way, the distance-dependent contact number ν_i' of the i^{th} residue is defined as

$$\nu_i' = \sum_{j \neq i}^{N} H(r_0 - r_{ij}) / r_{ij}^2$$

which defines a real-valued contact number, i.e., the integral contact number weighted by the square of the reciprocal distance between the contact pair. Due to the influence of

the cut-off radius r_0 (see results) and the fast decay of the factor $1/r_{ij}^2$ at large separation r_{ij}, the real-valued contact number can be simplified to

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

Figure 4 schematically illustrates the weighted contact number (or WCN) model. We will refer to ν as the CN, while ν' (Eq. 2) as the WCN.

Assessment of profiles comparisons with statistical analysis

The CN (or WCN) profile of a protein of N residues is defined as

$$\mathbf{w} = (\omega_1, \omega_2, \ldots \omega_N) \tag{3}$$

where ω_i is defined as the reciprocal contact number, i. e., $\omega_i = 1/\nu_i$ or $\omega_i = 1/\nu_i'$. The X-ray B-factor profile is denoted as

$$\mathbf{b} = (b_1, b_2, \ldots b_N) \tag{4}$$

where b_i is the B-factor of the Cα atom of the i^{th} residue taken from the PDB file.

For the purpose of easy comparison, we will normalize both the CN (or WCN) and the B-factor profiles to the corresponding z-scores:

$$z_{x_i} = (x_i - \bar{x})/\sigma_x \tag{5}$$

where $\bar{x}$ and σ_x are the mean and standard deviation of x. Here x designates b or ω. The normalized CN (or WCN) and the B-factor profiles are denoted by the vectors $\mathbf{z}_\omega$ and $\mathbf{z}_b$, respectively. In the CN model, the cut-off distance is set to 7.35Å, which corresponds to the second minimum of the average contact-pair distribution of protein

structures[14]. For the prediction assessment, we use two types of correlation coefficients between the profiles. One is the Pearson's linear correlation coefficient,

$$c = \frac{\sum \left(z_{b_i} - \bar{z}_b\right)\left(z_{w_i} - \bar{z}_w\right)}{\sqrt{\sum \left(z_{b_i} - \bar{z}_b\right)^2 \sum \left(z_{w_i} - \bar{z}_w\right)^2}} \tag{6}$$

if $c = 1$, two profiles are perfectly correlated; if $1 > c \geq 0.5$, two profiles are strongly correlated; if $0.5 > c > 0$, two profiles are weakly correlated; if $c = 0$, they are completely independent of each other; if $0 > c > -0.5$, two profiles are weakly anti-correlated; if $-0.5 \geq c > -1$, two profiles are strongly anti-correlated; if $c = -1$, they are perfectly anti-correlated. But we think the extreme values in profiles may affect Pearson's linear correlation coefficient results. To solve this problem, the other one is Spearman's rank correlation coefficient, ρ, which is based on the rank order of z-scores in profiles rather than its actual value,

$$\rho = 1 - \frac{6\sum \left(Z_{B_i} - Z_{W_i}\right)^2}{n\left(n^2 - 1\right)} \tag{7}$$

where n is the number of values in the dataset, B_i and W_i denote the ranks of corresponding values $\mathbf{z}_b$ and $\mathbf{z}_\omega$. Four indices of prediction assessment are used to compare the global performances of different methods for a data set: $\bar{c}$, the average Pearson's linear correlation coefficient, $c_{0.5}$, the fraction of number of structures with $c \geq 0.5$, $\bar{\rho}$, the average Spearman's rank correlation coefficient, and $\rho_{0.5}$, the fraction of number of structures with $\rho \geq 0.5$.

To know if the performances of correlation coefficients have significant differences between the models, we use Student t-test to check,

$$t = \frac{\bar{Z}_D - \mu_0}{S_D / \sqrt{N}} \tag{8}$$

where $\bar{Z}_D$ is the average and S_D is the standard deviation of differences in correlation coefficients produced by two of the models. The constant μ_0 is set to 0 because we want to know whether the average of the difference is significantly different than 0. N is the sample size. Once a t value is determined, a p-value can be found using a table of values from Student's t-distribution. If the p-value is lower than 0.05, it indicates that the distributions of the results are significantly different from each other.

Cross-correlation between residues by WCN model

The normalized correlation between fluctuations of atom i and j is defined as

$$C_{ij} = \frac{\langle \delta\mathbf{r}_i \cdot \delta\mathbf{r}_j \rangle}{\sqrt{\langle \delta\mathbf{r}_i \cdot \delta\mathbf{r}_i \rangle \langle \delta\mathbf{r}_j \cdot \delta\mathbf{r}_j \rangle}} \tag{9}$$

where $\delta\mathbf{r}_i$ and $\delta\mathbf{r}_j$ are the fluctuations of the atom i and j, respectively, around their equilibrium positions. In the framework of the WCN model, we formulate the correlation term W_{ij} between residue i and residue j as

$$W_{ij} = \left(\sum_{k \neq i,j}^{N} \frac{1}{r_{ik} r_{jk} / \log(e + r_{ij})} \right)^{-1} \left\langle \sum_{k \neq i,j}^{N} \hat{\mathbf{x}}_{ik} \cdot \hat{\mathbf{x}}_{jk} \right\rangle \tag{10}$$

where $\hat{\mathbf{x}}_{ik}$ and $\hat{\mathbf{x}}_{jk}$ are the unit vectors in the direction of $\mathbf{r}_i - \mathbf{r}_k$ and $\mathbf{r}_j - \mathbf{r}_k$, respectively. See Figure 5 for a schematically representation of Eq. 10. Note that when $i = j$, W_{ii} reduces to ω_i^{-1}. Then we denote the cross-correlation matrix as

$C = (w_{11}, w_{12}, w_{13}, ...w_{1n}, ...w_{NN})$ when the protein size is N. It must be noted that we WC model provides a straightforward way to computed the cross-correlation matrix directly from protein structure without the use of the Hessian matrix, whereas other method like the NMA or GNM needs to first diagonalize the Hessian matrix in order to obtain the cross-correlation matrix.

The WCN model also provides a way to compute the normal modes of protein motion. At first, the cross-correlation matrix C is diagonalized,

$$C = U\Lambda U^{-1} \tag{11}$$

where U and Λ is the eigenvectors and eigenvalues respectively. Since the cross-correlation matrix is the inverse of the Hessian matrix, the eigenvectors U are the amplitudes and phases and the inverse of eigenvalues Λ^{-1} are the frequencies of the residues motion vibration for each corresponding modes.

To compare the cross-correlation matrix generated by WCN model with the ones generated by NMA and GNM, we normalize the cross-correlation matrix $C = (w_{11}, w_{12}, w_{13}, ...w_{1N}, ...w_{NN})$ by the following equation:

$$W'_{ij} = \frac{W_{ij}}{\sqrt{W_{ii}W_{jj}}} \tag{12}$$

Dataset

We selected from PDB-REPRDB[24] 972 protein chains of length ≥ 60. Their structures are solved by X-ray crystallography with resolution ≤ 2.0 Å and R-factors ≤0.2. All chains are of pair-wise sequence identity ≤ 25%. The chains of the data

set are listed in the appendix. In the data set, the protein size ranges from 60 to 1520 with an average protein size around 300 residues. The resolution of the X-ray structures ranges from 0.73 Å to 2.0 Å with an average structural resolution 1.57 Å. The distribution of protein size (i.e., the number of residues) and structural resolution of the data set are shown in Figure 6.

RESULTS

Comparison of CN model and WCN model based on different cut-off distances

At first, to search which cut-off distance is suitable for WCN model and realize the influence of the cut-off distance on B-factor prediction, we use two definitions for cut-off distances to see the performance of average correlation coefficient in CN and WCN model. One definition is to set the cut-off distance from 3Å to 30Å; the other is to set the cut-off distance based on the percentage of a protein size which was calculated by the maximum Cα-Cα distance between residues in a protein.

Figure 7-10 show the curves of the Pearson's linear correlation coefficients and Spearman's rank correlation coefficients between the X-ray B-factors of Cα atoms and those computed by the CN model and the WCN model. In CN model, the curves of average correlation coefficients ascend when the cut-off distance increases at beginning, and they achieve the highest point at 15Å (the value is in contrast with the previous study[14] that the cut-of distance of CN model is 7.35 Å) or 25% protein size. After the highest point, the curves will descend if the cut-off distance still increases (see Figure 7 and Figure 8). In WCN model, the curves of the average correlation coefficients will ascend when the cut-off distance increases and reach a plateau (see Figure 9 and Figure 10). It implies that we don't need to worry which cut-off distance is appropriate when using WCN model to analyze protein dynamics. Finally, we don't set any cut-off distance as a default parameter for WCN model due to its property.

Comparison of CN model and WCN model based on all atoms and entire residue

The results of the above section are based on only Cα atoms. Under this criteria, in the WCN model with no cut-off distance, the average correlation coefficient are $\bar{c} = 0.61$ and $\bar{\rho} = 0.63$. The fraction of structures with a correlation coefficient ≥ 0.5 are $c_{0.5} = 79\%$ and $\rho_{0.5} = 82\%$. The CN model with the cut-off distance 7.35 Å, which is bases on previous study[14], yields poorer results: $\bar{c} = 0.51, \bar{\rho} = 0.49$, $c_{0.5} = 54\%$ and $\rho_{0.5} = 50\%$. The effect of the term $1/r_{ij}^2$, which is missing in the CN model, on the results is significant. But, if the average X-Ray B-factors for the entire residue are used, the WCN model yields $\bar{c} = 0.60$, $\bar{\rho} = 0.63$, $c_{0.5} = 79\%$ and $\rho_{0.5} = 83\%$. The CN model yields $\bar{c} = 0.50$, $\bar{\rho} = 0.50$, $c_{0.5} = 54\%$ and $\rho_{0.5} = 53\%$. These results are not much different from those based on the Cα atoms. These results are shown as a histogram of the Pearson's linear correlation coefficients and Spearman's rank correlation coefficients in Figure 12. To rely on this conclusion, we can simply use Cα atoms to represent entire residues.

To study the effects of all atoms on the computed B-factors, we calculate the WCN and CN B-factor profiles using all non-hydrogen atoms (i.e., C, N, O and S atoms) of proteins. If all heavy atoms are included in calculation in WCN model, the results are $\bar{c} = 0.62$, $\bar{\rho} = 0.65$, $c_{0.5} = 85\%$ and $\rho_{0.5} = 87\%$, while the all-atom CN model yields $\bar{c} = 0.56$, $\bar{\rho} = 0.58$, $c_{0.5} = 77\%$ and $\rho_{0.5} = 79\%$. Both results are better than those based on only Cα atoms. These results are shown as a histogram in Figure 11.

The computed B-factor profiles using CN model, GNM model and WCN model

The currently available GNM program[25] uses only Cα atoms for proteins in the calculation of the B-factors, therefore, for the sake of comparison, the following calculated results are based on the Cα atoms in both the WCN and CN models.

Figure 13 shows the histogram of the Pearson's linear correlation coefficients and Spearman's rank correlation coefficients between the X-ray B-factors of Cα atoms and those computed by the WCN model, the CN model and the GNM model. The GNM yields $\bar{c} = 0.56$, $\bar{\rho} = 0.57$, $c_{0.5} = 69\%$ and $\rho_{0.5} = 73\%$. This is in contrast with the previous study[14] that the CN model performs better than the GNM. It should be noted that, however, the previous study was conducted on a much smaller data set of 38 structures. Though the correlation-coefficient distributions of these models seems to look quite different, we perform additional Student t-test to check these distributions using the statistical package R[26]. The p-values of the WCN-GNM, GNM-CN and WCN-CN are all smaller than 2.20×10^{-16}, indicating that the distributions are significantly different from each other. On the other hand, we notice that a better correlation between the WCN model and the GNM ($c = 0.86$, $\rho = 0.91$) than that between the WCN and the CN model ($c = 0.67$, $\rho = 0.73$). These results are shown in Figure 14. Thought both the CN model and the GNM consider the contributions from any atoms to be identical as long as they are within the cut-off distance, the CN model completely ignores those atoms that are out of the cut-off range, while the GNM takes them into account implicitly through the network. The WCN model considers the contributions from any atoms with a weighting factor $1/r_{ij}^2$. In Figure 15 and Figure 16, we compare the B-factor profiles of flavocytochrome c3 (1Y0P:A) and human

ppGalNAcT-2 (2FFU:A) computed by 3 methods with each compared with the X-ray B-factor profile. The WCN and the GNM B-factor profiles agrees relatively well with the X-ray B-factors, but the CN B-factor profile appears to be much more rugged, probably due to the artificial cut-off effect.

The breakdown analysis for the accuracy of CN model, GNM model and WCN model

Classified in terms of the SCOP classes, the structures in the dataset has 111 all-α proteins, 181 all-β proteins, 245 α/β proteins, 193 α+β proteins, 15 multi-domain proteins, 11 membrane and call surface proteins and peptides, 22 small proteins, 5 coiled coil proteins, 1 designed protein and 188 undefined in SCOP. In Table 1 and Table 2, we compare the statistics of the performances of different models for the 4 major SCOP classes: all- α proteins, all- β proteins, α/β proteins and α+β proteins, since the other classes have much smaller sample size (1-22). All 3 models perform worst for the all-α proteins and perform better when the protein structure has β-sheets, especially for all-β proteins. In general, the trends of the performance of these methods appear to be similar. We compare the performance of all 3 methods as a function of protein size, X-ray resolution and R-factor in Figure 17, Figure 18 and Figure 19. We remove the proteins in the marginal regions (for example, protein size 600-1550) due to the smaller size in those regions. One notices that, while the average performance of the CN model shows slightly downward trend for proteins of larger size or lower resolution, the average performances of the WCN model and the GNM appear to be relatively unchanged in the range of protein properties studied.

Case studies for the discrepancy of B-factor profiles between theory and experiments at specific regions in some proteins

We note that WCN model doesn't always accurately reproduce the experimental B-factor profiles in some proteins, and even they have negative correlation with WCN predictions. For example, we show in Figure 20 the case with one of the bad correlation ($c = -0.11$, $\rho = 0.00$): 1GK8:I, the small chain of Rubisco from the green alga *Chlamydomonas reinhardtii*. On close inspection, we found that this structure is in fact part of a larger biological unit[27] (Figure 21A) and, if the complete biological unit is included in calculation, the WCN correlation of 1GK8:I is improved to $c = 0.75$ and $\rho = 0.79$ (Figure 21B). We show two more examples in Figure 22 and Figure 23, comparing WCN profiles with and without including biological units in calculation. By this inspection, we know some proteins in our data set are in fact part of larger biological units, which are the assumed functional form of the macromolecule. In the PDB, the biological units are built from the crystallographic space group using symmetry operation. Currently, the coordinates of biological units can be obtained from either PDB or PQS[28]. The PDB and PQS biological units agree on 82% of entries[29]. In this work, we use the PDB biological units for computation. We computed the WCN B-factor profiles of the same proteins of the data set with other parts of the whole biological units (if any) taken into consideration. The Figure 24 shows the histogram WCN model yields $\bar{c} = 0.65$, $\bar{\rho} = 0.69$, $c_{0.5} = 86\%$, and $\rho_{0.5} = 0.89$ which are better than the previous results.

Applications to large proteins

The WCN model can be readily applied to large proteins, since its memory requirement is of the order $O(N)$, where N is the size of protein. An example is the 50S ribosomal subunit (1YJW) comprising 3774 residues. Figure 25 shows its structure and computed B-factor profile using the WCN model. On the other hand, the oGNM[25], the web version of the GNM, is unable to return the B-factor profile of 1YJW.

The effect of the exponential value of the reciprocal distance between contact pair for weighting the contact number

According to the results which WCN model has a better performance than CN model, we think the "weight" plays an important role when we compute the protein B-factors. To verify the "weight" effect of WCN model on the B-factor prediction, we have a test using different exponential values for the weight in WCN model and try to find out which exponential value has the best performance in average correlation coefficients. From the Figure 26, we find WCN model reach the best performance when the exponential value is 2.3 in Pearson's linear correlation coefficient evaluation and 1.9 Spearman's rank correlation coefficient evaluation. This result is similar to the classical physical phenomenon which the effect of force always decays in the square of the reciprocal distance between the two objects. To keep the physical meaning, we still choose 2 as the exponential value.

The cross-correlation of fluctuations between residues

The knowledge of cross-correlated motion between residues is useful in understanding long-range communication[30,31] and large domain movements relevant to protein

function[32,33]. The cross-correlation matrix can be computed through the normal mode analysis (NMA)[19-21]. In this method, i.e. NMA, the protein structure must be first optimized through energy minimized. The second derivative matrix of the total potential function (also called the Hessian matrix) is computed from the optimized structure. The cross-correlation between atom fluctuations is computed from the eigenvalues (i.e. vibration frequencies) and the eigenvectors (i.e. vibration amplitudes and phases) of the Hessian matrix using $\left\langle \delta\mathbf{r}_i \cdot \delta\mathbf{r}_j \right\rangle \sim \sum_k U_{ik}U_{jk}/\lambda_k$, where λ_k is the eigenvalue of the k^{th} mode and U_{ik} is the i^{th} component of the eigenvector of the k^{th} mode. Instead of going through procedures of energy minimization and matrix diagonalization as in the case of NMA, we can compute the cross-correlation map directly from protein structure using Eq. 10. We eliminated those proteins whose residue sizes are larger than 1000 amino acids in the dataset because of the computational limit of the NMA program, or those ones whose structures can't be optimized through the energy minimization process. Finally, we get 961 proteins and calculate the correlation coefficients of the cross-correlation maps, eigenvalues and eigenvectors generated from WCN, GNM and NMA. Again, we use both of Pearson's linear correlation coefficients and Spearman's rank correlation coefficients to see if they have different results or conclusions.

In Table3, the results show the cross-correlation matrices computed by WCN correlates well with those by GNM or NMA, but the correlation between WCN and NMA are slightly better than WCN and GNM. Figure 27 shows the normalized cross-correlation map of 1RWH:A whose cross-correlation map was computed by WCN; and compares it with those of GNM and NMA. Currently, there is no "experimental" correlation map (except for the diagonal terms, which correspond to the X-ray B-factors)

as a reference standard. However, the similarity of these types of computed correlation maps indicates that the WCN model provides a quick alternative to GNM or NMA to compute the correlation of motions in proteins. For 1RWH:A, Pearson's linear correlation coefficient between WCN and NMA is $c_{WCN-NMA} = 0.85$; WCN and GNM, $c_{WCN-GNM} = 0.79$; and GNM and NMA, $c_{GNM-NMA} = 0.85$. Spearman correlation coefficients are $\rho_{WCN-NMA} = 0.88$, $\rho_{WCN-GNM} = 0.79$ and $\rho_{GNM-NMA} = 0.87$ .We also show more cross-correlation maps as examples in Figure 28-31.

The inverse of the eigenvalues, i.e., vibration frequencies, is equal to the eigenvalues of Hessian matrix, Table 4 only shows the Pearson's correlation coefficient because the inverse of eigenvalues (vibration frequencies) have been ranked and Spearman correlation coefficient will be 1. We find WCN has very high correlation coefficients with GNM ($\bar{c} = 0.99$) and NMA ($\bar{c} = 0.96$) when considering all modes. From the equation $\langle \delta \mathbf{r}_i \cdot \delta \mathbf{r}_j \rangle \sim \sum_k U_{ik} U_{jk} / \lambda_k$, the correlation between atom fluctuations is affected more when λ_k (vibration frequencies) is smaller. Therefore, we also consider the first 20 slowest modes to see the correlation, and the results of the first 20 slowest modes are slightly lower than the ones of all modes. Figure 32-35 show the inverse of eigenvalues of all modes and the first slowest 20 modes of 1CVR:A and 1ARB respectively.

Finally, we focus on the 3 sets of eigenvectors of modes with first 3 slowest modes. These eigenvector values represent amplitudes and the eigenvector phases represent the relative directions of slow vibration motions which can be regard as global scale motions. Table 5 shows the correlation coefficients of the eigenvectors generated from

WCN, GNM and NMA in first mode, second mode and third mode. WCN correlates well with GNM and NMA in the first mode, but the other two modes reveal more differences. Figure 36-38 show the ribbon diagrams which are colored by the positive (blue) and negative (red) phases and the eigenvectors distributions in first mode, second mode and third mode of 1CVR:A. Previous study[33] shows that the catalytic residues are usually immobilized in order to maintain the delicate arrangement of functional group. For 1CVR:A, Figure 39 shows the protein consists of the catalytic domain subdivided into A-subdomains (yellow) and B-subdomains (green), and the IgSF domain (purple)[34]. The subdomains of catalytic domain almost can be indentified by the phases of eigenvectors in first mode and second mode. The other example is 1ARB. Figure 40-42 show the eigenvectors distributions of 1ARB. For 1ARB, Figure 43 shows the catalytic triad comprises His57, Asp113, and Ser194 (green)[35]. The three residues lie on the interfaces of the phases in first mode, second mode.

DISCUSSION

A simulation of a polypeptide chain folding into a native structure must include many and complicated molecular force fields[1,2,4] . If people get protein dynamics directly from native protein structure properties, they will be able to avoid the complicated molecular simulation computation. Previous studies show CN model and PFP model can derive protein dynamics simply from protein structures (i.e., without the knowledge of protein sequences) and have good performance in B-factor prediction. In advance, we find the more close relationship between the thermal fluctuations of proteins and the distance-dependent protein contact number allows one to compute dynamic properties of proteins more accurately than CN model and PFP model. This method, i.e., the WCN model, does not presuppose a mechanical model[8-10] as well as the potential functions associated with that model as other methods: molecular dynamics is based on sophisticated molecular force field[1,2,4], while the GNM assumes a harmonic oscillator model for proteins with their structures described in terms of a collection of masses connected to each other through a spring of a uniform force constant. Besides, GNM or other methods based on molecular force field all need one or more parameters to perform a matrix creation or a molecular simulation.

We showed that the WCN model can produce more accurate B-factor profiles than other methods and just need the topological structure of proteins but without any parameter. Because of its simplicity, this method is very efficient when proteins have large sizes and is convenient to study the linkage between protein function and dynamics. From the study of biological units, we noted that the predicted B-factor will fall above the experimental B-factor at the regions of protein-protein interfaces because

of the lack of contacting effect between atoms. We think this may be a possible way to investigate protein-protein or protein-ligand interaction.

Traditionally, one who wants to use protein contact number as a feature to analyse protein structure dynamics or functions will need to decide an optical cut-off distance to catch appropriate neighboring atom information and assume these nightboring atoms contribute contact effect equally[9,14]. In WCN model, we don't need to find an optical cut-off distance but include all atoms contact effect which is scaled down by the square of the distance between atoms. In addition, we show for the first time that cross-correlation between residues can be computed directly from weighted protein contact number and correlates well with the ones computed by GNM or NMA. By directly computing eigenvalues and eigenvectors from the cross-correlation matrix, it suggests that it is possible to skip the Hessian matrix creation and obtain the normal mode motion from correlation maps computed by any method.

By inverting the Hessian matrix, (in GNM, it is the Kirchhoff matrix), one can obtain the cross-correlation matrix, and the diagonal elements of the matrix will correspond to the X-ray B-factors, from which the information of protein dynamics can be derived. On the other hand, the WCN model computes the correlation matrix (Eq. 10) directly from the protein structure without first computing the Hessian matrix, thus avoiding assuming any explicit form of the potential functions. In this way, the WCN model can be seen as complementary to the GNM.

Shih *et al.*[15] have recently showed that the atoms in proteins lying on the same spherical shell centered at the fixed point tend to have similar thermal fluctuations. We will refer to this model as the protein fixed-point (PFP) model. The PFP model assumes

that the protein centroid in the simple single-domain protein is the position of the smallest fluctuations, i.e., the fixed point. The PFP model, like the WCN model, provides a simple way to compute both auto-correlation and cross-correlation between residues in reasonable accuracy. It is not hard to show that the WCN model can reduce to the PFP model; the diagonal term of Eq. 10 reduces to the following form

$$W_{ij} = \left(\sum_{k \neq i,j}^{N} \frac{1}{r_{ik} r_{jk} / \log(e + r_{ij})} \right)^{-1} \left\langle \sum_{k \neq i,j}^{N} \hat{\mathbf{x}}_{ik} \cdot \hat{\mathbf{x}}_{jk} \right\rangle \rightarrow \left(\sum_{k \neq i}^{N} \frac{1}{r_{ik}^2} \right)^{-1} \tag{12}$$

Using the approximation,

$$\left(\sum_{k \neq i}^{N} \frac{1}{r_{ik}^2} \right)^{-1} \sim R_i^2$$

where R_i is the distance of the i^{th} Cα atom from the protein centroid. We obtain from Eq. 12 $W_{ii} = R_i^2$, which is the PFP B-factor[15]. But, if protein has multi-domains, the B-factor prediction results between the two models will have obvious difference because the protein may have more than one centroid[16] when the PFP model assumes a protein has only one centroid.

REFERENCES

1. Levitt M, Warshel A. Computer simulation of protein folding. Nature 1975;253(5494):694-698.

2. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267(5612):585-590.

3. Rueda M, Ferrer-Costa C, Meyer T, Perez A, Camps J, Hospital A, Gelpi JL, Orozco M. A consensus view of protein dynamics. Proc Natl Acad Sci U S A 2007;104(3):796-801.

4. Warshel A. Bicycle-pedal model for the first step in the vision process. Nature 1976;260(5553):679-683.

5. Warshel A. Molecular dynamics simulations of biological reactions. Acc Chem Res 2002;35(6):385-395.

6. Srinivasan R, Rose GD. LINUS: a hierarchic procedure to predict the fold of a protein. Proteins 1995;22(2):81-99.

7. Yue K, Dill KA. Folding proteins with a simple energy function and extensive conformational searching. Protein Sci 1996;5(2):254-261.

8. Tirion MM. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. Phys Rev Lett 1996;77(9):1905-1908.

9. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2(3):173-181.

10. Ming D, Kong Y, Lambert MA, Huang Z, Ma J. How to describe protein motion without amino acid sequence and atomic coordinates. Proc Natl Acad Sci U S A 2002;99(13):8620-8625.

11. Micheletti C, Banavar JR, Maritan A. Conformations of proteins in equilibrium. Phys Rev Lett 2001;87(8):088102.

12. Canino LS, Shen T, McCammon JA. Changes in flexibility upon binding: Application of the self-consistent pair contact probability method to protein-protein interactions. J Chem Phys 2002;117(21):9927-9933.

13. Pandey BP, Zhang C, Yuan X, Zi J, Zhou Y. Protein flexibility prediction by an all-atom mean-field statistical theory. Protein Sci 2005;14(7):1772-1777.

14. Halle B. Flexibility and packing in proteins. Proc Natl Acad Sci U S A 2002;99(3):1274-1279.

15. Shih CH, Huang SW, Yen SC, Lai YL, Yu SH, Hwang JK. A simple way to compute protein dynamics without a mechanical model. Proteins 2007;68(1):34-38.

16. Lu CH, Huang SW, Lai YL, Lin CP, Shih CH, Huang CC, Hsu WL, Hwang JK. On the relationship between the protein structure and protein dynamics. Proteins 2008;72(2):625-634.

17. Remington S, Wiegand G, Huber R. Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 A resolution. J Mol Biol 1982;158(1):111-152.

18. Wiegand G, Remington SJ. Citrate synthase: structure, control, and mechanism. Annu Rev Biophys Biophys Chem 1986;15:97-117.

19. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A 1983;80(21):6571-6575.

20. Kidera A, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. J Mol Biol 1992;225(2):457-475.

21. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 1985;181(3):423-447.

22. Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci U S A 1983;80(12):3696-3700.

23. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 1991;11(3):205-217.

24. Noguchi T, Akiyama Y. PDB-REPRDB: a database of representative protein chains from the Protein Data Bank (PDB) in 2003. Nucleic Acids Res 2003;31(1):492-493.

25. Yang LW, Rader AJ, Liu X, Jursa CJ, Chen SC, Karimi HA, Bahar I. oGNM: online computation of structural dynamics using the Gaussian Network Model. Nucleic Acids Res 2006;34(Web Server issue):W24-31.

26. Team R. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing;2007.

27. Taylor TC, Backlund A, Bjorhall K, Spreitzer RJ, Andersson I. First crystal structure of Rubisco from a green alga, Chlamydomonas reinhardtii. J Biol Chem 2001;276(51):48159-48164.

28. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci 1998;23(9):358-361.

29. Xu Q, Canutescu A, Obradovic Z, Dunbrack RL, Jr. ProtBuD: a database of biological unit structures of protein families and superfamilies. Bioinformatics 2006;22(23):2876-2882.

30. Budiman ME, Knaggs MH, Fetrow JS, Alexander RW. Using molecular dynamics to map interaction networks in an aminoacyl-tRNA synthetase. Proteins 2007;68(3):670-689.

31. Zheng W, Liao JC, Brooks BR, Doniach S. Toward the mechanism of dynamical couplings and translocation in hepatitis C virus NS3 helicase using elastic network model. Proteins 2007;67(4):886-896.

32. Ming D, Kong Y, Wakil SJ, Brink J, Ma J. Domain movements in human fatty acid synthase by quantized elastic deformational model. Proc Natl Acad Sci U S A 2002;99(12):7895-7899.

33. Yang LW, Bahar I. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure 2005;13(6):893-904.

34. Eichinger A, Beisel HG, Jacob U, Huber R, Medrano FJ, Banbula A, Potempa J, Travis J, Bode W. Crystal structure of gingipain R: an Arg-specific bacterial cysteine proteinase with a caspase-like fold. Embo J 1999;18(20):5453-5462.

35. Tsunasawa S, Masaki T, Hirose M, Soejima M, Sakiyama F. The primary structure and structural characteristics of Achromobacter lyticus protease I, a lysine-specific serine protease. J Biol Chem 1989;264(7):3832-3839.

36. Lee FS, Chu ZT, Warshel A. Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the POLARIS and ENZYMIX Programs. J Comp Chem 1993;14(2):161-185.

37. Fan Z-Z, Hwang J-K, Warshel A. Using simplified protein representaion as a reference potential for all-atom calculations of folding free energy. Theor Chem Acc 1999;103:77-80.

FIGURES AND TABLES

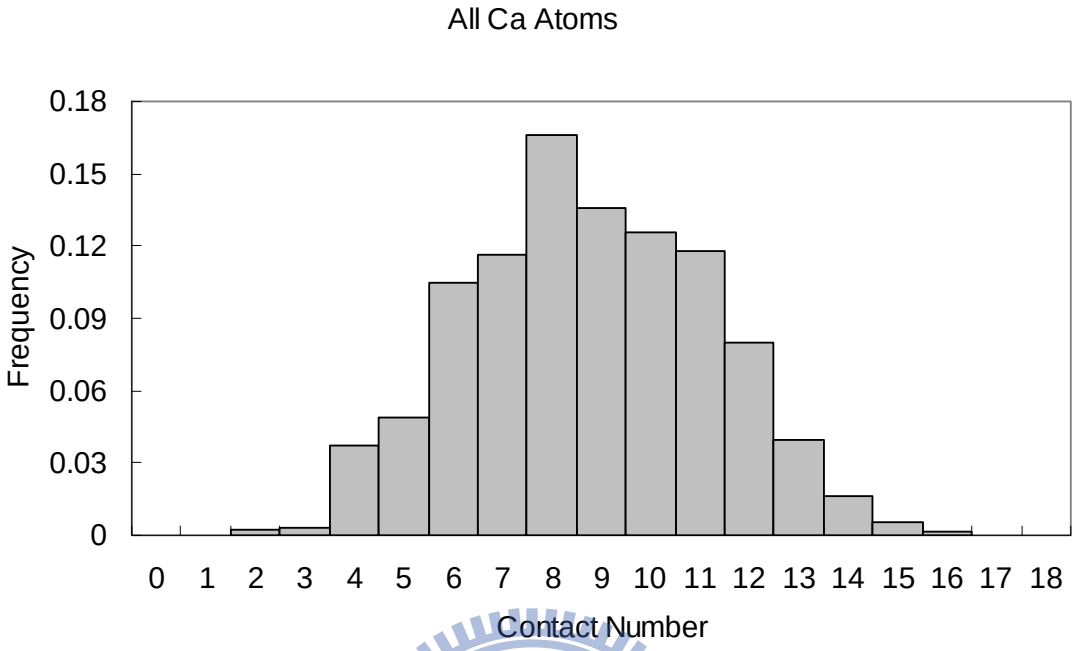


Figure 1. The distribution of the frequencies of residue contact numbers in proteins for the nonhomologous dataset which selected from PDB-REPRDB[24].

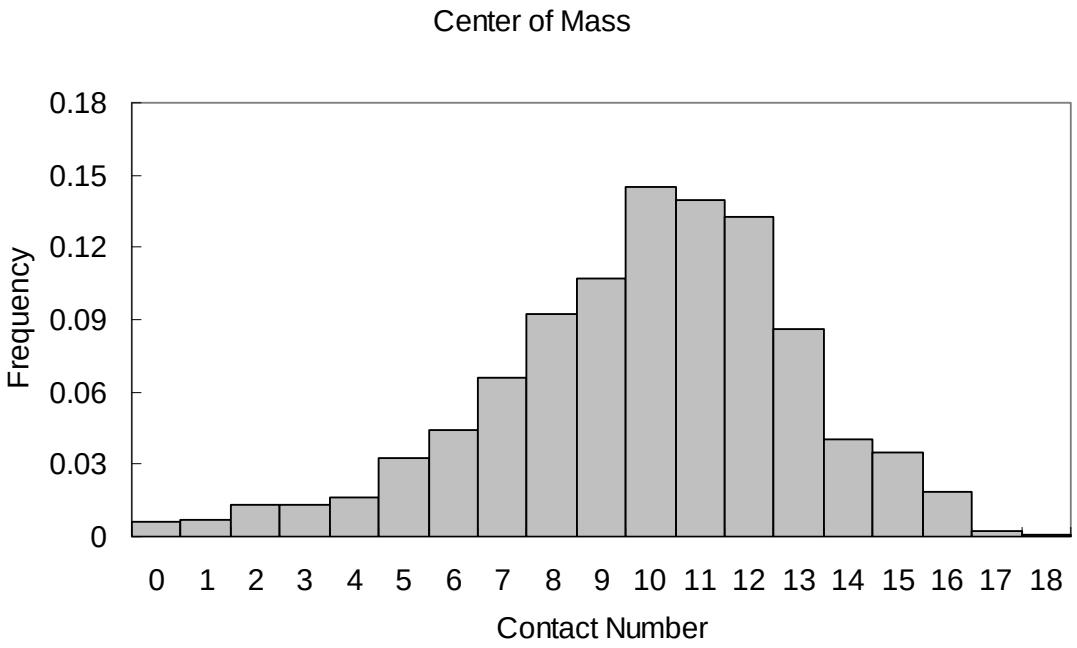


Figure 2. The distribution of the frequencies of centers of mass contact numbers in proteins for the nonhomologous dataset which selected from PDB-REPRDB[24].

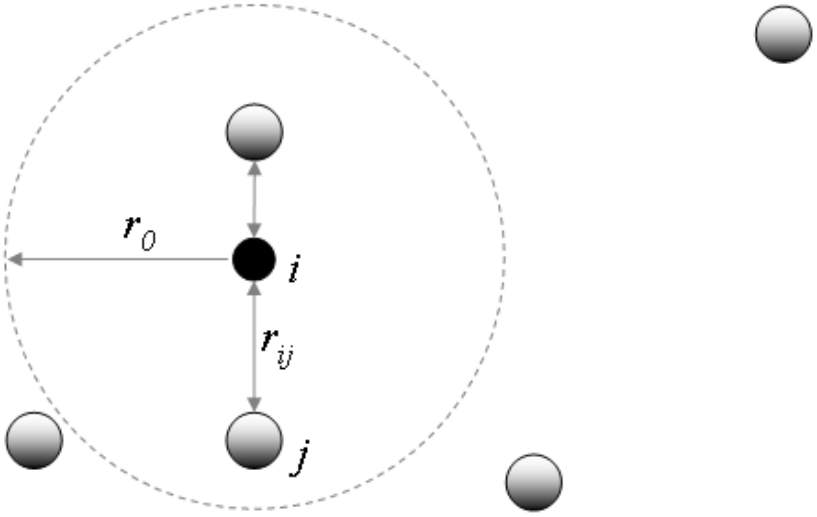


Figure 3. The schematic illustration of the CN model of a hypothetical 6-atom protein. We must define a cut-off distance r_0 first as the radius of contact sphere region of atom i. When the contact distance r_{ij} is smaller than r_0, we will think the atom j contacts atom i. In addition, the atoms in the contact sphere region contribute equally no matter how long the distance between atom i and atom j (r_{ij}) is.

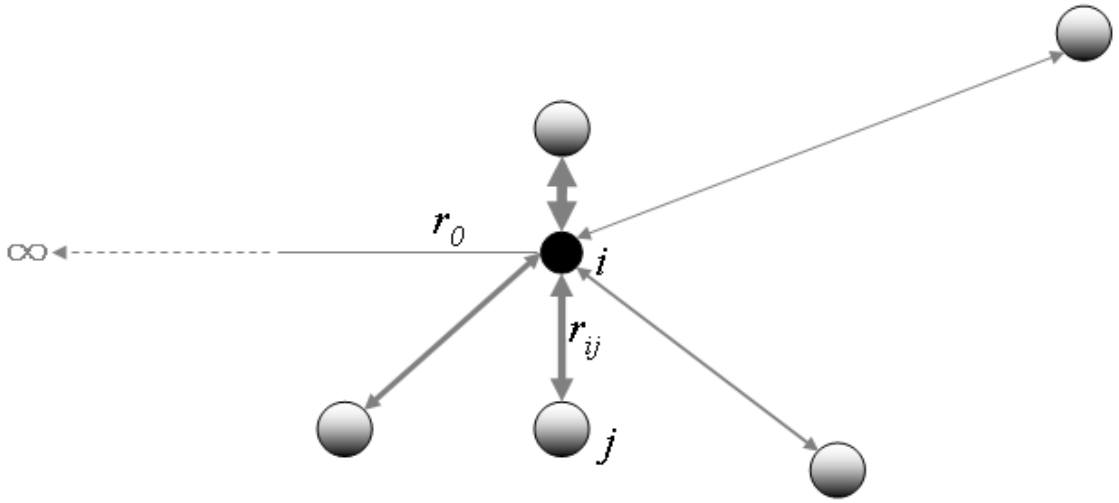


Figure 4. The schematic illustration of the WCN model of a hypothetical 6-atom protein. We don't need to define a cut-off distance r_0 (i.e, to regard r_0 as infinity) as the radius of contact sphere region of atom *i* but calculate weighted contact number of all atoms based on the results of cut-off distances effect screening. We think every atom will contact atom *i* but the contact effect is scaled down by the square of the distance between the contacting pair. The thickness of the arrow means the contact effect, and the thicker is the stronger.

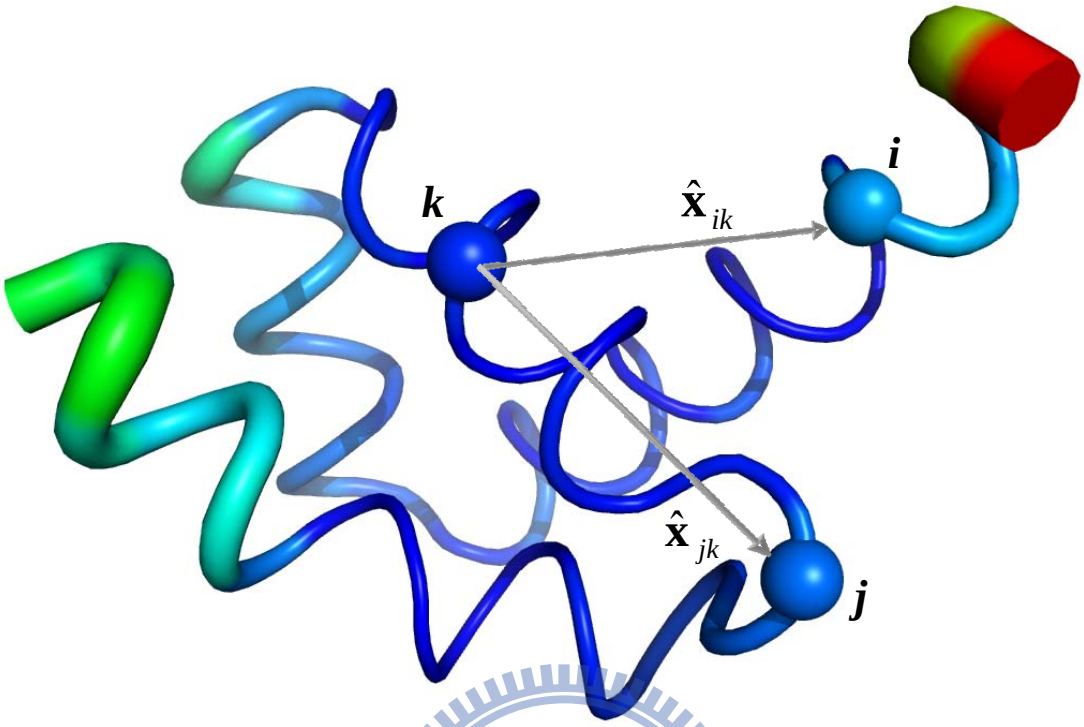


Figure 5. The structure of the human hyperplastic discs protein (1I2T:A) is presented in the cartoon putty representation, where the color is ramped by residue from blue at the lowest B-factor value to red at the highest B-factor value. This figure illustrates the cross-correlation between residues by WCN model. The balls represent the residues *i*, *j* and *k*. $\hat{\mathbf{x}}_{ik}$ and $\hat{\mathbf{x}}_{jk}$ are the unit vectors in the direction of $\mathbf{r}_i - \mathbf{r}_k$ and $\mathbf{r}_j - \mathbf{r}_k$, respectively.

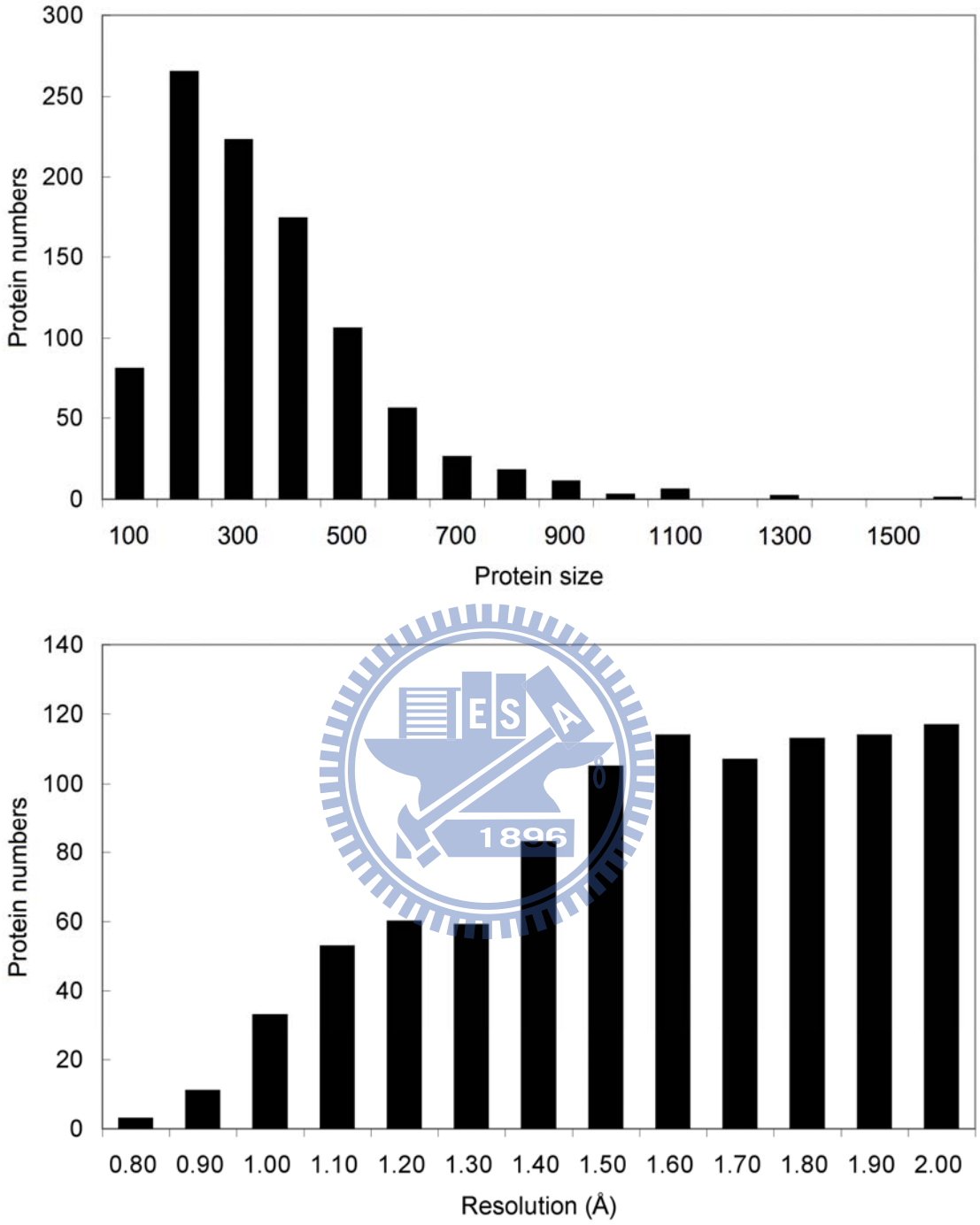


Figure 6. The distribution of (A) protein size and (B) structural resolution of the nonhomologous data set comprising 972 protein structures with resolution ≤ 2.0 Å and R-factors ≤0.2 selected from PDB-REPRDB.

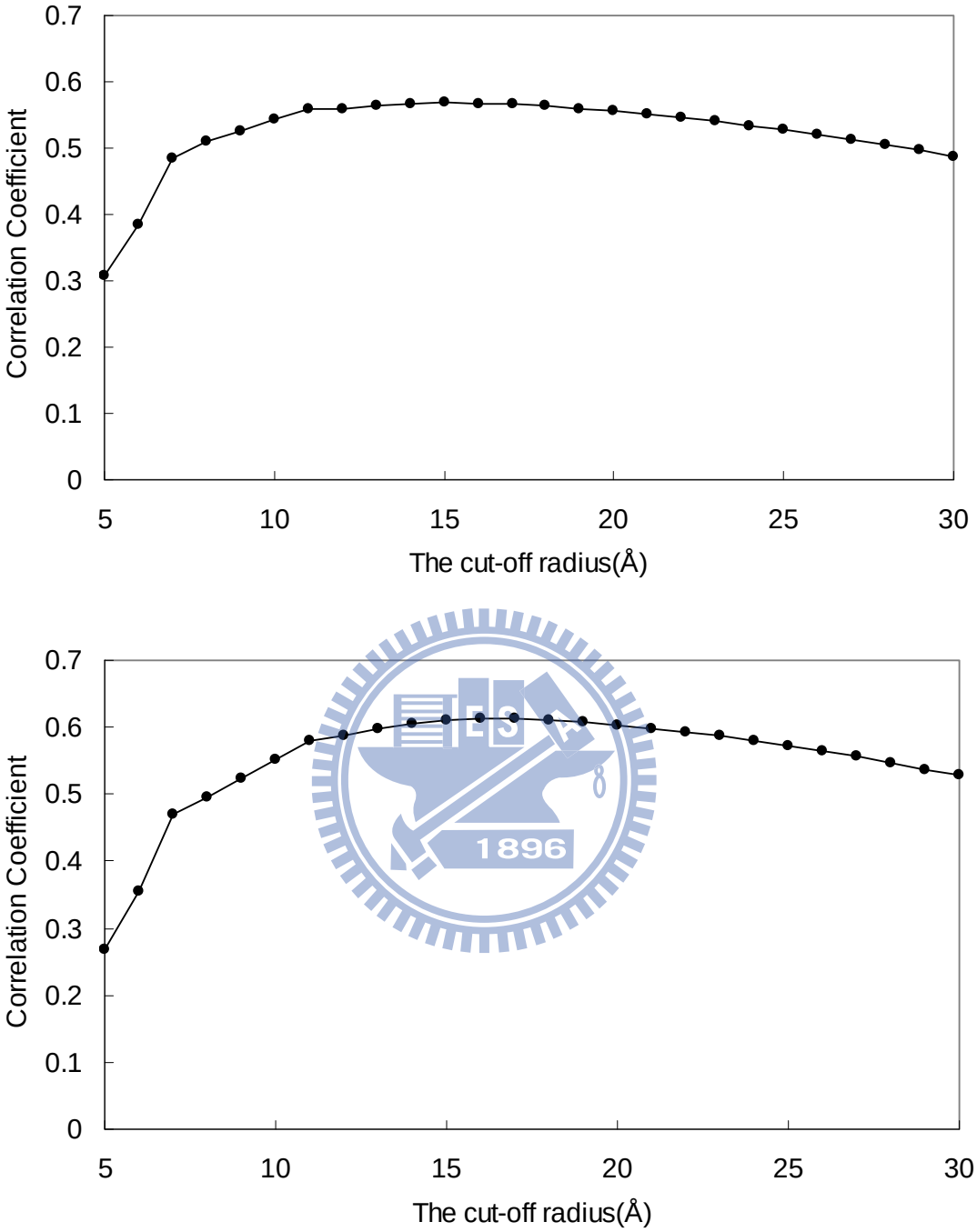


Figure 7. The performances of the correlation coefficients are in different cut-off radius values in CN model for the nonhomologous data set. The upper part is Pearson's linear correlation coefficients and the lower part is Spearman's rank correlation coefficients.

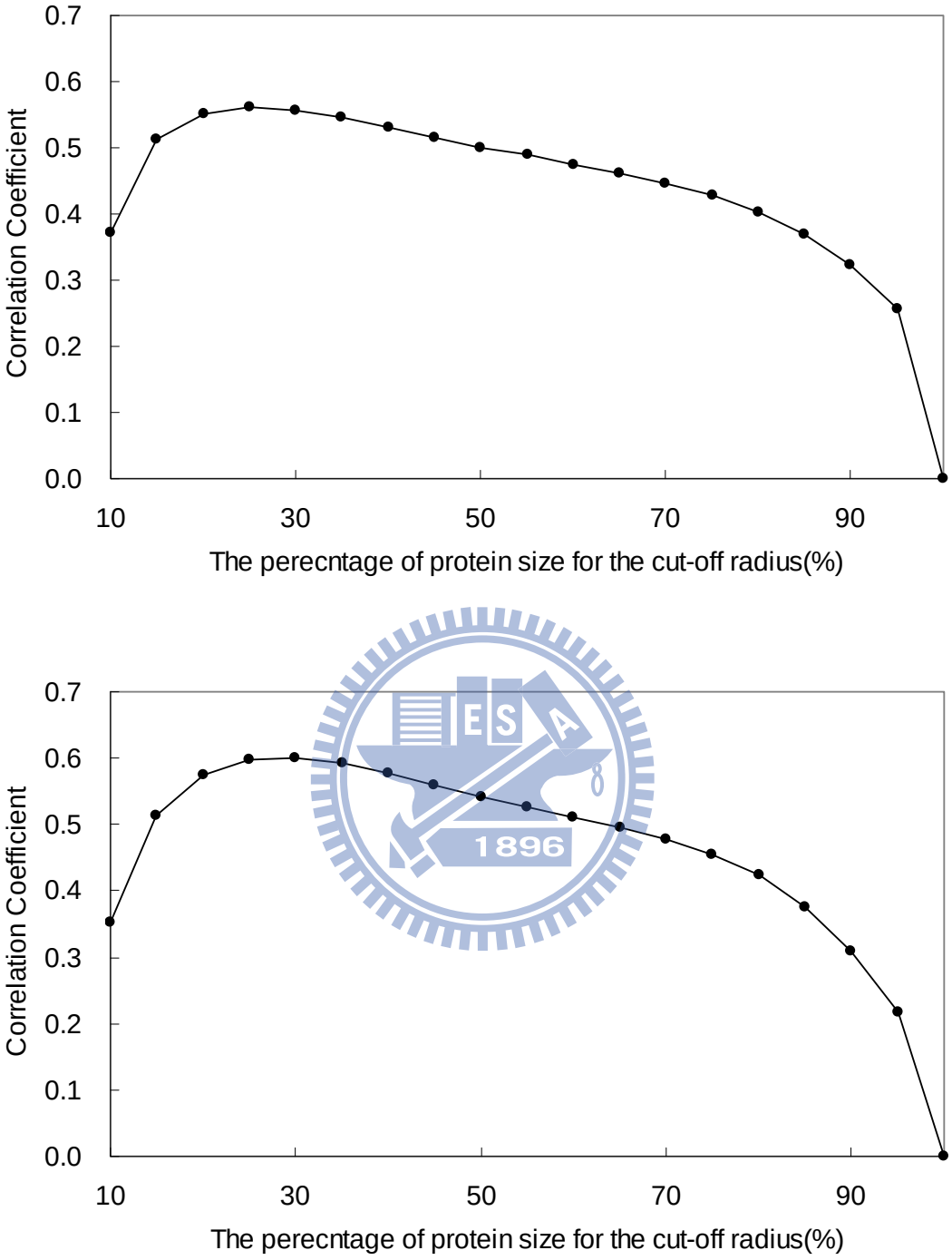


Figure 8. The performances of the correlation coefficients are in different percentage values of the protein size for the cut-off radius in CN model for the nonhomologous data set. The upper part is Pearson's linear correlation coefficients and the lower part is Spearman's rank correlation coefficients.

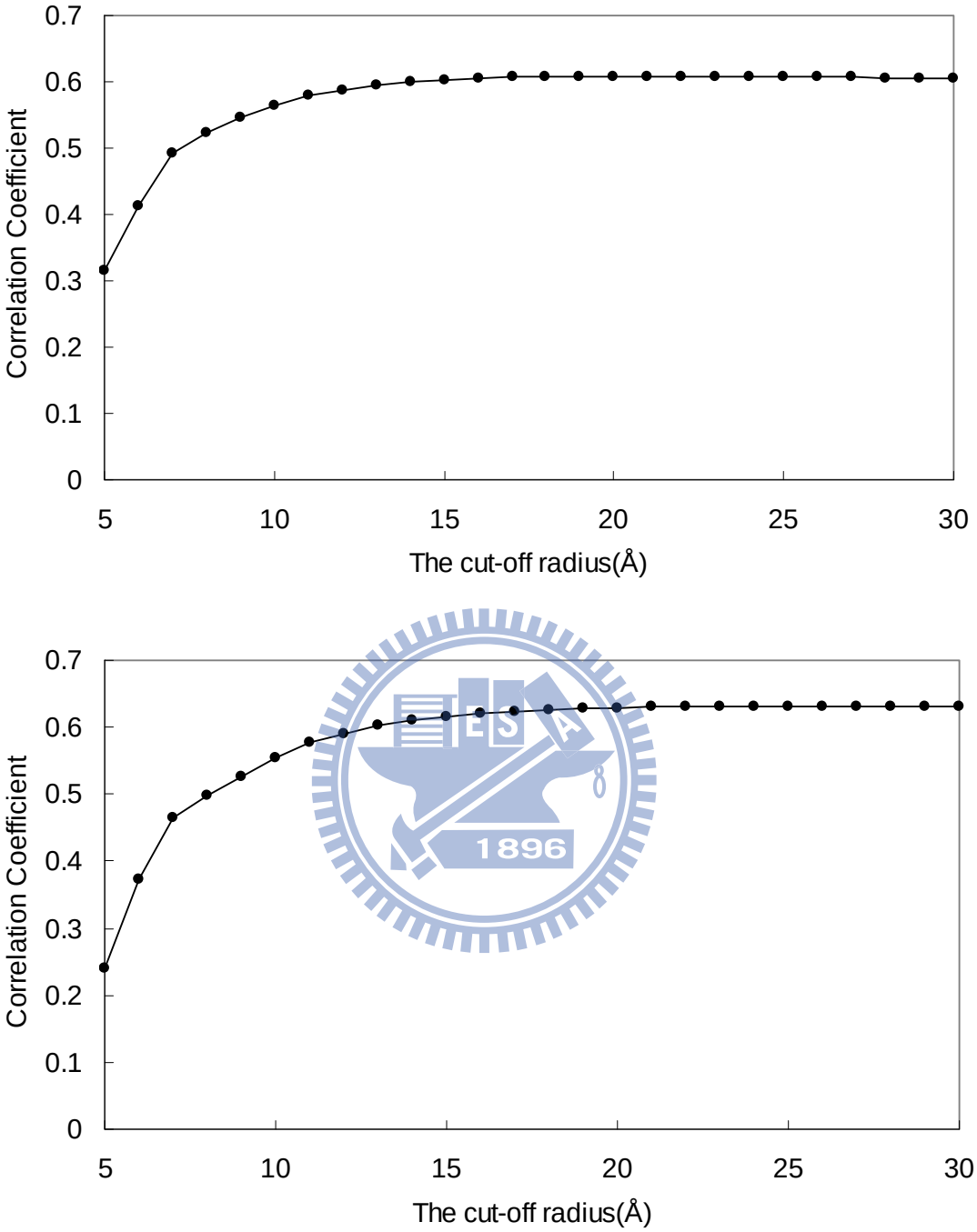


Figure 9. The performances of the correlation coefficients are in different cut-off radius values in WCN model for the nonhomologous data set. The upper part is Pearson's linear correlation coefficients and the lower part is Spearman's rank correlation coefficients.

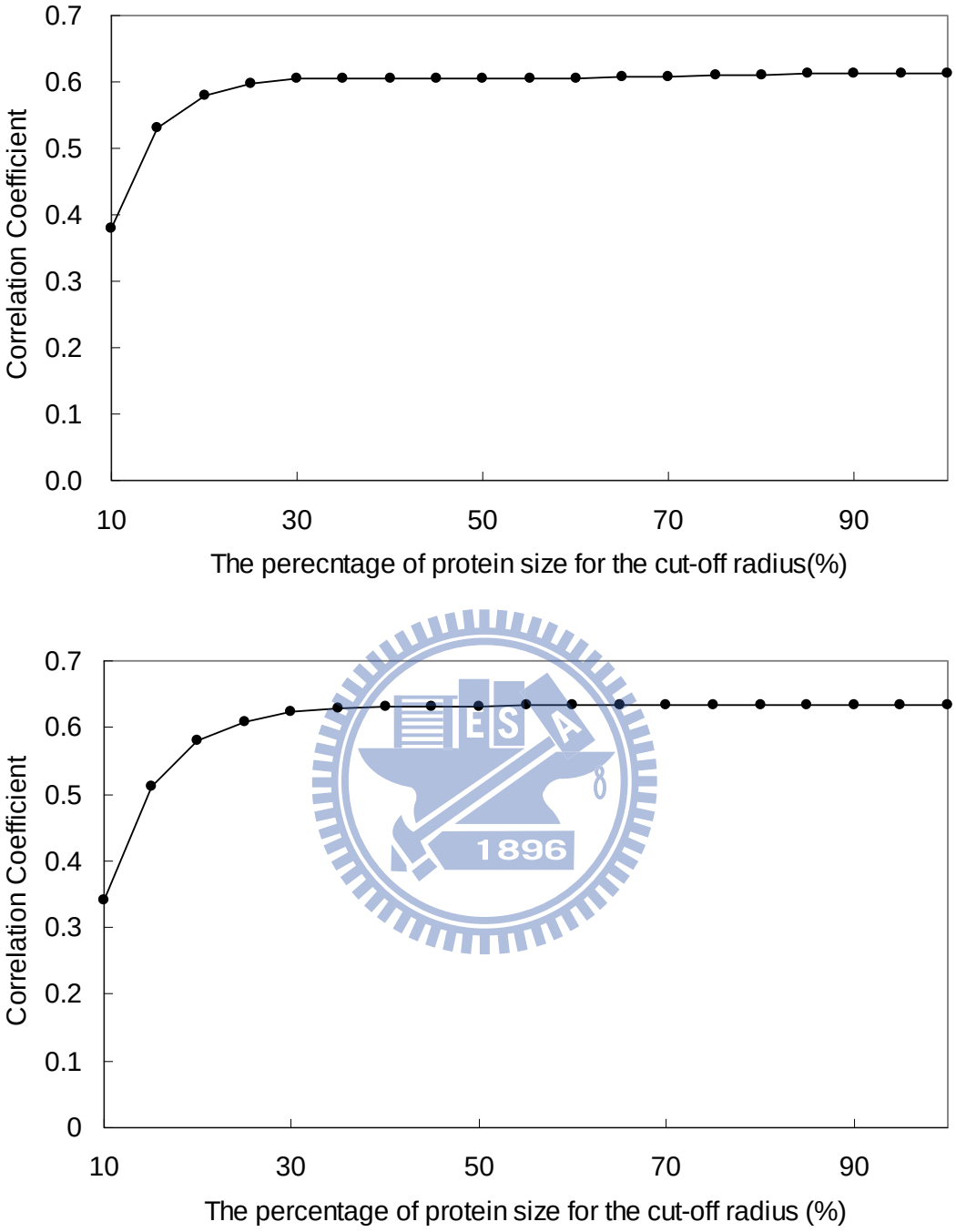


Figure 10. The performances of the correlation coefficients are in different percentage values of the protein size for the cut-off radius in WCN model for the nonhomologous data set. The upper part is Pearson's linear correlation coefficients and the lower part is Spearman's rank correlation coefficients.

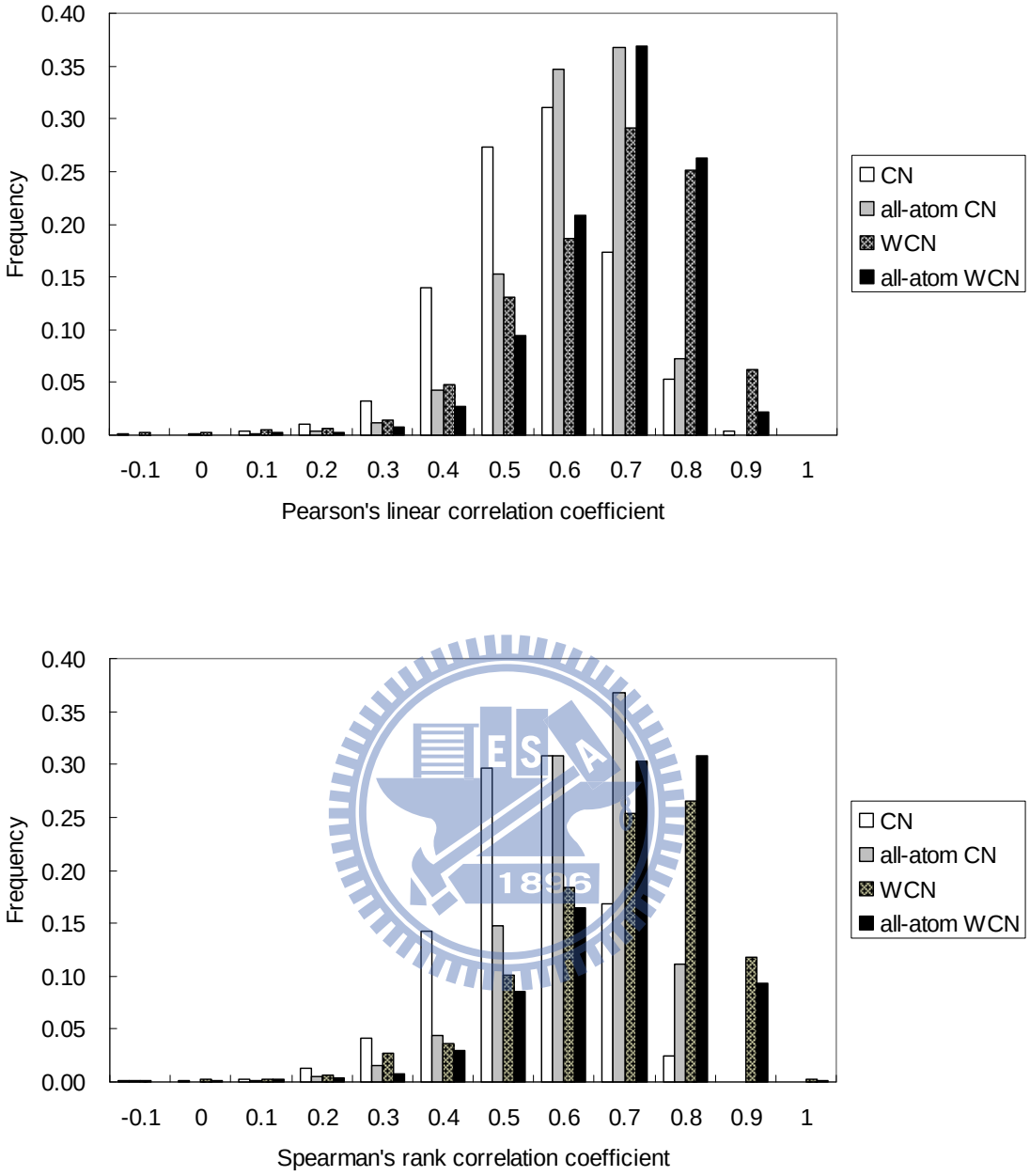


Figure 11. Comparison of the correlation coefficients between experimental and the computed B-factor profiles bases on the CN model (white), the all-atom CN model (grey), the WCN model (spot) and all-atom WCN model (black) for the nonhomologous data set.

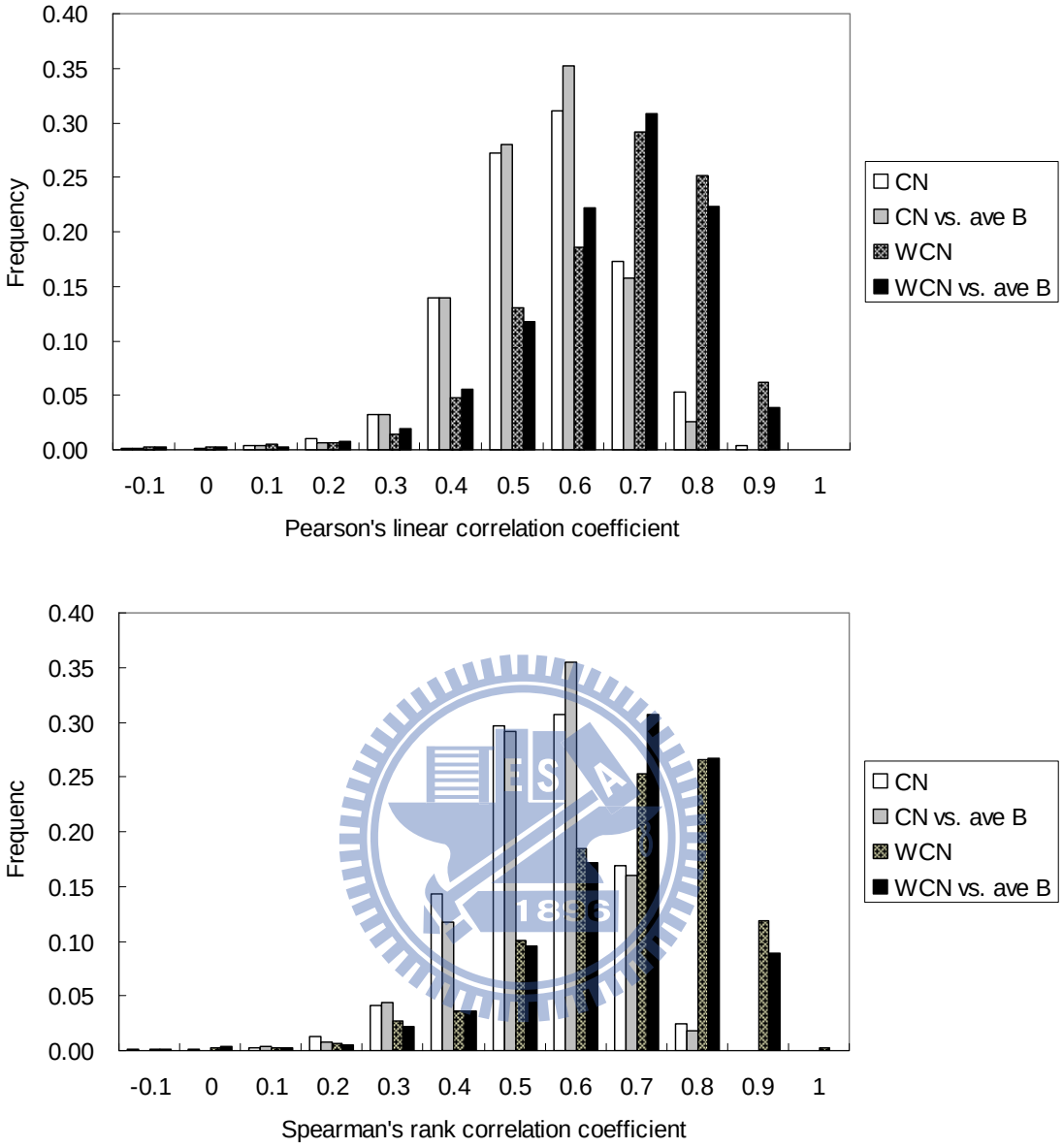


Figure 12. Comparison of the correlation coefficients between experimental and the computed B-factor profiles bases on the CN model (white), the CN model vs. average B-factor (grey), the WCN model (spot) and WCN model vs. average B-factor (black) for the nonhomologous data set.

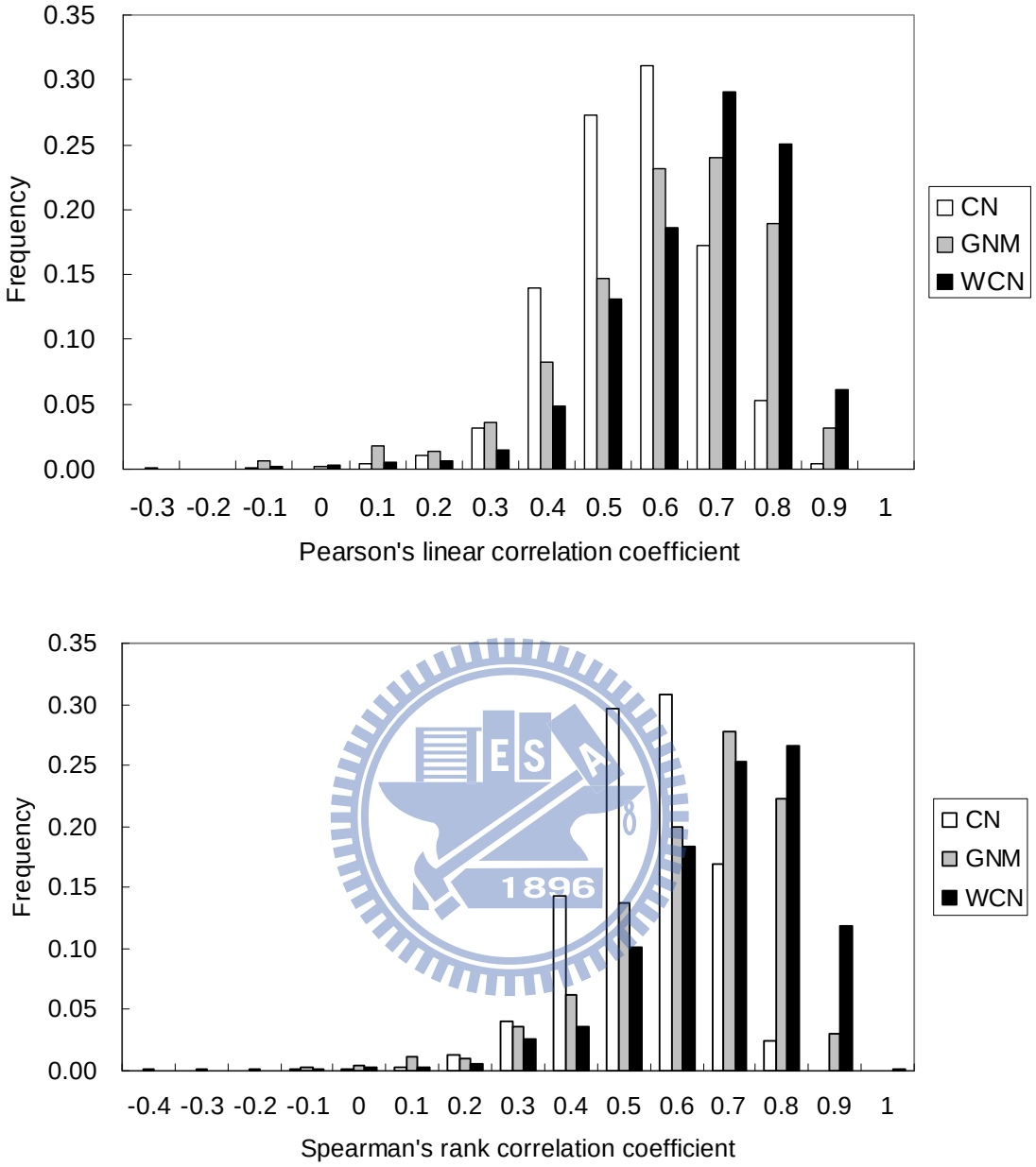


Figure 13. Comparison of the correlation coefficients between experimental and the computed B-factor profiles bases on the CN model (white), the GNM (grey) and the WCN model (black) for the nonhomologous data set.

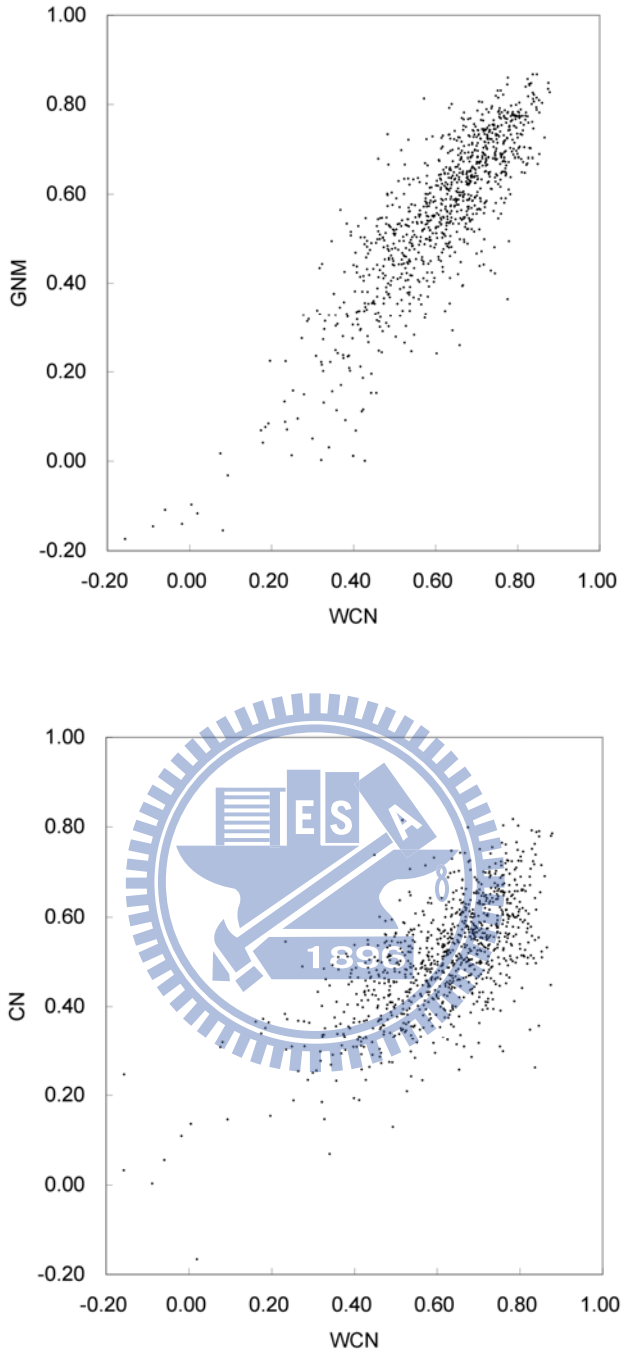


Figure 14. The correlation of model predictions (A) between the WCN model and the GNM, and (B) between the WCN model and the CN model for the nonhomologous data set.

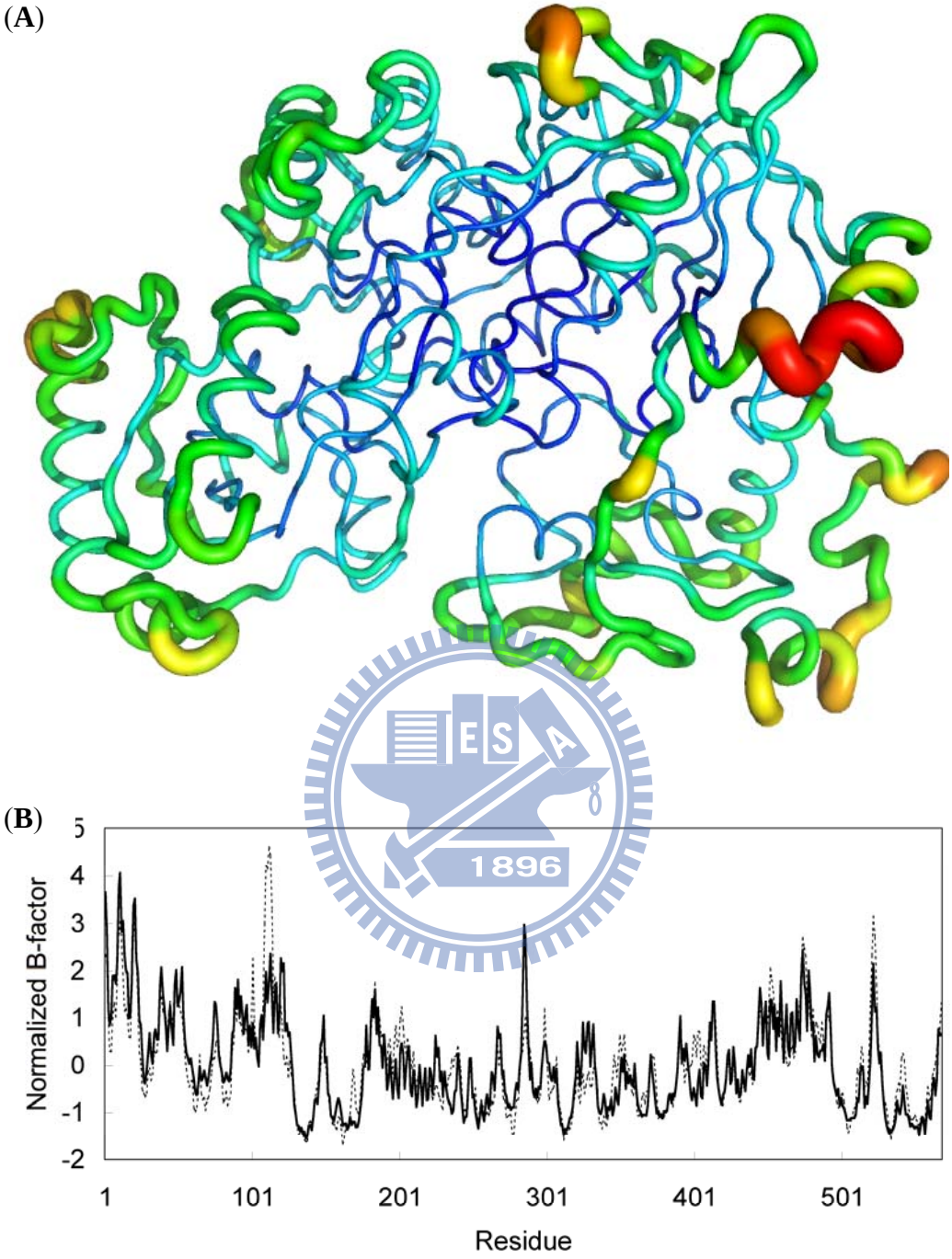
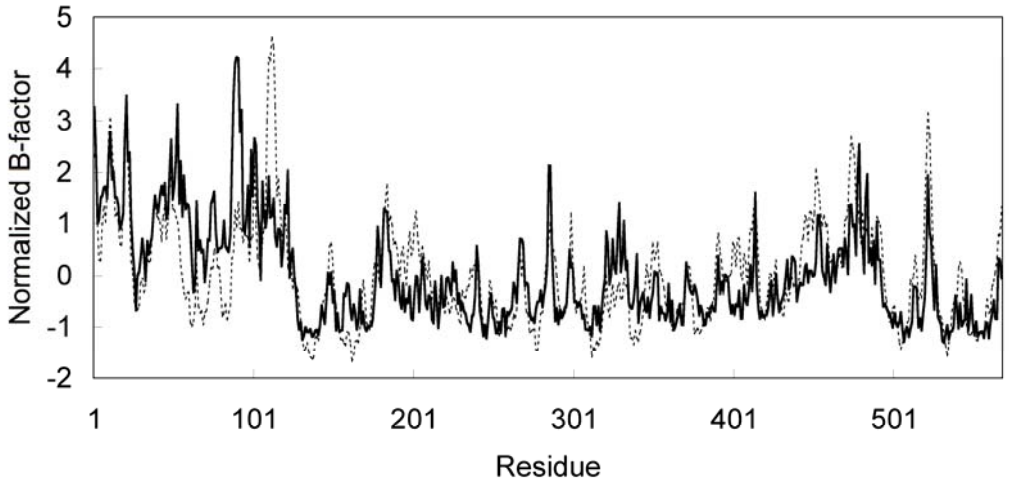


Figure 15. (A) The structure of the flavocytochrome c3 (1Y0P:A) which is presented in the cartoon putty representation. (**B**) The X-ray B-factor profile (dotted line) of compared with the computed B-factor profile (solid line) by the WCN model ($c = 0.85$, $\rho = 0.87$)

(**C**)



(**D**)

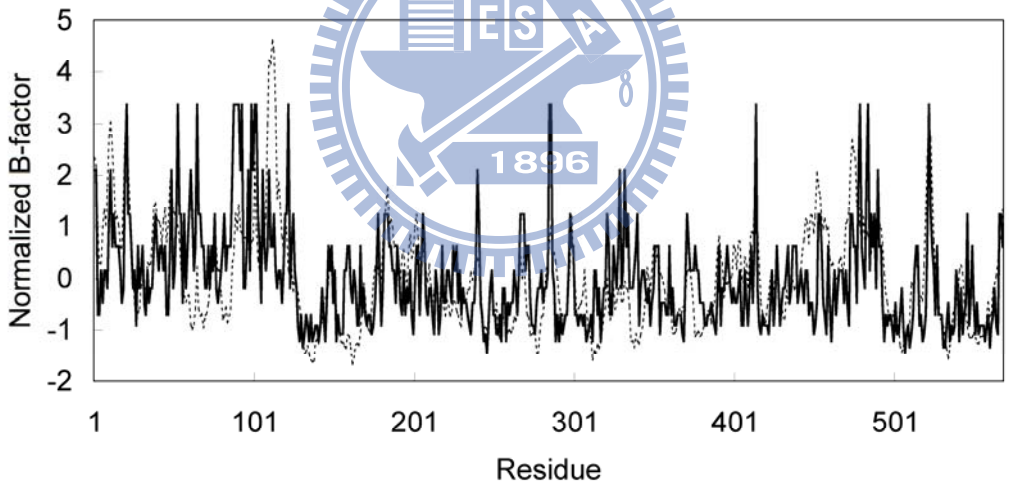
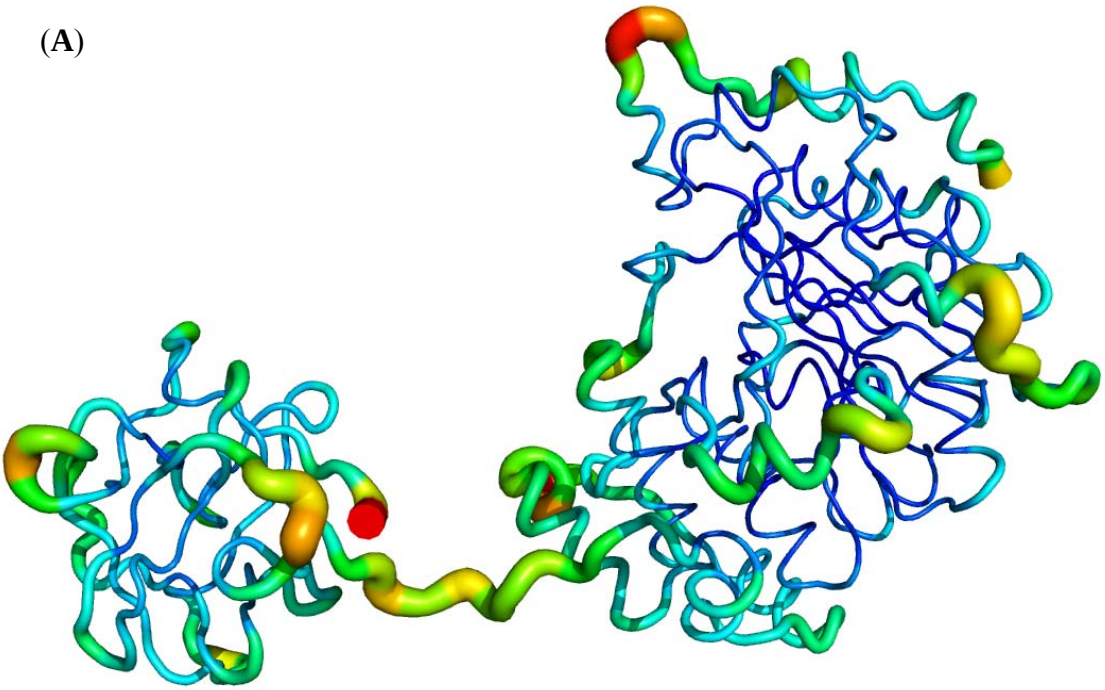


Figure 15. The X-ray B-factor profile (dotted line) of flavocytochrome c3 (1Y0P:A) compared with the computed B-factor profile (solid line) by (**C**) the GNM ($c = 0.69$, $\rho = 0.72$) and (**D**) the CN model ($c = 0.51$, $\rho = 0.51$).

(**A**)



(**B**)

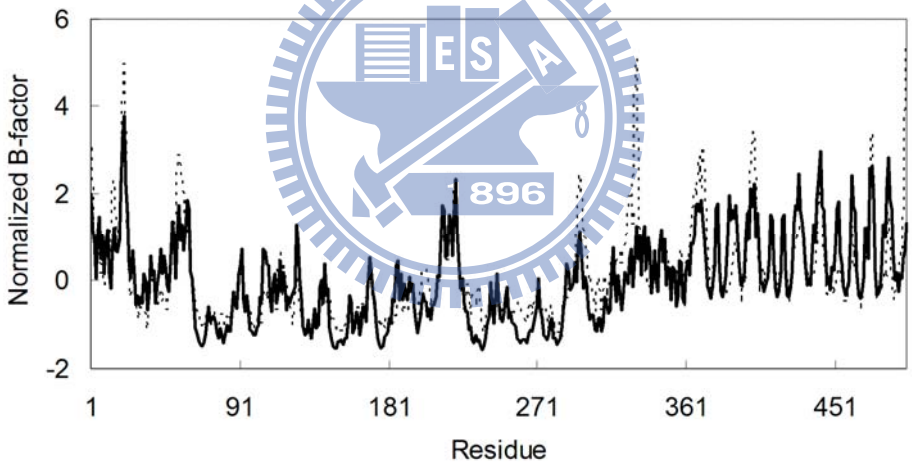
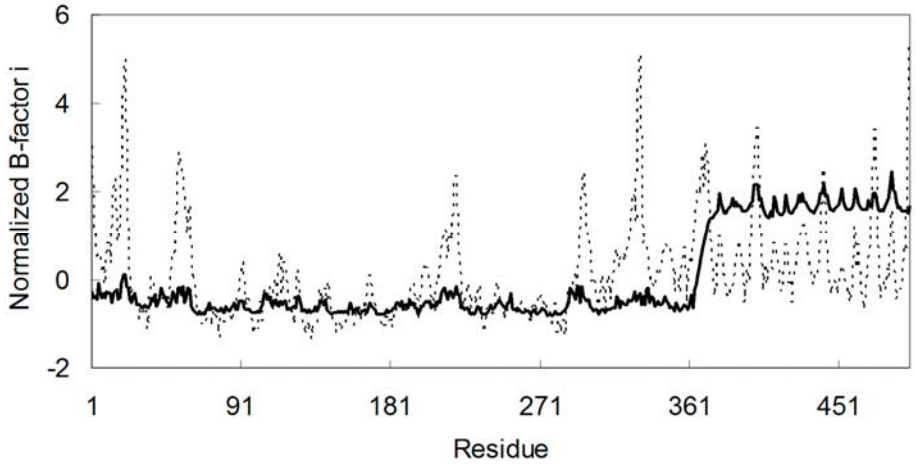


Figure 16. (A) The structure of the chain A of human ppGalNAcT-2 (2FFU:A) which is presented in the cartoon putty representation.. (**B**) The X-ray B-factor profile (dotted line) of compared with the computed B-factor profile (solid line) by the WCN model ($c = 0.78$, $\rho = 0.84$).

(**C**)



(**D**)

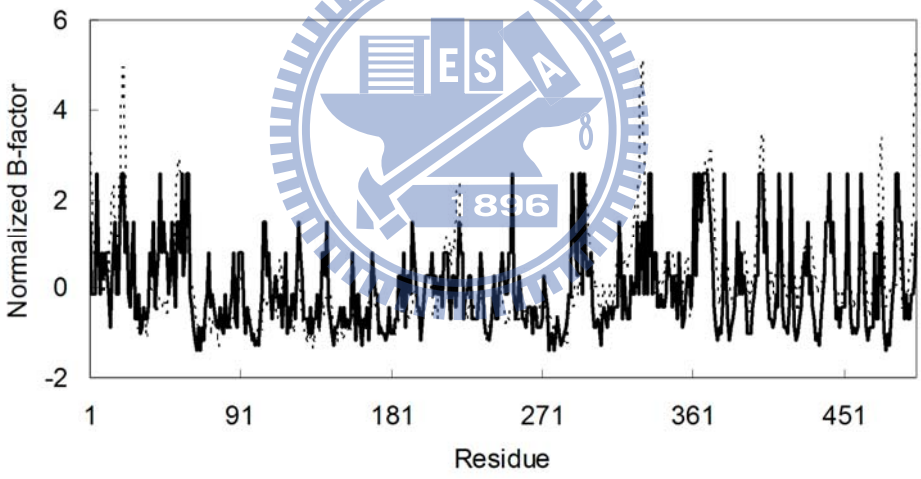


Figure 16. The X-ray B-factor profile (dotted line) of the chain A of human ppGalNAcT-2 (2FFU:A) compared with the computed B-factor profile (solid line) by (**C**) the GNM ($c = 0.36, \rho = 0.71$) and (**D**) the CN model ($c = 0.60, \rho = 0.60$) which are shown in upper and lower part, respectively.

Table 1. The performance breakdown of the WCN model, the CN model and the GNM by Pearson's linear correlation coefficient for the structures classified according to the SCOP classes

	WCN		CN		GNM	
SCOP classes	$\overline{c}$	$c_{0.5}$	$\overline{c}$	$c_{0.5}$	$\overline{c}$	$c_{0.5}$
All-α proteins	0.59	73%	0.47	43%	0.54	68%
All-β proteins	0.64	82%	0.51	58%	0.58	73%
α/β proteins	0.62	82%	0.49	51%	0.57	75%
α+β proteins	0.60	77%	0.49	51%	0.54	65%

Table 2. The performance breakdown of the WCN model, the CN model and the GNM by Spearman's rank correlation coefficient for the structures classified according to the SCOP classes

	WCN		CN		GNM	
SCOP classes	$\overline{\rho}$	$\rho_{0.5}$	$\overline{\rho}$	$\rho_{0.5}$	$\overline{\rho}$	$\rho_{0.5}$
All-α proteins	0.60	76%	0.45	30%	0.53	63%
All-β proteins	0.66	84%	0.50	54%	0.59	78%
α/β proteins	0.65	86%	0.48	45%	0.60	76%
α+β proteins	0.63	80%	0.49	48%	0.58	72%

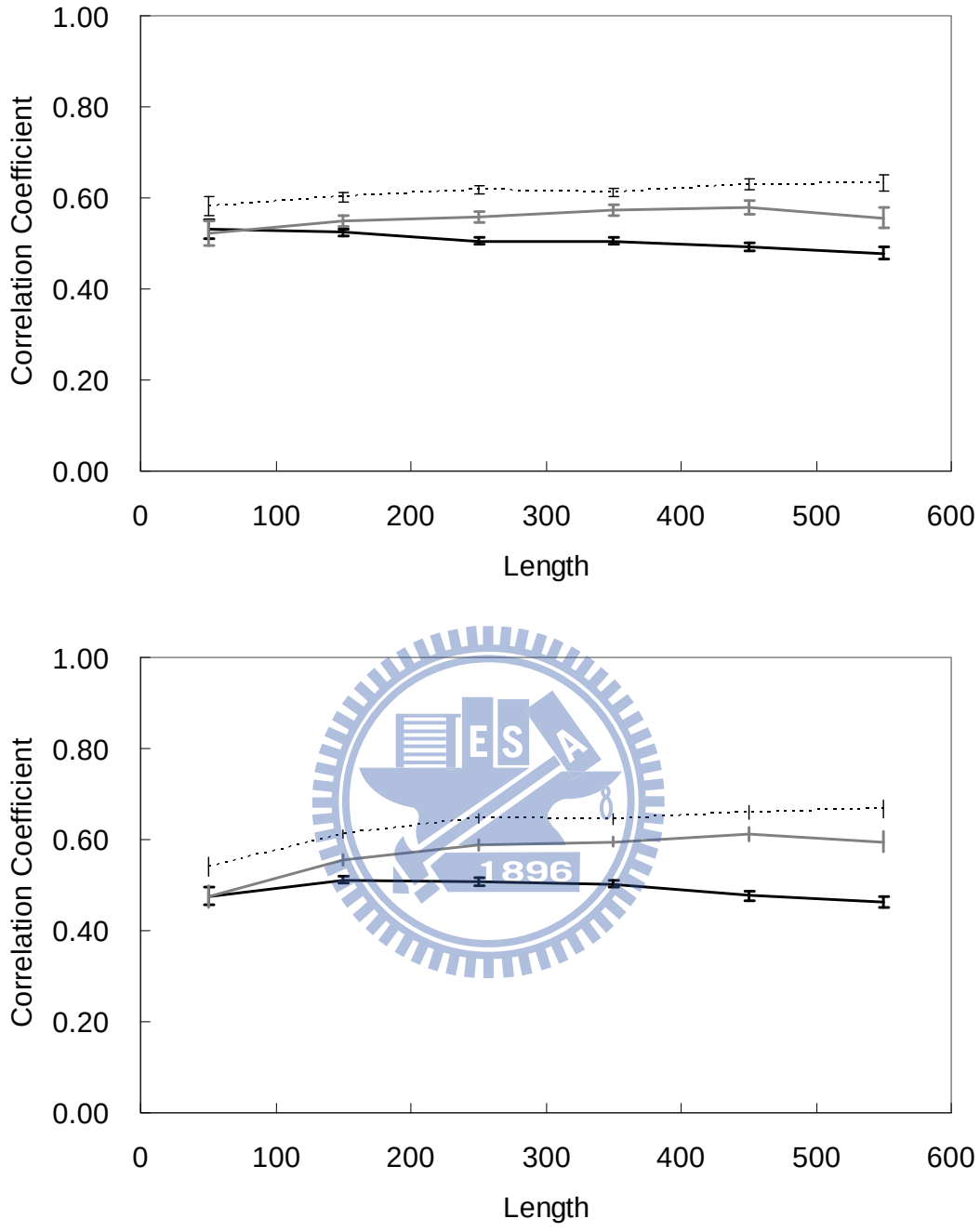


Figure 17. The correlation coefficient between the computed and the X-ray B-factors for the WCN model (dotted line), the CN model (solid line) and the GNM (grey line) as a function of the protein residue number. The upper part is Pearson's linear correlation coefficients and the lower part is Spearman's rank correlation coefficients.

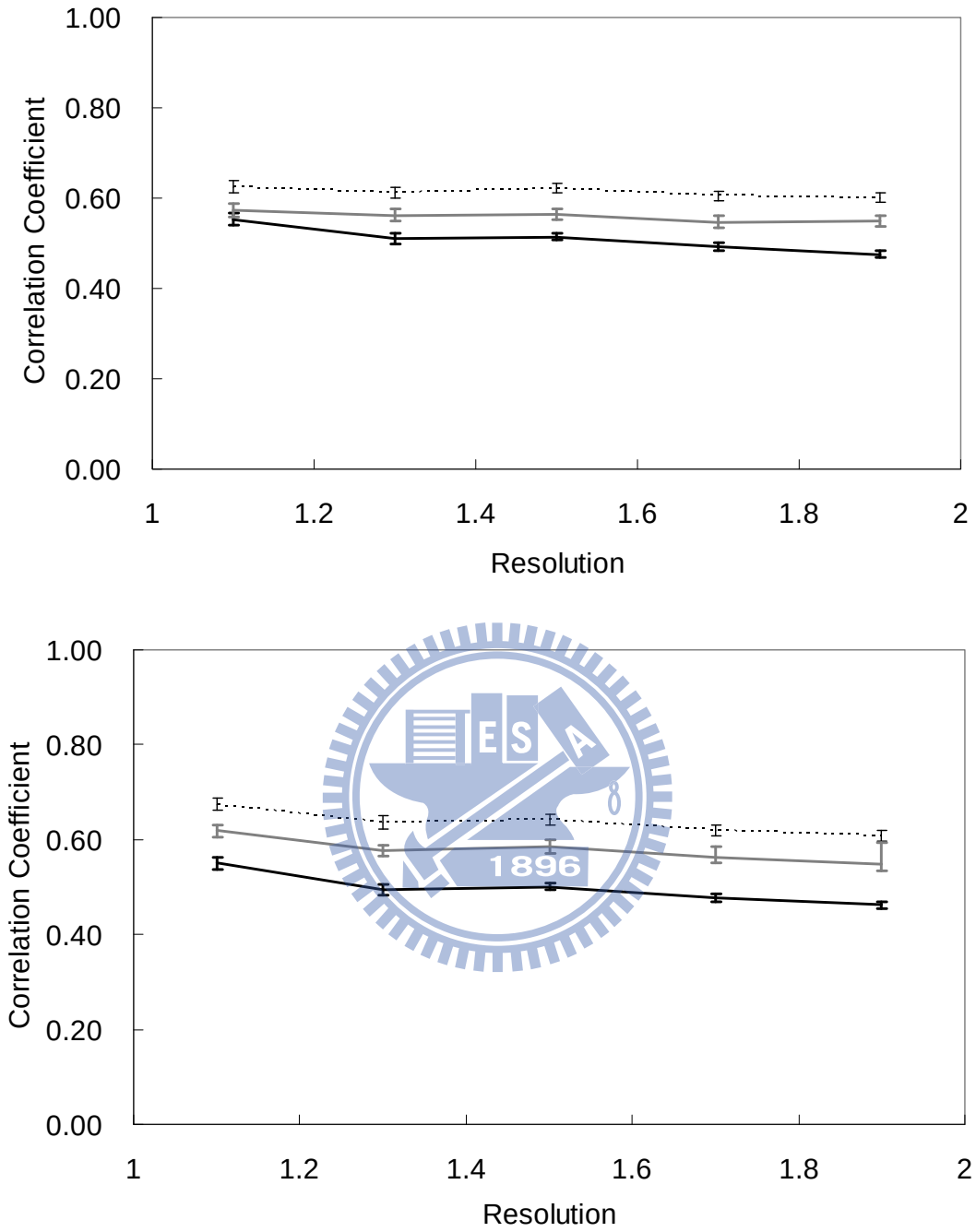


Figure 18. The correlation coefficient between the computed and the X-ray B-factors for the WCN model (dotted line), the CN model (solid line) and the GNM (grey line) as a function of the protein X-ray structure resolution in Å. The upper part is Pearson's linear correlation coefficients and the lower part is Spearman's rank correlation coefficients.

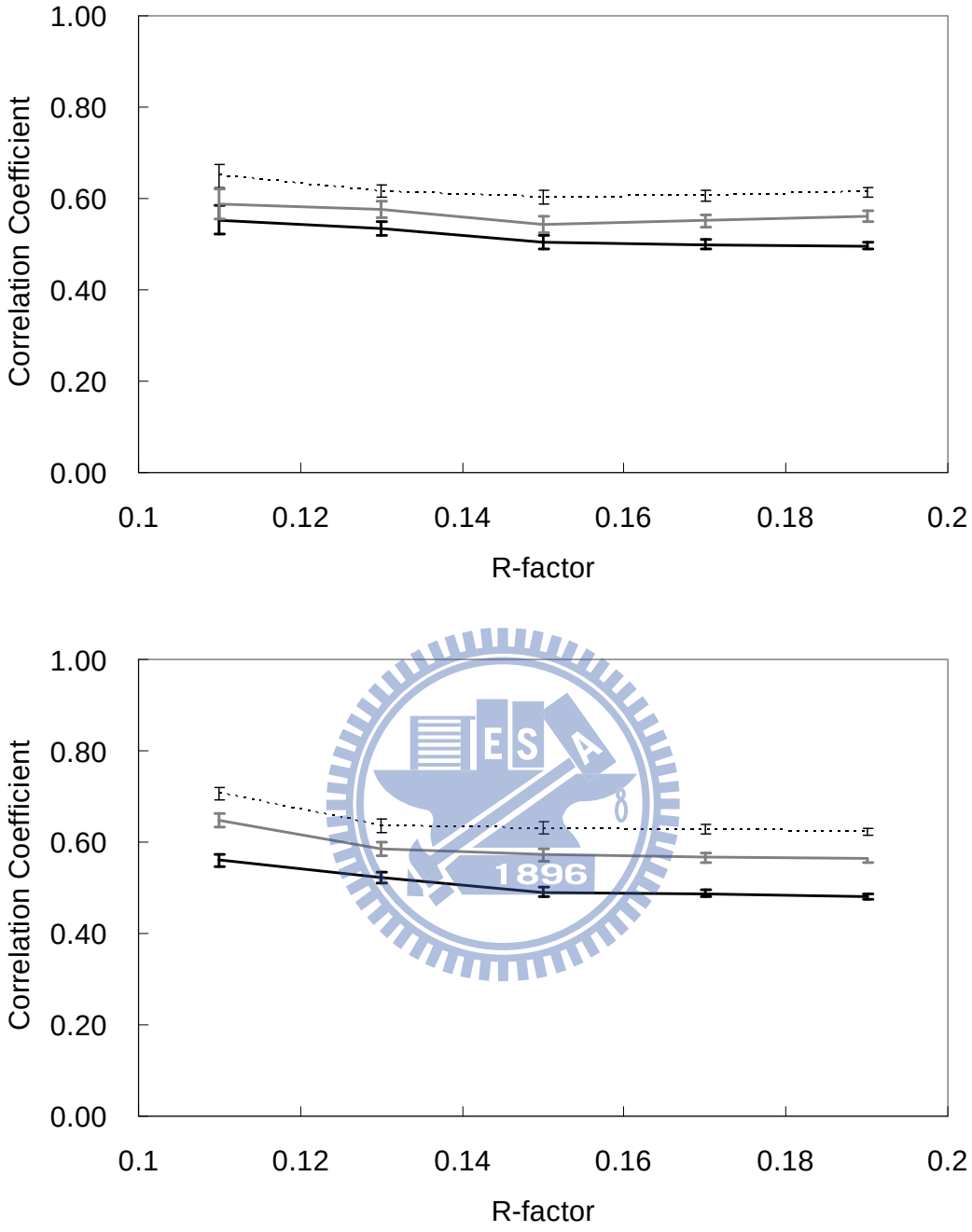
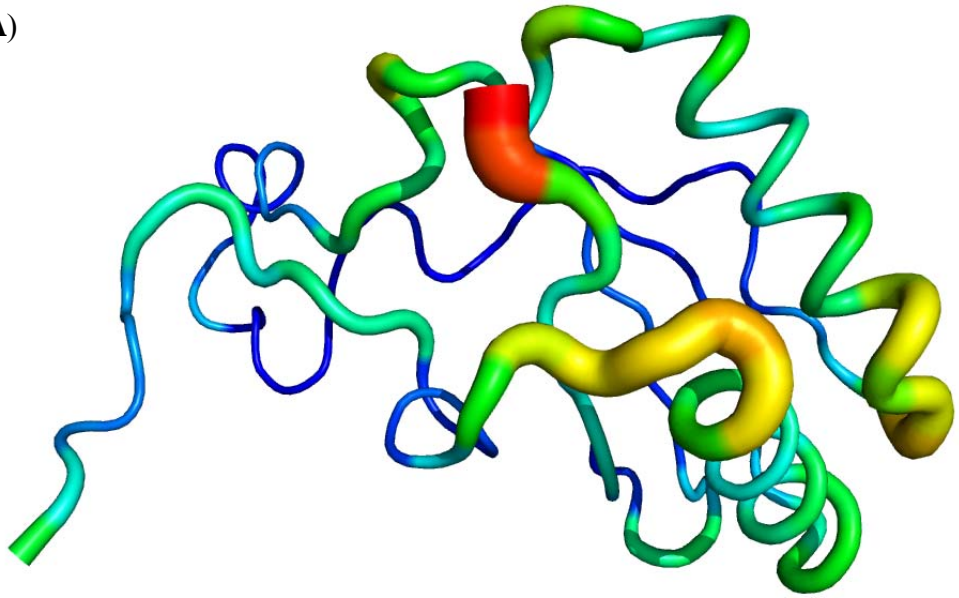


Figure 19. The correlation coefficient between the computed and the X-ray B-factors for the WCN model (dotted line), the CN model (solid line) and the GNM (grey line) as a function of the protein R-factor. The upper part is Pearson's linear correlation coefficients and the lower part is Spearman's rank correlation coefficients.

(**A**)



(**B**)

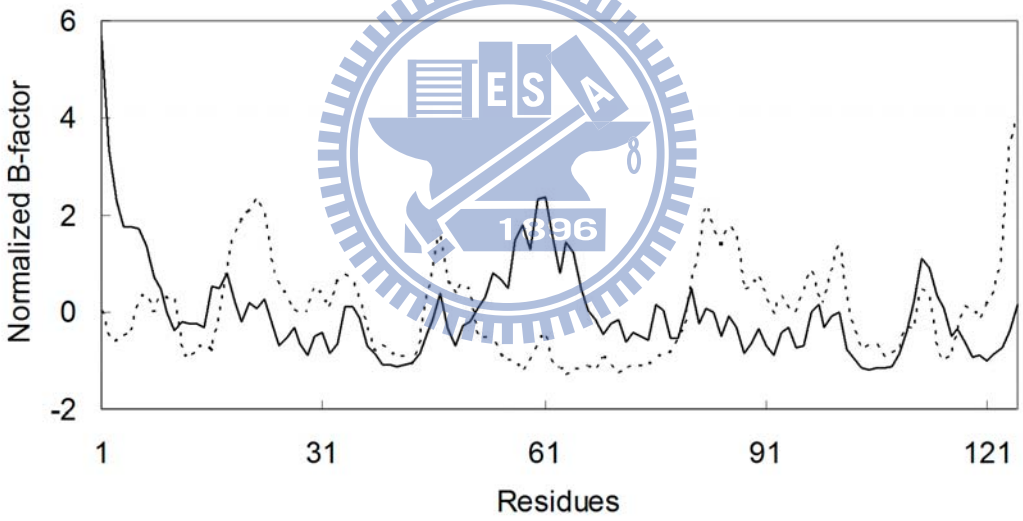


Figure 20. (A) The structure of the chain I of Rubisco from the green alga, *Chlamydomonas reinhardtii* (1GK8:I) which is presented in the cartoon putty representation. **(B)** The X-ray B-factor profile (dotted line) of 1GK8:I compared with the computed B-factor profile (solid line) by the WCN model ($c = -0.06$, $\rho = 0.00$).

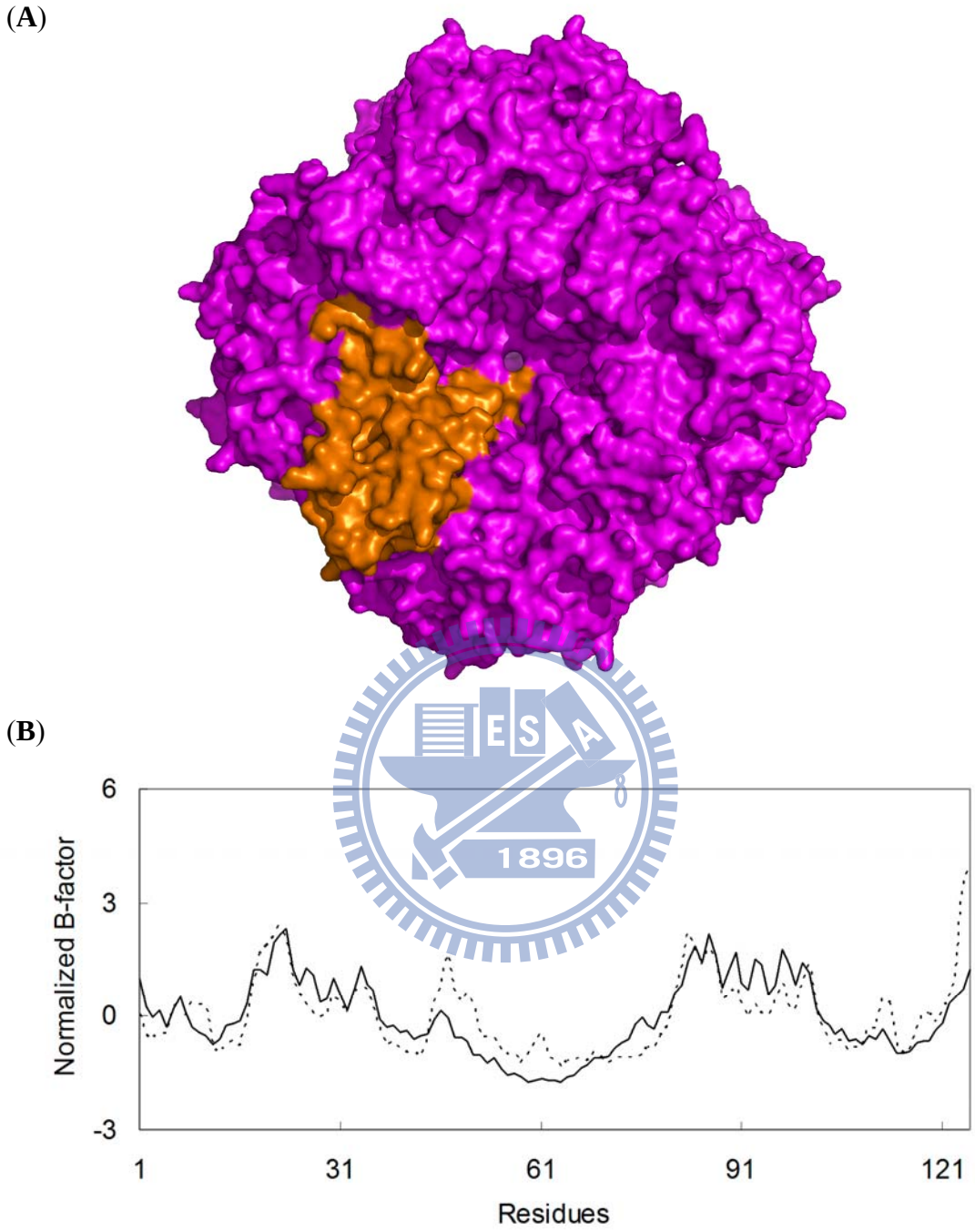


Figure 21. (A) The biological unit of the Rubisco (1GK8) in the surface representation. The chain I is colored in orange, while other chains are colored in magenta. (B) The X-ray B-factor profile (dotted line) of 1GK8:I compared with the WCN profile (solid line) when the biological units are included in WCN computation ($c = 0.77$, $\rho = 0.79$).

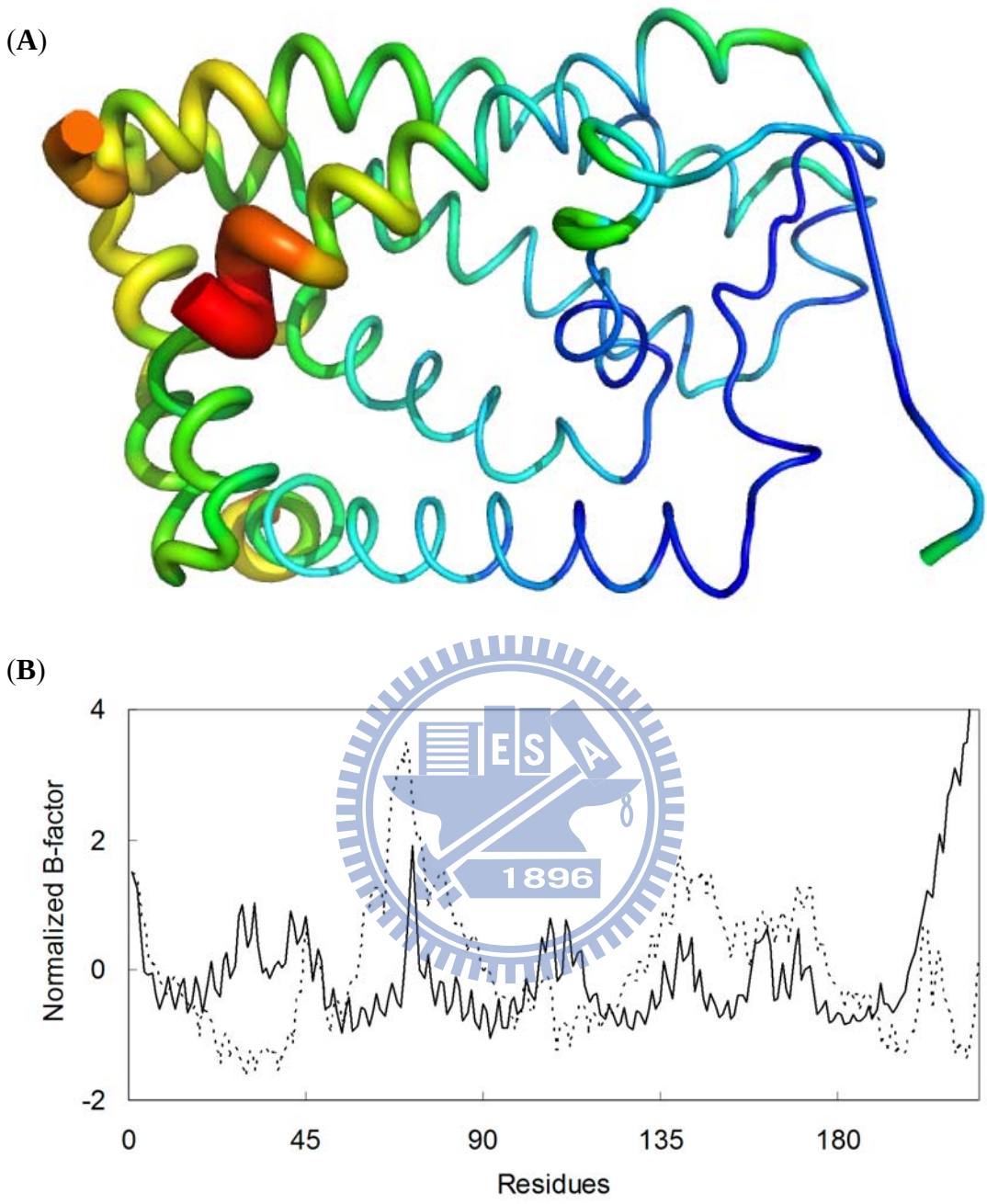


Figure 22. (A) The structure of the chain C of phosphocarrier protein (1Y5I:C) which is presented in the cartoon putty representation. **(B)** The X-ray B-factor profile (dotted line) of 1Y5I:C compared with the WCN profile (solid line) ($c = -0.09$, $\rho = -0.06$).

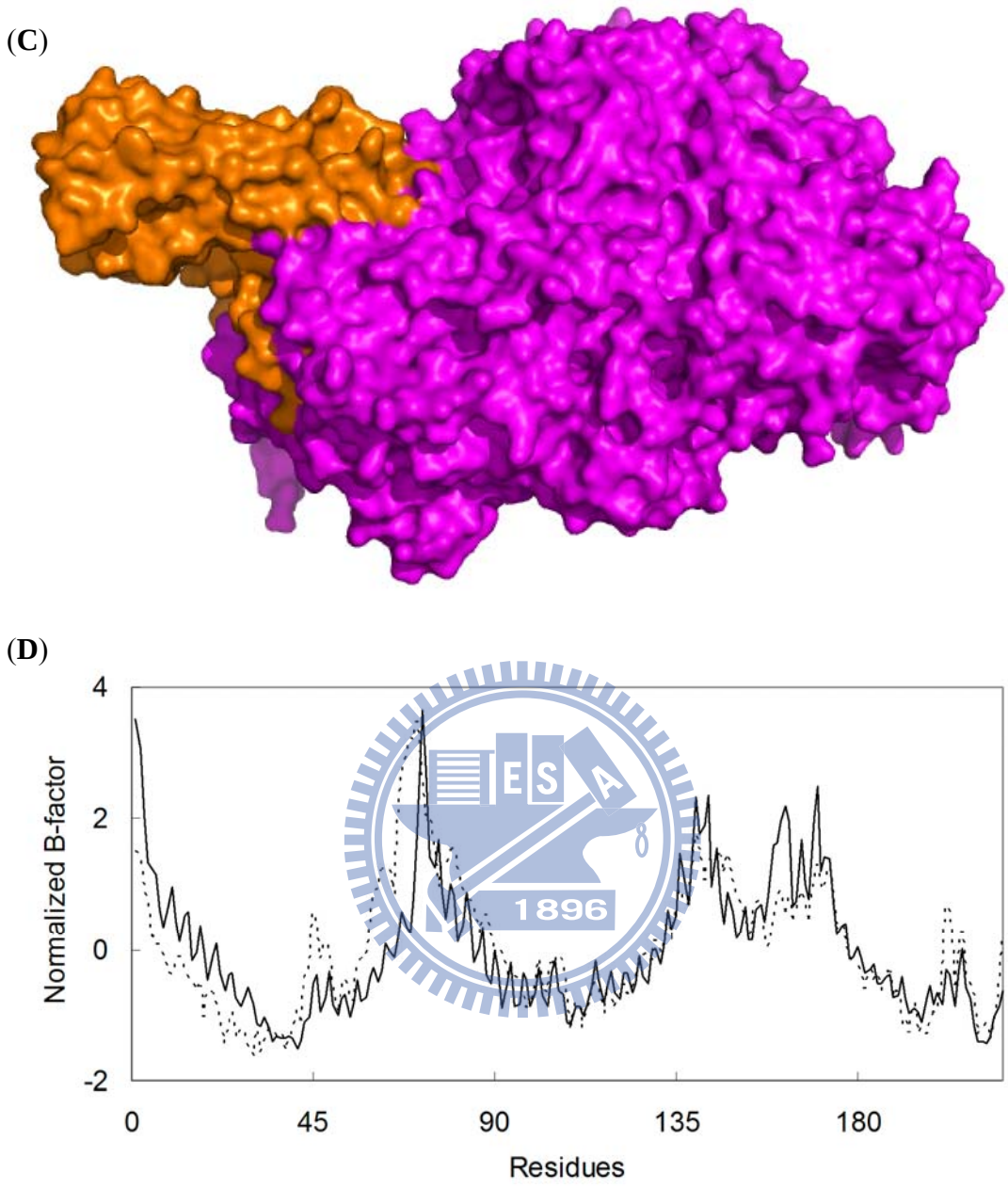


Figure 22. (C) The biological unit of phosphocarrier protein (1Y5I) in the surface representation. The chain C is colored in orange, while other chains are colored in magenta. **(D)** The X-ray B-factor profile of 1Y5I:C compared with the WCN profile when the biological units are included in WCN computation ($c = 0.76$, $\rho = 0.82$).

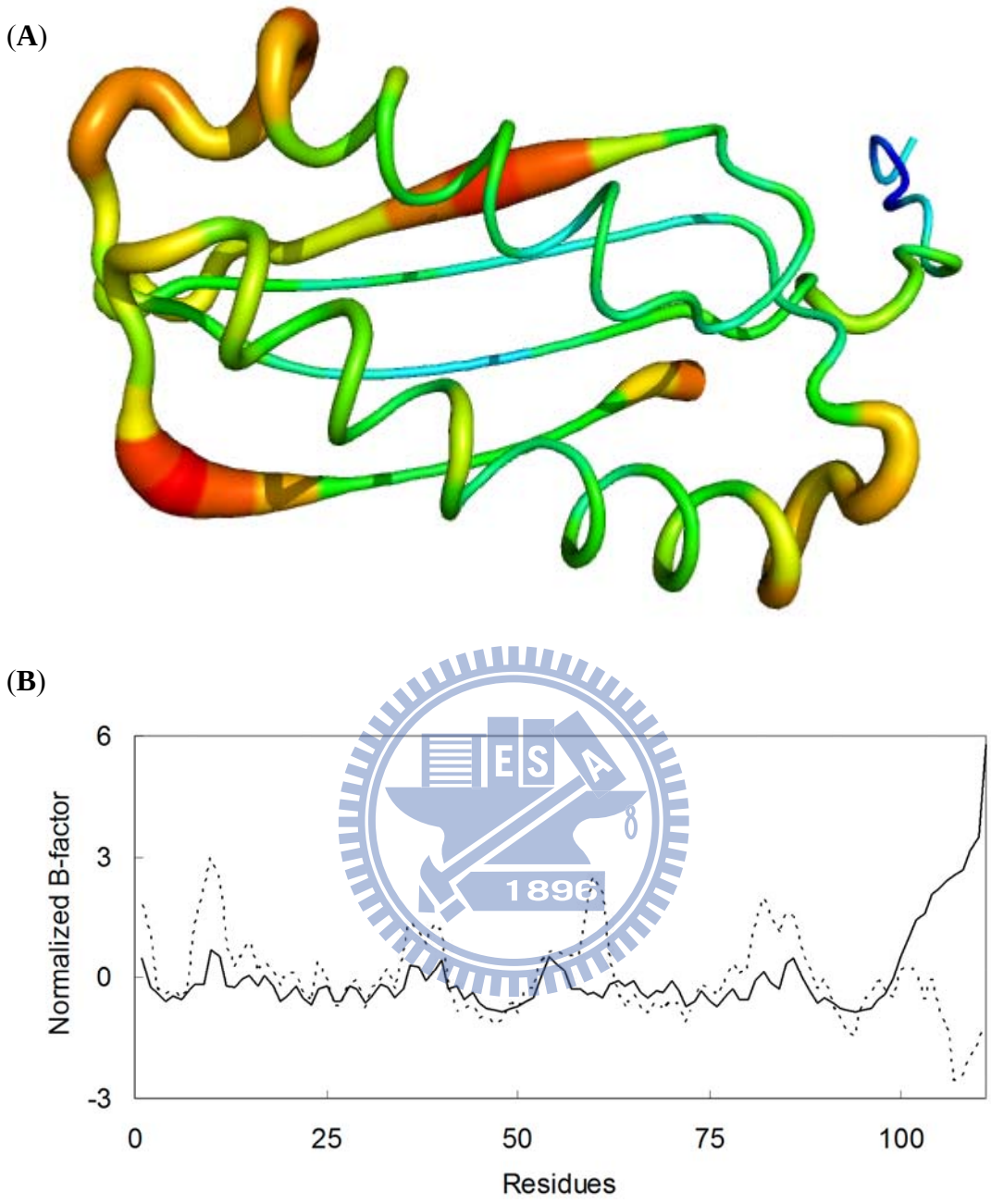
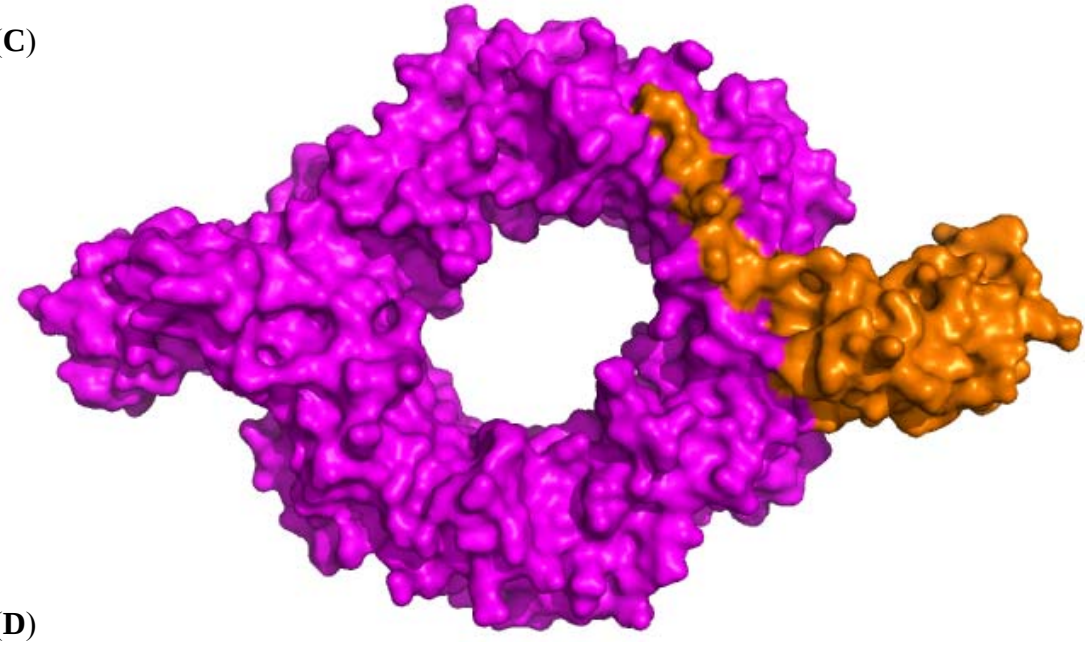


Figure 23. (A) The structure of the chain C of little finger fragment of DNA polymerase IV (1UNN:C) which is presented in the cartoon putty representation. **(B)** The X-ray B-factor profile (dotted line) of 1UNN:C compared with the computed B-factor profile (solid line) by the WCN model ($c = -0.16$, $\rho = 0.39$).

(**C**)



(**D**)

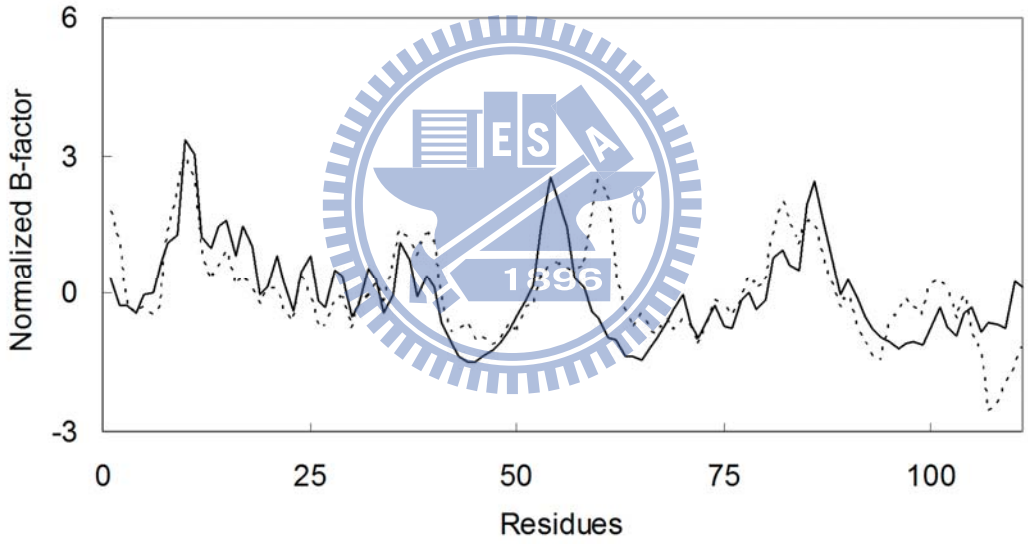


Figure 23. (C) The biological unit of little finger fragment of DNA polymerase IV (1UNN) in the surface representation. The chain C is colored in orange, while other chains are colored in magenta. **(D)** The X-ray B-factor profile of 1UNN:C compared with the WCN profile when the biological units are included in WCN computation ($c = 0.64$, $\rho = 0.67$).

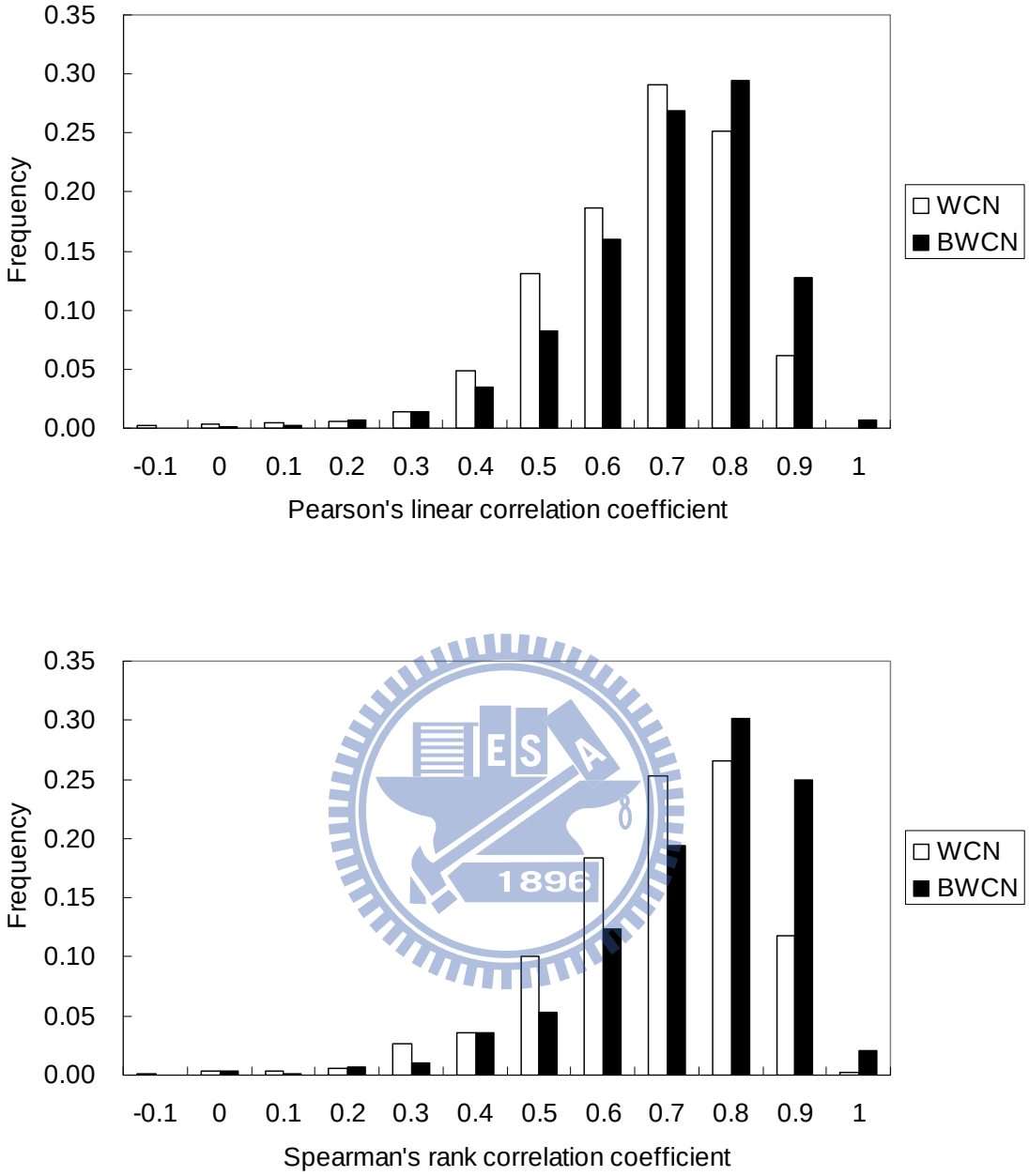
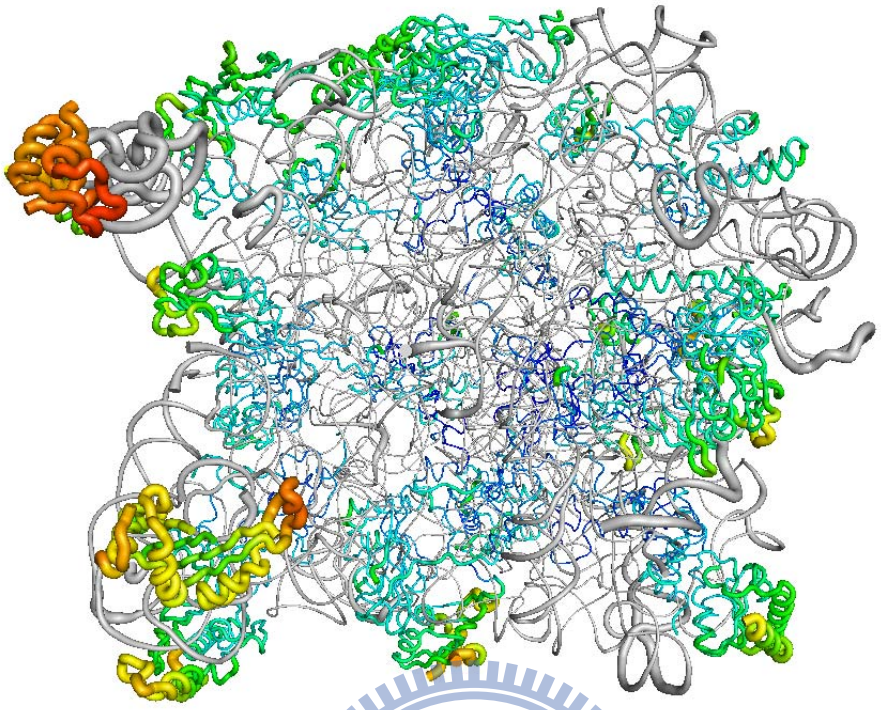


Figure 24. Comparison of the correlation coefficients between the experimental and the computed B-factor profiles bases on the WCN model (white) and the WCN model taking other parts of the whole biological units into consideration (black) for the nonhomologous data set.

(**A**)



(**B**)

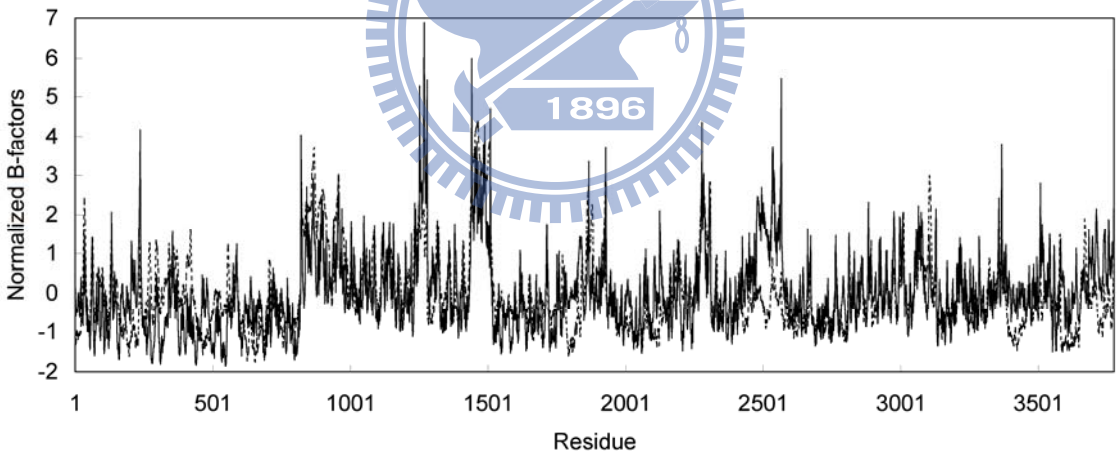


Figure 25. (**A**) The structure of the 50S ribosomal subunit (1YJW) which is presented in the cartoon putty representation. The gray parts are DNA structures. (**B**) The X-ray B-factor profile (dotted line) of compared with the computed B-factor profile (solid line). The correlation coefficient between the X-ray and the computed B-factors is $c = 0.60$ and $\rho = 0.49$.

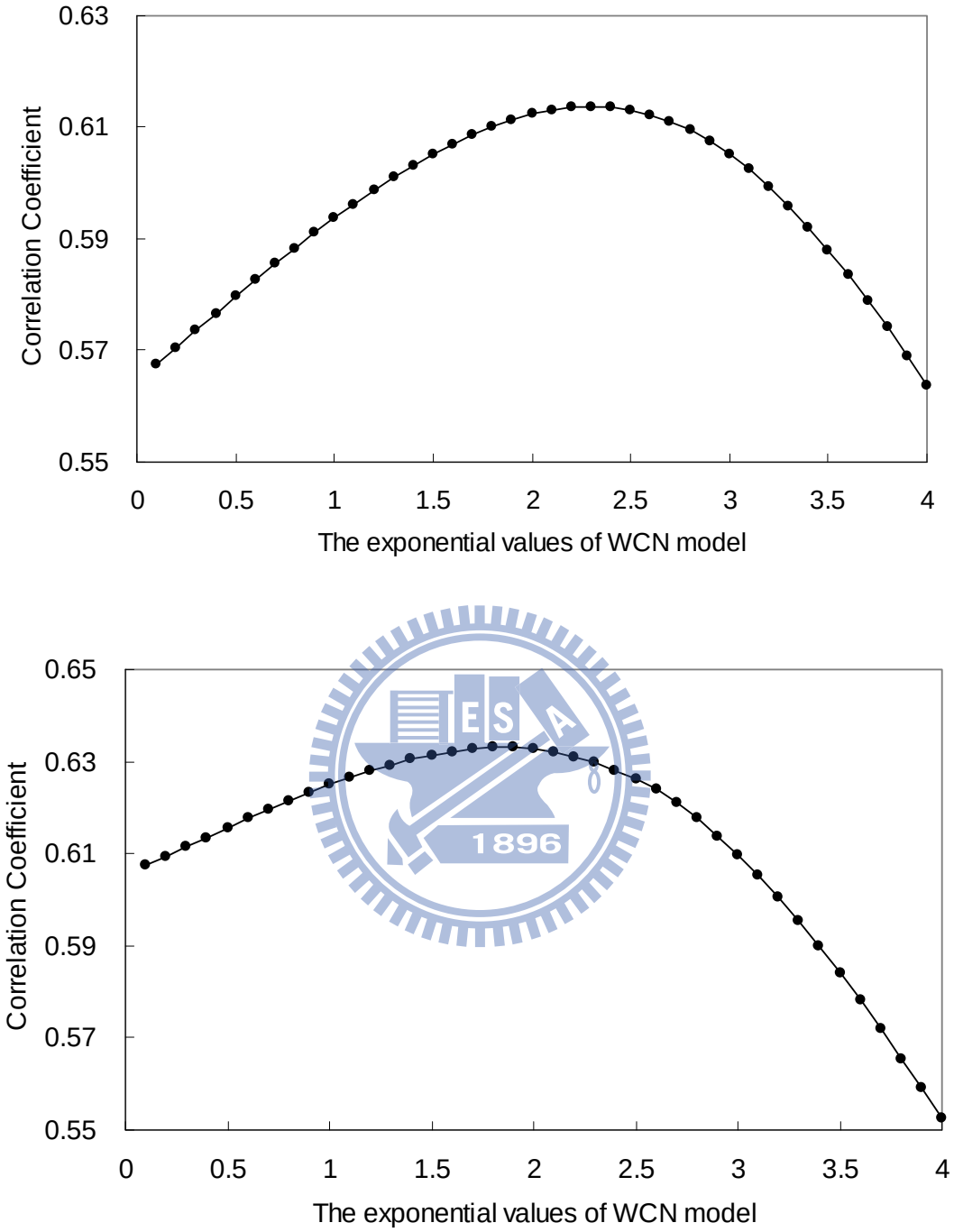


Figure 26. The performances of the correlation coefficients are in different exponent values of the reciprocal distance in WCN model. The upper part is Pearson's linear correlation coefficients and the lower part is Spearman's rank correlation coefficients.

Table 3. The correlation coefficients of cross-correlation maps between the WCN model, the GNM and the NMA.

	GNM		NMA	
	$\overline{c}$	$\overline{\rho}$	$\overline{c}$	$\overline{\rho}$
WCN	0.79	0.80	0.83	0.82
GNM	-	-	0.76	0.81

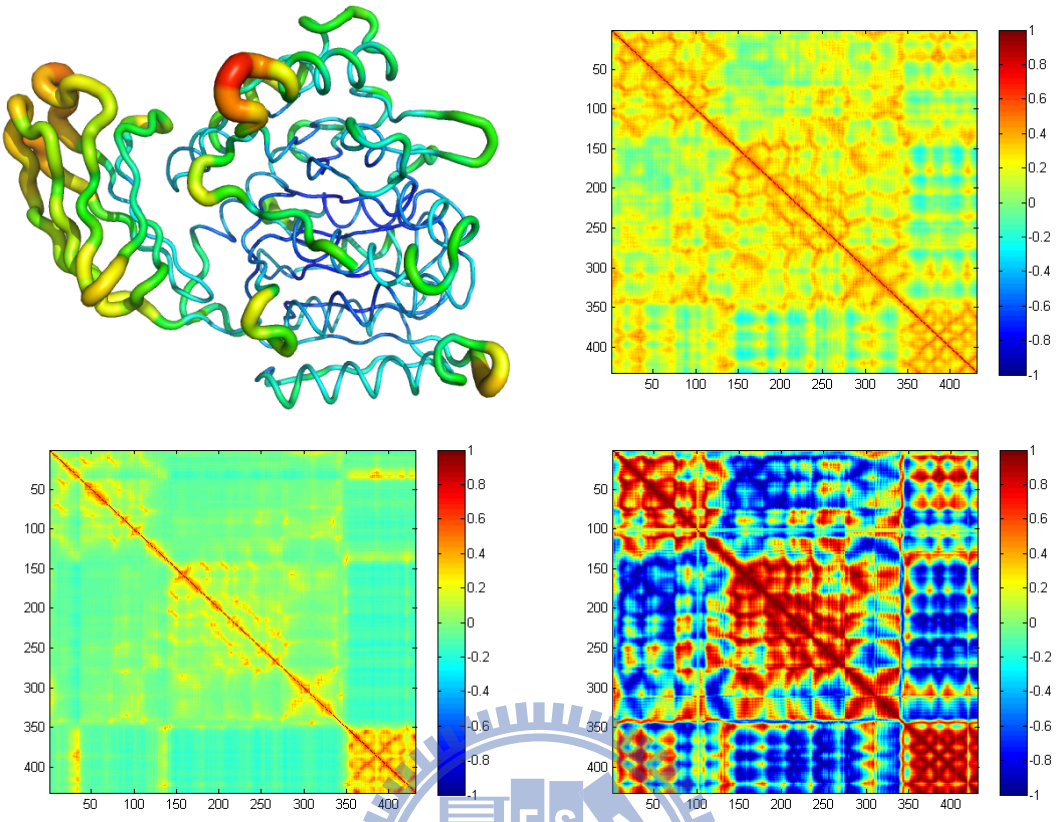


Figure 27. The cross-correlation maps for the chain A of Arg-specific cysteine proteinase gingipain R (1CVR:A). The map on the upper right is computed by Eq. 10 and normalized by Eq. 12, the map on the lower left by GNM and the map on the lower right by the NMA. The colors of the map ramp from red (positive correlation) to blue (negative correlation). NMA was performed using the simplified force field of ENZYMIX[36,37]. For 1RWH:A, Pearson's linear correlation coefficient between WCN and NMA is $c_{WCN-NMA} = 0.85$; WCN and GNM is $c_{WCN-GNM} = 0.79$; GNM and NMA is $c_{GNM-NMA} = 0.85$. Spearman correlation coefficients are $\rho_{WCN-NMA} = 0.88$, $\rho_{WCN-GNM} = 0.79$ and $\rho_{GNM-NMA} = 0.87$.

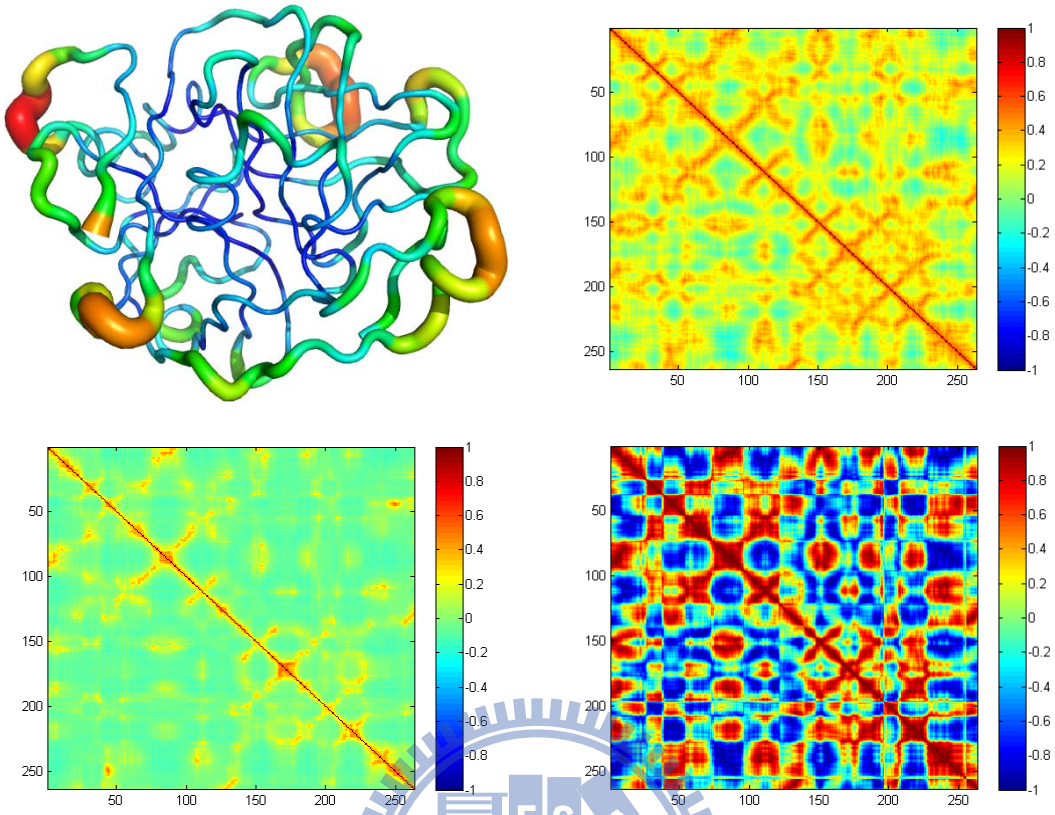


Figure 28. The cross-correlation maps for the achromobacter lyticus protease I (1ARB). The map on the upper right is computed by Eq. 10 and normalized by Eq. 12, the map on the lower left by GNM and the map on the lower right by the NMA. The colors of the map ramp from red (positive correlation) to blue (negative correlation). NMA was performed using the simplified force field of ENZYMIX[36,37]. For 1ARB, Pearson's linear correlation coefficient between WCN and NMA is $c_{WCN-NMA} = 0.80$; WCN and GNM, $c_{WCN-GNM} = 0.83$; and GNM and NMA, $c_{GNM-NMA} = 0.69$. Spearman correlation coefficients are $\rho_{WCN-NMA} = 0.83$, $\rho_{WCN-GNM} = 0.90$ and $\rho_{GNM-NMA} = 0.86$.

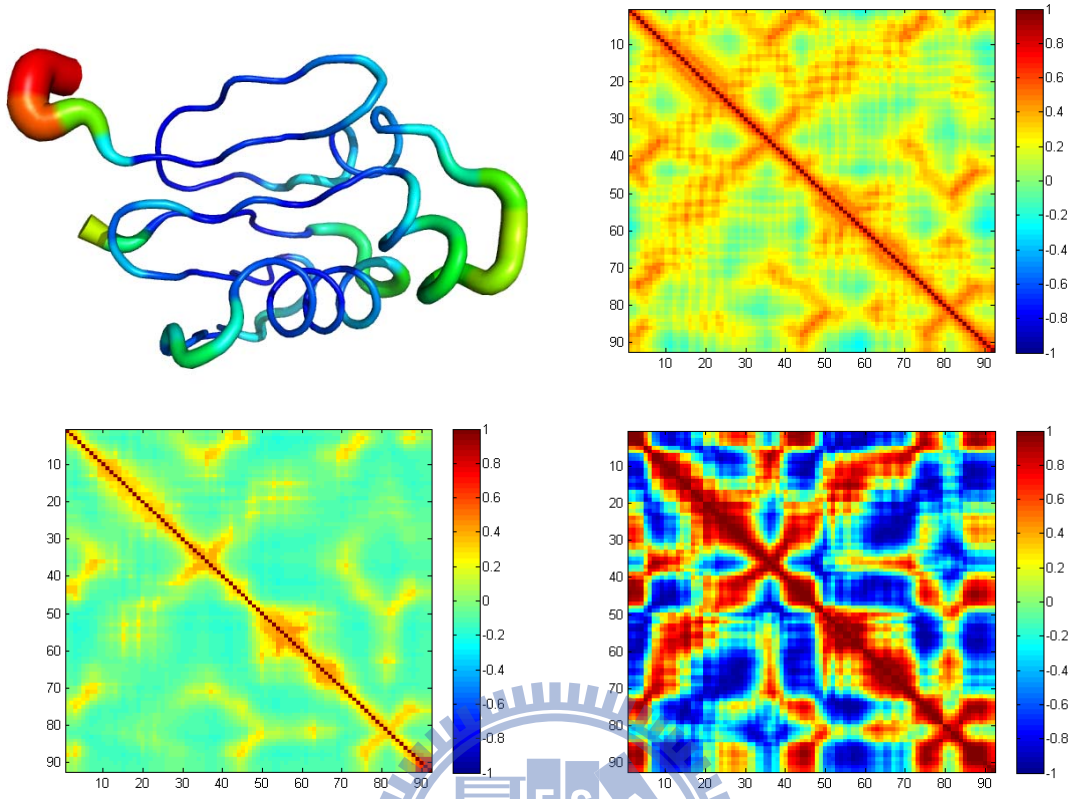


Figure 29. The cross-correlation maps for the chain A of toluene 4-monooxygenase catalytic effector protein (2BF5:A). The map on the upper right is computed by Eq. 10 and normalized by Eq. 12, the map on the lower left by GNM and the map on the lower right by the NMA. The colors of the map ramp from red (positive correlation) to blue (negative correlation). NMA was performed using the simplified force field of ENZYMIX[36,37]. For 2BF5:A, Pearson's linear correlation coefficient between WCN and NMA is $c_{WCN-NMA} = 0.81$; WCN and GNM, $c_{WCN-GNM} = 0.87$; and GNM and NMA, $c_{GNM-NMA} = 0.71$. Spearman correlation coefficients are $\rho_{WCN-NMA} = 0.84$, $\rho_{WCN-GNM} = 0.84$ and $\rho_{GNM-NMA} = 0.82$.

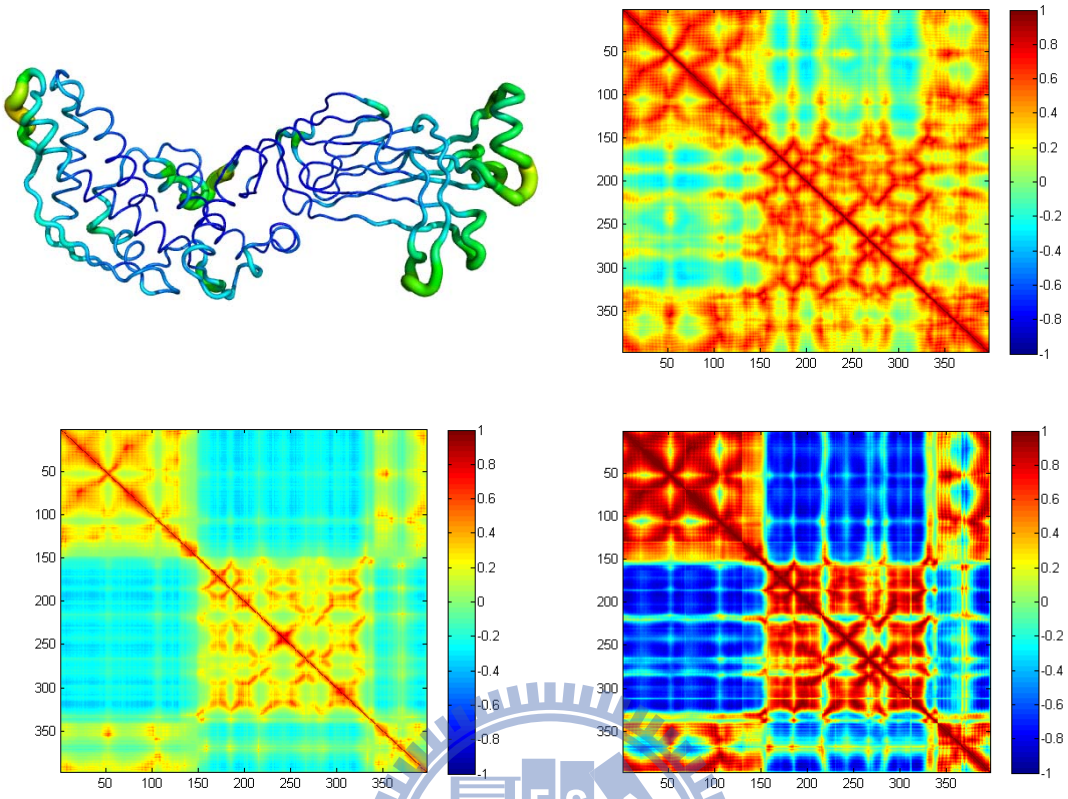


Figure 30. The cross-correlation maps for the chain A of major capsid protein of group A rotavirus (1QHD:A). The map on the upper right is computed by Eq. 10 and normalized by Eq. 12, the map on the lower left by GNM and the map on the lower right by the NMA. The colors of the map ramp from red (positive correlation) to blue (negative correlation). NMA was performed using the simplified force field of ENZYMIX[36,37]. For 1QHD:A, Pearson's linear correlation coefficient between WCN and NMA is $c_{WCN-NMA} = 0.88$; WCN and GNM, $c_{WCN-GNM} = 0.87$; and GNM and NMA, $c_{GNM-NMA} = 0.94$. Spearman correlation coefficients are $\rho_{WCN-NMA} = 0.90$, $\rho_{WCN-GNM} = 0.88$ and $\rho_{GNM-NMA} = 0.95$.

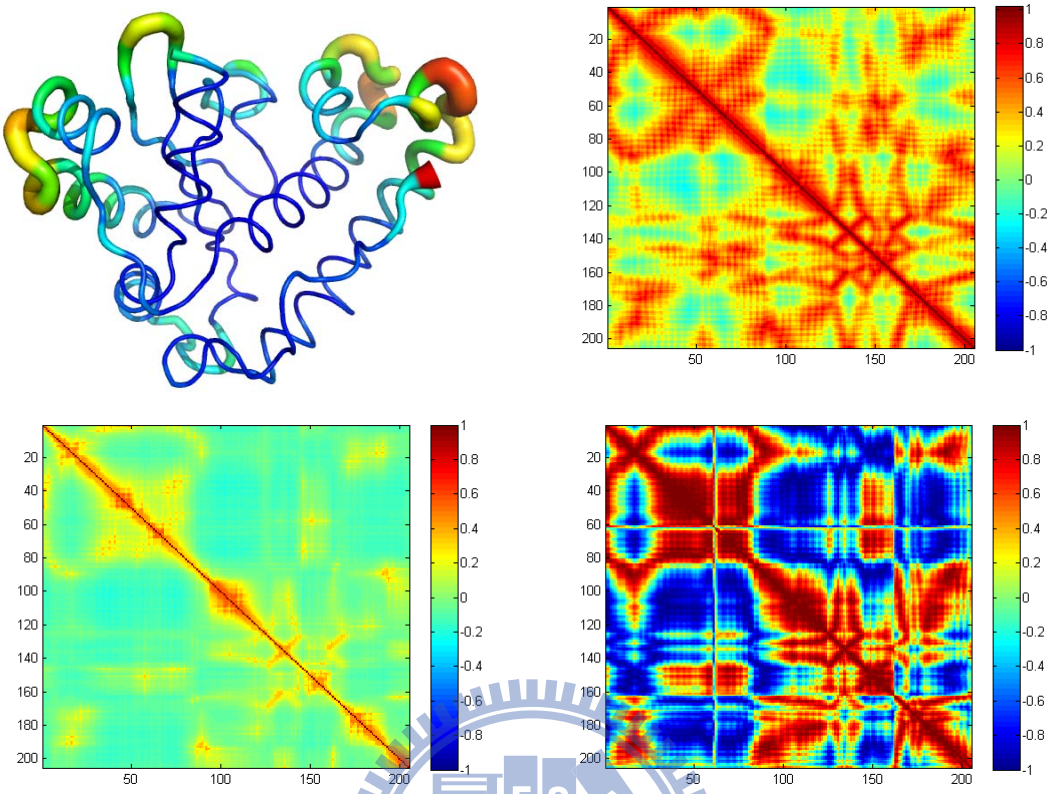


Figure 31. The cross-correlation maps for the chain A of E. Coli manganese(II) superoxide dismutase (1IXB:A). The map on the upper right is computed by Eq. 10 and normalized by Eq. 12, the map on the lower left by GNM and the map on the lower right by the NMA. The colors of the map ramp from red (positive correlation) to blue (negative correlation). NMA was performed using the simplified force field of ENZYMIX[36,37]. For 1IXB:A, Pearson's linear correlation coefficient between WCN and NMA is $c_{WCN-NMA} = 0.77$; WCN and GNM, $c_{WCN-GNM} = 0.81$; and GNM and NMA, $c_{GNM-NMA} = 0.73$. Spearman correlation coefficients are $\rho_{WCN-NMA} = 0.77$, $\rho_{WCN-GNM} = 0.85$ and $\rho_{GNM-NMA} = 0.85$.

Table 4. The correlation coefficients of the inverse of eigenvalues in all modes and first 20 slowest modes between the WCN model, the GNM and the NMA.

	GNM		NMA	
	all modes	first 20 slowest modes	all modes	first 20 slowest modes
WCN	0.99	0.93	0.96	0.85
GNM	-	-	0.95	0.97

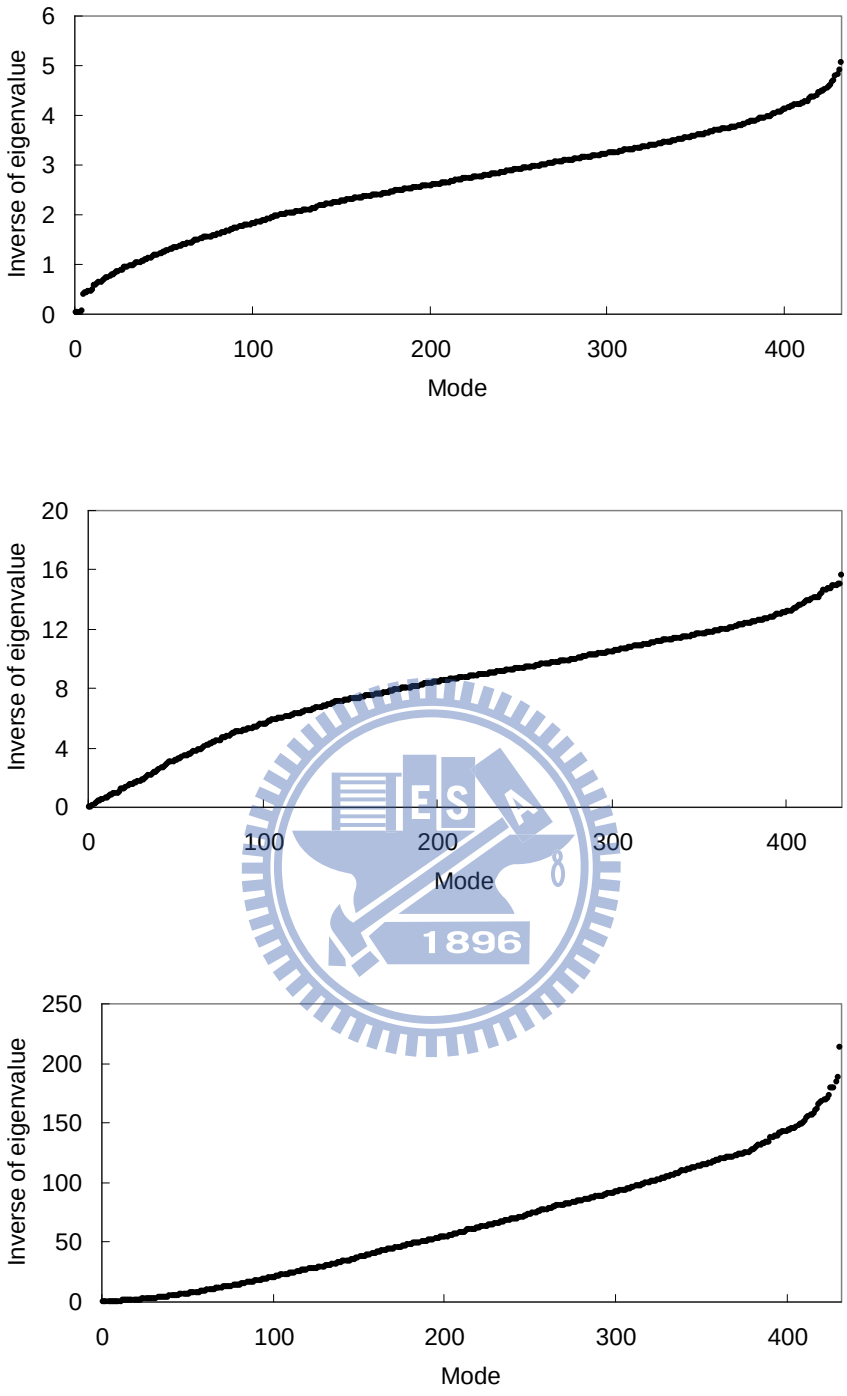


Figure 32. The inverse of eigenvalues of all modes for the chain A of Arg-specific cysteine proteinase gingipain R (1CVR:A). The upper one is plotted by diagonalizing the cross-correlation map from WCN, the middle one from GNM and the low one from NMA.

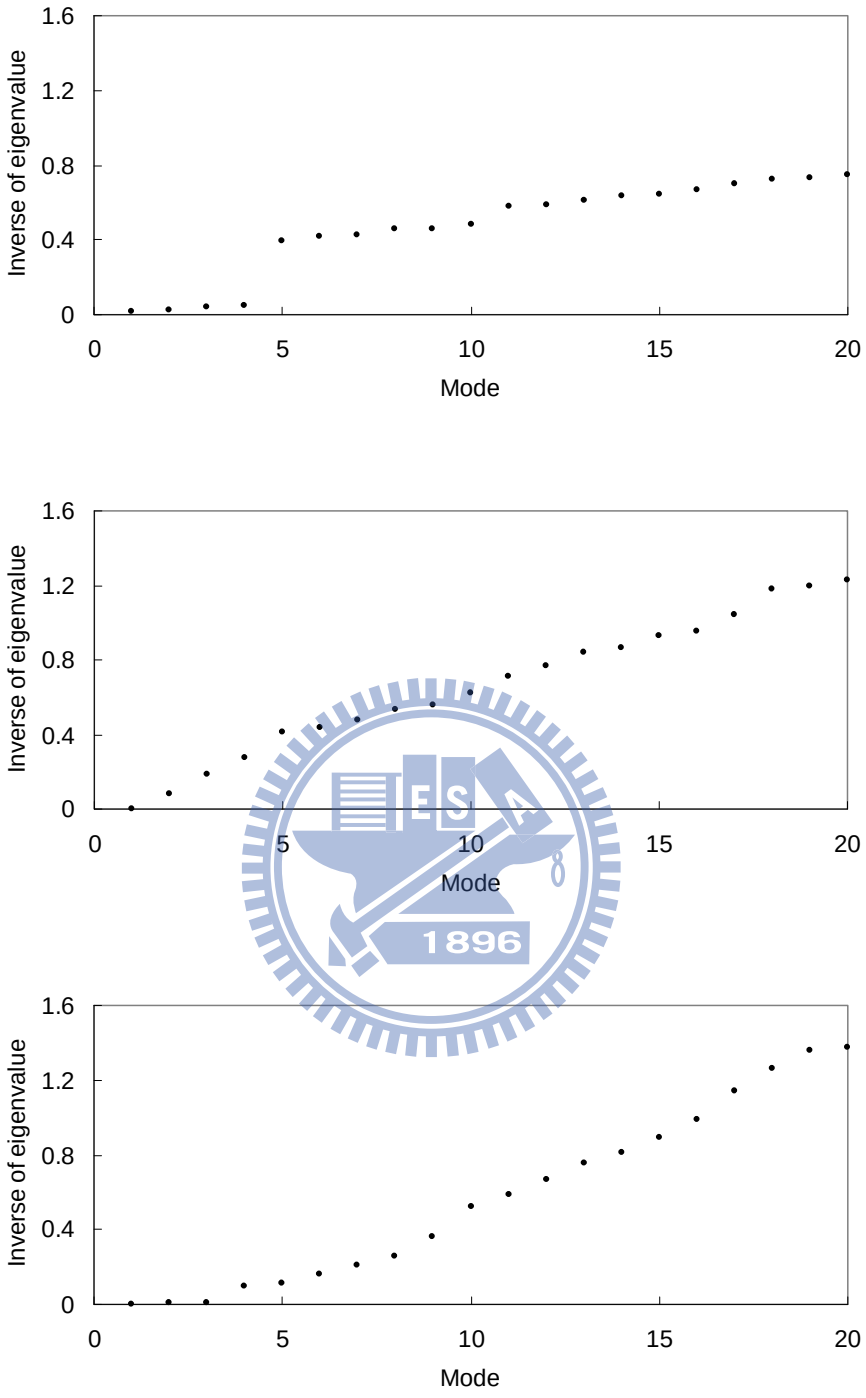


Figure 33. The inverse of eigenvalues of the first 20 slowest modes for the chain A of Arg-specific cysteine proteinase gingipain R (1CVR:A). The upper one is plotted by diagonalizing the cross-correlation map from WCN, the middle one from GNM and the low one from NMA.

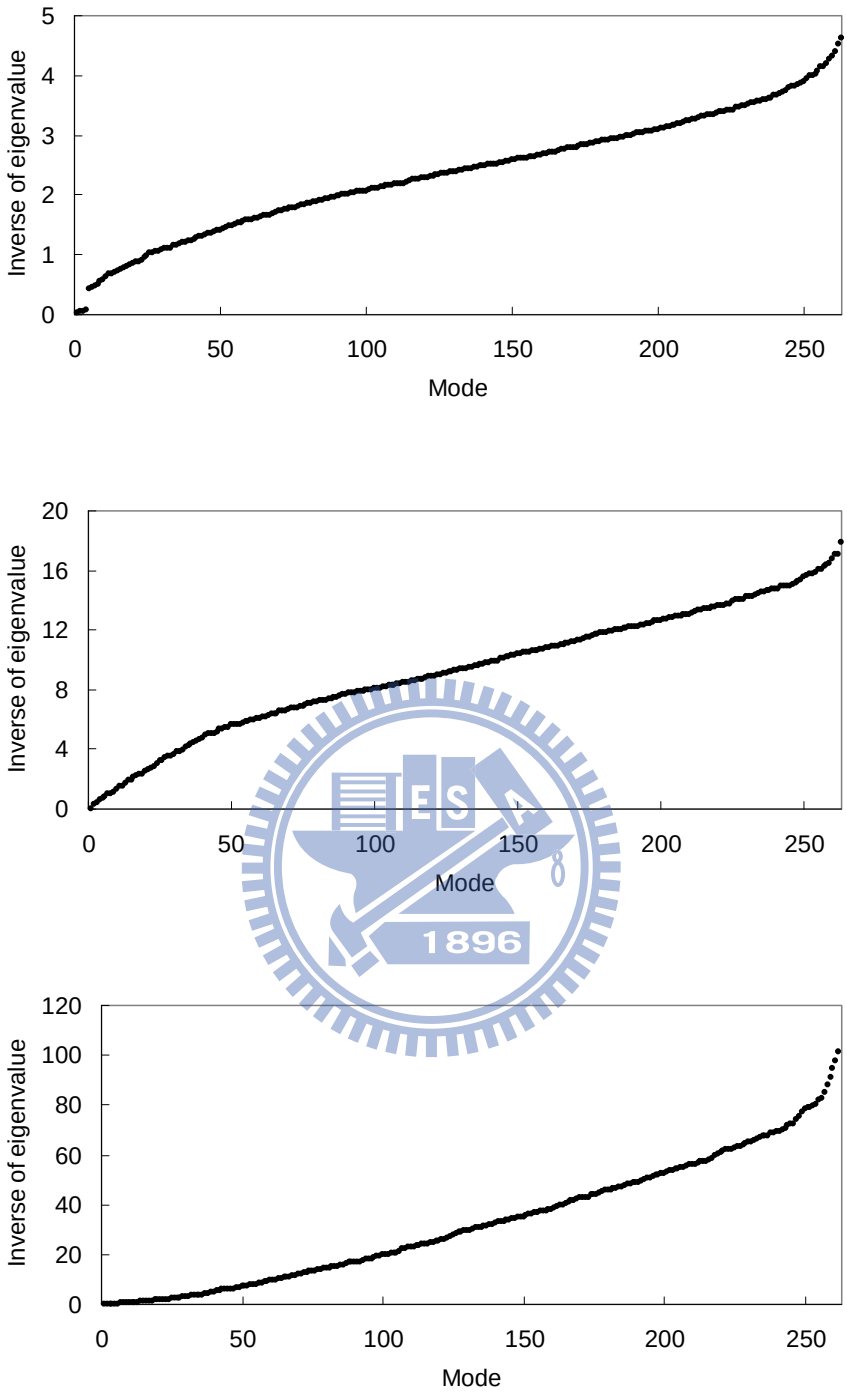


Figure 34. The inverse of eigenvalues of all modes for the achromobacter lyticus protease I (1ARB). The upper distribution is plotted by diagonalizing the cross-correlation map from WCN, the middle one from GNM and the low one from NMA.

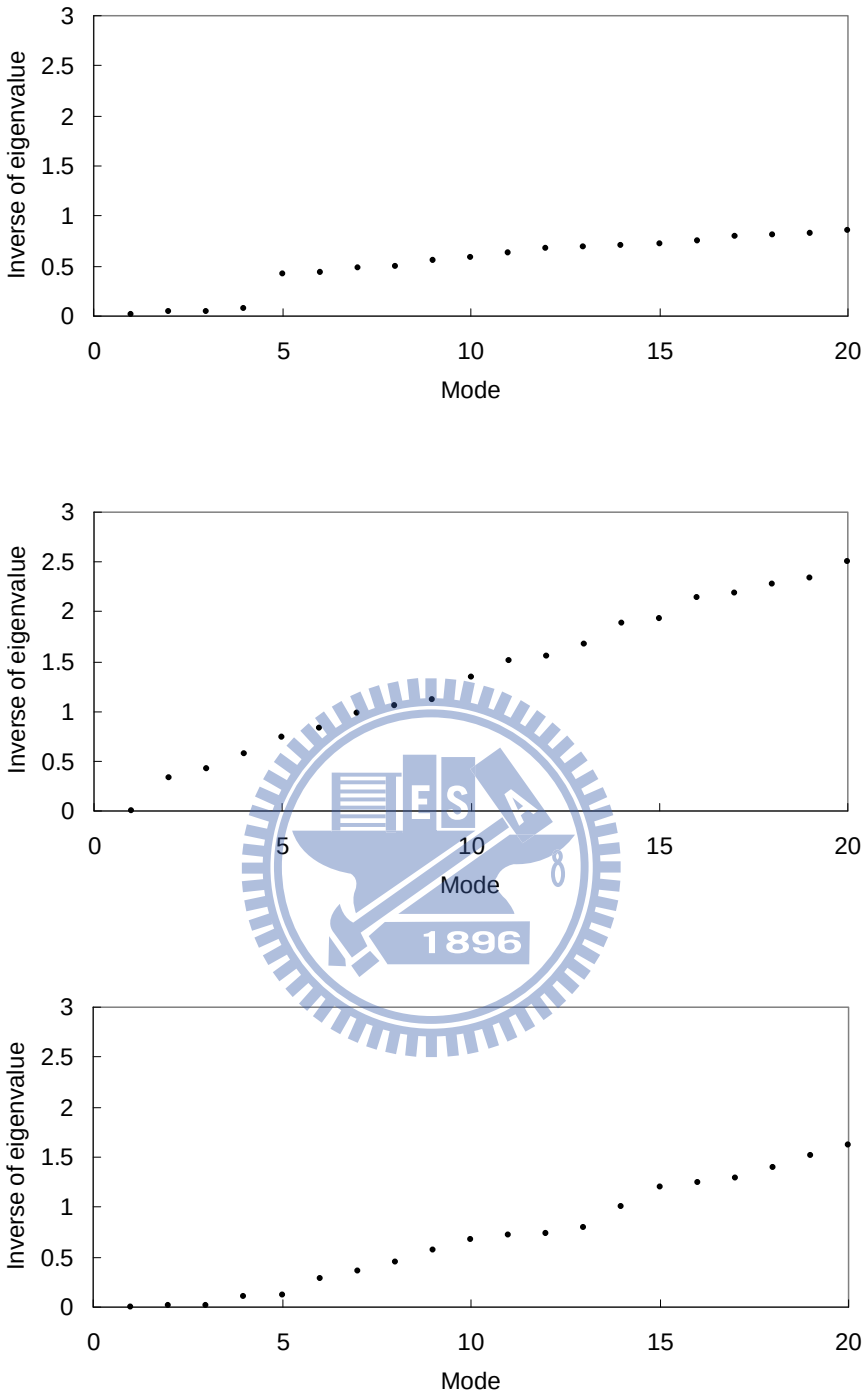


Figure 35. The inverse of eigenvalues of the first 20 slowest modes for the achromobacter lyticus protease I (1ARB). The upper distribution is plotted by diagonalizing the cross-correlation map from WCN, the middle one from GNM and the low one from NMA.

Table 5. The correlation coefficients of eigenvectors in first mode, second mode and third mode between the WCN model, the GNM and the NMA.

	GNM						NMA					
	First Mode		Second Mode		Third Mode		First Mode		Second Mode		Third Mode	
	$\bar{c}$	$\bar{\rho}$	$\bar{c}$	$\bar{\rho}$	$\bar{c}$	$\bar{\rho}$	$\bar{c}$	$\bar{\rho}$	$\bar{c}$	$\bar{\rho}$	$\bar{c}$	$\bar{\rho}$
WCN	0.81	0.83	0.61	0.61	0.46	0.46	0.82	0.83	0.63	0.64	0.59	0.61
GNM	-	-	-	-	-	-	0.76	0.77	0.52	0.53	0.41	0.43

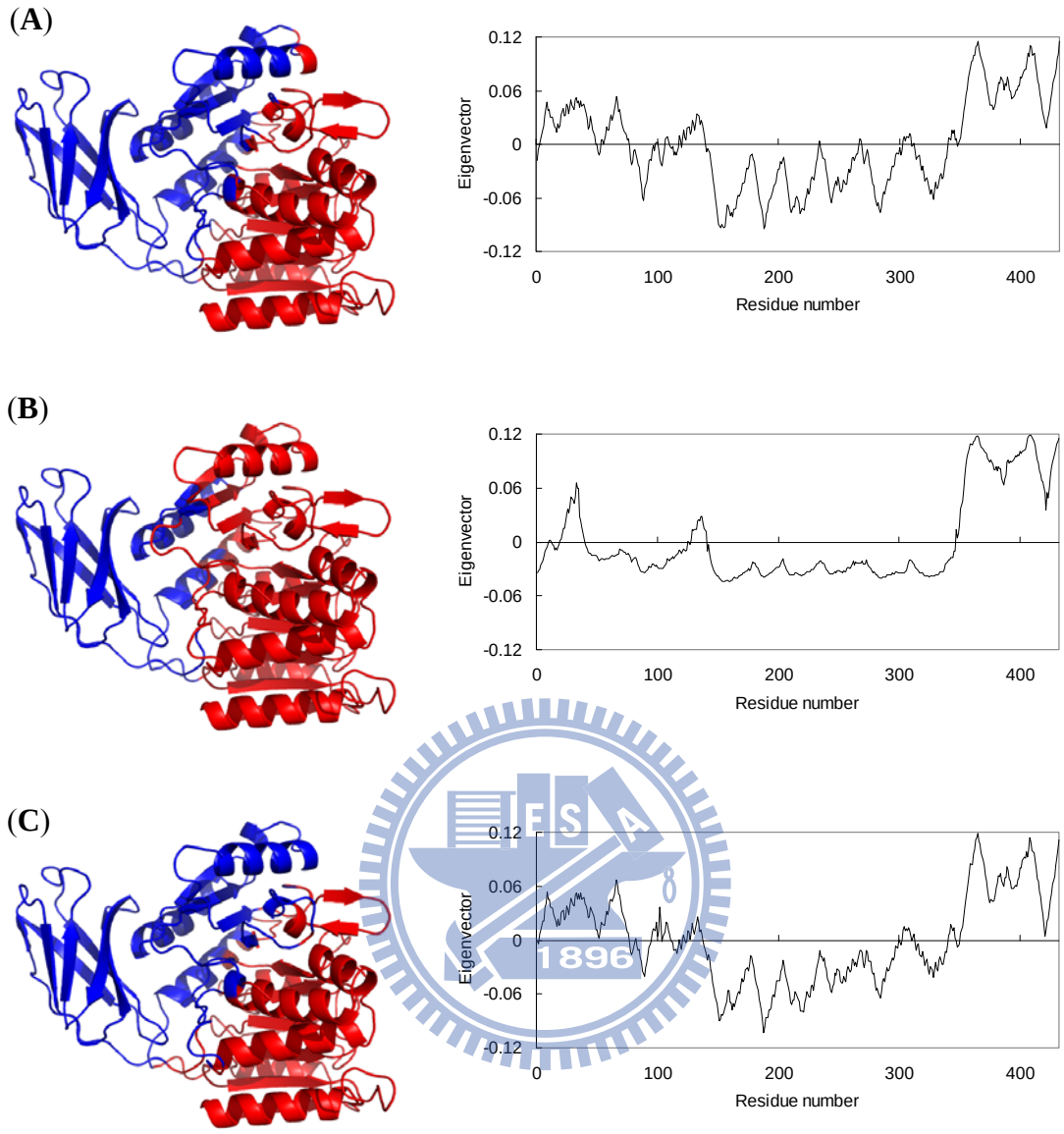


Figure 36. The regions subject to opposite direction displacements and the distribution of displacements along the first mode computed by (**A**) the WCN model for the chain A of Arg-specific cysteine proteinase gingipain R (1CVR:A). Regions colored in blue and red correspond to positive and negative displacements respectively. The diagram shows the vibration amplitudes and phases (positive or negative). The same illustrations of figure and diagram show the results by (**B**) the GNM and (**C**) the NMA.

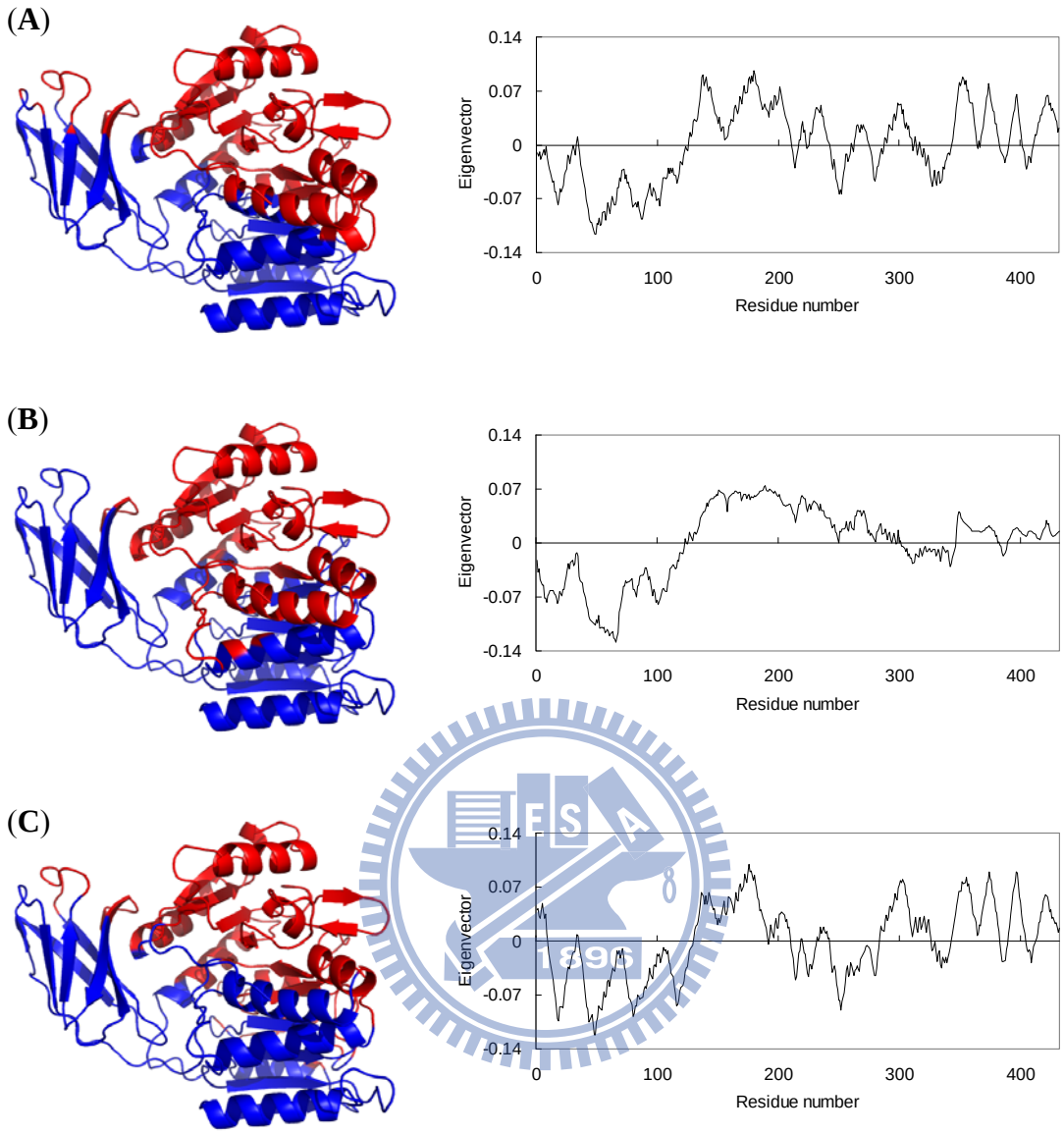


Figure 37. The regions subject to opposite direction displacements and the distribution of displacements along the second mode computed by (**A**) the WCN model for the chain A of Arg-specific cysteine proteinase gingipain R (1CVR:A). Regions colored in blue and red correspond to positive and negative displacements respectively. The diagram shows the vibration amplitudes and phases (positive or negative). The same illustrations of figure and diagram show the results by (**B**) the GNM and (**C**) the NMA.

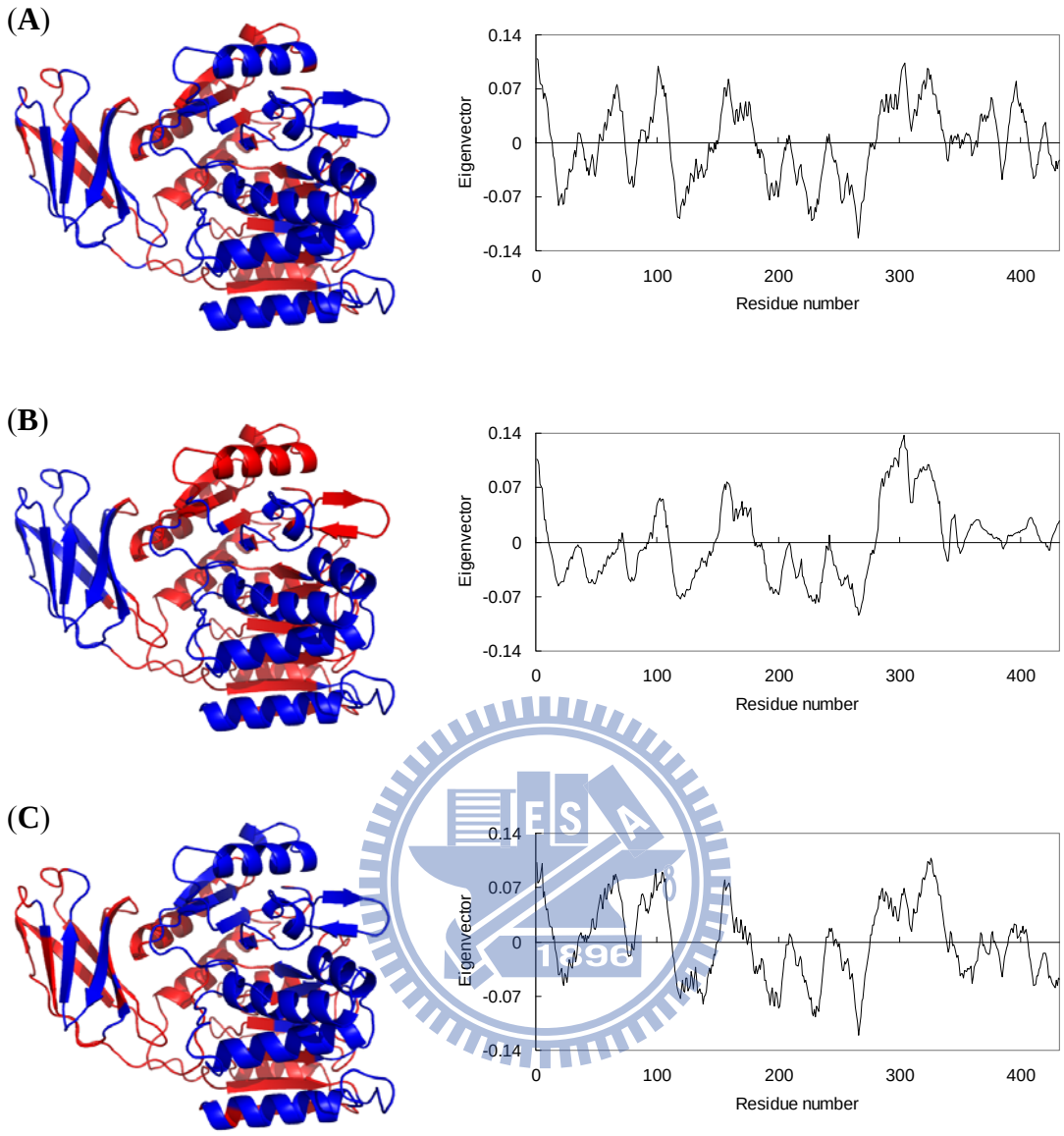


Figure 38. The regions subject to opposite direction displacements and the distribution of displacements along the third mode computed by (**A**) the WCN model for the chain A of Arg-specific cysteine proteinase gingipain R (1CVR:A). Regions colored in blue and red correspond to positive and negative displacements respectively. The diagram shows the vibration amplitudes and phases (positive or negative). The same illustrations of figure and diagram show the results by (**B**) the GNM and (**C**) the NMA.

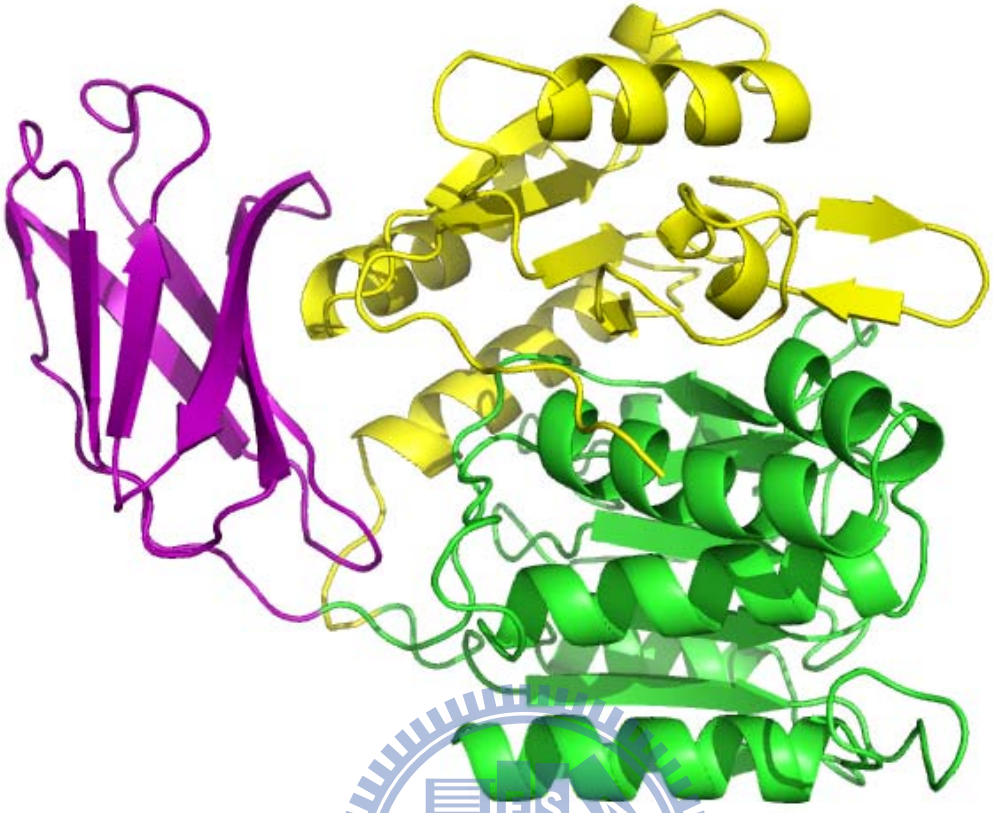


Figure 39. The chain A of Arg-specific cysteine proteinase gingipain R (1CVR:A) consists of the catalytic domain subdivided into A-subdomains (yellow) and B-subdomains (green), and the IgSF domain (purple). Each subdomain comprises a central β-sheet and a few additional hairpins flanked by helices on either side, as characteristic for α/β open-sheet structures[34].

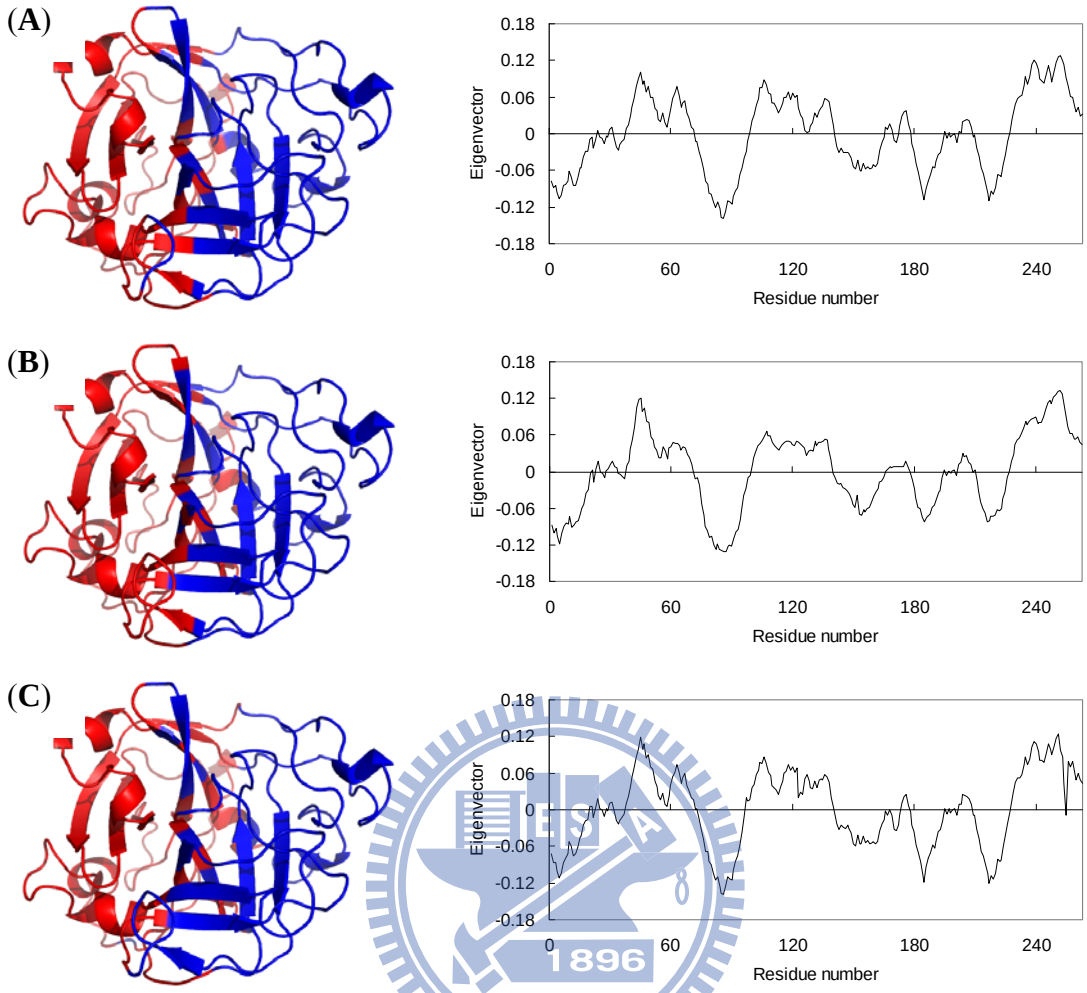


Figure 40. The regions subject to opposite direction displacements and the distribution of displacements along the first mode computed by (**A**) the WCN model for the achromobacter lyticus protease I (1ARB). Regions colored in blue and red correspond to positive and negative displacements respectively. The diagram shows the vibration amplitudes and phases (positive or negative). The same illustrations of figure and diagram show the results by (**B**) the GNM and (**C**) the NMA.

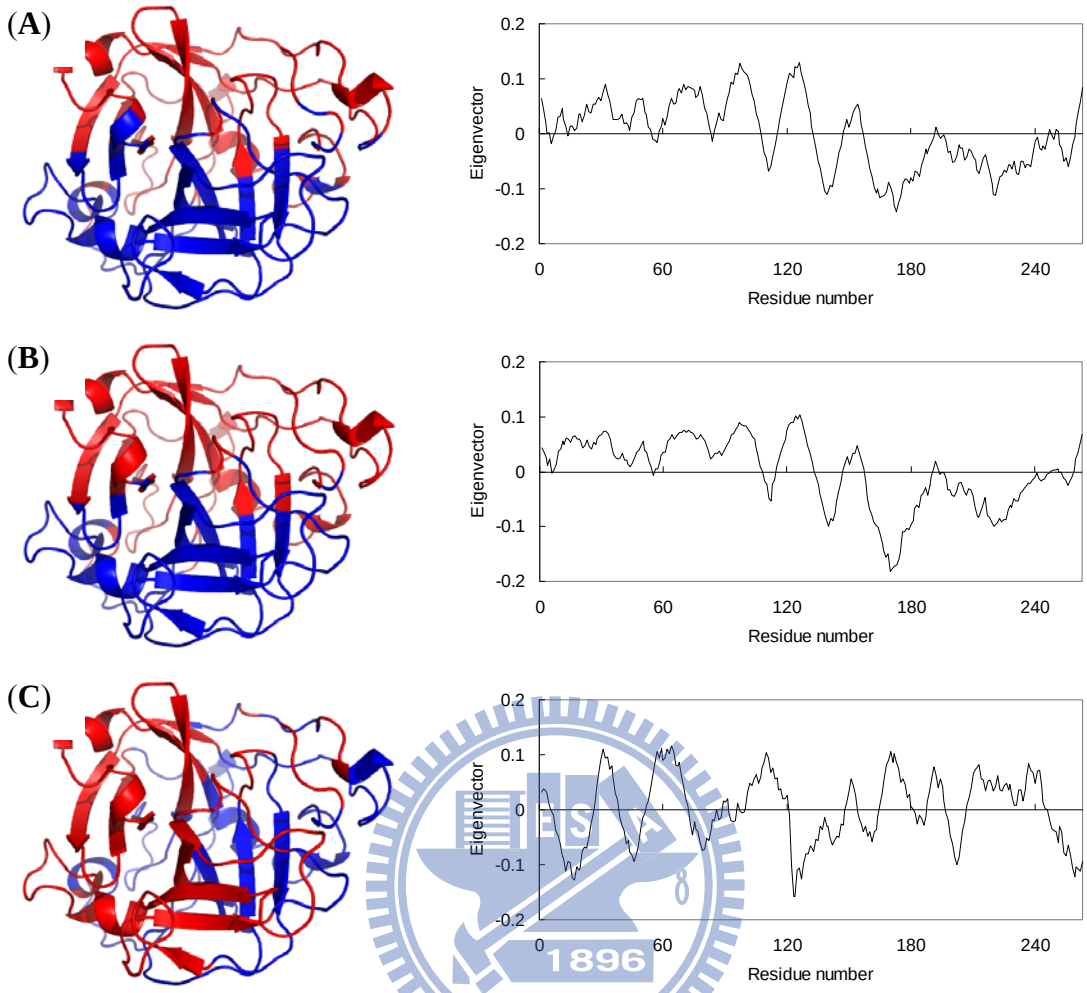


Figure 41. The regions subject to opposite direction displacements and the distribution of displacements along the second mode computed by (**A**) the WCN model for the achromobacter lyticus protease I (1ARB). Regions colored in blue and red correspond to positive and negative displacements respectively. The diagram shows the vibration amplitudes and phases (positive or negative). The same illustrations of figure and diagram show the results by (**B**) the GNM and (**C**) the NMA.

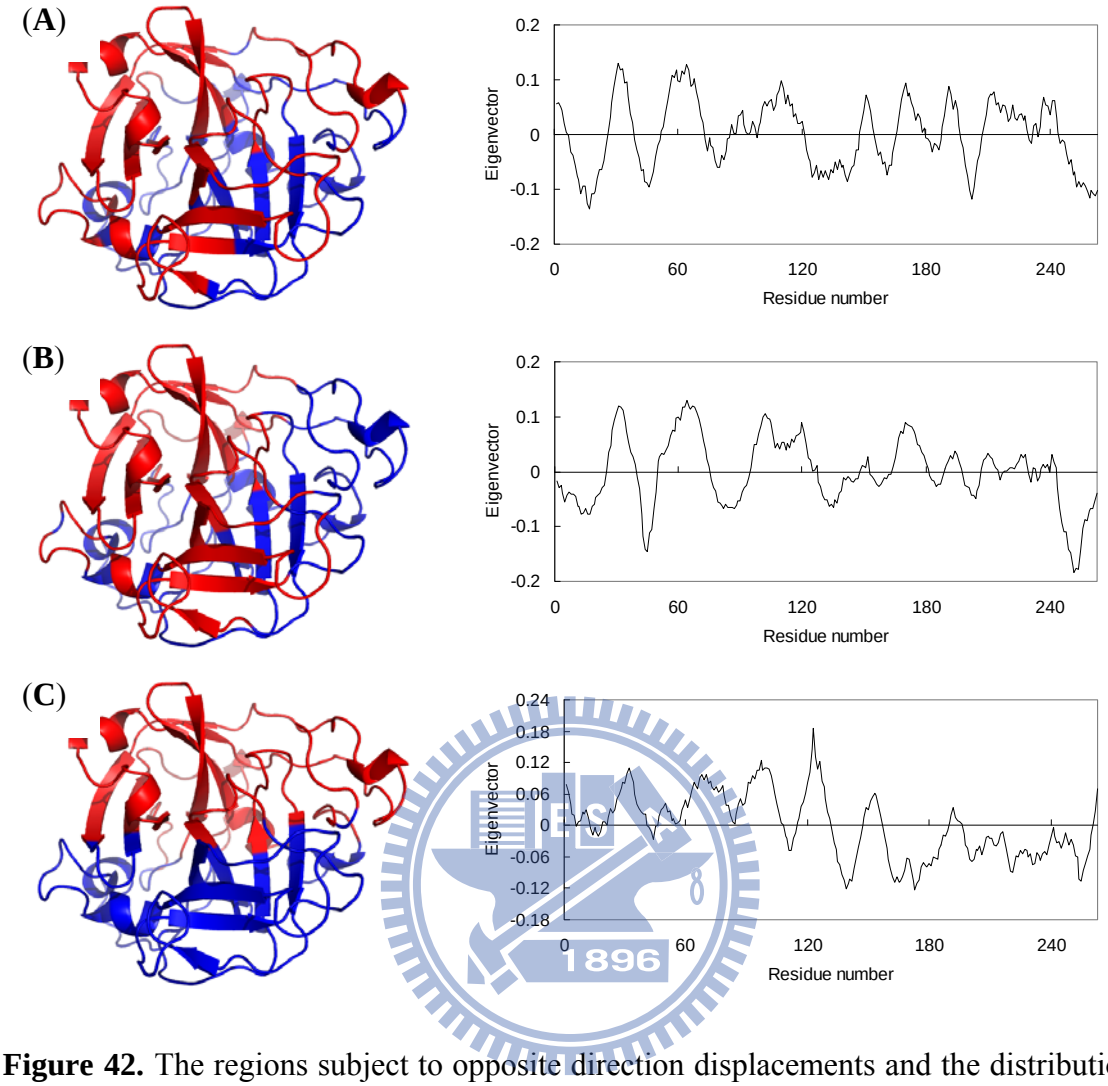


Figure 42. The regions subject to opposite direction displacements and the distribution of displacements along the third mode computed by (**A**) the WCN model for the achromobacter lyticus protease I (1ARB). Regions colored in blue and red correspond to positive and negative displacements respectively. The diagram shows the vibration amplitudes and phases (positive or negative). The same illustrations of figure and diagram show the results by (**B**) the GNM and (**C**) the NMA.

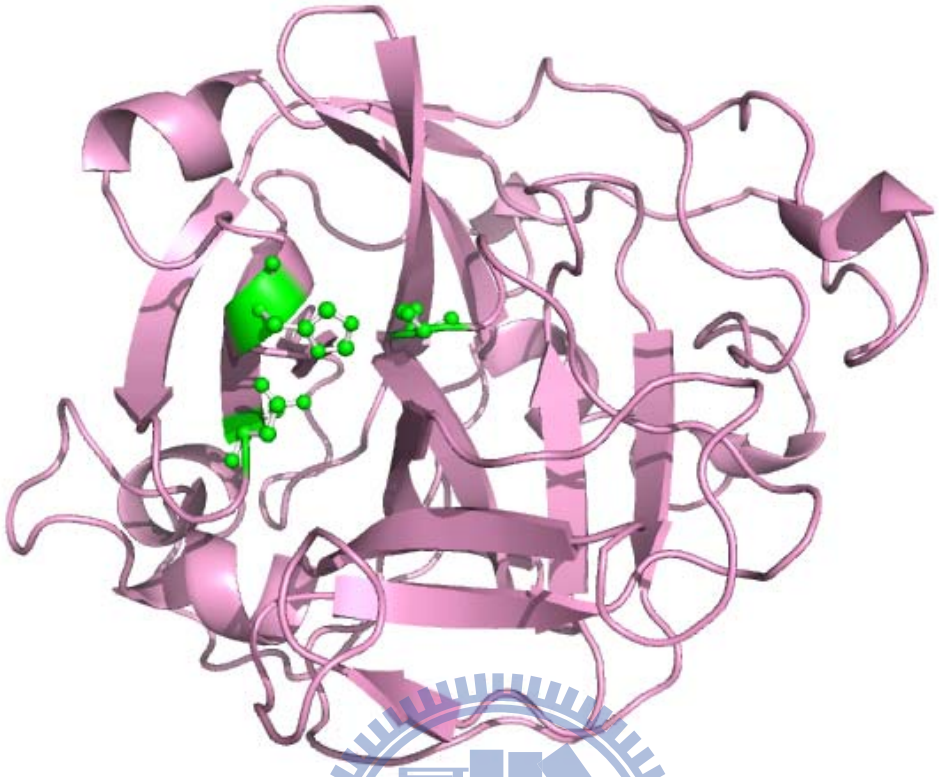


Figure 43. The structure of achromobacter lyticus protease I (1ARB). The catalytic triad comprises His57, Asp113, and Ser194 are colored in green and shown by the ball and stick model[35].

APPENDIX

The 972 protein chains of length ≥ 60 and their structures solved by X-ray crystallography with resolution ≤ 2.0 Å and R-factors ≤0.2 are selected from PDB-REPRDB[24].

1A1IA	1A53_	1A6M_	1A8D_	1A8I_	1A9XA	1ADEA	1ADOA
1AF7_	1AGJA	1AGQD	1AH7_	1AJSA	1AMM_	1AOCA	1AOP_
1APYA	1ARB_	1AW7A	1AY7B	1AYOA	1B3AA	1B5QA	1B65A
1B6TA	1B8EA	1BF2_	1BGF_	1BGVA	1BHTA	1BIF_	1BSLB
1BXEA	1C0PA	1C1KA	1C48A	1C5EA	1C7KA	1C7SA	1C9OA
1CB0A	1CC8A	1CCWA	1CCWB	1CCZA	1CFB_	1CHD_	1CMBA
1CQXA	1CQYA	1CRUB	1CSH_	1CV8_	1CVRA	1CZ9A	1CZFA
1D0DA	1D4OA	1D7PM	1D8DA	1DBFA	1DC1B	1DDT_	1DFMA
1DG6A	1DGWX	1DJ0A	1DJEA	1DJTA	1DLJA	1DMR_	1DOZA
1DQAA	1DQZA	1DS1A	1DUN_	1DUPA	1DXRM	1DY5A	1E1HA
1E4CP	1E6PB	1E6UA	1E7LA	1E9EA	1E9GB	1EB6A	1EBLA
1ECFB	1EDG_	1EDQA	1EEOA	1EEXA	1EH7A	1EJBA	1EJDA
1EKGA	1EKXA	1EL4A	1ELKA	1EPFB	1EQCA	1ES9A	1ESGB
1EU8A	1EUVA	1EX2A	1EXRA	1EXTA	1F1XA	1F20A	1F24A
1F4NA	1F86A	1F8EA	1FCQA	1FEHA	1FIUA	1FK5A	1FKMA
1FLTX	1FN9A	1FO8A	1FP3A	1FS7A	1FSGC	1FUPA	1G1TA
1G2BA	1G3P_	1G60A	1G61A	1G66A	1G8AA	1G8KA	1G9GA
1GBS_	1GCQC	1GCVB	1GD0A	1GK8I	1GK9A	1GK9B	1GKPA
1GMXA	1GNLA	1GNUA	1GOF_	1GQIA	1GQYB	1GTED	1GUIA
1GUQA	1GVKB	1GWEA	1GWMA	1GX5A	1GXMB	1GXUA	1H16A

1H1IB	1H2CA	1H32A	1H4GB	1H4YA	1H6FB	1H6KC	1H6KX
1HBNA	1HBNB	1HBNC	1HDKA	1HDOA	1HF8A	1HFES	1HG7A
1HH8A	1HP1A	1HPI_	1HQSA	1HS6A	1HT6A	1HYOB	1HZ4A
1HZ5B	1HZTA	1I19A	1I1DD	1I1NA	1I2TA	1I4UA	1I6TA
1I7QB	1I8OA	1I9ZA	1IAB_	1IC6A	1IDPA	1IFC_	1IIBA
1IPCA	1IQZA	1ITUA	1IU8A	1IUQA	1IV8A	1IV9A	1IXBA
1J0HA	1J0PA	1J2RA	1J34A	1J79B	1J8BA	1JAKA	1JBEA
1JD5A	1JDW_	1JEVA	1JG9A	1JIXA	1JM1A	1JNDA	1JNRA
1JPC_	1JPUA	1JRLA	1JU2A	1JUBA	1JZ7A	1JZTA	1K0EA
1K0MB	1K12A	1K3YA	1K4IA	1K55A	1K6ZA	1K7CA	1K7HA
1KAPP	1KBLA	1KD0A	1KEIA	1KG2A	1KHBA	1KHIA	1KJQB
1KNLA	1KOE_	1KPHB	1KQEA	1KQFB	1KQFC	1KQPA	1KS8A
1KT7A	1KUFA	1KV7A	1KVEA	1KWGA	1KWNA	1KYFA	1KZKB
1KZQA	1L2HA	1L3KA	1L6RA	1L7AA	1L8AA	1L9LA	1LAM_
1LATB	1LFWA	1LJ8A	1LK2A	1LK2B	1LKI_	1LKKA	1LL2A
1LLFA	1LML_	1LNIB	1LOVA	1LQVB	1LTM_	1LTSA	1LTZA
1LV7A	1LWBA	1LY2A	1LYVA	1LZJA	1M0KA	1M1NA	1M1NB
1M2DA	1M2XA	1M3KA	1M4IB	1M55A	1M65A	1M6JA	1M7YA
1M9XC	1M9ZA	1ME3A	1MG7B	1MIXA	1MJUL	1MK0A	1MKAA
1MKKA	1MN8D	1MOOA	1MPXA	1MQDA	1MQKH	1MRP_	1MTYB
1MTYD	1MUWA	1MXRA	1N0WA	1N13B	1N1BB	1N45A	1N62B
1N7SA	1N7SC	1N83A	1NC5A	1NKGA	1NKIA	1NLNA	1NOFA
1NOX_	1NQEA	1NQJA	1NSUB	1NTYA	1NU0A	1NV0A	1NVMG
1NWAA	1NWZA	1NYCA	1NYMA	1NYTA	1O08A	1O29A	1O4YA
1O6VA	1O7IA	1O7NB	1O83A	1O8XA	1O98A	1OAOC	1OBBB

1ODNA	1OE4A	1OEN_	1OEWA	1OFDA	1OFLA	1OFWA	1OGQA
1OGSA	1OI6B	1OI7A	1OJJA	1OJRA	1OK0A	1OLRA	1ON9D
1OOHA	1OOYB	1OR7C	1ORRA	1OWLA	1OX0A	1OZ2A	1P0KB
1P1JA	1P1MA	1P6OB	1PA7A	1PBJA	1PBYA	1PBYB	1PI1A
1PK6A	1PL3A	1PM1X	1PM4A	1PMHX	1PMI_	1PN0C	1POC_
1PSRB	1PSWA	1PT6B	1PV5A	1PVMB	1PWMA	1PX5B	1PXZA
1PYOC	1Q0NA	1Q0QA	1Q16A	1Q2OA	1Q33A	1Q40B	1Q63A
1Q6ZA	1Q7FB	1Q7LA	1Q7LB	1QB5D	1QF8B	1QFMA	1QFTA
1QGWB	1QGXA	1QH4A	1QH5A	1QHDA	1QHOA	1QIPA	1QKRB
1QKSA	1QMGA	1QNRA	1QOPB	1QOYA	1QR9A	1QSAA	1QTWA
1QUK_	1QV9A	1QW2A	1QW9A	1QWNA	1QWOA	1QWZA	1QX2A
1QXMA	1QXYA	1R0MA	1R1DA	1R29A	1R2QA	1R3DA	1R3SA
1R4PA	1R4XA	1R5LA	1R6JA	1R6XA	1R89A	1R8SA	1RA0A
1RA9_	1RC9A	1RCQA	1RG8A	1RGYA	1RHS_	1RIE_	1RJDC
1RKIA	1RKYA	1RLID	1RP0A	1RQHA	1RRO_	1RTQA	1RU4A
1RUTX	1RV9A	1RVAA	1RWHA	1RX0A	1RXQB	1RY9A	1RYAA
1RYIA	1S0AA	1S0IA	1S3EB	1S4BP	1S4KA	1S67L	1S7FA
1S7ZA	1S95B	1S99A	1S9RA	1SAUA	1SFSA	1SG4C	1SG6B
1SJWA	1SMBA	1SQEB	1SQSA	1SR4B	1ST0A	1SU8A	1SVB_
1SVFA	1SWXA	1T06A	1T0BH	1T0TV	1T1GA	1T1UA	1T2DA
1T46A	1T4BA	1T61D	1T6CA	1T6GA	1T7RA	1T92A	1T9HA
1TA3A	1TBFA	1TG5A	1TG7A	1TJYA	1TKEA	1TL2A	1TN6B
1TO2I	1TQ4A	1TQGA	1TT8A	1TU1A	1TU9A	1TUKA	1TWDB
1TXJA	1TXQB	1TZPA	1TZVA	1U07B	1U11B	1U3WA	1U5UA
1U69D	1U7GA	1U7IA	1U8FO	1U8VA	1U9DA	1UA4A	1UALA

1UASA	1UCDA	1UF5A	1UG6A	1UGHI	1UGNA	1UGPA	1UIRB
1UKUA	1UMGA	1UMKA	1UMZB	1UNNC	1UNQA	1UOHA	1UOWA
1UPGA	1UQ5A	1USCA	1UV4A	1UW1A	1UWCA	1UWFA	1UWKB
1UX6A	1UXZA	1UYLA	1UZKA	1V0EA	1V0LA	1V0WA	1V3EA
1V54A	1V54B	1V5FA	1V5IB	1V5VA	1V6PA	1V70A	1V7WA
1V82A	1VAJA	1VBKA	1VBLA	1VCLA	1VFYA	1VH5A	1VIYA
1VKPA	1VL9A	1VLBA	1VLS_	1VPSB	1VR7A	1VYBA	1VYIA
1VYKA	1VYRA	1VZIA	1W0HA	1W0NA	1W0OA	1W27A	1W2FA
1W2YA	1W4RA	1W5FA	1W66A	1W6GA	1W7LA	1W8OA	1W94A
1W96C	1WAKA	1WAPA	1WB4A	1WC2A	1WC3B	1WD3A	1WDCA
1WDDA	1WDPA	1WHI_	1WKQA	1WLDA	1WM3A	1WOFA	1WOYA
1WPNA	1WQ3A	1WRIA	1WS8A	1WT4B	1WU4A	1WUAA	1WUIL
1WUIS	1WV3A	1WVFA	1WWCA	1WY1A	1WYBA	1WYXB	1WZAA
1WZZA	1X09A	1X0CA	1X0JA	1X0RA	1X1NA	1X2JA	1X54A
1X6IB	1X6VA	1X82A	1XCLA	1XDNA	1XDZA	1XEOA	1XER_
1XFFA	1XFIA	1XG4A	1XGKA	1XH8A	1XHDA	1XJJA	1XKPB
1XKPC	1XOVA	1XQHA	1XQOA	1XSZA	1XTTA	1XUBA	1XWWA
1XZZA	1Y0EA	1Y0PA	1Y2TA	1Y3NA	1Y43B	1Y5IB	1Y5IC
1Y63A	1Y7BA	1Y8AA	1Y93A	1Y9GA	1Y9WA	1YB6A	1YDIA
1YFQA	1YGE_	1YGTA	1YHLA	1YI9A	1YIIA	1YJ1C	1YKDA
1YKUA	1YMIA	1YMTA	1YN9A	1YNPA	1YO3A	1YPHC	1YPHE
1YPQB	1YQZA	1YRKA	1YS1X	1YT3A	1YTBA	1YU8X	1YVIA
1Z05A	1Z0WA	1Z10A	1Z1SA	1Z2NX	1Z32X	1Z7XW	1Z84B
1ZCEA	1ZCJA	1ZCXA	1ZI9A	1ZJYA	1ZKPA	1ZL0B	1ZNDA
1ZO4B	1ZR6A	1ZUWC	1ZY7A	1ZZWA	2A14A	2A50A	2A50B

2A65A	2A6ZA	2AB0A	2AC7A	2ACFB	2ACVA	2AD6A	2AD6B
2AE0X	2AENB	2AEXA	2AFWA	2AGKA	2AGYD	2AHFA	2AIBA
2AIJX	2AJCA	2AKAA	2APXA	2AQ2B	2AQ5A	2AQ6A	2AQJA
2ARPF	2ARRA	2AUWB	2AVDA	2AWGA	2AWKA	2AXQA	2AXWA
2AYH_	2B06A	2B0TA	2B3FA	2B4HA	2B58A	2B5HA	2B61A
2B6DA	2B82A	2B97A	2BCGG	2BEMA	2BF5A	2BF6A	2BFDA
2BFDB	2BG1A	2BHUA	2BIBA	2BIIA	2BJFA	2BJKA	2BJRA
2BKFA	2BKXA	2BMOA	2BMWA	2BO9B	2BOGX	2BOPA	2BOQA
2BPTA	2BR6A	2BRAA	2BRFA	2BSWA	2BSYA	2BT9A	2BW3B
2BW4A	2BWQA	2BWVA	2BZUA	2C0NA	2C15A	2C1IA	2C1LA
2C1VA	2C2UA	2C3MA	2C4IA	2C4XA	2C5AA	2C6QB	2C71A
2C78A	2C9VA	2CARA	2CB2A	2CB5B	2CCAA	2CDBA	2CFUA
2CGLA	2CI1A	2CITA	2CIWA	2CJTC	2CK3D	2CK3G	2CKLA
2CKLB	2CL3A	2CN3B	2CNQA	2CTC_	2CVCA	2CVIA	2CWGA
2CXAA	2CXNA	2CXXC	2CYGA	2CZ1B	2D0OA	2DBBB	2DDSA
2DECA	2DKOB	2DQ6A	2ETGA	2EUTA	2EXVC	2F01B	2F2HA
2F2QA	2F4MA	2F4MB	2F5VA	2F5XB	2F6UA	2FA8C	2FBAA
2FBQA	2FE8A	2FFCA	2FFUA	2FH1B	2FHA_	2FHFA	2FHZA
2FIMB	2FL7A	2FP7B	2FPEA	2FRGP	2FSAA	2FSQA	2FSRA
2FWGA	2FY7A	2FYGA	2FYQA	2FZVB	2G29A	2G2WB	2G7CB
2G7OA	2G8OB	2GAGA	2GAGD	2GAIA	2GAKA	2GBAA	2GDQA
2GFOA	2GK4B	2GKEA	2GRHA	2GRRA	2GRRB	2GS5A	2GSOA
2GUDB	2H29A	2H6NB	2H7GX	2H88A	2H88D	2HALA	2HFT_
2HTS_	2IU1A	2IU4B	2IU5A	2IUWA	2IWAA	2IXMA	2KINA
2LISA	2MHR_	2NACA	2PGD_	2PTD_	2SQCA	2TGI_	3CHBF

3GRS_	3VUB_	4EUGA	4LZT_	4UBPC	7AHLB	7ATJA	7FABH
7FD1A	8A3HA	8ACN_	9GAFC