國立交通大學

生物資訊研究所

碩 士 論 文

利用彈性網路分子動力學研究蛋白質動態

The study of protein dynamics using Elastic Network Molecular Dynamics

研 究 生：顏士中

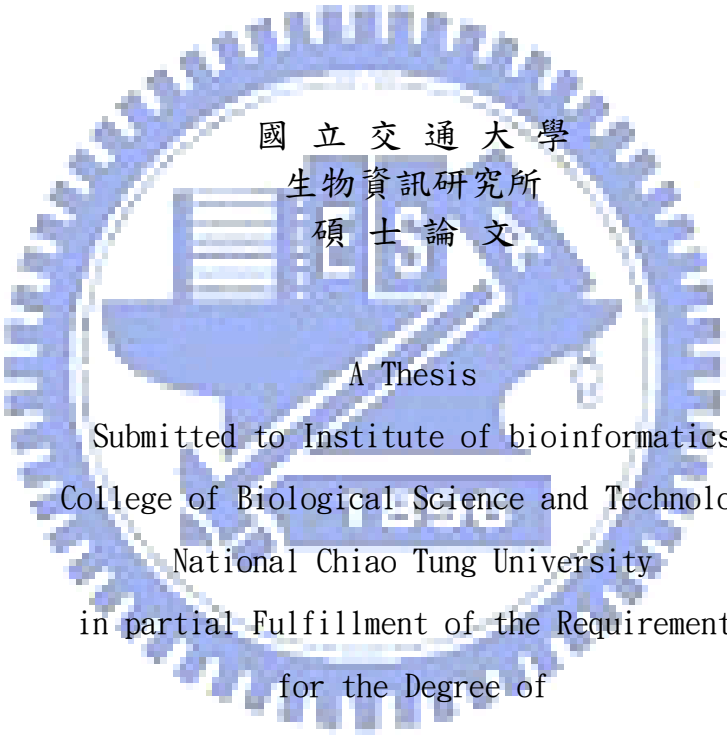
指導教授：黃鎮剛　教授

中 華 民 國 九 十 六 年 七 月

利用彈性網路分子動力學研究蛋白質動態
The study of protein dynamics using Elastic Network Molecular Dynamics

研 究 生：顏士中　　Student：Shih-Chung Yen

指導教授：黃鎮剛　　Advisor：Jenn-Kang Hwang



國 立 交 通 大 學

生物資訊研究所

碩 士 論 文

A Thesis

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

Master

in

Bioinformatics

July 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年七月

國　立　交　通　大　學

博碩士論文全文電子檔著作權授權書

（提供授權人裝訂於紙本論文書名頁之次頁用）

本授權書所授權之學位論文，為本人於國立交通大學 生物資訊研究所 ， 95 學年度第 2 學期取得碩士學位之論文。

論文題目：利用彈性網路分子動力學研究蛋白質動態

指導教授：黃鎮剛

■ 同意　□不同意

本人茲將本著作，以非專屬、無償授權國立交通大學與台灣聯合大學系統圖書館：基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學及台灣聯合大學系統圖書館得不限地域、時間與次數，以紙本、光碟或數位化等各種方法收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行線上檢索、閱覽、下載或列印。

論文全文上載網路公開之範圍及時間：	
本校及台灣聯合大學系統區域網路	■ 中華民國　年　月　日公開
校外網際網路	■ 中華民國　年　月　日公開

授 權 人：顏士中

親筆簽名：_____

中華民國　96　年　7　月　16　日

國　立　交　通　大　學

博碩士紙本論文著作權授權書

（提供授權人裝訂於全文電子檔授權書之次頁用）

本授權書所授權之學位論文，為本人於國立交通大學 生物資訊研究所 ， 95 學年度第 2 學期取得碩士學位之論文。

論文題目：利用彈性網路分子動力學研究蛋白質動態

指導教授：黃鎮剛

■ 同意

本人茲將本著作，以非專屬、無償授權國立交通大學，基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學圖書館得以紙本收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行閱覽或列印。

本論文為本人向經濟部智慧局申請專利(未申請者本條款請不予理會)的附件之一，申請文號為：＿＿＿＿＿＿＿＿＿＿＿＿，請將論文延至＿＿＿年＿＿＿月＿＿＿日再公開。

授權人：顏士中

親筆簽名：＿＿＿＿＿＿＿＿

民國 96 年 7 月 16 日

國家圖書館博碩士論文電子檔案上網授權書

ID:GT009451507

本授權書所授權之學位論文，為本人於國立交通大學 生物資訊研究所 ， 95 學年度第 2 學期取得碩士學位之論文。

論文題目：利用彈性網路分子動力學研究蛋白質動態

指導教授：黃鎮剛

茲同意將授權人擁有著作權之上列論文全文（含摘要），非專屬、無償授權國家圖書館，不限地域、時間與次數，以微縮、光碟或其他各種數位化方式將上列論文重製，並得將數位化之上列論文及論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

※ 讀者基於非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

授權人：顏士中

親筆簽名：_____

民國 96 年 7 月 16 日

1. 本授權書請以黑筆撰寫，並列印二份，其中一份影印裝訂於附錄三之二(博碩士紙本論文著作權授權書)之次頁；另一份於辦理離校時繳交給系所助理，由圖書館彙總寄交國家圖書館。

利用彈性網路分子動力學研究蛋白質動態

學生：顏士中　　指導教授：黃鎮剛

國立交通大學生物資訊研究所碩士班

摘　　要

在結構計算生物學上，蛋白質的動態一直與其功能有密切的關係。而在實驗上 X 光結晶繞射所得的溫度因數與核磁共振所得到的蛋白質動態多重模型也告知我們蛋白質並非靜態的。在過去，我們利用分子動力學來研究蛋白質動態，但是分子動力學的時間複雜度太過龐大且須巨大的計算量。在這篇論文中，首先我們試著建立一個粗略的模型來簡化分子動力學中過多的力學因子與及原子個數來計算分子的動態軌跡然後我們將其的振動來跟 X 光結晶繞射所得的溫度因數作比較。再來我們試著考慮各參數對這個模型的影響。並針對一個資料集合做整體的測試並分析其少數差距較大的蛋白質。最後我們證實蛋白質骨架的振動其實與其胺基酸種類並無太大的關係，而是由結構體決定其振動大小。

The study of protein dynamics using Elastic Network Molecular Dynamics

student：Shih-Chung Yen　　Advisors：Dr. Jenn-Kang Hwang

Institute of Bioinformatics
National Chiao Tung University

ABSTRACT

In computational structural biology, protein motion has relationships about function. Some experimental evidence such as temperature factor (B factor) or NMR multi-structure also shows that protein is not static. In traditionally, the molecular dynamics is useful to studying the protein motion, but there are so huge time complexity and computational scale on MD. In this article, first, we try to generate a coarse-grained model to simplify the force characteristics and simulate the trajectory of protein dynamics. Then we compare the result of fluctuation with B-factor from X-ray crystallography protein structure. Then we try to change the parameter to test the model and calculate a whole dataset. At last, we found that the protein fluctuation is decided by the structure.

誌謝

曾經有人跟我說過誌謝文只要隨便寫寫就好，但是我想，一篇論文最認真的地方應該就是誌謝文，尤其是只有這個才是真正屬於你自己的東西的時候。話說如果是六年前我的腦袋還不知道生物資訊的存在，如果是四年前，我對生物的了解大概只有分子生物學的中心法則，如果是三年前，我想我根本不奢望我會在這間實驗室。如果我這一生有做對甚麼選擇的話，那就是我來到了這裡。

如果說我這兩年來最感謝的是誰？我的指導教授黃鎮剛老師絕對是第一中的第一。對我而言他不只是一個教授，他是一個貴人，一個好友的存在。他對我的疼愛、付出、指導，絕對不是一篇簡單的誌謝文可以說盡的。他開啓我的眼界與對學術的認知，指引我方向。有人說一個好的老師不但交導你知識，還會教你如何做事，甚至如何做人。這一點黃老師做到而且還遠遠的超過。

除了老師外學長姐的幫忙與指教也是千言萬語道不盡的謝意。尤禎祥學長對機器的維護，玉菁學姐與景盛學長的關心(祝他們順利畢業)。還有草霸的幫忙，稅制的技術指導，以及建華、小操、少偉和蔚倫諸多的指教。

至於我的同輩，要感謝我的房東祥逢，啓文，還有兩位可愛的學妹整天讓我欺負逗我開心。當然更要感謝其他實驗是同學的幫助。尤其是楊老師實驗室的同學們。

最後就是我的父母，畢竟金主最大，沒有他們的資助與容忍我是不可能完成學業的。他們理應接受我所有的榮耀與上天的恩賜。最後要感謝的太多了，就感謝天吧。

Table of Contents

Chapter 1. Introduction

1.1 The importance of protein dynamics

Protein structure is dynamical and we can get other experimental evidence such as B-factor in X-ray structure and NMR structure.

The protein dynamics has relationship to protein function. This is consistent with the recent study[1], the catalytic residues have significantly lower positional fluctuations than other non-catalytic residues. Apparently, enzymatic activity is associated with the low translational mobility of the catalytic residues, which helps maintain the fine-tuned catalytic architecture. Then Wei-Chun Chiu[2] also use molecular dynamics to analyze the *N*-Acylamino acid racemase (NAAAR) and *N*-carbamoyl-d-amino-acid amidohydrolase (d-NCAase) and discover that the catalytic site are more stable than other residues.

Comparatively, the fluctuation of protein hinge region tends to move faster than others. This is support by Peter M. Jones[3] in Q-loop of ABC transporter. The structural diversity and generally high crystallographic temperature factors for atoms within the Q-loop suggest that it is flexible and may undergo conformational changes during the catalytic cycle. For detecting the protein hinge motivation, molecular simulation is also to be used to observe the atoms fluctuation[4].

1.2 The experimental evidence

In experiment, the evidence of protein motion are B-factor in X-ray structure and the multi-model structure from NMR.

The B-factor also called temperature factor. B-factor describes the thermal fluctuations of heavy atoms in the x-ray structure, its formula is

$$\left\langle (\mathbf{x}-\mathbf{x}_0)\bullet(\mathbf{x}-\mathbf{x}_0)\right\rangle = \frac{3}{8\pi^2}B$$

In figure 1, we can discover that the B-factor can reflect the fluctuation, the red means the fluctuation is large and blue means less. If the residues that are closed to the surface, the fluctuation would be large.

Besides B-factor of structure in X-ray crystallography, NMR (see Figure 2.) can get the protein large scales motion information. So we get a multi-models protein structure. Molecular dynamics can be used to calculate the motivation of proteins.

1.3 Molecular dynamics

For getting the information of protein motion, molecular dynamics (MD) is a well-known tool to calculate the trajectory of protein dynamics form of computer simulation. The atoms and molecules are allowed to interact for a period of time under known laws of physics. Because in general molecular systems consist of a large number of particles, it is impossible to find the properties of such complex systems analytically. MD simulation circumvents this problem by using numerical methods. It represents an interface between laboratory experiments and theory and can be understood as a virtual experiment. MD gained popularity in biochemistry and biophysics. In chemistry, MD serves as an important tool in protein structure determination and refinement using experimental tools such as X-ray crystallography and NMR. It has also been applied with limited success as a method of refining protein structure predictions. It is the physical principles of MD. One of the principal tools in the theoretical study of biological molecules is the method of molecular dynamics.

This computational method calculates the time dependent behavior of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. They are also used in the determination of structures from x-ray crystallography and from NMR experiments.

In computational structure biology, it is an important tool to study the protein and Nucleic acids molecular structure. In fact, because it is impossible that we can see how water molecular try to across the aquaporin[5]. We only can get the information by simulation the molecular trajectories. For getting the trajectories, we calculate the potential energy, and try to get the motion. The total potential energy of any molecule is the sum of terms allowing for bond stretching, bond angle bending, bond twisting, van der Waals interactions and electrostatics. Many properties of a biomolecular canbe simulated with such an empirical energy function.

1.4 Force field

In the context of molecular mechanics, a force field (also called a force field) refers to the functional form and parameter sets used to describe the potential energy of a system of particles (typically but not necessarily atoms). Force field functions and parameter sets are

derived from both experimental work and high-level quantum mechanical calculations. In fact, force field is just to define the environment in computer. So all force fields are based on numerous approximations and derived from different types of experimental data. Therefore they are called empirical. "All-atom" force fields provide parameters for every atom in a system, including hydrogen, while "united-atom" force fields treat the hydrogen and carbon atoms in methyl and methylene groups as a single interaction center. Amber[6] and GROMACS[7] are famous classical force field.

From before potential function, we can define each kind of energy in general molecular dynamics force field. The bond and angle terms are usually modeled as harmonic oscillators in force fields that do not allow bond breaking. The functional form is highly variable. It also include potentials for hydrogen bonds, another important torsion term to account for the planarity of aromatic rings and other conjugated systems, and "cross-terms" that describe coupling of different internal variables, such as dihedral angles and bond lengths. Nonbonding force is the main point to make the huge time complexity. These terms are most computationally intensive because they include many more pair wise interactions per atom. So, every step, we have to run the loop of each atom-pairs. If there are 3000 atoms in a protein, it will make 9000000 atoms pairs. Therefore, we always set a distance as cutoff, if the distance of a atom pair is over the cutoff. Because the force is too small, we will reduce the nonbonding force. In nonbond terms, the van der Waals term is usually computed with a Lennard-Jones potential and the electrostatic term with Coulomb's law, although both can be buffered or scaled by a constant factor to produce better agreement with experimental observation.

1.5 Elastic network molecular dynamics (EMD)

MD simulations on very large systems may require such large computer resources that they cannot easily be studied by traditional all-atom methods. The most disadvantage of MD is extremely computational intensive. When we want to simulate the macromolecular such as virus coat or molecular chaperon. It will take huge computational ability and time[8]. For this reason, many people try to build coarse-grained model to simplify the feature to solve the limit of running macromolecular[9] or use anisotropic normal mode analysis[10]. The coarse-grained model is like lattice model[11]. Instead of all complexity representing every atom or force in system, define the atoms as a kind of “pseudo-atoms” and take the force field only two kinds of force, the hydrophobic and hydrophilic. Therefore, the time complexity will

be reduced.

There are two examples for coarse-grained model, the discontinuous molecular dynamics (CG-DMD)[12] and Go-models[13]. We build a new method to calculate the fluctuation of proteins. We try to build a simplified structure model to run molecular dynamics for calculating the proteins fluctuations. We disable all non-bond force and connect all atoms in a cutoff by a simplified spring. Then we use molecular dynamics to simulate the movement of this network structure. Finally we can get the proteins fluctuation information and compare with the experimental temperature factor.

Chapter 2. Methods

The disadvantage of Molecular dynamics is that MD is extremely computational intensive. For example, non-bond force computation cost too much time complexity and space complexity, and we must add water molecular to simulate the real proteins environment. So we generate a elastic model to simplified some characters. We get only alpha carbon atoms and disable all non-bond force. Second, connect all atoms by a simplified spring. Third, use molecular dynamics to calculate the motivation. Forth, calculate the fluctuation of each atom. At last, compare with the experimental data.

2.2 Molecular dynamics and the elastic model

MD refers to the use of classical mechanics/Newtonian mechanics to describe the physical basis behind the models. Molecular models typically describe atoms (nucleus and electrons collectively) as point charges with an associated mass. The interactions between neighboring atoms are described by spring-like interactions (representing chemical bonds) and van der Waals forces[14]. In figure 3, the equation of potential energy and trajectory derives form the Newton's second law. In proteins, the total potential energy function includes energy from internal and external energy terms.

The internal energy terms describe the energy associated with changes in bond lengths, bond angles, and torsion angles. On the other hand, the external energy terms include salt-bridges, hydrogen bonds, and van der Waals interactions between atoms[15]. The empirical potential function reflects all the energy of the given protein structure. For example, (in figure 4) there are just five atoms but we have to calculated 6 non-bond force, 4 bond force, 2 torsion angle and 4 angles. Therefore, the time complexity has become an important

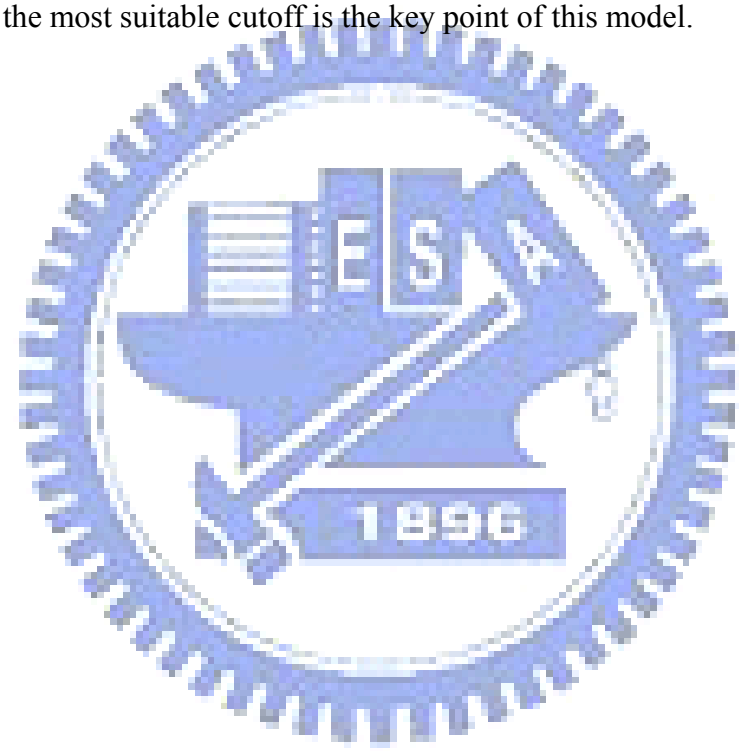
In our method, we try to keep the spring-like interaction and ignore the other force such as van der Waals interaction (non-bonded pairs), torsion angles, partial charges and bond angle bending force (in figure 5). So our total potential energy only considers all bonds force.

$$U = \sum \frac{1}{2} K_b (r - r_0)^2$$

Where *U* is the potential energy, *Kb* is the spring force constant (Hooken force constant). Besides this, we take only alpha-carbon of protein backbones (in figure 5).

In elastic network models (ENMs), the system is represented by a network of beads connected by elastic springs usually one bead per amino acid (although elastic networks have also been used together with all-atom descriptions, but in our method, we only take the C-alpha as the target). The extreme simplicity of the parameterization is balanced by the need to know the equilibrium reference configuration, from which only harmonic fluctuations are possible.

In figure 6, we try to generate a picture of elastic network model of 1crn. We can see that each CA is connected several bonding forces. The number of bonding force is decided by the cutoff. To set the most suitable cutoff is the key point of this model.



Chapter 3. Results

We generate this new methods, but there are three parameters can be discussed, the cutoff, force constant, and temperature. It is very interesting that some features are not like our anticipate results.

For compare these two data candidly, we transfer these data to z-scores and calculate the Pearson's correlation coefficient between the root mean square fluctuation (RMSF) from our method and the B-factor.

3.1 Effects of the cutoff on RMSF

First we use 1crn(see figure 7) and 5pti(see figure 8) as the example and set temperature is 500K, force constant is 1 and 1000000 steps. We can see that when the cutoff is lesser, the correlation is closer to the experimental results. In fact, the less cutoff means that the force between C-alpha to C-alpha become more and is close to the real environment. But when the cutoff is too less, the numbers of bond would decrease, in our tests, the cutoff may be between 4 to 6.

5pti[16] is bovine pancreatic trypsin inhibitor, also called BPTI, is a protein found in many tissues throughout the body. BPTI inhibits several of the serine protease proteins such as trypsin, kallikrein, chymotrypsin, and plasmin. BPTI is a member of the pancreatic trypsin inhibitor family, which is a family of serine protease inhibitors. These proteins usually have conserved cysteine residues that participate in forming disulfide bonds.

At last we try another 6 protein as example and get the same conclusion. Especially the main point is when the cutoff is too large, it will cover all the protein and make the number of bond force in every atom be the same one. So it makes the fluctuation calculated by our model almost to be a linear. (see figure 9 to figure 14)

3.2 Effects of the force constant on RMSF

In this work, we try to change the force constant in our model. It is interesting that we discover that the force constant is almost no influence to the results.

In figure 15, we added the force constant from 1 to 1000, and we can see that the

distribution of fluctuation is almost no transformation. The correlation has no changes.

In this experiment, we take 1aba[17](a/b) (see figure 15.) and 2omf[18](all beta) (see figure 16.) for examples. 1aba is the oxidized bacteriophage T4 glutaredoxin. 2omf is an integral membrane protein located in the outer membrane of the bacteria, Escherichia coli. OmpF porin is a non-specific transport channel that allows for the passive diffusion of small, polar molecules (600-700 Dalton in size) through the cell's outer membrane.

3.3 Effects of the temperature on RMSF

Temperature is very important to protein structure. We all know that the protein would be denaturing when the temperature is too high. On molecular dynamics, we also use it in this model, because we connect each C-alpha by a simplified spring. The bonding force may be more stable than native protein. We try to change the temperature and discover that the temperature like force constant is almost no influence to the results. In fact, the protein move more quickly but the structure would not be crash. So the correlation coefficient also changes nothing.

In this experiment, we also take 1aba(a/b)(see figure 17) and 2omf[18](all beta) (see figure 18) for examples.

3.4 Comparing with real molecular dynamics by GROMACS

We also take some examples to run the fine-grained simulation and compare the results with our method. Because of the so large time complexity, we just try a few proteins. In this case, the examples are 2omf, 1qr9a, 1c9oa, 1ucda and 1itua.

On environment in GROMACS, we test 2 nano-seconds (1000000 steps), as time and added water in our simulation. For getting the results, it cost me 3-5 days per target. Especially the 2omf, it almost takes 5 days but our method only takes 3 minutes.

After comparing the results with our method, we can get better correlation coefficient and take less time scale. In figure 19 to 23 and table 1, we can discover that the root mean square fluctuation calculated by our model is very closed to the results by fine-grained molecular dynamic from GROMACS. In this test, we use GROMACS 3.3 edition to run two examples, 2omf and 1crn. GROMACS[19] was developed in Herman Berendsens group, department of Biophysical Chemistry of Groningen University.

3.5 Non-homologous datasets

After trying the parameter, we test a non-homologous dataset to calculate the distribution of correlation coefficient between the B-factor and RMSF. We get the results support that our method can work and better. The mean of correlation coefficient is 0.5326(in figure 24). The all time scales take only 5 day, but a real molecular dynamics simulation by GROMACS may need 1000 folds of time complexity. In before example, the real molecular dynamics of 2omf take 5 days to get the result at the same situation, but our method takes only 2 minutes and 36 seconds. Of course reducing the atoms and nonbonding force may be the most important reason to lower the time complexity.

Our data is selected by the next condition. We selected from PDB-REPRDB 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%. (See in Table 2)

3.6 Discussion of failure cases

3.6.1 The protein of multiple chains

First, the datasets that we analyzed is separated proteins by chain. In fact, some proteins are working together and connect to each other to form a oligomer. We have to combine this identical chain to other subunit for getting the real fluctuation. In this case, we show the 1kqf, c-chain as the example. When we only calculate the c-chain, the RMSF is far away from the experimental situation. When we take the whole protein to build a elastic network model. The correlation coefficient between RMSF and B-factor tend to be match together. (see in figure 25)

3.6.2 The missing residue

We discover that there are several errors in PDB. For example, the 2fwg has 10 missing residue at the last tail which connected to the last 5 residues. Therefore it makes the results to be bad. When we cut the last 5 Histidine of the C-terminal, The correlation coefficient is updated from -0.07 to 0.52. (See in figure 26)

3.6.3 Biological unit or asymmetrical unit

In PDB, some researchers present the structure is unique. But in biological situation, they may connect the other subunit to become oligomer and work. In this case, we take 2bop as the example. 2bop is a DNA binding protein. When it works, it will connect to another one subunit to become a dimmer. So, we have to calculate the whole dimer's root mean square fluctuation. (see in figure 27)

Asymmetrical unit is situation of protein stacking in the in X-ray crystal. Because the B-factor is getting from the X-ray crystallography, the protein stack would affect the fluctuation of protein. When we calculate the RMSF, we have to take into consider the factor.

3.6.4 Several independent domains

Sometimes, one amino acid sequence tend to fold to several domains. When we face this problem, we try to separate the protein by domain knowledge and classification. Such as 1nty

and 2c4x. we can separate each protein as two domain by visualization check easily. For example, the 1nty can be separated to two independent domains. The correlation coefficient would be updated from 0.3 to 0.8. (see in figure 28 to 29)

3.6.5 PDB format error

At last, in fact, some researchers support the error data or uncompleted data to PDB. Such as 1ldd, there is no information about temperature factor. (See in figure 30)

Chapter 4. Conclusion

In this work, we develop a new coarse-grained model that combines molecular dynamics with elastic network model. According to our results, we found that our method can calculate the root mean square fluctuation better than MD, even we use less atoms and bonding force. It support us a new think of a way that the structure and the distance between atoms decide the backbone fluctuation. Although it is well-known that the side-chains traditionally decide the specificity in some cases, such as the trypsin and chymotrypsin, it has been suggested that the specificity of trypsin and chymotrypsin is decided by the structure[20].

Moreover, our method takes us around less than 1000-folds complexity to calculate the same results as MD does. In recent study, Rueda, M.[21] performed for all protein metafolds using the four most popular force fields (OPLS, CHARMM, AMBER, and GROMOS) to calculate motion and the RMSF. The time scale takes around 1.5 terabytes of data obtained using approximately 50 years of CPU. This paper shows that the difference of force fields would to make the same results. Our method also shows that even we use the simplest potential function, we can get the better results and faster.

Reference

1. Yang, L. W. & Bahar, I. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure 13, 893-904 (2005).
2. Chiu, W. C. et al. Structure-stability-activity relationship in covalently cross-linked N-carbamoyl D-amino acid amidohydrolase and N-acylamino acid racemase. J Mol Biol 359, 741-53 (2006).
3. Jones, P. M. & George, A. M. Mechanism of ABC transporters: a molecular dynamics simulation of a well characterized nucleotide-binding subunit. Proc Natl Acad Sci U S A 99, 12639-44 (2002).
4. van Aalten, D. M. et al. Engineering photocycle dynamics. Crystal structures and kinetics of three photoactive yellow protein hinge-bending mutants. J Biol Chem 277, 6463-8 (2002).
5. Page, E., Winterfield, J., Goings, G., Bastawrous, A. & Upshaw-Earley, J. Water channel proteins in rat cardiac myocyte caveolae: osmolarity-dependent reversible internalization. Am J Physiol 274, H1988-2000 (1998).
6. Case, D. A. et al. The Amber biomolecular simulation programs. J Comput Chem 26, 1668-88 (2005).
7. Van Der Spoel, D. et al. GROMACS: fast, flexible, and free. J Comput Chem 26, 1701-18 (2005).
8. Freddolino, P. L., Arkhipov, A. S., Larson, S. B., McPherson, A. & Schulten, K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. Structure 14, 437-49 (2006).
9. Arkhipov, A., Freddolino, P. L. & Schulten, K. Stability and dynamics of virus capsids described by coarse-grained modeling. Structure 14, 1767-77 (2006).

10. Poon, B. K. et al. Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-A crystallographic resolution. Proc Natl Acad Sci U S A 104, 7869-74 (2007).

11. Hinds, D. A. & Levitt, M. A lattice model for protein structure prediction at low resolution. Proc Natl Acad Sci U S A 89, 2536-40 (1992).

12. Ding, F., Borreguero, J. M., Buldyrey, S. V., Stanley, H. E. & Dokholyan, N. V. Mechanism for the alpha-helix to beta-hairpin transition. Proteins 53, 220-8 (2003).

13. Paci, E., Vendruscolo, M. & Karplus, M. Validity of Go models: comparison with a solvent-shielded empirical energy decomposition. Biophys J 83, 3032-8 (2002).

14. Levitt, M. The birth of computational structural biology. Nat Struct Biol 8, 392-3 (2001).

15. Michael Levitt , M. H., Ruth Sharon, Valerie Daggett. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. Computer Physics Communication 91, 215-231 (1995).

16. Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. J Mol Biol 180, 301-29 (1984).

17. Eklund, H. et al. Structure of oxidized bacteriophage T4 glutaredoxin (thioredoxin). Refinement of native and mutant proteins. J Mol Biol 228, 596-618 (1992).

18. Cowan, S. W. et al. The structure of OmpF porin in a tetragonal crystal form. Structure 3, 1041-50 (1995).

19. Kutzner, C. et al. Speeding up parallel GROMACS on high-latency networks. J Comput Chem 28, 2075-2084 (2007).

20. Ma, W., Tang, C. & Lai, L. Specificity of trypsin and chymotrypsin: loop-motion-controlled dynamic cor**relation as a determinant. Biophys J 89, 1183-93 (2005).**

21. Rueda, M. et al. A consensus view of protein dynamics. Proc Natl Acad Sci U S A 104, 796-801 (2007).

Table Contents

PBD	GROMACS	EMD
2omf	0.508	0.518
1ucda	0.555	0.678
1qr9a	0.758	0.884
1c9oa	0.385	0.542
1itua	0.655	0.710

Table 1. Comparison of the correlation coefficient of B-factor and RMSF that is calculated by GROMACS and EMD. The time scale of 2omf calculated by GROMACS is 5 days at one computer, but the EMD only take 2.5 minutes at the same computer.

priority	factor	constraint
1	Resolution	X > 2.0
2	R-factor	X > 0.2
3	Chain break	no
4	Ratio of non-standard amino acid residues	no
5	Ratio of residues with only CA coordinates	X > 0
6	Ratio of residues with only backbone coordinates	X > 0
7	Number of residues	X < 60
8	Mutant	include
9	Complex	include
10	Fragment	include
-	NMR	exclude
-	Membrane proteins	include

Table 2. PDB-REPRDB entries list sorting parameters. This is our non-homologous datasets

Figures Contents

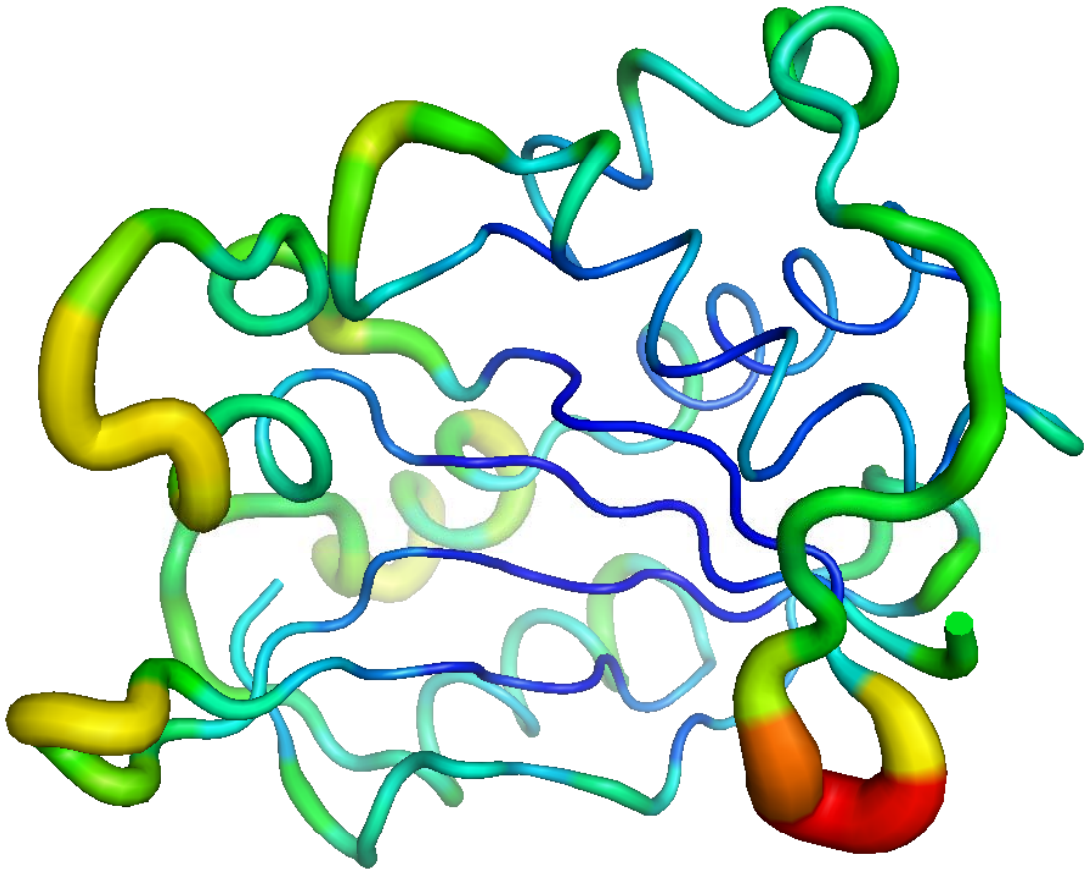


Figure 1. The B-factor showing picture of putty mode by pymol.

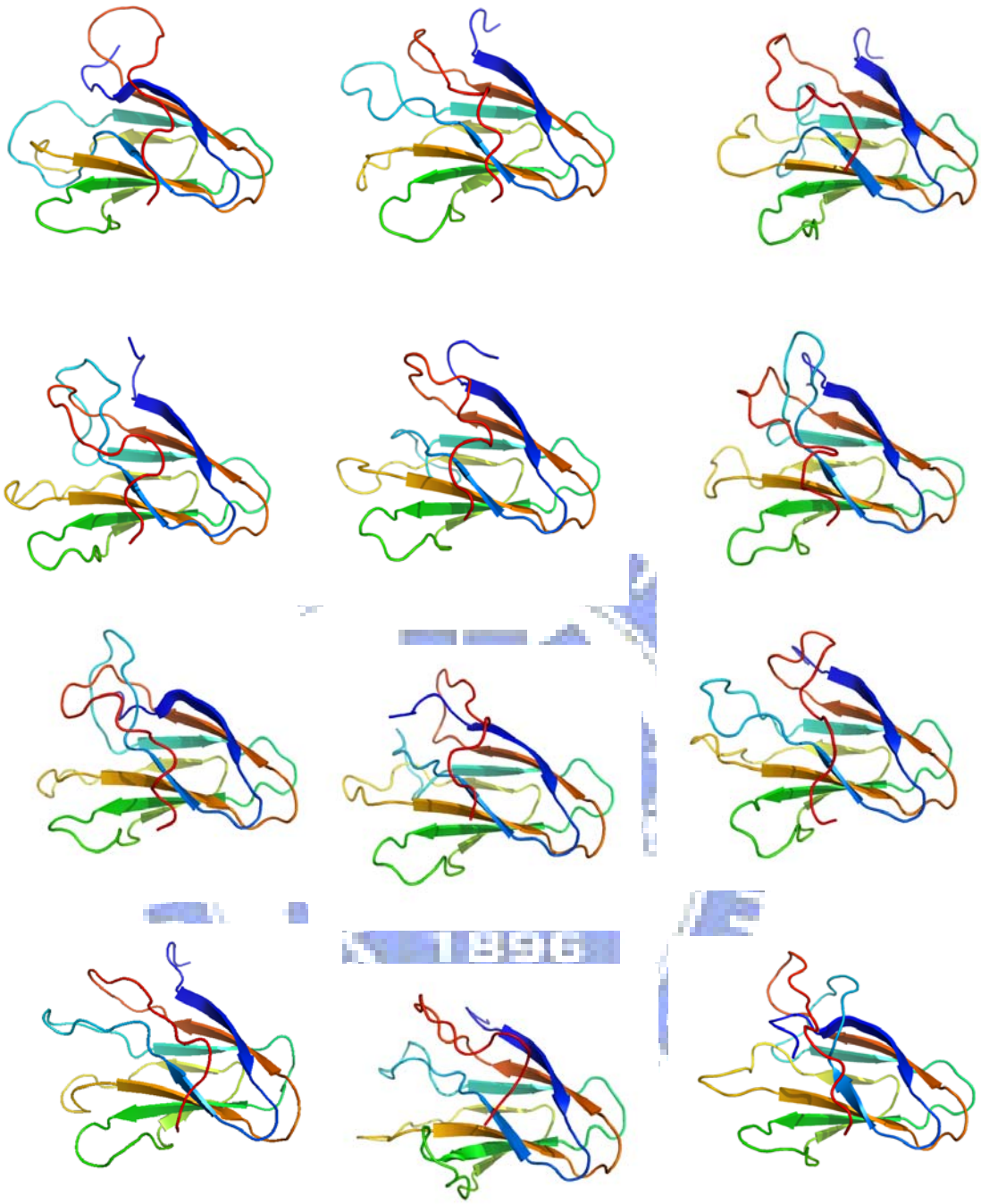


Figure 2. NMR structure of 1poq

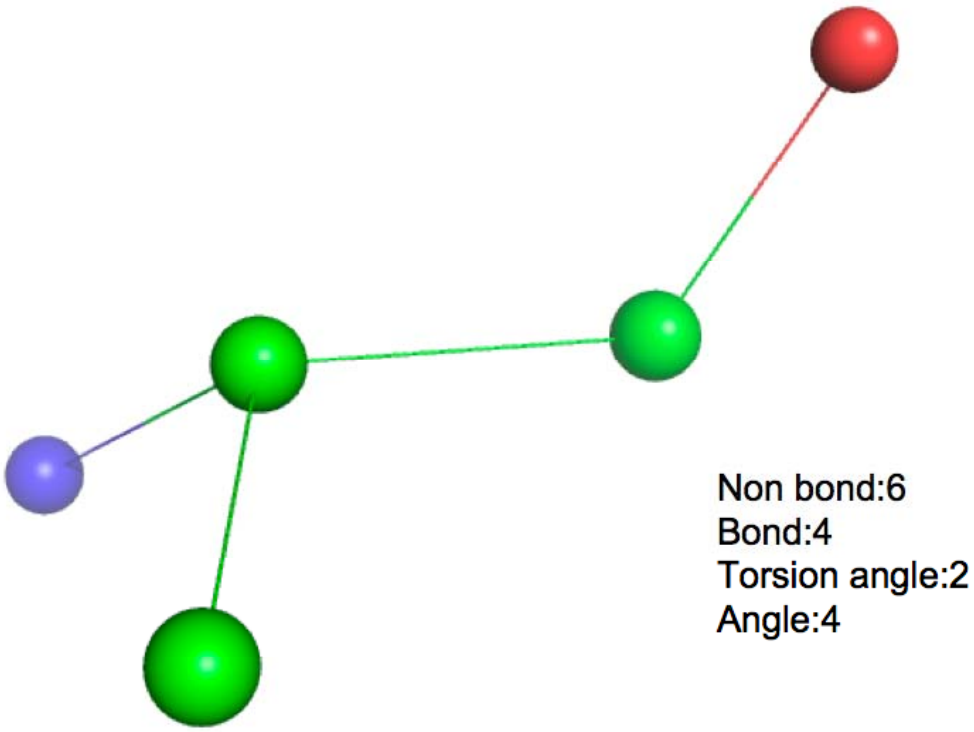


Figure 3. The numbers of potential energy of five atoms

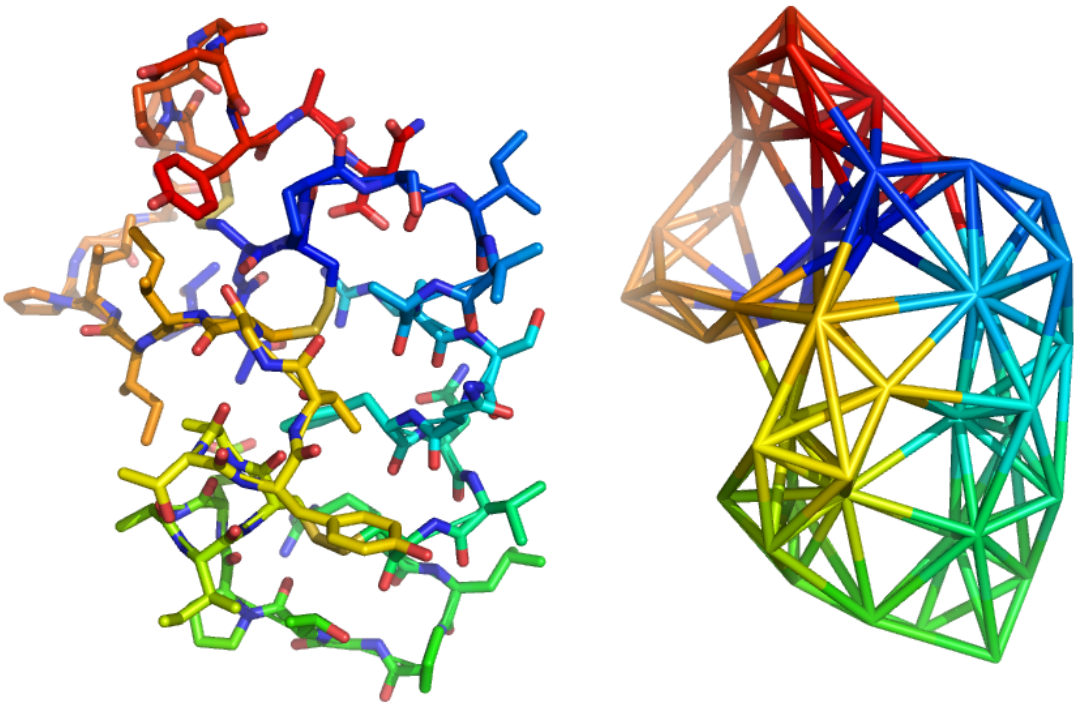


Figure 4. The elastic network model of 1crn, the cutoff is 7A.

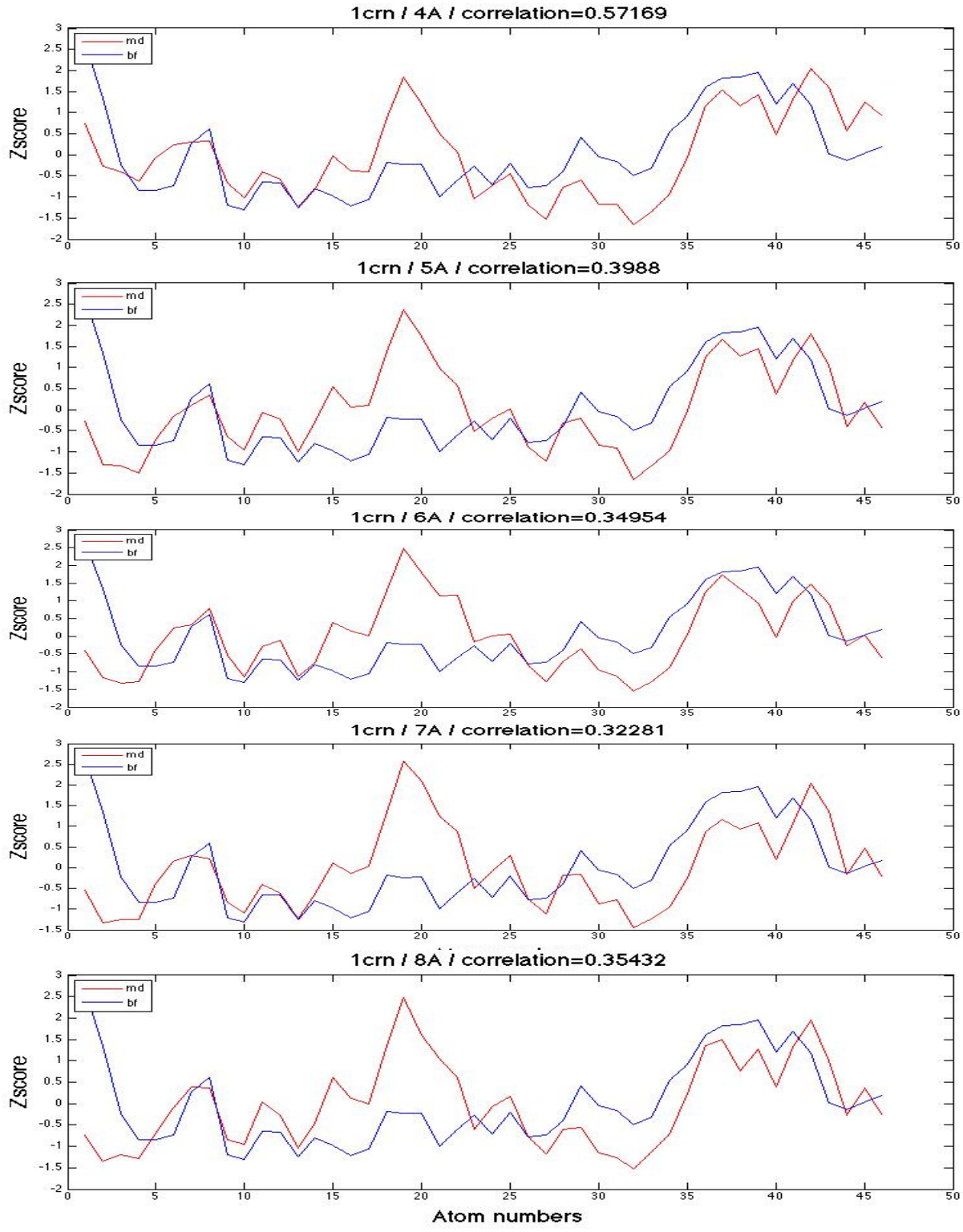
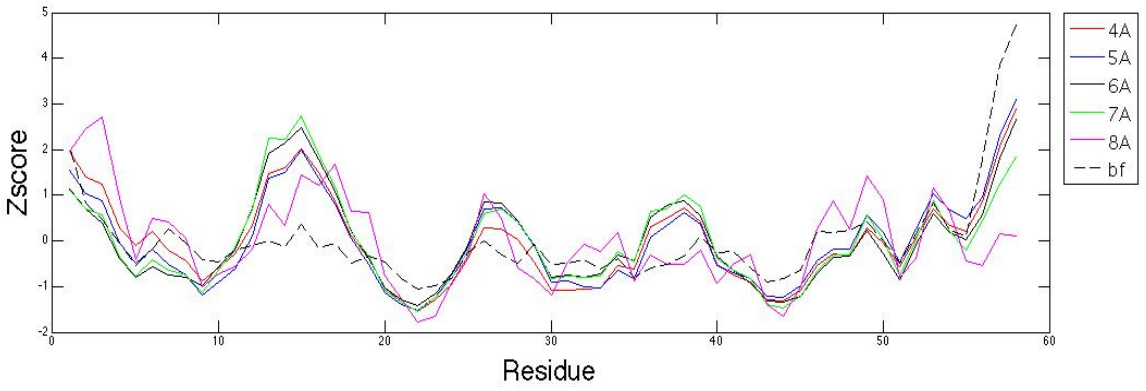


Figure 5. Effect of cutoff on RMSF in 1crn



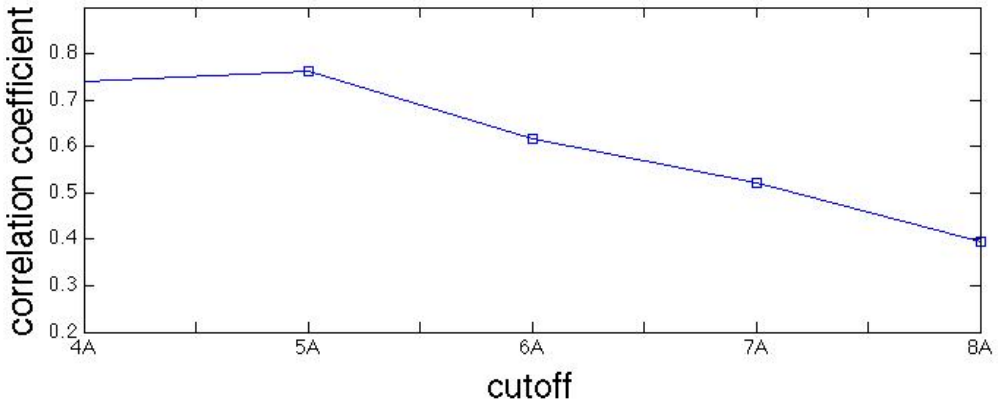


Figure 6. Effect of cutoff on RMSF in 5pti

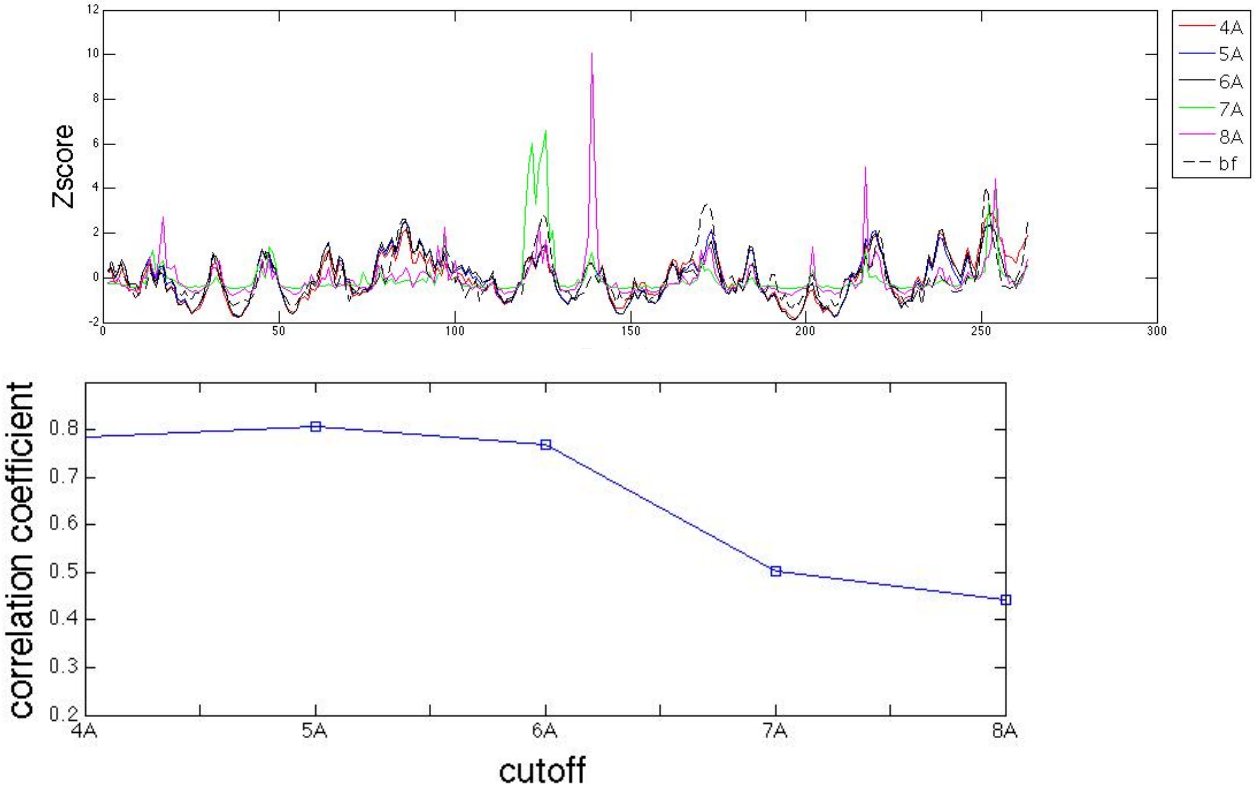


Figure 7. Effect of cutoff on RMSF in 1arb

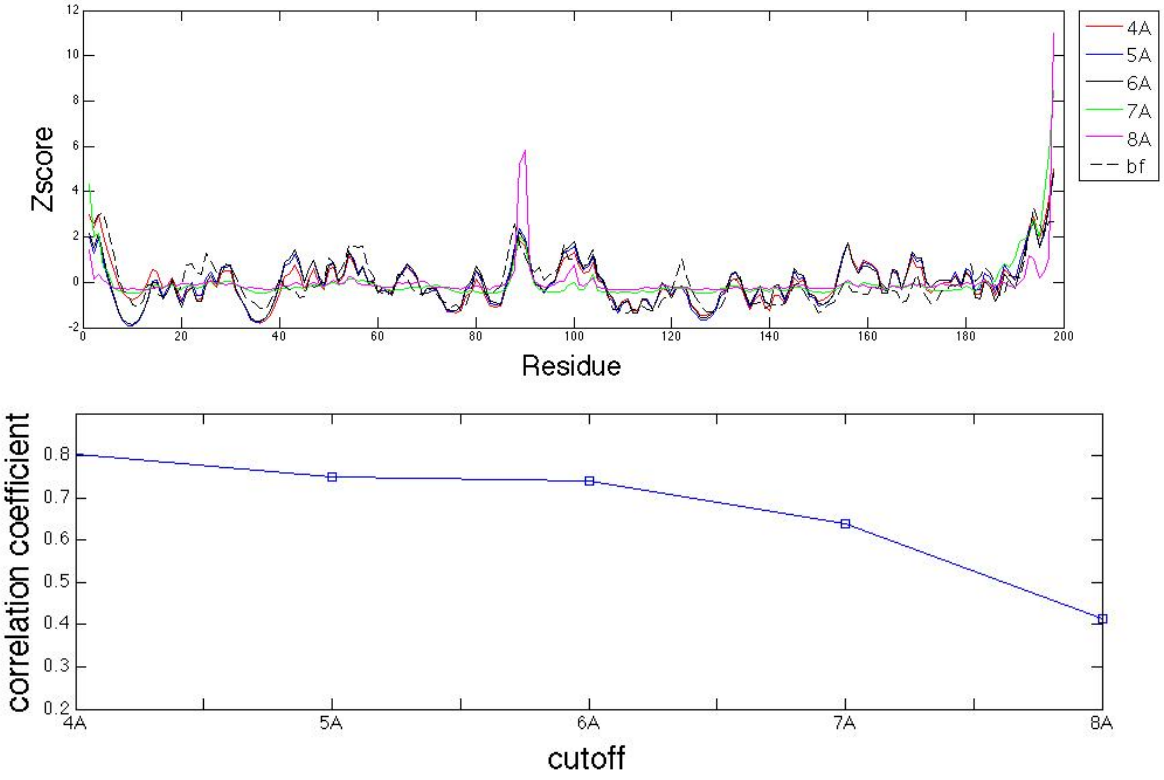


Figure 8. Effect of cutoff on RMSF in 1chd

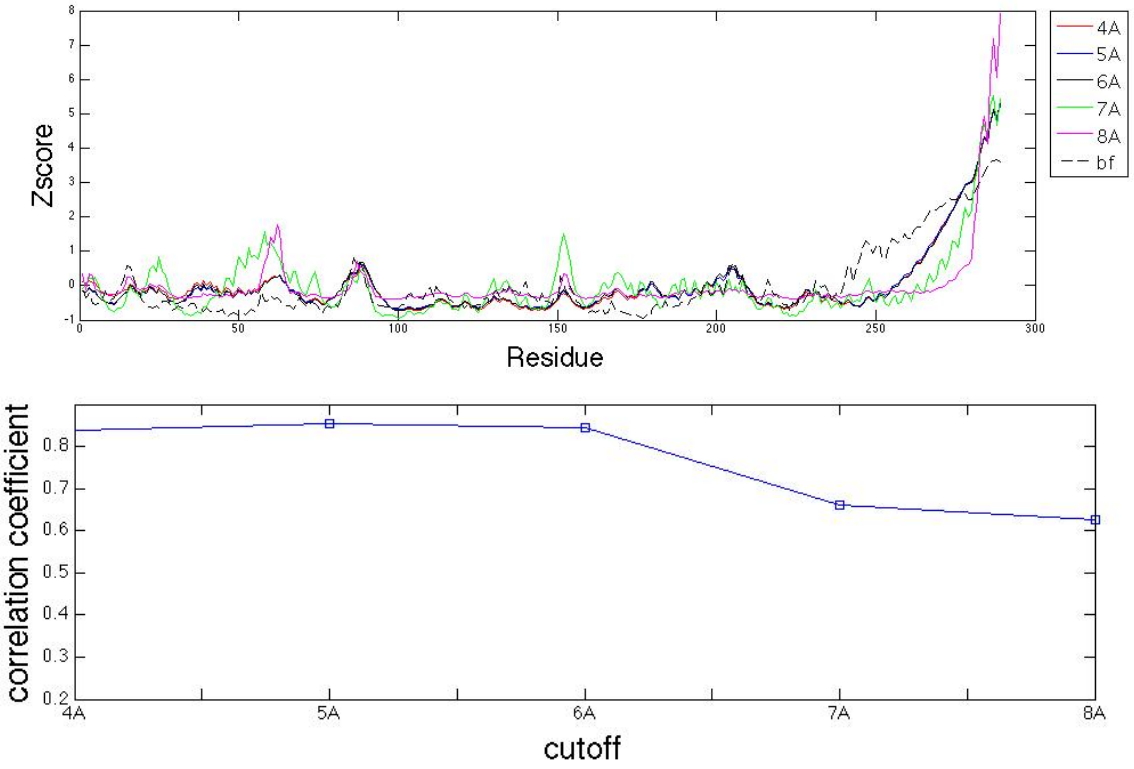


Figure 9. Effect of cutoff on RMSF in 1kqfb

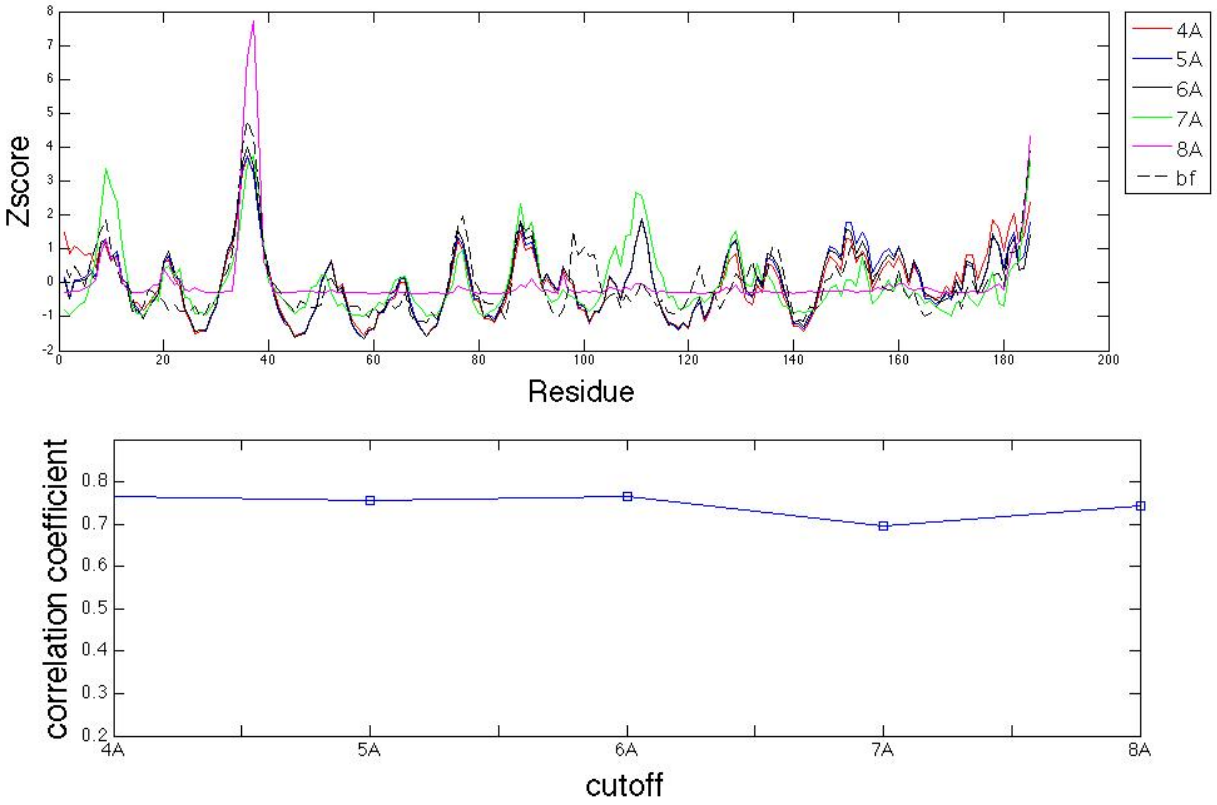


Figure 10. Effect of cutoff on RMSF in 1pl3a

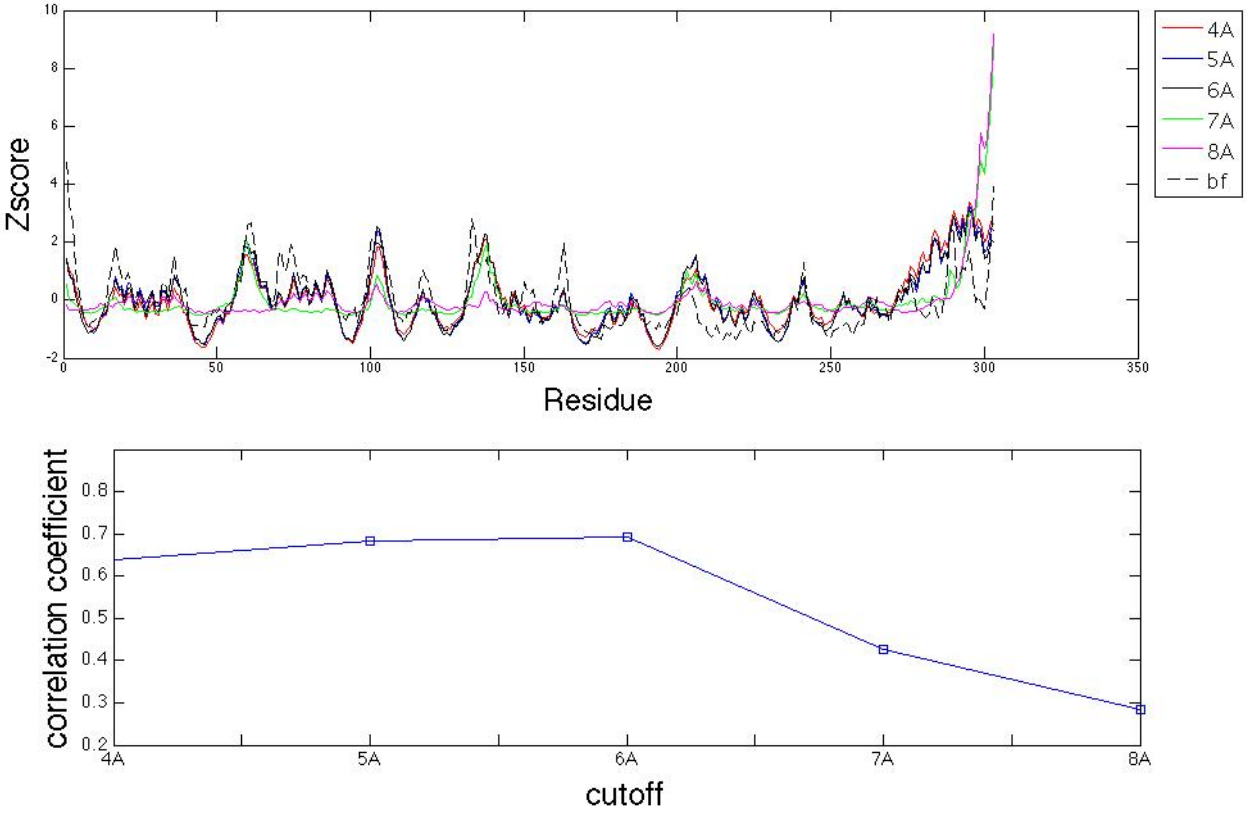


Figure 11. Effect of cutoff on RMSF in 1uf5a

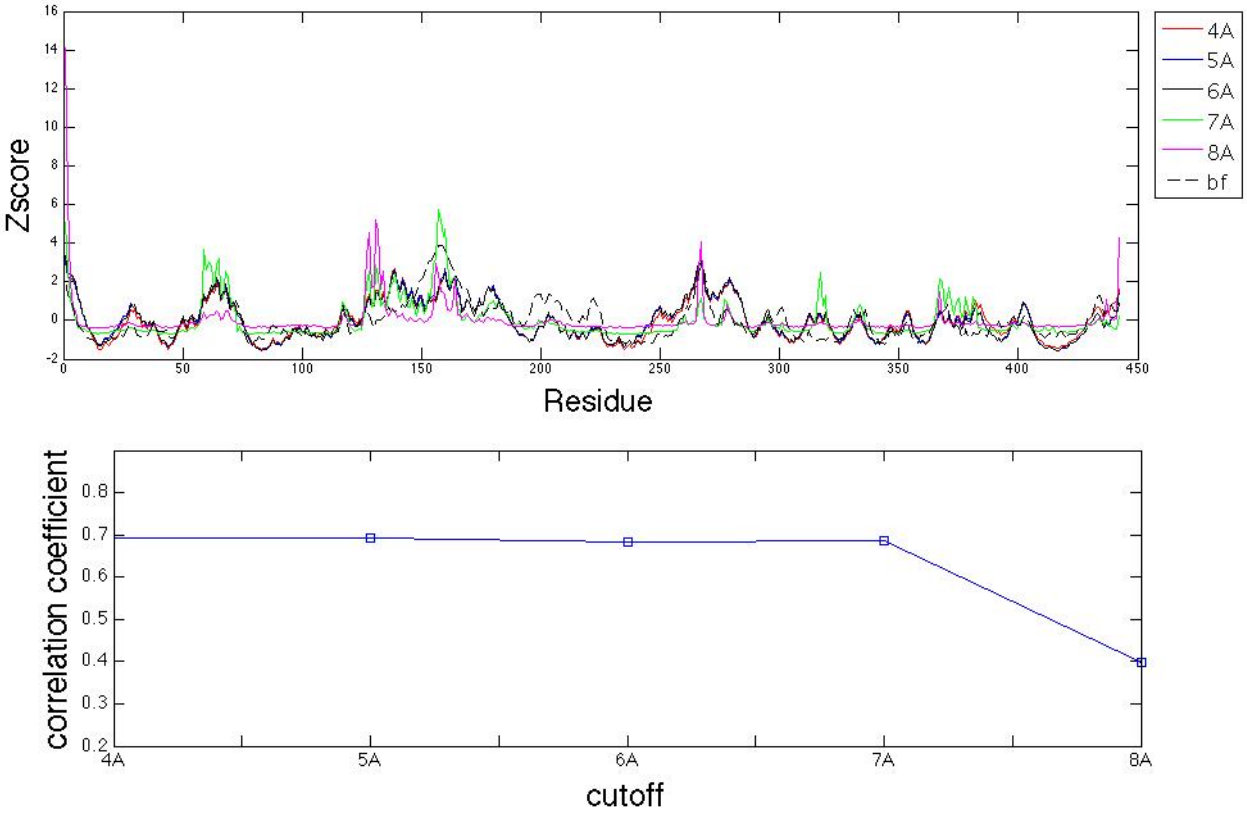


Figure 12. Effect of cutoff on RMSF in 2bcgg

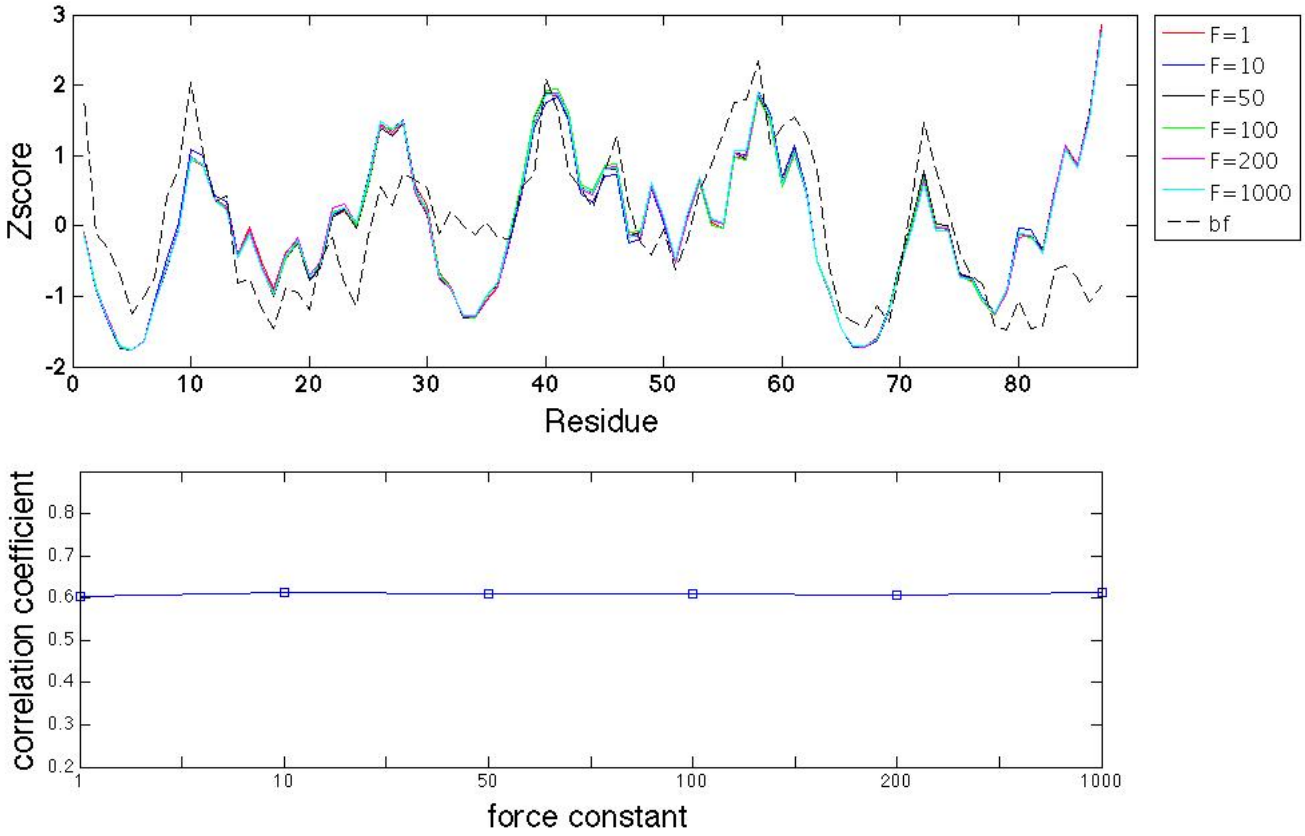


Figure 13. Effect of force constant on RMSF in 1aba, F1 means the force constant=1. F50 means the force constant =50.

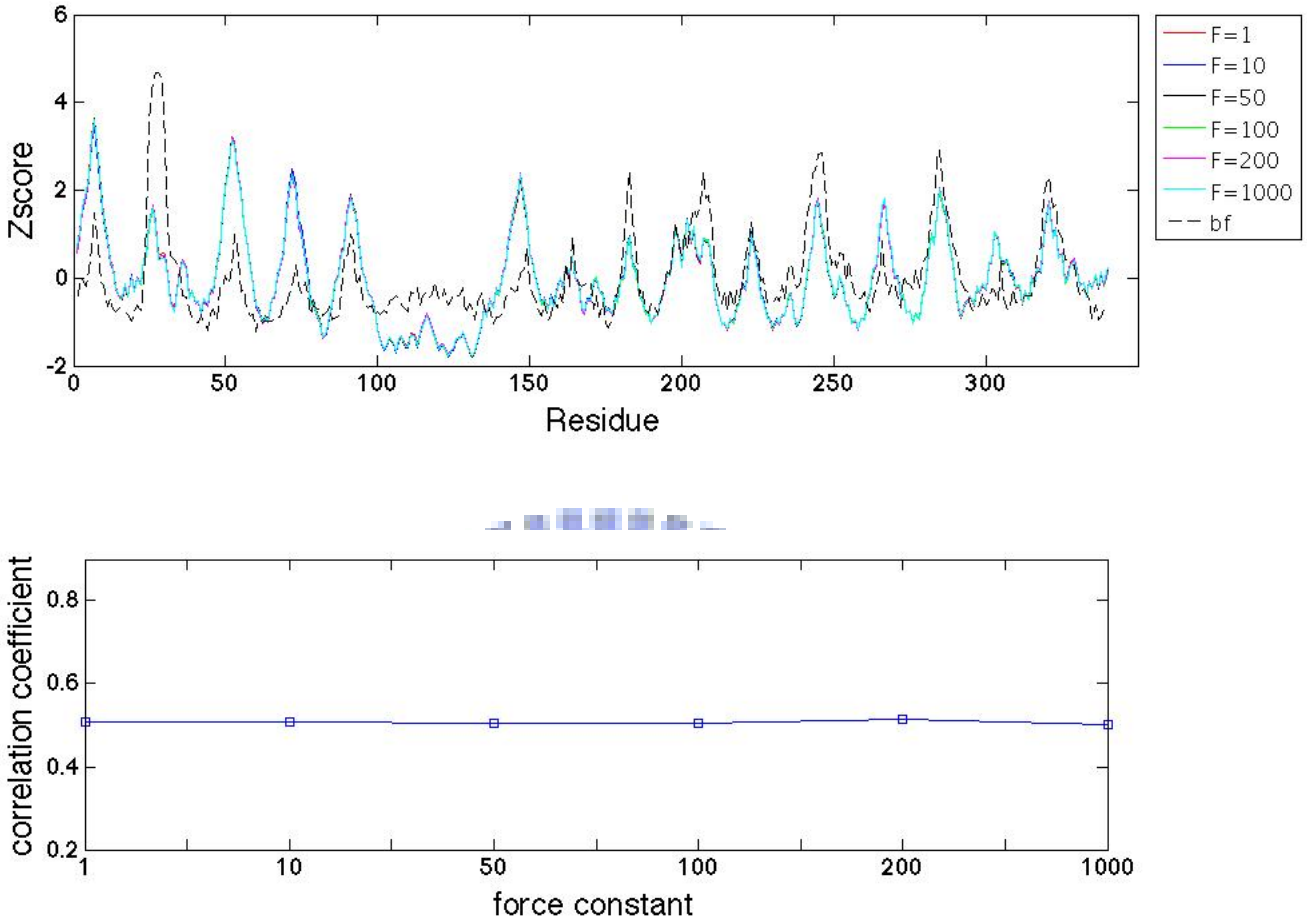


Figure 14. Effect of force constant on RMSF in 2omf, , F1 means the force constant=1. F50 means the force constant =50.

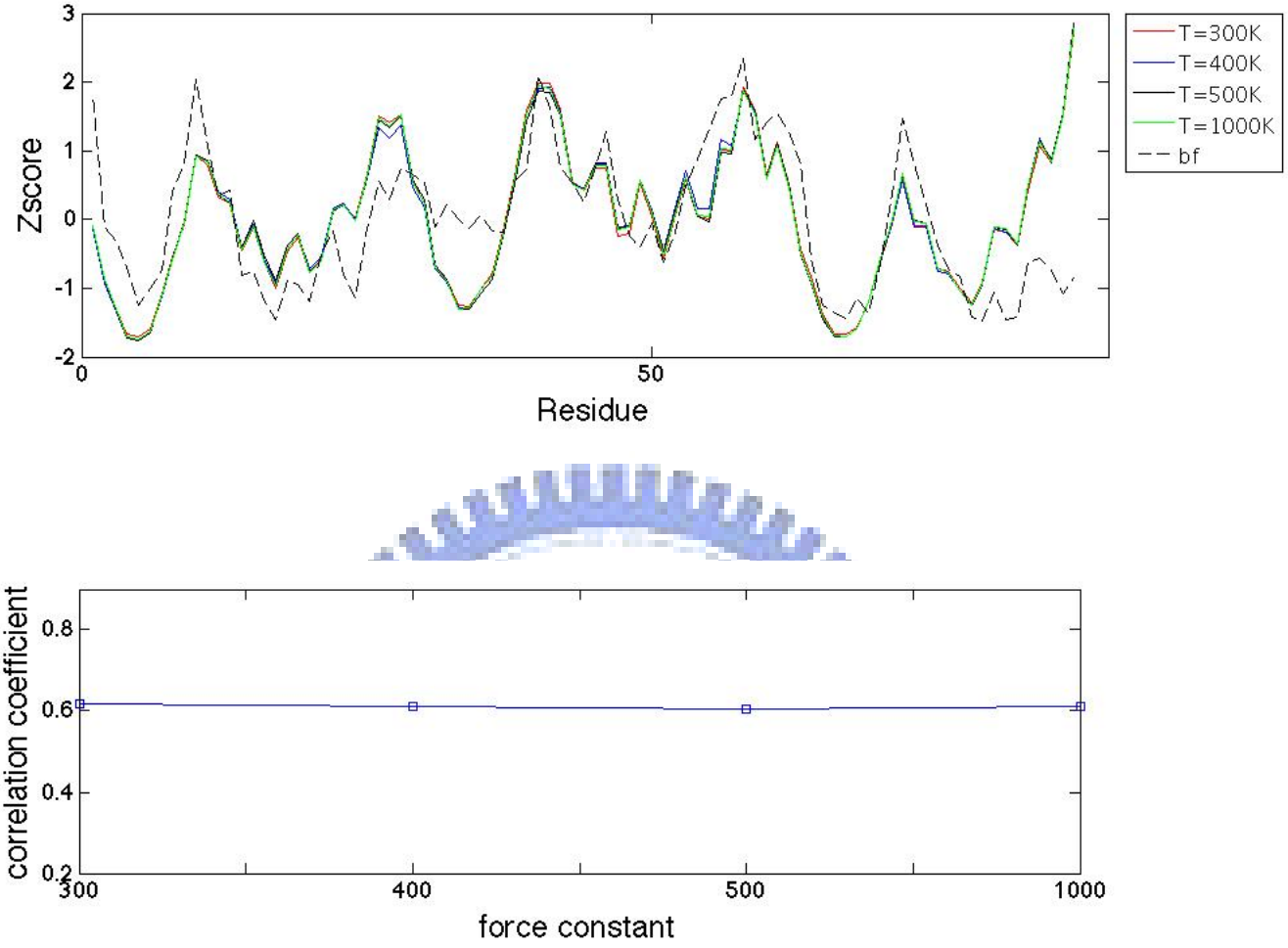
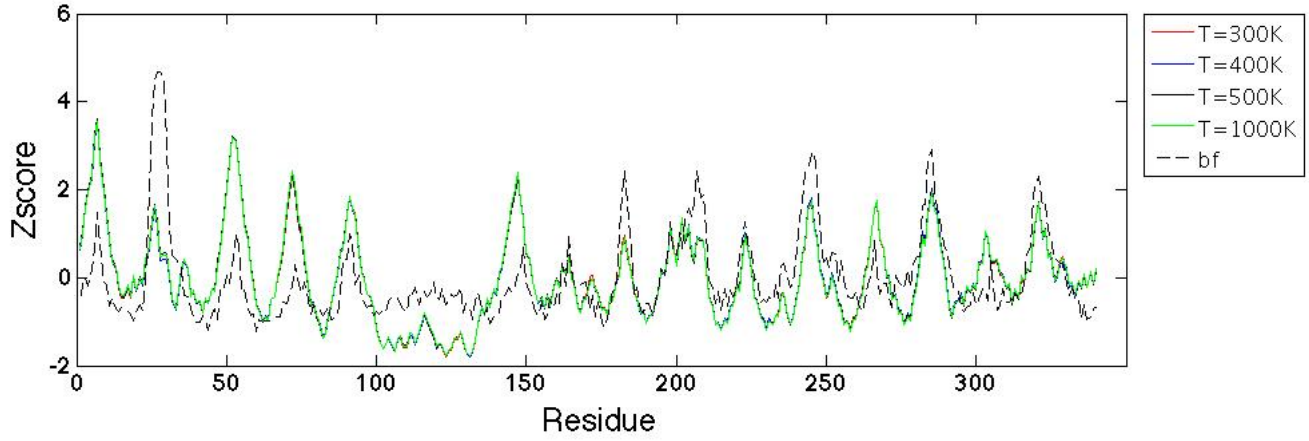


Figure 15. Effect of temperature on RMSF in 1aba



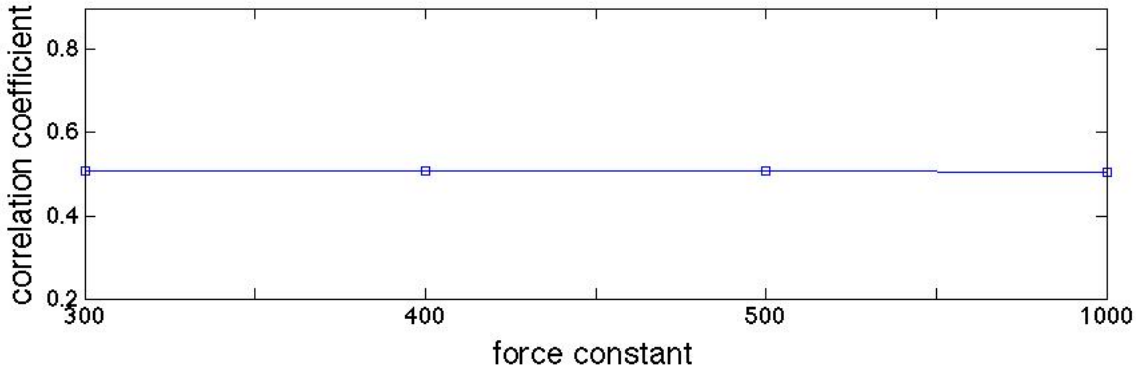


Figure 16. Effect of temperature on RMSF in 2omf

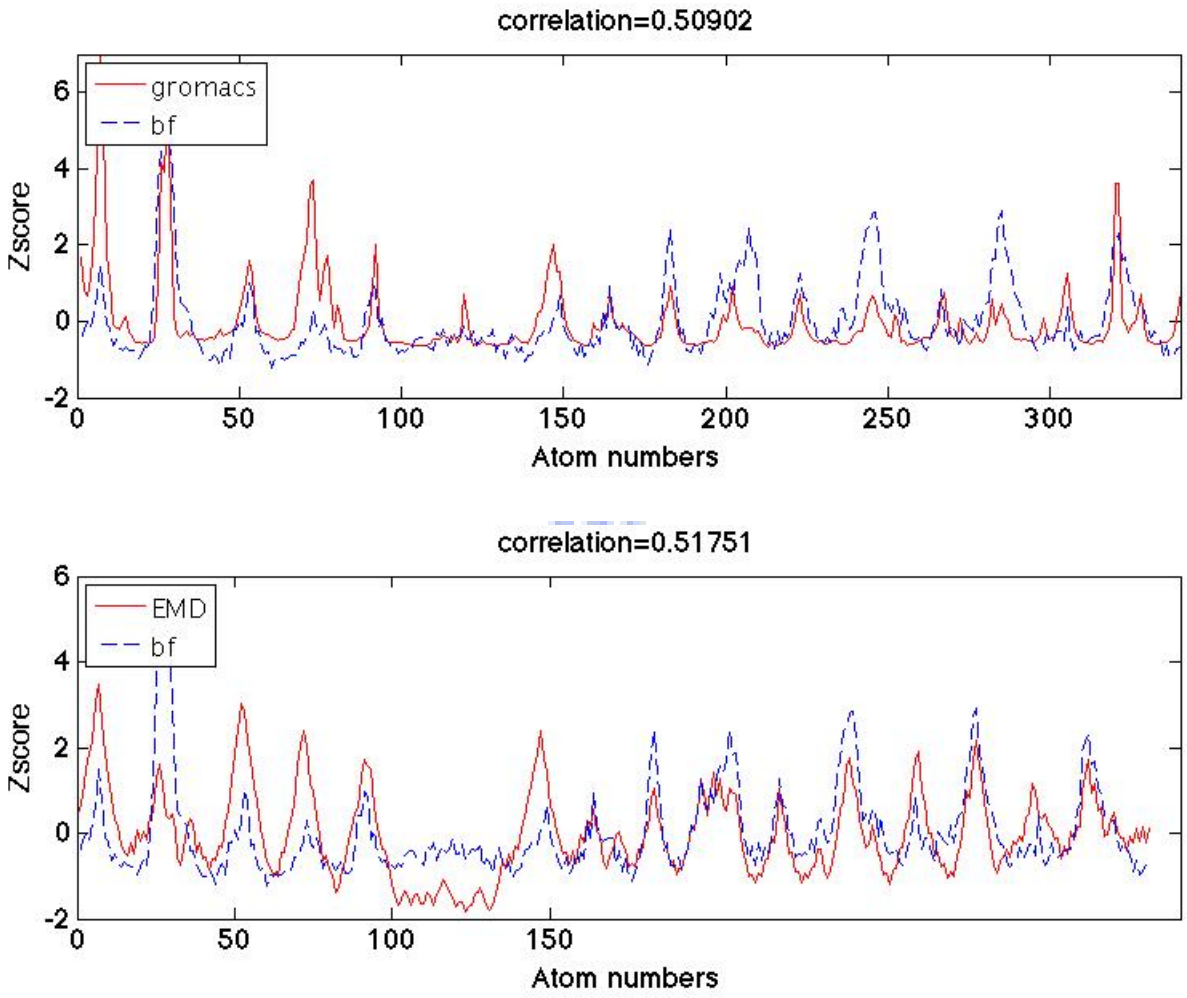


Figure 17. Comparing the correlation coefficient of B-factor with RMSF from GROMACS and EMD in 2omf.

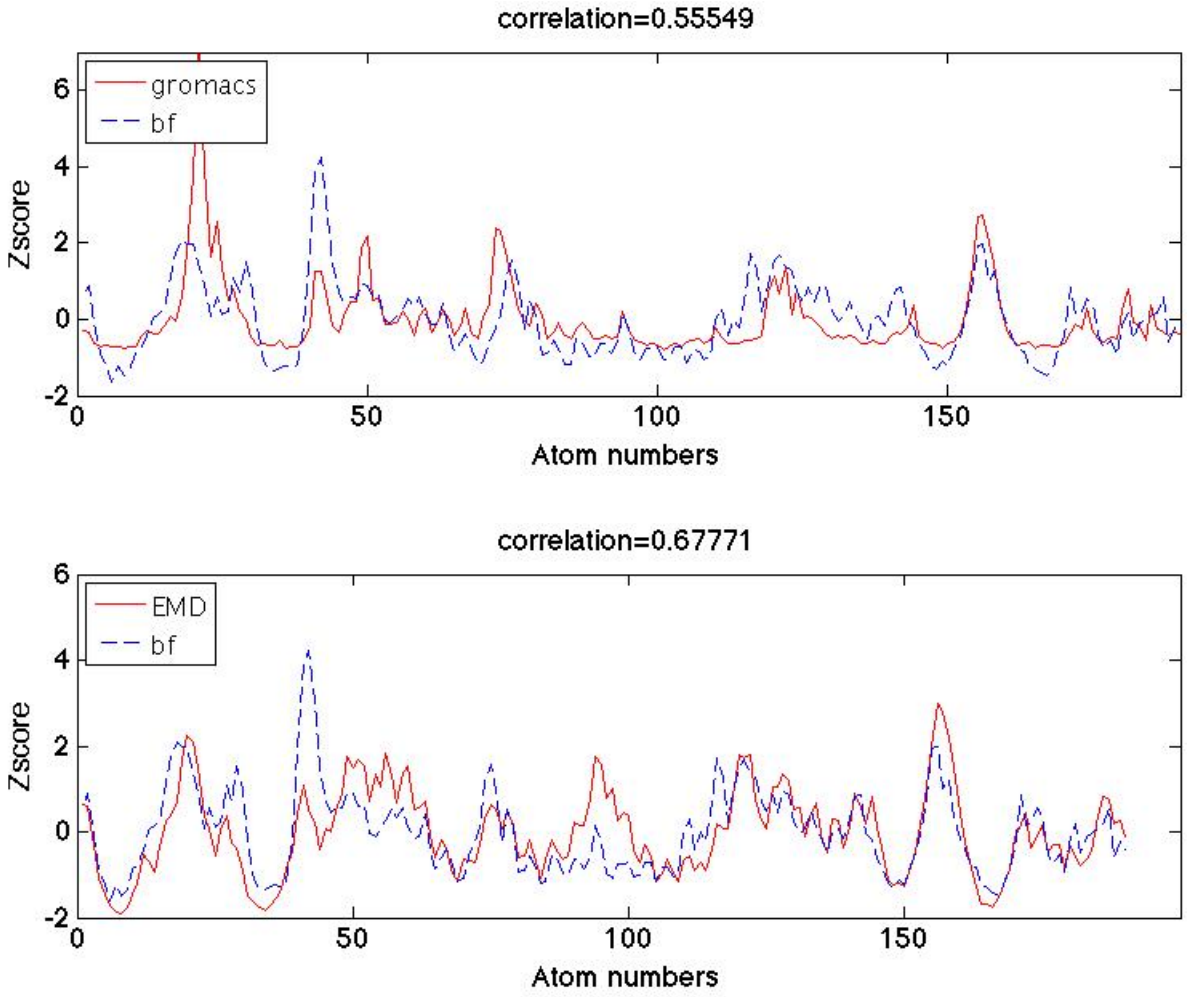


Figure 18. Comparing the correlation coefficient of B-factor with RMSF from GROMACS and EMD in 1ucda

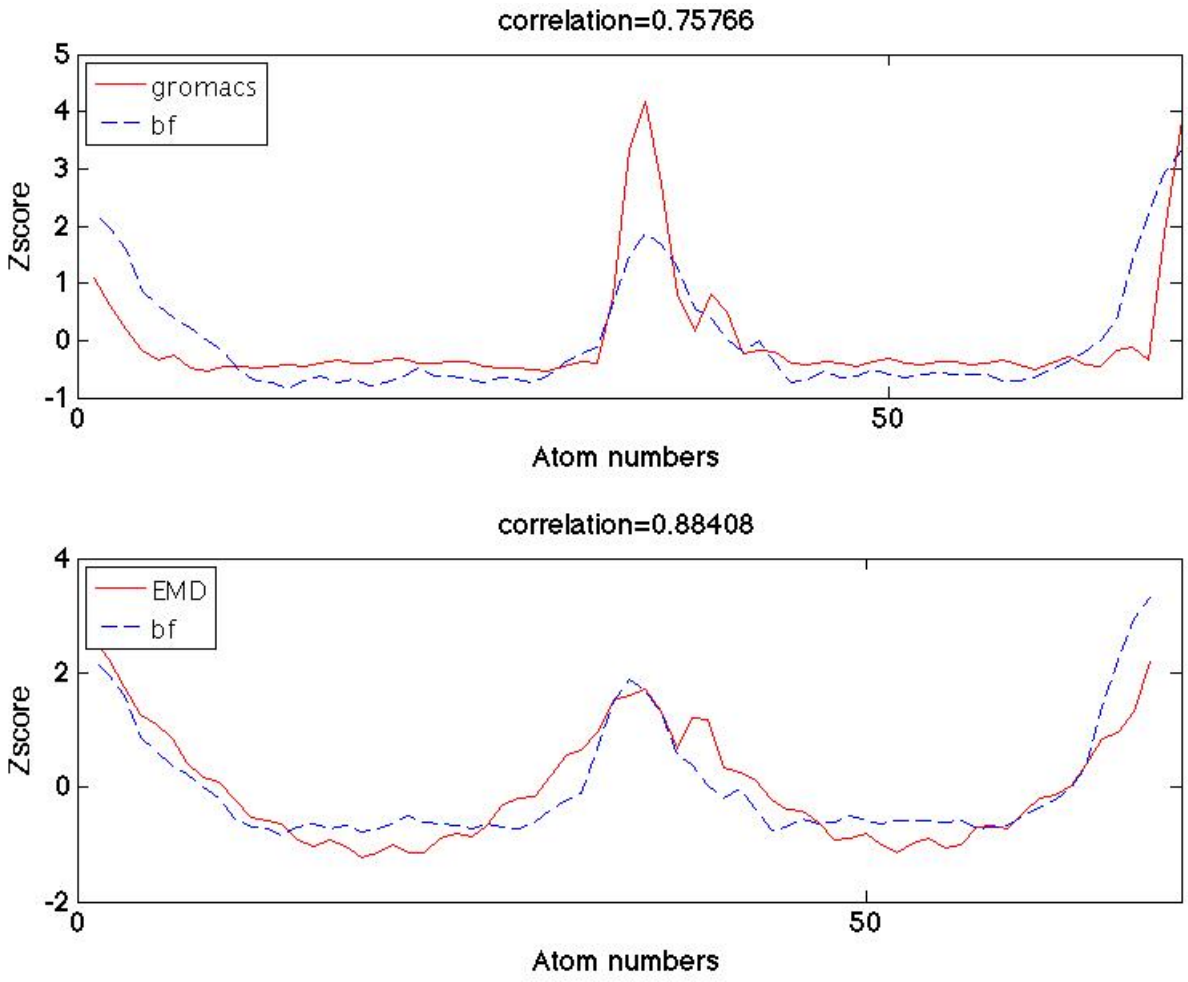


Figure 19. Comparing the correlation coefficient of B-factor with RMSF from GROMACS and EMD in 1qr9a

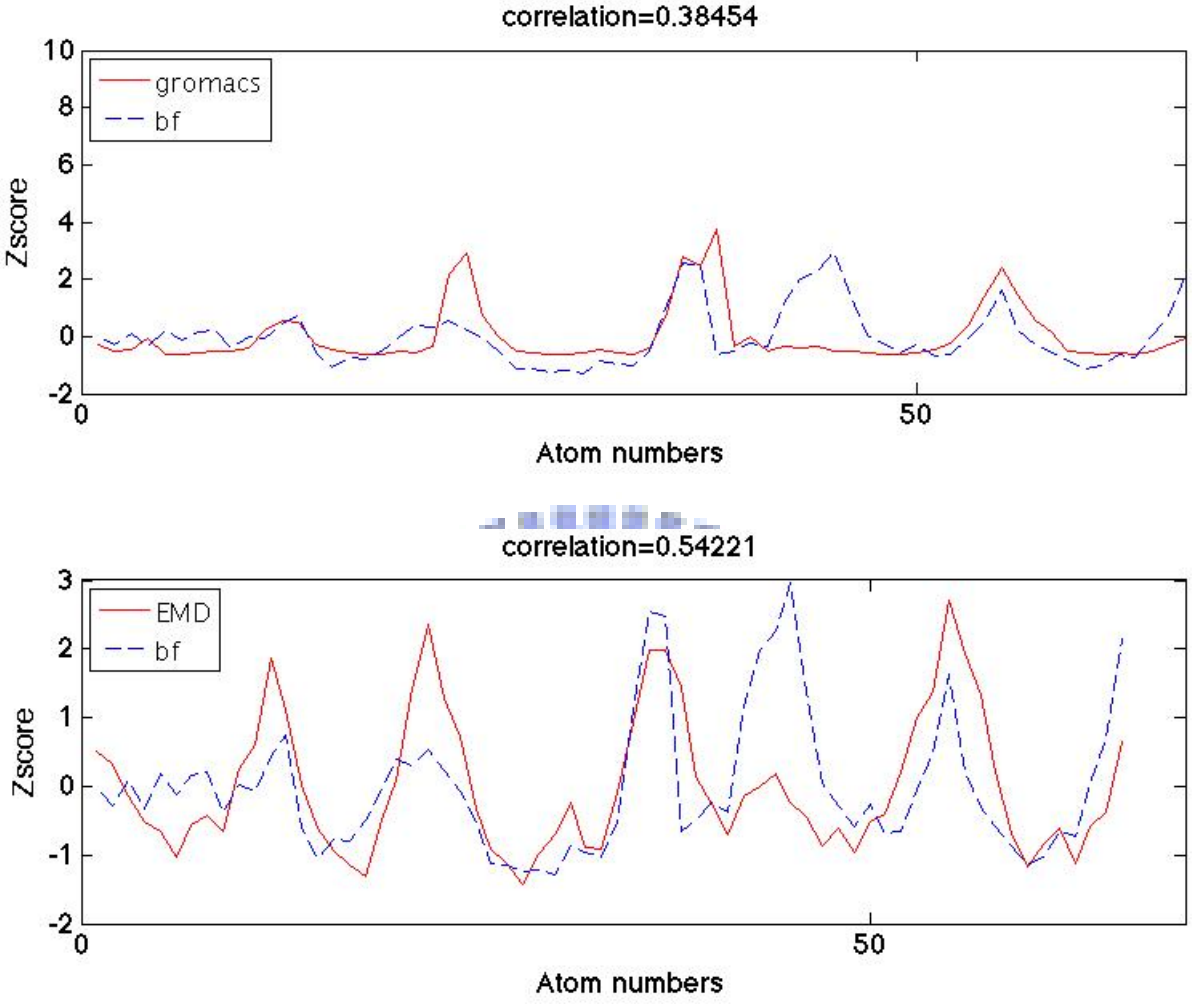


Figure 20. Comparing the correlation coefficient of B-factor with RMSF from GROMACS and EMD in 1c9oa

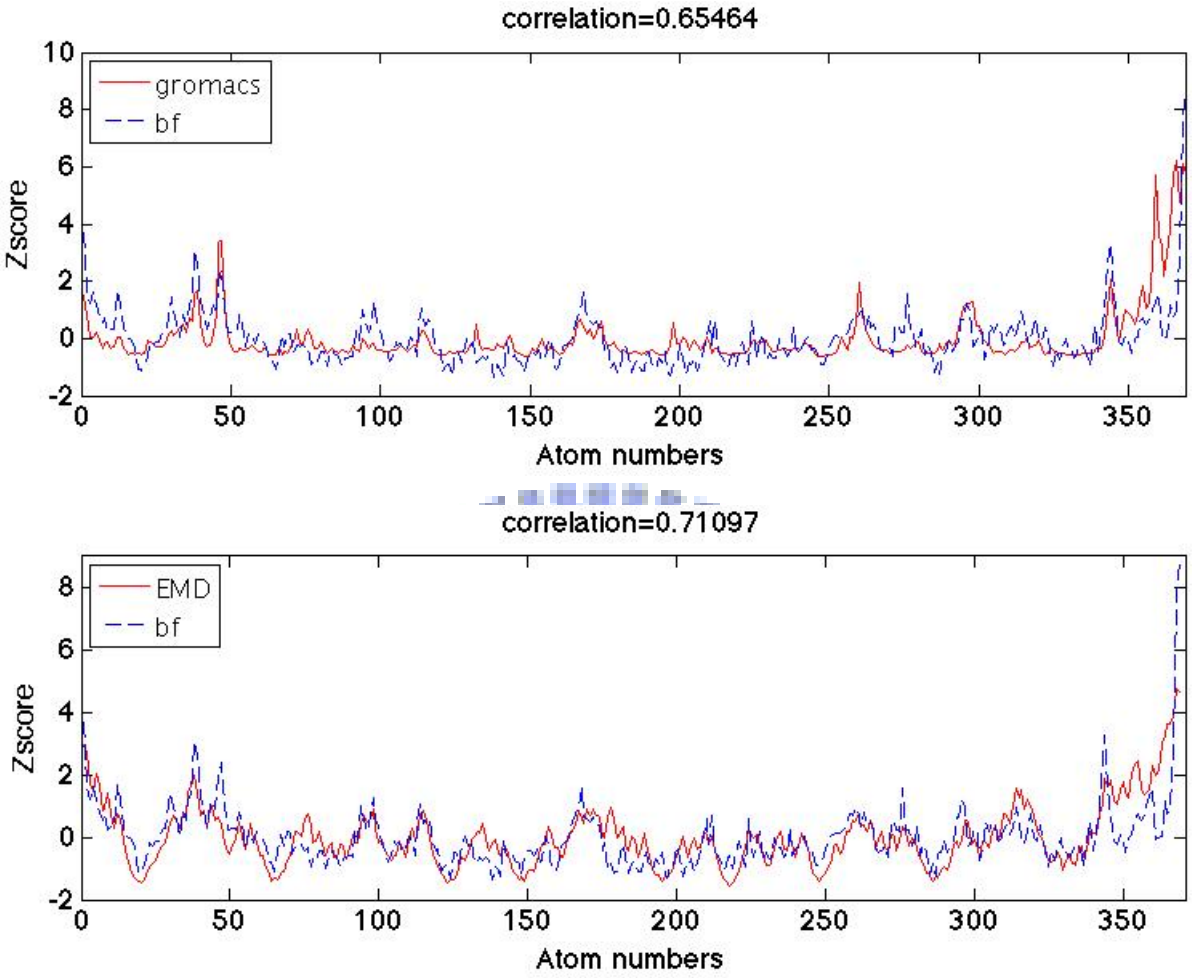


Figure 21. Comparing the correlation coefficient of B-factor with RMSF from GROMACS and EMD in 1itua

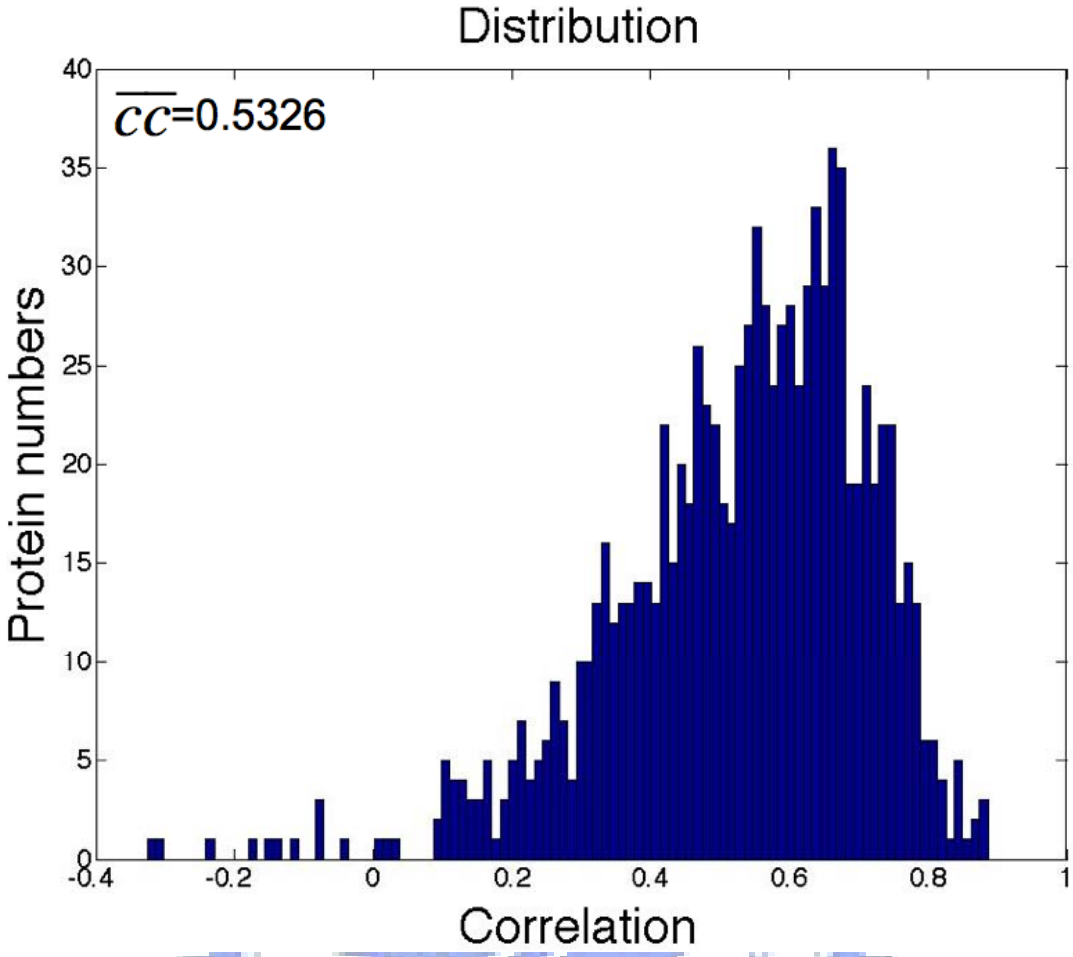


Figure 22. The distribution of correlation coefficient of RMSF from EMD and B-factor

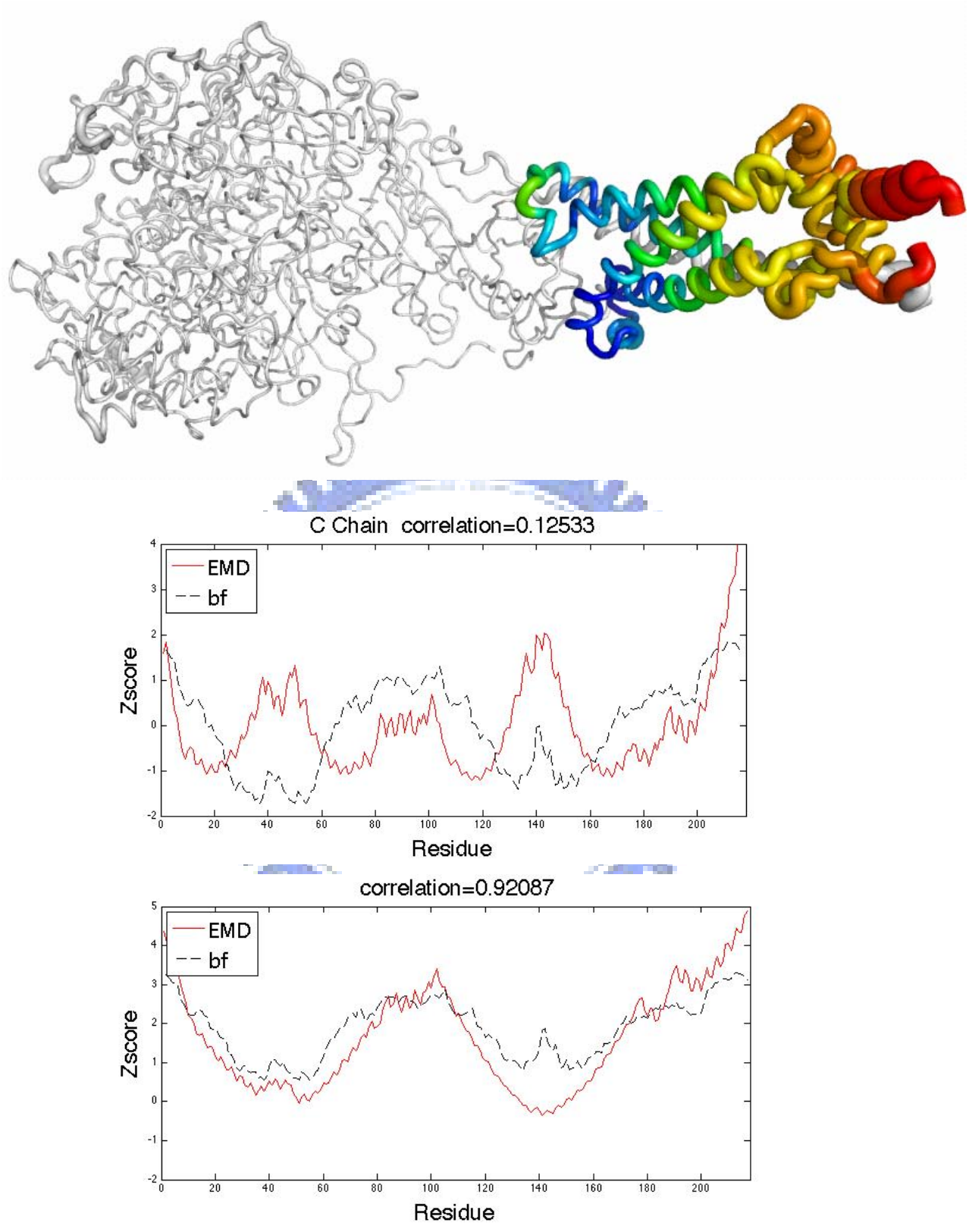


Figure 23. C-chain is a subunit of a oligomer in 1KQFC, we combine it to other subunit and get the overlap results

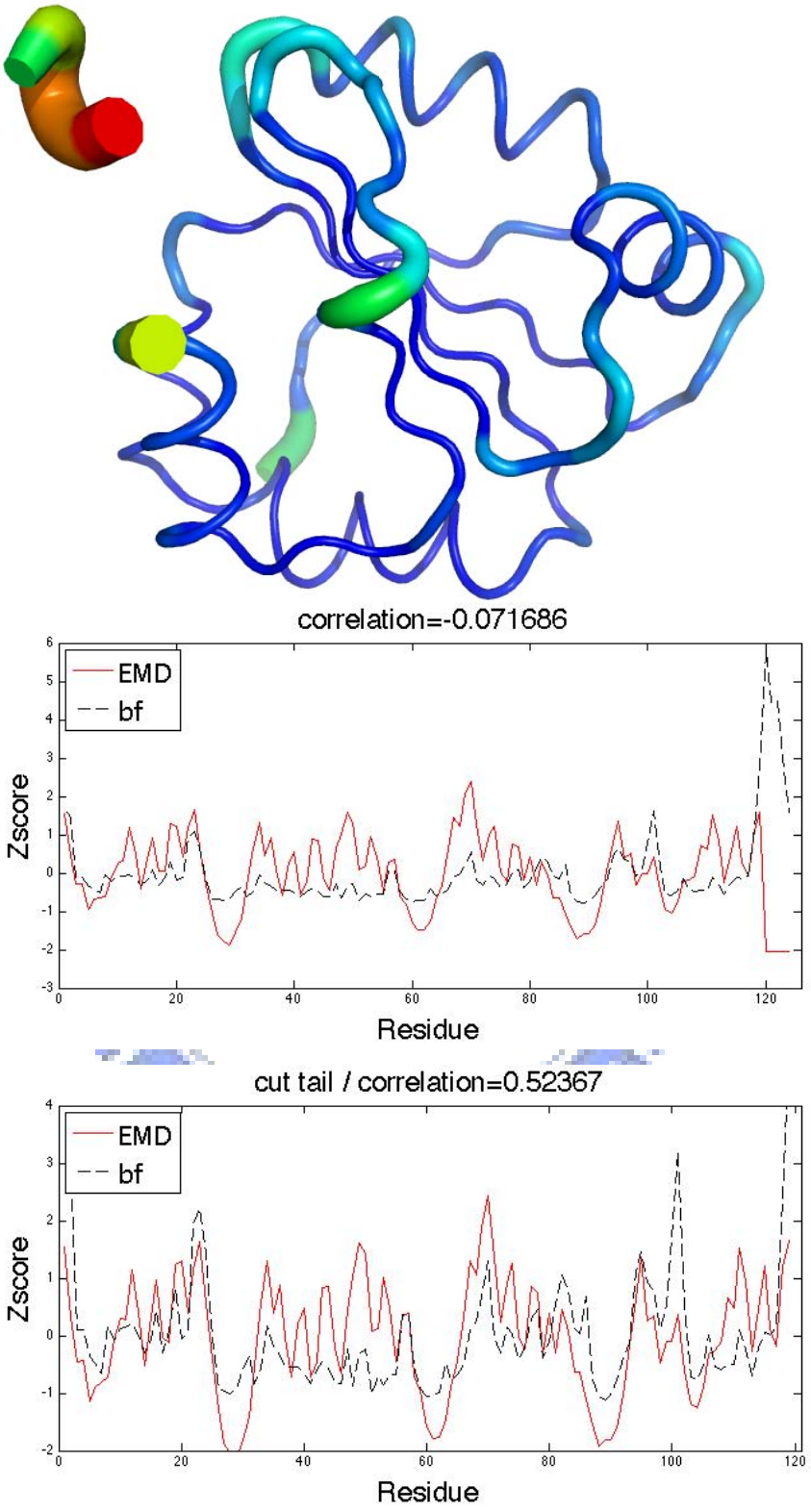


Figure 24. Effect of missing residue in 2fwg, we cut the five histidines of C-terminal

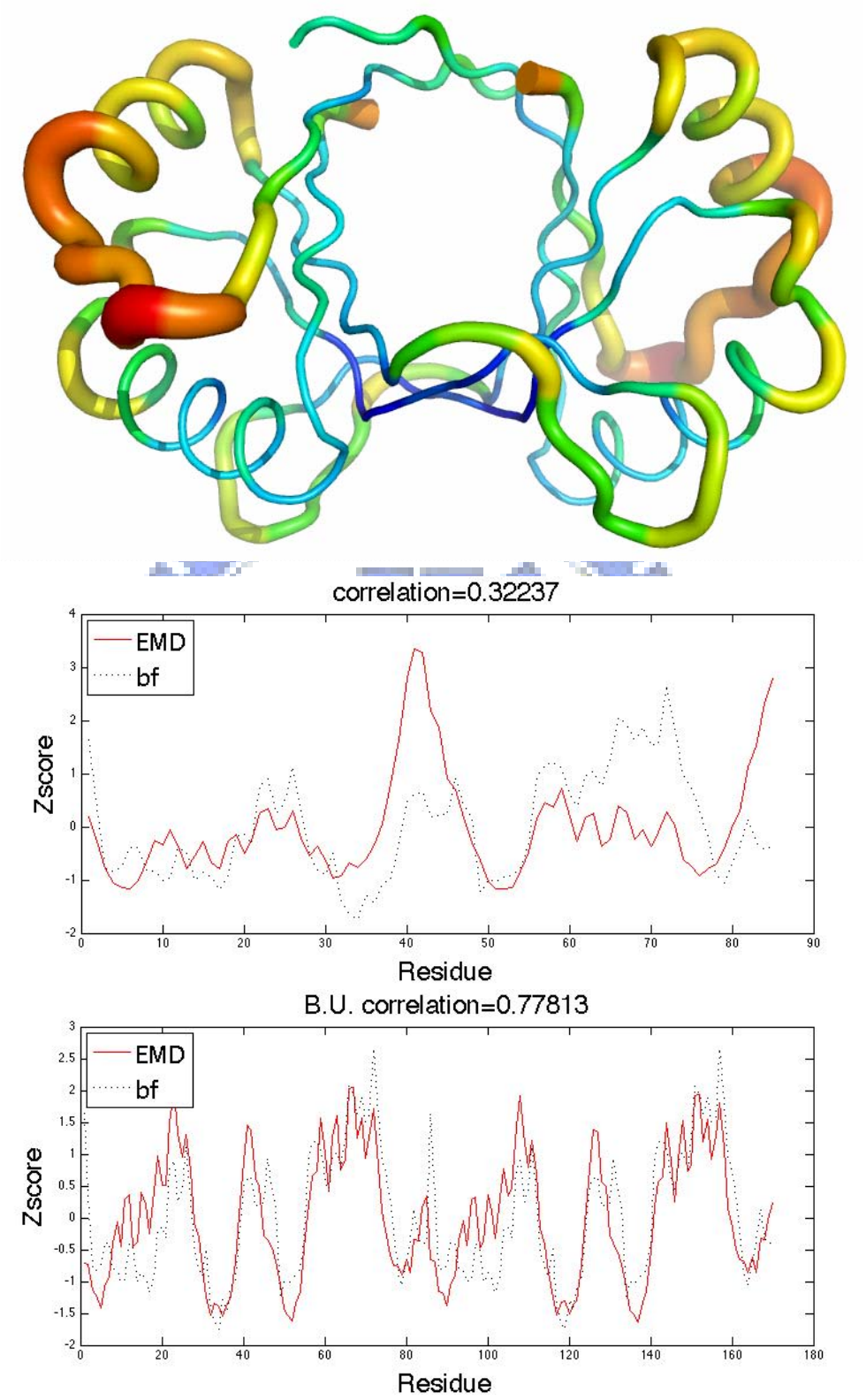


Figure 25. Effect of biological unit in 2BOP

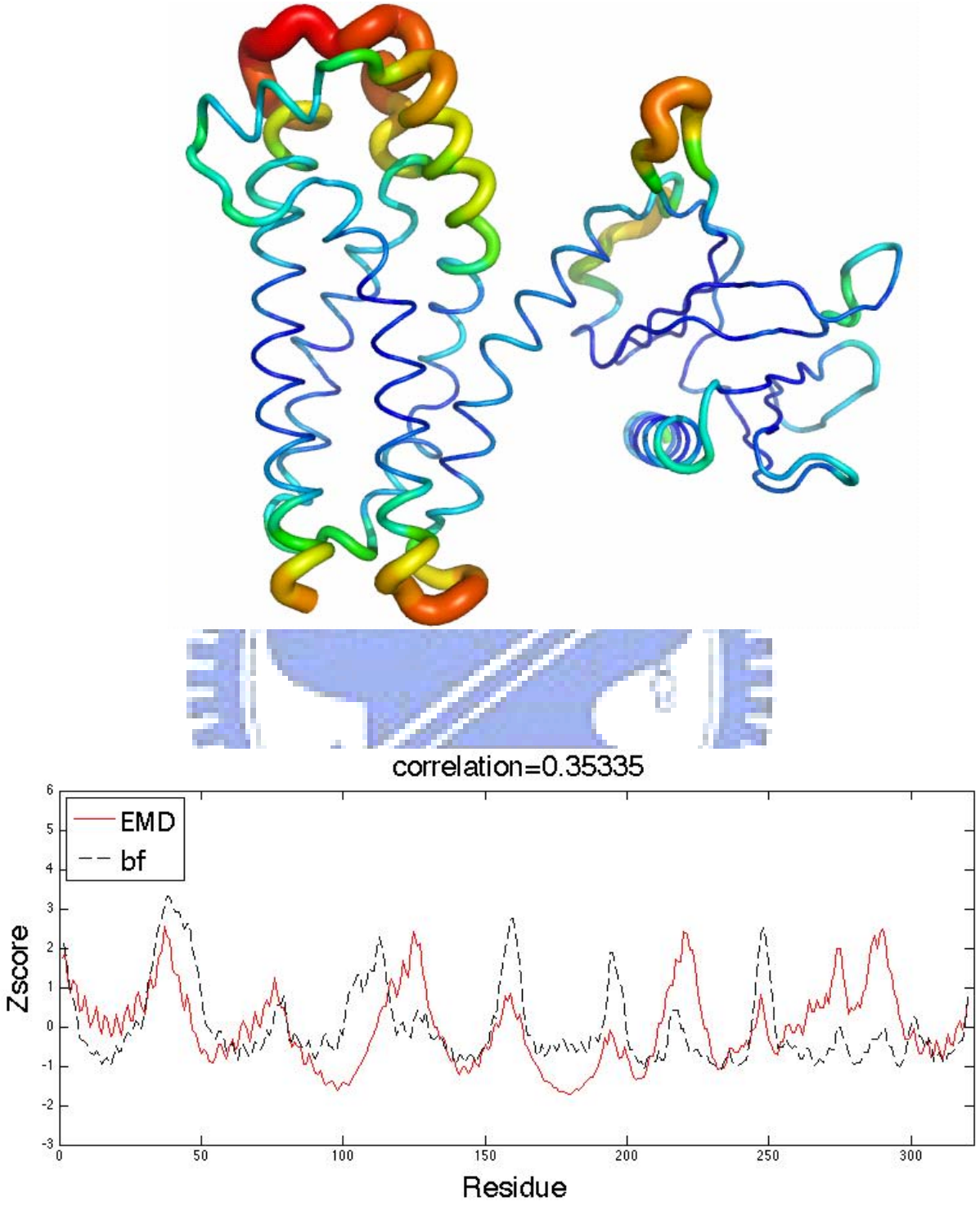


Figure 26. Effect of separating proteins of several domains in 1nty, testing all molecular.

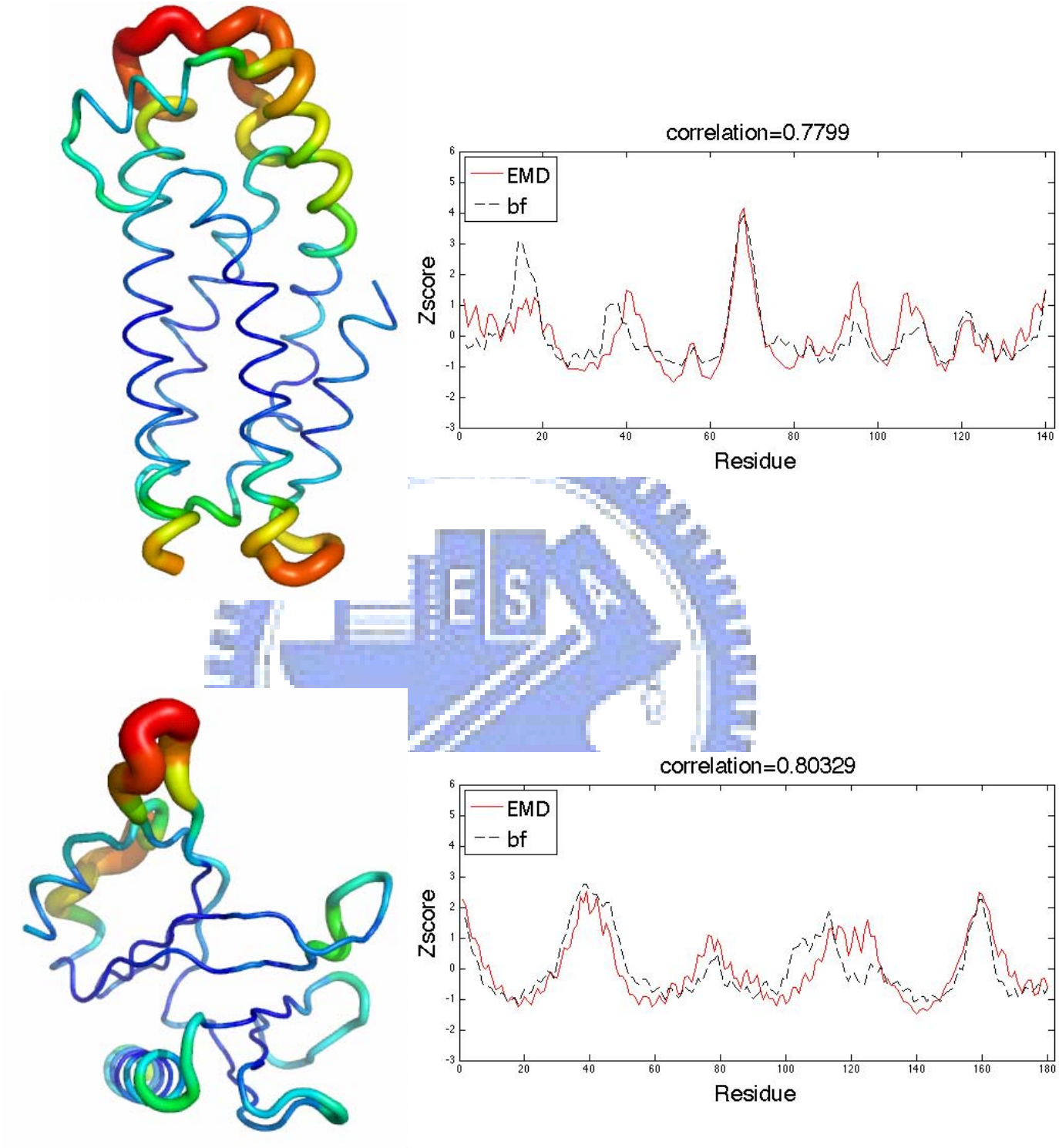


Figure 27. Effect of separating proteins of several domains in 1nty, separating two domains

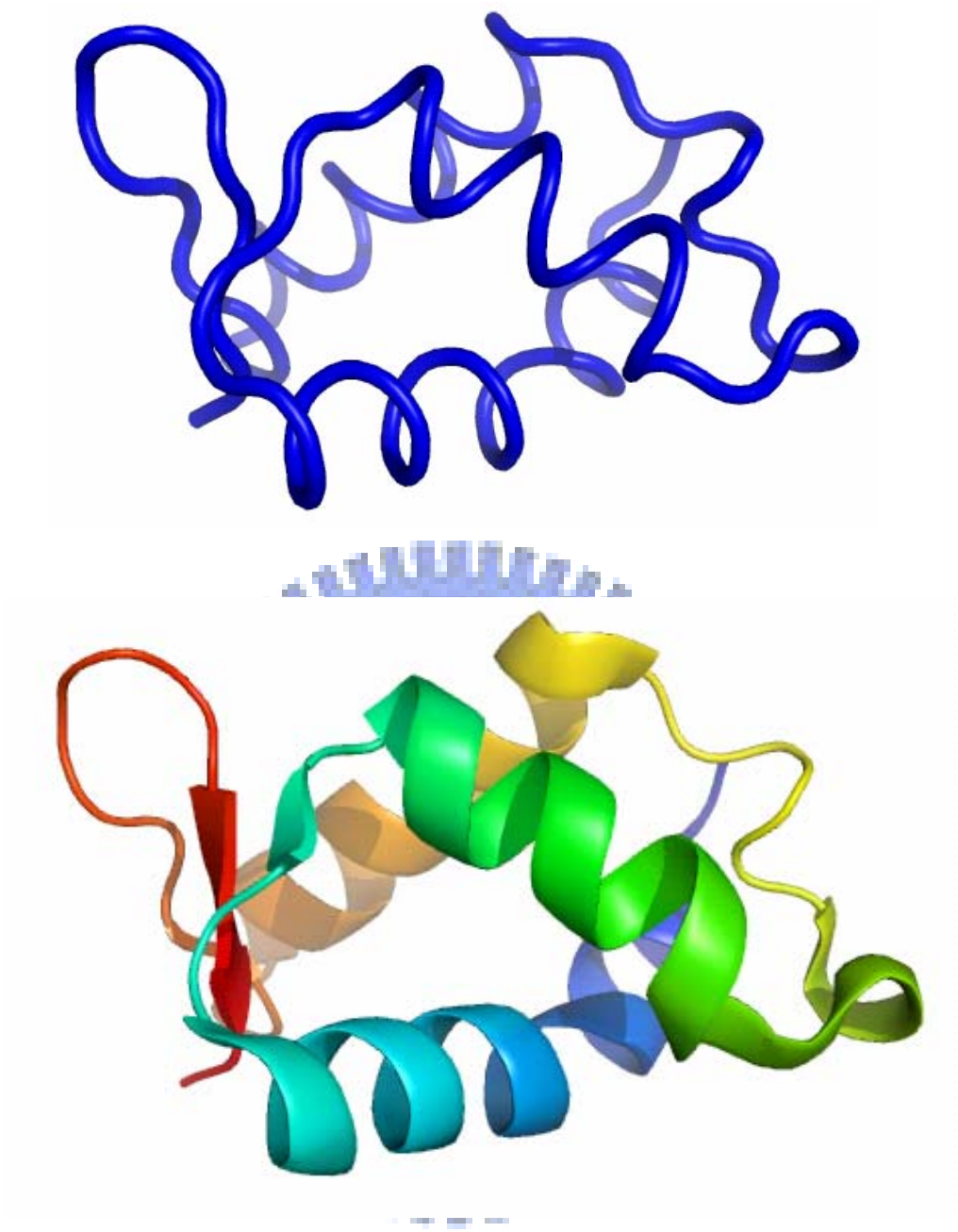


Figure 28. Example of error of experimental data

Appendix

The datasets pf PDB-REPEDB

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	1KQFA	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								