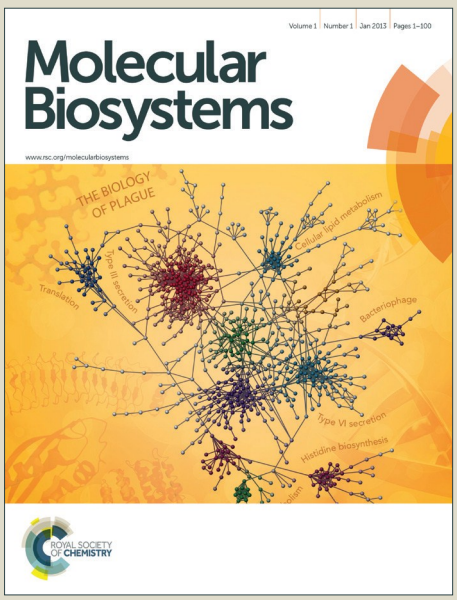
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Molecular BioSystems

PAPER

Exploring the Mechanism How tvMyb2 Recognizes and Binds ap65-1 by Molecular Dynamics Simulations and Free Energy Calculations

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Wei-Kang Li,[b] Qing-Chuan Zheng* [a,b] and Hong-Xing Zhang [b]

TvMyb2, one of the Myb-like transcriptional factors in Trichomonas vaginalis, which binds to two closely spaced promoter sites, MRE-1/MRE-2r and MRE-2f, on the ap65-1 gene. However, the detail dynamical structural characteristic of the tvMyb2-ap65-1 complex and the detail study of the protein in the complex have not been done. Focused on specific tvMyb2-MRE-2-13 complex (PDB code: 3OSF) and a serials of mutants K51A, R84A and R87A, we applied molecular dynamics (MD) simulation and molecular mechanics generalized Born surface area (MM-GBSA) free energy calculation to examine the role of tvMyb2 protein in recognition interaction. The simulation results indicate that tvMyb2 become stable when it binds the DNA duplex. A serials of mutants, K51A, R84A and R87A, have been followed, and the results of statistical analyses of the H-bond and hydrophobic contacts show that some residues have significant influence on recognition and binding to ap65-1 DNA. Our work can give important information to understand the interactions of tvMyb2 with ap65-1.

Introduction

The Myb family consist of a number of proteins, which are encoded by a series of proto-oncogene, with extensive functions in both animals and plants, such as being an important role in developing, cell survival, proliferation and homeostasis.[1] Furthermore, the Myb proteins have also been testified to be associated with leukemogenesis in several species including humans.[2, 3] In recently years, lots of evidences have been shown that Myb is tightly regulated by attenuation sequences which resides in the first intron.[4] The Myb domain normally has at most four imperfect amino acid sequence repeats (R). Each of the repeats has a very similar folding architecture, containing three well-defined helices, and the second and third helix in each repeat often forms a helix turn helix (HTH) structure. Both of the third helices of the second repeat (R2) and third repeat (R3) are involved in specific base recognition in the major groove of the DNA.[5, 6] Myb proteins are mostly operate transcriptional activator by highly conserved domains that binds to specified sequence t/cAACt/gG, which is known as the Myb binding site (MBS).[5, 7]
Trichomonas vaginalis (T. vaginalis), a kind of flagellated protozoan, causes the most prevalent non-viral sexually transmitted disease (STD) worldwide.[8] It has been demonstrated that the infection with the trichomonosis caused by the T. Vaginalis is associated with increased risk for adverse pregnancy outcome, cervical cancer, and HIV seroconversion.[9] Adherence to the vaginal epithelium by the T. vaginalis, a property key to colonization and infection, is a highly specific event that is mediated by adhesins. The adhesion protein 65-1 (ap65-1) gene encodes multiple homologous 65-kDa proteins. This protein have been detect on plasma membranes as part of adhesion complexes during host-parasite encounters in T. vaginalis.[10-12] In recent reports, the transcription of the ap65-1 gene is regulated in cis by two neighbouring MBS, MRE-1/MRE-2r and MRE-2f, along with several other neighbouring DNA regulatory elements in the ap65-1 promoter. The MRE-1/MRE-2r site comprises three overlapping DNA elements, which are the binding sites for three Myb-like proteins: tvMyb1, tvMyb2 and tvMyb3[13-16]. However, the tvMybs are not like the other member of the Myb family, tvMybs recognize the specified sequence a/gACGAT instead of the t/cAACt/gG. But at the same time, the tvMyb2 is just like the c-Myb in vertebrates, the DNA-free tvMyb2 has a very flexible conformation. Upon binding to the promoter DNA ap65-1, tvMyb2 undergoes conformational re-arrangement and structural stabilization[16]. The 3D structures of tvMyb2-MRE-2-13 complex has been determined by X-ray crystallography (PDB code: 3OSF), illustrated in Figure 1. Nonetheless, the stabilization and recognition mechanism by tvMyb are still unclear at the atomic level.

[a] *Key Laboratory for Molecular Enzymology and Engineering of the Ministry of Education, Jilin University, Changchun 130023, People's Republic of China.*
[b] *State Key Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, Jilin University, Changchun 130023, People's Republic of China.*

In order to clarify the details of the mechanisms and indicate the direction of suppressing DNA silencing or inhibit ap65-1 gene translation then lead the experts to develop treatments for the disease, Molecular dynamic (MD) simulation combined the molecular mechanics generalized Born surface area (MM-GBSA) method are employed to clarify the interactional details and energetic information between the tvMyb2 and MRE-2-13 gene at atomic level. And then, a series of MD simulations of mutates K51A, R84A, and R87A with DNA complexes have been done. All these MD simulations and MM-GBSA calculations complement the experiments [16] and help us to get better understanding of the recognition and binding mechanism of tvMyb2 and MRE-2-13 DNA complex by offering the binding details and experimentally inaccessible dynamic information.

Methods

Molecular Systems Preparation dures, Materials and methods, Crystallography

The initial crystal structure for this study was taken from the Protein Data Bank (PDB ID: 3OSF)[16]. We have chosen 2 different mutants K51A, and R87A to stand for the specific interaction and nonspecific interaction. And the R84A mutant is one of the most important mutant which lose all binding affinity in the experimental results. The structure of DNA-free tvMyb2 and the structures of mutants K51A, R84A, and R87A were generated based on the crystal structure using Discovery Studio 2.5 software[17]. All crystal water molecules were considered during MD simulations.

Molecular dynamics simulations

The MD simulations were performed by using AMBER 11 software package[18] with the classical force-field parameters AMBER99SB (ff99SB) force field, supplemented with the refined parmbsc0 parameters (ff99bsc0)[19, 20]. The sodium ions (Na^+) were added by the leap to be the explicit net neutralizing counterions based on a coulomb potential grid. For those complex that were further subjected to MD simulations in explicit solvent, each system was solvated with TIP3P waters in a truncated octahedron box[21], with an 8.0Å distance around the solute. Both the protein and the DNA were fixed with a 100kcal·mol^{-1} Å^{-2} positional restraints, and minimized the energy of the water and the irons for 1000 steps of steepest descent (SD) method and 2000 steps of conjugate gradient (CG) algorithms. Then the minimization was repeated with restraints on the proteins only for 1000 steps SD and 2000 steps CG. The minimization was repeated again for 2000 steps SD and 3000 steps CG without any restraints. Thereafter, the heating dynamic was performed with all restraints by a 10 kcal·mol^{-1} Å^{-2} weights. Thereby, the temperature was increased gently from 50 to 300K and then equilibrated for 300ps. Finally, a 25ns simulation for each system under NPT ensemble condition was carried out. All the systems were treated within periodic boundary conditions. Long rage electrostatic interactions were calculated with the Particle-Mesh Ewald (PME)[22] technique with a non-bonded cutoff of 12.0 Å to limit the direct space sum. An integration time step of 2 fs was used, and the SHAKE algorithm[23] was employed to constrain bonds involving hydrogen atoms during dynamics. All the data given in the tables and figures were obtained from the last 15 ns of the MD simulations unless otherwise mentioned. The PyMOL[24], Chimera[25], and VMD[26] software were used to visualize the trajectories and to depict structural representations.

MM-GBSA Calculations

Binding free energy for each complex was estimated by molecular mechanics generalized Born (MM-GBSA) approaches[27, 28], using the MM/PBSA protocol in the Amber11[18]. We extracted last 1000 snapshots taken from the trajectories of each simulation for the energy calculation. The interaction energy was calculated according to the following equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{DNA}} \quad (1)$$

Here, G_{complex}, G_{protein}, and G_{DNA} are the free energies of complex, protein, and DNA, respectively. The free energy ($G_{\text{x=complex, protein, DNA}}$) of each species can be estimated by using MM-GBSA methods:

$$G_{\text{x=complex, protein, DNA}} = E_{\text{MM}} + G_{\text{solv}} - \text{TS} \quad (2)$$

$$E_{\text{MM}} = E_{\text{ele}} + E_{\text{vdw}} + E_{\text{int}} \quad (3)$$

$$G_{\text{solv}} = G_{\text{gb}} + G_{\text{nonp}} \quad (4)$$

Here, the E_{MM} is the gas phase molecular mechanical energy, G_{solv} is the solvation free energy, and E_{ele}, E_{vdw}, and E_{int} are the electrostatic energy, the van der Waals interaction energy, and the internal energy, respectively. G_{solv} can be separated into an electrostatic solvation energy (G_{gb}) and a nonelectrostatic solvation energy (G_{nonp}). G_{gb} can be calculated with Generalized Born (GB) method. G_{nonp} is considered to be proportional to the molecular solvent accessible surface area (SASA) buried on binding. Normal-mode analysis was performed to estimate the change in conformational entropy upon DNA binding (TS) for all atoms using the nmode module of AMBER11[18]. Because the normal-mode analysis is computational expensive, we only took 100 snapshots from the last 10ns trajectory at an interval of 5 snapshots to calculate the entropy contribution.

For obtaining the detailed view of tvMyb2 and DNA interactions, we employed the MM-GBSA method to calculate the binding free energy of each residue as well. Every single residue can be partitioned into two parts in the calculation, backbone and side

chain. The snapshots used in the binding free energy decomposition are the same as those used in the binding free energy calculation.

RESULTS AND DISCUSSION

Structural and Dynamic Properties of each system

To access the stability of each system during the whole MD simulation, the average root-mean-square deviations (RMSD) of the backbone atoms referenced to the corresponding initial structure of each system was calculated and shown in Figure 2. All the systems achieved equilibrium quickly during the MD simulations except the DNA-free system. The average RMSD values are 1.65Å in 3OSF system, indicating good agreement with the X-ray crystal structures. The mean RMSD value of DNA-free system, which is 3.27Å, is very big compared with the 3OSF system, indicating that the DNA-free tvMyb2 has a very flexible conformation. The results mentioned above are in agreement with the experimental results[16].

To further understand the stability difference between the protein in 3OSF system and the protein in DNA-free system, we have analysed the cross-correlation of the protein in these two systems. In Figure 3, we have defined three regions R2-R2, R3-R3 and R2-R3. The R2 and R3 are the repeat 2 (amino acids 52-102) and repeat 3 (amino acids 103-154) of the tvMyb2. The R2-R2 means that the cross-correlation between amino acids of the R2 for instance. As shown in Figure 3(a), there are two high-relative regions which are R2-R2 and R3-R3, and one low-relative region R2-R3, indicating that the two repeat, R2 and R3, are independent from each other in the DNA-free system. In Figure 3(b), three regions, R2-R2, R3-R3 and R2-R3, are all high-relative, which means that the tvMyb2 protein is very stable in the 3OSF system. What's more, in both figures, the R2-R2 and R3-R3 are high-relative no matter if the protein bind the ap65-1 molecular or not.

To further illustrate the structural difference in every system, we analysed the secondary structure. See Figure 4, all the systems show six helixes. But for the R84A system, the fifth helix unwind after 10ns simulation, but the residue which we mutant, Arg84, locate on the third helix, which indicate that the mutation may cause the unwind processes by a long-range interaction. This unwinding leads us to conclude the Arg84 is very important for the stability of the binding system.

In addition, we analysed the extent of variation in flexibility of individual residue/nucleotide upon binding of each system to analysis the mobility of structural elements in each system, the per-residue/nucleotide root-mean-square fluctuations (RMSFs) of the protein Cα and DNA backbone with respect to the initial coordinates were computed over last 15ns simulations are shown in Figure 5. For the 3OSF system, the value is only high on the boundary of the protein and the loop between the R2 and R3. The DNA-free system is in very high flexibility compared with the 3OSF system from Figure 5. On the contrary, the value of the DNA-free system is high all the time, and only decreases a little on the middle of every helix. The RMSF values of the 3OSF and the mutant systems have approximate value of residues and nucleotides. The residues in R2 and R3 regions have low flexibility, and most high values of RMSF happened on the residues in loop region. After analysis the RMSF value of the protein of each system, we have also further analysis the RMSF for the DNA in the 3OSF and mutant systems. The RMSF value of the DNA is stable as well as the protein for the systems we analysed.

Energetic Analysis of the TvMyb2- MRE-2-13 complexes

Binding free energies analysis. Among the several solvation models, GBSA is regarded as an attractive approach for the biological molecules especially for the energy between protein and DNA. So the binding free energies for the complex systems were evaluated by MM-GBSA methodology, and run the decomposition analysis to obtain insights into the affinity between the protein and the DNA. The snapshot structure used for free energy calculations were extracted from the last 15 ns trajectory. The detailed contributions of various energy components computed by MM-GBSA and the entropy contributions from the normal-mode analysis are given in Table 1. As can be seen in Table 1, the binding free energies of the 3OSF system are estimated to be -86.21 kcal·mol^{-1}, and the K51A, R84A and R87A system are respectively estimated to be -75.03 kcal·mol^{-1}, -51.86 kcal·mol^{-1} and -61.07 kcal·mol^{-1}. The predicated results by MM-GBSA of the binding energies are not absolutely same to the experimental result, the different are caused by the deviation of MM-GBSA methods, and all the deviation are in acceptable range. However, all the results calculated by MM-GBSA have a similar tendency to the experimental results.

To further investigate the source of binding affinity as well as the loss of binding affinity in mutants, the detailed contributions of various energy components computed by MM-GBSA and the entropy contributions from the normal-mode analysis are given (Table 1). The major favourable contributions to the binding free energies (G_{bind}) come from the nonpolar solvation energies (G_{nonpol}), more specifically from the van der Waals energies (E_{vdw}). In contrast to the nonpolar solvation energies (G_{nonpol}), the polar solvation energies (G_{pol}) make unfavourable contribution to the binding energy. Actually, the direct intermolecular electrostatic interactions (E_{ele}) are highly favorable to the binding but their contributions are completely screened by the unfavourable stronger polar-electrostatic solvation energies (G_{gb}). In addition, the contributions of entropy changes (-TS) to the free energies impair the binding of the ap65-1 molecule to the protein tvMyb2. All the entropy change followed are stand for the –TS. In general, the favourable van der Waals (E_{vdw}) and nonelectrostatic solvation energies (G_{nonp}) mainly drive the binding

between the tvMyb2 protein and ap65-1. All the mutant systems have a very similar nonpolar energy, which is the major attractive interactions, and the nonpolar energy of all the mutant systems have increased 15 to 20 kcal·mol^{-1} compared with the 3OSF system. For K51A system, the main source for the binding affinity lost is that the nonpolar energy has increased about 15 kcal·mol^{-1}. For the R87A system, not only the nonpolar energy has increased, but also the entropy has increased about 10 kcal·mol^{-1} at the same time. For the R84A system, the polar energy also increase 10 kcal·mol^{-1}, except the values of the entropy and nonpolar energy have increased, compared with the 3OSF system. We could find that the R84A system getting unstable from every aspect by analysing the energy results. It happens that there is a similar case in the experiment results, the R84A system lose the binding affinity at all. And once more, the results we have got are in a same trend with the experiment up to a point[16].

A much better understanding of the tvMyb2–ap65-1 recognition is obtained via the interaction energy analysis. As it is shown in Table S1, the attractive interactions between tvMyb2 protein and the bases of ap65-1 are stronger than the corresponding interactions between the protein and the phosphate backbone. We could infer that the tvMyb2 is a sequence specific DNA binding protein.

Per-residual/nucleotide energy decomposition. To characterize and identify the key residues which are important for the recognition and binding of the tvMyb2-ap65-1 system, the free energy decomposition on residual basis was performed through MM-GBSA (Figure 6).

The values of $\Delta G_{MM+solv}$ had been decomposited on a per-residue basis into contributions from van der Waals energy, columbic interactions energy, polar solvation free energy, and the nonpolar solvation free energy. The residues with $\Delta G_{MM+solv} < -3\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ in the each system have been listed in TableS2.1~S2.4. In 3OSF system (Table S2.1), seven amino acid residues (Lys49, Lys51, Gln85, Arg87, Arg89, Arg120, and Trp122) make major contributions to the binding free energy with $\Delta G_{MM+solv} < -5\mathrm{kcal}\cdot\mathrm{mol}^{-1}$. These residues are very important for the tvMyb2-MRE-2-13 binding, which are in agreement with the previous experimental identification[16]. As seen from the Table S1.1 and Table S2.2-S2.4, three mutant systems have very different ways, for being unstable with the DNA molecular. In K51A system, the decomposition energy shows that the DNA molecular is the main reason for the unstable binding. See Table S1.2 ~ S1.3, the decomposition energy of the base of the K51A system is very similar to the 3OSF system, but the value of the backbone have risen about *10* kcal·mol^{-1}, some of the backbone of nucleotides (A(2'),T(3'),C(4')) have become to repel protein instead of attracting it. The real reason of the risen of the value of the binding free energy for K51A system compared with 3OSF system is that the mutation of Lys51 to Ala51 destroys the electronic attraction between the residues of protein and the backbone of the ap65-1. But in another word, this mutant influence little of the interaction between tvMyb2 and the base of DNA molecule which is the major recognition way. In R84A system, the reason of the unstable is very complicated. Just like the K51A system, the backbone of the DNA molecular repel protein, but at the same time, the decomposition energy of base have risen too, and what is really important is that the G (5'), which is the center of the specific base sequence, lose its binding affinity to the protein, it mean that the protein lose its recognize ability. The decomposition energy of R84A system could give another clue to the detail information of losing ability to recognize. Except Ala84 lose all the binding affinity because of the mutant, the Gln50 do not contribute the binding anymore, either. And some other residues (Lys49, Arg81, Lys128 and Asn136) have a significant rise of the value of decomposition energy. In R87A system, some of the residues (Lys49, Gln51, Arg81, Ala87, Lys91, Gln121, Trp122, Ala123 and Lys138) contribute less compare than the 3OSF system, and three of them, Lys49, Ala87 and Trp122 make the most important residues in the contribution of 3OSF system.

From TableS2.1~S2.4, it can be seen that most of the important residues of each system have a large polar interaction contribution while the nonpolar interaction has a relatively small influence. But for some of the other residues, the nonpolar interactions also have an obvious contribution, for example, Lys49, Arg84 and so on. Most of these polar interactions stem from the formation of hydrogen bonds, and the nonpolar interactions are come from the hydrophobic interactions. The detailed interaction between tvMyb2 protein and DNA molecular can be observed by monitoring the occupation probability of these hydrogen bonds and hydrophobic interactions during the MD process.

Intermolecular Interaction Analysis.

The protein-DNA interactions were investigated by focusing on hydrogen bonds and hydrophobic interactions. Protein-DNA hydrogen bonds include both nonspecific interactions (between tvMyb2 and DNA sugar/phosphate backbone), being significant for the stability of protein-DNA complex, and specific interactions (between tvMyb2 and DNA base), being important to the molecular recognition process in protein-DNA complex. Moreover, the hydrophobic contacts are necessary for the protein-DNA binding, but most of these interactions are nonspecific because the interactions are between residues and sugar backbone mostly. The criteria for hydrogen-bonded pairs are the distance shorter than 3.5Å and the binding angle greater than 120.0° between the hydrogen donor and acceptor atoms. The hydrophobic contact is defined as a distance between carbon atoms shorter than 4.5 Å in the present study.

All the complex systems become stable during the whole MD simulation, however, we could clearly find that 3 mutant systems are not as stable as the 3OSF system. To investigate the interaction characteristics between tvMyb2 and DNA and find the reason why the stability difference, both the hydrogen bonds and the hydrophobic interaction formed between tvMyb2 and ap65-1 in all complex systems were analyzed. By all these studies, we have drawn the contact maps of each complex, see Figure

7 and Figure 8. There are more stable hydrogen bonds and hydrophobic contacts in the 3OSF system than K51A, R84A or R87A systems. In the 3OSF system, there are 14 hydrogen bonds and 11 hydrophobic interactions between protein and DNA. However, the number of hydrogen bonds and hydrophobic interactions become less in three mutants K51A, R84A and R87A. For 3OSF system, Lys51, Lys49, Arg84, Tyr93 and Lys138 are involved in stable specific hydrogen bonds (Figure S1) between tvMyb2 and DNA. For K51A system, the hydrogen bond Lys51 - T (1') and hydrophobic contacts Lys51 – T (3') have been disappeared because of the mutation, and the Thr143 – G (-1') hydrogen bond have disappeared as well. Moreover, for R84A system, a lot of hydrogen bonds and hydrophobic interactions have been very unstable. We did not surprise by this result because of the experiment had already shown that the R84A system is very unstable[16]. Gln85, Arg87, Lys91, Ala123 and Thr143 lose their interactions with DNA molecular at all. For the last system, R87A, we found that all the hydrogen bonds or the hydrophobic interactions which are related to the Arg84 have been vanished or become not stable. Moreover, the hydrogen bonds Gln85 – C (3') and Lys91 – A (-2) have become unstable, and hydrophobic interactions Lys49 – T (3'), Lys138 – G (3) have totally disrupt either.

After analyzing the interaction network between the tvMyb2 protein and ap65-1 DNA molecular, we could infer that maybe the Lys49, Lys51 and Arg84 are the core residues for the recognizing, no matter from the energy aspect or the interaction aspect. All these residues involve with the specific interactions, and the decomposition energy of these residues make major contribution to the binding free energy with less than -3 kcal·mol^{-1}. Furthermore, the Arg87 is one of the important residue which form the nonspecific interaction between the phosphate backbone of the DNA molecular and the residue (see Figure7). Although the nonspecific interactions are not the most important interaction for the recognition for tvMyb2-DNA, we could clear find that the mutation of the Arg87 destroy not only the nonspecific interaction, but also a serial of the specific interaction. After all, the binding of the tvMyb2 and ap65-1 DNA molecular consist the specific interactions and the nonspecific interaction, and each of them are important to the binding interaction network.

CONCLUSION

Focused on the specific tvMyb2-MRE-2-13 DNA complex, the DNA-free tvMyb2, and 3 mutant complexes, we performed Molecular dynamic (MD) simulation to investigate the structural stability. Furthermore, MM-GBSA was used to analyze the detailed interaction and binding free energy in all 4 complexes systems.

The results from MD simulations indicate that the tvMyb2 protein in the DNA-free system are not as stable as the protein in the tvMyb2-MRE-2-13 system. Although the tvMyb2 are unstable and more flexibility when the DNA molecular are not binding with it, the R2 and R3 region are comparatively stable whenever the DNA binding or not. Furthermore, all four complexes systems are stable in the MD simulations, only slight structural changes of the protein and DNA molecular. But the R84A system suffers an unwinding of the fifth helix. It because of the mutant cause the DNA molecular undergo a slight structure change, but the slight change drag the residue Trp122, which locate on the fifth helix of the protein, away from the initial position by forming hydrogen bond Trp122-C (2) and hydrophobic contact Trp122-(2) with the DNA molecular, and the position change of the Trp122 may cause the unstable of the second structure. The structure change is one of the most important reason why the R84A system lose binding affinity compared with the 3OSF system.

MM-GBSA free energy calculations provide the following energetic information: the binding affinity for each complexes shows that the 3OSF protein binds DNA molecular best, and all 3 mutant complexes lose part of the binding affinity. The sequence of binding affinity for all 3 mutant systems are consist with the experimental results[16], which the K51A is the best, the R87A followed and the R84A is the worst. The extensive interactions between tvMyb2 and DNA base pair are the reason specific DNA-binding protein tvMyb2 is proposed, and the dominant contribution to the binding free energy of the tvMyb2-DNA is the nonpolar interaction. Seven conserved amino acid residues (Lys49, Lys51, Gln85, Arg87, Arg89, Arg120, and Trp122) make major contributions to the binding free energy with more than $5\,kcal/mol$ free energy for the tvMyb2-MRE-2-13 binding. Furthermore, the hydrogen bonds and hydrophobic interactions analysis suggest the key contact sites for the two complexes. At last, 3 key residues (Lys49, Lys51 and Arg84) for recognizing of tvMyb2-ap65-1 have been suggest which combined the energy analysis, the interaction analysis and the mutant results. MD simulations and MM-GBSA calculations provide the above dynamic structural information and energetic information, which are often inaccessible in the static crystal structure. This study will lead to a better understanding of tvMyb2-ap65-1 interactions.

Acknowledgements

This work is supported by Natural Science Foundation of China (Grant Nos. 21273095).

Keyword

tvMyb2; ap65-1; Molecular dynamic;

References

1. R. G. Ramsay and T. J. Gonda, *Nature reviews. Cancer*, 2008, **8**, 523-534.
2. M. R. Lidonnici, F. Corradini, T. Waldron, T. P. Bender and B. Calabretta, *Blood*, 2008, **111**, 4771-4779.
3. G. Anfossi, A. M. Gewirtz and B. Calabretta, *Proceedings of the National Academy of Sciences*, 1989, **86**, 3379-3383.
4. H. Hugo, A. Cures, N. Suraweera, Y. Drabsch, D. Purcell, T. Mantamadiotis, W. Phillips, A. Dobrovic, G. Zupi, T. J. Gonda, B. Iacopetta and R. G. Ramsay, *Genes, Chromosomes and Cancer*, 2006, **45**, 1143-1154.
5. K.-I. C. Ogata K, Sasaki M, Hatanaka H, Nagadoi A, Enari M, Nakamura H, Nishimura Y, Ishii S, Sarai A., *Nature Structural Biology* 1996, **[1996, 3(2):178-187]**.
6. L. Jia, M. T. Clegg and T. Jiang, *Plant physiology*, 2004, **134**, 575-585.
7. H. Biedenkapp, U. Borgmeyer, A. E. Sippel and K.-H. Klempnauer, *Nature*, 1988, **335**, 835-837.
8. F. Sorvillo, L. Smith, P. Kerndt and L. Ash, *Emerging infectious diseases*, 2001, **7**, 927.
9. M. A. Klebanoff, J. C. Carey, J. C. Hauth, S. L. Hillier, R. P. Nugent, E. A. Thom, J. M. Ernest, R. P. Heine, R. J. Wapner, W. Trout, A. Moawad, M. Miodovnik, B. M. Sibai, J. P. V. Dorsten, M. P. Dombrowski, M. J. O'Sullivan, M. Varner, O. Langer, D. McNellis, J. M. Roberts and K. J. Leveno, *New England Journal of Medicine*, 2001, **345**, 487-493.
10. J. F. Alderete and G. E. Garza, *Infection and immunity*, 1988, **56**, 28-33.
11. J. Alderete, J. L. O'Brien, R. Arroyo, J. A. Engbring, O. Musatovova, O. Lopez, C. Lauriano and J. Nguyen, *Molecular microbiology*, 1995, **17**, 69-83.
12. R. Arroyo, J. Engbring and J. Alderete, *Molecular microbiology*, 1992, **6**, 853-862.
13. C.-D. Tsai, H.-W. Liu and J.-H. Tai, *Journal of Biological Chemistry*, 2002, **277**, 5153-5162.
14. S. J. Ong, S. C. Huang, H. W. Liu and J. H. Tai, *Molecular microbiology*, 2004, **52**, 1721-1730.
15. H.-M. Hsu, S.-J. Ong, M.-C. Lee and J.-H. Tai, *Eukaryotic cell*, 2009, **8**, 362-372.
16. I. Jiang, C.-K. Tsai, S.-C. Chen, S.-h. Wang, I. Amiraslanov, C.-F. Chang, W.-J. Wu, J.-H. Tai, Y.-C. Liaw and T.-h. Huang, *Nucleic acids research*, 2011, **39**, 8992-9008.
17. D. Studio, *Accelrys Inc.: San Diego, CA, USA*, 2009.
18. C. Simmerling, J. Wang, R. Duke, R. Luo, W. Zhang, K. Merz and B. Roberts, *There is no corresponding record for this reference*.
19. J. Wang, P. Cieplak and P. A. Kollman, *Journal of Computational Chemistry*, 2000, **21**, 1049-1074.
20. A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophysical journal*, 2007, **92**, 3817-3829.
21. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of chemical physics*, 1983, **79**, 926-935.
22. T. Darden, D. York and L. Pedersen, *The Journal of chemical physics*, 1993, **98**, 10089-10092.
23. J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *Journal of Computational Physics*, 1977, **23**, 327-341.
24. W. DeLano, *There is no corresponding record for this reference*, 2002.
25. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of computational chemistry*, 2004, **25**, 1605-1612.
26. W. Humphrey, A. Dalke and K. Schulten, *Journal of molecular graphics*, 1996, **14**, 33-38.
27. J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *Journal of the American Chemical Society*, 1998, **120**, 9401-9409.
28. J. M. Swanson, R. H. Henchman and J. A. McCammon, *Biophysical journal*, 2004, **86**, 67-74.

Figure caption:

Figure 1. The overall structure of (A)3OSF complex and (B)the mutant residues.

Figure 2. Time dependence of the RMSD values (Å) of molecules of each system.

Figure 3. Cross-correlation matrices of the coordinate fluctuations for Cα atoms of (a) 3OSF and (b) DNA-free proteins around their mean positions.

Figure 4. Secondary structure variation along MD simulations of 3OSF, K51A, R84A and R87A mutants.

Figure 5. The RMSF per residues of protein (a) and nucleotides of the DNA (b) based on the initial structure.

Figure 6. Free energy decomposition analysis for 3OSF and three mutants.

Figure 7. Diagrams of the calculated key contact sites of tvMyb2 with ap65-2 DNA in different systems.

Figure 8. Hydrogen bonds occupancies along the simulation trajectories for every systems.

Table.1 Binding energy between tvMyb2 and ap65-1 in each complexes systems

(Kcal·mol^{-1})

	3OSF	K51A	R84A	R87A
E_{ele}	**-4892.54±69.18**	**-4354.28±78.68**	**-4196.02±69.51**	**-4397.73±77.76**
E_{vdw}	**-122.39±6.46**	**-109.66±6.25**	**-108.07±6.49**	**-102.88±6.84**
G_{gb}	**4901.12±65.54**	**4359.20±75.84**	**4213.57±66.47**	**4401.15±72.82**
G_{nonp}	**-20.39±0.34**	**-19.28±0.48**	**-18.94±0.70**	**-19.53±0.51**
G_{pol}[a]	**8.58**	**4.92**	**17.55**	**3.42**
G_{nonpol}[b]	**-142.78**	**-128.94**	**-127.01**	**-122.41**
H	**-134.20±8.57**	**-124.03±9.36**	**-109.46±8.17**	**-118.99±10.95**
-TS	**47.99±5.34**	**49.00±10.46**	**57.60±5.92**	**57.92±2.25**
G_{bind}[c]	**-86.21±10.10**	**-75.03±14.04**	**-51.86±10.09**	**-61.07±11.18**

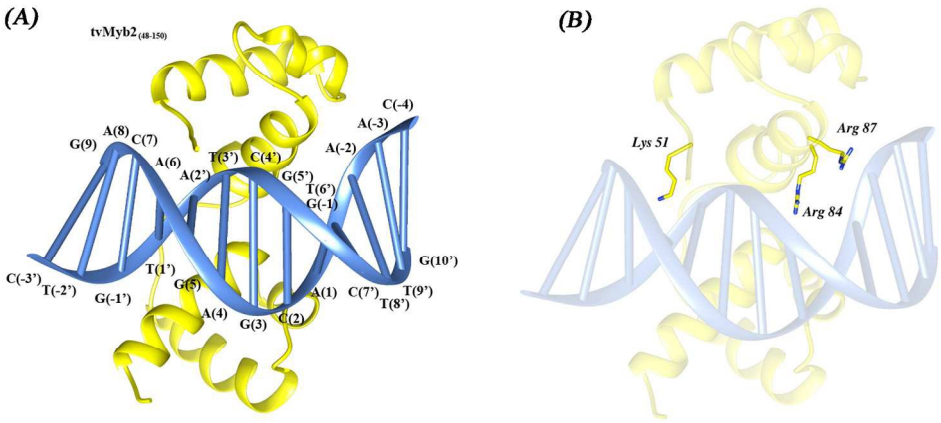
[a] $G_{pol} = E_{ele} + G_{gb}$. [b] $G_{nonpol} = E_{vdw} + G_{nonp}$. [c] $G_{bind} = G_{nonp} + G_{pb} - TS$.

Example caption

Table S1: Decomposition of the DNA molecular, include whole DNA, backbone and base.

Table S2: Decomposition energy of the important residues in each system.

Figure.S1 The important hydrogen bonds between tvMyb2 and ap65-1 in 3OSF system.



181x82mm (300 x 300 DPI)

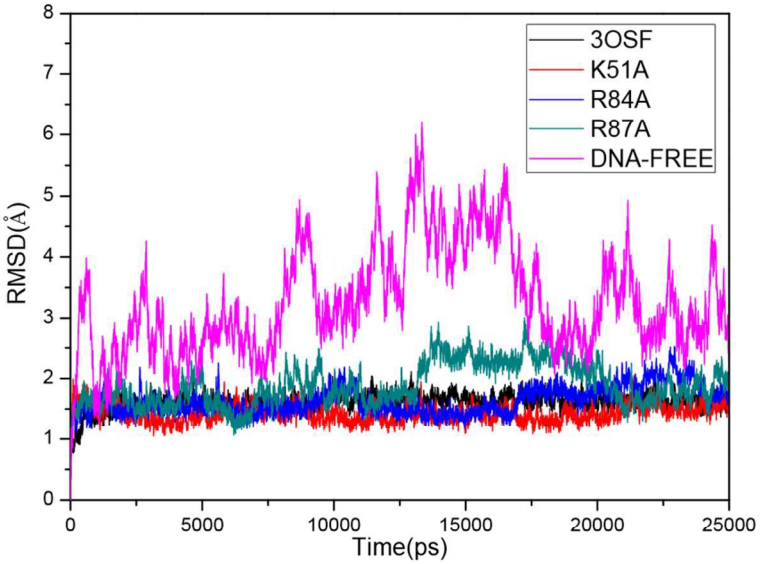


Figure 2. Time dependence of the RMSD values (Å) of molecules of each system. The black line correspond to RMSD of the 3OSF system. The red lines represent RMSD of K51A system. The blue line show RMSD of the R84A system. The dark cyan lines show RMSD of R87A system. The magenta lines show RMSD of DNA-free system.
300x212mm (95 x 95 DPI)

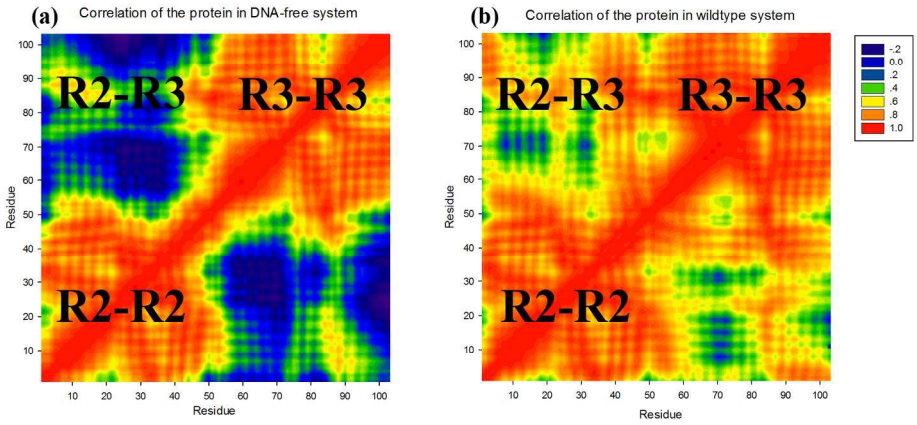


Figure 3. Cross-correlation matrices of the coordinate fluctuations for Cα atoms of (a) 3OSF and (b) DNA-free proteins around their mean positions. The extents of correlated motions and anticorrelated motions are labeled in different colors. Each picture has been divided into 3 region, R2-R2, R3-R3 and R2-R3.
180x86mm (299 x 299 DPI)

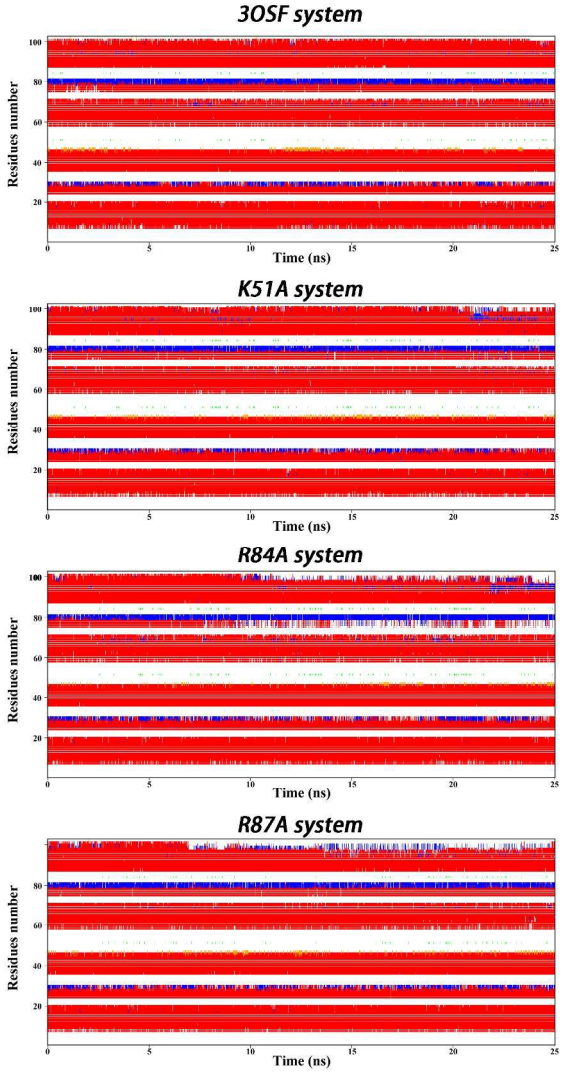


Figure 4. Secondary structure variation along MD simulations of 3OSF, K51A, R84A and R87A mutants. Residues in α-helix and 3-10 helix are shown in red and blue respectively. Residues in nonregular are shown in white.
874x1237mm (72 x 72 DPI)

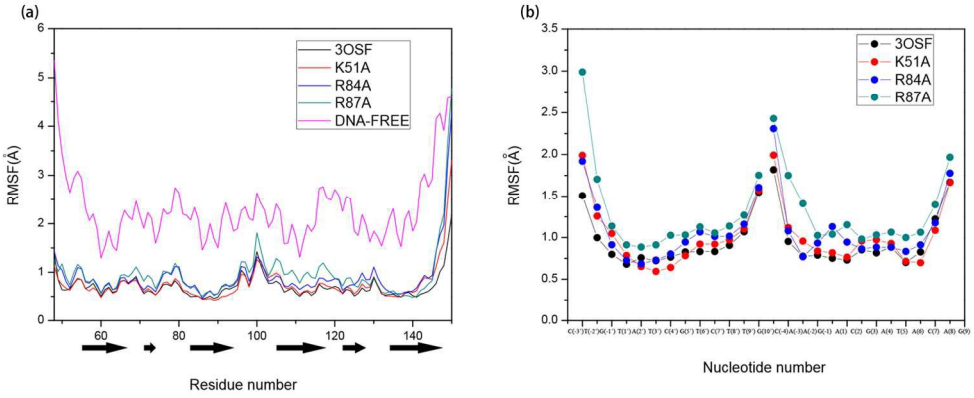


Figure 5. The RMSF per residues of protein (a) and nucleotides of the DNA (b) based on the initial structure. The RMSF values of free protein and protein/DNA in complex were shown with different color lines, respectively. The α-helixs are represented by black block arrows.
136x56mm (299 x 299 DPI)

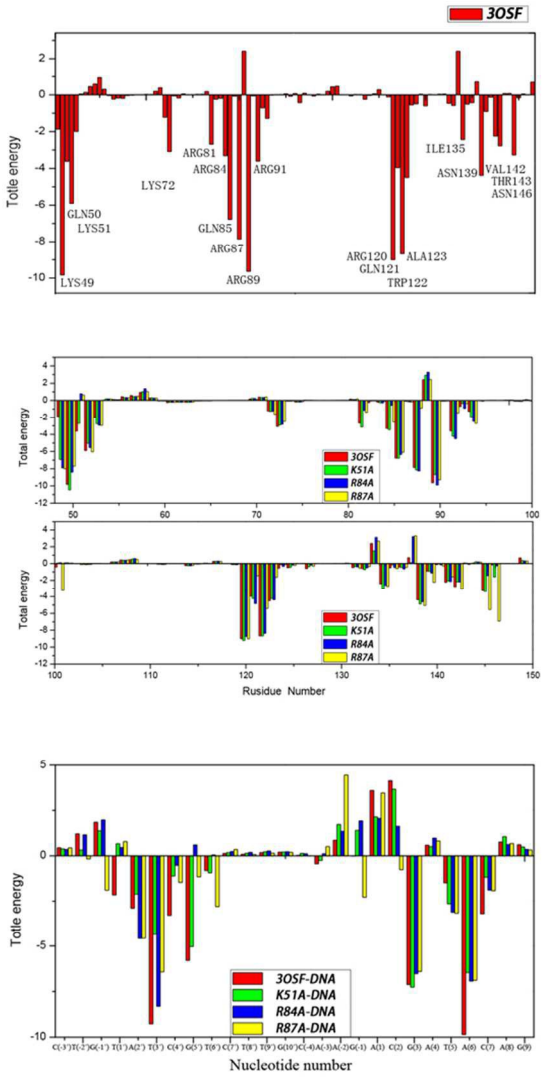


Figure 6. Free energy decomposition analysis for 3OSF and three mutants; residues contributing significantly for 3OSF are highlighted.
212x415mm (72 x 72 DPI)

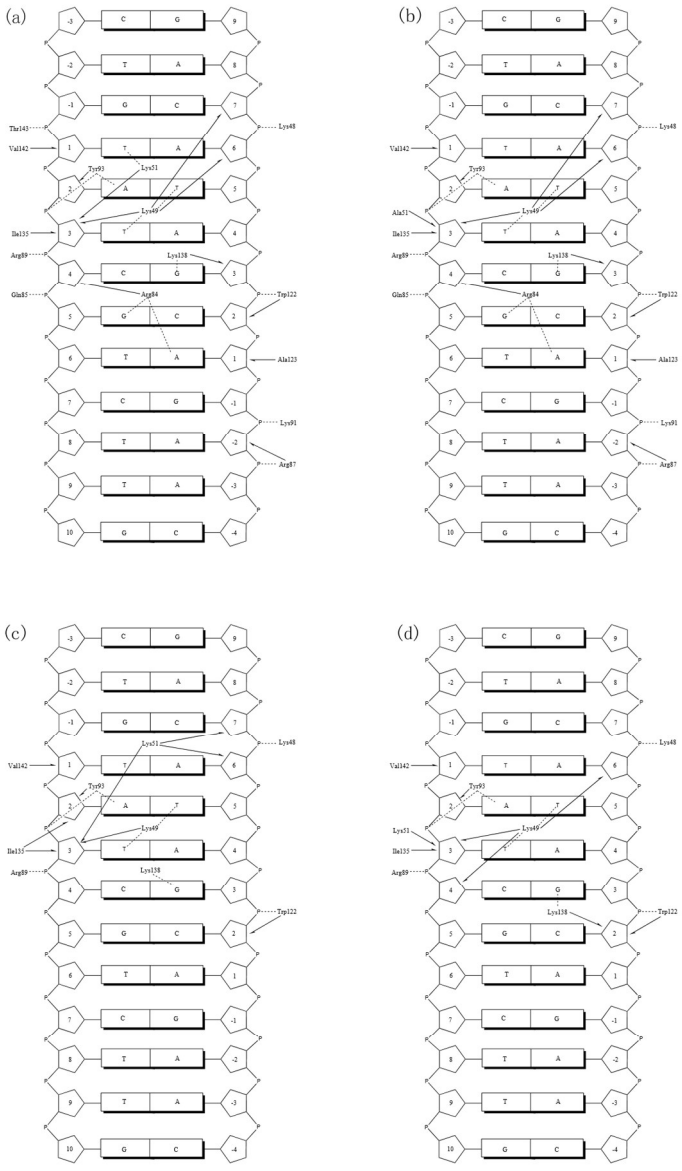


Figure 7. Diagrams of the calculated key contact sites of tvMyb2 with ap65-2 DNA in different systems, (a) the 3OSF system (b) the K51A system (c) the R84A system (d) the R87A system. Hydrogen bonds and hydrophobic interactions are shown with dashed lines and arrows, respectively.
421x720mm (72 x 72 DPI)

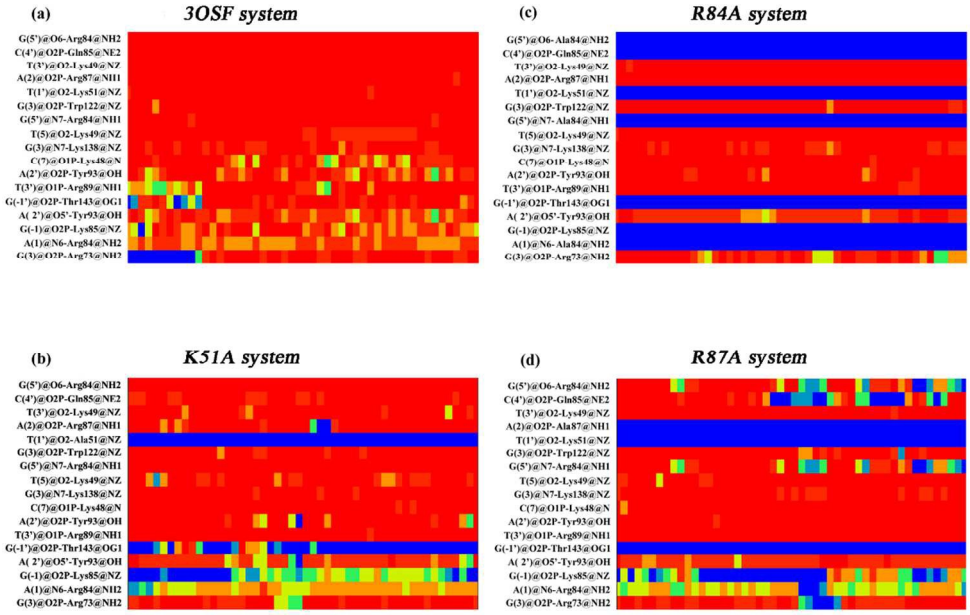


Figure 8. Hydrogen bonds occupancies along the simulation trajectories for 3OSF-DNA, K51A-DNA, R84A-DNA and R87A-DNA complexes.
146x97mm (219 x 219 DPI)