Cross-correlation analysis to salt-bridge dynamics in force-induced unfolding of titin kinase

Ming-Chya Wu*
Research Center for Adaptive Data Analysis,
National Central University,
Chungli 32001, Taiwan
Email: mcwu@ncu.edu.tw

Jeffrey G. Forbes
National Institute of Arthritis and
Musculoskeletal and Skin Diseases,
National Institutes of Health,
Bethesda, Maryland 20892, USA

Kuan Wang
Institute of Biological Chemistry,
Academia Sinica, Nankang,
Taipei 11529, Taiwan

***Abstract*—In this paper, a theoretical study on the salt-bridge dynamics of titin kinase is presented. We focus on the analysis of the spatial-temporal properties of the salt-bridge time series of titin kinase in force-induced unfolding simulated by steered molecular dynamics (SMD). Salt-bridge time series are defined from the SMD trajectories. Scaling analysis reveals two characteristics of the time series in short and long time scales, suggesting there is anti-persistent behavior in short-time scale less than 50 ps, while persistent behavior dominates in long-time scale larger than 100 ps. Using cross-correlation analysis, we study the dynamics of the salt-bridges. From analyzing the eigenvectors of the cross-correlation matrix constructed by the salt-bridge data, we classify salt-bridges into distinct groups. The knowledge of the grouping is useful in identifying force-relevant structural transitions.**

I. Introduction

Single molecule manipulation techniques have made possible experimental access to unfolding intermediates of proteins. It provides a route to investigate the physics at single molecule level and of a macroscopic system in thermodynamic limit. Such techniques usually operate in a typical time scale of 10^{-3} sec, which is much larger than the typical oscillatory frequency of molecules in the time scale of 10^{-12} sec. Detailed structures of the unfolding characters are unavailable in these experiments due to the limit of time resolution. Numerical simulations then have been used to work accompanying with such techniques. Among others, molecular dynamics (MD) is a practical tool to carry out simulations at desired resolution.

In this paper, we study the salt-bridge dynamics in the force-induced unfolding of human titin kinase simulated by steered molecular dynamics (SMD). Titin kinase is a domain at the sarcomeric M-band of the giant elastic protein titin. It is a catalytic domain consisting of 321 amino acid residues. Human titin kinase acts as a force sensor in biological responses to force-induced stress [1]. It converts force-induced stress into biochemical signal in a muscle cell. Such force-induced resistance is associated with the formation of salt-bridges in molecules [2], which contribute persistent features of force curve observed in force extension simulations.

By studying the SMD trajectories in the first 10000 ps simulation with pulling velocity of 0.0005 nm/ps, we define 55 salt-bridge time series which manifest properties of dynamic breaking and rejoining of the salt bridges. We analyze the spatial-temporal properties of these time series by scaling, and cross-correlation analyses. Scaling analysis of detrended fluctuation analysis (DFA) reveals the temporal correlation of the salt-bridge time series, while cross-correlation analysis is useful for understanding the grouping structure of the salt-bridge dynamics. With this knowledge, crucial set of salt-bridges to force-relevant structural transitions can be defined by selecting specified functionalities as references.

The rest of this paper is organized as follows. In Sec. II, we introduce the SMD simulation and the parameters used to simulate the force-induced unfolding of titin kinase in this study. Salt-bridge time series are defined from SMD trajectories. The analyses of temporal correlation and cross-correlation are respectively presented in Sec. III and Sec. IV. Finally, we summarize our results in Sec.V.

II. Steered molecular dynamics simulations

We used the NAMD package [3] with the CHARMM22 force field [4] to implement the SMD simulations. The SMD method involves the application of external forces to molecules in MD simulations. The simulations started from an experimentally solved X-ray structure of titin kinase, 1TKI, in the Protein Data Bank [5]. The structure of titin kinase was first solvated with the TIP3P model for water [6]. A water box was then constructed by using the water box module in the NAMD package. The titin kinase was placed in one-side of the water box. The resulting structure is a titin kinase molecule surrounded by a water box, which has the protein surface covered everywhere by water molecules throughout the simulation.

The water-protein system was gradually heated over 10 ps to 296.5 K, then equilibrated with a thermal bath at 296.5 K for another 10 ps. The free dynamics run exhibited a temperature fluctuation of 5 K and an RMSD of 1 Å from the backbone of the X-ray structure. Finally, 2 ps of free dynamics was performed. The same procedure was applied to solvate the titin kinase. The simulations were performed with a time step of 1 fs, a uniform dielectric constant of 1, and a cut-off of Coulomb forces with a switching function starting at a distance of 8 Å and reaching zero at 12 Å. SMD simulations were carried out by fixing one terminus of the domain, and applying external forces to the other terminus. The forces were applied

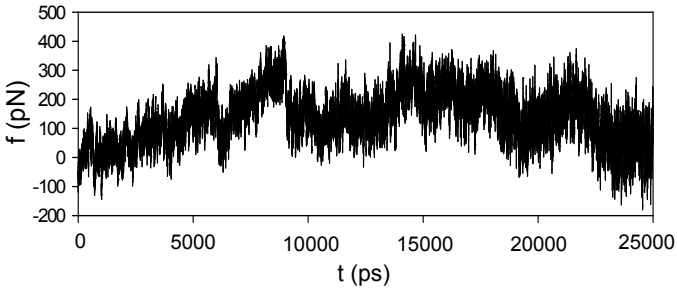

Fig. 1. A typical force curve of the force-induced unfolding of titin kinase simulated by 25000 ps SMD simulation with pulling speed 0.0005 nm/ps.

by restraining the pulled end to a restraint point and moving the restraint point with a constant velocity $v = 0.0005$ nm/ps in the desired direction. The forces experienced by the C_α atom of a pulled residue are

$$f = k\,(vt - y)\,. \tag{1}$$

Here y is the displacement of the pulled atom from its original position, and k is the spring constant. The pulling direction was chosen along the vector from fixed atom to pulled atom. The value of k was set at 10 $k_BT/$Å, corresponding to a spatial (thermal) fluctuation of the constrained C_α atom of $\delta x = \sqrt{k_BT/k} = 0.32$ Å at $T = 300$K. The trajectories of all atoms are stored. Figure 1 shows a typical force curve of the force-induced unfolding of titin kinase simulated by 25000 ps SMD with pulling speed $v = 0.0005$ nm/ps.

We used the VMD program [7] to analyze the output data. The salt-bridges in titin kinase are defined by an oxygen-nitrogen distance cut-off of 3.3Å. Here, an interaction between a pair of amino acids is considered as a salt-bridge as the distance of their oxygen-nitrogen pair is less than the cut-off during unfolding. The evolutions of the distances of these salt-bridges are calculated from the recording of the SMD trajectories. In one realization of the simulation, there are 55 salt-bridges involved in the unfolding of titin kinase. As titin kinase acts as a force sensor and functions when it is stretched, here we are particularly interested in the simulation of the first 10000 ps, in which the system is not stretched too much to lose its main structure. For simplicity, we use a three-letter symbol and three digits to indicate an amino acid residue and its location in linear sequence. Thus, notation GLU068 indicates the Glutamic acid at position 68. To explore properties of the salt-bridge time series, we consider each salt-bridge time series as a walk and define the corresponding return. For a walk $x(t)$, the return of the walk is defined as

$$r_\tau(t) = \frac{x(t+\tau) - x(t)}{x(t)}, \tag{2}$$

where τ is a multiple of the primary time sampling unit $\Delta t(= 1)$.

III. Temporal Correlation in Salt-Bridge Dynamics

In the following, we will focus on the data of $r_{\tau=1}(t)$, which will then be expressed as $r(t)$ for simplicity. We use DFA

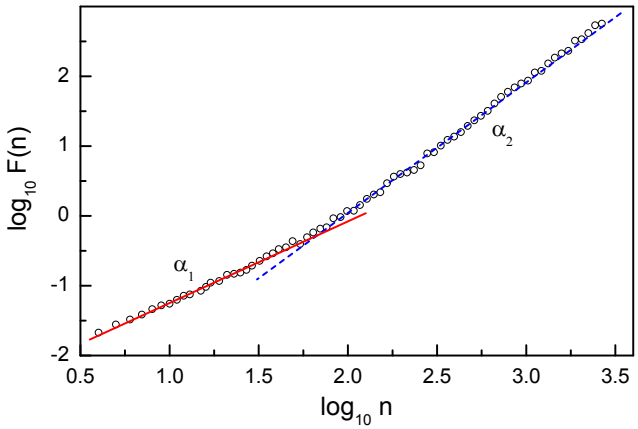

Fig. 2. Log-log plot for $F(n)$ as a function n, for the salt-bridge GLU068-LYS067. The slopes determined by linear fittings are $\alpha_1 = 1.11$ and $\alpha_2 = 1.83$.

[8] to assess the temporal correlation of the salt-bridge time series. The DFA is a analysis based on the measurement for fluctuation F of a time series in different scales. For a time series $r(t)$, the algorithm of DFA is given by [8]

$$F\,(n) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[r(t) - r_n(t)\right]^2} \propto n^\alpha, \tag{3}$$

where n is a scale factor for resampling $r(t)$, and $r_n(t)$ is the straight line with the slope determined by the best fitting of the n data points. The scale factor n is a moving window in the algorithm, and Eq.(3) describes the dependence of the fluctuation on the window size. Hence, the slope for the curve in the plot of F as a function of n in log-log scale is associated with the intrinsic correlation properties of the time series, characterized by the index α [8]. The DFA result for the salt-bridge GLU068-LYS067 is shown in Fig. 2. There are two characteristics of the time series in short and long time scales, respectively contributing to the *linear* parts in the two regimes. In short-time scale less than 50 ps, there is anti-persistent behavior such that $\alpha_1 = 1.11$, while persistent behavior dominates in long-time scale larger than 100 ps, with $\alpha_2 = 1.83$. We performed the same analysis to all salt-bridge data, and we have, in average, $\langle\alpha_1\rangle = 1.16$ and $\langle\alpha_2\rangle = 1.81$. These results show that for salt-bridge time series the effect of thermal fluctuation dominates in short-time scale, while the interactions associated with secondary structures is crucial in long-time scale.

IV. Cross-Correlation Analysis

Next, we study the spatial correlation among salt-bridges by using cross-correlation analysis [9], [10]. The salt-bridge time series $x(t)$ is here considered as a vector in a finite-dimensional space, with the dimension defined by the number of the salt-bridges. As the system is closed, the basis constructed from all the salt-bridges is complete. To implement the construction, we define a cross-correlation matrix from the salt-bridge time series. The fluctuation of the return time series

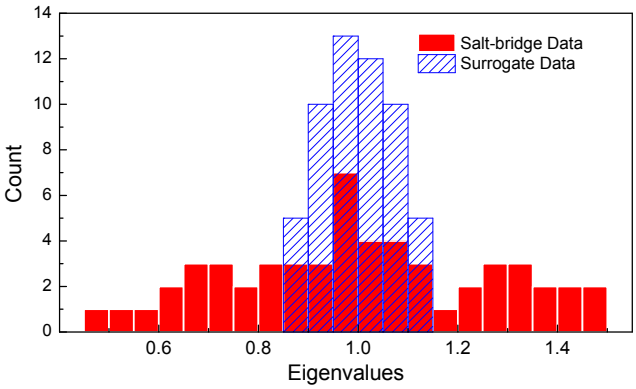

Fig. 3. Eigenvalue distributions for salt-bridge data and the surrogated data from an ensemble average of 10 realizations of shuffled data.

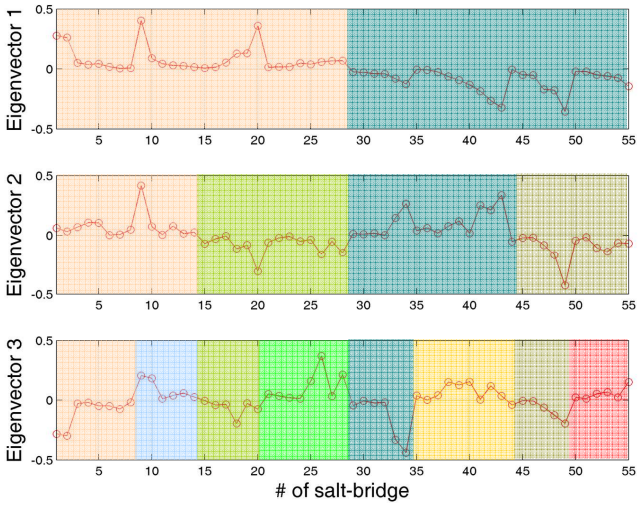

Fig. 4. Eigenvectors 1, 2 and 3 with grouping rearrangement.

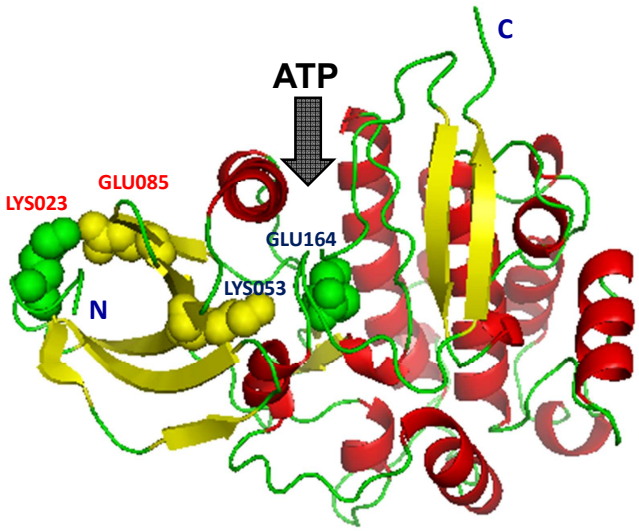

Fig. 5. Cartoon diagram of titin kinase (PDB code: 1TKI), and the amino acid residues in the salt-bridges GLU085-LYS023 and GLU164-LYS053 of the first group in the grouping. The residues are highlighted with space-filled presentation. The ATP binding site is highlighted by the arrow.

is $\delta r(t) = r(t) - \langle r(t) \rangle$, where $\langle \cdots \rangle$ denotes an average over time period T. The cross-correlation coefficient C_{ij} between the ith and jth salt-bridges is defined as

$$C_{ij} = \frac{\langle \delta r_i(t) - \delta r_j(t) \rangle}{\sigma_i \sigma_j}, \tag{4}$$

where $\sigma_i^2 = \langle \delta r_i^2(t) \rangle$. C_{ij} ranging from -1 (completely negative correlated) to $+1$ (completely positive correlated) is an entry of the cross-correlation matrix $\mathbf{C}$. By definition, the matrix $\mathbf{C}$ is symmetric. Its eigenvalues are all positive real numbers.

The salt-bridge time series consists of two components: one is noisy fluctuation originating from thermal effects, and the other is a response to atomic interactions in the molecule under a pulling force. Figure 3 shows the eigenvalue distributions for the original salt-bridge time series and the surrogated data, which is a result of the ensemble average of 10 realizations of shuffled data constructed by randomly shuffling the return time sequences. Two distributions are apparently different, implying there are indeed structure information in the salt-bridge dynamics.

We solved the 55×55 cross-correlation matrix numerically by using the algorithms of Numerical Recipes [11]. There are 55 eigenvalues and 55 eigenvectors. The signs of entries in the eigenvectors refer to the direction of the component of the salt-bridge tends to move in unfolding. For two entries having the same sign, the trends of the changes in their corresponding salt-bridges are substantially in phase. Likewise, the trends of the changes of two salt-bridges are out of phase if they have different signs. The absolute value of the entry is the magnitude of the movement. Because our aim is to group the salt-bridges with the same trends, this property is useful.

In the implementation of grouping, we simply divide the entries in a eigenvector into two groups, demarcated by zero. Following the order of eigenvalues from largest to smallest, we rearrange the eigenvectors such that the entries in the eigenvector corresponding to the largest eigenvalue are divided into two groups, one larger than zero, and the other smaller than zero; the entries in the eigenvector corresponding to the second largest eigenvalue are divided into four groups, the part corresponding to the positive part of the first eigenvector also has two divisions, one larger than zero and the other than zero; the same procedure is performed for the part corresponding to the negative part of the first eigenvector. Following such procedures, each salt-bridge is abstractly labeled with a number of indices, such that it is distinct from others. We use the first three largest eigenvalues to group salt-bridges into 2^3 groups. The results are shown in Fig. 4. Table I lists the eight groups of salt-bridges classified by the eigenvectors of first three largest eigenvalues.

In our study, the salt-bridges not formed from paired amino acids in neighbors in linear sequence are of special interest. These salt-bridges play essential roles in the maintenance of the structure of the molecule since the topological relation substantially defines compact dimension of the molecule. Dur-

TABLE I
THE EIGHT GROUPS OF SALT-BRIDGES CLASSIFIED BY THE EIGENVECTORS OF THE FIRST THREE LARGEST EIGENVALUES.

Group	Salt-bridges
1	ASP306-ARG315, GLU068-LYS053, GLU164-LYS053, ASP253-ARG257, ASP256-ARG257, GLU034-ARG032, GLU085-LYS023, GLU148-ARG146
2	GLU246-LYS245, GLU262-LYS261, ASP144-ARG323, GLU242-LYS294, GLU098-LYS161, ASP240-LYS294
3	GLU092-LYS056, GLU190-ARG263, GLU043-LYS048, GLU022-LYS023, ASP175-LYS172, GLU091-LYS058
4	GLU190-ARG266, GLU019-LYS047, GLU246-ARG291, GLU028-ARG040, ASP144-ARG146, GLU242-ARG291, GLU283-ARG119, GLU250-ARG284
5	ASP195-LYS334, GLU098-LYS048, GLU283-LYS281, GLU283-ARG284, GLU107-ARG315, GLU130-ARG078
6	ASP029-ARG032, GLU120-ARG108, GLU091-LYS056, ASP061-LYS058, GLU164-ARG323, GLU088-LYS066, GLU019-LYS018, GLU043-LYS023, GLU022-LYS018
7	GLU115-ARG156, ASP175-ARG169, GLU242-LYS245, GLU262-LYS264, GLU246-LYS288, GLU241-LYS245
8	ASP104-ARG315, GLU271-ARG257, ASP306-LYS305, GLU068-LYS067, GLU262-ARG263, ASP195-ARG263

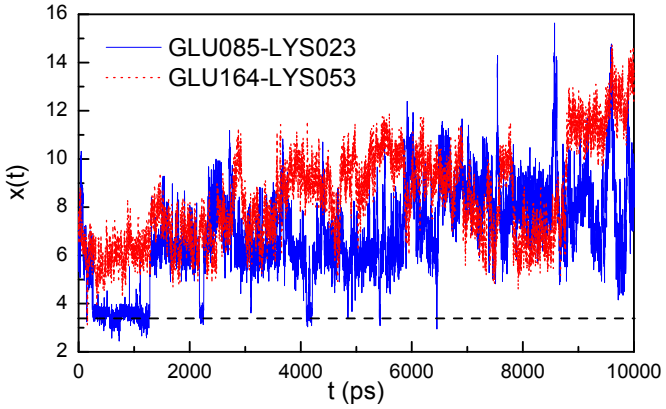

Fig. 6. Salt-bridges data of GLU085-LYS023 and GLU164-LYS053. The dashed line indicates the cut-off of 3.3 Å.

ing force-induced unfolding, these salt-bridges are expected to break once the structure breaks down, but this is not necessary true in our observations. We have observed that the salt-bridges form, break, and rejoin during unfolding. There are rich dynamic breaking and rejoining features in the process. To have a clear picture to the collective behavior of the salt-bridges classified in the same group, we show in Fig. 5 the locations of the salt-bridges GLU085-LYS023 and GLU164-LYS053 of the first group in Fig. 4. The salt-bridge data of GLU085-LYS023 and GLU164-LYS053 in Fig. 6 shows that a trend that the salt-bridges breaks quickly in the SMD simulation, where the corresponding time series exhibit similar behaviors at some epochs in the force curve shown in Fig. 1. For instance, the starting point of first four epochs of persistent increase in the force curve, corresponding to the collapse of the five-strand β sheets [1], are roughly consistent with rejoining of the salt-bridges at $t = 2200, 4100, 6600$ ps in Fig. 6. As the ATP biding site is active when titin kinase is moderately stretched, key residues can be defined with the help of experimental evidences of mutagenesis. The grouping suggests that a set of amino acid residues must be considered as a whole to take into account the collective behaviors of groups of salt-bridges. Novel proteins can be designed for biomedical applications via the analysis of simulated salt-bridge dynamics. For the limit of space, thorough discussions on this point are presented in a separate paper [12].

V. SUMMARY

In summary, we have carried out the SMD simulation to investigate the salt-bridge dynamics in the force-induced unfolding of titin kinase. By studying the SMD trajectories in the first 10000 ps simulation with pulling velocity of 0.0005 nm/ps, we defined 55 salt-bridge time series which manifest properties of dynamic breaking and rejoining of the salt bridges. We analyzed the spatial-temporal properties of these time series by scaling and cross-correlation analysis. Scaling analysis of DFA reveals that in short-time scale less than 50 ps, there is anti-persistent behavior, while the persistent behavior dominates in long-time scale larger than 100 ps. The former is contributed from thermal fluctuation, while the latter is mainly a result of the response to the interactions associated with secondary structures. We also analyzed the cross-correlation matrix constructed from the salt-bridge data, and used the properties of the eigenvectors of the first three largest eigenvalues to classify salt-bridges into 8 groups. The grouping provides information about which salt-bridges are associated with specified functions of titin kinase and mutagenesis should be considered as a whole in experiments.

ACKNOWLEDGMENT

This work was supported by the National Science Council of the Republic of China (Taiwan) under Grant Nos. NSC 96-2112-M-008-021-MY3, 97-2627-B-008-004, 98-2627-B-008-004, and 99-2627-B-008-002, and National Center for Theoretical Sciences at National Taiwan University.

REFERENCES

[1] E.M. Puchner, *et. al.*, Proc. Natl Acad. Sci. USA **105**, 13385 (2008).
[2] J. Forbes, *et al.*, J. Musc. Res. Cell Motil. **26**, 291 (2005).
[3] NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at UIUC. See M. Nelson, *et. al.*, J. Supercomputing Appl. **10**, 251 (1996)
[4] A.D. MacKerell Jr., *et. al.*, J. Phys. Chem. B **102**, 3586 (1998).
[5] H.M. Berman, *et. al.*, Nucleic Acids Research **28**, 235 (2000). The website of the Protein Data Bank is: http://www.rcsb.org/pdb/Welcome.do
[6] W.L. Jorgensen, *et. al.*, J. Chem. Phys. **79**, 926935 (1983).
[7] Humphrey, A. Dalke and K. Schulten, J. Mol. Graphics 14, 33 (1996).
[8] C.-K. Peng, *et. al.*, Chaos **5**, 82 (1995).
[9] L. Laloux, *et. al.*, Phys. Rev. Lett. **83**, 1467 (1999).
[10] V. Plerou, *et. al.*, Phys. Rev. Lett. **83**, 1471 (1999).
[11] W.H. Press, *et. al.*, *Numerical Recipes in Fortran. The Art of Scientific Computing*, Cambridge University Press, 2nd Ed. (1992).
[12] M.-C. Wu, J.G. Forbes, and K. Wang, *Salt-bridge dynamics in force-induced unfolding of titin kinase*, maunscript submitted (2011).