Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Perturbation-based Markovian Transmission Model for Macromolecular Machinery in Cell

Hsiao-Mei Lu and Jie Liang

Abstract— **The study of the dynamics of a complex system is an important problem that includes large macromolecular complexes, molecular interaction networks, and cell functional modules. Large macromolecular complexes in cellular machinery can be modeled as a connected network, as in the elastic or Gaussian network models as demonstrated by Bahar and colleagues. Here we propose the Perturbation-based Markovian Transmission Model for studying the dynamics of signal transmission in macromolecular machinery. The initial perturbation is transmitted by a Markovian processes, and the dynamics of the probability flow is analytically solved using the master equation. Due to the large size of macromolecular complexes, it is very difficult to obtain analytical time-dependent Markovian dynamics of all atoms from the first perturbation until stationary state. To overcome it, we decrease the level of complexity of the transition matrix using a Krylov subspace method. This method is equivalent to integrating all eigen modes, and we show it can provide a globally accurate solution to the dynamics problem of signal transmission for very large macromolecular complexes with reasonable computational time. We give results of the dynamics of the GroEL-GroES chaperone system by applying uniform perturbation to all residues. We are able to identify experimentally found important residues and provide a set of predicted pivot, messenger, and effector residues, each with distinct dynamic behavior. Further results of selective perturbation on the surface of ATP binding pocket identifies the path of maximal probability flow of signal. Our method can also be applied to other large systems, for example, virus capsid, ribosome, and large allosteric proteins.**

I. INTRODUCTION

The functional activities of cellular machinery often involves the dynamics of large macromolecules and their assembly, as seen in molecular interaction networks and cell functional modules. Understanding how the signal transmits dynamically in a complex molecular system from the knowledge of its molecular structures is therefore an important problem.

A challenging problem when studying large molecular assemblies is understanding the global dynamic behavior at multiple time scales. Simulation methods such as molecular dynamics study the molecules at atomic detail and cannot access the dynamic behavior of large systems at extended time scales. Instead, coarse grained models such as the Gaussian network model and elastic network model treat the molecules as connected graphs with nodes representing amino acid residues and edges representing contacts between residues [1–4]. By analyzing the dominant eigen modes of molecular motion and the key residue sites, much insight has been gained [1–4, 6].

It is difficult to decide *a priori* which modes are dynamically important, furthermore, the combination of selected subsets of eigen modes cannot guarantee the accurate description of the long term dynamics of the system. Analysis of different eigen modes often leads to different conclusions, especially when the system lacks dominant modes in the eigen spectrum. As different initial conditions are applied, different modes become important. It is possible that the slowest mode, frequently thought to be important, may contribute little to the overall dynamics of the system under certain initial conditions.

Here we study the dynamic behavior of large molecular systems by integrating the effects of all eigen modes at arbitrary initial conditions of perturbation. Our approach is to solve the master equation of the system, which is a technically challenging problem. It is very difficult to simultaneously follow the time-dependent dynamics of all atoms and residues of a very large macromolecular complex from initial perturbations until the stationary state [11, 12]. We solve this problem by decreasing the level of complexity of the transition matrix using the Krylov subspace method, which guarantees the solution to be accurate with in a specified error bound [6]. This approach is general and can be used to study very large macromolecular systems at multiple time scales, without the need of pre-clustering the residues.

In this paper, we present results of a study of the dynamics of the GroEL-GroES chaperone system. We first apply uniform perturbations to all residues in the chaperone system. This enables us to recover experimentally identified pivot, messenger, and effector residues, all with distinct dynamic behaviors. We also predict previously unknown residues playing the roles of pivot, messenger, and effector. We then perturb only the surface of ATP binding pocket of the GroEL-GroES system which enabled us to identify the path of maximal flow of signal transmission on the structure of GroEL-GroES.

This article is organized as follows. We first describe the models and the theory of the master equation. We then discuss the Perturbation-based Markovian Transmission Model. This is followed by a brief discussion of the Krylov subspace method. We then report our results on GroES-GroEL chaperone with 7 bound ADP molecules under the following initial perturbations: (1) uniform perturbation on all residues and (2) perturbation on the surface of ATP binding pocket. We conclude with a summary and remarks.

This work is supported by grants from NSF (CAREER DBI0133856), NIH (GM68958 and GM079804), and ONR (N000140310329).

Department of Bioengineering, SEO, MC-063 University of Illinois at Chicago 851 S. Morgan Street, Room 218 Chicago, IL 60607–7052, U.S.A. jliang@uic.edu

II. Model and Methods

A. Connection Model, Initial Perturbation, and Markovian Transition Model

In our model, the dynamic behavior of a large macromolecular complex is probed by applying a perturbation as initial condition. This perturbation can be regarded as a signal, which will be transmitted from the location of perturbation to the rest of the macromolecule. Here we follow others and model a large macromolecular complex as a network whose architecture solely depends on its three-dimensional structure. We use the Markovian transition model introduced by Chennubhotla and Bahar [2] to model how a given perturbation is transmitted. The dynamics of the macromolecular complex in our Perturbation-based Markovian Transmission (PMT) model is fully defined by the contacts.

In this model, each node represents a residue or an atom of the protein complex, and each edge connecting nodes represents the spatial contact in the structure. The perturbations are signals and are transmitted from one node to its neighboring nodes with a probability flow following a Markovian processes. In each step of the Markovian process, the initial perturbation is transmitted from residue i to residue j with the probability flow m_{ij} defined as follows: $m_{ij} = 0$ if there is no atom-atom contacts between residue i and j, *i.e.*, the distance between any two atoms is greater than 4.5 Å. $m_{ij} = \frac{n_{ij}}{d_j\sqrt{n_i n_j}}$ if there is at least one atom-atom contacts between residue i and j, where n_{ij} is the number of atom-atom contacts between residue i and j, n_i and n_j are number of atoms of residue i and j, respectively. $d_j = \sum_{i=1}^{N} \frac{n_{ij}}{\sqrt{n_i n_j}}$, where N is the total number of residues in the macromolecular structure. $m_{ij} = -\sum_{i \neq j} m_{ij}$ if $i = j$. The collection of m_{ij}s form the Markovian transition matrix $\mathbf{M} = \{m_{ij}\}$.

B. Master equation and Krylov Subspace Method

To analytically obtain the full dynamics of the PMT model, we discuss the master equation method and the Krylov subspace matrix reduction method for integrating the master equation.

We use a master equation to model the probability flow $\mathbf{p}_i(t)$ at time t for the i^{th} residue. Here $i = 1, ..., N$, where N is the total number of residues in the macromolecular complex. $\mathbf{p}_i(t)$ is described as the difference between the rates for transitions entering and leaving the residue i in a continuous time Markov process. We have:

$$\frac{dp_i}{dt} = \sum_{j=1}^{N} [m_{ji}p_j - m_{ij}p_i], \tag{1}$$

where m_{ij} and m_{ji} are rate constants for the respective transitions. The equivalent master equation in matrix form is:

$$\frac{d\mathbf{p}}{dt} = \mathbf{M} \cdot \mathbf{p}, \tag{2}$$

where $\mathbf{p}$ is the flow probability column vector $(\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N)^T$, $\mathbf{M}$ is the rate matrix, *i.e.*, Markovian matrix, defined as before.

A given initial perturbation $\mathbf{p}_0$ at $t = t_0$ whose components sum to 1 may be applied to any subset of nodes. The analytical solution of the master equation is

$$\mathbf{p}(t) = e^{\mathbf{M}(t-t_0)}\mathbf{p}_0, \tag{3}$$

and the stationary distribution is

$$\mathbf{p}_i(t = \infty) = \frac{d_i}{\sum j = 1^N d_j}, \tag{4}$$

which is independent to any initial condition. For a given Markovian matrix, the master equation has an exact solution, which provides an exact picture of the relaxation process of the initial perturbation in the complex macromolecular system. For any arbitrary perturbation on the system, the final stationary state of the system only depends on the connectivity of the network.

C. Krylov subspace method

Simultaneously monitoring the exact time evolution of probability flow for all individual residues from the first perturbation until reaching stationary state is a challenging task. Mathematically, the model of master equation is equivalent to a Markov chain, where the probability flow vector of residues at the k^{th} time step can be calculated by a straightforward matrix multiplication of $\mathcal{O}(\log kN^3)$ steps [6]. However, this becomes impossible for a large matrix. The analytical solution of $\mathbf{p}(t) = \sum_i C_i \mathbf{n}_i e^{-\lambda_i t}$ through diagonalization is also impractical, as it is only possible to calculate a few eigenvectors and eigenvalues for a large matrix [6, 11].

We seek an accurate solution without the approximation of macrostates, namely, merged states through clustering. Taking advantage of the sparsity of the Markovian matrix $\mathbf{M}$, we follow the approach of Sidje [8] and use the analytical solution of matrix exponential in Eq. (3) to expand $e^{\mathbf{M}t}\mathbf{p}_0$ in the Krylov subspace $\mathcal{K}_m$ defined as:

$$\mathcal{K}_m(\mathbf{M}t, \mathbf{p}_0) \equiv \text{Span}\{\mathbf{p}_0, \cdots, (\mathbf{R}t)^{m-1}\mathbf{p}_0\}. \tag{5}$$

Denoting $|| \cdot ||_2$ as the 2-norm of a vector or matrix, our approximation then becomes $\mathbf{p}(t) \approx ||\mathbf{p}_0||_2 \mathbf{V}_{m+1} e^{t\overline{\mathbf{H}}_{m+1}} \mathbf{e}_1$, where $\mathbf{e}_1$ is the first unit basis vector, $\mathbf{V}_{m+1}$ is a $(m + 1) \times (m + 1)$ matrix formed by the orthonormal basis of the Krylov subspace, and $\overline{\mathbf{H}}_{m+1}$ the upper Heisenberg matrix, both computed from an Arnoldi algorithm. The error can be bounded by $\mathcal{O}(e^{m-t||\mathbf{R}||_2}(t||\mathbf{R}||_2/m)^m)$. We now only need to compute explicitly $e^{\overline{\mathbf{H}}_{m+1}t}$. Because m is much smaller than the total number of states, this is a simpler problem. A special form of the Padé rational of polynomials instead of Taylor expansion is used [8]: $e^{t\overline{\mathbf{H}}_{m+1}} \approx N_{pp}(t\overline{\mathbf{H}}_{m+1})/N_{pp}(-t\overline{\mathbf{H}}_{m+1})$, where $N_{pp}(t\overline{\mathbf{H}}_{m+1}) = \sum_{k=0}^{p} c_k (t\overline{\mathbf{H}}_{m+1})^k$ and $c_k = c_{k-1} \cdot \frac{p+1-k}{(2p+1-k)k}$. In our calculation, we select $m = 60$.

III. Results

An application of the PMT model on the study of the dynamics of the x-ray crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperone complex in *Escherichia coli* (PDB id: 1AON) [10] is shown in this paper. There are 8,015 residues and 58,699 atoms in the GroEL-GroES protein complex with 14 and 7 identical chains in GroEL and GroES, respectively. This ATP regulated chaperone complex assists the folding of unfolded and misfolded proteins. The functional activity involved in the allosteric effect starts with 7 ATPs and a ligand (unfolded or misfolded protein) binding to the trans-ring chains, this is followed by conformational change of the trans-ring chains to the cis-ring chains. The effect continues with GroES binding to the top of the trans-ring chains and the releases of the ligand into the central cavity. In the meantime, the ATPs are hydrolyzed to ADPs. The whole process ends after another 7 ATPs and a ligand (the unfolded protein) bind to the trans-ring chains. It causes the cis-ring chains to take the same conformation as become trans-ring chains by releasing the GroES, ligand (the partial folded or folded protein), and 7 ADPs, and the cycle is resumed.

A. Uniform perturbation on all residues

We virtually perturb all residues in the GroEL-GroES-$(ADP)_7$ chaperone protein complex. We then assess the roles of different residues in transmitting signals.

a) Observation on the time trajectories of experimentally verified residues: Pro33 and Thr90 are on the surface of nucleotide binding pockets of the cis-ring chains, and Asp398 plays an important role in ATP hydrolysis by a large conformational change of its side chain atoms [5, 9, 10]. We observe these functional important resides experience significant periodic fluctuations. They have either large or small probabilities in the stationary state (in Fig. 1). Few other residues that are not on the nucleotide binding pockets experience these kind of periodic fluctuations.

Chennubhotla and Bahar [2] suggested that messenger residues play important roles by passing signals between domains or chains. For example, Glu461 and Arg197 are inter-chain and intra-chain messenger residues, respectively. As shown in Fig. 1, although these two residues do not make many contacts with other residues, the periodic fluctuations implies that they are involved in the constant transmission of signals. Unlike residues on nucleotide binding pockets, the messenger residues do not reach extreme probabilities in stationary state. That is once these residues receive perturbation signal, they simply pass it on and do not hold onto it.

Each chain of the GroEL structure consists of three domains: the equatorial, intermediate, and apical domain. These names reflect the spatial position in the GroEL-GroES chaperone complex structure. The three domains are connected by four pivot residues: Gly192, Gly375, Gly410 and Pro137. The first two connect the apical and intermediate domains, and the others connect the intermediate and equatorial domain. As shown in Fig. 2, Pro137 and Gly192 are found to have smooth time trajectories and extreme low

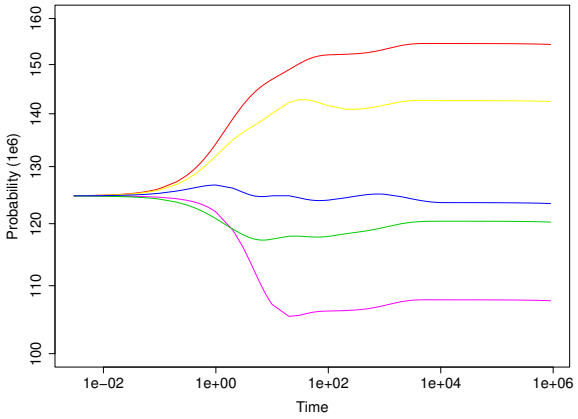

Fig. 1: Time trajectories of residues on ADP binding pocket: Pro33 (magenta), Asp398 (yellow), and Thr90 (red) and messenger residues: Glu461 (green) and Arg197 (blue).

probabilities in the stationary state. Although, only these two residues have the distinct dynamic behavior, the other two pivot residues are in immediate contact with these two residues.

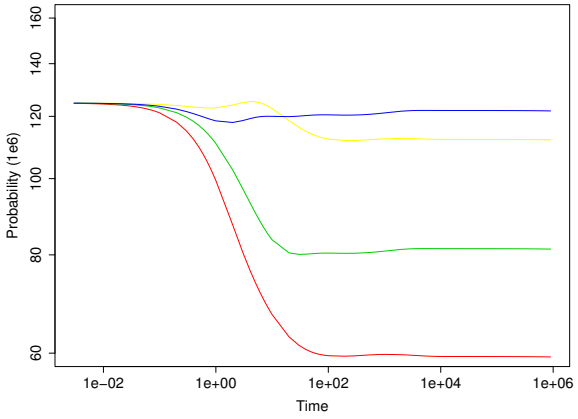

Fig. 2: Time trajectories of four pivot residues: Gly192 (red), Gly375(yellow), Pro137 (green), and Gly410 (blue).

b) Prediction on pivot, messenger, and effector residues: According to what is learned from the dynamic behavior of active site, messenger, and pivot residues, we are able to make predictions and identify additional functionally important residues.

We identify pivot residues by applying the rule of maximal change and lowest minimal probability, without any temporal increase in the time trajectory. By selecting the top 5% of residues in the cis-ring chains that obey the above rule, we identified thirteen pivot residues. Additional to the two structure pivot residues (Pro137 and Gly192) that were experimentally identified [10], seven pivot residues are found

on the equatorial domain, and the other four were found on the intermediate and apical domains. All thirteen residues are located on the surface of the structure and at the turn of helices. Because of the structural similarity of the cis-ring and trans-ring chains, residues on the equatorial domain have the same dynamic behavior in both of the cis-ring and trans-ring chains of GroEL. On the other hand, residues on the intermediate and apical domain in cis-ring chains are no longer buried in trans-ring chains structure. Indeed, the dynamics behavior is no longer as pivot-like (in Fig. 3).

Similarly, we identified pivot residues in GroES. We found residues Gly23, ILE25, Gly29, and Ala32, which make contacts with the cis-ring chains of GroEL and are located on the experimentally verified mobile loop of GroES (residues 16-33) [7, 10]. Additional pivots identified (Met1, SER75) are at the inter-chain area in GroES, suggesting that these residues may interact with neighboring chains.

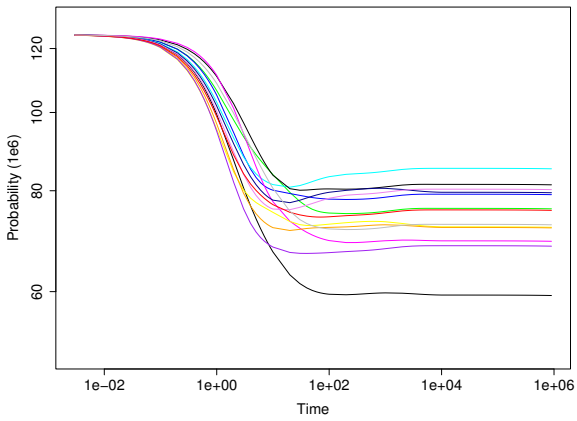

Fig. 3: Time trajectories of predicted pivot residues.

We also have identified effector residues, which greatly respond to the perturbations by experiencing maximal changes and attain highest maximal probability. By selecting the top 5% of residues in the cis-ring chains which obey the above rule, we identified thirteen effectors residues (in Fig. 4). Four of them are on the equatorial domain, two are on the intermediate domain, and the others are on the apical domain. Residues on the equatorial domain are near the nucleotide binding site and have the same dynamic behavior in both cis-ring and trans-ring. Effectors on the intermediate domain are at the bend and experience largest conformational change. Indeed, their dynamics also experience the largest change in the cis-ring and trans-ring chains. Effector residues identified on the apical domain in cis-ring chains are either facing the center cavity or on the helix interacting with GroES. It suggests that these effect residues may bind to unfolded proteins and/or GroES.

According to what we have learned about known messenger residues, we have identified additional messenger residues that transmit signals with periodic fluctuations during time evolution. We select the residues whose time

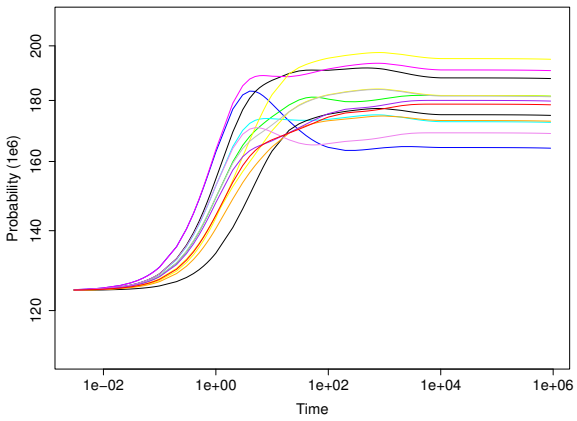

Fig. 4: Time trajectories of predicted effector residues.

trajectories have minimal changes but contain many local minimum and maximum. We further classify these residues into four clusters based on their dynamics (in Fig. 5).

(a) Messenger residues on the interface between GroES and GroEL.

Leu27 on the mobile chain of GroES and three residues (Ala241, Lys242, and Val271) on the helix of the cis-ring chains in GroEL communicate with GroES. They are selected as messenger residues. They make direct contacts to each other. According to the response time to the perturbation, we can see the signal is transmitted from the cis-ring chains of GroEL gradually to GroES. The much higher peak of Leu27 indicates it collects and holds the signal from GroEL before it transmits signal to GroES.

(b) Messenger residues on apical and intermediate domain.

They are messenger residues directly passing signal from equatorial domain to apical domain.

(c) Messenger residues on equatorial domain and near intermediate domain.

These are residues passing signal from the surface of nucleotide binding pocket toward intermediate domain (in rainbow color according to their response time).

(d) Messenger residues on equatorial domain and near the trans-ring chains of GroEL.

These are residues passing signal from the surface of nucleotide binding pocket toward the interface of cis-ring and trans-ring chains of GroEL. (in rainbow color according to their response time).

To sum up, our results suggest that the allosteric signal transmission is from the nucleotide binding pocket in the cis-ring chains of GroEL towards and propagates the GroES and the trans-ring chains of GroEL.

B. Pocket perturbation

By applying uniform perturbations only on the residues of nucleotide binding pocket, we can investigate the path of maximal flow of the signal by monitoring the residues with highest probability of fluctuation in each time point from the beginning of the perturbation to the stationary state. In

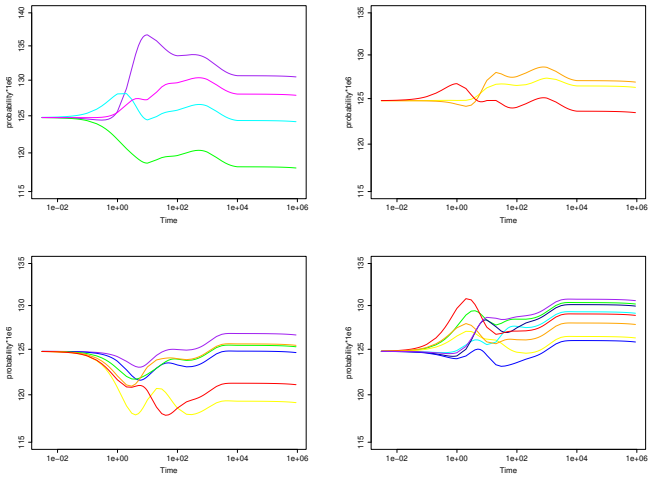

Fig. 5: Time trajectories of messenger residues. They are classified by the dynamics. Four classes from (a) to (d) described in the text are the figure in the order of upper left, upper right, lower left, and lower right.

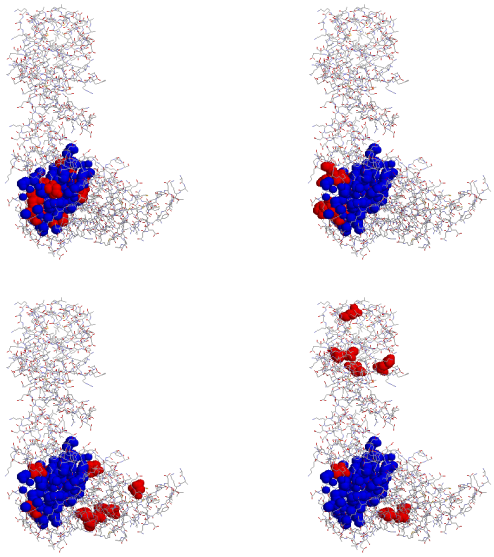

Fig. 6: The path of maximal flow transmission starts from the nucleotide binding pocket of the cis-ring of GroEL to both the trans-ring of GroEL and GroES. Residues on the nucleotide binding pocket are in blue. The top ten residues with the maximal probability of fluctuation at time $t = 0.1$ (upper left), 100 (upper right), $1,000$ (lower left), and $10,000$ (lower right) are in red.

Fig. 6, the top ten residues with the maximal probability of the fluctuation are selected at 4 time points. At time $t = 0.1$, immediately after the perturbation on the 58 residues on the nucleotide binding pocket, these residues respond to the perturbation greatly, as all are on the center of the perturbed surface. After time period $t = 100$, the signal starts to spread out and other residues which are not on the nucleotide binding pocket are influenced. At time $t = 1,000$, the majority of selected residues are not on the nucleotide binding pocket and some are at the interface between cis-ring and trans-ring chains of GroEL. At time $t = 10,000$, the signals are further transmitted toward the direction of the GroES. The highly fluctuated residues Asn265 is at the helix communicating with GroES.

IV. CONCLUSION

In this paper, we studied the dynamics of GroEL-GroES-$(ADP)_7$ chaperone protein complex with PMT model. The dynamics of the large macromolecular complex can be obtained by the PMT model where perturbations are transmitted by a Markovian processes on a connected network defined by the contacts between residues. Analytically solving the master equation with the Krylov subspace algorithm provides a global solution to the dynamics problem of very large macromolecular complexes from perturbation to stationary state within reasonable computational time.

We have perturbed all residues *in silico* in the chaperone structure. We have analyzed the response patterns of experimentally known pivot residues, messenger residues, and residues on the nucleotide binding pockets that are functionally important. Based on the distinct dynamic behavior of these residues, we are able to predict additional residues that may function as pivot, effector, and messenger residues. We also applied perturbations on the surface of nucleotide binding pocket and have identified the path of maximal flow of signal transmission. By monitoring the residues with maximal fluctuations at each time point, a clear pathway of signal transmission is identified from the center of the nucleotide binding pocket of the cis-ring chains of GroEL toward the direction to the GroES and the interface between cis-ring and trans-ring chains of GroEL.

The successful application of the PMT model on the chaperone protein complex, consisting of 8,015 residues, shows that the dynamics of signal transmission in large macromolecular machinery is accessible to computational studies. Although there is no current experimental confirmation, our predictions of pivot, effector, and messenger residues suggest a plausible new hypothesis that can be further tested experimentally. Our model can be easily applied to other systems, for example, virus capsid, ribosome, and allosteric proteins.

REFERENCES

[1] I. Bahar and A. Rader, "Coarse-grained normal mode analysis in structural biology." *Curr Opin Struct Biol*, vol. 15(5), pp. 586–92, 2005.

[2] C. Chennubhotla and I. Bahar, "Markov propagation of allosteric effects in biomolecular systems: application to groel-groes." *Mol Syst Biol*, vol. 2, p. 36, 2006.

[3] C. Chennubhotla, A. Rader, L. Yang, and I. Bahar, "Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies." *Phys Biol*, vol. 2(4), pp. S173–80, 2005.

[4] P. Doruker, R. Jernigan, and I. Bahar, "Dynamics of large proteins through hierarchical levels of coarse-grained structures." *J Comput Chem*, vol. 23(1), pp. 119–27, 2002.

[5] A. Horwich, W. Fenton, and T. Rapoport, "Protein folding taking shape. workshop on molecular chaperones." *EMBO Rep*, vol. 2(12), pp. 1068–73, 2001.

[6] S. Kachalo, H. Lu, and J. Liang, "Protein folding dynamics via quantification of kinematic energy landscape." *Phys Rev Lett*, vol. 96(5), p. 058106, 2006.
[7] S. Landry, A. Taher, C. Georgopoulos, and d. van, "Interplay of structure and disorder in cochaperonin mobile loops." *Proc Natl Acad Sci U S A*, vol. 93(21), pp. 11 622–7, 1996.
[8] R. B. Sidje and W. J. Stewart, "A numerical study of large sparse matrix exponentials arising in markov chains." *ACM Trans. Math. Softw.*, vol. 24(1), p. 130, 1998.
[9] P. Sigler, Z. Xu, H. Rye, S. Burston, W. Fenton, and A. Horwich, "Structure and function in groel-mediated protein folding." *Annu Rev Biochem*, vol. 67, pp. 581–608, 1998.
[10] Z. Xu, A. Horwich, and P. Sigler, "The crystal structure of the asymmetric groel-groes-(adp)7 chaperonin complex." *Nature*, vol. 388(6644), pp. 741–50, 1997.
[11] W. Zhang and S. Chen, "Exploring the complex folding kinetics of rna hairpins: I. general folding kinetics analysis." *Biophys J*, vol. 90(3), pp. 765–77, 2006.
[12] ——, "Exploring the complex folding kinetics of rna hairpins: Ii. effect of sequence, length, and misfolded states." *Biophys J*, vol. 90(3), pp. 778–87, 2006.