GroEL-GroES State Transition Analysis using Anisotropic Network Model

Haytham Zaami
Pennsylvania State University
Middletown, United States
haythamzaami12@gmail.com

***Abstract*—Predicting protein movements and fluctuations has been an ongoing study for many years. The GroEL/GroES chaperonin protein system found in *Escherichia coli* goes through several different transformations, also known as states; the tense state (T), relaxed state (R), and states with GroES attached (R', R'', etc.). Understanding whether these transitions are within the natural motions of individual conformational states remains a question in the study of protein analysis. In this study, we performed Anisotropic Network Model (ANM) analysis on the 14 chains of the GroEL protein (PDB: 4AAS, 1GR5). Using a coarse-grained approach by analyzing alpha carbon atoms of each chain in the relaxed state protein, we selected from the best 10 modes that most closely resembled the tense state for all chains. Our analysis reveals that the best-predicting modes achieve overlaps of 18–21%. Per-chain RMSD analysis reveals 4.3-fold asymmetry (7.45 Angstroms vs 1.72 Angstroms) between rings, providing quantitative evidence for negative cooperativity in GroEL's allosteric mechanism.**

I. Introduction

Proteins of all types tend to undergo several transformations. Understanding why and how certain proteins undergo these different changes is important, as the applications of these studies can impact health practices, biological impacts, and more. More specifically, using a theoretical study and comparing it to a transformation that happens in reality can give researchers a scope on just how practical an approach is. The GroEL-GroES protein system, often found in E. Coli, goes through different state changes for it's purpose is to correct misfolded proteins. Using the biological process ATP-Hydrolysis, proteins inside the double ringed body of GroEL to prevent aggregation, unfold, and refold a protein correctly before releasing the protein out of it's chamber. The Anisotropic Network Model allows the study of a protein's collective movements, predicting how they potentially vibrate, flex, and change shape for functions. By turning a state of the GroEL protein into an Anisotropic Network Model and seeing the different mode's fluctuations, we can see just how well the ANM can predict the transition. Taking individual ANMs for each chain of the GroEL protein and comparing the results to the actual transition, we can visualize and see a tangible difference between a predicted mode and actual state.

The GroEL-GroES system goes through three primary states: the R (relaxed) state, R' (intermediate) state, and T (tense) state. These transitions are important for the chaperonin's function, with the T to R transition representing the ATP-induced conformational change that opens the folding chamber and releases the substrate protein. This study particularly focuses on the T to R state transition which doesn't require ATP-Hydrolysis to be accounted for. The ANM requires a low energy minimum, and ATP-Hydrolysis violates that requirement, as the folding and unfolding of a protein inside the GroEL-GroES system requires a flux in energy.

The topic of GroES/GroEL transitions has been studied before, with a particular study being an adjusted approach to ANM called aANM, which explores the functional transitions. The aANM approach is a computationally efficient and physically plausible tool for analyzing potential transition pathways for a given state by using large complexes/assemblies. Having multiple structures as inputs and an adaptive spring constant for the connectivity between atoms, the aANM results in an accurate and intuitive approach.

By taking a different approach that focuses on verifiability of the ANM result rather than the strength of the prediction, less complex computation, the goal is to understand how close the ANM modes can replicate a state transition. Strictly using natural motions, finding an accurate overlap is not very intuitive, given the transition of GroEL's closed chamber form to an end having an opening. However, the goal is not to predict a structure, but compare intrinsic motion to natural motion. Using ANM with a coarse-grained C-alpha representation, we perform normal mode analysis on each of the 14 individual chains of the GroEL R-state (PDB: 4AAS) and calculate mode overlaps with the T-state (PDB: 1GR5). The coarse-grained approach uses a uniform spring constant of 1 and a cutoff distance of 7 Angstroms, which is commonly used in research. Of the modes calculated, we take the lowest 10 modes and select the one that conforms closest to the tense state chain's structure. This gives insight on individual chains having asymmetric transition dynamics. Our analysis reveals that the best-predicting modes achieve overlaps of 18-21%. Furthermore, we observe differences in Root Mean Square Deviation (RMSD) values between chain groups, suggesting asymmetric changes across the two GroEL rings. The RMSD calculates the average distance between atoms of our pseudoatom and the T state.

II. Literature Review

Proteins are not rigid structures, but are dynamic systems that take different formations based on several factors. In

order to complete a study that revolves around comparing transformations, several mathematical tools are used to allow for computation. Atilgan, A.R. et al. proposed the idea of Anisotropic Network Models, with the goal of estimating collective motions which make up the base of my study. The initial idea of using Gaussian Network Models (GNM) has worked in protein analysis, but having three dimensional vectors helps give direction for protein fluctuations. My initial approach used GNM, but did not give directional motions. After having understood the mathematical formula for ANM, relevant research has been introduced, with that being the aANM approach by Eyal et al., which does not retain the spring constant parameters but bases it on local structural environment [1]. While aANM improves accuracy and is efficient, the approach of strictly using ANM to see how well natural motions can replicate a true state transformation in GroEL is not as complex. To go more in depth, I reviewed three different publications revolving around the idea of GroEL-GroES states, ANM, and aANM.

A. Anisotropic Network Model

Anisotropic Network Models come from the umbrella the Elastic Network Models. By selecting the Carbon atom (Cα) of each residue in a given protein, we map the network of Cαs to each other, with a cutoff distance in mind to reduce computation time. By creating this less complex network of atoms interconnected to each other, we assume each connection between the Cαs is a spring at it's equilibrium. A specific formula that utilizes matrices will be introduced in the methodology section, but for now, the output of ANM is several eigenvalues and eigenvectors, where each eigenvalue is related to the frequency and stiffness of a mode and each eigenvector is a $3N$-dimensional vector where N is the number of atoms used [1]. Ultimately, we get the network of atoms' fluctuations and possible coordinates, which are used in the current study.

B. GroEL-GroES states

As stated previously, the GroEL-GroES chaperon protein assists in taking in improperly folded proteins and correctly folding them with ATP-Hydrolysis. The different states can be categorized into two different formal states with substates; 'T' for the tense/taut state, and 'R' for the relaxed [2]. The T state, which has no GroES cap, waits for the ATP to be put into the chamber before going into the R state. From here, the R state goes through different primes, where R' has the GroES co-chaperon attached along with ATP. After going through ATP-Hydrolysis, we get ADP in the GroEL chamber along with the cap on, before returning to the T state. Certain details like the salt bridge breaks are relevant to the state transitions, but in the study are not accounted for.

C. aANM

The concept of the adaptive anisotropic network model was introduced by Eyal et al [3]. to address a limitation in standard ANM; the assumption of uniform spring constants between all residue pairs within the cutoff distance. This parameter adaptation is achieved by analyzing a training set of protein structures and deriving residue-pair-specific spring constants that better reproduce experimental B-factors and low-frequency vibrational modes. Thus, the aANm is an optimal approach for understanding how these big groups of molecules move and switch between different states.

III. Methods

The relaxed (R) state structure was obtained from the Protein Data Bank (PDB) ID 4AAS and the tense (T) state from PDB ID 1GR5. GroEL consists of 14 chains in the form of two rings: Ring 1 (chains A-G, the cis ring) and Ring 2 (chains H-N, the trans ring). Chain matching was performed using the ProDy library's sequence alignment function, which identifies structurally equivalent residues present in both conformations.

From there, rather than taking the ANM of the entire protein and comparing it to our target T state, we perform ANM analysis on each chain. The methodology for doing so was done in the same manner as Atilgan, A.R. et al. Using the idea of the uniform spring constant, the formula used is:

$$U_{ANM} = \frac{1}{2} \sum \gamma (r_{ij} - rij^0)^2 \tag{1}$$

where $\gamma = 1$ is the uniform spring constant, $\mathbf{r}_{ij}$ is the distance vector between Cα atoms i and j, $\mathbf{r}_{ij}^0$ is the equilibrium distance from the crystal structure, and Γ_{ij} is the contact matrix.

A. Hessian Matrix and Contact Matrix

The cutoff distance $r_c = 7$ Angstroms was chosen based on standard practice in elastic network models of protein complexes. The Hessian matrix $\mathbf{H}$, representing the second derivative of potential energy, was constructed as a $3N \times 3N$ matrix (where N is the number of Cα atoms) with elements:

$$H_{ij} = \begin{cases} -\gamma \Gamma_{ij} & \text{for } i \neq j \\ -\sum_{k \neq i} H_{ik} & \text{for } i = j \end{cases} \tag{2}$$

where Γ_{ij} is the contact matrix:

$$\Gamma_{ij} = \begin{cases} 1 & \text{if } |\mathbf{r}_{ij}^0| < r_c \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Normal modes were obtained by solving the eigenvalue problem:

$$\mathbf{H}\mathbf{u}_k = \lambda_k \mathbf{u}_k \tag{4}$$

where $\mathbf{u}_k$ is the k-th eigenvector (mode shape) and λ_k is the corresponding eigenvalue. The six lowest eigenvalues correspond to trivial rigid-body translations and rotations and were discarded, as their values are near zero. Modes 7-16 (the first 10 non-trivial modes) were retained for analysis.

B. Mode Overlap Calculation

To quantify how well each mode predicts the T to R conformational transition, overlap between each mode and the actual structural change was calculated. First, the T state structure for each chain was aligned to the corresponding R state chain to remove rigid-body rotations and translations. The conformational change vector $\Delta\mathbf{r}$ was computed as:

$$\Delta\mathbf{r} = \mathbf{r}_T - \mathbf{r}_R \tag{5}$$

where $\mathbf{r}_T$ and $\mathbf{r}_R$ are the coordinates of the T and R states after alignment. Both $\Delta\mathbf{r}$ and each mode vector $\mathbf{u}_k$ were normalized to unit length. The overlap coefficient O_k for mode k was calculated as:

$$O_k = \left| \frac{\mathbf{u}_k \cdot \Delta\mathbf{r}}{|\mathbf{u}_k||\Delta\mathbf{r}|} \right| \tag{6}$$

Here, an overlap value of 1 indicates perfect alignment between the mode and conformational change. For each chain, the mode with the highest overlap coefficient was designated as the "best mode" for predicting the T to R transition.

C. Conformational Change Quantification

To quantify the conformational change for each chain during the T to R transition, Each chains' RMSD values were computed using ProDy's calcRMSD function. For each chain, the T state structure was aligned to the corresponding R state structure using least-squares superposition. The RMSD was calculated as:

$$\text{RMSD}_{RtoT}^{(c)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\mathbf{r}_R^{(c,i)} - \mathbf{r}_T^{(c,i)}|^2} \tag{7}$$

where $N = 515$ is the number of Cα atoms per chain, and superscript (c) is for chain identity. These RMSD values represent the actual displacement magnitude observed in experimental structures and provide scaling factors for pseudoprotein construction.

D. Pseudoprotein Construction

To validate whether ANM modes can predict the actual T state structure, a pseudoprotein was constructed by selecting each chain's best mode. The process involves three steps for each chain. First, the best mode (eigenvector with highest overlap to the T state) is identified. Next, the mode vector is scaled to match the actual conformational change. Specifically, for chain c with best mode index k_c, the scaling factor was:

$$\alpha_c = \frac{\text{RMSD}_{RtoT}^{(c)}}{\sqrt{\langle|\mathbf{u}_{k_c}|^2\rangle}} \tag{8}$$

This ensures the ANM displacement we predicted has the same magnitude as the observed conformational change, while direction comes purely from the normal mode. The coordinates were computed as:

$$\mathbf{r}_{pseudo}^{(c)} = \mathbf{r}_R^{(c)} + \alpha_c \cdot \mathbf{u}_{k_c} \tag{9}$$

This process was repeated for all 14 chains, using each chain's best mode and scaling factor. This resulted in 14 chains of coordinates which were assembled into a single pseudoprotein structure, keeping original residue numbers, and atom names from the R state to make matching and aligning simpler. This pseudoprotein represents the ANM's prediction of the T state structure based solely on intrinsic vibrational modes of the R state.

E. Validation Metrics

The pseudoprotein structure was compared to the actual T state structure (PDB: 1GR5) to assess prediction accuracy. Global alignment was performed by matching all 14 chains simultaneously based on chain identifiers. A total of 7210 Cα atoms were matched across all chains. The global validation RMSD was calculated as:

$$\text{RMSD}_{validation} = \sqrt{\frac{1}{N_{total}}\sum_{i=1}^{N_{total}} |\mathbf{r}_{pseudo}^{(i)} - \mathbf{r}_T^{(i)}|^2} \tag{10}$$

where $N_{total} = 7210$ is the total number of matched Cα atoms across all chains. This metric quantifies the overall accuracy of the ANM-based structural prediction.

F. Distinction Between RMSD Metrics

Two distinct RMSD calculations serve different purposes in this study. The per-chain conformational RMSD differs from the validation RMSD as it quantifies the magnitude of structural change each chain undergoes during the experimentally observed T to R transition. These values represent ground truth measurements ranging from 1.72 Angstroms (Ring 2) to 7.45 Angstroms (Ring 1) and are used as scaling factors for ANM mode displacements.

The global validation RMSD, more recently referenced, measures prediction accuracy by comparing the ANM-generated pseudoprotein to the actual T state structure.

IV. Results

From Fig. 1, we have the best calculated mode in bold for each chain. From our analysis, we see the highest related mode in the first ring is mode 2 for chain G. Ring 2 had chain N with the 10th mode being the largest overlap. From these details, we can see the first ring (A-G) experiences a majority conformation in the second mode, meaning the chain experiences a larger collective motion rather than individual. This makes sense as the first ring in the GroEL does not open up, where as the second chain does.

Fig. 3 displays mode overlap coefficients for all 14 chains across modes 1-10. Ring 1 shows a pattern coordination: six of seven chains (A, C, D, E, F, G) achieve highest overlap within Mode 2, ranging from 0.187 to 0.199. Chain B is the exception, achieving its highest overlap of 0.194 with Mode 4. The mean overlap for Ring 1 is 0.188 $\pm$ 0.004. Chain G exhibits the highest Ring 1 overlap at 0.199. This further strengthens our idea on the first ring moving collectively as it's not opening to take in any unfolded proteins.

Ring 2 displays dispersed behavior, with chains achieving maximum overlap with different modes: H and I with Mode 9, J with Mode 6, K, L, and M with Mode 4, and N with Mode 10. The mean overlap is 0.159, significantly lower than Ring 1. Chain N shows the highest Ring 2 overlap at 0.215, despite using Mode 10 rather than a low-frequency collective mode.

The use of Mode 2 in Ring 1 indicates this low-frequency collective mode captures large-scale domain rearrangements characteristic of the T to R transition in the cis ring. The dispersed mode selection in Ring 2, combined with lower overlaps, reflects minimal conformational changes in the trans ring, which is consistent with RMSD analysis. These findings align with previous studies showing that functional transitions in GroEL involve asymmetric allosteric mechanisms [3].

Modes	1	2	3	4	5	6	7	8	9	10
A	0.0368	**0.1876**	0.15	0.1394	0.0477	0.1262	0.075	0.0462	0.003	0.0911
B	0.0158	0.1485	**0.0861**	0.1943	0.0463	0.1258	0.0843	0.0389	0.0912	0.077
C	0.0369	**0.1877**	0.1498	0.1397	0.0479	0.1261	0.075	0.0462	0.003	0.0908
D	0.0394	**0.1886**	0.139	0.151	0.0566	0.1242	0.0732	0.0462	0.0004	0.0912
E	0.0368	**0.1889**	0.1467	0.1441	0.05	0.1254	0.0737	0.0464	0.0019	0.0916
F	0.0486	**0.1873**	0.1343	0.1522	0.0472	0.1272	0.076	0.0459	0.0012	0.0918
G	0.042	**0.1985**	0.1462	0.152	0.0292	0.115	0.0656	0.0446	0.016	0.0855
H	0.1269	0.0337	0.0693	0.0888	0.0835	0.0828	0.1168	0.0041	**0.131**	0.0662
I	0.0876	0.0299	0.1178	0.0719	0.0718	0.1142	0.0961	0.0006	**0.1808**	0.0443
J	0.0917	0.0395	0.1137	0.0638	0.1011	**0.1392**	0.0814	0.0166	0.1324	0.0761
K	0.0554	0.0545	0.0171	**0.1472**	0.1004	0.0957	0.0812	0.0387	0.126	0.0734
L	0.1017	0.0142	0.0708	**0.1885**	0.1107	0.0531	0.1312	0.0195	0.1164	0.0858
M	0.0951	0.0935	0.0212	**0.1722**	0.1024	0.1033	0.0496	0.0579	0.1277	0.1299
N	0.1188	0.0515	0.1068	0.0228	0.0829	0.1137	0.0708	0.0311	0.0148	**0.2153**

Fig. 1. Mode overlap values for all 14 chains across modes 1-10 from ANM analysis. Detailed heatmap visualization shown in Fig. 3

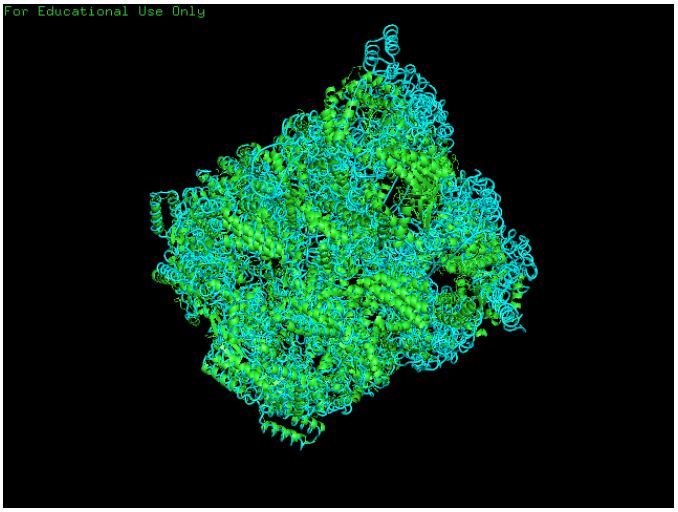

Fig. 2. Overlapping of pseudoprotein and 1GR5. Global validation RMSD is 51.9 Angstroms across 7210 Cα atoms.

V. Conclusions

Overall, this study's goal was to apply a theoretical study to real transformation that happens in the biological space. The GroEL-GroES protein system is vital to life, as without it there wouldn't be correction in misfolded proteins. Rather than attempting to predict the final structure with ANM, my approach is an attempt to give a different viewpoint for analyzing protein

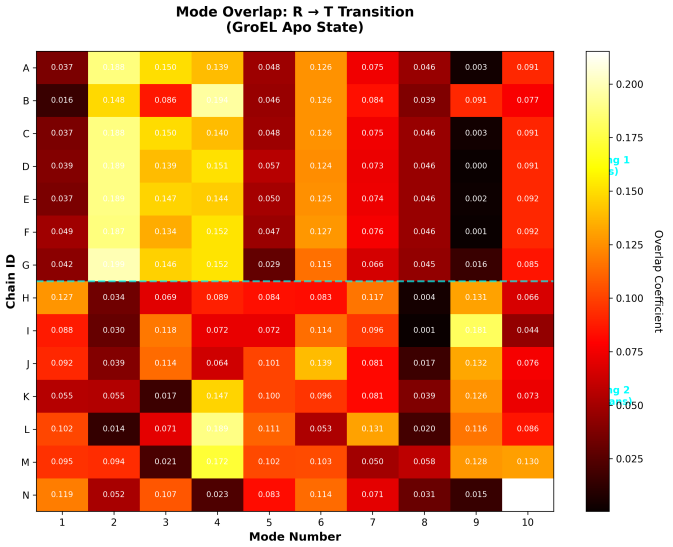

Fig. 3. Heatmap of mode overlap coefficients for all 14 chains across modes 1-10. Ring 1 (chains A-G, top) shows coordinated motion with six of seven chains achieving maximum overlap in Mode 2 (0.187-0.199), indicated by bright yellow. Ring 2 (chains H-N, bottom) exhibits dispersed mode selection with lower overlap values (0.131-0.215). Color intensity represents overlap magnitude, with warmer colors indicating stronger directional alignment between ANM modes and the actual T to R conformational change.

fluctuations on individual chains rather than the collective Cα atoms. Along with that, comparing theoretically derived low-energy collective motions with experimentally resolved conformational states gives us a tangible difference between a calculation found through theory and a real transition.

Our results show that, despite the simplicity of the ANM and its assumption of single energy minimum, certain low-frequency modes can display an overlap with the T state structures. The highest modes' overlaps fell into the range of 18–21%, indicating that a portion of the T to R transition is minimally encoded in the dynamics of the R state. While these overlaps are not enough to explicitly say the transition is a natural harmonic motion, they are consistent by having a collective change in the base ring of the GroEL protein (Chains A-G) where the shape stays relatively the same in both T and R states.

The observed variation in RMSD values and mode overlaps across individual chains highlights asymmetry in GroEL's transition dynamics, particularly between the two rings. This supports the idea that GroEL's functional motions are not uniformly distributed across each chain and that chains may contribute differently to the overall conformational change.

Acknowledgment

This document was prepared for Dr. Hyuntae Na's Computational Structural Biology Course at Pennsylvania State University.

References

[1] E. Eyal, L.-W. Yang, and I. Bahar, "Anisotropic network model: systematic evaluation and a new web interface," *Bioinformatics*, vol. 22, no. 21, pp. 2619–2627, 2006.

[2] X. Fei, D. Yang, N. LaRonde-LeBlanc, and G. H. Lorimer, "Crystal structure of a GroEL–ADP complex in the relaxed allosteric state at 2.7 Angstroms resolution," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 32, pp. E2958–E2966, 2013, doi: 10.1073/pnas.1311996110.

[3] Z. Yang, P. Májek, and I. Bahar, "Allosteric transitions of supramolecular systems explored by network models: Application to chaperonin GroEL," *PLoS Comput. Biol.*, vol. 5, no. 4, p. e1000360, 2009, doi: 10.1371/journal.pcbi.1000360.