Online Normal Mode Analysis with Bio3D WebApps

Lars Skjærven, Shashank Jariwala & Barry J. Grant

August 13, 2015

Background

Bio3D[1] is an established R package for structural bioinformatics that provides interactive tools for biomolecular structure, sequence and simulation data analysis [1,2]. The aim of this document is to introduce the new online WebApp interface to Bio3D functionality for **normal mode analysis of protein structures** [2].

Normal mode analysis (NMA) is one of the most popular simulation techniques used to predict and characterize large-scale collective motions in biomolecules. By large-scale collective motions we mean a process in which a substantial part of the protein moves as a unit relative to other parts. Many studies have now established the relevance of such motions to protein function [3,4,5] NMA models proteins as harmonic oscillating systems and classifies their possible deformations sorted with respect to their energetic cost. A commonly used technique for NMA is to treat the protein as a simplified network of C-alpha atoms connected by springs. This approximation facilitates the efficient calculation of normal modes and has shown remarkable agreement with experimental data [3].

Overview

The Bio3D **NMA WebApp** provides a rapid and rigorous tool for NMA of single protein structures. It is a one-page application with multiple sections, where each section (or row) corresponds to a separate analysis of the modes.

The first row provides options for user input. These include a user provided PDB code (left panel) and the force field specification (middle panel). This row also provides a rapid visualization of the input structure (right panel). Subsequent rows detail NMA results with each row providing a separate analysis or visualization of the calculation results. This includes an NMA derived **fluctuation profile** (row 1), **mode visualization** (row 2), **cross correlation analysis** (row 3), and finally **overlap analysis** (row 4).

---

[1] The latest version of the package, full documentation and further vignettes (including detailed installation instructions) can be obtained from the main Bio3D website.

[2] WebApps provide an online interface to select Bio3D functionality and negate the need for package installation and configuration on a local user machine.

(A) Normal mode analysis of the Ras protein

In the following example we will predict features of the functional motions of Ras GTPases with NMA. Ras proteins are important conformational switches that regulate signal transduction pathways related to cell growth and development. Their activity is regulated by a GTPase cycle that modulates the conformation of Ras and its affinity for binding and activating effector proteins that further propagate intracellular signaling.

1 INPUT protein identifier

To start our analysis, open a web browser and go to the **Bio3D NMA WebApp** (dcmb-grant-shiny. umms.med.umich.edu/nma-app). Note that the first section, or row, of the application is where initial user input to application is provided.

In the first panel of this row (**PDB Input Selection**), enter the the PDB code of Ras (4Q21) into the input text box (**Figure 1**).

Note: When the four characters have been entered the calculation will automatically start with a progress bar appearing at the very top of the screen to indicate that the server is working.

The second panel (**NMA parameters**) offers the selection of five different force fields to be used in the calculation of the normal modes. We will first use the default force field called calpha for this example.

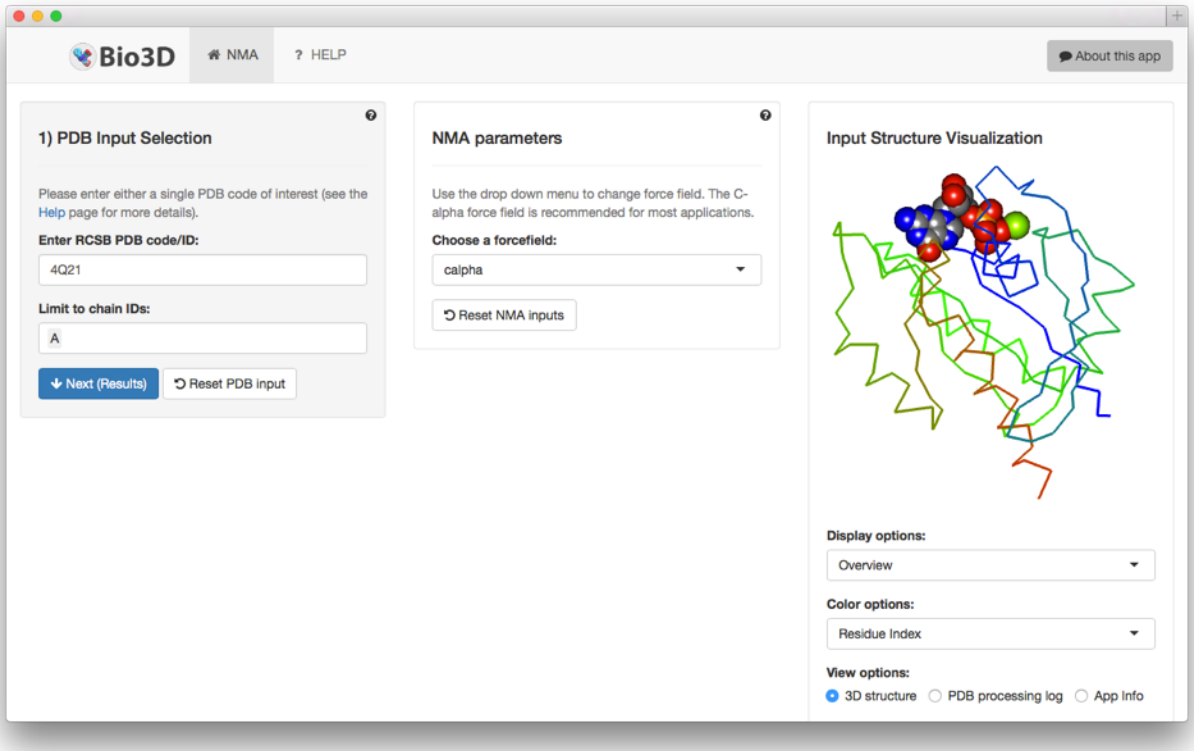


Figure 1: The Bio3D NMA WebApp is a one page application that is divided into five major steps, each represented by consecutive row. Background information on each tab and panel can be found by clicking the **About this app** button to the top right of the page and the small question mark icons in each panel.

The third and final panel of the first row (right hand side) provides a simple interactive **Input Structure Visualization**. Click and drag the mouse pointer over the protein to rotate, scroll to zoom, and select from the dropdown to apply different display and coloring options (**Figure 1**).

When the calculation of the normal modes is complete, click on the blue button **Next (Results)** or simply scroll down to explore the results of the first analysis, namely NMA derived fluctuations.

2 Fluctuation profile

The next row provides plotting options for the NMA derived **residue fluctuations**. Toggle the **Show B-factors** checkbox to compare the NMA derived fluctuations with the temperature beta factors obtained by X-ray crystallography.

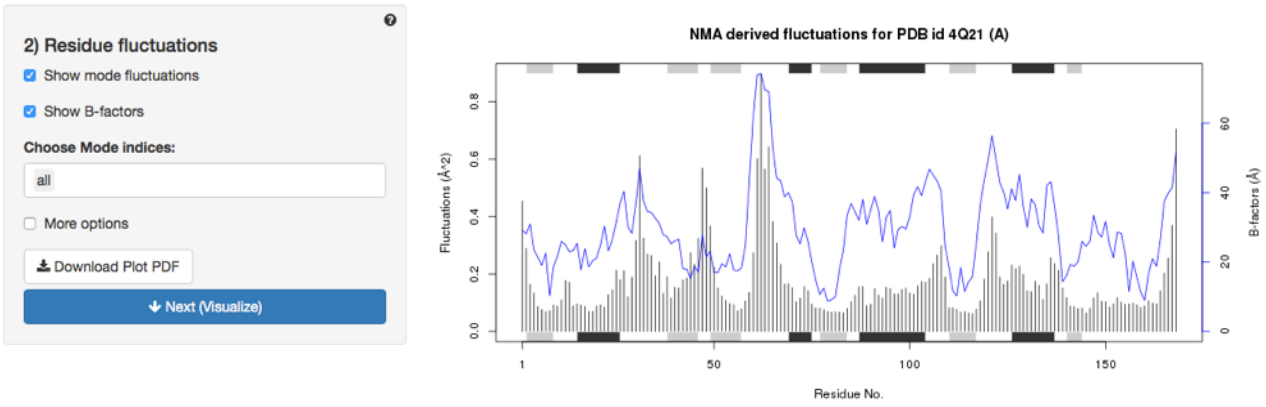


Figure 2: NMA derived fluctuations of the Ras protein 4Q21. Here the blue line depicts the experimentally obtained B-factor values, while the black bars depict the fluctuation values obtained from NMA. The light gray and black rectangles in the plot margins represent the locations of major secondary structure elements (gray for beta strands and black for alpha helices).

3 Normal Mode Visualization

This row offers visualization of the motions described by the calculated normal modes. Click and drag the mouse in the visualization window to get a feeling for the predicted motions. Change the **Color options** to Magnitude to color the structure by the magnitude of the motions, and increase the **Magnification factor** to amplify the displayed motions (**Figure 3**).

The individual normal modes can also be visualized in your favorite external molecular view (e.g. PyMOL or VMD). Click the **Download PDB Trajectory** button and open the trajectory in PyMOL. Note that the trajectory file contains only the C-alpha atoms and thus you may initially see isolated crosses or points displayed. To view the protein chain in PyMOL we can render with the ribbon representation by entering the following at the command line:

```
set ribbon_trace_atoms, 1
```

Click the play button in the lower window of PyMOL.

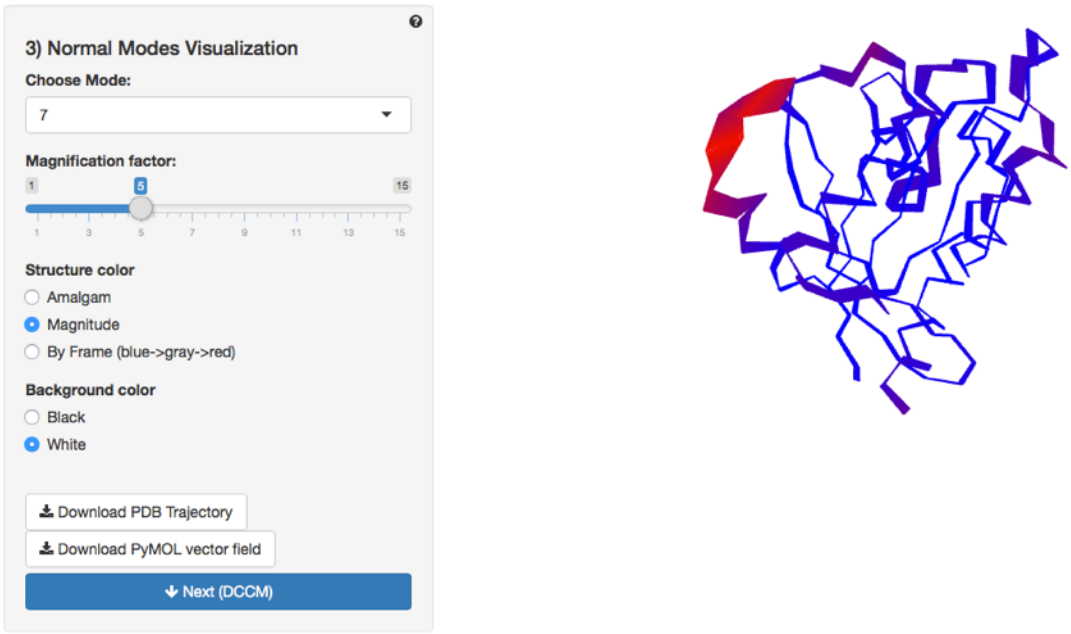


Figure 3: Trajectory view of the first normal mode of Ras. Here the structure is coloured according to the residual amplitudes of the motions.

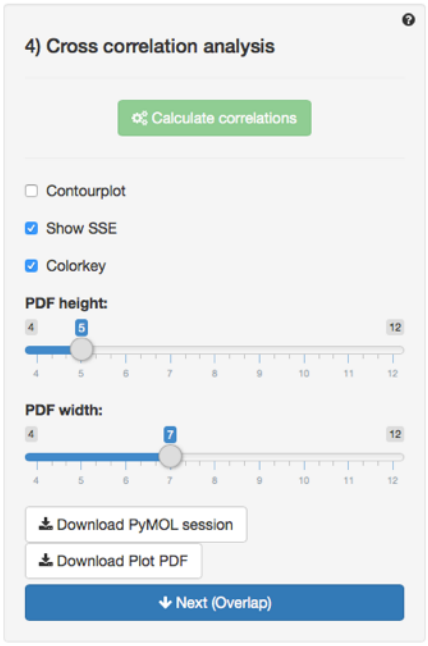
In the WebApp click the **Next (DCCM)** button to continue to the next analysis.

4 Cross correlation analysis

Cross correlation analysis aims to identify residues moving in a concerted manner, i.e. in the same direction at the same time. Detecting groups of amino acids moving in a correlated manner can aid in the understanding of the structure-dynamics-function relationship.

To calculate the cross-correlations click the **Calculate correlations** button. Once the calculation is complete the dynamic cross correlation matrix (DCCM) plot will be displayed on the right hand side (see Figure 4). This contour plot renders the values in the cross correlation matrix with values above and below 0.25 coloured according to a scale from pink (-1) to cyan (1). The marginal regions of the plot provide a schematic representation of major secondary structure elements.

Of particular interest is the off-diagonal elements showing positive correlation coefficient (e.g. > 0.5). We can visualise these correlations in the structure by dowloading the PyMOL session file (click the **Download PyMOL Session**) and open this in PyMOL (Figure 5).



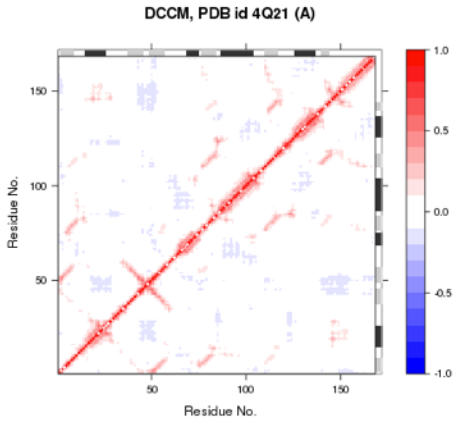


Figure 4: Cross correlation analysis of 4Q21. Toggle the **Contourplot** checkbox to obtain the plot in a slightly different variant.

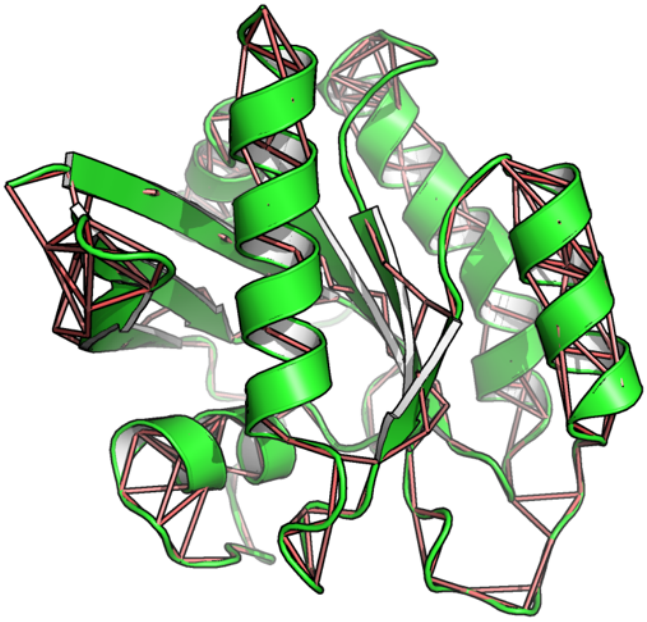


Figure 5: Cross-correlations visualised on-structure in PyMOL. Here the red lines depict a pair of residues with correlation coefficient > 0.4.

Click the blue **Next (Overlap)** button to continue to the next analysis.

5 Overlap analysis

Overlap analysis measures the similarity between a normal mode vector and a vector describing the difference between two known conformations of the same protein. An overlap value of 1 corresponds to identical vectors while a value of 0 correspond to two orthogonal vectors.

Click the **Launch PDB SEARCH** to identify structures in which the normal modes of our 4Q21 protein can be compared to. When the search is complete a list of available PDB structures is shown. Click the first entry (**1AA9_A**) and hit the **Calculate overlap** button. Once the calculation is complete a plot of the overlap values (black point) as well as the cumulative overlap values (red) are shown. In this particular case it seems like the first two normal modes of 4Q21 showing some similarity (dot product = ~0.2) to the structural difference between 4Q21 and 1AA9. Note that you can choose multiple structures from the table, and hit the **Calculate overlap** button. We will discus how to chose pairs of structures for such comparisons in the next workshop session on inter-conformer analysis.

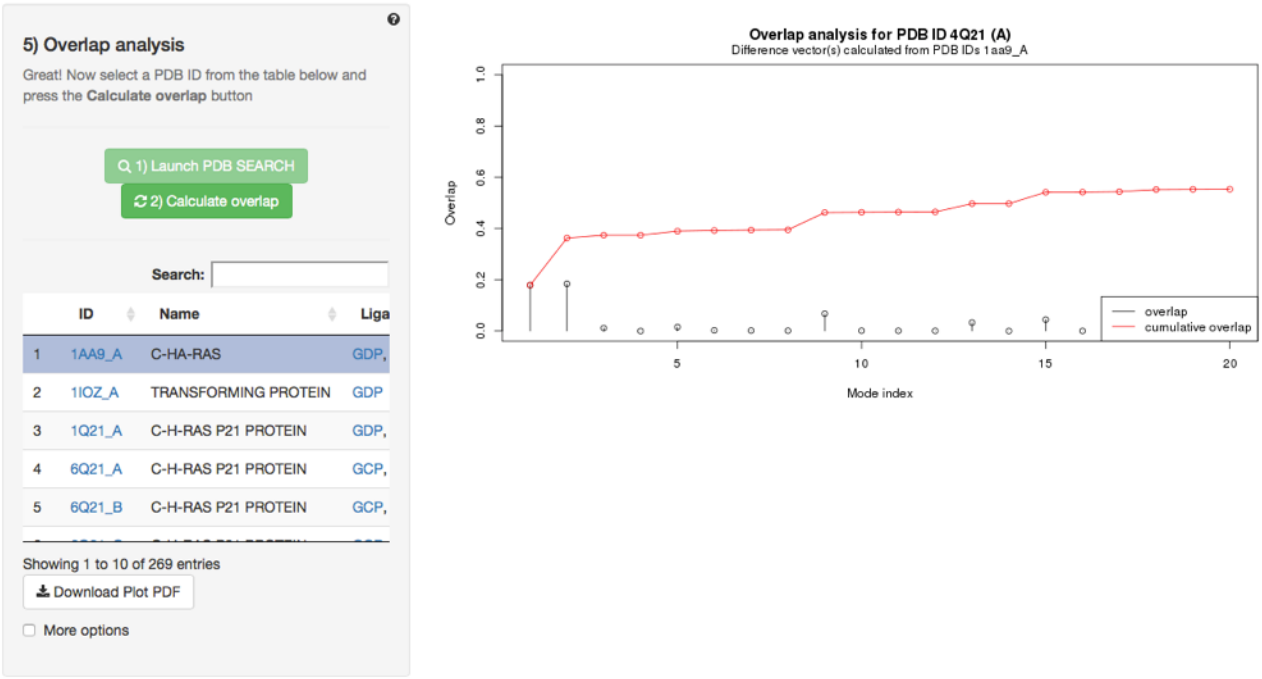
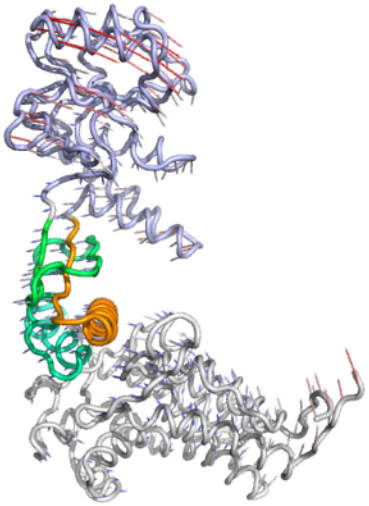


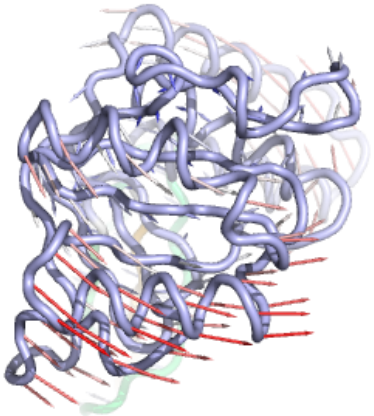
Figure 6: Overlap analysis shows that the two first modes of 4Q21 contribute the most to the conformational displacement between 4Q21 and 1AA9.

(B) Normal mode analysis of the GroEL subunit

In the following example we will explore the functional motions of the GroEL-GroES chaperonin on the subunit level. GroEL is an ATP-dependent molecular chaperone that assembles into a dual ring cylindrical structure (together with its co-chaperone GroES) to provide a protective chamber for protein folding. Each ring of the GroEL assembly consists of 7 identical subunits each of of 547 residues. The subunit folds into three distinct structural domains: the *equatorial*, *intermediate*, and *apical* domain. The apical domains are situated at the end of the cylinder and forms the entrance to the folding chamber. The equatorial domains contain the sites for nucleotide binding and are located at the equatorial region of the cylinder and therefore contain all inter-ring interaction points. The intermediate domain acts as a hinge between the apical and equatorial domains and provide flexibility to the assembly that enables large-scale conformational changes.



The functional cycle of GroEL-GroES is powered by ATP binding and hydrolysis, which drives a series of structural rearrangements that enable encapsulation and subsequent release of the substrate protein. In this tutorial we will explore these functional motions using NMA on an isolated subunit. We will investigate the normal modes through (1) mode visualization to illustrate the nature of the motions; (2) cross-correlation analysis to determine correlated regions; and (3) overlap analysis to determine which modes contribute to a given conformational change.



Here we will show an example of in which the normal modes shows an extraordinary similarity to the know conformational difference between the two states of the GroEL subunit.

- Enter **1SVT** as the input PDB ID and explore the first non-trivial mode (mode 7) in PyMOL and observe the rotation and downward motion of the apical domain. This corresponds very well with the motions described in the literature for the GroEL subunit (see [6]).
- Carry out the overlap analysis towards **1XCK_A** using the approach described in the previous section.

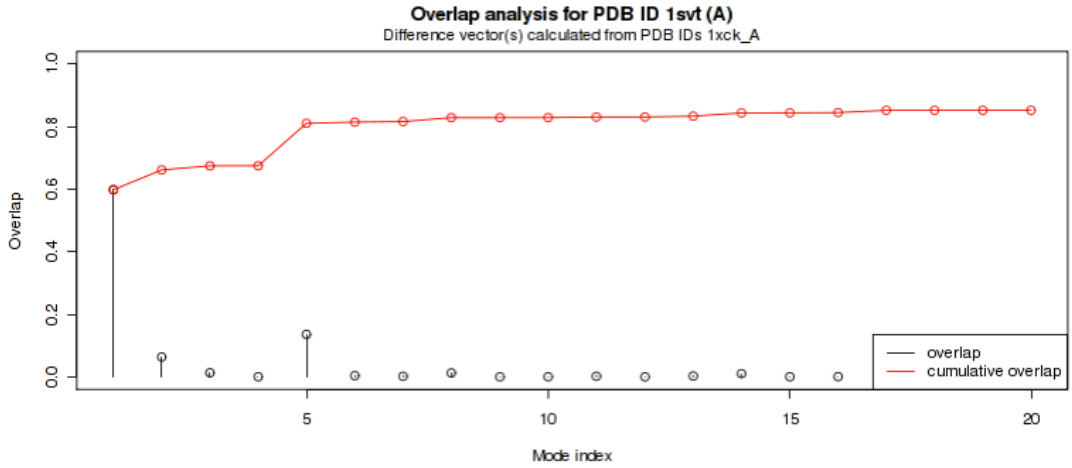


Figure 7: Overlap analysis between the normal modes of 1SVT (chain A) and 1XCK (chain A). The first normal mode shows an overlap value of 0.6 towards the conformational difference vector.

Summary

Here we have demonstrated how to perform and interpret **Bio3D WebApp NMA** with applications to two distinct protein systems, namely Ras and GroEL. In both systems conformational changes are critical for protein function. However, the nature and magnitude of these functional motions differ between each system. Conformational changes in Ras are more local - being largely restricted to active site loops and an associated short helix. In contrast, GroEL subunits undergo larger-scale collective motions that encompass large portions of the structure. Our NMA calculations performed extraordinarily well for GroEL (with predicted modes having a high similarity to the known conformational difference between two functional states). Despite the low overlap values obtained for our Ras example the regions predicted to be flexible are indeed those whose conformations change during Ras's functional cycle. These two examples where chosen to demonstrate the extremes of application utility in our quest to gain a deeper understanding of the mapping of structure to dynamics to function.

References

(1) Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD (2006) Bio3D: an R package for the comparative analysis of protein structures. Bioinformatics 22: 2695–6.

(2) Skjærven L, Yao Xq, Scarabelli G, Grant BJ (2014) Integrating protein structural dynamics and evolutionary analysis with Bio3D. BMC Bioinformatics 15: 1–11.

(3) Skjaerven, L, Hollup, SM, Reuter, N. (2009) Normal mode analysis for proteins. Journal of Molecular Structure: THEOCHEM. 898:42-48.

(4) Tama, F., & Sanejouand, Y. H. (2001). Conformational change of proteins arising from normal mode calculations. Protein Engineering, 14(1), 1–6.

(5) Bahar, I., Lezon, T. R., Yang, L.-W., & Eyal, E. (2010). Global dynamics of proteins: Bridging between structure and function. Annual Review of Biophysics, 39(1), 23–42.

(6) Clare, D. K., Vasishtan, D., Stagg, S., Quispe, J., Farr, G. W., Topf, M., Horwich, A. L., Saibil, H. R. (2012) ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin, Cell, 149(1), 113–23.