Supplementary Material

Folding cooperativity and allosteric function in the tandem-repeat protein class

Albert Perez-Riba*, Marie Synakewicz* and Laura S. Itzhaki

*Equally contributing authors
Department of Pharmacology, University of Cambridge
Tennis Court Road, Cambridge, CB2 1PD, United Kingdom

Contents

1 Supplementary Methods

1.1 Generation of Elastic Network Models

The Elastic Network Models (ENMs) were generated using the open-access ΔΔPT toolbox described in detail by Rodgers *et al.* [1]. PDB structures were reduced to C_α-atoms only and springs set between atoms i and j with an equilibrium distance, R, separated by a distance, r, and within a cut-off radius, R_c. The corresponding potentials are

$$V_{ij} = \begin{cases} \frac{k_{ij}}{2}(r_{ij} - R_{ij})^2 & R_{ij}^2 \leq R_c^2 \\ 0 & R_{ij}^2 > R_c^2 \end{cases}, \tag{1}$$

where k_{ij} is the spring constant of the potential. Unless otherwise stated, spring constants were set $k = 1$ kcal mol^{-1}Å^{-2} for all atom pairs, and the cut-off radius was set to $R_c = 12$. The calculated potential is used to construct a mass-weighted Hessian matrix, $\mathbf{D}$, with elements

$$\mathbf{D}_{i\alpha,j\beta} = \left.\frac{\partial^2 V}{\partial r_{i\alpha}\sqrt{m_i}\partial r_{j\beta}\sqrt{m_j}}\right|_R, \tag{2}$$

where m is the mass of the respective atom and α and β refer to the direction of motion. $\mathbf{D}$ is then diagonalized

$$\mathbf{e}^{-1}\mathbf{D}\mathbf{e} = \begin{pmatrix} \omega_1^2 & 0 & \cdots & 0 \\ 0 & \omega_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_n^2 \end{pmatrix} \tag{3}$$

and the eigenvectors of this matrix, $\mathbf{e}$, are the normal modes, ν, while the eigenvalues are the squares of the associated frequencies, ω_ν.

The frequencies from the eigenvalues can be used to calculate the free energy and entropy of each mode using

$$G_\nu = -k_B T \ln\left(\frac{1}{1-\exp(-\frac{\hbar\omega_\nu}{k_B T})}\right) \tag{4}$$

and

$$S_\nu = k_B\left(\frac{\frac{\hbar\omega_\nu}{k_B T}}{\exp(\frac{\hbar\omega_\nu}{k_B T})-1} - \ln\left(1-\exp\left(-\frac{\hbar\omega_\nu}{k_B T}\right)\right)\right), \tag{5}$$

where k_B is the Boltzmann constant and T the temperature which is set to $T = 298K$ [1].

Using the root mean square deviations (RMSDs) of the lowest 25 vibrational modes (ignoring the lowest 6 modes that account for rotation and translation) of an atom i, one can obtain a similar quantity to the crystallographic B-factor using

$$B_i = \frac{8k_B T\pi^2}{3m_i}\sum_\nu \frac{|\mathbf{e}_i^2(\nu)|}{\omega_\nu^2}. \tag{6}$$

Here, $k_B T$ only functions as a scaling factor. B_i gives a measure of the atom fluctuation about their equilibrium position, but since damping due to solvent or water does not exist, these fluctuations are very large. Therefore, they are scaled using average experimental and ENM B-factors of the whole structure and their respective root mean square displacements. The Pearson's Correlation Coefficient, r, between theoretical and experimental B-factors is used as a measure of the goodness of fit of the model to the experimental data, and are usually found to be > 0.5 [1, 2].

1.2 Qualitative comparison of protein dynamics

The $\Delta\Delta$PT toolbox can also be used to calculate the collectivity and cross-correlation of atoms within one protein [1]. Together they permit a qualitative description of atom motion, either with respect to other atoms or with respect to a normal mode.

The collectivity, κ, of a given mode, can be described using [3]:

$$\kappa_\nu = \frac{1}{N}\exp\left(-\sum_i^N \alpha|\mathbf{e}_i^2(\nu)|log(\alpha|\mathbf{e}_i^2(\nu)|)\right), \tag{7}$$

where N is the total number of atoms and α is the collectivity constant defined by $\sum_i^N \alpha|\mathbf{e}_i^2(\nu)| = 1$. Using κ_ν, it is possible to determine the fraction of atoms most affected by a mode ν. The lowest frequency modes tend to have $\kappa_\nu > 0.4$ [1].

The cross-correlation, C_{ij}, of atoms i and j over the lowest modes indicates how much they move into the same direction, and can be calculated using [4]:

$$C_{ij} = \sum_\nu \left(\frac{\mathbf{e}_i(\nu)\cdot\mathbf{e}_j(\nu)}{(|\mathbf{e}_i(\nu)|^2\ |\mathbf{e}_j(\nu)|^2)^{0.5}}\right). \tag{8}$$

For perfect correlation or anti-correlation $C_{ij} = 1$ or $C_{ij} = -1$, respectively. A value in between can arise from motion that is less correlated in terms of phase and/or period, or motions of atoms that are not (anti)parallel. If $C_{ij} = 0$, the atoms move with the same period and phase but their motions are orthogonal.

1.3 Quantitative comparison between ENM dynamics and conformational changes

The ENM eigenvectors of a given structure were compared to (i) the eigenvectors of, and (ii) the conformational changes between different structures. To calculate conformational changes in the backbones, structures were aligned and an RMSD was calculated using PyMol [5]. If this RMSD was significantly larger than the uncertainty in the backbone position (estimated by $\sqrt{\langle u^2\rangle} = \sqrt{B/(8\pi^2)}$ averaged over all backbone atoms, where B is the B-factor), a displacement vector, $\Delta\mathbf{R}_{ij}$, was obtained by simple vector addition $\Delta\mathbf{R}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ [6], where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the crystal coordinates of two different structures.

Comparisons between data sets (that is between (i) two eigenvectors $\mathbf{e}_i(\nu)$ and $\mathbf{e}_j(\nu)$ or (ii) an eigenvector, $\mathbf{e}_i(\nu)$ or $\mathbf{e}_j(\nu)$, and a conformational change, $\Delta\mathbf{R}_{ij}$) were made using the overlap, O, which is a measure of alignment between two vectors, as defined by:

$$O = \frac{|M \cdot N|}{||M||\ ||N||}, \tag{9}$$

where M and N are placeholders for any of the vectors $\mathbf{e}_i(\nu)$, $\mathbf{e}_j(\nu)$ or $\Delta\mathbf{R}_{ij}$ [6, 7]. If the directions align either in an antiparallel or parallel fashion, the overlap is 1; if they are exactly orthogonal, the overlap is 0. The cumulative overlap, CO, as defined by Yang *et al.* [7] is

$$CO(k) = \left(\sum_{j=1}^{k} O^2 \right)^{\frac{1}{2}} \tag{10}$$

which gives a measure of how well k normal modes overlap with a single vector, such as $\Delta\mathbf{R}_{ij}$. We chose to measure the cumulative overlap of the lowest five normal modes with conformational changes due to the mode swapping seen among different ENMs (see Fig. S3).

1.4 Preparation of PDB Structure files and optimization of ENMs

PDB files had to be modified to be suitable for ΔΔPT programs. If not specified otherwise, the binding partners and the peptide were removed to construct ENMs of RapH, RapF and RapJ, respectively. The structures had to be cut to their final size, which depended largely on which residues were shared among different structures of one protein. A particular challenge is imposed by downstream calculations such as overlap analyses, which require the ENMs to have the same number of atoms. The four Rap proteins share a large structural homology, particularly within the TPR domain [8]. Therefore, all four structures were aligned using PROMALS3D using both sequence and structural information (see Fig. S1) [9]. Missing residues that are conserved but not present in the structure were modelled using MODELLER [10]. Based on the PROMALS3D alignment, residues that (a) are flexible and cannot be modelled meaningfully using ENMs, (b) are not conserved and not represented in the crystal structure, or (c) are present in some structures but not others, were removed [1]. The final selection is displayed in Fig. S1 and Tab. 1. ENMs were also built for the TPR repeats only (Tab. 1).

Some structures had atoms with multiple possible positions between different crystal unit cells, e.g. due to different side chain rotamers. This is indicated by the occupancy, n, the likelihood of the atom adopting a given position. If $n \neq 1$, the atom position with the highest occupancy was kept and set to $n = 1$, while all others were deleted. If the occupancies for both positions equal to 0.50, the first position was kept, while the second one was deleted. Atoms with occupancies with $n \neq 1$ but only one given coordinate were kept and set to $n = 1$.

Optimal values for k and R_c were explored in the intervals of $0.5 < k < 2.0$ and $5 < R_c < 15$, respectively, by maximising the correlation between crystallographic and predicted B-factors. The spring constant was found to not have a measurable effect and correlations were maximal for $R_c > 11$Å for all proteins but RapJ. Correlations for the more extended conformations are consistently lower than 0.5, indicating that the experimental B-factors do not exhibit the large global flexibility seen in the network model, possibly due to crystallographic packing [11]. Therefore, final ENMs for all structures were built using default cut-off of $R_c = 12$Å for reasons of consistency and for finding an all-over optimum.

Table 1: The construct specific residues used to compute ENMs. See Fig. S1 for the corresponding structural alignment. The numbering is based on Uniprot sequences.

Protein	Chain	PDB ID	Residues removed prior to ENM	TPR domain
RapI	B	4i1a	1-13, 75-82, 375-391	100-391
RapJ	B	4gyo	1-7, 69-78, 88	96-373
RapH	A	3q15	1-7, 69-78, 88, 372-376	96-376
RapF	A	3ulq	1-7, 69-80, 90, 257, 375-381	98-381

2 Supplementary Figures

Figure S1: PROMALS3D structure-based sequence alignment of all four Rap proteins. The black boxes indicate residues that were removed before construction of ENMs, while the green line and arrow indicate the start of the TPR repeats.

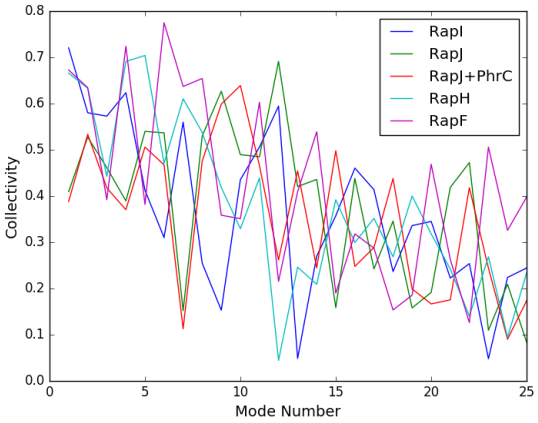


Figure S2: Collectivity of the lowest 25 normal modes, indicating the fraction of a protein involved in the motion of that particular normal mode. The collectivity is highest for the lowest normal modes. Values for RapJ and RapJ+PhrC models are consistently lower as some springs are set between sequence distant repeats due to the compact conformation, giving rise to a closed network and more isolated movements of the N- and C-termini.

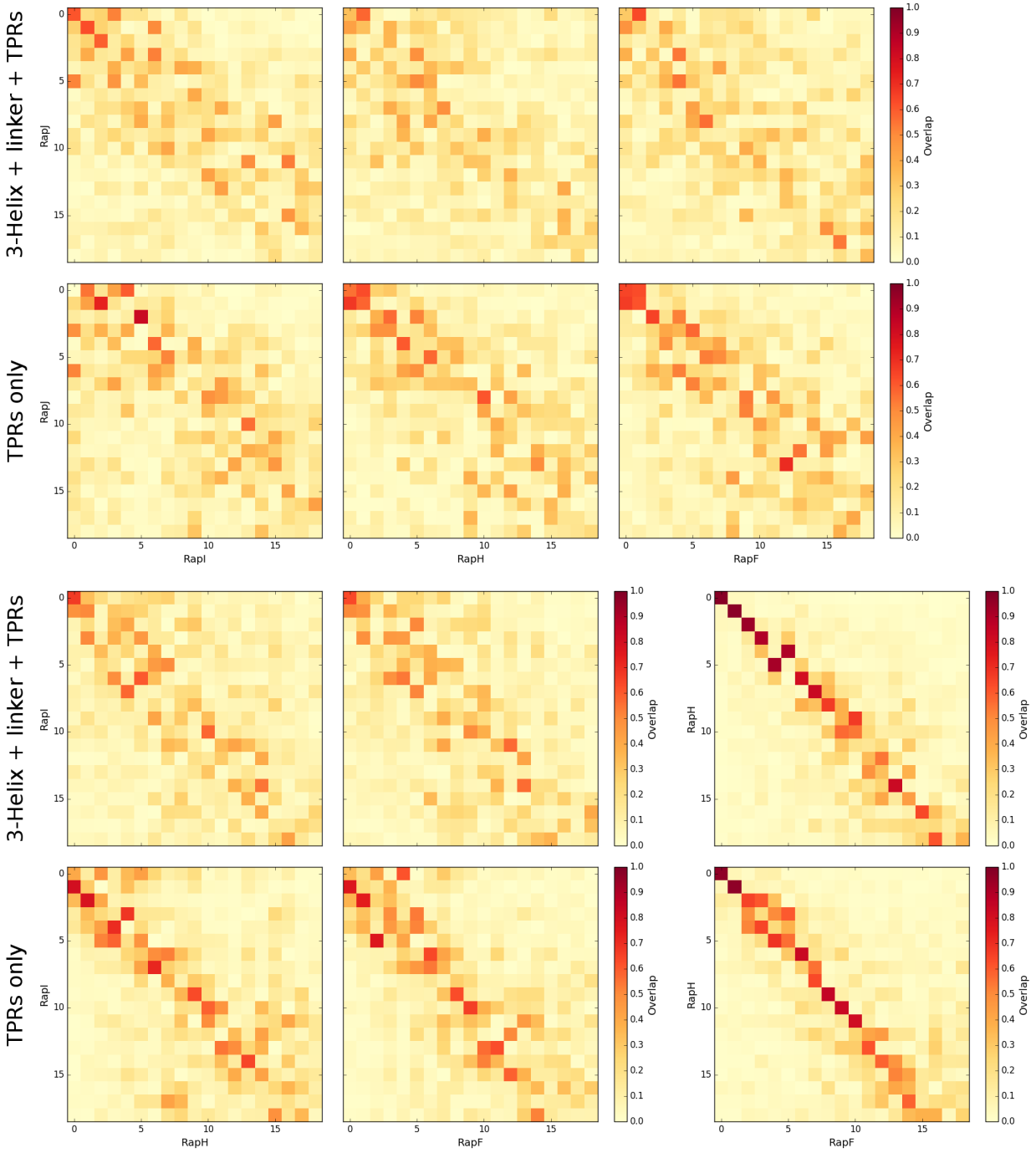


Figure S3: Overlaps between ENM normal modes of Rap proteins, for the whole proteins and TPR repeats only. If the normal modes of two proteins are nearly identical, the overlap matrix will have diagonal elements with $O \to 1$ and off-diagonal elements with $O \to 0$ (e.g. RapH vs RapF and Fig. S4). If normal modes are close in frequency their order can switch, giving rise to off-diagonal elements with a large overlap (e.g. RapI vs RapF/H). As expected, the structurally most similar Rap proteins, RapF and RapH, also exhibit large similarities in their dynamics. Due to its extended conformation RapI overlaps with RapF and RapH are high and concentrate in the diagonal. The dynamics of RapJ change dramatically upon compression of the superhelix and only the lowest normal modes overlap to a significant degree with the other structures ($O > 0.5$). When comparing ENMs of the TPR repeats only, similarities in motion are generally higher. This is expected, considering that the orientation and dynamics of the 3-helix bundle with respect to the TPR repeats differs between structures. The first six normal modes corresponding to rotational and translational motion have been omitted from this analysis.

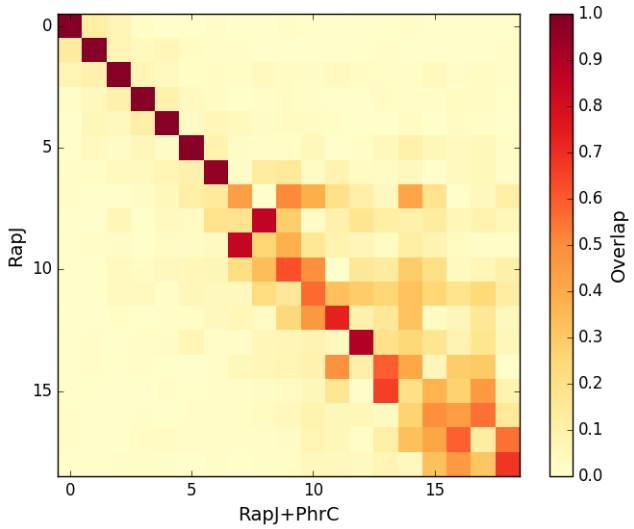


Figure S4: Overlaps between normal modes of ENMs of RapJ only and RapJ bound to the PhrC peptide. The presence of the peptide does not significantly influence the collective motion of the lowest 7 normal modes, most of them with overlaps of > 0.99. The peptide moderately affects the dynamics of the higher modes, which are localized to a few repeats at a time. Its presence is therefore negligible in comparisons of global dynamics between different Rap proteins.

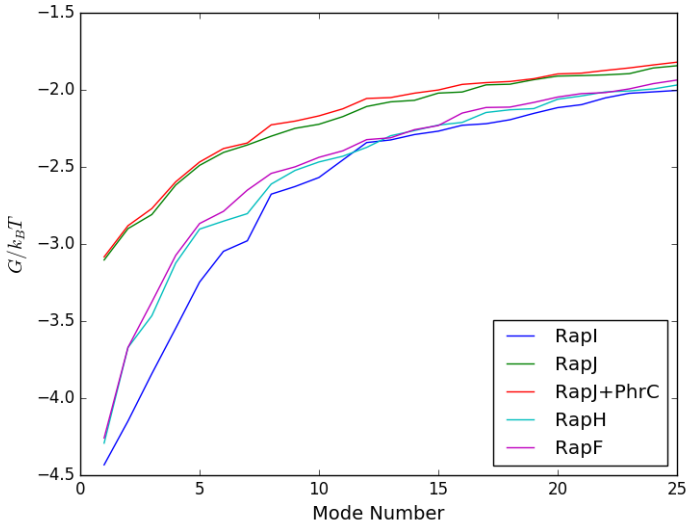
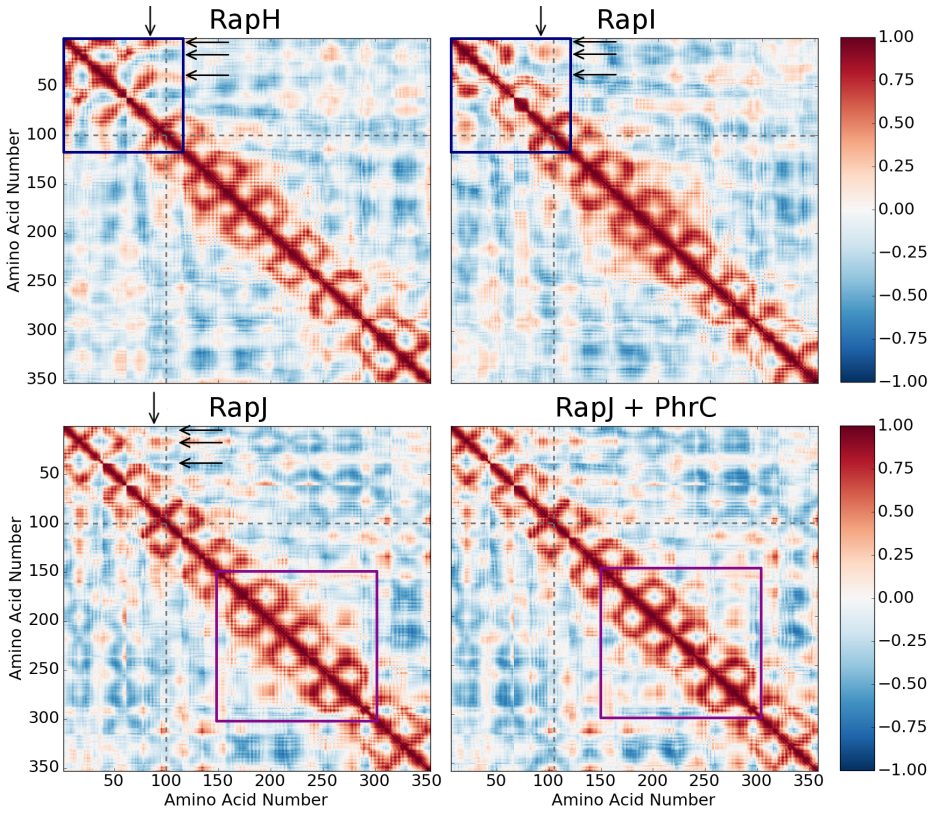
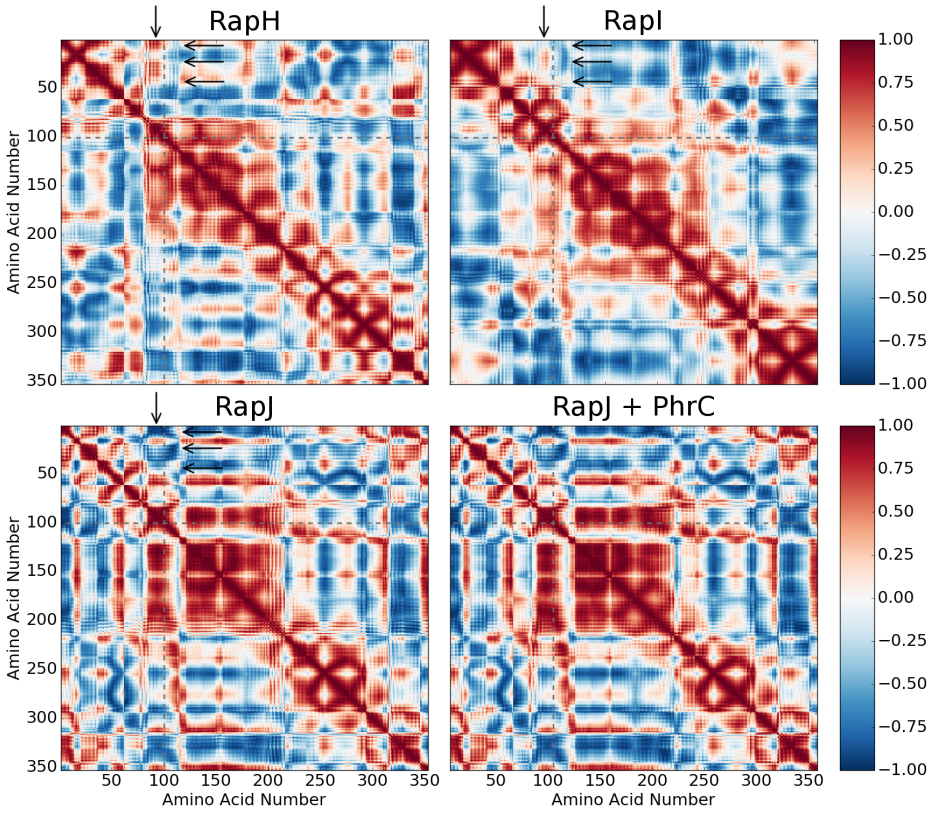


Figure S5: Free energy contribution of each mode for the respective Rap ENMs.



(a) Cross-correlation calculated over the slowest 10 normal modes



(b) Cross-correlation calculated over the slowest 3 normal modes

Figure S6: Different cross-correlation maps of Rap configurations calculated from the lowest 10 or lowest 3 vibrational normal modes. N-terminal helix bundle and TPR repeats are divided by grey dashed lines. The TPR repeats exhibit correlated motions only with their nearest neighbours, giving rise to the distinctive pattern of squares along the diagonal. However, the cross-correlations calculated from the lowest three modes exhibit a pattern in which the protein is divided into different subdomains, the borders of which shift depending on the conformational state. Most notably, the open conformation of RapI displays nearly symmetric correlation that are mirrored about the centre of the protein. As mentioned in the main text, movements of the rotated N-terminal 3-helix bundle, linker domain and first TPR motif (blue box) are non-TPR-like. However, the fewer high frequency modes are included to calculate the cross-correlation, (a) the more pronounced is the reversal in correlations once the N-terminal domain rotates (arrows), and (b) the less recognisable are any effects of peptide binding (purple box).

3 Supplementary Movies

3.1 RapI_mode1.mpg

This clip shows the motion of RapI along the lowest vibrational mode. The whole molecule bends along the central superhelical axis, changing the distance between the N- and C-terminal ends of the molecule.

3.2 RapI_mode2.mpg

This clip shows the motion of RapI along the second lowest vibrational mode. The whole molecule moves in a screw-like motion that loosens and tightens the superhelical twist.

3.3 RapI_mode3.mpg

This clip shows the motion of RapI along the third lowest vibrational mode. This motion involves largely the N-terminal three-helix bundle which twists in a screw-like manner orthogonal to the superhelical axis, while the C-terminal repeats simply open and close with respect to the superhelix.

References

[1] T. L. Rodgers, D. Burnell, P. D. Townsend, E. Pohl, M. J. Cann, M. R. Wilson, and T. C. McLeish. $\Delta\Delta$PT: a comprehensive toolbox for the analysis of protein motion. *BMC Bioinformatics*, 14(183):1–9, 2013.

[2] M. Delarue and Y.-H. Sanejouand. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *Journal of Molecular Biology*, 320:1011–1024, 2002.

[3] R. Brüschweiler. Collective protein dynamics and nuclear spin relaxation. *The Journal of Chemical Physics*, 102(8):3396–3403, 1995.

[4] T. Ichiye and M. Karplus. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Structure, Function, and Bioinformatics*, 11(3):205–217, 1991.

[5] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[6] F. Tama and Y.-H. Sanejouand. Conformational change of proteins arising from normal mode calculations. *Protein Engineering*, 14(1):1–6, 2001.

[7] L. Yang, G. Song, A. Carriquiry, and R. L. Jernigan. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure*, 16(2):321–330, 2008.

[8] V. Parashar, P. D. Jeffrey, and M. B. Neiditch. Conformational change-induced repeat domain expansion regulates rap phosphatase quorum-sensing signal receptors. *PLOS Biology*, 11(3):1–15, 2013.

[9] J. Pei, B.-H. Kim, and N. V. Grishin. PROMALS3D: a tool for multiple sequence and structure alignment. *Nucleic Acids Research*, 36(7):2295–2300, 2008.

[10] A. Šali and T. L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993.

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