Coarse-graining protein structures into their dynamic communities with DCI, a dynamic community identifier

Ambuj Kumar, Pranav M Khade, Karin S Dorman, Robert L Jernigan

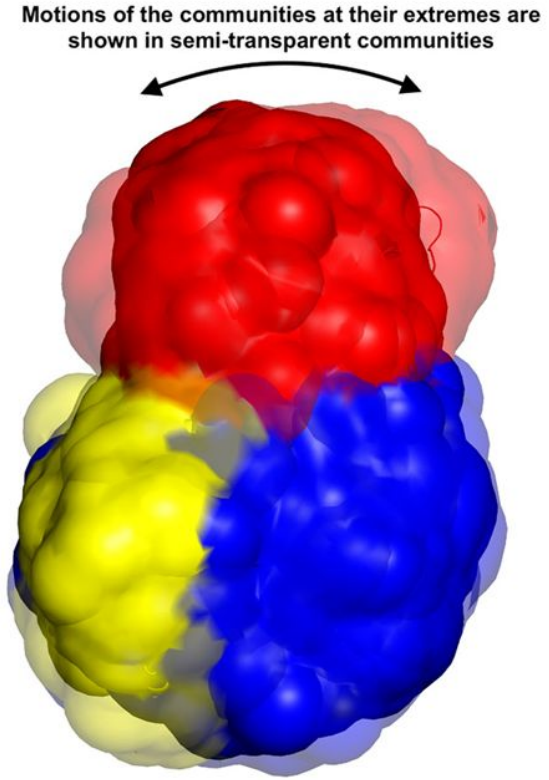
LifeLU reading group

presented by Özdeniz Dolu

14.11.2024

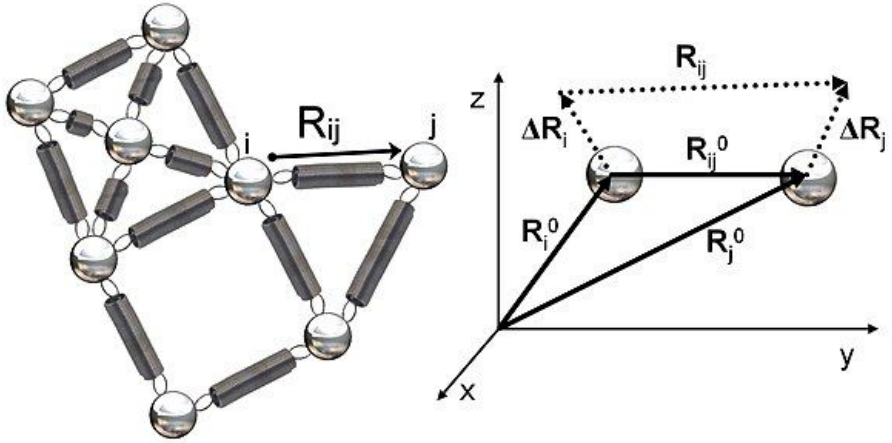
Core Idea

- Cluster the residues into “dynamical communities”.
- Dynamical communities are those in which the residues generally move together/coherently.
- An approach based on Gaussian Network Model and hierarchical clustering.



Interlude: Gaussian Network Model

- A coarse-grained method developed by Bahar, Atilgan, Haliloglu and Erman in 1997 that models the molecule as a network of springs and beads.
- Investigate the dynamics of “gaussian” fluctuations from equilibrium positions.
- Fluctuations are assumed “isotropic” i.e. equal in all directions.



Interlude: Gaussian Network Model

- Based on Kirchhoff matrix (Laplacian Matrix) (denoted with gamma).

$$V_{GNM} = \frac{\gamma}{2}\left[\sum_{i,j}^{N}(\Delta R_j - \Delta R_i)^2\right] = \frac{\gamma}{2}\left[\sum_{i,j}^{N}\Delta R_i \Gamma_{ij} \Delta R_j\right]$$

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \quad \text{and } R_{ij} \leq r_c \\ 0, & \text{if } i \neq j \quad \text{and } R_{ij} > r_c \\ -\sum_{j,j\neq i}^{N} \Gamma_{ij}, & \text{if } i = j \end{cases}$$

Interlude: Gaussian Network Model

- Normal mode analysis is done through the diagonalization $\mathbf{\Gamma = U \Lambda U}^T$ of the Kirchhoff matrix.
- First eigenvalue and eigenvector corresponds to the rigid body of the whole structure.
- Diagonalization extracts information about the cross-correlations between the fluctuations of various nodes.

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \quad \text{and } R_{ij} \leq r_c \\ 0, & \text{if } i \neq j \quad \text{and } R_{ij} > r_c \\ -\sum_{j,j\neq i}^{N} \Gamma_{ij}, & \text{if } i = j \end{cases}$$

Interlude: Anisotropic Network Model

- Also referred as Elastic Network Model, ANM is introduced in 2000 (Atilgan et al., 2001; Doruker et al., 2000) following the development of GNM.
- Still based on a modeling as a network of springs.
- Takes into consideration the directionality of fluctuations.
- Computationally more complex.

Interlude: Anisotropic Network Model

- ANM is based on Hessian matrix instead of the Kirchhoff matrix used in GNM.
- Hessian matrix is a matrix containing the second order partial derivatives of the harmonic potential of the springs between node i and node j.

$$\mathrm{H} = \begin{bmatrix} H_{ii} & H_{ij} \\ H_{ji} & H_{jj} \end{bmatrix}.$$

$$H_{ij} = \begin{bmatrix} \frac{\partial^2 V_{ij}}{\partial x_i \, \partial x_j} & \frac{\partial^2 V_{ij}}{\partial x_i \, \partial y_j} & \frac{\partial^2 V_{ij}}{\partial x_i \, \partial z_j} \\ \frac{\partial^2 V_{ij}}{\partial y_i \, \partial x_j} & \frac{\partial^2 V_{ij}}{\partial y_i \, \partial y_j} & \frac{\partial^2 V_{ij}}{\partial y_i \, \partial z_j} \\ \frac{\partial^2 V_{ij}}{\partial z_i \, \partial x_j} & \frac{\partial^2 V_{ij}}{\partial z_i \, \partial y_j} & \frac{\partial^2 V_{ij}}{\partial z_i \, \partial z_j} \end{bmatrix}.$$

$$V_{ij} = \frac{\gamma}{2}(s_{ij} - {s_{ij}}^o)^2$$

$$\frac{\partial^2 V_{ij}}{\partial x_i^2} = \frac{\partial^2 V_{ij}}{\partial x_j^2} = \frac{\gamma}{s_{ij}^2}(x_j - x_i)^2$$

$$\frac{\partial^2 V_{ij}}{\partial x_i \, \partial y_j} = \frac{-\gamma}{s_{ij}^2}(x_j - x_i)(y_j - y_i)$$

Back to the article: Methods

- Build the spring network with cutoff distance 7.0 A
- Calculate the Kirchhoff matrix.
- Calculate the pseudo-inverse of Kirschoff matrix. (pseudo-inverse: ignore the first eigenvalue and vector as it corresponds to rigid-body and its value is zero)
- From the inverse, calculate the correlations between residues.

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq 7.0\text{Å} \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > 7.0\text{Å} \\ -\sum_{i,\, i \neq j} \Gamma_{ij} & \text{if } i = j \end{cases} \qquad \Gamma^{-1} = \sum_{i=2}^{N} \frac{1}{\lambda_i} u_i u_i^T \,, \qquad C_{ij} = \frac{\Gamma_{ij}^{-1}}{\sqrt{\Gamma_{ii}^{-1} \times \Gamma_{jj}^{-1}}}$$

Methods

- Calculate dynamic distance matrix from the correlation values.
- Values of the dynamic distance matrix is used for the clustering algorithm.

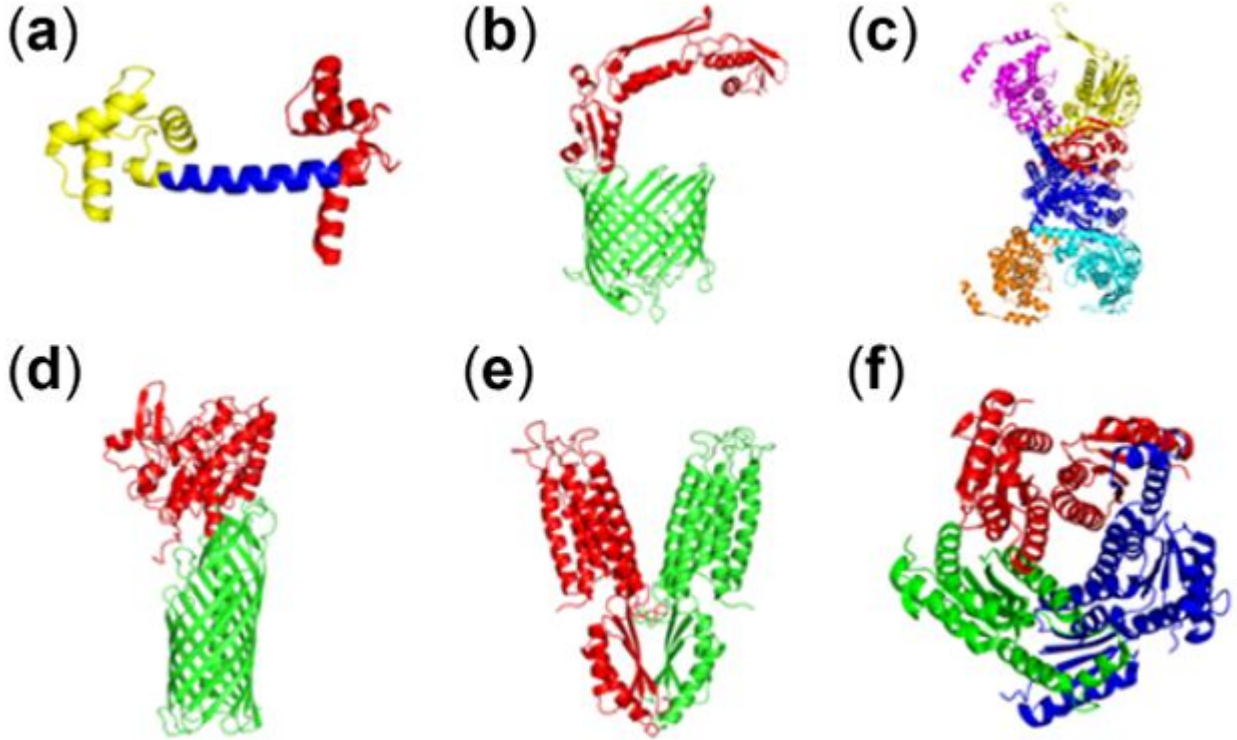
$$\mathcal{D}_{ij} = \sqrt{2\left(1 - C_{ij}\right)} \ ,$$

- Apply agglomerative hierarchical clustering algorithm to generate clusters.
- Number of communities (stopping condition for the algorithm) is decided (if not explicitly stated) automatically by the Calinski-Harabasz score (CH score for short).

$$CH = \frac{BCSS/(k-1)}{WCSS/(n-k)} \quad BCSS = \sum_{i=1}^{k} n_i \left|\left|\mathbf{c}_i - \mathbf{c}\right|\right|^2 \quad WCSS = \sum_{i=1}^{k} \sum_{\mathbf{x} \in C_i} \left|\left|\mathbf{x} - \mathbf{c}_i\right|\right|^2$$

Results

Selection of 6 results from DCI. A movie for (a) is supplied.

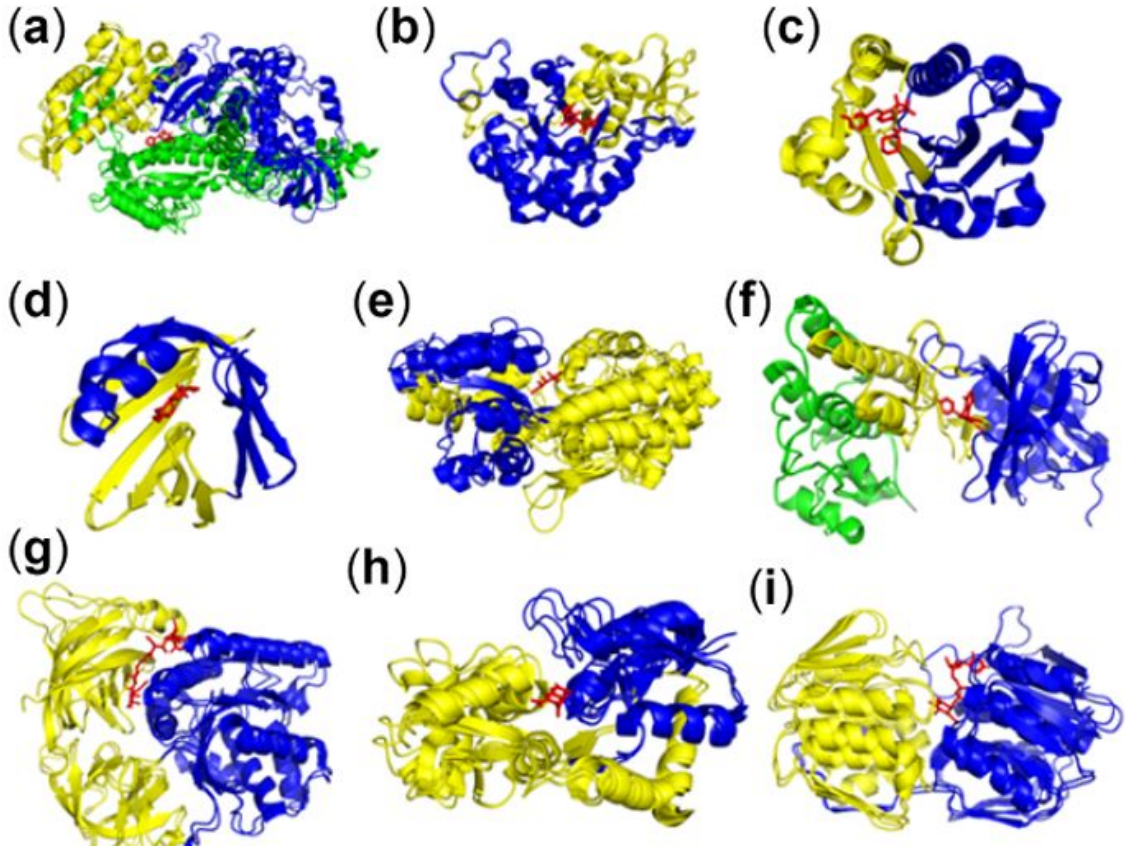


Results: Protein domain prediction using DCI

- In supplementary material table 1, it is shown that DCI method can identify 73 of the 98 domains for the globular proteins in SCOP database.
- Authors also remark that out of the remaining 25, by changing the number of communities (from the automatically selected optimal number), DCI was able to detect 23 of the remaining domains.

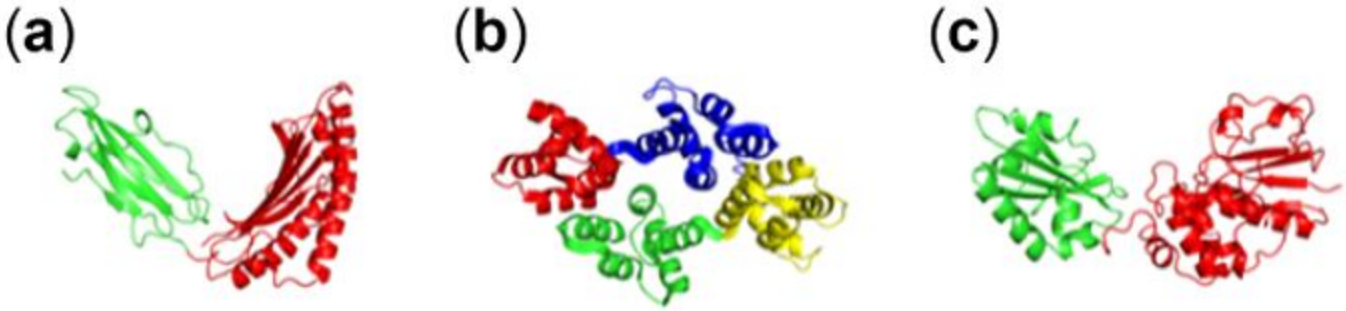
Results: Cryptic pocket comprises multiple communities

- For those proteins where apo (unbound) need to go through conformational changes to make its binding pocket accessible, pocket must be accompanied by multiple communities.
- Ligand is shown in red, and both holo and apo states are superimposed.



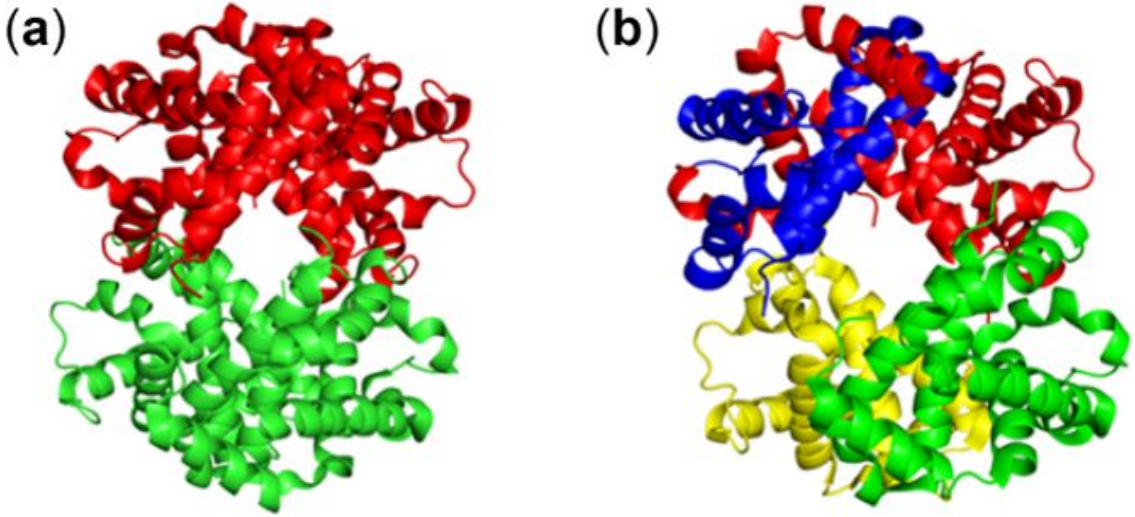
Results: Community boundaries indicate the hinge location for open-closed transitions

- Community boundaries found by DCI shows the hinge/linker locations.
- Movies 3-5 supplied for the corresponding proteins.



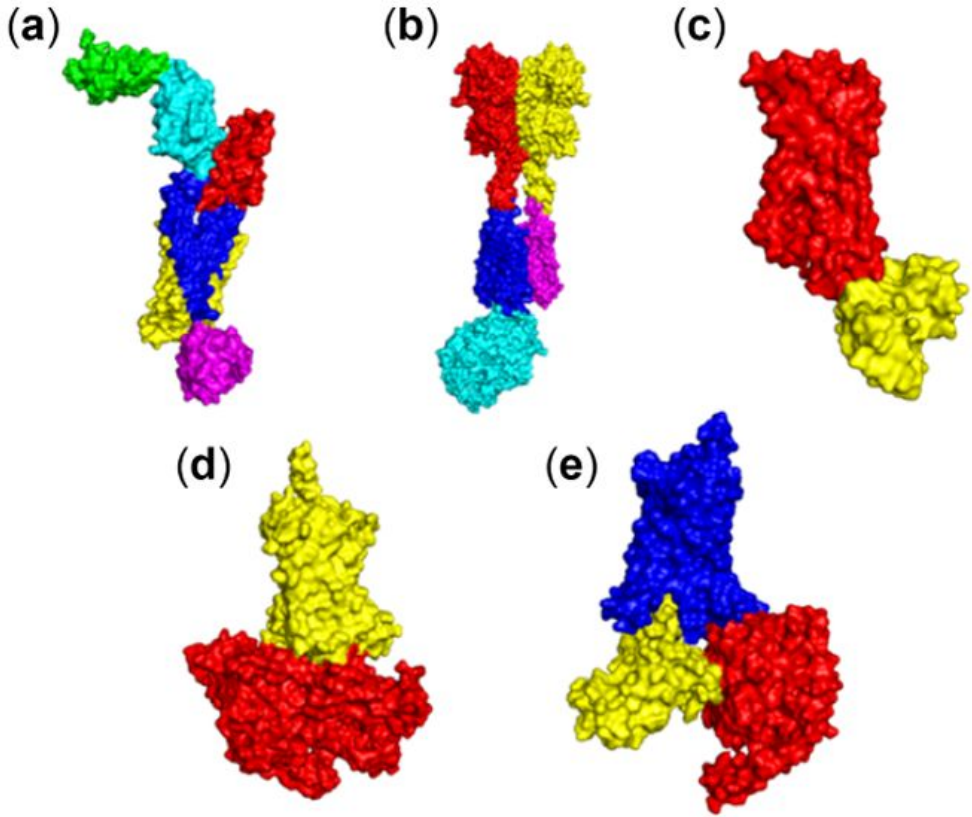
Results: Hemoglobin

- (a) shows communities for deoxy hemoglobin and (b) for oxy hemoglobin
- DCI is able to detect the increased degree of freedom in the existence of oxygen.



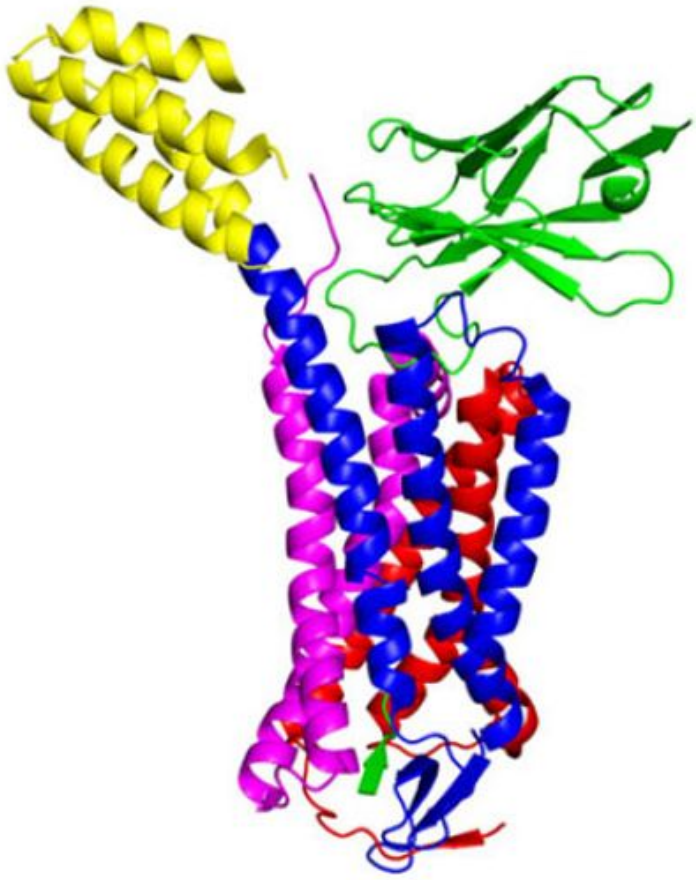
Results: G-Protein coupled receptors

- A selection of GPCR proteins and their dynamic communities



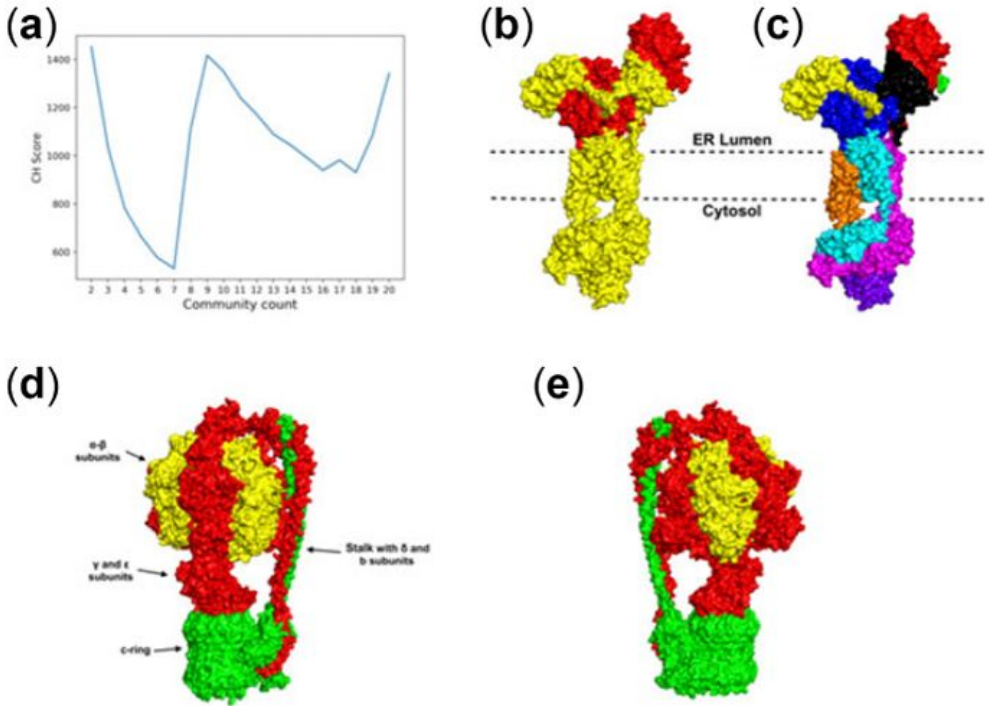
Results: G-Protein coupled receptors

- The structure where a membrane protein is bonded with a nanobody (shown in top-green).
- Note that bottom-green part belongs to the protein whereas top-green part belongs to the nanobody.
- Their allosteric relationship is captured by the dynamic community as they are both green.



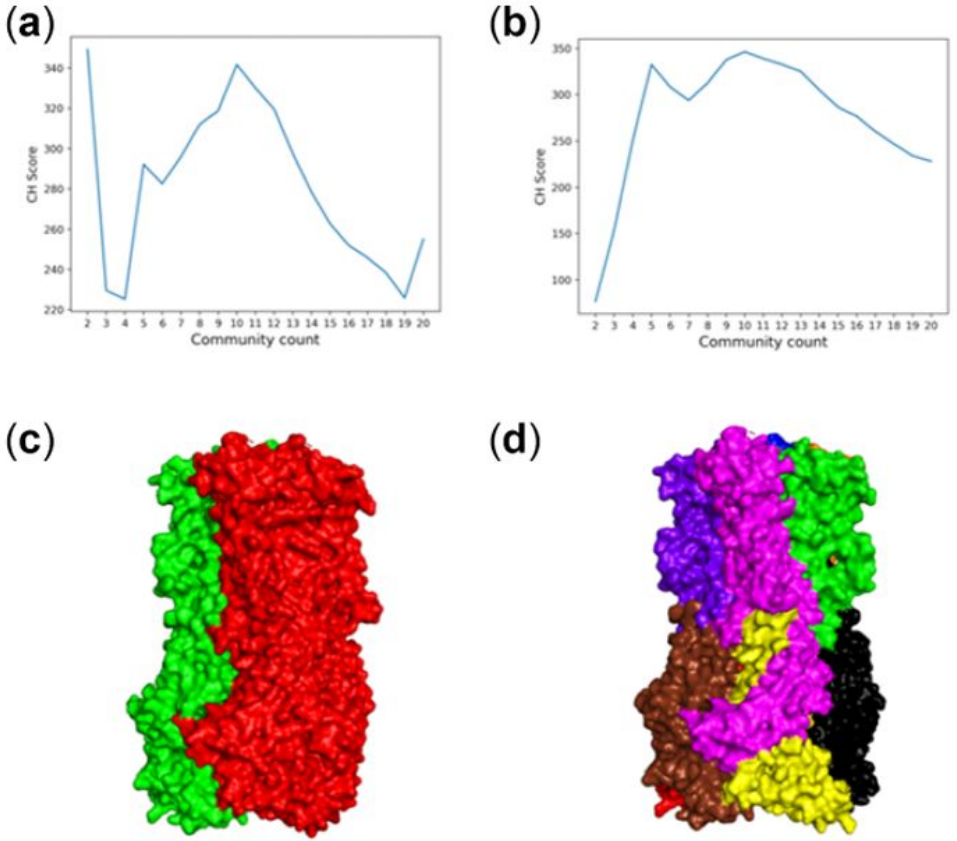
Results: Allosteric regulation among membrane protein communities

- 2 and 9 communities depending on 2 different CH scores is shown in (b) and (c).
- In (b) and (c), in the vertical, we see that some dynamic communities belong to more that 3 domains of the membrane protein showing an allosteric relationship between domains.
- In (d) and (e), a similar relationship is found between residues in physically distant domains



Results: Alpha4Beta2 nicotinic receptor

- Pentameric molecule found in various assemblages.
- (c) 3a:2b assemblage
- (d) 2a:3b assemblage
- Different assemblages have different binding affinities.



Conclusions

- DCI is a simple method extended upon GNM and hierarchical clustering and yet it can hint at various relationships between dynamical properties, function and structure.
- It can detect dynamical correlations between physically distant residues and allosteric relationships.
- It can find domain boundaries or hinge/linker locations.

Thanks for listening