# SUPPLEMENTARY MATERIAL FOR:

# Unravelling hierarchical levels of structure in lipid membranes

Alexandre Blanco-González[1,2,3], Rebeca García-Fandiño[2,*] Ángel Piñeiro[1,*]

**[1]Departamento de Física Aplicada, Facultade de Física, Universidade de Santiago de Compostela, E-15782 Santiago de Compostela, Spain. [2]Departamento de Química Orgánica, Center for Research in Biological Chemistry and Molecular Materials, Universidade de Santiago de Compostela, Campus Vida s/n, E-15782 Santiago de Compostela, Spain. [3]MD.USE Innovations S.L., Edificio Emprendia, 15782 Santiago de Compostela, Spain.**

## Content:

# Extended methods

### Local Reference Systems (LRS)

The LRS per grid as a function of time is described by a set of three orthonormal vectors:

The normal to each grid cell, which is identified with the normal to plane closest to the average position of the phosphate groups (PO4) of each cell and to the equivalent coordinates in the 8 adjacent neighbour cells. These 9 points are fitted to the equation of a plane:

$$a \cdot x + b \cdot y + c = z$$

That can be written in matrix form as:

$$P \cdot \boldsymbol{c} = \boldsymbol{z};$$

$$P = \begin{bmatrix} x_0 & y_0 & 1 \\ x_1 & y_1 & 1 \\ \dots & \dots & \dots \\ x_n & y_n & 1 \end{bmatrix}, \boldsymbol{c} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \boldsymbol{z} = \begin{bmatrix} z_0 \\ z_1 \\ \dots \\ z_n \end{bmatrix}$$

The nine points give an overdetermined system that can be fitted to get the coefficients $\boldsymbol{c}$ and the normal vector $\boldsymbol{n_i(t)}$ for each cell *i* as a function of time:

$$\boldsymbol{c} = (P^T \cdot P)^{-1} \cdot P^T \cdot \boldsymbol{z};$$

$$\boldsymbol{Z_i'(t)} = \boldsymbol{n_i(t)} = (1,0,a) \times (0,1,b)$$

The second vector of the orthonormal LRS is obtained by the normalized cross product of $\boldsymbol{Z_i'(t)}$ by the gradient of its projection on the Z axis of the global reference system:

$$\boldsymbol{X_i'(t)} = \left|\boldsymbol{Z_i'(t)} \times \vec{\nabla} \cdot \left(\boldsymbol{Z_i'(t)} \cdot \boldsymbol{Z}\right)\right|$$

Finally, the last compound of the basis is given by the cross product between the two first vectors:

$$\boldsymbol{Y_i'(t)} = \boldsymbol{Z_i'(t)} \times \boldsymbol{X_i'(t)}$$

### Singular Value Decomposition (SVD) analysis

SVD analysis allows to factorize a D(*m x n*) rectangular data matrix in two square - U(*m x m*) and $V^T$(*n x n*) - orthonormal matrices and a Σ(*m x n*) rectangular matrix:

$$D = U \cdot \Sigma \cdot V^T$$

From the previous equation, it follows that:

$$D \cdot D^T = U \cdot (\Sigma \cdot \Sigma^T) \cdot U^T$$

$$D^T \cdot D = V \cdot (\Sigma^T \cdot \Sigma) \cdot V^T$$

Thus, U contains the eigenvectors of ($D \cdot D^T$) and ($\Sigma \cdot \Sigma^T$) its eigenvalues, while V is equivalent to U upon replacing ($D \cdot D^T$) by ($D^T \cdot D$) and ($\Sigma \cdot \Sigma^T$) by ($\Sigma^T \cdot \Sigma$).

The data matrix used for our analysis has one row per grid cell and frame and one column per property associated to each cell. The number of grid cells per frame is 21x21 and the analysis is performed over 100 ns (4000 frames, considering that the data are stored for analysis every 25 ps) so m=21x21x4000=1764000. The number of properties determined for each cell are: Z-

coordinate of the average PO4 positions (topography), the three compounds of each of the vectors consisting the basis of the LRSs in absolute coordinates (9 scalar values), the angle between the surface normal vector and the Z axis of the global reference system, the angle between the vector tails and the same Z axis, the thickness and the average compounds of the lipid tails in the LRS. These properties amount for 39 columns in the data matrix for the SVD performed on the whole membrane. The matrix is filled with the values of these properties (P) normalized by their corresponding z-score ($p = \frac{P-\mu_P}{\sigma_P}$). The eigenvectors and eigenvalues of interest in this case will be $V^T$ and ($\Lambda = \Sigma^T \cdot \Sigma$), respectively, as they correspond to the local properties associated to each grid cell.

After the SVD is performed, the explained variance for each eigenvalue $\lambda_i$ in $\Lambda$is determined as:

$$\sigma_i^2 = 100 \cdot \frac{\lambda_i}{\sum_i^n \lambda_i}$$

And the cumulative variance:

$$C_{\sigma_i^2} = \sum_1^i \sigma_i^2$$

Finally, in order to determine which properties/variables are coupled to each other, we take advantage from the fact that $V^T$ is an orthonormal matrix so the sum of its squared coefficients $v_{i,j}^2$ both along its rows/eigenvectors *i* or columns/properties *j* equals 1. Therefore, the Relevance of each property along the eigenvector space can be defined as:

$$R_{ji} = R_{E=i}^{p=j} = \frac{v_{i,j}^2 \cdot \sigma_i^2}{\sum_i^n \left(v_{i,j}^2 \cdot \sigma_i^2\right)}$$

The variance values are employed to weight the compounds of $V^T$ in the definition of the relevance matrix because they represent the contribution of each eigenvector to the global variance of all the properties considered.

The columns of the Relevance matrix are then normalized, thus producing *n*-dimensional unitary vectors, and the coupling between different properties is calculated by means of the dot product of the different columns. An (*n x n*) coupling matrix for the different properties used in our analysis in then built:

$$\kappa_{rs} = (1 - \delta_{rs}) \cdot \sum_i^n \hat{R}_{ri} \cdot \hat{R}_{si}$$

where the term with the Kronecker delta is introduced to cancel self-couplings.

This analysis was repeated using the properties of each leaflet separately, thus using 20 columns for the D matrix, and for each lipid tail of each leaflet, thus using only 16 columns in the input matrix of the SVD analysis. All the results are included in this Supplementary Material and the employed properties are indicated in each plot.

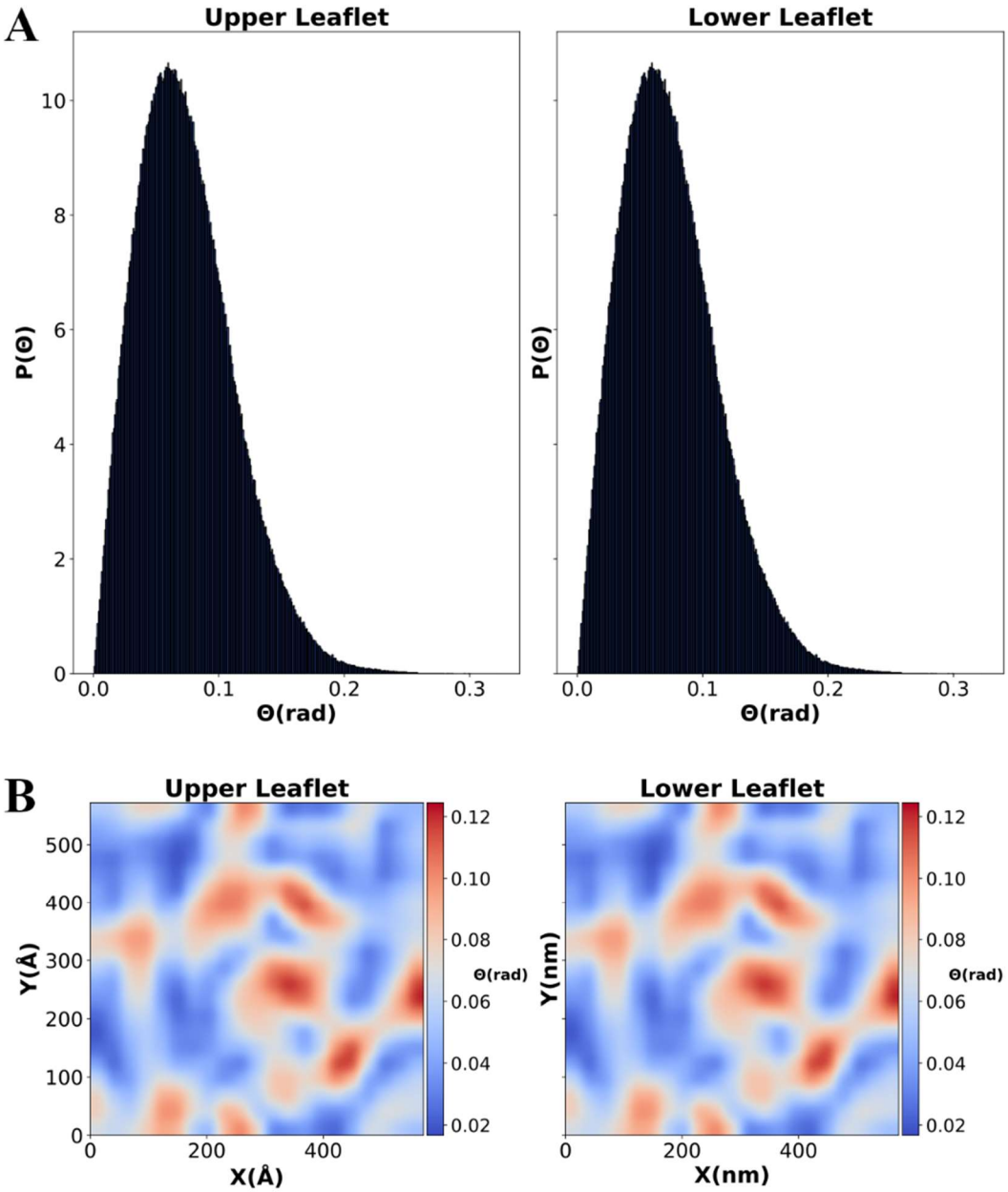


**Figure S1.-** Probability density distribution of the angles between the surface normal vector and the Z axis of the global reference system considering all the frames over the last 100 ns of the trajectory (A), and the associated heatmap in the XY plane obtained from the average of the same angles over the last ns of the trajectory (B).

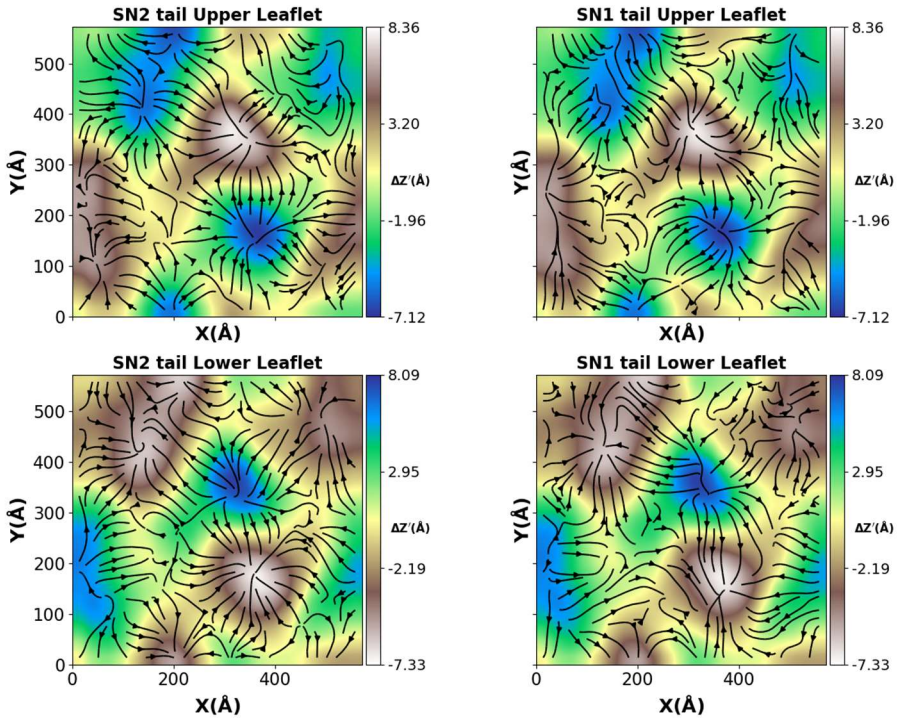


**Figure S2.-** Topographic heatmap and lipid tail vector fields for SN1 and SN2. Image obtained from the average of the vectors over the last ns of the trajectory.

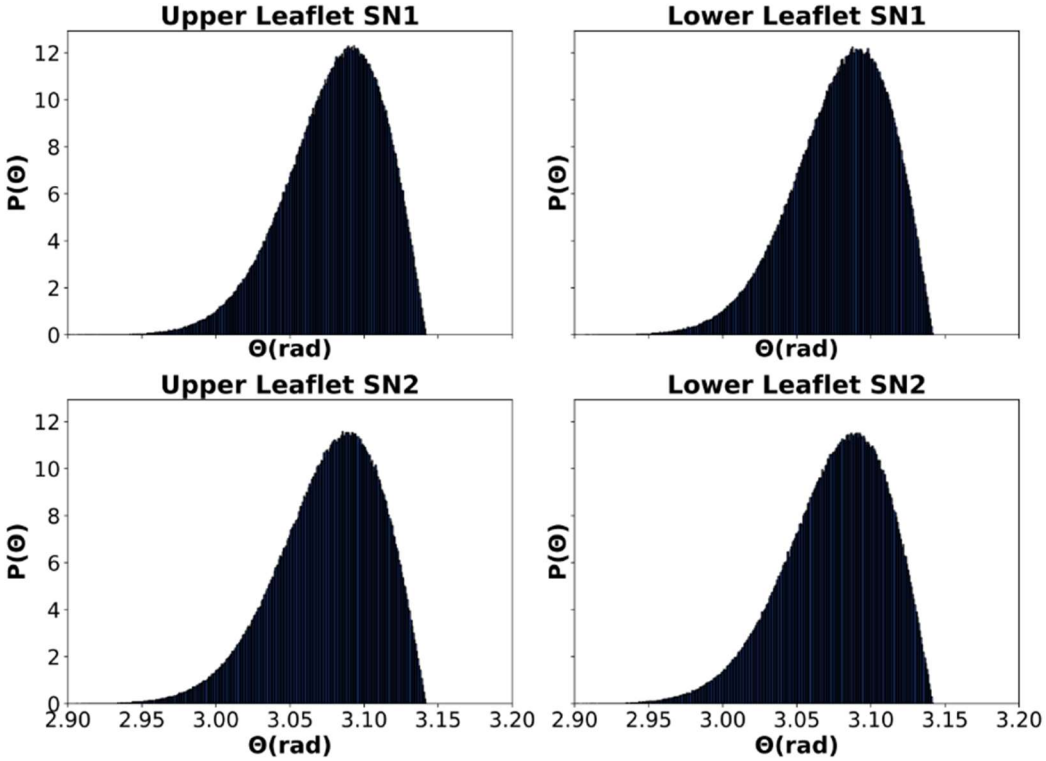


**Figure S3.-** Probability density distribution of the angles between the lipid tail vectors and the Z' axis of the LRS considering all the frames over the last 100 ns of the trajectory.

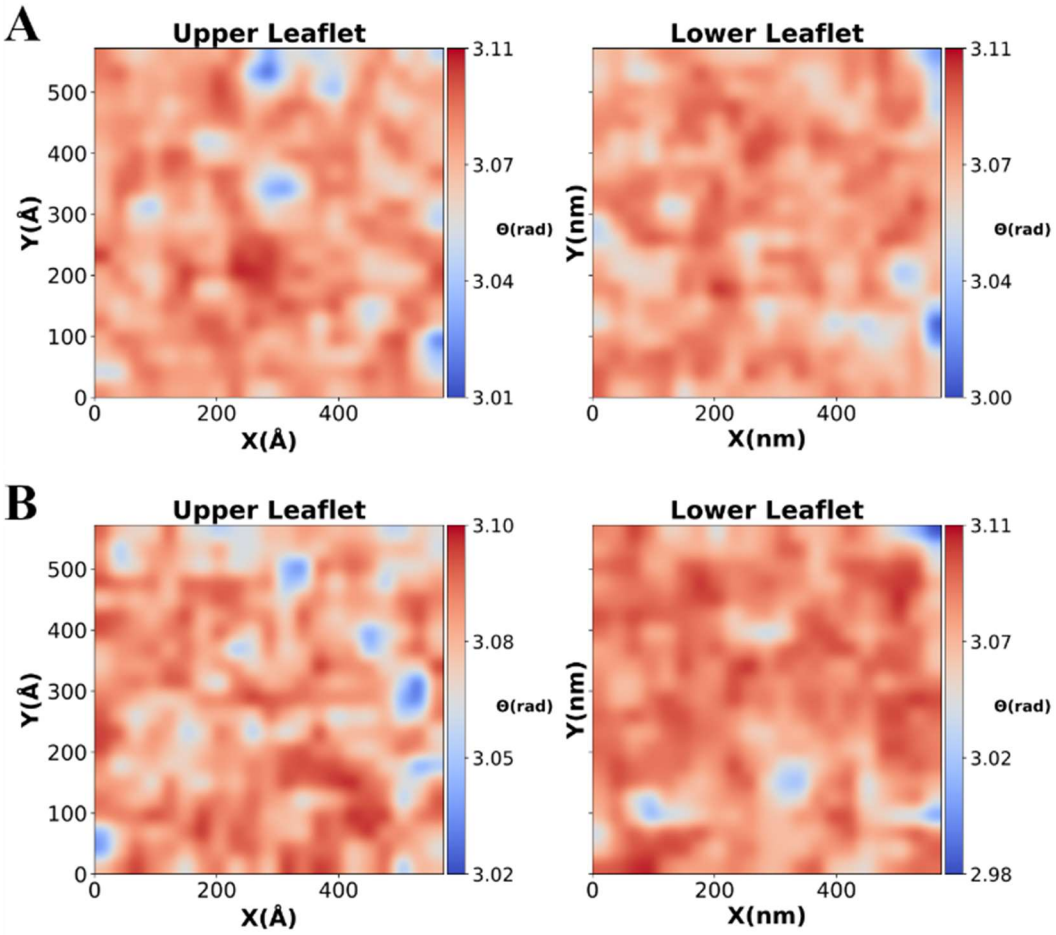


**Figure S4.-** Heatmap obtained from the average angles between the lipid tail vectors corresponding to SN1 (A) and SN2 (B), and the Z' axis of the local reference systems over the last ns of the trajectory.

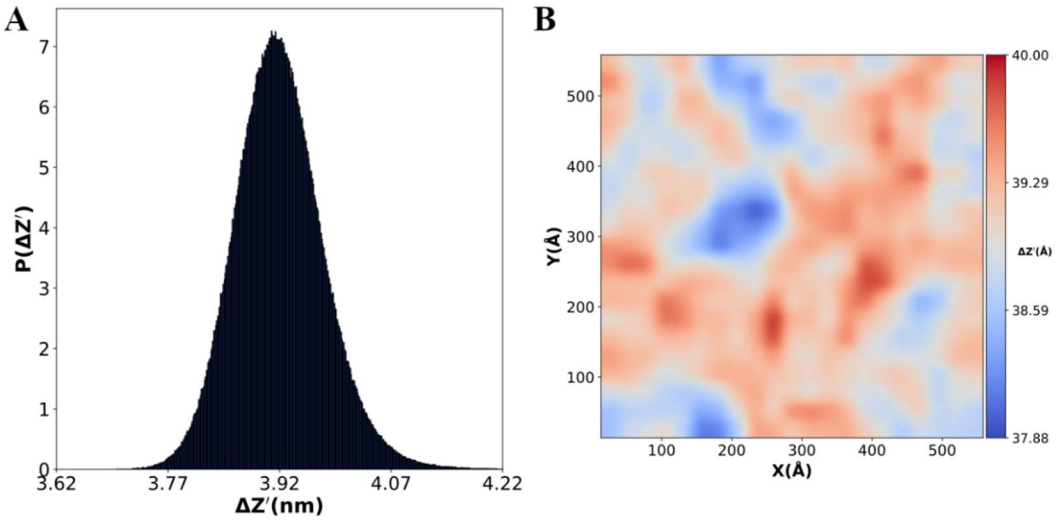


**Figure S5.-** Probability density distribution of the bilayer thickness obtained from the local reference systems, considering all the frames over the last 100 ns of the trajectory (A); and associated heatmap in the XY plane obtained from the average of the same thickness values over the last ns of the trajectory (B). Skewness and kurtosis of distribution are 0.46 and 1.96, respectively.

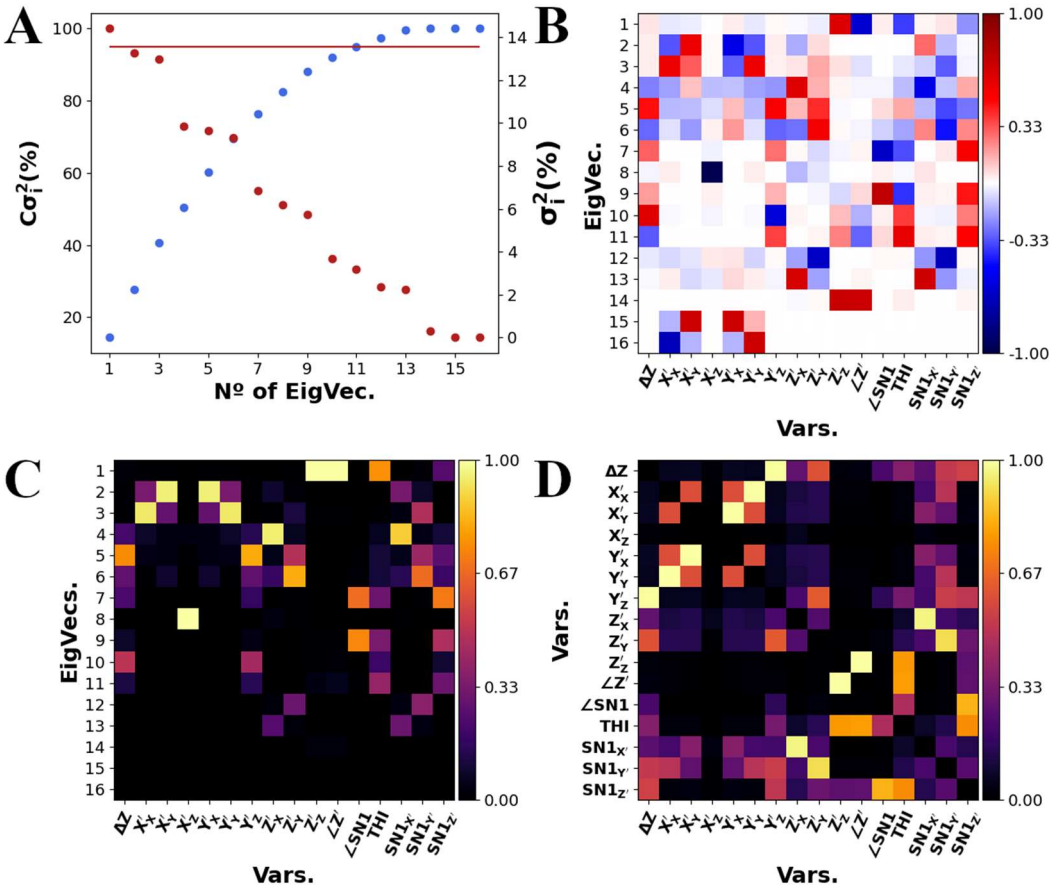


**Figure S6.-** Results for the SVD analysis performed on the upper leaflet and the SN1 tail. Variance (red) and cumulative variance (blue), 95% explained marked with a red line (A); $V^T$ matrix (B), relevance matrix (C), coupling matrix (D).

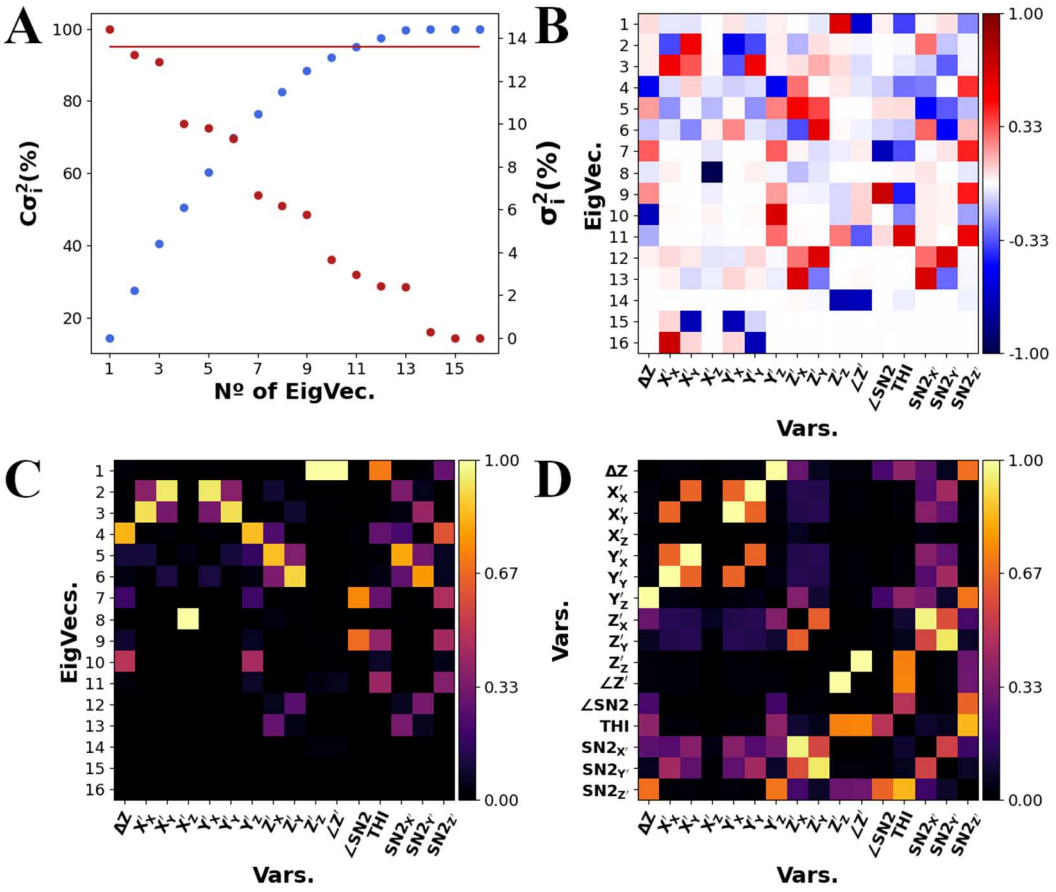


**Figure S7.-** Results for the SVD analysis performed on the upper leaflet and the SN2 tail. Variance (red) and cumulative variance (blue), 95% explained marked with a red line (A); $V^T$ matrix (B), relevance matrix (C), coupling matrix (D).

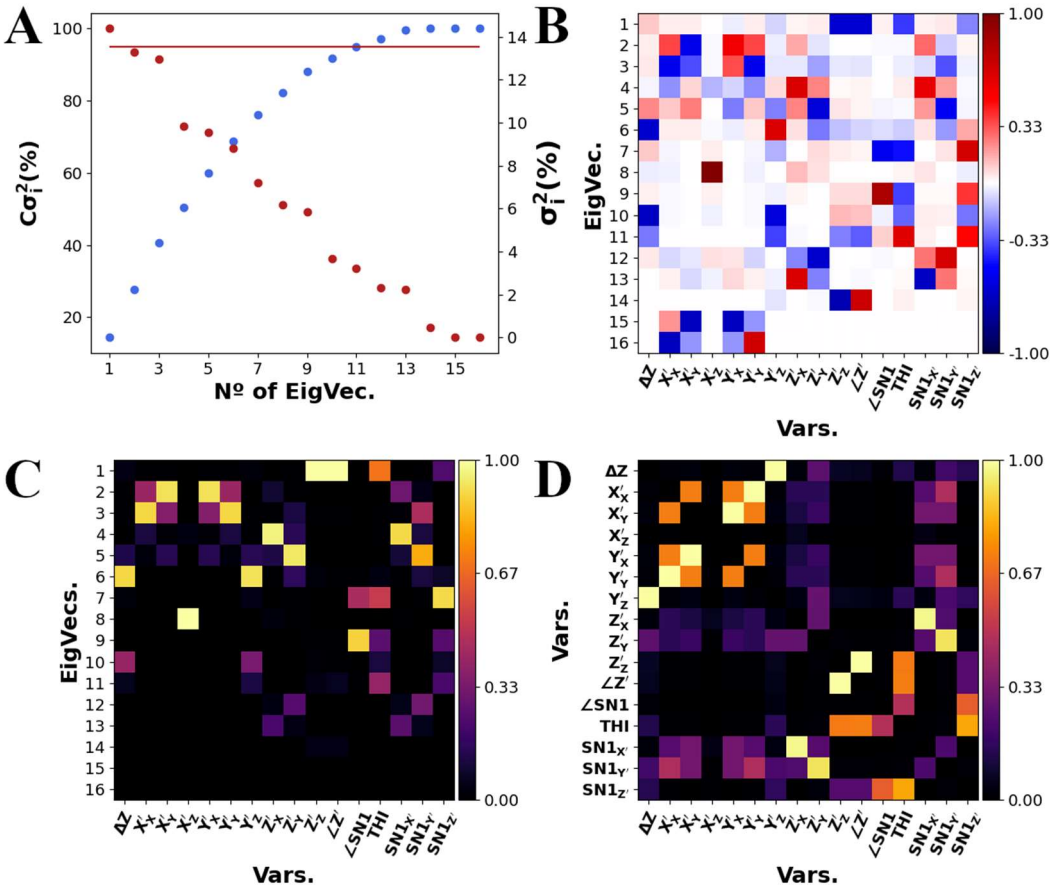


**Figure S8.-** Results for the SVD analysis performed on the lower leaflet and the SN1 tail. Variance (red) and cumulative variance (blue), 95% explained marked with a red line (A); $V^T$ matrix (B), relevance matrix (C), coupling matrix (D).

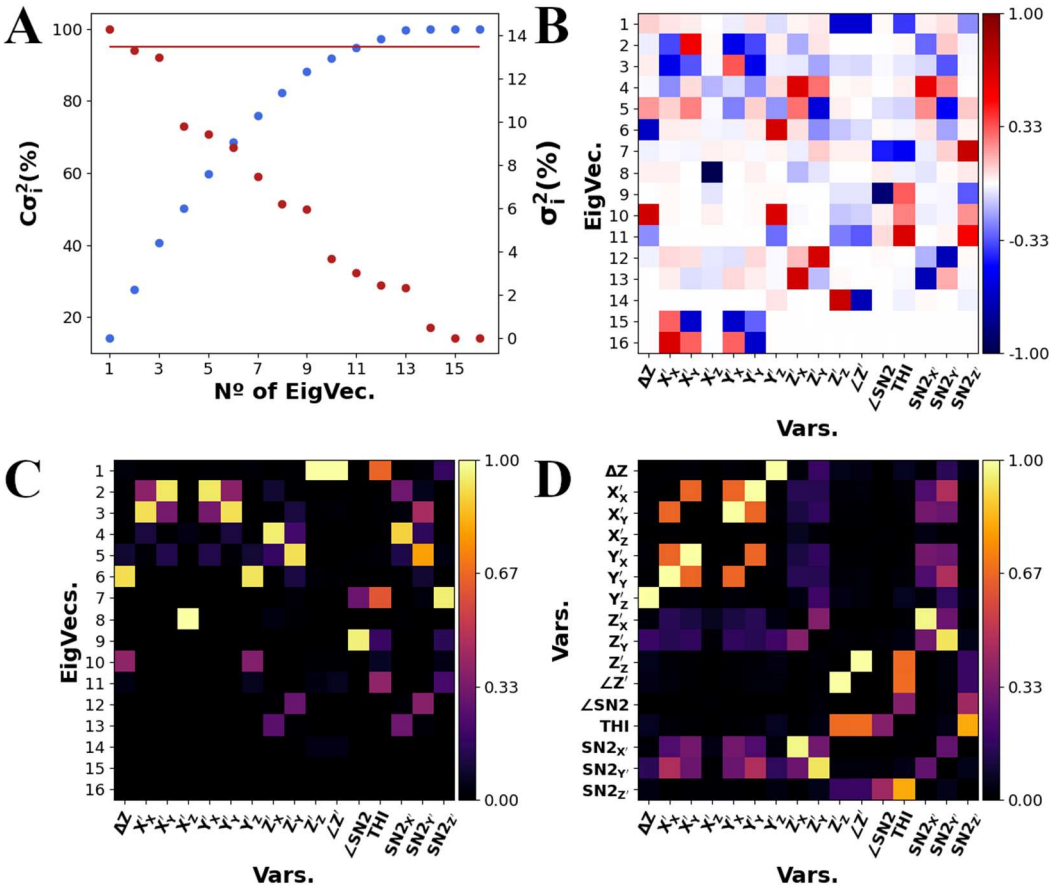


**Figure S9.-** Results for the SVD analysis performed on the lower leaflet and the SN2 tail. Variance (red) and cumulative variance (blue), 95% explained marked with a red line (A); $V^T$ matrix (B), relevance matrix (C), coupling matrix (D).

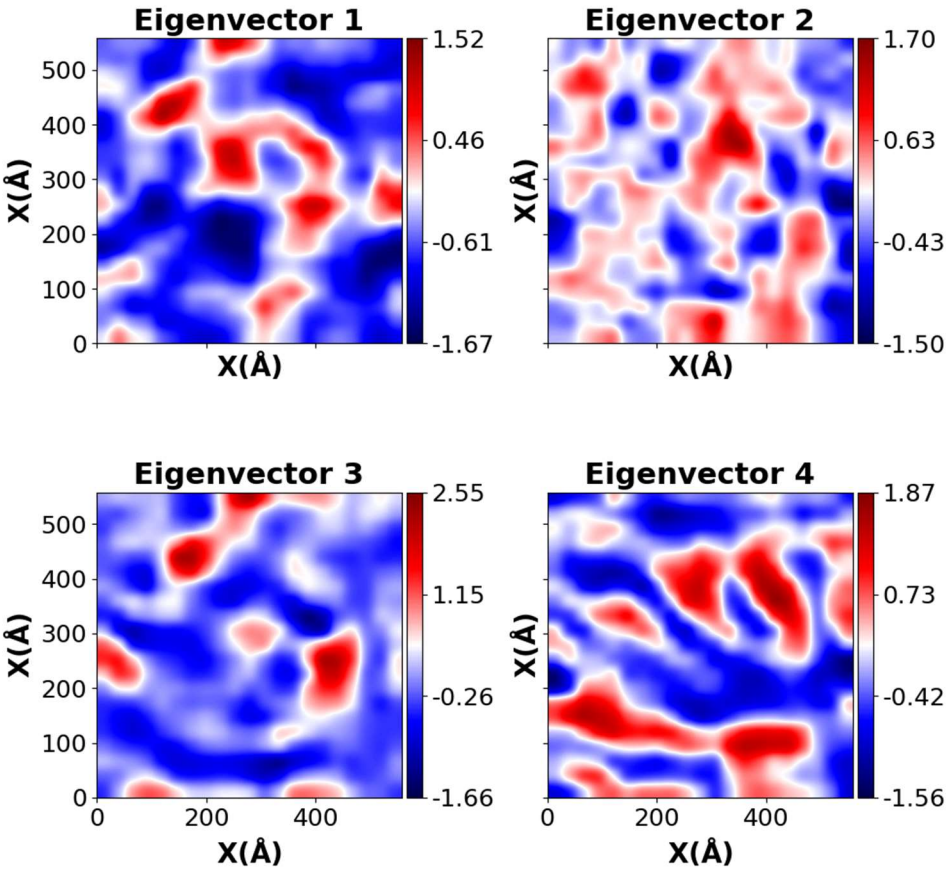


**Figure S10.-** Projection of the MD trajectory on 4 eigenvectors, as indicated on top of each panel. The results correspond to the average values over the last ns of our trajectory.

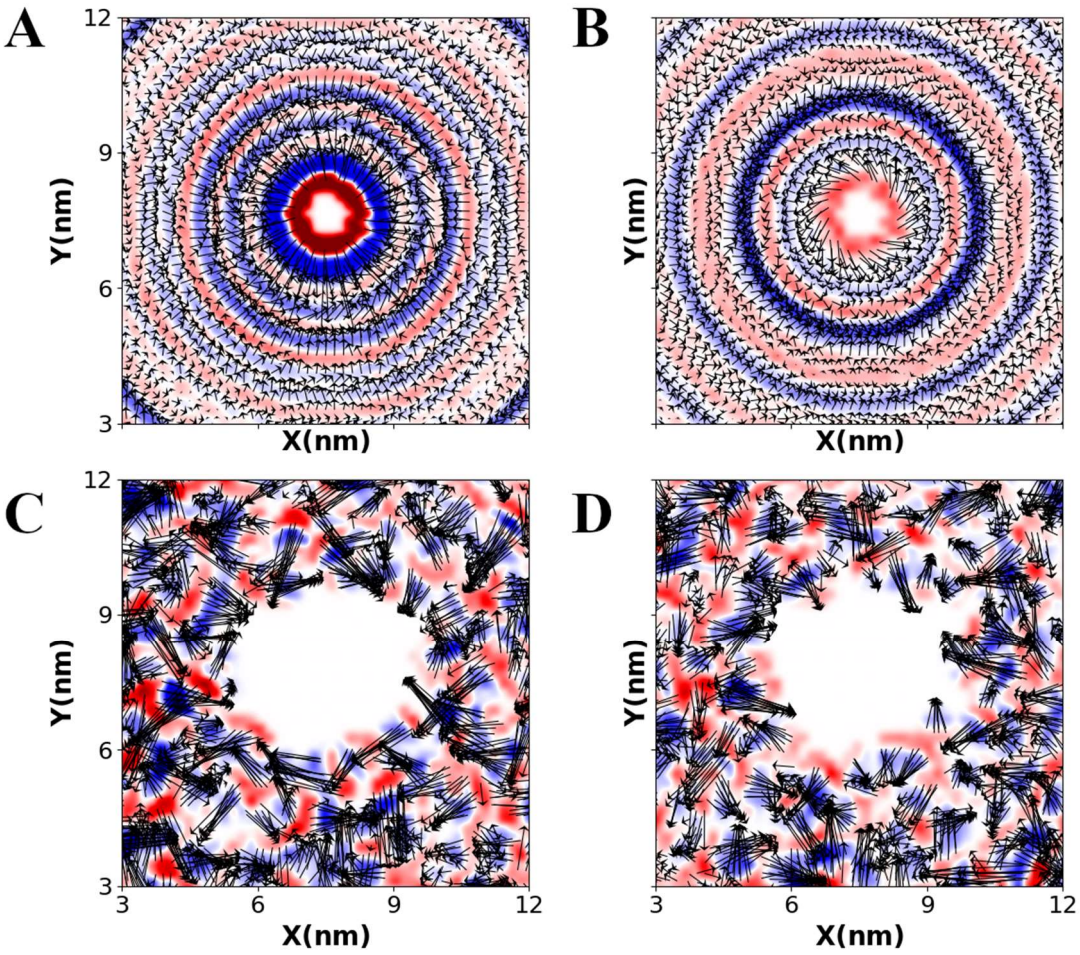


**Figure S11.-** Schematic representation of the quaternary structure in a DOPC lipid bilayer containing a carbon nanotube (A,B) and a b-barrel protein model (C,D). The vector fields obtained from the orientation of the lipid SN1 tail (A,C) and SN2 tail (B,D) in the lower leaflet are plotted together with the corresponding divergence represented in red-blue color gradient for divergence-convergence, respectively.