Computational and Structural Biotechnology Journal 21 (2023) 1390–1402

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Wordom update 2: A user-friendly program for the analysis of molecular structures and conformational ensembles

Angelo Felline [a,*], Simone Conti [b], Michele Seeber [a,*], Marco Cecchini [c], Francesca Fanelli [a,d,**]

[a] Department of Life Sciences, University of Modena and Reggio Emilia, via Campi 103, 41125 Modena, Italy
[b] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA
[c] Strasbourg Institute of Chemistry, UMR7177, CNRS, Strasbourg University, F-67083 Strasbourg Cedex, France
[d] Center for Neuroscience and Neurotechnology, via Campi 287, 41125 Modena, Italy

ARTICLE INFO

Article history:
Received 14 December 2022
Received in revised form 19 January 2023
Accepted 20 January 2023
Available online 27 January 2023

Keywords:
Molecular simulation analysis
Molecular Dynamics
Structural communication
Ion channel dynamics

ABSTRACT

We present the second update of Wordom, a user-friendly and efficient program for manipulation and analysis of conformational ensembles from molecular simulations. The actual update expands some of the existing modules and adds 21 new modules to the update 1 published in 2011. The new adds can be divided into three sets that: 1) analyze atomic fluctuations and structural communication; 2) explore ion-channel conformational dynamics and ionic translocation; and 3) compute geometrical indices of structural deformation. Set 1 serves to compute correlations of motions, find geometrically stable domains, identify a dynamically invariant core, find changes in domain-domain separation and mutual orientation, perform wavelet analysis of large-scale simulations, process the output of principal component analysis of atomic fluctuations, perform functional mode analysis, infer regions of mechanical rigidity, analyze overall fluctuations, and perform the perturbation response scanning. Set 2 includes modules specific for ion channels, which serve to monitor the pore radius as well as water or ion fluxes, and measure functional collective motions like receptor twisting or tilting angles. Finally, set 3 includes tools to monitor structural deformations by computing angles, perimeter, area, volume, β-sheet curvature, radial distribution function, and center of mass. The ring perception module is also included, helpful to monitor supramolecular self-assemblies. This update places Wordom among the most suitable, complete, user-friendly, and efficient software for the analysis of biomolecular simulations. The source code of Wordom and the relative documentation are available under the GNU general public license at http://wordom.sf.net.

1. Introduction

Wordom is a world-widely used software for fast manipulation and analysis of molecular structures and trajectories from molecular simulations. The original version of Wordom and its first update were published in 2005 [1] and 2011 [2], respectively. Other available software for structural dynamics analysis include: a) molecular simulation and analysis packages, like CHARMM [3,4], Gromacs [5,6], and Amber [7]; b) molecular viewers like VMD [8] and Pymol [9]; c) command-line analysis software and script-suites like MMTSB [10], carma [11], and pcazip [12]; and d) R or Python packages like Bio3D [13,14], MMTK [15], Biskit [16], ProDy [17], MD-TASK [18], MDTraj [19], and MDAnalysis [20].

For Wordom, the implementation of basic manipulation tools paved the way for the building of a modular framework to easily add analysis routines. The first update of Wordom implemented tools for secondary structure assignment, surface area calculations, and elastic network model (ENM) building, which were already available in other software packages or via web servers with impractical or time-consuming usage and/or complex input setup. Other modules introduced procedures and algorithms not available elsewhere, such as the protein structure network (PSN) analysis [21], the kinetic grouping analysis (KGA) [22], and the mincut-based free energy profile (cFEP) [23]. Significant technical improvements concerned simplification of the input syntax, the procedure for atom selection, gain in speed, and implementation of multi-threading for a subset of modules.

* Corresponding authors.
** Corresponding author at: Department of Life Sciences, University of Modena and Reggio Emilia, via Campi 103, 41125 Modena, Italy.
E-mail addresses: angelo.felline@unimore.it (A. Felline), michele.seeber@unimore.it (M. Seeber), fanelli@unimo.it (F. Fanelli).

https://doi.org/10.1016/j.csbj.2023.01.026

Table 1
New features in Wordom.

Module	Label	Function	Ref.
Distance correlation coefficient of atomic fluctuations (DiCC) and Atomic Movement Similarity (AMS)	CORR	Compute correlations of atomic fluctuations	[24,25]
Determination of geometrically stable domains	CORRCLUST	Finds dynamic protein domains that move as coherent units	[25]
Invariant core	INVCORE	Analyzes the variability of a conformational ensemble	[26]
Decomposition of atomic fluctuations	FLUDEC	Detects changes in domain-domain separation and mutual orientation	[27]
Wavelet analysis	WAVELET	Infers the multi-scale nature of protein motions	[28,29]
Tools for PCA	PCATOOLS	Processes PCA outputs	
Functional Mode Analysis	FMA	Finds collective motions related to function	[30]
Force constants	FORCE	Detects structurally stable residues	[31,32]
Overall fluctuations	FLUCT	Detects the intrinsic flexibility of the whole protein or of selected regions	[33]
Perturbation Response Scanning	PRS	Infers the role of residues in protein dynamics under external perturbations	[34–37]
Pore radius of ion channels	HOLE	Computes the radius of channel pore	[38,39]
Water or ion flux through a channel	FLUX	Monitors the number of ions or water molecules moving through a channel pore	[40,41]
Twist angle	TWIST	Computes the twist angle of an ion channel	[42]
Tilt angle	TILT	Computes the tilt angles of an ion channel	[42]
Angle between atom selections	SELEANGLE	Computes the angle between main axes	
Perimeter and area of a polygonal selection	POLYGON	Computes the perimeter and area of polygons whose vertices are the atom or centroid coordinates	
Volume	VOLUME	Computes the total, occupied, and free volume of a box	
β-sheet curvature	CURVATURE	Computes the curvature of a β-sheet	[43]
Radial distribution function	RDF	Computes atom density as a function of the distance from a selected point in space	[44]
Center of mass	COM	Computes the position of the center of mass of the selected atoms	[45]
Ring perception	RING	Allows identifying the number of cyclic structures from a contact map.	[46]
Geometric transformations	AXROTALIGN	Performs a number of coordinate transformations	

In this article, we illustrate the novelties added in this second update. We expanded the existing CORR module and added 21 new modules, divided in those suitable for the analyses of atomic fluctuations and structural communication in biomolecular systems, those specific for the study of ion channels, and those for computation of geometrical indices of structural deformation (see Table 1).

2. New implementations in Wordom

2.1. Biomolecule flexibility and structural communication

A relevant fraction of the new Wordom modules concerns the analysis of flexibility and structural communication in biomolecules. The new adds significantly extend the CORR module to infer correlated motions, identify regions that move as dynamic units, identify the dynamically stable core for proper superimposition and comparison of trajectory frames, distinguish intra-domain from inter-domain motions, infer the multi-scale nature of macromolecular motions by distinguishing low-frequency large-scale collective motions from high-frequency short-scale fluctuations, compare collective motions, infer collective motions related to function, identify mechanically stable residues that may be central in the allosteric communication and resist to mutation, and infer dynamically coupled regions and structural communication pathways.

2.1.1. CORR - correlations of atomic fluctuations

The CORR module computes the correlations of atomic fluctuations in a molecular dynamics (MD) trajectory. Two methods were already implemented in Wordom: dynamic cross-correlation (DCC) [47] and linear mutual information (LMI) [48]. The second update features two new methods: the distance correlation coefficient (DiCC) [24] and the atomic movement similarity (AMS) [25].

The DiCC method is able to detect long-range concerted motions in a protein and is a well-known and established measure of dependence between two random vectors [24]. DiCC is defined as the ratio between the distance covariance of two-atom positions along an MD trajectory and the product of their distance standard deviations as summarized in the following equation:

$$DiCC_{ij} = \frac{\frac{\sum_{k,l=1}^{N} I_{kl} J_{kl}}{N^2}}{\sqrt{\frac{\sum_{k,l=1}^{N} I_{kl}}{N^2} \frac{\sum_{k,l=1}^{N} J_{kl}}{N^2}}} \quad (1)$$

where $DiCC_{ij}$ is the distance correlation coefficient between atom i and j, respectively. N is the total number of trajectory frames, and I_{kl} and J_{kl} are defined as:

$$I_{kl} = i_{kl} - \langle i_{kx} \rangle - \langle i_{xl} \rangle + \langle i_{xx} \rangle$$
$$J_{kl} = j_{kl} - \langle j_{kx} \rangle - \langle j_{xl} \rangle + \langle j_{xx} \rangle \quad (2)$$

where $< i_{kx} >$ is the k^{th} row mean, $< i_{xl} >$ is the l^{th} column mean, and $< i_{xx} >$ is the grand mean of the distance matrix I_{kl}. In the same way, $< j_{kx} >$ is the k^{th} row mean, $< j_{xl} >$ is the l^{th} column mean, and $< j_{xx} >$ is the grand mean of the distance matrix J_{kl}.

DiCC values can vary from 0.0 to 1.0, which indicate completely uncorrelated and completely correlated atomic displacements, respectively.

The AMS algorithm detects concerted motions by focusing on each atom of a molecule and representing each atomic trace as a single trajectory. The AMS matrix is a matrix of similarity coefficients resulting from the comparison of each pair of atomic trajectories. The similarity of motion of atoms a and b is estimated by finding an isometry (translation plus rotation) of the trajectory of a that maximizes the correlation with the trajectory of b, i.e., an isometry that would bring the two trajectories as close as possible [25]. The translational part of such isometry is found by matching the initial coordinates of both trajectories, i.e., by matching the initial coordinates of both position vectors. The rotational part of isometry is found by representing rotations in quaternion space. Finding the rotation that would best superimpose the two objects is reduced to solving an eigenproblem of the following form:

$$M \cdot q = \lambda q \tag{3}$$

where M is a matrix defined as:

$$M = \sum_{\alpha=1}^{N} \begin{bmatrix} (x_\alpha - x'_\alpha)^2 & u_\alpha^T \\ u_\alpha & P_\alpha \end{bmatrix} \tag{4a}$$

$$u_\alpha = x_\alpha \times x'_\alpha \tag{4b}$$

$$P_\alpha = x_\alpha \cdot x'^T_\alpha + x_\alpha{}^T \cdot x'_\alpha \tag{4c}$$

where x_α and x'_α are the positions of the two atoms in a conformation α, and q and λ are the quaternion and the diagonal matrix, respectively. As a result of solving Eq. 3, four eigenvalues (λ_i) are obtained: the smallest one indicates the best superposition, while the largest one indicates the worst superposition. AMS values can vary from 0.0 to 1.0, which indicate completely uncorrelated and completely correlated atomic motions, respectively.

2.1.2. CORRCLUST - determination of geometrically stable domains

The CORRCLUST module of Wordom implements the geometrically stable substructures (GeoStaS) method, which can be used to find dynamic protein domains that move as coherent units in a simulation [25]. The identification of geometrically stable domains is founded on the clustering of atoms based on the similarity of their fluctuations, i.e., by using a matrix of correlation coefficients. The original method employed an AMS-based similarity matrix [25]. Wordom allows the choice of one of the four different algorithms implemented in the CORR module (i.e., DCC, LMI, DiCC, and AMS) for calculating the correlations of atomic fluctuations on conformational ensembles from molecular simulations or the DCC algorithm for computing correlations of motions on the pseudo-trajectories inferred from the Elastic Network Model-Normal Mode Analysis (ENM-NMA) on a single structure.

CORRCLUST groups the atoms by an agglomerative hierarchical merging (HM). Wordom allows the choice of four different linkage criteria for HM: average, complete, single, and ward. In each step, only the most similar clusters are merged. Finally, Wordom provides an automated method to stop the iterative merging on the basis of the Silhouette Index [51].

The GeoStaS method was used to elucidate the dynamics of a number of different multidomain proteins, including the bacterial elongation factor EF-Tu, the GroEL chaperone monomer bound to ATP, the capsid protein of Rous sarcoma virus, the HCV IRES domain IIa, and the periplasmic chaperone FIMC [25].

The complete HM clustering of atoms based on the correlated motions computed by the AMS algorithm was employed to predict geometrically stable domains from the MD trajectory of the wild type (WT) G protein transducin (Gt) (Fig. 1 A,B). Gt is a member of the α-family of heterotrimeric Ras GTPases, molecular switches that cycle between inactive, GDP-bound, and active, GTP-bound, states [52,53]. Gα proteins hold a Ras-like or GTPase domain (GD), characterized by a three-layer αβα -sandwich architecture and a helical domain (HD), characterized by a six-helix (from αA to αF) orthogonal bundle architecture (Fig. 1 A). Activation of Gα proteins by the cognate guanine nucleotide exchange factors (GEFs), which catalyze the exchange of bound GDP for GTP, requires detachment of the HD from the GD [50]. Remarkably, the estimated geometrically stable domains essentially coincide with the two structural domains of the protein (Fig. 1 A,B).

2.1.3. INVCORE - invariant core

The invariant core method is powerful in analyzing the variability of a conformational ensemble by distinguishing the most invariant region of a protein structure from the variant one(s). The invariant core is found using an iterative multi-step procedure [26] that works summarily as follows: 1) in the beginning, all atoms are considered to be members of the putative invariant core; 2) all atoms in the putative invariant core are used to superimpose all trajectory frames on the average structure obtained in the previous cycle (see point 4) or on the first frame in the first iteration; 3) the atom with the highest mobility is removed from the pool of atoms of the putative invariant core; 4) an average structure is computed after superimposing all trajectory frames using all but the most mobile atom identified in step 3; and 5) back to step 2. In all iterations, the mobility of each atom (step 3) is expressed in terms of the volume of an error-ellipsoid whose orientation and axis-lengths are computed from the eigenvalues obtained by diagonalizing the variance-covariance matrix of its coordinates:

$$C_i = R_i S_i R_i^{-1} \tag{5}$$

where R_i is the matrix that defines the ellipsoid-principal axis orientations and S_i is a diagonal matrix of eigenvalues. The volume of the error-ellipsoid is then computed by the following equation:

$$V_i = \frac{4}{3}\pi\sqrt{\lambda_x \lambda_y \lambda_z} \tag{6}$$

where V_i is the volume of the error-ellipsoid of atom i, and λ_x, λ_y, and λ_z are the eigenvalues.

The signal/noise ratio increases when geometrical analyses (e.g., distances, angles, dihedrals, etc.) are carried out using the most invariant region(s).

This method has been fruitfully employed for comparing protein structures undergoing large motions in sub-portions/domains by superimposing only the most invariant region rather than all atoms [13,26,54–57].

2.1.4. FLUDEC - decomposition of atomic fluctuations

When aligning the conformational ensemble of a multi-domain protein, a simple roto-translational alignment over the full protein to minimize the difference between a reference structure and trajectory frames does not separate motions within a domain from motions between domains. The FLUDEC module separates rigid-body domain motions from local fluctuations and can be useful to assess changes in domain-domain separation and mutual orientation [27].

Atomic displacements as a combination of domain and local fluctuations can be modeled according to the following equation:

$$U_t^{-1}\delta_t^i = U_t^{-1}(R_t - T_t) - [S_t^1 + W_t^1 R_0^1 \ldots S_t^d + W_t^d R_0^d] \tag{7}$$

where $U_t{}^{-1}$ and T_t are the inverse of the rotation matrix with respect to the reference structure in the t^{th} conformation and the overall translation vector, respectively. R_t and $R_0{}^d$ are the position vectors in the t^{th} conformation and the position vector of the atoms of the d^{th} domain in the reference structure, respectively. $S_t{}^d$ and $W_t{}^d$ are the translation and rotation of the d^{th} domain within the t^{th} conformation, respectively. The Wordom implementation of this method allows the user to divide the simulated protein in a number of domains using simple atom selections. Extracted domain and local fluctuations can be saved on new trajectory files for further analyses, and it is also possible to create a new trajectory file with only the fluctuations of the geometric centers of the defined domains.

This method was probed on the effects induced by the phosphorylation of Syk tyrosine kinase on the relative motion and rotation of the two SH2 domains [27]. Validation showed that a more accurate description of the fluctuations of a multi-domain protein was obtained by including two rigid-body alignments with more meaningful correlation coefficients compared to the overestimated correlated motion arising from a typical single-alignment procedure [27].

2.1.5. WAVELET - wavelet analysis

Wavelet analysis (WA) is a powerful signal processing technique widely used in several fields, including physics, chemistry, and

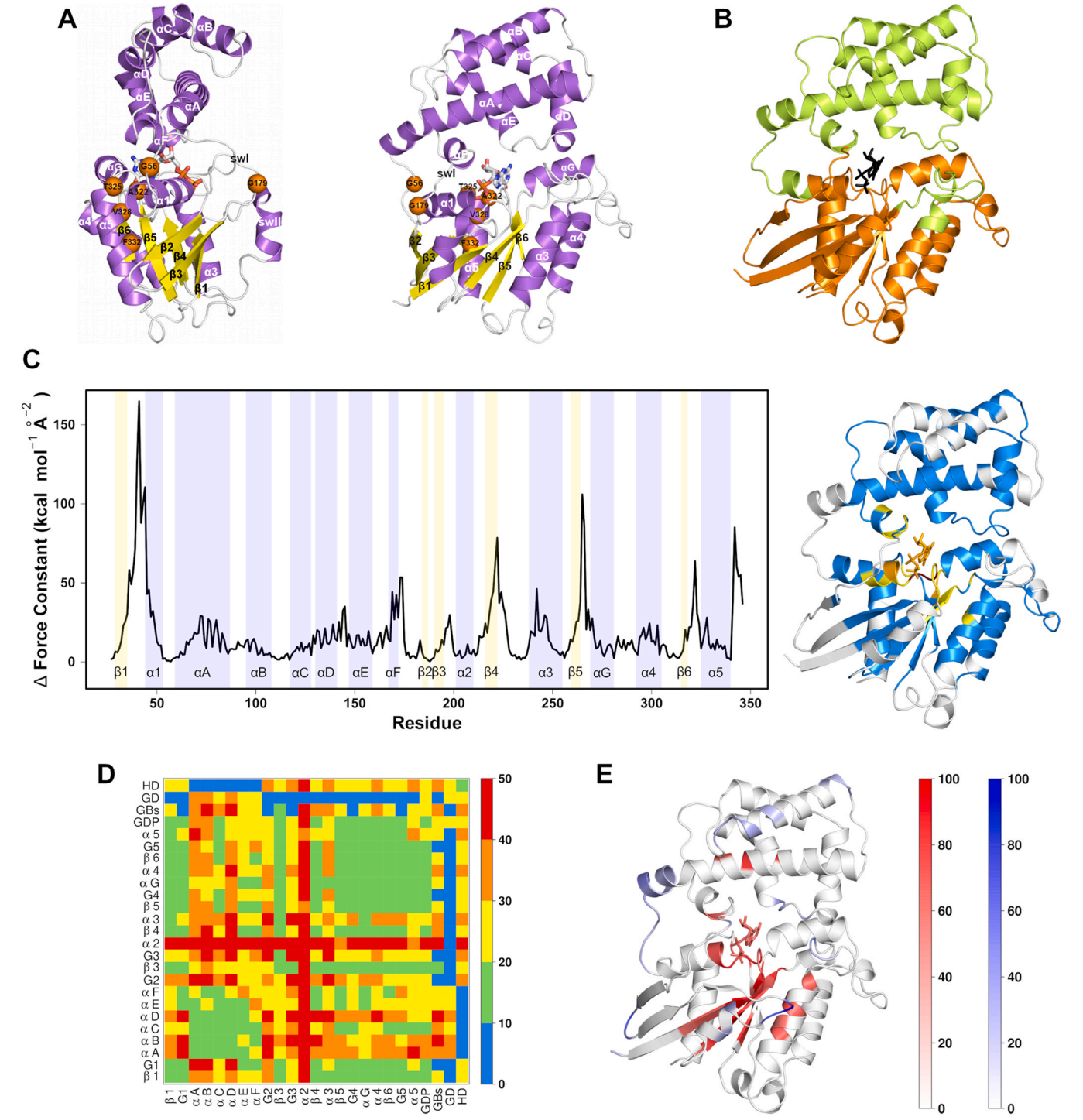

Fig. 1. Applications to the G protein transducin. A. Two views of the crystal structure of the Gα protein transducin (Gt) (PDB: 1TAG) are shown. Canonical α-helices and β-strands are violet and yellow, respectively. The secondary structure elements are labeled according to the Noel's nomenclature [49]. Gt holds a Ras-like or GTPase domain (GD), characterized by a three-layer αβα-sandwich architecture and an helical domain (HD), characterized by a six-helix (from αA to αF) orthogonal bundle architecture. Orange spheres indicate the location of the six constitutively active mutants (CAMs) G56P, G179P, A322S, T325A, V328A, and F332A subjected, together with the wild type (WT) form, to 500 ns MD simulations [50]. B. The dynamic domains inferred from the MD trajectories of WT Gt are colored in orange and green and are mapped on the structure closest to the average. C. On the left, the differences between the residue-force constant of WT Gt and the residue-force constants averaged over the force constants of the six CAMs (i.e., Δ Force) are plotted. Background is colored according to the corresponding secondary structures. On the right, Δ forces are mapped on the WT structure closest to the average and represented in cartoons. White, cyan, yellow, orange, and red correspond, respectively, to the following Δ-force values: $\Delta = 0$, $0 < \Delta \leq 25$, $25 < \Delta \leq 50$, $50 < \Delta \leq 75$, and $75 < \Delta \leq 100$. D. Comparisons between the overall fluctuations of T325A CAM and WT are plotted. They are expressed as % according to the following general formula: RelFuct%= ((CAM/WT)* 100)− 100, where CAM and WT indicate the Θ index of sets *A* and *B* in the CAM and WT forms, respectively. Blue, green, yellow, orange, and red indicate, respectively, RelFuct % values ≥ 0%, ≥ 10%, ≥ 20%, ≥ 30%, and ≥ 40%. G1, G2, G3, G4, and G5 indicate the single G-boxes, whereas the label GBs indicates all G-boxes. G2 and G3 are, respectively, called switch I and switch II (swI and swII); swII includes the α2-helix. E. The output of PRS analysis on the MD trajectory of the T325A CAM is shown, mapped on the T325A structure closest to the average. Red and blue shadows indicate, respectively, residues variably able to transmit signals and residues variably able to detect signals. In all panels, the nucleotide GDP is represented in sticks.

biology [28]. When applied to MD simulations, wavelet analysis is able to identify, time point, time length, and spatial location of statistically significant conformational changes with respect to a reference structure, thus quickly locating time-space regions where nonrandom motions occur. One interesting feature of WA is that it allows for straightforward comparisons of sets of simulations of one protein or simulations of protein variants [58]. WA consists in decomposing atom displacements into groups of orthogonal functions and associated coefficients, the latter describing the signal in the time-scale domain. Coefficients at fine scales reveal detailed features of the signal, while coefficients at coarse scales indicate global features. There are two types of wavelet transforms, discrete (DWT) and continuous (CWT), the latter being implemented in Wordom [59]. CWT, or simply WA, was employed in combination with statistically significant tests to analyze MD simulation databases (the Dynameomics database [29,58,60–63]), thus highlighting the multiscale nature of protein motion. WA was indeed able to distinguish long scale, low frequency, movements associated to large structural rearrangements and changes in secondary structure from short scale, high frequency movements characteristic of thermal vibrations. The continuous wavelet transformation of the displacements of an atom in an MD simulation is given by:

$$WT_{a,s} = IFT\,[D_a W_s] \tag{8}$$

where a is an atom, s is the wavelet scale, D_a is the frequency domain of the displacement time series of atom a, W_s is the frequency domain of the wavelet function at scale s and *IFT* denotes the inverse Fourier transform.

Application of WA on long time-scale (tens of μs) simulations of a DNA helix in combination with a simple clustering method proved able to distinguish and quantify the locations and time scales of significant motions [59]. Moreover, by changing the maximum time scale of WA, a more complete characterization of these motions could be obtained, allowing motions of different time scales to be identified or ignored as desired [59].

2.1.6. PCATOOLS - tools for PCA

This Wordom update adds a module, PCATOOLS, for post-processing the output of the principal component analysis (PCA) and ENM-NMA modules. This module allows to compare the conformational subspaces by a number of similarity metrics, i.e., the dot product of pairs of single or combined eigenvectors, the cumulative square overlap between two sets of eigenvectors, and the fraction of variance described by a set of eigenvectors. The module allows also calculation of the mode involvement coefficient, i.e. the overlap between an eigenvector and a deformation or displacement vector, which is useful to investigate conformational transitions.

PCATOOLS was used in a number of investigations that identified function-related similarity/differences in the collective motions inferred from MD trajectories [50,64,65] or from sets of crystallographic structures [66,67] of several biomolecular systems.

2.1.7. FMA - functional mode analysis

Functional mode analysis (FMA) is a powerful method that aims at elucidating collective motions directly related to a specific protein function. The analysis consists in finding possible correlations between the essential dynamics inferred from an MD trajectory or a set of crystallographic/NMR structures and a "functional quantity" computed on the same trajectory or set of structures [30]. The "functional quantity" can be quite general and can correspond to some structural descriptor or any variable relevant to the function of the protein. FMA quantifies the contributions of individual principal components to changes of the "functional quantity" and, by using the Pearson coefficient, finds the linear combination of principal components maximally correlated with the "functional quantity". The maximally correlated motion as a function of time is given by the projection along the normalized collective vector v of protein atoms:

$$p_{vt} = [x_t - \bar{x}]\cdot v \tag{9}$$

where x_t and $\bar{x}$ are the 3 N coordinates of the protein atoms at time t and their average calculated over all trajectory frames, respectively. The correlation between the normalized vector described above and the "functional quantity" is assessed as:

$$R = \frac{\mathrm{cov}(f_t, p_{vt})}{\sigma_f \sigma_v} \tag{10}$$

where f_t is the time series of the "functional quantity" (if computed on an MD trajectory) and σ_f and σ_v are the standard deviations of f_t and p_{vt}, respectively. The collective vector v can be derived by maximizing the Pearson coefficient R as a linear combination of the first d principal components:

$$v = \sum_{i=1}^{d} v_i e_i \tag{11}$$

where v_d are the coordinates of v with respect to the basis set $\{e_d\}$. Finally, the maximum of the absolute value of R can be found by solving:

$$\sum_{i=1}^{d} v_i \mathrm{cov}(p_i, p_j) = \mathrm{cov}(f, p_j), j = 1. ..d \tag{12}$$

The FMA modules of Gromacs or Wordom have been used to address a variety of mechanisms of proteins function, e.g., potassium channel gating [68], aquaporine water permeability and regulation [69], transthyretin amyloidogenesis [70], microtubule assembly [71], formation of the reactive state of the Holliday junction (HJ) resolving enzyme RecU [72], cadherin adhesion [73], mechanical signal transmission in a flagellar motor switch complex [74], protein-protein binding by allosteric switch [75], allosteric transition in hemoglobin [76], guanine nucleotide-exchange factor (GEF) binding and GEF-catalyzed nucleotide depletion by Ras GTPases [50,77], structural features of transducin mutations linked to disease [78], arrestin activation by the G protein coupled receptor (GPCR) vasopressin V2 receptor (V2R) [79].

2.1.8. FORCE - force constants

Computation of the force constant also named distance fluctuation stability (DFS) index allows inferring the mechanical properties of a protein structure, at single residue resolution [31,32]. The DFS index measures the energy cost of residue deformation during simulations. High values of DFS indexes are associated with stable residues that display small fluctuations in their distances to all other residues, while low values of this stability parameter point to more flexible sites that experience large deviations of their inter-residue distances. The DFS index k_i for each residue i is calculated from the average of the distances between the residues over the simulation trajectory using the following equation:

$$k_i = \frac{3k_B T}{\langle (d_i - \langle d_i \rangle)^2 \rangle} \tag{13}$$

where d_i is the average distance from residue i to all other residues j in the analyzed structure, angle-brackets denote the average over the simulation, k_B is the Boltzmann constant, and T is the simulation temperature [31,32]. Note that the position of each residue i can be characterized by a single atom, such as Cα; d_i can alternatively be obtained by averaging over the mean distances for each atom within each residue [32].

High values of the DFS index often indicate structurally stable centers and regulators of allosteric signals, whereas low values of the DFS index typically indicate highly fluctuating sites. In this framework, the index has been exploited to infer those residues that are

central in the allosteric communication in Hsp70 chaperones [80–82] and as a metric to assess both structural stabilities and allosteric propensities of the SARS-CoV-2 Spike protein in complex with ACE2 decoys, miniprotein inhibitors, and different classes of antibodies [83–85].

This method has been exploited to characterize the mechanical properties and detect functional sites in a number of proteins, including enzymes [31,64,65,86,87], globins [88,89], Rho guanine nucleotide exchange factors (RhoGEFs) of the Dbl family [64], V2R in complex with arrestin 2 [79], the native and mutated signal transducer Gal3 protein [90], with important mechanistic implications and direct relations with amino acid sequence conservation.

An exemplary application of the force constant analysis is the comparison of the MD trajectories of WT and six constitutively active mutants (CAMs) of Gt, which are characterized by increased intrinsic (i.e., GEF-independent) rate of nucleotide exchange [50]. In line with that, in CAMs those regions deputed to nucleotide depletion undergo reduction in structural rigidity compared to the WT (Fig. 1 C). The nucleotide GDP undergoes such reduction as well.

2.1.9. FLUCT - overall fluctuations

The overall fluctuation index Θ is a measure of the intrinsic flexibility of a whole protein or of a given sub-set of its residues [33]. The index is proportional to the extent of the conformational space explored in a simulation and can be used to compare the intrinsic flexibility of functionally important regions of the same protein in two different states or of two homologous proteins in the same state. Calculation of this index is based solely on internal distances and therefore does not require superposition of the trajectory frames to eliminate the roto-translations of the simulated system. Θ is defined as the root mean distance variance of each atom pair and is calculated by the following equation:

$$\Theta_{AB} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{\sum_{k=1}^{F}(d_{ij}^{k}-\bar{d}_{ij})^2}{F}}{N\times M}} \tag{14}$$

where A and B are two sets of residues, N and M are the total number of atoms in set A and set B, respectively, F is the total number of trajectory frames, d_{ij}^{k} is the distance between atom i from residue set A and atom j from residue set B in the k^{th} frame, and $\bar{d}_{ij}$ is the average distance between the same two atoms. If sets A and B coincide, intra-domain overall fluctuations can be computed.

Computations of the overall fluctuation index served to investigate structure-function relationships in a number of biosystems including six CAMs of Gt [50], two RhoGEFs in their free and RhoA-bound states [64], V2R in complex with arrestin 2 [79], and the Sec4 Rab GTPase [65]. Collectively, the analyses highlighted inter-molecular and inter-domain dynamic coupling related to function [64,65,79]. Comparisons between Θ of WT with those of six CAMs of Gt showed that, on average, the Gt portions undergoing increases of Θ in response to the six activating mutations include the five G boxes (i.e., G1-G5, highly conserved regions involved in nucleotide binding) taken singularly or as a whole, the secondary structures of GD and HD, and the nucleotide [50]. Alanine substitution for T325 increased significantly (i.e., by 32%) the overall fluctuations of the whole HD and GD compared to the WT (Fig. 1D). This is consistent with the finding that activating mutations, in particular T325A, prioritize those collective motions of the HD that overlap with the collective motions associated with GEF-catalyzed nucleotide release [50]. Remarkably, Gt portions like G2, G3, and α2, which are involved in conformational changes linked to function, undergo increases of Θ in tandem with almost all portions of the G protein. This may relate as well to the constitutive activity of T325A.

2.1.10. PRS - perturbation response scanning

Perturbation response scanning (PRS) [34] is a powerful technique to investigate the structural communication by assessing and quantifying the role of each protein residue in determining conformational changes as a consequence of external perturbations. Over time, the method has been subjected to several developments and expansions [35–37], all available in Wordom. Given two conformations of the same protein, an initial and a target structure, PRS is able to predict protein residues that, under the influence of an external perturbation, contribute to a conformational change in the direction of the target structure. Additionally, the method identifies and quantitatively assesses the *effectiveness* and *sensitivity* of each protein residue, i.e., the ability of a given residue to transmit signals when subjected to an external perturbation and the ability to detect signals transmitted by effectors, respectively. PRS, which is based on the linear response theory [91], can be used to analyze both single structures and conformational ensembles from MD simulations. PRS relies on systematically applying virtual random forces on each protein residue and detecting the linear response of the whole protein to these perturbations. According to the linear response theory, the response of a residue when a set of perturbations is applied to another residue is given by the following equation:

$$R_{ij} = \frac{\sum_{k=1}^{|F|} C_{ij}F_k}{|F|} \tag{15}$$

where i is the residue whose response is being tested, j is the perturbed residue, $|F|$ is the total number of applied perturbations, C is the variance-covariance matrix, and F_k is the k^{th} vector with the components of the externally applied force vectors.

PRS has been applied to the investigation of the allosteric communication in a wide variety of systems. Here we quote a selection of those studies [34–36,92–111].

The PRS analysis of the trajectories of WT and six Gt CAMs detected a crowding of signal transmitter residues in the nucleotide-binding site (Fig. 1E). A few signal detector residues fall on the linker regions participating in the inter-domain motion that leads to the activation of the G protein (Fig. 1E).

2.2. Ion channel dynamics

An important addition in Wordom is the implementation of a number of structural indices to investigate the function of oligomeric transmembrane (TM) proteins like ion channels. Although the new modules may work on any TM ion channel of class α (i.e., α-helical oligomers), here, pentameric ligand-gated ion channels (pLGIC) are presented as a case study.

pLGICs are oligomeric TM proteins that open an ion pore in response to increased levels of neurotransmitters at chemical synapses. Structural, functional, and modeling data indicate that these receptors transduce signals by cycling through distinct conformational states corresponding to the resting (R), active (A), and one or more desensitized (D) states [112]. Recent high-resolution structures in combination with all-atom MD opened to the structural characterization of the functional motions with atomic resolution [113]. In particular, it has been shown that receptor activation or *gating* is mediated by two major quaternary transitions involving a radial contraction or (un)-blooming of the extracellular domain to increase the affinity for the neurotransmitter and a global (un)-twisting to initiate pore opening [42,114]. These molecular motions provide meaningful reaction coordinates to monitor receptor activation [115].

Analysis tools like HOLE [39] to probe the physical dimensions of the ion pore have become popular to distinguish between open- and closed-channel structures. However, when more than one model can be annotated as open, other approaches based on all-atom MD

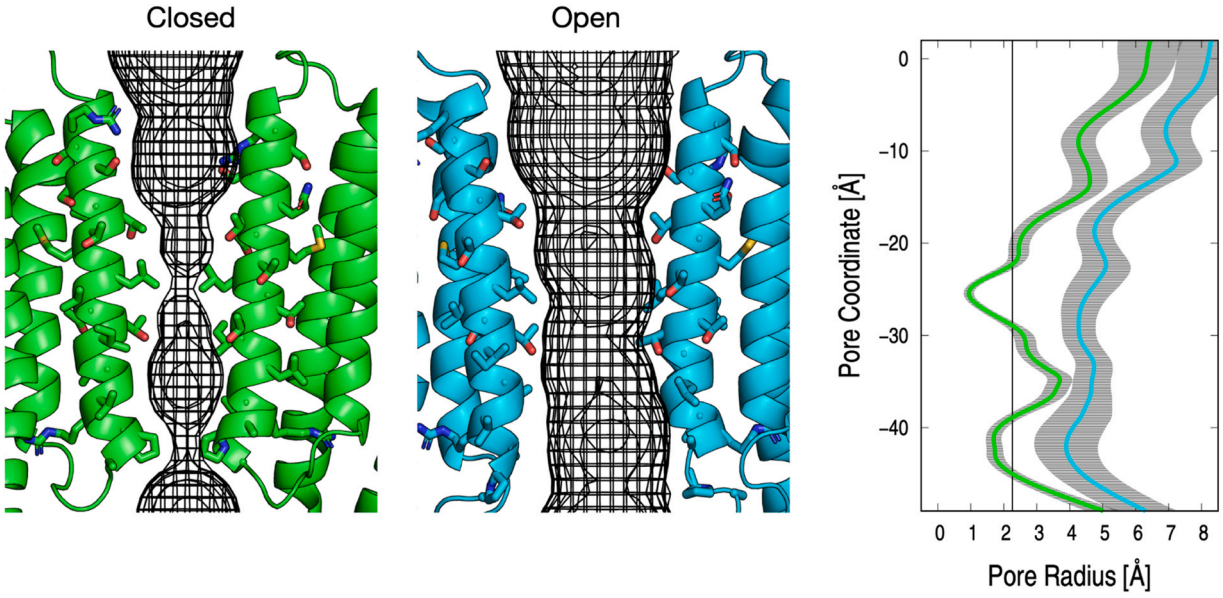

Fig. 2. Exemplar application of the HOLE module. Pore radius as a function of the pore coordinate for two recently solved structures of the glycine receptor (GlyR) α1 in complex with the partial agonist taurine. These structures, referred to as tau-closed (PDB: 6PM3) and tau-open (PDB: 6PM2), were simulated at room temperature for 100 ns by all-atom Molecular Dynamics with an explicit representation of the solvent and the membrane environment. The configuration of the pore was then analyzed using the HOLE module in Wordom. Assuming a radius of 2.26 Å for solvated chloride ions (black thin line), the simulations indicate that tau-closed (green) is a closed-channel state with two physical constrictions, one at L277 at position 9′ and one at P266 at position 2′. By contrast, the same analysis of tau-open (cyan) indicates that this structure is representative of an open channel with a minimum pore radius well above the Born radius of solvated chloride. In addition, analysis of the fluctuations of the pore-lining residues sampled by MD, here represented as horizontal error bars along the HOLE profile, show a much larger flexibility of the ion pore in the open-channel versus the closed-channel structure. The pore coordinate is chosen arbitrarily.

simulations need to be invoked. Recently, using an original combination of *in-silico* electrophysiology and polyatomic anion permeation simulations, we have provided evidence that none of the early cryo-EM structures of the glycine receptor is physiologically relevant and identified an alternative open-channel state by MD simulations, which was ion-conductive and anion-selective in semi-quantitative agreement with experiments [40]. Analysis tools to display the morphology of the ion pore and quantify both ion and water fluxes through it are critically important for a structure-to-function annotation as well as the exploration of the ion-permeation mechanism (s) [116].

To analyze the properties of ion channels in all-atom MD simulations, four new modules have been added in Wordom: 1) HOLE, to calculate the radius of an ion pore along the vector describing its direction; 2) FLUX, to compute the number of ions or water molecules translocating across the membrane during the simulation; 3) TWIST, and 4) TILT, to measure the global twisting and blooming of the receptor.

2.2.1. HOLE - Pore radius of ion channels

The HOLE module computes the radius of a channel pore along its principal direction, commonly referred to as a HOLE profile. It uses the definition of pore radius and the algorithm proposed by Smart and coworkers [38] implemented in the HOLE software [39]. Taking as an input one point inside the pore and a vector describing its approximate direction, the algorithm explores the plane perpendicular to the pore-axis by Monte Carlo moves, searching for the center of the largest sphere non-overlapping with the protein. Then, the point is moved along the pore axis and the search is repeated. By plotting the pore radius as a function of the pore coordinate, one can identify physical constrictions where the pore size decreases and evaluate if the channel is open or not (Fig. 2).

2.2.2. FLUX - water or ion flux through a channel

Albeit the pore of an ion channel can be physically open, it is not necessarily water- or ion-permeable [117]. The FLUX module counts the number of ions or water molecules that actually move through the pore during an MD simulation. For this purpose, selected ions or water molecules are labeled as being near the top, inside, or close to the bottom of the channel, or free in the bulk (away from the channel). Their position is monitored during the simulation and transition events, i.e., changes in labeling are counted (the direction is not distinguished, all events are counted together). The number of transitions per simulation snapshot is reported in output, as well as the complete list of transition events. Options are available to select the ion-pore region and filter out external transitions, e.g., water molecules permeating the lipid membrane away from the ion channel. This module allows computing the permeability of channels to water and to both monoatomic (e.g., Cl^-, K^+) and polyatomic ions (e.g., acetate or phosphate) [40,41].

2.2.3. TWIST - twist angle

The twist angle (τ) of an ion channel is defined as the clockwise rotation of the extracellular (EC) domain relative to the TM domain (Fig. 3 A). It is computed as the angle between the projections of the center of mass vectors of the EC and TM regions of one protein subunit over the plane perpendicular to the symmetry axis of the channel [42]. A selection of atoms in the system, e.g., the protein, is needed to define the rotational axis and the twist angle is computed between two user-defined selections, e.g., EC and TM regions of one subunit.

2.2.4. TILT - tilt angle

The tilt angle (θ) is defined as the angle between the axis of a user-defined selection of atoms and the principal axis of the protein. In the case of an ion channel, the user-defined selection can be usefully defined as one protein subunit to measure receptor blooming [114] or the axis of one pore-lining helix to monitor channel opening/closing [42]. Importantly, the tilt angle can be decomposed into polar (θ_p) and azimuthal (θ_a) contributions that allow separating the outward tilting of the user-defined selection from its tangential movement (Fig. 3B). These components are determined

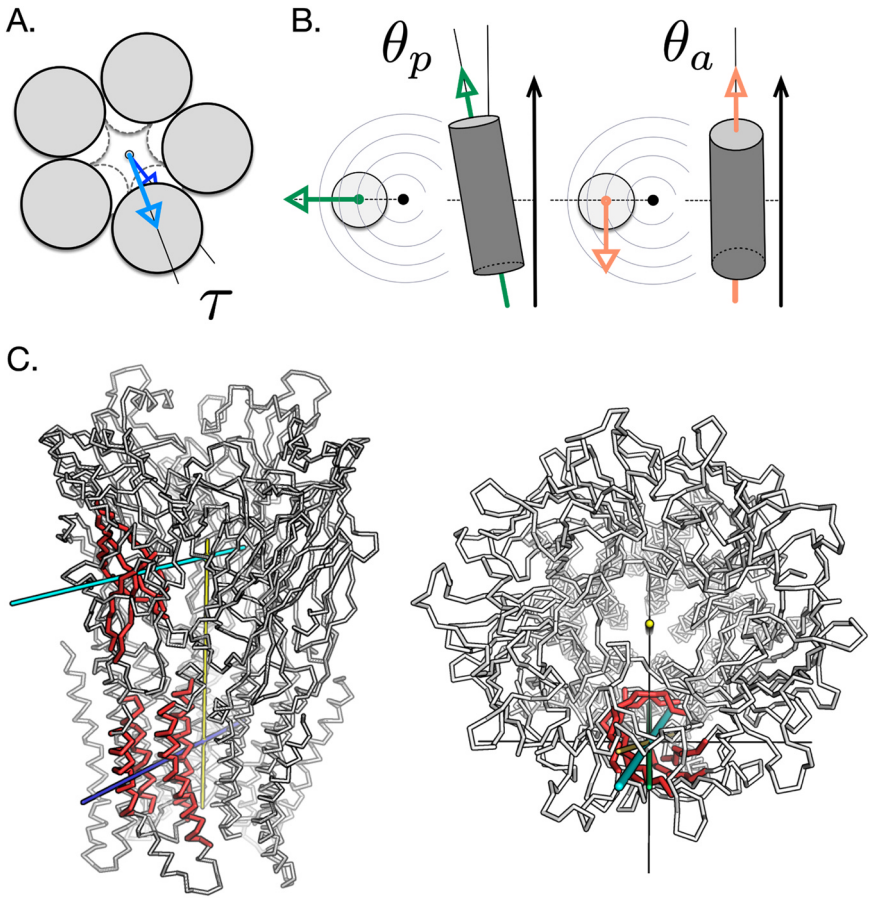

Fig. 3. The TILT-TWIST angles in ion α-helical oligomeric ion channels. Twist and tilt angles. (A) The twist angle **τ** is shown as the rotation of one domain (foreground) with respect to the other (background) about a common rotational axis perpendicular to the figure plane. (B) The polar (θ_p) and azimuthal (θ_a) components to the tilting angle of the selected domain (gray cylinder) relative to the reference axis (black arrow) are shown. (C) Visualization of the global twisting and blooming of the pentameric glutamate-gated chloride channel (GluCl) in the active state. On the left, the global twisting is shown as the twist angle (τ) between the extracellular and TM core regions (highlighted in red). The twisting arms of the extracellular and TM regions are cyan and blue, respectively, the axis of the receptor is yellow. On the right, blooming is quantified as the polar tilting (θ_p) of the extracellular β-sandwiches relative to the principal axis of the receptor. The tilting of one extracellular β-sandwich is shown as a thick cyan stick, while its polar and azimuthal components as thinner green and orange sticks. Structural representations of the twist and tilt angles were generated using the –VISUAL keyword of the TWIST and TILT modules in Wordom.

using a reference frame centered on the center of mass of the user-defined selection with the Z-axis parallel to the principal axis of the protein and the X-axis pointing outward along the radial direction. Using this convention, the polar tilt is measured as the angle between the projection of the selection axis on the XZ-plane and the Z-axis, while the azimuthal tilt is the angle between the projection of the selection axis on the XY-plane and the X-axis.

2.3. Miscellaneous indices of structural deformation

Wordom includes new modules for the computation of geometrical, size/shape indices like the angle between main axes, perimeter, area, volume, β-sheet curvature, radial distribution function and mass center to explore the conformational changes of a biomolecule or a supramolecular architecture more generally.

These indices can be used as hallmarks of protein functional states by allowing comparisons between different functional states of the same protein or between homologous proteins in the same functional state.

A ring perception module is also included, useful to monitor supramolecular self-assembly.

2.3.1. SELEANGLE - angle between atom selections

This module calculates the angle between the main axes of two atom selections. Atom selection can belong to the same structure/trajectory frame or to two distinct structures (e.g., a trajectory frame and a reference structure). The angle can be also computed between one of the three Cartesian axes and the main axis from an atom selection.

For graphical representations, the module can also create a file with the coordinates of the major axes.

2.3.2. POLYGON - perimeter and area of a polygonal selection

This module computes the perimeter and area of a polygon made by an arbitrary number of vertices. Each vertex can be independently defined by the coordinates of a single atom or by the geometrical center of a group of atoms (e.g., a whole amino acid, a single Cα-atom, the side chain of a residue, etc). Additionally, the length of each side of the defined polygon can be calculated using the three Cartesian coordinates or any combination of two of the three coordinates (i.e., XY, XZ, or YZ).

2.3.3. VOLUME - volume

This module computes the total, occupied, and free volume of a box defined by an atom selection. Wordom provides three different ways to define the box: a) as large as needed to contain all selected atoms; b) centered on the geometrical center of an atom-selection and with user-defined size; and c) the same as b) but the center of the box is directly defined by its three spatial coordinates. The box can be either updated at each frame of a trajectory or computed only once on a

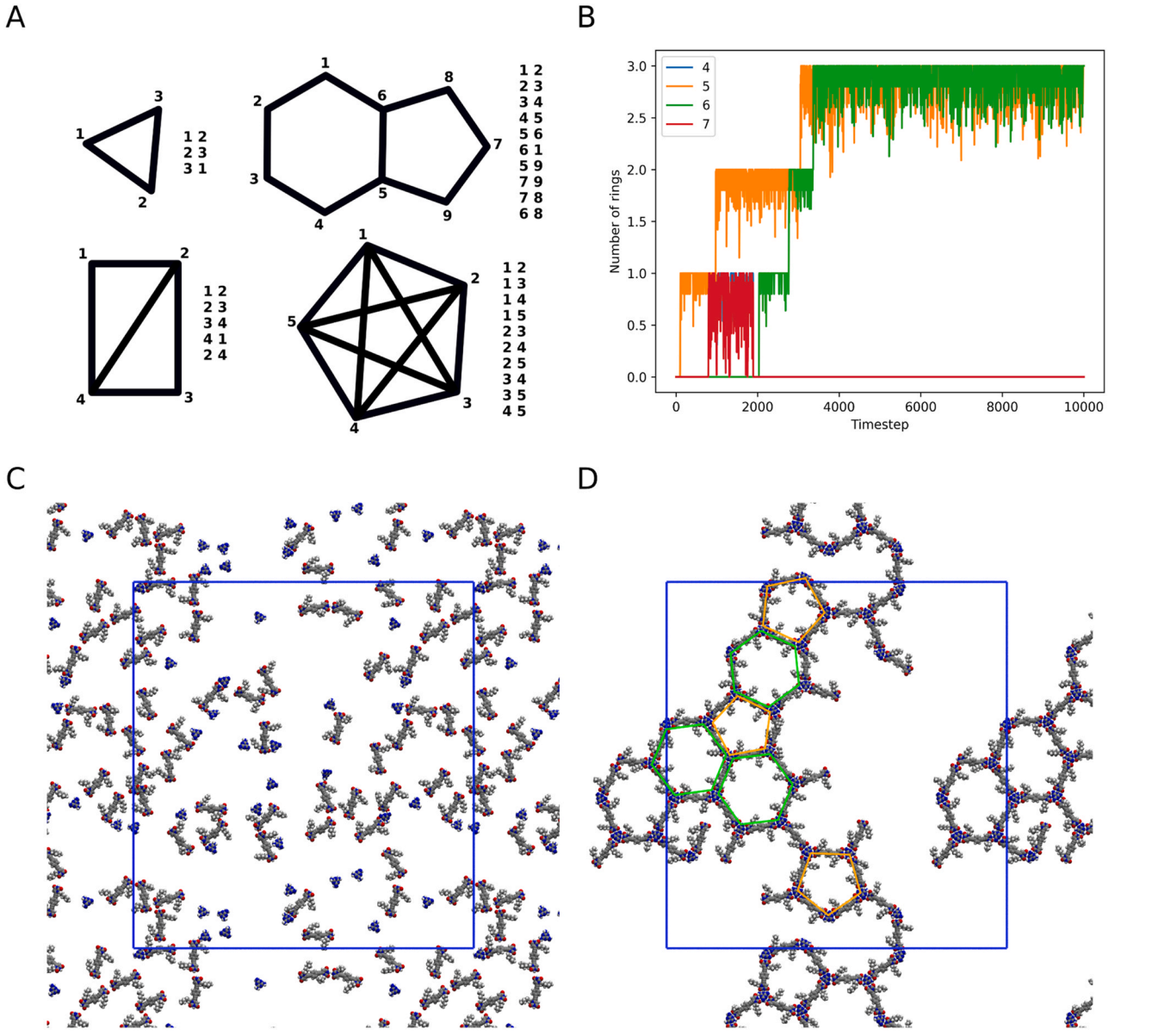

Fig. 4. RING perception. A) Four simple graphs possibly representing a supramolecular trimer, an indole molecule, or a graph construct (N = 8 complete graph, or "clique") are shown along with their contact map used as an input for RING. By performing ring perception on them, one gets: a) one 3-member ring in the first example; b) two 3-member rings and one 4-member ring in the second; c) one hexagon, one pentagon, and one nonagon in the third; and d) 37 rings in the fourth and last (10 triangles, 15 quadrilaterals, and 12 pentagons). B) Time series of the supramolecular polygons formed during MD of a system composed of 32 melamine (MEL) and 48 bis(N_1-hexyl-uracil) (URA) molecules on an implicit graphite substrate. Blue, orange, green, and red data points correspond, respectively, to the number of tetragons, pentagons, hexagons, and heptagons detected by RING. For this analysis, Wordom was used to compute contacts between MEL and URA using CONTACTS and the contact list was then passed to RING. The raw data were smoothed over a time window of 11 steps for clarity. C) Starting and D) ending snapshots of the MD simulation analyzed in B). From the "melt" or disordered state, MEL and URA molecules are shown to self-assemble into a partially ordered state. The three hexagons and three pentagons detected in (B) are highlighted in green and orange, respectively. Note a MEL molecule is missing to form a fifth hexagon. The blue square indicates the primary image in the 2D periodic system.

reference structure and kept fixed. The module properly estimates the contribution of the atoms located across the box boundaries, by considering only the fraction of their volumes inside the box.

2.3.4. *CURVATURE - βeta-sheet curvature*

The CURVATURE module computes the mean and Gaussian curvatures of a set of atoms by least-square fitting a quadratic surface over them. The quadratic surface z is defined as:

$$z = p_0 + p_1x + p_2x^2 + p_3y + p_4y^2 + p_5xy \quad (16)$$

where $p_0 \ldots p_5$ are fitted on the selected atomic coordinates using the Levenberg-Marquardt algorithm implemented in the levmar library (https://users.ics.forth.gr/~lourakis/levmar/). The local curvatures of a surface are computed from its Hessian matrix H. The mean curvature μ is defined as half of the trace of H, while the Gaussian curvature Γ is its determinant. In the case of a quadratic surface, such as in this case, the Hessian matrix is constant at any point, resulting in $\mu = p_2 + p_4$ and $\Gamma = 4p_2p_4 - p_5^2$. By default, the CURVATURE module will return the mean and Gaussian curvatures, along with the root mean square residual error between the fitted surface and the atom coordinates. Optionally, the $p_0 \ldots p_5$ fitted coefficients can be returned as well.

Measuring the curvature of a selection of atoms can be useful to determine the global twisting of β-sheets, which are often related to protein function such as the elastic deformations in motor proteins

like myosin and kinesin. One popular view is that the energy of ATP is stored as conformational free energy or strain of some structural elements, so-called hidden springs [117]. For example, the β-sheet curvature analysis was used to study the elastic properties of the 7-stranded β-sheet in the β-subunit of F1-ATP synthetase, a hexameric ring of three alternating α- and β-subunits surrounding a central coiled-coil γ-shaft [41].

2.3.5. RDF - radial distribution function

This module implements the classic radial distribution function, also known as pair correlation function, widely used in statistical mechanics [44]. This descriptor summarizes the density of atoms present in a series of evenly spaced concentric shells centered on a selected reference atom. In other words, the radial distribution function is a measure of the density as a function of distance from a selected point in space.

Example applications of this descriptor include the characterization of the solvation shells of a ligand in a binding site or the distribution of ions around a macromolecule, e.g., nucleic acids [118,119].

2.3.6. COM - center of mass

The COM module computes the position of the center of mass of selected atoms. The center of mass may be weighted either using atomic masses or arbitrary weights read from the B-factor column of the input PDB file. In the case of systems using periodic boundary conditions, the center of mass can be computed using the algorithm proposed by Bai & Breen for periodic systems [45].

2.3.7. RING - ring perception module

On the exploration of the mechanisms of molecular self-organization at surfaces and interfaces [120,121], it is interesting to investigate the structure of the supramolecular intermediates visited on the path to the 2D crystalline architecture with atomic resolution. For this purpose, in the case of 2D supramolecular networks the number of supramolecular rings as well as the number of building blocks per ring are useful parameters to be monitored along MD simulations of self-assembly [122]. The RING module in Wordom implements the ring perception algorithm by Hanser and coworkers [46], which allows to identify, enumerate, and characterize the number of cyclic structures (i.e., supramolecular hexagons, pentagons, or other polygons) from a contact map. The RING module takes a contact map as an input (i.e., a set of lines describing contacts between pairs of atoms or molecules) and yields a complete enumeration of detected rings in output (Fig. 4 A).

Interesting applications of RING include the automatic detection of supramolecular polygons in scanning tunneling microscopy (STM) images or atomistic models of 2D monolayers, or the time-dependent evolution of the 2D supramolecular architecture in MD simulations (assembly/disassembly). For illustration, the Fig. 4B-C shows the time evolution of the number of supramolecular polygons (i.e., tetragons, pentagons, hexagons, and heptagons) during an MD simulation of a mixture of melamine (MEL) and bis(N_1-hexyl-uracil) (URA) on graphite started from disordered [122]. The results show the incremental formation of pentagons and hexagons with temporary appearance of metastable tetragons and heptagons.

2.3.8. AXROTALIGN - geometric transformations

This module can be used to perform several geometric transformations of the coordinates of a structure/trajectory frame and save the transformed coordinates to a new structure or trajectory file. In more details, this module allows to align the major axis of a given atom selection to one of the three Cartesian axes, to rotate the structure around one of the three Cartesian axes by any degrees/radians, and to flip (i.e., reflect) the coordinates around an axis. Finally, Wordom can apply coordinate transformation by a 3 × 3 rotation matrix provided by the user.

3. Conclusions

The second update of Wordom brings a significant enrichment in tools for investigating biomolecule flexibility and allosteric communication as well as detecting structural deformations of biosystems. The new modules integrate an already rich platform for structural dynamics analysis, holding features for: a) coordinate reading/writing; b) analyzing molecular degrees of freedom (e.g., bond lengths and angles, dihedrals), structural flexibility by computation of RMSD (root mean square deviationa), RMSF (root mean square fluctuations), DRMS (distance root means square), correlation of motions, quasiharmonic entropy, collective motions by PCA, and clustering according to structural similarity; c) investigating structural deformations by monitoring secondary structures, molecular surface, order parameters; d) investigating structural communication by protein structure network (PSN) analysis; e) inferring the thermodynamics and kinetics of complex molecular processes by analyzing the free energy surface; and f) deciphering the dynamic features encoded in a protein structure by ENM building [1]. PSN analysis is now developed in a dedicated standalone software, PSNtools, that employs Wordom to read the atomic coordinates and compute the correlations of atomic fluctuations on MD trajectories or on pseudo-trajectories from ENM-NMA to infer the structural communication pathways [123]. The combination of Wordom and PSNtools constitutes also the engine of a webserver (webPSN) for high-throughput prediction of the allosteric communication in any bio-system [124].

Like the original project, the updated version is released under the general-public license (GPL, https://www.gnu.org/licenses/gpl-3.0.html). From 03–01–2008 to 01–01–2023 the software underwent 12073 downloads and, up to January 2023, the two articles describing Wordom [1] and its first update [2] received, respectively, 286 and 242 citations according to Google Scholar. This second update places Wordom among the most suitable, complete, user-friendly, fast, and efficient software for the analysis of biomolecular simulations. In the actual context, in which the scientific community is putting ever increasing efforts in sharing big data from molecular simulations [125,126], the availability and continuous development of analysis tools like Wordom acquire high value and perspective.

CRediT authorship contribution statement

Angelo Felline: Conceptualization, Formal analysis, Funding acquisition, Methodology, Software, Validation, Writing – original draft. **Simone Conti**: Software, Writing – review & editing. **Michele Seeber:** Conceptualization, Methodology, Software, Project. administration. **Marco Cecchini:** Conceptualization, Funding acquisition, Supervision, Visualization, Writing – review & editing. **Francesca Fanelli:** Conceptualization, Funding acquisition, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by a University of Modena and Reggio Emilia grant FAR2021 Dipartimentale to AF and a PRIN2017 MIUR grant [2017R5ZE2C] to FF.

References

[1] Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A. Wordom: a program for efficient analysis of molecular dynamics simulations. Bioinformatics 2007;23(19):2625–7.
[2] Seeber M, Felline A, Raimondi F, Muff S, Friedman R, Rao F, et al. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. J Comput Chem 2011;32(6):1183–94.
[3] Brooks BR, Brooks 3rd CL, Mackerell Jr. AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30(10):1545–614.
[4] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm - a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4(2):187–217.
[5] Munz M, Biggin PC. JGromacs: a Java package for analyzing protein simulations. J Chem Inf Model 2012;52(1):255–9.
[6] Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 2013;29(7):845–54.
[7] Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. J Comput Chem 2005;26(16):1668–88.
[8] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph Model 1996;14(1):33–8.
[9] Schrödinger L., DeLano W. PyMOL. Retrieved from http://www.pymol.org/pymol. 2020.
[10] Feig M, Karanicolas J, Brooks III CL. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model 2004;22(5):377–95.
[11] Glykos NM. Software news and updates carma: a molecular dynamics analysis program. J Comput Chem 2006;27:1765–8.
[12] Meyer T, Ferrer-Costa C, Perez A, Rueda M, Bidon-Chanal A, Luque FJ, et al. Essential dynamics: a tool for efficient trajectory compression and management. J Chem Theory Comput 2006;2(2):251–8.
[13] Grant BJ, Rodrigues A, ElSawy JM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 2006;22:2695–6.
[14] Grant BJ, Skjaerven L, Yao XQ. The Bio3D packages for structural bioinformatics. Protein Sci 2021;30(1):20–30.
[15] Hinsen K. The molecular modeling toolkit: a new approach to molecular simulations. J Comput Chem 2000;21:79.
[16] Grunberg R, Nilges M, Leckner J. Biskit—a software platform for structural bioinformatics. Bioinformatics 2007;23:769–70.
[17] Bakan A, Dutta A, Mao W, Liu Y, Chennubhotla C, Lezon TR, et al. Evol and ProDy for bridging protein sequence evolution and structural dynamics. Bioinformatics 2014;30(18):2681–3.
[18] Brown DK, Penkler DL, Sheik Amamuddy O, Ross C, Atilgan AR, Atilgan C, et al. MD-TASK: a software suite for analyzing molecular dynamics trajectories. Bioinformatics 2017;33(17):2768–71.
[19] McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernandez CX, et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. Biophys J 2015;109(8):1528–32.
[20] Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 2011;32(10):2319–27.
[21] Brinda KV, Vishveshwara S. A network representation of protein structures: implications for protein stability. Biophys J 2005;89(6):4159–70.
[22] Muff S, Caflisch A. Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β-sheet miniprotein. Prot. Struct. Funct. Bioinform. 2008;70:1185–95.
[23] Krivov SV, Karplus MJ. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. J Phys Chem B 2006;110:12689–98.
[24] Szekely GJ, Rizzo ML, Bakirov NK. Measuring and testing dependence by correlation of distances, annals of statistics. Ann Stat 2007;35:2769–94.
[25] Romanowska J, Nowinski KS, Trylska J. Determining geometrically stable domains in molecular conformation sets. J Chem Theory Comput 2012;8(8):2588–99.
[26] Gerstein M, Altman RB. Average core structures and variability measures for protein families: application to the immunoglobulins. J Mol Biol 1995;251(1):161–75.
[27] Roy A, Hua DP, Post CB. Analysis of multidomain protein dynamics. J Chem Theory Comput 2016;12(1):274–80.
[28] Torrence C, Compo GP. A practical guide to wavelet analysis. Bull Am Meteor Soc 1998;79:61–78.
[29] Benson NC, Daggett V. Wavelet analysis of protein motion. Int J Wavel Multiresolut Inf Process 2012;10(4).
[30] Hub JS, de Groot BL. Detection of functional modes in protein dynamics. PLoS Comput Biol 2009;5(8):e1000480.
[31] Sacquin-Mora S, Laforet E, Lavery R. Locating the active sites of enzymes using mechanical properties. Proteins-Struct Funct Bioinforma 2007;67(2):350–9.
[32] Lavery R, Sacquin-Mora S. Protein mechanics: a route from structure to function. J Biosci 2007;32(5):891–8.
[33] Munz M, Hein J, Biggin PC. The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. PLoS Comput Biol 2012;8(11):e1002749.
[34] Atilgan C, Atilgan AR. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. PLoS Comput Biol 2009;5(10):e1000544.
[35] Gerek ZN, Ozkan SB. Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. PLoS Comput Biol 2011;7(10):e1002154.
[36] Dutta A, Krieger J, Lee JY, Garcia-Nafria J, Greger IH, Bahar I. Cooperative dynamics of intact AMPA and NMDA glutamate receptors: similarities and subfamily-specific differences. Structure 2015;23(9):1692–704.
[37] Campitelli P, Ozkan SB. Allostery and epistasis: emergent properties of anisotropic networks. Entropy (Basel) 2020(6):22.
[38] Smart OS, Goodfellow JM, Wallace BA. The pore dimensions of gramicidin A. Biophys J 1993;65(6):2455–60.
[39] Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MS. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. J Mol Graph 1996;14(6):354–60. 76.
[40] Cerdan AH, Martin NE, Cecchini M. An ion-permeable state of the glycine receptor captured by molecular dynamics. Structure 2018;26(11):1555–62. e4.
[41] Harkat M, Peverini L, Cerdan AH, Dunning K, Beudez J, Martz A, et al. On the permeation of large organic cations through the pore of ATP-gated P2X receptors. Proc Natl Acad Sci USA 2017;114(19):E3786–95.
[42] Calimet N, Simoes M, Changeux JP, Karplus M, Taly A, Cecchini M. A gating mechanism of pentameric ligand-gated ion channels. Proc Natl Acad Sci USA 2013;110(42). E3987-96.
[43] Sun S, Chandler D, Dinner AR, Oster G. Elastic energy storage in beta-sheets with application to F1-ATPase. Eur Biophys J 2004;32:676–83.
[44] Levine BG, Stone JE, Kohlmeyer A. Fast analysis of molecular dynamics trajectories with graphics processing units-radial distribution function histogramming. J Comput Phys 2011;230(9):3556–69.
[45] Bai L, Breen D. Calculating center of mass in an unbounded 2D environment. J Graph Tools 2011:13.
[46] Hanser T, Jauffret P, Kaufmann g. A new algorithm for exhaustive ring perception in a molecular graph. J Chem Inf Comput Sci 1996;36:1146–52.
[47] Van Wynsberghe AW, Cui Q. Interpreting correlated motions using normal mode analysis. Structure 2006;14(11):1647–53.
[48] Lange OF, Grubmuller H. Generalized correlation for biomolecular dynamics. Proteins 2006;62(4):1053–61.
[49] Noel JP, Hamm HE, Sigler PB. The 2.2 Å crystal structure of transducin-a complexed with GTPgS. Nature 1993;366:654–63.
[50] Felline A, Mariani S, Raimondi F, Bellucci L, Fanelli F. Structural determinants of constitutive activation of Gα proteins: transducin as a paradigm. J Chem Theory Comput 2017;13:886–99.
[51] Rousseeuw PJ. A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
[52] Vetter IR, Wittinghofer A. The guanine nucleotide-binding switch in three dimensions. Science 2001;294(5545):1299–304.
[53] Colicelli J. Human RAS superfamily proteins and related GTPases. Sci STKE 2004;2004(250):RE13.
[54] Gerstein M, Krebs W. A database of macromolecular motions. Nucleic Acids Res 1998;26(18):4280–90.
[55] Grant BJ, McCammon JA, Caves LS, Cross RA. Multivariate analysis of conserved sequence-structure relationships in kinesins: coupling of the active site and a tubulin-binding sub-domain. J Mol Biol 2007;368(5):1231–48.
[56] Li H, Yao XQ, Grant BJ. Comparative structural dynamic analysis of GTPases. PLoS Comput Biol 2018;14(11):e1006364.
[57] Lukman S, Grant BJ, Gorfe AA, Grant GH, McCammon JA. The distinct conformational dynamics of K-Ras and H-Ras A59G. PLoS Comput Biol 2010;6(9).
[58] Benson NC, Daggett V. A comparison of multiscale methods for the analysis of molecular dynamics simulations. J Phys Chem B 2012;116(29):8722–31.
[59] Heidari Z, Roe DR, Galindo-Murillo R, Ghasemi JB, Cheatham 3rd TE. Using wavelet analysis to assist in identification of significant events in molecular dynamics simulations. J Chem Inf Model 2016;56(7):1282–91.
[60] Benson NC, Daggett V. Dynameomics: large-scale assessment of native protein flexibility. Protein Sci 2008;17(12):2038–50.
[61] Kehl C, Simms AM, Toofanny RD, Daggett V. Dynameomics: a multi-dimensional analysis-optimized database for dynamic protein data. Protein Eng Des Sel 2008;21(6):379–86.
[62] Simms AM, Toofanny RD, Kehl C, Benson NC, Daggett V. Dynameomics: design of a computational lab workflow and scientific data repository for protein simulations. Protein Eng Des Sel 2008;21(6):369–77.
[63] van der Kamp MW, Schaeffer RD, Jonsson AL, Scouras AD, Simms AM, Toofanny RD, et al. Dynameomics: a comprehensive database of protein dynamics. Structure 2010;18(4):423–35.
[64] Felline A, Belmonte L, Raimondi F, Bellucci L, Fanelli F. Interconnecting flexibility, structural communication, and function in RhoGEF oncoproteins. J Chem Inf Model 2019;59(10):4300–13.
[65] Felline A, Raimondi F, Gentile S, Fanelli F. Structural communication between the GTPase Sec4p and its activator Sec2p: Determinants of GEF activity and early deformations to nucleotide release. Comput Struct Biotechnol J 2022;20:5162–80.
[66] Raimondi F, Felline A, Fanelli F. Catching functional modes and structural communication in Dbl family Rho guanine nucleotide exchange factors. J Chem Inf Model 2015;55:1878–93.
[67] Raimondi F, Orozco M, Fanelli F. Deciphering the deformation modes associated with function retention and specialization in members of the Ras superfamily. Structure 2010;18(3):402–14.

[68] Kopec W, Rothberg BS, de Groot BL. Molecular mechanism of a potassium channel gating through activation gate-selectivity filter coupling. Nat Commun 2019;10(1):5366.
[69] Hall JE, Freites JA, Tobias DJ. Experimental and simulation studies of aquaporin 0 water permeability and regulation. Chem Rev 2019;119(9):6015–39.
[70] Yee AW, Aldeghi M, Blakeley MP, Ostermann A, Mas PJ, Moulin M, et al. A molecular mechanism for transthyretin amyloidogenesis. Nat Commun 2019;10(1):925.
[71] Igaev M, Grubmuller H. Microtubule assembly governed by tubulin allosteric gain in flexibility and lattice induced fit. Elife 2018:7.
[72] Khavnekar S, Dantu SC, Sedelnikova S, Ayora S, Rafferty J, Kale A. Structural insights into dynamics of RecU-HJ complex formation elucidates key role of NTR and stalk region toward formation of reactive state. Nucleic Acids Res 2017;45(2):975–86.
[73] Manibog K, Sankar K, Kim SA, Zhang Y, Jernigan RL, Sivasankar S. Molecular determinants of cadherin ideal bond formation: Conformation-dependent unbinding on a multidimensional landscape. Proc Natl Acad Sci USA 2016;113(39):E5711–20.
[74] Pandini A, Morcos F, Khan S. The gearbox of the bacterial flagellar motor switch. Structure 2016;24(7):1209–20.
[75] Smith CA, Ban D, Pratihar S, Giller K, Paulat M, Becker S, et al. Allosteric switch regulates protein-protein binding through collective motion. Proc Natl Acad Sci USA 2016;113(12):3269–74.
[76] Vesper MD, de Groot BL. Collective dynamics underlying allosteric transitions in hemoglobin. PLoS Comput Biol 2013;9(9):e1003232.
[77] Raimondi F, Portella G, Orozco M, Fanelli F. Nucleotide binding switches the information flow in ras GTPases. PLoS Comput Biol 2011;7(3):e1001098.
[78] Mariani S, Dell'Orco D, Felline A, Raimondi F, Fanelli F. Network and atomistic simulations unveil the structural determinants of mutations linked to retinal diseases. PLoS Comput Biol 2013;9(8):e1003207.
[79] Bellucci L, Felline A, Fanelli F. Dynamics and structural communication in the ternary complex of fully phosphorylated V2 vasopressin receptor, vasopressin, and beta-arrestin 1. Biochim Biophys Acta Biomembr 2020;1862(9):183355.
[80] Blacklock K, Verkhivker GM. Computational modeling of allosteric regulation in the hsp90 chaperones: a statistical ensemble analysis of protein structure networks and allosteric communications. PLoS Comput Biol 2014;10(6):e1003679.
[81] Stetz G, Tse A, Verkhivker GM. Ensemble-based modeling and rigidity decomposition of allosteric interaction networks and communication pathways in cyclin-dependent kinases: Differentiating kinase clients of the Hsp90-Cdc37 chaperone. PLoS One 2017;12(11):e0186089.
[82] Stetz G, Verkhivker GM. Computational analysis of residue interaction networks and coevolutionary relationships in the Hsp70 chaperones: a community-hopping model of allosteric regulation and communication. PLoS Comput Biol 2017;13(1):e1005299.
[83] Verkhivker GM, Agajanian S, Oztas DY, Gupta G. Allosteric control of structural mimicry and mutational escape in the SARS-CoV-2 spike protein complexes with the ACE2 decoys and miniprotein inhibitors: a network-based approach for mutational profiling of binding and signaling. J Chem Inf Model 2021;61(10):5172–91.
[84] Verkhivker GM, Agajanian S, Oztas DY, Gupta G. Dynamic profiling of binding and allosteric propensities of the SARS-CoV-2 spike protein with different classes of antibodies: mutational and perturbation-based scanning reveals the allosteric duality of functionally adaptable hotspots. J Chem Theory Comput 2021;17:4578–98.
[85] Verkhivker GM, Agajanian S, Oztas D, Gupta G. Computational analysis of protein stability and allosteric interaction networks in distinct conformational forms of the SARS-CoV-2 spike D614G mutant: reconciling functional mechanisms through allosteric model of spike regulation. J Biomol Struct Dyn 2021:1–18.
[86] Delalande O, Sacquin-Mora S, Baaden M. Enzyme closure and nucleotide binding structurally lock guanylate kinase. Biophys J 2011;101(6):1440–9.
[87] Sacquin-Mora S, Delalande O, Baaden M. Functional modes and residue flexibility control the anisotropic response of guanylate kinase to mechanical stress. Biophys J 2010;99(10):3412–9.
[88] Bocahut A, Bernad S, Sebban P, Sacquin-Mora S. Frontier residues lining globin internal cavities present specific mechanical properties. J Am Chem Soc 2011;133(22):8753–61.
[89] Colloc'h N, Sacquin-Mora S, Avella G, Dhaussy AC, Prange T, Vallone B, et al. Determinants of neuroglobin plasticity highlighted by joint coarse-grained simulations and high pressure crystallography. Sci Rep 2017;7(1):1858.
[90] Kar RK, Kharerin H, Padinhateeri R, Bhat PJ. Multiple conformations of gal3 protein drive the galactose-induced allosteric activation of the GAL genetic switch of saccharomyces cerevisiae. J Mol Biol 2017;429(1):158–76.
[91] Ikeguchi M, Ueno J, Sato M, Kidera A. Protein structural change upon ligand binding: linear response theory. Phys Rev Lett 2005;94(7):078102.
[92] Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR. Manipulation of conformational change in proteins by single-residue perturbations. Biophys J 2010;99(3):933–43.
[93] Atilgan AR, Aykut AO, Atilgan C. Subtle pH differences trigger single residue motions for moderating conformations of calmodulin. J Chem Phys 2011;135(15):155102.
[94] General IJ, Liu Y, Blackburn ME, Mao W, Gierasch LM, Bahar I. ATPase subdomain IA is a mediator of interdomain allostery in Hsp70 molecular chaperones. PLoS Comput Biol 2014;10(5):e1003624.
[95] Kim H, Zou T, Modi C, Dorner K, Grunkemeyer TJ, Chen L, et al. A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins. Structure 2015;23(1):34–43.
[96] Kumar A, Glembo TJ, Ozkan SB. The role of conformational dynamics and allostery in the disease development of human ferritin. Biophys J 2015;109(6):1273–81.
[97] Penkler D, Sensoy O, Atilgan C, Tastan Bishop O. Perturbation-response scanning reveals key residues for allosteric control in Hsp70. J Chem Inf Model 2017;57(6):1359–74.
[98] Larrimore KE, Kazan IC, Kannan L, Kendle RP, Jamal T, Barcus M, et al. Plant-expressed cocaine hydrolase variants of butyrylcholinesterase exhibit altered allosteric effects of cholinesterase activity and increased inhibitor sensitivity. Sci Rep 2017;7(1):10419.
[99] Ponzoni L, Bahar I. Structural dynamics is a determinant of the functional significance of missense variants. Proc Natl Acad Sci USA 2018;115(16):4164–74.
[100] Penkler DL, Atilgan C, Tastan Bishop O. Allosteric modulation of human Hsp90alpha conformational dynamics. J Chem Inf Model 2018;58(2):383–74.
[101] Ross CJ, Atilgan AR, Tastan Bishop O, Atilgan C. Unraveling the motions behind enterovirus 71 uncoating. Biophys J 2018;114(4):822–38.
[102] Stetz G, Tse A, Verkhivker GM. Dissecting structure-encoded determinants of allosteric cross-talk between post-translational modification sites in the Hsp90 chaperones. Sci Rep 2018;8(1):6899.
[103] Mikulska-Ruminska K, Shrivastava I, Krieger J, Zhang S, Li H, Bayir H, et al. Characterization of differential dynamics, specificity, and allostery of lipoxygenase family members. J Chem Inf Model 2019;59(5):2496–508.
[104] Astl L, Verkhivker GM. Atomistic modeling of the ABL kinase regulation by allosteric modulators using structural perturbation analysis and community-based network reconstruction of allosteric communications. J Chem Theory Comput 2019;15(5):3362–80.
[105] Mikulska-Ruminska K, Strzelecki J, Nowak W. Dynamics, nanomechanics and signal transduction in reelin repeats. Sci Rep 2019;9(1):18974.
[106] Xiao F, Song X, Tian P, Gan M, Verkhivker GM, Hu G. Comparative dynamics and functional mechanisms of the CYP17A1 tunnels regulated by ligand binding. J Chem Inf Model 2020;60(7):3632–47.
[107] Astl L, Verkhivker GM. Dynamic view of allosteric regulation in the Hsp70 chaperones by J-domain cochaperone and post-translational modifications: computational analysis of Hsp70 mechanisms by exploring conformational landscapes and residue interaction networks. J Chem Inf Model 2020;60(3):1614–31.
[108] Huang S, Mei H, Lu L, Kuang Z, Heng Y, Xu L, et al. Conformational transitions of caspase-6 in substrate-induced activation process explored by perturbation-response scanning combined with targeted molecular dynamics. Comput Struct Biotechnol J 2021;19:4156–64.
[109] Modi T, Risso VA, Martinez-Rodriguez S, Gavira JA, Mebrat MD, Van Horn WD, et al. Hinge-shift mechanism as a protein design principle for the evolution of beta-lactamases from substrate promiscuity to specificity. Nat Commun 2021;12(1):1852.
[110] Smith IN, Dawson JE, Krieger J, Thacker S, Bahar I, Eng C. Structural and dynamic effects of PTEN C-terminal tail phosphorylation. J Chem Inf Model 2022;62(17):4175–90.
[111] Campitelli P, Lu J, Ozkan SB. Dynamic allostery highlights the evolutionary differences between the CoV-1 and CoV-2 main proteases. Biophys J 2022;121(8):1483–92.
[112] Nemecz A, Prevost MS, Menny A, Corringer PJ. Emerging molecular mechanisms of signal transduction in pentameric ligand-gated ion channels. Neuron 2016;90(3):452–70.
[113] Cecchini M, Changeux JP. The nicotinic acetylcholine receptor and its prokaryotic homologues: Structure, conformational transitions & allosteric modulation. Neuropharmacology 2015;96(Pt B):137–49.
[114] Martin NE, Malik S, Calimet N, Changeux JP, Cecchini M. Un-gating and allosteric modulation of a pentameric ligand-gated ion channel captured by molecular dynamics. PLoS Comput Biol 2017;13(10):e1005784.
[115] Bergh C, Heusser SA, Howard R, Lindahl E. Markov state models of proton- and pore-dependent activation in a pentameric ligand-gated ion channel. Elife 2021:10.
[116] Cerdan AH, Peverini L, Changeux JP, Corringer PJ, Cecchini M. Lateral fenestrations in the extracellular domain of the glycine receptor contribute to the main chloride permeation pathway. Sci Adv 2022;8(41):eadc9340.
[117] Aryal P, Sansom MS, Tucker SJ. Hydrophobic gating in ion channels. J Mol Biol 2015;427(1):121–30.
[118] Lamichhane TR, Ghimire MP. Evaluation of SARS-CoV-2 main protease and inhibitor interactions using dihedral angle distributions and radial distribution function. Heliyon 2021;7(10):e08220.
[119] Xiao S, Zhu H, Wang L, Liang H. DNA conformational flexibility study using phosphate backbone neutralization model. Soft Matter 2014;10(7):1045–30.
[120] Palma CA, Cecchini M, Samori P. Predicting self-assembly: from empirism to determinism. Chem Soc Rev 2012;41(10):3713–30.
[121] Conti S, Cecchini M. Predicting molecular self-assembly at surfaces: a statistical thermodynamics and modeling approach. Phys Chem Chem Phys 2016;18(46):31480–93.
[122] Palma CA, Samori P, Cecchini M. Atomistic simulations of 2D bicomponent self-assembly: from molecular recognition to self-healing. J Am Chem Soc 2010;132(50):17880–5.

[123] Felline A, Seeber M, Fanelli F. PSNtools for standalone and web-based structure network analyses of conformational ensembles. Comput Struct Biotechnol J 2022;20:640–9.
[124] Felline A, Seeber M, Fanelli F. webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules. Nucleic Acids Res 2020;48(W1). W94–W103.
[125] Abraham M, Apostolov R, Barnoud J, Bauer P, Blau C, Bonvin A, et al. Sharing data from molecular simulations. J Chem Inf Model 2019;59(10):4093–9.
[126] Hospital A, Battistini F, Soliva R, Gelpi JL, Orozco M. Surviving the deluge of biosimulation data. Wiley Interdiscip Rev: Comput Mol Sci 2020;10. e1449.