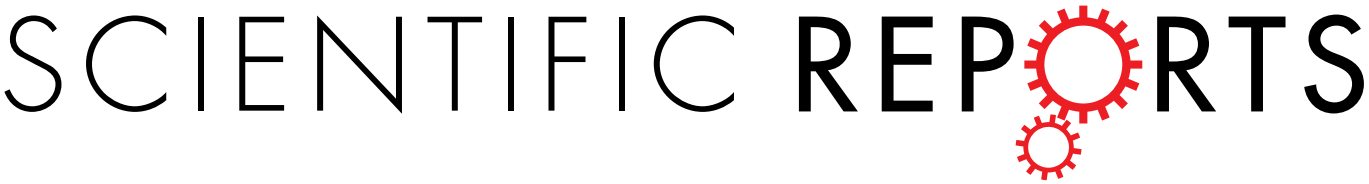

OPEN

An optimal distance cutoff for contact-based Protein Structure Networks using side-chain centers of mass

Received: 23 January 2017
Accepted: 28 March 2017
Published online: 06 June 2017

Juan Salamanca Viloria, Maria Francesca Allega, Matteo Lambrughi & Elena Papaleo

Proteins are highly dynamic entities attaining a myriad of different conformations. Protein side chains change their states during dynamics, causing clashes that are propagated at distal sites. A convenient formalism to analyze protein dynamics is based on network theory using Protein Structure Networks (PSNs). Despite their broad applicability, few efforts have been devoted to benchmarking PSN methods and to provide the community with best practices. In many applications, it is convenient to use the centers of mass of the side chains as nodes. It becomes thus critical to evaluate the minimal distance cutoff between the centers of mass which will provide stable network properties. Moreover, when the PSN is derived from a structural ensemble collected by molecular dynamics (MD), the impact of the MD force field has to be evaluated. We selected a dataset of proteins with different fold and size and assessed the two fundamental properties of the PSN, i.e. hubs and connected components. We identified an optimal cutoff of 5 Å that is robust to changes in the force field and the proteins. Our study builds solid foundations for the harmonization and standardization of the PSN approach.

Proteins are complex and highly dynamic entities attaining a myriad of different conformations in solution[1–5] that are often related to the protein function. Indeed, they can resemble bound states to a biological partner[6–10], active states of enzymes[11–14], or conformations that are stabilized by a post-translational modification (PTM)[6,11], as well as altered by a disease-related mutation[15].

An interesting property of proteins is that a perturbation (e.g. a binding event, a mutation or a PTM) occurring at a certain site of the structure can be transmitted over long distances to another location[16–19]. This long-range communication is often related to allostery and may affect critical distal sites for protein function.

At the atom-level, the perturbation from one protein site to a distal one can be propagated by a cascade of collisional clashes between residue side chains, which undergo changes of their rotameric states during protein dynamics[19,20]. Local rearrangements occurring in the intramolecular contacts during the protein dynamics are thus at the base of this long-range communication[19].

A convenient formalism to unravel the complexity behind long-range structural communication in proteins is the application of network theory to protein structure, i.e. the so-called Protein Structure Networks (PSNs). In a PSN, the protein residues become the nodes of the network connected by edges which can, for example, be described as the contact strength between each pair of residues[20–30]. Networks indeed are proper tools to link the local to global perturbations occurring during protein dynamics since they are by definition mediators of communication from local to global scales[19].

Nowadays, PSN-based strategies are very popular and used in structural biology, and a plethora of different methodologies has been proposed[25–28,31–37]. PSN approaches are often integrated to the dynamic description of proteins that all-atom molecular dynamics (MD) simulations or other sampling methods provide[21,31,38–45].

Despite their broad applicability, few efforts have been devoted so far to the benchmarking of PSN and PSN-MD methods, to define best practices in the field and to ultimately provide the community with clear rules to determine PSN optimal parameters. The definition of arbitrary cutoffs is one of the major weaknesses of contact-based networks applied to protein structure and dynamics[46,47]. As previously shown, many options are

Computational Biology Laboratory, Danish Cancer Society Research Center, Strandboulevarden 49, 2100, Copenhagen, Denmark. Correspondence and requests for materials should be addressed to E.P. (email: elenap@cancer.dk)

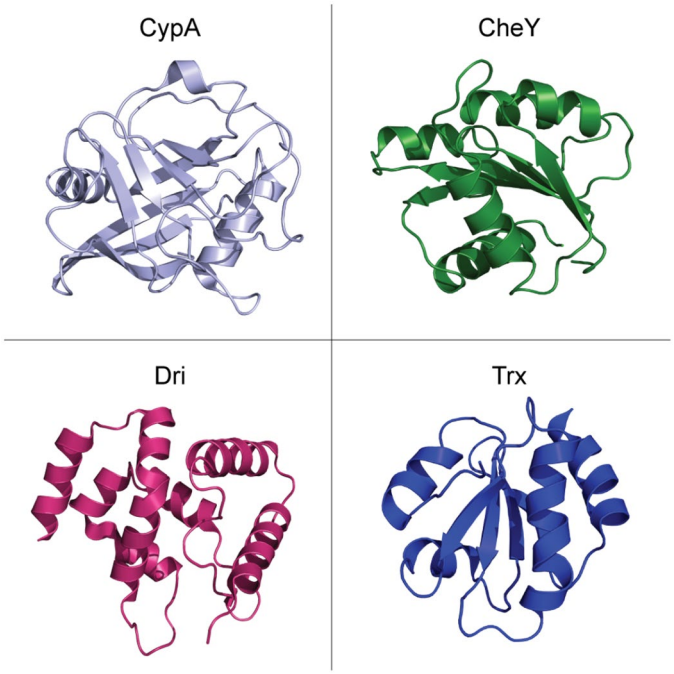

Figure 1. 3D of the selected proteins for molecular simulations. The 3D structures of the Cyclophilin A (CypA) from *H.sapiens* (PDB entry 3K0N), Chemotaxis protein (CheY) from *E.coli* (PDB entry 3CHY), the DNA-binding domain of the Dead ringer protein (Dri) from *D.melanogaster* (PDB entry 1C20) and the Thioredoxin (Trx) from *B.acidocaldarius* (PDB entry 1QUW) are shown as light blue, green, magenta and blue cartoon, respectively.

available to select suitable distance cutoffs for the prediction of residue contacts in protein structures[47]. Alternative solutions exist, i.e. using different principles for edge and weight definition such as energies or correlated motions. Nevertheless, a contact-based approach is still valuable especially if we consider the major advances that techniques such as atomistic biomolecular simulations have achieved in the last decade[48,49]. Indeed, MD simulations have now reached high accuracy in describing conformational changes even at the side-chain levels and occurring on different time scales, as attested by the agreement with experimental observables[4,50–53].

In many PSN-MD applications, it is convenient to use the centers of mass of residue side chains as PSN nodes, the distance between the centers of mass for edge definition and their occurrence as weight[20,31,41,54,55]. It becomes, thus, critical to evaluate the minimal distance cutoff between the centers of mass of two residues to include an edge in the PSN. Moreover, when the PSN is derived from a structural ensemble collected by MD simulations and not from experimental structures, it is mandatory to evaluate the impact of the physical model (i.e. force field) on the PSN parameters.

We selected a dataset of proteins with different architecture and size and assessed the distribution of the two fundamental properties of a PSN, i.e. the hubs and the connected components. We also evaluated the influence of the force field selection on the PSN parameters, and we propose an optimal distance cutoff for PSN based on distances between the centers of mass of protein residues. The cutoff here identified is robust independently on the protein size, fold, and the MD force field employed. Our study builds strong foundations toward the harmonization and standardization of PSN strategies and a framework to apply also more generally to the choice of parameters for other PSN-based approaches.

Results and Discussion

Selected protein structures for PSN-MD analyses. We selected four different three-dimensional (3D) structures of monomeric proteins of various size and fold (Fig. 1) and four different force fields (Table S1). In particular, we chose state-of-the-art physical models from each of the most used force-field families for MD simulations of proteins, i.e. CHARMM (CHARMM22*[56] and CHARMM36[57]), AMBER (Amber99SB*-ILDN[58,59]) and GROMOS (GROMOS54a7[60]). We carried out the MD simulations in explicit solvent for one μs so that they could reflect the MD sampling that is employed for PSN-MD studies[40,54]. For each MD ensemble, PSN based on distances between the side-chain centers of mass have been calculated as detailed in the Materials and Methods.

A distance cutoff of 5 Å allows a robust description of PSN properties independently on the protein and the MD force field employed. The choice of the distance cutoff is essential for the PSN definition. Indeed, the distance cutoff is used to discriminate which contact between two side chains has to be included or not as a link of the network, ultimately affecting the network topology. When the distance is calculated between the centers of mass of the residue side chains, the choice of the cutoff becomes even more critical. Indeed, we cannot arbitrarily assume that the distances commonly used in structural biology to define an interaction between

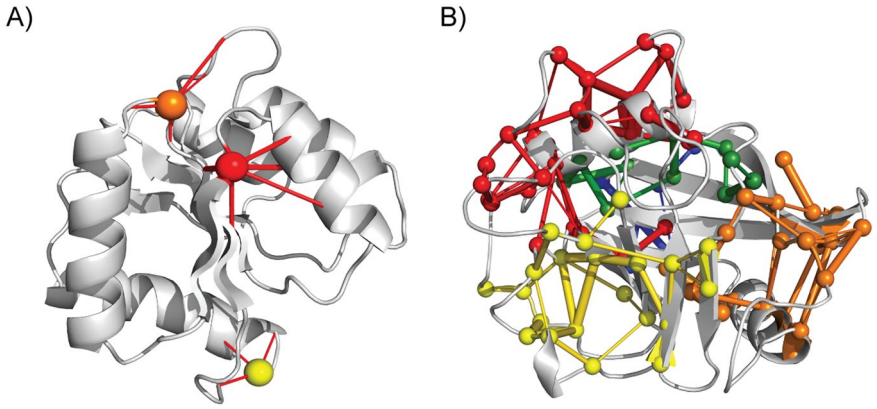

Figure 2. Schematic representation of hubs and connected components. Hubs are nodes that have a degree higher than the average degree of the nodes of a network. In PSN, we consider as hubs only those nodes having a number of edges greater than or equal to three. Hubs with a degree of three, four and five are shown in yellow, orange and red, respectively (**A**). The connected components are clusters of linked nodes with no edges in common with nodes that belong to the other clusters of the PSN. As an example, five connected components are shown (**B**).

two amino acids - such as 4 or 4.5 Å - are valid. The issue becomes even more cogent when a PSN is derived by an MD ensemble where each force field relies on different atomic masses.

The two most important properties of a PSN, which ultimately dictate how distant regions of the PSN are linked are the so-called hubs and connected components (also known as clusters of nodes) (Fig. 2).

Hubs are nodes that have a high degree of connectivity in a network. The highest degree of residue hubs is limited by steric constraints and it could vary from three to ten in PSN[27]. Protein structures are known to be made up of a significant number of strongly and weakly interacting residue hubs that stabilize the tertiary structure of the protein and provide resilience against random mutations[19,27].

A robust PSN should feature a certain amount of hub residues that have at least a node degree of three (i.e. connected with three or more other nodes by an edge in the PSN) and it should be composed of multiple connected components which are not too fragmented. Cluster fragmentation is particularly critical in the PSN definition. Other colleagues and we showed that central parameters that influence the size of the connected components are the p_{crit}[31,42] or I_{crit}[40,61,62], depending on the methods used for PSN construction. Indeed, edges that have extremely low weights would increase the noise and connect all the clusters into a single one. Conversely, if only high weights are retained, only sparsely populated and highly fragmented clusters will be observed with a minimal number of communication paths between distal regions.

In a PSN approach based on side-chain-side-chain contacts, the distance cutoffs used can affect the network in a similar way. Indeed, if a distance that is too short and restrictive is chosen, the network will appear as very fragmented with small separated clusters and few or virtually no hubs. If the distance is too long, each residue of the network will be connected, resulting in a single cluster that embraces the entire network. It is thus critical to find an optimal distance cutoff.

Moreover, since the PSN-MD approaches, as the one here employed, generally rely on extracting an average and static PSN from an MD trajectory, it becomes fundamental to assess the convergence of hubs and connected components over the simulation time.

We thus here evaluated: (i) the convergence of hubs and connected components in PSN derived by MD simulations using a Jackknife approach (see Materials and Methods) and (ii) the distribution of hubs and connected components at different distance cutoffs (Figs 3 and 4, Fig. S1). (iii) In the attempt of harmonizing the PSN protocol and allowing the reproducibility of the analyses, we also implemented a Python-based pipeline (PyInKnife.py) to automatize the steps described above, which can be used free of charge (see Materials and Methods for details).

At first, we evaluated whether hubs and connected components are stable properties in the MD ensembles here collected (Figs 3 and 4, Fig. S1). With regards to the distance cutoff, we identified common trends in the hubs and connected components distribution independently from the protein under investigation and the force field employed in the simulations. Indeed, in all the cases distance cutoffs lower than 5 Å resulted in a minimal number of hubs (less than four hub residues) where the connection degree was smaller than three (Fig. 3). On the contrary, distance cutoffs higher than 5 Å showed only one large cluster accounting for most of the protein residues (Fig. 4), indicating that this value is the more appropriate cutoff to employ for a PSN-MD where the contacts are calculated as distances between the centers of mass of residue side chains.

Localization of hubs and connected components on the 3D structure is conserved using the 5 Å distance cutoff. The 5 Å distance cutoff allows for similar general features of the PSN of the same protein described by different force fields (Figs 3 and 4). Despite this result is encouraging, we need to take into consideration that PSNs are employed to achieve residue-level details in structural biology. PSNs are used to identify the localization of the hub residues, the specific residues that belong to the same cluster or even the paths of communication between distal residues and their intermediate nodes. These are all important PSN properties that can,

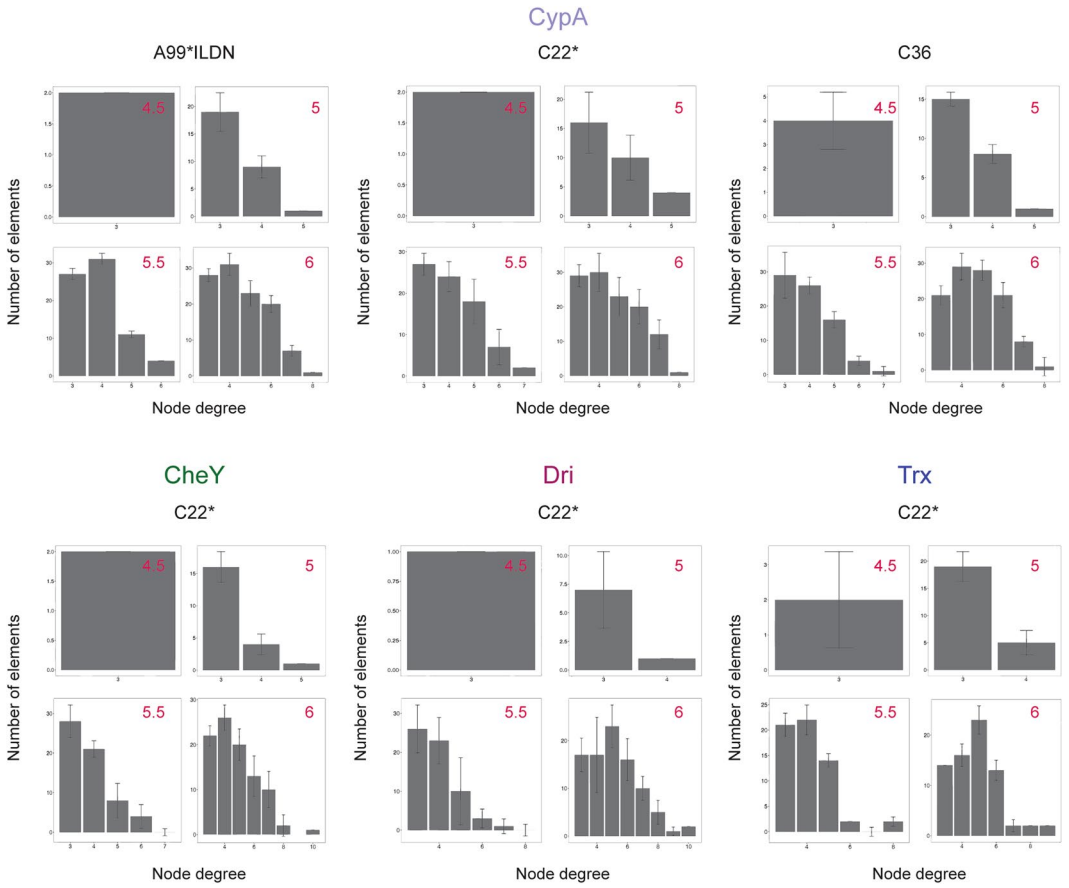

Figure 3. Hub distribution at different distance cutoffs used for the PSN-MD analyses. We evaluated the changes in the number of hubs and their node degree as a function of different distance cutoffs in the PSN derived from the entire MD trajectory (histogram values) and the associated standard deviations (error bars) calculated from the average PSNs obtained from the Jackknife resampled trajectories (see Materials and Methods). We noticed that hubs are virtually absent at distance cutoffs lower than 5 Å.

for example, be altered by interactions with biological partners[6,40,63] or mutations[21,40,42,51,64]. It is thus not enough to observe that the PSN description is robust regarding the overall distribution of hubs and connected components. Indeed, the PSNs collected for the same protein, but using different MD force fields, with the 5 Å distance cutoff might differ in the localization of hub residues in the 3D structure or in the individual residues that belong to the same cluster without affecting the total number of hubs and connected components. The same observation holds for the localization of hubs and connected components when the entire MD trajectory is compared to the resampled MD trajectories collected from the Jackknife approach.

We thus compared the hubs and connected components at the residue-level as derived by the PSN analyses of the entire MD trajectories or of the resampled MD trajectories obtained with the Jackknife procedure (see Materials and Methods). The analyses showed a reasonable convergence of hubs and connected components also at the residue-level with only minor discrepancies among the PSN calculated from the entire MD trajectory and few of the resampled trajectories (Figs S2 and S3).

Moreover, we analyzed the hub localization and their degree in the MD simulations of CypA where different force fields have been used (Figs 5 and 6A). We noticed that the localization of the hubs appears to be equally distributed on the 3D structure coming from different force fields, apart from minor changes in their node degree. Similar results were obtained for Trx using CHARMM22* and GROMOS54a7 force fields.

In parallel, we also mapped the first five more populated connected components onto the CypA sequence and 3D structure (Figs 6B and 7). The composition and distribution of the clusters are different only in CHARMM36 simulations. This apparent difference is only due to a splitting of the connected component number 1 in three smaller clusters, as well as to a different localization of the 5th cluster (i.e. the smallest one). Only subtle differences have been observed for Amber99SB*-ILDN and CHARMM22*, suggesting a robust description of the connected components with these two force fields, as also found in a recent PSN study of a dimer[54].

Conclusions

In the protein world, a perturbation occurring at a certain site of the protein structure can be transmitted over long distances to another site. These structural rearrangements can be propagated by a cascade of changes in the conformational states of the residue side chains. Local changes occurring in the residue-residue contacts during

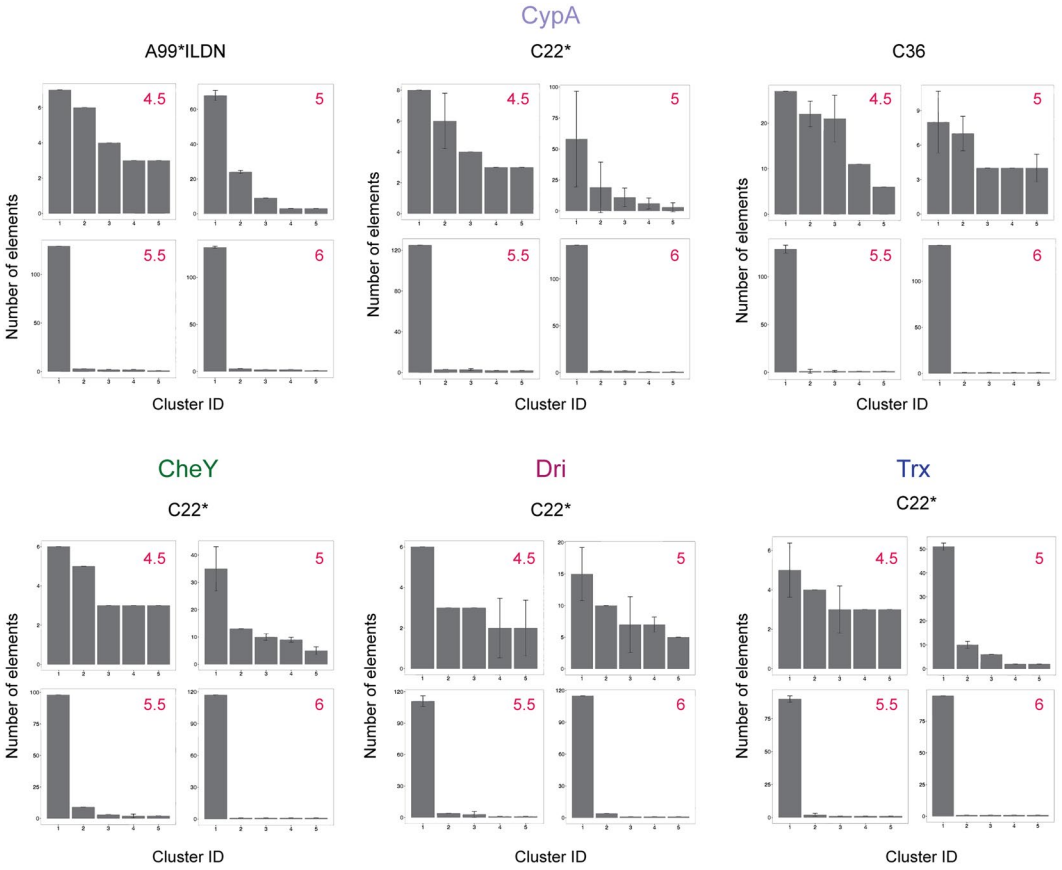

Figure 4. Connected component distribution at different distance cutoffs used for the PSN-MD analyses. We evaluated changes in the population of the connected components (i.e. clusters) as a function of different distance cutoffs in the PSN derived from the entire MD trajectory (histogram values) and the associated standard deviations (error bars) calculated from the average PSNs obtained from the Jackknife resampled trajectories (see Materials and Methods). We reported in the plot only the first five most populated clusters for sake of clarity. We observed that at distance cutoffs higher than 5 Å, most of the nodes of the PSN were located in the same cluster (cluster ID 1). This result suggests that 5 Å is an optimal distance cutoff to predict residue contacts in PSN.

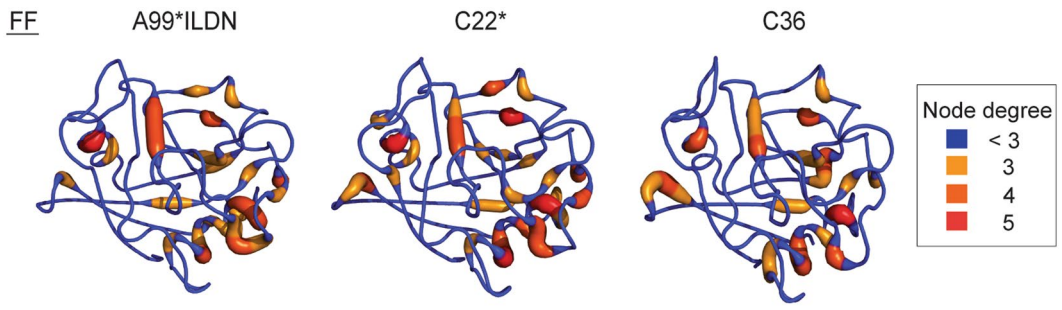

Figure 5. Location of the hub residues on the 3D structure of CypA. We mapped on CypA 3D structure the PSN hubs identified in CypA MD simulations with the three different force fields CHARMM22* (C22*), CHARMM36 (C36) and Amber99SB*-ILDN (A99*ILDN). The different colors and sizes represent the node degree, i.e. the number of edges for each residue.

the protein dynamics are thus at the base of this long-range communication. Network theory is a suitable formalism to evoke to analyze protein structures and to identify the paths of residues that can transmit the structural changes over long distances. In this context, a plethora of different approaches to define a PSN has been developed, often integrated with molecular dynamics simulations to account for the protein dynamics.

Despite the broad application of these methods, the community is missing clear rules and a solid framework to define the PSN parameters. It becomes thus critical to evaluate the minimal distance cutoff that can be used to

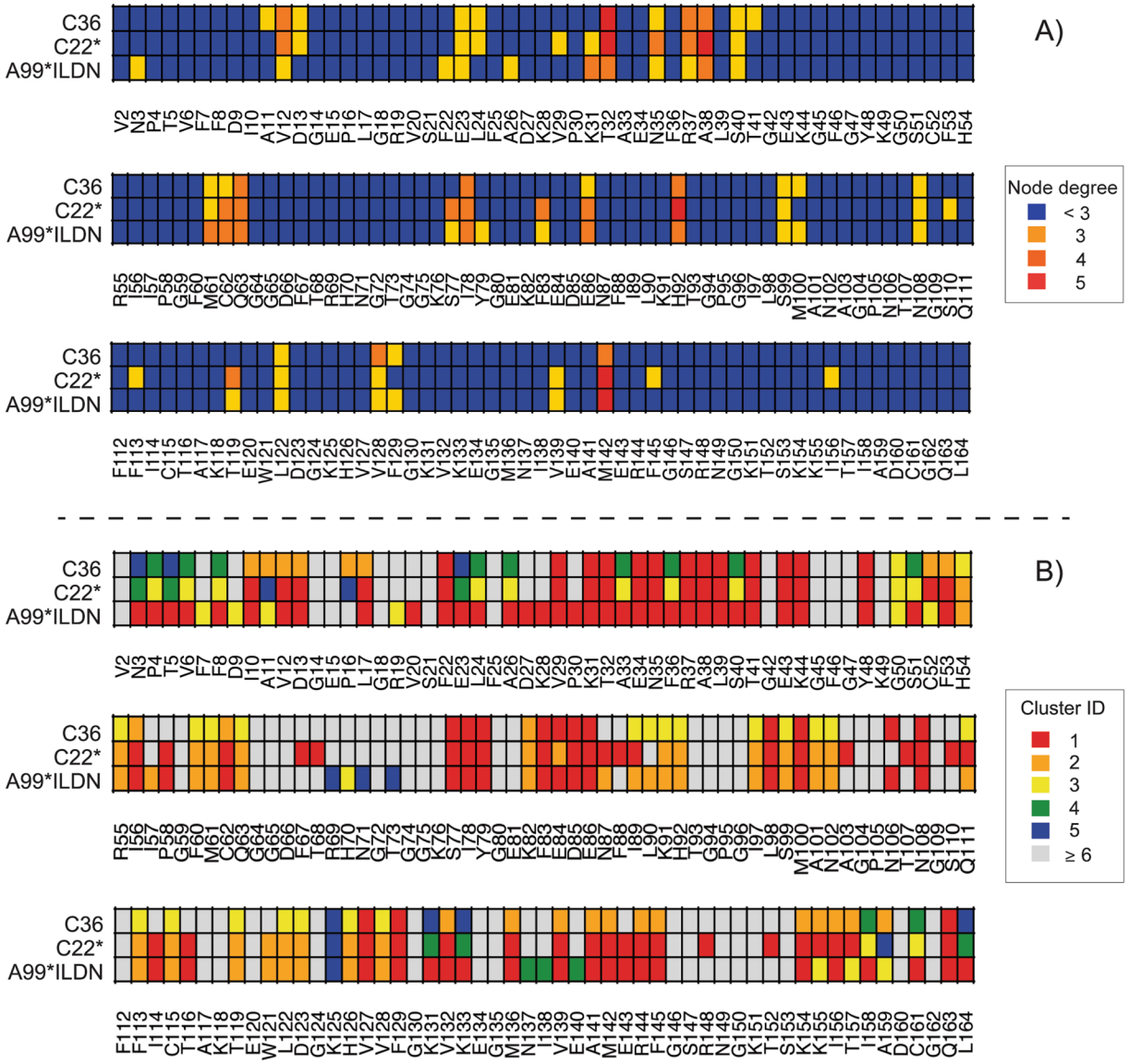

Figure 6. Heatmaps comparing hubs and connected components of CypA MD simulations. We used heatmaps to show the hubs (**A**) and the connected components (i.e. clusters) (**B**) identified in CypA MD simulations with the three different force fields CHARMM22* (C22*), CHARMM36 (C36) and Amber99SB*-ILDN (A99*ILDN). This representation allows to identify changes in node degree or in the cluster where a node is located at the residue-level.

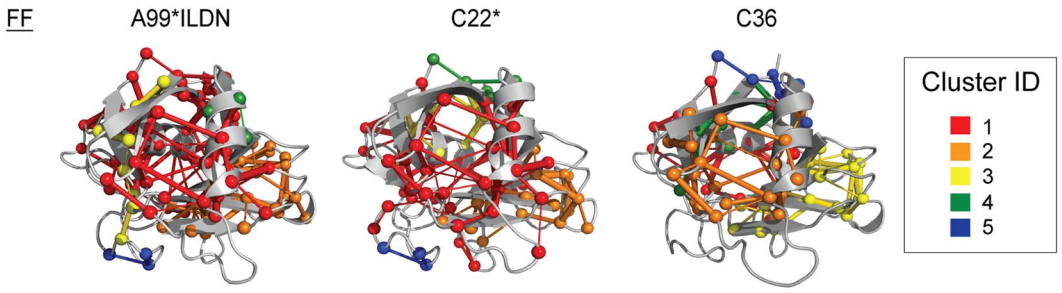

Figure 7. Location of the first five connected components on the 3D structure of CypA. We mapped on CypA 3D structure the PSN connected components identified in CypA MD simulations with the three different force fields CHARMM22* (C22*), CHARMM36 (C36) and Amber99SB*-ILDN (A99*ILDN). The different colors range from red to blue for the largest and smallest clusters, respectively. The nodes are represented as spheres while the edges are visualized as cylinders. The figure has been produced with the *xPyder* plugin[65] for Pymol.

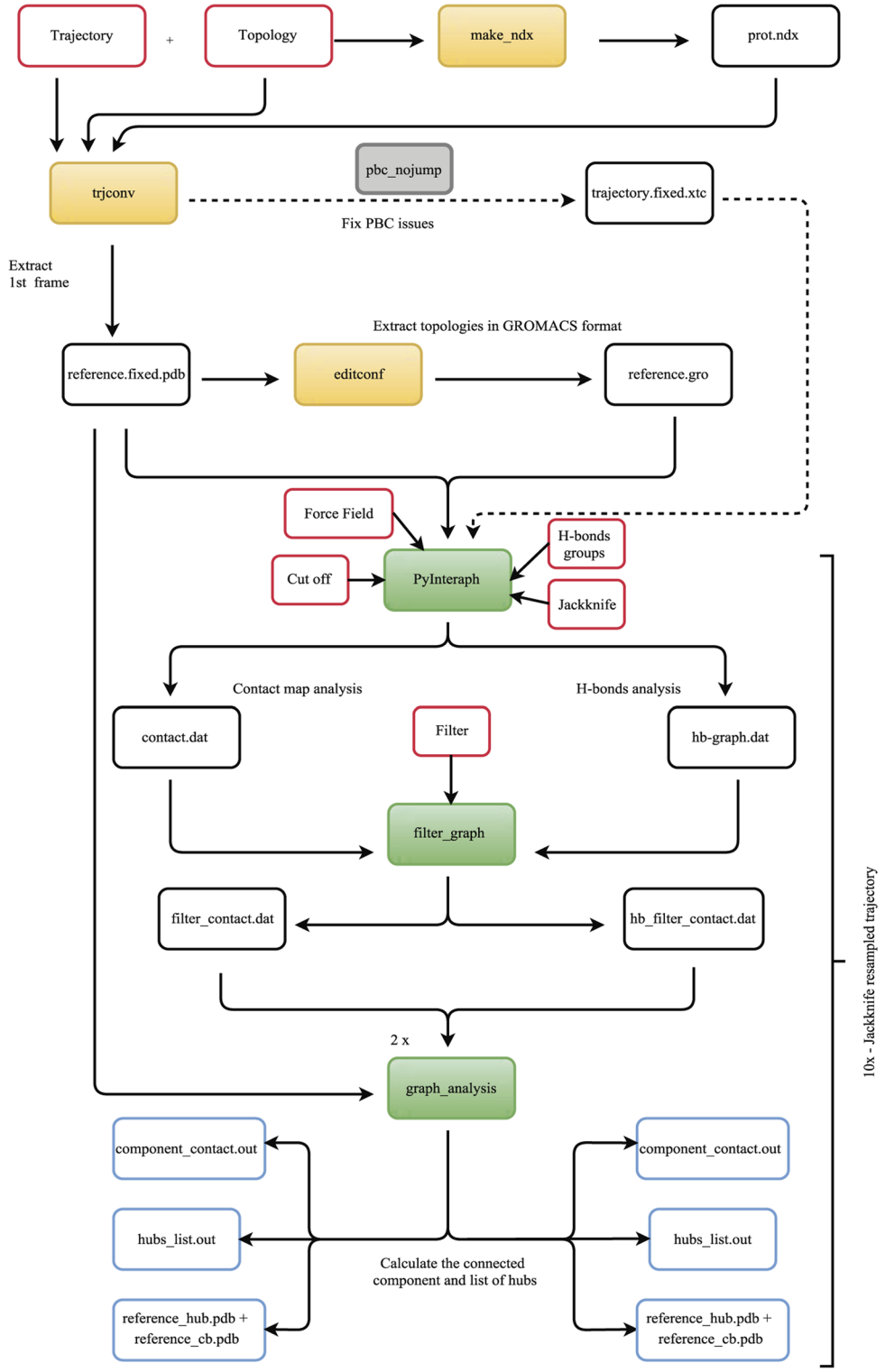

Figure 8. Flow chart for the *PyInKnife* pipeline for PSN analyses. The orange boxes represent the tools used from *GROMACS* software (*make_ndx*: to generate an index file consisting of the groups of interest, *trjconv*: to convert and manipulate trajectory files and *editconf*: to convert and manipulate structure files). White boxes with the red border are the main inputs of the pipeline while the boxes with blue border are the main outputs. An optional flag has been added to post-process the MD trajectories to remove artefacts related to periodic boundary conditions with *trjconv* (grey box). The green boxes represent the commands used from *PyInteraph*. The inputs required for the *PyInteraph* tool are the force field used in the MD simulation, the selected distance cutoff and the optional usage of the Jackknife resampling method. Moreover, the user can choose if to include or not the H-bonds in the PSN analysis. As an input, it is also possible to change the p_{crit} cutoff value used to remove the less informative interactions; by default this value is set to 20.

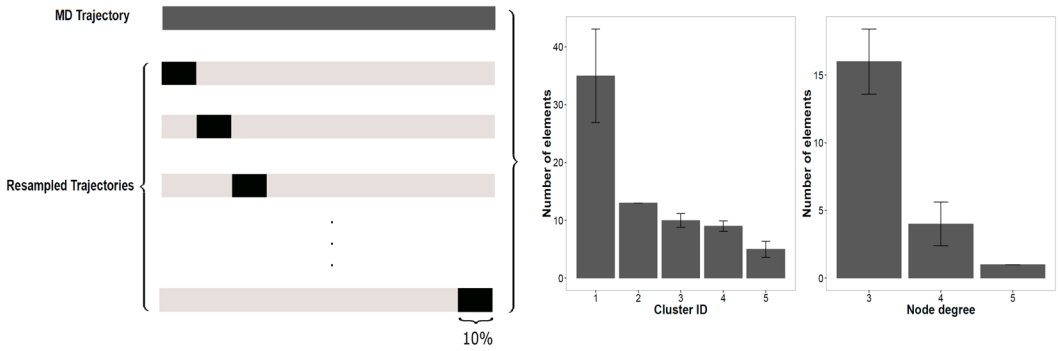

Figure 9. Jackknife resampling as implemented in *PyInKnife*. We show a schematic representation of the MD trajectory and the resampled trajectories used for the calculation of the convergence of PSN properties (left panel). The histograms show the two properties of the PSN analysis calculated for the whole trajectory, the connected components and the hubs (right panels). The example refers to the PSN analysis using a 5 Å distance cutoff of the CheY MD simulation with CHARMM22* force field. Error bars represent the Jackknife standard error from the resampled trajectories.

include an edge in the PSN and that provides stable network properties, as well as the influence of the physical model used to describe the protein in the simulations.

Indeed, there are not consolidated and uniform protocols in the PSN-MD field, especially when the edges are defined according to the distance between the centers of mass of protein side chains. Moreover, most of the PSN approaches have been optimized using datasets of static experimental structures from the Protein Data Bank. A careful evaluation of the PSN parameters in an MD ensemble of structures has been poorly applied. PSN parameters that are optimal for the network analyses of experimental crystallographic structures are not necessarily suitable for the analysis of an MD ensemble, as recently pointed out[40]. Most of the publications in which a PSN was calculated using the *PyInteraph* suite of tools, for example, employ very different distance cutoffs.

We thus selected a dataset of proteins to use as model systems to assess important PSN properties as a function of different distance cutoffs and physical models. In particular, we focused on two fundamental properties of the PSN, i.e. the hubs and the connected components. We identified an optimal value for the distance cutoff (5 Å) that is robust to changes in the MD force field and applicable to proteins with different sizes or folds. Our study provides a general framework to select PSN parameters and to improve reproducibility of the results thanks to a free-of-charge Python-based pipeline, *PyInKnife*. We here built the foundations toward the harmonization and standardization of the PSN-MD approach.

Materials and Methods

Molecular dynamics simulations. We performed explicit solvent MD simulations using the *GROMACS* software version 4.6[66] with different force fields and solvent models. A summary of the starting structures, protein size, force fields and solvent models used in this study is reported in Table S1. The MD simulation of Dri ARID domain has been published before[40] and here employed for the analyses. 500-ns simulations of CypA have been published before[51] and we here elongated them to achieve one μs of sampling. We collected the remaining simulations for the first time in this study at 300 K and 1 bar in the NVT ensemble with 150 mM of NaCl. We employed periodic boundary conditions and we set a distance equal or greater than 1.8 Å from the protein atoms and the box edges of a dodecahedral box of water molecules. Preparation steps have been carried out according to a protocol recently applied to other proteins[67]. We applied a 2-fs time step and the LINCS algorithm[68], as well as the Particle-Mesh Ewald (PME) summation scheme[69] to treat long-range electrostatic interactions. Van der Waals and short-range Coulomb interactions were truncated at 9 Å and conformations stored every 10 ps. We carried out productive MD simulations for one μs.

We calculated the minimal distance between each protein and its image to rule out artifacts due to periodic boundary conditions and artificial contacts between the protein and the corresponding image.

PSN definition. We used the *PyInteraph* suite of tools[31] to construct a PSN-MD based on side-chain contacts using all the residues except for glycines. The contacts are defined as distances between the centers of mass of side chains on the base of the atomic mass files provided by *PyInteraph*. Different distance cutoffs have been assessed in this study in the range of 4–6 Å (see below) to include a certain contact as edge of the network. Moreover, to derive a weighted network, the persistence of the contact in each MD ensemble was measured and a p_{crit} of 20% was employed to filter out meaningless interactions and to maintain the network structure, in agreement with previous applications of the same method[31,42,70]. We also used the *xPyder* plugin[65] for Pymol to map on the 3D structure the PSN connected components.

The PyInKnife pipeline. We developed a Python-based pipeline (which is available free of charge at https://github.com/ELELAB/PyInKnife) called *PyInKnife* in order to: (i) automatize the pre-processing of the trajectories for PSN analyses, (ii) sub-set the trajectories in shorter trajectory files that retain 90% of the frames (see below), (iii) run the different steps of *PyInteraph* on each trajectory subset and using different distance cutoffs, including

the creation of the PSN, calculation of hubs and connected components and their distribution, and (iv) generate a final report with publication-ready plots and figures. The pipeline is illustrated in Fig. 8.

PyInKnife requires the pre-processing of the MD trajectory to remove artefacts due to the periodic boundary conditions and to extract a reference structure along with the topology required for the PSN calculations. The pre-processing is carried out by three different *GROMACS* tools (www.Gromacs.org): *make_ndx, trjconv* and *editconf*. These tools allow us to generate the index file, convert and manipulate the trajectories and structures, respectively.

PyInKnife can be also used on trajectories obtained with other simulation packages, such as *Amber, CHARMM* and *NAMD* after conversion of the MD trajectory to the *GROMACS* format (.xtc or.trr file). This can be achieved with several tools such as *WORDOM*[71], the *MDAnalysis* package[72] and the *Catdcd* plugin (http://www.ks.uiuc.edu/Development/MDTools/catdcd/). The user can employ the GROMACS tool *editconf* to convert the PDB file of the starting structure, or one frame extracted from the trajectory, into the file format required by *PyInteraph* (*GROMACS*.gro file).

PyInKnife allows to automatize the analyses of contact-based PSN, hydrophobic interactions, and hydrogen bond networks implemented in *PyInteraph*. The user can specify from the command line the PyInteraph atomic mass databases, the distance cutoff values to be tested and other PSN parameters.

After the PSN for each MD trajectory is obtained, it is possible to calculate with *PyInKnife* the hubs and connected components for each class of interactions by using the *graph_analysis* tool of the *PyInteraph* suite.

PyInKnife also implements a pipeline to evaluate the convergence of the two most important PSN properties, i.e. hubs and connected components in the MD trajectory. We used the Jackknife resampling method[73] to calculate the deviation from resampled trajectories where a 10% has been discarded at regular intervals of the simulation frames. The resampled trajectories are calculated using the *GROMACS* tool *trjcat*. The procedure is illustrated in Fig. 9.

PyInKnife also includes R-based scripts to plot the results and produce publication-ready figures. To use the plotting R scripts, the R packages *ggplot, ggplot2* and *lattice* are required.

The Jackknife standard error is calculated as

$$SE(\hat{x})_{jack} = \left\{ \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right\}^{1/2}$$

where n is the number of resampled trajectories (10 as default), $\hat{\theta}$ is the estimator of the *ith* resampled trajectory and $\hat{\theta}_{(.)}$ is the empirical average of the estimator on the resampled trajectories

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}$$

References

1. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA* **102**, 6679–85, doi:10.1073/pnas.0408930102 (2005).
2. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **19**, 120–7, doi:10.1016/j.sbi.2009.03.004 (2009).
3. Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in proteins: Signaling and other functions. *Curr. Opin. Struct. Biol.* **20**, 142–147, doi:10.1016/j.sbi.2009.12.004 (2010).
4. Lindorff-Larsen, K. *et al.* Systematic validation of protein force fields against experimental data. *PLoS One* **7**, e32131, doi:10.1371/journal.pone.0032131 (2012).
5. Kovermann, M., Rogne, P. & Wolf-Watz, M. Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys.* **49**, e6, doi:10.1017/S0033583516000019 (2016).
6. Lambrughi, M. *et al.* DNA-binding protects p53 from interactions with cofactors involved in transcription-independent functions. *Nucleic Acids Res.* **44**, 9096–9109, doi:10.1093/nar/gkw770 (2016).
7. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–96, doi:10.1038/nchembio.232 (2009).
8. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5**, 808–814, doi:10.1038/nchembio.238 (2009).
9. Masterson, L. R. *et al.* Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat. Chem. Biol.* **6**, 821–8, doi:10.1038/nchembio.452 (2010).
10. Sumbul, F., Acuner-Ozbabacan, S. E. & Haliloglu, T. Allosteric Dynamic Control of Binding. *Biophys. J.* **109**, 1190–1201, doi:10.1016/j.bpj.2015.08.011 (2015).
11. Papaleo, E. *et al.* An Acidic Loop and Cognate Phosphorylation Sites Define a Molecular Switch That Modulates Ubiquitin Charging Activity in Cdc34-Like Enzymes. *PLoS Comput. Biol.* **7** (2011).
12. Papaleo, E. *et al.* Loop 7 of E2 Enzymes: An Ancestral Conserved Functional Motif Involved in the E2-Mediated Steps of the Ubiquitination Cascade. *PLoS One* **7** (2012).
13. Campbell, E. *et al.* Changes in protein dynamics optimize the active site during evolution of new enzyme function. *Nat. Chem. Biol.* **12**, 944–950, doi:10.1038/nchembio.2175 (2016).
14. Ma, B. & Nussinov, R. Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr. Opin. Chem. Biol.* **14**, 652–9, doi:10.1016/j.cbpa.2010.08.012 (2010).
15. Demir, Ö. *et al.* Ensemble-based computational approach discriminates functional activity of p53 cancer and rescue mutants. *PLoS Comput. Biol.* **7** (2011).
16. Guo, J. & Zhou, H. X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **116**, 6503–6515, doi:10.1021/acs.chemrev.5b00590 (2016).
17. Ribeiro, A. A. S. T. & Ortiz, V. A Chemical Perspective on Allostery. *Chem. Rev.* **116**, 6488–6502, doi:10.1021/acs.chemrev.5b00543 (2016).
18. Papaleo, E. *et al.* The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **116**, 6391–6423, doi:10.1021/acs.chemrev.5b00623 (2016).

19. Vuillon, L. & Lesieur, C. From local to global changes in proteins: a network view. *Curr. Opin. Struct. Biol.* **31**, 1–8, doi:10.1016/j.sbi.2015.02.015 (2015).
20. Papaleo, E. Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: strength in unity. *Front. Mol. Biosci.* **2**, 28, doi:10.3389/fmolb.2015.00028 (2015).
21. Angelova, K. *et al*. Conserved amino acids participate in the structure networks deputed to intramolecular communication in the lutropin receptor. *Cell. Mol. Life Sci.* **68**, 1227–39, doi:10.1007/s00018-010-0519-z (2011).
22. Di Paola, L., De Ruvo, M., Paci, P., Santoni, D. & Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chem. Rev.* **113**, 1598–1613, doi:10.1021/cr3002356 (2013).
23. Di Paola, L. & Giuliani, A. Protein contact network topology: a natural language for allostery. *Curr. Opin. Struct. Biol.* **31**, 43–8, doi:10.1016/j.sbi.2015.03.001 (2015).
24. Cheng, S., Fu, H. & Cui, D.-X. Characteristics Analyses and Comparisons of the Protein Structure Networks Constructed by Different Methods. *Interdiscip. Sci Comput Life Sci* **8**, 65–74, doi:10.1007/s12539-015-0106-y (2016).
25. O'Rourke, K. F., Gorman, S. D. & Boehr, D. D. Biophysical and computational methods to analyze amino acid interaction networks in proteins. *Comput. Struct. Biotechnol. J.* **14**, 245–251, doi:10.1016/j.csbj.2016.06.002 (2016).
26. Feher, V. A., Durrant, J. D., Van Wart, A. T. & Amaro, R. E. Computational approaches to mapping allosteric pathways. *Curr. Opin. Struct. Biol.* **25**, 98–103, doi:10.1016/j.sbi.2014.02.004 (2014).
27. Bhattacharyya, M., Ghosh, S. & Vishveshwara, S. Protein Structure and Function: Looking through the Network of Side-Chain Interactions. *Curr. Protein Pept. Sci.* **17**, 4–25, doi:10.2174/1389203716666150923105727 (2016).
28. van den Bedem, H., Bhabha, G., Yang, K., Wright, P. E. & Fraser, J. S. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat. Methods* **10**, 896–902, doi:10.1038/nmeth.2592 (2013).
29. Csermely, P., Nussinov, R. & Szilágyi, A. From allosteric drugs to allo-network drugs: state of the art and trends of design, synthesis and computational methods. *Curr. Top. Med. Chem.* **13**, 2–4, doi:10.2174/1568026611313010002 (2013).
30. Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **138**, 333–408, doi:10.1016/j.pharmthera.2013.01.016 (2013).
31. Tiberti, M. *et al*. PyInteraph: A Framework for the Analysis of Interaction Networks in Structural Ensembles of Proteins. *J Chem Inf Model* **54**, 1537–1551, doi:10.1021/ci400639r (2014).
32. Van Wart, A. T., Durrant, J., Votapka, L. & Amaro, R. E. Weighted implementation of suboptimal paths (WISP): An optimized algorithm and tool for dynamical network analysis. *J. Chem. Theory Comput.* **10**, 511–517, doi:10.1021/ct4008603 (2014).
33. Chakrabarty, B. & Parekh, N. NAPS: Network analysis of protein structures. *Nucleic Acids Res.* **44**, W375–W382, doi:10.1093/nar/gkw383 (2016).
34. Seeber, M., Felline, A., Raimondi, F., Mariani, S. & Fanelli, F. WebPSN: A web server for high-throughput investigation of structural communication in biomacromolecules. *Bioinformatics* **31**, 779–781, doi:10.1093/bioinformatics/btu718 (2015).
35. Stolzenberg, S., Michino, M., Levine, M. V., Weinstein, H. & Shi, L. Computational approaches to detect allosteric pathways in transmembrane molecular machines. *Biochim. Biophys. Acta - Biomembr.* **1858**, 1652–1662, doi:10.1016/j.bbamem.2016.01.010 (2016).
36. Nepomnyachiy, S., Ben-Tal, N. & Kolodny, R. CyToStruct: Augmenting the network visualization of CyToStruct with the power of molecular viewers. *Structure* **23**, 941–948, doi:10.1016/j.str.2015.02.013 (2015).
37. Niknam, N., Khakzad, H., Arab, S. S. & Naderi-Manesh, H. PDB2Graph: A toolbox for identifying critical amino acids map in proteins based on graph theory. *Comput. Biol. Med.* **72**, 151–159, doi:10.1016/j.compbiomed.2016.03.012 (2016).
38. Ghosh, A. & Vishveshwara, S. A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. USA* **104**, 15711–6, doi:10.1073/pnas.0704459104 (2007).
39. Karami, Y., Laine, E. & Carbone, A. Dissecting protein architecture with communication blocks and communicating segment pairs. *BMC Bioinformatics* **17**, 13, doi:10.1186/s12859-015-0855-y (2016).
40. Invernizzi, G., Tiberti, M., Lambrughi, M., Lindorff-Larsen, K. & Papaleo, E. Communication Routes in ARID Domains between Distal Residues in Helix 5 and the DNA-Binding Loops. *PLoS Comput. Biol.* **10**, e1003744, doi:10.1371/journal.pcbi.1003744 (2014).
41. Marino, V. & Dell'Orco, D. Allosteric communication pathways routed by Ca2+/Mg2+ exchange in GCAP1 selectively switch target regulation modes. *Sci. Rep.* **6**, 34277, doi:10.1038/srep34277 (2016).
42. Papaleo, E., Renzetti, G. & Tiberti, M. Mechanisms of intramolecular communication in a hyperthermophilic acylaminoacyl peptidase: a molecular dynamics investigation. *PLoS One* **7**, e35686, doi:10.1371/journal.pone.0035686 (2012).
43. Skjærven, L., Yao, X.-Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**, 399, doi:10.1186/s12859-014-0399-6 (2014).
44. Ribeiro, A. A. S. T. & Ortiz, V. MDN: A Web Portal for Network Analysis of Molecular Dynamics Simulations. *Biophys. J.* **109**, 1110–1116, doi:10.1016/j.bpj.2015.06.013 (2015).
45. Ribeiro, A. A. S. T. & Ortiz, V. Energy propagation and network energetic coupling in proteins. *J. Phys. Chem. A* **119**, 1835–1846, doi:10.1021/jp509906m (2015).
46. Ribeiro, A. A. S. T. & Ortiz, V. Determination of signaling pathways in proteins through network theory: Importance of the topology. *J. Chem. Theory Comput.* **10**, 1762–1769, doi:10.1021/ct400977r (2014).
47. Da Silveira, C. H. *et al*. Protein cutoff scanning: A comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. *Proteins Struct. Funct. Bioinforma.* **74**, 727–743, doi:10.1002/prot.v74:3 (2009).
48. Hertig, S., Latorraca, N. R. & Dror, R. O. Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations. *PLoS Comput. Biol.* **12**, 1–16, doi:10.1371/journal.pcbi.1004746 (2016).
49. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* **41**, 429–52, doi:10.1146/annurev-biophys-042910-155245 (2012).
50. Martín-García, F., Papaleo, E., Gomez-Puertas, P., Boomsma, W. & Lindorff-Larsen, K. Comparing Molecular Dynamics Force Fields in the Essential Subspace. *PLoS One* **10**, e0121114, doi:10.1371/journal.pone.0121114 (2015).
51. Papaleo, E., Sutto, L., Gervasio, F. L. & Lindorff-Larsen, K. Conformational Changes and Free Energies in a Proline Isomerase. *J. Chem. Theory Comput.* **10**, 4169–4174, doi:10.1021/ct500536r (2014).
52. Wang, Y., Papaleo, E. & Lindorff-Larsen, K. Mapping transiently formed and sparsely populated conformations on a complex energy landscape. *Elife* **5**, e17505, doi:10.7554/eLife.17505 (2016).
53. Lindorff-Larsen, K., Maragakis, P., Piana, S. & Shaw, D. E. Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J. Phys. Chem. B* acs.jpcb.6b02024, doi:10.1021/acs.jpcb.6b02024 (2016).
54. Nygaard, M. *et al*. The mutational landscape of the oncogenic MZF1 SCAN domain in cancer. *Front. Mol. Biosci*. doi:10.3389/fmolb.2016.00078 (2016).
55. Marino, V., Scholten, A., Koch, K. W. & Dell'Orco, D. Two retinal dystrophy-associated missense mutations in GUCA1A with distinct molecular properties result in a similar aberrant regulation of the retinal guanylate cyclase. *Hum. Mol. Genet.* **24**, 6653–6666, doi:10.1093/hmg/ddv370 (2015).
56. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100**, L47–9, doi:10.1016/j.bpj.2011.03.051 (2011).
57. Best, R. B. *et al*. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273, doi:10.1021/ct300400x (2012).

58. Best, R. B. & Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **113**, 9004–15, doi:10.1021/jp901540t (2009).
59. Lindorff-Larsen, K. *et al*. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–8, doi:10.1002/prot.22711 (2010).
60. Schmidt, C., Beilsten-Edmands, V. & Robinson, C. V. Insights into Eukaryotic Translation Initiation from Mass Spectrometry of Macromolecular Protein Assemblies. *J Mol. Biol* 1–13, doi:10.1016/j.jmb.2015.10.011 (2015).
61. Brinda, K. V. & Vishveshwara, S. A network representation of protein structures: implications for protein stability. *Biophys. J.* **89**, 4159–70, doi:10.1529/biophysj.105.064485 (2005).
62. Kannan, N. & Vishveshwara, S. Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* **292**, 441–64, doi:10.1006/jmbi.1999.3058 (1999).
63. Stetz, G. & Verkhivker, G. M. Probing Allosteric Inhibition Mechanisms of the Hsp70 Chaperone Proteins Using Molecular Dynamics Simulations and Analysis of the Residue Interaction Networks. *J. Chem. Inf. Model.* **56**, 1490–1517, doi:10.1021/acs.jcim.5b00755 (2016).
64. 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.* **9** (2013).
65. Pasi, M., Tiberti, M., Arrigoni, A. & Papaleo, E. xPyder: A PyMOL Plugin To Analyze Coupled Residues and Their Networks in Protein Structures. *J. Chem. Inf. Model.* **279**, 1–6, doi:10.1021/ci300213c (2012).
66. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435–447, doi:10.1021/ct700301q (2008).
67. Tiberti, M., Invernizzi, G. & Papaleo, E. (Dis) similarity Index To Compare Correlated Motions in Molecular Simulations. *J. Chem. Theory Comput.* **11**, 4404–14, doi:10.1021/acs.jctc.5b00512 (2015).
68. Hess, B., Bekker, H., Berendsen, H. & Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* **12**, 1463–1472 (1993).
69. Essmann, U. *et al*. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593, doi:10.1063/1.470117 (1995).
70. Papaleo, E., Pasi, M., Tiberti, M. & De Gioia, L. Molecular dynamics of mesophilic-like mutants of a cold-adapted enzyme: insights into distal effects induced by the mutations. *PLoS One* **6**, e24214, doi:10.1371/journal.pone.0024214 (2011).
71. Seeber, M. *et al*. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* **32**, 1183–94, doi:10.1002/jcc.21688 (2011).
72. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem* **32**, 2319–2327, doi:10.1002/jcc.21787 (2011).
73. Miller, R. G. The jackknife-a review. *Biometrika* **61**, 1–15 (1974).

Acknowledgements

The authors would like to thank Matteo Tiberti and Wouter Boomsma for fruitful discussion and suggestions.

Author Contributions

E.P. conceived and designed the research; J.S.V. and M.F.A. carried out the experiments; E.P., J.S.V. M.L., and M.F.A. discussed the data; J.S.V. and M.F.A prepared the figures and tables, E.P. wrote the manuscript with inputs from all the coauthors.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-01498-6

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.