



## OPEN ACCESS

EDITED BY  
Pietro Hiram Guzzi,  
Magna Græcia University, Italy

REVIEWED BY  
Alfredo Iacoangeli,  
King's College London, United Kingdom  
Andrea Villani,  
Bambino Gesù Children's Hospital  
(IRCCS), Italy

\*CORRESPONDENCE  
Tommaso Mazza,  
t.mazza@css-mendel.it

SPECIALTY SECTION  
This article was submitted to Network  
Bioinformatics,  
a section of the journal  
Frontiers in Bioinformatics

RECEIVED 15 September 2022  
ACCEPTED 05 October 2022  
PUBLISHED 19 October 2022

CITATION  
Petrizzelli F, Biagini T, Bianco SD,  
Liorni N, Napoli A, Castellana S and  
Mazza T (2022), Connecting the dots: A  
practical evaluation of web-tools for  
describing protein dynamics  
as networks.  
*Front. Bioinform. 2*:1045368.  
doi: 10.3389/fbinf.2022.1045368

COPYRIGHT  
© 2022 Petrizzelli, Biagini, Bianco,  
Liorni, Napoli, Castellana and Mazza.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Connecting the dots: A practical evaluation of web-tools for describing protein dynamics as networks

Francesco Petrizzelli<sup>1</sup>, Tommaso Biagini<sup>1</sup>,  
Salvatore Daniele Bianco<sup>1,2</sup>, Niccolò Liorni<sup>1,2</sup>,  
Alessandro Napoli<sup>1</sup>, Stefano Castellana<sup>1</sup> and Tommaso Mazza<sup>1\*</sup>

<sup>1</sup>Bioinformatics Laboratory, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, <sup>2</sup>Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy

Protein Structure Networks (PSNs) are a well-known mathematical model for estimation and analysis of the three-dimensional protein structure. Investigating the topological architecture of PSNs may help identify the crucial amino acid residues for protein stability and protein-protein interactions, as well as deduce any possible mutational effects. But because proteins go through conformational changes to give rise to essential biological functions, this has to be done dynamically over time. The most effective method to describe protein dynamics is molecular dynamics simulation, with the most popular software programs for manipulating simulations to infer interaction networks being RING, MD-TASK, and NAPS. Here, we compare the computational approaches used by these three tools—all of which are accessible as web servers—to understand the pathogenicity of missense mutations and talk about their potential applications as well as their advantages and disadvantages.

## KEYWORDS

protein structure networks, molecular dynamics simulations, network analysis, graph theory, dynamic residue interaction networks

## Introduction

Molecular dynamics (MD) simulation is one of the most effective methods for assessing how a system changes over time and is regarded as the most effective computing tool for this purpose, particularly in physics and chemistry (Karplus and McCammon 2002). It has also had great success in biology over the past 10 years, where it has frequently been employed to test hypotheses and, as a result, aid in providing specific answers to questions about the structural characteristics and dynamical mechanisms of biological systems (Biagini et al., 2017), like the impact of disease mutations on the protein functionality. Nowadays, we are dealing with a dramatically increased quantity and quality of simulation data due to improvements in hardware (Biagini et al., 2019) and software, as well as more frequent use of enhanced sampling techniques. This comes along

with the need for new and powerful analysis tools, capable of not only extracting information but also capturing key-properties fundamental to large-scale conformational changes (Melo et al., 2020).

One of these tools is the Protein Structure Network (PSN) (Greene 2012), which can model the spatial organization of proteins and record long-distance structural communications. In this model, nodes stand in for amino acids and are connected by edges, which can either represent the physical interactions between two residues or their spatial separation (Residue Interaction Network, RIN, or Protein Contact Network, PCN). The benefit of employing such a “nodes-edges” representation is that it makes it possible to resort to Graph Theory to analyze MD simulation results (del Sol et al., 2006), with the exploitation of a broad range of local (Borgatti 2005; Borgatti and Everett 2006; Mazza et al., 2012) (i.e., regarding nodes or edges) and global (Lozares et al., 2015; Mazza et al., 2010; Menniti et al., 2013) (i.e., regarding the entire network) network metrics to identify topologically important hubs, e.g., nodes fundamental for graph connectivity (betweenness) or identify residues located in functional regions (closeness). This aspect is brilliantly presented in (Liang et al., 2020), where the authors provided a thorough overview of protein network approaches and a description of the tools that are currently available for converting protein structures into graphs, with the aim of providing a new level of insight into seemingly unpredictable systems (Barabasi and Albert 1999). It has a history of successful applications: from the graph spectral methods used by (Kannan and Vishveshwara 1999) to identify side-chain clusters to the characterization of more complex molecular mechanisms (Karamzadeh et al., 2017), with important acknowledgements in drug design (Brown and Bishop 2017) and in the evaluation and prediction of disease mutations (Cheng et al., 2008; Doshi et al., 2016).

## How to turn a molecular dynamics trajectory into a network

The ability to build a network from conformational ensembles, such as snapshots taken from MD simulations, is crucial for accounting for those links that form or break as time progresses. Starting from a trajectory file, which is a collection of 3D coordinates of a protein structure in each of the various conformations investigated during the simulation time, we can either analyze the network properties of individual protein structures corresponding to each trajectory frame or work on the average structure derived over the course of the MD simulation to account for these atomic fluctuations.

Recent years have witnessed the development of several tools that integrate MD simulation data. The majority of these tools build simple unweighted PCNs using a geometry-based methodology, which consists of defining the contacts between

alpha/beta carbon pairs ( $C_{\alpha}$ ,  $C_{\beta}$ ) or between centroids of the amino acids of a protein. Contacts are established if such elements are within a predefined cut-off distance. This distance threshold, which typically ranges between 4.5 and 8.5 Å (Viloria et al., 2017), was carefully selected to map connections only for non-covalent intramolecular interactions, avoiding networks that are either poorly or excessively connected. When these networks are obtained from MD data, they are referred to as Dynamical Network Models (DNMs) or Dynamic Residue Interaction Networks (DRINs), and two nodes are connected only if their distance is less than a cut-off value in the range reported above for at least 65% of the simulation time. The advantage of creating these dynamic networks is that their properties, like dynamic residue-residue cross-correlations or their interaction frequencies, could be assigned as weights to the edges, providing a more accurate description of the system (Sethi et al., 2009).

Here, we evaluate and contrast a few tools made to summarize MD trajectories in a network, providing our general opinion on the benefits and drawbacks of each strategy. Many of these tools are standalone software packages (Felline et al., 2022), usually easy to use but requiring some technical knowledge to install and configure. In order to test their adaptability and usability (Table 1) for the specific task of assisting in the interpretation of the role of missense mutations in conformational changes, we only kept those that are available as web servers, i.e., NAPS, MD-TASK, and RING. This was done because one of the goals of this mini-review was to support widespread usage of this class of tools.

## Network based analysis of protein structures

Network based Analysis of Protein Structures (NAPS) is an online tool available at <https://bioinf.iit.ac.in/NAPS/>. It was originally built for the analysis and interactive visualization of PCN or RIN derived from static single proteins or protein complexes (Chakrabarty and Parekh 2016). Its key characteristic is the creation of various network types, both unweighted and weighted, from a single PDB, using  $C_{\alpha}$ ,  $C_{\beta}$ , geometric center, or center of mass distances to draw edges between residues. These features were greatly expanded in NAPS 2.0 (Chakrabarty et al., 2019) with the possibility to analyze MD trajectories, exported as .dcd files, and represent them as average networks of the ensemble of trajectories, dynamic cross-correlations (DCC), and bipartite networks. The web tool offers the analysis of various centrality measures (with distance-based weights or unweighted) computed using igraph (Csardi and Nepusz 2006) and NetworkX (Hagberg et al., 2008): *degree*, *average shortest path*, *closeness* and *betweenness*, *clustering coefficient*, *eigenvector centrality*, *eccentricity*, *average nearest neighbor degree* (ANN degree), and *edge betweenness*, in

TABLE 1 Web-tools for analyzing MD trajectories as networks. For each tool, we summarize the required input formats, different topologies that can be used to build a network, and which network centrality indices are computed. Finally, the main strengths of each tool are highlighted.

Tools	Url	Standalone	Trajectory requirements	Main strength
	<i>Network topology</i>		<i>Output - network centralities</i>	
NAPS	<a href="https://bioinf.iit.ac.in/NAPS/">https://bioinf.iit.ac.in/NAPS/</a> <i>C<sub>α</sub>, C<sub>β</sub>, geometric center and center of mass distances*, energy</i>	No	DCD format max 50 frames <i>Weighted/unweighted degree, shortest path, closeness, betweenness, clustering coefficient, eigenvector centrality, eccentricity, average nearest neighbor degree and edge betweenness</i>	Multiple network types
MD_TASK	<a href="https://mdtaskweb.rubi.ru.ac.za/">https://mdtaskweb.rubi.ru.ac.za/</a> <i>C<sub>β</sub> (C<sub>α</sub> for Glycine) distances*</i>	Yes	Multiple MD formats max 250 Mb <i>Betweenness, degree, eccentricity, averaged shortest paths, closeness, Katz, PageRank and eigencentality</i>	Great MD format support
RING3.0	<a href="https://ring.biocomputingup.it/">https://ring.biocomputingup.it/</a> <i>Closest atoms, lollipop (Center of mass), C<sub>α</sub> or C<sub>β</sub> of residues connected by multiple interaction type</i>	Yes	PDB and mmCIF format max 200 Mb <i>Cytoscape-compatible (.json) output format</i>	Probabilistic residue interaction networks

\*distances  $\leq$  user defined threshold (authors usually suggest 6.5–7.5 Å to avoid disconnected/highly connected graph)

addition to standard global properties, i.e. *number of nodes and edges, diameter, clustering coefficient, average degree, and average path length*. These can be computed for the network representing a specific simulation frame or for an “average network”, i.e., an individual network wiring nodes with edges only if the interactions they represent last more than 60% (default value) of the entire simulation time. This latter option is available by choosing “Ensemble Analysis” on the submission form, whereas “Timestep Analysis” allows users to compute and compare the centrality values pertaining to a particular time step to either those of other time steps or to those of the average network. Similarly, the shortest path can be obtained and compared for any pair of residues.

While we considered NAPS to be one of the most comprehensive web tools currently available in terms of options for network construction and analysis, we also ran into a number of usability issues, mostly pertaining to the MD section. In particular, the user is forced to limit the trajectory to 50 frames even though the web tool is intended to handle large MD trajectory files. Additionally, there are a few minor bugs in the web server that restrict user experience. We observe, in fact, numerous glitches when structures with more than 1000 residues are submitted, along with errors when ligands or heteroatoms are present in the input topology. The introduction of a standalone version and a more potent web-server would both be highly advantageous.

## MD-TASK

MD-TASK is a Python-based suite that can generate residue interaction networks from MD simulations and uses NetworkX

functions to compute average residue network centrality metrics. It is available as a downloadable program as well as a web server (<https://mdtaskweb.rubi.ru.ac.za/>), where it is integrated with the MODE-TASK suite, a normal mode and essential dynamics analysis toolkit (Brown et al., 2017; Amamuddy et al., 2021). Remarkably, MD-TASK supports many different MD file formats, including the most commonly used by AMBER (.netcdf), NAMD (.dcd) and Gromacs (.xtc) simulation frameworks. DRINs can be constructed using these trajectory files, with single residues as nodes and edges drawn when the distance between C<sub>β</sub> atoms (C<sub>α</sub> for Glycine) of two residues meets a user-defined cut-off (usually 6.5–7.5 Å).

It is noteworthy that the metrics are not computed on average networks, but rather the software returns changes in eight different centrality metrics to residues over a trajectory, which are obtained by aggregating the mean, median, and standard deviation of each frame’s residue metrics. The values that are obtained can either be displayed in the 3D structure with a color gradient or included in a downloadable csv file. MD-TASK also enables the construction of a weighted residue contact map from a trajectory, which is a weighted network graph with edges between a residue of interest and the other residues that are weighted in accordance with how frequently the interactions occur. The output is another network centered on the residue of interest and surrounded by the residues it interacts with. Finally, besides DRIN, the MD-TASK tool suite also deals with DCC and perturbation response scanning (PRS) techniques (Atilgan and Atilgan 2009).

When it comes to usability, we found MD-TASK to be fairly simple, with an easy-to-use submission form, a straightforward but sharp output visualization, and the option of direct comparison with previously submitted jobs. The developers

have provided a simple tool (<https://github.com/oliserand/MD-TASK-prep>) that makes it possible to reduce the trajectory size by keeping only heavy atoms, greatly increasing the number of frames that can be submitted. The 250 Mb limit for the trajectory size is undoubtedly a limitation, but it is partially overcome by this tool. The MD-TASK results page, on the other hand, is accompanied by a color legend, which helps identify significant regions or domains. However, the data interpretation remains difficult without prior knowledge of the altered molecular mechanisms.

## Residue interaction network generator

Residue Interaction Network Generator (RING) offers a simple way to build a network starting from an MD trajectory. The edges connecting the nodes are atom-specific physico-chemical interactions, such as disulfides, salt bridges, hydrogen bonds, aromatic interactions, or more general van-der-Waals contacts between residues. These interactions also rely on the concept of “distance”, computed using only geometrical criteria and after an exhaustive analysis of the entire PDB content (116568 X-ray and NMR structures as of April 2016). To represent strict and permissive parameters, two different distance thresholds were selected; the pair [2.84, 2.87] Å, for example, corresponds to interactions that stabilize the packing of different secondary structure elements, i.e., bridges between alpha-helices or turns; similar values for the main chain hydrogen bonds at [2.94, 2.98] correspond to interactions between adjacent strands in  $\beta$ -sheets, whereas [5.01, 5.04] Å identifies bonds in  $\alpha$ -helices separated by a turn. Beyond 5.6 Å, only spurious interactions are identified (Piovesan et al., 2016).

The most recent RING 3.0 version (Damiano et al., 2022) is available as a standalone package or in a completely redesigned web-portal at <https://ring.biocomputingup.it/>. It can process molecular dynamics simulations as multi-state files (in PDB and mmCIF format, up to 200 MB in size). Users can choose between four alternative types of network, and interactions involving the same residue but different atoms are sorted by energy and distance. In this case, the user may choose to retrieve only the most energetic interaction, a multigraph with all the interactions, or only one interaction for each type. For each node, a number of structural attributes are reported, including secondary structure, vertex degree (the number of directly connected nodes), experimental uncertainty for X-ray structures, conformational energy preferences, conservation (Shannon entropy), and cumulative mutual information (MI). From these, RING now creates probabilistic networks that take into account the frequency of connectivity between states (or snapshots): the edges have an associated weight (range: [0–1]) that represents the frequency at which the interaction was present in the conformational ensemble. Finally, RING

returns an interactive graph and a network that can be downloaded in a format, that is, optimized for Cytoscape. This dynamic layout makes it possible to quickly and effectively identify functional residues.

We consider RING to be by far the most interesting approach, encompassing different levels of granularity in the network construction. The output page, on the other hand, focuses primarily on the identification and description of significant interactions and does not include any analysis strategy based on network topological metrics that could aid the user in the localization of hotspots. Finally, the PDB format, which was probably chosen to speed up computation, poses a significant limitation due to its large size, thereby forcing the user to rely on the standalone version even for small trajectories.

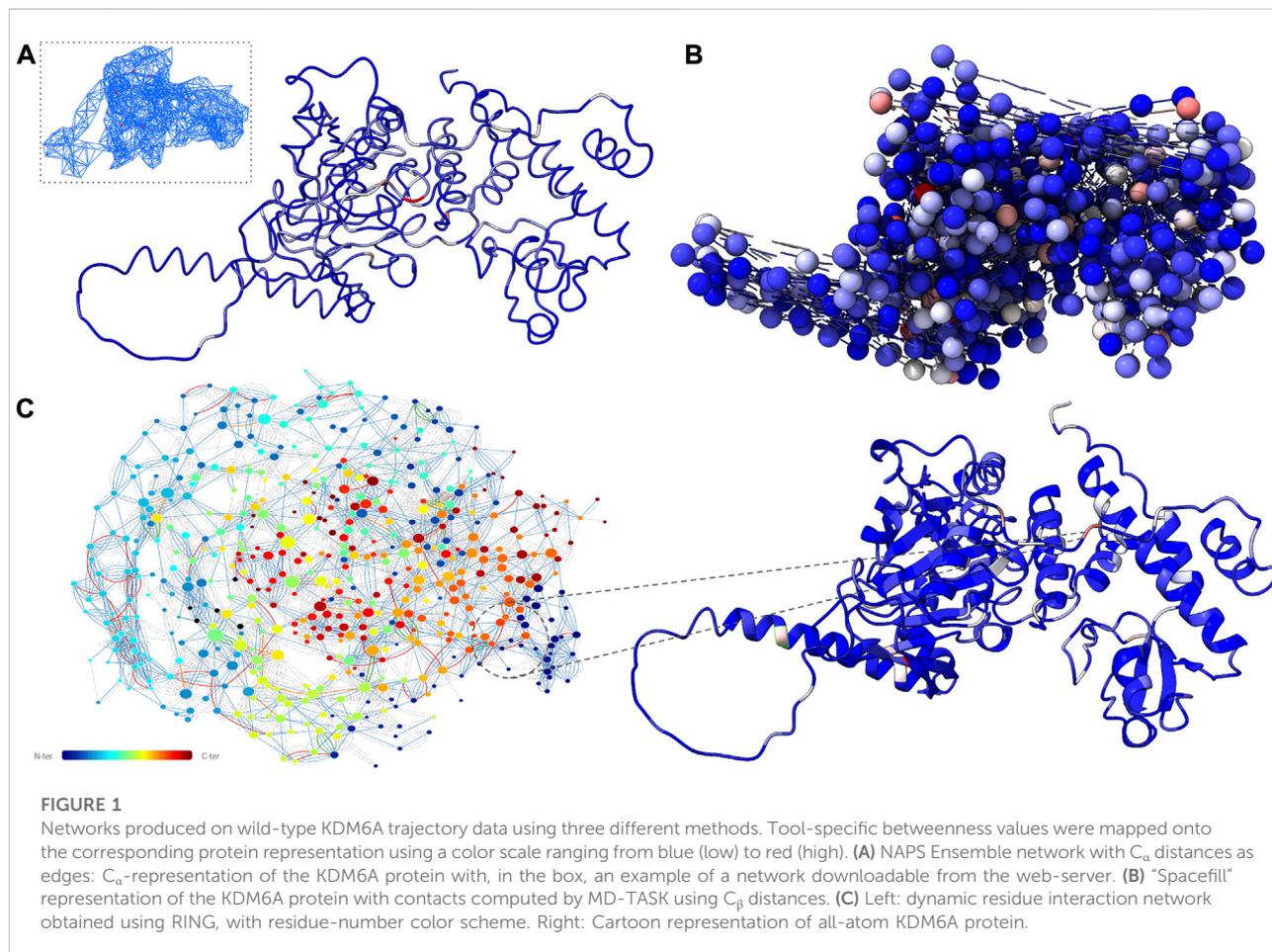
## How to use molecular dynamics-networks to understand the role of missense mutations?

In the last decade, molecular dynamics has become one of the most widely used computational approaches for the generation of hypotheses regarding the impact of missense mutations. From a network point of view, when these perturbations affect the most central nodes, we may observe a disruption in the transmission of information, with consequences for protein stability and the alteration of fundamental protein functions. There are several methods that can be used to investigate this aspect, such as PRS (Atilgan and Atilgan 2009) or comparing centrality metrics gleaned from various trajectory comparisons.

All of the tools mentioned above have been successfully used to achieve this goal. However, a performance-based evaluation poses conceptual difficulties because each approach might take into account different altered mechanisms (such as local impact and long-range effects) and, as a result, a direct comparison is unreliable or even impossible. Moreover, given the fact that fluctuation correlation is very hard to converge in molecular dynamics simulations (Hospital et al., 2015), it is often the case that the dynamical network is trajectory dependent or simulation time dependent, leading to inconclusive statements. (Li et al., 2019).

Here, to overcome these issues, we focused on the description of the ability of each strategy to identify key nodes and capture the impact caused by pathogenic mutations using already extensively described MD trajectories. We specifically used the wild-type and mutant trajectories of the catalytic Jumonji (JmjC) domain of KDM6A, a known cancer driver gene and the gene responsible for type 2 Kabuki Syndrome, to show how a particular set of missense mutations can be connected to the impairment of the interaction between the protein and the interacting protein (Petrizzelli et al., 2020; Biagini et al., 2022).

In [Supplementary Table S1](#), we report all the centrality values computed by each tool for the wild-type (WT) and one of the



most detrimental mutations, Arg1255Trp (R1255W). Default options were selected: NAPS "Ensemble Analysis" on  $C_{\alpha}$  distances was executed on 50 frames extracted from each trajectory; on the contrary, MD-TASK mean centralities were computed on  $C_{\beta}$  distances of 1000 frames obtained by taking advantage of the "trajectory-cutting" MD-TASK-prep tool. Finally, since RING does not offer metric analysis, we extracted the dynamic interaction network computed on 300 frames and used RINalyzer (Doncheva et al., 2011) of Cytoscape to compute the shortest path betweenness, closeness, and degree centralities using frequency as edge weight. Then, we examined the top 15% nodes for each metric, using a metric-specific color scale, with a special emphasis on the betweenness centrality (BC), the most widely used measure of centrality. In fact, betweenness measures how many shortest paths between any two nodes of a network pass through a given node given the total number of shortest paths in the network and, then, furnishes a measure of the overall importance of the given node.

Comparing the BC values of the missense variants located within the Jumonji domain (Supplementary Table S1 - DiseaseMutation),

we found that NAPS and MD-TASK identify as central some of these well-known pathogenic hotspots, which is consistent with the enrichment of central residues in the protein's core shown in Figure 1A,B, while none of these amino acidic positions but 1049 can be considered central in the interaction network obtained using RING. However, the poor correlation observed between pathogenic sites and central nodes could be mainly ascribed to the fact that their central role in structural rearrangements is not enough to determine the impact of a mutation on that site and supports the need for further attributes that account for the functional role of each residue, like evolutionary information (Liang et al., 2020).

On the other hand, we found that many wild-type top residues for all the generated networks had decreased betweenness when comparing the 100 top nodes between wild-type and R1255W trajectories. With a significant reduction in this region's betweenness across all the R1255W metrics (Supplementary Table S1 - RING), RING, in particular, was able to highlight the linker domain's central function more than the other tools (Figure 1C). This is consistent with our earlier observations that the alteration of the wild-type

conformational transitions was correlated with the loss of fundamental hydrogen bonds between the linker domain and its surrounding region.

In conclusion, each tool retains its unique characteristics, with networks produced by NAPS and MD-TASK able to prioritize those residues essential for the functionality of the protein as discernible from the enrichment of mutant hotspots in their central residues, and networks produced by RING able to identify the specific altered mechanism between wild-type and mutant dynamics.

## Final thoughts and future plans

*“Different packages can have different niche strengths, and their strengths are often complementary.”* - M. Hucka.

Creating network-based methods to model molecular systems is a rich area of study that capitalizes on several research fields, mainly transcriptomics, proteomics, and ecology, as well as in computer science applied to biological sciences (Mazza et al., 2016; Piepoli et al., 2012; Mazzoccoli et al., 2016; Palmieri et al., 2020; Mazza et al., 2017; Capocéfalo et al., 2018; Ballarini et al., 2009; Franco et al., 2006). However, these methods become especially fascinating when they are made to handle the massive amounts of data produced by enhanced or long-timescale MD simulations where, thanks to their vast potential, could help answer a variety of different questions, e.g., putative allosteric mechanism or protein-ligand interaction pathway, or also could be useful for challenging tasks like the identification of epistatic mutant sites (Castellana et al., 2015; Castellana et al., 2017;2021).

An ideal “MD to network” tool should be able to work with a variety of input formats, build trustworthy networks, analyze those networks’ topologies, and finally, provide simple visualization and interpretation of the results (Guzzi et al., 2022). Here, we provide a practical evaluation of the usability and applicability of three of the most popular web tools, NAPS, MD-TASK, and RING, to construct and analyze MD-based graphs. First of all, we would like to acknowledge the tremendous effort that has gone into developing and maintaining these web servers. We found them easy to use, even for beginners. But all of them have minor flaws, primarily because of the restrictions on input formats and size. When available, their standalone versions perform significantly better, but they cannot be proficiently used without the necessary programming skills and computing power. In addition, each software package exhibits a unique characteristic in the network construction, with varying degrees of granularity and specialization, and it offers the computation of the most common centrality metrics. For the identification of functionally significant residues, they all heavily rely on network/protein visualization, with many centrality metrics

that can be either highlighted on the structure or downloaded as a file in both MD-TASK and NAPS.

Finally, focusing on applicability, we discovered that each software was able to identify “central” residues in functional domains with varying sensitivity and that the DRINs obtained by RING were the ones that best captured the disruption of fundamental interactions during the mutant trajectory and the impact on the dynamics of harmful mutations. However, it can be challenging to determine which metric, or set of metrics, might be the most sensitive in describing the investigated functional mechanism, and we discovered a general lack of statistical consistency, with little that has been proposed to serve as a guide for selecting metrics or as evidence for comparing them (Foutch et al., 2021). For instance, application of comparative evaluation with synthetic networks may aid in assessing the strength of the community structure by determining the informational value of a metric or when the combination of multiple metrics would be advantageous (Oldham et al., 2019; Rajeh et al., 2021).

In conclusion, we believe that integrating various approaches is essential to more effectively exploring the information contained in the topology of an MD-based network. Additionally, the integration of conservation and evolutionary attributes could significantly enhance this information, and the use of more sophisticated metrics, such as key-player and group-centrality metrics, could support the agnostic identification of central “hubs,” or structured/regulatory regions (Dassi et al., 2012) necessary for the functionality of the protein, that could previously only be identified thanks to prior knowledge obtained from experimental or literature studies (Parca et al., 2020).

## Author contributions

FP; Conceptualization, design, writing—original draft. TB; Conceptualization, writing—review and editing; editing. SB; Methodology. NL; Investigation. AN; Investigation. SC; Investigation. TM; Writing—review and editing; editing, Conceptualization, Project administration.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Amamuddy, O. S., Glenister, M., Tshabalala, T., and Bishop, O. T. (2021). MDM-TASK-Web: MD-TASK and MODE-TASK web server for analyzing protein dynamics. *Comput. Struct. Biotechnol. J.* 19, 5059–5071. doi:10.1016/j.csbj.2021.08.043
- Atilgan, C., and Atilgan, A. R. (2009). Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput. Biol.* 5 (10), e1000544. doi:10.1371/journal.pcbi.1000544
- Ballarini, P., Forlin, M., Mazza, T., and Prandi, D. (2009). Efficient parallel structural model checking of biochemical networks. *EPTCS 14* 2009, 47–61. doi:10.48550/arXiv.0912.2551
- Barabasi, A.-L., and Albert, R. (1999). Emergence of scaling in random networks. *Science* 286 (5439), 509–512. doi:10.1126/science.286.5439.509
- Biagini, T., Chillemi, G., Mazzoccoli, G., Grottesi, A., Fusilli, C., Capocéfalo, D., et al. (2017). Molecular dynamics recipes for genome research. *Briefings Bioinforma.* 19 (5), 853–862. doi:10.1093/bib/bbx006
- Biagini, T., Petritzelli, F., Daniele Bianco, S., Liorni, N., Napoli, A., Castellana, S., et al. (2022). KDM6A missense variants hamper H3 histone demethylation in lung squamous cell carcinoma. *Comput. Struct. Biotechnol. J.* 20 (20), 3151–3160. doi:10.1016/j.csbj.2022.06.041
- Biagini, T., Petritzelli, F., Truglio, M., Cespa, R., Barbieri, A., Capocéfalo, D., et al. (2019). Are gaming-enabled graphic processing unit cards convenient for molecular dynamics simulation. *Evol. Bioinform. Online.* 15, 117693431985014. doi:10.1177/1176934319850144
- Borgatti, S. P. (2005). Centrality and network flow. *Soc. Netw.* 27 (1), 55–71. doi:10.1016/j.socnet.2004.11.008
- Borgatti, S. P., and Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Soc. Netw.* 28 (4), 466–484. doi:10.1016/j.socnet.2005.11.005
- Brown, D. K., and Bishop, Ö. T. (2017). Role of structural Bioinformatics in drug discovery by computational SNP analysis. *Glob. Heart* 12 (2), 151–161. doi:10.1016/j.ghheart.2017.01.009
- Brown, D. K., Penkler, D. L., Sheik Amamuddy, O., Ross, C., Atilgan, A. R., Atilgan, C., et al. (2017). MD-TASK: A software suite for analyzing molecular dynamics trajectories. *Bioinformatics* 33 (17), 2768–2771. doi:10.1093/bioinformatics/btx349
- Capocéfalo, D., Pereira, J., Mazza, T., and Jordán, F. (2018). Food web topology and nested keystone species complexes. *Complexity* 2018, 1–8. doi:10.1155/2018/1979214
- Castellana, S., Fusilli, C., Mazzoccoli, G., Biagini, T., Capocéfalo, D., Carella, M., et al. (2017). High-confidence assessment of functional impact of human mitochondrial non-synonymous genome variations by APOGEE. *PLoS Comput. Biol.* 13 (6), e1005628. doi:10.1371/journal.pcbi.1005628
- Castellana, S., Rónai, J., and Mazza, T. (2015). MitImpact: An exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants. *Hum. Mutat.* 36 (2), E2413–E2422. doi:10.1002/humu.22720
- Castellana, S., Tommaso, B., Francesco, P., Luca, P., Noemi, P., Viviana, C., et al. (2021). MitImpact 3: Modeling the residue interaction network of the respiratory chain subunits. *Nucleic Acids Res.* 49 (D1), 1282–1288. doi:10.1093/nar/gkaa1032
- Chakrabarty, B., Naganathan, V., Garg, K., Agarwal, Y., and Parekh, N. (2019). NAPS update: Network analysis of molecular dynamics data and protein-nucleic acid complexes. *Nucleic Acids Res.* 47 (W1), W462–W470. doi:10.1093/nar/gkz399
- Chakrabarty, B., and Parekh, N. (2016). Naps: Network analysis of protein structures. *Nucleic Acids Res.* 44 (W1), W375–W382. doi:10.1093/nar/gkw383
- Cheng, T. M. K., Lu, Y.-E., Vendruscolo, M., Lio, P., and Blundell, T. L. (2008). Prediction by graph theoretic measures of structural effects in proteins arising from non-synonymous single nucleotide polymorphisms. *PLoS Comput. Biol.* 4 (7), e1000135. doi:10.1371/journal.pcbi.1000135
- Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Inter. J. Complex Syst.* 1695, 1–9. <https://igraph.org>.
- Damiano, C., Del Conte, A., Monzon, A. M., Camagni, G. F., Minervini, G., Piovesan, D., et al. (2022). Ring 3.0: Fast generation of probabilistic residue interaction networks from structural ensembles. *Nucleic Acids Res.* 50 (W1), W651–W656. doi:10.1093/nar/gkac365
- Dassi, E., Malossini, A., Re, A., Mazza, T., Tebaldi, T., Caputi, L., et al. (2012). Aura: Atlas of UTR regulatory activity. *Bioinformatics* 28 (1), 142–144. doi:10.1093/bioinformatics/btr608
- del Sol, A., Fujihashi, H., Amoros, D., and Nussinov, R. (2006). Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families. *Protein Sci.* 15 (9), 2120–2128. doi:10.1110/ps.062249106
- Doncheva, N. T., Klein, K., Domingues, F. S., and Albrecht, M. (2011). Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* 36 (4), 179–182. doi:10.1016/j.tibs.2011.01.002
- Doshi, U., Holliday, M. J., Eisenmesser, E. Z., and Hamelberg, D. (2016). Dynamical network of residue–residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci. U. S. A.* 113, 4735–4740. doi:10.1073/pnas.1523573113
- Felline, A., Seeber, M., and Fanelli, F. (2022). PSNtools for standalone and web-based structure network analyses of conformational ensembles. *Comput. Struct. Biotechnol. J.* 20 (20), 640–649. doi:10.1016/j.csbj.2021.12.044
- Foutch, D., Pham, B., and Shen, T. (2021). Protein conformational switch discerned via network centrality properties. *Comput. Struct. Biotechnol. J.* 19 (19), 3599–3608. doi:10.1016/j.csbj.2021.06.004
- Franco, G., Guzzi, P. H., Manca, V., and Mazza, T. (2006). “Mitotic oscillators as MP graphs,” in *Membrane computing lecture notes in computer science* (Berlin, Heidelberg: Springer Berlin Heidelberg), 382–394.
- Greene, L. H. (2012). Protein structure networks. *Briefings Funct. Genomics* 11 (6), 469–478. doi:10.1093/bfpg/els039
- Guzzi, P. H., Di Paola, L., Giuliani, A., and Veltri, P. (2022). PCN-Miner: An open-source extensible tool for the analysis of protein contact networks. *Bioinformatics* 38, 4235–4237. doi:10.1093/bioinformatics/btac450
- Hagberg, A., Swart, P. J., and Chult, D. S. (2008). “Exploring network structure, dynamics, and function using NetworkX,” in *Proceedings of the 7th Python in Science Conference: SCIPY 08, Pasadena, August 21, 2008*.
- Hospital, A., Goñi, J. R., Orozco, M., and Gelpi, J. L. (2015). Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinform. Chem.* 8, 37–47. doi:10.2147/aabc.s70333
- Kannan, N., and Vishveshwara, S. (1999). Identification of side-chain clusters in protein structures by a graph spectral method 1 Edited by J. M. Thornton. *J. Mol. Biol.* 292 (2), 441–464. doi:10.1006/jmbi.1999.3058
- Karamzadeh, R., Karimi-Jafari, M. H., Sharifi-Zarchi, A., Chitsaz, H., Salekdeh, G. H., and Moosavi-Movahedi, A. A. (2017). Machine learning and network analysis of molecular dynamics trajectories reveal two chains of red/ox-specific residue interactions in human protein disulfide isomerase. *Sci. Rep.* 7 (1), 3666. doi:10.1038/s41598-017-03966-5
- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi:10.1038/nsb0902-646
- Li, Q., Luo, R., and Chen, H.-F. (2019). Dynamical important residue network (DIRN): Network inference via conformational change. *Bioinformatics* 35 (22), 4664–4670. doi:10.1093/bioinformatics/btz298
- Liang, Z., Verkhivker, G. M., and Hu, G. (2020). Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: Theory, tools and applications. *Briefings Bioinforma.* 21 (3), 815–835. doi:10.1093/bib/bbz029
- Lozares, C., López-Roldán, P., Bolibar, M., and Muntanyola, D. (2015). The structure of global centrality measures. *Int. J. Soc. Res. Methodol.* 18 (2), 209–226. doi:10.1080/13645579.2014.888238
- Mazza, T., Ballarini, P., Guido, R., and Prandi, D. (2012). The relevance of topology in parallel simulation of biological networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 9 (3), 911–923. doi:10.1109/tcbb.2012.27

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2022.1045368/full#supplementary-material>.

- Mazza, T., Copetti, M., Capocefalo, D., Fusilli, C., Biagini, T., Carella, M., et al. (2017). MicroRNA Co-expression networks exhibit increased complexity in pancreatic ductal compared to vater's papilla adenocarcinoma. *Oncotarget* 8 (62), 105320–105339. doi:10.18632/oncotarget.22184
- Mazza, T., Mazzoccoli, G., Fusilli, C., Capocefalo, D., Panza, A., Biagini, T., et al. (2016). Multifaceted enrichment analysis of RNA-RNA crosstalk reveals cooperating micro-societies in human colorectal cancer. *Nucleic Acids Res.* 44 (9), 4025–4036. doi:10.1093/nar/gkw245
- Mazza, T., Romanel, A., and Jordán, F. (2010). Estimating the divisibility of complex biological networks by sparseness indices. *Briefings Bioinforma.* 11 (3), 364–374. doi:10.1093/bib/bbp060
- Mazzoccoli, G., Colangelo, T., Panza, A., Rubino, R., Tiberio, C., Palumbo, O., et al. (2016). Analysis of clock gene-miRNA correlation networks reveals candidate drivers in colorectal cancer. *Oncotarget* 7 (29), 45444–45461. doi:10.18632/oncotarget.9989
- Melo, M. C. R., Bernardi, R. C., de la Fuente-Nunez, C., and Luthey-Schulten, Z. (2020). Generalized correlation-based dynamical network analysis: A new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *J. Chem. Phys.* 153 (13), 134104. doi:10.1063/5.0018980
- Menniti, S., Castagna, E., and Mazza, T. (2013). Estimating the global density of graphs by a sparseness index. *Appl. Math. Comput.* 224, 346–357. doi:10.1016/j.amc.2013.08.040
- Oldham, S., Fulcher, B., Parkes, L., Arnatkeviciūtė, A., Suo, C., and Fornito, A. (2019). “Consistency and differences between centrality measures across distinct classes of networks. *PLoS One* 14 (7), e0220061. doi:10.1371/journal.pone.0220061
- Palmieri, O., Mazza, T., Bassotti, G., Merla, A., Tolone, S., Biagini, T., et al. (2020). microRNA-mRNA network model in patients with achalasia. *Neurogastroenterol. Motil.* 32 (3), e13764. doi:10.1111/nmo.13764
- Parca, L., Truglio, M., Biagini, T., Castellana, S., Pettrizzelli, F., Capocefalo, D., et al. (2020). Pyntacle: A parallel computing-enabled framework for large-scale network biology analysis. *GigaScience* 9 (10), g1aa115. doi:10.1093/gigascience/g1aa115
- Petrizzelli, F., Biagini, T., Barbieri, A., Parca, L., Panzironi, N., Castellana, S., et al. (2020). Mechanisms of pathogenesis of missense mutations on the KDM6A-H3 interaction in type 2 Kabuki Syndrome. *Comput. Struct. Biotechnol. J.* 18 (18), 2033–2042. doi:10.1016/j.csbj.2020.07.013
- Piepoli, A., Tavano, F., Copetti, M., Mazza, T., Palumbo, O., Panza, A., et al. (2012). MiRNA expression profiles identify drivers in colorectal and pancreatic cancers. *PLoS One* 7 (3), e33663. doi:10.1371/journal.pone.0033663
- Piovesan, D., Minervini, G., and Tosatto, S. C. E. (2016). The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* 44 (W1), W367–W374. doi:10.1093/nar/gkw315
- Rajeh, S., Savonnet, M., Leclercq, E., and Cherifi, H. (2021). Characterizing the interactions between classical and community-aware centrality measures in complex networks. *Sci. Rep.* 11 (1), 1–15. doi:10.1038/s41598-021-89549-x
- Sethi, A., Eargle, J., Black, A. A., and Luthey-Schulten, Z. (2009). Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* 106 (16), 6620–6625. doi:10.1073/pnas.0810961106
- Viloria, J. S., Allega, M. F., Lambrugh, M., and Papaleo, E. (2017). An optimal distance cutoff for contact-based protein structure networks using side-chain centers of mass. *Sci. Rep.* 7 (1), 1–11. doi:10.1038/s41598-017-01498-6