RIP-MD: A tool to study residue interaction networks in protein molecular dynamics

Sebastián Contreras-Riquelme [1, 2, 3] , **Jose-Antonio Garate** [4] , **Tomas Perez-Acle** [Corresp., 1, 4] , **Alberto J.M. Martin** [Corresp. 3]

1 Computational Biology Laboratory (DLab), Fundacion Ciencia & Vida, Santiago, Chile

2 Facultad de Ciencias Biológicas, Universidad Andrés Bello, Santiago, Chile

3 Network Biology Laboratory, Centro de Genómica y Bioinformática, Facultad de Ciencias, Universidad Mayor, Chile

4 Centro Interdisciplinario de Neurociencia de Valparaíso, Valparaíso, Chile

Corresponding Authors: Tomas Perez-Acle, Alberto J.M. Martin
Email address: tomas@dlab.cl, proteinomano@gmail.com

Protein structure is not static; residues undergo conformational rearrangements and, in doing so, create, stabilize or break non-covalent interactions. Molecular Dynamics (MD) is a technique used to simulate these movements with atomic resolution. However, given the data-intensive nature of the technique, gathering relevant information from MD simulations is a complex and time consuming process requiring several computational tools to perform these analyses. Among different approaches, the study of Residue Interaction Networks (RINs) has proven to facilitate the study of protein structures. In a RIN, nodes represent amino-acid residues and the connections between them depict non-covalent interactions.

Here, we describe RIP-MD, a Visual Molecular Dynamics (VMD) plugin to facilitate the study of RINs using trajectories obtained from MD simulations of proteins. Our software generates RINs from MD trajectory files. The noncovalent interactions defined by RIP-MD include H-bonds, Salt bridges, VdWs, cation-π, π- π, Arginine-Arginine and Coulomb interactions. In addition, RIP-MD also computes interactions based on distances between C_αs and disulfide bridges. The results of the analysis are shown in an user friendly interface. Moreover, the user can take advantage of the VMD visualization capacities, whereby through some effortless steps, it is possible to select and visualize interactions described for a single, several or all residues in a MD trajectory. Network and descriptive table files are also generated, allowing their further study in other specialized platforms. Our method was written in python in a parallelized fashion. This characteristic allows the analysis of large systems impossible to handle otherwise. RIP-MD is available at http://www.dlab.cl/ripmd.

1 RIP-MD: A tool to study Residue Interaction
2 Networks in Protein Molecular Dynamics

3 Sebastian Contreras-Riquelme [1,3,4], Jose-Antonio Garate [2],
Tomas Perez-Acle [1,2,*] and Alberto J.M. Martin [4*]

4 October 23, 2018

5 [1]Computational Biology Laboratory (DLab), Fundacion Ciencia & Vida,
6 Santiago, Chile
7 [2]Centro Interdisciplinario de Neurociencia de Valparaíso, Valparaiso, Chile
8 [3]Facultad de Ciencias Biológicas, Universidad Andres Bello, Santiago, Chile.
9 [4]Network Biology Laboratory, Centro de Genómica y Bioinformática, Facultad
10 de Ciencias, Universidad Mayor, Santiago, Chile.
11 [*]To whom correspondence should be addressed.

12 1 Abstract

13 Protein structure is not static; residues undergo conformational rearrangements
14 and, in doing so, create, stabilize or break non-covalent interactions. Molec-
15 ular Dynamics (MD) is a technique used to simulate these movements with
16 atomic resolution. However, given the data-intensive nature of the technique,
17 gathering relevant information from MD simulations is a complex and time con-
18 suming process requiring several computational tools to perform these analyses.
19 Among different approaches, the study of Residue Interaction Networks (RINs)
20 has proven to facilitate the study of protein structures. In a RIN, nodes repre-
21 sent amino-acid residues and the connections between them depict non-covalent
22 interactions.
23 Here, we describe RIP-MD, a Visual Molecular Dynamics (VMD) plugin to
24 facilitate the study of RINs using trajectories obtained from MD simulations
25 of proteins. Our software generates RINs from MD trajectory files. The non-
26 covalent interactions defined by RIP-MD include H-bonds, Salt bridges, VdWs,
27 cation-π, π-π, Arginine-Arginine and Coulomb interactions. In addition, RIP-
28 MD also computes interactions based on distances between C_αs and disulfide
29 bridges. The results of the analysis are shown in an user friendly interface. More-
30 over, the user can take advantage of the VMD visualization capacities, whereby
31 through some effortless steps, it is possible to select and visualize interactions
32 described for a single, several or all residues in a MD trajectory. Network and
33 descriptive table files are also generated, allowing their further study in other
34 specialized platforms. Our method was written in python in a parallelized fash-
35 ion. This characteristic allows the analysis of large systems impossible to handle
36 otherwise. RIP-MD is available at `http://www.dlab.cl/ripmd`.
37 **Contact:** alberto.martin@umayor.cl or tomas@dlab.cl

38

2 Introduction

40 The function of proteins is determined by both their 3D structure and their
41 behavior. Therefore, the traditional dogma *sequence→structure→function* is
42 currently restated as *sequence→structure→dynamics→function*. Not surpris-
43 ingly, Molecular Dynamic (MD) methods have become essential tools to explore
44 protein dynamics with atomic resolution [38]. Notwithstanding, one of the ma-
45 jor caveats of MD is the huge amount of data produced by each simulation,
46 hindering both data handling and analysis. Thus, the generation of proper
47 tools to conduct structural and dynamic analyses, is an area of active research
48 and development [20, 19]. Among other methods, the production of Residue In-
49 teraction Networks (RINs) gained popularity due to their simple and intuitive
50 approach. RINs are a graph representation of protein structures in which nodes
51 represent Amino Acids (AAs) and the existence of an interaction between two
52 AAs is represented by edges. Some interesting usage of RINs in MD include
53 the identification of key AAs involved in several processes such as allosterism
54 [34, 28, 25], enzymatic activity [22], protein folding [39] and in protein-protein
55 interactions [7]. RINs have also been adopted for the prediction of the effect of
56 single point mutations in protein stability [15]; the analysis of protein stability
57 [5]; the superimposition and comparison of protein structures [18, 28]; assess-
58 ing the quality of predicted protein structures [37]; and the study of light dark
59 transitions in photoreceptors [25].

60 Many types of RINs can be defined depending on how the interactions be-
61 tween AAs are outlined. The most common definition is the euclidean distance
62 between atoms belonging to different residues, and thus interactions exists only
63 if the distance between AAs is shorter than a given threshold. When this dis-
64 tance is calculated between C_αs or C_βs, RINs are traditionally known as Contact
65 Maps (CMs) [34, 6]. RINs based on thresholded distances are simple graphs,
66 i.e., a pair of nodes can only be connected by a single edge. Nevertheless, there
67 is another approach to build RINs that considers explicitly all non-covalent in-
68 teractions pertaining a pair of AAs [29, 41], giving rise to a multigraph in which
69 nodes can be connected by several edges. In this way, each edge represents
70 a single non-covalent interaction. Non-covalent interactions depicted by edges
71 usually include salt bridges (SBs), hydrogen bonds (HBs), π-π, cation-π and
72 van der Waals (vdW) contacts among others. Multigraph RINs contain more
73 information than simple graphs built upon distance thresholds, since all inter-
74 actions are described in the graph. As a drawback, the existence of several
75 connections between pairs of nodes prevents the use of common algorithms and
76 metrics that can be applied to characterize and analyze simple graphs (see [4]
77 for a discussion of the calculation of shortest paths in multigraphs).

78 Most of the current approaches to depict protein structures as a network are
79 static [42], representing a single protein structure as a RIN disregarding the dy-
80 namic properties of the AAs and atoms forming it. Lately, several authors have
81 introduced RIN based methodologies on full MD trajectories, the so called dy-
82 namic RINs, allowing for the generation of consensus graphs that characterizes
83 and contains the (averaged) dynamic properties of each residue with respect to
84 the rest [3, 12, 31, 33]. Nonetheless, all the aforementioned methods are based

on distance thresholds between single atoms lacking the important information provided by the explicit inclusion of non-covalent interactions.

Herein, we present RIP-MD *(Residue Interaction Networks in Protein Molecular Dynamics)*, a software to generate both static and dynamics RINs. RIP-MD is available as a web server for RIN derivation using static protein structures (PDB files [2]), or as a Visual Molecular Dynamics (VMD) [21] plugin to obtain dynamic RINs derived from MD trajectories. RIP-MD is meant to be an intuitive and easy-to-follow visualization tool for (pairwise) residue interactions. The latter is particularly relevant for MD in which the generated data is enormous. However, due to the employment of networks, the users can take advantage of the mathematical formalism of network theory (i.e. by using Cytoscape) to further analyze their simulations. RIP-MD reads MD trajectory files in DCD binary format generating snapshots at fixed time intervals from which a dynamic RIN is built on. RIP-MD can be employed to perform several analyses, including cross-correlation studies between the different types of interactions; generation of a consensus RIN where edges exists if they are present in at least a given percentage of the snapshots; and the calculation of different node centralities. In the next sections a detailed description of RIP-MD is elaborated. Furthermore, the utility of RIP-MD is emphasized with two examples of MD trajectories: the analysis of possible interactions that stabilize the structure of a Gap Junction Channel and the structural changes on the Lymphocyte antigen 96 protein.

A web server for the creation of RINs from single PDB files, the stand alone and the VMD plug-in versions of RIP-MD can be accessed and downloaded from `dlab.cl/ripmd`. In addition, to facilitate the use of RIP-MD under any operative system we also provide it already installed in a VirtualBox® machine.

3 Methods

The general workflow of RIP-MD is summarized in Figure 1. RIP-MD starts either with a dynamic (MD trajectory) or static (PDB file) protein structure, and the parameters defining the interactions as input (Figure 1.A). The next step pre-processes the input to ensure it complies with the required format (Figure 1.B). Then, a search for interactions between all atoms is carried out in each snapshot from the MD trajectory or using the static structure from the PDB file (see Figure 1.C). In the last step (Figure 1.D), RIP-MD generates the output files including correlation maps and files defining a RIN, which can be further characterized in network visualization tools such as Cytoscape [35]. Each of these steps is described in detail below.

3.1 Input and pre-processing steps:

As previously mentioned, RIP-MD takes as input either a dynamic or a static protein structure (Figure 1.A). Structural analyses within RIP-MD are handled by *MDAnalysis* [30, 17], a python library to manipulate MD trajectories. The first action in this step is to delete heteroatoms maintaining only protein atoms. If the user wants to add missing atoms, such as hydrogens, these are added based on internal coordinates with *PDB2PQR* [9, 10]. After all these steps, additional parameters such as (partial) charges, Lennard-Jones parameters and secondary
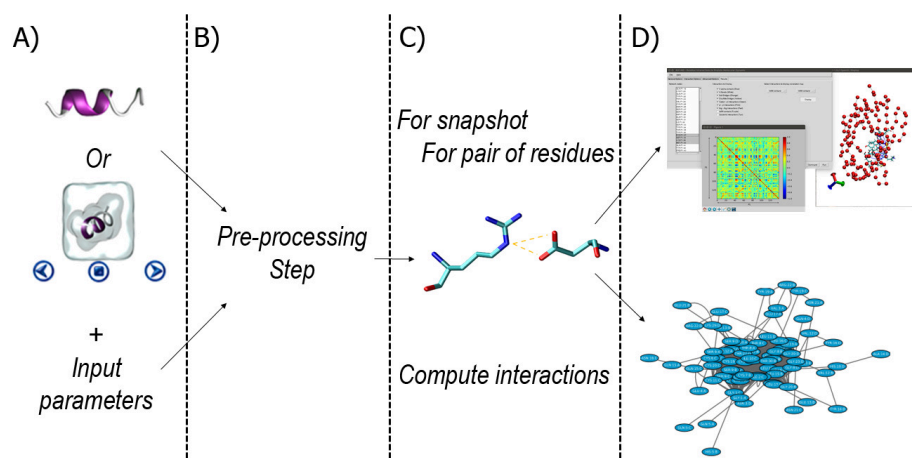
Figure 1: Workflow in *RIP-MD*. Panel A), input of structural information and analyses parameters. Panel B), pre-processing step. Panel C), definition of interactions (C_α contacts, H bonds, Salt bridges, Disulfide bonds, cation-π, π-π, Arginine-Arginine, Coulomb and Van der Waals contacts) according to the input parameters. Panel D), generation of RIN and output files.

structure are either defined or calculated. In the case of MD trajectories, partial charges are assigned from the topology file, *i.e.* PSF file, employed to perform the simulation. Charges for static structures and Lennard-Jones parameters are assigned from the parameters files of the CHARMM force field [26]. Secondary structure and solvent accessibility for each AA are defined using *DSSP* [23, 24]. For further details about input files and their format, please refer to the RIP-MD user manual.

3.2 Interactions defined in RIP-MD:

RIP-MD defines several types of interactions between the AAs in a protein structure: $C\alpha$ contacts, hydrogen bonds, salt bridges, disulfide bonds, cation-π, π-π, Arg-Arg, Coulomb and van der Waals contacts. Parameters defining each interaction, together with their mathematical formulation is explained in detail in Supp. File Text S1 and resumed in Table 1. Importantly, users may employ the provided by-default parameters defining each of these interactions or set them accordingly to his/her needs.

3.3 RIP-MD versions and availability:

RIP-MD is available free of charge at `http://www.dlab.cl/ripmd` in three options:

- Standalone program: This version is thought for users who want to take advantage of High Performance Computing (HPC) architectures to perform analysis of very large systems, such very long MD trajectories that are impossible to handle otherwise.

Cα contacts	$dist(C\alpha_i, C\alpha_j) \leq d$	$d = 8\mathring{A}$
Hydrogen Bonds	$dist(donor, aceptor) \leq d$	$d = 3\mathring{A}$
	$\theta(\overrightarrow{C-H}, \overrightarrow{aceptor}) \geq a$	$a = 120°$
Salt Bridges	contacts between NH/NZ groups of ARG/LYS and OE*/OD* in ASP/GLU $\leq d$	$d \leq 6\mathring{A}$
Disulfide Bonds	S atoms of two cysteins $\leq d$	$d \leq 3\mathring{A}$
	dihedral $\theta(C-S-S-C) \in [a,b]$	$a \geq 60°$ & $b \leq 90°$
Cation-π interactions*	distance between aromatic rings $\leq d$	$d \leq 6\mathring{A}$
π-π interactions*	$dist(aromatic\ ring, cation) \leq d$	$d = 7\mathring{A}$
	$\theta(\overrightarrow{normal\ vector\ ring}, \overrightarrow{ring\ center - cation}) = a$	$a \in [0°, 60°]$ or $a \in [120°, 180°]$
Arg-Arg	$dist(guanidine_1, guanidine_2) \leq d$	$d \leq 5\mathring{A}$
Coulomb interactions	charged-group based cut-off using a 1-4 potential	
van der Waals	12-6 Lennard-Jones potential	

Table 1: Summary of all interactions defined in RIP-MD. A detailed description can be found in Supp. File Text S1. * His residues are considered as a cation if they present a protonated nitrogen atom, and as π-system only if they are not protonated. Other π-systems considered are the aromatic rings of Phe, Tyr and Trp.

152 • *Visual Molecular Dynamics* (VMD)[21] plugin: This version benefits from
153   the graphical interface provided by VMD. This plugin performs a system
154   call to execute the standalone RIP-MD program.

155 • Webserver: This form of RIP-MD is for those users who want to analyze
156   a single PDB structure without installing the software locally. In this
157   web-server, users first upload their selected structure and then, after few
158   steps, the results can be easily downloaded.

159 It is important to note that results generated by the web-server and by the stand-
160 alone version of the program are compatible with the *VMD* plugin, making
161 possible to load and display these results into the *VMD* graphical user interface.
162 The installation guide for both the stand-alone and plug-in versions, together
163 with the user manual, is available at the RIP-MD webpage. In addition, users
164 can also download from the webpage a virtual machine with RIP-MD already
165 available which only requires the installation of all proprietary software that can
166 not be distributed by us.
167   Other scripts employed to generate the figures shown in this article are also
168 available at `http://www.dlab.cl/ripmd` or directly at `https://github.com/`
169 `networkbiolab/supp_script_ripmd`.

170 3.4 Output files.

171 RIP-MD provides two types of output: interaction graphs and Pearson correla-
172 tion plots.

173 3.4.1 Residue Interaction Networks (RINs):

174 Once the interactions are computed, RIP-MD generates network files that can
175 be visualized in specialized platforms such as Cytoscape [35]. In these networks,
176 each node represents an AA and each edge represents an interaction between
177 AAs. Several network files are generated; one for each type of interaction type
178 and a global network containing all interactions. Network files also describe
179 additional information about several properties of the AAs including secondary

structure and solvent accessible surface. These network files also contain edge attributes and the geometrical descriptors defining each interaction. To account for the dynamic behavior of node descriptors and interactions, RIP-MD calculates the percentage of frames in which they are present over the simulation.

3.4.2 Pearson correlation plots.

RIP-MD employs Pearson correlation to identify, for any given type of interaction, the existence of relationships between the dynamic behavior of two AAs. To do so, each AA is represented by a numerical vector in which each element contains the number of interactions of a given type in each frame. Once all vectors are obtained, the existence of a relationship between the rupture or formation of interactions over time is determined by calculating the Pearson correlation between these vectors. These results are provided in the form of correlation plots, square matrices of size N representing the total number of AAs, where each element is colored according to the correlation value.

3.5 Comparison with other methods

A comparison with two other approaches, Carma [16] and MD-TASK [6], used on our first example (the MD2 pocket closure, see next section) is explained in detail in Supp. File Text S1. These methods were chosen due to their availability, they are easy to install and they consider both inter and intra chain interactions, even thought both methods only generate a single type of interactions between residues (C_α and C_β contacts for Carma and MD-TASK respectively). Interested readers can also find in Supp. File Text S1 a brief description of other methods or libraries that can be used to generate RINs for MD simulations, even if, as Carma and MD-TASK, they use MDs instead of non-covalent interactions in their RIN definitions.

4 Application of RIP-MD to analyze MD simulations of proteins:

In the following section we explore two applications of RIP-MD to analyze MD simulations of proteins: conformational changes occurring in a soluble protein and the inter-monomeric interactions of a Gap-Junction Channel (Figure 2). Both analysis were carried out employing the by-default parameters of RIP-MD. The first example (Figures 3 and 4) is meant to show how RIP-MD can visually discriminate among the different stages of a conformational change and how correlated interactions guide the process and are lost when the "closed" state is reached, which might be a common trend in process guided by non-polar interactions. Our new analysis demonstrates new chemical information that was not reported in [14], the original studies in which for the first time the closing process was reported. The second example is meant to emphasize how interactions in x-ray protein crystals are not necessarily the same. Moreover, the visualization power of RIP-MD is highlighted as a very complex system (12 monomers) is very easily visualized by our method.
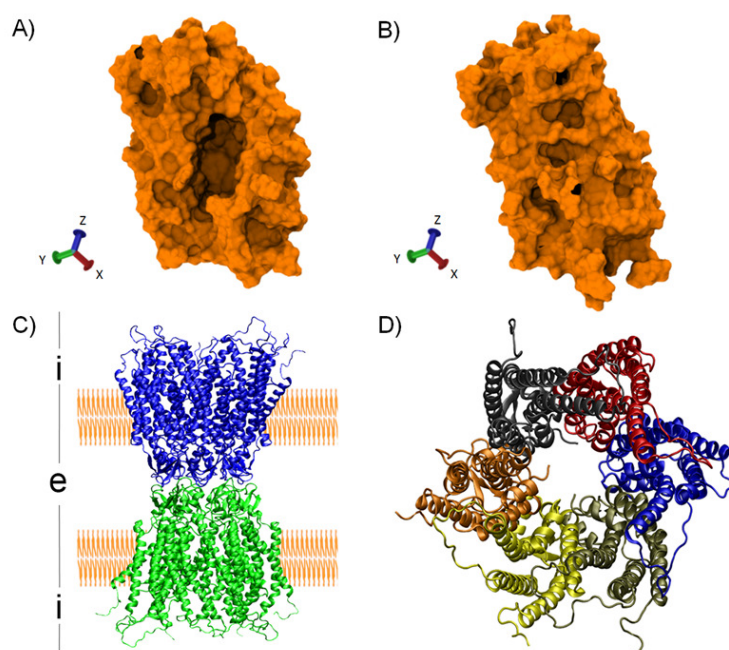
Figure 2: Protein structures used as case of study: First and last snapshots of the MD2 trajectory simulation (top) and Molecular structure of the human CX26 hemichannel and Gap Junction Channel (bottom). Panel (A), first snapshot of the MD showing the hydrophobic pocket in an open conformation. Panel (B), last snapshot of the MD where MD2 exhibit a closed conformation. Panel C), secondary structure representation of a GJC formed by the extracellular docking of two HCs (blue and green). The position of the plasmatic membrane appears rendered in orange. The extracellular space is denoted with an "i", while the intracellular space is denoted with an "e". Panel D), Secondary structure representation of a HC from the intracellular view. Each CX26 monomer is represented using different colors.
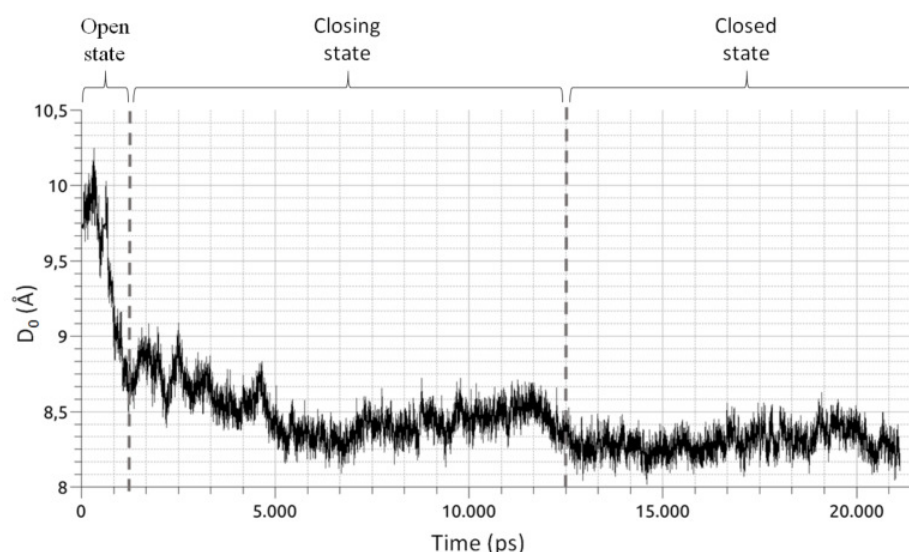
Figure 3: Time series of the collective variable D_0 (Eq.1). The graphic shows three distinct phases of the pocket closure, as marked by the vertical dashed lines. From 0 to 1,250 ps, the pocket is in an open conformation. From 1,250 to 12,500 ps, the closing process begins and is followed by a small opening of the pocket. From 12,500 ps until the end of the simulation, the pocket remains in a closed state.

4.1 MD2: an example of a conformational change.

MD2 is a soluble protein part of the innate immune response in humans. Upon binding of lipopolysaccharides belonging to Gram- bacteria, MD2 triggers an immune response via the interaction with Toll-Like receptors 4 [11, 32]. As shown by MD simulations [8, 14], MD2 possesses a hydrophobic cavity that rapidly closes upon ligand removal (Figure 2). The later is a clear example of a conformational change switching the RIN, suitable to be visualized with RIP-MD. To do so, we analyzed MD simulations of apo-MD2 obtained from [14]. Briefly, a 20 ns MD of a solvated MD2 was divided into three windows (Figure 3) which characterize the closing process. The closing event along the MD simulation was projected onto the collective variable described in Equation 1.

$$D_0 = \frac{1}{N} \sum_{n=1}^{N} \sqrt{(COM\beta_n - COM_{MD2})^2} \qquad (1)$$

were n stands for each of the 10 β strands of MD2, and $COM\beta_n$ and COM_{MD2} are the centers of mass of strand n and MD2 respectively. As reported in [14], D_0 unambiguously differentiates the three well defined stages of the closing process. We used RIP-MD to obtain RINs for both HBs and vdW interactions between AAs, employing default parameters (see Table 1 and Sup. Mat.). For each window, Pearson correlations were calculated for the number of HBs and vdW interactions for any given pair of AAs.
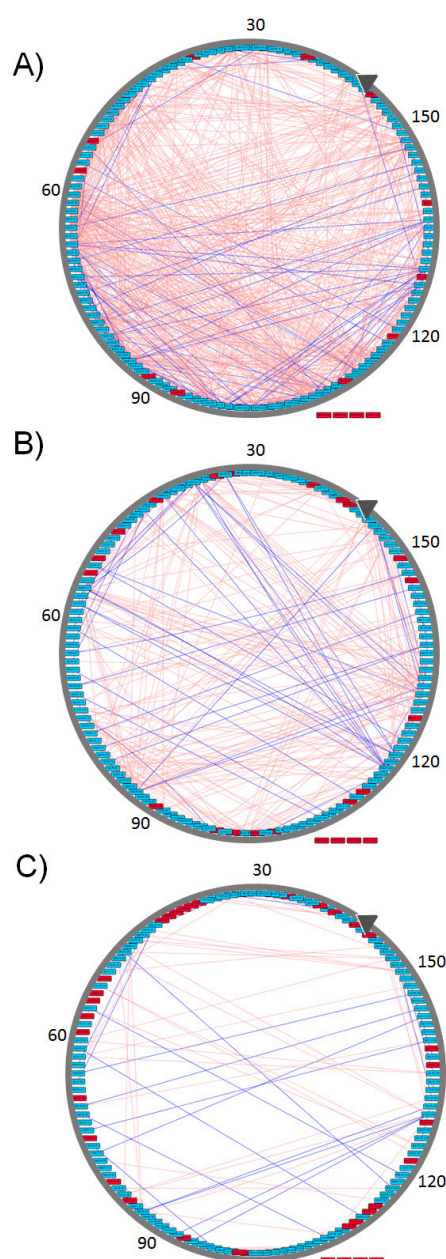
Figure 4: Graphic representation of changes occurring in the *RIN* of MD2 during the three windows of its closing process. All edges represent Pearson absolute correlation values where $|r| \geq 0.5$ in the open (A), closing (B) and closed conformation (C). Pink edges connect those AA interacting through vdW contacts while blue edges connect those AAs forming HBonds. Red nodes indicate residues with no $|r| \geq 0.5$, nodes outside the circle, Pro50, Met85, Lys125 and Pro142, do not form any interaction $|r| \geq 0.5$ in any of the three conformational states. Images were created after loading the resulting networks in Cytoscape with circular layout sorted according to AA numbering, first AA is indicated by the gray arrow.

	Window	≥ 0.5	≥ 0.6	≥ 0.7	≥ 0.8	≥ 0.9
HBs	open	1 (98)	1 (53)	1 (24)	1 (10)	1 (4)
	closing	0.64 (63)	0.55 (29)	0.54 (13)	0.7 (7)	0.25 (1)
	closed	0.38 (37)	0.40 (21)	0.54 (13)	0.90 (9)	1.25 (5)
vdW	open	1 (656)	1 (295)	1 (120)	1 (20)	* (0)
	closing	0.39 (257)	0.35 (103)	0.19 (23)	0.10 (2)	* (1)
	closed	0.12 (78)	0.08 (23)	0.07 (8)	0 (0)	* (0)

Table 2: Rate of correlated pairs over the closing process of MD2 according to their absolute Pearson correlation values over the three simulation windows: open, closing and closed conformations. Numbers indicate the rate between the number of correlated pairs of amino acids found in the first window (open state) an the number of correlated pairs in each of the other two windows with respect to the open state for Hydrogen Bonds and van der Waals interactions. In parenthesis the number of pairs of interacting AAs for each absolute Pearson correlation values. "*" indicates the absence of a interaction in the first window so it can not be computed for the other windows.

RINs obtained during the closing process of MD2 (Figure 3) for both HBs and vdW interactions between AAs, reflecting absolute Pearson correlations values $|r| \geq 0.5$ are shown in Figure 4. There is a higher amount of correlated pairs for both HBs and vdW interactions at the initial stage, as the closing process continues the number of correlated pairs decay (see Figure 3). The latter reveals an initial concerted action that triggers the closing event, but is lost after closure, a clear indication of the unspecific nature of non-polar interactions which are dominant within the hydrophobic cavity of MD2. To further quantify these changes, Table 2 presents the number of correlated pairs at each window normalized by the amount of correlated pairs of the first window. Overall, a monotonic decrease upon window increment, not withstanding any correlation threshold, is observed. Slight deviations from the latter regarding HBs occur at Pearson values equal or above 0.7 (third and fourth columns of Table 2). This increase is due to the formation of stable H-bonds upon MD2 closure (see Table 2). On the other hand, the number of vdW interactions in Table 2 is higher than that of H-bonds, as the latter are specific interactions depending on both the relative distance and orientations of the participant residues. Finally, while most of the pairs of residues are uncorrelated, the β strands tend to be coordinated with respect to the formation and break of both types of interactions, showing how all secondary structure elements are kept in the closure of MD2.

Interestingly, this behavior is not expected for other conformational changes such as folding, in which the formation of polar interactions will lead to the appearance of highly correlated pairs, *e.g.* the formation of an alpha-helix. In this way, RIP-MD served to quantify and visualize the counter-intuitive idea that a large conformational change can indeed lead to lower correlations, a consequence of the nature of the interactions that dominate a given structural process; in this case, non-polar VdWs interactions among residues within the MD2 cavity.

4.2 Gap Junction Channel: An example for "inter-monomeric" and "inter-molecular" interactions in a large system.

GJCs are intercellular hydrophilic channels connecting the cytoplasm of two adjacent cells [40]. GJCs allow the exchange of water, ions and small molecules of up to 1 kDa [36, 1]. GJCs are formed by the extracellular docking of two HCs, (Figure 2.A), where each HC is formed by six CXs monomers (Figure 2.B). In this example, we studied a 20ns all-atom MD simulation of a complete GJC formed by the human CX26 [13]. This molecular system contains twelve identical CX monomers in total, each one comprising 226 AAs. Due to the high computational cost of studying vdW and Coulombic interactions in such a large system, these were disregarded in this example. In addition, due to the covalent nature of disulfide bridges these were not considered in the following analyses. RIP-MD was run with default parameters, keeping only those appearing in at least 75% of the simulation time.

Table 3 exhibits the interactions found by RIP-MD over the human CX26 GJC simulation. All interaction types are found to be highly stable over the simulation with an overall persistence of 95%-100%, indicating that this molecular structure remains structurally stable at least over the 20ns of the MD simulation. Since C_α contacts describe spatial relationships between AAs with a plenty of available methods to study their interaction [34], only the other types of interactions calculated by RIP-MD are further analyzed.

	[75%,80%)	[80%,85%)	[85%,90%)	[90%,95%)	[95%,100%]	Total
C_αs	133	169	173	239	10311	11025
HBs	112	135	131	205	1557	2140
SBs	20	10	25	39	448	542
π-π	9	11	15	21	84	140
Arg-Arg	0	0	0	1	0	1
All	274	325	344	505	12400	13848

Table 3: Interactions present in different time intervals over the 20 ns GJC MD simulation

Table 4 shows those interactions occurring at least during the 75% of the MD simulation for all the following cases: interactions between AAs of the same chain for each of the two HCs (Intra-chain), between chains of the same HC (Inter-chain) and between AAs of different HCs (Inter-HCs). It should be noted that HBs and SBs are interactions appearing mainly between different chains, while π-π and Arg-Arg interactions are both intramonomeric interactions. Interestingly, only HBs appear at the interface of the two docked CXs (Fig 5). The role of these non-bonded interactions in the maintenance of the GJC quaternary structure were also described in the original paper reporting the CX26 crystal structure [27].

The information shown in the bottom section of Table 4 was then compared with the original static structure of the GJC after energy minimization (model based on PDB ID 2ZW3, see [1] for details). The top section of Table 4 resumes the number of interactions present in this structure. The first thing to notice when comparing the dynamic and static versions of the network is the lower number of interactions in the dynamic network. This reduction can be due

307 to their low stability at physiological temperatures, the replacement of intra-
308 hbonds with hbonds established with solvent molecules or a poor description of
309 a given interaction by the force-field, *e.g.* $\pi - \pi$ interactions. Notoriously, the
310 number inter-CXs HB interactions increases from 9 in the static structure to 26
311 in the dynamic network, highlighting the relevance of HBs to maintain the GJC
312 complex. It is also very relevant that most of the inter-CXs HBs observed in the
313 non-dynamic RIN are absent in the dynamic version, with only 2 HBs appearing
314 in the entire simulation (Table 5). Regarding the other types of interactions,
315 most of the intra and inter-chain HBs, SBs, Cation-π and $\pi - \pi$ interactions
316 appear in the dynamic and non-dynamic RINs.

		HC1		HC2		Inter-HC
		Intra-chain	Inter-chain	Intra-chain	Inter-chain	
Static	HBs	1312	101	1301	106	9
	SBs	168	106	174	104	0
	Cation-π	3	0	6	0	0
	π-π	70	11	65	10	0
	Arg-Arg	0	4	0	5	0
	Total	1553	222	1546	225	9
Dynamic	HBs	991	66	990	67	26
	SBs	183	101	170	88	0
	π-π	66	6	64	4	0
	Arg-Arg	0	0	0	1	0
	Total	1240	173	1224	160	26

Table 4: Number and type of interactions present in the human CX26 GJC. Interactions are divided into intra-chain, inter-chain in the same HC and inter-HC monomers. The top section displays the interactions in the dynamic network (interactions present in at least 75% of the simulation), and the bottom in the network derived from the static structure.

		HBs	SBs	Cation-π	π-π	Arg-Arg
Static	Intra-chain	948	89	9	43	0
	Inter-chain	136	44	0	14	9
	Inter-HC	7	0	0	0	0
MD	Intra-chain	316	100	0	38	0
	Inter-chain	62	23	0	3	1
	Inter-HC	24	0	0	0	0
Both	Intra-chain	1665	253	0	92	0
	Inter-chain	71	166	0	7	0
	Inter-HC	2	0	0	0	0

Table 5: Number of interactions of each type appearing only in the static, in the MD simulation and in both RINs of the human CX26 GJC. Interactions are separated by type and into intra-chain. inter-chain and inter-HCs.
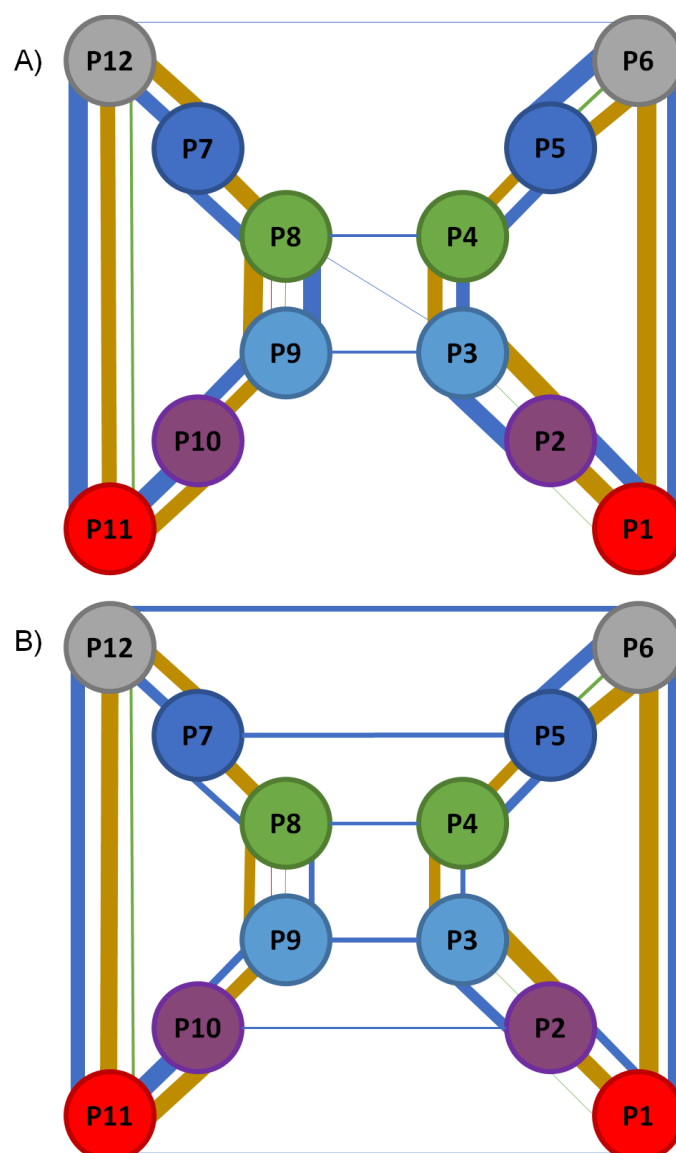
Figure 5: Representation of the RIN formed for the static and the dynamic structure of GJC. Panel A) show the network for the static structure and Panel B) for the MD simulation. Each circle represents a CX subunit using a color code for subunit in each of them: red for chain A (segments P1 and P11); purple for chain (segments P2 and P10); light blue for chain C (segments P3 and P9); green for chain D (segments P4 and P8); blue for chain E (segments P5 and P7); and gray for chain F (segments P6 and P12). Interactions thickness represent the quantity of interactions, colored with the following color code: red for Arg-Arg interactions; blue for HBs; gold for SBs; and green for π-π interactions. No self-interactions are represented.

5 Conclusion

MD simulation is a routinely employed technique to study the dynamic behavior of a system *i.e.* a protein. Therefore, tools that can extract relevant information in a simple and user-friendly way are urgently needed. Here we describe RIP-MD, a method that using graph theory approaches generates RINs for different types of electrostatic interactions in protein MD simulations. We are currently working on an improved version of RIP-MD to consider waters, ions and other non-AAs molecules to generate RINs. Using RIP-MD we were able to study MD simulations of two systems: MD2 and a GJC. Regarding MD2, we focused on the study of the protein movement as reflected by Pearson correlation plots of HBs and van der Waals contacts. This analysis showed notable differences between the different stages of the conformational change of the protein, revealing an initial concerted action at the beginning of the closing process, with an overall reduction of correlations for the closed state. In the case study of GJC, a comparison of the initial structure and a short MD simulation revealed that inter-chain interfaces are stabilized mainly by HBs and SBs, and that Arg-Arg and Cation-π interactions tend to disappear over the trajectory.

RIP-MD is freely available for the academic community, and it is distributed in three forms: a webserver, where users can analyze a single *PDB*; an standalone version that can take advantage of Multi-Core systems to generate these RINs; and a VMD plugin that executes the standalone version of the software and at the same time benefits from the graphical viewer of VMD. All these distributions, together with manuals and help files can be accessed from http://dlab.cl/ripmd.

Acknowledgements

The authors would like to acknowledge F. Villanelo and other DLab members for their useful comments, suggestions and discussions of the work presented here.

References

[1] R. Araya-Secchi, T. Perez-Acle, S. gu Kang, T. Huynh, A. Bernardin, Y. Escalona, J.-A. Garate, A. D. Martínez, I. E. García, J. C. Sáez, and R. Zhou. Characterization of a novel water pocket inside the human cx26 hemichannel structure. *Biophysical Journal*, 107(3):599 – 612, 2014.

[2] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[3] M. Bhattacharyya, C. R. Bhat, and S. Vishveshwara. An automated approach to network features of protein structure ensembles. *Protein Science*, 22(10):1399–416, 2013.

[4] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008.

[5] K. V. Brinda and S. Vishveshwara. A network representation of protein structures: implications for protein stability. *Biophysical journal*, 89(6):4159 – 4170, 2005.

[6] D. K. Brown, D. L. Penkler, O. S. Amamuddy, C. Ross, A. R. Atilgan, C. Atilgan, and Ö. T. Bishop. MD-TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics (Oxford, England)*, 27, 2017.

[7] A. del Sol and P. O'Meara. Small-world network approach to identify key residues in protein-protein interaction. *Proteins*, 58(3):672–82, 2005.

[8] M. L. DeMarco and R. J. Woods. From agonist to antagonist: Structure and dynamics of innate immune glycoprotein MD-2 upon recognition of variably acylated bacterial endotoxins. *Mol Immunol*, 2011.

[9] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker. Pdb2pqr: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research*, 35(Web Server issue):W522–5, 2007.

[10] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker. Pdb2pqr: an automated pipeline for the setup of poisson-boltzmann electrostatics calculations. *Nucleic acids research*, 32(Web Server issue):W665–7, 2004.

[11] R. Dziarski, Q. Wang, K. Miyake, C. J. Kirschning, and D. Gupta. Md-2 enables toll-like receptor 2 (tlr2)-mediated responses to lipopolysaccharide and enhances tlr2-mediated responses to gram-positive and gram-negative bacteria and their cell wall components. *The Journal of Immunology*, 166(3):1938–1944, 2001.

[12] J. Eargle and Z. Luthey-Schulten. Networkview: 3d display and analysis of protein·rna interaction networks. *Bioinformatics (Oxford, England)*, 28(22):3000–1, 2012.

[13] Y. Escalona, J.-A. Garate, R. Araya-Secchi, T. Huynh, R. Zhou, and T. Perez-Acle. Exploring the membrane potential of simple dual-membrane systems as models for gap-junction channels. *Biophysical journal*, 110(12):2678–88, 2016.

[14] J. A. Garate and C. Oostenbrink. Lipid a from lipopolysaccharide recognition: Structure, dynamics and cooperativity by molecular dynamics simulations. *Proteins: Structure, Function and Bioinformatics*, 81(4):658–674, 2013.

[15] M. Giollo, A. J. Martin, I. Walsh, C. Ferrari, and S. C. Tosatto. Neemo: a method using residue interaction networks to improve prediction of protein stability upon mutation. *BMC Genomics*, 15(Suppl 4):S7, 2014.

[16] N. M. Glykos. Software news and updates carma: A molecular dynamics analysis program. *Journal of Computational Chemistry*, 27(14):1765–1768.

[17] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, J. Domański, D. L. Dotson, S. Buchoux, I. M. Kenney, and O. Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In Sebastian Benthall and Scott Rostrup, editors, *Proceedings of the 15th Python in Science Conference*, pages 98 – 105, 2016.

[18] N. Gupta, N. Mangal, and S. Biswas. Evolution and similarity evaluation of protein structures in contact map space. *Proteins: Structure, Function and Genetics*, 59(2):196–204, 2005.

[19] S. Hayward and B. L. de Groot. Normal modes and essential dynamics. *Methods in Molecular Biology*, 443:89–106, 2008.

[20] J. S. Hub and B. L. De Groot. Detection of functional modes in protein dynamics. *PLoS Computational Biology*, 5(8), 2009.

[21] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[22] W. Y. Jianhong Zhou, G. Hu, and B. Shen. Amino acid network for prediction of catalytic residues in enzymes: a comparison survey. *Current Protein and Peptide Science*, 17:41–51, 2016.

[23] R. P. Joosten, T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander, and G. Vriend. A series of pdb related databases for everyday needs. *Nucleic Acids Research*, 39(suppl 1):D411–D419, 2011.

[24] W. Kabsch and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[25] R. Kaur Grewal, D. Mitra, and S. Roy. Mapping networks of light–dark transition in lov photoreceptors. *Bioinformatics (Oxford, England)*, 31(22):3608–3616, 2015.

[26] A. MacKerell, D. Bashford, M. Bellot, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.

[27] S. Maeda, S. Nakagawa, M. Suga, E. Yamashita, A. Oshima, Y. Fujiyoshi, and T. Tsukihara. Structure of the connexin 26 gap junction channel at 3.5 a resolution. *Nature*, 458(7238):597–602, 2009.

[28] N. Malod-Dognin and N. Pržulj. Gr-align: fast and flexible alignment of protein 3d structures using graphlet degree similarity. *Bioinformatics (Oxford, England)*, 30(9):1259–65, 2014.

[29] A. J. M. Martin, M. Vidotto, F. Boscariol, T. Di Domenico, I. Walsh, and S. C. E. Tosatto. Ring: networking interacting residues, evolutionary information and energetics in protein structures. *Bioinformatics (Oxford, England)*, 27(14):2003–5, 2011.

[30] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10):2319–2327, 2011.

[31] M. Münz and P. C. Biggin. Jgromacs: a java package for analyzing protein simulations. *Journal of chemical information and modeling*, 52(1):255–9, jan 2012.

[32] U. Ohto, K. Fukase, K. Miyake, and Y. Satow. Crystal structures of human md-2 and its complex with antiendotoxic lipid iva. *Science*, 316(5831):1632–1634, 2007.

[33] M. Pasi, M. Tiberti, A. Arrigoni, and E. Papaleo. xpyder: a pymol plugin to analyze coupled residues and their networks in protein structures. *Journal of chemical information and modeling*, 52(7):1865–74, jul 2012.

[34] A. Sethi, J. Eargle, A. A. Black, and Z. Luthey-Schulten. Dynamical networks in trna:protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6620–5, 2009.

[35] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.

[36] G. Söhl and K. Willecke. Gap junctions and the connexin protein family. *Cardiovascular Research*, 62(2):228–232, 2004.

[37] M. L. Tress and A. Valencia. Predicted residue–residue contacts can help the scoring of 3d models. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1980–1991.

[38] W. F. van Gunsteren, P. H. Hünenberger, A. E. Mark, P. E. Smith, and I. G. Tironi. Computer simulation of protein motion. *Computer Physics Communications*, 91(1-3):305–319, 1995.

[39] M. Vendruscolo, N. V. Dokholyan, E. Paci, and M. Karplus. Small-world view of the amino acids that play a key role in protein folding. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 65(6 Pt 1):061910, 2002.

[40] F. Villanelo, Y. Escalona, C. Pareja-Barrueto, J.-A. Garate, I. M. Skerrett, and T. Perez-Acle. Accessing gap-junction channel structure-function relationships through molecular modeling and simulations. *BMC cell biology*, 18(Suppl 1):5, 2017.

[41] K. Wolek, Á. Gómez-Sicilia, and M. Cieplak. Determination of contact maps in proteins: A combination of structural and chemical approaches. *Journal of Chemical Physics*, 143(24):243105, dec 2015.

[42] W. Yan, J. Zhou, M. Sun, J. Chen, G. Hu, and B. Shen. The construction of an amino acid network for understanding protein structure and function. *Amino acids*, 46(6):1419–39, 2014.