

RESEARCH ARTICLE

AlphaFold2 modeling and molecular dynamics simulations of an intrinsically disordered protein

Hao-Bo Guo^{1,2}*, Baxter Huntington^{1,3}, Alexander Perminov^{1,3}, Kenya Smith⁴, Nicholas Hastings⁴, Patrick Dennis¹, Nancy Kelley-Loughnane¹*, Rajiv Berry¹*

1 Material and Manufacturing Directorate, Air Force Research Laboratory, WPAFB, Mason, OH, United States of America, **2** UES Inc., Dayton, OH, United States of America, **3** Miami University, Oxford, OH, United States of America, **4** United States Air Force Academy, Colorado Springs, CO, United States of America

* These authors contributed equally to this work.

* haobo.guo.ctr@us.af.mil (HBG); nancy.kelley-loughnane.1@us.af.mil (NKL); rajiv.berry@us.af.mil (RB)

OPEN ACCESS

Citation: Guo H-B, Huntington B, Perminov A, Smith K, Hastings N, Dennis P, et al. (2024) AlphaFold2 modeling and molecular dynamics simulations of an intrinsically disordered protein. PLoS ONE 19(5): e0301866. <https://doi.org/10.1371/journal.pone.0301866>

Editor: Arabinda Ghosh, Gauhati University, INDIA

Received: December 5, 2023

Accepted: March 23, 2024

Published: May 13, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0301866>

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All data generated or analyzed during this study are included in this article and its [Supporting Information](#) files. R codes for residue-residue interaction network and secondary structure element analyses and

Abstract

We use AlphaFold2 (AF2) to model the monomer and dimer structures of an intrinsically disordered protein (IDP), *Nvjp-1*, assisted by molecular dynamics (MD) simulations. We observe relatively rigid dimeric structures of *Nvjp-1* when compared with the monomer structures. We suggest that protein conformations from multiple AF2 models and those from MD trajectories exhibit a coherent trend: the conformations of an IDP are deviated from each other and the conformations of a well-folded protein are consistent with each other. We use a residue-residue interaction network (RIN) derived from the contact map which show that the residue-residue interactions in *Nvjp-1* are mainly transient; however, those in a well-folded protein are mainly persistent. Despite the variation in 3D shapes, we show that the AF2 models of both disordered and ordered proteins exhibit highly consistent profiles of the pLDDT (predicted local distance difference test) scores. These results indicate a potential protocol to justify the IDPs based on multiple AF2 models and MD simulations.

1. Introduction

Intrinsically disordered proteins/regions (IDPs/IDPRs) are proteins or protein regions that lack well-folded three-dimensional (3D) structures [1–3]. These proteins are also referred to as “natively disordered/unstructured/unfolded proteins” in the literature [4–6]. The widespread presence of IDPs in the protein world and their importance to cell functions have been appreciated only starting from the dawn of this century [7]. Nonetheless, the vital roles IDPs play in cells are numerous, such as the formation of biomolecular condensates (also known as membrane-less organelles) [8], serving as protein-protein interaction hubs [9], in cellular signaling and regulation [10], as well as in the evolution of multicellular life form and cell type specifications [11].

Despite the importance of IDPs/IDPRs and the numerous studies in this field, the meaning of “disorder” is sometimes unclear. This is highlighted by the various descriptions, e.g.,

visualizations, as well as the 1000 AF2 (V2.2.2) models of both *Nvjp-1* and T7RdhA can be downloaded from the GitHub repository: <https://github.com/haoboguo/Nvjp-1>.

Funding: This work was supported by funding from the Office of the Under Secretary of Defense for Research and Engineering (OUSD R&E), Applied Research for Advancement of S&T Priorities (ARAP) Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interests.

“dancing protein clouds” [12], “fuzzy proteins” [13], and other names summarized in [6, 12], used to describe IDPs/IDPRs. Many studies use, a binary classification based on disorder predictions, i.e., a residue R_i is disordered if its disorder content (i.e., the probability of being disordered) $D_i > 0.5$, and a protein is an IDP if more than 50% of the residues are disordered. In reality, proteins dynamically engage in their functions, constituting a structure-function continuum [12, 14]. Therefore, a binary (or probability) classification may not be well-suited to determine whether or not a protein is an IDP.

The rationale for over 100 protein intrinsic disorder predictors [15] comes from the hypothesis that residual disorder content is primarily determined by the protein sequence, in line with Anfinsen’s protein sequence-structure dogma [16]. Anfinsen’s dogma is also the target of the CASP (Critical Assessment of techniques for protein Structure Prediction) biennial competition [17]. In 2020, CASP14 announced the grand breakthrough of AlphaFold2 (AF2), which achieved the goal of accurately predicting the protein structure simply from its primary sequence, with accuracy comparable to experimental structures including those by X-ray crystallography and cryoEM [18]. Inspired by CASP and its significant contributions to AlphaFold2 [18–20], in 2021, the Critical Assessment of protein Intrinsic Disorder prediction (CAID) biennial competition was established, for benchmarking the disorder predictors to improve the protein intrinsic disorder predictions [21, 22]. However, what is the exact definition of disorder and how it is justified experimentally or theoretically? The DisProt database [23], for example, is regarded as the ground truth for IDPs (a subset from DisProt was used as the target for CAID-1 [21]). From the prediction part, however, different predictors may yield significant variations (Fig 1A). It is worth noting that AF2 structure predictions, particularly, assisted by the pLDDT (predicted local distance difference test) scores, may serve as important measures of the per-residue disorder in proteins [24].

The present AF2 structure database contains over 200 million protein models covering nearly all known protein sequences [25], including the human proteome [26]. The protein conformational space of this database is enormous [27] (which is in line with a recent parallel perspective [28]): some match well with known experimental models deposited in the Protein Data Bank (PDB [29]) (the “good”); some deviate significantly from the PDB models (the “bad”), possibly owing to different conditions such as the native state conditions cannot be satisfied in the AF2 modeling [30]; moreover, many models are not well folded at all (the “ugly”). The “ugly” models, not surprisingly, are often related to IDPs.

The present work focuses on one example of an “ugly” models: *Nereis virens* jaw protein-1, or *Nvjp-1* [31], which is the predominant protein found in the marine polychaete *N. virens* jaw. The jaw of *N. virens* is made up of 90% (w/w) proteins, yet its mechanical properties in terms of hardness and stiffness are comparable to human dentin [32]. Moreover, these properties were shown to be modulated by Zn-binding [31], rendering *Nvjp-1* a unique candidate for dynamic sclerotization [33]. In the present work, we show that *Nvjp-1* is an IDP: first, the AF2 model of *Nvjp-1* possesses low per-residue pLDDT scores; second, high PAEs (predicted aligned errors) of all residues indicate the absence of correlated movements or interactions among the residues. Moreover, different AF2 models as well as protein conformations taken from the MD simulations deviate largely from each other. We also modeled a theoretical *Nvjp-1* homodimer using AlphaFold-Multimer [34]. Similar to the monomer models, the structures of homodimer models are not consistent with each other; even the two monomers in a selected dimer exhibit large root-mean-square deviation (RMSD) from each other after alignment. This is confirmed by structure snapshots from MD trajectories of the *Nvjp-1* dimer. The results from AF2 (and AlphaFold-Multimer) modeling and MD simulations, therefore, suggest an implication for “disorder” that is associated with the structural heterogeneity observed in the modeling and simulations.

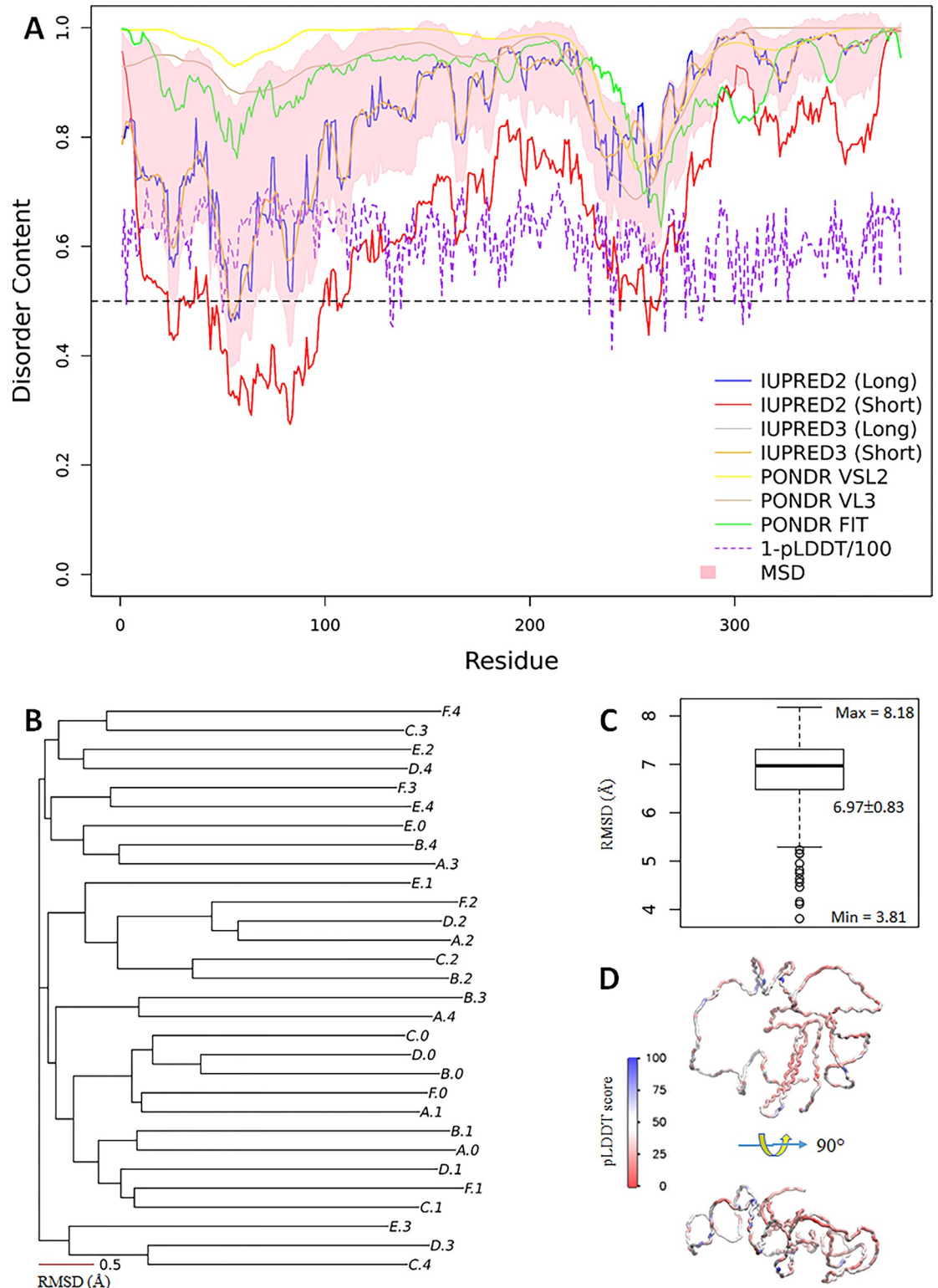


Fig 1. *Nvjp-1* is an intrinsically disordered protein. (A) Disorder contents (D_i) predicted by predictors IUPRED2A [51] (long/short), IUPRED3 [52] (long/short), PONDRL VSL2 [53], PONDRL VL3 [54], and PONDRL FIT [55]. The mean±sd (MSD) from all 7 predictors are shown in pink shade. The purple dashed line is (1-pLDDT/100) using one of the models (A.0) for a comparison. The black dashed line at $D_i = 0.5$ is used for reference. (B) A phylogenetic tree reconstructed using the RMSD matrix calculated for 30 AF2 models. The scale bar (0.5 Å) serves as a reference for “atomic resolution” observed in a well-folded proteins which has the mean RMSD of 0.5 Å, see Fig 7. (C) A boxplot of the RMSD distributions. (D) Front (top) and side (bottom) views of one of the AF2 models colored by the pLDDT scores of all residues.

<https://doi.org/10.1371/journal.pone.0301866.g001>

To represent ensembles of configurations of proteins by MD both sufficiently and correctly, long time-scale or enhanced sampling may be required [35] especially for addressing the ensemble-averaged experimental data [36]. A recent work indicated that a sufficient number of AF2 modeling can generate ensembles of protein conformations, containing “rare conformations” that might require considerably long time-scale MD to capture [37]. Here, besides the ensembles from independent MD simulations, 1000 AF2 models were further used to examine the structural heterogeneity of *Nvjp-1*, compared with 1000 AF2 models of a well-folded protein, the reductive dehalogenase *T7RdhA* [37]. The models of *T7RdhA* agree with each other (mean RMSD of 0.5 Å), whereas the *Nvjp-1* models deviate from other models (mean RMSD of 6.5 Å). Moreover, the results indicate that the pLDDT scores given by AF2 are highly consistent for both *Nvjp-1* and *T7RdhA*, which may therefore serve as a useful feature of the protein structures. We further constructed the residue-residue interaction networks (RINs) starting from the contact maps. The *Nvjp-1* RINs (1000 AF2 models) are dominated by transient interactions that are observed in less than 250 (<25%) AF2 models. In contrast, the *T7RdhA* RINs (1000 AF2 models) comprise mainly persistent interactions that are observed in more than 750 (>75%) AF2 models. Snapshots from MD trajectories exhibit the same patterns for the persistency of interactions in the RINs.

In the following, after summarizing the methods and simulation tools used (Section 2) we will show key results and discussion (Section 3) of this work. We show the intrinsic disorder profiles from different sequence-based predictors for *Nvjp-1*. The AF2 structure models of *Nvjp-1* significantly variate to each other; the *Nvjp-1* configurations from MD trajectories also differ significantly. The mean RMSD is ~7 Å of *Nvjp-1* between two AF2 models or two MD snapshots. Compared to a well-folded globular protein, *T7RdhA*, which has the mean RMSD of ~0.5 Å. Nevertheless, we show that the pLDDT profiles of both *Nvjp-1* and *T7RdhA* are highly consistent, indicating it may serve as a useful feature of proteins. We show a theoretical models of a doughnut-shaped *Nvjp-1* homodimer. We also show the distinct RIN patterns between the IDP (*Nvjp-1*) and the globular protein (*T7RdhA*). More analyses that support the conclusions of the present work (summarized in Section 4) can be found in the supporting information (SI). Potential future directions are also proposed in Section 4.

2. Methods and materials

2.1. Structure predictions and comparisons

The sequence of *Nvjp-1* protein (381 AA) is shown in the SI (S1 Fig). AlphaFold2 (V2.1) [18], together with AlphaFold-Multimer [34], are used for structure predictions of both the *Nvjp-1* monomer and dimer from the primary sequence. We constructed 1000 monomer and 25 dimer models in total. TM-align [38] and MM-align [39] are used to align the monomers and dimers, respectively, to calculate the root-mean-square deviations (RMSD) between each pair of proteins (single chain monomer or double chain dimer). We converted the RMSD matrices into tree-like representations (or clustering) using the R package *APE* [40], which utilize a neighbor-joining algorithm to inform trees from distance matrices in phylogenetics.

2.2. Molecular dynamics simulations

Molecular dynamics (MD) simulations are performed using the NAMD software [41, 42]. The CHARMM force field c36m [43]—which has been built to reflect residual flexibility in IDPs—is utilized for the protein, together with a modified TIP3P model for the solvent [44]. The water box for solvating is at least 15 Å larger than the protein in each of the six directions (X+, X-, Y+, Y-, Z+, and Z-). Zn^{2+} and Cl^- are used to neutralize the system at a concentration of 0.1 M. The Solvate and Autoionization packages of VMD software [45] are used for solvation and ionization, respectively. In the MD simulations, energy minimization was first performed

for 50,000 steps at 0 K. Then the system temperature is increased to 300 K at a rate of 0.001 K/step. A constant-pressure (1 atm), constant-temperature (300 K) NPT ensemble is used in the MD simulation maintained by the Langevin piston controls. The SHAKE algorithm is applied to constrain bonds with H atoms, and a 2 fs step length is used in the MD simulations. The non-bonded interaction cutoff switching is set at between 9 and 11 Å. The particle mesh Ewald summation with a grid spacing of 1.35 Å is applied for the long-range interactions. For each system, a 100 ns production run is performed after a 10 ns equilibration run. Approximately, the sidechain of a His residue has a pK_a of ~6 [46], hence at neutral or basic pH (e.g., >7), either $N_{\epsilon 2}$ or $N_{\delta 1}$ is protonated (neutral charge); whereas at acidic pH (e.g., <5), both $N_{\epsilon 2}$ and $N_{\delta 1}$ are protonated (charge +1). As an IDP, all residues in *Nvjp-1* tend to be exposed in the bulk water owing to their high fluctuations. Most of the simulations were conducted under a neutral pH (~7), under which we used the state where only $N_{\epsilon 2}$ is protonated but $N_{\delta 1}$ is not, which is referred to as **HSE**; we also performed MD of the *Nvjp-1* monomer under an acidic pH, under which both $N_{\epsilon 2}$ and $N_{\delta 1}$ of all His residues are protonated and this state is referred to as **HSP**. Nine monomer trajectories (8 HSE and 1 HSP) and three dimer trajectories (all in HSE), 100 ns each, have been analyzed. We extended the MD simulation for each system to longer time scales, (680 ns for HSE, 800 ns for HSP and 1250 ns for the dimer, see Table 1), and analysis of the final 500 ns trajectories were used for analysis.

Table 1. A summary of interaction numbers from MD simulations (9 for monomer and 3 for dimer). The models shown in the main text are highlighted in bold font.

Name ¹	State ²	Atom# ³	Length ⁴	R_{gyr} ⁵	State ⁶	Total ⁷	Persistent ⁷	Transient ⁷	Med. ⁷
Mono.A.0	Mono	280219	680 ns	28.9±1.5	HSE	4853	726	3376	751
Mono.A.0	Mono	277324	800 ns	41.3±0.9	HSP	5936	570	4818	548
Mono.C.2	Mono	391696	100 ns	34.1±1.1	HSE	4003	480	3048	475
Mono.C.3	Mono	220786	100 ns	28.1±0.6	HSE	3800	572	2573	655
Mono.C.4	Mono	221281	100 ns	26.6±0.4	HSE	3787	587	2554	646
Mono.D.0	Mono	381121	100 ns	31.2±1.0	HSE	3939	507	2829	603
Mono.D.2	Mono	388780	100 ns	41.5±0.7	HSE	3378	497	2309	572
Mono.D.3	Mono	226690	100 ns	26.6±0.8	HSE	3767	587	2498	682
Mono.D.4	Mono	216493	100 ns	25.1±0.6	HSE	4151	532	2975	644
Di.A.0	Dimer	123111	1250 ns	33.8±0.4	HSE	6847	2168	2838	1841
Di.C.2	Dimer	136282	100 ns	31.7±0.2	HSE	6154	1662	3131	1361
Di.D.0	Dimer	147683	100 ns	34.7±0.4	HSE	6380	1686	3334	1360
AF2 ⁸	Mono	-	-	49.2±12.4	-	5982	397	5388	197
T7RdhA ⁹	Mono	-	-	20.9±0.04	-	2352	1850	411	91

- The names (A.0 etc.) refer to the models shown in Figs 1 and 4 in the main text. Mono for monomers and Di for dimers, respectively.
- Only monomers and dimers are modeled in this work, despite higher order of oligomers may exist.
- The Atom number refers to the total number of atoms of the MD system, including water molecules and ions. It is notable that the monomer systems are significantly larger than the dimer systems, which is owing to the relatively large radius of gyration of the monomers, compared to the dimers.
- Length of the MD simulation
- The median±IQR of the radius of gyration (in Å) of the proteins (monomer or dimer) from the last 20 ns of the 100 ns MD trajectories, or the last 500 ns of the three long time-scale MD trajectories (in bold fonts).
- As illustrated in the main text that a binary states of the His residues have been considered: under high pH, all His are mono-protonated on $N_{\epsilon 2}$ (HSE state, neutral), whereas under low pH, all His are double-protonated at both $N_{\epsilon 2}$ and $N_{\delta 1}$ (HSP state, charge +1).
- Total edge numbers include all residue-residue interactions appear in MD trajectory; persistent edges (Strong) are those appear in >75% of all configurations of the MD trajectory (100 ns, or the last 500 ns for longer time-scale MD trajectories); transient edges are those only appear in <25% of all configurations of the MD trajectory; the medium (Med.) strength edges are all other interactions other than the strong and transient ones. The persistent, transient and medium strength edges are colored by red, blue and gray in the RIN figures.
- From 1000 *Nvjp-1* AF2 (V2.2.2) models.
- For a reference, data from 1000 AF2 (V2.2.2) models of a well-folded protein, T7RdhA [37].

<https://doi.org/10.1371/journal.pone.0301866.t001>

2.3. Residue interaction network

Residue interaction networks (RINs) are constructed based on the contact maps to capture the interactions contributed from different residues. The RINs are constructed for each AF2 models, or for the MD trajectories, RINs are calculated from the snapshots taken every 1 ns after equilibrations. Details of the RIN construction are summarized in a previous work [37]. Briefly, the distance between the two residues R_i and R_j (denoted as D_{ij}) is defined as the shortest distance between all non-hydrogen atoms the two residues. For each AF2 model or a configuration taken from MD trajectories, an adjacency matrix is estimated from the distance matrix $[D_{ij}]$ under the cutoff of 3.5 Å: $A_{ij} = 1$ if D_{ij} is less than 3.5 Å or $A_{ij} = 0$ otherwise. Without considering the directions (which may be important for certain interactions such as hydrogen-bonds) or weights (including attraction versus repulsion), this binary adjacency matrix can only be transformed into an undirected, unweighted network, or a contact map, in which the indices are amino acid residues and the edges are 1 for an interaction or 0 for no interaction. However, using ensembles of configurations (multiple AF2 models or snapshots from MD simulations), a weighted residue interaction network can be constructed by adding all binary adjacency matrices together. The *igraph* R package is used for network analysis [47]. The residue-residue distances are calculated using the *bio3D* R package [48]. An interaction is considered “persistent” if it appears in more than 75% of all models (either AF2 models or configurations from MD trajectories), and it is considered “transient” if it appears in less than 25% of all models. See Results and Table 1 for details.

2.4. Comparisons between *Nvjp-1* and a well-folded protein T7RdhA

Besides the 30 *Nvjp-1* models using the original AF2 code (V2.0.1, summarized in Table 1 and Fig 1) [18], we further built 1000 AF2 models using a more recent AF2 code (V2.2.2) [18] and compared with 1000 AF2 models of a well-folded protein, the reductive dehalogenase from *TMED77 acidimicrobiaceae* (T7RdhA) [37].

2.5. Other tools

The STRIDE program is used to predict the secondary structure elements in the proteins via the *bio3d* R package [49]. RGN2, a recent deep-learning tool using a language model is also applied to predict the *Nvjp-1* structure [50].

3. Results and discussion

3.1. *Nvjp-1* is an intrinsically disordered protein

We used different predictors to evaluate the intrinsic disorder contents (D_i) of *Nvjp-1* (Fig 1A). The disorder profiles are consistent: the Pearson’s correlation coefficients, $PCC = 0.44 \pm 0.51$ (median±IQR), and median p-value of 2.0×10^{-19} , with nearly all residues having a $D_i > 0.5$; the mean disorder profile from all 7 predictors has a $D_i = 0.85 \pm 0.09$, indicating that *Nvjp-1* protein is a fully disordered protein. We used AF2 to model the *Nvjp-1* structures: 6 independent runs (A to F) produced 5 models (0 to 4) each, yielding 30 AF2 models in total. We noticed that each of the 30 models structurally deviated from the other models. To verify this, we measured the RMSDs among all 30 protein models using TM-align, and converted the RMSD matrix (30×30) to a phylogenetic tree (Fig 1B). The RMSD matrix has values of 6.97 ± 0.83 Å, confirming the above observation. The structure of one of the models (A.0) is shown in Fig 1D, colored by the per-residue pLDDT scores, which indicates the low confidences given by AF2. Moreover, the pLDDT scores provided by AF2 do not correlate with any of the disorder profiles

assessed here. As a comparison, in a recent work [37], 320 AF2 models of a well-folded globular protein have overall RMSDs of 0.47 ± 0.15 Å, or at the “atomic accuracy” [19].

3.2. Structural heterogeneity of *Nvjp-1* monomer and the pH effect

For a globular protein, multiple AF2 models agree with the configuration space sampled by MD simulations [37], indicating the ability of AF2 for capturing different, and (potentially) biologically relevant, conformations. However, for *Nvjp-1*, it is difficult to tell which model generated by AF2 or which snapshot from MD simulation is of relevance. For the monomers, starting from a same AF2 structure (A.0 in Fig 1B), we perform two independent simulations with a) all His residues double protonated (i.e., in +1 charge states) or single protonated (neutral charge), to mimic acidic (pH 3) and basic (pH 8) conditions, respectively. It has been shown that *Nvjp-1* dissolves at low pH but precipitates at high pH, mainly owing to the high percentage (27 mol%) of histidine residues [33]. A recent report using MD simulation indicates that *Nvjp-1* has a larger radius of gyration (R_{gyr}) at low pH (3) compared to high pH (8) [56]. Consistent with this report [56], our simulations (taken from the final 500 ns from long MD trajectories) show that *Nvjp-1* monomer has a considerably larger R_{gyr} (41.3 ± 0.9 Å) at the HSP state (acidic pH) compared to HSE state (neutral pH, $R_{\text{gyr}} = 28.9 \pm 1.5$ Å), likely caused by the repulsions among the positively charged His residues (Fig 2A). Note that the Asp and Glu residues remain deprotonated in both conditions in this work, despite their pK_a values. However, the initial AF2 models (before MD simulations) have even larger R_{gyr} (49.2 ± 11.0 Å from 30 models generated by AF2 2.0.1, or 49.2 ± 12.4 Å from 1000 models generated by AF2 V2.2.2) values. The R_{gyr} of the initial AF2 and MD simulation models are larger than the R_{gyr} computed in the previous *Nvjp-1* monomer model (~ 22 Å) which utilized a polarizable force field to conduct the MD simulations [56]. For a comparison, the R_{gyr} of a well-folded protein, T7RdhA [37], is 20.9 ± 0.04 Å (Table 1).

We further compared the configurations taken from MD trajectories at low or high pH. The configurations significantly deviate from other configurations collected from the same MD trajectory (Fig 2B), and all pairs of structures show large RMSDs (7.14 ± 0.57 Å, Fig 2C). Similar to Fig 1, the scale bar in Fig 2B (0.5 Å) serves as a reference for the “atomic resolution”. The snapshots after 100 ns MD simulations of both high (top) and low (bottom) pH MD are shown in Fig 2D. Note that in contrast to previous studies [32, 56], little or no apparent secondary structures appear in *Nvjp-1* monomer during the MD simulations, supporting the fact that *Nvjp-1* is an IDP.

3.3. PAE map, distance map and residue interaction network

For the constructed models, AF2 provides the predicted aligned error (PAE) map to quantify the potential distance errors in the model: [18] if the i -th residue is aligned to the “real” (ground truth) model, the (potential) error, or the PAE value at the j -th residue is P_{ij} (in Å). The PAE matrix is asymmetric, i.e., P_{ij} and P_{ji} may not be equal. We previously suggested that the PAE map originates from the protein dynamics [57]. Here, we show that the PAE map is highly consistent with the distance map calculated from the MD trajectory (under both high and low pH, Fig 3A–3C). The distance maps can be converted to a binary contact map, then further visualized as residue interaction networks (RINs, see Methods section). In this network, the vertices are the residues and the edges are the interactions: an interaction exists if the shortest distance between the non-H atoms of residues R_i and R_j both residues is shorter than 3.5 Å [37]. The RINs indicate that in addition to the backbone-interactions (e.g., two neighboring residues are connected during the MD simulation), the majority of residue-residue interactions are short-lived, or transient, which are colored in blue in Fig 3D/3E, or S2 Fig in the SI. It

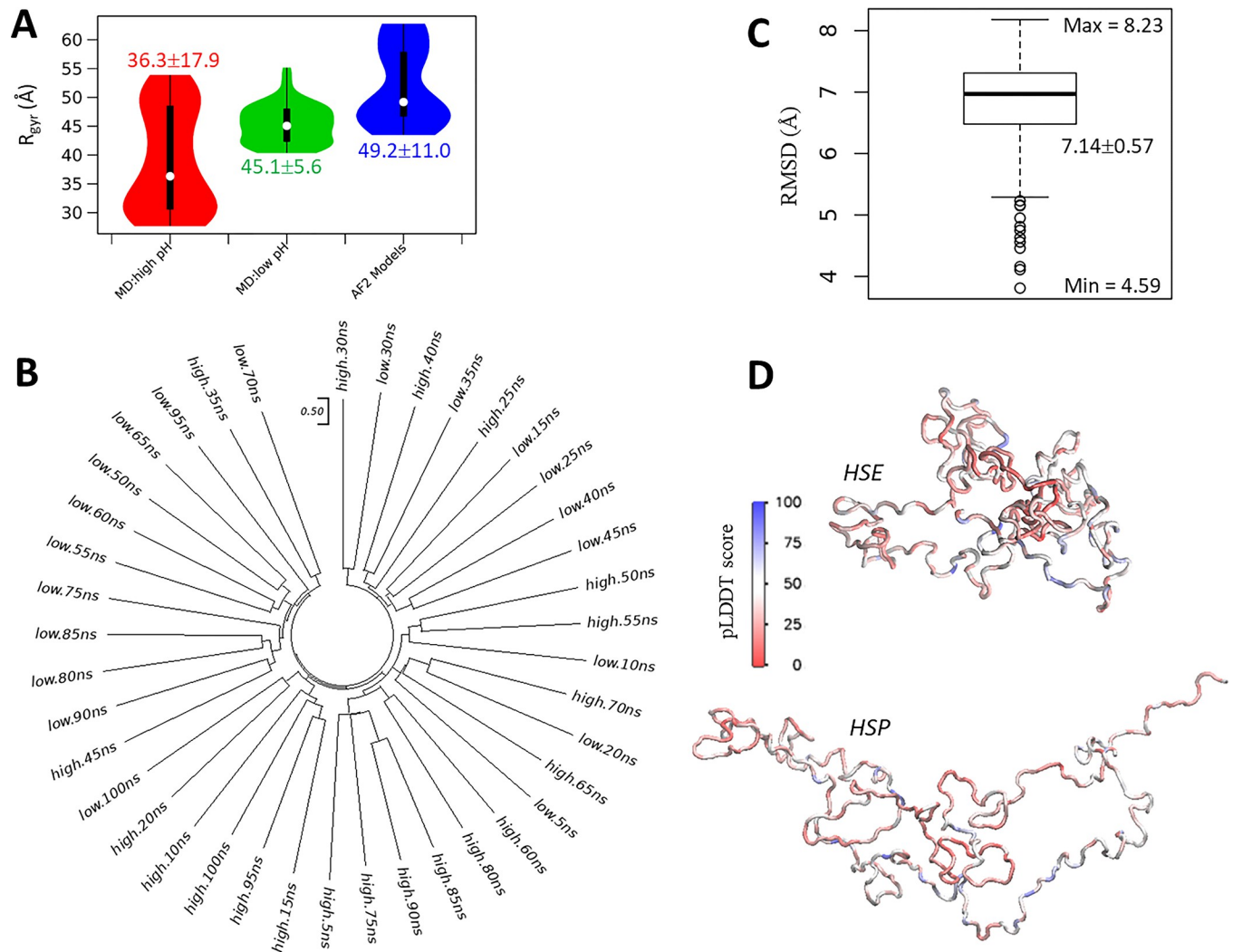


Fig 2. The effect of pH on Nvjp-1 monomer structure. (A) Violin plots for distribution of radius of gyration (R_{gyr} , Å) of protein structures taken from MD simulations under high pH (red, left), low pH (green, middle), compared with the AF2 models shown in Fig 1. (B) Structure-based phylogenetic tree using snapshots taken from 100 ns MD simulations under both HSE (neutral pH) and HSP (acidic pH) states. A scale-bar of RMSD = 0.50 Å is provided as a reference for the “atomic resolution” observed in well-folded proteins. (C) A boxplot of the RMSD values calculated from all pairs of structures in B. (D) Nvjp-1 monomer snapshots after 100 ns MD performed under HSE (neutral pH, top) and HSP (acidic pH, bottom) states. The structures are colored by the pLDDT scores of the original AF2 structure (A.0 in Fig 1) before the MD simulation.

<https://doi.org/10.1371/journal.pone.0301866.g002>

is interesting that the residue-residue interactions are dominated by the contributions from both Gly and His, the two most abundant residues in Nvjp-1. In addition, under HSE state (neutral pH), the R_{gyr} of Nvjp-1 monomer becomes smaller, and more residue interactions are observed compared to HSP state, or acidic pH conditions.

The results shown above are based on MD simulations from the A.0 structure in Fig 1B. Despite the structural heterogeneity, the results based on other structures are largely consistent, see S3 Fig in the SI for the results from two other models.

3.4. Nvjp-1 dimer: AF2 structure, distance map and RIN

In protein purification, previous experimental work (Dennis et al., unpublished) observed a potential dimer band from SDS-PAGE, which can sustain high temperature in SDS detergent

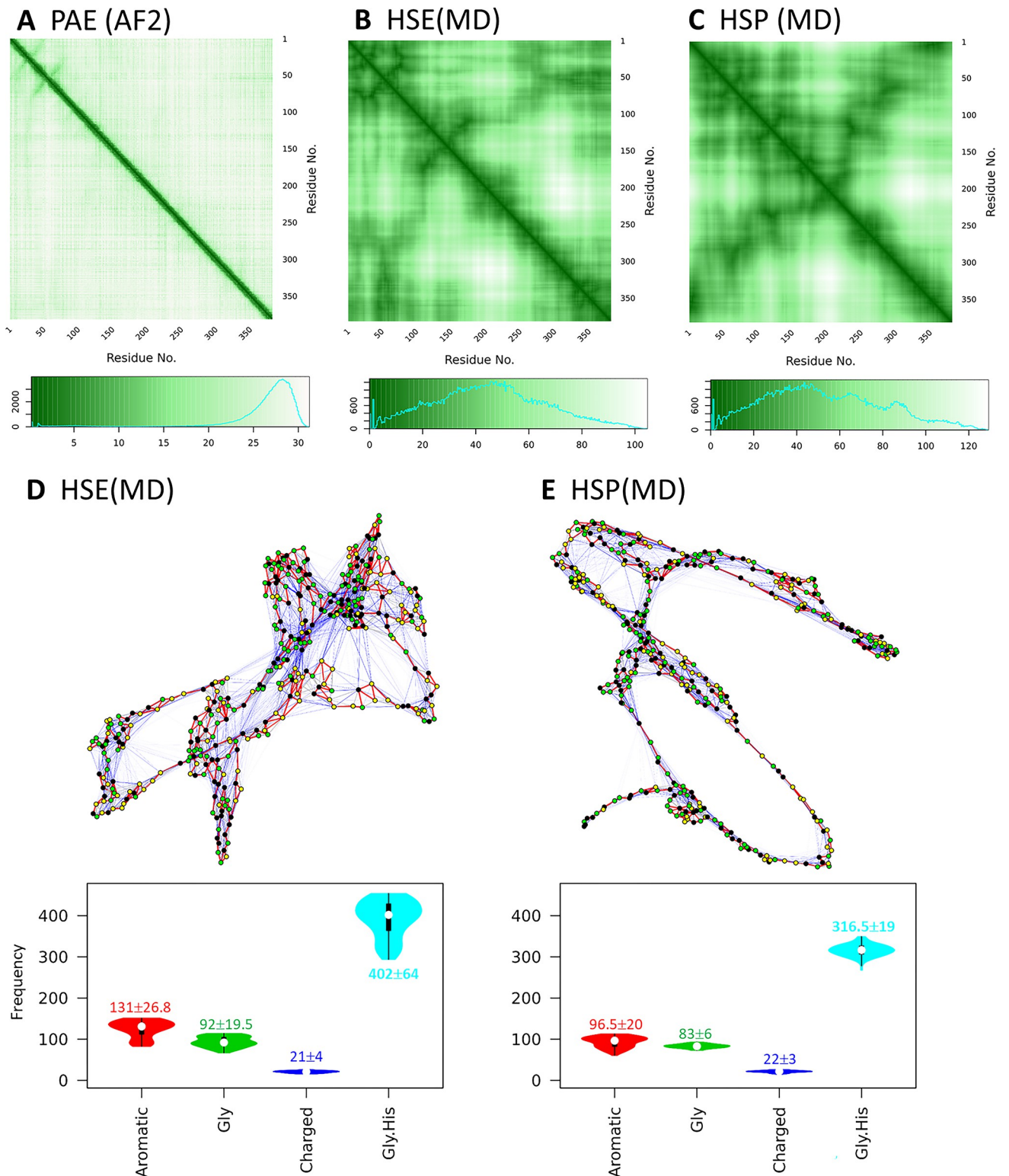


Fig 3. PAE, distance map and residue interaction network of *Nvjp-1* monomer. (A) PAE (predicted aligned error) map of one of the AF2 models (A.0 in Fig 1). The distance maps averaged from 100 ns MD simulations under (B) HSE (neutral pH) state and (C) HSP (acidic pH) state. Weighted residue interaction networks (RINs) under (D) HSE and (E) HSP states. In the RINs, green nodes are Gly, black nodes are aromatic (H, Y and F) residues of *Nvjp-1*. In the MD trajectory, the edges (3821 for HSE and 3295 for HSP) appear in >75% of all configurations are persistent interactions, and are shown in red (518, 13.6% for HSE; and 488, 14.8% for HSP), edges appear in <25% of all configurations are transient interactions and are in blue (2733, 71.5% for HSE;

and 2320, 70.4% for HSP), and those in between are in gray (570, 14.9% for HSE; and 487, 14.8% for HSP). A circular representation of the RINs can be found in S2 Fig of the SI. Violin plots of interaction frequencies contributed from aromatic residues (H, Y, F), Glycine, charged residues (D, E, R, K) or Glycine and Histidine are shown. Median \pm IQR are shown for the interactions during the 100 ns MD.

<https://doi.org/10.1371/journal.pone.0301866.g003>

(unpublished). Here, using AF2-multimer, theoretical *Nvjp-1* homodimer models were predicted by AlphaFold-Multimer, which are doughnut-shaped (Fig 4). Similar to the monomer, the dimeric structures (25 models in total) deviated from each other with an RMSD of 6.91 ± 0.65 Å (Fig 4A/4B). In addition, little or no apparent secondary structural elements were observed in the *Nvjp-1* dimers (Fig 4C). We compared the structural similarities among individual monomers constituting the dimers (50 monomers in total) and observed the same trend for the single *Nvjp-1* chains within the dimers, i.e., all individual dimer chains showed distinct structures, with an RMSD of 6.91 ± 0.72 Å (S4 Fig in SI).

The AlphaFold-Multimer models of *Nvjp-1* homodimers have a smaller R_{gyr} compared to the AF2 models of *Nvjp-1* monomers. Therefore, the *Nvjp-1* monomers are considerably larger than the dimers (Table 1). Interestingly, during the MD simulations, the R_{gyr} of the *Nvjp-1* homodimer is larger than the monomer at high pH, but is smaller than the monomer at low

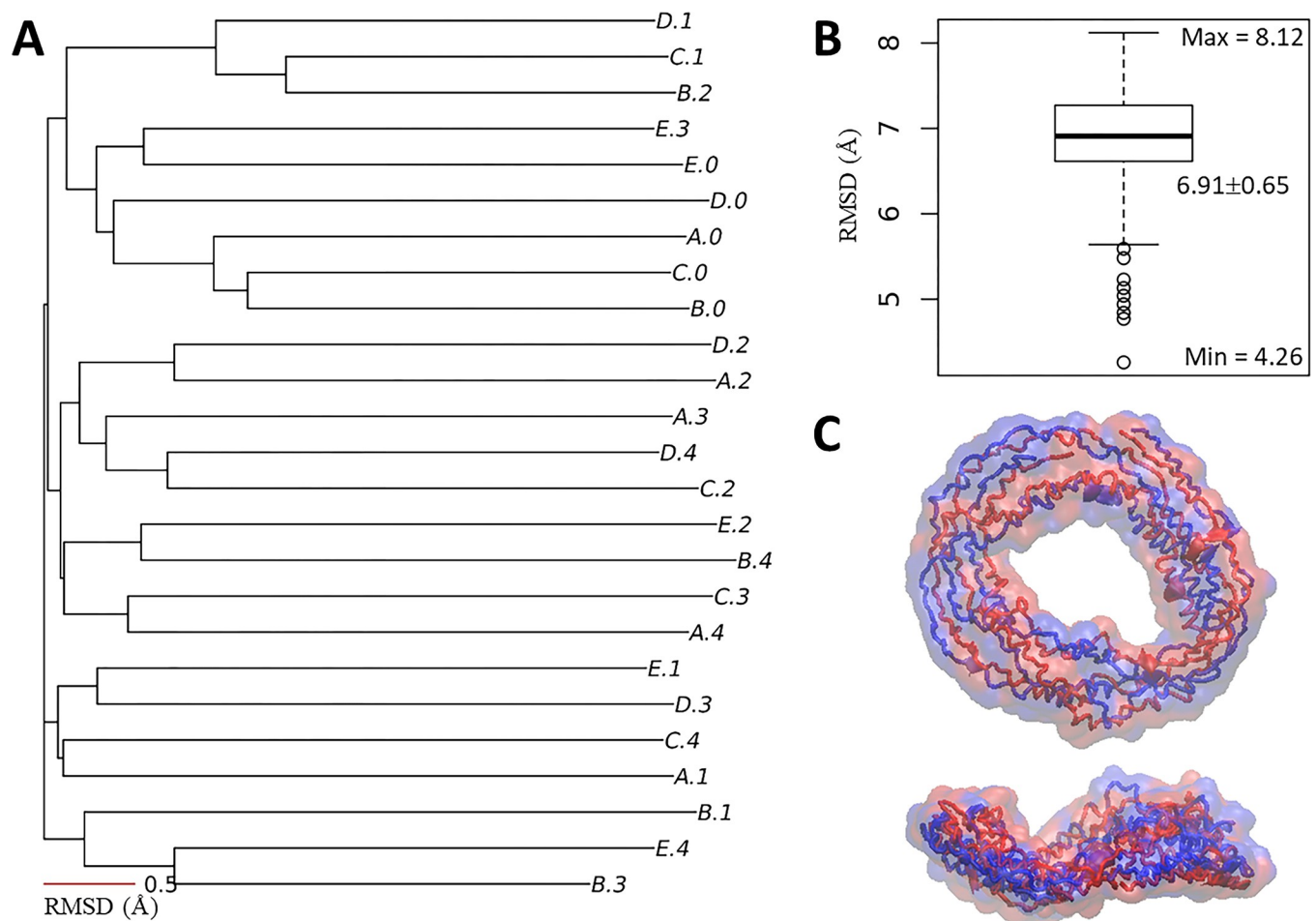


Fig 4. *Nvjp-1* homodimer structure. (A) Structure-based phylogenetic tree of 25 AF2 models from five independent runs. (B) A box plot of the RMSD matrix for the tree in A. (C) The doughnut-shaped *Nvjp-1* dimer structure; top: front view, bottom: side view. The two chains of the dimer are colored in blue and red, respectively.

<https://doi.org/10.1371/journal.pone.0301866.g004>

pH. However, the fluctuation of R_{gyr} of the *Nvjp-1* dimer is considerably smaller than that of the monomer under both pH conditions (Fig 5A), indicating that the *Nvjp-1* dimer is relatively rigid compared to the monomers. The R_{gyr} of the monomers and dimers of MD trajectories of

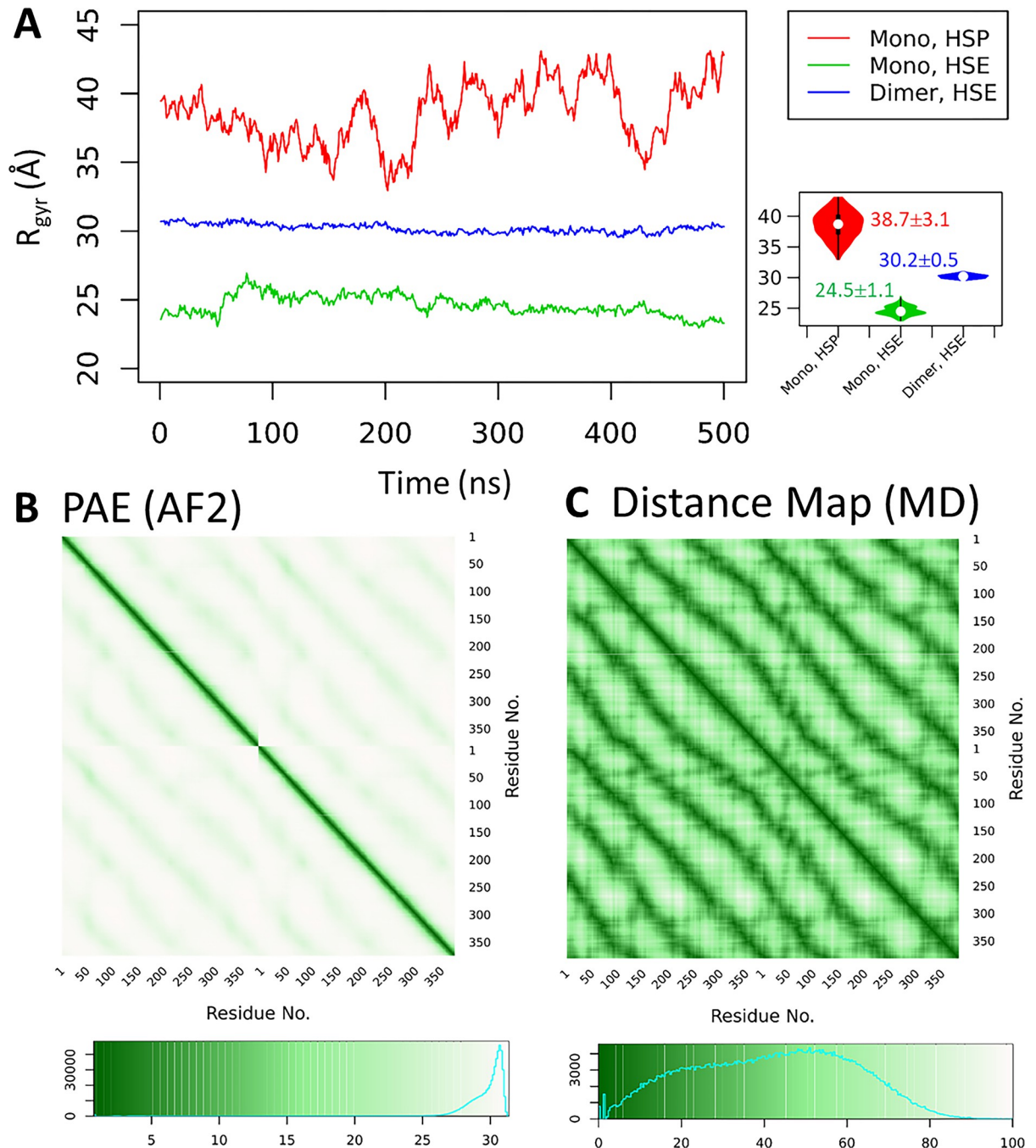


Fig 5. A rigid *Nvjp-1* dimer. (A) The radius of gyration (R_{gyr}) of the dimer is larger than the monomer under HSE state but smaller than the monomer under HSP state. Smaller fluctuations of R_{gyr} in the dimer indicates this structure is relatively rigid compared to the monomers (under both HSE and HSP states). The inset shows violin plots of the R_{gyr} . (B) The PAE map from AF2 and (C) the distance map averaged from the snapshots from a 500 ns MD exhibit a consistent pattern indicating each of the two chains loops for two circles in the doughnut-shaped *Nvjp-1* dimer.

<https://doi.org/10.1371/journal.pone.0301866.g005>

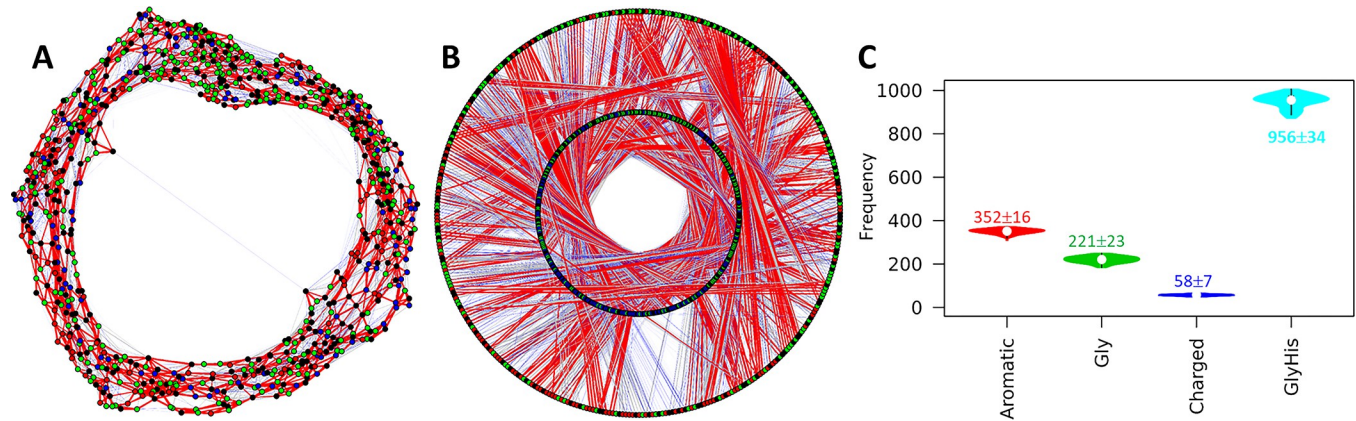


Fig 6. Residue interaction network (RIN) of *Nvjp-1* dimer. (A) A randomized RIN and (B) a circular representation of the RIN of *Nvjp-1* dimer. The Gly and His residues are in green and black, and other residues from chain A and B are in blue and red, respectively. In the circular presentation, nodes in the inner circle are from chain A and nodes in the outer circle are from chain B, respectively. The RIN has 6847 edges in total. In the MD trajectory, the edges that appear in >75% of configurations (strong interactions) are shown in red (2168, 31.7%), the edges that appear in <25% of all configurations (transient interactions) are in blue (2838, 41.4%), and those in between are in gray (1841, 26.9%). (C) Violin plots of interaction frequencies contributed from aromatic residues (H, Y, F), Glycine, charged residues (D, E, R, K) or Glycine and Histidine are shown. Median±IQR are shown for the interactions during the 100 ns MD. The RINs of two other *Nvjp-1* dimer models (Table 1) are shown in S5 Fig of SI.

<https://doi.org/10.1371/journal.pone.0301866.g006>

other systems are summarized in Table 1. The inconsistency of MD simulations may also be owing to the intrinsic disorder nature of the *Nvjp-1* protein, however, for the dimers, the relatively small fluctuations in R_{gyr} suggest a stabilization effect upon dimerization. The last 500 ns of the MD trajectories (680–1250 ns) were used for Fig 5A.

Further comparison of the *Nvjp-1* dimer PAE and distance maps demonstrated an apparent pattern (Fig 5B/5C). In addition to a strong diagonal line contributed by the interactions (or affiliations) of consecutive residues, which is also seen in the PAE and distance maps of the monomers (Fig 3A–3C), a periodic interaction trend emerges from each chain as it loops twice around the dimer doughnut (Fig 5B and 5C).

We constructed the RIN of the *Nvjp-1* dimer using the same criterion as the monomer (Fig 6) and observed the total interaction number of the dimer RIN was roughly double that of the monomer RIN (6847 for dimer model A.0, and 3821 for monomer model A.0 at HSE state). However, the interactions that persisted in over 50% of the configurations (i.e., strong interactions), taken from the MD trajectory, significantly increased in the dimer. After removing the backbone interactions (i.e., interactions from consecutive residues), the *Nvjp-1* dimer demonstrated 1408 strong interactions, which was over four times than those in the monomer (343 at high pH and 293 at HSP state). Moreover, the monomer had a higher ratio (71.5% at HSE state, 70.4% at HSP state) of transient interactions than that of the dimer (41.4%, HSE). See Table 1 for a summary of interaction numbers from the MD simulations. The RINs of two other *Nvjp-1* dimer models are shown in S5 Fig of SI.

3.5. Consistent pLDDT profiles predicted by AF2

Above discussions showed the inconsistency in the AF2 structure predictions of *Nvjp-1*, and possibly of other IDPs. In contrast, we recently found that multiple AF2 models of a well-folded protein, *T7RdhA*, are highly consistent, with mean RMSD lower than 1 Å, i.e., within “atomic resolution” [19]. We reconstructed 1000 AF2 models for both the disordered (*Nvjp-1*) and ordered (*T7RdhA*) proteins, and plotted the RMSD profiles against a randomly chosen model (Fig 7A). The large RMSD values of *Nvjp-1* models (6.5 ± 1.1 Å, also see Fig 2), compared to those of *T7RdhA* models (0.5 ± 0.1 Å) suggest that an ensemble of structures of an IDP (such

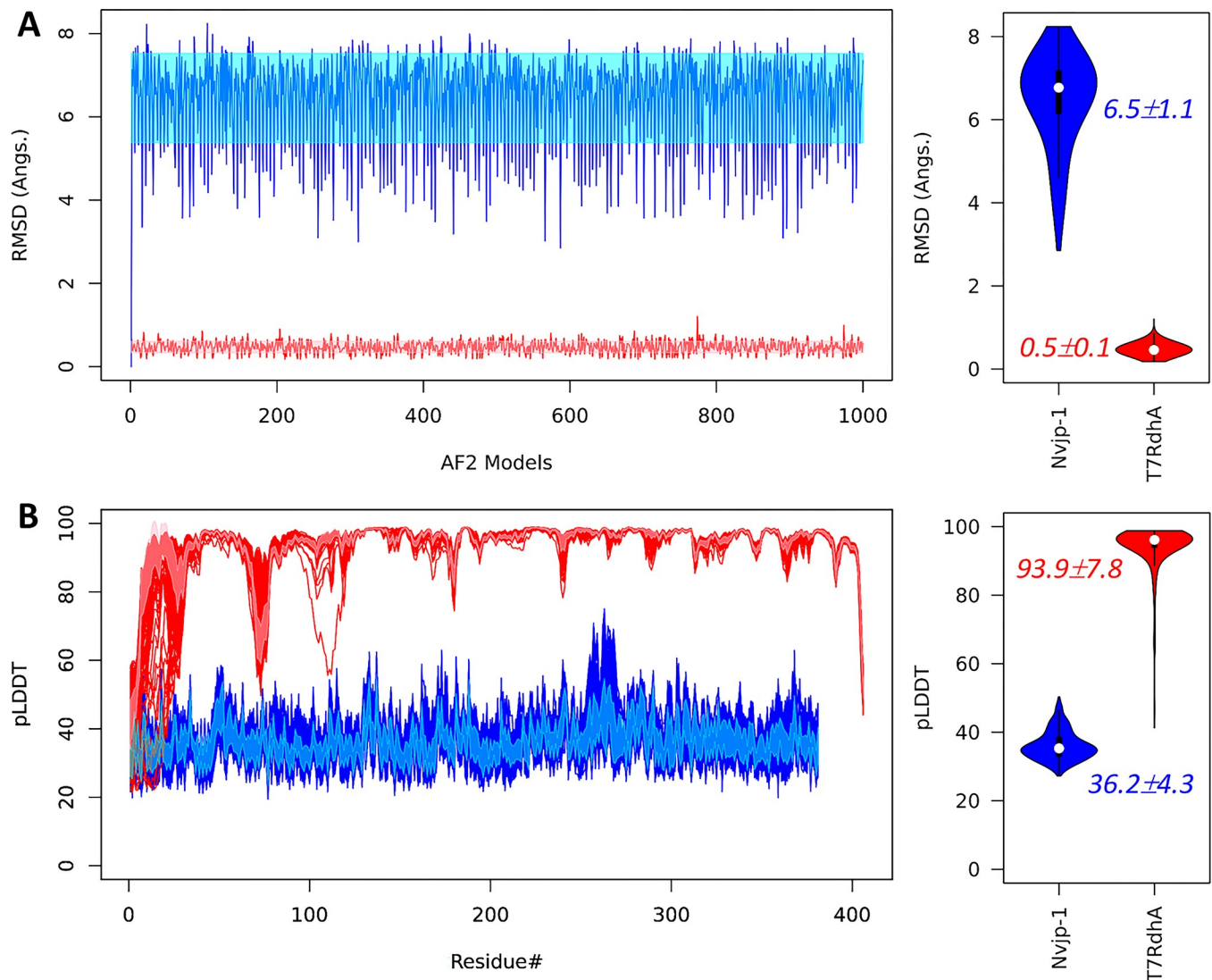


Fig 7. (A) RMSD fluctuations of 1000 AF2 models against a randomly selected model indicate high variation in structures of *Nvjp-1* (blue), and high consistency in structures of a well-folded protein, *T7RdhA* (red); (B) the per-residue pLDDT scores of *Nvjp-1* (381 AA, blue) and *T7RdhA* (406 AA, red) of 1000 AF2 models. The shaded areas of *Nvjp-1* (cyan) and *T7RdhA* (pink) profiles in the left panels illustrate the mean \pm sd. Violin plots of the RMSD and mean pLDDT profiles are shown in the right panels for both *Nvjp-1* (blue) and *T7RdhA* (red).

<https://doi.org/10.1371/journal.pone.0301866.g007>

as *Nvjp-1*), without inferences from other molecules, may sample a relatively diverse conformational space, either from AF2 structure modeling (Figs 1, 4 and 7) or from MD simulations (Fig 2). Whereas an ensemble of structures of an ordered protein (such as *T7RdhA*) exhibit relatively high consistency (Fig 7A). We also compared the distributions of the secondary structure elements in both *Nvjp-1* and *T7RdhA* in their respective 1000 AF2 models, as shown in S6 Fig in the SI. The *Nvjp-1* residues are mostly in the coil or turn states, whereas significant ratios of helices and sheets structures are observed in *TRdhA* residues.

Nevertheless, we found that the pLDDT profiles from AF2 models of both IDP (*Nvjp-1*) and ordered protein (*T7RdhA*) are highly consistent (Fig 7B). As discussed earlier that the pLDDT scores in AF2 models reflect the fluctuations of the residues [57], we suggest that, despite it is difficult to construct an ensemble of conformations to describe the characteristics

and functions of an IDP [58], the pLDDT profile may serve as a graphic, descriptive attribute, or feature.

In AF2 modeling, no multi-sequence alignment (MSA) hits was observed for *Nvjp-1*, i.e., it may be an orphan protein, or an orphan IDP. A recent language model RGN2 claimed that for the orphan proteins, it beats both AF2 [18] and RoseTTAFold [59] in both accuracy and speed [50]. The *Nvjp-1* models generated by RGN2 showed no apparent secondary structures (S7 Fig in the SI), in agreement with the AF2 models. Despite the potential “orphan” state and the intrinsic disorder of *Nvjp-1*, recent studies showed that, however, protein structures predicted from randomized protein sequences possess significant amount of ordered regions [57, 60]. This suggest that intrinsically disordered proteins do not appear by chance.

3.6. Ordered versus disordered, protein dynamics and AF2 models

As shown above, the residue-residue interactions in *Nvjp-1* are mainly transient. Even for different *Nvjp-1* AF2 models, we show that the majority of the calculated RINs are also transient. In contrast, the RINs in the AF2 models of the relatively well-folded protein, T7RdhA, are mainly persistent (Fig 8). For each RIN, the persistency is defined as the percentage of configurations in which the interaction can be observed. The *Nvjp-1* models has more residue-residue interactions than *T7RdhA* (Table 1), however, many of these interactions present in only a limited number of models (or a single model), but are absent in other models. Persistent interactions in *Nvjp-1* (Fig 8C, red bar) are mainly contributed by the peptide bonds between two

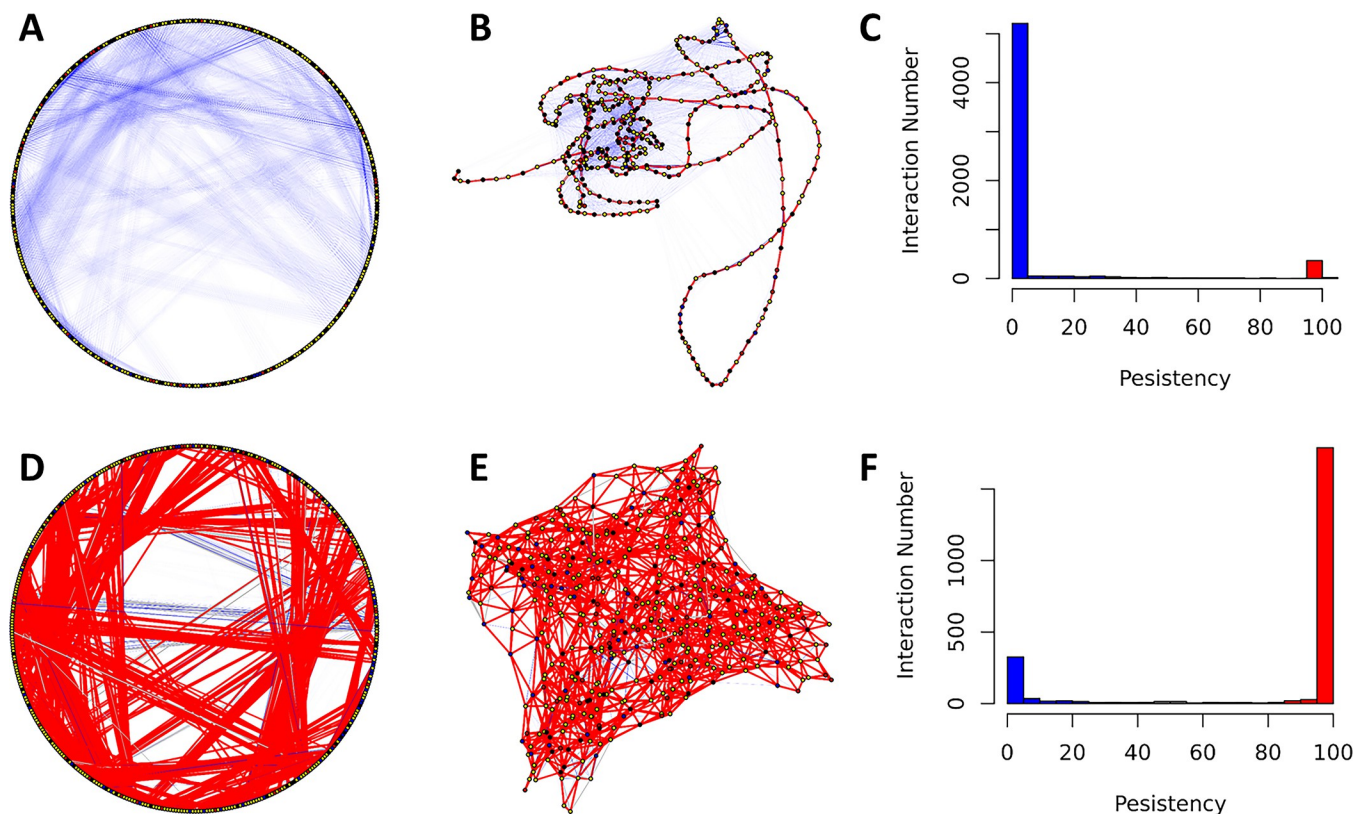


Fig 8. Ordered versus disordered: persistent and transient interactions in *Nvjp-1* (A-C, top) and *T7RdhA* (D-F, bottom). Circular (A and D) and randomized (B and E) RIN models are shown in which the persistent interactions (persistence $\geq 75\%$) are in red and transient interactions (persistence $\leq 25\%$) are in blue, and other interactions in gray. Histograms of the persistency (percentages) are also shown, similarly persistent in red and transient in blue.

<https://doi.org/10.1371/journal.pone.0301866.g008>

consequential residues. Instead, in T7RdhA (Fig 8F), majority of the interactions are persistent, and only a small number of interactions are transient.

We further analyzed the persistency of residue-residue interactions throughout the MD simulations (also see Figs 3 and 6). In this analysis, we take snapshots every single ns during the last 500 ns MD trajectories of either the monomer or dimer simulation, and the RINs were calculated based on these 500 configurations. Noticeable differences between AF2 models (Fig 8A–8C) and MD configurations (Fig 9A–9C) can be found, i.e., there are apparently more persistent interactions from MD trajectory than those in the AF2 models. The main reason contributing to this differences might be that all AF2 models have been “relaxed” using energy minimizations and hence represent locally optimized (yet static) structures on the potential energy surface. For the configurations from MD, however, they are in constantly moving states governed by physical laws, and the persistency of RINs is restrained by the interaction potentials which may hold certain interactions (e.g., salt-bridges) longer than others. Nevertheless, the RINs in T7RdhA are highly persistent (Fig 8F), in line with the high pLDDT scores predicted by AF2 (Fig 7B). For *Nvjp-1*, the dominance of transient RINs (Fig 8C) is also consistent with the low pLDDT scores (Fig 7B). Therefore, the pLDDT scores can be used to interpret the confidence of the structures, as well as residue-residue interactions.

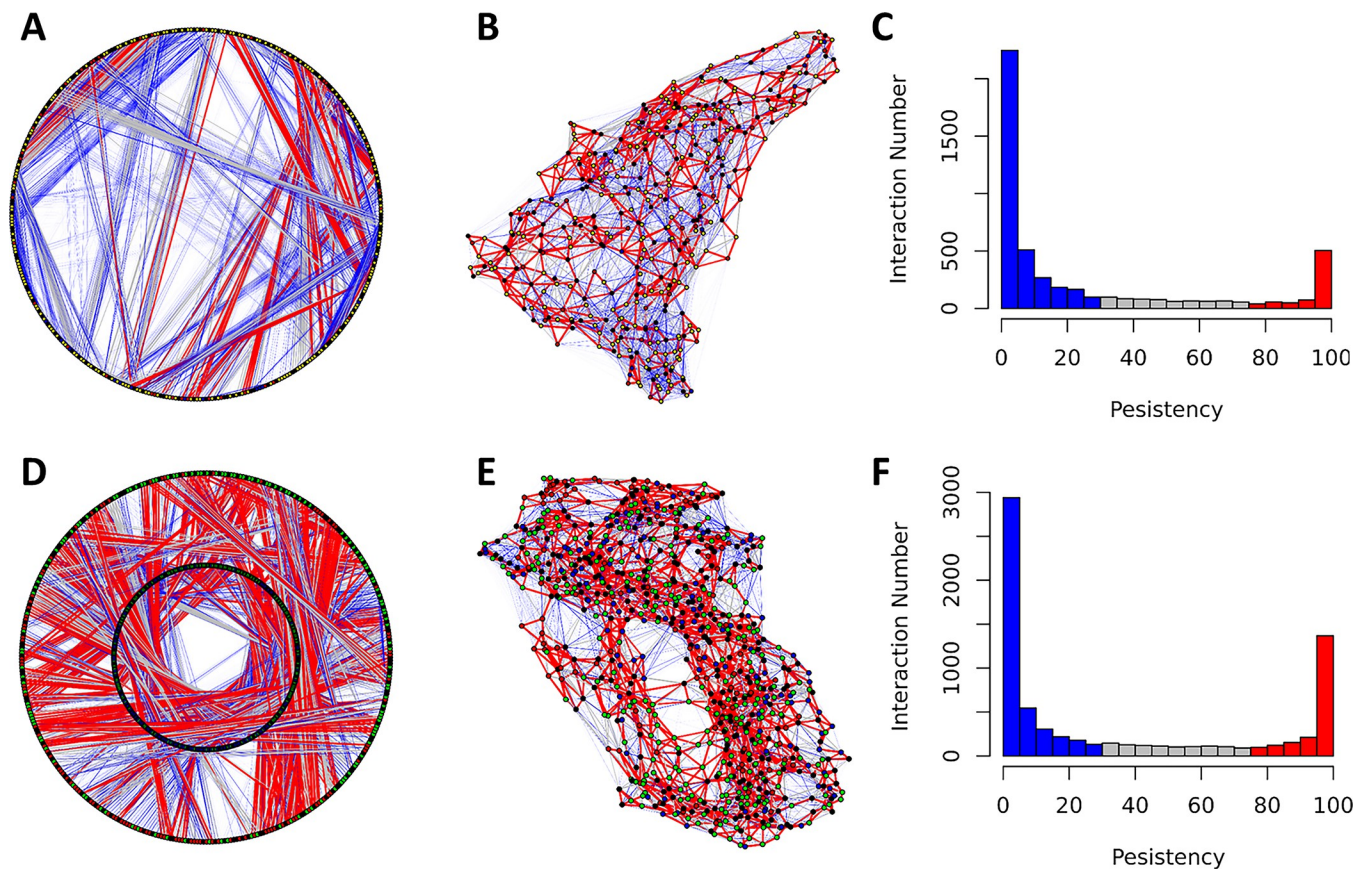


Fig 9. Residue-residue interaction networks from MD for *Nvjp-1* monomer (A-C, HSE) and dimer (D-F, HSE).

<https://doi.org/10.1371/journal.pone.0301866.g009>

4. Conclusions and future directions

Protein disorder is related to many factors such as flexibility [61], fuzziness [13], and entropy [60]. The residual disorder contents can be predicted from the protein sequence and hence are “intrinsic” [6]. Here, aided by AF2 and MD simulations, we propose an interesting aspect of protein disorder: 1) structure prediction is not reproducible by AF2, i.e., two independent AF2 predictions do not converge on a solution such as two models with low RMSD (at the atomic resolution); 2) the configurations from MD simulations do not overlap with each other (e.g., $\text{RMSD} < 1 \text{ \AA}$). The target of the present work, *Nvjp-1*, is an IDP in both monomeric and dimeric forms. Despite the relatively confined dynamics in the dimers, the *Nvjp-1* protein models (either from AF2 or from MD) are not similar to each other in both monomers and dimers, i.e., they are disordered. Using 1000 AF2 models, *Nvjp-1* is compared with *T7RdhA*, a well-folded globular protein. Unlike *Nvjp-1*, AF2 models of *T7RdhA* are consistent with the RMSD of $\sim 0.5 \text{ \AA}$. We observed that for both proteins, either disordered (*Nvjp-1*) or ordered (*T7RdhA*), AF2 produces highly consistent pLDDT profiles for all 1000 models, rendering it a potentially useful feature of the protein structures, or disorders. Moreover, in *Nvjp-1* AF2 models the residue-residue interactions are mostly transient, whereas the majority of those in *T7RdhA* are persistent.

The consistent residual pLDDT profiles predicted by AF2 for both disordered and ordered proteins suggest that the residual flexibilities and related protein dynamics may be encoded in the protein sequences [57]. Note that the 1000 AF2 models of *Nvjp-1* and *T7RdhA* were constructed by multiple AF2 runs with variable random seeds. It was suggested that this approach may uncover alternative functional configurations [18], including rare configurations [37]. With the recent advances, especially AF-cluster [62], we will be able to explore deeper in the folding spaces of interested proteins. Moreover, the physical adjacency-based RIN presented in this work will be applied to other systems, including the protein-ligand and protein-protein interactions, with potential applications to drug discoveries.

Supporting information

S1 Fig. The amino acid sequence of *Nvjp-1*.

(TIF)

S2 Fig. Circular representation of the residue interaction networks.

(TIF)

S3 Fig. Residue interaction networks from 100 ns MD trajectories of AF2 models.

(TIF)

S4 Fig. Comparison of the single chains in 25 *Nvjp-1* homodimers.

(TIF)

S5 Fig. Residue interaction network from the MD trajectory of *Nvjp-1* dimer models C.2 and D.0.

(TIF)

S6 Fig. Percentages of secondary structure elements (SSEs) estimated in 1000 AF2 models.

(TIF)

S7 Fig. The RGN2 model of *Nvjp-1* monomer.

(TIF)

Acknowledgments

The structural modeling and MD simulations were performed using the DoD HPC.

Author Contributions

Conceptualization: Hao-Bo Guo, Rajiv Berry.

Data curation: Hao-Bo Guo, Baxter Huntington, Alexander Perminov, Kenya Smith, Nicholas Hastings, Rajiv Berry.

Formal analysis: Hao-Bo Guo.

Investigation: Rajiv Berry.

Methodology: Hao-Bo Guo.

Project administration: Nancy Kelley-Loughnane.

Resources: Nancy Kelley-Loughnane, Rajiv Berry.

Software: Hao-Bo Guo.

Supervision: Rajiv Berry.

Validation: Rajiv Berry.

Visualization: Hao-Bo Guo.

Writing – original draft: Hao-Bo Guo.

Writing – review & editing: Baxter Huntington, Alexander Perminov, Patrick Dennis, Nancy Kelley-Loughnane, Rajiv Berry.

References

1. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*. 2000; 41(3):415–27. Epub 2000/10/12. [https://doi.org/10.1002/1097-0134\(20001115\)41:3<415::aid-prot130>3.0.co;2-7](https://doi.org/10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7) PMID: 11025552.
2. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*. 2002; 11(4):739–56. Epub 2002/03/23. <https://doi.org/10.1110/ps.4210102> PMID: 11910019; PubMed Central PMCID: PMC2373528.
3. Uversky VN. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci*. 2013; 22(6):693–724. Epub 2013/04/05. <https://doi.org/10.1002/pro.2261> PMID: 23553817; PubMed Central PMCID: PMC3690711.
4. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*. 1999; 293(2):321–31. Epub 1999/11/05. <https://doi.org/10.1006/jmbi.1999.3110> PMID: 10550212.
5. Fink AL. Natively unfolded proteins. *Curr Opin Struct Biol*. 2005; 15(1):35–41. Epub 2005/02/19. <https://doi.org/10.1016/j.sbi.2005.01.002> PMID: 15718131.
6. Dunker AK, Babu MM, Barbar E, Blackledge M, Bondos SE, Dosztanyi Z, et al. What's in a name? Why these proteins are intrinsically disordered: Why these proteins are intrinsically disordered. *Intrinsically Disord Proteins*. 2013; 1(1):e24157. Epub 2013/04/01. <https://doi.org/10.4161/idp.24157> PMID: 28516007; PubMed Central PMCID: PMC5424803.
7. Uversky VN, Kulkarni P. Intrinsically disordered proteins: Chronology of a discovery. *Biophys Chem*. 2021; 279:106694. Epub 2021/10/05. <https://doi.org/10.1016/j.bpc.2021.106694> PMID: 34607199.
8. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol*. 2017; 18(5):285–98. Epub 2017/02/23. <https://doi.org/10.1038/nrm.2017.7> PMID: 28225081; PubMed Central PMCID: PMC7434221.
9. Tompa P, Schad E, Tantos A, Kalmar L. Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struct Biol*. 2015; 35:49–59. Epub 2015/09/25. <https://doi.org/10.1016/j.sbi.2015.08.009> PMID: 26402567.

10. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol.* 2015; 16(1):18–29. Epub 2014/12/23. <https://doi.org/10.1038/nrm3920> PMID: 25531225; PubMed Central PMCID: PMC4405151.
11. Niklas KJ, Dunker AK, Yruela I. The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *J Exp Bot.* 2018; 69(7):1437–46. Epub 2018/02/03. <https://doi.org/10.1093/jxb/erx493> PMID: 29394379.
12. Uversky VN. Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J Biol Chem.* 2016; 291(13):6681–8. Epub 2016/02/07. <https://doi.org/10.1074/jbc.R115.685859> PMID: 26851286; PubMed Central PMCID: PMC4807255.
13. Fuxreiter M. Fuzzy protein theory for disordered proteins. *Biochem Soc Trans.* 2020; 48(6):2557–64. Epub 2020/11/11. <https://doi.org/10.1042/BST20200239> PMID: 33170209; PubMed Central PMCID: PMC7752076.
14. Uversky VN. Paradoxes and wonders of intrinsic disorder: Complexity of simplicity. *Intrinsically Disordered Proteins.* 2016; 4(1):e1135015. Epub 2017/02/25. <https://doi.org/10.1080/21690707.2015.1135015> PMID: 28232895; PubMed Central PMCID: PMC5314917.
15. Kurgan L. Resources for computational prediction of intrinsic disorder in proteins. *Methods.* 2022; 204:132–41. Epub 2022/04/04. <https://doi.org/10.1016/j.ymeth.2022.03.018> PMID: 35367597.
16. Anfinsen CB. Principles that govern the folding of protein chains. *Science.* 1973; 181(4096):223–30. <https://doi.org/10.1126/science.181.4096.223> PMID: 4124164
17. Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. *Proteins.* 2021; 89(12):1687–99. Epub 2021/07/05. <https://doi.org/10.1002/prot.26171> PMID: 34218458.
18. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021; 596(7873):583–9. Epub 2021/07/16. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844; PubMed Central PMCID: PMC8371605.
19. Jumper J, Hassabis D. Protein structure predictions to atomic accuracy with AlphaFold. *Nat Methods.* 2022; 19(1):11–2. Epub 2022/01/13. <https://doi.org/10.1038/s41592-021-01362-6> PMID: 35017726.
20. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics.* 2021; 89(12):1711–21.
21. Necci M, Piovesan D, Predictors C, DisProt C, Tosatto SCE. Critical assessment of protein intrinsic disorder prediction. *Nat Methods.* 2021; 18(5):472–81. Epub 2021/04/21. <https://doi.org/10.1038/s41592-021-01117-3> PMID: 33875885; PubMed Central PMCID: PMC8105172.
22. Lang B, Babu MM. A community effort to bring structure to disorder. *Nat Methods.* 2021; 18(5):454–5. Epub 2021/04/21. <https://doi.org/10.1038/s41592-021-01123-5> PMID: 33875888.
23. Quaglia F, Meszaros B, Salladini E, Hatos A, Pancsa R, Chemes LB, et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.* 2022; 50(D1):D480–D7. Epub 2021/12/02. <https://doi.org/10.1093/nar/gkab1082> PMID: 34850135; PubMed Central PMCID: PMC8728214.
24. Wilson CJ, Choy WY, Karttunen M. AlphaFold2: A Role for Disordered Protein/Region Prediction? *Int J Mol Sci.* 2022; 23(9). Epub 2022/05/15. <https://doi.org/10.3390/ijms23094591> PMID: 35562983; PubMed Central PMCID: PMC9104326.
25. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022; 50(D1):D439–D44. Epub 2021/11/19. <https://doi.org/10.1093/nar/gkab1061> PMID: 34791371; PubMed Central PMCID: PMC8728224.
26. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Zidek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature.* 2021; 596(7873):590–6. Epub 2021/07/23. <https://doi.org/10.1038/s41586-021-03828-1> PMID: 34293799; PubMed Central PMCID: PMC8387240.
27. Thornton JM, Laskowski RA, Borkakoti N. AlphaFold heralds a data-driven revolution in biology and medicine. *Nat Med.* 2021; 27(10):1666–9. Epub 2021/10/14. <https://doi.org/10.1038/s41591-021-01533-0> PMID: 34642488.
28. Parisi G, Palopoli N, Tosatto SCE, Fornasari MS, Tompa P. "Protein" no longer means what it used to. *Curr Res Struct Biol.* 2021; 3:146–52. Epub 2021/07/27. <https://doi.org/10.1016/j.crstbi.2021.06.002> PMID: 34308370; PubMed Central PMCID: PMC8283027.
29. Goodsell DS, Zardecki C, Di Costanzo L, Duarte JM, Hudson BP, Persikova I, et al. RCSB Protein Data Bank: Enabling biomedical research and drug discovery. *Protein Sci.* 2020; 29(1):52–65. Epub 2019/09/19. <https://doi.org/10.1002/pro.3730> PMID: 31531901; PubMed Central PMCID: PMC6933845.

30. Perrakis A, Sixma TK. AI revolutions in biology: The joys and perils of AlphaFold. *EMBO Rep.* 2021; 22(11):e54046. Epub 2021/10/21. <https://doi.org/10.15252/embr.202154046> PMID: 34668287; PubMed Central PMCID: PMC8567224.
31. Broomell CC, Chase SF, Laue T, Waite JH. Cutting Edge Structural Protein from the Jaws of *Nereis virens*. *Biomacromolecules.* 2008; 9:1669–77. <https://doi.org/10.1021/bm800200a> PMID: 18498188
32. Chou CC, Martin-Martinez FJ, Qin Z, Dennis PB, Gupta MK, Naik RR, et al. Ion Effect and Metal-Coordinated Cross-Linking for Multiscale Design of *Nereis* Jaw Inspired Mechanomutable Materials. *ACS Nano.* 2017; 11(2):1858–68. Epub 2017/02/07. <https://doi.org/10.1021/acsnano.6b07878> PMID: 28165707.
33. Gupta MK, Becknell KA, Crosby MG, Bedford NM, Wright J, Dennis PB, et al. Programmable Mechanical Properties from a Worm Jaw-Derived Biopolymer through Hierarchical Ion Exposure. *ACS Appl Mater Interfaces.* 2018; 10(38):31928–37. Epub 2018/08/31. <https://doi.org/10.1021/acsnano.6b07878> PMID: 30165014.
34. Evans R, O'Neill M, Pritzel A, Antropova N, Senior AW, Green T, et al. Protein complex prediction with AlphaFold-Multimer. *BioRxiv.* 2021.
35. Kasahara K, Terazawa H, Takahashi T, Higo J. Studies on Molecular Dynamics of Intrinsically Disordered Proteins and Their Fuzzy Complexes: A Mini-Review. *Comput Struct Biotechnol J.* 2019; 17:712–20. Epub 2019/07/16. <https://doi.org/10.1016/j.csbj.2019.06.009> PMID: 31303975; PubMed Central PMCID: PMC6603302.
36. Shrestha UR, Smith JC, Petridis L. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Commun Biol.* 2021; 4(1):243. Epub 2021/02/25. <https://doi.org/10.1038/s42003-021-01759-1> PMID: 33623120; PubMed Central PMCID: PMC7902620.
37. Guo H-B, Varaljay V, Kedziora G, Taylor K, Farajollahi S, Lombardo N, et al. Accurate prediction by AlphaFold2 for ligand binding in a reductive dehalogenase and implications for PFAS (per- and polyfluoroalkyl substance) biodegradation. *Scientific Reports.* 2023; 13:4028.
38. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 2005; 33(7):2302–9. Epub 2005/04/26. <https://doi.org/10.1093/nar/gki524> PMID: 15849316; PubMed Central PMCID: PMC1084323.
39. Mukherjee S, Zhang Y. MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucleic Acids Res.* 2009; 37(11):e83. Epub 2009/05/16. <https://doi.org/10.1093/nar/gkp318> PMID: 19443443; PubMed Central PMCID: PMC2699532.
40. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics.* 2004; 20(2):289–90. Epub 2004/01/22. <https://doi.org/10.1093/bioinformatics/btg412> PMID: 14734327.
41. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005; 26(16):1781–802. Epub 2005/10/14. <https://doi.org/10.1002/jcc.20289> PMID: 16222654; PubMed Central PMCID: PMC2486339.
42. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys.* 2020; 153(4):044130. Epub 2020/08/06. <https://doi.org/10.1063/5.0014475> PMID: 32752662; PubMed Central PMCID: PMC7395834.
43. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* 2017; 14(1):71–3. Epub 2016/11/08. <https://doi.org/10.1038/nmeth.4067> PMID: 27819658; PubMed Central PMCID: PMC5199616.
44. Price DJ, Brooks CL, 3rd. A modified TIP3P water potential for simulation with Ewald summation. *J Chem Phys.* 2004; 121(20):10096–103. Epub 2004/11/20. <https://doi.org/10.1063/1.1808117> PMID: 15549884.
45. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996; 14(1):33–8, 27–8. Epub 1996/02/01. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570.
46. Hansen AL, Kay LE. Measurement of histidine pKa values and tautomer populations in invisible protein states. *Proc Natl Acad Sci U S A.* 2014; 111(17):E1705–12. Epub 2014/04/16. <https://doi.org/10.1073/pnas.1400577111> PMID: 24733918; PubMed Central PMCID: PMC4035968.
47. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, complex systems.* 2006; 1695(5):1–9.
48. Grant BJ, Skjaerven L, Yao XQ. The Bio3D packages for structural bioinformatics. *Protein Sci.* 2021; 30(1):20–30. Epub 2020/08/01. <https://doi.org/10.1002/pro.3923> PMID: 32734663; PubMed Central PMCID: PMC7737766.
49. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics.* 1995; 23(4):566–79. <https://doi.org/10.1002/prot.340230412> PMID: 8749853

50. Chowdhury R, Bouatta N, Biswas S, Floristean C, Kharkare A, Roye K, et al. Single-sequence protein structure prediction using a language model and deep learning. *Nature Biotechnology*. 2022; 40:1617–23. <https://doi.org/10.1038/s41587-022-01432-w> PMID: 36192636
51. Meszaros B, Erdos G, Dosztanyi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res*. 2018; 46(W1):W329–W37. Epub 2018/06/04. <https://doi.org/10.1093/nar/gky384> PMID: 29860432; PubMed Central PMCID: PMC6030935.
52. Erdos G, Pajkos M, Dosztanyi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res*. 2021; 49(W1):W297–W303. Epub 2021/05/29. <https://doi.org/10.1093/nar/gkab408> PMID: 34048569; PubMed Central PMCID: PMC8262696.
53. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol*. 2005; 3(1):35–60. Epub 2005/03/08. <https://doi.org/10.1142/s0219720005000886> PMID: 15751111.
54. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*. 2006; 7:208. Epub 2006/04/19. <https://doi.org/10.1186/1471-2105-7-208> PMID: 16618368; PubMed Central PMCID: PMC1479845.
55. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*. 2010; 1804(4):996–1010. Epub 2010/01/27. <https://doi.org/10.1016/j.bbapap.2010.01.011> PMID: 20100603; PubMed Central PMCID: PMC2882806.
56. Bekele S, Singh K, Helton E, Farajollahi S, Naik RR, Dennis P, et al. Molecular Dynamics Investigation into pH Dependent Metal Binding of the Intrinsically Disordered Worm Jaw Protein, Nvj1-1. *J Phys Chem B*. 2022; 126(35):6614–23. Epub 2022/08/26. <https://doi.org/10.1021/acs.jpcc.2c02807> PMID: 36006408.
57. Guo HB, Perminov A, Bekele S, Kedziora G, Farajollahi S, Varaljay V, et al. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Sci Rep*. 2022; 12(1):10696. Epub 2022/06/24. <https://doi.org/10.1038/s41598-022-14382-9> PMID: 35739160; PubMed Central PMCID: PMC9226352.
58. Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol*. 2011; 21(3):426–31. Epub 2011/05/03. <https://doi.org/10.1016/j.sbi.2011.04.001> PMID: 21530234; PubMed Central PMCID: PMC3112268.
59. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021; 373(6557):871–6. <https://doi.org/10.1126/science.abj8754> PMID: 34282049
60. Guo HB, Ma Y, Tuskan GA, Qin H, Yang X, Guo H. A Suggestion of Converting Protein Intrinsic Disorder to Structural Entropy Using Shannon's Information Theory. *Entropy*. 2019;21(6). Epub 2019/06/14. <https://doi.org/10.3390/e21060591> PMID: 33267305; PubMed Central PMCID: PMC7515080.
61. Janin J, Sternberg MJ. Protein flexibility, not disorder, is intrinsic to molecular recognition. *F1000 biology reports*. 2013; 5.
62. Wayment-Steele HK, Ojoawo A, Otten R, Apitz JM, Pitsawong W, Homberger M, et al. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature*. 2024; 625(7996):832–9. Epub 2023/11/14. <https://doi.org/10.1038/s41586-023-06832-9> PMID: 37956700; PubMed Central PMCID: PMC10808063 authors declare no competing interests.