



HAL
open science

Afflecto: A web server to generate conformational ensembles of flexible proteins from AlphaFold models

Mátyás Pajkos, Ilinka Clerc, Christophe Zanon, Pau Bernadó, Juan Cortés

► **To cite this version:**

Mátyás Pajkos, Ilinka Clerc, Christophe Zanon, Pau Bernadó, Juan Cortés. Afflecto: A web server to generate conformational ensembles of flexible proteins from AlphaFold models. *Journal of Molecular Biology*, 2025, pp.169003. <10.1016/j.jmb.2025.169003>. <hal-04959947>

HAL Id: hal-04959947

<https://laas.hal.science/hal-04959947v1>

Submitted on 21 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

AFflecto: A web server to generate conformational ensembles of flexible proteins from AlphaFold models

Mátyás Pajkos¹, Ilinka Clerc¹, Christophe Zanon¹, Pau Bernadó², Juan Cortés^{1,*}

¹*LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France.*

²*Centre de Biologie Structurale, Université de Montpellier, INSERM, CNRS, Montpellier, France.*

* *Corresponding author: juan.cortes@laas.fr*

Abstract

Intrinsically disordered proteins and regions (IDPs/IDRs) leverage their structural flexibility to fulfill essential cellular functions, with dysfunctions often linked to severe diseases. However, the relationships between their sequences, structural dynamics and functional roles remain poorly understood. Understanding these complex relationships is crucial for therapeutic development, highlighting the need for methods to generate plausible IDP/IDR conformational ensembles. While AlphaFold (AF) excels at modeling structured domains, it fails to accurately represent disordered regions, leaving a significant portion of proteomes inaccurately modeled. We present AFflecto, a user-friendly web server for generating large conformational ensembles of proteins that include both structured domains and IDRs from AF structural models. AFflecto identifies IDRs as tails, linkers or loops by analyzing their structural context. Additionally, it incorporates a method to identify conditionally folded IDRs that AF may incorrectly predict as natively folded elements. The conformational space is globally explored using efficient stochastic sampling algorithms. AFflecto's web interface allows users to customize the modeling, by modifying boundaries between ordered and disordered regions, and selecting among several sampling strategies. The web server is freely available at <https://moma.laas.fr/applications/AFflecto/>.

1 Introduction

Structural biology has traditionally focused on the study of well-folded protein domains. However, intrinsically disordered proteins and regions (IDPs/IDRs), known for their conformational plasticity, are highly abundant across all three kingdoms of life, especially in eukaryotes. For instance, approximately 40% of proteins in the human proteome are estimated to contain IDRs longer than 30 residues [1]. The structural plasticity of IDRs is essential to their function, enabling them to adapt and interact with multiple partners, thereby serving as hubs in cellular networks and mediating complex processes, such as cell regulation and signaling [2]. More recently, IDRs have emerged as key actors in liquid-liquid phase separation processes in cells, where they temporally and spatially control key biological processes [3]. Due to their critical roles, IDRs are also implicated in many human diseases, making them important targets for drug development [4]. However, developing effective therapeutic interventions targeting IDRs is challenging due to their inherent flexibility, which complicates the structural studies needed for drug design.

Structural investigation of IDRs typically relies on a combination of experimental and computational methods [5,6]. Unlike well-folded domains with stable conformations, IDRs exist as ensembles of fluctuating conformations that reflect their structural diversity. While structure prediction tools like the widely used AlphaFold (AF) [7] are skillful in modeling structured domains, they fail to capture IDR ensembles, instead depicting them as single static conformers with limited physical relevance [8]. Note that methods based on AF, such as AF-Cluster [9], have been proposed to predict alternative conformations, typically in the case of fold-switching proteins. These AF-derived methods have proved effective in some particular cases, but generality problems have been highlighted, showing a strong dependence on structures learned during training [10]. Although methods to better exploit information contained in the multiple sequence alignment (MSA) used by AF [11] can improve the generation of conformers and potentially mitigate memorization effects, current approaches relying heavily on AF remain generally inapplicable to IDPs/IDRs. This limitation implies that a significant portion of protein regions in AF structures are not accurately modeled. Nevertheless, the confidence score of the AF models based on the predicted local distance difference test (pLDDT) can serve as indicator of disordered protein segments [12,13]. Additionally, it has been shown that in AF structures, IDRs modeled with (very) high-confidence are likely to be conditionally folded regions that, in many cases, closely resemble their bound state but they do not necessarily represent their structure in solution [14].

As a complement to structure predictors for well-folded proteins, various computational tools have been developed to model IDR ensembles relying on statistics-based or physics-based approaches and, more recently, on machine-learning methods [15–18]. However, most of them deal with individual disordered regions independently of the rest of the protein, ignoring the structural influence exerted by the molecular context. Only a few methods are capable of generating large conformational ensembles of proteins simultaneously considering both IDRs and rigid domains [19,20]. Nevertheless, these methods typically require significant computational resources or expertise, and lack user-friendly implementations, such as web servers, thereby limiting accessibility to a broad scientific community.

Here, we present the AFflecto web server, a new, user-friendly platform designed to generate large conformational ensembles of proteins that include both structured and disordered regions based on AF predictions. IDRs are identified using the pLDDT score and classified as tails, linkers or loops based on their molecular context. Additionally, we introduced a method to detect conditionally folded IDRs that are predicted as rigid regions in the AF models. To explore the conformational space efficiently, AFflecto incorporates our previously-developed IDR and loop sampling methods [21,22]. The web server allows users to perform customized analyses. Coupled with appropriate software, the generated ensemble

models can be used to interpret experimental measurements, such as NMR, SAXS or hydrodynamic data, or serve as initial states for molecular dynamics simulations.

2 Materials and methods

2.1 Region classification protocol

On the basis of the AF structural model, boundaries of rigid and flexible regions are defined using the following process. Per-residue pLDDT scores are extracted from the B-factor column of the PDB file. Residues with a pLDDT score lower than the specified cutoff (a default value of 0.7 is used, as recommended for disorder prediction [23], however this is adjustable by the *pLDDT_cutoff* parameter) are classified as flexible, while residues above this threshold are considered rigid. Next, the flexible residues are grouped into flexible regions through two steps: i) Flexible regions are defined as consecutive flexible residues with lengths of at least three residues for tails and linkers and at least five residues for loops. ii) These flexible regions are further merged if the number of rigid residues between them is smaller than the *shortest_rigid* parameter (three by default, but this is adjustable). The remaining regions of the sequence are identified as rigid regions.

Flexible regions are then classified as linker, loop or tail based on their structural context, as determined by analyzing the AF model. Flexible regions at the C- or N-termini, neighbored by only one rigid region, are classified as tails. Loops are defined as flexible regions where two adjacent rigid regions form at least 10 contacts. A contact is defined when the $C\beta$ atoms ($C\alpha$ for glycine) of residue pairs from each rigid region are at 8 Å or less. This distance is adjustable using the *contact_dist* parameter, which can be set in the range of 6–12 Å to effectively capture significant residue contacts associated with stabilizing interactions in folded proteins. Finally, linkers are classified as flexible regions that connect two rigid regions that do not form stable contacts. If no rigid region is detected, the protein is classified as a pure IDP.

Finally, secondary structural elements (SSEs) that do not form a significant number of tertiary contacts with other SSEs are identified as conditionally folded (CF) SSEs. Concretely, for a given SSE, the contacts between this SSE and all other SSEs in the protein are aggregated and the average number of contacts per residue is calculated. If this value falls below the optimized cutoff of 0.474, the SSE is identified as a CF SSE (see Supplementary Information for further details). The identified CF SSEs are sampled using strategies designed to capture their partial structures.

2.2 Conformational sampling of IDRs

To generate realistic conformers of IDRs, we integrated our previously-developed sampling method into the web server. This method uses structural information from three-residue fragments extracted from high-resolution structures to build conformers through two main strategies. The first, single-residue-based sampling (SRS), generates conformers in a residue-specific manner without considering neighboring residues. The second, three-residue-based sampling (TRS), considers the local sequence and the structural context, making it well suited to model partially structured regions in IDPs. Additionally, we provide an option to sample canonical SSEs such as α -helices, β -strands, and polyproline-II helices within flexible regions with a proportion (population??) defined by the user. This feature allows users to explicitly specify structural constraints, ensuring the sampled conformations include desired SSEs. The (ϕ, ψ) angles are sampled using a Gaussian distribution with mean set to $(-60^\circ, -45^\circ)$ for α -helices, $(-120^\circ, 120^\circ)$

for β -strands and $(-75^\circ, 145^\circ)$ for polyproline-II helices, and with a standard deviation of 5° for both angles in all cases.

Sampling is performed using computationally efficient stochastic algorithms, which have been shown to accurately reproduce NMR and SAXS experimental data measured in IDPs [21]. The sampling method manages loop closure constraints by integrating a robotics-inspired approach [22]. Collisions are checked using a grid-based approach [24] for the backbone atoms of all residues inside flexible and rigid regions. Residue-specific pseudo $C\beta$ atoms are considered at this stage instead of all-atom models of the side-chains in order to avoid their conformational sampling. Possible collisions between side-chains can be removed in an optional post-processing stage included in the server.

2.3 Web server implementation

AFflecto was integrated as a new tool of the MoMA software suite (<https://moma.laas.fr/>). The web server was developed using the Django python-based framework. For visualizing the protein structures, we used the PDBe implementation of Mol* [25], which is a web-based structure viewer offering simple navigation. The web server is fully functional with all modern browsers that support HTML5 and WebP formats.

3 Web server description

AFflecto takes AF models as input, identifies the IDRs and samples them to generate ensemble models of physically-plausible conformers. Users can modify the default settings, allowing for customized modelling of IDPs/IDRs. The web server is freely available at <https://moma.laas.fr/applications/AFflecto/>. This section provides guidance on its usage.

3.1 Submission page

On the submission page, users are required to provide a job title and structural input using a UniProt accession number or uploading a valid AF PDB file. This second option allows the use of AF models generated locally by the users. If a UniProt accession number is provided, the server automatically retrieves the corresponding PDB file from the AF Database (AFDB) [26]. After starting the job, the user will be redirected to the prediction page, unless the uploaded PDB file has an incorrect format or the pLDDT score is missing. In these cases, the job will be aborted.

3.2 Prediction page

The prediction page (Fig 1.) is divided into five main sections, each one offering distinct functionalities. These sections are described in detail below.

Part1 - Structure visualization: AFflecto classifies flexible regions into tails, loops or linkers. In the visualization section, these regions are mapped directly onto the AF model, presenting the structure with the flexible regions for visual inspection by users. Each type of flexible region is marked with a distinct color to guide users. This overview of flexible regions in the full protein context provides valuable insights, assisting users in customizing their analysis in the following sections.

Part2 - Parameter settings: This section provides access to adjustable parameters for the identification and classification of flexible regions. The available parameters are:

- *pLDDT_cutoff*: to discriminate rigid/flexible residues in the AF model.
- *contact_dist*: definition of contact between two rigid regions (see methods for details).
- *shortest_rigid*: minimum required length to define rigid regions.

After selecting the desired parameters, clicking the *Apply* button refreshes the entire prediction page based on the newly chosen configuration.

Part3 - Adjusting and redefining flexible regions: The classified regions can be modified by the user to customize the modeling. Complete regions can be removed, new linkers and loops can be added within rigid regions, and tails can be included at the termini. The region boundaries can be adjusted, and the *Update* button must be clicked afterward to save the changes. The continuity of boundary positions and the alternation between rigid and flexible regions are automatically checked and adjusted.

Part4 - Partially structured regions: In AF structures, conditionally folded (CF) regions are often represented by high-confidence secondary structural elements (SSEs) that are good models of the bound state [14]. In AFflecto, we introduced a method to identify CF SSE regions based on the assumption that, unlike SSEs in natively folded domains which form many contacts to stabilize the structure, CF SSEs lack significant tertiary contacts with other SSEs, as they primarily interact with their binding partners (Fig. S1) (see methods). These CF SSEs are treated as partially structured regions and are constructed using the TRS sampling strategy. Users can refine these regions by adjusting their boundaries, or adding and removing regions as needed.

Part5 - Canonical secondary structural element: Canonical SSEs, including α -helices, β -strands, and polyproline-II helices, can be defined by users within the specified tail, loop or linker regions. These canonical SSE are then explicitly sampled with a population specified by the user. This feature is particularly relevant when there is experimental evidence of secondary structure content in the protein.

Part 6 - Job submission: Users can specify the number of conformers to be generated (up to a maximum of 1000, with a default of 100). By enabling the *Add sidechains* option, side-chains will be added in the output PDB files based on the input AF structure; only the side-chains involved in collisions will be locally re-sampled. Otherwise, residue-specific pseudo $C\beta$ atoms will be considered for collision detection, and the output PDB files will only contain backbone atoms. Finally, clicking the *Launch* button initiates the ensemble generation process.

3.3 Result page

The conformational ensemble model of the submitted protein is generated on the server back-end using stochastic algorithms to sample IDR conformers (see methods). Once the calculation is completed, a notification email is sent (if the user has created an account) and the user can download the generated ensemble from the results page as a ZIP file, which contains the conformers as individual PDB files. Anonymous users can retrieve their results using a unique identifier assigned to them.

Part 1

Part 2

Afflecto parameters

PLDDT_cutoff: 0.7 (min = 0.0, max = 1.0, recommended = 0.7)

Contact_dist: 8 (min = 6, max = 12, recommended = 8)

Shortest_rigid: 3 (min = 3, max = 196 (sequence length), recommended = 3)

Part 3

Structural region classification (residues 1 to 196)

Start position	End position	Type	Action
1	24	TAIL	Delete
25	118	RIGID DOMAIN	+ LOOP + LINKER Delete
119	123	LOOP	Delete
124	192	RIGID DOMAIN	+ LOOP + LINKER Delete
193	196	TAIL	Delete

Part 4

Partially structured regions

Start position	End position	Action
120	122	Edit Delete

Part 5

Canonical secondary structure element

Start position	End position	Type	Proportion	Action
Add				

Part 6

Submit for ensemble generation

Number of conformations: 100 (min = 1, max = 1000, default = 100)

Add sidechains

Launch

Figure 1: Screenshot of the prediction page for the Q5W111 protein. The six sections with distinct functionalities are highlighted by red rectangles.

4 Use cases

4.1 Example 1: Measles Virus Nucleoprotein

The measles virus nucleoprotein (MV-N, 525 residues, UniProt ID: P0DXN6) assembles into nucleocapsids together with the viral genome and interacts with the P-protein, which is bound to the viral polymerase complex. These processes are mediated by MV-N IDRs and have been the subject of extensive experimental studies. MV-N consists of the N-core domain (residues 34-374), responsible for binding viral RNA and maintaining capsid structure, flanked by N- and C-terminal arms (N-arm and C-arm, 1-33 and 375-391, respectively) that stabilize nucleocapsid assembly by binding neighbouring protomers. The C-arm is followed by the disordered N-tail (residues 392-525), which includes the molecular recognition region (MoRE, residues 485-502) that binds to the P-protein [27-31].

Since viral proteins are not included in the AlphaFold Protein Data Bank (AFDB), the MV-N structure was predicted using the AlphaFold3 server [32]. Using default parameters, AFflecto identified a C-terminal long tail (residues 378-525), encompassing the C-arm and the disordered N-tail. [30]. In the N-tail region, two short CF SSEs were identified (residues 485-501 and 520-523) and flagged for sampling using the TRS strategy. When submitted to AFflecto, the ensemble showed partially helical conformations (Fig. 1). These findings align with experimental results, which describe α -helices within the conformational ensemble of the N-tail, particularly around the MoRE region [29]. Additionally, AFflecto identified an N-terminal tail (residues 1-26) associated with the N-arm. Within the N-arm region, a 15-residue-long CF region (2-16) was also identified by AFflecto. When sampled using the TRS strategy, this region showed α -helices in some of the conformers (Fig. 1). Notably, this helix has been demonstrated to stabilize nucleocapsid assembly by binding neighbouring MV-N protomers [33]. Note that AFflecto also identified two loops according to the AF prediction (133-143 and 206-210), which were sampled simultaneously with the other flexible regions.

Additionally, a flexible linker (residues 262-266) was inserted between the N- and C-terminal subdomains of the N-core domain (NTD and CTD) (Fig. 1). Although this region was predicted with high confidence in the AF structure and was not classified as a linker by default in the AFflecto model, the server allows users to manually define and insert such flexible regions. This feature provides a valuable advantage for customizing ensemble generation. The conformational ensemble generated by AFflecto displays the hinge motions between both folded subdomains. Note that this hinge motions, which are limited according to the modelling, have been associated with the open-closed transition during nucleocapsid formation [34, 35]. This use case exemplifies the capacity of AFflecto to model complex protein chains including the three types of flexible regions, and to integrate previous structural knowledge derived from high-resolution techniques.

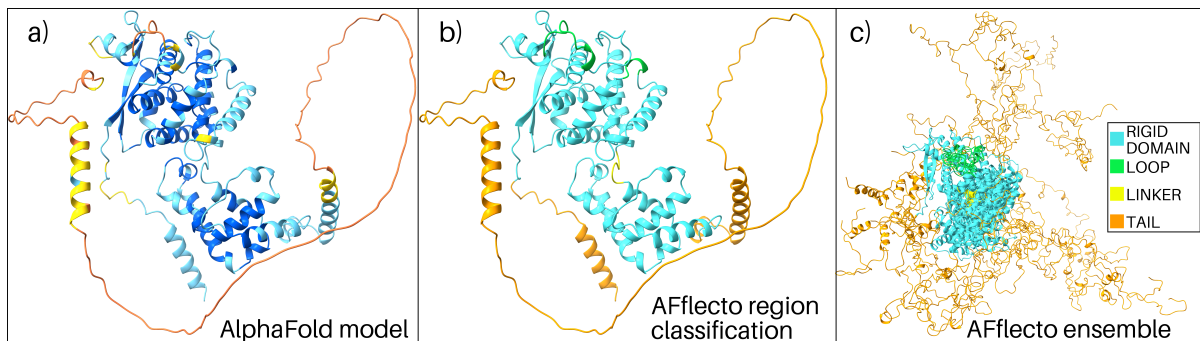


Figure 2: AFflecto model of the nucleoprotein from Measles Virus. (a) AF model of the protein, with colors corresponding to the AF pLDDT score coloring system. (b) AFflecto classification of rigid and flexible regions based on default parameters, and with the inclusion of an additional linker region connecting the two rigid subdomains. (c) Ten conformations generated for the protein using the definitions from panel (b).

4.2 Example 2: Human Tau protein

The human Tau40 protein (Tau-F, UniProt ID: P10636-8), a 441-residue isoform of the microtubule-associated protein Tau, stabilizes microtubules in neurons and regulates cytoskeletal structure [36]. Importantly, its aggregation into neurofibrillary tangles is a hallmark of neurodegenerative pathologies, such as Alzheimer’s disease [37]. The full-length protein can be divided into two main parts: an N-terminal projection domain (1-200) and a C-terminal microtubule (MT) binding region (MTBR) (residues 242-367), which encompasses four repeats that stabilize MTs through direct interaction [38]. Fluorescence correlation spectroscopy and acrylodan fluorescence screening revealed helical structure in the MTBR upon tubulin binding, indicating partially folded binding regions [39]. Thanks to extensive structural studies, Tau40 is considered as a prototypical example of IDP when not associated with MTs [40–43], and it has been annotated as such in the DisProt database under entry DP01100 [44].

The AF PDB file for Tau40 was generated using the AlphaFold3 server, as isoforms of this protein are not available in the AFDB. AF predicts several high-confidence, rigid regions in the human Tau40 protein, including long coil segments and an α -helix within the N-terminal projection domain (Fig. 2). However, using the developed region classification protocol, AFflecto identifies regions with falsely high confidence predictions in the AF structure as flexible, accurately presenting the Tau40 protein as a pure IDP (Fig. 2). In the AFflecto region classification, several short partially structured regions within the MTBR are correctly identified and designated for sampling using the TRS strategy. Additionally, a 14-residue-long partially-structured region (424–438) is identified within the C-terminal tail (Fig. 2), which contains both basic and acidic subregions and is associated with the indirect regulation of MT binding [38].

To further investigate the structural characteristics of the C-terminal tail of Tau40 (residues 410–441), we generated a conformational ensemble using the TRS sampling strategy for the region predicted as partially structured, and analyzed it with the WARIO tool [45]. WARIO identifies clusters of conformations that exhibit common structural patterns based on inter-residue contact maps. The analysis revealed that 22.5% of the conformers displayed α -helical conformations in the region spanning residues 425–430, indicating a higher propensity for partial helicity in C-terminal tail (Fig. S2) (for details see supplementary

materials). These findings suggest that the AFflecto-generated conformers effectively capture transient secondary structures.

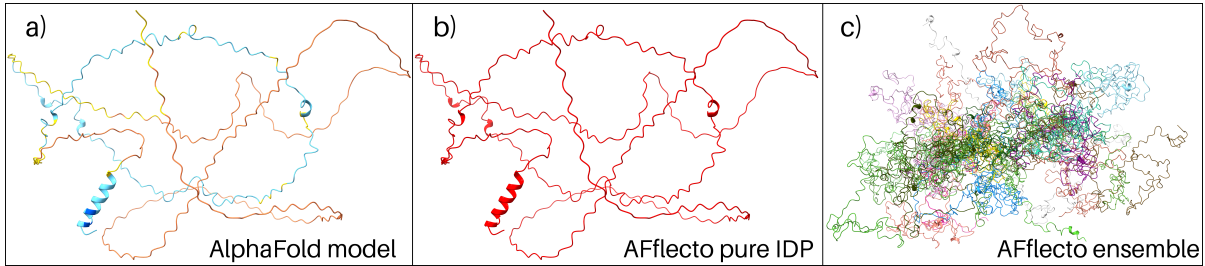


Figure 3: AFflecto model of the Tau40 protein. (a) AF model colored according to the pLDDT score. (b) AFflecto identifies this protein as a pure IDPs. (c) Twenty conformations generated for Tau40, modeled as a pure IDP. Each conformer is depicted in a different color.

4.3 Computational Efficiency

To assess computational performance, we analyzed the runtime for generating conformational ensembles of the two systems discussed in this study. For the measles virus nucleoprotein (two tails, two loops, and one linker, accounting for 195 flexible residues out of the full length of 525 amino acids) and the human Tau40 protein (a pure IDP containing 441 flexible residues), the runtimes for generating 100 conformers with default settings using 10 threads on a standard CPU were around 38 seconds and 20 seconds, respectively, demonstrating the methods's efficiency. Note however that the overall runtime on the server depends on the initialization step of the software, which can be relatively slow, and on server workload during execution. Additional delays may be caused by the size of the job stack if many users are simultaneously connected to the server. Despite these variations, AFflecto generally delivers results within minutes.

5 Conclusion

Many proteins are composed of both globular domains and flexible, disordered regions, such as linkers, tails and loops that, in solution, sample a vast conformational space. While machine-learning-based structure prediction methods, such as AF, provide accurate models for structured domains, they often depict disordered regions with a single, low-confidence conformation. To address this limitation, we present AFflecto, a web server that generates conformational ensemble models based on AF predictions. It samples the conformations of disordered regions using a data-driven stochastic algorithm. These ensembles offer a comprehensive representation of the diverse conformational states of flexible proteins in solution, making them valuable entry points for more detailed analyses to interpret experimental data or as starting points for molecular dynamics simulations.

The generated conformational ensembles provide a plausible picture of the protein in solution reporting on the overall size, the feasibility of long-range intramolecular contacts, and the extent of loop motions. Beyond these obvious applications of AFflecto, these models also provide valuable opportunities for various biological applications. First, they can be used in combination with experimental methods such as NMR and small-angle X-ray scattering (SAXS) to structurally characterize highly flexible proteins [5, 46]. When coupled to appropriate computational forward models, the backcalculated observables can be compared with the experimental ones in order to (in)validate the structural models. Alternatively, the theoretical profiles can be used for ensemble optimization or reweighting strategies [47, 48].

Second, the ensembles serve as excellent starting points for molecular dynamics simulations, facilitating more detailed investigations of protein flexibility and dynamics. Finally, these ensembles can be used for drug design initiatives, particularly in identifying and targeting disordered regions or transient conformations that play key functional roles.

While the current implementation of AFflecto relies on an in-house algorithm for defining flexible region types and boundaries based on AF predictions, users are encouraged to incorporate external information from relevant on-line resources, such as The Encyclopedia of Domains (TED) [49], or experimental data to refine classifications. The web interface allows users to make customized and informed decisions, complementing the default settings. Integrating AFflecto with this type of resources could be a very interesting extension for the future.

Additionally, we are developing a locally runnable version of the server that will allow users to generate larger ensembles and handle multiple chains, providing a more flexible and scalable solution for complex protein systems. All together, we believe that AFflecto will become a fundamental tool to investigate the flexibility of complex protein systems, which will provide new avenues to understand the role of disorder in molecular biology.

6 Acknowledgements

This work was supported by the French National Research Agency (ANR) under grant ANR-22-CE45-0003 (CORNFLEX project). It was also carried out in the framework of the COST ML4NGP action, supported by COST (European Cooperation in Science and Technology). The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively).

7 Author contributions

Mátyás Pajkos: Methodology, Software, Validation, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Ilinka Clerc:** Methodology, Software; **Christophe Zanon:** Software, Validation; **Pau Bernadó:** Conceptualization, Validation, Investigation, Writing - Review & Editing, Funding acquisition; **Juan Cortés:** Conceptualization, Methodology, Software, Validation, Investigation, Writing - Original Draft, Writing - Review & Editing, Supervision, Funding acquisition.

References

- [1] van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. *Chem Rev* **114**, 13, 6589–6631.
- [2] Tompa, P., Schad, E., Tantos, A., and Kalmar, L. (2015). Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struct Biol* **35**, 49–59.
- [3] Chong, P. A. and Forman-Kay, J. D. (2016). Liquid–liquid phase separation in cellular signaling systems. *Curr Opin Struct Biol* **41**, 180–186.
- [4] Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008). Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annu Rev Biophys* **37**, 215–246.
- [5] Milles, S., Salvi, N., Blackledge, M., and Jensen, M. R. (2018). Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments. *Prog Nucl Magn Reson Spectrosc* **109**, 79–100.
- [6] Cordeiro, T. N., Herranz-Trillo, F., Urbanek, A., Estaña, A., Cortés, J., Sibille, N., and Bernadó, P. (2017). Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr Opin Struct Biol* **42**, 15–23.
- [7] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., and others. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 7873, 583–589.
- [8] Ruff, K. M. and Pappu, R. V. (2021). AlphaFold and implications for intrinsically disordered proteins. *J Mol Biol* **433**, 20, 167208.
- [9] Wayment-Steele, H. K., Ojoawo, A., Otten, R., Apitz, J. M., Pitsawong, W., Hömberger, M., Ovchinnikov, S., Colwell, L., and Kern, D. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 7996, 832–839.
- [10] Chakravarty, D., Schafer, J. W., Chen, E. A., Thole, J. F., Ronish, L. A., Lee, M., and Porter, L. L. (2024). AlphaFold predictions of fold-switched conformations are driven by structure memorization. *Nat. Commun.* **15**, 1, 7296.
- [11] Bryant, P. and Noé, F. (2024). Improved protein complex prediction with AlphaFold-multimer by denoising the MSA profile. *PLoS Comput. Biol.* **20**, 7, 1–12.
- [12] Wilson, C. J., Choy, W.-Y., and Karttunen, M. (2022). AlphaFold2: A role for disordered protein/region prediction? *Int J Mol Sci* **23**, 9.

- [13] Ma, P., Li, D.-W., and Brüschweiler, R. (2023). Predicting protein flexibility with AlphaFold. *Proteins* **91**, 6, 847–855.
- [14] Alderson, T. R., Pritišanac, I., Kolarić, D., Moses, A. M., and Forman-Kay, J. D. (2023). Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proc Natl Acad Sci USA* **120**, 44, e2304302120.
- [15] Clerc, I., Sagar, A., Barducci, A., Sibille, N., Bernadó, P., and Cortés, J. (2021). The diversity of molecular interactions involving intrinsically disordered proteins: A molecular modeling perspective. *Comput Struct Biotechnol J* **19**, 3817–3828.
- [16] Liu, Z. H., Tsanai, M., Zhang, O., Forman-Kay, J., and Head-Gordon, T. (2024). Computational methods to investigate intrinsically disordered proteins and their complexes. *arXiv:2409.02240v1*.
- [17] Aupič, J., Pokorná, P., Ruthstein, S., and Magistrato, A. (2024). Predicting conformational ensembles of intrinsically disordered proteins: From molecular dynamics to machine learning. *J Phys Chem Lett* **15**, 32, 8177–8186.
- [18] Bernadó, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. H., and Blackledge, M. (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc Natl Acad Sci USA* **102**, 47, 17002–17007.
- [19] Liu, Z. H., Teixeira, J. M. C., Zhang, O., Tsangaris, T. E., Li, J., Gradinaru, C. C., Head-Gordon, T., and Forman-Kay, J. D. (2023). Local Disordered Region Sampling (LDRS) for ensemble modeling of proteins with experimentally undetermined or low confidence prediction segments. *Bioinformatics* **39**, 12, btad739.
- [20] Cao, F., von Bülow, S., Tesei, G., and Lindorff-Larsen, K. (2024). A coarse-grained model for disordered and multi-domain proteins. *Protein Science* **33**, 11, e5172.
- [21] Estaña, A., Sibille, N., Delaforge, E., Vaisset, M., Cortés, J., and Bernadó, P. (2019). Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database. *Structure* **27**, 2, 381–391.e2.
- [22] Barozet, A., Molloy, K., Vaisset, M., Simeon, T., and Cortés, J. (2020). A reinforcement-learning-based approach to enhance exhaustive protein loop sampling. *Bioinformatics* **36**, 4, 1099–1106.
- [23] Piovesan, D., Monzon, A. M., and Tosatto, S. C. E. (2022). Intrinsic protein disorder and conditional folding in AlphaFoldDB. *Protein Sci* **31**, 11, e4466.
- [24] de Angulo, V. R., Cortés, J., and Porta, J. M. (2012). Rigid-CLL: Avoiding constant-distance computations in cell linked-lists algorithms. *J Comput Chem* **33**, 3, 294–300.
- [25] Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S. K., Koča, J., and Rose, A. S. (2021). Mol* viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* **49**, W1, W431–W437.
- [26] Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Židek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., Birney, E., Steinegger, M., Hassabis, D., and Velankar, S. (2024). AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **52**, D1, D368–D375.

- [27] Desfosses, A., Milles, S., Jensen, M. R., Guseva, S., Colletier, J.-P., Maurin, D., Schoehn, G., Gutsche, I., Ruigrok, R. W. H., and Blackledge, M. (2019). Assembly and cryo-EM structures of RNA-specific measles virus nucleocapsids provide mechanistic insight into paramyxoviral replication. *Proc Natl Acad Sci USA* **116**, 10, 4256–4264.
- [28] Kingston, R. L., Baase, W. A., and Gay, L. S. (2004). Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. *J Virol* **78**, 16, 8630–8640.
- [29] Jensen, M. R., Communie, G., Ribeiro, Jr, E. A., Martinez, N., Desfosses, A., Salmon, L., Mollica, L., Gabel, F., Jamin, M., Longhi, S., Ruigrok, R. W. H., and Blackledge, M. (2011). Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci USA* **108**, 24, 9839–9844.
- [30] Longhi, S., Receveur-Bréchet, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., Yeo, R., Finet, S., and Canard, B. (2003). The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J Biol Chem* **278**, 20, 18638–18648.
- [31] Guseva, S., Milles, S., Jensen, M. R., Schoehn, G., Ruigrok, R. W. H., and Blackledge, M. (2020). Structure, dynamics and phase separation of measles virus RNA replication machinery. *Curr Opin Virol* **41**, 59–67.
- [32] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O’Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M., Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M. R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., Židek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 8016, 493–500.
- [33] Guryanov, S. G., Liljeroos, L., Kasaragod, P., Kajander, T., and Butcher, S. J. (2015). Crystal structure of the measles virus nucleoprotein core in complex with an N-terminal region of phosphoprotein. *J Virol* **90**, 6, 2849–2857.
- [34] Gutsche, I., Desfosses, A., Effantin, G., Ling, W. L., Haupt, M., Ruigrok, R. W. H., Sachse, C., and Schoehn, G. (2015). Structural virology. near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* **348**, 6235, 704–707.
- [35] Yabukarski, F., Lawrence, P., Tarbouriech, N., Bourhis, J.-M., Delaforge, E., Jensen, M. R., Ruigrok, R. W. H., Blackledge, M., Volchkov, V., and Jamin, M. (2014). Structure of nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nat Struct Mol Biol* **21**, 9, 754–759.
- [36] Wang, Y. and Mandelkow, E. (2016). Tau in physiology and pathology. *Nat Rev Neurosci* **17**, 1, 5–21.
- [37] Pîrșcoveanu, D. F. V., Pirici, I., Tudorică, V., Bălșeanu, T. A., Albu, V. C., Bondari, S., Bumbea, A. M., and Pîrșcoveanu, M. (2017). Tau protein in neurodegenerative diseases - a review. *Rom J Morphol Embryol* **58**, 4, 1141–1150.
- [38] Kellogg, E. H., Hejab, N. M. A., Poepsel, S., Downing, K. H., DiMaio, F., and Nogales, E. (2018). Near-atomic model of microtubule-tau interactions. *Science* **360**, 6394, 1242–1246.

- [39] Li, X.-H., Culver, J. A., and Rhoades, E. (2015). Tau binds to multiple tubulin dimers with helical structure. *J Am Chem Soc* **137**, 29, 9218–9221.
- [40] Smet, C., Leroy, A., Sillen, A., Wieruszeski, J.-M., Landrieu, I., and Lippens, G. (2004). Accepting its random coil nature allows a partial nmr assignment of the neuronal tau protein. *ChemBioChem* **5**, 12, 1639–1646.
- [41] Mylonas, E., Hascher, A., Bernadó, P., Blackledge, M., Mandelkow, E., and Svergun, D. I. (2008). Domain conformation of tau protein studied by solution small-angle X-ray scattering. *Biochemistry* **47**, 39, 10345–10353.
- [42] Mukrasch, M. D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2009). Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol* **7**, 2, e34.
- [43] Schwalbe, M., Ozenne, V., Bibow, S., Jaremko, M., Jaremko, L., Gajda, M., Jensen, M. R., Biernat, J., Becker, S., Mandelkow, E., Zweckstetter, M., and Blackledge, M. (2014). Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α -synuclein in solution from NMR and small angle scattering. *Structure* **22**, 2, 238–249.
- [44] Aspromonte, M. C., Nugnes, M. V., Quaglia, F., Bouharoua, A., DisProt Consortium, Tosatto, S. C. E., and Piovesan, D. (2024). DisProt in 2024: improving function annotation of intrinsically disordered proteins. *Nucleic Acids Res* **52**, D1, D434–D441.
- [45] González-Delgado, J., Bernadó, P., Neuvial, P., and Cortés, J. (2024). Weighted families of contact maps to characterize conformational ensembles of (highly-)flexible proteins. *Bioinformatics* **40**, 11, btae627.
- [46] Bernadó, P. and Svergun, D. I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol BioSyst* **8**, 151–167.
- [47] Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* **129**, 17, 5656–5664.
- [48] Pesce, F. and Lindorff-Larsen, K. (2023). Combining experiments and simulations to examine the temperature-dependent behavior of a disordered protein. *J Phys Chem B* **127**, 28, 6277–6286. PMID: 37433228.
- [49] Lau, A. M., Bordin, N., Kandathil, S. M., Sillitoe, I., Waman, V. P., Wells, J., Orengo, C. A., and Jones, D. T. (2024). Exploring structural diversity across the protein universe with the encyclopedia of domains. *Science* **386**, 6721, eadq4946.