EPJ Web of Conferences **366**, 01007 (2026) https://doi.org/10.1051/epjconf/202636601007
10th Complexity-Disorder Days 2025

Protein order and disorder: a quantitative in silico analysis

Alexandre G. de Brevern [1*]

[1] Université Paris Cité and Université de la Réunion, INSERM, EFS, BIGR U1134, DSIMB Bioinformatics team, 75015 Paris, France

Abstract. Proteins are fundamental biological macromolecules responsible for a wide range of cellular functions. Traditionally, a central paradigm linked the amino acid sequence to a unique, ordered three-dimensional structure that underlies biological activity. While it has long been recognized that proteins are not rigid and contain flexible regions necessary for function, it was only in the late 1990s that attention turned to intrinsically disordered regions (IDRs) and intrinsically disordered proteins (IDPs)—segments or entire proteins that lack stable tertiary structures and exist as dynamic ensembles. To analyse local conformations in structured proteins, we developed a structural alphabet, known as Protein Blocks (PBs). This tool enables a residue-level description of backbone geometry and has proven effective in applications such as structure prediction and the analysis of molecular dynamics simulations. Building on this framework, PB analysis was extended to disordered proteins and introduced an entropy-based gradient that characterizes regions along a continuum from rigid to fully disordered. This scale represents the first model to provide a unified description of structural dynamics, bridging the gap between ordered and disordered states. It has been successfully applied in various studies, offering new insights into the complex conformational behaviour of proteins.

* Corresponding author: alexandre.debrevern@u-paris.fr

1 Interest in 3D protein structures

Proteins are macromolecules composed of a sequence of amino acids that support most biological functions. The three-dimensional (3D) structure of proteins can be determined using various experimental approaches, including X-ray crystallography, nuclear magnetic resonance and cryo-electron microscopy [1]. This 3D information makes it possible to understand the mechanism of their various functions (in theory) at the atomic level [2]. For example, knowing the structure of an enzyme with its ligand provides a better understanding of the catalytic mechanism. Knowing the structure of a transmembrane transporter allows us to understand the passage of essential solutes. The use of this information in biotechnology or medicine is therefore particularly important. For example, knowing the structure involved in pathology can be used to propose molecules using a drug design approach to better propose drugs [3].

2 Protein flexibility

It was clear that proteins are not rigid bodies, but they have flexible regions (otherwise they would be just inert aggregates without function). From a structural point of view, X-ray crystallographic data contain the information of B-factors [4]. B-factors or atomic displacement parameters) quantify the uncertainty or mobility in the position of an atom within a protein crystal structure [5]. They are also named temperature factors, Debye–Waller factors, or isotropic/anisotropic displacement parameters. They correspond to the extent to which an atom deviates from its mean position due to (i) thermal motion and (ii) static disorder within the crystal lattice [6].

The B-factor belongs to the atomic scattering factor; it allows to better understanding the attenuation of X-ray scattering at higher diffraction angles [7]. A high B factor indicates greater atomic mobility or positional uncertainty, which can result from thermal vibrations, conformational flexibility, crystal defects or even data processing artefacts. Conversely, a low B-factor indicates rigid, well-ordered regions of the protein. They therefore serve as an indicator for protein flexibility, i.e. a high value means a flexible position, a low value, a rigid position, even if some limitations are well known [8,9].

B-factors are provided in Å² and are often used visually to analyse the protein dynamics [10]. For example, on average we can consider that backbone atoms in α-helices and β-sheets often show lower B-factors due to structural rigidity, while loops, termini, and solvent-exposed residues typically have higher B-factors due to increased flexibility or disorder [11]. Figure 1A shows the B-factor distribution on recombinant ubiquitin-conjugating enzyme (E2, PDB 2AAK [12]) structure. The central β-sheet is associated to low B-factor values and so considered as rigid (noted R), while some loops are in orange and red as associated to low B-factor values and so considered as flexible and highly flexible (noted F). As often, α-helices are between both.

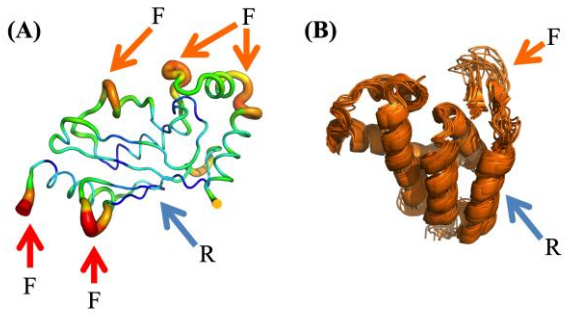

Fig. 1. Experimental flexibility. (A) B-factor values of recombinant ubiquitin-conjugating enzyme (PDB id 2AAK) and (B) NMR structure of N-terminal domain of HTLV-I CA1-134 (PDB id 1G03). Examples of regions considered as rigid are noted with R, and, as flexible with F. Visualization is done using PyMOL software [13].

In a similar way, Nuclear Magnetic Resonance (NMR) spectroscopy provides information on then 3D structure of proteins, but also its flexibility [14]. NMR principle relies on the magnetic properties of atomic nuclei—specifically nuclei that possess a non-zero nuclear spin, such as ^{1}H (proton), ^{13}C, ^{1}N, and ^{31}P. In a magnetic field, these nuclear magnets align either with or against the field, resulting in discrete energy levels. As each type of nucleus and its surrounding electronic environment affects the resonance frequency, it gives rise to chemical shifts, providing distance constraints. These constraints gives potential structural models, often more than 20, that all resolve the experimental constraints. NMR is particularly powerful for studying proteins in solution, preserving their natural conformational flexibility [15,16]. Portion with a lot of constraint will be considered as rigid, while the regions with few would be considered as flexible. Figure 1B shows the NMR structure of N-terminal domain of HTLV-I CA1-134 (PDB id 1G03 [17]). The authors have proposed 20 conformers. As seen, α-helices are proposed in a very similar way, i.e. they have a lot of good restraints and considered as rigid (noted R), while a loop is underlined with more flexibility (noted F), i.e. less restraints

3 Protein Disorder

However, it wasn't until the late 1990s that we began to consider protein regions that went beyond the notion of flexibility: protein disorder. These regions (or sometimes entire proteins) are so dynamic that they are associated with undetermined structures [18].

Protein disorder, or more precisely intrinsic disorder, means that some parts of a protein—or sometimes the whole protein—do not settle into a fixed 3D shape, even under normal conditions in the lab or in the body. Instead of folding into a stable structure, these proteins or regions stay very flexible and constantly shift between different shapes [19]. Through several important studies, Vladimir Uversky showed that some proteins don't need a fixed shape to work.

This went against the old idea that proteins must have a specific structure to function. Instead, he found that many proteins work just fine—and sometimes even better—when they are completely or partly flexible. At the time, this was a ground breaking discovery [20].

Uversky's research showed that intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) tend to have lots of polar and charged amino acids and fewer hydrophobic ones, which prevents them from folding into stable structures [21,22]. Their simple, repetitive sequences make them highly flexible, and despite lacking a compact shape, they stay dissolved in water thanks to their surface charges and ability to attract water molecules [23,24]. Later it was shown that some disorder regions can became ordered by interacting with their partner(s), i.e. namely disorder to order regions [25-27].

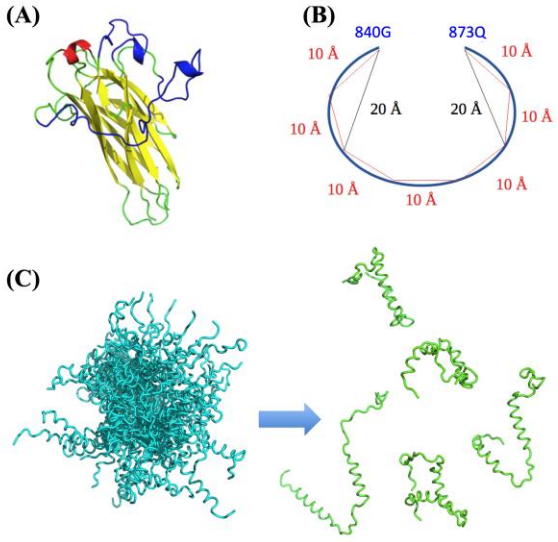

Fig. 2. IDRs and IDPs. (A) Completed model of the Calf-2 domain (PDB id 3FCS [28]). The model is represented with PyMOL [13]. Strands are in yellow, α-helices in red, and loops in green. The two missing regions are coloured in blue. (B) Set of constraints imposed on the second missing region of Calf-2 (positions 840G-873Q) using Modeller software [29]. (C) left, Structural ensemble of PaaA2 antitoxin from the human pathogen *E. coli* O157 with 50 structural models (PED00013 [30]), right: five extracted models [31].

From a structural perspective, disordered regions are therefore associated with unresolved regions (noted as missing residues in the PDB structure files [1]). Figures 2A and 2B underline the case of Calf-2 domain of integrin sub-unit αIIB (PDB id 3FCS [28]). A large IDR (characteristic of multiple integrins) exist, here missing residues are between residues 840 and 873, i.e. 33 residue length. For analysing it, it must be completed, but it is highly complex as all approaches try to compact it (as for globular protein), so multiple constraints must be added. This difficult is underlined on Figure 2B with the need of 7 constraints of 10Å and 2 extra ones of 20Å to succeed to build with Modeller [29] a correct IDR loop [28]. It provides a confident IDRs shown in blue of Figure 2A.

However, there are approaches to determining probabilistic ensembles of conformations. In this context, the Protein Ensemble Database (PED, https://proteinensemble.org/) contains a large amount of useful data [30]. These ensembles are calculated based on experimental data coming mainly from NMR and Small-angle X-ray Scattering (SAXS) measurements. Meaningful ensembles might also be constructed using Molecular Dynamics (MD) simulations. The latest update of September 14, 2023, includes 336, 170 for 306 different proteins. Figure 2C shows an example taken from PED, namely PED00013 [31]. It is composed of 50 structural models and highlights the high dynamics of the system. 5 distinct conformers are also individually shown to underline their differences.

4 Analysis of 3D structures and their dynamics

Pauling and Corey were among the first to predict the existence of regular local structures within proteins [32]. Two of these, which later became foundational elements of protein architecture, are the α-helix and the β-sheet [33,34]. The α-helix was shown to be both geometrically stable and energetically favourable, based on precise structural parameters derived from crystallographic studies of small molecules [35]. The β-sheet, in contrast, was described as an extended arrangement of polypeptide chains stabilized by hydrogen bonds between neighbouring strands, which could be either parallel or antiparallel [36]. This series of landmark publications significantly reshaped the understanding of protein structure.

Over time, the frequent observation of α-helices and β-sheets in experimentally resolved protein structures led to the establishment of the concept of secondary structure, typically divided into three categories: α-helix, β-strand, and coil—the latter representing regions that do not conform to regular helical or sheet patterns [37]. However, this classification oversimplifies the true structural diversity. Not all helices adopt the canonical α-helical form, β-sheets often display considerable variation and imperfection, and so-called coil regions are not truly random but encompass a range of conformational states with functional and structural significance [38-42].

We, like others decided to develop a structural alphabet able to approximate (and analyse) every local protein conformations, leaving no unassigned residue, and using only one type of generic rule. Developed in the early 2000s, this method, named Protein Blocks (PBs), allows complex three-dimensional protein structures to be translated into simplified one-dimensional sequences composed of a limited set of structural prototypes [43]. These prototypes were identified through a statistical analysis of backbone dihedral angles (ϕ and ψ) from known protein structures, resulting in sixteen representative local conformations labelled from PB *a* to PB *p*. Each PB corresponds to a five-residue fragment and captures a specific local geometry, including classical elements such as helices and strands, as well as more irregular or

transitional structures [44]. The 16 PBs (see Figure 3A) allows to translate a 3D structure into a series of PB letters, providing information as a 1D amino acid sequence, but corresponding to 3D information (see Figure 3B).

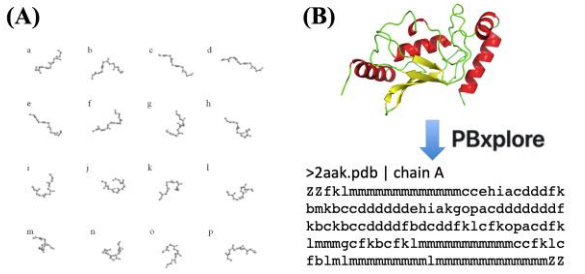

Fig. 3. The Protein Blocks. From left to right and top to bottom the 16 Protein Blocks (labelled from PB a to PB p) are shown. For each PB, the N-cap is on the left and the C-cap is on the right [45]. (B) Translation from 3D structure of recombinant ubiquitin-conjugating enzyme (PDB 2AAK) to Protein Blocks using PBxplore [46].

This approach provides a finer level of detail than the conventional division into α-helices, β-strands, and coils, which often oversimplifies the diversity of real protein structures. By encoding structural data as a string of PB letters, researchers can more easily compare proteins, predict local structure, and apply statistical or machine learning methods to structural bioinformatics tasks [47-51]. Additionally, Protein Blocks offer insight into local conformational variability and flexibility, making them a valuable tool for analysing structural features involved in protein function, dynamics, and evolution [46,52].

Originally dedicated to the analysis of 3D structures, PBs have been applied to many questions such as sequence-based prediction, superposition of 3D structures, definition of binding sites, and in recent years to the analysis of MD simulation trajectories [46].

In this context, each snapshot (time) of the protein simulation is translated into PBs. It is therefore possible to count the type of PB observed at each position. Figure 4A shows an example of this type of counting obtained with the PBxplore software [46]. A red square means that the position was seen 100% of the time with this PB assignment, something therefore quite rigid. Conversely, in this specific case, the other 15 PBs will be at 0%, therefore in blue. Figure 4B shows the same information using WebLogo software [53].

For MD, RMSD (Root-Mean-Square Deviation) quantifies structural variations during dynamics by comparing each frame to a reference structure, i.e., the starting frame. For each frame, an average of the differences between the reference positions and the positions of the current frame is performed in order to have an RMSD value per unit of time. The RMSF (Root-Mean-Square Fluctuation) is similar to the RMSD, by determining the fluctuation of each residue following the same principle as for the RMSD, i.e. a comparison with a reference. But this time, it is the average position of each residue and thus the measure of the difference between the current position and the average position in order to have a flexibility value per position.

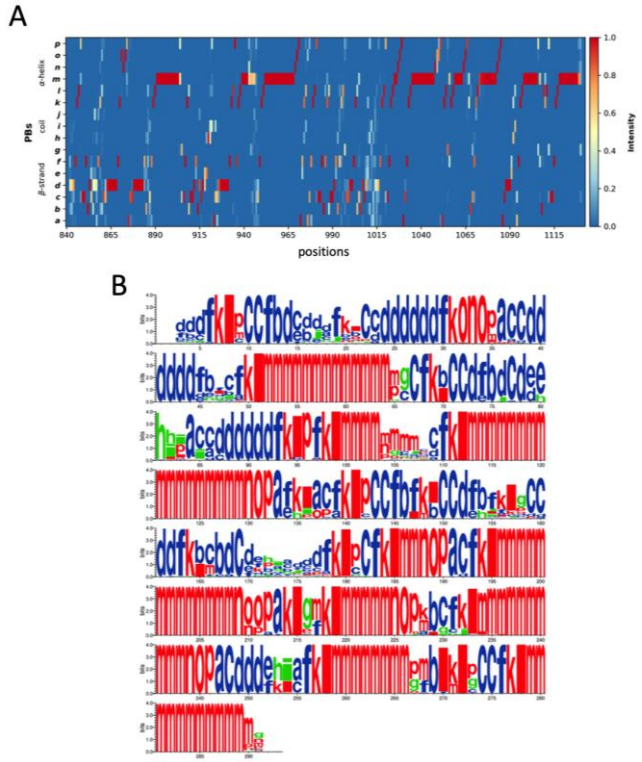

Fig. 4. Analysis of Molecular Dynamics simulations. (A) JH1 apo-pTyr1007 domain Molecular Dynamic simulations [54] translated into Protein Blocks [43] and (B) shown with WebLogo [53].

PB assignments are done for each residue of the protein and over every snapshot extracted from MD simulations. The equivalent number of PBs (N_{eq}) is a statistical measurement similar to entropy that represents the average number of PBs for a residue at a given position. N_{eq} is calculated as follows [43]:

$$Neq = \exp\left(-\sum_{x=1}^{16} f_x \ln f_x\right)$$

Where, f_x is the probability of PB x. A N_{eq} value of 1 indicates that only one type of PB is observed, while a value of 16 is equivalent to a random distribution.

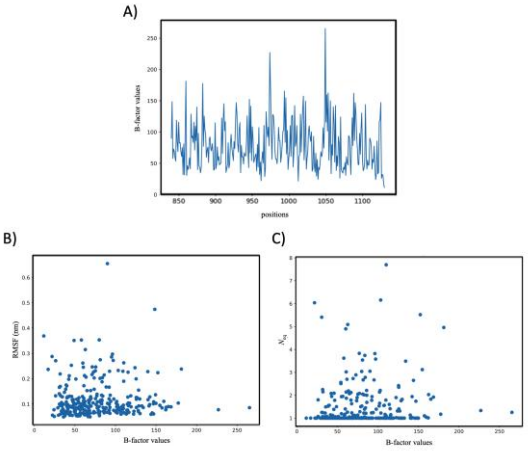

Fig 5. Analysis of B-factor values. (A) Raw values of B-factor, (B) B-factor values vs. RMSF values, and (C) B-factor values vs. RMSF values. Data taken from [54].

Interestingly, Figure 5 shows the comparison between experimental values of B-factors (see Figure 5A) and other computational measures. The correlation between B-factors and RMSF (Figure 5B) never exceeds 0.5, and that between B-factors and N_{eq} (Figure 5C) never exceeds 0.2. This result is logical because both experimental and computational approaches have limitations. RMSF is closer to B-factors because the comparison is global, while N_{eq} always remains local and therefore seems very distant. This is the interest of N_{eq}; it allows localizing rigid regions in very flexible areas, while RMSF can be minimal there, as shown in example [52,55-57].

It is so implemented in PBxplore tool [46] and recently in Menger_Curvature [58]. It had been used in numerous studies such as integrin [55-57,59], V_HHs (camel antibodies) [60-67], KISS1R [68], NMAD receptor channel [69], CALR [54], JAK2 [54], DARC [54,70] or *RHD* blood group proteins [71-73].

5 For a first link between order and disorder

With the availability of PBxplore [46], it became easy to use PBs to analyse molecular dynamics data. However, having no experience and the amount of structural data being initially small on IDRs/IDPs, it was difficult to tackle it.

A dedicated database will allow for a better understanding of this. Thus, the May 2018 release of PED3: Protein Ensemble Database, the database of conformational ensembles describing flexible proteins [30] included 24 entries describing different IDPs with different techniques, i.e. SAXS and NMR, NMR alone, SAXS alone and Molecular Dynamics [31,74-83]. Each set was encoded using the structural alphabet and analysed. The results are compared with the analysis of the ordered protein MD simulations previously performed (169 proteins) [42,84,85].

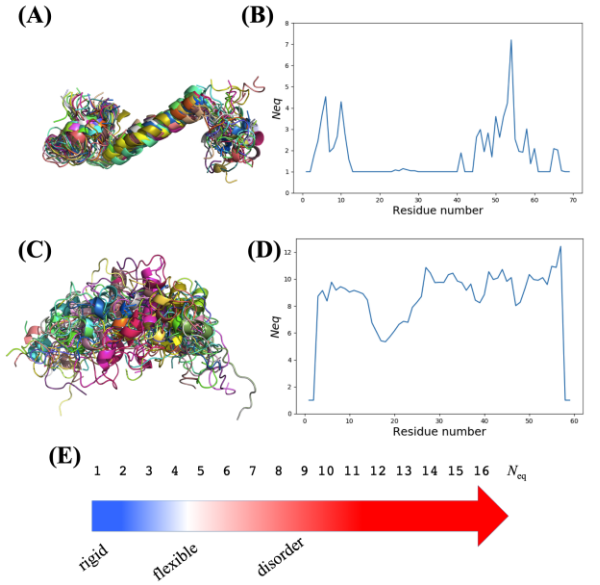

Fig 6. Analysis of PED database. (A) and (B) PED entry 2AAA, (C) and (D) PED entry 8AAC, with (A) and (C) superimposition of structural model ensemble, and (B) and (D) N_{eq} values. (E) N_{eq} gradient from order to pure disorder.

Figures 6A to 6D show two representative examples of what was observed. PED entry 2AAA is representative of PDIs that are not completely disorganised or even very flexible (see Figure 6B). The central part of 2AAA is a rigid helix (N_{eq} equal to 1), whereas N-terminus and especially C-terminus are more flexible (see Figure 6A). The maximum N_{eq} is only 7.

PED entry 8AAC is much more characteristic of what we expected, with very high N_{eq}, mostly above 8 (see Figure 6D), while even individual structures have locally constrained conformations (see Figure 6C). On average, the observations are predominantly driven by rigid local conformations, i.e. a N_{eq} value of 1.0, which corresponds to 60% of the residues in the ordered globular protein dataset [85] and 58% in the IDP dataset [86]. The trends change dramatically for $N_{eq} > 2.0$, as only 8% of the residues in the globular protein dataset and 36% of the residues in the IDP dataset belong to this group. Almost none of the residues in the globular protein dataset (i.e. 0.82%) have $N_{eq} > 4.0$ and none have $N_{eq} > 8.0$, whereas 31% of the IDP residues have $N_{eq} > 4.0$ and 15% have $N_{eq} > 8.0$. This suggests that the entropy index N_{eq} is relevant to the question of protein order and disorder, as we previously hypothesised [52].

Therefore, after visual inspection of each entry (see Figure 6E), we decided that a Neq greater than 8 should be considered fully disordered. By reviewing our many previous molecular dynamics simulations [52,55,56,84,85], we found that a N_{eq} close to 4 could be considered flexible. Extending this simple approach, a N_{eq} of 6 can be an intermediate state between high flexibility and disorder [87]. It was even used to analyse disorder prediction [88]. Since, this result had been applied in different cases, the most recent ones being the effect of humanisation of V_HH [60] and the impact of Ruxolitinib drug on Janus Kinase 2 domain JH1 [54].

Acknowledgments. This work is dedicated to the memory of Pr. Serge Hazout (Université Paris Diderot, Paris, France) and Pr. Narayanaswamy Srinivasan (MBU, IISc, Bangalore, India), both of whom left too soon. I would like to thank Prof. Jean-Claude Serge Levy (Université Paris Cité, Paris, France), who led ten Complexity-Disorder Meetings, enriching the work of all of us over the years. I would also like to thank my old colleagues, Pr. Catherine Etchebest (Université Paris Cité, Paris, France), Pr. Jean-Christophe Gelly (Université Paris Cité, Paris, France), Dr. Julien Diharce (Université Paris Cité, Paris, France), Pr. Bernard Offmann (Nantes Université, France), Dr. Vincent Jallu (APHP, Paris, France), Dr. Aline Floch (EFS Créteil, France), Pr. Nenad Mitic (Belgrade University, Serbia) and Dr. Joseph Rebehmed (LAU, Beirut, Lebanon) who are accompanying me in this research. I would also like to thank my parents, Georges and Françoise, and my wife, Hélène, as well as Snoopy, my mischievous little dachshund.

Funding. These works were supported by the France 2030 program through the Idex Université Paris Cité (ANR-18-IDEX-0001), and me by INSERM.

References

1. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. The protein data bank. Nucleic acids research **28**, 235-242 (2000)
2. V. Nallapareddy, N. Bordin, I. Sillitoe, M. Heinzinger, M. Littmann, V.P. Waman, N. Sen, B. Rost, C. Orengo. Cathe: Detection of remote homologues for cath superfamilies using embeddings from protein language models. Bioinformatics (Oxford, England) **39**, (2023)
3. R.L.M. van Montfort, P. Workman. Structure-based drug design: Aiming for a perfect fit. Essays in biochemistry **61**, 431-437 (2017)
4. O. Carugo. B-factor accuracy in protein crystal structures. Acta crystallographica. Section D, Structural biology **78**, 69-74 (2022)
5. O. Carugo. How large b-factors can be in protein crystal structures. BMC bioinformatics **19**, 61 (2018)
6. D. Ringe, G.A. Petsko. Mapping protein dynamics by x-ray diffraction. Progress in biophysics and molecular biology **45**, 197-235 (1985)
7. O. Carugo, P. Argos. Reliability of atomic displacement parameters in protein crystal structures. Acta crystallographica. Section D, Biological crystallography **55**, 473-478 (1999)
8. O. Carugo. How anisotropic and isotropic atomic displacement parameters monitor protein covalent bonds rigidity: Isotropic b-factors underestimate bond rigidity. Amino acids **53**, 779-782 (2021)
9. O. Carugo. Uses and abuses of the atomic displacement parameters in structural biology. Methods in molecular biology (Clifton, N.J.) **2449**, 281-298 (2022)
10. A. Kuzmanic, B. Zagrovic. Determination of ensemble-average pairwise root mean-square deviation from experimental b-factors. Biophysical journal **98**, 861-871 (2010)
11. A. Bornot, C. Etchebest, A.G. de Brevern. Predicting protein flexibility through the prediction of local structures. Proteins **79**, 839-852 (2011)
12. W.J. Cook, L.C. Jeffrey, M.L. Sullivan, R.D. Vierstra. Three-dimensional structure of a ubiquitin-conjugating enzyme (e2). The Journal of biological chemistry **267**, 15116-15121 (1992)
13. W.L. Delano. The pymol molecular graphics system on world wide web. Http://www.Pymol.Org. (2013)
14. M. Berjanskii, D.S. Wishart. Nmr: Prediction of protein flexibility. Nature protocols **1**, 683-688 (2006)
15. R. Ishima, D.A. Torchia. Protein dynamics from nmr. Nature structural biology **7**, 740-743 (2000)
16. K. Teilum, J.G. Olsen, B.B. Kragelund. Protein stability, flexibility and function. Biochimica et biophysica acta **1814**, 969-976 (2011)
17. C.C. Cornilescu, F. Bouamr, X. Yao, C. Carter, N. Tjandra. Structural analysis of the n-terminal domain of the human t-cell leukemia virus capsid protein. Journal of molecular biology **306**, 783-797 (2001)
18. V.N. Uversky, J.R. Gillespie, A.L. Fink. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins **41**, 415-427 (2000)
19. V.N. Uversky. Natively unfolded proteins: A point where biology waits for physics. Protein science : a publication of the Protein Society **11**, 739-756 (2002)
20. V.N. Uversky. Protein intrinsic disorder and structure-function continuum. Progress in molecular biology and translational science **166**, 1-17 (2019)
21. J.C. Hansen, X. Lu, E.D. Ross, R.W. Woody. Intrinsic protein disorder, amino acid composition, and histone terminal domains. The Journal of biological chemistry **281**, 1853-1856 (2006)
22. V.N. Uversky. A decade and a half of protein intrinsic disorder: Biology still waits for physics. Protein science : a publication of the Protein Society **22**, 693-724 (2013)
23. A.H. Phillips, R.W. Kriwacki. Intrinsic protein disorder and protein modifications in the processing of biological signals. Current opinion in structural biology **60**, 1-6 (2020)
24. R. Trivedi, H.A. Nagarajaram. Intrinsically disordered proteins: An overview. International journal of molecular sciences **23**, (2022)
25. A. Katuwawala, Z. Peng, J. Yang, L. Kurgan. Computational prediction of morfs, short disorder-to-order transitioning protein binding regions. Computational and structural biotechnology journal **17**, 454-462 (2019)
26. P. Mendoza-Espinosa, V. García-González, A. Moreno, R. Castillo, J. Mas-Oliva. Disorder-to-order conformational transitions in protein structure and its relationship to disease. Molecular and cellular biochemistry **330**, 105-120 (2009)
27. J. Song, L. Kurgan. Two decades of advances in sequence-based prediction of morfs, disorder-to-order transitioning binding regions. Expert review of proteomics **22**, 1-9 (2025)
28. J. Zhu, B.H. Luo, T. Xiao, C. Zhang, N. Nishida, T.A. Springer. Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. Molecular cell **32**, 849-861 (2008)
29. A. Sali, T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology **234**, 779-815 (1993)
30. M. Varadi, S. Kosol, P. Lebrun, E. Valentini, M. Blackledge, A.K. Dunker, I.C. Felli, J.D. Forman-Kay, R.W. Kriwacki, R. Pierattelli, *et*

al. Pe-db: A database of structural ensembles of intrinsically disordered and of unfolded proteins. Nucleic acids research **42**, D326-335 (2014)
31. Y.G. Sterckx, A.N. Volkov, W.F. Vranken, J. Kragelj, M.R. Jensen, L. Buts, A. Garcia-Pino, T. Jove, L. Van Melderen, M. Blackledge*, et al.* Small-angle x-ray scattering- and nuclear magnetic resonance-derived conformational ensemble of the highly flexible antitoxin paaa2. Structure **22**, 854-865 (2014)
32. D. Eisenberg. The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. Proc Natl Acad Sci U S A **100**, 11207-11210 (2003)
33. L. Pauling, R.B. Corey. The pleated sheet, a new layer configuration of polypeptide chains. Proc Natl Acad Sci U S A **37**, 251-256 (1951)
34. L. Pauling, R.B. Corey, H.R. Branson. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci U S A **37**, 205-211 (1951)
35. D.J. Barlow, J.M. Thornton. Helix geometry in proteins. Journal of molecular biology **201**, 601-619 (1988)
36. L. Regan. Protein structure. Born to be beta. Curr Biol **4**, 656-658 (1994)
37. W. Kabsch, C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22**, 2577-2637 (1983)
38. M. Bansal, S. Kumar, R. Velavan. Helanal: A program to characterize helix geometry in proteins. J Biomol Struct Dyn **17**, 811-819 (2000)
39. N. Fernandez-Fuentes, E. Querol, F.X. Aviles, M.J. Sternberg, B. Oliva. Prediction of the conformation and geometry of loops in globular proteins: Testing archdb, a structural classification of loops. Proteins **60**, 746-757 (2005)
40. Y. Mansiaux, A.P. Joseph, J.C. Gelly, A.G. de Brevern. Assignment of polyproline ii conformation and analysis of sequence--structure relationship. PloS one **6**, e18401 (2011)
41. P. Craveur, A.P. Joseph, J. Rebehmed, A.G. de Brevern. B-bulges: Extensive structural analyses of β-sheets irregularities. Protein science : a publication of the Protein Society **22**, 1366-1378 (2013)
42. P. Craveur, T.J. Narwani, N. Srinivasan, J.C. Gelly, J. Rebehmed, A.G. de Brevern. Shaking the β-bulges. IEEE/ACM transactions on computational biology and bioinformatics **19**, 14-18 (2022)
43. A.G. de Brevern, C. Etchebest, S. Hazout. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. Proteins **41**, 271-287 (2000)
44. A.G. de Brevern. New assessment of a structural alphabet. In silico biology **5**, 283-289 (2005)
45. M. Tyagi, C. Benros, J. Martin, A.G. de Brevern. Description of the local protein structure. Ii. Novel approaches. In *Recent research developments in protein engineering*, de Brevern, A.G., Ed. Research Signpost: Trivandrum 2007; pp 34-47.
46. J. Barnoud, H. Santuz, P. Craveur, A.P. Joseph, V. Jallu, A.G. de Brevern, P. Poulain. Pbxplore: A tool to analyze local protein structure and deformability with protein blocks. PeerJ **5**, e4013 (2017)
47. G. Faure, A.P. Joseph, P. Craveur, T.J. Narwani, N. Srinivasan, J.C. Gelly, J. Rebehmed, A.G. de Brevern. Ipbavizu: A pymol plugin for an efficient 3d protein structure superimposition approach. Source Code Biol Med **14**, 5 (2019)
48. Y. Ghouzam, G. Postic, P.E. Guerin, A.G. de Brevern, J.C. Gelly. Orion: A web server for protein fold recognition and structure prediction using evolutionary hybrid profiles. Sci Rep **6**, 28268 (2016)
49. B. Schneider, J. Cerny, D. Svozil, P. Cech, J.C. Gelly, A.G. de Brevern. Bioinformatic analysis of the protein/DNA interface. Nucleic acids research **42**, 3381-3394 (2014)
50. S. Leonard, A.P. Joseph, N. Srinivasan, J.C. Gelly, A.G. de Brevern. Mulpba: An efficient multiple protein structure alignment method based on a structural alphabet. J Biomol Struct Dyn **32**, 661-668 (2014)
51. L. Fourrier, C. Benros, A.G. de Brevern. Use of a structural alphabet for analysis of short loops connecting repetitive structures. BMC bioinformatics **5**, 58 (2004)
52. P. Craveur, A.P. Joseph, J. Esque, T.J. Narwani, F. Noël, N. Shinada, M. Goguet, S. Leonard, P. Poulain, O. Bertrand*, et al.* Protein flexibility in the light of structural alphabets. Frontiers in molecular biosciences **2**, 20 (2015)
53. G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner. Weblogo: A sequence logo generator. Genome Res **14**, 1188-1190 (2004)
54. H.N. Vu, R. Radjasandirane, J. Diharce, A.G. de Brevern. Impact of ruxolitinib interactions on jak2 jh1 domain dynamics. International journal of molecular sciences **26**, 3727 (2025)
55. M. Goguet, T.J. Narwani, R. Petermann, V. Jallu, A.G. de Brevern. In silico analysis of glanzmann variants of calf-1 domain of α(iib)β(3) integrin revealed dynamic allosteric effect. Sci Rep **7**, 8001 (2017)
56. V. Jallu, P. Poulain, P.F. Fuchs, C. Kaplan, A.G. de Brevern. Modeling and molecular dynamics of hpa-1a and -1b polymorphisms: Effects on the structure of the β3 subunit of the αiibβ3 integrin. PloS one **7**, e47304 (2012)
57. M.Y. Koker, N. Sarper, C. Albayrak, B. Zulfikar, E. Zengin, B. Saraymen, D.

Albayrak, B. Koc, H. Avcilar, M. Karakükcü, *et al.* New αiibβ3 variants in 28 turkish glanzmann patients; structural hypothesis for complex activation by residues variations in i-egf domains. Platelets **33**, 551-561 (2022)
58. E. Reboul, J. Marien, C. Prévost, A. Taly, S. Sacquin-Mora. Menger_curvature : A mdakit implementation to decipher the dynamics, curvatures and flexibilities of polymeric backbones at the residue level. bioRxiv, 2025.2004.2004.647214 (2025)
59. V. Jallu, P. Poulain, P.F. Fuchs, C. Kaplan, A.G. de Brevern. Modeling and molecular dynamics simulations of the v33 variant of the integrin subunit β3: Structural comparison with the l33 (hpa-1a) and p33 (hpa-1b) variants. Biochimie **105**, 84-90 (2014)
60. C. Martins, J. Diharce, A.A. Nadaradjane, A.G. de Brevern. Evaluation of the potential impact of in silico humanization on v(h)h dynamics. International journal of molecular sciences **24**, (2023)
61. C. Martins, F. Gardebien, A.A. Nadaradjane, J. Diharce, A.G. de Brevern. A simple analysis of the second (extra) disulfide bridge of v(h)hs. Molecules (Basel, Switzerland) **29**, (2024)
62. A. Melarkode Vattekatte, N.K. Shinada, T.J. Narwani, F. Noël, O. Bertrand, J.P. Meyniel, A. Malpertuy, J.C. Gelly, F. Cadet, A.G. de Brevern. Discrete analysis of camelid variable domains: Sequences, structures, and in-silico structure prediction. PeerJ **8**, e8408 (2020)
63. A.A. Nadaradjane, J. Diharce, J. Rebehmed, F. Cadet, F. Gardebien, J.C. Gelly, C. Etchebest, A.G. de Brevern. Quality assessment of v(h)h models. J Biomol Struct Dyn **41**, 13287-13301 (2023)
64. F. Noël, A. Malpertuy, A.G. de Brevern. Global analysis of vhhs framework regions with a structural alphabet. Biochimie **131**, 11-19 (2016)
65. A.M. Vattekatte, F. Cadet, J.C. Gelly, A.G. de Brevern. Insights into comparative modeling of v(h)h domains. International journal of molecular sciences **22**, (2021)
66. A.M. Vattekatte, J. Diharce, J. Rebehmed, F. Cadet, F. Gardebien, C. Etchebest, A.G. de Brevern. General trends of the camelidae antibody v(h)hs domain dynamics. International journal of molecular sciences **24**, (2023)
67. P. Vishwakarma, A.M. Vattekatte, N. Shinada, J. Diharce, C. Martins, F. Cadet, F. Gardebien, C. Etchebest, A.A. Nadaradjane, A.G. de Brevern. V(h)h structural modelling approaches: A critical review. International journal of molecular sciences **23**, (2022)
68. L. Chevrier, A. de Brevern, E. Hernandez, J. Leprince, H. Vaudry, A.M. Guedj, N. de Roux. Prr repeats in the intracellular domain of kiss1r are important for its export to cell membrane. Molecular endocrinology (Baltimore, Md.) **27**, 1004-1014 (2013)
69. M. Ladislav, J. Cerny, J. Krusek, M. Horak, A. Balik, L. Vyklicky. The lili motif of m3-s2 linkers is a component of the nmda receptor channel gate. Frontiers in molecular neuroscience **11**, 113 (2018)
70. A.G. de Brevern, H. Wong, C. Tournamille, Y. Colin, C. Le Van Kim, C. Etchebest. A structural model of a seven-transmembrane helix receptor: The duffy antigen/receptor for chemokine (darc). Biochimica et biophysica acta **1724**, 288-306 (2005)
71. A.G. de Brevern, A. Floch, A. Barrault, J. Martret, G. Bodivit, R. Djoudi, F. Pirenne, C. Tournamille. Alloimmunization risk associated with amino acid 223 substitution in the rhd protein: Analysis in the light of molecular modeling. Transfusion **58**, 2683-2692 (2018)
72. A. Floch, T. Galochkina, F. Pirenne, C. Tournamille, A.G. de Brevern. Molecular dynamics of the human rhd and rhag blood group proteins. Frontiers in chemistry **12**, 1360392 (2024)
73. A. Floch, F. Pirenne, A. Barrault, B. Chami, C. Toly-Ndour, C. Tournamille, A.G. de Brevern. Insights into anti-d formation in carriers of rhd variants through studies of 3d intraprotein interactions. Transfusion **61**, 1286-1301 (2021)
74. J.R. Allison, R.C. Rivers, J.C. Christodoulou, M. Vendruscolo, C.M. Dobson. A relationship between the transient structure in the monomeric state and the aggregation propensities of α-synuclein and β-synuclein. Biochemistry **53**, 7170-7183 (2014)
75. V.R. Bacot-Davis, J.J. Ciomperlik, H.A. Basta, C.C. Cornilescu, A.C. Palmenberg. Solution structures of mengovirus leader protein, its phosphorylated derivatives, and in complex with nuclear transport regulatory protein, rangtpase. Proceedings of the National Academy of Sciences **111**, 15792-15797 (2014)
76. A. De Biasio, A. Ibanez de Opakua, T.N. Cordeiro, M. Villate, N. Merino, N. Sibille, M. Lelli, T. Diercks, P. Bernado, F.J. Blanco. P15paf is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. Biophysical journal **106**, 865-874 (2014)
77. J.A. Marsh, J.D. Forman-Kay. Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. Journal of molecular biology **391**, 359-374 (2009)
78. H.D. Mertens, A. Piljic, C. Schultz, D.I. Svergun. Conformational analysis of a genetically encoded fret biosensor by saxs. Biophysical journal **102**, 2866-2875 (2012)

79. T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, M. Tyers, J.D. Forman-Kay. Structure/function implications in a dynamic complex of the intrinsically disordered sic1 with the cdc4 subunit of an scf ubiquitin ligase. Structure **18**, 494-506 (2010)
80. V. Ozenne, R. Schneider, M. Yao, J.R. Huang, L. Salmon, M. Zweckstetter, M.R. Jensen, M. Blackledge. Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. J Am Chem Soc **134**, 15138-15148 (2012)
81. M. Sanchez-Martinez, R. Crehuet. Application of the maximum entropy principle to determine ensembles of intrinsically disordered proteins from residual dipolar couplings. Phys Chem Chem Phys **16**, 26030-26039 (2014)
82. S.G. Sivakolundu, D. Bashford, R.W. Kriwacki. Disordered p27kip1 exhibits intrinsic structure resembling the cdk2/cyclin a-bound conformation. Journal of molecular biology **353**, 1118-1128 (2005)
83. S.D. Weeks, E.V. Baranova, M. Heirbaut, S. Beelen, A.V. Shkumatov, N.B. Gusev, S.V. Strelkov. Molecular structure and dynamics of the dimeric human small heat shock protein hspb6. J Struct Biol **185**, 342-354 (2014)
84. T.J. Narwani, P. Craveur, N.K. Shinada, A. Floch, H. Santuz, A.M. Vattekatte, N. Srinivasan, J. Rebehmed, J.C. Gelly, C. Etchebest*, et al.* Discrete analyses of protein dynamics. J Biomol Struct Dyn **38**, 2988-3002 (2020)
85. T.J. Narwani, P. Craveur, N.K. Shinada, H. Santuz, J. Rebehmed, C. Etchebest, A.G. de Brevern. Dynamics and deformability of α-, 310- and π-helices. Archives of Biological Sciences **70**, 21-31 (2018)
86. A. Melarkode Vattekatte, T.J. Narwani, A. Floch, M. Maljković, S. Bisoo, N.K. Shinada, A. Kranjc, J.C. Gelly, N. Srinivasan, N. Mitić*, et al.* Data set of intrinsically disordered proteins analysed at a local protein conformation level. Data in brief **29**, 105383 (2020)
87. M.V. Akhila, T.J. Narwani, A. Floch, M. Maljković, S. Bisoo, N.K. Shinada, A. Kranjc, J.-C. Gelly, N. Srinivasan, N. Mitić*, et al.* A structural entropy index to analyse local conformations in intrinsically disordered proteins. Journal of Structural Biology **210**, 107464 (2020)
88. A.G. de Brevern. Analysis of protein disorder predictions in the light of a protein structural alphabet. Biomolecules **10**, (2020)