Fold or flop: quality assessment of `AlphaFold` predictions on whole proteomes

Edoardo Sarti and Frédéric Cazals, Centre Inria at Université Côte d'Azur, France*

January 15, 2026

Abstract

Motivation: Reliability of `AlphaFold` predictions is mainly assessed using the predicted Local Distance Difference Test (pLDDT). For model organisms, 30–40% of residues fall into the low-confidence pLDDT range. Moreover, pLDDT sometimes fails to flag physically implausible structures. This raises two questions: can more robust reliability indicators be identified, and do unreliable predictions share common structural or biophysical features?
Results: We use packing-based dimensionality reduction and clustering to assess the quality of whole-proteome predictions in the AlphaFold Database (AFDB), tracing connections with domain annotations, intrinsic disorder and stereochamical measures. We thus chart pathological structural motifs and unsupported disorder predictions, that reveal strengths and limitations of current self-assessment metrics and allow us to define a novel, misfold-aware structure quality assessment score.
Availability: The code to compute arity maps is available within the Structural Bioinformatics Library.
See: AlphaFold analysis, and also Documentation, Applications, Installation guide.
The code and data for rerunning analyses are made available at doi.org/10.5281/zenodo.18216693
Contact: frederic.cazals@inria.fr
online.

Keywords: `AlphaFold`, `AlphaFold-DB`, Model Quality Assessment, Packing properties, Fold prediction, Intrinsically Disordered Regions.

1 Introduction

`AlphaFold` has revolutionized protein structure prediction, which in favorable cases achieves high accuracy comparable to experimental methods [1]. Its companion resource, the AlphaFold Protein Structure Database (`AlphaFold-DB`) now hosts more than 200 million predicted structures generated by `AlphaFold2`. `AlphaFold-DB` has rapidly become the primary reference for structural information on UniProt sequences, serving both as a comprehensive repository and as a springboard for derivative resources [2, 3] extending even more the range of predicted structures.

Each amino acid of any `AlphaFold` prediction is assigned a confidence score known as pLDDT (predicted Local Distance Difference Test). It is derived from the Local Distance Difference Test (LDDT [4]), a superposition-free score that evaluates how well interatomic distances in a predicted structure match those in a reference. During training, `AlphaFold` was supervised using known X-ray and CryoEM experimental structures, with a dedicated confidence head in the network predicting the expected LDDT score for each residue. The resulting pLDDT values are normalized from 0 to 1, where higher values indicate greater confidence in local geometry. For simplicity, four confidence categories have been defined based on pLDDT values: very high ($0.9 \leq$ pLDDT, dark blue), high ($0.7 \leq$ pLDDT < 0.9, cyan), low ($0.5 \leq$ pLDDT < 0.7, yellow), and very low (pLDDT < 0.5, orange). Importantly, pLDDT reflects local accuracy only, without assessing global

*Frederic.Cazals@inria.fr, Edoardo.Sarti@inria.fr

domain orientations or inter-chain relationships, which are instead captured by complementary metrics such as Predicted Alignment Error (PAE).

The structural features within AlphaFold predictions have been examined using the Geometricus algorithm [5], which represents protein structures through a series of steps: (i) a decomposition of the protein into structural fragments (based on k-mers or spherical neighborhoods), (ii) a discrete encoding of invariants (moments) of these fragments into shape-mers, and (iii) a counting step of shape-mers yielding an embedding of the protein. By clustering these embeddings – using non-negative matrix factorization – across 21 predicted proteomes from AlphaFold-DB, 250 structural clusters were identified, including 20 major clusters or superfamilies [6].

The pLDDT metric has been thoroughly analyzed, revealing its effectiveness as an indicator of result robustness. It has been observed that the low-pLDDT regions in AlphaFold predictions correlate with observed or predicted intrinsically disordered regions (IDRs) [7]. Indeed AlphaFold outperforms state-of-the-art methods such as IUPred2 on disordered region detection on a benchmark of DisProt targets and PDB controls [6]. Also, high mean pLDDT values are associated with minimal structural differences between AlphaFold and trRosetta models [6].

For the 49 model organisms whose entire proteome have been predicted and deposited on AlphaFold-DB, 30% to 40% of all amino acids fall within the low or very low pLDDT confidence categories (Table S1; Figures S1 to S3.). Although a considerable fraction could be explained by intrinsically disordered regions (IDRs), many low pLDDT amino acids are situated in the bulk, or are not part of low pLDDT segments. Moreover, pLDDT does not correlate with the local flexibility of globular proteins [8]. Even in regions predicted with high confidence, AlphaFold models can disagree with experimentally determined electron density maps concerning domain orientations, local backbone, and side-chain conformations [9]. Finally, AlphaFold lacks robustness to minor, biologically insignificant sequence perturbations. Adversarial sequences with only five altered residues can lead to drastically different 3D structures, yet AlphaFold often predicts these with similar confidence to original sequences [10].

Despite the evidence of faults in a small yet significant fraction of AlphaFold predictions, reports have so far remained anecdotal and linked to the context of specific studies. Is there a common denominator between AlphaFold incorrect predictions? Could we use this information to create a better confidence measure?

This study defines the arity map as a simple and effective dimensional reduction / clustering technique that can serve as proxy for structural quality and fold classification. The arity map is shown to distinguish between intrinsic disorder and misfold predictions, and it can be employed for meaningful global and per-residue quality assessment measure.

2 Methods

2.1 Packing analysis

2.1.1 Arity and arity signature

A natural parameter to identify compact regions in proteins in the number of neighbors of a C_α carbon. Let the *arity* of a C_α carbon be the number of neighboring C_αs within a distance range r. As classically done, we take $r = 10$Å to account for non-covalent contacts. We note in passing that such neighbors are sufficient to robustly infer protein domains using a direct application of spectral clustering [11]. Assuming the polypeptide chain has n amino acids, define:

- $L_a = \{a_n, \ldots, a_n\}$: arities of the n individual C_α carbons;
- $L = [A_1, \ldots, A_m]$: unique arities sorted by increasing value;

The number of C_α with a given arity is used to define the histogram H_{C_α}. We note in passing that the comparison of two normalized such histograms (unit cumulated sum) can be done using the mass transport solver from [12] implemented in the *Python Optimal Transport* package [13].

To simplify the presentation provided the H_{C_α} and the associated CDF_{C_α}, we define the following signature:

Definition. 1 *(Arity signature $Sig._{C_\alpha}(P)$ of a polypeptide chain P.) Consider the cumulative distribution functions associated with the number of C_α carbons, that is CDF_{C_α}. Consider a list $P_K = \{p_1, \ldots, p_K\}$ of increasing percentiles. For $p_k \in P_K$, let $\texttt{arity_}C_\alpha(p_k)$ be the smallest arity such that the CDF is larger than p_k, that is*

$$\texttt{arity_}C_\alpha(p_k) = CDF^{-1}_{C_\alpha}(p_k) := \arg\min_j CDF_{C_\alpha}(A_j) \geq p_k. \tag{1}$$

The C_α signature *is the increasing sequence $Sig._{C_\alpha}(P) = \{\texttt{arity_}C_\alpha(p_1), \ldots, \texttt{arity_}C_\alpha(p_K)\}$.*

Using two quantiles $\{q_1 = 0.25, q_2 = 0.75\}$, the arity signature refers to the two arity values required to gather 25% and 75% percent of the number of amino acids in the chain.

For a collection of structures, typically all predictions in a genome, we define:

Definition. 2 *Given two quantiles $q_1(= 0.25)$ and $q_2(= 0.75)$, the* arity map *is the map whose x and y axis are the arities at q_1 and q_2. A* bin/cell *of the map hosts all structures with a prescribed arity signature.*

2.1.2 Filtrations

The arity just defined can also be used to *build* the protein incrementally by inserting the a.a. in order of increasing arity. In doing so, we maintain the connected components, that is the contiguous stretches of a.a. along the sequence, which are formed once all a.a. up to a given arity have been inserted. Maintaining these connected components (and their number) is done using a Union-Find data structures [14].

To see the interest of this strategy, consider say a protein with two regions, one loose and one packed. Inserting the a.a. by increasing arity will sequentially reveal the former and the latter.

The *persistence diagram* associated with these components is the scatter plot whose x-axis (resp. y-axis) corresponds to the birth (resp. the death) of a component, in arity units. For more details on filtrations, see [15] and the companion paper [16].

2.2 Disorder predictors

`DisProt` We worked with the June 2025 version of the `DisProt` database which includes 413 disorder annotations for proteins that do not correspond to any PDB entry. The proteins come from 135 different species, spanning eukaryotes, prokaryotes and viruses. For each one of these sequences we compared the AFDB prediction and the `DisProt` annotation, where both were available, complete and unambiguous. For the remaining 371 proteins, we counted the percentage of residues that are associated with a pLDDT of less than 50 ("very low" confidence) but are not annotated as disordered in `DisProt`.

`AIUPred` `AIUPred` predictions were computed with standard parameters starting from the sequences of the `AlphaFold-DB` predictions. A disagreement between `AIUPred` disorder and `AlphaFold2` occurs when the per-residue `AIUPred` disorder score is ≥ 0.5 and its pLDDT is ≤ 50.

2.3 Structural quality assessment

`PROCHECK`. We ran the `PROCHECK` v.3.5.4 with resolution 1.00 A and collected the total G-factor of each prediction. The G-factor is a log-odds measure: a structure is considered and outlier when its G-factor is ≤ -1.

`PhiSiCal-Checkup`. We ran `PhiSiCal-Checkup` v.1.0 which provides a per-residue score for a wide range of dihedral angle distributions. By default, outlier values are defined having a score of ≤ -4. For each protein, we counted the number of residues having at least one outlier score.

`VoroMQA`. `VoroMQA` returns a per-structure normalized packing score. A score of < 0.5 determines a poor-quality structure in terms of packing. We ran `VoroMQA` on the whole human proteome and counted (per bin) the percentage of structures receiving a poor quality score.

2.4 Pathological structural motifs dataset

Within the `AlphaFold-DB` predictions with non-standard arity values, we find an overrepresentation of certain unlikely or nonphysical structural motifs. The two predominant ones are 100- to 300-residue long helices (often containing narrow bends), and beta solenoid structures. These structures have been observed and catalogued as two recurrent `AlphaFold2` hallucinations [17].

We manually collected examples of predictions containing unlikely structural motifs. The secondary structure elements (SSE) was defined using `DSSP`. Then, we calculated the curvature of the center line of each SSE.

To define the SSE center line with minimal complexity, we first calculate the center of mass of three successive intervals of N C_α shifted by k residues, where $N = 3, k = 2$ for beta sheets and $N = 4, k = 4$ for alpha helices. We then plot a B-spline $C(t)$ of degree 3 through the said center of masses.

The local curvature radius at any given point is then calculated with the usual formula [18]:

$$R(t) = \frac{||C'(t)||^3}{||C'(t)|| \times ||C''(t)||}.$$

Using the residue-wise curvature value as decision parameter, we then define a secondary structure element as pathological according to the following criteria:

- For alpha helices, whether the curvature radius of the helix center line over 12 residues is lower than 30 angstroms;
- For beta strands, whether the curvature radius of the beta sheet center line over 4 residues is lower than 5 angstrom, or whether the torsion of the beta sheet over 4 residues is greater than 40 degrees.

These structural motifs compose the negative dataset for structured domains. We composed a balanced positive dataset with TED high-quality, secondary structure rich domains.

We used about 3000 positive-labeled and 3000 negative-labeled residues extracted from 135 pathological motifs and an equal number of TED high-quality domains chosen randomly.

2.5 IDRs cross-validation dataset

For long, unstructured parts, we composed a parallel training dataset by equally sampling all arity boxes. Residues were given a normalized score based on their pLDDT and `AIUPred` as well as on the abundance f_{DisProt} of false positive disorder with respect to `DisProt` found in the corresponding arity cell:

- if predictions agree (pLDDT ≤ 50 and `AIUPred` ≥ 50 or vice-versa), the given score is $1 - f_{\text{DisProt}}$;
- if predictions disagree (pLDDT and `AIUPred` both ≥ 50 or ≤ 50), the given score is f_{DisProt}.

2.6 Novel, arity-based structure quality assessment method

From our balanced binary-labeled dataset of 6000 residues we calculated the following twenty two features:

- number of neighbors within balls or radii 6, 8, 10, 12, 14 angstroms;
- minimum and maximum arity in the intervals [0-10), [10-30), [30-50), [50-70), [70-90), [90-100];
- derivative of the arity at 10, 30, 50, 70, 90 percentiles.

These features were used to train a support vector machine (SVM) model with radial basis function (RBF) kernel on the data collected in our dataset.

For the unstructured parts, we trained a support vector regression (SVR) model with logit transformation.

We then tested on manually chosen predictions showing unfeasible motifs and control cases, none of which is included in the training dataset.

3 Results

3.1 Charting whole proteomes with the arity map

Arity is effective for distinguishing ordered, disordered, and mixed protein regions. Whereas proteins with canonical α or β folds typically exhibit arity values in the range of 15–20, unstructured regions show a markedly lower mode around 5–7 (Figure 1).

To visualize these properties at scale, we take the first and third quartiles (the 25th and 75th percentiles) of the neighbor distribution as coordinates and compile a histogram over a set of protein structures. This yields the *arity map*, a natural two-dimensional embedding that we use to probe the quality of `AlphaFold` predictions–Def. 2. The interpretation of this map enables proteome-wide assessment at a glance. We generated arity maps for all model proteomes in `AlphaFold-DB`. In the following, we present our analysis of the human proteome (Figure 2 A). Calculations on other genomes from `AlphaFold-DB` are easily conducted using the software accompanying this work, available within the Structural Bioinformatics Library [19], see AlphaFold analysis.

The x (resp. y) axis $\texttt{arity_}C_\alpha(q_{25})$ (resp. $\texttt{arity_}C_\alpha(q_{75})$) is geared towards loosely (resp. tightly) packed regions. For a given cell of this map, the distance to the diagonal $y = x$ encodes the uniformity of the overall packing of the protein. For example, for bins on the diagonal, a unique arity value suffices to move from a fraction of amino acids $< 25\%$ to a fraction $> 75\%$, stressing the uniformity of the overall 3D structure. This is never the case for native, non-trivial protein structures, which always present different quantiles q_{25} and q_{75} (Figure 1, Second row) due to their interchange between rigid and flexible regions. Such a collapse is observed on a minority of `AlphaFold` predictions (an example in Figure 2 A, AF-Q8IWJ2-F1).

For all proteomes, the densest region of the arity map locates in a short diagonal strip centered in $[15, 21]$ referred by *native sector*, or NS, in the following, which corresponds to well-folded proteins. Other densely populated regions are typically found in the wider $[5, 5] \times [15, 15]$ sector (*heterogenous sector*, or HS), but their occurrence and exact location largely depends on the species and the quality of predictions. For example, *P. falciparum* exhibits a continuum between the two sectors, whereas *E. coli* only present high density in the native sector. The case of *H. sapiens* is also remarkable as it presents a particularly populated bin at 7×13 in addition to the more typical denser region at very low arity values.

3.2 Structural characterization of arity map sectors

In order to confirm that well-folded proteins are mapped in the native sector, we compared the arity maps of the AF2-predicted human proteome and the set of all human proteins deposited on the PDB. As expected, the PDB structures only cluster in the native sector. (Figure 2 B). The arity map can thus cluster well-folded protein structures. In passing, we note that arity and pLDDT strongly correlate (Figure 2 C), confirming that the arity is also a good proxy for prediction robustness. Yet, pLDDT is not able to show the finer details of the structural repartition that emerges from the arity map.

Proteomes of all species also present considerable percentages of proteins with intrinsically disordered regions IDR, which in eukaryotes can often sum up to 20-30% [20]. `AlphaFold2` has been found to reliably predict intrinsic disorder, describing IDRs as unwound, maximally extended regions. Nonetheless, artifacts like the prediction for AF-Q8IWJ2-F1 (Figure 2 A) also present looser packing. Is the arity map able to separate correctly-predicted intrinsically disordered regions and proteins from artifacts?

In order to quantitatively characterize the arity map density population, we mapped the domains described in the recent TED database [3] on the human proteome (Figure 3, Top Row). TED domains are

grouped in three quality classes, depending on whether consensus has been found among three different approaches for recognizing structural domains.

For each protein, the fraction of amino acids assigned to each TED class is computed, with the fractions summing to one. Each cell of the arity map represents the average fraction associated with a given class.

As expected, high-quality (consensual) domains are enriched in the native sector, whereas medium- and low- quality domains are predominantly present in the heterogeneous sector, albeit in two different locations. The heterogeneous sector thus splits in three different sub-locations, one enriched in low-quality domains, one in medium-quality domains, and another where domain identification is scarce.

3.3 Arity as a fold classification

We now assess the ability of the arity map to isolate folds. In order to do so, we compare our results with the ones of `DPCstruct`, a recent unsupervised algorithm able to cluster structural motifs. The unsupervised character of the algorithm allows it to find structural similarities that have not yet been classified, while being in agreement with the known folds described in `SCOP` and `CATH` [21].

By running `DPCstruct` over the whole human genome, we found 324 metaclusters, whose size is distributed exponentially (60% of them having 20 members or less). Since `DPCstruct` classifies contiguous segments of protein sequences forming structural moduli, each prediction can contain multiple or no domains, with the vast majority containing 1-2 domains up to a maximum of 8 domains in the human proteome. Even so, a total of only 17212 domains are found for the 23391 human proteins. AF2 predictions deprived of `DPCstruct` domains are clustered in one region of the heterogeneous sector (Figure 3 Middle Row, Left).

By plotting the maximum number of domains found in the predictions of each arity bin we can thus have an idea of the structural complexity of the predictions (Figure 3 Middle Row). The variance of this distribution correlates with the maximum, implying that bins containing many multidomain structures also contain single domains, but not the inverse. The map has thus regions where only low-complexity structures are found.

Otherwise, `DPCstruct` cluster are uniformly spread in the rest of the arity map: the cluster label entropy per bin is similar everywhere, and the entropy of the structures associated to a certain label correlates with the the cluster's log-size.

3.4 AF2 misfolds do not always imply intrinsic disorder

Several recent works assume that in `AlphaFold`'s predictions, regions with low pLDDT and/or not displaying secondary nor tertiary structure are to be considered intrinsically disordered rather than just the product of insufficient or unexploitable information. The tools developed in the previous sections can help assessing whether and when a correlation between unstructured regions in `AlphaFold`'s prediction correlate with experimental evidences of disordered regions.

Analysis using `DisProt`. We first consider the `DisProt` database, which collects disordered regions from more than 1300 target protein sequences. Many of the benchmarked sequences correspond to proteins of known structure, present in the PDB. For this test, we focus on the 371 targets lacking structural data and having a predicted structure on `AlphaFold-DB`. The proteins are issued from 135 different species. Consider all amino acids of all proteins in a given arity map cell. We compute the fraction of false positive disorder found by `AlphaFold` with respect to the experimental annotations collected by `DisProt`. Notably, `DisProt` is known not to provide complete annotations, thus in principle `AlphaFold` could discover new unannotated IDRs on the `DisProt` targets. The simplest hypothesis consists of expecting an incompleteness proportional to the extent of the predicted disordered regions–regardless of other structural characteristics of `AlphaFold` predictions. However, while `AlphaFold` predicts a large number of amino acids as part of IDRs for the region of the arity map $\texttt{arity_}C_\alpha(0.25) < 7$, most of these are not annotated as such in `DisProt` (Figure 3 Bottom Row, Left). Predictions populating that part of the arity plot are thus likely to be of poor quality: this is not surprising, as proteins in this region often have a low number of homologous structures.

Analysis using AIUPred. AlphaFold2 predictions have been found to generally agree with state-of-the-art disorder predictors, among which AIUPred [22], a recent neural-network-based implementation of IUPred trained on IUPred predictions. AIUPred (resp. AlphaFold) predicts a residue as disordered when the AIUPred score (resp. pLDDT) is > 50 (resp. < 50.) We therefore use the arity map to study the (lack of) homogeneity between these two annotations. We ran AIUPred to predict IDRs on the whole human genome, and for each protein identified residues characterized by pLDDT ≥ 0.5 and a AIUPred score of > 0.5 (Figure 3 Bottom Row, Center). Note that such residues are false positives. The triangular region of the Heterogeneous Sector that was already tagged as domain-poor is enriched in predicted IDRs not clearly recognized by AlphaFold. Within the Heterogeneous Sector, this region is somehow complementary to the one highlighted by DisProt, and interestingly, the only neighborhood that both displays false positive and false negative IDR predictions is centered in the enriched 7×13 bin. We also plotted the normalized Hamming distance between the AIUPred and AlphaFold 0/1 encodings of a protein sequence, to count both false positive and false negatives, resulting in a comparable plot (data not shown).

3.5 AF2 misfolds are not caught by standard quality checks

We tested whether standard stereochemical quality checks such as PROCHECK [23] and PhiSiCal-Checkup [24] spot AF2 misfold predictions. Both algorithms flag a very small number of predictions as problematic, including physically implausible structures.

PhiSiCal-Checkup evaluates the stereochemistry of each residue by calculating a log-odd score for several joint dihedral distributions. Following the default definition, we classified as outliers residues having at least one log-odd score of -4 or less. We then calculate the average fraction of outlier residues per prediction in each bin of the arity map, and we see that it correlates with predicted disorder (Figure 4A).

PROCHECK evaluates the stereochemical quality of protein structures by means of G-factors, log-odds scoring functions that quantify how unusual certain stereochemical parameters are. A G-factor of -1 or less corresponds to a problematic structure. Like in the case of PhiSiCal-Checkup, highlighted structures concentrate in the unfolded part of the arity map, disregarding misfolds (Figure 4B).

If stereochemistry-based scoring function cannot grasp AlphaFold2 misfolds, packing-based methods could. A widely used technique in this respect is Voronoi tessellation. Voronoia is a widely used Voronoi tessellation package that also includes a quality assessment method called VoroMQA. The method returns a normalized packing score based on the deviation of the volume of the Voronoi cells centered on each residue from the distribution observed for experimentally determined, well-folded structures. A score of < 0.5 determines a poor-quality structure in terms of packing. We ran VoroMQA on the whole human proteome and counted the percentage of poor-quality packing score per arity map bin: the only region containing good-quality structures coincides with the Native Sector (Figure 4C). VoroMQA is thus unable to make the difference between AlphaFold's representation of truly disordered regions and loosely-packed misfolds.

The arity map has thus more discerning power than standard structure quality assessment measures, whether they are based on dihedral angle or packing distributions.

3.6 An arity-based, residue-wise structure quality assessment measure

We have so far characterized the regions of the arity map in order to provide a simple and efficient quality assessment evaluation for predicted protein structures. Yet, arity itself is a per-residue quantity: it is thus possible to compute a per-residue, local structural quality measure alike pLDDT. To do so we trained a support vector machine (SVM) on arity values of PDB structures, disordered structures, and implausible structures. The resulting per-residue score – on a scale from 0 (worst) to 100 (best) – is termed abstraqt, for Arity-Based STRuctural Arrangement Quality assessmenT.

With no additional information other than arity and its derivative for five values of neighborhood radii and arity percentiles – twenty two features in total – our measure distinguishes plausible and implausible structural motifs in the context of the queried protein (Figure 5). Two cases are worth commenting. The first one is that of curved alpha helices which get low abstraqt scores despite high pLDDT values (Figure 5(First row)). The second one consists of beta solenoid, whose medium abstraqt scores hint at implausible

motifs (Figure 5(Second row)). On the contrary, globular and non-globular crystallographic structures show uniform, high scores.

4 Discussion

For each `AlphaFold2` prediction, two quality assessment measures are available: the predicted local distance difference test (pLDDT), which predicts whether the distances between each residue and its neighbors are compatible with the observations, and the predicted alignment error (`PAE`), which estimates the uncertainty relative to the distance of two given residues, thus highlighting coherently predicted domains. pLDDT captures deviations at a local scale, whereas PAE evaluates the protein's global arrangement. Although these two measures have been proven robust and coherent, they still fail to catch a number of implausible predictions [17].

In this work, we introduce novel semi-global statistics characterizing packing properties at multiple scales. In particular, using two quantiles of the arity cumulative distribution function performs at once dimensionality reduction and a clustering of predicted structures. We use the resulting plot, termed the *arity map*, to study correlations with fold classification and disorder prediction. These analysis result in an unambiguous subdivision (Figure 6) which is instrumental to evaluate the quality of large sets of predictions as well as the *likely* reliability of single predictions. We also introduce a per amino acid regressor, which, based on twenty two features, refines insights provided by pLDDT values.

The arity map thus measures overall structural plausibility: the structure can be composed of plausible neighborhoods and domains but still be classified in one of the uncertain quality areas.

4.1 Code availability

The code to compute arity maps is available within the the Structural Bioinformatics Library [19]. See: AlphaFold analysis, as well as SBL portal, Documentation, Applications, Installation guide.
The code and data for rerunning analyses are made available at doi.org/10.5281/zenodo.18216693

Acknowledgments. This work has been supported by the French government, through the 3IA Côte d'Azur Investments (ANR-19-P3IA-0002), and the ANR project Innuendo (ANR-23-CE45-0019).

We wish to thank V. Mallet for insightful comments.

References

[1] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[2] Jingi Yeo, Yewon Han, Nicola Bordin, Andy M. Lau, Shaun M. Kandathil, Hyunbin Kim, Eli Levy Karin, Milot Mirdita, David T. Jones, Christine Orengo, and Martin Steinegger. Metagenomic-scale analysis of the predicted protein structure universe. *bioRxiv*, 2025.

[3] A. M. Lau, N. Bordin, S. M. Kandathil, I. Sillitoe, V. P. Waman, J. Wells, Orengo C. A., and D. T. Jones. Exploring structural diversity across the protein universe with the encyclopedia of domains. *Science*, 386(6721):eadq4946, 2024.

[4] V. Mariani, M. Biasini, A. Barbato, and T. Schwede. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013.

[5] Janani Durairaj, Mehmet Akdel, Dick de Ridder, and Aalt DJ van Dijk. Geometricus represents protein structures as shape-mers derived from moment invariants. *Bioinformatics*, 36(Supplement2):i718–i725, 2020.

[6] Mehmet Akdel et al. A structural biology community assessment of alphafold2 applications. *Nature Structural and Molecular Biology*, 29(11):1056–1067, 2022.

[7] Kathryn Tunyasuvunakool et al. Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873):590–596, August 2021.

[8] Oliviero Carugo. PLDDT values in AlphaFold2 protein models are unrelated to globular protein local flexibility. *Crystals (Basel)*, 13(11):1560, November 2023.

[9] Thomas C. Terwilliger, Dorothee Liebschner, Tristan I. Croll, Christopher J. Williams, Airlie J. McCoy, Billy K. Poon, Pavel V. Afonine, Robert D. Oeffner, Jane S. Richardson, Randy J. Read, and Paul D. Adams. Alphafold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nature Methods*, 21(1):110–116, November 2023.

[10] I. R. Alkhouri, S. Jha, A. Beckus, G. Atia, S. Jha, R. Ewetz, and A. Velasquez. Exploring the predictive capabilities of alphafold using adversarial protein sequences. *IEEE Transactions on Artificial Intelligence*, 5(07):3384–3392, 2024.

[11] F. Cazals, J. Herrmann, and E. Sarti. Simpler protein domain identification using spectral clustering. *Proteins: structure, function, and bioinformatics*, 93(7):1212–1225, 2025.

[12] Nicolas Bonneel, Michiel Van De Panne, Sylvain Paris, and Wolfgang Heidrich. Displacement interpolation using lagrangian mass transport. In *Proceedings of the 2011 SIGGRAPH Asia conference*, pages 1–12, 2011.

[13] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

[14] R. E. Tarjan. *Data Structures and Network Algorithms*, volume 44 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1983.

[15] H. Edelsbrunner and J. Harer. *Computational topology: an introduction*. AMS, 2010.

[16] F. Cazals and E. Sarti. Characterizing the fragmentation of alphafold predictions. *Submitted*, 2025.

[17] Olivia S. Pratt, Luc G. Elliott, Margaux Haon, Shahram Mesdaghi, Rebecca M. Price, Adam J. Simpkin, and Daniel J. Rigden. Alphafold2, but not alphafold3, predicts confident but unrealistic β-solenoid structures for repeat proteins. *bioRxiv*, 2024.

[18] M. do Carmo. *Differential geometry of curves and surfaces*. Prentice Hall, 1976.

[19] F. Cazals and T. Dreyfus. The Structural Bioinformatics Library: modeling in biomolecular science and beyond. *Bioinformatics*, 7(33):1–8, 2017.

[20] J. M. Lotthammer, G. M. Ginell, D. Griffith, R. J. Emenecker, and A. S. Holehouse. Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nat Methods*, 21:465–476, 2024.

[21] Federico Barone, Alessandro Laio, Marco Punta, Stefano Cozzini, Alessio Ansuini, and Alberto Cazzaniga. Unsupervised domain classification of alphafold2-predicted protein structures. *PRX Life*, 3:023009, 2025.

[22] Gábor Erdős and Zsuzsanna Dosztányi. AIUPred: combining energy estimation with deep learning for the enhanced prediction of protein disorder. *Nucleic Acids Research*, page gkae385, 2024.

[23] Roman A Laskowski, Malcolm W MacArthur, David S Moss, and Janet M Thornton. PROCHECK: a program to check the stereochemical quality of protein structures. *Applied Crystallography*, 26(2):283–291, 1993.

[24] Piyumi R Amarasinghe, Lloyd Allison, Craig J Morton, Peter J Stuckey, Maria Garcia de la Banda, Arthur M Lesk, and Arun S Konagurthu. PhiSiCal-Checkup: A bayesian framework to validate amino acid conformations within experimental protein structures. *Proceedings of the National Academy of Sciences*, 122(1):e2416301121, 2025.

[25] Alderson, T. Reid and Pritisanac, Iva and Kolaric, Desika and Moses, Alan M. and Forman-Kay, Julie D. Systematic identification of conditionally folded intrinsically disordered regions by alphafold2. *Proceedings of the National Academy of Sciences*, 120(44), October 2023.

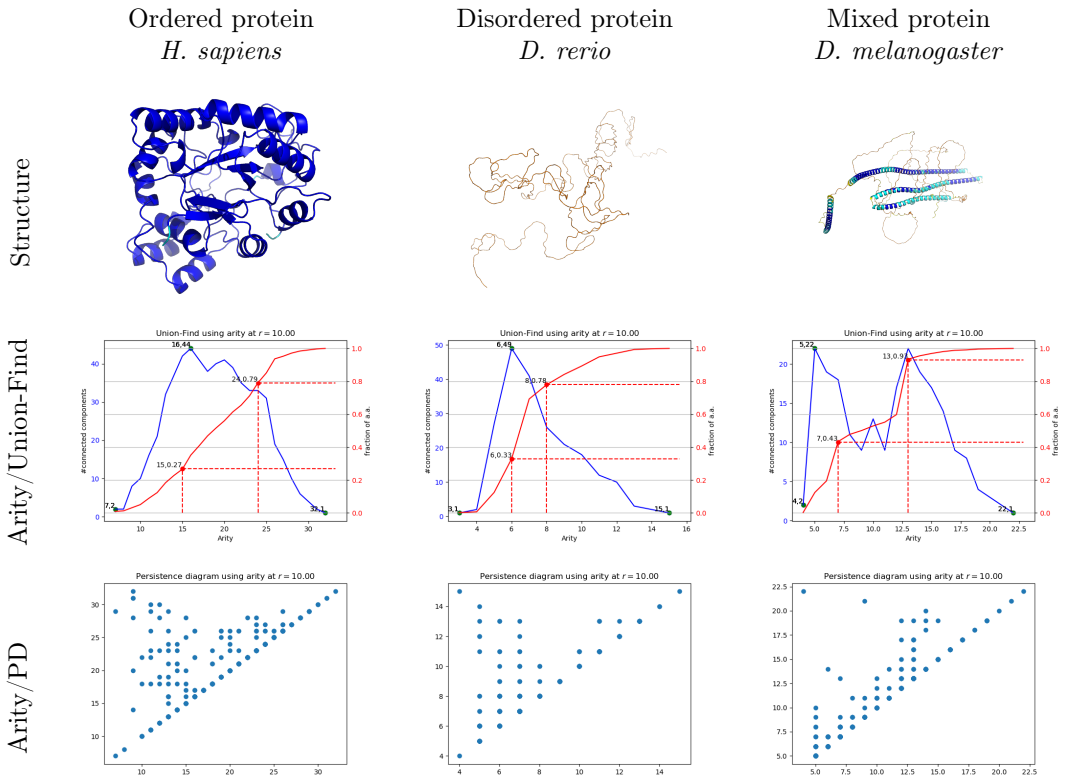


Figure 1: **Three prototypical examples corresponding to ordered, disordered and mixed proteins. Second row:** evolution of the number of connected components (blue curve) and cumulative distribution function of the arity (red curve). **Third row:** persistence diagram in the filtration defined from the arity. **(Left)** (H. sapiens; AF-P15121-F1-model_v4; 316 amino acids.) **(Middle)** (Zebrafish; AF-A0A0G2L439-F1-model_v4; 449 amino acids.) **(Right)** (D. Melanogaster; AF-Q9VQS4-F1-model_v4; 781 amino acids.)

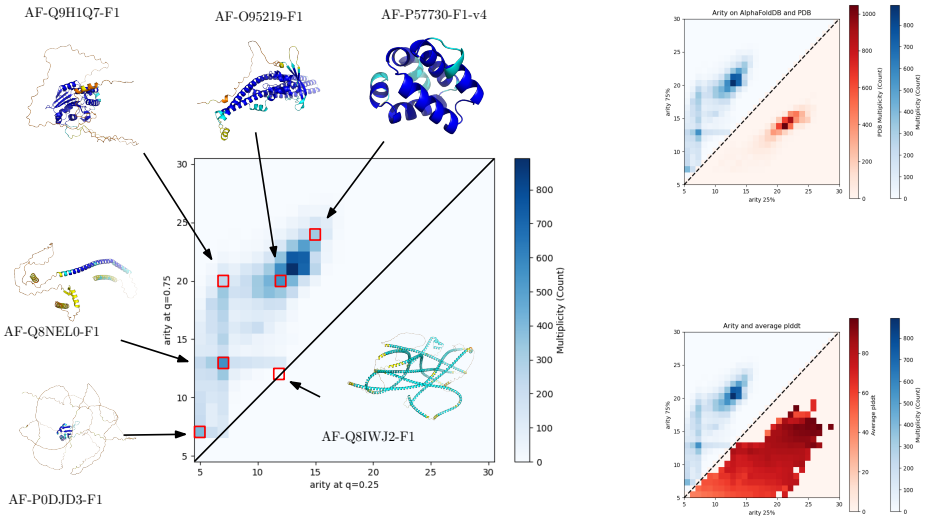


Figure 2: **H. sapiens: arity map.** The upper triangle of the arity map shows the protein count with a prescribed arity signature. **(A)** Each bin corresponds to all `AlphaFold` predictions harboring a given arity signature. **(B)** Lower triangle – mirrored with respect to the upper triangle: protein count from the PDB. Well-formed, ordered proteins only populate the (15,20) peak of the arity plot, confirming that the other two peaks visible in the arity plot of the `AlphaFold-DB` predictions must represent disordered structures or inaccurate predictions. **(C)** Lower triangle: average pLDDT per bin. Arity and pLDDT are strongly correlated: the higher the first and last quartile, the higher the averge pLDDT. Arity is thus a good estimator of structural complexity.

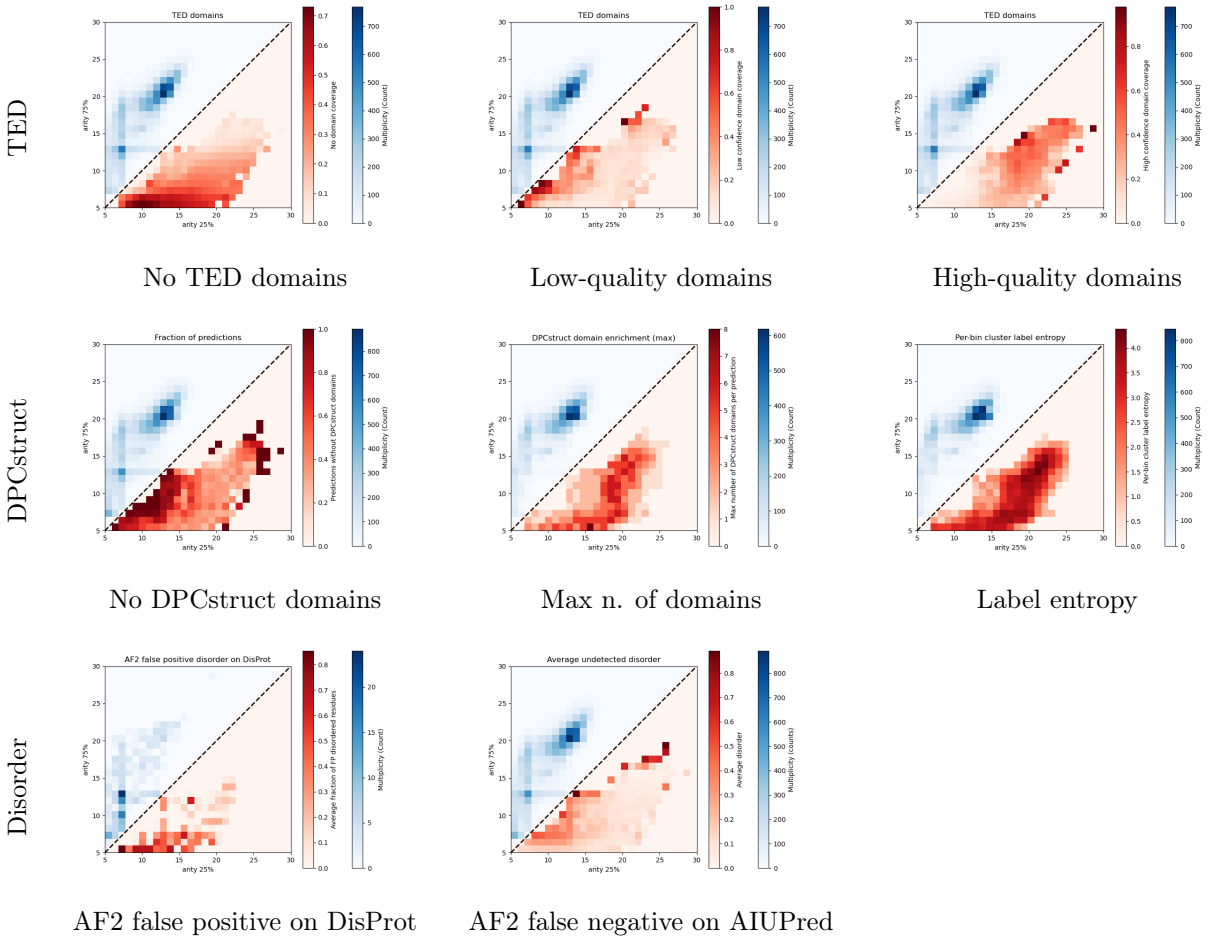
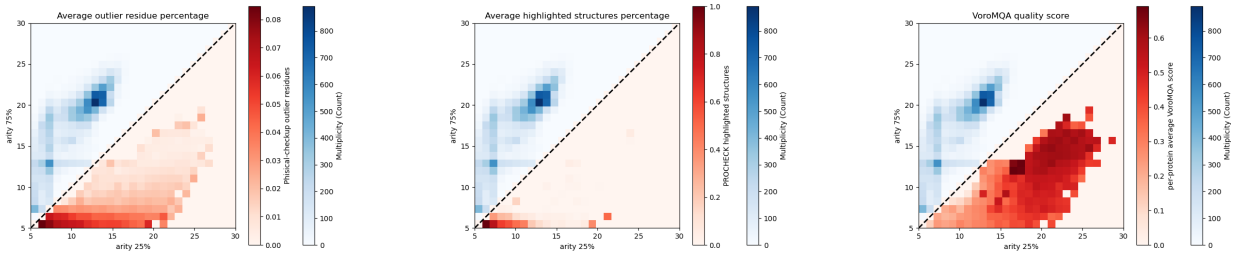


Figure 3: **Arity map characterization. First row:** density of predictions with **(Left)** no TED domains, **(Middle)** at least one high-quality TED domain, **(Right)** only low-quality TED domains. **Second row: (Left)** density of predictions with no DPCstruct domains, **(Middle)** maximum number of DPCstruct domains in a prediction, **(Right)** per-bin entropy of DPCstruct labels. **Third row:** fraction of **(Left)** residues with pLDDT < 50 not reported by `DisProt`, **(Middle)** residues with AIUPred > 50 and pLDDT > 50.



`PhiSiCal-Checkup` `PROCHECK` `VoroMQA`

Figure 4: **Comparison with structure quality assessment measures**. We plot on the arity map the results from two structure quality assessment algorithms. **(A)** For `PhiSiCal-Checkup`, we plot the average percentage of residues per predicted structure having at least one rotamer score of -4 or less. **(B)** For `PROCHECK`, we plot the percentage of structures having a global G-factor of -1 or less. **(C)** For `VoroMQA`, we plot the average VoroMQA score of each structure.

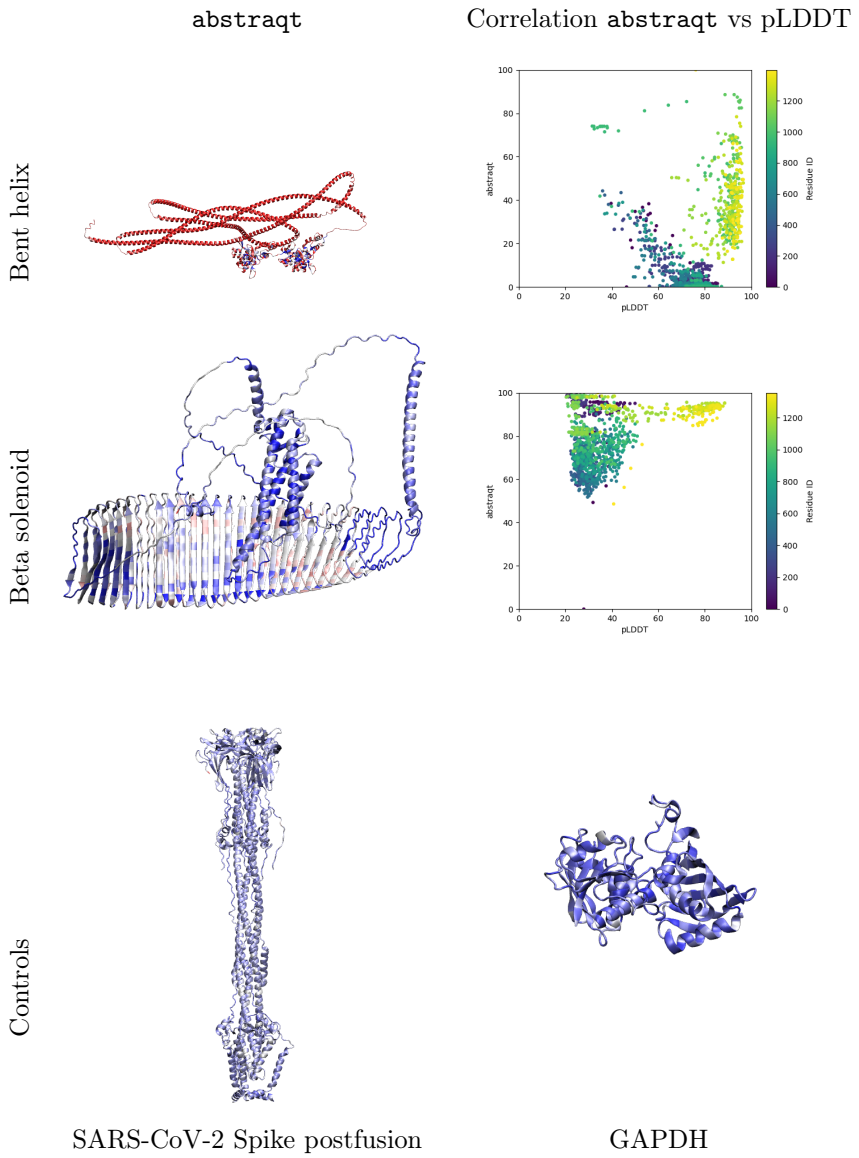


Figure 5: **The `abstraqt` per-residue score on pathological and control structures. Top and middle row, left.** Two predictions of *H. sapiens* proteins (UniProt codes P15924 and Q96Q06) displaying large implausible motifs are colored according to the `abstraqt` per-residue score. **Top and middle row, right.** For the two examples, the correlation between `abstraqt` and pLDDT is shown. Each point represents a residue in the prediction, and is colored according to its position in the polypeptide chain. **Bottom row.** Two control cases show that crystallographic globular (GAPDH, PDB 6GVE) and non-globular (postfusion Spike, PDB 8FDW) are associated with overall high and uniform `abstraqt` values.

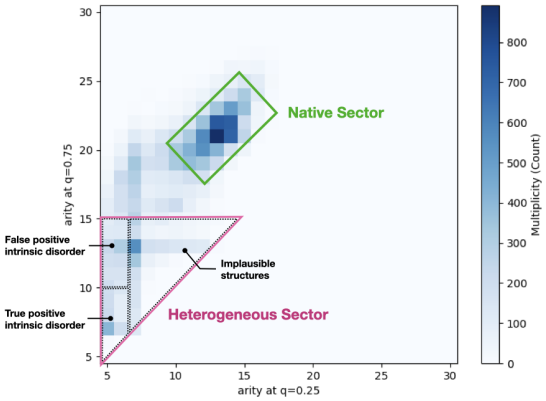


Figure 6: **Arity map enrichment areas.** The arity map can be partitioned in the Native Sector, containing most well-folded structures, and the Heterogeneous Sector, containing the majority of the remaining predictions. The analyses run in this work let us conclude the Heterogeneous Sector can be further divided into an area enriched in experimentally confirmed intrinsically disordered structures (True positive intrinsic disorder), experimentally disproved intrinsic disorder predictions (False positive intrinsic disorder), and predictions with physico-chemically implausible structures (Implausible).