

## Article

# pLDDT Values in AlphaFold2 Protein Models Are Unrelated to Globular Protein Local Flexibility

Oliviero Carugo <sup>1,2</sup> <sup>1</sup> Department of Chemistry, University of Pavia, 27100 Pavia, Italy; oliviero.carugo@univie.ac.at<sup>2</sup> Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, 1010 Wien, Austria

**Abstract:** Two non-redundant, high-quality sets of protein X-ray crystal structures from the Protein Data Bank (room temperature, 288–298 K, and low temperature, 95–105 K) were compared to structural predictions conducted using ColabFold/AlphaFold2. In particular, the relationship between B-factors and pLDDT values, which estimate the degree of prediction confidence, was investigated. It was observed that there is basically no correlation between these two quantities and, consequently, that the level of confidence in predictions does not provide information about the degree of local structural flexibility of globular proteins.

**Keywords:** AlphaFold; B-factor; ColabFold; pLDDT; protein data bank

## 1. Introduction

Structural biology has always been on a relentless pursuit to unravel the intricate details of molecular architecture and its implications for function [1]. For decades, traditional experimental techniques like X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy have paved the way for groundbreaking discoveries. However, despite their invaluable contributions, these techniques come with inherent limitations, often centered around resolution constraints, laborious sample preparation, and time-intensive analysis. This is changing thanks to artificial intelligence (AI)—a transformative force that is reshaping many scientific disciplines, including structural biology.

AlphaFold2, developed by DeepMind, stands as a paragon in this AI-driven revolution, showcasing unprecedented accuracy in predicting protein structures [2]. Its success in the 14th Critical Assessment of Structure Prediction (CASP14) elucidated the potential of machine learning models in decoding the protein folding problem, a challenge that had remained enigmatic for nearly half a century. The beauty of AlphaFold2 lies not just in its predictive precision but also in its capability to offer estimations of prediction confidence that limit the risk of overinterpretation.

Yet, the accessibility to such high-level AI models remained a barrier for many researchers, a gap that ColabFold sought to bridge [3], making predictions considerably faster and available to a broader spectrum of scientists irrespective of their computational expertise.

Evolutionary Scale Modeling (ESM) provides an additional approach, leveraging evolutionary data to create embeddings that capture protein sequence relationships and their structural implications through a large language model [4]. This melding of evolutionary biology with deep learning showcases the potential synergy of interdisciplinary integration in advancing our understanding of biomolecular systems.

Nevertheless, it is crucial to exercise prudence and avoid undue excitement [5]. Indeed, AI has the potential to generate unforeseen advancements that are now beyond our collective imagination. Consequently, it becomes imperative to undertake a rigorous assessment of AI-driven computational models to ensure their reliability and accuracy [6].



**Citation:** Carugo, O. pLDDT Values in AlphaFold2 Protein Models Are Unrelated to Globular Protein Local Flexibility. *Crystals* **2023**, *13*, 1560. <https://doi.org/10.3390/cryst13111560>

Academic Editor: Abel Moreno

Received: 11 October 2023

Revised: 24 October 2023

Accepted: 25 October 2023

Published: 1 November 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Several articles have recently emerged that have started to examine this particular issue. Numerous critical aspects of the AlphaFold2 models have been examined by Akdel and colleagues [7]. It was remarked that only half of the chalcogen bonds observed in high-resolution crystal structures are discernible in the corresponding models deposited in the AlphaFold protein structure database [8]. Additionally, Holcomb and colleagues have explored the potential application of these models in computational docking [9]. Buel and Walters have observed that AlphaFold2 appears to be unsuitable for predicting the structure of mutants [10]. Recently, Podvalnaya and colleagues were able to phase a protein crystal structure with an AlphaFold2 model [11]. Stevens and He examined the reliability of loops' structure predictions [12]. The performance of AlphaFold2 in predicting VHH structures has been compared to those of several other methods [13]. It is plausible to anticipate that a multitude of further studies of this kind will be undertaken and disseminated in future research.

This communication is centered on protein flexibility and on the information that ColabFold/AlphaFold2 models can provide on it.

All molecules, including proteins, are flexible, which has an obvious impact on both their thermodynamic stability and on their reactivity [14–20]. In crystal and cryo-EM structures, the local flexibility of globular proteins is monitored through the B-factors [21,22]. Regions characterized by high B-factors are considered to be more flexible and less rigid.

ColabFold/AlphaFold2 does not predict B-factors but computes a score for each residue named pLDDT, which indicates the estimated level of confidence of the predicted structure [2]. It is the predicted value of the Local Distance Difference Test (lDDT), which is the score that evaluates local distance differences of all atoms in a model [23]. Being based on internal coordinates—it compares the inter-atomic distances of one structure to the same distances in the other structure—it does not require the superposition of the two structures that are compared and it is, consequently, a robust tool for the assessment of structure predictions without manual intervention. pLDDT values range from 0 to 100 and  $pLDDT \geq 90$  indicates residues predicted with extremely high confidence, whereas residues with  $90 > pLDDT \geq 70$  are classified as confident. Residues with  $70 > pLDDT \geq 50$  are considered to be predicted with low confidence, and those whose pLDDT is less than 50 are predicted with extremely low confidence, which could mean that the prediction is impossible [2,24].

A correlation between pLDDT and conformational disorder has been recently observed. Ordered protein regions have, in general, pLDDT values larger than 80 while low pLDDT values are often observed in intrinsically disordered regions—pLDDT values below 50 are present in approximately half of them [25]. This is not surprising, since AlphaFold2 was designed to reproduce experimental observations. These AI-based methods, such as the well-known ChatGPT, reproduce the information encapsulated in the enormous amounts of data that were utilized in their training. The inherent scope of these methods allows for greater confidence in predicting the positions of well-ordered atoms/residues compared to disordered atoms/residues, for which structural information is limited.

Atoms that oscillate little around a unique equilibrium position exhibit a sharp and distinct electron density and their positions can be determined with high accuracy and small B-factors. On the contrary, atoms that oscillate with high amplitude around two or more equilibrium positions exhibit a blurry electron density and their position can be determined with lower accuracy and large B-factors. B-factors have actually been used, despite not being portable amongst structures [21], to estimate the positional standard errors of individual atoms in protein crystal structures [26,27]. It can be expected that the positions of well-ordered atoms are more reproducible than those of atoms that are more flexible. Consequently, one can expect that AlphaFold2 can predict the position of well-ordered atoms better than the positions of less-ordered atoms, and this should be reflected in greater pLDDT values for well-ordered atoms and in lower pLDDT values for less-ordered atoms. In other words, one may postulate that diminished levels of confidence may be associated with local molecular flexibility.

In order to verify the hypothesis that pLDDT values relate to local conformational flexibility, in this study, the B-factor and the pLDDT values were systematically compared. Two distinct sets of protein crystal structures were compiled using the data archived in the Protein Data Bank (PDB) [28–30]. These sets consist of high-quality structures that are both non-redundant. One set comprises structures obtained at ambient temperature within the range of 288–298 K, while the second set encompasses structures obtained at low temperature within the range of 95–105 K. Both the B-factors in their original form as deposited in the Protein Data Bank (PDB) and the B-factors normalized to have a zero mean and unit variance were used. Nevertheless, it was observed that the B-factors exhibited no correlation with the pLDDT value. This suggests that the pLDDT values do not convey any substantive physical information, but rather serve only their intended purpose of estimating the confidence in internal predictions.

Obviously, this does not imply that models made with ColabFold/AlphaFold2 are intrinsically unreliable. It simply means that the prediction confidence suggested by the pLDDT values is unrelated to local conformational flexibility.

## 2. Methods

All the X-ray crystal structures were taken from the Protein Data Bank—structures determined using other techniques were discarded [28–30]. They were divided into two groups: one, named the “room set”, containing structures determined at room temperature in the 288–298 K range; and the other, named “cryo set”, containing the structures determined at low temperature in the 95–105 K range. Each group was then processed as follows.

Structures refined at a resolution lower than 2 Å were discarded as well as the multi-model of C $\alpha$ -only refinements. Since individual atomic B-factors must be analyzed, TLS-refined structures and structures refined using non-crystallographic symmetry restraints were discarded. Since experimental protein structures must be compared to computational models, structures containing nucleic acids were discarded as well as structures containing too many (more than 5%) heteroatoms different to water and structures containing missing residues—those that are undetectable in the electron density maps.

Redundancy was then reduced at the sequence level by using CD-HIT (maximal pairwise sequence identity 40%) [31]. Only proteins of up to 150 residues were retained.

All these data mining procedures resulted in 22 structures in the room set and 308 structures in the cryo set (see supplementary material Table S1).

Since it is well known that B-factors reflect numerous structural features [21], ranging from atomic oscillation around the equilibrium position to conformational disorder, and since they depend on the crystallographic resolution too, it is necessary to normalize them when different crystal structures are compared. Here, the normalized BN-factors were computed as

$$BN_i = \frac{B_i - Bave}{Bstd} \quad (1)$$

where *Bave* and *Bstd* are the average value of the B-factors of the *n* protein atoms and the standard deviation of their distribution, respectively.

$$Bave = \sum_{i=1}^n B_i \quad (2)$$

$$Bstd = \sqrt{\frac{(B_i - Bave)^2}{n - 1}} \quad (3)$$

Thanks to this normalization, the BN-factors have zero mean and unit variance in each protein crystal structure.

The construction of all computational models was accomplished via ColabFold [3], employing AlphaFold2 [2] for the purpose of structure prediction, employing a total of three prediction cycles. ColabFold exhibits enhanced computational efficiency compared to

the original version of AlphaFold2 due to the use of a faster homology search methodology (MMseqs2) [32]. In all other aspects, ColabFold remains comparable to AlphaFold2. The use of templates sourced from the Protein Data Bank was permitted, followed by the application of Amber energy minimization. The model that achieved the highest ranking was selected for the purpose of comparing it with the experimental structure.

An invaluable feature of AlphaFold2 is the residue-based estimation of the prediction confidence. This is the pLDDT, which ranges from 0 to 100—though I have never seen this extreme value in reality. According to the AlphaFold2 manual, when the pLDDT value exceeds 90, it indicates that the predictions for both main-chain and side-chain atoms are regarded reliable. On the other hand, if the pLDDT value falls between 70 and 90, there is a higher likelihood of mistakes occurring essentially in the side-chain atoms. If the value of pLDDT falls in the range of 50 to 70, mistakes may occur in the main-chain atoms, and if the value of pLDDT is less than 50, the predictions might be considered very untrustworthy, indicating that making accurate predictions becomes unfeasible.

Secondary structures were assigned using Stride [33] and DSSP [34] and, since the results were substantially independent of the assignment procedure, those only produced using Stride are discussed.

Atomic solvent-accessible surface areas, measured in squared Å employing NACCESS [35], were used to monitor the level of solvent accessibility (*SASA<sub>atom</sub>*) of each individual atom. Using the same tools, the relative solvent-accessible surface areas were calculated for residues when all of their atoms are taken into account (*rel\_SASA<sub>residue</sub>*). It is a dimensionless quantity, it spans from 0 to around 100, and it is defined as the ratio (multiplied by 100) between the solvent-accessible surface area of residue X and the solvent-accessible surface area of the same residue in a G-X-G three-peptide in the extended conformation.

Crystal packing contacts were identified as previously described [36] with an inter-atomic distance threshold of 4.5 Å.

### 3. Results and Discussion

Some correlation between pLDDT and conformational disorder has been recently observed, since ordered protein regions have, in general, pLDDT values larger than 80 while low pLDDT values are often observed in intrinsically disordered regions [25]. Moreover, long segments of residues with poor confidence in the predicted structure are thought to be conformationally disordered [7]. This is of crucial importance because structural flexibility is essential to the thermodynamic stability and reactivity of proteins [14–20].

The pLDDT parameter, which has a range of 0 to 100, is used to estimate prediction confidence [2]. When the value is less than 50, it indicates that the forecast is not achievable, or that it is totally unreliable. Structural moieties with values in the 50–70 range should be carefully examined since they may significantly differ from the experimental structure. Particularly for the main-chain atoms, values in the 70–90 range suggest that the structure is probably comparable to an experimental structure. Values greater than 90 eventually suggest that the structure is probably almost exactly the same as the experimental structure.

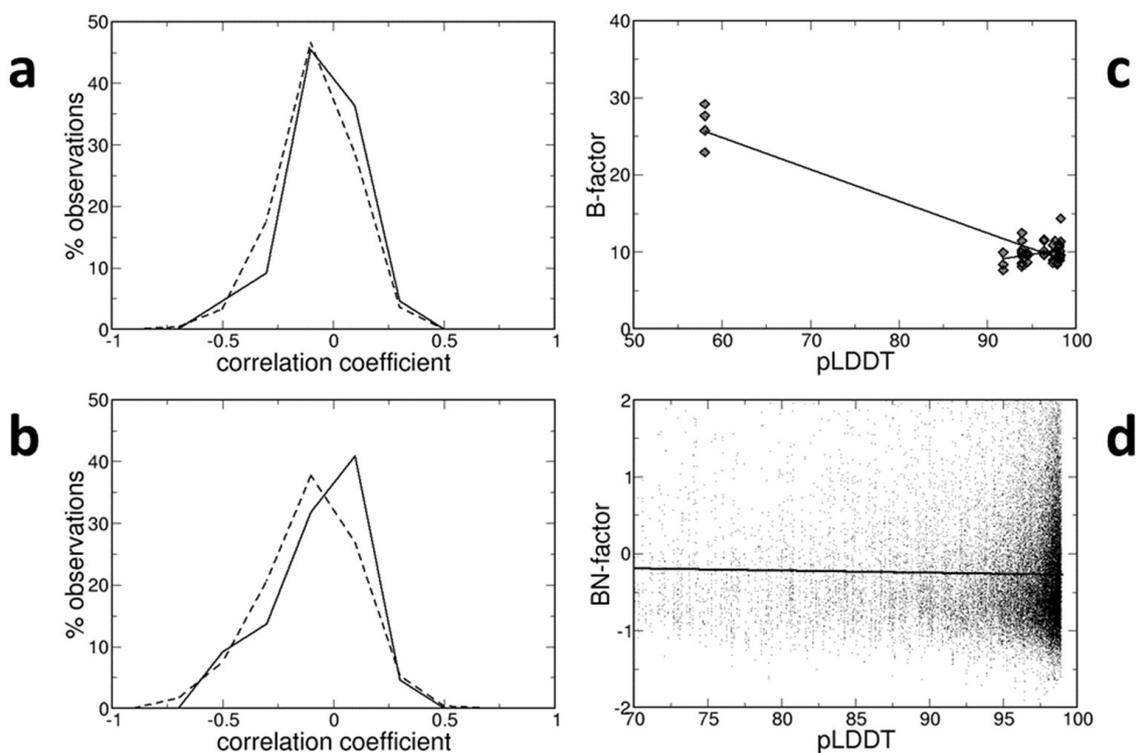
Since B-factors in crystal and cryo-EM structures monitor protein flexibility [21,22], it is important to calculate the degree of correlation between B-factor and pLDDT values to confirm whether computational models offer a straightforward way to gain knowledge about local flexibility in globular proteins.

In order to achieve this goal, a high-quality set of non-redundant protein must be examined. However, as the temperature at which experiments are conducted has significant effects on flexibility, two non-redundant sets of protein structures were built: one at room temperature (288–298 K) and the other at a low temperature (95–105 K). In both cases, only proteins up to 150 residues were retained. The latter is larger because, in modern times, diffraction measurements are usually conducted at 100 K. However, no particular discrepancies were found between the two sets of data, as can be seen below.

Furthermore, B-factors were normalized to have mean values equal to zero and a standard deviation equal to one because they are not very transferable from structure to structure and may depend on a variety of parameters, such as the crystallographic resolution [21,37]. However, no particular differences were found, by using raw B-factors or normalized B-factors, as can be seen below.

Pearson correlation coefficients were used to quantify the degree of correlation between pLDDT and B-factors. I start by quickly going over the examination of the relationship between the B-factor and pLDDT for every protein atom in each structure. Subsequently, I focus only on atoms for which the structure is highly confidently predicted (pLDDT > 70). Then, in order to rule out the impacts of data aggregation, I concentrate on 48 distinct types of atoms. Eventually, normalized B-factors are used for the analyses where all atoms of all proteins can be mixed into a unique large set of data.

Pearson correlation coefficients between B-factors and pLDDT values were computed for all atoms in each structure. Their distributions are shown in Figure 1a for both room-temperature and low-temperature structures. They are very similar and symmetric around 0. For room-temperature structures, they range from  $-0.431$  to  $+0.242$  with a mean value of  $-0.056$  ( $\pm 0.036$ ). For low-temperature structures, they range from  $-0.612$  to  $+0.364$  with a mean value of  $-0.067$  ( $\pm 0.009$ ). In both cases, the average correlation coefficient is negative, indicating that the smaller B-factors tend to be observed at increasing pLDDT values. However, the correlation is extremely weak.



**Figure 1.** (a) Distribution of the correlation coefficients between B-factors and pLDDT values for all atoms in structures determined at room temperature (continuous line) or at low temperature (broken line). (b) Like (a), but only for the  $C\alpha$  atoms, for which a high degree of confidence in their position is expected (pLDDT > 70). (c) Example of high correlation coefficient between B-factors and pLDDT values; regression lines for all points (long line) or only those with pLDDT > 70 (short line) show the relationship between B-factors and pLDDT values for the main-chain atoms of structure 2py0 (chain A); only the following atoms were considered: they belong to loop residues with a *rel\_SASA\_residue* > 5, their *SASA\_atom* > 2 Å, and they are close (< 4.5 Å) to at least one atom of another chain of the crystal structure. (d) Relationship between the normalized B-factors and the pLDDT values of  $C\alpha$  atoms, for which a high degree of confidence in their position is expected (pLDDT > 70).

The same analysis was conducted with a smaller group of atoms. Just one atom, the C $\alpha$  atom, was kept for each residue since the pLDDTs are linked to residues rather than atoms and because residues include a variety of atom types, some of which are even quite distinct from one another. In addition, atoms whose position was predicted with low confidence (pLDDT < 70) were discarded, since it is reasonable to assume that B-factors cannot monitor positional flexibility for completely incorrect positions—this would be experimentally observed in regions where there is no electron density.

The distributions of correlation coefficients, shown in Figure 1b, are not very different from those of Figure 1a. They are quite symmetric around 0. For room-temperature structures, they range from  $-0.493$  to  $+0.286$  with a mean value of  $-0.069$  ( $\pm 0.044$ ). For low-temperature structures, they range from  $-0.685$  to  $+0.401$  with a mean value of  $-0.105$  ( $\pm 0.012$ ). Therefore, both mean values are negative, indicating that B-factors tend to be smaller at larger pLDDTs, but their absolute values are very small and close to 0, indicating that such a correlation, if it exists, is very weak.

The correlation between B-factors and pLDDT values has been examined individually for 48 atom types, according to the following categorizations:

- Main- or side-chains atoms (two categories).
- Secondary structure of the residue to which it belongs (three categories: helix, strand, loop).
- Solvent accessibility of the residue to which it belongs (two categories: accessible if  $rel\_SASA\_residue > 5$  and inaccessible if relative  $rel\_SASA\_residue \leq 5$ ).
- Atomic solvent accessibility (two categories: accessible if atomic  $SASA\_atom > 2 \text{ \AA}^2$  and inaccessible if  $SASA\_atom \leq 2 \text{ \AA}^2$ ).
- Presence of interatomic contacts with other chains in the crystal structure, either within the same asymmetric unit or crystal packing contacts (two categories: presence or absence of inter-chain contacts).

Usually, the correlation coefficients are small and close to 0. Occasionally, they are surprisingly large. For example, in the PDB entry 2py0 (chain A), the correlation coefficient is equal to  $-0.925$ , very close to the minimal possible value, if the attention is focused on the main-chain and solvent-accessible atoms involved in inter-chain contacts and belonging to solvent-accessible residues located in loops. However, this is essentially due (see Figure 1c) to the presence of a small group of atoms with pLDDT  $\approx 60$  and B-factor  $\approx 25 \text{ \AA}^2$  and another set of atoms grouped at larger pLDDT values ( $\approx 95$ ) and smaller B-factors ( $\approx 10 \text{ \AA}^2$ ). The regression line ( $B = 49.51 - 0.41 \text{ pLDDT}$ ), therefore, simply links the two groups of atoms and does not indicate a genuine relationship between the two variables. If the analysis is limited to the atoms of the second, larger group, the correlation coefficient is much closer to 0 (0.292) and the regression line even has a slope with different sign ( $B = -7.10 + 0.18 \text{ pLDDT}$ ).

Due to the effect of resolution and other variables, B-factors are not portable and must be normalized for quantitative comparisons between different crystal structures [20]. The normalized BN-factors (see Methods) were computed for all C $\alpha$  atoms, the positions of which were predicted with high confidence (pLDDT > 70). There is no correlation between BN-factors and pLDDT values (see Figure 1d). Despite the fact that the correlation coefficient is negative ( $-0.020$ ), indicating that BN-factors decrease as pLDDT values rise, the regression line ( $B = 0.010 - 0.003 \text{ pLDDT}$ ) is essentially flat and horizontal. Note that only the information retrieved from the structures determined at low temperatures is shown in Figure 1d, though room-temperature structures provide the same findings.

#### 4. Conclusions

Understanding protein stability and reactivity requires knowledge of globular protein flexibility [14–20]. Since residues predicted with low confidence may be more flexible than residues predicted with high confidence, it was recently suggested that AI-based protein structure predictions made using AlphaFold2/ColabFold might provide information on protein flexibility [25]. This might be possible thanks to the pLDDT values, which were designed to estimate the degree of confidence of the predictions [2,3]. In fact, pLDDT

values larger than 80 are usually associated with ordered protein regions, while intrinsically disordered regions have low pLDDT values, often lower than 50 [25]. For this reason, the pLDDT values can be compared to the B-factors, which reflect local flexibility [21,22]. Unfortunately, the comparison detailed in the present scientific communication revealed no relationship between pLDDT and the B-factor. This is not to say that AI-based models are unreliable; rather, it just means that, in contrast to experimental structures, they are unable to provide direct information on local flexibility for time. This limitation does not lessen their value in structural biology, but it should be considered as a future challenge for AI-based structural bioinformatics. Nevertheless, it is plausible to assume that they may provide some insights into dynamics by allowing structure-based predictions of B-factor values [38–41].

This does not mean that artificial intelligence cannot be involved in the study of protein flexibility and dynamics. Machine learning, in particular, has been used to construct force fields based on the analysis of large amounts of data obtained through traditional and expensive molecular dynamics methods [42]. For example, in recent times, several research groups have collaborated in order to develop coarse-grained molecular potentials that are founded upon statistical mechanics by using artificial neural networks [43]. Based on a series of molecular dynamics simulations of approximately 9 ms for twelve proteins and polypeptides with different folds, a potential was constructed that allows for coarse-grained simulations, which are much faster than traditional methods based on examining all atoms. Even if the results of this study are still of modest practical use, it is reasonable to think that this is the way in which artificial intelligence can make a revolutionary contribution to the stage of protein flexibility.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/cryst13111560/s1>. Table S1. List of the protein crystal structures examined in the present study. They are identified by the four-letter Protein Data Bank identification code, followed by the chain identifier.

**Funding:** The author acknowledges support from the Ministero dell'Università e della Ricerca (MUR) and the University of Pavia through the program "Dipartimenti di Eccellenza 2023–2027".

**Data Availability Statement:** All data are available in public databases.

**Acknowledgments:** A. Cohen and C. Bartoli are gratefully acknowledged for helpful discussions and constant support. Open Access Funding by the University of Vienna is gratefully acknowledged.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Carugo, O.; Djinović-Carugo, K. Structural biology: A golden era. *PLoS Biol.* **2023**, *21*, e3002187. [[CrossRef](#)]
2. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 584–589.
3. Mirdita, M.; Schütze, K.; Moriawaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making protein folding accessible to all. *Nat. Met.* **2022**, *19*, 679–682.
4. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379*, 1123–1130. [[CrossRef](#)]
5. Read, R.J.; Baker, E.N.; Bond, C.S.; Garman, E.F.; van Raaij, M.J. AlphaFold and the future of structural biology. *Acta Crystallogr. Sect. D Struct. Biol.* **2023**, *79*, 556–558. [[CrossRef](#)]
6. Laurents, D.V. AlphaFold 2 and NMR Spectroscopy: Partners to Understand Protein Structure, Dynamics and Function. *Front. Mol. Biosci.* **2022**, *9*, 906437. [[CrossRef](#)]
7. Akdel, M.; Pires, D.E.V.; Pardo, E.P.; Jänes, J.; Zalevsky, A.O.; Mészáros, B.; Bryant, P.; Good, L.L.; Laskowski, R.A.; Pozzati, G.; et al. A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.* **2022**, *29*, 1056–1067.
8. Carugo, O.; Djinović-Carugo, K. Automated identification of chalcogen bonds in AlphaFold protein structure database files: Is it possible? *Front. Mol. Biosci.* **2023**, *10*, 1155629.
9. Holcomb, M.; Chang, Y.-T.; Goodsell, D.S.; Forli, S. Evaluation of AlphaFold2 structures as docking targets. *Protein Sci.* **2023**, *32*, e4530.
10. Buel, G.R.; Walters, K.J. Can AlphaFold2 predict the impact of missense mutations on structure? *Nat. Struct. Mol. Biol.* **2022**, *29*, 1–2.

11. Podvalnaya, N.; Bronkhorst, A.W.; Lichtenberger, R.; Hellmann, S.; Nischwitz, E.; Falk, T.; Karaulanov, E.; Butter, F.; Falk, S.; Ketting, R.F. piRNA processing by a trimeric Schlafen-domain nuclease. *Nature* **2023**, *622*, 402–409. [CrossRef]
12. Stevens, A.O.; He, Y. Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction. *Biomolecules* **2022**, *12*, 985. [CrossRef]
13. Nadaradjane, A.A.; Diharce, J.; Rebehmed, J.; Cadet, F.; Gardebien, F.; Gelly, J.-C.; Etchebest, C.; de Brevern, A.G. Quality assessment of V(H)H models. *J. Biomol. Struct. Dyn.* **2023**, 1–15. [CrossRef]
14. Teilum, K.; Olsen, J.G.; Kragelund, B.B. Functional aspects of protein flexibility. *Cell Mol. Life Sci.* **2009**, *66*, 2231–2247.
15. Nguyen, D.D.; Xia, K.; Wei, G.W. Generalized flexibility-rigidity index. *J. Chem. Phys.* **2016**, *144*, 234106.
16. Forrey, C.; Douglas, J.F.; Gilson, M.K. The fundamental role of flexibility on the strength of molecular binding. *Soft Matter*. **2012**, *8*, 6385–6392.
17. Scaramozzino, D.; Khade, P.M.; Jernigan, R.L.; Lacidogna, G.; Carpinteri, A. Structural compliance: A new metric for protein flexibility. *Proteins* **2020**, *88*, 1482–1492.
18. Carugo, O.; Argos, P. Correlation between side chain mobility and conformation in protein structures. *Protein Eng.* **1997**, *10*, 777–787. [CrossRef]
19. Vihinen, M. Relationship of protein flexibility to thermostability. *Protein Eng. Des. Sel.* **1987**, *1*, 477–480. [CrossRef]
20. Vander Meersche, Y.; Cretin, G.; de Brevern, A.G.; Gelly, J.C.; Galochkina, T. MEDUSA: Prediction of Protein Flexibility from Sequence. *J. Mol. Biol.* **2021**, *433*, 166882.
21. Carugo, O. Uses and abuses of the atomic displacement parameters in structural biology. *Meth. Mol. Biol.* **2022**, *2449*, 281–298.
22. Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M.T. Utility of B-factors in protein science: Interpreting rigidity, flexibility, and internal motion and engineering. *Chem. Rev.* **2019**, *119*, 1626–1665. [CrossRef]
23. Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. IDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **2013**, *29*, 2722–2728. [CrossRef]
24. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucl. Acids Res.* **2022**, *50*, D439–D444. [CrossRef]
25. Binder, J.L.; Berendzen, J.; Stevens, A.O.; He, Y.; Wang, J.; Dokholyan, N.V.; Oprea, T.I. AlphaFold illuminates half of the dark human proteins. *Curr. Opin. Struct. Biol.* **2022**, *74*, 102372. [CrossRef]
26. Gurusaran, M.; Shankar, M.; Nagarajan, R.; Helliwell, J.R.; Sekar, K. Do we see what we should see? Describing non-covalent interactions in protein structures including precision. *IUCr* **2014**, *1*, 74–81. [CrossRef] [PubMed]
27. Dinesh Kumar, K.S.; Gurusaran, M.; Satheesh, S.N.; Radha, P.; Pavithra, S.; Thulaa Tharshan, K.P.S.; Helliwell, J.R.; Sekar, K. Online\_DPI: A web server to calculate the diffraction precision index for a protein structure. *J. Appl. Cryst.* **2015**, *48*, 939–942. [CrossRef]
28. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Meyer, E.F.J.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542. [CrossRef] [PubMed]
29. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]
30. wwPDB Consortium. Protein Data Bank: The single global archive for 3D macromolecular structural data. *Nucleic Acids Res.* **2019**, *47*, D520–D528. [CrossRef] [PubMed]
31. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef] [PubMed]
32. Mirdita, M.; Steinegger, M.; Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **2019**, *35*, 2856–2858. [CrossRef] [PubMed]
33. Heinig, M.; Frishman, D. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **2004**, *32*, w500–w502. [CrossRef] [PubMed]
34. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637. [CrossRef] [PubMed]
35. Hubbard, S.J.; Thornton, J.M. NACCESS, Department of Biochemistry and Molecular Biology, University College London. 1993. Available online: <http://www.bioinf.manchester.ac.uk/naccess/nacdownload.html> (accessed on 10 October 2023).
36. Carugo, O.; Djinovic Carugo, K. Packing bridges in protein crystal structures. *J. Appl. Cryst.* **2014**, *47*, 458–461. [CrossRef]
37. Carugo, O. Atomic displacement parameters in structural biology. *Amino Acids* **2018**, *50*, 775–786. [CrossRef]
38. Bramer, D.; Wei, G. Blind prediction of protein B-factor and flexibility. *J. Chem. Phys.* **2018**, *149*, 135107. [CrossRef]
39. Yang, J.; Wang, Y.; Zhang, Y. ResQ: An Approach to Unified Estimation of B-Factor and Residue-Specific Error in Protein Structure Prediction. *J. Mol. Biol.* **2016**, *428*, 693–701. [CrossRef]
40. Pan, X.Y.; Shen, H.B. Robust prediction of B-factor profile from sequence using two-stage SVR based on random forest feature selection. *Protein Pept. Lett.* **2009**, *16*, 1447–1454. [CrossRef]
41. Weiss, M.S. On the interrelationship between atomic displacement parameters (ADPs) and coordinates in protein structures. *Acta Crystallogr.* **2007**, *D63*, 1235–1242. [CrossRef]

42. Husic, B.E.; Charron, N.E.; Lemm, D.; Wang, J.; Pérez, A.; Majewski, M.; Krämer, A.; Chen, Y.; Olsson, S.; de Fabritiis, G.; et al. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* **2020**, *153*, 194101. [[CrossRef](#)] [[PubMed](#)]
43. Majewski, M.; Pérez, A.; Thölke, P.; Doerr, S.; Charron, N.E.; Giorgino, T.; Husic, B.E.; Clementi, C.; Noé, F.; De Fabritiis, G. Machine learning coarse-grained potentials of protein thermodynamics. *Nat. Commun.* **2023**, *14*, 5739. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.