Applicability of AlphaFold2 in the modelling of coiled-coil domains

Rafal Madaj[a,§], Mikel Martinez-Goikoetxea[b,§], Kamil Kaminski[a], Jan Ludwiczak[a,†], and Stanislaw Dunin-Horkawicz[a,b,*]

[a] Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Zwirki i Wigury 101, 02-089, Warsaw, Poland

[b] Department of Protein Evolution, Max Planck Institute for Biology Tübingen, Max-Planck-Ring 5, 72076 Tübingen, Germany

[†] Present address: Prescient Design, Genentech Research & Early Development, Roche Group, Grenzacherstrasse 124, 4070 Basel, Switzerland.

[§] These two authors contributed equally to this work

[*] To whom correspondence should be addressed

E-mail: stanislaw.dunin-horkawicz@tuebingen.mpg.de

Abstract

Coiled coils are a common protein structural motif involved in cellular functions ranging from mediating protein-protein interactions to facilitating processes such as signal transduction or regulation of gene expression. They are formed by two or more alpha helices that wind around a central axis to form a buried hydrophobic core. Various forms of coiled-coil bundles have been reported, each characterized by the number, orientation, and degree of winding of the constituent helices. This variability is underpinned by short sequence repeats that form coiled coils and whose properties determine both their overall topology and the local geometry of the hydrophobic core. The strikingly repetitive sequence has enabled the development of accurate sequence-based coiled-coil prediction methods; however, the modelling of coiled-coil domains remains a challenging task. In this work, we present the outstanding accuracy of AlphaFold2 in modeling coiled-coil domains, both in modeling local geometry and in predicting global topological properties. Furthermore, we show that the prediction of the oligomeric state of coiled-coil bundles can be improved by using the internal representations of AlphaFold2, with a performance better than any previous state-of-the-art method (code available at https://github.com/labstructbioinf/dc2_oligo).

Introduction

Coiled coils are protein structural motifs consisting of two or more alpha helices, oriented parallel or antiparallel, that wind around a central axis to form rod-like bundles (Lupas *et al.*, 2017). This winding, also known as supercoiling, is caused by periodic interlocking of side chains localized in the hydrophobic core of the bundle. The basis of this interlocking, an interaction known as knobs-into-holes, is considered the hallmark of coiled coils. It places the side chain of a core residue (knob) of one helix into a cavity (hole) formed by residues of the opposite helix. At the sequence level, coiled-coil domains are formed by 7-residue repeats (heptad) in which residues are labeled from *a* through *g*. Residues at two positions, *a* and *d*, are typically hydrophobic and face the bundle axis, forming the hydrophobic core. Since the average spacing between hydrophobic core residues in a heptad is 3.5 (7/2) and the average periodicity of an undistorted alpha helix is 3.63 residues per turn, left-handed supercoiling is required to effectively reduce the helical periodicity to 3.5 and allow the formation of knobs-into-holes interactions along the helices (Figure 1A). Importantly, this effective reduction in periodicity occurs with respect to the bundle axis and does not affect the actual helical periodicity (Szczepaniak *et al.*, 2021).

Although heptads represent the vast majority of known coiled-coil domains, deviations from this canonical periodicity are possible and have been observed, very often as local non-canonical repeats interspersed between arrays of heptads, but also as global non-heptad coiled-coil domains (Martinez-Goikoetxea and Lupas, 2023). For example, one of the most common non-heptad repeats is the hendecad (Figure 1A), which is 11 residues long and contributes three residues to the hydrophobic core. Since the average spacing between core residues (11/3=3.67) is very close to the periodicity of an alpha helix, coiled-coil bundles based on hendecad repeats show almost no supercoiling. Unlike hendecads, pentadecads are 15 residues long, contribute four residues to the core (15/4=3.75), and cause right-handed supercoiling (3.75>3.63). These departures from the heptad periodicity are necessarily accompanied by the partial loss of the knobs-into-holes interactions, as some of them become

knobs-to-knobs, with the side chains of the core residues (knobs) pointing toward the central axis instead of toward a hole. The canonical knobs-into-holes packing can also be disrupted by axial rotation of the constituent helices, which does not alter supercoiling, but can lead to the formation of knobs-to-knobs interactions. For example, counterclockwise rotation of helices (viewed from their N-terminal ends) in an antiparallel 4-helix bundle of 7/2 periodicity by about 26° relative to the canonical packing results in the formation of an *a-d-e* core, with residues in positions *a* pointing toward the center of the bundle and residues in positions *d* and *e* flanking them (Figure 1B).

Due to their regular nature, coiled-coil structures can be fully described by parametric equations (Wood *et al.*, 2017; Grigoryan and Degrado, 2011; Strelkov and Burkhard, 2003; Szczepaniak *et al.*, 2021; Crick, 1953). Such descriptions allow structural parameters to be traced down to single-residue resolution, highlighting subtle differences within and between coiled-coil domains (Figure 1C). The most important parameters describing coiled-coil structures are their topology, i.e., the number and relative arrangement of helices (parallel vs. antiparallel), the degree of supercoiling (from left-handed to nearly straight to right-handed), and the axial rotation of helices, which defines the architecture of the hydrophobic core. Such a detailed view is important for relating the structural properties of coiled-coil domains to the function of the proteins that contain them. For example, the structural parameters of the HAMP domain, a small signaling 4-helix coiled coil, have been shown to be tightly coupled to the enzymatic activity of downstream domains (Ferris *et al.*, 2011).

The high frequency with which coiled-coil domains are found (Szczepaniak *et al.*, 2021) as well as their well-understood sequence-structure relationship have motivated the development of many computational tools for their detection and modeling. The accuracy of automated coiled-coil prediction tools such as COILS (Lupas *et al.*, 1991) or DeepCoil2 (Ludwiczak *et al.*, 2019) is remarkable, in part due to the convergent nature of coiled-coil sequence repeats (non-homologous coiled coils share similar general sequence patterns due to constraints imposed by

supercoiling). Despite our ability to predict coiled coils from sequence, their structural modeling remains a challenging problem that has been addressed by programs such as CCFold (Guzenko and Strelkov, 2018) and Rosetta (Das *et al.*, 2009), which are based on fragment modeling, and ISAMBARD (Wood *et al.*, 2017), CCBuilder (Wood and Woolfson, 2018), BeammotifCC (Offer *et al.*, 2002), and CCCP (Grigoryan and Degrado, 2011), which use the aforementioned parametric equations. However, none of these programs is universally applicable. Some are limited to certain oligomeric states or require additional data besides a sequence, while others can only model perfectly regular bundles without considering local discontinuities. These discontinuities are often essential for coiled-coil function, as demonstrated in examples such as signal transduction (Ferris *et al.*, 2012) and intracellular trafficking (Murray *et al.*, 2016).

Recent years have seen a number of breakthroughs in protein structure prediction with the development of deep learning-based methods, the most prominent of which is AlphaFold2 (Jumper *et al.*, 2021; Evans *et al.*, 2022). It is implemented as an end-to-end sequence-to-structure model that also exploits evolutionary information provided in the form of a multiple sequence alignment. Benchmarks have demonstrated its superiority over classical homology modeling approaches (Lupas *et al.*, 2021) and its applicability to the modeling of protein complexes (Akdel *et al.*, 2022) and peptides (McDonald *et al.*, 2023). In addition, although not designed for such tasks, AlphaFold2 has been reported to provide insight not only into protein structure but also into its dynamics (Winski *et al.*, 2024; del Alamo *et al.*, 2022; Stein and Mchaourab, 2022; Wayment-Steele *et al.*, 2022).

In this work, we present a systematic benchmark of modeling coiled-coil domains with AlphaFold2. Going into this project, there were a number of features of coiled coils that we thought might be challenging for AlphaFold2. For example, compared to globular proteins, coiled coils have a relatively small number of contacts, which means that their structure is determined by a comparatively small number of residues. This is reflected in the fact that coiled-coil topology (oligomerization state and helical orientation) can be altered by one or a

few mutations, typically near the hydrophobic core (Harbury *et al.*, 1993). Regarding the use of evolutionary information, it is worth noting that the structural features and repetitive nature of coiled coils impose strong constraints on their sequences, which has been identified as a problem because it often leads to false matches between non-homologous coiled-coil sequences (Mistry *et al.*, 2013). This could result in the MSA provided to AlphaFold2 being "contaminated" with non-homologous coiled-coil segments that could potentially have very different structural properties. With these potential challenges in mind, we set out to evaluate the ability of AlphaFold2 to predict coiled-coil features, including global topology and local geometry.

Methods

Datasets

The benchmark was to compare experimentally determined coiled-coil structures from CCdb (Szczepaniak *et al.*, 2021) with their corresponding AlphaFold2 predictions. We focused on structures containing dimeric, trimeric, and tetrameric coiled-coil bundles, filtering out structures with resolutions below 3.5 Å or with complex topologies other than regular bundles. Finally, we retained only structures in which coiled-coil regions, defined as residues involved in knobs-into-holes interactions detected by SOCKET (Walshaw and Woolfson, 2001), accounted for more than 50% of all residues. All the filtering steps were performed using the *localpdb* package (Ludwiczak *et al.*, 2022). This process resulted in an “automatic” benchmark set of 379 coiled-coil structures.

For the oligomer state prediction benchmark, we further refined this data set. We excluded heterooligomers and structures in which the heptad register could not be assigned with TWISTER (Strelkov and Burkhard, 2003) (unambiguous register assignment was critical because some of the benchmarked methods use this feature to improve prediction accuracy). This resulted in a set of 216 structures, with each oligomer class (38% dimers, 36% trimers, and 26%

tetramers) similarly represented. In addition to these sets, we constructed two additional sets based on manually curated examples, consisting of 19 parallel (Szczepaniak *et al.*, 2018) and 21 antiparallel (Szczepaniak *et al.*, 2014) coiled-coil structures, most of which were GCN4 variants. The four benchmark sets are summarized in Supplementary Table 1.

AlphaFold2 modelling

Modeling was performed with ColabFold (version 2.5.2) (Mirdita *et al.*, 2022) using the *alphafold2_multimer_v3* model, 5 rounds of recycling, and Amber minimization. It is important to note that although we modeled the entire sequences (i.e., including the non-coiled coil regions), we performed the structural analyses only on the coiled-coil regions.

Oligomerization state prediction

We selected to benchmark LOGICOL (Vincent *et al.*, 2013) and CoCoNat (Madeo *et al.*, 2023) as representatives of state-of-the-art coiled-coil oligomerization state prediction methods. Unlike AlphaFold2, both programs can be run with coiled-coil register information, which improves the robustness of the predictions. We benchmarked LOGICOL and CoCoNat in two different scenarios, a) with the structurally derived heptad registers obtained with TWISTER (Strelkov and Burkhard, 2003), and b) without explicitly providing register information (indicated by the "noregister" suffix in Figure 2). While CoCoNat implements its own register prediction, we opted to run LOGICOIL in the same way as its web server, with registers predicted by MARCOIL (Delorenzi and Speed, 2002).

To assess the extent to which AlphaFold2 can be used to predict the oligomeric state of coiled-coil domains, we used it to model each of the 216 sequences in the “oligomerization” benchmark set as a dimer, a trimer, and a tetramer, and checked whether the oligomerization state of the best-scoring model matched the oligomeric state observed in the experimental

structures. We tested several metrics, including pLDDT, pTM, and PAE. We further refined the PAE score (total_PAE) into the interchain PAE (inter_PAE), which can be thought of as the uncertainty in predicting the relative position of the helices. Confusion matrices and weighted F1 scores were calculated for each method (Figure 2).

Given the good accuracy of the AlphaFold2 quality metrics in predicting oligomeric state, and the fact that AlphaFold2 was not directly supervised for oligomeric state prediction (Evans *et al.*, 2022), we hypothesized that accuracy could be further improved by supervised learning on the internal representations of AlphaFold2. To test this, we extracted representations for each benchmark sequence modeled as a monomer (retrieved with the --save-single-representations option). The resulting representations were then averaged over the sequence length dimension, resulting in fixed size vectors of shape 5x256 (5 corresponds to each of the AlphaFold2 pre-trained models and 256 to the embedding size) per sequence. For visualization purposes, the dimensionality of the embeddings was reduced using PacMAP (Wang *et al.*, 2021) (Figure 3A, left panel).

To process these concatenated AlphaFold2 embeddings, we initially trained a simple neural network using fivefold cross-validation. The neural network architecture included multiple dense layers with batch normalization, dropout, and L2 regularization. The model was trained using the Adam optimizer (Kingma and Ba, 2017) and a sparse categorical cross-entropy loss function. Early stopping with a patience of 3 epochs was applied during training to avoid overfitting. The dimensionality of the internal representations (from the last layer) of the model was reduced using PacMAP (Wang *et al.*, 2021) and visualized (Figure 3A, right panel). We also considered simpler classifiers available in the scikit-learn Python package, and based on fivefold cross-validation, we found LogisticRegression to be the best performing. The internal classifier parameters, namely regularization strength and class balance weighting, were optimized, and the final model, available on GitHub, was fitted to all available data without cross-validation (https://github.com/labstructbioinf/dc2_oligo).

Bundle geometry prediction benchmark

We modeled each structure in the "automatic", "parallel", and "antiparallel" benchmark sets with the oligomerization state matching that of the experimental structure. For comparison with the corresponding experimental structures, we took the top-ranked model as representative of each prediction. We then superimposed the predicted and experimental structures using US-Align (Zhang *et al.*, 2022), checking that all chains were aligned in the correct orientation and excluding misaligned cases from further analysis. Finally, we used SamCC Turbo to detect the coiled-coil segments and compute their structural parameters (helical axial rotation and supercoiling). An example of the comparison of these parameters between an experimental and a predicted model is shown in Figure 1C. Details on the calculation of these parameters can be found in (Szczepaniak *et al.*, 2021).

Results and Discussion

Oligomerization state prediction

We began the benchmark by evaluating the performance of current methods for predicting the oligomerization state of coiled-coil domains. In addition to the sequence-based methods LOGICOIL (Vincent *et al.*, 2013) and CoCoNat (Madeo *et al.*, 2023), we also evaluated the predictive power of the AlphaFold2 model quality metrics pLDDT, pTM, and PAE. For predictions based on AlphaFold2 scores, we computed models for oligomerization states ranging from dimer to tetramer for a given sequence. Then, for each metric, the oligomerization of a model with the best score was selected as the prediction. For each approach, we computed a confusion matrix comparing the predicted number of helices with the number of helices in the experimental structures (Figure 2). For AlphaFold2, since it outputs a structural model, we also checked whether the helix orientation was correct (this was the case for 98% of the parallel and 76% of the antiparallel bundles). As expected, both LOGICOIL

and CoCoNat performed well in the oligomerization state prediction task when provided with structurally derived coiled-coil registers, with weighted F1 scores of 0.71 and 0.77, respectively, although LOGICOIL showed a strong bias towards predicting trimers, as did CoCoNat for dimers. When we did not provide the structurally derived heptad register (see Methods), the F1 scores dropped to 0.63 and 0.72, respectively (in addition to a larger number of incorrect predictions, some predictions were ambiguous, marked "?" in Figure 2).

Among the AlphaFold2 metrics tested, the best performing were pLDDT (F1=0.70) and pTM (F1=0.63), with the former showing a strong bias towards dimers and the latter towards tetramers. We also tested the PAE score, which is a 2D matrix representing the predicted error in the pairwise distances between all residues in all chains. In the benchmark, we considered the average of the entire PAE matrix (total_PAE) and the average of the PAE values between residues from different chains (inter_PAE). The average total_PAE (F1=0.71) was comparable to the pLDDT and slightly worse than the inter_PAE (F1=0.73), which additionally showed a much better ability to predict tetrameric bundles (Figure 2).

Intrigued by the fact that AlphaFold2 had comparable accuracy to state-of-the-art methods despite not being trained to predict oligomerization state, we investigated whether such predictions could be made directly from the internal representations of AlphaFold2, rather than from the scores associated with models computed for different oligomerization states. Projecting the AlphaFold2 representations of the “oligomerization” benchmark set cases into 2D space revealed some degree of separation between dimers, trimers, and tetramers (Figure 3A, left panel). Motivated by this observation, we trained a simple neural network to predict a coiled-coil oligomeric state based on its corresponding AlphaFold2 representation. Projection of the embeddings obtained from this neural network showed a clear separation between the groups (Figure 3A, right panel), suggesting the potential applicability of such an approach. In exploring alternative prediction models, we found that even a simple regression model can produce very good results: 5-fold cross-validation on the benchmark set showed significantly

improved performance (F1=0.82) without any noticeable bias in the preferred oligomeric state (Figure 3B). Note, however, that we evaluated this model using 5-fold cross-validation, so the results obtained are not directly comparable to those shown in Figure 2. The implementation, along with the training routines, has been deposited on GitHub (https://github.com/labstructbioinf/dc2_oligo).

Modeling of coiled-coil geometry

To perform a detailed comparison of experimental structures with AlphaFold2 models, we used SamCC Turbo (Szczepaniak *et al.*, 2021), a program designed to automatically detect coiled-coil domains and calculate their structural parameters, including the degree of bundle supercoiling and hydrophobic core geometry (Figure 1). Comparisons were made using three separate sets of benchmarks, each designed to highlight a specific challenge in coiled-coil modeling. The first and second sets focused on packing geometries in parallel and antiparallel 4-helix bundles, respectively, while the third set was more general and included bundles with different oligomerization states and degrees of supercoiling. For each benchmark, all structures were modeled in the correct oligomerization state obtained from the corresponding experimental structure, and the models were then analyzed using SamCC Turbo.

The first set, termed "parallel", consists of 19 canonical 7/2 parallel 4-helix bundles in which the helices interact by knobs-into-holes packing. Despite their overall similarity, these structures exhibit slight differences in the geometries of their hydrophobic cores. These differences are primarily manifested in the axial rotation of the helices, which is within ±5° (calculated relative to an idealized reference model of a coiled-coil employing knob-into-holes packing; Figure 1). Our previous work showed that these subtle differences are not artifacts of the experimental procedures but can be related to the sequence composition of the hydrophobic core (Szczepaniak *et al.*, 2018). All obtained models showed correct topology and their hydrophobic cores were modeled with high fidelity (Figure 4A). The only exception was the model of the

designed CC-Hex-II structure (PDB: 4h7r), in which the interhelical interfaces were incorrectly predicted.

The "antiparallel" set consisted of 21 four-helix antiparallel bundles with hydrophobic core geometries defined by rotations ranging from -26° to +10°. Structures with rotations up to ±10° are considered to have canonical knobs-in-holes packing (Figure 1B, left panel), whereas those with rotations below -10° are considered to have non-canonical packing (Figure 1B, right panel). Canonical packing is characterized by the interaction of two residues per heptad repeat within the hydrophobic core. In contrast, non-canonical packing involves the co-option of an additional heptad position e into the core (Lupas et al., 2017). In this benchmark, AlphaFold2 correctly modeled only 13 out of 21 structures. In most of the unsuccessful cases, the resulting models had the wrong topology (parallel instead of antiparallel), preventing a meaningful comparison with the experimental structures. However, the models with the correct topology were highly accurate (Figure 4B), except for the coiled-coil domain of the coronavirus S2 transmembrane fusion protein (PDB: 1zv7), where the core geometry was incorrectly predicted.

The two benchmark sets described above contained hand-picked structures, mostly GCN4 variants. To evaluate AlphaFold2 more comprehensively, we generated a third set containing 379 dimers, trimers, and tetramers of various topologies. Unlike the smaller benchmark sets, this set also included non-canonical bundles characterized by right-handed twist and periodicity >3.63 associated with repeats such as hendecads and pentadecads (Figure 1A). As with the two smaller sets, among the correct topology models (323 out of 379), the majority were accurate with respect to both helix packing and bundle periodicity (Figure 4C). Interestingly, the quality of modeling did not depend on the release date of the corresponding experimental structures, and those released after AlphaFold2 training (2018-04-30) were modeled equally well (compare red and green dots in Figure 4C).

Conclusions

While AlphaFold2 provides highly accurate predictions of the oligomeric state (Figures 2 and 3) and fine details such as supercoiling and core geometry (Figure 4), its ability to predict the correct orientation of the helices is limited. In all benchmarks, we observed the same tendency for AlphaFold2 to model antiparallel structures as parallel. For example, in the benchmark focused on predicting the oligomerization state, the helix orientation was correct in 98% of the parallel and 76% of the antiparallel test cases. Similarly, in the benchmark focused on a set of 21 antiparallel structures (Figure 4B), only 13 out of 21 were correctly modeled as antiparallel. The problems in predicting the correct helical arrangement were to be expected, since it is dictated by the long-range contacts, which can be difficult to capture due to the strong convergent evolutionary signal in coiled-coil domains. This problem has less impact on the prediction of the oligomeric state and core geometry, since these are clearly defined by the features of the interhelical interfaces, such as the presence of additional hydrophobic positions (Szczepaniak *et al.*, 2014) or the formation of multiple interfaces by a single helix (Lupas *et al.*, 2017).

As we looked at models that differed from the experimental structures, we realized that some of them might be more accurate. For example, the experimental structure of the human tetherin protein (PDB: 3mqc) is a tetramer, while the corresponding rank 1 AlphaFold2 model takes the form of two separate dimers (Figure 5A). The rank 2 model is a tetramer with the same helical arrangement as the experimental structure, but contains numerous helical distortions that significantly reduce the quality metrics of this model. Although the rank 1 model does not match the experimental structure, it is actually more accurate because the dimeric form of tetherin has been shown to be functional (Yang *et al.*, 2010). Another example is the human CtIP protein (Figure 5B); again, there is a clear discrepancy between the experimental structure (PDB: 7bgf) and the AlphaFold2 rank 1 model. While the experimental structure assumes a straight coiled coil, the AlphaFold2 model appears to be broken in the middle. The observed break in the model results from the presence of a functionally important

region, a zinc hinge, which is essential for the function of the protein (Morton *et al.*, 2021). Such examples highlight the possibility that some of the AlphaFold2 models can provide clues to protein function even when they do not agree with experimental structures.

Finally, it's worth highlighting the unexpected accuracy in predicting the oligomerization state using AlphaFold2 hidden representations. Such an approach proved to be not only effective (Figure 3), but also faster than modeling a given coiled-coil sequence in all possible oligomerization states (Figure 2). Recently, AlphaFold2 embeddings have also been explored in the context of predicting protein-ligand interactions (Gazizov *et al.*, 2023). Prediction of oligomerization states has also been attempted using ESM2 model embeddings (Sledzieski *et al.*, 2023). We also note that new AlphaFold2-based methods for predicting structures of homooligomers (Schweke *et al.*, 2024) and complex heteroligomeric assemblies (Shor and Schneidman-Duhovny, 2024) have recently been proposed.

In this work, we have highlighted the tremendous versatility of AlphaFold2, demonstrated here by its ability to accurately model important details of coiled-coil domains, and presented a tool capable of predicting the oligomeric state of coiled coils from sequence. We expect that our results will be useful to the coiled-coil community, for example to address challenging problems such as understanding the complex oligomerization propensities of some coiled-coil families, and to detect non-canonical coiled coils with divergent sequence properties.

Acknowledgments

S.D-H. and M.M.G. were supported by institutional funds of the Max Planck Society. This work was also supported by the "Excellence Initiative - Research University" program, an internal grant of the University of Warsaw to enhance the research potential of its staff.

Figure Legends

Figure 1. Coiled-coil domains and quantification of their structures. **(A)** Examples of coiled-coil structures with different periodicities. The most typical periodicity is 7/2, which induces the left-handed twist of a bundle, while others such as 11/3 or 15/4 are considered non-canonical. **(B)** Helical wheel diagrams of antiparallel coiled coils with heptad positions labeled *a-g*. Letters in the center of the helices indicate whether they are viewed from the N- or C-terminal side. The left diagram corresponds to the canonical *a-d* geometry, while the right diagram corresponds to the extended *a-d-e* geometry caused by a counterclockwise axial rotation of all helices by about 26 degrees, viewed from the N-terminus. **(C)** Experimental structure and corresponding AlphaFold2 model of the tropomyosin fragment. The results of the measurement of two structural parameters, helix axial rotation and periodicity, are shown below in scale. These parameters are defined separately for each individual layer of an input structure (layer is defined as a set of *n* residues, where *n* corresponds to the number of helices in the bundle, roughly localized on a plane perpendicular to the bundle axis).

Figure 2. Accuracy of AlphaFold2, LOGICOIL and CoCoNat in predicting the oligomeric state of coiled coils. LOGICOIL and CoCoNat were run in two modes, with and without ("_noregister" suffix) providing structurally derived heptad registers. AlphaFold2 predictions were made using the model quality metrics pLDDT, pTM and PAE, the latter divided into total_PAE and inter_PAE (see text for details). The accuracy of each method is shown as a confusion matrix with true and predicted labels on the y and x axes, respectively. Weighted F1 coefficients are also provided.

Figure 3. Application of AlphaFold2 representations for the prediction of the oligomeric state. **(A)** 2D plots of raw AlphaFold2 representations of the benchmark cases (left panel) and representations obtained from a downstream model (right panel). Colors represent oligomerization states. **(B)** A confusion matrix with true and predicted labels on the y and x axes, respectively, showing the performance of the final prediction model trained on AlphaFold2 representations. The weighted F1 score is also shown.

Figure 4. Comparison of coiled-coil parameters between experimental (PDB) and AlphaFold2 (model) structures. The scatter plots show the differences in average helix rotation and periodicity, while the accompanying plots (right) show detailed per-layer measurements for exemplary structures (model and experimental structures are shown in red and light blue, respectively). Panels A, B and C show the results for the "parallel", "antiparallel", and "automatic" benchmark sets, respectively.

Figure 5. Examples of successes and failures in modeling coiled-coil structures with AlphaFold2. See the text for details.

References

Akdel,M. *et al.* (2022) A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.*, **29**, 1056–1067.

del Alamo,D. *et al.* (2022) Sampling alternative conformational states of transporters and receptors with AlphaFold2. *Elife*, **11**, e75751.

Crick,F.H.C. (1953) The Fourier transform of a coiled-coil. *Acta Crystallogr.*, **6**, 685–689.

Das,R. *et al.* (2009) Simultaneous prediction of protein folding and docking at high resolution. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 18978–83.

Delorenzi,M. and Speed,T. (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics*, **18**, 617–25.

Evans,R. *et al.* (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.10.04.463034.

Ferris,H.U. *et al.* (2012) Mechanism of regulation of receptor histidine kinases. *Structure*, **20**, 56–66.

Ferris,H.U. *et al.* (2011) The mechanisms of HAMP-mediated signaling in transmembrane receptors. *Structure*, **19**, 378–85.

Gazizov,A. *et al.* (2023) AF2BIND: Predicting ligand-binding sites using the pair representation of AlphaFold2. *bioRxiv*, 2023.10.15.562410.

Grigoryan,G. and Degrado,W.F. (2011) Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.*, **405**, 1079–100.

Guzenko,D. and Strelkov,S. V (2018) CCFold: rapid and accurate prediction of coiled-coil structures and application to modelling intermediate filaments. *Bioinformatics*, **34**, 215–222.

Harbury,P.B. *et al.* (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, **262**, 1401–7.

Jumper,J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

Kingma,D.P. and Ba,J. (2017) Adam: A Method for Stochastic Optimization.

Ludwiczak,J. *et al.* (2019) DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics*, **35**, 2790–2795.

Ludwiczak,J. *et al.* (2022) Localpdb- a Python package to manage protein structures and their annotations. *Bioinformatics*.

Lupas,A. *et al.* (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–4.

Lupas,A. *et al.* (2017) The Structure and Topology of α-Helical Coiled Coils. In, *Sub-Cellular Biochemistry*., pp. 95–129.

Lupas,A.N. *et al.* (2021) The breakthrough in protein structure prediction. *Biochem. J.*, **478**, 1885–1890.

Madeo,G. *et al.* (2023) CoCoNat: a novel method based on deep learning for coiled-coil prediction. *Bioinformatics*, **39**.

Martinez-Goikoetxea,M. and Lupas,A.N. (2023) New protein families with hendecad coiled coils in the proteome of life. *J. Struct. Biol.*, **215**, 108007.

McDonald,E.F. *et al.* (2023) Benchmarking AlphaFold2 on peptide structure prediction. *Structure*, **31**, 111-119.e2.

Mirdita,M. *et al.* (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*, **19**, 679–682.

Mistry,J. *et al.* (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.*, **41**, e121.

Morton,C.R. *et al.* (2021) Structural basis for the coiled-coil architecture of human CtIP. *Open Biol.*, **11**, 210060.

Murray,D.H. *et al.* (2016) An endosomal tether undergoes an entropic collapse to bring vesicles together. *Nature*, **537**, 107–111.

Offer,G. *et al.* (2002) Generalized Crick equations for modeling noncanonical coiled coils. *J. Struct. Biol.*, **137**, 41–53.

Schweke,H. *et al.* (2024) An atlas of protein homo-oligomerization across domains of life. *Cell*.

Shor,B. and Schneidman-Duhovny,D. (2024) CombFold: predicting structures of large protein assemblies using a combinatorial assembly algorithm and AlphaFold2. *Nat. Methods*.

Sledzieski,S. *et al.* (2023) Democratizing Protein Language Models with Parameter-Efficient Fine-Tuning. *bioRxiv*, 2023.11.09.566187.

Stein,R.A. and Mchaourab,H.S. (2022) SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLOS Comput. Biol.*, **18**, e1010483.

Strelkov,S. V and Burkhard,P. (2003) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol.*, **137**, 54–64.

Szczepaniak,K. *et al.* (2021) A library of coiled-coil domains: from regular bundles to peculiar

twists. *Bioinformatics*, **36**, 5368–5376.

Szczepaniak,K. *et al.* (2014) Designability landscape reveals sequence features that define axial helix rotation in four-helical homo-oligomeric antiparallel coiled-coil structures. *J. Struct. Biol.*, **188**, 123–133.

Szczepaniak,K. *et al.* (2018) Variability of the core geometry in parallel coiled-coil bundles. *J. Struct. Biol.*, **204**, 117–124.

Vincent,T.L. *et al.* (2013) LOGICOIL--multi-state prediction of coiled-coil oligomeric state. *Bioinformatics*, **29**, 69–76.

Walshaw,J. and Woolfson,D.N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol.*, **307**, 1427–50.

Wang,Y. *et al.* (2021) Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization. *J. Mach. Learn. Res.*, **22**, 1–73.

Wayment-Steele,H.K. *et al.* (2022) Prediction of multiple conformational states by combining sequence clustering with AlphaFold2. *bioRxiv*.

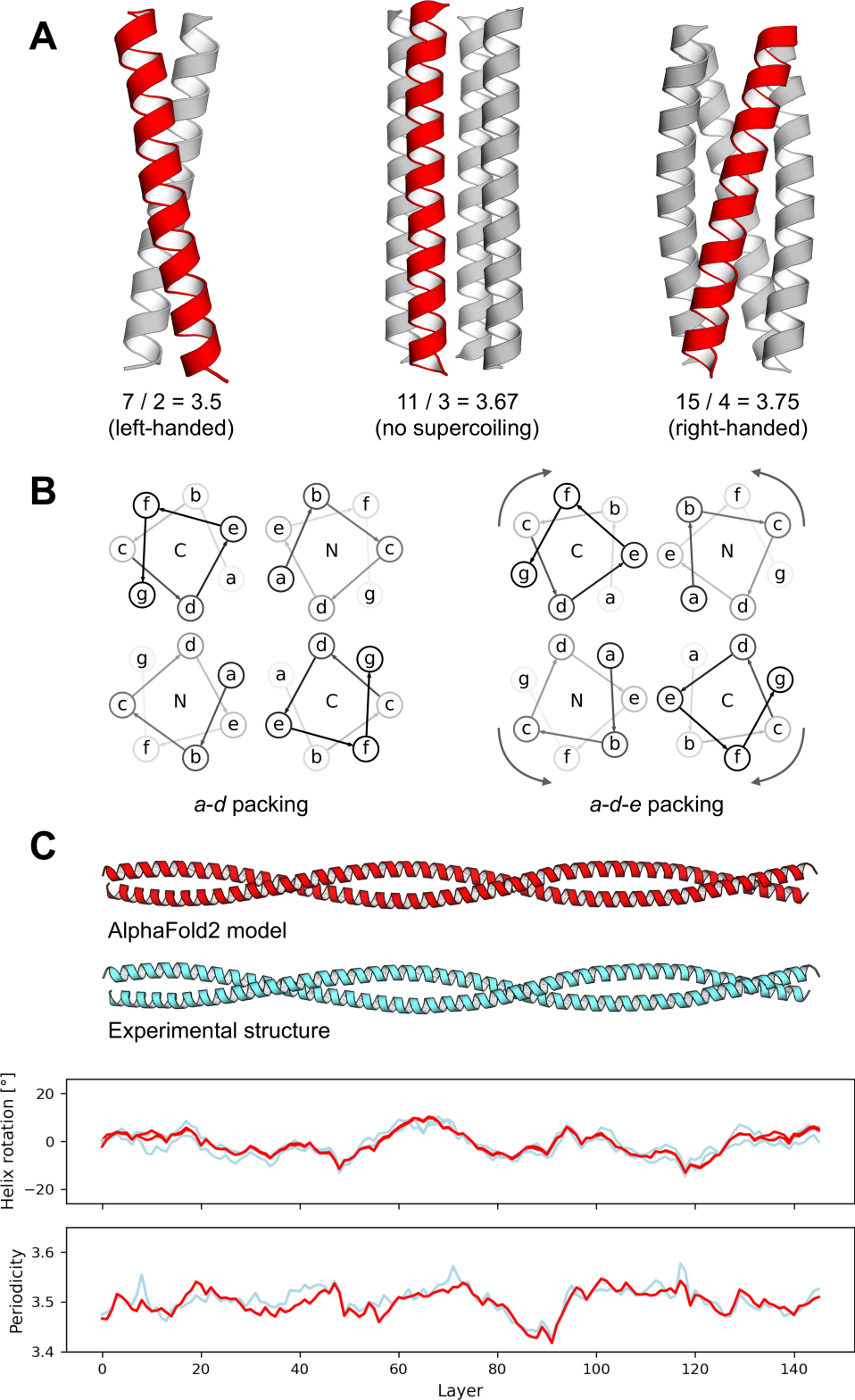
Winski,A. *et al.* (2024) AlphaFold2 captures the conformational landscape of the HAMP signaling domain. *Protein Sci.*, **33**, e4846.

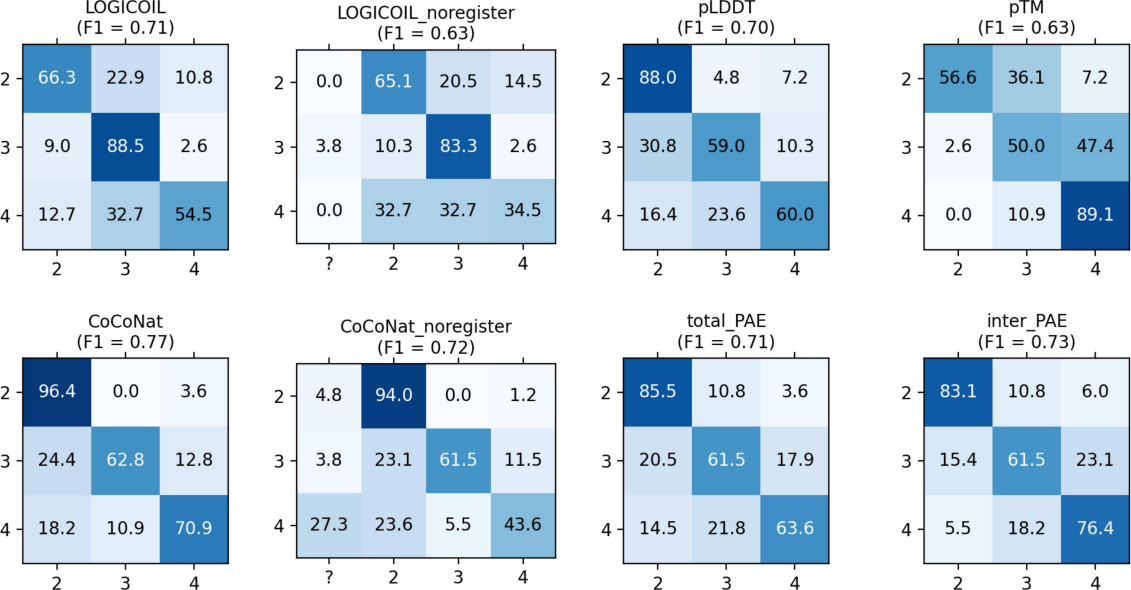
Wood,C.W. *et al.* (2017) ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design. *Bioinformatics*.

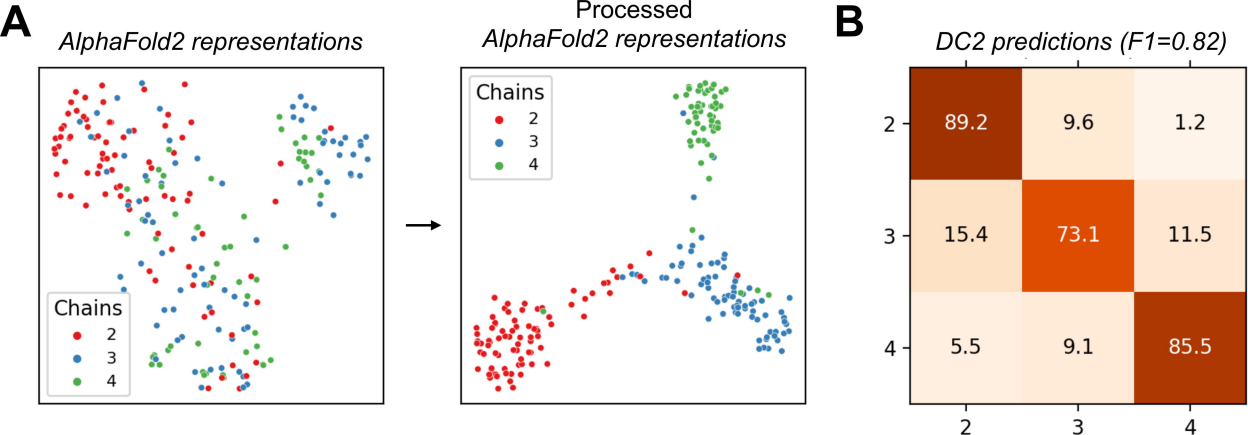
Wood,C.W. and Woolfson,D.N. (2018) CCBuilder 2.0: Powerful and accessible coiled-coil modeling. *Protein Sci.*, **27**.

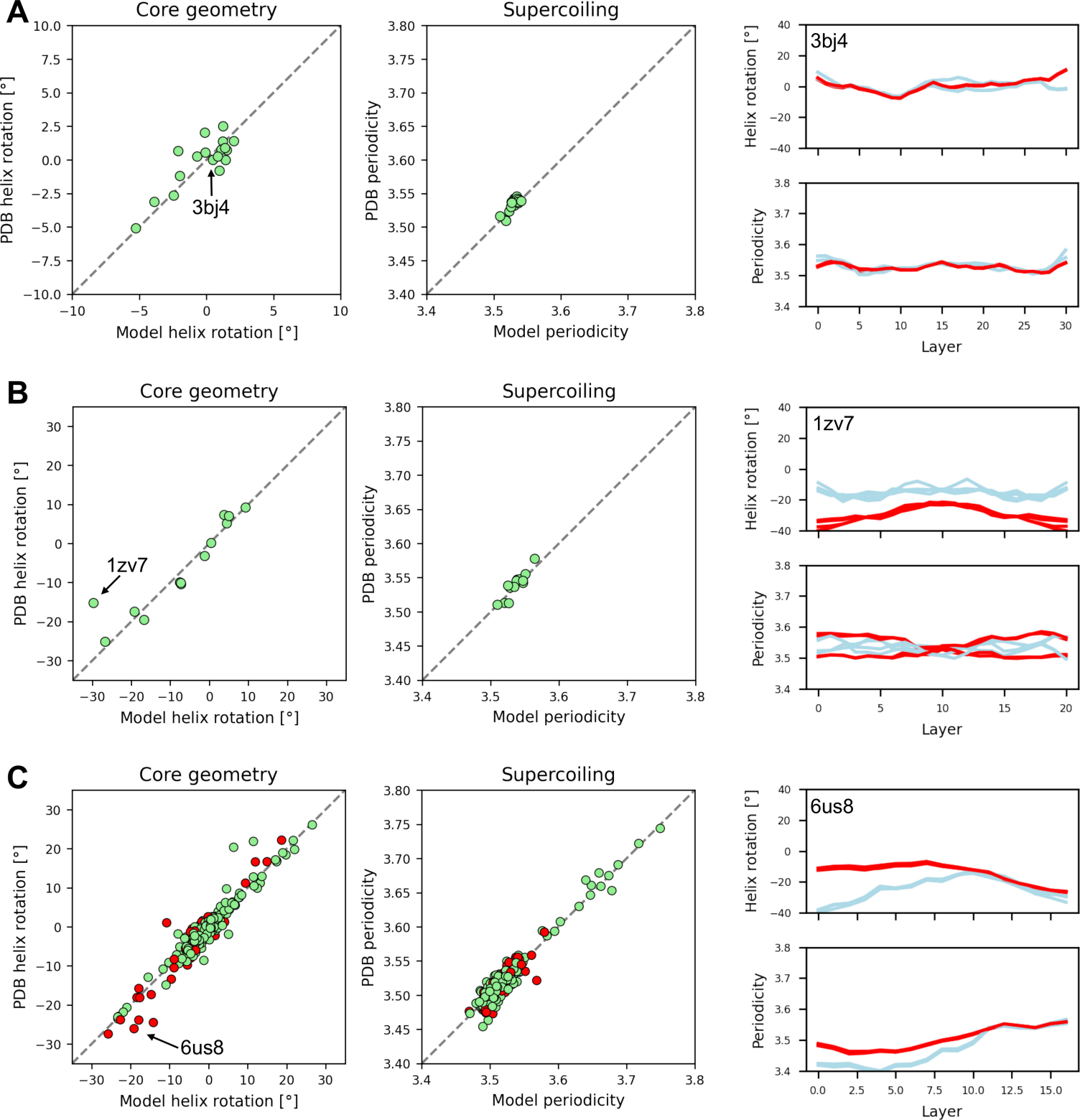
Yang,H. *et al.* (2010) Structural insight into the mechanisms of enveloped virus tethering by tetherin. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 18428–32.

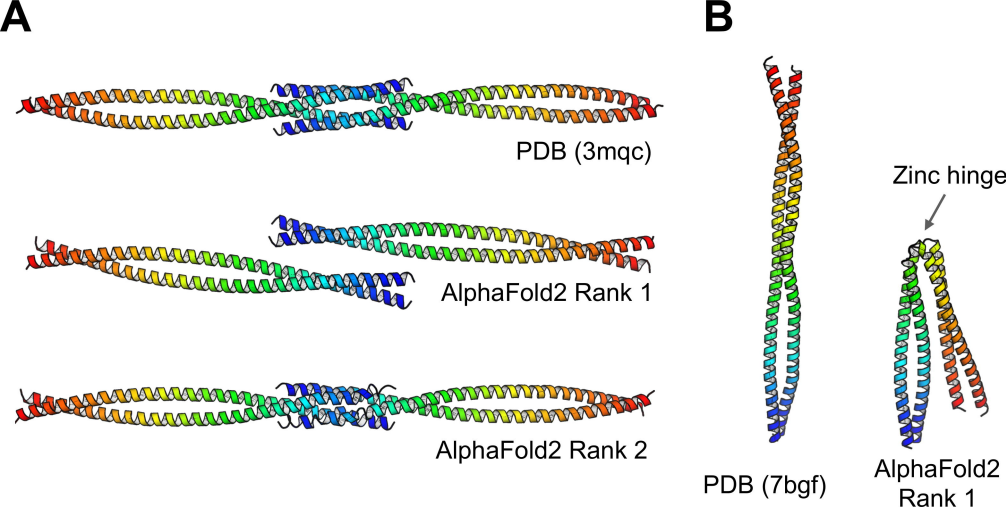
Zhang,C. *et al.* (2022) US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods*, **19**, 1109–1115.


A
7 / 2 = 3.5
(left-handed)
11 / 3 = 3.67
(no supercoiling)
15 / 4 = 3.75
(right-handed)
B
a-d packing
a-d-e packing
C
AlphaFold2 model
Experimental structure
Helix rotation [°]
Periodicity
Layer


LOGICOIL
(F1 = 0.71)
2 66.3 22.9 10.8
3 9.0 88.5 2.6
4 12.7 32.7 54.5
2 3 4
LOGICOIL_noregister
(F1 = 0.63)
2 0.0 65.1 20.5 14.5
3 3.8 10.3 83.3 2.6
4 0.0 32.7 32.7 34.5
? 2 3 4
pLDDT
(F1 = 0.70)
2 88.0 4.8 7.2
3 30.8 59.0 10.3
4 16.4 23.6 60.0
2 3 4
pTM
(F1 = 0.63)
2 56.6 36.1 7.2
3 2.6 50.0 47.4
4 0.0 10.9 89.1
2 3 4
CoCoNat
(F1 = 0.77)
2 96.4 0.0 3.6
3 24.4 62.8 12.8
4 18.2 10.9 70.9
2 3 4
CoCoNat_noregister
(F1 = 0.72)
2 4.8 94.0 0.0 1.2
3 3.8 23.1 61.5 11.5
4 27.3 23.6 5.5 43.6
? 2 3 4
total_PAE
(F1 = 0.71)
2 85.5 10.8 3.6
3 20.5 61.5 17.9
4 14.5 21.8 63.6
2 3 4
inter_PAE
(F1 = 0.73)
2 83.1 10.8 6.0
3 15.4 61.5 23.1
4 5.5 18.2 76.4
2 3 4


A
AlphaFold2 representations
Chains
2
3
4
Processed
AlphaFold2 representations
Chains
2
3
4
B
DC2 predictions (F1=0.82)
2
3
4
89.2
9.6
1.2
15.4
73.1
11.5
5.5
9.1
85.5
2
3
4


A
Core geometry
PDB helix rotation [°]
Model helix rotation [°]
3bj4
Supercoiling
PDB periodicity
Model periodicity
3bj4
Helix rotation [°]
Periodicity
Layer
B
Core geometry
PDB helix rotation [°]
Model helix rotation [°]
1zv7
Supercoiling
PDB periodicity
Model periodicity
1zv7
Helix rotation [°]
Periodicity
Layer
C
Core geometry
PDB helix rotation [°]
Model helix rotation [°]
6us8
Supercoiling
PDB periodicity
Model periodicity
6us8
Helix rotation [°]
Periodicity
Layer


A
PDB (3mqc)
AlphaFold2 Rank 1
AlphaFold2 Rank 2
B
Zinc hinge
PDB (7bgf)
AlphaFold2
Rank 1