THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Prediction of Intrinsic Disorder Using Rosetta ResidueDisorder and AlphaFold2

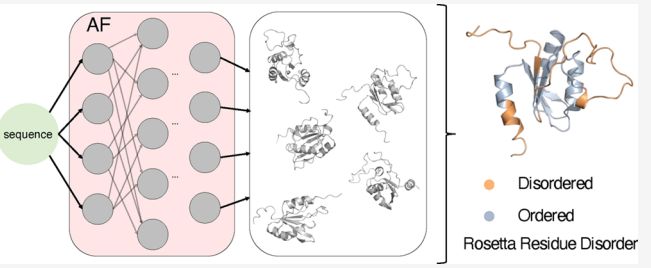
Jiadi He,[§] SM Bargeen Alam Turzo,[§] Justin T. Seffernick,[§] Stephanie S. Kim, and Steffen Lindert*

Cite This: *J. Phys. Chem. B* 2022, 126, 8439–8446

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The combination of deep learning and sequence data has transformed protein structure prediction and modeling, evidenced in the success of AlphaFold (AF). For this reason, many methods have been developed to take advantage of this success in areas where inaccurate structural modeling may limit computational predictiveness. For example, many methods have been developed to predict protein intrinsic disorder from sequence, including our Rosetta ResidueDisorder (RRD) approach. Intrinsically disordered regions in proteins are parts of the sequence that do not form ordered, folded structures under typical physiological conditions. In the original implementation of RRD, Rosetta *ab initio* models were generated, and disordered regions were predicted based on residue scores (disordered residues typically exist in regions of unfavorable scores). In this work, we show that by (i) replacing the *ab initio* modeling with AF (using the same scoring and disorder assignment approach) and (ii) updating the score function, the predictiveness improved significantly. Residues were better ranked by the order/disorder, evidenced by an improvement in receiver operating characteristic area-under-the-curve from 0.69 to 0.78 on a large (229 protein) and balanced data set (relatively even ordered versus disordered residues). Finally, the binary prediction accuracy also improved from 62% to 74% on the same data set. Our results show that the combined AF-RRD approach was as good as or better than all existing methods by these metrics (AF-RRD had the highest prediction accuracy).



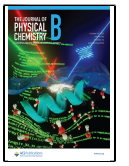
INTRODUCTION

Proteins and protein domains have been historically viewed as structures with rigid and stable features that play key roles in many biological functions.[1] However, there are numerous proteins that do not exist as a single, stable structure under physiological conditions, and instead sample a wide range of conformations.[2] The level of disorder in proteins is encoded by the amino acid sequence, favored by features such as lower sequence complexity, fewer hydrophobic amino acids, and more highly charged or highly hydrophilic amino acids to name a few.[3] These sequence properties cause relatively flat free energy surfaces, which allow the systems to transform between numerous conformations with low energy barriers. Because of this intrinsic nature of such systems, they are commonly known as intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs).[1] IDPs and IDRs are highly abundant in nature, and a large fraction are found in eukaryotic organisms.[4−6]

It has been shown that around 30% of the sequences (of proteins with 30 residues or more) in the human proteome have some IDRs.[7] Additionally, IDPs/IDRs play important roles in several biological functions such as transcriptional regulation, translation, phosphorylation, and cellular signal transduction.[3,8,9] Furthermore, IDPs/IDRs are also associated with human diseases such as various genetic disorders,[10] Alzheimer's disease,[11] as well as neurodegenerative[12] and cardiovascular[13] diseases. Because of their importance across various domains of biology and medicine, several experimental techniques have been utilized to identify IDPs/IDRs and determine intrinsic disorder (or lack thereof) of specific regions in a system. Nuclear magnetic resonance (NMR) spectroscopy,[14] circular dichroism (CD) spectroscopy,[15] small-angle X-ray scattering (SAXS),[16] and single-molecule fluorescence resonance energy transfer (smFRET)[17] are well established methods to study IDPs/IDRs. In addition to experimental techniques, fast and free computational methods to probe disorder in silico also play an important role.

There are many algorithms that have been developed to predict IDPs/IDRs.[18,19] RaptorX,[20] IUPred,[21−23] Metapredict,[24] and PrDOS[25] are a few popular computational methods available for disorder prediction from sequences. Generally, these methods use some combination of machine learning, sequence composition, energy function, and/or local secondary structure prediction to characterize the intrinsic disorder

Received: August 2, 2022
Revised: September 27, 2022
Published: October 17, 2022



https://doi.org/10.1021/acs.jpcb.2c05508
J. Phys. Chem. B 2022, 126, 8439−8446

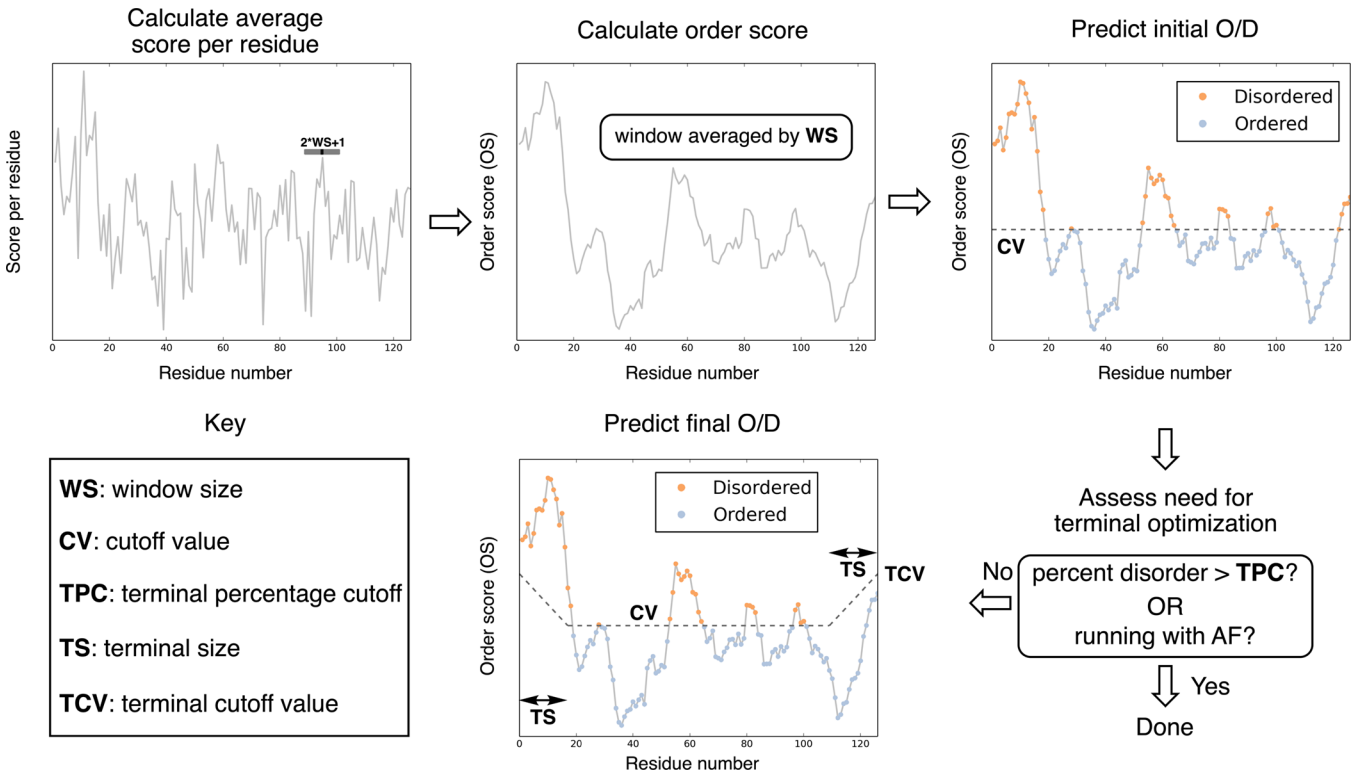


Figure 1. Overview of RRD and relevant parameters. Contains pictorial definitions of the window size (WS), cutoff value (CV), terminal percentage cutoff (TPC), terminal size (TS), and terminal cutoff value (TCV).

in proteins. In the following, we provide a brief explanation of how each of these disorder prediction method works. RaptorX is a sequence-based method for order/disorder prediction. RaptorX utilizes multiple sequence alignments as well as deep learning (deep convolutional neural fields) to categorize each residue into one of its secondary structure categories. IUPred is also a sequence-based method that predicts the tendency for each residue to be ordered/disordered. IUPred achieves this by estimating an energy based on an empirical force field. Metapredict is another sequence-based method for predicting the stability of residues. Metapredict uses a bi-directional recurrent neural network to predict each residue with a predicted disorder score. PrDOS uses a supervised machine learning technique (support vector machine) to make predictions from amino acid sequence and determines each residue as ordered/disordered. Recently, we developed the Rosetta ResidueDisorder (RRD)[26] application which can accurately predict the intrinsic disorder of proteins. Unlike most other methods, RRD utilizes structural modeling (such as the *ab initio*[27−30] structure prediction method in Rosetta) to first generate an ensemble of structures from the primary sequence. Next, RRD predicts the intrinsic disorder of each residue based on the average scores over the ensemble of predicted structures. We hypothesized that this method allows for the accounting of long-range interactions between residues that other purely sequence-based methods may not. In a benchmark of 245 proteins, the RRD approach outperformed other methods. Within RRD, we have also demonstrated that disorder can be predicted directly from the structure, if known, using the same methodology.[26,31] Furthermore, we examined conformational changes of proteins from molecular dynamics (MD) simulations with RRD and accurately predicted folding and unfolding events.[31]

Recent advances in protein structure modeling have been fueled by deep learning and sequence coevolution, evidenced in various highly accurate algorithms such as AlphaFold[32] and RoseTTAFold.[33] During the 14th edition of the Critical Assessment of Protein Structure Prediction in 2020 (CASP14), AlphaFold2 (AF)[32] significantly outperformed all other methods at predicting protein structures from sequences. Recent studies also showed that AF assigns a low predicted local-distance difference test (pLDDT)[32,34] score for disordered residues in IDPs/IDRs.[35−37]

Based on these exciting new developments, we proposed to build structural models for target sequences using state-of-the-art deep learning algorithms and utilize the structures within the RRD protocol. Using the same, balanced data sets employed in previous work, we benchmarked the performance of RRD with these new modeling approaches against existing methods such as RaptorX,[20] PrDOS,[25] IUPred,[22,23] Metapredict,[24] and the approach using AF pLDDT.[36] The prediction accuracy and receiver operating characteristic (ROC) curves were used to quantify the performance. Using our benchmark sets, we demonstrate that the combination of AF and RRD performed well compared to the other investigated methods with respect to predicting disordered regions, with an accuracy of 74% (about 10 percentage points higher than our previous work[26] with *ab initio* modeling) and an ROC curve area-under-the-curve (AUC) of 0.78 on a large (229 proteins), independent test set.

METHODS

RRD can be used to predict intrinsic disorder from a primary sequence (by assigning each residue as ordered or disordered), as described in detail in previous work[26] and schematically depicted in Figure 1. In short, the method works by first

https://doi.org/10.1021/acs.jpcb.2c05508
J. Phys. Chem. B 2022, 126, 8439−8446

predicting a set of tertiary structures (previously 100 models using the Rosetta *ab initio* protocol). Next, the average per residue scores (previous work used the Talaris2014 [t14] energy function[38]) of each residue are calculated from the models. These average residue scores are then smoothed by calculating a window average with the window size (WS), resulting in the order score (OS). In our previous work, we used WS = 5 for *ab initio* t14, that is, yielding the average score of N-5, N-4, ..., N, ... N + 4, and N + 5, where N is the residue of interest. Residues with OS greater than the cutoff value (CV = −1.0 for *ab initio* t14) are identified as disordered (indicating unfavorable Rosetta scores in the region), with residues having OS less than CV identified as ordered (indicating favorable Rosetta scores in the region). Finally, if the system is identified as mostly ordered after the first identification (disorder percentage less than terminal percentage cutoff [TPC, TPC = 60% for *ab initio* t14]), the cutoff is linearly increased for the terminal ends, and the disorder is predicted once more using the new cutoffs for the terminal size (TS, TS = 13% for *ab initio* t14) percentage of residues with the terminal ends reaching a terminal cutoff value (TCV, TCV = −0.3 for *ab initio* t14).

In this work, we utilized a variety of approaches to model structures. This is similar to the original use of *ab initio* modeling reported previously[26] for use in the RRD protocol. The modeling approaches include Rosetta *ab initio* (version 3.12),[39] RoseTTAFold (RF, version 1.0.0),[33] and Alphafold2 (AF, version 2.0.0).[32] The details of these methods are described below. We investigated RRD disorder prediction for each case using the REF2015 (r15) scoring function and the t14 scoring function. For comparison, we also used the native (crystal or solution NMR) structures to predict disorder (using the representative model). For each protein system tested, we generated 1 relaxed native model, 100 *ab initio* models, 1 RF model, and 5 AF models. The sets of each model were then relaxed in Rosetta (using the corresponding Rosetta scoring function, r15 or t14), then input into the RRD protocol.

We next updated the RRD parameters (WS, CV, TPC, TS, and TCV) to be compatible with different modeling protocols. These parameters were systematically optimized against a combination of AUC and percent accuracy (on the benchmark set) using an in-house python script. For example, because scores are lower (more negative) for the r15 scoring function in comparison with the t14 function, the CV was decreased. For each parameter selection, we optimized the parameters using the 16-protein benchmark set and later used the larger (229 proteins, both data sets are described in the following section), fully independent test set to verify the results.

The benchmark set-optimized parameters for each modeling case are shown in Table 1. For RRD using all Rosetta-based structure modeling, the same parameters were used (relaxed native, *ab initio*, and RF), while the parameters were different for the AF modeling. It should be noted that the terminal optimization procedure did not improve results for the AF modeling, so this step was disregarded for that modeling approach, as indicated in Figure 1 and Table 1 (TPC, TS, and TCV not applicable). Ultimately, we calculated the binary prediction accuracy for all residues (percentage of correct predictions) and the AUC for the ROC curve for each data set for each approach. The AUC value ranges from 0 to 1, with perfect prediction having AUC = 1 and a random prediction tool having AUC = 0.5.

Table 1. Parameters Used for RRD Predictions

method(s) for model generation	Rosetta RRD scoring function	WS	CV	TPC	TS	TCV
relaxed natives,[31] *ab initio*,[26] and RF	t14	5	−1.0	60%	13%	−0.3
relaxed natives, *ab initio*, and RF	r15	10	−1.5	40%	34%	−0.8
AF	t14	5	−1.2	NA	NA	NA
AF	r15	10	−1.8	NA	NA	NA

Data sets: Benchmark and Test Sets. In this work, we used the same two data sets that were employed in the previous RRD development. The data sets have been described in full detail previously.[26] In short, the benchmark data set contained 16 nonhomologous proteins, each with fewer than 150 residues, with experimentally identified disordered regions from the DisProt database.[40] The proteins in the benchmark data set covered a large range of the overall percent disorder (from very ordered [4.7% disorder] to very disordered [100% disordered]). Among the 16 proteins in the benchmark data set, there were 737 ordered residues and 648 disordered residues, a relatively balanced distribution between the order and disorder. The RRD parameters were optimized using the proteins in the benchmark set for the new modeling approaches in this work.

The independent test data set that was employed in previous work[26] was also used here to examine the prediction accuracy for an unbiased data set. The 229 test set proteins were obtained from the Protein Data Bank. We specifically chose structures determined with solution NMR. The percent disorder of this test set ranged from 0 to 100%. Among the 229 proteins, there were 11,172 disordered residues and 10,580 ordered residues, also balanced between ordered and disordered. The residues were identified as disordered if the root-mean-square fluctuation (RMSF) was greater than 2 Å; otherwise, they were identified as ordered.[41]

Structural Modeling for the RRD Input. To generate the 100 Rosetta *ab initio*[28,29] models, we used the same protocol described in detail from previous work.[26] Rosetta *ab initio* is a fragment-based method for protein structure prediction. In this type of modeling approach, the protein conformational space is searched by inserting fragments within a Monte Carlo sampling strategy. Therefore, a large number of sampled structures (thousands to tens of thousands) is usually required to sufficiently sample the conformational space. From our previous work, we found that generating 100 models was adequate to correctly predict disorder using our RRD approach.[26] For the RF and AF modeling, default parameters (as instructed in their respective GitHub repositories) were used. By default, AF predicts five models (that are typically very similar to one another) while RF produces one model. It is noted that the structural output of both AF and RF is rather deterministic as compared to that of Rosetta *ab initio*. For a given sequence, these methods predict similar structures from run to run. We have previously shown that adding a large ensemble of structures that are relatively similar as input into the RRD protocol often provides very little benefit (compared to a single structure). For example, when inputting relaxed natives, using the entire NMR ensemble only increased prediction accuracy by 0.6 percentage points compared to representative alone.[26] We thus decided to use five AF models and one RF model in our RRD protocol. After relaxation, the

https://doi.org/10.1021/acs.jpcb.2c05508
J. Phys. Chem. B 2022, 126, 8439−8446

scores were comparable to those of Rosetta-generated structures. For comparison purposes, C-α root-mean-square deviation (RMSD) was calculated for each model to the un-relaxed native structures using PyMol. To do this for any particular model, we calculated RMSD to each native structure (for NMR-derived structures with an ensemble of structures, there were multiple native pdb models), with the value reported being the minimum RMSD. Finally, the values reported are averages over the sets.

Applications of Other IDR Prediction Approaches. The performance of RRD using various modeling approaches was compared to five popular disordered region prediction protocols, namely, IUPred3,[23] RaptorX,[20] PrDOS,[25] Metapredict,[24] and the AF-predicted local-distance difference test (pLDDT) approach[36] (described in the following paragraph). All 245 protein sequences from both the benchmark and NMR data sets were submitted to the IUPred2A,[22] IUPred3,[23] RaptorX,[20,42] and PrDOS[25] online servers, while the Metapredict algorithm[24] was run locally. All predictions were performed with default settings. These methods predicted the probabilities of each residue being disordered, which were utilized in ROC curve analysis. For prediction accuracy comparison, a residue with disorder probability higher than 50% was identified as disordered and identified as ordered otherwise.

For the pLDDT approach, we used the five generated AF models using the default AF protocol. The pLDDT value, which acts as a confidence metric of the prediction, was extracted for each residue. This value indicates the predicted accuracy of the local structural environment of any modeled residue, ranging from 0 to 100, with 100 being the most confident. This pLDDT value has also been used to model disorder before. In the pLDDT disorder prediction approach,[36] the probability of a given residue being disordered is $1-0.01 \times$ pLDDT (as high confidence corresponds to low disorder and vice versa), resulting in a disorder probability for each residue. If the averaged disorder probability of a given residue was greater than 70, the residue is predicted as disordered.[36]

RESULTS AND DISCUSSION

In our previous work,[26] we developed Rosetta ResidueDisorder (RRD), which was designed to use 100 Rosetta *ab initio* structures modeled from the sequence to predict IDRs of a protein. The algorithm (see Figure 1) is dependent on the hypothesis and subsequent observation that residues in regions of favorable Rosetta scores are typically more ordered than residues in regions with unfavorable Rosetta scores. We then further demonstrated that replacing the 100 models with a known, relaxed native structure improved the accuracy of prediction.[31] Here, with the new innovations in structure prediction methods, through the use of deep learning and sequence coevolution, including AlphaFold (AF)[32] and RoseTTAFold (RF),[33] we sought to update the RRD methodology to take advantage of these state-of-the-art structure prediction methods. We also updated the application to use the current Rosetta score function. We demonstrate that the predictions were most accurate when using AF models as input into the RRD protocol. Additionally, we used various other structural modeling inputs as a comparison and updated RRD parameters for the different protocols. We also compared the results to those of some established disorder prediction methods.

The original RRD protocol with Rosetta *ab initio* modeling used the Talaris2014 (t14) scoring function in Rosetta. We first updated the RRD parameters (WS, CV, TPC, TS, and TCV) for use in *ab initio* modeling with Ref2015 (r15). This was needed because the typical ranges in residue scores varied between the two scoring functions. To do this, we re-optimized the parameters on the small test set (containing 16 proteins) and confirmed the predictiveness using the larger, independent benchmark set (containing 229 proteins), as shown in Table 1. Next, we used models generated using AF (5 models each) and RF (one model each) and compared to results using the methodology from previous work, Rosetta *ab initio* and relaxed native models. Prior to input into RRD, all generated models were relaxed in Rosetta. For RF, the same RRD parameters as those used for *ab initio* and relaxed native structures were used. For AF, the parameters were improved using the optimization strategy described above (note that no terminal optimization was necessary for AF modeling).

AF Produced Accurate Structural Results for both Data Sets. Prior to analyzing the disorder prediction results, we first quantified structural accuracy for the various modeling approaches as a baseline. We previously demonstrated that the more accurate relaxed native models produced better disorder predictions using the RRD score function-based method (compared to *ab initio* models[31]). Here, we calculated the average RMSD to un-relaxed natives for the structural models produced for each method in both the benchmark and test data sets: relaxed native (1 model, representative model), *ab initio* (100 models), RF (1 model), and AF (5 models). The structure prediction results matched as expected, as shown in Table 2. The relaxed natives had the lowest average RMSD,

Table 2. Structural Accuracy (Average RMSD in Å to Native) for Each Modeling Approach: Relaxed Native, *ab initio*, RF, and AF (Using the r15 Scoring Function)

modeling approach	benchmark set	test set
relaxed native	2.42 Å	4.23 Å
ab initio	8.70 Å	11.85 Å
RF	6.48 Å	5.91 Å
AF	4.96 Å	5.34 Å

followed by AF, RF, and *ab initio*, respectively. We note that high RMSDs for relaxed native models were due to the disordered proteins, for which the PDB structures were not expected to be stable; see Table S1. Notably, the AF modeling produced structures that were significantly more native-like than the previous *ab initio* modeling as evidenced by the RMSD values (4.96 versus 8.70 Å for the benchmark set and 5.34 versus 11.85 Å for the test set). Even with this improvement in the RMSD, highly disordered regions (predicted with AF) are often placed spherically around the protein or are artificially extended (examples from our data set are shown in Figure S1). These specific placements of atoms are not meant to be taken literally and thus have nearly zero physical meaning.[37] For these reasons, the resulting RMSD values (see Table 2) tend to be very high. For example, the average RMSD of AF predictions in the test set decreased from 5.34 Å overall to 1.78 Å when considering only ordered residues.

Ranking of the Residue Order via ROC Curve Analysis. To compare the effectiveness of the various modeling approaches when used as input for RRD, we first

https://doi.org/10.1021/acs.jpcb.2c05508
J. Phys. Chem. B 2022, 126, 8439−8446

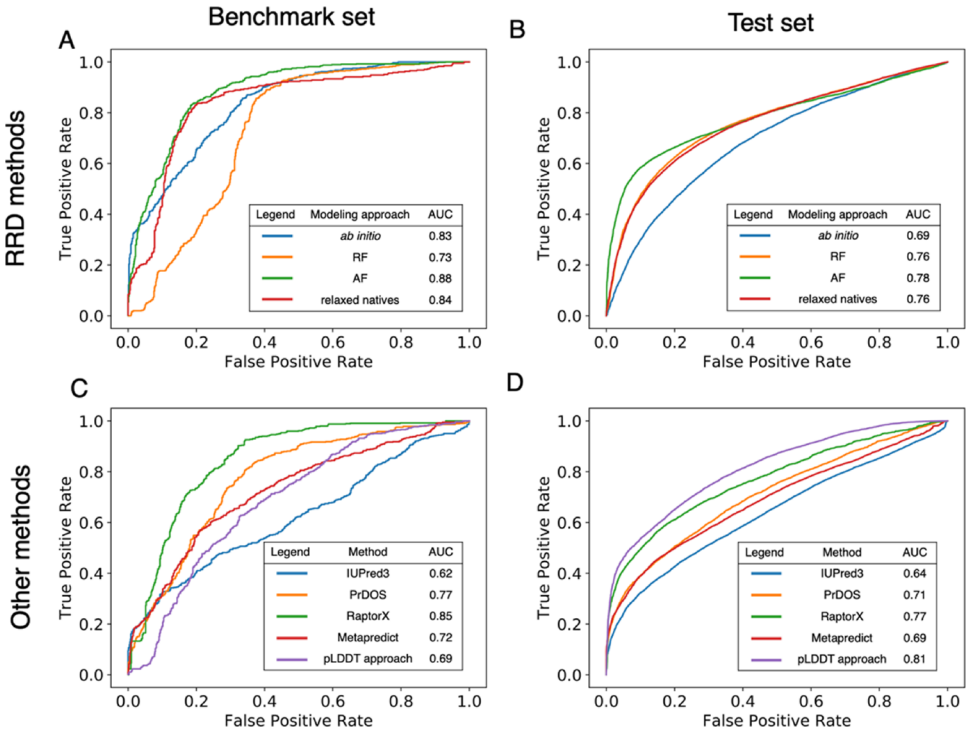


Figure 2. ROC curves for the RRD methods: relaxed native, *ab initio*, RF, and AF (A,B); and other methods: IUPred3, RaptorX, PrDOS, Metapredict, and the AF-based pLDDT approach (C,D). Data shown for both the benchmark (A,C) and test (B,D) data sets using the r15 scoring function.

examined the ranking of residues from the highest likelihood of being disordered to the highest likelihood of being ordered, based on the OS only. For this, we used ROC curves and calculated the AUC. A more accurate ranking is evidenced by the higher AUC. ROC curves for the various RRD approaches are shown in Figure 2A for the benchmark set. The data shown in the main manuscript tables and figures were calculated using the r15 score function. The native and AF had large AUC values (0.84 and 0.88, respectively), while *ab initio* and RF had slightly worse AUC values (0.83 and 0.73, respectively). While it was surprising that Rosetta *ab initio* outperformed RF, based on the structural accuracy (average RMSD of 8.70 versus 6.48 Å), we attribute this to a small sample size. Evidence for this can be found in the results from the test set, which is much larger than the benchmark set (229 versus 16 proteins); data shown in Figure 2B. AUC values for the test set were 0.76 for natives, 0.78 for AF, 0.76 for RF, and 0.69 for *ab initio*. In this larger, independent test set, RF predictably outperformed *ab initio*.

Furthermore, it was initially unexpected that AF slightly outperformed the relaxed natives for these cases. However, our results show that this was due to the inflated RMSD values for the AF models. While the AF predictions in unstructured regions should not be interpreted as structurally significant, they do provide useful information. This information was interpreted by RRD in the form of high Rosetta OSs (less negative or more positive) in these regions, which influences the prediction of disorder. For these reasons, the accurate predictions (high AUC values) for the AF-RRD combination make sense.

As a comparison, t14 results are shown in the SI; see Table S2 and Figure S2 for RMSD values and ROC curves. As there were two main improvements in the current work (use of better modeling approaches and Rosetta scoring function), we compare the t14 results to the r15 results for *ab initio*. Even before incorporating any of the new structural modeling approaches, the results improved when the r15 scoring function was used. For the benchmark set, the AUC improved from 0.76 to 0.83.

Furthermore, we compared the rankings with RRD to other disorder prediction methodology. To do this, we tested five additional methods using our data sets: IUPred3, RaptorX, PrDOS, Metapredict, and the AF-based predicted local-distance difference test (pLDDT) approach. The ROC curves are shown in Figure 2C,D for the benchmark and test sets, respectively. Apart from the pLDDT approach for the test set (AUC = 0.81), the AF-RRD approach produced an AUC that was better than the other methods (AUC = 0.88 for benchmark, AUC = 0.78 for test). It is to be noted that AF-RRD only slightly outperformed RaptorX (by a difference of 0.01 AUC units for the test set). However, compared to RaptorX (a computationally inexpensive sequence-based method), our approach additionally provides the benefit of producing a predicted 3D structure on which to visualize the prediction. These results were encouraging, but we also wanted to predict the binary order/disorder for each residue. For the RRD approach, this involves using the cutoffs and terminal optimization parameters, as described in the Methods section.

Disorder Prediction Results. Next, we used the OSs and cutoffs (see methods for full details) to predict each residue as ordered or disordered and compared the predictions to the known experimental data. Table 3 shows the percent accuracy for each prediction (corresponding to the ROC curves in Figure 2). Unsurprisingly, the binary prediction results followed the same trends as AUC-quantified rankings. Specifically, we highlight the high accuracy from the AF-RRD approach (81.8% in the benchmark set and 73.7% in the test set), which again remarkably outperformed even the

https://doi.org/10.1021/acs.jpcb.2c05508
J. Phys. Chem. B 2022, 126, 8439−8446

Table 3. Average Percent Accuracy in Disorder Prediction for Each RRD Modeling Approach (Relaxed Native, *ab initio*, RF, and AF) and the Previously Established Disorder Prediction Methods (IUPred3, RaptorX, PrDOS, Metapredict, and the AF-Based pLDDT Approach)

prediction method	benchmark set	test set
RRD: relaxed native	81.0%	70.9%
RRD: *ab initio*	76.0%	64.0%
RRD: RF	74.1%	71.7%
RRD: AF	81.8%	73.7%
IUPred3	61.6%	59.4%
RaptorX	71.9%	66.7%
PrDOS	62.7%	62.1%
Metapredict	63.5%	61.2%
pLDDT approach	59.8%	69.1%

relaxed native-RRD predictions (81.0% in the benchmark set and 70.9% in the test set).

The t14 prediction results are shown in Table S3. While new structural modeling approaches (RF and AF) certainly improved results, the choice of the Rosetta score function also played an important role. For example, when using *ab initio* (before including new modeling), the r15 results (Table 3) improved from the t14 results (Table S3) by 6.5 and 1.5 percentage points for the benchmark and test sets, respectively. We also examined whether RMSD to native corresponded with prediction accuracy for the AF-RRD approach on the individual protein level, as shown in Figure S3. For systems at less than 100% disorder, where the measure of RMSD to native is more meaningful, accuracy roughly correlated with the RMSD. All disorder predictions of 90% or greater had RMSD less than 5 Å.

Furthermore, the binary prediction results are also shown for the non-RRD methods in Table 3. While we previously highlighted one case where methods outperformed AF-RRD in the AUC, there were no cases that outperformed AF-RRD in the prediction accuracy, although the AF-based pLDDT approach in the test set was the closest non-RRD method in accuracy (69.1%). The AF-RRD results represent the highest binary prediction accuracy observed on these data sets (including when native structures were known). This is notable because this approach can be used in any case, regardless of prior knowledge of the native structure(s).

Two examples of the AF-RRD predictions are shown in Figure 3 (1Y8M and 2I4K). Shown in this figure are (i) the order score as a function of the residue number for the AF-RRD predictions, (ii) the AF model colored by predicted order/disorder, and (iii) the native model colored by experimental O/D. Based on the AF predictions, 1Y8M (29.9% disordered) had an accuracy of 86.8% (2.94 Å RMSD) and 2I4K (38.3% disordered) had an accuracy of 80.5% (4.48 Å RMSD). These examples demonstrate how RRD can identify disordered terminal ends (as was the case for 1Y8M) and a disordered loop within the chain (2I4K).

Additionally, we also calculated and compared the Mathew's correlation coefficient (MCC) for all disorder prediction approaches (Table S4). MCC explicitly accounts for all classifications (true positives, false positives, true negatives, and false negatives). Similar to what was observed with accuracy and ROC-AUC analysis of the test set, the MCC of AF-RRD outperformed all other disorder prediction methods, including IUPred3, PrDOS, RaptorX, Metapredict, and the pLDDT approach. Furthermore, one additional advantage of AF-RRD over sequence-based methods (IUPred3, PrDOS, RaptorX, and Metapredict) is that users are able to visualize the predictions on a 3D structure which could potentially be used for further calculations such as MD simulations, free energy calculations, and docking simulations.

CONCLUSIONS

This work expands the capabilities of RRD to accurately predict disordered regions of proteins from sequences. The prediction accuracy was tested on a large, independent set (test set, 229 proteins) that was well balanced between ordered and disordered residues (51% of the residues were ordered). First,

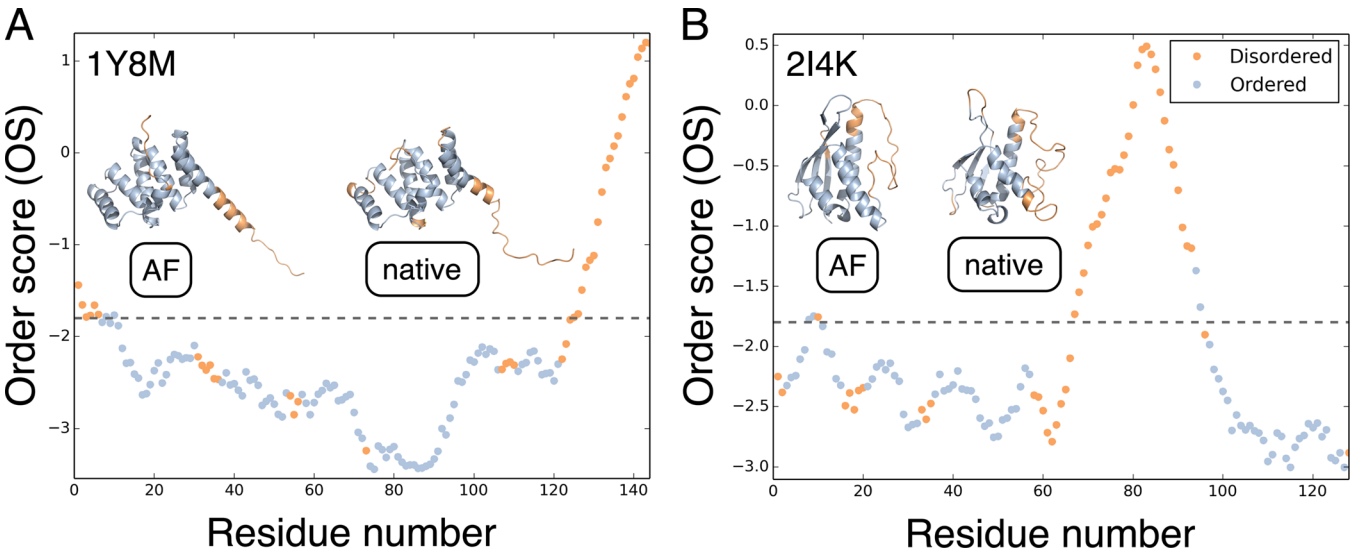


Figure 3. Prediction results for 1Y8M (A) and 2I4K (B). Each panel shows the order score as a function of residue number for the AF predictions using RRD, colored by experimental disorder (orange = disordered, blue = ordered). The gray, dotted cutoff line identifies the predictions. Points above the line are predicted as disordered and below predicted as ordered. The AF model is shown in a cartoon representation (1Y8M: 2.94 Å RMSD, 2I4K: 4.48 Å RMSD), colored by the predicted disorder along with the native structure, colored by the experimental disorder.

https://doi.org/10.1021/acs.jpcb.2c05508
J. Phys. Chem. B 2022, 126, 8439−8446

we updated parameters to utilize the REF2015 (r15) scoring function in Rosetta. When doing this, the prediction accuracy improved from 62% to 64% in the large test set when using Rosetta *ab initio* models as input into RRD. Next, we improved accuracy further by utilizing modeling approaches developed using deep learning and sequence coevolution data: AlphaFold and RoseTTAFold. Using the combined AF-RRD approach, the prediction accuracy improved from 64% with *ab initio* to 74%. These prediction accuracy results show the best binary prediction results seen on our two data sets. These results further indicate that the accuracy of RRD is dependent on the structural modeling accuracy. Additionally, AF-RRD also significantly reduced the time required for model generation over our previous method.[26] We evaluated the time required for generating models of proteins with three different sequence lengths from the test set. The results from our current work show that the model generation step was sped up by 1.2-, 3.6-, and 5.0-fold for the sequence lengths 61, 100, and 150, respectively, compared to the model generation step with Rosetta *ab initio*. We also used the same data sets to compare our prediction results to other available disorder prediction methods. While the AF-RRD approach outperformed all other tested methods in prediction accuracy, only the AF-based pLDDT approach[36] produced slightly higher AUC from ROC analysis for the test set. From these data, we conclude that the AF-RRD approach with r15 (i) significantly outperforms the previously developed usage of RRD (with *ab initio* and t14) and (ii) is at least as accurate as other available methods (if not more, specifically if the goal is binary prediction of a balanced data set). As a caveat, we note that RRD performed better on the benchmark (training) set compared to the test set. This could indicate some small amount of overfitting and suggests that training RRD on a larger training set might yield even higher prediction accuracy for the test set. Thus, for fairness, we compared the results to other methods using the larger test set. Furthermore, we deliberately chose to keep the benchmark and test data sets identical to our previous work[26] to better compare our current approach. This methodology represents another usage among many of the improved modeling capabilities available to the scientific community using AF. The benefit of our RRD modeling approach with respect to other sequence-based methods (such as RaptorX, IUPred3, Metapredict and PrDOS) is that along with accurate order/disorder predictions, users also obtain a predicted structure. While this procedure might have a slightly higher runtime compared to other methods, users obtain more total information in three simple steps of running AF for structure prediction, the Rosetta relax protocol for structure minimization, and RRD for order/disorder prediction. RRD using AF is freely available in Rosetta, and a tutorial can be found in the Supporting Information. Additionally, all figures and tables contained in this manuscript and its Supporting Information can be reproduced with data that can be found in Supplementary zip files.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c05508.

Variation of RMSD for relaxed native models over level of disorder (Table S1); structural accuracy (average RMSD in Å to native) for each modeling approach (Table S2); examples of AF-predicted structures from our data set with un-physical structural arrangements in the disordered regions (Figure S1); ROC curves for the RRD methods (Figure S2); average percent accuracy in disorder prediction for each RRD modeling approach (Table S3); MCC for both the benchmark and the test sets (Table S4); prediction accuracy as a function of RMSD to native for each protein in the test set (Figure S3); AF-RRD disorder prediction tutorial (PDF)

Files for data reproducibility (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Steffen Lindert − *Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States*; orcid.org/0000-0002-3976-3473; Phone: 614-292-8284; Email: lindert.1@osu.edu; Fax: 614-292-1685

Authors

Jiadi He − *Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States*

SM Bargeen Alam Turzo − *Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States*

Justin T. Seffernick − *Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States*

Stephanie S. Kim − *School of Biological Sciences, Seoul National University, Seoul 08826, South Korea*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c05508

Author Contributions

[§]J.H., S.M.B.A.T., and J.T.S. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank all, current and former, members of the Lindert lab for many useful discussions. We would like to thank the Ohio Supercomputer Center[43] for valuable computational resources as well as their computational support team including Summer Wang, Samuel Khuvis, Zhiqiang You along with others for their help with issues related to specific software installations and usage of the supercomputer. Additionally, we would like to thank Vikram K. Mulligan, Julia K. Leman, Rocco Moretti, along with many others in the Rosetta community for their help and guidance during the code development stage of the RDD application as well as reviewing the updated version of all code pushed into Rosetta. This research received no external funding.

■ REFERENCES

(1) Uversky, V. N. Introduction to intrinsically disordered proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6557−6560.

(2) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **2008**, *37*, 215−246.

(3) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.

(4) Tompa, P.; Davey, N. E.; Gibson, T. J.; Babu, M. M. A million peptide motifs for the molecular biologist. *Mol. Cell* **2014**, *55*, 161−169.

(5) Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guilliot, S.; Dunker, A. K. Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* **1998**, 437−448.
(6) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. Intrinsic protein disorder in complete genomes. *Genome Inf. Ser. Workshop Genome Inf.* **2000**, *11*, 161−171.
(7) Bondos, S. E.; Dunker, A. K.; Uversky, V. N. On the roles of intrinsically disordered proteins and regions in cell communication and signaling. *Cell Commun. Signal.* **2021**, *19*, 88−88.
(8) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573−584.
(9) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037−1049.
(10) Midic, U.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.; Uversky, V. N. Protein disorder in the human diseasome: unfoldomics of human genetic diseases. *BMC Genomics* **2009**, *10*, S12.
(11) Uversky, V. N.; Oldfield, C. J.; Midic, U.; Xie, H.; Xue, B.; Vucetic, S.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics* **2009**, *10*, S7.
(12) Raychaudhuri, S.; Dey, S.; Bhattacharyya, N. P.; Mukhopadhyay, D. The role of intrinsically unstructured proteins in neurodegenerative diseases. *PLoS One* **2009**, *4*, No. e5566.
(13) Cheng, Y.; LeGall, T.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* **2006**, *45*, 10448−10460.
(14) Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **2013**, *18*, 10802−10828.
(15) Chemes, L. B.; Alonso, L. G.; Noval, M. G.; de Prat-Gay, G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol. Biol.* **2012**, *895*, 387−404.
(16) Bernadó, P.; Svergun, D. I. Analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Methods Mol. Biol.* **2012**, *896*, 107−122.
(17) Nasir, I.; Onuchic, P. L.; Labra, S. R.; Deniz, A. A. Single-molecule fluorescence studies of intrinsically disordered proteins and liquid phase separation. *Biochim. Biophys. Acta, Proteins Proteomics* **2019**, *1867*, 980−987.
(18) Monastyrskyy, B.; Kryshtafovych, A.; Moult, J.; Tramontano, A.; Fidelis, K. Assessment of protein disorder region predictions in CASP10. *Proteins* **2014**, *82*, 127−137.
(19) Meng, F.; Uversky, V. N.; Kurgan, L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell. Mol. Life Sci.* **2017**, *74*, 3069−3090.
(20) Peng, J.; Xu, J. RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins* **2011**, *79*, 161−171.
(21) Dosztányi, Z. Prediction of protein disorder based on IUPred. *Protein Sci.* **2018**, *27*, 331−340.
(22) Mészáros, B.; Erdos, G.; Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329−W337.
(23) Erdős, G.; Pajkos, M.; Dosztányi, Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **2021**, *49*, W297−W303.
(24) Emenecker, R. J.; Griffith, D.; Holehouse, A. S. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* **2021**, *120*, 4312−4319.
(25) Ishida, T.; Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **2007**, *35*, W460−W464.
(26) Kim, S. S.; Seffernick, J. T.; Lindert, S. Accurately Predicting Disordered Regions of Proteins Using Rosetta ResidueDisorder Application. *J. Phys. Chem. B* **2018**, *122*, 3920−3930.
(27) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **1997**, *268*, 209−225.
(28) Bradley, P.; Misura, K. M.; Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **2005**, *309*, 1868−1871.
(29) Raman, S.; Vernon, R.; Thompson, J.; Tyka, M.; Sadreyev, R.; Pei, J.; Kim, D.; Kellogg, E.; DiMaio, F.; Lange, O.; et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* **2009**, *77*, 89−99.
(30) Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **2020**, *17*, 665−680.
(31) Seffernick, J. T.; Ren, H.; Kim, S. S.; Lindert, S. Measuring Intrinsic Disorder and Tracking Conformational Transitions Using Rosetta ResidueDisorder. *J. Phys. Chem. B* **2019**, *123*, 7103−7112.
(32) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(33) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871−876.
(34) Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **2013**, *29*, 2722−2728.
(35) Akdel, M.; Pires, D. E. V.; Porta Pardo, E.; 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 AlphaFold 2 applications. *bioRxiv* **2021** 2021.09.26.461876.
(36) Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590−596.
(37) Ruff, K. M.; Pappu, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. *J. Mol. Biol.* **2021**, *433*, No. 167208.
(38) O'Meara, M. J.; Leaver-Fay, A.; Tyka, M. D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; et al. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J. Chem. Theory Comput.* **2015**, *11*, 609−622.
(39) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031−3048.
(40) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; et al. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* **2007**, *35*, D786−D793.
(41) Wang, R. Y.; Han, Y.; Krassovsky, K.; Sheffler, W.; Tyka, M.; Baker, D. Modeling disordered regions in proteins using Rosetta. *PLoS One* **2011**, *6*, No. e22060.
(42) Wang, S.; Li, W.; Liu, S.; Xu, J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res.* **2016**, *44*, W430−W435.
(43) Ohio Supercomputer Center. **1987**. Ohio Supercomputer Center: Columbus, Ohio. http://osc.edu/ark:/19495/f5s1ph73.

https://doi.org/10.1021/acs.jpcb.2c05508
J. Phys. Chem. B 2022, 126, 8439−8446