THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 3920−3930

pubs.acs.org/JPCB

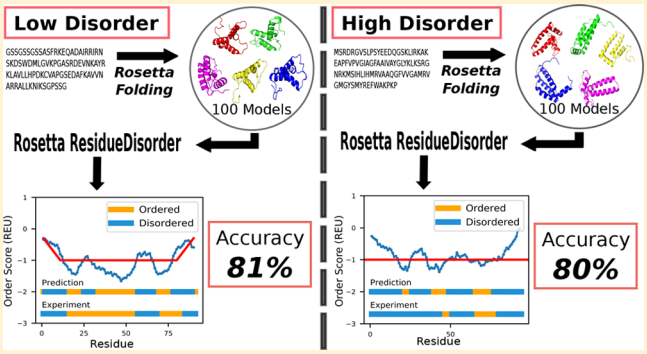
Accurately Predicting Disordered Regions of Proteins Using Rosetta ResidueDisorder Application

Stephanie S. Kim,[†] Justin T. Seffernick,[†] and Steffen Lindert*[,†]

[†]Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States

Supporting Information

ABSTRACT: Although many proteins necessitate well-folded structures to properly instigate their biological functions, a large fraction of functioning proteins contain regions—known as intrinsically disordered protein regions—where stable structures are not likely to form. Notable functional roles of intrinsically disordered proteins are in transcriptional regulation, translation, and cellular signal transduction. Moreover, intrinsically disordered protein regions are highly abundant in many proteins associated with various human diseases, therefore these segments have become attractive drug targets for potential therapeutics. Over the past decades, numerous computational methods have been developed to accurately predict disordered regions of proteins. Here we introduce a user-friendly and reliable approach for the prediction of disordered protein regions using the structure prediction software Rosetta. Using 245 proteins from a benchmark data set (16 DisProt database proteins) and a test data set (229 proteins with NMR data), we use Rosetta to predict the global protein structures and then show that there is a statistically significant difference between Rosetta scores in disordered and ordered regions, with scores being less favorable in disordered regions. Furthermore, the difference in scores between ordered and disordered protein regions is sufficient to accurately identify disordered protein regions. As a result, our Rosetta ResidueDisorder method (benchmark data set prediction accuracy of 71.77% and independent test data set prediction accuracy of 65.37%) outperformed other established disorder prediction tools and did not exhibit a biased prediction toward either ordered or disordered regions. To facilitate usage, a Rosetta application has been developed for the Rosetta ResidueDisorder method.



1. INTRODUCTION

Historically, proteins and protein domains have been viewed as stable, rigid, and well-defined three-dimensional structures. This notion was corroborated by the elucidation of more than 100,000 X-ray crystal and nuclear magnetic resonance (NMR) structures over the last decades.[1,2] From this data, the classical structure−function paradigm has emerged, relating biological functions of proteins to the amino acid sequence. The widely accepted theory is that the sequence uniquely encodes an energetically stable three-dimensional folded structure which uniquely encodes a specific biological function.[3] Although many proteins necessitate well-folded structures to properly instigate their biological functions, we now know that a large fraction of functioning proteins contain regions where stable structures are not likely to form.[4] These segments are known as intrinsically disordered regions (IDRs), intrinsically disordered protein regions (IDPRs), or also referred to as intrinsically disordered proteins (IDPs). These proteins do not exist as stable, rigid, globular, well-defined folded structures. Instead they exist as dynamic ensembles, where in equilibrium, many diverse folded structures exist because of their rather flat free energy surfaces. IDPs have unique sequence compositions, particularly low frequencies of bulky hydrophobic amino acids, and high frequencies of charged and hydrophilic amino acids.[5] IDRs and IDPs are surprisingly common.[6] It is estimated that over 40% of human proteins contain disordered segments of >30 amino acids in length, which are frequently associated with various signaling and regulatory biological functions.[7] However, due to the absence of structural constraints, identifying diverse functional roles of IDRs and IDPs based on structural homology has been a great challenge. In fact, 6.4% of protein-coding human genes' functional roles are unknown, and most of these genes with unknown functional roles contain IDRs.[4] Notably, IDPs have functional roles in transcriptional regulation, translation, and cellular signal transduction.[6] IDPs also play a major role biomedically. Many proteins involved in numerous human diseases contain regions of intrinsic disorder. For this reason, IDPs have become increasingly important in drug discovery, where small molecules are used to selectively bind to proteins, which change the free energy landscape. Several successful studies have shown small molecules successfully binding to proteins and inhibiting protein−protein

Received: February 20, 2018
Revised: March 18, 2018
Published: March 29, 2018

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

interactions involving IDPs.[3] Despite their importance, structure elucidation of IDPs is extremely challenging. Among experimental methods used for IDP structure elucidation, NMR spectroscopy has been one of the most successful. NMR has been used to determine structural information for highly flexible regions under near-physiological conditions at atomic resolution.[8] The discovery of disordered proteins' active participation in various biological functions and potential new biological functions linked to native disorder have motivated the development of computational methods for detecting disordered regions. These predictive methods have become valuable tools for uncovering biological roles and identifying protein interaction of native IDRs.[9] Thus, being able to accurately predict intrinsically disordered regions is important in characterizing protein structure and has the potential to lead to discovery of novel biological functions with deeper understanding of IDP's function.

The increasing functional significance of intrinsically disordered proteins prompted development of efficient computational methods that accurately predict disordered regions of proteins.[9,10] Over the past decades, numerous computational methods (such as scoring-function-based, structure-based, machine-learning-based, and meta predictors) have been developed to enhance the efficacy and accuracy of prediction.[9,10] The majority of the more recent prediction tools are either machine learning-based or meta predictors.[10] IUPred is a well-known prediction tool that estimates the energy of inter-residue interactions to characterize residue property.[11] The energetic contributions of favorable amino acid pairings are approximated by the statistics collected from a database of structured proteins. IUPred predicts the disordered residues of a protein sequence by calculating the energetic contribution of predefined sequential neighborhood residues. As a result, IUPred's algorithm characterizes residues as ordered or disordered without the actual conformation of a protein. PrDOS, a structure-based method, utilizes a local amino acid sequence predictor as well as templates of homologous proteins for the prediction of disordered residues. Its final prediction combines the machine learning model with a template-based approach.[12] DISOPRED is another method that utilizes machine learning techniques to predict disordered regions of proteins. After performing a PSI-BLAST[13] search over a database of sequences, DISOPRED encodes the information on each residue based on a window size of 15 residues. DISOPRED used a support vector machine learning technique to train the neural network based on disorder data from high resolution crystal structures.[14] PONDR also utilizes neural network machine learning methods, predicting disordered regions based on disorder propensities derived from protein sequences of IDPs.[15−17] PONDR VL3-H[18] searches the protein databank for homologous sequences (identified using PSI-BLAST). In an attempt to reduce predictor bias, the neural network was trained using a diverse training set, excluding sequences with >90% homology, which contained both ordered and disordered regions. As a result, PONDR was the first method to predict various lengths of IDRs with similar prediction accuracy.[10,19] Finally, meta-predictors, such as Meta-Disorder[20] and MFDp2,[21] combine results generated by several individual prediction methods as inputs to repredict disorder.[10] Despite the successes of the aforementioned prediction tools, there remain shortcomings of existing methods. Since most prediction tools utilize a local or sliding sequence window to make the disorder predictions, they may make substantial mistakes by over- or under-predicting the overall amount of disorder in the sequence,[22] resulting in average absolute errors ranging between 15% to 40%.[23,24] Moreover, depending on the algorithms and training sets utilized by different prediction tools, disorder predictors might predict certain ambiguous regions of IDPs to be totally disordered or totally ordered.[22] Most importantly, existing methods are either disregarding global protein structure completely or require the presence of a structural homologue. This might severely limit the prediction accuracy of disordered regions for proteins that exhibit low sequence similarity to proteins in the protein data bank. In fact, we are hypothesizing that contacts of residues that are distant in sequence reduce the disorder considerably. If no homologues are available, *de novo* protein structure prediction tools have the ability to probe these contacts and should be more accurate than sequence-based methods, which cannot account for sequence-distant residue contacts.

One such *de novo* protein structure prediction tool is Rosetta. The Rosetta software suite contains numerous computational methods and algorithms that can be used for a variety of tasks to model structures of proteins and nucleic acids. Some of the most commonly used methods are protein structure prediction (protein folding), protein−protein docking, small molecules interacting with proteins, and protein design.[25,26] While originally developed for *de novo* protein structure prediction,[27−29] Rosetta has more recently been used to solve structural problems in cryo-EM-guided structure prediction,[30−35] protein design,[36−40] protein−protein docking,[41−43] protein−small molecule docking,[44−46] and protein-nucleic acid structure prediction.[46,47] The Rosetta *de novo* protein structure prediction is a Monte Carlo process, where sampling is steered by the Metropolis criterion. Perturbations in structure are scored using the Rosetta energy function, which originally only accounted for residue environments and pairwise residue interactions using statistical potentials derived from the Protein Data Bank. Since then, additional low-resolution, centroid (packing, hydrogen-bonding, secondary-structure, and van der Waals interactions) and high-resolution, all-atom (atomic packing, orientation-dependent hydrogen-bonding, pairwise-additive implicit solvation model, statistically derived electrostatics, Lennard-Jones potential and backbone-dependent rotamer conformations) terms have been added.[48] Rosetta has been demonstrated to be one of the most reliable and accurate protein structure prediction tools.[49−52] Considering its strength in protein structure prediction, we hypothesized that the Rosetta energy function in combination with its Monte Carlo sampling would also be useful in prediction of intrinsically disordered protein regions. Rosetta's ability to directly sample and accurately score long-sequence-range structural contacts has the potential to make it more accurate than purely sequence-based methods, which cannot account for sequence-distant residue contacts. This idea of using Rosetta in modeling disordered regions in proteins has been explored previously.[53] Wang et al. tested a pair of approaches using Rosetta to predict structures of intrinsically disordered regions. The focus of this work was the structural modeling of disordered regions and the paper introduced two different approaches. The second approach relied on using external tools to predict the extension of these disordered regions; however, as part of the first approach, a method was suggested to predict the extension of disordered regions from within Rosetta.[53] This involved developing a free energy function, optimized using the

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

predicted lowest energy models, to identify disordered regions which required enumeration of possible ordered/disordered assignments. This proved to be helpful in predicting both disordered segments at termini and internal loops.

In this work, we explore an easier and more reliable approach to use Rosetta for the prediction of disordered regions in proteins. Since the Rosetta energy function accounts for a variety of features of ordered protein structures (such as amino acid packing, hydrogen-bonding, secondary-structure, and van der Waals interactions), we hypothesized that Rosetta per-residue scores would be lower (i.e., more favorable) in ordered regions than in intrinsically disordered protein regions. Here we show that Rosetta scores are indeed lower in ordered regions as compared to disordered regions. Using a benchmark and a large test data set, we show that the difference in scores between ordered and disordered protein regions is sufficient to accurately predict the disorder of protein regions. The accuracy of the Rosetta ResidueDisorder method is higher than that of other disorder prediction tools and does not exhibit a bias for prediction of either ordered or disordered regions.

2. METHODS

Using Rosetta *ab initio* folding (with energy function Talaris2014[54]), 100 *de novo* models were generated for each protein of interest. Subsequently, the average residue scores of the generated 100 models were calculated for each residue in the protein and smoothed over a window of length $w = 11$ residues, in order to reduce the noise of adjacent residue scores. Residues were predicted as disordered if the order score (defined as the window-averaged Rosetta residue score) was greater than or equal to a cutoff value of $E_{o/d} = -1.0$ REU. Both the residue window length w and the residue score cutoff $E_{o/d}$ were optimized using a benchmark protein set (see below). Additionally, to optimize the prediction accuracy of terminal residues, a sloped cutoff line was applied to proteins with less than 60% predicted disordered regions at the N- and C-termini (as opposed to a constant score cutoff value $E_{o/d}$). The fraction of predicted ordered residues was determined by applying the horizontal linear cutoff line ($E_{o/d} = -1.0$ REU) to the entire protein sequence prior to the application of terminal optimization. Finally, the percent accuracy of the prediction was calculated by the total number of correctly predicted residues (true positives) divided by the total number of residues across all benchmark set proteins.

2.1. Benchmark Set. A benchmark data set was assembled to optimize the parameters and performance of our developed Rosetta ResidueDisorder methodology to predict disordered protein regions. For the benchmark set, proteins from the DisProt database[19] were selected since it contained a large set of small to medium-sized proteins with various lengths of disordered regions. Since Rosetta is known to accurately predict the structure of small to medium sized protein from their sequence, proteins with fewer than 150 residues were chosen to maximize the accuracy of the predicted models. The resulting size-filtered proteins were categorized based on their fraction of disordered residues. For example, "0% disorder" category proteins contained disordered fractions ranging from 0 to 10% of the sequence length. To generate a benchmark set that comprised a wide range of disordered and ordered proteins, nonhomologous proteins from the 0%, 30%, 50%, 70%, and 100% disorder categories were selected. The DisProt database provided all the disordered residues of a protein that were measured with various experimental methods (solution NMR, X-ray crystallography, and circular dichroism). If a protein's disordered regions were determined using multiple experimental methods, DisProt database combined all of the individual experimental data and merged the fragmented disordered regions. In order to create a benchmark set of proteins that did not contain a biased proportion of disordered and ordered proteins as well as residues, the sum of total disordered residues and total ordered residues within the potential benchmark set was compared, and a subset of proteins was picked to ensure that the difference between the total disordered and ordered residues in the benchmark set were within 10% of each other. As a result, our benchmark set contained 16 diverse proteins comprising a total of 648 disordered and 741 ordered residues (Table 1).

Table 1. Benchmark Dataset Proteins[a]

protein	IDP %	disordered	ordered
DP00344	4.69	7	142
DP00741	7.58	5	61
DP00512	15.79	21	112
DP00180_C003	19.48	15	62
DP00289	28.15	29	74
DP00685	31.91	30	64
DP00641	31.96	31	66
DP00644	32.29	31	65
DP00288	42.86	33	44
DP00475	67.24	39	19
DP00201	70.37	76	32
DP00179	100	68	0
DP00293	100	64	0
DP00005	100	107	0
DP00148_C004	100	55	0
DP00004_C002	100	37	0
SUM		648	741

[a]List of 16 benchmark data set proteins selected from the DisProt database. Protein name is the DisProt database ID, "IDP %" represents the calculated fraction of disordered regions of each protein, and the "disordered" and "ordered" columns show the numbers of disordered and ordered amino acids of each protein.

2.2. Statistical Analysis of the Benchmark Set. The order scores of each residue in all 16 proteins were calculated and classified into ordered and disordered categories according to the experimental residue assignment. A P-value test was conducted on the two independent sets of disordered and ordered residues of the benchmark data set to measure the probability that the difference between Rosetta per-residue scores of disordered and ordered residues was not statistically significant (null hypothesis). The P-value of the two independent sets of disordered and ordered residues of the benchmark set was calculated with the t test tool of R.[55] Welch's t test was applied as the two samples had unequal variances and unequal sample sizes.[56]

2.3. Optimization of Prediction Parameters using Benchmark Set. Average per-residue Rosetta scores were calculated based on 100 *de novo* models for each protein. We used the benchmark protein set to optimize the window size (w) and score cutoff ($E_{o/d}$) parameters to maximize the accuracy of the prediction of ordered/disordered protein regions. The order score was calculated by averaging the residue scores over a window size of w. For example, when using a window size of 11 residues, the Rosetta residue score of

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

residue N was added to the Rosetta residue scores of 10 adjacent residues (5 N-terminal residues (N-1, N-2, N-3, N-4, N-5) and 5 C-terminal residues (N+1, N+2, N+3, N+4, N+5)) and the total sum was divided by the window size (w = 11). If the adjacent residues were less than $w/2$ residues from the N- and C-terminus of the protein, the window averaging only occurred over existing residues. For example, the adjacent residues for residue 2, used for the calculation of the order score, were residues 1, 3, 4, 5, 6, and 7. For the optimization of the window parameter, five different window sizes were tested (w = 1,3,5,7, and 11 residues). Depending on whether the order score was below or above a score cutoff $E_{o/d}$, residues were predicted as ordered or disordered, respectively. To find the cutoff score $E_{o/d}$ that most accurately predicted the disordered residues in the benchmark protein set, 40 different cutoff line values ranging from $E_{o/d}$ = 0.0 to −4.0 REU (with an increment of 0.1 REU) were tested.

2.4. Optimizing Terminal Residue Prediction. In order to optimize the prediction accuracy of terminal residues, a sloped cutoff line was applied at the N- and C-termini (as opposed to a constant score cutoff value $E_{o/d}$). To determine the optimal slope of the cutoff line, two components of the slope were optimized: the fraction of terminal residues for which a sloped cutoff was applied, and the score cutoff value at the termini. Terminal residue fractions, ranging from 0% to 15% of the protein sequence at both the N- and C-terminus, and $E_{o/d}$ score cutoff values at the termini, ranging from −1.0 REU to 1.0 REU (with an increment of 0.1 REU), were tested. Additionally, the sloped cutoff line was only applied to proteins with a minimum fraction of ordered residues. Minimum fractions, ranging from 0% to 100% disordered residues (with an increment of 10%), were investigated. The fraction of ordered residues was determined by applying the horizontal linear cutoff line ($E_{o/d}$ = −1.0 REU) to the entire protein sequence and calculating the fraction of residues with order scores below the score cutoff value. Terminal optimization was applied to proteins with less than 60% disordered regions and the sloped cutoff was used for the 13% most terminal residues at both the N- and C-terminus. The terminal $E_{o/d}$ score cutoff value at both the N- and C-terminus was −0.3 REU.

2.5. Test Set. In order to evaluate the unbiased performance of the developed Rosetta ResidueDisorder methodology to predict disordered protein regions, it was applied to a test data set that was distinct from the benchmark set of proteins. This also ensured that it was not overfitted to the benchmark set. For the test set, proteins from the protein data bank (PDB)[57] were assembled with the following selection criteria: (a) fewer than 150 residues, (b) single chained, and (c) structure verified by solution NMR experimental method. Proteins with larger than 70% sequence homology and proteins containing heteroatoms in their structure were excluded. As opposed to the DisProt database, the PDB database did not contain information on residue disorder. We defined disordered protein residues/regions as those that are associated with high structural variability in the NMR models, as is commonly done in the literature.[9,10,53] Residues with root-mean-square fluctuation (RMSF) greater than a threshold of 2 Å were considered as disordered. Following the same procedure applied to the benchmark set, the test set proteins were categorized based on their fraction of disordered residues, again focusing on 0%, 30%, 50%, 70%, and 100% disorder category proteins. To ensure a balanced number of disordered and ordered residues in the test set, a subset of proteins was picked to ensure that the difference between the total disordered and ordered residues in the test set were within 1% of each other. As a result, the test set contained 229 diverse proteins comprising a total of 10912 disordered and 10836 ordered residues (Table S1).

2.6. Application of other Methods for Prediction of Disordered Residues. The prediction performance of our method (Rosetta ResidueDisorder) was compared to six established protein disorder prediction tools that are based on different prediction methods. IUPred, PrDOS, PONDR VL3-H, DISOPRED, MFDp2, and Meta-Disorder were chosen as exemplary prediction tools for each of the four main types of disordered prediction methods used in the field (Score-based, structure-based, machine learning, and meta predictors). The sequences of all 245 proteins of the benchmark and test data sets were submitted to the IUPred, PrDOS, PONDR VL3-H, DISOPRED, MFDp2, and Meta-Disorder servers. The prediction tools generated output files that contained binary annotation of disordered and ordered (D and O), accompanied by confidence values ranging from 0 to 1 (Meta-Disorder values ranged from 0 to 9), indicating the prediction certainty for each predicted residue. The confidence value of each residue was used when evaluating the prediction performances by generating a receiver operating characteristic curve (ROC curve). For the ROC curve, residues were sorted by their confidence values. Most of the prediction servers defined a residue with confidence value greater than 0.5 as disordered. Thus, the ROC curve illustrates how well the confidence value represents the true positive (disordered residue) and the true negative (ordered residue) metric.

2.7. Analysis of Structural Consensus of the 100 *Ab Initio* Models of the Test Set Proteins. The structural consensus of the 100 *ab initio* models was analyzed to evaluate the structural alternative to the score-based prediction of the Rosetta ResidueDisorder method. The root-mean-square fluctuation (RMSF) of the aligned *ab initio* structures was utilized to assess the prediction of disordered residues. Residues with RMSF values greater than 2 Å were defined as disordered, which was the same criterion used to determine disordered regions in the NMR structures. The percent accuracy of the prediction was calculated by the total number of correctly predicted residues (true positives) divided by the total number of residues across all test set proteins.

2.8. Rosetta Application to Automate Prediction of Disorder. To automate the prediction process and to make the method user-friendly, a Rosetta application was developed. After independently generating the 100 *de novo* structures as previously described, the application was developed to calculate and output the residue-resolved prediction of order/disorder as well as the order score. The application requires a text file with the filenames of the 100 *de novo* structures (separated by new lines) as input and subsequently outputs the predictions using the previously described algorithm. A detailed description of how to use the Rosetta ResidueDisorder application to obtain the residue-resolved order/disorder prediction has been included in the Appendix S2.

3. RESULTS AND DISCUSSION

The Rosetta *de novo* protein structure prediction algorithm is a Monte Carlo process where structures are sampled using the Metropolis criterion. Generated structures are scored by the Rosetta energy function, which accounts for, among other factors, residue environments, residue-pair interactions, hydro-

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

gen-bonding, secondary-structure, and van der Waals interactions.[48] Since the Rosetta energy function explicitly accounts for features of ordered protein structures, we hypothesized that Rosetta per-residue scores would be lower (i.e., more favorable) in ordered regions than in intrinsically disordered protein regions and could thus be used to identify disordered protein regions. Additionally, unlike other sequence-based or amino acid property-based disordered prediction tools, Rosetta analyzes the predicted tertiary structure of a protein when scoring each residue of a model. By building three-dimensional protein models, Rosetta can sample and score potential long-range contacts (contacts between residues that are far apart in sequence). Long-range contacts and a resulting high contact order are a trademark property of ordered, stable proteins.[58] Since *de novo* sampling of long-range contacts is not part of other disordered region prediction tools, we hypothesized that a Rosetta-based prediction of disordered regions would be favorable in terms of prediction accuracy. Using large benchmark and test data sets, here we show that there is a statistically significant difference between Rosetta scores in disordered and ordered regions (scores are less favorable in disordered regions) and that the difference in scores between ordered and disordered protein regions is sufficient to accurately identify disordered protein regions. The Rosetta ResidueDisorder accuracy is higher than that of other disorder prediction tools.

3.1. Ordered Protein Regions Exhibit More Favorable Rosetta Scores than Disordered Protein Regions. The hypothesis underlying this work was that Rosetta per-residue scores would be lower (i.e., more favorable) in ordered protein regions than in disordered protein regions given that the Rosetta energy function is a weighted sum of terms that account for features of protein stability. To test this hypothesis, we built *de novo* models for a set of 16 proteins (benchmark data set) with experimentally verified disorders ranging from 4.7% to 100%. The benchmark data set was designed to represent all levels of protein disorder and contained four proteins in the 0% disorder category, four proteins in the 30% disorder category, one protein in the 50% disorder category, two proteins in the 70% disorder category, and five proteins in the 100% disorder category. 100 independent structural models were built for each of the 16 proteins in the benchmark data set. The average total Rosetta score for each of the proteins was calculated and normalized by the number of residues in the protein. Figure 1 shows a positive correlation (R^2 = 0.64889) between the fraction of disordered residues in a protein and size-normalized Rosetta residue scores. This suggests that as the fraction of disordered residues of a protein increases, the overall size-normalized Rosetta scores also increase (i.e., become less favorable). This data suggested that Rosetta scores have the potential to be used to predict the overall disorder of proteins. Furthermore, the trend shown in Figure 1 is in support of our hypothesis as proteins with a larger fraction of disordered residues indeed score worse than the proteins with smaller fraction of disordered residues. However, Figure 1 only compares the total protein disorder fraction to the total Rosetta score of a protein and does not address the ordered/disordered classifications of individual residues or their respective per-residue Rosetta scores. In order to explore whether the above trend extended to individual residues as well, the order score (window-averaged Rosetta per-residue score) for each residue in each of the 16 proteins was calculated and classified into ordered and disordered categories according to the experimental residue assignment. Figure 2 shows violin plots of the order score distributions for the 741 ordered residues and 648 disordered residues in the 16 proteins of the benchmark data set. The order scores of the ordered residues range from −1.78 to −0.17 Rosetta energy units (REUs) with an average of −1.08 REUs, while the order scores of the disordered residues range from −1.68 to 0.056 REUs with an average of −0.75 REUs. The difference between Rosetta order scores of disordered and ordered residues was statistically significant ($P = 2.2 \times 10^{-16}$). Based on the *t* test, we concluded that the Rosetta order scores of disordered and ordered residues can be exploited as a marker that distinguishes the two, and we thus explored whether these scores can be utilized predicting disordered regions of a protein.

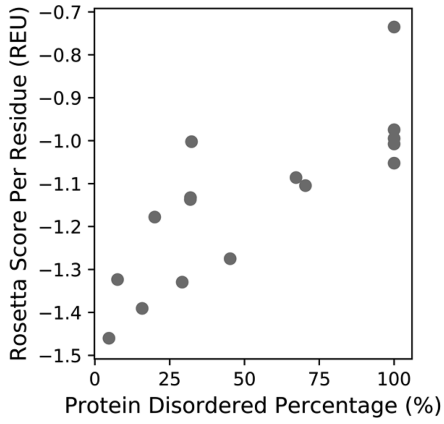


Figure 1. Correlation between degree of disorder and size-normalized per-residue Rosetta score. A positive correlation is observed between the degree of disorder (fraction of disordered residues in each protein) of 16 benchmark data set proteins and Rosetta score per residue of each protein. As the fraction of disordered residues increases, the size-normalized per-residue Rosetta score also increases (i.e., becomes less favorable).

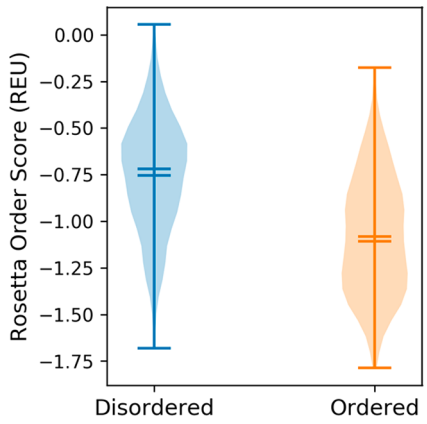


Figure 2. Comparison of individual Rosetta order scores of disordered and ordered residues. This figure compares the distributions of the order score (defined as the window-averaged per-residue Rosetta scores) of all disordered and ordered residues in the 16 protein benchmark data set. The two extreme points (maximum and minimum), mean, and the median are illustrated as tick marks.

3.2. Rosetta Score Can Be Used to Accurately Predict Disordered Protein Regions. To utilize the Rosetta order scores to predict disordered and ordered regions of a protein, a score cutoff had to be identified that separated disordered and

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

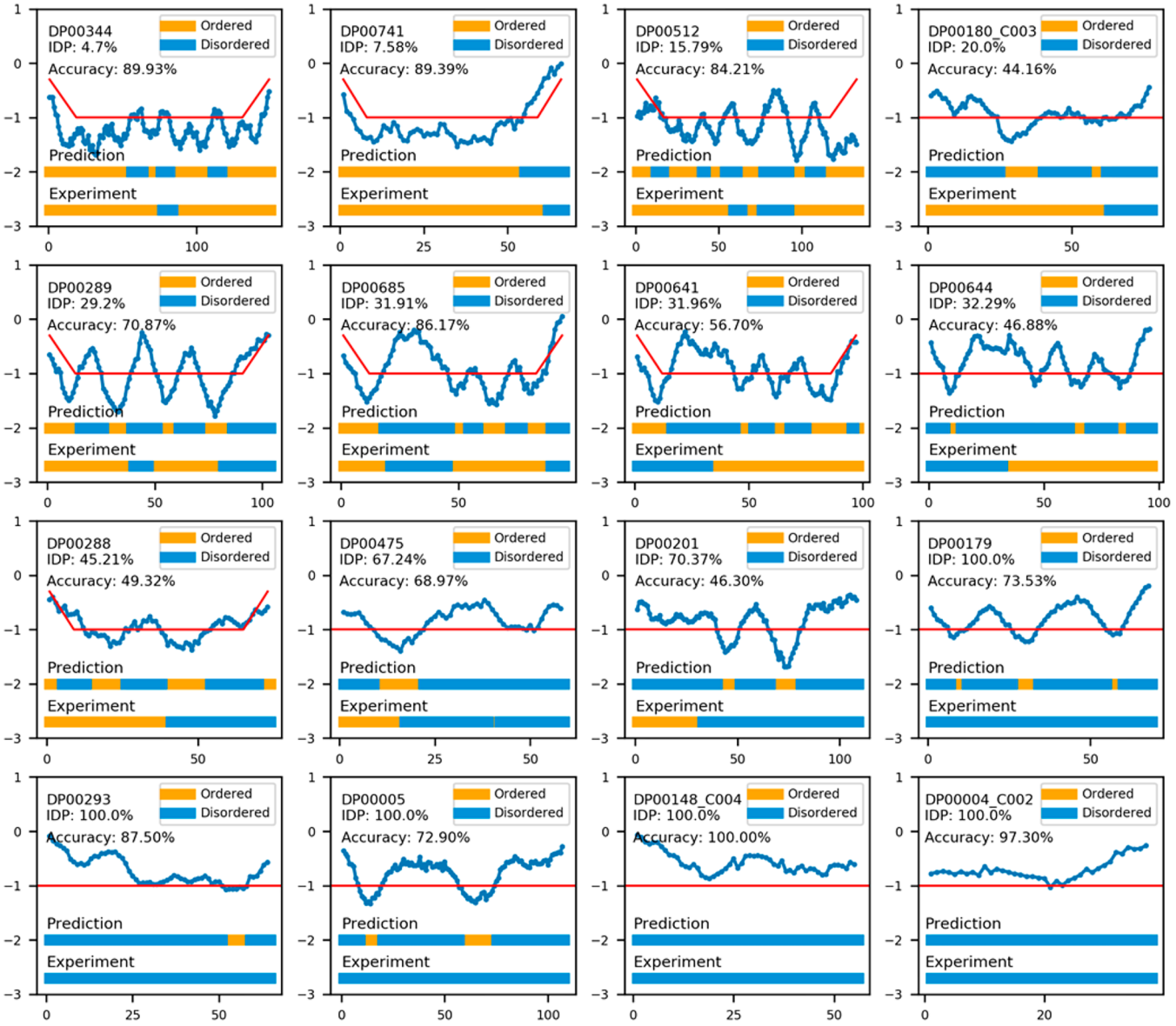


Figure 3. Optimization of terminal residues prediction of the benchmark data set. Rosetta ResidueDisorder disorder predictions of all 16 benchmark data set proteins are shown. The blue data points are Rosetta order scores calculated at a window size of 11 residues. Residues with order scores above the cutoff line (red line) are predicted as disordered residues, while residues with order scores below the cutoff line are predicted as ordered residues. The cutoff line is slopped for terminal residues in proteins with less than 60% predicted disordered residues using a flat cutoff line at −1.0 REU. The cutoff values are increased for the terminal 13% of the protein sequence with a maximum cutoff of −0.3 REU.

ordered residues most distinctly. Residues with order scores below (i.e., more favorable than) the cutoff would be labeled as ordered, while residues with order scores above (i.e., less favorable than) the cutoff would be labeled as disordered. Using the residue disorder assignments of the 16 protein benchmark data set, we optimized two parameters (window size and score cutoff) to enhance the prediction accuracy. A total of five different window sizes and 40 different score cutoff values were evaluated and the parameter combinations were ranked by the prediction accuracy across all benchmark proteins. A window size of 11 residues and a score cutoff value of −1.0 REUs were identified as the best parameters and yielded a prediction accuracy of 68.75% for correctly predicting the identities of ordered and disordered residues. Figure S1 shows the results of the prediction of ordered and disordered residues in the benchmark data set using the optimized parameters. Residues with order scores below −1.0 REUs are predicted as ordered while residues with order scores above −1.0 REUs are predicted as disordered. For comparison, Figure S1 also shows the locations and ranges of the experimentally determined disordered regions for all 16 proteins and compares them to our predictions. This illustrates that Rosetta order scores can be used to accurately predict disordered protein regions of various length. As can be seen in Figure S1, the Rosetta order scores of protein terminal residues were generally higher (less favorable) than nonterminal residues. As such, our method generally predicted those residues as disordered. Interestingly, a similar trend has been observed for other prediction methods as well.[9,10] Generally predicting terminal residues as disordered, independent of whether they are actually ordered or disordered, negatively affects the prediction accuracy of ordered proteins (0−50% disorder). We inves-

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

tigated whether the prediction accuracy for terminal residues can be improved by accounting for this effect and using increased score cutoff values for terminal residues. Instead of applying a horizontal linear cutoff value of −1.0 REUs for the determination of disorder of terminal residues, we implemented a sloped cutoff value at both N and C protein termini for proteins that exhibited at least 40% ordered residues in the original (constant cutoff value) prediction. Figure 3 shows the results of the corrected prediction. Using this approach, the prediction accuracy for the benchmark data set improved to 71.77%. Additionally, the prediction accuracy of terminal residues (i.e., those residues that are within 13% of the sequence on each the N- and C-terminal end) in the benchmark data set improved from 63.29% to 73.99%.

3.3. Rosetta ResidueDisorder Outperforms Other Prediction Tools for Benchmark Data Set. To evaluate the performance of our proposed method (referred to as Rosetta ResidueDisorder), one or two exemplary prediction tools from each of the four main types of disorder prediction methods were selected for comparison. Those included IUPred (scoring-function-based method), PrDOS (structure-based method), PONDR VL3-H (machine-learning-based method), DISOPRED (machine-learning-based method), MFDp2 (meta predictor), and Meta-Disorder (meta predictor). The sequences of the 16 benchmark proteins were submitted to each of the six established disorder prediction tools. To compare the performances of the individual methods, the prediction accuracy and the area-under-the-curve (AUC) values of ROC curves were calculated. Table 2 summarizes the results. The

Table 2. Comparison of Prediction Accuracies of Different Prediction Tools (Benchmark and NMR Datasets)[a]

	prediction accuracy (%)	
prediction tools	benchmark	NMR
Rosetta ResidueDisorder	71.77	65.37
IUPred	61.81	60.24
PrDOS	62.53	62.42
PONDR VL3-H	61.59	57.25
DISOPRED	63.75	59.46
MFDp2	60.43	64.87
Meta-Disorder	49.10	57.05

[a]This table shows the prediction accuracy of the benchmark and the test data set for six different prediction tools.

Rosetta ResidueDisorder method had the highest accuracy (71.77%), followed by DISOPRED (63.75%), PrDOS (62.53%), IUPred (61.81%), PONDR VL3-H (61.59%), MFDp2 (60.43%), and Meta-Disorder (49.1%). Figure 4 illustrates the average prediction accuracies in five different IDP categories (0%, 30%, 50%, 70%, and 100% disorder) for all seven different prediction tools (Rosetta ResidueDisorder, IUPred, PrDOS, PONDR VL3-H, DISOPRED, MFDp2, and Meta-Disorder servers). Standard deviations from the mean values are illustrated as error bars on the bar graphs. One trend observed in Figure 4 is that unlike the Rosetta ResidueDisorder method, other prediction tools tended to show biased prediction accuracy toward either ordered or disordered proteins. For example, IUPred exhibited a high accuracy for predicting ordered proteins (proteins in the category of 0% and 30% disordered, respective prediction accuracy = 74%, and 69.21%), whereas it performed poorly on heavily disordered (proteins in the category of 70% and 100% disordered, respective prediction accuracy = 65.72%, and 31.61%). A similar trend was observed for PrDOS, DISOPRED, and MFDp2 as well. The PrDOS average prediction accuracies of proteins in the category of 0% and 100% disorder were 79.31% and 45.21%, respectively. The DISOPRED average prediction accuracies of proteins in the category of 0% and 100% disorder were 83.09% and 53.38%, respectively. Finally the MFDp2 average prediction accuracies of proteins in the category of 0% and 100% disorder were 62.54% and 40.03%, respectively. Meta-Disorder and PONDR VL3-H exhibited a higher accuracy for correctly predicting disordered regions and performed worse on ordered protein regions. The Meta-Disorder average prediction accuracies of proteins in the category of 0% and 100% disorder were 18.92% and 75.72%, respectively, while the PONDR VL3-H average prediction accuracies of proteins in the category of 0% and 100% disorder were 49.10% and 81.65%, respectively. Compared to the biased trend of the known prediction tools, Rosetta ResidueDisorder showed consistent prediction accuracy throughout the wide variety of disordered and ordered proteins. The average prediction accuracy of proteins in the category of 0%, 30%, 70%, and 100% disorder were 69.55%, 70.41%, 69.37%, and 79.33%, respectively. The differences between each disordered category's average prediction accuracy were less than 10%. Using a balanced benchmark data set enabled the Rosetta ResidueDisorder method to predict disordered regions in proteins with various degrees of disorder with approximately consistent prediction accuracy.

In order to evaluate the ability of the Rosetta ResidueDisorder method to assess confidence of its predictions, a ROC curve analysis was performed and the area-under-the-curve (AUC) was calculated. The ROC curve was generated by combining and ranking residue predictions of all benchmark proteins. Residue predictions were ranked by their order scores, assigning a higher confidence to predictions further away from the score cutoff line (−1.0 REU). A similar analysis was performed for all other prediction tools. Figure 5 shows the ROC curves of all seven prediction tools. The Rosetta ResidueDisorder method and PrDOS exhibited the highest AUC (0.774) among all prediction tools. This suggests that Rosetta ResidueDisorder did not just predict ordered and disordered residues more accurately than the other prediction tools, but it also more accurately ranked them according to the Rosetta order score. As shown in Figure 5, PrDOS performs as well as the Rosetta ResidueDisorder method at ranking the true positive (i.e., disordered) residues by their confidence values. However, Rosetta ResidueDisorder (71.77% accuracy) outperforms PrDOS (62.53% accuracy) at accurately predicting disordered regions.

3.4. Analysis of Test Data Set. The superior performance of the Rosetta ResidueDisorder method on the 16 benchmark proteins might have been biased since the method's parameters had been tuned for accuracy on the benchmark data set. Thus, in order to verify that the Rosetta ResidueDisorder method was not overfitted to the initial benchmark set proteins, applying the approach to an independent, larger test data set was essential. The test data set proteins were collected from the PDB database, and we assembled 229 proteins with a total of 10902 disordered and 10825 ordered residues. Rosetta ResidueDisorder (using the same parameters employed on the benchmark data set) as well as the other six prediction tools were used to predict disordered regions for all 229 proteins. Results for this test are shown in Table 2. The Rosetta

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

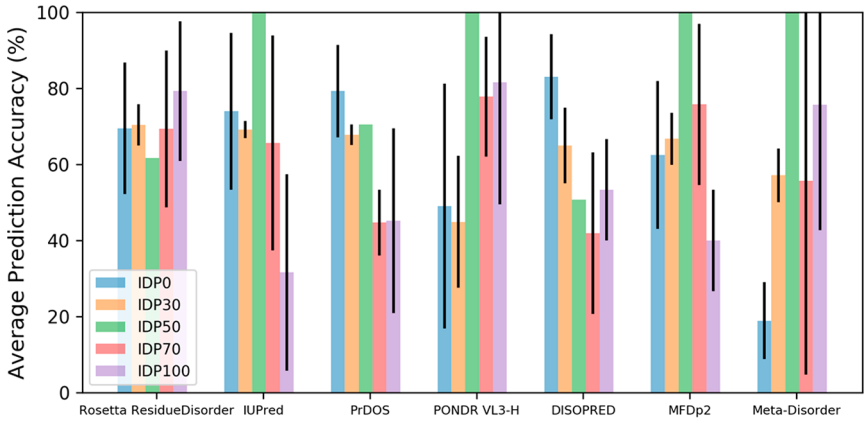


Figure 4. Comparison of six prediction tools' accuracy on 16 benchmark set proteins. Bar graph compares the average percent accuracy of 5 different IDP categories (0% disordered proteins (blue); 30% disordered proteins (yellow); 50% disordered proteins (green); 70% disordered proteins (red); 100% disordered proteins (purple)) for each of the six prediction tools. The error bar represents the standard deviation for each IDP category. IDP 50% bar graphs do not have error bars, because the IDP 50% category contained only one protein. A biased prediction accuracy can be observed toward long-length disordered regions for PONDR VL3-H and Meta-Disorder, and a biased prediction accuracy toward ordered regions for PrDOS, IUPred, DISOPRED, and MFDp2. Compared to the other tools, the Rosetta ResidueDisorder method shows consistent prediction accuracy throughout all levels of disorder.

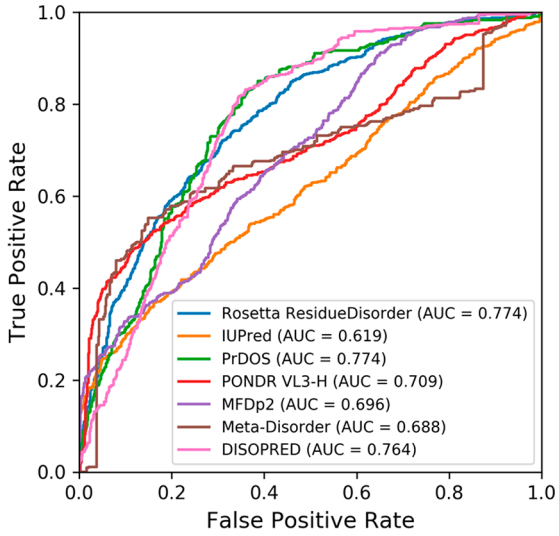


Figure 5. ROC curve analysis of the benchmark data set. ROC curves of six different prediction tools are shown: Rosetta ResidueDisorder (blue); IUPred (orange); PrDOS (green); PONDR VL3-H (red); MFDp2 (purple); Meta-Disorder (brown); DISOPRED (pink). AUCs are shown in the legend.

ResidueDisorder prediction accuracy on the test data set was 65.37%, which again was the highest accuracy among all seven prediction tools. However, not surprisingly, this value was lower than the 71.8% accuracy for the benchmark data set since Rosetta ResidueDisorder had not been optimized for the test data set.

Figure 6 illustrates the average prediction accuracies in five different IDP categories (0%, 30%, 50%, 70%, and 100% disorder) for all seven different prediction tools (Rosetta ResidueDisorder, IUPred, PrDOS, PONDR VL3-H, DISOPRED, MFDp2, and Meta-Disorder servers). Again, prediction standard deviations are illustrated as error bars on the bar graphs. The biased prediction trend observed for the benchmark data set is considerably more apparent for the larger test data set, where other prediction tools tended to show biased prediction accuracy toward either ordered or disordered proteins (Figure 6). For example, across all 229 proteins, IUPred exhibited a high accuracy for predicting ordered protein regions (proteins in the category of 0% and 30% disorder: prediction accuracy = 87.83% and 72.21%, respectively), whereas it performed rather poorly on the larger disordered regions (proteins in the category of 70% and 100% disorder, prediction accuracy = 41.39% and 30.38%, respectively). A similar biased trend was seen in PrDOS, DISOPRED, and MFDp2 as well. The PrDOS average prediction accuracies of proteins in the category of 0% and 100% disorder were 85.24%, and 27.12%, respectively, the DISOPRED average prediction accuracies for 0% and 100% disorder were 91.49% and 20.78%, respectively, while the MFDp2 average prediction accuracies were 81.98%, and 41.67%, respectively. Meta-Disorder and PONDR VL3-H exhibited the opposite, yet still biased trend with a high accuracy for predicting strongly disordered regions and performed worse on the ordered regions. The Meta-Disorder average prediction accuracies of proteins in the category of 0% and 100% disorder were 20.08% and 83.93%, respectively, while the PONDR VL3-H average prediction accuracies of proteins in the category of 0% and 100% disorder were 42.62% and 69.19%, respectively. Similar to what was seen for the benchmark data set, Rosetta ResidueDisorder showed consistent prediction accuracy throughout all levels of disorder. The average prediction accuracy of proteins in the category of 0%, 30%, 50%, 70%, and 100% disordered were 70.99%, 65.12%, 59.17%, 62.95%, and 65.68% respectively. The differences between each disordered category's average prediction accuracy were less than 12%. The difference between prediction accuracy of total disordered residues and total ordered residues for Rosetta ResidueDisorder method was 3%, which is the lowest difference among all prediction tools. For example, IUPred, PrDOS, and MFDp2 performed 64%, 61%, and 44% better at predicting ordered regions, respectively. Meta-Disorder and PONDR VL3-H performed 57% and 25% better at predicting disordered regions, respectively. In order to measure how accurately the Rosetta ResidueDisorder method predicted the percent disorder in a protein, we calculated the absolute difference between the known percent disorder and the predicted percent disorder. The absolute difference in

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

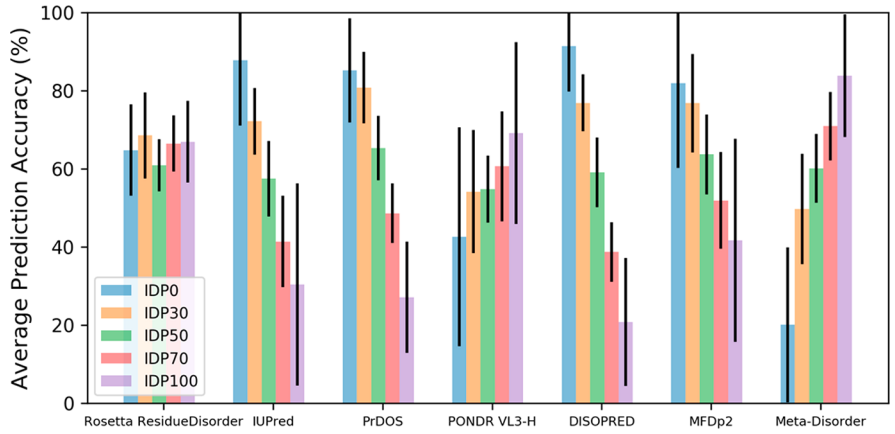


Figure 6. Comparison of six prediction tools' average prediction accuracy of the test data set. Bar graph compares the average percent accuracy of five different IDP categories (0% disordered proteins (blue); 30% disordered proteins (yellow); 50% disordered proteins (green); 70% disordered proteins (red); 100% disordered proteins (purple)) for each of six prediction tools. The error bar represents the standard deviation for each IDP category. The bar graph clearly illustrates a biased prediction accuracy toward long-length disordered regions for PONDR VL3-H and Meta-Disorder, and a biased prediction accuracy toward ordered proteins for PrDOS, IUPred, DISOPRED, and MFDp2. Compared to the other tools, the Rosetta ResidueDisorder method shows consistent prediction accuracy throughout all levels of disorder.

predicting the total percent disordered in proteins was 25% for the 229 NMR test set proteins. The Rosetta ResidueDisorder method overpredicted 51%, under-predicted 48%, and perfectly predicted 1% of the 229 NMR test set proteins' percent disorders.

To further assess the ability of the Rosetta ResidueDisorder method to correctly rank ordered and disordered residues according to the order score, a ROC curve analysis was performed, and the area-under-the-curve (AUC) was calculated. For comparison, a similar analysis was again performed for all other prediction tools. The ROC curves were generated by combining and ranking residue predictions of all test data set proteins. Figure 7 shows the ROC curves of all prediction tools. The Rosetta ResidueDisorder and DISOPRED methods exhibited the highest AUC (0.718) among all prediction tools suggesting that the trends observed for the benchmark data set hold true for the independent test data set. As shown in Figure 7, DISOPRED performs as well as the Rosetta ResidueDisorder method at ranking the true positive (i.e., disordered) residues by their confidence values. However, Rosetta ResidueDisorder (65.37% accuracy) outperforms DISOPRED (59.46% accuracy) at accurately predicting disordered regions.

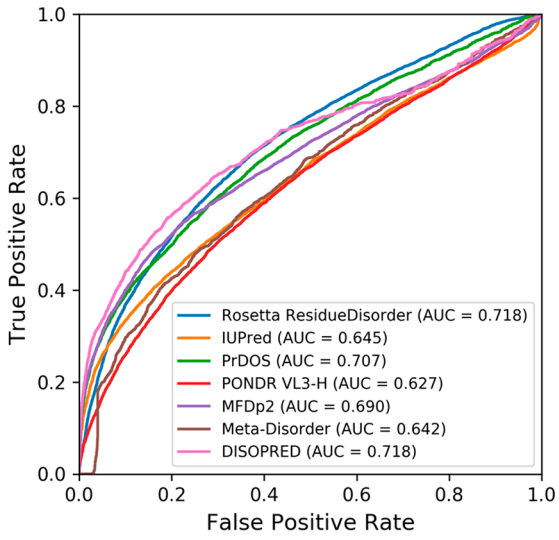


Figure 7. ROC curve analysis of the test data set. ROC curves of six different prediction tools are shown: Rosetta ResidueDisorder (blue); IUPred (orange); PrDOS (green); PONDR VL3-H (red); MFDp2 (purple); Meta-Disorder (brown); DISOPRED (pink). AUCs are shown in the legend.

Furthermore, the structural consensus of the 100 *ab initio* models was analyzed to evaluate the structural alternative to the score-based prediction of the Rosetta ResidueDisorder method. The overall prediction accuracy of the RMSF-based structure consensus evaluation method for the 229 test set proteins was 54.17%, which is lower than the order score-based Rosetta ResidueDisorder method (65.37%).

4. CONCLUSIONS

Here we presented the Rosetta ResidueDisorder method, an easy and accurate approach to use Rosetta for the prediction of disordered regions in proteins. Using 245 proteins from a benchmark data set (16 DisProt database proteins) and a test data set (229 proteins with NMR data), we demonstrated that ordered and disordered regions are distinguishable by Rosetta score. As a result, our Rosetta ResidueDisorder method (benchmark data set prediction accuracy of 71.77% and independent test data set prediction accuracy of 65.37%) outperformed other established disorder prediction tools. Most importantly, our method did not exhibit a biased prediction toward either ordered or disordered regions. We are hypothesizing that Rosetta's successful performance is based on its probing of long-range residue contacts, which reduce the disorder considerably. This sets it apart from other sequence-based methods, which cannot account for sequence-distant residue contacts. One limiting factor of the developed Rosetta ResidueDisorder method is the length of time it takes for a prediction. The Rosetta ResidueDisorder method requires 100 *de novo* structures to predict residue order scores and subsequently residue characteristics (ordered or disordered). Depending on the size of the protein, our method approximately takes 2 h for a complete analysis. This compares to predictions on the order of minutes for the other investigated prediction tools. To facilitate the usage of the Rosetta ResidueDisorder method as much as possible, it has

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

been developed as a Rosetta application and is freely available. In conclusion, Rosetta ResidueDisorder is a suitable and accurate method for prediction of disordered protein regions in cases when an unbiased prediction is favored, and the result does not have to be obtained within minutes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b01763.

Appendix S1: Rosetta *ab initio* folding options; Appendix S2: How to use the Rosetta ResidueDisorder application to predict order/disorder; Figure S1: Prediction accuracy of the benchmark set proteins; Figure S2: Prediction accuracy of the test data set; Table S1: 229 NMR test set proteins (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Address: Department of Chemistry and Biochemistry, Ohio State University 2114 Newman & Wolfrom Laboratory, 100 W. 18th Avenue, Columbus, OH 43210. 614-292-8284 (office); 614-292-1685 (fax); lindert.1@osu.edu (e-mail).

ORCID

Steffen Lindert: 0000-0002-3976-3473

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the members of the Lindert group for useful discussions. We would like to thank the Ohio Supercomputer Center for valuable computational resources.[59] Work in the Lindert laboratory is supported through NIH (R03 AG054904, R01 HL137015), NSF (CHE 1750666), and a Falk Medical Research Trust Catalyst Award.

■ REFERENCES

(1) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; et al. The Protein Data Bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58* (6), 899−907.

(2) Berman, H. M.; Coimbatore Narayanan, B.; Di Costanzo, L.; Dutta, S.; Ghosh, S.; Hudson, B. P.; Lawson, C. L.; Peisach, E.; Prlić, A.; Rose, P. W.; et al. Trendspotting in the Protein Data Bank. *FEBS Lett.* **2013**, *587* (8), 1036−1045.

(3) Uversky, V. N. Introduction to intrinsically disordered proteins (IDPs). *Chem. Rev.* **2014**, *114* (13), 6557−6560.

(4) van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **2014**, *114* (13), 6589−6631.

(5) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (1), 18−29.

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

(7) Pancsa, R.; Fuxreiter, M. Interactions via intrinsically disordered regions: what kind of motifs? *IUBMB Life* **2012**, *64* (6), 513−520.

(8) Jensen, M. R.; Zweckstetter, M.; Huang, J. R.; Blackledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* **2014**, *114* (13), 6632−6660.

(9) Bordoli, L.; Kiefer, F.; Schwede, T. Assessment of disorder predictions in CASP7. *Proteins: Struct., Funct., Genet.* **2007**, *69* (S8), 129−136.

(10) 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* (17), 3069−3090.

(11) Dosztányi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433−3434.

(12) Ishida, T.; Kinoshita, K. PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **2007**, *35*, W460−W464.

(13) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389−3402.

(14) Ward, J. J.; McGuffin, L. J.; Bryson, K.; Buxton, B. F.; Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **2004**, *20* (13), 2138−2139.

(15) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157* (1), 105−132.

(16) Williams, R. J. The conformational mobility of proteins and its functional significance. *Biochem. Soc. Trans.* **1978**, *6* (6), 1123−1126.

(17) Huang, F.; Oldfield, C. J.; Xue, B.; Hsu, W. L.; Meng, J.; Liu, X.; Shen, L.; Romero, P.; Uversky, V. N.; Dunker, A. Improving protein order-disorder classification using charge-hydropathy plots. *BMC Bioinf.* **2014**, *15* (Suppl 17), S4.

(18) Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K. Predicting intrinsic disorder from amino acid sequence. *Proteins: Struct., Funct., Genet.* **2003**, *53* (S6), 566−572.

(19) Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinf. Comput. Biol.* **2005**, *03*, 35−60.

(20) Kozlowski, L. P.; Bujnicki, J. M. MetaDisorder: A meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinf.* **2012**, *13*, 111.

(21) Mizianty, M. J.; Uversky, V.; Kurgan, L. Prediction of intrinsic disorder in proteins using MFDp2. *Methods Mol. Biol.* **2014**, *1137*, 147−162.

(22) Li, J.; Feng, Y.; Wang, X.; Liu, W.; Rong, L.; Bao, J. An overview of predictors for intrinsically disordered proteins over 2010−2014. *Int. J. Mol. Sci.* **2015**, *16* (10), 23446−23462.

(23) Hirose, S.; Shimizu, K.; Kanai, S.; Kuroda, Y.; Noguchi, T. POODLE-L: A two-level SVM prediction system for reliably predicting long disordered regions. *Bioinformatics* **2007**, *23* (16), 2046−2053.

(24) Jones, D. T., et al.The PSIPRED Protein Sequence Analysis Workbench; UCL Department of Computer Science: London, 1998.

(25) Bender, B. J.; Cisneros, A.; Duran, A. M.; Finn, J. A.; Fu, D.; Lokits, A. D.; Mueller, B. K.; Sangha, A. K.; Sauer, M. F.; Sevy, A. M.; et al. Protocols for molecular modeling with Rosetta3 and RosettaScripts. *Biochemistry* **2016**, *55* (34), 4748−4763.

(26) Das, R.; Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **2008**, *77*, 363−382.

(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* (1), 209−225.

(28) Simons, K. T.; Ruczinski, I.; Kooperberg, C.; Fox, B. A.; Bystroff, C.; Baker, D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins: Struct., Funct., Genet.* **1999**, *34* (1), 82−95.

(29) Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **2004**, *383*, 66−93.

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930

(30) DiMaio, F.; Song, Y.; Li, X.; Brunner, M. J.; Xu, C.; Conticello, V.; Egelman, E.; Marlovits, T. C.; Cheng, Y.; Baker, D. Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **2015**, *12* (4), 361−365.
(31) DiMaio, F.; Tyka, M. D.; Baker, M. L.; Chiu, W.; Baker, D. Refinement of protein structures into low-resolution density maps using rosetta. *J. Mol. Biol.* **2009**, *392* (1), 181−190.
(32) Lindert, S.; McCammon, J. A. Improved cryoEM-guided iterative molecular dynamics–Rosetta protein structure refinement protocol for high precision protein structure prediction. *J. Chem. Theory Comput.* **2015**, *11* (3), 1337−1346.
(33) Lindert, S.; Meiler, J.; McCammon, J. A. Iterative molecular dynamics-Rosetta protein structure refinement protocol to improve model quality. *J. Chem. Theory Comput.* **2013**, *9* (8), 3843−3847.
(34) Lindert, S.; Alexander, N.; Wotzel, N.; Karakas, M.; Stewart, P. L.; Meiler, J. EM-fold: de novo atomic-detail protein structure determination from medium-resolution density maps. *Structure* **2012**, *20* (3), 464−478.
(35) Leelananda, S. P.; Lindert, S. Iterative molecular dynamics-Rosetta membrane protein structure refinement guided by cryo-EM densities. *J. Chem. Theory Comput.* **2017**, *13* (10), 5131−5145.
(36) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003**, *302* (5649), 1364−1368.
(37) Leaver-Fay, A.; Jacak, R.; Stranges, P. B.; Kuhlman, B. A generic program for multistate protein design. *PLoS One* **2011**, *6* (7), e20937.
(38) Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E. M.; Wilson, I. A.; Baker, D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **2011**, *332* (6031), 816−821.
(39) Greisen, P.; Khare, S. D. Computational redesign of metalloenzymes for catalyzing new reactions. *Methods Mol. Biol.* **2014**, *1216*, 265−273.
(40) Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; Baker, D. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **2013**, *501* (7466), 212−216.
(41) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003**, *331* (1), 281−299.
(42) Chaudhury, S.; Berrondo, M.; Weitzner, B. D.; Muthu, P.; Bergman, H.; Gray, J. J. Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One* **2011**, *6* (8), e22477.
(43) Raveh, B.; London, N.; Zimmerman, L.; Schueler-Furman, O. Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS One* **2011**, *6* (4), e18934.
(44) Meiler, J.; Baker, D. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins: Struct., Funct., Genet.* **2006**, *65* (3), 538−548.
(45) Davis, I. W.; Baker, D. RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009**, *385* (2), 381−392.
(46) DeLuca, S.; Khar, K.; Meiler, J. Fully flexible docking of medium sized ligand libraries with RosettaLigand. *PLoS One* **2015**, *10* (7), e0132508.
(47) Das, R.; Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (37), 14664−14669.
(48) 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* (6), 3031−3048.
(49) Bradley, P.; Misura, K. M.; Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **2005**, *309* (5742), 1868−1871.
(50) Barth, P.; Schonbrun, J.; Baker, D. Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (40), 15682−15687.
(51) Kinch, L.; Yong Shi, S.; Cong, Q.; Cheng, H.; Liao, Y.; Grishin, N. V. CASP9 assessment of free modeling target predictions. *Proteins: Struct., Funct., Genet.* **2011**, *79* (S10), 59−73.
(52) Tai, C. H.; Bai, H.; Taylor, T. J.; Lee, B. Assessment of template-free modeling in CASP10 and ROLL. *Proteins: Struct., Funct., Genet.* **2014**, *82* (S2), 57−83.
(53) Wang, R. Y.; Han, Y.; Krassovsky, K.; Sheffler, W.; Tyka, M.; Baker, D. Modeling disordered regions in proteins using Rosetta. *PLoS One* **2011**, *6* (7), e22060.
(54) O'Meara, M. J.; Leaver-Fay, A.; Tyka, M. D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; Kuhlman, B. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J. Chem. Theory Comput.* **2015**, *11* (2), 609−622.
(55) R Development Core Team, *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
(56) Ruxton, G. D. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology* **2006**, *17*, 688−690.
(57) 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* (1), 235−242.
(58) Bonneau, R.; Ruczinski, I.; Tsai, J.; Baker, D. Contact order and ab initio protein structure prediction. *Protein Sci.* **2002**, *11* (8), 1937−1944.
(59) Ohio Supercomputer Center, 1987.

DOI: 10.1021/acs.jpcb.8b01763
J. Phys. Chem. B 2018, 122, 3920−3930