

Disorder Predictors Also Predict Backbone Dynamics for a Family of Disordered Proteins

Gary W. Daughdrill^{1*}, Wade M. Borchers², Hongwei Wu³

1 Department of Cell Biology, Microbiology and Molecular Biology and Center for Drug Discovery and Innovation, University of South Florida, Tampa, Florida, United States of America, **2** Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, Florida, United States of America, **3** Center for Drug Discovery and Innovation, University of South Florida, Tampa, Florida, United States of America

Abstract

Several algorithms have been developed that use amino acid sequences to predict whether or not a protein or a region of a protein is disordered. These algorithms make accurate predictions for disordered regions that are 30 amino acids or longer, but it is unclear whether the predictions can be directly related to the backbone dynamics of individual amino acid residues. The nuclear Overhauser effect between the amide nitrogen and hydrogen (NHNOE) provides an unambiguous measure of backbone dynamics at single residue resolution and is an excellent tool for characterizing the dynamic behavior of disordered proteins. In this report, we show that the NHNOE values for several members of a family of disordered proteins are highly correlated with the output from three popular algorithms used to predict disordered regions from amino acid sequence. This is the first test between an experimental measure of residue specific backbone dynamics and disorder predictions. The results suggest that some disorder predictors can accurately estimate the backbone dynamics of individual amino acids in a long disordered region.

Citation: Daughdrill GW, Borchers WM, Wu H (2011) Disorder Predictors Also Predict Backbone Dynamics for a Family of Disordered Proteins. PLoS ONE 6(12): e29207. doi:10.1371/journal.pone.0029207

Editor: Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

Received: November 4, 2011; **Accepted:** November 22, 2011; **Published:** December 15, 2011

Copyright: © 2011 Borchers et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the American Cancer Society (RSG-07-289-01-GMC; www.acs.org) and the National Science Foundation (MCB-0939014; www.nsf.gov). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gdaughdrill@usf.edu

Introduction

Intrinsically disordered proteins (IDPs) are widespread in eukaryotes and overrepresented in a number of human diseases, including cancer and neurodegenerative diseases [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15]. While the development of robust generalizations to describe the structure and function of IDPs is ongoing, a number of algorithms have been developed to identify IDPs and distinguish them from ordered proteins. This is achieved using the compositional differences in the amino acid sequences observed between ordered and disordered proteins [16,17,18,19]. These algorithms can identify disordered regions that are 30 amino acids or longer with 75–80% accuracy, but it is unclear whether they provide any information about backbone dynamics at the level of single amino acid residues.

Three popular disorder predictors, VL-XT (www.pondr.com), IUPred (iupred.enzim.hu), and VSL2B (ist.temple.edu) were selected to investigate whether there is a relationship between disorder probability and backbone dynamics. These algorithms were tested because they predict disorder using different principles. The VL-XT predictor integrates three feed forward neural network predictors (NNP). One NNP was trained using 8 long disordered regions identified from missing electron density in x-ray crystallographic studies, and 7 long disordered regions characterized by nuclear magnetic resonance (NMR) spectroscopy [19]. The other two NNPs were also trained using missing electron density from x-ray crystallographic data [20]. For these NNPs, N- and C-terminal disordered regions of 5 or more amino acids were used in the training set. The abbreviation, VL-XT, stands for the

use of Various methods to characterize Long disordered regions combined with Terminal disordered regions that were characterized using X-ray crystallography. When making predictions, VL-XT gives an output between 0 and 1 that is smoothed over a sliding window of 9 amino acids. If a residue value exceeds or matches a threshold of 0.5 the residue is considered disordered. IUPred is an abbreviation for Intrinsically Unstructured protein Predictor. This algorithm distinguishes ordered regions from disordered regions by estimating pairwise interaction energies. This interaction energy is determined using amino acid composition, the local sequence environment, and potential intramolecular interaction partners. When predicting long disordered regions, IUPred calculates interaction energies over a 100 residue sequential neighborhood. IUPred also provides an output that varies between 0 and 1 with a threshold for the transition between order and disorder of 0.5 [21,22]. VSL2B was also considered, it is a disorder prediction algorithm developed by the same group that developed VL-XT [23,24]. VSL2B stands for Various Short Long predictor and is the second version of VSL, and B denotes that it does not include the PSI-BLAST feature set. VSL2B was developed using a larger database of experimentally characterized IDPs than VL-XT, it incorporates 26 sequence-based features, including secondary structure prediction. VSL2B is more accurate than VL-XT at predicting short disordered regions; it accomplishes this by having a two tiered prediction method with separate prediction mechanisms for long or short disordered regions that are then combined by a Meta predictor [25,26]. VSL2B also provides an output scale from 0 to 1 with 1 being most disordered

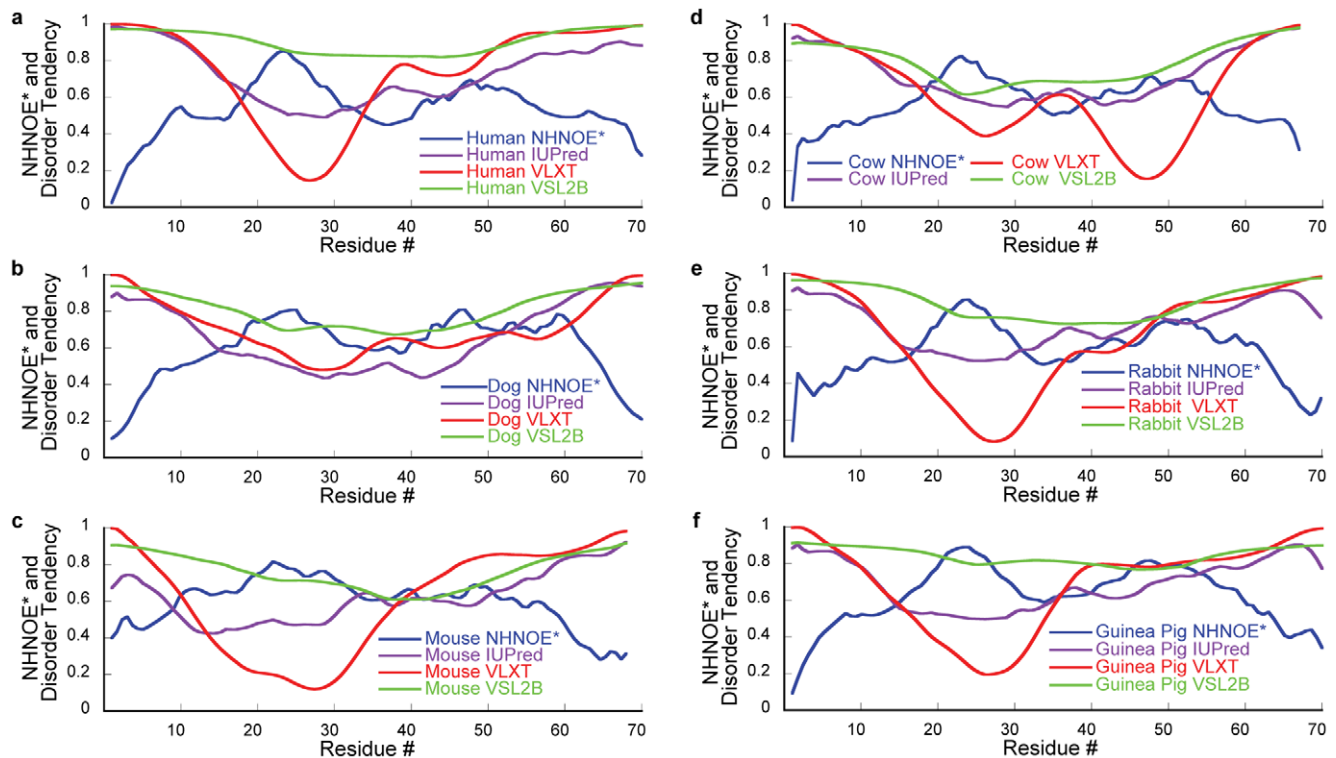


Figure 2. NHNOE* and Disorder probabilities smoothed over a 5 residue window. (a) Human, (b) Dog, (c) Mouse, (d) Cow, (e) Guinea Pig, and (f) Rabbit p53TAD.
doi:10.1371/journal.pone.0029207.g002

the full length and p53TAD sequences. The PSI-BLAST feature set was excluded since it incorporates sequence homology, and this dataset compares homologues already.

Visual inspection of the plots in Figure 2 show a strong negative correlation between the residue specific backbone dynamics and the disorder probabilities of residues 15–30 for all three predictors and all six homologues. Residues 15–30 form the MDM2 binding region of p53. This region forms transient helical structure in the absence of MDM2 that is stabilized upon binding [34,36,40]. Formation of transient helical structure in IDPs increases the local rotational correlation time, corresponding to a reduction in overall molecular motion and a higher NHNOE*. The NHNOE* values for the six homologues show a peak in this region, which is consistent with the previous observations of transient helical structure for human p53TAD. Transient helical structure is also observed for residues 15–30 of the non-human p53TAD homologues (data not shown). All three predictors are sensitive to this behavior and show a dip (to some extent) in the MDM2 binding region. This behavior was expected for VL-XT. It was previously shown that this predictor is able to discriminate regions of IDPs that become ordered upon binding to a protein partner [46,47]. The data presented in Figure 2 suggest that VL-XT, IUPred, and VSL2B are sensitive to regions of reduced flexibility in IDPs. These regions often correlate with protein binding regions and contain local elements of transient secondary structure. Although there is known transient secondary structure in these regions, VSL2B shows the smallest dips in the binding sites, which is surprising given its programming takes secondary structure prediction into account. However, the dips for VSL2B more accurately align with the NHNOE* peaks, whereas IUPred and VLXT are shifted toward the C-termini. Reasonable negative correlations between experiment and prediction were also

observed for the RPA70 binding region, which encompasses residues 40–60 of human p53TAD. This region is not as conserved as the MDM2 binding region (see Figure 1) and is more dynamic. However, there are still small peaks for the NHNOE* values and small dips for all of the disorder probabilities in this region, with the exception of the VL-XT probabilities for mouse p53TAD.

Regression Analysis

Linear regression was performed on the NHNOE* and disorder probabilities to assess the statistical significance of the correlations shown in Figure 2. Correlation coefficients and two-tailed p-values for the linear regression are shown in Table 1. All the p-values are less than 0.002 and in most cases are less than 0.0001, indicating the correlations are significant. While these three predictors were specifically designed to identify long disordered regions they can also accurately estimate the backbone dynamics of individual amino acids in these long disordered regions.

Figure 3 shows the correlation plot between NHNOE* and the disorder probabilities with the strongest correlation for each of the six homologues, respectively. There is no single predictor with the strongest correlation for a majority of the homologues; in fact the predictors split the homologues evenly, with each having the strongest correlation for two of the six homologues. Interestingly, the homologues appear to be apportioned according to their molecular flexibility. Based on the raw NHNOE data the overall molecular flexibility of the six homologues ranks in increasing disorder as follows: Guinea Pig < Rabbit < Mouse < Human < Cow < Dog. VLXT probabilities have the highest correlation with the two most dynamic homologues, dog and cow. VSL2B probabilities have the highest correlation with the two least dynamic homologues rabbit and guinea pig. IUPred probabilities have the highest correlation with the two homologues displaying

Table 1. Linear Regression of NHNOE* Versus Disorder Probability.

Species	Predictor	Sample Size	r	Two-tailed p-values
Human	IUPred	58	0.55	0.000008
Human	VL-XT	58	0.54	0.000012
Human	VSL2B	58	0.42	0.00103
Dog	IUPred	60	0.44	0.000435
Dog	VL-XT	60	0.65	<0.000001
Dog	VSL2B	60	0.49	0.000071
Mouse	IUPred	60	0.65	<0.000001
Mouse	VL-XT	60	0.58	0.000001
Mouse	VSL2B	60	0.51	0.000031
Cow	IUPred	59	0.59	0.000001
Cow	VL-XT	59	0.66	<0.000000
Cow	VSL2B	59	0.62	<0.000000
Guinea Pig	IUPred	61	0.58	0.000001
Guinea Pig	VL-XT	61	0.53	0.000014
Guinea Pig	VSL2B	61	0.71	<0.000000
Rabbit	IUPred	62	0.43	0.000524
Rabbit	VL-XT	62	0.43	0.000515
Rabbit	VSL2B	62	0.48	0.000065

doi:10.1371/journal.pone.0029207.t001

intermediate dynamics, human and mouse. While this result is interesting, the dataset is too small to assign any statistical significance to this partitioning.

Accuracy of residue specific correlations between backbone dynamics and disorder probability

Figure 4 shows the relative percentages for amino acid residues with VL-XT, IUPred, or VSL2B values that were $\geq 1\sigma$ from the mean value estimated using the linear regression equations. Relative percentages were determined by identifying the number of times a particular residue type had a VL-XT, VSL2B or IUPred value that was $\geq 1\sigma$ from the mean. This number was then divided by the total number of a given residue type in the six sequences and converted to a percentage. All disorder probabilities from each of the three predictors were within 2σ of the linear regression means, which is a common cutoff used to identify outliers. A cutoff of 1σ was chosen for the current analysis because values of disorder predictors that are $\geq 1\sigma$ from the mean are structurally significant. For instance, W53 in human p53TAD has a VL-XT disorder probability of 0.94, which is $\geq 1\sigma$ but $\leq 2\sigma$ from the mean, this disorder probability corresponds to an NHNOE* value of 0.27, compared with the measured NHNOE* value of 0.66. The difference in the predicted and measured NHNOE* values for W53 is structurally significant and corresponds to the difference between an amino acid residue that is completely disordered versus one that has some transient secondary structure.

In Figure 4, relative percentages are zero for P because the NHNOE cannot be measured due to its lack of an amide hydrogen. Relative percentages are also zero for all K residues, but in this case all three predictors were within 1σ of the mean for these residues. The predictors make an average of $22.0\% \pm 3.98\%$ errors, with disorder probabilities $\geq 1\sigma$ from the mean, excluding residues that do not occur at least once per homologue. Only the A, E, L, and V residues are above this range for all three predictors. This indicates that VL-XT, VSL2B, and IUPred are inaccurately predicting the dynamic behavior for these residue types. Probability values $\geq 1\sigma$ for A and E residues occur most

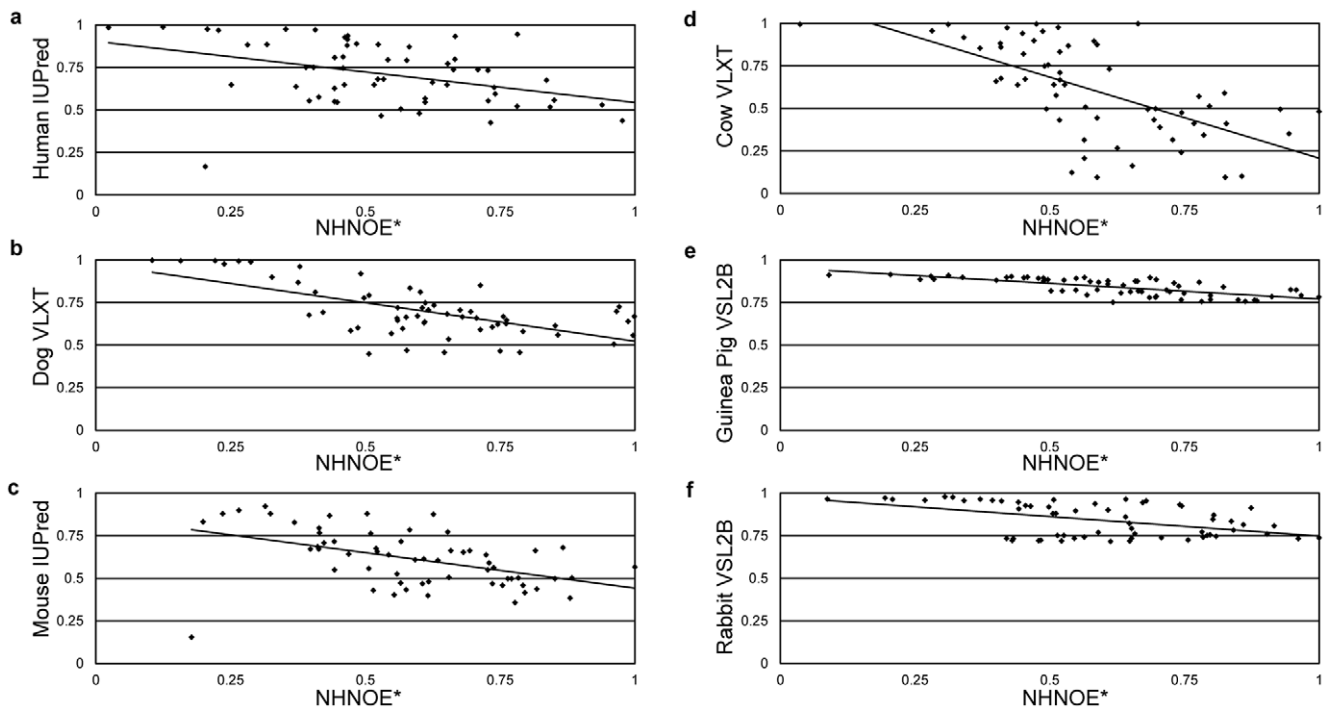


Figure 3. Correlation Plots of the Best fitting plots for Human IUPred. (a), Dog VL-XT (b), Mouse IUPred (c), Cow VL-XT (d), Guinea Pig VSL2b (e), and Rabbit VSL2b (f).

doi:10.1371/journal.pone.0029207.g003

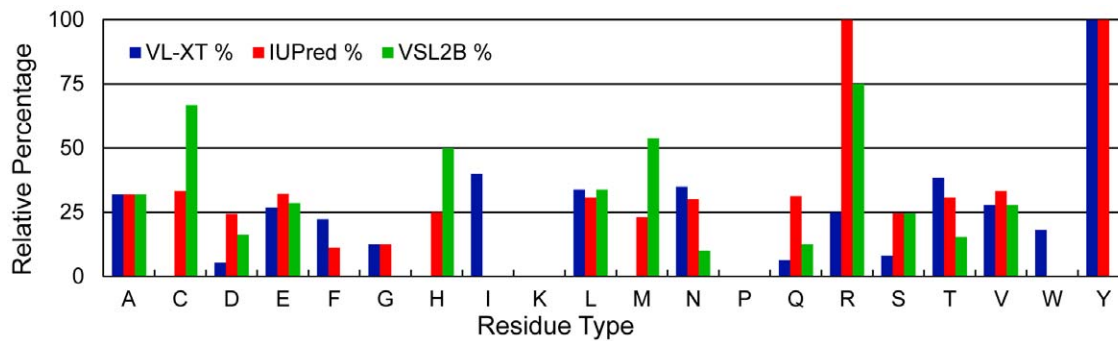


Figure 4. Relative percentage of incorrect predictions for VL-XT, IUPred, and VSL2B. An incorrect prediction is defined as a VL-XT, IUPred, or VSL2B value that is at least 1σ from the mean value estimated using the linear regression equations.
doi:10.1371/journal.pone.0029207.g004

frequently above the regression line, which indicates that all three algorithms predict these residue types to be less ordered than they actually are. For A residues there are 17 predictions that are greater than the mean and 6 that are less, while for E residues there are 29 predictions that are greater than the mean and 21 that are less. Conversely L and V probabilities $\geq 1\sigma$ from the mean are most often below the regression line, which indicates that all three algorithms predict these residue types to be more ordered than they actually are. For L residues there are 47 predictions that are less than the mean and 15 that are greater, while for V residues there are 10 predictions less than the mean and 7 that are greater.

L, and V were previously classified as order-promoting residues based on their low frequency of occurrence in a database of experimentally characterized disordered proteins [48]. A was classified as a neutral amino acid, while E was classified as a disorder promoting residue. The frequency of L and E residues in the p53TAD homologues are approximately 90% and 70% higher than the values observed in the database, respectively. A and V residues are approximately 20% and 30% less frequent in the p53TAD homologues than in the database, respectively. According to the NHNOE, L and V residues are more dynamic than predicted, while A and E residues are less dynamic than predicted. L is particularly interesting because over 75% of disorder probabilities $\geq 1\sigma$ from the mean value are predicted to be more ordered than the dynamics data indicate. The high frequency of an order-promoting residue may account for why VL-XT, VSL2B, and IUPred make so many incorrect predictions for L residues in these p53TAD sequences.

Summary

In this report, we show that three popular algorithms for predicting disorder probability based on amino acid sequence can accurately estimate the backbone dynamics of individual amino acids in long disordered regions from a family of disordered proteins. These findings are consistent with the results from three previous studies, which observed correlations between backbone dynamics and disorder probability using either molecular dynamics simulations or NMR dynamics measurements. In one of these studies, increased internal flexibility was suggested by both disorder predictors and molecular dynamics simulations [49]. In a second study, the insertion of a β -hairpin between two chimeric proteins resulted in a significant increase in predicted disorder probabilities that were subsequently confirmed using NMR dynamics measurements [50]. The third study found a strong negative correlation between the degree of predicted disorder and the stability of the protein complexes. In the third study, molecular

dynamics simulations were used to show that binding regions with higher predicted disorder probabilities correlated with weaker complex formation [51]. While the dataset used for this comparison is small, the correlations observed are robust, and demonstrates the accuracy of the disorder predictors for backbone dynamics at single amino acid resolution. As more experimental data on the backbone dynamics of IDPs is collected we predict this relationship will be refined so that in the future the NHNOE and other NMR measurements that provide information about residue specific structure and dynamics can be used to guide the development of disorder predictors.

Materials and Methods

Protein purification

Samples of human p53TAD (residues 1–73) that were uniformly labeled with either ^{15}N or ^{15}N and ^{13}C , were prepared as previously described [36]. Samples for dog (residues 1–77), mouse (residues 1–87), cow (residues 1–82), guinea pig (residues 1–88), and rabbit (residues 1–87) p53TAD were prepared using this same method.

NMR data collection and analysis

Resonance assignments for human p53TAD were previously reported [36]. Experiments on mouse, dog, cow, guinea pig, and rabbit p53TAD were carried out at 25°C on a Varian VNMRS 600 MHz spectrometer equipped with a triple resonance pulse field z -axis gradient cold probe. To make the amide ^1H and ^{15}N as well as $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{13}CO resonance assignments, sensitivity enhanced ^1H - ^{15}N HSQC and three dimensional HNCACB and HNCOC experiments were performed on the uniformly ^{15}N and ^{13}C labeled samples of dog (0.47 mM), mouse (0.36 mM), cow (0.38 mM), guinea pig (0.57 mM), and rabbit (0.341 mM) p53TAD in 90% H_2O /10% D_2O , PBS buffer, at a pH of 6.8. For the HNCACB experiment, data were acquired in ^1H , ^{13}C and ^{15}N dimensions using $8012.8 (t_3) \times 12000 (t_2) \times 2000 (t_1)$ Hz sweep widths, and $512 (t_3) \times 128 (t_2) \times 32 (t_1)$ complex data points. For the HNCOC, the sweep widths were $8012.8 (t_3) \times 3770 (t_2) \times 2000 (t_1)$ Hz, complex data points were identical to the HNCACB. The sweep widths and complex points of the HSQC were $8012.8 (t_2) \times 2000 (t_1)$ Hz and $512 (t_2) \times 128 (t_1)$, respectively. For dog p53TAD, processing and analysis of the HNCACB data resulted in 65 non-proline, amide ^1H , ^{15}N , $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments plus 10 proline $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments. 63 ^{13}CO resonance assignments were gained from HNCOC data analysis. For mouse p53TAD, 75 non-proline, amide ^1H , ^{15}N , $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments plus 12 proline $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$

resonance assignments were gained from HNCACB. 72 ^{13}C O resonance assignments were gained from HNCB data analysis. For cow p53TAD, processing and analysis of the HNCACB data resulted in 67 non-proline, amide ^1H , ^{15}N , $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments plus 12 proline $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments. 65 ^{13}C O resonance assignments were gained from HNCB data analysis. For rabbit p53TAD, processing and analysis of the HNCACB data resulted in 71 non-proline, amide ^1H , ^{15}N , $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments plus 12 proline $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments. 71 ^{13}C O resonance assignments were gained from HNCB data analysis. For guinea pig p53TAD, processing and analysis of the HNCACB data resulted in 73 non-proline, amide ^1H , ^{15}N , $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments plus 11 proline $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ resonance assignments. 74 ^{13}C O resonance assignments were gained from HNCB data analysis.

All NMR spectra were processed with nmrPipe and analyzed using nmrView software [36,52]. Apodization was achieved in the ^1H , ^{13}C and ^{15}N dimensions using a squared sine bell function shifted by 70° . Apodization was followed by zero filling to twice the number of real data points and linear prediction was used in the ^{15}N dimension of the HNCACB and HNCB. The ^1H carrier frequency was set on the water peak, and 4.753 ppm was used as the reference frequency in this report.

^1H - ^{15}N steady-state NOE experiments were recorded in the presence and absence of a 120° off-resonance ^1H saturation pulse

every 5 ms for a total of 3 s. A total of 512 (t_2) \times 128 (t_1) complex points were recorded with 128 scans per increment. The NHNOE values were determined by taking the quotient of the intensity for resolved resonances in the presence and absence of proton saturation. Three measurements were made on each protein and the values were averaged.

Disorder Prediction

For IUPred (iupred.enzim.hu) the N-terminal residues comprising the TAD were input using the long disordered region prediction setting, human 1–90, dog 1–77, mouse 1–87, cow 1–82, rabbit 1–87, and guinea pig 1–88. The VL-XT (www.pondr.com) predictions represent the full length predictions for each homologue. For the VSL2B (ist.temple.edu) predictions the full protein sequences were entered, using only the VSL2B feature set.

Acknowledgments

The authors thank Dr. Bina Nayak, Anne T. Powell, and Alex Barrett for critically reading this manuscript

Author Contributions

Conceived and designed the experiments: GWD WMB. Performed the experiments: WMB HW. Analyzed the data: GWD WMB HW. Wrote the paper: GWD WMB.

References

- Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK (2005) Natively Disordered Proteins. In: Buchner JaK, T, eds. Protein Folding Handbook. Darmstadt: WILEY-VCH. pp 275–357.
- Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41: 6573–6582.
- Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic Protein Disorder in Complete Genomes. *Genome Inform Ser Workshop Genome Inform* 11: 161–171.
- Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, et al. (1998) Thousands of Proteins Likely to Have Long Disordered Regions. *Pac Symp Biocomput*. pp 437–448.
- Tompa P (2002) Intrinsically unstructured proteins. *Trends in Biochemical Sciences* 27: 527–533.
- Tompa P (2003) Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays* 25: 847–855.
- Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. *Febs Letters* 579: 3346–3354.
- Tompa P (2010) Structure and Function of Intrinsically Disordered Proteins. Taylor and Francis Group, Boca Raton.
- Uversky VN (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Science* 11: 739–756.
- Uversky VN (2002) What does it mean to be natively unfolded? *European Journal of Biochemistry* 269: 2–12.
- Uversky VN, Gillespie JR, Fink AL (2000) Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins-Structure Function and Genetics* 41: 415–427.
- Uversky VN (2009) Intrinsic disorder in proteins associated with neurodegenerative diseases. *Frontiers in Bioscience* 14: 5188–5238.
- Vucetic S, Xie HB, Iakoucheva LM, Oldfield CJ, Dunker AK, et al. (2007) Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *Journal of Proteome Research* 6: 1899–1916.
- Xie HB, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, et al. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *Journal of Proteome Research* 6: 1882–1898.
- Xie HB, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, et al. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *Journal of Proteome Research* 6: 1917–1932.
- Ferron F, Longhi S, Canard B, Karlin D (2006) A practical overview of protein disorder prediction methods. *Proteins-Structure Function and Bioinformatics* 65: 1–14.
- Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, et al. (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins Supplement*. pp 566–572.
- Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, et al. (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry* 44: 1989–2000.
- Romero P, Obradovic Z, Dunker AK (1997) Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome Inform Ser Workshop Genome Inform* 8: 110–124.
- Li X, Romero P, Rani M, Dunker AK, Obradovic Z (1999) Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop Genome Inform* 10: 30–40.
- Simon I, Dosztanyi Z, Csizmok V, Tompa P (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of Molecular Biology* 347: 827–839.
- Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21: 3433–3434.
- Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* 61(Suppl 7): 176–182.
- Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 7: 208.
- Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. *Bmc Bioinformatics* 7.
- Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins-Structure Function and Bioinformatics* 61: 176–182.
- Dyson HJ, Wright PE (1998) Equilibrium NMR studies of unfolded and partially folded proteins. *Nature Structural Biology* 5: 499–503.
- Dyson HJ, Wright PE (2001) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Nuclear Magnetic Resonance of Biological Macromolecules Pt B* 339: 258–270.
- Dyson HJ, Wright PE (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Unfolded Proteins*. San Diego: Academic Press Inc. pp 311–340.
- Eliezer D (2007) Characterizing residual structure in disordered protein states using nuclear magnetic resonance. *Methods in Molecular Biology*. pp 49–67.
- Bargonetti J, Manfredi JJ (2002) Multiple roles of the tumor suppressor p53. *Current Opinion in Oncology* 14: 86–91.
- Dawson R, Muller L, Dehner A, Klein C, Kessler H, et al. (2003) The N-terminal domain of p53 is natively unfolded. *Journal of Molecular Biology* 332: 1131–1141.
- Kaustov L, Yi GS, Ayed A, Bochkareva E, Bochkarev A, et al. (2006) p53 Transcriptional activation domain - A molecular chameleon? *Cell Cycle* 5: 489–494.
- Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, et al. (2000) Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *Journal of Biological Chemistry* 275: 29426–29432.

35. Lowry DF, Stancik A, Shrestha RM, Daughdrill GW (2008) Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. *Proteins-Structure Function and Bioinformatics* 71: 587–598.
36. Vise PD, Baral B, Latos AJ, Daughdrill GW (2005) NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. *Nucleic Acids Research* 33: 2061–2077.
37. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, et al. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proceedings of the National Academy of Sciences of the United States of America* 105: 5762–5767.
38. Woods DB, Vousden KH (2001) Regulation of p53 function. *Experimental Cell Research* 264: 56–66.
39. Bochkareva E, Kaustov L, Ayed A, Yi GS, Lu Y, et al. (2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15412–15417.
40. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, et al. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274: 948–953.
41. Al-Lazikani B, Jung J, Xiang ZX, Honig B (2001) Protein structure prediction. *Current Opinion in Chemical Biology* 5: 51–56.
42. Baker D, Sali A (2001) Protein structure prediction and structural genomics. *Science* 294: 93–96.
43. Chothia C, Lesk AM (1986) The Relation Between the Divergence of Sequence and Structure in Proteins. *EMBO J* 5: 823–826.
44. Chothia C, Lesk AM (1987) The Evolution of Protein Structures. *Cold Spring Harb Symp Quant Biol* 52: 399–405.
45. Petsko GA, Ringe D (2004) *Protein Structure and Function*. New Science Press Ltd, London.
46. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, et al. (2006) Analysis of molecular recognition features (MoRFs). *Journal of Molecular Biology* 362: 1043–1059.
47. Oldfield CJ, Cheng YG, Cortese MS, Romero P, Uversky VN, et al. (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 44: 12454–12470.
48. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, et al. (2007) Intrinsic disorder and functional proteomics. *Biophysical Journal* 92: 1439–1456.
49. Li L, Uversky VN, Dunker AK, Meroueh SO (2007) A computational investigation of allostery in the catabolite activator protein. *J Am Chem Soc* 129: 15668–15676.
50. Kutysenko VP, Prokhorov DA, Timchenko MA, Kudrevatykh YA, Gushchina LV, et al. (2009) Solution structure and dynamics of the chimeric SH3 domains, SHH- and SHA-“Bergeracs”. *Biochim Biophys Acta* 1794: 1813–1822.
51. Liang S, Li L, Hsu WL, Pilcher MN, Uversky V, et al. (2009) Exploring the molecular design of protein interaction sites with molecular dynamics simulations and free energy calculations. *Biochemistry* 48: 399–414.
52. Johnson BA, Blevins RA (1994) NMR view—a computer-program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR* 4: 603–614.