Structural divergence is more extensive than sequence divergence for a family of intrinsically disordered proteins

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · OCTOBER 2013

Impact Factor: 2.63 · DOI: 10.1002/prot.24303 · Source: PubMed

4 AUTHORS, INCLUDING:

Gary W Daughdrill
University of South Florida
45 PUBLICATIONS **1,224** CITATIONS

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Structural divergence is more extensive than sequence divergence for a family of intrinsically disordered proteins

Wade Borcherds,[1,2] Stepan Kashtanov,[1,2] Hongwei Wu,[1,2] and Gary W. Daughdrill[1,2]*

[1] Department of Cell Biology, Microbiology, and Molecular Biology, University of South Florida, 3720 Spectrum Blvd., Suite 321, Tampa, Florida 33612

[2] The Center for Drug Discovery and Innovation, University of South Florida, 3720 Spectrum Blvd., Suite 321, Tampa, Florida 33612

ABSTRACT

The p53 transactivation domain (p53TAD) is an intrinsically disordered protein (IDP) domain that undergoes coupled folding and binding when interacting with partner proteins like the E3 ligase, MDM2, and the 70 kDa subunit of replication protein A, RPA70. The secondary structure and dynamics of six closely related mammalian homologues of p53TAD were investigated using nuclear magnetic resonance (NMR) spectroscopy. Differences in both transient secondary structure and backbone dynamics were observed for the homologues. Many of these differences were localized to the binding sites for MDM2 and RPA70. The amount of transient helical secondary structure observed for the MDM2 binding site was lower for the dog and mouse homologues, compared with human, and the amount of transient helical secondary structure observed for the RPA70 binding site was higher for guinea pig and rabbit, compared with human. Differences in the amount of transient helical secondary structure observed for the MDM2 binding site were directly related to amino acid substitutions occurring on the solvent exposed side of the amphipathic helix that forms during the p53TAD/MDM2 interaction. Differences in the amount of transient helical secondary structure were not as easily explained for the RPA70 binding site because of its extensive sequence divergence. Clustering analysis shows that the divergence in the transient secondary structure of the p53TAD homologues exceeds the amino acid sequence divergence. In contrast, strong correlations were observed between the backbone dynamics of the homologues and the sequence identity matrix, suggesting that the dynamic behavior of IDPs is a conserved evolutionary feature.

Proteins 2013; 81:1686–1698.

Key words: intrinsically disordered proteins; protein dynamics; structural evolution; structural divergence.

INTRODUCTION

The dynamic behavior of intrinsically disordered proteins (IDPs) is a poorly understood molecular property that is necessary for function and is probably under multiple levels of evolutionary selection.[1–5] IDPs have evolved to maintain a state of high conformational entropy.[1,6–13] They are highly dynamic, do not form tertiary structures, and contain variable amounts of transient secondary structure. A number of studies have concluded that disordered proteins generally evolve faster than their ordered counterparts, this behavior is thought to be related to the weaker structural constraints imposed on IDPs.[14–20] Nevertheless, there are only a couple of studies that have directly analyzed and compared the impact of this rapid evolution on the structure and dynamics of IDP families.[21,22] In contrast, there have been many case studies, and several systematic studies on the conservation of structure and dynamics in ordered proteins.[23–27]

In a previous study from our group, it was shown that the dynamic behavior for a family of intrinsically

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: The American Cancer Society; Grant number: RSG-07-289-01-GMC; Grant sponsor: The National Science Foundation; Grant number: MCB-0939014.

*Correspondence to: Gary W. Daughdrill, Department of Cell Biology, Microbiology and Molecular Biology, Center for Drug Discovery and Innovation, University of South Florida, 3720 Spectrum Blvd., Suite 321, Tampa, Florida 33612. E-mail: gdaughdrill@usf.edu

Received 14 December 2012; Revised 11 March 2013; Accepted 26 March 2013
Published online 22 April 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24303

disordered linker domains was conserved in the face of negligible amino acid sequence conservation.[21] This study represented the first experimental test of the evolutionary conservation of dynamic behavior for a family of IDPs and demonstrated that amino acid sequence conservation was not required for the conservation of dynamic behavior and presumably molecular function. The intrinsically disordered linker domains used in the previous study acted as a simple tether between two ordered domains and did not have any known binding partners. This is a commonly observed function of IDPs.

Another commonly observed function of IDPs is to fold upon binding to various ligands be it protein or DNA.[1,28–30] In particular, many IDPs form amphipathic helices when they bind to other proteins. The transactivation domain of the tumor suppressor p53 (p53TAD) is an example of an IDP that forms an amphipathic helix when it binds to the murine double minute 2 (MDM2) oncoprotein, or the 70 kDa subunit of replication protein A.[31–35] MDM2 forms a complex with p53 that inhibits transactivation. MDM2 is also an E3 ubiquitin ligase that targets p53 for degradation by ubiquitinating lysine residues near the carboxyl terminus.[36] The function of the RPA70-p53 complex is less well characterized but recent evidence suggest it plays a role in regulating homologous recombination DNA repair.[37]

Previous NMR spectroscopy studies of human p53TAD identified transient secondary structure in the MDM2 and RPA70 binding sites.[31,33] This is a common observation for IDPs that fold when they bind to another protein.[28–30,38–40] It is expected that the amount of transient helical structure in the free IDP may be conserved for a particular IDP family if these regions are important for function. However, a direct test of this hypothesis has not been performed.

In the current study, p53TAD was chosen as a model to test the relationship between sequence identity and structural similarity for a closely related family of IDPs that have multiple binding partners and undergo coupled folding and binding reactions with these partners. The p53TAD homologues shown in Figure 1(a) have a range of sequence identities, shown in Figure 1(b), which varies from 70% between cow and rabbit down to 44% between dog and mouse. There is a high probability that the topology will be conserved when ordered proteins exhibit sequence identities in this range.[41–47] The results from our investigation suggest that the relationship between sequence identity and structural similarity for IDP families is not so straightforward.

NMR spectroscopy was used to measure the transient secondary structure and backbone dynamics for the p53TAD homologues shown in Figure 1(a). NMR chemical shifts and backbone dynamics data of the p53TAD homologues were correlated with the sequence identity matrix. The correlations uniformly suggest that the structural properties of IDPs are at least as sensitive to amino acid substitutions as ordered proteins. Our data show weak correlations between secondary structure and the sequence identity matrix. In contrast, the backbone dynamics of the homologues showed a relatively high correlation with the sequence identity matrix.

MATERIALS AND METHODS

Protein purification

Samples of human p53TAD (residues 1–73) that were unlabeled or uniformly labeled with either ^{15}N or ^{15}N and ^{13}C, were prepared as previously described.[33] Samples for p53TAD from dog (residues 1–77), mouse (residues 1–87), cow (residues 1–82), guinea pig (residues 1–88), and rabbit (residues 1–87) were prepared using this same method.

NMR data collection and analysis

Resonance assignments for human p53TAD and the homologues from Oryctolagus cuniculus (aka rabbit), Cavia porcellus (aka guinea pig) Canis lupus familiaris (aka dog), and Mus musculus (aka mouse) were previously reported.[33,48] In summary, NMR experiments on the p53TAD homologues 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 ^{1}H and ^{15}N as well as $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ resonance assignments, sensitivity enhanced ^{1}H-^{15}N HSQC and three dimensional HNCACB experiments were performed on the uniformly ^{15}N and ^{13}C labeled samples of the homologues in 90%H2O/10% D2O, PBS buffer, at a pH of 6.8. For the HNCACB experiment, data were acquired in ^{1}H, ^{13}C, and ^{15}N dimensions using 7225.4335 (t_3) × 12063.4480 (t_2) × 1499.9813 (t_1) Hz sweep widths, and 512 (t_3) × 128 (t_2) × 32 (t_1) complex data points. The sweep widths and complex points of the HSQC were 7225.4335 (t_2) × 1500 (t_1) Hz and 512 (t_2) × 128 (t_1), respectively.

^{1}H-^{15}N steady–state NOE experiments were recorded in the presence and absence of a 120° off-resonance ^{1}H saturation pulse every 5 ms for a total of 3 s. A total of 512 (t_2) × 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.

All NMR spectra were processed with nmrPipe and analyzed using nmrView.[49] The ^{1}H, ^{13}C, and ^{15}N dimensions were adjusted using a squared sine bell function shifted by 70°. This was followed by zero filling to twice the number of real data points. Linear prediction was used in the ^{15}N dimension of the HNCACB and HNCO.

a

```
Human        MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPDEAPRMP-EAAPRVAPAPAAPTPAAPAPAPS
Guinea Pig   MEEPHSDLSIEPPLSQETFSDLWKLLPENNVLSDSLSPPMDHLLLSPEEVASWLGE--NPDGDGHVS-AAPVSEAPTSAGPALVAPAPATS
Rabbit       MEESQSDLSLEPPLSQETFSDLWKLLPENNLLTTSLNPPVDD-LLSAEDVANWLNE--DPEEGLRVP-AAPAPEAPAPAAPALAAPAPATS
Cow          MEESQAELNVEPPLSQETFSDLWNLLPENNLLSSELSAPVDDLLPY-TDVATWLDE--CPNEAPQMP-EPSAPAAPPPATP-----APATS
Dog          MEESQSELNIDPPLSQETFSELWNLLPENNVLSSELCPAVDELLLP-ESVVNWLDE--DSDDAPRMP-ATSAPTAPGPAP----------S
Mouse        MEESQSDISLELPLSQETFSGLWKLLPPEDILPSPHC--MDDLLLP-QDVEEFFEG---PSEALRVSG-APAAQDPVTETPGPVAPAPATP
             ***.::: .:: ******** **:*** :::*.      :*. :    .:  ::     ..   ::.  .. .  * .           .
```

b

Sequence Identity Matrix for p53TAD

Human	100	59	63	61	56	53
Guinea Pig	59	100	70	56	56	51
Rabbit	63	70	100	67	63	57
Cow	61	56	67	100	68	49
Dog	56	63	56	68	100	44
Mouse	53	51	57	49	44	100

Sequence Identity Matrix for MDM2 Binding Site

Human	100	100	100	92	85	77
Guinea Pig	100	100	100	92	85	77
Rabbit	100	100	100	92	85	77
Cow	92	92	92	100	92	69
Dog	85	85	85	92	100	69
Mouse	77	77	77	69	69	100

Sequence Identity Matrix for RPA70 Binding Site

Human	100	47	50	44	38	63
Guinea Pig	47	100	63	50	63	53
Rabbit	50	63	100	69	67	50
Cow	44	50	69	100	60	50
Dog	38	63	67	60	100	47
Mouse	63	53	50	50	47	100

c

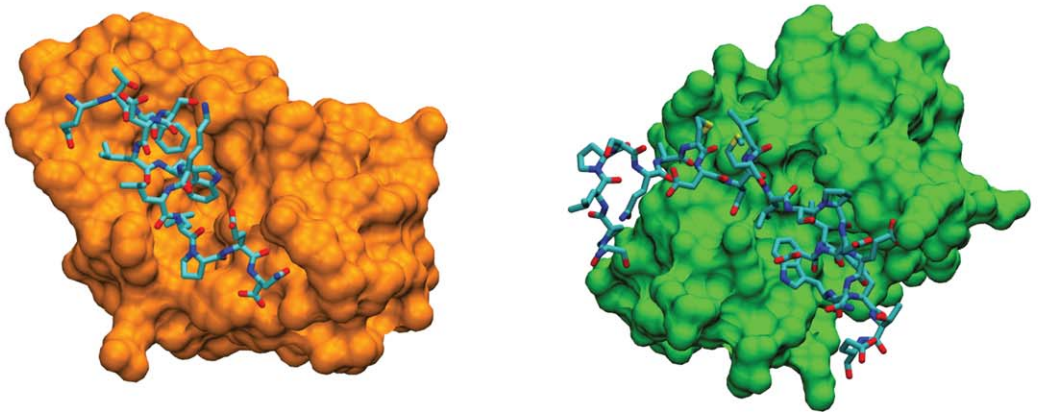

Figure 1

Sequence alignments and identity matrices for the p53 transactivation domain and poly-proline regions for the p53 homologues used in this study. This corresponds to residues 1–90 of human p53. **a**: Sequences are shown for Homo sapiens (aka human), Cavia porcellus (aka guinea pig), Oryctolagus cuniculus (aka rabbit), Bos taurus (aka cow), Canis lupus familiaris (aka dog), and Mus musculus (aka mouse). Orange text in the human sequence shows the position of the MDM2 binding site and green text shows the position of the RPA70 binding site. **b**: Sequence identity matrices for the full-length p53TAD homologues as well as the MDM2 and RPA70 binding sites. **c**: Surface images of MDM2 residues 17–125 bound to a peptide corresponding to residues 15–29 of human p53TAD is shown on the left, and on the right a surface image of RPA70 residues 1–120 bound to residues 41–59 of p53TAD is shown. PDB file 1YCR was used to generate the MDM2 image and 2B3G was used to generate the RPA70 image.[34,35]

The ^{1}H carrier frequency was set on the water peak, and 4.753 ppm was used as the reference frequency.

Correlation and clustering analysis

Pearson's r^2 correlation coefficient was used to compute pairwise correlation matrices for chemical shifts, NHNOEs, and secondary structure distributions (as determined by the d2D methodology), in order to test for similarities between the structural data and the sequence identity matrices of the homologues. The differences between these pairwise correlation matrices and the sequence identity matrices were determined by computing the Euclidian norm. Similarities between the structural data and the sequence identity matrices were also analyzed using a clustering approach. Principal component analysis (PCA) was used in combination with fuzzy c-means. The general algorithm for FCM was used as follows [2]:

set the data X with n data vectors consisting of d data points:

$$X = x_{ij},\ i = 1 \ldots n,\ j = 1 \ldots d;$$

set number of clusters: $2 \leq c \leq d$;
set degree of fuzziness: $w \in (1, \infty)$;
set the convergence threshold: δ (was used 10^{-6});

initialize the partition matrix $U = u_{ij} \in [0;1]$ with random values;

compute clusters' centers for each data vector:

$$c_{ij} = \frac{\sum_{j=1}^{d} (u_{ij})^w x_{nj}}{\sum_{j=1}^{d} (u_{ij})^w}, 1 \le i \le c$$

compute distance matrix D between centers using Euclidian norm and update partition matrix U:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{D(c_k, x_{nj})}{D(c_i, x_{nj})} \right)^{\frac{2}{w-1}}}, 1 \le i \le c, 1 \le j \le d$$

repeat steps and until convergence threshold is reached: $|U_{\text{step}} - U_{\text{step}-1}| \le \delta$

All calculations were performed in MATLAB.

RESULTS AND DISCUSSION

Evolution of transient helical secondary structure for p53TAD homologues

Figure 1(c) shows a surface image of MDM2 residues 17–125 (gold structure) bound to a peptide corresponding to residues 15–29 of human p53 and a surface image of RPA70 residues 1–120 (green structure) bound to a peptide corresponding to residues 40–57 of human p53. It is expected that the bound structures of the p53TAD homologues will be similar to the bound structures of the human homologue seen in Figure 1(c). This is a reasonable expectation, because their binding partners, MDM2 and RPA70, are very highly conserved, and as such their binding clefts should be nearly identical. Indeed, the percent identity for the both the RPA70 and MDM2 homologues is 85% or greater and there are no differences in the residues that make contact with p53TAD (data not shown).

Chemical shifts for the backbone atoms of the p53TAD homologues were measured to examine their secondary structure propensities at single residue resolution. Figure 2 shows plots of the alpha carbon secondary chemical shifts (CA$\Delta\delta$) for the p53TAD homologues. Shaded rectangles show the position of the MDM2 and RPA70 binding sites. CA$\Delta\delta$ values are sensitive to phi and psi backbone dihedral angles and provide a reliable estimate of transient secondary structure for IDPs.[50–52] Positive CA$\Delta\delta$ values are observed for residues in a helical conformation and negative values are observed for residues in a beta or extended conformation. Residue specific CA$\Delta\delta$ values for helical regions in ordered proteins range from 2–4 parts per million (ppm).[53] The smaller values observed in IDPs represent a population weighted average that can be used to estimate the precentage helicity.[50] For instance, the precentage helicity of the MDM2 binding site in human p53TAD is about 11.9% (calculated for residues 17–29). This value remains relatively constant for guinea pig (12.4%) and rabbit (12.5%) p53TAD and decreases for dog (5.7%), cow (5.7%), and mouse (3.4%). Inspection of the sequence alignment in Figure 1(a) reveals that the MDM2 binding sites (orange text in human sequence) are identical between human, guinea pig, and rabbit p53TAD but mouse, dog, and cow contain substitutions. For mouse, there is a glycine at position 21 compared to an aspartic acid for human, whereas dog has two substitutions compared to human, a glutamic acid at position 21 and an asparagine at position 24, and finally, cow also has an asparagine at position 24. The amino acid substitutions observed in the MDM2 binding site of p53TAD occur on the solvent exposed side of the amphipathic helix [see Fig. 1(c)] and should not interfere with any contacts made by residues at the binding interface.

Figure 2 also shows variation in the transient helical structure of the RPA70 binding site. In particular, guinea pig shows a strong preference for helical structure with a percentage helicity of 14.4% (calculated for residues 40–57). Rabbit has the next highest value at 3.7%. The percentage helicity values for human, mouse, cow, and dog p53TAD are close to zero. The sequences of the RPA70 binding site are more variable than the MDM2 binding site. As mentioned earlier, it is assumed that all of the p53TAD homologues will adopt similar topologies when bound to RPA70. This assumption is based on the high degree of sequence conservation observed for the corresponding RPA70 homologues.

Secondary structure predictors are not effective for p53TAD

In order to determine whether similar outputs are generated for the homologues, their amino acid sequences were analyzed using secondary structure prediction algorithms. The secondary structure prediction algorithms GOR IV and Jpred were used to analyze the sequences of the 6 p53TAD homologues.[54,55] Both algorithms successfully predicted some helical character in the MDM2 and RPA70 binding sites. The Jpred algorithm had a tendency to predict the helix for the MDM2 binding site 2–3 residues earlier than GOR IV. The CA$\Delta\delta$ plots show minimal helical character for these earlier residues. Although both predictors were able to predict some helical character in the RPA70 binding site, there was little correlation in how extensive the helices would be. Neither predictor was able to reproduce the more extensive helical characteristics observed in the guinea pig RPA70 binding site. These two predictors do a reasonable job of finding regions of transient helical structure but are not able to accurately define the extent of the regions. Additionally, the predictor AGADIR was

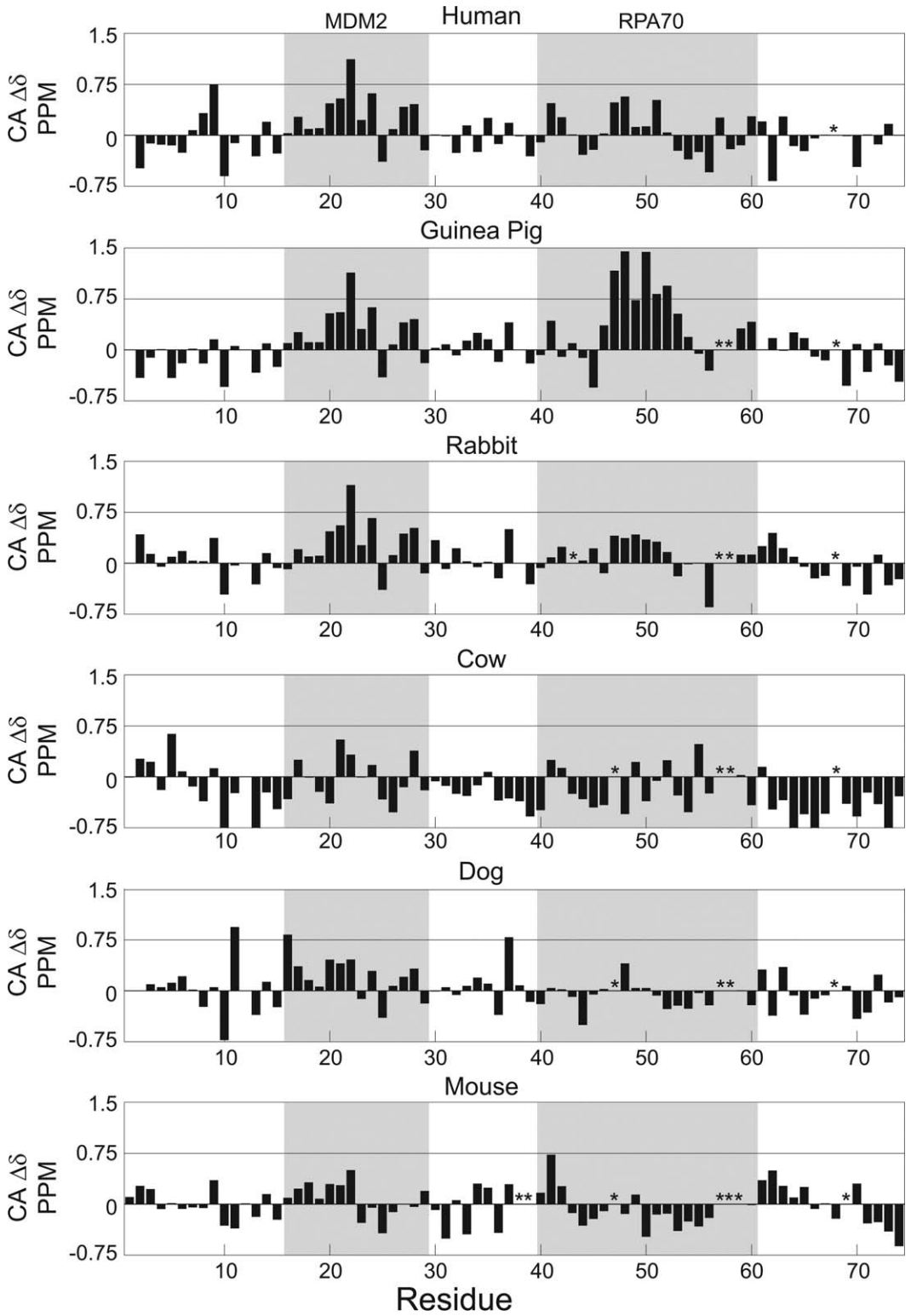

Figure 2

Evolution of transient helical secondary structure for the p53TAD homologues. Alpha carbon secondary chemical shifts (CAΔδ) are shown for **a**: human, **b**: guinea pig, **c**: rabbit, **d**: cow, **e**: dog, and **f**: mouse. Shaded boxes show the binding sites for MDM2 and RPA70 and the asterisks show the positions of gaps in the sequence alignment.

used to analyze these sequences; this predictor attempts to predict transient elements in solution, and gives a fractional output of each structural type for the overall sequence. The sequences of the full length polypeptides, as well as the isolated MDM2 and RPA70 binding regions, were tested and yielded the poorest results of all the predictors (data not shown).[56] These results highlight the necessity of increasing the available database of experimentally confirmed structural and dynamic data of IDPs.

Evolution of backbone dynamics for p53TAD homologues

NMR spectroscopy was also used to investigate the backbone dynamics of the p53TAD homologues. Figure 3 shows plots of the amide ^{1}H-^{15}N steady–state nuclear Overhauser effect (NHNOE). The NHNOE is sensitive to backbone motions on the nanosecond to picosecond timescale. This experiment is a reliable way to differentiate between ordered and disordered regions of a protein. NHNOE values for ordered proteins are positive and in the range of 0.8, reflecting relatively slow rotational motion of a compact, globular polypeptide. For IDPs, NHNOE values can be positive or negative. Positive NHNOE values are usually observed in regions that have transient secondary structure and negative values are observed in more flexible regions. It is important to note that NHNOE values will not be observed for proline residues because of their lack of an amide proton.

In general, there is reasonable agreement between the NHNOE values and the CAΔδ values for the MDM2 and RPA70 binding sites. Even the small changes in the transient helical structure of the MDM2 binding site are detectable in the NHNOE data. The mouse homologue has smaller positive values in this region relative to human and the dog homologue only has two residues with positive NHNOE values that are significantly greater than 0.02. The data presented in Figures 2 and 3 provide a view of the evolution of structure and dynamics for a family of closely related IDPs at single residue resolution and another demonstration that CAΔδ and NHNOE measurements are a sensitive indicator of changes in structure and dynamics for IDPs. Below we demonstrate that, even though the secondary structure and dynamics of the homologues shows some apparent concordance, a more detailed comparison of this data with the sequence identity matrix of the homologues reveals that the dynamic behavior is more conserved than secondary structure.

Distribution of secondary structure types in the p53TAD homologues

The d2D method, developed by Vendruscolo and coworkers, was used to calculate the probabilities of three different secondary structure types from the backbone chemical shifts of the p53TAD homologues.[57] Figure 4 shows the results from the d2D analysis of the p53TAD homologues. The probability for the alpha helix (blue bars), beta strand (red bars), and polyprolyl II helix (PPII) (green bars) secondary structure types are shown for all residues with chemical shift assignments. The remaining probabilities are in the random coil category. The sequences are shown along the *x*-axis and gaps are inserted based on a ClustalX alignment. Because the D2d method uses the chemical shifts from ^{1}H, ^{15}N, ^{13}CO, and ^{13}CB in addition to the ^{13}CA, it should provide a finer level of discrimination between the different secondary structure types than the CAΔδ analysis alone.

The data in Figure 4 shows more variation in secondary structure than would be expected based on the sequence identity between the homologues. Even the highly conserved MDM2 binding site shows big differences. For instance, the human homologue has six residues in the MDM2 binding site with an alpha helical probability between 0.09 and 0.30, whereas the mouse homologue has no residues in the MDM2 binding site with an alpha helical probably greater than 0.02. The RPA70 binding site shows more variation in secondary structure content than the MDM2 binding site. This is consistent with the greater variation in sequence identity for these regions but the differences cannot be explained based on this property alone. For instance, the sequence from L45-W53 for the guinea pig homologue shows a strong alpha helical signal but this signal is completely missing in the rabbit homologue. This is in contrast to the weak helical signal observed for the rabbit homologue based on just the CAΔδ values. This is interesting because the sequence identity of the RPA70 binding site for the two homologues is 63% and the most radical substitution in this region between the two sequences is at residue 47. In guinea pig this residue is a praline, whereas in rabbit it is an alanine. A proline would normally disrupt helical structure but in the guinea pig sequence the proline is in a position to aid in stabilizing an N-cap for the helix that forms when this region binds to RPA70. Also, the glutamic acid in position 49, has stronger helical proclivities than the aspartic acid in rabbit, this combination may explain why the helical signal for this homologue is so strong.[58]

A previous study on a construct of human p53, that included the transactivation domain and the proline rich region, estimated PPII content in the proline rich, and preproline regions of 66%, and 33%, respectively.[59] Our results using the d2D analysis show a maximum of about 12% PPII content at residue 70 of human p53TAD. We think the differences in PPII content between our study and the previous work are mostly due to differences in the size and composition of p53 fragments that were studied. The fragment used in the Fersht study contained p53 residues 1–92 and had a proline polymorphism at position 72, whereas our fragment contains residues 1–73

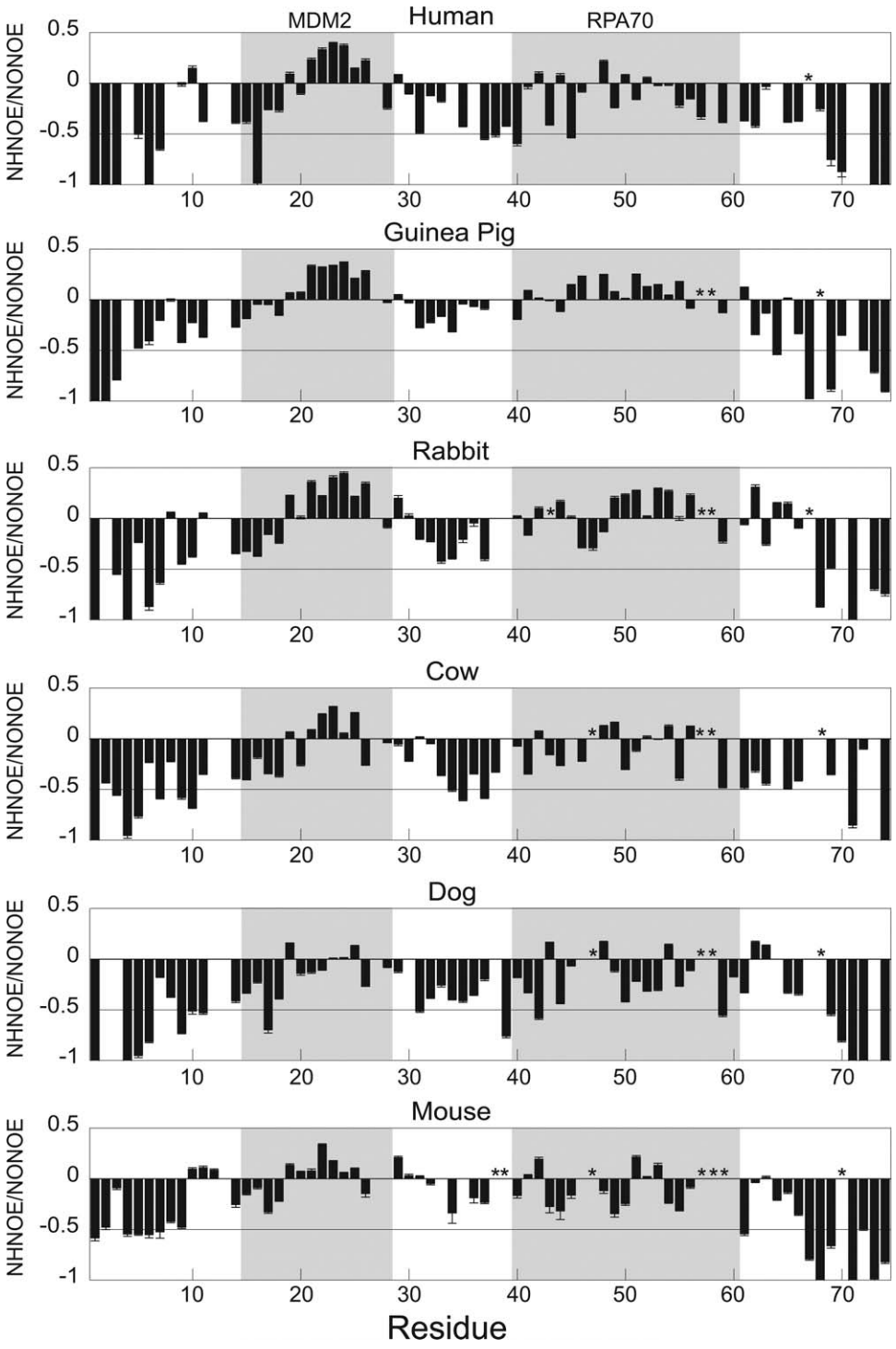

Figure 3

Evolution of backbone dynamics for the p53TAD homologues. ^{1}H-^{15}N steady–state NOE (NHNOE) values are shown for **a**: human, **b**: guinea pig, **c**: rabbit, **d**: cow, **e**: dog, and **f**: mouse. Shaded boxes show the binding sites for MDM2 and RPA70 and the asterisks show the positions of gaps in the sequence alignment.

and has an arginine polymorphism at position 72. As expected, our chemical shifts are nearly identical to the Fersht data for the pre-proline region, which corresponds to residues 60–70, and they are different for the residues surrounding the polymorphism. Using the d2D analysis on the chemical shifts from the Fersht study yields a slightly higher PPII content with a maximum of about 20% at position 70.[60] This value is lower than the 33%

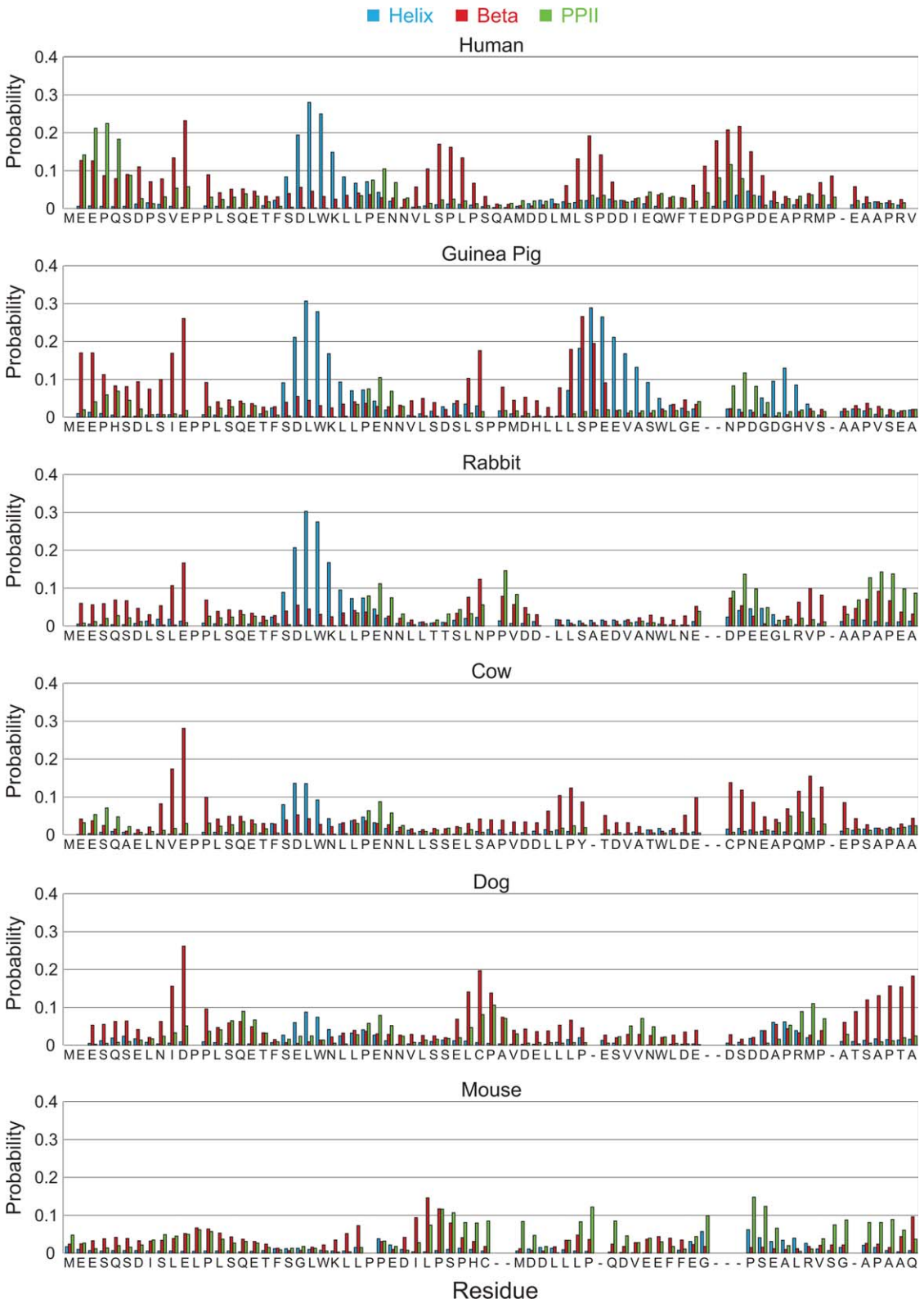

Figure 4

Distribution of secondary structure types for the p53TAD homologues. Probability of a particular secondary structure type is plotted on the y-axis for **a**: human, **b**: guinea pig, **c**: rabbit, **d**: cow, **e**: dog, and **f**: mouse. Data are shown for alpha helix (blue bars), beta strand (red bars), and PPII helix (green bars). The remaining probability is random coil.

reported from fitting Flexible Meccano models to RDC data.[59] It is unclear which approach is providing the more accurate estimate.

Correlating the structure and dynamics of the homologues with the sequence identity matrix

Pairwise correlation coefficients between the different homologues were calculated for the CAΔδ values, the chemical shifts of CA, beta carbon (CB), carbonyl carbon (CO), the amide nitrogen (N), and the amide proton (NH), the different secondary structure probabilities predicted by the d2D analysis, and the NHNOE values. This procedure results in a matrix of correlation coefficients for each chemical shift type, secondary structure category, or NHNOE data set that can be directly compared with the sequence identity matrix for the homologues.

For instance, the similarity between the correlation matrix for the NHNOE data from the homologues and the sequence identity matrix is shown in Figure 5(a). This figure shows the correlation coefficients for the NHNOE values below the diagonal and the correlation coefficients for the sequence identity values above the diagonal. In order to determine whether any of the experimental data sets reflect the phylogenetic relationships between the homologues, the Euclidian norms were computed between the correlation matrices and the percent identity matrix. Figure 5(b) shows a graph of these Euclidian norms. The Euclidian norm approaches zero as the correlation increases, that is, a lower value indicates a stronger correlation. According to this result, the backbone dynamics of the homologues is the data set with the highest similarity to the sequence identity matrix.

Evolutionary conservation of the backbone dynamics in mammalian p53TAD homologues

In order to further examine whether backbone dynamics, or any other structural features are conserved, principal component analysis (PCA) and fuzzy means clustering was performed on each of the correlation matrices. The clustering analysis also suggests that the backbone dynamics of the p53TAD homologues have the closest relationship to the percent identity matrix. The data sets were clustered using the aligned sequences of the mammalian homologues. Although there was some consistent clustering centered around the MDM2 and RPA70 binding sites for the CAΔδ values, as well as the chemical shifts of CA, CB, CO, and N(Supporting Information Fig. 1), the NHNOE data showed the most consistent and most extensive clusters [Fig. 6(a)]. Clustering of NHNOE data using the aligned sequences of mammalian homologues revealed regions of conserved dynamics, resulting in consistent clusters in the p53TAD sequence in the MDM2 and RPA70 interaction sites as seen in Figure 6(a). Figure 6(a) shows the result of clustering the

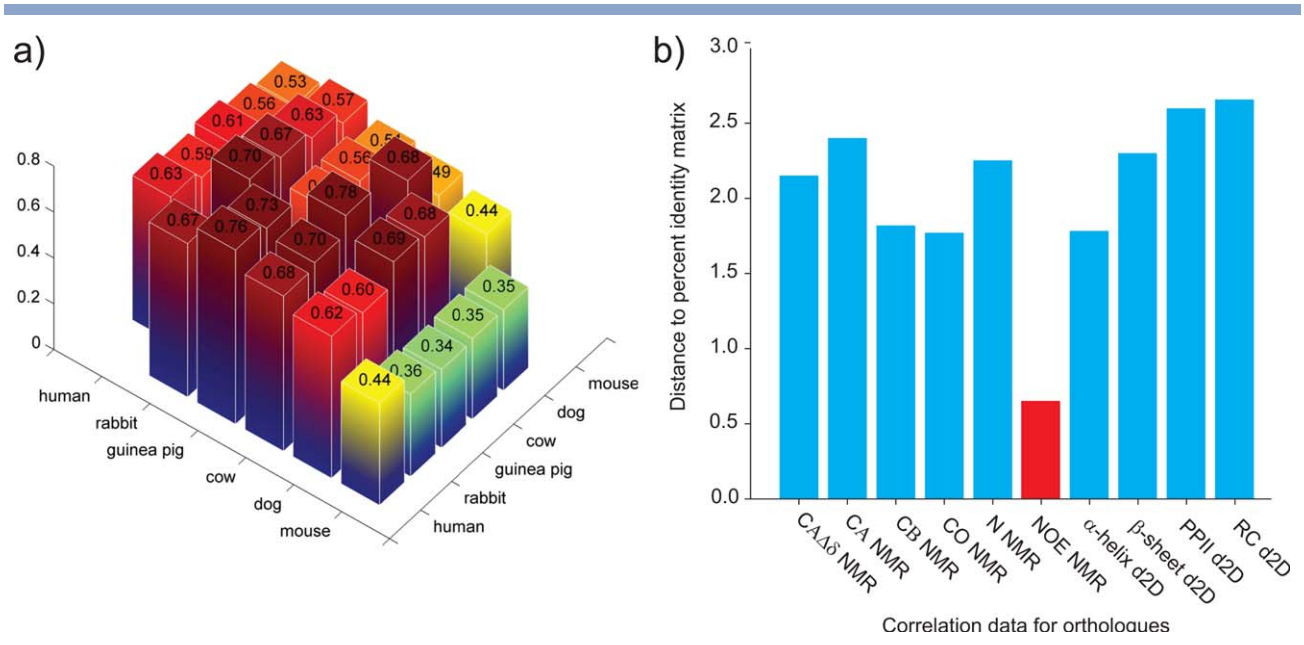

Figure 5

Correlation analysis for p53TAD homologues. **a**: Comparisons between the sequence identity matrix and correlation matrix from aligned NHNOE data. The values for the sequence identity matrix are shown above the diagonal and the values r^2 of the NHNOE correlation matrix are shown below the diagonal. Values on the diagonal were omitted for clarity. **b**: Euclidian norms between the correlation matrices of the different chemical shifts, secondary structure probabilities predicted by D2d, and the NHNOE values.

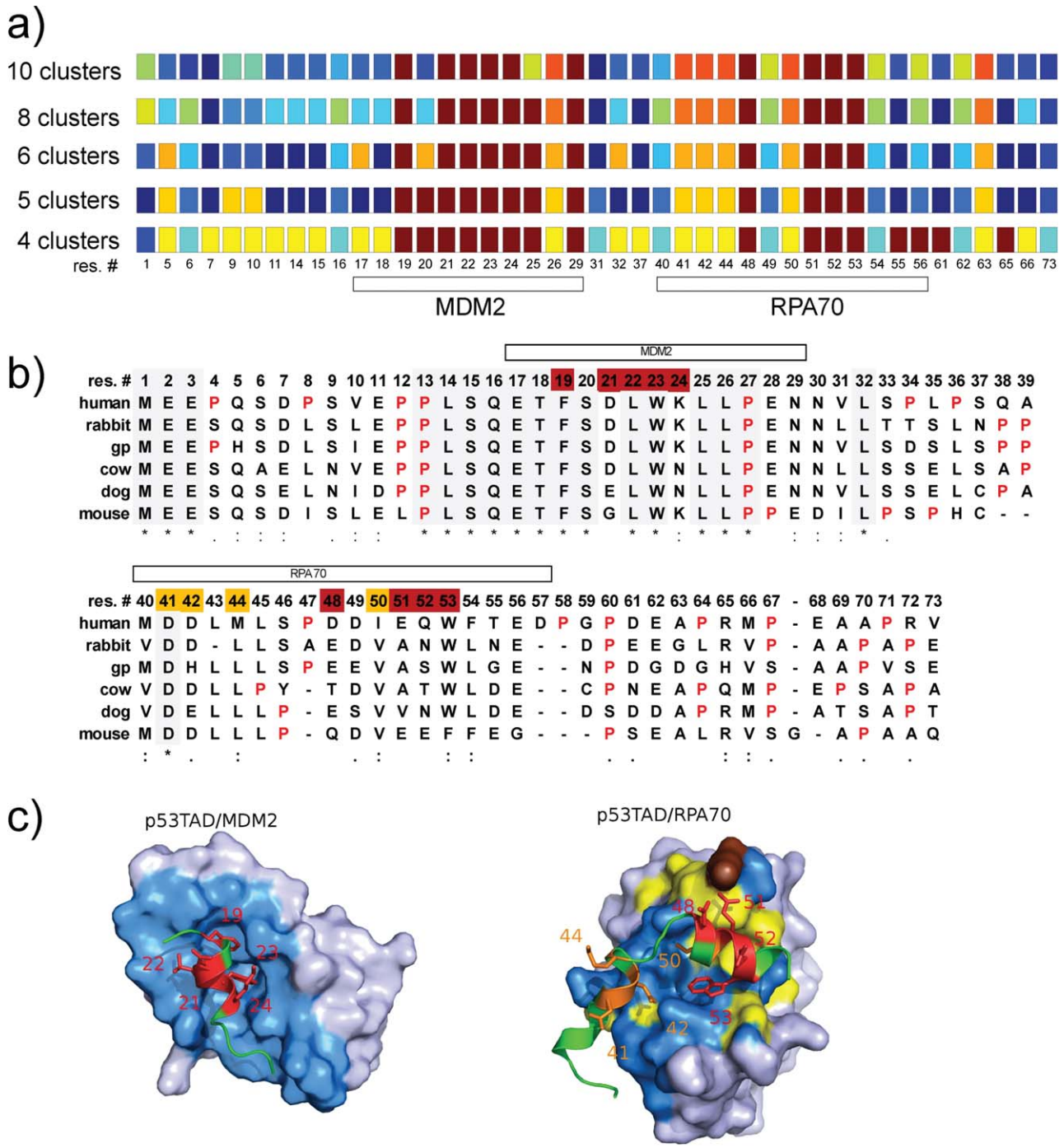

Figure 6

Clustering analysis of the aligned NHNOE data. **a**: Clustering of the aligned non-proline residues into 4,5,6,8,and 10 clusters. Each color represents a distinct cluster. Residue numbers correspond to human p53TAD sequence. **b**: Aligned sequences of mammalian homologues. Residues that cluster together within MDM2 and RPA70 binding sites are highlighted. **c**: Structures of human p53/MDM2 and p53/RPA70 complexes. Residues of human p53TAD which cluster together and fall within the binding site are highlighted in in Red and Orange according to their clustering from Figure 6(a). Binding pockets of both MDM2 and RPA70 are shown as surface areas with blue color. Known sites of specific p53TAD interaction in RPA70 are highlighted in yellow the loop containing K88 from RPA70 is marked in brown.

NHNOE data from the aligned mammalian sequences into 4, 5, 6, 8, and 10 clusters for aligned nonproline residues. Each color represents a cluster of residues, meaning the changes in the NHNOE value at that residue position changed in a similar manner between the homologues. If residues consistently cluster as the number of clusters increases then the correlation between the sequence and the NHNOE data is significant.

Figure 6(a) illustrates that consistent groups of clusters are clearly observable spanning residues 19, 21–24(I), 41,42,44,50 (II), 48,51,52,53 (III). As can be seen in Figure 6(b) which corresponds to the 10 cluster breakdown, group I (highlighted in red) is located within the MDM2 interaction site, whereas groups II (highlighted in yellow) and III (highlighted in red) are located in the interaction site of RPA70. This indicates the presence of two discrete groups of p53TAD residues interacting with RPA70. This behavior is typical for all homologues and thus preserved by evolution. Backbone dynamics of the residues in groups I and III are similar and so cluster together. The backbone dynamics of group II shows little similarity with groups I and III. The rest of the residues are more randomly clustered.

The MDM2 binding site is commonly considered as the region from residue 17–29, however, our clustering analysis outlines a narrower range spanning residues 19–24, which is in agreement with the NHNOE plots shown in Figure 3. These are the residues that fit into the binding pocket of MDM2 as depicted in Figure 6(c). In general, the MDM2 binding site is well conserved [Fig. 1(b)]. Nevertheless, there are notable differences in residues spanning 19–24. Residues at positions 21, for dog and mouse, and 24, for dog and cow, are substituted relative to the consensus. Despite these differences, the NHNOE data for these residues still clusters with the other residues in this range. It is reasonable to conclude that the backbone dynamics of this region is conserved for mammals even though some of the sites vary over time. It is also worth mentioning that these residues (21 and 24) do not interact with MDM2 directly but are still important for helix formation. Conversely, the clustering analysis shows that the backbone dynamics for residue S20 is not the same as the other residues that interact with MDM2 even though this residue is conserved.

The crystal structure of the human p53TAD/RPA70 complex revealed the existence of two helical regions on p53TAD, referred to as H1 and H2. Helix H2 is deeply embedded into the binding pocket and it plays an important role in the formation of the p53/RPA70 interface.[61] Our clustering analysis supports this conclusion. Residues that belong to group III (48, 51, 52, 53) reside in the same cluster as residues that interact with MDM2 from group I, which are also found to be deeply embedded when bound and are crucial for the p53-MDM2 interaction. The group III residues are also more conserved than the group II residues, and the residues outside both binding sites, though it is still not as well conserved as the MDM2 binding site. In the human p53TAD sequence this group contains residues D48, E51, Q52, and W53. Their position on helix H2 is depicted in red in Figure 6(c). It has already been reported that E51 and W53 in p53TAD stabilize the orientation of the H2 helix in the cleft of RPA70. Strong electrostatic interactions between D48 of p53 and K88 of RPA70 also stabilize the H2 helix in the binding cleft of RPA70.[34]

Helix H1 aligns with group II (residues 41, 42, 44, and 50) and clusters separately from H2, meaning its backbone dynamics do not correlate with the sequence variation in the same manner, but it is still conserved amongst the homologues. Previous NMR studies investigating p53TAD and RPA70 binding suggest that H1 plays a secondary role in the interaction.[61] The disparity in the backbone dynamics detected in our clustering analysis supports this proposed difference in functional behavior as seen in Figure 6(a). Positions of the residues in cluster II (D41,D42, M44, and I50) from the human sequence are shown on Figure 6(c) with light brown color. Although H1 plays a more secondary role in p53/RPA70 binding, altering the electrostatic interactions with RPA70, can still disrupt complex formation.[33,34]

CONCLUSIONS

In this study, the evolution of transient secondary structure and backbone dynamics in a closely related family of IDPs was investigated using NMR spectroscopy. The p53TAD homologues showed variation in their secondary structure and dynamics that could be directly related to amino acid substitutions. In particular, the MDM2 binding site of p53TAD was highly conserved. Three of the homologues studied had substitutions in this region. These substitutions all occur on the solvent exposed side of the amphipathic helix that forms when p53TAD binds to MDM2. However, the most common of these, K24N, does not appear to affect binding affinity, based on unpublished isothermal titration calorimetric data. The transient secondary structure and dynamics for the RPA70 binding site showed more variability than the MDM2 binding site. This behavior was consistent with the increased sequence variation observed for this region.

Clustering analysis of the NHNOE data revealed groups of residues that aligned well with helices that form when p53TAD is bound to either MDM2 or RPA70. Group I corresponded to the helix that forms when bound to MDM2. Groups II and III correspond to the H1, and H2 helices that form when bound to RPA70, respectively. It is interesting to note that these groups flank notable phosphorylation sites S20, and S46. Phosphorylation of these sites is associated with an apoptotic response to cellular stress.[62,63] The S46 is located in the

loop region between H1 and H2. The flexibility of this loop is important; it allows the backbone between the H1 and H2 helices to wrap around the binding pocket of RPA70. Conservation of the backbone dynamics for the residues surrounding phosphorylation sites may be necessary to maintain the phosphorylation-mediated damage response of p53.

It may be that the clusters represent local cooperative effects as they were observed in short regions corresponding to known binding sites that show measurable transient secondary structure. It is even tempting to make a connection between some long range cooperation between the clusters and prior evidence that the MDM2 and RPA70 binding regions make transient long-range contacts.[64,65] However, the data presented here are insufficient to make any definitive conclusions about this behavior.

Overall the clustering analysis shows greater divergence in the secondary structure data than expected based on sequence divergence. Nevertheless, strong correlations were found between the backbone dynamics and the sequence identity of the homologues. This result is consistent with a previous study from our group and suggests that the dynamic behavior of IDPs is a conserved evolutionary feature.[21]

ACKNOWLEDGMENTS

G.W.D. is funded by the American Cancer Society (RSG-07-289-01-GMC) and the National Science Foundation (MCB-0939014).

REFERENCES

1. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. Curr Opin Struct Biol 2011;21:441–446.
2. Liu J, Faeder JR, Camacho CJ. Toward a quantitative theory of intrinsically disordered proteins and their function. Proc Natl Acad Sci USA 2009;106:19819–19823.
3. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and combining predictors of mostly disordered proteins. Biochemistry 2005;44:1989–2000.
4. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004;337:635–645.
5. Nguyen Ba AN, Yeh BJ, van Dyk D, Davidson AR, Andrews BJ, Weiss EL, Moses AM. Proteome-wide discovery of evolutionary conserved sequences in disordered regions. Sci Signal 2012;5:rs1.
6. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK. Natively disordered proteins. In: Buchner J, Kiefhaber T, editors. Protein folding handbook, Vol. 3. Darmstadt: WILEY-VCH; 2005. pp 275–357.
7. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 2008;9 (Suppl 2):S1.
8. Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002;269:2–12.
9. Vendruscolo M. Determination of conformationally heterogeneous states of proteins. Curr Opin Struct Biol 2007;17:15–20.
10. Eliezer D. Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol 2009;19:23–30.
11. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208.
12. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999;293:321–331.
13. Tompa P. Structure and function of intrinsically disordered proteins. Boca Raton: Taylor and Francis Group; 2010. 331 p.
14. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK. Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol 2002;55:104–110.
15. Huntley M, Golding GB. Evolution of simple sequence in proteins. J Mol Evol 2000;51:131–140.
16. Lin YS, Hsu WL, Hwang JK, Li WH. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. Mol Biol Evol 2007;24:1005–1011.
17. Liu J, Zhang Y, Lei X, Zhang Z. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. Genome Biol 2008;9:R69.
18. Simon M, Hancock JM. Tandem and cryptic amino acid repeats accumulate in disordered regions of proteins. Genome Biol 2009;10:R59.
19. Tompa P. Intrinsically unstructured proteins evolve by repeat expansion. Bioessays 2003;25:847–855.
20. Toth-Petroczy A, Oldfield CJ, Simon I, Takagi Y, Dunker AK, Uversky VN, Fuxreiter M. Malleable machines in transcription regulation: the mediator complex. PLoS Comput Biol 2008;4:e1000243.
21. Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. J Mol Evol 2007;65:277–288.
22. Molloy RG, Ma WK, Allen AC, Greenwood K, Bryan L, Sacora R, Williams L, Gage MJ. *Aquifex aeolicus* FlgM protein exhibits a temperature-dependent disordered nature. Biochim Biophys Acta 2010;1804:1457–1466.
23. Keskin O, Jernigan RL, Bahar I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. Biophys J 2000;78:2093–2106.
24. Maguid S, Fernandez-Alberti S, Parisi G, Echave J. Evolutionary conservation of protein backbone flexibility. J Mol Evol 2006;63:448–457.
25. Maguid S, Fernandez-Alberti S, Echave J. Evolutionary conservation of protein vibrational dynamics. Gene 2008;422(1–2):7–13.
26. Maguid S, Fernandez-Alberti S, Ferrelli L, Echave J. Exploring the common dynamics of homologous proteins. Application to the globin family. Biophy J 2005;89:3–13.
27. Merlino A, Vitagliano L, Ceruso MA, Mazzarella L. Subtle functional collective motions in pancreatic-like ribonucleases: from ribonuclease A to angiogenin. Proteins-Structure Function and Genetics 2003;53:101–110.
28. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002;12:54–60.
29. Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol 2009;19:31–38.
30. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005;44:12454–12470.
31. Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH. Local structural elements in the mostly unstructured transcriptional activation domain of human p 53. J Biol Chem 2000;275:29426–29432.
32. Dawson R, Muller L, Dehner A, Klein C, Kessler H, Buchner J. The N-terminal domain of p53 is natively unfolded. J Mol Biol 2003;332:1131–1141.
33. Vise PD, Baral B, Latos AJ, Daughdrill GW. NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. Nucleic Acids Res 2005;33:2061–2077.

34. Bochkareva E, Kaustov L, Ayed A, Yi GS, Lu Y, Pineda-Lucena A, Liao JC, Okorokov AL, Milner J, Arrowsmith CH, Bochkarev A. Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. Proc Natl Acad Sci USA 2005;102:15412–15417.
35. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science 1996;274:948–953.
36. Momand J, Zambetti GP, Olson DC, George D, Levine AJ. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell 1992;69:1237–1245.
37. Serrano MA, Li Z, Dangeti M, Musich PR, Patrick S, Roginskaya M, Cartwright B, Zou Y. DNA-PK, ATM and ATR collaboratively regulate p53-RPA interaction to facilitate homologous recombination DNA repair. Oncogene 2012;257:1–11.
38. Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 2004;338:1015–1026.
39. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006;362:1043–1059.
40. Lee SH, Kim DH, Lee SH, Han JJ, Cha EJ, Lim JE, Cho YJ, Lee C, Han KH. Understanding Pre-Structured Motifs (PreSMos) in intrinsically unfolded proteins. Curr Prot Pep Sci 2011.
41. Al-Lazikani B, Jung J, Xiang Z, Honig B. Protein structure prediction. Curr Opin Chem Biol 2001;5:51–56.
42. Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001;294:93–96.
43. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. Embo J 1986;5:823–826.
44. Chothia C, Lesk AM. The evolution of protein structures. Cold Spring Harb Symp Quant Biol 1987;52:399–405.
45. Lesk AM, Chothia C. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. J Mol Biol 1980;136:225–270.
46. Lesk AM, Levitt M, Chothia C. Alignment of the amino acid sequences of distantly related proteins using variable gap penalties. Protein Eng 1986;1:77–78.
47. Petsko GA, Ringe D. Protein structure and function. In: Lawrence E, Robertson M, editors. Structure from Sequence. London: New Science Press Ltd; 2004. 195 p.
48. Daughdrill GW, Borcherds WM, Wu H. Disorder predictors also predict backbone dynamics for a family of disordered proteins. PLoS One 2011;6:e29207.
49. Johnson BA, Blevins RA. NMR view—A Computer-Program For The Visualization And Analysis Of Nmr Data. J Biomol NMR 1994;4:603–614.
50. Dyson HJ, Wright PE. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Adv Protein Chem 2002;62:311–340.
51. Wishart DS, Sykes BD. Chemical shifts as a tool for structure determination. Methods Enzymol 1994;239:363–392.
52. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 1991;222:311–333.
53. Wishart DS. Interpreting protein chemical shift data. Progress in nuclear magnetic resonance spectroscopy 2011;58(1–2):62–87.
54. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 1996;266:540–553.
55. Cuff JA, Barton GJ. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins 2000;40:502–511.
56. Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. J Mol Biol 1995;245:275–296.
57. Camilloni C, De Simone A, Vranken WF, Vendruscolo M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. Biochemistry 2012;51:2224–2231.
58. Aurora R, Rose GD. Helix capping. Prot Sci 1998;7:21–38.
59. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci USA 2008;105:5762–5767.
60. Camilloni C, De Simone A, Vranken WF, Vendruscolo M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. Biochemistry 2012;51:2224–2231.
61. Kaustov L, Yi GS, Ayed A, Bochkareva E, Bochkarev A, Arrowsmith CH. p53 Transcriptional activation domain—A molecular chameleon? Cell Cycle 2006;5:489–494.
62. Bulavin DV, Demidov ON, Saito S, Kauraniemi P, Phillips C, Amundson SA, Ambrosino C, Sauter G, Nebreda AR, Anderson CW, Kallioniemi A, Fornace AJ, Jr, Appella E. Amplification of PPM1D in human tumors abrogates p53 tumor-suppressor activity. Nat Genet 2002;31:210–215.
63. Hirao A, Kong YY, Matsuoka S, Wakeham A, Ruland J, Yoshida H, Liu D, Elledge SJ, Mak TW. DNA damage-induced activation of p53 by the checkpoint kinase Chk2. Science 2000;287:1824–1827.
64. Lum JK, Neuweiler H, Fersht AR. Long-range modulation of chain motions within the intrinsically disordered transactivation domain of tumor suppressor p53. J Am Chem Soc 2012;134:1617–1622.
65. Vise P, Baral B, Stancik A, Lowry DF, Daughdrill GW. Identifying long-range structure in the intrinsically unstructured transactivation domain of p53. Prot-Struct Funct Bioinformat 2007;67:526–530.