PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53

David F. Lowry, Amber Stancik, Ranjay Mann Shrestha, and Gary W. Daughdrill*

Department of Microbiology, Molecular Biology, and Biochemistry, University of Idaho, Life Science South Rm. 140, Moscow, Idaho 83844-3052

ABSTRACT

Internuclear distances derived from paramagnetic relaxation enhancement (PRE) data were used to restrain molecular dynamics simulations of the intrinsically unstructured transactivation domain of the tumor suppressor protein, p53. About 1000 structures were simulated using ensemble averaging of replicate molecules to compensate for the inherent bias in the PRE-derived distances. Gyration radii measurements on these structures show that the p53 transactivation domain (p53TAD) is statistically predominantly in a partially collapsed state that is unlike the open structure that is found for p53TAD bound to either the E3 ubiquitin ligase, MDM2, or the 70 kDa subunit of replication protein A, RPA70. Contact regions that potentially mediate the collapse were identified and found to consist of mostly hydrophobic residues. The identified contact regions preferentially place the MDM2 and RPA70 binding regions in close proximity. We show that our simulations thoroughly sample the available range of conformations and that a fraction of the molecules are in an open state that would be competent for binding either MDM2 or RPA70. We also show that the Stokes radius estimated from the average gyration radius of the ensemble is in good agreement with the value determined using size exclusion chromatography. Finally, the presence of a persistent loop localized to a PXP motif was identified. Serine residues flanking the PXP motif become phosphorylated in response to DNA damage, and we postulate that this will perturb the equilibrium population to more open conformations.

Proteins 2008; 71:587–598.
© 2007 Wiley-Liss, Inc.

Key words: intrinsically unstructured protein; paramagnetic relaxation enhancement; molecular modeling; tumor suppressor; dynamic structure; molecular function.

INTRODUCTION

Intrinsically unstructured proteins and protein domains (IUPs) participate in many essential biological processes.[1–10] Current efforts directed at identifying general relationships between their dynamic structure and molecular function are expanding and challenging the study of protein structure–function relationships. IUPs are thought to form a broad, heterogeneous ensemble of structures that undergo conformational fluctuations on multiple time-scales, and it is hypothesized that the molecular function of their dynamic structures depends on flexibility in the same way the molecular function of compact globular proteins depends on relative rigidity.[11] IUPs form linkers that regulate distances between structured domains, they mediate the specificity of protein/protein and protein/DNA interactions, and they can remodel their structures to interact with multiple binding partners. Based on empirical studies and computational analyses, it is now clear that proteins with intrinsically unstructured regions longer than 50 amino acids are common and that IUPs form a diverse set of protein families that can elicit biological function using a variety of mechanisms. Indeed, some groups have even suggested that IUPs represent the primordial or original structures in protein evolution.[12]

Despite compelling evidence for the biological significance of IUPs there are currently only a handful of three-dimensional models of their structures and no published work that investigates the conservation of dynamic structure for a homologous group of IUPs.[13–17] This is because IUPs form a rapidly interconverting ensemble of structures and they are resistant to crystallization as well as structure determination using traditional NOE (nuclear Overhauser effect) based NMR approaches.[1,18] In the absence of reliable atomic models for IUP structures, our ability to identify general relationships between their dynamic structure and molecular function is hindered. For instance, it is unclear whether the higher frequency of amino

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat

Grant sponsor: ACS; Grant number: IRG7700323; Grant sponsor: NIH; Grant numbers: P20 RR 16448, P20 RR 16454-02.

*Correspondence to: Gary W. Daughdrill, Department of Microbiology, Molecular Biology, and Biochemistry, University of Idaho, Life Science South Rm. 140, Moscow, ID 83844-3052.
E-mail: gdaugh@uidaho.edu

Received 30 April 2007; Revised 29 June 2007; Accepted 10 July 2007
Published online 30 October 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21721

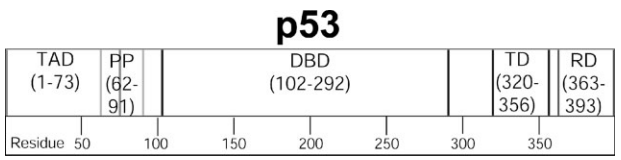

Figure 1

Schematic showing the domain structure of p53. DBD stands for DNA binding domain, TD is the tetramerization domain, RD is the regulatory domain, TAD is the transactivation domain, and PP is the polyproline domain. Amino acid numbering is relative to the human homologue.

acid substitutions observed for some IUPs is correlated with functional variation.[19] In addition, we do not know whether IUPs in the same functional families will adopt similar dynamic structures. In the absence of this basic information, it is impossible to predict the dynamic structure and molecular function of IUPs based on sequence data.

We have developed a project to investigate the structure and dynamics of the intrinsically unstructured transactivation domain of the human tumor suppressor protein, p53. The p53 protein regulates the cell cycle in response to a variety of stress signals through direct transcriptional activation or repression of specific target genes.[20–22] The specific stress signal as well as the cell type and context will dictate whether p53 induces transient growth arrest, senescence, or apoptosis. It also appears that p53 can play a direct role in DNA repair and replication.[23–25]

p53 has five structurally distinct domains with different functions (Fig. 1). From the N- to C-termini, these domains are (1) the transactivation domain (TAD; residues 1–73), (2) the polyproline subdomain (PP; residues 62–91), (3) the sequence specific DNA binding domain (DBD; residues 102–292), (4) the tetramerization domain (TD; residue 320–356), and (5) the regulatory domain (RD; residue 363–393). The TAD is responsible for regulating the transcriptional activity and the cellular stability of p53.[26] The PP contains five partially conserved PXXP repeats and is necessary for efficient growth suppression.[27] The DBD is responsible for site-specific DNA binding and is the domain where many cancer-associated mutations are located.[28,29] The TD appears to be necessary for the phosphorylation of certain sites in the TAD, but not for sequence-specific DNA binding.[30] The RD is thought to influence the function of the DBD.[22]

The molecular structures of the DBD and TD have been determined using NMR spectroscopy and X-ray crystallography.[31–37] These structures have provided important insights into the function of p53. In particular, the crystal structure of the DBD has provided a mechanistic explanation of many cancer-associated mutations. Conversely, the molecular structures of the TAD and RD are not known, which limits a detailed mechanistic explanation of mutations identified in these domains. It is known that ~50% of full length p53 is intrinsically unstructured.[38,39] Two groups have previously shown that p53TAD is one such unstructured region.[40–42]

The p53TAD contains 10 phosphorylation sites and numerous protein interaction sites. The presence of multiple sites for posttranslational modifications and protein interactions in a small domain like p53TAD is a hallmark of many IUPs.[43] Two of the protein binding sites present in p53TAD are for the E3 ubiquitin ligase, MDM2, and the 70 kDa subunit of replication protein A, RPA70. When bound to MDM2, p53 becomes ubiquitinated and targeted for proteosome-mediated degradation. When bound to RPA70, p53 may be transiently stabilized and available to amplify the cellular response to DNA damage.[42] In contrast to the important role p53TAD has in regulating p53 function, very little is know about its structure. As mentioned earlier, this is because p53TAD is intrinsically unstructured.

In this report, we have estimated internuclear distances for p53TAD based on paramagnetic relaxation enhancement (PRE) data. These distances were used to restrain torsion angle dynamics and calculate an ensemble of structures that represent the accessible conformations of p53TAD. Individual structures in the ensemble consistently show preferential association of the MDM2 and RPA70 binding regions. In addition, the ensemble of p53TAD structures was used to calculate amide nitrogen distance distributions over a six-residue length scale. All distributions show a bimodal character that may be unique to IUPs, and this analysis may be useful in future studies of IUPs to identify family membership. The amide nitrogen distance distributions also show the ensemble of p53TAD structures recovers previously identified features that are important for interactions with other proteins and also points to new regions that appear to be preferentially compact and are known to play a role in regulating the activity of p53.

MATERIALS AND METHODS

Protein purification and NMR data collection

Expression and purification of wild type and the cysteine mutants of p53TAD was previously described.[42,44] Four mutant proteins were used to generate PRE data for this study. Each mutant contained one cysteine substitution each at sites 7, 28, 39, and 61. All NMR experiments were performed as previously described using a Varian Inova spectrometer operating at a ^{1}H resonance frequency of 600 MHz and a sample temperature of 298 K.[42,44]

Resonance assignments and spectral analysis

The backbone resonance assignments of wild type p53TAD were reported previously.[42] Only small chemical

shift changes were observed for the resonances in the ^{15}N HSQC (heteronuclear single quantum coherence) spectra of the D7C, A39C, and D61C variants of p53TAD when compared with wild type. We have made the amide nitrogen and proton resonance assignments for these mutants based on the assumption of a minimal perturbation in ^{15}N HSQC spectrum. Whenever there is any ambiguity in the assignment of resonances using this principle, as was the case for E28C, the resonance assignments are made using the standard triple resonance methods that were applied to wild type p53TAD.[42] In the case of E28C, analysis of C_α and C_β chemical shifts was used to determine whether significant changes in the ^{15}N HSQC spectrum are accompanied by structural changes. This was not the case for E28C.

MTSL labeling and NMR data analysis

To induce PRE, the spin label MTSL is covalently attached to the free thiol of the cysteine residues in the p53TAD mutants. The cysteines are maintained in their reduced forms using a buffer containing dithiothreitol (DTT). The DTT is rapidly exchanged using a gravity flow-desalting column (PD10) and the MTSL is added directly to the protein fractions and incubated at room temperature for 1 h. Excess MTSL is removed using another PD10 column and the MTSL-labeled protein is concentrated for NMR analysis. For all the samples, the extent of the MTSL labeling is verified to be greater than 95% using mass spectrometry.

Three ^{15}N HSQC spectra are collected on each cysteine mutant. One spectrum is collected on the MTSL-labeled mutant, one spectrum is collected on the MTSL-labeled mutant after the spin label is reduced with ascorbic acid, and one spectrum is collected on a concentration-matched sample of the cysteine mutant without the spin label. The resonance intensity quotient is determined by measuring the peak intensities from all the three experiments and dividing the MTSL intensities by the reduced and unlabeled resonance intensities. Comparison of the resonance intensity quotients from the labeled and unlabeled cysteine mutants verified that a sufficient amount of ascorbic acid was added to reduce the spin label. Both resonance intensity quotients were identical to within ~5%. The resonance intensity ratios for the four cysteine mutants are presented in the supplementary materials.

Detecting interactions between free MTSL and wt p53TAD

A number of tests were performed to ensure that labeling the cysteine mutants with a hydrophobic molecule like MTSL did not induce or stabilize the long-range contacts that were observed. Free MTSL was titrated with a ^{15}N-labeled sample of wt p53TAD. Following the addition of free MTSL, a ^{1}H–^{15}N HSQC spectrum was collected to detect any chemical shift and resonance intensity changes. The presence of free MTSL induced mostly global and some local intensity changes. The free MTSL was subsequently reduced using an excess of ascorbic acid and another ^{1}H–^{15}N HSQC spectrum was collected. In this HSQC, some local chemical shift changes were observed. Most notably, combined ^{1}H and ^{15}N chemical shift changes between 0.02 and 0.05 ppm were observed for M1, V10, E11, W23-L26, L32, M44, L45, E51, Q52, T55, and V73. This result suggests that free MTSL may be binding to multiple sites on p53TAD. To ensure that these potential interactions were not the result of tight binding between MTSL and p53TAD, the sample was dialyzed back into the original buffer and a final ^{1}H–^{15}N HSQC spectrum was collected. In this spectrum, all chemical shifts returned to their original positions. The data are presented in the supplementary materials section.

Because some of the amide ^{1}H and ^{15}N chemical shift changes observed for p53TAD in the presence of free MTSL were similar to those observed in the MTSL-labeled cysteine variants, an independent test of the effect the MTSL has on the structure of p53TAD was developed. In this test, size exclusion chromatography (SEC) was used to measure the Stokes radius (R_s) of E28C before and after MTSL labeling. E28C was chosen for this test, because it had the largest chemical shift changes in the ^{1}H–^{15}N HSQC spectra for the MTSL labeled and unlabeled samples, relative to wild type. Within the error of the experiment (~0.02 nm), the R_s measurements for E28C before and after MTSL labeling were identical when compared with wt p53TAD.

Calculation of distances

To convert the PRE data into distances, the integrated intensities of resolved HSQC resonances were measured. The resonance intensity quotient for a particular residue was calculated as I_{ox}/I_{red}, where I_{ox} is the resonance intensity in the oxidized state and I_{red} is the resonance intensity of either the MTSL-labeled protein following the addition of ascorbic acid or the unlabeled protein. ^{15}N spin relaxation, longitudinal ^{1}H relaxation, and the difference between in-phase and anti-phase transverse ^{1}H magnetization were ignored, because these effects contribute only small corrections to the calculated electron–proton target distances. These approximations, together with a measurement of the intrinsic proton transverse relaxation rate in the reduced state, allow a fit of the contribution of the spin label to the proton transverse relaxation rate.[45] In our study, the intrinsic proton transverse relaxation rate was measured by fitting the detectable HSQC resonances to a Lorentzian function and calculating the average value (the individual linewidths are presented in the supplementary materials). In the HSQC experiments, the proton was transverse during the INEPT delays for a total of 9.8 ms.

Once the paramagnetic contribution to proton transverse relaxation is determined, the electron–proton distance is calculated using standard relaxation theory, assuming that the electron–proton correlation time is similar to the average overall rotational correlation time of the proton–nitrogen vectors obtained from reduced spectral density mapping of the ^{15}N relaxation data for the WT protein.[42] The overall correlation time for individual residues varied from 3 to 5 ns. However, the calculation of proton–electron distances is insensitive to changes in correlation time of 50%.[14] Based on our previous study, the average rotational correlation time from the reduced spectral density mapping was 3.3 ns.[42]

PRE-derived distance restraints

Two hundred seven distance restraints for structure calculations were generated as follows. Two classes of restraints were used. For residues that were determined to be less than 20 Å from the spin label, the restraint was a square-well potential centered on the target distance and 10 Å in width. For residues greater than 20 Å, a step potential was used with step at 20 Å and no upper bound. The distance was restrained from the C_α of the corresponding spin-labeled cysteine to the amide proton of residues that experience PRE. A C_α–H_N restraint was used because a backbone model of the protein was used in the simulated annealing step of the molecular dynamics.

Molecular dynamics simulations

The PRE-based distance measurements were used to calculate an ensemble of p53TAD structures. Structures were calculated with the program XPLOR-NIH 2.14 on a cluster supercomputer running 24 900 MHz and 39 1.2 GHz AMD Athlon Thunderbird processors. About 1000 structures were calculated starting from random extended structures. Torsion angle dynamics were initiated at 3500 oK and annealed to 100 oK in 15,000 steps. The energy term during annealing consisted of bond lengths, bond angles, impropers, steric repulsion, and PRE-derived distance restraints. The annealing is followed by 10,000 steps of torsion angle minimization and all-atom Cartesian minimization. There were no violations of the PRE-derived distance restraints for the 1000 member ensemble. However, three of the PRE-derived distance restraints could not be satisfied with the backbone model that was used during simulated annealing. These three restraints were between the C_α of spin label sites 7, 28, and 61 and the amide nitrogens at residues 9, 29, and 62, respectively. These unsatisfied restraints between near neighbors occur because the lower bound for the potential well is longer than the actual distance. Rather than remove these restraints from the simulations, these interatomic distances were restrained to be as long as the covalent bond geometry would allow. An analysis of the Ramachandran plots for all residues except glycines and prolines from the 984 structures in the ensemble showed that 31% of the dihedral angles were in the most favorable regions, 49% were in additionally allowed regions, 14.6% were in generously allowed regions, and 5.4% were in disallowed regions.

Statistical analyses of structures

GROMACS was used for clustering structures based on backbone RMSD (root mean square distance) and the distance between groups of residues, for measuring the gyration radii of structures and for generating residue contact maps that were averaged over the ensemble of structures.[46] Molecules were viewed and structure figures were made using VMD 1.8.4.[47]

Stokes radii measurements

Stokes radii for p53TAD and E28C were determined using SEC. A prepacked 16/60 Superdex 70 column with a V_o of 39.81 mL was attached to an AKTA FPLC system and equilibrated with five column volumes of 50 m*M* Tris-HCl (pH 7.5) and 300 m*M* KCl. Molecular weight standards were purchased from Sigma (MW-GF-70) and included aprotinin, bovine serum albumin, carbonic anhydrase, and cytochrome *c*. V_e was determined for each of the standards at a flow rate of 0.14 mL/min. A standard curve was generated by plotting log M_W for the standards versus V_e/V_o. Least squares fitting to the data of a linear function gave a standard equation with a correlation coefficient of 0.997. V_e was also determined for p53TAD, E28C, and E28C labeled with MTSL. Three replicate experiments were run for each sample and the average V_e values were used to calculate an apparent molecular weight and this was used to calculate the Stokes radii (R_s) based on an empirical relationship between the two quantities.[48] Average R_s for all three proteins were almost identical but did show a small trend toward compactness going from p53TAD to E28C. The R_s values in nanometer were 2.38 $\pm$ 0.02 for p53TAD, 2.36 $\pm$ 0.02 for E28C, and 2.36 $\pm$ 0.02 for E28C labeled with MTSL.

RESULTS AND DISCUSSION

Ensemble refinement and total number of structures

A major goal in the study of IUPs is to characterize the range of conformations consistent with the experimental data. To ensure that the distribution of conformations accurately reflects the experimental data, molecular dynamics simulations must proceed with restraints that are either time averaged over a single molecule or averaged at an instant in time over an ensemble of noninteracting molecules. Otherwise, the distribution of conformations tends to be shifted toward more compact

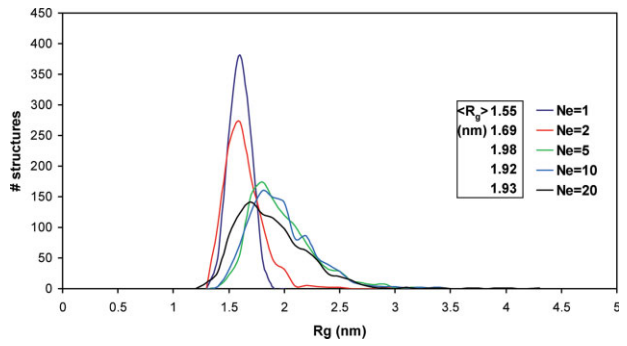

Figure 2

Smoothed histograms showing the distribution of R_g values for 1000 structures calculated using torsion angle dynamics restrained by PRE based distances and averaged over ensemble sizes (N_e) of 1, 2, 5, 10, and 20. The inset shows the average R_g values for each distribution.

structures.[16] In this report, ensemble averaging of noninteracting molecules was used based on a protocol developed by Clore and Schwieters,[49,50] with the exception that the torsion angle dynamics during the time-consuming annealing steps were restricted to two degrees of freedom per residue on the polypeptide backbone.

To obtain a faithful representation of the statistical distribution for a particular IUP, an appropriate size for the ensemble of noninteracting molecules and the total number of structures must be determined. It is expected that the ensemble size and total number of structures can be minimized to save time, but their optimum values depend on the size and sequence of the IUP and the number and nature of the restraints used. A report on the denatured state of bovine acyl-coenzyme A binding protein, which is almost double the size of p53TAD, showed that an ensemble size of 20 and 40,000 total structures was sufficient to sample the available distribution of structures.[16] In this report, optimal values for ensemble size and total number of structures were determined by increasing these parameters until there was no large change in the distribution of gyration radii (R_g). This approach addresses the problem of over-restraining, which would lead to an R_g distribution that is too narrow.[51] To some extent, the problem of under-restraining is addressed by the repulsive terms in the energy function applied during simulated annealing, but this may not be as effective as a full molecular dynamics force field that includes electrostatics.

First, to determine an appropriate ensemble size for the study of p53TAD, ~1000 total structures each were calculated with ensemble averaging using ensemble sizes of 1, 2, 5, 10, and 20 noninteracting molecules. Smoothed histograms of the R_g distributions for the structures from these calculations are shown in Figure 2. As expected, the mean R_g shifts to higher values as ensemble size is increased, but averaging over more than five molecules does not significantly change the average R_g value. Based on this comparison an ensemble size of five was chosen for this study.

Second, a determination of an appropriate total number of structures was made. An ensemble size of five was used and the total number of structures was increased from 1000 to 10,000 (final total numbers are different because some nodes on the cluster crashed during the structure calculations). The R_g distribution of the 1000 structures calculation was compared with that of the 10,000 structures calculation. It was found that the only difference between these two distributions was that for 10,000 structures, the distribution was smoother, with a slightly better defined tail at high R_g values. Lower moments of the distributions were nearly identical since the curves nearly superimpose on each other (Fig. 3). Consequently, for this study, 1000 total structures were sufficient to characterize the statistical distribution of structures for p53TAD.

Cluster analysis

There is currently no consensus on how to proceed with the analysis of an ensemble of structures that are likely to have limited structural similarity to one another. Clustering of structures in the ensemble based on the RMSD of backbone atoms is one procedure that a previous group performed to identify persistent structural features in the ensemble.[17] Cluster analysis was performed on the 1000 structures that were calculated by averaging the PRE distance restraints over five member ensembles. The resulting number of clusters and the number of structures per cluster was sensitive to the value chosen for the backbone RMSD cutoff. A cutoff of 0.88 nm resulted in a large number of sparsely populated clusters. A cutoff 0.90 nm resulted in a few heavily

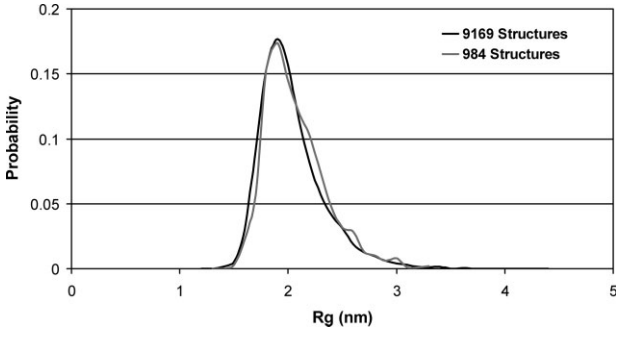

Figure 3

Overlay of smoothed histograms showing the distribution of R_g values for either 1000 (gray) or 10,000 (black) structures calculated using torsion angle dynamics restrained by PRE based distances that were averaged over an ensemble size of five. The final total number of structures differs because some nodes on the cluster failed during the calculations.

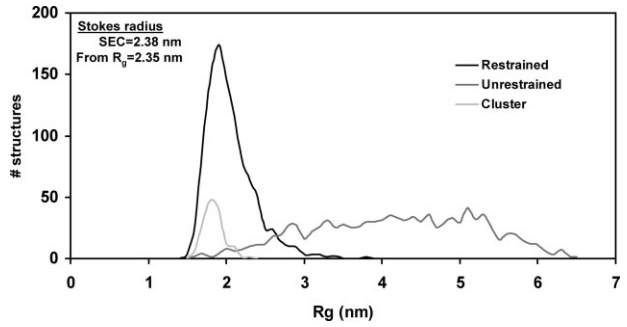

Figure 4

Smoothed histograms showing the distribution of R_g values for 1000 structures resulting from restrained dynamics simulations using the PRE based distances (restrained), a model for the random coil structure based on 1000 structures calculated without the PRE based distance (unrestrained) and the major cluster of PRE restrained molecules using a backbone RMSD cutoff of 0.889 nm (cluster). For all calculations, PRE restraints were averaged over five member ensembles.

populated clusters and a remainder of sparsely populated clusters. A cutoff of 0.89 nm resulted in five clusters with more than 10 members. The cluster sizes were 148, 34, 26, 20, and 20 structures of 1000 structures, which represent 25% of the total. The small number of clusters with more than 10 members indicates that the ensemble is rather broad and heterogeneous. Later in the manuscript, we will show that most of the heterogeneity appears to be localized to short segments of the protein.

Gyration radii for p53TAD structures

The radius of gyration (R_g) was calculated for the structures from the major cluster, the entire ensemble, and 1000 structures that represent a random coil model of p53TAD. The random coil model was generated by repeating the structure calculations without the PRE restraints. Smoothed histograms of these three R_g distributions are shown in Figure 4. A comparison of the restrained and unrestrained distribution shows that the range of R_g values is shifted from 1.41–3.72 nm to 1.59–6.42 nm and the average R_g is shifted from 2.0 to 4.1 nm. Similar structure calculations were performed on the IUP, alpha-synuclein.[13] In the alpha-synuclein study, average R_g values of 2.5 and 4.2 nm were observed for the restrained and unrestrained ensembles, respectively. This is interesting because alpha-synuclein is almost twice as long as p53TAD. Later, we describe some of the special sequence characteristics of p53TAD that may explain the larger than expected R_g values (even for an IUP). Figure 4 also shows that the major cluster is populated by more collapsed structures since its average R_g is shifted to 1.79 nm but that this cluster does not include the most collapsed structures since the range of R_g values in the cluster is 1.54–2.22 nm.

Stokes radii measurements

Recent studies on the Stokes radii (R_s) of native and denatured proteins permits the calculation of R_s based on the number of amino acids in the polypeptide chain.[52] Using the native state approximation, p53TAD is predicted to have an R_s value of 1.65 nm versus 2.55 nm in the denatured state. R_s can also be estimated using R_g based on a phenomenological relationship between the two quantities.[16] The estimated R_s values for p53TAD based on the average R_g values presented in Figure 4 are 2.35 nm for the PRE structures and 3.16 nm for the RC structures. The estimated R_s value of 2.35 nm for the PRE structures is in-between the predicted R_s values for p53TAD in the native and denatured states (see Fig. 4). This is the expected result for an IUP and alleviates some of our concerns about generating anomalously compact structures because of the inherent bias in the weighting function used to restrain the distances.[15,16] To partially validate the ensemble of structures calculated using PRE data, R_s was measured directly using SEC. The value for R_s obtained using this technique was 2.38 nm, which is in good agreement with the value estimated from the average R_g (see Fig. 4).

R_g distribution for the random coil model of p53TAD

The R_g distribution for the unrestrained structures shown in Figure 4 is broader and has a larger average value than expected for a random coil based on theoretical considerations.[53,54] The value of R_s calculated from the mean R_g value of the unrestrained structures is also greater than the value expected for a 73 residue polypeptide using the empirical relationship that was recently developed to estimate R_s for denatured proteins.[52] This appears to be because p53TAD is a 73 residue polypeptide that contains 13 prolines. It has two PXP motifs, one PXXP motif, one PP, and five single prolines. Prolines reduce flexibility and extend the polypeptide chain because of the ring structure formed between the side chain and backbone. To test how the presence of prolines influenced the distribution of R_g values for the random coil model of p53TAD, we performed unrestrained torsion angle dynamics on a sequence identical to p53TAD except all of the prolines were changed to alanines. The distribution of R_g values for the random coil model of p53TAD without prolines has an average value of 2.44 nm, which is much closer to the expected value for a theoretical random coil the size of p53TAD.

Representative structures from the ensemble

Space filling backbone models of representative structures taken from the restrained ensemble shown in Figure 4

are presented in Figure 5. These structures were chosen to span the observed range of R_g values. Shown above each structure is the R_g value for that structure as well as the number of structures that are present in the bin from which the structure was taken. Taken together, the 10 structures shown in Figure 5 are representative of 85% of the R_g values measured for the ensemble. For each structure shown in Figure 5, the binding regions for MDM2 (residues 17–26) and RPA70 (residues 40–54) are colored green. These two binding sites appear to maintain a close spatial proximity for structures with R_g values ranging from 1.54 to 2.22 nm. These data confirm an earlier report from our group on the presence of long-range structure in p53TAD.[44]

Contact maps as a function of R_g

To examine the predominant contacts in the p53TAD ensemble, contact maps were generated for structures in the different bins from the R_g distribution shown in Figure 4. The contact maps were based on the smallest distances between any two residue atoms for all residue pairs, and contacts were averaged over all structures in a particular bin. Contacts identified in the most collapsed structures (R_g = 1.6 nm) clearly persist in the ensemble up to an R_g of 2.0 nm. An average contact map was made using the 709 structures that span R_g values of 1.5–2.1 nm. A large cutoff for contacts of 1.86 nm was used since averaging over many structures shifts the threshold for off-diagonal patterns in the contact map. In this contact map, it is found that the PRE structures contain five predominant regions of contacts. These regions are between residues 16–23:37–41, 22–31:59–66, 25–31:50–55, 37–42:52–58, and 37–42:60–66. In addition, an average contact map was made using the 1000 structures from the unrestrained simulations. This contact map demonstrated that no long-range contacts form in the random coil model of p53TAD (data not shown).

Each of the regions that form the predominant contacts in the p53TAD ensemble contains one of three spin-labeled residues: residues 28, 39, and 61. The spin label at residue 7 did not generate any long-range restraints that might lead to a contact. This result provides an important internal control that ensures the spin label is not driving the formation of long range contacts. To ensure that the spin label did not stabilize the formation of existing long range contacts a number of tests were performed, including titrating free MTSL into an ^{15}N-labeled sample of wt p53TAD. A description of these additional tests is provided in the Materials and methods section. Based on the D7C result and the results from the additional tests it appears that the placement of the spin-labels was sufficient to detect the predominant long-range contacts, but their presence did not induce these contacts and only had minor effects on their stability.

Using the contact map to refine the cluster analysis

The regions containing the predominant contacts illustrated in Figure 6(a) were used to generate a new set of clusters. Clustering was performed using residues 16–31, 37–42, 50–66 and a backbone RMSD of 0.625 nm. Using these parameters, eight clusters were identified that contained 10 or more structures and the most popular cluster contained 62 structures. The eight clusters identified by excluding regions that do not give rise to the contacts shown in Figure 6 have a better convergence than the previous clusters, as evidenced by the lower RMSD between backbone atoms.

Distance distributions for the p53TAD ensemble

PRE analysis is not expected to restrain the molecule in a way that forms secondary structure, because at 1.0 nm the minimum length scale for the PRE interaction is too long. However, it is possible that long-range restraints might define collapsed or extended regions over a six-residue length scale. For typical secondary structure, turns and helices result in the amide nitrogen of residues 1 and 6 of a six-residue segment being between 0.5 and 1.0 nm apart. Extended regions or beta strands result in residues 1 and 6 being about 1.5 nm apart. Histograms showing the distribution of amide nitrogen distances for all amino acid pairs that are four residues apart were constructed for the 1000 member ensemble of p53TAD structures. Interestingly, all distributions of amide nitrogen distances between residues i:i + 5 were bimodal, with the lobes centered at ~1.1 and 1.4 nm. Thus, all six residue segments had some probability of being both collapsed and extended but the observed distances did not match the values expected for canonical secondary structures and different regions clearly had a higher probability of being either collapsed or extended. Figure 7(a) shows the extreme cases as well as the middle case, for all of the i:i + 5 amide nitrogen distance distributions. All other segments fit somewhere between these extremes. Figure 7(b) shows the results of the same analysis performed for these same residues in the random coil model of p53TAD. This comparison shows that the PRE restraints are responsible for inducing bimodality in these distance distributions because the random coil model yields nearly identical unimodal distributions.

Comparison of structures with previously identified nascent secondary structure

The propensity of different regions to be more collapsed or extended was examined by taking the ratio of the peak heights at 1.1 and 1.4 nm from the amide nitrogen distance distributions [like those shown in Fig. 7(a)] and plotting these ratios as a function of residue i. In

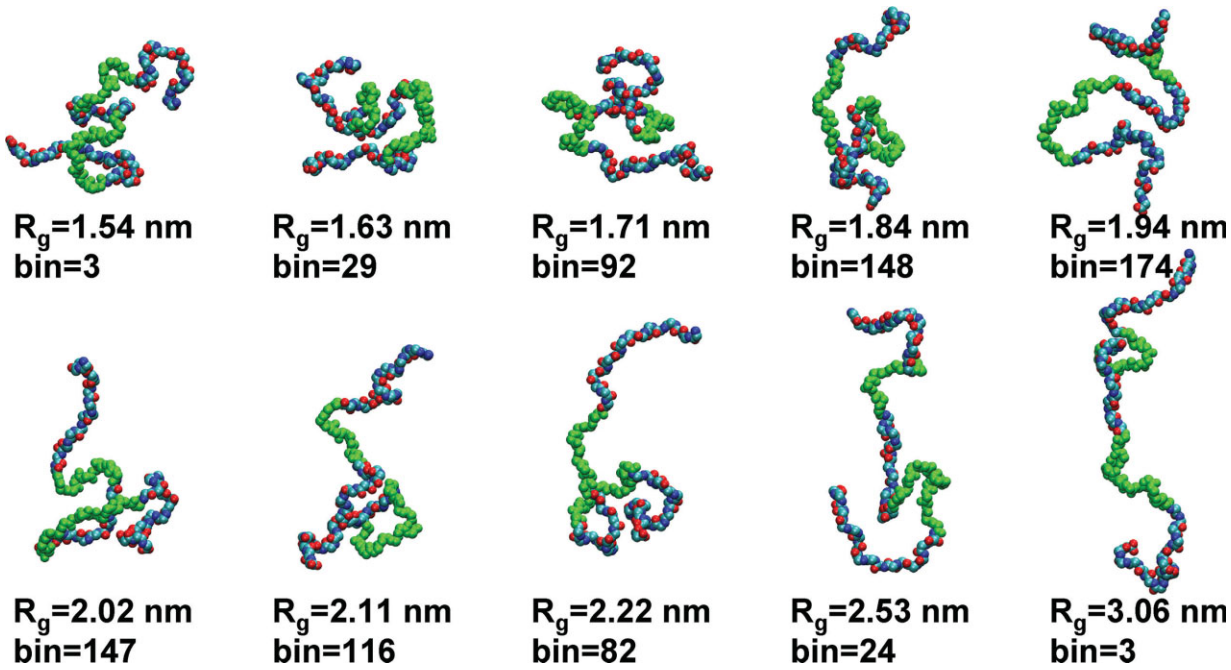

Figure 5

Space filling backbone models for structures in representative bins from the 1000 member ensemble shown in Figure 4. R_g and bin numbers are shown below each structure. The structures from these bins represent the R_g values for 85% of the structures in the ensemble. The MDM2 and RPA70 binding regions are colored green. These two binding regions maintain a close proximity in structures with R_g values $\leq$ 2.22 nm.

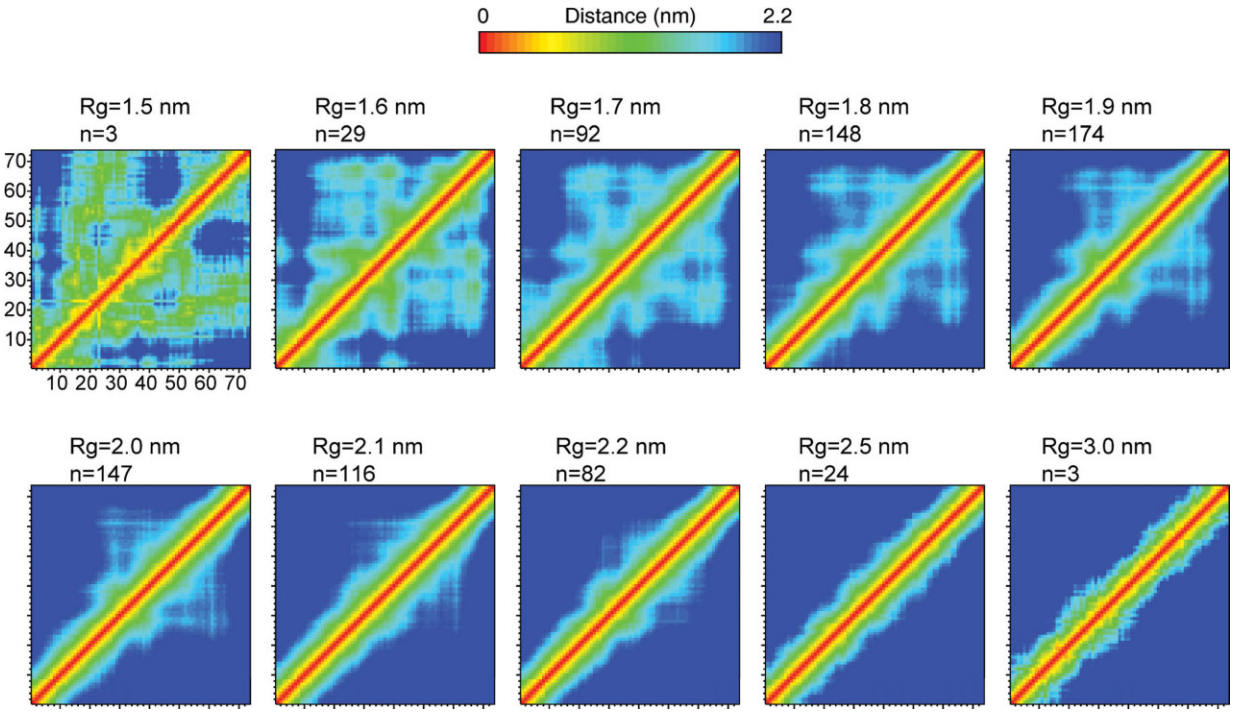

Figure 6

Contact maps showing the mean smallest distance between residues for structures from the bins in the R_g distribution shown in Figure 4. The distance scale is shown at the top of the figure and residue numbers are plotted for the first contact map. The R_g values shown represent the upper limit on the bin range.

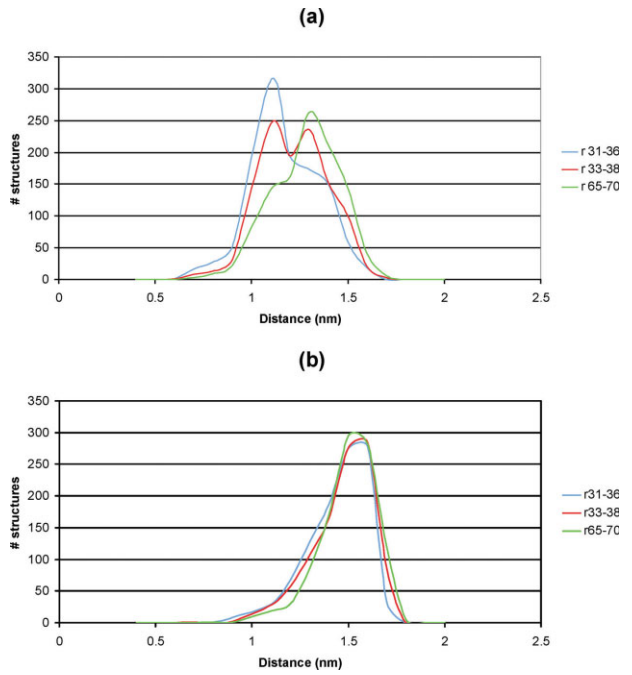

Figure 7

*Smoothed histograms showing the amide nitrogen distance distributions for pairs of residues that are four amino acids apart in the sequence. (**a**) The three histograms represent the two extremes and one middle cases of the bimodal distributions found for all residue pairs in the PRE-restrained p53TAD structures. (**b**) These three distributions were calculated for the same i:i + 5 pairs shown in (a) using the random-coil model of p53TAD.*

this analysis, a value of 1 means there is an equal probability of collapsed versus extended structures in the distribution. Values that are greater and less than 1 correspond to a probability of more collapsed and extended structures, respectively. The results of this analysis are shown in Figure 8(a) and allow a straightforward detection of regions that have a higher propensity for collapsed conformations. Using this approach, we are able to identify collapsed regions that overlap with the nascent helix that forms the primary MDM2 binding site and the two nascent turns that form the primary RPA70 binding site, which overlaps with two secondary MDM2 binding sites.[40,42,55–57] The region with the highest probability for collapse includes a PXP motif at residues 34–36. For comparison, Figure 8(b) shows the probabilities from two of the common disorder predictors, PONDR and IUPRED.[58,59] Both predictors show a weak tendency for order over the range of residues that have the greatest tendency for collapse (i.e., 16–55).

CONCLUSIONS

Identifying general relationships between the dynamic structure and molecular function of IUPs is hindered by a dearth of atomic models of their structural ensembles. In this report, we present the second ensemble of structures for an IUP based on PRE derived distance restraints. The first IUP structural ensemble was solved for a putative etiological agent of Parkinson's disease, alpha-synuclein.[13,17] Significant long-range interactions were observed for the structural ensembles of alpha-synuclein, and in one of the studies, the release of these interactions was shown to promote alpha-synuclein aggregation.[17] In the current study, a structural ensemble representing the native state of p53TAD was determined. Contrary to our expectation, biologically meaningful conclusions could be drawn from examining both individual structures from the ensemble as well as the entire ensemble.

Because the PRE based distances used to calculate the structures represent an average value from a dynamic and heterogeneous ensemble, there are a number of outstanding problems in their application to structure determination. For instance, the magnitude of the PRE effect varies as the inverse of the average sixth power of the distance between the unpaired electron and a nearby proton. This means that small populations in the ensemble

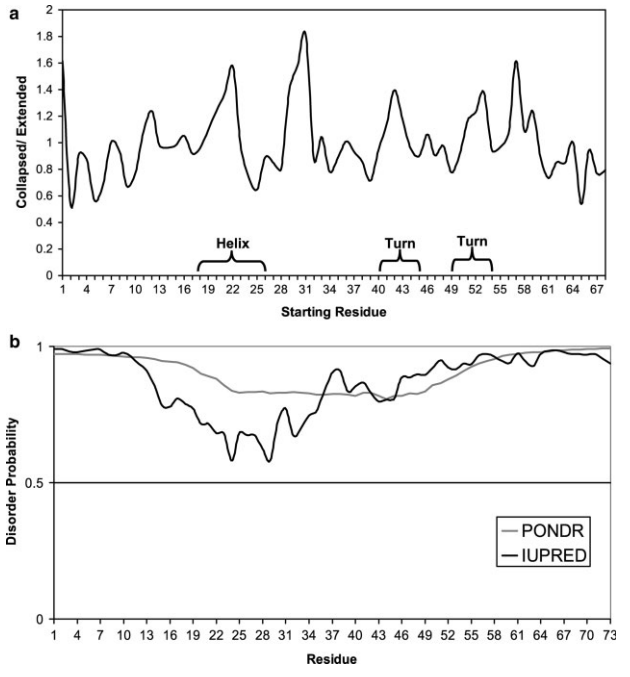

Figure 8

*(**a**) Ratio of collapsed versus extended structures for p53TAD based on distance distributions like those shown in Figure 6(a). The location of transient secondary structure elements that were previously identified using chemical shifts and J-coupling are shown with the parentheses. These regions overlap with the MDM2 (17–26) and RPA70 (40–54) binding sites. (**b**) Probabilities for two disorder predictors showing a weak tendency for order in the central region of p53TAD.*

with short distances will have a larger effect than large populations with long distances. We addressed this bias toward compact structures by using ensemble averaging. By varying the ensemble size between 1 and 20, we found that a modest ensemble size of five was sufficient to relax the R_g distribution to its equilibrium value. We also showed that when the average R_g value for the ensemble of structures was used to estimate the Stokes radius (R_s) it was in good agreement with R_s measurements (made using SEC) of wt p53TAD and E28C, with and without the MTSL.

Comparing calculated and measured R_s values does not generate a lot of confidence in the biological significance of individual structures in the ensemble. We examined individual structures from representative R_g bins in the ensemble and observed a striking correspondence in the spatial proximity of the MDM2 and RPA70 binding sites for structures with R_g values ranging from 1.54 to 2.22 nm. This result suggests that many of the p53TAD structures in the distribution would not be competent for binding to either MDM2 or RPA70 because of steric occlusion of the binding sites and that one of the more open conformations with R_g values in the range of 2.53–3.06 nm must be populated before binding.

An analysis of the entire ensemble of structures supports the contention that more extended structures must be populated for p53TAD to become competent for binding to MDM2 or RPA70. For instance, contact maps averaged over 709 structures in the ensemble that span R_g values of 1.5–2.1 nm showed predominant contacts occur between residues in the MDM2 and RPA70 binding sites. In addition, an analysis of the amide nitrogen distance distributions for the entire ensemble of structures revealed a preference for collapsed regions of p53TAD that overlap with the MDM2 and RPA70 binding sites. Although there is no direct evidence to support the contention that steric occlusion plays a roles in p53TAD binding to MDM2 and/or RPA70, it is not surprising that IUPs could regulate their function through the spatial organization of different binding sites.[11] In addition, there is some indirect evidence that steric occlusion plays a role in p53TAD binding to MDM2 and/or RPA70.[42,60–63]

Previous deletion and mutagenesis studies of binding between p53TAD and RPA70 identified p53TAD residues 40–60 as the most important for binding.[60–63] In these studies, a significant reduction in binding was also observed when p53TAD residues 20–40 were deleted and when p53TAD residues 22 and 23 were mutated. Consistent with these results, our recent study of the changes in the structure and dynamics of p53TAD when it binds to RPA70 showed significant chemical shift perturbations for p53TAD residues 20–40.[42] Our recent work also demonstrated that residues 40–60 of p53TAD are directly interacting with RPA70. Taken together, the deletion studies and our structural studies suggest that the binding of p53TAD to RPA70 is influenced by the presence of a compact dynamic structure.

The analysis of amide nitrogen distance distributions also provided evidence for a collapsed region between residues 28 and 38. In particular, there is a persistent loop centered on a PXP motif at residues 34–36. Interestingly, this PXP motif is flanked by two phosphorylation sites, S33 and S37. Both serines are phosphorylated in response to DNA damage, making p53 competent to activate apoptotic genes.[64] S33 is phosphorylated by the p38 kinase and S37 is phosphorylated by the ataxia telangiectasia mutated and homologous ATR proteins.[64] It is reasonable to suspect that phosphorylation at one or both of these serines will perturb the structure of the persistent loop we observe, which could increase the population of open conformations and promote the binding of MDM2 and RPA70. Whether this putative structural mechanism influences the apoptotic activity of p53 is unclear.

The structural biology of IUPs is at about the same point that the structural biology of compact globular proteins was 50 years ago. At that time there were a couple of low resolution protein structures available for identifying relationships between protein structure and function.[65,66] The highly dynamic nature of IUPs confounds the application of traditional methods for structure determination and new methods need to be developed. We hope this work will prompt other structural biologists to start solving the structural ensembles of their favorite IUP using internuclear distance constraints derived from PRE data. This will be necessary for a complete description of protein structure and function since IUPs are widespread in all organisms. More attempts at structure determination of IUPs will also prompt the development of new methods. Although the distance constraints provided by PRE data are not ideal, in the case of p53TAD, the ensemble of structures based on these restraints recovered a number of biologically important features as well as revealing new structural features that may impact function.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Brady Catherman and Rob Lyon for computer support at the University of Idaho Bioinformatics Core Facility and Dr. Charles Schwieters at NIH for assistance troubleshooting the parallel version of Xplor. The NMR data presented in this publication was collected at the University of Idaho Structural Biology Core Facility.

REFERENCES

1. 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.
2. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J,

Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model 2001;19:26–59.
3. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002;12:54–60.
4. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.
5. Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002;269:2–12.
6. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002;11:739–756.
7. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. J Mol Biol 1999; 293:321–331.
8. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41:6573–6582.
9. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 2005;579:3346–3354.
10. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208.
11. Vise P, Baral B, Stancik A, Lowry DF, Daughdrill GW. Identifying long-range structure in the intrinsically unstructured transactivation domain of p53. Proteins 2007;67:526–530.
12. Hou J, Sims GE, Zhang C, Kim SH. A global representation of the protein fold space. Proc Natl Acad Sci USA 2003;100:2386–2390.
13. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM. Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 2005;127:476–477.
14. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. J Mol Biol 1997;268:170–184.
15. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. J Mol Biol 1997; 268:158–169.
16. Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson CM, Poulsen FM, Vendruscolo M. Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. J Am Chem Soc 2004;126:3291–3299.
17. Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. Proc Natl Acad Sci USA 2005;102:1430–1435.
18. Dyson HJ, Wright PE. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Adv Protein Chem 2002;62:311–340.
19. 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.
20. Bargonetti J, Manfredi JJ. Multiple roles of the tumor suppressor p53. Curr Opin Oncol 2002;14:86–91.
21. Levine AJ. p53, the cellular gatekeeper for growth and division. Cell 1997;88:323–331.
22. Haupt Y, Robles AI, Prives C, Rotter V. Deconstruction of p53 functions and regulation. Oncogene 2002;21:8223–8231.
23. Zhou J, Ahn J, Wilson SH, Prives C. A role for p53 in base excision repair. EMBO J 2001;20:914–923.
24. Zhou J, Prives C. Replication of damaged DNA in vitro is blocked by p53. Nucleic Acids Res 2003;31:3881–3892.
25. Romanova LY, Willers H, Blagosklonny MV, Powell SN. The interaction of p53 with replication protein A mediates suppression of homologous recombination. Oncogene 2004;23:9025–9033.
26. Liu WL, Midgley C, Stephen C, Saville M, Lane DP. Biological significance of a small highly conserved region in the N terminus of the p53 tumour suppressor protein. J Mol Biol 2001;313:711–731.
27. Walker KK, Levine AJ. Identification of a novel p53 functional domain that is necessary for efficient growth suppression. Proc Natl Acad Sci USA 1996;93:15335–15340.
28. Hollstein M, Rice K, Greenblatt MS, Soussi T, Fuchs R, Sorlie T, Hovig E, Smith-Sorensen B, Montesano R, Harris CC. Database of p53 gene somatic mutations in human tumors and cell lines. Nucleic Acids Res 1994;22:3551–3555.
29. Hollstein M, Shomer B, Greenblatt M, Soussi T, Hovig E, Montesano R, Harris CC. Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation. Nucleic Acids Res 1996;24:141–146.
30. Shieh SY, Taya Y, Prives C. DNA damage-inducible phosphorylation of p53 at N-terminal sites including a novel site, Ser20, requires tetramerization. EMBO J 1999;18:1815–1823.
31. Kitayner M, Rozenberg H, Kessler N, Rabinovich D, Shaulov L, Haran TE, Shakked Z. Structural basis of DNA recognition by p53 tetramers. Mol Cell 2006;22:741–753.
32. Lee W, Harvey TS, Yin Y, Yau P, Litchfield D, Arrowsmith CH. Solution structure of the tetrameric minimum transforming domain of p53. Nat Struct Biol 1994;1:877–890.
33. Arrowsmith CH. Structure and function in the p53 family. Cell Death Differ 1999;6:1169–1173.
34. Arrowsmith CH, Morin P. New insights into p53 function from structural studies. Oncogene 1996;12:1379–1385.
35. Zhao K, Chai X, Johnston K, Clements A, Marmorstein R. Crystal structure of the mouse p53 core DNA-binding domain at 2.7 Å resolution. J Biol Chem 2001;276:12120–12127.
36. Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. Science 1994;265:346–355.
37. Ho WC, Fitzgerald MX, Marmorstein R. Structure of the p53 core domain dimer bound to DNA. J Biol Chem 2006;281:20494–20502.
38. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323:573–584.
39. Bell S, Klein C, Muller L, Hansen S, Buchner J. p53 contains large unstructured regions in its native state. J Mol Biol 2002;322: 917–927.
40. 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 p53. J Biol Chem 2000;275:29426–29432.
41. 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.
42. 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.
43. Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R. Extended disordered proteins: targeting function with less scaffold. Trends Biochem Sci 2003;28:81–85.
44. Vise PD, Baral B, Stancik AS, Lowry DF, Daughdrill GW. Identifying long-range structure in the intrinsically unstructured transactivation domain of p53. Proteins 2007;67:526–530.
45. Battiste JL, Wagner G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry 2000;39:5355–5365.
46. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem 2005;26:1701–1718.
47. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38, 27–38.
48. Uversky VN. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. Biochemistry 1993;32:13288–13298.

49. Clore GM, Schwieters CD. Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. Biochemistry 2004;43:10678–10691.
50. Clore GM, Schwieters CD. How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? J Am Chem Soc 2004;126:2923–2938.
51. Vendruscolo M. Determination of conformationally heterogeneous states of proteins. Curr Opin Struct Biol 2007;17:15–20.
52. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry 1999;38:16424–16431.
53. De Gennes PG. Scaling concepts in polymer physics. Ithaca, NY: Cornell University Press; 1979.
54. Flory PJ. Priciples of polymer chemistry. Ithaca, NY: Cornell University Press; 1953.
55. Schon O, Friedler A, Bycroft M, Freund SM, Fersht AR. Molecular mechanism of the interaction between MDM2 and p53. J Mol Biol 2002;323:491–501.
56. 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.
57. Chi SW, Lee SH, Kim DH, Ahn MJ, Kim JS, Woo JY, Torizawa T, Kainosho M, Han KH. Structural details on mdm2–p53 interaction. J Biol Chem 2005;280:38795–38802.
58. Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53 (Suppl 6):566–572.
59. Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol 2005;347:827–839.
60. Leiter LM, Chen J, Marathe T, Tanaka M, Dutta A. Loss of transactivation and transrepression function, and not RPA binding, alters growth suppression by p53. Oncogene 1996;12:2661–2668.
61. Dutta A, Ruppert JM, Aster JC, Winchester E. Inhibition of DNA replication factor RPA by p53. Nature 1993;365:79–82.
62. Li R, Botchan MR. The acidic transcriptional activation domains of VP16 and p53 bind the cellular replication protein A and stimulate in vitro BPV-1 DNA replication. Cell 1993;73:1207–1221.
63. Abramova NA, Russell J, Botchan M, Li R. Interaction between replication protein A and p53 is disrupted after UV damage in a DNA repair-dependent manner. Proc Natl Acad Sci USA 1997;94:7186–7191.
64. Roos WP, Kaina B. DNA damage-induced cell death by apoptosis. Trends Mol Med 2006;12:440–450.
65. Kendrew JC, Dickerson RE, Strandberg BE. Structure of myoglobin: a three-dimensional Fourier synthesis at 2 angstrom resolution. Nature 1960;206:757–763.
66. Perutz MF, Rossmann MP, Cullis AF, Muirhead H, Will G, North AC. Structure of haemoglobin: a three dimensional Fourier synthesis at 5.5 angstrom resolution, obtained by X-ray analysis. Nature 1960;185:416–422.