Published in final edited form as:
Proteins. 2014 December ; 82(12): 3373–3384. doi:10.1002/prot.24692.

Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins

Romel B. Perez[1], **Alexander Tischer**[2], **Matthew Auton**[2], and **Steven T. Whitten**[1,*]

[1]Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas

[2]Department of Internal Medicine, Division of Hematology, Mayo Clinic, Rochester, Minnesota

Abstract

Molecular transduction of biological signals is understood primarily in terms of the cooperative structural transitions of protein macromolecules, providing a mechanism through which discrete local structure perturbations affect global macromolecular properties. The recognition that proteins lacking tertiary stability, commonly referred to as intrinsically disordered proteins, mediate key signaling pathways suggests that protein structures without cooperative intramolecular interactions may also have the ability to couple local and global structure changes. Presented here are results from experiments that measured and tested the ability of disordered proteins to couple local changes in structure to global changes in structure. Using the intrinsically disordered N-terminal region of the p53 protein as an experimental model, a set of proline and alanine to glycine substitution variants were designed to modulate backbone conformational propensities without introducing non-native intramolecular interactions. The hydrodynamic radius (R_h) was used to monitor changes in global structure. Circular dichroism spectroscopy showed that the glycine substitutions decreased polyproline II (PP_{II}) propensities relative to the wild type, as expected, and fluorescence methods indicated that substitution-induced changes in R_h were not associated with folding. The experiments showed that changes in local PP_{II} structure cause changes in R_h that are variable and that depend on the intrinsic chain propensities of proline and alanine residues, demonstrating a mechanism for coupling local and global structure changes. Molecular simulations that model our results were used to extend the analysis to other proteins and illustrate the generality of the observed proline and alanine effects on the structures of intrinsically disordered proteins.

Keywords

intrinsically disordered protein; proline; alanine; polyproline II; hydrodynamic radius

Introduction

Intrinsic disorder is a common structural motif observed in many eukaryotic proteins most-widely found in transcription factors and other regulatory proteins that transduce molecular signals (1–4). Proteins, or protein domains, that have been classified as intrinsically

*To whom correspondence should be addressed: Department of Chemistry and Biochemistry, Texas State University, 601 University Drive, San Marcos, Texas 78666. Telephone: (512) 245-7893. Fax: (512) 245-2374. steve.whitten@txstate.edu.

disordered exhibit wide degrees of conformational freedom and flexibility (5), an absence of correlated dynamics between residue positions (6,7), and, overall, a persistent lack of tertiary stability (8,9). In addition to signal transduction, protein phosphorylation usually occurs within intrinsically disordered regions (10) and, accordingly, many critical biological processes are mediated by intrinsically disordered proteins (IDPs), such as transcriptional regulation, chromatin remodeling, and cell cycle control (11–13).

Induced folding studies (14–16) have led to quantitative descriptions of IDP signaling activity (17–19) in terms of allostery arising from the statistical thermodynamics of disorder-to-order transitions (i.e., cooperative folding). These allostery models have been useful for explaining distinctive signaling-related phenomena, such as conditional cooperativity (20) and agonism/antagonism switching (21). A key feature of these models is the induced folding of disordered structure from binding that results in the thermodynamic coupling of distant protein regions. Binding-induced folding has been observed for many IDPs (16,20–25) and characterized in detail, in terms of structures (23), binding kinetics (24), and interaction energies (25). Global organization of IDP structure in the absence of binding or folding has also been reported (26,27). The physicochemical mechanisms that organize disordered protein structures on a global scale without cooperative folding are not yet fully understood.

The study presented here quantitatively investigates coupled structural properties in an IDP in the absence of detectable folding. The intrinsically disordered N-terminal region of the p53 protein, p53(1–93), was used as the model IDP since this protein fragment is structurally dynamic (22,28,29), exhibits regulatory activity in the context of the full-length protein (30,31), and is proline-rich with high net charge; all common IDP properties (2–4). Motivation for the present work was the observation that substitution of all proline residues for glycine decreased the hydrodynamic radius (R_h) of p53(1–93) by 5.3 Å, though only a 2 Å reduction was expected from lower polyproline II (PP_{II}) propensities in the substitution variant (32). The experimental data showed that changes in intrinsic backbone conformational propensities could be coupled to changes in the global structure and, specifically, that changes in local PP_{II} propensities could cause changes in R_h greater than the predictions of structure-based models.

To assess the relationships between global structure, as reported by R_h, and position-specific conformational propensities, the structural characteristics of a set of proline (PRO) and alanine (ALA) to glycine (GLY) substitution variants of p53(1–93) were measured. The amino acid substitutions were designed to modulate intrinsic backbone conformational propensities without introducing non-native or long-range intramolecular interactions. Circular dichroism (CD) spectroscopy was used to estimate secondary structure content in each p53(1–93) variant while dynamic light scattering (DLS) and size exclusion chromatography (SEC) techniques were used to measure R_h. Analysis of the results estimated that per-residue PP_{II} propensities for ALA and PRO in the wild type were 48% and 76%, respectively. R_h dependence on intrinsic PP_{II} propensities followed an exponential trend, consistent with simulation data reported previously (32), showing an increased dependence on the amount of PP_{II} structure when R_h had expanded values and a reduced dependence with compacted R_h. To test if R_h compaction caused by the PRO and ALA to

GLY substitutions buried hydrophobic surface, fluorescence methods were used to measure solvent accessibility of tryptophan groups. Overall, the results showed that disordered structures can fine-tune R_h sensitivity to local structure perturbations through intrinsic backbone conformational propensities and thus regulate the linkage between local and global structure changes in the absence of cooperative folding. Extrapolation of the p53(1–93) results to other IDPs using R_h values from published reports suggested that most IDPs have similar structural characteristics in their disordered states.

Materials and Methods

Expression and purification of recombinant protein

Recombinant human p53(1–93) and *staphylococcal* nuclease were expressed in bacterial cells and isolated to >95% purity using the protocols described elsewhere (33,34). Purified bovine carbonic anhydrase, chicken albumin, and horse myoglobin were purchased from Sigma-Aldrich (St. Louis, MO) and further processed by ion exchange chromatography to remove residual contaminants.

Circular dicroism spectroscopy

CD spectra were recorded using a Jasco J-710 spectropolarimeter equipped with a PFD-425S peltier unit (Jasco, Easton, MD) and employed a 1-mm path-length quartz cuvette. Samples were equilibrated at each temperature for 10 min. Spectra were collected with a resolution of 0.5 nm, a scan rate of 20 nm/min, and were the average of eight scans. Reported spectra were baseline corrected for solvent and buffer contributions.

Dynamic light scattering

DLS readings used noninvasive backscatter optics and were measured using a Zetasizer Nano ZS with peltier temperature control from Malvern Instruments (Worcestershire, UK). Samples were filtered using 0.2 mm PVDF syringe-driven filters from EMD Millipore Corporation (Billerica, MA). All measurements used 1-cm path-length quartz cuvettes. Solvent viscosity was calculated using the solvent builder software program provided by Malvern, which calculates the viscosity of dilute solutions from the solution contents (e.g., percentage weight of solutes and protein) based on the program Sednterp (35).

Size exclusion chromatography

SEC experiments used Sephadex G-75 (GE Healthcare, Piscataway, NJ) equilibrated in 10 mM sodium phosphate, 100 mM sodium chloride, pH 7. Elution volumes were determined from chromatograms measured using a Bio-Rad BioLogic LP System equipped with a UV absorbance monitor (Hercules, CA). Each sample contained 0.2–0.5 mg/mL protein in 10 mM sodium phosphate, 100 mM sodium chloride, pH 7 with 0.3 mg/mL blue dextran and 0.03 mg/mL 2,4-dinitrophenyl-L-aspartate added as indicator dyes to determine the void and total column volumes, respectively.

Flourescence spectroscopy

Fluorescence measurements were performed using a Horiba Jobin-Yvon Fluorolog 3 spectrofluorometer equipped with a Wavelength Electronics Model LF1–3751 temperature controller. Spectra of 0.5 μM p53(1–93) and of the corresponding buffer were recorded at 20°C and averaged three times from 305–440 nm using excitation wavelengths of 280 and 295 nm. The step width was 1 nm, the integration time 1 s. Isothermal acrylamide quenching of the tryptophan fluorescence for 0.5 μM p53(1–93) used an excitation wavelength of 295 nm. Following addition of buffered acrylamide solution, the protein sample was equilibrated for 5 min under slight stirring and the fluorescence intensity was recorded for 20 s at 359 nm and averaged. Thermal scans with 0.5 μM p53(1–93) were performed using a scan rate of 2°C min^{-1} from 20–75°C under slight stirring.

Results and Discussion

Computer simulated structural relationship between PP_{II} content and R_h

R_h measured for p53(1–93) under normal solution conditions was 32.8 Å at 25°C, approximately 11 Å larger than statistical coil estimates for a 93-residue polypeptide (32). A statistical coil is defined as a polypeptide chain with no strongly preferred conformations (36). p53(1–93) shows a local maximum in its CD spectrum at 221 nm, indicating PP_{II} propensities that were estimated at 12–17% by spectral deconvolution (33). To determine if 12–17% PP_{II} propensities could explain the relatively large R_h that was measured for p53(1–93), simulations of randomly configured polypeptide chains using a hard sphere collision model (37) were performed to quantify the relationship between chain propensities for PP_{II} and R_h in disordered structures (32). The results of those computer simulations were condensed to a set of simple equations that are reproduced here. The mathematical relationships outlined by these equations frame the experimental strategy of the current study and accordingly were used for data analysis.

The first equation in this set relates R_h (in Å) to the number of residues, N, using the standard power-law scaling relationship,

$$R_h = R_0 \cdot N^v, \quad (1)$$

where R_o was found to be 2.16. The effects of N on R_h were insensitive to amino acid sequence in the simulations. The exponential term, v, was dependent on the applied sampling bias for PP_{II}, S_{PPII}, by,

$$v(S_{PPII}) = v_0 - 0.088 \cdot \ln(1 - S_{PPII}), \quad (2)$$

where v_o was the value of the exponential with no applied chain bias (i.e., for S_{PPII} = 0), found to be 0.509. Elimination of randomly generated structures containing van der Waals contact violations (i.e., removal of steric conflicts from the simulation) caused the fractional number of PP_{II} folded residues, f_{PPII}, to differ from S_{PPII} by a Gaussian function that can be represented as,

$$f_{PPII}=S_{PPII}-0.062\cdot e^{\left(-(S_{PPII}-0.63)^2/2\cdot 0.28^2\right)}. \quad (3)$$

Equations 1–3 provide an estimate of the dependence of R_h on PP_{II} content in disordered proteins of any size N.

From equations 1–3, PP_{II} propensities of 12–17% increase R_h for a 93-residue disordered protein to 23.0–23.6 Å from the statistical coil value of 21.7 Å (see Figure 1). Since R_h for p53(1–93) was measured to be 32.8 Å, these data indicated that either the simulation results were in gross error, the R_h measurements were wrong, PP_{II} propensities in p53(1–93) were much higher than 17%, or molecular properties other than PP_{II} propensities were dominant in determining R_h for this proline-rich protein (22 of 93 residues are proline in the wild type sequence). The CD-estimated PP_{II} propensities were supported by NMR analysis of the proline and pre-proline positions (38), whereas size exclusion chromatography and gel electrophoresis experiments indicated that the DLS-measured R_h values were reasonable (32). Surprisingly, the decrease in R_h, relative to the wild type, measured for a p53(1–93) variant that had all PRO residues substituted for GLY (referred to as PRO^-) supported the simulation data in terms of the predicted change in f_{PPII}. Figure 1 shows that an analysis of R_h measured for this PRO^- variant, found to be 27.5 Å at 25°C, was consistent with an 18% decrease in f_{PPII} relative to the wild type, in close agreement to the CD-estimated wild type PP_{II} propensities (32).

The simulation results thus seemed to capture some features of the relationship between intrinsic backbone conformational propensities and the global structural parameter R_h despite its simplicity. The simulation data also indicated that changes in R_h from incremental changes in f_{PPII} (i.e., $\partial R_h/\partial f_{PPII}$) were strongly dependent on the degree of R_h compaction. This can be seen in Figure 1 by noting that $\partial R_h/\partial f_{PPII}$ for wild type p53(1–93) was ~4X the statistical coil value. To test if R_h dependence on backbone conformational propensities could be affected by structural compaction in IDPs, two additional p53(1–93) variants were constructed. One had all 12 alanine residues in the wild type substituted for GLY (referred to as ALA^-). The other used the same ALA-to-GLY substitution strategy applied to the PRO^- variant (referred to as ALA^-PRO^-). ALA residues in disordered polypeptides show significant PP_{II} propensities, with per-residue estimates as high as 80% (39,40). Thus, ALA should contribute to R_h in a manner similar to PRO. These two additional substitution variants also tested additivity of the effects of PP_{II} structure on R_h in this system, by measuring the effects of ALA-to-GLY substitutions in the background of wild type and PRO^-, and similarly by measuring the effects of PRO-to-GLY substitutions in the background of wild type and ALA^-.

Proline and alanine contributions to PP_{II} propensities

CD spectroscopy was used to estimate relative PP_{II} propensities among the four p53(1–93) variants. The spectra measured under normal solution conditions for wild type, ALA^-, PRO^-, and ALA^-PRO^- from 5–85°C are provided in Figure 2. Spectra measured for ALA^-PRO^- were practically superimposable with the spectra measured for PRO^-. A local maximum centered at 221 nm was observed in the variants that contained proline residues

that was mostly absent in the variants that had no prolines (figure insets). A local CD maximum from 220–225 nm is often seen in proteins containing PP_{II} (41–43). For p53(1–93), this local maximum disappeared with increased temperatures, consistent with known heat effects on PP_{II} propensities (44). Overall, these data suggested that the proline-containing variants had significant PP_{II} content at normal temperatures, whereas the variants lacking proline had negligible PP_{II} content. Panel A of Figure 3 shows that the CD spectra at high temperatures differed noticeably between variants that contained proline residues (wild type and ALA^-) in comparison to variants that had no prolines (PRO^- and ALA^-PRO^-). Thus, at high temperatures where PP_{II} content should be minor, structural differences were noted owing to the presence or absence of proline residues in the sequence.

Spectral deconvolution using the CDPro software package (45) was performed to quantify PP_{II} content in the four p53(1–93) variants. Noted previously (32), CDPro predicted 12–17% PP_{II} propensities for the wild type spectrum, depending on the algorithm (e.g., CONTIN/LL, CDSSTR) and basis set used. The results of spectral analysis performed in this manner were surprisingly insensitive to differences in the p53(1–93) spectrum owing to heat and PRO-to-GLY and ALA-to-GLY substitutions. For example, the CONTIN/LL algorithm and SP22X basis set (46) showed PP_{II} content that ranged from 13.1–13.8% among the four variants at all temperatures (Figure S1). Since PRO-to-GLY substitutions (47) and increased temperatures (44) are known to decrease PP_{II} content, and this was not detected by spectral deconvolution, only qualitative differences in PP_{II} propensities among the p53(1–93) variants could be determined from CD readings. Using the local CD maximum at 221 nm as a metric for PP_{II} content, Figure 3 panel B shows that in terms of the height of this local maximum, wild type > ALA^- > PRO^- $\gtrsim$ ALA^-PRO^-. These CD data were thus interpreted qualitatively to indicate that at normal temperatures, wild type p53(1–93) had the highest levels of PP_{II} content, with ALA^- slightly less, and PRO^- and ALA^-PRO^- significantly less.

Proline and alanine contibutions to R_h

To quantify the structural size of p53(1–93), its diffusion coefficient, D, was measured by DLS and converted to an apparent hydrodynamic radius (R_h) using the Stokes-Einstein equation,

$$R_h = kT/(6\pi\eta D), \quad (4)$$

where k is the Boltzmann constant, T is absolute temperature, and η is solvent viscosity. The results from measuring R_h for each variant over a temperature range of 5–75°C using DLS are provided in Figure 4.

The three substitution variants each showed the same broad and gradual compaction in R_h caused by increased temperatures that was noted previously for the wild type (32), indicating heat-induced structural changes that were absent cooperative effects. At low temperatures, R_h for each variant was at least 7 Å larger than R_h calculated for a 93-residue statistical coil. At 75°C, R_h for PRO^- and ALA^-PRO^- approached statistical coil dimensions, whereas R_h for wild type and ALA^- remained significantly more expanded by ~5 Å. Similar to the spectroscopic observations at high temperatures using CD (Figure 3

panel A), these data indicated structural differences owing to proline residues even when proline-associated PP_{II} content should be small. Also, compaction in R_h from glycine substitution followed the trend that was observed in the CD-estimated changes in PP_{II} content. ALA$^-$, which was estimated to have slightly less PP_{II} content than the wild type, had a slightly smaller R_h at most temperatures. PRO$^-$ and ALA$^-$PRO$^-$, which were estimated to have significantly less PP_{II} content than the wild type, had significantly smaller R_h.

To test if the DLS methods measured R_h accurately, size exclusion chromatography (SEC) experiments were performed at room temperature using the p53(1–93) variants and a set of protein standards. Correlating elution volumes (V_e) measured for the protein standards to molecular weight using the distribution coefficient,

$$K_D=(V_e - V_0)/(V_t - V_0), \quad (5)$$

where V_t and V_o were the total and void volumes of the column, respectively, shows that the elution volume for each p53(1–93) variant was comparatively low, indicating R_h similar to folded proteins of much larger molecular weight, and R_h in a rank order of wild type > ALA$^-$ > PRO$^-$ $\gtrsim$ ALA$^-$PRO$^-$ (Figure 5 panel A). K_D that were measured for the p53(1–93) variants are provided in Table I. Though K_D for these variants did not trend with molecular weight when compared to folded proteins, K_D for all tested proteins trended with R_h (Figure 5 panel B). Here, R_h was estimated as one-half the maximum distance between any two C_α atoms in the reported crystallographic structure for each protein standard (48–51) and then used to correlate linearly with K_D. Plotting DLS-measured R_h against K_D followed the same linear trend for the p53(1–93) variants, showing that SEC and DLS methods yield consistent R_h values that likewise compare favorably to R_h estimated from high resolution structures. Because of this strong agreement, K_D values measured for the p53(1–93) variants were used to estimate R_h using the linear correlation that was observed with the protein standards. R_h determined in this manner yielded almost identical R_h values as the DLS measurements (Table I).

Forman-Kay and Marsh showed that R_h for IDPs scale with fractional proline content and net charge and can be predicted accurately from sequence (52). For p53(1–93), sequence predicted an R_h of 32.2 Å for the wild type and 24.3 Å for the variants lacking proline residues (Table I). For a 93-residue IDP with no proline residues and no net charge, the sequence-based equation predicted an R_h of 21.8 Å. These comparisons are important for a number of reasons. First, it shows that R_h measured for wild type p53(1–93) by DLS and SEC were consistent with general R_h trends observed in other IDPs – and thus the p53(1–93) values were not anomalous. Next, it demonstrates that estimates of statistical coil R_h from equations 1 and 2 (i.e., with S_{PPII} = 0) are reasonable since they agree with IDP R_h values absent proline-induced or net charge-induced structural expansion. For the specific case of p53(1–93), our data indicated that proline effects on R_h depend on the presence of alanine residues at the wild type ALA positions. This observation was not predicted by the Forman-Kay and Marsh equation that calculates an identical proline contribution to R_h (7.9 Å) for both the wild type and ALA$^-$.

R_h values measured for the p53(1–93) variants were also used to estimate per-residue PP_{II} propensities for ALA and PRO. Using averaged R_h for wild type and PRO^-, equations 1–3 predict that f_{PPII} changed from 0.572 to 0.393 with the PRO-to-GLY substitutions. This represents a change in fractional PP_{II} content of −0.179 and indicates a per-residue PP_{II} propensity of 76% for each PRO in wild type. An identical calculation using the averaged R_h for ALA^- gave a per-residue PP_{II} propensity of 48% for each wild type ALA. Other research groups have shown that alanine residues favor PP_{II} structure in disordered peptides, with propensities reported to be as low as 30% (53) and as high as 80% (39,40). Our measurements of ALA propensities for PP_{II} were within these (admittedly large) extremes.

Figure 1 predicted that R_h dependence on f_{PPII} should decrease with structural compaction. The experimental results clearly were in qualitative agreement with that prediction. The PRO-to-GLY substitutions compacted R_h by 5 Å when applied to the wild type, but caused only a 3 Å effect in ALA^-. Similarly, the ALA-to-GLY substitutions applied to the wild type compacted R_h by 2 Å, but had a negligible effect on R_h in PRO^-. Using the measured R_h values to compare changes in f_{PPII} among the p53(1–93) variants, however, demonstrated quantitative differences between the simulation model and the experimental results. The PRO-to-GLY substitutions decreased f_{PPII} by 0.179 when applied to the wild type and only by 0.117 when applied to ALA^-. If the simulation and experimental results were in full quantitative agreement, these values should have matched. Similarly, f_{PPII} decreased by 0.062 from wild type to ALA^-, but had no change from PRO^- to ALA^-PRO^-.

Contributions of hydrophobic burial to R_h compaction

The PRO-to-GLY and ALA-to-GLY substitutions compacted R_h for p53(1–93). Our analysis of the changes in R_h included only the effects of reduced PP_{II} content. Since unfolding almost universally reports increased R_h values for stable proteins (54), R_h compaction is typically associated with protein folding rather than lowered chain propensities for PP_{II}. To test for “folding” effects in R_h compaction of p53(1–93), experiments were performed to determine if hydrophobic surfaces were buried with the glycine substitutions. In this pursuit, intrinsic tryptophan fluorescence was measured in the presence of increasing acrylamide concentrations for wild type, ALA^-, PRO^-, and ALA^-PRO^-. Tryptophan residues are found at positions 23, 53, and 91 in p53(1–93). Acrylamide solutions quench tryptophan fluorescence by collisional contact and concomitant energy transfer, though this effect is more pronounced for exposed tryptophan groups relative to those that are structurally buried and protected from solvent (55–58).

The intrinsic fluorescence of each variant at 20°C is provided in Figure 6 panel A. The emission maximum for each was ~360 nm from excitation at 280 and 295 nm, indicating tryptophan residues in polar environments consistent with fully exposed groups (59,60). Adding acrylamide decreased p53(1–93) fluorescence proportional to the amount that was added. This is shown in panel B, where fluorescence is given in terms of F_o/F. F_o refers to fluorescence in the absence of acrylamide and thus the ratio F_o/F increases with increased quenching. For reference, acrylamide induced fluorescence quenching of equal molar amounts of N-acetyl tryptophanamide (NATA) was included. Quenching increased with higher temperatures (Figure S2), indicating a dynamic quenching mechanism (i.e.,

collisional frequency increases with added heat). If fluorescence quenching was caused by a static (i.e., binding) mechanism, then temperature increases would reduce the formation of bound complexes, thus reducing quenching, which was not observed for any of the p53(1–93) variants.

The extent of acrylamide induced fluorescence quenching, and thus tryptophan solvent accessibility, was assessed using the Stern-Volmer equation and the resulting Stern-Volmer constant, $K_{S\text{-}V}$ (61). Briefly, fluorescence intensity can be described as,

$$F_0/F = 1 + K_{S-V} \cdot [Q], \quad (6)$$

where [Q] is the quencher concentration. $K_{S\text{-}V}$ was 18.9 M^{-1} for wild type, 17.1 M^{-1} for ALA^{-}, 19.0 M^{-1} for PRO^{-}, 23.3 M^{-1} for $ALA^{-}PRO^{-}$, and 28.2 M^{-1} for the NATA reference. Disordered proteins (56) and proteins denatured by guanidine hydrochloride with reduced disulfides (62) characteristically have $K_{S\text{-}V}$ values ≥ 12 M^{-1}, while values of 1–2 M^{-1} are typical for folded proteins with tryptophan groups that are protected from solvent (55,56,62). $K_{S\text{-}V}$ measured for each p53(1–93) variant at 20°C thus was consistent with exposed tryptophan groups. Also, the substitution effects on $K_{S\text{-}V}$ values indicated mostly minor changes in tryptophan solvent accessibility. In general, the trends from PRO-to-GLY and ALA-to-GLY substitutions on $K_{S\text{-}V}$ were slight increases in tryptophan solvent accessibility, possibly owing to increased chain disorder from decreased PP_{II} content. The fluorescence quenching results did not seem to indicate significant burial of hydrophobic surface associated with R_h compaction owing to the PRO-to-GLY and ALA-to-GLY substitutions.

R_h trends among intrinsically disordered proteins

R_h that was measured for wild type p53(1–93) is similar in character to R_h measured for other IDPs, which is demonstrated in Figure 7 panel A. Here, R_h values for IDPs (26,63–80) and folded proteins (32,54,81) were plotted according to chain length. This dataset excluded IDPs with histidine tags, since histidine tags are known to compact R_h in IDPs (52). For reference, R_h for statistical coils were shown in this figure and provide a defining separation between folded and intrinsically disordered proteins. The folded proteins were all below the coil line and IDPs were all at or, more commonly, above this line.

To test if ALA content contributes to R_h for IDPs in general, as it does for p53(1–93), linear regression analysis was used to correlate fractional alanine content (f_{ALA}) with increases in R_h relative to the coil value ($R_{h,rel} = R_h/R_{h,coil}$). No correlation was observed using the set of IDPs shown in Figure 7 (coefficient of determination, $R^2 = 0.0001$, Table S1 in Supporting Information). Fractional proline content (f_{PRO}) and net charge (Q_{net}) each had positive correlations with $R_{h,rel}$ in this set ($R^2 = 0.242$ and $R^2 = 0.194$, respectively), in agreement with the report by Marsh and Forman-Kay (52). Since ALA effects on R_h for p53(1–93) were detected only in the variant that had the largest R_h value (i.e., wild type with high f_{PRO} and Q_{net}), multiple linear regression was used to determine if f_{ALA} in combination with f_{PRO} and Q_{net} could increase sequence-based correlations to $R_{h,rel}$. Regression analysis showed that $R_{h,rel}$ correlated with f_{PRO} and Q_{net} with $R^2 = 0.539$. Adding f_{ALA} as a fitting parameter increased R^2 to 0.542. The increase in R^2 owing to f_{ALA} was not meaningful, however, since

adding random numbers from 0 to 1 as the additional fitting parameter rather than f_{ALA} increased R^2 on average to 0.546 ± 0.008. Of note, if the set of IDPs used in regression analysis was limited to those with f_{PRO} above 10%, then f_{ALA} correlated strongly with $R_{h,rel}$ (R^2 = 0.472). Similarly, if regression analysis was limited to IDPs with Q_{net} > 10, f_{ALA} again correlated with $R_{h,rel}$ with high confidence (R^2 = 0.200). These data (tabulated in Table S1) suggest that ALA contributions to R_h may be minor for IDPs overall, but significant in combination with high f_{PRO} and/or high Q_{net}.

Figure 7A implies that, on average, IDPs are structurally expanded above coil estimates. Panel B in Figure 7 provides a possible selective advantage for disordered structures with expanded R_h – to regulate the coupling of local and global structure changes in IDPs. For example, the subtle effects of ALA-to-GLY substitutions on R_h were observed in wild type p53(1–93) and not in a structurally compacted variant. In Figure 7B, the reported R_h values were used to estimate $\partial R_h/\partial f_{PPII}$ (using equations 1–3) for each IDP and then divided by $\partial R_h/\partial f_{PPII}$ for a statistical coil to give a relative metric of how changes in f_{PPII} affect R_h (referred to as Relative Chain Sensitivity, RCS_{PPII}). An RCS_{PPII} of 1 represents the baseline effect of changes in f_{PPII} on R_h for a statistical coil. The IDP averaged RCS_{PPII} (dashed line in Figure 7B) was more than double the baseline coil value but the standard deviation was large (averaged RCS_{PPII} = 2.3 ± 0.8). Wide variations in RCS_{PPII} among IDPs could be related to protein-specific requirements for coordinating global structure changes to discrete local structure perturbations, such as phosphorylation, though experiments are needed to quantify the relationships between Q_{net}, f_{PRO}, f_{ALA}, and R_h. $R_{h,rel}$ correlated with Q_{net} (R^2 = 0.194), as noted above, suggesting that charge effects on IDP structure may be intertwined with f_{PRO} and f_{ALA}.

To test if charge effects on R_h depend on proline content, the set of IDPs in Figure 7 was split according to f_{PRO} and the relationships between $R_{h,rel}$ and Q_{net} were re-assessed. The data in Figure 8 show that the correlation between $R_{h,rel}$ and Q_{net} decreased to negligible levels (R^2 = 0.028 with $\partial R_{h,rel}/\partial Q_{net}$ = 0.002) in IDPs with f_{PRO} > 10%. In contrast, $R_{h,rel}$ correlated strongly with Q_{net} in IDPs with f_{PRO} < 10% (R^2 = 0.439 with $\partial R_{h,rel}/\partial Q_{net}$ = 0.008). These results predict variable charge effects, and presumably phosphorylation effects, in IDP structures dependent on proline content. A simple mechanism could account for this observation – proline promotes structural expansion in IDPs to weaken electrostatic-based interactions between charged groups. Accordingly, IDPs with high proline content, in general, should exhibit reduced coupling of changes in charge state with changes in R_h. IDPs with low proline content should demonstrate the opposite – increased coupling of changes in charge state with changes in R_h. Each IDP in Figure 7 was predicted to contain at least one phosphorylation site using the DISPHOS algorithm (Table S2 in Supporting Information), an algorithm that shows 80% accuracy in predicting protein phosphorylation sites from sequence (10). Owing to variable proline and alanine content among these IDPs, variable phosphorylation effects on R_h should be expected.

Conclusions

Proteins that are structurally dynamic often exhibit random coil characteristics (82–84), but IDPs under normal conditions seem to show non-random structural organization (23,26,28)

that is sensitive to sequence (52), charge (27,85,86), and temperature (32,87). To investigate the structural relationships between the global parameter R_h and intrinsic backbone conformational propensities in a disordered polypeptide chain, PRO-to-GLY and ALA-to-GLY substitutions were applied to the intrinsically disordered N-terminal region of the p53 protein, p53(1–93). The effects of GLY substitutions on R_h in p53(1–93) were compared to the results of molecular simulations of PP_{II} effects on disordered protein structures. The experimental results qualitatively matched the simulation predictions, but lacked quantitative agreement. Differences between computational and experimental data in our study may be from non-PP_{II} effects on disordered structure that were not included in the simulation. For example, charge-based interactions are known to affect the structural properties of IDPs (27,85,86). An analysis of IDP R_h values presented here indicated that proline content, and thus PP_{II} structure, and charge may act in concert to affect R_h. Specifically, R_h compaction (e.g., owing to glycine substitutions) should decrease the contributions of PP_{II} content to R_h and increase the effects of charge. ALA effects on R_h also seemed to be enhanced by high Q_{net} in IDPs.

The experimental results showed clearly that ALA content contributes to IDP R_h values, promoting structures larger than statistical coil estimates. Analysis of CD spectra indicated that ALA-to-GLY substitutions decreased PP_{II} content in p53(1–93), suggesting that ALA effects on R_h were primarily owing to changes in intrinsic PP_{II} propensities. The magnitude of R_h compaction in p53(1–93) from the ALA-to-GLY substitutions implied per-residue PP_{II} propensities of 48% for ALA in the wild type. Likewise, CD spectra indicated that PRO-to-GLY substitutions decreased PP_{II} content in p53(1–93), causing R_h compaction that was consistent with per-residue PP_{II} propensities of 76% for PRO. These results demonstrated that for IDPs, R_h compaction may trend with increased chain disorder, which, for p53(1–93), was caused by decreased PP_{II} content. This contrasts with R_h compaction associated with folded proteins (32,54), which typically follow disorder-to-order transitions (i.e., cooperative folding).

Overall, the results seemed to show that R_h sensitivity to discrete structural perturbations could vary significantly among IDPs. For example, the PRO-to-GLY substitutions caused a 5 Å compaction in R_h when applied to the wild type, but only a 3 Å compaction in R_h when applied to ALA^-. Also, the effects of ALA-to-GLY substitutions on R_h were context specific, causing a 2 Å compaction of the wild type, but no effect on R_h for PRO^-. In addition, correlations of $R_{h,rel}$ with Q_{net} in IDPs showed differences owing to proline content. These results predict that IDPs, though structurally dynamic and lacking in global conformational stability, can exhibit varied global structural responses to local structural perturbations. Accordingly, the effects of phosphorylation or cofactor interactions on disordered structure may differ substantially among IDPs, but likely in a predictable and sequence dependent manner.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants from the Research Corporation for Science Advancement 20039, the Texas Higher Education Coordinating Board 003615-0003-2011, and the NSF DMR-1205670 to STW and the National Heart Lung and Blood Institute of the National Institutes of Health HL109109 to MA. RBP was supported in part by NIGMS funds from the South Texas Doctoral Bridge program R25-GM-102783.

References

1. 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. [PubMed: 11381529]
2. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic disorder in transcription factors. Biochemistry. 2006; 45:6873–6888. [PubMed: 16734424]
3. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. Genome Inf Ser. 2000; 11:161–171.
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. [PubMed: 15019783]
5. Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol. 2007; 17:3–14. [PubMed: 17250999]
6. Kjaergaard M, Poulsen FM. Disordered proteins studied by chemical shifts. Prog Nucl Magn Reson Spectrosc. 2012; 60:42–51. [PubMed: 22293398]
7. Kosol S, Contreras-Martos S, Cedeño C, Tompa P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. Molecules. 2013; 18:10802–10828. [PubMed: 24008243]
8. Tompa P. Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci. 2012; 37:509–516. [PubMed: 22989858]
9. Eliezer D. Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol. 2009; 19:23–30. [PubMed: 19162471]
10. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 2004; 32:1037–1049. [PubMed: 14960716]
11. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005; 6:197–208. [PubMed: 15738986]
12. Cortese MS, Uversky VN, Dunker AK. Intrinsic disorder in scaffold proteins: getting more from less. Prog Biophys Mol Biol. 2008; 98:85–106. [PubMed: 18619997]
13. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophys J. 2007; 92:1439–1456. [PubMed: 17158572]
14. Kumar R, Lee JC, Bolen DW, Thompson EB. The conformation of the glucocorticoid receptor af1/tau1 domain induced by osmolyte binds co-regulatory proteins. J Biol Chem. 2001; 276:18146–18152. [PubMed: 11279138]
15. Wu P, Bolen DW. Osmolyte-induced protein folding free energy changes. Proteins. 2006; 63:290–296. [PubMed: 16453342]
16. Li J, Motlagh HN, Chakuroff C, Thompson EB, Hilser VJ. Thermodynamic dissection of the intrinsically disordered N-terminal domain of human glucocorticoid receptor. J Biol Chem. 2012; 287:26777–26787. [PubMed: 22669939]
17. Hilser VJ, Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proc Natl Acad Sci USA. 2007; 104:8311–8315. [PubMed: 17494761]
18. Motlagh HN, Hilser VJ. Agonism/antagonism switching in allosteric ensembles. Proc Natl Acad Sci USA. 2012; 109:4134–4139. [PubMed: 22388747]
19. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. Nature. 2014; 508:331–339. [PubMed: 24740064]

20. Garcia-Pino A, Balasubramanian S, Wyns L, Gazit E, De Greve H, Magnuson RD, Charlier D, van Nuland NA, Loris R. Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. Cell. 2010; 142:101–111. [PubMed: 20603017]
21. Ferreon AC, Ferreon JC, Wright PE, Deniz AA. Modulation of allostery by protein intrinsic disorder. Nature. 2013; 498:390–394. [PubMed: 23783631]
22. 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. [PubMed: 15824059]
23. Lawrence CW, Showalter SA. Carbon-Detected 15N NMR spin relaxation of an intrinsically disordered protein: FCP1 dynamics unbound and in complex with RAP74. J Phys Chem Lett. 2012; 3:1409–1413.
24. Rogers JM, Steward A, Clarke J. Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'. J Am Chem Soc. 2013; 135:1415–1422. [PubMed: 23301700]
25. Lawrence CW, Kumar S, Noid WG, Showalter SA. Role of ordered proteins in the folding-upon-binding of intrinsically disordered proteins. J Phys Chem Lett. 2014; 5:833–838.
26. Choi UB, McCann JJ, Weninger KR, Bowen ME. Beyond the random coil: stochastic conformational switching in intrinsically disordered proteins. Structure. 2011; 19:566–576. [PubMed: 21481779]
27. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci USA. 2013; 110:13392–13397. [PubMed: 23901099]
28. 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. [PubMed: 10884388]
29. 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. [PubMed: 17335006]
30. Levine AJ. p53, the cellular gatekeeper for growth and division. Cell. 1997; 88:323–331. [PubMed: 9039259]
31. Müller-Tiemann BF, Halazonetis TD, Elting JJ. Identification of an additional negative regulatory region for p53 sequence specific DNA binding. Proc Natl Acad Sci USA. 1998; 95:6079–6084. [PubMed: 9600920]
32. Langridge TD, Tarver MJ, Whitten ST. Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. Proteins. 2014; 82:668–678. [PubMed: 24150971]
33. Schaub LJ, Campbell JC, Whitten ST. Thermal unfolding of the N-terminal region of p53 monitored by circular dichroism spectroscopy. Protein Sci. 2012; 21:1682–1688. [PubMed: 22915551]
34. Whitten ST, García-Moreno EB. pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. Biochemistry. 2000; 39:14292–14302. [PubMed: 11087378]
35. Laue, TM.; Shah, BD.; Ridgeway, TM.; Pelletier, SL. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding, S.; Rowe, A.; Horton, JC., editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science. UK: Royal Society of Chemistry; 1992. p. 90-125.
36. Fitzkee NC, Rose GD. Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci USA. 2004; 101:12497–12502. [PubMed: 15314216]
37. Richards FM. Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng. 1977; 6:151–176. [PubMed: 326146]
38. 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. [PubMed: 18391200]
39. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci USA. 2002; 99:9190–9195. [PubMed: 12091708]

40. Shi Z, Chen K, Liu Z, Ng A, Bracken WC, Kallenbach NR. Polyproline II propensities from GGXGG peptides reveal an anticorrelation with beta-sheet scales. Proc Natl Acad Sci USA. 2005; 102:17964–17968. [PubMed: 16330763]
41. Woody RW. Circular dichroism and conformation of unordered polypeptides. Adv Biophys Chem. 1992; 2:37–79.
42. Rucker AL, Creamer TP. Polyproline II helical structure in protein unfolded states: lysine peptides revisited. Protein Sci. 2002; 11:980–985. [PubMed: 11910041]
43. Whittington SJ, Chellgren BW, Hermann VM, Creamer TP. Urea promotes polyproline II helix formation: implications for protein denatured states. Biochemistry. 2005; 44:6269–6275. [PubMed: 15835915]
44. Chen K, Liu Z, Kallenbach NR. The polyproline II conformation in short alanine peptides is noncooperative. Proc Natl Acad Sci USA. 2004; 101:15352–15357. [PubMed: 15489268]
45. Sreerama N, Woody RW. Estimation of protein secondary structure from CD spectra: comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. Anal Biochem. 2000; 287:252–260. [PubMed: 11112271]
46. Johnson WC. Analyzing protein circular dichroism spectra for accurate secondary structures. Proteins. 1999; 35:307–312. [PubMed: 10328265]
47. Rucker AL, Pager CT, Campbell MN, Qualls JE, Creamer TP. Host-guest scale of left-handed polyproline II helix formation. Proteins. 2003; 53:68–75. [PubMed: 12945050]
48. Stein PE, Leslie AG, Finch JT, Turnell WG, McLaughlin PJ, Carrell RW. Crystal structure of ovalbumin as a model for the reactive centre of serpins. Nature. 1990; 347:99–102. [PubMed: 2395463]
49. Saito R, Sato T, Ikai A, Tanaka N. Structure of bovine carbonic anhydrase II at 1.95 Å resolution. Acta Crystallogr D Biol Crystallogr. 2004; 60:792–795. [PubMed: 15039588]
50. Hynes TR, Fox RO. The crystal structure of Staphylococcal nuclease refined at 1.7 Å resolution. Proteins. 1991; 10:92–105. [PubMed: 1896431]
51. Zahran ZN, Chooback L, Copeland DM, West AH, Richter-Addo GB. Crystal structures of manganese- and cobalt-substituted myoglobin in complex with NO and nitrite reveal unusual ligand conformations. J Inorg Biochem. 2008; 102:216–233. [PubMed: 17905436]
52. Marsh JA, Forman-Kay JD. Sequence determinants of compaction in intrinsically disordered proteins. Biophys J. 2010; 98:2383–2390. [PubMed: 20483348]
53. Ferreon JC, Hilser VJ. The effect of the polyproline II (PPII) conformation on the denatured state entropy. Protein Sci. 2003; 12:447–457. [PubMed: 12592015]
54. 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. [PubMed: 10600103]
55. Eftink MR, Ghiron CA. Dynamics of a protein matrix revealed by fluorescence quenching. Proc Natl Acad Sci USA. 1975; 72:3290–3294. [PubMed: 810800]
56. Eftink MR, Ghiron CA. Exposure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching studies. Biochemistry. 1976; 15:672–680. [PubMed: 1252418]
57. Calhoun DB, Vanderkooi JM, Holtom GR, Englander SW. Protein fluorescence quenching by small molecules: protein penetration versus solvent exposure. Proteins. 1986; 1:109–115. [PubMed: 3130621]
58. Eftink MR, Hagaman KA. Fluorescence quenching of the buried tryptophan residue of cod parvalbumin. Biophys Chem. 1985; 22:173–180. [PubMed: 4052574]
59. Beechem JM, Brand L. Time-resolved fluorescence of proteins. Annu Rev Biochem. 1985; 54:43–71. [PubMed: 3896124]
60. Eftink MR. Fluorescence techniques for studying protein structure. Methods Biochem Anal. 1991; 35:127–205. [PubMed: 2002770]
61. Lakowicz, JR. Principles of Fluorescence Spectroscopy Third Edition. New York: Springer; 2006. p. 954

62. Tischer A, Madde P, Blancas-Mejia LM, Auton M. A molten globule intermediate of the von Willebrand factor A1 domain firmly tethers platelets under shear flow. Proteins. 2014; 82:867–878. [PubMed: 24265179]

63. Chong PA, Ozdamar B, Wrana JL, Forman-Kay JD. Disorder in a target for the smad2 mad homology 2 domain and its implications for binding and specificity. J Biol Chem. 2004; 279:40707–40714. [PubMed: 15231848]

64. Sivakolundu SG, Nourse A, Moshiach S, Bothner B, Ashley C, Satumba J, Lahti J, Kriwacki RW. Intrinsically unstructured domains of Arf and Hdm2 form bimolecular oligomeric structures in vitro and in vivo. J Mol Biol. 2008; 384:240–254. [PubMed: 18809412]

65. Danielsson J, Liljedahl L, Bárány-Wallje E, Sønderby P, Kristensen LH, Martinez-Yamout MA, Dyson HJ, Wright PE, Poulsen FM, Mäler L, Gräslund A, Kragelund BB. The intrinsically disordered RNR inhibitor Sml1 is a dynamic dimer. Biochemistry. 2008; 47:13428–13437. [PubMed: 19086274]

66. Yi S, Boys BL, Brickenden A, Konermann L, Choy WY. Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch. Biochemistry. 2007; 46:13120–13130. [PubMed: 17929838]

67. Paleologou KE, Schmid AW, Rospigliosi CC, Kim HY, Lamberto GR, Fredenburg RA, Lansbury PT Jr. Fernandez CO, Eliezer D, Zweckstetter M, Lashuel HA. J Biol Chem. 2008; 283:16895–16905. [PubMed: 18343814]

68. Baker, JMR. Structural characterization and interactions of the CFTR regulatory region (PhD Thesis). Toronto: Department of Biochemistry, University of Toronto; 2009.

69. Soragni A, Zambelli B, Mukrasch MD, Biernat J, Jeganathan S, Griesinger C, Ciurli S, Mandelkow E, Zweckstetter M. Structural characterization of binding of Cu(II) to tau protein. Biochemistry. 2008; 47:10841–10851. [PubMed: 18803399]

70. Lowry DF, Stancik A, Shrestha RM, Daughdrill GW. Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. Proteins. 2008; 71:587–598. [PubMed: 17972286]

71. Adkins JN, Lumb KJ. Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2. Proteins. 2002; 46:1–7. [PubMed: 11746698]

72. Uversky VN, Permyakov SE, Zagranichny VE, Rodionov IL, Fink AL, Cherskaya AM, Wasserman LA, Permyakov EA. Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: a natively unfolded protein. J Proteome Res. 2002; 1:149–159. [PubMed: 12643535]

73. Donaldson L, Capone JP. Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. J Biol Chem. 1992; 267:1411–1414. [PubMed: 1309782]

74. Haaning S, Radutoiu S, Hoffmann SV, Dittmer J, Giehm L, Otzen DE, Stougaard J. An unusual intrinsically disordered protein from the model legume Lotus japonicus stabilizes proteins in vitro. J Biol Chem. 2008; 283:31142–31152. [PubMed: 18779323]

75. Geething NC1, Spudich JA. Identification of a minimal myosin Va binding site within an intrinsically unstructured domain of melanophilin. J Biol Chem. 2007; 282:21518–21528. [PubMed: 17513864]

76. Permyakov SE, Millett IS, Doniach S, Permyakov EA, Uversky VN. Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. Proteins. 2003; 53:855–862. [PubMed: 14635127]

77. Magidovich E, Orr I, Fass D, Abdu U, Yifrach O. Intrinsic disorder in the C-terminal domain of the Shaker voltage-activated K+ channel modulates its interaction with scaffold proteins. Proc Natl Acad Sci USA. 2007; 104:13022–13027. [PubMed: 17666528]

78. Campbell KM1, Terrell AR, Laybourn PJ, Lumb KJ. Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. Biochemistry. 2000; 39:2708–2713. [PubMed: 10704222]

79. Sánchez-Puig N, Veprintsev DB, Fersht AR. Binding of natively unfolded HIF-1alpha ODD domain to p53. Mol Cell. 2005; 17:11–21. [PubMed: 15629713]

80. Sánchez-Puig N, Veprintsev DB, Fersht AR. Human full-length securin is a natively unfolded protein. Protein Sci. 2005; 14:1410–1418. [PubMed: 15929994]

81. Tcherkasskaya O, Davidson EA, Uversky VN. Biophysical constraints for protein structure prediction. J Proteome Res. 2003; 2:37–42. [PubMed: 12643541]

82. Chan HS, Dill KA. Polymer principles in protein structure and stability. Annu Rev Biophys Biophys Chem. 1991; 20:447–490. [PubMed: 1867723]

83. Tanford C, Kawahara K, Lapanje S. Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behavior. J Biol Chem. 1996; 241:1921–1923. [PubMed: 5947952]

84. Flory, PJ. Statistical Mechanics of Chain Molecules. New York: Wiley; 1969. p. 432

85. Müller-Späth S, Soranno A, Hirschfeld V, Hofmann H, Rüegger S, Reymond L, Nettels D, Schuler B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc Natl Acad Sci USA. 2010; 107:14609–14614. [PubMed: 20639465]

86. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci USA. 2010; 107:8183–8188. [PubMed: 20404210]

87. Wuttke R1, Hofmann H, Nettels D, Borgia MB, Mittal J, Best RB, Schuler B. Temperature-dependent solvation modulates the dimensions of disordered proteins. Proc Natl Acad Sci USA. 2014; 111:5213–5218. [PubMed: 24706910]

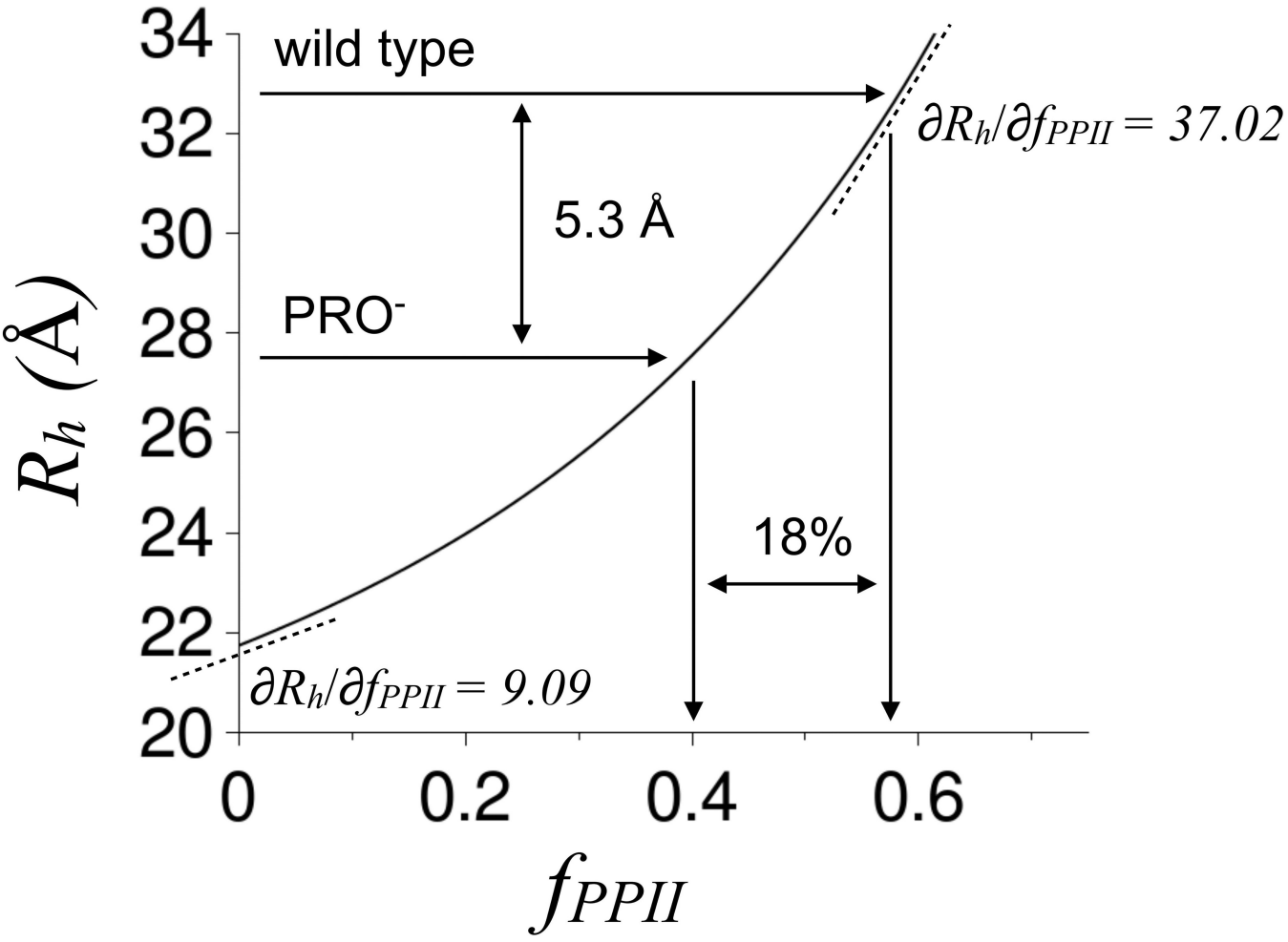

Figure 1. R_h dependence on f_{PPII} for a 93-residue disordered protein

R_h measured for wild type and PRO^- p53(1–93) are indicated in the figure. The curved line was calculated using equations 1–3 with $N = 93$. $\partial R_h/\partial f_{PPII}$ for a statistical coil was estimated empirically by calculating the change in R_h and f_{PPII} for the incremental change in S_{PPII} of 0 to 0.01. $\partial R_h/\partial f_{PPII}$ for wild type was estimated by a similar method. First, the wild type R_h (32.8 Å) was used to calculate a corresponding S_{PPII} ($S_{PPII,wt}$) using equations 1 and 2. Then $\partial R_h/\partial f_{PPII}$ for the wild type was estimated by calculating the change in R_h and f_{PPII} for the incremental change in S_{PPII} from $S_{PPII,wt}$ to $S_{PPII,wt} + 0.01$.

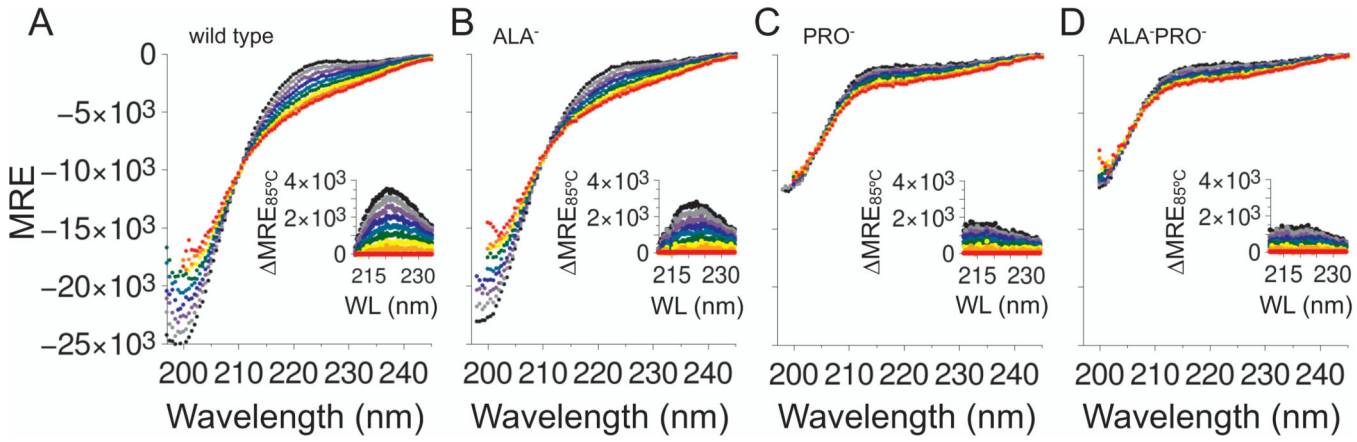

Figure 2. CD spectrum of p53(1–93) from 5–85°C

Panels show **A**) wild type, **B**) ALA^- **C**) PRO^-, and **D**) ALA^-PRO^-. All spectra were measured in 10 mM sodium phosphate, 100 mM sodium chloride, pH 7 using 0.1–0.2 mg/mL of protein and reported in molar residue ellipticity (MRE) units of deg cm^2 $dmol^{-1}$ res^{-1}. The data symbols (filled circles) were colored according to the temperature the spectrum was measured using: 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (blue–green), 55°C (green), 65°C (yellow), 75°C (orange), and 85°C (red). The insets show the difference in MRE between the spectra at each temperature relative to the 85°C spectrum.

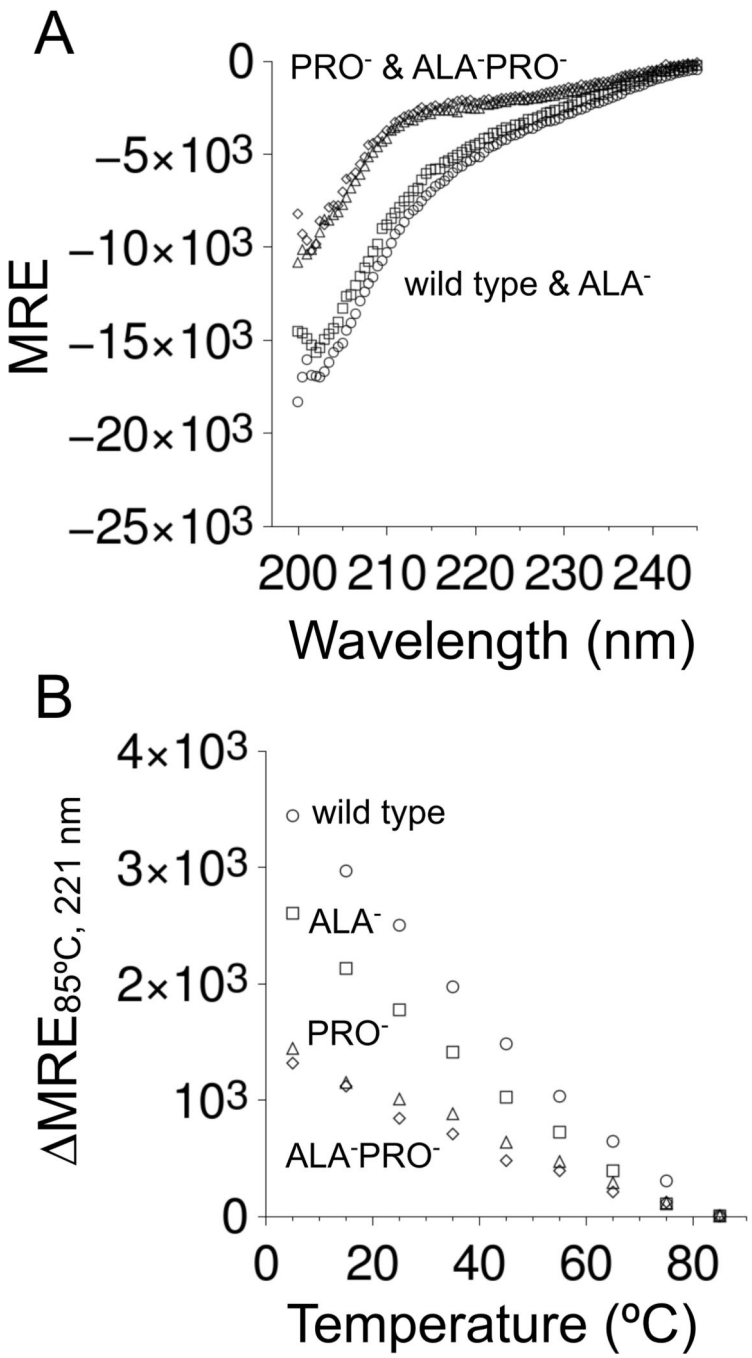

Figure 3. Effects of PRO-to-GLY and ALA-to-GLY substitutions on the CD spectrum of p53(1–93)

In both panels, wild type measurements were given with circles, ALA^- with squares, PRO^- with triangles, and ALA^-PRO^- with diamonds. Panel **A** shows the spectrum of each variant at 85°C. Panel **B** shows the temperature-dependence in the height of the local CD maximum that was observed at 221 nm (from insets of Figure 2).

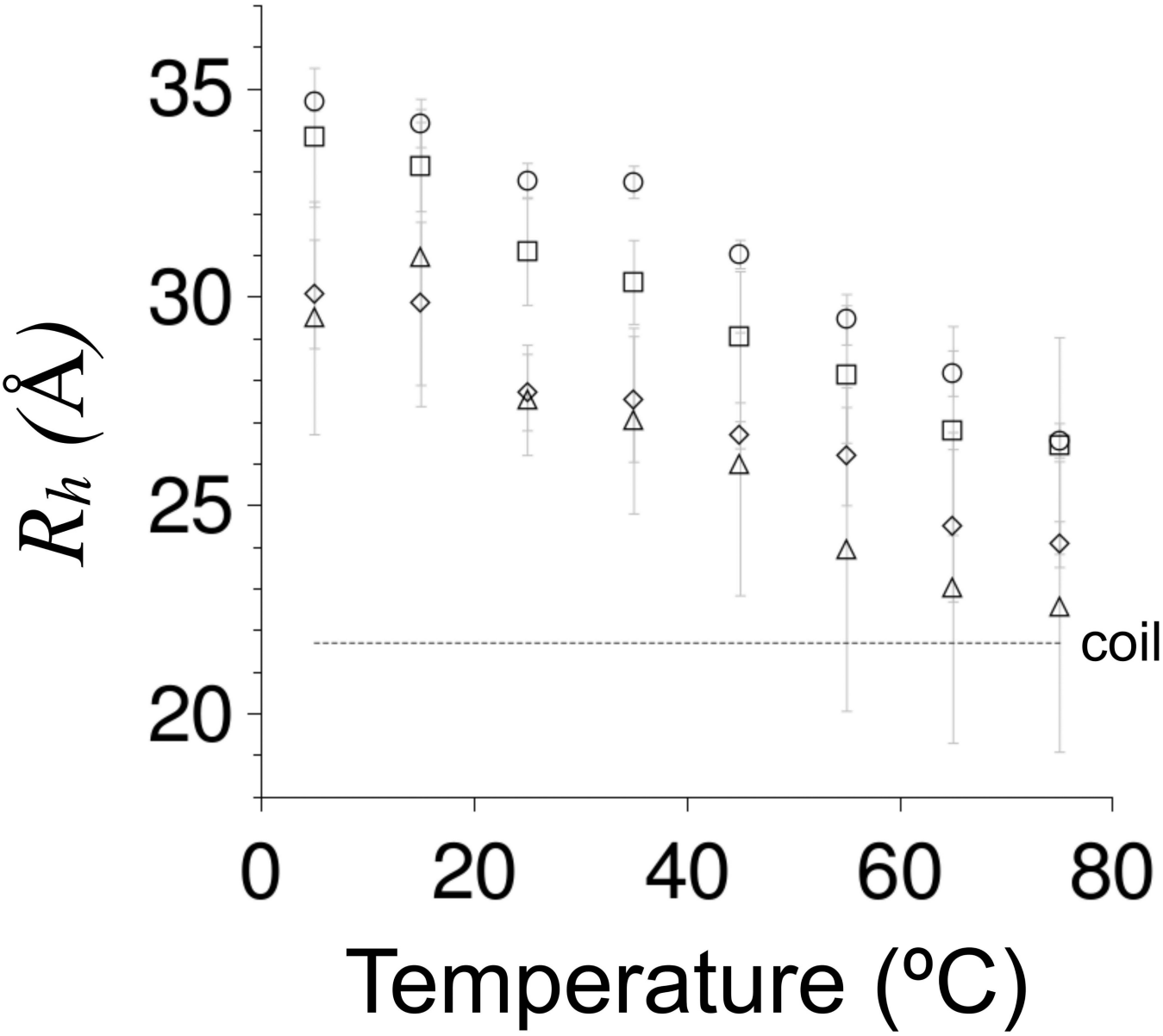

Figure 4. DLS-measured R_h for p53(1–93) from 5–75°C

R_h values were shown using circles for wild type, squares for ALA^-, triangles for PRO^-, and diamonds for ALA^-PRO^-. R_h were measured using 0.25–0.75 mg/mL samples of protein in 10 mM sodium phosphate, 100 mM sodium chloride, pH 7. The reported values are the average of at least 5 measurements performed at each temperature and error bars represent standard deviations. The stippled line labeled "coil" shows R_h calculated for a 93-residue disordered protein using equations 1 and 2 with $S_{PPII} = 0$.

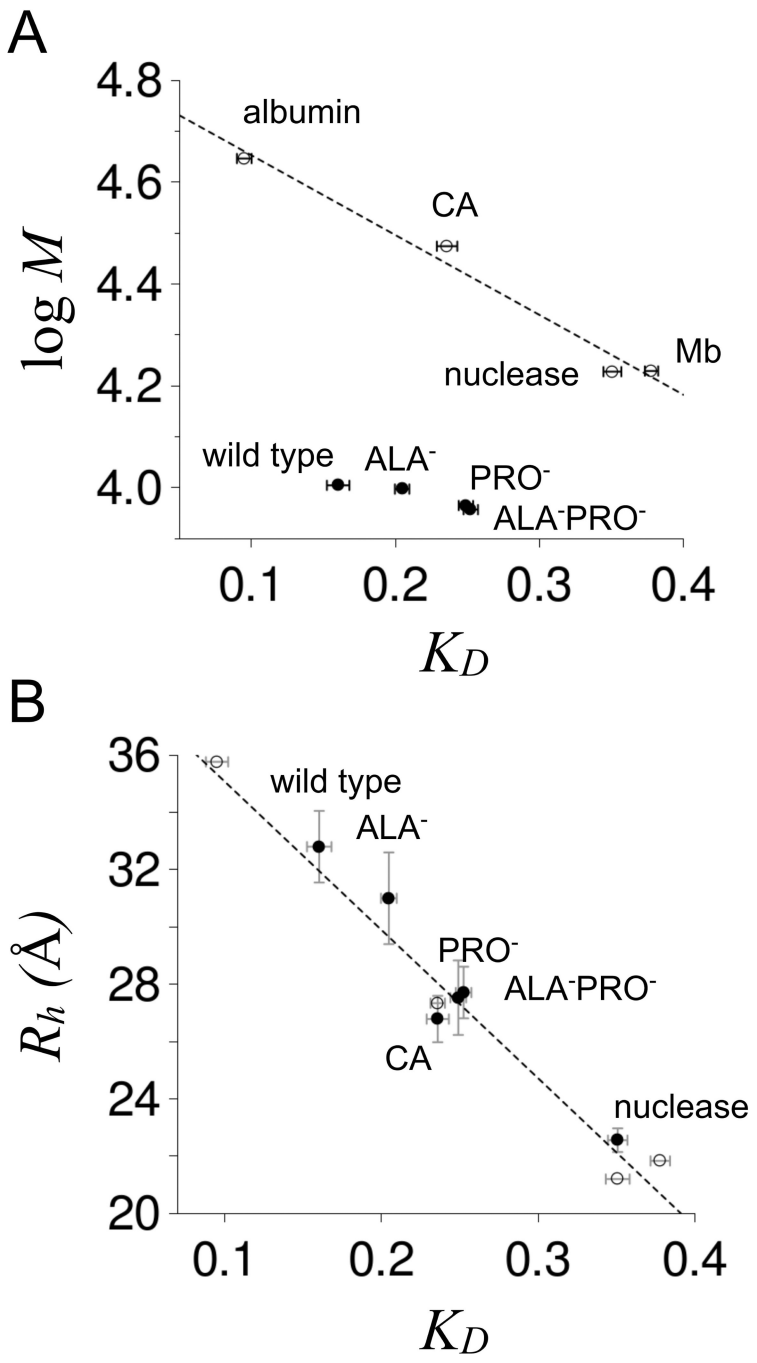

Figure 5. SEC-measured$K_D$for p53(1–93) at room temperature

Panel **A** shows a comparison of K_D to molecular weight for wild type, ALA^-, PRO^-, and ALA^-PRO^- (filled circles). Open circles are K_D measured for chicken albumin (44.3 kDa), bovine erythrocyte carbonic anhydrase (29.8 kDa, indicated as CA in the figure), *staphylococcal* nuclease (16.9 kDa), and horse myoglobin (16.95 kDa). Reported K_D were the average of at least 3 measurements for each protein with error bars representing the standard deviations. Panel **B** shows a comparison of K_D to R_h. Open circles are R_h estimated as one-half the maximum C_α-C_α distance in the crystallographic structures of albumin (48),

carbonic anhydrase (49), nuclease (50), and myoglobin (51). The dashed line is a linear fit of R_h to K_D applied to the open circles. Filled circles show R_h measured by DLS at 25°C for wild type, ALA^-, PRO^-, ALA^-PRO^-, carbonic anhydrase, and nuclease. K_D were measured using 0.2 mg/mL samples of protein in 10 mM sodium phosphate, 100 mM sodium chloride, pH 7. DLS measurements used identical solution conditions except for 0.25–0.75 mg/mL protein concentrations.

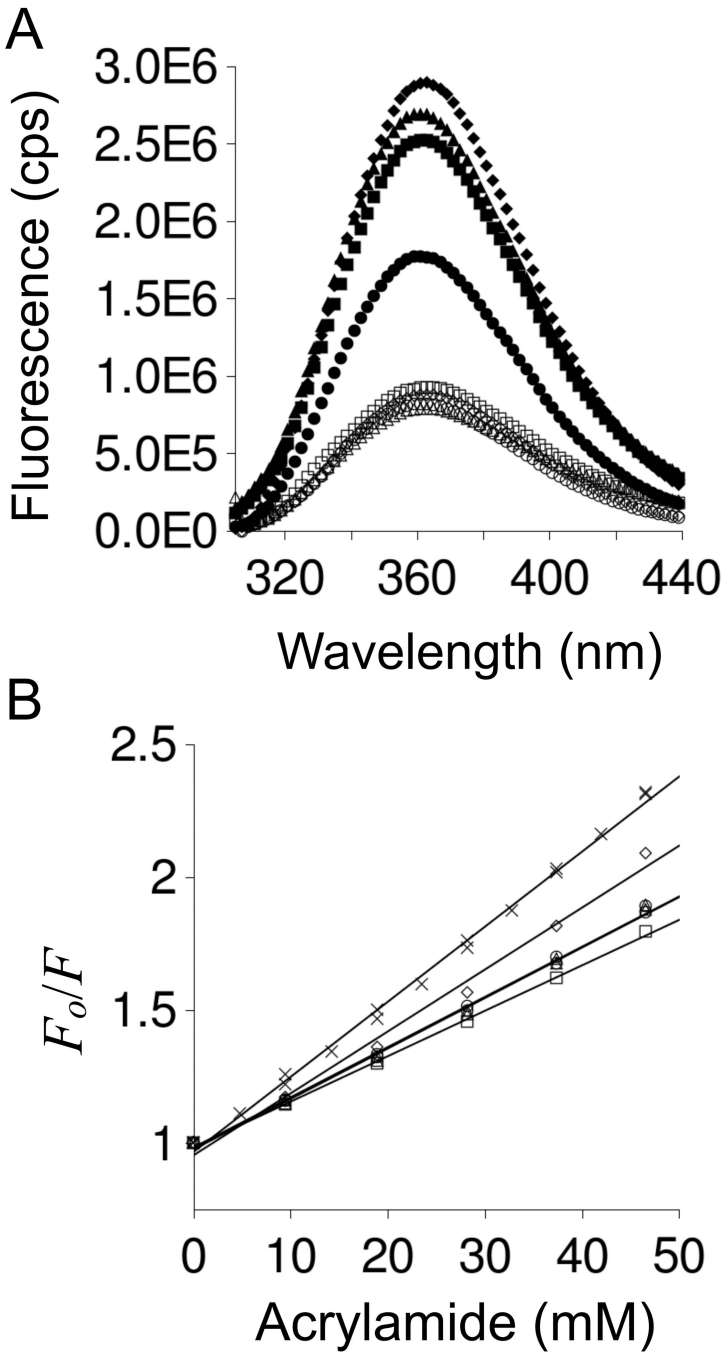

Figure 6. p53(1–93) fluorescence spectrum and acrylamide induced quenching of tryptophan fluorescence

Panel **A** shows the fluorescence spectrum of p53(1–93) measured at 20°C using excitation wavelengths of 280 nm (filled markers) and 295 nm (open markers). Panel **B** provides Stern-Volmer plots of acrylamide induced quenching of tryptophan at 20°C. Both panels used circles for wild type, squares for ALA^-, triangles for PRO^-, and diamonds for ALA^-PRO^-. NATA is shown in panel B using (X). In both panels, fluorescence was measured using 0.5

μM p53(1–93) or 1.5 μM NATA buffered at pH 7 with 10 mM sodium phosphate, 100 mM sodium chloride.

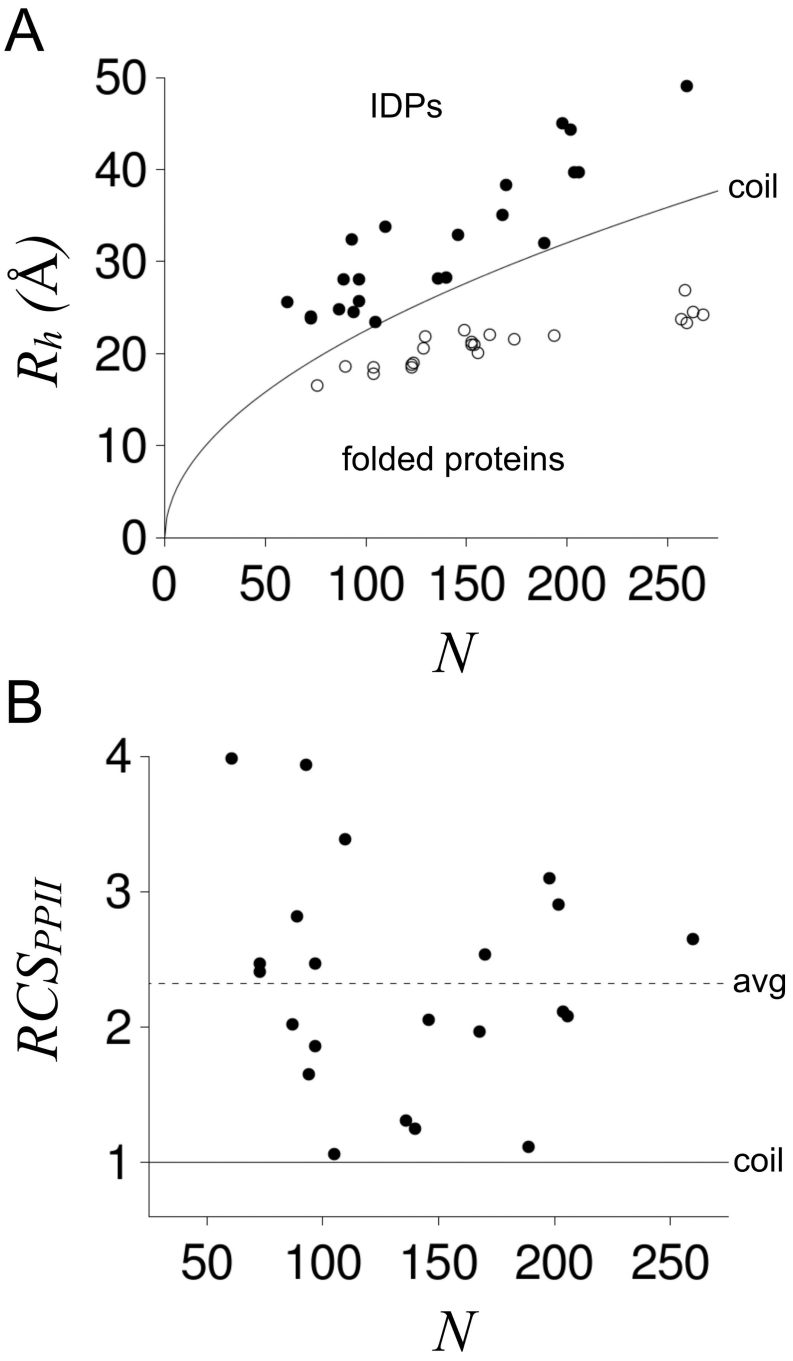

Figure 7. R_h comparisons to number of residues N for folded and intrinsically disordered proteins

This set was limited to $N > 50$ and used R_h from published reports (26,32,54,63–81). In panel **A**, the line labeled "coil" shows R_h calculated using equations 1 and 2 with $S_{PPII} = 0$. Filled circles are IDPs and open circles are folded proteins. In panel **B** R_h values for IDPs were converted to RSC_{PPII} RSC_{PPII} was determined for each IDP by first calculating the statistical coil $\partial R_h/\partial f_{PPII}$ (which varies with N) and IDP $\partial R_h/\partial f_{PPII}$ using the method given in Figure 1 for wild type p53(1–93). These two $\partial R_h/\partial f_{PPII}$ values were then divided (IDP/coil) to give RSC_{PPII}. For p53(1–93), the averaged R_h (32.4 Å) was used to determine

RSC_{PPII} rather than the DLS-measured R_h (32.8 Å). The dashed line shows the averaged RSC_{PPII} (2.3 ± 0.8) from the set of IDPs in panel A.

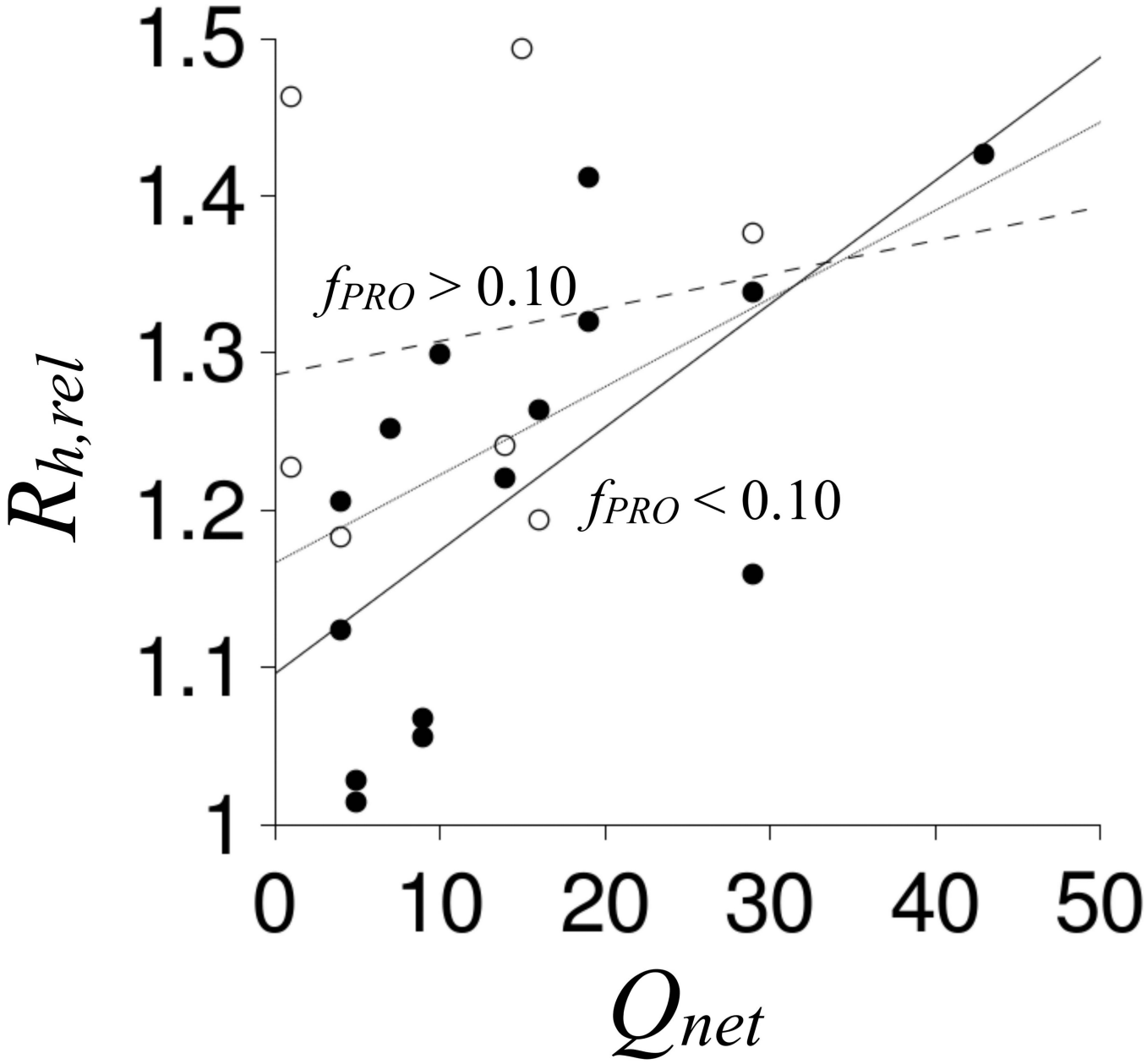

Figure 8. Comparison of $R_{h,rel}$ and Q_{net} for intrinsically disordered proteins

IDPs with $f_{PRO} > 0.10$ are shown using open circles. IDPs with $f_{PRO} < 0.10$ are shown with filled circles. Lines represent linear fits of $R_{h,rel}$ to Q_{net} for IDPs with $f_{PRO} > 0.10$ (dashed), IDPs with $f_{PRO} < 0.10$ (solid), and all IDPs (stippled).

Table I

R_h measured for p53(1–93) variants

R_h values are given in Å. $R_{h,DSL}$ refers to DLS measured values at 25°C (from Figure 4) and $R_{h,SEC}$ is room temperature measured values using SEC (from Figure 5B). $R_{h,DSL}$ and $R_{h,SEC}$ averaged for each variant is given by $R_{h,avg}$. $R_{h,seq}$ was calculated using the equation $R_h = (1.24 \cdot f_{PRO} + 0.904)(0.00759 \cdot Q_{net} + 0.963) \cdot 2.49 \cdot N^{0.509}$ determined by Marsh and Forman-Kay (52).

	$R_{h,DSL}$	K_D	$R_{h,SEC}$	$R_{h,avg}$	$R_{h,seq}$
wild type	32.80 ± 1.25	0.160 ±0.008	31.95	32.4	32.2
ALA^-	31.08 ± 1.28	0.205 ±0.005	29.65	30.4	32.2
PRO^-	27.53 ± 1.32	0.249 ±0.005	27.36	27.4	24.3
$ALAPRO^-$	27.71 ± 0.92	0.252 ±0.005	27.18	27.4	24.3