Supporting Information:

Structural Characterization of Protein Phosphatase 1 Regulatory Proteins

Barbara Dancheck[1], Angus C. Nairn[2] and Wolfgang Peti[1,*]

[1]Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, RI, 02912; [2]Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06508.

Sequence analysis. Intrinsically unstructured proteins (IUPs) can be readily detected by numerous bioinformatics programs because of their unique amino acid compositions (*1, 2*). IUPs contain a higher proportion of disorder inducing amino acids, such as charged, small uncharged hydrophilic and proline residues (Gln, Glu, Ser, Lys, Pro, Gly, Ala) (*3*). As shown in Table S1, 53 and 57% of the amino acids in I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$, respectively, are disorder inducing. In turn, many fewer hydrophobic residues, such as Val, Leu, Ile, Met, Phe, Trp and Tyr, are found in IUPs. The total percent of hydrophobic residues in I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ is below 25%, much smaller than for folded proteins. Very few aromatic residues, such as tryptophans, are an additional hallmark of IUPs. Tryptophan is often considered a folding nucleus and the low number of tryptophan residues in particular and the fewer numbers hydrophobic residues in general contribute to the unstructured behavior of IUPs.

Figures S1 and S2 show sequence comparisons of DARPP-$32_{1\text{-}118}$ and I-$2_{9\text{-}164}$ which are color-coded following the pattern of Dyson & Wright (*3*). The color scheme highlights the presence of disorder inducing amino acids in IUPs. For DARPP-32 (Figure S1) 100% sequence conservation (indicated by a red asterisk) can be identified for the PP1 binding motif and the PKA Thr^{34} phosphorylation site. Both sites are necessary for PP1 inhibition. A second phosphorylation site Ser^{102} (CK2), which when phosphorylated increases the rate of phosphorylation of Thr^{34} by PKA, shows much less sequence conservation, but as we show in this report forms a dynamic hotspot. A negatively charged amino acid patch immediately follows this phosphorylation site.

For I-$2_{9\text{-}164}$ (Figure S2), 100% sequence conservation can be detected around the GSK3 Thr^{72} phosphorylation site (indicated by a red asterisk). Thr^{72} is phosphorylated by GSK-3, which is important in the reactivation of the ATP-Mg-dependent I-2:PP1 phosphatase complex. A second conserved region is found in $helix_{127\text{-}154}$ (residues 138, 143, 144 and 146, 147) in I-$2_{9\text{-}164}$ (see description in text and below).

PKA Expression and Purification. A pET15b vector encoding the catalytic subunit α of Protein Kinase A was purchased from Addgene (Addgene plasmid 14921) (*25*). The plasmid was transformed into *Escherichia coli* strain BL21-CodonPlus (DE3)-RIL (Stratagene) and protein expression was carried out in LB after induction with 1 mM IPTG. Cells were lysed by high-pressure homogenization (Avestin C-3 Emulsiflex). For the first purification step the soluble proteins were loaded onto a HisTrap HP column (GE Healthcare) and were eluted with an imidazole gradient. Fractions containing the protein were pooled and dialyzed into low salt buffer. The protein was then loaded onto a MonoQ anion exchange column (GE Healthcare) and eluted in a salt gradient. Fractions containing the pure protein, as confirmed by SDS-PAGE, were pooled, buffer

exchanged into 20 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 2 mM DTT and concentrated to 2 mg/ml. 50% glycerol was added for freezing and storage at -80^{o} C. The activity of the purified PKA was tested by its ability to phosphorylate DARPP-32_{1-118}, which was confirmation by mass spectrometry (Voyager Pro DE MALDI-TOF mass spectrometer, Applied Biosystems).

Choice of databases for chemical shift indexing and secondary structure propensities.

It is important to choose a suitable random coil chemical shift database to reference the experimentally determined chemical shifts against. CSIs for I-2_{9-164} and DARPP-32_{1-118} were calculated using the RefDB (*4*), Wang and Jardetzky (*5*) and Wishart (*6*) random coil databases provided with the SSP program (Figure S3). The RefDB and Wang and Jardetzky databases are based on averaging chemical shifts from the BMRB, while the Wishart database is based on hexapeptides whose chemical shifts were measured in 1 M Urea. Because our proteins are intrinsically unstructured, not chemically denatured, we expected the former two databases to perform better for our proteins. The CSI results using the RefDB and Wang and Jardetzky databases agree outstandingly well. However, slight discrepancies were seen using the Wishart database, as expected. Therefore, all ssp scores were calculated using both the RefDB and Wang and Jardetzky databases as provided with the SSP program, with a seven-residue moving-average window size that excluded data from residues just before prolines.

The re-referencing script provided in the SSP program was used to confirm proper referencing of our shifts to DSS. Because the correction factors determined with the re-reference script were very small (0.075 and -0.106 for I-2_{9-164} using the RefDB and Wang and Jardetzky databases, respectively; 0.108 and -0.106 for Darpp-32_{1-118} using the RefDB and Wang and Jardetzky databases, respectively), results shown are not re-referenced. These determined correction factors are in the usual range, as described in the ssp paper (*7*). It is interesting to note, however, that the correction factors for re-referencing are almost equal in magnitude yet opposite in sign for the two separate databases.

It has been previously shown that neighboring residues may influence chemical shifts (*8, 9*); this is particularly critical in the case of unstructured proteins where it may cause false classification of transient secondary structure. We adjusted our chemical shifts using the Wang and Jardetzky random coil correction factors for neighboring residues (*8*). This database was chosen because it is based on chemical shifts reported in the BMRB, as opposed to peptides in chaotropic solutions. Differences between the sequence corrected ssp scores and the non-sequence corrected scores were limited to the regions of the protein which had ssp scores less than 0.2, i.e. regions with no significant transient structure. All regions defined as having transient secondary structure in I-2_{9-164} (ssp > 0.2 for 5 or more residues) remained the same after sequence correction. The only other area in which there was a difference between the corrected ssp scores and the non-corrected scores was in the last helix in I-2_{9-164} (residues 127 – 154) where the sequence corrected scores were lower (not the case in the two transient helices). The experimental parameters of helix$_{127–154}$ in I-2_{9-164} reflect in many ways a fully formed helix with the experimental CSI values identical to values seen in helices of folded proteins and restricted fast time-scale (ps-ns) backbone motions mirroring values

of helices in well-folded proteins. The Wang and Jardetzky database of correction factors for neighboring residues contains separate sets of correction factors for residues in random coil conformation than for residues in an α-helix. If the α-helix correction factors, instead of the random coil correction factors, were used in this area of the protein chain only, the sequence corrected shifts agreed well with the non-sequence corrected shifts. Thus, results shown are not sequence corrected.

Relaxation data. ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates and the [^{1}H, ^{15}N] heteronuclear NOE (hetNOE) were measured as described previously(*10-12*) at 500 MHz. All spectra were recorded with 2048 points in the direct dimension and 256 points in the indirect dimension. The sweep widths of the direct and indirect dimensions were 6010 Hz and 1140 Hz (I-2) and 5000 Hz and 1076 Hz (DARPP-32 and pDARPP-32). For I-$2_{9\text{-}164}$, T_2 was sampled at 20, 100, 200, 300, 400, 500, 600 and 800 ms; T_1 was measured at 20, 130, 275, 350, 400, 550, 650 and 850 ms. For DARPP-$32_{1\text{-}118}$ and pDARPP-$32_{1\text{-}118}$, T_2 was sampled at 20, 100, 250, 400, 500, 600, 700 and 850 ms; T_1 was measured at 20, 100, 250, 400, 550, 700, 850 and 950 ms. For the hetNOE measurement, an experiment with proton presaturation was interleaved with an experiment without proton presaturation. The relaxation delay for all T_1 and T_2 experiments was 3 s and 5 s for all hetNOE experiments. At least one time point was run in duplicate for each set of experiments.

To test for possible exchange contributions in the R_2 relaxation rates of I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ we re-measured them at 800 MHz with 2048 and 192 points in the direct and indirect dimensions, respectively. The sweep widths of the direct and indirect dimensions were 9615 Hz and 1824 Hz (I-2) and 8013 Hz and 1719 Hz (DARPP-32). I-$2_{9\text{-}164}$ was sampled at 20, 100, 200, 300, 400, 500, 600 and 800 ms and DARPP-$32_{1\text{-}118}$ at 20, 100, 250, 350, 450, 550, 650 and 800 ms. At least one time point was run in duplicate for each set of experiments.

The NMR spectra were processed using NMRPipe version 97.027.12.56 (http://spin.niddk.nih.gov/NMRPipe)(*13*). Processing included fourier transformation, zerofilling and apodization using a qsine multiplication in both dimensions. The final matrix size was 4096*512 points for T_1 and T_2 spectra at 500 MHz, 4096*384 points for T_2 spectra at 800 MHz and 4096*256 for the hetNOE spectra. For I-$2_{9\text{-}164}$, the resolution of the T_1 and T_2 spectra at 500 MHz was 0.73 Hz and 2.23 Hz in the direct and indirect dimensions, respectively; of the T_2 spectra at 800 MHz was 1.17 Hz and 5.34 Hz in the direct and indirect dimensions, respectively; of the hetNOE spectra was 1.47 Hz and 4.45 Hz in the direct and indirect dimensions, respectively. For DARPP-$32_{1\text{-}118}$ and pDARPP-$32_{1\text{-}118}$, the resolution of the T_1 and T_2 spectra at 500 MHz was 0.61 Hz and 2.10 Hz in the direct and indirect dimensions, respectively; of the T2 spectra at 800 MHz was 0.98 Hz and 5.04 Hz in the direct and indirect dimensions, respectively; of the hetNOE spectra was 1.22 Hz and 4.20 Hz in the direct and indirect dimensions, respectively.

Because I-2, DARPP-32 and pDARPP-32 are intrinsically unstructured, they are more prone to degradation than purified structured proteins. Thus, in order to obtain accurate relaxation data it is imperative to measure the relaxation experiments on freshly purified samples; all of our relaxation experiments were measured immediately after purification to ensure the most accurate and reliable data. In addition, the

relaxation experiments were repeated on a second batch of purified proteins; the data from the two sets of experiments agreed within 5% error and most outliers were residues with overlapped peaks.

Analysis of relaxation data. The amide proton chemical shift dispersion is poor in unstructured proteins, leading to a high degree of overlap in their spectra. Only the reasonable degree of nitrogen chemical shift dispersion enables us to use NMR spectroscopy for the quantitative analysis of these proteins. Although it is not common to include data from overlapped peaks in structured proteins, such practices would have eliminated a significant amount of data from I-$2_{9\text{-}164}$, DARPP-$32_{1\text{-}118}$ and pDARPP-$32_{1\text{-}118}$. Thus, we chose to present the data from all assigned residues, regardless of the degree of overlap of the peaks.

R_2 measurements were repeated on freshly prepared samples at 800 MHz (800 MHz spectrometer at the Brandeis University Nuclear Magnetic Resonance Facility, BrUNMR) (Figure S4) to detect possible exchange contributions. Only very few residues with exchange contributions were detected in I-$2_{9\text{-}164}$, including residues 26, 33, 63, 132, 134, 139 and 146. In DARPP-$32_{1\text{-}118}$, exchange contributions were detected for residues 5, 30, 31, 38, 39, 41, 42, 109, 116. These measurements at two magnetic fields (11.8T and 18.8T) show that we are not interpreting conformational exchange contributions in our R_2 rates.

Qualitative analysis of relaxation results. The segmental motion model is a simple model that describes the behavior of random coil polymers very well (*14, 15*). Its major implication is that in long polymer chains (such as denatured proteins) individual segments of a polymer move independently. This model considers only the influence of neighboring residues on R_2 relaxation rates as they decay exponentially with the distance from a certain residue. It has been reported that the R_2 rates of chemically denatured proteins often follow the very simple pattern of a flattened bell-shaped curve, where values in the central region of the protein sequence reach a maxima between 3.5 - 5 Hz while the N- and C-terminal R_2 rates are significantly smaller. This is strongly supported by ^{13}C R_1 measurements in atactic polystyrene (*16*), where the authors showed that for an increase in the polymer molecular weight from 2 to 10 kDa the R_1 rates increased 40%, while they stayed unchanged for larger polymers up to 860 kDa, clearly indicating that larger polymers behave like a number of small independent segments.

The segmental motion model (*14, 15*), using a chain persistence length of 7 and an R_2^{int} constant of 0.225 (depending mainly on temperature, solution viscosity) does not describe the R2 relaxation rates of I-$2_{9\text{-}164}$ or DARPP-$32_{1\text{-}118}$ within reasonable error (Figures S5 A and S6 A, respectively). They deviate significantly from a flat, bell-shaped curve characteristic for the random coil state. IUPs with transient structure elements, such as I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$, behave in a highly domain-like manner. Our data shows that areas without preferred conformations exhibit strict random coil-like behavior, following the 3.5 - 4 Hz prediction of the segmental relaxation model. However, for areas with transient secondary structure (ssp score based), large deviations from this behavior can be detected, proving that additional amino acid based

features need to be considered to describe the R_2 relaxation rates of these proteins in a more accurate manner.

To better understand the reason for this non-random coil behavior, we compared the R_2 rates of I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ to the amino acid specific characteristics hydrophobicity, side chain bulkiness and average area buried on transfer from standard state to folded protein (AABUF) (Figures S5 and S6, respectively). Because the relaxation rate profiles of DARPP-$32_{1\text{-}118}$ and pDARPP-$32_{1\text{-}118}$ are nearly identical, we focus here on the DARPP-$32_{1\text{-}118}$ data only. Hydrophobicity has previously been used to describe areas of restricted backbone motion in a stretch of intrinsically unstructured residues in staphylococcal nuclease (*17*) and denatured lysozyme (*18*); it is expected that hydrophobic residues might be involved in forming short range, highly localized rigid clusters. While hydrophobicity accounts for the region of restricted backbone motion near the second transient helix in I-$2_{9\text{-}164}$ (residues 94 -107), it only describes part of the motion near the first transient helix (residues 49 – 56, but not 39 – 48) and is extremely poor at describing the restricted motion in most of the final helix (residues 125 – 146). For DARPP-$32_{1\text{-}118}$, hydrophobicity describes the restricted backbone motion near the C-terminus (residues 104 – 114), however it shows a strong minima near the transient helix (residues 25 – 32) where the most restricted backbone motions in DARPP-$32_{1\text{-}118}$ occur and has a maxima in regions that are more flexible, such as residues 9 – 21 and 34 – 41. Radius of gyration has also been used to describe motional deviation from random coil behavior in unstructured proteins, particularly for gly and ala rich regions which tend to be more flexible than other areas (*19*). The bulkiness profile predicts the minima in the I-$2_{9\text{-}164}$ R_2 relaxation rates near residues ala-28, cys-85 and gly-122 reasonably well and overall is a better predictor of the R2 pattern than hydrophobicity. Bulkiness also predicts which residues form the more flexible regions of DARPP-$32_{1\text{-}118}$ well, including areas near gly-59 and gln-100 and again is a better indicator of the overall R_2 pattern than hydrophobicity. Finally, a third parameter, AABUF, which takes into account both hydrophobicity and side chain bulkiness, was shown to correlate with relaxation data from denatured apomyoglobin (*19*). Overall the AABUF profile is the best fit for the R_2 relaxation data from I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$, with both minima and maxima occurring at same points along the protein chain. These results provide strong evidence for the contribution of AABUF to the deviations from random coil behavior in the backbone motions of I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$.

In summary, our data strongly suggests that numerous amino acid type characteristics have to be taken into consideration to better understand local rigidity in the low populated transient structure elements of our IUPs. While hydrophobicity has been shown to be important for staphylococcal nuclease (*17*) and denatured lysozyme (*18*), side chain bulkiness seems to play the more dominate in our IUPs. For low populated states, such as $helix_{22\text{-}29}$ in DARPP-$32_{1\text{-}118}$ or $helix_{36\text{-}42}$ in I-$2_{9\text{-}164}$, AABUF profiles correlate well with the experimental R_2 profiles. Therefore, the dynamics of these low-populated states are dictated by the individual residue specific characteristics of side chain size and hydrophobicity. However, highly populated secondary structure elements, such as $helix_{127\text{-}154}$ in I-$2_{9\text{-}164}$, can only be correlated in profile, not in magnitude, clearly indicating that the correlated helical behavior dominates over individual amino acid properties.

Spectral density mapping. The program Relax version 1.2 (www.nmr-relax.com) was used to calculate the spectral densities of I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ at three frequencies, J(0), J(ω_N) and J(0.87ω_H), based on previous publications (*20, 21*) (Figures S7 (I-$2_{9\text{-}164}$) and S8 (DARPP-$32_{1\text{-}118}$)). All errors were set to 5%, based on our experimental R_1 and R_2 errors. The NH bond length was set to 0.102 nm, the CSA to -172 ppm and the spectrometer frequency to 500 MHz (^{1}H). 5000 Monte Carlo simulations were run in each experiment. As expected, R_2 makes the greatest contribution to J(0), with a correlation coefficient of 0.9994 for I-$2_{9\text{-}164}$ and 0.9981 for DARPP-$32_{1\text{-}118}$ when R_2 is plotted against J(0). Maxima occur in the plot of J(0) versus residue in the same regions as the transient helices in I-$2_{9\text{-}164}$ (residues 44 – 56, 95 – 108 and 126 – 156) and DARPP-$32_{1\text{-}118}$ (residues 23 – 32), where the height of J(0) scales with that of the R_2 rates and the ssp scores. J(0.87ω_H) shows less variation in both proteins. However J(0.87ω_H) tends to be at a minimum, indicating low backbone flexibility, when AABUF is at its highest for I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ (Figure S9). This correlation between J(0.87ω_H) minima and AABUF maxima was first described by Schwarzinger, *et al.* (*19*), who predict that the correlation is based on the existence of local hydrophobic interactions which lead to restricted backbone motions on the sub-nanosecond timescale. However, when the hydrophobicity is plotted and compared with the J(0.87ω_H) profile, only a few of the correlations can be observed. Side chain bulkiness, another amino acid specific parameter, also shows a good correlation with minima in the J(0.87ω_H) profile. Therefore, we believe that hydrophobicity itself is not the solitary reason for the AABUF J(0.87ω_H) profile correlation. Our data strongly suggests that the size of the amino acid side chains and their ability to sterically hinder certain rotations plays a critical role in the restriction of sub-nanosecond timescale motions in IUPs.

Spin labeling with MTSL. Raw data from spin labeling of I-$2_{9\text{-}164}$, DARPP-$32_{1\text{-}118}$ and pDARPP-$32_{1\text{-}118}$ is shown in Figures S10, S13 and S16, respectively. Overlap of [^{1}H, ^{15}N] HSQCs of C85S I-$2_{9\text{-}164}$ or C72S DARPP-$32_{1\text{-}118}$ with their respective WT and single cysteine mutant [^{1}H, ^{15}N] HSQCs are shown in Figures S11 and S14, respectively, and indicate that the mutations do not affect the overall conformations of I-$2_{9\text{-}164}$ or DARPP-$32_{1\text{-}118}$. Similarly, overlap of [^{1}H, ^{15}N] HSQCs with and without the MTSL label of WT and single cyteine mutant I-$2_{9\text{-}164}$ or DARPP-$32_{1\text{-}118}$ are shown in Figures S12 and S15, respectively, and indicate that the addition of MTSL does not alter the overall conformation of either protein. Analysis of paramagnetic relaxation enhancement data has to be done carefully since the enhancement is correlated with the inverse sixth power of the distance and larger enhancement can easily be over-interpreted. Furthermore, artifacts caused by nonspecific interactions between MTSL and aromatic residues might arise. By employing numerous spin labels throughout the sequence, artifacts such as non-specific interactions between MTSL and aromatic residues are easily detectable. Rapid purification and HSQC measurement was done to ensure the spectra were recorded before N- and/or C-terminal degradation of the proteins could occur. Protein concentrations were 0.1-1 mM; all spectra were recorded with 2048 points in the direct dimension and 512 points in the indirect dimension. The sweep widths of the direct and indirect dimensions were 6010 Hz and 1115 Hz (I-$2_{9\text{-}164}$)

and 6010 Hz and 1076 Hz (DARPP-$32_{1\text{-}118}$ and pDARPP-$32_{1\text{-}118}$). The diamagnetic and paramagnetic [^{1}H, ^{15}N] HSQCs were processed with NMRPipe version 97.027.12.56 (http://spin.niddk.nih.gov/NMRPipe), using the same parameters as the relaxation spectra (*13*). The final matrix size was 4096*1024 for all spectra. For I-$2_{9\text{-}164}$, the resolution of all diamagnetic and paramagnetic spectra was 0.73 Hz and 1.09 Hz in the direct and indirect dimensions, respectively. For DARPP-$32_{1\text{-}118}$ and pDARPP-$32_{1\text{-}118}$, the resolution of all diamagnetic and paramagnetic spectra was 0.73 Hz and 1.05 Hz in the direct and indirect dimensions, respectively. The peak intensities of the diamagnetic and paramagnetic spectra were analyzed with NMRView (*22*). The ratios of the peak intensities in the paramagnetic spectra to those in the diamagnetic spectra were normalized to 1 and plotted against residue number. As described for the relaxation data, a high degree of chemical shift overlap is found in unstructured protein spectra. We chose to exclude the data from residues with highly overlapped peaks only if the data points were extreme outliers.

Phosphorylation of Thr^{34} on DARPP-$32_{1\text{-}118}$ does not affect its overall conformation. The MTSL profile of pDARPP-$32_{1\text{-}118}$ (Figure S16A) closely resembles that of DARPP-$32_{1\text{-}118}$, indicating that long range contacts present in DARPP-$32_{1\text{-}118}$ are not altered by phosphorylation of Thr^{34}. Additionally, the NOE pattern of residues in the vicinity of Thr^{34} are largely unaffected by phosphorylation of Thr^{34} (Figure S16B).

I-$2_{9\text{-}164}$ are DARPP-$32_{1\text{-}118}$ are active in their unstructured forms. To experimentally verify the activity of I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ we tested their ability to form complexes with recombinantly expressed PP1 (Figure S17). It is important to note that identical constructs were used in numerous cell based studies as well (*23, 24*). A construct representing human PP1-α was subcloned into a pET28a-modified vector which encodes a $Thio_6His_6$ expression/purification tag and a TEV (tobacco etch virus protease) cleavage site (*25*). The plasmid was transformed into *Escherichia coli* strain BL21-CodonPlus (DE3)-RIL (Stratagene). The expression of PP1 was carried out by growing freshly transformed cells in LB media after induction with 1 mM IPTG. For purification, cells were lysed by high-pressure homogenization (Avestin C-3 Emulsiflex) and the soluble protein fraction was loaded onto a Ni-NTA column (Invitrogen); PP1 was eluted using an imidazole gradient. The purity of apo-PP1 was assayed by SDS-PAGE gel electrophoresis, and PP1 activity was confirmed using the commonly used pNPP assay. To produce the PP1:$I2_{9\text{-}164}$ complex, PP1 was bound to Ni-NTA resin; subsequently, purified $I2_{9\text{-}164}$ was incubated on this column and the complex was eluted with imidazole. The complex was then loaded onto a Superdex 75 26/60 size exclusion column (GE Healthcare) for final purification To produce the PP1:pDARPP-$32_{1\text{-}118}$ complex PP1 was directly eluted into purified, thiophosphorylated DARPP-$32_{1\text{-}118}$. A phenyl sepharose column was used as an intermediate purification step before loading the complex onto a Superdex 75 26/60 size exclusion column.

Hydrodynamic radius - dynamic light scattering measurements. As shown in Table S2, I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ have an experimentally determined compaction factor of ~30% when compared with chemically denatured polypeptides.

*The effect of cell crowding***.** The average protein concentration in a cell is very high (300-400 mg/ml). It has therefore been suggested that IUPs might behave differently, i.e. adopt secondary structure or even fold completely, in a native, crowded intracellular environment of metabolites and other macromolecules. Numerous methods for experimentally simulating the crowded cellular environment have been proposed. One of simplest ways is to examine the IUPs in the presence of highly concentrated glucose (2.2M), which simulates the high viscosity of the cell (*26*). We used CD and NMR spectroscopy to test if the overall structures of I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ change upon the addition of a crowded, highly viscous environment.

CD spectra were recorded in the presence and absence of 2.2 M glucose using a Jasco J-815 spectrometer (Jasco Inc, Easton, MD). Five experiments were averaged for each sample; all spectra were recorded from 205 nm to 260 nm with 0.5 nm intervals. Samples contained 5-10 μM protein in the typical NMR buffer: 20 mM $NaPO_4$ (pH 6.5), 50 mM NaCl for I-2; 50 mM $NaPO_4$ (pH 5.5), 50 mM NaCl for DARPP-32 with or without glucose. All measurements were performed at 298 K, using a cell with a path length of 0.2 cm. % α-helix content was calculated using % alpha-helix = (-[θ]222 nm + 3000)/39000) (*27*).

There are no significant differences between the CD spectra of I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$ in the presence or absence of 2.2 M glucose (Figure S18). In addition, there are no significant differences in the % α-helical content in any of the proteins: 26.7% and 27.8% for I-$2_{9\text{-}164}$ and 18.7% and 18.7% for DARPP-$32_{1\text{-}118}$ in the absence and presence of 2.2 M glucose, respectively. This data indicates that a crowded/viscous environment, mimicked by high glucose concentration, does not change the secondary structure properties of these three proteins.

Further, 2D [^{1}H, ^{15}N] HSQC spectra were recorded on I-$2_{9\text{-}164}$ (250 μM) and DARPP-$32_{1\text{-}118}$ (250 μM) in the presence of 2.2 M glucose and were compared to the same spectra recorded in the absence of glucose. 2048 points were recorded in the direct dimension and 512 points in the indirect dimension, with the same sweep widths as the relaxation experiments at 500 MHz. Although there was increased line broadening in the spectra recorded in the presence of glucose, as expected from slower overall tumbling in a higher viscosity solution, there was no overall change in topology of the 2D [^{1}H, ^{15}N] HSQC spectra recorded in glucose, as would be expected if the proteins adopted additional structure (Figure S19).

References

1. Romero, P., Obradovic, Z., and Dunker, A. K. (2004) Natively disordered proteins: functions and predictions, *Applied bioinformatics 3*, 105-113.
2. Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, *Bioinformatics (Oxford, England) 21*, 3433-3434.
3. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nature reviews 6*, 197-208.

4. Zhang, H., Neal, S., and Wishart, D. S. (2003) RefDB: a database of uniformly referenced protein chemical shifts, *Journal of biomolecular NMR 25*, 173-195.
5. Wang, Y., and Jardetzky, O. (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data, *Protein Sci 11*, 852-861.
6. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *Journal of biomolecular NMR 5*, 67-81.
7. Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation, *Protein Sci 15*, 2795-2804.
8. Wang, Y., and Jardetzky, O. (2002) Investigation of the neighboring residue effects on protein chemical shifts, *J Am Chem Soc 124*, 14075-14084.
9. Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts, *J Am Chem Soc 123*, 2970-2978.
10. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry 28*, 8972-8979.
11. Akke, M., and Palmer, A. G. (1996) Monitoring macromolecular motions on microsecond to millisecond time scales by R_{1p} - R_1 constant relaxation time NMR spectroscopy., *Journal of the American Chemical Society 118*, 2.
12. Palmer, A. G., and Case, D. A. (1992) Molecular dynamics analysis of NMR relaxation in a zinc-finger peptide., *Journal of the American Chemical Society 114*, 9.
13. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *Journal of biomolecular NMR 6*, 277-293.
14. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea, *Biochemistry 36*, 8977-8991.
15. Wirmer, J., Peti, W., and Schwalbe, H. (2006) Motional properties of unfolded ubiquitin: a model for a random coil protein, *Journal of biomolecular NMR 35*, 175-186.
16. Allerhand, A., and Hailstone, R. K. (1972) Carbon-13 fourier transform nuclear magnetic resonance. X. Effect of molecular weight on ^{13}C spin-lattice relaxation times of polystyrene in solution., *The Journal of Chemical Physics 56*, 3718-3720.
17. Alexandrescu, A. T., and Shortle, D. (1994) Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease, *Journal of molecular biology 242*, 527-546.
18. Buck, M., Schwalbe, H., and Dobson, C. M. (1996) Main-chain dynamics of a partially folded protein: 15N NMR relaxation measurements of hen egg white

lysozyme denatured in trifluoroethanol, *Journal of molecular biology 257*, 669-683.
19. Schwarzinger, S., Wright, P. E., and Dyson, H. J. (2002) Molecular hinges in protein folding: the urea-denatured state of apomyoglobin, *Biochemistry 41*, 12681-12686.
20. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1995) Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer, *Biochemistry 34*, 868-878.
21. Lefevre, J. F., Dayie, K. T., Peng, J. W., and Wagner, G. (1996) Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions, *Biochemistry 35*, 2674-2686.
22. Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules, *Methods in molecular biology (Clifton, N.J 278*, 313-352.
23. Huang, H. B., Horiuchi, A., Watanabe, T., Shih, S. R., Tsay, H. J., Li, H. C., Greengard, P., and Nairn, A. C. (1999) Characterization of the inhibition of protein phosphatase-1 by DARPP-32 and inhibitor-2, *J Biol Chem 274*, 7870-7878.
24. Lin, T. H., Huang, Y. C., Chin, M. L., Chen, Y. C., Jeng, H. H., Lin, F. M., Shiao, M. S., Horiuchi, A., Greengard, P., Nairn, A. C., and Huang, H. B. (2004) 1H, 15N, and 13C resonance assignments of DARPP-32 (dopamine and cAMP-regulated phosphoprotein, Mr. 32,000)--a protein inhibitor of protein phosphatase-1, *Journal of biomolecular NMR 28*, 413-414.
25. Peti, W., and Page, R. (2007) Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost, *Protein expression and purification 51*, 1-10.
26. Dedmon, M. M., Patel, C. N., Young, G. B., and Pielak, G. J. (2002) FlgM gains structure in living cells, *Proceedings of the National Academy of Sciences of the United States of America 99*, 12681-12684.
27. Woody, R. W. (1995) Circular dichroism, *Methods in enzymology 246*, 34-71.

Supporting Information – Tables

Table S1: Amino acid composition of DARPP-$32_{1\text{-}118}$ and I-$2_{9\text{-}164}$ sequences.

	I-$2_{9\text{-}164}$	% total protein	**DARPP-$32_{1\text{-}118}$**	% total protein
Gln	6		8	
Ser	14		14	
Pro	8		15	
Glu	19		14	
Lys	16		6	
Sum (Q,S,P,E,K)	63	*40.1*	57	*48.3*
Gly	7		5	
Ala	13		6	
Sum (G,A)	20	12.7	11	9.3
Sum (Q,S,P,E,K,G,A)	83	**52.9**	68	**57.6**
Val	2		4	
Leu	13		9	
Ile	8		4	
Met	8		2	
Phe	1		2	
Trp	1		0	
Tyr	5		2	
Sum (V,L,I,M,F,W,Y)	38	**24.2**	23	**19.5**

Table S2: Results of the hydrodynamic radius (R_h) measurements using dynamic light scattering.

	Residues	Calc. R_h folded [Å]	Calc. R_h unfolded [Å]	Exp. R_h [Å]	**Collapse factor**
DARPP-$32_{1\text{-}118}$	121	20.1	34	30±2	**~29%**
I-$2_{9\text{-}164}$	158	22.3	39.7	35±2	**~27%**

Supporting Information – Figures

Figure S1: Sequence alignment of known or proposed DARPP-$32_{1\text{-}118}$ sequences in different organisms. The color code was chosen according to Dyson and Wright to highlight the amino-acid-residue preferences of IUPs[8]. Green: small residues, uncharged hydrophilic residues and proline residues (Asn, Gln, Ser, Thr, Gly, Ala, Pro); yellow: hydrophobic residues (Val, Leu, Ile, Phe, Tyr); red: acidic residues (Asp, Glu); blue: basic residues (Lys, Arg, His); purple: low-frequency residues (Cys, Trp). We were able to find rat, human, mouse, bovine, frog and zebrafish sequences (see also Figure S2) for both IUPs in Genebank (http://www.ncbi.nlm.nih.gov/Genbank/). Alignments were performed using the program ClustLW[39].

Figure S2: Sequence alignment of known or proposed I-$2_{9\text{-}164}$ (human sequence was used in our studies) sequences in different organisms. Color code was chosen as described in Figures S1.

Figure S3. CSI scores of (A) I-$2_{9\text{-}164}$, (B) DARPP-$32_{1\text{-}118}$, and (C) pDARPP-$32_{1\text{-}118}$ indicate the presence of three transient helices in I-$2_{9\text{-}164}$, and one transient helix in DARPP-$32_{1\text{-}118}$; no additional transient secondary structure occurs upon phosphorylation of p-DARPP-$32_{1\text{-}118}$. The ($\Delta C\alpha - \Delta C\beta$) chemical shift delta is reported, calculated using the RefDB (blue)[11], Wang and Jardetzky (pink)[12], and Wishart (orange)[13] random coil databases. Cartoon representations of secondary structure (light grey cylinders correspond to transient α-helices) across the bottom of each are based on ssp scores > 0.2 over five consecutive residues or more.

Figure S4. Measurement of R_2 relaxation rates at two different magnetic field strengths indicate chemical exchange contributions to the R2 rates of only a few residues in I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$. R_2 relaxation rates of (A) I-$2_{9\text{-}164}$ and (B) DARPP-$32_{1\text{-}118}$ were measured at 500 MHz (orange) and 800 MHz (grey). Large differences in the R_2 rates between two different field strengths indicate the contribution of chemical exchange factors to the R_2 values. Residues with chemical exchange are designated by a red asterisk, and include residues 26, 33, 63, 132, 134, 139, and 146 in I-$2_{9\text{-}164}$ and residues 5, 30, 31, 38, 39, 41, 42, 109, and 116 in DARPP-$32_{1\text{-}118}$.

Figure S5. Qualitative analysis of the R_2 relaxation rates of I-$2_{9\text{-}164}$ indicates that AABUF is the best amino acid specific property to describe R_2 rates in an IUP with transient secondary structure. R_2 relaxation rates are shown as black squares; red lines indicate (A) segmental motion model, (B) hydrophobicity profile, (C) side-chain bulkiness profile, (D) average area buried upon folding (AABUF) profile. Cartoon representations of secondary structure (light grey cylinders correspond to transient α-helices) across the bottom of each are based on ssp scores > 0.2 over five consecutive residues or more.

Figure S6. Qualitative analysis of the R_2 relaxation rates of DARPP-$32_{1\text{-}118}$ indicates that AABUF is the best amino acid specific property to describe R_2 rates in an IUP with transient secondary structure. R_2 relaxation rates are shown as black squares; red lines indicate (A) segmental motion model, (B) hydrophobicity profile, (C) side-chain

bulkiness profile, (D) average area buried upon folding (AABUF) profile. Cartoon representations of secondary structure (light grey cylinders correspond to transient α-helices) across the bottom of each are based on ssp scores > 0.2 over five consecutive residues or more.

Figure S7. Deviations of relaxation rates and spectral density values from random coil show good agreement with the regions of transient structure in I-$2_{9\text{-}164}$. (A) R_2, (B) R_1, (C) hetNOE, (D) J(0), (E) J(ω_N) [J(50)], (F) J(0.87 ω_H) [J(435)]. Red line in (A) indicates R_2 rates calculated using the segmental random coil model. Positive hetNOEs and significant deviations of R_2 rates from random coil, indicating regions of restricted backbone motion, agree well with the presence of transient helices and are reflected in increased values of the spectral density at J(0) and decreased values of the spectral density at J(0.87 ω_H). R_1 rates and J(ω_N) are largely featureless. Cartoon representations of secondary structure (light grey cylinders correspond to transient α-helices) across the bottom of each are based on ssp scores > 0.2 over a stretch of five residues or more.

Figure S8. Deviations of relaxation rates and spectral density values from random coil show good agreement with the regions of transient structure in DARPP-$32_{1\text{-}118}$. (A) R_2, (B) R_1, (C) hetNOE, (D) J(0), (E) J(ω_N) [J(50)], (F) J(0.87 ω_H) [J(435)]. Red line in (A) indicates R_2 rates calculated using the segmental random-coil model. Positive hetNOEs and significant deviations of R_2 rates from random coil, indicating regions of restricted backbone motion, agree well with the presence of transient helices and are reflected in increased values of the spectral density at J(0) and decreased values of the spectral density at J(0.87 ω_H). R_1 rates and J(ω_N) are largely featureless. Cartoon representations of secondary structure (light grey cylinders correspond to transient α-helices) across the bottom of each are based on ssp scores > 0.2 over a stretch of five residues or more.

Figure S9. Minima in the spectral density at J(0.87 ω_H) corresponds to maxima in the to average area buried upon folding (AABUF) profiles for I-$2_{9\text{-}164}$ and DARPP-$32_{1\text{-}118}$. J(0.87 ω_H) of (A) I-$2_{9\text{-}164}$ and (B) DARPP-$32_{1\text{-}118}$ is plotted as black squares, AABUF is plotted as a red line. Minima in J(0.87 ω_H) correspond to maxima in AABUF, indicating that local interactions, mediated by both hydrophobicity and side chain volume, contribute to the spectral density at J(0.87 ω_H).

Figure S10. I-$2_{9\text{-}164}$ MTSL profiles show the presence of long range contacts in I-$2_{9\text{-}164}$. (A) G12C, (B) D47C, (C) C85 (WT), (D) S121C, and (E) N152C. Single cysteine mutants were prepared as described in the supporting material, and the paramagnetic spin label MTSL was attached to the cys residues. [^{1}H, ^{15}N] HSQC spectra were recorded in the presence (paramagnetic) and absence (diamagnetic) of MTSL. The ratio of peak intensities in the paramagnetic spectra to those in the diamagnetic spectra were normalized to 1 and plotted along the chain length. A ratio of 1 indicates no peak broadening by the spin label, thus that residue is not within 15 Å of the spin label; a ratio of 0 indicates significant peak broadening and that the residue spends time in very close

proximity to the spin label. Red asterisks indicate the position of the MTSL spin label (see details in text).

Figure S11. Single cysteine mutations on I-$2_{9\text{-}164}$ do not affect the overall structure of I-$2_{9\text{-}164}$. Peak shifts between the C85S [^{1}H,^{15}N] HSQC spectrum (blue) and those of WT, G12C, D47C, S121C, or N152C (black) are minor and are restricted to residues in the vicinity of the mutations. A few shifted peaks can be identified surrounding the single amino acid point mutation. This is due to the different CSA of the new amino acid and the changed local electrostatic and hydrophobic interactions. This is typical for 2D [^{1}H,^{15}N] HSQC of single amino acid point mutations. The overall chemical shift pattern is identical to WT.

Figure S12. Addition of the MTSL spin label to WT and single cysteine mutants of I-$2_{9\text{-}164}$ do not affect the overall structure of I-$2_{9\text{-}164}$. 2D [^{1}H,^{15}N] HSQC spectra without (blue) MTSL are superimposed with those with (black) MTSL. The black MTSL spectrum is plotted at lower peak heights, when compared to the blue without spectrum, to compensate for peak broadening due to the expected enhanced relaxation. This enables us to show more superimposed peaks. Peak broadening of resonances is due to paramagnetic relaxation enhancement of the free electron introduced via the MTSL label.

Figure S13. Darpp-$32_{1\text{-}118}$ MTSL profiles show the presence of long range contacts in DARPP-$32_{1\text{-}118}$. (A) G5C, (B) A36C, (C) S51C, (D) C72 (WT), and (E) S94C. Single cysteine mutants were prepared as described in the supporting material, and the paramagnetic spin label MTSL was attached to the cys residues. [^{1}H, ^{15}N] HSQC spectra were recorded in the presence (paramagnetic) and absence (diamagnetic) of MTSL. The ratio of peak intensities in the paramagnetic spectra to those in the diamagnetic spectra were normalized to 1 and plotted along the chain length. A ratio of 1 indicates no peak broadening by the spin label, thus that residue is not within 15 Å of the spin label; a ratio of 0 indicates significant peak broadening and that the residue spends time in very close proximity to the spin label. Red asterisks indicate the position of the MTSL spin label (see details in text).

Figure S14. Single cysteine mutations on DARPP-$32_{1\text{-}118}$ do not affect the overall structure of DARPP-$32_{1\text{-}118}$. Peak shifts between the C72S [^{1}H, ^{15}N] HSQC spectrum (blue) and those of WT, G5C, A36C, S51C, or S94C (black) are minor and are restricted to residues in the vicinity of the mutations. A few shifted peaks can be identified exactly surrounding the single amino acid point mutation. This is due to the different CSA of the new amino acid and the changed local electrostatic and hydrophobic interactions. This is typical for 2D [^{1}H,^{15}N] HSQC of single amino acid point mutations. The overall chemical shift pattern is identical to WT.

Figure S15. Addition of the MTSL spin label to WT and single cysteine mutants of DARPP-$32_{1\text{-}118}$ does not affect the overall structure of DARPP-$32_{1\text{-}118}$. 2D [^{1}H,^{15}N] HSQC spectra without (blue) MTSL are superimposed with those with (black) MTSL. The black MTSL spectrum is plotted at lower peak heights, when compared to the blue

without spectrum, to compensate for peak broadening due to the expected enhanced relaxation. This enables us to show more superimposed peaks. Peak broadening of resonances is due to paramagnetic relaxation enhancement of the free electron introduced via the MTSL label.

Figure S16. Phosphorylation of DARPP-32 Thr^{34} does not change the overall conformation of DARPP-32_{1-118}. (A) pDarpp-32_{1-118} MTSL profile does not differ significantly from that of DARPP-32_{1-118}. (B) ^{15}N-resolved [^{1}H,^{1}H] NOESY strips are presented. Left: DARPP-32_{1-118}; Right: pDARPP-32_{1-118}. NOE profiles for each residue in the vicinity of Thr^{34} remain largely unaffected upon phosphorylation of Thr^{34}.

Figure S17. I-2_{9-164} and DARPP-32_{1-118} are active in their unstructured states. Complexes between (A) I-2_{9-164} or (B) pDARPP-32_{1-118} and PP1 were formed (according to procedures described in the text) and loaded onto a size exclusion column.

Figure S18. High glucose concentrations, which have been proposed to mimic crowdedness due to high viscosity, do not affect the secondary structure of I-2_{9-164} or DARPP-32_{1-118} as shown by CD spectroscopy. (A) I-2_{9-164}, (B) DARPP-32_{1-118}. CD spectra of the proteins were recorded in buffer (black) and in the presence of 2.2 M glucose (pink). The black line indicates 222 nm. There are no significant differences between the spectra recorded in buffer and those recorded in 2.2 M glucose, indicating no change in secondary structure content of the proteins in the presence of 2.2 M glucose.

Figure S19. High glucose concentrations do not affect the overall topology of DARPP-32_{1-118}, as shown by NMR spectroscopy. [^{1}H, ^{15}N] HSQC spectra were recorded in NMR buffer (purple) or in the presence of 2.2 M glucose (blue). The lack of dispersion of the peaks in the ^{1}H dimension in both spectra implies that DARPP-32_{1-118} does not fold in the presence of 2.2 M glucose, a highly viscous solution that was proposed to mimic cellular crowding.

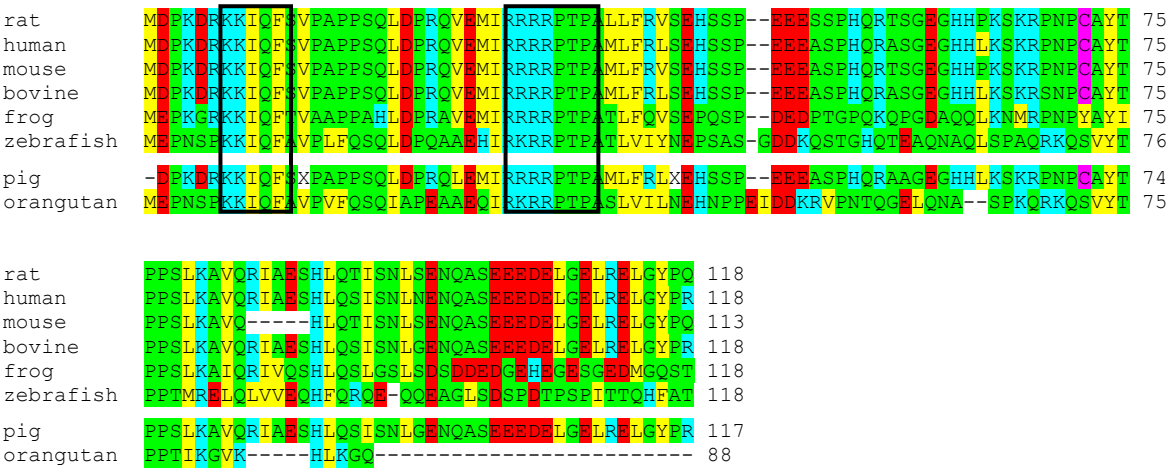
rat MDPKDRKKIQFSVPAPPSQLDPRQVEMIRRRRPTPALLFRVSEHSSP--EEESSPHQRTSGEGHHPKSKRPNPCAYT 75
human MDPKDRKKIQFSVPAPPSQLDPRQVEMIRRRRPTPAMLFRLSEHSSP--EEEASPHQRASGEGHHLKSKRPNPCAYT 75
mouse MDPKDRKKIQFSVPAPPSQLDPRQVEMIRRRRPTPAMLFRVSEHSSP--EEEASPHQRTSGEGHHPKSKRPNPCAYT 75
bovine MDPKDRKKIQFSVPAPPSQLDPRQVEMIRRRRPTPAMLFRLSEHSSP--EEEASPHQRASGEGHHLKSKRSNPCAYT 75
frog METKGRKKIQFTVAAPPAHLDPRAVEMIRRRRPTPATLFQVSEPQSP--DEDPTGPQKQPGDAQQLKNMRPNPYAYI 75
zebrafish METNSPKKIQFAVPLFQSQLDPQAALHIRKRRPTPATLVIYNEPSAS-GDDKQSTGHQTEAQNAQLSPAQRKQSVYT 76
pig -DPKDRKKIQFXVPAPPSQLDPRQLEMIRRRRPTPAMLFRLXEHSSP--EEEASPHQRAAGEGHHLKSKRPNPCAYT 74
orangutan METNSPKKIQFAVPVFQSQIAPEAAEQIRKRRPTPASLVILNEHNPPEIDDKRVPNTQGELQNA--SPKQRKQSVYT 75
rat PPSLKAVQRIAESHLQTISNLSENQASEEEDELGELREL GYPQ 118
human PPSLKAVQRIAESHLQSISNLNENQASEEEDELGELREL GYPR 118
mouse PPSLKAVQ-----HLQTISNLSENQASEEEDELGELREL GYPQ 113
bovine PPSLKAVQRIAESHLQSISNLGENQASEEEDELGELREL GYPR 118
frog PPSLKAIQRIVQSHLQSLGSLSDSDDEDGEHEGESGEDMGQST 118
zebrafish PPTMRELQLVVEQHFQRQE-QQEAGLSDSPDTPSPITTQHFAT 118
pig PPSLKAVQRIAESHLQSISNLGENQASEEEDELGELREL GYPR 117
orangutan PPTIKGVK-----HLKGQ------------------------- 88

Figure S1

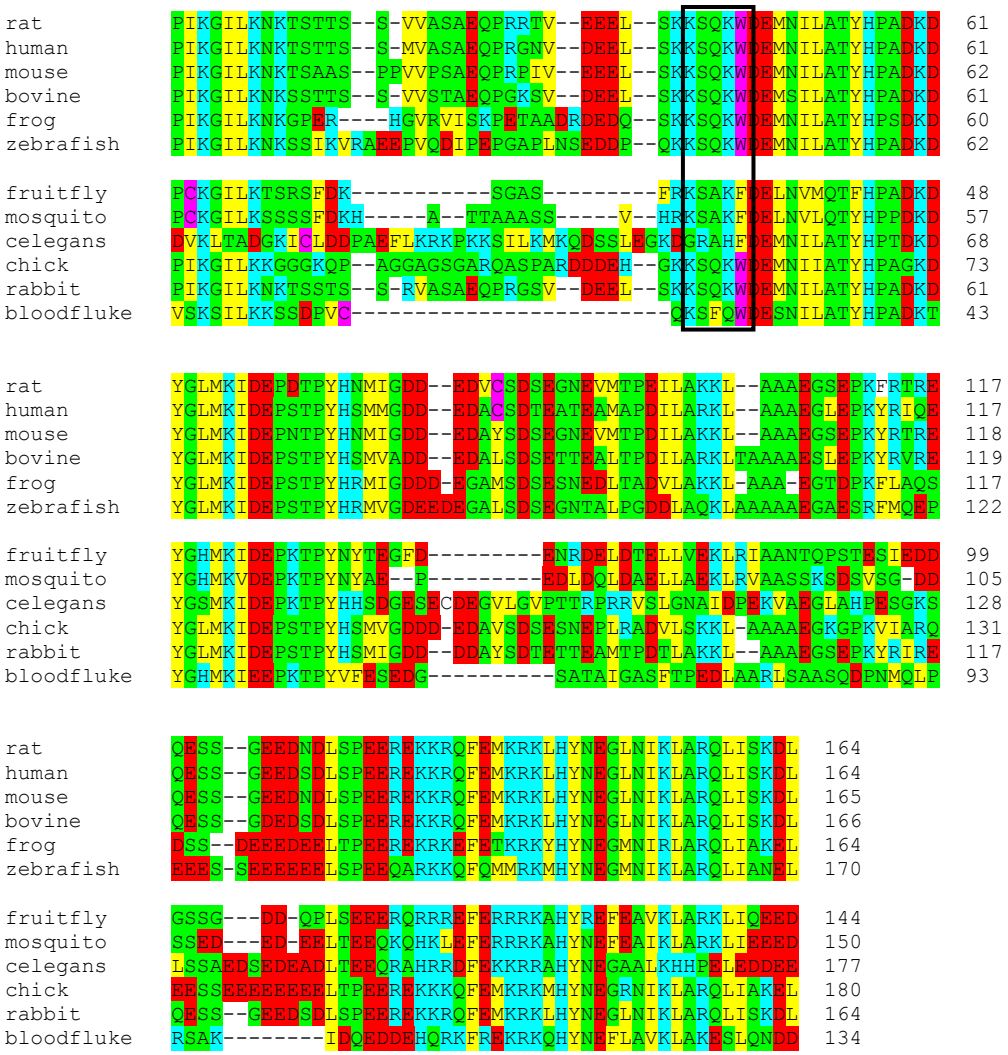
rat PIKGILKNKTSTTS--S-VVASAEQPRRTV--EEEL--SKKSQKWDEMNILATYHPADKD 61
human PIKGILKNKTSTTS--S-MVASAEQPRGNV--DEEL--SKKSQKWDEMNILATYHPADKD 61
mouse PIKGILKNKTSAAS--PPVVPSAEQPRPIV--EEEL--SKKSQKWDEMNILATYHPADKD 62
bovine PIKGILKNKSSTTS--S-VVSTAEQPGKSV--DEEL--SKKSQKWDEMSILATYHPADKD 61
frog PIKGILKNKGPER----HGVRVISKPETAADRDEDQ--SKKSQKWDEMNILATYHPSDKD 60
zebrafish PIKGILKNKSSIKVRAEEPVQDIPEPGAPLNSEDDF--QKKSQKWDEMNILATYHPADKD 62
fruitfly PCKGILKTSRSFDK----------SGAS---------FRKSAKFDELNVMQTFHPADKD 48
mosquito PCKGILKSSSSFDKH-----A--TTAAASS-----V--HRKSAKFDELNVLQTYHPPDKD 57
celegans EVKLTADGKICLDDPAEFLKRKPKKSILKMKQDSSLEGKDGRAHFDEMNILATYHPTDKD 68
chick PIKGILKKGGGKQP--AGGAGSGARQASPARDDDEH--GKKSQKWDEMNIIATYHPAGKD 73
rabbit PIKGILKNKTSSTS--S-RVASAEQPRGSV--DEEL--SKKSQKWDEMNILATYHPADKD 61
bloodfluke VSKSILKKSSDPVC----------QKSFQWDESNILATYHPADKT 43
rat YGLMKIDEPDTPYHNMIGDD--EDVCSDSEGNEVMTPEILAKKL--AAAEGSEPKFRTRE 117
human YGLMKIDEPSTPYHSMMGDD--EDACSDTEATEAMAPDILARKL--AAAEGLEPKYRIQE 117
mouse YGLMKIDEPNTPYHNMIGDD--EDAYSDSEGNEVMTPDILAKKL--AAAEGSEPKYRTRE 118
bovine YGLMKIDEPSTPYHSMVADD--EDALSDSETTEALTPDILARKLTAAAAESLEPKYRVRE 119
frog YGLMKIDEPSTPYHRMIGDDD-EGAMSDSESNEDLTADVLAKKL-AAA-EGTDPKFLAQS 117
zebrafish YGLMKIDEPSTPYHRMVGDEEDEGALSDSEGNTALPGDDLAQKLAAAAAEGAESRFMQEP 122
fruitfly YGHMKIDEPKTPYNYTEGFD---------ENRDELDTELLVEKLRIAANTQPSTESIEDD 99
mosquito YGHMKVDEPKTPYNYAE--P---------EDLDQLDAELLAEKLRVAASSKSDSVSG-DD 105
celegans YGSMKIDEPKTPYHHSDGESECDEGVLGVPTTRPRRVSLGNAIDPEKVAEGLAHPESGKS 128
chick YGLMKIDEPSTPYHSMVGDDD-EDAVSDSESNEPLRADVLSKKL-AAAAEGKGPKVIARQ 131
rabbit YGLMKIDEPSTPYHSMIGDD--DDAYSDTETTEAMTPDTLAKKL--AAAEGSEPKYRIRE 117
bloodfluke YGHMKIEEPKTPYVFESEDG----------SATAIGASFTPEDLAARLSAASQDPNMQLP 93
rat QESS--GEEDNDLSPEEREKKRQFEMKRKLHYNEGLNIKLARQLISKDL 164
human QESS--GEEDSDLSPEEREKKRQFEMKRKLHYNEGLNIKLARQLISKDL 164
mouse QESS--GEEDNDLSPEEREKKRQFEMKRKLHYNEGLNIKLARQLISKDL 165
bovine QESS--GDEDSDLSPEEREKKRQFEMKRKLHYNEGLNIKLARQLISKDL 166
frog DSS--DEEEDEELTPEEREKRKLFETKRKYHYNEGMNIRLARQLIAKEL 164
zebrafish EEES-SEEEEEELSPEEQARKKQFQMMRKMHYNEGMNIKLARQLIANEL 170
fruitfly GSSG---DD-QPLSEEEREQRRREFERRRKAHYREFEAVKLARKLIQEED 144
mosquito SSED---ED-EELTEEQKQHKLEFERRRKAHYNEFEAIKLARKLIEEED 150
celegans LSSAEDSEDEADLTEEQRAHRRQFEKKRRAHYNEGAALKHHPELEDDEE 177
chick EESSEEEEEEEELTPEEREKKKQFEMKRKMHYNEGRNIKLARQLIAKEL 180
rabbit QESS--GEEDSDLSPEEREKKRQFEMKRKLHYNEGLNIKLARQLISKDL 164
bloodfluke RSAK--------IDQEDDEHQRKFREKRKQHYNEFLAVKLAKESLQNDD 134

Figure S2

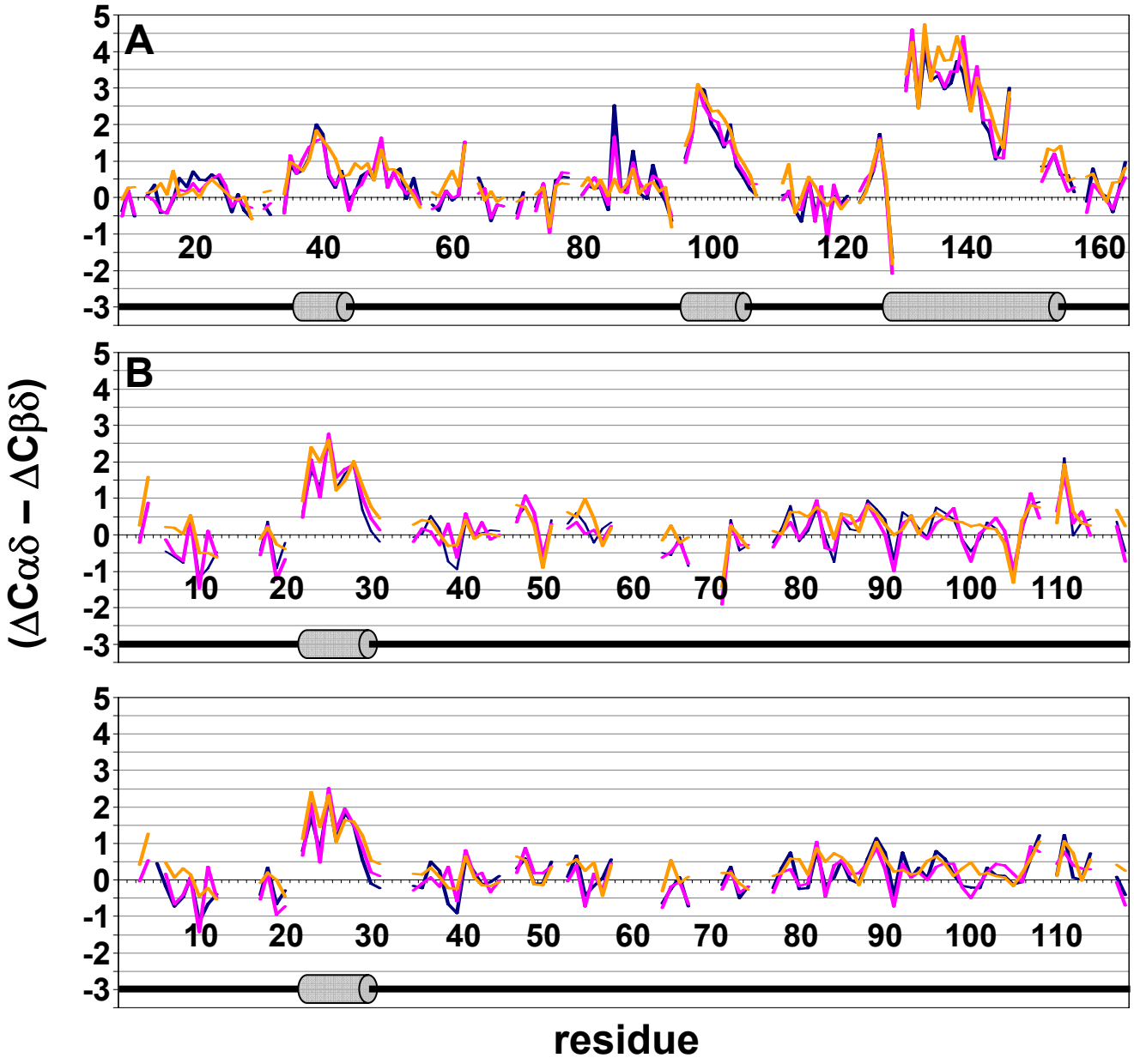
A
B
5
4
3
2
1
0
-1
-2
-3
20
40
60
80
100
120
140
160
10
20
30
40
50
60
70
80
90
100
110
(ΔCαδ – ΔCβδ)
residue

Figure S3

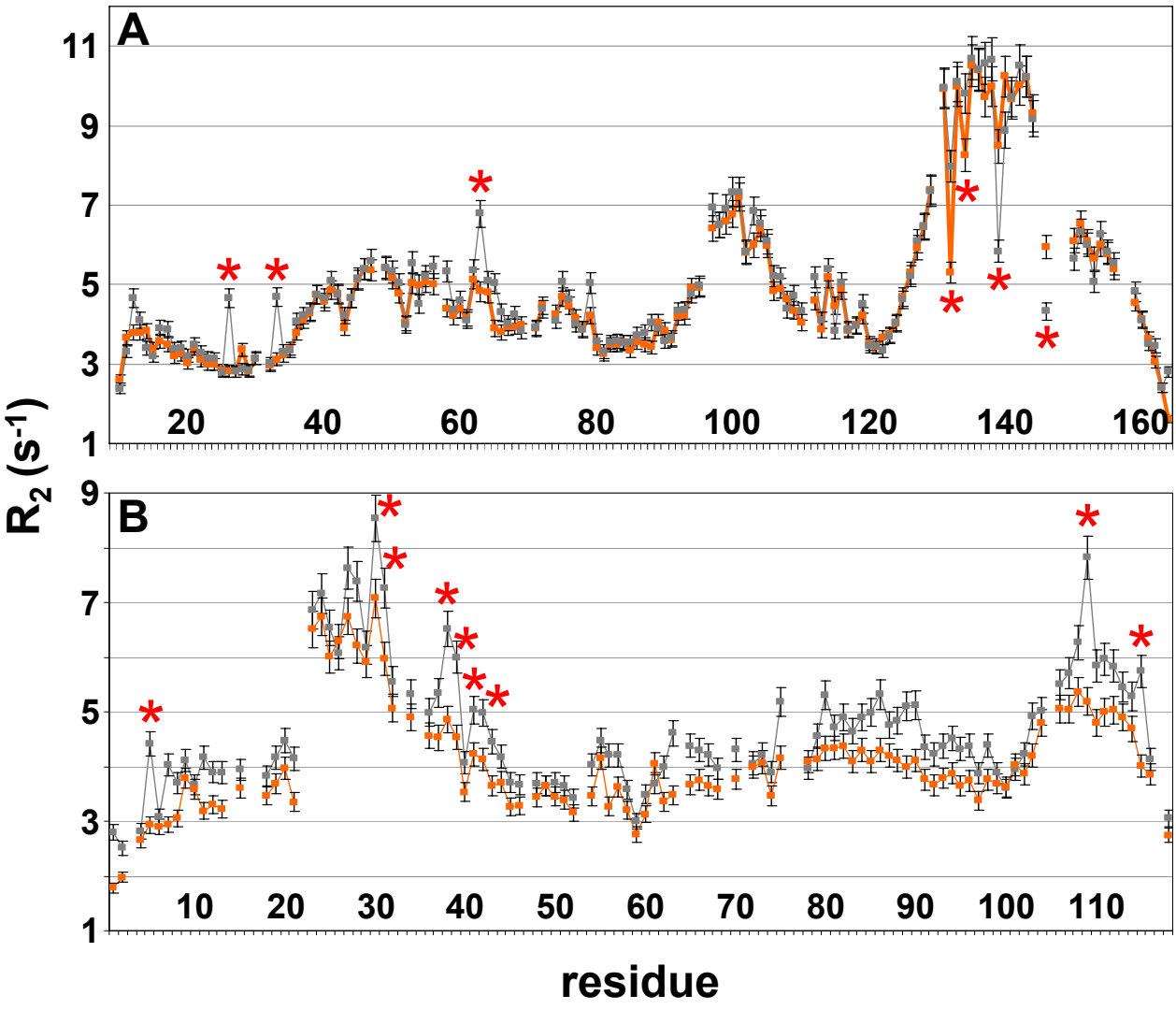

Figure S4

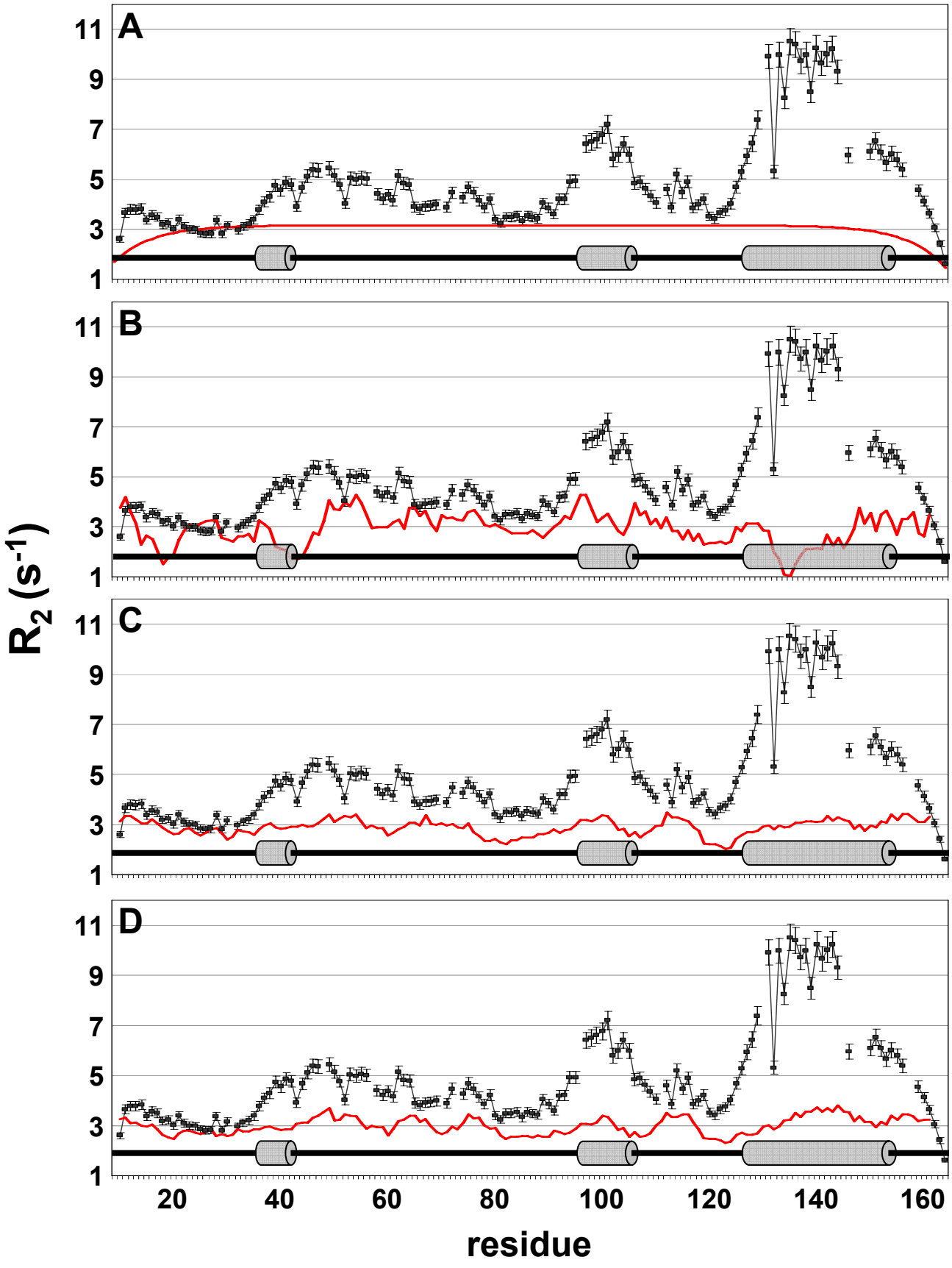

Figure S5

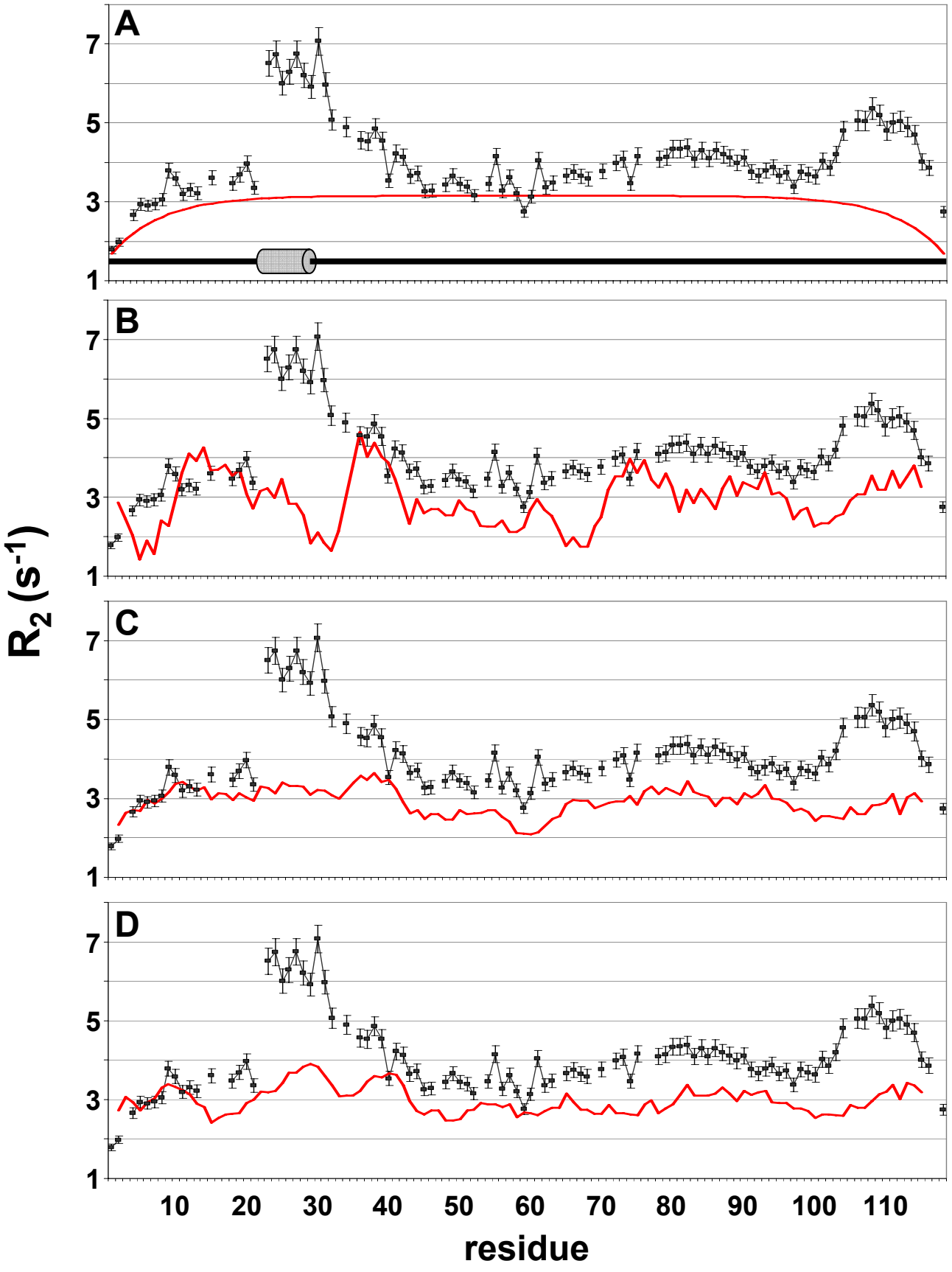

Figure S6

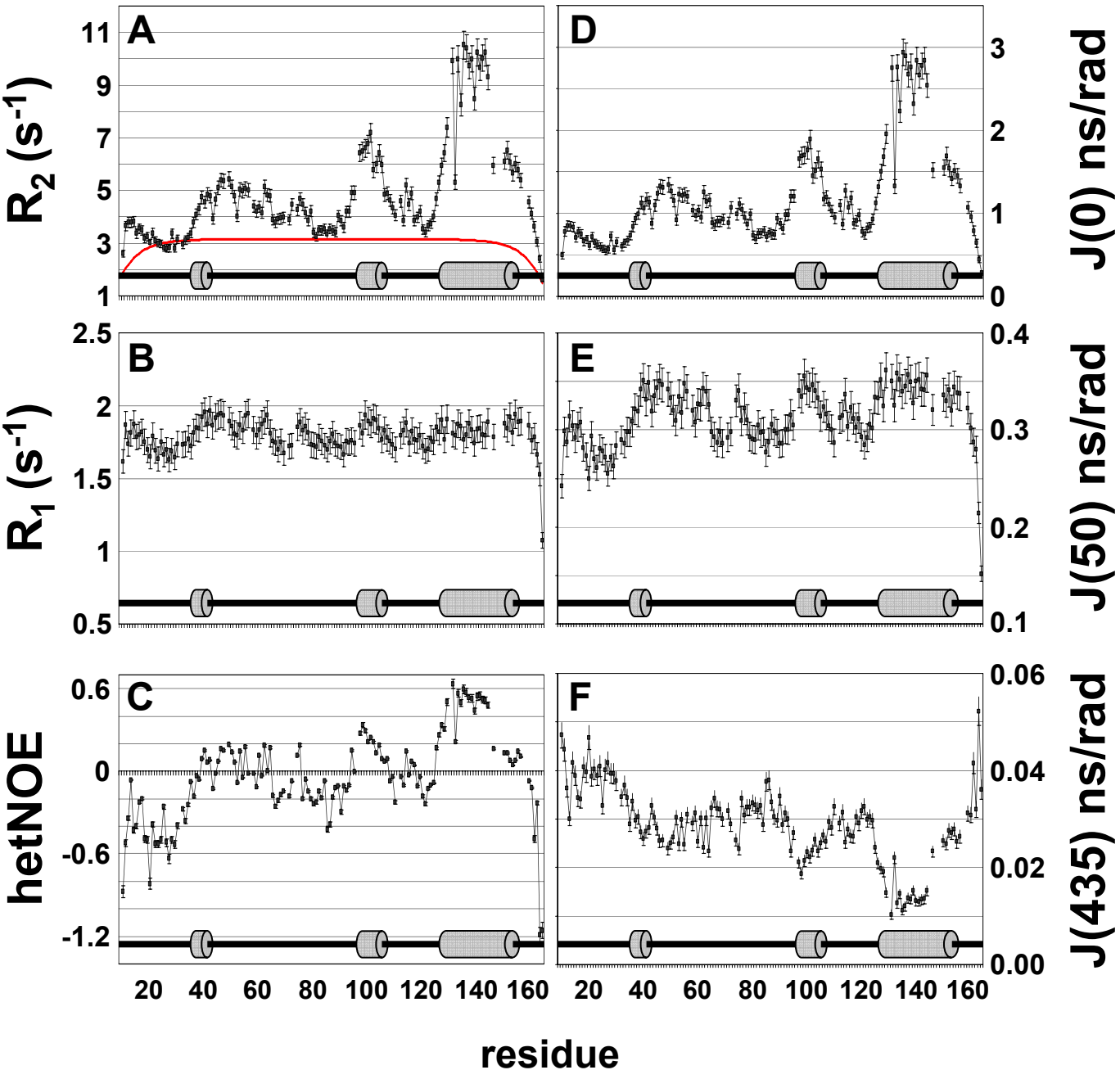
A
B
C
D
E
F
R2 (s-1)
R1 (s-1)
hetNOE
J(0) ns/rad
J(50) ns/rad
J(435) ns/rad
residue
11
9
7
5
3
1
2.5
2
1.5
1
0.5
0.6
0
-0.6
-1.2
3
2
1
0
0.4
0.3
0.2
0.1
0.06
0.04
0.02
0.00
20 40 60 80 100 120 140 160
20 40 60 80 100 120 140 160

Figure S7

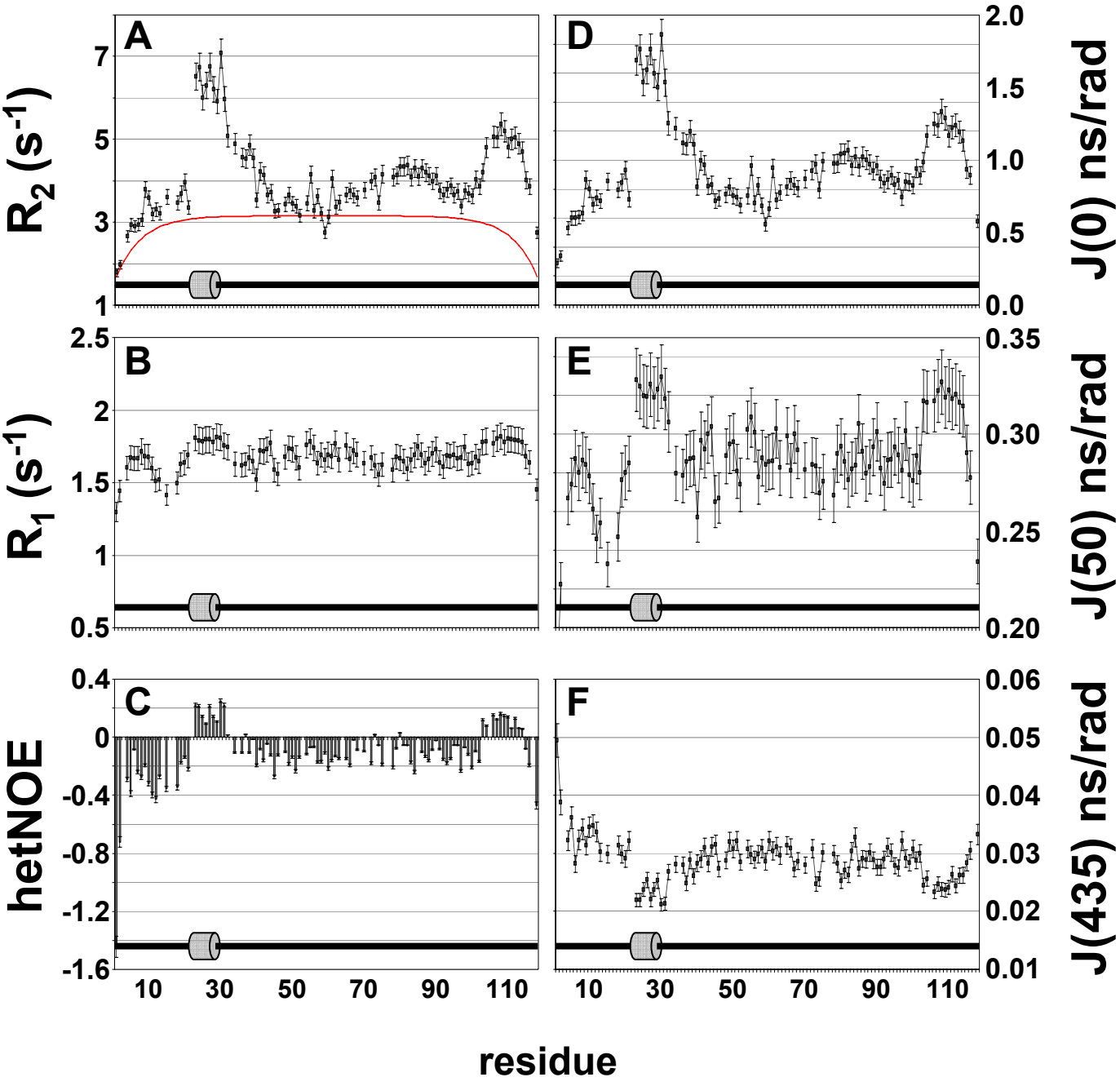
A
B
C
D
E
F
R2 (s-1)
R1 (s-1)
hetNOE
J(0) ns/rad
J(50) ns/rad
J(435) ns/rad
residue
10
30
50
70
90
110

Figure S8

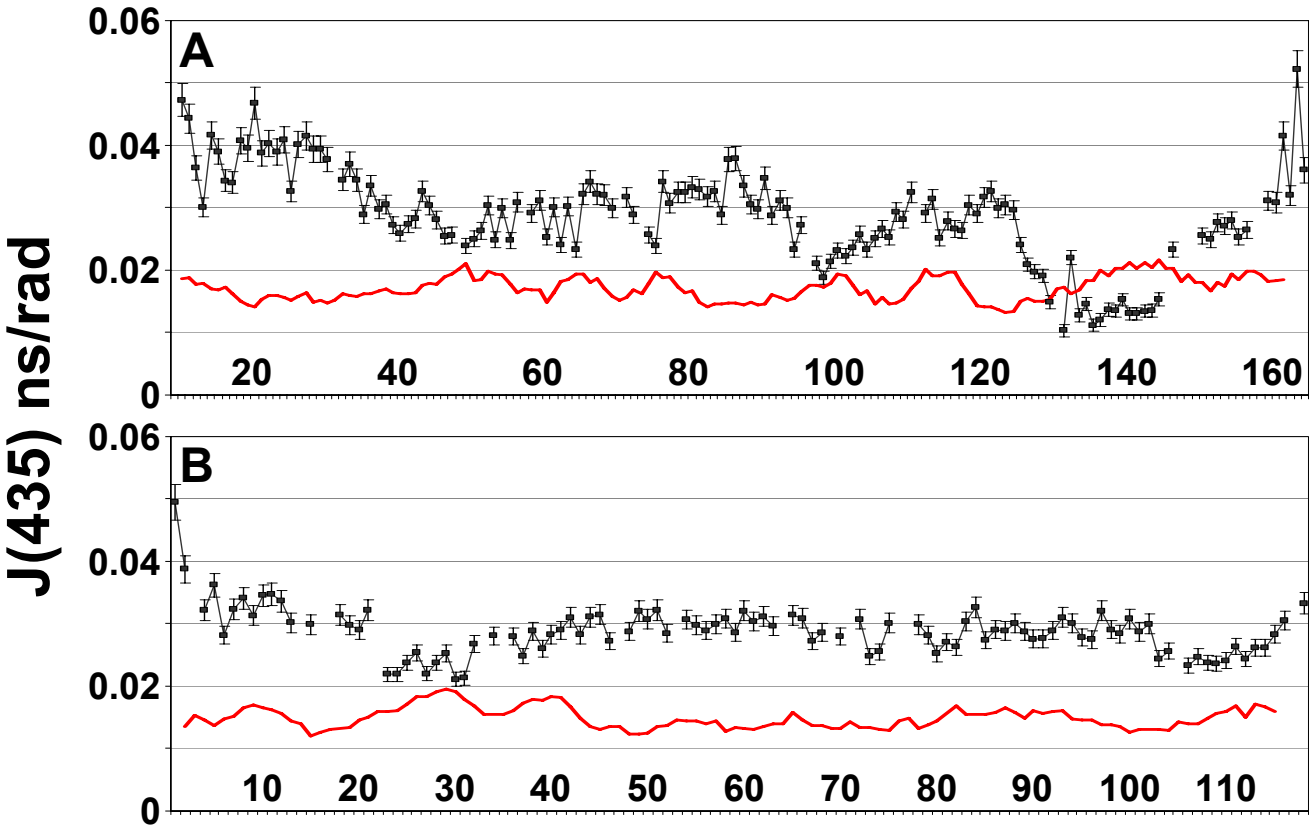
J(435) ns/rad
A
B
0.06
0.04
0.02
0
20
40
60
80
100
120
140
160
10
30
50
70
90
110

Figure S9

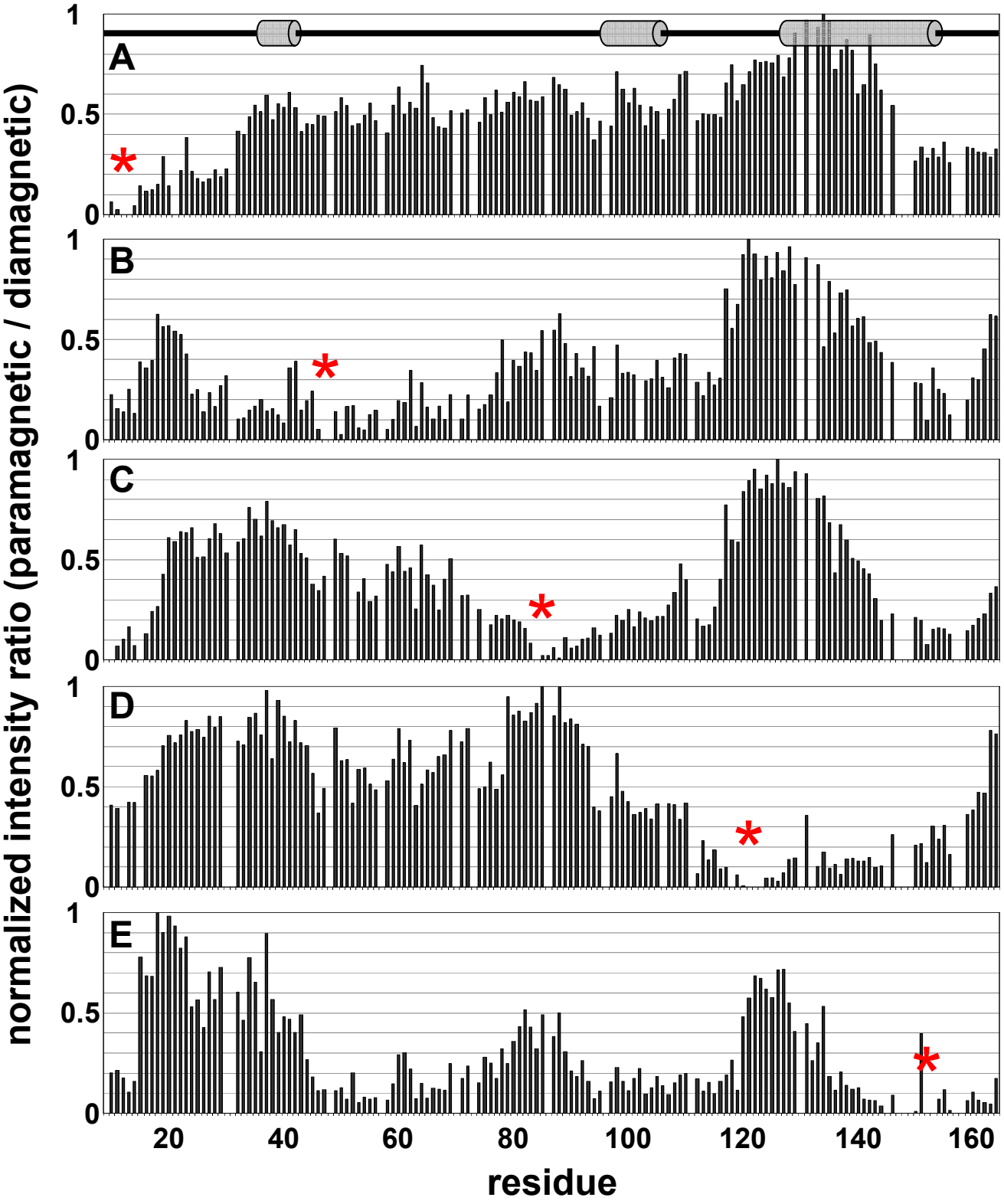
normalized intensity ratio (paramagnetic / diamagnetic)
A
B
C
D
E
1
0.5
0
20
40
60
80
100
120
140
160
residue

Figure S10

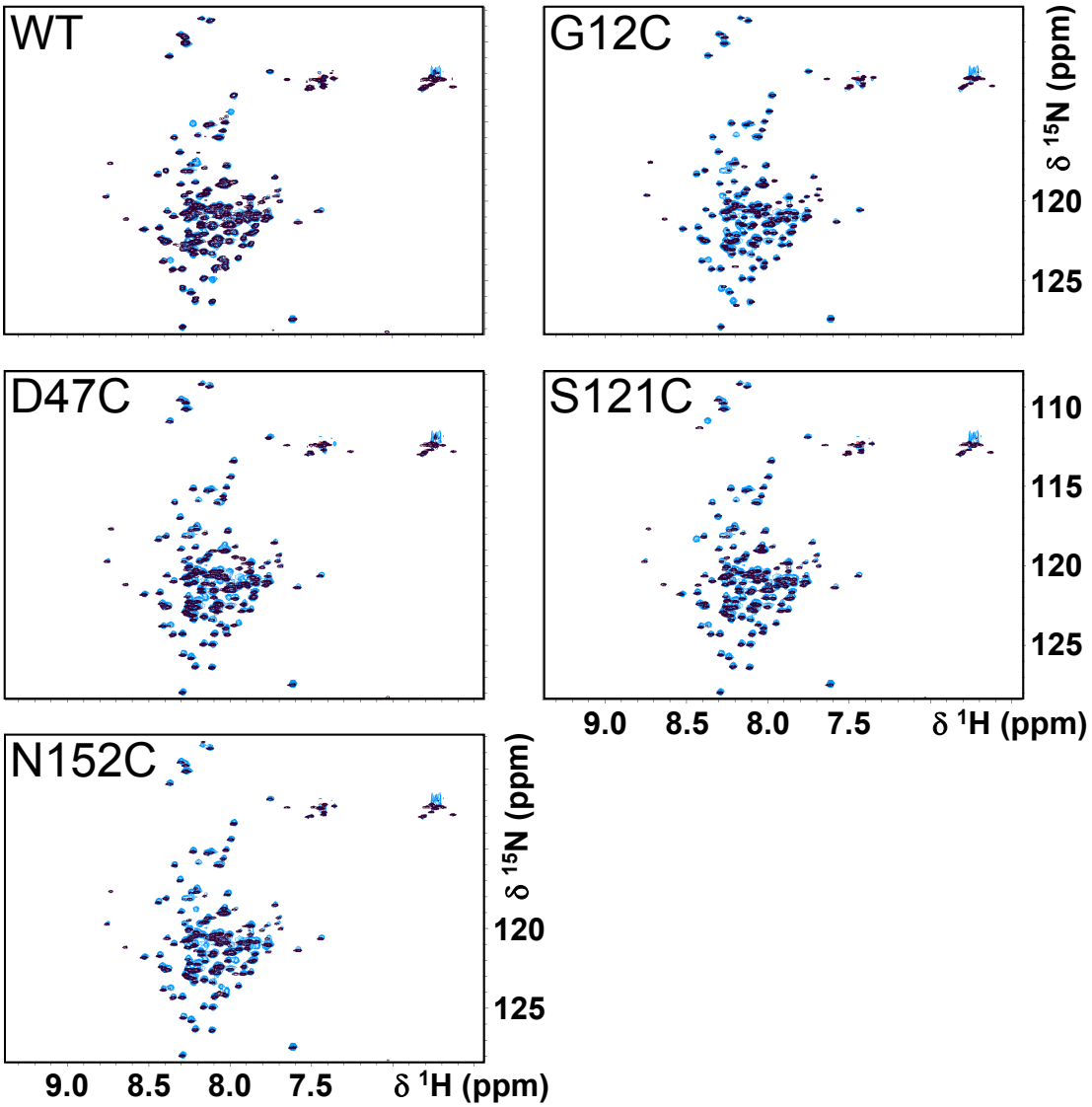
WT
G12C
D47C
S121C
N152C
δ 15N (ppm)
120
125
110
115
120
125
9.0
8.5
8.0
7.5
δ 1H (ppm)

Figure S11

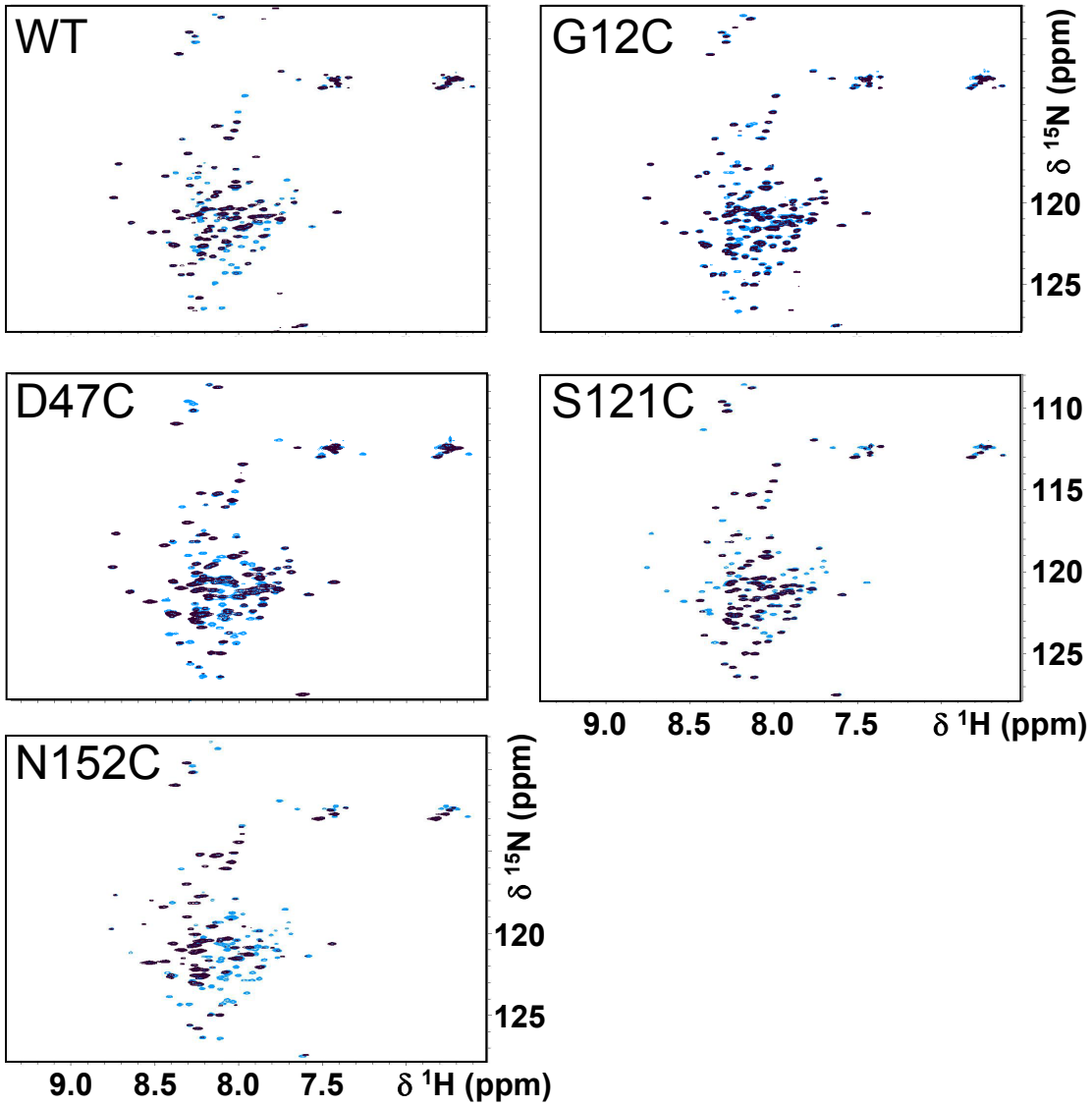
WT
G12C
D47C
S121C
N152C
δ 15N (ppm)
110
115
120
125
9.0
8.5
8.0
7.5
δ 1H (ppm)

Figure S12

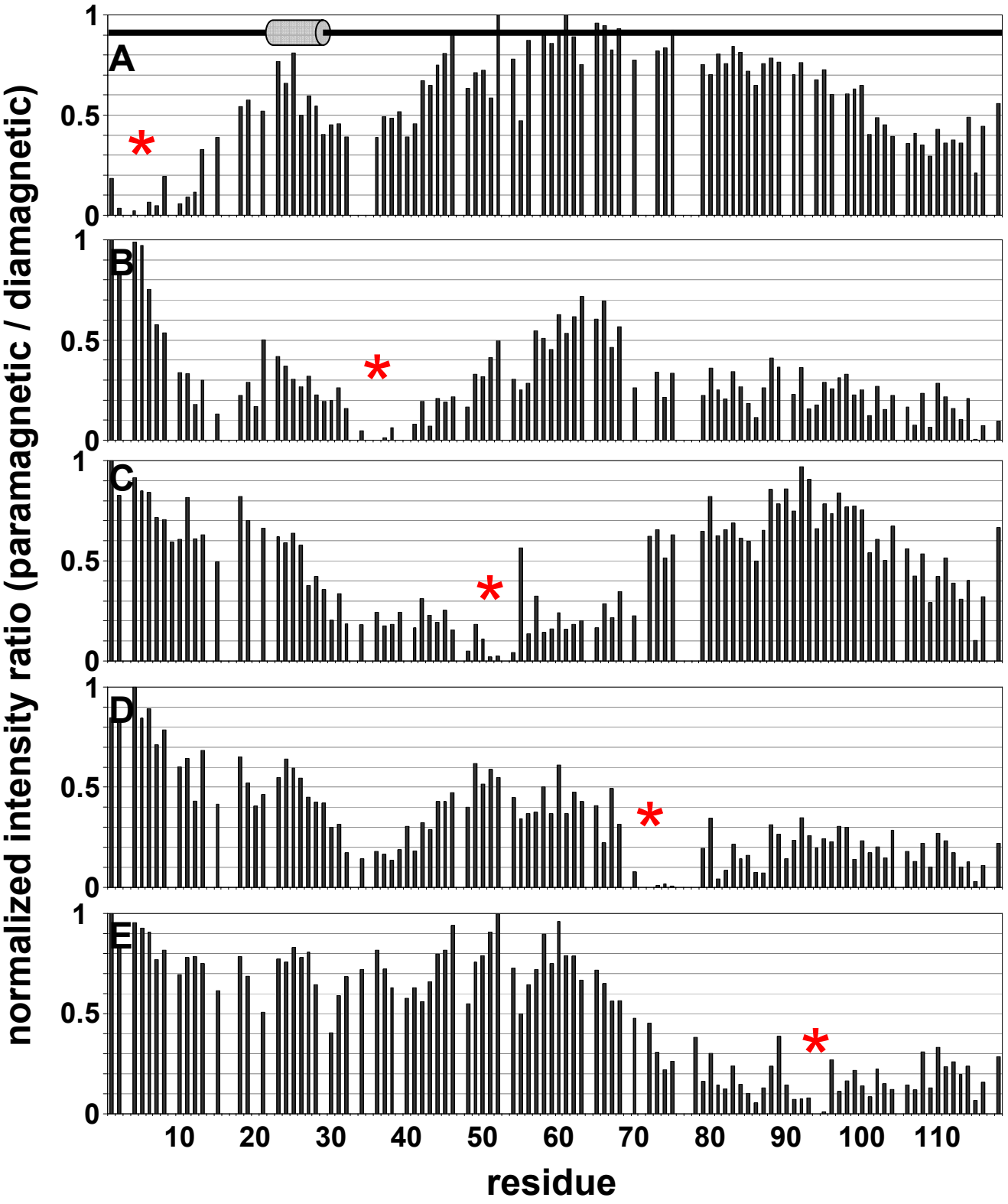
normalized intensity ratio (paramagnetic / diamagnetic)
A
B
C
D
E
1
0.5
0
10
20
30
40
50
60
70
80
90
100
110
residue

Figure S13

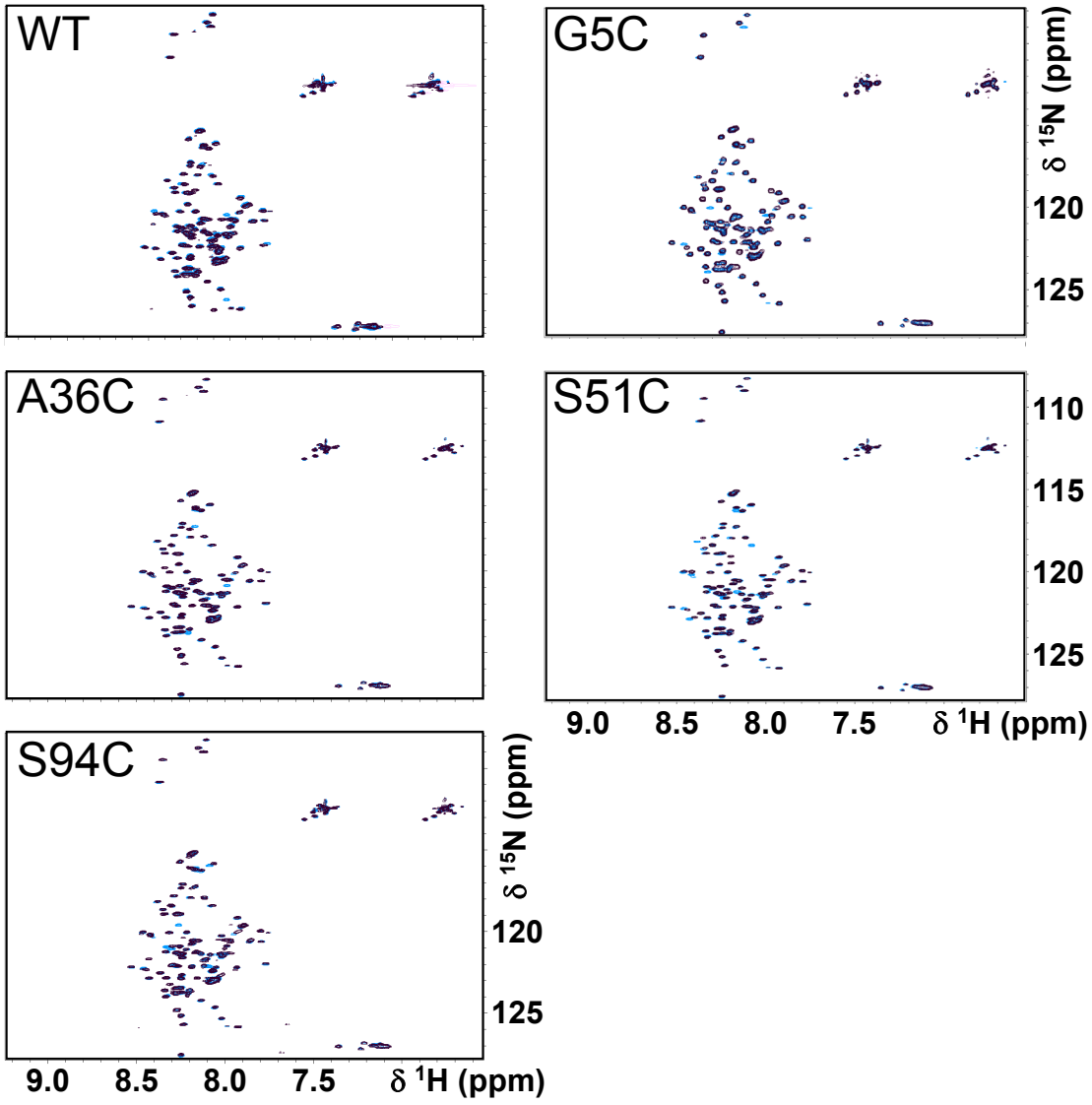
WT
G5C
A36C
S51C
S94C
δ ¹⁵N (ppm)
120
125
110
115
120
125
9.0
8.5
8.0
7.5
δ ¹H (ppm)

Figure S14

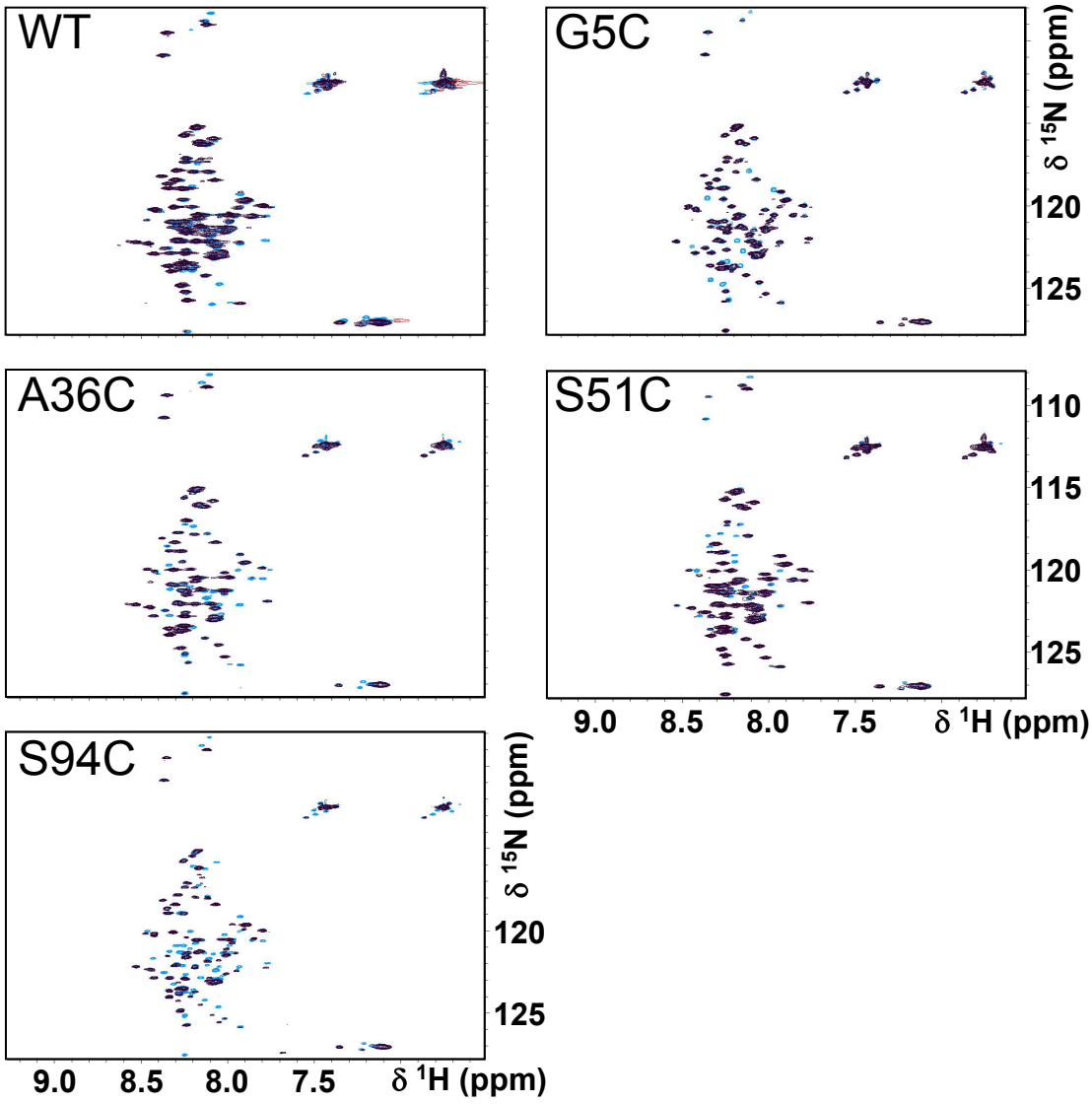
WT
G5C
A36C
S51C
S94C
δ ¹⁵N (ppm)
120
125
110
115
120
125
9.0
8.5
8.0
7.5
δ ¹H (ppm)

Figure S15

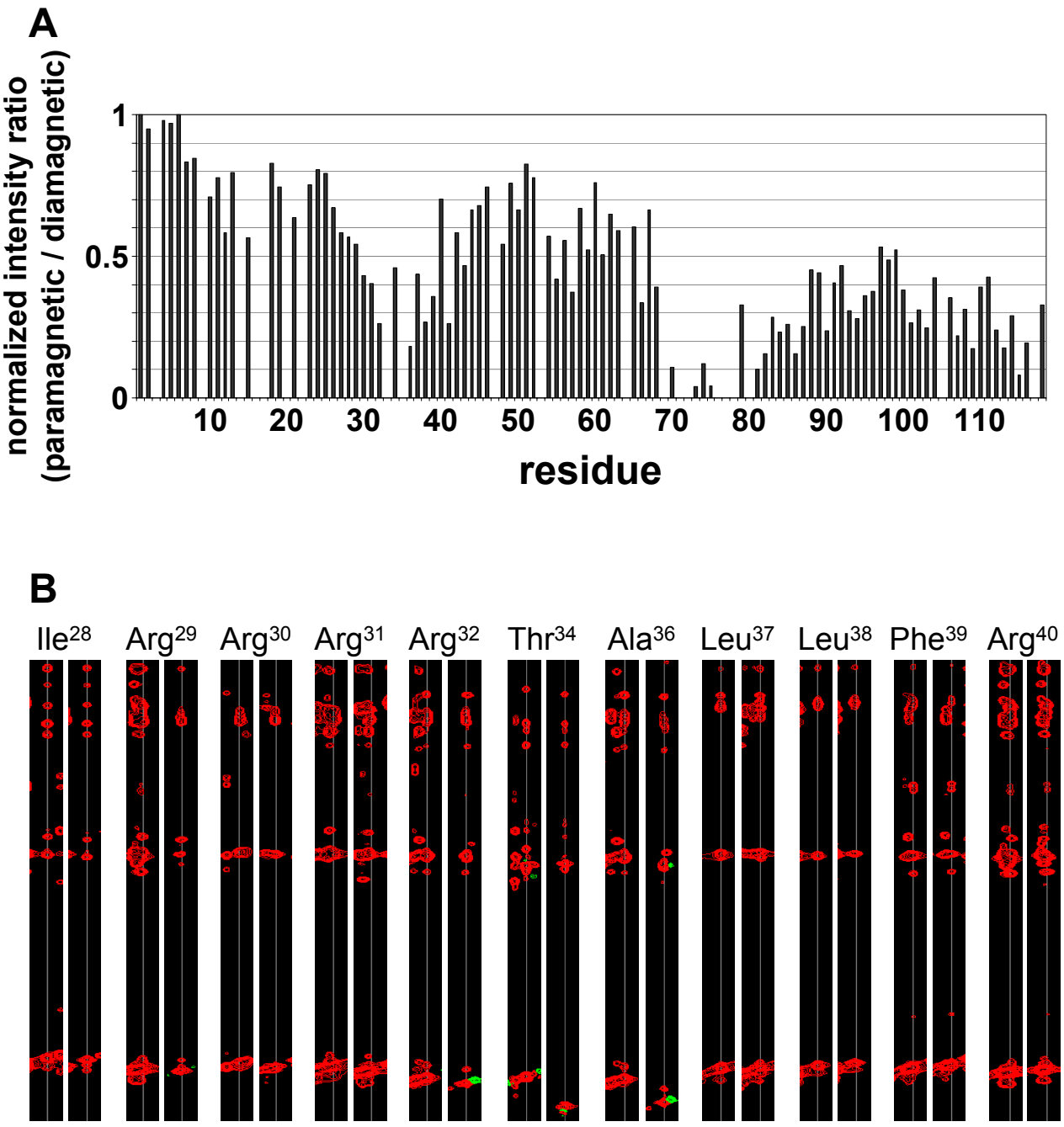
A
normalized intensity ratio
(paramagnetic / diamagnetic)
1
0.5
0
10
20
30
40
50
60
70
80
90
100
110
residue
B
Ile28
Arg29
Arg30
Arg31
Arg32
Thr34
Ala36
Leu37
Leu38
Phe39
Arg40

Figure S16

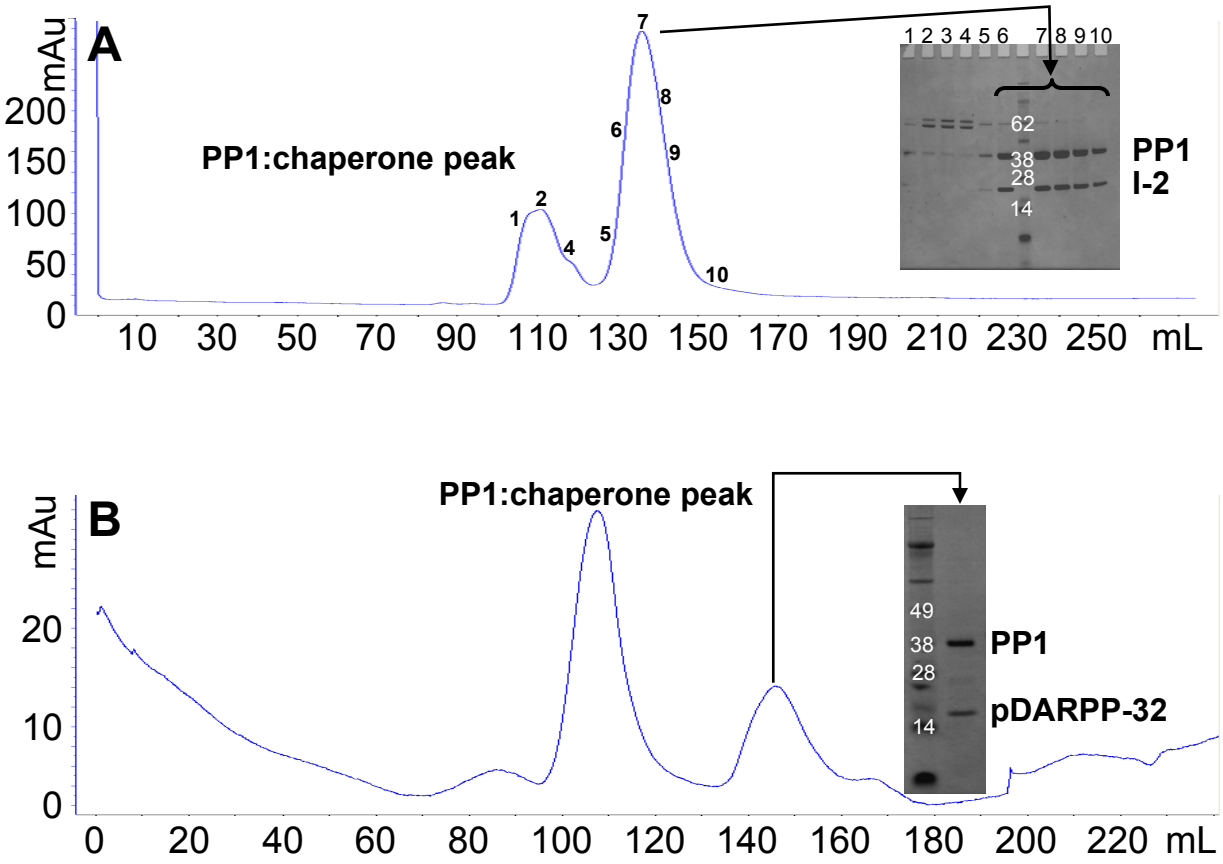
A
mAu
200
150
100
50
0
PP1:chaperone peak
1
2
4
5
6
7
8
9
10
1 2 3 4 5 6 7 8 9 10
62
38
28
14
PP1
I-2
10 30 50 70 90 110 130 150 170 190 210 230 250 mL
B
mAu
20
10
0
PP1:chaperone peak
49
38
28
14
PP1
pDARPP-32
0 20 40 60 80 100 120 140 160 180 200 220 mL

Figure S17

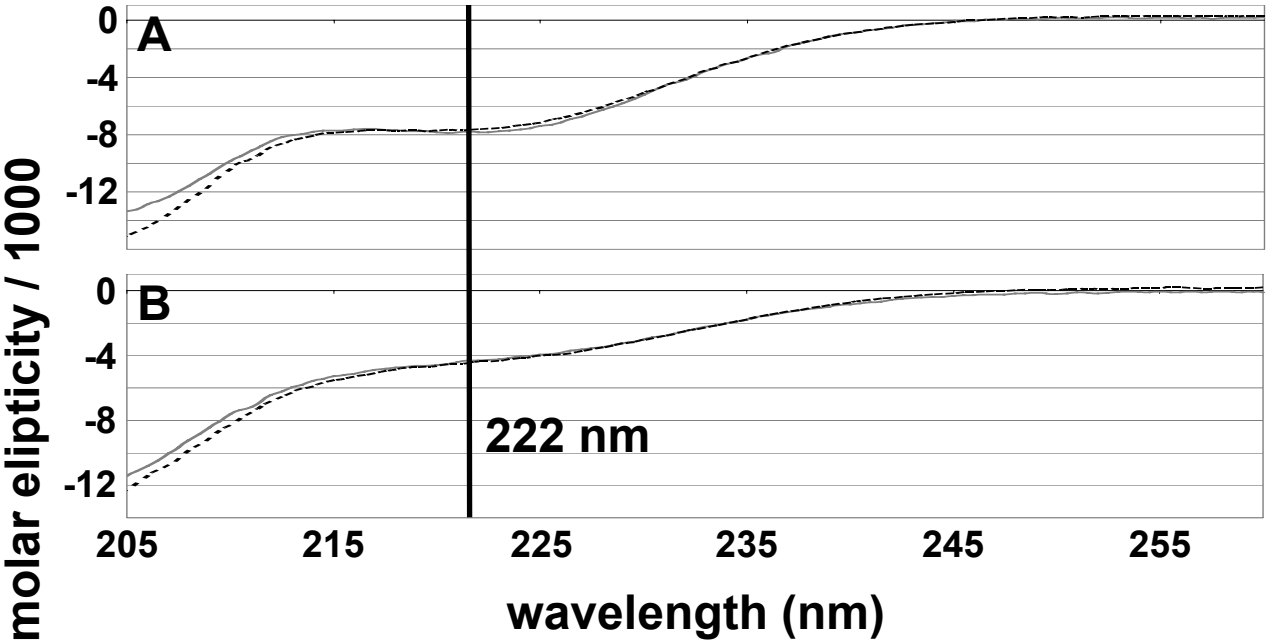
A
B
0
-4
-8
-12
222 nm
205
215
225
235
245
255
molar elipticity / 1000
wavelength (nm)

Figure S18

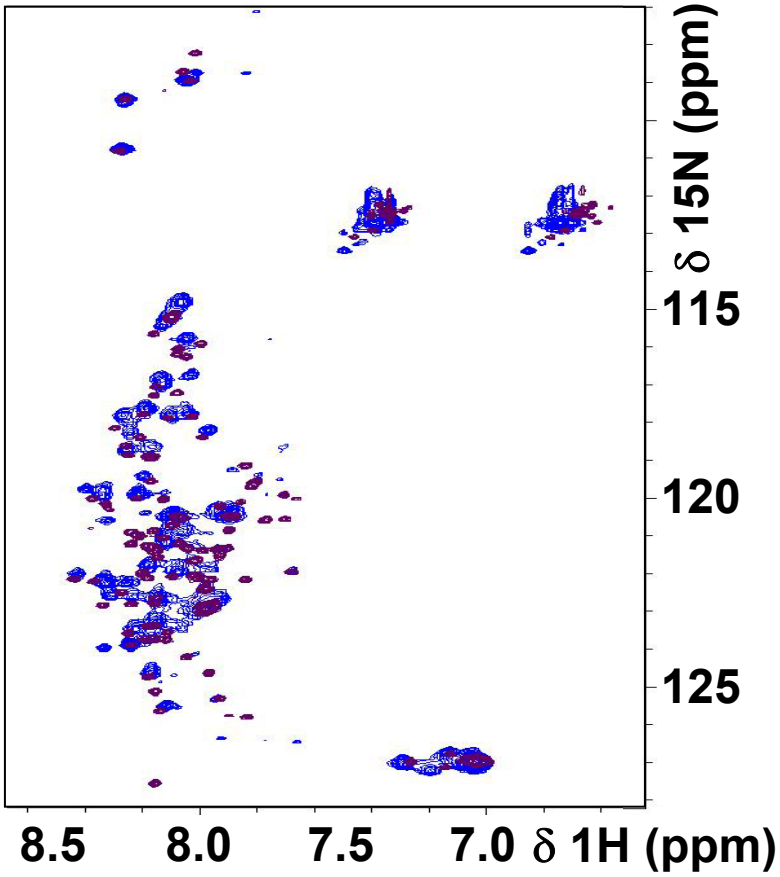

Figure S19