

Supplementary Information

Contribution of Long-Range Interactions to the Secondary Structure of an Unfolded Globin

Daria V. Fedyukina, Senapathy Rajagopalan[†], Ashok Sekhar, Eric C.
Fulmer[‡], Ye-Jin Eun[§] and Silvia Cavagnero*

*Department of Chemistry, University of Wisconsin - Madison,
Madison, Wisconsin 53706*

* To whom correspondence should be addressed

Email: cavagnero@chem.wisc.edu

Current addresses:

† The Methodist Hospital Research Institute, Diabetes Research, 6565 Fannin St,
F8-060, Houston, TX, 77030.

‡ Faculty of Earth and Life Sciences, VU University Amsterdam, De Boelelaan
1085, 1081 HV Amsterdam, the Netherlands

§ Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock
Dr., Madison, WI 53706

Supplementary Discussion

Far-UV circular dichroism. The far-UV circular dichroism (CD) of apoMb, fragments and circular permutants^{1,2} has been studied before at pH ca. 6.0. On the other hand, no prior studies on acid-unfolded apoMb N-terminal fragments were performed before.

Our far-UV CD data for pH unfolded (1-77), (1-119) and (1-153)apoMb are consistent with the NMR results. The CD spectra (Fig. S1) show that the helical content of (1-77)apoMb is nearly identical to that of (1-119)apoMb, on a per-residue basis. The random coil content, however, is larger for (1-119)apoMb due to the introduction of the highly unstructured FG region. Overall helicity increases and random coil content decreases, as chain length progresses from 1-119 to 1-153, upon addition of the residues encompassing the native H helix. Based on the NMR results, the increase in helicity upon chain elongation from 1-119 to 1-153 is likely attributable to the additional induced helicity in other regions of the protein (A, B and C regions) due to long-range interactions.

NMR data analysis. Figures S2-S4 show the ¹H-¹⁵N-HSQC spectra of (1-77), (1-119) and (1-153)apoMb with annotated N and H^N resonance assignments. Chemical shift values for the C^α, H^N, N, C' and selected C^β assigned resonances were deposited in BioMagResBank (see Supplementary Methods).

Secondary chemical shifts for the H^N and N backbone resonances of (1-77), (1-119) and (1-153)apoMb are provided in Figure S5. While C^α and C' secondary chemical shifts are significantly diagnostic for the presence of secondary structure¹⁻⁴, H^N and N shifts are known to be influenced by other structural and environmental factors^{3,4} and are therefore not good reporters for the presence of residual secondary structure in (1-77), (1-119) and (1-153)apoMb.

However, H^N and N secondary chemical shifts can still be regarded as qualitative reporters of variations in the surrounding electronic environment.

The secondary chemical shifts for the H^N and N backbone resonances (Fig. S5) for (1-77) and (1-119)apoMb are remarkably similar. While this result does not shed specific light on the chain length dependence of secondary structure, it supports the fact that the ensemble is populated with similar three dimensional conformations in both (1-77) and (1-119)apoMb, given that the electronic environment is similar. This suggests that the previously reported tertiary contacts involving the A and the G helix^{5,6} may not be as extensive.

The equilibrium unfolded ensemble ΔH^N SCS and ΔN SCS values (Fig. S5 and Fig. S6) confirm this result. Comparisons between each of the N-terminal fragments and the full-length protein show that introduction of the amino acids corresponding to the H helix into the polypeptide chain causes small perturbation in the chemical shift of the second portion of the B region. While it is hard to interpret the nature of the structural variations resulting from this effect, due to the heterogeneity of the unfolded ensemble, it is tempting to speculate that there may be significant B-H contacts underlying the increased helicity found in the A-B-C region as a result of introducing the H helix into the polypeptide chain.

Determination of percent helicity enhancements in the N-terminal region. The estimated helical populations for the A, B and C regions of the 1-119 fragment of apoMb and the full-length protein are shown in Figure 2c. Percent helicities were evaluated from the assigned C^α and C' secondary chemical shifts and the expected shifts for a 100% helix, as described in the figure legend. The data of Figure 2c explicitly illustrate the fact that the presence of the C-terminal residues corresponding to the native H helix enhance the apparent helicity of the ABC cluster by ca. 7%, i.e., from 5 to 12% in the absence and presence of the H region, respectively.

Additional comments on ΔC^α SCS values. It is interesting to notice that the ΔC^α and $\Delta C'$ SCS values smaller than the experimental error (see Fig. 2b) do not randomly fluctuate around zero but lay mostly on one side. We want to emphasize that this trend is clearly hard to interpret on solid ground, given that these ΔC^α and $\Delta C'$ SCS values are smaller than the reported uncertainties. On the other hand, we speculate that such monotonic trends may still be of some experimental significance, due to their nonrandom nature, and may possibly be due to extremely small second order effects, leading to tiny variations in secondary structure induced by long-range interactions. For instance, based on the data in Figure 2b, addition of the residues belonging to the G region may lead to extremely small decreases in helicity throughout the rest of the sequence (residues 1 to 77). Conversely, addition of the residues belonging to the native H helix may lead to extremely small increases in helicity throughout the rest of the sequence (residues 1 to 119).

Percent helicity enhancements in the N-terminal region. Estimated helical populations for the A, B and C regions of the 1-119 fragment of apoMb and the full-length protein are shown in Figure S7. Percent helicities were evaluated from the assigned C^α and C' secondary chemical shifts and the expected shifts for a 100% helix, as described in the figure legend. The data of Figure S7 explicitly illustrate the fact that the presence of the C-terminal residues corresponding to the native H helix enhance the apparent helicity of the ABC cluster by ca. 7%, i.e., from 5 to 12% in the absence and presence of the H region, respectively.

Linewidth analysis. We performed full width at half height (FWHH) analysis of the backbone ^{15}N nuclei of all three unfolded apoMb constructs analyzed in this work. The results are shown in Figure S8.

The sequence-specific FWHH trends observed for the full-length protein, (1-153)apoMb, are qualitatively consistent with the R2 profile reported by Yao *et al.* in reference 6 of the accompanying communication.

Interestingly, Figure S8 shows that the H-region-dependent enhancement in the secondary structure of the A-B-C region correlates with a selective line-broadening beyond detection in the A portion of the sequence. This broadening is primarily observed for the full-length (1-153)apoMb, bearing the residues corresponding to the native H helix and, most importantly, it is absent in (1-77) and (1-119)apoMb. Hence, the above tertiary structure-driven increase in secondary structure appears to be linked with the establishment of specific slow dynamic processes on the μ s-ms timescale.

No other major line-broadening effects are observed in other regions, as chain elongates.

As shown in Figure S8, the full length construct does not display any line-broadening beyond detection in the H region, despite the presence of the A-to-H contacts in some of the unfolded population and despite the observed extensive line-broadening in the A region.

The H region resonances 143 and 144, on the other hand, are fairly broad. The above observations suggest that line-broadening effects resulting from long-range contacts may, in general, not necessarily be symmetrical, i.e., they may not need to be present to the same degree for all the involved interaction counterparts.

While a more complete relaxation analysis is desirable, to more specifically identify slow-timescale R_{ex} contributions reporting on slow time-scale motions (μ s-ms) and reduced spectral densities, the FWHH data presented here are sufficient to highlight a correlation between slow motions and long-range-contact-driven secondary structure formation in the unfolded state.

Secondary structure analysis of acid-unfolded apoMb by amino acid type. The C^α secondary chemical shifts of pH-unfolded full-length apoMb were collected by amino acid type, as shown in Figure S9. The plot shows that, for each occurrences of the same amino acid in the sequence, the secondary chemical shift, and thus also the secondary structure, assume very different values. This result clearly indicates that the backbone secondary structure of each unfolded residue is not determined by amino acid type, and it is influenced by the presence of other residues in the sequence.

Supplementary Methods

Expression and purification of N-terminal apoMb fragments. Uniformly ^{13}C , ^{15}N -labeled N-terminal apoMb fragments were overexpressed in *E.coli* using Tuner DE3 pLacI (in the case of (1-77)apoMb) and BL21 DE3 cells (in the case of (1-119)apoMb and (1-153)apoMb) (Novagen, San Diego, CA) in M9 minimal medium containing ^{15}N - NH_4Cl (1.5g/L) and ^{13}C -glucose (2g/L). Cell growth and purification were carried out as described⁷.

Circular dichroism. Far-UV CD spectra were recorded at room temperature on an MOS-450 spectropolarimeter (Bio-Logic Science Instruments, Claix, France) using a 1 mm path length cuvette. All samples were in water adjusted to pH 2.4 with HCl. Sample concentrations ranged from 6 to 18 μM . The spectra displayed in Figure S1 are the average of two independent experiments. Two scans were averaged for each experiment. Each scan was performed with a 2 nm band-width, 1 nm step-size and 10 s averaging time per point.

NMR sample preparation. ^{13}C , ^{15}N -labeled lyophilized polypeptides (i.e., (1-77), (1-119), or (1-153)apoMb) were dissolved in a solution containing 5 mM CD_3COOH at pH 2.40. The pH was adjusted to 2.40 with HCl. The samples were then eluted through a 3-mL Sephadex G-25 (fine grade, Amersham Biosciences) spin column pre-equilibrated with 5 mM CD_3COOH at pH 2.40. After the spin column treatment, D_2O was added to each sample (final concentration: 5% v/v). The apparent pH was then readjusted to 2.40 with HCl. The final concentrations of (1-77), (1-119), and (1-153)apoMb were 313, 280, and 304 μM , respectively. Solutions containing 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) for external chemical shift referencing were prepared under identical solution conditions.

NMR data collection and processing. *General considerations and NMR data collection.* All the NMR experiments were performed on a Varian INOVA 600 MHz NMR spectrometer

equipped with a Varian $^1\text{H}\{^{13}\text{C}, ^{15}\text{N}\}$ triple resonance probe with triple axis gradients. Acquisition parameters for the 2D high resolution ^1H - ^{15}N - HSQC⁸ experiments and the triple-resonance experiments used for resonance assignments (HNCA⁹⁻¹², HN(CO)CA¹³, HNCACB^{11, 12, 14}, and HNCO⁹⁻¹²) are provided in Table 1. Relaxation delays were set to 1 s for all experiments. Chemical shift referencing in the ^1H dimension was done with by a 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) external standard. The ^{15}N and ^{13}C dimensions were referenced indirectly as described¹⁵ and gyromagnetic ratios of $^{13}\text{C}/^1\text{H} = 0.251449530$; $^{15}\text{N}/^1\text{H} = 0.101329118$ were used¹⁶. In all experiments, the ^1H carrier frequency was placed in correspondence of the HDO frequency. The spectrometer temperature was calibrated with neat methanol¹⁷.

Resonance assignments were carried out in close succession within 1.5 months for all three species in a single NMR session, at 25.0°C.

NMR data processing. The NMRPipe¹⁸ and NMRDraw¹⁸ software packages were used for NMR data processing. Linear prediction was performed once (36 points) on the ^{15}N dimension, in all the triple-resonance experiments. Time domain data were apodized with a 90°-shifted sine-bell squared window function in both the ^{13}C and ^{15}N dimensions, in all 3D experiments. Free induction decays were apodized with an unshifted Gaussian window function in the ^1H dimension of all 3D experiments and in both dimensions of the 2D experiments. A 90°-shifted sine-bell square window function was applied to ^{13}C and ^{15}N dimensions in all 3D experiments. Triple- and double-resonance time-domain data were zero-filled once and twice in all dimensions, respectively.

Backbone assignments were carried out by the Sparky-3 software package¹⁹.

The H^N , N, C' and C^α chemical shift assignments for (1-77), (1-119), and (1-153)apoMb, and assignments of selected C^β resonances, were deposited in the BioMagResBank (<http://www.bmrb.wisc.edu>). The BMRB accession numbers are 16499, 16500, and 16501 for (1-77), (1-119), and (1-153)apoMb, respectively.

NMR secondary chemical shift analysis of acid-unfolded apoMb fragments. Reference random coil chemical shift (RCCS) values, including corrections for neighboring residues, were calculated according to Schwarzinger *et al.*^{20,21}, following the relations

$$\delta_{\text{random coil-corrected}} = \delta_{\text{random coil}} + A + B + C + D ;$$

where A, B, C and D are corrections for amino acids $i-2$, $i-1$, $i+1$ and $i+2$, respectively.

Secondary Chemical Shifts (SCS) were calculated from experimental chemical shifts (CS) according to the simple relation SCS (ppm) = *experimental CS* (ppm) – *sequence-corrected RCCS* (ppm).

Error analysis of NMR chemical shifts. For (1-77)apoMb and (1-119)apoMb, the propagated uncertainties for NMR chemical shift differences, (see horizontal gray bars in Fig. 4 and Supplementary Fig. S6) were calculated from two independent resonance assignments, carried out on two separate samples of each species. In each case, NMR data were collected with the same acquisition parameters at the same temperature, at comparable pH and protein concentrations. For each assigned chemical shift of (1-77) and (1-119)apoMb, the standard deviation of the mean was calculated from the above repeats. The standard deviation for (1-153)apoMb was estimated as the average of the (1-77) and (1-119)apoMb standard deviations. The uncertainty in NMR chemical shift differences was then determined according to standard error propagation procedures²² as \pm the square root of the sum of the squares of the standard

deviation for each relevant species. No covariance was applied, as judged to be irrelevant in this case.

Acknowledgments

We are grateful to Dr. Charlie Fry, Dr. Marco Tonelli and Clement Chow for technical assistance and helpful discussions.

Supplementary Table

Table S1. Experimental acquisition parameters for the 2D HSQC and triple resonance experiments used for the resonance assignments of pH-unfolded (1-77) and (1-119) N-terminal apoMb fragments and pH-unfolded full-length apoMb (i.e., (1-53)apoMb). Data were collected on a 600 MHz NMR spectrometer as described in the Supplementary Materials and Methods.

Experiment	Number of complex points (t1 x t2 x t3), t1: ¹³ C; t2: ¹⁵ N; t3: ¹ H.	Sweep Widths (Hz)		
		¹ H	¹³ C	¹⁵ N
<u>(1-77)apoMb</u>				
2D HSQC	512 x 2048	6600	-	1500
3D HNCA	80 x 36 x 2048	6600	3620	1500
3D HN(CO)CA	64 x 36 x 2048	6600	3620	1500
3D HNCACB	64 x 36 x 2048	6600	8450	1500
3D HNCO	80 x 36 x 2048	6600	1000	1500
<u>(1-119)apoMb</u>				
2D HSQC	512 x 2048	6600	-	1500
3D HNCA	80 x 37 x 2048	6600	3620	1500
3D HN(CO)CA	64 x 36 x 2048	6600	3620	1500
3D HNCACB	68 x 36 x 2048	6600	8450	1500
3D HNCO	64 x 36 x 2048	6600	1000	1500
<u>(1-153)apoMb</u>				
2D HSQC	512 x 2048	6600	-	1500
3D HNCA	80 x 38 x 2048	6600	3620	1500
3D HN(CO)CA	64 x 36 x 2048	6600	3620	1500
3D HNCACB	64 x 36 x 2048	6600	8450	1500
3D HNCO	64 x 36 x 2048	6600	1000	1500

Supplementary Figure Legends

Figure S1. Far-UV CD spectra of (1-77), (1-119) and (1-153)apoMb at room temperature and pH 2.5. Data are shown as (a) ellipticity per unit concentration ($[\theta]/\text{Conc.}$) and (b) mean residue ellipticity ($[\theta]_{\text{MRE}}$).

Figure S2. $^1\text{H},^{15}\text{N}$ -HSQC spectrum of (1-77)apoMb at pH 2.4 and 25°C with annotated resonance assignments.

Figure S3. $^1\text{H},^{15}\text{N}$ -HSQC spectrum of (1-119)apoMb at pH 2.4 and 25°C with annotated resonance assignments.

Figure S4. $^1\text{H},^{15}\text{N}$ -HSQC spectrum of (1-153)apoMb at pH 2.4 and 25°C with annotated resonance assignments. Red peaks indicate folded peaks originating from the side chain amides.

Figure S5. (a) H^{N} and (b) N secondary chemical shifts of (1-77), (1-119) and (1-153)apoMb at pH 2.4 and 25°C. The helices of native full-length apoMb are mapped above the graph as black bars.

Figure S6. Differences between the (a) H^{N} and (b) N secondary chemical shifts of the species analyzed in this work, i.e., (1-119)apoMb - (1-77)apoMb, (1-153)apoMb - (1-77)apoMb, and (1-153)apoMb - (1-119)apoMb at pH 2.4 and 25°C. The native apoMb helices are mapped above

the graph as black bars. The horizontal gray bars denote experimental uncertainties (see Supplemental Materials and Methods).

Figure S7. Estimated helical populations in the N-terminal region of acid-unfolded (a) (1-77)apoMb, (b) (1-119)apoMb and (c) (1-153)apoMb expressed as percent helicities for the A, B and C regions of the sequence. Percent helicities were estimated from the assigned secondary chemical shifts, averaged over the pertinent portion of the sequence, divided by the experimental chemical shift expected for a 100% helix. The secondary chemical shifts corresponding to a 100% helix were taken to be 2.8 and 2.1 ppm for C^α and C', respectively²³. Estimated percent helicities are shown for data derived from C^α (light blue) and C' (blue) nuclei. Averages over the values derived from C^α and C' secondary chemical shifts are shown in black. The percent helicity of the ABC cluster was calculated by averaging over the secondary chemical shifts of all the cluster's residues, defined as amino acids 4-12 (A region), 20-35 (B region) and 46-52 (C region).

Figure S8. Linewidth analysis, reported as full width at half height (FWHH), of the ¹⁵N resonances of (a) (1-77)apoMb, (b) (1-119)apoMb and (c) (1-153)apoMb at pH 2.4 and 25°C. Dashed lines correspond to resonances broadened beyond detection. The missing resonances correspond to either prolines or peaks whose FWHH could not be reliably assessed due to spectral overlaps.

Figure S9. C^α secondary chemical shifts for full length apoMb at pH 2.4 and 25°C grouped according to amino acid type.

Supplementary Figures

Figure S1

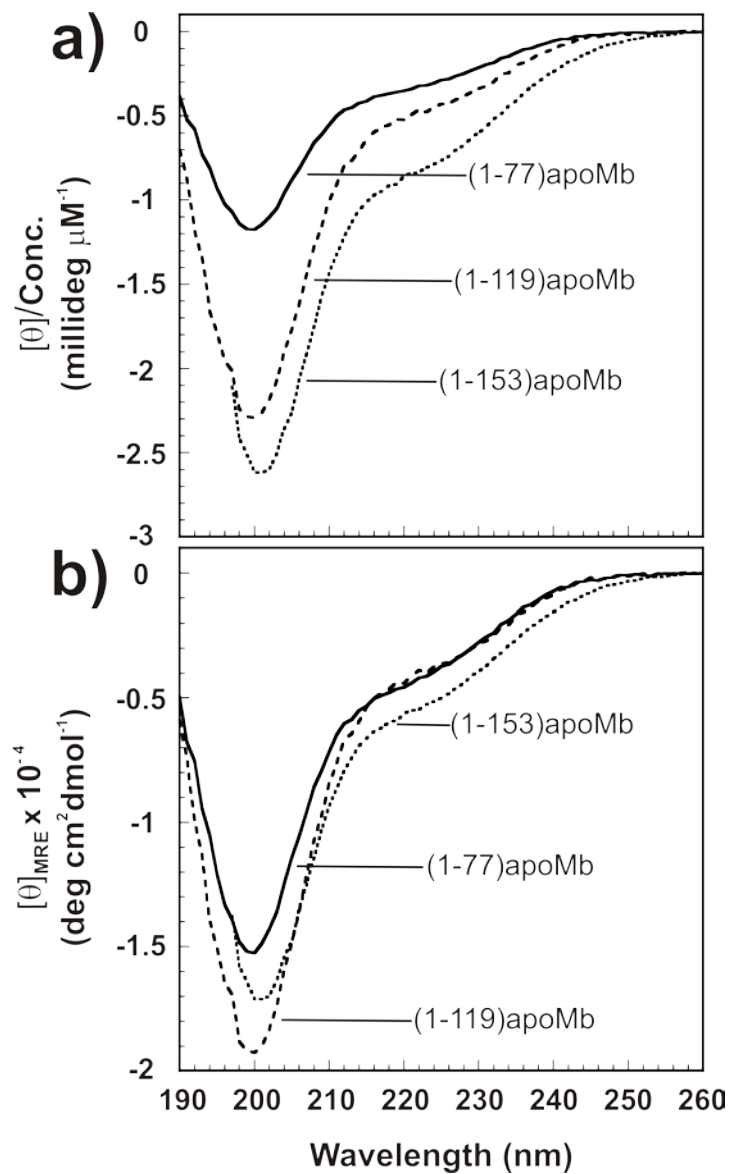


Figure S2

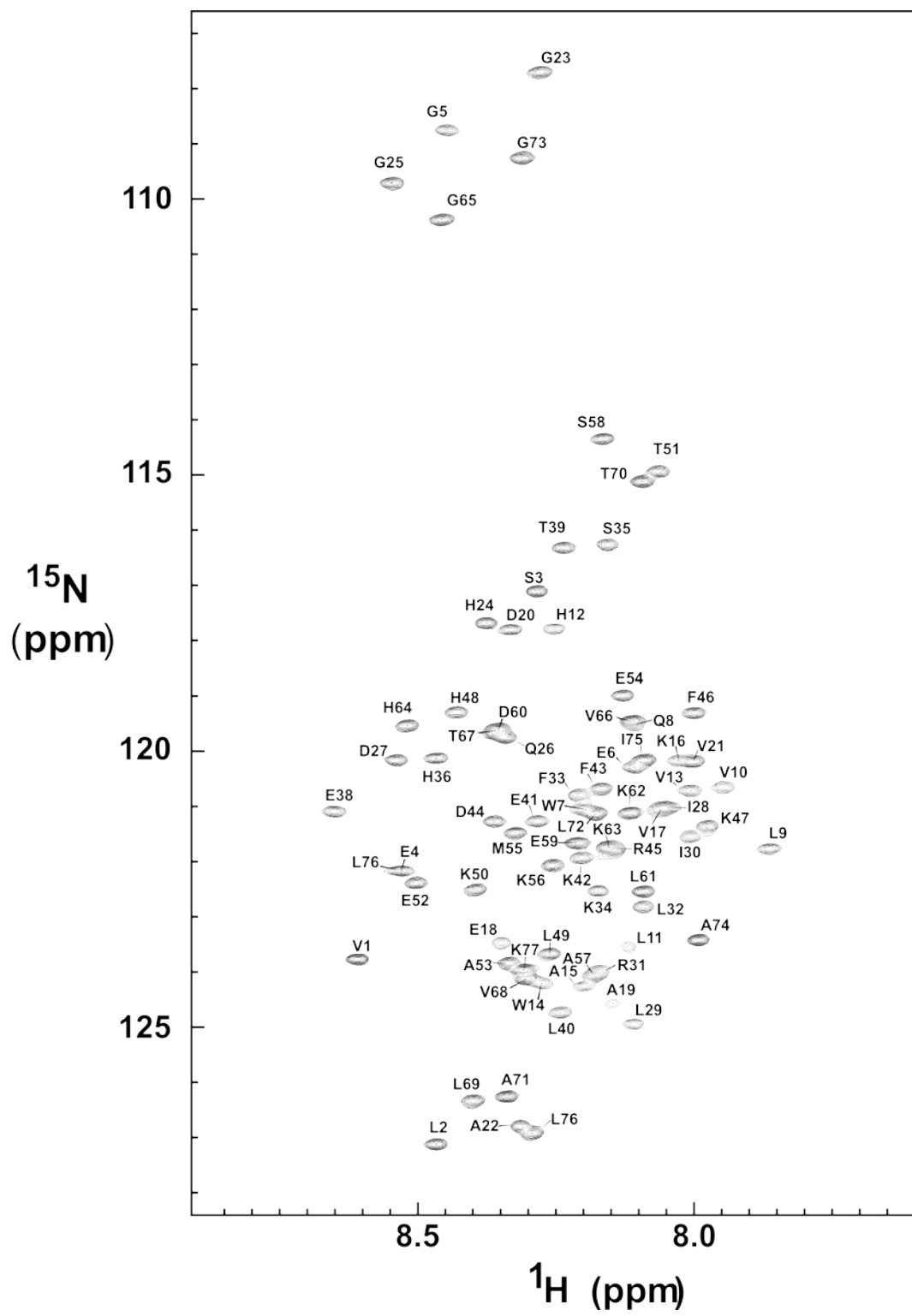


Figure S3

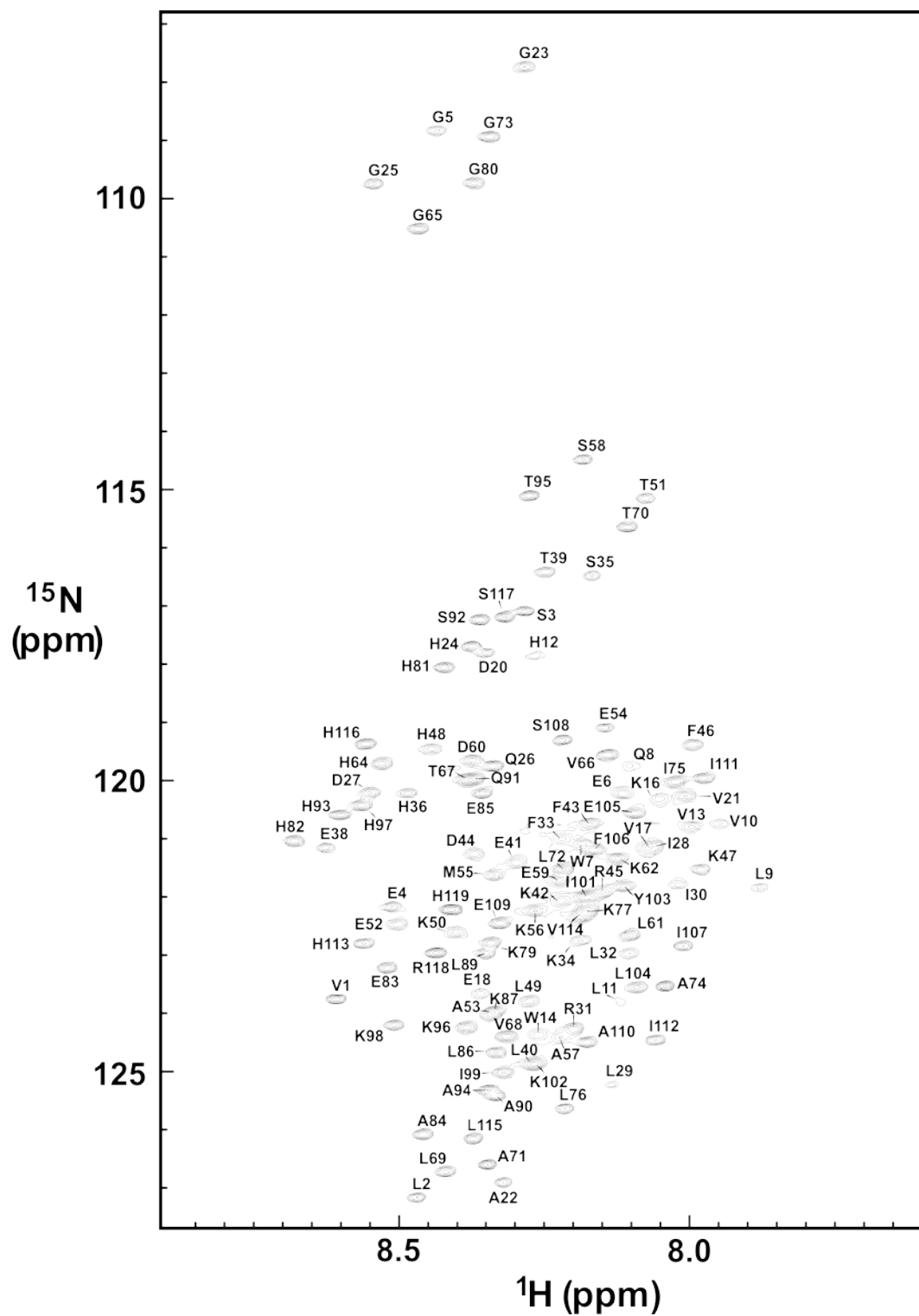


Figure S4

1

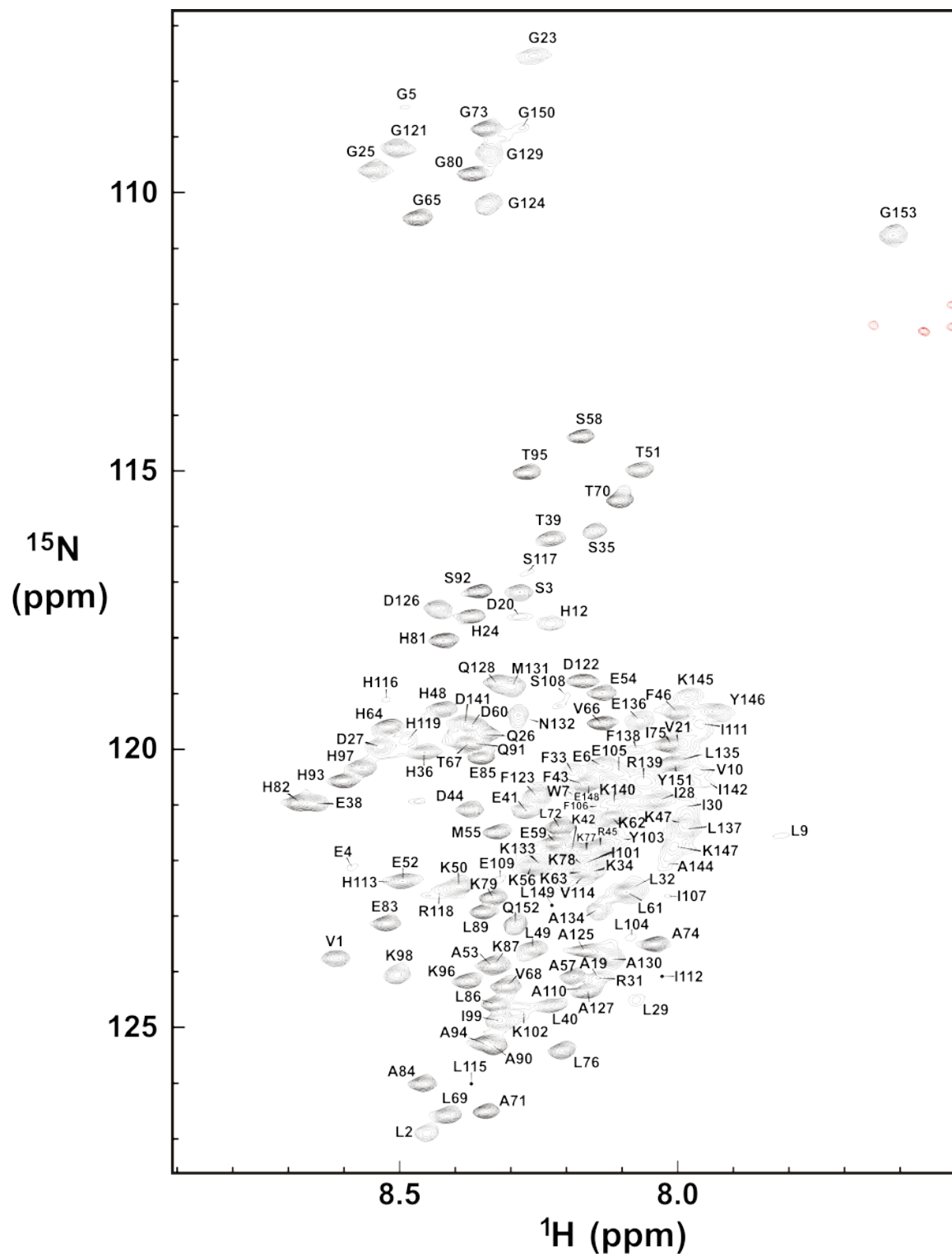


Figure S5

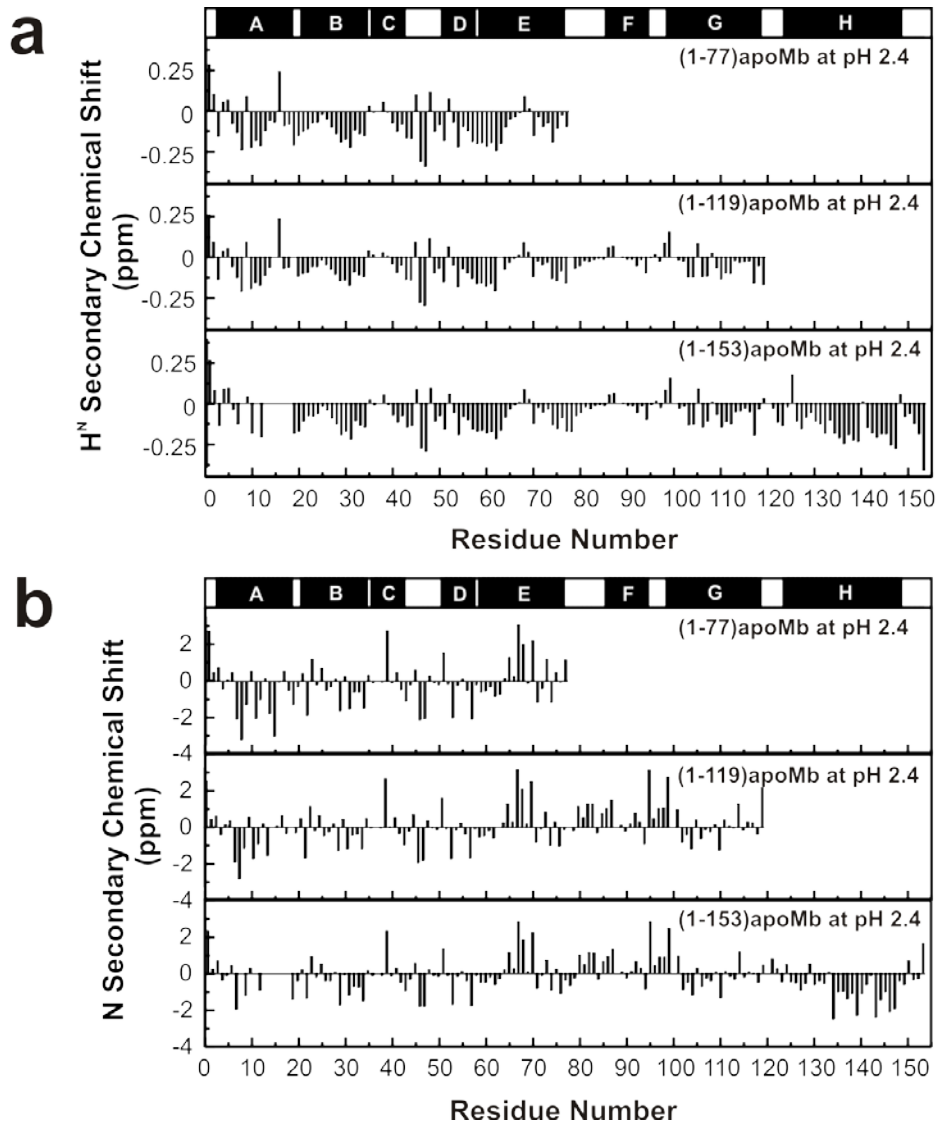


Figure S6

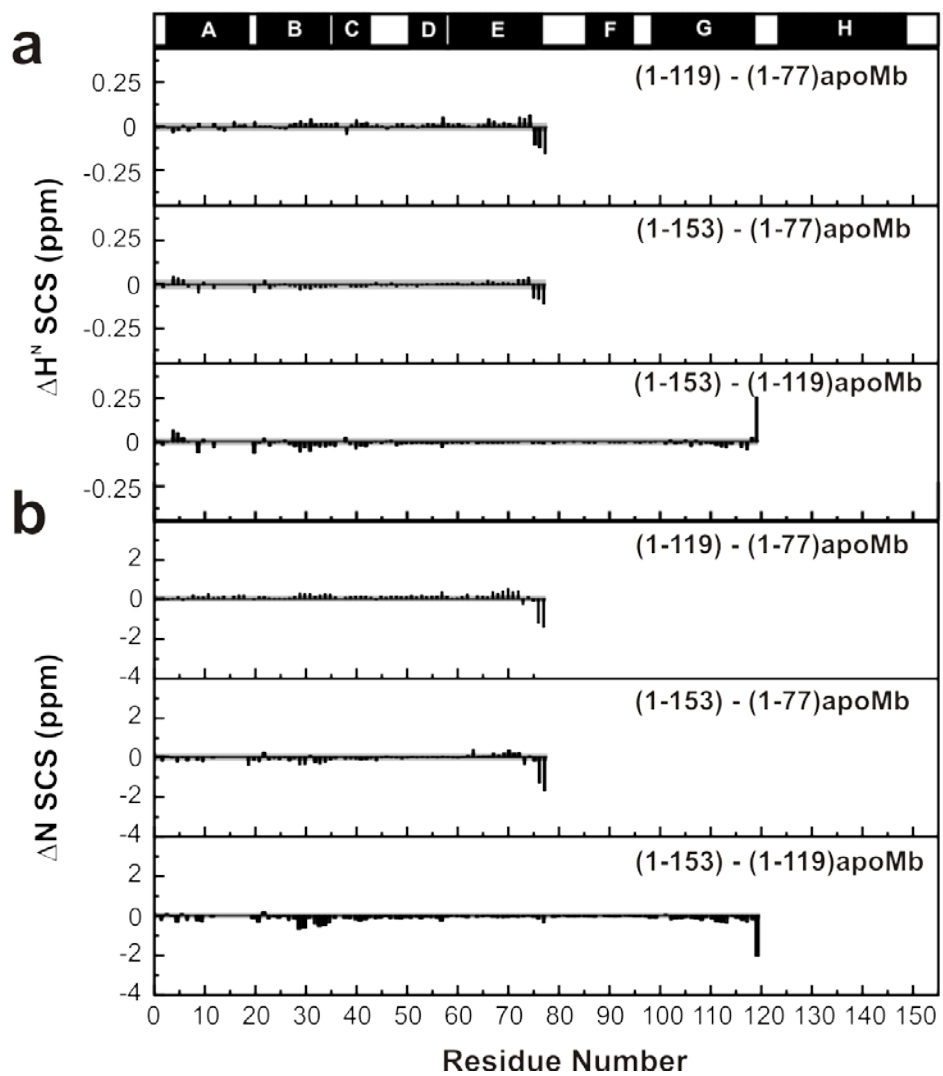


Figure S7

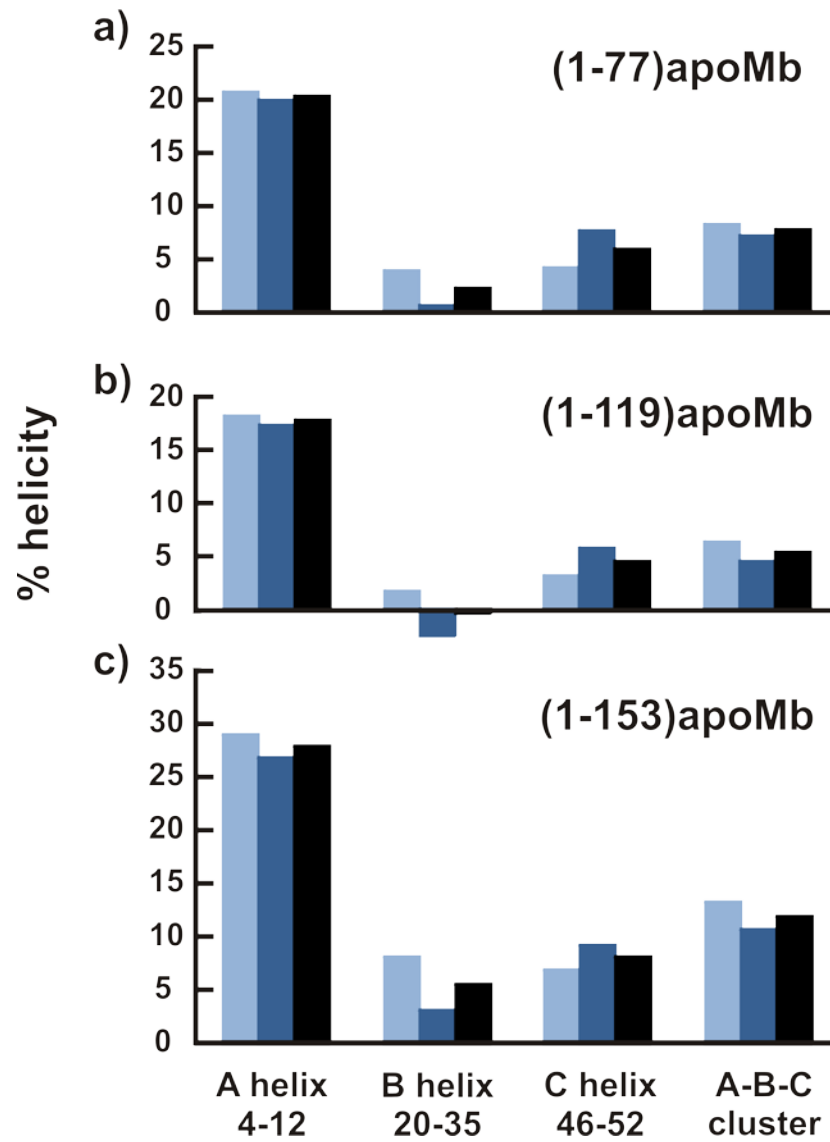


Figure S8

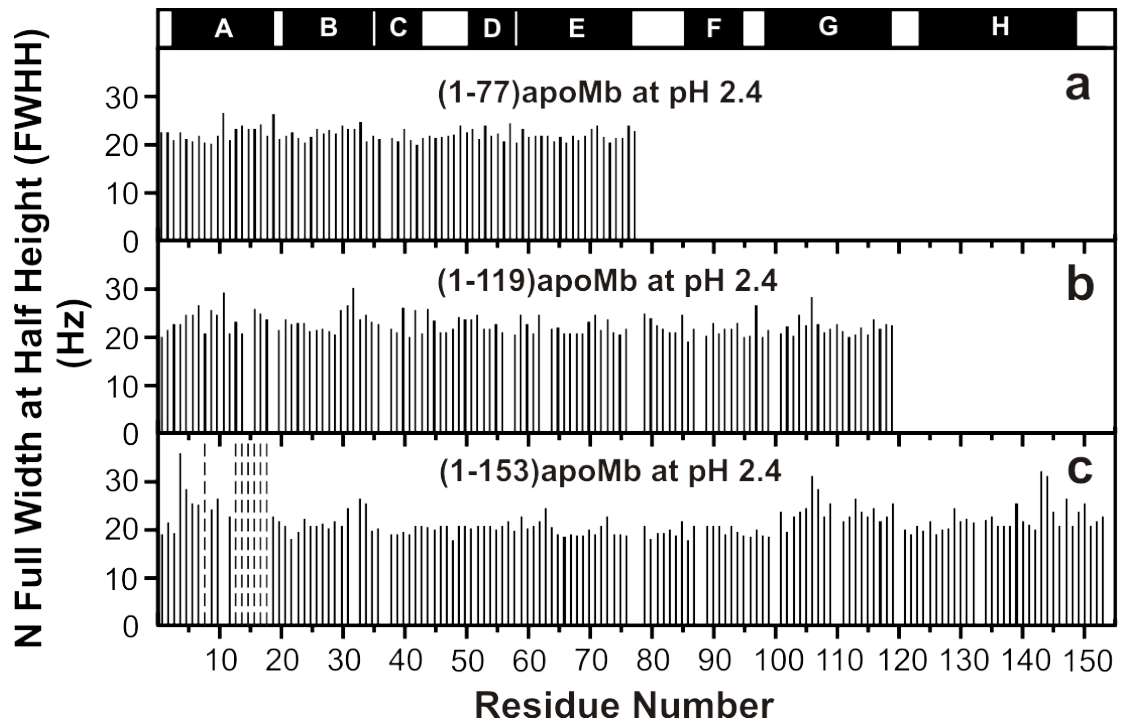
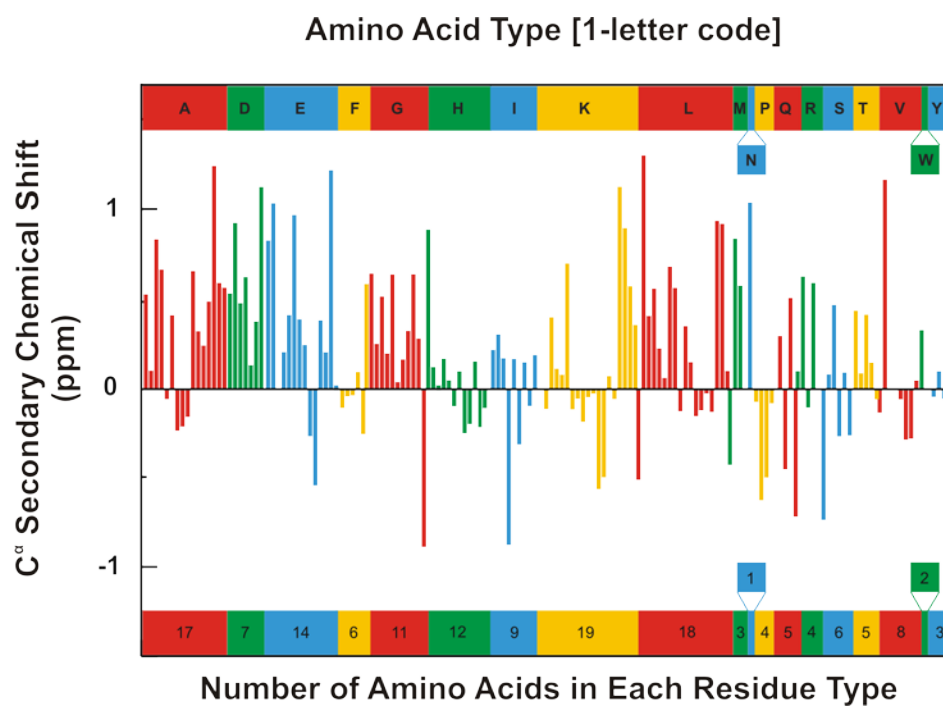


Figure S9



References

1. Ribeiro, E. A., Jr.; Ramos, C. H. I. Circular Permutation and Deletion Studies of Myoglobin Indicate that the Correct Position of Its N-Terminus Is Required for Native Stability and Solubility but Not for Native-like Heme Binding and Folding. *Biochemistry* **2005**, 44, (12), 4699-4709
2. Chow, C. C.; Chow, C.; Raghunathan, V.; Huppert, T. J.; Kimball, E. B.; Cavagnero, S. Chain Length Dependence of Apomyoglobin Folding: Structural Evolution from Misfolded Sheets to Native Helices. *Biochemistry*, **2003**, 42, (23), 7090-7099
3. Wishart, D. S.; Sykes, B. D.; Richards, F. M., Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **1991**, 222, (2), 311-33.
4. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **1995**, 6, (2), 135-40.
5. Lietzow, M. A.; Jamin, M.; Jane Dyson, H.; Wright, P. E., Mapping long-range contacts in a highly unfolded protein. *J. Mol. Biol.* **2002**, 322, (4), 655-662.
6. Felitsky, D. J.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E., Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, (17), 6278-6283.
7. Kurt, N.; Rajagopalan, S.; Cavagnero, S., Effect of Hsp70 Chaperone on the Folding and Misfolding of Polypeptides Modeling an Elongating Protein Chain. *J. Mol. Biol.* **2006**, 355, (4), 809-820.

8. Kay, L. E.; Keifer, P.; Saarinen, T., Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J. Am. Chem. Soc.* **1992**, 114, (26), 10663-10665.
9. Ikura, M.; Kay, L. E.; Bax, A., A Novel-Approach for Sequential Assignment of H-1, C-13, and N-15 Spectra of Larger Proteins - Heteronuclear Triple-Resonance 3-Dimensional Nmr-Spectroscopy - Application to Calmodulin. *Biochemistry* **1990**, 29, (19), 4659-4667.
10. Grzesiek, S.; Bax, A., Improved 3D Triple-Resonance NMR Techniques Applied to a 31-KDa Protein. *J. Magn. Reson.* **1992**, 96, (2), 432-440.
11. Muhandiram, D. R.; Kay, L. E., Gradient-Enhanced Triple-Resonance 3-Dimensional Nmr Experiments with Improved Sensitivity. *J. Magn. Reson. Ser. B* **1994**, 103, (3), 203-216.
12. Kay, L. E.; Xu, G. Y.; Yamazaki, T., Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H₂O Saturation. *J. Magn. Reson. Ser. A* **1994**, 109, (1), 129-133.
13. Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E., A Suite of Triple-Resonance Nmr Experiments for the Backbone Assignment of N-15, C-13, H-2 Labeled Proteins with High-Sensitivity. *J. Am. Chem. Soc.* **1994**, 116, (26), 11655-11666.
14. Wittekind, M.; Mueller, L., HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. *J. Magn. Reson. Ser. B* **1993**, 101, (2), 201-205.
15. Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wuthrich, K., Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Mol. Biol.* **1998**, 280, (5), 933-52.
16. Wishart, D. S.; Nip, A. M., Protein chemical shift analysis: a practical guide. *Biochem. Cell Biol.* **1998**, 76, (2/3), 153-163.

17. Van Geet, A. L., Calibration of methanol nuclear magnetic resonance thermometer at low temperature. *Anal. Chem.* **1970**, 42, (6), 679-80.
18. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* **1995**, 6, (3), 277-93.
19. Goddard, T.D. and Kneller, D.G., SPARKY 3, University of California, San Francisco
20. Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Wright, P. E.; Dyson, H. J., Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR* **2000**, 18, (1), 43-48.
21. Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J., Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *J. Am. Chem. Soc.* **2001**, 123, (13), 2970-2978.
22. Bevington, P. R.; Robinson, D. K., *Data Reduction and Error Analysis for the Physical Sciences*. Second ed.; McGraw-Hill, Inc.: New York, **1992**.
23. Wishart, D.S. and Sykes, B.D., Chemical shifts as a tool for structure determination. *Methods Enzymol.*, **1994**, 239, 363-392.