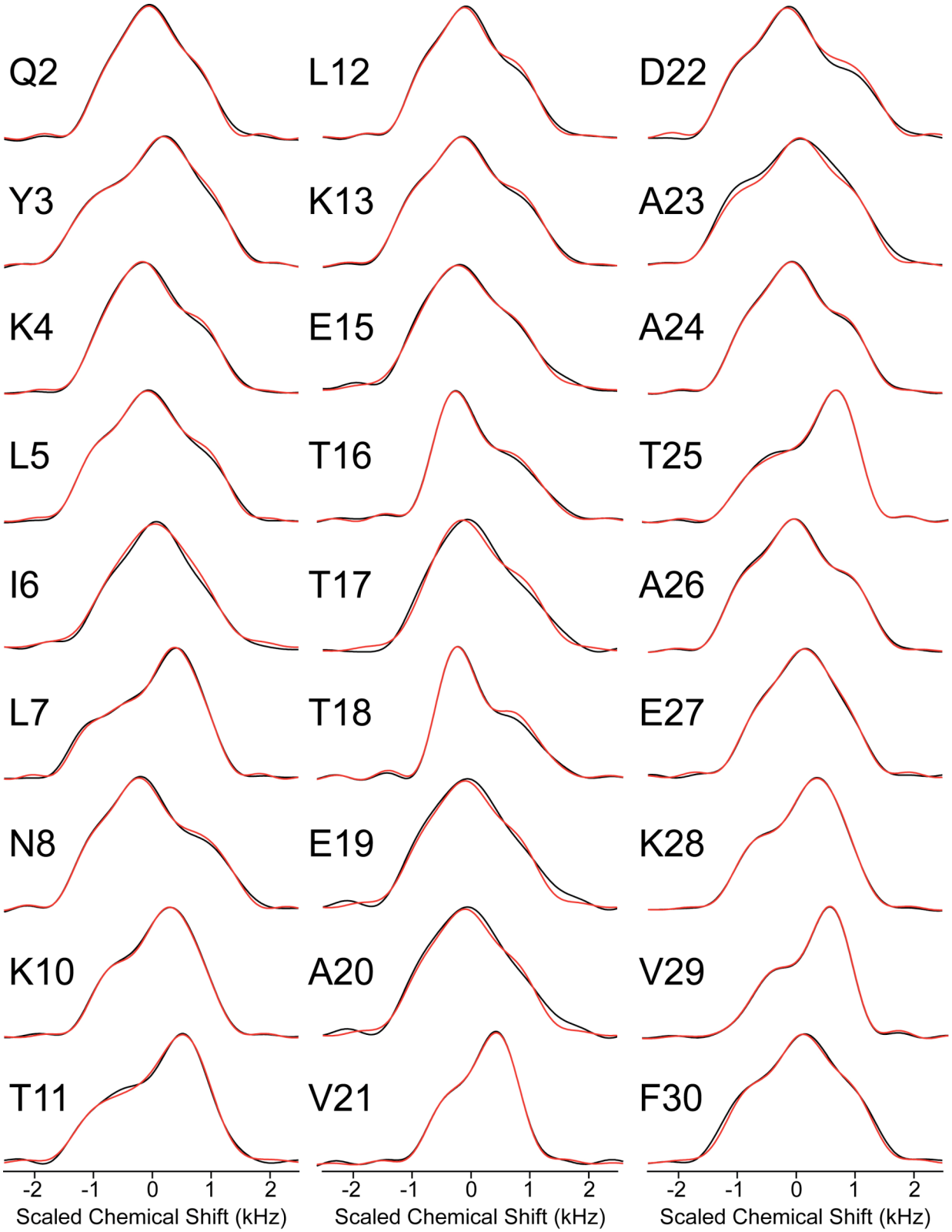
Protein Structure Refinement Using $^{13}C^{\alpha}$ Chemical Shift Tensors

Benjamin J. Wylie,[†] Charles D. Schwieters,[‡] Eric Oldfield[†] and Chad M. Rienstra[†]*

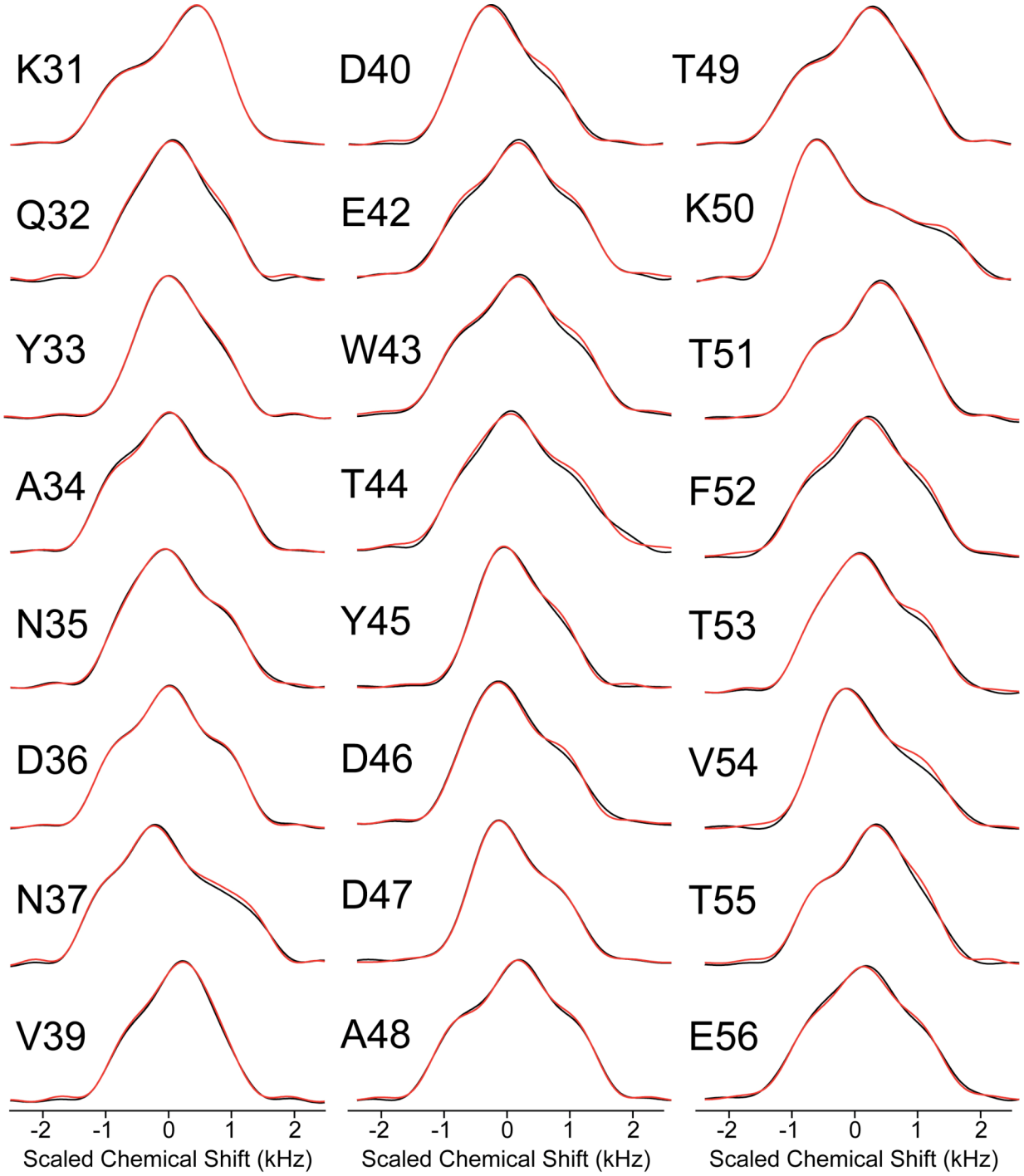
[†]Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801 and [‡]Imaging Sciences Laboratory, Center for Information Technology, National Institutes of Health, Building 12A, Bethesda, MD 20892-5624

E-mail: rienstra@scs.uiuc.edu

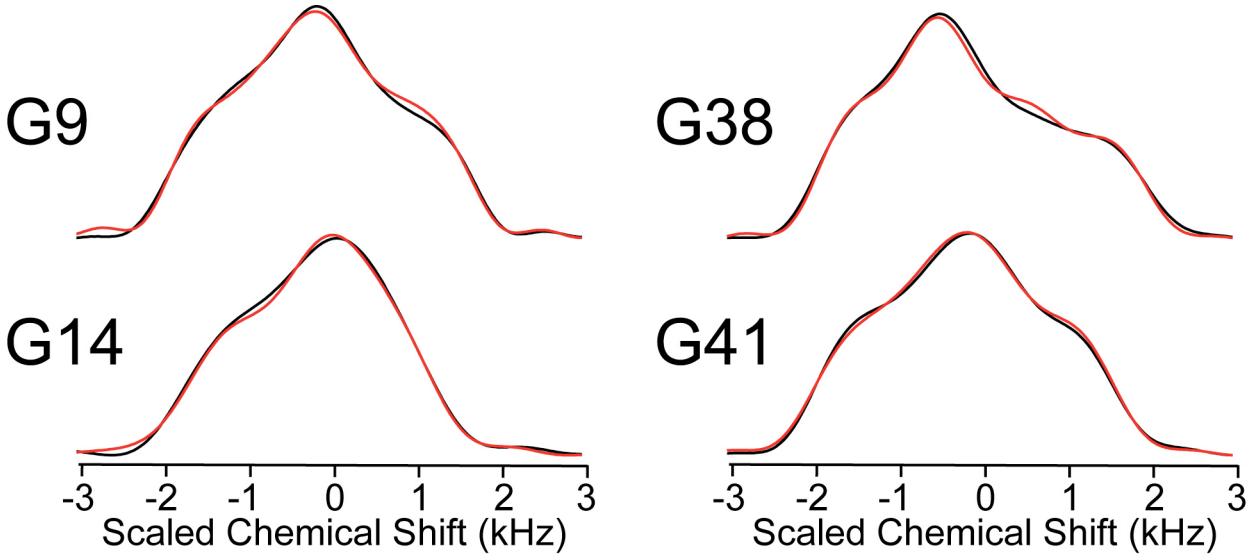
Supporting Information



Supporting Figure S1 continued on the next page.



Supporting Figure S1 continued on the next page.



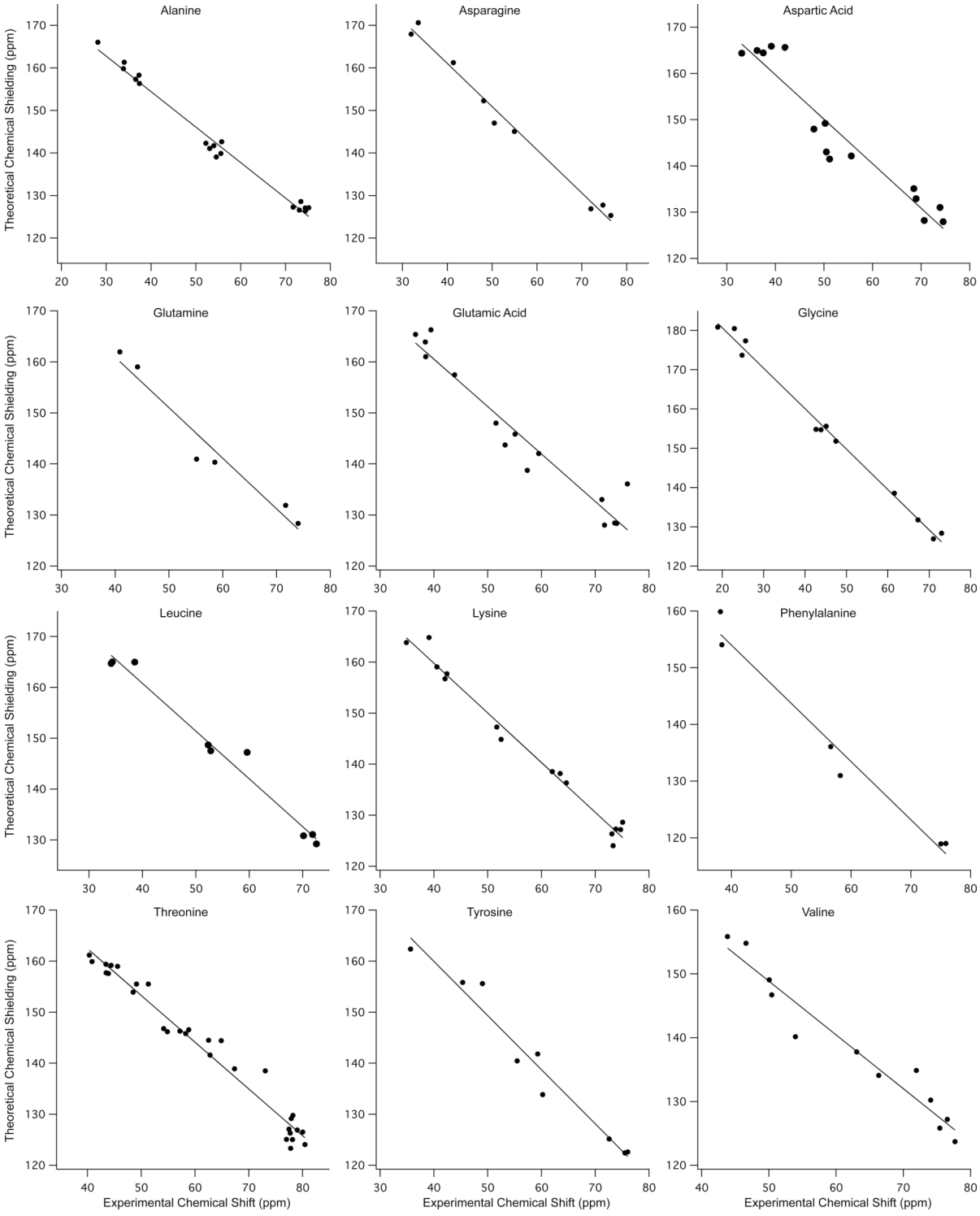
Supporting Figure S1. Experimental ROCSA line shapes and best-fit simulations. Experimental spectra are presented in black, with simulations in red. Glycine residues are pictured separately because their CST is slightly broader than other residue types. Root mean squared deviation between experiment and simulation are less than 3% in all cases.

Regression Analysis. The plot of the principal elements (Figure 4, main text) shows significant variations throughout the protein, arising from the specific dependences of CST parameters upon conformation. The trends are most prominent within a given residue type, where variations in tensor elements correlate well with secondary structure, as well known from previous studies.[1-7] For example, A20, located in a turn with near β-sheet geometry, has a larger span, Ω, compared to the other alanine residues, which possess α-helical geometries. Likewise V54, located in a β-strand, is 25% broader than V21 and V29, both of which adopt α-helical conformations.

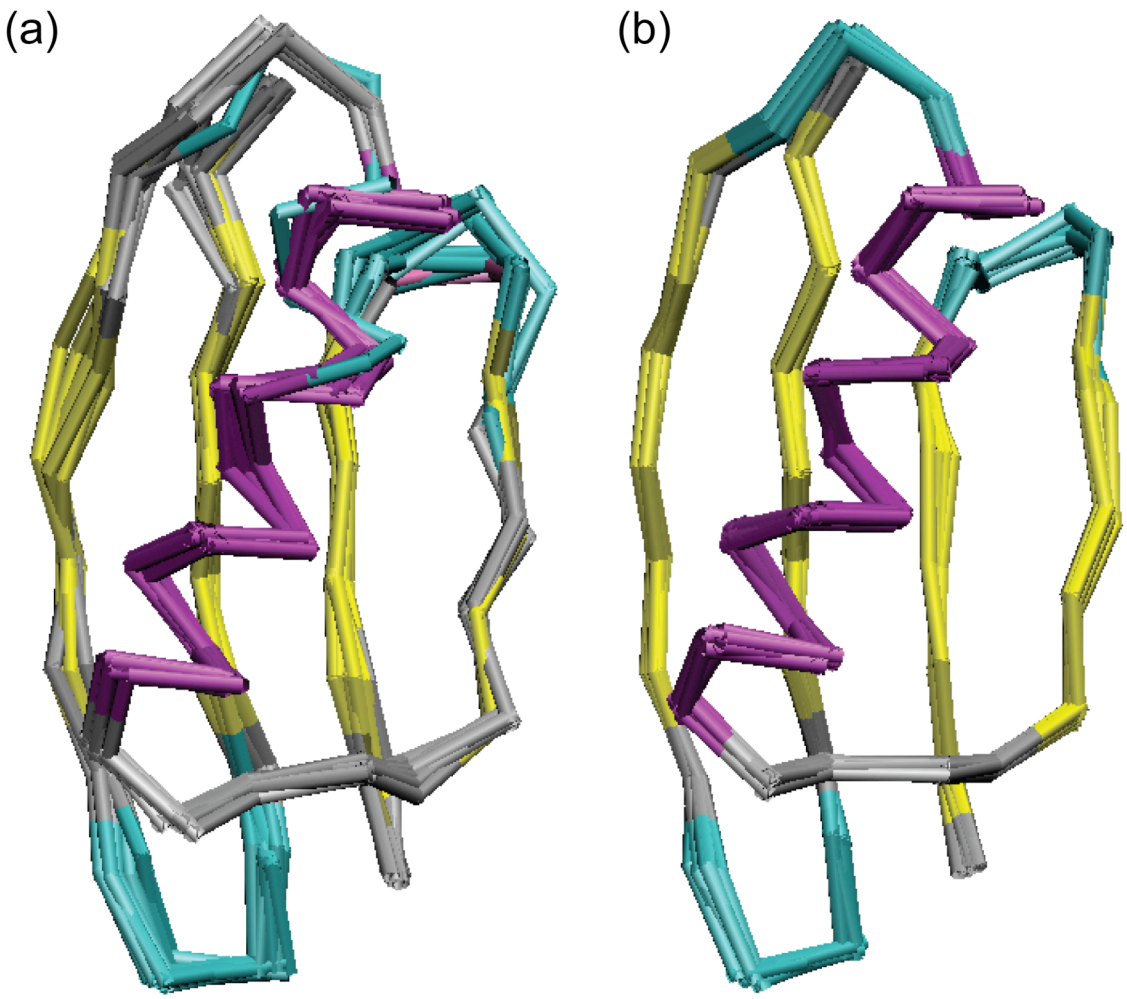
To quantify these trends, we performed regression analysis within each residue type, fitting experimental data versus *ab initio* shielding surfaces to generate regression plots (Table 2 and Supporting Information, Figure S2). (Residue types occurring fewer than three times in GB1 were omitted from this analysis, as was K50, which has an unusual positive ϕ.) Correlation coefficients (R^2) between experimental shifts and theoretical shieldings are 0.92 or greater for all residue types. RMSD values (experiment versus theory) are 3 ppm or less, except in the cases of Asp and Glu, which likely arise from differences in sidechain ionization state between the computed (assumed in the *ab initio* calculation to be a protonated carboxylic acid) and experimental conditions.

A potential source of disagreement between theory and experiment for some residue types is side chain conformation, since the Cα shift depends to some extent on sidechain rotameric state, in addition to its primary dependence upon the backbone torsion angles ϕ and ψ. The *ab initio* surfaces used in this study assume the statistically most populated rotameric states present in protein databases. Deviations from these idealized rotamers lead to small, systematic errors, but do not greatly impair the overall agreement between theory and experiment. For example, the three most common rotameric states for Thr and Val—*gauche*+, *gauche*-, and *trans*—are occupied in GB1, yet the agreement between theory and experiment is still within 2.4 ppm for each tensor element. This result demonstrates that even

assuming idealized side chain geometry, the current shielding surfaces are capable of restraining backbone conformation when properly integrated into a structure determination algorithm. Likewise, other simplifying assumptions (such as an ideal ROCSA scaling factor and the lack of backbone motion) and/or basis set deficiencies in the *ab initio* calculations give rise to additional, small errors. Therefore the overall properties of the experimental and simulated lineshapes are well suited for structure refinement.



Supporting Figure S2. Regression plots of theoretical chemical shielding vs. experimental chemical shifts for all residue types in GB1, computed by using the 2QMT geometry. RMSD between theory and experiment, R^2, p_{offset}, and p_{slope} are provided in Table 2 in the main text.



Supporting Figure S3. Ensembles of ten lowest energy structures of GB1, calculated as described in the main text. Structures presented are: (a) using only distances (structure 1 from Table 3); (b) using distances and TALOS dihedral angles (structure 2 from Table 3). The coloring is according to secondary structure (helix in purple, strands in yellow, turns in cyan, and coils in grey).

References

1. Wi, S.; Sun, H. H.; Oldfield, E.; Hong, M. *J. Am. Chem. Soc.* **2005,** *127*, 6451-6458.
2. Sun, H. H.; Sanders, L. K.; Oldfield, E. *J. Am. Chem. Soc.* **2002,** *124*, 5486-5495.
3. Oldfield, E. *Ann. Rev. Phys. Chem.* **2002,** *53*, 349-378.
4. Havlin, R. H.; Laws, D. D.; Bitter, H. M. L.; Sanders, L. K.; Sun, H. H.; Grimley, J. S.; Wemmer, D. E.; Pines, A.; Oldfield, E. *J. Am. Chem. Soc.* **2001,** *123*, 10362-10369.
5. Pearson, J. G.; Le, H. B.; Sanders, L. K.; Godbout, N.; Havlin, R. H.; Oldfield, E. *J. Am. Chem. Soc.* **1997,** *119*, 11941-11950.
6. Heller, J.; Laws, D. D.; Tomaselli, M.; King, D. S.; Wemmer, D. E.; Pines, A.; Havlin, R. H.; Oldfield, E. *J. Am. Chem. Soc.* **1997,** *119*, 7827-7831.
7. Havlin, R. H.; Le, H. B.; Laws, D. D.; deDios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1997,** *119*, 11951-11958.