Protein Structure Refinement Using Carbon-13 Nuclear Magnetic Resonance Spectroscopic Chemical Shifts and Quantum Chemistry

John G. Pearson,[†] Jin-Feng Wang,[‡] John L. Markley,[‡] Hong-biao Le,[†] and Eric Oldfield*,[†]

Contribution from the Department of Chemistry, University of Illinois at Urbana–Champaign, 505 South Mathews Avenue, Urbana, Illinois 61801, and Department of Biochemistry, University of Wisconsin at Madison, 420 Henry Mall, Madison, Wisconsin 53706

Received April 21, 1995[⊗]

Abstract: We show that protein structures in solution can be refined by using $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts. We investigated 12 alanine and 8 valine residues in the nuclease from *Staphylococcus aureus* and found that the ϕ, ψ, and χ values of these residues are closer to the X-ray ϕ, ψ, and χ values when nuclear Overhauser effect distance restraints and *J*-coupling torsion angle restraints are supplemented with chemical shift restraints. For ϕ in particular, the rmsd versus the X-ray structure is ~10° with chemical shift restraints versus ~20° without chemical shift restraints. Both "fully-restrained" families of structures, those determined with and without chemical shift constraints, had small numbers of minor NOE violations. However, the chemical shift restrained structures were consistent with experimental Ala and Val $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, whereas the structures determined without shift restraints yielded back calculated chemical shifts in poor accord with experiment. Carbon-13 chemical shifts therefore appear to be of use in protein structure refinement when used in conjunction with chemical shift surfaces computed using *ab initio* methods.

Introduction

Nuclear magnetic resonance (NMR) spectroscopy has become the leading technique for determining the structures of proteins in solution.[1,2] Although advances have been made in refining NMR structures derived from nuclear Overhauser effects (NOE) and coupling constants (*J*), these are based on relatively imprecise measurements of distances and dihedral angles. In this article, we demonstrate that the precision of local conformational parameters, such as the backbone torsion angles ϕ, ψ, and χ, can be improved by applying short-range constraints, obtained from (primarily ^{13}C) chemical shifts, within the structure refinement program X-PLOR.[3] Moreover, our results indicate that structures derived from a combination of NOE, *J*, and chemical shift restraints have ϕ, ψ, and χ angles closer to those deduced from X-ray diffraction.

Solution NMR structure determinations typically yield a family of structures with relatively large ϕ, ψ rmsd values (often about 20°) for pairwise comparisons of different NMR structures or one NMR structure and an X-ray structure of the same protein. In addition, NMR structures of the same protein determined in different laboratories under comparable solution conditions appear to vary more than do X-ray structures determined by different groups, with the solution NMR structures more closely resembling the X-ray structures than each other.[4] It is clear then that the quality of NMR structures needs to be improved. Recent studies indicate that carbon-13 chemical shifts (CS) offer some promise as structural constraints, and we and others have shown that the chemical shifts of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ atoms of amino acid residues are strong functions of ϕ, ψ,[5–7] and that it is possible to predict accurately chemical shifts from known structure,[5,8–10] or ϕ, ψ from known chemical shifts.[11,12] Because carbon-13 chemical shifts for proteins can be determined in the solid state as well as in solution, such shifts interpreted in structural terms should provide a way of assessing differences in local structure between the two states. Indeed, a preliminary investigation utilizing ^{13}C chemical shifts and the empirical all-residue ^{13}C $\Delta\delta$ surfaces of Spera and Bax[6] has been reported,[13] with an apparent improvement in the solution structure being obtained even with empirical shift restraints. Since the quantum mechanically computed shielding surfaces for individual amino acids vary considerably from residue-to-residue, and from the empirical surfaces, then even larger improvements might reasonably be expected when using individual, *ab initio* surfaces.

In this work, we incorporate both ^{13}C and ^{1}H chemical shift predicted ϕ, ψ values and back-calculated ^{13}C chemical shift restraints into a conventional NOE-based structure determination and analyze the refinements obtained. We chose to study the alanine and valine residues in nuclease A from *Staphylococcus aureus* (SNase), because this protein has a relatively large

[†] University of Illinois at Urbana–Champaign.
[‡] University of Wisconsin at Madison.
[⊗] Abstract published in *Advance ACS Abstracts*, August 15, 1995.

(1) Fesik, S. W. *J. Biomol. NMR* **1993**, *3*, 261–269.
(2) Wagner, G. *J. Biomol. NMR* **1993**, *3*, 375–385.
(3) Brünger, A. T. *X-PLOR*; Yale University Press: New Haven, CT, 1992.
(4) Smith, L. J.; Redfield, C.; Smith, R. A. G.; Dobson, C. M.; Clore, G. M.; Gronenborn, A. M.; Walter, M. R.; Naganbushan, T. L.; Wlodawer, A. *Nature Struct. Biol.* **1994**, *1*, 301–310.
(5) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *260*, 1491–1496.
(6) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490–5492.
(7) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171–180.
(8) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *J. Am. Chem. Soc.* **1993**, *115*, 9768–9773.
(9) Jiao, D.; Barfield, M.; Hruby, J. M. *J. Am. Chem. Soc.* **1993**, *115*, 10883–10887.
(10) de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1994**, *116*, 5307–5314.
(11) Le, H.; Pearson, J. G.; de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1995**, *117*, 3800–3807.
(12) Celda, B.; Biamonti, C.; Arnau, M. J.; Tejero, R.; Montelione, G. T. *J. Biomol. NMR* **1995**, *5*, 161–172.
(13) Kuszewski, J.; Qin, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson., Ser. B* **1995**, *106*, 92–96.

0002-7863/95/1517-8823$09.00/0

number of such residues in a variety of conformational environments, the crystal structure of nuclease A from the Foggi strain of *Staphylococcus aureus* is known,[14] the ^{13}C and ^{1}H NMR spectra of nuclease A from the V8 strain of *Staphylococcus aureus* (the two proteins are identical except for residue 124 which is histidine in the Foggi but leucine in the V8 variant) have been assigned independently by two groups,[15-18] $^{13}C^{\alpha}$, $^{13}C^{\beta}$ shielding surfaces for alanine and valine have already been determined,[8,10] and preliminary solid state ^{13}C shifts for SNase have been reported.[19] Our work is the quantum chemical analog of the semiempirical approaches described by Case et al.,[20] in which solely H^{α} chemical shifts were used, and Kuszewski et al.,[13] who used empirical C^{α}, C^{β} shift surfaces.

Results and Discussion

In order to investigate the effects of chemical shift restraints on structure, it is necessary to develop a strategy for using chemical shift information in structure prediction, refinement, and validation. We have used two methods in this work, with the results being described below. The first method uses ^{13}C and ^{1}H chemical shifts to predict ϕ, ψ values for alanine residues, with these predicted values then being used as dihedral angle restraints in X-PLOR. The second method uses ^{13}C chemical shifts directly as restraints of molecular conformation through the incorporation of alanine and valine $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift surfaces into X-PLOR. Alternative approaches to better define ϕ, ψ angles can also be provided by coordinate analysis of available one-, two-, and three-bond coupling data,[21] which has not yet been implemented as a refinement method, and was not pursued in this work.

Elsewhere, we suggested a method, the *Z*-surface approach,[11] in which quantum mechanically evaluated chemical shielding surfaces[8] are used to predict the most likely backbone torsion angles, (ϕ, ψ), for residues in proteins. Briefly, this was done by evaluating a "*Z*-surface" for each residue:

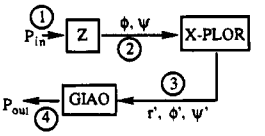
$$Z = \exp(-(\delta - \delta(\phi,\psi))^2/W)$$

where δ is an experimental shift, $\delta(\phi,\psi)$ is the corresponding shift surface, and W is twice the expected mean-square-deviation between the shift surface and experiment, as estimated by the mean-square deviation found between surface predicted chemical shifts and experimental chemical shifts for sites in proteins with rigorously assigned ^{13}C spectra and high resolution X-ray structures. Conformations most consistent with the experimental shift will yield the highest value of *Z*, with the maximum of the Z^3 surface

$$Z^3 = Z^{C^{\alpha}} Z^{C^{\beta}} Z^{H^{\alpha}}$$

yielding the ϕ and ψ values most consistent with the three experimental shifts. These ϕ, ψ values can then be used as additional restraints in an NOE-based structure determination, to yield new torsion angles, ϕ', ψ', which can be used to "back-calculate" the ^{13}C NMR spectrum for comparison with experiment. This method is similar to that used by Celda et al.,[12] except that we are using calculated residue-specific $\delta(^{13}C)$ surfaces and we are applying restraints to residues regardless of secondary structure. A structure that is compatible with NOE-derived restraints and permits correct prediction of the experimental chemical shifts can reasonably be thought to be closer to the actual solution structure than one that is compatible with NOE data but incompatible with the experimental chemical shifts. X-ray structures were used as standards for comparison, since they appear to be more accurate.[4] For example, two independently determined X-ray structures of interleukin-4 have been shown to have a ~0.53 Å rmsd (for residues 4−127), while two similar independently determined NMR structures have a 2.01 Å rmsd when compared on the same basis. Moreover, both solution NMR structures were closer to either X-ray structure than to each other. Also, the restraints used in determining our solution structures are varied between structure families, so a single standard structure is desirable for comparison.

Our first strategy for using chemical shift information to predict, refine, and validate structure can readily be seen from the following flow chart:



We use experimental input parameters (P_{in}, e.g., $\delta(^{13}C^{\alpha})$, $\delta(^{13}C^{\beta})$, $\delta(H^{\alpha})$), one, combined with shielding surfaces and the *Z*-surface approach, to predict ϕ, ψ values. These ϕ, ψ values, two, and NOE data are input into X-PLOR, and the resulting output structures, three, represent the new chemical shift restrained structures. These structures are then used to back-calculate chemical shifts, four, for structure validation purposes.

One of the principal motivations for developing the *Z*-surface method was to convert chemical shift data into restraints which are compatible with X-PLOR, allowing the use of chemical shift data in conjunction with more "traditional" NOE and *J*-coupling results. This requires the torsion angle restraints to be targeted upon single torsion angles, ϕ and ψ, placing stringent limitations upon use of the method, for ^{13}C chemical shifts are complex functions of multiple torsion angles, and multiple solutions for ϕ, ψ can occur. Requiring restraints to be based upon single torsion angles and their associated uncertainties, $\phi \pm \Delta\phi$, $\psi \pm \Delta\psi$, etc., therefore limits the application of the method to simple residues, such as alanine, and requires the use of an empirical $\delta(^{1}H^{\alpha})$ surface which restricts residues to conformations consistent with well-populated regions of (ϕ,ψ) space, as shown in Figure 1. Predicted torsion angles and uncertainties for alanine in the SNase H124L mutant are reported in Table 1, with uncertainties being chosen to exclude all improbable (according to the chemical shift) regions of conformational space. We note that while many of the *Z*-surface predictions are reasonably close to the X-ray values, there are a few instances where the *Z*-surface predicted torsion angles and the X-ray structure disagree, especially at Ala 109. This may be due to errors in either the *Z*-surface method or the X-ray structure, or it may be due to real differences between solution and crystal structures of the protein. Comparison of the *Z*-surface predicted values against those found in the X-ray structure and shift unrestrained X-PLOR structures, Figure 2, demonstrates the potential utility of chemical shift data in refining protein local structure.

A more rigorous approach would be to restrain protein conformation directly with chemical shifts from *ab initio*

(14) Loll, P. J.; Lattman, E. E. *Proteins: Struct., Funct., Genet.* **1989**, *5*, 183−201.
(15) Yamazaki, T.; Torchia, D. Private communication.
(16) Wang, J.; Hinck, A. P.; Loh, S. N.; Markley, J. L. *Biochemistry* **1990**, *29*, 102−113.
(17) Wang, J.; Hinck, A. P.; Loh, S. N.; LeMaster, D. M.; Markley, J. L. *Biochemistry* **1992**, *31*, 921−936.
(18) Wang, J.; Dzakula, Z.; Zolnai, Z.; Markley, J. L. **1995**, manuscript in preparation.
(19) Sparks, S. W.; Cole, H. B. R.; Torchia, D. A. *Chem. Scri.* **1989**, *29A*, 31.
(20) Ösapay, K.; Theriault, Y.; Wright, P. E.; Case, D. A. *J. Mol. Biol.* **1994**, *244*, 183−197.
(21) Edison, A. S.; Weinhold, F.; Westler, W. M.; Markley, J. L. *J. Biomol. NMR* **1994**, *4*, 543−551.

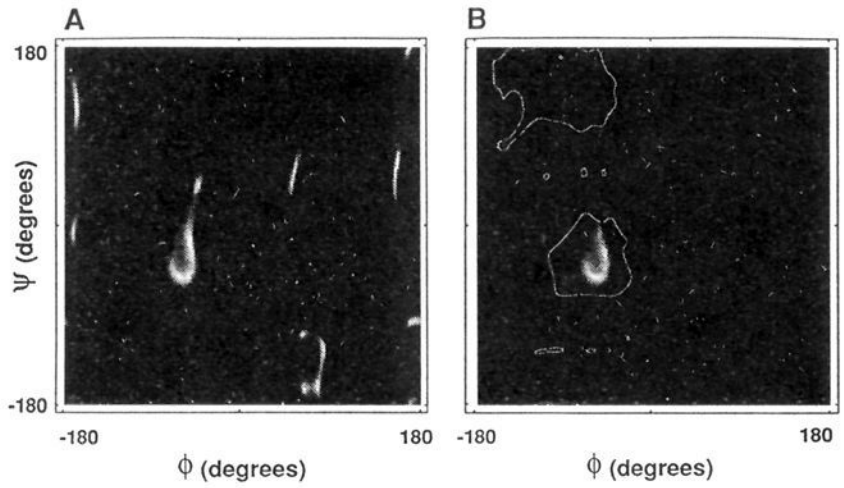


Figure 1. *Z*-surfaces for Ala 60 of Snase: (A) the $Z(C^{\alpha})Z(C^{\beta})$ surface, which indicates several allowable regions; (B) the $Z(C^{\alpha})Z(C^{\beta})Z(H^{\alpha})$ surface. The empirical H^{α} $\delta(\phi,\psi)$ surface was constructed from 169 Ala H^{α} experimental chemical shifts, with the white line indicating the regions in which the empirical surface is well behaved. Based in part on ref 11.

Table 1. Chemical Shift[a] Predicted Torsion Angles (ϕ,ψ) and Torsion Angle Uncertainties ($\Delta\phi,\Delta\psi$) for Ala Residues in Staphylococcal Nuclease (+pdTp,+Ca^{2+})

residue	X-ray ϕ (deg)[b]	ϕ (deg)[c]	$\Delta\phi$ (deg)[d]	X-ray ψ (deg)[b]	ψ (deg)[c]	$\Delta\psi$ (deg)[c]
12	−148	−148	5.0	161	173	10.0
17	−81	−76	7.5	132	133	20.0
58	−62	−45	7.5	−43	−40	15.0
60	−66	−65	5.0	−41	−48	15.0
69	−73	−84	5.0	171	143	22.5
90	−149	−148	7.5	171	166	15.0
94	−119	−97	5.0	−110	114	12.5
102	−56	−50	7.5	−44	−41	20.0
109	−163	−131	7.5	156	163	15.0
112	−148	−143	5.0	172	152	15.0
130	−63	−58	10.0	−39	−40	12.5
132	−60	−44	5.0	−39	−31	10.0

[a] Chemical shifts were for the nuclease H124L-pdTp-Ca^{2++} ternary complex.[16-18] [b] Torsion angles from the structure of Loll and Lattman.[14] [c] Torsion angles are from Le, et al.[11] [d] Uncertainties in ϕ, ψ are given as the half-height half-widths of Gaussian distributions centered about ϕ_i, ψ_i.

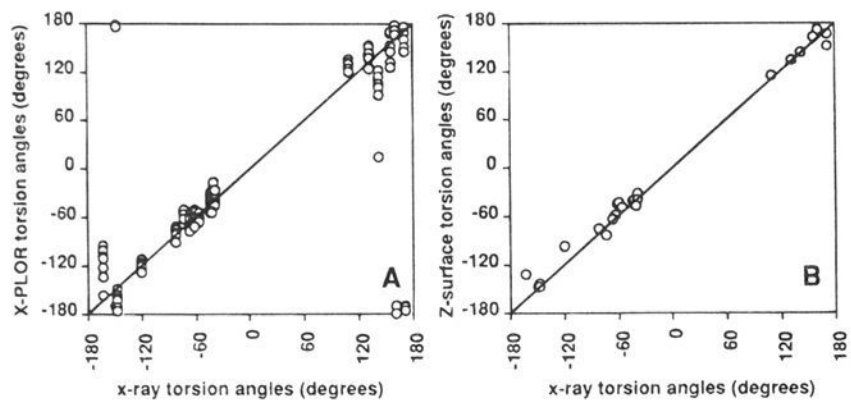


Figure 2. Comparisons of predicted backbone torsion angles with torsion angles from the X-ray structure. (A) X-PLOR predicted torsion angles (from structures generated using N, J, and H sets, as described below) vs X-ray ϕ's and ψ's. The line indicates perfect agreement with X-ray ϕ, ψ results. B) Z-surface predicted torsion angles vs X-ray ϕ's and ψ's. The line again indicates perfect agreement with X-ray ϕ, ψ results.

shielding or shift surfaces, removing the approximations in the *Z*-surface method and eliminating the need to use empirical H^{α} shielding surfaces. We have previously calculated *ab initio* shielding surfaces, $\sigma(\phi,\psi)$, for Ala and Val $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei[8,10] using Pulay's gauge-including atomic orbital (GIAO) algorithm,[22] and these may be converted into shift surfaces as

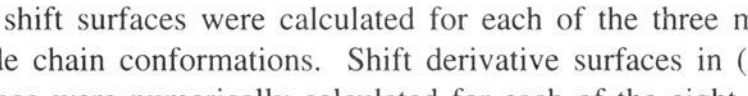
(22) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251−8260.

Table 2. Chemical Shielding Surface Calibration Coefficients and Resulting Chemical Shift Surface Mean Square Deviation and Coefficient of Determination, R^2, from Experimental Chemical Shifts

atom	offset (ppm)	slope	msd (ppm^2)	R^2
A $^{13}C^{\alpha}$	−188.72	−0.782	0.41	0.92
A $^{13}C^{\beta}$	−185.29	−0.584	1.22	0.77
V $^{13}C^{\alpha}(\chi = 180°)$	−184.73	−0.755	1.97	0.82
V $^{13}C^{\alpha}(\chi = -60°)$	−184.73	−0.755	1.97	0.82
V $^{13}C^{\alpha}(\chi = 60°)$	−184.73	−0.755	1.97	0.82
V $^{13}C^{\beta}(\chi = 180°)$	−183.65	−0.574	0.59	0.70
V $^{13}C^{\beta}(\chi = -60°)$	−191.32	−0.744	0.61	0.91
V $^{13}C^{\beta}(\chi = 60°)$	−191.32	−0.744	0.61	0.91

[a] Calibration was performed against experimental chemical shifts from SNase,[16-18] dihydrofolate reductase,[26] and a vertebrate calmodulin.[27] For reasons discussed elsewhere,[11] data from residue 109 of SNase were deleted. All calmodulin valine side chain torsions were set to 180° as suggested by the solution *J* couplings.[10]

described elsewhere.[11] The direct use of the chemical shift as a restraint can be achieved by incorporation of the calculated shift surfaces directly into X-PLOR:[13]



allowing the chemical shifts to be included directly as restraints. Although the all-residue empirical surface was found to be of limited utility,[13] it seems reasonable to believe that calculated residue specific surfaces will be more useful, due to the marked difference in $\delta(\phi,\psi)$ for different residues.[8]

The use of ^{13}C chemical shifts as direct restraints of molecular conformation was achieved by incorporating chemical shift surfaces, $\delta(\phi,\psi)$'s, (and their derivatives with respect to torsion angles), into the source code of X-PLOR, creating a non-standard version of the program dubbed "delta X-PLOR" (D-PLOR). A chemical shift surface, $\delta(\phi,\psi)$, is created by calibrating an ab initio chemical shielding surface, $\sigma(\phi,\psi)$, versus experimental chemical shifts, with the calibration coefficients for alanine and valine $\delta(^{13}C^{\alpha})$ and $\delta(^{13}C^{\beta})$ reported in Table 2. We note that chemical shift and shielding scales are not referenced to the same compound, generating a large zero point offset between the two scales. We also note that the presence of non-ideal slopes in calibration plots with reasonably high correlation coefficients ($R > 0.8$) is probably due to correlation of the local charge field with residue conformation, for the slopes of correlation plots approach ideality (−1.0) upon the addition of point charges in neighboring residues. By using the slope and offset to convert $\sigma(\phi,\psi)$ to $\delta(\phi,\psi)$:

$$\delta(\phi,\psi) = \left\{\frac{\sigma(\phi,\psi) - \text{offset}}{\text{slope}}\right\}$$

We are essentially changing the scale zero point to the resonance frequency of TMS and accounting for the effects of point charges in nearest neighbor residues. Since both $\delta(^{13}C^{\alpha})$ and $\delta(^{13}C^{\beta})$ of valine are functions of the side chain torsion angle, χ, shift surfaces were calculated for each of the three major side chain conformations. Shift derivative surfaces in (ϕ,ψ) space were numerically calculated for each of the eight (two Ala, six Val) shift surfaces, and all shift and shift derivative surfaces were incorporated into X-PLOR. Because of the very large number of calculations needed to generate a full $\sigma(\phi,\psi,\chi)$ hypersurface for valine, we were limited to using the valine shift surfaces of the closest staggered side chain conformations

Table 3. Comparison of D-PLOR Structure Families Against the X-ray Structure of Ternary Complex Staphylococcal Nuclease (+pdTp, Ca^{2+}), together with Back-Calculated Chemical Shift Comparisons

structure family[a]	global SP (Å)[b]	A,V-rich SP (Å)[b]	Ala,Val ϕ (deg)[d]	Ala,Val ψ (deg)[d]	Val χ (deg)[d]	NOE[c]	$\delta(C^{\alpha})$ (ppm)[f]	(C^{β}) (ppm)[f]
N	0.98	0.80	33.6	24.3	18.7	0.9	3.0	3.2
NJ	0.88	0.72	19.4	**15.8**[g]	15.4	0.4	2.2	2.5
NH	0.89	0.66	22.7	19.7	14.5	1.8	2.1	2.8
NZ	1.04	0.80	25.9	20.4	16.0	1.1	1.7	1.7
NS	1.00	0.82	33.7	20.3	12.2	1.2	**0.8**	**0.7**[g]
NJH	**0.73**[g]	**0.55**[g]	19.8	**15.8**[g]	14.8	**0.1**[g]	1.9	2.5
NJZ	0.96	0.79	18.8	16.2	11.7	0[g]	1.5	1.8
NJS	0.88	0.69	**13.5**[g]	17.0	18.5	**0.1**[g]	**0.8**	**1.0**[g]
NHZ	0.86	0.60	20.7	18.6	12.0	1.2	1.5	1.7
NHS	0.95	0.68	32.4	21.8	**10.9**[g]	2.9	1.0	1.2
NJHZ	**0.78**[g]	0.61	16.0	**15.8**[g]	11.8	**0.1**[g]	1.6	1.7
NJHS	0.78[g]	**0.58**[g]	**13.2**[g]	**15.7**[g]	**11.6**[g]	0.2	**0.8**	**1.0**[g]
X-ray	na	na	na	na	na	na	1.4	1.6

[a] D-PLOR structure families were named to indicate the restraint sets used, as described in the text. [b] Average distance rmsd's between D-PLOR and X-ray[14] structure superpositions of backbone atoms for residues 8−42 and 56−141. [c] Average distance rmsd's between D-PLOR and X-ray[14] structure superpositions of backbone atoms for residues 89−115. [d] Average values for torsion angle comparisons between D-PLOR and X-ray[14] structures. [e] Average number of NOE based proton−proton distance restraints violated by between 0.1 and 0.5 Å in non-fatally flawed structures. [f] Average rmsd values between chemical shift surface back-calculations using X-ray[14] and D-PLOR structures and experimental chemical shifts.[16−18] [g] Value given represents a first or second place ranking in the set of 12 structure refinements.

to calculate the chemical shifts, with the χ portion of the restoring force generated by a finite diffence method. Harmonic potentials were used, with the force constants selected to reproduce the root-mean-square deviation of the calibration plots.

Solution structures of the inhibited ternary complex of SNase (SNase + pdTp, Ca^{2+}) were calculated in X-PLOR and D-PLOR via the simulated annealing protocol of Nilges *et al.*[23] with minor modifications. In each simulation, the initial structure of an extended polypeptide strand with the SNase H124L mutant primary sequence was subjected to 24 000 molecular dynamics steps at 2000 K, then cooled to 100 K over 24 000 steps. The restraints used were grouped into five separate sets (abbreviated): 2032 NOE-derived proton-proton distances (N);[18] 228 interatom distances corresponding to 114 NOE-derived hydrogen bonds (H);[18,24] 183 non-shift derived torsion angle restraints (118 ^3J coupling derived ϕ restraints, 20 ^3J coupling derived χ restraints, and 45 dihedral angle restraints on the ligated sugar) (J);[18] 12 pairs of chemical shift derived alanine ϕ and ψ restraints (Z);[11] and 40 ^{13}C chemical shift restraints,[16−18] applied at the α and β carbons of 12 alanines and 8 valines (those Ala and Val residues which are X-ray resolved,[14] minus Val 51, that is believed to be mobile)[19] (S). A harmonic potential was used for both the chemical shift derived torsional restraints and the chemical shift restraints. Any final proton−proton distance more than 0.1 Å outside the NOE derived expected distance range was considered an "NOE violation". NOE violations of more than 0.5 Å were considered fatal (i.e., the structure was discarded). Structures were generated in families of 10, with each family labeled by the restraints used. The families generated were N, NH, NJ, NZ, NS, NHJ, NHZ, NHS, NJZ, NJS, NHJZ, and NHJS, allowing us to analyze the effects of restraints other than traditional NOE-based interproton distances upon the generated structures when compared to the X-ray structure. For example, the effect of the H restraint set is defined as the average change of the ensemble N → NH, NJ → NJH, NZ → NHZ, NS → NHS, NJZ → NJHZ, and NJS → NJHS. All sets contain the N restraint set. This approach was adopted in order to minimize fluctuations due to the small number of individual data points.

Structure families were examined for rms deviation of backbone atoms from the X-ray structure, rms deviation of Ala and Val torsion angles from those found in the X-ray structure, number of minor NOE violations per structure, and rms deviation of back-calculated chemical shifts from experimental shifts. We report average values for each of these criteria for each structure family in Table 3.

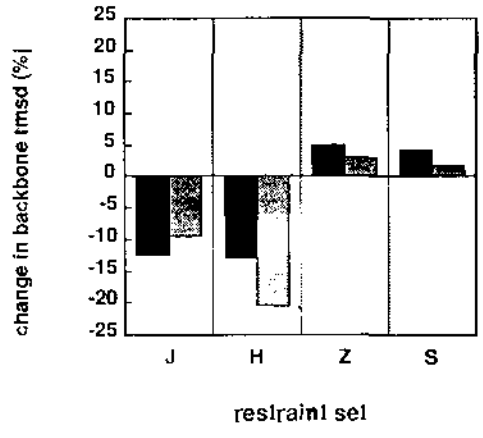


Figure 3. Average percent change of backbone rmsd from the X-ray upon the addition of different restraint sets. Results for superposition of backbone atoms in residues 8−42 and 56−141 are shown in black, results for superposition of residues 89−115 are shown in gray.

Backbone rmsd calculations were performed for both the immobile portions of the protein, residues 8−42 and 56−141, and an alanine and valine rich subset, residues 89−115, to maximize our sensitivity to the effects of chemical shift restraints. We show in Figure 3 the average percent change of rms deviation of backbone atoms in residues 8−42 and 56−141 upon the addition of each of the restraint sets J, H, Z, and S to the NOE-based structure, as determined above. Since negative changes in rmsd represent closer agreement with the X-ray structure, the results in Figure 3 show that the addition of both J and H restraints improves backbone conformational accuracy (i.e., they bring it closer to the X-ray) by about 12%. Conversely, both Z and S restraints degrade backbone conformation by about 4%. Examining the aliphatic rich region, residues 89−115, where 9 of the 27 residues are either alanine or valine, we find similar results, with J and H restraints improving the structures, while Z and S restraints cause a weak degradation of the structures. However, we note that the negative effects of shift-based restraints are less evident in the region where the restraints are most frequently applied (shown in gray in Figure 3).

We show in Figure 4 the average percent change in rms deviation of alanine and valine torsion angles upon the addition of each restraint set J, H, Z, and S. Since negative changes in rmsd again represent stronger agreement with the X-ray

(23) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* 1988, *229*, 317.

(24) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; J. Wiley: New York, 1986.

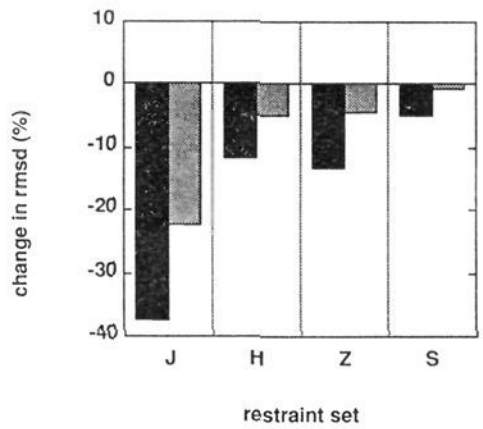


Figure 4. Average change in rms deviation of Ala and Val backbone torsion angles upon addition of different restraint sets. Results for ϕ are shown in black, ψ are shown in gray.

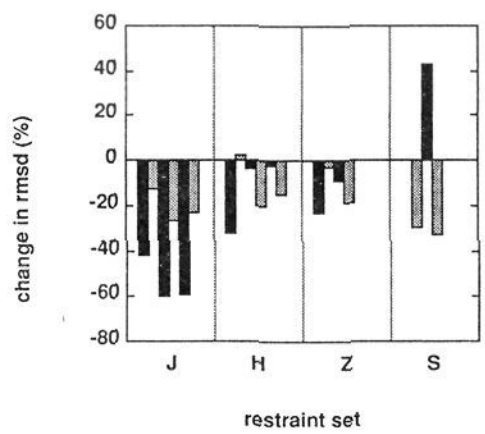


Figure 5. Change in rmsd of Ala and Val ϕ upon the addition of different restraint sets. Initial restraint sets for J were the following (left to right): N (black), NH (gray), NS (black), NZ (gray), NHS (black), NHZ (gray). Initial sets for H were the following: N (black), NJ (gray), NS (black), NZ (gray), NJS (black), NJZ (gray). Initial sets for both Z and S were the following: N (black), NJ (gray), NH (black), NJH (gray), again using a left to right convension. The result for adding S to N is essentially zero, hence the "missing" bar in the figure.

structure, we find that application of each restraint set improves the torsional structures, primarily by improving ϕ. Torsion angle restraints from scalar couplings provided the best average improvement, followed by chemical shift derived (Z) torsion angle restraints and NOE derived hydrogen bond distances. Surprisingly, the direct application of chemical shift restraints only improved ϕ by an average of 5%.

Examining the effects of adding restraint sets in individual cases, Figure 5, we find that the effects of each restraint set depend upon the restraints already present. The addition of the J-restraint set improved Ala and Val ϕ's markedly, as would be expected with the application of ϕ restraints upon all 20 residues of interest, in addition to most of their nearest neighbors. The addition of the Z restraint set also improved ϕ's found in all structure families, but the effect was diluted by the restraints being applied to only 12 of the 20 residues. The addition of the H restraint set improved ϕ when added to the most weakly restrained set as well as when used with the Z set, but had little effect otherwise. The addition of the S set caused no noticeable overall improvement upon the N family of structures, seriously degraded the NH family, but markedly improved the NJ and NJH families.

The difference in behavior between the Z and S sets, and their interaction with the H restraint set, originates from the manner in which the restraint sets were derived. The Z set, by definition, restrains structure conformations about single points in ϕ,ψ space, while the S set allows many possible conformations. Thus, the direct application of a small number of chemical shifts to a relatively poorly restrained structure results in sporadic selection of alternate conformations not allowed in the Z-surface method. Conversely, the direct application of a small number of chemical shifts to refine an already relatively well restrained structure generates dramatic improvements in ϕ, yielding the family of structures most closely resembling the X-ray torsion angles. We illustrate this behavior in Figure 6, where we show comparisons between X-ray determined ϕ,ψ values for 12 Ala and 8 Val residues in SNase with those deduced for families of D-PLOR based structures with varying combinations of restraints. As can be seen in Figure 6A, which illustrates the addition of S to N (i.e., N $\rightarrow$ NS), most of the torsion angles are either improved or unaffected by the addition of chemical shift restraints, but ϕ's in two of the residues are severely degraded. While torsion angles for only one structure from each family are shown in Figure 6A, the trend is generally true for all members of both families, with an individual Ala or Val ϕ being more than twice as likely to be closer to the X-ray in the NS family than in the N family. Indeed, the improvement found in the majority of the residues upon addition of S to N masks the occasional selection of an "improper" conformation. In the case of adding S to NH there was not sufficient room for improvement for the majority of residues to mask a few "improper" conformations, so the effect is more noticeable. The occasional selection of an alternate (incorrect) conformation is fortunately suppressed by additional restraints, as shown in Figure 6B.

We note that both of the families represented in Figure 6B consistently determine ϕ for Ala 109 of SNase to be about 30° higher than that found in the X-ray structure. The good general agreement between the non-shift restrained structures, the shift restrained structures, and the Z-surface predicted conformation (Table 1) suggests to us that the actual solution conformation of Ala 109 in SNase is, therefore, significantly different from that reported in the X-ray structure.

We show in Figure 7A the change in the average number of minor NOE (N set) violations per structure (violations greater than 0.1 Å and less than 0.5 Å) accompanying the addition of the restraint sets J, H, Z, and S. As can be seen from the results in Table 3, there is a marked degradation for N $\rightarrow$ NS and NH $\rightarrow$ NHS structures, leading to poor results. We show in Figure 7B the number of structures containing at least one "fatal" N violation, where families without fatally flawed structures are not shown. Fatally flawed structures result from structures becoming trapped in unfavorable conformations during simulated annealing, an effect which may be reduced through variation of constraint weighting in the simulated annealing algorithm. Again, we note that the NS family is the most troublesome (i.e., had the most fatally flawed members), but all other unacceptable structures were from families generated with the Z set. This is not surprising, since the Z-surface predicted ϕ,ψ restraints are much more rigid than δ restraints (i.e., the region of conformational space allowed by the Z-surface restraints is only a small subset of those allowed by the δ restraints), causing structures generated with the Z set to be more error-prone during simulated annealing. This again indicates that it may be necessary to gradually "ramp in" restraints based upon chemical shifts, especially in less constrained structures.

We next investigated the utility of each structure in back-calculation of Ala and Val $^{13}C^{\alpha}$, $^{13}C^{\beta}$ chemical shifts. Since *ab initio* calculations are very lengthy, and given 12 different combinations of D-PLOR restraints plus the X-ray structure, for a total of 13 "families", we have used the shift surfaces alone to predict chemical shifts, since previously we have shown that good agreement with experiment is obtained using this approach.[10,11] Results are given in Table 3, with the average change in rmsd upon the addition of each restraint set shown in Figure 8.

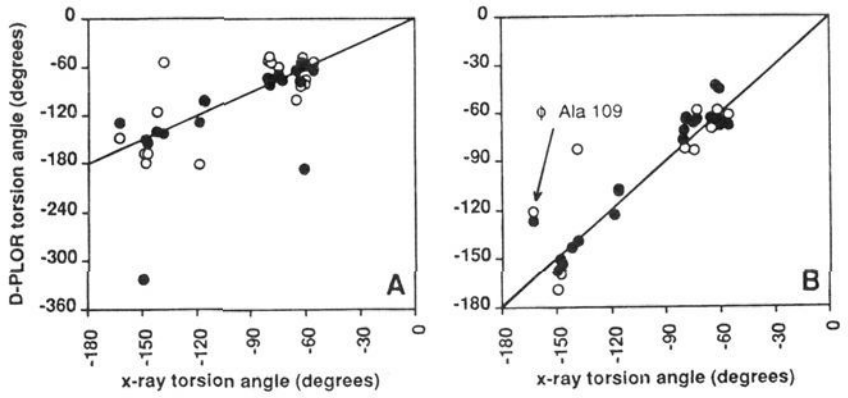


Figure 6. Comparison of NMR solution structure ϕ's with X-ray ϕ's. For clarity, only results for single structures (all with zero minor NOE violations) from each family are shown. (A) Results for two D-PLOR structures, one from the N family (open circles) and one from the NS family (filled circles). The line indicates perfect agreement with the X-ray ϕ results. (B) Results for structures from the NJH (open circles) and NJHS families (filled circles). The line again indicates perfect agreement with the X-ray ϕ results.

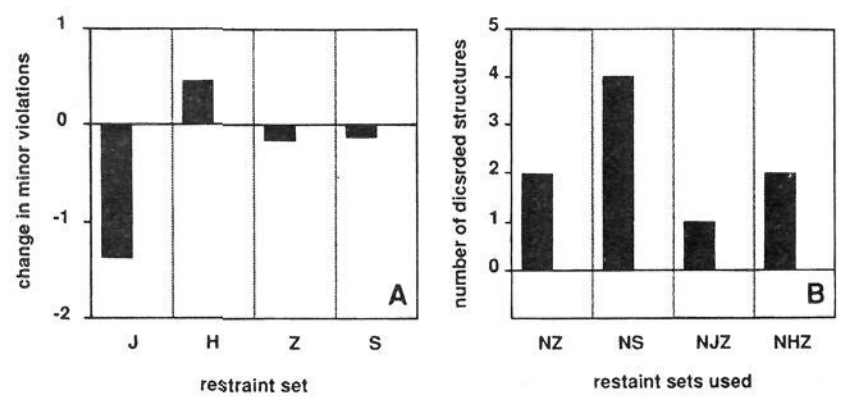


Figure 7. The compatibility of each restraint set with the N set, as measured by (A) the change in minor NOE violations per structure, averaged over acceptable structures, and (B) the number of structures out of 10 that contained at least 1 major NOE violation, and were therefore discarded. Sets not shown had no major violations in any of the structures.

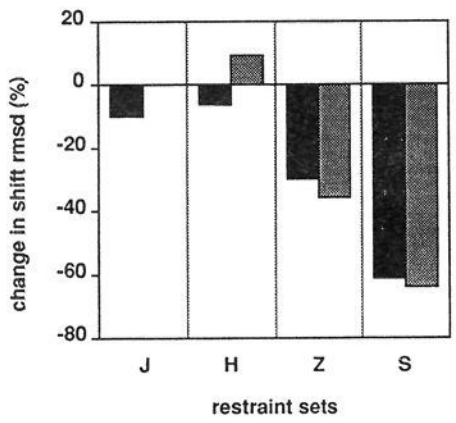


Figure 8. Average change in rmsd of surface back-calculated shifts upon the addition of different restraint sets. Results for $^{13}C^{\alpha}$ are shown in black, those for $^{13}C^{\beta}$ are shown in gray.

Shifts back-calculated from non-shift restrained X-PLOR structures yield poor agreement with experiment, with rmsd's of between 2 and 3 ppm being typical (Table 3). Shifts back-calculated from X-ray structure yield better agreement, with rmsd's of about 1.5 ppm. Shifts back-calculated from shift restrained X-PLOR structures (i.e., the Z set was used) were slightly worse than the X-ray results, with rmsd's between 1.5 and 1.8 ppm. Shifts back-calculated from δ restrained D-PLOR structures (i.e., the S set was used) yield excellent agreement with experiment, with rmsd's between 0.7 and 1.0 ppm being typical. The average change of the shift rmsd's upon the addition of each restraint set (Figure 8) is equally dramatic, with both J and H sets having little effect when compared with the effects caused by incorporation of the Z and S sets. The back-calculated shifts for the two "best" structure families, NJH and NJHS, are shown in Figure 9. At least in the region of the 20 Ala and Val residues we have investigated, even the best shift-

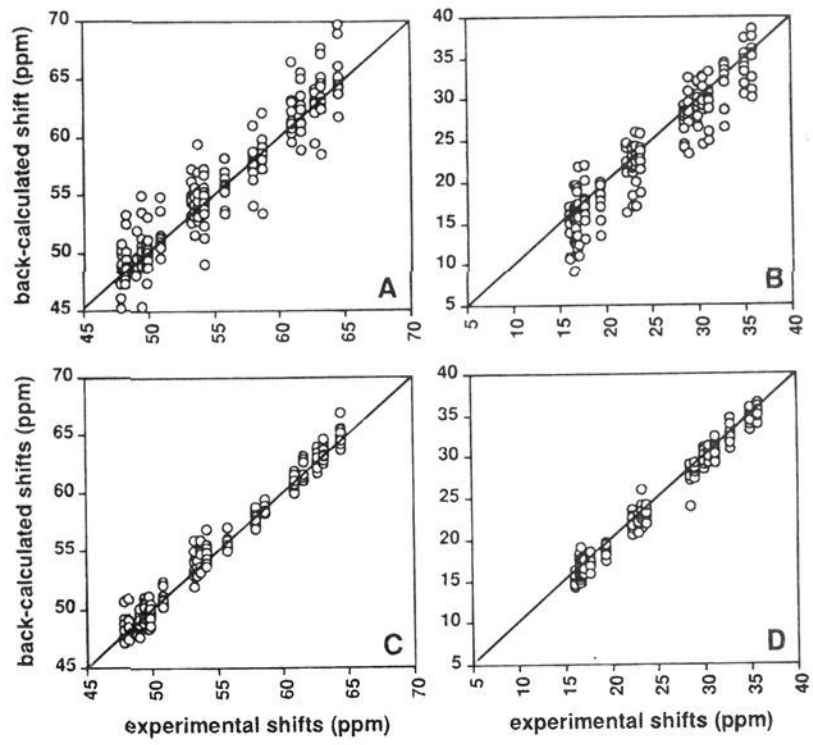


Figure 9. Comparison between back-calculated chemical shifts with experimental chemical shifts for structures in the NJH and NJHS families: (A) back-calculated NJH $^{13}C^{\alpha}$ shifts; (B) back-calculated NJH $^{13}C^{\beta}$ shifts; (C) back-calculated NJHS $^{13}C^{\alpha}$ shifts; (D) back-calculated NJHS $^{13}C^{\beta}$ shifts.

unrestrained structures used in this work are poorly related to the experimental chemical shifts, while the X-ray and shift-restrained structures give generally good agreement with experimental shifts. This is an important result since it implies that shift-restrained solution NMR structures are very close to X-ray structures.

Finally, we investigated the use of structure averaging under NMR constraints and structure minimization under "purely physical" constraints (i.e., no NMR constraints), and found that neither method improved agreement within structure sets or between solution and X-ray structures (data not shown).

When taken together, our results suggest that it is necessary to take into account the extent of the restraint sets available when choosing the optimal method for applying ^{13}C chemical shift restraints to protein structure refinement. If the molecular conformation is poorly restrained, then the Z-surface method can be useful, especially when used to check resonance assignments and otherwise verify restraint sets. This may be seen by comparing the results found for the structure families N and NZ, where application of the Z set helps refine the local structure with only a marginal degradation of the global structure. If the molecular conformation is already well restrained, then the direct application of chemical shift restraints is particularly useful, as can be seen by comparing the results found for the structure families NJH and NJHS. Although the structures of the NJH family are slightly closer to the X-ray in terms of global structure (0.73 Å vs 0.78 Å), as measured by backbone rmsd's over the immobile regions of the protein, the addition of direct chemical shift restraints brings the local structure closer to the X-ray, as measured by Ala and Val torsion angle rmsd's. While the increase in distance rmsd upon application of the S restraints should be noted, the observation that the increase is less for the aliphatic-rich region between residues 89 and 115 (33% Ala or Val versus 17% in regions 8−42 and 56−141) suggests that addition of more shift restraints will reduce or even eliminate this effect.

Since the improvement in local structure is accompanied by much better agreement with the experimental chemical shifts and no significant change in the average number of NOE violations, the addition of a relatively small number of chemical shift restraints to a fairly "high-resolution" SNase solution

structure results in a much better description of its local structure. Application of chemical shift restraints to more poorly restrained structures would require the introduction of a more arbitrary discriminating restraint, as used in the *Z*-surface approach with the inclusion of empirical surfaces. In other words, the inclusion of restraints derived from Ala and Val chemical shifts, together with NOE and *J*-coupling derived restraints, reduces the conformational space available for the predicted SNase structure, with the S restraint set being highly compatible with the more conventional restraints. The resulting solution structure of SNase thus appears closer to the X-ray structure than do NOE and *J*-coupling NMR structures. Indeed, addition of chemical shift restraints appears to also improve agreement between Val χ_1 torsion angles in the solution structures and those found in the X-ray. Since the Val chemical shift surfaces only differentiate between three χ_1 conformers, this improvement must be a scale effect of improved backbone conformation (i.e., backbone and side-chain conformations are correlated in Val residues). It is also reassuring that in cases where the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts predict local conformations different from the X-ray (e.g., Ala 109 of SNase), the NOE and *J*-coupling restraints confirm this difference. Removal of Ala 109 torsion angle deviations from the results in Table 3 would reduce the ϕ rmsd's for the more highly restrained structures by 3−4°, bringing the average rmsd's for the NJHS family to 9.8° for ϕ and 15.7° for ψ, in very close agreement with the X-ray result.

Inclusion of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift restraints therefore allows the prediction of solution state protein structures that are more consistent with the chemical shifts observed experimentally and with the X-ray structure, with further improvements expected as shift restraints for more residue types become available. The NOE, *J*-coupling, hydrogen-bond, and chemical shift based structural results for SNase shown in Table 3 (NJHS) indicate to us that incorporation of the maximal number of restraints yields a structure or structure set in best overall accord with the X-ray distance, ϕ, ψ, χ torsion angle, and experimentally observed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts. The NJH structure is marginally closer to the X-ray in terms of rms deviation (by 0.05 Å over residues 8−42 and 56−141, 0.03 Å over residues 89−115), but it is clearly much worse in ϕ, resulting in poor chemical shift predictions (1.9/2.5 ppm for $^{13}C^{\alpha}/^{13}C^{\beta}$ vs 0.8/1.0 ppm for NJHS). Of the twelve different types of structure refinement, Table 3, the shift restrained (NJHS) set rates as having the best overall agreement with all available structural parameters, ranking in the top two for seven out of eight criteria, while the conventional structure set (NJH) ranks second (four out of eight first or second place rankings). As more surfaces and even hypersurfaces ($\delta(\phi,\psi,\chi,...)$) become available, we believe that ^{13}C chemical shifts will help resolve the apparent existing differences between solution and X-ray structures,[4] especially since, as Torchia, et al.[25] have shown, solid-state ^{13}C (and ^{15}N) shifts of crystalline proteins, including SNase, can now be obtained. This should lead to a more definitive resolution of crystal−solution structural problems using the methods outlined above.

Acknowledgment. We thank Angela Gronenborn for helpful suggestions. This work was supported by the United States National Institutes of Health (Grant No.s HL-19481 and GM-50694) and in part by the American Heart Association with funds provided in part by the American Heart Association, Illinois Affiliate, Inc. (Grant No. 92-013340), and by equipment grants from the International Business Machines Corp., Shared University Research Instrumentation Program, and the Graduate Research Board of the University of Illinois. J.G.P. was supported by a United States Public Health Service Postdoctoral Fellowship (Grant No. GM-14545) and by an American Cancer Society Postdoctoral Fellowship (Grant No. PF-4141). H.L. was supported by a University of Illinois Graduate Fellowship. J.L.M. and J.F.W. received support from NIH Grant No. GM-35976.

JA951269Z

(25) Cole, H. B. R.; Sparks, S. W.; Torchia, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6362−6365.

(26) Soteriou, A.; Carr, M. D.; Frenkiel, T. A.; McCormick, J. E.; Bauer, C. J.; Sali, D.; Birdsall, B.; Feeney, J. *J. Biomol. NMR* **1993**, *3*, 535−546.

(27) Ikura, M.; Bax, A. Private communication.