NMR of hydrogen bonding in cold-shock protein A and an analysis of the influence of crystallographic resolution on comparisons of hydrogen bond lengths

ANDREI T. ALEXANDRESCU,¹ DOUG R. SNYDER,¹ AND FRITS ABILDGAARD²

¹Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269, USA

²Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

(RECEIVED April 12, 2001; FINAL REVISION June 6, 2001; ACCEPTED June 14, 2001)

Abstract

Hydrogen bonding in cold-shock protein A of *Escherichia coli* has been investigated using long-range HNCO spectroscopy. Nearly half of the amide protons involved in hydrogen bonds in solution show no measurable protection from exchange in water, cautioning against a direct correspondence between hydrogen bonding and hydrogen exchange protection. The N to O atom distance across a hydrogen bond, R_{NO} , is related to the size of the $^3J_{NC'}$ *trans* hydrogen bond coupling constant and the amide proton chemical shift. Both NMR parameters show poorer agreement with the 2.0-Å resolution X-ray structure of the cold-shock protein studied by NMR than with a 1.2-Å resolution X-ray structure of a homologous cold-shock protein from the thermophile *B. caldolyticus*. The influence of crystallographic resolution on comparisons of hydrogen bond lengths was further investigated using a database of 33 X-ray structures of ribonuclease A. For highly similar structures, both hydrogen bond R_{NO} distance and C α coordinate root mean square deviations (RMSD) show systematic increases as the resolution of the X-ray structure used for comparison decreases. As structures diverge, the effects of coordinate errors on R_{NO} distance and C α coordinate root mean square deviations become progressively smaller. The results of this study are discussed with regard to the influence of data precision on establishing structure similarity relationships between proteins.

Keywords: Through hydrogen-bond scalar coupling; HN chemical shift; C α RMSD; structure conservation; structure homology; hydrogen exchange

Hydrogen bonds are integral components of protein structure. Although the hydrophobic effect provides the main driving force for the collapse of polypeptides into compact conformations, hydrogen bonds, among other factors, are key determinants of structural specificity (Dill 1990). To date, most information on hydrogen bonds in solution has come from hydrogen/deuterium isotope exchange experiments. Hydrogen exchange protection offers unmatched

sensitivity to conformational dynamics and stability (Bai et al. 1995), but cannot independently identify hydrogen bonding partners or provide structural information about hydrogen bonds. The recently described long-range HNCO experiment enables the direct identification of individual N–H:O=C hydrogen bond donors and acceptors in solution, through *trans* hydrogen bond $^3J_{NC'}$ scalar couplings (Cordier and Grzesiek 1999). The $^3J_{NC'}$ scalar coupling constant appears to be exquisitely sensitive to the N to O atom distance across hydrogen bonds (Cornilescu et al. 1999a; Scheurer and Brüschweiler 1999; Cordier et al. 2000). The availability of a precise and independent measure of hydrogen bond lengths makes it possible to further evaluate the extent to which hydrogen bonds are conserved between proteins.

Reprint requests to: Andrei T. Alexandrescu, Department of Molecular and Cell Biology, University of Connecticut, 75 N. Eagleville Road, U-3125, Storrs, CT 06269-3125, USA; e-mail: andrei@uconnvm.uconn.edu; fax: (860) 486-4331.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1101/ps.14301>.

Here we describe an NMR characterization of hydrogen bonding in cold-shock protein A (CspA) of *Escherichia coli*, a protein with an all β -sheet fold. CspA belongs to a large family of homologous proteins. The HSSP database (Sander and Schneider 1991) currently lists 170 sequences with >30% homology to CspA. At the level of more distant structural relationships, the five-stranded β -barrel core structure of CspA is assigned in the SCOP classification of protein structures to the OB-fold (Murzin et al. 1995). The OB-fold is comprised of 62 protein structural domains, distributed among 16 structural families, and seven structural superfamilies. Part of our interest in CspA is as a model system to investigate whether conserved protein structures reflect conserved mechanisms of protein folding (Alexandrescu et al. 1999).

Results and Discussion

Comparison of CspA hydrogen bonds in solution and in the crystalline state

Hydrogen bonding of CspA in solution was characterized using long-range HNCO spectroscopy. The long-range HNCO experiment is optimized for the detection of small ^{15}N to $^{13}\text{C}'$ scalar coupling constants of ~ 1 Hz, and for the suppression of the larger ~ 15 Hz interresidue $^1J_{\text{N}(i)\text{C}'(i-1)}$ couplings that traverse the peptide bond (Cordier and Grzesiek 1999; Cornilescu et al. 1999b). The small ~ 1 Hz couplings detected in the long-range HNCO experiment include both intraresidue $^2J_{\text{N}(i)\text{C}'(i)}$ and through hydrogen bond $^3J_{\text{N}(i)\text{C}'(j)}$ couplings (Fig. 1). Quantitative determination of coupling constants from cross-peak intensities requires an additional reference HNCO experiment in which only the large $^1J_{\text{N}(i)\text{C}'(i-1)}$ couplings are observed (Cordier and Grzesiek 1999; Cornilescu et al. 1999b). The size of through-hydrogen $^3J_{\text{N}\text{C}'}$ coupling constants can be calculated from

$$^3J_{\text{N}\text{C}'} \approx \frac{[(I_{\text{lr}}/I_{\text{ref}})(N_{\text{ref}}/N_{\text{lr}})]^{1/2}}{(2\pi T)} \quad (1)$$

where the ^{15}N to $^{13}\text{C}'$ INEPT refocusing period T is 66.6 msec, and I_{lr} and I_{ref} are the sizes of the $^3J_{\text{N}\text{C}'}$ and $^1J_{\text{N}\text{C}'}$ correlations in the long-range and reference experiments, respectively (Cordier and Grzesiek 1999). The factor $(N_{\text{ref}}/N_{\text{lr}})$ corrects for differences in the number of transients averaged per free induction decay in the two experiments (e.g., $N_{\text{ref}} = 2$, $N_{\text{lr}} = 8$). Representative strips from reference ("R") and long-range ("H") 3D-HNCO experiments recorded on the 70-residue (7.4 kD) protein CspA are shown in Figure 2. The $^3J_{\text{N}\text{C}'}$ coupling constants determined for CspA are summarized in Table 1.

Hydrogen bonds in the 2.0-Å resolution X-ray structure of CspA (PDB code: 1MJC, Schindelin et al. 1994) were

Fig. 1. Scalar couplings between ^{15}N and $^{13}\text{C}'$ nuclei (A) and parameters defining the geometry of $\text{NH}:\text{O}=\text{C}$ hydrogen bonds (B). Note that the signs of $^1J_{\text{N}\text{C}'}$, $^2J_{\text{N}\text{C}'}$, $^3J_{\text{N}\text{C}'}$ coupling constants are all negative (Cornilescu et al. 1999a). For simplicity, the negative scalar couplings between ^{15}N and $^{13}\text{C}'$ nuclei in this work are reported as absolute values. The geometry of hydrogen bonds between amide and carbonyl groups is defined by three parameters. R_{NO} is the N to O atom distance across the hydrogen bond. $R_{\text{NH-O}}$ is the amide proton NH to O atom distance across the hydrogen bond. $\angle \text{NHO}$ is the angle formed between the N, H, and O atoms. The covalent N to H atom distance, R_{NH} , is assumed fixed at 1.03 Å. From the law of cosines the relationship between the parameters defining the hydrogen bond is: $R_{\text{NO}} = [(R_{\text{NH}})^2 + (R_{\text{NH-O}})^2 - 2(R_{\text{NH}})(R_{\text{NH-O}})\cos(\angle \text{NHO})]^{0.5}$.

defined using a $R_{\text{NH-O}}$ distance threshold shorter than 2.5 Å, and an $\angle \text{NHO}$ angle between 120° and 180° . For hydrogen bonds with multiple amide proton donors or carbonyl acceptors (e.g., bifurcated), only the hydrogen bond with the shortest R_{NO} distance was considered. Based on these criteria, there are 34 backbone hydrogen bonds in the 1MJC X-ray structure. Of these, 23 were identified by NMR (Table 1). The analysis of three hydrogen bonds (residues 10, 13, and 20) was precluded by short ^{15}N T2 relaxation times (Feng et al. 1998). A further six hydrogen bonds could not be unambiguously identified because of resonance overlap in the 3D spectrum. There were no hydrogen bonds observed in solution that are not present in the crystal. Two hydrogen bonds in the 1MJC X-ray structure were not observed within the signal-to-noise limits of the NMR experiment: 4N:2O ($R_{\text{NO}} = 3.20$ Å) and 49N:46O ($R_{\text{NO}} = 3.06$ Å). These hydrogen bonds are part of a half-turn at the N-terminus of the protein, and a reverse turn in the loop between β -strands 3 and 4, respectively. In addition to long R_{NO} distances, the two hydrogen bonds involve residues

Fig. 2. ^1H - $^{13}\text{C}'$ strips from reference ("R") and long-range ("H") 3D CT-HNCO spectra of CspA. Spectra were recorded at 750 MHz on a 2.3-mM $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled CspA sample at pH 5.6, and a temperature of 20°C. The reference spectrum (R) shows exclusively $^1J_{\text{N}(\text{G})\text{C}'(\text{G}-1)}$ couplings. These are labeled according to the sequence number of the preceding carbonyl carbon (e.g., 6'). In the long-range experiment (H) the $^1J_{\text{N}(\text{G})\text{C}'(\text{G}-1)}$ couplings are suppressed, and smaller intrasidue $^2J_{\text{N}(\text{G})\text{C}'(\text{G})}$ (labeled "2J") and through-H-bond $^3J_{\text{N}(\text{G})\text{C}'(\text{G})}$ cross-peaks (indicated by rectangles) are observed. The strips from the two spectra are plotted at the same contour levels. The long-range and reference spectra were acquired with eight and two transients averaged per FID, respectively. Consequently, R and H cross-peaks of equal intensities in this figure are effectively fourfold larger in the reference compared to the long-range experiment.

that, based on NMR relaxation data (Feng et al. 1998), are subject to motional averaging on ps to ns time scales. The S^2 order parameters for residues 49 and 46 are 0.9 and 0.58, respectively. The S^2 values for Ser2 and Lys4 were not determined; however, the intervening residue Gly3 has an order parameter of 0.33 (Feng et al. 1998).

There are four backbone to side-chain hydrogen bonds in the 1MJC X-ray structure: 15N:13OD1 ($R_{\text{NO}} = 3.09 \text{ \AA}$), 27N:25OD1 ($R_{\text{NO}} = 3.20 \text{ \AA}$), 40N:39OD1 ($R_{\text{NO}} = 2.82 \text{ \AA}$), and 46N:49OE1 ($R_{\text{NO}} = 2.70 \text{ \AA}$). Three of these are observed in solution (Table 1). Of the three hydrogen bonds, 46N:49OE1 gives a $^3J_{\text{NC}'}$ value that corresponds to a longer R_{NO} distance (2.96 $\text{\AA}$) than observed in the X-ray structure (2.70 $\text{\AA}$). As previously mentioned, the S^2 order parameter of 0.58 for residue Asp46 (Feng et al. 1998) indicates that this site experiences large-amplitude motional averaging on fast time scales. The 40N:39OD1 hydrogen bond is not observed, but has an NHO angle in the 1MJC X-ray structure that is exactly at the 120° threshold used to identify hydrogen bonds. Within experimental uncertainty, the 40N:39OD1 interaction might not correspond to a hydrogen bond interaction in the CspA crystal.

Hydrogen bonding and protection from solvent exchange

In a previous study, the protection of CspA amide protons from solvent exchange was examined as a function of the denaturant urea and of the stabilizing osmolyte TMAO, tri-

methylamine *N*-oxide (Jaravine et al. 2000). Hydrogen exchange measurements as a function of urea concentration suggested that exchange of amide protons from strands $\beta 1$ – $\beta 4$ occurs through an all-or-none global unfolding mechanism. By contrast, amide protons from the last strand, $\beta 5$, show no detectable protection from exchange in water (Feng et al. 1998; Jaravine et al. 2000). Protection becomes measurable in the presence of TMAO, an osmolyte that stabilizes proteins by raising the free energy of the denatured state compared to that of the native state. The TMAO dependence of exchange protection suggested that amide protons in strand $\beta 5$ exchange through a segmental unfolding of this region of the protein (Jaravine et al. 2000). Part of the motivation for the present study was to verify that the rapidly exchanging amide protons from strand $\beta 5$ are hydrogen bonded in solution. The long-range HNCO experiment unambiguously establishes that all amide protons in strand $\beta 5$ (residues 63–69) that are hydrogen bonded in the X-ray structure of CspA are also hydrogen bonded in solution (Table 1).

Of the 23 hydrogen bonds observed in the long-range HNCO experiments, 10 have amide protons that exchange with rates too fast to measure in water (Table 1). Only three hydrogen-bonded backbone amide protons (36N:33O, 37N:34O, 38N:67O) show no protection at TMAO concentrations above 0.2 M. At concentrations larger than 0.2 M TMAO, however, weak protection is also observed for the amide protons of residues 28, 29, and 45, which are not involved in intramolecular hydrogen bonds in solution or in

Table 1. $^3J_{NC'}$ coupling constants, $^3J_{NC'}$ -derived R_{NO} distances, and hydrogen exchange protection factors for hydrogen bonds in CspA^a

H-bond	$^3J_{NC'}$ (Hz)	R_{NO} (Å)	$P_f(D_2O)^b$	$P_f(0M\ TMAO)^c$
4N:2O	<0.20	>3.15	<230	N.D. ^d
5N:53O	0.70 ± 0.04	2.84 ± 0.03	<220	81 ± 22
7N:51O	0.56 ± 0.06	2.90 ± 0.04	450 ± 80	480 ± 20
8N:22O	0.41 ± 0.09	2.97 ± 0.07	290 ± 30	340 ± 20
9N:49O	ov ^e		ov	ov
10N:20O	lb ^f		320 ± 40	400 ± 30
13N:18O	lb		<609	520 ± 250
17N:13O	0.70 ± 0.11	2.84 ± 0.05	730 ± 100	1010 ± 120
19N:32O	0.66 ± 0.28	2.86 ± 0.11	810 ± 130	970 ± 90
20N:11O	0.48 ± 0.13	2.93 ± 0.08	260 ± 130	480 ± 100
21N:30O	ov		600 ± 150	ov
22N:8O	0.87 ± 0.05	2.79 ± 0.03	1080 ± 60	1210 ± 90
24N:6O	0.36 ± 0.09	3.00 ± 0.08	110 ± 20	140 ± 10
26N:23O	ov		<210	150 ± 10
30N:21O	ov		450 ± 190	ov
31N:62O	lb		260 ± 60	410 ± 80
32N:19O	0.55 ± 0.12	2.90 ± 0.07	460 ± 110	560 ± 40
34N:17O	0.53 ± 0.09	2.91 ± 0.06	<130	47 ± 20
36N:33O	0.28 ± 0.08	3.07 ± 0.09	<340	N.D.
37N:34O	0.25 ± 0.09	3.10 ± 0.11	<30	N.D.
38N:67O	ov		<120	N.D.
48N:9O	0.63 ± 0.07	2.87 ± 0.04	490 ± 90	580 ± 40
49N:46O	<0.15	>3.22	730 ± 70	1000 ± 140
51N:7O	0.51 ± 0.09	2.92 ± 0.06	400 ± 30	480 ± 50
52N:68O	0.65 ± 0.05	2.86 ± 0.04	690 ± 90	710 ± 30
53N:5O	0.53 ± 0.13	2.91 ± 0.07	320 ± 40	420 ± 40
54N:65O	ov		350 ± 60	410 ± 30
56N:63O	0.69 ± 0.08	2.84 ± 0.04	<30	12 ± 6
63N:56O	0.53 ± 0.13	2.91 ± 0.07	<100	7 ± 4
64N:31O	0.80 ± 0.05	2.81 ± 0.03	<170	26 ± 12
65N:54O	0.50 ± 0.10	2.92 ± 0.06	<320	N.D.
67N:36O	0.47 ± 0.13	2.94 ± 0.08	<70	83 ± 20
68N:52O	0.64 ± 0.12	2.86 ± 0.06	<100	N.D.
70N:50O	0.22 ± 0.10	3.13 ± 0.13	9 ± 2	12 ± 1
15N:13OD1	0.22 ± 0.06	3.13 ± 0.09	<90	N.D.
27N:25OD1	0.24 ± 0.05	3.10 ± 0.08	<590	N.D.
40N:39OD1	<0.15	>3.22	<180	N.D.
46N:49OE1	0.42 ± 0.04	2.96 ± 0.05	<60	N.D.

^a Uncertainties in $^3J_{NC'}$ and R_{NO} were estimated according to Equations 5 and 6, respectively, as described in the Materials and Methods section.

^b Protection factors for hydrogen exchange in D₂O calculated from the data of Jaravine et al. (2000) for CspA at pD* 5.4, 5°C.

^c Protection factors, estimated from the TMAO concentration dependence of hydrogen exchange extrapolated to zero concentration of TMAO. Data are from Jaravine et al. (2000).

^d Not determined due to fast hydrogen exchange rates.

^e Not determined due to spectral overlap.

^f Coupling constant not determined due to HN line broadening.

the crystal (P_f [0M TMAO], protection factors extrapolated to zero concentration of TMAO are 70, 80, and 80, respectively). The four amide protons that form backbone to side-chain hydrogen bonds in CspA show no protection in water or at the highest TMAO concentrations studied. There is no discernable relationship between the size of the $^3J_{NC'}$ coupling constant and the magnitude of protection (Table 1). For example, the $^3J_{NC'}$ coupling constant for the 49N:46O hydrogen bond is estimated to be below 0.15 Hz, whereas the protection factor (P_f) for the amide proton of residue 49

in water is ~700. Conversely, residues 63 to 68, which serve as hydrogen bond donors from strand $\beta 5$, give $^3J_{NC'}$ coupling constants above 0.45 Hz but show no protection from exchange in water and only weak protection in the presence of TMAO (Table 1). Published $^3J_{NC'}$ coupling constant data (Cordier and Grzesiek 1999; Cornilescu et al. 1999a) were also compared to hydrogen exchange data (Pan and Briggs 1992; Orban et al. 1995) for ubiquitin and the B1 domain of protein G (not shown). The latter two proteins also show no apparent correlation between the sizes of

$^3\text{H}_{\text{NC}'}$ coupling constants and protection factors (for ubiquitin exchange data at $\text{pD}^* = 3.5$, 22°C , $R = -0.3$, $\rho = 0.1$; for protein G exchange data at $\text{pD}^* = 5.7$, 25°C , $R = 0.2$, $\rho = 0.2$).

Taken together, these observations caution against a direct relationship between hydrogen bonding and protection from solvent exchange. Hydrogen bonds can fail to afford protection if the free energy differences between exchange-susceptible and resistant conformations are small (Alexandrescu et al. 1996). Conversely, amide protons can be shielded from solvent exchange when not involved in intramolecular hydrogen bonds.

Comparison of hydrogen bond lengths in CspA determined by NMR and crystallography

Overall, the hydrogen bonds identified from long-range HNCO data on CspA in solution are in excellent agreement with the 2.0-Å resolution X-ray structure of the protein (PDB code 1MJC). By contrast, N to O atom hydrogen bond lengths calculated from either $^3\text{H}_{\text{NC}'}$ coupling constants or HN chemical shifts, are only weakly correlated with hydrogen bond lengths in the 2.0-Å resolution X-ray structure. Cornilescu et al. (1999a) described the empirical relationship

$$R_{\text{NO}} = 2.75 - 0.25 \ln(^3\text{H}_{\text{NC}'}) \pm 0.06 \text{ \AA} \quad (2)$$

between the size of $^3\text{H}_{\text{NC}'}$ H-bond-mediated couplings in protein G, and N to O atom hydrogen-bond distances in three X-ray structures of protein G variants. The X-ray structures analyzed were a 1.9-Å resolution structure of a protein that differs by two substitutions from the protein studied by NMR (1PGB); and a 1.1-Å resolution structure

of a protein that differs by eight substitutions, refined with isotropic (1IGD) and anisotropic B-factors (2IGD). Equation 2 is based on comparison of the NMR data to N–O distances averaged over all three X-ray structures. Comparison to any of the individual structures gave a poorer correlation than to the average (Cornilescu et al. 1999a).

Equation 2 was used to calculate R_{NO} distances from $^3\text{H}_{\text{NC}'}$ coupling constants measured for CspA (Table 1). Hydrogen bond R_{NO} distances derived from $^3\text{H}_{\text{NC}'}$ couplings are weakly correlated with R_{NO} distances in the 2.0-Å resolution X-ray structure of the protein, 1MJC, and give an R_{NO} distance RMSD of 0.126 Å (Table 2). One factor that might blur the agreement between solution and crystal data for CspA is the moderate 2.0-Å resolution of the 1MJC X-ray structure. Unfortunately, the 1MJC structure is the only X-ray structure available for a cold-shock protein identical to the one studied by NMR. There are four crystal structures available of CspA homologs, solved at crystallographic resolutions ranging from 1.17 Å to 2.70 Å. The 2.5-Å 1CSP and 2.7-Å 1CSQ structures, represent two different crystal forms of a *Bacillus subtilis* cold-shock protein (Schindelin et al. 1993). This protein has 61% amino acid identity to CspA. The 1.17-Å resolution 1C9O:A and 1C9O:B structures (Mueller et al. 2000) are two independent molecules in the asymmetric unit cell of a crystal of cold-shock protein from the thermophile *Bacillus caldolyticus*. This protein has 58% identity to CspA. Pair-wise Cα RMSDs between the five X-ray structures are all below 1.4 Å.

Table 2 compares R_{NO} distances calculated from $^3\text{H}_{\text{NC}'}$ values for CspA in solution, to the corresponding R_{NO} distances in the five X-ray structures of cold-shock protein homologs. The agreement between the two parameters improves with increasing resolution and decreasing R -factors of the X-ray structures. The R_{NO} distances obtained for

Table 2. Agreement between R_{NO} hydrogen bond distances derived from $^3\text{H}_{\text{NC}'}$ coupling constants and HN chemical shifts for CspA in solution, and R_{NO} distances in X-ray structures of cold-shock protein homologs

Structure	Resolution (Å) ^a	R -factor ^b	R_{NO} from $^3\text{H}_{\text{NC}'}$			R_{NO} from $\Delta(\delta\text{HN})$		
			RMSD (Å) ^c	R^d	ρ^e	RMSD (Å)	R	ρ
1C9O:A			0.069	0.68	10^{-4}	0.093	0.70	10^{-5}
1C9O:B			0.088	0.56	0.006	0.112	0.65	10^{-4}
<1C9O>	1.17	0.124	0.074	0.65	10^{-3}	0.094	0.71	10^{-5}
1MJC	2.00	0.187	0.126	0.52	0.01	0.169	0.50	0.004
1CSP	2.50	0.195	0.149	0.39	0.12	0.159	0.34	0.08
1CSQ	2.70	0.220	0.341	−0.28	0.19	0.246	0.20	0.36
Averages over different crystal structures								
<1C9O, 1MJC>			0.072	0.69	10^{-4}	0.103	0.54	0.01
<all>			0.108	0.33	0.14	0.129	0.27	0.23

^a Crystallographic resolution, a parameter describing the highest resolution X-ray data used in building the structural model.

^b Crystallographic R -factor, a measure of the agreement between the structural model and the X-ray data.

^c RMSD, root-mean-square deviation.

^d R , the Pearson coefficient for linear correlations between X-ray and NMR derived ($^3\text{H}_{\text{NC}'}$, HN shift) R_{NO} values.

^e ρ is the probability that the slope of a linear correlation is zero. Values of $p < 0.05$ are considered statistically significant.

CspA in solution are in better agreement with those from the 1.17-Å resolution X-ray structure (1C9O) of the protein from *B. caldolyticus*, than with those from the 2.0-Å resolution X-ray structure (1MJC) of the same protein studied by NMR (Fig. 3). The closer correspondence to the 1.17-Å resolution structure is observed despite the fact that the *B. caldolyticus* protein has only 58% amino acid identity to the *E. coli* protein.

The HN chemical shift of amide hydrogen bond donors has been shown to vary with NH–O distances across hydrogen bonds (Wagner et al. 1983; Wishart et al. 1991; Cordier and Grzesiek 1999). Empirical functions describing the relationship between HN shifts and $R_{\text{NH-O}}$ hydrogen bond distances have been previously reported for BPTI (Wagner et al. 1983) and thioredoxin (Wishart et al. 1991). The present work considers N–O, rather than NH–O distances, for consistency with the analysis of $^3\text{J}_{\text{NC'}}$ coupling constants, and because N and O atom coordinates are more accurately defined than H atom coordinates in protein X-ray structures. An analysis of hydrogen bonds in a database of four ultra high-resolution (<1.0 Å) X-ray structures for which NMR assignments are available (Fig. 4), suggests the empirical relationship:

$$R_{\text{NO}} = \left[\frac{\Delta(\delta\text{HN}) + 4.01}{106.98} \right]^{(-1/3)} \pm 0.12 \text{ Å} \quad (3)$$

where $\Delta(\delta\text{HN})$ is the amide proton chemical shift corrected for its intrinsic “random coil” value ($\delta\text{HN} - \delta\text{random coil}$), and R_{NO} is the N to O atom distance across a hydrogen bond. A similar calibration using $R_{\text{NH-O}}$, rather than R_{NO} distances, gave

$$R_{\text{NH-O}} = \left[\frac{\Delta(\delta\text{HN}) + 2.46}{20.67} \right]^{(-1/3)} \pm 0.13 \text{ Å} \quad (4)$$

with $R = 0.77$. The coefficients for Equation 4 are in good agreement with those previously obtained from an analysis of a 1.5-Å resolution structure of BPTI (Wagner et al. 1983; Deisenhofer and Steigemann 1975).

A comparison between R_{NO} distances calculated from HN chemical shifts of CspA using Equation 3, and X-ray structures of CspA homologs is given in Table 2. Although HN chemical shifts appear to define R_{NO} hydrogen bond distances less precisely than $^3\text{J}_{\text{NC'}}$ values, the HN chemical shifts of CspA also correspond more closely with the 1.17-Å resolution structure of the *B. caldolyticus* protein than with the 2.0-Å resolution structure of the *E. coli* protein studied by NMR (Fig. 3). Averaging of R_{NO} distances over structures in different crystals does not improve the agreement with either HN chemical shifts or $^3\text{J}_{\text{NC'}}$ values (Table 2), indicating that the closer match to the 1C9O structure re-

flects the better quality of this structure. The larger variance of hydrogen bond R_{NO} distances in the 2.0-Å resolution 1MJC structure is evident in the spread of data points along the y-axes of Figure 3A and C.

The R_{NO} distance RMSD over 31 backbone hydrogen bonds conserved between the X-ray structures of the *E. coli* (1MJC) and *B. caldolyticus* (1C9O) cold-shock proteins is 0.130 Å. The values obtained for the *E. coli* protein from $^3\text{J}_{\text{NC'}}$ NMR coupling constants (0.074 Å RMSD for 22 H-bonds) and HN chemical shifts (0.094 Å RMSD for 29 H-bonds) are in better agreement with the (1C9O) structure than the 1MJC X-ray structure (Table 2). The NMR data suggest that hydrogen bond lengths in the *E. coli* and *B. caldolyticus* CspA are more similar to each other than is apparent from the 2.0-Å resolution 1MJC X-ray structure of the *E. coli* protein.

Analysis of ribonuclease A structures

To further characterize the effects of crystallographic resolution on measurements of structural similarity, an analysis was performed on a database of 33 X-ray structures of the protein ribonuclease A (RNase A). Only proteins with the wild-type RNase A sequence (100% sequence identity) and those containing electron density for all 124 residues in the protein sequence were included in the database. Figure 5A (filled circles) shows RMSDs between hydrogen bond R_{NO} distances calculated (Eq. 3) from the published HN chemical shifts of RNase A (Shimotakahara et al. 1997) and R_{NO} distances from RNase A X-ray structures. Figure 5B (filled circles) shows a similar analysis obtained from a comparison of R_{NO} distances from the highest resolution RNase A structure (Wlodawer et al. 1988), 7RSA (1.26 Å), to all RNase A X-ray structures. The agreement between X-ray and NMR-derived R_{NO} distances, as well as the internal agreement between distances obtained from crystallography, shows an apparent exponential improvement with improved resolution of the X-ray structures. To assess the significance of the R_{NO} distance RMSDs, the R_{NO} distances of the 58 hydrogen bonds in the 7RSA structure were compared to 50 simulated data sets of distances randomly distributed between 2.5 and 3.5 Å (the range of distances used to define hydrogen bonds). The resulting average RMSD of 0.33 ± 0.02 Å suggests the limit for comparisons to random R_{NO} distances within hydrogen bonding range. Comparisons between the 7RSA structure and RNase A structures solved at resolution poorer than 2.5 Å start to approach the RMSD limit for uncorrelated hydrogen bond lengths (Fig. 5A,B).

The open squares in Figure 5A and B compare R_{NO} distances in RNase A structures to R_{NO} distances from the 1.5-Å resolution structure of angiogenin 1AGI (Acharya et al. 1995), and to R_{NO} distances derived from the published HN chemical shifts of angiogenin (Reisdorf et al. 1994).

Fig. 3. Comparison of R_{NO} hydrogen bond lengths calculated from solution NMR data on *E. coli* CspA, and R_{NO} distances in the X-ray structures of the *E. coli* (1MJJC) and *B. caldolyticus* (1C9O) CspAs. 1MJJC is a 2.00-Å resolution structure of the same protein studied by NMR. 1C9O is a 1.17-Å resolution structure of a protein with 58% sequence identity to the protein studied by NMR. Each data point represents an individual hydrogen bond. Solid lines indicate the relation $y = x$. The 1C9O X-ray structure contains two molecules in the asymmetric unit cell. Hydrogen bond lengths are averaged over the two structures ($\langle 1C9O \rangle$) unless otherwise indicated. (A) R_{NO} distances calculated from $^3\text{H}_{\text{J}_{\text{NC}}}$ values (Eq. 2) compared to the 1MJJC X-ray structure. (B) R_{NO} distances from $^3\text{H}_{\text{J}_{\text{NC}}}$ values compared to the $\langle 1C9O \rangle$ average structure. (C) R_{NO} distances calculated from HN chemical shifts (Eq. 3) compared to the 1MJJC structure. (D) R_{NO} distances calculated from HN chemical shifts compared to the $\langle 1C9O \rangle$ structure. (E) Correlation of R_{NO} distances between the 1MJJC and $\langle 1C9O \rangle$ X-ray structures. (F) Correlation of R_{NO} distances between the two molecules (1C9O:A, 1C9O:B) in the asymmetric unit of the 1C9O crystal. Correlation parameters are given in Table 2.

Angiogenin is a protein with 36% sequence identity to RNase A. As with comparisons between CspA homologs, and internal comparison between RNase A structures, the

similarity between hydrogen bond distances in angiogenin and RNase A structures depends on crystallographic resolution. The R_{NO} distance RMSDs between angiogenin and

Fig. 4. Correlation between R_{NO} distances in a database of four X-ray structures solved at <1.00 -Å resolution, and conformational (native–random coil) NMR chemical shifts of the corresponding amide proton hydrogen bond donors. The solid line indicates a least-squares fit (Eq. 3) to a $1/R_{NO}^3$ dependence of the HN conformational chemical shift ($R = 0.72$ for 213 hydrogen bonds).

RNAse A structures show a weaker dependence on resolution than internal comparisons between RNAse A structures. As the similarity between the proteins compared decreases, random errors in atomic coordinates have a progressively smaller influence on RMSDs.

Structural similarity between proteins is typically expressed in terms of C α atom coordinate RMSDs. Whereas most hydrogen bonds participate in conserved interactions in the protein core, global C α RMSDs can be dominated by contributions from poorly conserved surface-exposed loops. The dependence of C α RMSDs on resolution is less pronounced and more difficult to interpret than hydrogen bond distance RMSDs. The filled circles in Figure 5C show pairwise C α RMSDs between X-ray structures of RNAse A (residues 3–121) and the highest 1.26-Å resolution structure, 7RSA. The plot suggests a linear dependence of RMSD on crystallographic resolution with an increase of about 0.27 Å in C α coordinate RMSDs for every 1-Å decrease in resolution. The open circles in Figure 5C show pairwise C α RMSDs between all RNAse A structures and the lowest 2.7-Å resolution structure 1RBJ. Compared to alignments to the 1.26-Å resolution 7RSA template, the alignments to the 2.7-Å resolution 1RBJ template show a higher y-intercept of 0.5 and a smaller slope of 0.08. Finally, the open squares in Figure 5C show C α RMSDs between the 1.5-Å resolution 1AGI structure of angiogenin and the RNAse A structures. The pairwise C α RMSDs are closely distributed around 1.50 Å (1.47 to 1.65 Å), and are essentially independent of the precision of the RNAse A

structures (slope = 0.015). In summary, the data in Figure 5C suggest that the effects of structural precision on C α RMSDs are most evident when the protein structures compared are highly similar (e.g., C α RMSDs below 0.5 to 0.75 Å).

Structural similarity and model quality

Structural similarity is used to infer evolutionary relationships between proteins, to determine the effects of sequence perturbations or intermolecular interactions on structure, and in some cases to provide initial clues about proteins with unknown functions. The effects of model quality on measures of structural similarity have been largely unexplored. The high precision afforded by atomic resolution X-ray data, and NMR $^3J_{NC'}$ coupling constants, help to bring into better focus the conservation of hydrogen bond lengths between *E. coli* and *B. thermophilus* cold-shock proteins. That hydrogen bond lengths and C α coordinate RMSDs of RNAse A structures show systematic variations as a function of crystallographic resolution, suggests that differences in data precision account for at least some of the differences within this family of protein structures.

Figure 6 summarizes the distribution of crystallographic resolution parameters for protein X-ray structures in the December 2000 release of the PDB (Berman et al. 2000). The median and mean resolution of 9116 protein X-ray structures in the PDB is 2.1 Å (Fig. 6). The fraction of protein structures with resolution better than 1.50 Å is 3%. There are 36 structures determined at resolution of 1.00 Å or better, accounting for 0.4% of the database. Crystallographic *R*-factors are weakly correlated with crystallographic resolution ($R = 0.43$, $\rho < 0.0001$ for 8900 protein structures); however, the *R*-factor is a parameter that depends on several details of refinement that can make the relationship between crystallographic *R*-factors and resolution ambiguous (Kuriyan et al. 1987).

The analysis in this work has focused primarily on X-ray structures. Quality factors for X-ray structures are relatively standardized and accessible. Typically, NMR structures tend to diverge more from high-resolution X-ray structures than most X-ray structures. The 2AAS NMR structure of RNAse A (Santoro et al. 1993), for example, gives a C α RMSD of 1.1 Å to the 7RSA structure, a value nearly double that of any X-ray structures of the wild-type protein. With the exception of domain-swapped RNAse dimers (1BSR, 11BG) and of a desiccated RNAse A (1C0C), C α RMSDs above 1.1 Å are observed only for proteins with less than 50% sequence identity to the wild-type protein. Similarly, the 3MEF NMR structure of CspA (Feng et al. 1998) gives a C α RMSD of 1.54 Å to the 2-Å resolution 1MJC X-ray structure of this protein. Pair-wise comparisons between any of the X-ray structures of CspA homologs (1C9O:A, 1C9O:B, 1MJC, 1CSP, 1CSQ), give C α RMSDs

Fig. 5. Influence of crystallographic resolution on comparisons of RNase A structures. (A) R_{NO} distance RMSDs between 33 RNase A structures and hydrogen bond lengths calculated from NMR chemical shifts (Eq. 3) for RNase A (filled circles) and angiogenin (open squares). The curves illustrate least-squares fits of the data to exponential functions (RNase A, solid curve, $y = 0.07 * \exp^{(0.45x)}$, $R = 0.81$; angiogenin, dotted curve, $y = 0.12 * \exp^{(0.24x)}$, $R = 0.73$). (B) Same as in A, except R_{NO} hydrogen bond lengths are compared to those in the highest resolution (1.26 Å) 7RSA X-ray structure of RNase A (curve fit parameters: RNase A, solid curve, $y = 0.02 * \exp^{(0.93x)}$, $R = 0.90$; angiogenin, dotted curve, $y = 0.07 * \exp^{(0.39x)}$, $R = 0.81$). (C) Dependence of $C\alpha$ coordinate RMSDs on crystallographic resolution. Filled circles, solid line—pair-wise $C\alpha$ RMSDs obtained after least-squares superposition of individual RNase A structures onto the highest resolution structure, 7RSA ($y = -0.12 + 0.27x$, $R = 0.69$, $\rho = 10^{-5}$). Open circles, dashed line: pair-wise $C\alpha$ RMSDs obtained after least-squares superposition of individual RNase A structures onto the lowest (2.7-Å) resolution structure, 1RBJ ($y = 0.50 + 0.07x$, $R = 0.39$, $\rho = 0.03$). Open squares, dotted line: pair-wise $C\alpha$ RMSDs obtained after least-squares superposition of individual RNase A structures onto the 1.5-Å resolution structure of angiogenin, 1AGI ($y = 1.49 + 0.02x$, $R = 0.16$, $\rho = 0.36$).

below 1.40 Å. High $C\alpha$ RMSDs between NMR and X-ray structures may be due to genuine differences between proteins in solution and in crystals. Portions of the polypeptide chain that are disordered in solution, such as loops, are improperly modeled by a single conformation and can contribute disproportionately to global $C\alpha$ RMSDs. A comparison of hydrogen bond lengths obtained from $^3J_{NC'}$ coupling constants of CspA (Table 1) with those of the best representative conformer in the 3MEF ensemble of NMR structures, gives a hydrogen bond R_{NO} distance RMSD of 0.69 Å. Four of the 23 hydrogen bonds identified in the long-range HNCO experiment (24N:6O, 34N:17O, 36N:33O, 37N:34O) have R_{NO} distances outside hydrogen bonding range, reflecting insufficient experimental restraints. Hydrogen bond restraints in NMR structures are typically inferred from NOE connectivities and hydrogen exchange protection, and the amide protons of the four hydrogen bonds are not protected under the conditions (30°C, pD* 6.0) used to determine the 3MEF structure (Feng et al. 1998). When only R_{NO} distances between 2.5 and 3.5 Å in the 3MEF structure are considered, the R_{NO} RMSD of 0.24 Å to the $^3J_{NC'}$ data approaches that obtained for comparisons to simulated random hydrogen bond distances (0.32 ± 0.03 for 23 hydrogen bonds). This observation suggests that the ag-

Fig. 6. Histogram summarizing the crystallographic resolution of 9116 protein X-ray structures in the December 2000 release of the PDB (Berman et al. 2000).

gregate of restraints used in NMR structure calculations leave hydrogen bond lengths largely undefined. Hydrogen bonds in NMR structure calculations are usually restrained to uniform ranges (e.g., 1.8 to 2.3 Å for $R_{\text{NH-O}}$, 2.7 to 3.3 for R_{NO}). Tighter restraints based on $^3\text{J}_{\text{NC}'}$ coupling constants or HN shifts could potentially improve the definition of hydrogen bond lengths in NMR structures.

The emphasis in structural biology has recently shifted towards increasing the automation and speed of structure determination. The motivation for this shift is that the database of protein sequences far exceeds that of known protein structures. Access to a larger number of experimentally determined protein structures should lead to a greater coverage of protein fold space through comparative modeling. A protein structure in outline may suffice for many purposes. Some problems will continue to require the highest quality models available, and will continue to benefit from methodologies that improve the accuracy of structural data. The analysis of the relationship between hydrogen bond length and NMR parameters such as the $^3\text{J}_{\text{NC}'}$ coupling constant and the HN chemical shift represents one such example. The quality of protein structure models has yet to reach a plateau.

Materials and methods

NMR spectroscopy

3D HNCO experiments were recorded on a 2.3 mM $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled CspA sample at pH 5.6 and a temperature of 20°C, using the 750 MHz Bruker instrument of the National Magnetic Resonance Facility at Madison. The CspA sample was prepared and purified as previously described (Chatterjee et al. 1993; Alexandrescu and Rathgeb-Szabo 1999), with the modification that BL21(DE3) *E. coli* cells harboring the pET11-CspA vector were grown in a 99.9% D_2O solution of modified New Minimal Medium (Budisa et al. 1995): 55 mM KH_2PO_4 , 100 mM K_2HPO_4 , 10 mM Na_2SO_4 , 1 mM MgSO_4 , 1 mg/L Ca^{2+} , 1 mg/L Fe^{2+} , 1 µg/L Cu^{2+} , 1 µg/L Mn^{2+} , 1 µg/L Zn^{2+} , 1 µg/L MnO_4^{2-} , 10 mg/L thiamine, 10 mg/L biotin, 100 µg/mL ampicillin, supplemented with $^{15}\text{NH}_4\text{Cl}$ (1 g/L) and d_7 - $^{13}\text{C}_6$ -D-glucose (3 g/L). The level of ^2H enrichment in the CspA sample was estimated to be above 90% based on comparison of aliphatic and amide proton regions of 1D NMR spectra. A standard ($T = 16.7$ msec) 3D HNCO experiment (Kay et al. 1994) was used to extend previously reported $^1\text{H}^{\text{N}}$, ^{15}N , and $^{13}\text{C}'$ resonance assignments for CspA at pH 6.0, 30°C (Feng et al. 1998) to the conditions of the present study. Side-chain carbonyl and carboxyl ^{13}C resonance assignments were obtained from a 2D H(CA)CO experiment (Yamazaki et al. 1994) recorded on a 1-mM ^{13}C -labeled sample of CspA using a 600-MHz Bruker spectrometer of the University of Basel.

Trans hydrogen bond $^3\text{J}_{\text{NC}'}$ scalar coupling constants were measured using long-range ($T = 66.6$ msec) and reference ($T = 50$ msec) 3D TROSY CT-HNCO experiments (Wang et al. 1999). Both experiments were collected with $1024 \times 58 \times 40$ complex points and spectral widths of $8893 \times 2632 \times 2381$ Hz for ^1H , ^{15}N , and ^{13}C frequencies, respectively. The long-range HNCO data were recorded with eight transients averaged per FID, for a total experiment time of 52 h. The reference HNCO data were

recorded with two transients averaged per FID, for a total experiment time of 13 h. NMR data were processed with Felix 2000 (MSI). The indirectly detected time domains were extended by linear prediction prior to digital filtering and zero filling. The ^{15}N dimension was doubled using mirror image linear prediction (Zhu and Bax 1990). Forward-backward linear prediction (Zhu and Bax 1992) was used to extend the ^{13}C dimension by 50%. Linear prediction in a given dimension was performed only after the other two dimensions were transformed, as described by Kay et al. (1991). Specifically, following the processing of the ^1H dimension, the ^{15}N dimension was preprocessed by applying a cosine window truncated at 5%, prior to zero filling and Fourier transformation. After linear prediction, apodization, and Fourier transformation were applied to the ^{13}C dimension; the ^{15}N dimension was inverse Fourier transformed, and a digital filter corresponding to the inverse of the previously used truncated cosine function was applied. Subsequent linear prediction, apodization, and Fourier transformation of the ^{15}N dimension completed the processing. Lorentzian-to-Gaussian apodization was used in all three dimensions.

The sizes of $^3\text{J}_{\text{NC}'}$ coupling constants were calculated from the peak intensities of one-bond N (i) to $\text{C}'(i-1)$ correlations in the reference HNCO spectrum, and three-bond N(i) to $\text{C}'(j)$ correlations in the long-range HNCO spectrum, according to Equation 1. Peak intensities were determined from the cross-peak matrix point with the largest value. In cases where hydrogen bonds were present in the X-ray structure of CspA but not observed in the long-range HNCO experiment, lower limits on the size of $^3\text{J}_{\text{NC}'}$ (Table 1) were estimated from the root mean square value of the baseline noise in the long-range HNCO experiment.

Experimental uncertainties in $^3\text{J}_{\text{NC}'}$ were estimated according to

$$\sigma(^3\text{J}_{\text{NC}'}) = 0.5 * (^3\text{J}_{\text{NC}'}) * [(\sigma(I_{\text{R}}))/I_{\text{R}}] \quad (5)$$

where $\sigma(I_{\text{R}})$ is the root mean square baseline noise in the long-range HNCO experiment. This formula is derived from standard error propagation of Equation 1, with the assumption that uncertainties in peak intensities of the strong $^1\text{J}_{\text{NC}'}$ correlations, measured in the reference spectrum, have a negligible effect on the uncertainty of $^3\text{J}_{\text{NC}'}$ (Jaravine et al. 2001). Equation 1 is an approximation valid for $|2\pi ^3\text{J}_{\text{N(i)C'(j)}}T| \ll 1$ and $^1\text{J}_{\text{N(i)C'(i-1)}} \sim 15$ Hz. Use of this approximation introduces an uncertainty of less than 0.03 Hz in the derived coupling constants (Cordier and Grzesiek 1999). Consequently, the uncertainties in $^3\text{J}_{\text{NC}'}$ reported in Table 1 include a uniform contribution of 0.03 Hz to account for the approximate nature of Equation 1.

To evaluate the uncertainties in the R_{NO} distances calculated from $^3\text{J}_{\text{NC}'}$ coupling data, it is necessary to account for the uncertainties of the empirically derived coefficients $A = 2.75$ and $B = -0.25$ of Equation 2. The data for the protein G variants reported by Cornilescu et al. (1999a) were used for a nonlinear least-squares fit to Equation 2, with the coefficients A and B as fitted parameters. The variance-covariance matrix for the fit gives estimates for the uncertainties $\sigma(A) = 0.022$ and $\sigma(B) = 0.027$ as well as for the covariance of A and B , $\sigma(AB)^2 = 0.521 \times 10^{-3}$. The uncertainty in R_{NO} can be evaluated using error propagation of Equation 2:

$$\sigma(R_{\text{NO}}) = \sqrt{\sigma(A)^2 + [\sigma(B)^2 * (\ln(^3\text{J}_{\text{NC}'}))^2] + \left[\frac{B * \sigma(^3\text{J}_{\text{NC}'})}{^3\text{J}_{\text{NC}'}} \right]^2 - [2 * \sigma(AB)^2 * (\ln(^3\text{J}_{\text{NC}'}))]} \quad (6)$$

where the covariance term $\sigma(AB)$ has been included because of the relatively high correlation between the parameters A and B: $\sigma(AB)^2/[\sigma(A)\sigma(B)] \sim 0.9$.

Analysis of hydrogen bonds in X-ray structures

Hydrogen atoms were added to heavy atoms with the program INSIGHT 2000 (MSI) using idealized covalent geometry and an N–H bond length of 1.03 Å (Creighton 1993). High-resolution X-ray structures often include hydrogen atoms. An analysis of eight structures solved at better than 1.0 Å resolution indicated that covalent N–H bond distances were shorter than 1 Å, and showed systematic differences between structures, suggesting uncertainty in the precise location of hydrogen atoms even in the highest resolution X-ray structures. For example, the mean and standard deviations of N–H covalent bond distances in the 1BXO structure of penicillopepsin were 0.8600 ± 0.0004 Å. The corresponding values for the 1AHO structure of scorpion toxin II were 0.908 ± 0.004 Å. Because of this consideration the analysis of hydrogen bonds in this work focuses on N to O atom distances, rather than NH to O distances. For X-ray structures in which hydrogen atom coordinates were reported, the hydrogen atoms were removed and replaced by hydrogen atoms with idealized N–H bond geometry and a fixed N–H covalent bond length of 1.03 Å. The 1.03 Å covalent NH distance is more consistent with recent NMR studies of ubiquitin aligned in dilute liquid crystalline media in solution (R_{NH} of 1.041 ± 0.006 Å), and values from neutron diffraction studies on model compounds in the solid state (Ottiger and Bax 1998).

Hydrogen bonds were identified using thresholds of a NH–O atom distance shorter than 2.5 Å, and an NHO angle between 120° and 180° with the program INSIGHT 2000. N to O atom distances (R_{NO}) were calculated from R_{NH-O} distances and $\angle_{NHO}$ angles obtained from INSIGHT using the law of cosines and a fixed R_{NH} distance of 1.03 Å (Fig. 1B), or from N and O atom coordinates. In cases of bifurcated hydrogen bonds only the interaction with the shortest R_{NO} distance was considered. Except where noted, hydrogen bond parameters were averaged in cases where multiple independent copies of the same protein occupied the asymmetric unit cell of a crystal. Statistical analyses were performed with the programs Kaleida Graph 3.5 (Synergy) and S-Plus 5 (MathSoft).

Comparison of hydrogen bond lengths in CspA homologs

Structure-based sequence alignments were calculated with the CE program (Shindyalov and Bourne 1998). The 1C9O, 1CSP, and 1CSQ structures have no alignment gaps. Compared to the other cold-shock proteins, *E. coli* CspA (1MJC) contains three extra residues at its N-terminus, and a one-residue insertion (Asp24). Using the sequence numbering for the *E. coli* protein, the alignment of the 1MJC structure to the other cold shock protein structures is 1MJC(4–23):X(1–20), 1MJC(25–70):X(21–66). Comparisons to the NMR structure of *E. coli* CspA, used model 1 of the PDB entry 3MEF (Feng et al. 1998), which is identified as the best representative conformer in the PDB file header. Calculations of hydrogen bond R_{NO} distances from conformational HN chemical shifts (Eq. 3) were based on NMR assignments for the native (Feng et al. 1998; BMRB 4269) and urea denatured (BMRB 4108) forms of *E. coli* CspA.

Calibration of the R_{NO} dependence of the HN chemical shift

Equations 3 and 4 were derived from a database of diamagnetic proteins solved at better than 1.00-Å resolution, for which HN resonance assignments are available in the BMRB database (Seavey et al. 1991). The four proteins used for the analysis were BPTI (PDB: 5PTI, BMRB: 485), crambin (PDB: 1EJG, BMRB: 4509), cutinase (PDB: 1CEX, BMRB: 4104), and superoxide dismutase (PDB: 1MEF, BMRB: 4202). Hydrogen atoms were added to the X-ray structures as previously described. Native-state HN chemical shifts were corrected for “random coil” values (Wishart et al. 1995). An additional correction was included for ring current contributions to the native state chemical shift. The ring current shift calculations were done with the program MOLMOL 2K.1 (Koradi et al. 1996), using the Johnson-Bovey (1958) method with a distance cutoff of 5 Å. Ring current corrections were included only for the derivation of Equations 3 and 4, and not for subsequent calculations of R_{NO} distances from HN chemical shifts.

Analysis of ribonuclease A structures

X-ray structures of RNase A used for the analysis of the crystallographic resolution dependence of structural similarity were identified from the FSSP database of protein structural alignments (Holm and Sander 1996). Only PDB entries with 100% sequence identity to wild-type bovine RNase A (%IDE = 100) and only structures that could be aligned over the entire length of the RNase A sequence (LALI = 124) were selected. The highest resolution X-ray structure of any ribonuclease A (PDB: 1DY5, 0.87 Å), is that of a mutant in which Asn67 is replaced by an isoaspartyl residue (Cappaso et al. 1996), and this structure was excluded from the database. HN chemical shift-derived R_{NO} hydrogen bond distances are in good agreement with crystallographic R_{NO} distances (RMSD = 0.126 Å). The Cα atom RMSD value of 0.74 Å to the 7RSA structure is higher than that of any of the wild-type RNase A structures. The high global RMSD is dominated by differences in the loop delimited by the 65–72 disulphide bond, which contains the site of the Asn67 mutation (Cappaso et al. 1996). Structures representing complexes with large molecules (ribonuclease inhibitor protein, antibodies) were excluded from analysis, as well as structures based on combined neutron and X-ray diffraction data. The series of PDB files 1RAT to 9RAT, which are part of a crystallographic temperature dependence study, were also excluded from analysis. The resulting subset of 25 RNase A X-ray structures correspond to PDB entries: 1BEL, 1RBJ, 1RBN, 1RBX, 1RCA, 1RCN, 1RHA, 1RHB, 1RNC, 1RND, 1RNM, 1RNQ, 1RNW, 1RNX, 1RNY, 1RNZ, 1ROB, 1RPF, 1RPG, 1RPH, 1RTA, 1RTB, 1RUV, 3RN3, and 7RSA. An additional eight PDB entries (1AFK, 1AFL, 1AFU, 1EOS, 1QHC, 1RBB, 1XPS, and 1XPT) have two molecules in the asymmetric unit of the crystal. In these cases, structural parameters were averaged over the two molecules in the asymmetric unit. The 1AGI structure was used for comparisons between RNase A and angiogenin.

$\Delta(\delta_{HN})$ -derived hydrogen bond R_{NO} distances (Eq. 3) for RNase A and angiogenin were calculated using the respective native state NMR assignments (Reisdorf et al. 1994; Shimotakahara et al. 1997), and a “random coil” chemical shift database (Wishart et al. 1995). Superposition of Cα atoms in RNase A structures (residues 3–121) and calculations of pair-wise Cα coordinate RMSDs were performed with the program Insight 2000. The CE program (Shindyalov and Bourne 1998) was used to align RNase A structures onto the 1AGI angiogenin structure.

Acknowledgments

We thank Klara Rathgeb-Szabo for preparing the CspA samples used for NMR spectroscopy. NMR studies were carried out in the National Magnetic Resonance Facility at Madison, which is supported by grants from the NIH Biomedical Research Technology Program (RR02301), the NSF Biological Instrumentation Program (DMB-8415048), the NSF Shared Instrumentation Program (RR02781), the U.S. Department of Agriculture, and the University of Wisconsin.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Acharya, K.R., Shapiro, R., Riordan, J.F., and Vallee, B.L. 1995. Crystal structure of bovine angiogenin at 1.5-Å resolution. *Proc. Natl. Acad. Sci.* **92**: 2943–2953.
- Alexandrescu, A.T. and Rathgeb-Szabo, K. 1999. An NMR investigation of solution aggregation reactions preceding the misassembly of acid-denatured cold shock protein A into fibrils. *J. Mol. Biol.* **291**: 1191–1206.
- Alexandrescu, A.T., Dames, S.A., and Wiltsccheck, R. 1996. A fragment of staphylococcal nuclease with an OB-fold structure shows hydrogen-exchange protection factors in the range reported for "molten globules." *Protein Sci.* **5**: 1942–1946.
- Alexandrescu, A.T., Jaravine, V.A., Dames, S.A., and Lamour, F.P. 1999. NMR hydrogen exchange of the OB-fold protein LysN as a function of denaturant: The most conserved elements of structure are the most stable to unfolding. *J. Mol. Biol.* **289**: 1041–1054.
- Bai, Y., Sosnick, T.R., Mayne, L., and Englander, S.W. 1995. Protein folding intermediates: hydrogen-state hydrogen exchange. *Science* **269**: 192–197.
- Berman, H.M., Westbrook, J., Zeng, F., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The protein data bank. *Nucleic Acid Res.* **28**: 235–242. {<http://www.rcsb.org/pdb/>}
- Budisa, N., Steipe, B., Demange, P., Eckersorn, C., Kellermann, J., and Huber, R. 1995. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur. J. Biochem.* **230**: 788–796.
- Capasso, S., Di Donato, A., Esposito, L., Sica, F., Sorrentino, G., Vitagliano, L., Zagari, A., and Mazzarella, L. 1996. Deamidation in proteins: The crystal structure of bovine pancreatic ribonuclease with an isoaspartyl residue at position 67. *J. Mol. Biol.* **257**: 492–496.
- Chatterjee, S., Jiang, W., Emerson, S.D., and Inouye, M. 1993. The backbone structure of the major cold-shock protein CS7.4 of *Escherichia coli* in solution includes extensive β -sheet structure. *J. Biochem.* **114**: 663–669.
- Cordier, F. and Grzesiek, S. 1999. Direct observation of hydrogen bonds in proteins by interresidue $^3\text{J}_{\text{NC}'}$ scalar couplings. *J. Am. Chem. Soc.* **121**: 1601–1602.
- Cordier, F., Wang, C., Grzesiek, S., and Nicholson, L.K. 2000. Ligand-induced strain in hydrogen bonds of the c-Src SH3 domain detected by NMR. *J. Mol. Biol.* **304**: 497–505.
- Cornilescu, G., Hu, J.S., and Bax, A. 1999b. Identification of the hydrogen bonding network in a protein by scalar couplings. *J. Am. Chem. Soc.* **121**: 2949–2950.
- Cornilescu, G., Ramirez, B.E., Frank, M.K., Clore, G.M., Gronenborn, A.M., and Bax, A. 1999a. Correlation between $^3\text{J}_{\text{NC}'}$ and hydrogen bond length in proteins. *J. Am. Chem. Soc.* **121**: 6275–6279.
- Creighton, T.E. 1993. *Proteins: Structures and molecular properties*, 2nd ed. W.H. Freeman and Company, New York.
- Deisenhofer, J. and Steigemann, W. 1975. Crystallographic refinement of the structure of bovine pancreatic trypsin at 1.5 Å resolution. *Acta Crystallogr.* **B31**: 238–250.
- Dill, K.A. 1990. Dominant principles in protein folding. *Biochemistry* **29**: 7133–7155.
- Feng, W., Tejero, R., Zimmerman, D.E., Inouye, M., and Montelione, G.T. 1998. Solution NMR structure and backbone dynamics of the major cold shock protein (CspA) from *Escherichia coli*: Evidence of conformational dynamics in the single-stranded RNA-binding site. *Biochemistry* **37**: 10881–10896.
- Holm, L. and Sander, C. 1996. Mapping the protein universe. *Science* **273**: 595–602. {<http://www2.ebi.ac.uk/dali/fssp/>}
- Jaravine, V.A., Alexandrescu, A.T., and Grzesiek, S. 2001. Observation of the closing of individual hydrogen bonds during TFE-induced helix formation in a peptide. *Protein Sci.* **10**: 943–950.
- Jaravine, V.A., Rathgeb-Szabo, K., and Alexandrescu, A.T. 2000. Microscopic stability of cold shock protein A examined by NMR native state hydrogen exchange as a function of urea and trimethylamine N-oxide. *Protein Sci.* **9**: 290–301.
- Johnson, C.E. and Bovey, F.A. 1958. Calculation of nuclear magnetic resonance spectra of aromatic hydrocarbons. *J. Chem. Phys.* **29**: 1012–1014.
- Kay, L.E., Ikura, M., Zhu, G., and Bax, A. 1991. Four-dimensional heteronuclear triple-resonance NMR of isotopically enriched proteins for sequential assignment of backbone atoms. *J. Magn. Reson.* **91**: 422–428.
- Kay, L.E., Xu, G.Y., and Yamazaki, T. 1994. Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. *J. Magn. Reson.* **A109**: 129–133.
- Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**: 51–55.
- Kuriyan, J., Karplus, M., and Petsko, G.A. 1987. Estimation of uncertainties in x-ray refinement results by use of perturbed structures. *Proteins* **2**: 1–12.
- Mueller, U., Perl, D., Schmid, F.X., and Heinemann, U. 2000. Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. *J. Mol. Biol.* **297**: 975–988.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**: 536–540. {<http://scop.mrc-lmb.cam.ac.uk/scop/>}
- Orban, J., Alexandrer, P., Bryan, P., and Kahare, D. 1995. Assessment of stability differences in the protein G B1 and B2 domains from hydrogen-deuterium exchange: Comparison with correlation data. *Biochemistry* **34**: 15291–15300.
- Ottiger, M. and Bax, A. 1998. Determination of relative N-H^{N} , $\text{N-C}'$, $\text{C}^{\alpha}\text{-C}'$, and $\text{C}^{\alpha}\text{-H}^{\alpha}$ effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**: 12334–12341.
- Pan, Y. and Briggs, M.S. 1992. Hydrogen exchange in native and alcohol forms of ubiquitin. *Biochemistry* **31**: 11405–11412.
- Reisdorf, C., Abergel, D., Bontems, F., Lallemand, J.-Y., and Decottignies, J.-P. 1994. Proton resonance assignments and secondary structure of bovine angiogenin. *Eur. J. Biochem.* **224**: 811–822.
- Sander, C. and Schneider, R. 1991. Database of homology-derived protein structures. *Proteins* **9**: 56–68. {<http://www.sander.ebi.ac.uk/hssp/>}
- Santorio, J., Gonzalez, C., Bruix, M., Neira, J.L., Nieto, J.L., Herranz, J., and Rico, M. 1993. High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **229**: 722–734.
- Scheurer, C. and Brüschweiler, R. 1999. Quantum-chemical characterization of nuclear spin-spin coupling across hydrogen bonds. *J. Am. Chem. Soc.* **121**: 8661–8662.
- Schindelin, H., Jiang, W., Inouye, M., and Heinemann, U. 1994. Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **91**: 5119–5123.
- Schindelin, H., Marahiel, M.A., and Heinemann, U. 1993. Universal nucleic acid-binding domain revealed by the crystal structure of the *B. subtilis* major cold shock protein. *Nature* **364**: 164–168.
- Seavey, B.R., Farr, E.A., Westler, W.M., and Markley, J.L. 1991. A relational database for sequence-specific protein NMR data. *J. Biomol. NMR* **1**: 217–236. {<http://www.bmrwisc.edu/>}
- Shimotakahara, S., Rios, C.B., Laity, J.H., Zimmerman, D.E., Scheraga, H.A., and Montelione, G.T. 1997. NMR structural analysis of an analog of an intermediate formed in the rate-determining step of one pathway in the oxidative folding of bovine ribonuclease A: Automated analysis of ^1H , ^{13}C , and ^{15}N resonance assignments for wild-type and [C65S, C72S] mutant forms. *Biochemistry* **36**: 6915–6929.
- Shindyalov, I.N. and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**: 739–747. {<http://cl.sdsc.edu/ce.html>}
- Wang, Y.-X., Jacob, J., Cordier, F., Wingfield, P., Stahl, S.J., Lee-Huang, S., Torchia, D., Grzesiek, S., and Bax, A. 1999. Measurement of $^3\text{J}_{\text{NC}'}$ connectivities across hydrogen bonds in a 30 kDa protein. *J. Biomol. NMR* **14**: 181–184.
- Wagner, G., Pardi, A., and Wüthrich, K. 1983. Hydrogen bond length and ^1H NMR chemical shifts in proteins. *J. Am. Chem. Soc.* **105**: 5948–5949.

- Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. 1995. ^1H , ^{13}C , and ^{15}N random coil NMR chemical shifts of the common amino acids: Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5**: 67–81.
- Wishart, D.S., Sykes, B.D., and Richards, F.M. 1991. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**: 311–333.
- Wlodawer, A., Svensson, L.A., Sjolín, L., and Gilland, G.L. 1988. Structure of phosphate-free ribonuclease A refined at 1.26 Å. *Biochemistry* **27**: 2705–2717.
- Yamazaki, T., Nicholson, L.K., Torchia, D.A., Wingfield, P., Stahl, S.J., Kaufman, J.D., Eyermann, C.J., Hodge, C.K., Lam, P.Y.S., Ru, Y., et al. 1994. NMR and x-ray evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the proteinase and a nonpeptide cyclic urea based inhibitor. *J. Am. Chem. Soc.* **116**: 10791–10792.
- Zhu, G. and Bax, A. 1990. Improved linear prediction for truncated signals of known phase. *J. Magn. Reson.* **90**: 405–410.
- Zhu, G. and Bax, A. 1992. Two-dimensional linear prediction for signals truncated in both dimensions. *J. Magn. Reson.* **98**: 192–199.