Solution NMR Structure and Backbone Dynamics of the Major Cold-Shock Protein (CspA) from *Escherichia coli*: Evidence for Conformational Dynamics in the Single-Stranded RNA-Binding Site[†,‡]

Wenqing Feng,[§] Roberto Tejero,[§,||] Diane E. Zimmerman,[§] Masayori Inouye,[⊥] and Gaetano T. Montelione*[,§]

Center for Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854-5638, Departamento de Quimica Fisica, Universidad de Valencia, Avenida Dr. Moliner, 50, 46100 Burjassot (Valencia), Spain, and Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854-5638

Received February 3, 1998; Revised Manuscript Received May 19, 1998

ABSTRACT: The major cold-shock protein (CspA) from *Escherichia coli* is a single-stranded nucleic acid-binding protein that is produced in response to cold stress. We have previously reported its overall chain fold as determined by NMR spectroscopy [Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S. D., Inouye, M., and Montelione, G. T. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5114−5118]. Here we describe the complete analysis of ^{1}H, ^{13}C, and ^{15}N resonance assignments for CspA, together with a refined solution NMR structure based on 699 conformational constraints and an analysis of backbone dynamics based on ^{15}N relaxation rate measurements. An extensive set of triple-resonance NMR experiments for obtaining the backbone and side chain resonance assignments were carried out on uniformly ^{13}C- and ^{15}N-enriched CspA. Using a subset of these triple-resonance experiments, the computer program AUTOASSIGN provided automatic analysis of sequence-specific backbone N, C^{α}, C', H^{N}, H^{α}, and side chain C^{β} resonance assignments. The remaining ^{1}H, ^{13}C, and ^{15}N resonance assignments for CspA were then obtained by manual analysis of additional NMR spectra. Dihedral angle constraints and stereospecific methylene H^{β} resonance assignments were determined using a new conformational grid search program, HYPER, and used together with longer-range constraints as input for three-dimensional structure calculations. The resulting solution NMR structure of CspA is a well-defined five-stranded β-barrel with surface-exposed aromatic groups that form a single-stranded nucleic acid-binding site. Backbone dynamics of CspA have also been characterized by ^{15}N T_1, T_2, and heteronuclear ^{15}N−^{1}H NOE measurements and analyzed using the extended Lipari−Szabo formalism. These dynamic measurements indicate a molecular rotational correlation time τ_m of 4.88 ± 0.04 ns and provide evidence for fast time scale (τ_e < 500 ps) dynamics in surface loops and motions on the microsecond to millisecond time scale within the proposed nucleic acid-binding epitope.

Many organisms have specific physiological mechanisms for acclimating to thermal stress. These include changes in intracellular protein production and function in response to both "heat shock" (for a review, see ref *1*) and "cold shock" (for recent reviews, see refs *2* and *3*). In *Escherichia coli*, an environmental temperature downshift from 37 to 10−15 °C results in a lag period in cell growth, during which "cold-shock proteins" believed to provide cell-growth adaptation under cold-stress conditions are induced (*2*, *4*). The most abundant of these, the major cold-shock protein CspA,[1] is transiently produced. Expression of CspA peaks at more

[†] This work was supported by grants from the National Institutes of Health (GM-47014) and the National Science Foundation (MCB-9407569), a National Science Foundation Young Investigator Award (MCB-9357526), and a Camille Dreyfus Teacher-Scholar Award. Computing facilities were provided by a grant from the W. M. Keck Foundation. R.T. also acknowledges travel support from the University of Valencia.

[‡] A listing of ^{1}H, ^{15}N, and ^{13}C resonance assignments for CspA have been submitted to the BioMagRes Database (Madison, WI). Atomic coordinates for these 16 NMR structures of CspA, together with listings of experimental constraints used in this study, have been deposited in the Brookhaven Protein Data Base (access code 2mef).

* Address correspondence to Prof. G. T. Montelione, CABM-Rutgers University, 679 Hoes Ln., Piscataway, NJ 08854. Phone: (732) 235-5321. Fax: (732) 235-4850. E-mail: guy@nmrlab.cabm.rutgers.edu.

[§] Rutgers University.

[||] Universidad de Valencia.

[⊥] University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School.

[1] Abbreviations: 2D, two-dimensional; 3D, three-dimensional; CF, correction factor; CSA, chemical shift anisotropy; CspA, cold-shock protein A from *E. coli*; CspB, cold-shock protein B from *B. subtilis*; DIANA, program for computing 3D structures of polypeptides and proteins from NMR data; FID, free induction decay; GS, generic spin system; ^{1}H−^{2}H exchange, amide hydrogen−deuterium exchange; HYPER, grid search program for obtaining ranges of ϕ, ψ, and χ^{1} dihedral angles in polypeptides; IPTG, isopropyl β-thiogalactopyranoside; NOE, nuclear Overhauser effect; HNOE, heteronuclear NOE; NOESY, NOE spectroscopy; OB-fold, oligonucleotide- and/or oligosaccharide-binding fold; PFG, pulsed-field gradient; pH*, uncorrected pH meter reading of a $^{2}H_2O$ solution; rmsd, root-mean-square deviation; RNP, ribonucleoprotein; SSE, sum-squared errors; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; ss, single-stranded; T_1 and R_1, longitudinal relaxation time and rate, respectively; T_2 and R_2, transverse relaxation time and rate, respectively; TOCSY, total correlation spectroscopy.

S0006-2960(98)00269-4 CCC: $15.00
Published on Web 07/17/1998

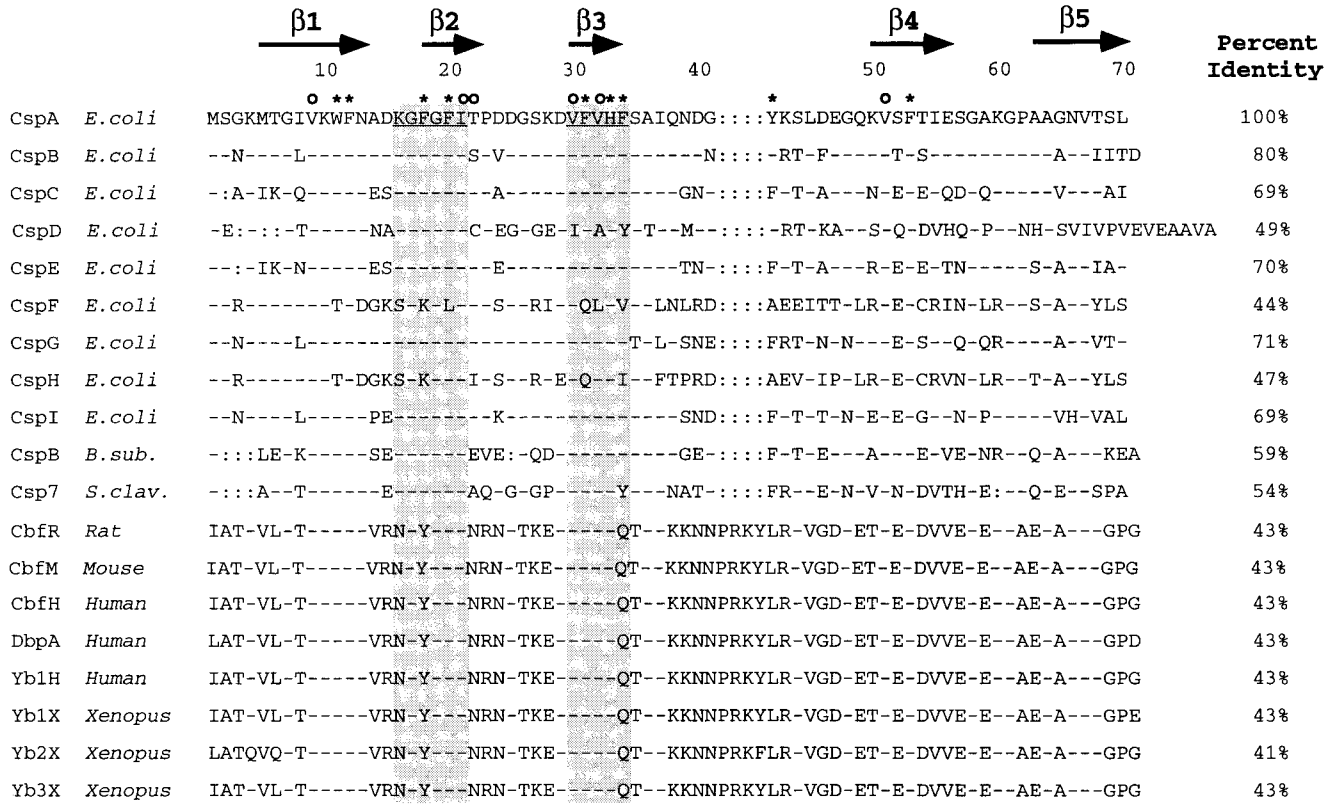

FIGURE 1: Sequence alignments for members of the CspA family, together with the "cold-shock domains" from eukaryotic Y-box proteins. Identical residues are denoted by dashed lines (−), and proposed gaps in the sequence alignments are indicated with colons (:). The RNP1 and RNP2 sequence motifs are underlined in the CspA sequence, and the corresponding residues are shaded for all the sequences. Nine highly conserved aromatic residues are indicated with asterisks (*). The locations of six residues with the slowest-exchanging amides are labeled with open circles. Five β-strands are denoted by arrows.

than 10% of total cellular protein production during the first hour following the temperature downshift and then diminishes to basal levels as the cells acclimate to growth at the lower temperature (*5*). The gene encoding the 70-amino acid CspA protein was the first of the so-called cold-shock protein gene family to be identified in *E. coli* (*5*); recently, eight other members of the *cspA* gene family have been found (*3*, *6*). These *E. coli* proteins (Figure 1) CspB, -C, -D, -E, -F, -G, -H, and -I have sequences 44−80% identical to that of CspA but are not all produced in response to cold stress. Interestingly, these cold-shock proteins also are significantly homologous with nucleic acid-binding domains of eukaryotic Y-box proteins (*7*, *8*) that are involved in transcriptional regulation and RNA packaging. Indeed, this "cold-shock domain" is one of the most highly conserved protein sequences in nature, exhibiting greater than 43% sequence identity in homologues ranging from bacteria to humans (Figure 1).

CspA is a Greek-key β-barrel protein that appears to function as a single-stranded nucleic acid-binding protein (*9*−*13*). Key clues to this biochemical function have come from structural analysis. The low-resolution solution NMR structure (*9*) and a 2.0 Å X-ray crystal structure (*10*) of *E. coli* CspA have been described. Similar chain folds have also been described for the cold-shock protein CspB from *Bacillus subtilis* (*14*, *15*). These CspA and CspB structures belong to the "OB-fold family" (*16*, *17*) that also includes several oligonucleotide- and/or oligosaccharide-binding proteins. The structural analysis also revealed putative RNP1 and RNP2 sequence motifs that are commonly found in ssRNA-binding proteins (*18*, *19*). These structural features first suggested that CspA may function as a single-stranded DNA and/or RNA binding protein (*9*, *10*, *14*). This hypothesis has been verified by chemical shift perturbation analysis of complexes between CspA and ssDNA (*9*), ssRNA-binding gel-shift assays (*12*, *13*), Trp fluorescence quenching studies (*20*), and site-directed mutagenesis of the ssDNA-binding function of *E. coli* CspA (*20*) and *B. subtilis* CspB (*11*).

The precise role of CspA's single-stranded nucleic acid-binding function in the cell biology of the cold-shock response is still not fully understood (for reviews, see refs *2* and *3*). Most efforts to understand its cellular function in vivo focus on its now well-characterized in vitro ssRNA-binding activities (*3*, *12*, *13*). The wide-spread occurrence of this nucleic acid recognition module in a variety of genes from many species suggests that it is a fundamental structural motif and makes CspA a good target for detailed biophysical studies aimed at better characterizing and understanding the atomic basis for its molecular recognition properties.

As part of our ongoing efforts to better understand the mechanisms of nucleic acid recognition in this system, we have undertaken a detailed NMR study of the structure and dynamics of CspA in the absence of nucleic acid ligands. In this paper, we present nearly complete ^{1}H, ^{13}C, and ^{15}N resonance assignments for *E. coli* CspA together with the refined solution NMR structure. Relative to our initial description of the overall chain fold of CspA (*9*), these refined coordinates provide a better characterization of the β-sheet core, of the backbone structure of the interstrand polypeptide loops, and of many side chain conformations in

the solution structure. In addition, ^{15}N relaxation measurements reveal that portions of the interstrand loops of CspA are mobile on the sub-nanosecond time scale. Relaxation measurements also provide evidence for conformational dynamics on the millisecond to microsecond time scale of atoms that contribute to the nucleic acid-binding epitope.

MATERIALS AND METHODS

Protein Purification. Recombinant CspA was expressed in *E. coli* and purified using an optimized protocol based on a previously described purification procedure (*21*). *E. coli* strain BL21(DE3) cells containing the gene for T7 RNA polymerase under control of the isopropyl β-thiogalactopyranoside (IPTG) inducible *lacZ* promoter were transformed with plasmid pET11-CspA, in which the expression of the *cspA* gene was controlled by the T7 RNA polymerase promoter (*22*). A single colony from these transformed cells was then isolated and used to inoculate 200 mL of M9 casamino acid media (*23*). The cell culture was grown overnight at 37 °C and then diluted into 4 L of M9 casamino acid media and grown at 37 °C for about 2 h to the early logarithmic phase (i.e., OD_{600} of 0.4−0.6). Overexpression of CspA was then induced by 1 mM IPTG, and the cells were grown at 37 °C for another 1−1.5 h. The cells were then centrifuged at 4500*g* for 30 min, washed with ∼5 mL of 20 mM sodium phosphate buffer at pH 7.0, and ruptured by French press at 14 000 psi. The resulting suspension was then centrifuged at 1500*g* for 10 min; the pellet was removed, and the supernatant was centrifuged at 100000*g* for 2.5 h at 4 °C. The resulting supernatant was collected, and solid ammonium sulfate was slowly added while the mixture was stirred until 90% saturation. This suspension was then centrifuged at 16000*g* for 20 min, the pellet dissolved in ∼20 mL of 10 mM Bis-Tris buffer at pH 7.0, and the resulting solution dialyzed against 2 × 4 L of the same buffer. This protein solution was next diluted to a concentration of ∼3 mg/mL for further purification by FPLC using Pharmacia Q-Sepharose and Mono-S column chromatography. The solution was first applied to a Q-Sepharose column (2.5 cm × 13.8 cm) equilibrated in 10 mM Bis-Tris (pH 7.0) and then eluted at 1 mL/min using a 400 mL gradient of 0 to 300 mM NaCl. The CspA-containing fractions were pooled, dialyzed against 2 × 4 L of 50 mM succinic acid at pH 4.0, and then loaded onto a Mono-S FPLC column equilibrated with 50 mM succinic acid at pH 4.0 and eluted at 2 mL/min using a 400 mL gradient of 0 to 500 mM NaCl. The major protein fractions were pooled and prepared for NMR studies as described below.

Uniformly ^{15}N- and ^{15}N-,^{13}C-enriched CspA samples were produced using the same procedure with M9 minimal media without casamino acids. The only sources of nitrogen and carbon in this case were [^{15}N]ammonium chloride and unenriched or [U-^{13}C]glucose. Each liter of M9 media contained 1 g of $^{15}NH_4Cl$ and 4 g of unenriched glucose or 2 g of [U-^{13}C]glucose. The final yield of purified isotope-enriched protein was approximately 8 mg per liter of cell culture.

Protein Characterization. These purified CspA samples exhibited a single band (>99%) on SDS−polyacrylamide gel electrophoresis. In addition, Superdex 75 gel filtration column chromatography was carried out to examine the association state of the protein in 50 mM K_3PO_4 and 200 mM NaCl at pH 6.0 and room temperature. In these chromatograms, CspA exhibits a single peak with an apparent molecular mass of ∼7 kDa, demonstrating that under the solution conditions used for these NMR studies the protein exists as a monomer. Laser desorption mass spectroscopic analysis showed molecular ions with masses corresponding to both the whole 70-amino acid protein and the protein lacking its N-terminal methionine residue, in ratios of ∼1:4. The corresponding masses were 7400 and 7269, 7485 and 7353, and 7816 and 7679 Da for unenriched, [^{15}N]-, and [^{13}C,^{15}N]CspA, respectively; the corresponding percentages of isotopic enrichment were ∼90 and ∼98% for [^{15}N]- and [^{13}C,^{15}N]CspA, respectively. Amino acid composition analysis of unenriched CspA samples demonstrated molar ratios for each amino acid that match values expected from the primary sequence, except that for Met which was consistently lower than expected. A molar extinction coefficient ϵ_{280} of 8600 M^{-1} cm^{-1} was calculated on the basis of quantitative amino acid analysis.

Trace proteolytic or enzymatic contamination in CspA samples produced with an earlier version of the purification protocol (*21*) resulted in the slow appearance of new resonances in the NMR spectra collected at 30 °C. This chemical instability of some CspA samples was monitored in two ways. In freshly prepared samples, the single Trp-11 indole proton gives rise to a well-resolved peak in the one-dimensional (1D) spectrum at ∼10.4 ppm, and peaks appearing at ∼10.0 ppm provide a means of monitoring the time-dependent changes in the sample homogeneity. In addition, this chemical heterogeneity can be observed by Superdex 75 gel-filtration chromatography with detection at 214 nm at room temperature. Degradation of CspA results in species with both shorter and longer gel-filtration elution times compared with those of native CspA. The improved purification scheme described above produces CspA samples that exhibit good chemical stability in these assays.

NMR Sample Preparation. Pooled column fractions of purified CspA were prepared for NMR analysis by dialysis against a buffer containing 50 mM potassium phosphate and 1 mM NaN_3 at pH 6.0, and then ultrafiltration to protein concentrations of 1−4 mM using Centricon-3 devices. After the protein solution was concentrated, the appropriate amounts of a stock EDTA solution and ^2H_2O were added. The pH was checked and, if necessary, readjusted to 6.0. The final NMR sample contained 50 mM potassium phosphate, 1 mM NaN_3, and 0.1 mM EDTA in 90% ^1H_2O/10% ^2H_2O at pH 6.0 ± 0.1. ^2H_2O samples were prepared similarly by first dialyzing the protein against distilled water to remove salts, lyophilizing the protein twice from ^2H_2O solvent, and then dissolving it in ^2H_2O buffer containing 50 mM potassium phosphate, 1 mM NaN_3, and 0.1 mM EDTA at pH* 6.0. Volumes of ∼250 μL were then transferred into 5 mm susceptibility-matched Shigemi NMR tubes (Shigemi Co., Ltd) for NMR measurements.

NMR Spectroscopy. All of the NMR data sets were collected at 30 °C (except where noted) on a Varian Unity 500 spectrometer equipped with three channels and a fourth frequency synthesizer for carbonyl decoupling. Schematic diagrams of the 3D NMR pulse sequences used in this work (Figures S1−9 in the Supporting Information) are available. All multidimensional NMR experiments for obtaining ^1H,

^{13}C, and ^{15}N assignments were collected using pulsed-field gradients, constant-time evolution of ^{13}C and ^{15}N, and sensitivity enhancement. In some of the 3D experiments, "C−H"- and "C−C"-type tunings were incorporated to obtain phase information useful for spin-system-type identification (*24*−*26*). The number of complex data points, spectral widths, number of scans per increment in the indirect dimensions, additional phase labeling delays or mixing times, and corresponding literature references are described for all of the data sets used in this study in Table S1 in the Supporting Information. Carrier positions in all of the experiments were set to 116.9 ppm for ^{15}N, 174.3 ppm for ^{13}CO, 54.7 ppm for $^{13}C^{\alpha}$, 39.0 ppm for $^{13}C^{\alpha}/^{13}C^{\beta}$, and 8.62 ppm for $^{1}H^{N}$. Backbone ^{15}N and narrow-band $^{13}C'$ decoupling was achieved using GARP-1 (*27*) and SINC waveforms, respectively. In all of the experiments, quadrature detection in the indirectly detected dimensions was carried out by combining N- and P-type spectra selected with pulsed-field gradients (*28, 29*).

All of the spectra were processed with VNMR (Varian Associates) processing software. The FIDs were zero-filled to 512, 256, and 1024 complex points in the t_1, t_2, and t_3 dimensions, respectively, and multiplied with appropriately shifted Gaussian window functions. The resulting Fourier-transformed frequency domain data sets were imported into the NMRCompass program (Molecular Simulations, Inc.) for peak picking. The threshold for automatic peak picking with NMRCompass was set so that the total number of the picked peaks in each spectrum was about 10% greater than the expected numbers. Each 3D plane was then examined manually on a graphics display and edited to remove obvious spectral artifacts (e.g., sinc wiggles). No additional manipulations of the peak lists were done prior to automated analysis. Proton chemical shifts were referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonic acid. ^{13}C and ^{15}N chemical shifts were referenced indirectly using appropriate ratios of ^{13}C:^{1}H and ^{15}N:^{1}H gyromagnetic ratios (*30*).

Automated Analysis of Resonance Assignments Using AUTOASSIGN. AUTOASSIGN is an expert system providing automated analysis of backbone and C^{β} resonance assignments from triple-resonance spectra of ^{13}C- and ^{15}N-enriched proteins (*31*, *32*). The input for AUTOASSIGN includes peak lists from 2D $^{15}N-^{1}H^{N}$ HSQC and 3D HNCO spectra along with peak lists from three intraresidue (CANH, CBCANH, and HCANH) and three interresidue [CA(CO)NH, CBCA(CO)NH, and HCA(CO)NH] experiments which correlate the C^{α}, C^{β}, and H^{α} resonances of residues *i* and *i* − 1, respectively, with the backbone amide $^{15}N-^{1}H^{N}$ resonances of residue *i*. Details of the AUTOASSIGN software used here have been described elsewhere (*32*). The program outputs a list of sequence-specific backbone ^{1}H, ^{13}C, ^{15}N, and side chain $^{13}C^{\beta}$ resonance assignments.

Manual Analysis of Side Chain Resonance Assignments. Side chain resonance assignments were obtained using PFG HCCNH−TOCSY and PFG HCC(CO)NH−TOCSY (*24*, *33*−*35*) and homonuclear TOCSY experiments recorded with multiple mixing times of 22, 36, 45, 54, 71, and 90 ms (*36*). Side chain proton and carbon resonance assignments were easily identified from these spectra once the backbone resonance assignments were obtained with AUTOASSIGN.

Calibration of NOE Intensities. Interatomic distance constraints were derived from the following three NOESY data sets: 2D NOESY (*37*) and 3D ^{15}N-edited NOESY−HSQC (*38*) spectra recorded with a mixing time τ_m of 60 ms of a CspA sample dissolved in 90% H_2O/10% $^{2}H_2O$ and a 2D NOESY spectrum recorded with a mixing time τ_m of 50 ms of a sample dissolved in 100% $^{2}H_2O$. The intensity loss due to differential relaxation during the HSQC period of the NOESY−HSQC experiment was corrected by introducing correction factors (CF) for each $^{15}N-^{1}H^{N}$ cross-peak; CF is defined as the individual peak intensity divided by the maximum intensity of a well-resolved peak in an HSQC spectrum (I_i/I_{max}). In cases of overlapping NOESY cross-peaks, CF values of 1.0 were assigned as the most conservative correction. Each NOESY cross-peak intensity in the 3D NOESY−HSQC spectrum was then corrected by dividing its intensity by the CF value of the corresponding $^{15}N-^{1}H^{N}$ HSQC correlation peak. The resulting cross-peak intensities were then converted into interproton distance constraints using appropriate distance calibrations. NOE intensities in H_2O spectra were calibrated using the sequential $H^{\alpha}_i-H^{N}_{i+1}$ NOEs and intraresidue $H^{\alpha}_i-H^{N}_i$ NOEs in β-sheets of CspA, which for ideal β-sheet conformations correspond to distances of 2.2 and 2.9 Å, respectively. These calibrations made in H_2O NOESY spectra were then transferred to $^{2}H_2O$ NOESY spectra by comparing intensities of 16 well-resolved NOEs between aliphatic protons. Distance constraints between H^{N}, H^{α}, and H^{β} atoms were derived from this calibration method, while all other NOESY-derived distance constraints were assigned upper-bound values of 5.0 Å.

Vicinal Coupling Constant Measurements. Relative values of methylene $^{3}J(H^{\alpha}-H^{\beta})$ coupling constants were determined from the relative intensities of $H^{\alpha}-H^{\beta}$ cross-peaks in a TOCSY spectrum recorded with a mixing time of 36 ms. Vicinal $^{3}J(H^{N}-H^{\alpha})$ coupling constants were estimated using a series of 2D $^{15}N-^{1}H^{N}$ HSQC-J experiments (*39*, *40*). The values of defocusing delays τ were set to 23, 25, 28, 31, 36, 42, 50, 62.5, and 83 ms, corresponding to nulls in the transfer function for $^{3}J(H^{N}-H^{\alpha})$ coupling constants of 11, 10, 9, 8, 7, 6, 5, 4, and 3 Hz, respectively.

Stereospecific $C^{\beta}H_2$ Methylene Proton Assignments Using HYPER. Stereospecific assignments of methylene H^{β}s were made by analysis of local NOE and vicinal coupling constant data using the program HYPER (*41*). HYPER is a conformational grid search program used for determining stereospecific $C^{\beta}H_2$ methylene proton assignments and for defining the ranges of dihedral angles ϕ, ψ, and χ^1 that are consistent with the local experimental NMR data for each amino acid in the polypeptide. The program compares experimental NMR data with analytical functions describing the ϕ, ψ, and χ^1 dihedral angle dependence of interproton distances and Karplus relationships of homo- and heteronuclear coupling constants. HYPER determines the allowed conformational spaces from the intersection of dihedral angle ranges evaluated from each of the available intraresidue and sequential constraints. Random ϕ, ψ, and χ^1 values generated within this "allowed conformational space" are then used as the starting conformations for subsequent structure generation calculations. HYPER also provides stereospecific assignments of some methylene H^{β} resonances. First, it assigns the upfield H^{β} as $H^{\beta 2}$ and downfield H^{β} as $H^{\beta 3}$. Under this assumption, HYPER evaluates the consistency of all the local constraints available for a given amino acid residue. Next, it inverts the assumption, assigning upfield H^{β} as $H^{\beta 3}$

and downfield H^{β} as $H^{\beta 2}$. Stereospecific assignments are made if only one of these two assumptions results in conformational space(s) that satisfies all of the constraints.

Amide Hydrogen−Deuterium Exchange Rate Measurements. Samples of ^{15}N-enriched CspA were prepared for amide $^{1}H-^{2}H$ exchange rate measurements by first dialyzing against H_2O and then lyophilizing. Seven 2D $^{15}N-^{1}H^{N}$ HSQC spectra were then collected immediately after dissolving the sample. The resulting protein sample contained 50 mM potassium phosphate, 1 mM NaN_3, and 0.1 mM EDTA at pH* 6.0 in $^{2}H_2O$. Each spectrum was acquired in 22 min. The peak intensities $I(t)$ for each observed amide were fit using the Kaleidagraph program (Abelbeck Software) to a single-exponential decay, $I = a + b \exp(-kt)$, where k is the pseudo-first-order exchange rate constant.

DIANA Calculations. Structure generation calculations were carried out with the DIANA program (*42*, *43*), version 2.8 (TRIPOS, Inc.), using R8000 processors of a Silicon Graphics Onyx workstation. The program generates protein conformations by minimizing a variable target function. Contributions to this target function include upper- and lower-bound NOE-derived distance constraints, dihedral angle constraints, and hydrogen bond constraints identified from amide $^{1}H-^{2}H$ exchange measurements. For these DIANA calculations, the NOE-derived distances calibrated as described above were rounded up to the nearest 0.5 Å, and when stereospecific assignments were not available, pseudoatom corrections (*44*) were applied. The resulting distances were then used as upper-bound distance constraints for DIANA calculations. Starting conformations were generated randomly within the ϕ, ψ, and χ^1 ranges defined by the HYPER program. The minimization protocol consisted of three parts. First, relative weighting factors for upper-bound, explicit lower-bound, steric lower-bound, and dihedral angle constraints were set to 1.0, 1.0, 0.2, and 10.0 units, respectively. Sixteen minimization steps with increasing residue ranges ($i - j$) from 0 to 10 were applied using 500 iterations for each step. Second, the same protocol was repeated using 36 minimization steps of ~500 iterations each with residue ranges ($i - j$) from 0 to 70. Finally, the relative weighting factor for the steric constraints was increased from 0.2 to 1.0 units, and 10 sets of ~500 iterations of minimization were executed using all constraints. This last cycle of the protocol was repeated four times to effectively minimize the residual constraint violations.

Backbone ^{15}N Relaxation Measurements. Measurements of backbone ^{15}N T_1 and T_2 relaxation times and steady-state heteronuclear $^{15}N-^{1}H^{N}$ NOEs (HNOEs) were carried out using 2D proton-detected heteronuclear NMR spectroscopy on uniformly ^{15}N-enriched CspA at a spectrometer frequency of 500 MHz, as described previously (*45*, *46*) except that pulsed-field gradients were used in all of the experiments for heteronuclear coherence selection and solvent suppression. Recycle delays of 1.4 s were employed in both T_1 and T_2 experiments. Twelve T_1 experiments were recorded with delays T of 20, 50, 100 (two times), 200, 300 (two times), 400, 500 (two times), 700, and 900 ms, while ten T_2 experiments used delays T of 15, 29, 44, 73, 102 (two times), 146, 219, 292, and 365 ms. The cross-peak intensities $I(t)$ in the T_1 measurements were fit to three parameters (I_∞, I_0, and $R_1 = 1/T_1$)

$$I(t) = I_\infty - (I_\infty - I_0) \exp(-R_1 t) \quad (1)$$

and the T_2 measurements were fit to two parameters (I_0 and $R_2 = 1/T_2$)

$$I(t) = I_0 \exp(-R_2 t) \quad (2)$$

Uncertainties in the R_1 and R_2 curve fits were derived from Monte Carlo simulations using baseline noise values as an estimate of uncertainties in peak intensities (*47*). In each of the T_1 and T_2 spectra, the rms noise of intensity measurements was estimated from random baseline regions of six ω_2 traces.

Heteronuclear NOEs (HNOEs) were calculated from the ratios of cross-peak intensities in two spectra collected with (I_{sat}) and without (I_{eq}) broad-band amide proton saturation for a period of 5 s using GARP-1 (*27*)

$$\mathrm{HNOE} = \frac{I_{sat}}{I_{eq}} \quad (3)$$

A total recycle delay of 15 s was used in the HNOE measurements to ensure complete relaxation of water magnetization at the beginning of each scan (*45*). The uncertainty of the HNOE value, σ_i, measured for each residue i, was computed by propagation of the baseline white noise in HNOE spectra recorded with ($I_{b,i,sat}$) or without ($I_{b,i,eq}$) proton saturation (*48*).

$$\sigma_i = \sqrt{\left(\frac{I_{b,i,sat}}{I_{i,sat}}\right)^2 + \left(\frac{I_{b,i,eq}}{I_{i,eq}}\right)^2} \quad (4)$$

Analysis of Internal Motions. The relaxation properties of an amide ^{15}N nuclear spin in a diamagnetic protein are generally dominated by dipolar interactions with the directly bonded amide proton spin and contributions due to chemical shift anisotropy. With this assumption, relaxation parameters R_1, R_2, and HNOE can be expressed as linear combinations of the spectral density functions $J(\omega)$

$$R_1 = \frac{d^2}{4}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \quad (5)$$

$$R_2 = \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{6}[4J(0) + 3J(\omega_N)] + R_{ex} \quad (6)$$

$$\mathrm{HNOE} = 1 + \left(\frac{d^2}{4R_1}\right)\left(\frac{\gamma_H}{\gamma_N}\right)[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad (7)$$

in which

$$d = \frac{\mu_0 h \gamma_N \gamma_H}{8\pi^2}\left\langle \frac{1}{r_{NH}^3} \right\rangle$$

$$c = \frac{\omega_N}{\sqrt{3}}(\sigma_{||} - \sigma_\perp)$$

where μ_0 is the permeability of free space, h is Plank's constant, γ_H and γ_N are the gyromagnetic ratios of ^{1}H and

Table 1: Expressions of Spectral Density Functions for the Five Models

model	spectral density functions	optimized parameters
1	$J(\omega) = {}^2/_5[S^2\tau_m/(1+\omega^2\tau_m^2)]$	S^2
2[a]	$J(\omega) = {}^2/_5[S^2\tau_m/(1+\omega^2\tau_m^2) + (1-S^2)\tau_e'/(1+\omega^2\tau_e'^2)]$	S^2, τ_e
3	$J(\omega) = {}^2/_5[S^2\tau_m/(1+\omega^2\tau_m^2)]$	
	$R_{2(obs)} = R_2 + R_{ex}$	S^2, R_{ex}
4	$J(\omega) = {}^2/_5[S^2\tau_m/(1+\omega^2\tau_m^2) + (1-S^2)\tau_e'/(1+\omega^2\tau_e'^2)]$	
	$R_{2(obs)} = R_2 + R_{ex}$	S^2, τ_e, R_{ex}
5[b]	$J(\omega) = {}^2/_5[S^2\tau_m/(1+\omega^2\tau_m^2) + S_f^2(1-S_s^2)\tau_s'/(1+\omega^2\tau_s'^2)]$	S_f^2, S_s^2, τ_e

[a] $\tau_e' = \tau_m\tau_e/(\tau_m+\tau_e)$. [b] $\tau_s' = \tau_m\tau_s/(\tau_m+\tau_s)$; $S^2 = S_f^2S_s^2$.

^{15}N, respectively, the N−H bond length r_{NH} is 1.02 Å, ω_H and ω_N are the Larmor frequencies of ^{1}H and ^{15}N, respectively, and the chemical shift anisotropy $\sigma_{||} - \sigma_\perp = -160$ ppm (*49*). Using an extended Lipari−Szabo formalism (*50*−*52*), the spectral density function $J(\omega)$ can be modeled assuming motions on different time scales. The generalized motional parameters are S^2 (the generalized order parameter), τ_m (the overall rotational correlation time), τ_e (the effective correlation time), and R_{ex} (the chemical exchange term). Using the program MODELFREE (*47*, *53*), data were fit to the five models summarized in Table 1 with the following adjustable parameters: S^2 (model 1); S^2 and τ_e (model 2); S^2 and R_{ex} (model 3); S^2, τ_e, and R_{ex} (model 4); and S_f^2, S_s^2, and τ_e (model 5). MODELFREE calculates Lipari−Szabo motional parameters and their uncertainties from ^{15}N relaxation parameters R_1, R_2, and HNOE and the corresponding experimental uncertainties. The global rotational correlation time, τ_m, was initially estimated from the average ratio of R_2/R_1. Using this fixed τ_m value, the MODELFREE program minimizes the target function of sum-squared error (SSE) for all five models. The SSE value (eq 8) is defined as a function of the back-calculated relaxation parameters (R_1*, R_2*, and HNOE*), the experimental relaxation parameters (R_1, R_2, and HNOE), and the uncertainties in the experimental measurements (σ_1, σ_2, and σ_{HNOE}):

$$\mathrm{SSE} = \frac{(R_1 - R_1^*)^2}{\sigma_1^2} + \frac{(R_2 - R_2^*)^2}{\sigma_2^2} + \frac{(\mathrm{HNOE} - \mathrm{HNOE}^*)^2}{\sigma_{\mathrm{HNOE}}^2} \quad (8)$$

The optimal parameter values for each residue are the ones that provide the smallest SSE value, and the simplest model consistent with goodness of fit criteria is chosen. Additional criteria in model selection require that $S_f^2 > 0.90$, $\tau_e > 30$ ps, and $R_{ex} > 0.5$ Hz; otherwise, simpler models are chosen. Values of the global rotational correlation time τ_m were optimized iteratively, as described elsewhere (*46*). Uncertainties in the derived motional parameters were estimated by Monte Carlo methods (*47*, *54*).

RESULTS

Automated Analysis of Backbone ^{1}H, ^{13}C, and ^{15}N and Side Chain $^{13}C^\beta$ Resonance Assignments. Figure 2 shows the 2D ^{15}N−^{1}H^N HSQC spectrum of CspA at pH 6.0 and 30 °C. As indicated in Materials and Methods, the majority of CspA protein is expressed without the N-terminal methionine, and therefore, CspA contains 69 amino acids, including two prolines. Under these conditions of pH and temperature, there are 58 well-resolved backbone ^{15}N−^{1}H^N peaks and four partially overlapping ^{15}N−^{1}H^N peaks subsequently assigned to residues Lys-4/Ile-8, Thr-6/Gly-26, Asp-15/Asp-25, and Glu-47/Ile-55 in the 2D ^{15}N−^{1}H^N correlation spectrum. ^{15}N−^{1}H^N cross-peaks were observed for residues Lys-4−Leu-70 (excluding prolines) in all of the spectra, while the backbone ^{15}N−^{1}H^N cross-peaks for residues Ser-2 and Gly-3 were not always observed. In the course of this work, an optimized strategy was developed that uses a specific set of triple-resonance NMR experiments as input to the automated resonance assignment program AUTOASSIGN (*32*); this program was developed in part with data sets for CspA described here. It was observed that the following data sets were sufficient to provide reliable and robust assignments with AUTOASSIGN: ^{15}N−^{1}H^N HSQC, HNCO, HNCA, CA(CO)NH, HA(CA)NH, HA(CA)(CO)NH, CBCANH, and CBCA(CO)NH. The results provided by AUTOASSIGN analysis of the peak-picked 2D and 3D NMR spectra are summarized in Table 2. The reliability of these assignments was ascertained through manual analysis of the same data in tandem with the development of the automated analysis program. As neither the N-terminal residues (Met-1 or Ser-2) nor the two proline spin systems exhibit intraresidue amide ^{15}N−^{1}H^N cross-peaks, only 66 backbone ^{15}N−^{1}H^N peaks, which define "roots" of the generic amino acid spin systems (GSs) of CspA in AUTOASSIGN, are expected in the HSQC spectrum. GS roots are defined as the backbone ^{15}N−^{1}H^N resonances of each spin system that is compiled by AUTOASSIGN from the triple-resonance spectra. Two spin systems are considered "degenerate" when their GS roots are simultaneously within 0.025 ppm in the ^{1}H^N dimension and 0.35 ppm in the ^{15}N dimension. The number of "assigned GSs" (Table 2) refers to the number of spin systems that could be assigned to sequence-specific residues. The number of "assigned residues" is generally greater than the numbers of assigned GSs as the chemical shifts of C$^\alpha$,

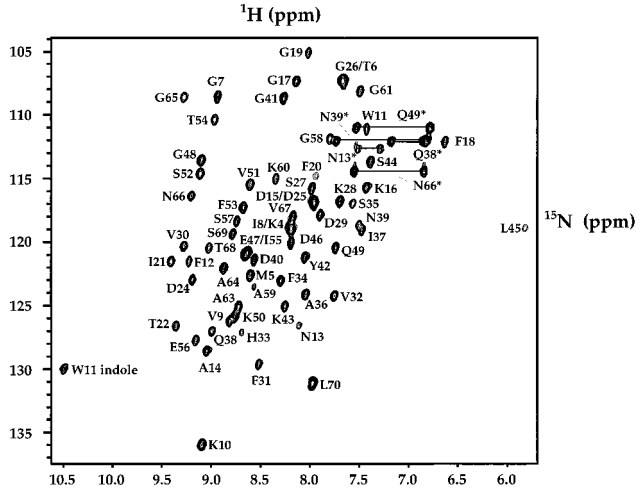

FIGURE 2: 2D ^{15}N−^{1}H^N HSQC spectrum of CspA at pH 6.0 and 30 °C. Backbone amide ^{15}N−H^N correlation peaks are labeled using the one-letter amino acid code followed by the sequence position number. Side chain amide ^{15}N−^{1}H^N correlation peak pairs are connected with solid lines and indicated with an asterisk (*). The side chain indole ^{15}N$^\epsilon$−^{1}H^N correlation of Trp-11 is also labeled.

Table 2: Summary of AUTOASSIGN Analysis for CspA Triple-Resonance NMR Data

residues	69
GSs expected	66
GSs observed	67
degenerate GS roots	8
assigned GSs	65
extra GSs	2
assigned residues	68
percent assigned residues	99%
execution time (s)	16

number of assignments (expected[a])	AUTOASSIGN analysis	manual analysis
backbone		
H^N (67)	65	66
H^α (79)[b]	77	79
^{15}N (67)	65	66
$^{13}C^\alpha$ (69)	67	69
$^{13}C'$ (66)	64	66
$^{13}C^\beta$ (59)	49	59
side chain		
^{15}N (6)	6	6
H^N (11)[c]	11	11

[a] Expected number refers to the total count of possible resonances from Ser-2 to Leu-70 in CspA. [b] Number of expected assignments includes glycines of degenerate H^αs and nondegenerate H^αs. [c] Side chain $^1H^N$ and ^{15}N includes those from three Asns, two Gln, and one Trp.

C^β, and H^α nuclei of residue $i-1$ can be detected on the $^{15}N-^1H^N$ resonances of the GS assigned to residue *i*. In addition to the 65 assigned GSs, two additional GSs were observed that appear to be due to chemical or conformational heterogeneity present in some samples of ^{13}C- and ^{15}N-enriched CspA. In general, manual interpretation of these 3D spectra can require weeks of analysis. The execution time of AUTOASSIGN in determining sequence-specific backbone resonance assignments from these CspA data sets is ~16 s on a Sun Sparc 20 workstation.

One significant problem encountered in the analysis of resonance assignments for CspA was the slow appearance of new cross-peaks in the spectra due to apparent chemical decomposition in CspA samples purified according to the previously published protocol (*21*). To avoid misinterpretation of assignments due to decomposition products, the peak lists from 3D triple-resonance spectra were filtered by the AUTOASSIGN program to exclude cross-peaks with $^{15}N-^1H^N$ values different from those observed in a $^{15}N-^1H^N$ HSQC spectrum recorded on a freshly prepared sample.

The assignments determined by parallel manual analysis of these same triple-resonance spectra were found to be consistent with those obtained by the AUTOASSIGN program (Table 2). However, the program had some difficulties in completing resonance assignments for residues Ser-2 and Gly-3. These difficulties are related to the fact that the CspA samples are heterogeneous with respect to the N-terminal residue Met-1, which is absent in ~80% of the molecules (see Materials and Methods). This heterogeneity resulted in multiple spin systems for residues Ser-2 and Gly-3. Moreover, the intensities of H^N resonances in this N-terminal polypeptide segment were very weak due to solvent-saturation transfer effects.

Manual Analysis of Side Chain Resonance Assignments. These backbone ^1H, ^{13}C, and ^{15}N and C^β resonance assignments were next extended by manual analysis of 3D PFG CC(CO)NH−TOCSY, H(CC)(CO)NH−TOCSY, CCNH−TOCSY, and H(CC)(CO)NH−TOCSY spectra to provide side chain ^1H, ^{13}C, and ^{15}N resonance assignments. In addition, 2D $^1H-^1H$ TOCSY experiments on samples dissolved in both ^1H_2O and ^2H_2O were analyzed. These 2D experiments were required to complete resonance assignments of aromatic protons and were complementary to the 3D data sets because they had much higher resolution in aliphatic proton dimensions. The resulting list of ^1H, ^{15}N, and ^{13}C resonance assignments is presented in Table S2 of the Supporting Information. The intraresidue and sequential ^1H and ^{13}C connectivities that were observed in the triple-resonance experiments and used to establish these assignments are summarized in Figure 3. Individual assignments of side chain amide protons for three Asns and two Gln were made using the relative amplitudes of $H^\beta-H^\delta$ and $H^\gamma-H^\epsilon$ NOE peak intensities, respectively (*55*). For the 33 nondegenerate methylene H^βs, 10 pairs of stereospecific H^β assignments were obtained using the HYPER program.

Characterization of Secondary Structure in CspA. Secondary structural elements of CspA were identified using a combination of data, including medium- and long-range interstrand NOEs, $^3J(H^N-H^\alpha)$ coupling constants, amide $^1H-^2H$ exchange rates, and characteristic $\Delta\delta C^\alpha$ and $\Delta\delta C^\beta$ values [where $\Delta\delta C^{\alpha,\beta} = \delta C^{\alpha,\beta}{}_{obs} - \delta C^{\alpha,\beta}{}_{rc}$ (*56*)]. These data are summarized in Figure 3. From this information, five β-strands corresponding to polypeptide segments of residues 5−13, 18−22, 30−33, 50−56, and 63−70 were identified. Interstrand hydrogen bonds were determined on the basis of the interstrand backbone−backbone NOEs and the locations of slowly exchanging amides. CspA contains five antiparallel β-strands (Figure 4). β-Strands 1−4 form a Greek-key topology. Two bulges were defined by the patterns of slowly exchanging amide protons and of unambiguous H^N-H^N and $H^\alpha-H^\alpha$ NOE's in the β-sheets; they are at positions Lys-10 on β-strand β1 and Asn-66 on β-strand β5.

Amide $^1H-^2H$ Exchange Rates. Amide $^1H-^2H$ exchange rates for CspA at pH* 6.0 and 30 °C are unusually fast for a folded protein; nearly all of the amide proton sites are fully exchanged within about 2 h. Twenty-five backbone amide sites which were observed after ~5 min of $^1H-^2H$ exchange are considered in this study to be "slowly exchanging". Among these slowly exchanging backbone amide protons, the six slowest (Val-9, Ile-21, Thr-22, Val-30, Val-32, and Val-51) exhibit exchange rate constants, *k*, ranging from 0.024 to 0.055 min^{-1}.

Experimental Conformational Constraints. In total, 699 NMR constraints were included in the structure refinement. An additional 279 upper distance limits were found to be nonrestrictive on the conformations generated by DIANA and were not used in the calculations. Table 3 summarizes the three kinds of experimental constraints used in the input: (i) NOE-derived upper-bound $^1H-^1H$ distance constraints, (ii) hydrogen-bond constraints, and (iii) constraints on dihedral angles ϕ, ψ, and χ^1. The distribution of the NOE-derived constraints along the protein primary sequence is shown in Figure 5A; the average number of distance constraints per residue is 10.4. Amide N−H hydrogen bond donors were identified as slowly exchanging amide protons, and acceptors were identified from the constraints imposed by interstrand NOEs. Dihedral angle constraints were obtained from the program HYPER as outlined in Materials

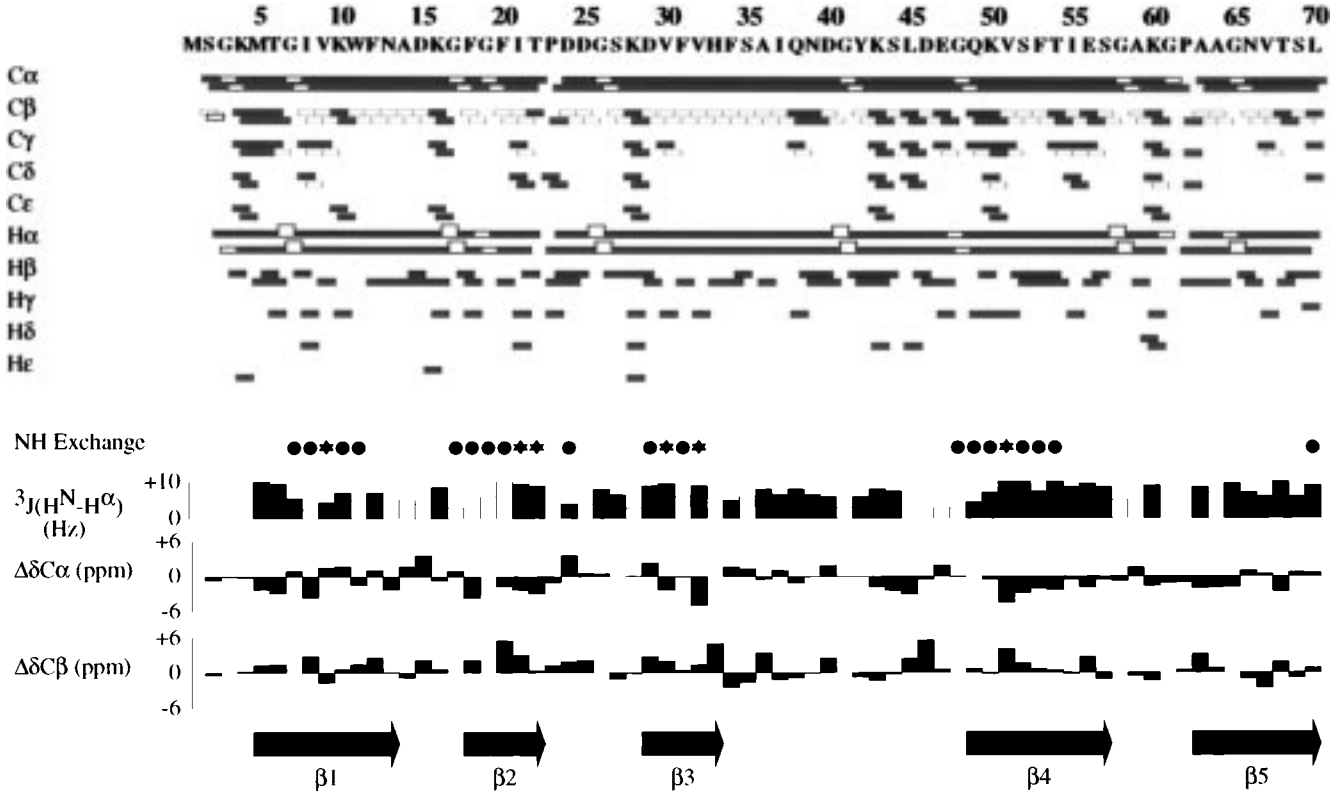

FIGURE 3: (Top) Summary of information derived from triple-resonance data sets and used for establishing intraresidue and sequential correlations of spin systems. Data derived from CA-type (intraresidue correlations) experiments are indicated with red bars beneath the amino acid sequence, while data from CO-type (sequential correlations) experiments are indicated with blue bars. (Bottom) Summary of NMR data used to identify secondary structural elements in CspA. Slowly exchanging backbone amides ($t_{1/2}$ > 3 min at pH 6.0 and 30 °C) are indicated by filled circles ($t_{1/2}$ < 30 min) or stars ($t_{1/2}$ > 30 min). Values of $^3J(H^N-H^\alpha)$ coupling constants are indicated by vertical bars; filled bars indicate that the data provided a useful estimate (±0.5 Hz) of the corresponding coupling constant, while open bars indicate that the experimental data provide only an upper bound on its value. Values of conformation-dependent secondary shifts $\Delta\delta C^\alpha$ and $\Delta\delta C^\beta$ are plotted with solid bars. The locations of five β-strands in the sequence of CspA are indicated with arrows.

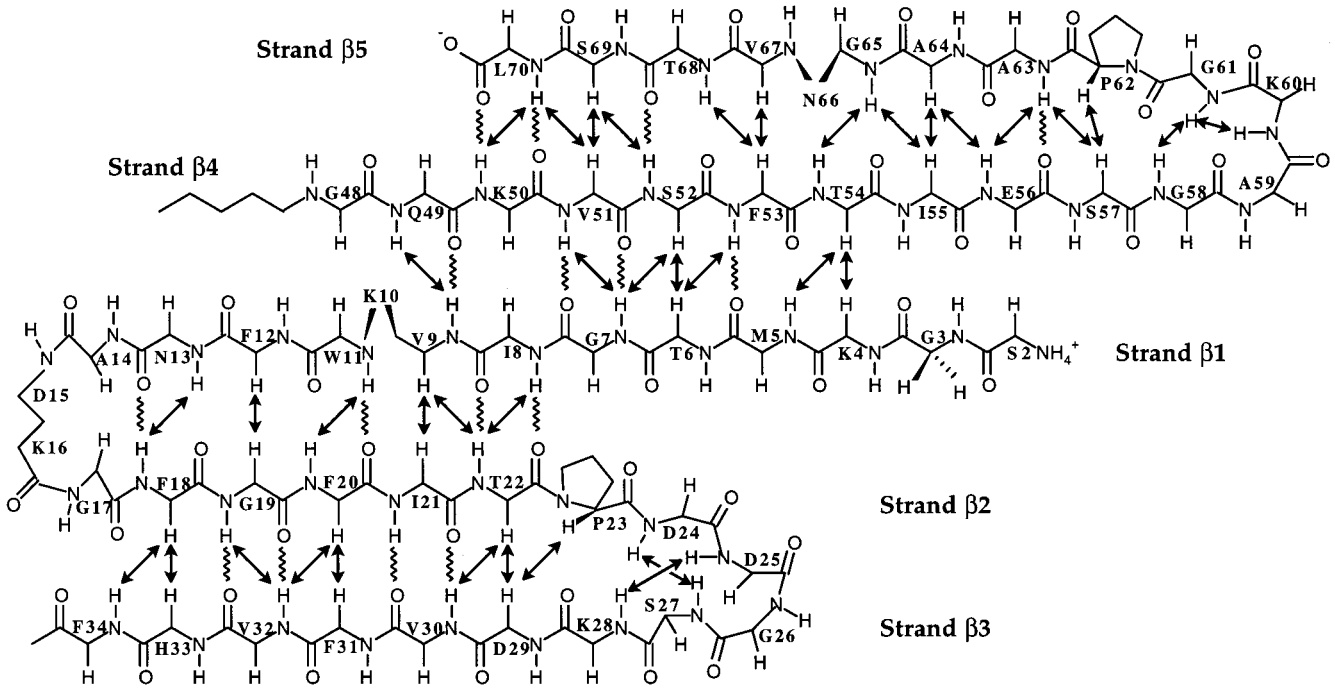

FIGURE 4: Schematic diagram of the five β-strands of CspA. Unambiguous interstrand backbone NOEs are indicated with arrows, and locations of interstrand hydrogen bonds involving slowly exchanging amide protons are denoted by squiggly lines.

and Methods. Structure generation calculations were then carried out using the program DIANA as described in Materials and Methods. Finally, 16 conformations with the lowest target function values were chosen to represent the solution structure of CspA. Each resulting structure satisfies all of the distance constraints to within ~0.5 Å and all of

Table 3: Summary of Input Constraints for Structure Calculations of CspA

total number of NMR constraints	699
total number of upper-bound NOE constraints	531
intraresidue constraints ($i = j$)	90
interresidue constraints ($i - j = 1$)	160
backbone−backbone constraints	50
backbone−side chain constraints	4
side chain−side chain constraints	106
interresidue constraints ($1 < i - j \leq 5$)	48
backbone−backbone constraints	8
backbone−side chain constraints	7
side chain−side chain constraints	33
interresidue constraints ($i - j > 5$)	233
hydrogen bond constraints[a]	62
total number of dihedral angle constraints[b]	106
ϕ	36
ψ	36
χ^1	34
pairs of stereospecifically assigned $C^\beta H_2$ resonances	10

[a] Pairs of upper- and lower-bound hydrogen bond constraints. [b] Pairs of upper- and lower-bound dihedral angle constraints.

the dihedral angle constraints to within ~5°.

Analysis of the Solution Structure of CspA. Overall, the solution structure of CspA is reasonably well-defined from these NMR data, except for the polypeptide segment from Ser-2 to Lys-4, and portions of surface loops corresponding to polypeptide segments from Asn-39 to Gly-48 and from Gly-58 to Gly-61. Superpositions of the backbone atoms for 16 conformers of CspA generated in these DIANA calculations are shown in Figure 6A. Backbone and heavy atom rmsds for these superposed structures are listed in Table 4, and the corresponding dihedral angle order parameters (*57*), $S(\phi)$ and $S(\psi)$, are summarized in panels A and B of Figure 7. The stereochemical quality of these structures was also examined using the program Procheck-NMR (*58*). Figure S10 in the Supporting Information shows a Ramachandran plot for the structures, excluding residues with an angular order parameter $S(\phi)$ or $S(\psi)$ smaller than 0.8. Except for residues in poorly defined regions of the protein structure with low dihedral angle order parameters, all $\phi - \psi$ values occur in low-energy regions in the Ramachandran plot. A stereoview of a representative CspA conformer is presented in Figure 6D.

Backbone Dynamics of CspA. Quantitative analysis of the R_1, R_2, and HNOE data was possible for 58 well-resolved backbone amide groups and one Trp side chain group. The decay of the intensities was monoexponential for each residue in the R_1 and R_2 curve-fitting procedure. The average values of R_1, R_2, and HNOE over these 59 sites are 2.39 s^{-1}, 7.40 s^{-1}, and 0.59 unit, respectively. Plots of HNOE, T_1, and T_2 values obtained for each N−H site are shown in panels B−D of Figure 5. The values of R_1, R_2, and HNOE, along with their measurement uncertainties, are also listed in Table S3 of the Supporting Information. Backbone ^{15}N atoms of residues Gly-3, Asp-46, and Gly-61 have R_1 values of less than 2.0 s^{-1}, while the ^{15}N of residue Ser-35 has the longest R_1 value, 2.9 s^{-1}. Large R_2 values (>10 s^{-1}) were observed for backbone N−H sites of Lys-10−Asn-13, Phe-31, His-33, Thr-54, and Lys-60. ^{15}N atoms in polypeptide residues (or polypetide segments) Gly-3, Asn-39−Lys-43, Asp-46, and Gly-58−Gly-61 have HNOE values of less than 0.5 unit; the HNOE for the N−H site of residue Gly-3 is negative.

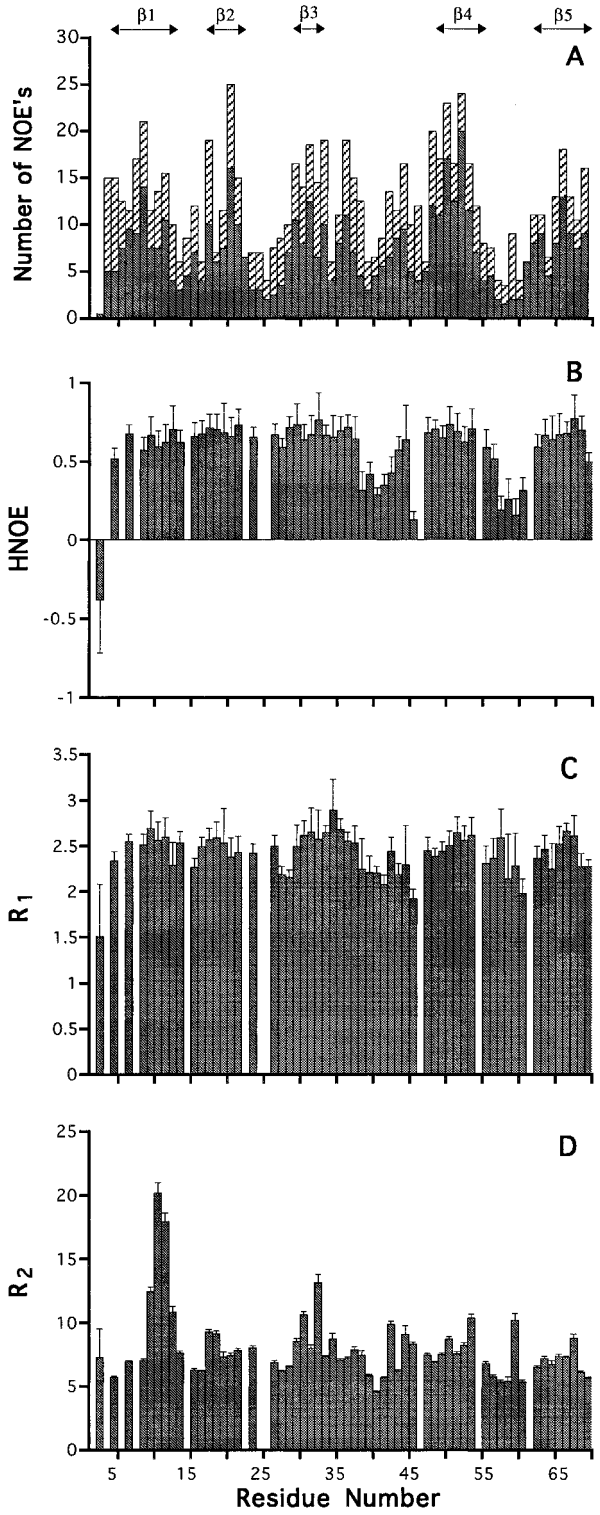

FIGURE 5: Plots of the distributions in the sequence of (A) $^1H-^1H$ NOE constraints, (B) HNOE values, (C) backbone ^{15}N R_1 values, and (D) backbone ^{15}N R_2 values, for residues Gly-3−Leu-70. Intraresidue and interresidue constraints are indicated in panel A with solid bars and hatched bars, respectively. Arrows indicate the locations of β-strands in the sequence. Uncertainties in estimates of HNOE, R_1, and R_2 measurements are also plotted.

The principal components of the inertia tensor derived from unsolvated atomic coordinates of the NMR structure of CspA are 0.90:1.00:1.00. The corresponding ratios of rotational correlation times about each of these principal axes, computed assuming oblate ellipsoidal hydrodynamics (*59*), are 0.96:1.00:1.00. Accordingly, the isotropic rotational

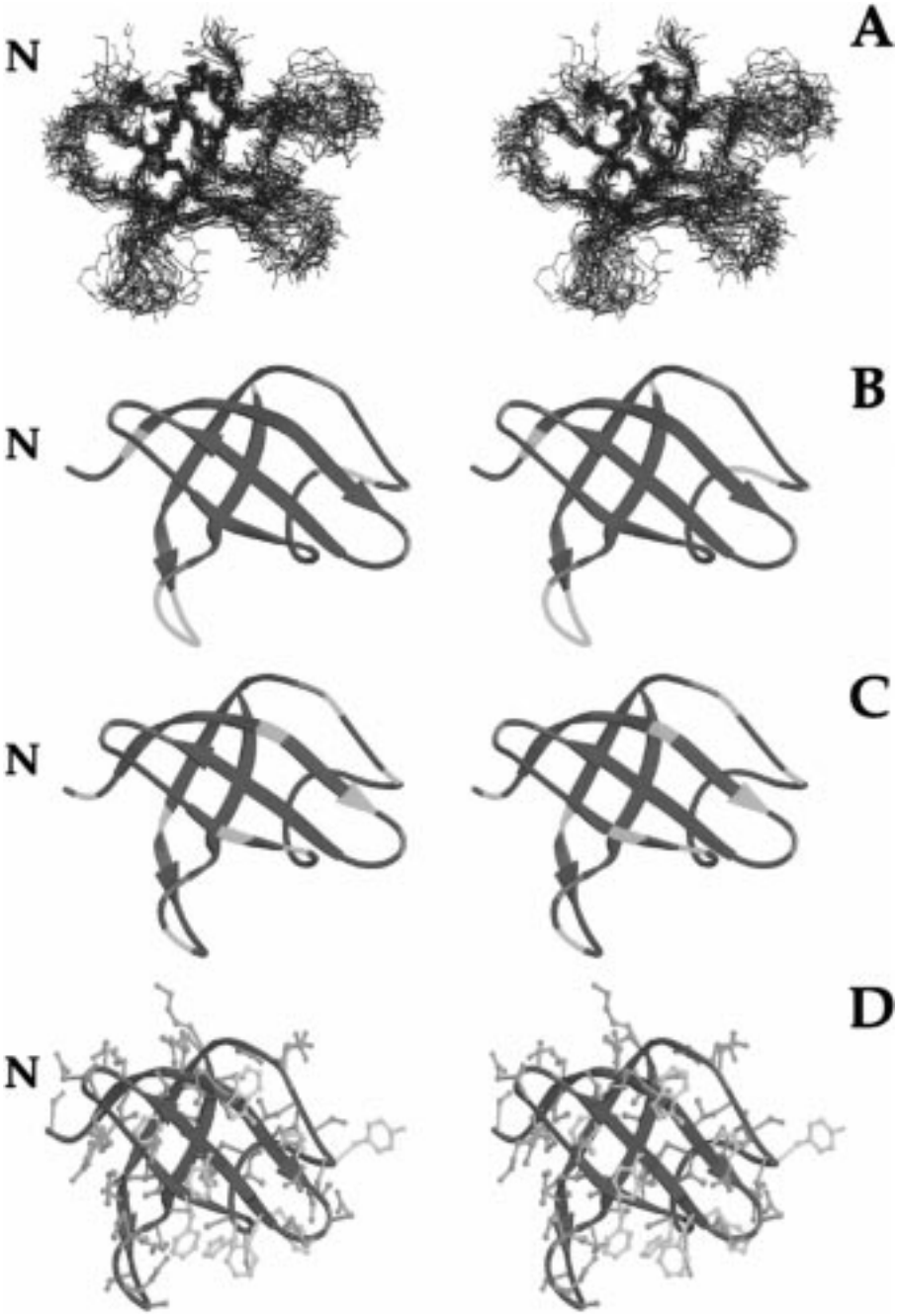

FIGURE 6: Stereoviews of the solution NMR structure of *E. coli* CspA. The N termini are labelel in each stereoview. (A) Atomic coordinates of backbone (N, C^{α}, and C′) atoms of 16 solution NMR structures, optimally superimposed with respect to the average coordinates of the β-core backbone structure (i.e., residues 5−9, 19−22, 50−56, and 63−69). Also shown (in pink) is a superposition of backbone atoms of the crystal structure (*10*) of CspA on these same reference coordinates. (B) Summary of backbone N−H bond vector order parameters, S^2, on a ribbon diagram (*95*) of a representative solution NMR structure of CspA; the backbone ribbon is shown in blue for sites where $S^2 \geq 0.75$, pink where $0.60 \leq S^2 < 0.75$, red where $S^2 < 0.60$, and gray for sites for which motional parameters could not be obtained. (C) Summary of backbone R_{ex} terms and order parameters S^2 on the same NMR structure of CspA; the backbone ribbon is shown in blue for sites where $0 \leq R_{ex} < 3.0$ Hz, pink where 3.0 Hz $\leq R_{ex} <$ 10 Hz, red where $R_{ex} \geq 10$ Hz, and gray for sites for which motional parameters could not be obtained. (D) Location of the previously characterized single-stranded nucleic acid-binding epitope on the current structure of CspA (for details of these chemical shift perturbation measurements, see ref *9*). Side chain atoms are shown in orange except for the nine aromatic groups which are shown in yellow; the nucleic acid-binding epitope includes seven surface aromatic side chains and accounts for much of the internal dynamics of the molecule. The backbone ribbon is shown in blue for backbone N−H sites exhibiting no chemical shift change upon complex formation, red for sites with significiant chemical shift changes upon complex formation, and gray for sites where the chemical shift perturbation could not be reliably characterized.

tumbling model used in this dynamical analysis of CspA is reasonably valid, although the relaxation properties at some sites may be affected to a small degree by anisotropic tumbling effects.

In our experience, the model selection and optimized motional parameters derived from Lipari−Szabo analysis of relaxation data depend significantly on the initial value of the overall correlation time, τ_m, used in the analysis. MODELFREE analysis started with an initial estimate of τ_m based on the average ratio $\langle R_2/R_1 \rangle$, excluding sites with ratios differing by more than two standard deviations from the mean value. This initial value $\langle R_2/R_1 \rangle$ was iteratively

Table 4: Atomic rmsds for the Final 16 CspA Structures[a]

residue range	backbone atoms	heavy atoms
all residues (5−69)	1.45 (1.16−1.77)[b] Å	2.01 (1.63−2.26) Å
ordered regions[c]	0.84 (0.68−1.27) Å	1.38 (1.12−1.47) Å
core regions[d]	0.62 (0.41−0.78) Å	1.07 (0.88−1.47) Å

[a] rmsds are relative to the mathematical average structure. [b] The ranges of rmsd values computed are shown in parentheses. [c] Well-defined regions refer to the parts in the structure that have a sum for the ϕ and ψ order parameters of >1.5. [d] Core regions refer to β-sheet residues 5−9, 19−22, 50−56, and 63−69.

Table 5: Rotational Correlation Time (τ_m) for Three Cycles of MODELFREE Calculations

cycle	$\langle R_2/R_1\rangle$[a]	τ_{m-init} (ns)	τ_{m-opt} (ns)
1	3.03[b] ± 0.23	5.37 ± 0.35	4.90 ± 0.04
2	2.85[c] ± 0.20	5.11 ± 0.30	4.90 ± 0.04
3	2.70[d] ± 0.22	4.88 ± 0.34	4.88 ± 0.04

[a] The value $\langle R_2/R_1\rangle$ is the mean ratio of R_2/R_1. The initial value of $\langle R_2/R_1\rangle_{all}$ is 3.31 ± 1.02 averaged over all 59 N−H sites for which R_1 and R_2 data were obtained. [b] $\langle R_2/R_1\rangle$ ratio for 49 N−H sites that satisfy the condition $|R_2/R_1 - \langle R_2/R_1\rangle_{all}| < 1.00$. The 10 excluded residues are G3, K10, W11, F12, N13, H33, G41, D46, G58, and K60. [c] $\langle R_2/R_1\rangle$ ratio for 38 N−H sites from cycle 1 that were fit with $R_{ex} < 1.5$ Hz. The 11 excluded residues are F18, G19, D24, V30, F31, N39, K43, G45, V51, T54, and T68, along with the 10 residues listed in footnote b. [d] All of the R_2 values were corrected using R_{ex} values computed in cycle 2: $R_{2corr} = R_2 - R_{ex}$. The resulting value of $\langle R_2/R_1\rangle_{all,corr}$ is 2.68 ± 0.18 averaged over all 59 residues. $\langle R_2/R_1\rangle$ was then calculated using 44 residues which satisfy the condition $|R_2/R_1 - \langle R_2/R_1\rangle_{all,corr}| < 0.18$. The 15 excluded residues are G3, M5, G17, F20, G41, D46, Q49, S52, E56, S57, G58, K60, A64, G65, and N66.

refined using three cycles of MODELFREE calculations by a process analogous to that described by Li and Montelione (*46*) in the analysis of ^{15}N relaxation data for type-α transforming growth factor. This process of iterative refinement of τ_m and associate Lipari−Szabo motional parameters, summarized in Table 5, converged on an optimal global value τ_m of 4.88 ± 0.04 ns.

The resulting motional parameters (S^2, τ_e, and R_{ex}) derived from the extended Lipari−Szabo analysis are plotted in panels C−E of Figure 7, and are also listed Table S4 of the Supporting Information. The numbers of backbone N−H sites that were best fit to models 1−5 were 4, 17, 11, 22, and 4, respectively. Data for residues Gly-41 and Asp-46 did not fit well with any of the five models. In the final calculation, the data were fit with model 5 for Gly-41 and model 4 for Asp-46. The generalized order parameters S^2 ranged from 0.33 to 0.98 unit. Lower S^2 values were observed for ^{15}N atoms located in residues (or polypetide segments) Gly-3, Asn-39−Lys-43, Asp-46, Ser-57−Gly-61, and Leu-70. The locations of sites with low values of S^2 (Figure 6B) indicate that the N- and C-terminal regions, and the two polypeptide loops connecting β-strands $\beta3-\beta4$ and $\beta4-\beta5$, undergo relatively large amplitude internal motions on the sub-nanosecond time scale. N−H sites of residues Met-5, Gly-17, Gly-41, and Leu-70 appear to be particularly dynamic, and could only be fit by assuming a multiple-time scale model of the internal motions (model 5). Forty-three ^{15}N sites were fit to models with effective correlation times for internal motion τ_e ranging from 30 to 300 ps, and the ^{15}N atom of residue Gly-3 was fit with a τ_e value of ∼500 ps. However, Monte Carlo uncertainty analysis indicates that these larger τ_e values have very large uncertainties.

For several backbone N−H sites, the ^{15}N relaxation data could be fit only by including an additional contribution to the R_2 relaxation rate, R_{ex}. Thirteen ^{15}N sites (Figure 6C) were fit with R_{ex} terms of >2 Hz: Gly-3, Lys-10, Trp-11, Phe-12, Asn-13, Phe-18, Gly-19, Phe-31, His-33, Lys-43, Asp-46, Thr-54, and Lys-60. The R_{ex} values of these N−H sites suggest the presence of multiple diamagnetic environments within CspA that interconvert on the micro- to millisecond time scale.

Tryptophan-11 Backbone and Side Chain Dynamics in CspA. The cross-peak of the single tryptophan side chain indole N−H (Trp-11) is also well-resolved in the 2D ^{15}N−^{1}H^{N} HSQC spectrum of CspA (Figure 2), allowing accurate determination of its ^{15}N relaxation parameters. Under these conditions, the Trp-11 indole N$^{\epsilon1}$ atom has an R_1 of 2.08 ± 0.10 s^{-1}, an R_2 of 6.88 ± 0.13 s^{-1}, and an HNOE of 0.64 ± 0.07 unit. Both the side chain and backbone Trp-11 N−H relaxation data fit model 4. The best-fit values of S^2, τ_e, and R_{ex} for the indole N−H are 0.72 ± 0.04, 34 ± 23 ps, and 1.31 ± 0.30 Hz, while for the backbone N−H, the corresponding values are 0.88 ± 0.05, 168 ± 154 ps, and 13.39 ± 0.92 Hz. The Trp-11 N$^{\epsilon1}$−H bond vector exhibits a relatively high S^2 value for a side chain group, indicating that even though it is on the surface of CspA this indole ring exhibits relatively restricted sub-nanosecond mobility. However, the requirement for a significant R_{ex} contribution to fit both the Trp-11 backbone and side chain relaxation data indicates that the environments of these N−H bonds are modulated on the microsecond to millisecond time scale either by slow motions of neighboring groups or by slow motions of the Trp-11 indole ring itself. The R_{ex} contribution required to fit the Trp-11 backbone N−H, 13.4 Hz, is especially dramatic. These dynamics probably have functional significance, as Trp-11 is within the single-stranded nucleic acid-binding epitope of CspA (*9*−*11*, *20*).

DISCUSSION

Automated Analysis of Backbone Resonance Assignments from Triple-Resonance NMR Data. Triple-resonance NMR data are especially well suited for automatic analysis with computer software (*60*). The AUTOASSIGN program (*32*) uses constraint propagation methods from artificial intelligence to narrow the domain of possible sequence-specific assignments for each spin system to unique positions in the primary amino acid sequence. The particular set of triple-resonance experiments used here included 2D ^{15}N−^{1}H^{N} HSQC and 3D HNCO data for defining the ^{15}N−^{1}H^{N} "roots" of each spin system, together with 3D HACA(CO)NH, HACANH, CA(CO)NH, CANH, CBCA(CO)NH, and CBCANH data to establish intraresidue and sequential correlations of the spin systems. Back-to-back data collection on identical samples of all these experiments is crucial for achieving the smallest and most accurate match tolerances. The total collection time for the eight experiments was approximately 7 days, and the analysis time was significantly reduced from a few weeks of manual work to the few hours of time required for semiautomated peak picking of these spectra plus ∼16 s for AUTOASSIGN execution. In addition, once the backbone (and C$^{\beta}$) resonance assignments were determined by AUTOASSIGN, the manual analysis of the peripheral side chain assignments was quite straightforward to complete. In addition to CspA, AUTOASSIGN has

been applied successfully in analysis of resonance assignments for several other proteins, with molecular masses ranging from 7 to 17 kDa (*32*, *61*−*63*). The primary limitation for small proteins such as CspA involves the commercial peak-picking software used in this analysis, as certain spectral artifacts cannot yet be reliably edited without some human intervention.

Description of the Solution NMR Structure of CspA at pH 6.0. The structure of CspA is a closed β-barrel formed by two β-sheet surfaces. It contains five strands, β1 through β5, corresponding to polypeptide segments of residues 5−13, 18−22, 30−33, 50−56, and 63−70. The first four β-strands have a Greek-key fold topology. Two residues, Lys-10 and Asn-66, in these β-sheets (Lys-10 in β1 and Asn-66 in β5) form β-bulges. These two bulges break strands β1 and β5 into two small pieces, namely $\beta1'-\beta1''$ and $\beta5'-\beta5''$. Two β-sheet surfaces, β-sheet 1 and β-sheet 2, are formed by strands $\beta1''-\beta2-\beta3$ and $\beta1'-\beta4-\beta5'-\beta5''$. The slowest exchanging backbone amide protons are those of residues Val-9, Ile-21, Thr-22, Val-30, Val-32, and Val-51. The side chain groups of these residues (except Thr-22) point to the interior of the protein, forming the hydrophobic core of the protein. The antiparallel β-sheets are stabilized by the backbone hydrogen-bonding network within the β-sheets and hydrophobic interactions between the two β-sheets. Surface loops connecting strands β1 and β2, and β2 and β3, in β-sheet 1 and strands β4 and β5 in β-sheet 2 are quite short, while the polypeptide loop connecting β-sheets 1 and 2 (i.e., between strands β3 and β4) is about 13 residues long. The current structure determination is much more precise, accurate, and complete than the initial description of the chain fold that we presented previously (*9*).

An especially important feature of the CspA structure is the presence of seven solvent-exposed aromatic residues, including side chains of residues Trp-11, Phe-18, Phe-20, Phe-31, His-33, Phe-34, and Tyr-42 (shown in yellow in Figure 6D). These aromatic residues are highly conserved among various members of the prokaryotic cold-shock protein and Y-box protein families (Figure 1). RNP1 and RNP2 RNA-binding sequence motifs (*8*, *18*, *19*), Lys-Gly-Phe-Gly-Phe-Ile and Val-Phe-Val-His-Phe, are located in antiparallel strands β2 and β3, respectively. The aromatic side chains of these evolutionarily conserved RNP sequence motifs (Phe-18, Phe-20, Phe-31, and His-33) are on the surface of the CspA protein. The aromatic side chain of conserved residue Trp-11 is also turned onto the surface by the β-bulge at Lys-10 in β-strand β1. Aromatic residue Phe-34 which is highly conserved among the bacterial cold-shock proteins (Figure 1) is largely solvent accessible in CspA. The resulting evolutionarily conserved surface aromatic cluster (shown in Figure 6D) appears to play important roles in the single-stranded nucleic acid-binding epitope (*9*−*11*, *20*).

Comparison of Solution NMR and X-ray Crystal Structures of CspA. A comparison of the atomic coordinates for the CspA backbone chain fold of the 16 NMR conformers with the corresponding coordinates for the CspA backbone atoms of the 2.0 Å X-ray crystal structure (*10*) is also shown in Figure 6A. The NMR and X-ray structures of *E. coli* CspA have essentially identical chain folds. For backbone atoms of the β-sheet core region (i.e., residues Met-5−Val-9, Gly-19−Thr-22, Lys-50−Glu-56, and Ala-63−Ser-69), the rmsd of the mathematical average coordinates of the 16 NMR structures to the X-ray structure is 0.61 Å. Two β-bulges at positions Lys-10 and Asn-66 are well-defined by the interstrand NOEs and hydrogen bonds in the solution structure. Similar β-bulges are also found in the crystal structures of CspA from *E. coli* (*10*) and in homologous positions in CspB from *B. subtilis* (*14*, *15*).

The atomic coordinates of the crystal and solution NMR structures of *E. coli* CspA are most different for surface loops between β strands β3 and β4, and between β4 and β5 (Figure 6A), where considerable motion is indicated by the ^{15}N relaxation data. In the crystal structure of CspA, the conformation of the polypeptide segment of residues Asn-39−Tyr-42 between strands β3 and β4 could not be modeled reliably due to high thermal factors for atoms in this portion of the structure (*10*); i.e., the average *B* value for backbone atoms in this polypeptide segment is 72 Å^2, while the average *B* value for all backbone atoms is 27.5 Å^2. The high thermal factors in this loop suggest that it is partially disordered in the crystallization environment. These observations are consistent with the NMR data which reveal a relatively low number of constraints (Figure 5A), low values of backbone dihedral angle order parameters (Figure 7A,B), low values of the generalized order parameter S^2 (Figure 7C), and relatively large τ_e values (Figure 7D) in this region of the protein structure.

Comparison of the Solution Structure of CspA with Other Proteins. CspA is a member of the oligonucleotide/oligosaccharide-binding (OB) fold family (*9*, *10*, *14*−*17*). Many other single-stranded nucleic acid-binding proteins utilize similar β-barrel-like structures to form their nucleic acid-binding epitopes. Using the refined CspA coordinates described here, structural database searches of the Brookhaven Protein Data Base (PDB) with the program DALI using standard search parameters (*64*) identify a long list of proteins (or domains) of structural homologues. Aside from the published X-ray and NMR structures of CspA and CspB, these other structural homologues have little or no detectable sequence homology with CspA. Significantly, however, many of these proteins have biochemical functions that involve single-stranded DNA or RNA binding. Structural homologues of CspA that have little or no sequence identity but exhibit similar biochemical functions include portions of the RNA-binding domain of *E. coli* polyribonucleotide nucleotidyltransferase (*65*), the human mitochondrial ssDNA-binding protein (*66*), *E. coli* translation initiation factor 1 (*67*), the ssDNA-binding protein from gene V of filamentous bacteriophages M13 and f1 (*68*−*71*), the ssDNA-binding protein from filamentous *Pseudomonas* phage Pf3 (*72*), elongation factor G from *Thermus thermophilus* (*73*), a domain of yeast aspartyl tRNA synthetase (*74*), a domain of *E. coli* lysyl tRNA synthetase (*75*), human replication protein A (*76*), staphylococcal nuclease (*77*), and a domain of *E. coli* topoisomerase I (*78*). Some of these structures are shown in Figure 8. The chain fold of CspA is also strikingly similar to those of ribosomal protein S17 from *Bacillus stearothermophilus* (*79*) and the RNA-binding domain of *E. coli* rho transcriptional termination factor (*80*). It is remarkable that this wide range of gene products uses different amino acid sequences to generate similar three-dimensional structural motifs that have homologous functions for binding single-stranded oligonucleotides.

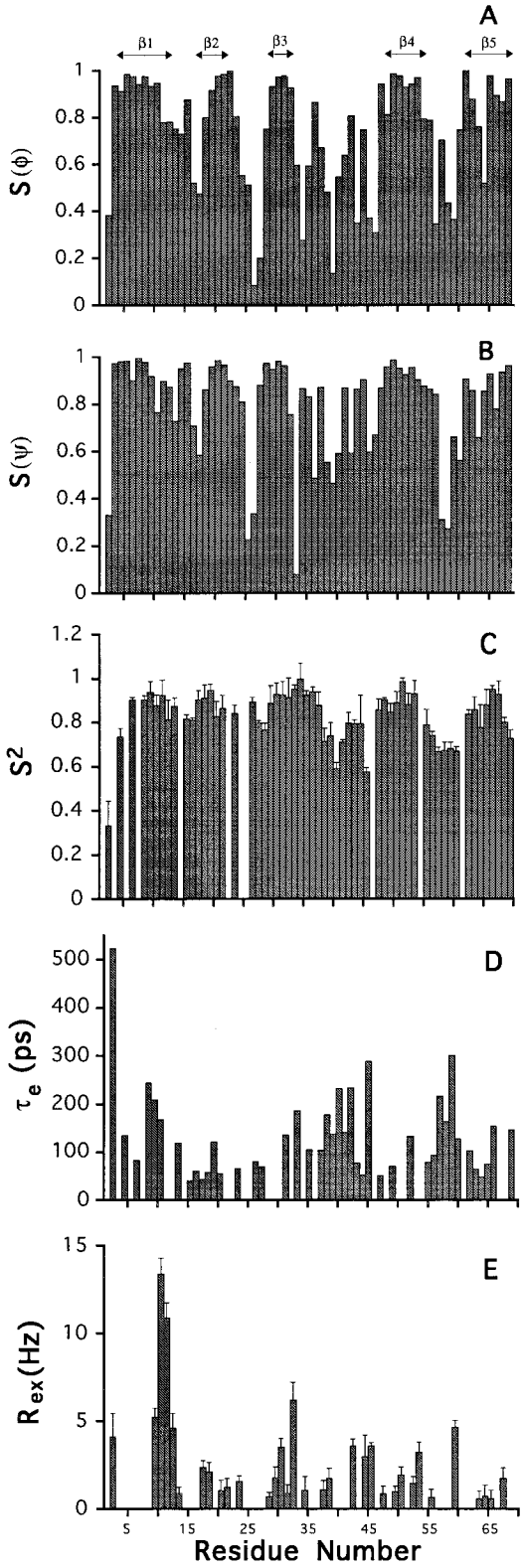

FIGURE 7: Plots of the distributions in the sequence of (A and B) backbone dihedral angle order parameters (*57*) for ϕ and ψ, together with Lipari−Szabo motional parameters (C) S^2, (D) τ_e, and (E) R_{ex}, for residues Gly-3−Leu-70. Arrows indicate the locations of β-strands in the sequence. Uncertainties in estimates of S^2 and R_{ex} are also indicated; uncertainties in estimates of τ_e were generally quite large and listed in Table S4 in the Supporting Information. Reanalysis of the relaxation data with a CSA value of −170 ppm affected the S^2 and R_{ex} values by <3% and the τ_e values by <15%.

Structural database searches with programs such as DALI (*64*) provide an important bioinformatics approach for identifying proteins (or domains) that have similar biochemical functions (e.g., single-stranded nucleic acid-binding functions) but no easily detected amino acid sequence homology. CspA provides an especially striking example of using structural homology as a basis for identifying gene products that have related biochemical functions. Moreover, information on the *biochemical* functions of CspA has provided important insights into its potential *cellular* functions. For example, in part on the basis of the evidence for its in vitro ssRNA-binding activity, it has been proposed that CspA may function in the cellular physiology of cold-shock adaptation as an RNA "chaperone" (*13*). According to this model, a sudden drop in temperature induces duplex formation of mRNA molecules, disrupting the translational machinery. Working together with a cold-shock-induced RNA helicase CsdA (*81*), CspA is proposed to bind single-stranded mRNA molecules with low affinity and specificity, maintaining the RNA in single-stranded states at low temperatures so that translation can resume. CspA may also function as a transcriptional regulator through its mRNA-binding activities (*12*, *13*).

Similarities of biochemical functions that can be deduced by structural homology searches are valuable for developing hypotheses regarding cellular functions but do not provide direct information about the cellular roles of gene products in physiological processes. Moreover, as is true of homo-

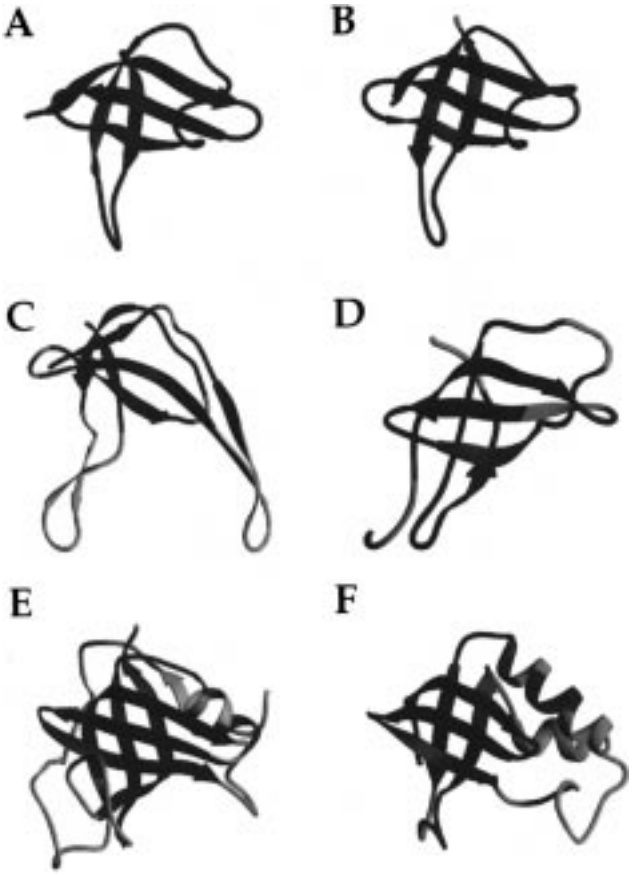

FIGURE 8: Representative ribbon diagram (*95*) of protein structures that are both structurally and functionally homologous with (A) *E. coli* CspA, including (B) *B. subtilis* CspB (*14*), (C) the ssDNA-binding protein from gene V of filamentous bacteriophage f1 (*70*), (D) *E. coli* translation initiation factor 1 (*67*), (E) residues 68−217 of yeast aspartyl tRNA synthetase (*74*), and (F) staphylococcal nuclease (*77*). The regions of these proteins that are structurally homologous (as defined in ref *64*) to CspA are shown in blue.

logues identified by sequence similarity, functional homologies suggested from structural homologies are not always reliable or easily understood. For example, in addition to identifying structural homologues of CspA with single-stranded nucleic acid-binding functions, DALI also identifies several structural homologues with apparently unrelated biochemical functions, including staphylococcal toxic shock syndrome toxin-1 (*82*), turnip leaf cytochrome *f* (*83*), pertussis toxin (*84*), *E. coli* verotoxin-1 (*85*), and several others. Some of these proteins have biochemical functions involving oligosaccharide binding that may be physiochemically (and evolutionarily) related to the nucleic acid-binding mechanism of CspA. It is not clear why some of the oligonucleotide-binding proteins share a similar fold with oligosaccharide-binding proteins; however, structural similarities of these other proteins (or domains) and CspA may be a manifestation of an ancient functional divergence within a single protein (or domain) fold family or may represent simple coincidence.

Relationship to the RNP RNA-Binding Structural Motif. The most widely found and best-characterized single-stranded RNA-binding motif is the RNP motif (for reviews, see refs *8* and *19*). The key identifying feature of the RNP motif is the RNP (also called the RRM−RNA recognition motif) consensus sequence, comprised of RNP1 and RNP2 sequence motifs together with other well-conserved structurally important residues (*8, 19*). Three-dimensional structures are now available for RNP domains from several ssRNA-binding proteins (*86*−*92*). Invariably, these domains exhibit $\beta\alpha\beta\beta\alpha\beta$ secondary structures, in which the four β-strands form a single β-sheet surface with the RNP1 and RNP2 sequence motifs oriented opposite one another in two central antiparallel hydrogen-bonded β-strands (strands β3 and β1, respectively). The β-sheet face of the RNP structural motif, together with residues in polypeptide loops, forms the ssRNA-binding epitope (*89*−*91, 93*), with the two α-helices on the opposite side of the molecule. It is quite remarkable that CspA, an all β-sheet protein with no obvious evolutionary relationship with the RNP motifs, also uses similar RNP1 and RNP2 sequence motifs on antiparallel hydrogen-bonded β-strands (β2 and β3) to form part of its single-stranded nucleic acid-binding epitope. Significantly, the order of RNP1 and RNP2 sequence motifs along the amino acid sequence is distinct for CspA-like folds (where RNP1 is followed by RNP2 sequence motifs as in Figure 1) and RNP folds (where RNP2 is followed by RNP1 sequence motifs). The CspA-like and RNP structural motifs thus appear to represent examples in which different gene families have converged on a similar molecular recognition epitope from very different evolutionary directions.

Single-Stranded Nucleic Acid-Binding Epitope. Although it has not yet been possible to determine the 3D structure of a CspA−nucleic acid complex, the location of the single-stranded nucleic acid-binding epitope is well-established. Chemical shift perturbation studies reveal that complex formation with single-stranded nucleic acids affects the environments of backbone amide ^{15}N and H^N resonances that are mostly clustered together on one face of the CspA molecule (*9*), as shown in Figure 6D. Supporting evidence for this nucleic acid-binding epitope comes from electostatic field calculations on both *E. coli* CspA and *B. subtilis* CspB indicating that this face of these homologous molecules contains a strongly polarized positively charged electrostatic field appropriate for attracting polyanionic nucleic acid molecules (*10*). Mutational analysis of CspB−oligonucleotide interactions has confirmed the location of this binding epitope by demonstrating that surface aromatic or basic groups of residues (CspA numbering) Lys-10, Trp-11, Lys-16, Phe-18, Phe-20, Phe-31, Phe-34, and His-33 are required for formation of cross-linked CspB−ssDNA complexes (*11*). Mutational analysis of the free energies of CspA−ssDNA interactions further demonstrates roles for surface aromatic residues Phe-18, Phe-20, and Phe-31 in forming the nucleic acid-binding epitope (*20*). In addition, fluorescence quenching experiments demonstrate significant changes in the solvent accessibility of the indole side chain of residue Trp-11 upon complex formation (*20, 94*). Taken together, these results demonstrate that the single-stranded nucleic acid-binding epitope of CspA is localized within β-strand segments Val-9−Asn-13 (in β1″), Phe-18−Ile-21 (in β2), and Val-30−His-33 (in β3), together with surface loops Ala-14−Gly-17, Phe-34−Gln-49, and Gly-58−Gly-61. The surface aromatic and basic residues within this single-stranded oligonucleotide binding epitope are also strongly conserved in the family of the cold-shock domain that spans a wide range of organisms from bacteria to humans (Figure 1), and include the strongly conserved RNP1 and RNP2 sequence motifs. Conserved surface-exposed aromatic residues of strands β1″, β2, and β3 within this binding epitope (Figure 6D) form a hydrophobic face suitable for aromatic/base stacking interactions with splayed conformations of single-stranded nucleic acids.

Functional Significance of Internal Motions in CspA. The ^{15}N relaxation studies complement our solution structure analysis by providing evidence for internal motions within this nucleic acid-binding epitope of unliganded CspA. In particular, polypeptide segments Asn-39−Lys-43 and Gly-58−Gly-61, corresponding to two surface loops within the epitope, exhibit relatively small values of backbone N−H bond vector order parameters (Figures 6B and 7), an indication of relatively large motions on the sub-nanosecond time scale. In addition, backbone ^{15}N atoms within polypeptide segment Lys-10−Asn-13 exhibit unusually large R_{ex} terms indicative of chemical exchange broadening (Figures 6C and 7). Several other N−H sites within the RNP1 and RNP2 sequence motifs (including Phe-18, Gly-19, Phe-31, and His-33) also exhibit significant (2−6 Hz) R_{ex} values. The backbone ^{15}N atom of residue Lys-60 also has a relatively high R_{ex} value, 4.6 Hz.

Although the precise motions that give rise to these dynamic features of the NMR data are not yet determined, one reasonable mechanism that accounts for much of the apparent exchange broadening of ^{15}N resonances involves conformational fluctuations of the Trp-11 side chain conformation on the microsecond to millisecond time scale. For each of the corresponding multiple side chain conformers of Trp-11, aromatic ring current shifts would result in significant chemical shift differences for backbone amide resonances in polypeptide segment Lys-10−Asn-13 and within the nearby RNP1 and RNP2 sequence motifs. These different side chain conformers would also exhibit different $^{15}N^{\epsilon}-^{1}H$ chemical shifts, and hence exchange broadening, for the side chain indole $^{15}N^{\epsilon}-^{1}H$ resonances. Regardless of the precise nature of these fast (i.e., sub-nanosecond) and

slow (i.e., microsecond to millsecond) time scale motions of surface loops and side chains within the nucleic acid-binding epitope, it is likely that these dynamics will be altered by nucleic acid interactions. The resulting changes in dynamics upon complex formation will contribute to changes in entropy associated with complex formation, and therefore to the overall thermodynamics of the binding process.

Conclusion. This paper describes nearly complete ^{1}H, ^{13}C, and ^{15}N resonance assignments and a solution NMR structure of *E. coli* Csp A. Structural database searches reveal that the CspA chain fold is used by many other single-stranded nucleic acid-binding proteins and demonstrates how, in the absence of clear sequence homology, analysis of structural homology can provide important clues about relationships between protein functions. ^{15}N relaxation data reveal relatively large amplitude sub-nanosecond motions of surface loops within the single-stranded nucleic acid-binding epitope of CspA. Slower motion dynamics, on the microsecond to millisecond time scale, are also observed for backbone N−H vectors in the vicinities of residues Trp-11 and Lys-60. These portions of CspA are all within the nucleic acid-binding epitope. Changes in these dynamics of CspA are proposed to contribute to changes in entropy upon complex formation, and thus to modulate binding affinity and function. Further studies are now in progress to better characterize the kinetics of these dynamics in free and nucleic acid-bound CspA and to determine their precise role in the thermodynamics of nucleic acid recognition.

ACKNOWLEDGMENT

We thank M. Andrec, B. Celda, K. Foygel, Y. Huang, C. B. Ríos, P. Sahasrabudhe, G. V. T. Swapna, and R. Watson for helpful discussions and comments on the manuscript. Mass spectroscopy and amino acid composition analyses were carried out by the Cornell University Biotechnology Analytical Protein Chemistry Facility.

SUPPORTING INFORMATION AVAILABLE

Detailed information on NMR data collection parameters, pulse sequence diagrams, resonance assignments (^{1}H, ^{13}C, and ^{15}N), ^{15}N relaxation data, motional parameters, and Ramachandran plot (32 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Lindquist, S., and Craig, E. A. (1988) *Annu. Rev. Genet. 22*, 631−677.
2. Jones, P. G., and Inouye, M. (1994) *Mol. Microbiol. 11*, 811−818.
3. Yamanaka, K., Fang, L., and Inouye, M. (1998) *Mol. Microbiol. 27*, 247−256.
4. Jones, P. G., Van Bogelen, R. A., and Neidhardt, F. C. (1987) *J. Bacteriol. 172*, 2092−2095.
5. Goldstein, J., Pollitt, N. S., and Inouye, M. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 283−287.
6. Lee, S. J., Xie, A., Jiang, W., Etchegaray, J. P., Jones, P. G., and Inouye, M. (1994) *Mol. Microbiol. 11*, 833−839.
7. Wolffe, A. P., Tafuri, S., Ranjan, M., and Familari, M. (1992) *New Biol. 4*, 290−298.
8. Wolffe, A. P. (1994) *BioEssays 16*, 245−251.
9. Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S. D., Inouye, M., and Montelione, G. T. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5114−5118.
10. Schindelin, H., Jiang, W., Inouye, M., and Heinemann, U. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5119−5123.
11. Schröder, K., Graumann, P., Schnuchel, A., Holak, T. A., and Marahiel, M. A. (1995) *Mol. Microbiol. 16*, 699−708.
12. Jiang, W., Fang, L., and Inouye, M. (1996) *J. Bacteriol. 178*, 4919−4925.
13. Jiang, W., Hou, Y., and Inouye, M. (1997) *J. Biol. Chem. 272*, 196−202.
14. Schindelin, H., Marahiel, M. A., and Heinemann, U. (1993) *Nature 364*, 164−168.
15. Schnuchel, A., Wiltscheck, R., Czisch, M., Herrler, M., Willimsky, G., Graumann, P., Marahiel, M. A., and Holak, T. A. (1993) *Nature 364*, 169−171.
16. Murzin, A. G., and Chothia, C. (1992) *Curr. Opin. Struct. Biol. 2*, 895−903.
17. Murzin, A. G. (1993) *EMBO J. 12*, 861−867.
18. Bandziulis, R. J., Swanson, M. S., and Dreyfuss, G. (1989) *Genes Dev. 3*, 431−437.
19. Burd, C. G., and Dreyfuss, G. (1994) *Science 265*, 615−621.
20. Hillier, B. J., Rodriguez, H. M., and Gregoret, L. M. (1998) *Folding Des. 3*, 87−93.
21. Chatterjee, S., Jiang, W., Emerson, S. D., and Inouye, M. (1993) *J. Biochem. 114*, 663−669.
22. Studier, F. W., and Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113−130.
23. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning*, Cold Spring Harbor Laboratory Press, Plainview, NY.
24. Tashiro, M., Rios, C. B., and Montelione, G. T. (1995) *J. Biomol. NMR 6*, 211−216.
25. Feng, W., Rios, C. B., and Montelione, G. T. (1996) *J. Biomol. NMR 8*, 98−104.
26. Rios, C. B., Feng, W., Tashiro, M., Shang, Z., and Montelione, G. T. (1996) *J. Biomol. NMR 8*, 345−350.
27. Shaka, A. J., Barker, P. B., and Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
28. Nagayama, K. (1986) *J. Magn. Reson. 69*, 508.
29. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
30. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR 6*, 135−140.
31. Zimmerman, D. E., Kulikowski, C. A., Wang, L., Lyons, B. A., and Montelione, G. T. (1994) *J. Biomol. NMR 4*, 241−256.
32. Zimmerman, D. E., Kulikowski, C. A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C. Y., Powers, R., and Montelione, G. T. (1997) *J. Mol. Biol. 269*, 592−610.
33. Logan, T. M., Olejniczak, E. T., Xu, R. X., and Fesik, S. W. (1992) *FEBS Lett. 314*, 413−418.
34. Montelione, G. T., Emerson, S. D., and Lyons, B. A. (1992) *Biopolymers 32*, 327−334.
35. Lyons, B. A., and Montelione, G. T. (1993) *J. Magn. Reson. 101*, 206−209.
36. Celda, B., and Montelione, G. T. (1993) *J. Magn. Reson. B101*, 189−193.
37. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
38. Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., and Gronenborn, A. M. (1990) *Biochemistry 29*, 3542−3556.
39. Billeter, M., Neri, D., Otting, G., Qian, Y. Q., and Wüthrich, K. (1992) *J. Biomol. NMR 2*, 257−274.
40. Moy, F. J., Li, Y.-C., Rauenbuehler, P., Winkler, M. E., Scheraga, H. A., and Montelione, G. T. (1993) *Biochemistry 32*, 7334−7353.
41. Tejero, R., Monleon, D., and Montelione, G. T. (1998) *J. Biomol. NMR* (submitted for publication).
42. Braun, W., and Go, N. (1985) *J. Mol. Biol. 186*, 611−626.
43. Güntert, P., Braun, W., and Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.
44. Wüthrich, K., Billeter, M., and Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.
45. Li, Y.-C., and Montelione, G. T. (1994) *J. Magn. Reson. B105*, 45−51.

46. Li, Y.-C., and Montelione, G. T. (1995) *Biochemistry 34*, 2408−2423.
47. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) *J. Mol. Biol. 246*, 144−163.
48. Nicholson, L. K., Kay, L. E., Baldisseri, D. M., Arango, J., Young, P. E., Bax, A., and Torchia, D. A. (1992) *Biochemistry 31*, 5253−5263.
49. Hiyama, Y., Niu, C., Silverton, J. V., Bavoso, A., and Torchia, D. (1988) *J. Am. Chem. Soc. 110*, 2378−2383.
50. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4558.
51. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559−4570.
52. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 4989−4991.
53. Palmer, A. G., Rance, M., and Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.
54. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986) *Numerical Recipes*, Cambridge University Press, Cambridge, England.
55. Montelione, G. T., Arnold, E., Meinwald, Y. C., Stimson, E. R., Denton, J. B., Huang, S.-G., Clardy, J., and Scheraga, H. A. (1984) *J. Am. Chem. Soc. 106*, 7946−7958.
56. Spera, S., and Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490−5492.
57. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) *Protein Sci. 1*, 736−751.
58. Laskowski, R. A., MacArthur, M. W., Moss, M. D., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
59. Cantor, C. R., and Schimmel, P. R. (1980) *Biophys. Chem. 2*, 560−565.
60. Zimmerman, D. E., and Montelione, G. T. (1995) *Curr. Opin. Struct. Biol. 5*, 664−673.
61. Shimotakahara, S., Rios, C. B., Laity, J. H., Zimmerman, D. E., Scheraga, H. A., and Montelione, G. T. (1997) *Biochemistry 36*, 6915−6929.
62. Laity, J. H., Lester, C. C., Shimotakahara, S., Zimmerman, D. E., Montelione, G. T., and Scheraga, H. A. (1997) *Biochemistry 36*, 12683−12699.
63. Tashiro, M., Tejero, R., Zimmerman, D. E., Celda, B., Nilsson, B., and Montelione, G. T. (1997) *J. Mol. Biol. 272*, 573−590.
64. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
65. Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M., and Murzin, A. G. (1997) *Cell 88*, 235−242.
66. Yang, C., Curth, U., Urbanke, C., and Kang, C. (1997) Brookhaven Protein Data Base access code 3ull.
67. Sette, M., van Tilborg, P., Spurio, R., Kaptein, R., Paci, M., Gualerzi, C. O., and Boelens, R. (1997) *EMBO J. 16*, 1436−1443.
68. Folkers, P. J. M., Nilges, M., Folmer, R. H. A., Konings, R. N. H., and Hilbers, C. W. (1994) *J. Mol. Biol. 236*, 229−246.
69. Folmer, R. H. A., Nilges, M., Folkers, P. J. M., Konings, R. N. H., and Hilbers, C. W. (1994) *J. Mol. Biol. 240*, 341−357.
70. Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. M., Gray, C. W., Konings, R. N. H., Wang, A. H. J., and Terwilliger, T. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2071−2075.
71. Su, S., Gao, Y. G., Zhang, H., Terwilliger, T. C., and Wang, A. H. J. (1997) Brookhaven Protein Data Base access code 1gvp.
72. Folmer, R. H. A., Nilges, M., Konings, R. N. H., and Hilbers, C. W. (1995) *EMBO J. 14*, 4132−4142.
73. Aevarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Y., Al-Karadaghi, S., Svensson, A., and Liljas, A. (1994) *EMBO J. 13*, 3669−3677.
74. Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J. C., and Moras, D. (1994) *EMBO J. 13*, 327−337.
75. Onesti, S., Miller, A. D., and Brick, P. (1995) *Structure 3*, 163−176.
76. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997) *Nature 385*, 176−181.
77. Loll, P. J., and Lattman, E. E. (1989) *Proteins: Struct., Funct., Genet. 5*, 183−201.
78. Yu, L., Zhu, C. X., Tse-Dinh, Y. C., and Fesik, S. W. (1995) *Biochemistry 34*, 7622−7628.
79. Golden, B. L., Hoffman, D. W., Ramakrishnan, V., and White, S. W. (1993) *Biochemistry 32*, 12812−12820.
80. Briercheck, D. M., Allison, T. J., Richardson, J. P., Ellena, J. F., Wood, T. C., and Rule, G. S. (1996) *J. Biomol. NMR 8*, 429−444.
81. Jones, P. G., Mitta, M., Kim, Y., Jiang, W., and Inouye, M. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 76−80.
82. Acharya, K. R., Passalacqua, E. F., Jones, E. Y., Harlos, K., Stuart, D. I., Brehm, R. D., and Tranter, H. S. (1994) *Nature 367*, 94−97.
83. Martinez, S. E., Huang, D., Szczepaniak, A., Cramer, W. A., and Smith, J. L. (1994) *Structure 2*, 95−105.
84. Stein, P. E., Boodhoo, A., Armstrong, G. D., Cockle, S. A., Klein, M. H., and Read, R. J. (1994) *Structure 2*, 45−57.
85. Stein, P. E., Boodhoo, A., Tyrrell, G. J., Brunton, J. L., and Read, R. J. (1992) *Nature 355*, 748−750.
86. Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., and Keene, J. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495−2499.
87. Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G., and Mueller, L. (1992) *Biochemistry 31*, 6254−6265.
88. Lee, A. L., Kanaar, R., Rio, D. C., and Wemmer, D. E. (1994) *Biochemistry 33*, 13775−13786.
89. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) *Nature 372*, 432−438.
90. Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D., and Varani, G. (1996) *Nature 380*, 646−650.
91. Nagai, K. (1996) *Curr. Opin. Struct. Biol. 6*, 53−61.
92. Xu, R.-M., Jokhan, L., Cheng, X., Mayeda, A., and Krainer, A. R. (1997) *Structure 5*, 559−570.
93. Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L., and Dreyfuss, G. (1992) *EMBO J. 11*, 3289−3295.
94. Feng, W. (1997) Ph.D. Dissertation, Rutgers University, Piscataway, NJ.
95. Carson, M. (1991) *J. Appl. Crystallogr. 24*, 958−961.

BI980269J