NMR Studies of the Phosphotransfer Domain of the Histidine Kinase CheA from *Escherichia coli*: Assignments, Secondary Structure, General Fold, and Backbone Dynamics†

Hongjun Zhou,‡ David F. Lowry,‡ Ronald V. Swanson,§,|| Melvin I. Simon,§ and Frederick W. Dahlquist*,‡

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, and Department of Biology, California Institute of Technology, Pasadena, California 91125

Received May 25, 1995; Revised Manuscript Received August 17, 1995⊗

ABSTRACT: Multidimensional heteronuclear NMR techniques were applied to study the phosphotransfer domain, residues 1−134, of the histidine kinase CheA, from *Escherichia coli*, which contains the site of autophosphorylation, His48. Assignments of the backbone amide groups and side chain protons are nearly complete. Our studies show that this protein fragment consists of five α-helices (A−E) connected by turns. Analysis of NOE distance restraints provided by two-dimensional (2D) ^{1}H−^{1}H and three-dimensional (3D) ^{15}N-edited NOESY spectra using model building and structure calculations indicates that the five helices form an antiparallel helix bundle with near-neighbor connectivity. The amino-terminal four helices are proposed to be arranged in a right-handed manner with helix E packing against helices C and D. From ideal hydrophobic helical packing and structure calculations, the site of autophosphorylation, His48, is nearly fully exposed to the solvent. We measured the NMR relaxation properties of the backbone ^{15}N nuclei using inverse detected two-dimensional NMR spectroscopy. The protein backbone dynamics studies show that $CheA_{1-134}$ is formed into a tight and compact structure with very limited flexibilities both in helices and turns. Structural implications of titration and phosphorylation experiments are briefly discussed.

Escherichia coli is able to respond to gradients of chemical stimuli by altering its swimming behavior to effect net migration toward attractants and away from repellents. This process is controlled by the chemotaxis system in which the histidine autokinase CheA plays a central role. The signal transduction cascade begins with changes in ligand binding to the periplasmic domains of the membrane-spanning receptors which form ternary complexes with the kinase CheA and the coupling protein CheW (Gegner et al., 1992). Changes in the receptor−ligand binding state are transmitted to the cytoplasmic domains of the receptors, and the autophosphorylation rate of CheA is presumed to be regulated through conformational changes within the protein complex (Milburn et al., 1992; Borkovich et al., 1989; Borkovich & Simon, 1990; Ninfa et al., 1991). Phosphorylated CheA donates its phosphoryl group to the response regulator proteins CheY and CheB which control the direction of flagellar rotation and the level of receptor adaptation, respectively (Hess et al., 1988; Wylie et al., 1988).

The chemotaxis system belongs to the two-component family in which CheA, CheY, and CheB share homology with proteins in other bacterial regulatory pathways [for reviews, see Stewart and Dahlquist (1987), Stock et al. (1989), Bourret et al. (1991), and Parkinson and Kofoid (1992)]. Intact CheA protein from *E. coli* has 654 residues, which is large for a high-resolution NMR[1] approach. It was shown that CheA has five functionally distinct and separable regions linearly encoded along its polypeptide chain (Swanson et al., 1993; Bourret et al., 1990; Stock et al., 1989; Parkinson & Kofoid, 1992; Morrison & Parkinson, 1994): phosphotransfer (P1), CheY binding (P2), kinase activity, receptor coupling, and CheW coupling. Isolated domain fragments retain their respective functions (Swanson et al., 1993). For instance, the amino-terminal 134 residues ($CheA_{1-134}$) encode the phosphotransfer function, residues 160−227 encode the CheY-binding function, and phosphorylated $CheA_{1-233}$ is able to bind CheY and transfer its phosphoryl group to CheY (Swanson et al., 1993; Morrison & Parkinson, 1994). This property makes it possible to separately study these functional domains in aqueous solution by high-resolution NMR techniques.

We have expressed and studied the phosphotransfer domain $CheA_{1-134}$, which carries the autophosphorylation site His48, by using heteronuclear two- and three-dimensional NMR techniques. In this paper we report progress in backbone and side chain assignments, structure determination, and backbone dynamics studies of the protein fragment.

† D.F.L. was supported by a Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation Fellowship DRG-1195. R.V.S. was supported by a National Research Service Award Fellowship GM14767 from NIH. This work was supported by Research Grants GM33677 (F.W.D.) and AI-19296 (M.I.S.).

‡ University of Oregon.

§ California Institute of Technology.

* Author to whom correspondence should be addressed.

|| Present address: Department of Molecular Biology, Industrial BioCatalysis Inc., 512 Elmwood Ave., Sharon Hill, PA 19079.

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1995.

[1] Abbreviations: HSMQC, heteronuclear single-multiple-quantum coherence; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; DQF-COSY, double-quantum-filtered correlation spectroscopy; 2D, two dimensional; 3D, three dimensional; NMR, nuclear magnetic resonance; TPPI, time-proportional phase incrementation; SA, simulated annealing.

0006-2960/95/0434-13858$09.00/0 © 1995 American Chemical Society

MATERIALS AND METHODS

Sample Preparation. To express uniformly ^{15}N-labeled $CheA_{1-134}$, strain M15/pRS1-6/pREP4 (Swanson et al., 1993) was inoculated into five 1 L flasks of ^{15}N minimal media [per liter: 10.5 g of K_2HPO_4, 4.5 g of KH_2PO_4, 1 g of (^{15}-$NH_4)_2SO_4$, 0.5 g of citrate, 5 mL of 0.1% thiamine, 1 mL of 1 M $MgCl_2$, 10 mL of 50% glucose, 1 mL of 0.5% $FeCl_3$, and 0.1 mL of 0.1 M $CaCl_2$] and grown at 37 °C. At OD_{600} = 0.5, the cultures were induced with the addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 2 mM. Growth was continued for an additional 5 h following induction. Cells were harvested by centrifugation and frozen. $CheA_{1-134}$ was purified essentially as described by Swanson et al. (1993). Cell pellets were resuspended in 50 mM Tris, pH 7.5, 2 mM 2-mercaptoethanol, 5 mM EDTA, 0.1 mg/mL lysozyme, and 1 mM phenylmethanesulfonyl fluoride (PMSF) and then sonicated briefly. Lysates were cleared by centrifugation at 123000g for 30 min. Pelleted material was discarded, and 0.314 g/mL $(NH_4)_2SO_4$ was added to the supernatant to bring it to 50% saturation. This solution was maintained on ice for 3 h and pelleted, and the supernatant was then brought to 60% saturation by the addition of 0.065 g/mL $(NH_4)_2SO_4$. The pellet was collected by centrifugation. Precipitated protein was resuspended and dialyzed against Tris, pH 8.0, 2 mM 2-mercaptoethanol, and 5 mM EDTA. Fractions were assayed by SDS−PAGE, and those containing $CheA_{1-134}$ judged to be pure were pooled and concentrated using a Centriprep-10 concentrator. Purified protein was dialyzed against 50 mM $NaPO_4$, pH 8.0.

To express [^{15}N]leucine-labeled $CheA_{1-134}$, strain D139/pREP4/pRS1-6 was inoculated into five 1 L flasks of defined rich media (Muchmore et al., 1989) containing 100 mg/L ^{15}N-labeled leucine (Isotec). Cells were grown at 37 °C until reaching an OD_{600} = 0.8, at which time protein production was induced by the addition of IPTG to a final concentration of 2 mM. After 2 h of induction, cells were harvested and frozen. [^{15}N]Leu-labeled $CheA_{1-134}$ was purified as described above.

Protein samples were concentrated to 500 μL with a final concentration of approximately 2 mM each. Five percent D_2O and 0.02% sodium azide were added to the final NMR samples. Uniformly labeled sample was later exchanged through a Sephadex G-25 spin column to pH 6.3, 50 mM $NaPO_4$ buffer.

NMR Data Collection and Processing. All data reported in this paper were collected on a General Electric Omega 500 MHz spectrometer at pH 6.3, 30 °C, except as otherwise stated. The chemical shifts were set to an external proton reference of sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) at 0.0 ppm and an external nitrogen standard of ^{15}-NH_4Cl at 24.93 ppm relative to $^{15}NH_3$. In all experiments, proton excitation was applied at water proton frequency of 4.76 ppm, and ^{15}N excitation was applied at 116 ppm. Water suppression was usually achieved by selective weak presaturation (Zuiderweg et al., 1986). In the ^{15}N experiments 4.5 ms was used as the nominal $1/2J_{NH}$. Data were processed using FELIX software (Hare Research, Biosym Technologies, San Diego, CA).

Two-dimensional (2D) ^{1}H−^{15}N correlation spectra were collected by a heteronuclear single-multiple quantum coherence (HSMQC) sequence (Zuiderweg, 1990). A total of 256 real t_1 (^{15}N) and 1024 complex t_2 (^{1}H) points were collected with spectral widths of 2082.6 and 6666.7 Hz, respectively. Zero-filling was employed to yield an absorptive 2D matrix of 256 × 1024 data points.

2D ^{15}N-edited NOESY experiments were carried out as described by McIntosh et al. (1990) at pH 8 and 6.3, with NOE mixing times of 150 or 120 ms. A total of 512 real t_1 (^{15}N) and 1024 complex t_2 (^{1}H) points were collected with spectral widths of 2083.2 and 6666.7 Hz, respectively. Zero-filling was applied to yield a final real matrix of 512 × 1024 data points.

For all 3D ^{15}N-edited NOESY-HSMQC (Kay et al., 1989b; Zuiderweg et al., 1989) and TOCSY-HSMQC (Cavanagh et al., 1989; Marion et al., 1989b) experiments, 256 t_1 (^{15}N) and 64 t_2 (^{1}H) real points were collected in the transient ^{15}N and ^{1}H dimensions with spectral widths of 1041.6 and 6666.7 Hz, respectively, and 1024 complex points were collected in FIDs with a spectral width of 6666.7 Hz. Sixteen scans were applied for each transient real (t_1,t_2) point. The NOESY mixing times were 150 or 100 ms. The TOCSY mixing was achieved by a MLEV-17 pulse train and was applied for 60 ms. Quadrature detection in t_1 and t_2 was achieved using TPPI (Marion & Wüthrich, 1983) or States−TPPI (Marion et al., 1989b) methods. Zero-filling was applied to give absorptive 3D matrixes of 256 × 64 × 1024 data points each.

To measure $^3J_{HN\alpha}$ coupling constants, ^{1}H−^{15}N HMQC-J (Kay & Bax, 1990) and HSQC-J (Billeter et al., 1992) experiments were employed. In the HMQC-J experiment a total of 1024 real t_1 (^{15}N) and 1024 complex t_2 (^{1}H) data points were collected with spectral widths of 2083.2 and 6666.7 Hz, respectively. Zero-filling was employed to yield an absorptive 2D matrix of 1024 × 1024 data points. In the HSQC-J experiment, a series of HSQC-J data sets were collected with eight time delays: 15, 30, 50, 70, 80, 90, 115, and 130 ms. For each spectrum a total of 256 real t_1 (15N) and 1024 complex t_2 (^{1}H) points were used with spectral widths of 1041.6 and 6666.7 Hz, respectively. Sine bell 30° and 40° shifts were applied along ^{1}H and ^{15}N dimensions, respectively, and zero-filling was employed to yield 2D matrixes of 512 × 1024 real data points each.

To calculate the coupling constants from the HSQC-J experiments, $^3J_{HN\alpha}$ was varied from 0 to 15 Hz in a 0.1 Hz step size, and peak intensities were fit according to (Billeter et al., 1992)

$$I = I_0 \cos(\pi^3 J_{HN\alpha}\tau) \cos(\pi^3 J_{HN\alpha}\tau_0) \exp(-\tau/T) \quad (1)$$

where I_0 is the maximal peak intensity, τ is the variable time delay, $\tau_0 = 1/2J_{HN}$, and T is the effective decay constant. I_0 and T were fitted at a fixed $^3J_{HN\alpha}$ value. The best $^3J_{HN\alpha}$ was taken as the one that gave the least square deviation $\chi^2(J)$. $^3J_{HN\alpha}$ values at $2\chi^2(J)$ were considered as the lower and upper bounds of the best fit value. Typically, only larger than 5 Hz splitting can possibly be resolved in the HMQC-J experiment, while as small as 1 Hz of $^3J_{HN\alpha}$ can be detected in the HSQC-J experiment though with a 2−3 Hz error range. For J > 5 Hz, the error range is approximately ±0.3 Hz around the best fit value.

Backbone dynamics studies measured the NMR relaxation properties of the backbone ^{15}N nuclei using inverse-detected two-dimensional NMR spectroscopy (Kay et al., 1989a; Clore et al., 1990; Barbato et al., 1992). Pulse sequences

used were those of Barbato et al. (1992) with slight modifications. To obtain T_1 relaxation rates, seven delays of 28, 70, 140, 210, 308, 505, and 841 ms were employed. To obtain T_2 relaxation rates seven delays of 7.6, 23, 46, 69, 99, 122, and 161 ms were employed. Thirty-two scans were acquired for each real t_1 (^{15}N) point. To measure ^{15}N-{^{1}H} NOE, two spectra were recorded. For the spectrum with the NOE, protons were kept in a saturated state during the 3 s predelay by a high power 120° pulse train. For the spectrum without the NOE effect, water signal suppression was achieved by combined use of two 10 ms high power scrambling pulses (Messerle et al., 1989; Barbato et al., 1992) of identical phases separated by a 1 ms delay and followed by a 100 ms weak water saturation right before the first ^{15}N pulse. Sixty-four scans were collected for each real t_1 point in the NOE measurement. All T_1, T_2, and NOE data sets were acquired with 256 real t_1 points and 1024 complex t_2 points. The spectral widths along the t_1 (^{15}N) and t_2 (^{1}H) dimensions are 1041.6 and 6666.7 Hz, respectively. Each data set was processed with sine bell apodization along the ^{1}H and ^{15}N dimensions, and each dimension was zero-filled to give a final 512 × 1024 real matrix. Maximum peak height was searched and a simple parabolic interpolation was applied to estimate the real peak height. This was carried out by an in-house program written to run within FELIX 2.30 (Biosym Technologies). T_1 and T_2 values were then fit by a simple exponential decay function using Levenberg−Marquardt nonlinear fitting scheme (Press et al., 1990).

Relaxation data were analyzed using the "model free" approach of Lipari and Szabo (1982). Equations and parameters relating T_1, T_2, and NOE factor of an amide ^{15}N spin are identical to eqs 1−3 of Kay et al. (1989a). In the interpretation of the relaxation data we have assumed that the relaxation properties of the ^{15}N spins are governed solely by the ^{1}H−^{15}N dipolar coupling and the chemical shift anisotropy (CSA) interaction. We also have ignored any anisotropy in the molecular motion. Under these assumptions, for a spherical molecule with overall rotational correlation time τ_m and effective correlation time τ_e of rapid internal motions, the spectral density function is given by (Lipari & Szabo, 1982)

$$J(\omega) = S^2\,\tau_m/[1 + (\omega\tau_m)^2] + (1 - S^2)\tau/[1 + (\omega\tau)^2] \quad (2)$$

where $1/\tau = 1/\tau_m + 1/\tau_e$, and S^2 is the generalized order parameter, which describes the amplitude of the internal motions. To a first-order approximation, the contribution of the fast internal motions to the spectral density can be ignored (Kay et al., 1989a; Clore et al., 1990; Barbato et al., 1992). This enabled us to calculate τ_m from T_1/T_2 ratio. In the present study, we derived S^2 and τ_e, by minimizing the function given by (Barbato et al., 1992)

$$f(S^2,\tau_e) = [(T_{1,calc} - T_{1,meas})/T_{1,calc}]^2 + [(T_{2,calc} - T_{2,meas})/T_{2,calc}]^2 + [(NOE_{meas} - NOE_{calc})/NOE_{calc}]^2/4 \quad (3)$$

where "meas" and "calc" refer to measured and calculated values, respectively. This function represents the total relative deviation of theoretical values from experimental values. The contribution from the NOE data is scaled down by a factor of 4 because of the larger experimental error in the measured NOE.

A series of ^{1}H−^{15}N HSMQC experiments (Zuiderweg, 1990) were used to identify slowly exchanging amide protons. A total of 256 real t_1 (^{15}N) points and 1024 complex t_2 (^{1}H) points were collected with spectral widths of 1041.6 and 6666.7 Hz, respectively. Data acquisitions for the HSMQC spectra were started 5, 24, 43, 80, 116, 185, 331, and 474 min after the protein was transferred into D_2O through a Sephadex G-25 spin column. ^{1}H−^{1}H phase-sensitive double-quantum-filtered COSY (DQF-COSY; Shaka & Freeman, 1983; Rance et al., 1984), and NOESY (Macura & Ernst, 1980; Kumar et al., 1980; Bodenhausen et al., 1984) spectra were carried out on this sample. For both spectra, 512 real t_1 and 1024 complex t_2 points were collected with spectral widths of 6666.7 Hz in both dimensions. Zero-filling was used to yield two spectra of 512 × 1024 real data points each.

Structure Calculations. Structural studies of $CheA_{1-134}$ were performed both by simple model building and structure calculations using hybrid distance geometry−dynamical simulated annealing methods (Nilges et al., 1988) carried out by the program X-PLOR 3.1 (Brünger, 1992). In the structure calculations, the dihedral angles ϕ and ψ of the helical regions, as judged from the NOE data and $^3J_{HN\alpha}$ values, were restricted within the typical ranges for ideal α-helices. The minimum ranges employed for ϕ and ψ were ±30° and ±50°. All well resolved and assigned NOEs, including approximately 600 sequential ($|i - j| = 1$), 200 medium-range ($1 < |i - j| < 5$), and 100 long-range ($|i - j| > 5$) NOEs, were used as distance restraints. The NOEs were initially classified as strong, medium, and weak corresponding to interproton distance restraints of 1.8−2.9, 1.8−3.5, and 1.8−5.0 Å, respectively. The distance restraints that lie at the boundary of two distance ranges and were systematically violated in the initial stages of simulated annealing (SA) refinement were subsequently modified according to Powers et al. (1993). Hydrogen bonding restraints for those amides that are protected from exchange with solvent were introduced subsequent to the initial set of calculations. These amides were considered as those still present in the first HSMQC spectrum taken after the protein sample was transferred into D_2O. Two distance restraints were used for each hydrogen bond: one between the hydrogen and acceptor atom of 1.5−2.3 Å and one between the donor heavy atom and the acceptor atom of 2.4−3.3 Å.

The structure calculations began with generation of an initial set of substructures incorporating only N, C, C^αH, C^β, nonterminal C^γ and C^δ atoms, and a pseudoatom for the aromatic rings (Nilges et al., 1988), followed by distance-geometry simulated annealing with all atoms. Finally, iterative simulated annealing refinement was employed to gradually improve the qualities of the initially calculated structures. The target function that is minimized during simulated annealing and conventional Powell minimization comprises only quadratic harmonic terms for covalent geometry, square-well quadratic potentials for the experimental distance and dihedral angle restraints, and a quartic van der Waals repulsion term for the nonbonded contacts. All peptide bonds were restrained to be planar and trans. The final force constants for the NOE-derived distance restraints, the dihedral angle restraints, and the van der Waals repulsion term were set to 50 kcal mol^{-1} Å^{-2}, 200 kcal mol^{-1} rad^{-2}, and 4 kcal mol^{-1} Å^{-4}, respectively. The van der

Waals hard sphere radii were set to 0.75 times those in the X-PLOR parameter file.

RESULTS

Identification of ^1H Spin Systems and Assignments of Side Chain Groups. The identification of the majority of the spin systems from the 2D $^1H-^1H$ DQF-COSY spectrum taken on the sample in D_2O was extremely difficult because of the overcrowding of the spectrum. In spite of this difficulty, the 17 leucines were easily identified by comparing the $^1H-^{15}N$ HSMQC spectrum from the ^{15}N-uniformly labeled protein sample with that from the [^{15}N]leucine specifically labeled sample. In addition, 16 Ala, six Gly, two Val, and four out of seven Thr residues were identified in the 3D ^{15}N-edited TOCSY-HSMQC spectrum by their unique spin system patterns in the early stages of the assignment process. From the TOCSY data all the $C^\alpha H$ protons of the identified residues, and a great majority of the $C^\beta H$ protons of the AMX spin systems (Ser, Cys, Asp, Asn, Phe, Tyr, and His), which account for 35 residues of the $CheA_{1-134}$, were readily assigned. For those residues with long side chains (Glu, Gln, Met, Arg, Lys, Leu, and Ile) the assignments of the TOCSY cross peaks to specific side chain protons were tentative. Because of the inefficient coherence mixing, some of the cross peaks are simply missing in the ^{15}N-edited TOCSY. Some of these missing peaks were indirectly assigned in the 3D ^{15}N-edited NOESY after the sequential connectivity was established (indicated by "*" in Table S1 in the Supporting Information).

In the 3D ^{15}N-edited NOESY spectrum, the amide protons of the eight Phe and three Tyr residues and their C-terminal adjacent residues carry cross peaks in the range of 6.5−7.5 ppm, which cannot be traced to any other NH or NH_2 groups. These cross peaks were assigned to the amide protons and the ring δ protons of the respective Phe and Tyr residues. These assignments are supported by the $^1H-^1H$ NOESY spectrum, in which clear NOE cross peaks were observed between these ring protons and the $C^\beta H$ and $C^\alpha H$ protons of the respective Phe and Tyr residues. Accordingly, all the $C^\delta H$ and $C^\epsilon H$ protons of these residues were assigned despite substantial chemical shift degeneracy in the aromatic region. The $C^\zeta H$ protons of the eight Phe residues have not been identified.

For the three histidines, H26, H48, and H67, only the $C^\delta H$ protons of H26 and H67 display weak NOEs to their $C^\alpha H$ and $C^\beta H$ protons in the 2D $^1H-^1H$ NOESY spectrum. In the 2D and 3D ^{15}N-edited NOESY spectra, the $C^\delta H$ proton of His48 displays weak NOEs to the amide protons of His48 and Ser49. Recently, all the $C^\delta H$ and $C^\epsilon H$ protons and δ and ϵ nitrogens of the imidazole rings have been identified through the titration experiments by taking 2D $^1H-^{15}N$ HMQC spectra at various pH values (H. Zhou and F. W. Dahlquist, unpublished results). It was observed that, at pH 6.3, the chemical shifts of the three $C^\delta H$ and three $C^\epsilon H$ protons from the imidazole rings are nearly degenerate, respectively, at 7.27 and 8.27 ppm. The nitrogen bonded protons have not been observed.

The amide side chain groups of Asn and Gln were assigned to specific residues from the observation of intraresidue NOEs from the NH_2 protons to the $C^\beta H$ and/or $C^\gamma H$ protons, respectively, in a 3D ^{15}N-edited NOESY spectrum.

Sequence-Specific Residue Assignments. Figure 1 shows a $^1H-^{15}N$ correlation spectrum labeled with the assignments of the backbone amide groups and some of the NH_2 groups of the Asn and Gln residues. The assignments of the backbone amide groups of $CheA_{1-134}$ were made from examination of the 2D $^1H-^{15}N$ NOESY, 3D ^{15}N-edited NOESY and TOCSY spectra taken at pH 6.3 and 8. Since the residue type identification was made mainly from [^{15}N]-leucine specific labeling and the 3D ^{15}N-edited TOCSY and was incomplete and ambiguous for some of the spin systems at the early stage of the assignment process, we have taken the following approach for the sequence specific residue assignments of $CheA_{1-134}$ (Englander & Wand, 1987; Wüthrich, 1986).

First, many stretches of sequentially connected spin systems were identified through their strong NH_i-NH_{i+1} NOEs from the 2D and 3D ^{15}N-edited NOESY spectra. Sequential $C^\alpha H_i-NH_{i+1}$ and $C^\beta(C^\gamma)H_i-NH_{i+1}$ NOEs were checked for consistency. Sequence directionality was mostly distinguished by the direction of $C^\alpha H_i-NH_{i+1}$ and $C^\beta H_i-NH_{i+1}$ NOEs. Then, Leu, Ala, Gly, Val, and Thr residues were used as major landmarks, and the sequentially connected spin systems were placed inside the sequence. In most of the cases the placements of these stretches were unique. For nearby segments that were not sequentially connected by amide−amide NOEs, either structurally or by chance of chemical shift proximity to its nearby residues, $C^\alpha H_i-NH_{i+1}$ and $C^\beta(C^\gamma)H_i-NH_{i+1}$ NOEs were examined to build the right connections between these sequence stretches. Finally, gaps within the sequence were filled in by examination of the unassigned peaks to match the observed sequential NOEs.

The five longest sequence segments, which are D14−Q31 (with five Leu, two Ala, and one Val), A37−F59 (with one Leu, five Ala, four Gly, and one Thr), M69−Q82 (with two Leu, one Ala, and one Gly), D86−D96 (with two Leu and one Thr), and A122−E133 (with three Leu, three Ala, and one Gly) were assigned confidently in the early stages of the assignment process. Short sequence stretches, V61−T65 and T66−L68, were also assigned since they contain the easily identifiable amino acids, valine, leucine, and threonine. Up to this point, two Leu, three Ala, and two Thr residues were left unassigned. By examination of the $C^\alpha H_i-NH_{i+1}$ and $C^\beta(C^\gamma)H_i-NH_{i+1}$ NOEs, one leucine was assigned to L83 and the other to L102. The two Ala residues that carry strong NH_i-NH_{i+1} NOEs were assigned to the sequence segment A113−A114. Strong sequential NOEs were also observed for an isolated residue pair containing an alanine. They were assigned to E33−A34, which are between two prolines. The sequence segments S6−Q10 and T11−F13 were identified later in the assignment process by matching the unassigned threonines with the sequence. The NH_i-NH_{i+1} NOEs for residues F13−D14, Q10−T11, and Q101−D103 were not obvious because of their chemical shift proximity, but their sequential connections were confirmed by the observed $C^\alpha H_i-NH_{i+1}$ and $C^\beta(C^\gamma)H_i-NH_{i+1}$ NOEs within the respective sequence stretches.

S60 was assigned by comparing its amide NOESY cross peaks with those of F59 and V61. Though the sequential NH−NH NOEs between F59 and S60 are weak and close to the diagonal peaks, there is a unique $C^\alpha H_i-NH_{i+1}$ NOE cross peak from F59 to S60. An NOE cross peak between the amide proton of S60 and the aromatic ring $C^\delta H$ proton of F59 was also observed.

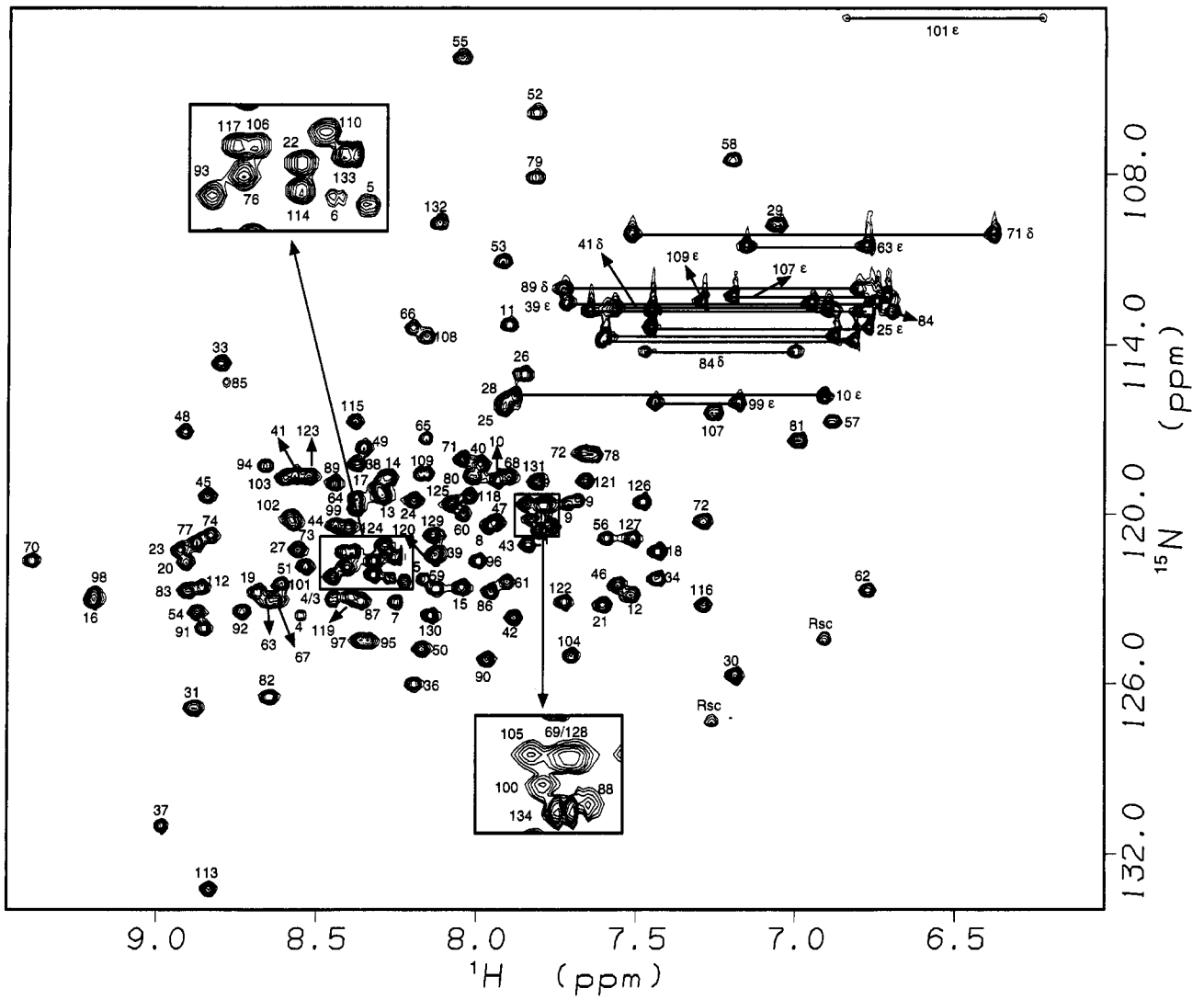

FIGURE 1: $^{1}H-^{15}N$ HSMQC spectrum of ca. 2 mM uniformly ^{15}N-labeled $CheA_{1-134}$ in 95% H_2O/5% D_2O, 50 mM $NaPO_4$, and 0.02% sodium azide, at 30 °C and pH 6.3. The spectrum is labeled with the assignments of the backbone amides and some of the side chain NH_2 groups from Asn and Gln residues. Two cross peaks labeled with Rsc are from the side chains of arginine residues and are folded into the spectrum from the ^{15}N chemical shift of ~85 ppm.

N84 was identified by observing NOEs between its NH proton and NH protons of L83, D86, and I87. T85 was not observed at pH 8 and was a weak peak at pH 6.3 in the $^{1}H-^{15}N$ HSMQC spectra. Its assignment came from comparison of its NOESY cross peak patterns with those of nearby residues.

D36, K109, and A112 were identified at the final stage of the assignment process by comparing the NOE cross peaks of the unassigned amide groups with those of the N- or C-terminal residues of the assigned sequence segments.

No clear sequential NOEs were observed for D4–I5. These residues were assigned according to their TOCSY cross peak patterns. At pH 6.3 the amide groups of residues D4–Y9 appear as doublets in the $^{1}H-^{15}N$ correlation spectrum. At pH 8 these cross peaks are very weak or invisible. This suggests that the N-terminal region is exposed to the solvent and is involved in slow conformational exchange. Residues M1–S2 were not observed.

In summary, over 70% of the residues in $CheA_{1-134}$ are assigned to helical structure, which causes a great deal of chemical shift overlap yet aids in the process of identifying sequential connectivities. The identification of the 17 leucines through [^{15}N]leucine-specific labeling has enabled the residue specific assignments to proceed in the early stages of the assignment process.

Regular Secondary Structure. The most striking feature of the secondary structure of the $CheA_{1-134}$ fragment is its high helical content. Figure 3 shows the backbone amide–amide and side chain–amide NOE correlations along with data from the hydrogen exchange experiments and $^{3}J_{HN\alpha}$ measurements. Analysis of the 2D and 3D NOESY spectra indicates that this protein contains five helices and no β strands. The helical structure extends from approximately the sixth residue of the N-terminal region to near the C-terminus interrupted by a few turns. Though due to overlap or chemical shift proximity some NOE correlations are not well resolved, the continuous helical structure within the five sequence segments is clear. Extensive strong NH_i-NH_{i+1}, $C^{\alpha}H_i-NH_i$, and $C^{\beta}H_i-NH_{i+1}$ NOEs, some medium strong or weak NH_i-NH_{i+2}, $C^{\alpha}H_i-NH_{i+3}$, and $C^{\alpha}H_i-NH_{i+4}$ NOEs were observed in the five helical regions. These sequential and medium-range NOEs indicate helical structure (Wüthrich, 1986; Englander & Wand, 1987). This conclusion is supported by hydrogen exchange experiments and

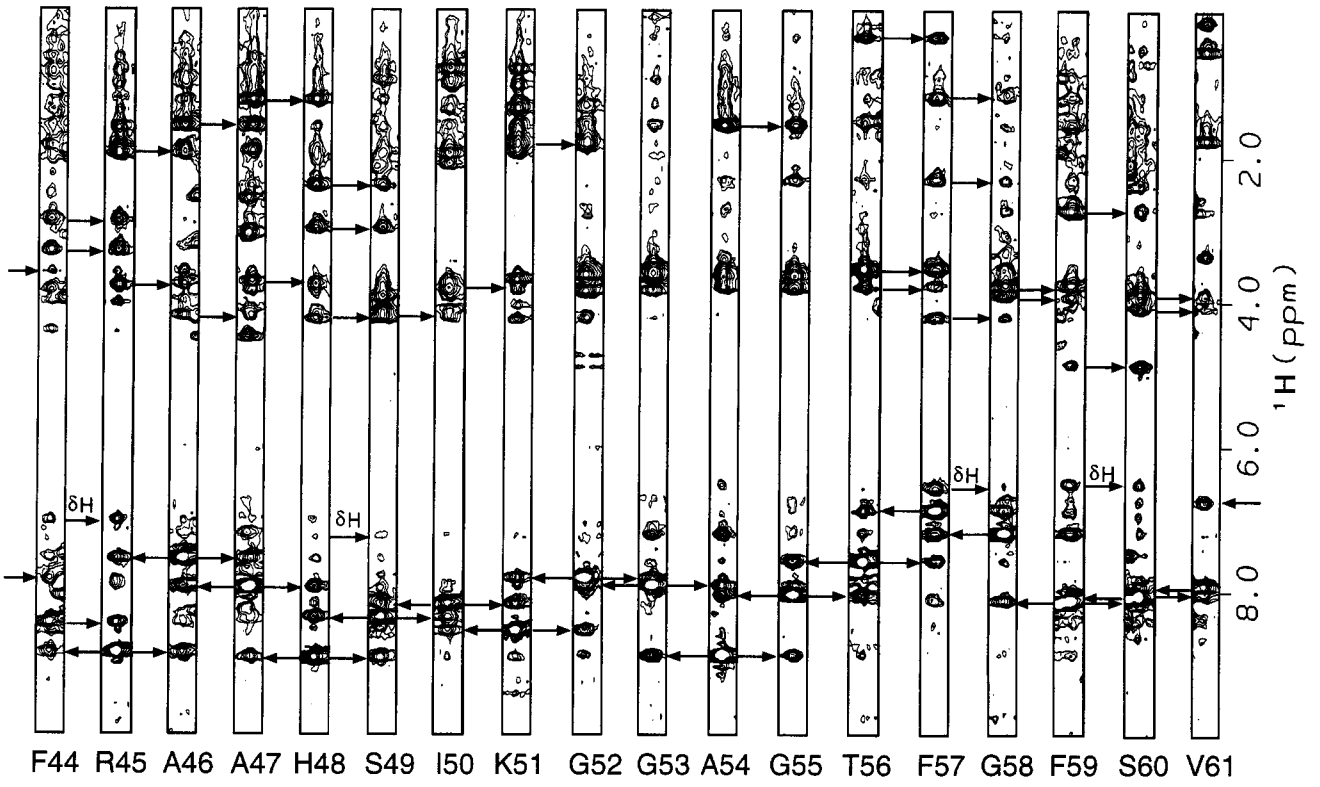

FIGURE 2: Selected strips from a 3D ^{15}N-edited NOESY with a mixing time of 150 ms, illustrating the NOE correlations involving some of the residues in helix B and residues in the well defined loop region connecting helices B and C. Arrows indicate the sequential NH_i-NH_{i+1}, $C^{\alpha}H_i$-NH_{i+1}, and $C^{\beta}H_i$-NH_{i+1} connectivities. Also shown are the NOEs between the aromatic ring $C^{\delta}H$ protons of the Phe and His residues and the amide protons of their own and the adjacent residues. The two strong cross peaks at 7.21 ppm on the strips of Gly53 and Ala54 are from the aromatic ring protons of Tyr8. The weak cross peak at 6.53 ppm on the strip of Ala54 is from the $C^{\delta}H$ proton of Phe59.

measurements of $C^{\alpha}H$ proton to NH proton coupling constants, which show five continuous slowly exchanging regions with small $^3J_{HN\alpha}$ values (<6 Hz) interspersed by several fast-exchanging segments with large $^3J_{HN\alpha}$ values (>7 Hz). As these experiments suggest, the five helices are defined approximately as follows: helix A, T11−H26; helix B, D36−K51; helix C, S60−R77; helix D, D86−K106; helix E, D112−K129.

Turns and Loops. Another interesting feature of the secondary structure is the well defined long loop connecting helices B and C, which has an unusually glycine rich sequence -GGAGTFGF-. Strong and continuous NH_i−NH_{i+1} NOEs were observed from A37 to Q82 except that medium strong NOEs showed around S60. G52−T56 all carry small $^3J_{HN\alpha}$ values. The turn connecting helices B and C is made by T56−F59, where $^3J_{HN\alpha}$ values of F57 and F59 are approximately 8 Hz. Therefore, A37−T56 is one continuous helical structure with the carboxyl end switched smoothly from α-helical structure toward possibly 3_{10}-helical structure and followed by a final turn around G58. NOEs were observed between residues in this region and residues in the N-terminal region of helix A and the turn region connecting helices D and E.

Other turn or loop regions were defined approximatley as L27−D36, L83−D86, and Q109−A112. T85 was not observed at pH 8, and its amide group displayed a weak cross peak at pH 6.3. NOEs were observed between the amide proton of N84 and the amide protons of L83, D86, and I88. We also observed NOEs between the $C^{\gamma}H$ protons of I87 and the amide proton of N84. These NOEs suggest that L83−D86 makes a sharp turn with T85 fully exposed to the solvent. Weak NH−NH NOEs were observed between Q31 and D36. The locations of these turns or loops are also supported by the topology of the 3D structure implied by the observed long-range NOEs.

Global Fold. The difficulty in assigning the cross peaks in the aliphatic region of the overcrowded 2D ^{1}H−^{1}H NOESY and DQF-COSY spectra hindered us from solving the detailed three-dimensional structure of $CheA_{1-134}$ at this stage. Nevertheless, a low-resolution structure of the protein resulted from studies of the available medium- and long-range NOEs derived from the NOESY spectra. Most of the identified long-range NOEs are between aromatic ring protons and other hydrophobic side chain protons (Figure 4).

As suggested by the NOE data, hydrogen exchange experiments, backbone dynamics studies, and measurements of $C^{\alpha}H$ proton to NH proton coupling constants, the protein is structured in a repeated helix−turn (or loop)−helix motif. Given the presence of five long helices as well as the topological restrictions imposed by the turns and the requirement of hydrophobic packing, only a small number of long-range NOEs are needed to define the general fold. Through simple model building, it was shown that all the available long-range NOEs are consistent with an antiparallel helix-bundle. Specifically, NOEs observed between the residue pairs Y105−T56, Y105−V61, Y105−L62, D112−V61, F116−V61, F116−E64, and F116−T65 indicate helices C and D are antiparallel and helices C and E are parallel. NOEs observed between the residue pairs Q31−D36, F12−G55, F12−F57, F12−F59, F8−G53, F8−A54, and F12−L62 indicate that helices A and B are antiparallel and helices A and C are parallel. The ring protons of F13 and the amide protons of K106 and Q107 carry NOE correlations with a

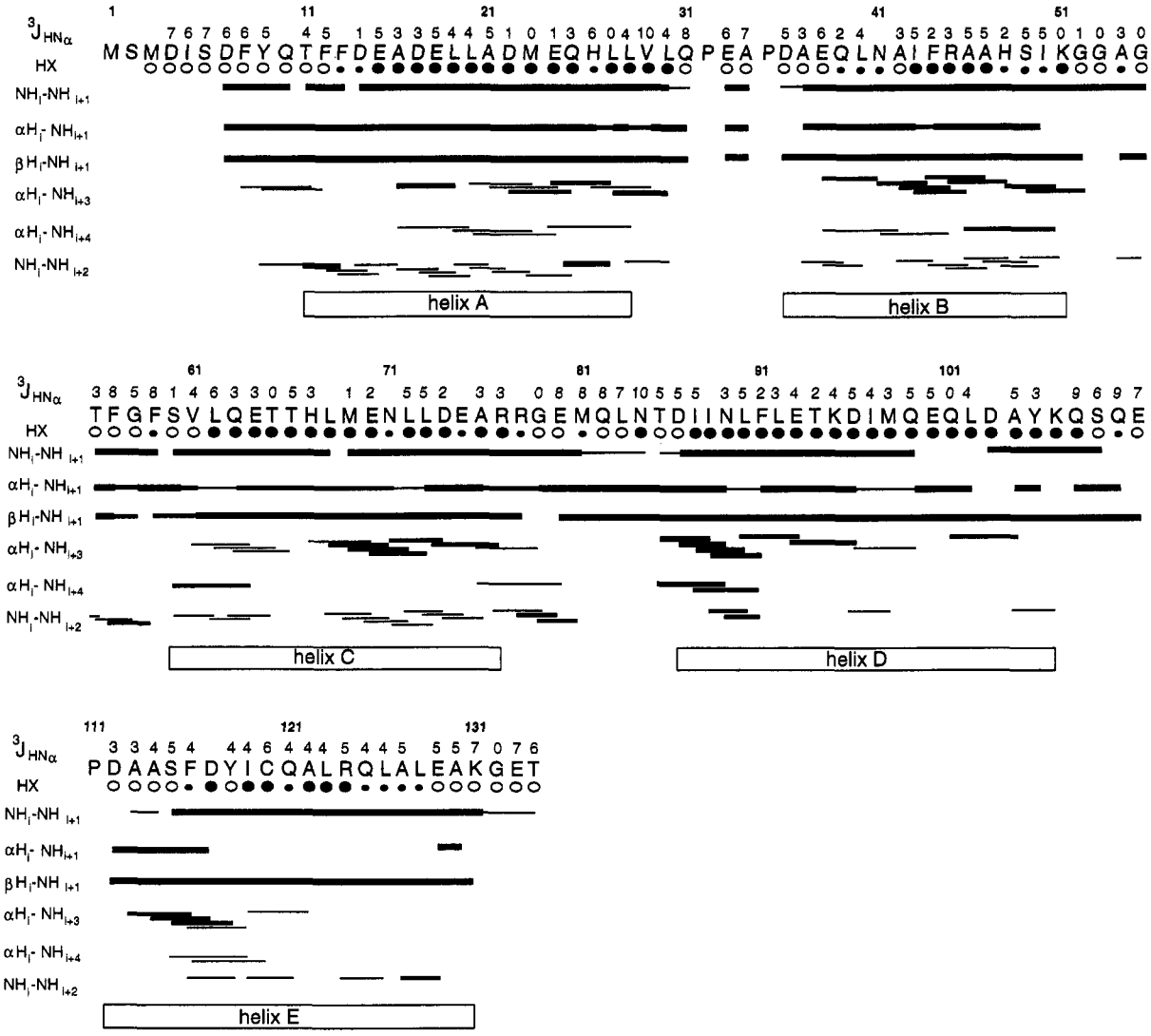

FIGURE 3: Summary of the sequential NOEs involving NH, $C^{\alpha}H$, and $C^{\beta}H$ protons together with data from hydrogen exchange and $^3J_{HN\alpha}$ measurements. Open circles denote those amides absent from the HSMQC spectrum after 10 min exchange with D_2O. Big black circles denote those amides still present in the spectrum after 1 h of exchange. Small black circles indicate those amides that are partially protonated after 10 min of exchange but are nearly fully deuterated after 1 h of exchange. The thickness of the line for the NOEs represents the intensity of the NOE. The $^3J_{HN\alpha}$ coupling constants (Hz), determined from the HSQC-J experiments, are shown above each residue.

common proton, which is most likely from the side chain of L102. Overall these NOEs suggest that the protein has an up-and-down helical connection. This type of topology is also consistent with the observed interhelical NOE correlations.

For antiparallel four-α-helix bundle proteins, the possible interhelical connectivity has been classified into six categories, either right-handed or left-handed (Presnell & Cohen, 1989). In our case we decomposed the problem into the determination of the handedness of the helix cluster formed by helices B, C, and D and the relative positions of helices A and E. From the examination of the available NOE data, the topological restraints confined by ideal local geometry, and patterns of hydrophobic packing, only one type of arrangement of the five helices is highly possible. This is the simplest structure of antiparallel helix clusters, with an up-and-down, near neighbor connectivity. In this structure helices B, C, and D are arranged in a right-handed manner. Helix A packs against helices B and D, and helix E packs against helices C and D (Figure 4a).

The right-handed arrangement of helices B, C, and D can be shown by comparison with a left-handed structure. If one looks down from the carboxyl terminal end of helix A, the observed NOE distance restraints can be more easily satisfied if helices B, C, and D are arranged in a right-handed manner (Figure 4a). In a left-handed helix bundle, there is strong competition among various NOE restraints involving residues in helices B, C, and D (Figure 4b). Additionally, it is impossible to satisfy the NOE restraints between helices C and E and the NOEs between the pairs F116−Q101 and F44−D74 if helices B, C, and D are arranged in a left-handed fashion.

Once the handedness of helices B, C, and D is established, the positions of helices A and E are relatively easy to determine. The NOEs observed between the residue pairs F116−V61, F116−E64, F116−T65, and F116−Q101 suggest that helix E packs against helices C and D. The NOEs observed between residue pairs F13−L102 and F12−L102 suggest that helix A packs against helices B and D. Both helix A and helix E have obvious hydrophobic sides. The proposed positions of helices A and E also satisfy the requirement of hydrophobic packing. Obviously, with helix A packing against helices B and C, the autophosphorylation

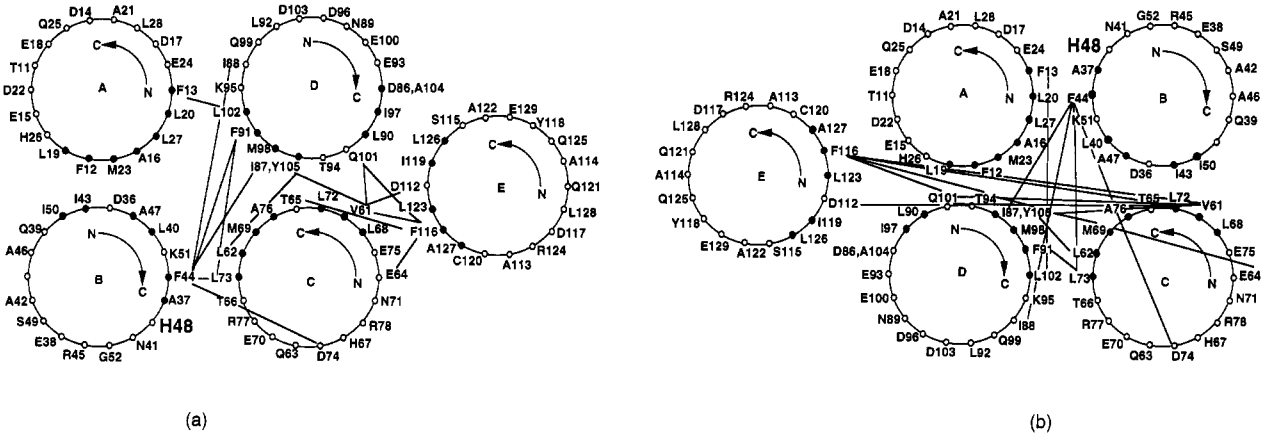

FIGURE 4: Helical wheel representations of the general fold of $CheA_{1-134}$ and summary of the interhelical NOE contacts derived from analysis of the 2D $^1H-^1H$ and 3D ^{15}N-edited NOESY spectra. The residues that form the hydrophobic surfaces are indicated by black circles. Helices A, C, and D are viewed from their C- termini, while helices B and C are viewed from their N- termini. (a) The proposed right-handed antiparallel helix bundle. (b) A typical unfavored left-handed antiparallel helix bundle. Hydrophobic interactions and topological restrictions imposed by the turns have been taken into account in making these arrangements. Many of the NOE restraints clearly cannot be satisfied in panel b. Other structures which have helices B, C, and D arranged in a left-handed manner are similarly unfavored. The NOE between Ile43 and His26, assigned during the structure calculation process, and the NOEs involving the residues in the turn regions are not shown.

site His48 and many other charged groups will very likely be buried.

Structure calculations using the program X-PLOR (Brünger, 1992) support the right-handed arrangement of the helices. During this process over 100 structures have been generated and subjected to simulated annealing refinement. In over 90% of the structures, helices B, C, and D are arranged in a right-handed manner. Other arrangements, which cannot be categorized as right-handed or left-handed structures as shown in Figure 4, significantly increase the van der Waals energy and the number of NOE violations. After further SA refinement, these structures appear to rearrange and form right-handed helix bundles. A typical left-handed arrangement of the helices, as shown in Figure 4b, has not been generated, indicating this type of arrangement is highly unlikely. The position of helix A depends on the NOEs between the residue pairs F13−L102 and F12−L102 being included as the distance restraints. Without these restraints, helix A tends to pack against the hydrophilic surfaces of helices B and C. This is very unlikely and is caused by the distance restraints between the residues in the N-terminal region of helix A and those in the turn region connecting helices B and C. When the restraints between helices A and D are included in the structure calculations, helix A prefers to pack against the hydrophobic surfaces of helices B and D. This change does not affect individual energy terms, showing that the packing of helix A against the hydrophilic surfaces of helices B and C is trivial and comes from the lack of NOE restraints between helices A and D. After SA refinement using all restraints, a total of 40 low energy structures, which have no dihedral angle violations larger than 5° and NOE distance violations larger than 0.8 Å and have at most a single NOE violation larger than 0.5 Å, were chosen and examined. All these structures are right-handed and carry a common fold even though the root-mean-square deviation from the average structure is relatively high at this resolution. An NOESY cross peak, which was not included in current calculations, was assigned to the amide proton of I43 and the $C^\beta H$ proton of H26 during the refinement process. This NOE is consistent with the 3D structure we currently have.

The right-handed arrangement of helices B, C, and D is also consistent with data from the pH titration and phosphorylation experiments (H. Zhou and F. W. Dahlquist, unpublished results). We observed that the chemical shifts of some backbone amides, including those of the three histidines and nearby residues, are sensitive to pH changes. These amides report apparent pK_a values which reflect the pK_a values of the nearby histidine imidazole rings. The imidazole ring of His48 and the backbone amides of His48 and its nearby residues report a common and unique pK_a value of 7.8. The amides of Glu70 and Asn71 are the only ones remote in the sequence from His48 that were observed to report an apparent pK_a of 7.8. Other amides mostly report apparent pK_a values of about 7.0 or 6.4, which are close to the pK_a values of His26 and His67, respectively. Upon phosphorylation of His48, Asn71 and Asp74 are among the few residues with relatively large chemical shift changes. These properties are consistent with a right-handed structure, in which Glu70, Asn71, and Asp74 are close to His48 and are apparently affected by its protonation and/or phosphorylation. In a left-handed structure these residues are exposed on the opposite side of the helix cluster with respect to His48. It is unlikely that in such a structure phosphorylation of His48 only affects the remote Glu70, Asn71, and Asp74 without inducing changes to the environments of other residues.

Figure 4a shows a helical wheel representation of the proposed general fold of $CheA_{1-134}$. The restrained minimized average structure of 40 SA structures is shown in Figure 5. This structure has no NOE distance violations larger than 0.5 Å and has a low van der Waals energy of 80 kcal/mol. The root-mean-square deviations of the SA structures about the average structure are 1.9 Å for the backbone atoms of helices A to D and 3.9 Å for all nonhydrogen atoms of residues 13−117. An overlay of 15 NMR structures is shown in Figure 6. The structure of $CheA_{1-134}$ carries the following general features:

(1) Most of the hydrophobic groups are buried while the polar and charged groups are mostly exposed on the surface. It seems that helix A has more hydrophobic contact with helix B than helix D, and helix E has more contact with helix D than helix C.

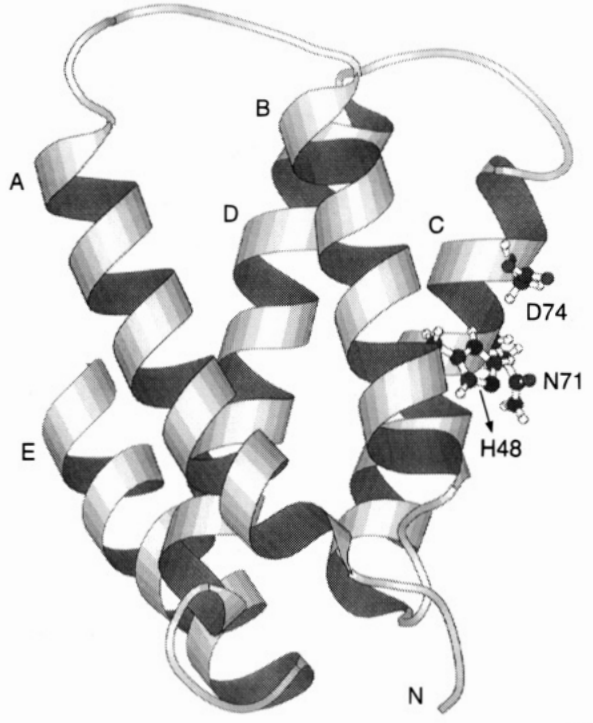

FIGURE 5: Ribbon diagram of the restrained minimized mean NMR structure of $CheA_{1-134}$. The average structure was obtained by averaging the coordinates of 40 structures best fit to N, C^{α}, and C atoms of residues 13−117. This structure was then subjected to restrained minimization with the van der Waals radii gradually increasing from a quarter of their full values to the final values. This process removed a number of bad nonbonded contacts in the original average structure. The site of autophosphorylation, His48, resides at the outer edge of the interface between helix B and helix C, with the imidazole ring tilted outward. Asn71 and Asp74 were observed to be affected by the phosphorylation of His48.

(2) Two hydrophobic clusters are observed. One makes the hydrophobic core of the protein, formed by the aromatic rings of F44 and F91, and the side chain groups of L72, L73, I87, and I88. The other one is formed by the side chain groups of F59, F57, Y105, F116, V61, L62, L102, F12, F13, and T58 and very likely having Y9 and F8 in close proximity. The latter resides at one end of the helix bundle.

(3) The position of His48, the site of autophosphorylation of CheA, is one of the major concerns. His48 resides near the carboxyl end of helix B. This position is consistent with its pK_a value of 7.8 (H. Zhou and F. W. Dahlquist, unpublished results). The protonated state of His48 is possibly stabilized partially by the effective negative charge at the carboxyl terminus of the helix (Armstrong & Baldwin, 1993; Lodi & Knowles, 1993). Evidence also indicates the imidazole ring of His48 is involved in hydrogen bonding (H. Zhou and F. W. Dahlquist, unpublished results). The rigid local environment of His48 is also consistent with the very limited and localized structural changes upon phosphorylation of His48, reflected by the small chemical shift changes of the backbone amides (H. Zhou and F. W. Dahlquist, unpublished results). In the 3D structure generated, His48 is located at the outer edge of the interface between helices B and C. No long-range NOEs leading to the imidazole ring have been identified. The backbone position of His48 implies the imidazole ring is most likely exposed to the solvent. This is supported by the weak NOEs observed between the amide proton of Ser49 and the ring $C^{\delta}H$ proton of His48 at both pH 6.3 and 8.0 and also strongly supported by the little structural changes upon phosphorylation of His48.

(4) Most of the upfield and downfield proton chemical shifts can be explained structurally by the aromatic ring current effects (Wüthrich, 1986). Upfield shifts of $C^{\gamma}H$ protons of V61 and T56 (0.17 and 0.38 ppm respectively) are due to the combined effects of several aromatic rings, including those of F13, F57, F59, Y105, and F116, clustered at one end of the helix bundle. The ring current effects of F44 and F91 also cause upfield shifts of the C^{γ} methyl protons of L72 and I87. Their cross peaks are distinct in the otherwise crowded $^1H-^1H$ NOESY and COSY spectra.

(5) Except for F116, whose aromatic ring is close to the γ methyl group of V61, the amide protons of E64 and T65, and the side chain NH_2 protons of Q101, none of the residues in helix E have been observed to have NOE correlations with any part of the amino-terminal four helices. This is partially due to the fact that only a small number of isolated NOE cross peaks in the 2D aliphatic proton region can be reliably assigned. Helix E obviously has a hydrophobic surface. Structural studies suggest that its N-terminal region packs against helices C and D and its central region packs against part of the hydrophobic surface of helix D. Though the orientation of helix E is not well defined, it is unlikely that helix E sticks out toward the solvent. However, our speculation is that the packing of helix E with helices C and D is possibly loose and helix E might not play an important role in stabilizing the helix cluster. Evidence shows that $CheA_{1-122}$, a protein fragment only including the N-terminal four helices and the turn region connecting helices D and E, is stable and can be phosphorylated nearly as well as $CheA_{1-134}$ (Morrison & Parkinson, 1994; T. Morrison and J. S. Parkinson, personal communication). This protein possibly corresponds to the 16 kDa fragment that was released after further partial proteolytic digestion of the 18 kDa phosphorylated phosphotransfer domain, as observed by Hess et al. (1988). These results suggest that the major part of helix E is neither important for the stability of the helix cluster nor critical for the phosphorylation reaction. Helix E is possibly related to other functional roles of the kinase CheA or involved in cooperating interdomain functional activities. The possible loose packing of helix E against other helices is also reflected in the smaller chemical shift deviation of the $C^{\alpha}H$ protons of the residues in this helix relative to the random coil chemical shifts (Wishart et al., 1995).

(6) The five helices show significantly different degrees of protection of the amide protons from exchange with the solvent. Residues 16, 19, 20, 21, 23, and 28 from helix A, residues 69, 70, 72, and 73 from helix C, residues 88, 89, 91, 92, 93, 94, 95, 97, 98, 100, 101, 102, 103, and 105 from helix D, and residue 123 from helix E were observed to be strongly protected from hydrogen exchange with the solvent. The amides of these residues still display strong cross peaks in the HSMQC spectrum taken after 8 h of exchange with D_2O (Figure 7). The strong and continuous protection of helix D from hydrogen exchange supports the current structure in which helix D has extensive contact with the other four helices. It is worth noting that none of the residues from helix B and only one residue from helix E show such strong protection, indicating very different structural and dynamic features of these helices. The intensities of the amide cross peaks of H48 and S49 were significantly reduced

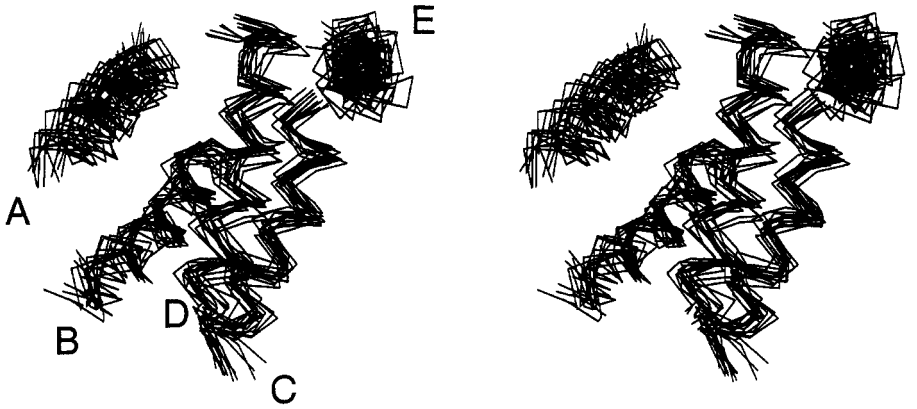

FIGURE 6: Stereoview of the helical regions of 15 NMR structures best fit to N, C^{α}, and C atoms of residues 13−117. The turns and N- and C-terminal regions are omitted from the diagram for more clarity. The N-terminus is located in the upper left corner.

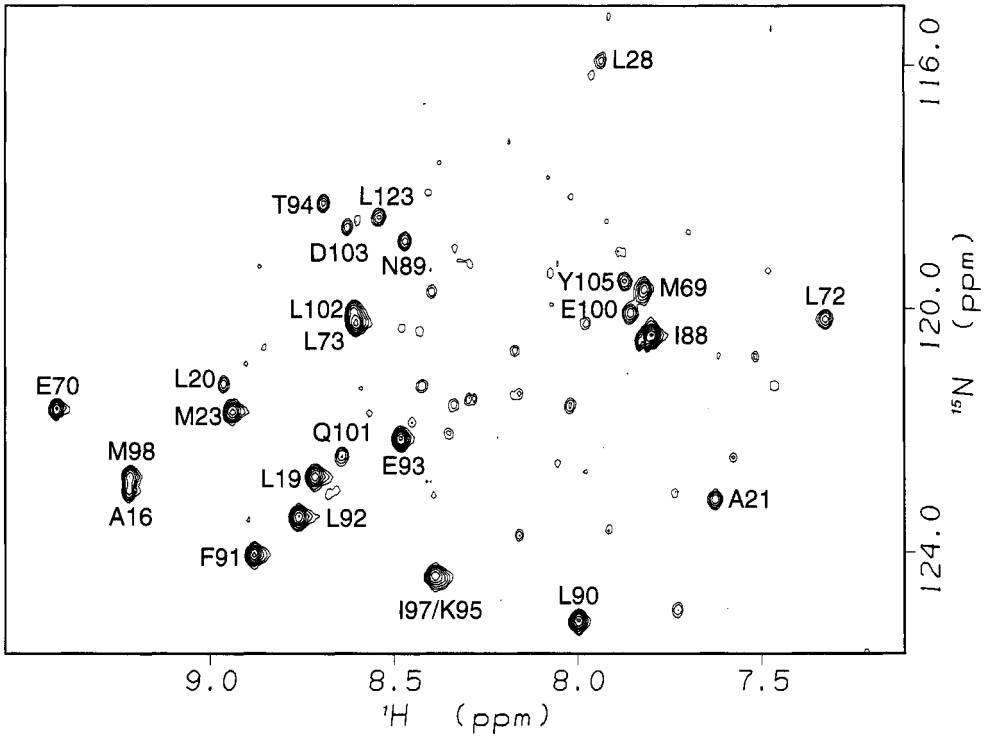

FIGURE 7: $^{1}H-^{15}N$ HSMQC spectrum of ^{15}N-uniformly labeled $CheA_{1-134}$ after approximately 8 h of exchange with D_2O, showing the residues strongly protected from exchange with the solvent.

after about 10 min exchange with D_2O. The hydrogen exchange of these residues may be enhanced by a number of nearby charged groups, which may play significant roles in the phosphotransfer reaction.

Backbone Dynamics. Shown in Figure 8 are the measured T_1 and T_2 values and NOE factors, and the calculated overall correlation times τ_m, order parameters S^2, and effective correlation times τ_e of internal motions. A total of 120 T_1 and T_2 values for ^{15}N spins have been extracted from the well resolved peaks. The average relaxation times for $CheA_{1-134}$ are $\langle T_1 \rangle = 543 \pm 32$ ms and $\langle T_2 \rangle = 102 \pm 10$ ms. The τ_m's calculated from T_1/T_2 ratios give $\langle \tau_m \rangle = 8.13 \pm 0.31$ ns. Only the well defined five helical regions, which have a total of about 79 residues, were included in the $\langle \tau_m \rangle$ calculation, while residues 13−129 were included to calculate the average T_1 and T_2. The small spread in the measured τ_m suggests that there is little anisotropy in the rotational reorientation of the molecule.

A single rotational correlation time $\langle \tau_m \rangle = 8.13$ ns was used in the calculation of S^2 and τ_e. The S^2 values mostly fall in the range of 0.85−0.95, with $\langle S^2 \rangle = 0.87 \pm 0.06$ for residues 13−129 and $\langle S^2 \rangle = 0.90 \pm 0.03$ for the helical regions. The τ_e values for internal motions, however, cover an extremely large range. The central regions of helices A, D, and E exhibit small τ_e values ($<10^{-2}$ ps), while the central regions of helices B and C and the turn regions primarily have τ_e values in the 10−150 ps range.

As expected, the N- and C-terminal residues exhibit high mobility reflected by the small S^2 values and large τ_e values. They are also the only residues whose cross peaks are negative in the spectrum with the $^{15}N\{^{1}H\}$ NOE. Overall it shows that, except for the two ends of the protein, internal motions are fast enough to be ignored in the calculation of τ_m.

The five helices have similar but potentially different average effective correlation times: helix A (~15 residues), $\langle \tau_m \rangle = 8.4 \pm 0.4$ ns; helix B (~13 residues), $\langle \tau_m \rangle = 7.9 \pm 0.2$ ns; helix C (~14 residues), $\langle \tau_m \rangle = 7.9 \pm 0.3$ ns; helix D (~19 residues), $\langle \tau_m \rangle = 8.4 \pm 0.2$ ns; helix E (~14 residues), $\langle \tau_m \rangle = 8.0 \pm 0.3$ ns. The slightly smaller effective

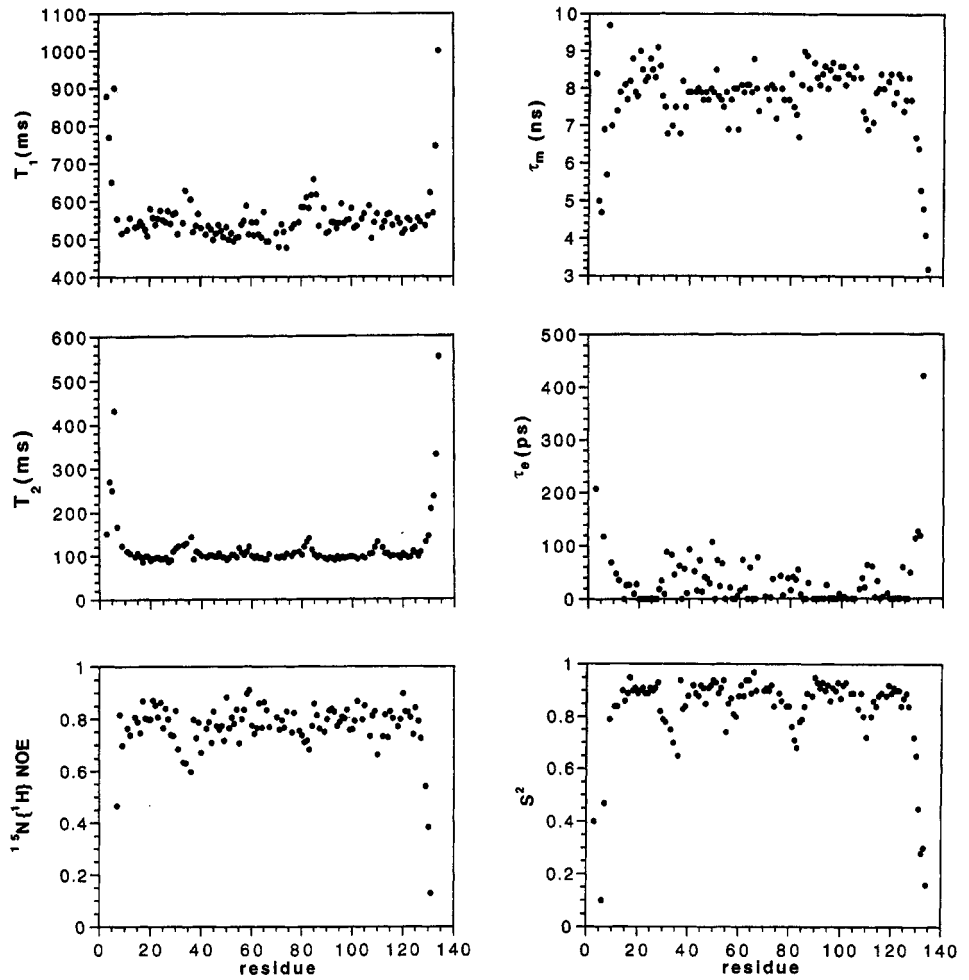

FIGURE 8: Plots as a function of residue number of the measured T_1 and T_2 values, and the $^{15}N\{^1H\}$ NOE, the rotational correlation time derived from T_1/T_2, the order parameter, S^2, and the correlation time of internal motion, τ_e. The τ_e values of Asn71 and Asp74, 1.7 and 2.0 ns, respectively, are omitted for proper scale of the data.

correlation times of helices B and C are possibly due to the higher internal motions of this region as observed in the larger τ_e values.

Besides the residues near the amino and carboxyl ends, three regions in the protein, L31−D36, G81−N84, and K109−A112, show significantly longer T_2, smaller motional correlation times and order parameters, than the average values. This suggests these turn regions are relatively flexible. Dynamics parameters of G52−G58 only show slight differences from the average values. These values are consistent with the NOE correlations observed, which indicate this region is formed in a well defined structure. N71 and D74 from helix C have unusually long τ_e values, 1.7 and 2.0 ns, respectively, indicating unique dynamics properties of these residues. D74 also exhibits a S^2 value of 0.83, which is small compared to the average of 0.90 for the helical regions. These residues are possibly involved in interactions with the site of phosphorylation, H48.

DISCUSSION

We have presented the backbone amide and side chain proton assignments, secondary structure, and folding topology of the protein fragment $CheA_{1-134}$, based on analysis of 3D ^{15}N-edited TOCSY, 2D and 3D ^{15}N-edited NOESY, 2D $^1H-^1H$ NOESY and COSY, $^3J_{HN\alpha}$ measurements, and NH exchange data. ^{15}N labeling of the leucine residues was shown to be critical for the unambiguous identification of the 17 leucines at the early stage of the assignment procedure. Through analysis of long-range and medium-range NOE data and patterns of hydrophobic packing by simple model building and structure calculations, we have determined the up-and-down and near-neighbor connectivity of the five helices. On the basis of analysis of the available interhelical NOEs and requirements of hydrophobic packing, we further propose the N-terminal four helices to be packed in a right-handed fashion with helix E packed against helices C and D. At this stage, the orientation of the C-terminal helix has not been firmly established. We also have reported backbone dynamics studies of $CheA_{1-134}$. The relaxation parameters determined well reflect degree of secondary structure and local flexibility. It was shown that $CheA_{1-134}$ is a compact protein with very limited flexibilities both in helices and turns.

Because of extensive chemical shift overlap, as commonly encountered in NMR studies of proteins with high helical content, the 2D $^1H-^1H$ NOESY did not yield enough unambiguous NOE distance restraints for a high-resolution structure determination. Three- or four-dimensional ^{13}C or $^{15}N/^{13}C$ edited NOESY spectra are required for detailed structure studies. The mixing times used in the NOESY experiments, 100 or 150 ms, are quite long for a protein of this size. To minimize possible errors in the evaluation of

NOE restraints resulting from spin diffusion effect, we have examined NOEs at both mixing times and have used conservative distance ranges for the distance restraints. The distance ranges were also adjusted at different stages of the structure calculations, with the longest NOE distance assigned fixed at 5.0 Å. These approaches did not appear to affect the overall topology of the best structures obtained. In the backbone dynamics studies, the $^{15}N\{^1H\}$ NOE factors might be slightly overestimated due to solvent−saturation transfer effects during the short presaturation period in the experiment without NOE effect (Li & Montelione, 1994; Sørensen et al., 1995). These possible errors were taken into account simply by reducing the contribution of the NOE data in the fitting of S^2 and τ_e in eq 3.

As discussed above, residues 122−134 may not play an important role in the phosphotransfer reaction and in stabilizing the helix cluster. Structure and dynamics studies of the protein fragment $CheA_{1-233}$ (H. Zhou, M. M. McEvoy, D. F. Lowry, R. V. Swanson, M. I. Simon, and F. W. Dahlquist, unpublished results), which sequentially encodes the phosphotransfer domain (P1), a linker region L1, and the CheY binding domain (P2), show that the linker region is highly flexible and there is no stable contact between the two domains. The only detectable changes in the amide proton and nitrogen chemical shifts from the values given by $CheA_{1-134}$ were observed for several residues around M98 and most of the residues in the C-terminal helix. However, no significant secondary structure changes have been observed. These results suggest that the structural changes in or around helix E do not affect the environments of the N-terminal three helices, and that the major part of helix E possibly only interacts strongly with helix D. Whether these subtle changes play any significant functional role is not clear. It is interesting to note that M98 is the start site of the $CheA_S$, the short form of the proteins encoded by the *E. coli cheA* gene that lacks the N-terminal 97 residues of the full length $CheA_L$ (Kofoid & Parkinson, 1991; Smith & Parkinson, 1980). Complexes of $CheA_S$ with CheZ have been observed, whereas $CheA_L$/CheZ complexes are not observed (McNally & Matsumura, 1991). This is very likely due to the structural differences in the N-terminal phosphotransfer domain of CheA, especially for those residues in helices D and E. Residue 98 occurs in the middle of helix D. The remaining 10 residues in helix D could form a shortened helix in $CheA_S$ or might be disordered.

In order to function, His48 must interact with both the kinase domain to be phosphorylated and with bound CheY or CheB to transfer its phosphate. The observed rigidity of the P1 domain suggests that internal motions within P1 are not likely to account for both interactions. It is likely that the flexible linker joining P1 and P2 allows motion of His48 between these sites of interaction. Our future studies aim at obtaining a high-resolution structure of $CheA_{1-134}$.

ACKNOWLEDGMENT

We thank Megan M. McEvoy for helpful discussions and comments on the manuscript.

SUPPORTING INFORMATION AVAILABLE

Two tables containing the chemical shifts of the amide groups and side chain protons of 128 residues and some of the NH_2 groups of Asn and Gln residues as well as the values and errors of the experimental ^{15}N T_1, T_2, and NOE relaxation data together with the calculated rotational correlation times, τ_c, order parameters, S^2, and correlation times for internal motions, τ_e (6 pages). Ordering information is given on any current masthead page.

REFERENCES

Armstrong, K. M., & Baldwin, R. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11337−11340.
Barbato, G., Ikura, M., Kay, L. E., Paster, R. W., & Bax, A. (1992) *Biochemistry 31*, 5269−5278.
Billeter, M., Neri, D., Otting, G., Qian, Y. Q., & Wüthrich, K. (1992) *J. Biomol. NMR 2*, 257−274.
Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 5485−5489.
Borkovich, K. A., & Simon, M. I. (1990) *Cell 63*, 1339−1348.
Borkovich, K. A., Kaplan, N., Hess, J. F., & Simon, M. I. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1208−1212.
Bourret, R. B., Borkovich, K. A., & Simon, M. I. (1991) *Annu. Rev. Biochem. 60*, 401−441.
Brünger, A. T. (1992) *X-PLOR 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
Cavanagh, J., Chazin, W. J., & Rance, M. (1989) *J. Magn. Reson. 87*, 110−131.
Clore, G. M., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 7387−7401.
Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953−5958.
Garrett, D. G., Powers, R., March, C. J., Frieden, E. A., Clore, G. M., & Gronenborn, A. M. (1992), *Biochemistry 31*, 4547−4553.
Gegner, J. A., Graham, D. R., Roth, A. F., & Dahlquist, F. W. (1992) *Cell 70*, 975−982.
Hess, J. F., Bourret, R. B., & Simon, M. I. (1988) *Nature 336*, 139−143.
Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110−126.
Kay, L. E., Torchia, D. A., & Bax, A. (1989a) *Biochemistry 28*, 8972−8979.
Kay, L. E., Marion, D., & Bax, A. (1989b) *J. Magn. Reson. 84*, 72−84.
Kumar, A., Wüthrich, K., & Ernst, R. R. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
Li, Y. C., & Montelione, G. T. (1994) *J. Magn. Reson. B 105*, 45−51.
Lipari, G., & Szabo, A. (1992) *J. Am. Chem. Soc. 104*, 4546−4558.
Lodi, P. J., & Knowles, J. R. (1993) *Biochemistry 32*, 4338−4343.
Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95−117.
Marion, D., & Wüthrich, K. (1983) *Biochem. Res. Commun. 113*, 967−974.
Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., & Gronenborn, A. M. (1989a) *Biochemistry 28*, 6150−6156.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989b) *J. Magn. Reson. 85*, 393−399.
McIntosh, L. P., Wand, A. J., Lowry, D. F., Redfield, A. G., & Dahlquist, F. W. (1990) *Biochemistry 29*, 6341−6362.
McNally, D. F., & Matsumura, P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6269−6273.
Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 608−613.
Milburn, M. V., Privé, G. G., Milligan, D. L., Scott, W. G., Yeh, J., Jancarik, J., Koshland, D. E., & Kim, S.-H. (1991) *Science 254*, 1342−1347.
Morrison, T., & Parkinson, J. S. (1994) *Biochemistry 33*, 5485−5489.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44−73.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
Ninfa, E. G., Stock, A., Mowbray, S., & Stock, J. (1991) *J. Biol. Chem. 266*, 9764−9770.
Parkinson, J. S., & Kofoid, E. C. (1992) *Annu. Rev. Genet. 26*, 71−112.
Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1993) *Biochemistry 32*, 6743−6762.

Presnell, S. R., & Cohen, F. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6592−6596.
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Hannery, B. P. (1992) *Numerical Recipes*, Cambridge University Press, Cambridge, U.K.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1984) *Biochem. Biophys. Res. Commun. 117*, 479−482.
Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169−173.
Smith, R. A., & Parkinson, J. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5370−5374.
Sørensen, M. D., Kristensen, S. M., & Led, J. J. (1995) *J. Magn. Reson. B 107*, 83−87.
Stewart, R. C., & Dahlquist, F. W. (1987) *Chem. Rev. 87*, 997−1025.
Stock, J. B., Ninfa, A. J., & Stock, A. M. (1989) *Microbiol. Rev. 53*, 450−490.
Swanson, R. V., Schuster, S. C., & Simon, M. I. (1993) *Biochemistry 32*, 7623−7629.
Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., & Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wylie, D., Stock, A., Wong, C.-Y., & Stock, J. (1988) *Biochem. Biophys. Res. Commun. 151*, 891−896.
Zuiderweg, E. R. P. (1990) *J. Magn. Reson. 86*, 346−357.
Zuiderweg, E. R. P., & Fesik, S. W. (1989) *Biochemistry 28*, 2387−2391.
Zuiderweg, E. R. P., Hallenga, K., & Olejniczak, E. T. (1986) *J. Magn. Reson. 70*, 336−343.

BI951174+