Reexamination of the Secondary and Tertiary Structure of Histidine-Containing Protein from *Escherichia coli* by Homonuclear and Heteronuclear NMR Spectroscopy†

Philip K. Hammen,‡ E. Bruce Waygood,§ and Rachel E. Klevit*,‡

Department of Biochemistry, University of Washington, Seattle, Washington 98195, and Department of Biochemistry, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0, Canada

Received August 2, 1991; Revised Manuscript Received October 16, 1991

ABSTRACT: Analysis of the histidine-containing protein (HPr) from *Escherichia coli* by two-dimensional homonuclear and heteronuclear nuclear magnetic resonance techniques has been performed, extending the work originally reported [Klevit, R. E., Drobny, G. D., & Waygood, E. B. (1986) *Biochemistry 25*, 7760–7769; Klevit, R. E., & Drobny, G. P. (1986) *Biochemistry 25*, 7770–7773; Klevit, R. E., & Waygood, E. B. (1986) *Biochemistry 25*, 7774–7781]. Two-dimensional homonuclear total coherence spectroscopy (TOCSY) allowed for more complete assignments of the side-chain spin systems than had been possible in the original studies. As well, two-dimensional ^{15}N–^{1}H heteronuclear spectroscopy was used to resolve a number of ambiguities present in the homonuclear spectra due to resonance redundancies. These analyses led to the correction of a number of resonance assignments that were made with the spectra that could be collected with the technology that existed 6 years ago. In addition, amide exchange rates and $^{3}J_{NH}$ coupling constants have been measured, extending the original analysis and yielding new structural information. All these data have been used to reexamine the folding topology of *E. coli* HPr. Structure calculations showed that the topology derived from the earlier NMR data, i.e., a four-stranded β-sheet with three α-helices running along one side of the sheet, was essentially unchanged, although at the present level of analysis, a well-defined "helix B" could not be established with high confidence. In addition, the data reported here revealed the existence of two slowly-exchanging side-chain hydroxyl protons belonging to Ser^{31} and Thr^{59}. Their behavior strongly suggests that these side chains are involved in hydrogen bonds. These two residues are both at the edges of the β-sheet and may contribute to the extreme stability of this structure.

Five years ago the folding topology for the histidine-containing protein (HPr) from *Escherichia coli* was reported (Klevit & Waygood, 1986). On the basis of a qualitative analysis of two-dimensional homonuclear ^{1}H NOESY[1] NMR spectra, we concluded that the 85-residue phosphotransfer protein consists of a four-stranded antiparallel β-sheet, with three α-helices running approximately parallel to the sheet and antiparallel to each other. Subsequent to that report, a similar analysis was performed for a related HPr from the Gram-positive bacterium *Bacillus subtilis* (Wittekind et al., 1990). Again, the NMR data indicated the presence of a four-strand antiparallel β-sheet as well as two of the three α-helices indicated in the *E. coli* protein. However, the tertiary structure derived from X-ray diffraction studies on *E. coli* HPr was significantly different from the topology suggested by the NMR results, consisting of two two-stranded β-sheets that are approximately parallel to each other and two α-helices that run perpendicular to these sheets (El-Kabbani et al., 1987).

In an attempt to address the apparent discrepancy between the NMR and X-ray structures, we set out to determine a detailed structure for HPr from NMR data. As part of this process, additional ^{1}H–^{1}H homonuclear and ^{15}N–^{1}H heteronuclear 2D spectra were collected. From these, additional ^{1}H resonance assignments and ^{15}N assignments have been made. During this procedure, some of the original spectral assignments were found to be incorrect. The extended and corrected resonance assignments are reported here. Preliminary structural determination has been carried out using these assignments and other new information reported here. The calculated structures have the same topology as that originally proposed on the basis of the original qualitative analysis.

Materials and Methods

Sample Preparation. HPr was purified from *E. coli* as previously described (Sharma et al., 1991). Protein that was uniformly labeled with ^{15}N was generated by growing the cells on minimal medium with 0.2% glucose in which the sole source of nitrogen was [^{15}N]NH_4Cl. Purified proteins were dialyzed against buffer containing 5 mM potassium phosphate and 0.01 mM EDTA at pH 6.5. The solutions were lyophilized and dissolved in 90% H_2O/10% D_2O [or 99.96 atom % D_2O (Sigma)] to give solutions that were ca. 4 mM protein, 50 mM potassium phosphate, and 0.1 mM EDTA.

NMR Spectroscopy. Two-dimensional NMR spectra were acquired on a Bruker AM500 spectrometer. All spectra were recorded at 30 °C and pH 6.5. The two-dimensional spectra were obtained in the pure-phase absorption mode using time-proportional phase incrementation (Marion & Wüthrich,

† Supported by NIH Grant DK-35187 (R.E.K.) and MRC of Canada Operating Grant MT-6147 (E.B.W.). R.E.K. was supported by an American Heart Association Established Investigatorship Award.

* Author to whom correspondence should be addressed.

‡ University of Washington.

§ University of Saskatchewan.

[1] Abbreviations: HPr, histidine-containing protein; NMR, nuclear magnetic resonance; EDTA, ethylenediaminetetraacetic acid; 1D, one-dimensional; 2D, two-dimensional; NOE, nuclear Overhauser effect; TOCSY, two-dimensional total coherence spectroscopy; NOESY, two-dimensional NOE spectroscopy; DQF-COSY, two-dimensional double-quantum-filtered *J*-correlated spectroscopy; HMQC, two-dimensional heteronuclear multiple quantum coherence spectroscopy; RELAY, two-dimensional relayed coherence transfer spectroscopy.

0006-2960/91/0430-11842$02.50/0

1983). The water signal was suppressed by saturation between acquisitions and during NOESY mixing times. Data were processed using the software FELIX 1.0 (Hare Research, Woodinville, WA) on a Silicon Graphics 4D-35 workstation or FTNMR (Hare Research) on a VAX 11/780. Two-dimensional data sets were generally 600 × 2K points, processed with sine-bell filters skewed toward $t = 0$, shifted by $\pi/3$ in both dimensions and zero-filled to produce 2K × 2K matrices. Polynomial baseline correction of third or fourth order was used in the f_1 transform. A chemical shift reference for ^1H of 4.8 ppm (relative to TMS) for the solvent signal was used. For ^{15}N chemical shift assignments, a value of 24.9 ppm for NH_4^+, relative to liquid NH_3, was used. The resonance frequency of NH_4^+ relative to protein was established by ^{15}N detection of a sample that was 2.9 M $^{15}NH_4Cl$ in 1.0 M HCl.

DQF-COSY spectra were acquired using composite pulses to achieve signal filtering (Mueller et al., 1986). The MLEV-17 mixing sequence (Bax & Davies, 1985) was used for D_2O TOCSY spectra with mixing times of 53 and 81 ms. In H_2O, clean-TOCSY spectra (Griesinger et al., 1988) were obtained with mixing times of 46 and 70 ms. NOESY spectra were obtained using a 32-step phase cycle and mixing times of 5, 25, 50, 75, and 100 ms in H_2O and 100 ms in D_2O. Scalar contributions to the signal were suppressed by randomization of the mixing time.

Heteronuclear 2D ^{15}N–^1H spectroscopy was carried out with proton detection, using a 5-mm inverse probe. Heteronuclear decoupling was achieved during acquisition with the GARP-1 composite pulse (Shaka et al., 1985). An HMQC-NOESY (Gronenborn et al., 1989) spectrum was obtained with a mixing time of 150 ms and had dimensions of 452 × 4K. It was processed with sine-bell filters shifted by $\pi/2$ and was zero-filled to yield a 512 × 2K matrix.

An HMQC-J experiment (Driscoll et al., 1990) was carried out, with collection of 1058 serial files and a t_1(max) of 0.21 s. It was processed using a skewed sine-bell filter shifted by $\pi/3$ for the t_2/f_2 transform. Individual data-containing columns were then extracted from the data matrix. Extrapolation of the signal was carried out using the Burg algorithm for linear prediction (Burg, 1967). A total of 100 complex data points were added to the 529 experimental points as described by 30 linear prediction coefficients. The resultant signals were processed with a sine-bell filter skewed toward t_1(max). This function reached a maximum at the end of the experimental data and fell off rapidly in the extrapolated portion of the signal, thereby minimizing the chance of errors caused by overemphasis of extrapolated points (Olejniczak & Eaton, 1990). Use of linear prediction in this way gave maximum weight to the experimental data at longer times, while bringing the signal smoothly to zero. Sinc ringing associated with the Fourier transform of a truncated signal was therefore avoided.

Detection of amide protons resistant to exchange with solvent was achieved by the rapid measurement of heteronuclear correlation spectra, after ^{15}N-labeled protein was dissolved in D_2O. The Bruker AM500 spectrometer was modified with a timing device (Tschudin Associates, Kensington, MD) which allowed for the acquisition of 2D spectra without storing data onto a disk after every serial file (Marion & Bax, 1989). It was possible to collect a full 2D data set in 15 min (2 scans per serial file). The experiment used a refocused INEPT type of transfer between ^1H and ^{15}N (Clore et al., 1990). Ten minutes after D_2O was added to the protein, the first experiment was begun and subsequent experiments were performed over a 32-h period. Spectra were collected 17, 32, 47, 62, 86, 135, 255, 382, 600, 1295, 1455 and 1955 min after mixing. Each experiment is referred to by the time that had elapsed from mixing to the midpoint of that experiment. For experiments at times greater than 1 h, the data were collected with 4 scans per serial file to provide adequate signal-to-noise. Data sets of 256 × 2K points were collected. The data were processed using a skewed sine-bell filter shifted by $\pi/3$ for both transformations. The data were zero-filled to 1024 points in f_1, to give a 1K × 1K data matrix. Volume integration was carried out on each of the cross-peaks and the scaled information was plotted as a function of time. The resulting curves were fitted to a single-exponential decay to calculate the apparent solvent exchange rates for each amide proton.

Structure Calculation. Determination of the folding topology of HPr was carried out using a variable target function algorithm, as implemented in the program DIANA (Güntert et al., 1991), which was licensed from Professor Kurt Wüthrich at ETH-Zurich. A minimal set of interproton distance constraints were identified according to the following criteria: (1) protons were not the same residue, unless they were separated by more than 4 bonds; (2) cross-peak assignment was unambiguous, i.e., there was no spectral overlap in either dimension; and (3) where applicable, NOESY interactions were confirmed in the HMQC-NOESY. A total of 117 such constraints were identified: two correlated protons within the same residue, 67 correlated protons from adjacent residues, and 48 correlated protons from nonsequential residues.

Distance estimates were established on the basis of relative cross-peak intensity. The intensities were divided into three categories: strong (2.0–3.0 Å), medium (2.0–4.0 Å), and weak (2.0–5.0 Å). Values of $^3J_{NH}$ obtained from analysis of the HMQC-J experiment were used to place constraints on the value of the dihedral angle ϕ (Bystrov 1976). For residues with $^3J_{NH}$ of 2–4 Hz, ϕ was constrained to values of $-60° \pm 15°$. When $^3J_{NH}$ = 4–6 Hz, ϕ was constrained to values of $-60° \pm 30°$. For $^3J_{NH}$ values of 8–9 Hz, ϕ was constrained to $-120° \pm 30°$. For $^3J_{NH}$ values >9 Hz, ϕ was constrained to $-120° \pm 15°$. A total of 64 constraints were established on the basis of the measured J values. All of the remaining ϕ dihedrals except those of glycine and proline were held within the range $-(30°–190°)$.

Hydrogen bond interactions were identified for amide protons that exhibited exchange half-times greater than 30 min (see Table II). The hydrogen bond acceptor, in each case, was the oxygen atom expected in the regular secondary structure that was predicted by the NOE pattern and the magnitude of $^3J_{NH}$. When such a clear assignment could not be made, no hydrogen bond constraint was used. A total of 25 hydrogen bonds were identified according to these criteria.

The distance constraints, dihedral angle constraints, and hydrogen bond interactions (entered as distance constraints, 1.7–1.9 Å) were used as the primary data input in DIANA. The calculations were done such that intraresidue and sequential distance constraints were taken into account first. The sequence separation of protons considered in the calculation was gradually increased until all constraints were included. Structures obtained from DIANA were displayed and plotted using QUANTA (Polygen Corp., Waltham, MA) on a Silicon Graphics 4D-35 workstation.

RESULTS

Resonance Assignments. The original sequence-specific resonance assignments were based on the 2D spectra obtainable in 1985: magnitude COSY and RELAY and phase-sensitive NOESY spectra (Klevit et al., 1986). Subsequent application of newer experiments such as TOCSY and DQF-COSY has allowed for more complete assignments of the

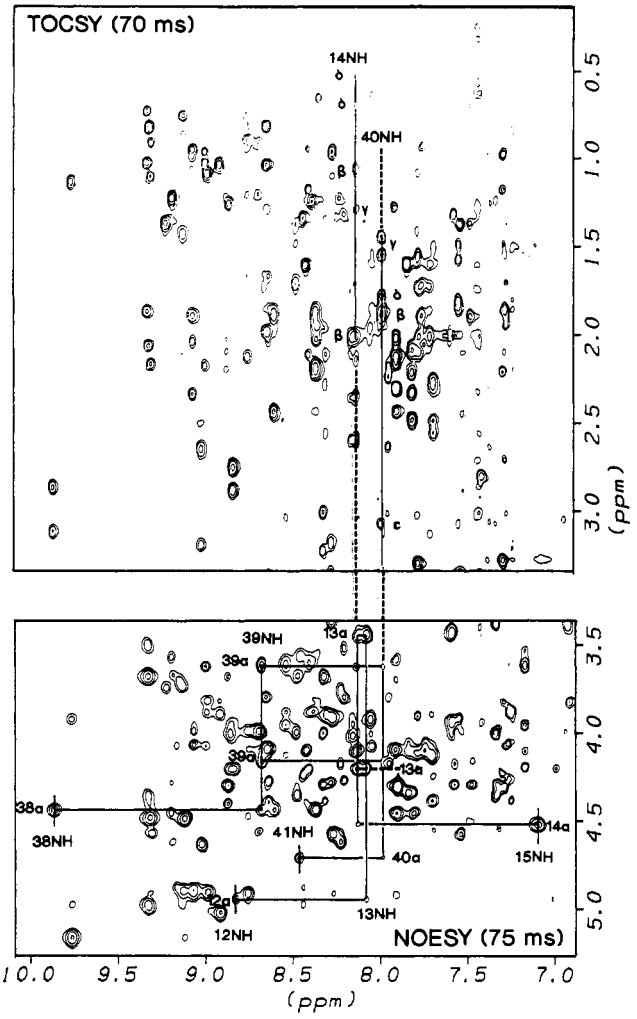

FIGURE 1: Homonuclear ^1H NMR spectra of *E. coli* HPr. This figure shows regions of TOCSY (upper pannel) and NOESY (lower panel) spectra obtained for HPr in H_2O. The vertical lines that run through both spectra indicate the positions of the amide proton resonances for Leu[14] and Lys[40]. Originally, the more downfield resonance was assigned to Lys[40]. The TOCSY spectrum reveals that the upfield resonance exhibits the pattern expected for a lysine spin system, while the downfield amide exhibits the spin system pattern for a leucine. The sequential connectivities for these two amide resonances are outlined in the NOESY spectrum in the lower panel. In both cases, the amide resonances in question give cross-peaks to both of the $C^{\alpha}H$ resonances from a glycine spin system.

side-chain proton resonances. Several TOCSY spectra were collected with different mixing times to account for the mixing time dependence of cross-peak intensities (Remerowski et al., 1989; Cavanaugh et al., 1990). In most cases, complete spin system assignments were obtained from these spectra.

The complete spin system assignments led to the resolution of a number of ambiguities in the earlier spectra and revealed incorrect assignments. An example of how the TOCSY spectra resolved ambiguities is illustrated in Figure 1 for the resonances of Asn[12]-Gly[13]-Leu[14] and Asn[38]-Gly[39]-Lys[40]. Translated into the level of spin system identification available during the original analysis, both of the tripeptide segments are AMQX-Gly-long side chain. Thus, distinguishing these two segments requires the ability of distinguish a leucine spin system from a lysine, which is straightforward with a TOCSY but virtually impossible with only COSY and RELAY spectra. The TOCSY indicated that the resonance originally assigned to Leu[14] gives the pattern for a lysine spin system, while that assigned to Lys[40] is clearly a leucine spin system. Thus, the original assignments for these two segments must be reversed.

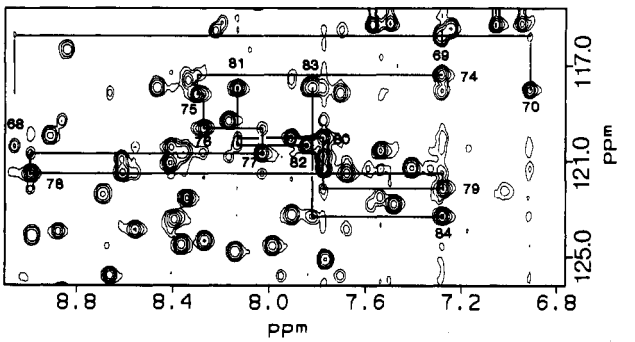

FIGURE 2: Region of the ^{15}N–^1H HMQC-NOESY spectrum. This figure illustrates how the HMQC-NOESY spectrum was useful in resolving ambiguities due to spectral overlap in the ^1H–^1H NOESY spectrum. Peaks labeled 69, 74, 79, and 84 are the HMQC peaks for the four amide resonances with δ_{NH} = 7.27, showing that the resonances are resolved in the ^{15}N dimension. Solid lines indicate the NOESY cross-peaks that correlate each ^{15}N resonance to its neighboring amide ^1H frequencies. (In the figure, the labeled peaks correspond to the HMQC cross-peak and the unlabeled peaks to the NOESY cross-peaks.) In all cases, the corresponding d_{NN} cross-peaks were observed in the ^1H–^1H NOESY and the HMQC-NOESY unambiguously revealing the correct connectivities among them.

There were breaks in the original sequential connectivity map on both sides of the tripeptide Asn[38]-Gly[39]-Lys[40] (see Figure 4). With the corrected assignments, a $d_{\beta N}$ connectivity now connects Ser[37] to Asn[38] and a $d_{\alpha N}$ connectivity connects Lys[40] to Ser[41]—additional confirmation that these are indeed the correct assignments.

A second example of how the complete spin system assignments led to a change in assignments is the tripeptide segment originally assigned as Thr[52]-Leu[53]-Gly[54]. Unambiguous spin system assignments were made for the threonine and glycine from COSY and RELAY spectra, but no specific identification other than long side chain could be made for the middle resonance in this set of resonances connected by $d_{\alpha N}$ [see Figure 6 of Klevit et al. (1986)]. However, there is another tripeptide segment in HPr of the type Thr-long side chain-Gly, i.e., Thr[56]-Gln[57]-Gly[58]. The TOCSY showed that the peak originally assigned to Leu[53] belongs to a six-proton spin system of the type Glu, Gln, Met, indicating that the original assignment was in error. Again, gaps in the original sequential connectivities could be filled in with these assignments, supplying further support for these assignments.

All the sequential assignments for HPr were checked in this way and corrected where necessary. However, even with the additional TOCSY information some ambiguities remained due to overlap in the ^1H spectrum. Heteronuclear 2D ^{15}N–^1H HMQC and HMQC-NOESY spectra obtained for uniformly-labeled ^{15}N-HPr were used to overcome this problem and to confirm assignments. For example, there are four amide protons and two aromatic protons that resonate at 7.27 ± 0.01 ppm. The homonuclear NOESY spectrum contains eight d_{NN} cross-peaks at 7.27 ppm, making unambiguous assignment of the sequential connectivities extremely difficult. In the HMQC spectrum, all four of the amide cross-peaks with δ_{NH} = 7.27 are resolved in the ^{15}N dimension. Thus, since protons that give d_{NN} peaks in a homonuclear NOESY give a distinctive cross-peak pattern in an HMQC-NOESY [i.e., peaks at ($^{15}N_i$, N^1H_{i+1}) and ($^{15}N_{i+1}$, N^1H_i)], the proper connectivities involving these overlapping amide protons could be unambiguously distinguished. This is illustrated in Figure 2.

Table I presents the updated ^1H resonance assignments for *E. coli* HPr. Backbone resonances for 33 residues have been reassigned. As well, complete assignments for the backbone ^{15}NH resonances were obtained, as listed in Table II.

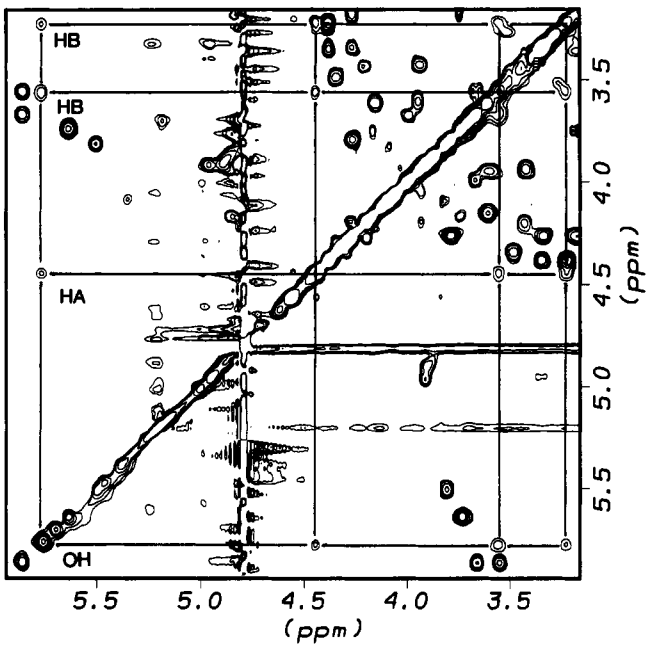

FIGURE 3: Spin system of Ser-31 (NH not shown) of *E. coli* HPr in a clean-TOCSY spectrum with 70-ms mixing period in 90% H_2O (see Materials and Methods for details).

Following the analysis described above, several protons remained difficult to assign. The C^{α} protons of His^{15}, Ser^{37}, and Ser^{43} generated no NH–H^{α} cross-peaks in coherence spectra, due to their coincidental saturation by the irradiation that was used to saturate the H_2O signal. The chemical shifts of these protons were deduced from coherence involving other protons (especially $C^{\beta}H$'s) or from NOESY spectra. In the process, the chemical shift of the His^{15} amide proton was reassigned. The side-chain protons of Lys^{27}, Lys^{45} and Leu^{47} were not completely assigned because the necessary coherence cross-peaks were not observed in any of our spectra. In addition, some side-chain resonances cannot be uniquely assigned to specific protons due to chemical shift coincidence (for example, some $C^{\beta}H_2$'s, $C^{\gamma}H_2$'s, etc.). The residues for which these ambiguities exist are noted in Table I.

There is sparse evidence for several sequential residue relationships. Specifically, the connectivities for Arg^{17}-Pro^{18}, Ser^{46}-Leu^{47}, Thr^{52}-Leu^{53}, Thr^{62}-Ile^{63}, and Ala^{82}-Glu^{83} are poorly represented in the spectra. The latter three connections are ambiguous due to the coincidence of resonance frequencies (see Table I). The Ser^{46}-Leu^{47} connection is difficult to observe, in part because of the inability to observe all the Leu^{47} resonances. With the exception of the Glu^{83}-Leu^{84}-Glu^{85} segment, strong cases could be made for the sequential connectivities on either side of each of the sequence ambiguities.

Side-Chain Hydroxyl Protons. In the process of checking assignments, we noted two unusual resonances. Both appear in the region just downfield of the H_2O signal, where $C^{\alpha}H$ protons resonate, but are observable only in H_2O. This suggests that they arise from exchangeable protons; however, there is no evidence in the ^{15}N–^{1}H spectra that either exchangeable proton is attached to nitrogen. One of the resonances, at 5.77 ppm, gives cross-peaks to the $C^{\alpha}H$ and both $C^{\beta}H$'s of Ser^{31} in TOCSY spectra (Figure 3). The other resonance, at 5.31 ppm, gives TOCSY cross peaks to the $C^{\alpha}H$ and/or $C^{\beta}H$ (their chemical shifts coincide) and the $C^{\gamma}H3$ resonance of Thr^{59}. An investigation of the temperature dependence of these resonances revealed that their chemical shifts are much more sensitive to temperature than either of the other ($C^{\alpha}H$) resonances in this region of the spectrum or the aromatic resonances. Taken together, these observations lead to the rare identification of these resonances as hydroxyl protons of Ser^{31} and Thr^{59}, respectively. The fact that their exchange rates are slow enough to make them observable in the H_2O spectra suggests that they are involved in hydrogen bonds. Consistent with this proposal, NOESY cross-peaks are observed between the hydroxyl resonance of Ser^{31} and the amide proton resonances of Asp^{69} and Glu^{70}. A slowly exchanging hydroxyl proton has also been observed for Ser^{31} in an analogous HPr molecule, from *Staphylococcus aureus* (H. R. Kalbitzer, personal communication). A brief search of the recent literature revealed the report of one other hydroxyl proton in the H_2O spectrum of a protein: that of Thr^{77} in reduced *E. coli* thioredoxin (Dyson et al., 1989).

Additional Spectral Characterization. Although identifications of slowly-exchanging amide protons were made in the original report (Klevit & Waygood, 1986), these were obtained from a COSY spectrum that took 16 h to acquire. With newer technical advances, 2D spectra can be collected in just under 20 min, allowing finer time resolution of the exchange behavior (see Materials and Methods). Spectra were collected on a sample of HPr that had been freshly dissolved in D_2O at pH 6.5, and exchange half-times under these conditions were measured (see Table II). Amide protons were characterized as having fast, medium, or slow exchange rates on the basis of their having half-times shorter than 10 min, between 10 and 30 min, and longer than 30 min, respectively (see Figure 4).

Twenty-six amide protons had been identified as slow exchangers on the basis of the original data. As shown in Figure 4, there are 30 slowly-exchanging amide protons observed in the newer measurements. These are all in regions of regular secondary structure (i.e., α-helices and β-sheet), as originally identified (Klevit & Waygood, 1986). Changes in resonance assignments resulted in the change of identity of several resonances originally labeled as slow exchangers: Ala^{82} and Glu^{83} (originally slow exchangers) are medium exchangers and Gln^{71} exchanges rapidly (the cross-peak originally assigned as Gln^{71} was reassigned to Ile^{63}). The exchange half-times for amide protons in helices A (residues 18–28) and C (residues ~70–84) and β-strands A (residues 1–8), B (residues 31–37), and D (residues 60–67) indicate that these sections of the structure are stabilized by a network of hydrogen bonds. Exchange half-times decrease near the ends of these elements of secondary structure. The exchange rates are also faster in the secondary structure elements previously defined as helix B (residues 46–52) and β-strand C (residues 41–44).

Coupling constants between amide and C^{α} protons were measured from a ^{15}N–^{1}H HMQC-J spectrum collected on a uniformly ^{15}N-labeled sample (see Materials and Methods and Table II). As expected for a protein with regular secondary structure, there are continuous stretches of residues that have large coupling constants ($^{3}J_{NH} \geq 8$ Hz) and stretches that exhibit small couplings ($^{3}J_{NH} \leq 6$ Hz). In the former category are the four previously identified β-strands and in the latter category are the two long α-helices, α-helix A and α-helix C. The sequence originally identified as helix B exhibits somewhat greater coupling constant variation.

Secondary Structure. The originally reported sequential connectivities and those derived from the updated assignments are compared in Figure 4. The majority of sequential connectivity types remain the same, particularly in regions with regular secondary structure, as characterized by continuous stretches of $d_{\alpha N}$ or d_{NN} connectivities. On the basis of the new assignments, many of the breaks in the original sequential connectivity map were filled in. Most of the differences due

Table I: ^{1}H Resonance Assignments for HPr[a]

residue	NH	C$^\alpha$H	C$^\beta$H	C$^\gamma$H	C$^\delta$H	other
M1	*b*	4.63	2.30, 2.10	2.50, 2.50	*c*	
F2	9.01	4.83	3.19, 2.65		7.28[c]	C$^\epsilon$H 7.39; C$^\zeta$H 7.47
Q3	7.76	5.76	1.90, 1.75	2.00, 2.09		N$^\epsilon$H$_2$ 7.46, 6.72
Q4	8.59	4.43	2.44, 1.89	2.44, 1.89		N$^\epsilon$H$_2$ 7.50, 6.91
E5	8.36	5.47	1.89, 1.84	2.21, 2.17		
V6	9.29	4.57	2.18	0.91, 1.10		
T7	8.25	4.91	3.90	0.96		
I8	8.75	3.98	2.13	1.66, 1.20, 0.89	0.76	
T9	8.68	4.56	4.44	1.20		
A10	7.53	4.58	1.37			
P11		4.20*	2.38, 1.95	2.91, 2.19	3.93, 4.04	
N12	8.83*	4.93*	2.75, 2.88			N$^\delta$H$_2$ 7.52, 6.96
G13	8.07*	4.20*, 3.43*				
L14	8.13*	4.52*	2.02, 1.06	1.29	0.70, 0.53	
H15	7.07*	4.79	3.27, 3.51		7.29	C$^\epsilon$H 7.98
T16	8.19	3.79	4.26	1.32		
R17	8.63	4.38	2.04, 1.99	1.84, 1.67	3.34, 3.34	*c*
P18		4.56*	2.35[d]	3.18	4.09, 3.84	
A19	8.38*	3.87*	1.24			
A20	8.42	4.10	1.60			
Q21	7.69	4.12	2.30, 2.27	2.48, 2.54		N$^\epsilon$H$_2$ 7.92, 6.92
F22	8.53	3.95	3.59, 3.03		7.16, 7.16	C$^\epsilon$H 7.30, 7.30; C$^\zeta$H 6.96
V23	8.40	4.20	2.11	0.95, 1.18		
K24	7.78	3.89	2.06, 1.99	1.56, 1.40[d]	1.74, 1.74	C$^\epsilon$H 3.02, 3.02
E25	8.15	4.01	1.99, 2.02	2.36, 2.61		
A26	8.84	3.99	1.26			
K27	8.05	3.93	2.10, 1.92	*c*	*c*	C$^\epsilon$H 3.07, 3.07
G28	7.58	3.42, 3.93				
F29	7.42*	4.78	3.39, 2.81		7.32, 7.32	C$^\epsilon$H 6.92, 6.92; C$^\zeta$H 6.63
T30	11.01	4.20	4.27	1.32		
S31	9.18*	4.44	3.23, 3.55			O$_\gamma$H 5.77
E32	7.89	4.49	2.09, 2.21	2.44, 2.44		
I33	9.11	5.15	1.44	0.75, 0.86, 1.64	0.57	
T34	9.75	4.97	3.91	1.14		
V35	9.31	4.89	1.87	1.03, 0.73		
T36	8.97	5.50	3.81	1.08		
S37	9.13	4.82	3.82, 3.70			
N38	9.85*	4.43*	3.12, 2.86			N$^\delta$H 7.57, 7.05
G39	8.67*	4.15*, 3.61*				
K40	7.98*	4.71*	1.91, 1.86	1.55, 1.44	1.88, 1.78	C$^\epsilon$H 3.08, 3.08
S41	8.45	5.86	3.66, 3.55			
A42	9.18	4.61	1.22			
S43	8.22	4.87	4.16, 3.76			
A44	8.43	4.23	1.33			
K45	7.70	4.37*	1.50, *c*	*c*	*c*	*c*
S46	7.24*	4.86*	3.87, 3.87			
L47	9.16*	4.07*	*c*	*c*	*c*	
F48	8.30	4.38	3.22, 3.00		7.26, 7.26	C$^\epsilon$H 7.39, 7.39; C$^\zeta$H 7.34
K49	7.71	4.12*	2.01, 2.01	1.61, *c*	1.87, 1.73	C$^\epsilon$H 3.12, 3.12
L50	8.64	3.79	1.85, 1.73	1.64	0.81, 0.81	
Q51	7.94	4.18	2.18, 2.14	2.63, 2.26		N$^\epsilon$H$_2$ 7.35, 6.83
T52	7.56	4.29	4.36	1.32		
L53	7.49*	4.30*	1.92, 1.90	1.38	0.72, 0.86	
G54	8.32*	3.99*, 3.67*				
L55	8.47*	4.35*	1.72, 1.49	1.37	0.70, 0.79	
T56	7.33*	4.30*	3.93	1.24		
Q57	8.86*	3.66*	2.03, 2.03[d]	2.10, 2.21		N$^\epsilon$H$_2$ 7.31
G58	9.33*	4.34*, 3.49*				
T59	7.90*	4.09*[d]	4.09	1.28		O$^\gamma$H 5.31
V60	8.63	4.47	1.98	0.81, 1.03		
V61	9.30	5.01	2.07	0.83, 0.83		
T62	8.91	4.90	3.88	1.03		
I63	9.06	4.88*	1.90	1.53, 1.06, 0.95	1.00	
S64	9.11	5.64	3.73, 3.73			
A65	9.21	5.70	1.37			
E66	8.36	5.38	2.04, 1.94	2.21, 2.17		
G67	10.19	4.87, 3.97				
E68	9.06*	4.06*	2.04, 2.04	2.33, 2.33		
D69	7.28	4.95	2.69, 3.35			
E70	6.91	3.82	1.87, 2.15	2.37, 2.76		
Q71	7.77*	3.30*	1.56, 1.47	1.56, 1.09		N$^\epsilon$H$_2$ 7.50, 6.60
K72	7.53*	3.91*	1.79, 1.55	1.58, 1.50	1.74, 1.74[d]	C$^\epsilon$H 3.07, 3.07
A73	8.35	1.95*	0.66			
V74	7.28	3.28	2.21	0.97, 1.17		
E75	8.30	3.87	1.95, 2.11	2.27, 2.56		
H76	8.27	4.27	3.18, 3.36		6.78	C$^\epsilon$H 8.15
L77	8.04	4.07	1.87, 1.62	1.50	0.63, 0.28	

Table I (Continued)

residue	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	$C^{\delta}H$	other
V78	8.98	3.60	2.17	1.03, 0.97		
K79	7.26	4.05	1.98, 1.73	1.56, 1.37[b]	1.87, 1.87	$C^{\epsilon}H$ 3.00, 3.00
L80	7.75	4.09*	2.14, 1.89	1.72	0.94, 0.94	
M81	8.12	4.09	1.71, 2.13	2.60, 2.56, 2.35	*c*	
A82	7.83*	4.33*	1.59			
E83	7.81*	4.45*	2.21, 2.09[d]	2.33, 2.48		
L84	7.28*	4.29*	1.49, 1.38	2.10		
E85	7.89*	4.32*	2.02, 2.15[d]	2.31, 2.31		

[a] Asterisks indicate resonances with corrected assignments. [b] Not observed. [c] Not assigned. [d] Assigned according to conventional chemical shift relationship (e.g., $C^{\gamma}H$ downfield of $C^{\beta}H$ in Glu, Gln).

Table II: ^{15}N Amide Chemical Shifts, ^{15}N–^{1}H Coupling Constants, and Amide Exchange Rates for *E. coli* HPr at pH 6.5 and 30 °C

residue	^{15}NH[a]	$^{3}J_{NH\alpha}$[b]	$t^{1/2}$ (min)[c]	residue	^{15}NH[a]	$^{3}J_{NH\alpha}$[b]	$t^{1/2}$ (min)[c]
F2	131.0	8.8		A44	128.4	5.0	
Q3	125.1	8.4		K45	109.6	8.9	13
Q4	121.4	7.5	188	S46	115.6	9.7	<10
E5	124.5	8.7		L47	133.1		
V6	122.9	8.6	702	F48	117.6	3.0	
T7	124.3	9.0		K49	118.2	5.8	
I8	128.3	5.6	172	L50	125.7	<2.5	
T9	122.4	9.4		Q51	113.8	6.4	
A10	132.6	2.5		T52	111.4	7.3	
P11				L53	122.7	7.1	<10
N12	116.4	8.9		G54	109.9		
G13	108.6	5.5		L55	126.8		
L14	124.7	9.1		T56	121.3		
H15	122.1			Q57	123.9	<2.5	
T16	115.2			G58	117.6	6.5	
R17	121.0	3.8		T59	120.0	5.1	<10
P18				V60	129.7	9.0	
A19	123.4	3.8		V61	127.4	9.9	*d*
A20	121.0	4.4		T62	120.0	9.1	326
Q21	121.5	5.9		I63	131.0	9.5	799
F22	123.8	3.2	31	S64	122.9	9.9	461
V23	120.5	3.9	*d*	A65	125.8	9.2	308
K24	121.1	3.9	92	E66	120.6	9.5	*d*
E25	119.3	6.2	75	G67	120.4	6.6	<10
A26	126.3	3.0	108	E68	120.4	11.4	
K27	114.4	3.8	89	D69	115.9	10.5	
G28	107.5	5.9/6.6		E70	118.1	<2.5	75
F29	121.3	10.6	97	Q71	121.3	5.6	
T30	120.7	6.9		K72	120.6	5.5	<10
S31	121.8	<2.5	<10	A73	122.6	5.9	800
E32	127.1	8.8	30	V74	117.4	5.4	1619
I33	131.0	11.4		E75	118.3	<2.5	668
T34	127.2	11.0	233	H76	119.6	5.2	38
V35	127.9	9.7	*d*	L77	120.7	5.9	282
T36	124.1	10.5	130	V78	121.5	5.2	1411
S37	120.8	7.3	10	K79	122.1	5.2	186
N38	129.3			L80	120.0	2.9	363
G39	104.8	5.9/6.9		M81	118.0	4.4	306
K40	124.5	9.6	<10	A82	120.4	5.8	<10
S41	118.0	9.5		E83	117.9	9.0	<10
A42	125.3	8.3	20	L84	123.3	5.9	
S43	115.5	3.6		E85	123.9	8.3	

[a] Chemical shift relative to liquid NH_3. [b] Blank entries were not observed. [c] Blank entries exchanged too rapidly for measurement. [d] Quantitation was not possible due to resonance overlap.

to assignment changes are in irregular regions, such as residues 13–15, 46–60, and 68–71, as illustrated in Figure 4.

The characteristic spectral properties for β-strands include strong $d_{\alpha N}$ connectivities, large $^{3}J_{NH}$ coupling constants, and slow amide exchange rates. There are four stretches of sequence in *E. coli* HPr that exhibit these characteristics: residues 1–8, 32–36, 40–42, and 60–66. We originally identified the four β-strands as residues 1–9 (strand A), 31–37 (strand B), 41–43 (strand C), and 61–67 (strand D). Thus, this interpretation has not changed significantly in this subsequent analysis.

NOEs between protons on neighboring strands define the topology of a β-sheet. In our original analysis, 20 such cross-strand NOEs were identified: 9 $C^{\alpha}H$–$C^{\alpha}H$ NOEs, 6 NH–NH NOEs, and 5 $C^{\alpha}H$–NH NOEs [see Figure 6 of Klevit and Waygood (1986)]. *None* of the resonances involved in these 20 cross-strand NOEs have been reassigned in the present analysis (Table I). In fact, only two resonances in the four β-strands were reassigned, the NH of Ser^{31} and the $C^{\alpha}H$ of Ile^{63}, and neither of these were involved in cross-strand NOEs. Therefore, the topology of the β-sheet originally described (A-D-B-C) is consistent with the updated assignments.

The spectral characteristics of α-helices include sequential d_{NN} and $d_{\beta N}$ connectivities, medium-range $i, i+3$ connectivites, small $^{3}J_{NH}$ coupling constants, and slow amide exchange rates. There are two stretches of sequence that clearly exhibit this behavior, namely, residues 19–28 and 70–84. These correspond to helices A and C in the original report. We had

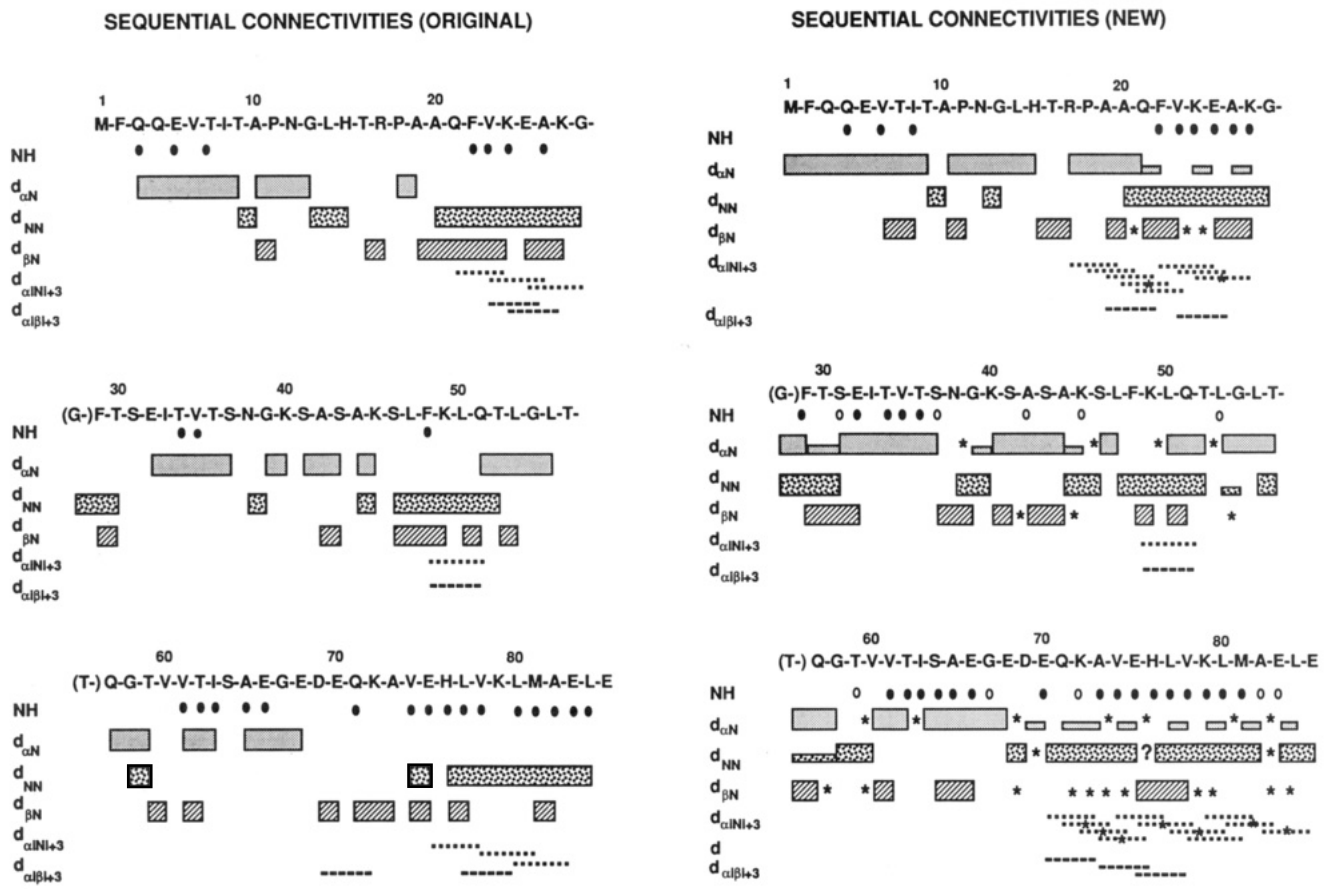

FIGURE 4: Summary of the sequential connectivities observed for *E. coli* HPr. (Left panel) Sequential connectivities originally reported in Klevit and Waygood (1986). (Right panel) Sequential connectivities based on the expanded and corrected assignments reported here. Horizontal bars signify the presence of a NOESY cross-peak of the type indicated. Narrower bars indicate weak cross-peaks. Asterisks (*) indicate that cross-peak intensity is observed at this position, but spectral overlap exists, making the assignment ambiguous. A question mark (?) indicates that the two resonances in question are too close together to observe the expected peak. Filled circles indicate amide protons with exchange half-times greater than 30 min; open circles indicate amide protons with exchange half-times between 10 and 30 min.

also proposed that a third helix exists at residues 46–52, called helix B. This proposal was based on the observation of a series of sequential d_{NN} connectivities from residues 46–52, i, i + 3 connectivities between residues 48 and 51, and a slowly exchanging amide proton at residue 48. As illustrated in Figure 4, our more recent analysis also shows d_{NN} connectivities for residues 47–52 and the i, i + 3 connectivities between residues 48 and 51, but the identification of Phe[48] as a slowly exchanging amide proton was in error. Instead, Leu[53] is now identified as a medium-exchanging proton. Finally, the $^3J_{NH}$ values are all fairly small for residues 48–51 (Table II). Therefore, it appears that there is some sort of α-helical turn structure involving the residues from 47 to 53, but since this represents only two helical turns, it is likely to be a less stable structure than the longer helices A and C (hence the faster amide exchange rates). A similar conclusion was reached on the basis of the behavior of the residues in this region in the *B. subtilis* HPr protein (Wittekind et al., 1990).

Structure Determination. The significant number of new resonance assignments revealed by our recent studies raised the issue of their impact on the folding topology proposed on the basis of the old assignments. Therefore, distance constraints based on observed NOESY cross peaks, angle constraints based on the measured coupling constants, and hydrogen bond constraints based on solvent exchange properties were used as input for a variable target function algorithm program, DIANA (see Materials and Methods). In order to avoid introducing any of our own structural bias into this calculation, a conservative approach was adopted in dealing with the data. In particular, only those cross-peaks for which completely unambiguous assignments were available (i.e., no spectral overlap in *either* dimension) were used. For example, the cross-peak that was identified as Ser[31] C^αH–Gly[67] C^αH in our original analysis (Waygood & Klevit 1986) could instead be Asn[38] C^α–Gly[67] C^αH due to spectral overlap. Even though on the basis of the predicted β-sheet topology the former assignment would be the only one that is structurally feasible, this cross-peak was *not* used in the structure calculation. Intraresidue constraints, in general, were not used since they contribute little information to the folding topology. Although this protocol resulted in a minimal set of distance constraints (117 distance constraints, 64 ϕ angle constraints, and 25 hydrogen bond constraints; included as supplementary material), we felt that this was the most accurate representation of what was actually known from the existing data. While these minimal constraints not surprisingly resulted in an underdetermined structure, they were sufficient to define the overall folding topology of the protein.

In all, 75 structures were calculated using DIANA. They were ranked on the basis of the number of major deviations from the input constraints. A major deviation was considered to be one that was out of range by 0.75 Å for an upper distance limit, 1.0 Å for lower distance limits and van der Waals contacts, and 15° for dihedral angles. Portions of the twelve structures with the fewest major constraint violations are depicted in Figure 5. These representations show the α-carbon positions of residues involved in the β-sheet and the two well-defined α-helices. The superimposed structures reveal

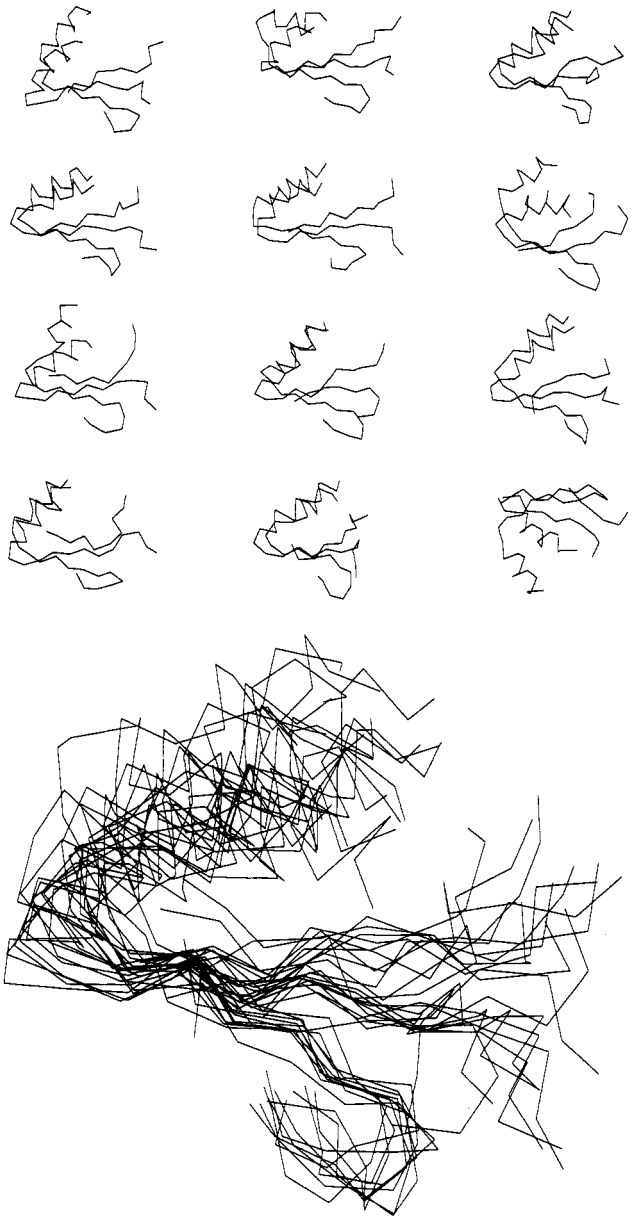

FIGURE 5: Representative structures calculated using a minimal set of constraints. Only the defined regions of the structures are depicted (residues 1–10, 10–42, and 57–82). The positions of the α-carbon atoms of the β-sheet residues were fit by a pairwise least squares calculation, resulting in RMSD of ca. 2 Å for each pair. (Top panel) Individual representation of the 12 structures containing the fewest major constraint violatons (see text). The view of each structure is identical. (Bottom panel) Superposition of 11 structures with the same overall symmetry, from the same perspective used in the top panel.

the consistent assembly of the β-sheet in the topology described above. The precise orientation of the α-helices relative to the β-sheet varies among these structures as shown in the individual representations. This is undoubtedly due to the paucity of helix–sheet constraints used in the calculation. The two helices run antiparallel to one another as originally proposed. The topology shown in the lower right corner of Figure 5 (top panel) appears quite different from the others but differs merely by the symmetry operation of reflection. In this case, the segments connecting the elements of β and α secondary structure make turns in the sense opposite to the turns in the other structures.

It is important to point out that HPr represents a special class of protein structures that can be defined by such a minimal set of constraints. Of the 48 nonsequential constraints used, 23 were between protons in the β-sheet and unambiguously define the relative positions of the four β-strands in the sheet, while 9 were between protons in the two well-defined α-helices, fixing the relative orientation of these two elements of secondary structure. In addition, the topology of the β-sheet itself and the β-α-β supersecondary structure of this protein effectively constrain the overall folding pattern of the molecule, since right-handed crossovers of the helical segments between the β-strands will occur. The fold that is consistent with the constraints is a four-stranded β-sheet with the helices crossing over one face of the sheet.

Discussion

Data collected since the early reports of 2D NMR studies on *E. coli* HPr have extended the analysis of this protein. New 2D homonuclear ^{1}H spectra and 2D heteronuclear ^{15}N–^{1}H spectra resolved ambiguities and revealed incorrect assignments that had been made on the basis of the more limited data set that was attainable with the technology available at that time. In addition, a more detailed analysis of the solvent-exchange rates of backbone amide protons has been performed and $^3J_{NH}$ coupling constants were measured. These data offer additional information regarding the solution structure of HPr.

Since the new analysis resulted in a significant number of assignment changes, it was imperative to ascertain the impact of these changes on the proposed folding topology of HPr. On the basis of the patterns of sequential connectivities, medium-range connectivities, amide exchange rates, and coupling constants, the main conclusions regarding the secondary structure of *E. coli* HPr remain the same. This is not surprising since most of the assignment changes were in regions that exhibit irregular connectivity patterns, i.e., the stretches of the protein that do not exist in regular secondary structure and serve to connect the long elements of regular secondary structure. The one exception is the short helix B, originally proposed to exist between residues 46 and 52. While its spectral properties are similar to those of a helical segment, the paucity of confirmatory medium-range connectivities and slowly-exchanging amide protons suggest that this region of the polypeptide is either not a regular α-helix or is at least more dynamical in nature.

In the process of the more recent spectral analysis, two unusual resonances were observed whose behavior leads to the conclusion that they arise from slowly-exchanging side-chain hydroxyl protons. The two residues with observable OH resonances are Ser^{31} and Thr^{59}. Both are at the edges of the β-sheet, at the N-terminal ends of the two internal strands, β-strand B (Ser^{31}) and β-strand D (Thr^{59}). Their spectral behavior strongly suggests that these side-chain hydroxyl groups are involved in hydrogen bonds, probably within the β-sheet structure. These may contribute to the extreme thermal stability of HPr, as a sequence comparison of HPrs from a variety of sources reveals that the serine at position 31 is fairly highly conserved, while position 59 is often either a threonine or a serine residue (Waygood et al., 1989). Studies of variants and mutants lacking these hydroxyl-containing side chains may shed light on their function and importance.

Because the 2D homonuclear ^{1}H NOESY spectrum suffers from a fair degree of spectral overlap, a quantitative determination of the three-dimensional structure of HPr in solution will require the use of 3D spectra. There are, however, enough unambiguous NOESY cross-peaks, coupling constants, and slowly-exchanging amide protons discernible from the 2D spectra to uniquely define the folding topology of the protein. The preliminary structure calculations reported here indicate that the topology is as originally described, i.e., a four-stranded antiparallel β-sheet of topology A-D-B-C with α-helices run-

ning along one face of the sheet. At this stage, we feel it is prudent to propose that there are two α-helices in the structure, rather than the three originally proposed, as there are not enough helical constraints observable for residues 47–52. Thus, this makes the topology virtually identical to that more recently reported on the basis of 2D NMR analysis of *B. subtilis* HPr (Wittekind et al., 1990) and *S. aureus* HPr (H. R. Kalbitzer, personal communication). With respect to the differences between tertiary structures of *E. coli* HPr derived from 2D NMR and X-ray diffraction, an assessment by epitope mapping indicates that the structure derived from NMR data is the physiologically relevant one (Sharma et al., 1991). A more detailed discussion of the structure must await a well-defined structure, obtained using many more distance constraints interpretable only from higher-dimensioned NMR spectra.

Acknowledgments

Thanks go to Dr. Tom Pratum for providing the linear prediction program. We thank Joan Smallshaw for investigating the growth conditions for ^{15}N-HPr and Alice Leung for her protein purification.

Supplementary Material Available

Two tables, listing upper distance limits for distance constraints and angle ranges used for dihedral constraints (7 pages). Ordering information is given on any current masthead page.

References

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185–189.
Burg, J. P. (1967) Proceedings of the 37th Meeting of the Society of Exploration Geophysicists, October 31, 1967.
Bystrov, V. F. (1976) *Prog. Nucl. Magn. Reson. Spectrosc. 10*, 41–81.
Cavanaugh, J., Chazin, W. J., & Rance, M. (1990) *J. Magn. Reson. 87*, 110–131.
Driscoll, P. C., Gronenborn, A. M., Wingfield, P. T., & Clore, G. M. (1990) *Biochemistry 29*, 4668–4682.
Dyson, H. J., Holmgren, A., & Wright, P. E. (1989) *Biochemistry 28*, 7074–7087.
El-Kabbani, O. A. L., Waygood, E. B., & Delbaere, L. T. J. (1987) *J. Biol. Chem. 262*, 12926–12929.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.
Gronenborn, A. M., Bax, A., Wingfield, P. T., & Clore, G. M. (1989) *FEBS Lett. 243*, 93–98.
Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517–530.
Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110–126.
Klevit, R. E., & Drobny, G. P. (1986) *Biochemistry 25*, 7770–7773.
Klevit, R. E., & Waygood, E. B. (1986) *Biochemistry 25*, 7774–7781.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989) *J. Magn. Reson. 85*, 397–399.
Muller, N., Ernst, R. R., & Wüthrich, K. (1986) *J. Am. Chem. Soc. 108*, 6482–6492.
Olejniczak, E. T., & Eaton, H. L. (1990) *J. Magn. Reson. 87*, 628–632.
Remerowski, M. L., Glaser, S. J., & Drobny, G. P. (1989) *Mol. Phys. 68*, 1191–1218.
Shaka, A. J., Barker, B. P., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547–552.
Sharma, S., Georges, F., Delbaere, L. T. J., Lee, J. S., Klevit, R. E., & Waygood, E. B. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4877–4881.
Waygood, E. B., Sharma, S., Bhanot, P., El-Kabbani, O. A. L., Delbaere, L. T. J., Georges, F., Wittekind, M. G., & Klevit, R. E. (1989) *FEMS Microbiol. Rev. 63*, 43–52.
Wittekind, M. G., Reizer, J., & Klevit, R. E. (1990) *Biochemistry 29*, 7191–7200.

CORRECTION

Functional Modulation of the Isolated Glycoprotein Ib Binding Domain of von Willebrand Factor Expressed in *Escherichia coli*, by Mitsuhiko Sugimoto, George Ricca, Michael E. Hrinda, Alain B. Schreiber, George H. Searfoss, Enrica Bottini, and Zaverio M. Ruggeri*, Volume 30, Number 21, May 28, 1991, pages 5202–5209.

Page 5205. In column 2, line 6, between 30 and 50 μM should read between 30 and 50 nM.