Two-Dimensional ^1H NMR Studies of Histidine-Containing Protein from *Escherichia coli*. 1. Sequential Resonance Assignments†

Rachel E. Klevit* and Gary P. Drobny

Department of Chemistry, University of Washington, Seattle, Washington 98195

E. Bruce Waygood

Department of Biochemistry, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0, Canada

Received December 6, 1985; Revised Manuscript Received June 5, 1986

ABSTRACT: Two-dimensional NMR studies at 500 MHz have been performed on the histidine-containing protein (HPr) from *Escherichia coli*. HPr is one of the phosphocarrier proteins involved in the bacterial phosphoenolpyruvate:sugar phosphotransferase system that is responsible for the concomitant phosphorylation and translocation of a number of sugars. Sequential resonance assignments of HPr are complete. The conventional method of sequential assignments involving *J*-correlated spectroscopy (COSY) and nuclear Overhauser spectroscopy (NOESY) has been supplemented by optimized relayed coherence transfer spectroscopy (RELAY) to help overcome the spectral overlap that is inevitable in the spectra of proteins the size of HPr. RELAY experiments were performed in H_2O to obtain NH–$C^\beta H$ connectivities and in D_2O to obtain $C^\alpha H$–$C^\gamma H$ connectivities. The abundance of relayed coherence transfer peaks in the two experiments greatly aided in the assignment process of the complicated protein spectrum. The assignments lay the groundwork for the determination of the solution structure of HPr, as described in the accompanying paper [Klevit, R. E., & Waygood, E. B. (1986) *Biochemistry* (third paper of three in this issue)].

The phosphoenolpyruvate:sugar phosphotransferase system (PTS)[1] is responsible for the simultaneous uptake and phosphorylation of disaccharides, hexoses, and hexitols in many anaerobic and facultatively anaerobic bacteria [for a review, see Postma and Lengeler (1985)]. In general, the PTS system consists of four proteins: histidine-containing protein (HPr) is a phosphocarrier protein that transfers a phosphoryl group from enzyme I to either another phosphocarrier protein, factor III, or enzyme II. All HPr proteins from various bacteria have a conserved site of phosphorylation–dephosphorylation that is the imidazole N-1 of His-15 (Anderson et al., 1971; Simoni et al., 1973; Kalbitzer et al., 1982).

Many of the protein components of the PTS have been purified, but only preliminary information has been available concerning the tertiary structure of HPr and the mechanism of phosphoryl transfer (Waygood et al., 1985). While X-ray diffraction techniques are being used to study HPr (Delbaere et al., 1982), two-dimensional ^1H NMR (2DNMR) offers the advantage of obtaining a solution structure at a pH that is more physiological than the conditions under which the crystals were grown. Moreover, the instability of the 1-phosphohistidinyl residue formed at the active site of HPr (Waygood et al., 1985) seems to preclude an X-ray diffraction determination of the tertiary structure of phospho-HPr, while 2DNMR may allow a determination of this protein containing a phosphoramidate bond. In this series of papers, we describe the determination of the solution structure of HPr by 2DNMR. In the first paper, the sequence-specific assignments for the amide, $C^\alpha H$, and $C^\beta H$ resonances of the HPr spectrum are reported. In the second paper, the application of two-dimensional multipulse experiments that generate long-range relayed coherence to obtain sequence-specific assignments for $C^\delta H_3$ resonances of leucine and isoleucine residues is described. In the third paper, the assignments are used to identify both short-range nuclear Overhauser effects (NOEs) and long-range NOEs from which both the secondary structure and the tertiary fold of the protein are obtained.

Materials and Methods

Sample Preparation. HPr was purified from *Escherichia coli* P650 as previously described (Waygood & Steeves, 1980). Following the last step in the purification, the protein was lyophilized from H_2O. To prepare HPr samples for the NMR experiments, the protein was dissolved in 4 mL of 5mM potassium phosphate buffer, pH 6.5, and dialyzed exhaustively against the same buffer. The sample was again lyophilized and dissolved in 0.4 mL of either 99.96% D_2O or 90% H_2O/10% D_2O, centrifuged to remove any insoluble material, and put into a 5-mm NMR tube. The final concentration of HPr in the NMR sample was ~5 mM, in 50 mM potassium phosphate buffer, pH 6.5.

NMR Spectroscopy. NMR spectra were obtained on a Bruker WM-500 spectrometer. Two-dimensional nuclear Overhauser effect spectra (NOESY) were obtained in the

† This research was supported by National Institutes of Health Grants AM-35187-01 (R.E.K.) and GM-32681-01 (G.P.D.) and an operating grant from the MRC of Canada (E.B.W.). The 500-MHz spectrometer at the University of Washington is supported by instrumentation grants from the Murdock Foundation, the National Science Foundation (PCM80-18053), and the National Institutes of Health (GM28764-01S1).

* Address correspondence to this author at the Department of Biochemistry, SJ-70, University of Washington, Seattle, WA 98195.

[1] Abbreviations: PTS, phosphoenolpyruvate:sugar phosphotransferase system; HPr, histidine-containing protein; 2DNMR, two-dimensional NMR spectroscopy; NOE, nuclear Overhauser effect; COSY, two-dimensional *J*-correlated spectroscopy; NOESY, two-dimensional NOE spectroscopy; RELAY, two-dimensional relayed coherence transfer spectroscopy.

0006-2960/86/0425-7760$01.50/0

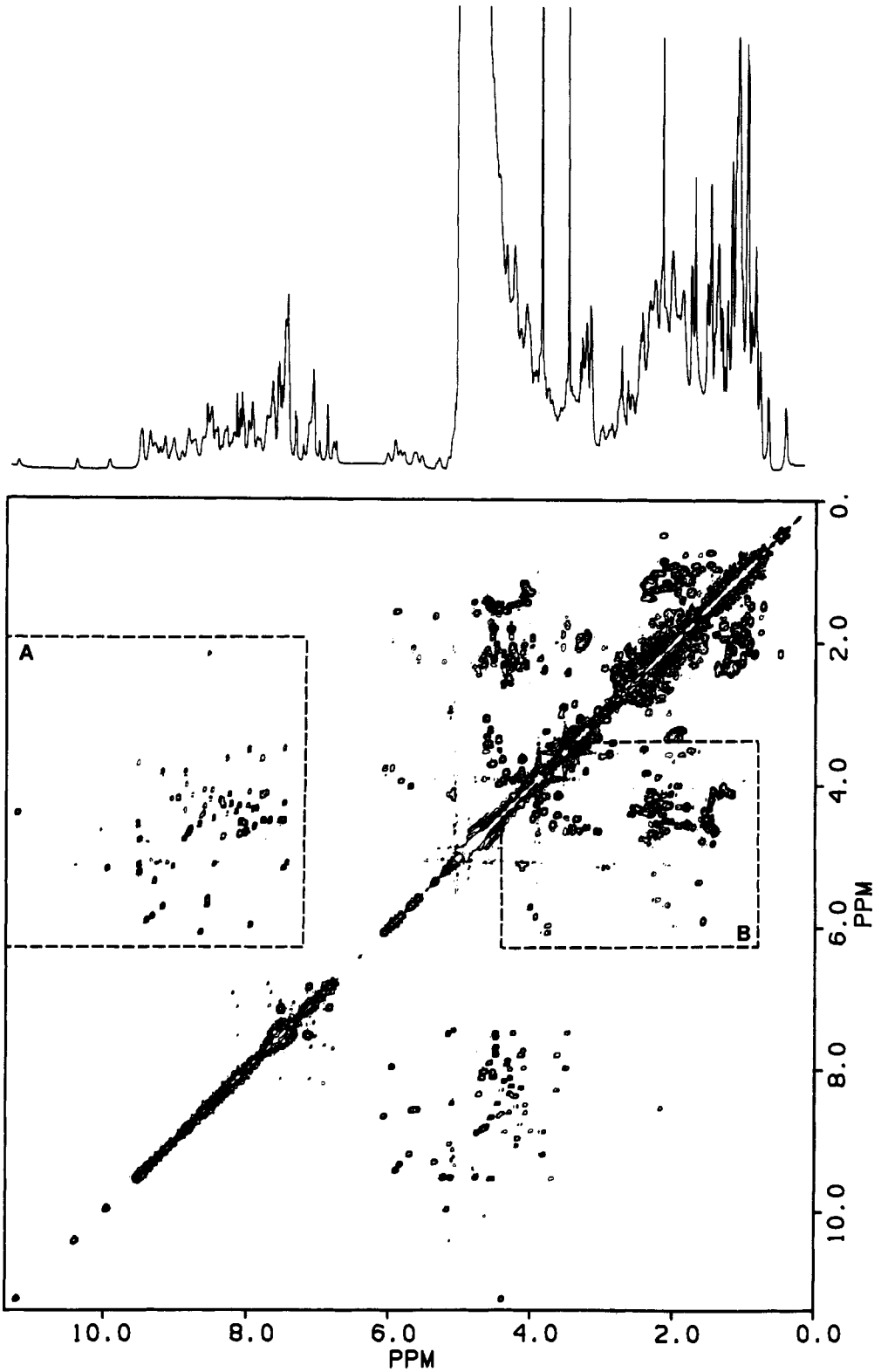

FIGURE 1: The 500-MHz COSY spectrum of HPr in H_2O. The full COSY spectrum of HPr in H_2O is shown. The regions marked "A" and "B" contain the NH–$C^{\alpha}H$ and $C^{\alpha}H$–$C^{\beta}H$ cross peaks, respectively.

phase-sensitive mode as described by States et al. (1982). NOESY spectra were obtained with a range of mixing times, from 50 to 200 ms, to monitor the effects of spin diffusion. All NOESY spectra shown in this paper were collected with a mixing time of 200 ms, where the cross peaks were the strongest. The two-dimensional J-correlated (COSY) and RELAY spectra were obtained as previously described (Weber et al., 1985). The mixing times, τ, used in the RELAY experiments were chosen to optimize the relayed coherence intensity for the spin systems of interest, according to the time dependences described by Bax and Drobny (1985). All spectra were collected over 6410 Hz, with quadrature detection into 1024 complex points, which gave a digital resolution of 6.3 Hz/point in t_2. For each t_1 value, 64 scans were signal-averaged, with a recycle time of 2.0–2.3 s. The residual HDO signal, which was placed on-resonance, was saturated during the recycle time. Normally, about 300–400 t_1 experiments were collected. Chemical shift values are referenced to sodium

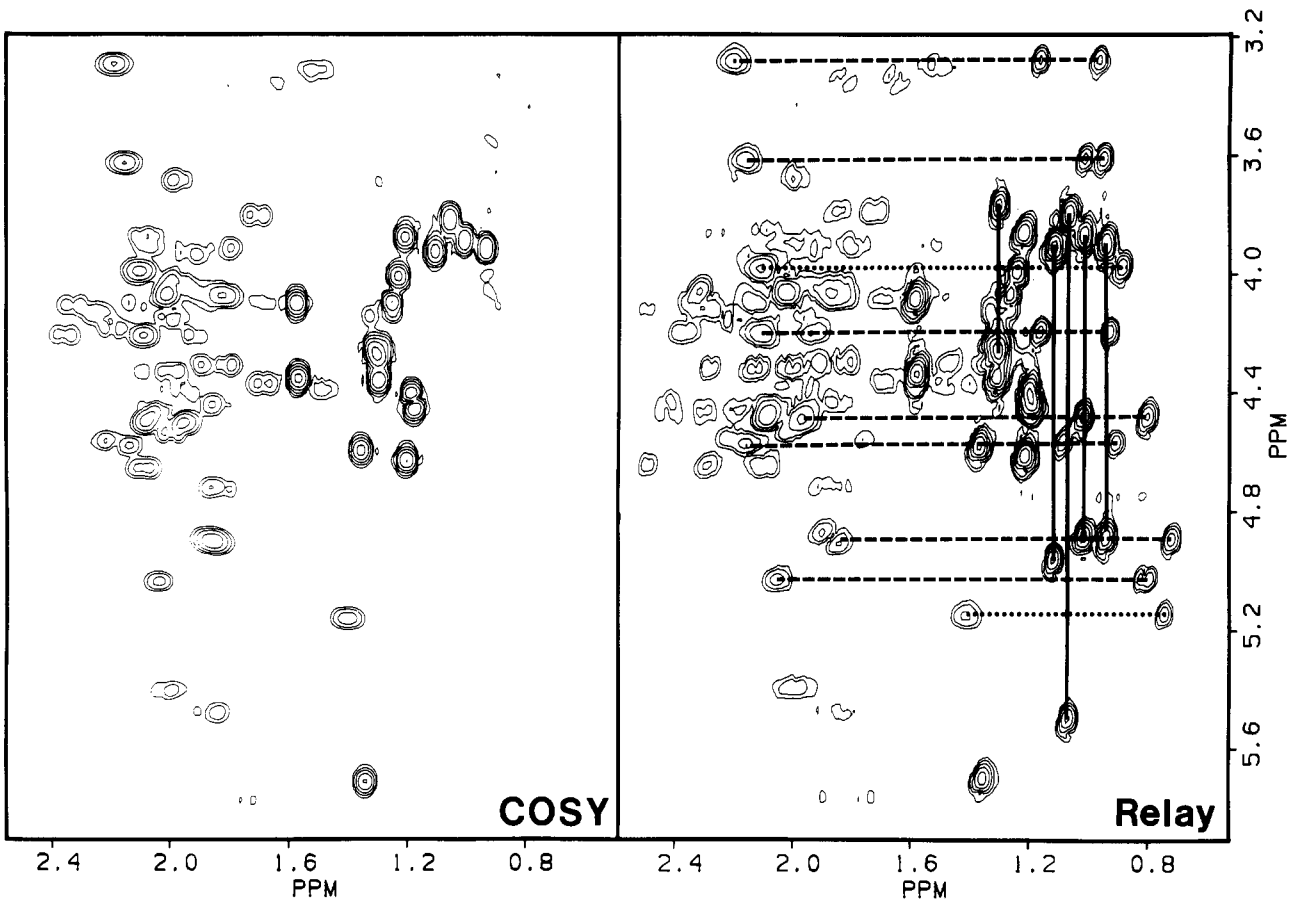

FIGURE 2: Relay experiment in D_2O. The region containing $C^{\alpha}H-C^{\gamma}H$ relay cross peaks is shown from a RELAY spectrum obtained with a mixing time, τ, of 26 ms (right). The identical region of the D_2O COSY spectrum (left) is shown for comparison. (The region corresponds to the right half of the box marked "B" in Figure 1.) The dashed lines connect $C^{\alpha}H-C^{\beta}H$ COSY cross peaks to $C^{\alpha}H-C^{\gamma}H_3$ RELAY cross peaks for valines; dotted lines connect $C^{\alpha}H-C^{\beta}H$ COSY cross peaks to $C^{\alpha}H-C^{\gamma}H_3$ RELAY cross peaks for isoleucines; solid lines connect $C^{\beta}H-C^{\gamma}H_3$ COSY cross peaks to $C^{\alpha}H-C^{\gamma}H_3$ RELAY cross peaks for threonines.

4,4-dimethyl-4-silapentane-1-sulfonate (DSS) at 0.0 ppm.

The two-dimensional spectra were copied onto magnetic tape and transferred to a VAX 11/780 to be processed with software developed by Dr. Dennis Hare. The NOESY spectra were apodized with a sine bell window function, phase shifted by 35°, and skewed toward the beginning of the free induction decay in t_2 and with a sine bell phase shifted by 35° in t_1. COSY and RELAY spectra were processed with a sine bell function in each dimension. After Fourier transformation, t_1 streaking was reduced when necessary by the subtraction method (Klevit, 1985).

RESULTS

Assigning the Spin Systems. The first task in assigning a complicated protein spectrum is to connect spins within an amino acid residue (or spin system) and relate these to a particular type of amino acid. The information necessary for this is contained in through-bond couplings. The cross peaks in a COSY spectrum correspond to spins that are coupled through two to three bonds, e.g., $NH-C^{\alpha}H$, $C^{\alpha}H-C^{\beta}H$, etc. If the COSY spectrum is completely resolved, all the spin systems in the protein can be identified from this experiment. In most cases, including the spectrum of *E. coli* HPr (Figure 1), spectral overlap leads to ambiguities in connecting the spin systems. Two-dimensional relayed coherence transfer spectroscopy (RELAY) helps to overcome the problem of spectral overlap, effectively by moving the vital connectivity information (i.e., cross peaks) out of crowded regions of the spectrum.

Bax and Drobny (1985) have shown that the intensity of a relay cross peak is dependent on both the coupling that creates the original coherence and the coupling through which it is transferred. It is therefore possible to optimize the RELAY experiment to select for a desired type of spin system on the basis of known coupling constants (Weber et al., 1985). Two types of optimized RELAY experiments were used in assigning the spin systems in the HPr spectrum: (1) a RELAY performed in D_2O to obtain $C^{\alpha}H-C^{\gamma}H$ cross peaks and (2) a RELAY performed in H_2O to obtain $NH-C^{\beta}H$ cross peaks.

The D_2O RELAY experiment in Figure 2 was optimized to give maximal relayed intensity for valine, isoleucine, and threonine $C^{\alpha}H-C^{\gamma}H_3$ spins. [Since each of these three spin systems involves the relay of coherence through a single proton to a methyl group, the time dependence of the transfer is similar (Bax & Drobny, 1985).] The corresponding region of the COSY spectrum is also shown in Figure 2, and it is clear that the RELAY experiments results in many new cross peaks. HPr contains six valines, three isoleucines, and nine threonines. The spectrum contains strong relay cross peaks for all the valines, two isoleucines, and five threonines. (The region of the spectrum shown contains the $C^{\alpha}H-C^{\beta}H$ COSY cross peaks for valines and isoleucines and the $C^{\beta}H-C^{\gamma}H_3$ COSY cross peaks for threonines. Thus, the $C^{\alpha}H-C^{\gamma}H_3$ RELAY peaks give horizontal connectivities for valines and isoleucines but vertical connectivities for threonines.) As was predicted from the dependence of relay intensity on mixing time, the relay cross peaks for these spin systems are actually stronger than the $C^{\alpha}H-C^{\beta}H$ COSY cross peaks. This was not the case for

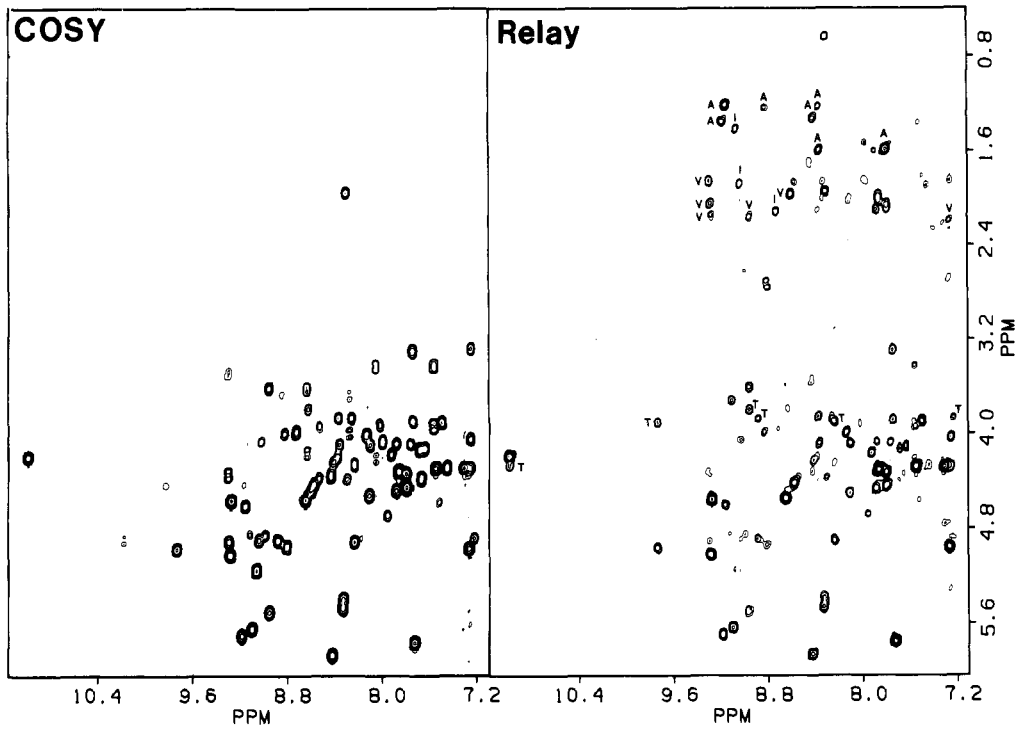

FIGURE 3: Relay experiment in H_2O. The region containing NH–$C^{\beta}H$ relay cross peaks is shown on the right. The identical region of the H_2O COSY spectrum (box A in Figure 1) is shown on the left for comparison. The cross peaks labeled T, A, V, and I were identified as being derived from threonine, alanine, valine, and isoleucine spin systems, respectively.

the optimized relay cross peaks obtained on the *cro* repressor protein (Weber et al., 1985) and is probably due to the longer effective T_2 for HPr. In addition to the spin systems for which the experiment was optimized, a number of other relay cross peaks appear in the spectrum. These are useful in identifying the $C^{\gamma}H$ resonances of protons on long side chains.[2] In the most upfield region of the RELAY spectrum there are several relay cross peaks that appear between pairs of methyl resonances, thus identifying these as belonging to leucine side chains (not shown).

An H_2O RELAY spectrum is shown in Figure 3. This experiment was optimized to give maximal relayed intensity between amide protons (A) and C^{β} protons (X) of AMX spin systems (i.e., those residues with a single C^{β} proton) according to the expression described by Bax and Drobny (1985).[3] As illustrated in Figure 4, although the actual relayed intensity observed depends on the two coupling constants $^3J_{NH-C^{\alpha}H}$ and $^3J_{C^{\alpha}H-C^{\beta}H}$, the optimal mixing time is nearly the same for all residues of this type. Comparison of the RELAY spectrum with the corresponding region of the H_2O COSY spectrum reveals 30 new cross peaks that correspond to NH–$C^{\beta}H$ relayed coherence. Since HPr contains only 19 AMX spin systems (nine threonines, seven valines, and three isoleucines), it is clear that some AMX3 and AMQX spin systems also evolved relay intensity during the optimized mixing time for AMX systems. The time dependence for relay intensity shown in Figure 4 for alanines (AMX3) shows that although the optimal mixing time for the relay is shorter than the mixing time of 52 ms used here, the maximal intensity is much higher and there will still be significant intensity remaining at the longer τ. Indeed, NH–$C^{\beta}H_3$ cross peaks were observed for

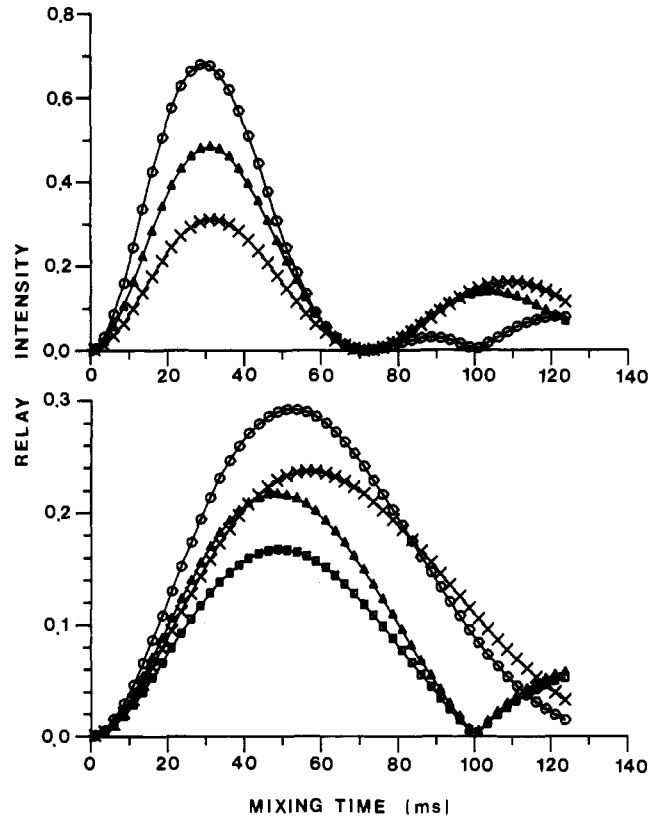

FIGURE 4: Dependence of relay intensity on mixing time. The dependence of relay intensity on mixing time (Bax & Drobny, 1985) is plotted for AMX spin systems (bottom) and AMX3 spin systems (top). The different lines in each plot were generated for different sets of coupling constants ($^3J_{NH-C^{\alpha}H}$, $^3J_{C^{\alpha}H-C^{\beta}H}$). (Bottom plot) (○) (7 Hz, 7 Hz) the random-coil values for valine; (×) (7 Hz, 5 Hz) the random-coil values for threonine; (△) (4 Hz, 10 Hz) values for α-helical residues; (□) (10 Hz, 3 Hz) values for β-sheet residues. (Top plot) (○) (10 Hz, 7 Hz) the β-sheet values for alanine; (×) (4 Hz, 7 Hz) values for α-helical alanines; (△) (6.5 Hz, 7 Hz) random-coil values for alanine (Bundi & Wüthrich, 1979).

[2] In the context of these NMR studies, glutamic acid, glutamine, lysine, leucine, methionine, and arginine are considered long side chains.

[3] The spin system notation used in this paper is the same as that used by Bax and Drobny (1985). In particular, the spin closest to the backbone is named "A". Thus, in D_2O the $C^{\alpha}H$ is spin A and in H_2O the NH is spin A.

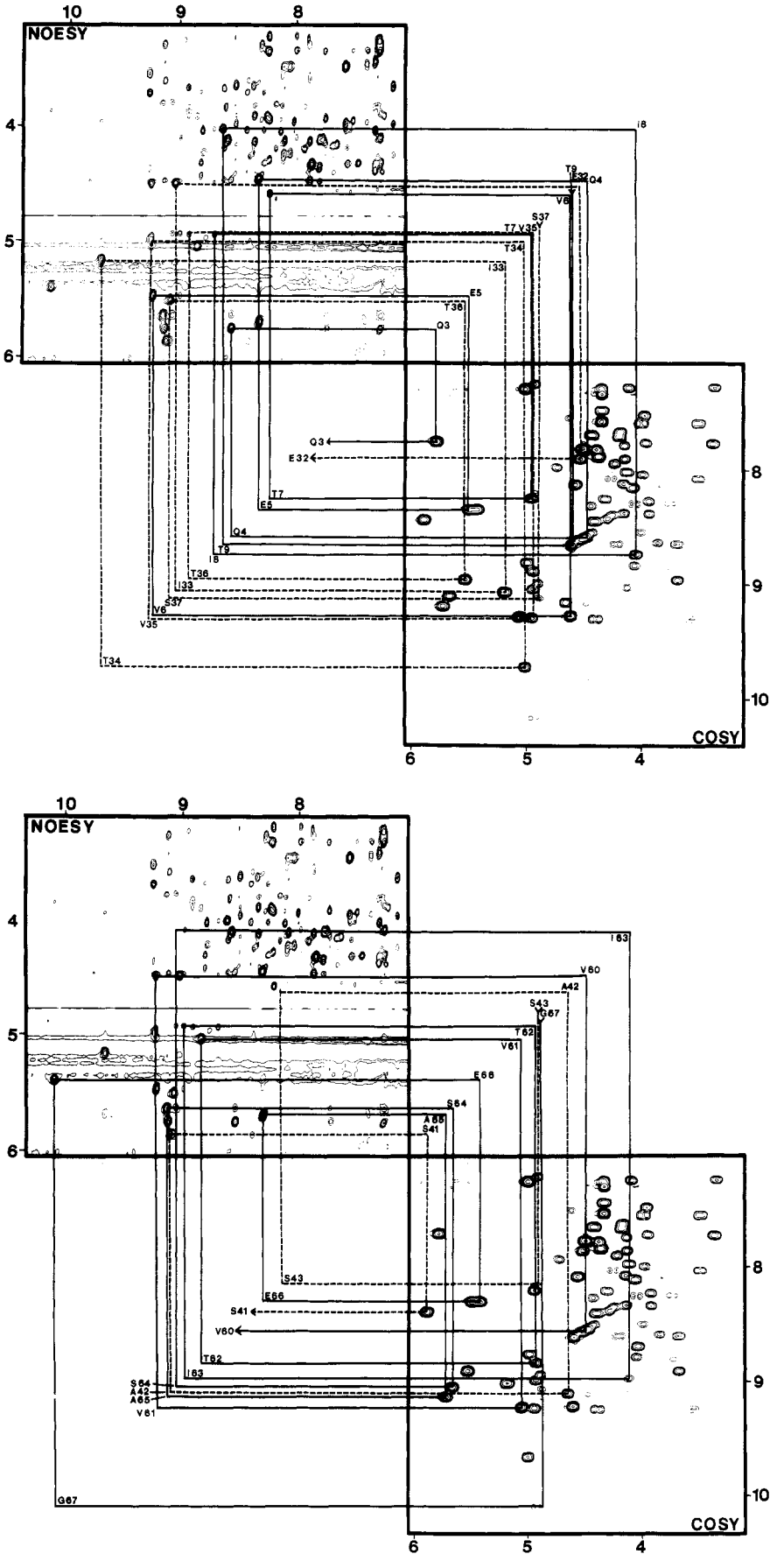

FIGURE 5: $d_{\alpha N}$ connectivities for HPr. $d_{\alpha N}$ connectivities begin at the $C^{\alpha}H$ of the C-terminal residue of a segment (arrowheads in the upper quadrant) and are traced to the N-terminal residue of the segment (arrowheads in the lower left quadrant). The positions labeled along the diagonal in the upper right quadrant are the $C^{\alpha}H$ resonance positions, and the positions labeled in the lower left quadrant are the NH resonance positions. (Top) $d_{\alpha N}$ connectivities from Thr-9 through Gln-3 and from Ser-37 through Glu-32 are followed. (Bottom) $d_{\alpha N}$ connectivities from Ser-43 through Ser-41 and from Gly-67 through Val-60 are followed.

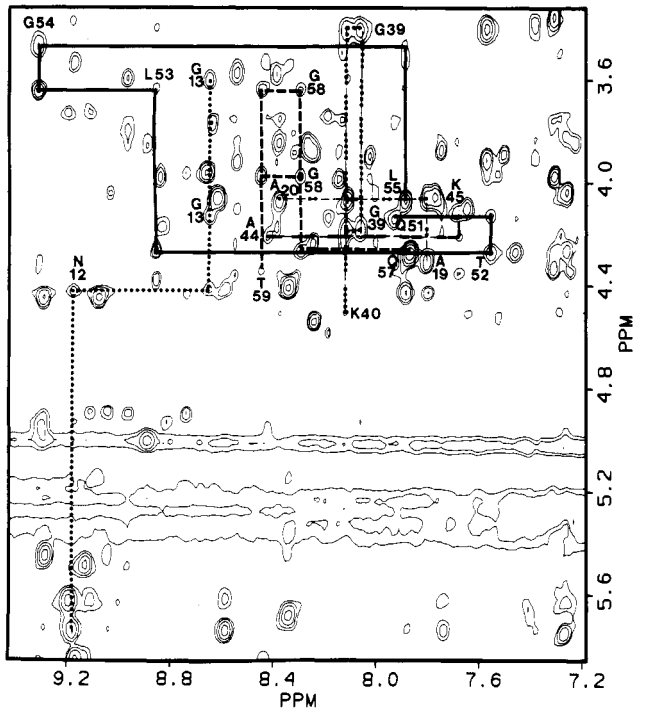

FIGURE 6: $d_{\alpha N}$ connectivities for HPr. The second type of $d_{\alpha N}$ connectivities are traced in the H_2O NOESY in this figure. The spectral region shown is the same region as box A in Figure 1, in a NOESY spectrum. Labeled cross peaks correspond to the $d^i_{\alpha N}$ NOEs. The solid line follows the longest stretch of connectivities, from Leu-55 to Gln-51. The long dashed line follows the segment from Thr-59 to Gln-57. Short dashed lines indicate the single dipeptide connectivities that are observed.

seven of the nine alanines in HPr. As predicted, in an H_2O RELAY experiment performed with a shorter mixing time, the relative intensities for the alanine relays increased while those of the other spin systems decreased (results not shown). These 30 relay connectivities, used in conjunction with the COSY spectra and D_2O RELAY spectra, were used to identify a majority of the spin systems in the protein spectrum.

Sequential Assignments. The sequential assignment strategy for proteins takes advantage of the fact that the backbone amide proton of a given residue i will always be in close proximity to at least one of the following protons of its nearest neighbor residue, $i-1$: NH (d_{NN}), C$^\alpha$H ($d_{\alpha N}$), or C$^\beta$H ($d_{\beta N}$)[4] (Wüthrich et al., 1982; Billeter et al., 1982). The assignments were therefore made from NOESY spectra obtained from H_2O solutions where the exchangeable amide proton resonances were observed. Since there are three different possible nearest-neighbor connectivities for a given dipeptide step, several regions of the NOESY spectrum must be analyzed. These are described separately below.

(*i*) $d_{\alpha N}$ *Connectivities.* The regions of the two-dimensional spectra used to establish the $d_{\alpha N}$ connectivities in the HPr spectrum are shown in Figures 5 and 6. In addition to the type of $d_{\alpha N}$ connectivities that have been observed and described for other proteins assigned by 2DNMR techniques thus far [for example, bovine pancreatic trypsin inhibitor (Würthrich et al., 1982) and *cro* repressor protein (Weber et al., 1985)], another type of $d_{\alpha N}$ connectivity was observed in the HPr spectrum. The two types of behavior can be generally described as (1) those that consisted of strong $d_{\alpha N}$ cross peaks and weak or no $d^i_{\alpha N}$ cross peaks (Figure 5), indicative of β-strand structure, and (2) those that consisted of both $d_{\alpha N}$ and $d^i_{\alpha N}$ cross peaks of similar intensities (Figure 6). Four segments of the first type of $d_{\alpha N}$ connectivities were observed, as followed in Figure 5 and six segments of the second type of $d_{\alpha N}$ connectivities were observed, as followed in Figure 6. In each case, as the segments were followed, the identity of NH–C$^\alpha$H peaks that had been unambiguously established by the COSY/RELAY spectra was noted so that the connected segment could be matched to the primary sequence of HPr (Weigel et al., 1982; Powers & Roseman, 1984; De Ruese et al., 1985). In cases where the identity of the COSY cross peak could not be uniquely determined, the side-chain type (i.e., AMQX, etc.) was noted.

The first stretch of $d_{\alpha N}$ connectivities in the top plot of Figure 5 is seven residues long and ends with the sequence V-T-I-T, which identifies the segment as Gln-3 through Thr-9 (see the sequence in Figure 9). The second stretch in the top plot is six residues long and contains the sequence I-T-V-T. This segment was assigned to Glu-32 through Ser-37. The first stretch in the bottom plot is only three residues long, with an alanine in the middle, surrounded by two AMQX side chains and was assigned to Ser-41 through Ser-43. The longer segment traced in the bottom plot is eight residues long, beginning with the sequence V-V-T-I, and was assigned to Val-60 through Gly-67. Although for the sake of clarity this segment has been drawn as a continuous connectivity path in the figure, the nearest-neighbor $d_{\alpha N}$ NOE cross peak for Ile-63 to Ser-64 is missing. As the predicted location for this cross peak was not obscured by the solvent line or any other resonance, the absence of this expected cross peak must be due to an interruption in the regular β-strand structure.

The second type of $d_{\alpha N}$ connectivities are traced in Figure 6, with the $d^i_{\alpha N}$ cross peaks, i.e., those that coincide exactly with NH–C$^\alpha$H COSY cross peaks, labeled. The longest stretch of this type of $d_{\alpha N}$ connectivities in HPr is five residues long, with a threonine in the second position and a glycine in the fourth position, and was assigned to Gln-51–Leu-55. A shorter stretch containing a glycylthreonine dipeptide was found and was assigned to Gln-57–Thr-59. Thus, there is a nine-residue segment starting with Gln-51 and ending with Thr-59 that gives $d_{\alpha N}$ connectivities, with a break at Thr-56. At the end of this segment, another break occurs between Thr-59 and the Val-60–Gly-67 segment assigned by the β-strand-type connectivities above. A three-residue segment that contains a glycine at its C-terminus and an AMQX residue in the middle was also found. The N-terminal C$^\alpha$H resonance in this tripeptide overlaps exactly with that of Gln-3, which was already assigned at the N-terminal end of a β-strand connectivity (above). The COSY spectrum only showed a single NH–C$^\alpha$H cross peak at this resonance position, but the one-dimensional spectrum confirmed that two protons resonate at this frequency (the second peak from the left in the group of C$^\alpha$H resonances that are directly downfield of the solvent signal in the spectrum in Figure 1). These observations indicated that the C$^\alpha$H resonance was derived from a proline residue, and since there is a tripeptide sequence Pro-Asn-Gly in HPr, the $d_{\alpha N}$ connectivity was assigned to Pro-11–Gly-13.

In addition to the long stretches of $d_{\alpha N}$ connectivities, there were four single $d_{\alpha N}$ connections, corresponding to dipeptide segments, in the spectrum (Figure 6): (1) a glycine–long side-chain pair, (2) an alanine–long side-chain pair, (3) a dipeptide of two long side chains, and (4) an alanine to a C$^\alpha$H resonance with no corresponding amide resonance. There are six glycines in HPr, only three of which have long side-chain

[4] To avoid ambiguities in the type of NOEs being described, the following nomenclature has been adopted for this paper. The three distances used in the sequential assignment strategy, explicitly, $d_{\alpha_{i-1}N_i}$, $d_{N_{i-1}N_i}$, and $d_{\beta_{i-1}N_i}$, are written in the short-hand form suggested by Wüthrich et al. (1982), $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$, respectively. NOEs between protons within a single amino acid residue are denoted as $d^i_{\alpha N}$, $d^i_{\beta N}$, etc.

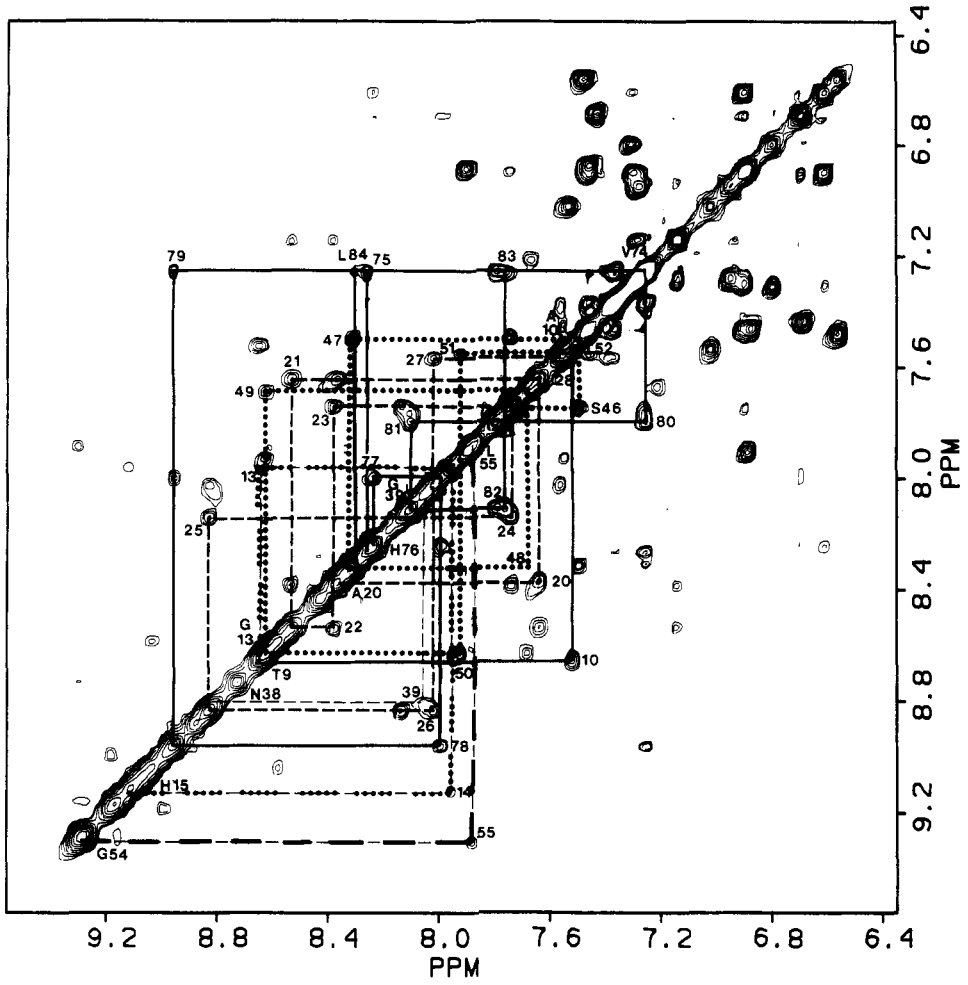

FIGURE 7: d_{NN} connectivities for HPr. The region of the H_2O NOESY containing d_{NN} cross peaks is shown. The ends of each segment are labeled on the diagonal. To follow the connectivities, begin at the C-terminal end of a segment (i). The first cross peak encountered is labeled $i-1$ and gives the resonance position for NH_{i-1} when followed to the diagonal. The next cross peak identifies $i-2$ and so on until the N-terminal resonance in the segment is reached. Three long stretches of d_{NN} connectivities are followed from Gly-28 through Ala-20, from Thr-52 through Ser-46, and from Leu-84 through Val-74. Shorter stretches and isolated dipeptide connectivities are indicated. The unlabeled cross peaks (mainly in the downfield region) are nonsequential d_{NN} cross peaks between residues across β-sheets.

residues after them, Gly-13, Gly-39, and Gly-67. Since both Gly-67 and Gly-13 had already been assigned as part of other $d_{\alpha N}$ segments, the glycyl dipeptide described above must contain Gly-39, thereby assigning Lys-40 as well. There are nine alanines in HPr, five of which are followed by long side-chain residues, Ala-20, Ala-26, Ala-44, Ala-65, and Ala-82. Ala-65 was assigned in the $d_{\alpha N}$ segment Val-60–Gly-67. As described below, Ala-20, Ala-26, and Ala-82 appear in long d_{NN} stretches, thus leaving Ala-44–Lys-45 as the second $d_{\alpha N}$ dipeptide. The third dipeptide was assigned on the basis of d_{NN} connectivities described below, and the fourth dipeptide was assigned to Pro-18-Ala-19. In addition, a $d_{\alpha N}$ connectivity was found from Gly-67 to an amide that did not give a NH–C^{α}H COSY cross peak. The amide did give an NH–C^{α}H NOESY cross peak at a chemical shift where an extra $C^{\alpha}H$–$C^{\beta}H$ COSY peak had been found. The amide was assigned to Glu-68, thus extending the long $d_{\alpha N}$ segment that began with Val-61 by one residue. The $d_{\alpha N}$ connectivities accounted for 39 of the 85 residues in HPr.

(*ii*) d_{NN} *Connectivities*. The region of the H_2O NOESY spectrum that contains d_{NN} cross peaks is shown in Figure 7. Except at the ends of segments, an NH_i resonance showing d_{NN} behavior will have two NH–NH cross peaks associated with it, to NH_{i-1} and to NH_{i+1}. Therefore, unlike $d_{\alpha N}$ connectivities, the polarity of the connection cannot be determined by the direction of the connectivities. The identity of spin systems at several steps along the d_{NN} segment however, allows for unambiguous sequential assignment of the segment. There were three long stretches of d_{NN} connectivities in the HPr spectrum. The first was nine residues long and had an alanine at one end and a glycine at the other. With the additional information that there is a phenylalanine two residues away from the end alanine and an alanine two residues away from the end glycine, this segment was assigned to the sequence Ala-20 through Gly-28. The next segment was 11 residues long and contained a histidine with a valine residue two removed. This segment was assigned to Val-74 through Leu-84. (The Glu-75-His-76 cross peak was not observed because it was too close to the diagonal.) The third stretch of d_{NN} connectivities was seven residues long and had a threonine at one end and an AMQX side chain at the other end. The remaining residues in this segment were long-chain residues that were difficult to identify, but with the additional observation that the sequence in question did contain a glycine, it was unequivocally assigned to Ser-46 through Thr-52. These three d_{NN} segments resulted in the assignment of 27 residues of HPr.

Inspection of Figure 7 reveals that these three segments accounted for nearly all of the strong NH–NH cross peaks. However, as was the case with the $d_{\alpha N}$ connectivities, there were also some isolated d_{NN} cross peaks that connected di- and tripeptide units. The strongest of these connected Thr-9 (the

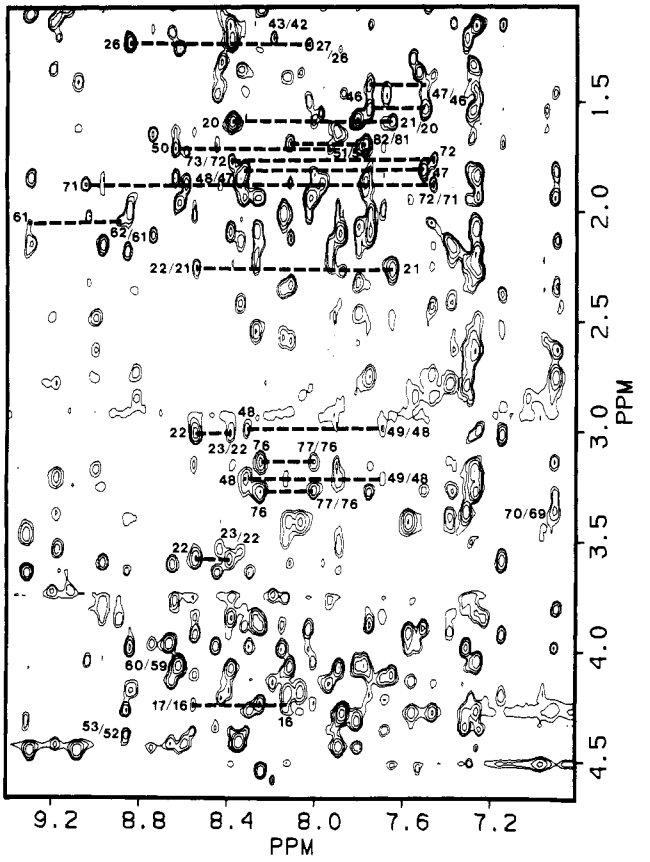

FIGURE 8: $d_{\beta N}$ connectivities in HPr. The region of the H_2O NOESY spectrum containing $d_{\beta N}$ cross peaks is shown (essentially box A in Figure 1 extended upfield). The dashed lines connect $d^{i}_{\beta N}$ cross peaks, which are labeled with a single number, $i-1$, to $d_{\beta N}$ cross peaks, which are labeled with two numbers, $i/i-1$. Those $C^{\beta}H$ resonances that do not give $d^{i}_{\beta N}$ cross peaks are also labeled.

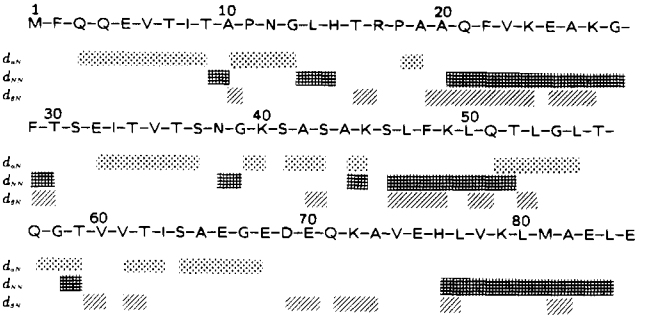

FIGURE 9: Summary of sequential connectivities observed for HPr. The three types of nearest-neighbor connectivities used for the sequential assignments are summarized. ($d_{\alpha\delta}$ and $d_{\beta\delta}$ connectivities involving proline residues are shown as $d_{\alpha N}$ and $d_{\beta N}$ connectivities).

last residue in a long $d_{\alpha N}$ stretch) to Ala-10. Two weak d_{NN} peaks appeared for a long side-chain residue, one of which connected to Gly-13, which had been assigned by a $d_{\alpha N}$ connectivity to Asn-12, and the other of which connected to a residue that gave no observable COSY NH–$C^{\alpha}H$ cross peak. Thus, the d_{NN} cross peak that involved Gly-13 identified Leu-14, and the second d_{NN} cross peak than connected Leu-14 to the amide of His-15. The $C^{\beta}H$ resonances of His-15 were found by NOEs to the imidazole C4H, and the $C^{\alpha}H$ resonance position could then be identified in the COSY spectrum. The $C^{\alpha}H$ resonates very near the H_2O resonance position and was therefore bleached out of the H_2O spectra by the solvent-suppression pulse. A weak cross peak appeared between two resonances that also gave a strong $d_{\alpha N}$ peak (Figure 6), Gly-

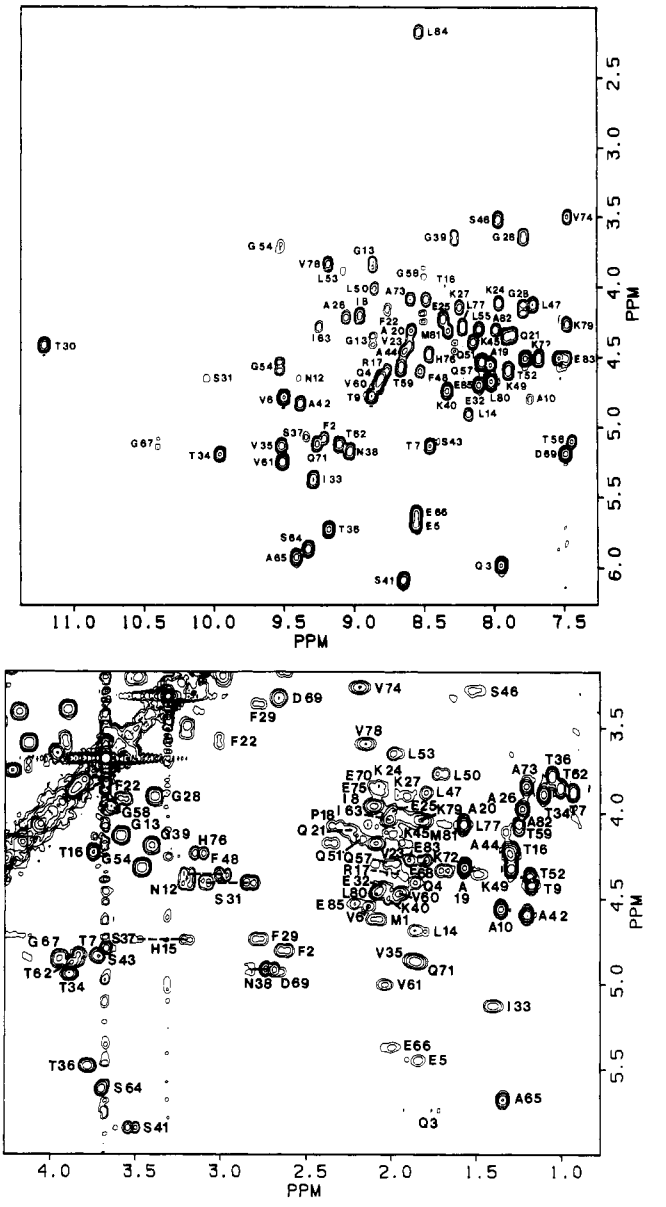

FIGURE 10: Assignments in the spectrum. (Top) The fingerprint region of the H_2O COSY spectrum (box A in Figure 1) with the assignments of the NH–$C^{\alpha}H$ cross peaks is shown. (Bottom) The $C^{\alpha}H$–$C^{\beta}H$ region of the COSY spectrum (box B in Figure 1) with the results of sequential assignments is shown.

58-Thr-59. For the sake of resolution in the published figure, the three most downfield-shifted amide resonances in the HPr spectrum have not been included in Figure 7. There was, however, one additional strong d_{NN} cross peak that occurred between the most downfield NH resonance [identified as a threonine from the RELAY experiment (Figure 3)] and an NH resonance at 9.18 ppm. The remaining weak cross peaks in the d_{NN} plot connected nonsequential NH resonances that had been previously assigned by their $d_{\alpha N}$ connectivities in the section above and were due to cross-strand β-sheet NOEs.

(*iii*) $d_{\beta N}$ *Connectivities.* The third type of nearest-neighbor NOEs, $d_{\beta N}$, are from the NH of a residue to the $C^{\beta}Hs$ of the preceding residue. These will occur for α-helical residues and thus should be observed for resonances already assigned by their d_{NN} connectivities. In such cases, $d^{i}_{\beta N}$ cross peaks will also be observed. $d_{\beta N}$ connectivities may also occur in certain types of turns and in regions that lack regular secondary structure. As illustrated in Figure 8 and summarized in Figure 9, $d_{\beta N}$ cross peaks were indeed observed for many of the

Table I: Sequence-Specific Resonance Assignments of the *E. coli* HPr ^{1}H NMR Spectrum[a]

residue	NH	$C^{\alpha}H$	$C^{\beta}H$	residue	NH	$C^{\alpha}H$	$C^{\beta}H$
Met-1	n.o.	4.60	2.09	Ala-44	8.42	4.23	1.30
Phe-2	8.99	4.85	3.17, 2.62	Lys-45	7.68	4.12	1.98
Gln-3	7.72	5.75	1.86, 1.71	Ser-46	7.74	3.28	1.50, 1.42
Gln-4	8.57	4.42	1.84	Leu-47	7.49	3.88	1.79
Glu-5	8.33	5.45	1.83	Phe-48	8.30	4.37	3.20, 2.98
Val-6	9.27	4.56	2.12	Lys-49	7.68	4.34	1.47
Thr-7	8.23	4.90	3.84	Leu-50	8.62	3.77	1.84, 1.70
Ile-8	8.73	3.97	2.10	Gln-51	7.92	4.16	2.16
Thr-9	8.64	4.55	4.41	Thr-52	7.55	4.28	4.35
Ala-10	7.51	4.57	1.34	Leu-53	8.85	3.66	1.99
Pro-11		5.75		Gly-54	9.30	4.34, 3.48	
Asn-12	9.16	4.42	3.50, 3.20	Leu-55	7.88	4.07	2.09
Gly-13	8.64	4.15, 3.61		Thr-56	7.22	4.87	3.86
Leu-14	7.95	4.67	1.85	Gln-57	7.86	4.30	1.98
His-15	9.14	4.72	3.47, 3.18	Gly-58	8.28	3.98, 3.66	
Thr-16	8.12	3.76	4.22	Thr-59	8.43	4.34	4.05
Arg-17	8.53	4.36	2.00, 1.95	Val-60	8.60	4.49	1.94
Pro-18		4.06	2.46, 2.32	Val-61	9.21	5.01	2.02
Ala-19	7.79	4.32	1.56	Thr-62	8.88	4.88	3.83
Ala-20	8.36	4.08	1.57	Ile-63	9.02	4.05	2.01
Gln-21	7.64	4.11	2.26	Ser-64	9.10	5.63	3.69, 3.69
Phe-22	8.53	3.92	3.58, 3.01	Ala-65	9.18	5.69	1.33
Val-23	8.38	4.18	2.09	Glu-66	8.33	5.39	1.98, 1.92
Lys-24	7.74	3.88	2.11	Gly-67	10.17	4.88, 3.98	
Glu-25	8.13	3.99	1.99	Glu-68	7.31	4.28	1.80
Ala-26	8.83	3.98	1.22	Asp-69	7.26	4.96	3.32, 2.66
Lys-27	8.02	3.91	1.97	Glu-70	6.90	3.81	2.33
Gly-28	7.56	3.92, 3.41		Gln-71	9.04	4.89	1.84
Phe-29	9.17	4.73	3.35, 2.78	Lys-72	7.45	4.28	1.80
Thr-30	10.99	4.18	4.24	Ala-73	8.37	3.85	1.20
Ser-31	9.82	4.42	3.09, 2.83	Val-74	7.25	3.27	2.18
Glu-32	7.88	4.47	2.05	Glu-75	8.26	3.85	1.84
Ile-33	9.06	5.14	1.40	His-76	8.24	4.24	3.27, 3.12
Thr-34	9.73	4.97	3.87	Leu-77	8.00	4.06	1.52
Val-35	9.29	4.90	1.85	Val-78	8.96	3.60	2.13
Thr-36	8.95	5.50	3.77	Lys-79	7.25	4.03	1.84
Ser-37	9.11	4.77	3.65	Leu-80	7.79	4.44	2.07
Asn-38	8.80	4.93	2.86, 2.71	Met-81	8.10	4.08	1.68
Gly-39	8.05	4.19, 3.41		Ala-82	7.76	4.07	1.24
Lys-40	8.11	4.51	1.98	Glu-83	7.25	4.28	1.85
Ser-41	8.42	5.85	3.63, 3.52	Leu-84	8.31	1.95	0.62
Ala-42	9.16	4.59	1.19	Glu-85	8.05	4.51	2.20
Ser-43	8.17	4.87	4.13, 3.71				

[a] ppm relative to DSS ± 0.03 ppm. Values are for HPr at 30 °C, pH 6.5.

residues that were found in long d_{NN} stretches. In addition, $d_{\beta N}$ cross peaks were observed for resonances that had not been assigned by either of the other two sequential connectivities. A $d_{\beta N}$ cross peak was observed from a threonine to an unassigned long side chain. Of the 10 threonines in HPr, Thr-16, Thr-30, and Thr-56 could not be assigned by $d_{\alpha N}$ or d_{NN}, but only Thr-16 appears next to an unassigned long side-chain residue. Therefore, the peak was due to an NOE between Thr-16 and Arg-17. The most downfield NH resonance (a threonine) gave $d_{\beta N}$ cross peaks to a pair of β-protons that also gave cross peaks to a phenylalanine ring resonance. (The downfield threonine NH resonance has been omitted from Figure 8 to give a clearer view of this crowded region of the NOESY spectrum.) The NH resonance at 9.18 ppm, which gave a d_{NN} cross peak to this threonine, also gave NOEs to these β-protons. In a D_2O COSY spectrum, these $C^{\beta}H_2$ resonances were found to be coupled to a $C^{\alpha}H$ resonance at 4.75 ppm. In the H_2O COSY spectrum, there was no NH–$C^{\alpha}H$ cross peak at 4.75 ppm due to bleaching of this region by the solvent-suppresion pulse. On the basis of the connectivities described above, it was possible to identify the position where the peak should appear. Of the threonines not previously assigned, only Thr-30 is preceded by a phenylalanine residue, making the invisible NH resonance that of Phe-29. A strong $d_{\beta N}$ cross peak was observed between Thr-59 and Val-60. As shown in Figure 9, Thr-59 was assigned at the end of a $d_{\alpha N}$ stretch, and Val-60 was assigned at the beginning of a $d_{\alpha N}$ stretch, but a break occurs between the two residues. The $d_{\beta N}$ connectivity fills in the gap and suggests that a turn configuration occurs at this point.

A six-residue segment of HPr, Asp-69–Val-74, did not give any $d_{\alpha N}$ or d_{NN} cross peaks, as shown in Figure 9. Ala-73 could be assigned by the process of elimination as the last remaining unassigned alanine. Ala-73 was used to search for possible sequential connectivities among the remaining unassigned resonances. As shown in Figure 8, $d_{\beta N}$ connectivities could be traced from Ala-73 through Asp-69.

The sequential connectivities for *E. coli* HPr are summarized in Figure 9. One phenylalanine residue, Phe-2, did not give a sequential connectivity through its NH, $C^{\alpha}H$, or $C^{\beta}Hs$. However, a strong NOE was observed from a phenylalanyl ring proton to the $C^{\alpha}H$ of Gln-3. The same ring proton also gave cross peaks to a pair of $C^{\beta}H$ resonances that had not been previously assigned and that were J-coupled to a $C^{\alpha}H$ that was in turn J-coupled to an amide that also had not been previously assigned. On the basis of these observations, the resonances described were assigned to Phe-2. The N-terminal residue of the protein, Met-1, did not have an observable amide resonance presumably because the protons exchange too rapidly with solvent protons. The $C^{\alpha}H$–$C^{\beta}H$ COSY cross peak for Met-1 was identified on the basis of it not having a corresponding NH–$C^{\alpha}H$ COSY cross peak, and this assignment was later

confirmed on the basis of cross-strand $C^{\alpha}H-C^{\alpha}H$ β-sheet NOEs (Klevit & Waygood, 1986).

At this stage, only Glu-85 remained unassigned. In the $C^{\alpha}H-C^{\beta}H$ region of the COSY, there was a cross peak that had not yet been assigned, and following the resonance position of the $C^{\alpha}H$ back into the $NH-C^{\alpha}H$ region pointed to a very weak cross peak. An amide at this position did give an NOE to a side-chain proton of Leu-84, consistent with it being Glu-85. Thus, all of the 85 residues of *E. coli* HPr were assigned. The $NH-C^{\alpha}H$ assignments and the $C^{\alpha}H-C^{\beta}H$ assignments are in the COSY spectrum shown in Figure 10 and are summarized in Table I.

Discussion

Sequence-specific resonance assignments are necessary if one wishes to use NMR spectroscopy to determine the three-dimensional structure of a protein in solution. By use of the sequential assignment strategy (Wüthrich et al., 1982; Billeter et al., 1982) together with optimized RELAY spectroscopy, sequence-specific assignments for all 85 residues of *E. coli* HPr are complete as summarized in Figures 9 and 10. Although the sequential connectivity method reveals information primarily about the backbone proton resonances of the protein spectrum, many side-chain resonances were assigned in the process as well. The NOEs among side-chain protons are crucial for determining the tertiary fold of a protein and are discussed in more detail in the third paper in this series (Klevit & Waygood, 1986).

HPr, which has 85 residues, is the largest protein for which complete sequential assignments have been made. The spectrum of *E. coli* HPr suffers from the spectral overlap that is inevitable with the increasing size of macromolecules being studied by ^1H 2DNMR. The usefulness of RELAY spectroscopy in resolving the ambiguities caused by overlap cannot be over-emphasized. It is clear from the RELAY spectra of HPr that when the experiments are performed under optimized conditions, large numbers of strong relay cross peaks appear in the spectrum, even in H_2O solutions. If we are to study larger proteins by 2DNMR, optimized RELAY spectroscopy must become as routine as experiment as the COSY and NOESY experiments now are.

This is the first in a series of papers describing our studies of *E. coli* HPr by ^1H NMR. The second paper will describe the application of long-range coherence transfer experiments to obtain side-chain assignments. The third paper will describe the use of short- and medium-range NOEs to determine the secondary structure and long-range NOEs to determine the tertiary fold of the protein in solution. The spectral properties of the phosphoprotein are currently under investigation.

Acknowledgments

We thank Prof. Brian Reid for his generousity in providing NMR instrumentation and laboratory facilities as well as for his enthusiasm and encouragement for this project. We also thank Drs. David Wemmer and Paul Weber for helpful discussions and suggestions, Jon Herriott for his critical comments on the manuscript, Clem Furlong for making the initial contacts that made this work possible, and Ossama El Kabbani and Louis Delbaere for their continuing interest in this project.

References

Anderson, B., Weigel, N., Kundig, W., & Roseman, S. (1971) *J. Biol. Chem. 246*, 7023–7033.

Bax, A., & Drobny, G. (1985) *J. Magn. Reson. 61*, 306–320.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285–297.

Delbaere, L. T. J., Bruse, L. M., & Waygood, E. B. (1982) *J. Mol. Biol. 157*, 161–162.

DeRuese, H., Roy, A., & Danchin, A. (1985) *Gene 35*, 199–207.

Kalbitzer, H. R., Hengstenberg, W., Rosch, P., Muss, P., Bernsmann, P., Engelmann, R., Dorschug, M., & Deutscher, J. (1982) *Biochemistry 21*, 2879–2885.

Klevit, R. E. (1985) *J. Magn. Reson. 62*, 551–555.

Klevit, R. E., & Waygood, E. B. (1986) *Biochemistry* (third paper of three in this issue).

Postma, P. W., & Lengeler, J. (1985) *Microbiol. Rev. 49*, 232–269.

Powers, D. A., & Roseman, S. (1984) *J. Biol. Chem. 259*, 15212–15214.

Simoni, R. D., Nakazawa, T., Hays, J. B., & Roseman, S. (1973) *J. Biol. Chem. 248*, 932–940.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.

Waygood, E. B., & Steeves, T. (1980) *Can. J. Biochem. 58*, 40–48.

Waygood, E. B., Erickson, E., El Kabbani, O. A. L., & Delbaere, L. T. J. (1985) *Biochemistry 24*, 6938–6945.

Weber, P. W., Drobny, G., & Reid. B. R. (1985) *Biochemistry 24*, 4549–4552.

Weigel, N., Powers, D. A., & Roseman, S. (1982) *J. Biol. Chem. 257*, 14499–14509.

Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 115*, 311–319.

Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.