Two-Dimensional ^{1}H NMR Studies of Histidine-Containing Protein from *Escherichia coli*. 3. Secondary and Tertiary Structure As Determined by NMR†

Rachel E. Klevit*

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

E. Bruce Waygood

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

Received June 5, 1986; Revised Manuscript Received August 12, 1986

ABSTRACT: Sequence-specific resonance assignments of the ^{1}H NMR spectrum of the 85-residue histidine-containing phosphocarrier protein (HPr) are complete [Klevit, R. E., Drobny, G. P., & Waygood, E. B. (1986) *Biochemistry* (first paper of three in this issue)]. Additional side-chain assignments have been made with long-range coherence transfer experiments [Klevit, R. E., & Drobny, G. P. (1986) *Biochemistry* (second paper of three in this issue)]. In this paper, the NMR assignments were used to determine the secondary structure and the tertiary folding of HPr in solution. The secondary structural elements of the protein were determined by visual inspection of the pattern of nearest-neighbor nuclear Overhauser effects (NOEs) and the presence of persistent amide resonances. *Escherichia coli* HPr consists of four β-strands, three α-helices, four reverse turns, and several regions of extended backbone structure. Long-range NOEs, especially among side-chain protons, were used to determine the tertiary structure of the protein by use of the secondary structural components. The four β-strands form a single antiparallel β-pleated sheet. The hydrophobic faces of the α-helices interact to form a hydrophobic core and sit above the hydrophobic face of the β-sheet, forming an open-face β-sheet sandwich structure. The active site histidine, His-15, is on a short kinked segment of backbone that is accessible to the solvent. The positively charged phosphorylation site (His-15 and Arg-17) interacts with the negatively charged carboxyl terminus of the protein (Glu-85). The resulting model of HPr was obtained without the use of a structure derived from X-ray diffraction studies and is the largest obtained thus far by two-dimensional NMR techniques.

The transport of sugars by the bacterial phosphoenolpyruvate:sugar phosphotransferase system (PTS)[1] occurs via a series of phosphoryl group transfer steps, all of which involve phosphoprotein intermediates. The PTS protein known as HPr accepts a phosphoryl group from enzyme I at the N-1 position of the imidazole ring of histidine residue 15. Phospho-HPr in turn donates the phosphoryl group to the N-3 position of a histidine residue of either factor III^{sugar} or enzyme II^{sugar} [for a review, see Postma and Lengeler (1985)].

At present, little is known about the structure of PTS proteins. As part of a program aimed at understanding the mechanism of phosphotransfer, we have used two-dimensional ^{1}H NMR (2DNMR) techniques to determine the structure of *Escherichia coli* HPr in solution. In the first paper in this series (Klevit et al., 1986), sequence-specific resonance assignments were obtained for the amide, C$^\alpha$, and C$^\beta$ protons of the 85 residues of HPr. Additional side-chain assignments were obtained by long-range coherence transfer spectroscopy (Klevit & Drobny, 1986). In this paper, we extend the sequence-specific assignments to other side-chain resonances. The secondary structure, as determined by short-range NOEs, and the tertiary structure, as determined by medium- and long-range NOEs, are reported. The functional ramifications of the structure will be discussed.

Materials and Methods

The sample preparation and spectroscopic methods used to obtain the spectra shown in this paper have been described in detail (Klevit et al., 1986).

Results

Assignment of Aromatic Resonances. E. coli HPr contains six aromatic residues: four phenylalanines and two histidines. Aromatic spin systems can be traced according to their through-bond connectivities in COSY and RELAY spectra. The four-bond couplings between the C2 and C4 protons of both histidines were observed in the D_2O COSY spectrum shown in Figure 1. The aromatic connectivities for the four phenylalanines of HPr could also be traced in the COSY spectrum and were confirmed by the C2,6H–C4H cross peaks observed in the RELAY spectrum shown in Figure 1. Although ring protons are not coupled strongly enough to their own C$^\beta$-protons to be observed in a COSY spectrum, the protons are usually near enough in space to give NOESY cross peaks. Since the C$^\beta$H resonances have been assigned sequentially, it is possible to assign the aromatic spin systems

† This work was supported by NIH AM-35187-01 (R.E.K.) and an MRC operating grant (E.B.W.). The 500-MHz spectrometer at the University of Washington is supported by instrumental 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 nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser effect; COSY, two-dimensional *J*-correlated spectroscopy; NOESY, two-dimensional NOE spectroscopy; RELAY, two-dimensional relayed coherence transfer spectroscopy.

0006-2960/86/0425-7774$01.50/0

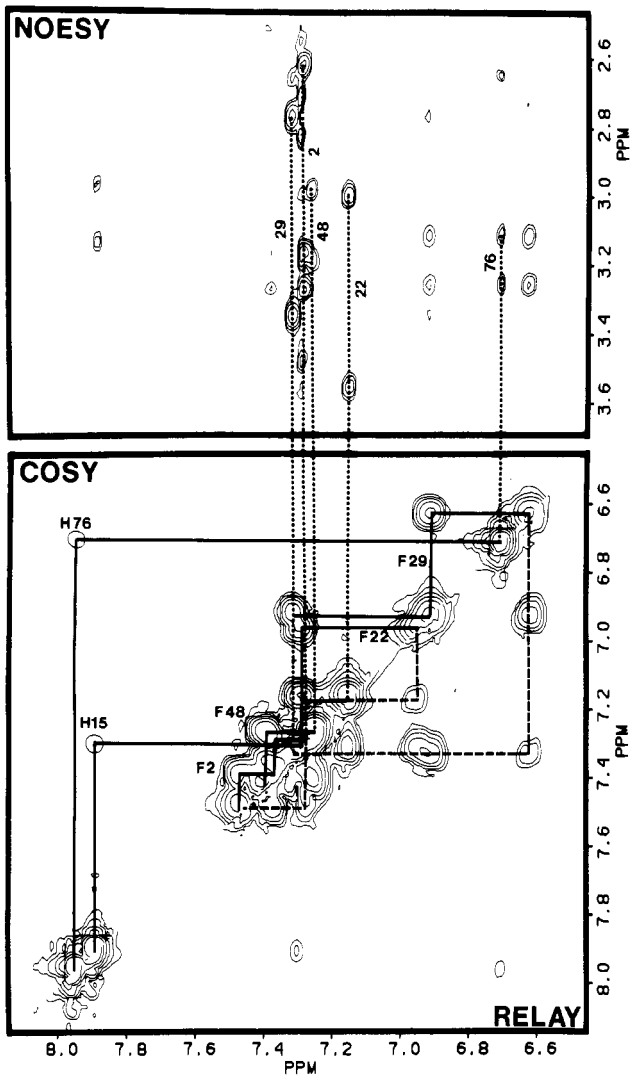

FIGURE 1: Connectivities for aromatic spin systems in HPr. The aromatic ring connectivities for the two histidines and four phenylalanines were followed in the COSY and RELAY spectra shown in the bottom half of the figure. The ring systems were then attached to the sequentially assigned $C^{\beta}H$ resonances via their NOEs, as shown in the upper half.

on the basis of such NOEs. The relevant region of the NOESY spectrum and the assignments are also shown in Figure 1.

The three most upfield aromatics in the spectrum all gave strong NOEs to the two $C^{\beta}H$ resonances of His-76. The middle of these three aromatic resonances was assigned to a histidine proton on the basis of the weak COSY cross peak to another singlet resonance and therefore must be from the C4H of His-76. The other aromatic resonances that gave NOEs to the $C^{\beta}H_2$s of His-76 are part of the Phe-29 ring system and must therefore be due to a long-range contact resulting from the tertiary fold of the protein.

Assignment of Methyl Group Resonances. The region of a D_2O COSY spectrum that contains the methyl cross peaks from valine, leucine, and isoleucine residues is shown in Figure 2. The $C^{\gamma}H_3$ resonances of valines and isoleucines could be assigned directly from a $C^{\alpha}H$-to-$C^{\gamma}H_3$ RELAY experiment. The C^{γ}-methyl resonances for all six valines and all three isoleucines in HPr were identified from the RELAY and long-range coherence transfer experiments (see the accompanying papers.) The resulting sequence-specific assignments of the methyl resonances are summarized in Figure 2. Furthermore, the $C^{\delta}H_3$ resonances of four of seven leucines

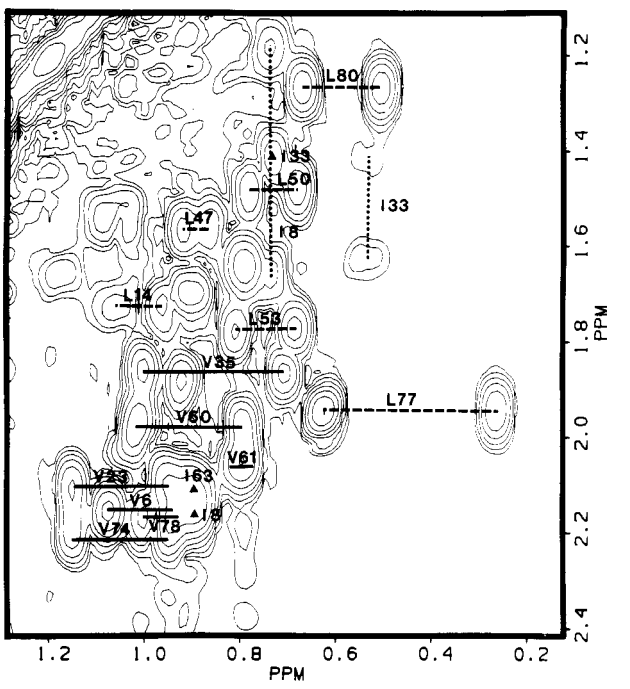

FIGURE 2: Methyl region of a COSY spectrum. The assignments of the methyl resonances of HPr are shown. The solid lines indicate the $C^{\beta}H$–$C^{\gamma}H_3$ cross peaks for valines, identified sequentially via $C^{\alpha}H$–$C^{\gamma}H_3$ RELAY cross peaks. The dashed lines identify $C^{\gamma}H$–$C^{\delta}H_3$ cross peaks of leucines, identified sequentially via long-range coherence transfer spectra (Leu-77, Leu-80, Leu-50, and Leu-53) or intraresidue NOEs (all others). (▲) indicate the positions of $C^{\beta}H$–$C^{\gamma}H_3$ cross peaks of isoleucines, as identified from $C^{\alpha}H$–$C^{\gamma}H_3$ RELAY peaks, while vertical dotted lines connect the pairs of $C^{\gamma}H$–$C^{\delta}H_3$ cross peaks for isoleucines.

(Leu-50, Leu-53, Leu-77, and Leu-80) and one isoleucine (Ile-8) were connected to their corresponding $C^{\alpha}H$ resonances in the long-range coherence transfer spectrum. These sequence-specific assignments are also illustrated in Figure 2. The remaining C^{δ}-methyl resonances were connected to sequentially assigned resonances on the basis of intraresidue NOEs. In each case, the methyl resonance gave an NOE cross peak to either the NH or the $C^{\alpha}H$ resonance of the leucine side chain.

In certain configurations, it is possible to obtain stereospecific assignments of the two $C^{\gamma}H_3$ resonances of valine residues. This is accomplished by comparing the intensities of the NH–$C^{\gamma}H_3$ and $C^{\alpha}H$–$C^{\gamma}H_3$ NOESY cross peaks: the methyl group that gives the stronger NOE to its own amide proton is assigned to the γ_1-methyl group. As previously shown for valines found in α-helices (Zuiderweg et al., 1984), the $C^{\gamma_1}H_3$ is closer both to its own amide proton and to the i + 1 amide proton. Stereospecific assignments can also be obtained for valines in β-sheets. In these structures, the $C^{\gamma_1}H_3$ is closer to its own amide proton while the $C^{\gamma_2}H_3$ is closer to the i + 1 amide proton. Both methyl groups will be close enough to the valine $C^{\alpha}H$ to give NOEs to that proton. The relevant region of an H_2O NOESY spectrum of HPr is shown in Figure 3. The β-sheet pattern of NOESY cross peak intensity was observed for Val-6, Val-35, and Val-60, while the α-helical pattern of intensities was observed for Val-23. Spectral overlap, either of the two methyl resonances or of NH resonances, precluded obtaining stereospecific assignments for the other three valines in HPr. The stereospecific assignments are included in Table I.

Persistent NH Resonance. In order to observe the resonances of slowly exchanging amide protons, COSY and NOESY experiments were performed on HPr samples that

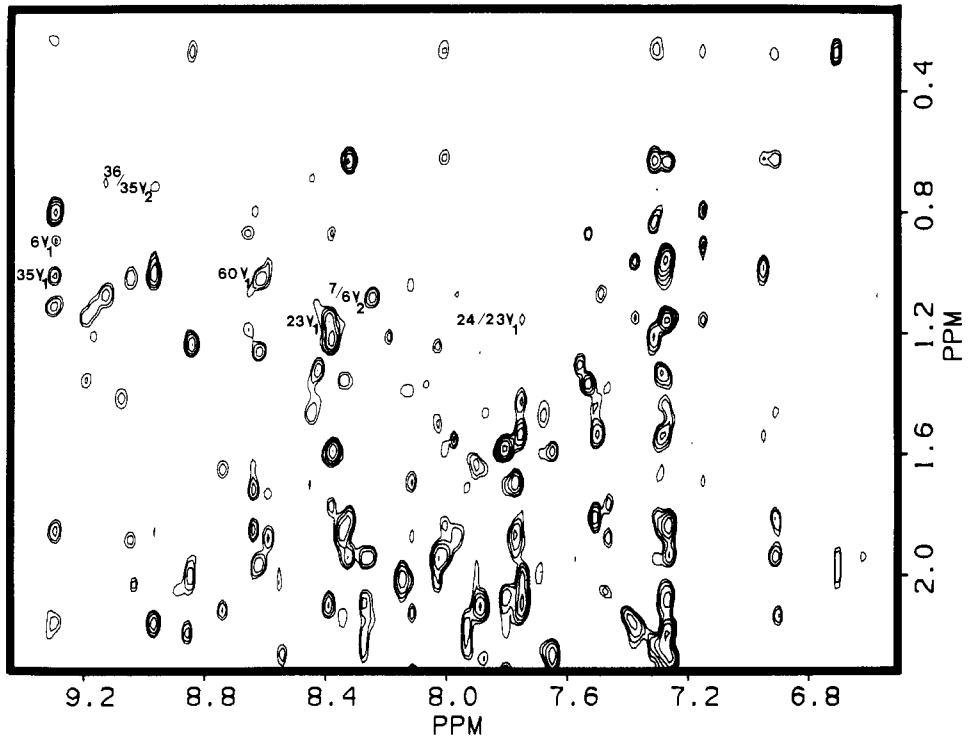

FIGURE 3: Stereospecific NOEs. The region of an H_2O NOESY spectrum that contains the stereospecific cross peaks for valine methyl groups is shown. Cross peaks labeled with a single number (e.g., $35\gamma_1$) are intraresidue cross peaks from amide to $C^{\gamma}H_3$ protons. Cross peaks labeled with two numbers (e.g., $7/6\gamma_2$) are interresidue cross peaks between the amide proton of the first residue (i.e., Thr-7) to the $C^{\gamma}H_3$ protons of the second residue (i.e., Val-6).

Table I: Side-Chain Resonance Assignments of the *E. coli* HPr ^1H NMR Spectrum[a]

residue	$C^{\gamma}H$	$C^{\delta}H$	$C^{\epsilon}H$	$C^{\zeta}H$	residue	$C^{\gamma}H$	$C^{\delta}H$	$C^{\epsilon}H$	$C^{\zeta}H$
Met-1	2.46, 2.26	1.98			Phe-48		7.24	7.27	7.27
Phe-2		7.27	7.37	7.27	Lys-49		0.90		
Glu-5	2.14				Leu-50	1.46	0.78, 0.60		
Val-6	0.90 (γ_1), 1.07 (γ_2)				Gln-51	2.62			
Thr-7	0.92				Thr-52	1.17			
Ile-8	0.86 ($C^{\gamma}H_3$), 1.65, 1.16	0.74			Leu-53	1.76	0.81, 0.68		
Thr-9	1.16				Gln-57	2.24			
Pro-11	3.00	4.01, 3.89			Thr-59	1.24			
Leu-14	1.71	1.06, 0.96			Val-60	1.00 (γ_1), 0.79 (γ_2)			
His-15		7.90	7.29		Val-61	0.78, 0.78			
Thr-16	1.28				Thr-62	0.99			
Arg-17	1.82, 1.67	3.12, 2.98			Ile-63	0.88 ($C^{\gamma}H_3$)			
Gln-21	2.47				Glu-66	2.15			
Phe-22		7.16	7.30	6.95	Lys-72	1.45			
Val-23	1.15 (γ_1), 0.96 (γ_2)				Val-74	1.15, 0.95			
Glu-25	2.58, 2.35				Glu-75	2.25			
Phe-29		7.28	6.87	6.60	His-76		7.91	6.68	
Thr-30	1.28				Leu-77	1.93	0.63, 0.26		
Glu-32	2.16, 2.29				Val-78	1.00, 0.95			
Ile-33	0.73 ($C^{\gamma}H_3$), 1.41, 1.61	0.54			Leu-80	1.25	0.66, 0.50		
Thr-34	1.09				Met-81		1.51		
Val-35	1.00 (γ_1), 0.90 (γ_2)				Leu-84	1.34	1.85, 1.68		
Thr-36	1.05				Glu-85	1.75, 1.66			
Leu-47	1.54	0.90, 0.86							

[a] ppm relative to DSS ± 0.03 ppm. Values are for HPr at 27 °C, pH 6.5.

had been lyophilized from H_2O buffer and then redissolved in D_2O. The NH resonances that appeared in these experiments are denoted with closed diamonds in Figure 4 (top), which also summarizes the nearest-neighbor connectivities used to obtain the sequential assignments reported in the first paper in this series. In general, under the experimental conditions used (pH 6.5, 30 °C), only those protons that are hydrogen-bonded will have exchange rates slow enough to be observed for the duration of the experiment. As expected, amides in α-helical regions (i.e., those that gave d_{NN} and $d_{\beta N}$ connectivities) and in β-sheets (i.e., those that gave strong $d_{\alpha N}$ connectivities) were slowly exchanging.

Secondary Structural Elements. It has been demonstrated that the regular secondary structural elements of proteins can be identified by visual inspection of the patterns of sequential connectivities observed in a given region of the primary structure (Wüthrich et al., 1984). Stretches of d_{NN} and $d_{\beta N}$ connectivities are indicative of α-helices, while stretches of $d_{\alpha N}$ connectivities are indicative of extended structures such as β-strands. Type I and type II β-turns can also be identified by their pattern of sequential connectivities. The observation of persistent amides and medium-range connectivities such as $d_{\alpha_i N_{i+3}}$ and $d_{\alpha_i \beta_{i+3}}$ add further support to the secondary structure assignments. These rules have been applied to a number of

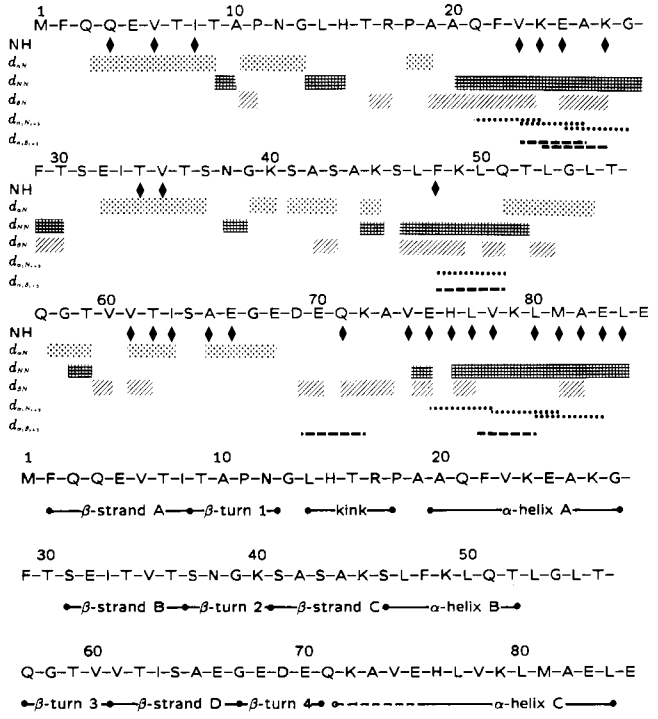

FIGURE 4: (Top) Summary of sequential connectivities observed for HPr. The three types of nearest-neighbor connectivities used for the sequential assignments (Klevit et al., 1986) are summarized along with additional connectivity information used to deduce the secondary structure of the protein. ($d_{\alpha\delta}$ and $d_{\beta\delta}$ connectivities involving prolines are indicated as $d_{\alpha N}$ and $d_{\beta N}$). The diamonds indicate slowly exchanging amide protons. The horizontal dotted and dashed lines connect residues for which medium-range $d_{\alpha_i N_{i+3}}$ and $d_{\alpha_i \beta_{i+3}}$ connectivities were observed. (Bottom) Secondary structure of *E. coli* HPr as deduced from the short-range and medium-range connectivities shown above.

proteins with known crystal structures [e.g., bovine pancreatic trypsin inhibitor (Wüthrich et al., 1984), *cro* repressor (Weber et al., 1985), and cytochrome *c* (Wand & Englander, 1985)] and have been shown to accurately identify regular secondary structures. The secondary structure of HPr, as deduced from the sequential connectivity patterns summarized in Figure 4 (top) is shown in Figure 4 (bottom).

On the basis of stretches of $d_{\alpha N}$ connections, we conclude that HPr contains four β-strands: strand A, Gln-3–Thr-9; strand B, Glu-32–Ser-37; strand C, Ser-41–Ser-43; strand D, Val-61–Gly-67. Strand A had persistent amides at every other residue, which is the pattern expected for a strand at the edge of a β-sheet. On the other hand, strands B and D had persistent amides next to each other in the sequence, indicating that these two strands are internal in the sheet. There were no persistent amides observed in strand C. Under the experimental conditions used, amides near the ends of the strands were not observed, and since strand C is particularly short, persistent amides would not expected.

There are three stretches of d_{NN} connectivities in HPr, which are now denoted as helix A, Ala-20–Gly-28, helix B, Ser-46–Thr-52, and helix C, Val-74–Leu-84. The identification of these regions as α-helices was further confirmed by the $d_{\beta N}$, $d_{\alpha_i N_{i+3}}$, and $d_{\alpha_i \beta_{i+3}}$ connectivities and the persistent NHs observed. In α-helices, the hydrogen bonds are between the carbonyl oxygen of residue *i* and the amide proton of residue *i* + 4, so the N-terminus of a helix can be deduced from the position of the first persistent NH resonance observed. Thus, the NH of Val-23 in helix A is H-bonded to Ala-19, the NH of Phe-48 in helix B is H-bonded to Ala-44, and the NH of Val-74 in helix C is H-bonded to Glu-70.

Table II: Medium- and Long-Range NOEs in *E. coli* HPr

residues belonging to	NOEs	residues belonging to	NOEs
helix A–β-sheet	Phe-22–Val-6	helix B–active site	Thr-52–Thr-16
	Phe-22–Val-61		
	Phe-22–Thr-62	helix C–β-sheet	Val-74–Phe-2
helix A–turn 4	Phe-29–Lys-72		Val-74–Ser-64
	Phe-29–Ala-73	turn 1	Ile-8–Ala-10
helix A–helix B	Ala-20–Lys-49	turn 1–turn 3	Ile-8–Thr-59
helix A–helix C	Phe-22–Leu-77		Ala-10–Gln-57
	Phe-22–Met-81	active site	Leu-14–Thr-16
	Val-23–Met-81		His-15–Arg-17
	Val-23–Val-78	C-terminus–active site	Leu-14–Glu-85
	Ala-26–Leu-77		His-15–Leu-84
	Phe-29–His-76		His-15–Glu-85
	Phe-29–Leu-77		Arg-17–Glu-85
helix B–β-sheet	Leu-50–Thr-36	turn 4	Asp-69–Lys-72
	Leu-50–Ser-41	turn 4–β-sheet	Glu-68–Ala-65
	Leu-50–Ala-42		Glu-68–Phe-2
helix B–turn 2	Thr-52–Gly-39		

HPr contains four tight turns, as evidenced by short- and medium-range NOEs (see Table II and Figure 4): turn 1, Thr-9–Asn-12, connects β-strand A to the segment containing the active site His-15; turn-2, Ser-37–Lys-40, is a type II β-turn that connects β-strands B and C; turn 3, Gln-57–Val-60, connects an extended segment at the end of helix B to β-strand D; turn 4, Glu-68–Gln-71 reverses the backbone between β-strand D and α-helix C so that these two elements are antiparallel. The amide proton of Gln-71 is one of the most persistent protons in the molecule (unpublished observation). At the present level of refinement, it is not possible to tell whether its persistence arises from a H-bond to a backbone carbonyl within the turn structure or to one of the many side-chain oxygens found in this region of the protein.

There are several regions of the protein that do not possess any regular secondary structure. In particular, His-15, the site of phosphorylation, appears to be on a turn or kink (see below). The region from Leu-53 to Leu-55 is in an extended conformation that leads into a reverse turn that leads into the β-strand beginning with Val-61.

Determination of Tertiary Structure of HPr. The nearest-neighbor NOEs used in the sequential assignment strategy represent only a fraction of the cross peaks that appeared in a NOESY spectrum. Many cross peaks connect resonances that are due to protons that are far apart in the primary sequence of the protein but that are in close proximity as a consequence of the tertiary fold of the molecule. Once assigned, these long-range NOEs can be used to determine the solution structure of a protein to a fairly high resolution (Kaptein et al., 1985; Havel & Wüthrich, 1985). Although the distance geometry and molecular dynamic techniques require a large number of such long-range NOEs to be assigned, it is possible to determine the overall fold of the protein on the basis of a fairly limited number of such NOEs.

Building β-Sheet. The alignment of β-strands in a β-sheet could be determined unambiguously from NOEs that occur between protons from adjacent strands. Figure 5 shows the region of a D_2O NOESY spectrum containing the strong $C^\alpha H$–$C^\alpha H$ cross peaks due to these short cross-strand distances. A total of nine cross-strand $C^\alpha H$–$C^\alpha H$ cross peaks were observed for HPr. In addition, weaker cross-strand NOEs were observed from six amide resonances to $C^\alpha H$ resonances as well as six cross-strand NH–NH cross peaks. There is exactly one way to align the four β-strands that is consistent with all the NOEs observed, as shown in Figure 6. Thus, the four β-strands of HPr form a single antiparallel β-sheet, with the connectivity A–D–B–C. The observation of

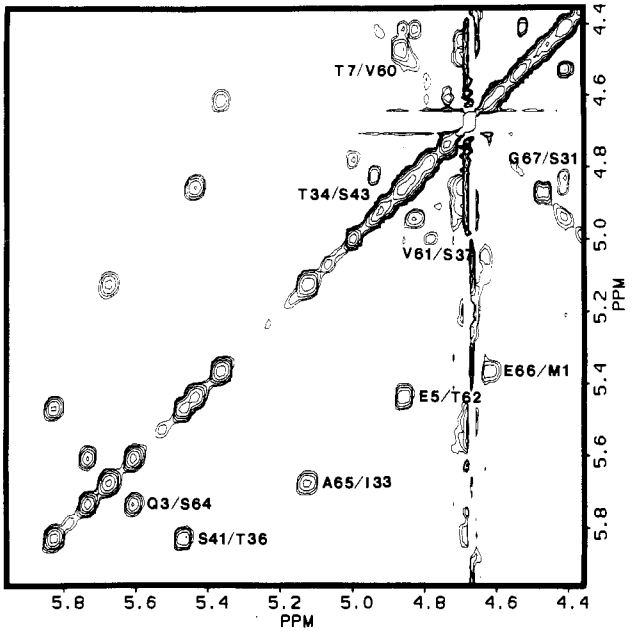

FIGURE 5: Cross-strand NOEs. The region of a D_2O NOESY spectrum that contains the $C^{\alpha}H$–$C^{\alpha}H$ cross peaks is shown. The cross peaks are labeled with the two sequentially assigned $C^{\alpha}H$ resonances that the peaks connect.

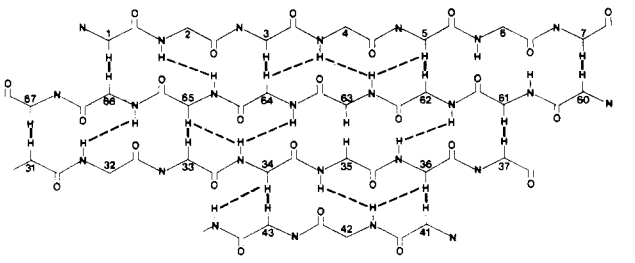

FIGURE 6: β-Sheet of HPr. The structure of the antiparallel β-sheet as determined from long-range NOEs is shown. Short solid lines connect pairs of C^{α}Hs for which cross-strand NOEs were observed (see Figure 5). Diagonal dashed lines connect the amide and C^{α}-protons for which long-range NOEs were observed. Note the lack of expected connectivities around residues 35 and 63, indicating a distortion in the sheet structure.

a cross-strand $C^{\alpha}H$–$C^{\alpha}H$ NOE between Met-1 and Glu-66 reveals that β-strand A actually begins at Met-1, rather than Gln-3 as assigned on the basis of $d_{\alpha N}$ connectivities. The β-sheet construction shown is also consistent with the pattern of persistent amides observed (Figure 4, top).

There was a missing nearest-neighbor $d_{\alpha N}$ connectivity between Ile-63 and Ser-64 in the middle of strand D (Figure 4, top). Furthermore, there is one $C^{\alpha}H$–$C^{\alpha}H$ NOE that is predicted from the sheet structure but not observed, between Ile-63 and Val-35. Also, the two internal amide protons of Thr-36 and Ser-64 were not persistent. Taken together, these observations are suggestive of a β-bulge structure. However, as Figure 6 shows, the proper cross-strand NOEs are observed on either side of Ile-63, so the bulge is not the consequence of having an extra residue on one strand but is instead caused by irregularities in the backbone. This is likely due to the presence of the branched side chains of Val-35 and Ile-63 on neighboring strands. In order for these two side chains to pack well, one of the must adopt an unfavorable side-chain configuration, or alternatively, the main chain must twist (Richardson, 1981).

When the side chains in the four-strand β-sheet are categorized by their hydrophobicity and hydrophilicity, an alternating pattern emerges. Since side chains in the sheet point

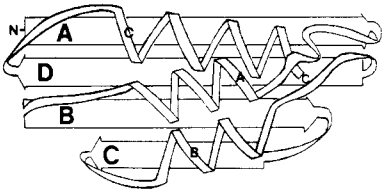

FIGURE 7: Tertiary folding of *E. coli* HPr. A model of the structure of HPr as determined from the 2DNMR data is shown. The β-strands and α-helices are labeled A–D and A–C, respectively, starting at the N-terminus of the molecule.

alternately up and down, the sheet has one hydrophobic face and one hydrophilic face.

Placing α-Helices. With the β-sheet determined, the positions of the three α-helices relative to the sheet were defined by treating the helices as semirigid units. On the basis of a small number of contacts, a model was built that could then be tested by searching for NOEs predicted to appear by the model. This process was repeated until the structure was consistent with all the identified NOEs. The long-range NOEs among assigned side-chain protons used in this process are summarized in Table II.

The NOEs observed between Phe-22 near the N-terminal end of helix A and protons of Val-6, Val-61, and Thr-62 on the hydrophobic face of the β-sheet gave the first point of contact. Since there are only two residues between the end of helix A (Gly-28) and the beginning of β-strand B (Ser-31), the helix could be placed relative to the β-sheet, stretching from Val-6/Val-61 to Glu-32. The position of helix B relative to the β-sheet was defined by the NOEs between the methyl protons of Leu-50 and the protons of Thr-36, Ser-41, and Ala-42, on the hydrophobic face of the sheet. These, with an additional NOE between Thr-62, the last residue in helix B, and Gly-39, in the middle of the β-turn between strands B and C, defined the position of the C-terminal end of the helix. Since the helix begins with residue 46, with only Lys-45 between it and the end of β-strand C, the position of the N-terminus of this helix could be fixed. A check on the structure thus far was afforded by the observation of an NOE between Ala-20, at the N-terminus of helix A, and Lys-49, near the C-terminus of helix B. Thus, these helices are approximately antiparallel to each other. The NOEs observed from the methyl protons of Val-74, at the N-terminal end of helix C, to the ring protons of Phe-2 and to Ser-64 placed this end of helix C. The C-terminus of helix C could not be placed on the basis of NOEs to the β-sheet. Instead, contacts between helix C and helix A were used to define the position of helix C. In particular, NOEs were observed among Val-78 and Met-81 on the hydrophobic face of helix C and Phe-22/Val-23 on the hydrophobic face of helix A.

The next group of contacts used involved the ring protons of Phe-29. This residue is not actually part of the regular α-helical structure but instead forms a turn with Thr-30 that links helix A to β-strand B. The aromatic protons of Phe-29 gave NOEs to side-chain protons of His-76 and Leu-77 on helix C. Since the methyl protons of Leu-77 also gave NOEs to the ring protons of Phe-22, this places the two helices relative to each other, with the side chain of Leu-77 coming between the two phenylalanines at either end of helix A. Final adjustment of helix C was made on the basis of NOEs observed from the C-terminus to the active site, as described below. The placement of the three helices on top of the β-sheet according to these NOEs is shown schematically in Figure 7.

Defining Phosphorylation Site. The backbone segment surrounding His-15, the site of phosphorylation in HPr, is one of three regions of the molecule that does not possess regular

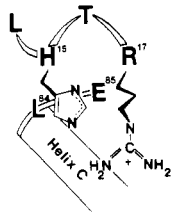

FIGURE 8: Phosphorylation site of *E. coli* HPr. Representation of the configuration around the active His-15 residue that is consistent with the short-, medium-, and long-range NOEs observed.

secondary structure. However, the backbone conformation could still be defined on the basis of short-range NOEs. As shown in Figure 4 (top), $d_{\alpha N}$ connectivities were observed for the pairs Pro-11/Asn-12 and Asn-12/Gly-13, indicating that the backbone is in an extended conformation coming out of the reverse turn that starts with residue 9. Next, d_{NN} connectivities were observed for the pairs Gly-13/Leu-14 and Leu-14/His-15, indicating a turn or coil structure. No nearest-neighbor connectivity was observed between His-15 and Thr-16. Finally, a $d_{\beta N}$ connectivity was observed between Thr-16 and Arg-17. In addition to the $i \rightarrow i + 1$ connectivities, two $i \rightarrow i + 2$ connectivities were observed: an NOE between one of the $C^{\delta}H_3$ groups of Leu-14 and the amide proton of Thr-16 and a $d_{\beta N}$ connectivity between His-15 and Arg-17. Taken together, these short-range NOEs indicate that the backbone is in a kinked conformation around the active histidine.

Short-range side-chain NOEs were observed that define the disposition of the imidazole ring: the C2 proton resonance of His-15 gave NOEs to both the $C^{\beta}H$ and $C^{\delta}H$ resonances of Arg-17, and the C4 proton gave an NOE to one of the methyl resonances of Leu-14, consistent with the conformation shown in Figure 8.

As mentioned earlier, long-range NOEs were observed from protons at the extreme C-terminus of HPr and the site of phosphorylation (see Table II). One of the $C^{\beta}H$ resonances of His-15 gave a weak NOE to the $C^{\beta}H$ resonance of Leu-84. Together with the observation that the methyl groups of Leu-14 gave NOEs to the amide, C^{β}, and C4 protons of His-15, this indicates that there is a hydrophobic patch on one side of the phosphorylation site. The imidazole C4 proton of His-15 also gave an NOE to a $C^{\gamma}H$ resonance of Glu-85, the last residue in HPr, while the $C^{\alpha}H$ of the terminal residue gave weak NOEs to the $C^{\beta}H$ and $C^{\delta}H$ of Arg-17. On the other side of the active site, a contact was observed between the methyl group of Thr-16 and the C-terminal residue of helix B, Thr-52. A picture of the phosphorylation site consistent with all these observations is shown in Figure 8.

Defining Remaining Structure. With the β-sheet, the α-helices, and the site of phosphorylation defined, only the short stretches of backbone that link these units together remained. As mentioned earlier, there are four regions that give short-range NOE patterns associated with turn structures: (1) Thr-9 through Asn-12 links β-strand A to the kinked region around His-15, (2) Asn-37 through Lys-40 is a reverse turn that links strands B and C together, (3) Gln-57 through Val-60 links a short extended stretch of backbone that comes out of helix B to β-strand D, and (4) Glu-68 through Gln-71 reverses the backbone as it comes out of the β-sheet at strand D. NOEs observed among protons in turns 1 and 3 (Table II) define the relative positions of these two turns. Similarly, contacts between Glu-68, in turn 4, and β-sheet protons of Phe-2 and Ala-65 place this turn over the hydrophobic edge of the sheet. Finally, the stretch of extended backbone from Leu-53 to Thr-56 links helix B to turn 3. Both β-turns and the extended regions are represented by ribbons in the HPr structure shown in Figure 7.

Discussion

A model of the solution structure of *E. coli* HPr has been determined from only primary sequence information and 2DNMR data. This was accomplished in three steps: (1) the sequence-specific assignments were determined, (2) the secondary structure was deduced from the patterns of short-range NOEs, and (3) the tertiary structure was obtained by determining the relative positions of the elements of secondary structure on the basis of medium- and long-range NOEs.

The secondary structures of HPr as determined from the NMR include three α-helices, four β-strands, and four reverse turns. The circular dichroic spectrum of *E. coli* HPr (Weigel et al., 1982) is indicative of a high degree of secondary structure, with a mean residue ellipticity at 222 nm of approximately $-18\,000$ deg cm^2 dmol^{-1}. The secondary structure of HPr had been predicted from the primary sequence of the protein (Weigel et al., 1982).[2] The algorithm used (Chou & Fasman, 1974) correctly predicted the three α-helices, two of the four β-strands, (strands II and IV), and all four turns but incorrectly predicted a fourth helix, extending from residue 1 to residue 8, which actually constitute β-strand A and a "random-coil" configuration in the vicinity of His-15 rather than the kink observed. Secondary structure predictions using the method of Garnier et al. (1978) gave results similar to those presented by Weigel et al. (1982).

The site of phosphorylation, His-15, is in the middle of a stretch of polypeptide that is set apart from the rest of the backbone by the two prolines, Pro-11 and Pro-18. This six-residue segment adopts a special kinked conformation that allows the histidine side chain to interact with the side chain of Arg-17. The long-range NOEs observed between His-15/Arg-17 and Leu-84/Glu-85 reveal that the C-terminus of HPr interacts with the phosphorylation site, placing a significant amount of negative charge in the otherwise positively charged site. The lack of long-range side-chain NOEs for the other residues in the phosphorylation site indicates that the kink is lying on the surface of the protein.

Since the pK_a of His-15 is 5.6 (Doiijewaard et al., 1979; Kalbitzer et al., 1982), the imidazole side chain is deprotonated at the pH at which the structure was determined (as it is at physiological pH). Therefore, the negatively charged glutamate is more likely to be interacting with the positively charged arginine than with the neutral histidine. This arrangement leaves the deprotonated histidine, which is a better nucleophile than its protonated counterpart, available to accept a phosphoryl group at its N-1. Recently, Waygood et al. (1985) have studied the hydrolysis of the 1-phosphohistidinyl residue, and the results are satisfactorily explained by the inclusion of an acidic residue at the active site. Waygood et al. (1985) proposed that the acidic residue was Glu-66 on the basis of a set of X-ray diffraction data that are now known to be unreliable. Interpretation of more recent X-ray diffraction data is now consistent with the model proposed on the basis of the 2DNMR results, placing Glu-85 at the active site (O. A. L. El-Kabbani, E. B. Waygood, and L. T. J. Delbaere, unpublished results). Thus, the interpretation of the active site mechanism remains essentially the same except the glutamate

[2] The secondary structure prediction in Weigel et al. (1982) was carried out on the sequence of *S. typhimurium* HPr, which is identical with that of the *E. coli* protein. Several corrections to the original sequence have been subsequently made (Powers & Roseman, 1984; De Reuse et al., 1985).

is residue 85. pH dependence of the hydrolysis of the 1-phosphohistidinyl residue does not distinguish between the involvement of the side chain or backbone carboxyl group of Glu-85.

Dooijewaard et al. (1979) showed that the pK_a of His-15 increases to 7.8 upon phosphorylation. It is tempting to speculate that in the phosphorylated protein the glutamate carboxylate moves over to interact with the protonated imidazole, thus stabilizing it. The presence of Glu-85 in the site of phosphorylation could serve to stabilize the protonated imidazole, which in turn destabilizes the phosphoramidate bond. Thus, the structure of the site is designed to enhance the favorability of both the phosphate acceptor and the phosphate donator roles of the protein. Although this model can only be proved by determining the structure of the phosphorylated protein (a project that is currently under way), there is already some evidence in the literature that supports it. Dooijewaard et al. (1979) have published 1D spectra of unphosphorylated and phosphorylated *E. coli* HPr and pointed out several peaks that were clearly sensitive to the state of the protein. On the basis of the sequential assignments, it is now possible to go back to those spectra and identify several of the sensitive peaks, in particular, the methyl resonances of Leu-80 (0.66 and 0.70 ppm) and the δ-methyl peak of Met-81 (1.51 ppm). These residues lie on the same face of helix C, one turn away from Leu-84 which interacts with His-15. Thus, it appears that phosphorylation of His-15 does result in a change in the environment of helix C. Furthermore, Doiijewaard et al. (1979) pointed out that a similar spectral change could be induced by dropping the pH of unphosphorylated HPr down to where the imidazole ring would be protonated.

E. coli HPr has an "open-face sandwich" structure (Richardson, 1981) with a four-strand antiparallel β-sheet base and three α-helices that cover the hydrophobic face of the sheet, resulting in a compact ellipsoid shape. There are several features that are revealed by model building. With reference to Figure 7, the left-hand end of the molecule is rich in acidic residues and thus gives a negatively charged surface topography. The forward upper central portion is rich in lysines that give a positively charged surface area, which potentially covers Ser-46, the site of ATP-dependent protein kinase phosphorylation in Gram-positive bacteria (Deutscher et al., 1986). The right-hand end of the molecule up to the active site is a neutral to hydrophobic area that may be important to the phosphoryl transfer interaction with the sugar-specific phosphocarrier protein $III^{glucose}$ in which a 3-phosphohistidine is formed. This active site histidine, His-91, is found in a hydrophobic sequence (Doerschug et al., 1984; Nelson et al., 1984). There are four phenylalanine residues whose locations are such that they could exist as two pairs: Phe-2 and Phe-29 and Phe-22 and Phe-48. Burley and Petsko (1985) have pointed out the potential for such pairs of aromatic residues to act as stabilizers of structure. The only direct interaction between aromatic rings observed for *E. coli* HPr, however, was between Phe-29 and His-76. The lack of potential effects on the active site of the deamidation of glutamine residues resulting in the formation of HPr-1 and HPr-2 has been previously discussed (Waygood et al., 1985). In contrast, the structure presented here suggests that a deamidation of Gln-57 could result in some interaction with the active site His-15. 2DNMR spectra of HPr-1 indicate that the residues around Gln-57 as well as Leu-14 and His-15 experience a change in their environment (as judged by their resonance positions) in the deamidated protein (R. E. Klevit and E. B. Waygood, unpublished results).

HPr molecules from Gram-positive bacteria (Beyreuther et al., 1977; Deutscher et al., 1986) have a high degree of sequence homology with *E. coli* HPr in the active site region (residues 12–18). However, the only C-terminal amino acid residue that is well established is in the *Streptococcus faecalis* HPr, which has a C-terminal glutamine (residue 89), but is accompanied by Glu-88.[3] Thus, either the carboxyl terminus or the glutamate carboxyl groups are candiates if the active site of this HPr molecule has the same active site arrangement as the *E. coli* protein. For reasons given below, it is plausible that the different HPr molecules are all similarly constructed. Several studies with HPrs from Gram-positive bacteria have invoked a tyrosyl residue as a component of the active site with Tyr-37 being the most likely (Kalbitzer et al., 1982, 1985). This residue is a serine in *E. coli* HPr, and it is not directly involved with the active site according to the 2DNMR data.

There is considerable homology throughout the sequences of HPrs. Residues 39–41 appear to involve a conserved turn; residues 45–47 are a conserved site for ATP-dependent phosphorylation; residues 52–55 are a conserved strand that is next to the active site strand in the model; residues 58–70 involve the central β-sheet strand D and the beginning of a very hydrophilic turn and are also highly conserved. If one compares the differences in amino acids in the three known sequences, there are few differences that involve considerable change in either the chemical nature of the residue or the hydropathic nature of the residue. In fact, the three sequences when compared by hydropathic plots (Hopp & Woods, 1981) show that the three HPrs have considerable overall similarities except in the N-terminal region (not shown). The N-terminal region of *E. coli* HPr is slightly hydrophobic in keeping with its β-sheet structure. However, *S. faecalis* HPr has a very hydrophilic N-terminal region while the *Bacillus subtilis* HPr N-terminal region is intermediate between the other HPrs. It is the N-terminal region that has recently been characterized as the epitope for two monoclonal antibodies against *E. coli* HPr (Waygood & Lee, 1986). The substitution of Lys-4 for Gln-4 as occurs in HPr from *Salmonella typhimurium* strain SB3093 (Beneski et al., 1982) resulted in one antibody having no detectable reactivity and the other having 10-fold less binding ability. HPrs from *Staphylococcus aureus* and *S. faecalis* had no detectable reaction with either antibody. 2DNMR spectra of SB3093 HPr revealed that the single amino acid substitution does not disrupt the β-sheet structure of the protein (R. E. Klevit and E. B. Waygood, unpublished results). An antibody to a β-pleated sheet structure is unusual.

The C-terminal α-helix and the hydrophilic turn proceeding it appear to be very highly conserved among the HPrs with respect to their hydropathic properties. Yet the amino acid sequences of the C-terminal portion do show considerable divergence. As this region is involved in the active site, perhaps the lack of sequence homology in this region accounts for the inability of PTS components from Gram-positive and Gram-negative bacteria to successfully complement each other (Simoni et al., 1973).

The ATP-dependent protein kinase activity that phosphorylates HPrs from Gram-positive bacteria has been shown to phosphorylate Ser-46 (Deutscher et al., 1986). This seryl residue and its neighbor Lys-45 are highly conserved, yet no acid-stable ATP-dependent phosphorylation has been detected in HPrs from Gram-negative bacteria (Waygood et al., 1984), and the purified kinase does not phosphorylate *E. coli* HPr

[3] The *S. aureus* HPr sequence (Beyeuther et al., 1977) is known to contain omissions and errors (W. Hengstenberg, personal communication).

(Reizer et al., 1984). The transfer of phosphoryl groups to serine residues has been correlated with the proximity of basic residues either on the substrate or at the active site of kinase (Shlyapnikov et al., 1975; Kochetov et al., 1977). From model building, lysine residues 27, 45, and 49 can all be oriented over Ser-46 in *E. coli* HPr. In Gram-positive bacteria, Lys-49 is not present, and Lys-27 is replaced by a conserved Ser-27, which is curiously next to conserved Lys-28. Perhaps either the number or the placement of these lysines is the determining function for the ATP-dependent protein kinase.

In summary, the model for the solution structure of *E. coli* HPr presented here is consistent with all the assigned NOEs. It was determined from only 2DNMR data along with the primary sequence of the protein. While this work was being done, analysis of X-ray diffraction data was being performed concurrently (O. A. L. El-Kabbani, E. B. Waygood, and L. T. J. Delbaere, unpublished results), and the results from this work, along with a comparison of the solution and crystal structure, should be available in the near future. The solution structure as defined by proton–proton NOEs is currently being refined with constrained molecular dynamics that include the observed NOEs as constraints (F. Hoesel, G. Robilliard, E. B. Waygood, and R. Klevit, unpublished results).

Several structural features of the HPr molecule are presently being investigated to determine the structure of phospho-HPr by 2DNMR and to determine the relevant roles of specific amino acids by site-directed mutagenesis. By this combined approach, it is hoped that information about protein structure and phosphoryl-transfer mechanism will be forthcoming.

ACKNOWLEDGMENTS

We thank Ossama El Kabbani and Louis Delbaere for their continuing interest in this project and Jon Herriott and Gary Drobny for many useful discussions and critical comments on the manuscript.

REFERENCES

Beneski, D. A., Nakazawa, A., Weigel, N., Hartman, P. E., & Roseman, S. (1982) *J. Biol. Chem. 257*, 14492–14498.

Beyreuther, K., Raufuss, H., Schrecker, O., & Hengstenberg, W. (1977) *Eur. J. Biochem. 75*, 275–286.

Burley, S. K., & Petsko, G. A. (1985) *Science (Washington, D.C.) 229*, 23–28.

Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 222–245.

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

Deutscher, J., Pevec, B. Beyeuther, K., Kiltz, H.-H., & Hengstenberg, W. (1986) *Biochemistry* (in press).

Doerschug, M., Frank, R., Kalbitzer, H. R., Hengstenberg, W., & Deutscher, J. (1984) *Eur. J. Biochem. 144*, 113–119.

Dooijewaard, G., Roosien, F. F., & Robillard, G. T. (1979) *Biochemistry 18*, 2996–3001.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281–294.

Hopp, T. P., & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3824–3828.

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

Kalbitzer, H. R., Muss, H. P., Engelmann, R., Klitz, H. M., Stuber, K., & Hengstenberg, W. (1985) *Biochemistry 24*, 4562–4569.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179–182.

Klevit, R. E., & Drobny, G. (1986) *Biochemistry* (second paper of three in this issue).

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

Kochetov, S. N., Bulgargina, T. V., Sashchenko, L. P., & Severin, E. S. (1977) *Eur. J. Biochem. 81*, 111–118.

Nelson, S. O., Schuitema, A. R.-J., Benne, R., van der Ploeg, L. H. T., Plijter, J. S., Aan, F., & Postma, P. W. (1984) *EMBO J. 3*, 1587–1593.

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

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

Reizer, J. Novotny, M. J., Hengstenberg, W., & Saier, M. H., Jr. (1984) *J. Bacteriol. 160*, 333–340.

Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 168–363.

Shlyapnikov, S. V., Arutyunyan, A. A., Kurochkin, S. N., Memolova, L. V., Nersterova, M. V., Sashchenko, L. P., & Severin, E. S. (1975) *FEBS Lett. 53*, 316–319.

Wand, A. J., & Englander, S. W. (1986) *Biochemistry 25*, 1100–1106.

Waygood, E. B., & Lee, J. S. (1986) *Proc. Can. Fed. Biol. Soc. 29*, 194.

Waygood, E. B., Mattoo, R. L., & Peri, K. G. (1984) *J. Cell. Biochem. 25*, 139–159.

Waygood, E. B., Erickson, E., El Kabbani, O. A. L., & Delbaere, L. T. J. (1985) *Biochemistry 24*, 6938–6845.

Weber, P. L., Wemmer, D. E., & Reid, B. R. (1985) *Biochemistry 24*, 4553–4562.

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

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

Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1984) *Biopolymers 23*, 601–611.