Tanaka, S., & Koike, K. (1977) *Biochim. Biophys. Acta 479*, 290-299.
Villani, G., Fay, P. J., Bambara, R. A., & Lehman, I. R. (1981) *J. Biol. Chem. 256*, 8202-8207.
Weaver, D. T., & DePamphilis, M. L. (1982) *J. Biol. Chem. 257*, 2075-2086.
Wernette, C. M., & Kaguni, L. S. (1986) *J. Biol. Chem. 261*, 14764-14770.
Weymouth, L. A., & Loeb, L. A. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1924-1928.
Wolstenholme, D. R., & Clary, D. O. (1985) *Genetics 109*, 725-744.
Wolstenholme, D. R., Goddard, J. M., & Fauron, C. M.-R. (1979) in *Extrachromosomal DNA* (Cummings, D., Dawid, I. B., Borst, P., Weissman, S., & Fox, C. F., Eds.) pp 409-425, Academic, New York.
Yamaguchi, M., Matsukage, A., & Takahashi, T. (1980a) *Nature (London) 285*, 45-47.
Yamaguchi, M., Matsukage, A., & Takahashi, T. (1980b) *J. Biol. Chem. 255*, 7002-7009.

Sequence of Cloned Enzyme II$^{N\text{-acetylglucosamine}}$ of the Phosphoenolpyruvate:*N*-Acetylglucosamine Phosphotransferase System of *Escherichia coli*†

Krishna G. Peri and E. Bruce Waygood*
Department of Biochemistry, University of Saskatchewan, Saskatoon, S7N 0W0 Canada
Received November 12, 1987; Revised Manuscript Received March 2, 1988

ABSTRACT: In *Escherichia coli*, *N*-acetylglucosamine (nag) metabolism is joined to glycolysis via three specific enzymes that are the products of the *nag* operon. The three genes of the operon, *nagA*, *nagB*, and *nagE*, were found to be carried by a colicin plasmid, pLC5-21, from a genomic library of *E. coli* [Clarke, L., & Carbon, J. (1976) *Cell (Cambridge, Mass.) 9*, 91-99]. The *nagE* gene that codes for enzyme II$^{N\text{-acetylglucosamine}}$ of the phosphoenolpyruvate:sugar phosphotranferase system (PTS) was sequenced. The *nagE* sequence is preceded by a catabolite gene activator protein binding site and ends in a putative rho-independent termination site. The amino acid sequence determined from this DNA sequence shows 44% homology to enzymes II^{glucose} and III^{glucose} of the PTS. Enzyme II$^{N\text{-acetylglucosamine}}$, which has 648 amino acids and a molecular weight of 68 356, contains a histidine at residue 569 which is homologous to the active site of III^{glc}. Sequence homologies with enzymes II^{glucose}, II$^{\beta\text{-glucoside}}$, and II^{sucrose} indicate that residues His-190, His-213, and His-295 of enzyme II^{nag} are also conserved and that His-190 is probably the second active site histidine. Other sequence homologies among these enzymes II suggest that they contain several sequence transpositions. Preliminary models of the enzymes II are proposed.

The bacterial phosphoenolpyruvate:sugar phosphotransferase system (PTS)[1] carries out the concomitant translocation and phosphorylation of various sugars in a variety of bacterial species. The PTS was first described in *Escherichia coli* by Kundig et al. (1964), and considerable information about the components of the PTS and their interactions has since been gathered. The PTS in *E. coli* is important as the predominant way by which many hexoses enter the cell as carbon sources for metabolism; as a system involved in the regulation of adenylate cyclase, various permeases, and other enzymes; and as a system involved in chemotaxis [see review by Postma and Lengeler (1985)].

The PTS has both soluble and membrane-bound components. The integral membrane component, enzyme II^{sugar}, is often an inducible sugar-specific enzyme that carries out both the translocation and phosphorylation of the sugar. There are two classes of enzyme II^{sugar}: those that interact directly with the common phosphocarrier protein, HPr, and those that interact with specific phosphocarrier proteins, III^{sugar}. Citing several lines of evidence, Saier et al. (1985) proposed that these two clases are functionally and evolutionarily related. The enzymes II of the former class have molecular weights close to 65 000, while the latter class has a similar molecular weight when the enzyme II^{sugar} and III^{sugar} pair are considered together. It was proposed that the former class is a fused protein with dual function of the latter class, a single gene split into two. Protein phosphorylation (Waygood et al., 1984; Peri et al., 1984) and phosphoryl-transfer experiments (Begley et al., 1982) indicated that all the PTS components, and in particular the enzymes II, could be phosphorylated. This led to the proposal that an enzyme II that interacts directly with P-HPr has a triple-displacement mechanism for phosphoryl transfer and thus should contain two phosphorylation sites (Saier et al. 1985).

Enzyme II$^{N\text{-acetylglucosamine}}$ has been identified as an integral membrane enzyme (M_r 65 000), which interacted directly with P-HPr (Waygood et al, 1984). Enzyme II^{mannitol} interacts directly with P-HPr and forms a phosphoprotein similar to enzyme II^{nag} (Waygood et al, 1984). Enzyme II^{nag} therefore

† This work was supported by an operating grant to E.B.W. from the Medical Research Council of Canada and a Studentship Award to K. G.P. from the Saskatchewan Health Research Board.
* Author to whom correspondence should be addressed.

[1] Abbreviations: PTS, phosphoenolpyruvate:sugar phosphotransferase; HPr, histidine-containing phosphocarrier protein of the PTS; NAG or nag, *N*-acetylglucosamine; III^{sugar}, a sugar-specific phosphocarrier protein (also called enzyme III or factor III). Enzymes II^{sugar} have sugars abbreviated: glc, glucose; man, mannose; mtl, mannitol; gut, glucitol; scr, sucrose; bgl, β-glucoside.

0006-2960/88/0427-6054$01.50/0

had the potential to show structural and functional homologies to enzyme II^{mtl}, which had been sequenced and purified (Jacobson et al., 1983b; Lee & Saier, 1983). In addition, enzyme II^{nag} had the potential to show homologies to either the glucose- or mannose-specific PTS enzymes II and III^{sugar} because of overlapping sugar specificities (White, 1970; Adler & Epstein, 1974; Curtis & Epstein, 1975; Stock et al., 1982).

Considerable information on the phosphotransferase systems of glucose, mannose, sucrose, glucitol, and β-glucoside has subsequently become available from cloning and sequencing (Erni & Zanolari, 1986; Erni et al., 1987; Bramley & Kornberg, 1987a,b; Ebner & Lengeler, 1987; Schnetz et al., 1987; Yamada & Saier, 1987). In order to seek information about homologies to other enzymes II^{sugar} and III^{sugar} and to describe a membrane-bound component of the PTS, the NAG operon was cloned (details to be published elsewhere). The nucleotide sequence of the gene, *nagE*, that codes for enzyme II^{nag} is reported here.

Materials and Methods

Materials. Restriction endonucleases, DNA polymerase (Klenow fragment), 2′,3′-dideoxynucleotide 5′-triphosphates, mung bean nuclease, exonuclease III, T4 DNA ligase, and M13 universal primer were obtained from Pharmacia. Sequenase, a modified T7 DNA polymerase, was obtained from U.S. Biochemical Corp. Adenosine 5′-[^{35}S]thiotriphosphate (sp act. >1300 Ci/mmol), *N*-acetyl[1-^{14}C]-D-glucosamine (sp act. 35.6 mCi/mmol), and 2-deoxy[^{14}C(U)]glucose (sp act. 286 mCi/mmol) were obtained from New England Nuclear. *N*-Acetylglucosamine, bovine serum albumin, ampicillin, chloramphenicol, and tetracycline were from Sigma. Bacterial medium was from Difco. Agarose was from Bethesda Research Laboratories, and other electrophoretic reagents were from Bio-Rad. HPr and enzyme I were prepared as previously described (Waygood & Steeves, 1980).

Bacterial Strains, Plasmids, and Growth Conditions. The bacterial strains used were the following: *E. coli* strain ZSC114, *ptsMg1k7rpsL* (Curtis & Epstein, 1975); strains KPN14 and KPN16 that were selected from ZSC114 as strains incapable of fermenting 1% NAG contained in MacKonkey's agar, following *N*-methyl-*N*′-nitrosoguanidine mutagenesis (Miller, 1972). KPN14 and KPN16 were subsequently characterized as lacking enzyme II^{nag} activity by a sugar phosphorylation assay (Table I). Plasmid pLC5-21 from the *E. coli* genomic library of Clarke and Carbon (1976) was identified as containing the *nag* operon (K. G. Peri & E. B. Waygood, unpublished results). Plasmid pLC5-21 was digested with restriction endonuclease *Bgl*II to yield several products including a 5.5-kb fragment that was cloned into the *Bam*HI site in the tetracycline resistance gene of pBR322 to yield pKP1. This fragment contained the *nag* operon. A subclone, pKP1.3, was produced, and it contained the *nagE* gene flanked by *Cla*I and *Bgl*II restriction sites. Strains containing plasmids derived from pBR322 were grown in the presence of 100 μg/mL ampicillin. Enzyme II^{nag} activity was measured in cells grown on nutrient broth with 0.2% NAG. Plasmid enrichments were carried out by using 200 μg/mL chloramphenicol (Maniatis et al., 1982). M13-derived phage were grown by using the recommended strains and procedures (Messing, 1983).

PTS Assays. Assays were carried out according to the general methods previously described (Waygood et al., 1979). Specific conditions for enzyme II^{nag} were 1 mM [^{14}C]NAG, sp act. 10000cpm/nmol, and 5 μM HPr; those for enzyme II^{mannose} were 1 mM 2-deoxy[^{14}C(U)]glucose, sp act. 10000 cpm/nmol, and 5 μM HPr.

Table I: Enzyme II^{nag} and Enzyme II^{man} Activities

strain and relevant genotype	enzyme II sp act.	
	N-acetylglucosamine[a]	mannose[a]
ZSC17 (wild type)	11.8	27.5
ZSC114 (*ptsM*)	13.0	1.1
KPN14 (*ptsM nagE*)	0.6	0.7
KPN16 (*ptsM nagE*)	1.5	1.6
pLC5-21/KPN14	37.2	2.4
pLC5-21/KPN16	28.0	1.8

[a] In units of nmol min^{-1} (mg of membrane protein)$^{-1}$.

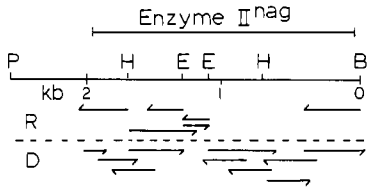

FIGURE 1: Sequencing of *nagE*. The sequence of the *nagE* gene was determined by using DNA fragments (R) generated from restriction sites shown. Sequencing from fragments generated by exonuclease III deletions was as indicated (D). The restriction endonuclease sites were as follows: P, *Pst*I; H, *Hin*cII; E, *Eco*RI; B, *Bgl*II.

Protein Determination. Protein was determined by the microbiuret method using bovine serum albumin as the standard (Layne, 1957).

DNA Sequencing. Sequencing was carried out by using the dideoxynucleotide chain-termination method of Sanger et al. (1977), using DNA polymerase (Klenow fragment), and in duplicate by the Sequenase methodology. Initial sequencing was on restriction fragments generated as shown in Figure 1. Overlapping fragments and smaller fragments were obtained by the generation of deletions using exonuclease III (Henikoff, 1984). All fragments were propagated in M13-derived phage for the purposes of sequencing.

Computer Programs. Sequence homologies, codon utilization, and sequence alignments were investigated by using a Beckman Microgenie program. The hydropathy plot was produced by using an IBI program.

Results

Cloning and Production of Enzyme II^{nag}. Plasmid pLC5-21, which contains the *nag* operon, restored the ability of *nagE* strains to ferment NAG. Plasmid pKP1 contained a 5.5-kb insert, and when this was subcloned, plasmid pKP1.3 contained the *nagE* gene on a 3.2-kb insert (details to be published elsewhere). The activity of enzyme II^{nag} in wild-type cells and plasmid-containing strains grown in the presence of NAG is shown in Table I. In *nagE* strains containing the plasmid pLC5-21, enzyme II^{nag} was found to be about 3-fold higher than wild type. Some activity was present in the supernatant fraction, and this remains to be characterized.

DNA and Amino Acid Sequence. The DNA insert in pKP1.3 was sequenced by the dideoxynucleotide chain-termination method of Sanger et al. (1977). Both strands were sequenced and the overlaps are shown in Figure 1. The DNA sequence is shown in Figure 2, and preceding the N-terminal methionine codon, a putative ribosomal binding site was found. The next methionine codon in the reading frame is at methionine-96. Two potential −35 and −10 RNA polymerase promoter sequences are identified in Figure 2. One of them is part of a catabolite gene activator protein (CAP) binding site. The intergenic region between *nagB* and *nagE* is 333 bp, and transcription of these two genes occurs in opposite directions (Peri and Waygood, unpublished results). The C-terminal amino acid codon is followed by one nonsense codon

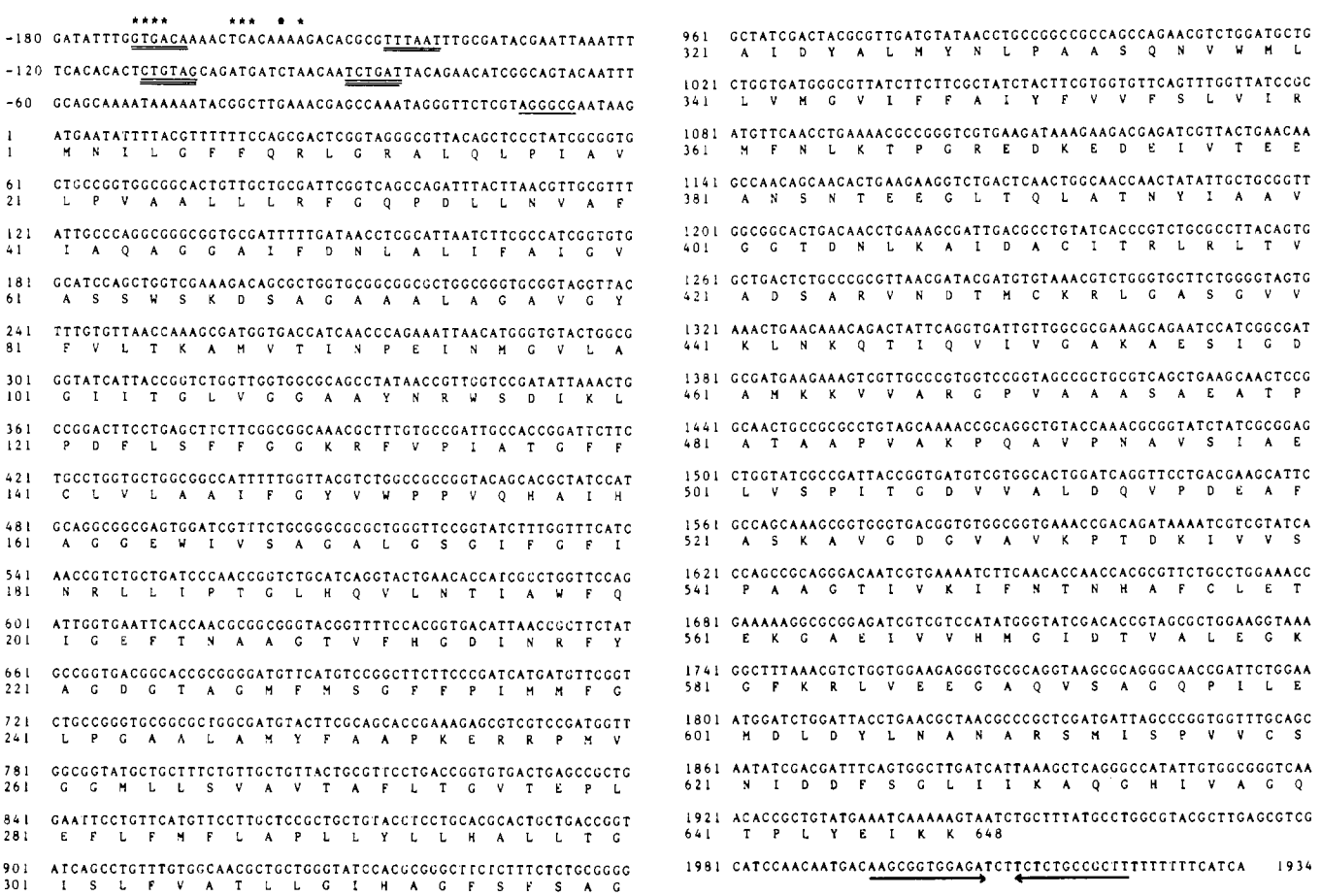

FIGURE 2: DNA sequence of *nagE*. The DNA sequence for *nagE* is presented. The first residue is about the middle of the intergenic region between *nagE* and *nagB*. The following putative regulatory regions are indicated: ribosomal binding site (single underline); RNA polymerase −35 and −10 promotor sites (double and triple underlines, respectively); CAP binding site homology (*); rho-independent terminator (← →).

and a putative rho-independent termination site. The C-terminal amino acid is further defined because there is considerable homology in the C-terminal amino acids of enzyme II^{nag} to most of III^{glc} including residues leading to the C-terminal residue of III^{glc} (Figure 3) (see below).

The codon utilization of enzyme II^{nag} was compared to that found for the genes of other PTS proteins and *E. coli* proteins (Ikemura, 1981). Codon utilization for enzyme II^{nag} resembles that of enzymes II^{glc} and II^{mtl} and does not display the propensity for low utilization codons found in the genes for enzymes II^{bgl} (Bramley & Kornberg, 1987a; Schentz et al., 1987) and II^{scr} (Ebner & Lengler 1987). The amino acid sequence (648 aa) is shown in Figure 2. The amino acid composition has 39% hydrophobic amino acids. The protein obtained from this sequence has a molecular weight of 68 356, which, for a membrane-bound protein, is in good agreement with the M_r 65 000 determined by sodium dodecyl sulfate/polyacrylamide gel electrophoresis of the phosphorylated protein (Waygood et al., 1984).

A hydropathy plot (Kyte & Doolittle, 1982) generated from the amino acid sequence revealed a protein that is comprised of a hydrophobic region (0–350 aa) and an essentially neutral region (350–648 aa) as shown in Figure 4. The enzyme II^{nag} sequence from 0 to 350 residues may contain as many as nine protential transmembrane helices; however, the enzyme II^{glc} hydropathy plot (Erni & Zanolari, 1986) indicated seven helices as shown in Figure 4.

The N-terminal sequence of enzyme II^{nag} could form an amphipathic helix (Figure 5). This is similar to enzyme II^{glc} (Erni & Zanolari, 1986) although the sequence homology is poor. Yamada and Saier (1987) found similar helices in enzymes II^{gut} and II^{mtl}. By analogy to melittin, such sequences are considered to be membrane-seeking (Eisenberg, 1984).

Table II: Enzyme II^{sugar} Sequences Containing Histidinyl Residues Homologous to the Active Site Histidine in III^{glc}

Protein	Sequence [a]
	91
III^{glc}	I E S D S G I E L F V H F G I D T V
	569
II^{nag}	L E T E K G A E I V V H M G I D T V
	547
II^{bgl}	I E S D D G V E I L I H V G I D T V

[a] Sequences in the C-terminal region of enzymes II^{nag} and enzyme II^{bgl} (Bramely & Kornberg, 1987a,b; Schnetz et al., 1987) are compared to the active site histidinyl residue of III^{glc} (Dorschug et al., 1984; Nelson et al., 1983). The histidinyl residue position is numbered.

Sequence Homologies to Other Enzymes II^{sugar}. Enzyme II^{nag} has extensive sequence homology (44% identity or 64% similarity) with enzymes II^{glc} and III^{glc} (Figures 3 and 4) such that it is essentially an enzyme that is a fusion of enzyme II and phosphocarrier protein. A similar homology to III^{glc} was evident in the sequence of enzyme II^{bgl} (Bramley & Kornberg, 1987a; Schentz et al., 1987) and has been reported (Bramley & Kornberg, 1987b). Both enzymes II^{nag} and II^{bgl} contain histidinyl residues homologous to the active site histidine of III^{glc} (Table II). Ebner and Lengeler (1987) have shown that enzyme II^{scr}, which interacts with III^{glc}, has extensive homology

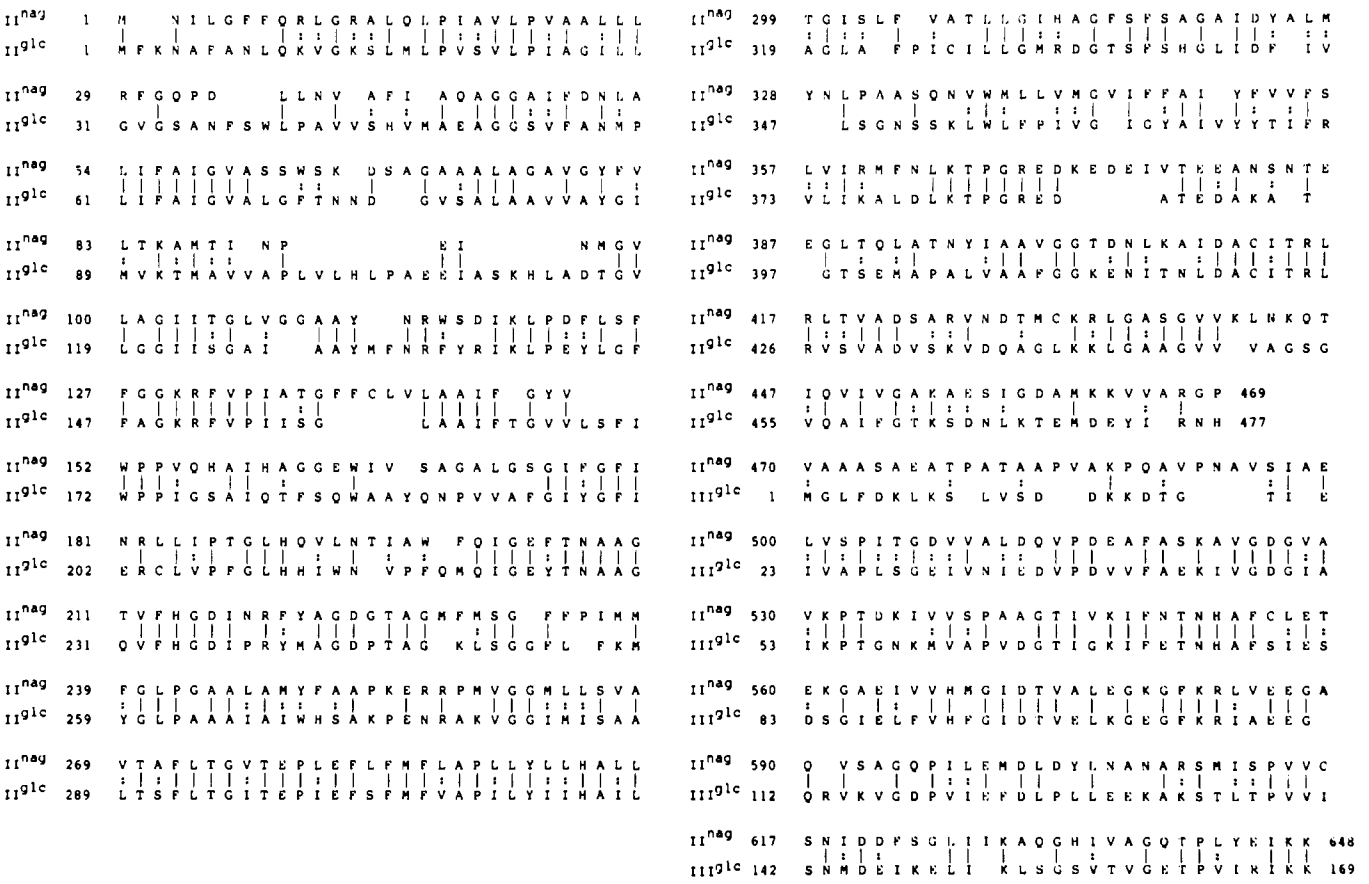

FIGURE 3: Amino acid sequence homology between enzyme II^{nag} and enzyme II^{glc} with III^{glc}. The homologies were generated with the aid of the alignment by similarity algorithm of Dayhoff (1972). (|) marks identities, (:) marks similarities.

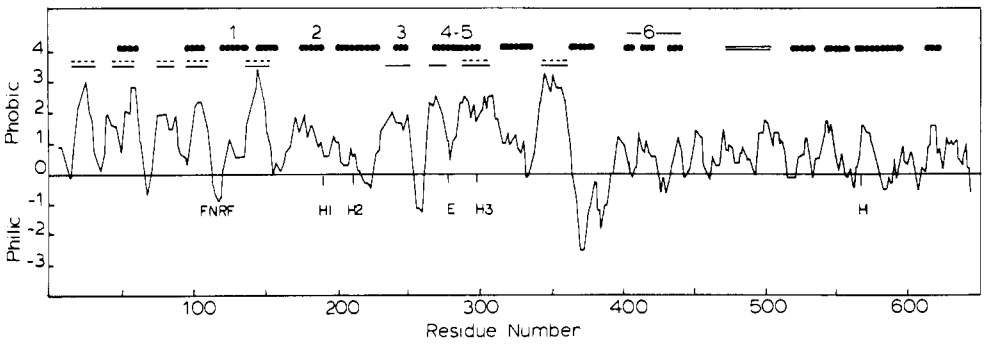

FIGURE 4: Hydropathy plot of enzyme II^{nag}. The hydropathy plot was generated according to the method of Kyte and Doolittle (1982). The values were generated from a window of nine amino acids, and the data are compressed 2-fold with respect to length. Potential transmembrane sequences are identified for enzyme II^{nag} (—), and the equivalent sequences in enzyme II^{glc} are (---). The following are identified: sequences in enzyme II^{nag} with 75% amino acid homology to enzyme II^{glc} and III^{glc} (●●●); the conserved histidine residues, H1 to H3 (Table III); the III^{glc} active site histidine, H; the conserved glutamate in sequence 4 (Table IV), E; the conserved sequences in Table IV, labeled 1–6; the -FNRF- sequence (Table IV); and the hinge region, Table V (═).

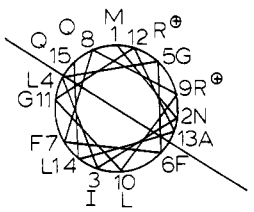

FIGURE 5: N-Terminal amphipathic sequence. The helical wheel representation of the N-terminal sequnce of enzyme II^{nag}.

to enzyme II^{bgl}. Attempts have been made to show homology between these four enzyme systems as well as with other enzymes II (Bramley & Kornberg, 1987b; Ebner & Lengeler, 1987; Saier et al., 1988). The results indicate that it is difficult to align the enzyme II^{sugar} sequences as the homologies are poor except for the pairs of enzymes II (II^{glc}/II^{nag} and II^{scr}/II^{bgl}). However, Bramley and Kornberg (1987b) identified a sequence near the C-terminal of enzyme II^{glc} that had homology to the N-terminal sequence of enzyme II^{bgl} (see sequence 6 in Table IV). Ebner and Lengeler (1987) found a similar sequence at the N-terminal of enzyme II^{scr} from *E. coli* and *Bacillus subtilis*, and in enzyme II^{nag} this sequence is present at the same position as found in enzyme II^{glc}.

In order to identify other sequence homologies, the following procedure was used. Amino acid sequence homologies between enzymes II^{glc} and II^{nag} were identified by alignment (Figure 4). Homologies between enzyme II^{scr} and II^{bgl} were found separately. Similarities between the two sets of homologies were found by inspection. Then other enzymes II were searched for sequences similar to those identified in the above

Table III: Conserved Enzyme II Sequences Containing Histidine

Enzyme II[a]	Site 1	Site 2	Site 3
	190	213	295
nag	L L I P T G L H Q V L N T	A A G T V P H G D I N F Y	P L L Y L L H A L L T G
	211	235	315
glc	C L V P F G L H H I W N V	A A G Q V F H G D I P R Y	P I L Y I I H A I L A G
	306	183	183
bgl	I F V M F G L H W G L V P	I G G A L V H P L I L T	G G A L V H P L I L T
	308	187	187
scrA	V I V I T G I H H S F H A	L G G I L T H P A L T N	G G I L T H P A L T N
	309	188	188
scrP	L I V L T G V H H S F H A	I G G I M I H P N L L N	G G I M I H P N L L N
	256		
mtl	I H F L G G I H E I Y F P		

[a] Sequences for enzymes II were obtained from the following references: enzyme II^{glc}, Erni and Zanolari (1986); enzyme II^{bgl}, Bramely and Kornberg (1987a), Schnetz et al. (1987); enzyme II^{scrA} from *E. coli* and enzyme II^{scrP} from *B. subtilis*, Ebner and Lengeler (1987); enzyme II^{mtl}, Lee and Saier (1983). The histidinyl residue is numbered.

Table IV: Some Conserved Sequences in Enzymes II[a]

		Sequence 1		Sequence 2
		* *		(Is site 1, Table III)
nag	120	L P D F L G F F A G K R F V P I A T G	132	
glc	140	L P E Y L G F F A G K R F V P I I S G	152	
bgl	155	F P I I L G Y T A G K R F G G N P	169	
scrA	159	L P I L I G F T A A R E F G G N P	173	
scrP	160	L P I L I G V S A S K E F G S N P	175	
mtl	60	L P L L I G Y T G G K L V G G E R G G	76	

		Sequence 3		Sequence 4	
		*		* * * *	
nag	242	P G A A L A M Y F	250	266 S V A V T A F L T G V T E P L E F	282
glc	262	P A A A I A I W	269	286 S A A L T S F L T G I T E P I E F	302
bgl	338	V G A A L G V F L	345	359 G S A A L T S L F G I T E P A V Y	373
scrA	341	G G A C L A V W F	349	362 P S A F A M L G I T E A A I F	376
scrP	339	G G A G L A V F F	347	360 P A A F A F L G I T E P V I F	376

		Sequence 5		Sequence 6	
		(Is site 3, Table III)		** * **** * * * *	
nag	391	QLATNYIAAVGGTDNLKAIDACITRLRLTVAD SARVND TMCKRL GASGVVKLNKQT IQIQVIVGA	453		
glc	400	EMAPALVAAFGGKENITNLDACITRLRVSVADVS KVDQ AGLKKL GAAGVVVAG SG VQAIFGT	461		
bgl	3	ELARKIVAGVGGADNIVSLMHCATRLRFKLKDES KAQAE VLKKTPGIIMVV ESGGQFQVVIGN	65		
scrA	5	QISCSLLPLLGGKENIASA ALATRLRLVLVDDSL ADQQA IGKVEGVKGCF RNAG QMQIIFGT	67		
scrP	5	ETAKRLIELLGGKENIISAAHCATRLRLVMKDES KIDQAQ VEELDGVKGAF SSSGQYQIIFGT	68		

[a] References for sequences are in Table III footnote. (*) indicates conserved residues.

Table V: Potential Hinge Sequences

PTS Component		Sequence[a]	
III^{man}	133	P V E K A A A P A P A A A A P K A A P T P A K P M G P	158
II^{nag}	469	G P V A A A S A E A P A T A A P V A K P Q A V P N A V S I E L V S P	504
II^{bgl}	458	P A Q G A P Q E K T P E V I T P P E Q G G I C S P	482

[a] III^{man} sequence from Erni et al. (1987).

comparison. In addition, homologies in the DNA sequences were examined by computer methods using various degrees of stringency.

In the comparison between enzymes II^{glc} and II^{nag}, three conserved histidinyl residues were found (Table III), other than the active site histidinyl residue of III^{glc} (Table II). Comparison between enzymes II^{bgl} and II^{scr} revealed two conserved histidinyl residues at about residues 185 and 307 (Table III). It is apparent that while convincing similarities exist within each pair, the sequence similarities become poor when comparisons are made with other enzymes II. The alignment shown as site 1 (Table III) was influenced by the -G(L/I)H- sequence. Site-directed mutagenesis of enzyme II^{bgl} His-306, which is aligned as site 1 (Table III), produced an enzyme II^{bgl} incapable of supporting sugar fermentation [see Saier et al. (1988)]. The site 2 (Table III) alignment is the most highly conserved histidine-containing sequence in enzymes II^{glc} and II^{nag} (8/13 identical residues). This is also the only histidine-containing sequence that has high DNA sequence conservation (75% in 20 nucleotides). Unfortunately, both site 2 and site 3 of enzymes II^{glc} and II^{nag} can be aligned with the second conserved histidine in enzymes II^{bgl} and II^{scr} such that it is difficult to assign a preference on the basis of sequence homologies.

A few other sequence homologies (similarities) were identified, and these are shown in Table IV. Sequence 1 has either a lysyl residue or an arginyl residue in a hydrophilic sequence with glycyl and prolyl residues, which probably forms a loop between transmembrane-spanning regions (Figure 4). Sequence 4 contains the only conserved glutamate residue in the enzyme II sequences excluding the III-like portions. Bramley and Kornberg (1987b) have aligned Glu-257 of enzyme II^{mtl} with this sequence. However this is also part of the site 1 histidine alignment (Table III). Sequence 4 may also be located on a loop (Figure 4). In enzymes II^{bgl} and II^{scr}, sequence 4 is associated with a hydrophilic region at residues 360–370 (Ebner & Lengeler, 1987). Sequence 6 (Table IV) contains a conserved aspartyl residue and two conserved arginyl residues. This sequence is after the major hydrophilic region (residues 360–385) in enzymes II^{glc} and II^{nag} and is clearly part of a large nonhydrophobic domain (Figure 4). Sequence 6 is the transposition identified by others in enzymes II^{bgl} and II^{scr} (Bramley & Kornberg, 1987b; Ebner & Lengeler, 1987).

DNA Homologies. DNA homologies should be reflected in amino acid homologies. These homologies were examined by using a restricted window, 75% identity in sequences of 20 nucleotides. This produced the following results: (1) A fifth of the DNA of enzymes II^{nag} and II^{glc} is conserved. (2) This conserved DNA includes sequences 1, 4, and 5 of Table IV and only site 2 of Table III. (3) There is no DNA sequence homology to III^{glc} using this window for either enzyme II^{nag} or enzyme II^{bgl}. (4) There is however some sequence homology between the III-like parts of enzymes II^{nag} and II^{bgl}. (5) There are a few sequence homologies between other enzymes II, but these involve mainly hydrophobic sequences (see below).

A less restricted window (50% identity in 50 nucleotides) gave almost complete homology between enzyme II^{nag} and enzyme II^{glc}/III^{glc}, except for about 200 nucleotides that correspond to the amino acid sequence that joins the enzyme II and III parts. However, this potential hinge region in enzymes II^{nag} and II^{bgl} was identified as homologous by their DNA sequences. The amino acid sequences of the hinge regions are given in Table V along with the potential hinge region identified in III^{man} (Erni et al., 1987). Conservation of DNA sequences in enzymes II^{bgl} and II^{scr} is limited to 156 nucleotides that code for sequences 1–2 (Table IV) and 240 nucleotides containing sequences 3–5 (Table IV). Thus enzymes II^{bgl} and II^{scr} are not as closely related as enzymes II^{glc} and II^{nag} are to each other. Comparison of enzymes II^{scr} and II^{glc} indicated that the DNA sequence which codes for amino acid residues 60–140 (approximately) was conserved. How-

Table VI: -FNRF- Sequence

Enzyme II^{sugar}	Sequence [a]
	114
II^{nag}	A A Y N R W S D I
	134
II^{glc}	Y M F N R F Y R I
	99
II^{bgl}	N L L N R F V Y V
	103
II^{scrA}	N P F Q R I A R L
	104
II^{scrP}	N P A A R F A K T
	115
II^{mtl}	K H F D R W V D G

[a] The arginyl residue is numbered. References given in Table III footnote.

ever, the amino acid sequences are dissimilar except for the -FNRF- sequence shown in Table VI. This sequence, once identified, was found in a number of enzymes II and contains a conserved arginyl residue on a probable loop (Figure 4). The analysis of DNA sequences of enzymes II^{nag} and II^{glc} with enzymes II^{scr} and II^{bgl} does not identify the major transposition (sequence 6, Table IV).

Extension of this type of analysis to the other sugar-specific PTS components, enzyme II^{mtl}, enzymes II^{gut} and III^{gut}, and enzymes II^{man} (*manM* and *manP*) and III^{man}, coupled with amino acid homology searches produced results that were remarkable for the paucity of identities. Only a few observations were made: (1) In general, amino acid comparisons identified homologies in sequences rich in prolines, glycines, alanines, valines, isoleucines, and leucines (e.g., sequence 3, Table IV). Presumably, these represent a structural motif required for transmembrane segments. However, such sequences were neither long nor very similar. (2) The hinge region in enzyme II^{nag} shares some homology with the potential hinge region of III^{man} (Table V). (3) Enzyme II^{mtl} has sequence similarity for sequence 1 (Table IV) in particular when compared to enzymes II^{bgl} and II^{glc}.

Sequence Homologies to Other Proteins. Erni and Zanolari (1986) found few sequence homologies for enzyme II^{glc} to other proteins, and this is the case for enzyme II^{nag}. However, enzyme II^{nag} sequence 475–491, which contains three prolines and six alanines, shows 60–65% homology to several eukaryotic viral proteins as does the leucine-rich enzyme II^{nag} sequence 280–314.

Discussion

The gene, *nagE*, which codes for enzyme II^{nag}, has been sequenced. The expression of this gene may occur through both cAMP-dependent and independent mechanisms. Two possible promoter sequences and a CAP binding site can be identified prior to the open reading frame. However, the same CAP binding site may be involved with the regulation of *nagB* which is transcribed in a direction opposite to that of *nagE* (Peri and Waygood, unpublished results). The *nagE* gene has been considered to be part of a *nag* operon (Postma & Lengeler, 1984), but it is transcribed separately from *nagB* and *nagA*, suggesting that the genes form a regulon (Peri and Waygood, unpublished results). The *nagE* gene ends with a putative rho-independent terminator that has been identified previously (Yamao et al., 1982), thus placing *nagE* next to *glnS* on the *E. coli* chromosome. This is a correction of the current map and confirms the gene order reported by Plumbridge (1987).

The amino acid sequence of enzyme II^{nag} has 44% homology to enzyme II^{glc} and III^{glc} and is clearly a fusion of these two proteins. This is similar to the arrangement described for enzymes II^{bgl} and II^{scr} (Bramley & Kornberg, 1987b; Ebner & Lengeler, 1987) and confirms the proposal of Saier et al. (1985) that enzymes II with molecular weights of about 65 000 would contain an enzyme II like part and a III-like part. Implicit in the model presented by Saier et al. (1985) was the proposal that such an enzyme would have two sites of phosphorylation equivalent to the active sites of the distinct parts. In both enzymes II^{nag} and II^{bgl} (Bramley & Kornberg, 1987b), the active site histidine of III^{glc} is contained in a highly conserved sequence (Table II). This active site histidine in III^{glc} is a 3-P-histidine (Meadow & Roseman, 1982), and a similar phosphorylation was identified for enzyme II^{nag} (Waygood et al., 1984). The second active site histidine would be located in the enzyme II part, and it has been suggested that this would be a 1-P-histidine (Saier et al., 1985). Peri et al. (1984) identified a phosphorylation in enzyme II^{glc} that behaves as a 1-P-histidine. For these reasons, considerable attention has been paid to conserved histidinyl residues in various enzymes II that are now sequenced (Bramley & Kornberg, 1987a,b; Ebner & Lengeler, 1987; Saier et al., 1988). In addition, it has been thought that the enzymes II should show significant sequence homologies. However, it appears that the conservation of sequences in all enzymes II is very limited. The sequences in Table IV represent about 110 amino acids out of about 470 in an enzyme II sequence. Despite efforts to show sequence homologies between the various enzymes II, the only convincing alignments can be made between enzymes II^{nag} and II^{glc} with III^{glc} (Figures 3 and 4), and between enzymes II^{bgl} and II^{scr} with III^{glc} (Ebner & Lengeler, 1987). The sequence transposition identified by Bramley and Kornberg (1987b), in which the sequence near the C-terminal of enzyme II^{glc} is found at the N-terminal of enzyme II^{bgl}, can also be identified in enzymes II^{nag} and II^{scr} (sequence 6, Table IV). As enzymes II^{nag}, II^{glc}, II^{bgl}, and II^{scr} all interact with III^{glc}, either as a separate protein or within their own structure, it would seem plausible that these four enzymes should have some common components in the enzyme II active site.

A considerable effort has been made to identify active site histidines and similarities in all of the enzymes II (Saier et al., 1988). We have tried a somewhat different approach which is based upon the following. First, the enzymes II^{glc}, II^{nag}, II^{bgl}, and II^{scr} should have some similarity due to the reasons given above. Second, our work on the tertiary structure of HPr has implicated the involvement of acidic (glutamate) and basic (arginine) residues in the formation or phosphoryl-transfer mechanism of the 1-P-histidine present in HPr (Waygood et al., 1985; Klevit & Waygood, 1986; Bhanot et al., 1987; El Kabbani et al., 1987). Third, because membrane proteins like enzyme II are transmembrane, the orientation of a conserved sequence with respect to the side of the membrane is of considerable importance.

In this paper a few sequences have been identified in the enzymes II that have similarities to each other (Tables III–VI). Hydropathy plots (Figure 4; Lee & Saier, 1983; Erni & Zanolari, 1986; Ebner & Lengeler, 1987) can be used to tentatively assign transmembrane sequences. The folding of enzyme

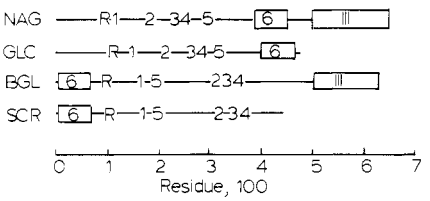

FIGURE 6: Ordering of conserved sequences in enzymes II^{sugar}. The position of the numbered (1–6) conserved sequences in Table IV and the -FNRF- sequence (Table VI), labeled R, are shown for enzymes II^{nag}, II^{glc}, II^{bgl}, and II^{scr}. The III box represents the III^{glc} homologous region.

II^{mtl} into seven transmembrane sequences as suggested by Lee and Saier (1983) is compatible with the assigned transmembrane sequences for enzyme II^{glc} (Figure 4). Given the high degree of sequence conservation between enzymes II^{glc} and II^{nag}, a similar folding pattern is suggested, even though as many as nine transmembrane sequences can be assigned to enzymes II^{nag}. This places the III-like portions of enzyme II^{mtl} and enzyme II^{nag} on the cytoplasmic side of the membrane. There is good evidence for this in enzyme II^{mtl} (Stephen & Jacobson, 1986). This folding pattern results in the site 1 histidine (Table III), sequence 4 (Table IV) containing a conserved glutamate, and the major transposition sequence 6 (Table IV) all being cytoplasmically oriented. Sequence 1 (Table IV) and the -FNRF- sequence (Table V) are oriented to the exterior in enzymes II^{glc} and II^{nag}. However, the -FNRF- sequence in enzyme II^{mtl} is located on the cytoplasmic side. Because of the transposition of several sequences between the enzymes II (Figure 6), the folding pattern of enzymes II^{bgl} and II^{scr} may be different as shown in Figure 7. It proved difficult to fold enzymes II^{bgl} and II^{scr} in a similar manner to the other enzymes II, while maintaining some compatibility with the hydropathy plot data. Preservation of the C-terminal III-like region and the conserved site 1 histidine (H1) and sequence 4 (E) on the cytoplasmic side suggests two transmembrane helices between residues 375 and 460. The location of sequence 6 (transposition) at the cytoplasmic side requires an even number of transmembrane segments between residues 100 and 280. This causes sequence 1 (K) to be cytoplasmically oriented. However, the assignment of the second conserved histidine of enzymes II^{bgl} and II^{scr} to the site 3 histidine (Table III) is preferred because the histidine is then located in the same position relative to the membrane as is found in the other enzymes II (Figure 7).

The inability to arrive at a consensus location for the -FNRF- sequence suggests that this sequence may not be a true homology. The inability to arrive at a conserved location for sequence 1 (K) is more problematic because it is one of the better conserved sequences. The location of conserved residues like sequence 1 (K) and the site 3 histidine on or toward the exterior suggests that these may be involved in some common mechanism of sugar recognition and/or translocation. The cytoplasmically oriented residues like site 1, histidine, sequence 4 (E), and sequence 6 (R, R, and D residues) may have a role in the phosphoryl-transfer reaction. In particular, those enzymes II that interact with a soluble III^{sugar} should have the phosphorylation site readily accessible to the cytoplasm.

Our analysis of histidinyl residues is in essential agreement with that of Saier et al. (1988) except for our assignment of His-256 to the active site histidine in enzyme II^{mtl}. Site-directed mutagenesis of Gly-253 does inactivate enzyme II^{mtl} (Saier et al., 1988), providing some support for the assignment of histidine-256 (Table III). It is apparent that enzyme II^{man}/III^{man} is very different from the other enzymes II (Saier et al., 1988) and that enzyme II^{gut}/III^{gut} may only have similarities to enzyme II^{mtl} (Yamada & Saier, 1987). There are some similarities between enzyme II^{mtl} and the four III^{glc}-dependent enzymes II (Tables IV and VI). The mannose and glucitol PTS components were investigated, and no compelling similarities could be detected.

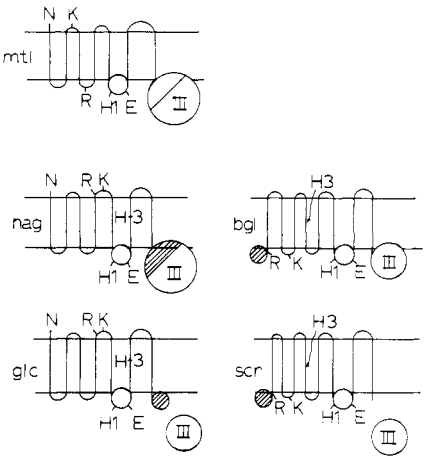

FIGURE 7: Possible structures for enzymes II. The enzyme II^{mtl} structure (mtl) is that previously proposed by Lee and Saier (1983). Enzyme II^{glc} structure (glc) is derived from the transmembrane segments identified by Erni and Zanolari (1986). Enzyme II^{nag} structure (nag) is derived from the hydropathy plot (Figure 4), but the number of potential transmembrane sequences is reduced in order to arrive at a structure that is similar to that of enzyme II^{glc}. The enzymes II^{bgl} and II^{scr} structures (bgl and scr, respectively) are derived from the hydropathy plots of Ebner and Lengeler (1987) and are influenced by the location of conserved residues. The location of identified conserved sequences is shown as follows: site 1 and site 3 histidines (Table III), H1 and H3; sequence 1 (Table IV) lysine, K; sequence 4 (Table IV) glutamate, E; sequence 6 (Table IV) transposition, crosshatched area; -FNRF- sequence (Table VI), R; III^{glc}, III, except in enzyme II^{mtl} where it is a III-like portion (Saier et al., 1988).

While the obvious sequence homologies that exist between enzymes II^{nag} and II^{glc}/III^{glc}, and between enzymes II^{bgl} and II^{scr}/III^{glc} (Ebner & Lengeler, 1987), confirm some of the proposals made by Saier et al. (1985) concerning the evolutionary relationships within the PTS, extension from these pairs to larger groups of PTS proteins has proved difficult. The addition of the sequence of enzyme II^{nag} to the accumulated data has helped to delineate some sequences that may contain important functional residues. Clearly, much work on both the individual residues identified and the topographical arrangements must be carried out to confirm the suggested roles.

ACKNOWLEDGMENTS

We thank Drs. M. H. Saier, Jr., M. Yamada, J. Lengeler, H. L. Kornberg, and H. Bramley for preprints of their papers and access to their sequence data prior to publication. We thank Dr. J. Weiner for providing the Clarke–Carbon DNA library, Dr. H. Goldie for his advice on cloning protocols, and Dr. K. Adachi for his help and advice on sequencing. We thank Dr. P. Bhanot for his comments on the sequence homologies and many useful discussions. In particular, we thank Dr. J. Rodger and Dr. D. Soll for providing their sequence of *nagE* prior to publication. Comparison of the two sequences led to minor corrections in both.

REFERENCES

Adler, J., & Epstein, W. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2895–2899.

Begley, G. S., Hanson, D. E., Jacobson, G. R., & Knowles, J. R. (1982) *Biochemistry 21*, 5552–5556.

Bhanot, P., Georges, F., & Waygood, E. B. (1987) *Proced. Can. Fed. Biol. Soc. 30*, 152.

Bramley, H. F., & Kornberg, H. L. (1987a) *J. Gen. Microbiol. 133*, 563–573.
Bramley, H. F., & Kornberg, H. L. (1987b) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4777–4780.
Clarke, L., & Carbon, J. (1976) *Cell (Cambridge, Mass.) 9*, 91–99.
Curtis, S. J., & Epstein, W. (1975) *J. Bacteriol. 122*, 1189–1199.
Dayhoff, M. O. (1972) *Atlas of Protein Structure*, National Biomedical Research Foundation, Washington, DC.
Dorschug, M., Frank, R., Kalbitzer, H. R., Hengstenberg, W., & Deutscher, J. (1984) *Eur. J. Biochem. 144*, 113–119.
Ebner, R., & Lengeler, J. W. (1987) *Mol. Microbiol. 2*, 9–17.
Eisenberg, D. (1984) *Annu. Rev. Biochem. 53*, 595–623.
El-Kabbani, O. A. L., Waygood, E. B., & Delbaere, L. T. J. (1987) *J. Biol. Chem. 262*, 12926–12929.
Erni, B., & Zanolari, B. (1986) *J. Biol. Chem. 261*, 16398–16403.
Erni, B., Zanolari, B., & Kocher, H. P. (1987) *J. Biol. Chem. 262*, 5238–5247.
Henikoff, S. (1984) *Gene 28*, 351–359.
Ikemura, T. (1981) *J. Mol. Biol. 151*, 389–409.
Jacobson, G. R., Kelly, D. M., & Finlay, D. R. (1983a) *J. Biol. Chem. 258*, 2955–2959.
Jacobson, G. R., Lee, C. A., Leonard, J. E., & Saier, M. H., Jr. (1983b) *J. Biol. Chem. 258*, 10748–10756.
Klevit, R. E., & Waygood, E. B. (1986) *Biochemistry 25*, 7774–7781.
Kundig, W., Ghosh, S., & Roseman, S. (1964) *Proc. Natl. Acad. Sci. U.S.A. 52*, 1067–1074.
Kyte, J., & Doolittle, R. G. (1982) *J. Mol. Biol. 157*, 105–132.
Layne, E. (1957) *Methods Enzymol. 3*, 447–453.
Lee, C. A., & Saier, M. H., Jr. (1983) *J. Biol. Chem. 250*, 10761–10767.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Messing, J. (1983) *Methods Enzymol. 101*, 20–78.
Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Nelson, S. O., Schuitema, A. R. J., Renne, R., van der Ploeg, L. H. T., Plijter, J. J., Aan, F., & Postma, P. W. (1984) *EMBO J. 3*, 1587–1593.
Peri, K. G., Kornberg, H. L., & Waygood, E. B. (1984) *FEBS Lett. 170*, 55–58.
Plumbridge, J. (1987) *Mol. Gen. Genet. 209*, 618–620.
Postma, P. W., & Lengeler, J. W. (1985) *Microbiol. Rev. 49*, 232–269.
Saier, M. H., Jr., Grenier, F. C., Lee, C. A., & Waygood, E. B. (1985) *J. Cell. Biochem. 27*, 43–56.
Saier, M. H., Jr., Yamada, M., Erni, B., Suda, K., Lengeler, J., Ebner, R., Argos, P., Rak, B., Schnetz, K., Lee, C. A., Stewart, G. C., Breidt, F., Jr., Waygood, E. B., Peri, K. G., & Doolittle, R. F. (1988) *FASEB J. 2*, 199–208.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5477.
Schnetz, K., Toloczyki, C., & Rak, B. (1987) *J. Bacteriol. 169*, 2579–2590.
Stephen, M. M., & Jacobson, G. R. (1986) *Biochemistry 25*, 8230–8234.
Stock, J. B., Waygood, E. B., Meadow, N. D., Postma, P. W., & Roseman, S. (1982) *J. Biol. Chem. 257*, 14543–14552.
Waygood, E. B., & Steeves, T. (1980) *Can. J. Biochem. 58*, 40–48.
Waygood, E. B., Meadow, N. D., & Roseman, S. (1979) *Anal. Biochem. 95*, 293–304.
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*, 6938–6945.
White, R. J. (1970) *Biochem. J. 118*, 89–92.
Yamada, M., & Saier, M. H., Jr. (1987) *J. Biol. Chem. 262*, 5455.
Yamao, F., Inokuchi, H., Cheung, A., Ozeki, H., & Soll, D. (1982) *J. Biol. Chem. 257*, 11639–11643.