JOURNAL OF BACTERIOLOGY, Apr. 1989, p. 1923–1931
0021-9193/89/041923-09$02.00/0

Vol. 171, No. 4

Nucleotide Sequence of the Osmoregulatory *proU* Operon of *Escherichia coli*†

J. GOWRISHANKAR

Centre for Cellular and Molecular Biology, Hyderabad 500 007, India

Received 19 September 1988/Accepted 9 January 1989

The sequence of 4,362 nucleotides encompassing the *proU* operon of *Escherichia coli* was determined. Three open reading frames were identified whose orientation, order, location, and sizes were in close accord with genetic evidence for three cistrons (*proV*, *proW*, and *proX*) in this operon. Similarities in primary structure were observed between (i) the deduced sequence of ProV with membrane-associated components of other binding-protein-dependent transport systems, in the nucleotide-binding region of each of the latter proteins, and (ii) that of ProW with integral membrane components of the transport systems above. The DNA sequence data also conclusively established that ProX represents the periplasmic glycine betaine-binding protein. Two copies of repetitive extragenic palindromic sequences were identified beyond the 3′ end of the *proX* gene. The primer extension technique was used to identify the 5′ ends of *proU* mRNA species that are present in cells grown at high osmolarity; the results suggest that at least some of the osmotically induced *proU* transcripts have a long leader region, extending as much as 250 base pairs upstream of the *proV* gene. Evidence was also obtained for the existence of a sequence-directed bend in DNA in the upstream regulatory region of the *proU* operon.

The *proU* locus in *Escherichia coli* and *Salmonella typhimurium* encodes a transporter for active uptake of two solutes, glycine betaine and L-proline, whose intracellular accumulation is important in the process of water stress adaptation in these organisms (3, 6, 8, 13, 18–20, 39). The expression of *proU* is induced approximately 200-fold, and the transporter activity is also stimulated, upon growth of these bacteria in media of elevated osmolarity (6, 11, 13, 14, 18, 39). In both *E. coli* and *S. typhimurium*, a periplasmic glycine betaine-binding protein has been shown to be a product of the *proU* locus (3, 4, 14, 27, 39), indicating that the ProU transporter is one among the class of multicomponent binding-protein-dependent transport systems characterized in the enterobacteria (1, 19).

In the accompanying paper (11), C. S. Dattananda and I adduced genetic evidence for the presence of three genes (designated *proV*, *proW*, and *proX*) in the *proU* locus, organized in a single operon; their respective gene products were shown to be 44-, 35-, and 33-kilodalton proteins, the last of which was localized in the periplasm. In this paper, I present the nucleotide sequence of the *proU* locus along with results from experiments directed towards characterization of its *cis* regulatory region.

MATERIALS AND METHODS

Recombinant DNA and M13 phage techniques. The methods for restriction enzyme digestion, ligation, transformation, and gel electrophoresis of DNA fragments were those described by Maniatis et al. (38). Techniques for work with recombinant M13 phages and their host, JM101, for cloning and DNA sequence determination have been described (40).

Strategy for DNA sequence determination of *proU*. My colleagues and I had previously established that a 5-kilobase-pair (kb) segment of chromosomal DNA clockwise of a *Bgl*II site and extending up to the site of a mini-Mu phage insertion cloned in the plasmid pHYD58 (Fig. 1) encompasses the entire *proU* locus (20). The data of Bremer and co-workers (14, 39) and the results presented in the accompanying paper (11) suggested further that the *proU* operon is situated to the right of the *Eco*RV site shown in Fig. 1. The complete nucleotide sequence of the chromosomal region on pHYD58 to the right of the *Eco*RV site was, therefore, determined.

The strategy for this effort is described in the legend to Fig. 1. It entailed first the cloning into the polylinker region of the M13 phage vector tg131 (31) of three fragments from pHYD58: a 3.8-kb *Eco*RV-*Nsi*I fragment and two *Nsi*I fragments, 0.11 and 0.95 kb long (Fig. 1). The DNA sequence was then determined on both strands, either directly or after the generation of overlapping exonuclease III-generated deletions (22). The sequence across the two *Nsi*I sites in this region was determined after an *Rsa*I fragment from this region was cloned into tg131, as shown in the figure.

DNA sequence compilation and analysis. The computer program packages of Staden (50, 51) were used for compilation of the *proU* DNA sequence and for its analysis. Hydropathicity profiles of the deduced protein sequences were determined as described by Kyte and Doolittle (35), and the search for homologies within the protein sequence database of the Protein Identification Resource, National Biomedical Research Foundation (Washington, D.C.), was accomplished with the aid of the Lipman-Pearson algorithm (37).

5′-End mapping of mRNA. The method for determining the 5′ ends of mRNA was essentially that involving the technique of reverse transcriptase-directed primer extension on mRNA, described earlier (30). Radiolabeled single-stranded DNA probe primers were prepared on appropriate recombinant M13 phage templates by Klenow-directed extension from universal primer followed by restriction enzyme digestion; they were then purified after electrophoretic separation on urea-polyacrylamide gels. Each probe was hybridized under stringent conditions with total RNA isolated from (i)

† Dedicated to Pushpa M. Bhargava on his 60th birthday.

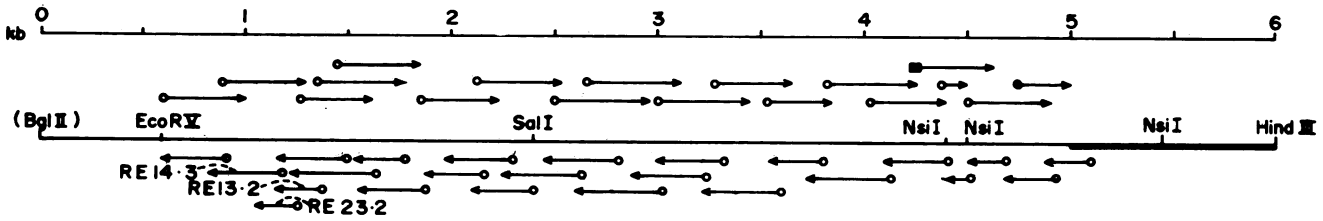

FIG. 1. Strategy for DNA sequence determination of *proU*. The map of insert DNA of pHYD58 (20) is shown, and relevant restriction sites are marked; a kilobase scale is included. The insert includes 5 kb of DNA clockwise of the *Bgl*II site from the *E. coli proU* locus (thin line) and 1 kb of Mu *c* DNA (thick line); the *Bgl*II site was lost in the process of construction of pHYD58 and is therefore shown within parentheses. The sequence to the right of the *Eco*RV site, marked at approximately 0.6 kb in the figure, was determined. An ordered, overlapping series of deletions starting from the end nearest the site of hybridization of universal primer (31) was generated with the aid of exonuclease III and S1 nuclease (22) in the 3.8-kb *Eco*RV-*Nsi*I fragment in each of two tg131 recombinant clones that had this fragment cloned in opposite orientations with respect to the M13 sequence. The open circles above and below the map indicate the deletion endpoints in the clones of the two sets, respectively, and the arrows leading from them delineate the direction and extent of sequence determined by the dideoxynucleotide chain-termination method (40) in each of these clones. Specific clones that were used for probe preparation in the primer extension experiment described in the text have been identified in the figure. The 0.11-kb *Nsi*I fragment was sequenced directly after it had been cloned in both orientations into tg131. The exonuclease III strategy was again followed in determining the sequence of the bottom strand on the 0.95-kb *Nsi*I fragment, whereas the top strand of this fragment was sequenced directly from the *Nsi*I end with the use of a synthetic oligonucleotide 17-mer primer complementary to the sequence in the region indicated by the solid circle, as shown. Sequencing across the two *Nsi*I sites (shown by the arrow leading from the solid square) was accomplished after an *Rsa*I fragment spanning this region was cloned in the appropriate orientation into tg131.

strain MC4100 (Δ*lac rpsL*) grown in half-strength minimal salts medium supplemented with 0.2% glucose and 0.5% Casamino Acids (low-osmolarity medium) or (ii) strain GJ157 (MC4100 Δ*putPA proP proX*::*lac*) grown in LB + 0.2 M NaCl (high-osmolarity medium). The hybridized probe was then extended on the RNA template with avian myeloblastosis virus reverse transcriptase (Bio-Rad). The products were run on a urea-polyacrylamide gel and sized against a sequence ladder generated with universal primer on the cognate M13 template DNA.

RESULTS

Nucleotide sequence of the *E. coli proU* operon. The sequence of the 4,362-nucleotide region of chromosome encompassing the *proU* operon was determined (Fig. 2). The end of this sequence marks the junction of chromosomal and Mu phage DNA in pHYD58, as indicated from a comparison of the sequence obtained in this study (data not shown) with that published of the Mu *c* region (47).

Deduced protein products of *proU* operon. Three long open reading frames were identified in the *proU* sequence, all on the same strand of DNA. The putative translation initiation site in each of them was localized on the basis of features expected of such sites in *E. coli* (17), and the inferred amino acid sequences of the three corresponding gene products are shown in Fig. 2. In view of the close correlation between the genetic data on *proU*, described earlier and in the accompanying paper (11, 18, 20), and the three gene products identified herein from the nucleotide sequence, I have designated these three gene products ProV, ProW, and ProX, respectively.

The translation of *proV* is shown in Fig. 2 to begin at nucleotide 689. This particular reading frame in fact remains open over an additional length of 276 upstream nucleotides, but no other site of translation initiation can be predicted in this upstream region from the empirical consensus rules that are in current use for this purpose (17); furthermore, the size of a truncated polypeptide obtained from a *proV*::Tn*1000* plasmid (11) is consistent with the translational initiation site marked here, and Faatz et al. (14) have shown that a Tn*5* insertion 0.55 kb downstream of the *Eco*RV site does not disrupt *proV*. ProV is predicted to be a 400-amino-acid-long polypeptide, relatively hydrophilic (Fig. 3), with an M_r of 44,162; interestingly, it is devoid of any tryptophanyl residues. The predicted *proV* coding sequence extends beyond the *Sal*I site at position 1810 for another 26 codons; consistent with this identification is the observation by my colleagues and myself in maxicell experiments that a plasmid (pHYD56 [20]) in which this *Sal*I end has been ligated with the *Sal*I site of pBR322 (so that the open reading frame terminates three codons downstream [38]) encodes a protein that is 2 kilodaltons smaller than the native ProV protein (K. Rajkumari, unpublished). The fact that pHYD56 is *proV*$^+$ in complementation experiments (20) would indicate that the C-terminal residues of the native protein are not essential for its function as a component of the betaine/L-proline transporter.

The inferred amino acid sequence of ProV shows significant similarity in two regions to HisP, a component of the L-histidine transporter of *S. typhimurium* (24) (Fig. 4). These same regions of HisP are in turn known to be homologous with corresponding regions in one component of each of the other binding-protein-dependent transport systems (1, 15, 26, 49) and also with several other ATP-binding proteins, in each of which they are believed to constitute the so-called fingers of a nucleotide-binding fold (25, 26, 56). Indeed, the expected similarity was also observed between ProV and each of these other proteins (data not shown).

ProW is deduced to be a hydrophobic polypeptide 354 amino acids long (M_r 37,619). There is an 8-nucleotide overlap between the end of *proV* and the start of *proW*, suggesting that there may be translational coupling in the expression of the two genes; a ribosome-binding site is also present upstream of the *proW* initiation site (Fig. 2), a feature that has been shown to be necessary for such coupling to operate in other pairs of genes (9).

A set of amino acid residues, which has previously been identified as conserved across the integral membrane components of binding-protein-dependent transport systems and situated approximately 90 residues from the C-terminal end in each of these proteins (10, 28), is also conserved in the ProW sequence (Fig. 5); furthermore, its location in ProW relative to the C-terminal end of the polypeptide, viewed in

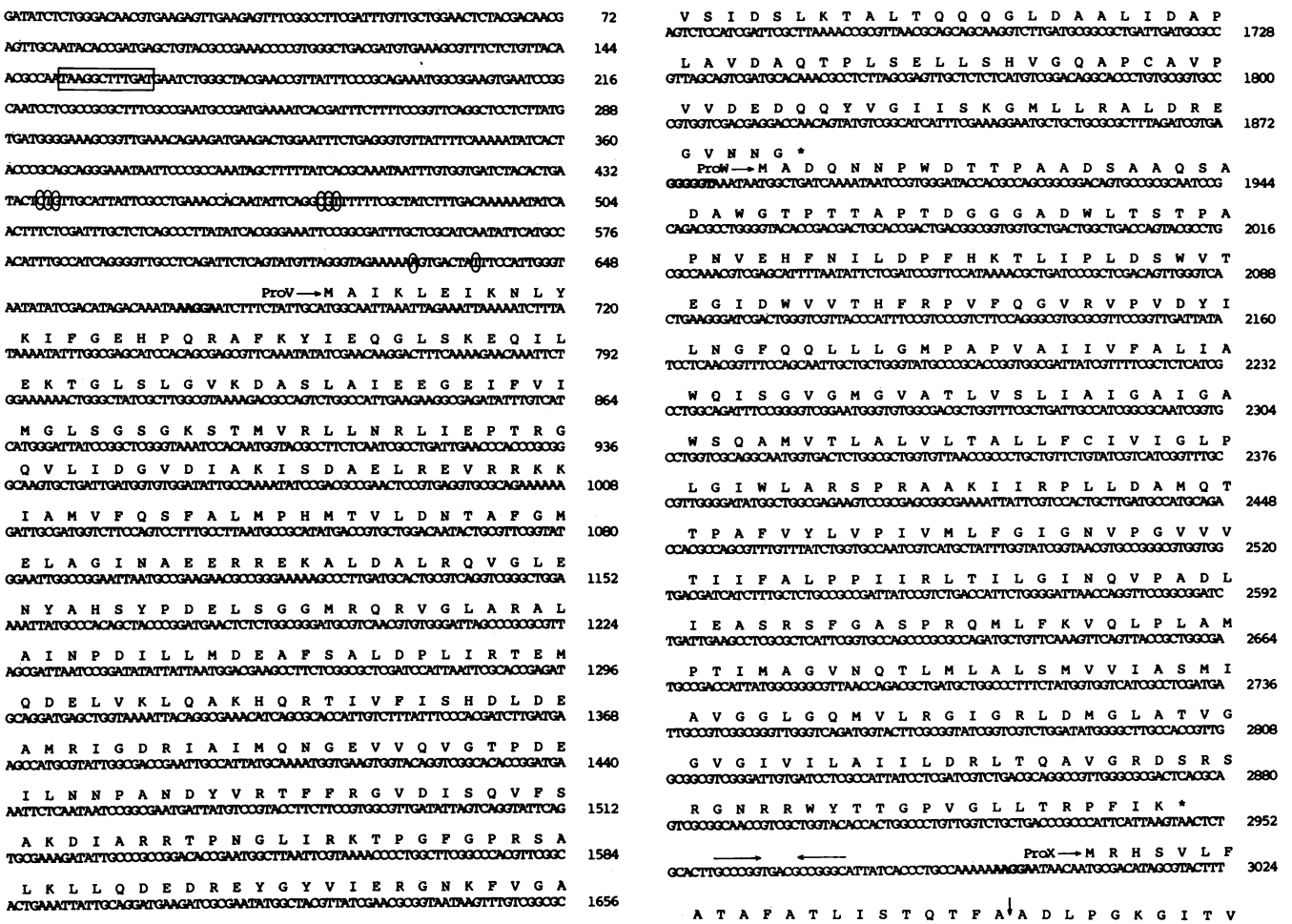

FIG. 2. Nucleotide sequence of *proU* region in *E. coli*. The sequence of the noncoding DNA strand is given, beginning at the *Eco*RV site and proceeding clockwise on the chromosome; the numbering is indicated at the right end of each line. The derived sequences of the three translation products from this locus, ProV, ProW, and ProX, have been identified in the figure and denoted in the one-letter amino acid code; putative ribosome-binding sites for the initiation of translation of the three polypeptides have been identified in boldface in the nucleotide sequence. The predicted site of cleavage of the signal sequence of ProX is indicated by a vertical arrow. The 5′ ends of *proU* mRNA identified in the primer extension experiments (Fig. 7) are encircled. A sequence corresponding to the consensus for integration host factor binding (7, 36) is boxed. The positions of an inverted repeat sequence in the *proW-proX* intergenic region and of two REP sequences (REP-1 and REP-2) distal to *proX* are marked by overhead arrows.

terms of both the residue number (Fig. 5) and the hydropathicity profile (Fig. 3), is similar to that reported earlier for this group of proteins (28). An interesting similarity between a different region of ProW and the α subunit of the acetylcholine receptor protein (45) was also observed (Fig. 6); the region of homology includes 32 identical or conserved positions within 72 amino acid residues. This similarity in sequence might relate to the fact that the choline moiety of acetylcholine, which binds the α subunit, is similar chemically to glycine betaine, which is a substrate for the ProU transporter.

The *proW-proX* intergenic sequence is 57 nucleotides long and includes a region of potential secondary structure (ΔG,

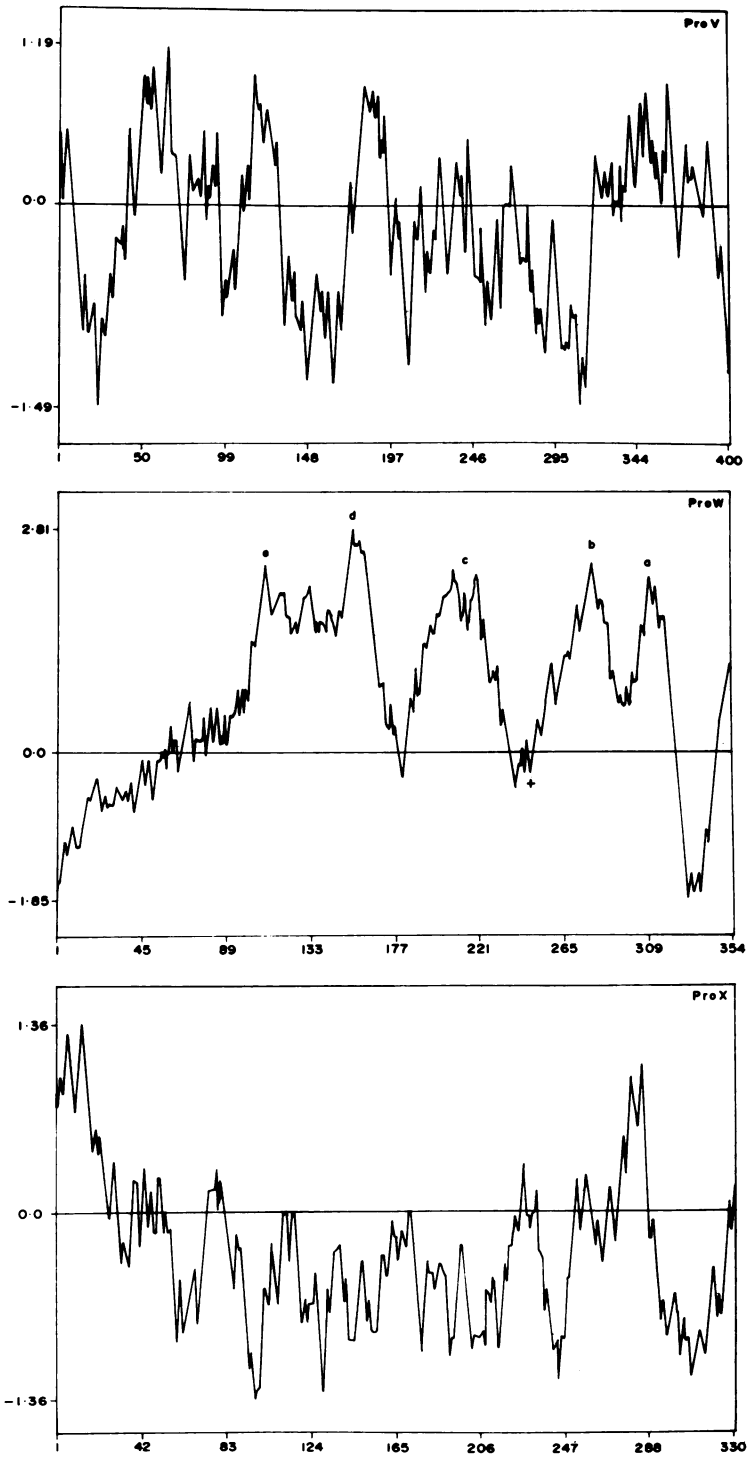

FIG. 3. Hydrophobicity plots of ProV, ProW, and ProX proteins, as obtained by the method of Kyte and Doolittle (35); a span length of 19 amino acid residues was used. The hydrophobic peaks designated a to e in ProW correspond to those identified in integral membrane components of other binding-protein-dependent transport systems in reference 28; the + symbol identifies the location of its region of similarity with the latter components, depicted in Fig. 5.

−18.2 kcal/mol [ca. −76.1 kJ/mol], calculated according to reference 54) between positions 2959 and 2976 (Fig. 2).

The predicted ProX polypeptide is 330 amino acids long, hydrophobic at its N-terminal end, and hydrophilic thereafter (Fig. 3). The periplasmic betaine-binding protein of the ProU transporter has recently been purified, and the published sequence of its N-terminal 13 residues (3) exactly matches that of the inferred ProX sequence from residue 22 onwards; the sequence of the first 21 amino acids of ProX has the characteristics typical of a leader signal peptide,

```
ProV: (26)EQGLSKEQILEKTGLSLGVKDASLAIEEGEIFVIMGL
          +: +   ::  : +    :+  ++    +::  + +
HisP:  (4)EMKLHVIDLHKRYGGHEVLKGVSLQARAGDVISIIGS

ProV: SGSGKSTMVRLLNRLIEPTRGQVLIDGVDIAKISDAE(99)
      +++++++ :+ :+ +  +: + ::::+ :+  : + :
HisP: SGSGKSTFLRCINFLEKPSEGAIIVNGQNINLVRDKD(77)

ProV: ...(144)KALDALRQVGLENYAHS YPDELSGGMRQRVGL
               :++  +  ++::: +   ++  ++++  +++ :
HisP: ...(132)RALKYLAKVGIDERAQGKYPVHLSGGQQQRVSI

ProV: ARALAINPDILLMDEAFSALDPLIRTE(202)
      +++++ :++:++ ++  +++++ :  +
HisP: ARALAMEPDVLLFDEPTSALDPELVGE(191)
```

FIG. 4. Similarity between ProV and HisP. Two regions of one protein are aligned with two regions of the other, and the sequence numbers of the N- and C-terminal residues of each region are given in parentheses. Individual amino acids are represented in the one-letter code, with the symbols + and : being used to indicate identity and conservative substitution (within one or another of the following groups: D, E, N, Q; R, K; S, T; and I, V, L), respectively, between the two proteins. Where necessary, gaps have been introduced in the sequence to maximize the homology in alignment.

which is expected to be present in a protein destined for the periplasm and to be cleaved in the process of translocation (41). ProX represents, therefore, the periplasmic betaine-binding protein, and the calculated M_r for the 309-amino-acid-long mature polypeptide is 33,729.

Primer-extension mapping of 5′ ends of *proU* mRNA. The data presented in the accompanying paper (11) indicated that *proV*, *proW*, and *proX* are organized in a single operon whose osmoresponsive expression is controlled by *cis* regulatory elements upstream of *proV*. In an effort to identify these elements, the primer extension technique was used to map the 5′ ends of mRNA species that are osmotically induced in the *proU* operon.

Five radiolabeled single-stranded DNA probes were purified (Table 1) that had their respective 3′ ends at nucleotide positions 512, 571, 606, 675, and 678 on the bottom strand of the *proU* sequence (complementary to that shown in Fig. 2). One sample from each of them was hybridized in one tube to 10 μg of total RNA prepared from a culture in which the expression of *proU* was maximally induced, and an equivalent amount of probe was hybridized in another tube to 10 μg of RNA from an uninduced culture. The conditions were so chosen that the amount of *proU*-specific mRNA was limiting

```
ProW: E A S R S F G A S P R Q M L F K V Q L P - 97
      + + +     + +   + +   + + :   + +
MalF: E A S A M D G A G P F Q N F F K I T L P - 94
MalG: E A A A L D G A T P W Q A F R L V L L P - 87
HisQ: E A A T A F G F T H G Q T F R R I M F P - 84
HisM: E A A R A Y G F S S F K M Y R C I I L P - 82
PstC: E S A Y G I G C T T W E V I W R I V L P - 98
PstA: E A A Y A L G T P K W K M I S A I T L K - 92
OppC: E A A Q V G G V S T A S I V I R H I V P - 86
```

FIG. 5. Similarity between ProW and integral membrane components of other binding-protein-dependent transporters. The protein designations are marked on the left, and the relevant sequences are shown in the one-letter code. The boxed-in regions correspond to the homology previously identified within this group of proteins (10), to which a segment from ProW has now been compared. In the context of the alignment between ProW and MalF, the symbols + and : have the same meaning as explained in the legend to Fig. 4. The distance to the C-terminal end from the last residue marked for each protein is given at the right end of the corresponding line.

```
ProW: (147)VTLALVLTALLFCIVIGLPLGIWLARSPRAAKIIRPLL
            : :::  +++  : ++   :       :    +  ++
Achα: (238)FVVNVIIPCLLFSFLTGLVFYLPTDSGEKMTLSISVLLSLTV

ProW:       DAMQTTPAFVYLVPIVMLFG IGNVPGVVVTIIFA(218)
            :   :+   + +:   +++  +  :  :::+::
Achα: FLLVIVELIPSTSSAVPLIGKYMLFTMIFVISSIIITVVVI(320)
```

FIG. 6. Similarity between ProW and the α subunit of the *Torpedo californica* acetylcholine receptor (Achα). For explanation of symbols used, see legend to Fig. 4. The alignment was identified first through the Lipman-Pearson database search (37) and was then optimized with the aid of the ALIGN program supplied by the Protein Identification Resource, National Biomedical Research Foundation, Washington, D.C. The score for the observed alignment was significantly higher ($P = 0.02$) than that expected between random pairs of peptides with similar amino acid composition.

in the hybridization reactions. The probe primer was then extended with reverse transcriptase on the mRNA to which it had hybridized, and the sizes of the run-off transcripts were measured on a urea-polyacrylamide sequencing gel.

Four sets of extension bands were identified in this experiment (Fig. 7), all of which were present only in the culture grown at elevated osmolarity. They correspond to 5′ mRNA ends at nucleotide positions 437 to 439, 473 to 475, 629, and 637 of the *proU* sequence shown in Fig. 2.

Evidence for sequence-directed DNA bending in *proU* upstream regulatory region. Sequence-directed DNA bending is known to occur in regions with multiple homopolymeric stretches of A and T residues so spaced that they are situated along a common phase of the double helix (34). Bent DNA has been shown in several instances to be an important recognition feature for the binding of particular proteins in *E. coli* (42, 52). My data suggest that the sequence in the upstream regulatory region of *proU* has the characteristics of bent DNA.

Plaskon and Wartell (46) have recently described a theoretical method to assess the propensity for a given DNA sequence to bend; by their scoring criteria, the region between the nucleotides 390 and 510 in the *proU* sequence was predicted to have a bent DNA conformation (data not shown). The conventional experimental hallmark for bent DNA has been the demonstration of anomalously slow mobility of the concerned restriction fragments upon electrophoresis in polyacrylamide gels at low temperature as well as the demonstration of the correction of this anomaly upon electrophoresis at high temperature (34, 42). In the case of *proU*, this feature was tested with restriction enzyme-digested fragments of replicative-form DNA from recombinant M13 phage that carried the 3.8-kb *Eco*RV-*Nsi*I region of the *proU* locus. Fragments carrying the upstream *proU* region from a variety of restriction enzyme digestions

TABLE 1. List of probes used in primer-extension experiments

Probe no.	Parental M13 template	3′ end of probe generated by:	Position[a] of 3′ end	Length[b] of probe (bases)
1	RE13.2	*Hin*fI	678	124
2	RE13.2	*Hin*fI (filled-in)	675	127
3	RE23.2	*Hin*fI	606	73
4	RE23.2	*Sfa*NI	571	108
5	RE14.3	*Taq*I	512	100

[a] The nucleotide position number corresponds to that in Fig. 2, but in this case the 3′ end is on the strand complementary to the one whose sequence is given in the figure.

[b] The probe length includes 17 bases of universal primer and an additional 17 bases of tg131 sequence to the 5′ side of *proU*-specific DNA.

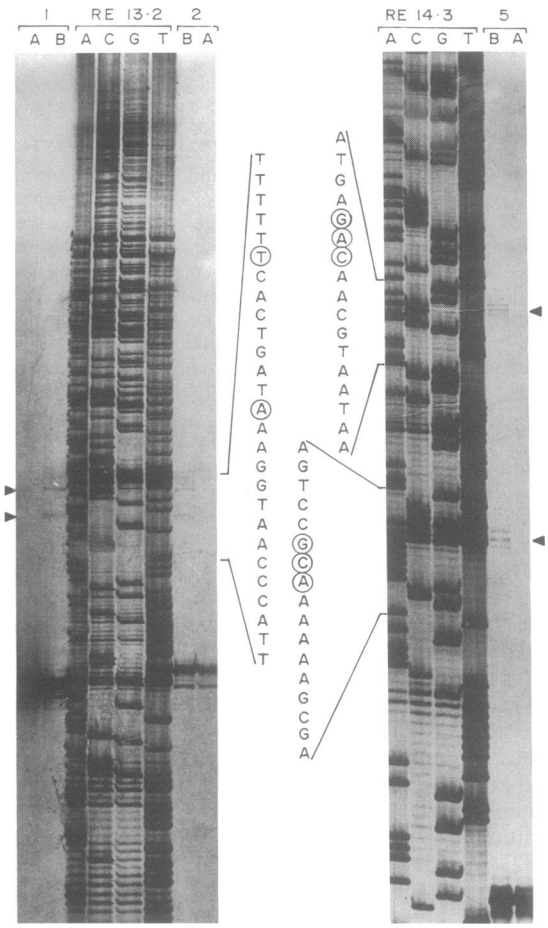

FIG. 7. 5′-End mapping of mRNA from the *proU* locus by primer extension analysis. Each of the probes, 1, 2, and 5 (described in Table 1), was mixed with total RNA from (A) strain MC4100 grown in low-osmolarity medium or (B) strain GJ157 grown in high-osmolarity medium and then analyzed further as described in the text. The corresponding lanes on the autoradiograph are identified at the top of the figure. The lanes representing the sequence ladders of two M13 clones, RE13.2 and RE14.3, run as markers on the same gel are also marked above the figure. The intense bands towards the bottom of each of the pairs of test lanes correspond to the labeled probe itself, and the positions of the extension products (seen only on lanes B) are marked by arrowheads. The relevant portions of the marker sequences are indicated, and the nucleotides within each sequence whose sizes correspond to those of the probe extension products are encircled. The marker sequences indicated here are from the strand complementary to that shown in Fig. 2. Extension products corresponding to those seen in lane 5B were also obtained with probe 4 (Table 1) after hybridization to GJ157 RNA (data not shown).

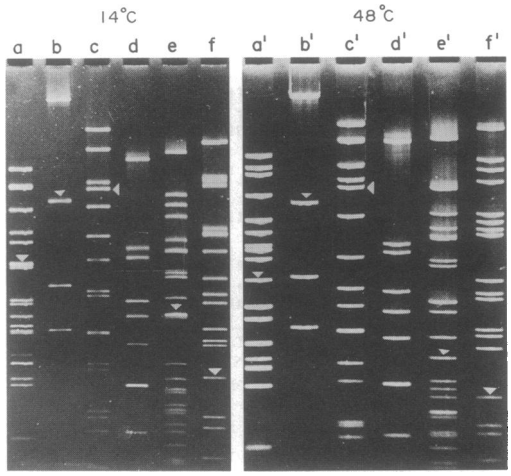

FIG. 8. Polyacrylamide (5%) gel electrophoresis (run in 90 mM Tris-borate–1 mM EDTA and visualized after staining with ethidium bromide [38]) of restriction enzyme-digested DNA from recombinant M13 phage carrying the 3.8-kb *Eco*RV-*Nsi*I insert from *proU*. Each digested sample was run on a pair of gels at 14 or 48°C, as marked; the corresponding lanes have been identified by a common letter, with prime symbols denoting those run at 48°C. The enzymes used were: a, *Hin*fI; b, *Bgl*I; c, *Hae*III; e, *Hpa*II; and f, *Dde*I. The fragments corresponding to the upstream region of *proU* in the different lanes are marked by arrowheads, and their apparent sizes are indicated in Table 2. Lanes d and d′ represent *Hin*fI fragments of pBR322 DNA, run as size markers on the gels above.

showed a consistent 22 to 29% retardation in mobility (from that expected of their size) upon electrophoresis at 14°C, and this retardation was largely corrected when the digested fragments were electrophoresed at 48°C (Fig. 8 and Table 2); the 0.19-kb *Dde*I fragment between nucleotides 332 and 521 of the *proU* sequence was the smallest identified from this region that exhibited anomalous mobility, in accord with the prediction above. Fragments encompassing two other A+T-rich regions from the M13 vector DNA (between nucleotides 360 and 670 and between 5870 and 6030, in the numbering system of reference 55) also exhibited 12 to 15% anomalous mobility in this experiment. None of the other regions of vector or of insert in the 11-kb molecule exhibited any consistent anomaly in electrophoretic mobility larger than 4% with the different restriction enzyme digests (data not shown).

Features of interest towards the 3′ end of *proU*. Beyond the *proX* gene in the operon, between nucleotides 4004 and 4089, there exist two copies of so-called repetitive extragenic palindromic (REP) sequences (16, 53) organized in inverse orientation to one another (Fig. 2). As is the case with other regions in which REP sequences have been found, this region in *proU* is also capable of alternative forms of extensive, stable secondary structure (not shown). Arguing further by analogy with the other systems in which REP sequences have been described (16, 53), it may be expected that this part of *proU* is also transcribed and that it defines either the 3′ noncoding region or an intergenic region within the operon.

We had attempted to determine the 3′ end of *proU* mRNA by the S1 nuclease mapping technique, but our results (C. S.

TABLE 2. Anomalous electrophoretic mobility of DNA fragments carrying the *proU* regulatory region[a]

Restriction fragment	Extent[b]	Actual size (base pairs)	Apparent size (base pairs) at:		% Retardation at:	
			14°C	48°C	14°C	48°C
*Hin*fI	209 to 602	393	495	413	26.0	5.1
*Bgl*I	147 to 827	680	880	715	29.4	5.1
*Hae*III	38 to 835	797	975	830	22.3	4.1
*Hpa*II	268 to 546	278	350	285	25.9	2.5
*Dde*I	332 to 521	189	230	194	21.7	2.6

[a] Calculations in this table are based on data from the gel electrophoresis shown in Fig. 8.

[b] The numbers correspond to nucleotide positions in Fig. 2.

Dattananda and J. Gowrishankar, unpublished) in fact suggest that the transcript from this locus extends beyond the right extremity of the chromosomal DNA segment obtained in the primary cloning of *proU* (that is, beyond nucleotide 4362 of the sequence reported here). The 3′ extent of the operon, therefore, remains uncharacterized.

DISCUSSION

The nucleotide sequence of the *proU* locus reported here is, for the major part, in agreement with the restriction maps of this region obtained by us (11, 20) and by Bremer and co-workers (14, 39), and also with the data of Kohara et al. (32) on the physical map of the entire *E. coli* chromosome (within which we have localized *proU* in the region around 2,815 kb). The orientation, size, and location of the three open reading frames identified in the sequence are all in close agreement with the data presented in the accompanying paper on the genes in this locus, their direction of transcription, and their respective products (11). Faatz et al. (14) have speculated on the presence of a fourth gene within this region, but the nucleotide sequence does not support this prediction.

Several genetic lines of evidence indicate that the sequence reported here includes the majority, if not all, of the *cis* information necessary for the known features of *proU* function and regulation. (i) All chromosomal mutations in *proU* that have so far been mapped physically are located within this segment of DNA (11, 14). (ii) The data presented in the accompanying paper establish that this region is sufficient for the expression of all facets of the ProU$^+$ phenotype (sensitivity to 3,4-dehydroproline; osmoprotection by both glycine betaine and L-proline) in a variety of *proU* null mutants, including a Δ*proU* strain (11). (iii) The region of *proU* DNA downstream of the *Eco*RV site, when borne on a multicopy plasmid, is sufficient to confer osmotic inhibition of growth (11), implying that the *cis* elements necessary for osmoresponsivity of expression are also present in this region. (iv) Finally, May et al. (39) and we (11) have shown that the region of *proU* cloned downstream of the *Eco*RV site is sufficient to permit osmoresponsive expression of β-galactosidase from plasmids bearing *proU*::*lacZ* gene fusions. It should also be noted in this context that our own results (11) were obtained with the plasmid pHYD151, which was in fact constructed by subcloning from the M13tg131 derivative used herein for DNA sequence determination.

Two questions, however, still remain open. One is whether the sequence upstream of the *Eco*RV site is also involved in *cis* regulation of *proU*. Although our data (11) suggest that the sequence cloned downstream of *Eco*RV is entirely sufficient for instantaneous osmotic induction of *proU*, May et al. (39) have reported that it does not appear to regulate steady-state expression of *proU* over the full range of osmolarity to which the chromosomal gene is subject; however, interpretation in their case is also complicated by the fact that regulation with the foreshortened sequence was studied on a multicopy plasmid, under conditions where growth might have been affected by overexpression of a hybrid protein product (11). The second question is whether additional downstream genes exist in the *proU* operon; if indeed there are any, they would (for the reasons discussed above) either define new functions for the ProU porter or be nonessential for its transport function. We are at present attempting to address these questions.

ProU and other binding-protein-dependent transport systems: similarities and differences. In many respects, the ProU transporter is similar to other binding-protein-dependent transport systems that have so far been studied in both *E. coli* and *S. typhimurium* (reviewed in reference 1). Thus, (i) ProU is also a multicomponent porter, with a periplasmic substrate-binding protein being one of its components; (ii) the genes encoding the porter are organized in a single operon; (iii) ProW has the features of an integral membrane protein (with several hydrophobic stretches capable of spanning the membrane) and shows the same conserved sequence motif previously identified in corresponding polypeptides of the other transport systems; and (iv) ProV also shares primary sequence similarity with the nucleotidyl triphosphate-binding domains of corresponding component proteins of the other porters. Arguing again by analogy, therefore, one may predict that the processed ProX protein binds substrate in the periplasm and presents it to the membrane components of the porter for transport across the inner membrane and that ProV is a peripheral membrane protein, found on the cytoplasmic surface of the membrane, which is involved in the coupling between high-energy phosphate bond hydrolysis and the work done by the porter.

There are two differences between ProU and the majority of other binding-protein-dependent transport systems. One is that ProU is composed of only three component polypeptides, whereas all other transporters (with the exception of AraFGH [49]) have a minimum of four polypeptide components. In each of the latter cases, there is evidence of gene duplication within the operon (1), so that some of the polypeptides are homologous to one another and perhaps function as hetero-oligomers in the fully assembled transporter; the corresponding polypeptides in the ProU porter may instead be functional as homo-oligomers.

The second difference is that ProX, the binding protein component of the ProU porter, is encoded by the third gene in the operon, whereas in the case of each of the other transport systems (with the exception of the Rbs transporter [5] and perhaps also of the vitamin B_{12} transporter [15]), the binding protein is the product of the first gene in the operon. It is possible that the above preference for a first-gene arrangement reflects a need for the binding protein to be expressed in far greater molar proportion than the membrane components of the porter and yet to remain subject to the same pattern of regulation in response to environmental signals (29, 43, 44). If one assumes, in the case of ProU as well, that the periplasmic protein is synthesized in larger quantity than the other two polypeptides, then this could be achieved either by differential rates of translation or by differential stabilities of mRNA from the three coding regions of the operon. In this context, it is significant that whereas β-galactosidase activity in mutants with *lac* operon fusions in each of the three genes is similar, the activity from *proX*::*lac* gene fusions is much higher than that from *proW*::*lac* gene fusions (11). The Shine-Dalgarno sequences upstream of *proV* and of *proX* are identical, but the latter is more optimally spaced from the ATG start codon (17) (Fig. 2); this might contribute to less efficient initiation of translation of *proV* (and also of *proW*, which appears to be translationally coupled to *proV*). Furthermore, analysis of synonymous codon usage (a parameter that might affect rate of polypeptide chain elongation on mRNA) in the three genes gives the following percentage values, respectively, for rare codons and for infrequently used codons (as defined in reference 33): *proV*, 9.8 and 26.2; *proW*, 6.8 and 17.5; and *proX*, 4.5 and 13. The values observed in *proX* are similar to the average for all *E. coli* genes, whereas those for *proV* are close to the values observed in poorly expressed genes such

as *dnaG* (33). With regard to differential mRNA stability, one possibility is that endonucleolytic cleavage occurs in the *proW-proX* intergenic segment of the transcript (similar to that described in the *pap* operon [2]) and that the *proX* messenger segment alone is then stabilized by the REP sequences at its 3′ end (43, 44).

***cis* regulatory elements in the upstream region of *proU*.** An unexpected finding from the primer extension experiments in this study was that at least some *proU* transcripts have an unusually long leader region, extending as much as 250 nucleotides upstream of the initiation codon of the *proV* gene. The extension bands identified in Fig. 7 might correspond (i) to sites of transcription initiation from osmoresponsive promoters or (ii) to 5′ mRNA ends generated after nucleolytic cleavage of transcripts from an upstream promoter. Additional lines of evidence from in vivo promoter cloning and in vitro transcription experiments are required to determine which, if any, of the identified bands are explained by (i) above. A perusal of the DNA sequence around each of the four 5′-end positions indicates that a reasonable match with the consensus *E. coli* promoter sequence (21, 48) exists in three of the cases (corresponding to the ends at 437–439, 473–475, and 629; data not shown).

Several alternative possibilities exist for inverted-repeat structures in the DNA immediately upstream of nucleotide 437 (data not shown); the sequence between nucleotides 152 and 163 also matches the consensus sequence for binding of integration host factor (7, 36), a protein that is believed to influence the expression of several genes in *E. coli* (7, 12). The role, if any, for these sequences or for the bent-DNA motif observed in this region, with regard to the *cis* osmotic regulation of *proU*, remains to be determined.

Higgins et al. (23) have suggested that the induction of *proU* is the direct result of increased DNA supercoiling, which in turn is a consequence of intracellular K^+ accumulation under conditions of water stress. If their model is correct, then the sequence and structures identified in the *proU* upstream region might contribute either to an increased local supercoiling effect in response to changes in intracellular ion concentration or to an increased promoter sensitivity to the general superhelicity change (23).

ACKNOWLEDGMENTS

It is with pleasure and gratitude that I acknowledge the contributions from the following people to this work: Arna Andrews, K. Chandrasekaran, C. S. Dattananda, Sam Ganesan, K. Guruprasad, Ajay Kumar, Saroja Nagaraj, G. Narasaiah, Judyta Praszkier, K. Rajkumari, T. A. Thanaraj, and Ji Yang. I would especially like to thank Jim Pittard for many discussions and advice.

Some of the work reported here was done at the University of Melbourne, where it was supported by a Biotechnology Career Fellowship award from the Rockefeller Foundation. Partial support from the Department of Science and Technology, Government of India, for this work is also acknowledged.

LITERATURE CITED

1. **Ames, G. F.-L.** 1986. Bacterial periplasmic transport systems: structure, mechanism, and evolution. Annu. Rev. Biochem. **55**:397–425.
2. **Baga, M., M. Göransson, S. Normark, and B. E. Uhlin.** 1988. Processed mRNA with differential stability in the regulation of E. coli pilin gene expression. Cell **52**:197–206.
3. **Barron, A., J. U. Jung, and W. Villarejo.** 1987. Purification and characterization of a glycine betaine binding protein from *Escherichia coli*. J. Biol. Chem. **262**:11841–11846.
4. **Barron, A., G. May, E. Bremer, and M. Villarejo.** 1986. Regulation of envelope protein composition during adaptation to osmotic stress in *Escherichia coli*. J. Bacteriol. **167**:433–438.
5. **Bell, A. W., S. D. Buckel, J. M. Groarke, J. N. Hope, D. H. Kingsley, and M. A. Hermodson.** 1986. The nucleotide sequences of the *rbsD*, *rbsA*, and *rbsC* genes of *Escherichia coli* K12. J. Biol. Chem. **261**:7652–7658.
6. **Cairney, J., I. R. Booth, and C. F. Higgins.** 1985. Osmoregulation of gene expression in *Salmonella typhimurium*: *proU* encodes an osmotically inducible betaine transport system. J. Bacteriol. **164**:1224–1232.
7. **Craig, N. L., and H. A. Nash.** 1984. E. coli integration host factor binds to specific sites in DNA. Cell **39**:707–716.
8. **Csonka, L. N.** 1982. A third L-proline permease in *Salmonella typhimurium* which functions in media of elevated osmotic strength. J. Bacteriol. **151**:1433–1443.
9. **Das, A., and C. Yanofsky.** 1984. A ribosome binding site sequence is necessary for efficient expression of the distal gene of a translationally-coupled gene pair. Nucleic Acids Res. **12**:4757–4768.
10. **Dassa, E., and M. Hofnung.** 1985. Sequence of gene *malG* in *E. coli* K12: homologies between integral membrane components from binding protein-dependent transport systems. EMBO J. **4**:2287–2293.
11. **Dattananda, C. S., and J. Gowrishankar.** 1989. Osmoregulation in *Escherichia coli*: complementation analysis and gene-protein relationships in the *proU* locus. J. Bacteriol. **171**:1915–1922.
12. **Dorman, C. J., and C. F. Higgins.** 1987. Fimbrial phase variation in *Escherichia coli*: dependence on integration host factor and homologies with other site-specific recombinases. J. Bacteriol. **169**:3840–3843.
13. **Dunlap, V. J., and L. N. Csonka.** 1985. Osmotic regulation of L-proline transport in *Salmonella typhimurium*. J. Bacteriol. **163**:296–304.
14. **Faatz, E., A. Middendorf, and E. Bremer.** 1988. Cloned structural genes for the osmotically regulated binding-protein-dependent glycine betaine transport system (ProU) of *Escherichia coli* K-12. Mol. Microbiol. **2**:265–279.
15. **Friedrich, M. J., L. C. DeVeaux, and R. J. Kadner.** 1986. Nucleotide sequence of the *btuCED* genes involved in vitamin B_{12} transport in *Escherichia coli* and homology with components of periplasmic-binding-protein-dependent transport systems. J. Bacteriol. **167**:928–934.
16. **Gilson, E., J.-M. Clement, D. Brutlag, and M. Hofnung.** 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. EMBO J. **3**:1417–1421.
17. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo.** 1981. Translational initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
18. **Gowrishankar, J.** 1985. Identification of osmoresponsive genes in *Escherichia coli*: evidence for participation of potassium and proline transport systems in osmoregulation. J. Bacteriol. **164**: 434–445.
19. **Gowrishankar, J.** 1988. Osmoregulation in *Enterobacteriaceae*: role of proline/betaine transport systems. Curr. Sci. **57**:225–234.
20. **Gowrishankar, J., P. Jayashree, and K. Rajkumari.** 1986. Molecular cloning of an osmoregulatory locus in *Escherichia coli*: increased *proU* gene dosage results in enhanced osmotolerance. J. Bacteriol. **168**:1197–1204.
21. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
22. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351–359.
23. **Higgins, C. F., C. J. Dorman, D. A. Stirling, L. Waddell, I. R. Booth, G. May, and E. Bremer.** 1988. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in S. typhimurium and E. coli. Cell **52**:569–584.
24. **Higgins, C. F., P. D. Haag, K. Nikaido, F. Ardeshir, G. Garcia, and G. F.-L. Ames.** 1982. Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *S. typhimurium*. Nature (London) **298**:723–727.
25. **Higgins, C. F., I. D. Hiles, G. P. C. Salmond, D. R. Gill, J. A. Downie, I. J. Evans, I. B. Holland, L. Gray, S. D. Buckel, A. W. Bell, and M. A. Hermodson.** 1986. A family of related ATP-

binding subunits coupled to many distinct biological processes in bacteria. Nature (London) **323**:448–450.
26. **Higgins, C. F., I. D. Hiles, K. Whalley, and D. J. Jamieson.** 1985. Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. EMBO J. **4**:1033–1040.
27. **Higgins, C. F., L. Sutherland, J. Cairney, and I. R. Booth.** 1987. The osmotically regulated *proU* locus of *Salmonella typhimurium* encodes a periplasmic betaine-binding protein. J. Gen. Microbiol. **133**:305–310.
28. **Hiles, I. D., M. P. Gallagher, D. J. Jamieson, and C. F. Higgins.** 1987. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium*. J. Mol. Biol. **195**:125–142.
29. **Horazdovsky, B. F., and R. W. Hogg.** 1987. High affinity L-arabinose transport operon: gene product expression and mRNAs. J. Mol. Biol. **197**:27–35.
30. **Hudson, G. S., and B. E. Davidson.** 1984. Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K12. J. Mol. Biol. **180**:1023–1051.
31. **Kieny, M. P., R. Lathe, and J. P. Lecocq.** 1983. New versatile cloning and sequencing vectors based on bacteriophage M13. Gene **26**:91–99.
32. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole E. coli chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
33. **Konigsberg, W., and G. N. Godson.** 1983. Evidence for use of rare codons in the *dnaG* and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80**:687–691.
34. **Koo, H.-S., H.-M. Wu, and D. M. Crothers.** 1986. DNA bending at adenine · thymine tracts. Nature (London) **320**:501–506.
35. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
36. **Leong, J. M., S. Nunes-Duby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy.** 1985. The ø80 and P22 attachment sites: primary structure and interaction with *Escherichia coli* integration host factor. J. Biol. Chem. **260**:4468–4477.
37. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435-1441.
38. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **May, G., E. Faatz, M. Villarejo, and E. Bremer.** 1986. Binding protein dependent transport of glycine betaine and its osmotic regulation in *Escherichia coli* K12. Mol. Gen. Genet. **205**: 225–233.
40. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
41. **Michaelis, S., and J. Beckwith.** 1982. Mechanism of incorporation of cell envelope proteins in Escherichia coli. Annu. Rev. Microbiol. **36**:435–465.
42. **Mizuno, T.** 1987. Static bend of DNA helix at the activator recognition site of the *ompF* promoter in *Escherichia coli*. Gene **54**:57–64.
43. **Newbury, S. F., N. H. Smith, and C. F. Higgins.** 1987. Differential mRNA stability controls relative gene expression within a polycistronic operon. Cell **51**:1131–1143.
44. **Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles, and C. F. Higgins.** 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell **48**:297–310.
45. **Noda, M., H. Takahashi, T. Tanabe, M. Toyosato, Y. Furatini, T. Hirose, M. Asai, S. Inayama, T. Miyata, and S. Numa.** 1982. Primary structure of α-subunit precursor of *Torpedo californica* acetylcholine receptor deduced from cDNA sequence. Nature (London) **299**:793–797.
46. **Plaskon, R. R., and R. M. Wartell.** 1987. Sequence distributions associated with DNA curvature are found upstream of strong *E. coli* promoters. Nucleic Acids Res. **15**:785–796.
47. **Priess, H., D. Kamp, R. Kahmann, B. Brauer, and H. Delius.** 1982. Nucleotide sequence of the immunity region of bacteriophage Mu. Mol. Gen. Genet. **186**:315–321.
48. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
49. **Scripture, J. B., C. Voelker, S. Miller, R. T. O'Donnell, L. Polgar, J. Rade, B. F. Horazdovsky, and R. W. Hogg.** 1987. High-affinity L-arabinose transport operon: nucleotide sequence and analysis of gene products. J. Mol. Biol. **197**:37–46.
50. **Staden, R.** 1977. Sequence data handling by computer. Nucleic Acids Res. **4**:4037–4051.
51. **Staden, R.** 1978. Further procedures for sequence analysis by computer. Nucleic Acids Res. **5**:1013–1015.
52. **Stenze, T. T., P. Patel, and D. Bastia.** 1987. The integration host factor of Escherichia coli binds to bent DNA at the origin of replication of the plasmid pSC101. Cell **49**:709–717.
53. **Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37**:1015–1026.
54. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. D. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
55. **van Wezenbeek, P. M. G. F., T. J. M. Hulsebos, and J. G. G. Schoemakers.** 1980. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. Gene **11**:129–148.
56. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.