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, 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 added 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 BglII 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 EcoRV site shown in Fig. 1. The complete nucleotide sequence of the chromosomal region on pHYD58 to the right of the EcoRV 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 tgl31 (31) of three fragments from pHYD58: a 3.8-kb EcoRV-NsiI fragment and two NsiI 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 NsiI sites in this region was determined after an Rsal fragment from this region was cloned into tgl31, 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.
strain MC4100 (Δlac 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 ΔputPA 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 pHYS8, 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::Tn1000 plasmid (11) is consistent with the translational initiation site marked here, and Faatz et al. (14) have shown that a Tn5 insertion 0.55 kb downstream of the EcoRV site does not disrupt proV. ProV is predicted to be a 400-amino-acid-long polypeptide, relatively hydrophilic (Fig. 3), with an M, of 44,162; interestingly, it is devoid of any tryptophanyl residues. The predicted proV coding sequence extends beyond the Sall 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 (pHYS8 [20]) in which this Sall end has been ligated with the Sall 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 pHYS8 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 hydrophilic polypeptide 354 amino acids long (M, 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 EcoRV 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.

In terms of both the residue number (Fig. 5) and the hydrophaticity 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, -52 kcal/mol).
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: \(2(6)\) EQLSKEYLTKSCLCVDASLAIREGEFHVINGL
HisP: \(4(7)\) EMELYVLHEKFRGQVLEQQLQGARQGVWISIIGS
ProV: SGCGESYVLWWLVPITLGQVLDGDIARSDAK(99)
HisP: SGCGESTFLRCGIFLFLXPSGAGALLYQGHNMLWECBD(77)
ProV: \(\ldots\) KALDALRQVGLENYARS(144)
HisP: \(\ldots\) KALDALRQVGLENYARS(144)
ProV: ARAAI1PIPLLNDAEASLPLITE(202)
HisP: ARAANEDPVDLDEPTSLPYLE(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 +: 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 \(M\), for the 309-amino-acid-long mature polypeptide is 33,729.

**Primer-extension mapping of S' 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 S' 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 S' 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 \(\mu\)g 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 \(\mu\)g of RNA from an uninduced culture. The conditions were so chosen that the amount of proU-specific mRNA was limiting 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 osmolality. They correspond to S' mRNA ends at nucleotide positions 437 to 439, 473 to 475, 623, 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 EcoRV-NsiI 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

<table>
<thead>
<tr>
<th>Probe no.</th>
<th>Parental M13 template</th>
<th>3' end of probe generated by:</th>
<th>Position* of 3' end</th>
<th>Length* of probe (bases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RE13.2</td>
<td>HinfI</td>
<td>678</td>
<td>124</td>
</tr>
<tr>
<td>2</td>
<td>RE13.2</td>
<td>HinfI (filled-in)</td>
<td>675</td>
<td>127</td>
</tr>
<tr>
<td>3</td>
<td>RE23.2</td>
<td>HinfI</td>
<td>606</td>
<td>73</td>
</tr>
<tr>
<td>4</td>
<td>RE23.2</td>
<td>SfaNl</td>
<td>571</td>
<td>108</td>
</tr>
<tr>
<td>5</td>
<td>RE14.3</td>
<td>TaqI</td>
<td>512</td>
<td>100</td>
</tr>
</tbody>
</table>

* 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.

* The probe length includes 17 bases of universal primer and an additional 17 bases of tgl31 sequence to the 5' side of proU-specific DNA.
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 regiona |
|----------------------------------|---------------------------------|------------------|--------------------------|-----------------|-----------------|
| Restriction fragment | Extentb | Actual size (base pairs) | Apparent size (base pairs) at: | % Retardation at: |
|-----------------------|---------|--------------------------|-----------------------------|-----------------|----------------|
| Hinfl | 209 to 602 | 393 | 495 | 413 | 26.0 | 5.1 |
| BglII | 147 to 827 | 680 | 880 | 715 | 29.4 | 5.1 |
| HaeIII | 38 to 835 | 797 | 975 | 830 | 22.3 | 4.1 |
| HpaII | 268 to 546 | 278 | 350 | 285 | 25.9 | 2.5 |
| DdeI | 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.

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 and 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 EcoRV-NorI 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, Hinfl; b, BglII; c, HaeIII; e, HpaII; and f, DdeI. 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 Hinfl 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 DdeI 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
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; osmoregulation by both glycine betaine and L-proline) in a variety of proU null mutants, including a ΔproU strain (11). (iii) The region of proU DNA downstream of the EcoRV site, when borne on a multicopy plasmid, is sufficient to confer osmotic inhibition of growth (11), implying that the cis elements necessary for osmoregulation 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 EcoRV site is sufficient to permit osmoregulation of 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 EcoRV site is also involved in cis regulation of proU. Although our data (11) suggest that the sequence cloned downstream of EcoRV 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 non-essential 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. These are two different functions between ProU and 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 ProUporter, is encoded by the third gene in the operon, whereas in 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 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, 8.8% 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 osmosensitive 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, A. Chandrasekaran, C. S. Dattananda, Sam Ganesh, K. Guruprasad, Ajay Kumar, Saroja Nagaraj, G. Narasiah, Judyta Praszker, K. Rajkumari, T. A. Thanaraj, and J. 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


