

The Leucine Binding Proteins of *Escherichia coli* as Models for Studying the Relationships Between Protein Structure and Function

Tammy K. Antonucci, Robert Landick, and Dale L. Oxender

Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109

The genes encoding the leucine binding proteins in *E. coli* have been cloned and their DNA sequences have been determined. One of the binding proteins (LIV-BP) binds leucine, isoleucine, valine, threonine, and alanine, whereas the other (LS-BP) binds only the D- and L-isomers of leucine. These proteins bind their solutes as they enter the periplasm, then interact with three membrane components, livH, livG, and livM, to achieve the translocation of the solute across the bacterial cell membrane. Another feature of the binding proteins is that they must be secreted into the periplasmic space where they carry out their function. The amino acid sequence of the two binding proteins is 80% homologous, indicating that they are the products of an ancestral gene duplication. Because of these characteristics of the leucine binding proteins, we are using them as models for studying the relationships between protein structure and function.

Key words: amino acid transport, binding proteins, secretion, gene duplication, oligonucleotide-directed mutagenesis

There are three major systems operating in *Escherichia coli* for the transport of the branched-chain amino acids. Two of these systems consist of a binding protein and three membrane components while the third system appears to involve only a single membrane component. In this discussion we will describe the biochemical and genetic properties of the two binding protein-dependent transport systems, emphasizing the multiple functions of the binding protein components of the transport systems and, as a consequence, the value of the binding proteins for understanding the relationship(s) between protein structure and function.

The LIV-I system transports L-leucine, L-isoleucine, and L-valine with K_d values of approximately 10^{-6} to 10^{-7} μM [1-4]. In addition, this system is responsible for the transport of threonine and alanine, but with somewhat lower affinities [2].

Robert Landick's present address is the Department of Biological Sciences, Stanford University, Stanford, CA 94305.

Received June 6, 1985; revised and accepted August 6, 1985.

The other transport system to be considered is the leucine-specific transport system (LS). The LS system transports L-leucine with a K_d value of approximately 10^{-6} μM ; D-leucine is also an acceptable substrate but with a lower affinity [1-4]. All structural genes for the LS and LIV-I transport systems are located at the minute 76 region on the *E coli* chromosome [4].

The genes encoding the components of the branched-chain amino acid transport systems LIV-I and LS in *E coli* have been cloned [5]. The gene organization is shown in Figure 1. There are three functional operons with their directions of transcription indicated by the arrows. The *livJ* and *livK* gene products are the periplasmic binding proteins, called LIV-BP and LS-BP, respectively. The LIV-BP binds isoleucine, leucine, and valine with high affinity and threonine and alanine with a somewhat lower affinity. The LS-BP binds only isomers of leucine. The *livH*, *livM*, and *livG* genes encode proteins that have been localized to the inner membrane of the bacterial cell [6]. These proteins presumably interact with the BP-amino acid complex to effect the transport of the amino acid into the cytoplasm of the cell. Thus the LIV-I and LS transport systems consist of individual binding proteins that have different substrate specificities yet interact with the same membrane components during the transport process. The *livL* operon produces a 17 kD protein but its role in transport, if any, has not been determined [7].

CHARACTERISTICS OF THE LIV-BP AND THE LS-BP

The DNA sequence of the *livJ* and *livK* genes encoding the two binding proteins has been determined [8] and is shown in Figure 2. A comparison of the coding regions of the two sequences shows that the two binding protein genes are 80% homologous, indicating that they are the products of an ancestral gene duplication [8-11]. While the sequences of the 5' flanking regions and coding regions revealed no strong homology with the known ribosome-binding sites, the pattern of codon usage in both genes implies that they are translated with average efficiency.

The amino acid sequence of the LIV-BP has been previously reported [12] and is consistent with the DNA sequence except for two positions [8]. Moreover, the crystal structure of the mature LIV-BP was determined to a resolution of 2.0 Å [13]. The crystal structure reveals that the LIV-BP has two domains that fold to form a cleft [13]. One domain of LIV-BP is composed of the N-terminal half of the protein, while the other domain is formed by most of the C terminal portion of the molecule. The secondary structure obtained from the crystallographic analysis shows that each domain is formed by an alternating series of alpha helices and beta sheets [8]. Figure 3 shows the predicted secondary structural features superimposed onto a drawing of the crystal structure. The filled-in dots represent differences in amino acid sequences between the two binding proteins. The two arrows in Figure 2 point out regions where amino acid insertions occur in the LS-BP. A disulfide bridge can be positioned between segments beta-2 and beta-3.

Substrate Binding

One of the major functions carried out by the BPs is binding of substrate. As mentioned above, the LIV-BP and the LS-BP have different substrate specificities. Saper and Quioco [13] have observed that when L-leucine is diffused into crystals

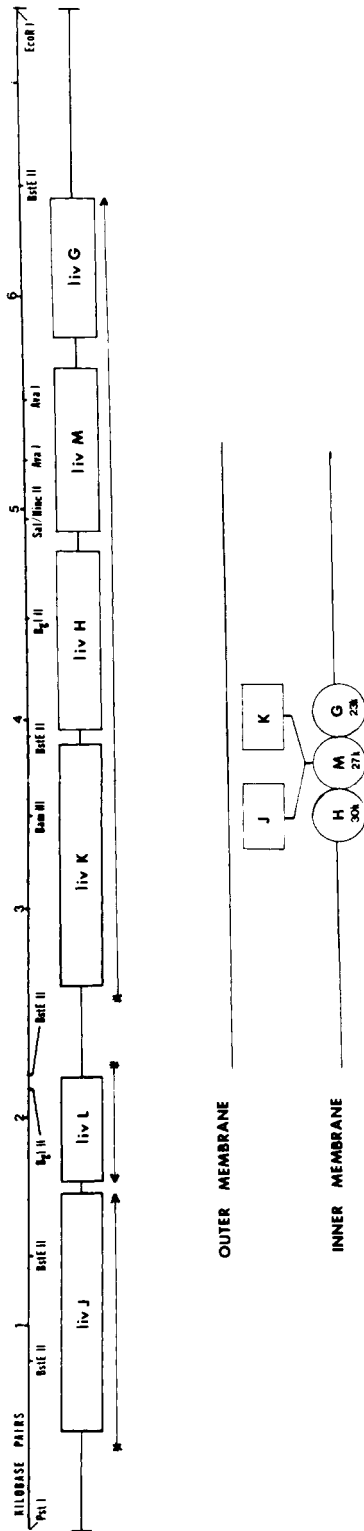


Fig. 1. Organization of the high affinity branched-chain amino acid transport genes in *E. coli*. Arrows represent the directions of transcription for the operons.

```

-20          -10          -1  1
ATG AAC ACA AAG GGC AAA GCG TTA CTG GCA GGA TTG ATC GCG CTG GCA TTC AGC AAT ATG GCT CTG GCA GAA GAT ATT AAA GTC GCG GTC
Met Asn Thr Lys Gly Lys Ala Leu Ala Gly Leu Ile Ala Leu Ala Phe Ser Asn Met Ala Leu Ala Met Ala Asp Asp Ile Lys Val Ala Val
Met Lys Arg Asn Ala Lys Thr Ile Ile Ala Gly Met Ile Ala Leu Ala Ile Ser His Thr Ala Met Ala Asp Asp Ile Lys Val Ala Val
ATG AAA CGG AAT GCG AAA ACT ATC ATC GCA GGG ATG ATT GCA CTG GCA ATT TCA CAC ACC GCT ATG GCT GAC GAT ATT AAA GTC GCC GTT

          10          20          30
GTG GGC GCA ATG TCC GGT CCG GTT GCG CAG TAC GGT GAC CAG GAG TTT ACC GGC GCA GAG CAG GCG GTT GCG GAT ATC AAC GCT AAA GGC
Val Gly Ala Met Ser Gly Pro Val Ala Gln Tyr Gly Asp Gln Glu Phe Thr Gly Ala Glu Gln Ala Val Ala Asp Ile Asn Ala Lys Gly
Val Gly Ala Met Ser Gly Pro Ile Ala Gln Trp Gly Ile Met Glu Phe Asn Gly Ala Glu Gln Ala Ile Lys Asp Ile Asn Ala Lys Gly
GTC GGC GCG ATC TCC GGC CCG ATT GCC CAG TGG GGC ATA ATG GAA TTT AAC GGC GCG GAG CAG GCG ATT AAA GAC ATT AAT GCC AAA GGG

          40          50          60
GGC ATT AAA GGC AAC AAA CTG CAA ATC GCA AAA TAT GAC GAT GCC TGT GAT CCG AAA CAG GCG GTT GCG GTG GCG AAC AAC GTC GTT AAC
Gly Ile Lys Gly Asn Lys Lys Leu Gln Ile Ala Lys Tyr Asp Asp Ala Cys Asp Pro Lys Gln Ala Val Ala Val Ala Asn Lys Val Val Asn
Gly Ile Lys Gly Asp Lys Leu Val Gly Val Glu Tyr Asp Asp Ala Cys Asp Pro Lys Gln Ala Val Ala Val Ala Asn Lys Ile Val Asn
GGA ATT AAG GGC GAT AAA CTG GTT GGC GTG GAA TAT GAC GAC GCA TGC GAC CCG AAA CAA GCC GTT GCG GTC GCC AAC AAA ATT GTT AAT

          70          80          90
GAC GGC ATT AAA TAT GAT ATT GGT CAC CTC TGT TCC TCA TCA ACG CAG CCT GCG TCG GAT ATC TAC GAA GAC GAA GGC ATT TTG ATG ATC
Asp Gly Ile Lys Tyr Val Ile Gly His Leu Cys Ser Ser Ser Thr Gln Pro Ala Ser Asp Ile Tyr Glu Asp Gly Gly Ile Leu Met Ile
Asp Gly Ile Lys Tyr Val Ile Gly His Leu Cys Ser Ser Ser Thr Gln Pro Ala Ser Asp Ile Tyr Glu Asp Gly Gly Ile Leu Met Ile
GAC GGC ATT AAA TAC GTT ATT GGT CAT CTG TGT TCT TCT ACC CAG CCT GCG TCA GAC ATC TAT GAA GAC GAA GGT ATT CTA ATG ATC

          100          110          120
ACC CCA GCG GCA ACC GCG CCG GAG CTG ACC GCC CGT GGC TAT CAG CTG ATC CTG CGA ACC ACC GGC CTG GAT TCC GAC CAA GGC CCG ACG
Thr Pro Ala Ala Thr Ala Pro Glu Leu Thr Gln Arg Gly Tyr Gln His Ile Leu Arg Thr Thr Gly Leu Asp Ser Ser Ser Gln Gly Pro Thr
Ser Pro Gly Ala Thr Ala Pro Glu Leu Thr Gln Arg Gly Tyr Gln His Ile Met Arg Thr Ala Gly Leu Asp Ser Ser Ser Gln Gly Pro Thr
TCG CCG GGA GCG ACC GCG CCG GAA CTA ACC CAA CGC GGT TAT CAA CAC ATT ATG CGT ACT GCC GGG CTG GAC TCT TCC CAG GGG CCA ACG

          130          140          150
GCT GCC AAA TAT ATT CTT GAG AAA GTG AAA CCG CAG CGT ATT GCT ATC GTT CAC GAC AAA CAG CAA TAC GGC GAA GGT TCG GCG CGA GCG
Ala Ala Lys Tyr Ile Leu Glu Lys Val Lys Pro Gln Arg Ile Ala Ile Val His Asp Lys Gln Gln Tyr Gly Glu Gly Leu Ala Arg Ala
Ala Ala Lys Tyr Ile Leu Glu Thr Val Lys Pro Gln Arg Ile Ala Ile Ile His Asp Lys Gln Gln Tyr Gly Glu Gly Leu Ala Arg Ser
GCG GCA AAA TAC ATT CTT GAG ACG GTG AAG CCC CAG CGC ATC GCC ATC ATC CAG GAC AAA CAA CAG TAT GCG GAA GGC GTG GCG GGT TCG

          160          170          180
GTG CAG GAC GGC CTG AAG AAA GGC AAT GCA AAC GTG GTG TTC TTT GAT GGC ATC ACC GCC GGG GAA AAA GAT TTC TCA ACG CTG GTG GCG
Val Gln Asp Gly Leu Lys Lys Gly Asn Ala Asn Val Val Phe Phe Asp Gly Ile Thr Ala Gly Glu Lys Asp Phe Ser Ala Leu Ile Ala
Val Gln Asp Gly Leu Lys Ala Ala Asn Ala Asn Val Val Phe Phe Asp Gly Ile Thr Ala Gly Glu Lys Asp Phe Ser Ala Leu Ile Ala
GTG CAG GAC GGG CTG AAA GCG GCT AAC GCC AAC GTC GTC TTC TTC GAT GGT ATT ACC GCC GGG GAG AAA GAT TTC TCC GCG CTG ATC GCC

          190          200          210
CGT CTG AAA AAA GAG AAT ATC GAC TTC GTT TAC TAC GGC GGT TAT CAC CCG GAA ATG GGG CAA ATC CTG CGT GAC GCA CGC GCG GCA GGG
Arg Leu Lys Lys Glu Asn Ile Asp Phe Val Tyr Tyr Gly Gly Tyr Tyr Pro Glu Met Gly Gln Met Leu Arg Gln Ala Arg Ser Val Gly
CGC CTG AAA AAA GAA AAC ATC GAC TTC GTT TAC TAC GGC GGT TAC TAC CCG GAA ATG GGG CAG ATG CTG CCG CAG GCG GCT TCC GTT GGC

          220          230          240
CTG AAA ACT CAG TTT ATG GGG CCG GAA GGT GTG GCT AAC GTT TCG CTG TCT AAC ATT GGG GGC GAA TCA GCG GAA GGG CTA CTG GTG ACC
Leu Lys Thr Gln Phe Met Gly Pro Glu Gly Val Ala Asn Val Ser Leu Ser Asn Ile Ala Gly Glu Ser Ala Glu Gly Leu Leu Val Thr
Leu Lys Thr Gln Phe Met Gly Pro Glu Gly Val Gly Asn Ala Ser Leu Ser Asn Ile Ala Gly Asp Ala Glu Gly Met Leu Val Thr
CTG AAA ACC CAG TTT ATG GGG CCG GAA GGT GTG GGT AAC GCG TCG TTG TCG AAC ATT GCC GGT GAT GCC GCG GAA GGC ATG TTG GTC ACT

          250          260          270
AAA CCG AAG AAC TAC GAT CAG GTT CCG GCG AAC AAA CCC ATT GTT GAC GCG ATC AAA GCG AAA AAA CAG GAC CCA AGT GGC GCA TTC GTT
Lys Pro Lys Asn Tyr Asp Gln Val Pro Ala Asn Lys Pro Ile Val Asp Ala Ile Lys Ala Lys Lys Gln Asp Pro Ser Gly Ala Phe Val
Met Pro Lys Arg Tyr Asp Gln Asp Pro Ala Asn Gln Gly Ile Val Asp Ala Leu Lys Ala Asp Lys Lys Asp Pro Ser Gly Pro Tyr Val
ATG CCA AAA CCG TAT GAC CAG GAT CCG GCA AAC CAG GGC ATC GTT GAT GCG CTG AAA GCA GAC AAG AAA GAT CCG TCC GGG CCT TAT GTC

          280          290          300
TGG ACC ACC TAC GCC GCG CTG CAA TCT TTG CAG GCG GGC CTC AAT CAG TCT GAC GAT CCG GCT GAA ATC GCC AAA TAC CTG AAA
Trp Thr Thr Tyr Ala Ala Leu Gln Ser Leu Gln Ala Gly Leu Asn Gln Ser Asp Asp Pro Ala Glu Ile Ala Lys Tyr Leu Lys
Trp Ile Thr Tyr Ala Ala Val Gln Ser Leu Ala Thr Ala Leu Glu Arg Thr Gly Ser Asp Glu Pro Leu Ala Leu Val Lys Asp Leu Lys
TGG ATC ACC TAC GCG GCG GTG CAA TCT CTG GCG ACT GCC CTT GAG CGT ACC GGC AGC GAT GAG CCG CTG GCG CTG AAA GAT TTA AAA

          310          320          330
GCG AAC TCC GTG GAT ACC GTA ATG GGC CCG CTG ACC TGG GAT GAG AAA GGC GAT CTG AAA GGC TTT GAG TTC GGC GTA TTT GAC TGG CAC
Ala Asn Ser Val Asp Thr Val Met Gly Pro Leu Thr Trp Asp Glu Lys Gly Asp Leu Lys Gly Phe Glu Phe Gly Val Phe Asp Trp His
Ala Asn Gly Ala Asn Thr Val Ile Gly Pro Leu Asn Trp Asp Glu Lys Gly Asp Leu Lys Gly Phe Asp Phe Gly Val Phe Gln Trp His
GCT AAC GGT GCA AAC ACC GTG ATT GGG CCG CTG AAC TGG GAT GAA AAA GGC GAT CTT AAG GGA TTT GAT TTT GGT GTG TTC CAG TGG CAC

          340
GCC AAC GGC ACG GCC ACC GAT GCG AAG LIV-BP
Ala Asn Gly Thr Ala Thr Asp Ala Lys
Ala Asp Gly Ser Ser Thr Ala Ala Lys LS-BP
GCC GAC GGT TCA TCC ACG GCA GCC AAG
    
```

Fig. 2. Nucleotide sequences of the genes encoding LIV-BP and LS-BP. Positive numbering indicates the mature protein sequences, while negative numbering designates the signal sequences. Homologous amino acids are noted by asterisks.

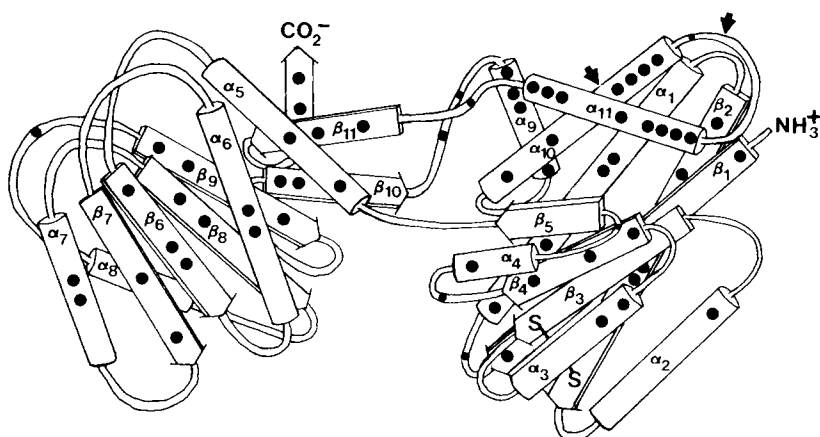


Fig. 3. Structure of the LIV-BP. Predicted secondary structures are superimposed on the three-dimensional structure reported by Saper and Quioco [13]. Filled circles indicate the positions of amino acid differences between the LIV-BP and LS-BP sequences. The two dark arrows indicate the region where a two-amino acid insertion occurs in the LS-BP sequence. A disulfide bridge is indicated between segments beta-2 and beta-3.

of the LIV-BP, it binds to the inside face of the N-terminal domain. It is reasonable to suggest [8] that amino acid differences between the two BP sequences within the cleft between the two domains are responsible for the different substrate specificities of the LIV-BP and LS-BP. Nonhomologous residues are present near the cleft in segments beta-1-alpha-1, beta-4-alpha-4, alpha-9, alpha-10, beta-9, and beta-10. Thus, these areas of the BPs may be involved either in forming the structure of the substrate binding site or in directly interacting with the substrate.

INTERACTION OF BPS WITH MEMBRANE COMPONENTS

Both the LIV-BP-substrate complex and the LS-BP-substrate complex must interact with the same three membrane components to effect the transport of the substrate. Therefore, it is interesting to note that some of the regions of high amino acid sequence conservation include the C terminus of beta-2, alpha-2, beta-3, alpha-3, the C terminus of beta-8, and alpha-8. Because of their sequence conservation and position on the BPs, these regions may function in interactions with the membrane components. It is tempting to speculate that this interaction may be the signal that initiates the membrane translocation process.

SECRETION OF THE BPS INTO THE PERIPLASM

Both the LIV-BP and LS-BP are synthesized as precursors with a 23-amino acid signal sequence that is processed in association with the BP secretion into the periplasm [7, 10, 14]. Using the rules of Chou and Fasman [15], the secondary structure of the N-terminal portion of the precursor would be predicted to form a helical hairpin structure. These two helices would form from the signal sequence and the mature N-terminal regions and would contain hydrophobic amino acids in the sections that might span the membrane. A model for the secretion of the BPs is shown in Figure

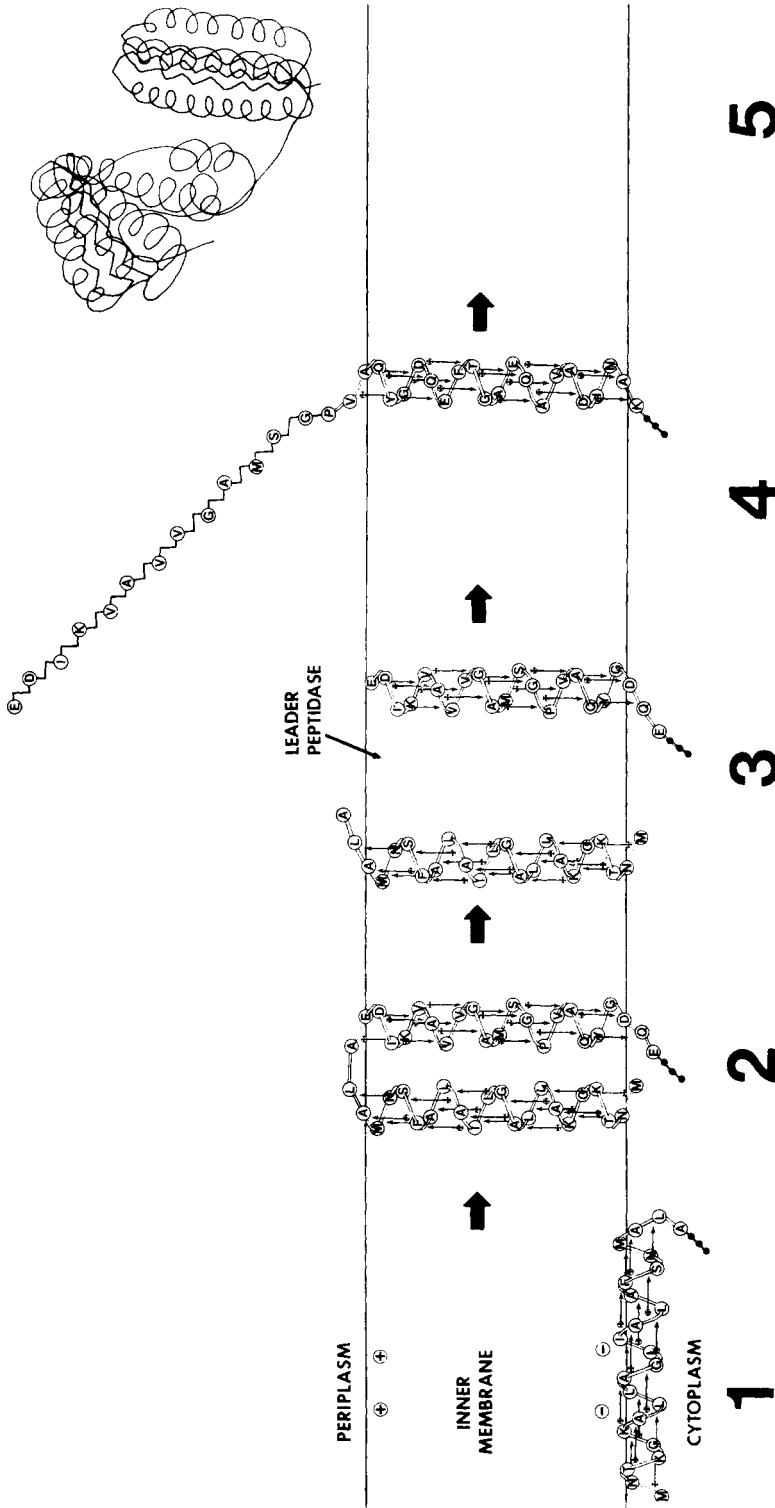


Fig. 4. Model for the export of the periplasmic leucine-binding proteins. The amino acids are represented by single-letter codes in the circles. The small arrows indicate the directions (positive to negative) of helix dipoles. Step 1 represents the initial association of the BP with the inner membrane. Step 2 indicates the formation of a helical hairpin structure between the leader peptide and the N-terminal portion of the mature protein. This helical hairpin assumes a transmembrane orientation in the presence of a transmembrane potential. Step 3 indicates the cleavage of the leader sequence by the leader peptidase. Finally, steps 4 and 5 correspond to a refolding of the mature protein and its translocation into the periplasmic space.

4. In this model, a major role of the membrane potential is suggested for the proper membrane orientation of the proposed helix hairpin formed by the signal sequence and the first segment of the mature portion of the BP [16,17].

Recent work on the secretion of the BPs has been directed toward identifying the sequences required to achieve secretion. To this end, mutants of LS-BP carrying various deletions or sequence alterations have been constructed [17]. It was found that these altered binding proteins were processed and secreted normally. Moreover, hybrid molecules were constructed between the tryptophan synthetase alpha-subunit and various lengths of the N-terminal portion of the LS-BP [18]. The results showed that almost the entire C-terminal half of the LS-BP can be removed with no adverse effect on processing or secretion of the hybrid protein. Thus, the signal sequence together with a portion of the N-terminal portion of the LS-BP is sufficient to direct its export into the periplasmic space.

THE BPS AS MODELS IN STUDYING STRUCTURE-FUNCTION RELATIONSHIPS

Because the LIV-BP and LS-BP have three known functions in *E coli*, they are proving useful in examining the relationships between protein structure and function. For example, using oligonucleotide-directed mutagenesis, the nonhomologous residues in the clefts of the BPs are being altered. If an alteration corresponds with a change in either substrate affinities or substrate specificities, then knowledge will be gained about the mechanism of substrate binding. Other changes are being directed toward the regions of high amino acid homology on the BPs that are potential regions for interacting with the membrane components. These altered BPs will be especially useful for study when the individual membrane proteins have been isolated. Finally, work in progress also involves analyzing additional LS-BP-trpA hybrid molecules containing larger C-terminal deletions to identify more precisely the N-terminal sequences required for export of the BPs into the periplasmic space. Since the LIV-BP and LS-BP are the products of gene duplication, a comparison of their structure-function relationships should help illuminate the potential constraints placed on their evolutionary divergence.

In conclusion, the leucine binding proteins of *E coli* have been thoroughly studied by biochemical, genetic, and recombinant DNA approaches. As a consequence, much is already known about their molecular structure and their relationship to each other. Because these proteins must be secreted and interact with the same membrane components yet exhibit divergent substrate specificities, they provide a model system for studying the relationship of protein structure to protein function.

REFERENCES

1. Piperno JR, Oxender DL: J Biol Chem 243:5914-5920, 1968.
2. Rahmanian M, Claus DR, Oxender DL: J Bacteriol 116:1258-1266, 1973.
3. Wood JM: J Biol Chem 250:4477-4485, 1975.
4. Anderson JJ, Oxender DL: J Bacteriol 136:168-174, 1978.
5. Oxender DL, Anderson JJ, Daniels CJ, Landick R, Gunsalus RP, Zurawski G, Selker E, Yanofsky C: Proc Natl Acad Sci USA 77:1412-1416, 1980.
6. Nazos PN: Ph.D. Thesis, University of Michigan, Ann Arbor, 1984.
7. Landick R: Ph.D. Thesis, University of Michigan, Ann Arbor, 1983.

216:JCB Antonucci, Landick, and Oxender

8. Landick R, Oxender DL: *J Biol Chem*, 260:8257–8261, 1985.
9. Landick R, Anderson JJ, Mayo MM, Gunsalus RP, Mavromara P, Daniels CJ, Oxender DL: *J Supramolec Struct* 14:527–537, 1981.
10. Oxender DL, Anderson JJ, Daniels CJ, Landick R, Gunsalus RP, Zurawski G, Yanofsky C: *Proc Natl Acad Sci USA* 77:2005–2009, 1980.
11. Riley M, Aninlionis A: *Annu Rev Microbiol* 32:519–560, 1978.
12. Ovchinnikov YA, Aldanova NA, Grinkevich VA, Arzamazova NM, Moroz IN, Nazimov IV: *FEBS Lett* 78:313–316, 1977.
13. Saper M, Quioco FA: *J Biol Chem* 258:11057–11062, 1983.
14. Daniels CJ, Anderson JJ, Landick R, Oxender DL: *J Supramolec Struct* 14:305–311, 1980.
15. Chou TY, Fasman GD: *Annu Rev Biochem* 47:251–276, 1978.
16. Daniels CJ, Bole DG, Quay SC, Oxender DL: *Proc Natl Acad Sci USA* 78:5396–5400, 1981.
17. Landick R, Duncan JR, Copeland BR, Nazos PM, Oxender DL: *J Cell Biochem* 24:331–344, 1984.
18. Su TZ, Copeland BR, Landick R, Oxender DL: *J Biol Chem* 1985, (submitted).