JOURNAL OF BACTERIOLOGY, Mar. 1985, p. 904–908
0021-9193/85/030904-05$02.00/0
Copyright © 1985, American Society for Microbiology
Vol. 161, No. 3

Nucleotide Sequence of the Gene for the Vitamin B_{12} Receptor Protein in the Outer Membrane of *Escherichia coli*

KNUT HELLER† AND ROBERT J. KADNER*

Department of Microbiology, The University of Virginia School of Medicine, Charlottesville, Virginia 22908

Received 10 September 1984/Accepted 5 December 1984

The nucleotide sequence of a 2220-base-pair fragment containing the *btuB* gene of *Escherichia coli* was determined. There was a single open reading frame which was translated into a 614-amino-acid polypeptide, the first 20 amino acids of which comprised a typical leader sequence. The putative mature or processed form had a molecular weight (66,400) and a composition in close agreement with that determined for the purified receptor. The distribution of amino acids in the receptor protein was similar to that of other outer membrane proteins, showing a fairly even distribution of charged residues and the absence of extensive hydrophobic stretches. The *btuB451* mutation appears to alter the receptor to eliminate its ability to function in vitamin B_{12} uptake without affecting its ligand binding properties. The sequence of the DNA from this mutant was determined and revealed a leucine-to-proline (C-to-T transition) change in the eighth amino acid of the mature form.

The uptake of vitamin B_{12} in *Escherichia coli* is initiated by the binding of vitamin B_{12} to a receptor protein in the outer membrane. Release of vitamin B_{12} into the periplasm or cytoplasm is dependent on the function of the *tonB* product, which is also necessary for the energy-dependent phases of outer membrane-dependent transport processes (20, 23). The structure of the vitamin B_{12} receptor is of interest for several reasons. Phage BF23, colicin A, and the E colicins bind to the vitamin B_{12} receptor competitively with each other and with vitamin B_{12} (4, 7). The location on the receptor protein of the binding sites for these ligands and the relationship between these binding sites might help explain the nature of this competitive binding by such different types of molecules. There may also be a binding site on the receptor for the *tonB* product to mediate the energy coupling of the transport process. The receptor does appear to carry out a transport process different from that mediated by the porins, because vitamin B_{12} uptake shows considerable substrate specificity and energy dependence. Comparison of the structure and topology of the vitamin B_{12} receptor with those of the porins and other outer membrane proteins might reveal the structural features common to all outer membrane proteins and necessary for their export and maintenance in that membrane.

To study the structure of the receptor, its gene, *btuB*, was cloned on a 2.3-kilobase (kb) fragment (11). In this study is reported the nucleotide sequence of the fragment and the predicted amino acid sequence of the receptor protein.

A mutant has been described (1, 15) in which vitamin B_{12} transport was completely defective, although the binding of vitamin B_{12} and all other ligands to the receptor was normal in affinity and amount. The receptor functioned normally for transmission of the E colicins and phage BF23, and vitamin B_{12} still protected cells from their lethality. Complementation tests indicated that the mutation in this strain was in *btuB*, and the suggestion was made that this mutation affected the domain on the receptor which was responsible for the response to the *tonB* product (3). The interaction of the *tonB* product with any of the receptor proteins dependent on its presence has not been demonstrated, and it is premature to claim that this *btuB451* mutation affects the TonB binding site, rather than a site involved in vitamin B_{12} translocation across the outer membrane. Whatever the biochemical basis for the phenotype of this mutant, the mutational lesion was determined by nucleotide sequencing. A single amino acid substitution near the amino terminus of BtuB was found.

MATERIALS AND METHODS

Bacterial strains and plasmids. Most of the bacterial strains and plasmids were described in the accompanying paper (11). Plasmid pKH3-3 carries a 2.3-kb insert which contains the *btuB* gene and was generated by partial *Sau*3A digestion; this plasmid was the source of DNA for determination of the sequence of the wild-type *btuB* gene. The region carrying the *btuB451* mutation was obtained as follows. Strain RK4775 (RK5173 *btuB451*) was lysogenized with phage λ d*arg*13 ($btuB^+$) with selection for growth on 5 nM vitamin B_{12}. A phage lysate from pooled lysogens was prepared by thermal induction and lysis by the addition of $CHCl_3$. The lysate was used to infect strain RK4936A [Δ(*arg-btuB*)], which was obtained after excision of *btuB*::Tn*10*. arg^+ transductants were selected at 30°C. All Arg^+ transductants were sensitive to phage BF23, and approximately 10% were unable to utilize 5 nM vitamin B_{12}, i.e., carried *btuB451*. One of these transductants produced phage upon induction and $CHCl_3$ lysis and was used to prepare phage DNA. This phage DNA was digested with *Hin*dIII and *Sal*I, and the appropriate 163-base-pair (bp) fragment was cloned into phage M13mp8.

Genetic techniques. Restriction endonuclease digestion, ligation, and transformation were described in the accompanying paper (11). The location of the wild-type allele of the *btuB451* mutation was determined by introducing into strain RK4775 (*btuB451*) derivatives of plasmid pKH3-5 from which portions had been removed by restriction endonuclease-generated deletion. The ability of these transformants to give rise to Btu^+ recombinants was determined.

DNA sequence determination. DNA sequence analysis was performed by the enzymatic method of Sanger et al. (26), using an $\alpha^{35}S$-labeled deoxynucleoside triphosphate for la-

* Corresponding author.
† Present address: Lehrstuhl Mikrobiologie, Universität Tübingen, Tübingen, Federal Republic of Germany.

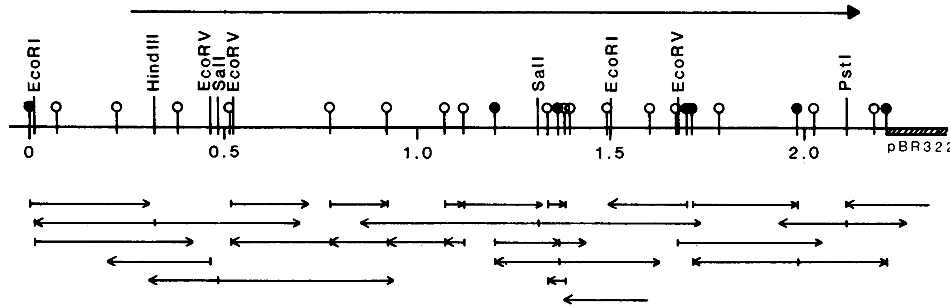

FIG. 1. Restriction map and sequencing strategy of the *btuB* region. The 2.3-kb insert in plasmid pKH3-3 is shown with the restriction sites used for sequencing. The filled circles are sites for *Sau*3A; the open circles are *Hpa*II. The arrow represents the location and direction of the BtuB reading frame. Below the restriction map are the fragments from which the nucleotide sequence was read.

beling (2). Cloning of fragments into M13 derivatives and their use as templates for DNA sequence analysis was performed as described by Messing and Vieira (17). Sequence analysis was performed with the aid of the computer program described by Staden (28).

RESULTS

DNA sequence of *btuB*. The nucleotide sequence of the 2220-bp fragment carrying *btuB* was determined by cloning smaller restriction fragments of the insert into the mp8 or mp9 derivatives of phage M13 (17). The restriction map and regions sequenced are shown in Fig. 1. The nucleotide sequence and its translation are presented in Fig. 2. The sequence of both DNA strands was determined, except for regions of 28 residues (residues 491 to 518), 144 residues (residues 931 to 1075), and 10 residues (residues 1703 to 1712). In these three regions, several restriction fragments were sequenced, and no uncertainties were encountered. Examination of the six possible reading frames within the 2220-bp fragment revealed only one of significant length, extending from residues 195 to 2153. There were two potential ATG initiation codons, located at residues 270 and 312. It is likely that the codon at position 312 is the actual initiation site for the vitamin B_{12} receptor because it is preceded by a potential Shine-Dalgarno sequence (GTG GATG) (27) and has the sequence characteristics of translation initiation sites (29). This 1842-bp open reading frame encodes a polypeptide with 614 amino acids, the first 20 of which are very similar to the leader peptide (signal sequence) of other outer membrane proteins (18).

The putative leader peptide was followed by a polypeptide with a calculated molecular weight of 66,400, which is in good agreement with the size of the vitamin B_{12} receptor determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (11, 12, 25). The composition of the polypeptide determined from the DNA sequence was in excellent agreement with the amino acid composition determined for the purified receptor (12). The overall sequence of the protein was typical of other outer membrane proteins, in that it was not appreciably nonpolar (22% charged residues) and that there were only a few long stretches composed only of hydrophobic amino acids. There were no cysteine and only two methionine residues in the mature region. Tryptophan (2.3%), tyrosine (8.0%), histidine (1.8%), and threonine (9.1%) were present more frequently than in total *E. coli* proteins by at least a factor of two.

The codon usage for the precursor vitamin B_{12} receptor was very similar to the codon usage pattern of other weakly expressed *E. coli* proteins (10). All but 4 of the possible 60 codons (only one Cys residue is present) were used. This was in contrast to the situation with the more abundant outer membrane proteins which employ a limited range of codons corresponding to the more abundant tRNA species (6, 13).

The direction of translation of the *btuB* gene determined from the sequence agreed with that proposed from the properties of Tn*1000* insertions, i.e., from the *Hin*dIII site to the *Pst*I site (11). The promoter and upstream regulatory regions were carried on the sequenced fragment because the expression of the cloned gene was still subject to repression by vitamin B_{12} (11, 14). Four regions of palindromic structure were found in the region 5′ to the coding segment (24), but their role cannot be assessed until the start site for transcription has been determined.

Sequence of the *btuB451* mutation. The *btuB451* mutation confers the BtuI (1), or BtuA (15) phenotype, in which the mutant strain lacks vitamin B_{12} uptake activity, although all other receptor functions remain normal. Plasmid pKH3-5 was able to complement the defect in vitamin B_{12} utilization in both *rec*$^+$ and *recA* strains carrying the *btuB451* mutation. Derivatives of this plasmid carrying deletions between restriction sites in *btuB* were constructed and introduced into the *rec*$^+$ *btuB451* strain. The formation of Btu$^+$ recombinants was determined (Fig. 3). Several independent isolates of each construction were examined. Plasmids in which the *Hin*dIII fragment (e.g., pKH35-H1) or both *Sal*I fragments (e.g., pKH35-S1) were deleted did not give rise to *Btu*$^+$ recombinants. In contrast, plasmids deleted for the smaller *Sal*I fragment (e.g., pKH35-S2) did give Btu$^+$ recombinants. This marker rescue experiment indicates that the wild-type allele of *btuB451* resides within the small segment between the *Hin*dIII and *Sal*I sites.

The *btuB451* mutation was transferred onto phage λ d*arg*13 by homologous recombination. The 163-bp *Hin*dIII-*Sal*I fragment was isolated from a digest of this phage DNA and was cloned into M13mp8. The nucleotide sequence of this fragment was determined in parallel with the corresponding fragment from the wild type (Fig. 4). The mutant differed from the wild type at only a single position. The mutation was associated with a T-to-C transition at residue 394 in Fig. 2. This nucleotide change predicts the substitution of proline for leucine at the eighth amino acid residue of the mature polypeptide (amino acid 28 in Fig. 2).

DISCUSSION

The nucleotide sequence of the 2.2-kb fragment, shown in the accompanying paper to contain the *btuB* gene (11), allowed prediction of the amino acid sequence of the vitamin B_{12} receptor protein. The composition of this polypeptide agreed very well with that reported for the colicin E3 receptor by Imajoh et al. (12). The size of the translated

```
                                                                                                   100
GATCTTAGATGAATTCCTATTTGTGAGCTACGTCTGGACAGTAACTTGTTACAACCTGTAGCATCCACTTGCCGGTCCTGTGAGTTAATAGGGAATCCAGTGCGAATCTGGAGCTGACGC
                                                                                 200
GCAGCGGTAAGGAAAGGTGCGATGATTGCGTTATGCGGACACTCGCCATTCGGTGGGAAGTCATCATCTCTTAGTATCTTAGATACCCCTCCAAGCCCGAAGACCTGCCGGCCAACGTCG
                                                             300
CATCTGGTTCTCATCATCGCGTAATATTGATGAAACCTGCGGCATCCTTCTTCTATTGTGGATGCTTTACA ATG ATT AAA AAA GCT TCG CTG CTG ACG GCG TGT TCC
                                                                        Met Ile Lys Lys Ala Ser Leu Leu Thr Ala Cys Ser
                                                            400
GTC ACG GCA TTT TCC GCT TGG GCA CAG GAT ACC AGC CCG GAT ACT CTC GTC GTT ACT GCT AAC CGT TTT GAA CAG CCG CGC AGC ACT GTG
Val Thr Ala Phe Ser Ala Trp Ala Gln Asp Thr Ser Pro Asp Thr Leu Val Val Thr Ala Asn Arg Phe Glu Gln Pro Arg Ser Thr Val
                                                                        500
CTT GCA CCA ACC ACC GTT GTG ACC CGT CAG GAT ATC GAC CGC TGG CAG TCG ACC TCG GTC AAT GAT GTG CTG CGC CGT CTT CCG GGC GTC
Leu Ala Pro Thr Thr Val Val Thr Arg Gln Asp Ile Asp Arg Trp Gln Ser Thr Ser Val Asn Asp Val Leu Arg Arg Leu Pro Gly Val
                                                                                    600
GAT ATC ACC CAA AAC GGC GGT TCA GGT CAG CTC TCA TCT ATT TTT ATT CGC GGT ACA AAT GCC AGT CAT GTG TTG GTG TTA ATT GAT GGC
Asp Ile Thr Gln Asn Gly Gly Ser Gly Gln Leu Ser Ser Ile Phe Ile Arg Gly Thr Asn Ala Ser His Val Leu Val Leu Ile Asp Gly
                                                                                                700
GTA CGC CTG AAT CTG GCG GGG GTG AGT GGT TCT GCC GAC CTT AGC CAG TTC CCT ATT GCG CTT GTC CAG CGT GTT GAA TAT ATC CGT GGG
Val Arg Leu Asn Leu Ala Gly Val Ser Gly Ser Ala Asp Leu Ser Gln Phe Pro Ile Ala Leu Val Gln Arg Val Glu Tyr Ile Arg Gly

CCG CGC TCC GCT GTT TAT GGT TCC GAT GCA ATA GGC GGG GTG GTG AAT ATC ATC ACG ACG CGC GAT GAA CCC GGA ACG GAA ATT TCA GGA
Pro Arg Ser Ala Val Tyr Gly Ser Asp Ala Ile Gly Gly Val Val Asn Ile Ile Thr Thr Arg Asp Glu Pro Gly Thr Glu Ile Ser Gly
800
GGG TGG GGA AGC AAT AGT TAT CAG AAC TAT GAT GTC TCT ACG CAG CAA CAA CTG GGG GAT AAG ACA CGG GTA ACG CTG TTG GGC GAT TAT
Gly Trp Gly Ser Asn Ser Tyr Gln Asn Tyr Asp Val Ser Thr Gln Gln Gln Leu Gly Asp Lys Thr Arg Val Thr Leu Leu Gly Asp Tyr
            900
GCC CAT ACT CAT GGT TAT GAT GTT GTT GCC TAT GGT AAT ACC GGA ACG CAA GCG CAG ACA GAT AAC GAT GGT TTT TTA AGT AAA ACG CTT
Ala His Thr His Gly Tyr Asp Val Val Ala Tyr Gly Asn Thr Gly Thr Gln Ala Gln Thr Asp Asn Asp Gly Phe Leu Ser Lys Thr Leu
                    1000
TAT GGC GCG CTG GAG CAT AAC TTT ACT GAT GCC TGG AGC GGC TTT GTG CGC GGC TAT GGC TAT GAT AAC CGT ACC AAT TAT GAC GCG TAT
Tyr Gly Ala Leu Glu His Asn Phe Thr Asp Ala Trp Ser Gly Phe Val Arg Gly Tyr Gly Tyr Asp Asn Arg Thr Asn Tyr Asp Ala Tyr
                             1100
TAT TCT CCC GGT TCA CCG TTG CTC GAT ACC CGT AAA CTC TAT AGC CAA AGT TGG GAC GCC GGG CTG CGC TAT AAC GGC GAA CTG ATT AAA
Tyr Ser Pro Gly Ser Pro Leu Leu Asp Thr Arg Lys Leu Tyr Ser Gln Ser Trp Asp Ala Gly Leu Arg Tyr Asn Gly Glu Leu Ile Lys
                                      1200
TCA CAA CTC ATT ACC AGC TAT AGC CAT AGC AAA GAT TAC AAC TAC GAT CCC CAT TAT GGT CGT TAT GAT TCG TCG GCG ACG CTC GAT GAG
Ser Gln Leu Ile Thr Ser Tyr Ser His Ser Lys Asp Tyr Asn Tyr Asp Pro His Tyr Gly Arg Tyr Asp Ser Ser Ala Thr Leu Asp Glu
                                                1300
ATG AAG CAA TAC ACC GTC CAG TGG GCA AAC AAT GTC ATC GTT GGT CAC GGT AGT ATT GGT GCG GGT GTC GAC TGG CAG AAA CAG ACT ACG
Met Lys Gln Tyr Thr Val Gln Trp Ala Asn Asn Val Ile Val Gly His Gly Ser Ile Gly Ala Gly Val Asp Trp Gln Lys Gln Thr Thr
                                                            1400
ACG CCG GGT ACA GGT TAT GTT GAG GAT GGA TAT GAT CAA CGT AAT ACC GGC ATC TAT CTG ACC GGG CTG CAA CAA GTC GGC GAT TTT ACC
Thr Pro Gly Thr Gly Tyr Val Glu Asp Gly Tyr Asp Gln Arg Asn Thr Gly Ile Tyr Leu Thr Gly Leu Gln Gln Val Gly Asp Phe Thr
                                                                        1500
TTT GAA GGC GCC AGA CGC AGT GAC GAT AAC TCA CAG TTT GGT CGT CAT GGA ACC TGG CAA ACC AGC GCC GGT TGG GAA TTC ATC GAA GGT
Phe Glu Gly Ala Arg Arg Ser Asp Asp Asn Ser Gln Phe Gly Arg His Gly Thr Trp Gln Thr Ser Ala Gly Trp Glu Phe Ile Glu Gly
                                                                                    1600
TAT CGC TTC ATT GCT TCC TAC GGG ACA TCT TAT AAG GCA CCA AAT CTG GGG CAA CTG TAT GGC TTC TAC GGA AAT CCG AAT CTG GAC CCG
Tyr Arg Phe Ile Ala Ser Tyr Gly Thr Ser Tyr Lys Ala Pro Asn Leu Gly Gln Leu Tyr Gly Phe Tyr Gly Asn Pro Asn Leu Asp Pro

GAG AAA AGC AAA CAG TGG GAA GGC GCG TTT GAA GGC TTA ACC GCT GGG GTG AAC TGG CGT ATT TCC GGA TAT CGT AAC GAT GTC AGT GAC
Glu Lys Ser Lys Gln Trp Glu Gly Ala Phe Glu Gly Leu Thr Ala Gly Val Asn Trp Arg Ile Ser Gly Tyr Arg Asn Asp Val Ser Asp
1700
TTG ATC GAT TAT GAT GAT CAC ACC CTG AAA TAT TAC AAC GAA GGG AAA GCG CGG ATT AAG GGC GTC GAG GCG ACC GCC AAT TTT GAT ACC
Leu Ile Asp Tyr Asp Asp His Thr Leu Lys Tyr Tyr Asn Glu Gly Lys Ala Arg Ile Lys Gly Val Glu Ala Thr Ala Asn Phe Asp Thr
        1800
GGA CCA CTG ACG CAT ACT GTG AGT TAT GAT TAT GTC GAT GCG CGC AAT GCG ATT ACC GAC ACG CCG TTG TTA CGC CGT GCT AAA CAG CAG
Gly Pro Leu Thr His Thr Val Ser Tyr Asp Tyr Val Asp Ala Arg Asn Ala Ile Thr Asp Thr Pro Leu Leu Arg Arg Ala Lys Gln Gln
                1900
GTG AAA TAC CAG CTC GAC TGG CAG TTG TAT GAC TTC GAC TGG GGT ATT ACT TAT CAG TAT TTA GGC ACT CGC TAT GAT AAG GAT TAC TCA
Val Lys Tyr Gln Leu Asp Trp Gln Leu Tyr Asp Phe Asp Trp Gly Ile Thr Tyr Gln Tyr Leu Gly Thr Arg Tyr Asp Lys Asp Tyr Ser
                            2000
TCT TAT CCT TAT CAA ACC GTT AAA ATG GGC GGT GTG AGC TTG TGG GAT CTT GCG GTT GCG TAT CCG GTC ACC TCT CAC CTG ACA GTT CGT
Ser Tyr Pro Tyr Gln Thr Val Lys Met Gly Gly Val Ser Leu Trp Asp Leu Ala Val Ala Tyr Pro Val Thr Ser His Leu Thr Val Arg
                                        2100
GGT AAA ATA GCC AAC CTG TTC GAC AAA GAT TAT GAG ACA GTC TAT GGC TAC CAA ACT GCA GGA CGG GAA TAC ACC TTG TCT GGC AGC TAC
Gly Lys Ile Ala Asn Leu Phe Asp Lys Arg Tyr Glu Thr Val Tyr Gly Tyr Gln Thr Ala Gly Arg Glu Tyr Thr Leu Ser Gly Ser Tyr
                                          2200
ACC TTC TGA ACCACGTCCCACCGTGCTGGTGTTTGACTCCGGCGTCGGTGGGTTGTCGGTCTATGACGAGATC
Thr Phe ***
```

FIG. 2. Nucleotide sequence of the *btuB* region.

polypeptide was close to that observed for the vitamin B_{12} receptor, and the polypeptide started with a typical leader, or signal, sequence. Although the amino-terminal sequence of the mature polypeptide in the outer membrane has not yet been determined, analogy with the sites of leader peptidase action on other secreted proteins suggests that this cleavage occurs after the sequence Ala_{18}-Trp_{19}-Ala_{20} (18).

The vitamin B_{12} receptor is similar in several respects to other outer membrane proteins (6, 13, 19, 22). Like them,

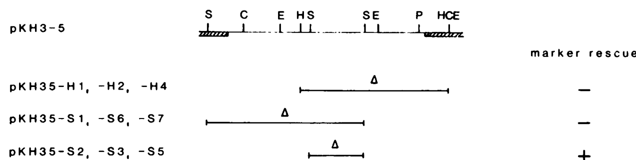

FIG. 3. Localization of the *btuB451* mutation by marker rescue. Plasmids carrying deletions in the *btuB* region were generated by cleavage with the indicated restriction enzyme and treatment with DNA ligase. Plasmids deleted for the regions shown were tested for their ability to give rise to Btu^+ recombinants in a strain carrying *btuB451*. Restriction enzyme cleavage sites are as follows: C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sal*I.

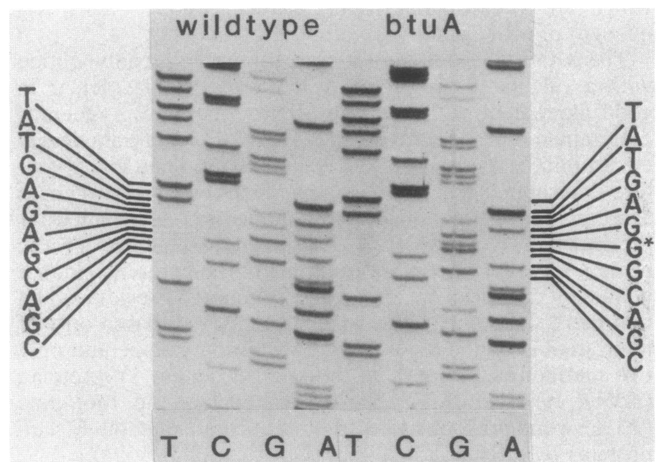

FIG. 4. Determination of the nucleotide change in the *btuB451* mutation. A portion of the sequencing gel obtained with the 163-bp *Hin*dIII-*Sal*I fragments from the wild-type (left) and the *btuB451* mutant (right) is shown. The region covers nucleotides 369 to 416 on the 3′ strand.

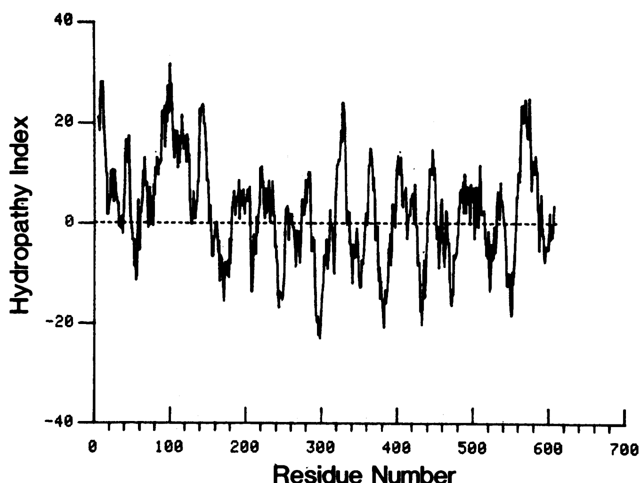

FIG. 5. Hydropathy profile of the BtuB protein. The hydropathy index was determined at a span setting of 13 with the algorithm presented by Kyte and Doolittle (16). Hydrophobic regions extend above the dotted line.

BtuB contains an appreciable content of charged amino acids (22%) and does not exhibit extensive segments of hydrophobic residues. Charged residues were evenly spaced and provide negative charge along the length of the protein. There were seven stretches with 13 or more uncharged amino acids. The hydrophobicity profile (16) for the protein confirmed that the protein was not appreciably nonpolar in overall composition, although there were two regions of significant hydrophobic character, one near the amino terminus (residues 80 to 160) and another near the carboxy terminus (residues 570 to 600) (Fig. 5). The remainder of the protein revealed alternating regions of polar and nonpolar character. It is possible that the polypeptide chain spans the outer membrane bilayer numerous times. Polar or charged residues were interspersed throughout these hydrophobic regions, showing that there are no long nonpolar α-helical segments spanning the bilayer, as is seen with some transmembrane proteins in the cytoplasmic membrane of bacteria or in the membranes of eucaryotic cells or viruses (8).

Chou-Fasman analysis (5) of secondary structure predicted extensive regions of β-structure comprising 55 to 60% of the length of the mature polypeptide (data not shown). The predicted content of α-helical structure was much lower, comprising the signal sequence and 6 to 10% of the mature portion. A high content of β-sheets has been found in other outer membrane proteins (9) and may be a general feature of them.

Considerable amino acid sequence homology exists within the family of porins OmpC, OmpF, and PhoE (13, 19, 22), but much less homology is apparent between the porins and other outer membrane proteins OmpA and LamB (6). Nikaido and Wu (21) have described short regions of partial homology that are located at similar, but not identical, sites within all these proteins. Visual comparison of the BtuB sequence with those of other outer membrane proteins revealed no appreciable overall homology and only weak homology to some of the common consensus regions found by Nikaido and Wu (21). Examples are presented in Fig. 6. Perhaps the best fit was to homology region b, which was suggested to be involved in the protein export process. Whereas this region lies near the amino terminus of the mature sequence of other outer membrane proteins, in BtuB it was near the carboxyl end. Since the match in these regions is weak, proposals concerning their function would be of questionable significance. The major conclusion is that there are no strongly conserved sequences common to outer membrane proteins.

Since this is, to our knowledge, the first report of the sequence of a *tonB*-dependent or a vitamin B_{12}-binding protein, it is not possible to use homologies to identify possible functional domains. However, examination of the

Region b	PhoE:	27	Ala	Ser	Lys	Asp		Gly	Asp	Gln	Ser	Tyr	Ile	Arg	Phe	Gly		
	OmpC:	29	Lys	Asp	Val	Asp		Gly	Asp	Gln	Thr	Tyr	Met	Arg	Leu	Gly		
	OmpF:	32	Tyr	Gly		Gly	Asn	Gly	Asp	Met	Thr	Tyr	Ala	Arg	Leu	Gly		
	LamB:	32	Tyr	Arg	Leu	Gly	Asn	Glu	Cys	Glu	Thr	Tyr	Ala	Glu	Leu	Lys		
	OmpA:	1	Ala	Pro	Lys	Asp	Asn	Thr	Trp	Tyr	Thr	Gly	Ala	Lys	Leu	Gly		
	BtuB:	512	Tyr	Asp	Phe	Asp	Trp	Gly		Ile	Thr	Tyr	Gln	Tyr	Leu	Gly		
Region d	PhoE:	144	Gly		Leu	Asn	Leu	Thr	Leu	Gln	Tyr	Gln	Gly	Lys		Asn	Glu	Asn
	OmpC:	142	Gly		Leu	Asn	Phe	Ala	Val	Gln	Tyr	Gln	Gly	Lys		Asn	Gly	Asn
	OmpF:	150	Gly		Leu	Asn	Phe	Ala	Val	Gln	Tyr	Leu	Gly	Lys		Asn	Glu	
	LamB:	185	Gly	Thr	Leu	Glu	Leu	Gly	Val	Asp	Tyr		Gly	Arg	Ala	Asn	Leu	
	OmpA:	160	Gly	Met	Leu	Ser	Leu	Gly	Val	Ser	Tyr			Arg	Phe		Gly	Gln
	BtuB:	399	Gly	Gln	Leu		Tyr	Gly	Phe		Tyr		Gly	Asn	Pro	Asn	Leu	
	OmpF:	86	Ala	Gly	Leu		Lys		Tyr	Ala	Asp	Val	Gly	Ser	Phe	Asp	Tyr	Gly
	LamB:	127	Ala	Gly	Leu		Glu		Asn	Ile	Asp	Val	Gly		Phe			Gly
	OmpA:	41	Gly	Gly		Tyr	Gln	Val	Asn	Pro	Tyr	Val	Gly		Phe	Glu	Met	Gly
	BtuB:	338	Thr	Gly	Ile	Tyr	Leu	Gly	Leu	Gln	Gln	Val	Gly	Asp	Phe			
							Thr											

FIG. 6. Homology between BtuB and other outer membrane proteins. The regions of homology (b and d) in the major outer membrane proteins are those identified by Nikaido and Wu (21). Alignments to provide best fit with these regions are shown and include deletion or insertion of an amino acid. The coordinates are for the processed form of the protein.

distribution of certain amino acids throughout the sequence revealed sections which might be ligand-binding sites. Particularly noteworthy was the case of tyrosine. There were six 20 to 25 amino acid regions which contained at least four Tyr residues (amino acids 241 to 256, 287 to 307, 403 to 425, 455 to 474, 539 to 558, and 592 to 613). In comparison, the porins have roughly the same tyrosine content (13, 19, 22) but, at most, only two tyrosine clusters of the type seen in BtuB. Five of the six regions were very hydrophilic in character and contained at least five charged or hydroxyl-containing residues. One of the clusters was at the carboxy terminus of the protein, which has been implicated in vitamin B_{12} binding based on the altered vitamin B_{12} binding and transport properties of the mutant generated by deletion from the *Pst*I site (11). This altered receptor did allow entry of phage BF23, but the binding of this phage was not blocked by vitamin B_{12}.

The six tyrosine-rich regions were compared for homology at the nucleotide and amino acid levels. Although the clusters were rich in Asp, Gly, and Ser, the longest amino acid sequence shared by any two clusters was only three amino acids long. There was no significant homology at the nucleic acid level. Thus, it is unlikely that these clusters arose by genetic duplication. Experiments to identify binding domains by isolation of mutations in the cloned gene are in progress.

Also of interest is the amino-terminal region of the putative processed protein. This region was affected by the *btuB451* mutation, which blocked vitamin B_{12} uptake but not any other receptor functions (1). One possibility is that this domain is involved in interaction with the *tonB* product. Comparison of this region with that of other *tonB*-dependent receptors should prove to be useful. If this domain is involved in interaction with TonB, then this region should face the periplasm. Future studies are directed toward examination of the topology of the BtuB protein in the membrane and its interaction with the *tonB* product.

ACKNOWLEDGMENTS

We are extremely indebted to M. M. Smith for providing the computer facilities and instruction needed for this study.

This study was supported by research grant GM19078 from the National Institute of General Medical Sciences. K.H. was supported by a stipend from the Deutsche Forschungsgemeinschaft.

LITERATURE CITED

1. **Bassford, P. J., Jr., and R. J. Kadner.** 1977. Genetic analysis of components involved in vitamin B_{12} uptake in *Escherichia coli*. J. Bacteriol. **132**:796–805.
2. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer-gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. U.S.A. **80**:3963–3965.
3. **Bradbeer, C.** 1979. Transport of vitamin B_{12} in *Escherichia coli*, p. 711–723. *In* B. Zagalak and W. Friedrich (ed.), Vitamin B_{12}. W de Gruyter, Berlin.
4. **Chai, T.-J., V. Wu, and J. Foulds.** 1982. Colicin A receptor: role of two *Escherichia coli* outer membrane proteins (OmpF protein and *btuB* gene product) and lipopolysaccharide. J. Bacteriol. **151**:983–988.
5. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47**:251–276.
6. **Clement, J. M., and M. Hofnung.** 1981. Gene sequence of the λ receptor, an outer membrane protein of *E. coli* K12. Cell **27**:507–514.
7. **DiMasi, D. R., J. C. White, C. A. Schnaitman, and C. Bradbeer.** 1973. Transport of vitamin B_{12} in *Escherichia coli*: common receptor sites for vitamin B_{12} and the E colicins on the outer membrane of the cell envelope. J. Bacteriol. **115**:506–513.
8. **Eisenberg, D.** 1984. Three-dimensional structure of membrane and surface proteins. Annu. Rev. Biochem. **53**:595–623.
9. **Garavito, R. M., J. A. Jenkins, J. M. Neuhaus, A. P. Pugsley, and J. P. Rosenbusch.** 1982. Structural investigations of outer membrane proteins from *Escherichia coli*. Ann. Microbiol. (Inst. Pasteur) **133A**:37–41.
10. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
11. **Heller, J., B. J. Mann, and R. J. Kadner.** 1985. Cloning and expression of the gene for the vitamin B_{12} receptor in the outer membrane of *Escherichia coli*. J. Bacteriol. **161**:896–903.
12. **Imajoh, S., Y. Ohno-Iwashita, and K. Imahori.** 1982. The receptor for colicin E3. Isolation and some properties. J. Biol. Chem. **257**:6481–6487.
13. **Inokuchi, K., N. Mutoh, S.-I. Matsuyama, and S. Mizushima.** 1982. Primary structure of the *ompF* gene that codes for a major outer membrane protein of *Escherichia coli* K-12. Nucleic Acids Res. **10**:6957–6968.
14. **Kadner, R. J.** 1978. Repression of synthesis of the vitamin B_{12} receptor in *Escherichia coli*. J. Bacteriol. **136**:1050–1057.
15. **Kadner, R. J., and G. L. Liggins.** 1973. Transport of vitamin B_{12} in *Escherichia coli*: genetic studies. J. Bacteriol. **115**:514–521.
16. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
17. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
18. **Michaelis, S., and J. Beckwith.** 1982. Mechanism of incorporation of cell envelope proteins in *Escherichia coli*. Annu. Rev. Microbiol. **36**:435–465.
19. **Mizuno, T., M.-Y. Chou, and M. Inouye.** 1983. A comparative study on the genes for three porins of the *Escherichia coli* outer membrane. DNA sequence of the osmoregulated *ompC* gene. J. Biol. Chem. **258**:6932–6940.
20. **Neilands, J. B.** 1982. Microbial envelope proteins related to iron. Annu. Rev. Microbiol. **36**:285–309.
21. **Nikaido, H., and H. C. P. Wu.** 1984. Amino acid sequence homology among the major outer membrane proteins of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **81**:1048–1052.
22. **Overbeeke, N., H. Bergmans, F. van Mansfeld, and B. Lugtenberg.** 1983. Complete nucleotide sequence of *phoE*, the structural gene for the phosphate limitation inducible outer membrane pore protein of *Escherichia coli* K-12. J. Mol. Biol. **163**:513–532.
23. **Reynolds, P. R., G. P. Mottur, and C. Bradbeer.** 1980. Transport of vitamin B_{12} in *Escherichia coli*. Some observations on the role of the gene products of *BtuC* and *TonB*. J. Biol. Chem. **255**:4313–4319.
24. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
25. **Sabet, S. F., and C. A. Schnaitman.** 1973. Purification and properties of the colicin E3 receptor of *Escherichia coli*. J. Biol. Chem. **248**:1797–1806.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
27. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequences of *Escherichia coli* 16S rRNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. **71**:1342–1346.
28. **Staden, R.** 1980. A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. **8**:3673–3694.
29. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10**:2971–2996.