The Net Charge of the First 18 Residues of the Mature Sequence Affects Protein Translocation across the Cytoplasmic Membrane of Gram-Negative Bacteria

ANDREY V. KAJAVA,[1]* SERGEY N. ZOLOV,[2] ANDREY E. KALININ,[2]†
AND MARINA A. NESMEYANOVA[2]

*Center for Molecular Modeling, CIT, National Institutes of Health, Bethesda, Maryland 20892,[1] and Laboratory of
Protein Secretion in Bacteria, Skryabin Institute of Biochemistry and Physiology of Microorganisms,
Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia[2]*

This statistical study shows that in proteins of gram-negative bacteria exported by the Sec-dependent pathway, the first 14 to 18 residues of the mature sequences have the highest deviation between the observed and expected net charge distributions. Moreover, almost all sequences have either neutral or negative net charge in this region. This rule is restricted to gram-negative bacteria, since neither eukaryotic nor gram-positive bacterial exported proteins have this charge bias. Subsequent experiments performed with a series of *Escherichia coli* alkaline phosphatase mutants confirmed that this charge bias is associated with protein translocation across the cytoplasmic membrane. Two consecutive basic residues inhibit translocation effectively when placed within the first 14 residues of the mature protein but not when placed in positions 19 and 20. The sensitivity to arginine partially reappeared again 30 residues away from the signal sequence. These data provide new insight into the mechanism of protein export in gram-negative bacteria and lead to practical recommendations for successful secretion of hybrid proteins.

Protein export in cells is initiated by an N-terminal hydrophobic signal sequence that routes the protein into the secretory pathway. The signal sequence contains information for interaction with protein components of the secretory machinery (for reviews, see references 15 and 42), membrane phospholipids (12, 20, 37), and signal peptidase (9, 25). A signal peptide is not, however, sufficient to mediate the export of any attached polypeptide. The fusion of a signal peptide to a normally cytoplasmic protein has frequently failed, in gram-negative bacteria, to induce secretion into the periplasm (4, 7, 19, 34, 41). Moreover, despite similarity in the signal sequences, not all of proteins normally secreted in eukaryotic cells can be exported in bacteria even with a prokaryotic signal peptide. These facts suggest that some features of the mature protein either contribute to or constrain the secretory process, at least in bacteria. Indeed, it was shown that the mature part of the exported protein should not include a highly hydrophobic membrane anchor sequence (11). Furthermore, Li et al. (29), Yamane and Mizushima (53), MacIntyre et al. (33), and Geller et al. (16) have shown that positively charged residues can block export when introduced directly after the signal sequences of exported proteins. A statistical study also revealed that most prokaryotic proteins have a negative or neutral net charge in the region (about five residues) immediately downstream of the signal sequence (51). However, analysis of the latest releases of sequence databases shows that this rule does not hold for all exported prokaryotic proteins. It was also suggested that net positive or neutral charge difference between the N- and C-terminal regions of the signal peptide may be required for protein secretion (51). Further mutational analysis has shown that this N-C charge imbalance may not be obligatory for translocation of proteins across the cytoplasmic membrane of bacteria (5, 37, 43). Several experimental works suggested that the charge-sensitive mature sequence critical for protein export in bacteria may be extended to 15 (47), 20 (27), and even 30 residues (1). Nevertheless, the wide scatter of the critical region lengths and the absence of additional evidence made it impossible to provide precise answers to the following questions: (i) where does the export domain end, (ii) what is the property of the mature sequence which is critical for protein export, and (iii) how general are these requirements? Thus, despite numerous indications of the importance of the charged residue distribution in the mature region adjacent to the signal sequence, the precise explanation of these observations is not known.

The number of known amino acid sequences has now increased sufficiently to make detailed statistical studies feasible for secreted proteins from various species. In this work, we have analyzed systematically the net charge distribution in a region of the mature proteins adjoining to the signal sequence. This analysis shows that almost all exported proteins of gram-negative bacteria have either neutral or negative net charge in the region of the first 16 ± 2 residues of the mature sequences. At the same time, neither eukaryotic nor gram-positive bacterial exported proteins have this charge bias. We also report examination of the secretion of mutants of *Escherichia coli* alkaline phosphatase. These mutant proteins were designed to verify the hypothesis that the observed charge bias is critical for protein translocation across the cytoplasmic membrane of gram-negative bacteria.

MATERIALS AND METHODS

Selection of sequences for statistical analysis. Sequences from gram-negative bacteria were taken from SwissProt 32.0 (3) using Sequence Retrieval System

---

* Corresponding author. Mailing address: Center for Molecular Modeling CIT, NIH, Bldg. 12A Room 2011, Bethesda, MD 20892. Phone: (301) 402-3043. Fax: (301) 402-2867. E-mail: kajava@helix.nih.gov.
† Present address: National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892.

TABLE 1. Mutagenic oligonucleotides

Oligonucleotide structure[a]	Substitution(s)
5′-GGCATTTCTGGC<u>CTA</u>CCGGGCTTTTG-3′	T2→amber
5′-CCAGAACAGGC<u>TA</u>TTCTGGTGTC-3′	M5→amber
5′-TATCGCCCT<u>A</u>AGCAGCCCG-3′	Q14→amber
5′-GAGCACCGCC<u>CTA</u>TGCAGTAATATCGC-3′	P20→amber
5′-GCGGCAGTCT<u>A</u>ATCACCCGTTAAACG-3′	Q30→amber
5′-ACAGGCATTTC<u>TTTTT</u>TCCGGGCTTTTG-3′	Amber2→K/P3→K
5′-GTTTTCCAGAAC<u>TTTTTT</u>TTCTGGTGTCC-3′	Amber5→K/P6→K
5′-CAGTAATATCGCC<u>TTTTTT</u>AGCCCGGTTTTC-3′	A13→K/amber14→K
5′-CGAGCACCGCC<u>TTTTTT</u>AGTAATATCGCCCTG-3′	A19→K/amber20→K
5′-GAGCGGCAGT<u>TTTTTT</u>TACCCGTTAAACG-3′	D29→K/amber30→K
5′-GTTTTCCAGAAC<u>ACGACG</u>TTCTGGTGTCCG-3′	Amber5→R/P6→R
5′-CAGTAATATCGCC<u>ACGACG</u>GAGCCCGGTTTTC-3′	A13→R/amber14→R
5′-CGAGCACCGCC<u>ACGACG</u>AGTAATATCGCCCTG-3′	A19→R/amber20→R
5′-GAGCGGCAGT<u>ACGAC</u>GACCCGTTAAACG-3′	D29→R/amber30→R

[a] Substituted residues are underlined.

software (http://www.ebi.ac.uk/srs/) and then checked manually. The sequences are for 110 proteins of *E. coli* (68 with known and 42 with well-predicted cleavage sites) and 81 proteins of other gram-negative bacteria with known cleavage sites. The collection did not include highly homologous sequences with more than 80% identity. Anomalous signal sequences (those whose lengths of the hydrophobic core did not fall to the range between 7 and 17 residues) and proteins, secreted by other or modified secretion machineries (hydrogenases, having RRxxFxK pattern within the signal sequence [39], pili [42], and lipoproteins), were also excluded. The collection of the 191 sequences is available over the World Wide Web (http://cmm.info.nih.gov/kajava/). The data sets of the exported proteins from gram-positive bacteria and humans were taken from the SIGNALP database (38). The sequences were randomized using the RandSeq option of the EXPASY server (http://www.expasy.ch).

Bacterial strains and plasmids. *E. coli* E15 (Hfr Δ*phoA8 fadL701 tonA22 garB10 ompF627 relA1 pit-10 spoT1T2*) (2) was used as a host strain for the expression of wild-type and mutant *phoA* genes cloned in plasmids. *E. coli* Z85 [*thi* Δ(*lac-proAB*) Δ(*srl-recA*) hsdR::Tn*10* (F′ *traD proAB lacI*^qΔZM15)] (54) was used to construct mutant *phoA* genes. Phagemid pPHOA12 (24), produced by cloning of the wild-type alkaline phosphatase gene (*phoA*) in the vector p15SK(+), was used to construct and express mutant *phoA* genes. Helper phage R408 was used to isolate single-strand recombinant phagemids. Plasmid harboring the amber suppressor tRNA^{Ala} gene of *E. coli* in the vector pGFIB (26) was provided by J. Miller.

Media and culture conditions. Bacteria for cloning and oligonucleotide-directed mutagenesis were grown on Luria-Bertani (LB) or 2YT medium at 37°C. All media were supplemented with chloramphenicol (25 μg/ml) to either select for or maintain *phoA*-containing plasmids. To screen for colonies expressing active alkaline phosphatase, *E. coli* cells were grown on agar plates made of LB medium free of inorganic phosphate and containing 40 μg of XP (5-bromo-4-chloro-3-indolylphosphate) per ml (17). The cells were grown on minimal medium (50) with 1 mM K₂HPO₄ for repression of enzyme synthesis. For its derepression, *E. coli* cells were grown to the mid-log phase, collected by centrifugation, washed with 0.14 M NaCl at 4°C, resuspended in the same medium lacking orthophosphate, and grown for additional hour.

Oligonucleotide-directed mutagenesis. To generate mutant forms of *phoA*, we used a new two-step method which allowed us to omit hybridization with labeled nucleotides during selection of clones containing mutant genes (21). First, an amber codon in a place of codon corresponding to positions of interest was introduced by oligonucleotide-directed mutagenesis using oligonucleotides 1 to 5 (Table 1). Colonies expressing active alkaline phosphatase become blue when growing on agar plates containing XP, a chromogenic substrate of the enzyme. Insertion of an amber codon resulted in a premature termination of protein synthesis; therefore, colonies bearing the mutant genes remain white on the plates with XP. This allowed us to identify mutant clones. The amber mutation was confirmed by recovery of the alkaline phosphatase phenotype (blue colonies) after introduction into cells of a plasmid carrying the Ala2 amber suppressor. Next, the codons of the desired amino acids (Arg or Lys) were introduced in place of the amber codons by using oligonucleotides 6 to 14 (Table 1). That led to recovery of alkaline phosphatase phenotype in cells lacking of amber suppressor (blue colonies). Oligonucleotide-directed mutagenesis was carried out by the protocol of Promega (Madison, Wis.). Isolation of single-strand phagemid DNA and plasmid DNA, electrophoresis of DNA fragments in agar gels, phosphorylation of oligonucleotides, and transformation of *E. coli* cells were performed by standard procedures (44). Mutations were confirmed by DNA sequencing (45).

Alkaline phosphatase maturation. Pulse-chase experiments were used to analyze the alkaline phosphatase maturation. *E. coli* cells grown to the mid-log phase in the minimal medium with 1 mM K₂HPO₄ were harvested, washed, and

incubated for 10 min in the same medium without orthophosphate to induce alkaline phosphatase synthesis. The cells were labeled with [³⁵S]methionine (50 μCi/ml) for 60 s and chased for 0.1, 1.0, 5.0, or 60.0 min by addition of unlabeled methionine to a final concentration of 0.05%. Proteins were precipitated with 10% trichloroacetic acid. Alkaline phosphatase and its precursor were immunoprecipitated with rabbit antibodies and separated by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) followed by autoradiography. Proteins were quantified with an LKB UltroScan laser densitometer. The relative quantity of mature alkaline phosphatase and its precursor was calculated with adjustment for the difference in number of methionine residues between the precursor and mature form.

Alkaline phosphatase isoforms. Cells expressing alkaline phosphatase were harvested and converted to spheroplasts in 20 mM Tris-HCl (pH 7.5)–10 mM EDTA–50 mM sucrose–1 mg of lysozyme per ml for 15 min at 0°C. The periplasmic fraction was separated from the cell debris by centrifugation at 12,000 × *g* for 5 min. The samples were analyzed by nondenaturing PAGE in a 7.5% gel (10). Staining of the alkaline phosphatase isoforms was performed by incubation of the gel with α-naphthyl phosphate (Sigma N-7255) and fast red dye TR (Chemapol, Praha, Czech Republic) (31).

Analytical methods. Protein SDS-PAGE was performed in 10% gels (28). Immunoprecipitation was carried out as described elsewhere (35). Alkaline phosphatase activity was determined by measuring the rate of *p*-nitrophenylphos-
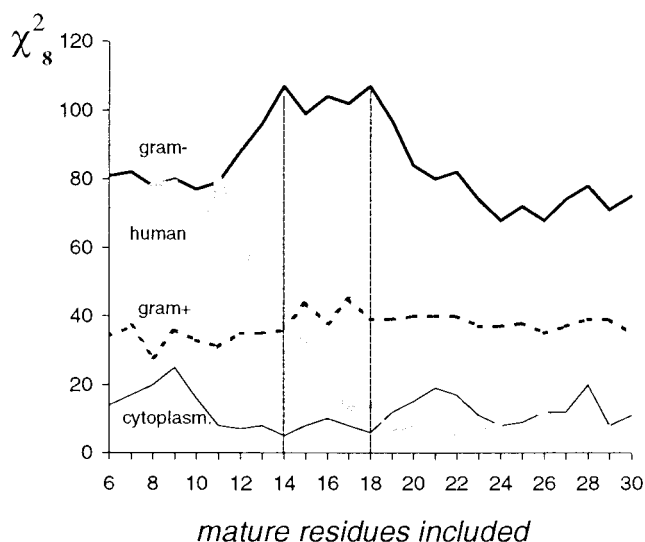

FIG. 1. Deviation between observed and expected net charge distribution (estimated as χ^2_8) for a collection of 191 exported proteins from gram-negative bacteria (thick line), 138 exported proteins from gram-positive bacteria (dotted line), 415 exported human proteins (gray line), and 128 cytoplasmic proteins from *E. coli* (thin line) plotted as a function of the length of the N-terminal region of the mature protein. Probability of having $\chi^2_8 > 20$ by chance is 0.01.
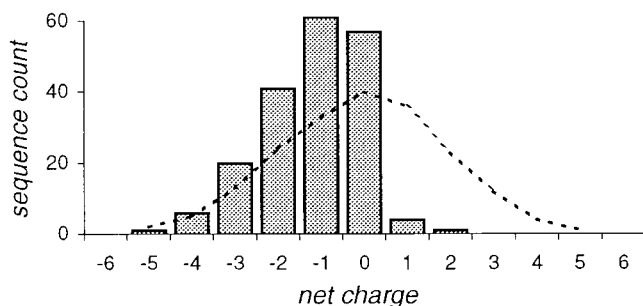
FIG. 2. Observed (histograms) and theoretical (broken line) net charge distributions in the N-terminal 16-residue region of the 191 proteins from gram-negative bacteria.

phate hydrolysis, taking the hydrolysis of 1 μmol of substrate per min at 37°C as a unit of enzymatic activity. Protein was assayed by the Lowry method (32).

RESULTS

Statistical analysis of net charge distribution. We have analyzed the net charge distribution in the N-terminal 30-residue-long mature part of the proteins exported by the Sec-dependent pathway. First, for a window of length n residues starting at the first position of the mature protein, the net charge was calculated for each of the 191 selected proteins (see Materials and Methods), counting arginine and lysine as +1 and aspartate and glutamate as −1. As a result, a distribution of the 191 net charges was obtained. The same procedure was repeated for the window shifted to the second, through to the $30 - n$ position, which resulted in the $30 - n$ net charge distributions. All possible windows of lengths equivalent to 6 to 30 residues were subsequently tested. As controls, theoretical net charge distributions expected for random sequences, with overall amino acid composition as in 30-residue-long domains of the mature proteins, were derived for all windows. The analysis reveals that the deviation between the observed and expected net charge distributions (evaluated using χ^2 analysis) is always maximal when the windows start at the first position. Of particular interest is the conclusion that the first 14 to 18 residues of the mature sequences have the highest deviation (Fig. 1). Comparison of this definitely nonrandom net charge distribution with an expected theoretical one has shown a markedly higher incidence of acidic than basic residues in these regions (Fig. 2). The net charge usually is between −3 and 0, and there are only a few proteins with a net positive charge in the N-terminal 14- to 18-residue region. Altogether, only 8 sequences (versus an expected 76 for sample with random sequence) have

a positive net charge when the first 16 residues are included in the calculation.

The charge bias is equally observed for periplasmic, outer membrane, and extracellular proteins of gram-negative bacteria. In contrast, such a high deviation between the observed and expected net charge distributions does not hold for exported eukaryotic proteins (based on the analysis of human proteins) and gram-positive bacteria proteins (Fig. 1). Indeed, in gram-positive bacteria, exported proteins have a much weaker preference for negative charges, and this preference is almost uniform over 6- to 30-residue windows. Human exported proteins have a relatively high deviation between the observed and expected net charge distributions at the beginning of the mature sequence. However, this deviation is caused not by a high incidence of acidic than basic residues but by a frequent occurrence of charged residue clusters of both signs. The analysis also shows that cytoplasmic proteins of gram-negative bacteria do not have the charge bias (Fig. 1).

Examination of the secretion of a series of mutant alkaline phosphatases. To systematically study the effect on translocation of positively charged residues placed at different distances from a signal sequence, we analyzed mutants derived from the *E. coli* alkaline phosphatase (PhoA). The secretion of this protein is well studied and is typical for Sec-dependent proteins of gram-negative bacteria (18, 25, 35, 37). The mature alkaline phosphatase has 0 net charge in the region of the first 14 or 15 residues and −1 net charge in the 16- to 18-residue regions. Our statistical analysis suggested that introduction of additional negative charge into this early mature region should not inhibit the export. Indeed, in a previous mutational analysis, we have shown that substitution of Arg in position +1 with Glu, which shifts the net charge of the 14- or 15-residue region from 0 to −2, has no effect on translocation and processing of alkaline phosphatase (23, 25). On the other hand, introduction of two positively charged residues gives a positive sign to the net charge of the 14- to 18-residue regions, and as is predicted by the sequence analysis, this can reduce the secretion. We thus constructed five mutant proteins where two residues at positions 2 and 3, 5 and 6, 13 and 14, 19 and 20, or 29 and 30 were replaced by two lysines and four mutant proteins with residues at positions 5 and 6, 13 and 14, 19 and 20, or 29 and 30 replaced by two arginines (Fig. 3).

All mutant proteins were active in the cells (Fig. 4). Thus, the mutants were translocated across cytoplasmic membrane, as it is known that alkaline phosphatase becomes active only after translocation into the periplasm, where disulfide bond formation and enzyme dimerization take place (35). However, the level of enzyme activity in the cells secreting mutant pro-
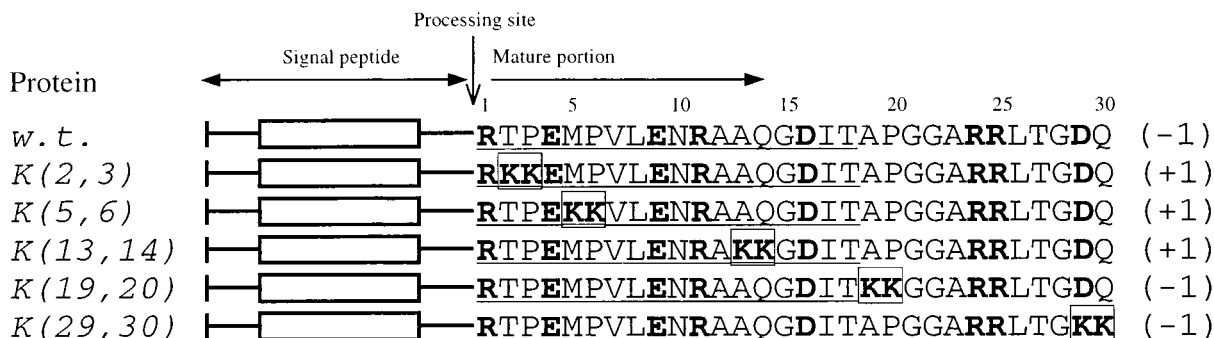

FIG. 3. N-terminal sequences of the first 30 residues of mature wild-type (w.t.) and mutant alkaline phosphatases (lysine series). Amino acid substitutions are in boxes. Charged residues are in bold. The first 18 residues are underlined. The corresponding net charge of the 18-residue region is on the right (in parentheses).
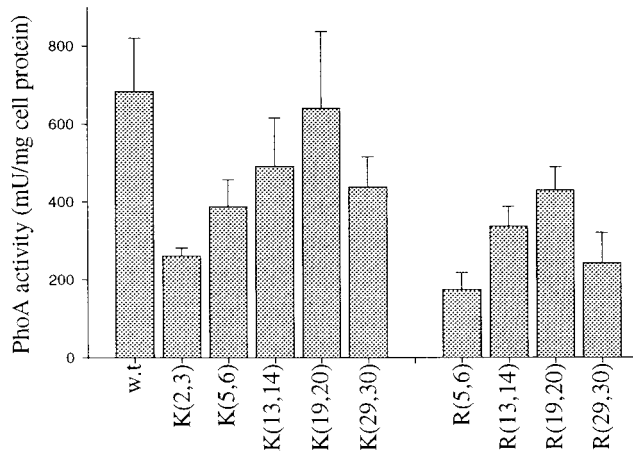
FIG. 4. Alkaline phosphatase activity in cells producing wild-type (w.t.) or mutant enzyme forms.

and R[29,30]. It is worth mentioning that all corresponding mutants of the arginine series were less active than mutants of the lysine series.

Study of protein secretion based on the level of enzyme activity is a reasonable approach if the mutations do not affect the catalytic properties of alkaline phosphatase. Indeed, analysis of the three-dimensional structure shows that the mutated residues are located far apart from the active center and dimerization surface of the enzyme. However, we could not completely exclude changes in the catalytic properties of the mutant alkaline phosphatases. Therefore, the effect of the substitutions on alkaline phosphatase translocation was also assessed by the rate of conversion of pulse-labeled mutant protein precursors into the mature forms in vivo using the standard pulse-chase method. As shown in Fig. 5, translocation of the wild-type alkaline phosphatase was essentially completed within a 1-min pulse (corresponding to 0 min of chase). In contrast, most amino acid substitutions located within the first 14 residues resulted in an inhibition of protein translocation. Similar to the results with enzyme activity, the most severe effect was observed when a couple of lysines or arginines was placed closest to the signal peptide. The further away the residues substituted by lysine were from the signal peptide, the less the effect of the inhibition. When lysines were placed at distance 19 residues or more, the translocation was completely restored (Fig. 5A and C). The arginine series of mutant proteins had a similar dependence of the translocation inhibition (Fig. 5B and D), although the inhibition was stronger. In contrast to the lysine mutant proteins, the arginine constructs R[5,6] and R[13,14] had almost the same level of inhibition,

teins was lower than that in cells secreting the wild-type protein. Most effect was observed in proteins having the amino acid substitutions for positions near the signal peptide: positions 2 and 3 with Lys and positions 5 and 6 with Arg. The least effect was observed when either Lys or Arg was placed in positions 19 and 20. Moreover, the level of alkaline phosphatase activity in cells producing mutant protein K[19,20] was equal to that in cells producing wild-type protein. Surprisingly, we found that the activity decreased again for both K[29,30]
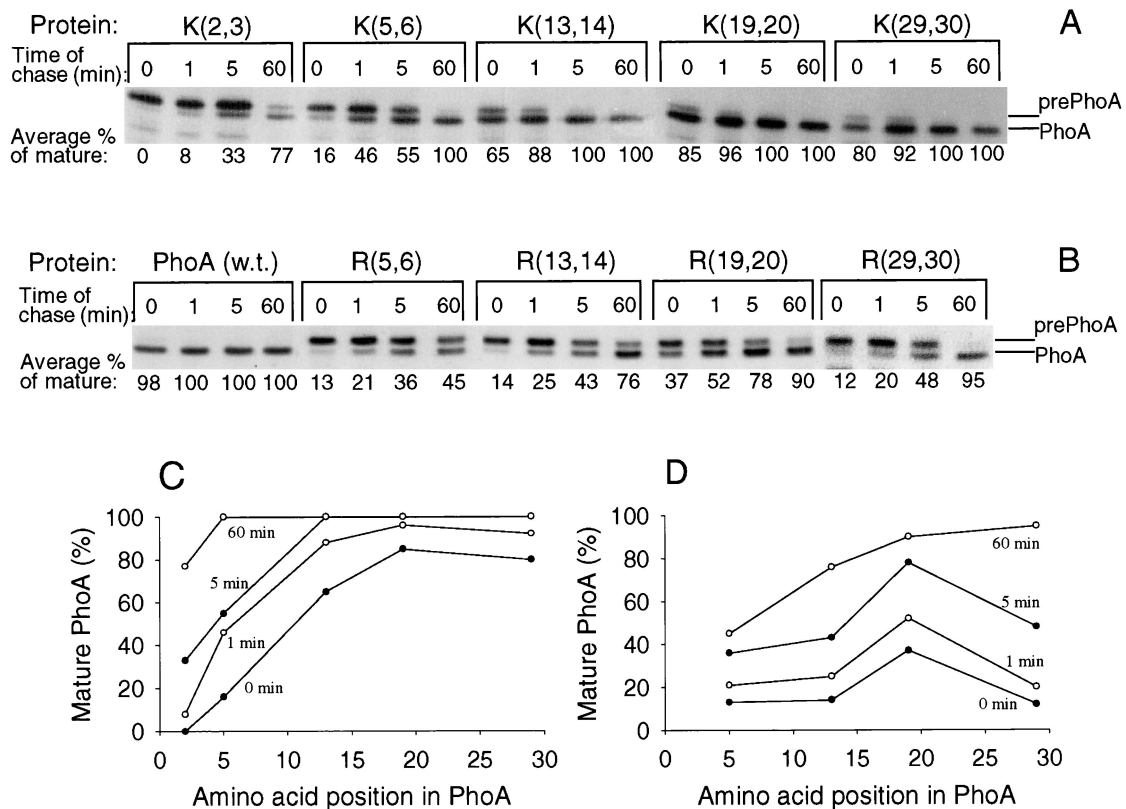

FIG. 5. The dynamics of maturation of wild-type and mutant alkaline phosphatases. (A and C) Lysine series of mutant proteins. (B and D) Arginine series of mutant proteins. (C and D) Graphical representations of average percentage of mature alkaline phosphatases (PhoA) at various time of chase depending on the positions of the mutations.
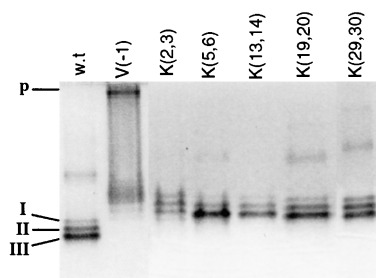
FIG. 6. Multiple forms of wild-type (w.t.) and mutant alkaline phosphatases. Samples were analyzed by PAGE (7.5% gel) under nondenaturing conditions, and the active enzyme was revealed by incubation of the gel with α-naphthyl phosphate as the alkaline phosphatase substrate and fast red dye TR. Isoforms of wild-type alkaline phosphatase (I, II, and III) and active precursor (p) are indicated.

followed by more sharp restoration of translocation for R[19,20]. Unlike the study of enzyme activity, the pulse-chase experiments did not reveal the inhibition for K[29,30] and R[29,30] mutant proteins 60 min after the pulse. However, this secondary inhibition was observed for R[29,30] for the first 5 min after the pulse (Fig. 5B and D). This suggests that the introduction of arginines at positions 29 and 30 affects kinetic of the translocation.

To determine whether the mutations affect protein translocation or precursor processing, we visualized alkaline phosphatase isoforms in gels after electrophoresis under nondenaturing conditions. Active alkaline phosphatase can be stained in the gel by treatment with the enzyme substrate α-naphthyl phosphate and an appropriate dye. It is known that the periplasmic alkaline phosphatase is active regardless of whether it is transformed into the mature form or remains as a translocated precursor with an uncleaved signal peptide due to mutation. On the other hand, the cytoplasmic precursor does not have such enzymatic activity (6). Enzymatic activities of the mature protein and the translocated precursor can be individually estimated by staining the gel after nondenaturing electrophoresis, since mature protein isoforms and the translocated precursor have different electrophoretic mobilities (22, 25). The precursor translocated across the membrane can be found at the top of the gel, in forms probably resulting from

protein aggregation due to the presence of hydrophobic signal peptide. Such unprocessed mutant protein with the substitution of Val at position −1 was used as a control [Fig. 6, V(−1)]; its activity after translocation across the cytoplasmic membrane was shown previously (22). Our experiment showed that only mature isoforms were found for all mutant proteins of the lysine and arginine series (Fig. 6). This allowed us to conclude that the mutations did not inhibit efficiency of processing but inhibited efficiency of protein translocation.

DISCUSSION

The results of the mutagenesis confirm that the theoretically observed bias of the net charge in the 16 ± 2-residue region is associated with protein translocation. Two consecutive basic residues change the net charge of this region to +1, +2 and reduce translocation when placed inside it but not when placed only a few residues away. It is known that SecA protein interacts with both signal peptide and the mature part of the protein precursor at the initiation of its export (15, 30, 46). Therefore, one can assume that the described nonrandom distribution of charged residues of the mature part is important for specific interactions with SecA or any other protein of the secretory machinery. However, the fact that we did not find specific sequence patterns within this region of the exported protein, but a net charge bias, makes the hypothesis of specific protein-protein interactions highly improbable. The detection of the net charge bias implies a mechanism based on the sign of the membrane potential rather than on specific ionic interactions with components of the export machinery. Remarkably, the 16 ± 2-residue length of the mature region containing the charge bias coincides with the combined length of the N-terminal positively charged region and the hydrophobic part of the signal sequence. This correlation can be explained within the frame of the loop or, more precisely, the α-helical hairpin model of the export initiation domain (8, 13, 52) (Fig. 7). The insertion of the hydrophobic signal peptide into the lipid bilayer may proceed either directly or (as shown in Fig. 7, left) in association with SecYEG translocation complex, with or without the help of SecA protein (15, 20). In a mechanism proposed here, this insertion of the signal peptide favors spontaneous diving of the adjoining mature region into the transmembrane environment (into the lipid bilayer or SecYEG
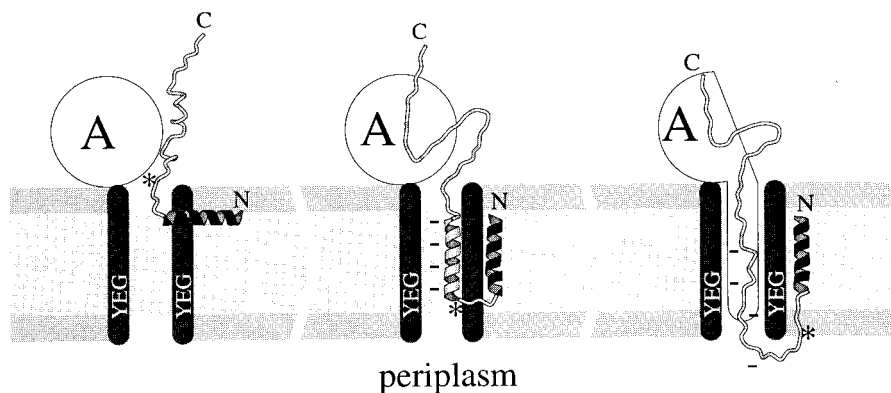

FIG. 7. Schematic representation of the initial steps of protein secretion. (Left) insertion of the signal peptide into the membrane; (center) spontaneous insertion of a mature part of the protein and formation of an α-helical hairpin oriented within the SecYEG translocase in a manner favorable for the subsequent ATP-dependent, SecA-driven translocation; (right) SecA-driven translocation of the protein throughout the transmembrane translocase. SecA protein and SecYEG translocase complex are marked "A" and "YEG," correspondingly. The N-terminal hydrophobic region of the signal peptide is outlined in black. Asterisks mark the cleavage site of the leader peptidase. The region of the mature sequence containing the net charge bias is marked by dashes. The polar head layers of the membrane are in darker gray compared with the middle nonpolar layer. The dimensions of the membrane and the lengths of protein chain regions are shown in scale (thicknesses of the polar and nonpolar layers of the membrane are taken as 8 and 30 Å, respectively).

complex). Then, at the instant the first 14 to 18 residues of the mature part sink into the membrane and adopt an α-helical conformation, the transmembrane potential, which is positively charged at the periplasmic side, does not push out this negatively charged fragment (Fig. 7, center). We cannot rule out the possibility that during this stage, a mature region of the precursor, which is remote from the signal peptide, may interact with SecA. However, an important point is that the adjoining mature domain inserts into the membrane without the help of SecA. Thus, the mechanism assumes that in gram-negative bacteria during Sec-dependent protein translocation, there is a step of spontaneous insertion, which is mainly controlled by the sequence of the export initiation hairpin. The occurrence of the negative net charge in the mature sequence suggests a simple solution for a major problem of the hairpin models that is to understand how the second (hydrophilic) helix manages to insert and remain into the nonpolar membrane. The suggested mechanism agrees well with previous finding that it is the membrane electrical potential $\Delta\psi$ which inhibits the initiation of translocation of outer membrane protein A precursor in *E. coli* when two positively charged residues were inserted immediately after the signal peptide, whereas it has the opposite effect on the mutant protein with these two positions occupied by negatively charged residues (16). Our data also support the earlier observation (47) that the further from the amino terminus the positive charges, the less blocking effect on export they have. In the frame of our mechanism, this observation can be explained by more efficient spontaneous insertion of the second helix, which has positive charge at its end and therefore overtakes a smaller barrier of insertion compared with the same helix but having positive charges at the beginning. It is important to mention here, that in both cases the second helix tends to position itself 16 ± 2 residues inside the membrane in order to be available for next step. Therefore, the distribution of the charged residues is important for the efficiency of the insertion, while the net charge of the 16 ± 2 residues controls a hairpin position, which can be permissive for the subsequent ATP-dependent, SecA-driven protein translocation (Fig. 7, right).

Our experimental data support previous observations (48) that arginine is more efficient than lysine for reducing of the translocation, (Fig. 4 and 5). This suggests that basic residues near the N terminus of a mature part of a secreted protein may be deprotonated if orderly export is to occur. The difference can be also explained by hydrogen bonding of the basic residues to the phospholipids, because arginine side chain is able to form more hydrogen bonds compared to lysine side chain.

The sensitivity to positively charged residues partially reappeared again 30 residues away from the signal sequence. The reappearance of the inhibition at position 30 may indicate that there is the second charge-dependent process, which can reduce the efficiency of the secretion. This finding may explain the discrepancy of the export domain length suggested in previous studies (15 to 20 residues [27, 47] versus 30 residues [1]).

An important conclusion is that the net charge of the first 18 residues of the mature sequence is probably critical for Sec-dependent translocation only in gram-negative bacteria, while this requirement does not hold for exported proteins from gram-positive bacteria or eukaryotes. It is already known that there are differences between signal sequences and secretory machineries of gram-negative bacteria, gram-positive bacteria, and eukaryotes (15, 36). For example, it was shown that the translocase subunits of *E. coli* and *Bacillus subtilis* cannot be unconditionally exchanged (49). All this evidence suggests that the spontaneous insertion controlled by the net charge of the

first 18 residues of the mature part may be a step specific to gram-negative bacteria.

In parallel with a more detailed insight into the molecular mechanism of protein export, the demarcation of the region critical for secretion leads to a practical recommendation. Cytoplasmic proteins, as well as proteins secreted in eukaryotic cells and gram-positive bacteria, frequently have a positive net charge exceeding +2 in the region of interest. Although it has been known for some time that exclusion of positively charged residues in the early mature region may promote efficient export of heterologous or cytoplasmic proteins in gram-negative hosts, the precise length of this early mature region and the value of the net charge which permits the export were not known. Our study suggests that for successful expression of these proteins in gram-negative bacteria, it is reasonable to optimize the sequence, which begins after the hydrophobic core of the signal peptide and ends at the 18th position of the mature protein, so that the net charge of this region becomes electronegative or neutral.

REFERENCES

1. **Andersson, H., and G. von Heijne.** 1991. A 30-residue-long "export initiation domain" adjacent to the signal sequence is critical for protein translocation across the inner membrane of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **88:**9751–9754.
2. **Bachmann, B. J.** 1987. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1120–1219. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
3. **Bairoch, A., and R. Apweiler.** 1997. The SwissProt protein sequence data bank and its supplement TREMBL. Nucleic Acids Res. **25:**31–36.
4. **Bassford, P. J., Jr., T. J. Silhavy, and J. R. Beckwith.** 1979. Use of gene fusion to study secretion of maltose-binding protein in *Escherichia coli* periplasm. J. Bacteriol. **139:**19–31.
5. **Bosch, D., P. de Boer, W. Bitter, and J. Tommassen.** 1989. The role of the positively charged N-terminus of the signal sequence of *Escherichia coli* outer membrane protein PhoE in export. Biochim. Biophys. Acta **979:**69–76.
6. **Boyd, D., C. D. Guan, S. Willard, W. Wright, K. Strauch, and J. Beckwith.** 1987. Enzymatic activity of alkaline phosphatase precursor depends on its cellular location, p. 89–93. *In* A. Torriani-Gorini, F. G. Rothman, S. Silver, A. Wright, and E. Yagil (ed.), Phosphate metabolism and cellular regulation in microorganisms. American Society for Microbiology, Washington, D.C.
7. **Boyd, D., and J. Beckwith.** 1990. The role of charged amino acids in the localization of secreted and membrane proteins. Cell **62:**1031–1033.
8. **Engelman, D. M., and T. A. Steitz.** 1981. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell **23:**411–422.
9. **Dalbey, R. E., and G. von Heijne.** 1992. Signal peptidases in prokaryotes and eukaryotes—a new protease family. Trends Biochem. Sci. **11:**474–478.
10. **Davis, B. J.** 1964. Disc electrophoresis. II. Method and application to human serum proteins. Ann. N. Y. Acad. Sci. **121:**404–427.
11. **Davis, N. G., and P. Model.** 1985. An artificial anchor domain: hydrophobicity suffices to stop transfer. Cell **41:**607–614.
12. **de Vrije, T., R. L. de Swart, W. Dowhan, J. Tommassen, and B. de Kruijff.** 1988. Phosphatidylglycerol is involved in protein translocation across *Escherichia coli* inner membranes. Nature **334:**173–175.
13. **DiRienzo, J. M., K. Nakamura, and M. Inouye.** 1978. The outer membrane proteins of Gram-negative bacteria: biosynthesis, assembly, and functions. Annu. Rev. Biochem. **47:**481–532.
14. **Duffaud, G., and M. Inouye.** 1988. Signal peptidases recognize a structural feature at the cleavage site of secretory proteins. J. Biol. Chem. **263:**10224–10228.
15. **Fekkes, P., and A. J. Driessen.** 1999. Protein targeting to the bacterial cytoplasmic membrane. Microbiol. Mol. Biol. Rev. **63:**161–173.
16. **Geller, B., H. Y. Zhu, S. Cheng, A. Kuhn, and R. E. Dalbey.** 1993. Charged residues render pro-OmpA potential dependent for initiation of membrane

translation. J. Biol. Chem. **268:**9442–9447.

17. **Inouye, H., S. Michaelis, A. Wright, and J. Beckwith.** 1981. Cloning and restriction mapping of alkaline phosphatase structural gene (*phoA*) of *Escherichia coli* and generation of deletion mutant in vitro. J. Bacteriol. **146:** 668–675.

18. **Ito, K., P. J. Bassford, Jr., and J. Beckwith.** 1981. Protein localization in *E. coli*: is there a common step in the secretion of periplasmic and outer-membrane proteins? Cell **24:**707–717.

19. **Kadonaga, J. T., A. E. Gautier, D. R. Straus, A. D. Charles, M. D. Edge, and J. R. Knowles.** 1984. The role of the beta-lactamase signal sequence in the secretion of proteins by *Escherichia coli*. J. Biol. Chem. **259:**2149–2154.

20. **Kajava, A. V., M. V. Bogdanov, and M. A. Nesmeyanova.** 1991. Stereochemical analysis of interaction of signal peptide with phospholipids at the initiation of protein translocation across the membrane. J. Biomol. Struct. Dyn. **9:**143–157.

21. **Kalinin, A. E., N. I. Mikhaleva, A. L. Karamyshev, Z. N. Karamysheva, and M. A. Nesmeyanova.** 1999. Interaction of mutant alkaline phosphatase precursors with membrane phospholipids *in vivo* and *in vitro*. Biochemistry (Moscow) **64:**1021–1029.

22. **Karamyshev, A. L., A. E. Kalinin, I. M. Tsfasman, V. N. Ksenzenko, and M. A. Nesmeyanova.** 1994. Study of the biogenesis and secretion of alkaline phosphatase and its mutant forms in *Escherichia coli*. II. Effect of amino acid substitutions in the processing site and N-terminus of mature polypeptide chain on its biogenesis. Mol. Biol. (Moscow) **28:**245–252.

23. **Karamyshev, A. L., A. E. Kalinin, M. I. Khmel'nitsky, M. G. Shlyapnikov, V. N. Ksenzenko, and M. A. Nesmeyanova.** 1994. Study of the biogenesis and secretion of alkaline phosphatase and its mutant forms in *Escherichia coli*. III. Substitutions of alkaline phosphatase N-terminal amino acid affect enzyme biogenesis. Mol. Biol. (Moscow) **28:**253–258.

24. **Karamyshev, A. L., M. G. Shlyapnikov, M. I. Khmel'nitsky, M. A. Nesmeyanova, and V. N. Ksenzenko.** 1994. Study of the biogenesis and secretion of alkaline phosphatase and its mutant forms in *Escherichia coli*. I. Introduction of directed mutations into the alkaline phosphatase gene. Mol. Biol. (Moscow) **28:**150–157.

25. **Karamyshev, A. L., Z. N. Karamysheva, A. V. Kajava, V. N. Ksenzenko, and M. A. Nesmeyanova.** 1998. Processing of *Escherichia coli* alkaline phosphatase: role of the primary structure of the signal peptide cleavage region. J. Mol. Biol. **277:**859–870.

26. **Kleina, L. G., J.-M. Masson, J. Normally, J. Abelson, and J. H. Miller.** 1990. Construction of *E. coli* amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. J. Mol. Biol. **213:**705–717.

27. **Kuhn, A., D. Kiefer, C. Kohne, H. Y. Zhu, W. R. Tschantz, and R. E. Dalbey.** 1994. Evidence for a loop-like insertion mechanism of pro-Omp A into the inner membrane of *Escherichia coli*. Eur. J. Biochem. **226:**891–897.

28. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680–685.

29. **Li, P., J. Beckwith, and H. Inouye.** 1988. Alteration of the amino terminus of the mature sequence of a periplasmic protein can severely affect protein export in *E. coli*. Proc. Natl. Acad. Sci. USA **85:**7685–7689.

30. **Lill, R., W. Dowhan, and W. Wickner.** 1990. The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. Cell **60:**271–280.

31. **Lojda, Z., R. Gossrau, and T. H. Schibler.** 1979. Enzyme histochemistry: A laboratory manual. Springer-Verlag, Berlin, Germany.

32. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193:**265–275.

33. **MacIntyre, S., M. L. Eschbach, and B. Mutschler.** 1990. Export incompatibility of N-terminal basic residues in a mature polypeptide of *Escherichia coli* can be alleviated by optimising the signal peptide. Mol. Gen. Genet. **221:** 466–474.

34. **Moreno, F., A. V. Fowler, M. Hall, T. J. Silhavy, I. Zabin, and M. Schwartz.** 1980. A signal sequence is not sufficient to lead beta-galactosidase out of the cytoplasm. Nature **286:**356–359.

35. **Michaelis, S., H. Inouye, D. Oliver, and J. Beckwith.** 1983. Mutations that alter the signal sequence of alkaline phosphatase in *Escherichia coli*. J. Bacteriol. **154:**366–374.

36. **Navarre, W. W., and O. Schneewind.** 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol. Mol. Biol. Rev. **63:**174–229.

37. **Nesmeyanova, M. A., A. L. Karamyshev, Z. N. Karamysheva, A. E. Kalinin, V. N. Ksenzenko, and A. V. Kajava.** 1997. Positively charged lysine at the N-terminus of the signal peptide of the *Escherichia coli* alkaline phosphatase provides the secretion efficiency and is involved in the interaction with anionic phospholipids. FEBS Lett. **403:**203–207.

38. **Nielsen, H., J. Engelbrecht, G. von Heijne, and S. Brunak.** 1996. Defining a similarity threshold for a functional protein sequence pattern: the signal peptide cleavage site. Proteins **24:**165–177.

39. **Niviere, V., S. L. Wong, and G. Voordouw.** 1992. Site-directed mutagenesis of the hydrogenase signal peptide consensus box prevents export of a beta-lactamase fusion protein. J. Gen. Microbiol. **138:**2173–2183.

40. **Overbye, L. J., M. Sandkvist, and M. Bagdasarian.** 1993. Genes required for extracellular secretion of enterotoxin are clustered in *Vibrio cholerae*. Gene **132:**101–106.

41. **Palva, I., M. Sarvas, P. Lehtovaara, M. Sibakov, and L. Kaariainen.** 1982. Secretion of *Escherichia coli* beta-lactamase from *Bacillus subtilis* by the aid of alpha-amylase signal sequence. Proc. Natl. Acad. Sci. USA **79:**5582–5586.

42. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. **57:**50–108.

43. **Puziss, J. W., S. M. Strobel, and P. J. Bassford.** 1992. Export of maltose-binding protein species with altered charge distribution surrounding the signal peptide hydrophobic core in *Escherichia coli* cells harboring *prl* suppressor mutations. J. Bacteriol. **174:**92–101.

44. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, New York, N.Y.

45. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

46. **Schiebel, E., A. J. Driessen, F. U. Hartl, and W. Wickner.** 1991. Δμ H+ and ATP function at different steps of the catalytic cycle of preprotein translocase. Cell **64:**927–939.

47. **Summers, R. G., and J. R. Knowles.** 1989. Illicit secretion of a cytoplasmic protein into the periplasm of *Escherichia coli* requires a signal peptide plus a portion of the cognate secreted protein. Demarcation of the critical region of the mature protein. J. Biol. Chem. **264:**20074–20081.

48. **Summers, R. G., C. R. Harris, and J. R. Knowles.** 1989. A conservative amino acid substitution, arginine for lysine, abolishes export of a hybrid protein in *Escherichia coli*. Implications for the mechanism of protein secretion. J. Biol. Chem. **264:**20082–20088.

49. **Swaving, J., K. H. van Wely, and A. J. Driessen.** 1999. Preprotein translocation by a hybrid translocase composed of *Escherichia coli* and *Bacillus subtilis* subunits. J. Bacteriol. **181:**7021–7027.

50. **Torriani, A.** 1966. Alkaline phosphatase from *E. coli*, p. 224–234. *In* G. L. Cantori and R. Davis (ed.), Procedures in nucleic acid research. Harper and Row, Publishers, New York, N.Y.

51. **von Heijne, G.** 1986. Net N-C charge imbalance may be important for signal sequence function in bacteria. J. Mol. Biol. **192:**287–290.

52. **von Heijne, G., and C. Blomberg.** 1979. Trans-membrane translocation of proteins. The direct transfer model. Eur. J. Biochem. **97:**175–181.

53. **Yamane, K., and S. Mizushima.** 1988. Introduction of basic amino acid residues after the signal peptide inhibits protein translocation across the cytoplasmic membrane of *E. coli*. J. Biol. Chem. **263:**19690–19696.

54. **Zaitsev, E. N., E. M. Zaitseva, I. V. Bakhlanova, V. N. Gorelov, N. P. Kuzmin, V. M. Krykov, and V. A. Lantsov.** 1986. Cloning and characterization of recA gene from *Pseudomonas aeruginosa*. Genetics (Moscow) **22:** 2721–2727.