Combinatorial Analysis of the Structural Requirements of the *Escherichia coli* Hemolysin Signal Sequence*§§*

(Received for publication, August 31, 1999, and in revised form, October 14, 1999)

David Hui§§, Carla Morden§§, Fang Zhang, and Victor Ling†***

*From the British Columbia Cancer Research Centre, Vancouver, British Columbia, V5Z 1L3, Canada and the †Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, V6T 1Z3, Canada

We have investigated the substrate specificity of the *Escherichia coli* hemolysin transporter system. Translocation of hemolysin is dependent on a C-terminal signal sequence located within the last 60 amino acids of this protein. Previous comparative studies of the signal sequence have revealed a conserved helix(α1)-linker-helix(α2) motif, suggesting that secondary structure is important for transport. In this study, we generated three random libraries in the α1, linker, and α2 regions, as well as an α1-amphiphilic helical library to identify features buried within the structural motif that contribute to transport. Combinatorial variants were generated by altering the primary sequence of specific regions, and correlation between the genotype and phenotype of the mutant populations allowed us to objectively identify any functional features involved. It was found that the α1-amphiphilic helix and the linker are both important for function. To our surprise, the second helix of the conserved structural motif was not essential for transport. The finding that a predicted amphiphilic helix and hydrophobicity, rather than primary sequence, contribute to transport in the α1 region allows us to speculate on the mechanism of multiple substrate recognition. This may have implications for understanding the broad substrate specificity common among other ATP-binding cassette transporters.

The transport of substrates across biological membranes is an essential function of all cells. The ATP-binding cassette (ABC)* transporter superfamily of proteins plays an active role in this process. They are found in all kingdoms of life, and are involved, for example, in amino acid uptake, phosphate import, protein secretion, polysaccharide export, ion transport, and cellular drug efflux (1, 2). Because of the high degree of sequence homology among members of this superfamily, it is believed that the fundamental mechanism of transport among ABC proteins is likely to be similar. One of the most captivating questions in the field of ABC transporters is the phenomenon of broad substrate specificity, either as a group or within single members. For instance, in the latter case, two of the best characterized ABC proteins, P-glycoprotein and multidrug resistance protein, actively pump a wide variety of chemotherapeutic agents out of cells (3). Overexpression of these proteins contributes to the multidrug resistance phenotype in many types of cancer in humans.

In the present study, we have investigated the substrate specificity of the *Escherichia coli* hemolysin system. The hemolysin system is a well characterized bacterial ABC exporter (4). Studies to investigate substrate specificity in this system have been facilitated by the fact that hemolysin is a protein substrate, which allows for easy genetic manipulation. Secretion of hemolysin is directed by a signal sequence that is located within the C-terminal 60 amino acids. This short peptide can be secreted by itself (5), or as demonstrated by fusion protein studies, is sufficient to guide the translocation of foreign proteins directly from the cytoplasm to the outside of cells (6–9).

Defining the specificity of hemolysin transport has been elusive, however. Although the system apparently has one unique substrate, it is able to tolerate a wide variety of primary sequences. Extensive point mutations and minor deletions have been created in the signal sequence but few have had a dramatic effect on transport efficiency (7, 10). This has led to the proposal that there exists a handful of critical residues scattered throughout the signal sequence that are specifically recognized by the complex to trigger the transport process (11, 12). Thus, it is assumed that the other amino acids in the signal sequence are able to accommodate a wide range of changes.

At the same time, work done in our laboratory has demonstrated that the signal sequence of hemolysin can be replaced by that of *Pasteurella hemolytica* leukotoxin and still retain wild-type secretion, providing evidence that the two C-terminal signal peptides are functionally equivalent (13). Comparison of the two signal sequences revealed that they share very little primary sequence homology; however, there appears to be a common predicted helix-linker-helix motif (Fig. 1). This structure has been confirmed by circular dichroism (9) and 15N nuclear magnetic resonance studies (14). Both of these peptides exhibit similar biophysical properties in that they appear unstructured in an aqueous environment, but they assume an amphiphilic helical secondary structure under certain membrane mimetic environments. A similar motif has also been reported for the C-terminal 56 residues of *Erwinia chrysanthemi* protease G (15), which issecreted by a transport system analogous to the hemolysin system. This conservation of secondary structure among different organisms has led to the hypothesis that the helix-linker-helix structural motif within the signal sequence may be a prerequisite for transport. Such a secondary structure may be required for the signal sequence to
**EXPERIMENTAL PROCEDURES**

**Bacterial Strains and Plasmids**—The **E. coli** strain Top10F' (F' lacI, Tn10 (TetR) merA Δmrr-hsdRMS-merC) Δ88I lacZAM15 ΔlacX74 deoC recA araD139 Δara-lav7697 galU galK rpsL(str8) endA1 supG (Invitrogen), DH5α (supE44 lacZΔM15 (Δ88I lacZAM15) hsdS recA1 endA1 gyrA96 thi-1 relA1), and JM38 (λ araD30lac-pro-lac rpsL thi Δ80 ilvD ΔlacI58) were used for all cloning procedures. pUC494, pUC494BN, and pUC494-adaptor all encode hemolysin. pUC494 was used for the generation of the α1 and α2 random libraries. pUC494BN and pUC494-adaptor, the two derivatives of pUC494, were used for the construction of the linker random library and the α1-amphiphilic helical library, respectively. All pUC494 variants were selected with ampicillin at 50 μg/ml. **E. coli** harboring pLGBCD was transformed with mutated versions of pUC494 and assayed on blood agar plates. pGBCD, which contains hlyB, hlyD, and hlyC, was selected with chloramphenicol at 34 μg/ml.

**Construction of Plasmids**—The pUC494 plasmid containing the **hlyA** gene was constructed as follows: pUC19 was digested with SalI and HindIII, blunt end ligated, and then digested with BanHI and KpnI to facilitate the directional insertion of a BanHI-KpnI fragment (2.9 kilobases encoding **hlyA** from pLG58dc8/02) (9).

The plasmid pLGBCD containing hlyB, hlyD, and hlyC was constructed as follows: pLG575 (16), a pACYC derivative encoding HlyB and HlyD, was digested with SalI, and the ends were fixed with Klew. The blunt ended product was subsequently ligated with fragments from PruU-digested pLG570 (16), a plasmid that contains hlyA, hlyB, hlyC, and hlyD, and transformed into cells harboring pUC494. Plasmid sequences were obtained from transgenic and (thus contain a full complement of the hemolysin genes) and re-transformed into E. coli to facilitate the isolation of pLGBCD.

Random oligonucleotide mutagenesis involved insertion of a cassette containing random sequence into the coding region of the signal sequence at two flanking restriction sites. For practical reasons, we limited ourselves to assemble cassettes of 185 bp or less (maximum length of each oligonucleotide synthesized by Life Technologies, Inc. is 100 bp, 15 bp of overlap for annealing. This implies that the distance between the two restriction sites must be less than 185 bp.) To meet this requirement in the construction of linker random library, two sites were engineered within pUC494 to create pUC494BN.

Construction of pUC494BN from pUC494 consisted of two steps. The first step involved the creation of a BstBI restriction site in the α1 region of the hemolysin signal sequence. Primers AsalC494 (5'-TGATCTGACCACTTCTTGAAACGTTACGGC-3') and HlyA-BstBI-R (5'-CAGCTGAAATGTGTTGCAAATTTAATATGTAAGCAGC-3') were used to generate a 97-bp fragment from pUC494. This PCR product was used directly as a primer, in conjunction with M13R1 (5'-AAAAGGACGGCCAGGTAATTC-3') to amplify a 159-bp fragment from pUC494. The resulting reaction product was purified with QIAquick PCR purification kit (Qiagen), then digested with KpnI and SalI, and finally inserted into the pUC494 at these two restriction sites to generate pUC494A.

The second step involved creation of an Nhel site in the α2 region of pUC494B. Primers HlyA-Nhel-F (5'-GAAAGATCTCCGCGCTAGCTTATTGACATGTTCTCC-3') and HlyA-Nhel-R (5'-GGAAGATCTCCGCGCTAGCTTATTGACATGTTCTCC-3') were used to generate a 97-bp fragment from pUC494B. This reaction product was purified with a QIAquick PCR purification kit, and ligated into the KpnI and BglII sites of pUC494B. The resulting vector, pUC494BN, differed from pUC494 in that it contained two additional unique restriction sites, BstBI (in the α1 region) and Nhel (in the α2 region). These sites permitted assembly of the signal amide acid sequence remained unaltered. The hemolytic zone of pUC494B and pUC494BN were determined to be identical to that of pUC494A.

The plasmid pUC494-adaptor was generated by replacing 89 bp of the hemolysin signal sequence with a 27-bp adapter. The two oligonucleotides, forward adapter (5'-TGCCGCGCGCCGCGTTTTAATGAT-GCAG-3'), underlined region represents complementary sequence required for transport, while the amphiphilic nature of the first helix (α1) is a critical determinant of function.
nucleotide restriction site to appear at the same position), amplification of a PCR fragment and subsequent digestion with a selected restriction enzyme would allow the selection of clones with a PCR product that cannot be cleaved, and thus likely contain a random sequence. For the α1, linker, and α2 random libraries, a 579-bp fragment was amplified from each colony using primers HA24 (5'-AGGAGGAAATCTTNNNNNNNNNNNNNNNNNNNNTGATTAGATGTCGCTC-3') and M13R1, and then subjected to digestion with PstI, KpnI, and HpaII, respectively. All positive colonies were grown overnight and plasmid DNA of each variant was prepared with the Quantum miniprep kit (Bio-Rad) following the manufacturer's instructions. DNA sequencing was performed on each variant to obtain the genotype.

Cloning of Amphiphilic Helical Variants—The cloning procedure of α1-amphiphilic helical variants was slightly different from that of random variants. Four oligonucleotides were mixed at equimolar concentrations and incubated at 95 °C for 3 min, then left to cool at room temperature. The annealed oligonucleotides were incubated at 70 °C for 1 h with Taq DNA polymerase, purified with a Qiagen II Gel Isolation Kit (Qiagen), and then digested with BglII and Sall. The double-stranded, digested cassette was purified using the QIAquick Nucleotide Removal Kit (Qiagen). Table I illustrates the cloning plasmid, restriction sites, and oligonucleotide sequences for the construction of the amphiphilic helical library. The ligation product was transformed into JM83 cells containing pLBGCD and plated on blood agar plates.

All transformants were picked on a plate regardless of hemolytic zone and subjected to colony PCR. Unlike the random variants, selection of α1-amphiphilic helical variants was based on the size of PCR fragment rather than the absence of a restriction site. This was facilitated by the use of pUCAC494-adaptor as a cloning vector, which contained a 63-bp deletion within the signal sequence coding region. Insertion of an α1 cassette into this plasmid would restore its length to α1 size. The primers HA23 (5'-GACGGCAGGTAATCACA-3') and M13R1 were used to amplify a 411-bp fragment from a positive candidate and a 348-bp fragment from a negative clone. All colony forming units on a given plate were analyzed by PCR, regardless of the size of the hemolytic zone in order to generate an unbiased library of helices. Upon identification of amphiphilic helical clones, the plasmid DNA was isolated (mixture of pUCAC494 and pLBGCD) and re-transformed into JM83 selected with ampicillin in order to obtain a pure source of plasmid for DNA sequencing.

DNA Sequencing—Plasmid DNA was quantitated with the use of an SSF-600 solid state fluorimeter (Tyler Research Instruments Corp.) by the ethidium bromide fluorescence assay described in the instruction manual. DNA samples were prepared with ABI PRISM BigDye terminator cycle sequencing ready reaction kit according to manufacturer's instructions, and analyzed on a 310 Genetic Analyzer (PE Biosystems). The forward sequence was obtained for all variants using sSEQ (5'-GACGGCAGGTAATCACA-3') as primer. The reverse sequence of selected variants was obtained using M13R1.

Blood Agar Plate Assay—The hemolysin secretion level of each variant was determined using blood agar plate assay. Five different plating conditions were tested: 20 ml of LB agar with 1, 2, or 5% defibrinated sheep blood (PML Microbiologicals), or two-layered plates with 10 ml of plain LB bottom agar and 10 ml of 2 or 5% blood LB top agar. In general, high secretors could be resolved better on high percent blood plates. Plates with double layers were found to provide a better contrast. The optimal condition was established to be 10 ml of LB bottom agar with 10 ml of 5% blood LB top agar, and these plates were used to determine the phenotype of all clones.

Since an average of 100 variants were assayed for each library at the same time, the blood agar plate assay proved to be the only feasible method for phenotype determination because it is quick and convenient. Plasmid DNA isolated from each variant was transformed into JM83 E. coli containing pLBGCD and spread on blood agar plate in triplicate. After an incubation period of 19 h at 37 °C, each variant was assayed as a zone range from 0 (no hemolysin) to 6 (wild-type) by comparing to a set of standards. To control for lysis, cells transformed with pUCAC494 and pLBGCD (i.e. without HlyB) had a zone size of zero (no secretion), and were identical to cells transformed with pLBGCD only (i.e. without HlyA). Attributes such as hemolytic zone size, brightness, and colony size were all taken into account. All plates were examined twice separately. Altogether, six readings were taken for each variant. The average and standard deviation were calculated to provide an indication of the secretion level and variability, respectively. Assignments were most consistent at the two extremes (0, 1, and 6), with the variability greatest for clones secreting at levels 3 and 4. It should be emphasized that the scale for zone assignment (rank 0 to 6) was non-linear, and that the numbers obtained were only semiquantitative.

All variants within one library were assayed the same day to minimize any inconsistencies. In addition, selected variants from each of the three random libraries, as well as all amphiphilic helical variants, were plated together in a final round to facilitate a direct comparison between the different libraries. To determine the reproducibility of this assay, averages taken on the first and second occasions were analyzed. Out of a total of 54 random variants, 11 were given the same assignments as before. Thirty-nine mutants had averages that differed by less than one, and five other variants had averages that differed by more than one. The Pearson correlation coefficient between the two sets of measurements was 0.96 (data not shown).

Data Analysis—DNA sequences were translated and analyzed with the Wisconsin Package® Version 9.1, Genetics Computer Group (GCG), Madison, WI. Additional Perl scripts were written by Dr. Eric Cabot and David Hui to facilitate bulk analysis. Peptool version 1.1 (Biotools Inc., Edmonton, Canada) was used for helix hydrophobic moment determination.

SDS-Polyacrylamide Gel Electrophoresis and Western Blotting—To determine the amount of endogenous hemolysin, JM83 bacteria were transformed with plasmids encoding the hemolysin variants and harvested at A600 = 0.85 ± 0.05. Following centrifugation (7,000 × g for 15 min), cell pellets were resuspended in STE buffer (10 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA) supplemented with various protease inhibitors. For each sample, an equivalent of 200 μl of cells was boiled for 3 min and ran on an SDS-polyacrylamide gel (7.5% separating) under

| TABLE I | Sequence of oligonucleotides used to assemble combinatorial cassettes |
|---------|---------------------------------------------------------------------|
| α1 random cassette: inserted into pUCAC494 at the SalI and BglII sites |
| HlyA-α1-F: 5'-GGGAAGTACCCGTCACATATTGACCCAGAATCTTATA-3' |
| HlyA-α1-R: 5'-GGCGAATCTTCCCTTTATTACAGAAGCTACCC-3' |
| Linker random cassette: inserted into pUCAC494B at the BetB1 and NheI sites |
| HlyA linker-F: 5'-CACTTAATATGAACTTTGCAAGAATCCATGTCGCA-3' |
| HlyA linker-R: 5'-GGGACACTGCAATATTGAGCCGCCCCGCC-3' |
| α2 Random cassette: inserted into pUCAC494 at the BglII and KpnI sites |
| HlyA-α2-F: 5'-GGGGAAGATCTTCCCTTTATTACAGAAGCT-3' |
| HlyA-α2-R: 5'-GACGTTGCTACATTAGTCTGACCAATATTGACGAAATCC-3' |
| α1 amphiphilic helical cassette: inserted into pUCAC494-adaptor at the SalI and BglII sites |
| Oligo A: 5'-ATGATGCTCGAATATTGACGACAACTTATGAAATCC-3' |
| Oligo B*: 5'-AAAATGAAATCTTACAAATCCATATGAAATCC-3' |
| Oligo C: 5'-GGGAAGATCTTCCCTTTATTACAGAAGCT-3' |
| Oligo D: 5'-GGCGAATCTTCCCTTTATTACAGAAGCT-3' |

* B, g/a/c; 3c, a/t/c; Bc, g/t/c.
removal of the supernatant, anti-hemolysin antiserum (diluted 5,000

e-arginine and 0.05% Tween 20 and incubated at 37 °C for 2 h. Upon

pre-blocked with phosphate-buffered saline containing 1% bovine se-

applied to Immulon-2 microtiter plates (Dynatech Labs) that had been

Centricon-30 filtration columns (Amicon). The concentrate was then

replaced by goat anti-rabbit antibody (diluted 10,000

e-arginine albumin and 0.05% Tween 20) and incubated at room temperature for 2 h. Following this, the primary antibody was

instructions.

the ABTS (Sigma) method was used following the manufacturer’s

Tween 20) and incubated at room temperature for 1 h. For detection,

results (Table II). Based on \( \chi^2 \) statistics, it was deter-

reducing condition. After standard Western transfer and blotting pro-

cedure (with anti-hemolysin antiserum at 20,000

e-arginine antibody at 10,000 \times \) dilution and goat

anti-rabbit antibody at 10,000

\( \times \) dilution and goat

3

With the exception of Ser and Thr, all polar amino acids were represented by the
degenerate codon BaN (where B = g/a/t/c). Ser and Thr residues were substituted by the codon aBN. The 11th and 12th amino acids in the a1 helix

are alanines, and were replaced by Jct (where J = g/a/t/c), and gBcBc (where Bc = g/t/c), respectively, in order to maintain the amphiphilicity of the

het.

TABLE II

Design of a1 random and amphiphilic helical variants

| Wild-type a1 helix | Random variants | Helical variants |
|-------------------|-----------------|-----------------|
| DNA | DNAa | Amino acid | DNAa | Amino acid | DNA | Amino acid |
| Amino acid | tta | att | aat | gaa | atc | age | aat | ate | atc | att | tea | gct | gca |
| L | I | N | E | I | S | K | I | S | A | A |
| Random variants | X | X | X | X | X | X | X | X | X | X |
| DNAc | DNA | Amino acid | DNAc | Amino acid |
| N | T | N | BaN | BaN | aBN | BaN | N | N | T | N |
| T | M | F | F | K | N | M | F | N | K | M | F | N | K |
| H | V | V | V | V | R | R |
| Helical variants | X | X | X | X | X | X | X | X | X | X |

\( ^{a} \) N represents any of the 4 nucleotides (g/a/t/c).

\( ^{b} \) X represents any of the 20 amino acids or the stop codon.

\( ^{c} \) N-polar amino acids utilized the degenerate codon NtN. With the exception of Ser and Thr, all polar amino acids were represented by the
degenerate codon BaN (where B = g/a/t/c). Ser and Thr residues were substituted by the codon aBN. The 11th and 12th amino acids in the a1 helix

are alanines, and were replaced by Jct (where J = g/a/t/c), and gBcBc (where Bc = g/t/c), respectively, in order to maintain the amphiphilicity of the

helix.

reducing condition. After standard Western transfer and blotting pro-

cedure (with anti-hemolysin antiserum at 20,000

e-arginine antibody at 10,000 \( \times \) dilution and goat

anti-rabbit antibody at 10,000

\( \times \) dilution and goat

\( ^{b} \) N-polar amino acids utilized the degenerate codon NtN. With the exception of Ser and Thr, all polar amino acids were represented by the
degenerate codon BaN (where B = g/a/t/c). Ser and Thr residues were substituted by the codon aBN. The 11th and 12th amino acids in the a1 helix

are alanines, and were replaced by Jct (where J = g/a/t/c), and gBcBc (where Bc = g/t/c), respectively, in order to maintain the amphiphilicity of the

helix.
protein increased with the hemolytic zone. Thus, the size of the halos could be used as an indicator for secretion efficiency. In this study, a hemolytic zone assignment of 4 (~50%) or above was considered to be efficient transport.

Thirty-four α1 random variants were generated to investigate the functional role of this region. If the α1 amphiphilic helix plays a critical role in transport, most α1 random mutants will be expected to be secreted at low levels since they will not contain this specific structural feature. However, if the α1 region does not contain any critical elements, then most of the random variants will be secreted at a level equivalent to wild-type. The hemolytic zones for α1 random mutants ranged from 2 to 3.5. The population was relatively homogenous, with a mean and a median of 2.8, suggesting that secretion was greatly hampered (Fig. 4A). Since the rest of the signal sequence was intact, any changes in hemolysin transport could be attributed to modifications in the α1 region. More specifically, the dramatic reduction in secretion indicated that some feature(s) essential for efficient transport must be located in this region.

A random library was also created in the 10-amino acid region just downstream of the α1 helix (linker). The hemolytic zones for the 45 linker random variants ranged from 1.3 to 4, with a mean of 2.4, and a median of 2.3. Secretion in these linker variants was obviously reduced despite the presence of an intact α1 helix (Fig. 4B). The dramatic decrease in transport upon random mutagenesis of this region suggested that the linker, like the α1 region, also contains some important elements that may be required for efficient transport.

The secretion pattern of α2 random variants was expected to resemble that of the α1 random variants since both regions are conserved. However, examination of the 32 mutants from the α2 random library revealed a different distribution. The hemolytic zones for α2 random mutants ranged from 1 to 6, with a mean of 5.1, and a median of 5.5. Twenty-four of the 32 full-length mutants secreted at 5 or higher, with only one transporting at lower than 2.5 (Fig. 4C). This surprising distribution provided strong evidence that the α2 region can tolerate almost any combination of amino acids without having a major effect on transport.

**Secretion Efficiency of α1-Amphiphilic Helical Variants**—Results from our α1 random library suggested that the α1 region contains features important for recognition and transport. As a result, an α1-amphiphilic helical library was generated to determine if an amphiphilic helix in this region could support efficient transport regardless of primary sequence. This approach was based on a strategy described by Kamtekar et al. (17) in which the sequence locations of polar and non-polar residues were specified explicitly, but the precise identities of the side chains were not constrained and varied extensively (Table II). By maintaining the periodicity of polar and non-polar residues within the α1 region the same as that of wild-type, all amphiphilic helical variants were predicted to retain an amphiphilic helical structure. A total of 1.5 × 10^8 different amino acid sequences were theoretically possible in this library.

Examination of the 22 α1-amphiphilic helical variants allowed us to establish a relationship between structure and function in the α1 region. If an amphiphilic helical structure in the α1 region is sufficient for transport, then most α1-amphiphilic helical variants would be secreted at high levels since they are all predicted to contain this specific structural feature. The hemolytic zones for α1-amphiphilic helical mutants ranged from 1.3 to 5.7, with a mean of 4.0 and a median of 4.3. Only one variance.

### TABLE IV

| Amino acid | α1 Random library | Linker Random library | α2 Random library |
|------------|-------------------|-----------------------|-------------------|
|            | Observed %        | Predicted %           | Observed %        | Predicted %           | Observed %        | Predicted %           |
| G          | 4.41              | 4.27                  | 10.17             | 9.27                  | 2.82              | 3.89                  |
| A          | 4.90              | 5.15                  | 10.51             | 10.49                 | 2.56              | 2.56                  |
| V          | 3.19              | 3.76                  | 2.54              | 3.92                  | 7.04              | 7.76                  |
| L          | 6.00              | 6.41                  | 6.61              | 5.31                  | 12.55             | 12.49                 |
| I          | 4.29              | 5.23                  | 1.86              | 1.99                  | 8.07              | 8.82                  |
| P          | 6.99              | 6.21                  | 11.36             | 11.88                 | 1.92              | 1.69                  |
| F          | 1.23              | 1.42                  | 2.03              | 0.79                  | 8.96              | 8.11                  |
| W          | 0.86              | 0.78                  | 1.69              | 1.19                  | 2.18              | 1.53                  |
| Y          | 2.82              | 2.84                  | 1.86              | 1.35                  | 4.48              | 5.75                  |
| T          | 8.95              | 9.03                  | 7.29              | 7.65                  | 4.10              | 3.63                  |
| S          | 9.68              | 7.76                  | 7.46              | 7.64                  | 6.91              | 8.00                  |
| C          | 1.84              | 1.62                  | 1.02              | 1.86                  | 5.89              | 4.06                  |
| M          | 1.59              | 1.36                  | 0.85              | 0.57                  | 2.94              | 2.17                  |
| Q          | 5.51              | 5.14                  | 4.07              | 4.03                  | 1.92              | 1.73                  |
| N          | 5.88              | 5.66                  | 2.20              | 2.33                  | 3.46              | 4.08                  |
| D          | 2.45              | 3.23                  | 2.54              | 3.20                  | 3.46              | 2.88                  |
| E          | 2.82              | 4.26                  | 4.24              | 3.56                  | 2.43              | 2.62                  |
| K          | 8.21              | 7.47                  | 2.88              | 2.60                  | 5.12              | 3.72                  |
| R          | 8.46              | 9.41                  | 12.54             | 14.05                 | 3.46              | 5.19                  |
| H          | 3.31              | 3.89                  | 3.73              | 3.62                  | 1.92              | 1.90                  |
| Stop codon | 6.62              | 5.11                  | 2.54              | 2.38                  | 7.81              | 7.40                  |

Sample size: 816, 590, 781

p value: 0.49, 0.21, 0.08

- The observed amino acid frequencies were determined for all random and stop variants within each library.
- The predicted amino acid frequencies were calculated based on the observed nucleotide frequencies from Table III.
- The sample size equals to the total number of random and stop variants multiplied by the number of amino acids in each target region.

**Fig. 3. Hemolytic zone size.** *E. coli* JM83 bacteria harboring the plasmid pLGBCD, which encodes HlyB and HlyD (transporters) as well as HlyC (toxin activation), were transformed with plasmid DNA encoding a hemolysin signal sequence variant. Transformants were plated on blood agar plates in triplicate and incubated at 37 °C for 19 h. Each plate was read twice independently by comparing to a set of standards and assigned a rank from 0 (no transporter) to 6 (wild-type activity). A total of 6 readings for each variant was averaged. The percentage of secretion relative to wild-type for each zone size was approximately: rank 6, 100%; rank 5, 90%; rank 4, 50%; rank 3, 30%; rank 2, 10%; rank 1, 2%; rank 0, 0%.
proposed critical residue, Glu-46, has been reported in this region (11). However, we have isolated a number of amphiphilic helical mutants that do not contain this residue at this position but are still able to support efficient transport, suggesting that this residue may not be essential. Rather, the finding that the average secretion level of α1-amphiphilic helical mutants was significantly higher than α1 random mutants (which had a mean secretion of 2.8) strongly suggests that an amphiphilic helical structure in the α1 region is supportive of efficient transport (Fig. 4D).

Helix Hydrophobic Moment and Hydrophilicity of Random and Amphiphilic Helical Variants—The presence or absence of important elements with each region can be determined by the secretion distribution of the corresponding random variants. Once a region is found to be important, the actual feature(s) can be identified by correlating the genotype of the combinatorial population to the phenotype. To achieve this end, a number of biophysical properties, including helix hydrophobic moment and Kyte-Doolittle hydrophilicity, were calculated for the mutated region of each random and amphiphilic helical mutant and analyzed in the context of transport efficiency (Table V).

The helix hydrophobic moment is an indicator of amphiphilic helix formation (18). This value was relatively low for random mutants (average of 0.14) compared with wild-type (0.29), suggesting that most random variants do not contain an amphiphilic helical structure. The Spearman correlation coefficient between secretion and helix hydrophobic moment was 0.38 (p < 0.05) for α1 random variants. This weak degree of correlation implied that an amphiphilic helix in the α1 region contributes partly to efficient transport.

The average hydrophilicity differed substantially among the three random libraries. This could be a result of the uneven distribution of nucleotides. Nevertheless, there existed a wide range of hydrophilicity values within each library. The Spearman correlation coefficient between secretion and hydrophilicity for α1 random variants was relatively strong (−0.53, p < 0.01), providing some evidence that hydrophilicity in the α1 region may be important for efficient transport.

When the helix hydrophobic moment and hydrophilicity were calculated for the linker and α2 random variants, no relationships were observed between these values and secretion level. This was expected for α2 random mutants since sufficient secretion was obtained regardless of the primary sequence. In the case of the linker region, it appeared that helix hydrophobic moment and hydrophilicity were not particularly important for transport.

The average helix hydrophobic moment for α1-amphiphilic helical mutants was 0.25 (Table V). This supported that the amphiphilic helical design indeed yielded variants with the intended structure. Interestingly, when the secretion level was correlated with the helix hydrophobic moment for α1-amphiphilic helical variants, no specific relationship was observed. It was possible that the helix hydrophobic moment was saturated, and that a stronger amphiphilic helix might not necessarily improve transport. Consistent with results from the α1 random variants, there also appeared to be some negative correlation (−0.38) between hemolytic zone size and hydrophilicity for the amphiphilic helical variants.

Secretion Efficiency of α1, Linker, and α2 Stop Variants—The stop mutants in each of the random libraries were also analyzed. These variants contain a stop codon in the targeted region, resulting in a truncated signal sequence with a random tail. For example, α1 stop mutants were identical up to but not including the α1 region, and the rest of the signal sequence (49 amino acids in this case) was substituted with a random string of zero to 11 amino acids. Analysis of stop mutants from different random libraries provided some insights into the functional properties of the extreme C terminus.

Thirty-four α1 stop variants and 15 linker stop variants were generated. Their secretion ranged from 1 to 3, with a mean of 1.9 and 1.7, respectively. Since the majority of the signal sequence was removed, it was not surprising to see the dramatic reduction in hemolysin secretion (Fig. 4, A and B). Rather, it was interesting to note that no null mutants were obtained despite the drastic deletions.

Examination of the 53 stop mutants in the α2 random library revealed a skewed distribution (Fig. 4C). The zone size for α2 stop mutants ranged from 1 to 4, with a mean of 1.5, and a median of 1. It should be pointed out that 13 out of 53 stop mutants had a stop codon at position 1 of the α2 region, and
thus share the same genotype. Phenotype determination by blood agar plate assay was consistent, they all had an assign-
ment of 1, with only a single exception (i.e., 1.2). The low level of secretion of α2 stop mutants suggested that the residues
downstream of the linker must contain some important ele-
ment(s) for efficient transport. The α2 region could be excluded
since it could tolerate almost any amino acids, thus it is likely
that some required element(s) lies within the region after the
second helix.

Interpretation of the stop mutants is complicated by the fact
that two parameters, length and composition of the extreme C
terminus, were changed concurrently. When the hemolytic
zone was plotted against the position of the stop codon for α1
and linker stop variants, no obvious trend was observed (data
not shown). However, when the same plot was performed for α2
variants, there appears to be some dependence on length (Fig.
5). Furthermore, when the length element was controlled by
comparing mutants with a stop codon at the same position,
those that contained positively charged residues between posi-
tions –2 to –8 consistently transported at levels lower than
their counterparts. Taken together, the results suggested that
the length and hydrophobicity of the last few amino acids could
be important.

DISCUSSION

The hemolysin transporter complex is dedicated to the active
export of the 107-kDa hemolysin directly from the cytoplasm to
the outside of the cell. Although this transporter system has a
single natural substrate, it is able to recognize and transport
many other different primary sequences, from heavily mutated
versions of hemolysin (7, 10) to protein toxins from other spe-
cies (13). To investigate the substrate specificity of the hemol-
ysin transporter complex, we have attempted to identify fea-
tures within the hemolysin signal sequence that are important
for transport. The conserved helix-linker-helix motif within the
targeting signal was divided into three regions (α1-linker-α2)
and subjected to random oligonucleotide mutagenesis. Using
this approach, we have shown that the α1 and linker regions
are sensitive to random changes in the primary sequence. For
the α1 region, it appears only sequences that result in a pre-
dicted amphiphilic helical structure and appropriate hydropho-
bicity can support efficient transport. In contrast, the α2 region
has no specific primary or secondary structural requirements.
This is a surprise since this region represents the second helix
of the conserved helix-linker-helix motif. While the α2 helix
may not be important for secretion, it is possible that this
structure is required for some yet unidentified function.

A slight variation in the design of combinatorial library
allowed us to investigate the structural requirements of the
signal sequence at a higher resolution. We have created an
amphiphilic helical library to determine the contribution of
secondary structure to transport in the absence of other fea-
tures in the α1 region. Our study revealed that any one of
a large number of sequences that yielded a predicted amphiphilic
helix in this region is sufficient for transport, demonstrating
that there is no specific requirement at the primary sequence
level. This is consistent with past reports that the α1 region
can tolerate a wide variety of point mutations (10, 11, 19). Based
on our data, it is believed that the higher-order structure is more
important for function than primary sequence.

In essence, results from the α1 combinatorial libraries dem-
onstrated that while the hemolysin transporter system can
accommodate many variations of the substrate, certain restric-
tions apply. This observation seems to hold for the linker region
as well. While the leukotoxin signal sequence could be trans-
ported efficiently by the hemolysin transporter system despite
a lack of sequence similarity in the equivalent linker region,
most hemolysin linker random variants were not secreted at
high levels. By analogy to the α1 region, the important features
for transport in the linker region are likely to be some occult
elements buried within the primary sequence, rather than the
specific identity of individual residues.

Determination of the identity of the functional features
within the hemolysin signal sequence provides the first step in
understanding the broad substrate specificity of hemolysin
transporter complex. In the case of the α1 region, the conserved
amphiphilic helical structure was found to be important for
function. Distribution of the 22 amphiphilic helical mutants
points to the likelihood that as many as 1.5 × 10⁸ different
variants of the α1 region could be transported at greater than

### TABLE V

| Helix hydrophobic moment | Range | Average | Correlation |
|--------------------------|-------|---------|-------------|
| α1 Random                | 0.08 to 0.21 | 0.14  | 0.38*       |
| Linker random            | 0.05 to 0.23 | 0.12  | 0.04        |
| α2 Random                | 0.09 to 0.23 | 0.14  | 0.10        |
| Helical                  | 0.21 to 0.30 | 0.25  | −0.09       |

| Kyte-Doolittle hydrophilicity | Range | Average | Correlation |
|-------------------------------|-------|---------|-------------|

- a The helix hydrophobic moment was calculated using Peptool, v.1.1. A window of 9 residues was used, resulting in a value for each residue within the mutated signal sequence.
- b The Kyte-Doolittle hydrophilicity was determined with Wisconsin Package™ Version 9.1, Genetics Computer Group (GCG). A window of 7 residues was used, resulting in a value for each amino acid within the mutated signal sequence.
- c The helix-hydrophobic moment or hydrophilicity values for residues within the mutated region of each mutant were averaged. Following that, a mean was obtained for each library.
- d The average helix-hydrophobic moment and hydrophilicity values for each mutant were correlated with the hemolytic zone size. The values shown above represent the Spearman correlation coefficients.
- e Significant at the 0.05 level (2-tailed).
- f Significant at the 0.01 level (2-tailed).

**FIG. 5.** Effect of changing the length and amino acid composition of the C terminus on secretion as demonstrated by α2 stop mutants. The position of the stop codon of 39 α2 stop mutants is plotted against hemolytic zone size. Note that the random region becomes the extreme C terminus in stop mutants. For instance, α2 variants that have a stop codon at position 11 contain a 10-amino acid random tail. The probability of secretion appears to increase with length. Furthermore, when mutants of the same length were compared, those with a relatively hydrophobic tail consistently secreted at higher levels.
50% of wild-type. It is intriguing that the hemolysin transporter system can tolerate such a large number of primary sequences. If secondary structure is the only functional requirement of the α1 region, what role does it play in transport? Two mechanistically distinct, but not mutually exclusive, planar models exist. It has been proposed that the α1-amphiphilic helical structure may be involved in interacting with the membrane (19). This allows hemolysin to diffuse in a two-dimensional plane rather than a three-dimensional space, greatly enhancing the chance of substrate binding to the transporter complex. If the functional role of the α1 region is limited to the membrane level, primary sequence does not have to be specific as long as an amphiphilic helix is present.

It is also possible that the α1 region is directly involved in interacting with the transporter complex. In this case, the amphiphilic helical structure may be required to present residues within the α1 region in the proper orientation to the binding pocket of the transporter. The ability of the transporter complex to recognize many different primary sequence variations can be explained if this binding pocket contains multiple “contact” residues, all of which could act as potential “docking” sites for residues within the signal sequence. Interaction with any one of several possible combinations of these contact residues would trigger the transport process. Following this rationale, hemolysin and leukotoxin, two proteins that share little primary sequence similarity but which are secreted at equally high efficiency, may actually interact with different sets of contact residues within the binding pocket. Indeed, two point mutants have been isolated in the hemolysin ABC transporter, HlyB, in which the transport of hemolysin was unaffected but the secretion of leukotoxin was reduced (20, 21). This observation could be explained if the two mutated residues specifically recognize leukotoxin and not hemolysin. The large number of specific regional mutants generated in this study could also be used for further genetic complementation analyses to elucidate the intricate nature of this transporter-substrate interaction.

HlyB is a member of the ABC transporter superfamily, which is responsible for the translocation of many important substrates across biological membranes. Interestingly, many ABC proteins, such as P-glycoprotein and multidrug resistance protein, exhibit broad substrate specificity. Although the hemolysin transporter system is dedicated to the translocation of one natural substrate, our study clearly demonstrates that it can recognize and transport a wide range of signal sequence variants. This raises the possibility that multiple substrate specificity could be an intrinsic property of ABC transporters, and thus the principles of transport proposed for the hemolysin system could be applied in a more general sense. For example, substrates of P-glycoprotein are relatively amphiphilic or hydrophobic in nature. This property is believed to facilitate their partitioning into the plasma membrane before interacting with the transporter (22). The high substrate concentration in the membrane environment may help to overcome the specificity issue. It has also been suggested that there are multiple functionally distinct but potentially overlapping binding sites within P-glycoprotein that are responsible for drug efflux (23).

This is a parallel interpretation to the multivalent binding pocket of the hemolysin system, in which different substrates may interact with different sets of contact residues to promote transport.

In a broader context, combinatorial analysis provides a powerful tool to investigate regions of interest in other biological systems. Random oligonucleotide mutagenesis allows drastic alterations to be engineered into a well defined region while maintaining the natural spatial relationships of the target and non-target regions. Furthermore, a large number of variants can be generated, allowing for a greater degree of confidence in the interpretation of results. The distribution of variants can provide a clear-cut answer regarding the presence of any important features within the target region. Once a region is determined to be important, more in-depth correlation analysis can be carried out on the mutant population to objectively identify the exact features involved. Furthermore, as demonstrated by our helical library, a more specialized design of combinatorial library allows us to dissect primary and secondary structural elements within the same region, and to evaluate them separately for their biological contributions.

**Acknowledgments**—We are grateful to Dr. Yan Li, Angie Ma, and Stephen Ralph for contributing their time to assist in this study. We thank Drs. Mark Blight and Barry Holland who generously provided the anti-hemolysin antiserum, Dr. Eric Cobat for help in programming, and Dr. Nhu Le for assistance in statistical analysis. We also thank Drs. Jonathan Sheps and Douglas Hogue for critical review of this manuscript.

**REFERENCES**

1. Childs, S., and Ling, V. (1994) in *Important Advances in Oncology* (DeVita, V. T., Hellman, S., and Rosenberg, S. A., eds). pp. 21–36, J. B. Lippincott Co., Philadelphia.
2. Dean, M., and Allikmets, R. (1995) *Curr. Opin. Genet. Dev.* 5, 779–785.
3. Chan, H. S., DeBoer, G., Thorner, P. S., Haddad, G., Gallie, B. L., and Ling, V. (1994) *Hematol. Oncol. Clin. North Am.* 8, 383–410.
4. Sheps, J., Zhang, F., and Ling, V. (1996) in *Membrane Protein Transport*, (Rothman, S., ed) Vol. 5, pp. 81–118, J Press Ltd., London.
5. Jarcho, T., Chakraborty, T., Garcia, F., and Goebel, W. (1994) *Mol. Gen. Genet.* 245, 53–60.
6. Hess, J., Gentschev, I., Goebel, W., and Jarcho, T. (1990) *Mol. Gen. Genet.* 224, 201–208.
7. Koronakis, V., Koronakis, E., and Hughes, C. (1989) *EMBO J.* 8, 595–605.
8. Mackman, N., Baker, K., Gray, L., Haigh, R., Nicaul, J. M., and Holland, I. B. (1987) *EMBO J.* 6, 2835–2841.
9. Zhang, F., Yin, Y., Arrowsmith, C. H., and Ling, V. (1995) *Biochemistry* 34, 4193–4201.
10. Chervaux, C., and Holland, I. B. (1996) *J. Bacteriol.* 178, 1232–1236.
11. Kenny, B., Taylor, S., and Holland, I. B. (1992) *Mol. Microbiol.* 6, 1477–1489.
12. Kenny, B., Chervaux, C., and Holland, I. B. (1994) *Mol. Microbiol.* 11, 99–109.
13. Zhang, F., Greig, D. I., and Ling, V. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 4211–4215.
14. Yin, Y., Zhang, F., Ling, V., and Arrowsmith, C. H. (1995) *FEBS Lett.* 366, 1–5.
15. Wolf, N., Ghiyo, J. M., Delepelaire, P., Wandersman, C., and Delepierre, M. (1994) *Biochemistry* 33, 6792–6801.
16. Mackman, N., Nicaul, J.-M., Gray, L., and Holland, I. B. (1985) *Mol. Gen. Genet.* 201, 282–288.
17. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) *Science* 262, 1680–1685.
18. Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 140–144.
19. Stanley, P., Koronakis, V., and Hughes, C. (1991) *Mol. Microbiol.* 5, 2391–2403.
20. Zhang, F., Sheps, J. A., and Ling, V. (1993) *J. Biol. Chem.* 268, 19889–19895.
21. Sheps, J. A., Cheung, J., and Ling, V. (1995) *J. Biol. Chem.* 270, 14829–14834.
22. Shapiro, A. B., and Ling, V. (1998) *Eur. J. Biochem.* 254, 181–188.
23. Shapiro, A. B., and Ling, V. (1997) *Eur. J. Biochem.* 250, 130–137.